Computational Chemistry and Materials Modeling

Lecture
Computational Chemistry of Molecules: part 1

Dmitry Aksenov (lead instructor)
Sergey Levchenko (co-instructor)
Alexander Kvashnin (co-instructor)
Alexander Shapeev (co-instructor)
Arseniy Burov **(**TA**)**

Outline

- *Practical guide to most common calculations:*
- Molecular structure
- Total energies
- Electronic structure (analysis)

Practical synopsis I

You need to calculate electronic structure of XXX molecule and compare it to experiment

CC is the most accurate method in the world!!!

Use DFT: it is mighty and practical!

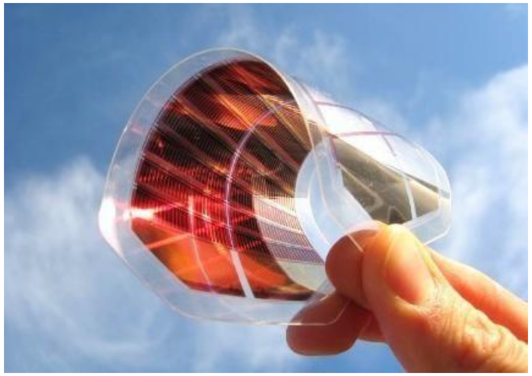
Practical synopsis II

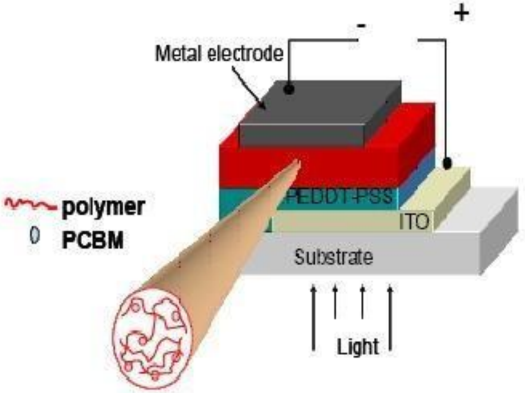
- There is a multitude of methods and codes, and the obvious choice does not always exist
- There is a law of “conservation of difficulty”, and there is a trade-off between efficiency and transferability
- Use physical sense and chemical intuition to maximize (result)/(spent time), i.e. performance/price ratio

Philosophy of large molecule calculations

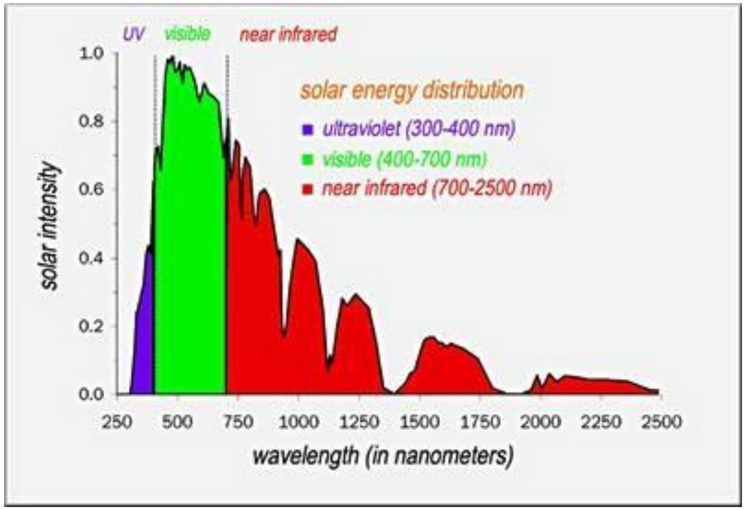
- High accuracy approaches are not feasible
- Benchmarks on small molecules are not usually applicable
- Direct and indirect comparisons with experiment are necessary
- Get the fundamental physics first
- Look to the numbers

Example: Organic solar cells





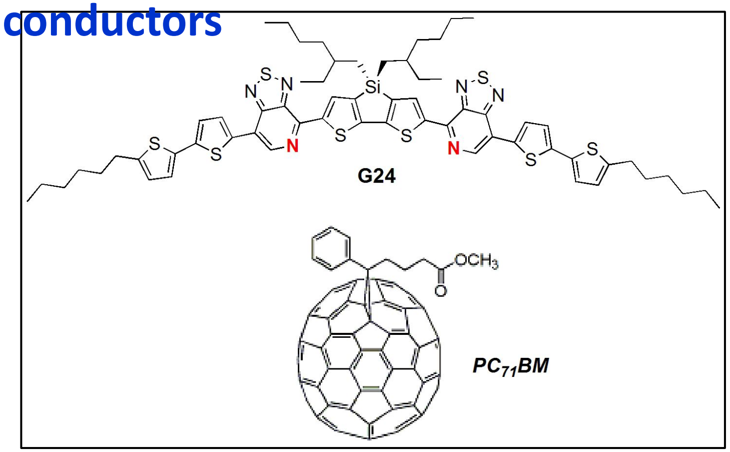
- **Light creates an exciton**
- **Exciton produces charges**



Usually you will not be asked to calculate something, instead, the problem may be formulated like this: ***Need new photoactive materials for bulk-heterojunction solar cells*** *(broad absorption spectrum matching the spectrum of solar light + large exciton diffusion length + efficient charge separation at the heterojunction + high hole mobility + …)*

Example: Organic solar cells

Usually a mix of 2 components: electron and hole conductors

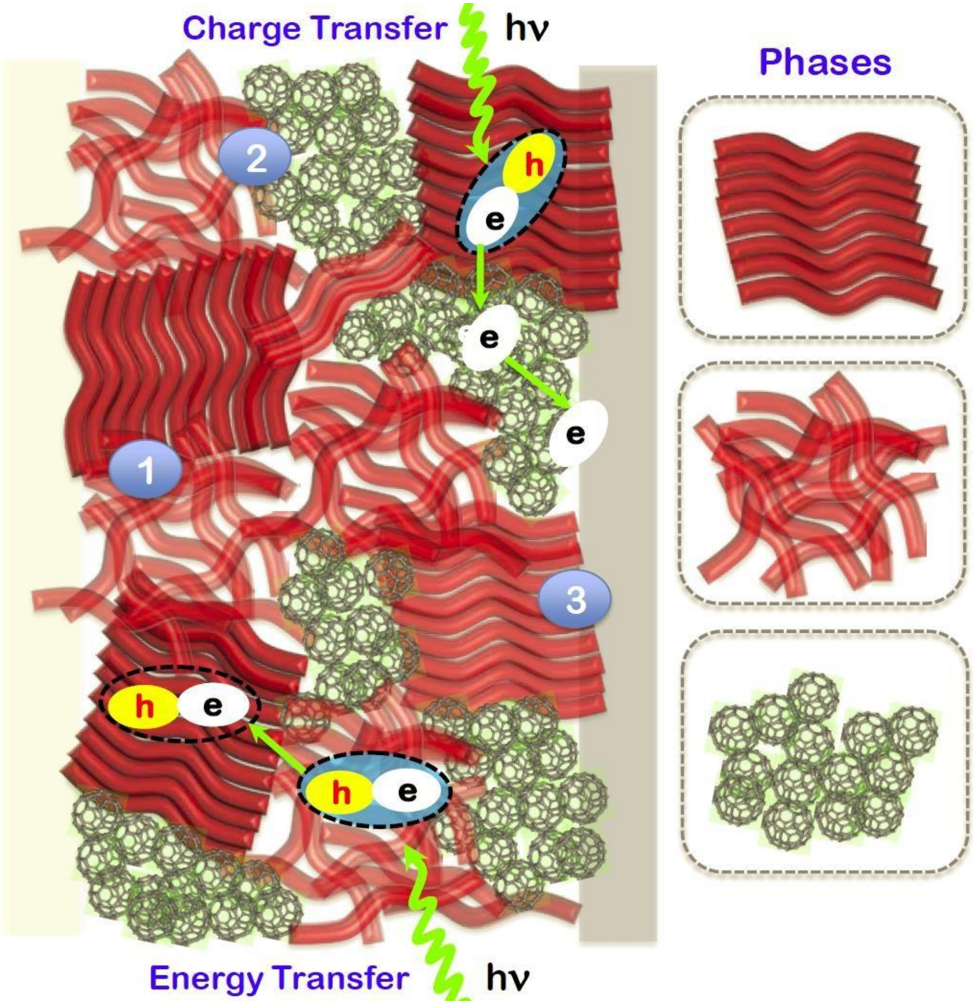


Fundamental processes in 'soft' organic materials: light collection, energy transfer, charge separation, charge transport:

-Conversion of excitation energy (exciton) into electrical/chemical energy (charges) and *vice versa*.

-Exciton and charge transfer processes as a function of molecular conformations and packing at the interface.

-Carrier injection/extraction from organic materials.



