

HPC, AI & The Design of Fusion Reactors

HPC-AI Advisory Council 8th Annual UK Conference

Andrew Davis, Head of Computational Engineering & Materials

14th October 2026

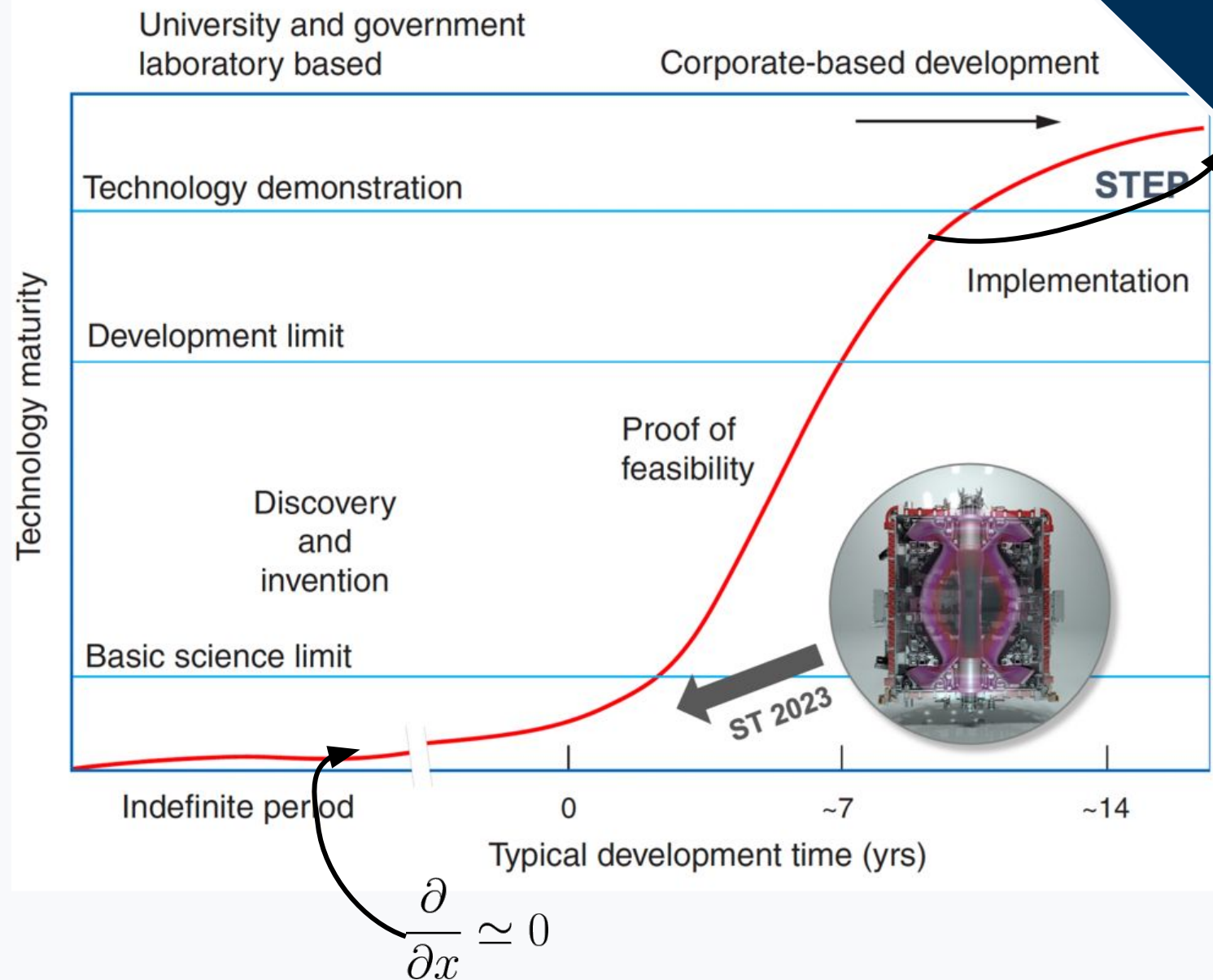
Evolutionary design nature is patient, humans are not

The process of design is going through a number of steps;

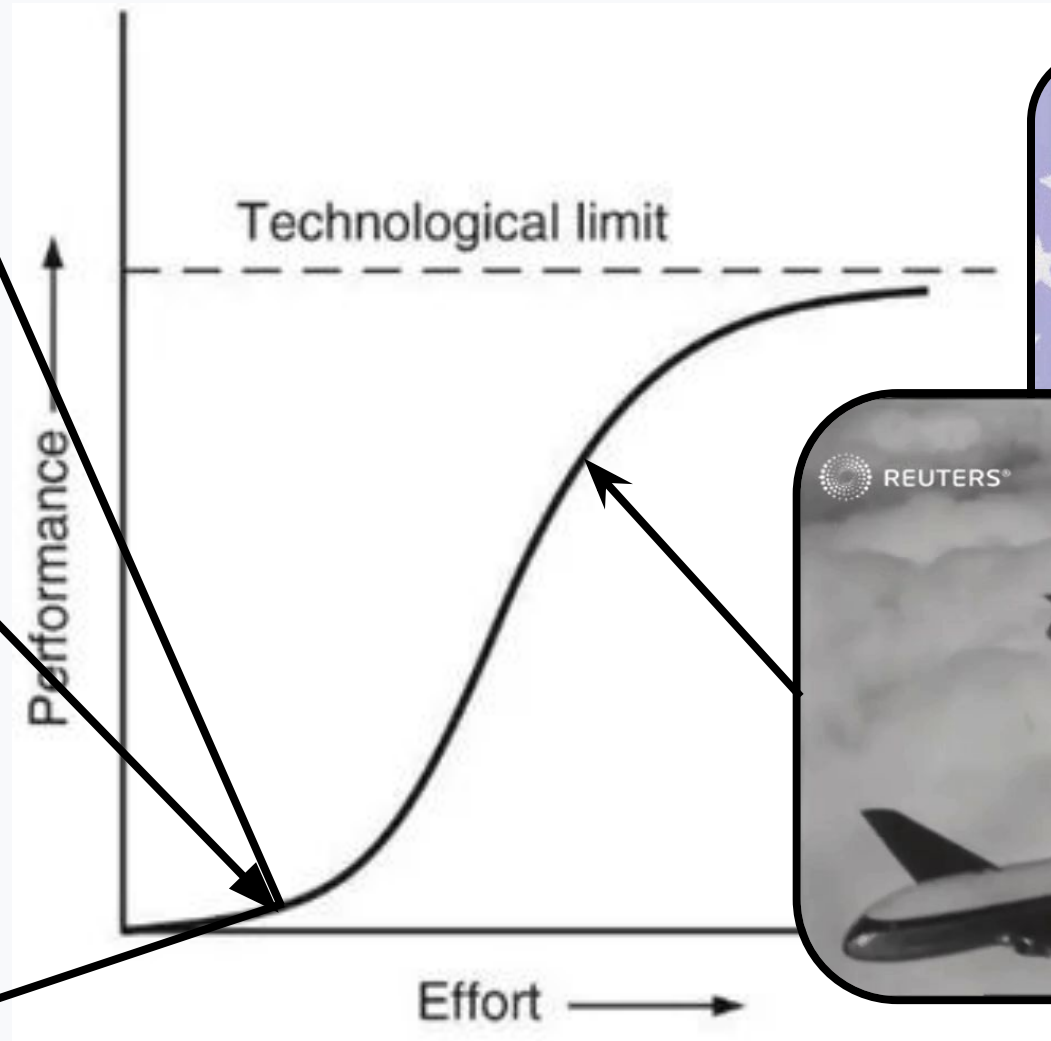
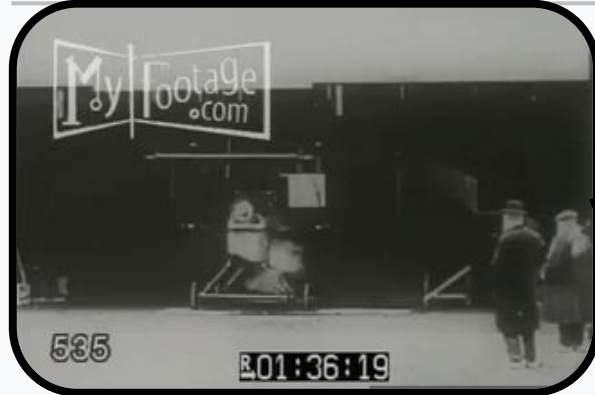
1. Define the Problem
2. Collect Information
3. Analyse
4. Develop Solutions
5. Gather Feedback
6. Improve

Historically (in most fields) designs are progressed through a series of iterative steps

Most developers of most technologies typically develop at the mature end of the TRL scale, where the gradient is flat - "flatliners" - presents cultural challenges



Design is historically achieved through painful experience



Design is historically achieved through painful experience

Iterative design results in successive failures, information gathering, post mortems to determine root causes of failure.

Usually where the technology is well understood and characterised, towards the top of the 'S' curve.

If failure is cheap we can learn tremendous amounts of things in this process, however as the cost creeps up, design by failure starts to become too expensive

Rapid iteration can lead to rapid learning, if failure is cheap or information rich.



