Center for
Molecular Modeling

# Faster Potential Energy Surfaces: The *art* of Making Force Fields

*Toon Verstraelen*

Center for Molecular Modeling, Ghent University, Belgium

IPAM Workshop IV: Physical Frameworks for Sampling Chemical Compound Space

May 2011, Los Angeles

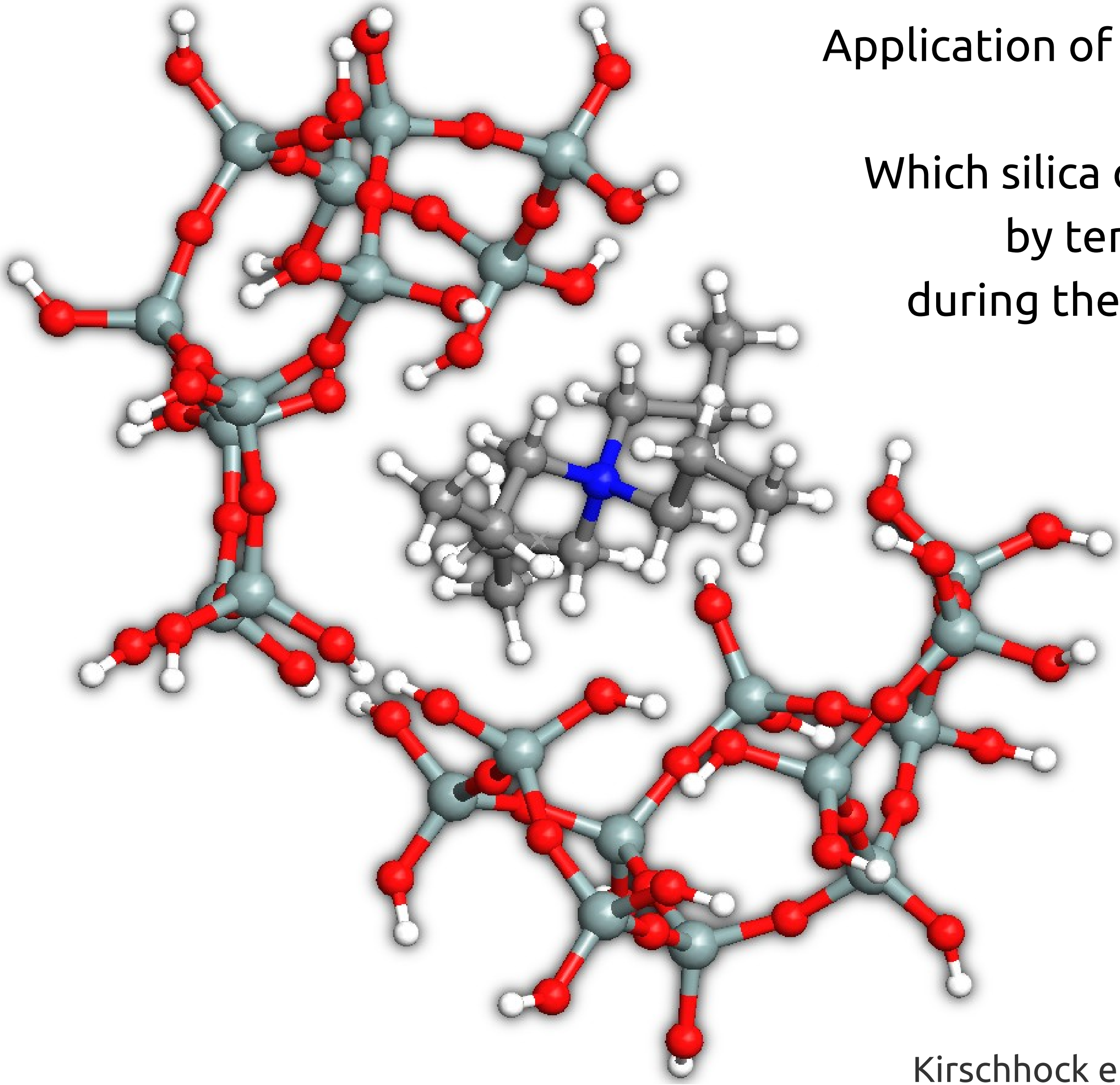
FACULTY OF ENGINEERING

UNIVERSITEIT
GENT

# Overview

| | |
|---|---|
| **Introduction** | Exploring Chemical Compounds<br>Force Field 101<br>Making Force Fields |
| **Atoms in Molecules** | Partitioning of Electron Density<br>Iterative Hirshfeld<br>ESP Fitted charges<br>Benchmarks |
| **The Hydrogen Bond** | AIM Application to FF development<br>Charge Transfer Matters! |
| **Split Charge Equilibration** | Partitioning of Polarization<br>Split Charges<br>Benchmarks |
| **Harmonic Models** | Force Constants & Rest Values |
| **Conclusion** | |

