Stochastic Quantum Chemistry:

Towards fully-quantum electronic structure and dynamics of very large molecules and clusters

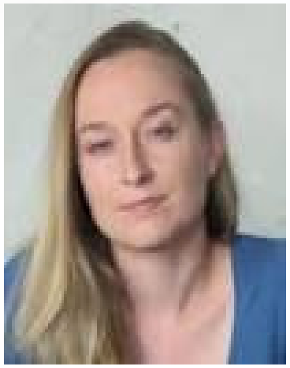
DANNY NEUHAUSER (UCLA) (**NSF, DOE; PNNL**)

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
- Bio. simulations: pico/nano sec.
- 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.

<u>Stochastic QM both: Computationally efficient and physically rich</u>

Mathematical guidelines:

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

$$F(\boldsymbol{r},\boldsymbol{r}') \to \eta(\boldsymbol{r})\langle\eta|F|\zeta\rangle\zeta(\boldsymbol{r}')$$

- Replace **sums over orbitals by random averages**:

$$\int\sum_{ijkl} \dots \phi_i(r) \dots . \phi_i(r') \dots . dr dr' = \int\{\dots . \eta(r) \dots \eta(r') \dots\} d\boldsymbol{r} d\boldsymbol{r}'$$

- If ϕ_i is **occupied** (energy below μ) η <u>**should be filtered:**</u>

$$\eta = \theta(\mu - h)\eta_{w.n.}$$

- **Energy denominator → 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

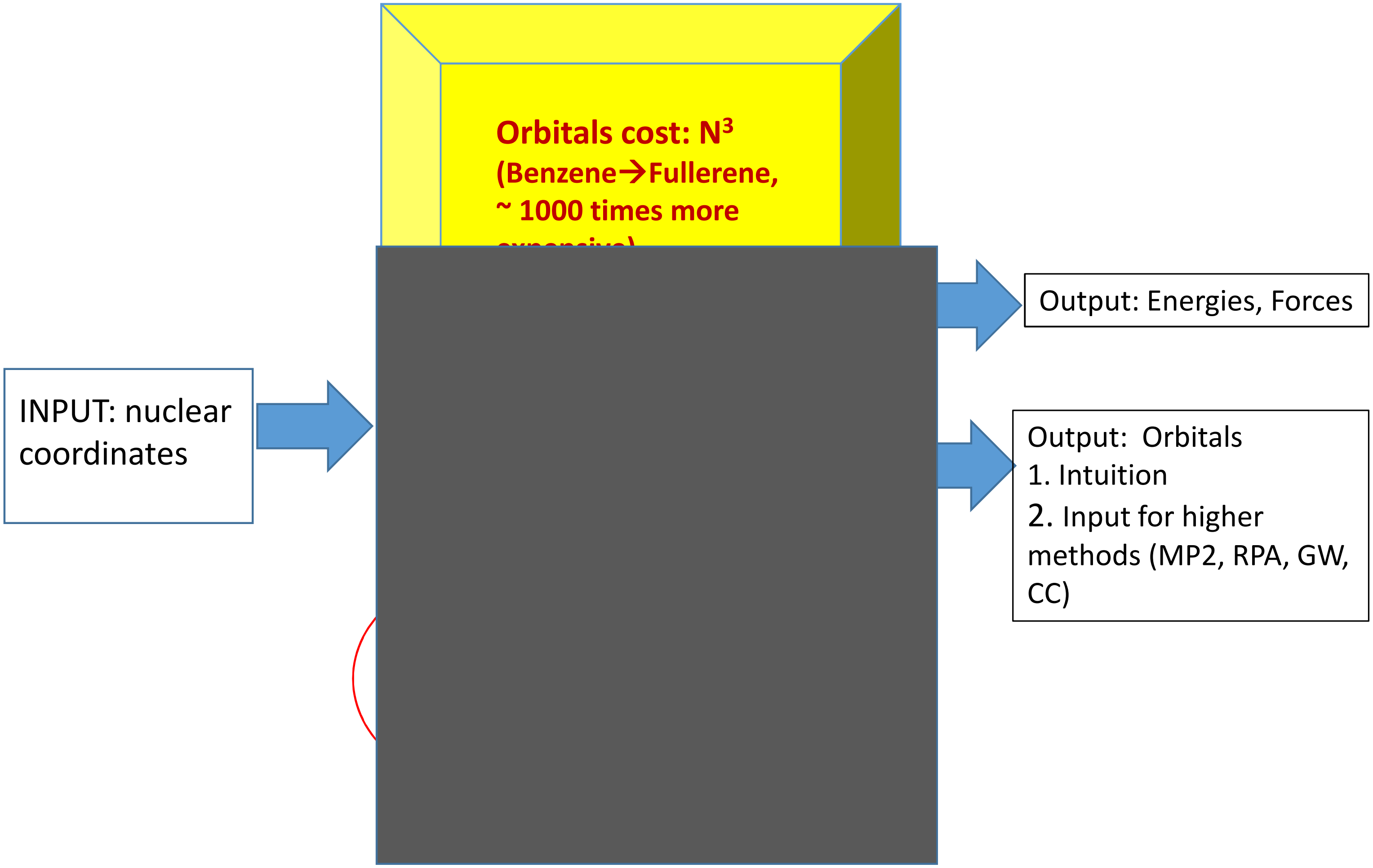
HISTORY:

- 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):



Orbitals cost: N^3
(Benzene→Fullerene,
~ 1000 times more
expensive)

INPUT: nuclear coordinates

Output: Energies, Forces

Output: Orbitals
1. Intuition
2. Input for higher methods (MP2, RPA, GW, CC)

Traditional method scale steeply.

