Name of Bidding Organisation:	University of Exeter
Contract Title:	Examining the performance of Nektar++ for fusion applications

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 2.3.

We propose a schedule of work to develop proxyapps to investigate and quantify the performance of spectral element methods in the context of modelling anisotropic heat transport in the tokamak plasma edge in close proximity to a complex, 3D first wall, part of the NEPTUNE fusion modelling system. Specifically, we will target the computational and algorithmic developments required to construct a proxyapp for a sample 2D model of anisotropic heat transport (output 2.1.5), by combining efforts on two separate fronts:

- The development of bespoke high-order mesh generation techniques that are capable of faithfully representing the physics found in tokamak plasma edge, both in terms of handling X-point geometries (output 2.1.1) and highly accurate surface meshing (output 2.1.2).
- Extending the capability of the *Nektar++* spectral/*hp* element framework (www.nektar.info, [1, 2]), in collaboration with UKAEA, the successful bidder of call T/NA083/20, and other stakeholders of the NEPTUNE and ExCALIBUR projects, to enable the development of proxyapps for related problems in edge plasma physics, taking into account factors such as performance portability, separation of concerns and sustainable code development (outputs 2.1.3 and 2.1.4).

To enable these developments, the investigatory team will draw on unique expertise with *Nektar++* and the high-order mesh generator *NekMesh*. In the following sections, we outline how the objectives of this work package (and broader goals of the NEPTUNE project) can be met through the development of high-order methods, and how the goals and background of the investigatory team map onto those objectives.

Background. High-order methods are increasingly being viewed as an enabling technology for bespoke high-fidelity modelling of challenging physical systems. The use of high-order polynomials within elements, when compared to classical linear and second-order methods, offers advantages from both a numerical analysis and implementation perspective. The 'resolution power' of high-order discretisations is far greater than at lower orders, with favourable dispersion and diffusion characteristics meaning that complex spatial structures and phenomena spanning multiple length scales can be accurately resolved and tracked across long timescales. This accuracy is critical in highly sensitive and complex systems, such as fluid dynamics, where energetic and inherently unsteady phenomena such as vortex interactions require accurate resolution. From an implementation perspective, the modern many-core hardware architectures that will power upcoming exascale platforms are increasingly reliant on algorithms and methods that are *arithmetically intense*: that is, they perform a sufficiently large number of floating-point operations (FLOPS) per byte of data transferred from memory in order to realise the full potential of the available hardware. This trend towards arithmetic intensity holds on any of the traditional CPU (Intel Skylake or AMD Threadripper), energy-efficient architectures (e.g. ARM ThunderX2) or GPU based platforms (e.g. Nvidia Tesla) and, to date, lower order methods have struggled to realise the full potential of these platforms. Since the resolution power of high-order methods comes at the cost of additional FLOPS, this makes high-order methods ideally placed as an enabling technology for exascale simulations.

Enabling high-order methods within NEPTUNE. The primary goal of the work we propose in this project is to develop the necessary algorithmic and software improvements required to enable high-order simulations to be performed for the problems of interest to the NEPTUNE community, and demonstrate that high-order methods meet a significant number of the requirements **P1** through **P8** that have been identified in the call document. We highlight that high-order methods meet some of these requirements already, for example:

- the accurate solution of the compressible Navier-Stokes simulations as a hyperbolic system **(P1)** for direct numerical simulation of high-Reynolds number flows, as shown in [3];
- more broadly, high-order discontinuous Galerkin formulations are used to ensure e.g. conservation of mass in the above **(P6)**;
- the accurate solution of elliptic problems **(P2)** such as the pressure Poisson equation are routinely solved in complex geometries as part of e.g. velocity or pressure correction schemes for the incompressible Navier-Stokes equations, which feature highly anisotropic dynamics **(P3)** in boundary layer regions and in modelling vortex interactions [4].

The focus of this project will therefore be to demonstrate that these properties are not restricted to e.g. only the fluid dynamics community, but can be observed in the plasma-edge specific systems, by considering the overriding problem at the heart of this call: developing a proxyapp, based on high-order spectral/*hp* element methods, that can work towards the goal of anisotropic heat transport in the tokamak plasma edge, accounting for a complex 3D wall geometry. To meet this objective, we separate the goals of this project into two tasks which will be required to meet the identified outputs from this work package.

<u>Task 1: Curvilinear mesh generation.</u> One of the significant bottlenecks in developing and running a high-order simulation is the ability to generate high-quality, boundary conforming meshes. Unlike linear meshes, high-order meshes must be curved and deformed so that they align with the boundary, which makes the generation of slender anisotropic grids (such as in a boundary layer) a significant challenge. Creating a valid and accurate mesh for the simulation of, for example, a full car or aircraft configuration, therefore remains an unresolved challenge, posing a significant hurdle to the uptake of these methods. To work towards addressing this, the *Nektar++* team has invested significant academic development of the *NekMesh* mesh generator [5, 6], in order to establish a set of pre-processing modules for the generation of high-quality curvilinear meshes for *Nektar++*. This is an area where both commercial and academic codes are scarce. In this task, the *NekMesh* generator will be augmented to support the geometries that are critical in the modelling of tokamak edge regions:

- **Task 1.1: Generation of tokamak edge region meshes:** To support the clustering of elements around X-point configurations and other divertor configurations (output 2.1.1), we will leverage and significantly extend the variational framework outlined in [5], in order to produce refined anisotropic grids that can accurately model the X-point configuration. This will extend and strengthen existing efforts in this area for *rp*-adapted grids [7] which leverage this approach in order to refine grids for shock capturing purposes. **<u>Deliverable 1.1</u>**: A set of mesh generation routines, embedded in NekMesh, for modelling the tokamak edge region.
- **Task 1.2: Quad-based mesh generation for 2D configurations:** One of the main themes in the use of high-order methods in exascale applications is the use of *matrix-free methods*. In this approach, the generation of either large, sparse global matrices (for the whole grid), or smaller, locally-dense matrices (for each element) is avoided by recasting the action of the matrix as a summation and reduction operation. This therefore reduces the memory bandwidth requirements and allows high-order methods to attain 50-70% of the peak performance of modern CPU and GPU platforms [8, 9]. This is most effective when combined with a tensor contraction technique called sum-factorisation alongside a mesh of quadrilateral or hexahedral elements [10], which readily admit such a decomposition. In this deliverable we will leverage a proof-of-concept 2D quadrilateral generation technique, developed in NekMesh [11]. Although this method shows great potential, it requires a concerted software development effort to increase its robustness. This task will support this development and integrate the changes of D1.1 to generate high-quality, adapted grids suitable for the tokamak edge region. **<u>Deliverable 1.2</u>:** Augmentation of the NekMesh generator to provide quad-based meshes for 2D configurations.

- **Task 1.3: Accurate surface mesh generation:** In this task, we will first validate surface quality for representative meshes supplied by UKAEA, by generating meshes using the existing NekMesh high-order generation techniques and investigating important metrics such as surface normal accuracy. Where necessary, we will improve the surface mesh generation technology contained with *NekMesh* to ensure that sufficiently accurate grids by, for example, examining whether the metrics used to optimise node positions for the high-order mesh can be improved, or if change within the CAD system will improve vertex placement. We will ensure that the requirements outlined in output 2.1.2 are met for representative geometries. As well as those supplied by UKAEA, we will also validate meshes from other related calls, such as the successful bidders for call T/NA083/20, and other bidders who require high-order meshes as part of their tasks where required. **Deliverable 1.3:** Report outlining validation process for accurate surface meshes, and/or required improvements to NekMesh to deliver this.
- **Task 1.4: End-user interfaces and workflow integration:** A key theme within this call is the requirement for software that is not only exascale-capable (or in the case of mesh generation, that supports an exascale-capable end-goal), but can align with user requirements and is easy to deploy, use and develop. Although *NekMesh* and *Nektar++* already strive for excellence in this area through the provision of high-quality documentation, continual integration and code review, in this deliverable, we will ensure that these themes are translated across to these mesh generation developments through the provision of documentation, tutorials and training sessions. Working with UKAEA and other call partners, we will ensure that these can be embedded within workflows as required with the overall NEPTUNE project. To this end we will extend and solidify a Python interface to enable advanced scripted usage of *NekMesh* in a variety of pre-processing environments, as well as to embed this within existing and future workflows. **Deliverable 1.4:** Python interface, documentation and continual integration for the NekMesh generator for the developments undertaken in task 1.

Task 2: Developing flexible and performance portable proxyapps. The primary objective will be to construct a proxyapp for 2D anisotropic heat transport, as required in output 2.1.5. As part of this, we will extend the existing x86 algorithmic kernels with kernels for ARM and GPU architectures to demonstrate the code will be performant on a range of potential exascale platforms. This task is therefore broken down into the following stages:

- **Task 2.1: Implementation and validation of baseline x86 proxy-app for anisotropic 2D heat transport.** This will be built within the *Nektar++* framework, leveraging high-order spectral elements and using the existing matrix-based local elemental operators. This initial proxy-app will be used to verify correctness of the initial algorithmic implementation and used as a baseline case for performance optimisations undertaken in the later part of this task. **Deliverable 2.1:** Baseline proxy-app, built on Nektar++, for solving the anisotropic heat transport equation for x86 architectures.
- **Task 2.2: Matrix-free anisotropic Laplacian kernel.** Matrix-free kernels enable high-order methods to achieve high arithmetic intensity. We will develop matrix-free kernels for the Laplacian kernel which minimise data movement and can exploit the vectorisation capabilities of modern x86 processors. In particular, we will focus on how best to incorporate the second-order coefficient tensor into this operator which allows for the representation of anisotropic media. Benchmarking will be performed to measure performance gain over baseline case, percentage of peak FLOPS, with results characterised using roofline plots. **Deliverable 2.2**: Matrix-free kernel for the anisotropic Laplacian operator for x86 architectures.
- **Task 2.3: Matrix-free kernels for ARM and GPU.** Spectral/hp element kernels will be developed for the ARM architecture to allow performant vectorised execution of the proxy-app on the power-efficient ARM architecture. We will also develop corresponding kernels for NVidia GPU architectures, enabling use of a range of potential exascale platforms. Our previous work [12, 13] have demonstrated we can achieve performance portability on GPUs with this approach, but a more considered approach to e.g. memory layout and thread distribution & scheduling must be employed to achieve peak performance. The performance of these kernels will be benchmarked against the baseline case on a flops-per-watt basis. **Deliverable 2.3:** Collection of matrix-free kernels for ARM and GPU architectures.

- **Task 2.4: Preconditioning for highly anisotropic systems.** The 2D heat transport test case is strongly ill-conditioned due to the strong anisotropy required. Iterative solvers, used for large scalable parallel runs, will consequently converge slowly. We will identify the best choice of preconditioners for use with this problem, with input from the preconditioning work undertaken by the winning bidders of task T/NA084/20. **Deliverable 2.4**: Report summarising benchmarking and preconditioner assessment undertaken in Task 2.

Task 3: Engagement with NEPTUNE partners and the broader proxyapp ecosystem. A critical part of this call is to ensure engagement with the wider NEPTUNE and ExCALIBUR community, particularly around the development of the proposed proxyapp within this call, the work undertaken within closely related calls T/NA083/20 (development of plasma fluid referent model) and T/NA084/20 (preconditioning), as well as to align with the overriding goals of enabling performant, sustainable and easily usable software. In this task, we lay out a series of deliverables to deliver coordination of our efforts with those of other calls and the wider NEPTUNE and ExCALIBUR projects.

- **Task 3.1: Coupling with other proxyapps.** To enable the connection of the proxyapps being developed in Task 2 with those being developed in other work package partners, we will leverage the coupling technology (CWIPI) already in Nektar++ to enable the coupling of different proxyapps. A proof-of-concept approach has been demonstrated in [14] for the integration of an aeroacoustic model together with a combustion LES code. In this task we will undertake software development work to solidify this interface and provide a documented example case to highlight its usage. **Deliverable 3.1:** Proof-of-concept coupling example that demonstrates the use of the proxyapp with another CWIPI-compatible solver.
- **Task 3.2: Training materials.** As a fundamental issue in the coordination of work package efforts will be to communicate progress on the various objectives throughout this project. Furthermore, ensuring the usability and longevity of the software requires a concerted effort to thoroughly document the resulting proxyapps and developments. To tackle this challenge, the focus of this continuing task is to document progress in an evolving manner throughout the project. **Deliverable 3.2:** training materials for the developments undertaken in tasks 1 and 2.
- **Task 3.3: Coordination efforts.** To engage with UKAEA and other work package partners, we propose to hold community engagement events, through e.g. workshops/training days, tutorials, etc. We will aim to align this with other projects, such as PRISM and ELEMENT (noted above) to highlight broader engagement with ExCALIBUR as well as other codes in development. To support these efforts, we have requested a small travel budget for coordination purposes. Although we anticipate that in-person travel will suffer from restrictions due to Covid-19 in the short-term, these funds will be used wherever possible and appropriate to encourage collaboration between ourselves and other work package partners. The PRISM project will also supplement the travel budget should further travel or conference attendance become possible. **Deliverable 3.3:** activity such as an end-of-project workshop to highlight the developments in all tasks.

Alignment with ExCALIBUR goals. The work we propose in this project closely aligns with the goals and pillars of the ExCALIBUR project, as well as other ExCALIBUR-funded and EPSRC-funded projects. The work we propose under task 1 on mesh generation relates directly to the *ELEMENT* project (EP/V001345/1), funded as a priority use case under the phase 1 EPSRC calls for ExCALIBUR. Additionally, our team are also investigators of the closely aligned *Platform for Research in Simulation Methods (PRISM*) project. In terms of ExCALIBUR's pillars, we closely align with:

- **Co-design:** The investigators form a highly interdisciplinary team, who sit at the interface of applied mathematics, scientific computing and applications in engineering and physics. Through projects such as ELEMENT and PRISM, as well as e.g. Sherwin's role as Director of Research Computing Service (RCS) at Imperial College London, the team has a wide range of experience in coordinating large-scale scientific computing projects, such as the project we propose here.

- **Investing in people:** The mesh generation and proxyapp development we outline here relies on the investment in scientific computing researchers, in order to develop their RSE skillset and coordinate with other work package leaders. The investigatory team is committed to supporting work in this area, as evidenced by Sherwin's position as RCS director, as well as the PRISM project, which for several years has underpinned RSE activities within a wide range of codes, including Nektar++, and allowed researchers to develop interdisciplinary careers in research software engineering. As a concrete example, the project's recent continual integration upgrade to a modern container-based environment, which will benefit the testing and quality this project significantly, was supported through the RSE team at Imperial College.
- **Data Science:** A significant part of Task 3 is to ensure that the proxyapp developed in Task 2 is rigorously documented and, moreover, can align with the workflows of other work packages in this call. Moreover, we will invest effort in Task 3 into ensuring our solvers can be integrated with other codes using coupling frameworks, allowing data science and other researchers to use these tools in the broader context of NEPTUNE.
- **Separation of concerns:** The implementation of the high-order kernels in Task 2 will be implemented in Nektar++ in a manner so that the performance portability aspects between architectures can be preserved, but at a higher level, modellers can still interface with these routines in the same manner without concern of specific performance considerations.

Sustainability and usability of software. All codes will be developed within, or upon, the Nektar++ framework and NekMesh tools and therefore adhere to good programming practices already applied within these software packages. Consequently, this will lead to sustainable software which can be easily accessed, extended and integrated into new workflows. Codes will leverage the build system and packaging tools used within the Nektar++ project and therefore the developed proxyapp will be easily built and deployed. Performance portability is addressed through platform-specific kernels and support for multiple compiler toolchains in the build system.

References

[1] C. D. Cantwell *et al.*, 'Nektar++: An open-source spectral/hp element framework', *Comput. Phys. Commun.*, vol. 192, pp. 205–219, 2015, doi: 10.1016/j.cpc.2015.02.008.

[2] D. Moxey *et al.*, 'Nektar++: Enhancing the capability and application of high-fidelity spectral/hp element methods', *Comput. Phys. Commun.*, vol. 249, p. 107110, Apr. 2020, doi: 10.1016/j.cpc.2019.107110.

[3] D. de Grazia, D. Moxey, S. J. Sherwin, M. A. Kravtsova, and A. I. Ruban, 'DNS of a compressible boundary layer flow past an isolated three-dimensional hump in a high-speed subsonic regime', *Phys. Rev. Fluids*, vol. 3, p. 024101, 2018, doi: 10.1103/PhysRevFluids.3.024101.

[4] J.-E. W. Lombard, D. Moxey, S. J. Sherwin, J. F. A. Hoessler, S. Dhandapani, and M. J. Taylor, 'Implicit large-eddy simulation of a wingtip vortex', *AIAA J.*, vol. 54, no. 2, pp. 506–518, 2016, doi: 10.2514/1.J054181.

[5] M. Turner, J. Peiró, and D. Moxey, 'Curvilinear mesh generation using a variational framework', *Comput.-Aided Des.*, vol. 103, pp. 73–91, 2018, doi: 10.1016/j.cad.2017.10.004.

[6] D. Moxey, M. D. Green, S. J. Sherwin, and J. Peiró, 'An isoparametric approach to high-order curvilinear boundary-layer meshing', *Comput. Methods Appl. Mech. Eng.*, vol. 283, pp. 636–650, 2015, doi: 10.1016/j.cma.2014.09.019.

[7] J. Marcon, G. Castiglioni, D. Moxey, S. J. Sherwin, and J. Peiró, 'rp-adaptation for compressible flows', *Int. J. Numer. Methods Eng.*, vol. n/a, no. n/a, 2020, doi: 10.1002/nme.6529.

[8] M. Kronbichler and K. Kormann, 'Fast matrix-free evaluation of discontinuous Galerkin finite element operators', *ACM Trans. Math. Softw. TOMS*, vol. 45, no. 3, pp. 1–40, 2019.

[9] M. Kronbichler and W. A. Wall, 'A performance comparison of continuous and discontinuous Galerkin methods with fast multigrid solvers', *ArXiv Prepr. ArXiv161103029*, 2016.

[10] D. Moxey, R. Amici, and M. Kirby, 'Efficient matrix-free high-order finite element evaluation for simplicial elements', *SIAM J. Sci. Comput.*, vol. 42, no. 3, pp. C97–C123, 2020.

[11] J. Marcon, D. A. Kopriva, S. J. Sherwin, and J. Peiró, ‘A high resolution PDE approach to quadrilateral mesh generation’, *J. Comput. Phys.*, vol. 399, p. 108918, Dec. 2019, doi: 10.1016/j.jcp.2019.108918.
[12] J. Eichstädt, M. Green, M. Turner, J. Peiró, and D. Moxey, ‘Accelerating high-order mesh generation with an architecture-independent programming model’, *Comput. Phys. Commun.*, vol. 229, pp. 36–53, 2018, doi: 10.1016/j.cpc.2018.03.025.
[13] J. Eichstädt, M. Vymazal, D. Moxey, and J. Peiró, ‘A comparison of the shared-memory parallel programming models OpenMP, OpenACC and Kokkos in the context of implicit solvers for high-order FEM’, *Comput. Phys. Commun.*, vol. 255, p. 107245, Oct. 2020, doi: 10.1016/j.cpc.2020.107245.
[14] K. Lackhove, A. Sadiki, and J. Janicka, ‘Efficient Three Dimensional Time-Domain Combustion Noise Simulation of a Premixed Flame Using Acoustic Perturbation Equations and Incompressible LES’, presented at the ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition, Aug. 2017, doi: 10.1115/GT2017-63050.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

The objectives and deliverables for this work plan are listed as follows. Please refer to the description of tasks in section 2 for further details of each deliverable.

