A comprehensive environment for property prediction and force field development

Hannes Schweiger, Paul Saxe
Materials Design Inc.
Agenda

- **MEDEA software platform**
- **Properties**
- **Examples:**
  - Vapor Pressure and VLE
  - Azeotropes
  - Phase Diagram
  - Thermal conductivity
  - Viscosity

- **Intrinsic Issues in Calculations**
- **Challenges for industrial R&D**
  - Time Pressure
  - Multiple Projects
  - Restricted Environment
- **Assessment**
MEDEA SOFTWARE PLATFORM
Purpose

Predict and understand materials properties

Approaches:
- Ab initio electronic structure calculations
- Forcefield simulations
- Statistical mechanics
- Analytical theory
- Empirical correlations
- Experimental data of existing materials (databases) as reference
Materials modeling: MedeA capabilities

- Device
- Finite element
- Macroscopic strain/stress
- Heat flux, ...
- Force field
- Atomistic
- Semi empirical
- First principles
  - Density Functional (DFT)
  - Hybrid Methods
  - Hartree-Fock
- LAMMPS
- Gibbs ensemble Monte Carlo
- MOPAC
MedeA’s Three Tier Architecture

User Interface

Model creation and analysis of results

Local databases (expt. and computed)

Job Server

Job control and job databases

Task Servers

Local or remote

Computations
VASP  GIBBS  LAMMPS  MOPAC
Integration with
Queuing System: PBS, LSF, …
PROPERTIES
Forcefield Methods in MedeA

- Focus on materials properties including:
  - Density and structure
  - Vapor pressure
  - Solubility/miscibility
  - Liquid-vapor phase diagrams, critical points
  - Thermal conductivity
  - Viscosity
  - Mechanical properties
  - Diffusion
- Selection, editing, and control of forcefields
- Automated atom-typing and forcefield assignment
- Flow-chart control of computational stages
- Full use of parallel architectures
- Property-oriented analysis tools
- Estimation of (statistical) error bars
Forcefield Control

- **Automated atom type assignment** using the templates section of Materials Design’s forcefield file *.frc
- **Wildcards**: major simplification in the angle and torsion terms – critical for more complex simulations with GIBBS and LAMMPS
- **Atom type equivalences** for nonbonds, bonds, angles, torsions, etc.
- **Versioning**: each parameter has its own version, so updates do not remove older parameters but override them
- **“Include” capability**: a user can modify a forcefield by including the original, adding parameters and, by using version numbers, override parameters in the original in a well organized and controlled manner
Classes of Materials

- Gases and liquids including high pressure and temperature
- Organic materials ranging from small molecules to polymers
- Inorganic materials (e.g. crystalline, amorphous materials, glasses, molten materials)
- Semiconductor materials (Si, Ge, III-V, ...)
- Metals and alloys
- Nanostructures
- Interfaces
**Computed Materials Properties**

**Structural properties**
- Molecular structures
- Crystal structures
- Surface structures
- Structure around defects
- Adsorption geometries
- Structures of interfaces
- Liquids and amorphous systems

**Thermo-Mechanical properties**
- Elastic moduli
- Speed of sound
- Vibrational properties
- Thermal expansion coefficients
- Fracture

**Thermodynamic properties**
- $\Delta U$, $\Delta H$, $\Delta S$, $\Delta G$, heat capacity
- Binding energies
- Solubility
- Melting temperature
- Vapor pressure
- Miscibility
- Phase diagrams
- Surface tension

**Chemical properties**
- Chemical reaction rates in gases and condensed phases
- Reactivity on surfaces
- Solid-solid reactions
- Pressure-induced reactions
- Photochemical reactions

**Transport properties**
- Mass diffusion coefficient
- Permeability
- Thermal conductivity
- Viscosity

**Electronic, optical, and magnetic properties**
- Electron density distribution - electrical moments
- Polarizabilities, hyperpolarizabilities
- Optical spectra
- Dielectric properties
- Piezoelectric properties
- Electrostatic potential
- Spin density distribution, magnetic moments
- Energy band structure - metal, semiconductor, insulator, superconductor
- Band gaps, band offsets at hetero-junctions
- Ionization energies and electron affinities
- Work function
EXAMPLES
Carbon Dioxide

- Vapor pressure
- Density gas-liquid
- Ethane-CO$_2$ azeotrope
- CO$_2$-SO$_2$ phase diagram (Lachet et al., 2009)
- Viscosity
- Thermal conductivity