# Exploring Chemical Compounds



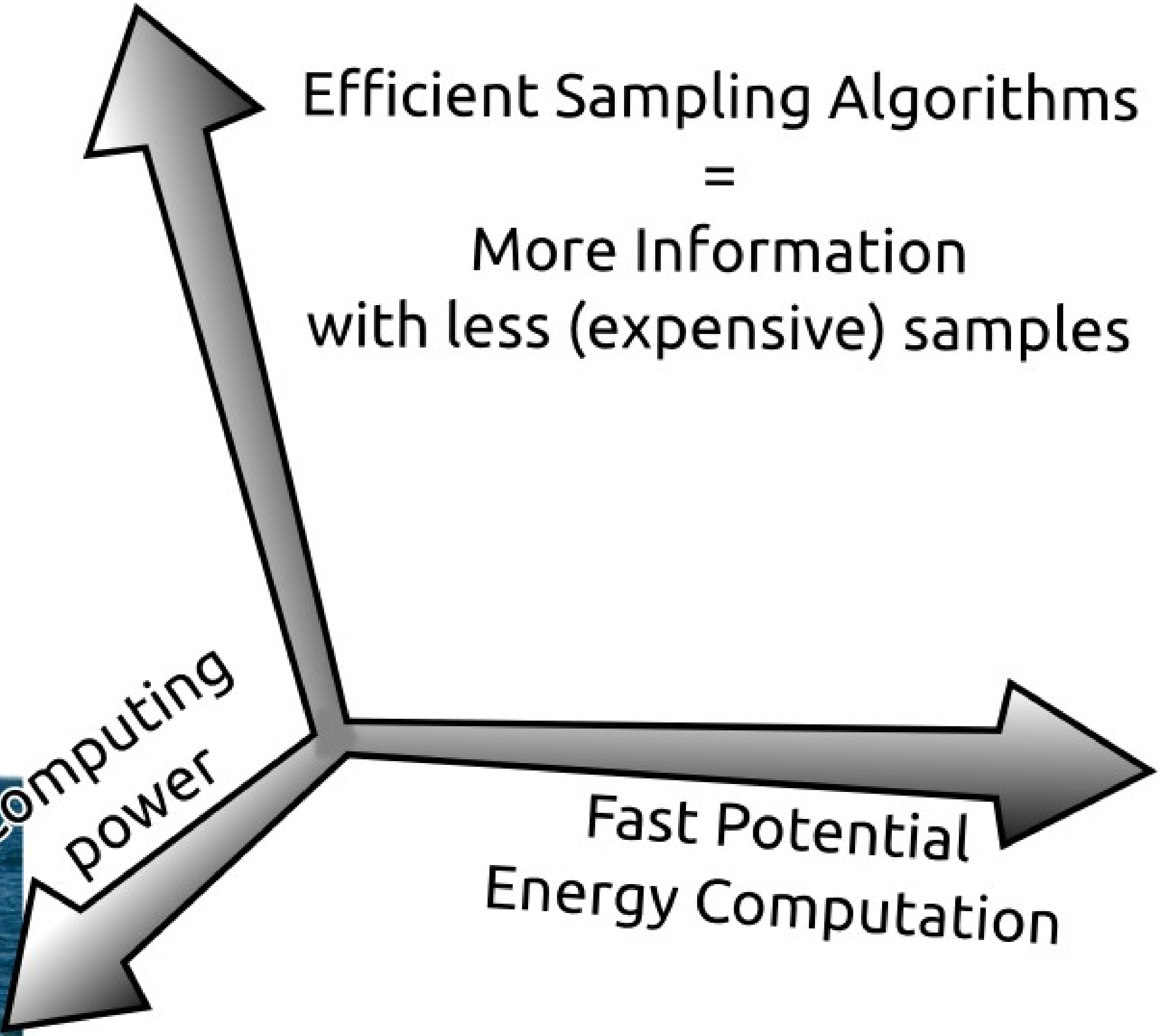
Application of interest: Zeolite synthesis

Which silica oligomers are promoted
by template molecules
during the prenucleation phase?

Work of Catlow & Lewis

Kirschhock et al. *Chem. Eur. J.* **2005**, 11, pp4306

# Exploring Chemical Compounds



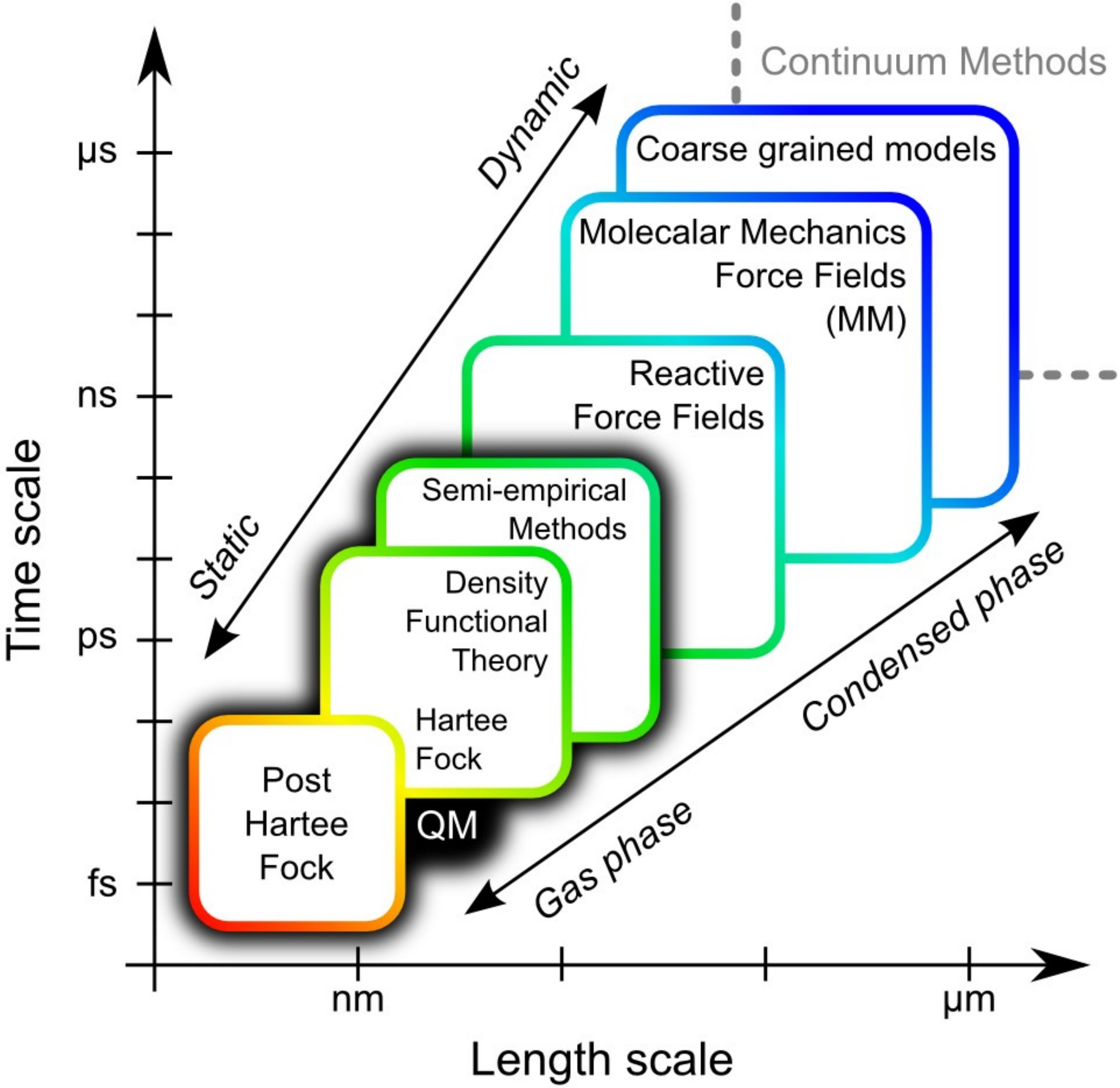
Efficient Sampling Algorithms
=
More Information
with less (expensive) samples

