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 Implementation Phase Project

Grant Agreement Number: RI-312763

D7.3.2
UEABS Benchmarking Results

Final

Version: 1.0
Author(s): Mark Bull, EPCC
Date: 20.2.2014

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: RI-312763**	
	Project Title: PRACE Third Implementation Phase Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: < **D7.3.2**>	
	Deliverable Nature: Report	
	Deliverable Level: PU	**Contractual Date of Delivery:** 28 / February / 2014
		Actual Date of Delivery: 28 / February / 2014
	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: UEABS Benchmarking Results**	
	ID: D7.3.2	
	Version: 1.0	**Status:** ***Final***
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2007	
	File(s): D7.3.2.docx	
Authorship	**Written by:**	Mark Bull, EPCC
	Contributors:	Dimitris Dellis (GRNET) Jorge Rodriguez (BSC) Andrew Emerson (CINECA) Xu Guo (EPCC) Jussi Enkovaara (CSC) Tristan Cabel, Gabriel Hautreux, Tyra Van Olmen (CINES) Tilman Dannert (RZG) Desislava Ivanova, Plamena Borovska (NCSA) Charles Moulinec (STFC)
	Reviewed by:	Nuzhet Dalfes, UHeM Dietmar Erwin, PMO
	Approved by:	MB/TB

Document Status Sheet

Version	**Date**	**Status**	**Comments**
0.1	07/Feb/2014	Draft for Review	
1.0	20/Feb/2014	For Approval	

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Benchmark Suite

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] PRACE 2IP Deliverable D7.4 "Unified European Application Benchmark Suite"
[2] PRACE PP Deliverable D6.3.2 "Final Benchmark Suite".
[3] PRACE 1IP Deliverable D7.4.2 "Benchmarking and Performance Modelling on Tier-0 Systems".
[4] DEISA Benchmark Suite http://www.deisa.eu/science/benchmarking
[5] Unified European Application Benchmark Suite http://www.prace-ri.eu/ueabs

List of Acronyms and Abbreviations

BSC	Barcelona Supercomputing Center (Spain)
CEA	Commissariat à l'Energie Atomique (represented in PRACE by GENCI, France)
CFD	Computational Fluid Dynamics
CINECA	Consorzio Interuniversitario, the largest Italian computing centre (Italy)
CINES	Centre Informatique National de l'Enseignement Supérieur (represented in PRACE by GENCI, France)
CPU	Central Processing Unit
CSC	Finnish IT Centre for Science (Finland)
CSCS	Swiss National Computing Centre (Switzerland)
DEISA	Distributed European Infrastructure for Supercomputing Applications. EU project by leading national HPC centres.
DFT	Density Functional Theory
EPCC	Edinburgh Parallel Computing Centre (represented in PRACE by EPSRC, United Kingdom)
FZJ	Forschungszentrum Jülich (Germany)
GFlop/s	Giga (= 10^9) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s
GRNET	Greek Research and Technology Network (Greece)
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
IBM	Formerly known as International Business Machines
JSC	Jülich Supercomputing Centre (FZJ, Germany)
LRZ	Leibniz Supercomputing Centre (Garching, Germany)
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
QCD	Quantum Chromodynamics
RZG	Rechenzentrum Garching (Germany)
STMV	Satellite Tobacco Mosaic Virus
TD-DFT	Time Dependent Density Functional Theory
TFlop/s	Tera (= 1012) 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
UEABS	Unified European Application Benchmark Suite

Executive Summary

This deliverable present the results of running the Unified European Application Benchmark Suite (UEABS), a set of 12 application codes selected in the PRACE Second Implementation Phase (PRACE-2IP) project from the pre-existing PRACE and DEISA application benchmark suites, on PRACE Tier-0 architectures. This work has been undertaken by Task 7.3 “Support for European Researchers” in the PRACE Third Implementation Phase (PRACE-3IP) project and is an extension of the work carried out in Task 7.4 “Unified European Applications Benchmark Suite for Tier-0 and Tier-1” in the PRACE Second Implementation Phase (PRACE-2IP) project. The results demonstrate that for some benchmarks there are significant differences in the performance obtained on the different architectures, and no one architecture gives the best performance on all the benchmarks.

1 Introduction

The Unified European Application Benchmark Suite [1] is a set of 12 application benchmarks, which was selected from the pre-existing PRACE [2], [3] and DEISA [4] application benchmark suites. The selection process and a description of the contents of the suite can be found in [1], and the suite is available online at http://www.prace-ri.eu/ueabs [5].

The UEABS application codes are: **ALYA**, **Code_Saturne**, **CP2K**, **GADGET**, **GENE**, **GROMACS**, **GPAW**, **NAMD**, **NEMO**, **QCD**, **Quantum Espresso** and **SPECFEM3D**.

The main purposes of the UEABS are the following:

- To provide a resource of application codes and datasets that PRACE partners can draw on for procurement purposes.
- To provide performance data on existing PRACE systems to assist users when choosing which system to apply for time on.
- To provide data for “currency conversion” of CPU hours between PRACE systems.

This deliverable is intended primarily to fulfil the second of these purposes, by reporting benchmark data from the current set of PRACE Tier-0 architectures. Some initial sample data from both Tier-0 and Tier-1 systems was included in [1]. Where appropriate (i.e. in cases where the systems have not changed significantly) this data has been re-used in this deliverable, rather than generated anew. Table 2 in Section 2 shows, for each application/architecture pair, whether the data was generated in Task 7.4 of PRACE-2IP or in Task 7.3 of PRACE-3IP.

Section 2 presents the benchmarking data, and comments on the obtained results, while Section 3 draws some conclusions and outlines future work on the benchmark suite.

2 Benchmark Results

Table 1 lists the systems used to obtain the results presented in this Section. Note that the last three systems in the list, HECToR, HYDRA and Lindgren are classed as PRACE Tier-1 systems. They have been used in these studies because their architectures are essentially the same as the Tier-0 systems HERMIT and MareNostrum, on which only limited CPU resources were available.

Details of the codes and datasets can be found in [1]. For each code in the suite there are either one or two test case datasets. Where there are two test cases, the smaller one (Test Case A) is designed to run on Tier-1 systems and scale up to around 1000 x86 cores, or equivalent. The larger one (Test Case B) is designed to run on Tier-0 systems and scale up to 10000 x86 cores or equivalent. Where there is only one test case, it is suitable for both classes of system, and should also scale to 10000 x86 cores or equivalent. In this study, since we are using Tier-0 architectures, we have used the larger datasets.

Table 2 shows the systems on which each of the applications codes were run, and the source of the data (Task 7.4 of PRACE-2IP or Task 7.3 of PRACE-3IP). Due to constraints on both CPU and human resources, and, in some cases, porting difficulties, it was not possible to run every benchmark code on each Tier-0 architecture.

For each code we present two different figures for each test case: first, a scaling plot, which is the execution time as a function of the number of CPUs (where “CPU” is used synonymously for “core”). Secondly, we show the performance (the reciprocal of execution time) per Peak-TFlop/s as function of the partition size in Peak-TFlop/s. In this figure the y-axis values are actually not meaningful but there are two reasons to show the data in this way:

(i) the shape of the curves is of interest: scaling is better if this curve does not decline too much, and ideal scaling is represented by a horizontal line, and

(ii) the performance of codes can be more fairly compared between systems with different processors.

Note that it is not possible to compare different codes on one platform using this metric.

Name	Partner	System	Total # of cores	Processor	Cores per node	Clock rate (GHz)	Peak Gflop/s per core
JUQUEEN	FZJ	IBM BlueGene/Q	458752	IBM PowerPC A2	16	1.60	12.8
FERMI	CINECA	IBM BlueGene/Q	163840	IBM PowerPC A2	16	1.60	12.8
CURIE (thin nodes)	GENCI	Bull X cluster	80640	Intel Sandy Bridge EP	16	2.70	21.6
HERMIT	HLRS	CRAY XE6	113472	AMD Interlagos	32	2.30	9.2
SuperMUC	LRZ	IBM iDataPlex	147456	Intel Sandy Bridge EP	16	2.70	21.6
MareNostrum	BSC	IBM iDataPlex	48896	Intel Sandy Bridge EP	16	2.60	20.8
HECToR	EPCC	CRAY XE6	90112	AMD Interlagos	32	2.30	9.2
HYDRA	RZG	IBM iDataPlex	9904	Intel Sandy Bridge EP	16	2.60	20.8
Lindgren	KTH	Cray XE6 MC	36384	AMD Magny-Cours	24	2.10	8.4

Table 1 List of systems used for benchmarking tests

	IBM BlueGene/Q	Bull X Cluster	IBM iDataPlex	Cray XE6
ALYA	-	CURIE	MareNostrum	Lindgren
Code_Saturne	JUQUEEN	CURIE	-	HECToR
CP2K	JUQUEEN	CURIE	SuperMUC	HECToR
GADGET	JUQUEEN	-	-	-
GENE	-	-	HYDRA	HERMIT
GPAW	-	CURIE	MareNostrum	HECToR
GROMACS	JUQUEEN	CURIE	SuperMUC	HECToR
NAMD	JUQUEEN	CURIE	SuperMUC	HECToR
NEMO	JUQUEEN	CURIE	MareNostrum	HECToR
QCD	JUQUEEN	CURIE	-	HECToR
Quantum Espresso	FERMI	CURIE	-	HECToR
SPECFEM3D	-	CURIE	-	HECToR

Table 2 Systems used for benchmarking results. Entries in blue indicate re-used data from Task 7.4 PRACE-2IP, entries in purple are new data from Task 7.3 PRACE-3IP.

2.1 ALYA

Figure 1 and Figure 2 show the results of running ALYA using Test Case B, a 552.9 million element mesh of generic elements. ALYA scales well on all three architectures, but performs significantly better on the Cray XE6 than on the other systems.

