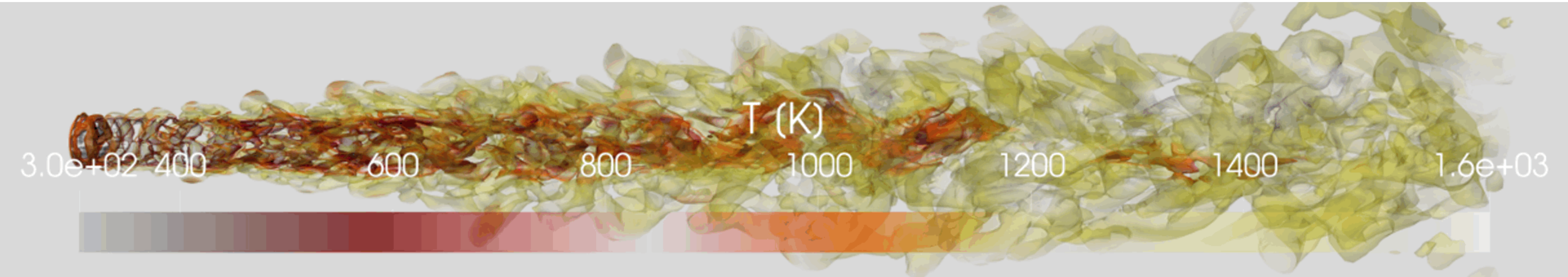


CYPHER data challenge: Benchmarking ML-enhanced turbulent combustion closures

Lorenzo Piu, Pasquale Eduardo Lapenna, Tamara Osseily, Giuseppe Indelicato, Kosuke Shigematsu, Albina Tocilla, Jonid Kazani, Antonio Attili, Alessandro Parente

ULB UNIVERSITÉ LIBRE DE BRUXELLES
THE UNIVERSITY of EDINBURGH
CYPHER
SAPIENZA UNIVERSITÀ DI ROMA
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY
BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación
UNIVERSITETI POLIS
独立行政法人国立高等専門学校機構 大分工業高等専門学校 National Institute of Technology, Oita College
Funded by the European Union

CYPHER Cost Action objective

“ Propel the **collaborations** between European researchers and industrial stakeholders to foster the use of **digital-physical** systems and ultimately promote a safe and sustainable adoption of **renewable synthetic fuels** to decarbonise **energy-intensive industries**”

CYPHER

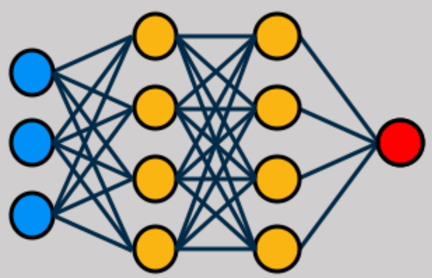
CYPHER Cost Action objective

" Propel the **collaborations** between European researchers and industrial stakeholders to foster the use of **digital-physical** systems and ultimately promote a safe and sustainable adoption of **renewable synthetic fuels** to decarbonise **energy-intensive industries**"

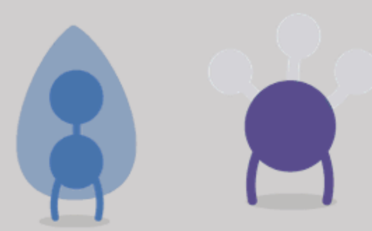
CYPHER

1. Collaboration

2. Machine Learning



3. Renewable fuels



4. Energy-intensive industries

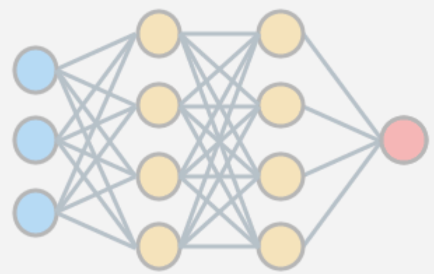
CYPHER Cost Action objective

" Propel the **collaborations** between European researchers and industrial stakeholders to foster the use of **digital-physical** systems and ultimately promote a safe and sustainable adoption of **renewable synthetic fuels** to decarbonise **energy-intensive industries**"

CYPHER

1. Collaboration

2. Machine Learning

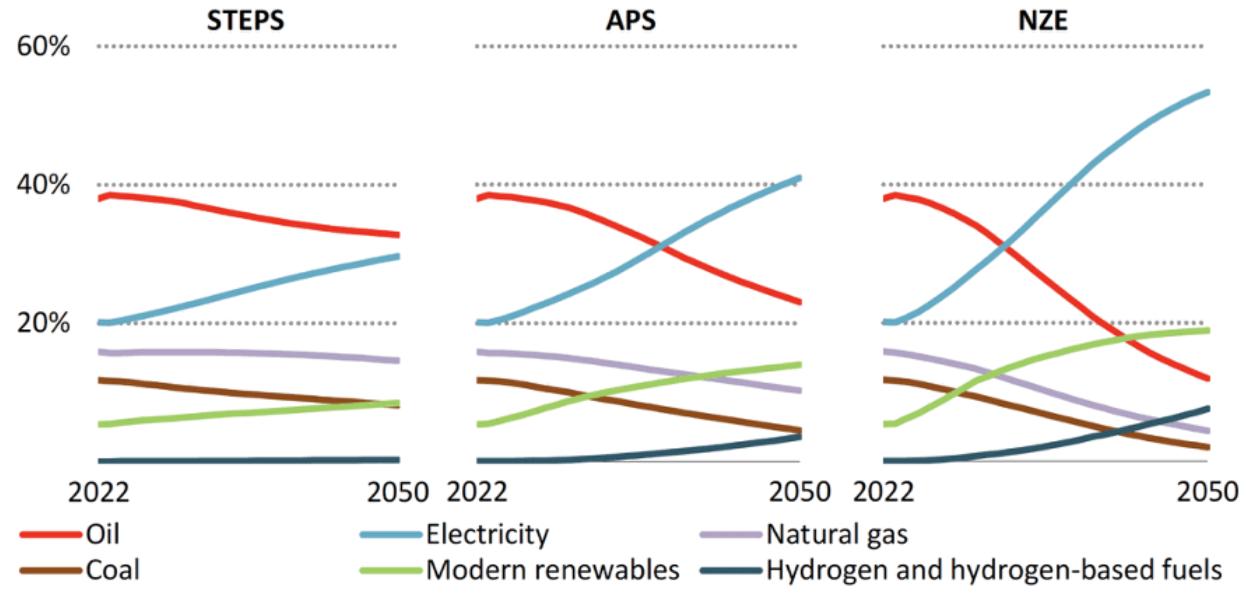


3. Renewable fuels

4. Energy-intensive industries

Combustion will play an important role in the energy transition

At least 40% of the world's energy supply will be based on hydrogen, hydrogen-based fuels, and natural gas [1].

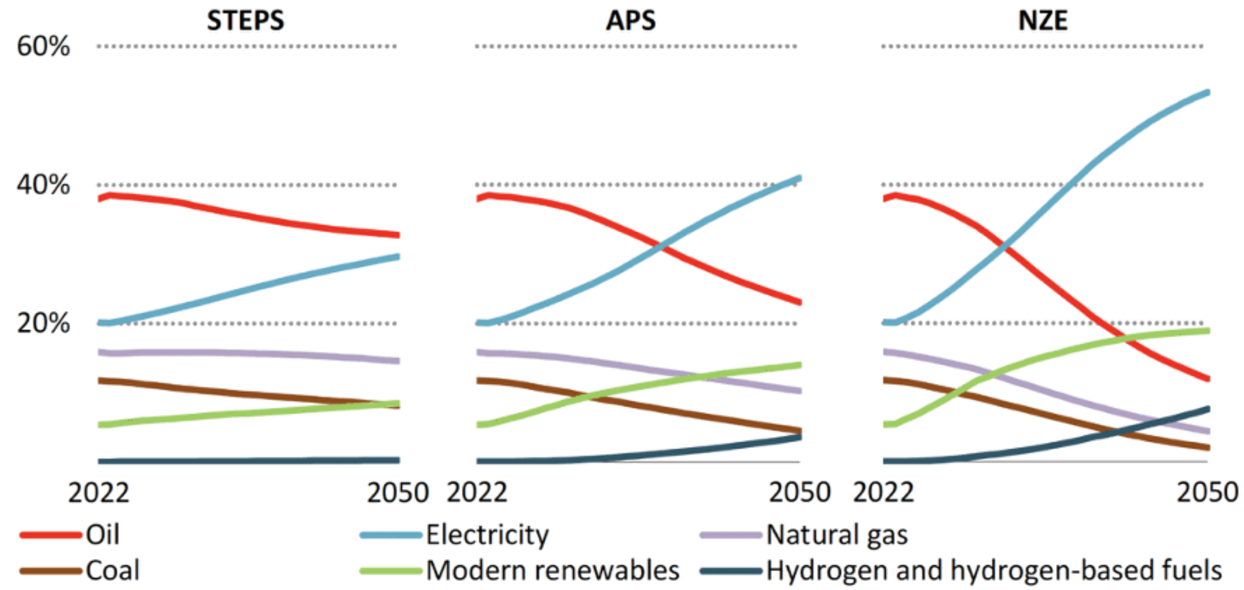


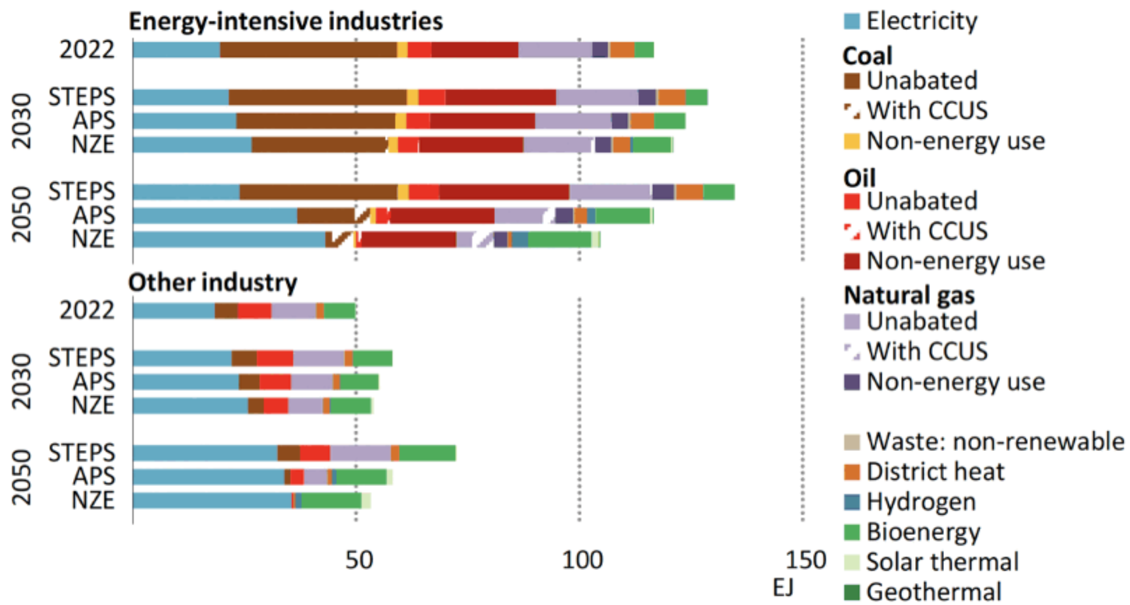
[1] Pitsch, H. (2024). *The transition to sustainable combustion: Hydrogen- and carbon-based future fuels and methods for dealing with their challenges*. *Proceedings of the Combustion Institute*, **40**(1–4), Article 105638. https://doi.org/10.1016/j.proci.2024.105638

[2] U.S. Energy Information Administration, International energy outlook 2023, 2023, https://www.eia.gov/outlooks/ieo/.

