Chemically-specific multiscale mechanical simulation of nanomaterials: improving tractability with machine learning

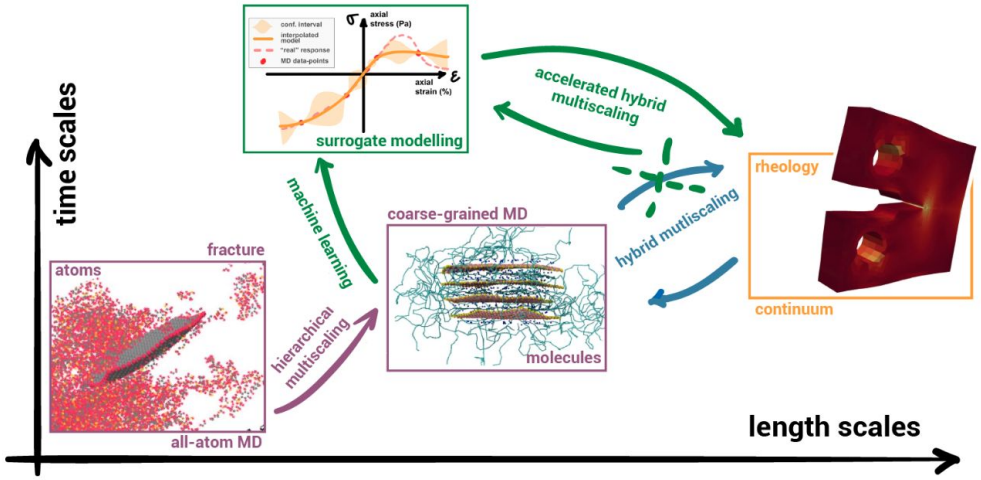
Maxime Vassaux
Institut de Physique de Rennes
Univ. Rennes, CNRS - UMR 6251

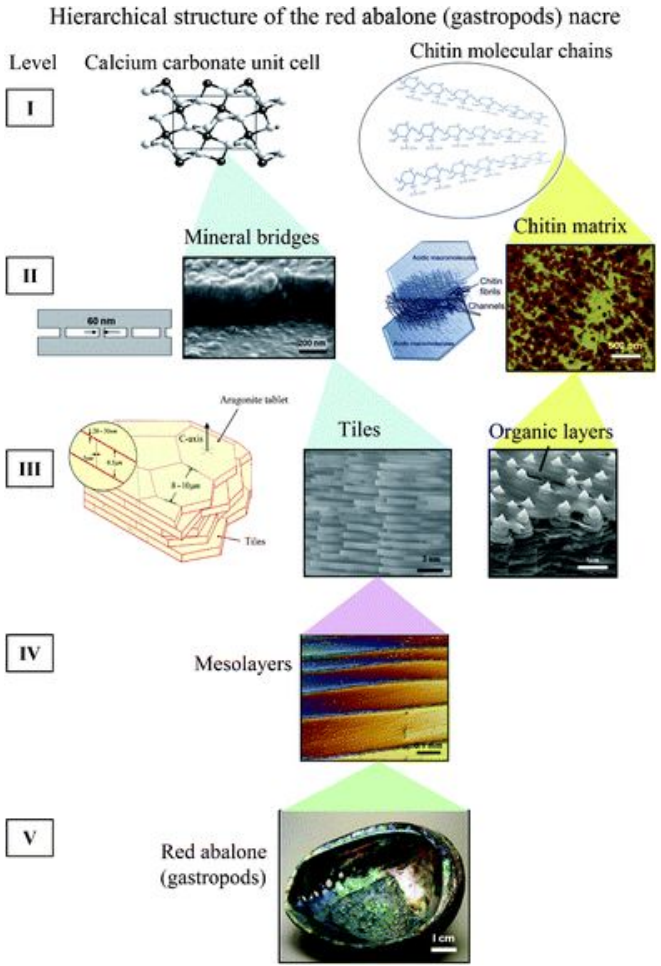
Peter V. Coveney
Centre for Computational Science
University College London

@mxvassaux
maxime.vassaux@cnrs.fr

Emergence of mechanical properties across scales

- hierarchy of scales more or less separated



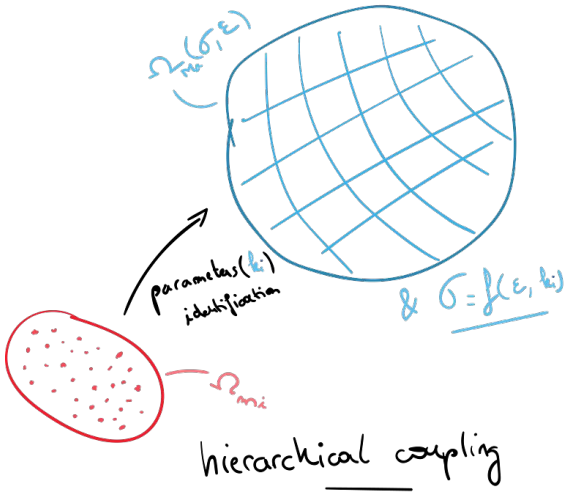


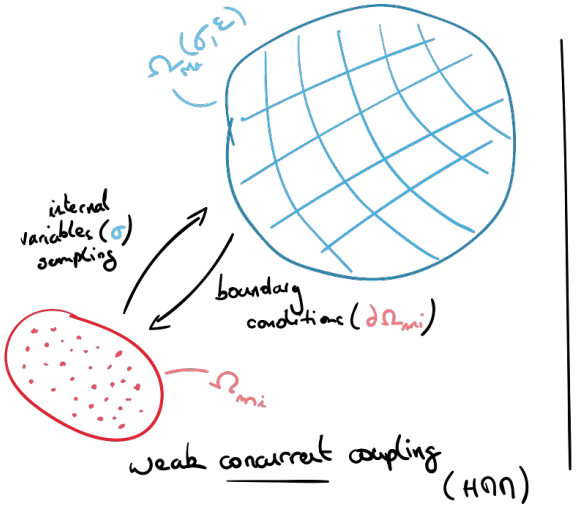
Vassaux, et al. *Adv. Theo. Sim.* (2019) 3.1

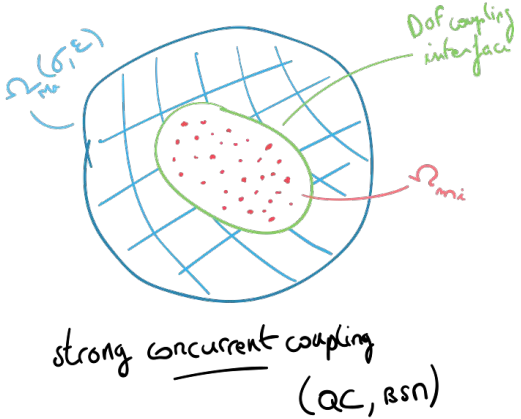
Sun, et al. *RSC Adv.* (2012) 2

A wide spectrum of multiscale strategies

- **different** feedback **frequency** and **resolution**
  - hierarchical
  - semi-concurrent
  - concurrent

