Institution: University of Cambridge		
Unit of Assessment: 10 Mathematics		
Title of case study: Advancing High Performance Computing and Visualisation through the collaboration between the COSMOS supercomputer team, Hewlett Packard Enterprise and Intel		
Period when the underpinning research was undertaken: 1/1/2000 – 1/6/2018		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Professor E.P.S. Shellard	Director, Centre for Theoretical Cosmology (CTC)	1/1/2000 – present
Dr Kacper Kornet	COSMOS IPCC Manager, Faculty HPC	1/3/2015– present
Dr Juha Jäykkä	COSMOS System Manager	1/3/2015– 31/12/2017
James Briggs	COSMOS Parallel Programmer	1/3/2013 – 31/3/2016
Dr James Fergusson	CTC Lecturer, CDT Deputy Director	1/10/2009 - present
Dr Pau Figueras	EPSRC & Stephen Hawking Advanced Fellow	1/10/2010 – 30/09/2015
Dr Ulrich Sperhake	University Reader	1/10/2012 – present
Period when the claimed impact occurred: 11/8/2013 – 31/7/2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact (indicative maximum 100 words)

The Centre for Theoretical Cosmology (CTC) at the University of Cambridge is engaged in long-standing industrial collaborations with both Intel and HPE/SGI, which have brought significant benefits to these commercial partners through co-design and improvement of high performance computing (HPC) hardware and software.

[I1] CTC has collaborated with Hewlett Packard Enterprise (HPE), resulting in new capabilities and improvements for the SGI UV2 and UV3 supercomputers. These were incorporated in the HPE Superdome Flex server, HPE's flagship in-memory system deployed by commercial customers with large data flows worldwide.

[I2] CTC has collaborated with Intel on advancing hardware acceleration on HPC systems, as well as creating in-situ visualisation capabilities. Collaborative innovations have been released in open-source libraries/applications used worldwide and incorporated in 150-200 commercial graphics tools, and in Top 10 Supercomputers.

2. Underpinning research (indicative maximum 500 words)

The CTC at the University of Cambridge has facilitated development of the necessary HPC ecosystem for 20 years through hosting and operating the COSMOS National Cosmology Supercomputer on behalf of the wider UK research community.

2.1 Context: An important component of CTC research is developing high performance computing (HPC) capabilities to advance the confrontation between mathematical theory and new observational data provided by huge ongoing cosmological surveys. The CTC has contributed *via* the Planck satellite analysis pipeline MODAL [R1, R2], designed to analyse the cosmic microwave background, and the computer code GRChombo [R3], designed to study black holes. Both contributions were only possible because of the effective use of cutting-edge HPC. The HPE/Intel collaboration has enabled the co-design of hardware to meet CTC requirements, with early access to the latest technologies and HPC software advances with expert developer support.

2.2 Co-design of in-memory systems: Impetus for new HPC architecture designs came from the extreme challenges faced by CTC research on Planck satellite maps of the cosmic microwave background, i.e. the MODAL pipeline. In 2010, Fergusson and Shellard published a general methodology [R1], which they applied to satellite maps produced by the Wilkinson Microwave Anisotropy Probe in 2012. However, they faced a huge computational task scaling this up by nearly two orders of magnitude for the Planck satellite analysis. The MODAL pipeline contained conflicting requirements; the analysis of the cosmic microwave background (pink, fig.1) required terabytes of shared memory, while theoretical modelling (blue, fig.1) needed low-memory compute-intensive integrations. This required the CTC, with its industrial collaborators, to define a unique new HPC architecture that combined a large shared-memory server from SGI, and hardware acceleration from Intel's many-core Xeon Phi co-processors. On delivery in 2012, the hybrid COSMOS UV2 supercomputer was the largest shared-memory system in the world (15TB) and the first to incorporate hardware acceleration. The resulting papers enabled by this system became some of the highest cited Planck science results [R2].