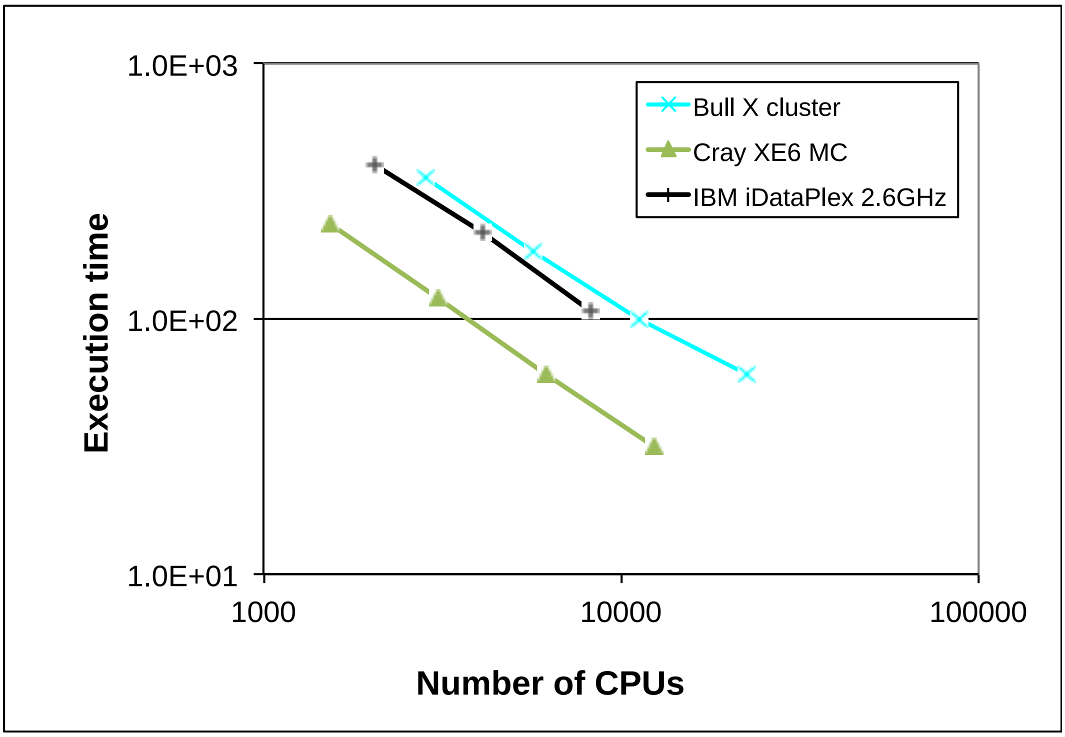

Figure 1 Execution time of ALYA for Test Case B

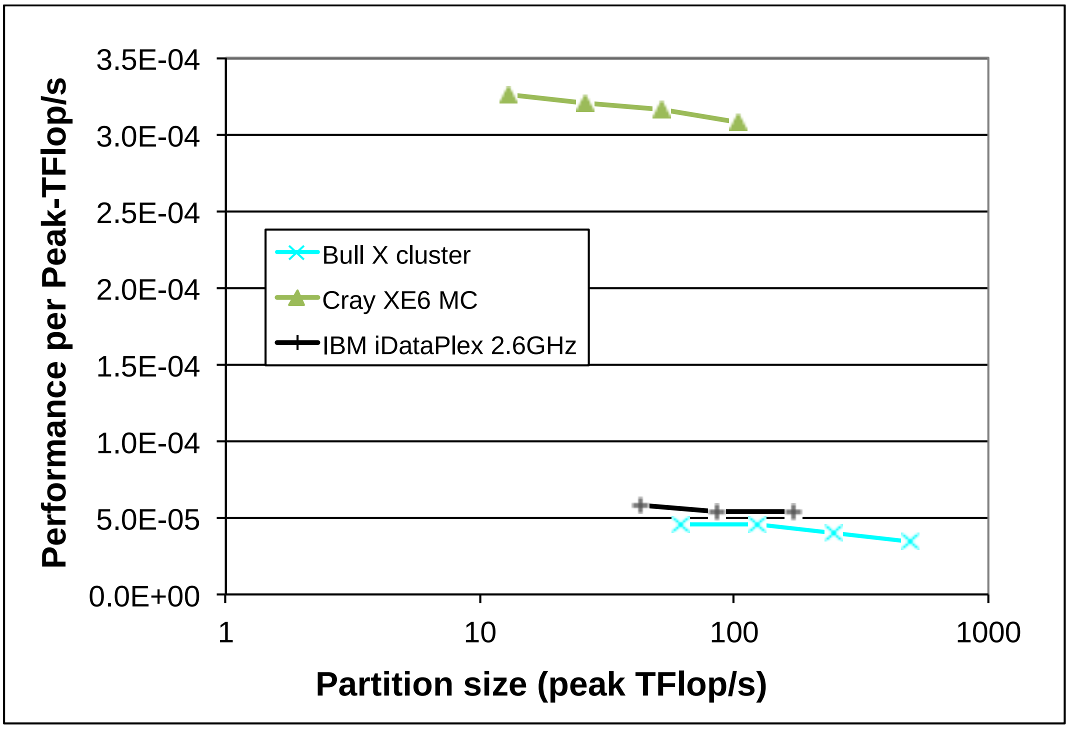

Figure 2 Performance per Peak-TFlop/s of ALYA for Test Case B

2.2 Code_Saturne

Figure 3 and Figure 4 show the results of running Code_Saturne, using Test Case A, a mesh of 51 million of hexahedral cells, with the Space-Filling Curve (SFC) partitioner. The Bull X cluster and IBM BlueGene/Q give similar performance. The Cray XE6 has a long execution time at low core counts, but scales much better, and is likely to give the best performance for larger core counts (above 4k cores, or 40 peak Tflop/s).

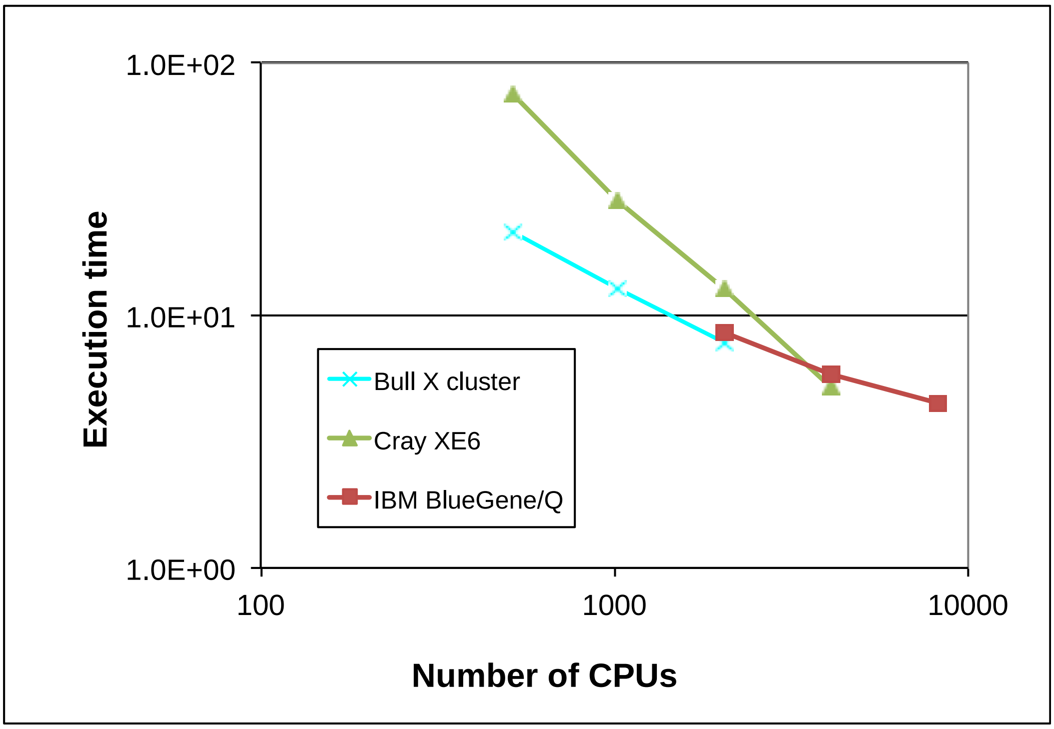

Figure 3 Execution time of Code_Saturne for Test Case A, SFC partitioner

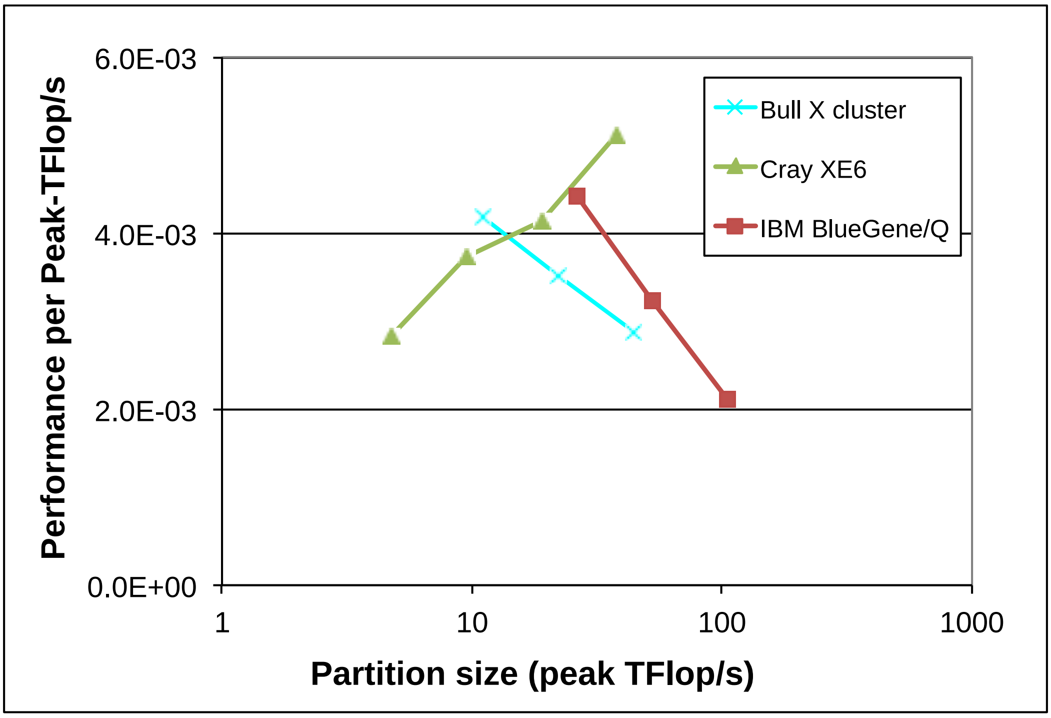

Figure 4 Performance per Peak-TFlop/s of Code_Saturne for Test Case A, SFC partitioner

2.3 CP2K

Figure 5 and Figure 6 show the results of running CP2K using Test Case B, a 216 LiH system with Hartree-Fock Exchange. The IBM iDataPlex system gives the best execution times, though in terms of performance per peak Tflop/s the Cray XE6 is a competitive. The IBM BlueGene/Q is not a good choice for this benchmark.