Traditionallly:

$$\rho(\boldsymbol{r}) = \boldsymbol{2}\left\{|\boldsymbol{\psi_1}(\boldsymbol{r})|^{\boldsymbol{2}} + |\boldsymbol{\psi_2}(\boldsymbol{r})|^{\boldsymbol{2}} + \cdots + |\boldsymbol{\psi_{HOMO}}(\boldsymbol{r})|^{\boldsymbol{2}}\right\}$$

MO: expensive – due to Pauli principle:

$$\langle\psi_i|\psi_j\rangle = 0 \quad if \quad i \neq j.$$

Getting all occupied orbitals: scales cubically

$$10\ times\ bigger\ system\ \rightarrow 1000\ times\ more\ effort.$$

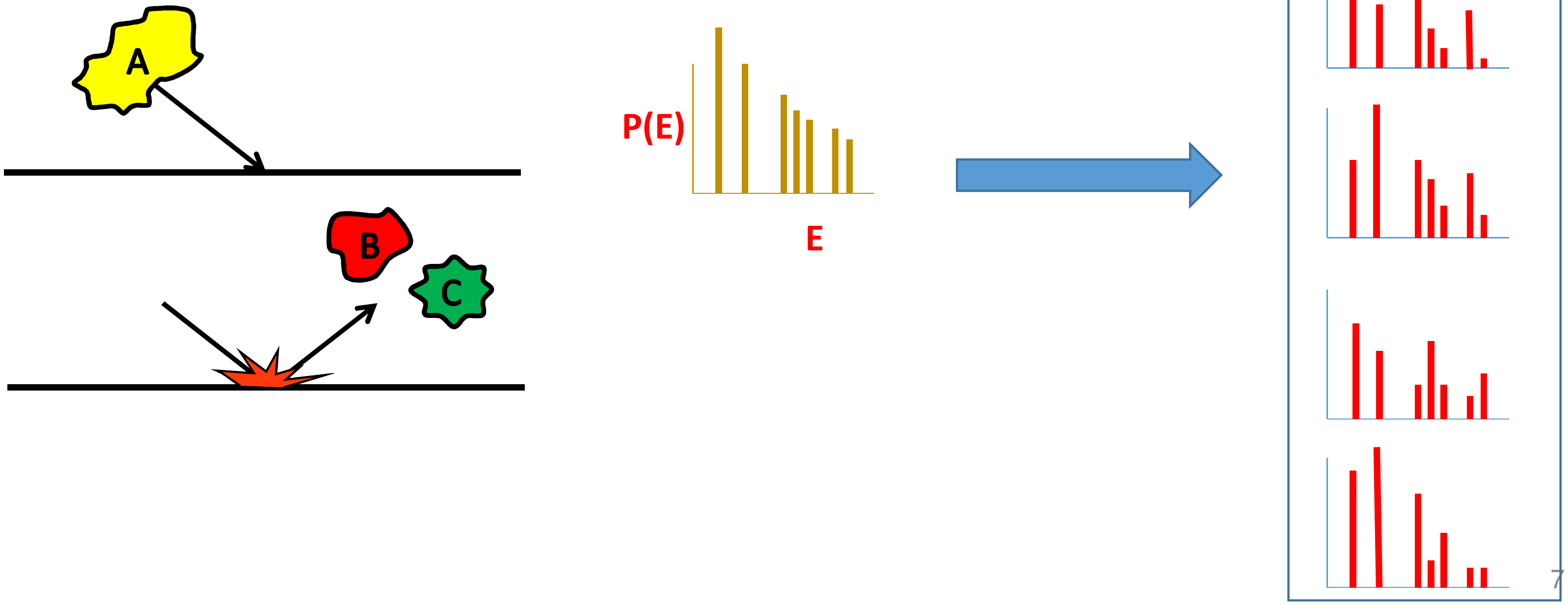
Alternatives:

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!

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

Instead of summing over millions (or more) of states,
~100 collisions give correct reactions rate



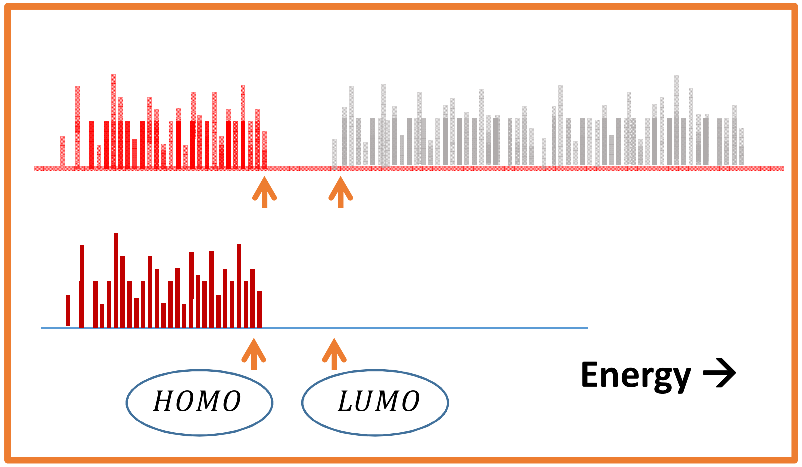
Stochastic averaging for quantum chemistry:

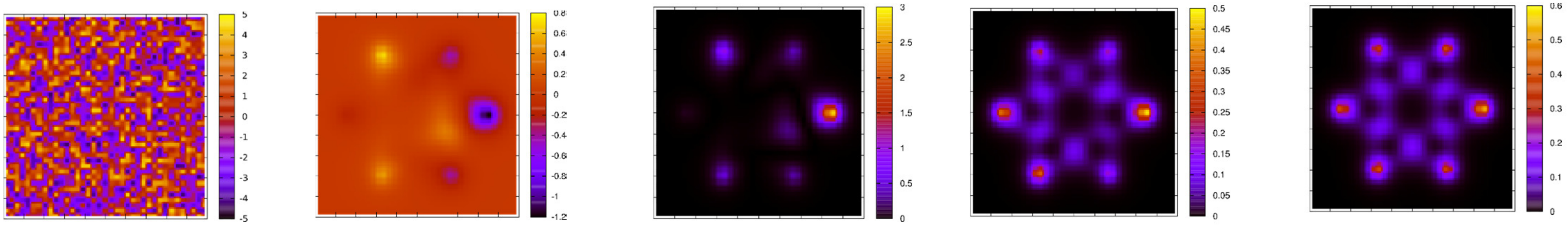
$$\xcancel{\rho(\boldsymbol{r}) = 2\{|\psi_1(\boldsymbol{r})|^2 + |\psi_2(\boldsymbol{r})|^2 + \cdots + |\psi_{HOMO}(\boldsymbol{r})|^2\}}$$