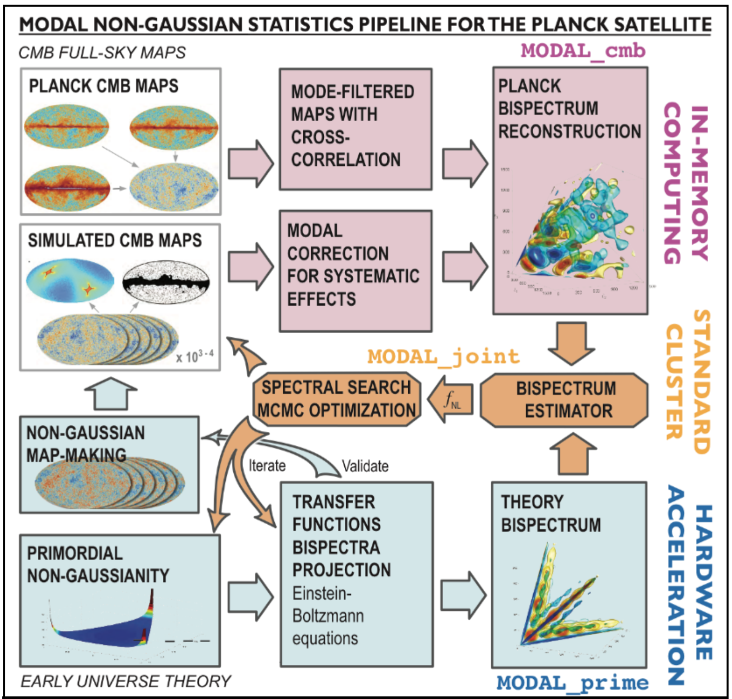

Figure 1. Planck workflow adapted from Shellard

2.3 Exploiting hardware acceleration: Apart from the codes MODAL [R1], designed to analyse Planck satellite maps of the cosmic microwave background, and GRChombo [R3], designed to study black holes, the CTC also developed the WALLS code [R4], designed to explore *topological defects* that resulted from phase transitions in the early Universe. To be run effectively, all three codes required dramatic speed-ups, which were to be achieved using new offload programming models for the latest Intel accelerators hosted on the hybrid COSMOS platform. This succeeded (in collaboration with Intel software engineers) through vectorisation and parallel algorithms [R5, R6], while identifying new offload capabilities vital for in-memory acceleration [R4, E3]. A key highlight was a speed-up exceeding a factor of 100 for the Planck MODAL pipeline [R6] (which received the HPCwire Award in 2015 [E10]). Impressive speedups were also achieved both in WALLS and in GRChombo, which since 2018 has become an open source project supported by a growing number of numerical relativity/cosmology developers [R3]. These codes have been used to gain new physical insights in areas previously intractable.

2.4 Developing in-situ visualisation: Visualisation is an essential component of CTC research, but requires unprecedented computing resources. This need has driven the CTC collaboration with Intel (since 2014) to pioneer *in-situ visualisation:* a new paradigm, which analyses data “on the fly”, as it is generated, rather than *post hoc* as was traditionally the case (see section [I2]). In-situ visualisation greatly reduces analysis timescales and has led to a wealth of scientific results, including new insights into the conjectured exponential expansion of the early Universe gleaned from Planck data [R2, R7], and the dynamics of *cosmic strings* simulated with GRChombo [R8].

3. References to the research (indicative maximum of six references)

[R1] J. R. Fergusson, L. Liguori and E.P.S. Shellard, General CMB and primordial bispectrum estimation: Mode expansion, map making and measures of F_{NL}, Physical Review D 82, 023502 (2010). arXiv:0912.5516, DOI: 10.1103/PhysRevD.82.023502.
[R2] Planck Collaboration: Ade, P., *et al.* (including J. Ferguson and E.P.S. Shellard), *Planck* 2015 results XVII. Constraints on primordial non-Gaussianity", Astronomy & Astrophysics, 594 A17 (2016). arXiv:1502:01592, DOI: 10.1051/0004-6361/201525836
[R3] K. Clough, Pau Figueras, H. Finkel, M. Kunesch, E. A. Lim, S. Tunyasuvunakool, GRChombo : Numerical relativity with adaptive mesh refinement, Class. Quant. Grav. 32 (2015) 24, 245011. arXiv:1503.03436.
[R4] J. Briggs, S. J. Pennycook, E.P.S. Shellard, C.J.A.P. Martins, Unveiling the Early Universe: Optimizing Cosmology Workloads for Intel Xeon Phi Coprocessors in an SGI UV2000 System, Tech.Rep.(SGI/Intel White Paper) (2014).
[R5] J. Briggs, S. J. Pennycook, J. R. Fergusson, J. Jäykkä, E. P. S. Shellard, "Cosmic Microwave Background Analysis: Nested Parallelism In Practice”,
Chapter 10 in *High Performance Parallelism Pearls Volume Two*, 1st Edition, Multicore and Many-core Programming Approaches, Jim Jeffers, James Reinders (eds.) July 2015, Morgan Kaufmann Print Book (ISBN: 9780128038192).
[R6] J. P. Briggs, S. J. Pennycook, J. Jäykkä, J. R. Fergusson, E. P. S. Shellard, Separable projection integrals for higher-order correlators of the cosmic microwave sky: Acceleration by factors exceeding 100, J. Comput. Phys. 301, 285 – 300 (2016). arXiv:1503.08809, DOI: 10.1016/j.jcp.2016.01.019
[R7] A. Lazanu, T. Giannantonio, M. Schmittfull, E.P.S. Shellard, Matter bispectrum of large-scale structure: Three-dimensional comparison between theoretical models and numerical simulations, Physical Review D 93, 083517 (2016). arXiv:1510.04075,
DOI: 10.1103/PhysRevD.93.083517
[R8] A. Drew, E. P.S Shellard, Radiation from Global Topological Strings using Adaptive Mesh Refinement: Methodology and Massless Modes(2019), pre-print arxiv:1910.01718v1