Computing power

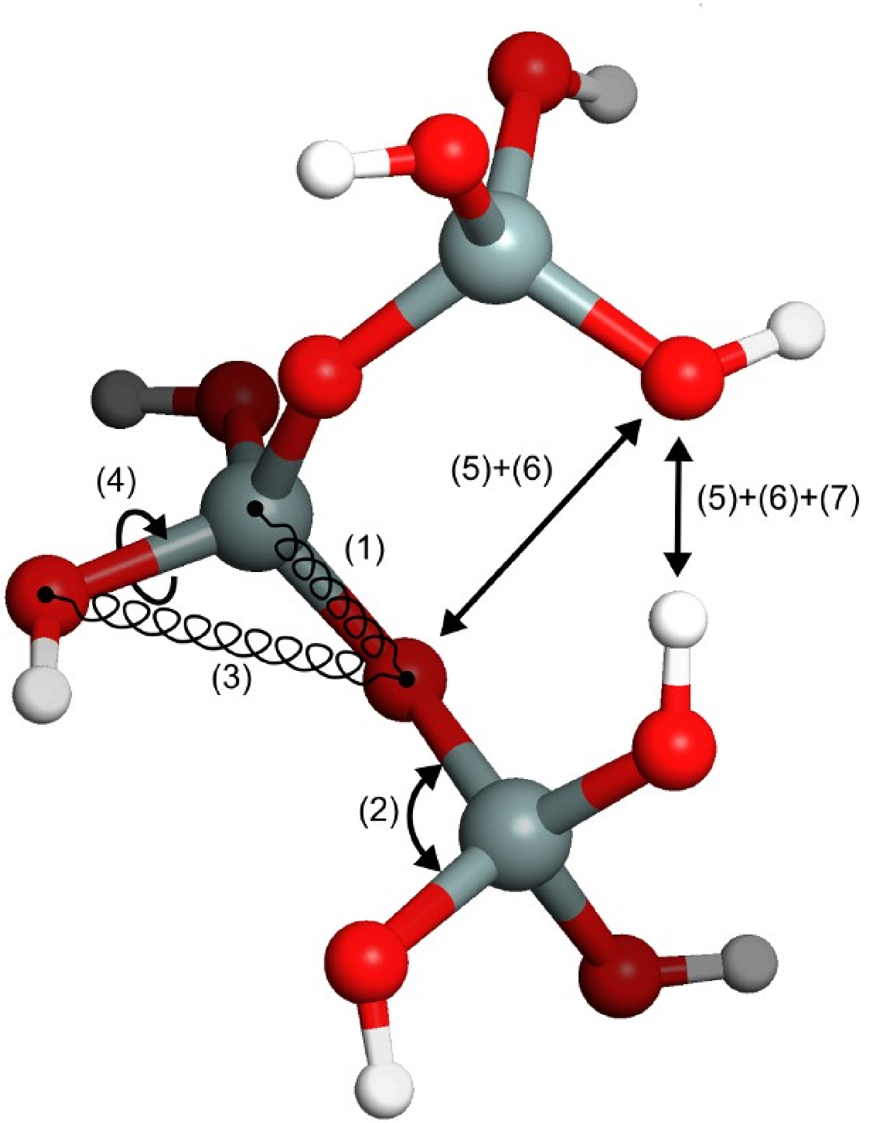
Fast Potential
Energy Computation

Moore's law

# Exploring Chemical Compounds



# Force Field (FF) 101



## (1,2,3,4) Valence Force Fields

$$\sum_k \frac{1}{2} K \left( q_k - q_k^{(0)} \right)^2$$

## (5) Electrostatics

$$\sum_{i>j} \frac{q_i q_j}{r_{ij}}$$

## (6) London Dispersion & Pauli Repulsion

$$\sum_{i>j} A_{ij}^{(p)} \exp\left( -\frac{r_{ij}}{B_{ij}^{(p)}} \right) - C_{ij}^{(6)} r_{ij}^{-6}$$

## (7) Hydrogen Bonding

$$\sum_{i>j} -A_{ij}^{(hb)} \exp\left( -\frac{r_{ij}}{B_{ij}^{(hb)}} \right)$$

# Exploring Chemical Compounds with FF's

## Chemical bonds are fixed during an FF simulation!

* Terms in the valence FF

* Exclusion rules for non-bonding interactions

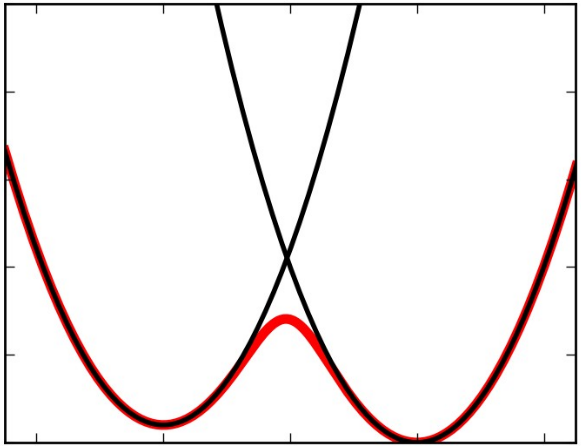
☛ Is chemical sampling possible?

### Reactive FF's

- Use empirical model for the bond order
- Computationally more expensive than traditional FF
- e.g. Tersoff, Brenner, ReaxFF, ...

### Enveloping Distribution Sampling

Christ et al, *JCP* **2007,** 126, p184110



$$-\frac{\log(\exp(-\beta V_1)+\exp(-\beta V_2))}{\beta}$$

### FF Interpolation

$$(1-\lambda)V_1(\mathbf{r}^N) \\ +\lambda V_2(\mathbf{r}^N)$$

(e.g. alchemical change)

Reference energy may be included in FF.

# Making FF's

**Old school approach:**

* Start from known parameters or make a guess
* Try and fix parameters repetitively until some target is reached

Drawbacks: labor intensive & boring

**Better approach:**

* Define an objective function that measures the FF quality
* Minimize the objective numerically

Drawbacks: no physical insight & rank deficiency

**Even better:**

* Associate FF parameters with QM 'observables'
  (where possible, avoid parameter fitting)
* Carefully designed cost functions

# Overview

| | |
|---|---|
| **Introduction** | Exploring Chemical Compounds<br>Force Field 101<br>Making Force Fields |
| **Atoms in Molecules** | Partitioning of Electron Density<br>Iterative Hirshfeld<br>ESP Fitted charges<br>Benchmarks |
| **The Hydrogen Bond** | AIM Application to FF development<br>Charge Transfer Matters! |
| **Split Charge Equilibration** | Partitioning of Polarization<br>Split Charges<br>Benchmarks |
| **Harmonic Models** | Force Constants & Rest Values |
| **Conclusion** | |

# Atoms In Molecules

**FF = Potential energy for atoms in molecules**

⇒ Obvious question: **"What are atoms in molecules?"**

Strict **definition is not possible**, but several *partitioning schemes* exist.