$$\rho(\boldsymbol{r}) = \{|\psi(\boldsymbol{r})|^2\}$$

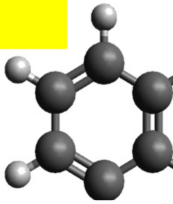
Stochastic Occupied Orbital

- Contains: ALL occupied MOs
- Each occupied MO has same weight (1) <u>in average.</u>





white noise → stochastic occ. orbital → density (1 stoch.orbital) → average 10 → average 100



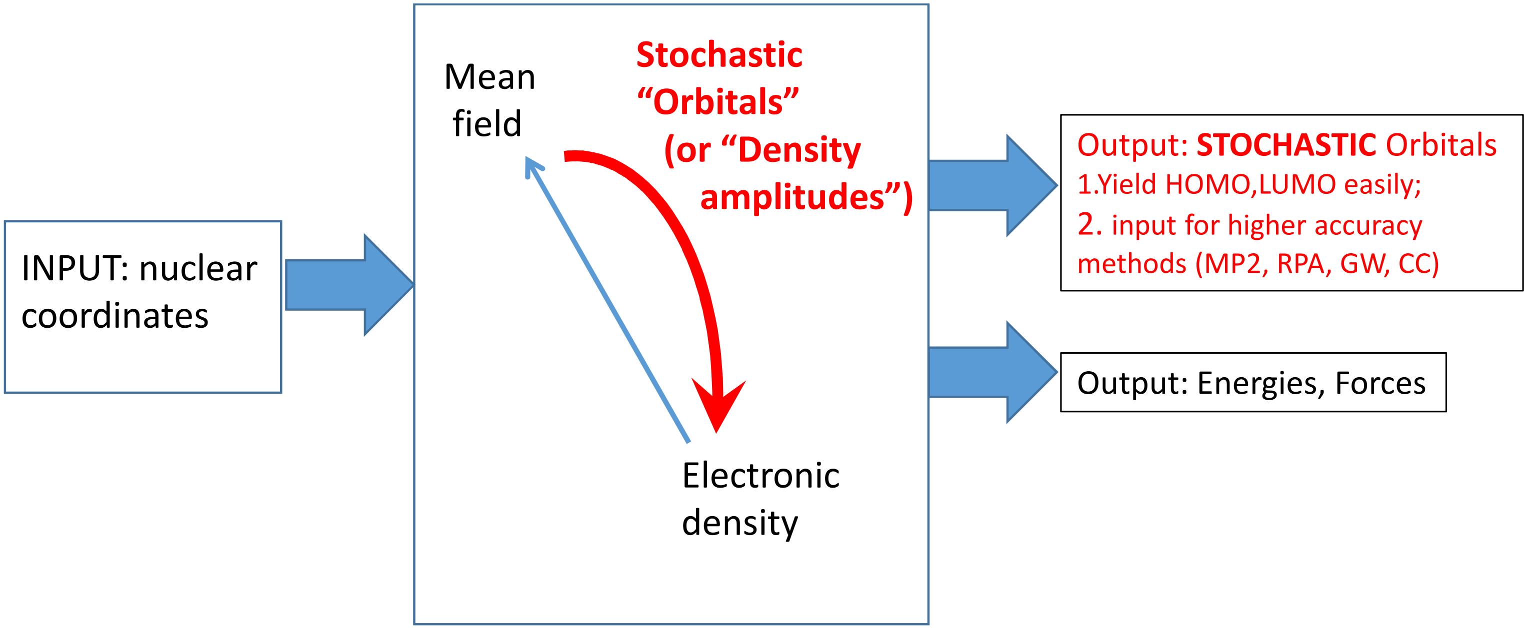
Stochastic DFT

Baer, DN, Rabani, PRL 111, 106402 (2013)

Average over:

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

$$\rho(\boldsymbol{r}) = \left\{ \left|\psi_{filtered}(\boldsymbol{r})\right|^2 \right\}$$



INPUT: nuclear coordinates

Mean field

Stochastic “Orbitals” (or “Density amplitudes”)

Electronic density

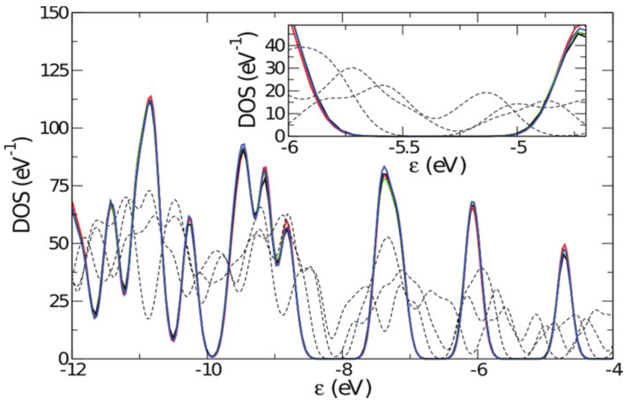
Output: **STOCHASTIC** Orbitals
1.Yield HOMO,LUMO easily;
2. input for higher accuracy methods (MP2, RPA, GW, CC)

Output: Energies, Forces

Embedded-Fragment (JCP,2014)

- Originally: $\rho(r) = \left\{\left|\boldsymbol{\psi}_{filtered}(\boldsymbol{r})\right|^2\right\}$
- Instead: use **putty:** $\rho(\boldsymbol{r}) = \rho_0(\boldsymbol{r}) + \rho(\boldsymbol{r}) - \rho_0(\boldsymbol{r})$

$$\rho(\boldsymbol{r}) = \boldsymbol{\rho_0}(\boldsymbol{r}) + \left\{\left|\boldsymbol{\psi}_{filtered}(\boldsymbol{r})\right|^{\boldsymbol{2}} - \left|\boldsymbol{\psi}_{\boldsymbol{0},filtered}(\boldsymbol{r})\right|^{\boldsymbol{2}}\right\}$$



