SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2012-2.3.1 – Third Implementation Phase of the European High Performance Computing (HPC) service PRACE

PRACE-3IP

PRACE Third Phase Implementation Phase Project

Grant Agreement Number: RI-312763

D8.1.1-resubmit
Technical Specifications for the PCP and for Phase 1

Final

Version:	2.2
Author(s):	Dirk Pleiter, JUELICH
	Marcus Richter, JUELICH
	Mark Parsons, EPCC
Date:	24.09.2013

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: RI-312763**	
	Project Title: PRACE Third Phase Implementation Phase Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: D8.1.1	
	Deliverable Nature: Report	
	Deliverable Level: PU	**Contractual Date of Delivery:** 30 / September /2013
		Actual Date of Delivery: 30 / September /2013
	EC Project Officer: Leonardo Flores Añover	

* - 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: Technical Specifications for the PCP and for Phase 1**	
	ID: D8.1.1	
	Version: <2.2 >	**Status:** ***Final***
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2007	
	File(s): D8.1.1-resubmit.docx	
Authorship	**Written by:**	Dirk Pleiter, JUELICH Marcus Richter, JUELICH Mark Parsons, EPCC
	Contributors:	Pascal Moussier, GENCI Stephane Requena, GENCI François Robin, CEA Florian Berberich, JUELICH Philippe Segers, GENCI
	Reviewed by:	Aleksandar Belic, IPB; Dietmar Erwin, FZJ
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	08/November/2012	Draft	Outline of the document
0.2	03/December/2012	Draft	Revised draft
0.7	18/December/2012	Final draft	Almost complete
1.0	20/December/2012	Final version	Complete
1.1	14/August/2013	Draft	First draft of revised document
2.0	09/September/2013	Final revised version	Complete
2.1	18/September/2013	Comments of EC reviewers incorporated	Complete
2.2	24/September/2013	Comments of François Robin included	Complete

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Pre Commercial Procurement

Table of Contents

List of Tables

List of Figures

References and Applicable Documents

[1] European Commission PCP website http://cordis.europa.eu/fp7/ict/pcp/home_en.html

[2] International Exascale Software Programm (IESP) http://www.exascale.org/mediawiki/images/2/20/IESP-roadmap.pdf

[3] European Exascale Software Initiative http://www.eesi-project.eu/pages/menu/publications/final-report-recommendations-roadmap.php

[4] DARPA HPCS Programme http://www.darpa.mil/Our_Work/MTO/Programs/High_Productivity_Computing_Systems_(HPCS).aspx

[5] Legal basis for PCP http://www.ertico.com/assets/Activities/P3ITS/P3ITS-D2.1-Analysis-of-public-Pre-Commercial-Procurementv1.8.pdf

[6] www.green500.org, June 2013

[7] C. Bekas and A. Curioni, *A new energy aware performance metric*, Comput. Sci. Res. Dev. (2010) 25, pp. 187-195

[8] www.deep-project.eu

[9] www.montblanc-project.eu

[10] Peter Kogge et al., *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems*, 2008, http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf

[11] Rob Schreiber et al., *Exascale Technology Roadmap Meeting. Node Architecture and Power Group*, 2009.

[12] PRACE-PP Project, *D7.6.1 Procurement Strategy*

[13] PRACE-1IP Project, *D2.2.4 PRACE Operational and Procurement Model*

[14] PRACE-1IP Project, *D9.3.3 Report on Prototypes Evaluation*

[15] PRACE-PP Project, *D5.4 Report on the Application Benchmarking Results of Prototype Systems*

List of Acronyms and Abbreviations

AISBL	Association sans but lucrative (legal form of the PRACE-RI)
ARM	Computer processors designed by ARM Holdings
CEA	Commissariat à l'énergie atomique et aux énergies alternatives
CINECA	Consorzio Interuniversitario, the largest Italian computing centre (Italy)
CSC	Finnish IT Centre for Science (Finland)
DARPA	Defense Advanced Research Projects Agency
DoE	US Department of Energy
DoD	US Department of Defence
EC	European Community
EPCC	Edinburg Parallel Computing Centre (represented in PRACE by EPSRC, United Kingdom)
EPSRC	The Engineering and Physical Sciences Research Council (United Kingdom)
Exaflop/s	Exa (= 10^{18}) Floating point operations (usually in 64-bit, i.e. DP) per second
FP	Floating-Point

FPGA	Field Programmable Gate Array
FPU	Floating-Point Unit
FZJ	Forschungszentrum Jülich (Germany)
GB	Giga (= 2^{30} ~ 10^9) Bytes (= 8 bits), also GByte
Gb/s	Giga (= 10^9) bits per second, also Gbit/s
GB/s	Giga (= 10^9) Bytes (= 8 bits) per second, also GByte/s
GDP	Gross domestic product
GENCI	Grand Equipement National de Calcul Intensif (France)
Gigaflop/s	Giga (= 10^9) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s
GHz	Giga (= 10^9) Hertz, frequency =10^9 periods or clock cycles per second
GoP	Group of Procurers
GPGPU	General Purpose GPU
GPU	Graphic Processing Unit
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPCS	High productivity computing systems. A DARPA programme in the USA.
HPL	High Performance LINPACK
JSC	Jülich Supercomputing Centre (FZJ, Germany)
KB	Kilo (= 2^{10} ~10^3) Bytes (= 8 bits), also KByte
LINPACK	Software library for linear Algebra
MB	Mega (= 2^{20} ~ 10^6) Bytes (= 8 bits), also MByte
MB/s	Mega (= 10^6) Bytes (= 8 bits) per second, also MByte/s
Megaflop/s	Mega (= 10^6) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MHz	Mega (= 10^6) Hertz, frequency =10^6 periods or clock cycles per second
Mop/s	Mega (= 10^6) operations per second (usually integer or logic operations)
NDA	Non-Disclosure Agreement. Typically signed between vendors and customers working together on products prior to their general availability or announcement.
PCP	Pre-Commercial Procurement
Petaflop/s	Peta (= 10^{15}) Floating point operations (usually in 64-bit, i.e. DP) per second, also PF/s
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE-RI	PRACE Research Infrastructure
R&D	Research and development
SMEs	Small and medium enterprises
SNIC	Swedish National Infrastructure for Computing (Sweden)
SoC	System on a chip
SP	Single Precision, usually 32-bit floating point numbers
STRATOS	PRACE advisory group for STRAtegic TechnOlogieS
TB	Tera (= 240 ~ 1012) Bytes (= 8 bits), also TByte
Teraflop/s	Tera (= 10^{12}) Floating-point operations (usually in 64-bit, i.e. DP) per second, also TF/s
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure would host the Tier-0 systems; national or topical HPC centres would constitute Tier-1