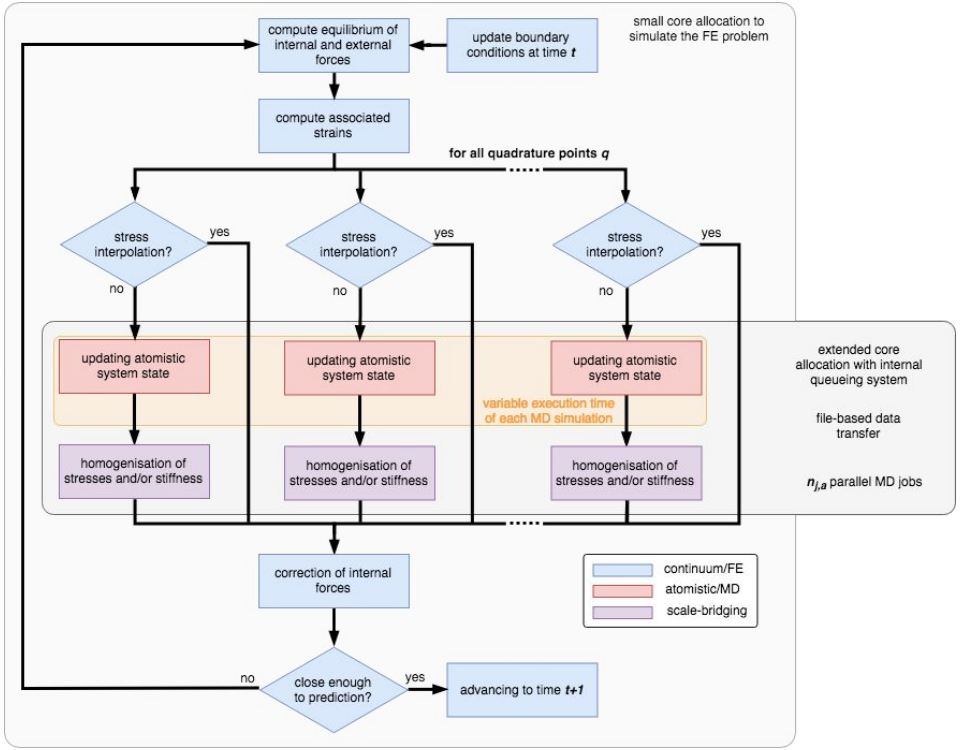


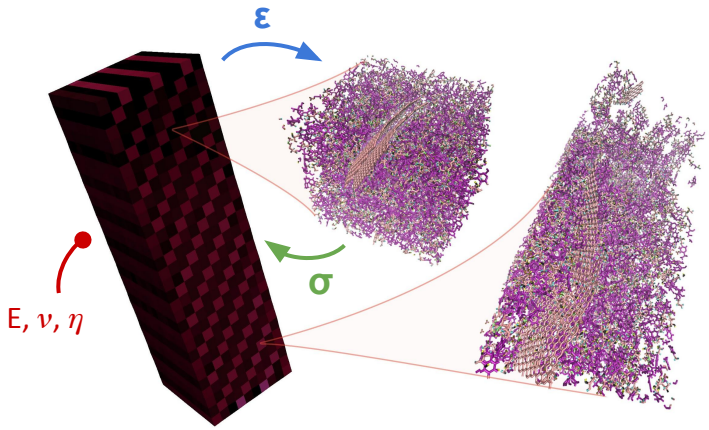




Shenoy et al. *J. Mech. Phys. Solids* (1999) 47.3
E et al. *Phys. Rev. B: Condens. Matter* (2003) 67.9

SCEMa: Simulation Coupling Environment for Materials



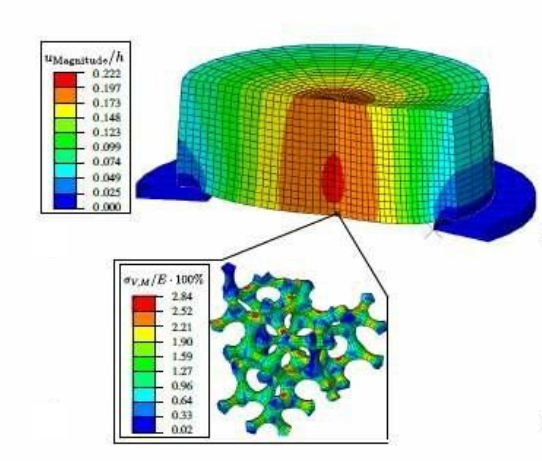


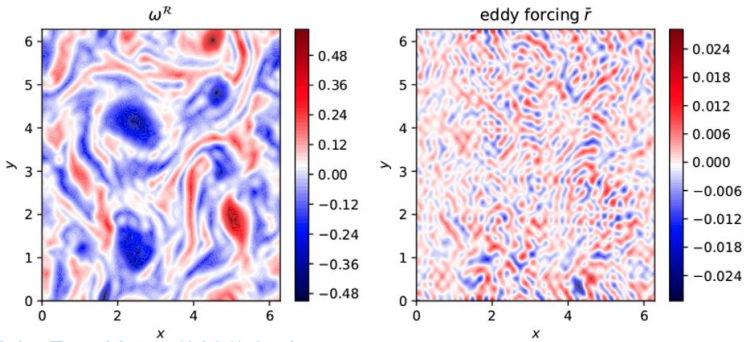
SCEMa : https://github.com/UCL-CCS/SCEMa

Vassaux, Richardson and Coveney. *Philos. Trans. R. Soc. London, Ser. A.* (2019) 377.2142

Ubiquity of semi-concurrent approaches

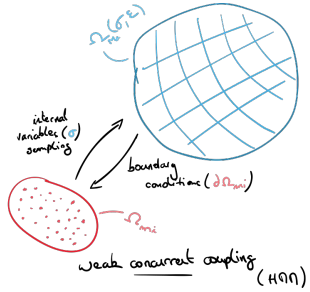
- FE^2, FE-MD, …

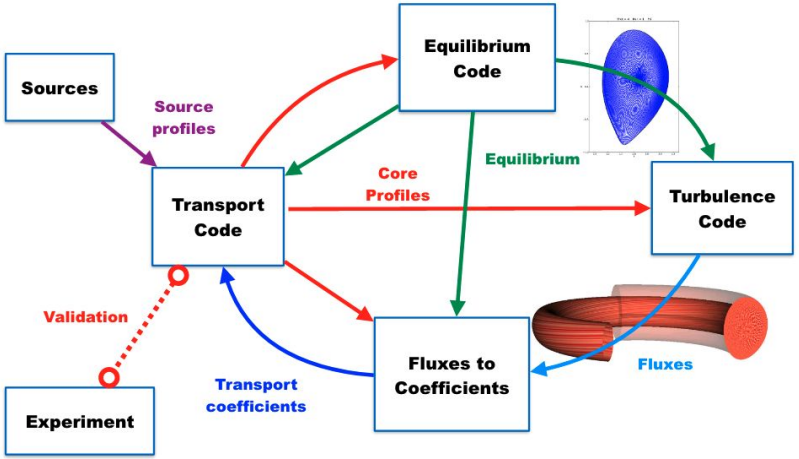




Lange, et al. *Adv. Eng. Mater.* (2021) 24.2

Edeling, et al. *Comput. Fluids.* (2020) 201





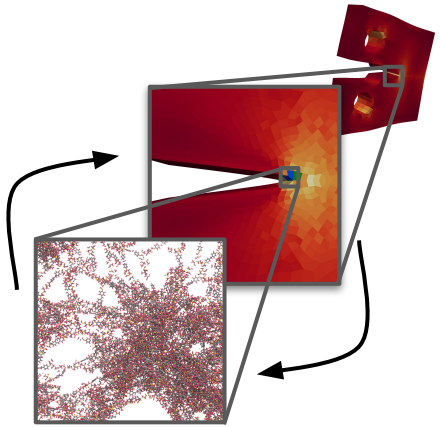
- mechanics but also fusion, climate applications

