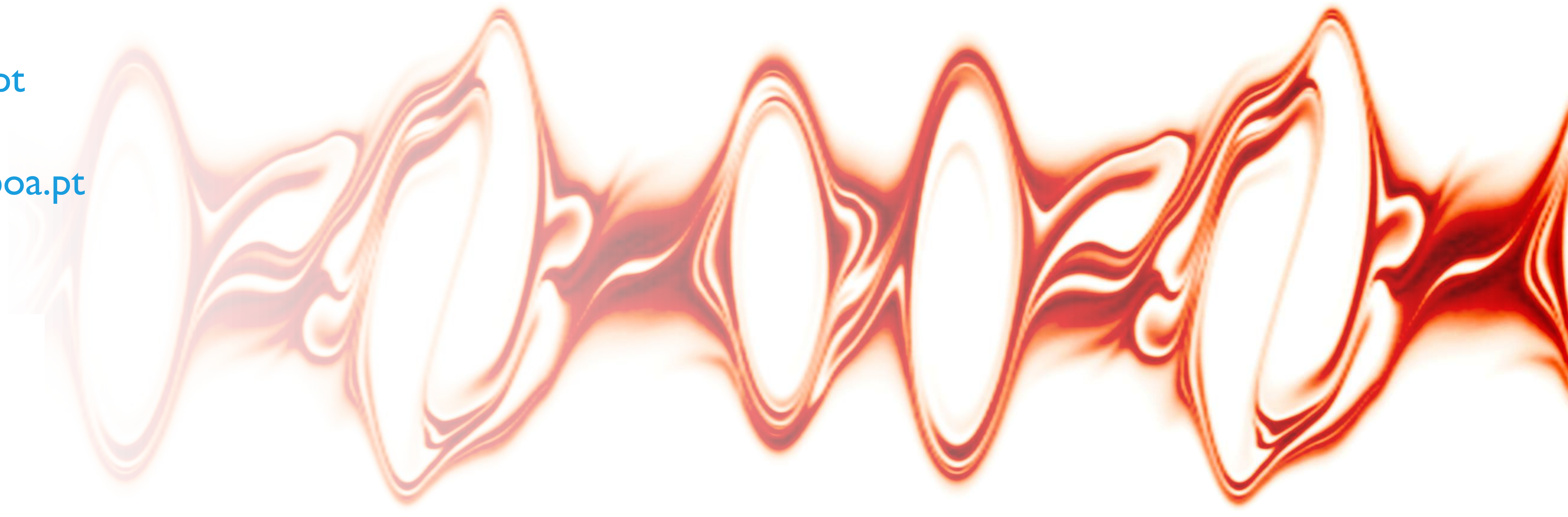
The AI Physicist: learning equations from data

Frederico Fiúza
frederico.fiuza@tecnico.ulisboa.pt
web.ist.utl.pt/frederico.fiuza
https://astroplasmas.tecnico.ulisboa.pt



ipfn
INSTITUTO DE PLASMAS E FUSÃO NUCLEAR

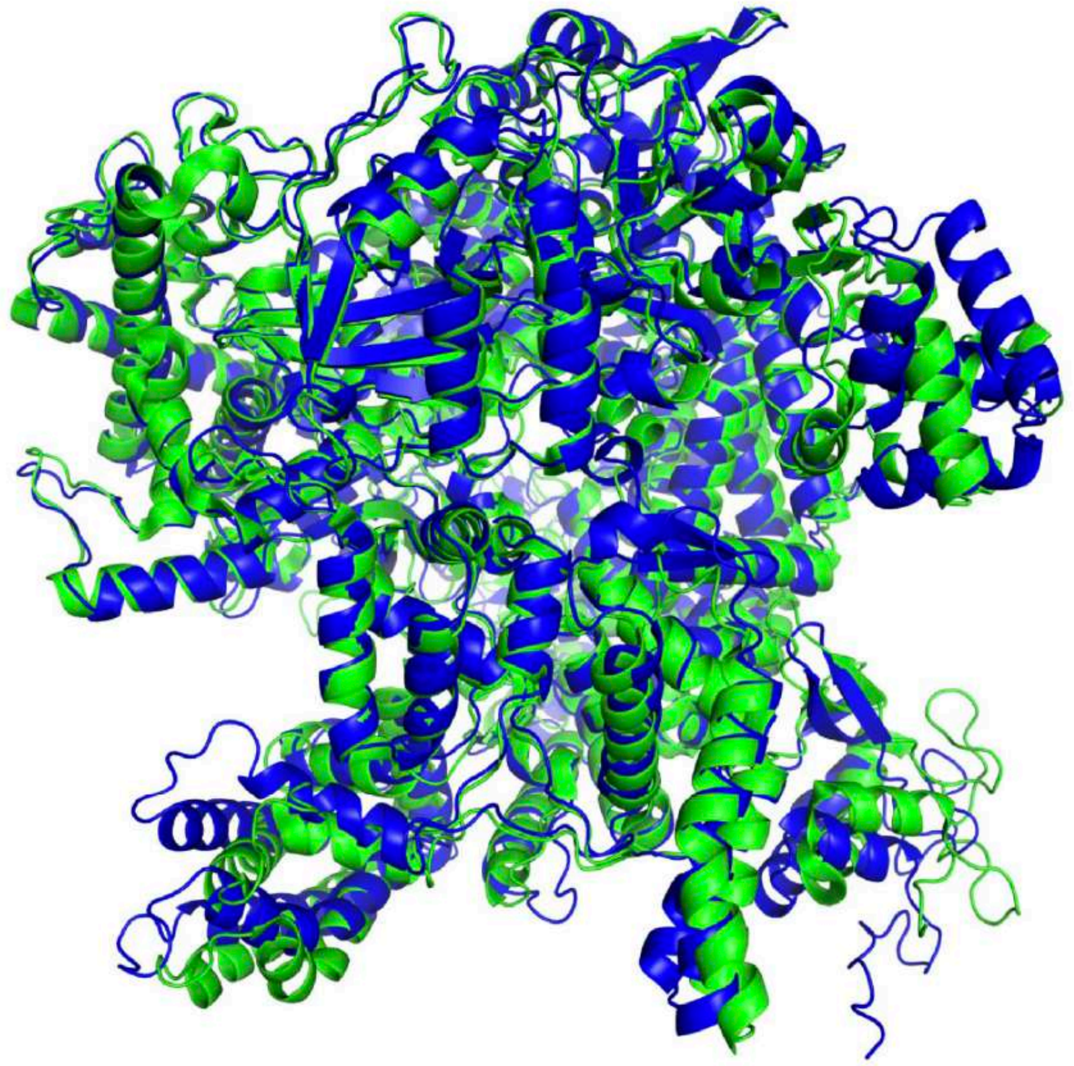
TÉCNICO LISBOA

erc
European Research Council
Established by the European Commission

Artificial intelligence is revolutionizing prediction in science

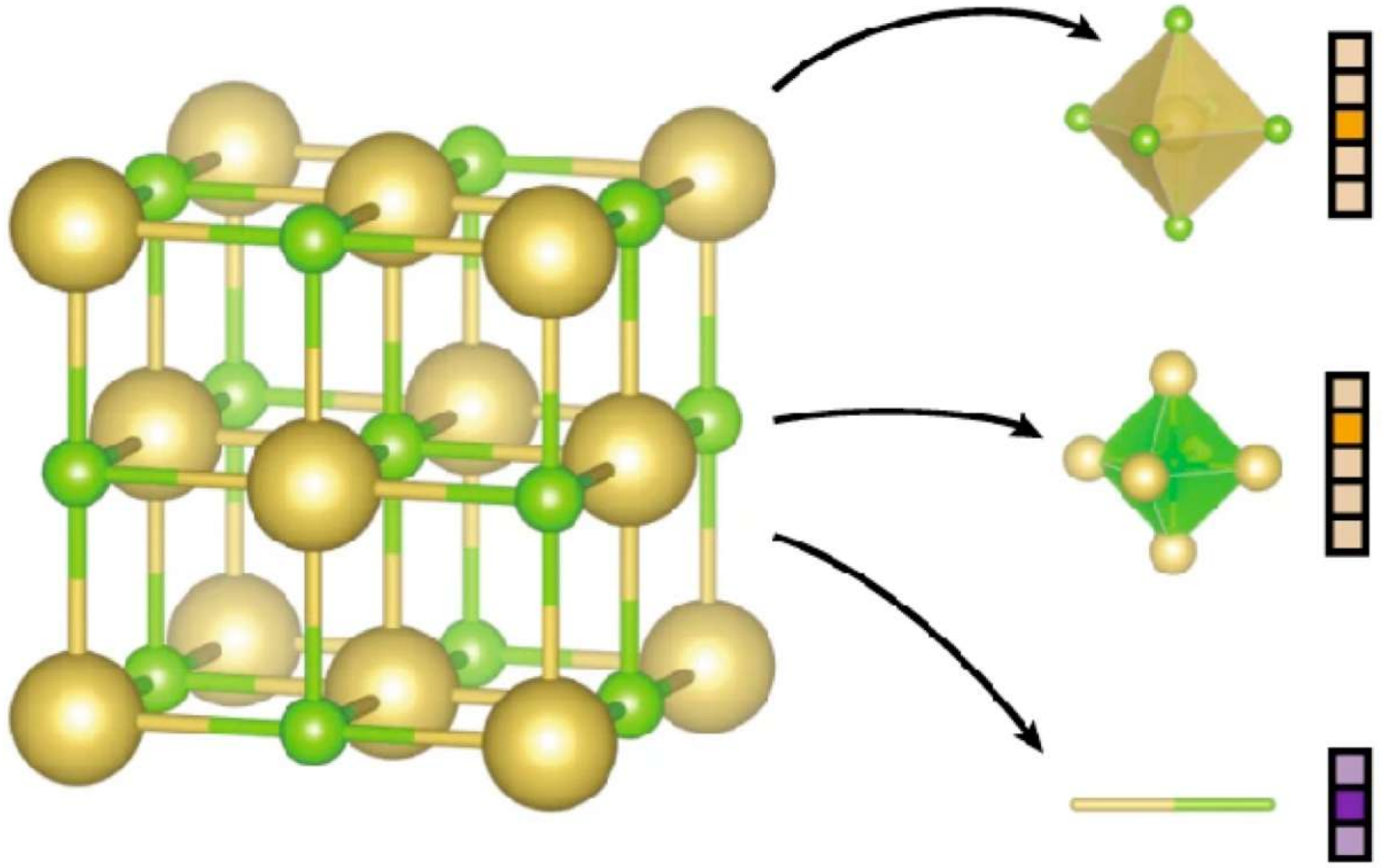
Neural networks (deep learning) are extraordinary at "learning" complex nonlinear patterns in large datasets