Executive Summary

Pre-Commercial Procurement (PCP) is a new model of procurement that is being promoted by the European Commission (EC) and gaining usage in many European Union member states in order to foster new innovative products prior to their commercial release, especially in the IT market and reduce the gap between Europe and countries like USA.

HPC has been identified as an area where it should be used and where basic R&D coupled with PCP can drive European innovation.

In the PRACE-3IP project, a consortium of partners called the Group of Procurers (GoP) will launch a joint Pre-Commercial Procurement (PCP) pilot on "Whole System Design for Energy Efficient HPC" addressing one of the major obstacles towards future multi-Petascale supercomputers. It will be the first time in Europe that such a PCP procedure will be used in the field of HPC. For the partners involved in this exercise it will represent a clear assessment of the potential of such procedure for future procurements.

The goal of the PCP is to drive innovation towards HPC solutions, which on the one hand are suitable for operation within the PRACE infrastructure of leadership class systems for scientific computing, and on the other hand significantly improves on energy efficiency. The formulated technical requirements define minimum requirements with respect to the performance of such leadership systems. Bidders are given the freedom to propose different solutions with respect to how to achieve improvements in terms of energy efficiency. This increases the openness and attractiveness of this PCP to SMEs, who may wish to provide highly innovative solutions that concern only parts of an overall system design.

PCP aims to foster innovation for economic growth to ensure sustainable high-quality public services in Europe. The PCP framework has been devised to enable public procurers to source research and development services in a competitive and open manner, bringing the best solutions to the public agencies running the procurement. The research and development services must result in pilot systems that demonstrate that the outcome meets the set of technical requirements.

In the previous PRACE projects, PRACE-PP, PRACE-1IP and PRACE-2IP, the development and purchasing of prototype technologies that may subsequently be deployed within the PRACE-RI has been undertaken by the successful prototyping-oriented work packages. These work packages, supported at times by the work of the legal-aspects work packages, have successfully implemented a series of prototyping activities, where appropriate using co-design approaches to technology development with external suppliers and existing procurement processes. Considerable existing knowledge and experience has already been gained and shared between partners.

During the first six months of the current PRACE-3IP project, the Group of Procurers detailed the specification of these technical requirements based on the overall goals as described in the project proposal. "Whole System Design for Energy Efficient HPC" has been chosen as the focus of this PCP because of the need to take a holistic approach to technology developments in HPC as we approach the multi-Petascale and Exascale eras. Simply optimising one component of the system, be it cooling system, processor or memory, will not deliver the necessary reduction in energy usage we will require in the next decade to operate HPC systems at the pinnacle of performance.

Through advertising in targeted mailing lists and websites, PRACE encouraged interested suppliers to participate in an open technical dialogue. A public meeting was held in Brussels and subsequent feedback received from those who expressed an interest in the dialogue. The

feedback resulting from this process has been invaluable in finalising the technical requirements of the forthcoming procurement.

These finalised technical requirements have been used to form the tender documents for the PCP and these will constitute the core of the subsequent public tender. To compare the different bids, evaluation criteria have been derived based on these technical requirements, and will be used as a part of the tendering process. Here we present detailed explanations and justifications for the technical criteria PRACE has selected for the PCP.

1 Introduction

Pre-Commercial Procurement (PCP) is a new model of procurement that is being promoted by the European Commission (EC) and is gaining usage in many European Union member states. The EC have proposed the PCP model [1] in order to tackle a difference between how Europe and the USA benefit from their basic and applied R&D expenditure. Although the EU and US have similar levels of spending on basic and applied R&D (2% of GDP in Europe versus 2.5% of GDP in the USA) the USA have, over the past 30 years, followed a public procurement policy that encourages early procurement by the public sector of new innovative products prior to their commercial release. This has led to an enormous difference between public procurement of R&D results between Europe and the USA (€2.5 billion in Europe versus €50 billion in the USA, annually). This difference, with the USA investing twenty times more in the public procurement, is felt by the EC to be a major contributing factor in Europe's inability to build large, innovative companies, particularly in the IT sector.

The EC have therefore proposed PCP as a procurement model in order to address this innovation gap. Information Technology is seen as a key area where this model has worked well in the USA, and the EC would like to see it being used in Europe. HPC has been identified as an area where it could be used successfully and where R&D coupled with PCP can drive European innovation. This is the fundamental rational behind the pilot of PCP currently being undertaken by the PRACE-3IP project, part of which is described in this document.

Energy efficiency has been clearly identified as one of the major challenges to address in the design and the operation of future multi-Petascale and Exascale HPC systems by multiple international expert reports (IESP [2], EESI [3]).

Consequently, a subset of PRACE-3IP project partners, participating as the Group of Procurers in this context, have decided to focus the PCP on "Whole System Design for Energy Efficient HPC". This has enabled them to exploit the experience of the partners and the results of prototype work undertaken during previous PRACE projects. It will be the first time in Europe that the PCP process will be used in the field of HPC. For the partners who are involved in this exercise it will represent a clear assessment of the potential of such procedures for future procurements.

