

# PACE Project board – Meeting 4

## Minutes

---

Date: 11/12/2020

**Present:** Chair: Russell Ewings (RE), Toby Perring (TP), Christopher Marooney (CM), Ray Osborn (RO), Jon Taylor (JT), Steve Hayden (SH), Lamar Moore (LM), Philip King (PK), Lesley Mansfield (LAM)

**Apologies:** Gordon Brown

### Agenda:

No.	Time	Item	Speaker / paper
1	14:30	<b>Welcome, Introductions</b>	Russell Ewings
2	14:35 – 14:45	<b>Minutes of last meeting &amp; actions</b>	All
3	14:45 – 15:05	<b>Progress report</b>	Chris Marooney, Toby Perring
4	15:05 – 15:35	<b>Discussion</b>	Russell Ewings
5	15:35 – 15:50	<b>Financial summary</b>	Toby Perring
6	15:50 – 16:00	<b>Risk register update</b>	Toby Perring
7	16:00 – 16:15	<b>Communication Plan</b>	Toby Perring
8	16:15 – 16:30	<b>Any other business</b>	All

### 1. Minutes of last meeting and actions

- There were no comments on the minutes from meeting 3.
- 1.5.1 and 1.5.2 Not started. Tessella contractors are now in place and established. They now have an understanding of the code base and the technical specifications. This will be discussed later in this meeting.
- 3.1.1 - LAM spoke to LM about this after the last PMB. LM was reluctant to share this documentation without first talking to GB.
- 3.2.1 This is a complex issue and probably will not be resolved for a while. Need to look at least restrictive option, GPL dependencies and commercial impact. It was expressed that we should follow BSD licensing. Not to do so will cause problems for US labs (it has been raised as an issue in connection with Mantid, for example). This action can be closed here but needs to be followed up in the technical meetings.
- 3.2.2 Completed.
- 3.2.3 Completed. Communication Plan to be discussed later in this meeting.
- 3.2.4 Completed. Milestones to be discussed later in this meeting.
- 3.2.5 Completed. MVP to be discussed later in this meeting.

- 3.2.6 Completed. Dependencies to be discussed later in this meeting.
- 3.3.1 Completed. This risk has been added to the Register.
- 3.3.2 Completed. TP did seek advice on the command line, but found it was not helpful as opinion is divided. The Python/MATLAB interface works, so this may be unnecessary?
- 3.3.3 Completed. Communication Plan to be discussed later in this meeting.
- 3.4.1 Closed. This is no longer an issue. Greg Tucker is leaving and we will be recruiting to replace him.
- 3.4.2 Closed. Corona Virus impact assessment no longer required.
- 3.5.1 Completed. Communication Plan to be discussed later in this meeting
- 3.5.2 Completed.
- 3.5.3 Completed. TP has been in contact.

## 2. Progress report

- CM gave a presentation summarising the progress of PACE to date, and planned objectives
- PACE Minimum Viable Product:
  - Interface to third party simulation codes
    - o SpinW (now a separate project)
    - o Phonon calculations (Euphonic)
    - o API for generic user (MATLAB, Python and C++), and 3rd party codes
  - Optimisation and resolution convolution algorithms
    - o Parallel Tobyfit
    - o New approaches
  - Performance and usability framework for
    - o Parallel and distributed computing architecture: IDaaS, SCARF, but will also work on high end laptop/desktop
    - o MATLAB and Python user interface
    - o Handling large datasets out of memory
    - o Generic projections
- SpinW developer to start in December. TP added that there is to be a joint SpinW project with ESS. Coordinated plans are upcoming.
- Euphonic tasks completed since March 2020
  - Major refactor of Euphonic's API completed (more user friendly and maintainable, easy handling of units, can output any Euphonic object as .json file)
  - Addition of spherical averaging tools for powder averaging (contributed by Abins developer Adam Jackson)
  - More robust and user-friendly command-line tools e.g. dispersion.py is now euphonic-dispersion
  - Has been successfully validated against other computational codes
  - Graduate effort (James King) – 3 month project improving testing
  - Migrating to Pytest
  - Adding Windows/Mac continuous integration nodes
- Euphonic Tasks for the upcoming year:
  - Work towards publication on Euphonic
  - Alongside publication aiming for:
    - o 1.0.0 release of Euphonic