Combustion will play an important role in the energy transition

At least 40% of the world's energy supply will be based on hydrogen, hydrogen-based fuels, and natural gas [1].





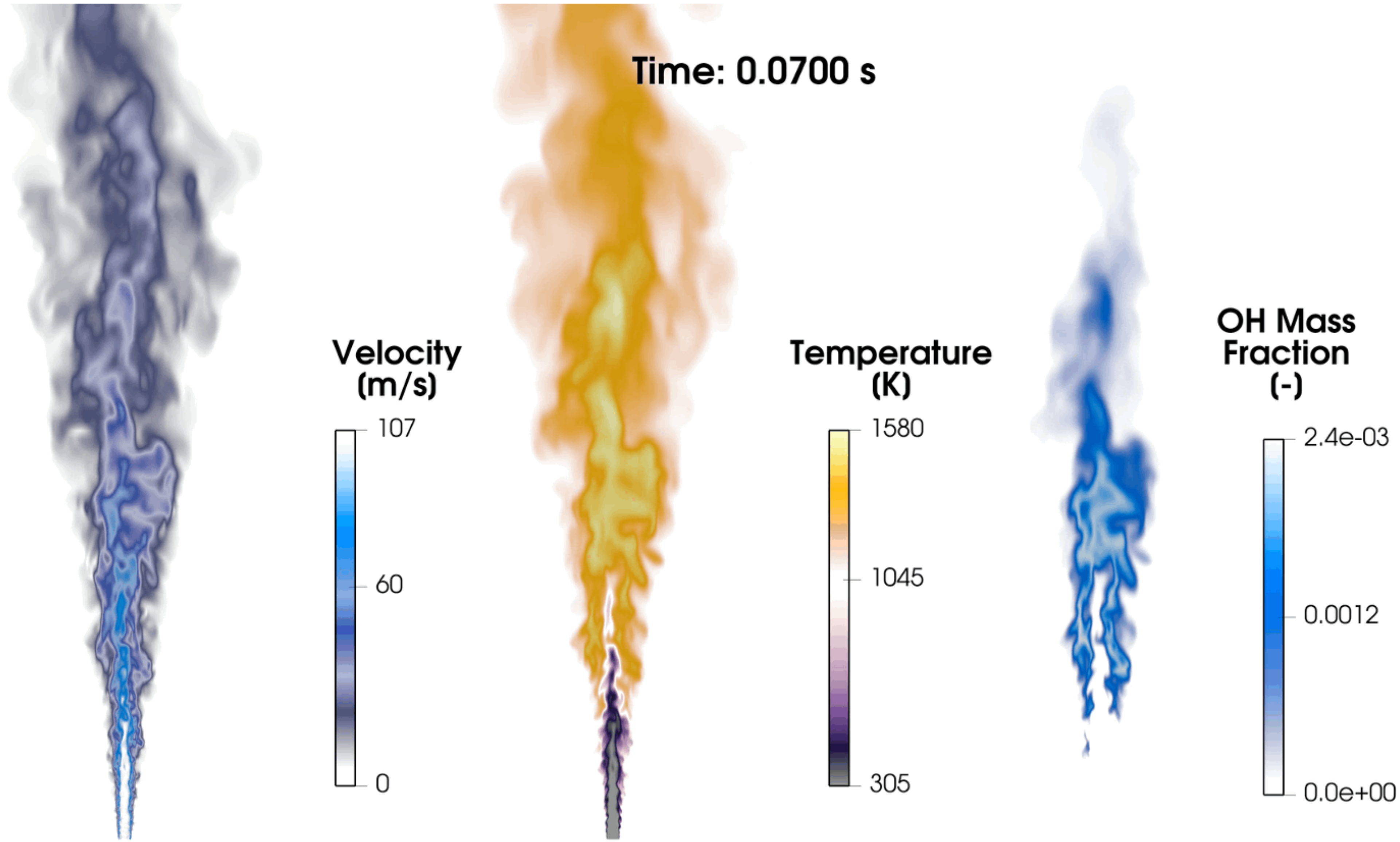
The vast majority of the energy request comes from the *hard-to-abate* or *energy-intensive* industries [2].

These industries need **high-temperature heat**, which is difficult to electrify and hence has to be provided in large parts by **combustion** processes.

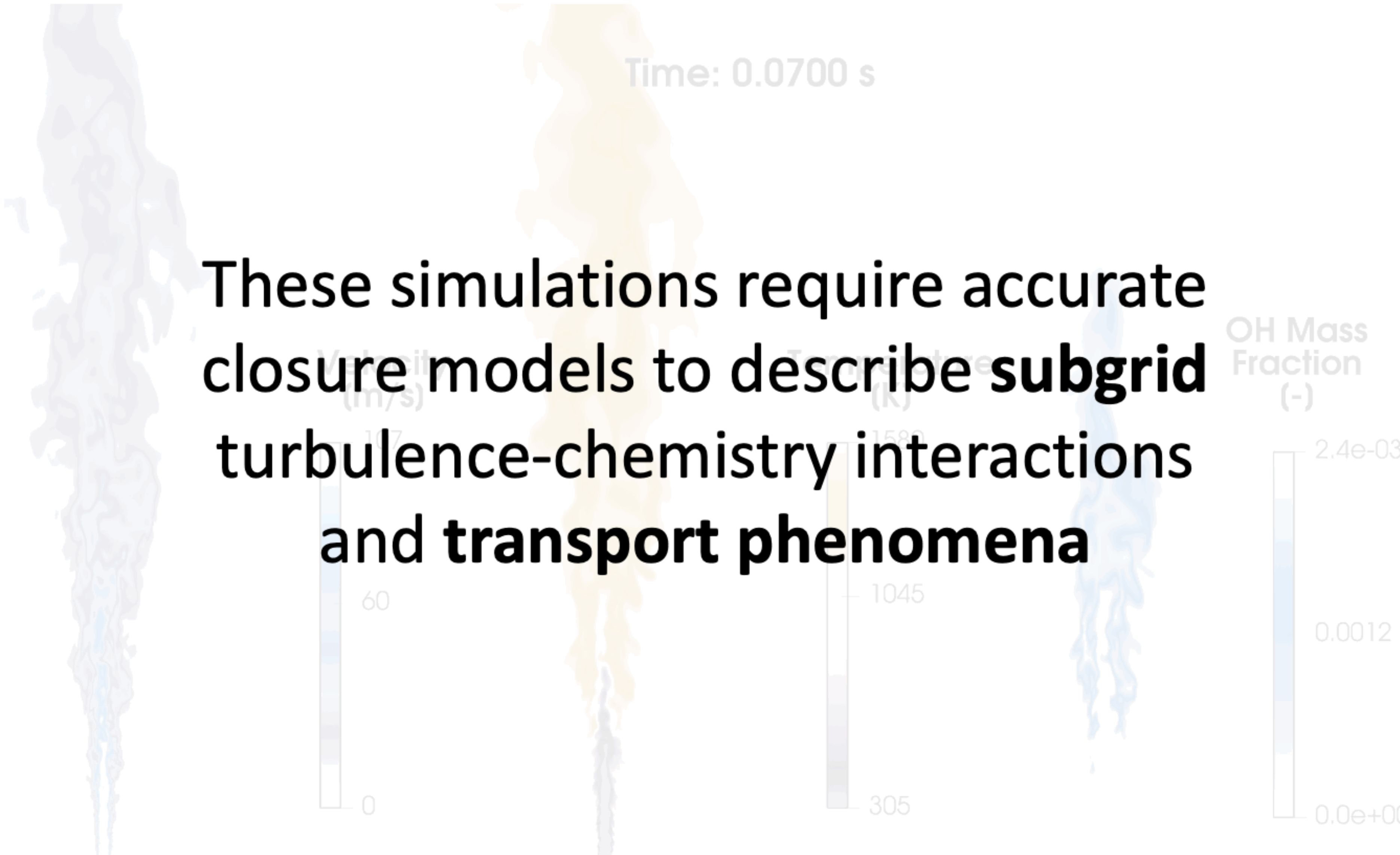
[1] Pitsch, H. (2024). *The transition to sustainable combustion: Hydrogen- and carbon-based future fuels and methods for dealing with their challenges*. *Proceedings of the Combustion Institute*, **40**(1–4), Article 105638. https://doi.org/10.1016/j.proci.2024.105638

[2] U.S. Energy Information Administration, International energy outlook 2023, 2023, https://www.eia.gov/outlooks/ieo/.

