Stochastic Quantum Chemistry: Towards fully-quantum electronic structure and dynamics of very large molecules and clusters

DANNY NEUHAUSER (UCLA) (NSF, DOE; PNNL)

Stochastic: traditional quantum chemistry (DFT, MP2, RPA, GW, TDDFT, BSE, GF2) for very large systems

ROI BAER (Hebrew U.)

ERAN RABANI (Berkeley)

Recently: DOMINIKA ZGID (Michigan)
Stochastic Quantum Chemistry:
Towards fully-quantum electronic structure and dynamics of very large molecules and clusters

1992:
- $50,000 per Gigaflop
- QC: 100-150 electrons

2014:
- GPU: Gigaflop ~$5 or less
- Bio. Sim: ~micro/mili/more sec.
- Quantum Chemistry: ~1000-2000 electrons (DFT) (less for correlation)

New paradigm clearly needed for Quantum Chemistry.
Stochastic QM both: Computationally efficient and physically rich
Mathematical guidelines:

- **Stochastic compression:** \( f(r, t) = \eta(r, t)\langle \eta(t) | f(t) \rangle \eta \) random noise; \( \eta_{\text{w.n.}}(r) = \pm 1 \)

- **Multidimensional: stochastic compression:**
  \[ F(r, r') \rightarrow \eta(r) \langle \eta | F | \zeta \rangle \zeta(r') \]

- Replace **sums over orbitals by random averages:**
  \[ \int \sum_{ijkl} \ldots \phi_i(r) \ldots \phi_i(r') \ldots dr dr' = \int \{ \ldots \eta(r) \ldots \eta(r') \ldots \} dr dr' \]

- If \( \phi_i \) is **occupied** (energy below \( \mu \)) \( \eta \) should be **filtered:**
  \[ \eta = \theta(\mu - h)\eta_{\text{w.n.}} \]

- **Energy denominator \( \rightarrow \) correlation function**
  \[ \frac{\phi_k(r')}{E_i - E_j} \phi_j(r'') = -i \int_0^\infty \phi_k^*(r', t)\phi_j(r'', t) dt = -i \int_0^\infty \zeta^*(r', t)\eta(r'', t) dt \]
**OVERVIEW**

- DFT first
- Correlation next:
  - s-MP2
  - s-GW
  - s-GW

**HISTORY:**

- **Difference from QMC:** no moving walkers.

- Historically: DFT closest to Sankey’s approach in the 90’s – but realization now is “embedded fragment” and look for forces, not total energy

- Almlof had initial ideas for MP2 correlations.

**Relation to machine learning:**

- Usually looking for high accuracy.

- But want to sample many molecules.

- So instead eventual goal: do multiple low-resolution study (millions of high-noise calculations) – let the machine find the regions where less noise is desired.
Existing DFT codes (Gaussian, VASP, more):

- **INPUT:** nuclear coordinates

  **Orbitals cost:** $N^3$
  
  (Benzene $\rightarrow$ Fullerene, ~ 1000 times more expensive)

  **Output:** Energies, Forces

  **Output:** Orbitals

  1. Intuition
  2. Input for higher methods (MP2, RPA, GW, CC)
Traditional method scale steeply.

Traditionally:

\[ \rho(r) = 2\{ |\psi_1(r)|^2 + |\psi_2(r)|^2 + \cdots + |\psi_{HOMO}(r)|^2 \} \]

MO: expensive – due to Pauli principle:

\[ \langle \psi_i | \psi_j \rangle = 0 \quad \text{if} \quad i \neq j. \]

Getting all occupied orbitals: scales cubically

10 times bigger system \( \rightarrow \) 1000 times more effort.

Alternatives:

1) Avoid orbitals, concentrate on “Density Matrices” – OK for small Gaussian basis sets, cannot describe collective effects requiring orbitals away from HOMO/LUMO.
Instead of extracting and summing MO, use **Stochastic Averaging**: Practiced by large molecules all around you!

\[
\text{Reaction Rate} = \sum_{\text{level } j} P(\text{reactant level } j) \times \text{Rate(level } j, \text{reactant } \rightarrow \text{ product)}
\]