- Release version of Euphonic-Horace-Interface
- Benchmark Euphonic performance against itself, CASTEP phonon tools, other codes etc: much has been done but it needs to be formalized
- Allow use of brille from Euphonic
- New integration repository for Horace-Euphonic interaction – Basis for testing, regressions
- RE: How do you get the powder averaging information?
- TP: Powder averaging can be done in Euphonic. Adam fully involved in the project, and Euphonic is a core library for Abins.
- SH: In the report it mentions the interface between CASTEP and phonopy – are there only these two interfaces?
- TP: Euphonic via phonopy should cover all cases identified as used by ISIS Excitations Group and MoSpec group apart from one.
- SH: Is ELK covered?
- TP: I can't guarantee – but pretty certain it is. Need to confirm with Adam
- JT: Phonopy covers most of Ab initio code [*checks web page; includes ELK*]
- SH: CASTEP is not good for metals
- TP: We are aware of this and that PACE will have to evolve.
- JT: Have you considered pushing Euphonic into phonopy?
- TP: No. Euphonic is a separate package from Horace. We are making sure everything works within PACE before we start to consider pushing elsewhere.
- JT: Might be useful to consider in the future?
- RE: Are phonopy developers aware of Euphonic?
- TP: Yes. Becky has been in contact and Adam knows the developers.
- Brille updates:
  - Refactoring, handling improvements:
    - Proper handling of symmetry for interpolated vectors, matrices, and phonon eigenvectors
    - Equivalent mode caching
    - Some consequent performance impact offset by shared memory arrays
  - Usability:
    - C++ library namespace encapsulation
    - Automated Python module build and publish pipeline
    - Automated hybrid Sphinx and Doxygen documentation
  - Example of use with AI project using SpinW
    - Training time for neural nets reduced by order of magnitude to feasible size: now 7000 CPU-hours
  - Staffing:
    - Greg Tucker (PDRA) leaving for ESS Lund, will continue to work on brille at a reduced level
    - Process of formalising this ongoing
- RO: Can't find documentation on brille.
- TP: There is a program of work to ensure that the code is sorted out, but documentation is currently being updated.
- RE: Documentation needs to be findable
- The documentation needs to be made publicly available. Make good documentation a parallel strand of work over the next 5 months in time for the Excitations data analysis workshop at the end of May 2021.
- **ACTION 4.2.1 : CM/TP to add Documentation task to Project Plan**

- Tobyfit refactoring preparation:
  - MATLAB instrument component and detector classes
    - o Rewritten and incorporated into a refactored Tobyfit.
    - o Now part of new SQW object redesign (see framework, next slide)
  - Parameter optimisation application
    - o Design document finalised.
    - o Generalises the fit functions, allows for different swappable core fitting engines.
    - o Allows construction of more complex fit function
    - o Can be implemented once the core SQW object rewrite is completed.  
The legacy Tobyfit version has been removed
- RE: Have you considered changing the name? Tobyfit name does not obviously state what the function is.
- TP: Possibly. C good time to change the name is when PACE is fully developed into a collection for the first PACE distribution (as opposed to Horace+friends).
- RO: Is Tobyfit standalone?
- TP: No, not quite sure how easy that would be to manage.
- Framework tasks completed:
  - Support for large datasets (not fitting in memory) – implementation in progress
  - MPI framework developed to support the parallelisation of a number of operations to significantly improve performance
  - Compiled MATLAB now available
  - Continuous Integration (CI) set up using ANVIL service - Captures cross-platform build and test functionality previously missing or executed by hand
  - Design for major update of the SQW object complete – implementation in progress
- RE: Are there any resource requirements for build server hardware?
- CM: Periodically we have outages, but these are due to issues rather than demand.
- TP: There is only one person looking after anvil, which is concerning. This needs resolving but will have to stick with this at the moment as there is no internal (ISIS) solution.
- JT: Disk.io - was this developed against parallel file system or generic hardware?
- TP: Generic. Takes advantage of parallel file store but not necessary.
- JT: Is it optimised for pfs?
- TP: This has not been considered. Default hardware is in a dynamic state. Using STFC Cloud (IDAaaS). There are 5000 CPUs across the nodes. Not sure what the requirements will be in 18 months.
- JT: This needs to be considered.
- TP: We are trying to keep things generic. Our needs have changed since the last PMB meeting – especially with remote working now the default.
- SQW:
  - RO: How is this serialised? How do you save the objects?
  - CM: Written in binary format. This is in development.
  - TP: SQW files can be large. Temporary files are created as independent jobs, then combined. There are multiple stages of read/write. Not sure how far developed this is. This is being replace by HDF5.
  - RO: HDF5 performs better than anything MATLAB produces. John Readey at HDF5 is a useful person to talk to about this. (email provided: jreadey@hdfgroup.org)
- Project Plan:
  - JT: When is the MVP release planned for?