Water clusters:

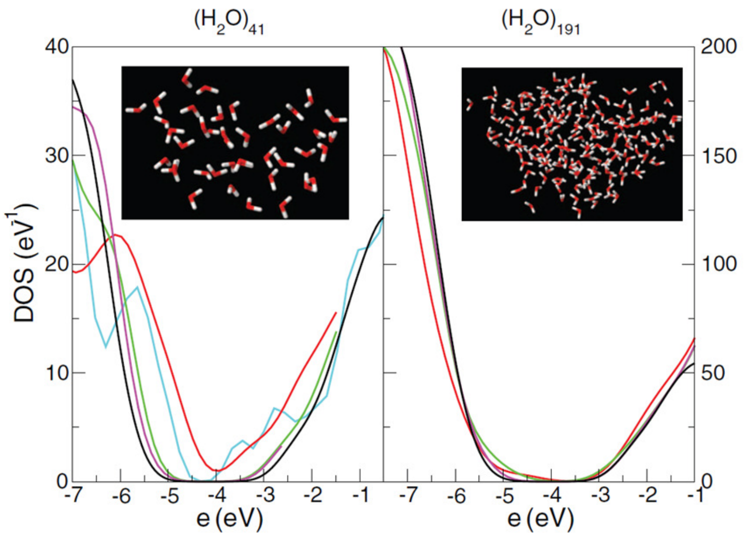
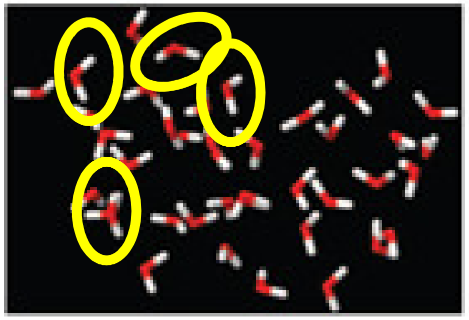
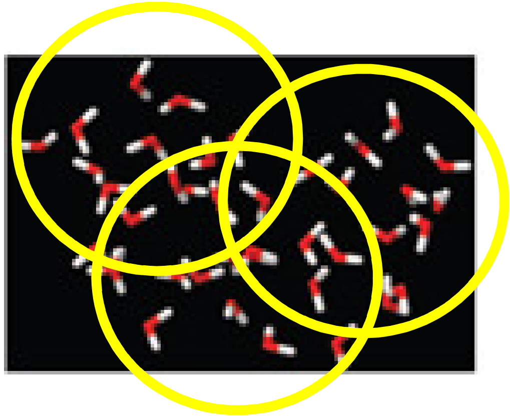


FIG. 2. The density of states near the highest occupied and lowest unoccupied KS eigenvalues of $(H_2O)_{41}$ (left panel) and $(H_2O)_{191}$ (right panel) using sDFT with $I = 320$ (cyan) and efsDFT with $I = 80$ (red), 160 (green), and 320 (magenta) stochastic orbitals. The solid black curve represents the deterministic DFT calculation.

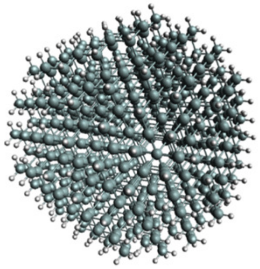
- **At present: single-molecule fragments**

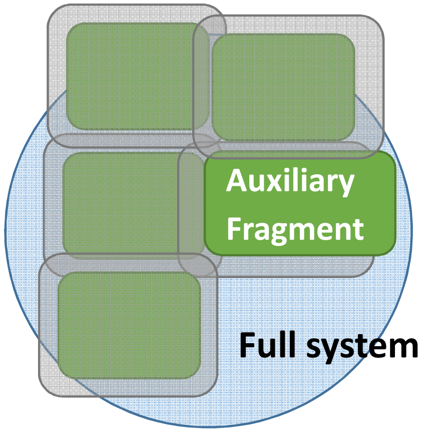


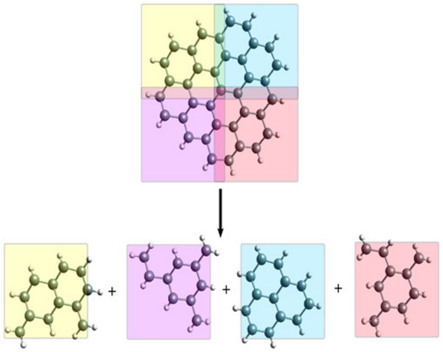
- **Better : bigger fragments**



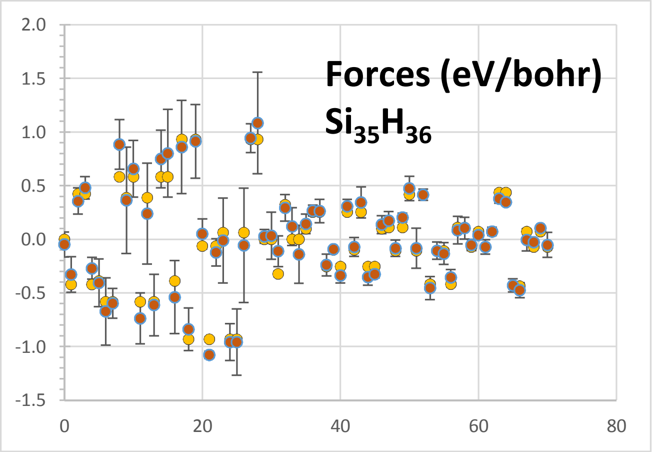
Embedded fragment: works for covalently held large systems