Molecular structure

- How to get initial guess for geometry
- Geometry optimization (PES extremum)
- If there are multiple minima (PES scan)
- Transition state (saddle point)

How to get initial geometry

- Databases: NIST, COD, CCDC, Materials Project (clean up, remove disorder, add H)
- Generate from asymmetric unit
- Use internal coordinates (Z-matrix)
- Use molecular/crystal builders

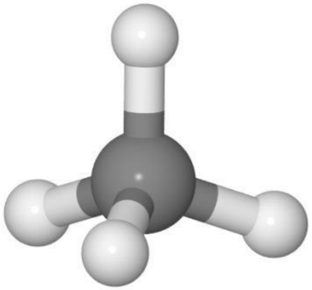
- Enumerate atoms wisely and consistently
- Orient and symmetrize if appropriate
- Store geometries as separate documented files in commonly recognized formats (XYZ or CIF)

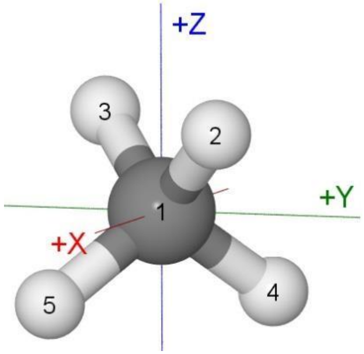
Wrong initial geometry will nullify all calculations

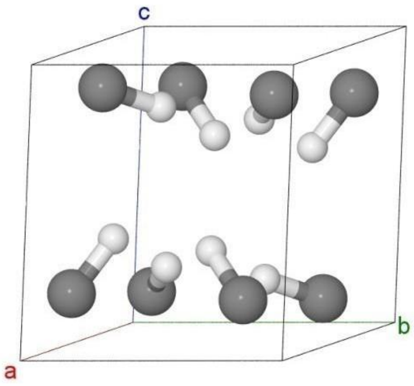
Examples of bad initial geometries

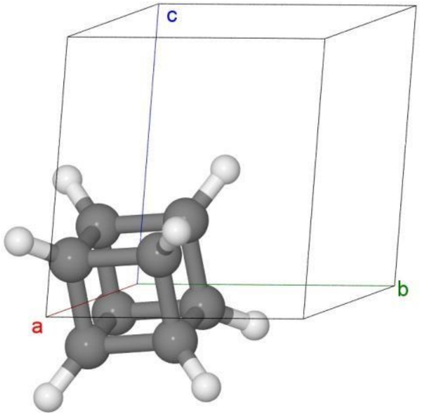
Bad

Good



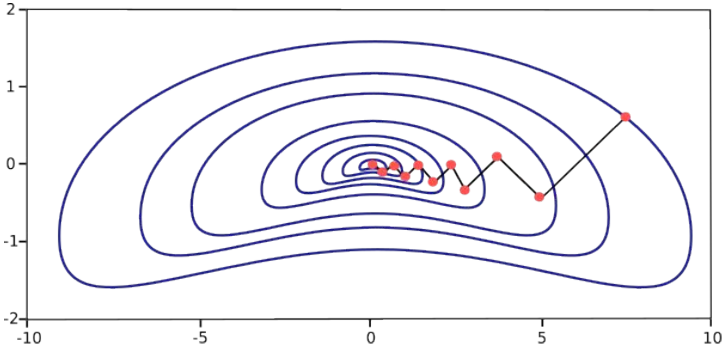


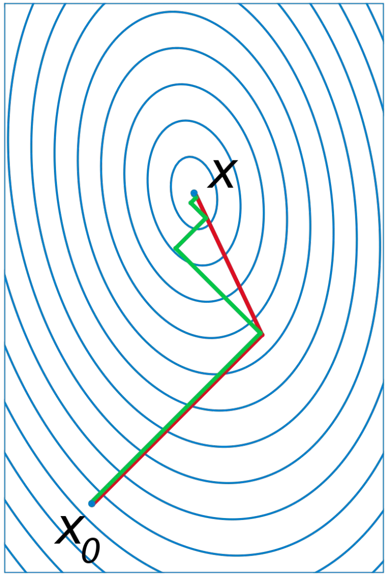




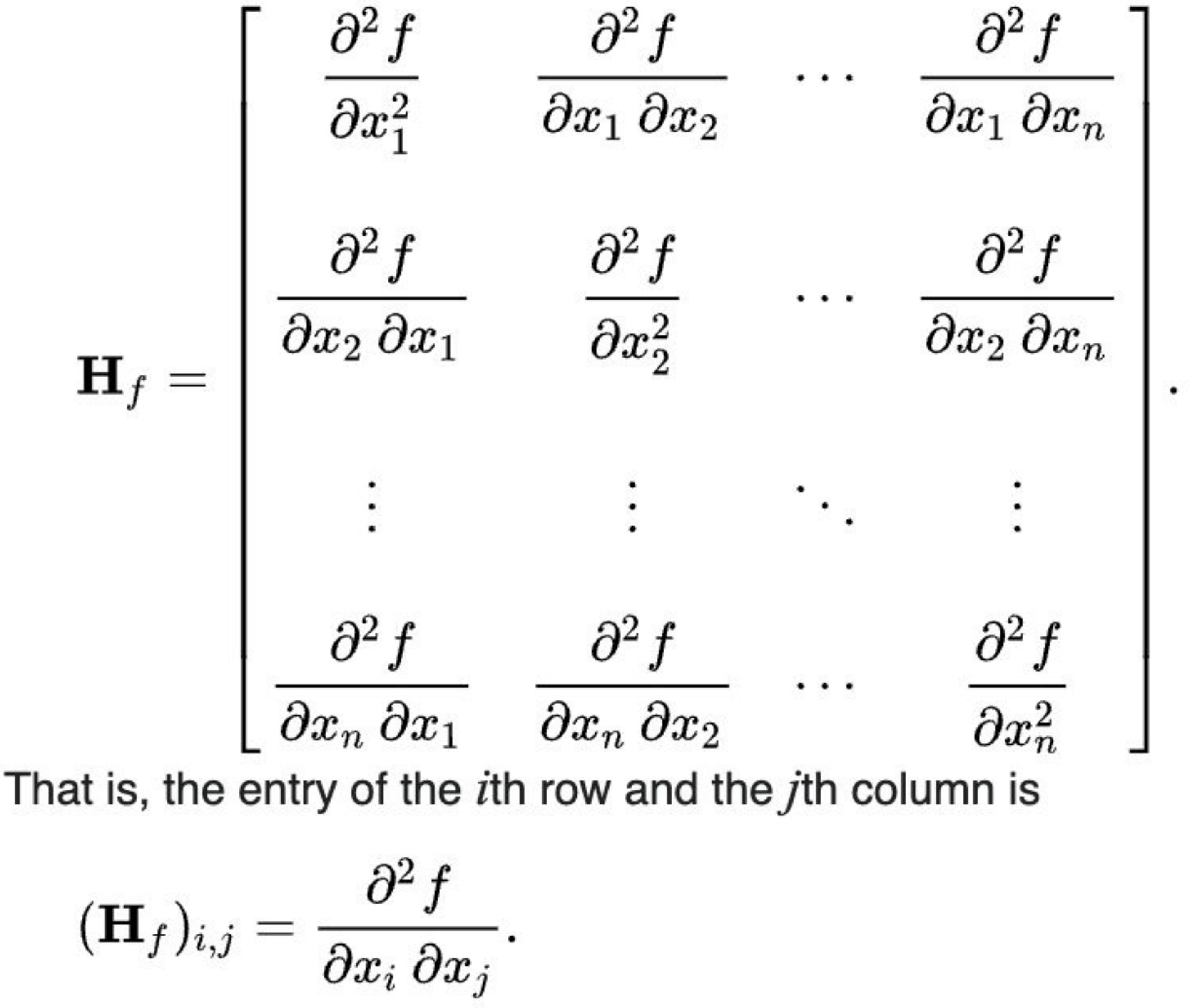
Geometry optimization: theory

- Usually you will find local extremum on PES (Hessian-based methods can find minimum)
- Apply appropriate method:
  - steepest decent
  - conjugated gradients
  - quasi-Newton methods
  - least squares extrapolation methods (DIIS)
  - damped dynamics
- Use analytic gradients and Hessians
- Constrained optimization (by symmetry, by fixing some coordinates)
- Unit cell optimization is a separate problem





Hessian - the second derivative of energy



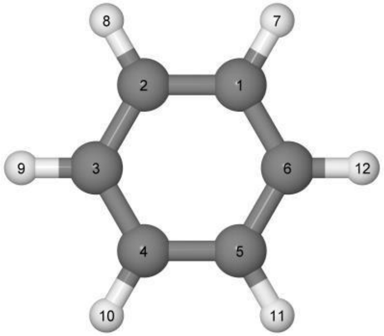
$$\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

That is, the entry of the *i*th row and the *j*th column is

$$(\mathbf{H}_f)_{i,j} = \frac{\partial^2 f}{\partial x_i\,\partial x_j}.$$

Geometry optimization: practice

- Save geometry after optimization (XYZ/CIF file)
- Do single point to get final energy and gradient
- Check Hessian if feasible
- Is there a symmetry breaking?
- Symmetrize if not done by optimizer
- Use reasonably tough stopping criteria
  (e.g. intermolecular or for Hessian)
- If optimization fails (bad convergence or result):
  - determine the origin of the problem
  - for wave-function convergence see next slide
  - check initial geometry, modify if needed
  - use other algorithm or other coordinates
  - select "best" geometry (min E, min G)



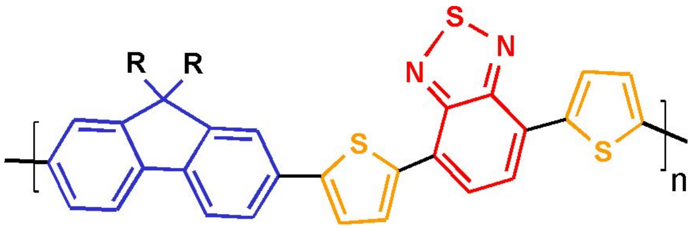
SCF convergence problems

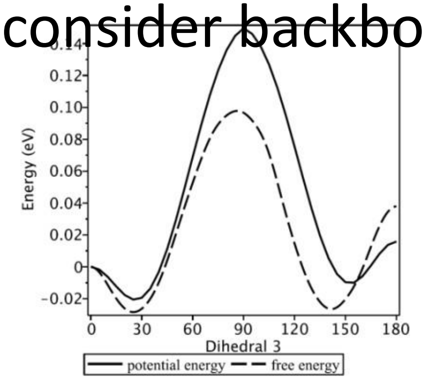
- Determine the origin of the problem
- Check initial guess for MOs, modify if needed
- Use other algorithm or tune parameters
- If gap is small use proper Fermi level smearing
- Preconverge by other method
- Converge from other geometry

