Representability problems for coarse-grained potentials

Ard Louis
Emergent physics/chemistry and coarse-graining

"More is Different".
P.W. Anderson,
Science 177, 393-396 (1972).

-Micellar phases emerge from geometry:
-SAW and polymer scaling
-Hard-spheres and structure/crystallization
How to model DNA self-assembly?

- Atomistic models
  - orders of magnitude too slow
- Bottom-up coarse-graining
  - Representability problems

- We use **top-down coarse-graining** instead
  - Self-assembly primarily determined by: chain-like molecule with specific binding
Creating a Minimal Model - a Top-Down Approach

Choose a minimal set of interactions and degrees of freedom.

Design a model which incorporates these interactions.

Is the discrepancy qualitative or quantitative?

Test - does it reproduce known properties?

Apply it to understand other systems.
Creating a Minimal Model - a Top-Down Approach

Choose a minimal set of interactions and degrees of freedom.

Design a model which incorporates these interactions.

Test - does it reproduce known properties?

- yes: Apply it to understand other systems
- no: Is the discrepancy qualitative or quantitative?

qualitative
quantitative
Creating a Minimal Model - a Top-Down Approach

Choose a minimal set of interactions and degrees of freedom.

Design a model which incorporates these interactions.

Is the discrepancy qualitative or quantitative?

Test - does it reproduce known properties?

yes

no

Apply it to understand other systems.
In DNA competition of 2 length-scales leads to double helix

In DNA single strands are flexible and can stack.

Competition between two length-scales leads to the double helix.

Simple coarse-grained DNA model

- Interactions
  - H-bond between complementary bases
  - Stacking between bases
  - Backbone: FENE spring
  - Helicity emerges naturally
    - Propellor twist emerges naturally
  - But no minor/major groove

- Disordered single strand
- Stacked single strand
- Hybridized double strand
Duplex formation & length dependence

Good agreement of $T_m$ with $L$ is a measure of the cooperativity of the transition – influenced by the single strand cooperativity.

- Melting Temperature / K
- Number of Bases
The width of the transition is related to how well you can predict the concentration dependence of the melting temperatures.
Free-energy profile for duplex formation

fraying

formation of a 15mer duplex
Mechanical properties

- Duplex ~ 125 bp
- Unstacked single strand ~ 2-4 bases
- Fully stacked single strand ~ 64 bases
- Twist persistence length of duplex ~ 3.74°/bp
Mechanical properties – many subtleties we can’t get

dsDNA undertwists upon initial stretching
Sequence dependent elastic properties are very very subtle – need a much better representation of excluded volume etc. . . .
Strand displacement reaction

(a) (b) (c)

Y

X(n)

n

S

Displacement (or strand-exchange) reactions are key to dynamical behaviour of DNA nanodevices, and DNA computation.

Rates change by orders of magnitude as toehold length increases, before leveling out at \( nt \sim 5 \)

David Yu Zhang and Erik Winfree, JACS, 131, 17305 (2009)
Displacement (or strand-exchange) reactions are key to dynamical behaviour of DNA nanodevices, and DNA computation.
Coaxial stacking is important; displacement slowed by several effects.
DNA computing OR gate

Neural network computation with DNA strand displacement cascades

Simulation on GPU equiv ~ 10 min real time
Just another pretty movie?

The first principle is that you must not fool yourself--and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that.

-- R.P. Feynman,

“Cargo Cult Science” (1974)
Einstein’s Razor

Make things as simple as possible, but no simpler.

What he really said:
It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

The Herbert Spencer Lecture, delivered at Oxford (10 June 1933)

Survival of the fittest
Einstein’s Razor & Coarse-grainer’s nightmare

Make things as simple as possible, but no simpler.

You work really hard to derive a tractable coarse-grained system, only to find …..

Survival of the fittest
Einstein’s Razor & Coarse-grainer’s nightmare:

Wrong: too simple!!!

Make things as simple as possible, but no simpler.

Wrong: too simple!!!
No free lunch theorems

- Effective-potentials for depletion systems
  - Representability and potentials
    - A) Two and three-body potentials
    - B) Isotropic model for water
  - Dynamics?
  - DNA?
Case-study 1: Depletion interactions

Coarse-grain by equating partition functions (or all distribution functions of CG variables)

Only maps to an *effective Hamiltonian system* with effective potentials in grand-canonical or semi-grand ensembles.