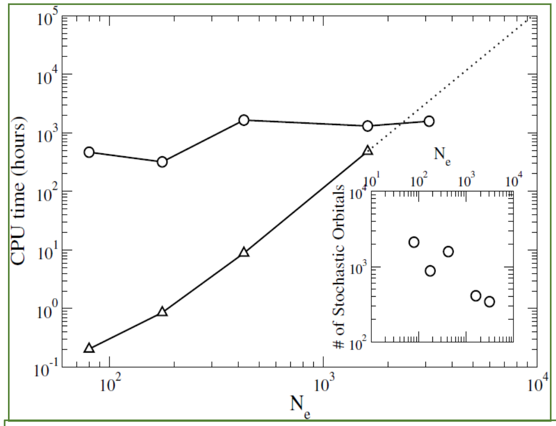




$Si_{705}H_{300}$ (80 iterations)	δE_{gap}(eV)	$\frac{\delta E_{tot}}{N}$(eV)
No fragments	No gap	0.025
Tiny fragments (~5 Si)	0.10	0.002
Small fragments (~20 Si)	0.03	0.001



System	*Iteratio*	E/N_e (eV)	
		Trad.	Stoch.
$Si_{16}H_{16}$	30	-24.023	-24.025
$Si_{35}H_{36}$	30	-23.921	-23.924
$Si_{87}H_{76}$	30	-23.550	-23.552
$Si_{353}H_{196}$	30	---	-23.847
$Si_{705}H_{300}$	30	---	-23.4489



Effort scales gently since required # of stochastic orbitals decreases with system size

Large systems:

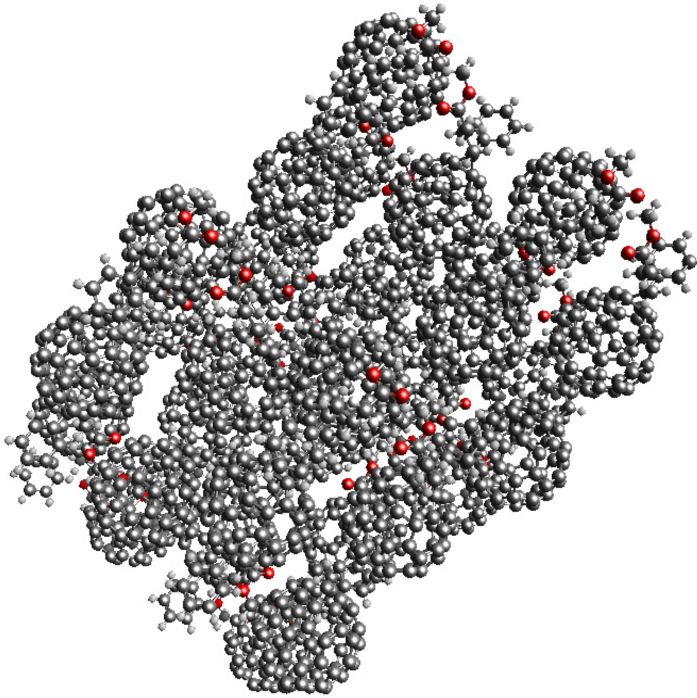
Timing:

~ **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 1.000,000electrons.



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)

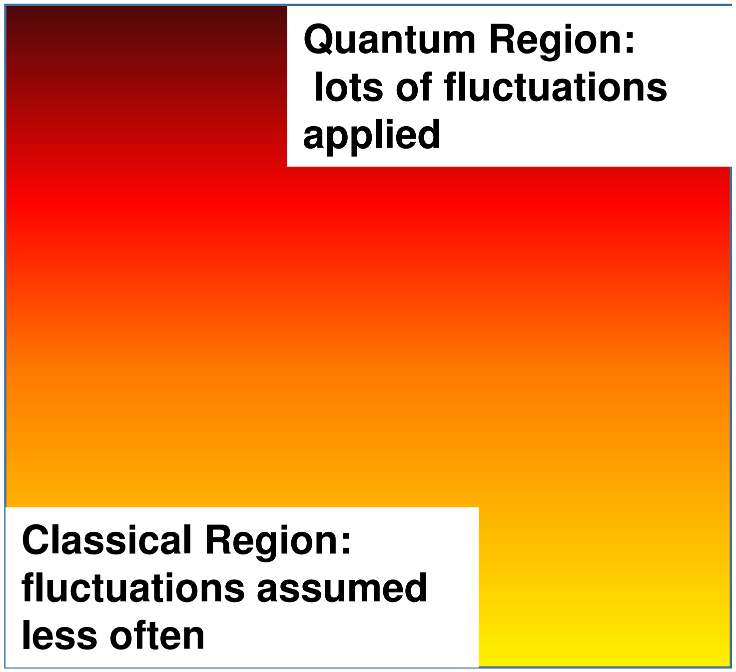
E(Exchange) easily calculated.
Underlying hybrid exchange potentials (B3LYP): officially not suitable (need fast-acting potentials).
We're Working on stochastic versions of:

- **"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;
<u>The more classical a region is, the less often fluctuations are applied on it.</u>


Quantum Region:
lots of fluctuations applied

Classical Region:
fluctuations assumed less often

Correlation: there's life beyond (Stochastic) DFT!

