EUROfusion

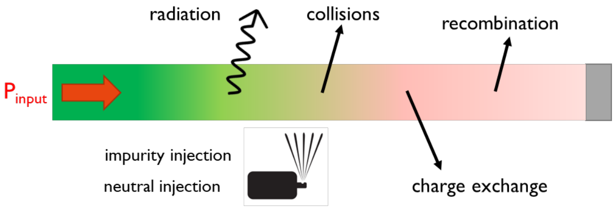
Multivariable feedback control of radiative loss-processes using multi-spectral imaging

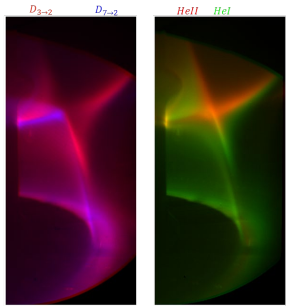
M. van Berkel (P.I.)[1], A. Perek[2], J.T.W. Koenders[1,3], E. Huett[2], C. Galperti[2], B.P. Duval[2], O. Février[2], T.A. Wijkamp[1,3], I.G.J. Classen[1], M. O'Mullane[4], J. Citrin[1], E. Westerhof[1], C. Theiler[2] and the TCV Team[*]
Supported by: K. Verhaegh, B. Dudson, L. van Leeuwen, G.L. Derks, J. Caballero, L. Martinelli
[1]DIFFER – Dutch Institute for Fundamental Energy Research, Eindhoven, The Netherlands
[2]École Polytechnique Fédérale de Lausanne (EPFL), Swiss Plasma Center (SPC), Lausanne, Switzerland
[3]Eindhoven University of Technology, Control Systems Technology, Eindhoven, The Netherlands
[4]University of Strathclyde, Glasgow, United Kingdom
[*]See author list of S. Coda et al. 2019 Nucl. Fusion 59 112023

DIFFER EPFL University of Strathclyde Glasgow TU/e

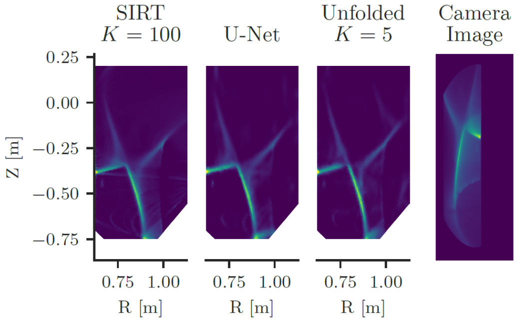
This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Project overview (and status)

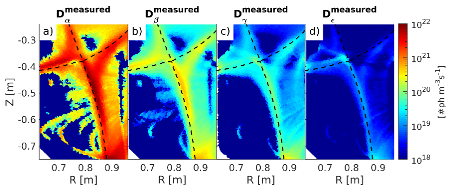




Scaled emission line intensities encoded into red, green and blue channels. MANTIS on the TCV tokamak; #63974, t = 1.32s



line-ratios



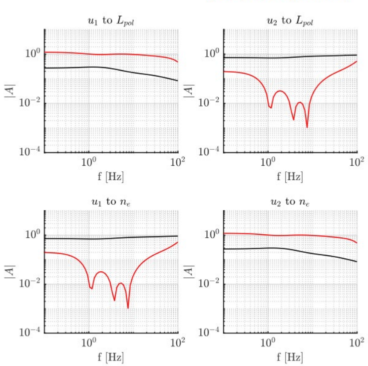
Bayesian interf.

$n_e, T_e, n_o, I_{rate}, R_{rate}, CX_{rate}, P_{exc}, P_{rec}$

feedback + feedforward

digital control system (SCD)

real-time signal processing



MIMO-decoupling

demonstration September 2023
(Project end June 2024)

Overview complete project

Main objective: MIMO control of the divertor state using multiple MANTIS camera’s

Sub-objectives

- **Real-time (millisecond range) tomographic reconstruction of MANTIS images.**
  - ✓ Is achieved using a machine learning accelerated approach (2 ms)
  - ✓ Awaiting a new GPU for implementation (<2 ms) -> new GPU tested and largely integrated
  - ✓ Aim for an experimental demonstration before 2023 (not on tokamak)
- **Real-time inference of recombination, ionization and impurity radiation power losses**
  - ✓ Basic version for inference of ionization, recombination, and divertor $T_{electron,}n_{electron,}$ n_0 from filtered camera images [1]
  - ✓ Further development necessary to improve Bayesian inference and validate results, e.g., ionization
  - ✓ Aim for an experimental demonstration before 2023 (not on tokamak)
- **Control-oriented modelling for MIMO exhaust control**
  - ✓ 1 dimensional dynamic SOL Model DIV1D was benchmarked against SOLPS-ITER in steady-state [2]
  - ❑ Ongoing benchmark against dynamic experiments
- **MIMO system identification + feedback control (and integration in SCD)**
  - ✓ MIMO sys.id. and control of line-averaged electron density and NII emission front position [3]
  - ❑ Repeat of above with real-time inferred processes (ionization, etc.) from MANTIS camera’s -> September 2023

Recommendations from committee

- It should be interesting to indicate the expected ways to obtain a real-time control on 1 ms time scale and the possible issues encountered to reach it before the end of the project -> detail explanation next slide
- Precise the possible issues to apply MiMO real-time control on different devices such as W7X, JT60-SA, ITER, ... -> very valid point but a deep and complicated question where a clear distinction needs to made between fusion producing devices (ITER, DEMO) and W7X and JT60SA for the exhaust, preparing a publication on this topic

Many thanks for the input!

