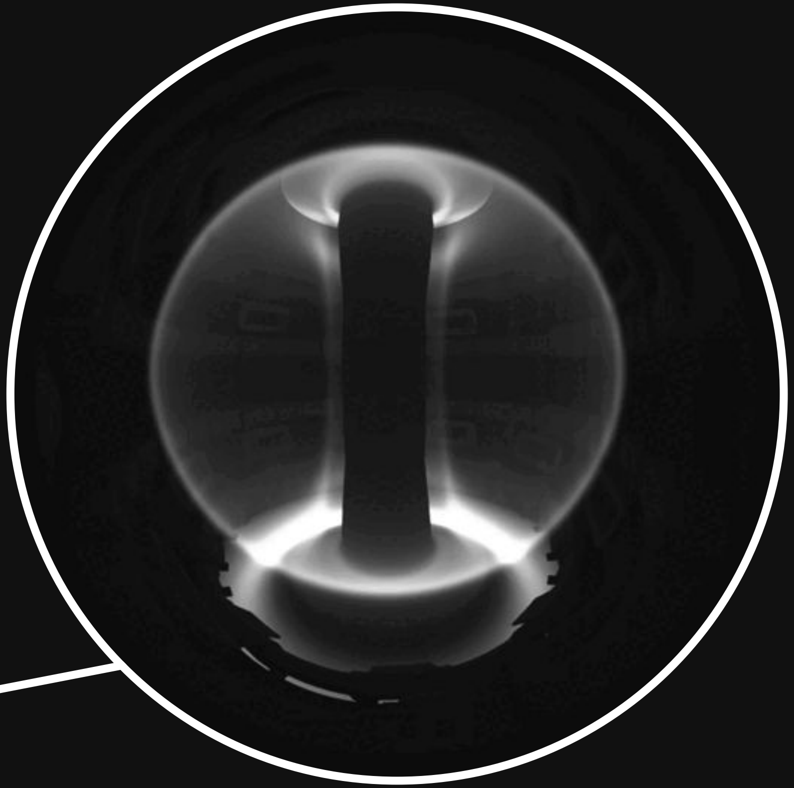
High Performance Multiphysics driven design for fusion systems

Prof. Andrew Davis, Head of Advanced Engineering Simulation,
UKAEA
4th Fusion HPC Meeting BSC



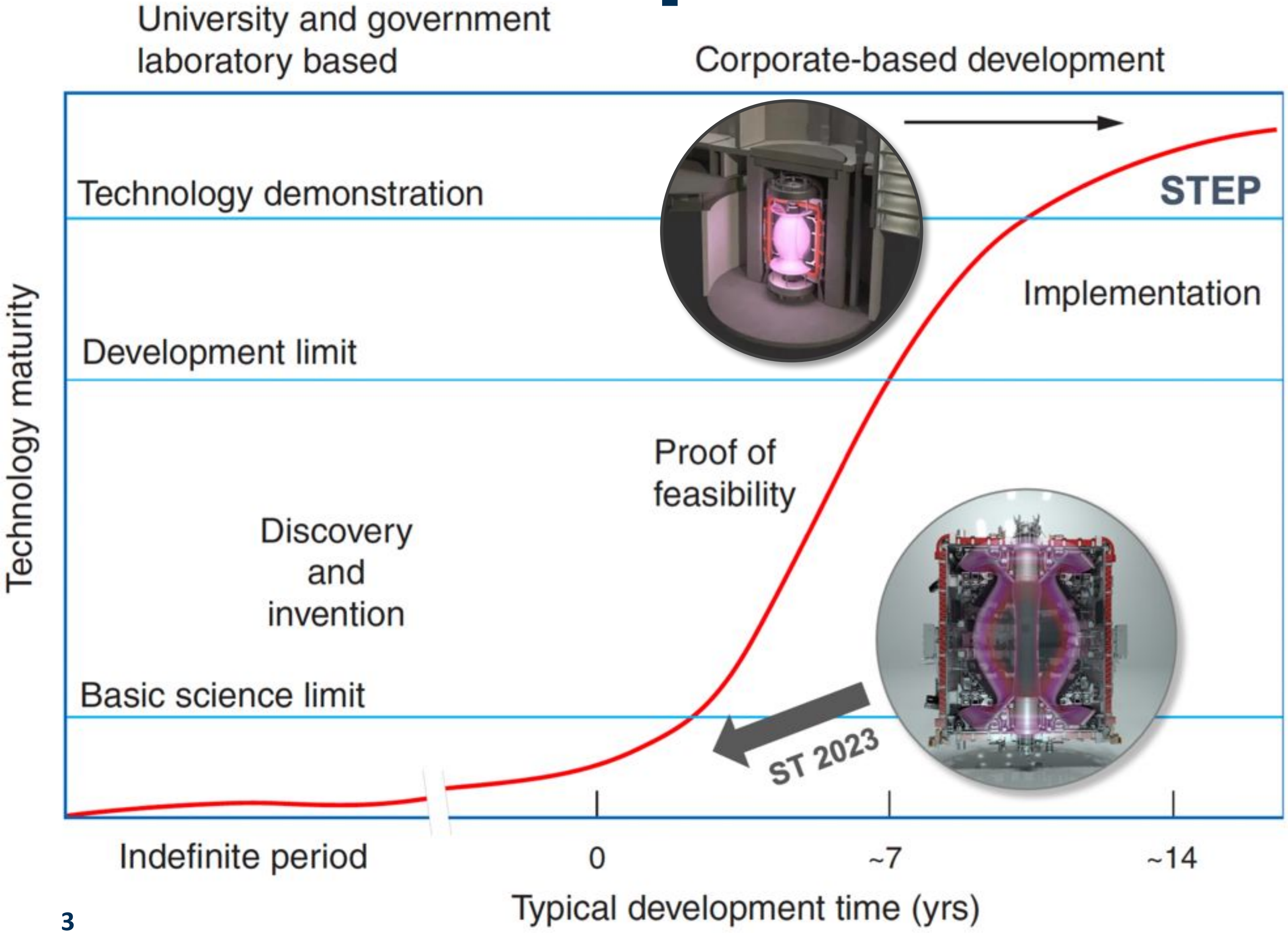
MAST-Upgrade (UKAEA, Oxfordshire, UK)
Radius: 1.5m
Temperature: 15,000,000 degrees Celsius
Power output: zero (negative)

Sun
Radius: 696,000 km
Temperature: 15,000,000 degrees Celsius
Power output: 385 Million Exawatts
(0.385 Octillion Watts)

ADVANCED ENGINEERING
SIMULATION

Kassel Lab

Challenge: catapulting fusion up the "S-curve" to help deliver Net Zero



CODAS & IT
Service / process delivery

TRL Valley of Death
Gaining mastery over systems / product (e.g., HPC & AI for engineering design or fusion technology itself)

Advanced Computing
R&D in HPC, AI & Big data

CD COMPUTING DIVISION

Complete Physical Testing is too expensive

UK Atomic Energy Authority

Fusion devices (like fission plants) are capital intensive

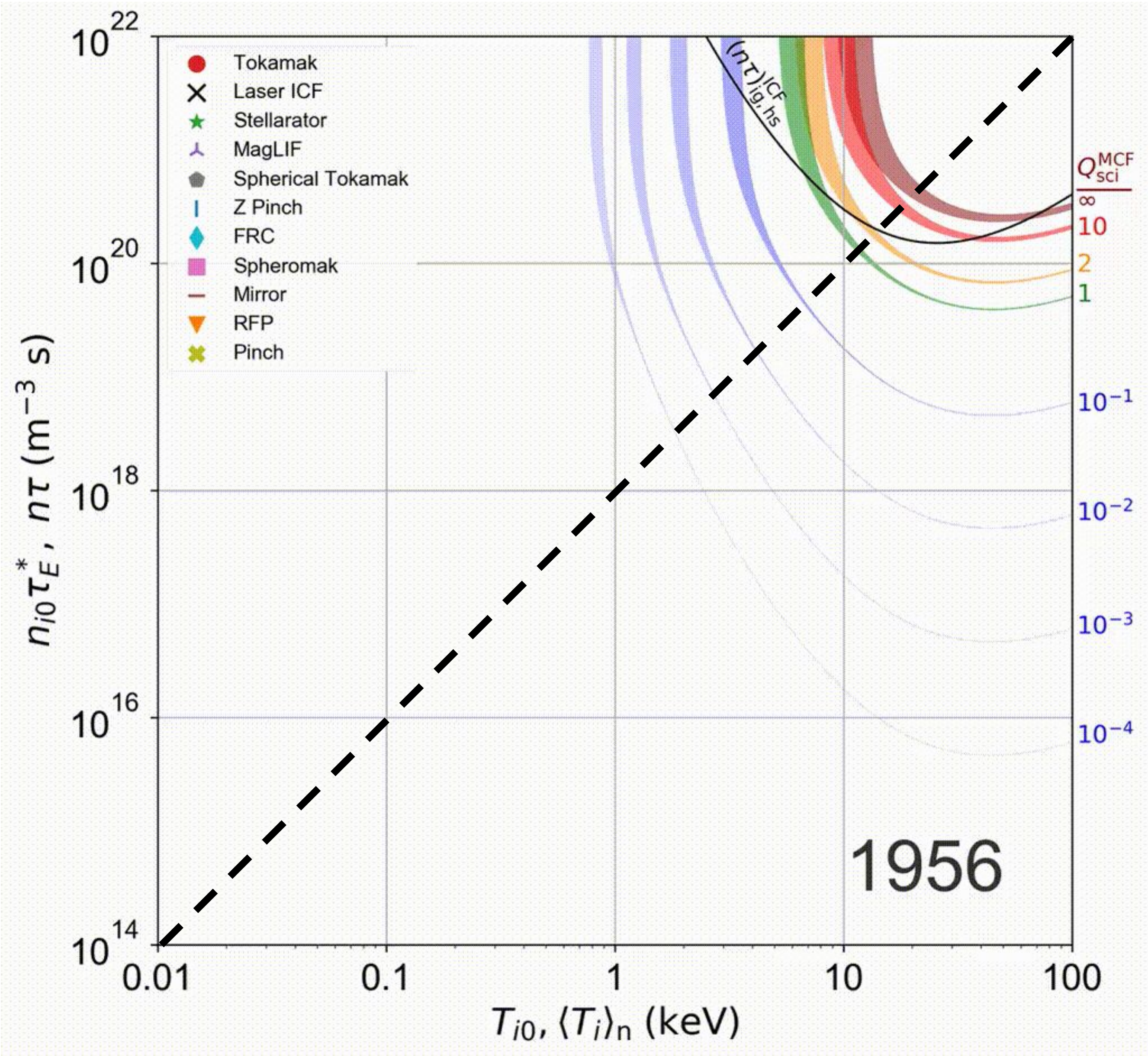
- ITER cost $22B (FOAK)
- Fission plant $6B
- SMR $2B

Much of the cost is insurance/concrete; magnets are expensive, steel is expensive, many fusion materials have no large existing supply chains