- **Task 1:**
  - **Deliverable 1.1**: A set of mesh generation routines, embedded in NekMesh, for modelling the tokamak edge region.
  - **Deliverable 1.2:** Augmentation of the NekMesh generator to provide quad-based meshes for 2D configurations.
  - **Deliverable 1.3:** Report outlining validation process for accurate surface meshes, and/or required improvements to NekMesh to deliver this.
  - **Deliverable 1.4:** Python interface, documentation and continual integration for the NekMesh generator for the developments undertaken in task 1.
- **Task 2:**
  - **Deliverable 2.1:** Baseline proxy-app, built on Nektar++, for solving the anisotropic heat transport equation for x86 architectures.
  - **Deliverable 2.2**: Matrix-free kernel for the anisotropic Laplacian operator for x86 architectures.
  - **Deliverable 2.3:** Collection of matrix-free kernels for ARM and GPU architectures.
  - **Deliverable 2.4**: Report summarising benchmarking and preconditioner assessment in Task 2.
- **Task 3:**
  - **Deliverable 3.1:** Proof-of-concept coupling example that demonstrates the use of the proxyapp with another CWIPI-compatible solver.
  - **Deliverable 3.2:** Training materials for the developments undertaken in tasks 1 and 2.
  - **Deliverable 3.3:** Activity such as an end-of-project workshop to highlight the developments in all tasks.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

None.

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

Asides from any potential restrictions/risks (and the corresponding mitigations) noted elsewhere in the document (the most notable due to Covid-19), there are no constraints that we are aware of.

4 Approach

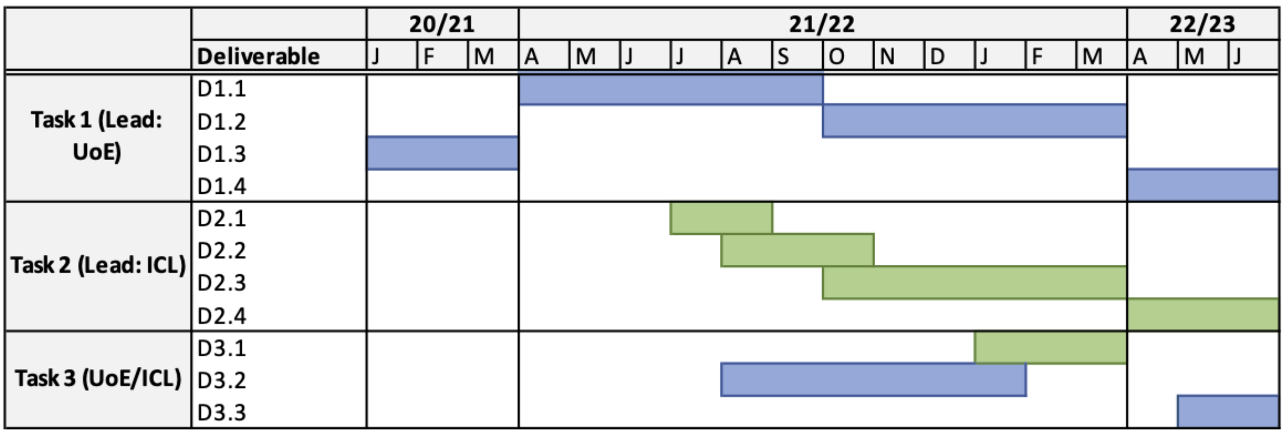
Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

Software development process. As set out in section 2, the work to be undertaken in this project will be implemented within the Nektar++ spectral/hp element framework and its constituent mesh generation software, NekMesh. To achieve this work package's objectives, we will rely on the extensive background and experience of the investigatory team in leading the numerical and computational development of spectral element methods.

Since software developments form the central deliverables of the project, we will leverage the extensive work and existing processes undertaken within Nektar++ to ensure sustainable software is delivered. To be specific, the resulting software will:

- use the Nektar++ build system, based on CMake, in order to be both cross-platform and offer a user-friendly and easy to build environment;
- leverage our existing continual integration platform, utilising Docker images and GitLab CI to ensure the rigorous testing of the developed software;
- similarly, leverage the continuous delivery mechanism of Nektar++ to support automatic generation of Docker containers of latest developments;
- maintain performance portability is addressed through platform-specific kernels and support for multiple compiler toolchains in the build system.

Project structure. Organisation of the project's deliverables, which we set out in the following chart, have been timed to ensure a linear progression of the project from mesh generation through to. Consideration has also been given to the requirements of other calls; for example, the generation of meshes under D1.1 has been timed to be completed by the half-way stage of the project, to enable their potential usage in e.g. call T/NA083/20.



Personnel. The work to be undertaken in this project will be supported through two PDRAs: PDRA1 will be based in the Department of Engineering at the University of Exeter, employed from 02/20 (to allow sufficient time for recruitment) until the end of the project. They will be supervised by Dr. David Moxey, who will also coordinate the project as PI. Primarily, PDRA1 will focus on the development of mesh generation (task 1). They will also coordinate with PDRA2 on solver aspects under task 2 as required. PDRA2 will be based in the Department of Aeronautics at Imperial College London. They will be supervised by Dr. Chris Cantwell and Prof. Spencer Sherwin, where they will focus on developing the flexible proxyapp/kernels under task 2, in collaboration with PDRA1 as appropriate. PDRA1 and PDRA2 will both assist in coordination efforts under task 3.

Coordination efforts and travel. In light of the ongoing coronavirus pandemic, the initial stages of the project will be coordinated via regular online meetings, both for internal project meetings between the investigators/PDRAs, as well as with UKAEA and other winning bidders of related calls where appropriate. Although the restrictions to travel imposed by Covid-19 are likely to endure for some time into the project, we hope for (and anticipate) the ability to conduct some in-person collaboration in FY 21/22.

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible Owner	Required Data
Task/deliverable 2.4 will leverage developments from the winners of call T/NA084/20 on preconditioning. Should this input not be possible or should the task be delayed, we will leverage our previous experience on the development of preconditioning approaches for fluid dynamics systems.	T/NA084/20	Preconditioning
Tasks/deliverables 3.1 & 3.3 will rely on input from other work packages; primarily we believe from the winners of call T/NA083/20 on fluid-referent models. Should input from these bidders not be possible, we will aim to demonstrate the potential of our developments using the proxyapp developed in task 2.	T/NA083/20	Other proxyapp development

6 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

	Activity No	**1**
Activity: **Task 1: Curvilinear mesh generation** **Assignee:** **PDRA1 / David Moxey (DM)**		
Objective 1: Enable the generation of 2D tokamak edge geometries with quadrilateral meshes.		

<u>Objective 2:</u> Ensure that surface mesh generation is accurate for required simulations. **<u>Objective 3:</u>** Software development to enable workflow integration & widespread usage of NekMesh.		
<u>Key Deliverables:</u> **Deliverables D1.1, D1.2, D1.3 and D1.4.**	**<u>Start and Completion date:</u>** **D1.1: 01/04/21 - 30/09/21** **D1.2: 01/10/21 - 31/03/22** **D1.3: 04/01/21 - 31/03/21** **D1.4: 01/04/22 - 30/06/22**	**<u>Assignee:</u>** **PDRA1 will lead with support from DM.**
<u>Milestones towards deliverables:</u> Reporting via call requirements.	**<u>Completion date:</u>** As above.	**<u>Assignee:</u>** As above.

	Activity No	**2**
<u>Activity:</u> Task 2: Developing flexible and performance portable proxyapps **<u>Assignee:</u> PDRA2 / Chris Cantwell (CC) & Spencer Sherwin (SJS)**		
<u>Objective 1</u>: Implementation and validation of baseline x86 proxy-app for anisotropic 2D heat transport **<u>Objective 2:</u>** Matrix-free anisotropic Laplacian kernel for ARM/GPU **<u>Objective 3:</u>** Consideration of preconditioning for highly anisotropic systems		
<u>Key Deliverables:</u> **Deliverables D2.1, D2.2, D2.3 and D2.4.**	**<u>Start and Completion date:</u>** **D2.1: 01/07/21 - 31/08/21** **D2.2: 01/08/21 - 31/10/21** **D2.3: 01/10/21 - 31/03/22** **D2.4: 01/04/22 - 30/06/22**	**<u>Assignee:</u>** **PDRA2 will lead with support from CC & SJS.**
<u>Milestones towards deliverables:</u> Reporting via call requirements.	**<u>Completion date:</u>** As above.	**<u>Assignee:</u>** As above.

	Activity No	**3**
<u>Activity:</u> Task 3: Engagement with NEPTUNE partners and the broader proxyapp ecosystem **<u>Assignee:</u> PDRAs & investigators (DM, CC, SJS)**		
<u>Objective 1:</u> Coupling with other proxyapps. **<u>Objective 2:</u>** Production of training materials for tasks 1 and 2. **<u>Objective 3:</u>** Coordination efforts (e.g. workshop or seminar) to engage with other work package coordinators and UKAEA.		
<u>Key Deliverables:</u> **Deliverables D3.1, D3.2 and D3.3.**	**<u>Start and Completion date:</u>** **D3.1: 01/01/22 - 31/03/22** **D3.2: 01/08/21 - 31/01/22** **D3.3: 01/05/22 - 30/06/22**	**<u>Assignee:</u>** **PDRA1 will lead D3.2 and D3.3 with support from DM. PDRA2 will lead D3.1 with support from CC/SJS.**
<u>Milestones towards deliverables:</u> Reporting via call requirements.	**<u>Completion date:</u>** As above.	**<u>Assignee:</u>** As above.

7 Resource Plan

Research Plan Roles and Responsibilities

Name	Title	Organisation /institution	Required Role	Required Responsibility	Cost	Confirmation of payment source (Paid from the grant award / in kind/other funding)
David Moxey	Dr	University of Exeter	PI	Overall project coordination, supervision of PDRA1 under tasks 1/2		
Chris Cantwell	Dr	Imperial College London	Co-I	Supervision of PDRA2 under tasks 2/3		
Spencer Sherwin	Prof	Imperial College London	Co-I	Supervision of PDRA2 under tasks 2/3		
PDRA1	-	University of Exeter	PDRA	Task 1 + 3		
PDRA2	-	Imperial College London	PDRA	Task 2 + 3		

Warwick University: PIC methods.

Note that this is a modified summary of the bid document accounting for modifications that occurred after initial submission, so is intended to present information about Warwick's bid in an easy-to-follow way. That is, this is not a redacted version of the agreed contract document.

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

There are two major products:

1) An exploratory proxy-app, in which disruptive PIC numerical methods will be implemented, and which will also enable UQ exploration.
2) Reports on the feasibility and limitations of the disruptive methodologies employed, and on UQ aspects in PIC.

The single activity, which involves carrying out the research plan, comprises several parts, which may temporally overlap:

1) Literature survey, consultation with relevant experts (e.g. at the Met Office and UKAEA), and draft specification of additional features for the proxy-app.
2) Development of the proxy-app features by the project team.
3) Exploitation of the proxy-app, to determine the practicality and limitations of the potentially disruptive features, and UQ studies.
4) Reporting on these aspects, to enable this work to be integrated within the ongoing NEPTUNE project.

Timing of individual work elements is described in detail in the list of milestones, the overall project leads will be responsible for ensuring on-time delivery of milestones, and redeploying effort as required to ensure full delivery before the end of the 9 month project period.

Resources required are the following

1) A proportion of the time of each of the project participants.
2) Access to journals, which will be provided by the University of Warwick
3) Access to desktop PCs and workspaces, which would normally be provided by the University.
4) Access to workstation-level HPC, which is provided by the university.

There are two major objectives, which we will now provide contextual information on, including some relevant literature, and describe in detail.

Objective 1: Co-design.

We will work together, early-on in the project, with other successful co-bidders to determine whether the proposed particle methods will be compatible with the proposed fluid methods and adapt them as needed. Due to the very short timeline of the project, it is however necessary to work in parallel on other aspects of the project on the assumption that such adaptions will be possible and relatively minor. We will also need to, as part of the work towards this milestone, develop familiarity with the spectral-hp method and the chosen software infrastructure in which it is embedded (e.g. Nektar++). This milestone will assume the availability and willingness of the lead organisation and other relevant successful bidders to commit time and resources early-on in the project. We note that coupling between PIC methods and fluid methods has previously been achieved, and we think that the advanced-PIC aspects and advanced-fluid aspects of the project should in-principle be compatible. The basic PIC algorithm itself is quite flexible in terms of grid and field representation (e.g. high order finite element methods are common e.g. ORB5), and the proposed PIC extensions are not highly sensitive to the choice of field discretisation.

We will also work with the other successful bidders to examine how the gyrokinetic theory of interest may be implemented in a kinetic scheme, and which features of PIC algorithms need to feed in to the design of the uncertainty quantification.

This work will initially be expected to involve literature search, discussions with other partners, and training of staff, for example on PIC and Spectral-HP methods. This will allow the best approaches to the couplings between these NEPTUNE subprojects be identified, which will be described in the milestone report.

Objective 2: Implicit timestepping and transformative PIC algorithms

The finite-difference time-domain (FDTD) method is the de facto industry standard time-integration algorithm for particle in cell (PIC) codes. However, its temporal stability constraints render it inefficient for modelling systems with a large disparity in timescales, where implicit [e.g. 1, 2] or semi-implicit [3] methods are preferable. Additionally (semi-)implicit methods can provide physically desirable behaviour, such as energy conservation, which explicit time-integration methods cannot.

Leveraging implicit PIC methods on Exascale systems is unlikely to be straightforward. In addition to the algorithmic considerations relevant to explicit PIC, implicit PIC requires the efficient inversion of a non-linear system. Interestingly, both Chen et al. [1] and Markidis and Lapenta [2] utilise the Jacobian-free Newton-Krylov (JFNK) method. JFNK methods do not explicitly form the Jacobian matrix, thus avoiding the high-memory requirements of conventional implicit methods. A third possible time-integration method is the semi-implicit method of Lapenta [3], which offers the advantages of implicit methods (no temporal stability constraints, exact energy conservation) whilst maintaining the lower algorithmic complexity of explicit methods.

Unfortunately decoupling the performance benefits of existing methods from the differences arising from data-structures, linear-solvers, and numerous other implementation details is at best difficult, and at worst impossible. In order to disentangle these issues, we are proposing to produce a prototype mini-app to assess the relative merits of different algorithms, whilst using the same underlying computational framework. The choice of time-integration algorithm is complex. Even restricting the discussion to JFNK based time-implicit methods using particle enslavement, one must decide on a method to solve for the enslaved motion, for example the Picard method [1] and the Newton-Raphson method [2] have both been used. Furthermore design decisions regarding spatial filtering, particle substepping, and even the possibility of preconditioning the JFNK system must be made. As JFNK methods don’t explicitly form the matrix, these methods typically require bespoke physics-based preconditioners (e.g. [4-5]).

The questions raised in the preceding paragraph are far beyond the scope of this initial nine month study. Instead this work package will deliver a key first step for further work, an extendable, easily deployed, time-implicit mini-app, upon which future investigations and comparisons can be built.

Especially in a tokamak context, the rapid gyration time can be an equally or more severe limit on the particle timestep as an explicit field solve. To overcome this limitation, several methodologies have been developed, including substepping and gyrokinetic techniques. Particle substepping is known to be compatible with implicit methods in general, and, although it increases in computational complexity per particle per timestep, does not require significant additional memory bandwidth so is appealing on the pathway to exascale. Both a reduced (simplified, long-wavelength) gyrokinetic particle tracking method and direct substepping will be implemented in the PIC mini-app and compared, as a full-scale integrated edge code will likely have particles of such widely varying gyrofrequency/gyroradius that both types of methods will be required.

Low-noise particle methods generally depend on improved sampling of the particle velocity space and on knowledge of an appropriate underlying smooth distribution function. One way to do this is via fluid-entrainment techniques. A control-variates method has been implemented in EPOCH, and this will be ported to the EPOCH-miniapp; an extension to allow background evolution consistent with the fluid equations (closed with PIC moments) will be implemented and investigated.

[1] Chen, Guangye, Luis Chacón, and Daniel C. Barnes. "An energy-and charge-conserving, implicit, electrostatic particle-in-cell algorithm." *Journal of Computational Physics* 230.18 (2011): 7018-7036.

[2] Markidis, Stefano, and Giovanni Lapenta. "The energy conserving particle-in-cell method." *Journal of Computational Physics* 230.18 (2011): 7037-7052.

[3] Lapenta, Giovanni. "Exactly energy conserving semi-implicit particle in cell formulation." *Journal of Computational Physics*334 (2017): 349-366.

[4] Knoll, Dana A., and David E. Keyes. "Jacobian-free Newton–Krylov methods: a survey of approaches and applications." *Journal of Computational Physics* 193.2 (2004): 357-397.

[5] Chen, Guangye, et al. "Fluid preconditioning for Newton–Krylov-based, fully implicit, electrostatic particle-in-cell simulations." *Journal of computational physics* 258 (2014): 555-567.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 2.3.

This submission directly works towards several of the objectives in Part 1 Section 2.1.

Firstly, the issue of timestep limitations in particle tracing will be tackled by both advanced particle update schemes and an implicit time-solver for the coupled field-plasma problem that is, it addresses the ‘task work’ described in section 2.3 and 2.4. In addition, low noise techniques, including low-dispersion sampling, and control variates will be explored, probing the questions in section 2.2 of the task work.

As required by the call, each disruptive approach tested will be examined by performing appropriate literature surveys, and by implementing proof-of-principle code in the PIC proxy-app. The overall findings will be documented to determine their applicability in NEPTUNE.

We will provide the (existing) PIC proxy-app to the broader community and support its application and use by NEPTUNE participants to deliver the broader project aims; in particular, this will allow HPC experts to explore the HPC and scaling aspects of PIC algorithms.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

1) Write a report addressing co-design aspects related to PIC, particularly with respect to spectral-HP.

2) Implement an implicit electromagnetic solver, advanced low-noise methods, and advanced particle tracing (to mitigate restrictions due to the gyro-time) within a PIC mini-app, report on issues, and identify potential solutions.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

1) Examine coupling of fluid methods to PIC.

2) Examine interface of PIC with spectral-HP elements scheme. This may be performed outside the bid framework as a PhD student is currently investigating advanced Galerkin-type representations for the purpose of representing anisotropic variations.

3) Some aspects of PIC data storage and sub-MPI task tiling to mitigate memory-throughput related performance limitations were investigated on historical HPC using the EPOCH mini-app [1] [2]. Reviewing and extending work on data layout and locality with reference to advanced HPC would thus be a natural extension of the proposed work but we don’t have experience staff available to perform this at the moment.

[1] Bird, Robert F., et al. "Mini-app driven optimisation of inertial confinement fusion codes." *2015 IEEE International Conference on Cluster Computing*. IEEE, 2015.

[2] Bird, Robert F., et al. "Performance Optimisation of Inertial Confinement Fusion Codes using Mini-applications." *The International Journal of High Performance Computing Applications* 32.4 (2018): 570-581.

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

- This project is designed to feed in to a collaborative activity and potential fusion of the PIC method explored here with a range of other approaches for specification of a complete edge model. This bid was prepared without a clear picture of these elements, and in the absence of a clear strategy for this process to occur.
- The tight project schedule, with certain milestones due within the first three months, probably precludes any significant modification to the project plan, or significant renegotiation of milestones etc., after the start date, without substantially impacting on the scope of the work.

4 Approach

Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

Numerical science will commence with a review of the relevant literature, for work-packages where there is not an existing identified numerics strategy. This will be a collaborative process chaired by Ben McMillan and Tony Arber. To give an example where value for money can be provided by short-circuiting this process, the implicit field solver ties in with an existing project of a project member (Tom Goffrey) and this literature review and identification of the numerical methods is already complete; the implementation in the PIC proxy-app thus can begin immediately.