1955: Mulliken Population Analysis
1970: Löwdin Population Analysis
1977: Hirshfeld Partitioning
1985: Natural Population Analysis
1994: Bader Partitioning
2007: Iterative Hirshfeld Partitioning

**Properties that can be derived:**

* Atomic charges (& multipole expansion)
* Bond orders
* Pairwise electrostatic interactions
* Atomic spin density
* Overlap of atomic densities
* Condensed reactivity descriptors
* ...

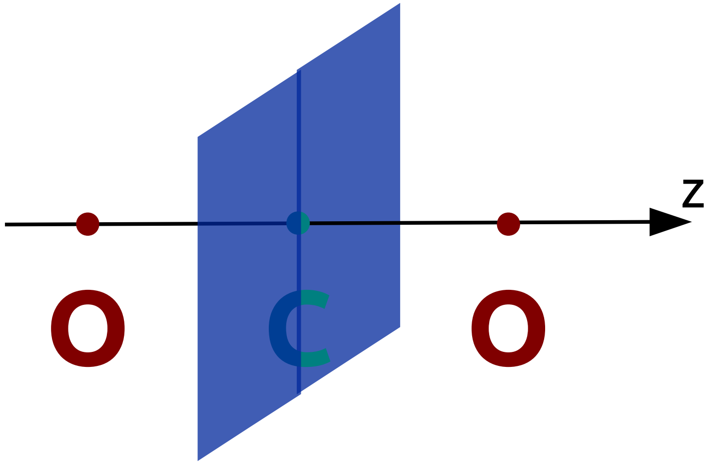
# Hirshfeld partitioning

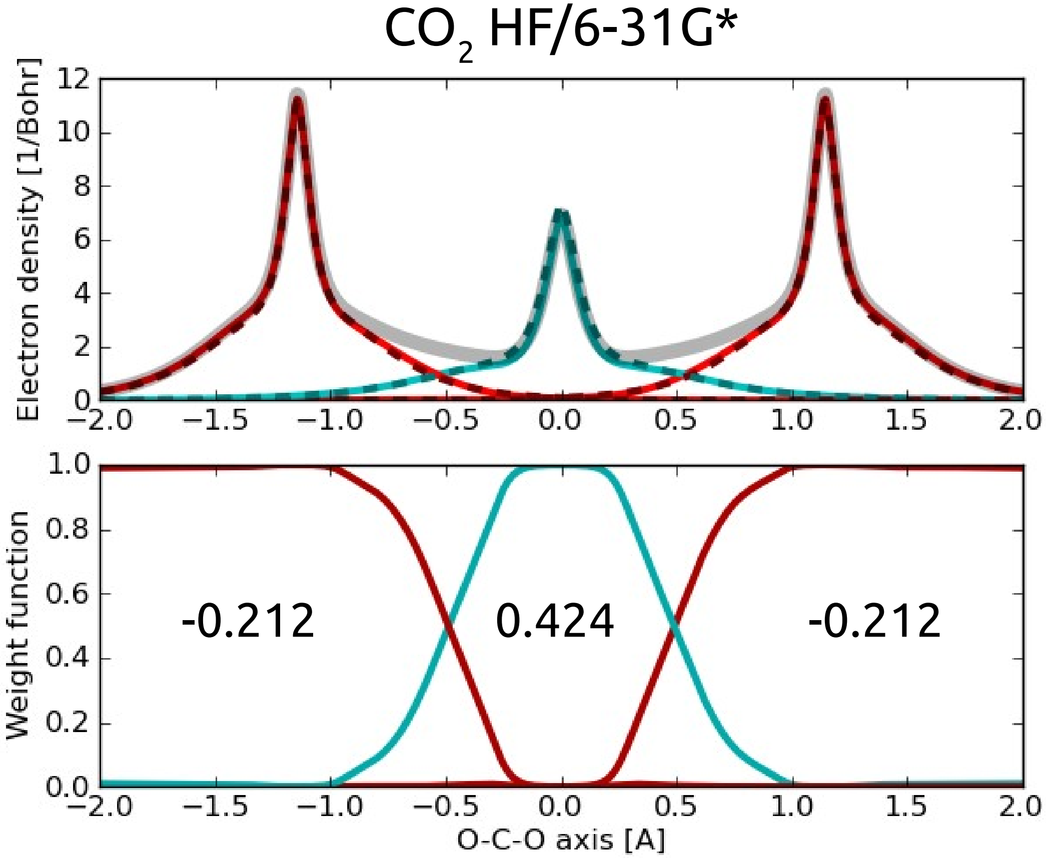
## Density partitioning in general:

$$\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho_{\mathrm{mol}}(\mathbf{r}) \qquad \sum_A w_A(\mathbf{r}) = 1 \qquad N_A = \int \rho_A(\mathbf{r})d\mathbf{r}$$

## Hirshfeld partitioning:

$$w_A(\mathbf{r}) = \frac{\rho_{0,A}(|\mathbf{r}-\mathbf{R}_A|)}{\sum_B \rho_{0,B}(|\mathbf{r}-\mathbf{R}_B|)}$$





Hirshfeld, *Theoret. Chim. Act.* **1977**, 44, p129

# Iterative Hirshfeld Partitioning

## Density partitioning in general:

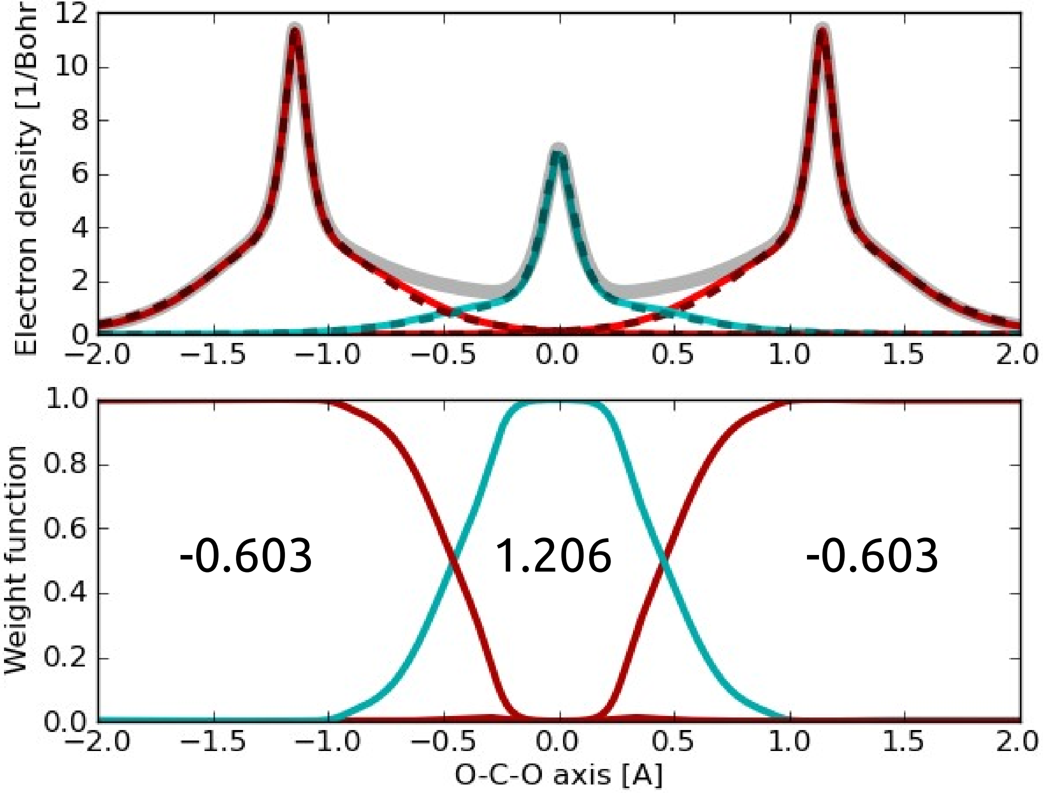
$$\rho_A(\mathbf{r}) = w_A(\mathbf{r})\rho_{\mathrm{mol}}(\mathbf{r}) \qquad \sum_A w_A(\mathbf{r}) = 1 \qquad N_A = \int \rho_A(\mathbf{r})d\mathbf{r}$$

