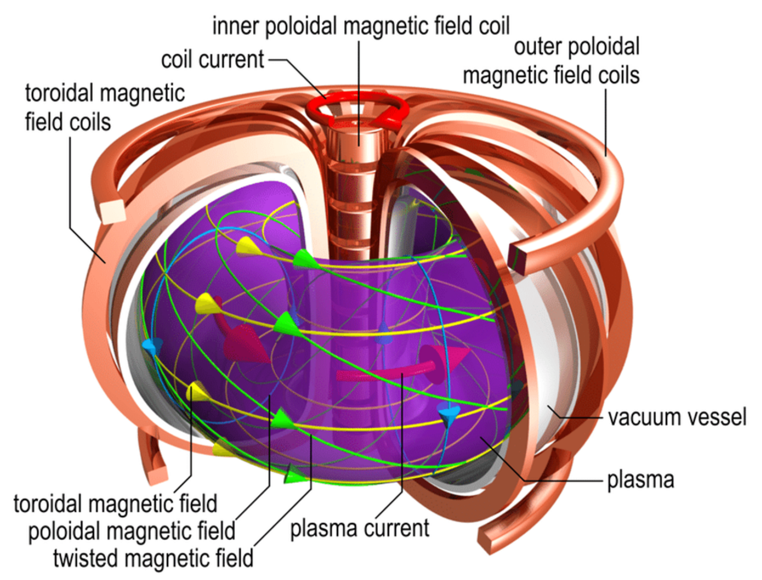
A RL-BASED VERTICAL STABILIZATION SYSTEM FOR THE EAST TOKAMAK

G. De Tommasi[1,2], S. Dubbioso[1,2], Y. Huang[3], Z. P. Luo[3], A. Mele[4], B. J. Xiao[3]

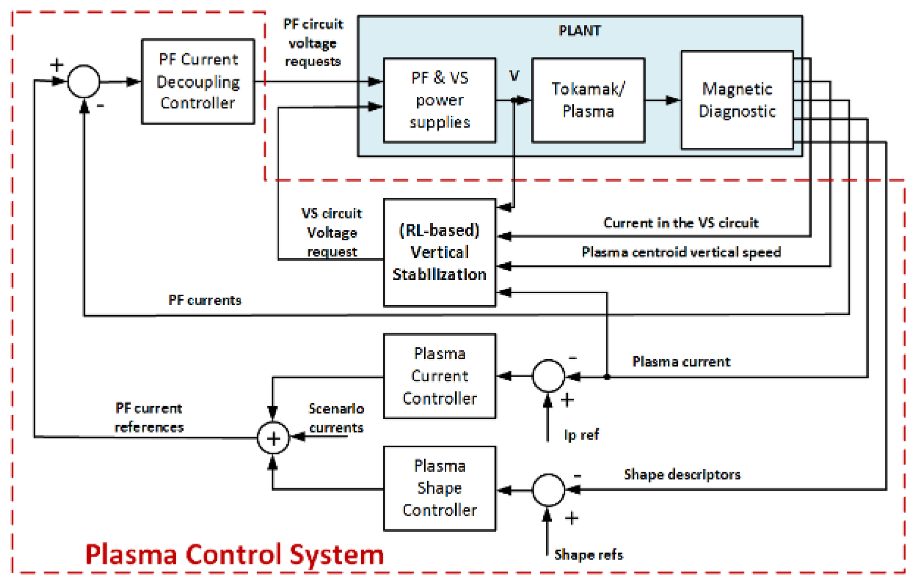
[1] Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, via Claudio 21, 80125, Napoli, Italy
[2] Consorzio CREATE, via Claudio 21, 80125, Napoli, Italy
[3] Dipartimento di Economia, Ingegneria, Società e Impresa, Università degli Studi della Tuscia, Viterbo, Italy
[4] Institute of Plasma Physics, Chinese Academy of Sciences, Hefei 230031, People's Republic of China

