9th Asia-Pacific Conference on Plasma Physics, 21-26 Sep, 2025 at Fukuoka

AI and data solutions for experiment design and control

Adriano Agnello[1], Alasdair Ross[1], Aran Garrod[1], Pedro O. Cavestany[1], George K. Holt[1], Nicola C. Amorisco[2], Nam Nguyen[1], Matthijs Mars[3], Kamran Pentland[2], Abigail Keats[1], Charles Vincent[2], Graham McArdle[2], Stanislas Pamela[2]

[1] STFC Hartree Centre, Sci-Tech Daresbury, Warrington (United Kingdom),

[2] United Kingdom Atomic Energy Authority, Culham Science Centre, Abingdon (UK)

e-mail (speaker): adriano.agnello@stfc.ac.uk

The efficient sampling of simulations has applications throughout fusion research. For example, large synthetic datasets are needed to train emulators for real-time control of tokamak plasmas, while the efficient traversal of parameter spaces governing simulations is necessary to find optimal configurations or to fit experiment data with model predictions. Considerable effort is under way to assemble interoperable multi-machine experiment databases [1,2], but those do not necessarily span the whole range of parameters relevant to upcoming experiments or new machines. Challenges arise from the dimensionality of the parameter space, and numerical instability or computational cost of simulations.

The *MIDAS/HiveMind* suite [3] provides Monte Carlo samplers and optimisers, also via asynchronous multi-agent implementations that are easily distributable on HPC and cloud systems. *HiveMind* has been used to build a large synthetic set of Grad-Shafranov equilibria to train virtual circuit emulators, for use in plasma shape control on the MAST-U machine [4,5]. An application of the asynchronous "multi-hopper" is the optimized design of engineering components, e.g., in the divertor region. The same algorithms can also be used to tune hyperparameters of machine learning models, and to sample expensive-evaluation functions or experiment setups while minimizing overall uncertainties.

Another example of synthetic datasets arises in reinforcement learning (RL) for plasma shape control, whose agents are trained on several simulated or historic experimental instances of plasma dynamics. RL based on experiment databases has been trained on machines with a long history of campaigns [6,7], but simulation-based RL [e.g. 8,9] is needed for new or upcoming machines. The *FreeGSNKE-RL* codebase [10] seamlessly integrates the RL environments with the *FreeGSNKE* codebase to simulate plasma dynamics [11,12]. In this context, the data requirements depend on the choices of observation and action spaces, as well as on the chosen agent architecture and training regime.

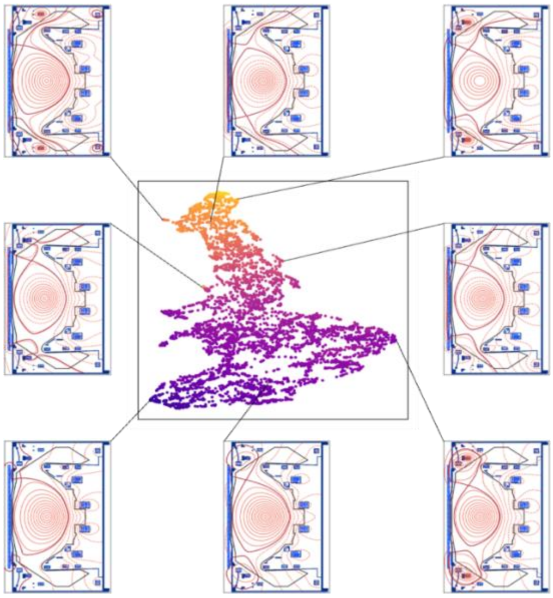

Figure 1: UMAP embedding of a synthetic library of 5×10^6 MAST-U equilibria, containing both limited and diverted configurations, colour-coded by the radial coordinate of the lower divertor strike-point.

References:

[1] Jackson, S., Khan, S, Cummings, N, et al. (2024), doi: 10.1016/j.softx.2024.101869
[2] https://crea-psfc.github.io/open-fair-fusion/
[3] https://gitlab.stfc.ac.uk/optimising-expensive-optimisation/HiveMind **[4]** Agnello, A., et al. (2024), doi:10.1063/5.0187822
[5] Cavestany, P. et al. (2025, IEEE CCTA2025 subm.)
[6] Seo, J., et al. (2021), doi: 10.1088/1741-4326/ac121b
[7] Zhang, Y. C., Wang, S., Yuan, Q. P., et al. (2024),. doi: 10.1088/1361-6587/ad3749
[8] Degrave, J., et al (2022),,doi: 10.1038/s41586-021-04301-9
[9] Tracey, B. D., et al (2024),doi: 10.1016/j.fusengdes.2024.114161
[10] gitlab.stfc.ac.uk/farscape-ws3/freegsnkerl
[11] Amorisco, N.C., et al. (2024), doi: 10.1063/5.0188467
[12] https://github.com/ FusionComputingLab/freegsnke/tree/main/freegsnke