We note that there is an existing proxy-app based on a heavily simplified version of the EPOCH PIC code, developed by several of the team members: this combination of experience and existing code-base allows rapid pursuit of the advanced numerics targets. Indeed, this code has previously been used to examine scaling and data-arrangement issues on earlier-generation architecture.

The numerical algorithms chosen to meet the milestones will be fed into a streamlined software engineering process to allow agile development of proxy-app extensions. A lightweight code specification will be developed by the lead on each milestone, and approved by the lead scientists.

5 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

We note that Annex B was not included as part of the bid documents: the work schedule may be inferred from the below.

Duplicate this table for each Activity	Activity No	1
Activity:		

Exascale PIC **Assignee:** Ben McMillan		
Objective 1: UQ for PIC methodology. **Objective 2:** Implicit timestepping.		
Key Deliverables: **1) Report on findings of codesign process and implications for further algorithm development.** **2) Implement advanced methods (low-noise methods, implicit field solver, advance particle tracing) within a PIC mini-app.**	**Start and Completion date:** **1) 01/01/2021 – 20/03/2021** **2) 01/01/2021 - 01/09/2021**	**Assignee:** **1) Ben McMillan** **2) Tom Goffrey**
Milestones towards deliverables: **1) Report on co-design.** **2) Implicit field solver baseline implementation in mini-app.** **3) Advanced particle orbit solver implementation in mini-app.** **4) Low noise techniques implementation in proxy-app.**	**Completion date:** **1) 20/03/2021** **2) 01/09/2021** **3) 20/03/2021** **4) 01/09/2021**	**Assignee:** **1) Ben McMillan** **2) Tom Goffrey** **3) Ben McMillan** **4) Ben McMillan**

10 Financial Schedule

10.1 Payment schedule

Activity number	Milestone number	Payment Value	On completion of associated activities (please provide relevant activity number) *	Start Date	End Date
2	3			04/01/2021	01/09/2021
			Associated Output of Milestone: Mini-app with implicit field solver baseline implemented. Dependencies: none		
	4			04/01/2021	20/03/2021
			Associated Output of Milestone: Mini-app with advanced particle orbit solver implemented. Dependencies: None		
	5			04/01/2021	20/03/2021

			Associated Output of Milestone: Low-noise control variates implementation in proxy-app. Dependencies: None		
1	6			04/01/2021	20/03/2021
			Associated Output of milestone: Report on findings of co-design process, assessing the impact of spectral-hp fluid and gyrokinetic/drift-kinetic model on design of advanced PIC methods. Dependencies: none		
	Total Value:				

Name of Bidding Organisation:	University College London
Contract Title:	**T/NA080/20**

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 2.3.

Our proposal features the following benefits:

- We have experience with developing and applying uncertainty quantification (UQ) methods applied to a wide range of domains including weather and climate forecasting, advanced materials, epidemiology, drug discovery, personalised medicine, human migration and, of most pertinence to the present proposal, fusion energy applications. The open source, open development software we have developed, called the VECMA Toolkit (VECMAtk) is freely available from the VECMA project (www.vecma.eu) at www.vecma-toolkit.eu. With such tools and experience at hand we are already able to assess the robustness of the codes that will be developed in FM-WP1.

- For the aforementioned applications, we have implemented within EasyVVUQ [1,2] a range of non-intrusive and semi-intrusive UQ methods featuring quasi-Monte Carlo, Polynomial Chaos Expansion and Stochastic Collocation, as well as dimension adaptive sampling, and surrogate models.

- We have established a taxonomy alongside a systematic and in-depth analysis of UQ workflows which we refer to as "UQ patterns"; analysing the pattern(s) of the NEPTUNE project workflow will immediately provide us with a refined idea of candidate UQ methods (already implemented with EasyVVUQ) which can then be applied very easily.

- We have decades of experience in developing scalable HPC methods for multiscale, multiphysics simulations. We have developed two very promising candidate exascale applications SCEMa (https://github.com/UCL-CCS/SCEMa) and HemeLB (hemelb.org). We have developed a third application, the Binding Affinity Calculator for ligand-protein binding free energy prediction, which at its core features large ensemble simulations, fully equipped with UQ capabilities [3] and runs at scale all over Europe and USA.

- Most interestingly we have developed scalable HPC methods specifically for VVUQ methods which run on several of Europe's largest supercomputers (SuperMUC-NG, Eagle, Marconi). VECMAtk features two further pieces of software which are highly relevant and immediately applicable to the FM-WP1 objectives: MUSCLE3 enables the (exa)scalable execution of multi-model workflows (P7.a/b) while EasySurrogate is intended to perform dynamic training of surrogate models for UQ such as arise within multi-model simulations (P7.c).

- We have experience in developing standalone, versatile ProxyApps: EasyVVUQ, FabSim, MUSCLE3 and EasySurrogate (as part of the larger toolkit VECMAtk [4,5]). EasyVVUQ has been developed in order to be able to readily perform UQ on any model, a design feature which fits well the "separation of concerns" approach of the NEPTUNE project.

- We have experience with error propagation in chaotic dynamical systems, whether originating in fluid turbulence or particle based codes such as molecular dynamics; indeed, we have recently discovered a new floating point pathology in such systems when simulated on digital computers which will require investigation in the present context as well [6]. This pathology cannot be removed by increasing the precision of the floating-point numbers and requires further investigation.

- We will participate fully within the community built around the NEPTUNE project, providing training for our UQ tools, experience and know-how for organising (physical and virtual) workshops, and we can readily lead projects involving a large number of development teams.

- We have many years of scientific project management experience leading international multidisciplinary research projects involving ultra-high-end high performance computing.

- EasyVVUQ fits the requirements of the NEPTUNE charter as it is implemented in Python; the same applies to MUSCLE3, as it is implemented in C++. We are experienced at handling the standard HDF5 data format.

[1] Wright, D. W., Coveney, P.V., et al. "Building confidence in simulation: Applications of EasyVVUQ." *Advanced Theory and Simulations* (2020): 1900246.
[2] Richardson, R. A., Coveney, P.V., et al. "EasyVVUQ: A Library for Verification, Validation and Uncertainty Quantification in High Performance Computing." *Journal of Open Research Software* 8.1 (2020).
[3] Al Saadi, A., Coveney, P.V., et al. “IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads.” *arXiv preprint* (2020):2010.06574.
[4] Groen, D., Coveney, P.V., et al. "VECMAtk: A Scalable Verification, Validation and Uncertainty Quantification toolkit for Scientific Simulations." *arXiv preprint* (2020):2010.03923.
[5] Groen, D., Coveney, P.V., et al. "Introducing VECMAtk-Verification, validation and uncertainty quantification for multiscale and HPC simulations." International Conference on Computational Science. Springer, (2019).
[6] Boghosian, B. M., Coveney, P.V., Wang, H. "A New Pathology in the Simulation of Chaotic Dynamical Systems on Digital Computers", *Advanced Theory and Simulations* (2019): 1900125.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

We expect to produce the following three main outputs as deliverables from this project:

1. A clear and concise literature summary of VVUQ methods.
2. The organisation and running of two events, one at the start and one at the end of the project.
3. A concise summary of recommendations of methods to perform efficient and scalable uncertainty quantification (UQ) of the plasma edge multi-model code.

In more detail, we will:

- write a concise literature review of UQ methods suited for multi-model workflows and in particular simulating the physics of the plasma edge; the review will feature methods we already have experience with (advanced sampling, adaptive sampling, surrogate modelling) and also prospective methods among which some have been introduced during the Data Assimilation and UQ workshop organised by ExCALIBUR (https://excalibur-sle.github.io/workshop2.html); we will prioritise UQ methods that enable scalable computations on existing multipetascale and emerging exascale architectures for high-dimensional problems.
- exploit already implemented methods within EasyVVUQ and perform initial UQ and sensitivity analysis (SA) campaigns on available and the early-stage (or toy) models developed in FM-WP2 and FM-WP3 and on the existing European Boundary Code and B2-EIRENE.
- organise two event(s) to gather the community of NEPTUNE model developers, initially: (i) for them to explain the algorithmic structure of their models, what they would need to perform UQ on and for us to explain our capabilities and to demonstrate the use of EasyVVUQ by means of hands-on tutorials; (ii) we will present the methods we recommend at a final meeting and to confirm that the community is in collective agreement with our proposals.
- report on community-wide agreement as to what approach will be taken in terms of the use of VVUQ methods in the next phase of the NEPTUNE project; this outcome will be captured in our final report on recommendations concerning the optimal path to pursue.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

None.

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

We will face four main constraints:

1. Limited time in which to implement and test the shortlisted methods;
2. Access to suitable compute resources, especially if we intend to test scalability;
3. Availability of the codes which will require instrumentation with UQ methods from various NEPTUNE partners;
4. The likelihood that we shall have to convene meetings online.

However, the present project is of limited duration (just six months) and is not intended to involve significant software development. Thanks to the usability and flexibility of the VECMAtk, we expect to be able to perform rapid prototyping to demonstrate feasibility by working in collaboration with code owners for around five codes in this period of time.

For demonstration purposes, we expect to have access to sufficient computing resources from our own end, in order to be able to accommodate some basic performance testing in this period.

4 Approach

Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

Introduction:

It is widely recognised that uncertainty quantification (UQ) is essential for any model simulation to be actionable, i.e. to have genuine predictive value. Ensemble-based methods lie at the heart of techniques enabling the evaluation of the uncertainty associated with a model parameter. A combinatorially large number of simulations of the model must be performed to probe the effects of the parameters comprising it. As a result, UQ is intrinsically expensive. Nevertheless, many methods can be used to accelerate UQ and reduce the computational costs in comparison to brute force quasi-Monte Carlo methods. Their efficiency depends on the model equations, the simulation workflow, and the number of parameters in play.

A distinction is made between intrusive UQ methods, where one substitutes the original model with a form of stochastic representation, and non-intrusive methods, where the original model is used as a black-box. Intrusive methods are efficient and relatively easy to apply to linear models. This, however, represents only a relatively small class of models. They can be applied to non-linear models as well, but the solution of the resulting equations may then become very demanding. Non-intrusive methods can be applied to any type of non-linear model. However, if a single model run requires large execution times, these UQ methods may be relatively ineffective, and at worst computationally intractable. In between, there exists a family of semi-intrusive UQ algorithms for multiscale models. These methods are called semi-intrusive since they are intrusive only on the level of the multiscale model workflow, that is, in the way the single scale components are coupled together. The single scale components themselves, however, are treated as black-boxes.

Non-intrusive methods rely primarily on advanced sampling methods such as Polynomial Chaos and Stochastic Collocation. Sampling efficiently the probabilistic distribution of the model's parameters enables one to reduce the number of samples required to build trustworthy estimates of the uncertainty. Furthermore, dimension-adaptive sampling methods [1] enable one to refine the number of samples for parameters with higher sensitivity and thus evade the curse of dimensionality whereby the computational cost would scale exponentially in the number of dimensions of the parameter space.

Intrusive and semi-intrusive methods frequently depend on surrogate modelling. Expensive parts of a model are replaced by cheaper stochastic approximations. Gaussian processes and neural networks are key tools to build such models. In turn, such models can be simulated multiple times at far lower computational cost, while preserving the full model's sensitivity to uncertain parameters.

There is forward propagation of uncertainty which originates in the uncertainty of model parameters, and the stochasticity in e.g. chaotic systems. When it comes to perform the inverse approach, to infer uncertainty from measured outputs, inverse Bayesian methods become relevant [2,3].

Background:

EasyVVUQ, the tool we intend to use to investigate different UQ methods, is designed for multiscale and multiphysics models, as well as single scale models. It is part of a larger toolkit, the VECMAtk, which also features tools for centralised execution on multiple petascale supercomputers (QCG), improved job array allocation management (PilotJob), efficient data management between geographically distributed e-infrastructure (FabSim) along with coordination of submodel execution and data transfer within complex workflows (MUSCLE3). Overall, investigations of methodologies to perform efficient UQ will benefit from the existing testbed we have available for this project (see table 1), which will ensure scalability and ease of implementation.

Site	**Architecture**	**Processors**	**Number of cores**
PSNC	AMD	AMD Opteron (Interlagos)	5,500
	Intel Xeon	Intel Xeon E5 v3	36,650
	Intel Xeon	Intel Xeon E5 v4	1,760
	ARM	Toradex based on NVIDIA Tegra 3(Apalis T30), Samsung Exynos 5, and APM ARM 64bit	400
	GPU	Nvidia Tesla M2050	150,000
LRZ	Intel Xeon	SandyBridge EP E5-2680	150,000
	Intel Xeon	Haswell E5	96,768
	Intel Xeon Phi	Knights Landing	9,600

Table 1: Overview of established HPC resources available at PNSC (Poznan, Poland) and LRZ (Leibniz, Germany). These resources together comprise the VECMA VVUQ testbed.

EasyVVUQ already features the aforementioned non-intrusive UQ methods (PCE, SC, and adaptive sampling). We have applied those to a well-known epidemiological model featuring more than nine hundred parameters, which at first sight might be thought to suffer from the “curse of dimensionality”. By means of sensitivity analysis (SA) and adaptive sampling we have been able to reduce to a mere handful the parameters which dominate the variance in the code’s output data; in fact, over 60% of the variance in the model’s behaviour is attributable to a mere three parameters. It also serves as a very user-friendly environment for the aggregation and thorough analysis of data output from codes, including of course reporting on sensitivity analysis and UQ. We have also initiated the development of EasySurrogate to enable rapid substitution of expensive submodels with efficient, trained and verified surrogate models, as part of semi-intrusive UQ campaigns. It already features stochastic neural network surrogates. EasySurrogate is publicly available as an alpha version part of the latest VECMAtk release.

Outstanding applications of EasyVVUQ currently include enabling full *in silico* trials as part of the EU-funded INSIST project, as well as assessing a well-known epidemiological model featuring more than nine hundred parameters at the behest of RAMP [4]. The adoption of the individual VECMA tools (QCG-PJ, EasyVVUQ-QCGPJ, FabSim3, EasyVVUQ) has been increasing continuously in the research community for the last three months (see figure 1). Early adoption has been most significant for EasyVVUQ which has now been downloaded some three thousand times. Rapid increases in the number of views and downloads have followed the organisation of special events such as a workshop held at the Alan Turing Institute in January 2020 (https://www.turing.ac.uk/events/reliability-and-reproducibility-computational-science) and the VECMA-ICCS-SIAM Conference held online in June 2020 (https://www.vecma.eu/siamuq20_iccs2020/) which attracted over 300 attendees.

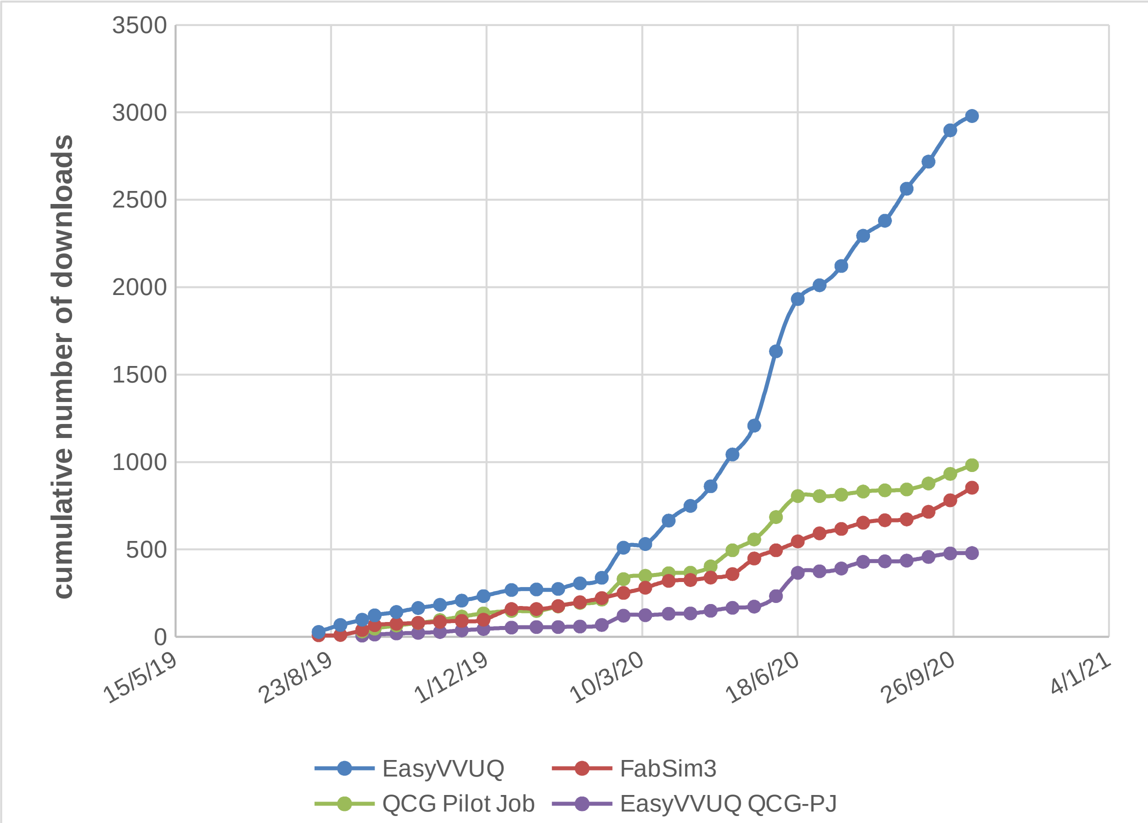


Figure 1: Cumulative number of individual downloads of the VECMAtk components (QCG-PJ, EasyVVUQ-QCGPJ, FabSim3, EasyVVUQ) since the initial release of the VECMAtk on the 23 August, 2019.

We have significant experience in designing ProxyApps which are compatible with various multiscale and multiphysics computational workflows. We have acquired proven expertise driving successive FET-HPC EU-funded projects over the last decade: MAPPER (http://mapper.pub.lab.nm.ifi.lmu.de/web/guest), ComPat (https://www.compatproject.eu/) and currently VECMA (http://https://www.vecma.eu). We have designed generic optimisation tools for the execution of multiscale simulations, as well as for the execution of large-scale UQ campaigns. One of the key aspects of these projects is their versatility; that is, we have designed tools that are efficient for a large spectrum of applications. Of most interest here is fusion, but these also include climate modelling, which suits the needs of the other major use case of ExCALIBUR. Indeed, the VECMAtk is being used to perform non-intrusive and intrusive UQ on the plasma simulation workflow developed at MPG-IPP [5] and the climate model developed at CWI [6].

In the current NEPTUNE project, the models and equations are already specified, so we can perform shortlisting for methods particularly suited for the plasma turbulence code. In particular, we are familiar with such codes either from personal experience with chaotic systems [7] and our longstanding collaboration with MPG-IPP.

The VECMAtk has been developed from the outset for use with large supercomputers. Therefore it has been designed to handle efficiently the following performance-critical aspects:

- the sampling of parameter values and creation of large numbers of simulation inputs;
- the submission, execution and retrieval of large ensembles of simulation jobs;
- the efficient movement of data between local and remote resources;
- the efficient movement of data between coupled models.

Methodology:

- We will investigate the prospective structure of the plasma edge code defined in the Science Plan and in the Equations for NEPTUNE Proxyapps document. We will compare the structure of the code with the UQ Patterns we have defined in our previous work in the context of the VECMA project. Based on the identified patterns, we will immediately be able to propose a selection of UQ methodologies suited for the plasma edge code. We anticipate that the structure of the plasma edge code will be similar to the MPG-IPP plasma turbulence code, which we already have substantial experience with and have performed non-intrusive UQ on using Marconi, the PRACE Tier-0 supercomputer at CINECA in Italy.