Research outputs are published in peer-reviewed journals.

4. Details of the impact (indicative maximum 750 words)
The CTC has longstanding industrial collaborations with Intel (since 2003) and HPE/SGI (since 1997 with SGI acquired by HPE in 2016). These collaborations are characterised by the CTC’s continual, intimate involvement in development of both hardware and software. CTC codes are used “*to validate and stress test these early access systems*” [E1]. Impacts from this collaboration include:

[I1] Transforming HPC hardware

In-memory SGI systems with hardware acceleration: CTC played a key part in SGI’s development of successive in-memory server architectures through design specification and early-access testing. CTC’s requirement for integration of Intel hardware acceleration (section 2.2) in the next shared-memory SGI COSMOS supercomputer led to the new *MG-blades* design, which was deployed in all SGI UV2 systems. Through optimising scientific codes for this architecture (section 2.3), the CTC COSMOS team, collaborating with Intel, contributed to improvement of the software stack support for *MG-blades*. These advances gave SGI important guidance in designing the next generation UV3 systems, which included the co-designed MG-blade accelerator architecture [E1, E2].
“This collaborative work with the COSMOS team had an important impact on the definition of SGI UV2 and UV3 architectures, which were the flagship In-Memory HPC servers for SGI

from 2012-2016 […]. Beyond validating our early-access In-Memory architectures, the co-designed 'MG-blades' […] became a generic feature available on all SGI UV2000 systems." HPE Fellow/Vice-President (former Chief Engineer at SGI) [E1].
The takeover of SGI by HPE in 2016 for USD275,000,000 was significantly influenced by CTC-led innovations:
The core technology for the In-Memory platforms is what drove the interest for HPE to acquire SGI in 2016. […] The value of combining In-Memory computing with tightly coupled accelerators continues today. HPE Fellow/Vice-President (former Chief Engineer at SGI) [E1].

Next-generation HPE Superdome Flex: In late 2017, CTC and HPE co-designed and delivered the world's first, and largest, Superdome Flex server with tightly coupled hardware acceleration [E1]. HPE Senior VP and Chief Technology Officer, AI confirms that [E2] "*the hybrid UV2 co-designed with inbuilt accelerators, influenced the definition of subsequent in-memory architectures, including the UV3 system and now the HPE Superdome Flex."*

HPE VP stressed the importance of Superdome Flex in HPE's roadmap for the Mission Critical/HPC Division, for which annual revenues are USD3,200,000,000:
"The Superdome Flex server has become a core element in HPE's Mission Critical product range [...] It is a key platform for database users including Oracle and Microsoft SQL, while rapidly became the dominant platform supporting SAP HANA. Superdome Flex has been commercially successful around the globe on a wide variety of sectors/use-cases, with publicly announced examples including India Oil Company (Manufacturing), Burger King Russia (Retail), Kubo.com in Japan (Financial sector), Danfoss (Manufacturing), Travelport (Retail)" [E1].