High-fidelity simulations are becoming a fundamental tool to gain insights into complex combustion systems

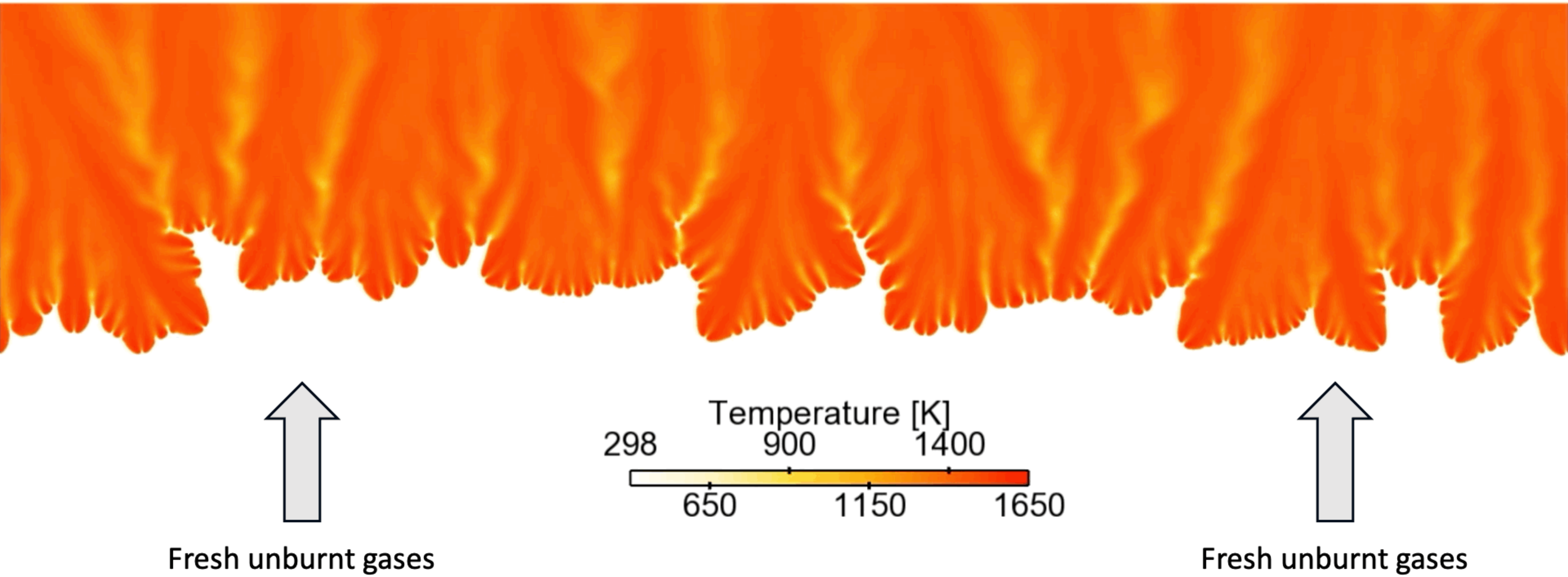


High-fidelity simulations are becoming a fundamental tool to gain insights into complex combustion systems



These simulations require accurate closure models to describe **subgrid** turbulence-chemistry interactions and **transport phenomena**

Hydrogen is a seemingly simple fuel, but modeling differential diffusion can be challenging even for laminar flames



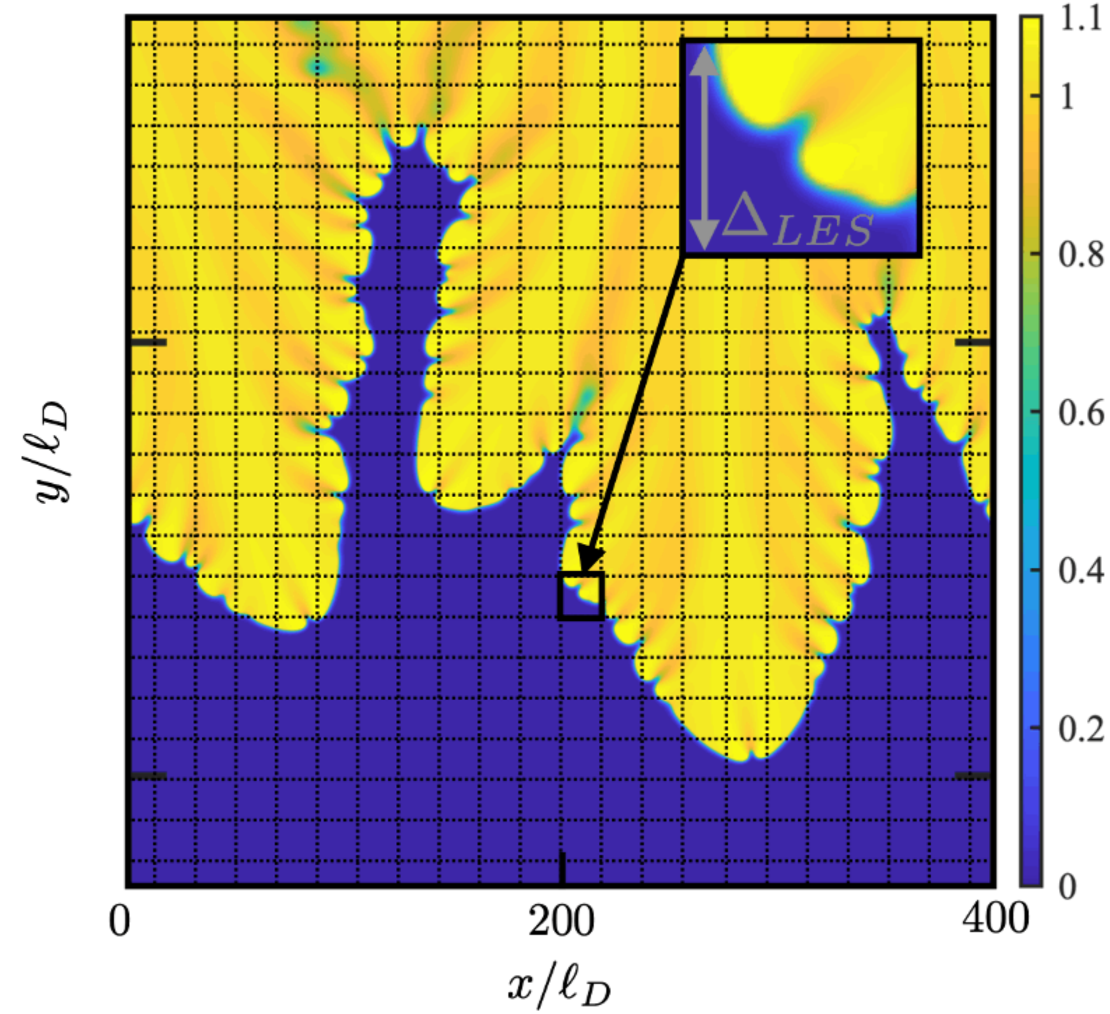
[3] Berger, L., Kleinheinz, K., Attili, A., & Pitsch, H. (2019). *Characteristic patterns of thermodiffusively unstable premixed lean hydrogen flames*. *Proceedings of the Combustion Institute, 37*(2), 1879–1886. https://doi.org/10.1016/j.proci.2018.06.072

Modeling challenges for intrinsic flame instabilities in Large Eddy Simulations (LES)

1.Flame structure dimensionality: minimal/optimal number of independent scalars needed to describe the flame structure with acceptable accuracy.

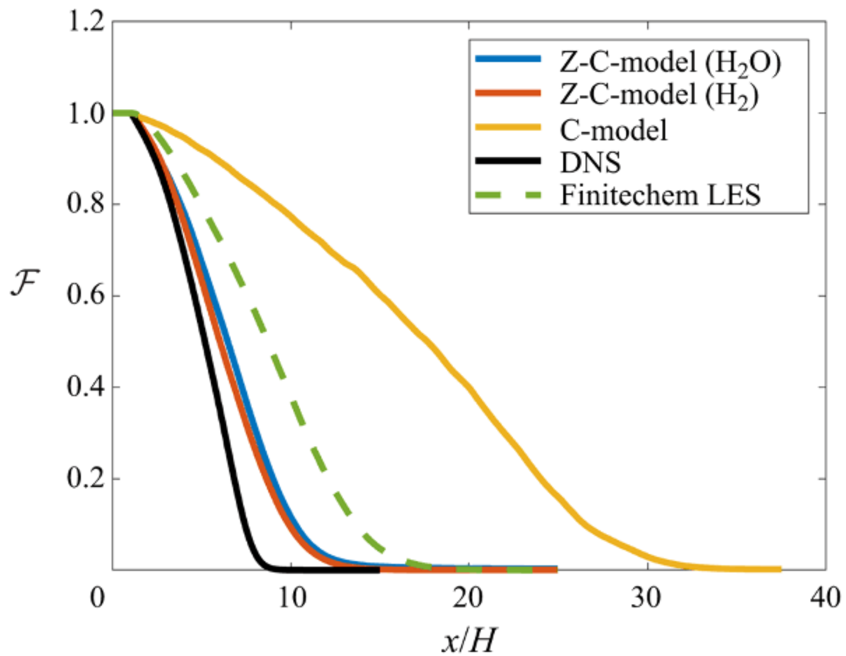
2.Unresolved cellular wrinkling: corrugations due to IFI partially unresolved by LES grid and present at sub-filter scale level.