Gibbs ensemble Monte Carlo
Biased configurational averages

LAMMPS
Results for CO₂

Vapor pressure

Density of liquid and gas phase

Critical point
**Ethane-CO$_2$ Azeotropic**

**C$_2$H$_6$ – CO$_2$**

Standard GIBBS AUA potentials

Computed concentration of CO$_2$ of azeotrope at 34 bar:
[CO$_2$]=0.64
boiling point of -0.6 °C

Experiment (Nordstad et al.):
[CO$_2$]=0.7
boiling point of -6 °C
Thermodynamic behavior of the CO$_2$ + SO$_2$ mixture: experimental and Monte Carlo simulation studies

Véronique Lachet$^a$*, Theodorus de Bruin$^a$, Philippe Ungerer$^a$, Christophe Coquelet$^c$, Alain Valtz$^c$, Vladimir Hasanov$^b$, Frederick Lockwood$^b$, Dominique Richon$^c$

$^a$IFP, 1-4 avenue de Bois Préau, 92852 Rueil-Malmaison, France
$^b$Air Liquide, 1 chemin de la Porte des Loges, Les Loges-en-Josas BP 126, 78354 Jouy-en-Josas, France
$^c$Mines Paris-Tech, CEP/TEP, Centre Energétique et Procédés CNRS FRE 2861, 35 rue Saint Honoré, 77305 Fontainebleau, France
Thermal Conductivity

**CO₂**

- \( P = 500 \text{ bar} \)
- \( V = 2.89 \times 2.89 \times 8.68 \text{ nm}^3 \)
- \( T = 298.2 \text{ K} \)
- \( \rho = 1.0341 \text{ g/cm}^3 \)

5 ns molecular dynamics

- \( \lambda_{\text{comp}} = 0.141 \pm 0.017 \text{ W m}^{-1} \text{K}^{-1} \)
- \( \lambda_{\text{expt.}} = 0.1385 \text{ W m}^{-1} \text{K}^{-1} \)

Reverse non-equilibrium molecular dynamics

MedeA-LAMMPS

COMPASS forcefield
CO$_2$

$P = 494$ bar
$V = 2.89 \times 2.89 \times 8.68$ nm$^3$
$T = 298.2$ K
$\rho = 1.0341$ g/cm$^3$

0.1 ns molecular dynamics

$\eta_{\text{comp}} = 1.360 \times 10^{-4}$ Pa s
$\eta_{\text{expt.}} = 1.3466 \times 10^{-4}$ Pa s (NIST)

Reverse non-equilibrium molecular dynamics
MedeA-LAMMPS
COMPASS forcefield
INTRINSIC ISSUES WITH CALCULATIONS
<table>
<thead>
<tr>
<th>Liquid</th>
<th>Temperature (K)</th>
<th>Density (expt) (g/cm³)</th>
<th>Density LAMMPS/OPLSAA (g/cm³)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Butane</td>
<td>273</td>
<td>0.6013</td>
<td>0.6010 ± 0.0026</td>
<td>-0.05%</td>
</tr>
<tr>
<td>Isobutane</td>
<td>273.2</td>
<td>0.58052</td>
<td>0.6046 ± 0.0039</td>
<td>4.15%</td>
</tr>
<tr>
<td>Isobutane</td>
<td>298.2</td>
<td>0.55059</td>
<td>0.5709 ± 0.0056</td>
<td>3.69%</td>
</tr>
<tr>
<td>Pentane</td>
<td>298.2</td>
<td>0.62074</td>
<td>0.6156 ± 0.0032</td>
<td>-0.83%</td>
</tr>
<tr>
<td>Isopentane</td>
<td>298.2</td>
<td>0.61516</td>
<td>0.6261 ± 0.0032</td>
<td>1.78%</td>
</tr>
<tr>
<td>Neopentane</td>
<td>298.2</td>
<td>0.58435</td>
<td>0.6276 ± 0.0036</td>
<td>7.41%</td>
</tr>
<tr>
<td>Methanol</td>
<td>298.2</td>
<td>0.786</td>
<td>0.7797 ± 0.0021</td>
<td>-0.80%</td>
</tr>
<tr>
<td>Ethanol</td>
<td>298.2</td>
<td>0.78509</td>
<td>0.7955 ± 0.0021</td>
<td>1.30%</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>300</td>
<td>0.7795</td>
<td>0.8088 ± 0.0036</td>
<td>3.76%</td>
</tr>
<tr>
<td>1,2-butanediol</td>
<td>373.2</td>
<td>0.9394</td>
<td>0.9591 ± 0.0037</td>
<td>2.10%</td>
</tr>
<tr>
<td>Glycerol</td>
<td>373.2</td>
<td>1.209</td>
<td>1.1893</td>
<td>-1.60%</td>
</tr>
</tbody>
</table>
Analyzing the Results

Decane Thermal Conductivity OPLS

Converged?