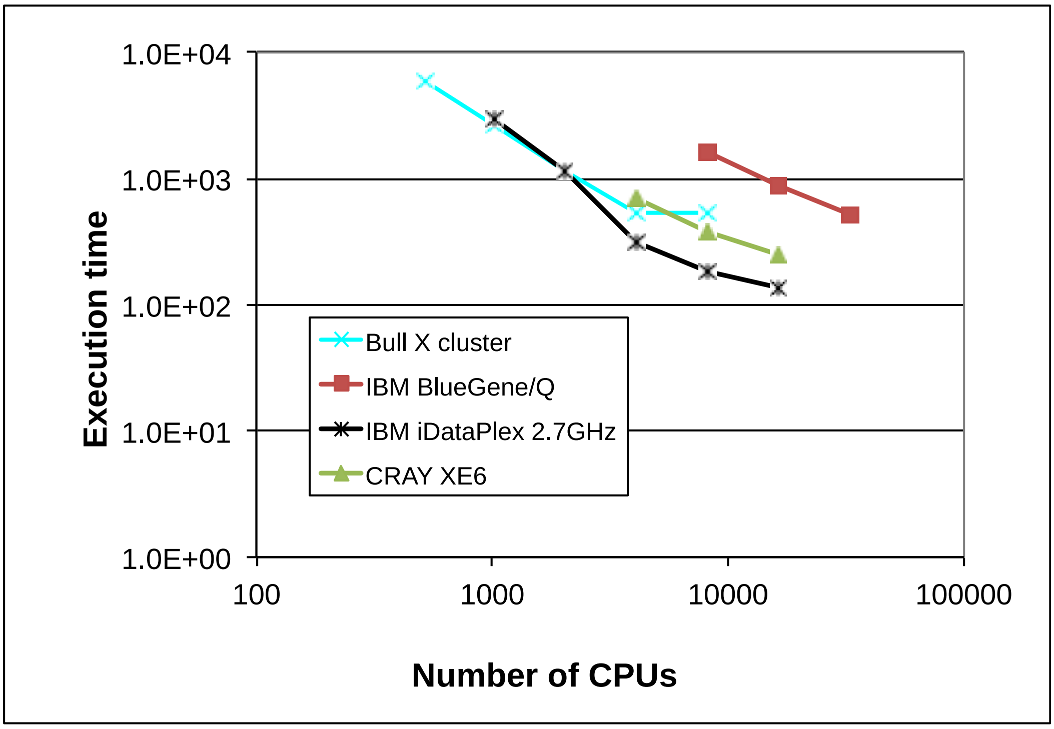

Figure 5 Execution time of CP2K for Test Case B

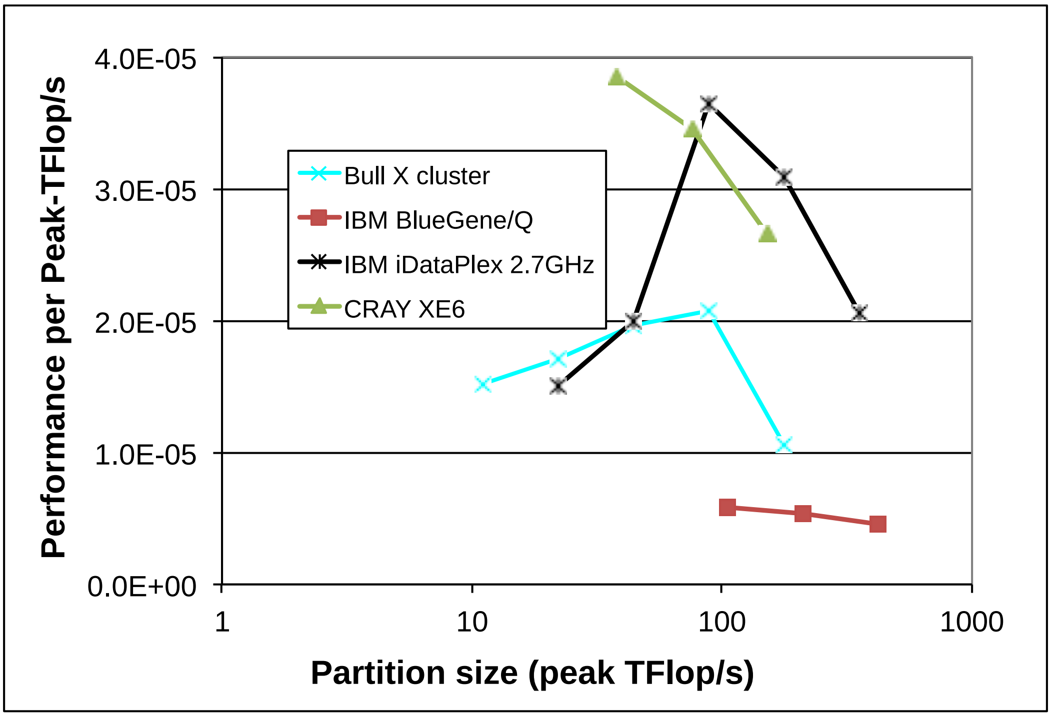

Figure 6 Performance per Peak-TFlop/s of CP2K for Test Case B

2.4 GADGET

Figure 7 and Figure 8 show the results of running GADGET, using Test Case A, a simulation consisting of 135 million particles. Scalability on the IBM BlueGene/Q up to 1024 cores is good, but data from other systems could not be obtained.

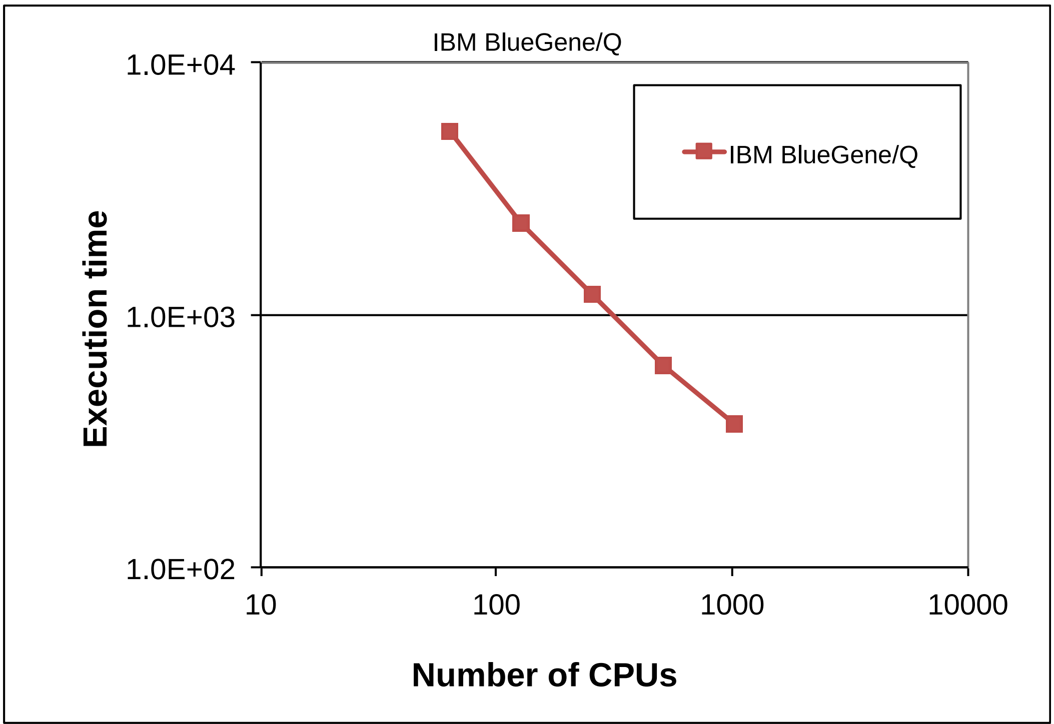

Figure 7 Execution time of GADGET for Test Case A

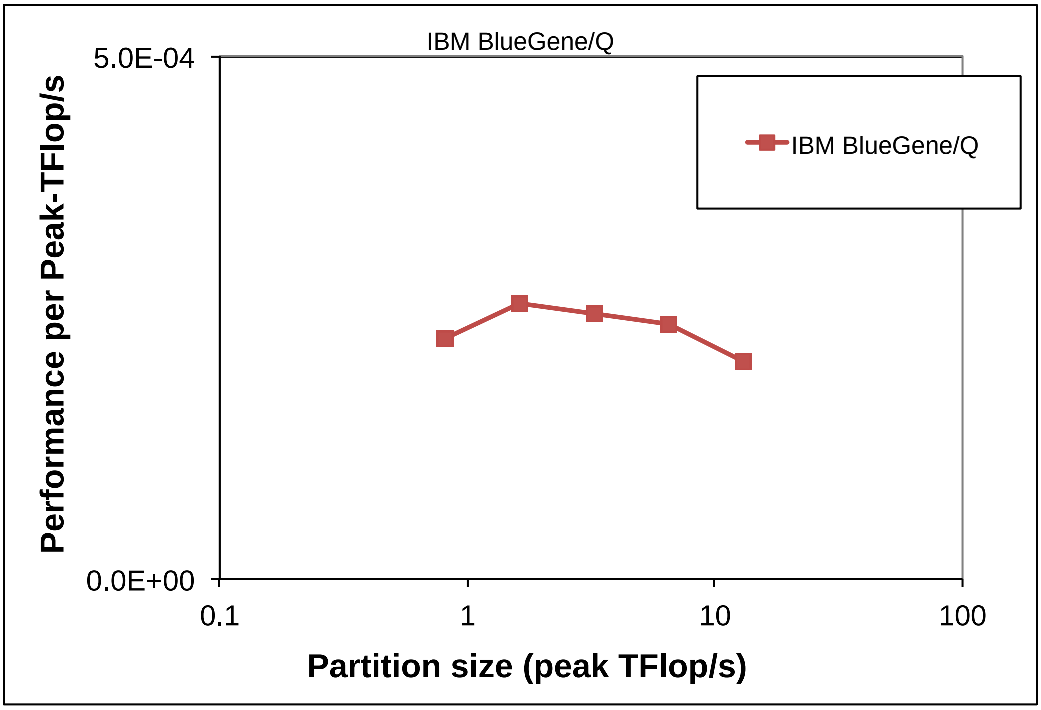

Figure 8 Performance per Peak-TFlop/s of GADGET for Test Case A

2.5 GENE

Figure 9 and Figure 10 show the results of running GENE using Test Case B, a global simulation of ion-scale turbulence in JET. The Bull X cluster and IBM iDataPlex show similar performance, which is significantly better than that on the Cray XE6.