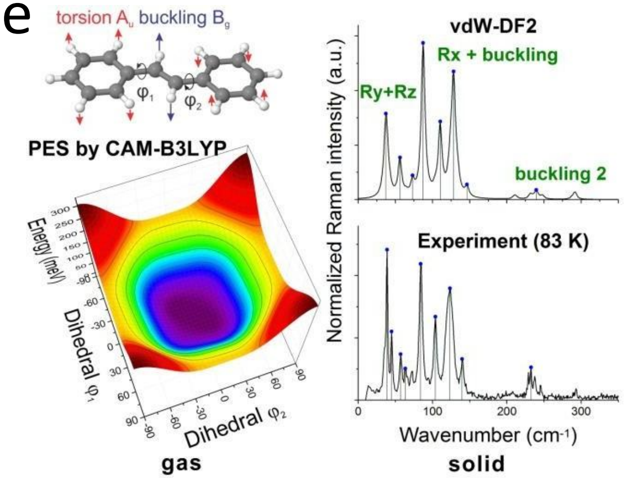
- Do not weaken convergence stopping criteria!
- In contrast to geometry optimization, it is often impossible to converge wave-function by a given method in a given program without re-coding

If there are multiple minima

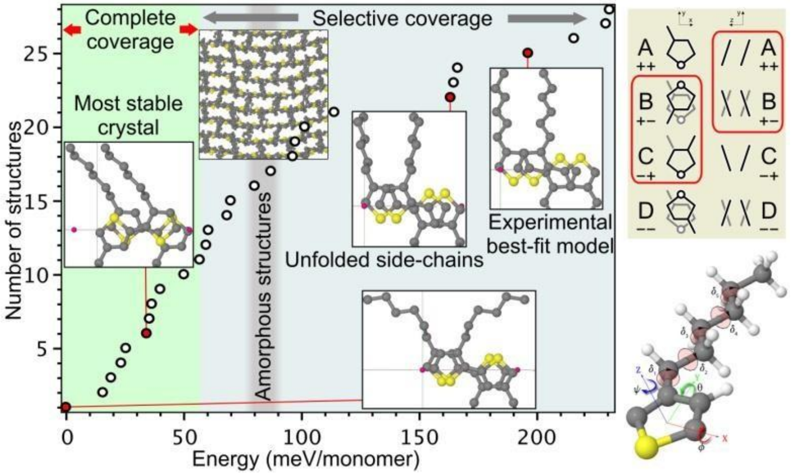
- Small distortion – use high symmetry structure
- Noncritical side-chains – consider backbone







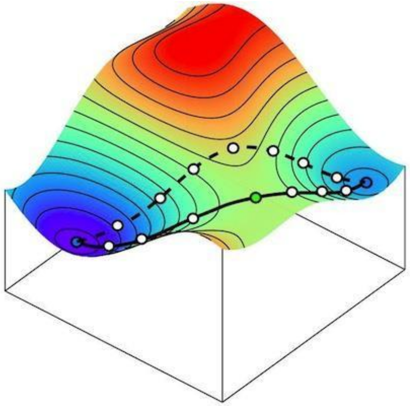
- Few coordinates – scan PES
- Several coordinates – enumerate



- Huge number of conformers:
  - use proper thermodynamic energy
  - use MD or MC to sample
  - use special software, e.g. USPEX code

Transition state

- Use ordinary optimizer if close to transition
- Use path optimizer if single path:
  - Nudged Elastic Band
  - Quadratic Synchronous Transit
- Use special methods if complex PES



- Be careful if the band gap is small at the transition (electron transfer, chemical reactions, ethylene)

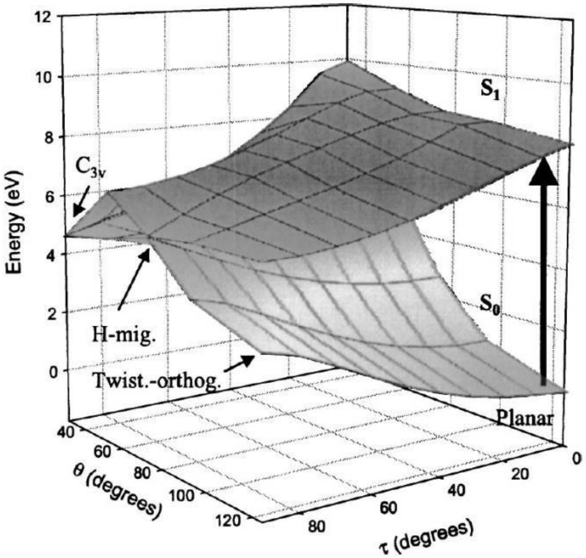
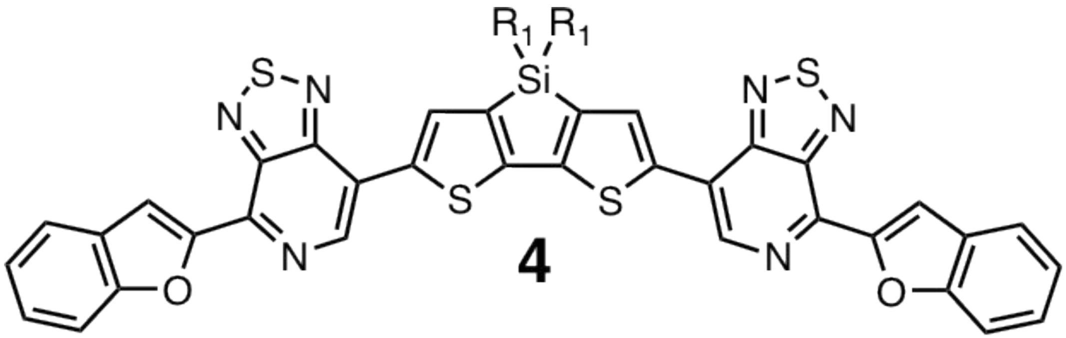


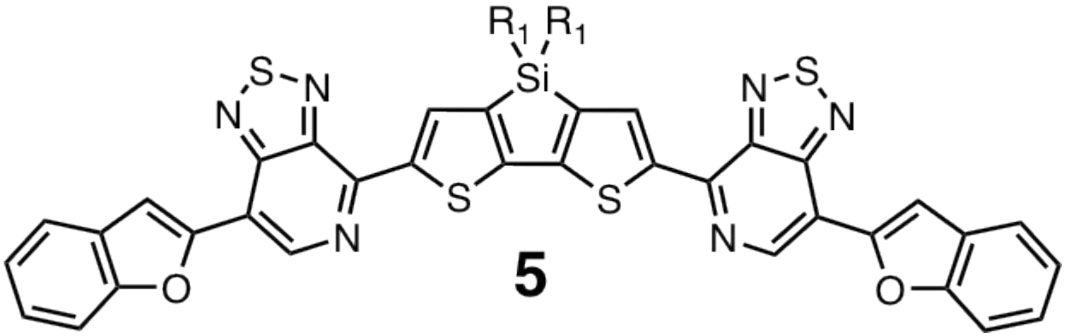
FIG. 9. S_0 and S_1 potential energy surfaces in dependence of the H migration (θ) and the torsion (τ) angles optimized for S_1 state. Calculations performed at the same level as given in Fig. 5. The planar ground-state ethylene energy is the energy zero. DOI: 10.1063/1.1807378

Example: Barriers and conformers

Experimentalist: I have synthesized a few new molecules, one seems to crystallize well and the second is not. What is going on?



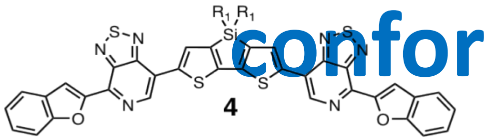
4



5

Master Student, how do you approach?