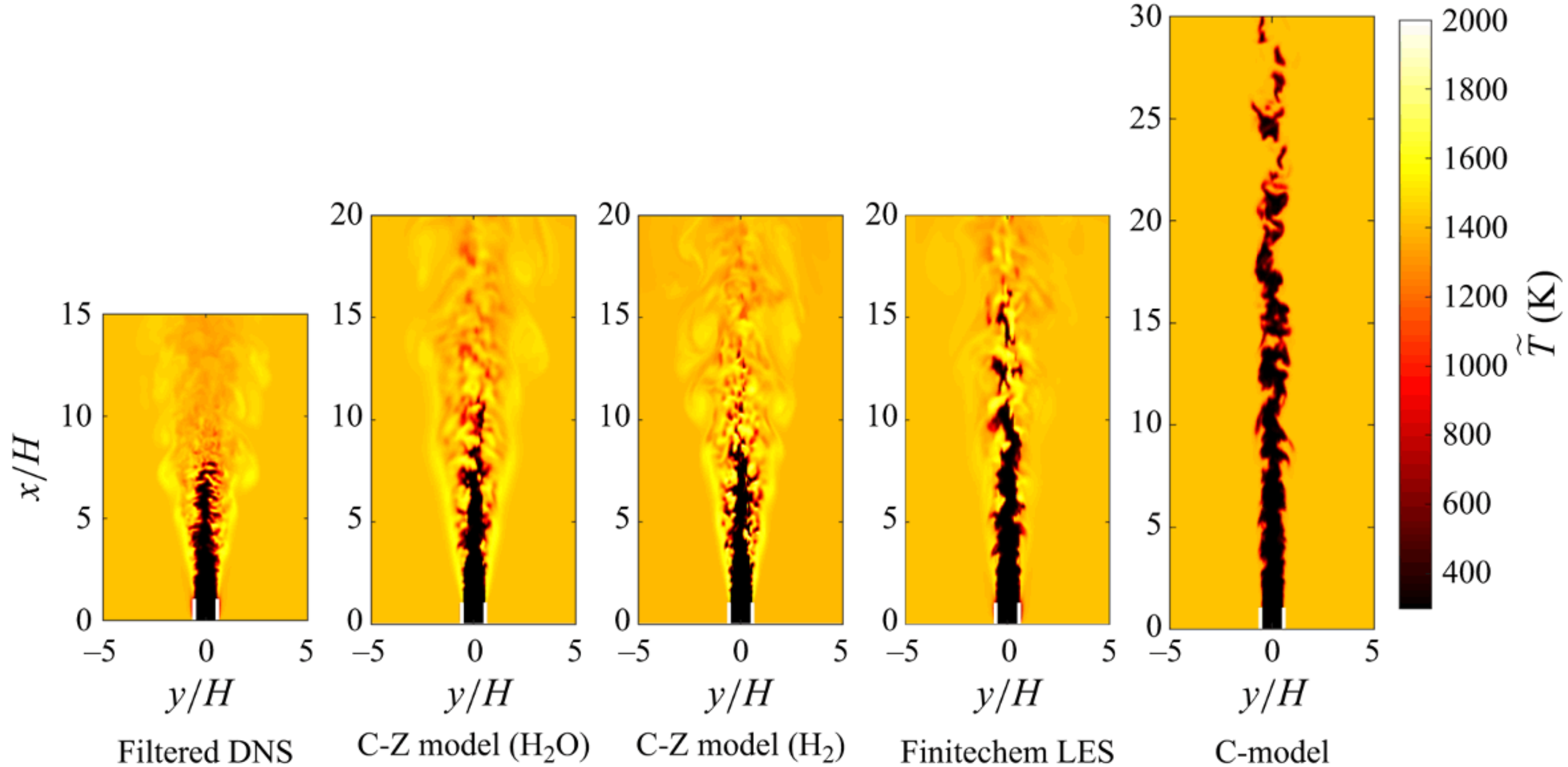
3.Turbulence-IFI interactions: synergistic effect, additional increment of the consumption speed.



[4] Lapenna, P. E., Berger, L., Creta, F., & Pitsch, H. (2023). Hydrogen laminar flames. In E.-A. Tingas (Ed.), Hydrogen for future thermal engines (pp. 93–139). Springer. https://doi.org/10.1007/978-3-031-28412-0_3

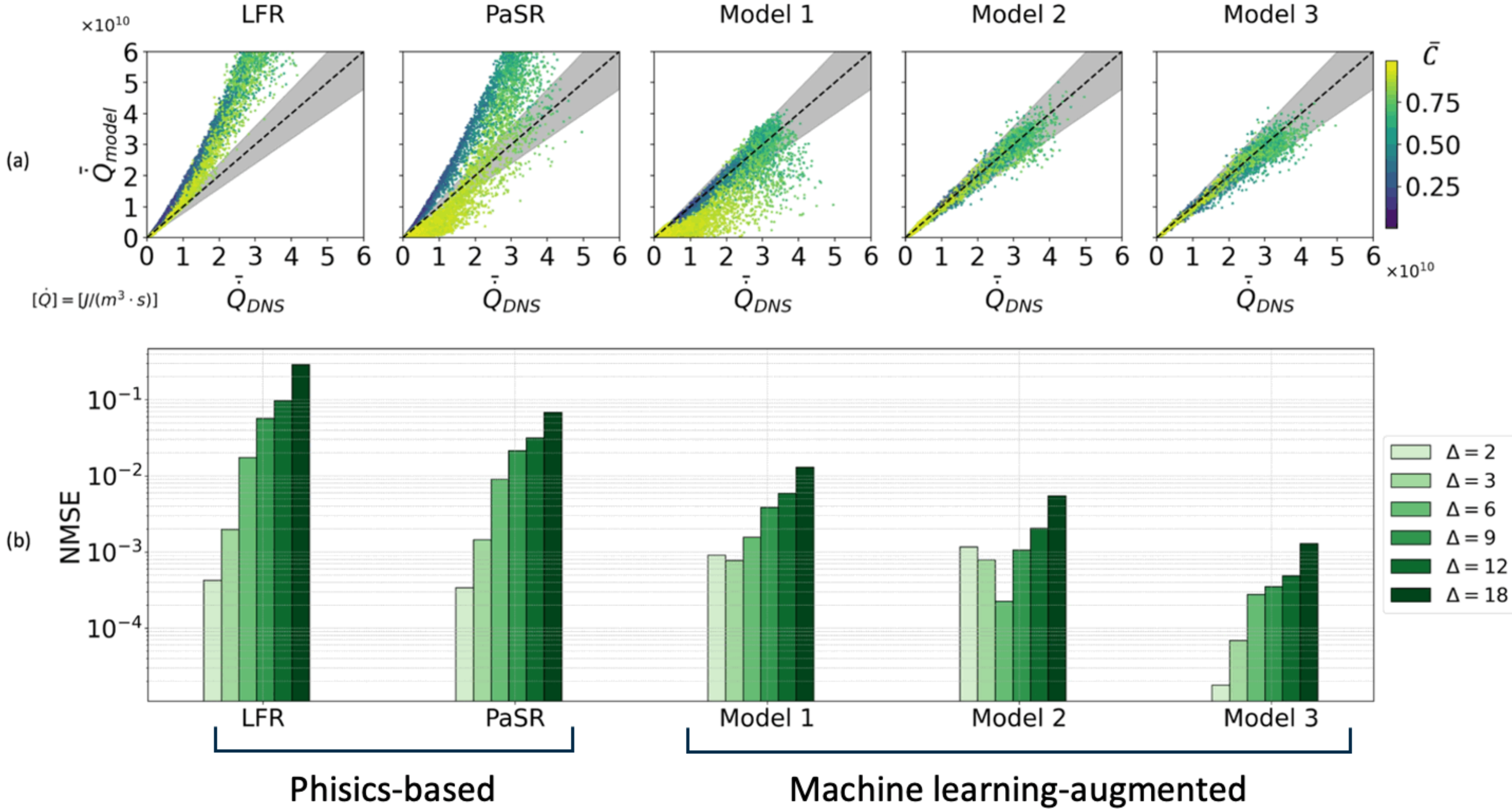
Neglecting thermo-diffusive instabilities can lead to macroscopic errors in the simulations





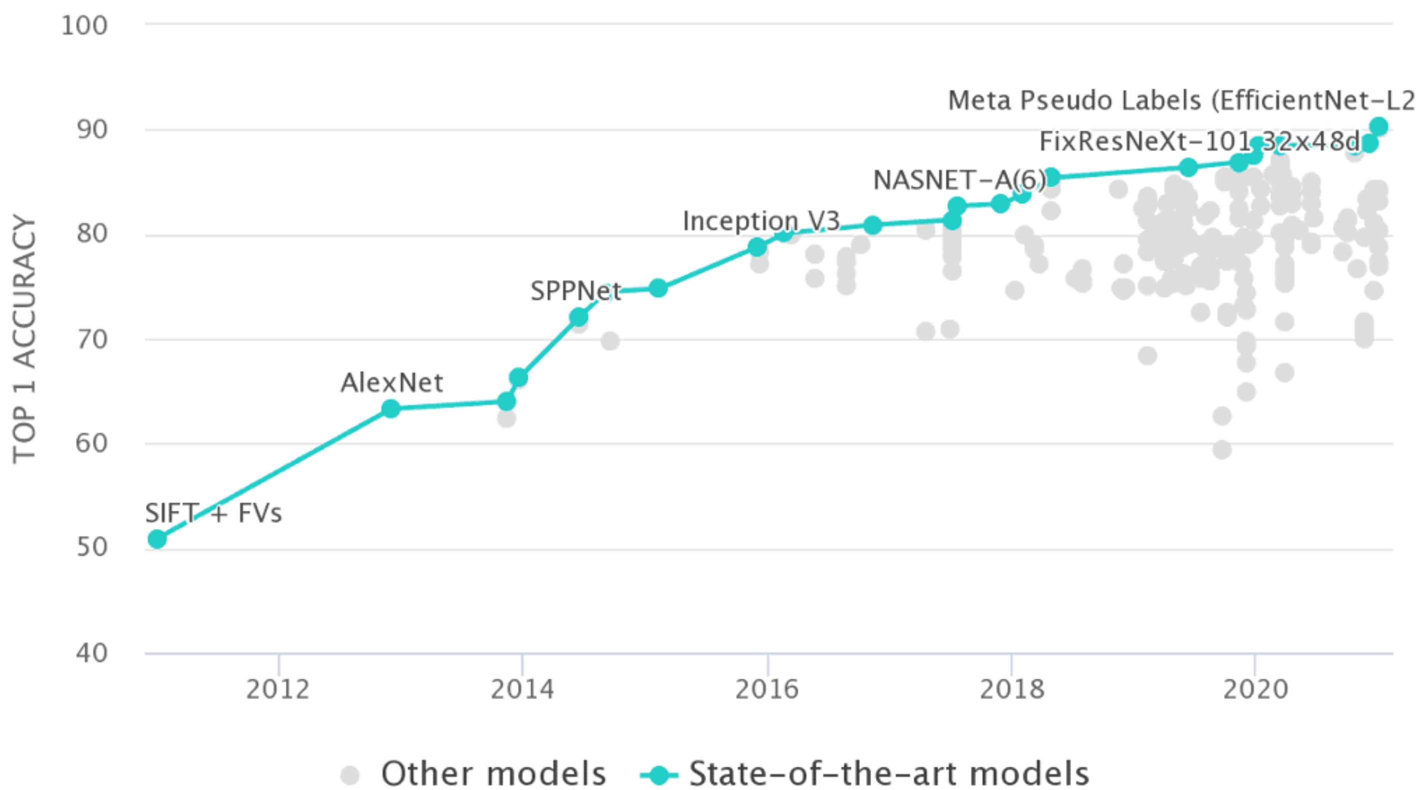
[5] Berger, L., Attili, A., Gauding, M., & Pitsch, H. (2025). *LES combustion model for premixed turbulent hydrogen flames with thermodiffusive instabilities: a priori and a posteriori analysis*. Journal of Fluid Mechanics, 1003, A33. doi:10.1017/jfm.2024.1201