The decision on which technical innovation will lead to the most significant improvement in terms of energy efficiency is left to the bidder. Technically we expect the following aspects to be most relevant:

1. Energy efficient computers: to meet the energy constraints of multi-Petascale and subsequent Exascale systems, hardware suppliers must develop and integrate energy efficient processors, chipsets, low-power memory and interconnect technologies, possibly including energy-efficient accelerator technologies, and drawing on the expertise from Europe's existing HPC and embedded computing sectors.
2. Extreme cooling efficiency: current direct liquid-cooled high-temperature HPC cooling systems are able to remove up to 85% of the heat generated under normal computer centre conditions. The remaining 15% may be tackled by targeting technological developments related to thermal and infrared heat dissipation and at the reduction in the number of components not subject to water cooling (power supplies, disk storage, network components).
3. Systemware efficiency: overall system efficiency improvement through scalable file systems and system software optimization, including operating system support for application based fault tolerance and resiliency aspects.

The main expected goals of the PCP call can be summarised as:

- The transformation of HPC Centre efficiency through the production of energy efficient hardware and systemware components, demonstrated through the benchmarking of a set of representative PRACE HPC applications;
- The demonstration of an integrated whole system design for low total energy-to-solution values, as enabled by highly energy-efficient computing and extreme cooling technologies.

It should be noted that the previous PRACE projects and the very good performance of scientific applications using the PRACE-RI facilities show that several European teams have developed application and software know-how that enable them to engage fruitfully with vendors. However, PRACE partners cannot bid as part of a vendor response to the PCP tender and this is one aspect of the process that causes concern in the context of PCP and stimulating co-design of HPC technologies in Europe.

The objective of Task 8.1 in the PRACE-3IP work package WP8 is to define the technical requirement specifications and evaluation criteria for the PCP. The deliverable D8.1.1 specifies the technical criteria for Phase 1 of the PCP. It describes the expected solution of the PCP and defines the evaluation criteria.

In this document, Section 2 illustrates the background of the PCP pilot by PRACE, Section 3 underpins the importance of energy efficiency for future HPC technologies, Section 4 describes the technical dialogue with vendors which resulted in the final technical requirements specified in Section 5, and finally Section 6 draws conclusions of the process which led to the specifications of these technical requirements. The annex contains the presentations from the ‘Open Dialog with Vendors’ workshop. The vendor feedback is included in a confidential annex that is omitted from the public version of the deliverable.

2 Background

2.1 Legal basis

The legal basis for PCP is quite complex. A good explanation can be found in Ref. [3] (see Chapters 3 and 5). In summary, there is an exemption from the normal rules on how public organisations can purchase research services in the Directive 2004/18/EC (chapter 16/f) for "*research and development services other than those where the benefits accrue exclusively to the contracting authority for its use in the conduct of its own affairs, on condition that the service provided is wholly remunerated by the contracting authority."* This means that there are no restrictions on the way a public organisation can purchase research services in the EU, provided certain constraints are met. PCP is a recommendation on the usage of this exemption. It defines a safe way to implement this usage and lowers the risk of problems during procurement. However, it is a recommendation, not a regulation, so there is flexibility in the implementation by individual EU member states – only some of which have currently adopted the practice.

2.2 The PCP Model

PCP is seen as a phased model whereby initial basic and applied R&D, possibly funded by the EC Framework Programme or EU member state research funding, is subsequently commercialised through a phase of PCP.

One of the best-known recent examples of an equivalent approach has been DARPA's HPCS Programme in the USA [4]. This programme has funded IBM, CRAY, SUN and other US companies to develop the next generation of HPC hardware and software (particularly next generation HPC languages).

Other more recent examples in the field of Exascale computing are the two "PCP-like" procurements launched in 2012 and 2013 by US DoE, called DesignForward and FastForward. Both aim to deploy early investment for processor, memory and storage technologies, and more globally Exascale technologies that are required to move research technology into vendors' products.

During the latest SC'12 conference, W. Harrodd (Advanced Scientific Computing Research at DoE) presented some information about the FastFoward procurement with early information about the companies recently awarded:

Vendor	SCOPE	Value
AMD	Processor / Memory	$12,600,000
IBM	Memory	$10,476,714
Intel	Processor / Memory	$18,963,437
NVIDIA	Processor	$12,398,893
WhamCloud (Intel)	Storage & I/O	$7,996,053
Total		**$62,435,097**

Table 1: Companies selected by the FastForward procurement by DoE

The diagram below shows the EC model for PCP:

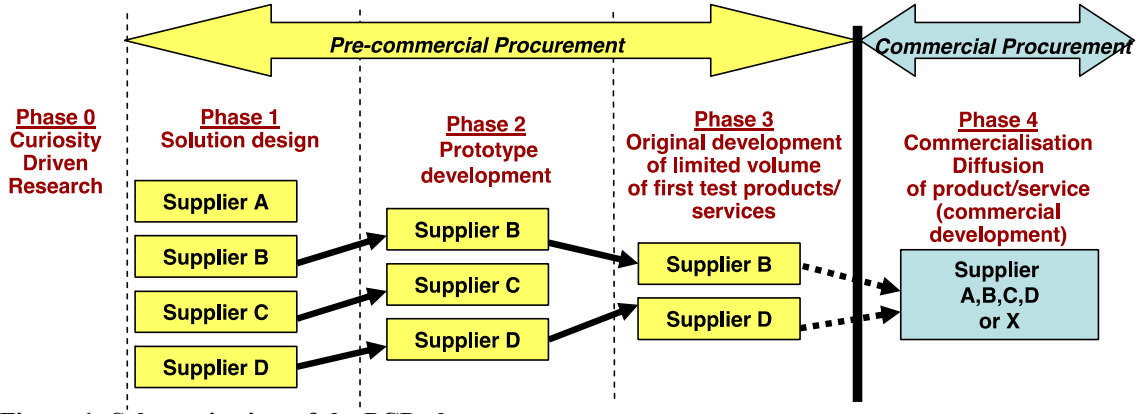

Figure 1: Schematic view of the PCP phases

Key components of the model include:

- Pooling the efforts of multiple procurers;
- Procuring from suppliers *"so as to stimulate companies to locate a relevant portion of the R&D and operational activities related to the PCP contract in the European Economic Area or a country having concluded a Stabilisation and Association Agreement with the EU";*
- Ensuring that specific activities within the project focus on *"defining the mid-to-long term solution requirements for the required public service innovation".*

2.3 PCP in PRACE-3IP

In the previous PRACE-PP, PRACE-1IP and PRACE-2IP projects, the development and purchasing of prototype technologies that could subsequently be deployed within the PRACE-RI has been undertaken by the successful prototyping work packages. These work packages, supported at times by the work of the legal aspects work packages, have successfully implemented a series of prototyping activities, where appropriate using co-design approaches to technology development with external suppliers and existing procurement processes. The results can be found in Refs. [14] and [15].