Prediction of protein structure



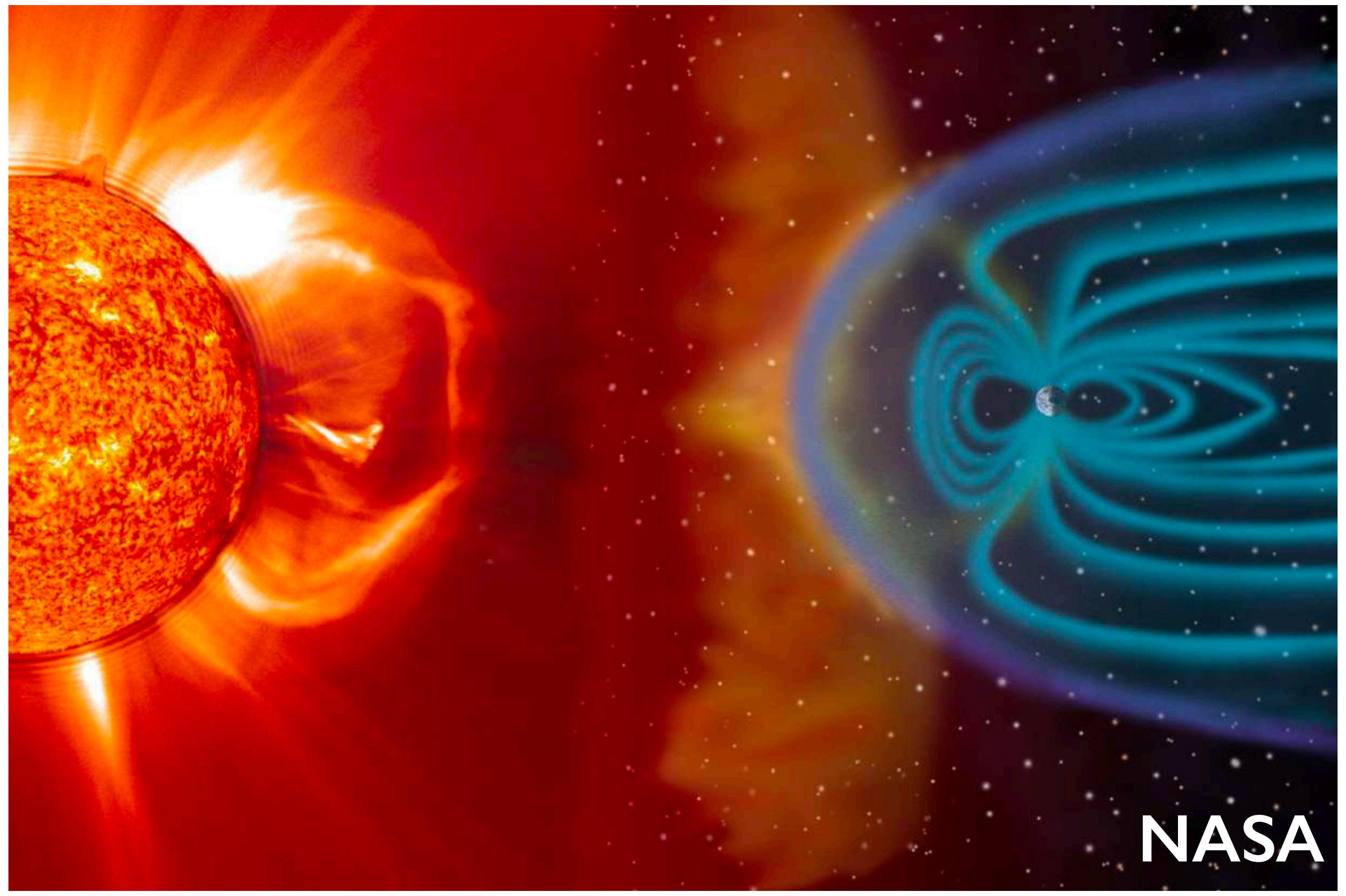
Chemistry Nobel Prize 2024

Discovery of new stable materials



Schmidt et al. 2019

Space weather prediction

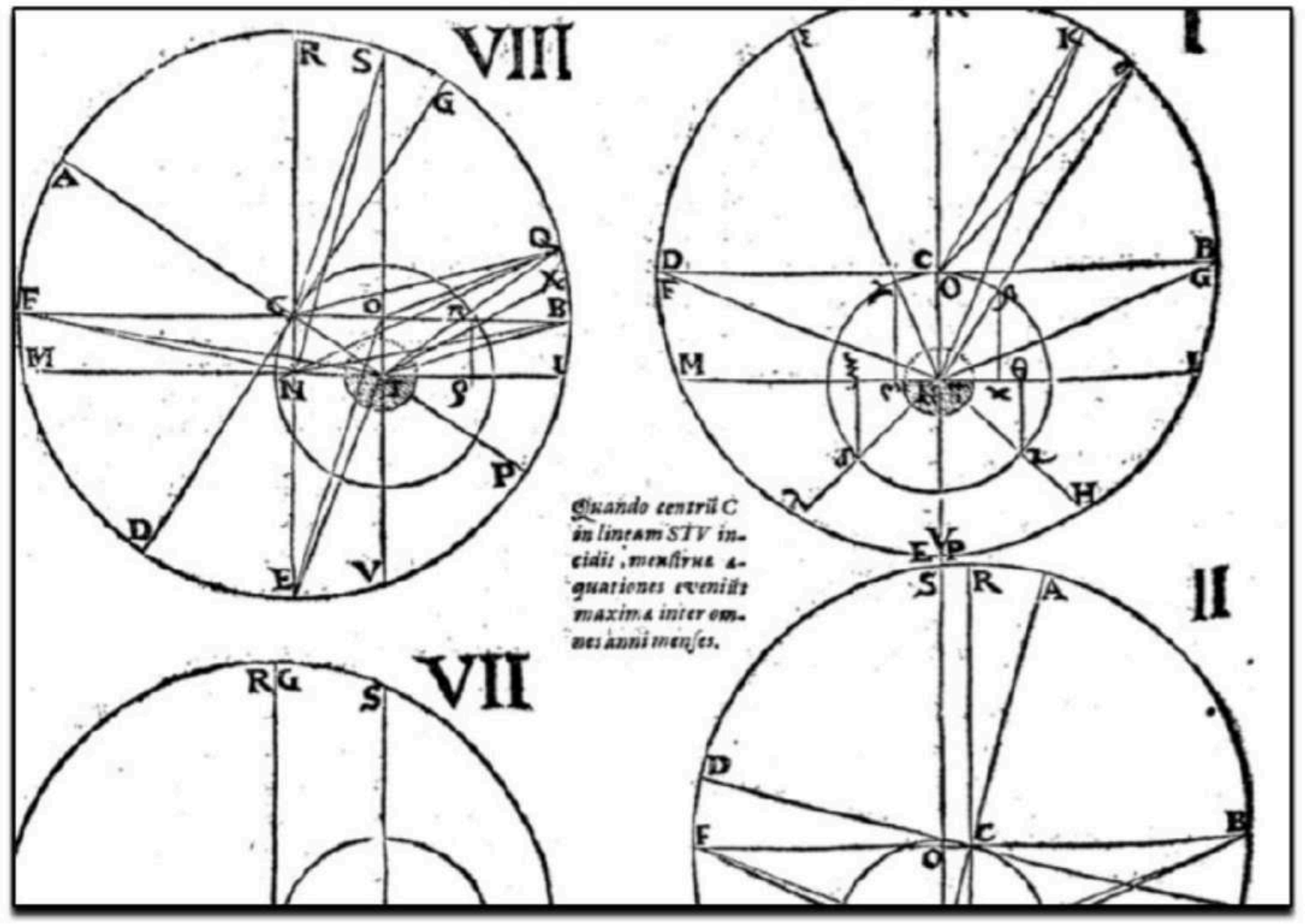


Camporeale et al. 2018

High prediction accuracy often comes at the expense of poor interpretability (and generalization)

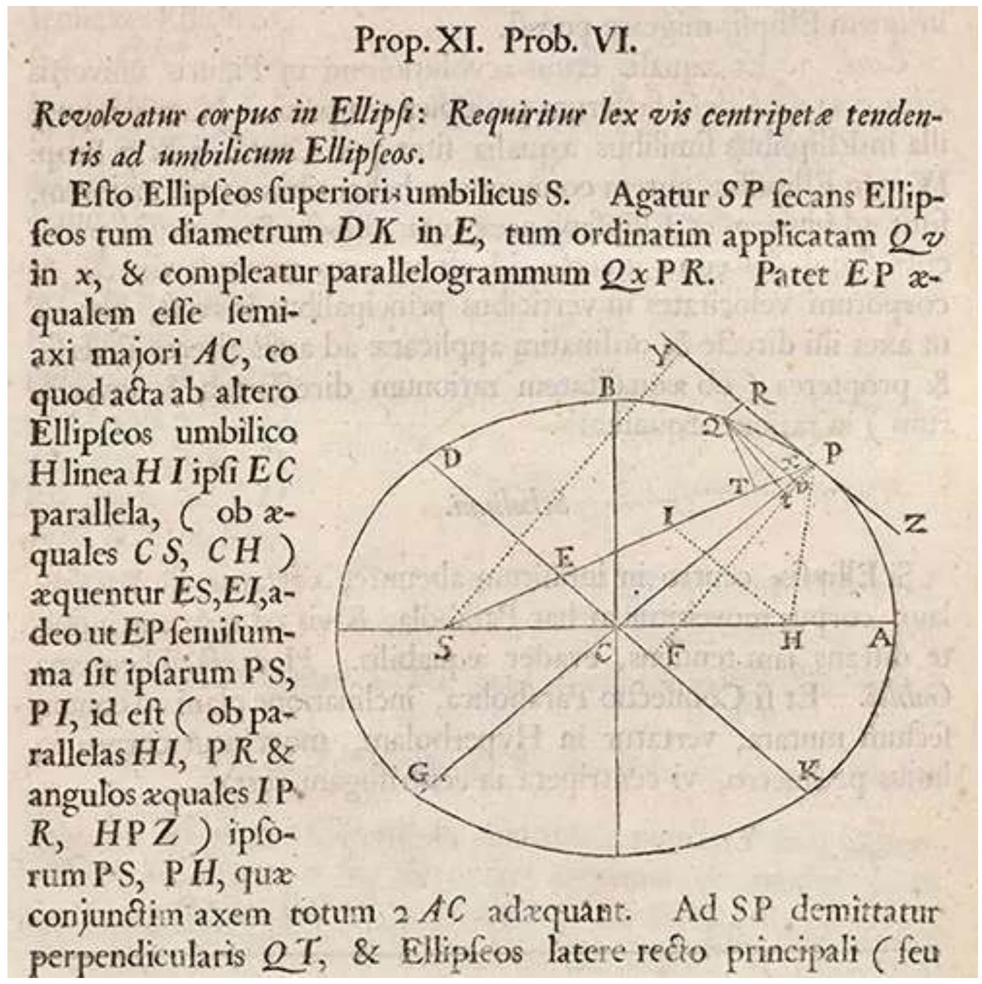
From data science to governing equations

Johannes Kepler (1571-1630): an early big data scientist



By carefully analyzing the best and most well-guarded astronomical data for 11 years, he posited the elliptical nature of planetary orbits

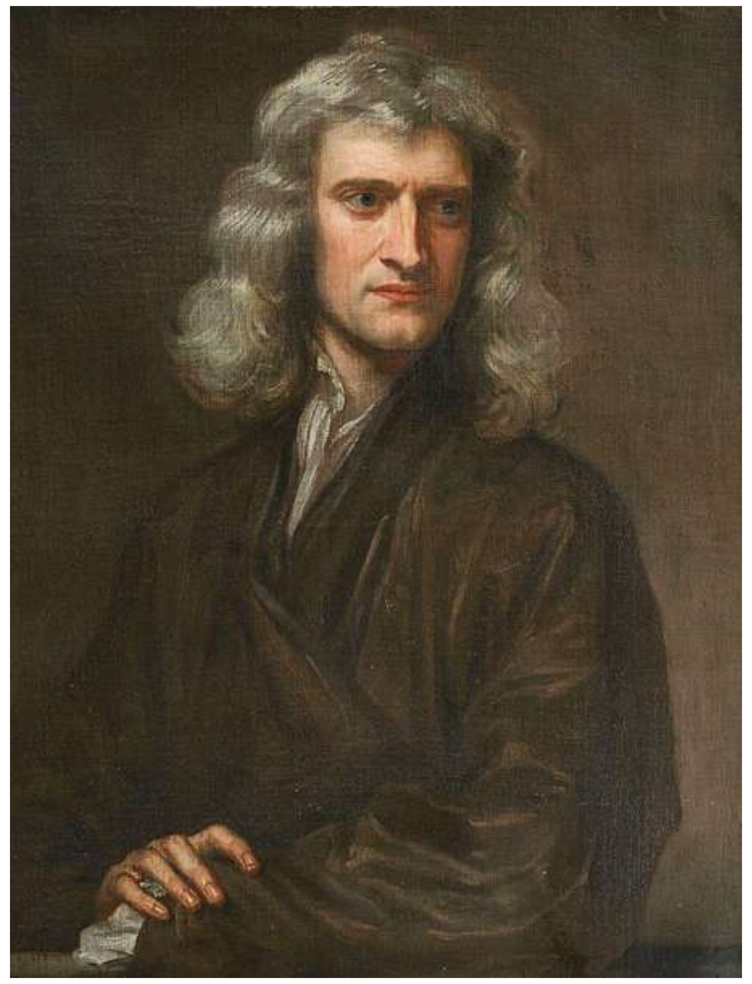
Isaac Newton (1643-1727)



Prop. XI. Prob. VI.

Revolvatur corpus in Ellipsi: Requiritur lex vis centripetæ tendentis ad umbilicum Ellipseos.

Esto Ellipseos superioris umbilicus S. Agatur SP secans Ellipseos tum diametrum DK in E, tum ordinatim applicatam Qv in x, & compleatur parallelogrammum QxPR. Patet EP æqualem esse semiaxi majori AC, eo quod acta ab altero Ellipseos umbilico H linea HI ipsi EC parallela, (ob æquales CS, CH) æquentur ES, EI, adeo ut EP semisumma sit ipsarum PS, PI, id est (ob parallelas HI, PR & angulos æquales IPR, HPZ) ipsorum PS, PH, quæ conjunctim axem totum 2AC adæquant. Ad SP demittatur perpendicularis QT, & Ellipseos latere recto principali (seu



Newton's law of gravitation explained attractive force between objects and why planets move in elliptical orbits

$$F = G\frac{m_1 m_2}{r^2}$$

Can we use AI to discover governing equations directly from data?