Correlation methods (better e-e interaction):

- Absorption spectra, HOMO and LUMO, van-der-Waals. <u>(More important for bigger systems!</u>
- Usually **scale terribly** $(N^4 - N^6)$, but in <u>Stochastic QM they all **scale ~linearly**</u>

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

$$E_{vdw} = \frac{1}{2} \sum_{\substack{ab:unocc.\\ rs:occu.}} \frac{I_{abrs}(2I_{abrs} - I_{asbr})}{\epsilon_a + \epsilon_b - \epsilon_r + \epsilon_s} \simeq \int V(t)V(0)dt$$

$$V(t) = \int \zeta(\boldsymbol{r},t)\alpha(\boldsymbol{r},t)\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}\eta(\boldsymbol{r}',t)\beta(\boldsymbol{r}',t)\boldsymbol{dr}\boldsymbol{d}r'$$

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

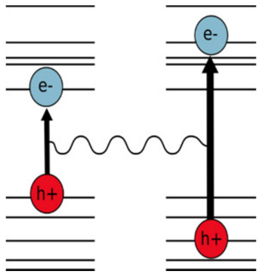
$$\alpha = \big(1 - \theta(\mu - h)\big)\alpha_{w.n.}$$

Physically:

$$E_{MP2} \simeq -\sum_{\substack{ij\ \mathrm{occ.} \\ ab\ \mathrm{unoccu.}}} \frac{\left|Interaction\left(\mathbf{Exciton}(\mathbf{j},\mathbf{a}),\mathbf{Exciton}(\mathbf{i},\mathbf{b})\right)\right|^2}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b}$$

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

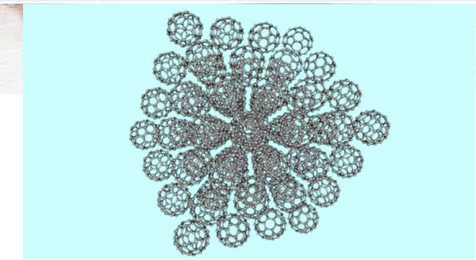
Stochastic: **Instead of summing: roll dice!**



$$\mathbf{E(MP2)} = -\left[\begin{array}{l}\mathbf{V(exc.(1),exc.(2),t=0)}\times \\ \displaystyle\int \mathbf{V(exc.(1),exc.(2),t)\ dt}\end{array}\right]$$

$$E_{MP2} = \sum_{ijab} \frac{(2\langle ab|ij\rangle - \langle ab|ji\rangle)\langle ij|ab\rangle}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b}$$

Molecule	Basis Set Size	$E_{MP2}/N_e\ (eV)$		SE	$E\ per$ $teration$
		Tradit.	**Stoc.**		
Benzene	**30**	0.521	**0.522**	0.022	5.6
Naphthalene	**48**	0.558	**0.557**	0.017	4.5
Fullerene	**240**		**0.678**	0.012	2.7
Fullerene x 36 $3\times3\times4$	**8640**		**0.704**	0.009	2.2
Fullerene x 64 $4\times4\times4$	**15360**		**0.696**	0.008	2.0

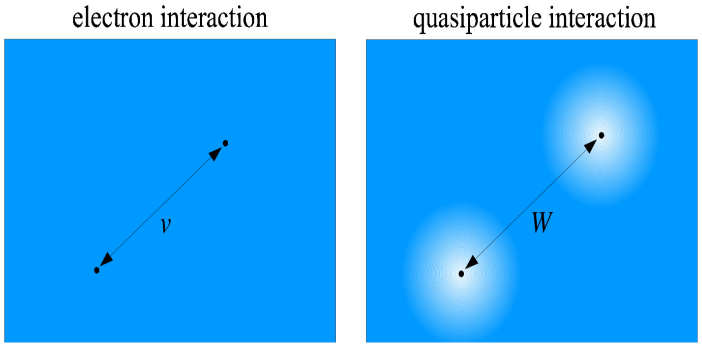


Accurate structure/dynamics: **need: Susceptibility, $\chi(\boldsymbol{r}_1, \boldsymbol{r}_2, t)$**:

Disturb at $\boldsymbol{r}_1$ at t=0,

density change at $\boldsymbol{r}_2$ at t.

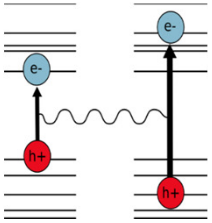
Related to Effective Interaction, $W(\boldsymbol{r}_1, \boldsymbol{r}_2, t)$



LeChateliers' principle:

Other e^- **polarize**, mask perturbation

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



2. Absorption:
frequency-dependent response, $\chi(\boldsymbol{r}_1, \boldsymbol{r}_2, \omega)$:

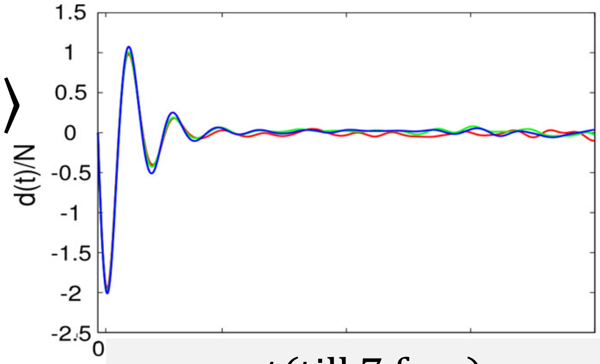
3. GW: correct ionization/affinity.
G: Green's function:
W: effective interaction.

HOMO-LUMO correction from: $\boldsymbol{\Sigma}(\mathbf{t}) \sim \mathbf{G}(\mathbf{t})\mathbf{W}(\mathbf{t})$

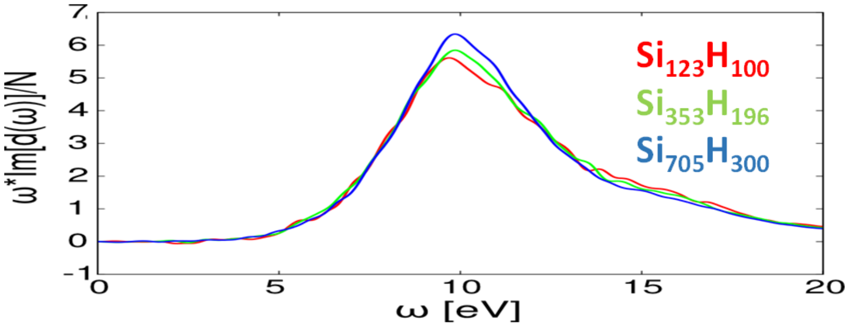
Stochastic susceptibility

- <u>Need: propagate disturbance in t: TDDFT.</u>
- **Stochastic TDDFT: enough to propagate few vectors (fewer for larger systems!):**
  - $\chi(\omega)$: <u>absorption: </u>for large systems!
  - Crucial ingredient in RPA and GW below.