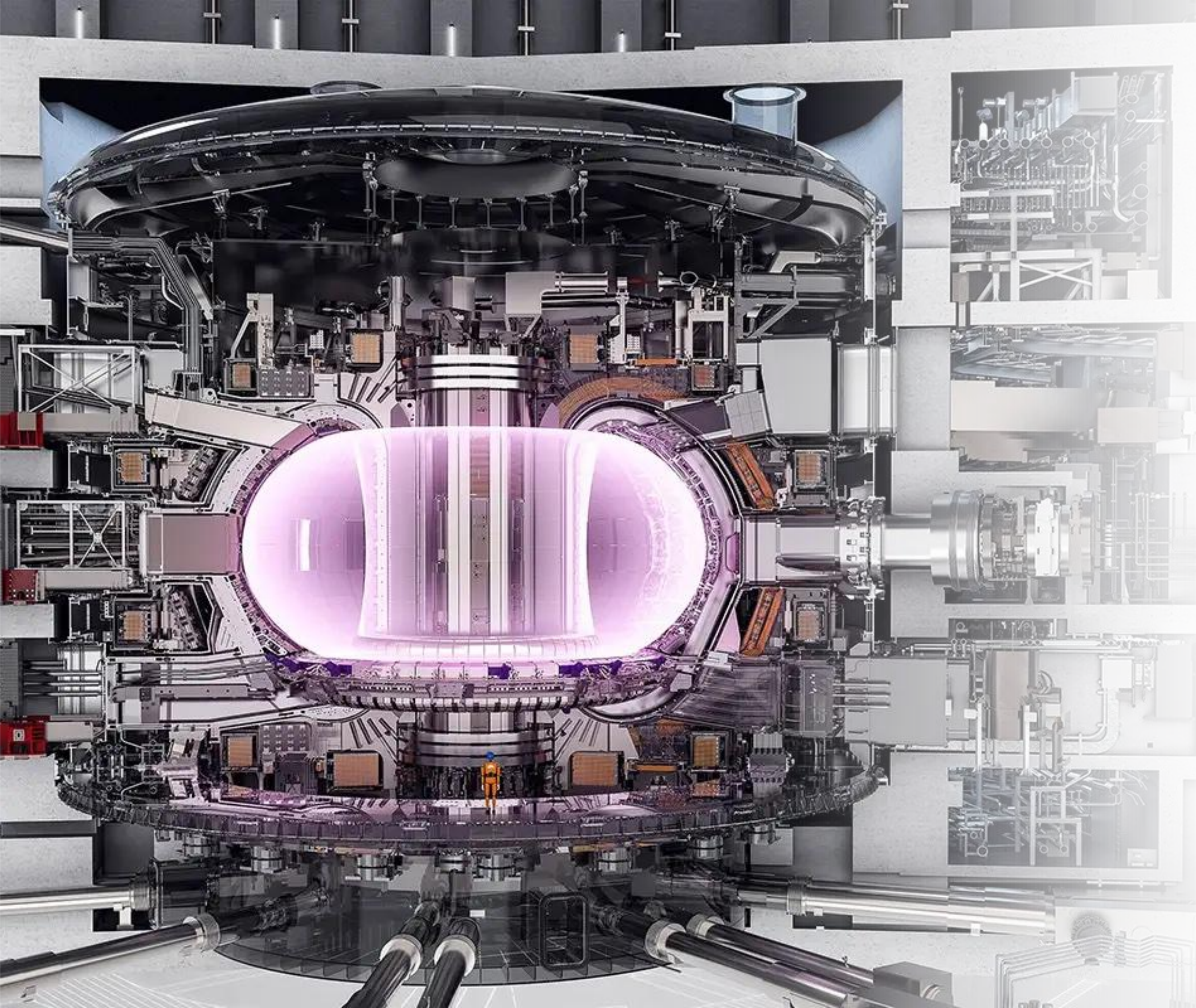
Clearly, design by iteration eventually becomes too expensive at some point for tokamaks (device) cost scales as B^2R^3.

As we approach Q ~ 1 - $$$$$$

Realistic fusion environment difficult to achieve artificially, high radiation, high temperature, high magnetic field



Sam Wurzel - https://twitter.com/swurzel/status/1534556521744457731



Simulation:

Everything, everywhere, all at once…

- **Radiation**
- **Electromagnetism**
- **Structural forces and Gravity**
- **Heat transport** (conductivity, thermal hydraulics)
- **Chemical transport** (Diffusion, radio-nuclide transport through fluids)
- **Temporal coupling –** transmutation, radiation induced embrittlement, fatigue, tribology/wear

Experiments *aren't* going to ride to the rescue?

UK Atomic Energy Authority

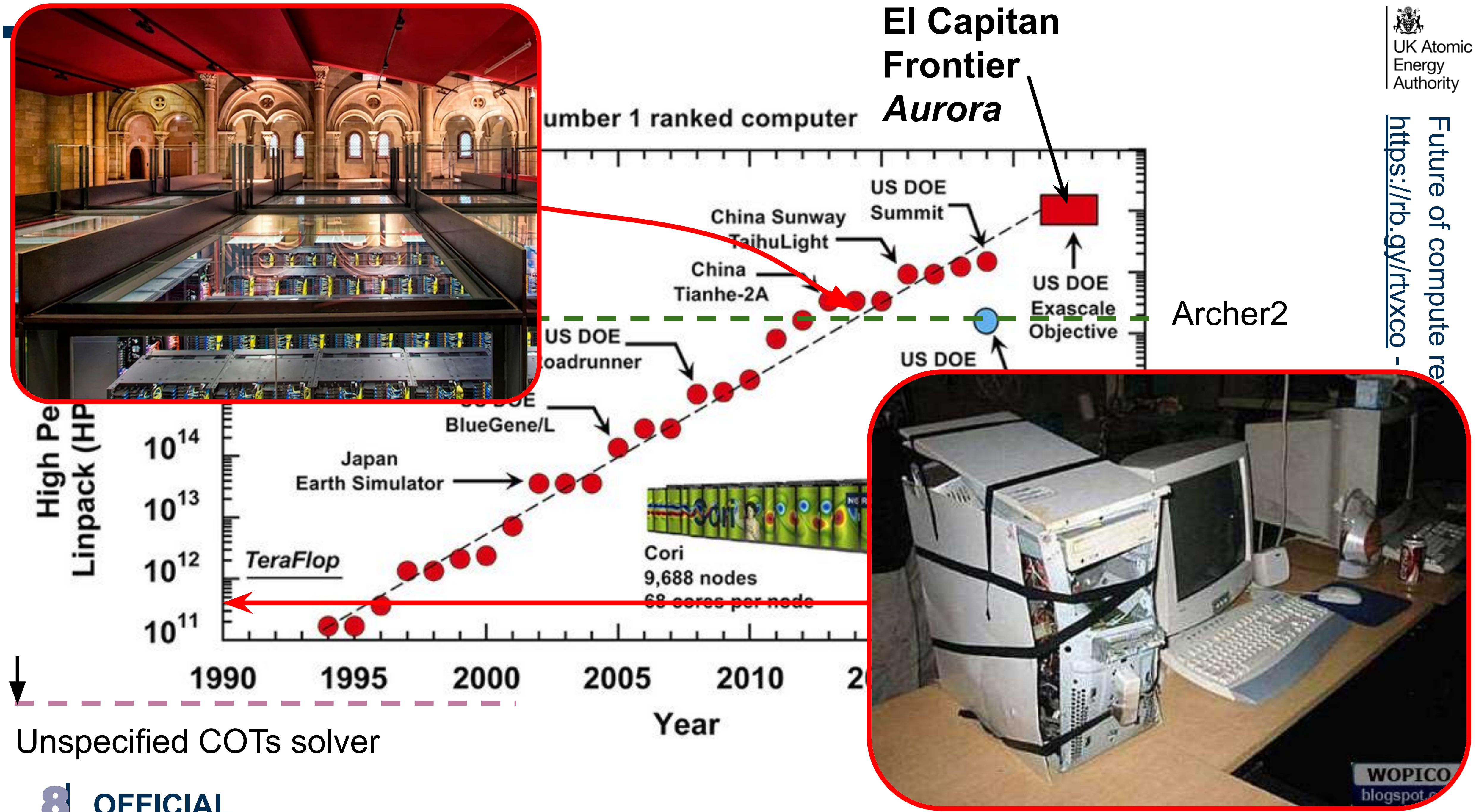
- Traditionally we used a design-by-failure approach
  - tokamaks are getting too **expensive** for the trend to continue
  - there will **not** be enough of the **right** experiments to fully de-risk tokamaks pre-2050
- The normal regulatory approach is demonstrate safety through margins and design codes; both are determined either through experiment or through long and numerous demonstrations of robustness
- The only way to deliver a demonstration tokamak that has been de-risked relative to today is to replicate all physics ***in silico***. This will require significant effort to advance state of the art simulation to the point where we can simulate a complete tokamak (i.e. not 'just' the plasma)

“

I do not fear computers, I fear the lack of them

”

Isaac Asimov


El Capitan
Frontier
Aurora
umber 1 ranked computer
US DOE
Summit
China Sunway
TaihuLight
China
Tianhe-2A
US DOE
Exascale
Objective
Archer2
US DOE
oadrunner
US DOE
BlueGene/L
Japan
Earth Simulator
Cori
9,688 nodes
TeraFlop
High Pe
Linpack (HP
10^{14}
10^{13}
10^{12}
10^{11}
1990
1995
2000
2005
2010
Year
Unspecified COTs solver
WOPICO
blogspot
UK Atomic Energy Authority
Future of compute re
https://rb.gy/rtvxco -

There are (at least two) Exascales

- The first is the exascale that Frontier achieved, with an R_{max} of 1.194 EFlop/s (out of a possible R_{peak} 1.679 EFlop/s.
  - Well done ORNL :)
  - This is great for non-memory bound problems, e.g. monte carlo which should be able to scale a long way
  - LLMs are going to have a fun time
- The second is the exascale that Frontier didn't achieve with an HPCG result of 0.014 Eflops, i.e. 1 Linkpack Flop = 0.0083 HPCG flops
  - HPCG is a better reflection of how real world distributed memory bound problems behave, like those we have in engineering behave
- Current trends indicate that 1 HPCG GPU flop = 3 HPCG CPU flops (HBM2)
  - What does that mean for us?
  - High Productivity Computing

I have anxiety around the AI trend

- I have anxiety around the AI trend, but not for the reasons that you might think
  - Not a luddite
  - Not particularly conservative
  - Not worried that AI is going to steal my job
- GPUs rose to prominence in HPC, not because of HPC and their end use case but due to the success of the home gaming market, they happened to be accidently useful for some applications
- Now we see the rise of AI applications, and push for reduced precision within those AI specific operations from 64->32->16->8 bit
  - If HPC becomes driven by the AI trend, and the GPU market responds, is it plausible to imagine reduced precision GPUs, with fewer CPUs per node, and reduced precision where will I run my code?!

A brief diversion....

UK Atomic Energy Authority

TOP500 Linpack

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	1,679.82	22,703
2	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	4,742,808	585.34	1,059.33	24,687
3	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Microsoft Azure United States	1,123,200	561.20	846.84	
4	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
5	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107

Top500 HPCG

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	HPCG (TFlop/s)
1	4	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	16004.50
2	1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	14054.00
3	5	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	4586.95
4	6	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN EuroHPC/CINECA Italy	1,824,768	238.70	3113.94
5	7	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	2925.75

Look at the ratio of HPCG/Linpack

UK Atomic Energy Authority

HOME › HPC › Is This The End Of The Line For NEC Vector Supercomputers?