## Hirshfeld-I partitioning:

$$w_A(\mathbf{r}) = \frac{\rho'_{0,A}(|\mathbf{r}-\mathbf{R}_A|, N_A)}{\sum_B \rho'_{0,B}(|\mathbf{r}-\mathbf{R}_B|, N_B)}$$

$$\begin{aligned}&\rho'_{0,A}(r, N_A) = \\ &(N_A - \lfloor N_A \rfloor)\rho'_{0,A}(r, \lfloor N_A \rfloor) + \\ &(\lceil N_A \rceil - N_A)\rho'_{0,A}(r, \lceil N_A \rceil)\end{aligned}$$

CO$_2$ HF/6-31G*



Bultinck et al., *JCP* **2007**, 126, p144111

# Other Properties

*Hirshfeld* | *Hirshfeld-I*

## Atomic multipole expansion [a.u.]

*Hirshfeld*

| | c | dz | qzz | ozzz |
|---|---|---|---|---|
| **O1** | -0.212 | -0.114 | 0.036 | 0.506 |
| **C2** | 0.424 | 0.0 | -1.160 | 0.0 |
| **O3** | -0.212 | 0.114 | 0.036 | |

*Hirshfeld-I*

| | c | dz | qzz | ozzz |
|---|---|---|---|---|
| **O1** | -0.603 | 0.217 | -0.001 | 0.219 |
| **C2** | 1.206 | 0.0 | -0.292 | 0.00 |
| **O3** | -0.603 | -0.217 | -0.001 | 0.0 |

## Pairwise atomic Coulomb interaction [$E_h$]

$$\iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}'$$

*Hirshfeld*

| | O1 | C2 |
|---|---|---|
| **C2** | -0.771 | |
| **O3** | 0.023 | -0.771 |

*Hirshfeld-I*

| | O1 | C2 |
|---|---|---|
| **C2** | -0.783 | |
| **O3** | 0.093 | -0.783 |

## Overlap of atomic densities [a.u.]

$$\int \rho_A(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r}$$

*Hirshfeld*

| | O1 | C2 |
|---|---|---|
| **C2** | 0.253 | |
| **O3** | 0.002 | 0.253 |

*Hirshfeld-I*

| | O1 | C2 |
|---|---|---|
| **C2** | 0.214 | |
| **O3** | 0.004 | 0.214 |

## Bond orders

$$\iint [\rho^{(2)}(\mathbf{r},\mathbf{r}') - \rho(\mathbf{r})\rho(\mathbf{r}')]w_A(\mathbf{r})w_B(\mathbf{r}')d\mathbf{r}d\mathbf{r}'$$

*Hirshfeld*

| | O1 | C2 |
|---|---|---|
| **C2** | 2.46 | |
| **O3** | 0.30 | 2.46 |

*Hirshfeld-I*

| | O1 | C2 |
|---|---|---|
| **C2** | 2.08 | |
| **O3** | 0.39 | 2.08 |

# Why Hirshfeld-I?

**Pragmatic reasons (benchmarks on large sets of molecules)**

- Charges reproduce the ESP
  - some small molecules (like $CO_2$) are exceptions
  - 10% to 20% error on ESP with charges
  - 5% to 10% error on ESP with charges and dipoles
- Not sensitive to basis set (e.g. diffuse functions)
- Robust with respect to conformational changes
- Does not suffer from buried atom problem (like ESP charges)
- 'Easy' to compute

**Conceptual reason**

- Hirshfeld-I atoms are very similar to atoms in vacuum

**Don't get too enthusiastic!**
**Hirshfeld-I is still just a scheme.**

Bultinck et al, *CPL* **2007**, 444, p205 Van Damme et al, *JCTC* **2009**, 5, p334 Catak et al, *JOC* **2010**, 75, p4530

# ESP Fitted charges

**ESP Fitted charges minimize:**

$$\chi^2 = \sum_{g=1}^{N_g} w_g \left( V_{\mathrm{AI}}(\mathbf{r}_g) - \sum_{a=1}^{N_a} \frac{q_a}{\|\mathbf{r}_a - \mathbf{r}_g\|} \right)^2 + \text{some penalty}$$

Often used for FF development (e.g. RESP for AMBER)

but ...

1) Buried charges are ill-defined, sensitive to choice of grid
2) Sensitive to conformational changes

because ...

1) Ill-conditioned least-squares system
2) Atomic multipole expansions are truncated after the monopole.
   This error is compensated by overfitting the charges.