Design by iteration no longer scales

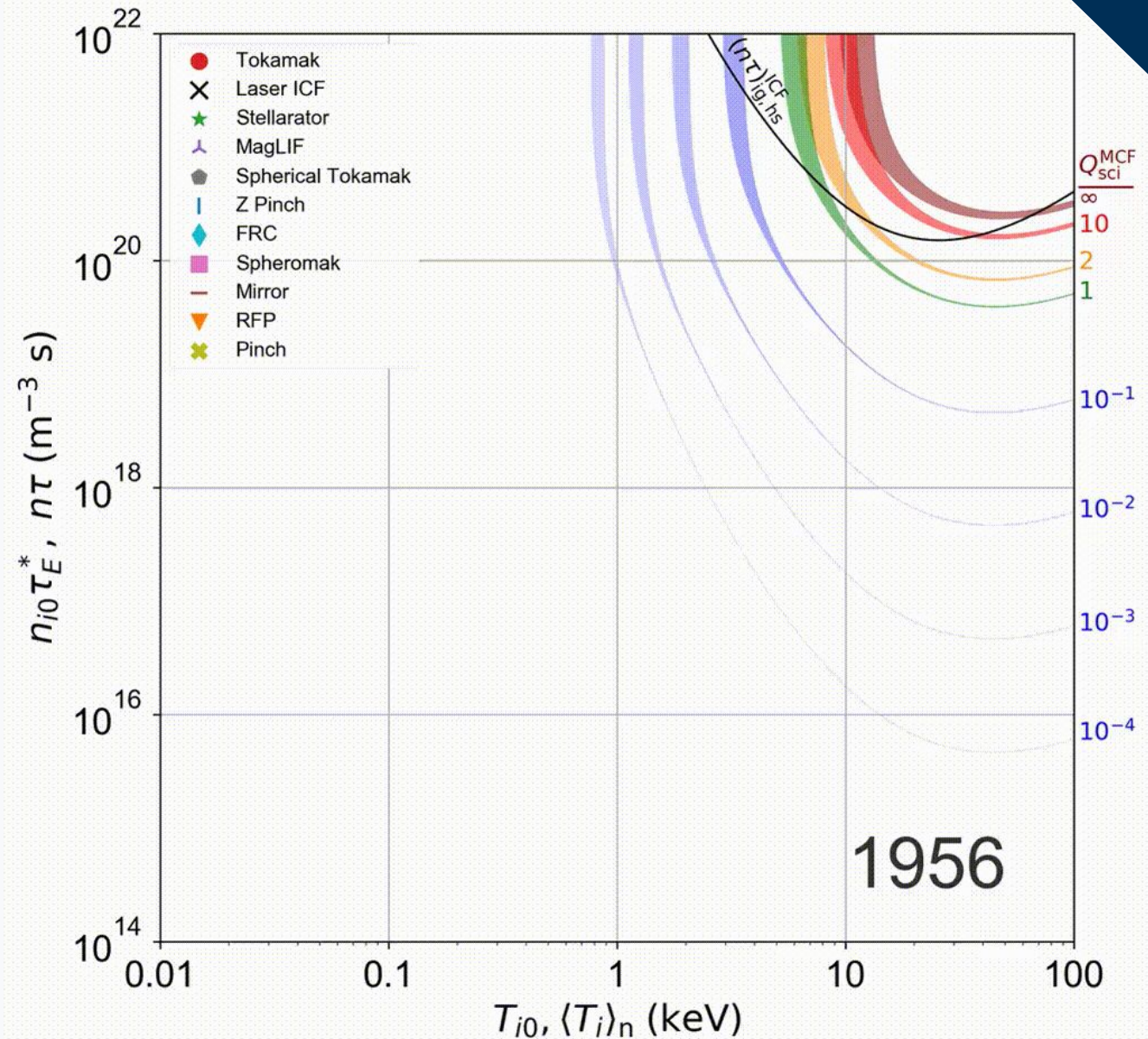
Fusion devices (like fission plants) are capital intensive,

- ITER cost £22B (Construction only)
- Hinkley Point C (£40B)
- SMR < £3B

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 $B^2 R^3$.

As we approach $Q \sim 1$ - \$\$\$\$\$\$



Experiments not sufficient require faster solutions

- All fusion devices (with the exception of those 8 NIF shots and JET) are low (relatively speaking temperature non-nuclear devices)
 - Neutron fluxes too low to cause large amounts of transmutation
 - Neutron damage too low to damage materials
 - Temperature less than around 100 °C in structural members
 - Strongly transient devices - 10'000 k pulses
 - Fusion pilot plants will be steady state or weakly transient
 - High neutron flux 10^{15} n/cm²s
 - High Damage > 50 DPA/FPY
 - Higher operational temperature (500-1000 °C)
 - We do not have good models (or data) for materials that will survive more than 20 and at the right operational temperature
 - All material properties strong functions of neutron dose
 - No global phenomenological model exists - working on it
- Recent work suggests Ferritic Martensitic steel may reach their end of life at 15 DPA; due to radiation induced thermal creep and rapid fracture radiation reduced fracture toughness after 1 year of 1.2 MW/m² of neutron irradiation ([link](#))

Complex interacting systems require solutions that can evolve



Kadlytics (defunct) recognised physics interferes with an otherwise standard design process

Fusion design sits in an evolutionary space, where experiments, test rigs and non-fusion devices validate and provide (physics) feedback. AI can cut through the noise.

Tension in requirements during the design process cause long integration times and frequent rework



Agentic workflow and GPU compute is the saviour we need

THE NEW STACK

AI AGENTS / AI STRATEGY / LARGE LANGUAGE MODELS

Andrej Karpathy says only 5,000 people have seen what AI can really do

Andrej Karpathy says the AI capability gap is widening. Box CEO Aaron Levie says agent swarms are next, but most enterprises aren't ready to use them yet.

Oct 12th, 2026 11:54am by [Amanda Caswell](#)

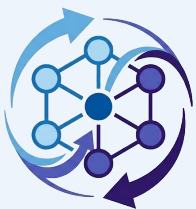
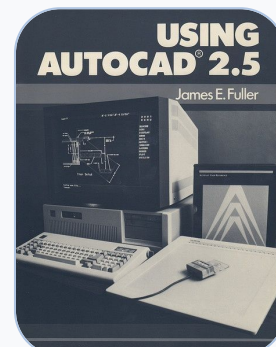
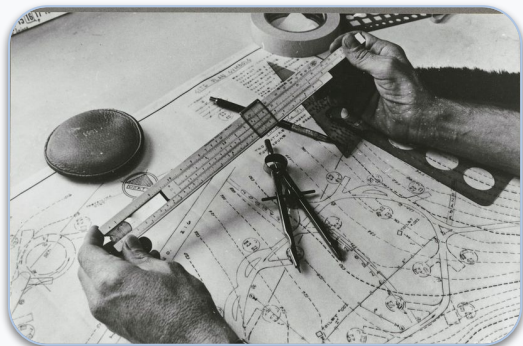


If I am not one of the 5000 then I for one welcome our robot overlords (I have no idea who Andrej Karpathy is)

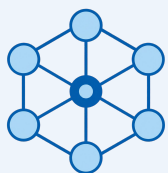
Vibe Engineering - human describes intent and checks



Engineering & design is approaching an inflexion point



**Acceleration of
change**



**AI has already
changed much**



AI Native

Prompt:

Design a deuterium-tritium magnetically confined fusion reactor, it needs to be tritium self sufficient, maintenance should happen no more than once every two years. The output should be in the form of a PDF with supporting CAD files ready for manufacturing. Make no mistakes.

CHOOSE YOUR SUPERPOWER



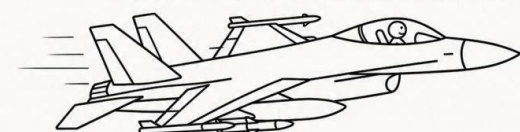
BOSS



LEADER



INTROVERT



ADHD INTROVERT WITH AI

HPC enabled agents can answer questions, learn and interact

Director Agent: “Design a breeding blanket that has a $TBR > 1.1$, do not use any beryllium.”

Project Manager Agent: “Ensure all sub agents are producing useful information”

Concept Design Agent: “Scan literature for blanket concepts that meet the TBR requirement”

Geometry Agent: “Given the concept, instance geometry”

Meshing Agent: “Given the geometry, instance a mesh”

Physics Agent: “Given the instructions from CDA, produce estimates of TBR”

Self reinforcing design cycles, equivalent to automated DOE

May start as human assisted inner loops, increasingly expanding to more AI control

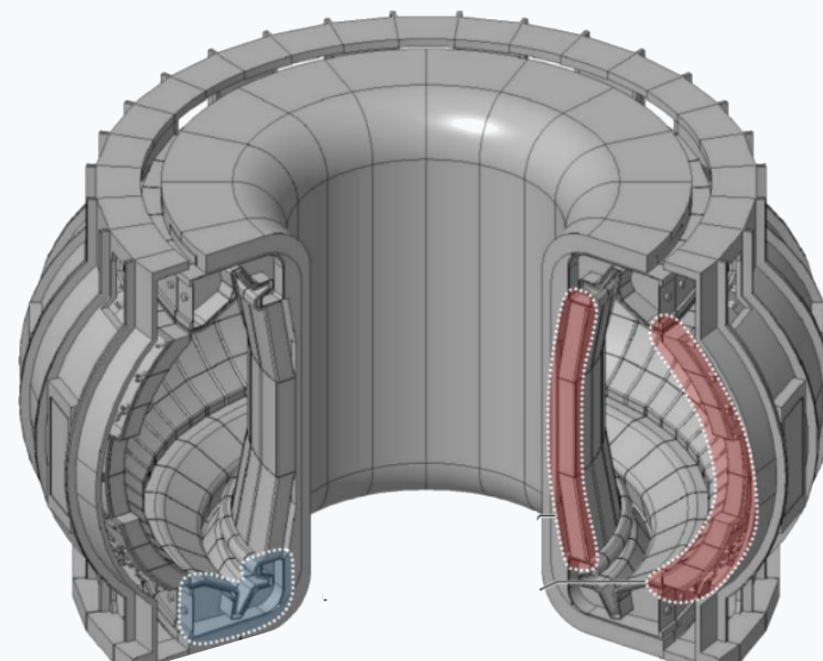
Open source, APIs and adaptability are key to exploitation given the rate of change



OpenMC



coreform



GPU Acceleration is now a requirement not a nice to have....

*** INTELLIGENCE PER GPU

OCTOBER 2026

How many GPUs did it take?

Each model's training cluster. 1 square = 100 GPUs.

Mistral Large 4 Mistral · Grace Blackwell GPUs

0

GPT-6 Astra OpenAI · Grace Blackwell GPUs

0

Grok 4.7 xAI · GPUs

EST 0

Claude Opus 5.5 Anthropic · Trainium2 chips

EST 0

Index: Artificial Analysis Intelligence Index, best setting, 6 Oct 2026. GPU counts: Mistral and OpenAI published theirs. xAI and Anthropic did not, so we use our estimates. Grok 4's Colossus cluster and Amazon's Project Rainier.
Sources: mistral.ai, OpenAI, xAI, Amazon, artificialanalysis.ai

With 100 GPUs we can run inference on the largest frontier models that exist today.