Example: Barriers and conformers

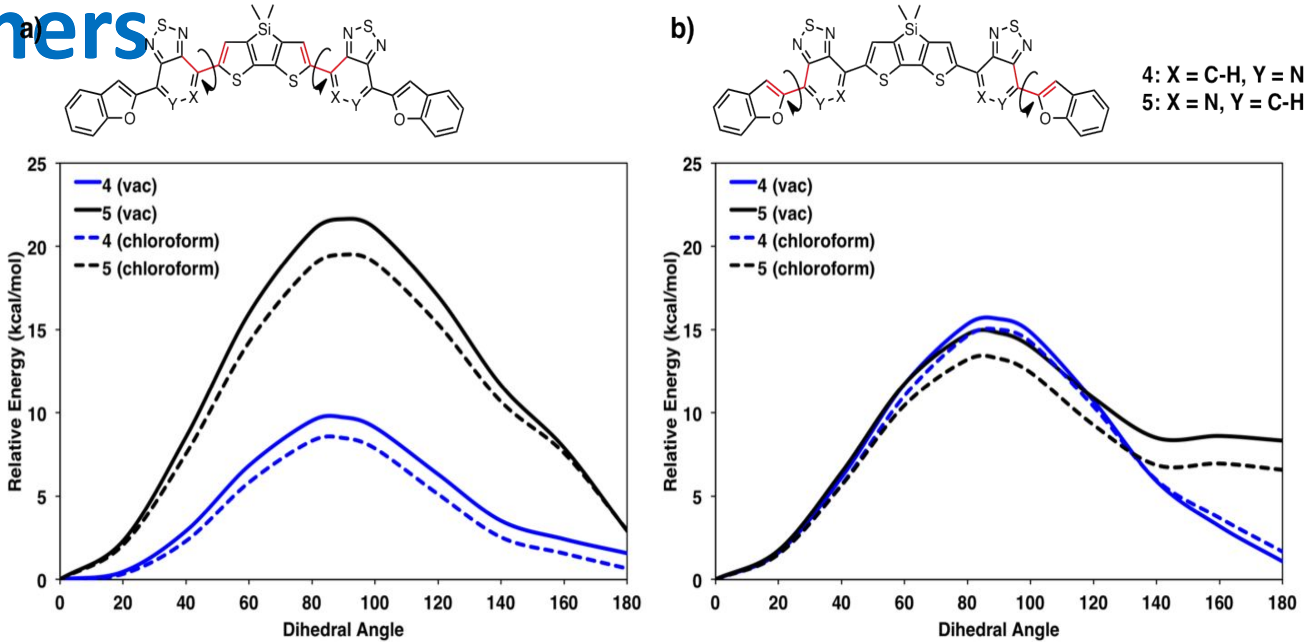


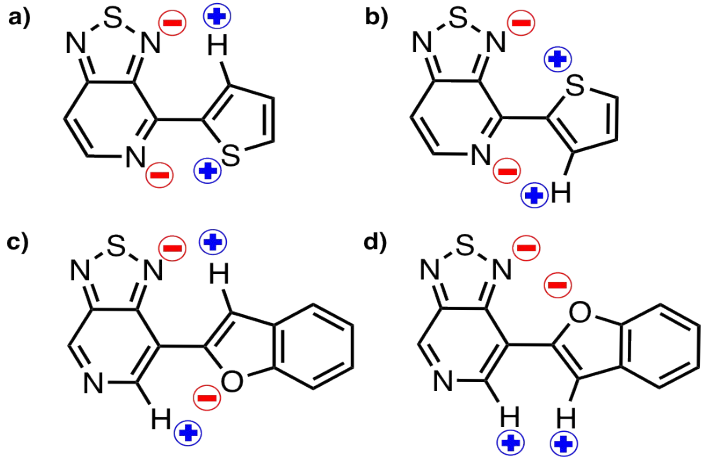
4

5

Very different energies for rotamers for compound 4, but similar for compound 5. Possibly will affect crystallization (multiple conformers in solution will not assemble well!)

Why? Combination of steric and electrostatic interactions.





Discussion

1. What if after geometry optimization (a) there are several imaginary vibrational frequencies; (b) several frequencies are close to 0?

2. How do you check if your geometry corresponds to a transition state?

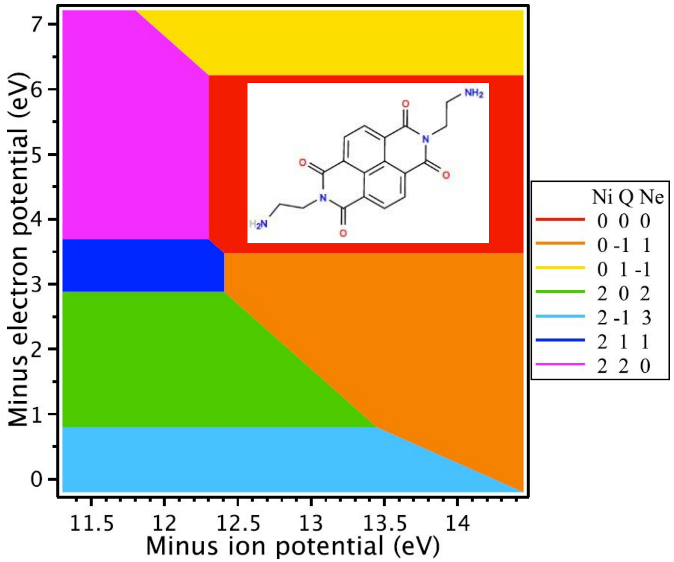
3. In DFT calculations with gradient-based optimization, your final geometry had lower symmetry than the initial one. How could the program code break the symmetry?

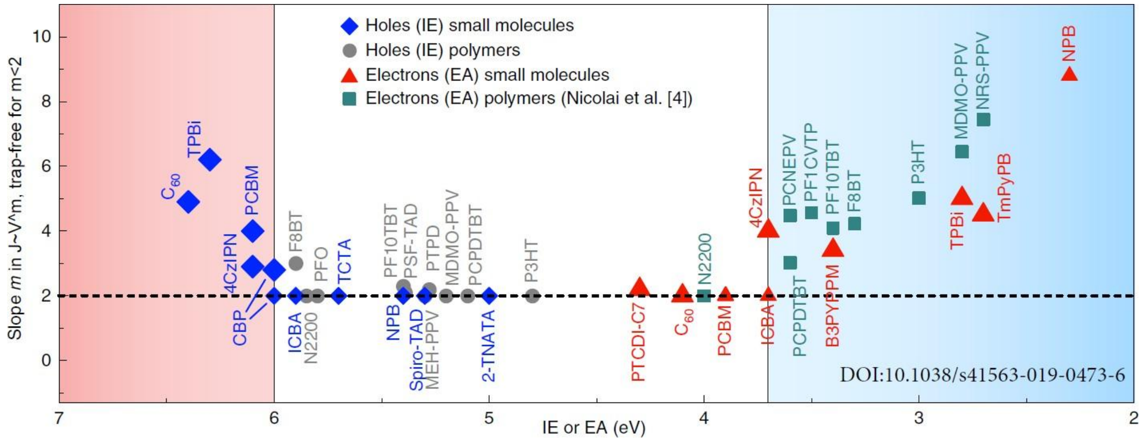
Total energy

- Use relative energies, not absolute – idea behind composite methods
- Use the same method for all comparisons
- Zero-point energy and vibrational free energy
- Avoid basis set superposition error, e.g. by counterpoise correction
- Study dependence on method (basis set, density functional)
- Higher level of theory can be used to benchmark your result

Varying number of atoms and electrons

- To compare energies use chemical potential (μ):
  $\mu(X)+E(A)=E(XA)$
- Be careful comparing μ in different methods
- What reference value of μ can we use:
  - Relevant to problem, e.g. experimental or lowest in your dataset
  - From Standard Hydrogen Electrode (SHE):
    $2p(aq)+2e=H_2(g)$, μ(e)=-4.44 eV, μ(p)≈-11.37 eV[DOI:10.1063/1.5000799]
    this μ(e) fits bandgap of semiconductors, μ(p) depends on pH



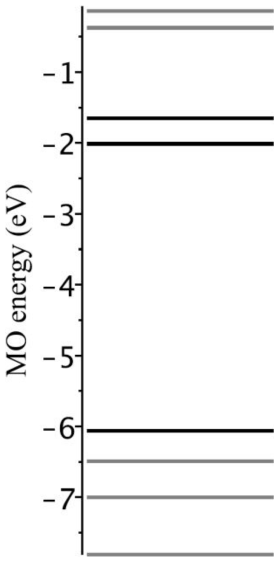


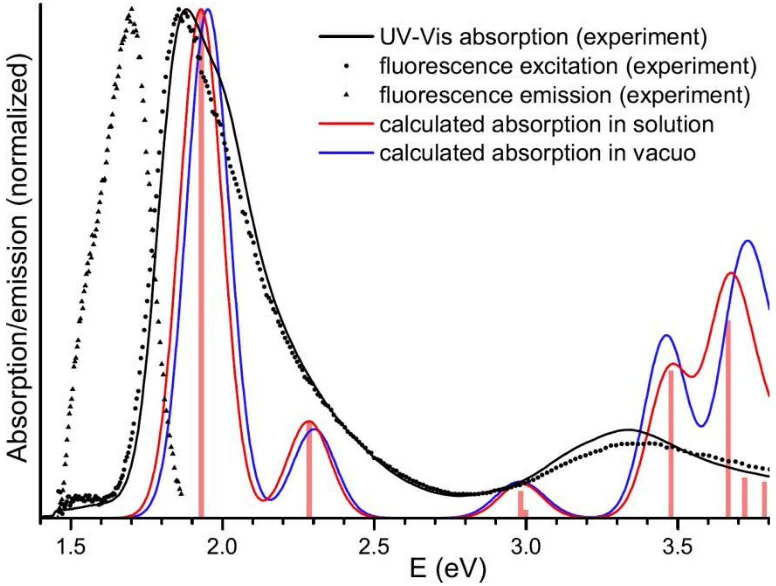
DOI:10.1038/s41563-019-0473-6

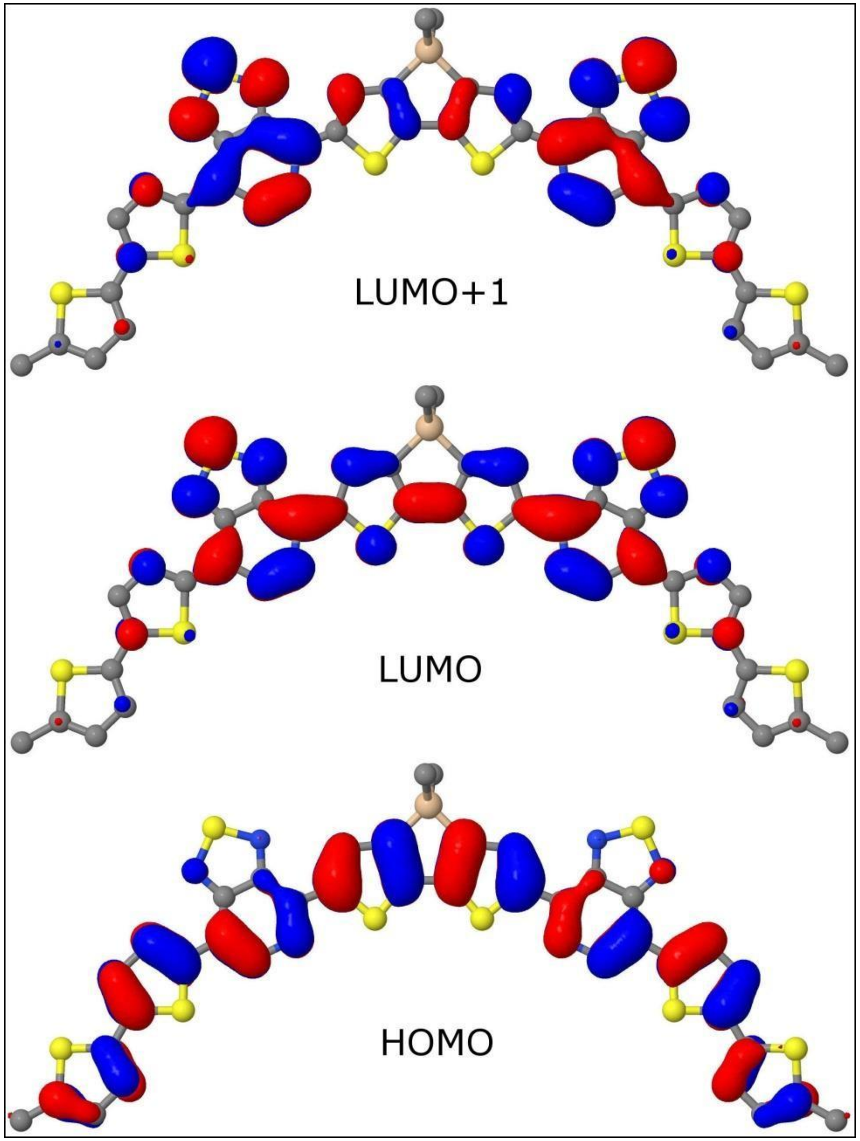
Electronic structure (analysis)