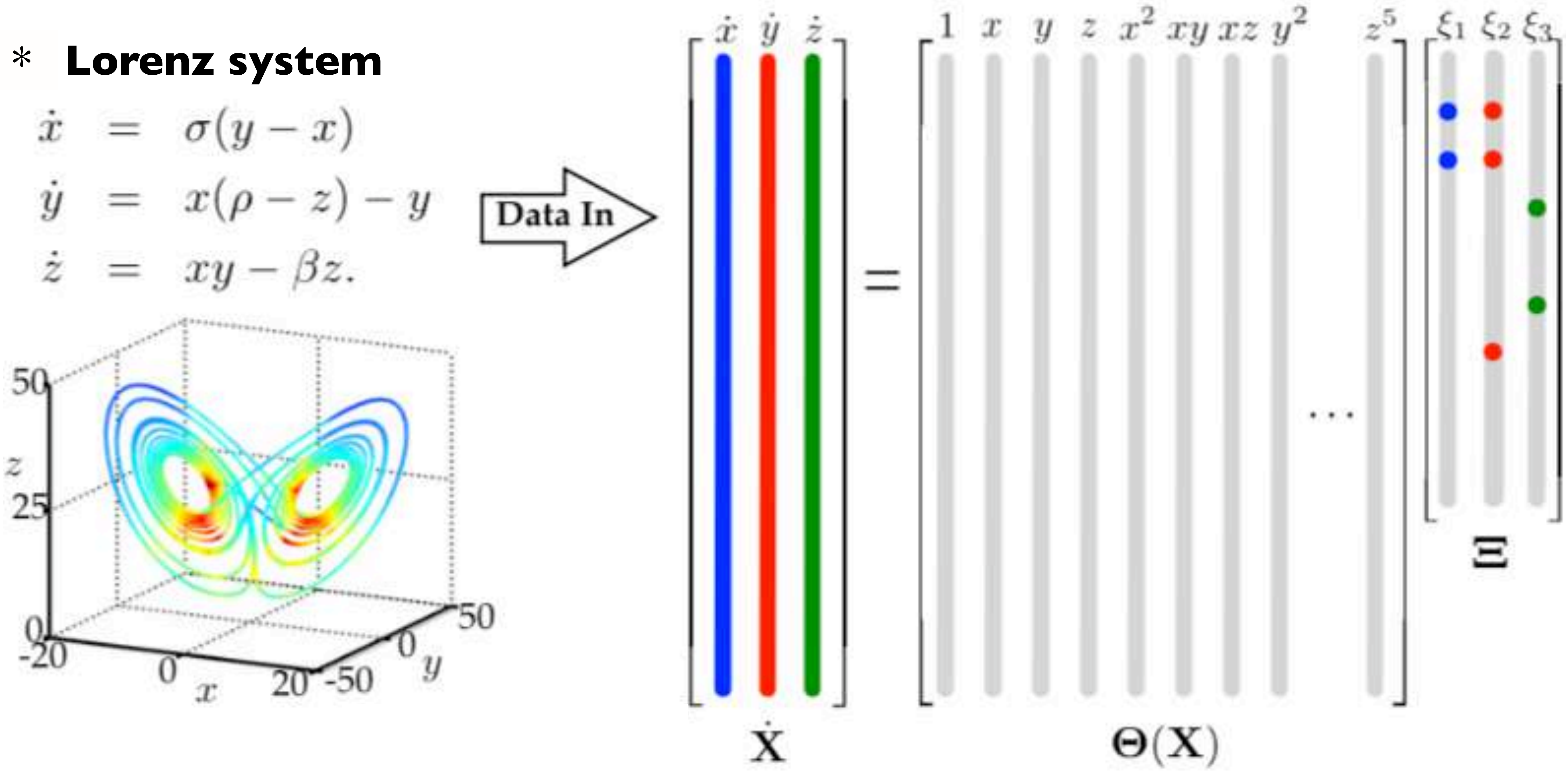
TÉCNICO LISBOA

Sparse regression methodology

- ODE/PDE identification by selecting from library of candidate terms
- Use sparsity-promoting regularizers to select important terms

* **Lorenz system**

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z.$$

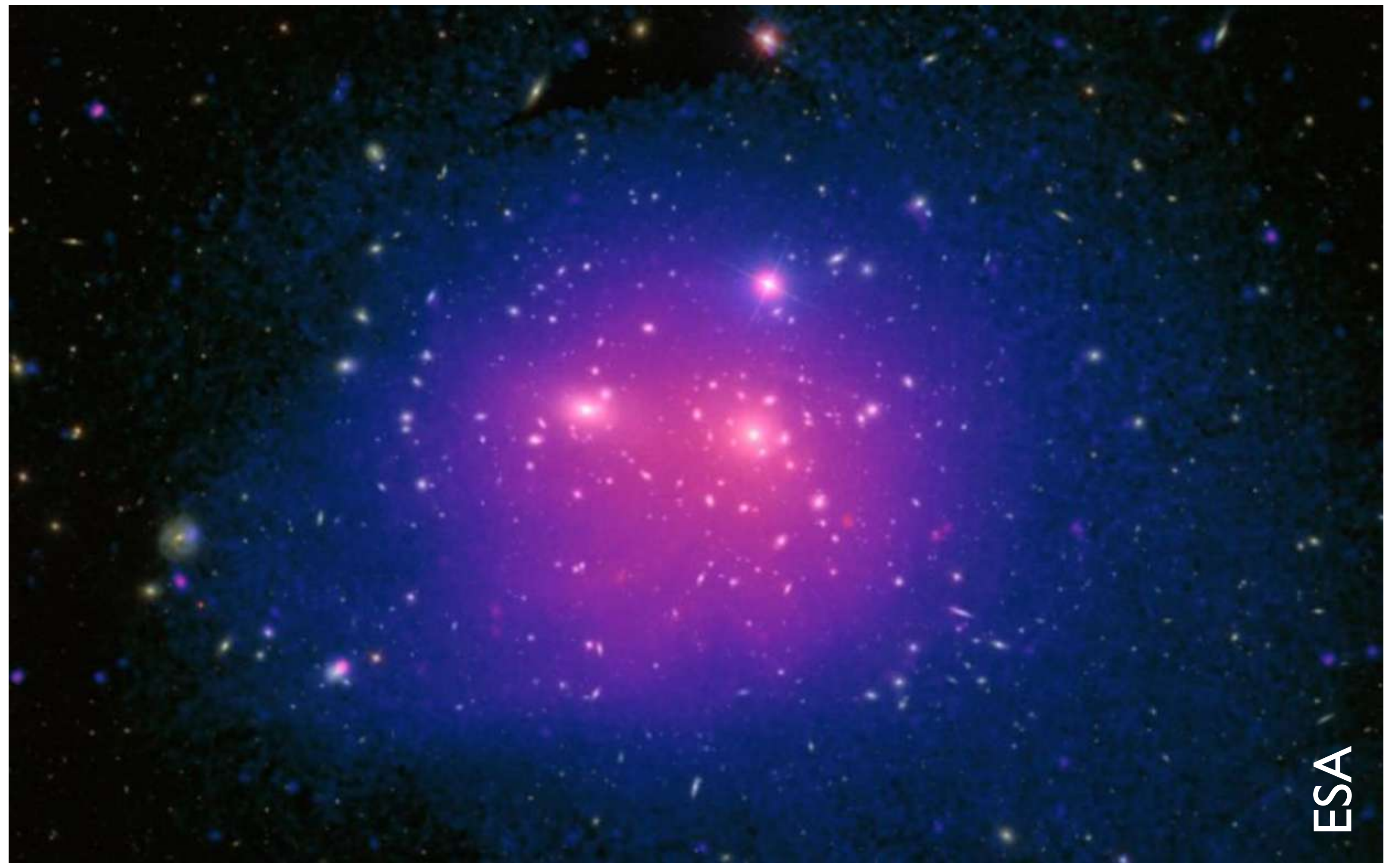


Wang et al., PRL (2011); *Brunton et al., PNAS (2016); Schaeffer, PRS A (2017); Rudy et al. Science (2017)

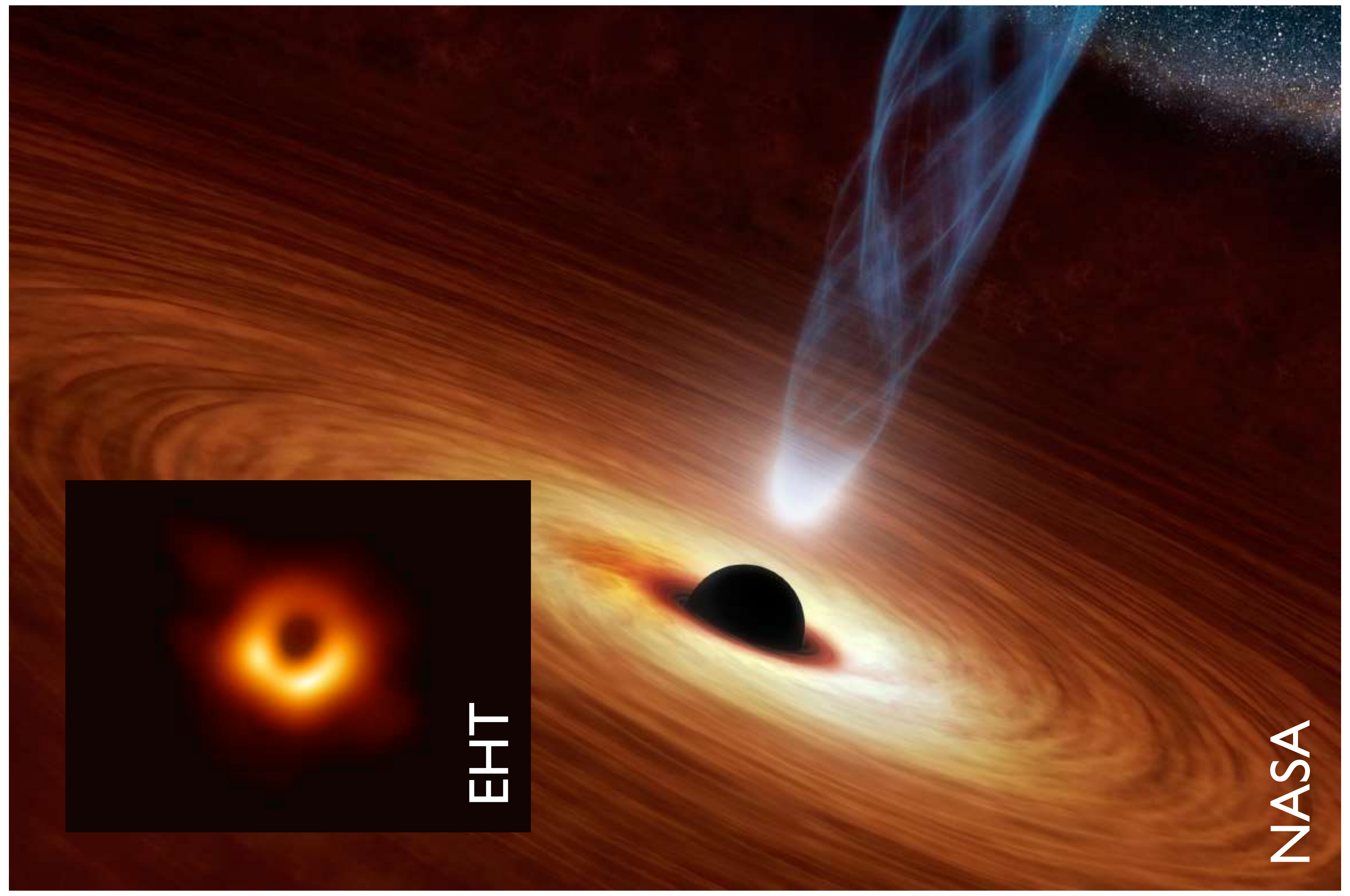
Ability to efficiently handle multi-dimensional, multi-variate data makes sparse regression a potentially suitable approach for complex nonlinear dynamics, such as in plasma physics problems