- We will complement this shortlist by means of a literature review and consultation with other partners funded within the NEPTUNE project. We have already initiated discussions at a DA and UQ workshop organised by the ExCALIBUR project held online the 24-25th of September, in which we were active participants (https://excalibur-sle.github.io/workshop2.html).

- We will organise a first workshop with representative partners from FM-WP1, FM-WP2, and FM-WP3 for them to explain the algorithmic structure of their models. We will also make sure that representatives from FM-WP4, WC-WP1 and XC-WP1&2 are able to attend for community building purposes. We will introduce and explain our existing UQ capabilities and demonstrate the use of EasyVVUQ by means of hands-on tutorials, which should be of interest to all attendees. Discussions should also help us to refine our shortlist of UQ methods based on expectations from model developers.

- According to the document FMS/0021, only the code to compute the Braginskii coefficients is already available. We expect to be able to rapidly gather toy models provided by partners involved in FM-WP2 and FM-WP3 during the first meeting. We would also hope to be able to use the existing European Boundary Code and B2-EIRENE which the future developments in FM-WP2 and FM-WP3 will rely on. In the same order of priority, we will immediately test the shortlisted UQ methods. Based on the similarity with the applications we have already instrumented with UQ, most of the shortlisted UQ methods will be already available within VECMAtk, and more precisely EasyVVUQ. These include the non-intrusive Polynomial Chaos Expansion and Stochastic Collocation sampling methods. We would expect to try adaptive SC sampling, as it is expected that the future plasma edge code will feature a large number of input parameters and therefore suffer from the curse of dimensionality. We will also investigate multi-level Monte Carlo methods (MLMC), computing the uncertainty at different level of accuracy. MLMC methods are currently not available within EasyVVUQ, but they can be implemented readily if required.

- In the meantime, we will perform SA of the available codes using the adaptive sampling method, which will enable us to determine the most influential input parameters of these plasma edge codes. We have designed template generation tools to automatically find and list parameters of a given code. This will enable us to perform an exhaustive SA of every single parameter from the codes we investigate. From the resulting SA, we will be able to establish a reduced list of parameters to perform UQ on, hindering the “curse of dimensionality” in future analysis. While performing SA, we will also test the scalability of our methodology. Such large-scale UQ campaigns, which we have performed in the context of applications in epidemiology and drug discovery, typically require scalable workflows that already push the limits of petascale infrastructures [8].

- Computations on toy models will be performed on local clusters at UCL, which do not have any resource usage restrictions. When it comes to perform UQ and SA of larger codes, we will access infrastructures on which we already have substantial allocations. Our allocations are largely sufficient for the execution of this project. The HPC resources we have access to are associated with the following supercomputers: ARCHER (EPCC), Eagle (PSNC), Marconi (HPC-Cineca, where MPG-IPP performs production runs) PizDaint (CSCS), and SuperMUC-NG (LRZ).

- We will convene a second workshop involving presumably many of the same partners from the initial meeting; our deeper connections established during the first meeting should also encourage us to update the list of partners involved in the community. We will report on our experience with our shortlist of UQ methods, the SA performed on available codes, and their computational efficiency in the context of emerging exascale UQ computing. We will also report on the structure of the established workflows to enable discussions on optimal computer architectures required to perform such computations. To that extent, we will consider the requirements of the submodel simulations, multi-model workflow execution as well as UQ workflows.

- The conclusions of our discussions and analysis of the most suitable UQ methods, preliminary SA of available codes and UQ workflows architectures will be disseminated in a final report as a set of recommendations for subsequent work within the NEPTUNE project.

- We expect to extend our UQ of the plasma edge code with verification and validation (VV). EasyVVUQ also provides a set of V and V patterns that enable *inter alia* automated verification of the accurate resolution of a model's mathematical equations as well as comparison with experimental results. For the former, we will be able to determine automatically the level of refinement (e.g. discretisation in space and time) of the solution algorithms for convergence and accuracy of the code. For the latter purpose, we will be able to choose between several relevant metrics, such as the Hellinger distance [9], the Jensen–Shannon distance, which is a symmetrized and smoothed version of the Kullback–Leibler divergence [10], and the Wasserstein metrics for comparison of predicted and experimental data [11]**.**

[1] Gerstner, T. and Griebel, M. "Dimension–adaptive tensor–product quadrature." *Computing 71* (2003): 65–87.
[2] Edeling, W. N., et al. "Predictive RANS simulations via Bayesian Model-Scenario Averaging. " *Journal of Computational Physics*, 275 (2014): 65–91.
[3] Marzouk, Y. and Xiu, D. "A stochastic collocation approach to Bayesian inference in inverse problems." *Communications in Computational Physics*, 6 (2009): 826–847.
[4] Edeling, W., Coveney, P.V., et al. "Model uncertainty and decision making: Predicting the Impact of COVID-19 Using the CovidSim Epidemiological Code." (2020), *in review*.
[5] Lakhlili, J., et al. "Uncertainty Quantification for Multiscale Fusion Plasma Simulations with VECMA Toolkit." *International Conference on Computational Science*. Springer, Cham (2020).
[6] Edeling, W. N., and Crommelin, D. "Towards data-driven dynamic surrogate models for ocean flow." *Proceedings of the Platform for Advanced Scientific Computing Conference* (2019).
[7] Fazendeiro, L., Coveney, P. V., et al. "Unstable periodic orbits in weak turbulence." *Journal of Computational Science 1.1* (2010): 13-23.
[8] Al Saadi, A., Coveney, P.V., et al. "IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads." *arXiv preprint* (2020):2010.06574.
[9] Nikulin, M. S. "Hellinger distance." *Encyclopedia of mathematics* 78 (2001).
[10] Kullback, S., and Leibler, R. A. "On information and sufficiency." *The annals of mathematical statistics* 22.1 (1951): 79-86.
[11] Villani, C. "*Optimal transport: old and new*." Vol. 338. Springer Science & Business Media (2008).

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible Owner	Required Data
Prototypes or toy models of the turbulence submodels code	PI of projects part of FM-WP2/3	Executables of toy models and source codes

6 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

<table>
<tr><td colspan="2" align="right">Activity No</td><td>1</td></tr>
<tr><td colspan="3"><u>Activity:</u>
Report the knowledge gathered during the initial literature review, the discussions during the two workshops as well as the execution of preliminary UQ and SA campaigns.
<u>Assignee:</u> Peter V. Coveney</td></tr>
<tr><td colspan="3"><u>Objective 1:</u> write a concise set of recommendations as to which UQ methodologies to develop
<u>Objective 2:</u> explicitly describe the architecture of UQ workflows for co-design purposes toward exascale</td></tr>
<tr><td><u>Key Deliverables:</u>

1) First report
2) Second final report</td><td><u>Start and Completion date:</u>

1) 04/01/2021 to 31/01/2021
2) 01/07/2021 to 31/07/2021</td><td><u>Assignee:</u>

1) Maxime Vassaux
2) Maxime Vassaux</td></tr>
<tr><td><u>Milestones towards deliverables:</u>

1) All codes for which UQ will be considered will be specified by the end of month 2
2) All (VV)UQ evaluations will be performed by the end of month 5</td><td><u>Completion date:</u>

1) 01/03/2021

2) 01/06/2021</td><td><u>Assignee:</u>

1) Peter V. Coveney

2) Peter V. Coveney</td></tr>
</table>

<table>
<tr><td colspan="2" align="right">Activity No</td><td>2</td></tr>
<tr><td colspan="3"><u>Activity:</u>
Organisation of two meetings with the ExCALIBUR partners from FM-WP1, FM-WP2, and FM-WP3, and if possible WP4, WC-WP1 and XC-WP1&2
<u>Assignee:</u> Peter V. Coveney</td></tr>
<tr><td colspan="3"><u>Objective 1:</u> hold the first meeting to introduce EasyVVUQ capabilities, provide hands-on tutorials and partners to present the structure of their codes and UQ requirements
<u>Objective 3:</u> hold the second meeting to present our shortlist of recommended UQ methods</td></tr>
<tr><td><u>Key Deliverables:</u>

1) Initial workshop
2) Final workshop</td><td><u>Start and Completion date:</u>

1) 14/01/2021
2) 01/07/2021</td><td><u>Assignee:</u>

1) Peter V. Coveney
2) Peter V. Coveney</td></tr>
<tr><td><u>Milestones towards deliverables:</u>

1) confirmation of more than 15 attendees
2) confirmation of more than 15 attendees</td><td><u>Completion date:</u>

1) 07/01/2021

2) 24/06/2021</td><td><u>Assignee:</u>

1) Peter V. Coveney

2) Peter V. Coveney</td></tr>
</table>

7 Resource Plan

Redacted

7.1 Relevant Experience

See CVs attached for Professor Peter Coveney and Dr Maxime Vassaux.

8 IPR

Redacted

9 Initial Risks

(NB: See annex A for risk scoring, please include exit plan risks. Add more rows as required)

Risk No	Risk Description	Cause (Describe what it is that might occur. IF … BECAUSE…)	Effect / Impact (What will be impacted as a result of this risk? RESULTING in…)	Mitigating actions (What actions will you take to reduce the risk/increase the opportunity)	Risk owner (Name of individual responsible for the management and control of this risk)	Probability (delete as appropriate)	Impact (delete as appropriate)	Score (insert score from risk scoring matrix)
01	Schedule incompatibility to organise two meetings during the course of the project	Unavailability of key partners when the workshop must be scheduled, that is the first and the last months of the project	Only part of the knowledge to be gathered during the workshops, and dissemination of our VVUQ toolkit will be lessened.	We will revert to one-to-one partner discussions, or at least revert to a list of essential partners to gather information on the codes to perform UQ on.	Peter V. Coveney	M	L	8
02	No toy model, existing model can be gathered during the project to perform initial SA and UQ	Partners in WP1, WP2 and WP3 have not had sufficient time to reflect on code implementations to simulate the models provided in the document CD/EXCALIBUR-FMS/0021.	Our application testbed will be reduced, limiting the applicability of our conclusions	We will revert to legacy code to perform initial UQ analysis	Peter V. Coveney	M	M	14

10 Financial Schedule

10.1 Payment schedule

Redacted

10.2 Breakdown

Redacted

Resourcing Breakdown

Redacted

T&S breakdown

Redacted

Direct / Indirect Costs Breakdown

Redacted

10.3 Value for Money Statement

Redacted

11 HPC Plan

Redacted

Annex A

Risk Scoring Matrix

	Probability				
	Unlikely but possible event **< 5%**	Realistic possibility **5-30%**	Likely to happen over a period of time **30-50%**	Likely to happen **50-75%**	Likely & imminent **> 75%**
	Very Low	**Low**	**Medium**	**High**	**Very High**
Time Delay (months)	< 1 month	1 to 2 months	2 to 3 months	3 to 6 months	> 6 months
Financial Impact on Project Budget (whichever is the highest)	<£20k -5% of project budget	£20k to £100k 5 - 10% of project budget	£100k to £200k 11 - 15% of project budget	£200k to £1m 16 - 20% of project budget	>£1m 20% + of project budget
Reduced Benefit (Less Revenue / Failure to deliver)	Minor impact on one Group/ Programme objective	Significantly impact upon one Group/Programme objective	Significantly impact upon more than one Group/ Programme objective	Failure to deliver one Group/ Programme objective	Failure to deliver more than one Group/Programme objective
Risk causes a Business Continuity Event leading to PR Issue or customer complaint / impact	Loss of potential customer/contract Interruption of critical services > 1/2 hr	Unhappy customer Interruption of critical services > 1 hour	Loss of minor customer/contract Interruption of critical services > 3 hours	Loss of a key (top 100) customer/contract Interruption of critical services > 6 hours	Loss of several key (i.e. top 100) customers/contracts Interruption of critical services > 12 hours
	Very Low	**Low**	**Medium**	**High**	**Very High**

		Probability						
		Very Low	**Low**	**Medium**	**High**	**Very High**		
Impact	**Very High**	15	33	56	78	100	**Very High**	Impact
	High	6	17	28	39	50	**High**	
	Medium	3	9	14	19	25	**Medium**	
	Low	1	4	8	10	13	**Low**	
	Very Low	1	2	4	5	7	**Very Low**	

Key	
High risk requiring immediate attention	**value > 32**
Medium risk requiring action	**13 < value <= 32**
Low risk – assurance required	**value <= 13**

Name of Bidding Organisation:	University College London (UCL)
Contract Title:	Steps towards Accurate Fusion modelling at the Exascale: Model Order Reduction (SAFE-MOR)

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 2.3.

- We will review and investigate concretely methods and codes for Model Order Reduction (MOR) in the context of complex UQ of multi-physics and multi-fidelity codes at the Exascale for fusion. This effort benefits from our ongoing Alan Turing Institute (2018-2021) project "Uncertainty quantification of multi-scale and multi-physics computer models" and our EPSRC ExCALIBUR software working group "Exascale Computing for System-Level Engineering" (EP/V001396/1 & EP/V001531/1) in which S. Guillas is the UQ lead coinvestigators, as Professor of Statistics, specialist in Uncertainty Quantification of complex computer models, Group Leader in the Data-Centric Engineering programme at the Alan Turing Institute and founder of the Turing UQ interest group. This aligns with the main objective of building a UQ strategy within the overarching project of fusion at the exascale.
- PI Serge Guillas is also the Met Office Joint Chair in Data Sciences for Weather and Climate, which allows synergies in defining the strategies for Reduced Order Modelling of complex fusion and weather/climate models, therefore providing robustness in the developed infrastructure for UQ at the exascale. The PI S. Guillas was the lead organiser for the ExCALIBUR Software Workshop on "Data Assimilation and Uncertainty Quantification at the exascale", 24-25 September 2020 (https://excalibur-sle.github.io/workshop2.html), with STFC, Met Office, UKAEA and international leaders in the field from Japan and the USA invited in talks and panels to discuss and establish pathways towards actionable UQ at the exascale, in which models of high dimensions and complexity need to be reduced.
- We will quantify shortcomings and benefits of Reduced Order Modelling approaches, some already implemented, including making used of reduced dimensions and complexities to design computer experiments of small sizes at high fidelity (for some or all the components). This aligns with the objective of reduction of sizes of the ensembles, and the objective that the eventual platform is actionable.
- We will provide tangible pathways to solutions for the Bayesian calibration of reduced order models against disparate and sparse measurements. It aligns with the objectives that the eventual platform is actionable, and the accuracy/robustness of the developed infrastructure becomes fully established.
- We have made use of the HDF5 data formats in other contexts for data compression (e.g. earthquake-generated tsunamis, climate models), to facilitate data exchanges as required in NEPTUNE.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

Our key deliverables are:

1. A summary of benefits and shortcomings of three different families of approaches to ROM, Proper Orthogonal Decomposition (POD), Proper Generalised Decomposition (PGD), Reduced Basis Methods (RB) in the specific context of promise of suitability for HPC/Exascale implementation particularly for a spectral element code.
2. Small development tasks to test candidate methods for accuracy, stability and HPC scalability potential, in the specific context of promise of suitability for HPC/Exascale implementation particularly for a spectral element code. Key outcomes would be suitability to reduced designs of computer experiments for multi-physics and multi-scale exascale computing. We will also assess the suitability of ROM calibration against data, one of NEPTUNE overarching goal, when the model (or some sub-parts) are reduced.
3. A joint strategy for ROMs as part of the overarching project, through interactions with the community (the other NEPTUNE projects, UKAEA, Met Office). The emphasis will be on developing both coherent and bespoke strategies across ROM and UQ NEPTUNE projects.

These investigations will be carried out by means of literature and code surveys and consultation with experts, especially the ones funded by the other NEPTUNE calls.

Implementations and suitability will be examined through co-design with NEPTUNE project partners and initial implementation only in a restricted class of machine architectures.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

None.

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

None.

4 Approach

Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

Introduction

Reduced Order Models (ROM) lessen the computational and mathematical complexity of challenging codes. For instance, inputs can be reduced to a smaller set, either by ruling out parameters or in terms of dimensionality, see e.g. Morris (1991) or Liu & Guillas (2017). Similarly, outputs can be shrunk to a more manageable representation in terms of basis expansion of either time series of outputs or spatial representations.

Our experience in ROM for UQ exactly targets the kind of complex, multi-physics, multi-scale codes that UKAEA will employ at the exascale. Our key ingredient is to use Gaussian Processes (GPs) as surrogates (also known as emulators) of the computer models. These are comparable to Polynomial Chaos (PC) approaches in terms of accuracy (Owen et al., 2017). GPs can be nowadays parallelised for large input/output spaces, as we are doing in the released Alan Turing Institute Package for fitting Gaussian Process Emulators to multiple output computer simulation results. It is named the *Multi-Output Gaussian Process Emulator (MOGP)* and relies on Research Software Engineering to manage, organise, and accelerate/parallelise some computations. For instance, in 2020, two postdocs in the group of S. Guillas have created up to 500,000 emulators in parallel representing each grid cell in an unstructured mesh output of a Finite Volume code ran on the GPU cluster Wilkes-2 on CSD3 at Cambridge.

The Turing platform MOGP and its few existing tools for ROM can interoperate with other UQ platforms such as the now mainstream EasyVVUQ: both are Python libraries with a similar philosophy but complement each other in terms of workflows (actually much more elaborate in EasyVVUQ), sampling (more elaborate in MOGP), surrogates (PC for EasyVVUQ, GP for MOGP) which are now well understood and interchangeable in workflows.

Our approach

We will first explore the various possible methods (or more precisely families of approaches and their extensions that include active subspace approaches) that can lead to a reduced order model. A literature review of the various strands will summarise the strengths and weaknesses of these approaches for the multiple types of models that may emerge in FM-WP1, FM-WP2, FM-WP3 and FM-WP4. We will describe these carefully to help the various communities of users.

We will work then together with the modellers of these Proxyapps in order to investigate potential strategies and methods to the early stages. This will enable feedbacks on the kind and size of inputs and outputs that can be reduced, with a quantification of the trade-offs in terms of losses in approximation, gains in computational burden and/or surrogate building complexity. It will also offer space for discussions of the critical leaps towards exascale which will have the most influence on the next methodological and computational research required to provide ROM for fusion modelling.

Key questions will be model-dependent, but always aim at addressing the trade-off between computational burden, accuracy of the model and input-output dependencies across the whole system. We endeavour to embed the concern of ROM into the initial development of approaches in FM-WP1, FM-WP2, FM-WP3 such as:

- **spectral/hp element methods**: integration of the choices of polynomial order (p-adaptivity), mesh refinement (h-adaptivity), etc into the ROM
- **Particle methods**: in one study, Nicolini et al. (2019) developed a proper orthogonal decomposition technique, applied to a finite-element time-domain particle-in-cell (PIC) algorithm for the simulation of kinetic plasmas. However, it can also lead to spurious results if proper care is not taken when assembling the reduced-order system. Hence, we will focus on exploring prospective strategies in terms of quantification of uncertainties for ROM that can likely mitigate these issues.
- **5D gyrokinetic models**: the computational burden is the main obstacle and thus ROM ought to give preference to this aspect while ensuring validity and stability. Truncations of spherical harmonics (as we did in Chang & Guillas 2019) are one option, among others to be explored.

FM-WP4 is devoted to **Code structure & coordination** and will have a specific status as the integrator of the various sub-models. We will investigate how our ROM approaches can be tailored to the computational implementation to further reduce data flows, as well as frequencies and latencies of I/O exchanges across sub-models. We developed the first surrogate models of feed-forward suites of models (Ming & Guillas 2020), which have analytical expressions, and will thus be able to illustrate the promise of ROM in this context. Key questions will be around the ability to identify the valuable coupling strategies that reduce dimensionality of data exchanges across the sub-models or the value of information propagated from an additional experiment at the sub-model level to further enhance the overall ROM of the system. We anticipate cooperating with the Met Office work package WC-WP1, as concerns are shared.