In a *pure canonical ensemble*, one cannot decompose the potential into pairwise, three-body etc. terms… see e.g. AAL, J. Phys.: Condens. Matter **14**, 9187 (2002)

Case-study 1: Depletion interactions: Lessons?

1. Coarse-graining by equating partition functions or distribution functions is really a re-summation -- formally this could be done by tracing out any variable.

2. Be careful which ensemble you work in if you want to *map* to an effective potential or Hamiltonian system. [Canonical is suspect].

3. Phase-diagrams and thermodynamics can be worked out, but be careful to include zero-body and one-body terms – remember McMillan-Mayer.

4. Dynamics must be treated with care.

Case-study 1b: Debye-Huckel potentials

Charge-neutrality means Canonical ensemble is natural & is effectively a one-component system

\[ \beta v_{DH}(r; \rho) = \frac{Z^2}{r} \exp[-\kappa(\rho)r]. \]

\[ \kappa(\rho) = \sqrt{4\pi Z \rho} \]

\[ Z_c = \frac{\beta P}{\rho} = \int_0^\rho \frac{\partial \beta P(\rho^{'})}{\partial \rho^{'}} \frac{d\rho^{'}}{\rho} = \int_0^\rho \left[ 1 - \rho^{' \hat{c}}(k = 0; \rho^{'}) \right] \frac{d\rho^{'}}{\rho} \]

In RPA approximation \( c(r) = -\beta v(r) \)

\[ \beta \hat{v}(k = 0; \rho) = \frac{4\pi Z^2}{\kappa^2} = \frac{Z}{\rho}, \]

\[ Z_c^{RPA} = 1 + Z, \]

\[ Z_{vir} = \left( 1 + \frac{1}{2} Z \right) - \frac{2}{3} \beta \pi \rho \int r^2 \left[ h(r) \left( r \frac{\partial v_{DH}(r)}{\partial r} \right) - g(r) 3 \rho \frac{\partial v_{DH}(r)}{\partial \rho} \right]. \]

These terms reduce \( Z_{vir} \) further
No free lunch theorems

- Effective-potentials for depletion systems
- Representability and potentials
  - A) Two and three-body potentials
  - B) Isotropic model for water
- Dynamics?
Case study 2: Representability problems for pair potentials

Coarse-graining throws away information (No free lunch)

Question: when does this matter?

Intuitions for effective potentials $v^{\text{eff}}(r)$ …:

representability problems: You can’t simultaneously represent all the properties of the underlying system at one state point with one coarse-grained potential. AAL, J. Phys.: Condens. Matter 14, 9187 (2002)

Contrast with transferability: different $v^{\text{eff}}(r)$ at different state points
Original system

\[ H = K + \sum_{i<j} w^{(2)}(r_{ij}) + \sum_{i<j<k} w^{(3)}(r_{ij}, r_{ik}, r_{ki}) \]

Coarse-grained system

\[ v^{\text{eff}}(r) = w^{(2)}(r) + \delta v(r) \]

Energy route:

\[ U(N, V, T) = \frac{1}{2} \rho^2 \int d\mathbf{r}_1 d\mathbf{r}_2 g(r_{12}) w^{(2)}(r_{12}) \]
\[ + \frac{1}{6} \rho^3 \int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 g^{(3)}(r_{12}, r_{23}, r_{31}) \times w^{(3)}(r_{12}, r_{23}, r_{31}) \]

Structure route:

Same \( g(r) \) in coarse-grained system

\[ v^{\text{eff}}(r) = w^{(2)}(r) + \delta v_g(r) \text{ (unique)} \]

Comparison

\[ \frac{\delta v_U(r)}{\delta v_g(r)} = \frac{1}{3} + O[(w^{(3)})^2; \rho^2] \]

Representability problems: one potential can’t simultaneously represent multiple properties of the system

**H₂O with spherically symmetric pair potentials?**

\[ v_g(r) \leftrightarrow g_{OO}(r) \text{ - "bottom up" from Tip4pEw water} \]