Machine learning has shown potential to enhance existing turbulent combustion closures in LES



[6] Piu, L., Péquin, A., Freitas, R. S. M., Iavarone, S., Pitsch, H., & Parente, A. (2025). *A data-driven approach to refine the partially stirred reactor closure for turbulent premixed flames*. Flow, Turbulence and Combustion, 115(3), 1235–1260. https://doi.org/10.1007/s10494-024-00626-3

Shared databases enabled fast and standardized progress in machine learning, for example in image recognition tasks

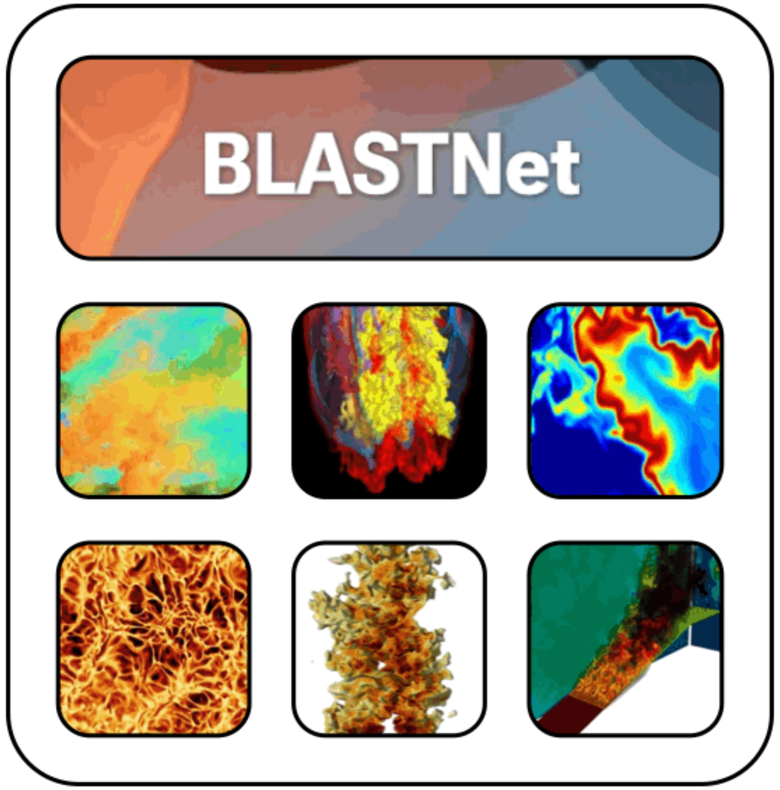


- **Ready to use**, labeled data
- Easy to handle, **accessible** dataset

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

Shared databases enabled fast and standardized progress in machine learning, for example in image recognition tasks

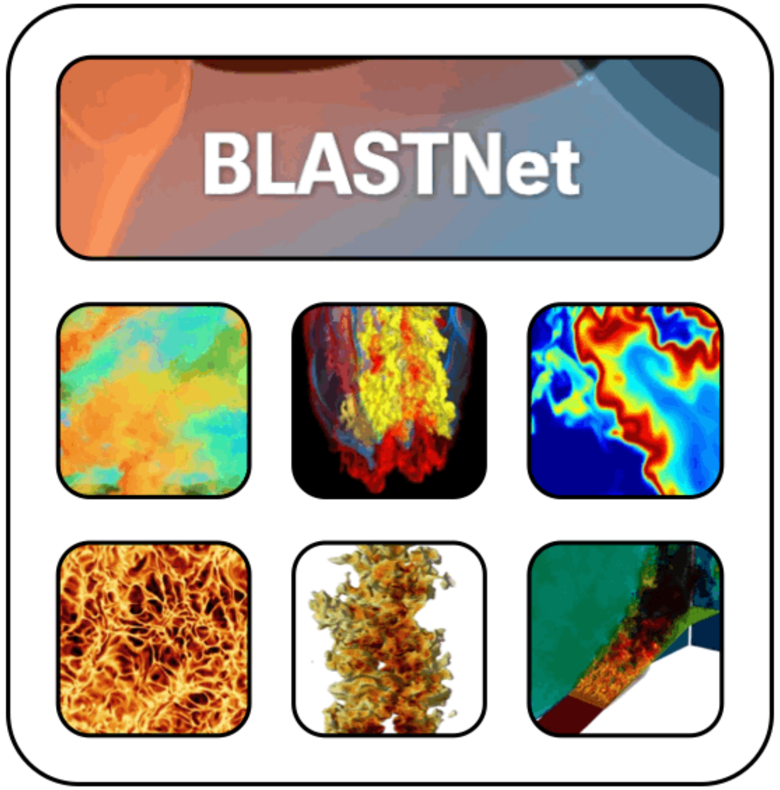


- Data to be **processed** (e.g. sub-filter closures)
- Large dimensions (up-to **100s GB** per simulation timestep)

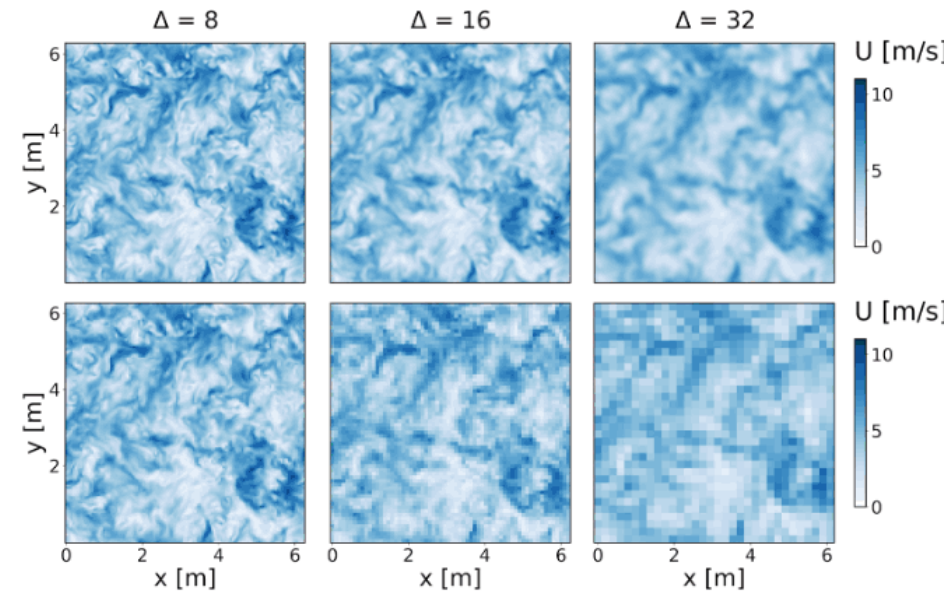
- **Ready to use**, labeled data
- Easy to handle, **accessible** dataset

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

Processing Direct Numerical Simulations (DNS) can be challenging from a practical standpoint



Data processing

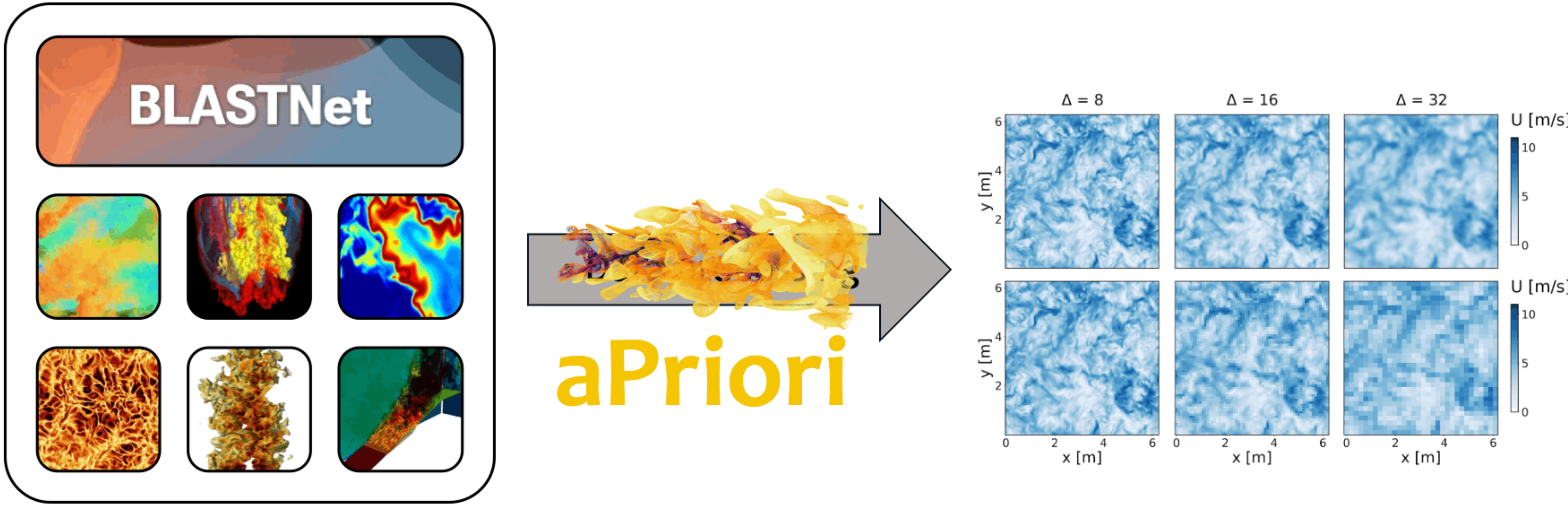


- Data to be **processed** (e.g. sub-filter closures)
- Large dimensions (up-to **100s GB** per simulation timestep)

- **Ready to use** sub-filter data to model closures with machine learning
- Easy to handle, **accessible** datasets

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

Processing Direct Numerical Simulations (DNS) can be challenging from a practical standpoint



- Data to be **processed** (e.g. sub-filter closures)
- Large dimensions (up-to **100s GB** per simulation timestep)

- **Ready to use** sub-filter data to model closures with machine learning
- Easy to handle, **accessible** datasets