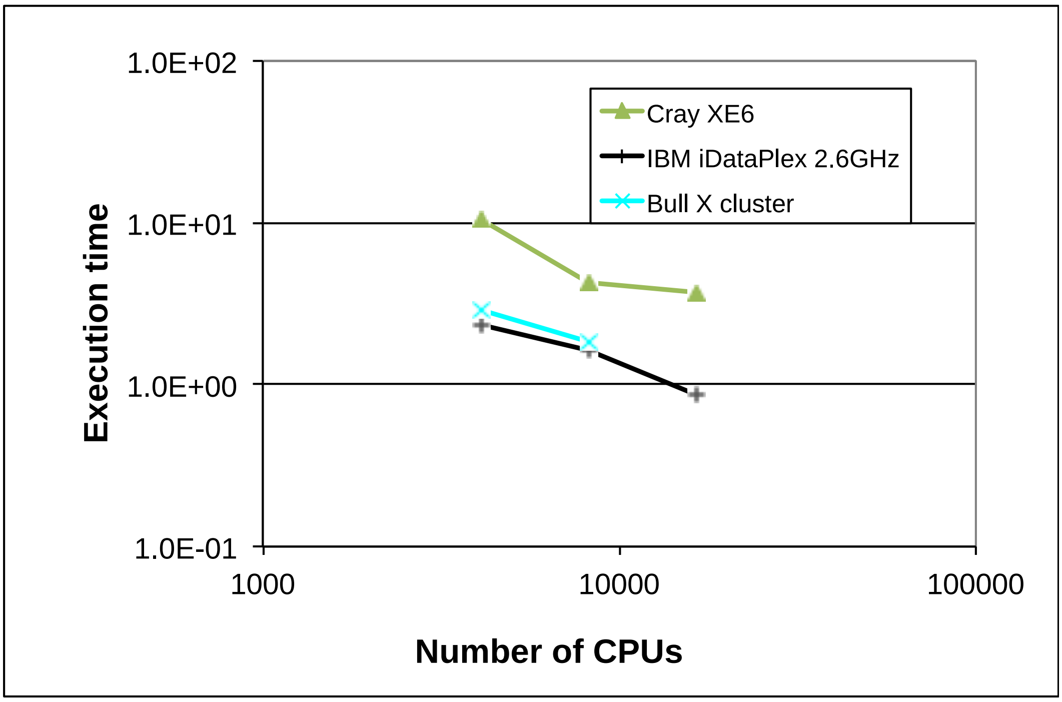

Figure 9 Execution time of GENE for Test Case B

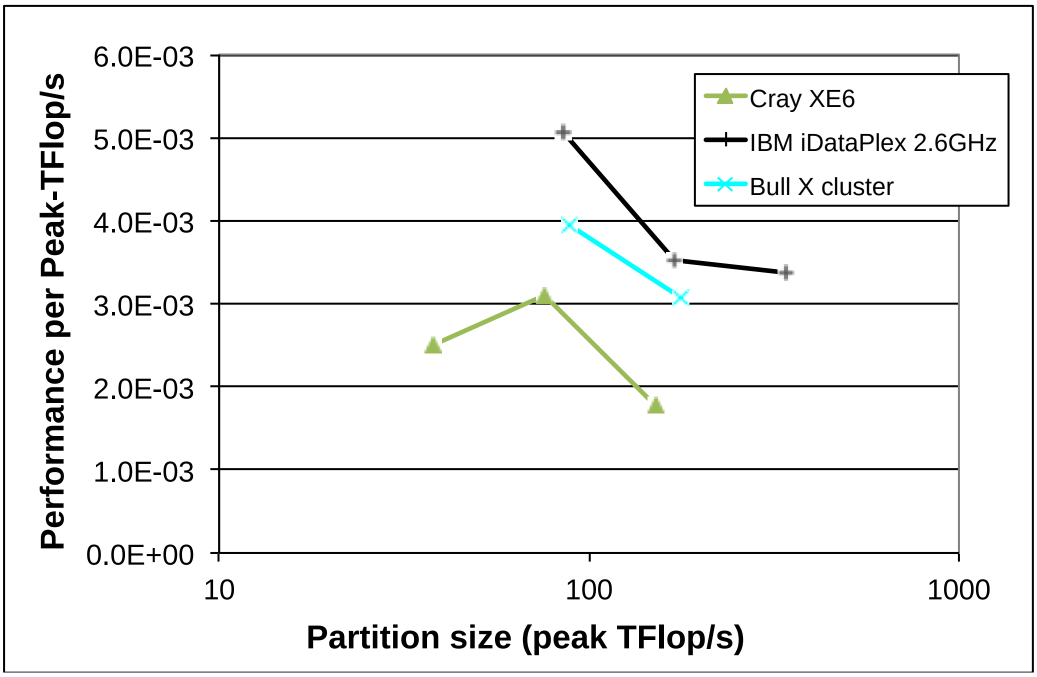

Figure 10 Performance per Peak-TFlop/s of GENE for Test Case B

2.6 GPAW

Figure 11 and Figure 12 show the results of running GPAW using Test Case B, a large TD-DFT computation. All three systems show very good scalability, with the Bull X cluster delivering the lowest execution times.

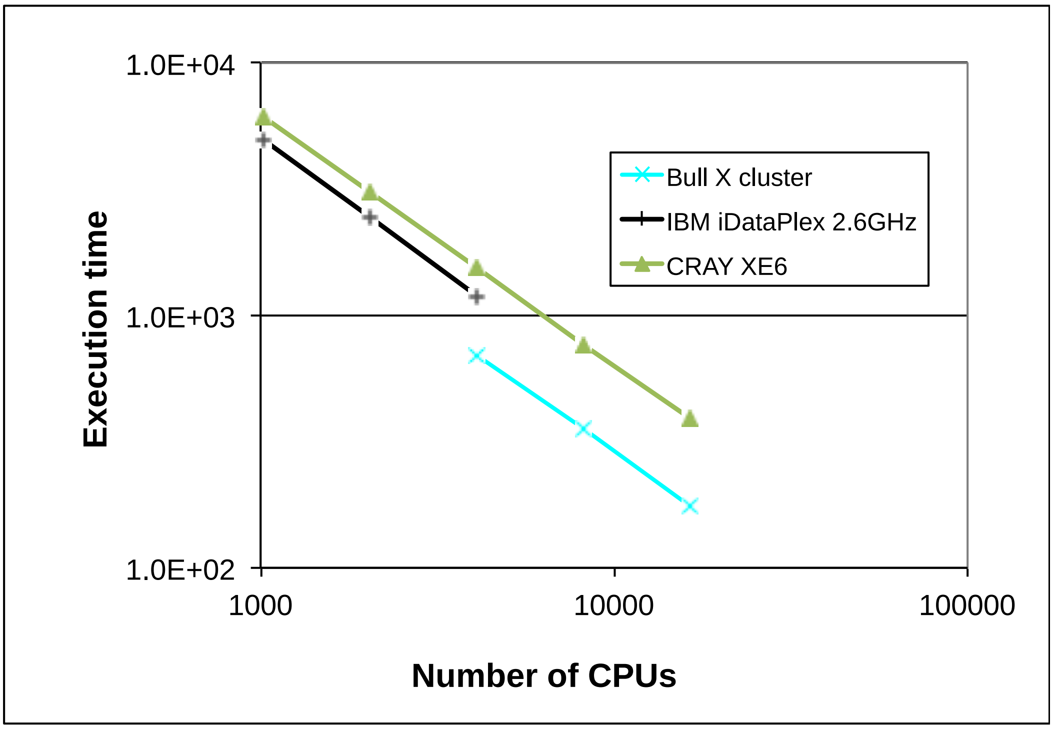

Figure 11 Execution time of GPAW for Test Case B

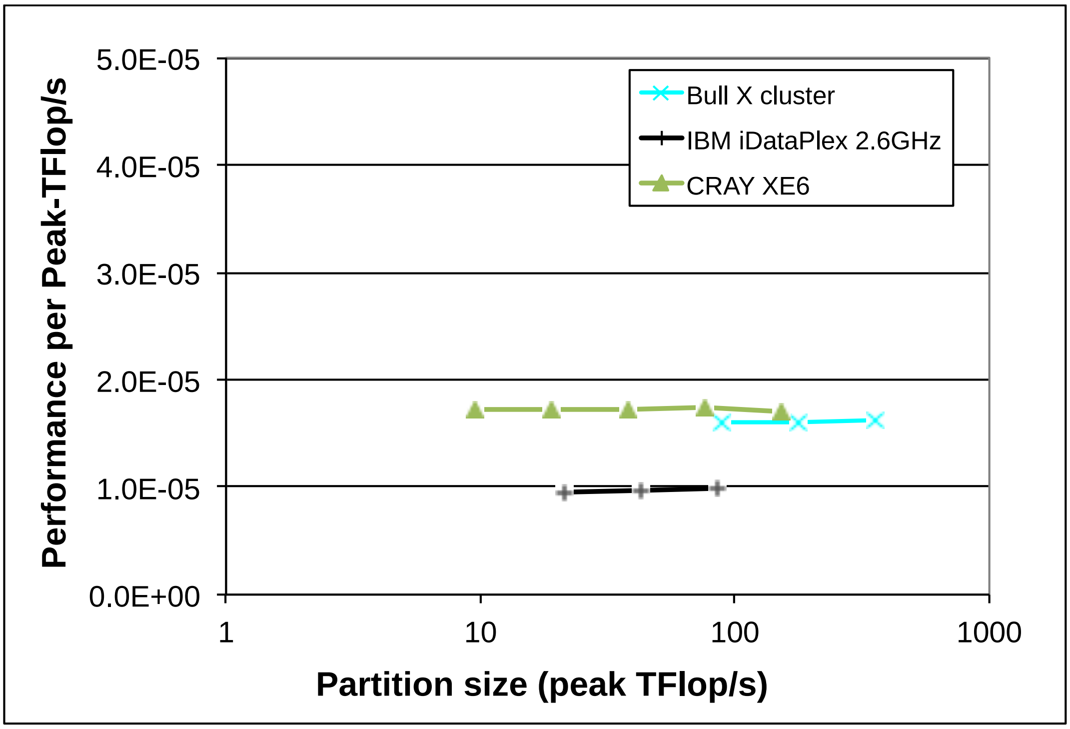

Figure 12 Performance per Peak-TFlop/s of GPAW for Test Case B

2.7 GROMACS

Figure 13 and Figure 14 show the results of running GROMACS using Test Case B, a 3.3M atom model of cellulose and lignocellulosic biomass in an aqueous solution. For this benchmark, the Bull X cluster gives the best performance and the IBM BlueGene/Q the poorest.