Plasmas are ubiquitous and intrinsically multi-scale systems

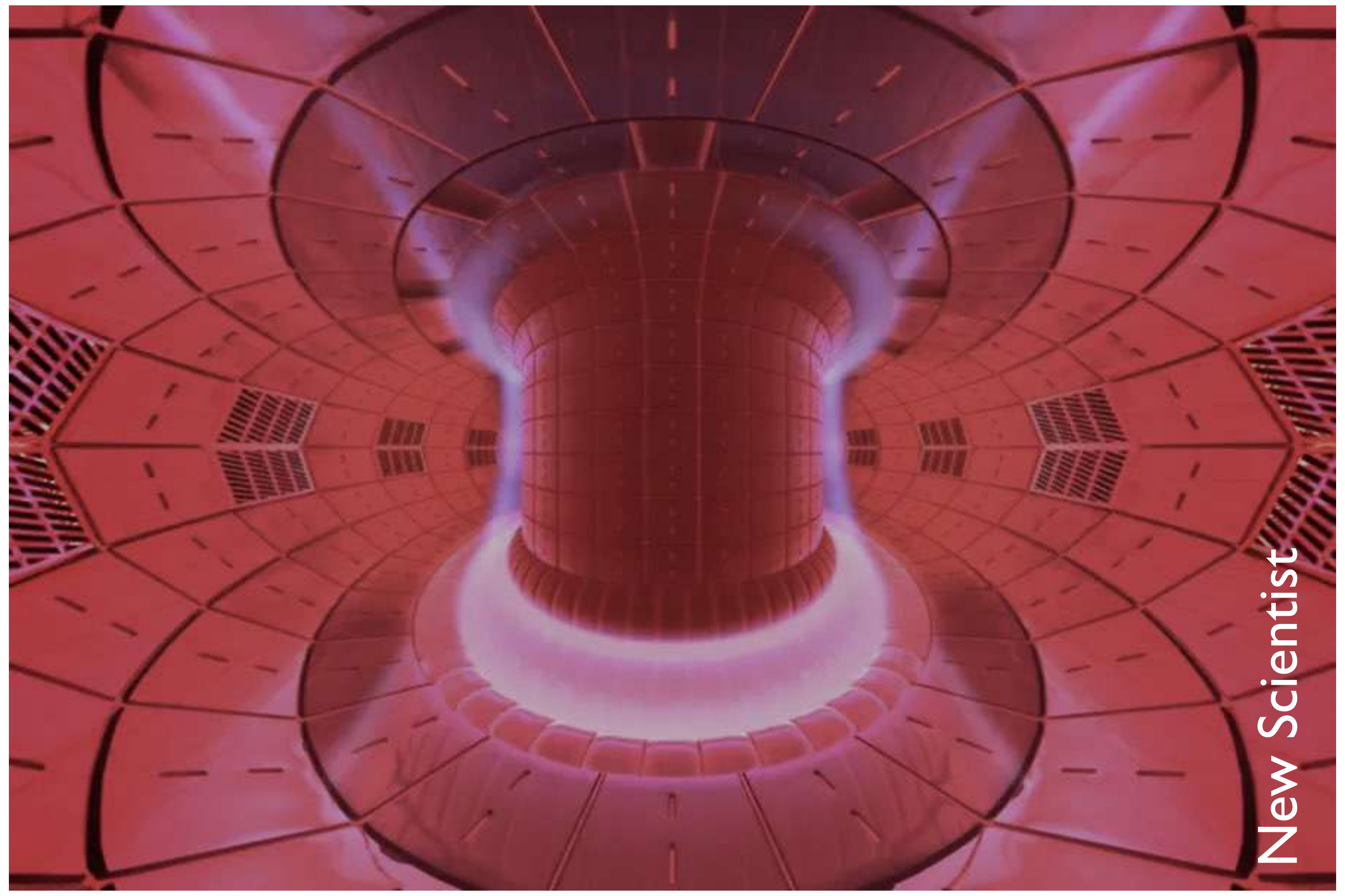
99.9% of visible universe is plasma



Extreme plasmas near black holes



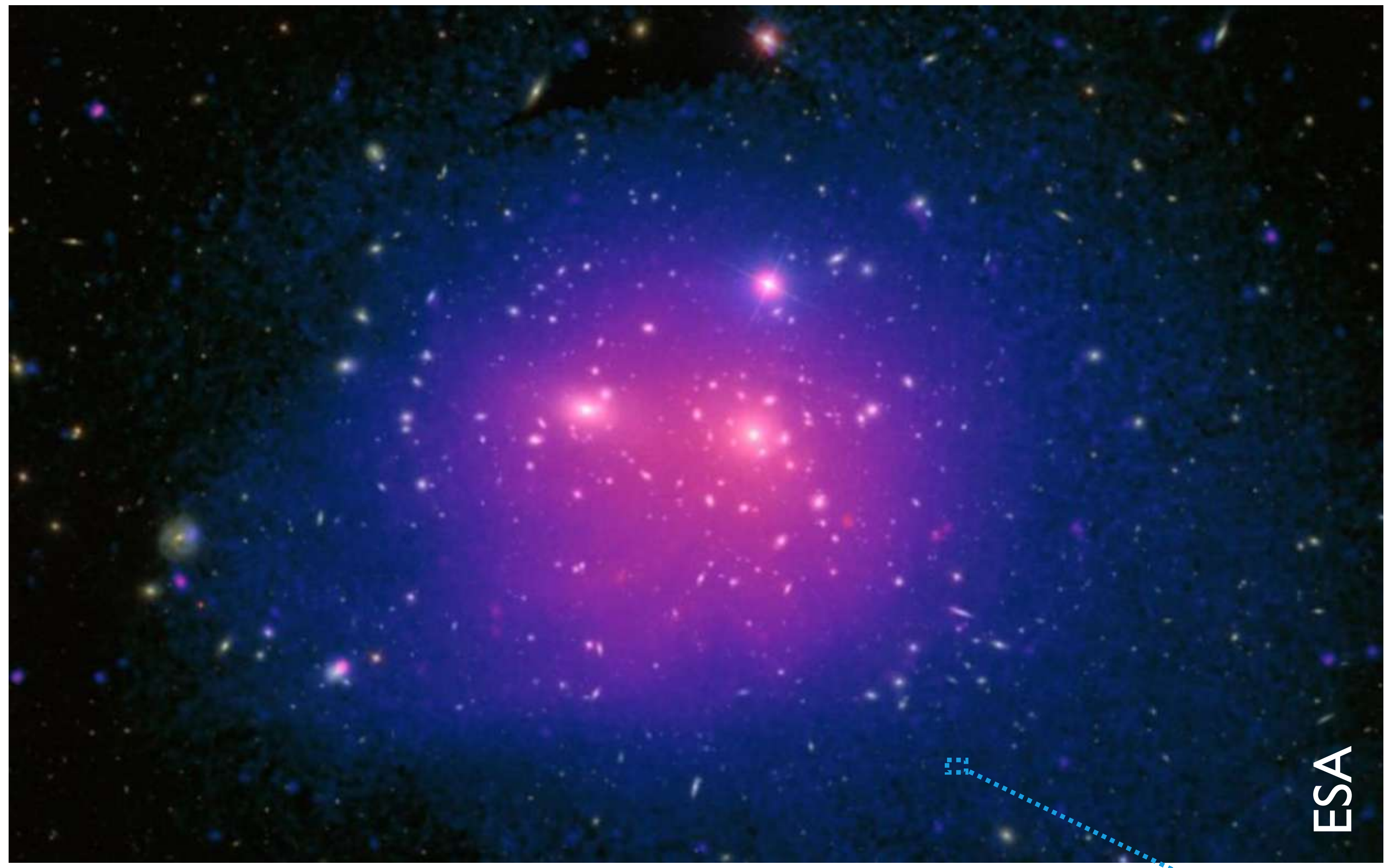
Fusion plasmas



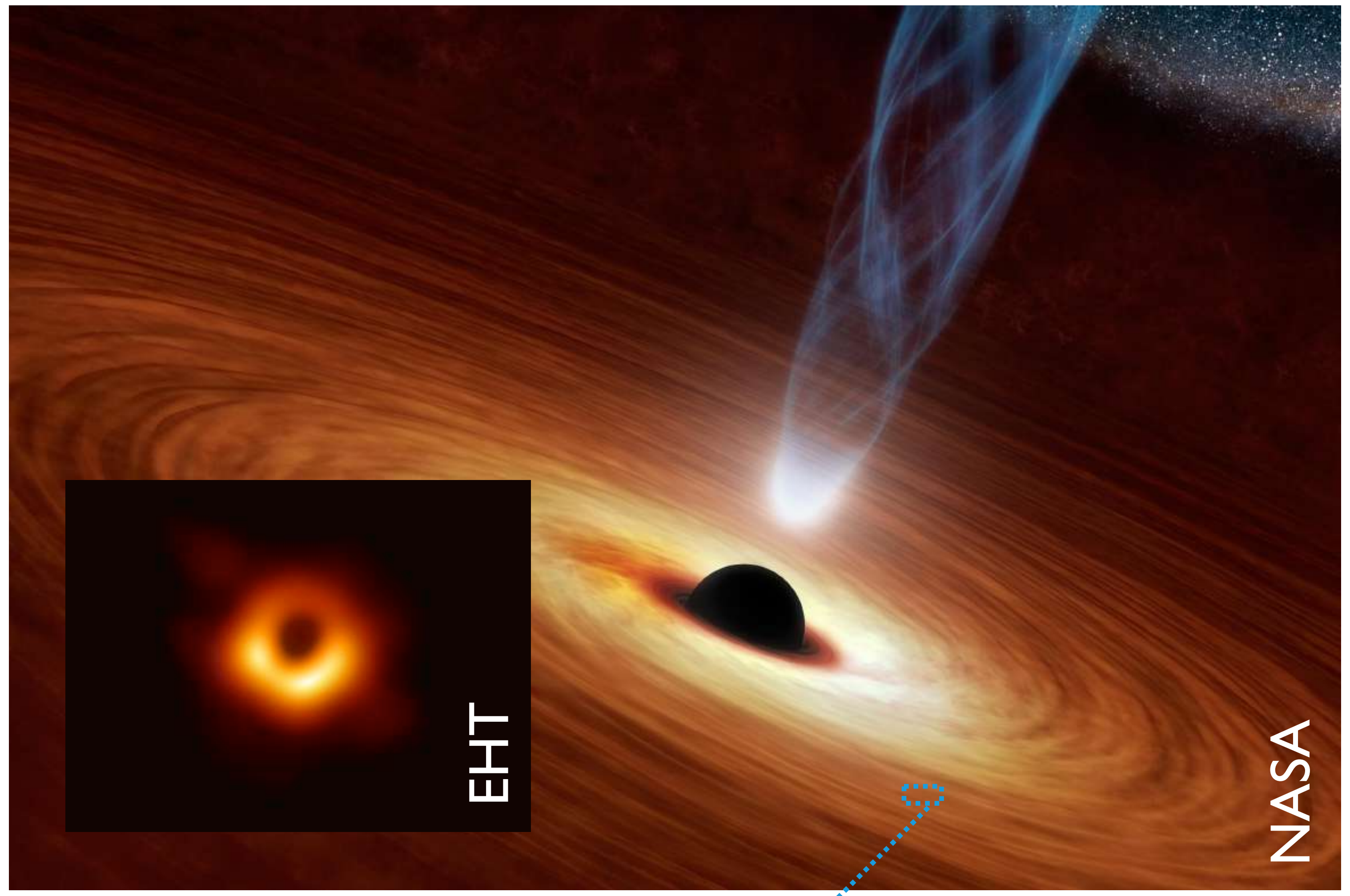
Plasmas are ubiquitous and intrinsically multi-scale systems