- CM: probably around June 2021
- JT: What is MVP:
- CM: Parallelisation, Large Dataset support, not-to-byfit rework, Parameter optimisation, Interfaces with SpinW and Euphonic, Generic versions of these.
- JT: How will this be deployed?
- CM: We have a Release and Deployment pipeline. Will be on GitHub.
- JT: Will it be on IDAaaS?
- TP: Horace is on IDAaaS, as is mantid, and PACE will be too as part of the bespoke VM for Excitations instrument users.
- JT: Need to think about how it will be deployed to users.
- SH: Is there MVP documentation? Is there a milestone for this?
- CM. Yes, it will be completed with the release.
- SH: Will there be interaction with the user community to let them know?
- TP: Possibly as part of the PACE workshops. We can confirm. Currently there are WIKI pages for Horace.
- PK: Are there key users that you want involved?
- TP: Yes, certain people have been considered.
- RE: MVP must have supporting documentation
- SH: Users need to know when product and documentation are going to be released. This needs a milestone. Feedback is very important, but users need warning so that they are lined up to participate. They can also factor this into their research programmes, for example by making student effort available to test and help develop
- Consider the ways of making PACE citeable.
  - Papers are fine (Euphonic, brille), but what about PACE as a whole: a citable DOI ?
  - What is STFC policy on licensing?
- **ACTION 4.2.2: CM to add a milestone to engage with users to inform them and prepare them for the release.**
- **ACTION 4.2.3: CM to add milestone for MVP release.**
- JT: Need to look at DOIs for PACE, and MVP so that it can be referenced as a unique object.
- **ACTION 4.2.4: TP to follow up on DOIs**

### 3. Financial Summary

- TP provided the Financial Summary update, including the recent extension of the Tessella contract to February 2021
- RE: Have you included Becky's recent promotion in the figures?
- TP: This is an addition of approximately £16k, so not a huge change.
- RE: What about extra funding from ISIS?
- TP: there has been a capital underspend across ISIS, which enabled us to extend the Tessella contract.
- JT: What about the Euphonic secondment?
- TP: Becky is in SCD, as are Chris and Jacob. We already have a close relationship with SCD and we are currently in the process of coordinating closer engagement across various projects, not just PACE.

### 4. Risk register update

- Two risks have been closed, as they are no longer considered risks to the project:
  - PACE fixed term staff departure (the staff member has left, and we will be recruiting a replacement on an open-ended contract)

- CASTEP dependency
- One new risk has been added:
  - Inability to use product because of poor GUI
- TP: Our main concern now is HPC on-demand availability and implications of the hardware architecture

## 5. Communication Plan update

- RO: What about the user community?
- TP: There have been discussions to engage them through the focus groups that will take place in early 2021.
- RE: What about SNS and ILL?
- **ACTION: 4.5.1 - TP/LAM to add User Engagement with SNS and ILL**
- RE: It will be interesting to see how the workshops go, and if we do need to engage a User Experience person.