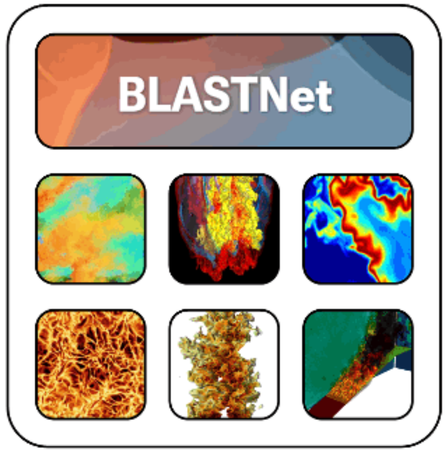
[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

Processing Direct Numerical Simulations (DNS) can be challenging from a practical standpoint

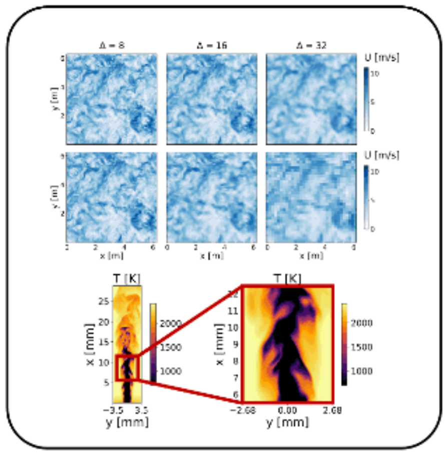
aPriori

Documentation website

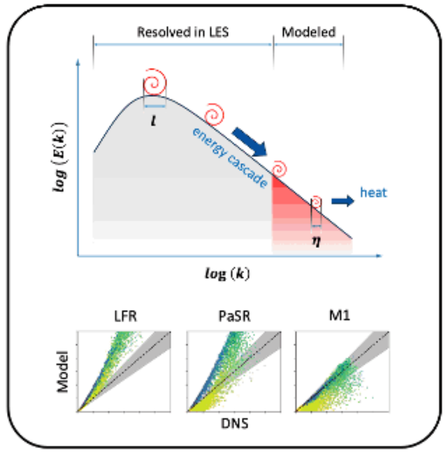
1. Handling large datasets from open-source repositories



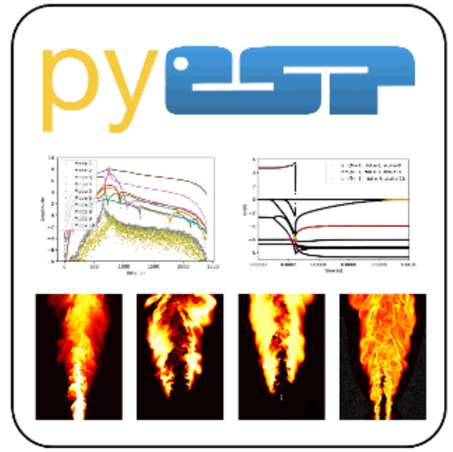
2. Filtering, downsampling, and sub-domain extraction



3. Sub-filter quantities and model assessment



4. CSP and complex chemical analyses for reactive flows



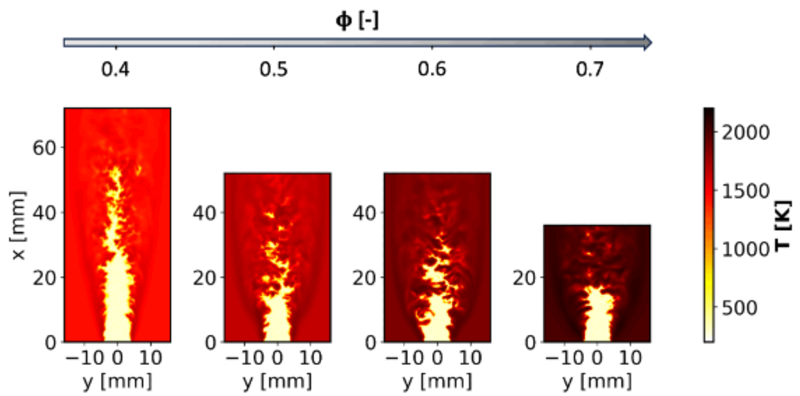
[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

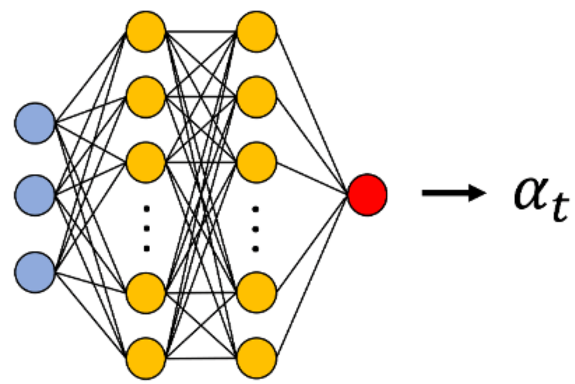
The CYPHER data challenge aims at modeling sub-filter scalar flux of $\tilde{C}$ in lean premixed H_2 flames using machine learning

Codabench

1. Training data processing



2. Online model training



3. Closure model evaluation

$$\frac{\partial}{\partial t}\left(\bar{\rho}\tilde{C}\right) + \nabla\cdot\left(\bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}\right) = \nabla\cdot\left[\bar{\rho}(\alpha_C + \alpha_t)\nabla\tilde{C}\right] + \bar{\rho}\dot{\omega}_C$$

[9] Malé, Q., Lapeyre, C. J., & Noiray, N. (2025). *Hydrogen reaction rate modeling based on convolutional neural network for large eddy simulation*. Data-Centric Engineering, *6*, e11. doi:10.1017/dce.2025.1

The CYPHER data challenge aims at modeling the sub-filter scalar flux of $\tilde{C}$ in lean premixed H_2 flames using machine learning

$$\tilde{C} = \frac{\tilde{Y}^u_{H_2} - \tilde{Y}_{H_2}}{\tilde{Y}^u_{H_2} - \tilde{Y}^b_{H_2}} \quad (\tilde{Y}^b_{H_2} \to 0)$$

Filtered progress variable

sub-filter scalar flux of $\tilde{C}$

$$\boldsymbol{q}_{\tilde{C}} = \bar{\rho}\widetilde{\boldsymbol{u}C} - \bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}$$

Combustion closure models
(Thickened flame model, Flamelet, FPV, Flame surface density…)

$$\frac{\partial \bar{\rho}\tilde{C}}{\partial t} + \nabla \cdot \left(\bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}\right) + \overbrace{\nabla \cdot \left(\bar{\rho}\widetilde{\boldsymbol{u}C} - \bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}\right)} = \overbrace{\overline{\nabla \cdot (\rho D \nabla C) + \dot{\omega}}}$$

Filtered progress variable **transport equation**

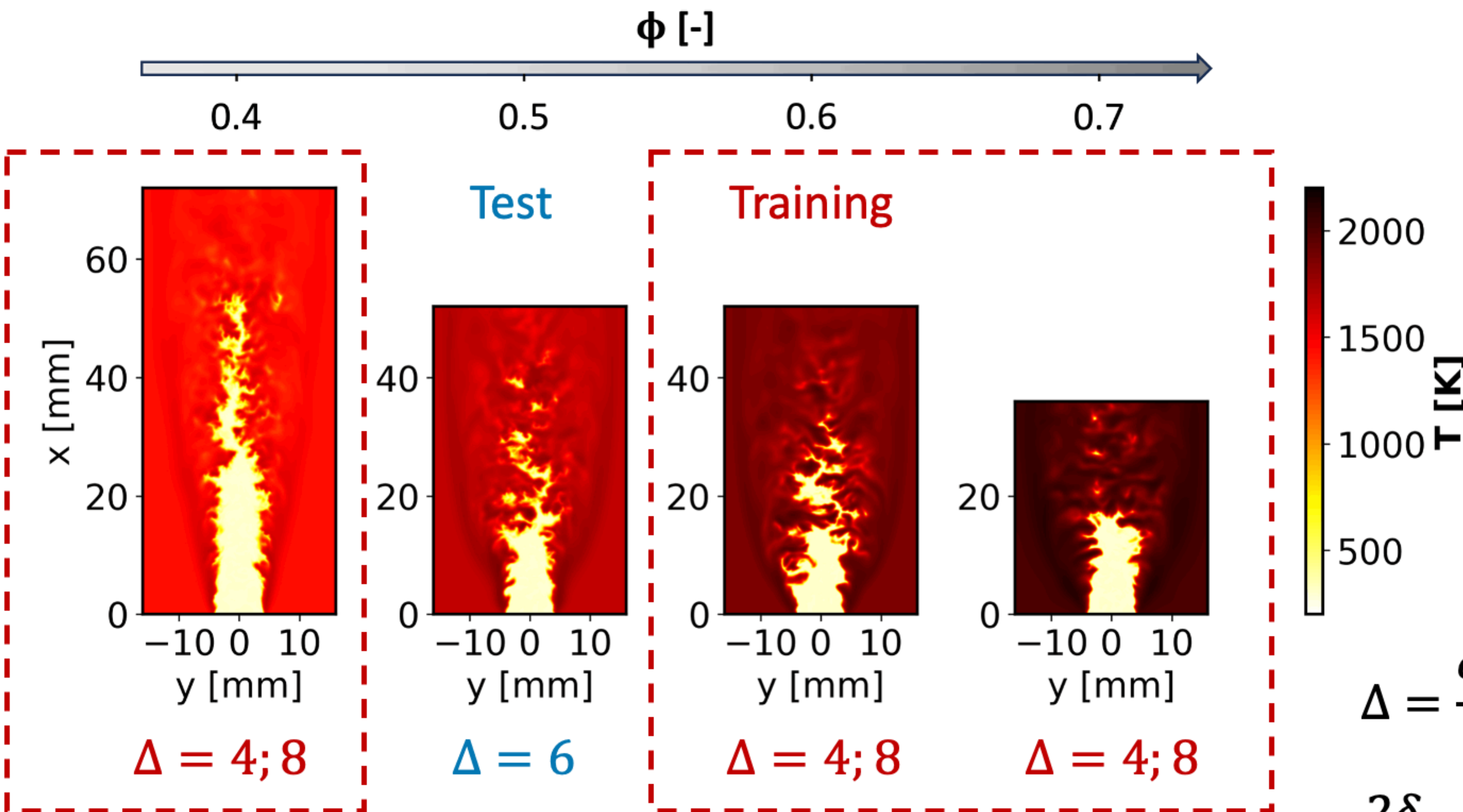
$$\nabla \cdot \left(\bar{\rho}\widetilde{\boldsymbol{u}C} - \bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}\right) \approx \nabla \cdot \bar{\rho}\,\boxed{\alpha_t}\,\nabla\tilde{C}$$