IS THIS THE END OF THE LINE FOR NEC VECTOR SUPERCOMPUTERS?

March 23, 2023 Timothy Prickett Morgan

Matsuoka said in a tweet. "This could be the final nail in the coffin for their SX vector processors history. At 6.8 TB/sec mem BW target, MI300 would have buried them already in 2023… Not just MI300 but also the

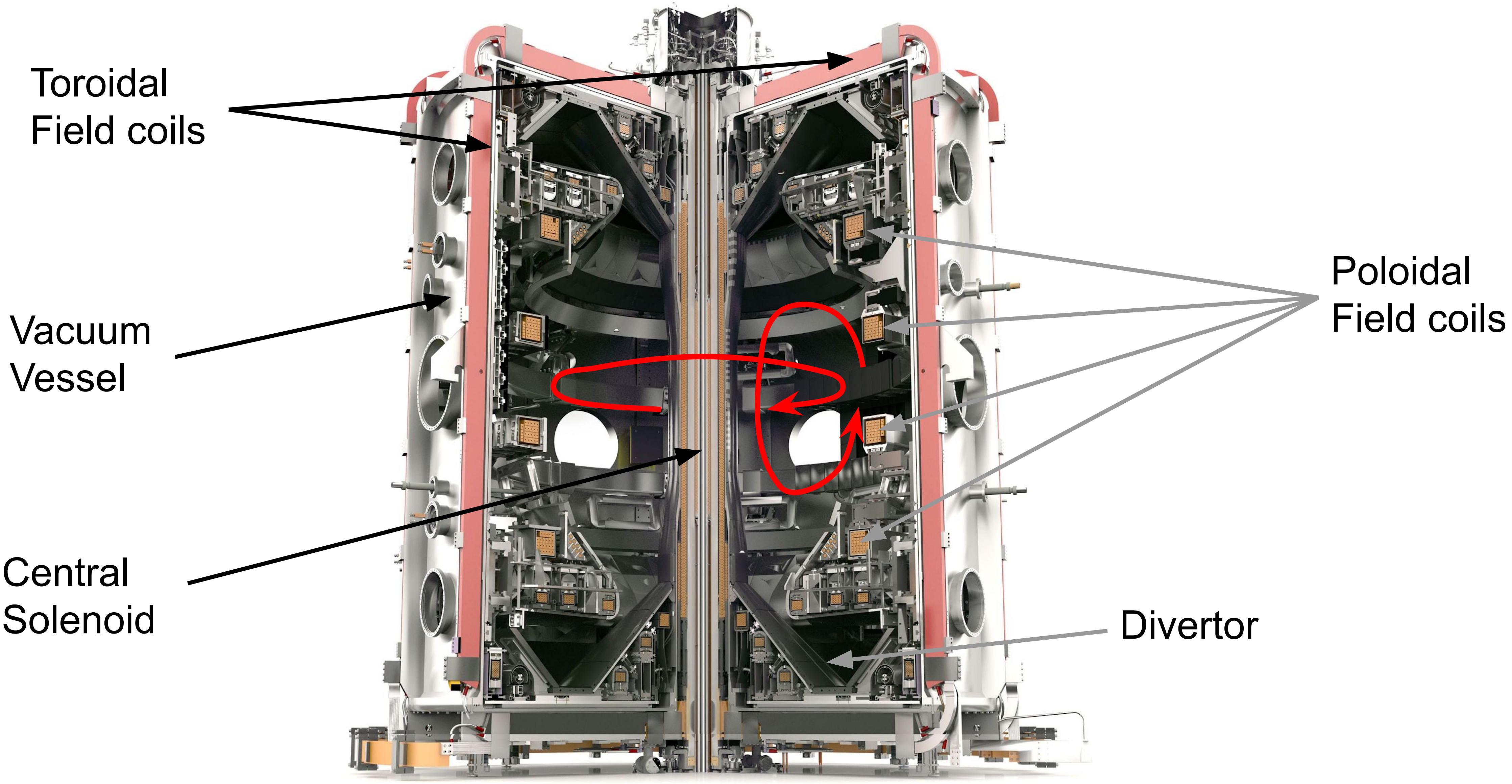
"I have consulted with our business teams, and they have indicated that although NEC is discontinuing efforts to develop low-power consumption accelerators with Japan's New Energy and Industrial Technology Development Organization's (NEDO) Green Innovation Fund, NEC is continuing its High Performance Computing business, including the NEC SX-Aurora TSUBASA."

- I had never heard of the NEC Vector Engine - have you?