Nuclear fusion & Tokamaks

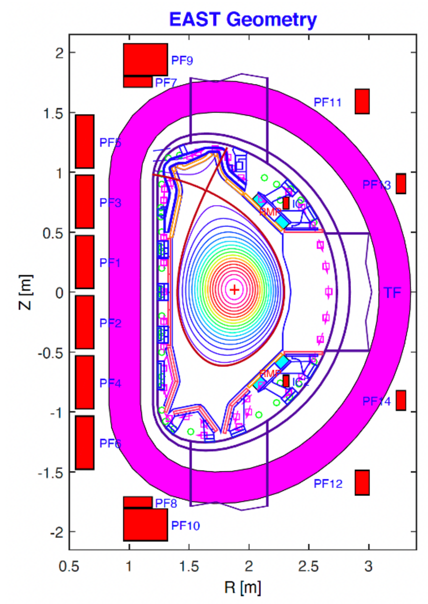
- Nuclear fusion is foreseen as a promising source of clean and sustainable energy for the next century
- Tokamak are experimental devices aimed at producing energy from nuclear fusion reactions that occur in a **magnetically confined** hot **plasma**
- *EAST* is a superconductive tokamak located in Hefei, People's Republic of China



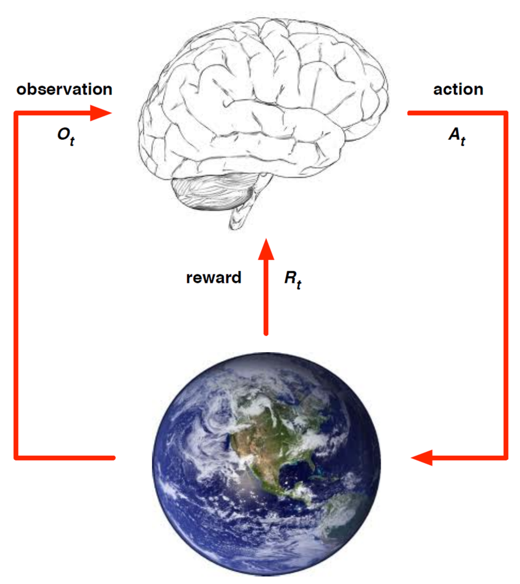
Plasma Magnetic Control & Vertical Stabilization Stabilization



- **Plasma magnetic control** aims at controlling the current, position and shape of the plasma column inside the vacuum vessel by means of **external magnetic** fields generated by the so called Poloidal Field coils (PFC)
- High performance plasmas, as the ones achieved at the EAST tokamak, have elongated poloidal cross-section which turn to be **vertically unstable** (like a ball on the top of a hill)
- A **Vertical Stabilization (VS)** system is needed to run any modern tokamak



Reinforcement Learning approach for the EAST VS



Main idea: train an agent by making it interact with a state-space linearized model of the EAST plasma and surrounding coils dynamics (1) (RL environment), so that the agent learns how to solve the VS problem

$$\delta\dot{x}(t) = A\delta x(t) + B\delta u(t) + E\delta\dot{w}(t) \quad (1a)$$
$$\delta y(t) = C\delta x(t) + F\delta w(t)\,, \quad (1b)$$

The offline training based on control of random Vertical Displacement events (VDE) in the range ± 5 cm permit to retrieve a static input-output table for the VS control system, which represents the VS control strategy

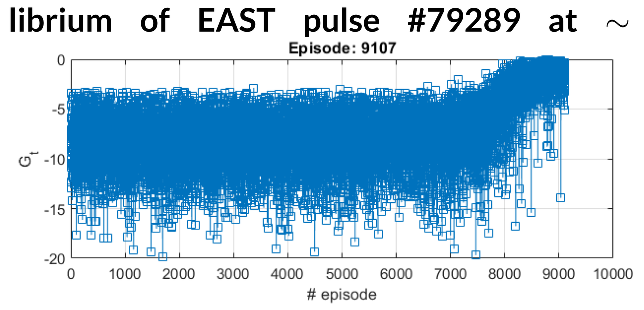
The **RL-based VS agent** can then be included in the whole magnetic control architecture and implement the policy to select the voltage request to the in-vessel coils V_{IC} based on observation coming from the plant

Q-learning for the training of the VS agent

Reward function

$$R(s\,,a) = -k_1 \cdot \left(\frac{\dot{Z}_c}{\dot{Z}_{c_{\max}}}\right)^2 - k_2 \cdot \left(\frac{I_{IC}}{I_{IC_{\max}}}\right)^2 - k_3 \cdot \left(\frac{V_{IC}}{V_{IC_{\max}}}\right)^2\,, \quad (2)$$