TÉCNICO LISBOA

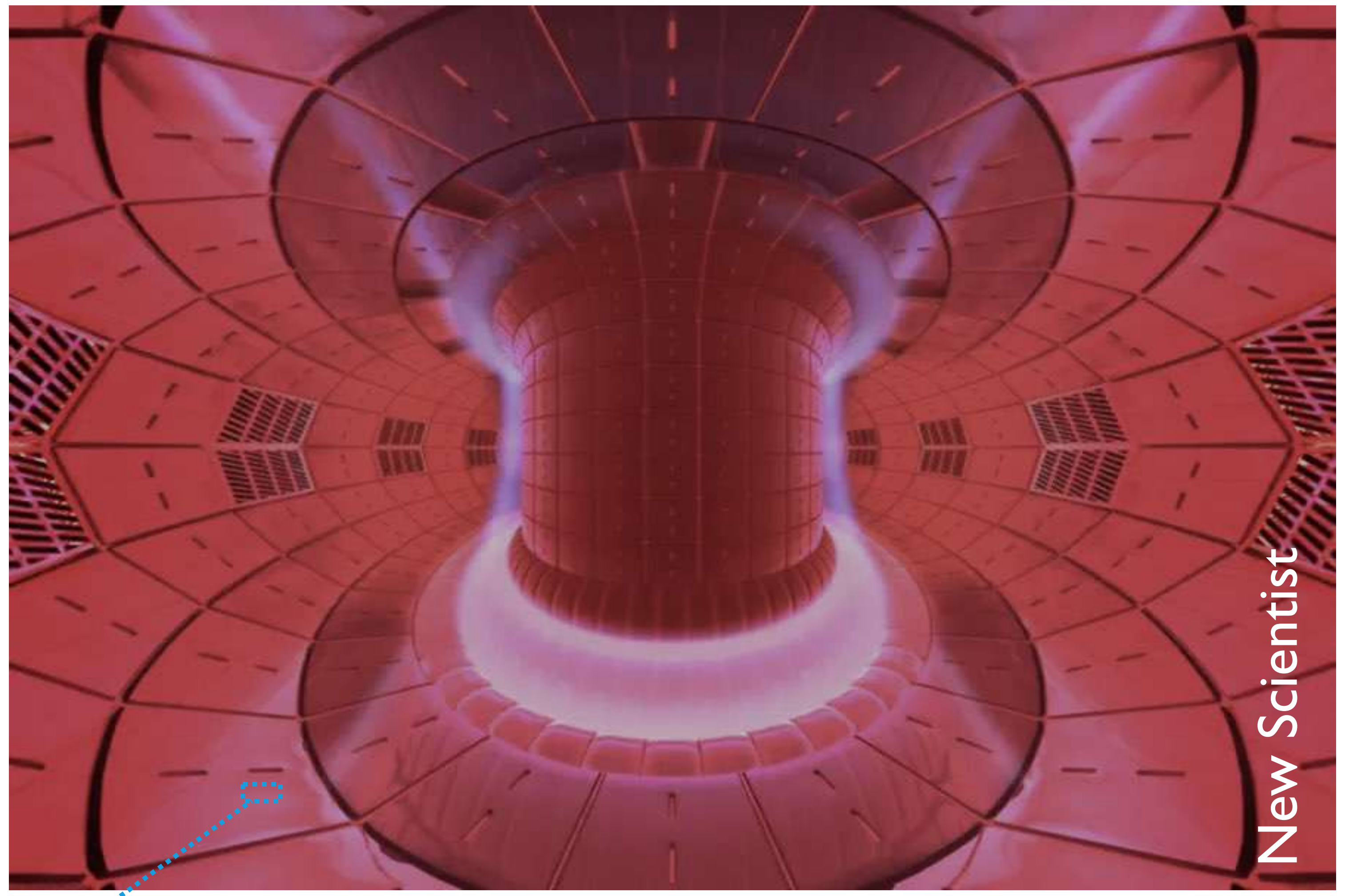
99.9% of visible universe is plasma

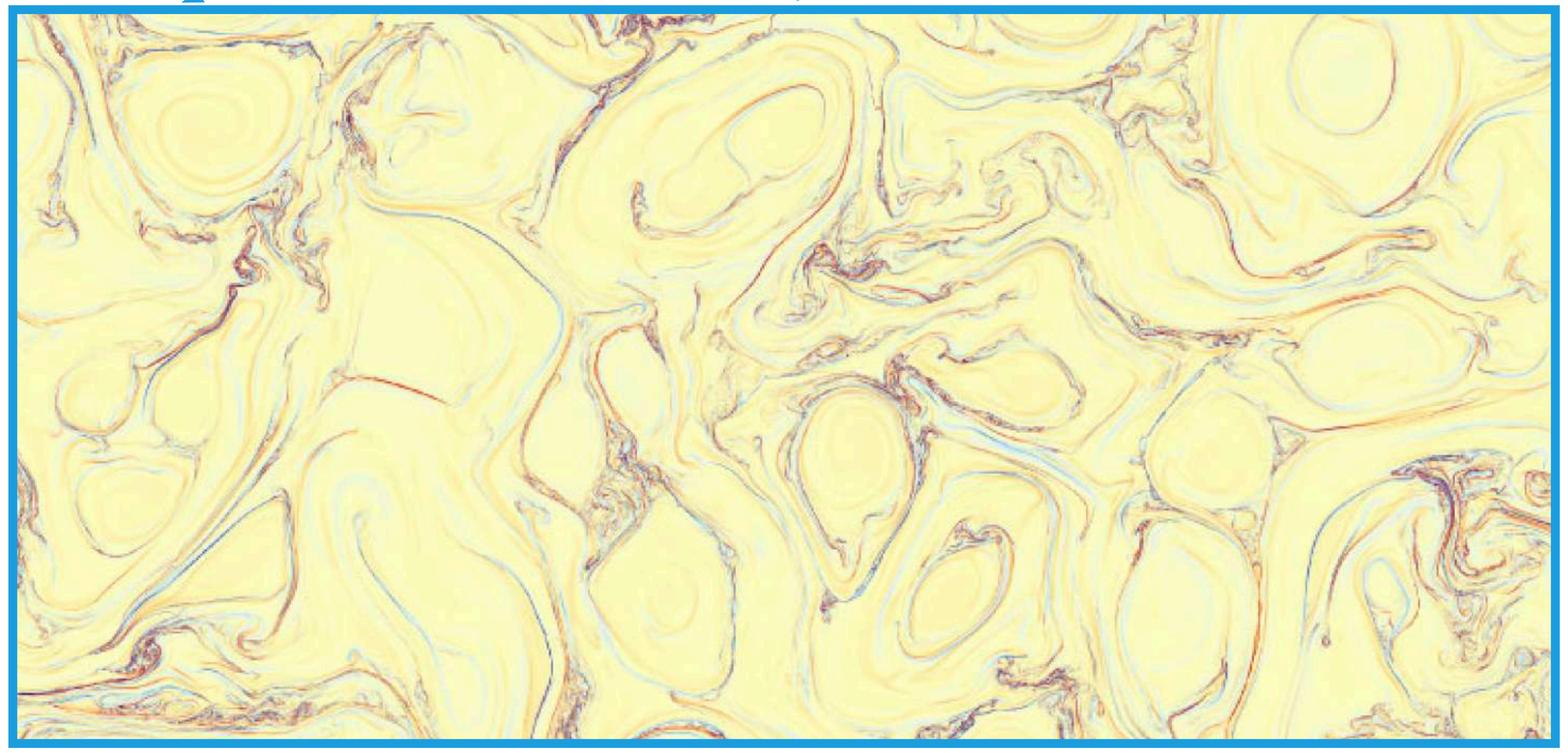


Extreme plasmas near black holes



Fusion plasmas

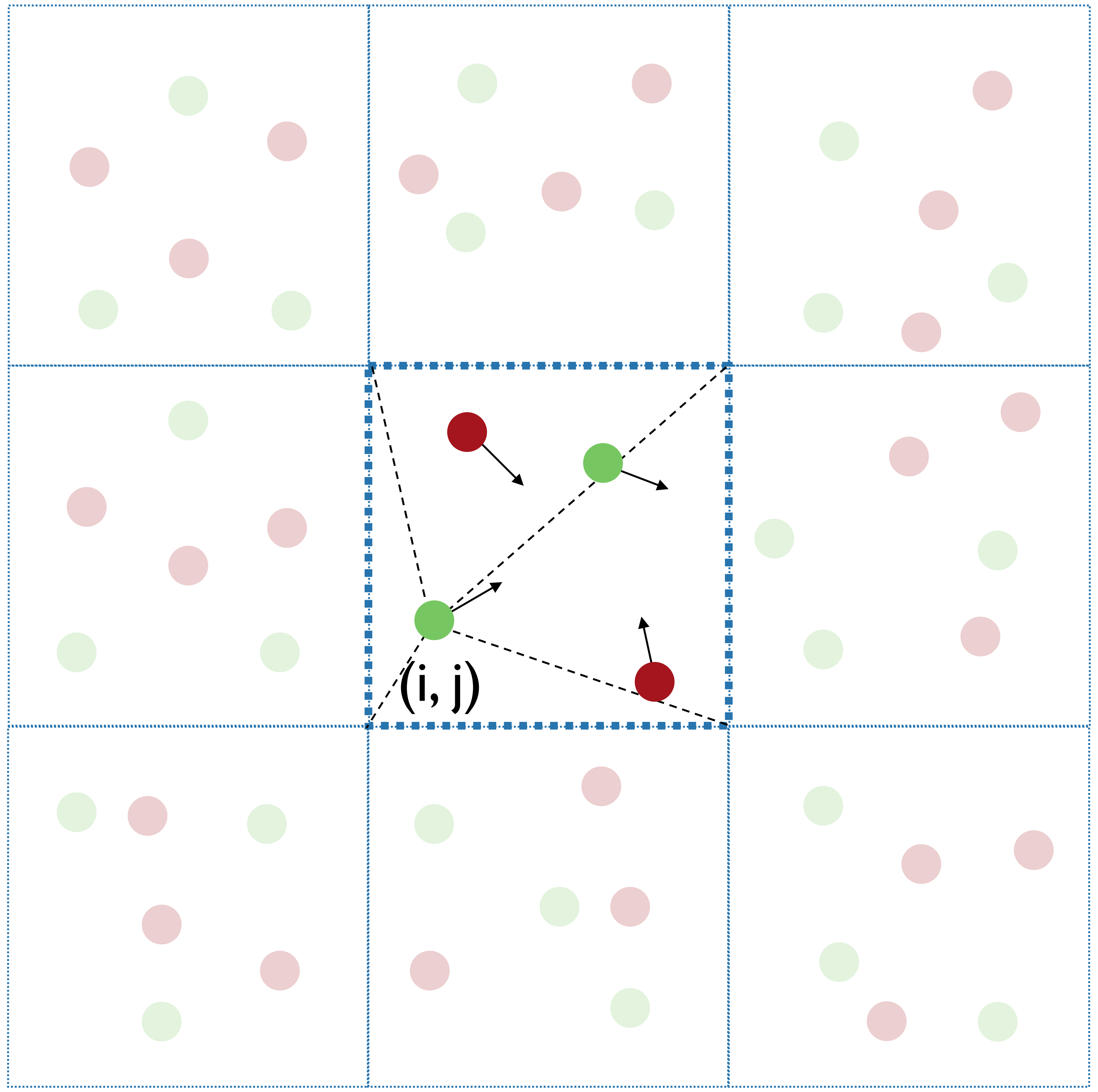




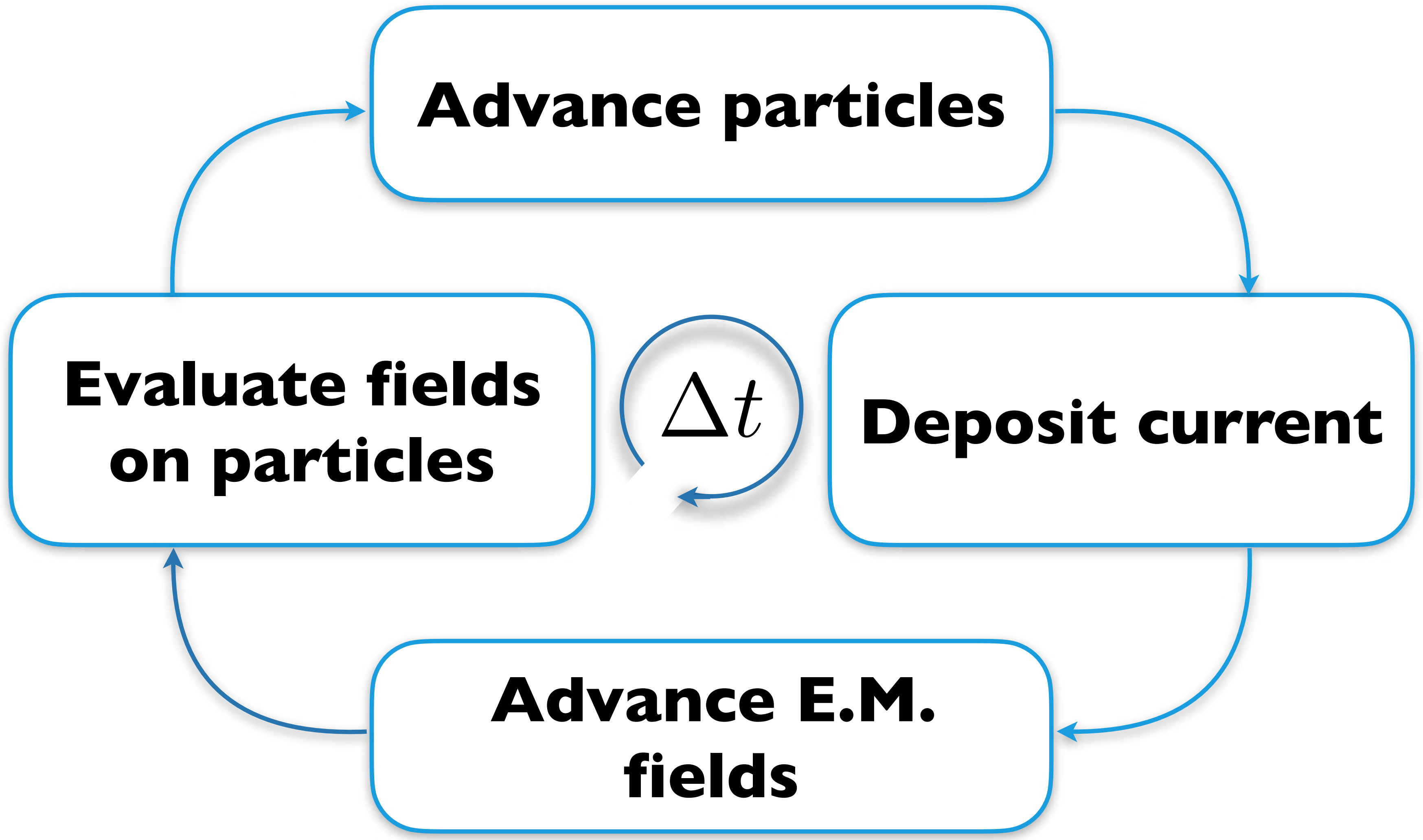
Dong et al. 2018

Data-driven discovery of reduced models from first-principles simulations

Particle-in-cell (PIC) method provides first-principles description of a plasma



$$\frac{d\mathbf{p}}{dt} = q\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\right)$$