Finally, the testing against data is essential to validate models for e.g. the critical high fidelity modelling the edge region of the tokamak plasma. The aim is mainly to tune unknown parameterisations or e.g. boundary conditions. With two decades of research since the seminal paper of Kennedy & O'Hagan (2001), the calibration against observations of complex models with ROM has made progress: for instance Chang & Guillas (2019) made use of reduced spatial representations over the globe for a high exascale code by targeting the principal modes of variability of the field (dominant spherical harmonics) and incorporated the key non-stationary behaviour (e.g. local regimes), in terms of spatial variations of both variance and correlations, in the Bayesian calibration, the technique that provides the best uncertainties in the tuning, albeit at some cost of running MCMC. We will illustrate potential for these and other approaches, especially for the calibration of multi-physics and multi-scale models against complex/missing/sparse/multivariate/high dimensional/asynchronous measurements. This effort will be co-designed with the modellers and experimentalists in FM-WP1, FM-WP2, FM-WP3. Other real-time alternatives of adjustment of the model using Data Assimilation will also be explored (e.g. Meldi & Poux 2017 for Kalman filtering, and others for particle filtering).

References:

Chang, K. L., & Guillas, S. (2019). Computer model calibration with large non-stationary spatial outputs: application to the calibration of a climate model. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 68(1), 51-78

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464

Liu, X., & Guillas, S. (2017). Dimension reduction for Gaussian process emulation: An application to the influence of bathymetry on tsunami heights. *SIAM/ASA Journal on Uncertainty Quantification*, *5*(1), 787-812.

Meldi, M., & Poux, A. (2017). A reduced order model based on Kalman filtering for sequential data assimilation of turbulent flows. *Journal of Computational Physics*, *347*, 207-234.

Ming, D., & Guillas, S. (2020). Integrated Emulators for Systems of Computer Models. arXiv preprint, under review at *SIAM/ASA Journal on Uncertainty Quantification*

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161-174.

J. L. Nicolini, D. Na and F. L. Teixeira, "Model Order Reduction of Electromagnetic Particle-in-Cell Kinetic Plasma Simulations via Proper Orthogonal Decomposition," in *IEEE Transactions on Plasma Science*, vol. 47, no. 12, pp. 5239-5250, Dec. 2019, doi: 10.1109/TPS.2019.2950377.

Owen, N. E., Challenor, P., Menon, P. P., & Bennani, S. (2017). Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators. *SIAM/ASA Journal on Uncertainty Quantification*, *5*(1), 403-435.)

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible	Required Data

	Owner	
Edge plasma simplified Proxyapp models	UKAEA & partners	Inputs, codes for each sub-component to be run by UCL for design and surrogate modelling
Pseudo or real observations	UKAEA	Data and attributes (nature, time, space) in HDM5 format or equivalent with readme files and explanation, feedback if questions arise. To be used by UCL for calibration of the model against data

6 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

Duplicate this table for each Activity		**Activity No**	**1**
Activity: **Report on suitability and potential of ROM to fusion models** **Assignee:** Prof. Serge Guillas			
Objective 1: Summary of strengths and weaknesses of three families of existing ROM approaches (POD, PGD, RB) and their current and ongoing extensions (e.g., active subspace). **Objective 2:** Description of ROM potential for the reduction of size of inputs, size of design, size of outputs, and computational benefits for multi-physics fusion modelling at multiple scales for toy models and ProxyApps in discussions with FM-WP1, FM-WP2, FM-WP3 and FMWP4			
Key Deliverables: **1) deliver a progress report with preliminary findings** **2) Final report**	**Start and Completion date:** **1) 4 January-20 March 2021** **2) 20 March-31 July 2021**	**Assignee:** **1) D. Ming** **2) D. Ming**	
Milestones towards deliverables: **1) literature and codes identified** **2) ROM benefits on toy model**	**Completion date:** **1) 1 March 2021** **2) 15 May 2021**	**Assignee:** **1) S. Guillas** **2) S. Guillas**	

Duplicate this table for each Activity	**Activity No**	**2**
Activity: **Exploration of ROM for individual codes and their UQ** **Assignee:** Prof. Serge Guillas		
Objective 1: co-organisation of two workshops with the T/NA080/20 activity for UQ, and FM-WP1, FM-WP2, FM-WP3 and FM-WP4 to integrate ROM within the overarching UQ framework. Scope:		

(workshop 1) gather evidence of needs and potential; (workshop 2) display possible methods and implementations. **Objective 2:** direct interaction with researchers in FM-WP1, FM-WP2, FM-WP3 and FM-WP4 to scope the reduction of the number of spatial dimensions across models of different complexities and other ROM topics such as bespoke sensitivity analysis, data compression, reduced basis representations across fidelities and scales.

Key Deliverables:	**Start and Completion date:**	**Assignee:**
1) workshop 1 2) workshop 2 3) potential quantified cost/benefit for a few key codes through scientific interactions	**1) 4 January-15 February 2021 2) 1 July-15 July 2021 3) 15 February -15 July 2021**	**1) S. Guillas 2) S. Guillas 3) D. Ming**
Milestones towards deliverables:	**Completion date:**	**Assignee:**
1) 10 participants secured in workshop 1 2) 1 set of codes and developers engaged amongst one of: FM-WP1, FM-WP2, FM-WP3	**1) 15 January 2021 2) 1 May 2021**	**1) S. Guillas 2) S. Guillas**

7 Resource Plan

Research Plan Roles and Responsibilities

Name	Title	Organisation /institution	Required Role	Required Responsibility	Cost	Confirmation of payment source (Paid from the grant award / in kind/other funding)
Serge Guillas	Prof	UCL	PI	Direction of research, reporting and meetings		
Deyu Ming	Dr	UCL	Postdoc	Researcher		

Name of Bidding Organisation:	University of York
Contract Title:	**T/NA083/20 - Fluid Referent Models**

1 Purpose of Document

2 Benefits and alignment to Work Package objectives

- **Holistic, collaborative design:** Our approach will be to develop a series of proxyapps, with associated tests and documentation, structured around the 5 tasks specified in the call document. These will be made available to the community, and used to undertake studies of issues such as preconditioning, UQ, and DSL design, in collaboration with the successful bidders to the relevant calls, UKAEA and the NEPTUNE community. Lessons learned from these studies will be written up (see deliverables below), and used to inform the following tasks.
- **Task delivery**: The Work Package objectives set out five tasks, involving increasingly complex models, and integration with other work packages. To address these tasks with limited resources, we have designed a set of ambitious but realistic activities and deliverables. Where there is insufficient knowledge to enable a reliable resource estimate to be carried out, a pathway to obtaining the required knowledge has been identified, and the deliverable includes a suggested roadmap to achieving the objective.
- **Performance portability**: Researchers involved in this bid will have access to a range of workstations and high performance computing facilities at York and Exeter universities. This includes for example the York Viking cluster (Intel Xeon x86, NVIDIA TESLA v100 GPUs) and the new N8 Bede (IBM Power 9 CPUs and NVIDIA Volta GPUs). This will enable proxyapps to be developed and tested on a range of architectures. Access to these facilities is an in-kind contribution to the project.
- This submission aligns with the Excalibur pillars**:**
    - **Separation of Concerns:** The most complete proxyapp in this call is Task 5, which must integrate a number of different aspects of the project (e.g. DSLs, UQ), including outputs from bidders to other calls. To do this we will explore and develop capabilities in simpler proxyapps, and then use these to test integration with other aspects of the project. We will use both BOUT++ and Nektar++ to develop these proxyapps, to ensure that our solutions are not tightly coupled to a single framework.
    - **Co-design:** This proposal directly involves domain experts in plasma physics, computer scientists, Research Software Engineers with expertise of coordinating large-scale open source scientific research code development (e.g. BOUT++, GS2), and experts in the Nektar++ spectral/hp library. We have also identified that close coordination with other NEPTUNE and Excalibur work packages will be essential at an early stage, and will be an iterative process of improvement. Specific activities, deliverables and resources are described below, in order to carry out the coordination and training required.
    - **Data Science:** By providing functioning code which solves the referent systems of equations at an early stage of the project, we will enable other researchers to develop new work flows to analyse the data in-situ and in post-processing. Part of our coordination efforts will be in teaching other groups to use the codes and data we produce.
    - **Investing in People:** To deliver the goals of the NEPTUNE project, it is our aim to build a team of researchers and software engineers who understand holistically the issues involved, from plasma physics, algorithms, sustainable software engineering, to the hardware likely to be used at exascale. We have brought together a collaboration including engineers, physicists and computer scientists, who between them have the required expertise. During the course of this project the RSEs and researchers involved will work together, to spread and exchange their domain knowledge. To help build careers in this interdisciplinary work, we request funds to support attendance at high profile international conferences. We recognise that collaboration between successful bidders on the various NEPTUNE projects will also be essential, so have requested travel funding to support these collaborations.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

The deliverables of this work package will consist of the following types:

- **Test cases :** A report describing the test cases to be used to compare the accuracy, efficiency and scalability of implementations. This to be accompanied by data and scripts (written in python) to enable testing of new implementations. Where possible these will include analytic solutions, or manufactured solutions (MMS) for convergence testing in realistic complex cases.
- **Reference implementations**, implemented using BOUT++, to provide a baseline for comparison with other implementations, and a platform which can be used for initial exploration of preconditioning, intrusive UQ, and the development of DSLs. Our vision is to replace these BOUT++ implementations with Spectral/hp or other high-order implementations, but while those are being developed we will make use of the capabilities which already exist in BOUT++ to accelerate progress in other areas.
- **Spectral/hp implementations,** implemented using Nektar++. These will be tested using the same test cases as the reference implementations (in addition to implementation-specific unit tests). The accuracy and efficiency of the codes will be compared using the methodology to be established during the first three months of the call (deliverable 0.1).
- **Exploratory mini-apps**, which implement an algorithm in a higher-level language, for example Python or Julia. These would be used to investigate the accuracy and robustness of algorithm choices, for example phase space representation, time stepping methods and preconditioning. We see Task 3 (system 2-4) in particular as a suitable system with which to explore some of these design decisions which will inform the full 5D gyrokinetic implementation.

This work package brings together and interfaces with multiple other work packages, in particular calls T/N/A078/20 [Performance of spectral elements], T/NA084/20 [preconditioners], T/NA080/20 [Study of Uncertainty Quantification Techniques] and T/NA086/20 [DSL techniques]. We therefore propose an activity, Task 0, to carry out the required coordination and project management.

Task 0 : Coordination and management

- ○ **Deliverable 0.1** : Set up environment e.g version control, continuous development/integration, automated testing and documentation services, coding standards. Set up testing framework for evaluating parallel scaling on e.g. Archer2 / Viking / Bede
- ○ **Deliverable 0.2** : A report on coordination with UKAEA and other work groups, including any training activities carries out, during FY 2021/22
- ○ **Deliverable 0.3** : A report on coordination with UKAEA and other work groups, including any training activities carries out, during FY 2022/23

The remaining deliverables are organised under the tasks specified in section 1.3 of the call document:

- **Task 1** : System 2-2, 2D elliptic solver in complex geometry
    - o **Deliverable 1.1:** Test cases for system 2-2.
    - o **Deliverable 1.2:** Reference implementation in BOUT++, using PETSc and Hypre. This early implementation will allow testing of the performance of Hypre, including on GPU systems (Bede), making use of ongoing collaborations with LLNL.
    - o **Deliverable 1.3:** Spectral/hp implementation in Nektar++, to be developed in collaboration with the winning bidders of call T/N/A078/20 [Performance of spectral elements].
    - o **Deliverable 1.4:** A report on progress in preconditioning of this system, developed in collaboration with the winning bidder of call T/NA084/20 [preconditioners]. Initial work may

make use of the BOUT++ implementation, but will switch to the Spectral/hp implementation when that becomes available.

- **Task 2** : System 2-3, 1D fluid solver with UQ and realistic boundary conditions
  - o **Deliverable 2.1:** Test cases for system 2-3
  - o **Deliverable 2.2:** Reference BOUT++ implementation. This will build on existing models which have been used in production (SD1D).
  - o **Deliverable 2.3:** Nektar++ implementation of the 1D system of fluid equations, including the sonic outflow (plasma sheath) boundary condition and sources. This will be used to assess numerical accuracy and stability at this boundary.
  - o **Deliverable 2.4:** Report on intrusive UQ, in collaboration with the winning bidder of project T/NA080/20 [Study of Uncertainty Quantification Techniques]. We expect that this initial work will be based on the BOUT++ implementation, and will include a roadmap for its extension to higher order implementations and eventually the full NEPTUNE system.
  - o **Deliverable 2.5**: A report on intrusive UQ, including an assessment of its likely impact on performance, and the number of moments/polynomials needed to represent the error due to boundary conditions.
- **Task 3** : System 2-4, 1D plasma model with velocity space effects
  This task is a 1D kinetic system, and serves as a useful test-bed for exploring algorithms which will be required to build the full gyrokinetic NEPTUNE system.
  - o **Deliverable 3.1**: Test cases for system 2-4
  - o **Deliverable 3.2**: Report on phase space representation through exploratory mini-apps. This will draw on previous work on kinetic (Vlasov) solvers [e.g. Arber&Vann, https://doi.org/10.1006/jcph.2002.7098] and more recent work on the Gkeyll gyrokinetic code, in particular basis function choice, and positivity preserving reconstructions [Hakim et al. https://doi.org/10.1063/1.5141157]. System 2-4 will be implemented, likely in a high-level language such as Python or Julia, to explore these approaches.
  - o **Deliverable 3.3:** Report on time-stepping methods for kinetic systems, using implementations in high-level languages such as Python or Julia: Compare the accuracy and robustness of time integration, with an assessment of the impact of this choice on parallel scaling to Exascale. The baseline comparison will be between an explicit SSP-RK scheme, and a fully implicit Jacobian-free method such as LSODE or CVODE. If time permits, we would explore Semi-Lagrangian methods, which have potential stability advantages. A further extension would be to explore asynchronous or local time-stepping methods, which have the potential to reduce or eliminate global communications, and so may scale more favourably to Exascale hardware.
- **Task 4 :** System 2-5, Spatially 1D multispecies plasma model
  - o **Deliverable 4.1** : Test cases for system 2-5
  - o **Deliverable 4.2** : Reference BOUT++ implementation. During the last 2-3 years we have developed and implemented methods to modularise and generalise simulations codes to multiple species. We will build on this as-yet unpublished work on multifluid versions of the SD1D and the Hermes-3 BOUT++ models.
  - o **Deliverable 4.3** : Report on DSL adoption, and approaches to code design which are flexible in the number of equations/species solved. This will be done in collaboration with the successful bidder to T/NA086/20 [DSL and code generation], to test the methods developed under that call in a complex setting. We will work with the successful bidders of T/N/A078/20

[Spectral Elements] to explore how this same DSL could drive both BOUT++ and Nektar++ solvers.

- **Task 5 :** System 2-6, 2D plasma model with velocity space effects
This referent model will be important for further development, and is an ambitious undertaking: It must incorporate research and development carried out during the course of previous tasks within this project, and also work carried out by other projects. In particular the outputs of the T/N/A078/20 [Spectral Elements], T/NA086/20 [DSL and code generation techniques] and T/NA084/20 [preconditioners].
    - **Deliverable 5.1 :** Test cases for system 2-6
    - **Deliverable 5.2:** A report on progress in FY 2021/22, detailed assessment and roadmap for completion of this proxyapp. At the current time the resources required to carry out this task are uncertain. Work on deliverables 1.3 and 2.3, and coordination with the other NEPTUNE groups, will enable an assessment of the steps required to implement system 2-6 in a Spectral/hp framework, and any obstacles to be overcome.
    - **Deliverable 5.3:** A report on progress in FY 2022/23 towards implementation in Nektar++ or other Spectral/hp framework, all source code and documentation produced, and an updated roadmap leading to the full NEPTUNE code.

3.2 Exclusions

3.3 Constraints

- A number of deliverables are dependent on, or would benefit significantly from, the outputs from other successful bidders. These are highlighted in the description of deliverables above, and include the outputs of T/N/A078/20 [Performance of spectral elements], T/NA084/20 [preconditioners], T/NA080/20 [Study of Uncertainty Quantification Techniques] and T/NA086/20 [DSL techniques]. We will actively engage with these groups, and have allocated time and resources to coordination activities. To mitigate the risk that this implies, we have identified fall-back activities within each deliverable, so that we can ensure that progress is made towards the project goals.

4 Approach

This work package links together the work of several work packages, and so activities and associated resources to form and coordinate a community around the NEPTUNE project is included as deliverables 0.1-0.3. The proposed approach is based on our experience in building and maintaining large open-source plasma physics simulation codes, and coordination of communities around them. Most relevant to this call is our expertise and experience with BOUT++ and GS2: In the first FY (deliverable 0.1) we will use online meetings with the other successful bidding groups and UKAEA, to establish contacts and create a community. The outcome of these meetings will be to establish agreed community guidelines and software development practices to be used in developing our proxyapps, building on the NEPTUNE charter. Our expectation at this stage is that proxyapps will be made available under a BSD 3-clause licence; development will be carried out in public Github repositories, using a system of pull requests and code review, with automated tests using a continuous integration/development service, an approach successfully used in BOUT++ development. In addition to these processes which increase the development efficiency, speed and reliability, we will also establish a performance testing methodology, in discussion with UKAEA, so that performance regressions can be identified and quickly addressed. To do this we have access to HPC clusters at both York and Exeter, including GPU nodes, x86 and Power 9 processors, to ensure performance portability and scalability of the software algorithms.

Once the development environment and connections are established, it will be necessary to continue coordination between work packages, so that we can make use of the outputs of these other groups, and ensure that they benefit from our work. During FY 2021/22 and 2022/23 we have therefore included coordination activities across groups (deliverables 0.2 and 0.3). These will be used to ensure that development across the groups converges on common standards, APIs, libraries, build systems, and vocabularies/ontologies, all of which will be vital for the overall success and sustainability of NEPTUNE. COVID-related restrictions on travel and gathering are likely to be in place for a significant fraction of this project, and so the majority of seminars and meetings will be held online.

Management, oversight and reporting of the project activities will be led by the PI Ben Dudson, working closely with the other researchers involved. The initial work on Nektar++ implementations will be led by Dr David Moxey at the University of Exeter, supervising a PDRA to do this work. Some University of York staff time is also included in the Nektar++ implementation development: This will be used initially to enable Dr Peter Hill to participate in deliverables 1.3 and 2.3, before widening the number of researchers involved as focus shifts to Spectral/hp implementations. This will transfer plasma physics knowledge to the University of Exeter staff, in particular the PDRA involved, and enable the University of York RSEs and staff to gain experience of applying Nektar++ to plasma physics applications. Work on DSLs, performance portability and UQ aspects of the deliverables will be led by Dr Steven Wright, who will coordinate with the successful bidders to call T/NA086/20 [DSL techniques]. This methodology will enable us to build a team with the capability to work collaboratively across disciplines, in order to deliver the ambitious objectives of NEPTUNE and ExCALIBUR projects.