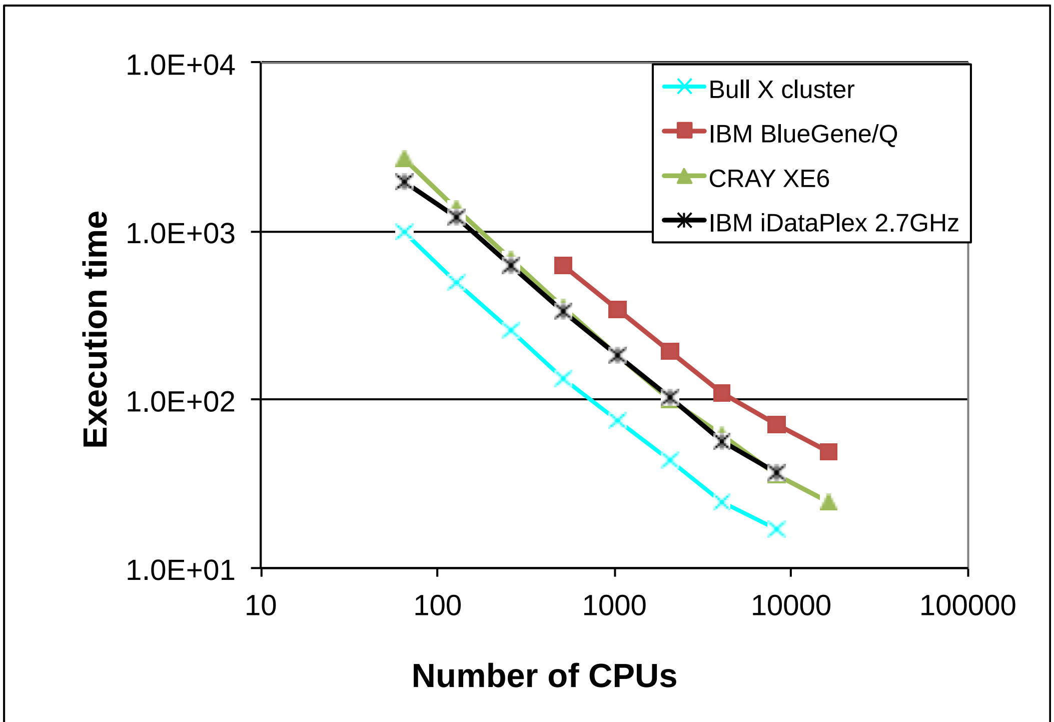

Figure 13 Execution time of GROMACS for Test Case B

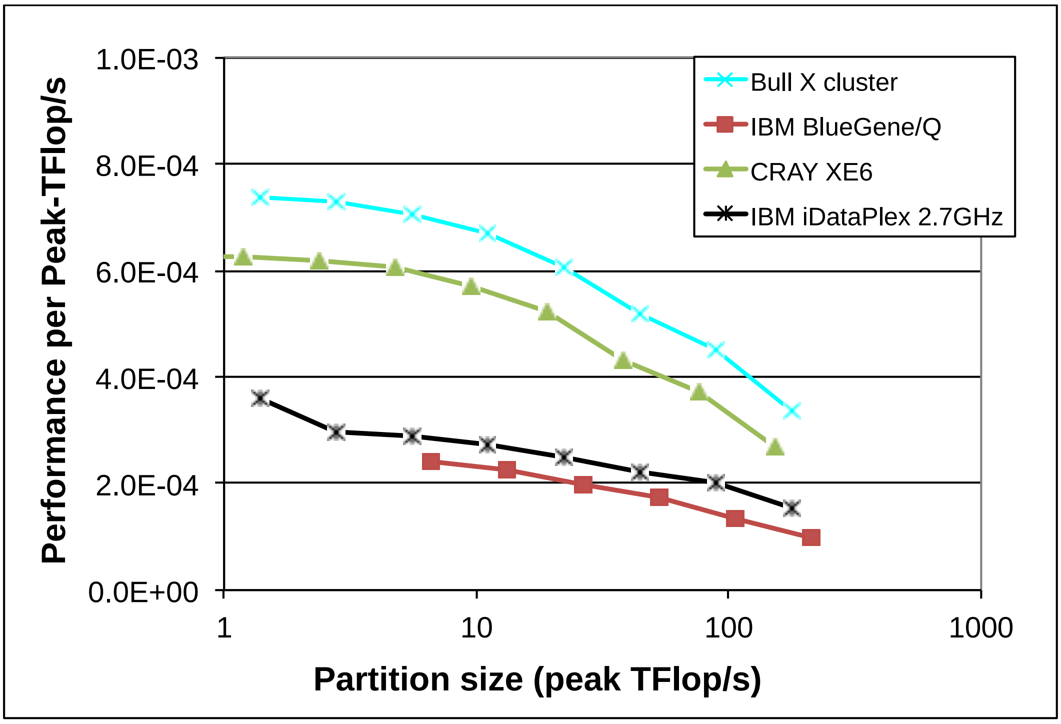

Figure 14 Performance per Peak-TFlop/s of GROMACS for Test Case B

2.8 NAMD

Figure 15 and Figure 16 show the results on running NAMD using Test Case B, a 3x3x3 replication of the "Satellite Tobacco Mosaic Virus (STMV)" dataset from the official NAMD site. On this test case, the Bull X cluster and the IBM iDataPlex give the lowest execution times on a given core count, though the Bull X cluster shows some loss of scalability above 4k cores. In terms of performance per peak Tflop/s, though, the Cray XE6 does best, while the IBM BlueGene/Q is the least effective architecture by both metrics.

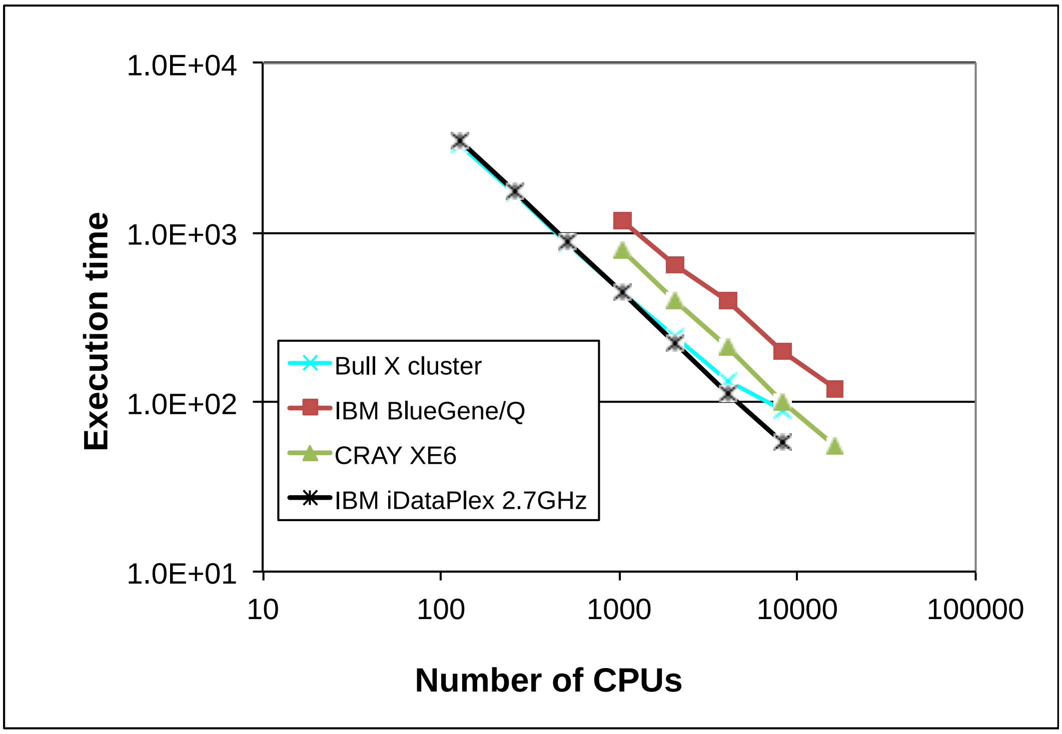

Figure 15 Execution time of NAMD for Test Case B

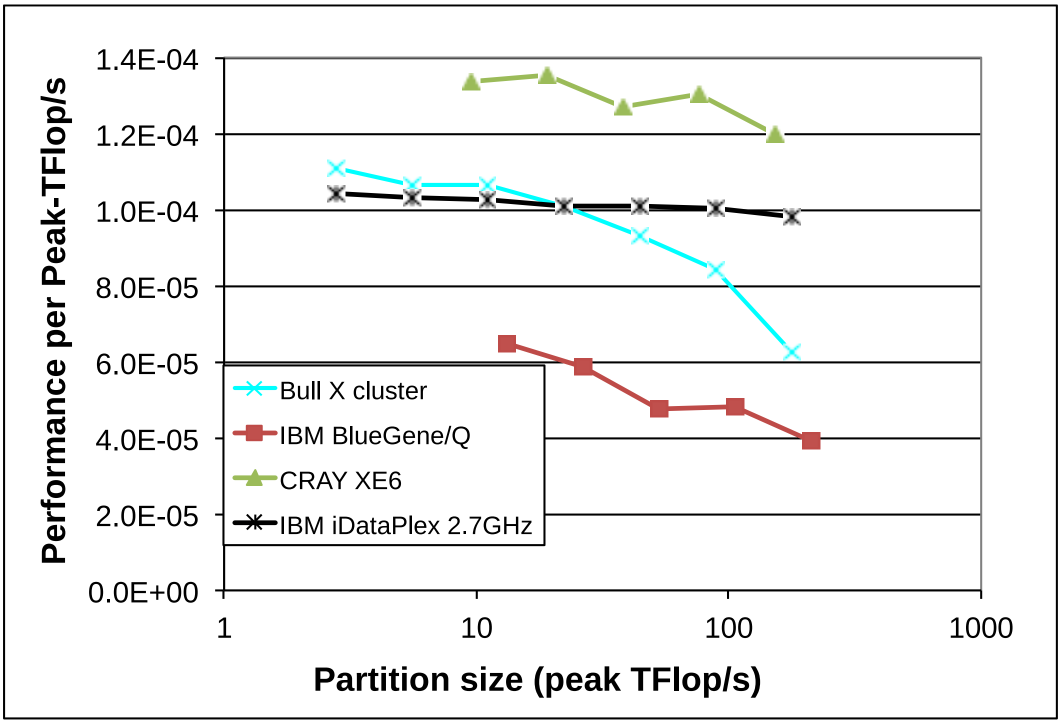

Figure 16 Performance per Peak-TFlop/s of NAMD for Test Case B

2.9 NEMO

Figure 17 and Figure 18 show the results on running NEMO using Test Case A, a 1/12° global configuration with 75 vertical levels. This benchmark does not scale especially well on any of the architectures. The lowest execution times and best performance per peak Tflop/s are on the Bull X cluster and IBM iDataPlex.

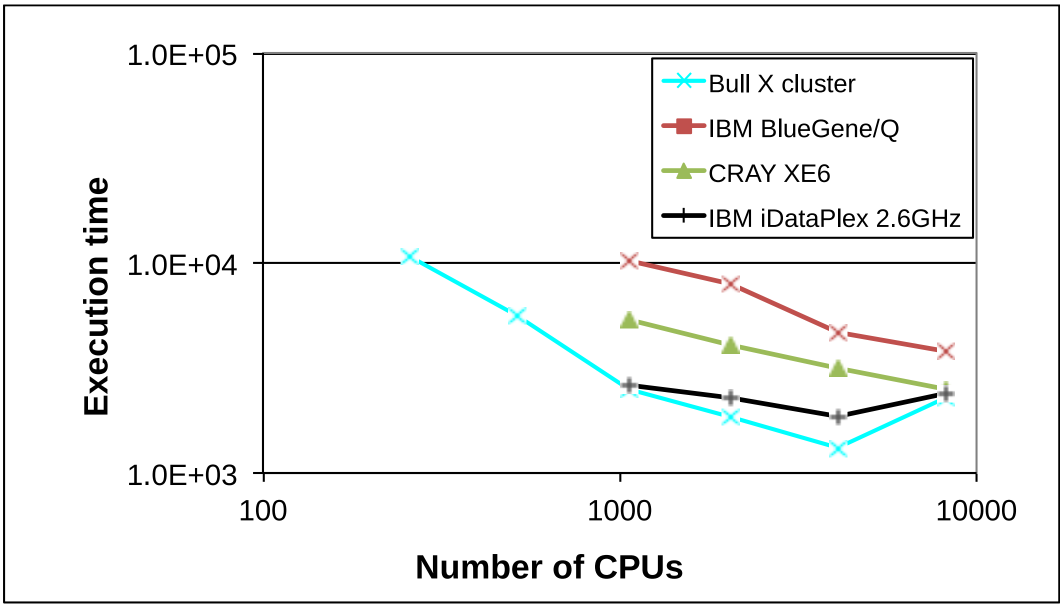

Figure 17 Execution time of NEMO for Test Case A

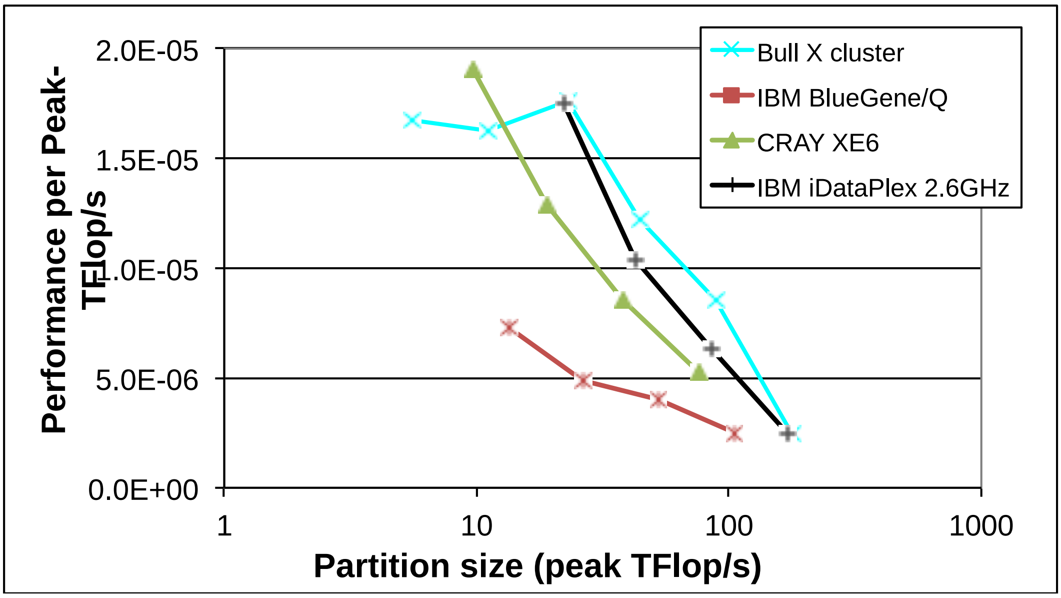

Figure 18 Performance per Peak-TFlop/s of NEMO for Test Case A

2.10 QCD

The QCD benchmark is, unlike the other benchmarks in the PRACE application benchmark suite, not a full application but a set of 5 kernels that are representative of some of the most compute-intensive parts of QCD calculations.