Considerable existing knowledge and experience has already been gained and shared between the partners. For example, work packages WP7 of PRACE-PP and WP2 of PRACE-1IP have studied good procurement practices (see Refs. [12] and [13]). Furthermore, PCP was briefly studied in the original PRACE-PP project's deliverable D7.6.1 [12]. Interaction with industry partners has been successfully mediated through STRATOS (a PRACE advisory group for Strategic Technologies) and the activity of the prototyping work packages.

Since the introduction of Pre-Commercial Procurement by the European Commission at the end of 2007, it has not so far been applied to the field of High Performance Computing in Europe, whereas the equivalent approach is widely used by US agencies (DoE, DoD) for funding leading R&D services.

As HPC is now considered as a strategic activity by many European countries, as well as the European Commission, it has become important to assess the feasibility of such an instrument in the context of a multi-partner multi-national consortium. The PRACE-3IP project accepted the responsibility to undertake the pilot; it is driven by its work package WP8, with the strong support of Task 2.1 of the work package WP2.

3 Energy efficiency

3.1 Exascale challenges

The primary constraint of future HPC architectures will be power consumption. Already today the restrictions in power consumption have stopped any further increase of core clock frequencies leading to a significant increase in on-chip parallelism to achieve further increase of performance per computing device – the so-called "multicore revolution". While power consumption is mainly limited due to constraints in the technology (as well as the costs of this technology), energy consumption becomes a major issue due to electricity costs, which in future could significantly exceed the costs for the system's hardware itself. For this reason different Exascale studies have limited the power consumption of an Exascale system to 20-30 MW. It is important to note, however, that no system design exists today which can deliver a system within this power envelope. We stress that also the DARPA study published in 2008 [10] indicates a power requirement in excess of 60MW for their best design.

The energy per floating-point operation is not the most critical part of the energy budget. Today energy costs are about 30 pJ/Flop for high-performance and up to 2 times less for low-power designs. They are expected to drop to 7 pJ/Flop or less around 2018 [10]. This means an Exascale system would consume up to 7 MW for executing 1 Exaflop/s. But this is only part of the story.

From an overall systems perspective, the real energy (and therefore also power) challenges lie in low-energy data transport. Data needs to be moved within a processor chip, off-chip within a node, e.g. between compute chip and an external (high capacity) volatile memory, between different nodes as well from nodes to non-volatile storage devices (e.g. external, high capacity

storage systems or integrated high-performance Storage Class Memory). The power consumption of electrical links scales (to a first approximation) linearly with the signalling rate and the wire length squared. Today short-distance copper links consume about 10-20 pJ/bit, which is expected to improve to 2 pJ/bit by 2018 [11]. This means that moving the (double precision) input arguments and the result of a floating-point add-multiply over a short distance of copper links will cost almost 40 times more energy than the arithmetic operations itself.

At system level, power efficiency as defined by the Green500 project [6] recently exceeded 3 Gigaflop/s/W, reaching 3.2 Gigaflop/s/W in the most recent list published in June 2013 [6]. This corresponds to a power consumption of 0.3 MW for 1 Petaflop/s and would result in 312.5MW when linearly extrapolated to a compute performance of 1 Exaflop/s. According to average 2011 electricity prices for industrial consumers in the EU countries, this would translate into operational costs of 300.8 million EUR per year for electricity costs alone.

While the Green500 list is widely accepted and plays a valuable role in promoting awareness concerning power and energy efficiency, it does, however, have a number of shortcomings. Most importantly, the Green500 list by definition does not rank according to energy usage but power efficiency. The entries are based on a measurement of a (relatively constant) compute performance and power consumption of a system while executing the High Performance LINPACK (HPL). Using HPL as a standardized load has some advantages from a methodological point of view, but it is not a representative load and its performance signature is far off many real-world applications. For many HPC centres, the highest-ever recorded power load on their systems is when HPL is run on them for benchmarking and testing purposes.

Power consumption of the overall systems is, according to the Green500 list's "run rules", extrapolated by measuring the power consumption of a part of the system, e.g. a single rack, for which power consumption can be measured in isolation. As a result, the full computing centre's power consumption is not taken into account, e.g. power consumed by external storage systems, computing centre communication networks, or the cooling infrastructure are neglected.

Various power and energy efficiency metrics have been proposed in the literature. Taking only operational costs for power into account, measuring the energy required to accomplish a certain task seems to be most natural. This, however, does not take the need to finish a given task within a certain amount of time into account. In Ref. [5] therefore a function of time-to-solution times energy (FTTSE) metric was proposed, where energy consumption is multiplied by a weight function $f(T)$ that depends on the total execution time T.

In summary, energy efficiency has become the main obstacle for future supercomputer architectures on the way to Exascale in 2018-20. It implies that radical changes in the design of computer hardware and software compared to today's technology are necessary to build multi-Petaflop/s systems and to break the Exaflop/s barrier.

3.2 Previous work on power consumption

Based on this analysis, previous PRACE projects (PRACE-1IP, PRACE-2IP) have:

- Invested significant effort in measuring power and energy consumption for architectures and applications which are part of, or are commonly used, within the PRACE infrastructure;
- Promoted optimisation of power consumption in future system architectures by implementing and evaluating various prototypes based on new technologies (see Fig. 2 for an overview).