Real-time control time scales

1 ms is and NEVER has been a target for the feedback control scheme

TV analogy: *refresh rate vs. frame rate*:

TV's refresh rate 50 Hz (maximum frames to be displayed by TV), blue-ray 25 Hz how many new frames are shows

Overall control system performance is determined by slowest time-scales and loop-delay:

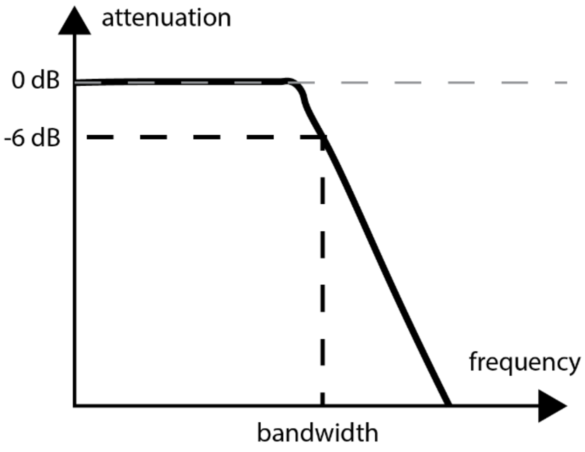
- ➢ 800 Hz inherent frame rate (MANTIS) camera's
  (Camera's 1.25 ms delay, 2.6 ms processing, 0.5 ms communication)
- ➢ 50 Hz bandwidth of the gas-valve at TCV (deadtime: signal to action ~ 1 ms)
- ➢ ~100 Hz observable plasma dynamics due to signal-to-noise ratio (plasma dynamics)

❖ Rule of thumb: sampling rate (1 ms) X 8~10 faster than bandwidth (100 Hz)

loop-delay τ_{loop} determines bandwidth ~ $\left(2\pi\tau_{loop}\right)^{-1}$

Real-time control time scales (cont'd)

- Rule of thumb for bandwidth (real-time control time-scale):
  - sampling rate X 8~10 faster than bandwidth
  - bandwidth (BW) ~ $(2\pi\tau_{loop})^{-1}$



reference -> measured

Current set-up **theoretical** bandwidth:

- $\tau_{loop} = \mathrm{CS} + \mathrm{IP} + \mathrm{COM} + \mathrm{GV} \sim \text{bandwidth}$
- τ_{loop} = 1.25 + 2.6 + 0.5 + 1 ~ 30 Hz
- τ_{loop} = 1.25 + 0/1 + 0.5 + 1 ~ 57.9 Hz (42 Hz)
- τ_{loop} = 1.25 + 2.6 + 0.5 + 0 ~ 37 Hz (logical first step)

Practical bandwidths

- SISO exhaust controllers TCV bandwidth ~ 8 Hz -> MIMO normally slower, we aim at ~ 10 Hz
- Other devices even slower (due to gas-pipes): AUG ~ 1 Hz, ITER > 0.1 Hz

τ_{loop} can only be broken by different control, e.g., model predictive control!

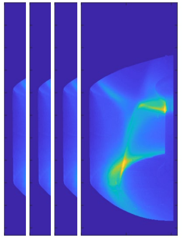
ENR WPs overview and progess

Main goal: setting up for MIMO control with different MANTIS camera's

Progress	Project
P1 (largely finished) development only in terms of P3	MANTIS development to determine loss-processes in 2D
P2 (continued evolution, especially with respect with control)	Detachment analysis, scenario selection, setting control requirements
P3 (largely finished)	Conversion from off-line to real-time camera analysis (incl. machine learning)
P4 (largely finished)	MIMO system identification
P5 (ongoing)	Dynamic modelling for MIMO-control
P6 (ongoing)	MIMO feed-back control (and integration)

WP3: Offline to real-time analysis conversion

Plasma emission



~5 CPUs per frame

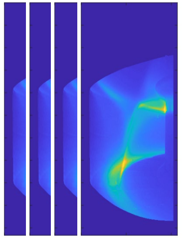
Tomographic inversions

~4000 CPUs per time step

Bayesian Plasma parameter inference [1]

I_{rate}, R_{rate}

Plasma emission



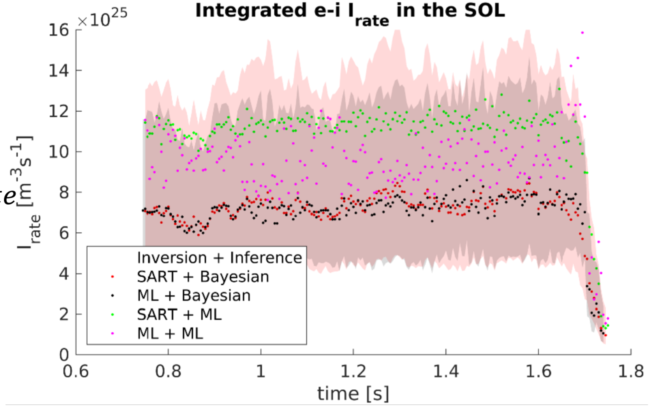
Machine learning for tomographic inversions