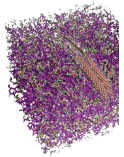
Lakhlili, et al. *LNCTS.* (2020) 12143

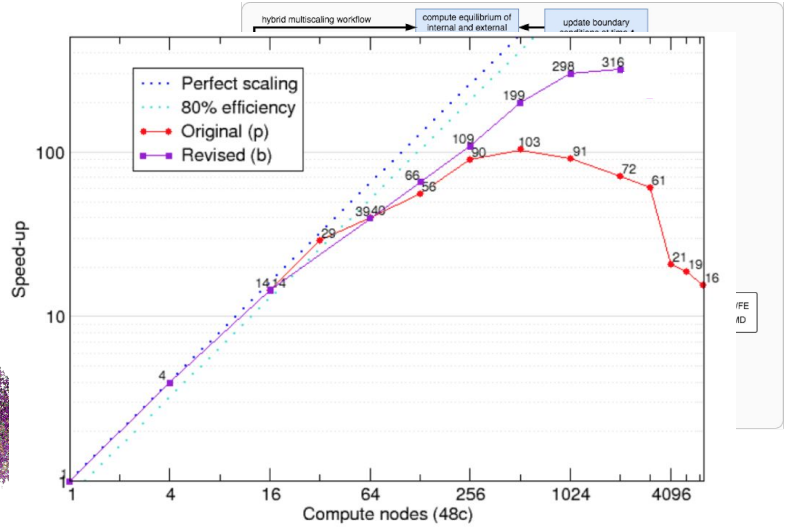
SCEMa: Simulation Coupling Environment for Materials

- Prediction of advanced structural materials properties at the meter/second scale
- Concurrent model coupling asynchronously
  - a continuum macroscale model (FE)
  - a molecular microscale model (MD)

- Computational cost
  - 0.1 core hour per microscale simulation
  - 10,000 cells x 8qps/cell x 5 replicas/qp
  - → 400,000 independent microscale simulations per macroscale simulation time-step
- Scalability benchmark on **full partition**
  - 311,000 CPUs of SuperMUC-NG
  - 316x peak speed-up on 2048 nodes



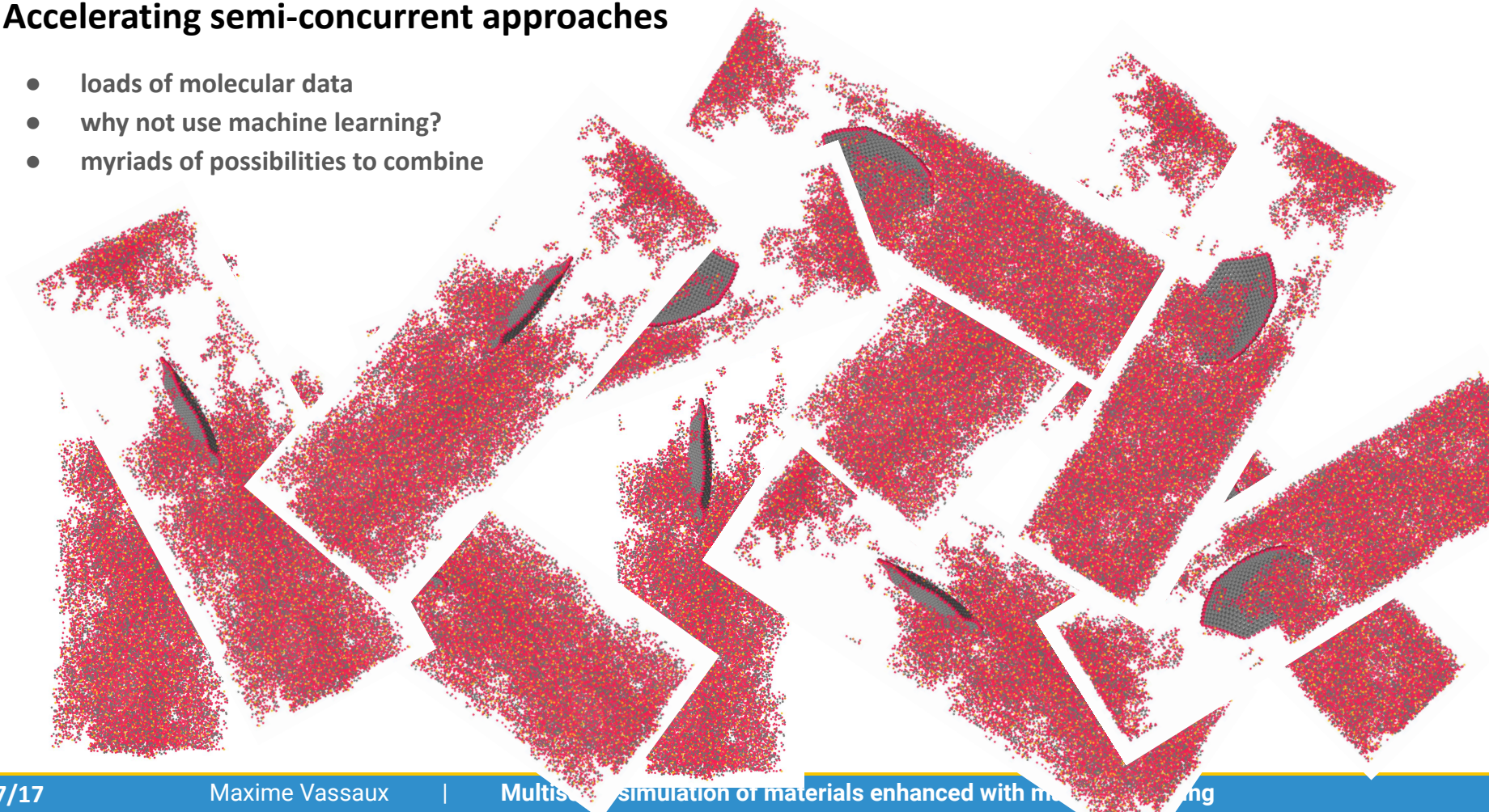




Vassaux, Richardson and Coveney. *Philos. Trans. R. Soc. London, Ser. A.* (2019) 377.2142