Examples which we want to highlight and which are shown in the diagram below are (i) the SHAVE-PRACE prototype, (ii) an ARM+GPU based system and (iii) a prototype which makes use of direct liquid free cooling:

i. The SHAVE-PRACE prototype utilises the Streaming Hybrid Architecture Vector Engine (SHAVE) processor architecture developed by the European company Movidius to provide the rapidly increasing computational capabilities demanded by next generation mobile video applications. The heart of the prototype platform consists of 4 identical nodes, each featuring 8 Movidius Fragrak SoCs with 128 cores/node. The nodes are connected by a special-purpose network implemented in FPGA hardware;

ii. The ARM+GPU system is based on the fact that a growing number of the Top500 systems today are built on multi-core chips coupled with GPGPU accelerators. In this prototype the low-power counterpart of such systems is investigated by using ARM multicore processors and mobile GPGPU accelerators;

iii. Direct Liquid Free Cooling leverages the opportunity of raising the outlet temperatures of cooling circuits by attaching it directly to the systems hot spots. High temperatures facilitate free cooling without the use of chillers thus reducing the power consumed by the cooling sub-system. At sufficiently high temperatures heat reuse becomes an option, e.g. for cooling by means of an adsorption chiller.

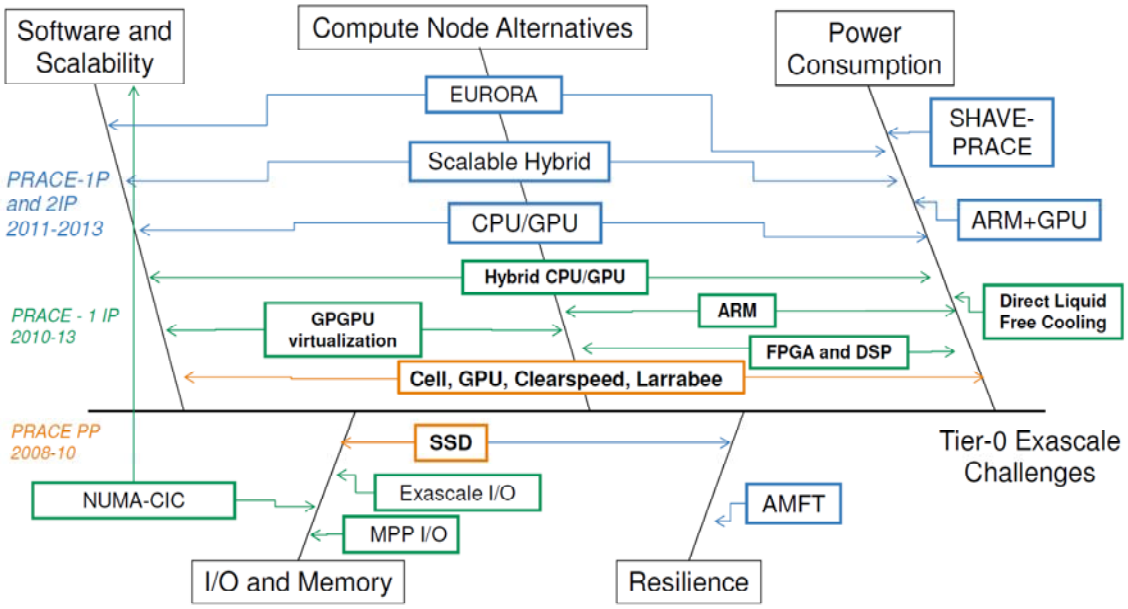

Figure 2: Diagram of the different PRACE prototypes

Furthermore, the DEEP [8] and Mont-Blanc [9] projects are Exascale initiatives, funded by the EU 7th Framework Programme, which use these technologies. The Dynamical Exascale Entry Platform (DEEP) project employs direct liquid cooling to enable free cooling of both the cluster system as well as the many-core accelerator based booster system. The Mont Blanc project designs a next-generation HPC system based on low-power commercially available embedded technologies like ARM multi-core processors and mobile GPUs.

Some others prototypes targeted application resilience or I/O and memory technologies which may also impact the overall energy efficiency of a whole system.

3.3 Progress beyond state-of-the-art

In this part of the PRACE-3IP project we are making progress beyond current state-of-the-art in two key aspects:

3.3.1 Technical progress

The criteria to evaluate different systems will be purely based on energy efficiency. The metric that will be adopted is similar to the FTTSE metric as it mandates an upper limit for time-to-solution. Secondly, the PCP will result in pilot system(s), i.e. systems able to be tested in an HPC centre with real applications in pre-production mode.

3.3.2 Legal and procedural progress

The validity, or otherwise, of PCP as a tool for stimulating innovative research and development in the European HPC sector will be tested and documented. If in the future PCP is to be used as a tool to publicly procure large-scale high performance systems, the process that the PRACE-3IP project is exploring with this activity will provide vital knowledge of what works and what doesn't.

4 Technical Dialogue with Vendors

Before launching the formal PCP procedure, all interested major suppliers were invited to start a technical dialogue on both the technical challenges related to energy efficiency and the practical implementation of a PCP process in the context of HPC. To start the dialogue, a workshop was organised in Brussels on September 21, 2012, where representatives from all institutions that intend to constitute the Group of Procurers (GoP), i.e. CINECA, CSC, EPCC, GENCI, JSC, SNIC, and the PRACE AISBL participated.

During this conference, Lieve Bos gave a talk on behalf of the European Commission, which explained to the participants the role of PCP as a funding instrument in Europe. The stage for the technical part of the conference was set by a presentation of Lennart Johnsson on the achievements with respect to energy efficiency prototyping activities in PRACE-2IP.

The main goal of the conference was to present to the participating vendors the current vision of the organisation of the planned PCP (presented by François Robin), as well its technical goals (Mark Parsons) in order to receive their feedback during and after the workshop.