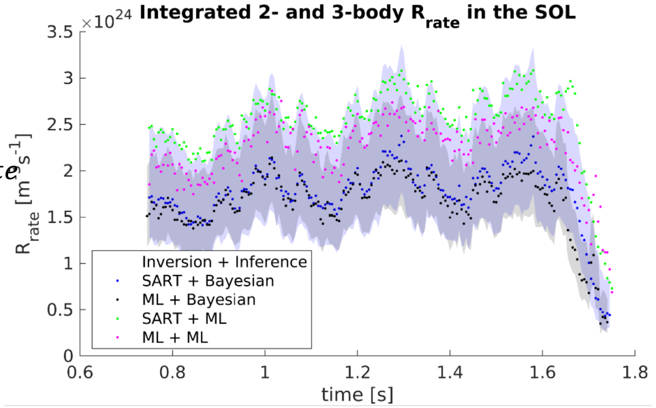
Machine learning for plasma parameter inference

I_{rate}, R_{rate}

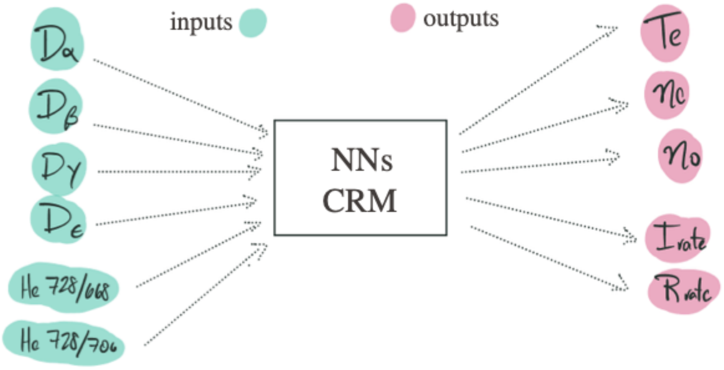
~0.002 CPUs per frame

~0.03 CPUs per time step





Offline to real-time analysis conversion: Parameter inference

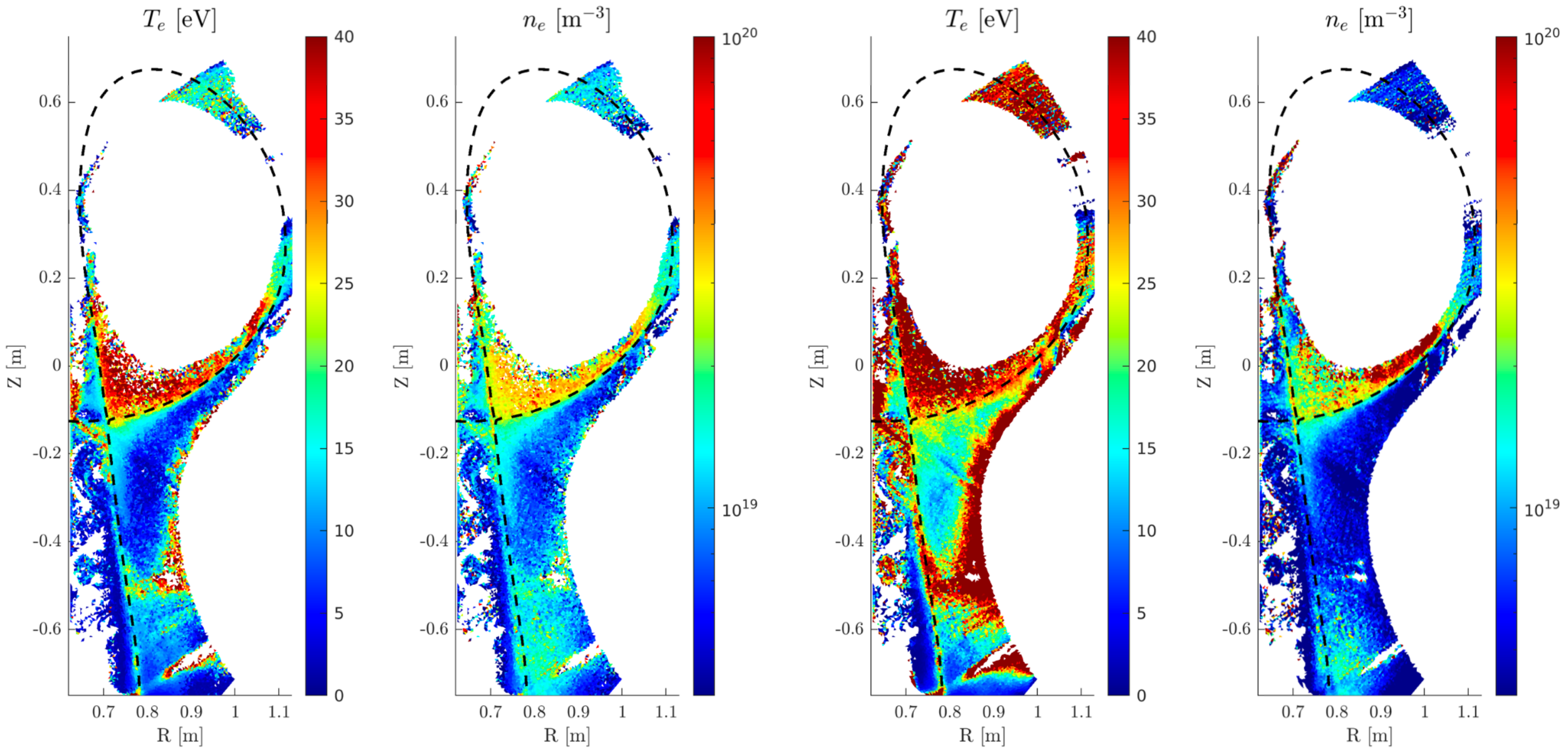


- This work focuses on majority Deuterium plasmas, since power exhaust its only relevant in Deuterium plasmas. Helium its also taken in consideration, since it is present at TCV and its present in reactor grade plasmas
- For **Hydrogen** the **ADAS CRM** is used, and for **Helium** the **Goto CRM**. Given the research question, the objective is to develop an inverse mapping of these models
- This work uses neural networks (NN) since they are universal approximators capable of learning complex non linear relationships **provided they are supplied enough representative data**
- A relevant parameter space is taken into account given the scenarios expected at TCV
- This approach is backup by similar works in which NN were able to learn inverse mappings of CRM, although for accelerating simulations and not real-time diagnosis [6, 7]