- We are talking about 1e properties
- Analysis of 1e Hamiltonian:
  - molecular orbitals (MO), frontier MO, LMO
  - DOS, band structure E(k)
- Analysis of 1e density matrix and 1e density:
  - population analysis, atomic charges and bond orders
  - natural orbitals and natural transition orbitals
  - natural bond orbital (NBO) analysis
- 1e density -> electrostatic field (multipoles)

Example: organic solar cells

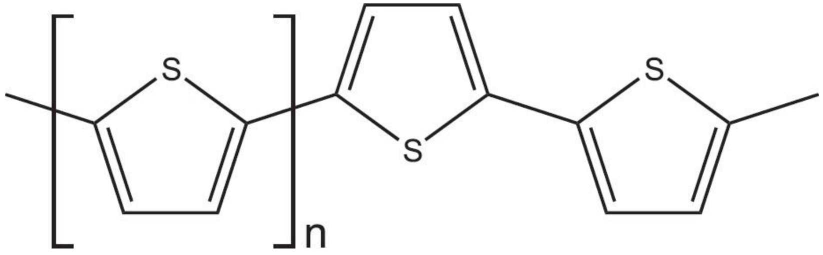


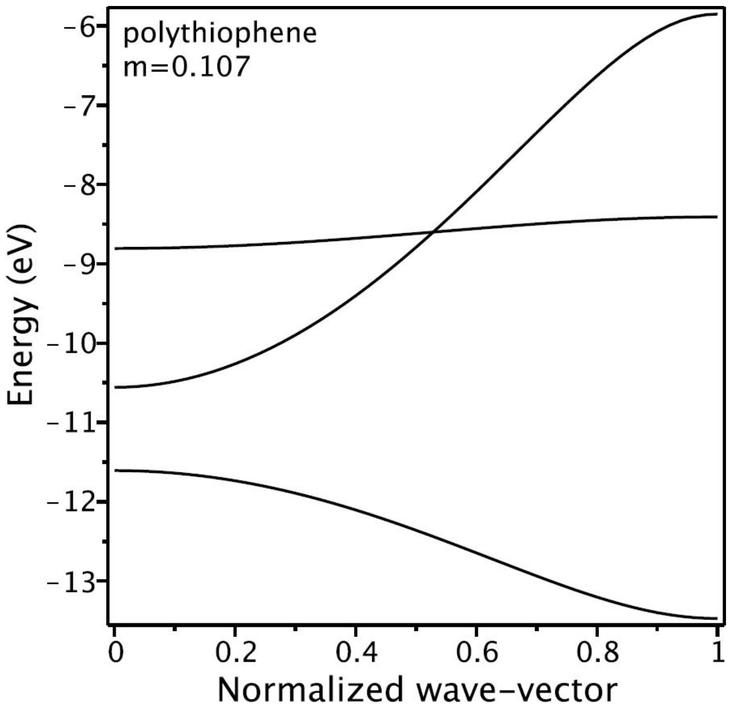


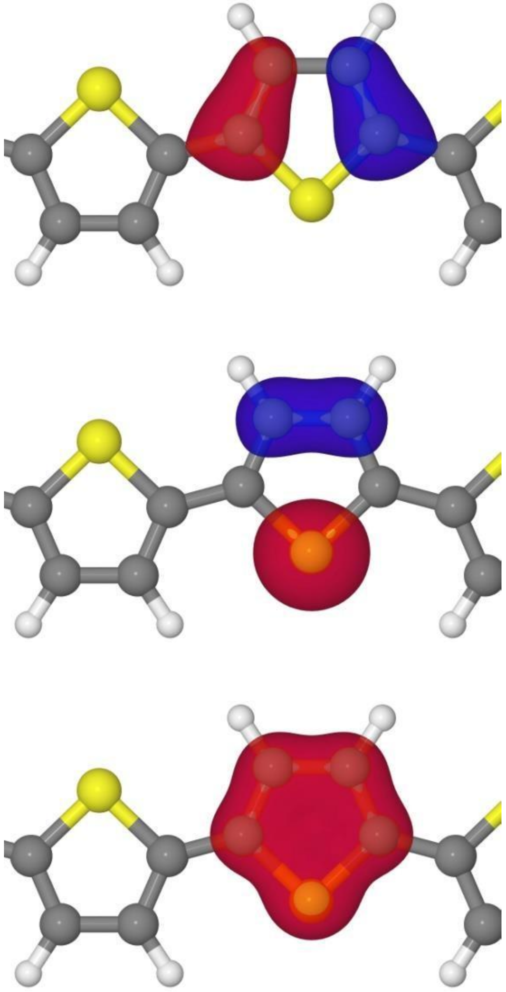


Example: LMO in polymers

If unit cell has >20 atoms band structure is not informative







1e orbitals: definitions

- molecular orbitals (MO) – eigenfunctions of one-electron Hamiltonian (HF/DFT)
- localized molecular orbitals (LMO) – a rotation of MOs localizing each orbital in space
- natural orbitals (NO) – eigenfuntions of one-electron density matrix $\rho^{1e}_{\Psi}(\xi;\eta) = N \int \Psi(\xi,\zeta_2,\ldots,\zeta_N)\overline{\Psi}(\eta,\zeta_2,\ldots,\zeta_N)\,\mathrm{d}\zeta$
- natural transition orbitals (NTO) – the same for transition density matrix $\rho^{1e}_{\Psi\Phi}(\xi;\eta) = N \int \Psi(\xi,\zeta_2,\ldots,\zeta_N)\overline{\Phi}(\eta,\zeta_2,\ldots,\zeta_N)\,\mathrm{d}\zeta$

Atomic charges and bond orders

It is the simplest yet usually sufficient analysis of 1e density matrix

Basic approach – Mulliken charges and bond orders:

- Atomic charge

$$\rho_{\mathrm{A}} = \sum_{\alpha\in \mathrm{A}}^{M_{\mathrm{basis}}} \sum_{\beta}^{M_{\mathrm{basis}}} D_{\alpha\beta} S_{\alpha\beta} \qquad Q_{\mathrm{A}} = Z_{\mathrm{A}} - \rho_{\mathrm{A}}$$

- Bond order between two atoms

$$\mathrm{BO}_{\mathrm{AB}} = \sum_{\alpha\in \mathrm{A}}^{M_{\mathrm{basis}}} \sum_{\beta\in \mathrm{B}}^{M_{\mathrm{basis}}} (\mathbf{DS})_{\alpha\beta} (\mathbf{DS})_{\beta\alpha}$$

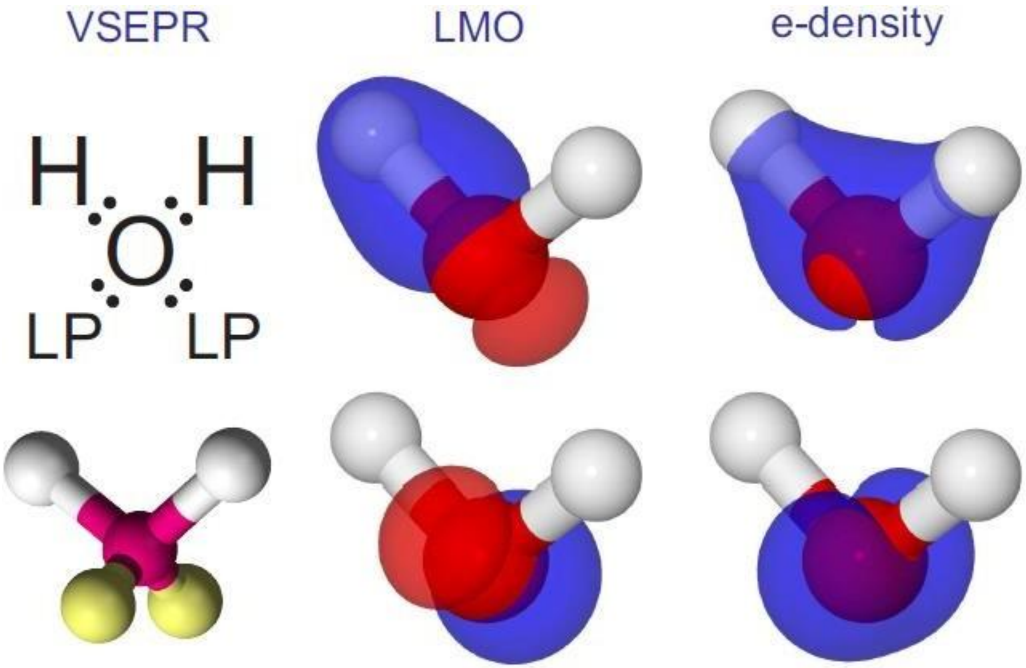
- But sensitive to basis set

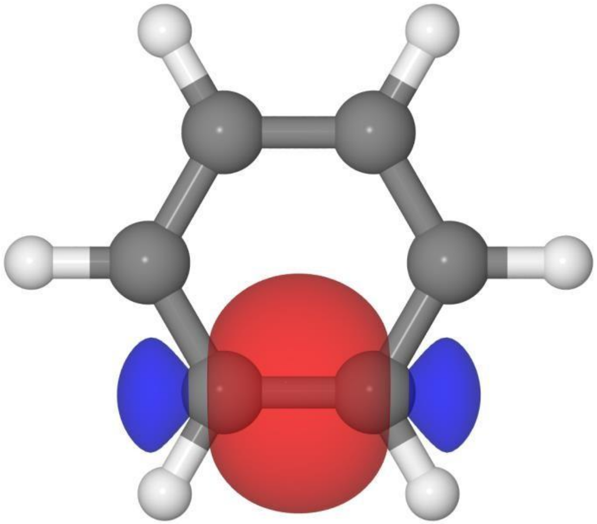
Advanced – Weinhold’s Natural Bond Orbital (NBO) analysis:

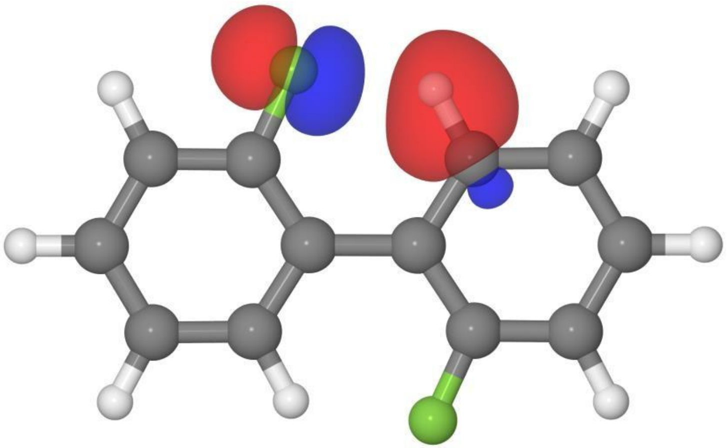
- NAO (Natural Atomic Orbitals): orthogonalized orbitals localized on atoms → comprehensive analysis of AO populations (e.g. 3s, 4dxy)
- NBO (Natural Bond Orbitals): orthogonalized orbitals localized on bonds (e.g. σ, π*) or atoms (e.g. LP, core, Ry)

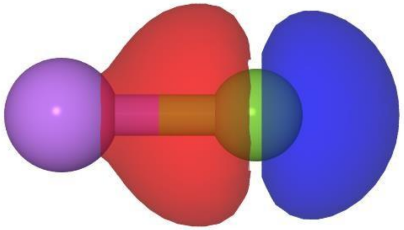
From NBO website: More precisely, NBOs are an orthonormal set of localized "maximum occupancy" orbitals whose leading N/2 members give the most accurate possible Lewis-like description of the total N-electron density.

Example: LMO/NBO analysis of bonding

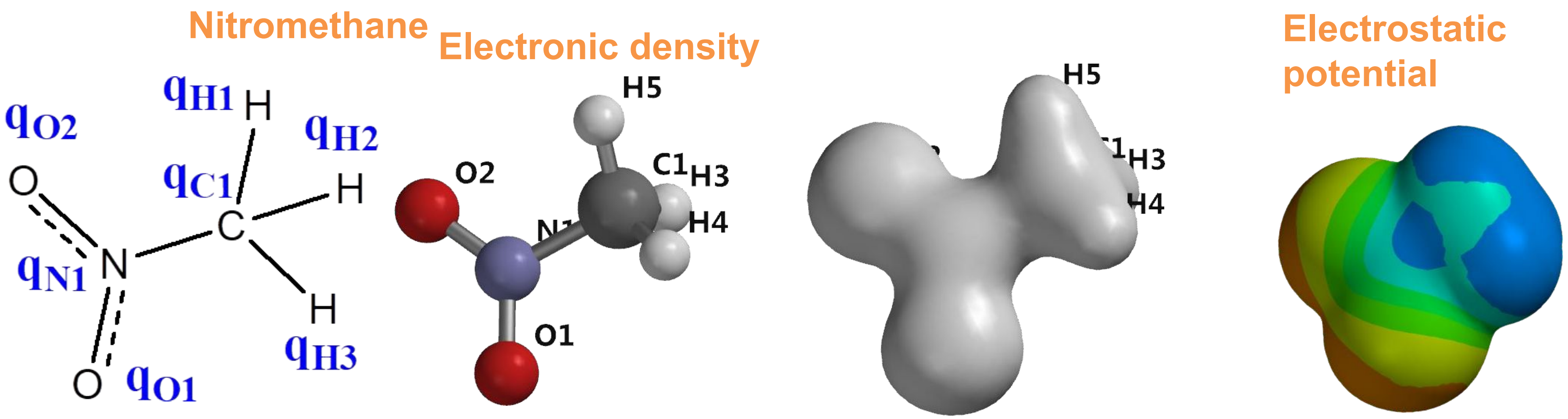








Atomic charges from charge density



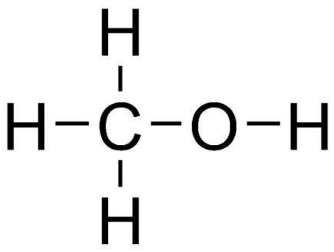
Used for force field parameterization, IR spectra etc

There are different charge partitioning schemes:

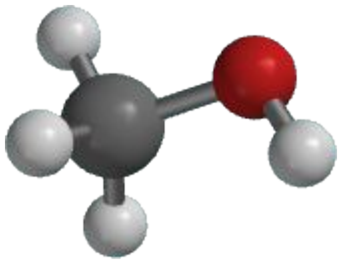
- ***Mulliken and NBO:*** not designed to represent charge density
- ***Hirshfeld***: among first charges (1977) to work well in force fields
- ***Charge model 5 (CM5):*** improved Hirshfeld, better dipoles (2012)
- ***Electrostatic potential fitting (ESP)***: MSK (1984), CHELPG,

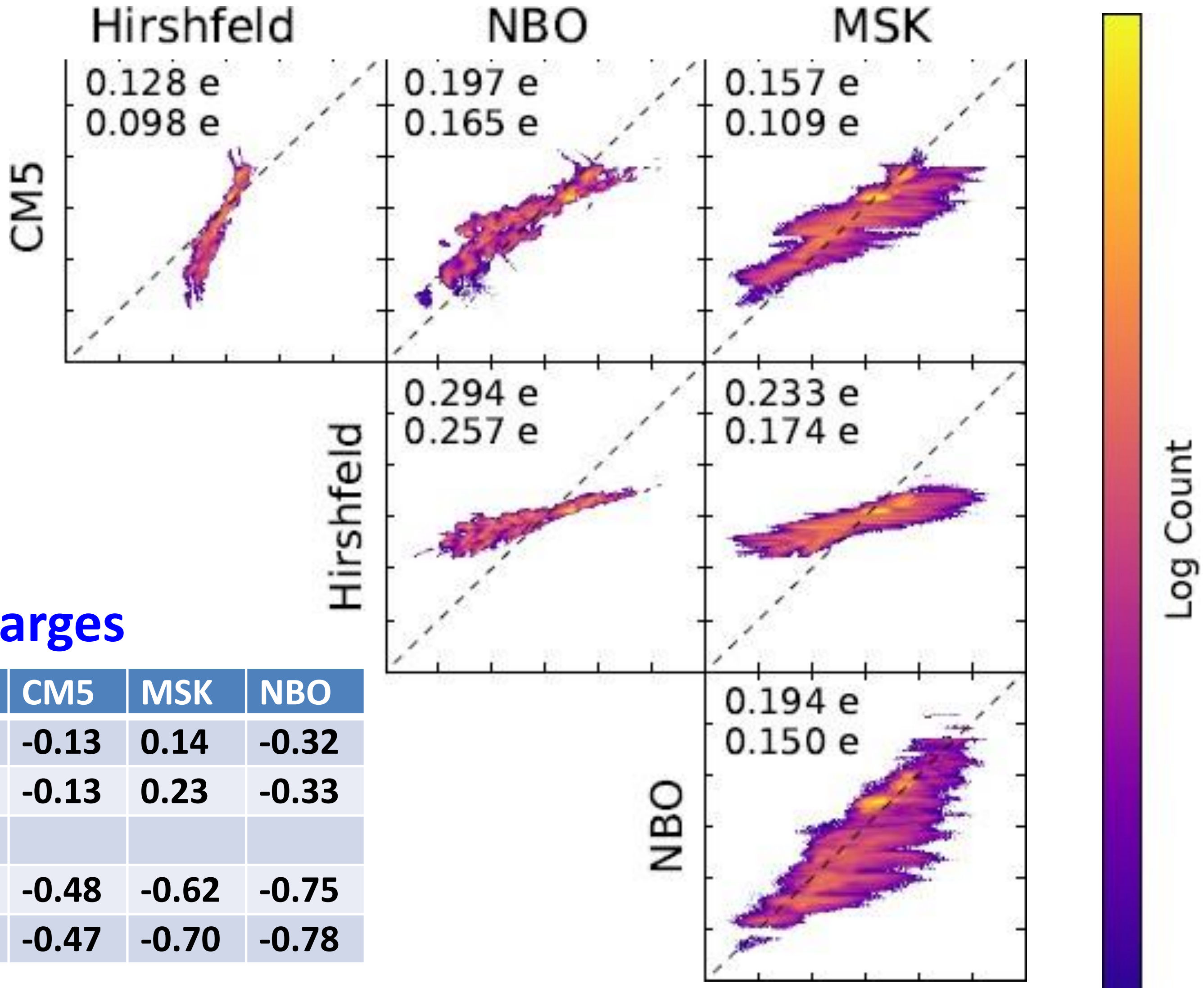
  charges depend on what region to fit
- ***Machine learning:*** see next slides [J Phys Chem Lett 9, 4495 (2018)]

