Machine learning non-equilibrium potential energy surfaces sampled in virtual reality

David Glowacki, glowacki@bristol.ac.uk
Dept of Chemistry & Dept. of Computer Science
www.glow-wacky.com
Experimenting with the latest computational tools to enable new modelling workflows\textsuperscript{1}

Part I

Sampling non-equilibrium geometries in virtual reality using a real-time interactive MD framework

Part II

Fitting the sampled geometries to accurate Neural Net functional forms which enable efficient MD

\textsuperscript{1}Aspuru-Guzik, Lindh, Reiher, \textit{The Matter Simulation (R)evolution}, \textit{ACS Cent. Sci.}, 4, 2, 144-152
The application focus of this talk: CN scattering off liquid squalane surfaces
The application: CN scattering off liquid squalane surfaces
The challenge of modelling systems like these

• CN at squalene liquid surfaces involves thousands of degrees of freedom, with a wide range of possible reaction pathways
• The electronic structure is a challenge (DFT & HF approaches bread down) and we therefore require the use of methods which can accurately treat electron correlation
• Accurate modeling requires good statistics (thousands of trajectories)
• The experimental measurements are carried out in a regime which is far from equilibrium
Part I: Sampling reaction pathways using real-time interactive molecular simulation in VR
VR for accelerating computational tasks
Larry Wallen, Dreamworks CEO

• In 2009, DreamWorks made a multi-million dollar strategic decision to go beyond the mouse & keyboard, investing in HCI and VR technologies so that digital animators could reach into an animation scene, manipulate an object, and see how their modifications impacted the animation sequence.

• From 2009 – 2014, Dreamworks saw a 3x reduction in the computational cost and man-power required to produce a commercial-grade digital animation sequence, saving tens of millions every year

Keynote lecture, Supercomputing 2014
What kind of task is each computer best suited to solve?
VR has a strong track record for improving scientific workflows in the medical field.
You definitely want a VR-trained surgeon!

Figure 3. Mean duration of operative procedure for the VR and ST groups.

Figure 4. Total error number for each error type. LOP, lack of progress; GBI, gallbladder injury; LI, liver injury; intraperitoneal, incorrect plane of dissection; BNT, burn nontarget tissue; TT, tearing tissue; IOV, instrument out of view; AT, attending ...

Figure 5. Total number of errors scored per procedure for VR and ST groups. The mean number of errors per procedure was significantly greater in the ST group than in the VR group ($P < .006$).
O’Connor, Deeks et al., arXiv:1801.02884, Science Advances, 2018
Interactive Molecular Dynamics using Virtual Reality

O’Connor, Deeks et al., arXiv:1801.02884, Science Advances, 2018
Using external forces to bias molecular dynamics simulations

Velocity Verlet Integration algorithm, for simulating atomic dynamics

\[
\begin{align*}
x(t + \Delta t) &= x(t) + v(t)\Delta t + \frac{F(t)}{2m} \Delta t^2 \\
v(t + \Delta t) &= v(t) + \frac{F(t + \Delta t) + F(t)}{2m} \Delta t
\end{align*}
\]

where \( F(t) = -\frac{dV}{dq} \)

Split up the force into two separate components

\[
\begin{align*}
V &= V_{\text{int}} + V_{\text{ext}} \\
\frac{dV_{\text{ext}}}{dq} &= \frac{mj}{\sigma^2} \left( q_j - g_i \right) e^{-\|q_j - g_i\|^2/2\sigma^2} \\
\frac{dV_{\text{int}}}{dq} &\in \text{classical (CHARMM, AMBER, GROMACS)} \\
&\quad \text{quantum (DFTB, semiempirical, DFT)}
\end{align*}
\]

\[
F(t) = -\frac{dV_{\text{int}}}{dq} - \frac{dV_{\text{ext}}}{dq} = F_{\text{int}} + F_{\text{ext}}
\]

Glowacki et. al, Faraday Discussion 169, 2014
O’Connor, Deeks et al., arXiv:1801.02884, Science Advances, 2018
Software is available for download on itch.io


O’Connor, Deeks et al., arXiv:1801.02884, Science Advances, 2018
Trypsin + benzamidine binding

Interactively Docking Benzamidine Into Trypsin

O’Connor, Deeks et al., arXiv:1801.02884
Modular Design

- Generic server/client model. Does not have to be a molecular simulation!
- Connect to existing simulation codes:
  - GROMACS
  - OpenMM
  - LAMMPS
  - DFTB+
  - DL-POLY
- Containerized design for easy deployment on generic cloud architectures
CPU & GPU multi-threading allows for very fast force calculations, to reach interactive latencies.