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible Owner	Required Data
Deliverable 1.4 depends on work carried out under work package reference T/NA084/20, where preconditioners will be developed, and applied to the systems implemented in this project. The required activity is scheduled to run in the second half of FY 2021/22. If T/NA084/20 is able to provide input then we will build on that work; if not then we have experience with physics-based preconditioning which we would employ to perform initial exploration of preconditioning.	Successful bidder to work package T/NA084/20	
Deliverable 2.4 depends on work package reference T/NA080/20 [Study of Uncertainty Quantification Techniques]. This activity is scheduled to start in April 2022, and will benefit from work carried out under that project. If for some reason such input is not available, we will draw on literature, and discussions with UKAEA and the ExCALIBUR/NEPTUNE community to guide our efforts.		
Deliverable 4.3 depends on work package T/NA086/20 [DSL and code generation techniques]		

Deliverable 5.3 brings together much of the previous work under this work package, together with work under other work packages.	NEPTUNE team	

6 Activity Plan

In all following activity descriptions the deliverables constitute milestones towards the overall task objective. Within each financial year, progress towards deliverables will be reported according to the schedule specified in Part 1 of the call document: Monthly and quarterly reports will summarise status against our expectations, and annual reports will report deliverables due to be completed in the corresponding financial year.

Task 0: NEPTUNE community coordination	**Activity No**	0
Activity: To set up and maintain the software repository, testing, documentation, and performance evaluation environment for proxyapp development; Organise meetings with NEPTUNE partners to exchange knowledge and build a community; Coordinate activities and training; disseminate results and engage collaborators through conference & workshop attendance. **Assignee:** Dr Peter Hill		
Objective 1: To establish community standards and strong links between NEPTUNE groups **Objective 2:** Set up and maintain the software development environment, coordinating with other NEPTUNE groups. **Objective 3:** Disseminate findings of other activities, within and outside the NEPTUNE project partners.		
Key Deliverables: **1)** Deliverable 0.1 **2)** Deliverable 0.2 **3)** Deliverable 0.3	**Start and Completion date:** 1) 4/1/2021 - 31/3/2021 2) 1/4/2021 - 20/3/2022 3) 1/4/2022 - 20/6/2023	**Assignee:** 1) Peter Hill 2) Peter Hill 3) Peter Hill

Task 1 : System 2-2, 2D elliptic solver in complex geometry	**Activity No**	1
Activity: To define a set of test cases and performance metrics; implement the system in both BOUT++ and Nektar++ frameworks and test; apply and evaluate preconditioning methods developed in other work packages, or if not available carry out development work. **Assignee:** Dr Ben Dudson		
Objective 1: Assess the performance of a Nektar++ implementation in simplified tokamak geometry against a reference BOUT++ case **Objective 2:** Assess the additional numerical cost of realistic tokamak geometry solves with detailed geometry (Nektar++ only)		

Objective 3: Evaluate preconditioning strategies, building on work performed under work package T/NA084/20 or developed under this work package.		
Key Deliverables: 1) Deliverable 1.1 2) Deliverable 1.2 3) Deliverable 1.3 4) Deliverable 1.4	**Start and Completion date:** **1)** 4/1/2021 - 20/3/2021 2) 4/1/2021 - 20/3/2021 3) 1/4/2021 - 20/3/2022 4) 1/4/2021 - 20/3/2022	**Assignee:** 1) Ben Dudson 2) Ben Dudson 3) David Moxey 4) Ben Dudson

Task 2 : System 2-3, 1D fluid solver with UQ and realistic boundary conditions	**Activity No**	2
Activity: Define a set of tests and performance measures; implement the system in BOUT++ and Nektar++ frameworks; use these implementations to explore intrusive UQ techniques. **Assignee:** Dr Ben Dudson		
Objective 1: Evaluate the performance and stability of Spectral/hp methods with sheath boundary conditions **Objective 2:** Compare the performance of finite difference and spectral/hp implementations **Objective 3:** Explore the implementation and implications of intrusive UQ		

Key Deliverables:	**Start and Completion date:**	**Assignee:**
1) Deliverable 2.1 2) Deliverable 2.2 3) Deliverable 2.3 4) Deliverable 2.4 5) Deliverable 2.5	1) 4/1/2021 - 20/3/2021 **2)** 1/4/2021 - 20/3/2022 **3)** 1/4/2021 - 20/3/2022 **4)** 1/4/2021 - 20/3/2022 5) 1/4/2022 - 20/6/2023	1) Ben Dudson 2) Ben Dudson 3) David Moxey 4) Steven Wright 5) Steven Wright

Task 3 : System 2-4, 1D plasma model with velocity space effects		**Activity No**	3
Activity: Define tests to exercise physics of 1D kinetic model of the tokamak SOL; compare phase space representations in terms of accuracy and scaling; compare time stepping approaches, with a view to their scaling to large problems and exascale hardware. **Assignee:** Dr David Dickinson			
Objective 1: Create a set of test cases against which implementations can be compared for accuracy, performance and scaling **Objective 2:** Provide guidance on the choice of phase space representation for kinetic systems, in collaboration with other work packages **Objective 3:** Provide guidance on choice of time stepping method for kinetic systems, to inform the design of future NEPTUNE models.			
Key Deliverables: 1) Deliverable 3.1 2) Deliverable 3.2 3) Deliverable 3.3	**Start and Completion date:** 1) 1/4/2021 - 20/3/2022 2) 1/4/2021 - 20/3/2022 3) 1/4/2021 - 20/3/2022	**Assignee:** 1) David Dickinson 2) David Dickinson 3) David Dickinson	

Task 4 : System 2-5, Spatially 1D multispecies plasma model		**Activity No**	4
Activity: To define correctness tests and performance metrics; implement system 2-5 in BOUT++, building on and extending design patterns for modular code reuse; apply and evaluate DSL techniques in collaboration with the NEPTUNE DSL and code generation work package. **Assignee:** Dr Steven Wright			
Objective 1: Define a set of tests and performance metrics against which implementations can be compared **Objective 2:** Implement a reference BOUT++ implementation **Objective 3:** Guide the choice DSL and code design pattern for NEPTUNE, through implementation in a plasma model and exploration of the performance and code maintenance trade-offs.			
Key Deliverables: 1) Deliverable 4.1 2) Deliverable 4.2 3) Deliverable 4.3	**Start and Completion date:** 1) 1/4/2021 - 20/3/2022 2) 1/4/2021 - 20/3/2022 3) 1/4/2021 - 20/3/2022	**Assignee:** 1) Ben Dudson 2) Ben Dudson 3) Steven Wright	

Task 5 : System 2-6, 2D plasma model with velocity space effects	Activity No	5
Activity: Define tests and performance metrics; evaluate the steps required to implement system 2-6, integrating the outputs of earlier tasks and other NEPTUNE work packages; begin implementation of the code base and documentation; design and refine the roadmap to completion of the implementation. **Assignee:** Dr David Moxey		
Objective 1: Define tests and performance metrics for system 2-6 **Objective 2:** Begin implementation in Nektar++, with a design taking into account earlier results and other work package outputs **Objective 3:** Plan a recommended roadmap to completion of the implementation		
Key Deliverables: 1) Deliverable 4.1 2) Deliverable 4.1 3) Deliverable 4.1	**Start and Completion date:** 1) 1/4/2021 - 20/3/2022 2) 1/4/2021 - 20/3/2022 3) 1/4/2022 - 20/6/2023	**Assignee:** 1) David Moxey 2) David Moxey 3) David Moxey

Name of Bidding Organisation:	UKRI Science and Technology Facilities Council
Contract Title:	HT03467 NA084 Matrix-Preconditioning

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 2.3.

At the heart of the NEPTUNE Programme is the solution of coupled partial differential equations (PDEs) that exhibit challenging properties when trying to solve them. Following discretisation of the PDE, a large linear system ***Ax=b*** must be solved or a series of large linear systems: the size of the systems to be solved mean that iterative methods are preferable over direct factorisation methods but, without preconditioning to improve rate of convergence, the number of iterations needed to solve the system is too large and it is necessary to stop the iterative process before the required levels of accuracy are met. This loss in accuracy can have a detrimental effect on the convergence properties of any overarching iterative methods that are, for example, dealing with the non-linearity of the PDEs being solved. By using a preconditioner that is suitable for exascale computations, is relatively cheap to initialise and apply, and significantly speeds up the convergence towards the solution of ***Ax=b***, simulations should become faster to run and have improved accuracy. This allows the computational simulation of models that were previously infeasible to become feasible to run.

In the past couple of decades, a lot of research has been done with the aim of efficiently solving a range of individual partial differential equations (PDEs) for different discretisation methods. These preconditioners are often developed by considering the underlying mathematical problem being solved, for example, multigrid methods [1] [2]. The first part of our proposal involves **literature surveys** to identify the most appropriate existing preconditioners specifically designed for elliptical and hyperbolic PDEs identified within the NEPTUNE Programme. **We will collate information about their convergence properties, suitability for exascale and whether codes already exist that can be utilised within prototype tests. Consultation within the NEPTUNE collaboration will allow us to identify which PDEs to target and, following the literature survey, which of these PDE-based preconditioners are of most interest.**

There has also been a lot of development in more generalised preconditioning methods that only assume knowledge of the entries of the matrix A or the action of that matrix when multiplying the matrix with a series of vectors, for example, preconditioners based on sparse approximate inverses [8] [9] or incomplete factorisations [10] [11]. Of particular interest are Markov Chain Monte Carlo Matrix Inverse (MCMCMI) preconditioners, which are a class of sparse approximate inverse preconditioners. Alexandrov et al have shown that MCMCMI preconditioners very suitable for predicted exascale architectures and perform well [12] [13]. In particular, the MCMCMI preconditioner has been shown to be effective when reduced precision is used to form and apply the preconditioner via GPGPUs, which makes this preconditioner a cheap choice. Therefore, we will also **perform a literature and code survey on sparse approximate inverse preconditioners with a particular focus on MCMCMI preconditioners due to their expected suitability for exascale computing.**

The information from these literature and code surveys will be disseminated to other areas of the NEPTUNE Programme to help inform their decisions about numerical scheme choices, as required by this task.

When *n* PDEs are coupled together, as is required within the NEPTUNE Programme, the linear system will have a block-structure of the following form [3]:

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{bmatrix}, \tag{1}$$

where $A_{i,i}$ represents the discretisation of the *i*-th PDE and $A_{i,j}$ $(i \neq j)$ couple the *i*-th and *j*-th PDEs. As observed within the SD1D example, which uses BOUT++ [4], a simple yet effective preconditioner for a single PDE can be no longer effective when coupling is introduced and, hence, other tactics are required. Thorne (née Dollar) has extensive experience in designing, analysing and implementing preconditioners for block-structured systems, in particular, designing highly effective implicit-factorisation preconditioners [5] [6], which were very successfully used when some of the sub-blocks were discretisations of PDEs and the overall preconditioner incorporated known preconditioners for these PDEs [7]. **These implicit-factorisation preconditioners are related to the block-factorisation approach proposed by Ferronato *et al* [3] and, by combining these approaches, we will aim to form preconditioners that significantly improve the convergence of the iterative method being used to solve the system but are also relatively cheap to initialise. We will ensure that our implicit-factorisation preconditioning framework will allow us to precondition both close-coupled and loose-coupled problems.** It may also allow us to use reduced-subspace iterative methods based on iterative methods that are not immediately considered suitable for applying to (1) but have superior properties [5] [6]. **This approach will allow the Programme to solve the coupled systems in a manner suggested by the underlying physics and meets the associated task package objective.**

We will compare at least two classes of preconditioners for non-coupled problems: SPAI preconditioners and PDE-based preconditioners. For coupled problems, we will form MCMCMI preconditioners for the whole of ***A***, and also use produce a hybrid method where the MCMCMI preconditioners are used within the sub-blocks of our proposed implicit-factorisation preconditioners: these preconditioners will be compared with the implicit-factorisation preconditioners that feature PDE-based preconditioners. **Hence, we will be comparing at least three different preconditioning approaches for coupled problems and will compare the differences with regards whether the problem is close-coupled or weak-coupled.** Due to the complex nature of the coupled problems within the NEPTUNE Programme, this comparison will help us understand whether PDE-based preconditioners are required, which could require a different preconditioner for each PDE, or if a more generalised preconditioner can be used. If the latter is possible, then this will simplify the task of developing preconditioners for more complex coupled problems. **By drawing together the results from our literature and code surveys, implicit-factorisation preconditioner design (theoretical and exascale considerations), and numerical tests, we will be able to categorise the suitability of different approaches with regards to robustness, efficiency, suitability with respect to predicted exascale computing architectures, and how quickly they can be developed within the NEPTUNE Programme. This will then be used to provide our recommendations and, hence, the final outcome of this project will guide the Programme in its choice of preconditioner implementations within its proxyapp developments, which completes the overall objective for this NEPTUNE Task.**

References

[1] U. Trottenberg, C. W. Oosterlee and A. Schuller, Multigrid, Elsevier, 2000.

[2] Y. Shapira, “Matrix-Based Multigrid for Locally Refined Meshes,” in *Matrix-Based Multigrid*, Springer, 2003, pp. 133-166.

[3] M. Ferronato, A. Franceschini, C. Janna, N. Castelletto and H. A. Tchelepi, “A general preconditioning framework for coupled multiphysics problems with application to contact-and poro-mechanics,” *Journal of Computational Physics,* vol. 398, p. 108887, 2019.

[4] B. D. Dudson, M. V. Umansky, X. Q. Xu, P. B. Snyder and H. R. Wilson, “BOUT++: A framework for parallel plasma fluid simulations,” *Computer Physics Communications,* vol. 180, no. 9, pp. 1467-1480, 2009.

[5] H. S. Dollar, N. I. M. Gould, W. H. A. Schilders and A. J. Wathen, “Implicit-Factorization Preconditioning and Iterative Solvers for Regularized Saddle-Point Systems,” *SIAM Journal on Matrix Analysis and Applications,* vol. 28, no. 1, pp. 170-189, 2006.

[6] H. S. Dollar, N. I. Gould and A. J. Wathen, “On implicit-factorization constraint preconditioners,” *Large-scale nonlinear optimization,* pp. 61-82, 2006.

[7] T. Rees, H. S. Dollar and A. J. Wathen, “Optimal Solvers for PDE-Constrained Optimization,” *SIAM Journal on Scientific Computing,* vol. 32, no. 1, pp. 271-298, 2010.

[8] J. D. F. Cosgrove, J. C. Diaz and A. Griewank., “Approximate inverse preconditionings for sparse linear systems,” *International journal of computer mathematics,* vol. 44, no. 1-4, pp. 91-110, 1992.

[9] H. Anzt, T. K. Huckle, J. Bräckle and J. Dongarra, “Incomplete sparse approximate inverses for parallel preconditioning,” *Parallel Computing,* vol. 71, pp. 1-22, 2018.

[10] D. S. Kershaw, “The incomplete Cholesky—conjugate gradient method for the iterative solution of systems of linear equations,” *Journal of Computational Physics,* vol. 26, no. 1, pp. 43-65, 1978.

[11] J. D. Booth, K. Kim and S. Rajamanickam, “A comparison of high-level programming choices for incomplete sparse factorization across different architectures,” in *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2016.

[12] A. Lebedev and V. Alexandrov, “On Advanced Monte Carlo Methods for Linear Algebra on Advanced Accelerator Architectures,” in *2018 IEEE/ACM 9th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (scalA)*, 2018.

[13] M. E. Şahin, A. Lebedev and V. Alexandrov, “Empirical Analysis of Stochastic Methods of Linear Algebra,” in *nternational Conference on Computational Science*, 2020.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

A. Report: Literature survey of preconditioning approaches for elliptic problems, which are based on the underlying mathematical structure of the PDEs being solved and discretisation method. Information of gathered within the survey will include: theoretical convergence results for applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes.

B. Report: Literature survey of preconditioning approaches for hyperbolic problems, which are based on the underlying mathematical structure of the PDEs being solved and discretisation method. Information of gathered within the survey will include: theoretical convergence results for applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes.

C. Report: Literature survey of preconditioning approaches based on sparse approximate inverses, in particular, focussing on Markov Chain Monte Carlo approximate inverses. These are already being developed with exascale architecture consideration [12] [13] and use the entries of the matrix **A** to compute a preconditioner with no prior knowledge of the underlying mathematical problem. Information of gathered within the survey will include: theoretical convergence results for applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes.

D. Report: For both close-coupled problems and weak-coupled problems, the development of suitable implicit-factorisation preconditioners and reduced subspace iterative methods building on the work of Thorne [5] [6] [7] and Ferronato *et al* [3]. These implicit-factorisation preconditioners will be able to incorporate the preconditioners identified from our literature surveys. Theoretical convergence results will be developed.

E. Report and Prototype Code: Following consultation with the consortia, tests will be performed to solve selected close-coupled problems through the development of a prototype test code. Comparisons will be made between the use of block/implicit factorisation preconditioners and applying the Markov Chain Monte Carlo approximate inverse preconditioner to the whole system, as well as with weak-coupled versions of the problems. From this, recommendations will be made to allow deferred implementations to be facilitated within the next stage of the NEPTUNE Programme. *Note: the reports for D and E may be combined into 1 report*.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

- Development of proxyapp implementation of preconditioners that have been identified as promising. This is something that we would like to do but is out of scope due to the timescales.
- Comparison of all combinations of promising preconditioners from the literature surveys [ADD]

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

- The available funding limits the amount time that can be devoted to this project, so it will be important to restrict the comparison to the higher priority coupled and non-coupled problems.

4 Approach

Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

Project Delivery Groups are responsible for developing and delivering the individual projects under each of the different programmes and funding streams. This includes identifying and mitigating potential risks; resolving issues within the remit of the project; developing detailed implementation plans and taking day-to-day responsibility for implementation of the project against the agreed scope once approved.
STFC project management is broadly based around the Association for Project Management, however the Hartree Centre's approach to project delivery additionally aims to combine the flexibility and responsiveness of AGILE with the governance of PRINCE2 when deemed appropriate for the type of project. The Hartree Centre is ISO 9001 accredited,

by demonstrating the ability to successfully complete projects and consistently meet customer and regulatory requirements. The key principles relating to the delivery of projects are:

- All individual projects complete internal management reviews and the 'Quality at Entry' (QaE) process before being submitted for Management Board approval.
- Projects that the Management Board feel warrant a more cautious approval process must include subsequent Decision Reviews at agreed points to assess the ongoing feasibility of the project with a Go/No-Go decision.
- Once approved, projects will be managed on a day to day basis by exception by an appropriate PMO representative and/or Programme Manager.
- Prioritisation and effort will be applied to ensuring the delivery of requirements, outputs, and acceptance criteria using the MoSCoW1 technique.
- Progress and success will usually be measured by completion of outputs rather than the completion of individual tasks.
- Each project must confirm the minimum acceptable outcomes (MUSTs) the project will deliver at the point of approval. These will form the minimum base line for project outcomes.
- Any change in 'Must have' requirements, timescales or budgets after a project has been approved will trigger an exception and must be approved by the Management Board via the Hartree Centre's Change Control process.
- The Operations board can approve changes to "Should have" requirements providing they are within the tolerance boundaries set for time and budget.
- The Project Manager can include or de-scope 'Could have' requirements as long as they are within the base lines set for time and budget.
- Any tolerances for budget, time and quality (Outcomes) will be defined and agreed for each project during the approval phase. Tolerances will be agreed at project approval stage and may only be changed by submission of a change request to the Management Board.
- Projects will review progress regularly and report progress or exceptions to the Operations Board at minimum on a monthly basis.

Customer updates and progress reports will be agreed as part of the project planning phase and will be completed as a minimum on a monthly basis.