Training a sovereign frontier model would be closer to 10k GPUs operating for months.

HPC Architectural Shift: Modern supercomputing architecture is fundamentally GPU-driven. CPU clusters simply cannot provide the power efficiency or compute density required for next-generation scientific computing. If a code isn't GPU-native, it cannot effectively utilize the vast majority of compute cycles available on modern national or institutional clusters.

The VRAM Ceiling: Workstation hardware in 2026 allows massive datasets, complex 3D meshes, and tensor networks to reside entirely on the GPU. For domains like fluid dynamics, materials science, or liquid-metal magnetohydrodynamics, eliminating the latency of shuffling data across the PCIe bus is transformative.

AI and Multiphysics Convergence: Scientific software is increasingly blending traditional physics solvers with AI surrogate models, frameworks like MOOSE, Cardinal, or OpenMC require a unified hardware environment where both the heavy physics kernels and the machine learning inferences can run concurrently on the GPU without context switching.

GPUs are a key enabler for science

we can do more than was ever possible in the past; we can produce virtual environments to train AIs, we can do simulation at a scale unimaginable even 5 years ago.

Predicting and steering the future of systems of systems



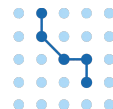
Empirical / Data-Driven (AI)



Experiments



Diagnostics



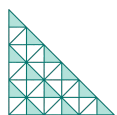
Operational
Data



AI /
Inference

$$\frac{dq}{dt} = \mathcal{R}(q, t)$$

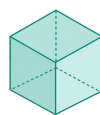
First Principles
Physics



Numerical
Models



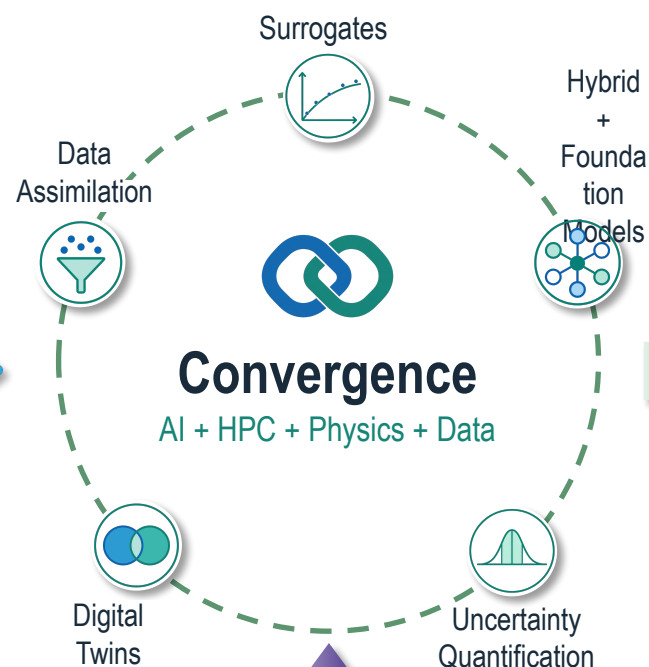
Simulation



Reduced-Order
Models



Theory / Model-Based (HPC)



Quantum-Classical Methods



For hard *ab initio* materials & chemistry kernels

Actionable Predictive Capability



Tokamak State
Prediction



Control &
Optimisation



Design of Fusion
Power Plants



Trusted Decision
Support

Where AI meets HPC, theory, and experimentation.

The Culham AI Growth Zone



UKAEA Culham Campus announced as the UK's first AI Growth Zone by the Prime Minister in Jan 2025

SUNRISE

World's most powerful supercomputer dedicated to fusion – “application readiness” for Culham AIGZ



UK Atomic
Energy
Authority

DAY
10



DELL
Technologies

intel®



UK Atomic
Energy
Authority



UNIVERSITY OF
CAMBRIDGE



WEKA®



AMD

DELL
Technologies

intel


UK Atomic
Energy
Authority

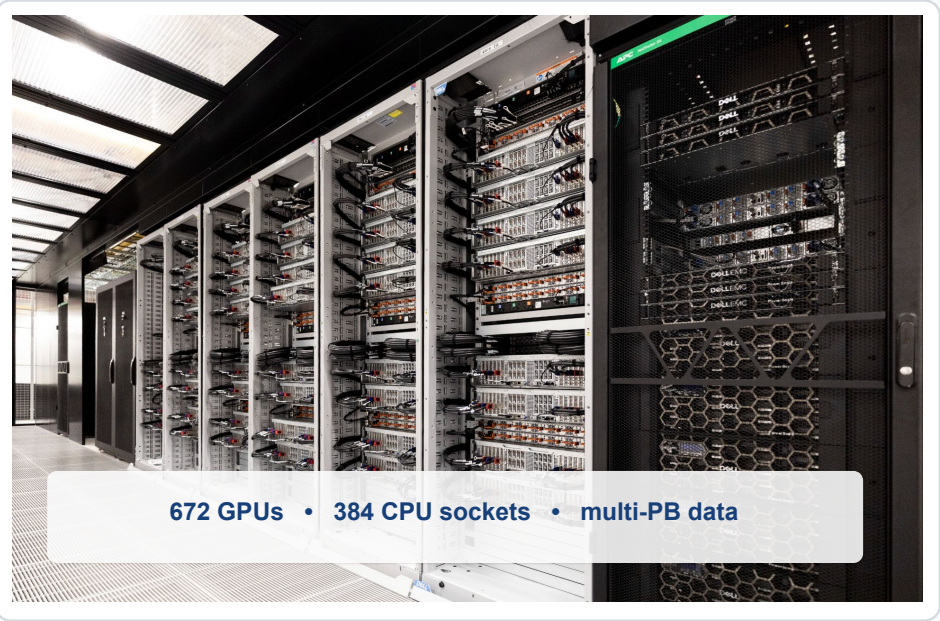
 UNIVERSITY OF
CAMBRIDGE

 WEKA

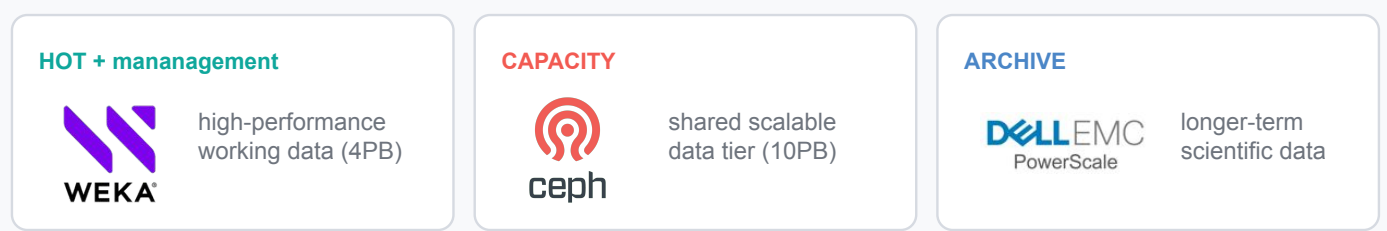
SUNRISE — UKAEA's AI/HPC platform for fusion

heterogeneous compute • high-performance data • built to federate

COMPUTE



DATA



One platform for simulation + AI + data-intensive R&D

52.8 PF FP64 peak
672 GPUs

384 CPU sockets

67	Sunrise - PowerEdge XE9785L, AMD EPYC 9575F 64C 3.3GHz, AMD Instinct MI355X, Infiniband NDR400, Rocky Linux, DELL UKAEA - University of Cambridge United Kingdom	156,672	34.77	46.16
68	Zenith - PowerEdge XE9785L, AMD EPYC 9575F 64C 3.3GHz, AMD Instinct MI355X, Infiniband NDR400, Rocky Linux, DELL University of Cambridge United Kingdom	156,672	34.77	46.16

SUNRISE #67 • ZENITH #68

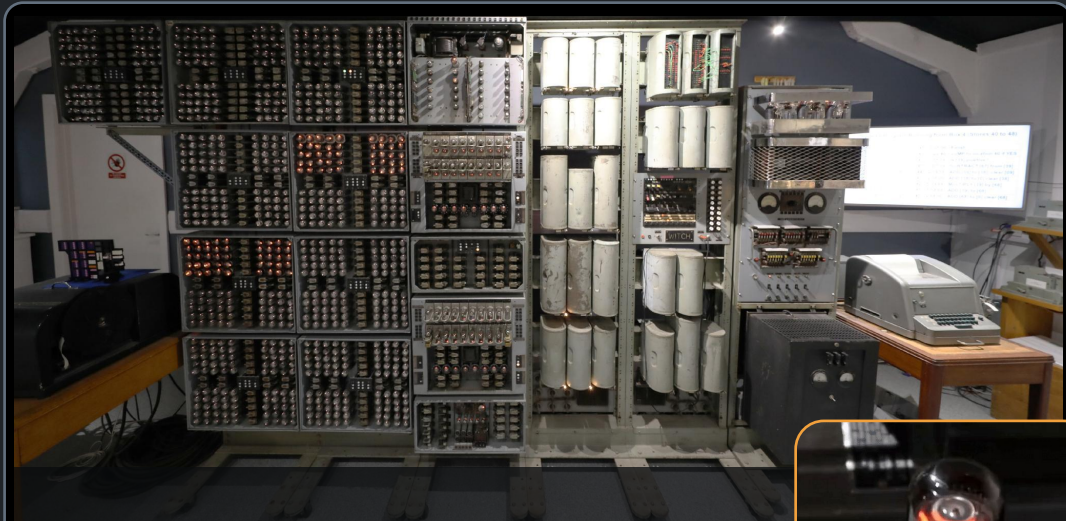
TOP500 • June 2026

SUNRISE + ZENITH signing

sister systems • Cambridge

A tangible UK capability today — and a natural building block for a federated US–UK digital estate.