Cloud Latency

![Graph showing latency times for different cable connections between Bristol and other locations. The x-axis represents time in milliseconds, and the y-axis represents probability. The graph compares Bristol-Frankfurt, Bristol-Ashburn, and Bristol-Phoenix cable connections. The Bristol-Frankfurt cable has a higher probability of lower latency times compared to the other two connections.](image-url)
Simulation Size is Not (Always) the Bottleneck

Current VR Rendering Limit!
Discovering reaction pathways in VR

Quickly generating reaction pathways for CN + squalane in VR (20,000 geometries in total)

Part II: fitting geometries along the reaction pathways to efficient functional forms
Sample reactive pathways in VR at low level of theory (semi-empirical/DFTB)

Refine the pathways at a higher level of theory (embedded coupled cluster/DFT)

Fit the high level pathways to an efficient representation (Neural Nets or MS-EVB)

Run dynamics

Analyze
CN + squalane potential energy surface close to the minimum energy path [CCSD/CBS]
CN + squalene potential energy [CCSD(T)] far from the minimum energy path
Neural Network Fitting (Coulomb Matrix) far from the minimum energy path

Mean absolute error (MAE) = 7 kJ/mol

Neural Network Fitting (Global SLATM) far from the minimum energy path

MAE = 2.2 kJ/mol

Huang & von Lilienfield, The “DNA” of chemistry: Scalable quantum machine learning with amons arXiv, 2017
Neural Network Fitting (Local SLATM) far from the minimum energy path

MAE = 1.2 kJ/mol

Huang & von Lilienfield, The “DNA” of chemistry: Scalable quantum machine learning with amons arXiv, 2017
Neural Network Fitting (atom-centred symmetry functions) far from the minimum energy path

MAE = 1.6 kJ/mol

J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials
JCP 134, 074106 (2011)
Conclusions

• New technologies lets us imagine new workflows for accelerating molecular research workflows – e.g., machine learning on human-sampled structures

• There’s lots of interesting opportunities to explore using this technology, in both research & education

• User-controlled HCI studies testing various aspects of this framework are encouraging

• We’re working to connect the VR iMD framework to machine learned forces

• By ‘gamifying’ reaction discovery in VR, can we use reinforcement learning to develop new search algos for kinetic network mapping?
That’s all!
www.glow-wacky.com
glowacki@bristol.ac.uk
Or else find me on google
Thanks very much to the funders
Opportunities: PhDs
Opportunities: Post-Docs
What should we be using VR for?

- What can we do in VR that we cannot do in any other way?
  - Touching world we cannot otherwise touch
  - Quickly manipulating objects that we could not otherwise manipulate
  - Quickly & clearly communicating & understanding concepts that would otherwise take a long time
  - Effective training

- I think it’s very important to try & make an effort to use VR for those tasks that are a struggle for conventional displays
Commodity VR now solves the problem of 3d co-location

- **Co-location** refers to a users’ ability to integrate visual & proprioceptive signals to simultaneously align position in physical space with position in virtual (or simulated) space
- Tablet interfaces solve the problem of 2d co-location
- Solving the problem of real-time 3d co-location has been a long-standing problem
- Many studies suggest that co-location is more important to user experience than haptic feedback

Swapp, Pawar, Loscos, *Virtual Reality*, 2006, 10, p 24
Another application: Linking European Biotech networks
Global descriptor

1 configuration = 1 descriptor

Distance matrix
Local descriptor

1 configuration with \textbf{N atoms} = \textbf{N} descriptors

\begin{tabular}{ccc}
A & B & C \\
\hline
A & 0 & 1.5 & 2.1 \\
B & 1.6 & 0 & 2.0 \\
C & 0.6 & 1.5 & 0 \\
\end{tabular}

2-Body terms
Local descriptor

1 configuration with \textbf{N atoms} = \textbf{N} descriptors

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th></th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1.5</td>
<td>2.1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1.7</td>
</tr>
<tr>
<td>B</td>
<td>1.6</td>
<td>0</td>
<td>2.0</td>
<td></td>
<td>0</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0.6</td>
<td>1.5</td>
<td>0</td>
<td></td>
<td>1.1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

2-Body terms

3-Body terms
Descriptors

Global

• Inverse distance matrix
• Coulomb matrix
• SLATM

Local

• Atom Centred Symmetry Functions
• aSLATM

Spectrum of London and Axilrod-Teller-Muto