Eddy viscosity assumption to model the sub-filter scalar fluxes of the filtered progress variable

$$\boxed{\alpha_t} = \frac{\left(\overbrace{\bar{\rho}\widetilde{\boldsymbol{u}C}} - \bar{\rho}\tilde{\boldsymbol{u}}\tilde{C}\right)}{\bar{\rho}\nabla\tilde{C}}$$

Sub-filter information available in the **DNS**

The dataset comprises flames at different equivalence ratios and was filtered at various filter sizes



$$\Delta = \frac{\delta_{filter}}{\delta_{DNS}}$$

$$2\delta_{DNS} = \eta_k$$

[9] Malé, Q., Lapeyre, C. J., & Noiray, N. (2025). *Hydrogen reaction rate modeling based on convolutional neural network for large eddy simulation*. Data-Centric Engineering, *6*, e11. doi:10.1017/dce.2025.1

The evaluation metric includes both the model's accuracy and the inference time

- The **Codabench** platform [9] hosts the competition and allows **automatic evaluation** of the submitted models

- The evaluation metric accounts for model's **accuracy** and **inference time**

$$\epsilon = \underbrace{\left\| \boldsymbol{q}_{\tilde{C}}^{model} - \boldsymbol{q}_{\tilde{C}}^{DNS} \right\|_2}_{\text{Accuracy}} + \underbrace{\beta_t \, t_{inf}}_{\text{Computational cost}}$$

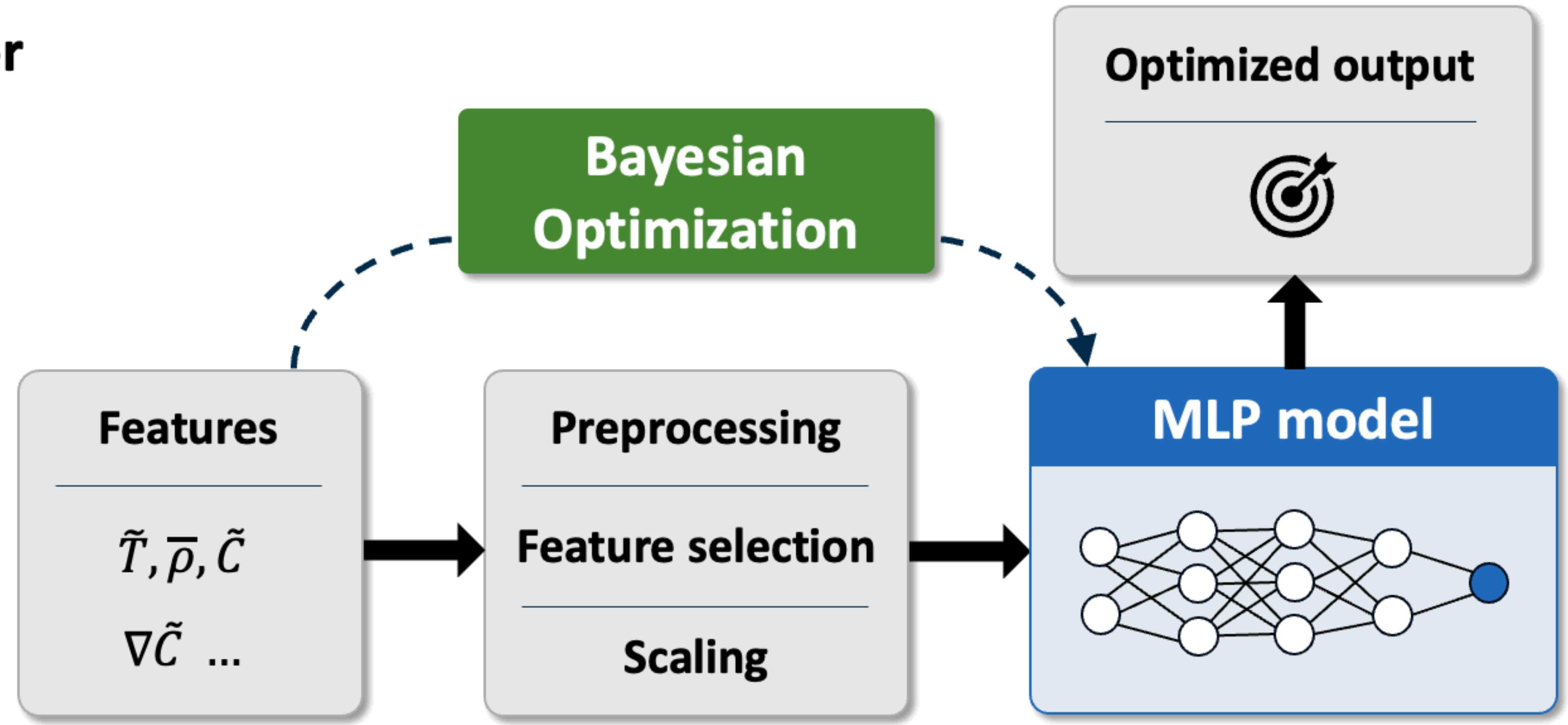
- β_t is a coefficient to balance the relative weight of the two terms during the evaluation phase

[10] Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W. W., Yao, Q., Zhao, H., & Guyon, I. (2022). *Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform*. Patterns (New York, N.Y.), *3*(7), 100543. https://doi.org/10.1016/j.patter.2022.100543

Participants submitted different model architectures, promoting an heterogeneous comparison

Bayesian optimization for MLP feature selection

- Wide exploration of the input features
- Simple baseline model structure
- Relatively fast inference time



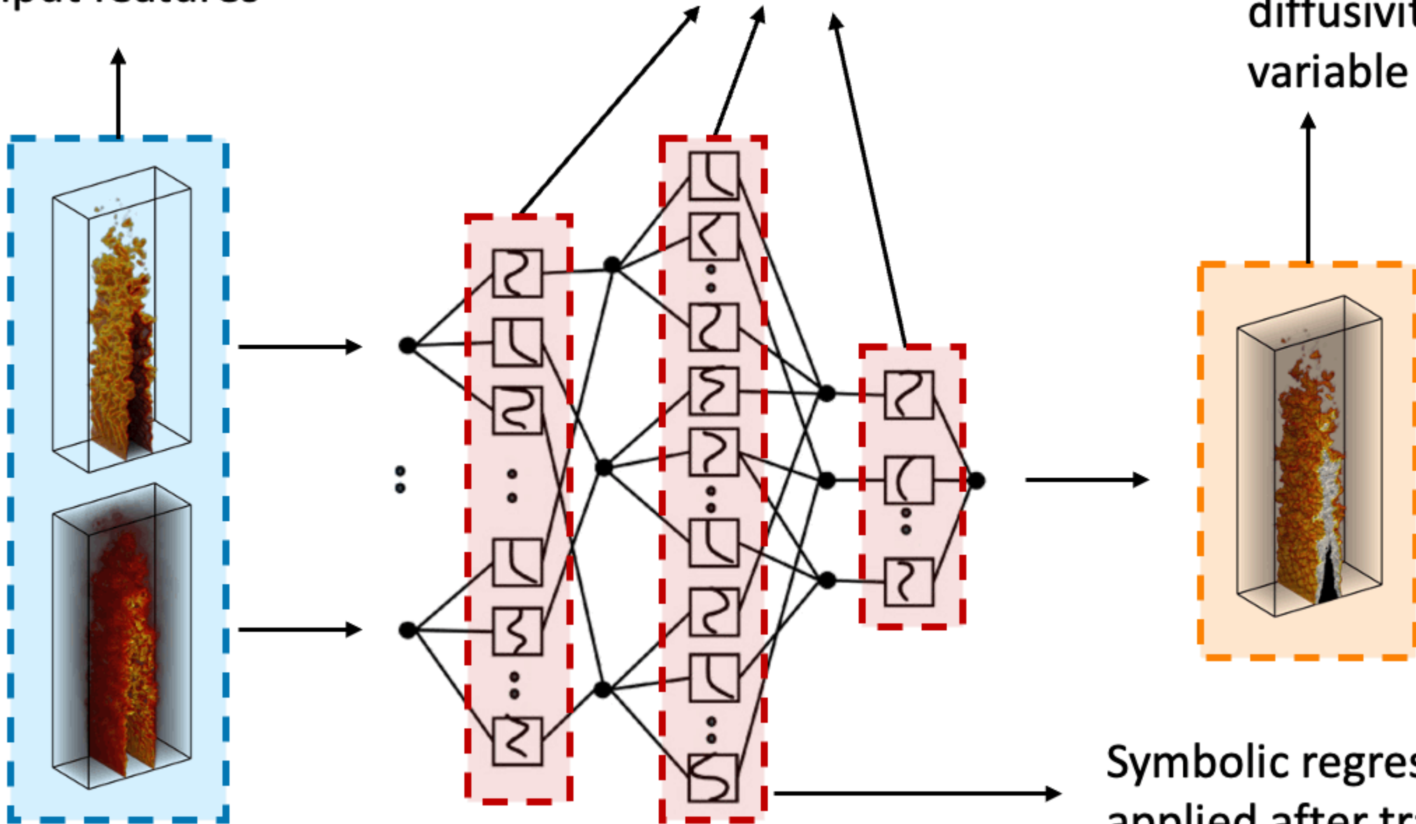
Prof. Kosuke Shigematsu,
Oita National Institute of Technology, Japan

独立行政法人国立高等専門学校機構
大分工業高等専門学校
National Institute of Technology, Oita College

Participants submitted different model architectures, promoting an heterogeneous comparison

Kolmogorov-Arnold Networks (KANs)