From WITCH to SUNRISE — 75 years of scientific computing



1951 · AERE HARWELL
HARWELL WITCH
Dekatrons · relays · valve electronics

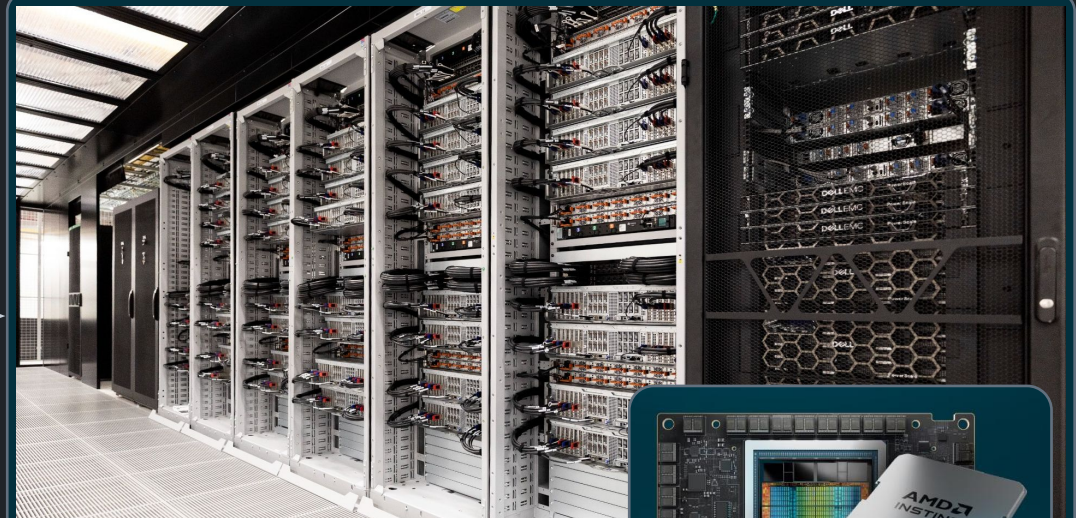
Still working today at TNMOC



Dekatron
tubes

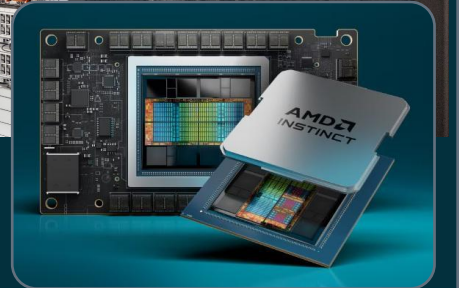
75
YEARS

same mission



2026 · UKAEA
SUNRISE
672 × AMD MI355X · AI + HPC for fusion

≈105.7 PFLOP/s FP32 peak



AMD MI355X
GPUs

1 SECOND
on SUNRISE



10–20 BILLION YEARS*
of WITCH computing



AGE OF EARTH
same order as the age of the Universe

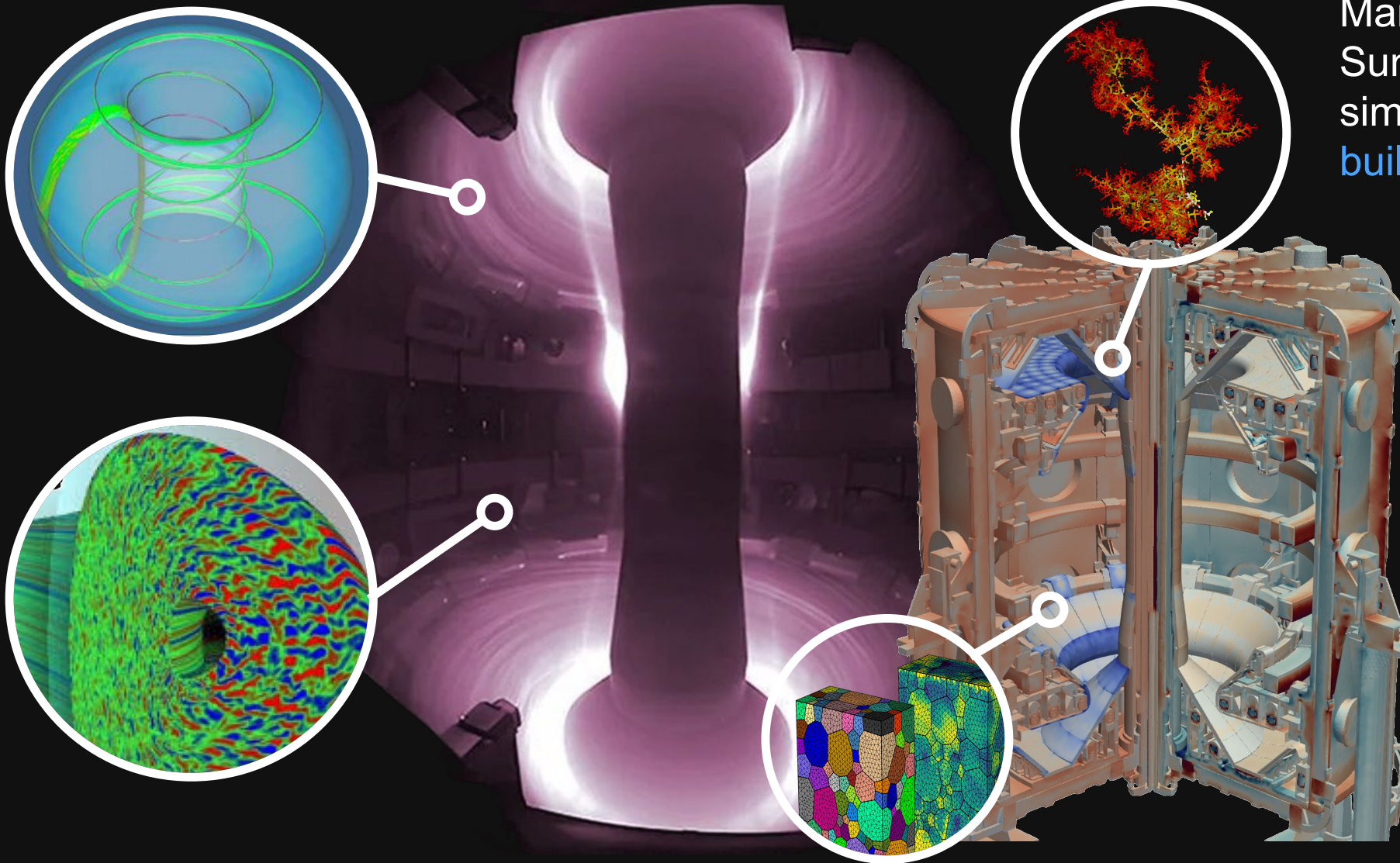
Same ambition. Different era. Extraordinary machines extending what scientists and engineers can do.

*Illustrative peak-arithmetic comparison, not a workload-equivalence benchmark. Uses 672 × MI355X at 157.3 TFLOP/s FP32 and historical WITCH arithmetic timings of several seconds per operation.

SUNRISE Workflows exemplifying convergence of HPC & AI

Many domains can make use of Sunrise to perform classical simulation work, with a view to building AI models....

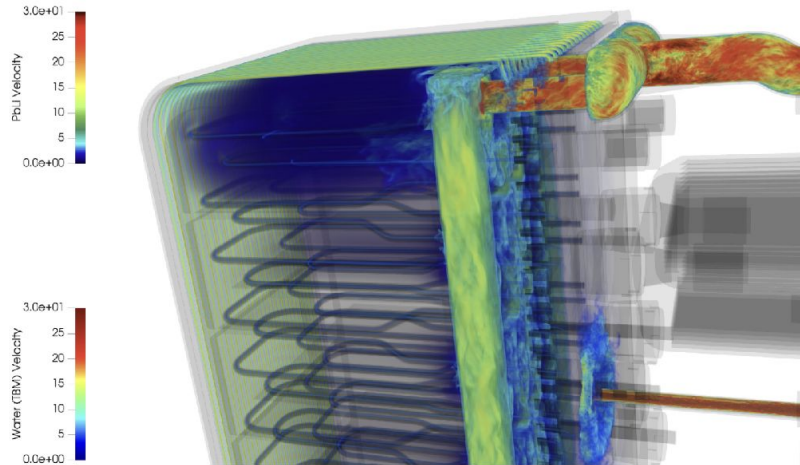
Gyrokinetics
Edge Turbulence
Fast Ions
CFD & MHD
Radiation Transport
Finite Element
Atomic Cascades
Dislocation Dynamics
Training sets
Foundation Models