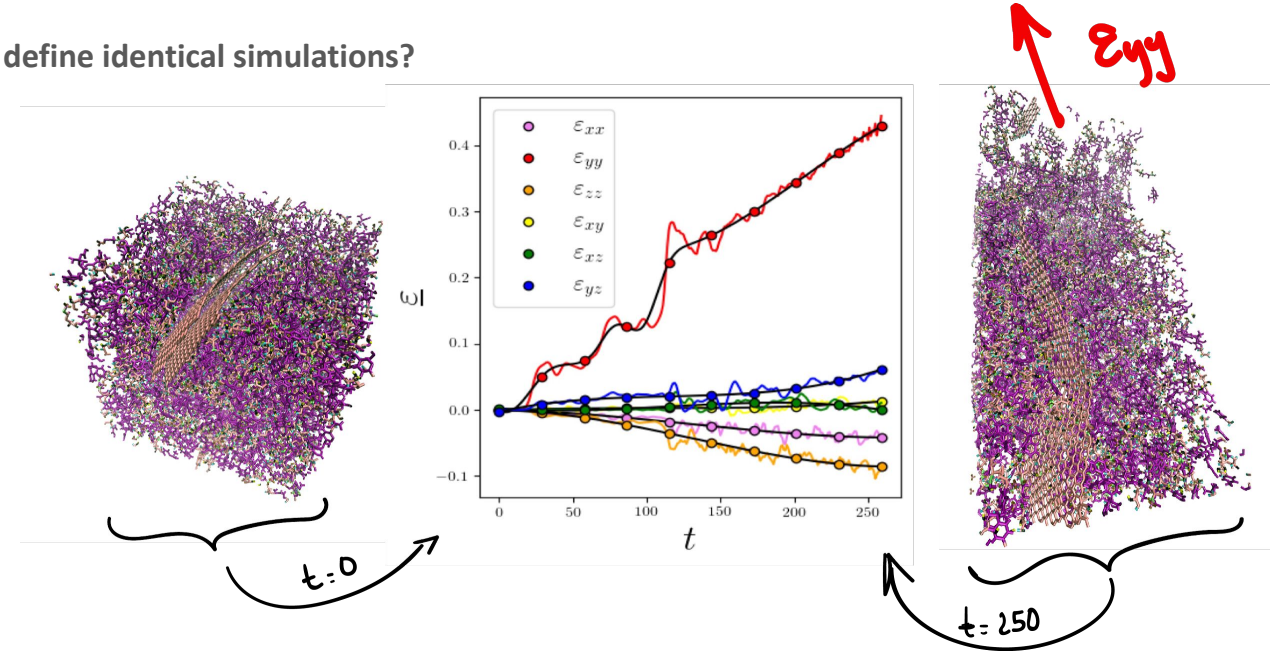
Accelerating semi-concurrent approaches

- loads of molecular data
- why not use machine learning?
- myriads of possibilities to combine



Determine identical simulations using unsupervised learning

- how do we define identical simulations?



- fitting 7D splines (history matters, not rate)
- similarity of a pair of splines inversely proportional to their euclidean distance

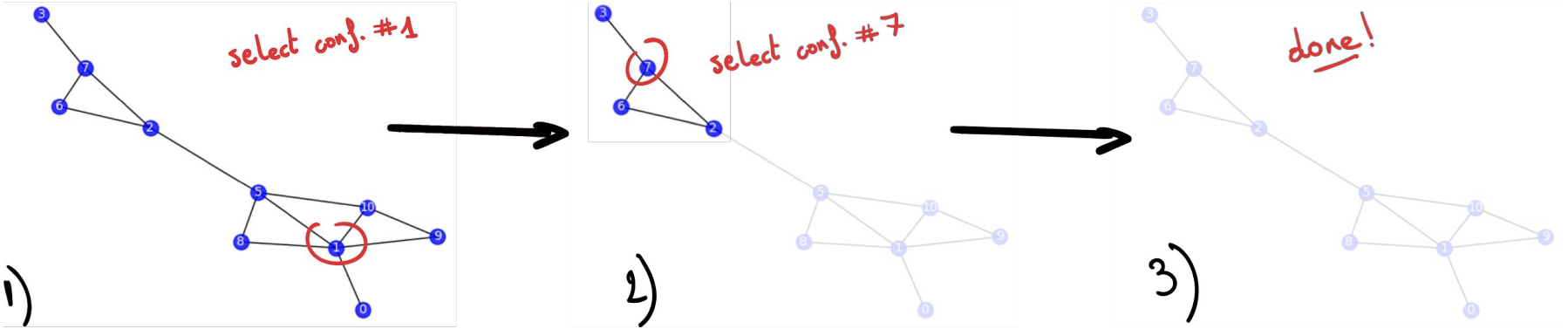
$$\frac{1}{s_{1,2}} \propto \sum_{n=1}^{n_{cp}} \sum_{i=1}^{6} \sqrt{\left(\tilde{\epsilon}_{1,n,i} - \tilde{\epsilon}_{2,n,i}\right)^2}$$

Vassaux, et al. *Adv. Theor. Simul.* (2020) 4.2

Compute as few molecular simulations as possible

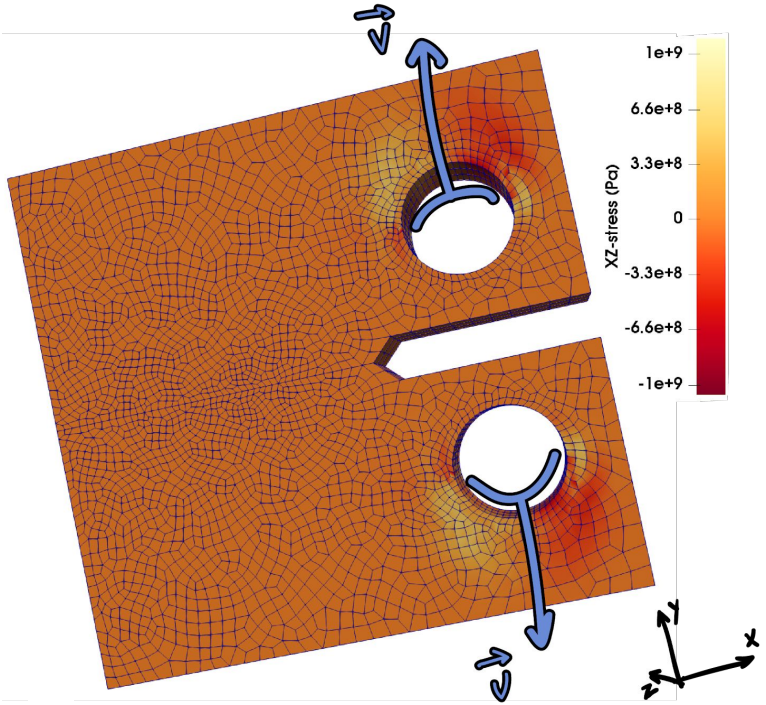
- **clustering** to minimise computational effort
- splines are considered similar if below arbitrary threshold **α**
- algorithm
  - build the graph of similar configurations
  - compute the most connected configurations
  - remove simulated configurations and similar ones from the graph
  - return to 1. (until no more configurations to compute the stress from)