[6] Mathews, A., et al. "Deep modeling of plasma and neutral fluctuations from gas puff turbulence imaging." Review of Scientific Instruments 93.6 (2022).

[7] Vander Wal, Michael D., Ryan G. McClarren, and Kelli D. Humbird. "Neural network surrogate models for absorptivity and emissivity spectra of multiple elements." Machine Learning with Applications 8 (2022).

Offline to real-time analysis conversion: Parameter inference



(a) $\mathrm{BNN}[\mathrm{He}^{\mathrm{LRT}}] \rightarrow \{T_e, n_e\}$ **(b)** $\mathrm{HeMSI} \rightarrow \{T_e, n_e\}$

Jaime Caballero et al., Machine learning accelerated plasma parameter inference for multispectral imaging on TCV, in preparation for Plasma Physics Controlled Fusion

Hardware and software implementation for real-time analysis

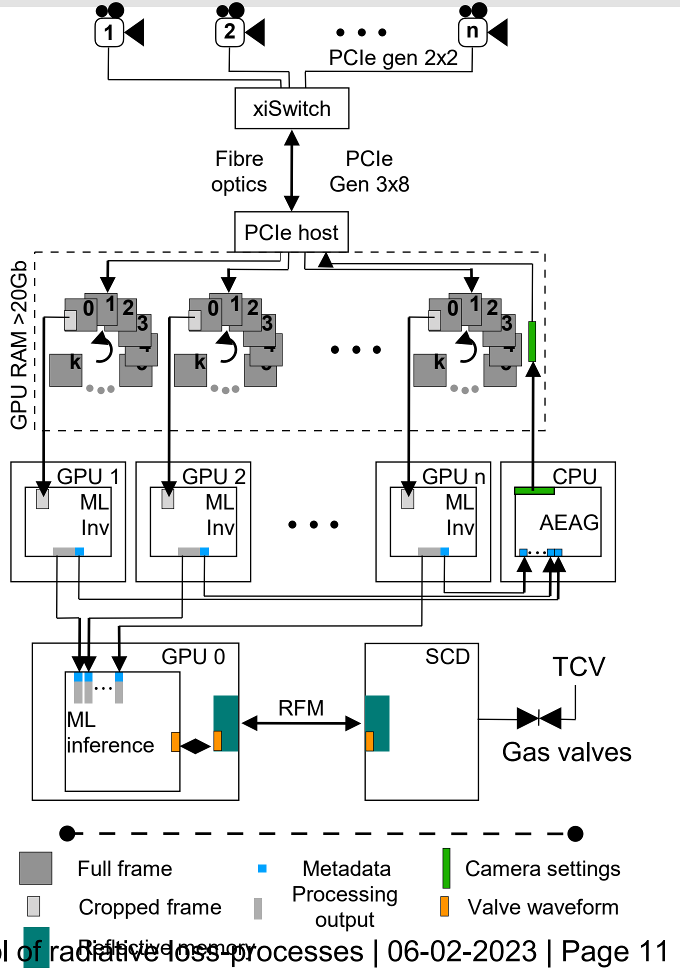
Strategy:

- Move our current CPU processing into a GPU for performance in ML applications.
- Move from our in-house developed C/C++ code to the F4E MARTe2 framework for maintainability and compatibility with other fusion experiments.
- Incorporate Nvidia Triton Inference Server **to divide a single physical GPU into n parallel virtual GPUs.**

GPU requirements:

- **Supports remote direct memory access (RDMA) for the cameras to stream directly to the GPU**
- At least 1Gb of GPU RAM for 1s of frames per camera, at least 20Gb total.
- **Sufficiently powerful to deliver the required performance under the Nvidia Triton Inference Server**

- We are working with the Nvidia Science Team to test our networks on their servers and determine the most suitable GPU for our applications. Currently deciding between Nvidia V100 and **A100** models (up to 15k€)
- Implementation and testing with MANTIS cameras.
- Aiming for the first experiments ~~by the end of 2022 or~~ the second half of 2023



Hardware and software implementation for real-time analysis

Strategy:

- Move our current CPU processing into a GPU for performance in ML applications.
- Move from our in-house developed C/C++ code to the F4E MARTe2 framework for maintainability and compatibility with other fusion experiments.
- Incorporate Nvidia Triton Inference Server **to divide a single physical GPU into n parallel virtual GPUs.**

GPU requirements:

- **Supports remote direct memory access (RDMA) for the cameras to stream directly to the GPU**
- At least 1Gb of GPU RAM for 1s of frames per camera, at least 20Gb total.
- **Sufficiently powerful to deliver the required performance under the Nvidia Triton Inference Server**

- We are working with the Nvidia Science Team to test our networks on their servers and determine the most suitable GPU for our applications. Currently deciding between Nvidia V100 and **A100** models (up to 15k€)
- Implementation and testing with MANTIS cameras.
- Aiming for the first experiments ~~by the end of 2022 or~~ the second half of 2023