Cumulative reward (Training agains equilibrium of EAST pulse #79289 at $\sim$ 3 s



Action and state discretization parameters

	V_{IC}	I_{IC}	$\dot{Z}_c$
Points number	17	21	21
Max absolute value	300 V	6 kA	30 m/s
Bonus threshold	–	50 A	0.5 m/s

Input

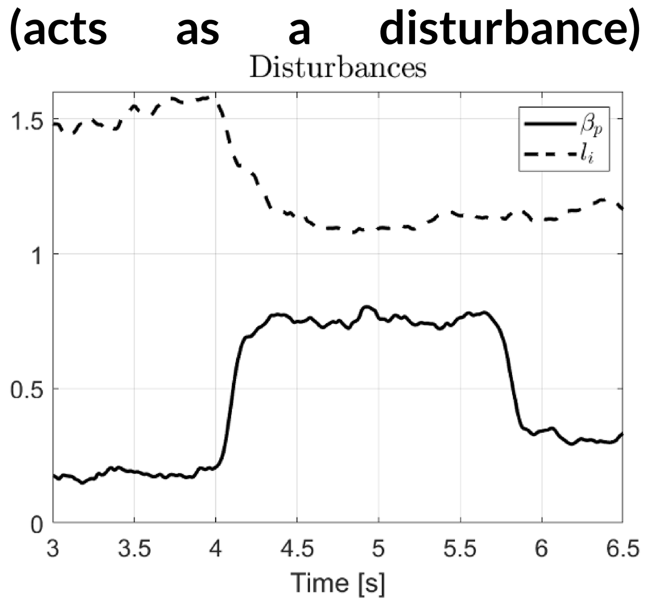
- state $s = (\dot{Z}_c\,, I_{IC})$ and action $a = V_{IC}$ discretized spaces
- Maximum state values $\dot{Z}_{c_{\max}}\,, I_{IC_{\max}}$
- Bonus assignation threshold $\dot{Z}_b\,, I_{IC_b}$
- Bonus b
- step size $\alpha \in [0, 1)$
- discount factor $\gamma \in [0, 1)$
- initial exploration parameter $\epsilon \in [0\,, 1)$
- ϵ-decay factor $\delta \in [0\,, 1)$

Specify reward function $R(s\,,a)$ according to (2)
Initialize $Q(s\,,a) \leftarrow R(s\,,a)$ for all state-action pairs
foreach *episode* **do**
  Initialize s_0 with a random VDE in the range $[-5\,,5]$ cm
  Initialize the cumulative reward $G \leftarrow 0$
  foreach *step* t *in an episode* **do**
    Choose $a_t \in \mathcal{A}$ given $s_t \in \mathcal{S}$ according to the ϵ-greedy policy applied on current Q table
    Simulate plasma linearized model starting from state s_t applying action a_t
    Observe the new state s_{t+1}
    $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\left[R(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\right]$ \$
    Update $s_t \leftarrow s_{t+1}$
    (* Evaluate bonus and episode terminating condition *)
    Initialize the current reward $r \leftarrow R(s_t\,, a_t)$
    if $|\dot{Z}_{c_t}| < \dot{Z}_b$ ***and*** $|I_{IC_t}| < I_{IC_b}$ **then**
      $r \leftarrow r + b$
    end
    else if $|\dot{Z}_{c_t}| >= \dot{Z}_{\max}$ ***or*** $|I_{IC_t}| >= I_{IC_{\max}}$ **then**
      $r \leftarrow r - 10 * b$
      terminate the episode
    end
    Update the episode cumulative reward $G \leftarrow G + \gamma^t r$
    Update $\epsilon \leftarrow \delta\epsilon$
  end
end