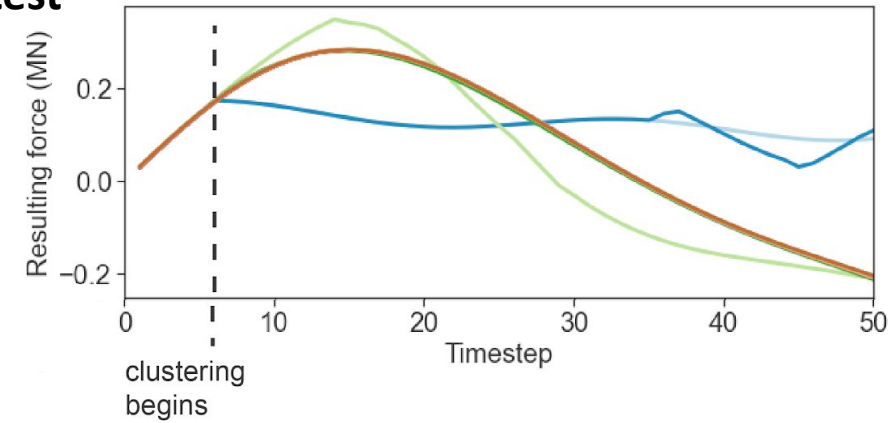
$$\frac{1}{s_{1,2}} \propto \sum_{n=1}^{n_{cp}} \sum_{i=1}^{6} \sqrt{\left(\tilde{\epsilon}_{1,n,i} - \tilde{\epsilon}_{2,n,i}\right)^2}$$

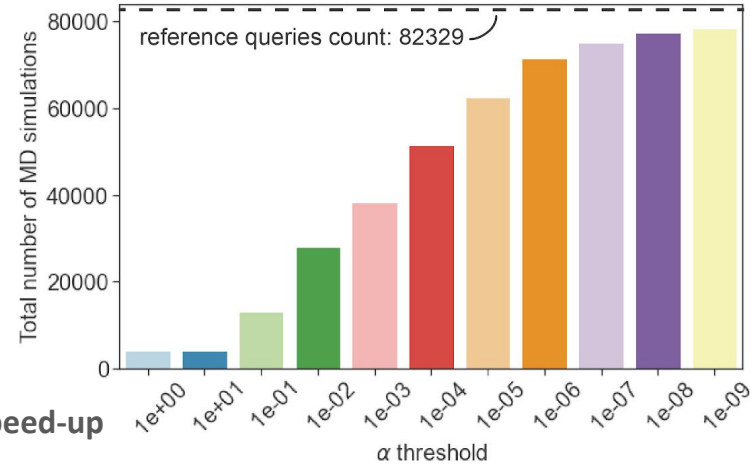


Vassaux, et al. *Adv. Theor. Simul.* (2020) 4.2

Multiscale simulation of a compact-tension test





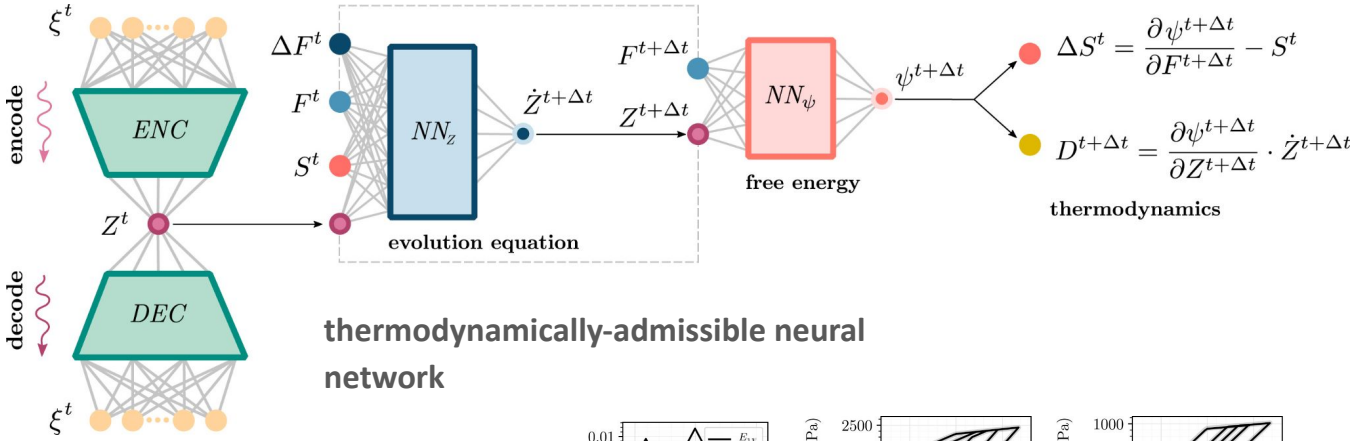


➥ Accurate global prediction with threshold **α < 0.01 → x3.03 speed-up**

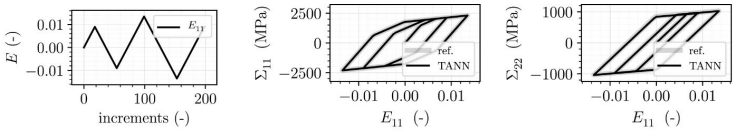
Vassaux, et al. *Adv. Theor. Simul.* (2020) 4.2

Make it faster using interpolation (rather than database look up)