AND NOW FOR SOMETHING
COMPLETELY DIFFERENT

Map of a Tokamak

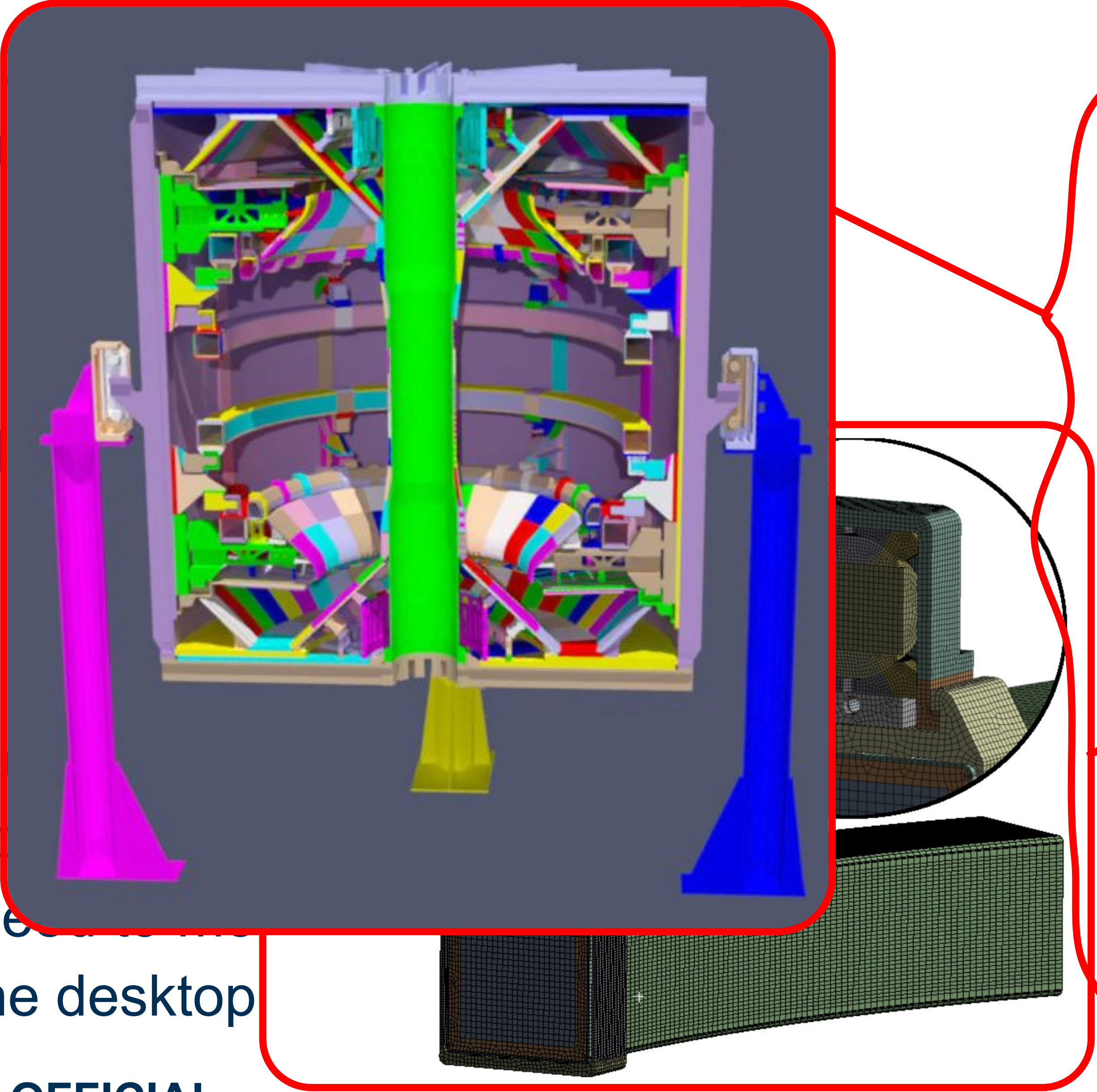
UK Atomic Energy Authority

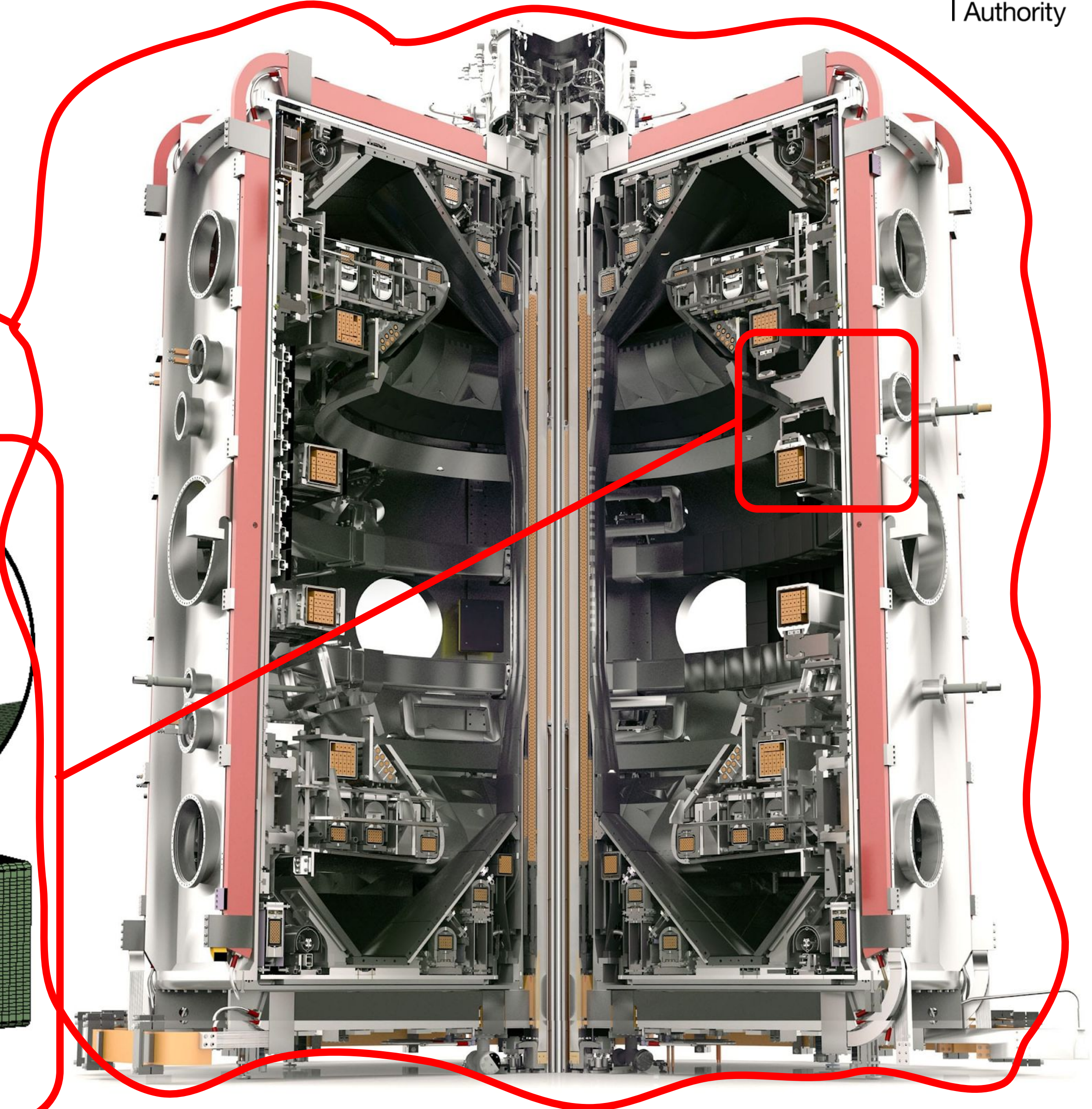


Traditional techniques lack scale

UK Atomic Energy Authority

- S
  th
  -
  -
- M
  -
  -
  -
- C
  sc
- Ne
  the desktop





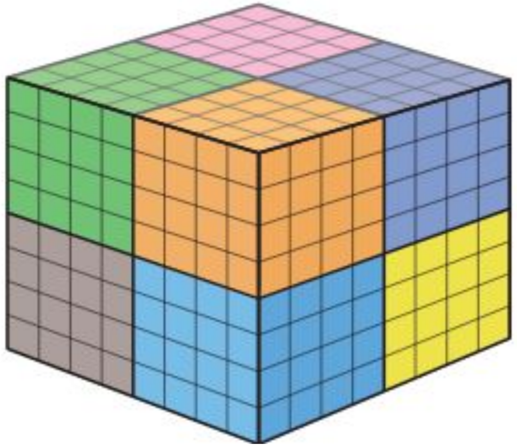
Large scale multiphysics

- If we are to fully simulate complete systems, need scalable coupled framework, it must cover:
    - Computational Solid Mechanics
        - Including dynamic contact
        - fracture mechanics
        - Micro-mechanics?
    - Computational Chemistry
        - DFT, Damage
    - Computational Electromagnetics
    - Computational Fluid Dynamics
    - Computational Radiation Transport
    - Heat Transfer
    - Microstructure (phase field, grains)
    - Diffusion (and reaction) for tritium transport
- Massively scalable ➡️100,000’s of CPUs
- Exascale gazing (considering support for GPU)

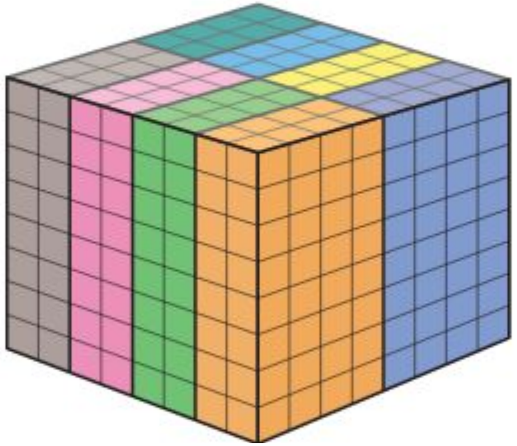
We are trying to provide a scalable multi-physics software framework

UK Atomic Energy Authority

- We need to perform efficient multiphysics simulations
    - Experimenting with tightly coupled framework rather than federated model
    - This means we take a more ***holistic*** view of how we stick different codes together
    - It may be the case that taking the best/fastest physics packages and sticking them all together does not lead to the most scalable solution
        - communication of mesh or solutions could begin to dominate
        - for massively decomposed problems (where we need to be for performance)
        - Time stepping between fast/slow physics may dominate



Ideal partitioning for a CFD problem

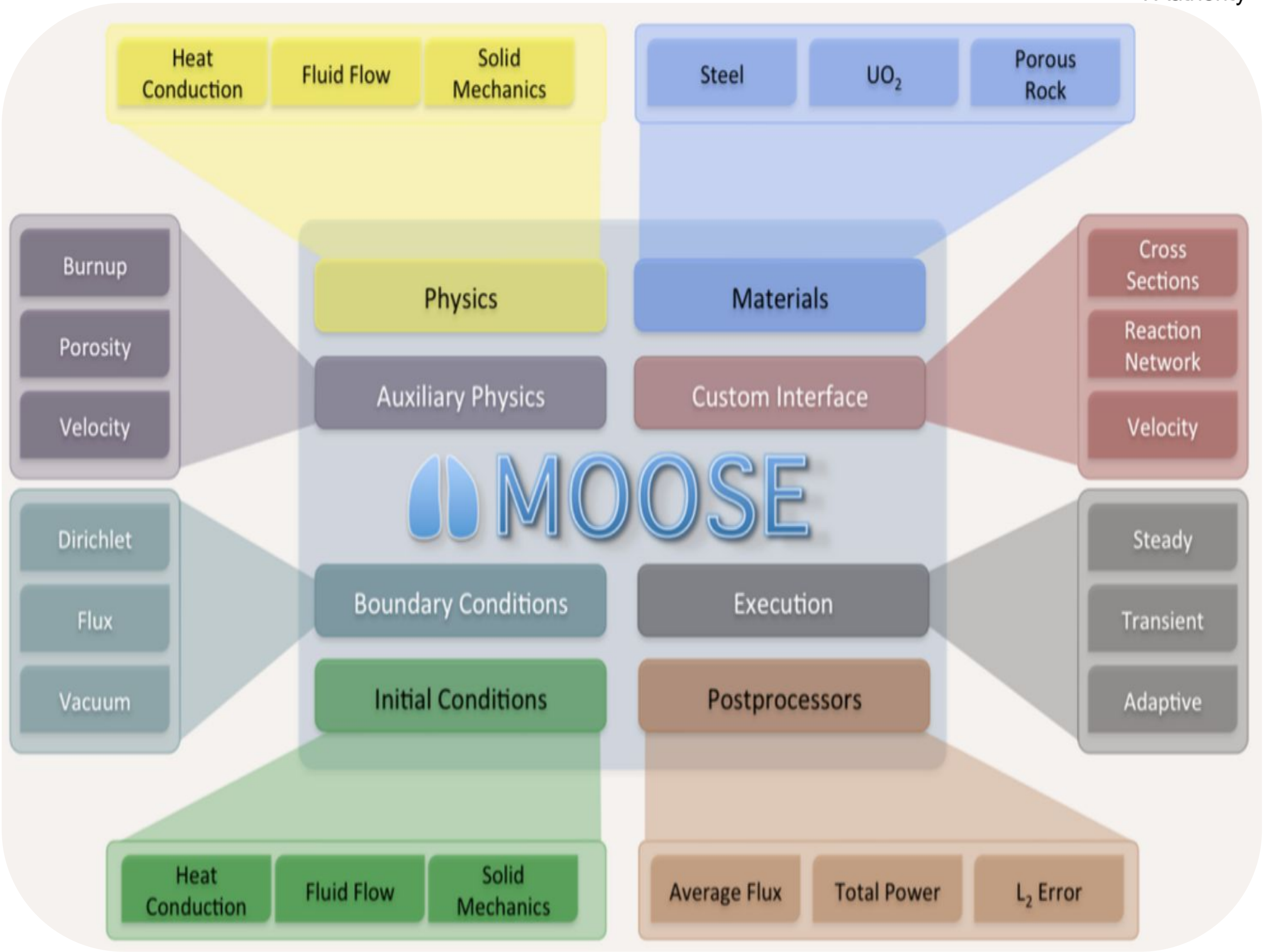


Ideal partitioning for a neutral particle problem

The Pantheon (of our applications)

UK Atomic Energy Authority

- Using the MOOSE framework as the basis for a range of applications to solve fusion engineering problems
  - E.g. neutronics, tritium transport, fluids, electromagnetism, optimisation etc
- Sets of prebuilt physics that can be assembled to fit a wide variety of our problems
- NQA-1 Validated Applications

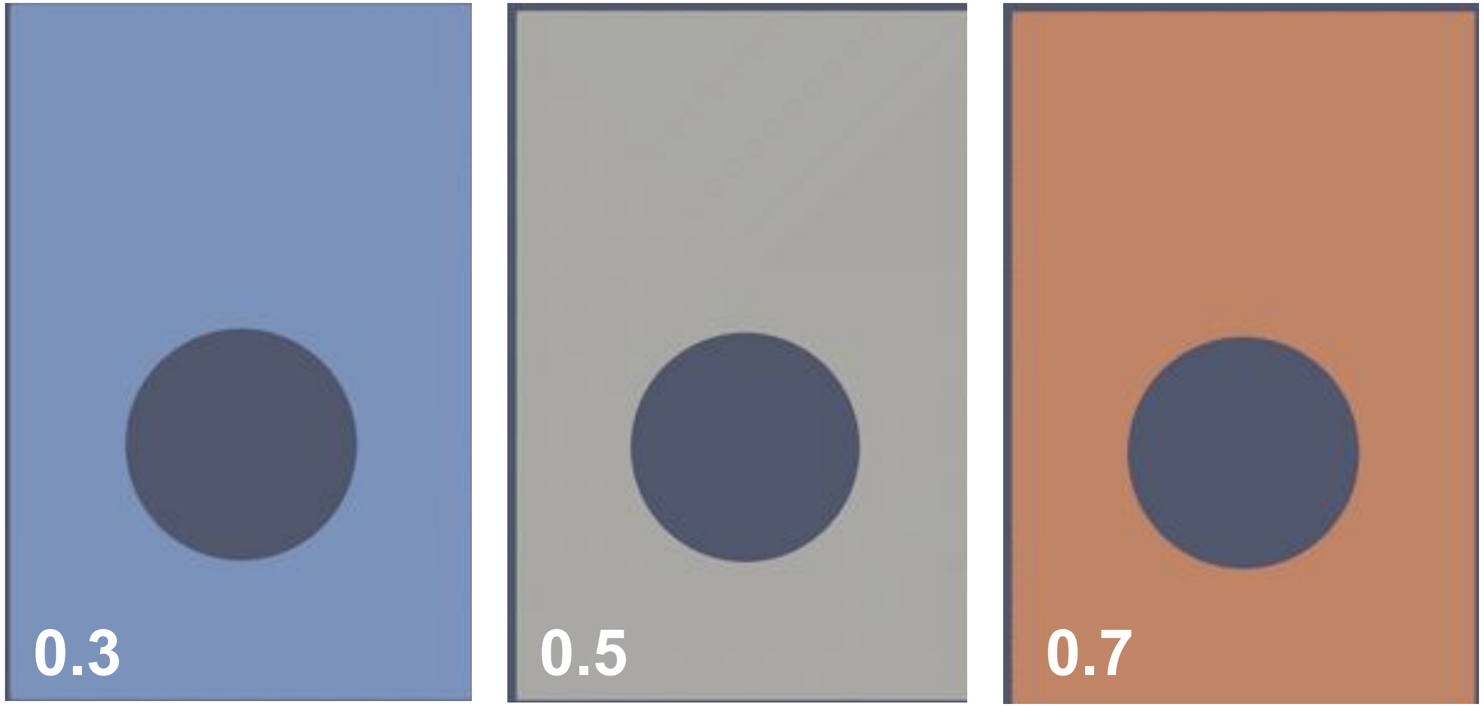


Why MOOSE (c.f. Dubas et al)