This feedback has been and is being used by the GoP to define the setup of the PCP and adjust the technical goals such that the overall goals of the project can be achieved. During two Q&A sessions, the participating suppliers were given the chance to ask questions, formulate comments, or present their views. All invited vendors, i.e. not only the participating ones, were encouraged to provide written feedback after the meeting. Due to the confidentiality, which has to be ensured, this public document only contains a summary of the feedbacks, while full details can be found in the confidential appendix.

The technical goals presented during this meeting were chosen to be both comprehensive and informative enough as to ensure a concise discussion of the technical and commercial viability by the participants with the organisers. The technical talk given at the meeting is included as an Appendix to this document. It is important to note that following these discussions and the feedback from the vendors, considerable changes were made to the technical requirements of the PCP. This was exactly the intended response to the dialogue process, which has been very beneficial to the development of the PCP.

Anticipating the most power-efficient computing devices to be many-core (and perhaps heterogeneous) devices, development work targeting a tighter-than-today coupling to network and storage devices was suggested. For a demonstration of the overall power efficiency of the system it was proposed to mandate a submission to the Green500 list with an efficiency of at least 4 Gigaflop/s/W. Other architectural features included the modularity of the design to allow for flexible optimization of systems, either for compute-intensive or data-intensive applications. Furthermore, it was suggested to the vendors to foresee liquid cooling to reduce the power consumed by the cooling subsystem and to achieve the target peak compute performance of 1 Petaflop/s (double precision) per rack in a compute-capability configuration. The architecture should be such that systems could scale to 100 Petaflop/s sustained performance.

Another area of technical goals dealt with data-intensive capabilities of the system building blocks. This could be achieved by integrating storage class memory or equivalent, which would allow very high IOPS rates. However, it would require significant effort to fully integrate such storage devices in the system software stack, e.g. to integrate such devices in parallel file systems or support efficient check pointing.

In order to be successful, the PCP must ensure that the outcome is an architecture and implementation that shortly after the PCP will lead to a product which could be procured by PRACE members and deployed within the PRACE-RI, but also be successful on the HPC market in general. Therefore, a working pilot system or systems must be developed in the framework of this project. It must meet reliability, availability and serviceability requirements to be suitable for production use in a PRACE HPC centre, although the pilot itself may still consist of components that have not yet passed full product qualification.

The main concern of the vendors was the challenge of reaching agreement between the number of competing options, the scope of development required and the demonstration of systems capabilities within the available budget limits. The general feedback from the vendors was that the overall budget is relatively small in comparison to the requested features. The concern was raised that the requirement of providing an integrated design and to stimulate development of innovative technologies could be in conflict. It was suggested to reduce the number of technical focus topics to very few main topics of interest, despite that the presented technical goals were considered both relevant and valuable areas for investment in research and development. One of the vendors stressed the need for the customer to make choices to ensure the whole PCP process is customer-driven, making it more likely that the outcome meets market expectations and leads to competitive products.

Other concerns raised included the formulated power efficiency targets, referring to significantly more ambitious targets formulated in the DARPA call. Furthermore, it was noted that there are significant differences with respect to the needs of companies doing chip development (e.g., high non-recurring engineering costs during early development phases) or those doing system integration (e.g., high costs for building the pilot system, i.e. the last phase of the planned PCP).

In general, the feedback from vendors and the technical dialogue process was viewed as extremely useful and informative by the GoP.

5 PRACE PCP Technical Goals

5.1 Introduction

On the basis of the feedback from the vendors the technical goals and the derived technical requirements were reformulated such that the PCP will result in innovative HPC solutions that can be operated within PRACE while leaving vendors the freedom to target their R&D efforts such that maximum progress in terms of reduction of energy-to-solution across the overall system is achieved. In this way the scope of the challenge becomes manageable for potential suppliers.

The process that resulted in the list of technical goals documented below had been complex. Different ideas and strategies on how to promote innovative solutions to the challenge of Energy Efficiency in Whole System Design had to be brought together and balanced with what is realistically achievable within the available funding and time constraints.

5.2 Agreement on Technical Goals

The list of goals given below describes at a high level what the PCP wants to achieve in technical terms. It includes a rationale for each of the goals. These goals will also be presented to each bidder together with a set of precisely defined technical requirements that are derived from this list of goals.[1] Solutions proposed by each bidder must conform to the technical requirements in order for the bid to have a chance of being successful.

On 22nd November 2012 the GoP met in Brussels to consolidate the preparatory work we had done to date, the outcome of the vendor dialogue process, and the specific feedback we had received from that process. During this meeting the GoP unanimously agreed on a set of PRACE PCP Technical Goals.

In the following text, each of the agreed Technical Goals is given in **bold**. The rational behind each of the requirements is then given below this.

Technical Goals

5.2.1 Energy efficiency in whole system design

TG-1-1 All proposed solutions should target energy efficiency in whole system design. Vendors can choose which components of a system to optimise. One or more of computational or input/output (I/O) performance or other sub-systems, such as interconnection network, cooling or power supply, can be targeted.

Here we state the main target of the PCP that was included in the original PRACE-3IP project proposal. Based on the feedback received and our own analysis we have concluded that insufficient funding is available for vendors to optimise all of the components in a whole system for energy efficiency. They may therefore decide which of the components of a system to optimise and, based on the companies that have expressed interest in the PCP, we expect this to be computational, I/O, cooling or the performance of any other sub-system.

TG-1-2 Vendors will propose the minimum improvement in energy efficiency at the bidding stage that they will demonstrate through their pilot system.

We want vendors to consider at the outset what the goals of their technical developments are and what they expect the minimum improvement in energy efficiency to be. This will allow

[1] See: PRACE-3IP PCP Pre Commercial Procurement on "Whole System Design for Energy Efficient HPC", Technical Requirements. This document will be published with the Tender Documents.

us to discriminate between bids and also identify those technologies in between the first and second phases that have the most promise.

TG-1-3 The energy performance model should demonstrate how close the system design is to a linear scaling in terms of energy usage at least up to the performance defined in TG-3-2.