	A100 80GB PCIe	A100 80GB SXM
FP64	9.7 TFLOPS	
FP64 Tensor Core	19.5 TFLOPS	
FP32	19.5 TFLOPS	
Tensor Float 32 (TF32)	156 TFLOPS | 312 TFLOPS*	
BFLOAT16 Tensor Core	312 TFLOPS | 624 TFLOPS*	
FP16 Tensor Core	312 TFLOPS | 624 TFLOPS*	
INT8 Tensor Core	624 TOPS | 1248 TOPS*	
GPU Memory	80GB HBM2e	80GB HBM2e
GPU Memory Bandwidth	1,935 GB/s	2,039 GB/s
Max Thermal Design Power (TDP)	300W	400W ***
Multi-Instance GPU	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 10GB

https://www.nvidia.com/en-us/data-center/a100/

MANTIS 5 - a testbed for GPU computing development

ASUS Prime Z690-P
Seasonic Focus GX 1000 W
Kingston FURY Beast RGB - 64GB
Intel Core i9-12900K LGA 1700, 3.20 GHz, 16 -Core
Rocky Linux 8.6

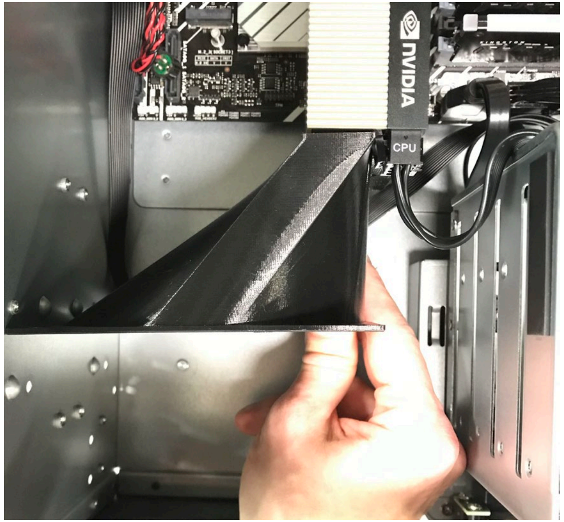
The default server rack cooling fan was insufficient to prevent an idle thermal crash (95 deg C).

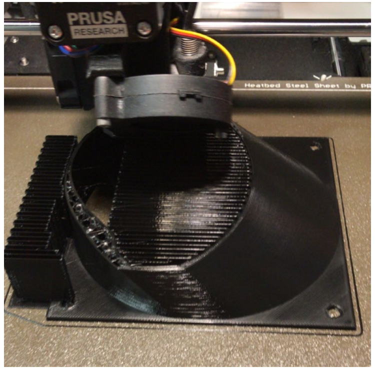
An additional fan at the card's outlet and a streaming funnel combined with a directly mounted fan was enough to prevent a thermal crash when idle (70/300W) while still thermally throttling (85 deg C).

OD1238-12HBVXC

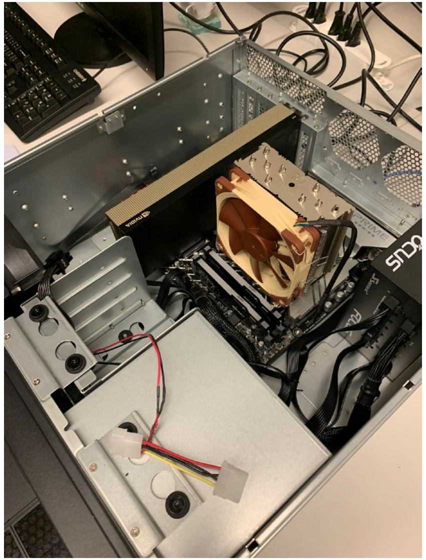
DC 120x120x38mm

12V, 384m³/h, 50W





https://www.thingiverse.com/thing:5145188



MANTIS 5 - a testbed for GPU computing development

The default server rack cooling fan was insufficient to prevent an idle thermal crash (95 deg C).

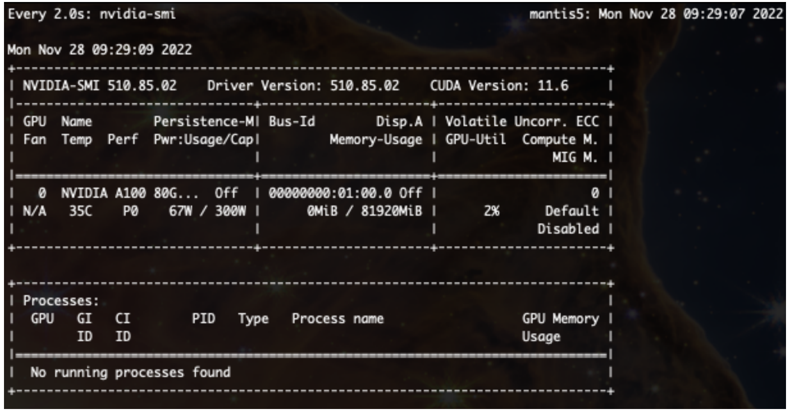
An additional fan at the card's outlet and a streaming funnel combined with a directly mounted fan was enough to prevent a thermal crash when idle (70/300W) while still thermally throttling (85 deg C).