Open Source	**HPC scalable**	**Extensible**	**Community**	**Productivity**
Plug and Play Physics	**Capability**	**Scalable**	**Community**	**Quality Assurance**

- Recent feature example, Topology (Thermal & Mechanical) recently added fully parallelised 2/3D optimisation capability.
- **Want** to do more **multiphysics** optimisation
- Coupled T-M common fusion problem

Topology Optimisation - thermal - diverter-ish



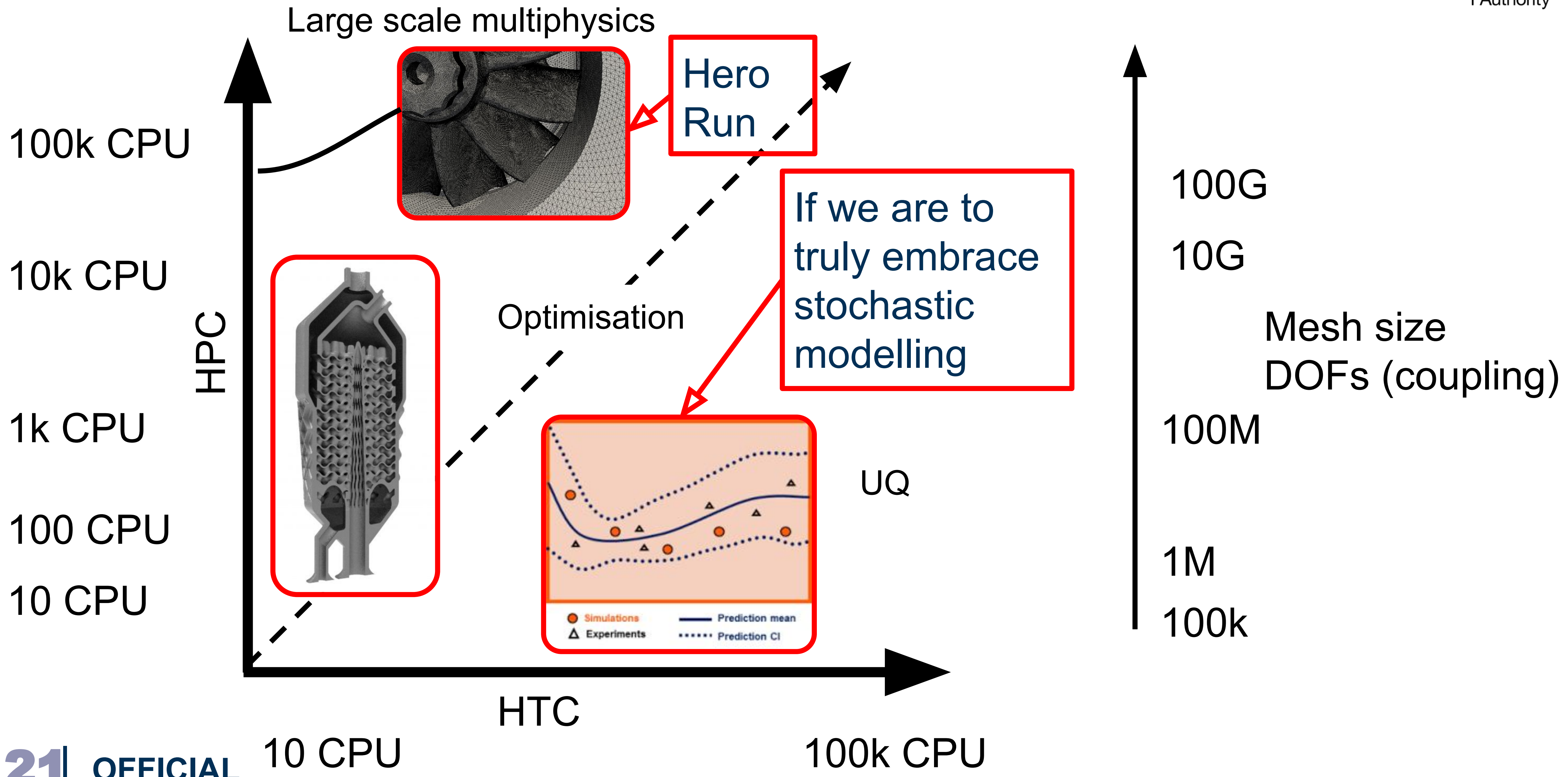
Void fraction

Why is Open so important?

UK Atomic Energy Authority

- In the field of nuclear (anything) but specifically fission based codes are very often subject to e.g. US Export Control, ITAR, or other regulations
  - Often means analysis software needs extensive background checks & Single Site Single User licences 😢 (any MCNP users out there?) - **no Marconi for you**
- It means easy collaboration, we can easily work on projects together to jointly improve software without complex legal agreements
- It means every gets the benefit of the software (no barriers)
- It often means higher quality of code (it also sometimes doesn't :))
- It means more users for your code
- For publicly funded works it is philosophically the right thing to do
- Can be deployed anywhere - take advantage of compute anywhere
- Public development means fewer repetition of doubled development
- Development costs can easily be shared

Scalable in multiple dimensions

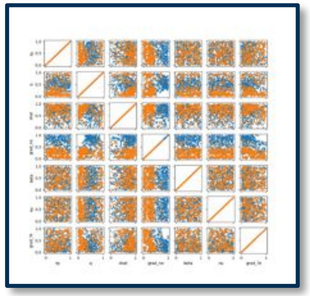


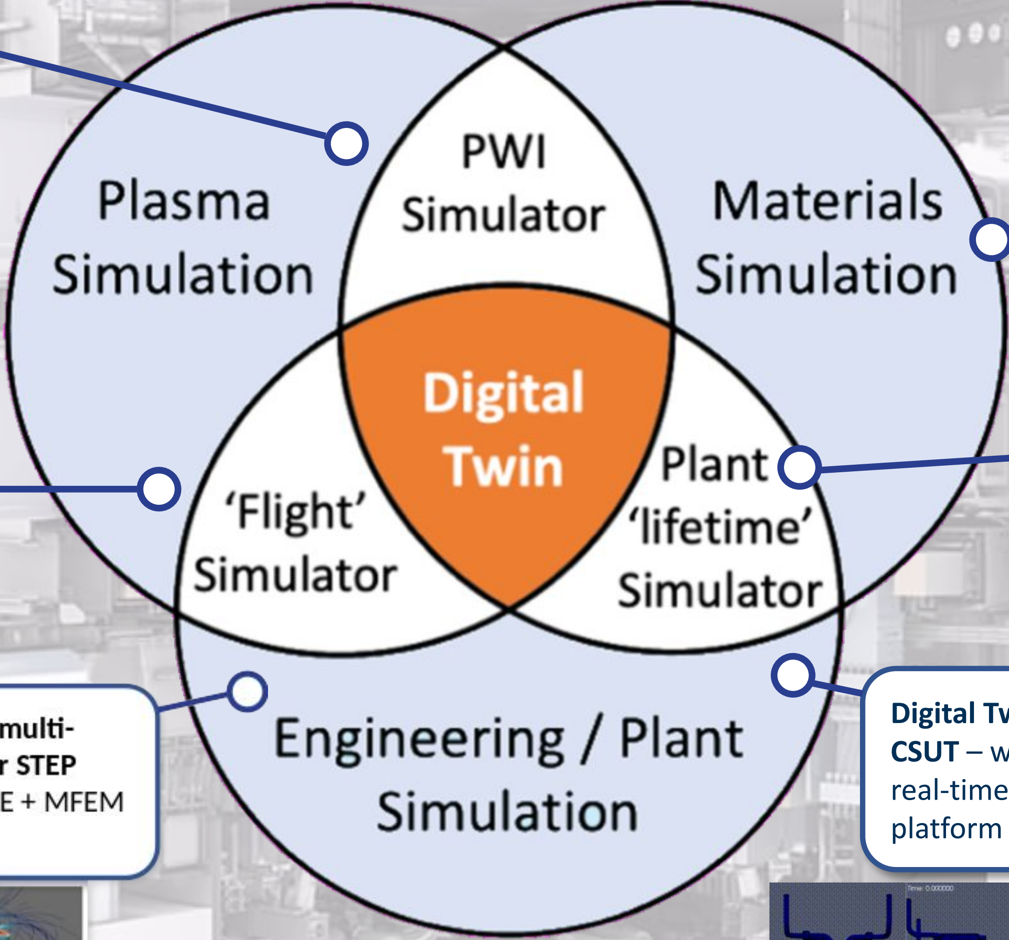
Towards a full plant Digital Twin

UK Atomic Energy Authority

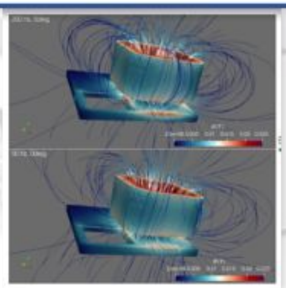
Building **exascale targeted, performance portable plasma simulation** capability

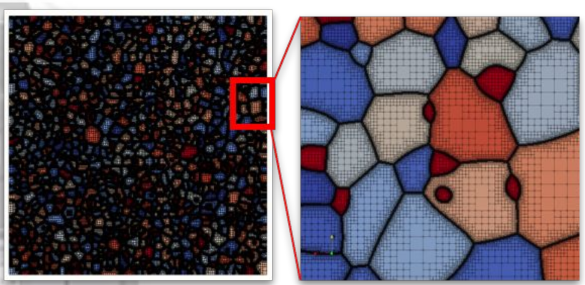
Gaussian Processes, Active Learning etc. to construct **Gyrokinetic emulators**





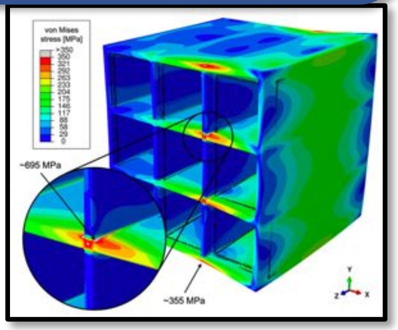
Building a scalable multi-physics platform for STEP (based upon MOOSE + MFEM from LLNL)



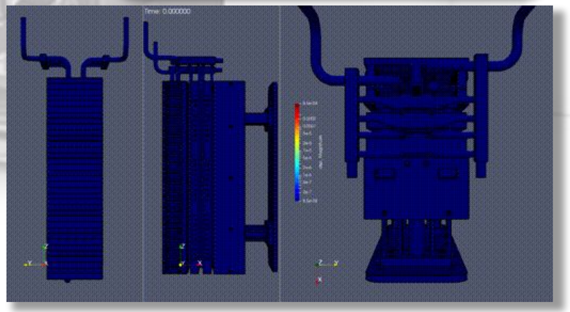


Coupling Multi-Scale materials modelling materials modelling necessarily has many scales to it