**This is also just a scheme,**
**i.e. one to partition the ESP.**

Francl et al, *Rev. Comput. Chem.* **2000**, 14, p1

# Penta Alanine Benchmark

## Computational details

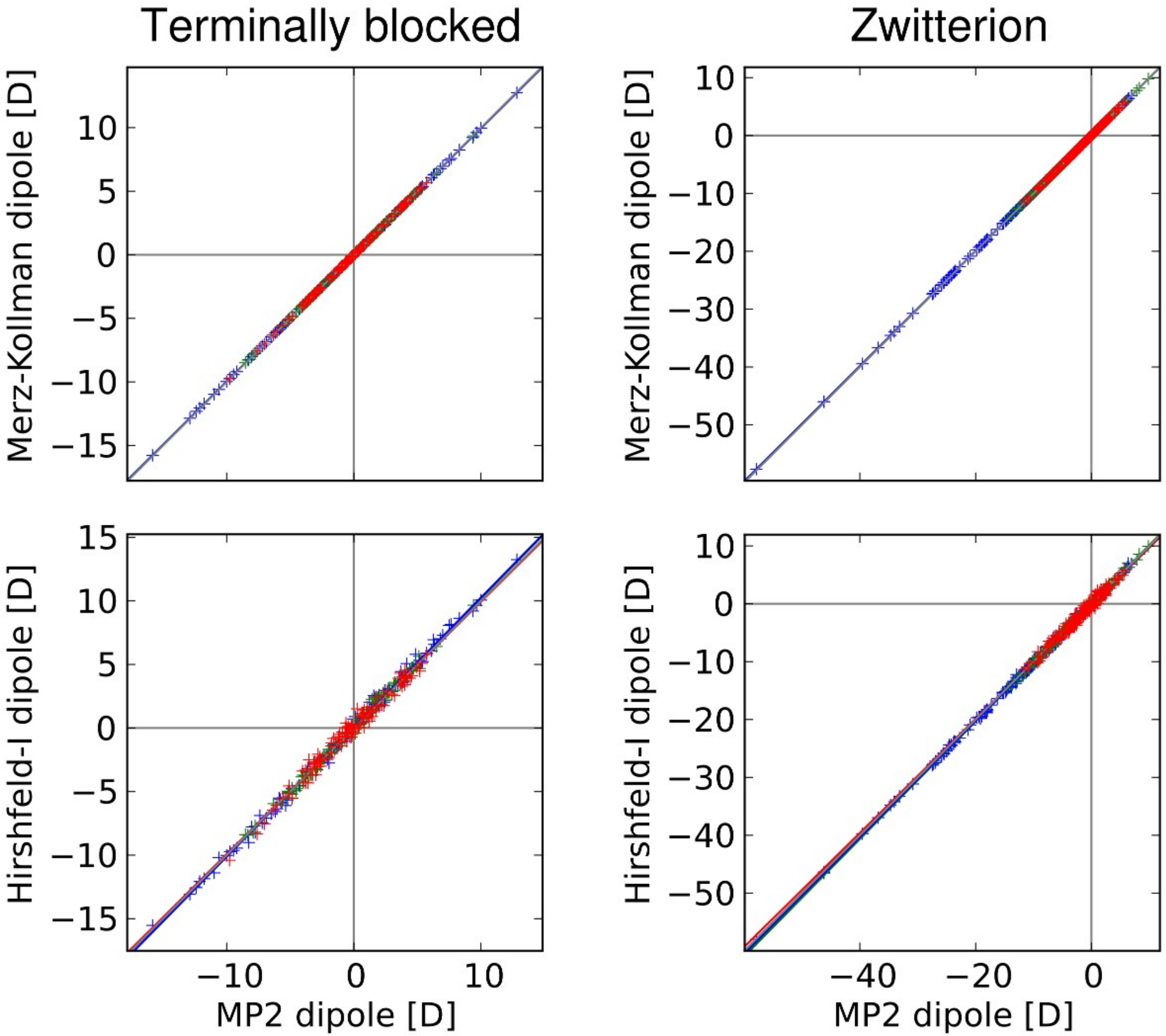
1) Random penta alanine conformers
    * 103 Terminally blocked
    * 134 Zwitterionic
2) B3LYP/6-31Gd geometry optimizations
3) MP2/Aug-cc-pVTZ single point computations
4) Charges derived with a selection of schemes:
    * Merz-Kollman
    * Hirhsfeld-I

## Benchmarks

1) Comparison of MP2 dipole with charge-derived dipole
2) Sensitivity of charges to conformational changes

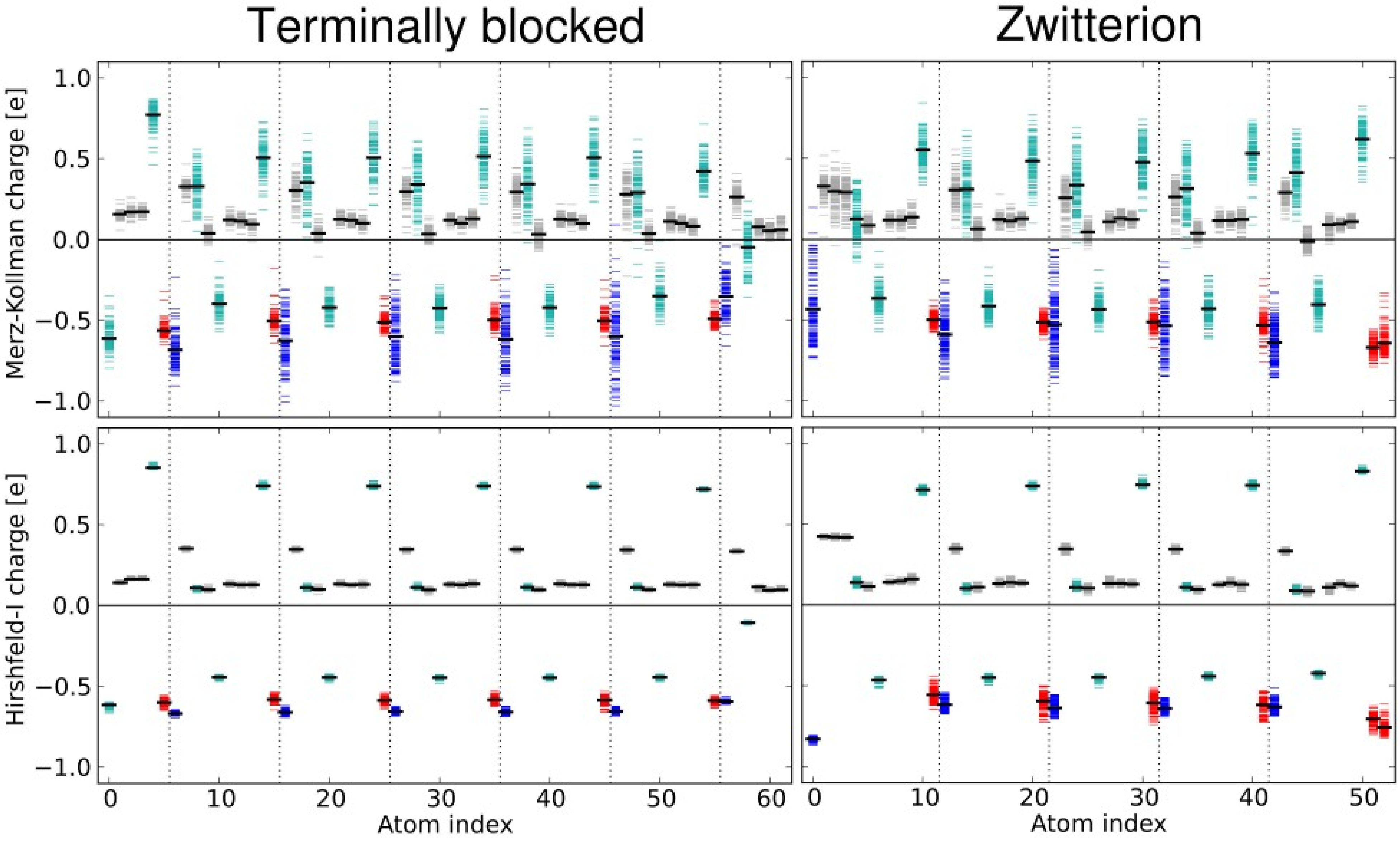
(paper in preparation)