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible Owner	Required Data
Test matrices and right-hand sides to test preconditioners. This could be in the form of individual test matrices or codes that generate these matrices	UKAEA	Test matrices and right-hand sides for non-coupled and coupled problems
Interaction with project partners for prioritisation of PDEs	UKAEA	PDE prioritisation list

6 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

	Activity No	**1**
Activity: Literature survey: preconditioners for elliptic PDEs from fusion modelling **Assignee:** Sue Thorne		
Objective 1: Identify key elliptic PDEs for fusion modelling and proposed discretisation methods **Objective 2:** Perform literature and code surveys to find preconditioners for the identified elliptic PDEs and their discretisations. **Objective 3:** Catalogue each preconditioner in a report, recording theoretical convergence results for applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes.		
Key Deliverables: **D1.1)** Report cataloguing preconditioners and codes for prioritised elliptic PDEs and associated discretisation methods	**Start and Completion date:** **1)** 04/01/2021-26/02/2021	**Assignee:** **1)** Sahin & Thorne
Milestones towards deliverables: **M1.1)** Prioritisation of elliptic PDEs and discretisation methods **M1.2)** Complete literature and code survey for preconditioners applicable to prioritised elliptic PDEs and associated discretisation methods	**Completion date:** **1)** 20/01/2021 **2)** 19/02/2021	**Assignee:** **1)** Sahin & Thorne **2)** Sahin

	Activity No	**2**
Activity: Literature survey: preconditioners for hyperbolic PDEs from fusion modelling **Assignee:** Sue Thorne		
Objective 1: Identify key hyperbolic PDEs for fusion modelling and proposed discretisation methods **Objective 2:** Perform literature search to find preconditioners for the identified hyperbolic PDEs and their discretisations. **Objective 3:** Catalogue each preconditioner in a report, recording theoretical convergence results for		

applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes.		
Key Deliverables: **D2.1)** Report cataloguing preconditioners and codes for prioritised hyperbolic PDEs and associated discretisation methods	**Start and Completion date:** **1)** 04/01/2021-26/02/2021	**Assignee:** **1)** Sahin & Thorne
Milestones towards deliverables: **M2.1)** Prioritisation of hyperbolic PDEs and discretisation methods **M2.2)** Complete literature and code survey for preconditioners applicable to prioritised hyperbolic PDEs and associated discretisation methods	**Completion date:** **1)** 20/01/2021 **2)** 19/02/2021	**Assignee:** **1)** Sahin & Thorne **2)** Sahin

Activity No	3	
Activity: Literature survey: sparse approximate inverse (SPAI) preconditioners **Assignee:** Sue Thorne		
Objective 1: Perform literature search to identify classes of sparse approximate inverses **Objective 2:** Catalogue each preconditioner in a report, recording theoretical convergence results for applicable iterative methods; cost of setting-up the preconditioner and executing the preconditioner; suitability of the preconditioner for Exascale computation; availability of existing codes. Particularly focus on MCMCMI Preconditioners.		
Key Deliverables: **D3.1)** Report cataloguing SPAI preconditioners and codes	**Start and Completion date:** **1)** 09/02/2021-05/03/2021	**Assignee:** **1)** Alexandrov, Sahin & Thorne
Milestones towards deliverables: **M3.1)** Complete literature and code survey for SPAI preconditioners	**Completion date:** **1)** 26/02/2021	**Assignee:** **1)** Sahin

Activity No	4
Activity: Theoretical development of implicit-factorisation preconditioners for coupled problems **Assignee:** Thorne	

Objective 1: Understand the differences, in general, between the problem set-up for loosely-coupled and strongly-coupled problems from NEPTUNE. This will be key to developing appropriate implicit-factorization/block-based preconditioners. **Objective 2:** Develop implicit-factorization preconditioners and any associated convergence properties for inclusion in Key Deliverable from Activity 5.		
Key Deliverables: **D4.1)** Section on implicit-factorisation preconditioners for inclusion in Activity 6's report or separate report	**Start and Completion date:** **1)** 08/03/2021-31/03/2021	**Assignee:** **1)** Sahin & Thorne
Milestones towards deliverables: **M4.1)** Set of proposed implicit-factorisation preconditioners and initial theoretical results	**Completion date:** **1)** 20/03/2021	**Assignee:** **1)** Sahin &Thorne

	Activity No	**5**
Activity: Development of prototype code: implicit-factorisation preconditioners for coupled problems **Assignee:** Thorne		
Objective 1: Develop prototype code for proposed implicit-factorisation preconditioners including documentation for code		
Key Deliverables: **D5.1)** Prototype code **D5.2)** Documentation including readme file, installation documentation and simple user documentation	**Start and Completion date:** **1)** 01/04/2021-31/07/2021 **2)** 01/04/2021-31/07/2021	**Assignee:** **1)** Sahin **2)** Sahin
Milestones towards deliverables: **M5.1)** Prototype code for loosely & strongly coupled problems (Version 1) **M5.2)** Prototype code and documentation for loosely & strongly coupled problems (Final version)	**Completion date:** **1)** 07/05/2021 **2)** 31/07/2021	**Assignee:** **1)** Sahin & Thorne **2)** Sahin

	Activity No	6

Activity:
Comparison of preconditioners for fusion modelling
Assignee: Thorne

Objective 1: Comparison of preconditioners for non-coupled elliptic and hyperbolic problems from NEPTUNE Programme. Comparison will include scalability results for each preconditioner, preconditioner set-up times, preconditioner execution time for single application of preconditioner to vector, number of iterations for chosen iterative method and time to solution.
Objective 2: Comparison of preconditioners for coupled elliptic and/or hyperbolic problems from NEPTUNE Programme to understand effective of proposed approaches: SPAI-based preconditioner to be applied to whole system assuming no knowledge of block structure; implicit-factorization based preconditioner incorporating PDE-based preconditioner; implicit-factorization based preconditioner incorporating SPAI-based preconditioner. Comparison will include scalability results for each preconditioner, preconditioner set-up times, preconditioner execution time for single application of preconditioner to vector, number of iterations for chosen iterative method and time to solution.
Objective 3: Use comparison data to provide recommendations as to which preconditioners should be incorporated into the proxyapps being developed as part of the NEPTUNE programme.

Key Deliverables:	Start and Completion date:	Assignee:
D6.1) Test code(s) for preconditioner comparisons	**1)** 11/05/2021-31/07/2021	**1)** Sahin
D6.2) Data from preconditioner comparison tests for non-coupled problems	**2)** 11/05/2021-04/06/2021	**2)** Sahin & Thorne
D6.3) Data from preconditioner comparison tests for coupled problems	**3)** 24/05/2021-09/07/2021	**3)** Sahin & Thorne
D6.4) Report containing data analysis and recommendations	**4)** 11/05/2021-31/07/2021	**4)** Alexandrov, Sahin & Thorne

Milestones towards deliverables:	Completion date:	Assignee:
M6.1) Test code(s) & documentation for preconditioner comparison (Version 1) finished and added to repository	**1)** 21/05/2021	**1)** Sahin
M6.2) Creation of non-coupled test set and added to repository	**2)** 21/05/2021	**2)** Sahin & Thorne
M6.3) All data from non-coupled test runs added to repository	**3)** 04/06/2021	**3)** Sahin
M6.4) Creation of coupled test set and added to repository	**4)** 11/06/2021	**4)** Sahin & Thorne
M6.5) All data from coupled test runs added to repository	**5)** 09/07/2021	**5)** Sahin
M6.6) Test code(s) for preconditioner comparison (Final Version) finished and added to repository	**6)** 31/07/2021	**6)** Sahin

Name of Bidding Organisation:	University of Oxford
Contract Title:	FM-WP2 Plasma Multiphysics model: Development of gyro-averaged referent model (Project NEPTUNE)

1 Purpose of Document

This document provides a statement of how and when the Research Plan's objectives would be achieved, by showing the major products, activities and resources required of the Research Plan.

2 Benefits and alignment to Work Package objectives

Please evidence below how the submission aligns to the Work Package objectives outlined in Part 1 Section 1.3.

- We will develop different drift kinetic models applicable to the plasma edge and compare them in various simplified settings to propose a referent model, which is one of the stated purposes of Work Package FM-WP2.
- The different drift kinetic approaches will be tested in proxy-apps that will solve the drift-kinetic equivalent of the following problems (numbering corresponds to the one used in the document Fusion Modelling System Science Plan):

  4. Spatially 1D plasma model incorporating velocity space effects.

  6. Spatially 2D plasma model incorporating velocity space effects.

  These problems are of interest for Work Package FM-WP2.
- We will include a simplified model for neutrals to see how it affects the plasma behaviour. This will facilitate future interaction with FM-WP3.
- As asked in section 1.3 of part 1, we will consider geometries (i)-(iv) (slab-twisted/toroidal) by ultimately constructing a model for helical magnetic fields.
- As asked in section 1.3 of part 1 of the bid, we will construct our models so that they transition smoothly from core δf gyrokinetics to drift kinetics and from drift kinetics to fluid equations as the "magnetisation" parameter δ_s and normalised collision frequency ν^* change across the edge plasma.
- As asked in section 1.3 of part 1 of the bid, we will develop boundary conditions to connect to the non-neutral sheaths that cover the walls of the tokamak, and we will discuss how to connect to fluid equations and δf gyrokinetics.

3 Scope

3.1 Key Deliverables and/or Desired Outcomes

- Report describing the important regimes in the tokamak plasma edge and a review of the state-of-the-art, including prioritisation of key physical processes for modelling and for further research, as requested in section 2.1 of part 1 of the bid.
- Proxy-app that compares different drift kinetic models for a spatially 1D plasma.
- Proxy-app that compares different drift kinetic models for a spatially 2D plasma.
- Report on the full drift kinetic model for helical magnetic field, including material to assist numerical implementation and a discussion of possible extensions of the model.

3.2 Exclusions

Activities/topic areas that are out of scope of the Bid and which will not be undertaken (may also include things that Bidder would like to do but are not currently in scope)

- We will develop a model for helical magnetic fields to ensure that we include toroidicity into the problem, but we do not propose to develop a model for a full tokamak magnetic field. We have two reasons to justify this strategy:

  (i) helical magnetic fields are the current state-of-the-art for edge kinetic codes [N. R. Mandell et al, *J. Plasma Phys.* **86**, 905860109 (2020)], and

  (ii) very recent work suggests that the radial components of the curvature and ∇B drifts have been treated improperly in previous edge plasma models [J. F. Parisi et al, accepted in *Nucl. Fusion*, https://arxiv.org/abs/2004.13634].

  More research, outside of the scope of this bid, is needed to determine the correct way to introduce the radial components of the curvature and ∇B drifts into drift kinetic models for the edge.

3.3 Constraints

Restrictions that affect proposals of the project by imposing limitations such costs, resources or project schedule, which may affect the execution of the Bid.

- The proposed time commitment of Professors Felix I. Parra Diaz and Michael Barnes has been calculated to ensure that they are able to be relieved of some of their teaching duties for the duration of the project. Any reduction in the funds might lead to a very sharp reduction in the deliverables.

4 Approach

Describe how will the work be undertaken, including a definition of methodology that will be used in the project to deliver the work package and call objectives.

We propose the development of an electrostatic drift-kinetic model for a magnetized plasma in a helical magnetic field composed of one ion species, electrons and one neutral species. Drift kinetics is an expansion in the “magnetisation” parameter δ_s that eliminates the fast gyro-orbit timescale and one of the velocity space coordinates: the gyrophase [1]. Our proposal addresses four challenges that are crucial for drift kinetic treatments of the edge plasma:

- Unfortunately, a piece of the electric field (the radial component) cannot be determined with the lowest order drift kinetic equations [2]. Higher order terms have to be kept, and the most common way to do so is to split the distribution function and the electromagnetic fields into a background, slowly varying piece and a fast-evolving, small piece: δf gyrokinetics [3, 4]. This procedure is not valid for edge plasmas. As a result, terms of different order in the expansion must be included in the model for edge plasmas, leading to equations with terms of very different size and hence to numerical constraints. For example, the time step size in electrostatic drift kinetics is heavily constrained by an uninteresting and unphysical wave known as the electrostatic shear Alfven wave [5, 6].
- Drift kinetic formulations are valid everywhere in the edge, but they are overly complicated for the core, where δf gyrokinetics should be applied, and for the far scrape-off layer, where collisions are frequent and fluid models can be used [7]. It would be highly desirable to have an edge drift kinetic

model that smoothly converts into these other simpler, faster models for the tokamak core and the far scrape-off layer.

- To our knowledge, the drift kinetic expansion in the presence of a wall has only been considered recently [8], and it needs to be developed further.
- To our knowledge, drift kinetic expansions have not been performed including neutral physics. The inclusion of neutrals breaks the usual drift kinetic expansion because neutrals are not magnetised and hence can introduce dependence on the gyrophase back into the equations.

To address the first two points, we propose a new method to derive drift kinetics. Instead of working with the distribution function $f(\mathbf{x}, \mathbf{v}, t)$ that depends on position $\mathbf{x}$, velocity $\mathbf{v}$ and time t, we will develop a gyrokinetic set of equations that will evolve the normalized distribution function $F(\mathbf{x}, \mathbf{w}, t) = (v_{th}^3/n) f(\mathbf{x}, \mathbf{w}v_{th} + \mathbf{u}, t)$ that depends on the normalized velocity $\mathbf{w} = (\mathbf{v} - \mathbf{u})/v_{th}$. Here n is the particle density, $\mathbf{u}$ is the average velocity and v_{th} is the thermal speed. The values of n, $\mathbf{u}$ and v_{th} will be obtained from fluid-like equations. We expect this approach to have three advantages:

- The higher order terms required to obtain the radial component of the electric field will only appear in one of the equations. This equation can be numerically evolved with implicit time-stepping methods to avoid numerical problems that stem from terms of very different size co-existing in the same equation.
- Using the fluid equations for n, $\mathbf{u}$ and v_{th}, it is easier to impose conservation of global energy and momentum. Unlike other moment approaches based on polynomial expansions of f, our proposed approach will ensure that f remains positive.
- The fact that n, $\mathbf{u}$ and v_{th} are evolved independently makes matching with fluid equations and with δf gyrokinetics straightforward.

There are risks associated with the development of a new approach to drift kinetics. To mitigate those risks, in the initial stages of the project, we will compare the new approach with more conventional drift kinetic equations by implementing both approaches in one proxy-app. In this way, we can easily revert to conventional drift kinetics if necessary. We propose to build four proxy-apps to test the most important features of the new set of equations:

1. **1D model along straight magnetic field lines with periodic boundary conditions.** The idea is to start with a simple proxy-app to test the new drift kinetic approach. The usual 1D drift kinetic equation will be compared with three other models:
    a. One in which the density n is evolved using a fluid-like equation.
    b. One in which both the density n and the average velocity $\mathbf{u}$ are evolved using fluid-like equations.
    c. One in which the density n, the average velocity $\mathbf{u}$ and the thermal speed v_{th} are evolved using fluid-like equations.

    With this approach, we will test whether our proposed method to obtain drift kinetic equations leads to unforeseen numerical problems. We will also test different implicit time evolution algorithms by treating implicitly only some terms in the fluid-like equations for n, $\mathbf{u}$ and v_{th}. If we find numerical problems that are insurmountable, we will revert to the usual derivation of drift kinetics.

    The piece of the electric field that necessitates the inclusion of higher order terms in the drift kinetic equation is perpendicular to the magnetic field. Thus, in this 1D proxy-app we will not need to include terms of very different sizes.

    The proxy-app will be benchmarked against analytical solutions for Landau damping of ion acoustic waves.

2. **1D model along straight magnetic field lines with wall boundary conditions.** We want to develop a drift kinetic model in the presence of walls. This means imposing boundary conditions at the end of magnetic field lines that represent such walls [8]. This proxy-app will test the compatibility between the wall boundary conditions and the new drift kinetic approach with fluid-like equations for density, average velocity and thermal speed.

   The piece of the electric field that necessitates the inclusion of higher order terms in the drift kinetic equation is perpendicular to the magnetic field. Thus, in this 1D proxy-app we will not need to include terms of very different sizes.

   The proxy-app will be benchmarked against analytical steady state solutions obtained for particular particle sources [9, 10], and against analytical time-evolving solutions obtained for particular initial conditions [11, 12].

3. **Axisymmetric 2D model with helical field lines and wall boundary conditions.** This 2D model does not include turbulence due to axisymmetry and hence all the transport will be collisional in nature. To avoid having to include finite orbit width effects, we will assume that collisions with neutrals produce most of the transport. This limit is relevant to tokamaks, as it corresponds to the charge exchange mean free path being larger than the ion gyroradius.

   This 2D model will allow us to test the compatibility of the new drift kinetic approach with 2D geometry. The wall boundary conditions determine the piece of the electric field that necessitates the inclusion of higher order terms in the drift kinetic equation. Thus, in this proxy-app we will not need to include terms of very different sizes.

   We have decided to limit our proxy-app to helical magnetic field lines because this is the current state-of-the-art for edge kinetic codes [13] and because in this way we avoid having to choose an ordering for the radial components of the curvature and ∇B drifts. These drifts have been neglected in the past, but recently their radial components have been proven to be important in JET pedestals [14]. Further research is needed to determine the importance of these radial components, and this is outside of the scope of this proposal.

   We will benchmark this 2D model with analytical solutions in the high collisionality regime (fluid limit) [7].

4. **Axisymmetric 2D model with helical field lines and periodic boundary conditions.** The periodic boundary conditions imitate the closed flux surfaces in the tokamak core. Due to the periodic boundary conditions, there is a perpendicular component of the electric field (the radial component) for which we need higher order terms. We will test whether our new approach to drift kinetics facilitates the calculation of this component.

   We will benchmark this 2D model with analytical solutions in the high collisionality regime (fluid limit) [7].

During each stage, we will re-evaluate the numerical approach taken to ensure that it is efficient. We will endeavour to design each proxy-app to be as modular as possible to enable testing of different models/numerical approaches and to improve portability.

In addition to the two axisymmetric 2D proxy-apps with different boundary conditions, we will consider an axisymmetric plasma with two spatial regions: one with periodic boundary conditions (tokamak core) and another with wall boundary conditions (Scrape-Off Layer). We will determine the appropriate equations for the transition between these two regions. Note that the radial electric field is determined differently on each side of the 'separatrix' between the two regions.

The interactions between the neutrals and the charged particles will be modelled using Krook operators. By using simplified models for collisional interactions between charged particles and neutrals, we are able to focus on the drift kinetic treatment, while still keeping important neutral effects.

The final result of our process will be a 3D electrostatic referent drift kinetic model for helical magnetic fields. Note that we will have only built 2D proxy-apps, but the model will be fully 3D. We believe that we have chosen proxy-apps that will uncover most of the numerical problems that can be associated with the new drift kinetic approach. If the new approach to drift kinetics does not work due to unforeseen numerical problems, we will derive a new drift kinetic model with all the higher order terms needed to calculate every component of the electric field.

In addition to providing an electrostatic referent model for helical magnetic fields, we will write a report discussing three possible extensions: the inclusion of gyroaverages to convert the drift kinetic model into a gyrokinetic model, the inclusion of radial magnetic drifts, and the inclusion of electromagnetic effects.

References

[1] R. D. Hazeltine, *Plasma Phys.* **15**, 77 (1973).
[2] F. I. Parra and P. J. Catto, *Plasma Phys. Control. Fusion* **50**, 065014 (2008).
[3] E. A. Frieman and L. Chen, *Phys. Fluids* **25**, 502 (1982).
[4] I. G. Abel et al, *Rep. Prog. Phys.* **76**, 116201 (2013).
[5] W. W. Lee, *J. Comput. Phys.* **72**, 243 (1987).
[6] M. Barnes, F. I. Parra and M. Landreman, *J. Comput. Phys.* **391**, 365 (2019).
[7] P. Helander, S. I. Krasheninnikov and P. J. Catto, *Phys. Plasmas* **1**, 3174 (1994).
[8] A. Geraldini, F. I. Parra and F. Militello, *Plasma Phys. Control. Fusion* **60**, 125002 (2018).
[9] L. Tonks and I. Langmuir, *Phys. Rev.* **34**, 876 (1929).
[10] E. R. Harrison and W. B. Thompson, *Proc. Roy. Soc.* **74**, 145 (1959).
[11] J. E. Allen and J. G. Andrews, *J. Plasma Phys.* **4**, 187 (1970).
[12] N. St. J. Braithwaite and L. M. Wickens, *J. Plasma Phys.* **30**, 133 (1983).
[13] N. R. Mandell et al, *J. Plasma Phys.* **86**, 905860109 (2020).
[14] J. F. Parisi et al, accepted in *Nucl. Fusion*, https://arxiv.org/abs/2004.13634.