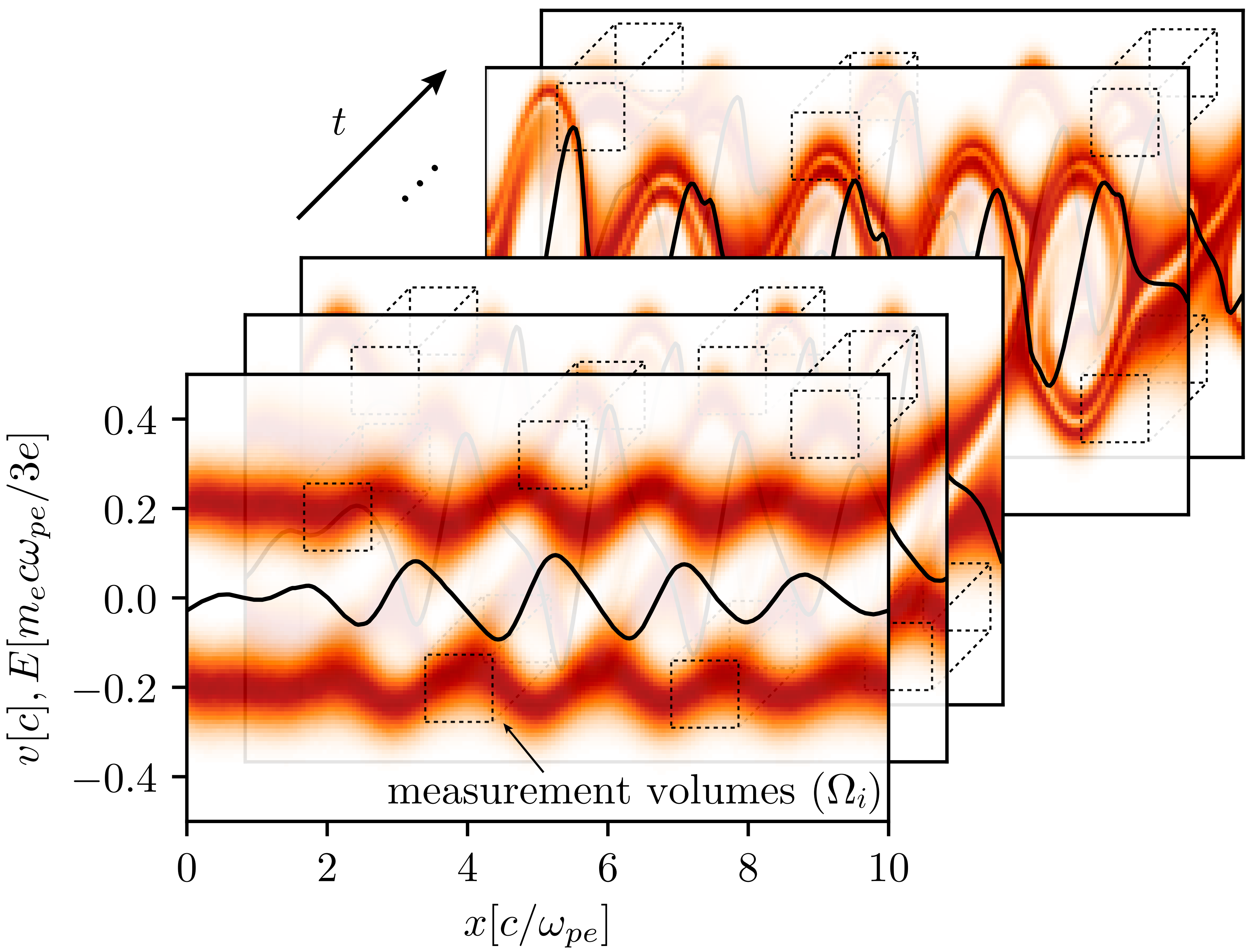
$$\frac{\partial \mathbf{E}}{\partial t} = c\nabla \times \mathbf{B} - 4\pi \mathbf{j}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E}$$

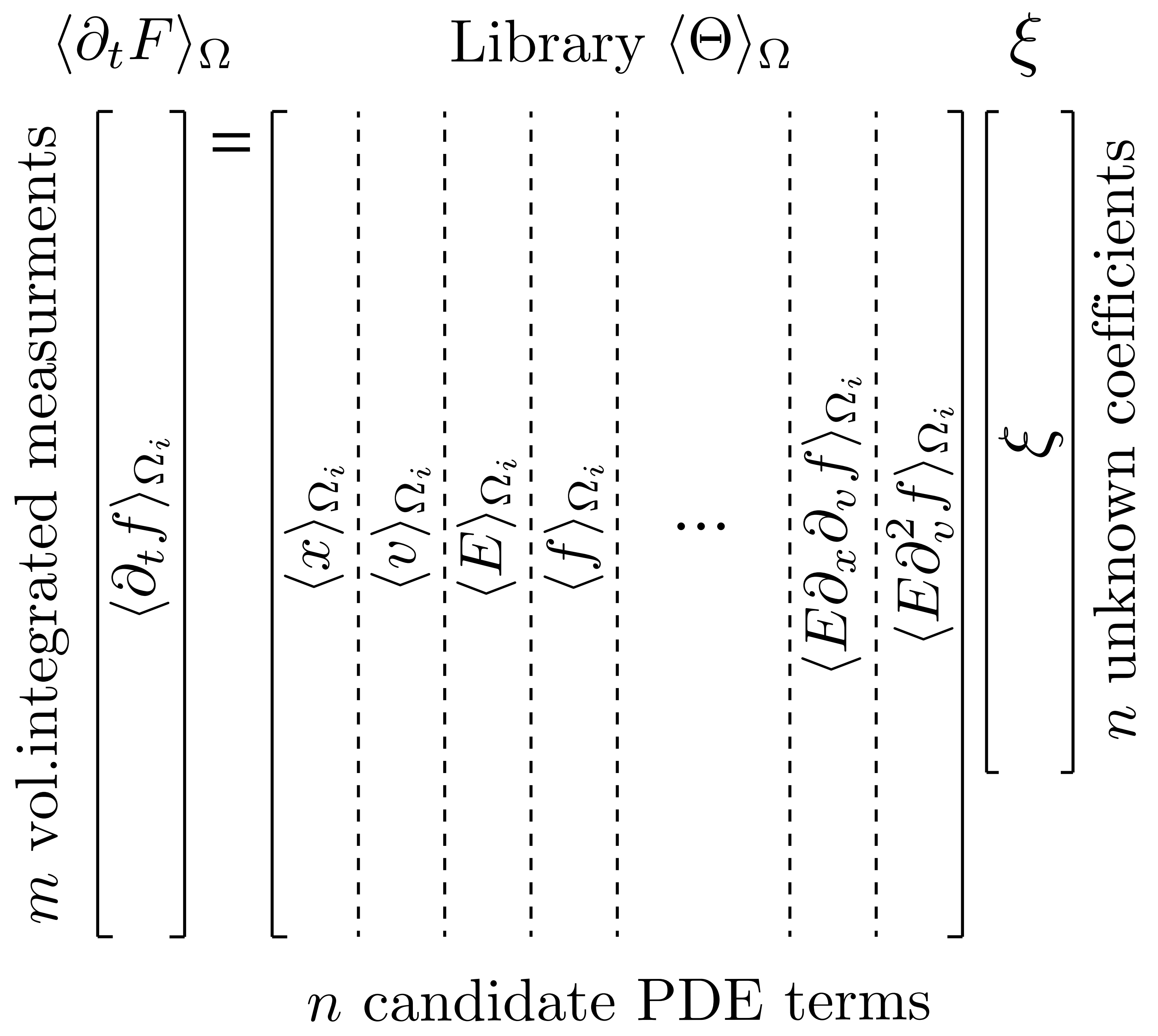
Can we "discover" reduced plasma models using sparse regression?

TÉCNICO LISBOA

1. Sampling measurement volumes from data



2. PDE discovery as a regression problem



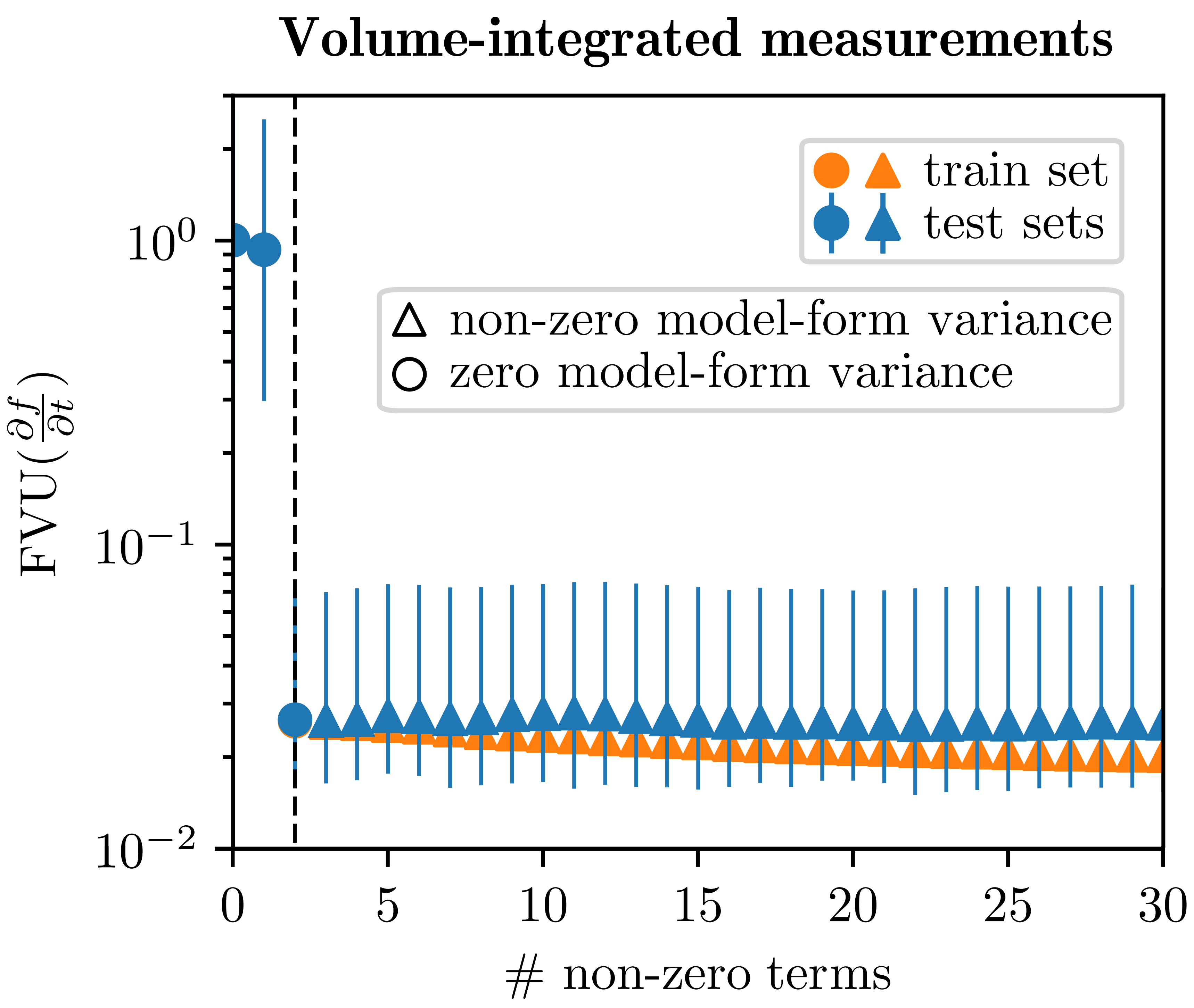
3. Solve sparse optimization problem

$$\text{argmin}_\xi ||\langle \partial_t F \rangle_\Omega - \langle \Theta \rangle_\Omega \xi||_2^2 + \lambda ||\xi||_0$$

E. P. Alves and F. Fiuza, Phys. Rev. Research 4, 033192 (2022)

Kinetic Vlasov equation robustly recovered

Pareto analysis reveals steep Pareto-front



Accurate recovery of the underlying PDE

Inferred PDE:

$$\partial_t f = -1.006 v \partial_x f + 0.997 E \partial_v f$$

True Vlasov eq.:

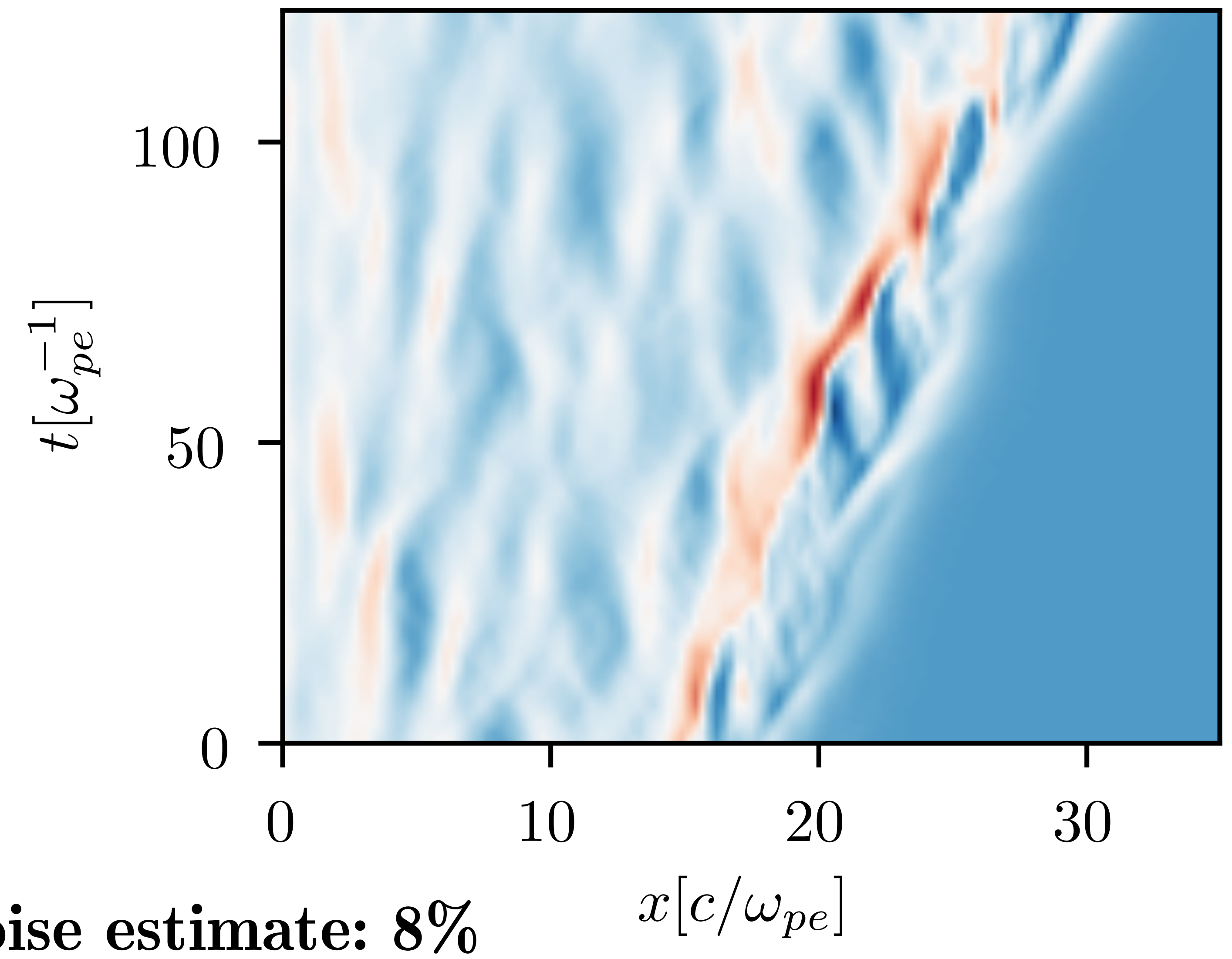
$$\partial_t f = -v \partial_x f - \frac{q}{m} E \partial_v f$$

- <1% error in inferred coefficients
- **Vlasov equation was not directly solved in simulation**

E. P. Alves and F. Fiuza, Phys. Rev. Research 4, 033192 (2022)

MHD equations "discovered" from fully-kinetic particle data

TÉCNICO LISBOA

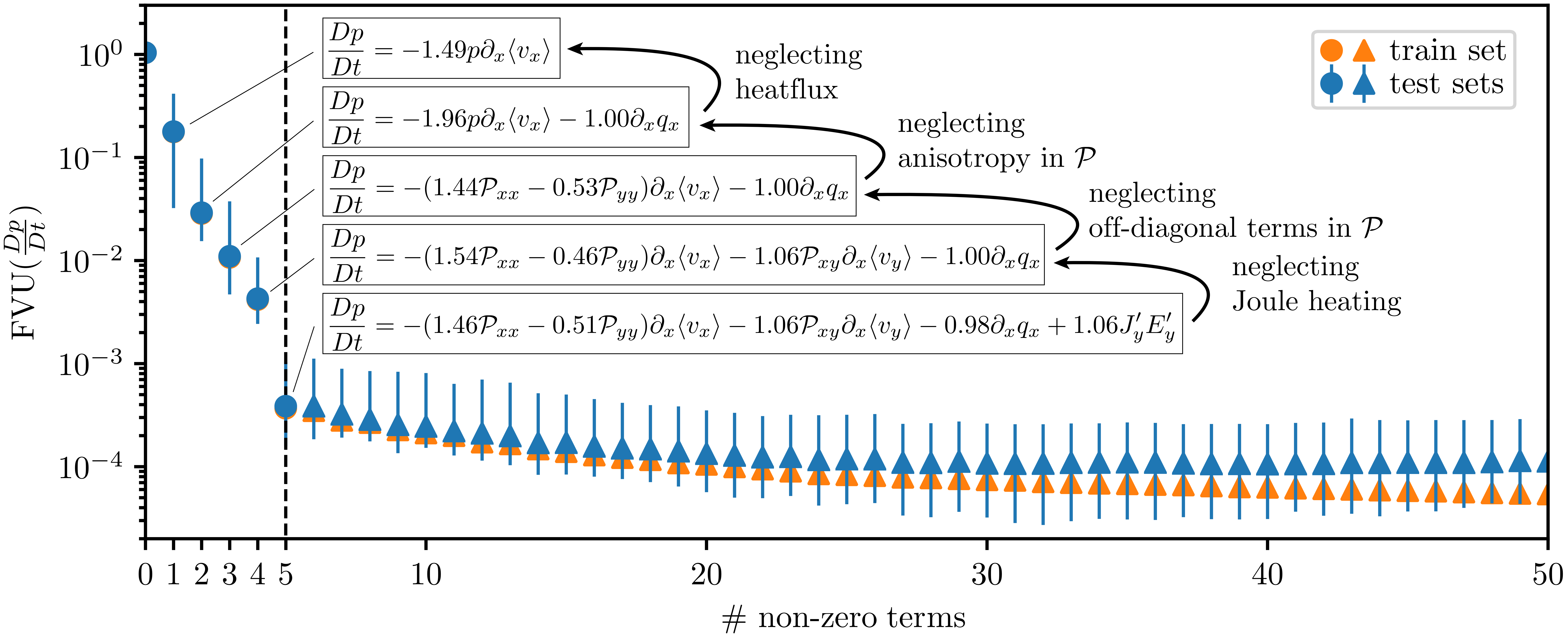
	Data	PDEs	PDE identification and inference accuracy: a) Point strategy	PDE identification and inference accuracy: b) Integral strategy
Single-fluid equations	**Collisionless magnetized shock dynamics**  **noise estimate: 8%**	**Continuity eq.** $\partial_t \rho_m = -\nabla \cdot (\rho_m \langle \mathbf{v} \rangle)$	Unsucessful identification	**Sucessful identification** **Mean coeff. error: 1%**
		Momentum eq. $\partial_t (\rho_m \langle \mathbf{v} \rangle) = -\nabla \cdot (\rho_m \langle \mathbf{vv} \rangle) + \rho_c \mathbf{E} + \mathbf{J} \times \mathbf{B}$	Unsucessful identification	**Sucessful identification** **Mean coeff. error: 1%**
		Energy eq. $\frac{f}{2}\partial_t p = -\frac{f}{2}\nabla \cdot (p\langle \mathbf{v} \rangle) - (\mathcal{P} \cdot \nabla) \cdot \langle \mathbf{v} \rangle - \nabla \cdot \mathbf{q} + \mathbf{J}' \cdot \mathbf{E}'$	Unsucessful identification	**Sucessful identification** **Mean coeff. error: 3.7%**

E. P. Alves and F. Fiuza, Phys. Rev. Research 4, 033192 (2022)

Hierarchy of Pareto-optimal models can help identify suitable approximations

TÉCNICO LISBOA

Pareto analysis of the MHD energy equation from magnetized shock dynamics



E. P. Alves and F. Fiuza, Phys. Rev. Research 4, 033192 (2022)

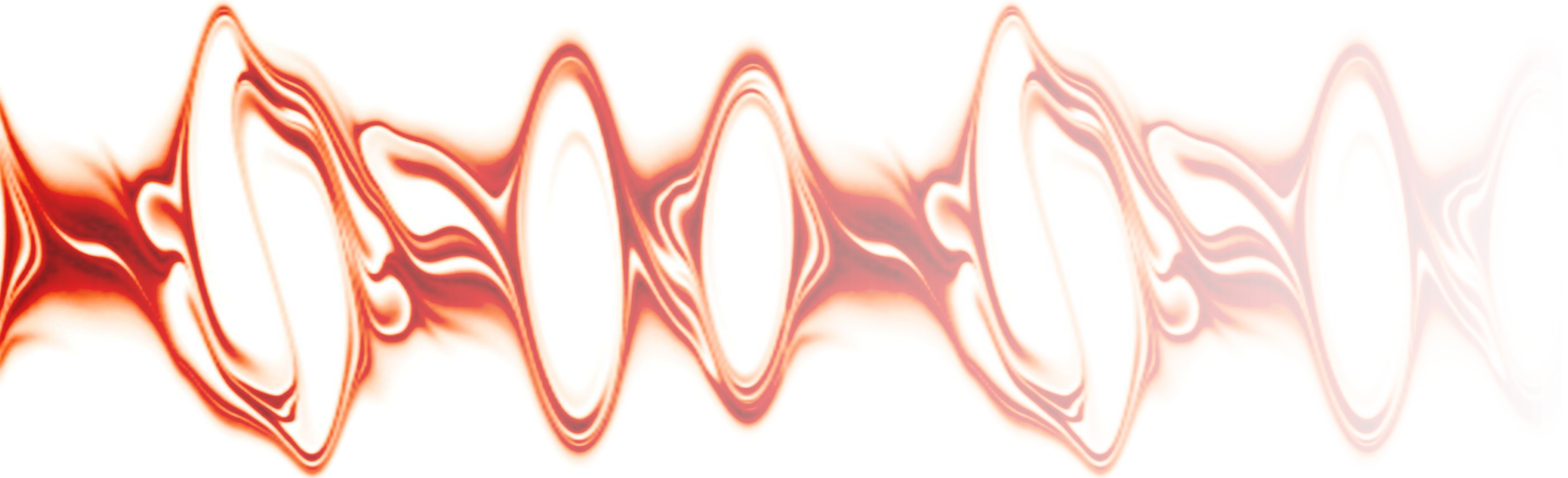
Take aways

AI / machine learning can be a powerful tool to discover governing equations from data:

Takes advantage of increasingly large datasets available

Interpretable tools, such as sparse regression, promote scientific insight and can stimulate theoretical efforts to "reverse engineer" learned equations from lower-level frameworks