Each of the 5 kernels has one test case:

Kernel A is derived from BQCD (Berlin Quantum ChromoDynamics program), a hybrid Monte-Carlo code that simulates Quantum Chromodynamics with dynamical standard Wilson fermions. The computations take place on a four-dimensional regular grid with periodic boundary conditions. The kernel is a standard conjugate gradient solver with even/odd pre-conditioning. Lattice size is 32^2 x 64^2.

Kernel B is derived from SU3_AHiggs, a lattice quantum chromodynamics (QCD) code intended for computing the conditions of the Early Universe. Instead of "full QCD", the code applies an effective field theory, which is valid at high temperatures. In the effective theory, the lattice is 3D. Lattice size is 256^3.

Kernel C Lattice size is 8^4. Note that Kernel C can only be run in a weak scaling mode, where each CPU stores the same local lattice size, regardless of the number of CPUs. Ideal scaling for this kernel therefore corresponds to constant execution time, and performance is simply the reciprocal of the execution time.

Kernel D consists of the core matrix-vector multiplication routine for standard Wilson fermions. The lattice size is 64^4 .

Kernel E consists of a full conjugate gradient solution using Wilson fermions. Lattice size is 64^3 x 3.

Figure 19 through to Figure 28 show the results of running the five kernels. On Kernels A and E there is little to choose between the three systems tested. On Kernel B the Bull X cluster is fastest at low core counts, but does not scale as well as the other two systems. On Kernels C and D, there is little to choose between the Bull X cluster and Cray XE6, but the IBM BlueGene/Q shows significantly poorer performance.