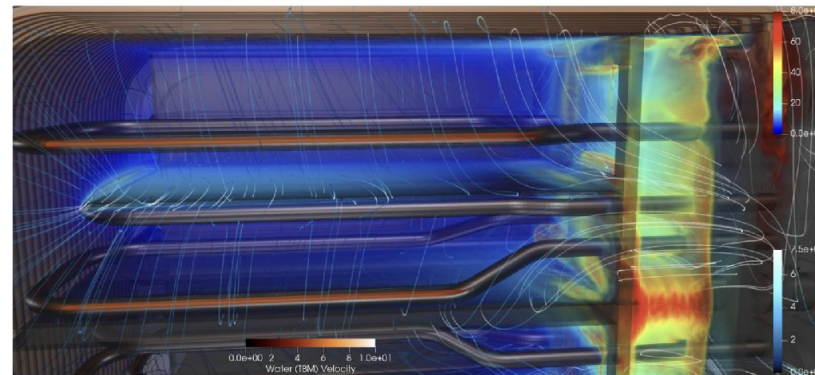
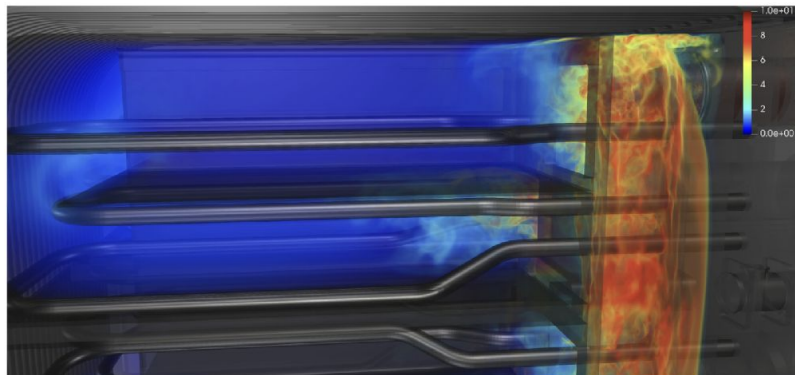
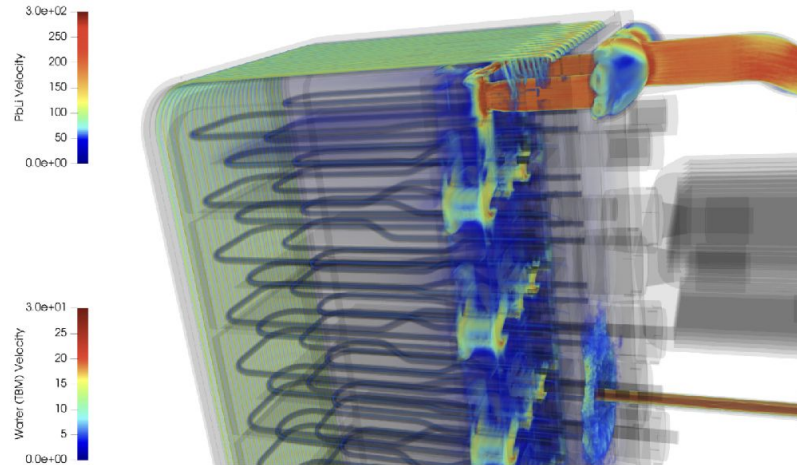
Exascale sets the fidelity bar - NekRS + OpenMC



MHD off



MHD on



EXASCALE REFERENCE MODEL

The practical fidelity ceiling

2.06T
GRID POINTS

6.9T
DOF

≈ 1,000
SECONDS / RUN

Practical for design. UQ needs reduced models.

COUPLED PHYSICS IN THIS RUN

- 3D CFD / thermal hydraulics in PbLi + water
- Conjugate heat transfer + buoyancy
- Liquid-metal MHD
- Neutronics + shielding + nuclear heating
- Tritium breeding source + water activation

The reference will show where strong coupling changes the answer and where simpler models remain trustworthy.

NOT IN THIS RUN

Tritium transport, FIV and irradiation-driven stress/strain.

Exascale reference

Coupling sensitivity

ROMs + surrogates

Engineering + UQ

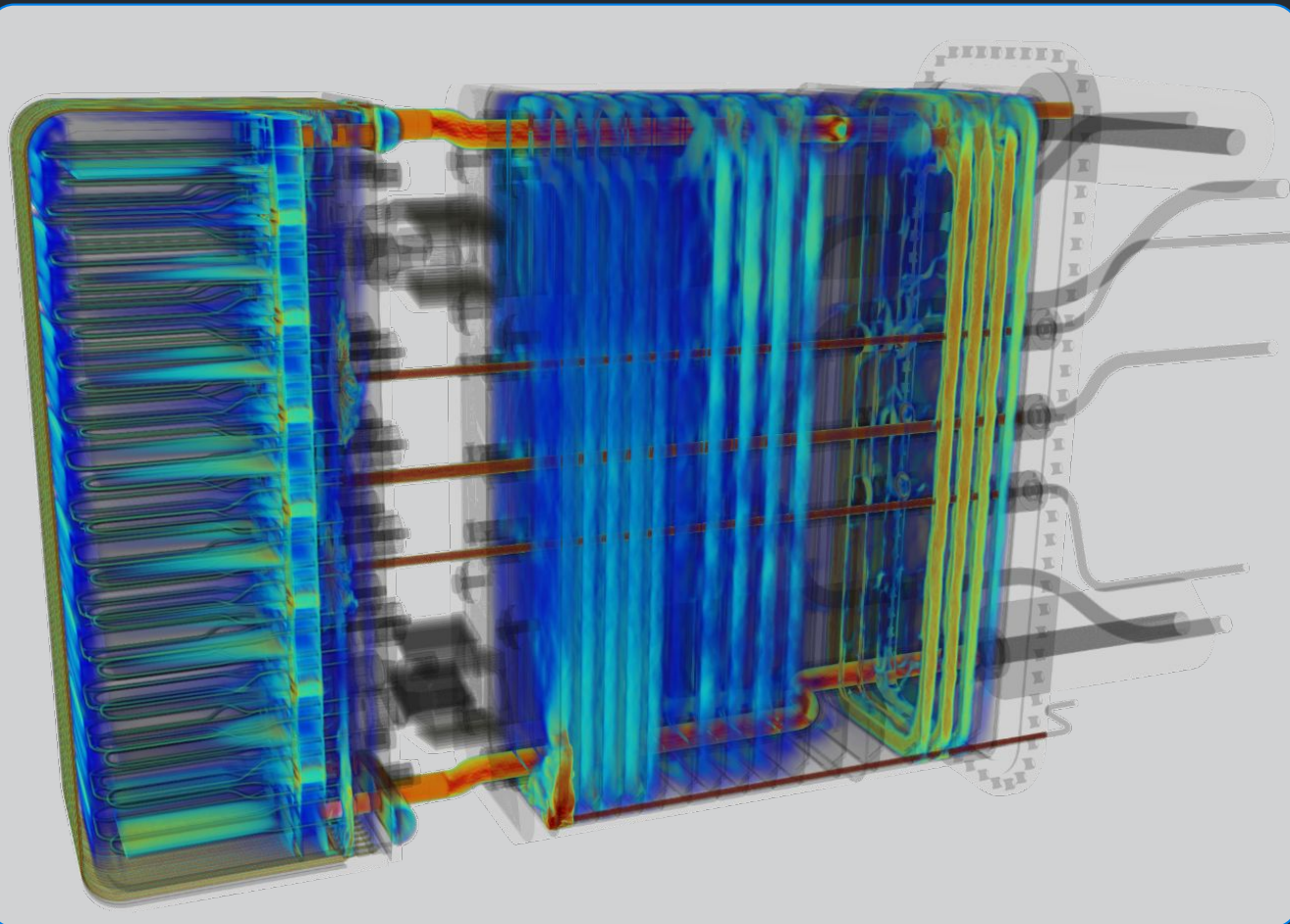
1

2

3

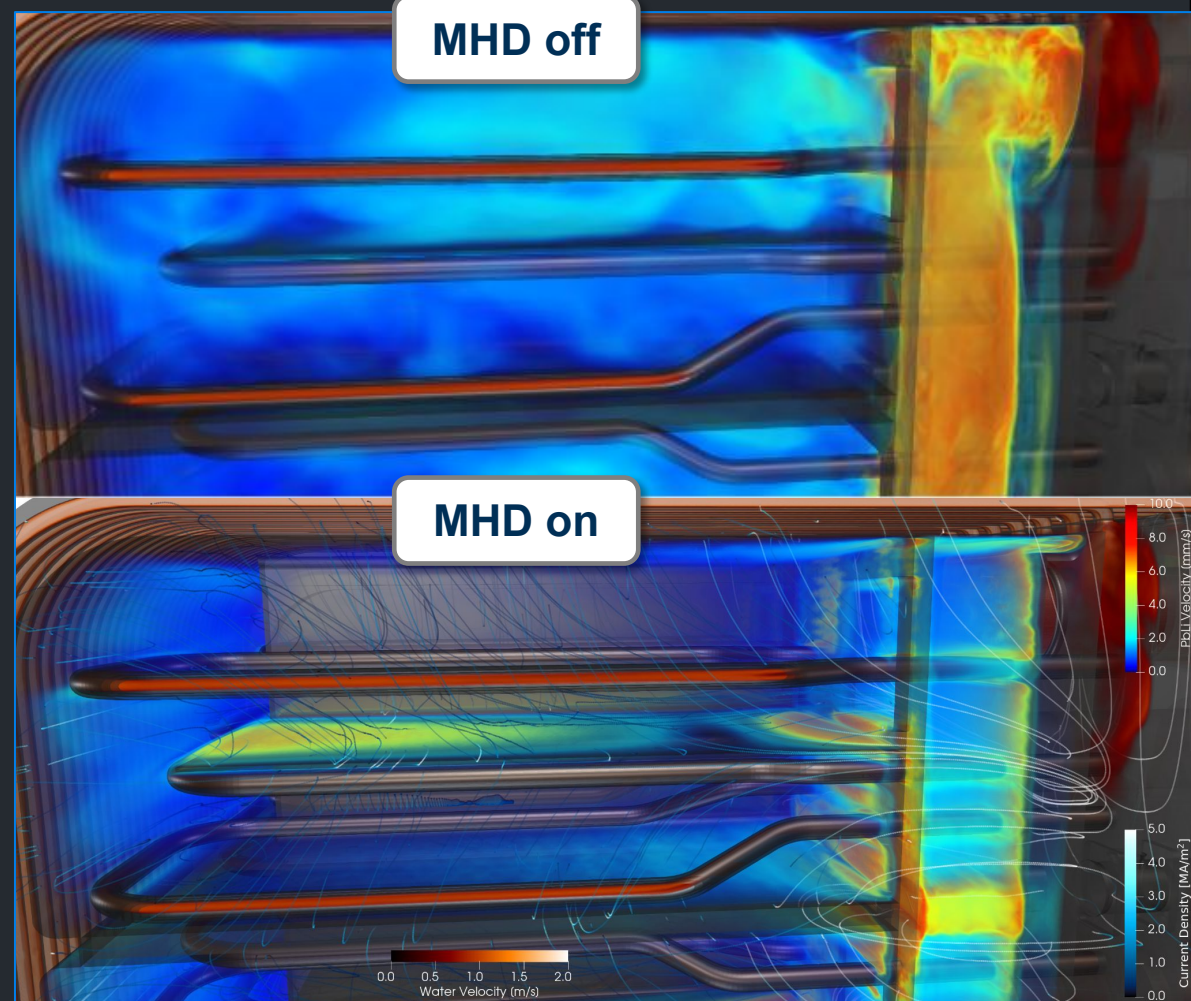
4

Whole component simulation - Paradigm Evolution



Three working fluids simulated at the same time; helium gas, lithium-lead eutectic with MHD and high pressure water. Run on two the world's largest supercomputers, using some 10k GPUs

Being able to simulate a whole component allows emergent phenomena to be evaluated and designs updated to account for them.