5 External Dependencies

Information about potential dependencies on other activities/organisations involved eg. Data that would need to have access to as part of the research, what historical data would be available to run case studies, that the Bid would benefit from

Dependency Description	Responsible Owner	Required Data

6 Activity Plan

Identify activities plans for the Research Plan (please add and use as many activity templates as required into the document and complete Annex B with schedule). Please include any relevant planning assumptions.

Duplicate this table for each Activity	**Activity No**	**1**
<u>Activity:</u>		

Development of drift kinetic referent model **Assignee:** Felix I. Parra Diaz		
Objective 1: Identify the theoretical work that needs to be done to obtain a complete drift kinetic referent model for the edge. **Objective 2:** Propose a drift kinetic referent model for helical magnetic field lines. **Objective 3:** Propose a suite of theoretical tests for the drift kinetic referent model.		
Key Deliverables:	**Start and Completion date:**	**Assignee:**
1) Report describing the important regimes in the tokamak plasma edge and a review of the state-of-the-art, including prioritisation of key physical processes for modelling and for further research.	**1)** Jan. 2021 – March 2021	**1)** Felix I. Parra Diaz
2) Report on 1D model along magnetic field lines with periodic boundary conditions and with wall boundary conditions	**2)** Jan. 2021 – May 2021	**2)** Felix I. Parra Diaz
3) Report on axisymmetric 2D model with wall boundary conditions	**3)** March 2021 – Sept. 2021	**3)** Felix I. Parra Diaz
4) Report on axisymmetric 2D model with periodic boundary conditions	**4)** Aug. 2021 – Nov. 2021	**4)** Felix. I. Parra Diaz
5) Report on axisymmetric 2D model with 'separatrix'	**5)** Nov. 2021 – Jan. 2022	**5)** Felix I. Parra Diaz
6) Report on the 3D electrostatic referent model for a helical magnetic field	**6)** Jan. 2022 – May 2022	**6)** Felix I. Parra Diaz
7) Report on extensions to the 3D electrostatic referent model for helical magnetic field	**7)** March 2022 – June 2022	**7)** Felix I. Parra Diaz
Milestones towards deliverables:	**Completion date:**	**Assignee:**
1) Report on 1D model along magnetic field lines with periodic boundary conditions	**1)** Jan. 2021	**1)** Felix I. Parra Diaz
2) Report describing the important regimes in the tokamak plasma edge and a review of the state-of-the-art, including prioritisation of key physical processes for modelling and for further research.	**2)** March 2021	**2)** Felix I. Parra Diaz
3) Report on wall boundary conditions	**3)** May 2021	**3)** Felix I. Parra Diaz

4) Report on 2D model with wall boundary conditions **5)** Report on fluid analytical solutions to use as benchmarks for the 2D model with wall boundary conditions **6)** Report on 2D model with periodic boundary conditions **7)** Report on 2D model with 'separatrix' **8)** Report on the 3D electrostatic referent model for helical magnetic field **9)** Modifications to the 3D electrostatic referent model for helical magnetic field suggested by proxy-app results **10)** Report on extensions to the 3D electrostatic referent model for helical magnetic field	**4)** July 2021 **5)** Sept. 2021 **6)** Nov. 2021 **7)** Jan. 2022 **8)** March 2022 **9)** May 2022 **10)** June 2022	**4)** Felix I. Parra Diaz **5)** Felix I. Parra Diaz **6)** Felix I. Parra Diaz **7)** Felix I. Parra Diaz **8)** Felix I. Parra Diaz **9)** Felix I. Parra Diaz **10)** Felix I. Parra Diaz

Duplicate this table for each Activity	**Activity No**	2
Activity: Implementation and testing of the drift kinetic referent model **Assignee:** Michael Barnes		
Objective 1: Compare the numerical performance of the novel, moment-based drift kinetic model to a more standard drift kinetic model for a variety of problems. **Objective 2:** Determine potential bottlenecks in the numerical treatment of the drift kinetic models **Objective 3:** Develop appropriate numerical algorithms for treating the wall boundary		
Key Deliverables: **1)** Proxy-app for 1D drift kinetic model **2)** Report on numerical issues and findings associated with the 1D drift kinetic models **3)** Proxy-app for 2D drift kinetic model **4)** Report on numerical issues and findings associated with the 2D drift kinetic model	**Start and Completion date:** **1)** Jan. 2021 **–** Sept. 2021 **2)** Jan. 2021 **–** Oct. 2021 **3)** Oct. 2021 **–** June 2022 **4)** Oct. 2021 **–** June 2022	**Assignee:** **1)** Michael Barnes **2)** Michael Barnes **3)** Michael Barnes **4)** Michael Barnes

5) Final report on the developed Proxy-apps	**5)** Oct. 2021 – June 2022	**5)** Michael Barnes
<u>Milestones towards deliverables:</u>	**<u>Completion date:</u>**	**<u>Assignee:</u>**
1) Report on numerical issues/findings for standard 1D drift kinetic code with periodic boundary conditions	**1)** Feb. 2021	**1)** Michael Barnes
2) Report on numerical issues/findings for simplest (only density) moment-based 1D model with periodic boundary conditions	**2)** April 2021	**2)** Michael Barnes
3) Report on numerical comparison between standard and moment-based 1D models	**3)** June 2021	**3)** Michael Barnes
4) Report on numerical issues/findings for standard 1D drift kinetic code with wall boundary conditions	**4)** Aug. 2021	**4)** Michael Barnes
5) Report on numerical issues/findings for moment-based 1D model with wall boundary conditions	**5)** Oct. 2021	**5)** Michael Barnes
6) Report on numerical issues/findings for standard 2D model with wall boundary conditions	**6)** Dec. 2021	**6)** Michael Barnes
7) Report on numerical issues/findings for moment-based 2D model with wall boundary conditions	**7)** Feb. 2022	**7)** Michael Barnes
8) Report on numerical issues/findings for 2D model with periodic boundary conditions	**8)** April 2022	**8)** Michael Barnes
9) Final report on numerical issues/findings from the developed Proxy-apps	**9)** June 2022	**9)** Michael Barnes

T/NA086/20
Code structure and coordination

Benefits and alignment to Work Package objectives

- **Alignment with other work packages**: This submission represents the core work in the Fusion Modelling System's fourth work package, *Code structure and coordination*. In particular this project is well aligned with the proxy applications being developed in the second and third work packages. Close coordination between these three groups will therefore be vital to the success of this project. Additionally, the output from the first work package, *numerical representation*, will feed into this project by influencing the design of data structures that remain performant at scale.

- **Benefits to other work packages**: The *code structure and coordination* work package will directly benefit the other NEPTUNE groups by highlighting approaches to software development that are appropriate for Exascale systems.

- **Interdisciplinary team with broad experience**: The researchers involved in this bid have a wide range of backgrounds in high performance computing and research towards exascale computing, in collaboration with industry, national and international universities, UK and US national labs: The lead (Dr Steven Wright) and Warwick Co-I (Dr Gihan Mudalige) have extensive experience of DSL development for modern scientific libraries, and application to HPC for performance portability. The team also includes the lead maintainers of BOUT++, a world-leading framework for plasma simulations which makes extensive use of performant abstractions to combine a DSL for easy implementation of equations, with high performance. By including domain experts in both fluid and kinetic plasma systems, research software engineers with experience of software sustainability, and performance portability experts, we can ensure that the recommendations arising from this work are suitable for NEPTUNE applications and go beyond the current state-of-the-art.

- **Close links to broader ExCALIBUR effort**: Through the Warwick Co-I (Dr Gihan Mudalige) this proposed work package will have direct links to ExCALIBUR working groups on code generation and high-level abstractions for exascale applications. This will ensure that the work here is aligned and complementary to the broader ExCALIBUR programme, benefiting from those working groups while focussing more specifically on NEPTUNE applications to tokamak plasma physics.

Scope

Key Deliverables and/or Desired Outcomes

The deliverables of the work package will be:

- **Survey of available technologies:** The first task will be to survey available and upcoming hardware that are likely to be present in exascale systems, and identify appropriate programming models that will allow full exploitation of these systems. Particular focus will be paid to performance portable programming models, such that NEPTUNE does not become limited to particular systems.

- **Identification of Representative Applications**: To evaluate the portability of the identified software approaches, representative algorithms and datasets will be identified for evaluation. In particular, proxy apps that implement the equations described in FMS0021 (Equations for NEPTUNE Proxyapps) will be identified or developed using the software frameworks described above such that the performance portability of the approaches can be evaluated. An example of such a representative application is system 2-5, to be implemented by the successful bidders to work package T/NA083/20.

- **Evaluation of Approaches to Performance Portability**: The applications identified previously, and/or those developed by the successful bidders of T/NA078/20 and T/NA083/20, will be evaluated on a range of hardware (e.g., Bede [V100+Power9], Viking [x86], Isambard [Arm], Archer2 [AMD]) to identify key performance indicators and to assess the performance portability of the above identified approaches.

- **Identification and Dissemination of Best Practices**: Key to this work package is working with the developers of the two proxyapps being developed by the related projects, T/NA078/20 and T/NA083/20, but also with other NEPTUNE work packages and ExCALIBUR groups to ensure any software developed for NEPTUNE is fit for purpose and resilient to shifts in hardware. Therefore this work package will involve analysis of performance evaluations using appropriate metrics to identify a set of best practices, and will focus on the dissemination of these best practices through documentation and training resources.

The specific tasks identified to achieve these deliverables are:

- **Task 1**: Survey of available technologies
  - **Activity 1.1**: Survey of available and upcoming hardware that is or is likely to be present in pre- and post-Exascale HPC systems.
  - **Activity 1.2**: Survey of domain specific languages, performance portable programming models and accelerated scientific libraries.
  - **Activity 1.3**: Survey of parallel file systems, I/O libraries and approaches to in-memory data management.
  - **Deliverable**: A report summarising the findings of activities 1.1-1.3, outlining the opportunities and potential risks of these technologies to NEPTUNE. The report will include specific details on achieved performance, as reported in the latest literature for the above identified programming models, DSLs and libraries. Wherever available, the measure of achieved performance and/or performance portability with these technologies will be detailed.

- **Task 2**: Identification of testbed platforms and applications
  - **Activity 2.1**: Identification of a broad range of proxy applications that have computationally similar algorithms to those in FMS0021 (Equations for NEPTUNE Proxyapps).
  - **Activity 2.2**: Identification of testbed platforms, seeking systems that cover the range of alternative technologies that are likely in exascale systems.
  - **Activity 2.3**: Acquire data sets that will give a representative view of application performance at scale on a variety of systems.
  - **Deliverable**: A repository of benchmarks (with a number of alternative implementations using various DSLs, performance portable programming models, etc) and data sets for assessing the various technologies identified in Task 1.

- **Task 3**: Evaluation of approaches to future-proofed applications
    - **Activity 3.1**: Set up testbed system environment, build dependencies, configure build systems etc.
    - **Activity 3.2**: Evaluate the performance of the previously identified proxy applications in a range of configurations and using a broad range of approaches to portability at scale.
    - **Activity 3.3**: Analyse the performance portability of the selected applications using appropriate metrics, [e.g. Pennycook et al. https://doi.org/10.1016/j.future.2017.08.007].
    - **Deliverable**: Report on the approaches analysed providing recommendations on best approaches for Exascale software development.

- **Task 4**: Dissemination of Best Practices
    - **Activity 4.1**: Coordination with other groups, in particular successful bidders of T/NA078/20 and T/NA083/20 to gather requirements, and relevant proxy applications.
    - **Activity 4.2**: Coordination throughout project with NEPTUNE groups to facilitate performance portable implementation of proxy applications.
    - **Deliverable**: Final report on best approaches to scientific software development for portable performance on post-exascale HPC systems.

Exclusions

- **Performance assessment of parallel file systems**: While this project seeks to evaluate a small number of parallel file systems and I/O libraries, access to test bed systems with alternative file systems is likely to be limited. While the assessment of parallel file systems will form some part of this project, it is likely to be limited in scope.
- **Performance of programmable hardware**: A number of the programming models and DSLs being investigated in this project can target programmable devices (i.e., FPGAs). However, again there is likely to be limited availability of these systems on which to evaluate performance. A best-effort approach will be taken to evaluate as many alternative platforms as possible.

Constraints

- This project would benefit directly from the outputs from other successful bidders. In particular, the outputs of T/N/A078/20 [Performance of spectral elements], T/NA079/20 [Optimal use of particles], and T/NA083/20 [Plasma fluid referent model] will provide much better proxies of performance than other open-source alternatives. We will engage with these groups to encourage performance portable development, and also to ensure our own evaluations are as representative as possible, by using their developed applications where possible.
- During the project, we should have access to a range of HPC platforms, allowing a wide range of hardware options to be assessed. However, full coverage of likely post-Exascale systems may not be possible -- to mitigate this, we will leverage access to systems at Warwick, York, the N8 and EPCC (if available) to ensure as broad a spectrum of systems as possible are evaluated.

Approach

The key determinant in understanding the limitations and opportunities of the current UKAEA/NEPTUNE codebase when targeting exascale systems is through the identification of algorithms of interest within this workload and ascertaining their performance on current and emerging HPC architectures. As such the project will use a repository of proxy applications, developed as part of NEPTUNE or taken from publicly available benchmark suites, to evaluate approaches to performance portable software development. These evaluations will be conducted on the HPC systems available at the University of York (Viking), University of Warwick (Avon, Orac) and within the N8 (Bede), as well as through collaborations with the ExCALIBUR groups at EPCC (Archer2) and the University of Bristol (Isambard).

Our performance evaluations will be based on a series of “ground rules” for assessing performance portability in a fair way. This will take the form of the metrics outlined by Pennycook et al. [https://doi.org/10.1016/j.future.2017.08.007].

To ensure the evaluation is relevant to the NEPTUNE project, we will use our domain knowledge of the plasma physics systems for NEPTUNE as outlined in the call document, and work with other successful bidders to use the proxy applications that they are developing (or have been developed) as the basis of our evaluation. If these proxy applications are not yet ready for performance portable evaluation, we will identify key kernels in existing open source applications which are representative of NEPTUNE systems of equations.

Each task outlined above builds incrementally on the previous tasks and activities, and where applications are not available, open source alternatives will be found such that we can evaluate as many options of post-exascale software development as possible. While each task culminates in a deliverable report, these will be living documents -- continually being updated throughout the project as new hardware emerges and as support for software evolves.

The work in this proposal will take direction from and inform the direction of several of the NEPTUNE work packages, and so coordination between projects will be vitally important to its success and is included as Task 4 (Dissemination of Best Practices) which will run throughout the project.

As travel may be restricted throughout the project, due to the COVID-19 pandemic, many of the coordination activities will necessarily have to take place virtually. However, should travel restrictions be lifted, we would hope to attend relevant conferences, such as ISC High Performance 2021. To support this, we request a travel and subsistence budget of [redacted]. Additionally, should in-person meetings become possible during the project, we request an additional [redacted] for UK travel and subsistence.

Relevant Experience

A brief summary of the relevant expertise and track record of the people involved in this bid is given below. Further information can be found in the CVs submitted with this bid.

- **Dr Ben Dudson** is a Reader in the Department of Physics at the University of York, specialising in tokamak boundary plasma physics. Lead designer and author of BOUT++ (2007-present), a high performance, object-oriented C++ framework used worldwide as a basis for plasma simulations. 72 publications in experimental, theoretical and computational plasma physics, with an h-index of 29 (Google Scholar). Chair of the EPSRC Software Outlook Working Group, and deputy director of the York Plasma Institute, with extensive experience of mentoring students, postdocs and junior staff. Funding application and project management lead, as PI and Co-I, with a record of sustained funding income from a range of funding agencies, in particular UK EPSRC and EU EuroFusion (Horizon 2020). He leads a team at the University of York working on plasma edge and divertor modelling for STEP, under contract with UKAEA.

- **Dr Steven Wright** is a Lecturer in Computer Science at the University of York. He has worked in High Performance Computing (HPC) for 10 years, particularly in the areas of performance analysis, benchmarking and optimisation of HPC systems. His Ph.D. was completed at the University of Warwick in collaboration with collaborators at Lawrence Livermore National Laboratory and Los Alamos National Laboratory, analysing and optimising the performance of parallel file systems and I/O in HPC applications. Before joining the University of York, he was a postdoctoral research fellow at the University of Warwick working with UK AWE on the Computational Collaborative Project in Plasma Physics [EP/M011534/1], assessing the sustainability and adaptability of the Odin Arbitrary Lagrangian-Eulerian (ALE) code. Following a secondment as a visiting research fellow at Sandia National Laboratories (NM), he has been collaborating on the ElectroMagnetic Plasma in Realistic Environments (EMPIRE) project, focussed on the development of a performance portable unstructured FEM-PIC code.
- **Dr Peter Hill** is a research software engineer (RSE) in plasma physics and currently provides support to several computational projects at the York Plasma Institute (YPI) totalling more than £6M, including as Researcher Co-I on the Turbulent Dynamics of Tokamak Plasmas (TDoTP) EPSRC Programme grant (EP/R034737/1). His current interests are in software sustainability and reproducible research, both of which are vital to maintaining trust in research. Peter is a lead maintainer on two world-class plasma codes, BOUT++ and GS2, and regularly contributes to many other software projects, including external open-source projects such as PETSc, netCDF and CMake, collectively used in millions of projects.
- **Dr David Dickinson** is a Lecturer in the Department of Physics at the University of York with a primary physics interest in core plasma turbulence and associated instabilities. He has a research focus of first-principles computational analysis of gyrokinetic/gyrofluid instabilities and brings experience with using and developing world-class tools for core, pedestal and SOL gyrokinetic/gyrofluid simulation. He has been a leading developer and maintainer for the gyrokinetic code GS2 since 2012 and has also made substantial contributions to BOUT++ since 2016. He is Co-I on the Turbulent Dynamics of Tokamak Plasmas (TDoTP) EPSRC Programme grant (EP/R034737/1) and leads work at the University of York for core turbulence modelling for STEP, under contract with UKAEA.
- **Dr Gihan Mudalige** is an Associate Professor in the Department of Computer Science at the University of Warwick. His research focuses on the development of next-generation High Performance Computing numerical simulation software libraries through the utilization of domain– specific high-level abstraction frameworks. In 2018, he was awarded a four-year Royal Society Industry Fellowship with Rolls-Royce (INF/R1/180012), focusing on developing future-ready massively-parallel CFD simulations for Exascale HPC systems. He is also a co-investigator in the £6.8M EPSRC Prosperity Partnership (EP/S005072/1) for Advanced Simulation and Modelling of Virtual Systems (ASiMoV) aiming to realise new simulation technologies at ultrahigh resolution and extreme scales. Dr. Mudalige is a CoI on two of the recently awarded ExCALIBUR working groups, both on applying code-generation

and high-level abstractions for exascale software. The research software developed by Dr. Mudalige and his team is now utilized in projects carried out by a range of academic and commercial organizations, including EPSRC (EP/V00140X/1), Rolls-Royce plc., STFC, AWE plc., the Numerical Algorithms Group (NAG), UCL, ATI London, ETH Zurich and the Universities of Nottingham and Southampton.