- supervised machine learning can fit mechanical data

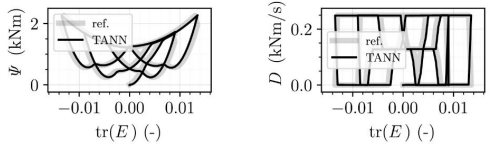


thermodynamically-admissible neural network



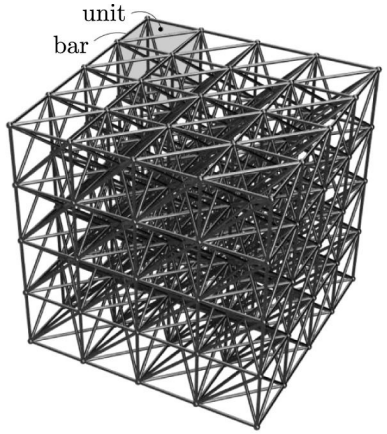
(a) strain path

(b) stress-strain behavior



(c) free-energy and dissipation rate in function of volume change

microscale model for training data generation



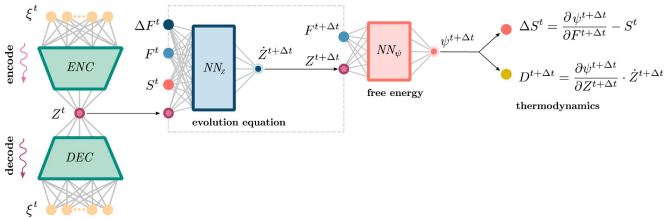
3D lattice material structure

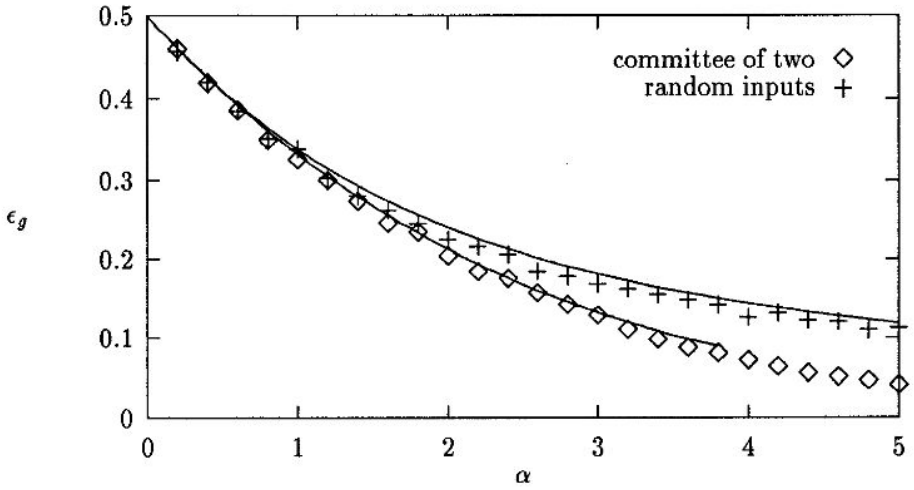
validation of the ML-based model

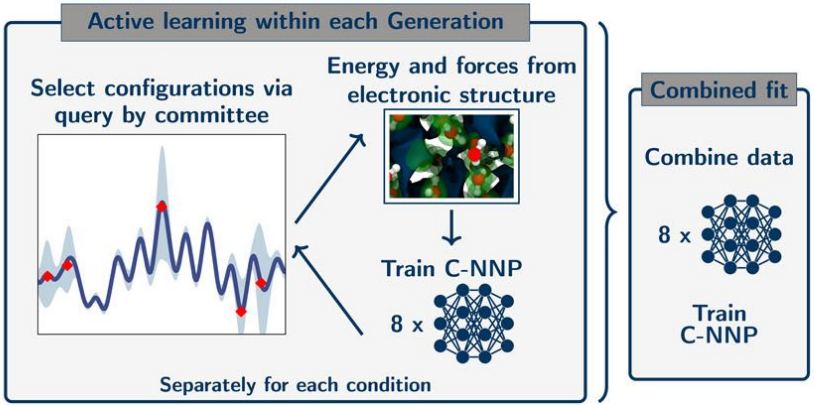
Masi, et al. *CMAME*, (2022) *298*

But how to quantify accuracy of ML-model prediction?

- ML models most often do not have uncertainty estimates…
- Query-by-Committee
  - from Active Learning
  - minimise size of training datasets
  - improve generalization of ML model







Seung, et al. *COLT*. (1992)

Schran, et al. *J. Chem. Phys.* (2020) 153

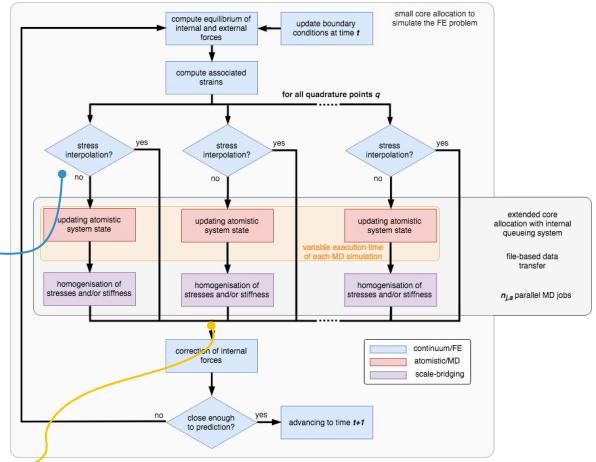
Adaptation of the concurrent multiscale workflow

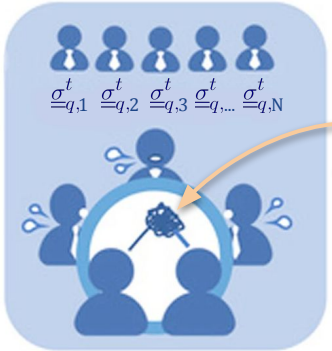
- ask the Committee: do we need to compute a new MD ensemble?

$$(\underline{\underline{\epsilon}}_q^{t-1}, \underline{\underline{\sigma}}_q^{t-1}, \Delta\underline{\underline{\epsilon}}_q^{t})_{inputs} \rightarrow (\underline{\underline{\sigma}}_q^{t})_{outputs}$$

$t \leftarrow$ timestep;
$q \leftarrow$ quadrature point;
$\sigma_{acc} \leftarrow$ MD stress accuracy;
$N_C \leftarrow$ committee size;
$trainset \leftarrow []$;
for $i \leftarrow 0$ **to** N_C **do**
  $\widetilde{\sigma}_{q,i}^t = Committee(i, \epsilon_q^{t-1}, \epsilon_q^t, \sigma_q^{t-1})$
end
if $Var[\widetilde{\sigma}_{q,i}^t]_i < \sigma_{acc}$ **then**
  $\sigma_q^t = E[\sigma_{q,i}^t]_i$
else
  $\sigma_q^t = \mathrm{MD}(\epsilon_{q,i}^t)$;
  $trainset \leftarrow (\epsilon_q^{t-1}, \epsilon_q^t, \sigma_q^{t-1}, \sigma_q^t)$
end

$freq \leftarrow$ training frequency ;
if $trainset.size()\%freq = 0$ **then**
  $Committee = TrainCommittee(trainset)$
else
end