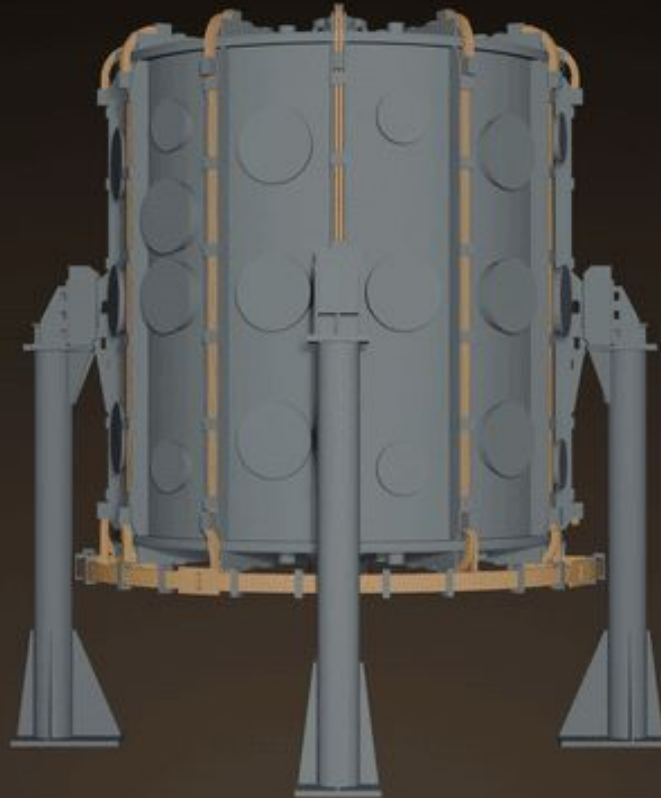


AI across scales — from fusion neutrons to structural lifetime

Neutron transport • atomistic damage • surrogate material response • structural FEM

Multiscale simulations of a full fusion tokamak: neutron transport, atomistic, and finite element simulations

L. Reali, A.K. Gray, M. Boleininger, W.M.E. Ellis, A. Davis, J. Buchanan, S.L. Dudarev



● self-interstitial defects

● vacancy lattice sites

Winger-Seitz defect identification

Illustrative low-dose cascade regime

THE MULTISCALE BRIDGE

1

Fusion neutron field

14.1 MeV D-T neutrons → energy-resolved transport

2

Atomistic damage

PKA cascades → vacancies, interstitials and swelling

3

Surrogate material response

local damage state → eigenstrain / stress-strain response

4

Structural evolution

FE cells → evolving stress, strain and component lifetime

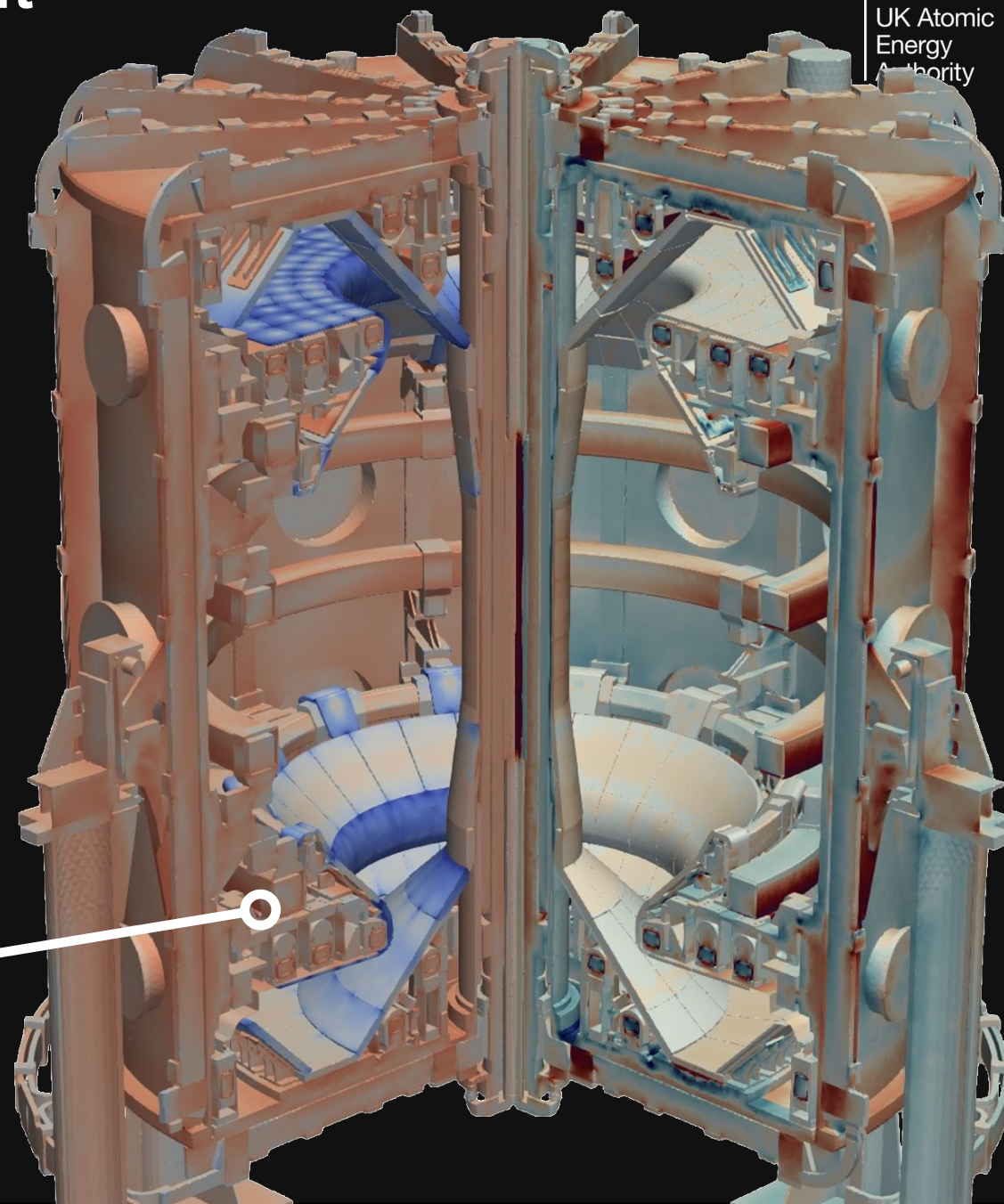
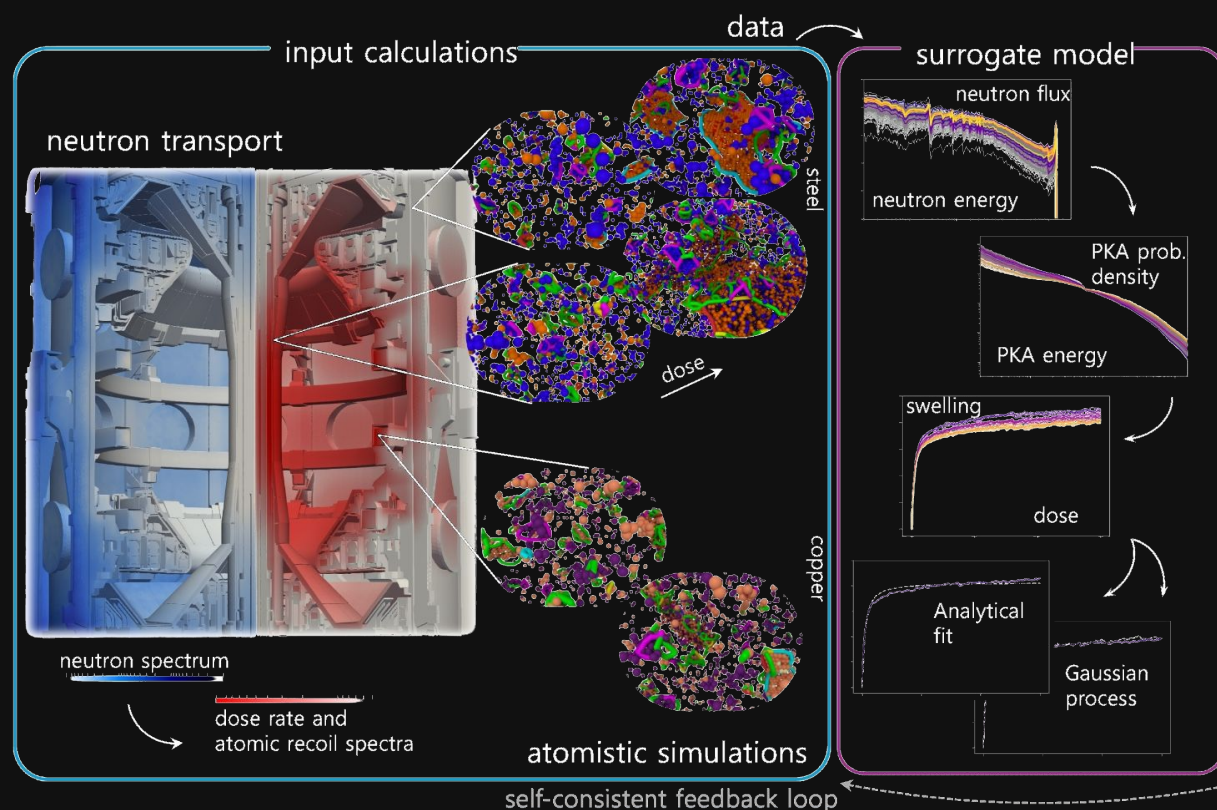
Surrogate = bridge between scales

Train on expensive atomistic physics once; deploy the learned local material response across ~160 million FE elements.