Instead of summing over millions (or more) of states, ~100 collisions give correct reactions rate
Stochastic averaging for quantum chemistry:

\[ \rho(r) = \frac{1}{2} \left( |\psi_1(r)|^2 + |\psi_2(r)|^2 + \cdots + |\psi_{\text{HOMO}}(r)|^2 \right) \]

\[ \rho(r) = \left\{ |\psi(r)|^2 \right\} \]
Stochastic DFT

Average over:
- Choose WHITE NOISE (random) ORBITAL.
- FILTER IT TO REMOVE HIGH ENERGIES (NON-OCCUPIED)
- Add contribution from filtered-white noise

\[
\rho(r) = \left\{|\psi_{\text{filtered}}(r)|^2\right\}
\]

INPUT: nuclear coordinates

Output: STOCHASTIC Orbitals (or "Density amplitudes")
1. Yield HOMO, LUMO easily;
2. Input for higher accuracy methods (MP2, RPA, GW, CC)

Output: Energies, Forces

Baer, DN, Rabani, PRL 111, 106402 (2013)
Embedded-Fragment  (JCP,2014)

• Originally:
  \[ \rho(r) = \left| \psi_{\text{filtered}}(r) \right|^2 \]

• Instead: use putty:
  \[ \rho(r) = \rho_0(r) + \rho(r) - \rho_0(r) \]
  
  \[ \rho(r) = \rho_0(r) + \left\{ \left| \psi_{\text{filtered}}(r) \right|^2 - \left| \psi_{0,\text{filtered}}(r) \right|^2 \right\} \]
Water clusters:

- At present: single-molecule fragments
- Better: bigger fragments
Embedded fragment: works for covalently held large systems

<table>
<thead>
<tr>
<th>System</th>
<th>Iteration</th>
<th>$E/N_e$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Trad.</td>
</tr>
<tr>
<td>$Si_{16}H_{16}$</td>
<td>30</td>
<td>$-24.025$</td>
</tr>
<tr>
<td>$Si_{35}H_{36}$</td>
<td>30</td>
<td>$-23.924$</td>
</tr>
<tr>
<td>$Si_{87}H_{76}$</td>
<td>30</td>
<td>$-23.552$</td>
</tr>
<tr>
<td>$Si_{353}H_{196}$</td>
<td>30</td>
<td>---</td>
</tr>
<tr>
<td>$Si_{705}H_{300}$</td>
<td>30</td>
<td>---</td>
</tr>
</tbody>
</table>

Si$_{705}$H$_{300}$
(80 iterations)

<table>
<thead>
<tr>
<th>$\delta E_{gap}$ (eV)</th>
<th>$\delta E_{tot}$ (eV) $/N$</th>
</tr>
</thead>
</table>
| No fragments           | No gap                      | 0.025
| Tiny fragments (~5 Si) | 0.10                        | 0.002
| Small fragments (~20 Si)| 0.03                        | 0.001

Forces (eV/bohr) $Si_{35}H_{36}$

Effort scales gently since required # of stochastic orbitals decreases with system size.
Large systems:

Timing:

~ 2 hours on a single GPU * 40 GPUs for 3000 e’s;

Estimated ~30 hours *40 GPUs when 100,000 e’s

(Spectrum, error)

200 hours*40 GPUs for 1,000,000 electrons.
Interim Conclusions: Stochastic DFT

Scaling: **linear or below**

*Trivially parallelize* (each “stochastic orbital” essentially not influenced by others. **No orthogonalization!**)

*Insulators, Semiconductors; metals too* (≈20 times more effort for metals at room temperature)

---

**WHAT ABOUT:**

1) **Stochastic Fluctuations of E(total)**
   
   Non issue here, since force directly from density.

2) **Exchange (hybrid potentials)**
   
   E(Exchange) easily calculated.
   
   Underlying hybrid exchange potentials (B3LYP): **officially not suitable** (need fast-acting potentials).
   
   We’re Working on stochastic versions of:
   
   • “Optimized Effective Potentials”.
   
   • LDA+U
   
   • **Stochastic Exchange.**
Side note:

Stochastic DFT: Opens new ways to think about (and apply) classical-quantum embedding:

All regions really quantum; The more classical a region is, the less often fluctuations are applied on it.
Correlation: there’s life beyond (Stochastic) DFT!

