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


A wide spectrum of multiscale strategies

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

SCEMa: Simulation Coupling Environment for Materials


SCEMa: https://github.com/UCL-CCS/SCEMa
Ubiquity of semi-concurrent approaches

- $\text{FE}^2$, FE-MD, ...

- mechanics but also fusion, climate applications


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

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

\[
\frac{1}{S_{1,2}} \alpha \sum_{n=1}^{n_{cp}} \sum_{i=1}^{6} \sqrt{\left(\tilde{e}_{1,n,i} - \tilde{e}_{2,n,i}\right)^{2}}
\]

Compute as few molecular simulations as possible

- **clustering** to minimise computational effort
- splines are considered similar if below arbitrary threshold $\alpha$
- **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)

\[
\frac{1}{s_{1,2}} \propto \sum_{n=1}^{n_{cp}} \sum_{i=1}^{6} \sqrt{(\tilde{c}_{1,n,i} - \tilde{c}_{2,n,i})^2}
\]

Multiscale simulation of a compact-tension test

Accurate global prediction with threshold $\alpha < 0.01 \rightarrow x3.03$ speed-up

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

- supervised machine learning can fit mechanical data

\[
\Delta S^t = \frac{\partial \psi^{t+\Delta t}}{\partial F^{t+\Delta t}} - S^t
\]
\[
D^{t+\Delta t} = \frac{\partial \psi^{t+\Delta t}}{\partial Z^{t+\Delta t}} \cdot \dot{Z}^{t+\Delta t}
\]

thermodynamically-admissible neural network

3D lattice material structure

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)
Adaptation of the concurrent multiscale workflow

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

\[
\begin{pmatrix}
\varepsilon^{t-1}_q, \
\sigma^{t-1}_q, \
\Delta \varepsilon^t_q
\end{pmatrix}_{\text{inputs}} \rightarrow \begin{pmatrix}
\sigma^t_q
\end{pmatrix}_{\text{outputs}}
\]

\[
t \leftarrow \text{timestep};
q \leftarrow \text{quadrature point};
\sigma_{\text{acc}} \leftarrow \text{MD stress accuracy};
N_C \leftarrow \text{committee size};
\text{trainset} \leftarrow [];
\]  
for \( i \leftarrow 0 \) to \( N_C \) do
\[
\begin{align*}
\tilde{\sigma}^t_{q,i} &= \text{Committee}(i, \varepsilon^{t-1}_q, \varepsilon^t_q, \sigma^{t-1}_q) \\
\end{align*}
\]
end
if \( \text{Var}[\tilde{\sigma}^t_{q,i}] < \sigma_{\text{acc}} \) then
\[
\sigma^t_q = E[\tilde{\sigma}^t_{q,i}]
\]
else
\[
\sigma^t_q = \text{MD}(\varepsilon^t_q);
\text{trainset} \leftarrow (\varepsilon^{t-1}_q, \varepsilon^t_q, \sigma^{t-1}_q, \sigma^t_q)
\]
end
Can we trust the committee?

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

Can we trust the committee?

<table>
<thead>
<tr>
<th>committee agrees</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.16</td>
<td>1.1</td>
</tr>
<tr>
<td>False</td>
<td>95</td>
<td>3.4</td>
</tr>
</tbody>
</table>

prediction is accurate
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 ...