# Penta Alanine Benchmark



(paper in preparation)

# Penta Alanine Benchmark



(paper in preparation)

# Hirshfeld-I limitations

**One limitation observed so far ...**

Poor point charge ESP's in very polar systems:

- Zeolites
- Oxide part in metal-organic frameworks

**Hirshfeld-I is still young, and needs more testing**

(paper in preparation)

# Overview

| | |
|---|---|
| **Introduction** | Exploring Chemical Compounds<br>Force Field 101<br>Making Force Fields |
| **Atoms in Molecules** | Partitioning of Electron Density<br>Iterative Hirshfeld<br>ESP Fitted charges<br>Benchmarks |
| **The Hydrogen Bond** | AIM Application to FF development<br>Charge Transfer Matters! |
| **Split Charge Equilibration** | Partitioning of Polarization<br>Split Charges<br>Benchmarks |
| **Harmonic Models** | Force Constants & Rest Values |
| **Conclusion** | |

# Hirshfeld-I & FF development

$$q_A, d_{A,x}, d_{A,y}, d_{A,z}, \ldots$$

$$\iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'$$

**Atomic densities (or weight functions) provide...**

1) Breakdown of the electrostatic interactions
    - *MM charges*
    - *Polarizable FF's (charge equilibration, inducible dipoles)*
    - *Intermolecular charge-transfer*

$$\int \rho_A(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r}$$

2) Pauli repulsion ~ overlap of (singlet state) densities
    - *Overlap of partitioned atoms -> pairwise potentials*

$$\iint [\rho^{(2)}(\mathbf{r},\mathbf{r}') - \rho(\mathbf{r})\rho(\mathbf{r}')] w_A(\mathbf{r}) w_B(\mathbf{r}') d\mathbf{r}d\mathbf{r}'$$

3) Bond order in dimers
    - *Pairwise breakdown of weak covalent bonding*

4) Partitioning of linear response
    - *Changes in charges/dipoles due to external field*
        *=> Second order parameters in polarizable FF*

cfr. WOFF: Rotenberg, B. et al *PRL* **2010**, 104, 138301

# Water dimer

**Two scans of the water dimer**

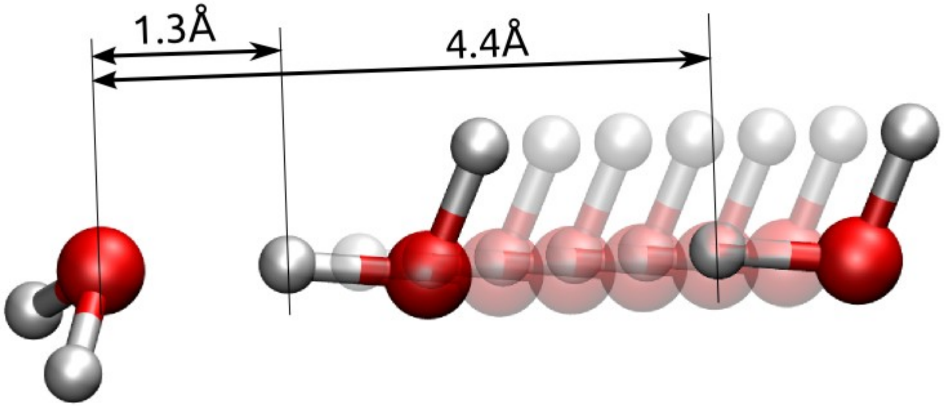
- Both go through optimal geometry
- PBE/DGTZVP
- cp. cor. Interaction energies

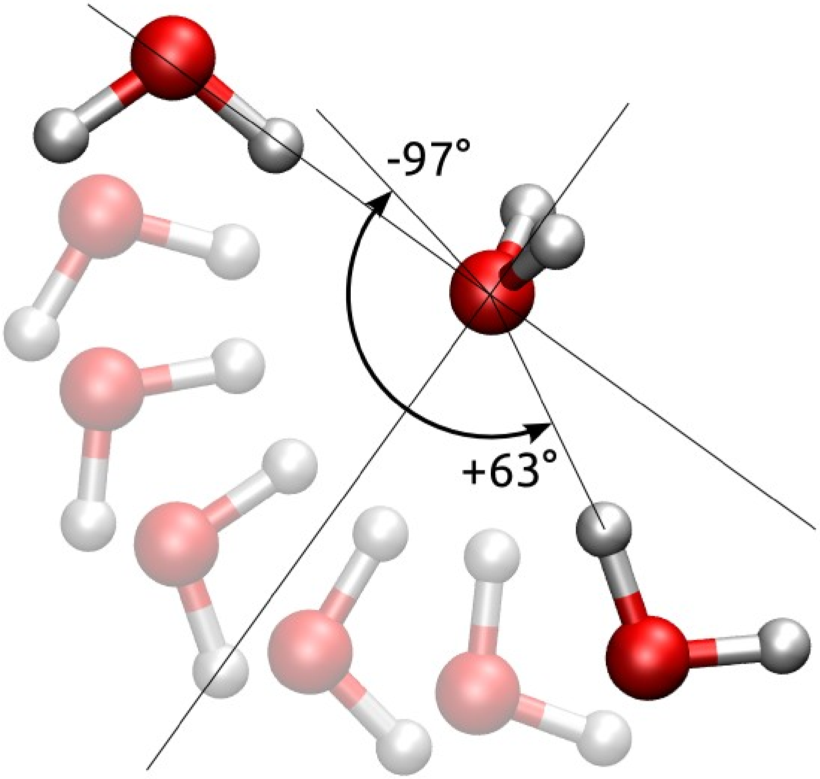
**AIM properties computed at each point**