Coupling neutronics to FE (Crystal Plasticity Finite Elements) to optimise stress-strain across load assembly



Digital Twin of CHIMERA CSUT – within a factor 4 of real-time using MOOSE platform



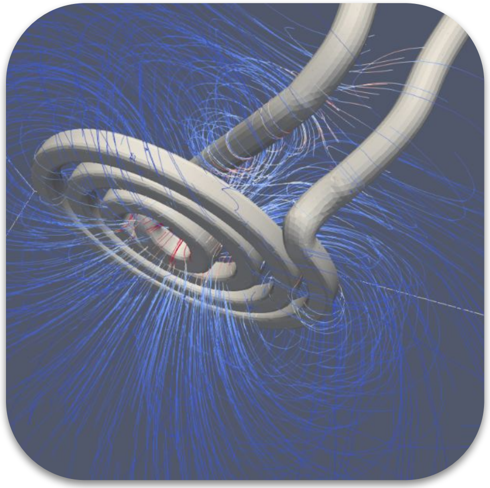
Chimera:

(Greek mytholo

...walking before running – construction of sub-system proxies

UK Atomic Energy Authority

HIVE



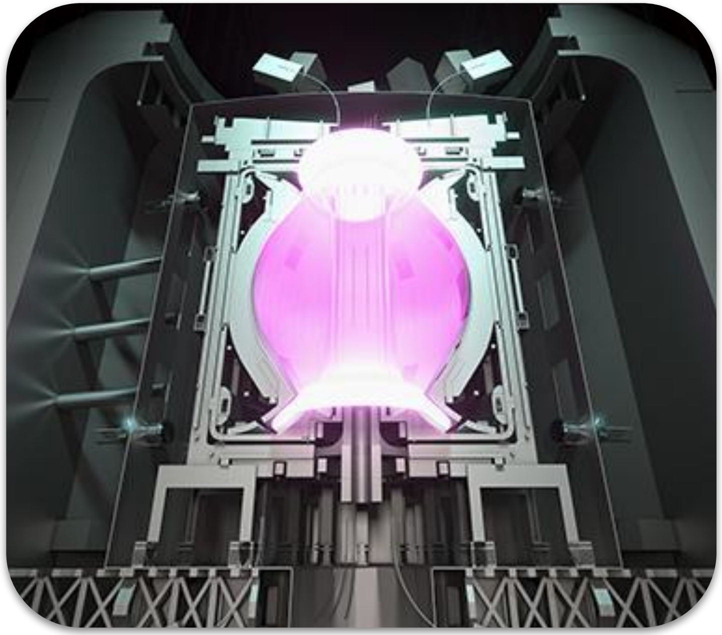
Component / small rig Level

Digital Twin focus in near term

Thermal Hydraulics (CHIMERA)

Existing Facility: Sub-system proxy

STEP Powerplant



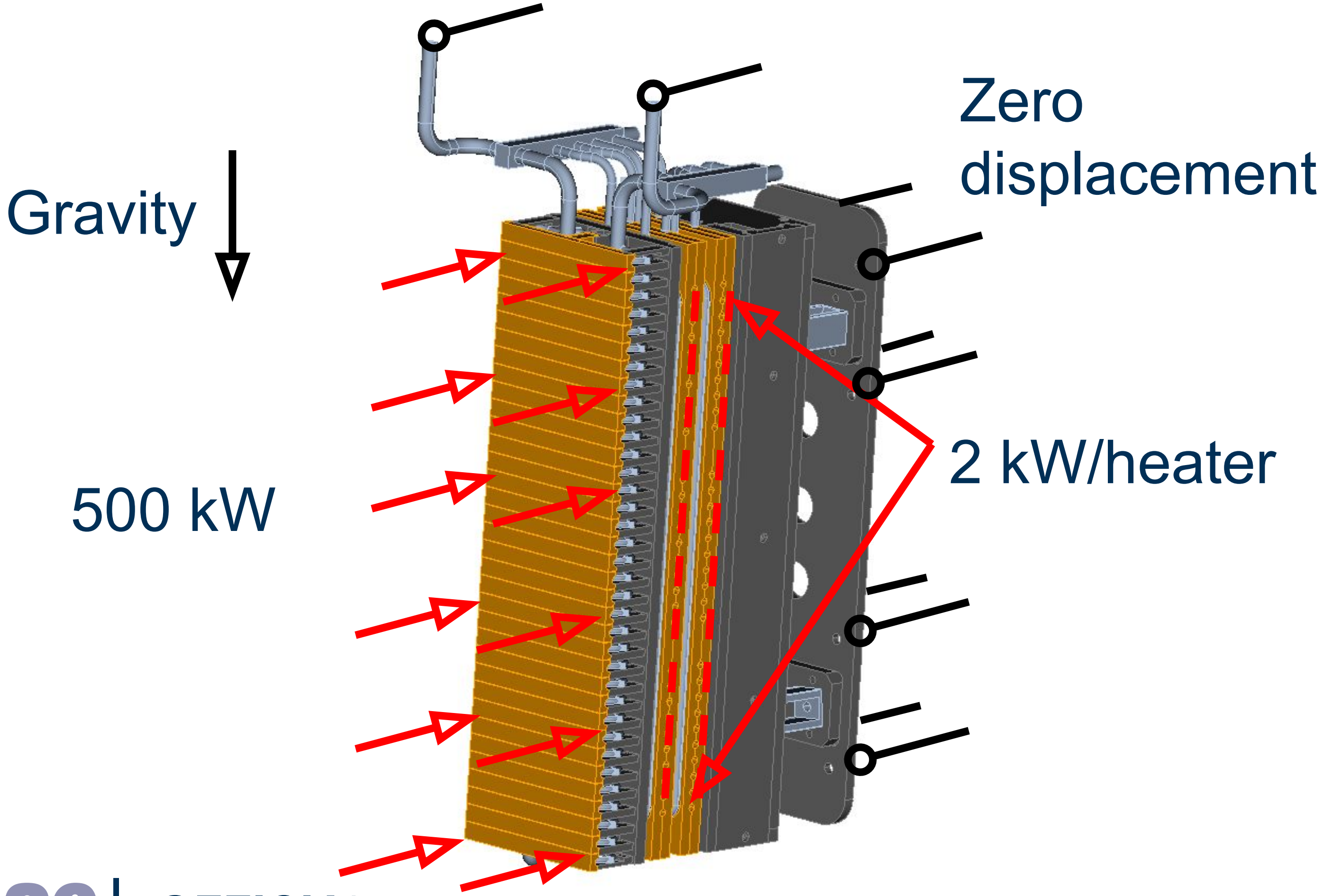
Integrated Plant Design

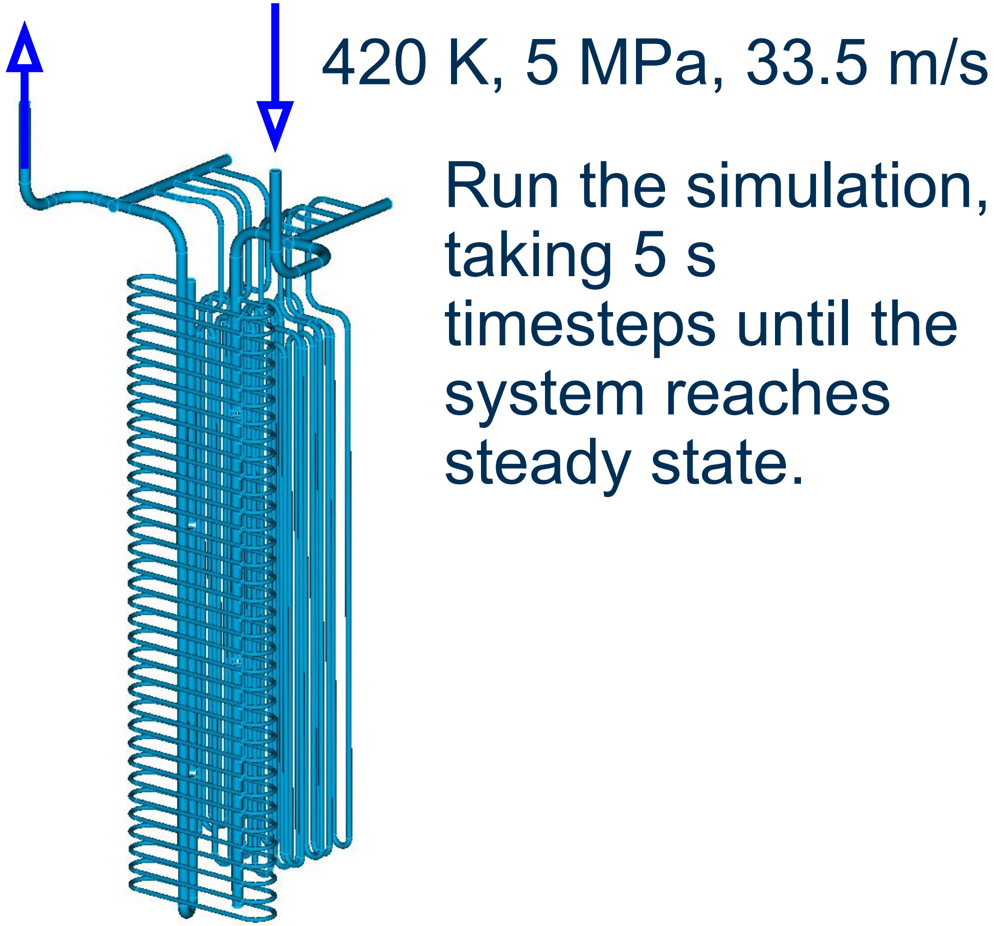
Small scale directed testing necessary, but validating some physics in isolation of others

CHIMERA Definition

UK Atomic Energy Authority

- CHIMERA load case, currently approximated atmospheric convective HTC boundary condition. There are other CHIMERA cases, e.g. the magnetic sample, this focuses on the thermo-mechanical sample