Different system designs vary in terms of how their energy use scales as more and more racks of the hardware are joined together into a full system. Some designs will scale linearly, others, largely those with high performance networks, may not. We would like to understand how close to a linear scaling each pilot system that is delivered achieves.

5.2.2 Self-contained pilot system

TG-2-1 The technology or technologies for improving energy efficiency which vendors propose to develop and implement must be integrated and deployed within the PCP as part of a pilot system. This pilot system should be able to be tested in an HPC centre with real applications as a "pre-production" system.

Although we only expect vendors to focus on one or two aspects of energy efficiency, we do still want them to deliver a working pilot system, which we can use to assess and test their technology. It will not be sufficient merely to develop and deliver a new energy efficient component. This pilot system, integrating the new energy efficient components into a 2016 state of the art HPC system, must be capable of operation in a PRACE HPC centre as a pre-production system.

TG-2-2 The pilot system will be self-contained and, in particular, not rely on external data stores.

External data stores would be one way in which vendors could hide poor energy efficiency performance caused by deficiencies in the I/O sub-system. We therefore require that the proposed systems are self-contained and do not rely on external data stores.

5.2.3 Architecture suitable for PRACE

PRACE partners aim to purchase and operate the largest, most powerful HPC systems in Europe for their users.

TG-3-1 Pilot systems delivered at the end of the PCP should provide a minimum of 1 Petaflop/s peak.

This requirement sets a clear scale for the vendors to deliver against. By 2016 we fully expect this to be attainable with ease using the technologies that are on most vendors' roadmaps.

TG-3-2 Pilot systems should be designed to be scalable to 100 Petaflop/s peak.

By 2016 we expect there to be at least one 100 Petaflop/s system installed somewhere in the world. We therefore want to ensure that the technologies being developed by the PCP have the potential to reach sufficient scale to compete in the global marketplace as Tier-0 systems.

5.2.4 Energy measurement capabilities

TG-4-1 The pilot system must feature reliable measurement of energy consumed by the whole system with a granularity of 10 seconds or less.

This feature is required to enable verification of the improvement in energy efficiency. Currently available systems typically only allow the measurement of power with a temporal granularity which is not sufficient to obtain a reliable estimate of the power consumption integrated over time. Furthermore, for current systems measurements of power (and energy) consumption can usually only be determined for a part of the system. Even in cases where there are comprehensive measurement data, the results from different sub-systems are typically not aggregated to provide a complete and reliable power-profile of the system.

TG-4-2 Improvements in energy efficiency must be demonstrated through the use of real production application codes and a benchmark in use by PRACE. These codes will stress both compute and I/O performance and be supplied with both small and large representative datasets. Vendors will be provided with four PRACE application codes selected by PRACE and accompanied by the High Performance LINPACK benchmark. We have chosen to allow vendors to modify no more than 10% of the software code (in number of lines of source code) in order to give them the opportunity to take advantage of the technical advances present in their pilot systems without wasting large amounts of effort on application optimisation.

In developing new technologies, PRACE must focus on those technologies that are of value to our users. We therefore want to evaluate the improvements to energy efficiency using real production scientific codes in use by PRACE users and also by a common benchmark (see next requirement). The codes will stress compute, I/O and network performance of the systems in order to ensure that we can compare the systems that have focused on different aspects of energy efficiency. The application codes will be provided with both large and small datasets in order to facilitate system development and to aid this comparison.

Four application codes have been selected from the ones running daily in production on the current generation of Tier-0 systems and applications studied by the application work packages in the PRACE implementation phase projects (1IP, 2IP and 3IP). These applications will be accompanied by the HPL benchmark as well as a set of synthetic I/O benchmarks.

We have chosen to allow vendors to modify no more than 10% of the software code (in number of lines of source code) in order to give them the opportunity to take advantage of the technical advances present in their pilot systems without wasting large amounts of effort on application optimisation. Every modification should be documented, and it's rational explained. The algorithms used should not be modified nor the accuracy of the computation (64 bits computation should remain 64 bits computation). Usage of other libraries could be permitted, but require agreement of the PCP team. The numerical results should be the same even if small differences should be allowed (rounding errors, etc…). It is up to the PCP team to decide if modifications are acceptable and if small differences in the results are acceptable.

TG-4-3 Vendors must compare their total wall-clock time and total energy consumed during execution by each of the codes while running each of the representative datasets against benchmark measurements taken by PRACE on current (2013) PRACE systems.

In order to have a reference point, PRACE will benchmark each of these codes on current (2013) PRACE systems including various HPC architectures (MPPs, clusters of SMP thin, fat or hybrid nodes). We will measure total energy consumed by each application and also the total wall-clock time (see next requirement). This is to avoid energy efficiency being mainly achieved by reducing clock speed.

TG-4-4 All applications must be demonstrated on the final pilot systems executing to completion in equal or less wall-clock time than the original measurements performed by PRACE.

We want vendors to provide their technology in working, pre-production pilot systems. Therefore, all of the applications must be demonstrated on the final pilot systems and none of them must execute in more wall-clock time than the 2013 measurement.

5.2.5 Energy efficiency models

TG-5-1 Vendors should develop a model of performance that enables total energy consumed for each code on a 100 Petaflop/s peak system to be predicted. Vendors may replicate the four PRACE application codes across the model system as if they were executing as an ensemble. Vendors should quote a result for HPL scaled to the full system.

Although the delivered pilot systems will only provide a minimum of 1 Petaflop/s peak, we want to be able to forecast and understand the total energy consumed for each code if executed on a 100 Petaflop/s system. This is a complex requirement because few if any of today's PRACE application codes will scale, in terms of parallelism, to such a system. In this requirement we therefore allow the vendors, in their models, to simply assume they can run multiple stand-alone copies of the four PRACE applications as if they were operating as an ensemble analysis. Because the HPL benchmark will scale to this level of parallelism (it is trivially parallel), we expect the vendors to quote a result for HPL across an entire 100 Petaflop/s system.