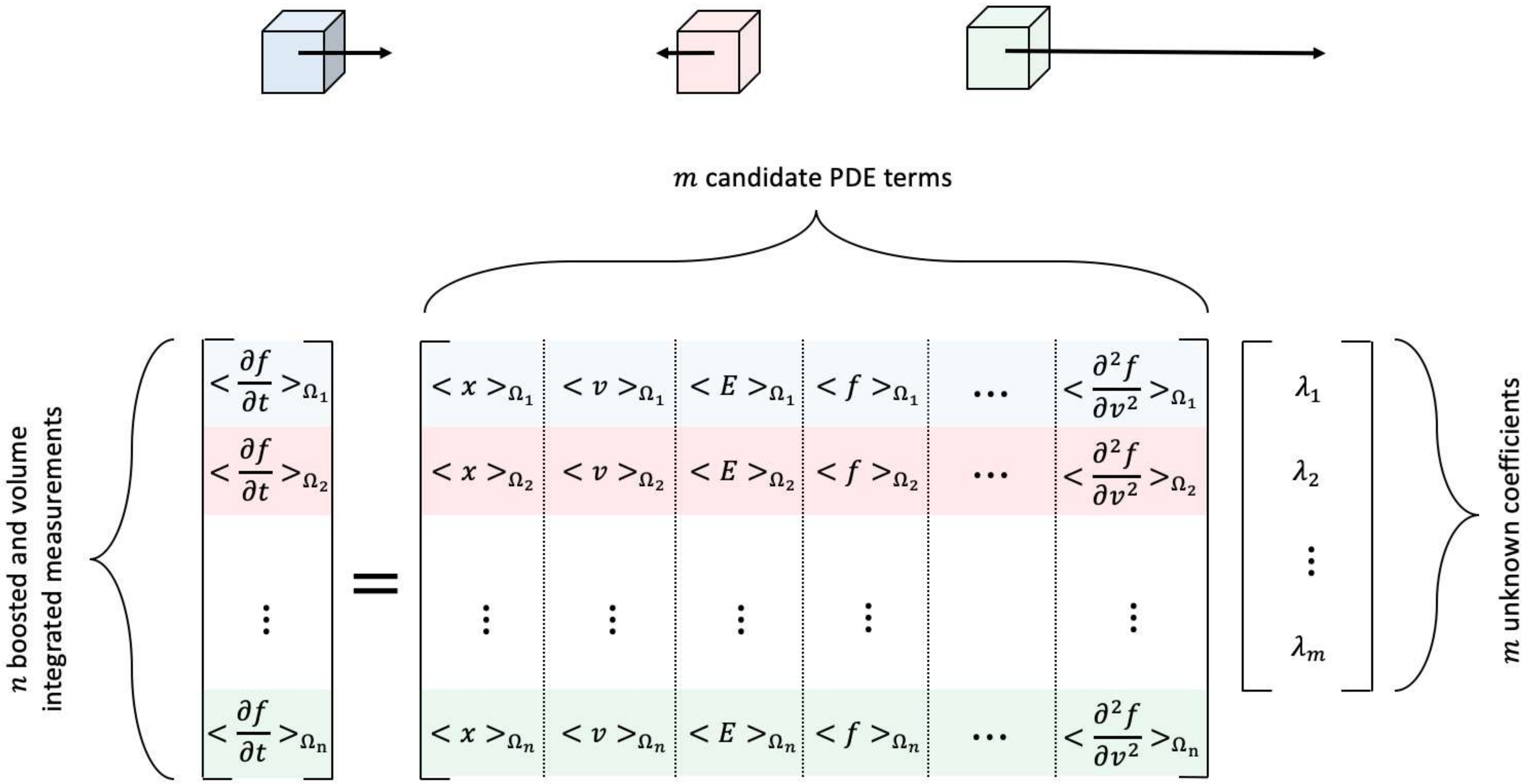
Can accelerate development of reduced models for different multi-scale, nonlinear systems



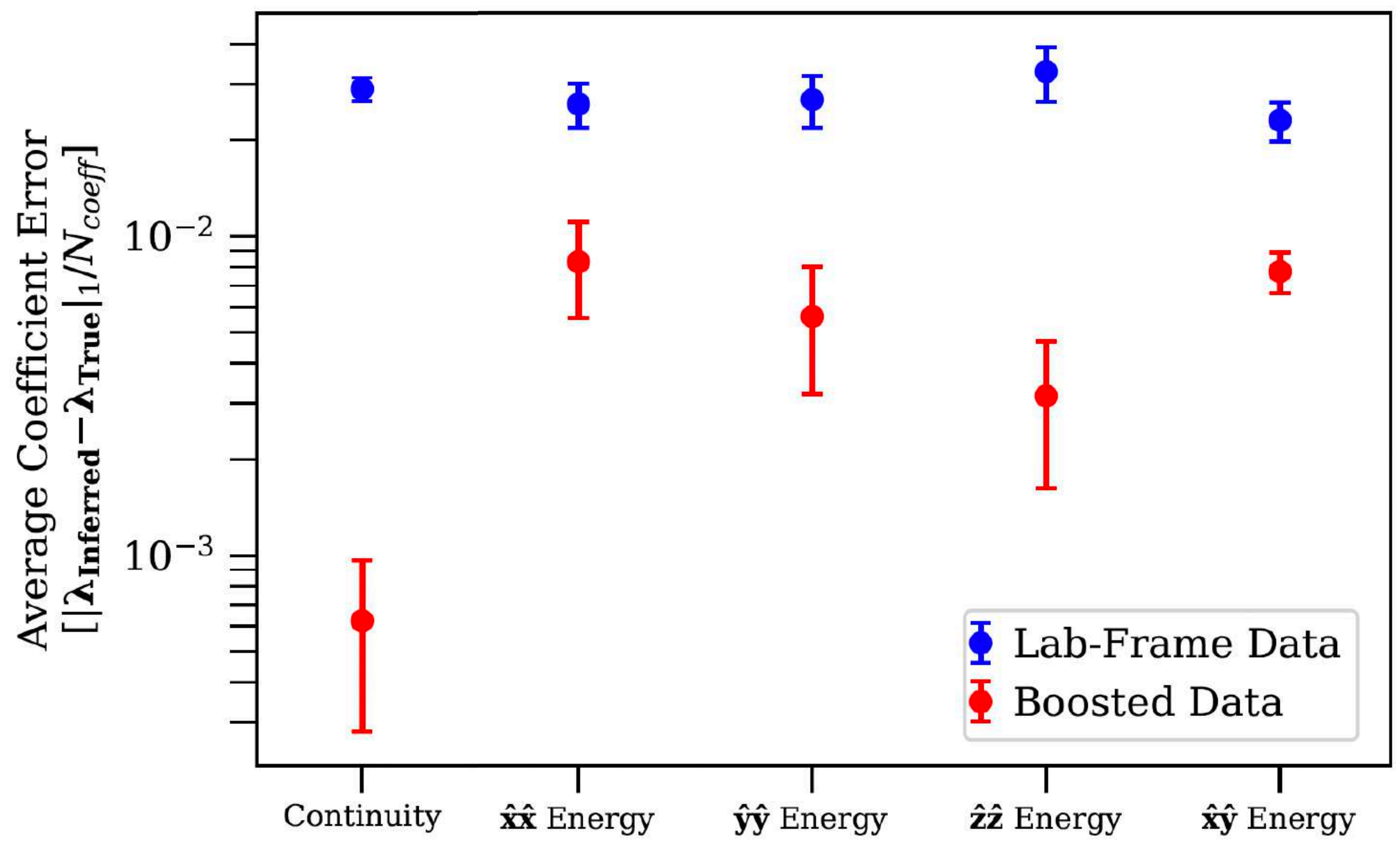
Embedding conservation laws in reduced models

Example: Lorentz covariance

Different colors indicate quantities observed from different Lorentz-boosted frames



Large reduction in coefficient errors

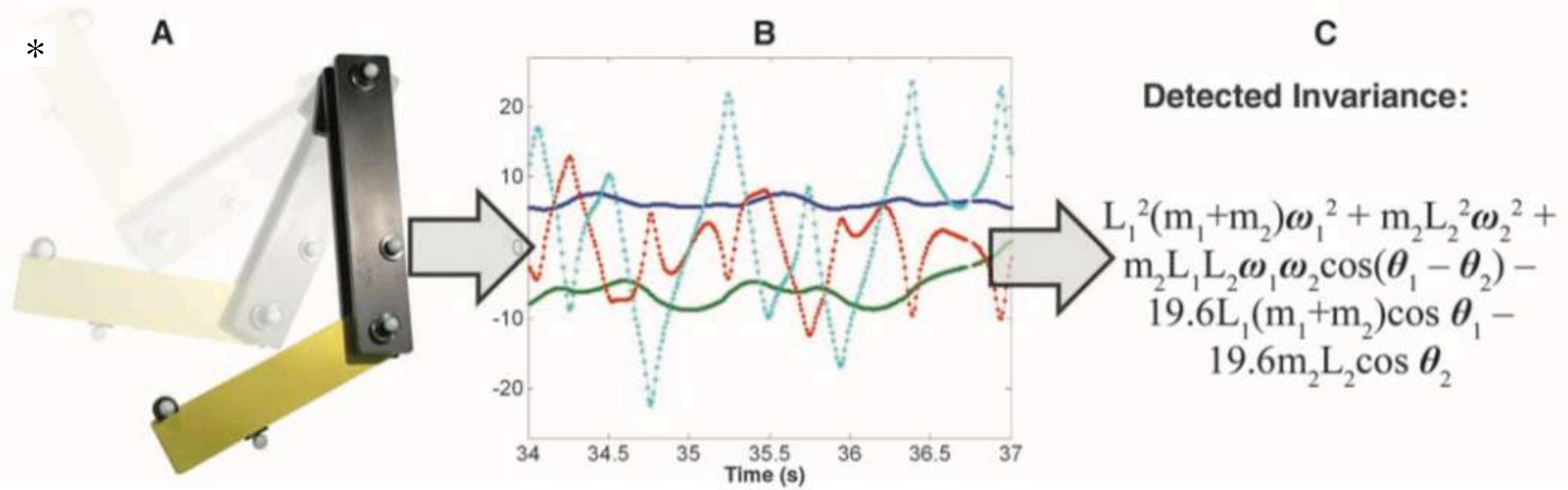


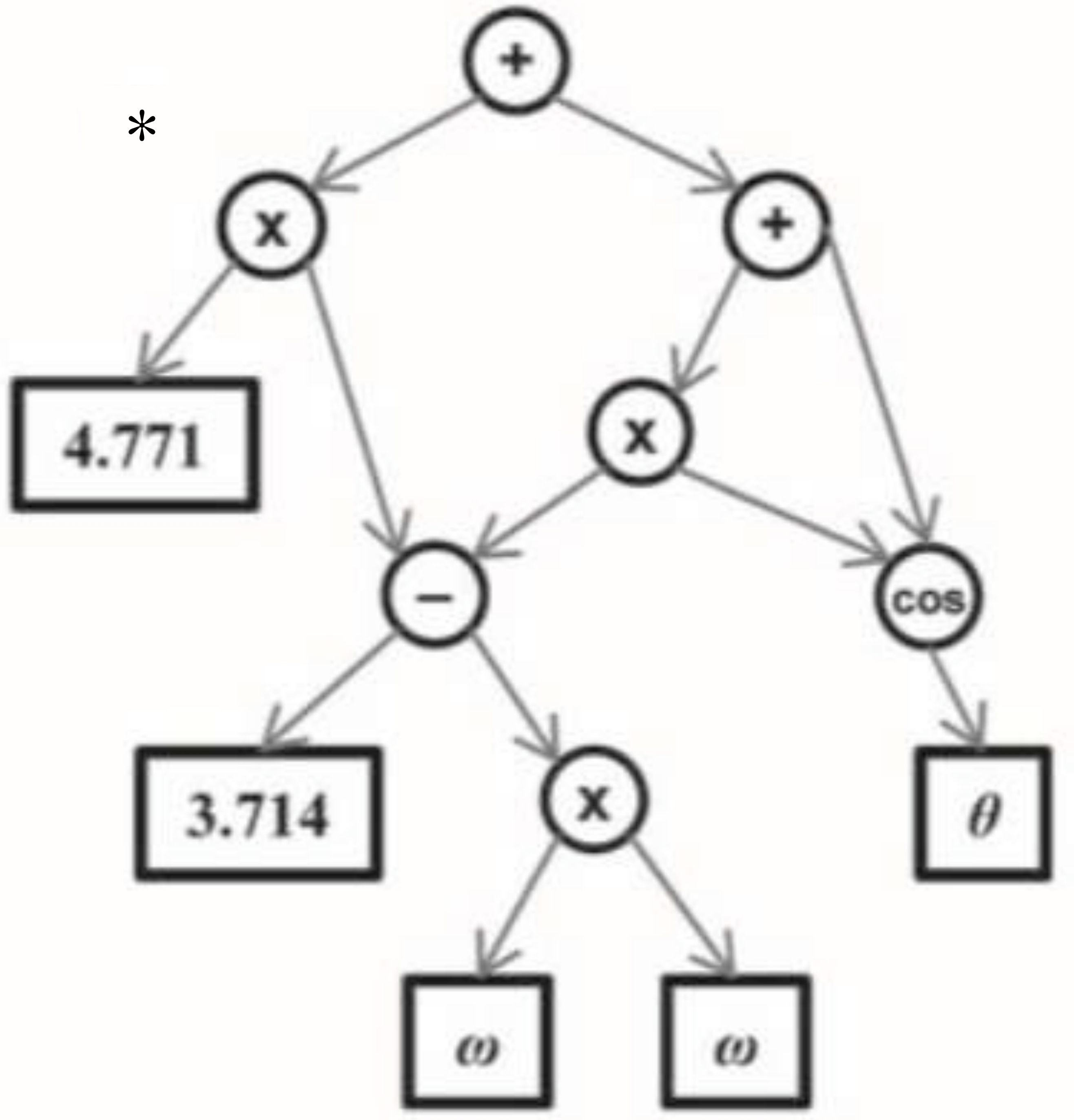
M. McGrae-Menge *et al, arXiv:2506.14048* (2025)

Can we use AI to discover governing equations directly from data?

Symbolic regression methodology

- Candidate symbolic expressions generated via genetic programming
- Expressions are evaluated on the data
- Evolutionary algorithms to optimize expression accuracy while balancing complexity





Bongard & Lipson, PNAS (2007)
*Schmidt & Lipson, Science (2009)

Combinatorially large search space and does not scale well to multi-variate high-dimensional systems