Correlation methods (better e-e interaction):

- Absorption spectra, HOMO and LUMO, van-der-Waals. *(More important for bigger systems!*

- Usually **scale terribly** \((N^4 - N^6)\),
  but in Stochastic QM they all **scale ~linearly**

**Stochastic MP2/MP3/MPn**
**Stochastic RPA**
**Stochastic GW**
We did not find yet Stochastic Coupled Clusters!
*Side note: both basis sets and plane-waves!**
Warm-up: MP2: simplest perturbation theory

\[ E_{vdw} = \frac{1}{2} \sum_{\text{ab: unocc.}} \frac{l_{abrs}(2l_{abrs} - l_{asbr})}{\epsilon_a + \epsilon_b - \epsilon_r + \epsilon_s} \approx \int V(t)V(0)dt \]

\[ V(t) = \int \zeta(r, t)\alpha(r, t) \frac{1}{|r - r'|} \eta(r', t)\beta(r', t)drdr' \]

\[ \zeta = \theta(\mu - h)\zeta_{w.n.} \]

\[ \alpha = (1 - \theta(\mu - h))\alpha_{w.n.} \]
Physically:

\[ E_{MP2} \approx - \sum_{ij \text{ occ.}} \left| Interaction(\text{Exciton}(j, a), \text{Exciton}(i, b)) \right|^2 \]

\[ \varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b \]

**Traditionally:** *EXPENSIVE!* \( N^4 - N^5 \) Often gives all accuracy needed.

**Stochastic:** Instead of summing: roll dice!

\[ E(\text{MP2}) = - \left[ V(\text{exc. (1)}, \text{exc. (2)}, t = 0) \times \int V(\text{exc. (1)}, \text{exc. (2)}, t) \, dt \right] \]

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Basis Set Size</th>
<th>( E_{MP2}/N_e ) (eV)</th>
<th>SE</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzene</td>
<td>30</td>
<td>0.521</td>
<td>0.022</td>
<td>5.6</td>
</tr>
<tr>
<td>Naphthalene</td>
<td>48</td>
<td>0.558</td>
<td>0.017</td>
<td>4.5</td>
</tr>
<tr>
<td>Fullerene</td>
<td>240</td>
<td>0.678</td>
<td>0.012</td>
<td>2.7</td>
</tr>
<tr>
<td>Fullerene x 36</td>
<td>3 x 3 x 4</td>
<td>0.678</td>
<td>0.009</td>
<td>2.2</td>
</tr>
<tr>
<td>Fullerene x 64</td>
<td>4 x 4 x 4</td>
<td>0.521</td>
<td>0.008</td>
<td>2.0</td>
</tr>
<tr>
<td>Fullerene x 64</td>
<td>4 x 4 x 4</td>
<td>0.521</td>
<td>0.008</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Accurate structure/dynamics: need:

Susceptibility, $\chi(r_1, r_2, t)$:
Disturb at $r_1$ at $t=0$,
density change at $r_2$ at $t$.

Related to Effective Interaction,
$W(r_1, r_2, t)$

1. Van-der Waals: “RPA”:
Interaction of pairs through effective interaction

2. Absorption:
   frequency-dependent response, $\chi(r_1, r_2, \omega)$:

3. GW: correct ionization/affinity.
   $G$: Green’s function:
   $W$: effective interaction.

   \[ HOMO-LUMO \text{ correction from: } \Sigma(t) \sim G(t)W(t) \]

LeChateliers’ principle:
Other e⁻ polarize, mask perturbation
Stochastic susceptibility

• Need: propagate disturbance in t: TDDFT.

• Stochastic TDDFT: enough to propagate few vectors (fewer for larger systems!):
  • $\chi(\omega)$: absorption: for large systems!
  • Crucial ingredient in RPA and GW below.

$\langle \chi(t) \rangle$

Susceptibility and absorption: from 64 random amplitudes (vs. up to 1560 occupied orbitals!)
Highlight: quasiparticle energies with stochastic GW

Why GW? DFT fails to give proper ionization/affinity (HOMO/LUM). GW corrects!