[I2] Exploiting hardware acceleration and revolutionising visualization

Exploiting hardware acceleration: The CTC was among the first UK sites awarded Intel Parallel Computing Center (IPCC) status in 2014 for their pioneering involvement in code innovation and modernisation on novel Intel many-core and hybrid architectures (section 2.3) [E3]. CTC's exemplar research codes, WALLS [R4], GRChombo [R3] and MODAL [R1, R2], were used to validate and stress test the new Xeon Phi coprocessor (Knights Corner, KNC) and processor (Knights Landing, KNL). Optimisation of these codes, in collaboration with Intel "*provided key insights for both Intel and our broad ecosystem on best practices and application of the most advanced methods to deliver profound increases in performance",* confirms an Intel Senior Director [E3]. Intel's Xeon Phi processors have played a key role in recent HPC, with the second-generation KNL processor being deployed in the world's fastest supercomputers since 2016. By November 2017, KNL powered four of the top 10 systems and 19 of the top 100 [E3]. CTC's work also helped improve Intel's Manycore Software Stack (MPSS) and software developer tools [E3] with these key technical advances enabling hardware acceleration on UV2 and Superdome Flex.

In-situ visualisation: Since 2014, CTC's collaborative research with Intel has been at the frontier of visualisation technology, with intimate involvement in the development of open-source visualisation libraries (OSPray [E4], Catalyst), and the creation of *in-situ* visualisation capabilities on HPC systems [E5, E6]. Milestones of CTC research demonstrating these in-situ capabilities have included:

- A benchmark real-time animated visualisation of 10TB of cached WALLS data using the OSPRay ray tracing framework co-developed with Intel [E5];
- First visualisations on next generation Xeon Phi KNL processors using the Planck bispectrum, which required scaling OSPRay to many-core systems [E7];
- Together with Kitware Inc, creators of the popular ParaView graphics application, CTC combined *in-situ* visualisation frameworks (the Catalyst library) with OSPRay's HPC scalability. This was extended from fixed grids (WALLS [E5]) to complex unstructured grids (GRChombo/Adaptive Mesh Refinement) [E6].

OSPray is now at the core of the Intel OneAPI Rendering Toolkit, which is widely used in medical data applications, the oil and gas industry, the defense industry, interior space planning software, and Hollywood CGI [E8].
"[OSPray open-source library use is estimated] in the 150-200 range from commercial applications like Autodesk's "AutoCAD" to Dreamworks "MoonRay" Renderer used to make "How To Train Your Dragon: Hidden World" […] to Blender, which is itself used for everything from movies and television to automotive design. [We are also using them] in a large-scale public outreach programme with the Discovery Channel, 'Universe Unravelled with the Stephen Hawking Centre'," Intel Senior Director [E8].

In 2020, in recognition" of their important extended contributions to HPC visualisation techniques", the CTC was awarded the status of Intel Graphics and Visualisation Institute of Xellence [E8].
Working together with Kitware, Intel and CTC have been using OSPRay and Catalyst libraries to create *in-situ* visualisation capabilities on HPC systems, including incorporating unstructured grids for their GRChombo code [E5, E6]. These features are now available in Kitware's open source, multi-platform data analysis and visualisation application ParaView, which supports applications ranging from laptops to remote visualisations on Top 10 Supercomputers [E9]:

"Thanks to our collaboration with Intel and your Center, these […] capabilities have been incorporated into ParaView and are now available to our worldwide user base. New versions of ParaView are downloaded over one hundred thousand times every year. Commercial applications including aerospace, energy and automotive industries ranging from wind turbine to space craft to fusion reactor design. Many of these applications directly benefit from the work created by our collaboration with Intel and CTC." Kitware Senior Director of Computing.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[E1] HPE Letter of support, dated 10/03/20 from HPE Fellow/Vice-President.
[E2] HPE Letter of support, dated 23/04/2020 from Chief Technology Officer, HPC & AI.
[E3] Intel Letter of support dated 10/03/20 from Intel Senior Director, Planning–Cloud, Enterprise & Government.
[E4] "OSPRay–A CPU Ray Tracing Framework for Scientific Visualization", IEEE Vis 2016 (CTC visuals)
[E5] "In-Memory Large Scale Visualization" SGI White Paper 2016
[E6] "Visualizing the Physics of the Cosmos" Intel/CTC Collaboration Case Study original version 2018 (updated late 2019).
[E7] ISC 2015 Intel Booth KNL demonstration of collaborative Intel/CTC visualisation https://www.youtube.com/watch?v=yrw_11Tc8JQ
[E8] Intel Letter of support, dated 17/03/2020 from Senior Director, Senior Principal Engineer, "Advanced Rendering and Visualization", Intel Corporation.
[E9] Kitware Letter of support, dated 7/04/20 from Senior Director of Scientific Computing.
[E10] "Hawking CTC Receives Honors in HPCwire Readers' and Editors' Choice Awards", 2015.