ASUS Prime Z690-P
Seasonic Focus GX 1000 W
Kingston FURY Beast RGB - 64GB
Intel Core i9-12900K LGA 1700, 3.20 GHz, 16 -Core
Rocky Linux 8.6

OD1238-12HBVXC

DC 120x120x38mm 12V 384m³/h

50W…

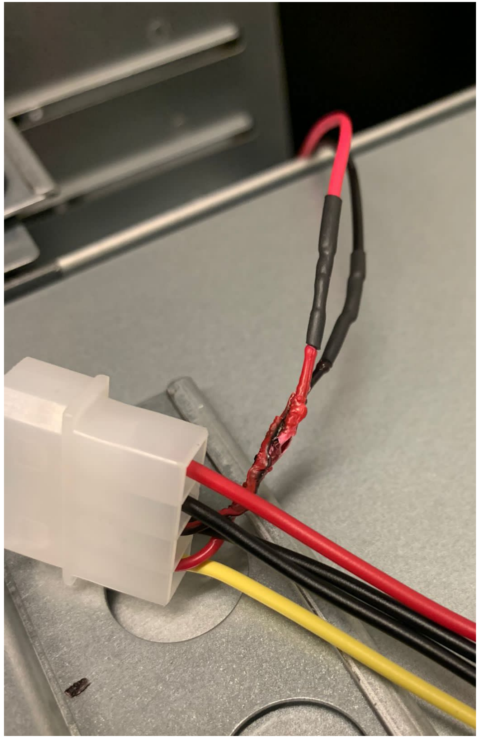


```
Every 2.0s: nvidia-smi                                    mantis5: Mon Nov 28 09:29:07 2022

Mon Nov 28 09:29:09 2022
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 510.85.02    Driver Version: 510.85.02    CUDA Version: 11.6     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA A100 80G...  Off  | 00000000:01:00.0 Off |                    0 |
| N/A   35C    P0    67W / 300W |      0MiB / 81920MiB |      2%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

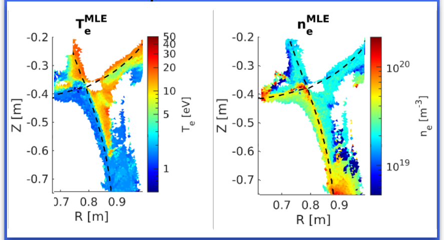
67 deg C at full power in a steady state.



Offline to real-time analysis conversion

H

Uses emissivities of 4 Balmer lines:
$D_{3\to2}, D_{4\to2}, D_{5\to2}, D_{7\to2}$
to infer maps of:

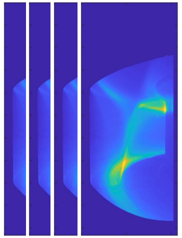


Main conclusions:
- The 2D map of the ionisation and recombination rates can be inferred in the divertor leg.
- Molecular contributions to the Balmer series can significantly skew the results and must be included in the analysis.

[1] A. Perek et al. 2022 Nucl. Fusion **62** 096012

~0.02 GPUs per frame | ~30 GPUs per frame

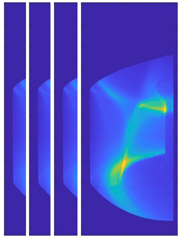
Plasma emission



~5 CPUs per frame | ~4000 CPUs per time step

→ Tomographic inversions → Bayesian Plasma parameter inference [1] → I_{rate}, R_{rate}

Plasma emission



→ Machine learning for tomographic inversions → Machine learning for plasma parameter inference → I_{rate}, R_{rate}

~0.002 GPUs per frame | ~0.0006 GPUs per frame

2.6 ms

Tested with Nvidia A100

Hardware and software implementation for real-time analysis

Does the GPU Direct Memory Access work?