GW

GW: apply self-energy correction to the quasiparticle energy.

$$\delta E_{\text{HOMO}} = \langle \phi_{\text{HOMO}} | \Sigma | \phi_{\text{HOMO}} \rangle$$

$$\Sigma(r_1, r_2, t) = G(r_1, r_2, t) W(r_1, r_2, t)$$

Pictorially:

Propagate electron from one point to another, while having the tamed interaction.
Stochastic GW:

\[ \eta_{occ}: \text{stochastic occupied orbital} \]

\[ \eta_{occ}(t) \]

\[ G_0(r_2, r_1, t) \simeq \{ \eta_{occ}(r_2, t) \eta_{occ}(r_1) \} \]

Putting it all together:

\[ \zeta_{filtered}(r_2) \]

\[ \eta_{occ}(r_2, t) \]

\[ \eta_{occ}(r_1) \]

\[ \zeta_{filtered}(r_1) \]

\[ W(r_2, r_1, t) \]

Stochastic TDDFT
(Source: \( \eta^* \zeta_{filtered} \))
Side note: stochastic compression in GW

TDDFT: gives \((W_{\text{casual}} g)(r, t) = \int W^{*}_{\text{casual}}(r, r', t) g(r', t)dr'\)

\[ W_{TO}(r, r', \omega) = \begin{cases} W^*_{\text{casual}}(r, r', \omega) & \omega \leq 0 \\ W_{\text{casual}}(r, r', \omega) & \omega > 0 \end{cases} \]

Stochastic compression:
\[(W_{\text{casual}} g)(r, t) \rightarrow \eta(r)c(t)\]

\[ c(t) \equiv \langle \eta(r) | W(t) g \rangle \rightarrow c(\omega) \rightarrow c_{TO}(\omega) = \begin{cases} c^*(\omega) & \omega \leq 0 \\ c(\omega) & \omega > 0 \end{cases} \rightarrow c_{TO}(t) \]
SILICON CLUSTERS

Present

Prev. Work

Stochastic GW Results

PCBM (fullerene with add-ons)

GW(TDH): \( \text{EHOMO} = 7.1 \pm 0.1 \text{ eV} \)

Experiment: \( \text{EHOMO} = 7.17 \text{ eV} \),

GW(TDDFT): \( \text{ELUMO} = 2.5 \pm 0.1 \text{ eV} \)

Experiment: \( \text{EHOMO} = 2.63 \text{ eV} \),

FULLERENES

In progress: GW for large organic clusters (<20,000 CPU core hours for (PCBM)_{18})

Timing: 5000 cpu-core hours for cluster with 3120 electrons; linear scaling.

<table>
<thead>
<tr>
<th>System</th>
<th>( N_e )</th>
<th>( N_g )</th>
<th>( I_{DFT} )</th>
<th>( N_e )</th>
<th>( \beta_{GW}^{-1} (E_h) )</th>
<th>( E_{qp}^{QP} (\text{eV}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si_{36}H_{36}</td>
<td>176</td>
<td>60^3</td>
<td>3000</td>
<td>16</td>
<td>0.020</td>
<td>6.19 \pm 0.13</td>
</tr>
<tr>
<td>Si_{47}H_{76}</td>
<td>424</td>
<td>64^3</td>
<td>1600</td>
<td>16</td>
<td>0.012</td>
<td>4.82 \pm 0.14</td>
</tr>
<tr>
<td>Si_{147}H_{100}</td>
<td>688</td>
<td>70^3</td>
<td>800</td>
<td>16</td>
<td>0.010</td>
<td>4.10 \pm 0.13</td>
</tr>
<tr>
<td>Si_{323}H_{196}</td>
<td>1088</td>
<td>90^3</td>
<td>400</td>
<td>16</td>
<td>0.008</td>
<td>2.99 \pm 0.13</td>
</tr>
<tr>
<td>Si_{708}H_{300}</td>
<td>3120</td>
<td>108^3</td>
<td>200</td>
<td>16</td>
<td>0.007</td>
<td>2.24 \pm 0.15</td>
</tr>
</tbody>
</table>
Conclusions

• **Paradigm shift:**
  • away from exact orbitals, density matrices, to:
  • **stochastic density amplitudes.**
• **Linearizes** effort
• **Almost all traditional Quantum Chemistry** recasted.
• Makes sense **only once system is large** (>5000 electrons for DFT, >100 electrons for correlation), **but then quickly becomes only viable approach**
• Handles forces and gradients by **adiabatic transform**
• **DFT of systems with million electrons – in sight now!**
• **GW for accurate IP/EA (HOMO/LUMO)** – required accuracy (~0.05-0.1 eV) easily reached for thousands of e’s and more.
• Beyond Quantum Chemistry, extends to other fields: electron dynamics, molecular vibrations.