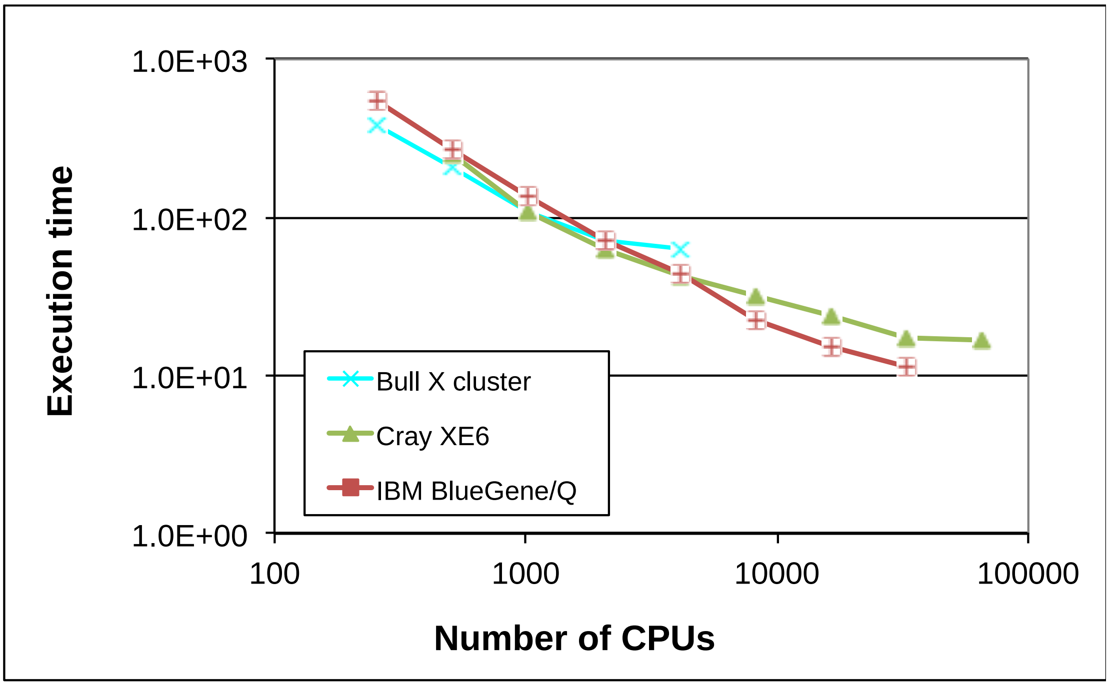

Figure 19 Execution time of QCD for Kernel A

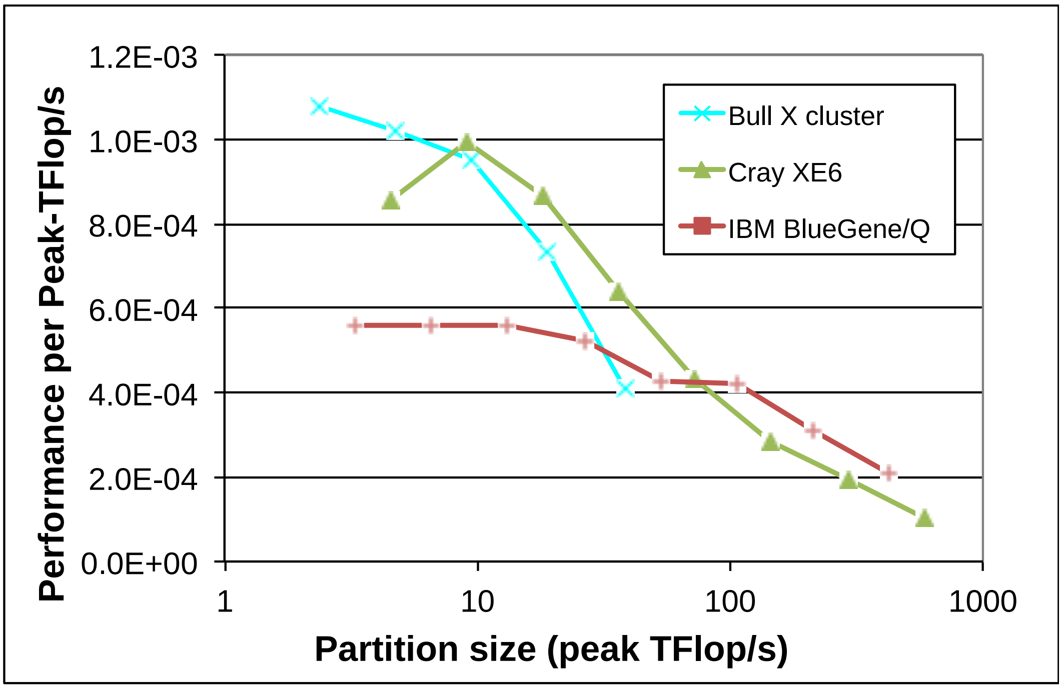

Figure 20 Performance per Peak-TFlop/s of QCD for Kernel A

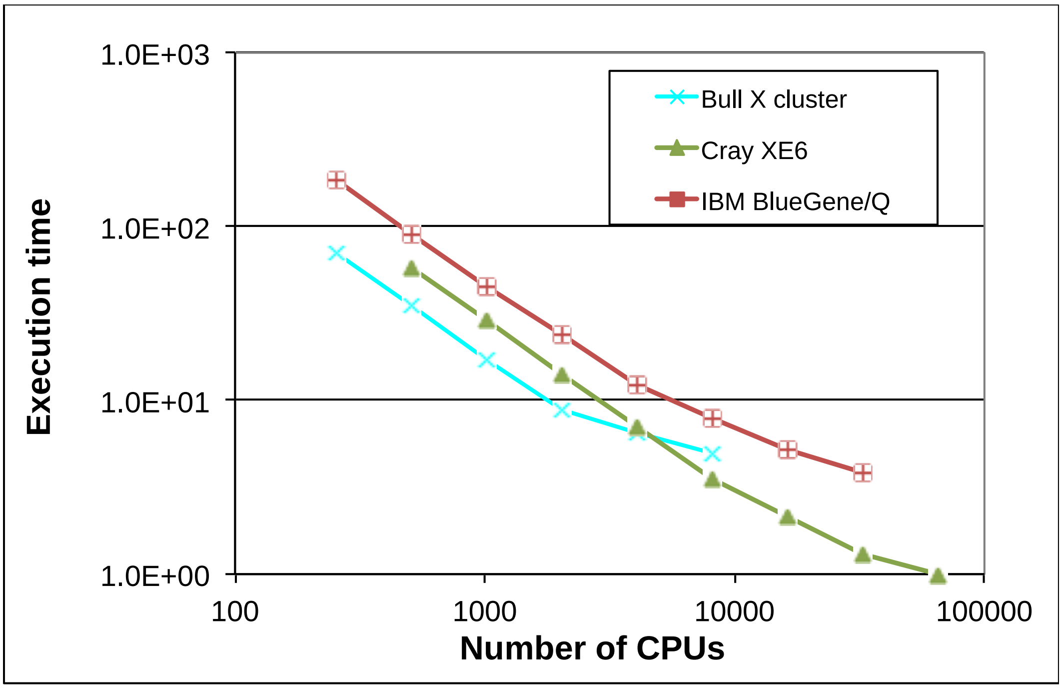

Figure 21 Execution time of QCD for Kernel B

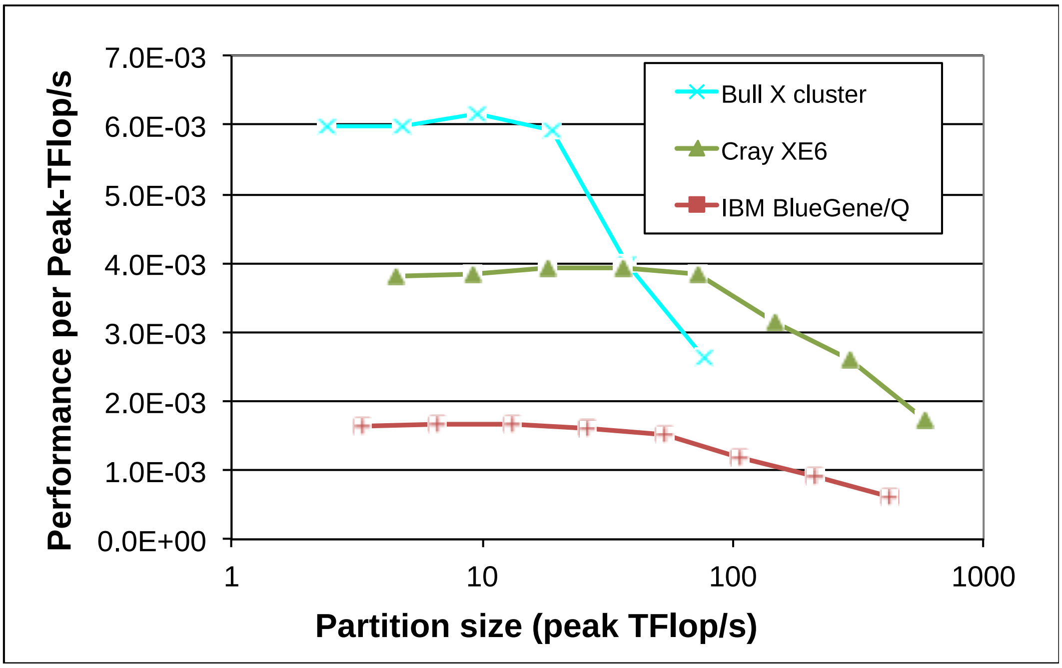

Figure 22 Performance per Peak-TFlop/s of QCD for Kernel B

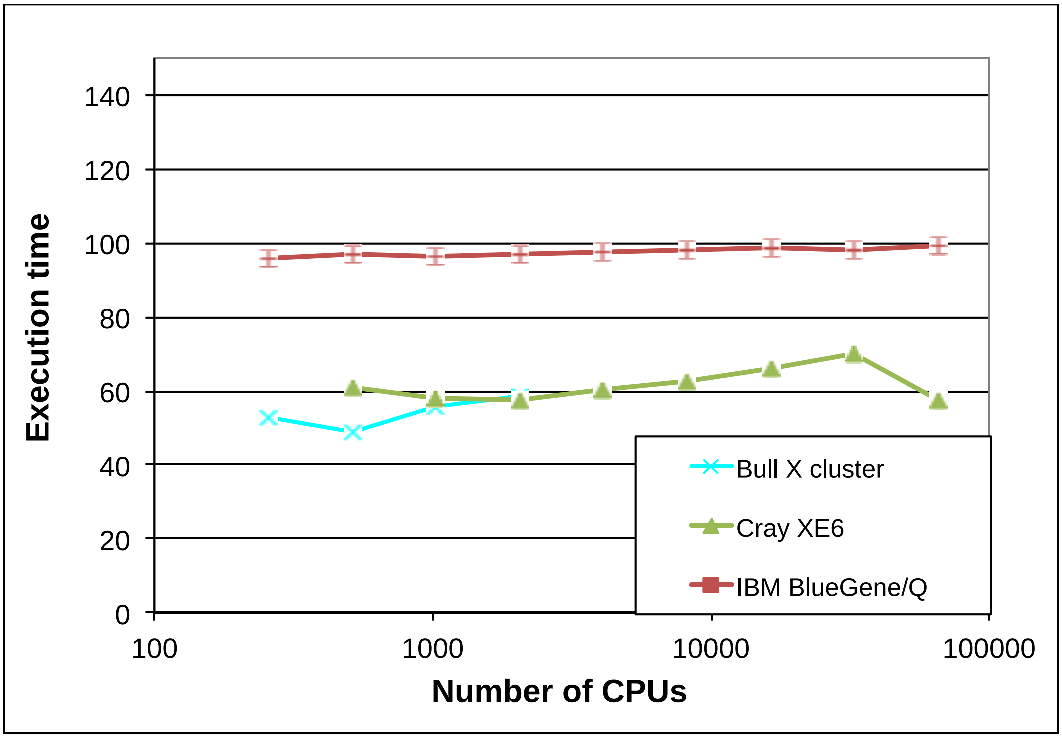

Figure 23 Execution time of QCD for Kernel C

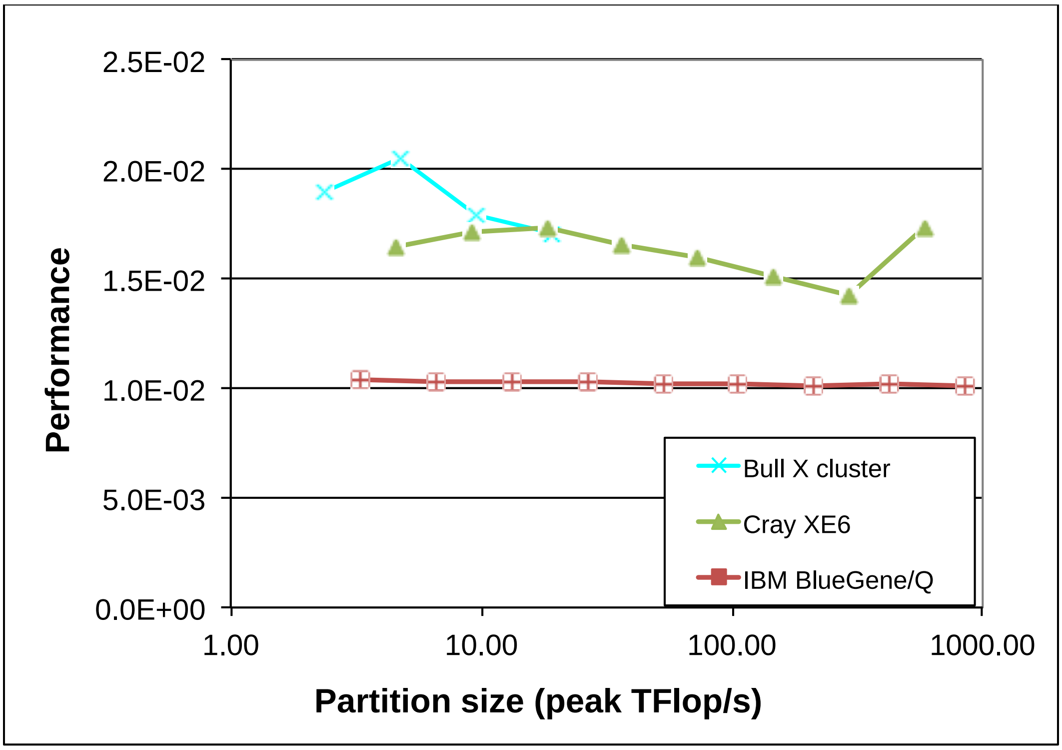

Figure 24 Performance of QCD for Kernel C

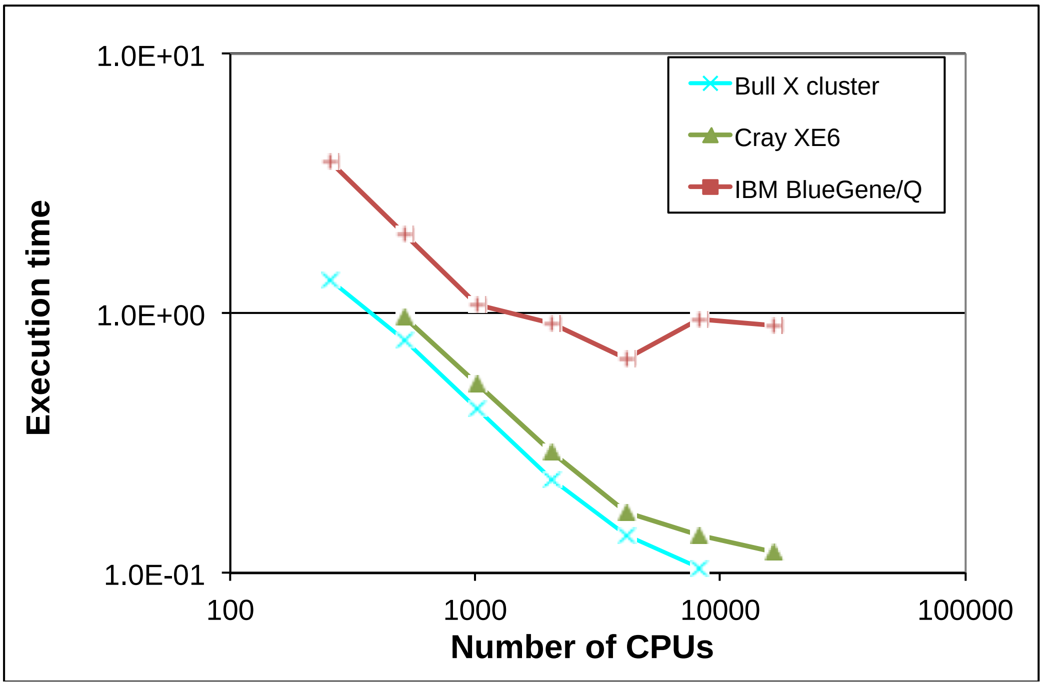

Figure 25 Execution time of QCD for Kernel D

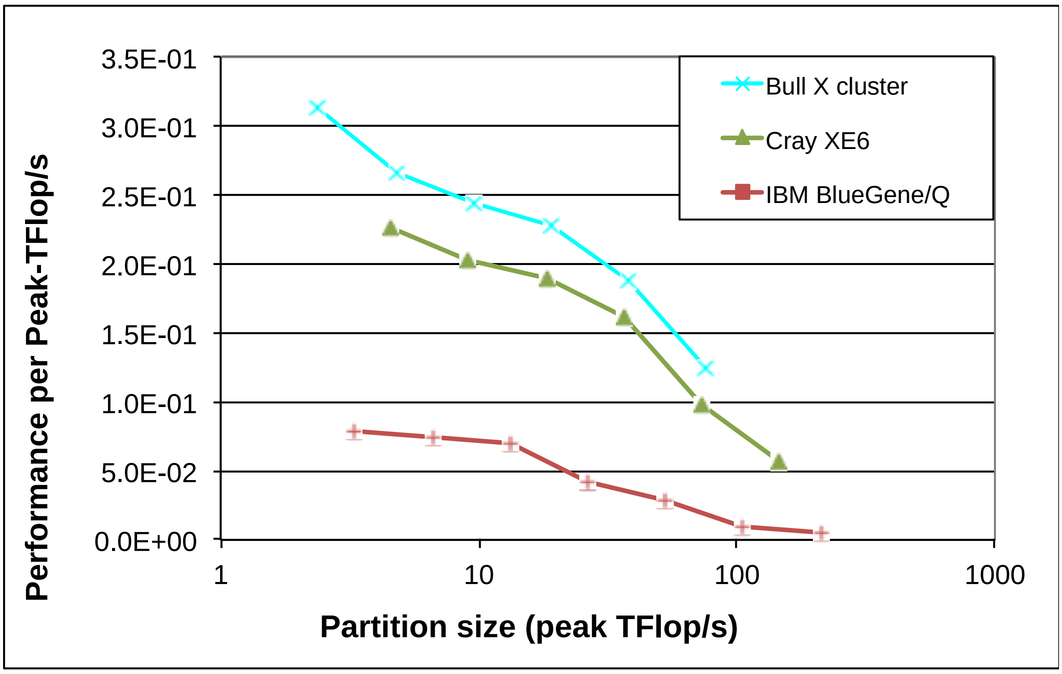

Figure 26 Performance per Peak-TFlop/s of QCD for Kernel D

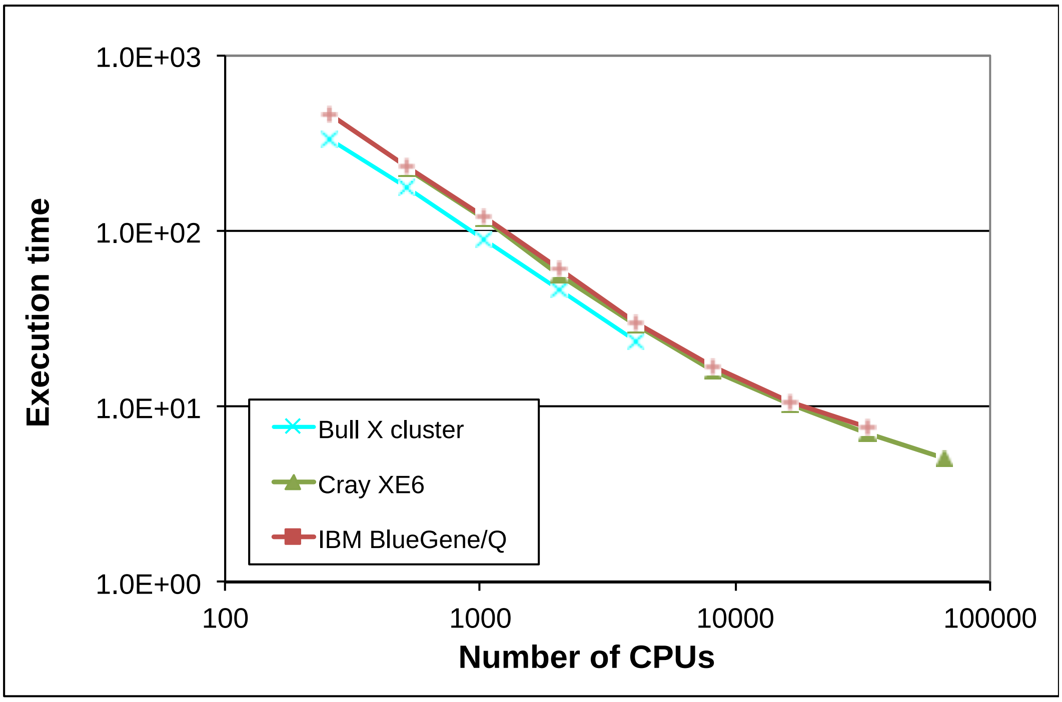

Figure 27 Execution time of QCD for Kernel E

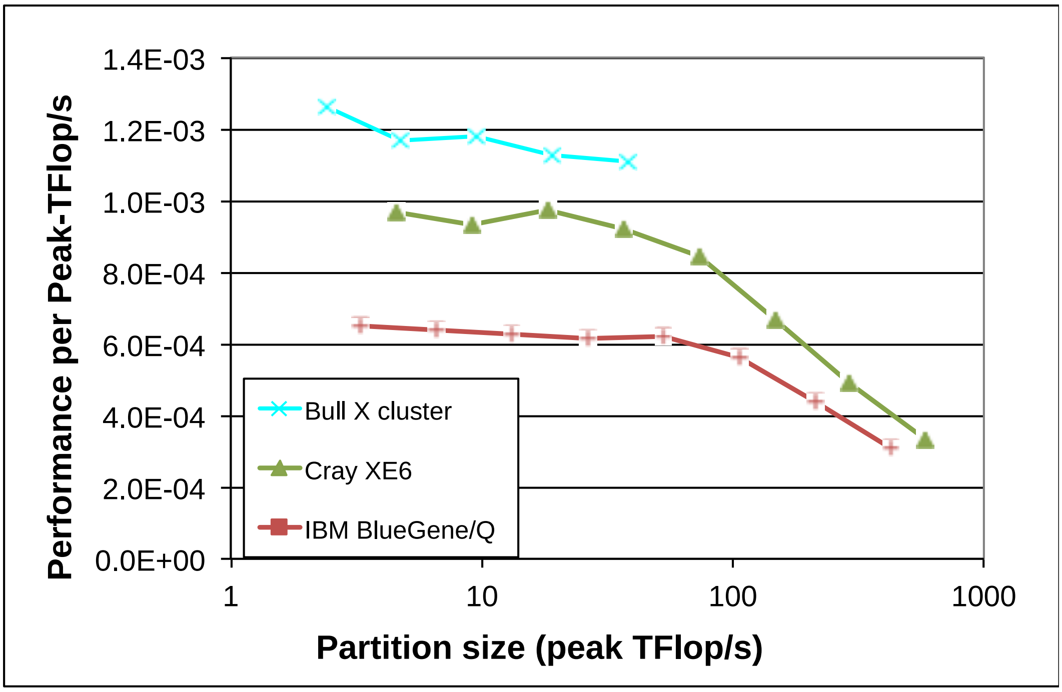

Figure 28 Performance per Peak-TFlop/s of QCD for Kernel E

2.11 Quantum Espresso

Figure 29 and Figure 30 show the results of running Quantum Espresso using Test Case B, which is based on two iterations of a SCF calculation on a functionalised carbon nanotube with a total of 1532 atoms. This test case does not scale very well at high core counts. The lowest execution times, but also the poorest scalability, are obtained on the Cray XE6.

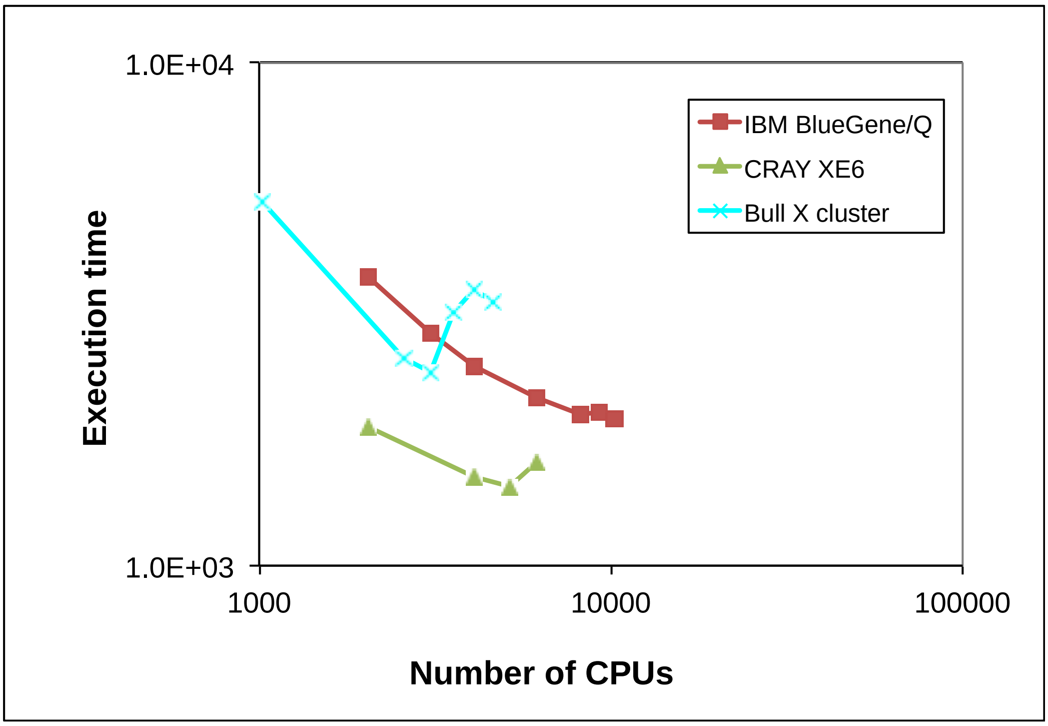

Figure 29 Execution time of Quantum Espresso for Test Case B

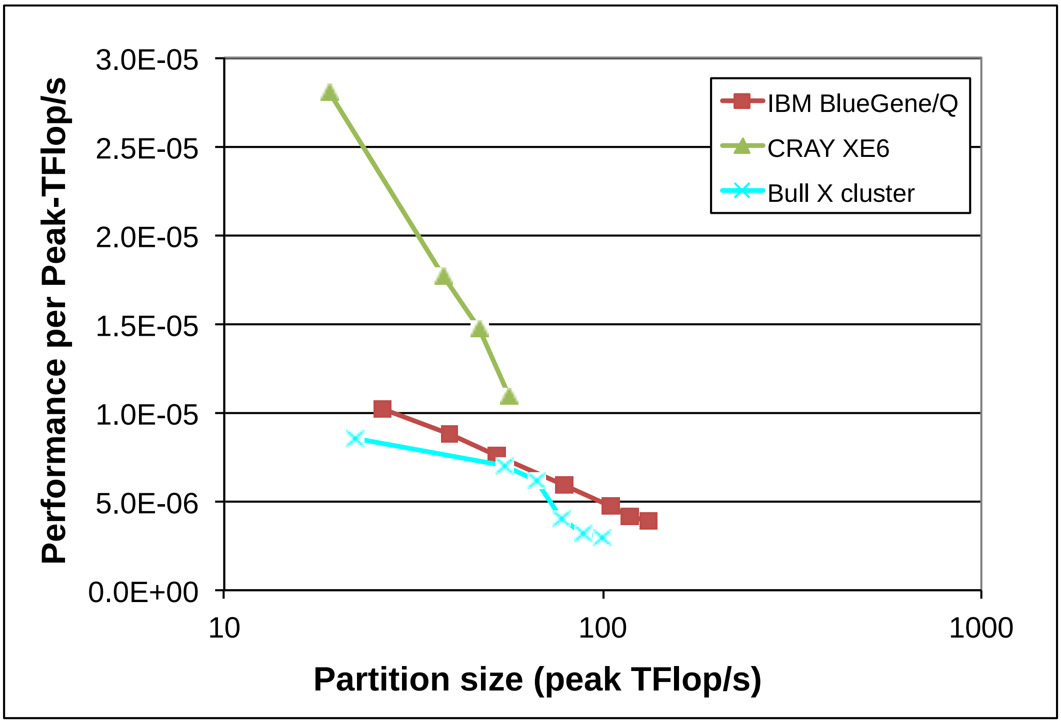

Figure 30 Performance per Peak-TFlop/s of Quantum Espresso for Test Case B

2.12 SPECFEM3D