- Yes, I tested it on Mantis2b PC.

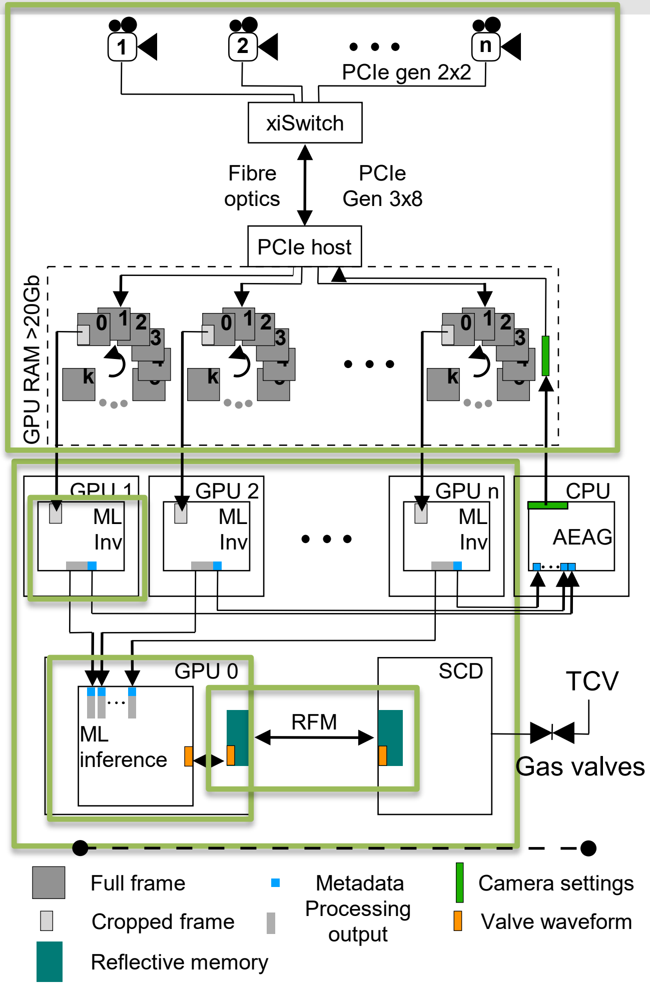
Can we perform the real-time tomographic inversion?

Can we perform the real-time parameter inference?

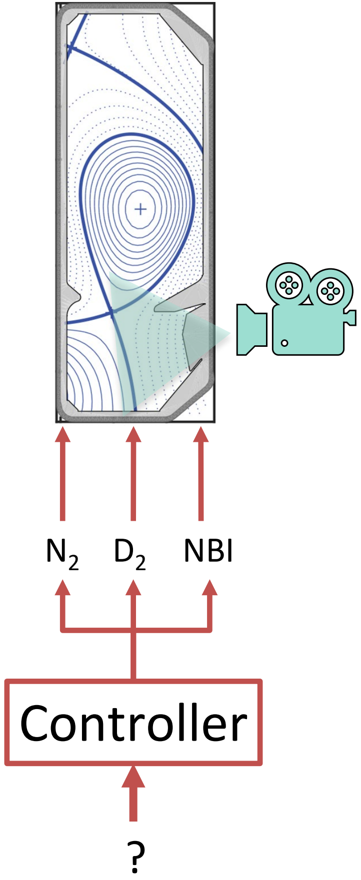
Can we setup parallel processing streams on the GPU using Multi-Instance GPU?

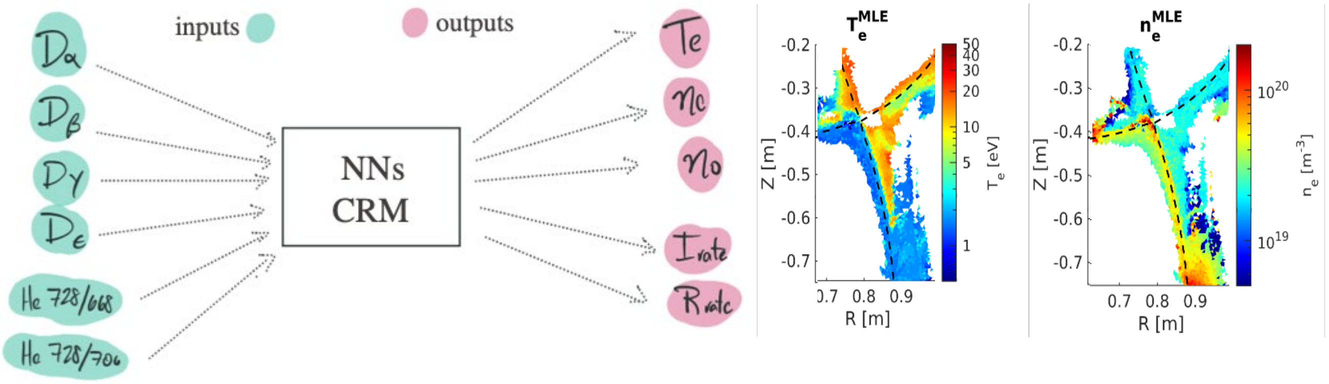
Can we perform operations within a <1ms jitter to ensure real-time performance?

There are no known obstacles to performing integration for real-time control but further development is taking place



WP2/6: Real-time control of the heat-exhaust



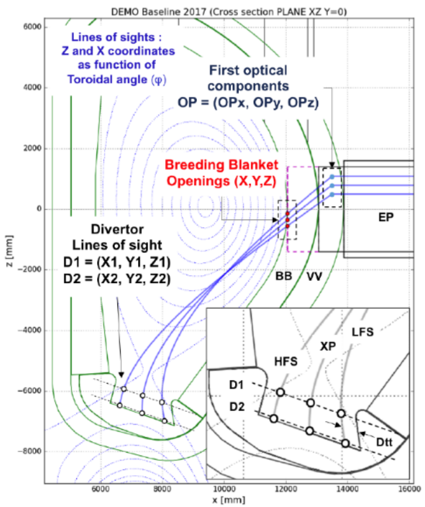


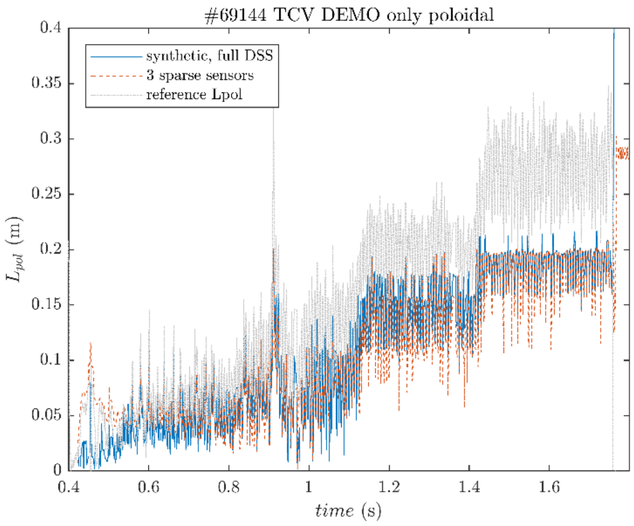
What do we wish to control?

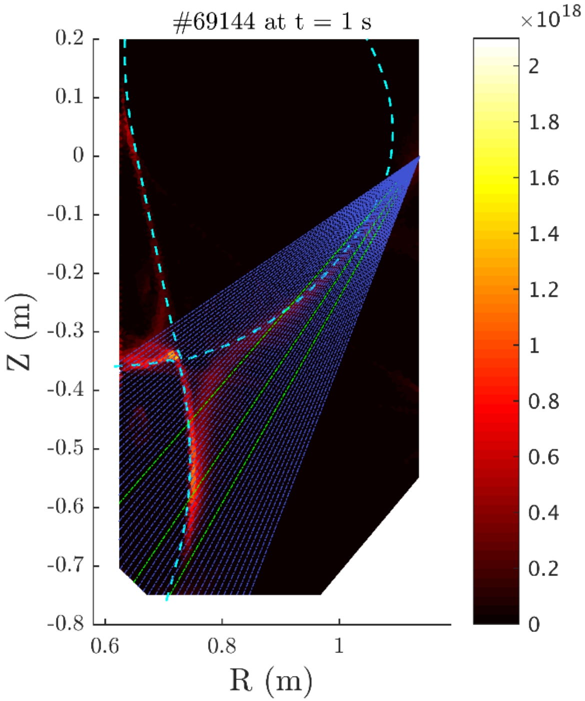
- Start with control of total I_{rate} and R_{rate} in divertor
- Analyse response of T_e, n_e, n_0, I_{rate}, R_{rate} to actuators
- Identify possible 'region of interests' to define the exhaust solution in other parameters -> improved control parameters, go from 0D to 1D
- Combine with e.g. FIR: extend to larger Multi-input / Multi-output schemes

WP6: Synthetic spectroscopy for exhaust in DEMO

- Spectroscopy on DEMO (WPDC/WPENR)
- Different lines-of-sight configuration than currently available: synthetic diagnostic using MANTIS







Where do we stand now: deliverables?

Made deliverables (from 2022):

- **Simple-MIMO identification demonstration**: Demonstrate MIMO system identification algorithms developed in this proposal for the simplified case of D2 and N2. (initial testing RT-alg.).

Deliverables (from 2023) almost made:

- **Loss-process measurements in 2D: Quantitative real-time algorithms for the observation of the loss-processes (based on MANTIS + other diagnostics):** Successful first publication on off-line quantitate modelling, heavy investment in ML algorithms to speed up, implementation on hardware GPU progressing well