5.2.6 Energy efficient technology with sustainable market

TG-6-1 Vendors must present an analysis that shows that a sustainable market for the technology developed within the PCP exists from 2016 onwards.

There is no point in vendors developing technology for its own sake. PRACE wants to be sure that its investment will lead to a sustainable market for each company or consortium who bids beyond the scientific HPC market. This is a key component of PCP – providing the funding to perform research and development and ensuring that a market (beyond the initial procurers) exists for the technology that is developed. This is a skill that the USA has had for many years, and which Europe must learn to emulate.

Evaluation Criteria

For the evaluation of the bids provided during the Tendering Stage as well as for Phase II and III during the Execution Stage a set of overall technical criteria and corresponding weights have been defined:

Criteria	Description	Weight
Quality of R&D and level of innovation	Quality of the offered R&D services and the solution's ability to innovate and improve substantially the scope of operation in which it is intended to be inserted.	30%
Technical requirements compliance	Level of compliance of the solution (in terms of quality and completeness) to the functional and performance requirements.	20%
Progress in terms of energy efficiency	Solution's ability to progress energy efficiency beyond state-of-the-art.	30%
Project quality and feasibility	Quality of the project (work planning, risk management etc.) as well as feasibility and reproducibility of the solution using an industrial process proper respect to the reference market.	20%

Table 2: Evaluation Criteria

These criteria and the weights will not change during the PCP, although the specific sub-criteria used for rating the bids will change during the progress of the innovation process. These criteria will also be used as references for the evaluation of the performance of the successful bidders at the end of the phases of the Execution Stage.

The evaluation of the initial bids comprises a technical and financial evaluation. The weight for the technical and financial evaluation is 90% and 10%, respectively. This choice reflects the goal to maximize the outcome of the PCP by promoting the provision of high quality R&D services.

The technical quality of the bids will be evaluated on the basis of a set of technical documents which the bidders are requested to submit:

- Technology concepts;
- High-level architecture and pilot system description;
- Power-efficiency analysis;
- Work-plan for phase 1, 2 and 3;
- Risk analysis;
- Market analysis from 2017 onwards;
- Critical components supply documentation;
- Human resources and place of performance documentation.

This technical documentation must be at a level of detail that allows an assessment of the compliance of the proposed solution with the technical requirements. Furthermore, for each of these technical documents one or more criteria have been defined to rate both the quality of the solution as well as the quality of the technical documentation that is provided.

6 Management of Intellectual Property Rights

A key aspect of the PCP process is that both risks and benefits associated with innovative R&D are shared between suppliers and the procurer. At the same time the process should be such that there is a high incentive for all involved parties to pursue wide commercialisation and to take up of the new solutions. Therefore, an Intellectual Properties Rights (IPR) scheme was chosen where the ownership of IPR remains with the supplier that generated it unless these rights are not exploited within a period of three years after the PCP has ended. In any case, the procurer will retain a worldwide free and non-exclusive licence to the IPR generated in the PCP plus any relevant background IPR.

Background IPR here refers to information or technology that was under control of the supplier before start of the PCP or was generated outside the PCP. Relevant background IPR includes all intellectual properties needed for effectively being able to use the foreground IPR generated within this project. In order for the procurer to benefit from the results of the PCP the supplier may not charge any member of the Group of Procurers for the license to this foreground IPR.

7 Conclusions

Pre-Commercial Procurement (PCP) is a new model of procurement that allows the public sector to procure new products prior to their commercial release with the aim of stimulating innovation. In this project this model is applied for the first time in Europe to the field of HPC in order to address the problem of energy efficiency. In the past, similar methods have been successfully applied in the USA. For example, DARPA (Defense Advanced Research Projects Agency, DoD) has funded a series of PCP-like activities such as the HPCS (High Productivity Computing Systems) in 2007-2010, and more recently a $20 million PCP awarded to nVIDIA project called PERFECT (Power Efficiency Revolution For Embedded Computing Technologies). DoE has funded the FastForward and the Design Forward PCPs launched in 2012 and 2013.

The PCP is being executed by a subset of PRACE-3IP partners, the Group of Procurers (GoP), under the umbrella and in close collaboration with PRACE-3IP. During three phases of the PCP several vendor consortia will compete against each other. Preparing the PCP requires the definition of a set of technical requirements and the corresponding evaluation criteria in order to compare the bids of the vendors and the outcome of the different phases.

The main challenge addressed in this PCP is the energy efficiency of future HPC architectures. Today power consumption has become a major source of restriction in the design space of many critical components. Without a significant boost to energy efficiency at the system level, costs for electricity will become unaffordable for Exascale systems. To ensure that the PCP results in solutions which could become part of a commercial procurement process immediately or soon after the PCP completes, the improvements in terms of energy efficiency must be demonstrated by building and deploying pilot systems, able to be operated in pre-production mode at PRACE sites.

An early important step while preparing the PCP was the initiation of an open technical dialogue with interested vendors to inform them about the planned PCP and expose them to the ideas on how to organise the PCP, as well as on the technical requirements. Almost all relevant vendors in Europe could be involved in this process and their feedback provided valuable input to finalize the technical requirements. The main concern of the vendors was the challenge of reaching consistency between the relatively small budget, the initially larger number of competing architectural options, the scope of development required, and the demonstration of system's capabilities by deploying a pilot system.

Based on this feedback a set technical requirements has been developed which is documented in this deliverable. The technical requirements, developed over one year, have been designed to present an achievable challenge to the vendors in order to encourage as many bids as possible, whilst still presenting a complex research and development challenge for the companies. Any solution proposed by a vendor must target energy efficiency in whole system design. The evaluation of improvements in energy efficiency will be verified through the use of real production applications codes in use by PRACE today, as well as by HPL benchmark.

Next steps concerning the technical aspects of this PCP concerns the benchmarking of the selected set of applications. For this set of applications and input data sets, execution times as well as energy consumption will be measured on a representative set of systems currently operated at PRACE sites.