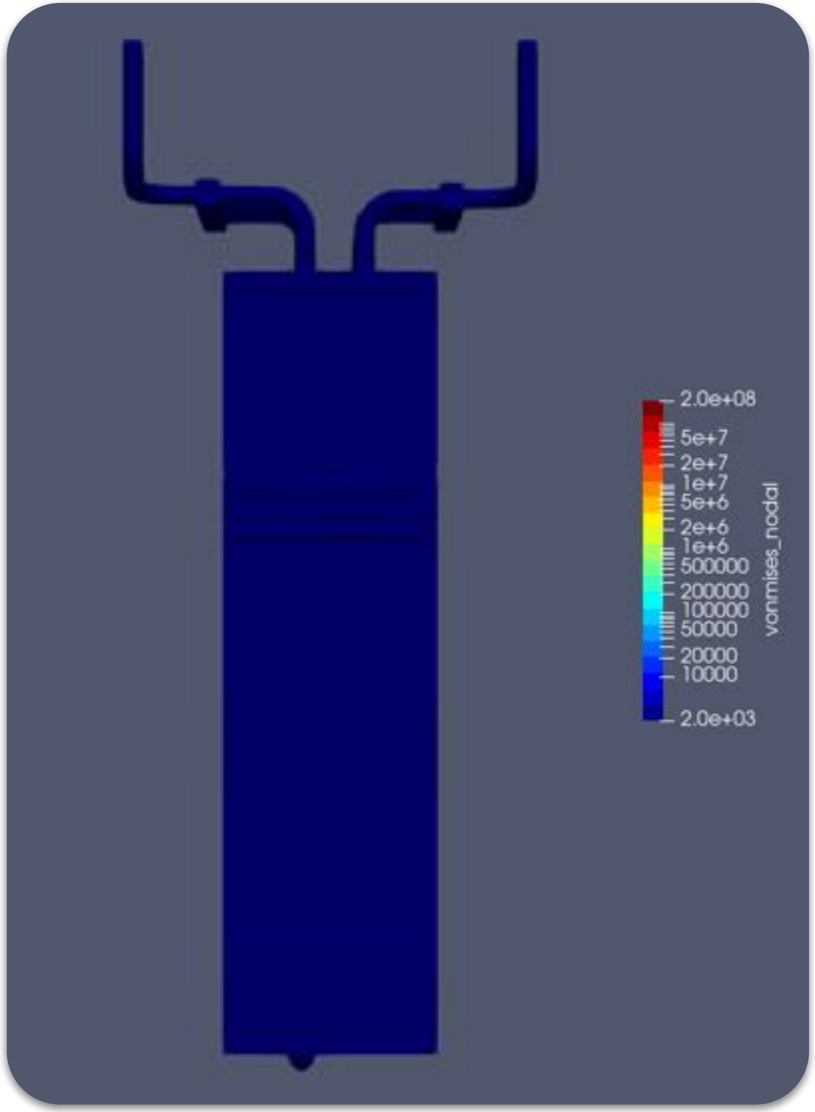


Run the simulation, taking 5 s timesteps until the system reaches steady state.

CHIMERA Simulations

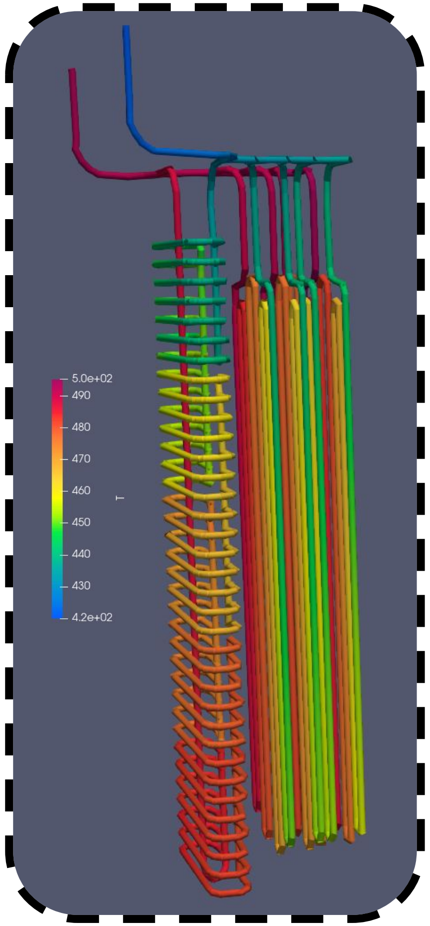
UK Atomic Energy Authority

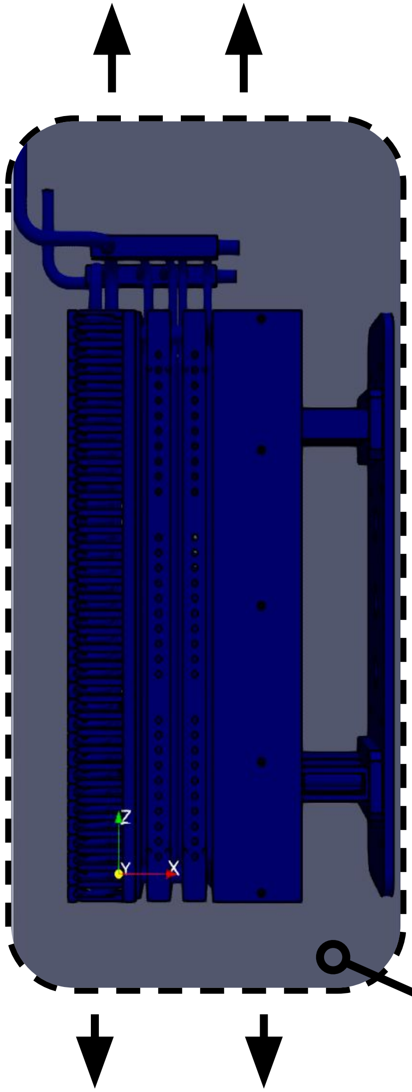
- New Intel Sapphire Rapids machines at CSD3 have improved throughput of simulations by a factor roughly 6x
  - reduced the simulation time from ~12 hours → **1.8 hours**
  - This brings our current best efforts to within a factor of 4x of real time, i.e. 1 second of real time takes 4 seconds to simulate
- There is potential to reduce the runtime potentially further, e.g. switch to GPU solve via Kokkos could get further **2x** speedup
- Need to increase fidelity further, to include impact of thermal radiation, cryostat and so on
- All other toolchain in MOOSE ecosystem benefit, e.g. STEP blankets, HIVE etc.



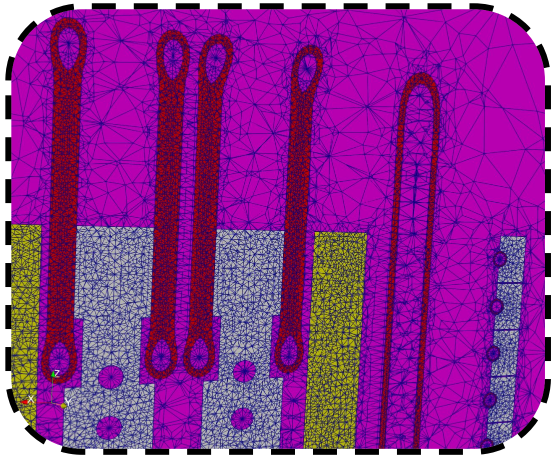
Specific CHIMERA Plans & Challenges

- Thermal simulations
- EM simulations
- Link with systems?
- UQ - likely MC based
- Thermal-Hydraulic network
- CFD Coupling
- Reduce BCs
- Higher order reps
- IGA based version
- More detail and more fidelity
- Fewer assumptions, let more emergent behaviour arise





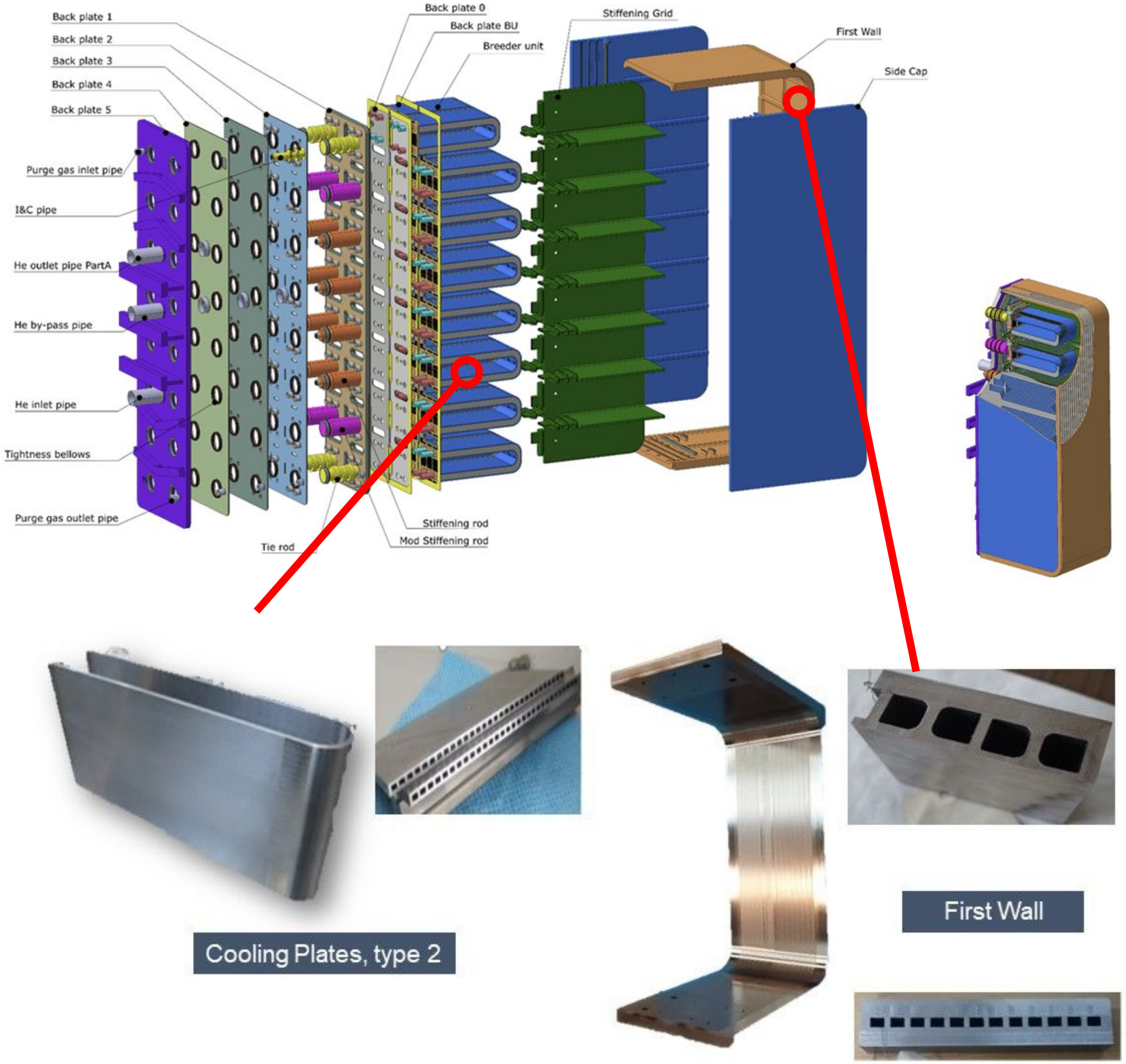
Move assumptions further away from the CSUT assembly



Vacuum mesh generation in the gap (thermal radiation & EM modelling

CHIMERA serves as a useful surrogate...