Atomic charge schemes do not agree!



$$\mathrm{H{-}\overset{H}{\underset{H}{C}}{-}O{-}H}$$





Methanol charges

Carbon	Mulliken	Hirshfeld	CM5	MSK	NBO
6-31g*	-0.23	-0.007	-0.13	0.14	-0.32
6-31+g*	-0.36	-0.01	-0.13	0.23	-0.33
Oxygen					
6-31g*	-0.62	-0.26	-0.48	-0.62	-0.75
6-31+g*	-0.66	-0.25	-0.47	-0.70	-0.78

A. Sifain, N. Lubbers, B. Nebgen, J.S. Smith, A.Y. Lokhov, O. Isayev, A.E. Roitberg, K. Barros, S. Tretiak, J. Phys. Chem. Lett. 9, 4495 (2018)

ML Affordable Charge Assignment (ACA)

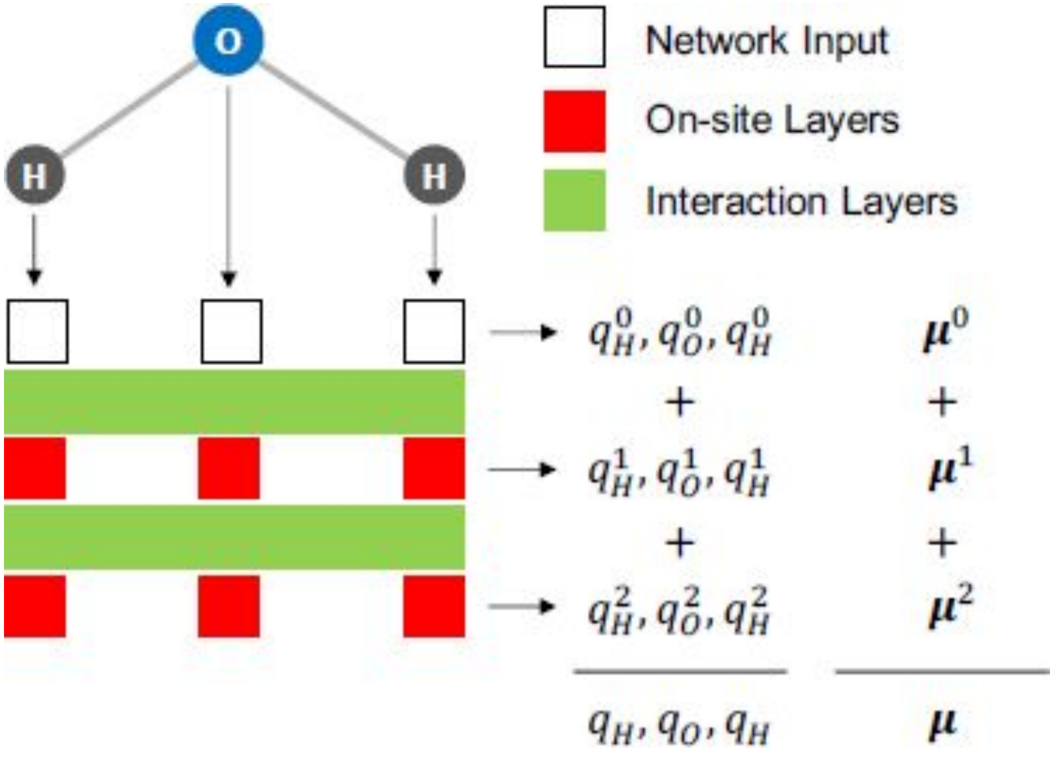
Can we choose atomic charges such that the point charge calculation of the dipole reproduces the *ab initio* dipole?

Hip-NN neural nets

ANI-1X training set

Reference DFT:

wB97x/6-31g(d)



$$\boldsymbol{\mu}^{ML} = \sum_{i}^{N_{atoms}} q_i \boldsymbol{r}_i$$

$$\boldsymbol{\mu}^{REF} = \langle \psi | \rho \boldsymbol{r} | \psi \rangle$$

Cost function

$$\mathcal{L} = \langle (\boldsymbol{\mu}^{REF} - \boldsymbol{\mu}^{ML})^2 \rangle_{Dataset}$$

Only ground state dipoles were learned!

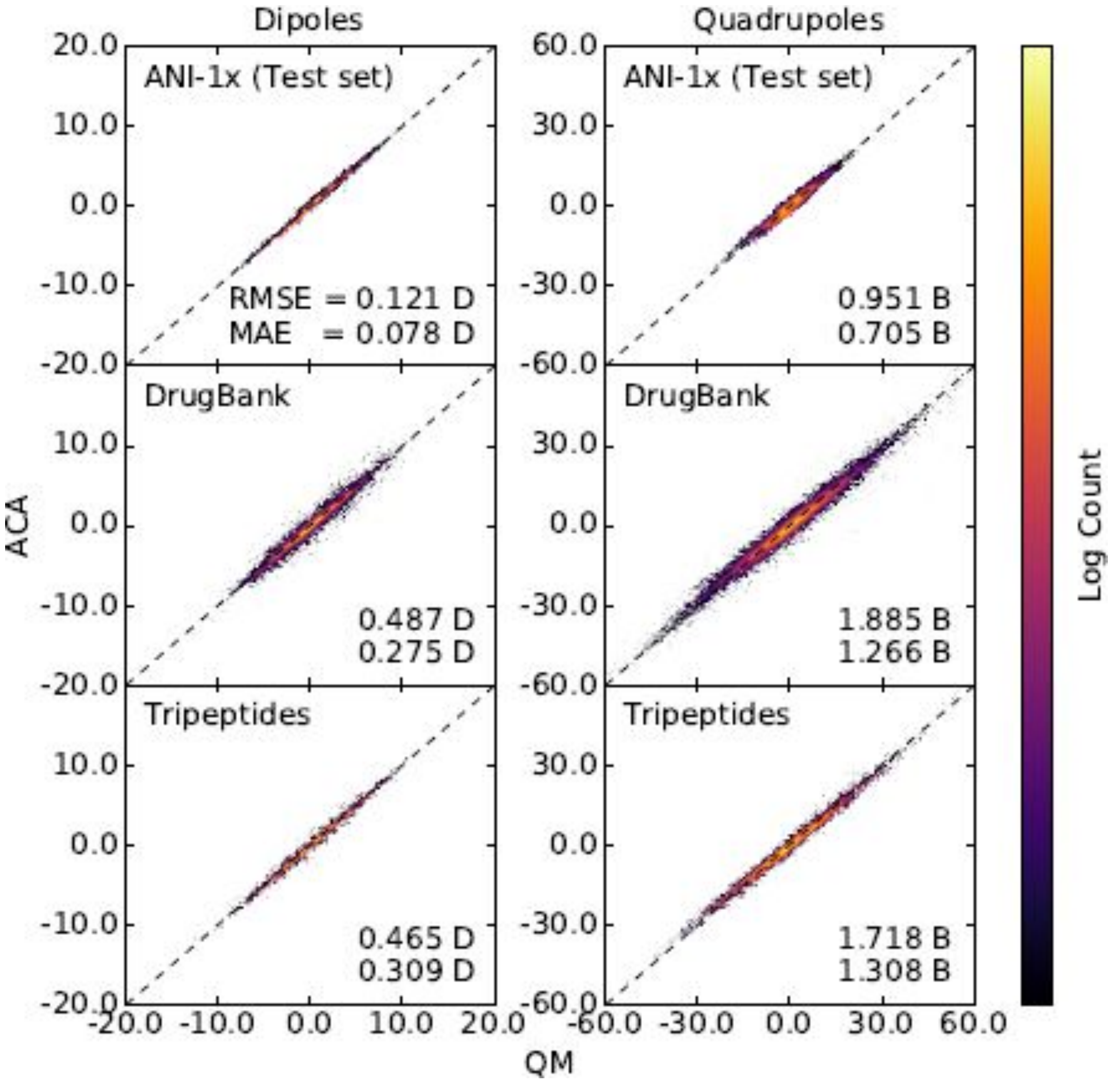
A. Sifain, N. Lubbers, B. Nebgen, J.S. Smith, A.Y. Lokhov, O. Isayev, A.E. Roitberg, K. Barros, S. Tretiak, J. Phys. Chem. Lett. 9, 4495 (2018)

Dipole and quadrupole predictions

Dipole accuracy ~0.1-0.5 Debye

Quadrupoles are 'for free' - never trained. The accuracy ~1-2 Barns (C*m^2)

The accuracy 'per atom' is the same across all datasets!