Aim: predict how irradiation-driven microstructural damage becomes engineering stress and strain.

Whole device simulation - Paradigm Shift

By combining traditional HPC enabled modelling with simulation trained Gaussian processes, efficient multiscale modelling can be brought to bear on tough challenges with modest computational requirements



TokaMind – one model, many plasma tasks

- MULTI-MODAL
- MULTI-PURPOSE
- SCHEMA-FLEXIBLE
- BENCHMARKED
- OPEN

Pretrain broadly on heterogeneous tokamak data → adapt efficiently to many downstream tasks

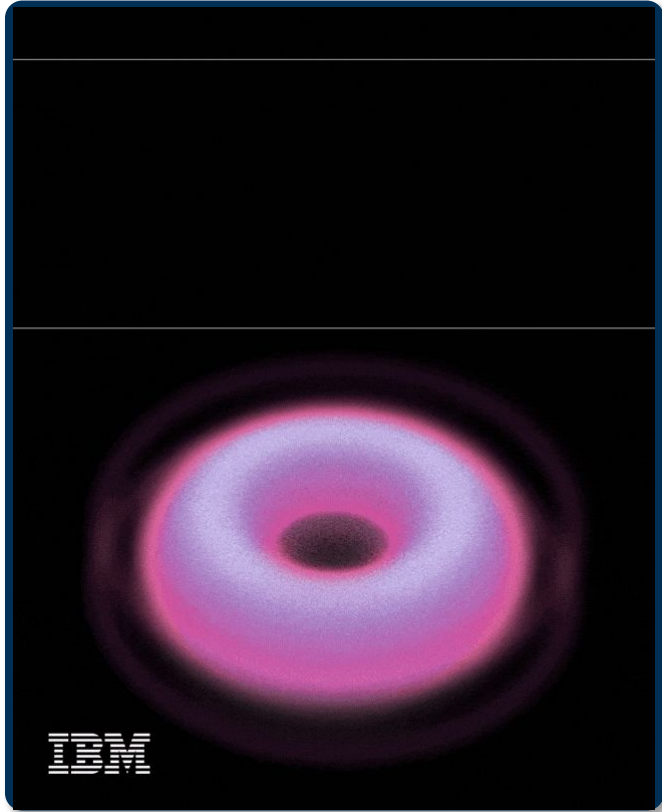
- 1

Different diagnostics → tokens
time series • profiles • video • actuators
- 2

Add meaning & context
signal • modality • role • recency
- 3

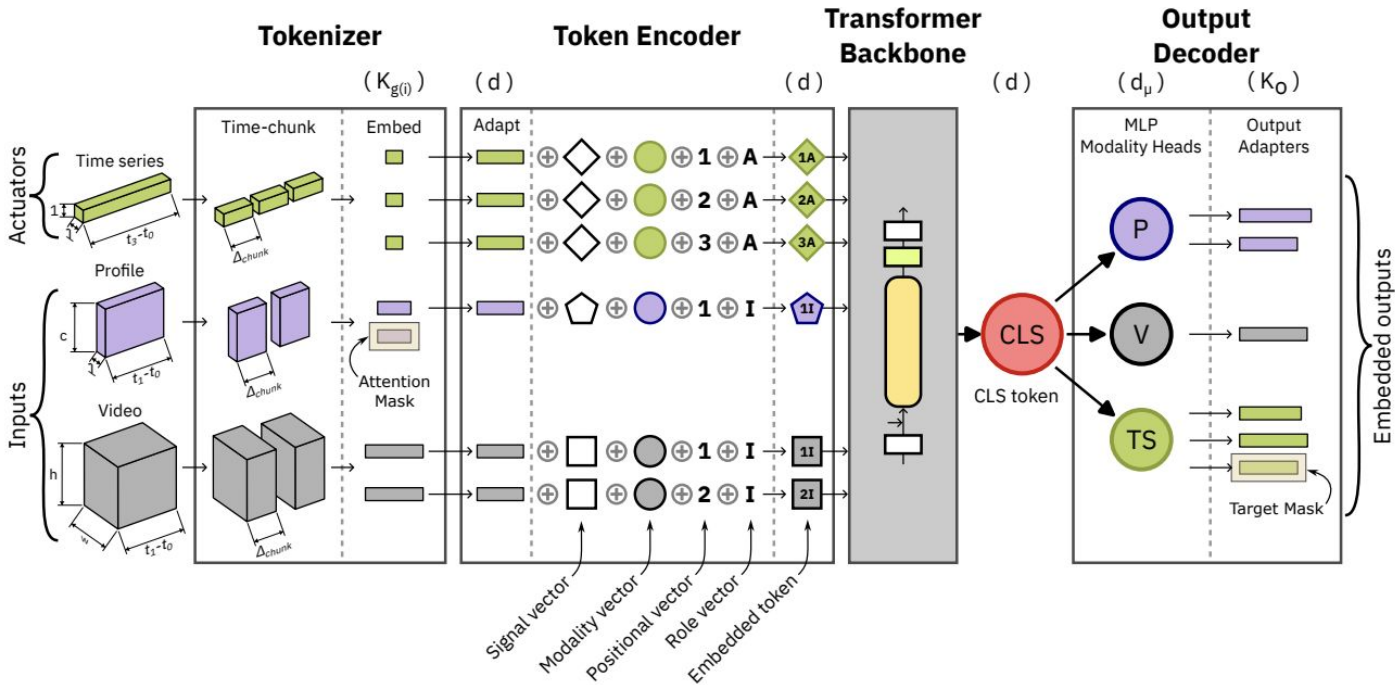
Learn a shared plasma state
Transformer links what is happening together
- 4

Adapt to many outputs
new tasks via small heads / adapters



FOUNDATION MODEL ≠ ONE TASK

One reusable pretrained representation
→ many reconstruction & forecasting tasks



US-UK opportunity

MAST / MAST-U ↔ NSTX-U: test cross-device transfer

→ candidate experimental-data backbone / building block for IGNOS?

11,573

MAST discharges

39

signals

3

modalities

13 / 14

benchmark tasks beat best baseline

52 / 57

output signals beat best baseline

OPEN

code + model weights

Hugging Face

GyroSwin – nonlinear turbulence at AI speed

Keep the nonlinear physics. Lose the computational cost

5D

NONLINEAR

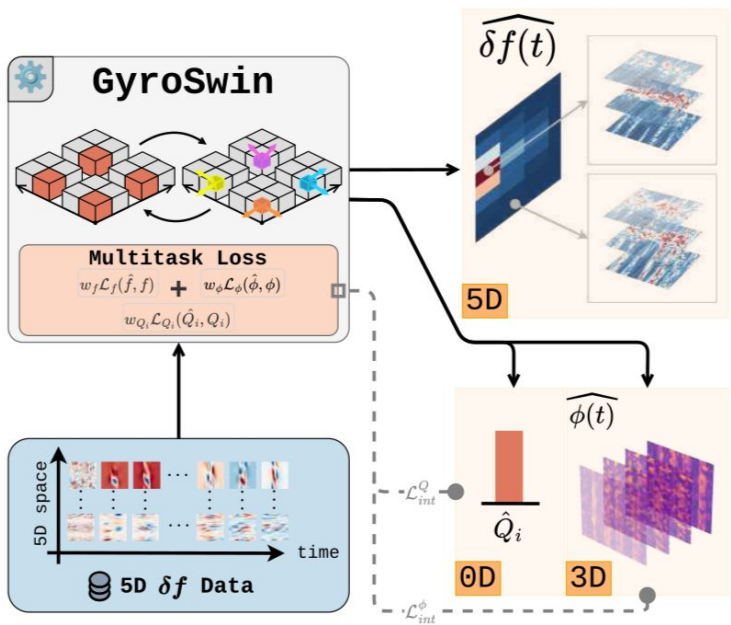
PHYSICS-AWARE

SCALABLE

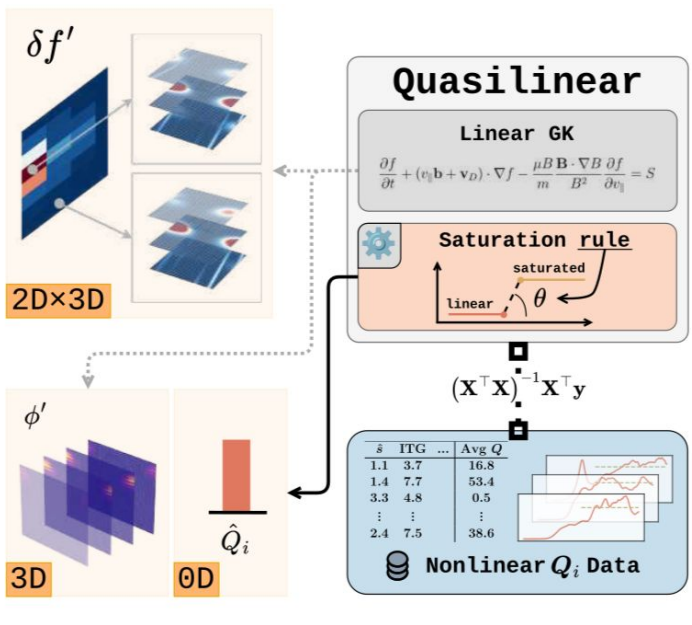
VERIFIABLE

Train on expensive nonlinear gyrokinetics → deploy as a fast surrogate in design & optimisation loops

GyroSwin: learn the full nonlinear 5D state



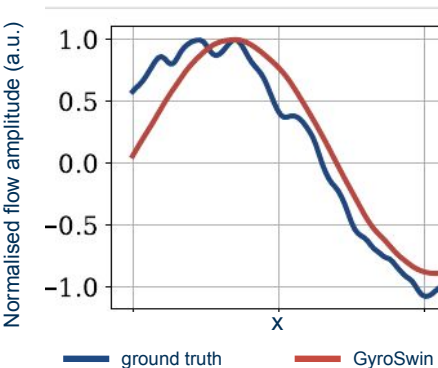
Traditional shortcut: linear physics + saturation rule