- **Dynamic detachment models: Control-oriented (hybrid) models useful for time-dependent detachment simulations and control development:** First publication G.L. Derks, static maps with DIV1D, dynamic models based on data also ready, improved model under development

Next to making progress on the project (presentations/publications being prepared):

Just submitted: L. van Leeuwen et al., Machine learning accelerated tomographic reconstruction for multispectral imaging on TCV, Nuclear Fusion

In preparation: Machine learning accelerated plasma parameter inference for multispectral imaging on TCV, J. Caballero et al., Plasma Physics Controlled Fusion

Multi-input multi-output control of the plasma exhaust and line-integrated electron density in a tokamak plasma, J.T.W. Koenders, Nuclear Fusion

Conferences: Application of a sparse sensor placement technique to the limited availability of experimental data in DEMO, J. Raukema et al. (ECPD)

Multispectral Advanced Narrowband Tokamak Imaging Systems (MANTIS), A Perek et al. (ECPD)

Multivariable feedback control of radiative loss-processes using multi-spectral imaging, M. van Berkel et al. (IAEA FEC in prep.)

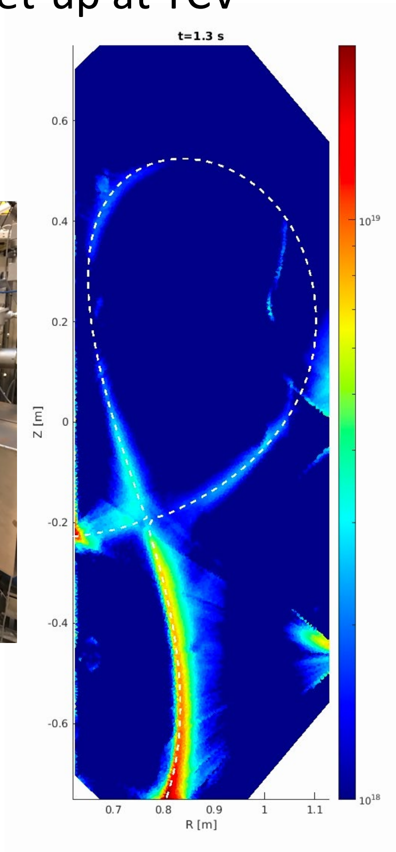
Contributions outside ENR

- ENR contributes to MAST-U SXD multi-wave imaging
- New GPU methods ALSO speed up calculations for post-processing, inter-shot inference is now possible for WPTE campaign physics studies
- Jointly with WPDC we develop synthetic diagnostics and MANTIS can be used for DEMO control like sightlines experiments
- Full plasma diagnostics 3 MANTIS camera’s (next slide)

Thank you for your attention

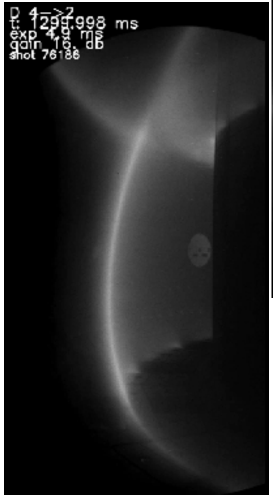
3 MANTIS set-up at TCV

MANTIS 2b



D_β emissivity, TCV #76186

Mantis 1



Mantis 2a