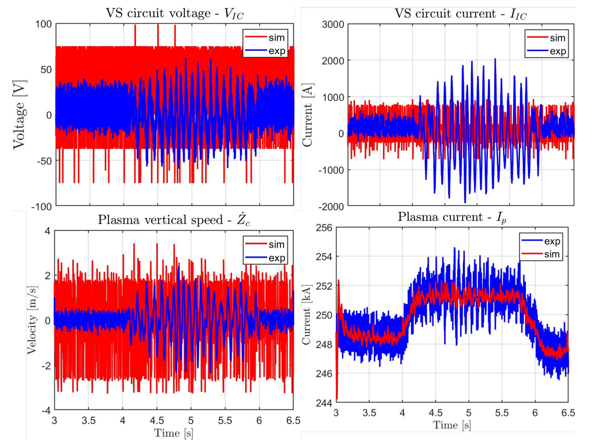
ster

Simulation results & Conclusions

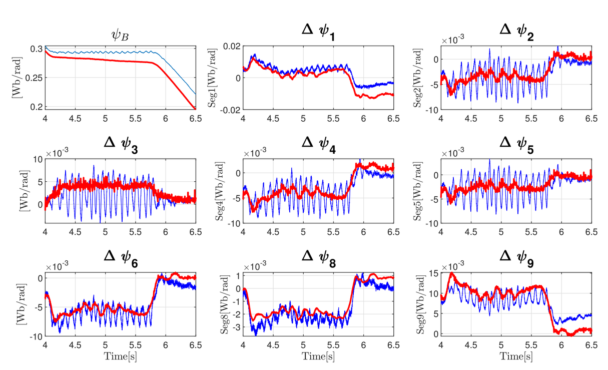
Playback of EAST pulse #92141 1 MW of additional heating power injected (acts as a disturbance)



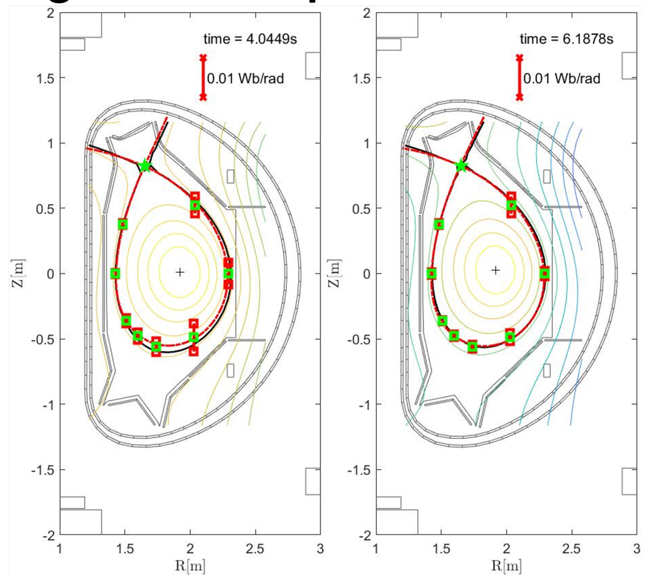
Agent I/O and Plasma current



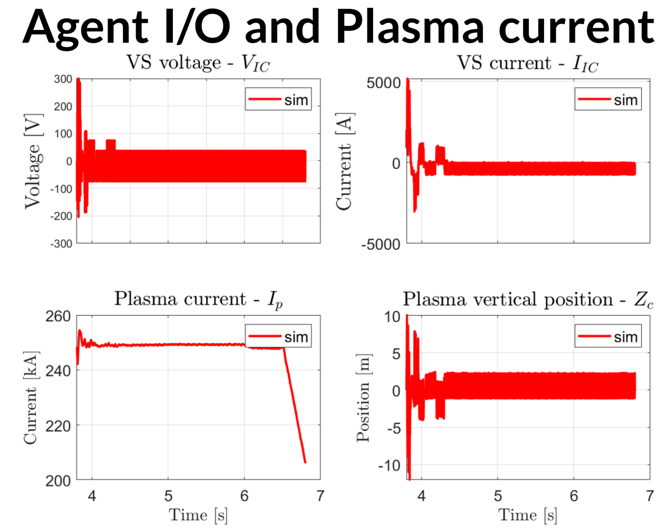
Plasma shape descriptors



Vertical displacement events (VDE) applied at $\sim$4 s during EAST pulse #79289



Agent I/O and Plasma current



- The preliminary assessment of the data-driven VS for the EAST proved to be robust with respect to models and scenarios not used during the training
- The robustness can be increased by considered a set of different equilibria (different plasma internal profiles, triangularity, elongation, ecc.). . .
- . . .however a non-tabular approach such as Deep Deterministic Poloidal Gradient (DDPG) must be pursued in order to contain the computational burden