$\langle\chi(t)\rangle$



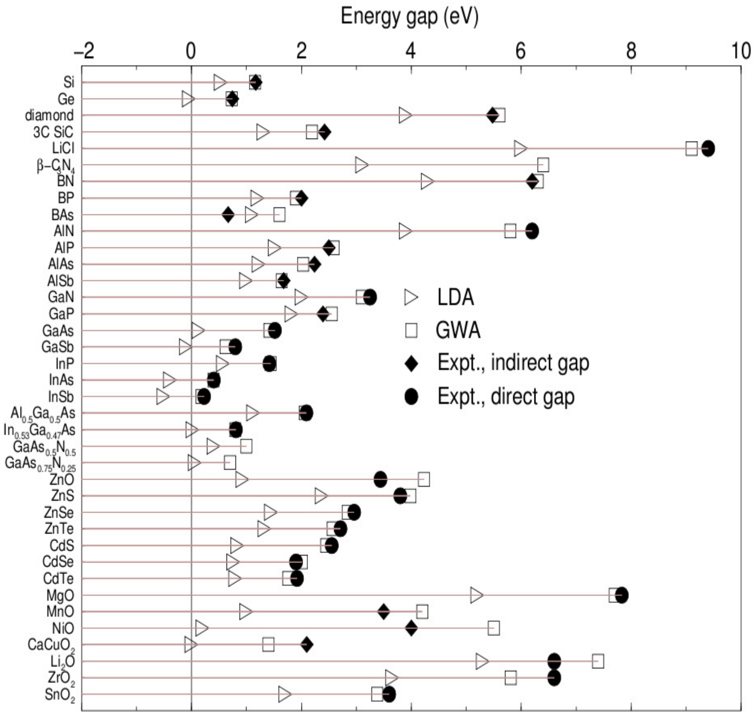
t (till 7 fsec)



Susceptibility and absorption:
from 64 random amplitudes
(vs. up to 1560 occupied orbitals!)

Highlight :quasiparticle energies with stochastic GW

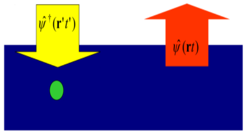
Why GW? DFT fails to give proper ionization/affinity (HOMO/LUM). GW corrects!

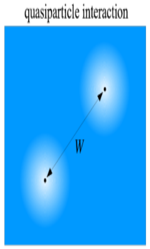


GW

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

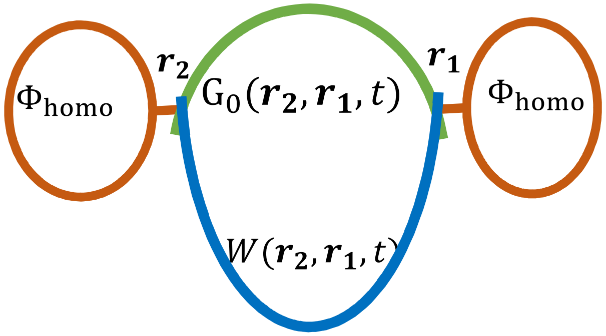
$$\delta E_{HOMO} = \langle \phi_{HOMO} | \Sigma | \phi_{HOMO} \rangle$$





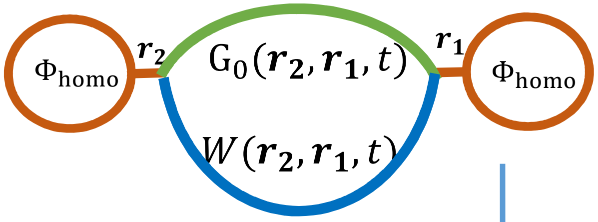
$$\Sigma(\boldsymbol{r}_1, \boldsymbol{r}_2, t) = G(\boldsymbol{r}_1, \boldsymbol{r}_2, t)\, W(\boldsymbol{r}_1, \boldsymbol{r}_2, t)$$

Pictorially:



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

Stochastic <u>G</u>W:



η_{occ}: stochastic occupied orbital

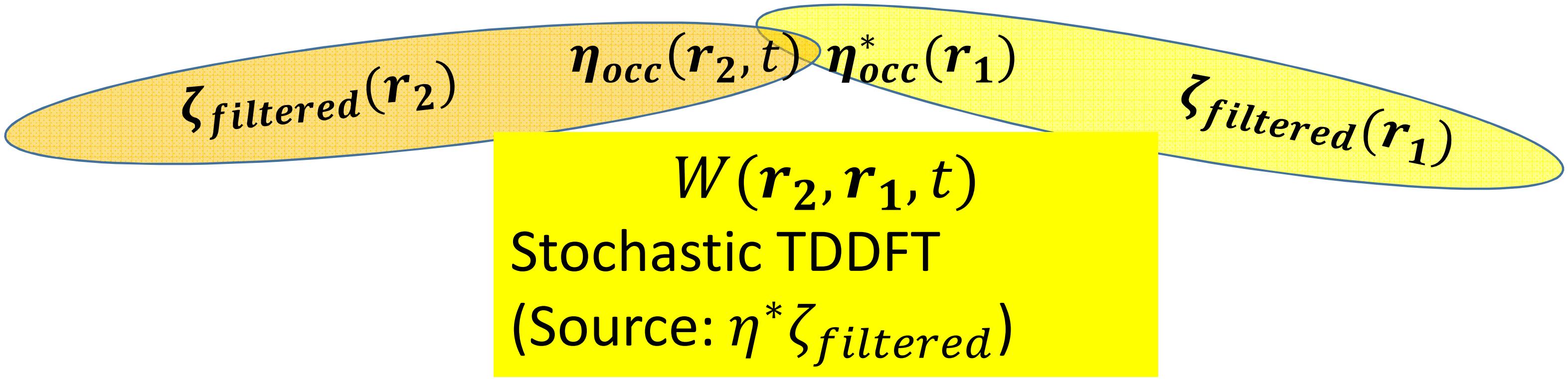
$\eta_{occ}(t)$

$$\mathrm{G}_0(\boldsymbol{r_2},\boldsymbol{r_1},t) \simeq \{\eta_{occ}(\boldsymbol{r_2},t)\eta_{occ}(\boldsymbol{r_1})\},$$