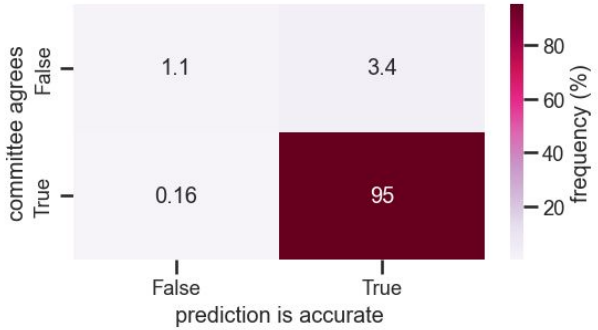


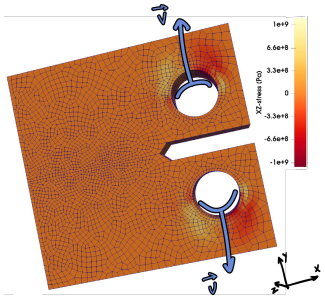


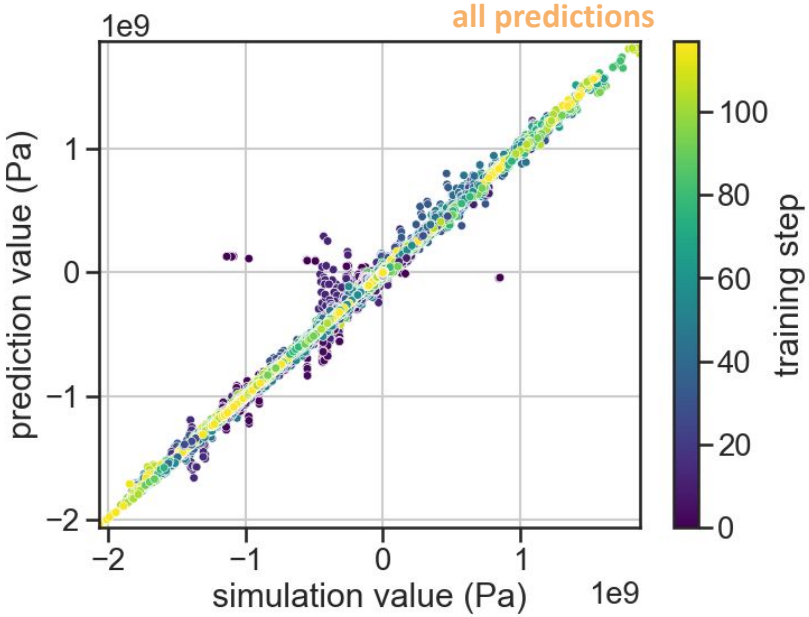
$$Var_i[\sigma_{q,i}^t]$$

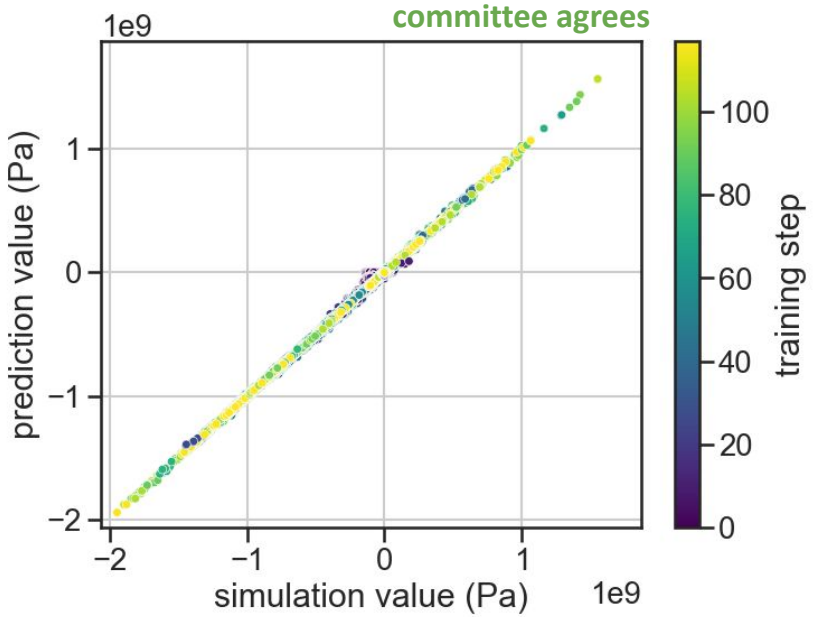
Can we trust the committee?

- reference configuration
  - committee size: 4
  - strain accuracy: 30 MPa
  - train frequency: 100 samples



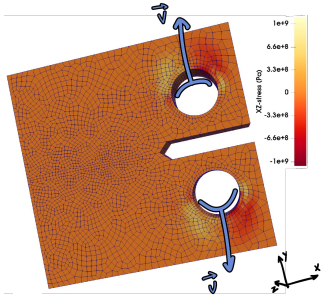


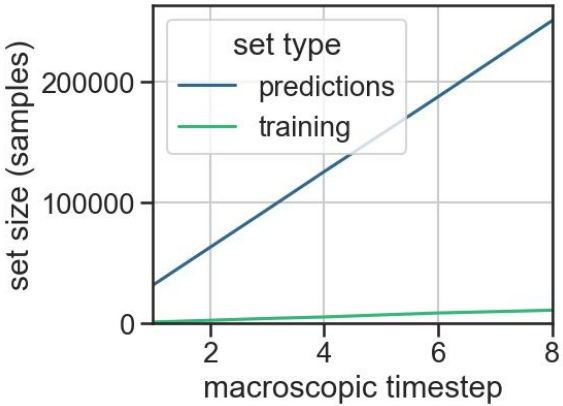


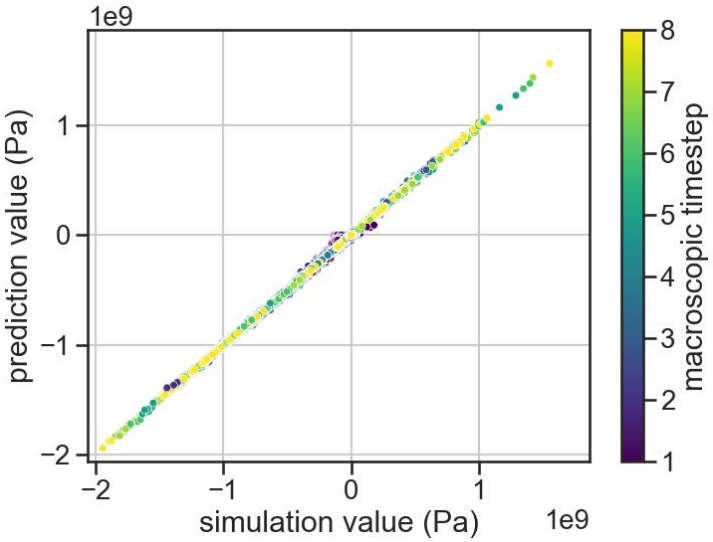


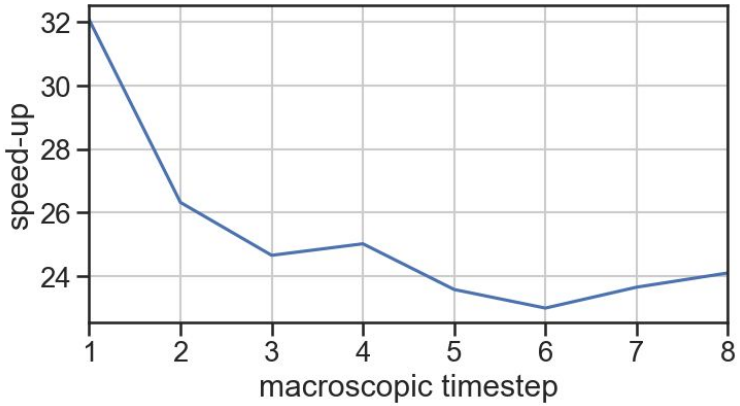
Accuracy, data requirements and speed-up

- reference configuration
  - committee size: 4
  - strain accuracy: 30 MPa
  - train frequency: 100 samples