What Value?

Thermal Conductivity (W/m.K)

J0Jt

lambda (W/m/K)

t (fs)
Which ARE the Right Results?

Decane at 480 K and 178 atm

<table>
<thead>
<tr>
<th>SHAKEing?</th>
<th>COMPASS</th>
<th>OPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Yes</td>
<td>0.16</td>
<td>0.13</td>
</tr>
<tr>
<td>(experiment)</td>
<td>0.10015</td>
<td>W/m.K</td>
</tr>
</tbody>
</table>

SHAKEing C-H bonds and H-C-H.

Even at 480 K C-H stretch and the H-C-H and C-C-H bends will not be active. What about C-C-C?

How does this affect calculated thermal conductivity?
CHALLENGES FOR INDUSTRIAL R&D
Immediate Proof of Concept
- Can we calculate “x” for “y”?
  - Very short timeframe – hours or perhaps a day or two
  - What – exactly – is “y” in terms of an atomic model?
  - Need to answer “What is the accuracy of the result?”
- Wide range of properties and systems – no one tool fits all

Projects
- Tight schedules
- Aren’t simple, by definition!
- Usually require a large number of calculations – 100’s
- One of many simultaneous projects

Proprietary
- Not easy to find related work without revealing topic
- How to get help with problems with a specific system?
Challenges of Industrial Environment

- Lack of control over computing resources
  - Shared resources
  - IT restrictions
    - Type of hardware
    - Network access
    - Security
    - Changing passwords monthly

- Software Environment
  - Must be robust, robust, robust!
  - Handling of a large number (>10,000/yr) of a wide range of calculations
  - Well automated, including analysis
  - Checking for errors
  - Providing a history
  - Error bars

As Anne Chaka of NIST comments: “We give away the data for free – but we charge for the error bars!”
From code to property

- **Programmer’s Mind**
  - `#kpoints`
  - Optimize program
    - Platform specific
    - Special cases:
      - Non-magnetic
      - Independent blocks
      - Block/group per node
  - Expert in code
  - Full Focus

- **Engineers Mind**
  - Similar `k-spacing`
  - General approach
    - Platform independent
    - Keep procedure
      - Change molecule
      - Change process property
      - Use more computing power
  - Expert
  - One of many tools
Assessment

- Computer simulations have become a viable source of many industrially important thermodynamic property data providing
  - Consistent and complete dataset
  - Resolution of conflicting experimental data
  - Extrapolation to regions where experiments are too difficult, costly, time consuming, or dangerous
  - Understanding on deep level
- We are still in an early stage; accuracy and cost-efficiency will accelerate deployment
- The key bottleneck is “cognitive access”,
  - scientists and engineers who understand industrial needs
  - have the skills to formulate meaningful simulations
  - and control the error bars
Our technological world faces enormous challenges
The solutions – for better or worse – must rely on technology
We must always do better!
Atomistic modeling can play an important, growing role
Two Routes Forward

- Open Source/Content
- Commercial Companies
Integrating these methods and tools is much more work than developing them in the first place – 10x?

This is science, not software

The intellectual investment is enormous

Think chips and computers, not Linux or Wikipedia
Example: Forcefields

- Forcefields are highly interconnected: new parameters must work with and extend preceding parameters
- Require large effort to validate
- Require consistency
- Are a dynamic compromise between coverage and accuracy
- Typically the work of a single group, headed by one person
Organic Forcefields

• No publicly available accurate organic forcefields!
  • OPLS … sort of, but quite uneven
  • MM4 … ?
  • COMPASS … commercial

• No longer actively developed
  – Last major efforts were at Accelrys > 10 years ago
  – Allinger and MM4 8-10 years ago

• Academic forcefield work focussing on new frontiers
  – Reactive forcefields
  – Polarizable forcefields
  – Etc.

  – Where do the next generation of forcefields come from?
What Industry Needs

• Modeling tools that are so useful that they cannot be ignored!

• Robust, competitive commercial products

• Leading edge academic research moving into products over time

• Intel and AMD to compete and compete and...

• National Labs to demonstrate large scale use, also advance high computing