A. Sifain, N. Lubbers, B. Nebgen, J.S. Smith, A.Y. Lokhov, O. Isayev, A.E. Roitberg, K. Barros, S. Tretiak, J. Phys. Chem. Lett. 9, 4495 (2018)

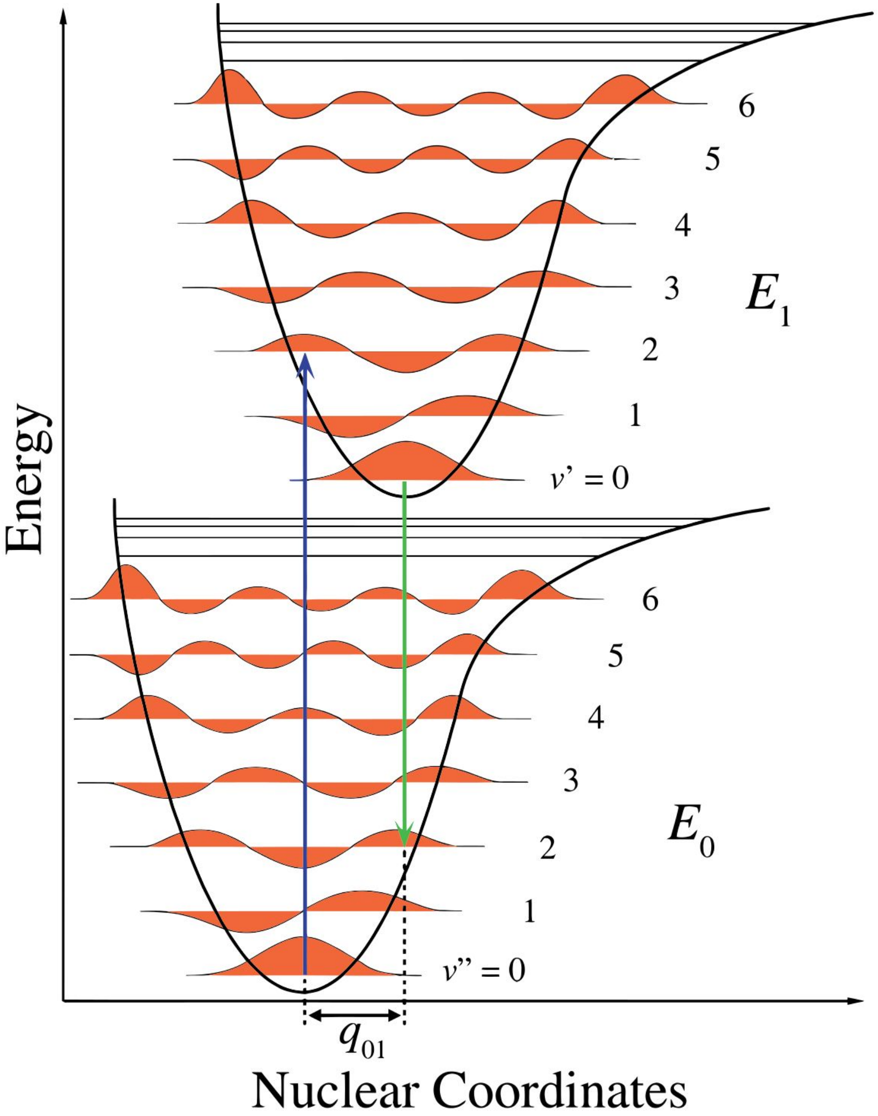
Discussion

1. Define bonding and antibonding MOs.
2. Why one can take a superposition of MOs, for example, to form LMOs?

Vertical and adiabatic IP

- IP=E(Q=+1)-E(Q=0)
- Koopman's theorem:

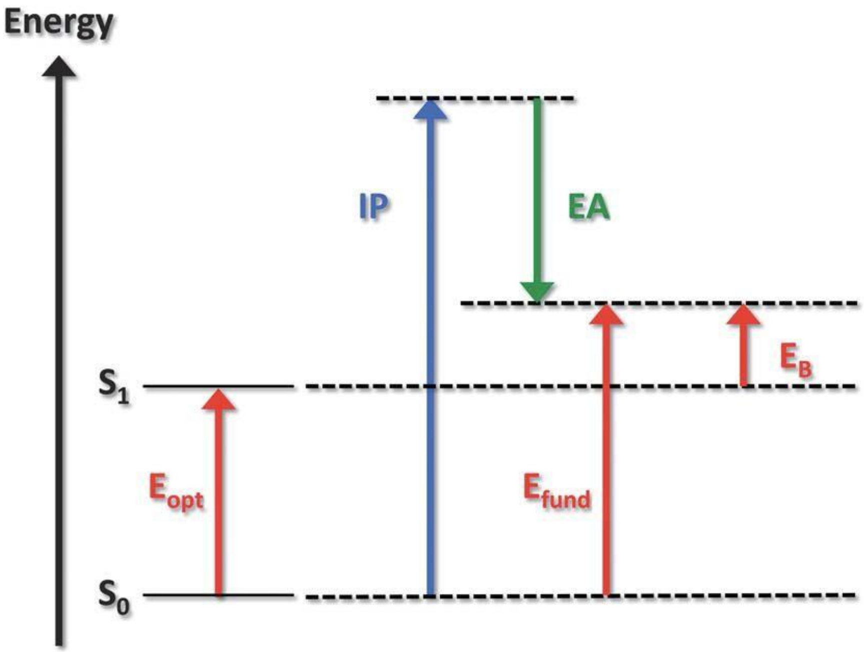
  IP_{vert}=-E(HOMO)



Mind the gap

J Bredas, Mind the gap, Mater Horiz 1, 17 (2014)

- Charge gap (IP-EA)
  (fundamental or transport gap)
- Optical gap (S1-S0)
  (absorption edge)
- Spin gap (T1-S0)
- HOMO-LUMO gap
  (!experimentalists interpret it either as optical or charge gap!)
- Band gap center (-IP-EA)/2
  (minus Mulliken electronegativity)



- For closed shell systems charge gap > optical gap > spin gap
- For strong correlations one of the gaps approaches zero
- By default all gaps are vertical (only electronic relaxation), but adiabatic gaps (also nuclei relaxation) are also meaningful

Calculation of heat of formation: molecule in gas phase

Basic thermodynamics

- Thermal energy $E(S,V) = E_0 + E_{\text{rot+vib}}(T) + E_{\text{translat}}(T,V)$
- Enthalpy $H(S,p) = F + pV$
- Free energy $F(T,V) = E - TS = -T \ln Z$
- For ideal gas $pV = NT$, $E_{\text{translat}} = (3/2)T$, and

$$\frac{S}{N} = \frac{5}{2} + \ln\left[\left(\frac{m_{\text{molecule}}}{2\pi\hbar^2}\right)^{3/2} \frac{T^{5/2}}{p}\right] + \ln(\text{spin multiplicity})$$

Reference state

- Standard conditions: $T = 298.15$ K, $p = 1$ atm
- Atomization energy – fully separated atoms
- Heat (enthalpy) of formation – atoms in their ground states[1]

---

[1] H – H_2 gas, O – O_2 gas, Br – Br_2 liquid, C – graphite, S – rhombic sulfur

Calculation of heat of formation: practical aspects

Challenge: get 1 kcal/mol accuracy, while absolute error in total energy calculation is usually $\gg 1$ kcal/mol

Approach 1: Atomization

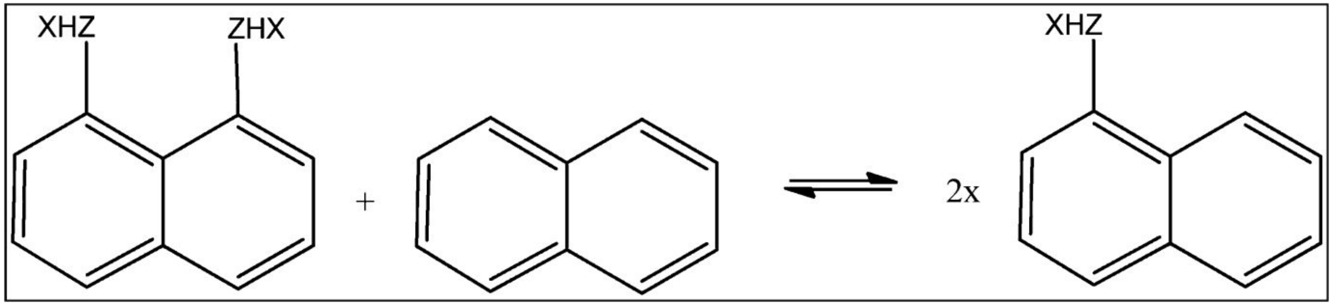
- $X = \sum_i n_i A_i$
- $\Delta H(X) = \sum_i n_i \left[\Delta H(A_i) - E_0(A_i)\right] - \frac{5}{2}NT + H(X)$
- Take $\Delta H(A_i)$ from reference tables
- Calculate $H(X)$ accurately, e.g. by composite G1-G4 methods
- Use precalculated $E_0(A_i)$

Approach 2: Isodesmic reaction

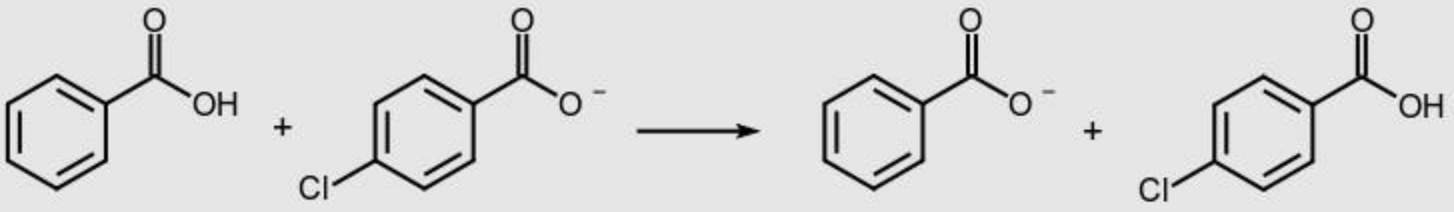
- $X = \sum_i \nu_i Y_i$, ν_i may be negative and rational
- $\Delta H(X) = \sum_i \nu_i \left[\Delta H(Y_i) - H(Y_i)\right] + H(X)$
- Take $\Delta H(Y_i)$ from reference tables
- Calculate $H(X)$ and $H(Y_i)$ with DFT-level accuracy

See also Accurate Thermochemistry for Large Molecules with Modern Density Functionals

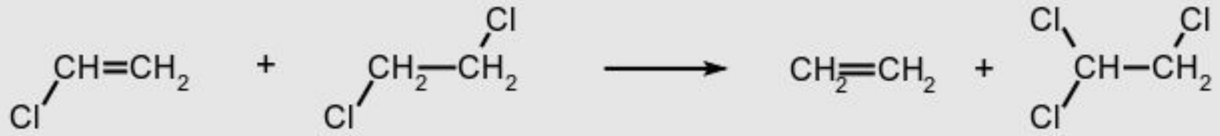
Example: isodesmic reactions



a)



b)



See also Error-cancelling balanced reactions

Individual studies:

- ***Reading.***

Required:

Additional: *Cramer (different chapters)*