- Interpretablity
- Symbolic model form
- Easy to implement in CFD solvers



Tamara Osseily,
Université Libre de Bruxelles
(ULB), Bruxelles

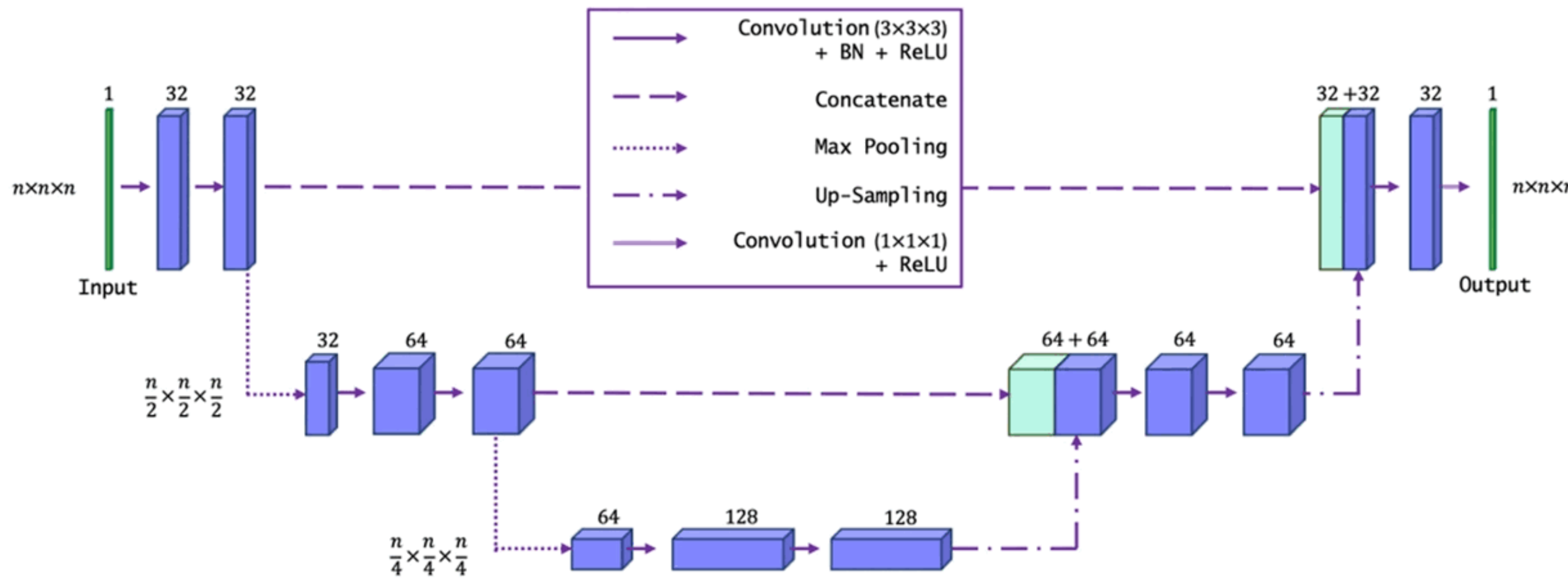
ULB UNIVERSITÉ LIBRE DE BRUXELLES

Participants submitted different model architectures, promoting an heterogeneous comparison

Convolutional U-Net architecture

- Best model submitted
- Convolutional layers embed spatial correlation
- Notably low inference time

[11] Arumapperuma, G., Sorace, N., Jansen, M., Grenga, T., Attili, A., & Pitsch, H. (2025). Extrapolation performance of convolutional neural network-based combustion models for large-eddy simulation: Influence of Reynolds number, filter kernel and filter size. *Flow, Turbulence and Combustion, 115*, 1261–1290. https://doi.org/10.1007/



Dr. Giuseppe Indelicato
Barcelona Supercomputing Center, Spain

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

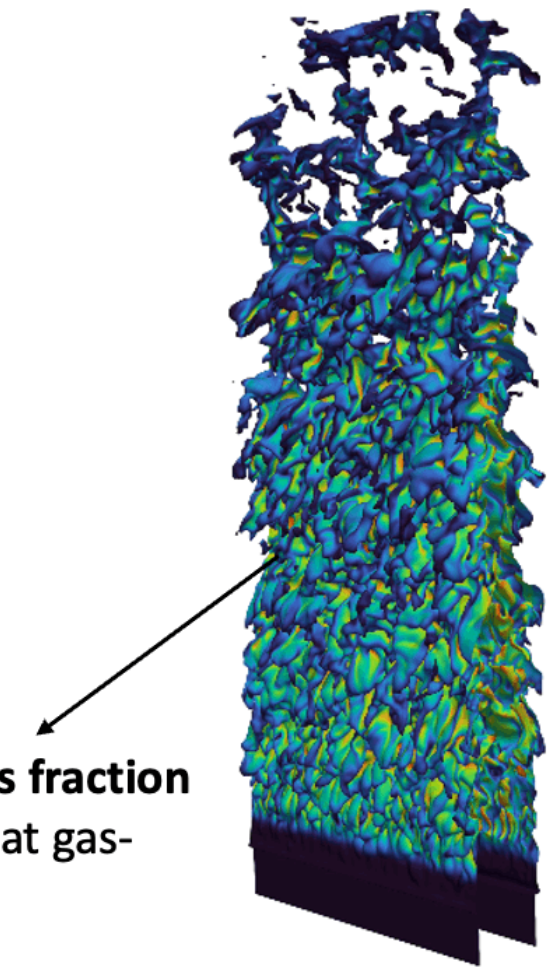
Dr. S. Shubham
Barcelona Supercomputing Center, Spain

Second part of the challenge: modeling of the unclosed reaction rates in LES of premixed lean hydrogen flames

- Goal:
  Train a **machine learning** model to predict the **filtered reaction rates** $\overline{\dot{\omega}}_{H_2}$

$$\overline{\dot{\omega}}_{H_2} = f\left(\tilde{Y}_{H_2}, \tilde{Y}_{H_2O}\right)$$

- The dataset features a premixed jet flame at conditions that are of interest for practical **gas turbine applications**
- All the interested participants to this community effort are encouraged to **reach out by email** to pasquale.lapenna@uniroma1.it.



Isosurface of **hydrogen mass fraction** coloured with reaction rate at gas-turbine conditions (700 K and 20 atm) [10]

[12] Kassar, S. A., Cantagalli, S., Lauder, W., Arumapperuma, G., & Attili, A. (2026). Turbulent hydrogen premixed flames at high pressure and high temperature [Preprint, presented at the 12th European Combustion Meeting]. arXiv. https://doi.org/10.48550/arXiv.2601.15848

Organizing committee (CYPHER)

Professor
PASQUALE LAPENNA

Assistant Professor at Sapienza University of Rome, Italy

Professor
ALESSANDRO PARENTE

Full Professor at Université Libre de Bruxelles, Belgium

Professor
ANTONIO ATTILI

Lecturer at University of Edinburgh, United Kingdom

Professor
PAOLA CINNELLA

Full Professor at Sorbonne Université, France

Professor
FEDERICA FERRARO

Junior Professor at TU Braunschwwig, Germany

Professor
ANH KHOA DOAN

Assistant Professor at TU Delft, Netherlands

Bibliography

[1] Pitsch, H. (2024). *The transition to sustainable combustion: Hydrogen- and carbon-based future fuels and methods for dealing with their challenges*. *Proceedings of the Combustion Institute*, 40(1–4), Article 105638. https://doi.org/10.1016/j.proci.2024.105638

[2] U.S. Energy Information Administration, International energy outlook 2023, 2023, https://www.eia.gov/outlooks/ieo/.

[3] Berger, L., Kleinheinz, K., Attili, A., & Pitsch, H. (2019). *Characteristic patterns of thermodiffusively unstable premixed lean hydrogen flames*. *Proceedings of the Combustion Institute, 37*(2), 1879–1886. https://doi.org/10.1016/j.proci.2018.06.072

[4] Lapenna, P. E., Berger, L., Creta, F., & Pitsch, H. (2023). Hydrogen laminar flames. In E.-A. Tingas (Ed.), Hydrogen for future thermal engines (pp. 93–139). Springer. https://doi.org/10.1007/978-3-031-28412-0_3

[5] Berger, L., Attili, A., Gauding, M., & Pitsch, H. (2025). *LES combustion model for premixed turbulent hydrogen flames with thermodiffusive instabilities: a priori and a posteriori analysis*. Journal of Fluid Mechanics, 1003, A33. https://doi.org/10.1017/jfm.2024.1201

[6] Piu, L., Péquin, A., Freitas, R. S. M., Iavarone, S., Pitsch, H., & Parente, A. (2025). *A data-driven approach to refine the partially stirred reactor closure for turbulent premixed flames*. Flow, Turbulence and Combustion, 115(3), 1235–1260. https://doi.org/10.1007/s10494-024-00626-3

[7] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database*. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[8] Chung, W. T., Jung, K. S., Chen, J. H., & Ihme, M. (2022). *BLASTNet: A call for community-involved big data in combustion machine learning*. Appl. in En. and Comb. Sc.

[9] Malé, Q., Lapeyre, C. J., & Noiray, N. (2025). Hydrogen reaction rate modeling based on convolutional neural network for large eddy simulation. Data-Centric Engineering, 6, e11.

https://doi.org/10.1017/dce.2025.1

[10] Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W. W., Yao, Q., Zhao, H., & Guyon, I. (2022). Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns (New York, N.Y.), 3(7), 100543. https://doi.org/10.1016/j.patter.2022.100543

[11] Arumapperuma, G., Sorace, N., Jansen, M., Grenga, T., Attili, A., & Pitsch, H. (2025). *Extrapolation performance of convolutional neural network-based combustion models for large-eddy simulation: Influence of Reynolds number, filter kernel and filter size*. Flow, Turbulence and Combustion, 115, 1261–1290. https://doi.org/10.1007/s10494-025-00643-w

[12] Kassar, S. A., Cantagalli, S., Lauder, W., Arumapperuma, G., & Attili, A. (2026). *Turbulent hydrogen premixed flames at high pressure and high temperature* [Preprint, presented at the 12th European Combustion Meeting]. arXiv. https://doi.org/10.48550/arXiv.2601.15848

[13] Piu, L., Pitsch, H., & Parente, A. (2024). *aPriori: a Python package to process direct numerical simulations*. [Computer software]. GitHub. https://github.com/LorenzoPiu/aPriori

Thank you for your attention!

Scan to **download** the **slides** and for the link to the **data challenge** websites

Acknowledgements

The CYPHER Data Challenge is organized within the CYPHER COST Action, promoting collaboration between European researchers and industrial stakeholders to accelerate digitalization in sustainable fuels combustion systems.

Lorenzo Piu has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101072779 (ENCODING). The results of this publication/presentation reflect only the author(s) view and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.