Test Cases A and B for SPECFEM3D simulate the earthquake of June 1994 in Northern Bolivia at a global scale with the spherically symmetric isotropic IASP91 model using respectively 4.60 and 48.2 billion degrees of freedom. The version of the program used is SPECFEM3D_GLOBE_V5.1.5. It has been possible to run both test cases on the Bull X Cluster. On this system, Test Case A had a mean elapsed time per time step of 0.89 second on 864 cores and Test Case B of 0.1 second on 11616 cores. Test Case A was also run on the Cray XE6 using 864 cores, with a mean elapsed time per time step of 2.63 seconds.

3 Conclusions and Future Work

We have presented the results of running the UEABS benchmark suite on the current set of Tier-0 architectures. We have shown that the benchmark performance can vary significantly between the different architectures, both in terms of execution time on a given number of cores, and in terms of performance on a system partition with a given peak floating point performance. Furthermore, no one system gives the best results on all the benchmarks. It is therefore important for users to choose the best system for a given application in order to maximise the scientific output for a given amount of compute resource.

The remaining work in Task 7.3 on the benchmark suite will be to conduct a review of the contents, with the intention of keeping versions of codes (in the cases where code is hosted on the PRACE web site, and not on the applications own web site) and versions of datasets up-to-date and relevant to the scientific community. The revised version of the benchmark suite will be published on the PRACE web site, in fulfilment of Milestone MS73 “Update on Unified European Applications Benchmark Suite”.