## 6. AOB

- PK: think the Project Management has been done really well – I like how the slippages and the milestones have been shown
- JT: Very impressive; looks like the project is progressing well. What about BREXIT? And European funding?
- PK: About to apply for a new round of funding (closing date is 26<sup>th</sup> January 2021). Beyond this is unknown.
- Hardware agnosticism:
  - Can we abstract the MPI framework so that specific filestore (CEPH, Panasus, Spectrum Scale) and cluster (STFC Cloud, other facility HPC, AWS etc.) implementations have specifics isolated in as thin a layer as possible.
  - What efficiency compromises will this entail?
  - What about using hdf5 object store?
  - These points are really the subjects for a technical board to advise about. Potential members include Jon Taylor (ESS), Ray Osborn (Argonne), Stuart Campbell (Brookhaven), Andy Richards (Diamond), Lamar Moore, SCD CEPH and SCARF teams (what is the hardware roadmap in SCD?), whoever will install PACE at SNS and ESS to come together to talk about these issues.
- JT: Science side seems well controlled as is the software development, but the architecture needs thinking about. Do you have advice or an area to test your architecture?
- TP: I think a long discussion is needed – nothing too formal
- JT: Needs input to your hardware providers
- RE: Yes, need a long term plan from SCD as technologies change
- JT: Stuart Gamble might be a good contact
- RO: Also need to consider who will be happy installing prototype elsewhere e.g. SNS
- TP: can set up a three hour slot to explore the technical aspects
- **ACTION 4.6.1: TP to set a technical discussion**
- Where do we stand on future European initiatives e.g., Digital Twins, Horizon Europe?
- Next PMB meeting will be in 6 months

## Ongoing Actions:

Identifier	Action	Owner	Status
1.5.1	Determine composition of technical review panel	TP	Ongoing
1.5.2	Schedule meetings of technical review panel and project board after the project specifications have been completed	TP	Ongoing
3.1.1	LM to circulate the IRIS bid documentation	LM	Ongoing
4.2.1	CM/TP to add Documentation task to Project Plan	CM/TP	New
4.2.2	CM to add a milestone to engage with users to inform them and prepare them for the release.	CM	New
4.2.3	CM to add milestone for MVP release	CM	New
4.2.4	TP to follow up on DOIs	TP	New
4.5.1	TP/LAM to add User Engagement with SNS and ILL	TP/LAM	New
4.6.1	TP to set up a technical discussion	TP	New

## Completed Actions:

Identifier	Action	Owner
2.1.1	Add lack of storage capacity on STFC cloud to risk register	TP
2.2.1	Contact Hartree centre for advice on distributed framework	TP
2.3.1	TP to contact JT and/or RO to discuss distributed data libraries or frameworks which PACE could use.	TP
2.5.1	TP to talk to G. Brown about mitigating PACE staffing risks	TP
2.5.2	LM to give panel details of the IRIS bid of SCD for increasing the capacity of the DAaaS system.	LM
2.6.1	TP to amend Project Brief to state that running DFT calculations is outside PACE's scope	TP
2.6.3	TP to ask A. Butts to talk to LM about parallel NXS file I/O	TP
3.2.1	LM, TP and Greg Tucker to discuss licence issue for Brille	TP
3.2.2	TP to add QMC and MD to consideration of the generic 3 <sup>rd</sup> party code interface	TP
3.2.3	TP and LAM to look at Communication Plan to ensure that User Engagement over the ISIS shutdown period has been considered.	LAM
3.2.4	LAM to look at providing a milestone chart (or converting current excel spreadsheet to a Project file)	LAM
3.2.5	TP and LAM to discuss intermediate user releases or MVP	LAM
3.2.6	LAM to look at task dependencies (with TP input)	LAM
3.3.1	TP/LAM to add Risk: non-update of product due to inability to use. Or regarded as no benefit or value	TP
3.3.2	TP to seek advice on command line use	TP
3.3.3	TP/LAM to add User Engagement for GUI to Communication Plan	LAM
3.4.1	TP to look at provide a list of critical resources	TP
3.4.2	TP to look at an action plan for the impact of Corona virus on the team	TP

<b>3.5.1</b>	TP/LAM to review Communication plan – add workshop, and non-ISIS User Engagement	LAM
<b>3.5.2</b>	RE to review Benefits Realisation of Business Case	RE
<b>3.5.3</b>	TP to contact Olivier Delaire, Kate Ross and Martin Mourigal regarding SNS user community engagement	TP