**The potential is unique**


Thermodynamics through compressibility route

\[
\rho k_B T \chi_T = 1 + 4 \pi \rho \int r^2 [g_{AB}(r) - 1] dr,
\]

\[ v_g(r) \text{ changes with density,} \]
\[ \Rightarrow \text{ Transferability problems for } g(r) \]

Representability problems are severe

**Virial eos**

**internal energy**

\[ \delta U(r) = \frac{\sum'_{\alpha,\beta} g_{\alpha\beta}(r)v_{\alpha\beta}(r)}{g_{AB}(r)} \]

\( Z/Bp/p \)

**\( v_g(r) \) structure route**

**\( v_U(r) \) – energy route**
Representability problems are severe

B. Structure

By construction, our coarse-graining procedure generates a family of unique state-dependent single site potentials that reproduce the oxygen-oxygen pair correlations exactly. To compare the three-body correlations of the TIP4P-Ew and family of coarse-grained models, we measure the bond angle distribution as an integral over the full three-body correlation function and measure the distribution of angles generated by the neighbors of each molecule within a specified radius, $R_c$.

$$
\frac{\text{b}^2}{\text{H}_2\text{O}} = \frac{8}{\text{H}_2\text{O}} \int_0^{R_c} r_{12}^2 r_{13}^2 \sin^2 \theta_{123} \, dr
$$

where the triplet correlation of Eq. for a spherically symmetric potential is now dependent only on the magnitude of the vectors connecting two nearest neighbors to a central particle, and the angle between the two vectors and normalizes $b$ to a probability distribution. In Fig. 5 we compare the bond angle distributions for TIP4P-Ew and for the isotropic family. While the isotropic potentials do generate a peak at the tetrahedral angle, they show a marked increase in close-packed configurations at 60° corresponding to a defective network structure. Even though the absolute three-body correlations are very different between the isotropic family and TIP4P-Ew due to this defective structure, both models show the same trends with density, namely, a loss of structural order under compression. Thus it is evident that in contrast to the virial pressure which, besides having the incorrect value at a given state point, also showed physically incorrect trends when comparing state points, the family of isotropic potentials does exhibit a structurally anomalous region, if the appropriate potential is taken at each state point. By contrast, if a single isotropic potential is taken and used at different state points, it typically does not reproduce these structural anomalies.

C. Diffusion

While the coarse-grained particles' translation diffusion is an order of magnitude faster than the TIP4P-Ew water molecules, both models display an anomalous increase in diffusivity with compression at several isotherms. In Fig. 6 we show that the TIP4P-Ew and isotropic potentials.

FIG. 5. Bond angle distributions at $T=235.5$ K as a function of density for (a) TIP4P-Ew and (b) isotropic potentials. $R_c$ was chosen as 3.4 Å.
Compromise by fitting to multiple properties?

\[ U(N, V, T) = 2 \pi \frac{N^2}{V} \int_0^\infty g_{AB}(r)v_{U}^{\text{eff}}(r)r^2 dr, \]
\[ P = \rho k_B T - \frac{2\pi \rho^2}{3} \int r^3 dr g(r) \frac{dv(r)}{dr} \]
\[ \rho k_B T \chi_T = 1 + 4\pi \rho \int r^2 [g_{AB}(r) - 1] dr, \]

Representability:
A single potential can’t represent all properties simultaneously – All potentials are at best compromises

Fitting too tightly to one property (e.g. structure) may increase errors in another (e.g. thermodynamics)

See also Kremer/Mainz army on H\_2O
Corrections to virial equation from density dependence?

\[ Q(N, V, T) = \frac{\Lambda^{-3N}}{N!} \int dr^N \exp\left\{ -\beta \sum_{i<j} v(r_{ij}; \rho) \right\} \]

\( \Lambda \) is the usual thermal de Broglie wavelength. The volume derivative in

\[ \beta P = \left( \frac{\partial \log Q(N, V, T)}{\partial V} \right)_{N,T} \]

\[ Z_{vir}^\rho = \frac{\beta P}{\rho} = 1 - \frac{2}{3} \beta \pi \rho \int_0^\infty r^2 \left\{ r \frac{\partial v(r; \rho)}{\partial r} - 3 \rho \frac{\partial v(r; \rho)}{\partial \rho} \right\} g(r) \, dr \]