The “ears”: Φ_{homo}

$$\boldsymbol{\Phi_{HOMO} \sim filter \cdot \zeta_{random}}$$

Putting it all together:



$\boldsymbol{\zeta_{filtered}(r_2)}$ $\boldsymbol{\eta_{occ}(r_2,t)}$ $\boldsymbol{\eta^{*}_{occ}(r_1)}$ $\boldsymbol{\zeta_{filtered}(r_1)}$

$W(\boldsymbol{r_2},\boldsymbol{r_1},t)$

Stochastic TDDFT

(Source: $\eta^{*}\zeta_{filtered}$)

Side note: stochastic compression in GW

TDDFT: gives $(\boldsymbol{W_{casusal}g})(\boldsymbol{r},t) = \int \boldsymbol{W}_{\text{"}\boldsymbol{causal}\text{"}}(\boldsymbol{r},\boldsymbol{r}',t)g(\boldsymbol{r}',t)d\boldsymbol{r}'$

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

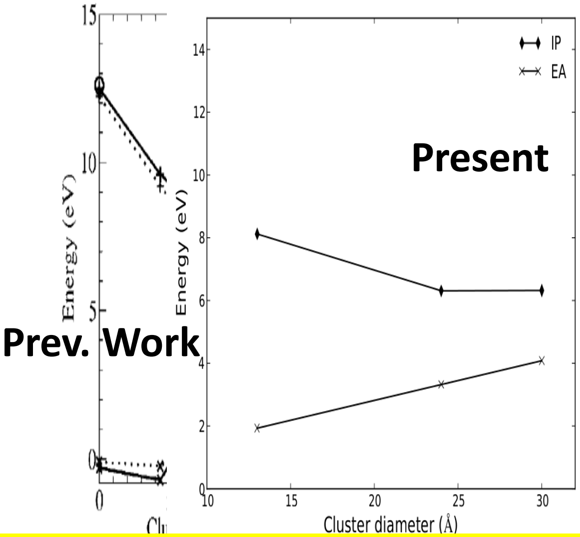
Stochastic compression:

$$(W_{casuaul}g)(\boldsymbol{r},t) \rightarrow \eta(\boldsymbol{r})c(t)$$

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

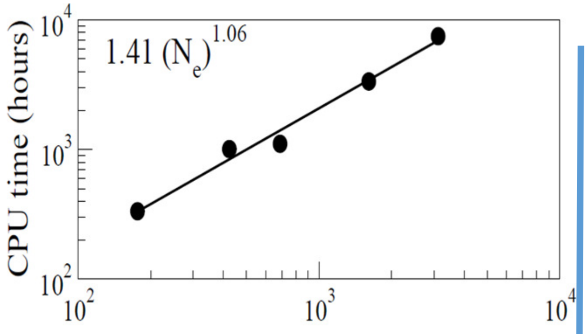
Stochastic GW Results

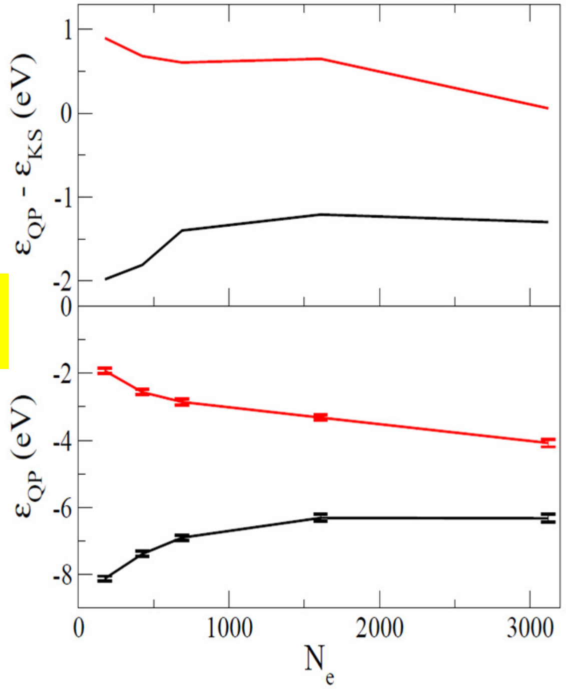
SILICON CLUSTERS



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

System	N_e	N_g	I_{sDFT}	N_φ	$\beta^{-1}_{GW}\,(E_h)$	E^{QP}_{gap} (eV)
$Si_{35}H_{36}$	176	60^3	3000	16	0.020	6.19 ± 0.13
$Si_{87}H_{76}$	424	64^3	1600	16	0.012	4.82 ± 0.14
$Si_{147}H_{100}$	688	70^3	800	16	0.010	4.10 ± 0.13
$Si_{353}H_{196}$	1608	90^3	400	16	0.008	2.99 ± 0.13
$Si_{705}H_{300}$	3120	108^3	200	16	0.007	2.24 ± 0.15





FULLERENES

PCBM (fullerene with add-ons)

GW(TDH): EHOMO = 7.1±0.1 eV

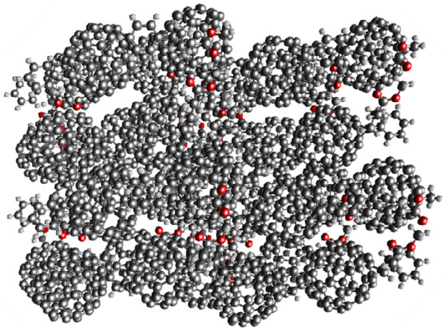
Experiment: EHOMO = 7.17 eV,

GW(TDH): ELUMO = 3.4±0.1 eV

GW(TDDFT): ELUMO = **2.5±0.1 eV**

Experiment: EHOMO = **2.63 eV,**

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



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.