≈ 1,000×

faster than nonlinear GKW

while evolving the full 5D turbulence state



Emergent physics survives

zonal-flow profile reproduced on an unseen OOD simulation

SURROGATE ≠ REDUCED PHYSICS

learn the expensive nonlinear solution rather than approximate it with a fitted saturation rule

FULL PHYSICS → FAST SURROGATE → PREDICTIVE CAPABILITY

US–UK opportunity jointly validate & generalise high-fidelity surrogates across machines and codes → integrated modelling • optimisation • digital twins

5D → 5D
full turbulence state

~1,000×
faster than GKW

>100
stable rollout steps

OOD
generalisation tested

~1B
parameter scaling

ZONAL FLOWS
emergent nonlinear physics

Collaboration in 2026 – communication



STELLAR-AI  **PPPL**
PRINCETON
PLASMA PHYSICS
LABORATORY

 **Genesis Mission + IRI, AmSC etc.**



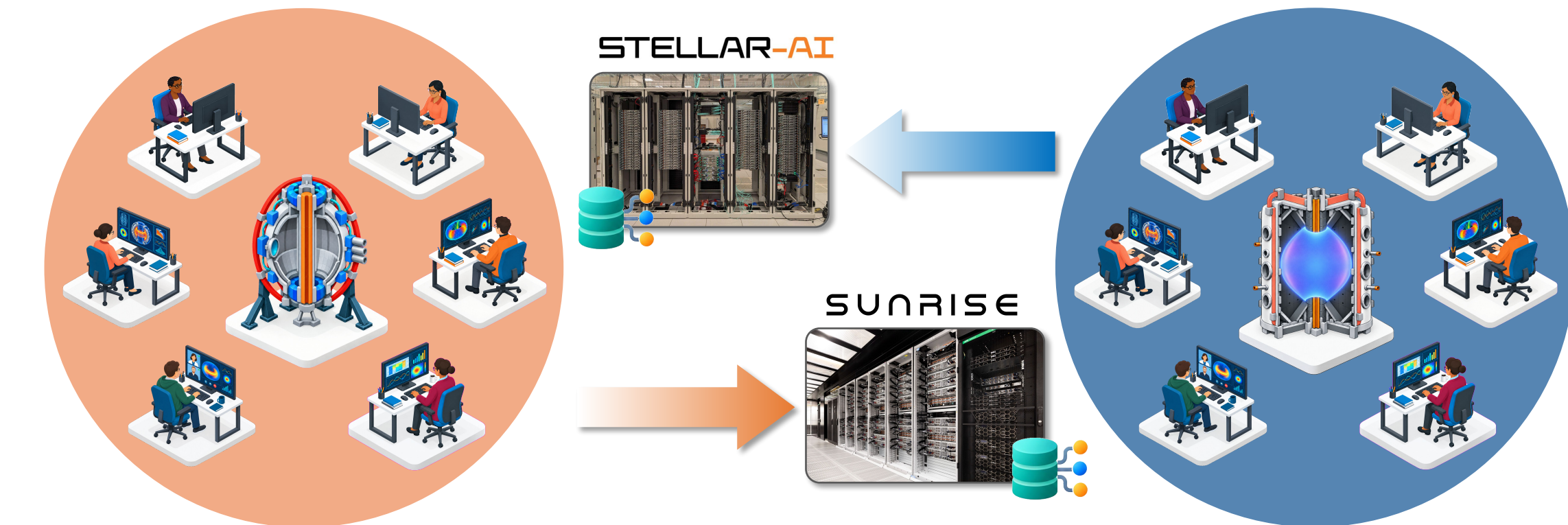
**Video
Conferencing**



SUNRISE

AIRR, DRI, NCR etc. etc.

BUT: Blockers – broader digital estate/tools – data & compute



STELLAR-AI

PPPL
PRINCETON
PLASMA PHYSICS
LABORATORY

Genesis Mission + IRI, AmSC etc.

Federated
Compute

FAIR
DATA

SUNRISE

AIRR, DRI, NCR etc. etc.

1. The “Fabric” – supporting a shared, digital estate...

Shared Digital Fabric

APIs • semantics • identity & policy • workflows • portability

US ECOSYSTEM AI FOR FUSION DISCOVERY & DELIVERY



STELLAR-AI 
 Genesis Mission + IRI, AmSC etc.

DATA
MODELS
TOOLS
WORKFLOWS

Metadata & Semantics
vocabularies • schemas • meaning

Identity, Access & Policy
trusted access • agreed policy

**Workflow
Orchestration**

**SHARED /
DIGITAL FABRIC**

**Portable
Environments**
reproducible stacks • containers

Common APIs
reusable services • integration

common services • aligned policy
portable workflows • reusable components

UK ECOSYSTEM AI FOR FUSION DISCOVERY & DELIVERY



SUNRISE 
AIRR, DRI, NCR etc. etc.

DATA
MODELS
TOOLS
WORKFLOWS

Make the sum greater than the parts



Faster
Discovery



More Ambitious
Science



Reuse, Not
Rebuild



Trusted &
Reproducible



Economy
of Scale



Complementarity

2. Compute – co-designing digital twin infrastructure...

Federated AI/HPC Compute Platform

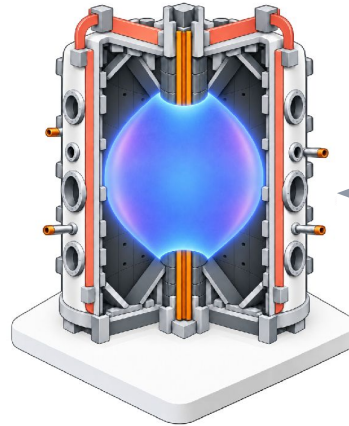
shared capacity • HPC simulation • AI training • inference & analytics

US ECOSYSTEM AI FOR FUSION DISCOVERY & DELIVERY



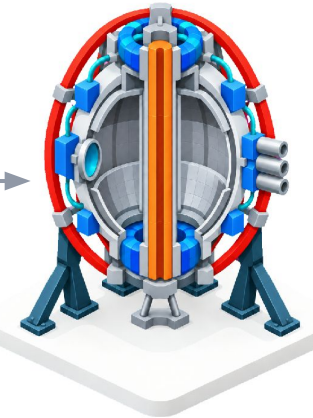
STELLAR-AI **PPPL**
Genesis Mission + IRI, AmSC etc.

DATA
MODELS
TOOLS
WORKFLOWS



shared where beneficial • protected where necessary

- Federated Resource Sharing**
shared capacity • workload placement
- HPC Simulation**
large-scale modelling • ensembles
- AI Training**
foundation models • scientific ML
- Inference & Advanced Analytics**
inference • big-data analytics



DATA
MODELS
TOOLS
WORKFLOWS

UK ECOSYSTEM AI FOR FUSION DISCOVERY & DELIVERY

SUNRISE
AIRR, DRI, NCR etc. etc.

enabling • digital twins • foundation models • AI surrogates • actionable predictive capability

Make the sum greater than the parts



Faster
Discovery



More Ambitious
Science



Reuse, Not
Rebuild



Trusted &
Reproducible



Economy
of Scale



Complementarity

3. Data – Findable, Accessible, Interoperable, Reusable

CATALOGUE & DISCOVERY • CURATED DATA PRODUCTS • PROVENANCE & LINEAGE

Catalogue & Discovery

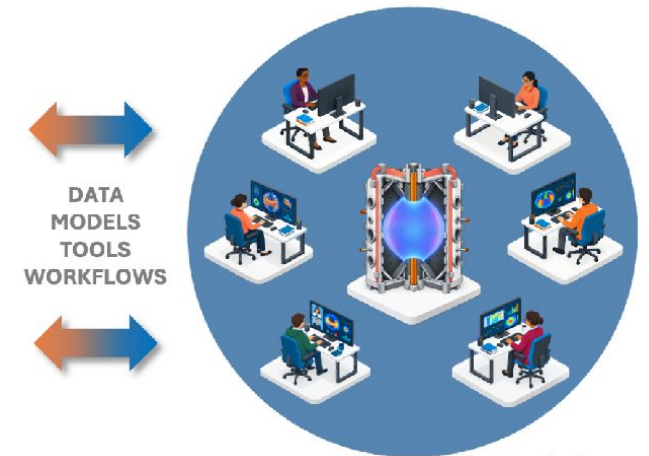
search • identifiers • metadata

Curated Data Products

harmonised schemas • units • quality

UK ECOSYSTEM

AI FOR FUSION DISCOVERY & DELIVERY



SUNRISE

AIRR, DRI, NCR etc. etc.

DATA
MODELS
TOOLS
WORKFLOWS

DATA
MODELS
TOOLS
WORKFLOWS

FUSION DATA
COMMONS

Provenance & Lineage

source • version • derivation

digital thread across experiments, data products, models and evidence



Make the sum greater than the parts

US ECOSYSTEM

AI FOR FUSION DISCOVERY & DELIVERY



STELLAR-AI PPPL

Genesis Mission + IRI, AmSC etc.



shared where beneficial • protected where necessary



Faster
Discovery



More Ambitious
Science



Reuse, Not
Rebuild



Trusted &
Reproducible



Economy
of Scale



Complementarity

Summary – Areas of Strategic Collaboration



SUNRISE

platform interoperability • cross-platform workflows • AI tools



Digital Twins for Fusion

predictive modelling • operational support • virtual testing



AI for Fusion Power Plant Design

AI-enabled engineering optimisation • integration • reliability



Advanced Computing Innovation

HPC • accelerators • novel architectures • scientific workflow infrastructure



VVUQ

automated simulation workflows • quantified experimental and model uncertainty



Physics Model Validation & AI Surrogates

validated surrogates for turbulence, transport and macro-stability



Fusion Foundation Models

large-scale AI across plasma, engineering, materials and operations



Industry Engagement & Development

workshops • staff exchanges • student programmes • collaborative research initiatives

Architecting the digital foundation for the commercial fusion era - Sunrise and its successor system will give fusion the digital backbone it needs, HPC, AI and Advanced Computing will accelerate the path to fusion.

andrew.davis@ukaea.uk We'd love to collaborate with you on Sunrise, reach out and see what we can do together.