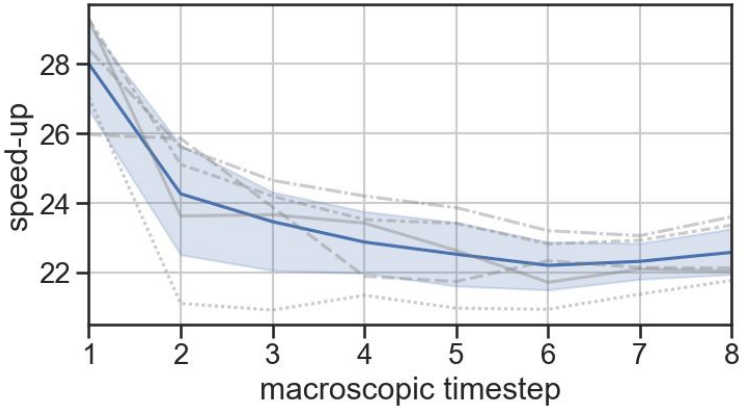


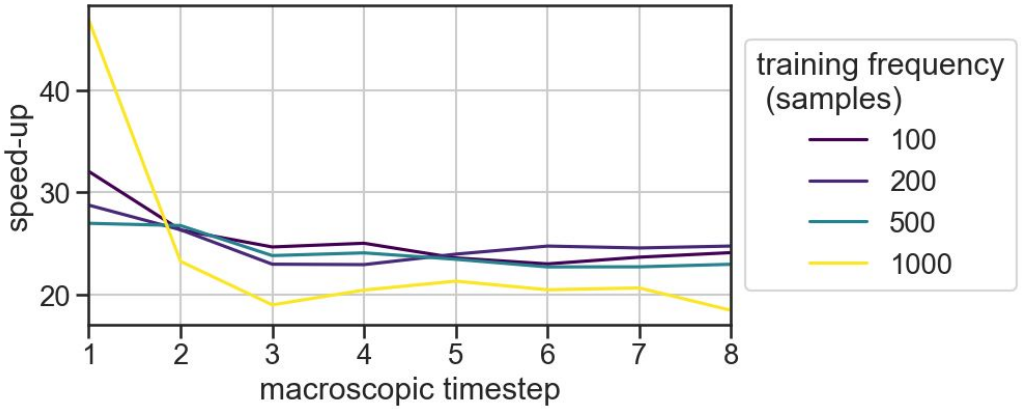


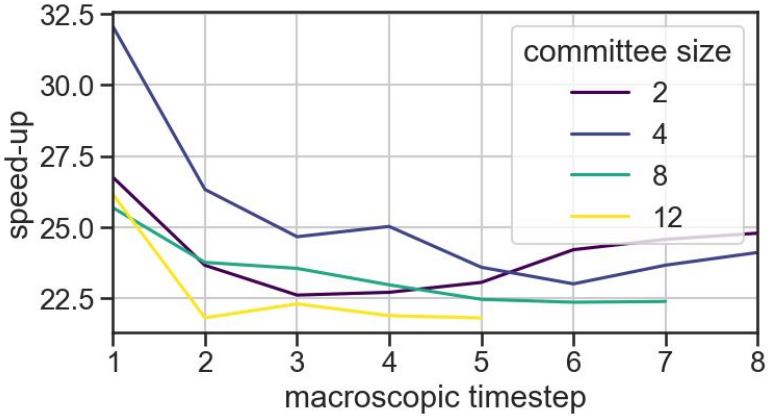


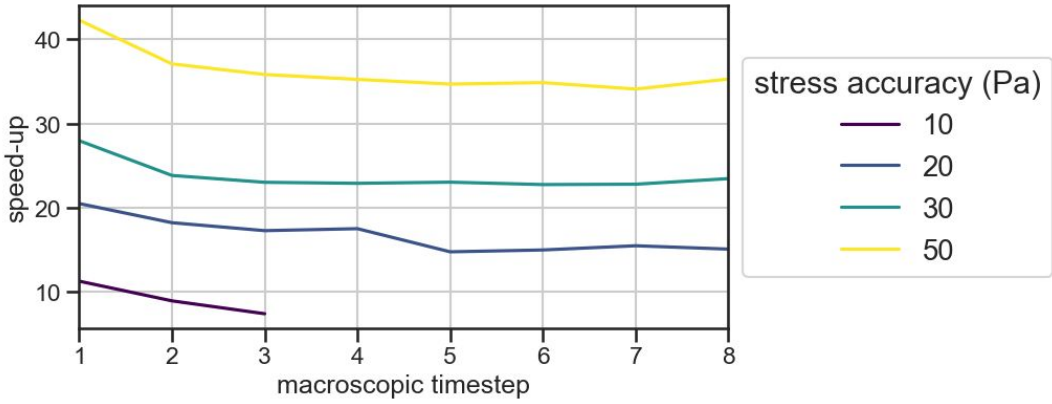


Influence of parameters: seed, committee size, accuracy and training frequency

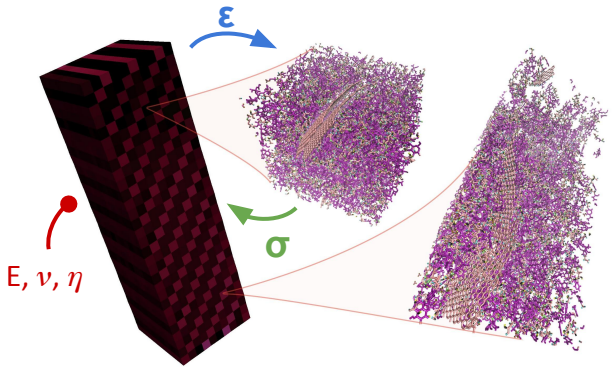






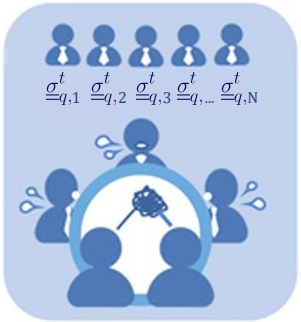


Conclusions



- **MD is already a powerful tool**
  - to capture some material properties
  - to understand physics at hardly accessible scales
  - **but!**
    - **MD has limited (scale) capacities**
    - **MD needs to be made more reliable (VVUQ, see other talk)**

- **multiscale strategies and data-based methods help**
  - **hierarchical**: cheap, efficient but limited transferability
  - **semi-concurrent**: expensive but widely applicable
  - **compromise**: using machine-learning approaches



- **unsupervised and supervised acceleration**
  - *a priori* assessment of ML model uncertainty
  - **10^1 to 10^2** theoretical speed-up of multiscale simulations

Acknowledgements

- Éric Robin
- Peter Coveney
- Werner Müller
- James Suter
- Wouter Edeling
- Robin Richardson
- Robert Sinclair
- and many others ...

CompBioMed
compbiomed.eu/

Horizon 2020
European Union Funding
for Research & Innovation

VECMA
vecma.eu/

SEAVEA

UK COMES
Consortium on Mesoscale Engineering Sciences

UKRI
Engineering and
Physical Sciences
Research Council

ucl.ac.uk/mesoscale-modelling-consortium/

lrz

epcc