- CHIMERA serves as an experimental surrogate for blankets and divertors
  - The experimental conditions aren't equivalent but component complexity comes close
  - Common physics between them, similar challenges with time stepping fast/slow physics
- CHIMERA load case is not equivalent to a full fusion load case (no ionising radiation)
  - However, despite that if we cannot model CHIMERA then we are going to face issues when attempting full fusion systems
- Ultimately, will include MHD

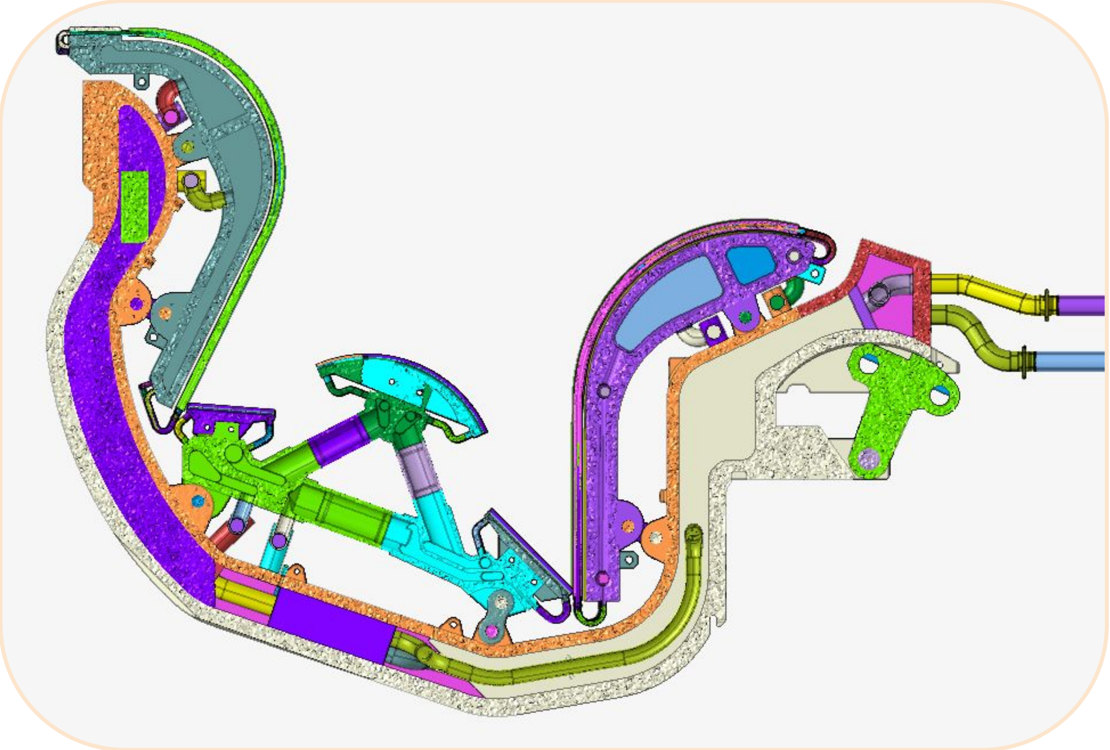


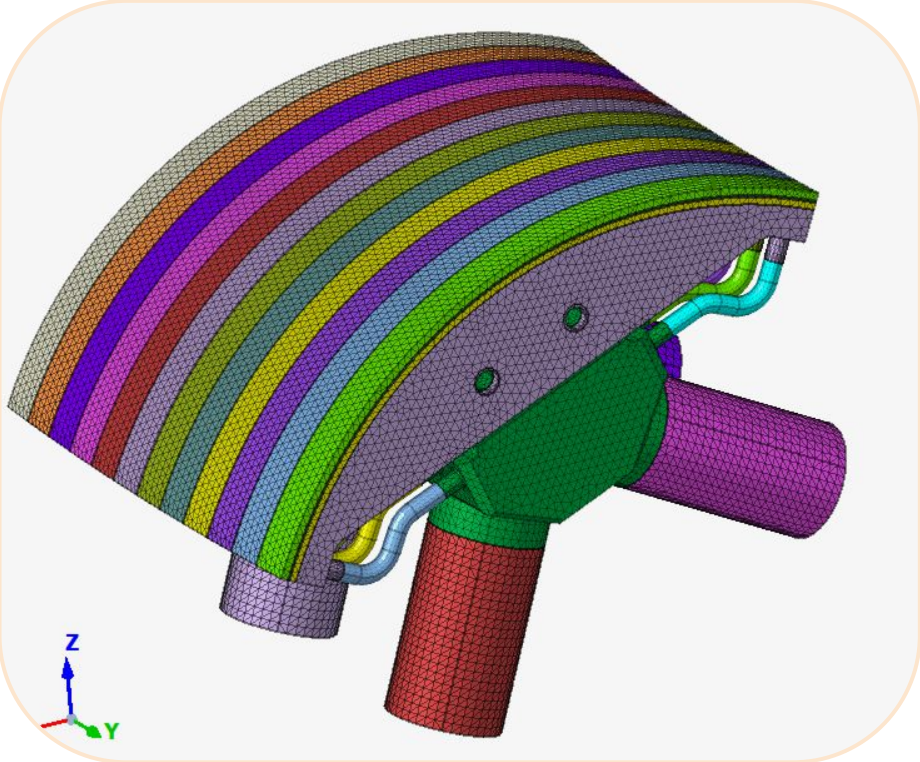
“I do not fear computers, I fear the lack of them”

Isaac Asimov

It is not however, all plain sailing...

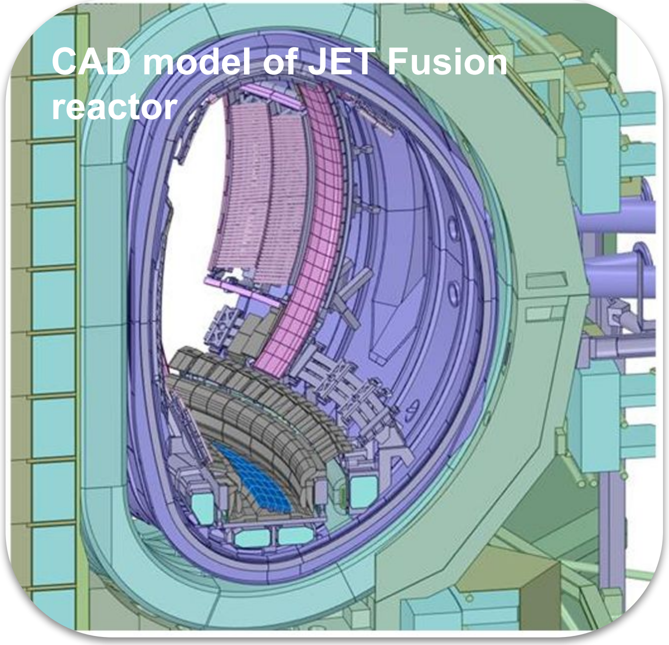
- I'm fairly confident that if we can mesh a geometry, then we can solve the problem with enough brute force parallelisation & AMR - see ITER divertor
- However, preprocessing of complex geometry
  - Takes a long time ….
  - I've had good success with SimLab - performant even on large geometries

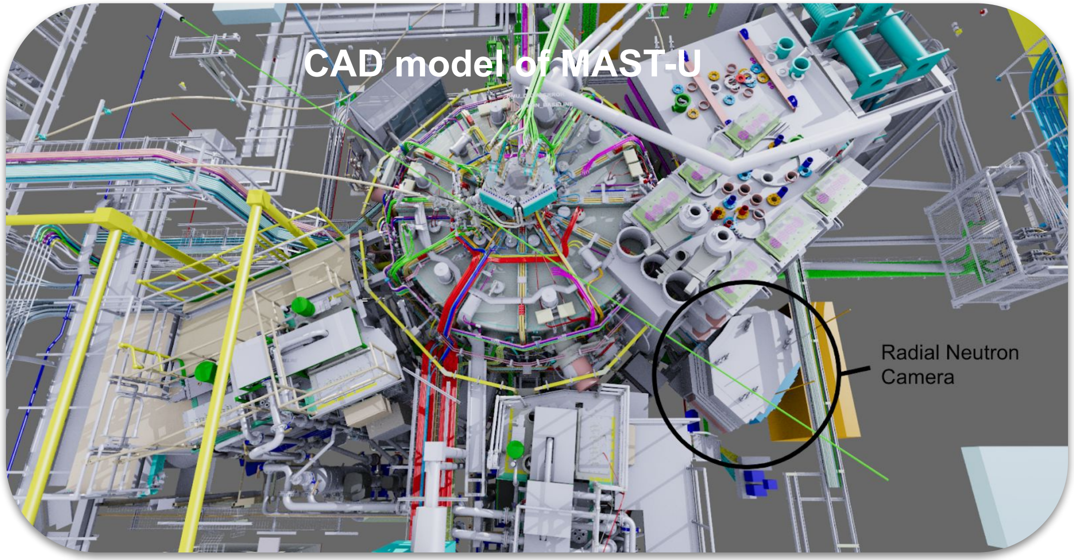




Geometry processing dominates ....

- …over simulation runtime
- As geometry complexity grows; geometry pre-processing time grows something like something like $O(N^2)$ where N is the number of components
  - Some degree of automation is required
  - CAD is error prone and often inconsistent with reality - how to fix?
  - Very large geometries can be difficult to check
  - Not only require no geometry overlaps, some analysis requires consistent gaps e.g. mechanical contact
- Fire-and-forget Mesh processing efficiently parallelised e.g. `BoxerMesh` or `psculpt` - meshes are often not conformal
- We need a focussed effort on CAE, including meshing for HPC, including metadata





UK Atomic Energy Authority

“

The secret of change is to focus all of your energy not on fighting the old, but on building the new

”

Socrates

How can we get engineers using HPC?

- The STEP programme aims to deliver fusion energy on the UK electricity grid in the 2040 timeframe
- If we are to do so, we will need to deploy potentially 10's of thousands of engineers to perform detailed design work
    - **Problem:** Fusion engineering calculations are quite strongly coupled multi-scale multi-physics simulations
    - **Problem:** COTS software does not deliver all the physics needed to simulate fusion engineering
    - **Problem:** COTS software cost model does not scale
    - **Problem:** COTS software is a black box - difficult to prove correctness
    - **Problem:** Many COTS software does not hardware scale

How can we get engineers using HPC?

UK Atomic Energy Authority

- Only a few issues to solve there, shouldn't be too bad…..

  - **Solution:** Use a software that allows multi-scale multi-physics problems to be solved
  - **Solution:** Use software that can couple in other physics, or is setup to solve arbitrary PDEs
  - **Solution:** Use free/open software
  - **Solution:** Use open source software
  - **Solution:** Use software that can scale

- Great, let's use that and….
  - **Problem:** Is it validated for my use case?
  - **Problem:** Does it have a GUI?
  - **Problem:** What's UQ?
  - **Problem:** Where do I get training from?
  - **Problem:** How do I use a terminal?
  - **Problem:** Whats Docker?

Conclusions

UK Atomic Energy Authority

- The timelines to deliver fusion are short, with insufficient time to perform experimental driven validation
- Simulation must be used as the 3rd mode of discovery
  - Can only be done if simulation is actionable
  - Requires VVUQ
- The simulation we use must be **open** anyone who has suffered export controlled software will understand the pain
  - We need to deploy this kind of software to 10k engineers to deploy on fusion problems, needs to be useable, needs to be freely, available, needs to be validated
  - If we can get industry using open HPC scalable tools, it is they who will really drive innovation and deploy these tools to solve societal problems, build better bridges etc
- We have demonstrated real problems running on small amounts of hardware, now need to the scales to 100k cores, with a vision to running simulations for whole reactor systems