—Coarse-grained potentials do not generate a Hamiltonian system

“We record our opinion that the use of density dependent effective pair potentials can be misleading unless it is recognized that these are mathematical constructs to be used in specified equations rather than physical quantities”

New equations with your coarse-graining scheme?

\[
U(N, V, T) = 2\pi \frac{N^2}{V} \sum_{\alpha, \beta} \int v_{\alpha\beta}(r)g_{\alpha\beta}(r)r^2dr, \quad \delta v_U(r) = \frac{\sum'_{\alpha, \beta} g_{\alpha\beta}(r)v_{\alpha\beta}(r)}{g_{AB}(r)},
\]

\[
U(N, V, T) = 2\pi \frac{N^2}{V} \int_0^\infty g_{AB}(r)v_{U}^{\text{eff}}(r)r^2dr, \quad v_{U}^{\text{eff}} = v_{AB}(r) + \delta v_U(r),
\]

Representing properties differently in your CG scheme, e.g. take into account:

1) Entropy differences etc…
2) One-body terms?
3) Phase-diagrams
CG moves phase-boundaries ...
Thoughts on representability

1. Effective potentials are not Hamiltonian potentials
   a) Beware of using equations that treat them as such
      - E.g. Corrections to virial equation from density dependence – mostly wrong
      - But better thermodynamic equations might be derivable from coarse-graining scheme
   b) To map to something analogical to a Hamiltonian, care must be taken with ensembles
      (canonical ensemble may not be ideal)
   c) Be careful about one-body terms & missing degrees of freedom
   d) “Energetic” and “Entropic” character can change

2. Effective potentials are compromises – there is no free lunch
   a) It’s probably better not to fit too tightly to just one property (e.g. structure), but rather
      multiple properties at once.
   b) STOP & THINK: no substitute for physical insight, e.g. symmetries etc…
      a) Effective emergent models
      b) Nearby phase boundaries
   c) “Automated” coarse-graining techniques and black box methods should be treated with suspicion.

1. Representability and transferability are probably related. (State dependence)

Further thoughts on representability?

- “Emergent models” v.s. models for fine-graining
  - Different approaches?

- Fitting to thermodynamic quantities?

- Testing representability issues

There is no such thing as a free lunch! but cost can be brought down by physical insight
No free lunch theorems

• Effective-potentials for depletion systems

• Representability and potentials
  • A) Two and three-body potentials
  • B) Isotropic model for water

• Dynamics?
More difficulties with dynamics ...

Diffusion coefficient of assemblies ……

Telescoping down?

Flattened energy land-scapes

D ~ R^{-1}

D ~ R_g^{-1}

Brownian Dynamics D ~ N^{-1}

Coarse-graining: Telescoping time-scales

All you need is time-scale separation

developing down

Colloid of diameter \( \mu \)m

\( \tau_{col} \) - fluid collision
\( \tau_F \) - fluid de-correlation
\( \tau_{FP} \) - Fokker-Planck

\( t_{cs} \) - sonic = \( R/c_s \)

\( \tau_B \) - Brownian (Langevin) = \( M/\xi \)
\( \tau_v \) - kinematic = \( R^2/\nu \)

\( \tau_D \) - Colloid diffusion = \( R^2/D_{col} \)

\( \tau_B \) has no independent physical significance for colloids: beware of simple Langevin equation
How do I map to physical time/length scales?

1. mapping to diffusion time
2. mapping to kinematic time
3. mapping potentials

$10^{-3}$ -- $\tau_{FP}$-Fokker-Planck
$10^{-2}$ -- $t_{cs}$-sonic=$R/c_s$
$10^{-1}$ -- $\tau_\nu$-kinematic=$R^2/\nu$
$10^0$ -- $\tau_D$-Colloid diffusion= $R^2/D_{col}$
Coarse-graining changes energy landscapes

Extremely unlikely that all barriers are lowered to give the same time-scales – so multiple time-scales mappings in one simulation …

Figure from C. Peter