- Charges

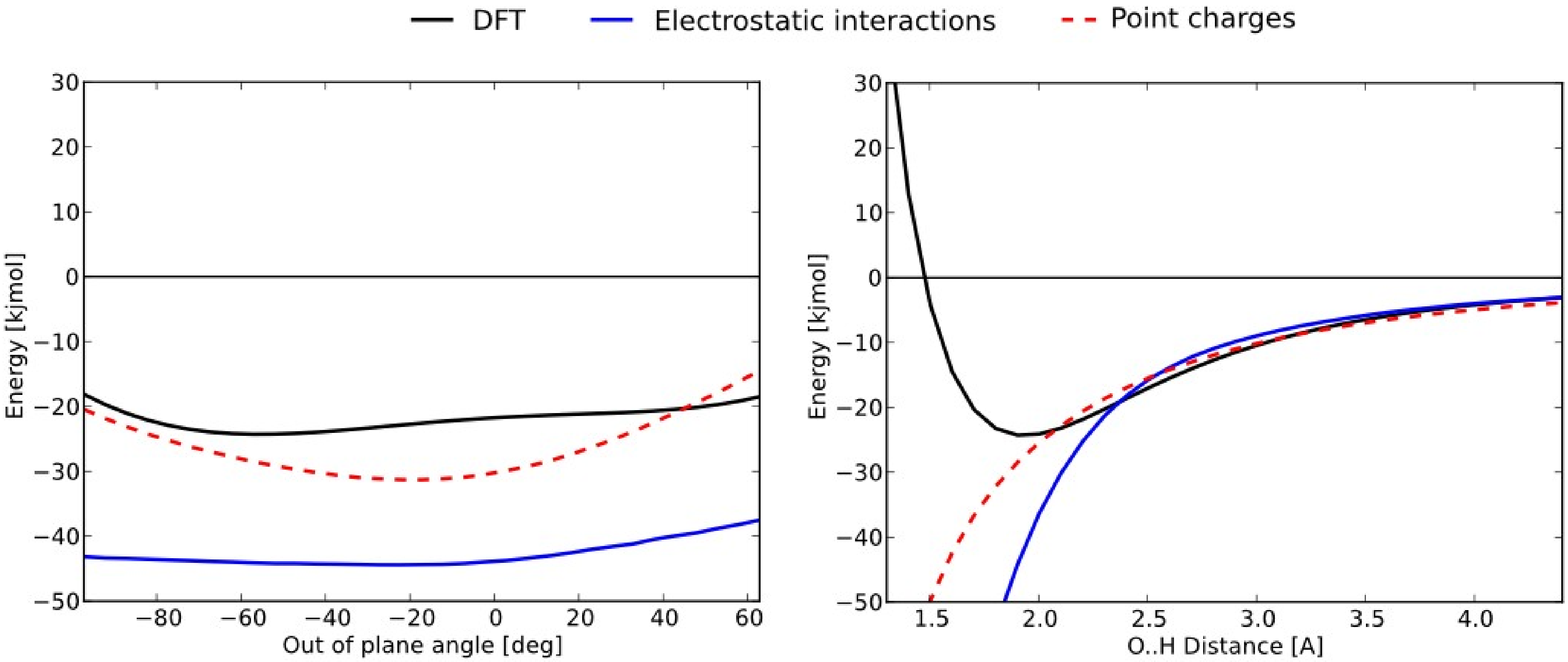
**AIM pairwise properties**

- Electrostatic interaction
- Bond orders
- Pauli Overlap





# Electrostatic Interactions

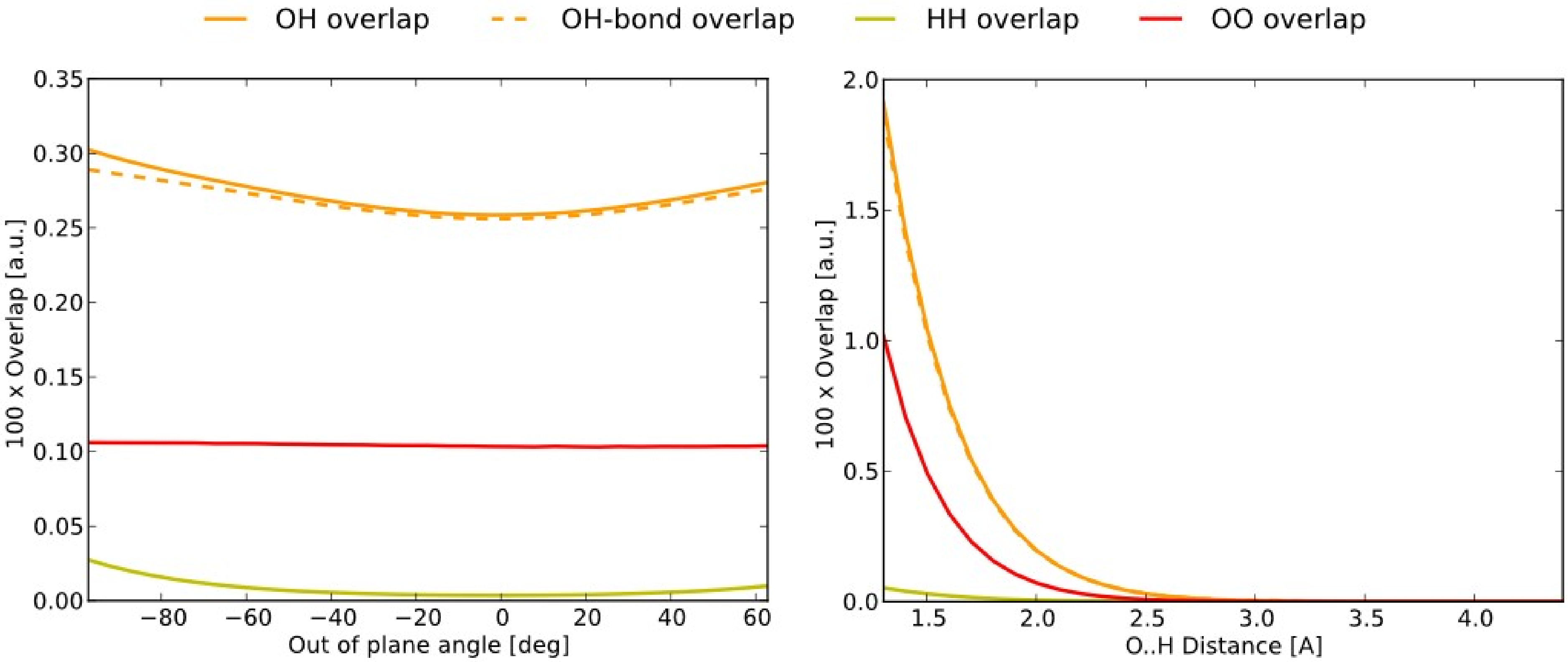


* Electrostatic interactions OK at long distances

* Hirshfeld-I dipoles (& multipoles) are needed at short distance.

* Subtle angular variations are missing

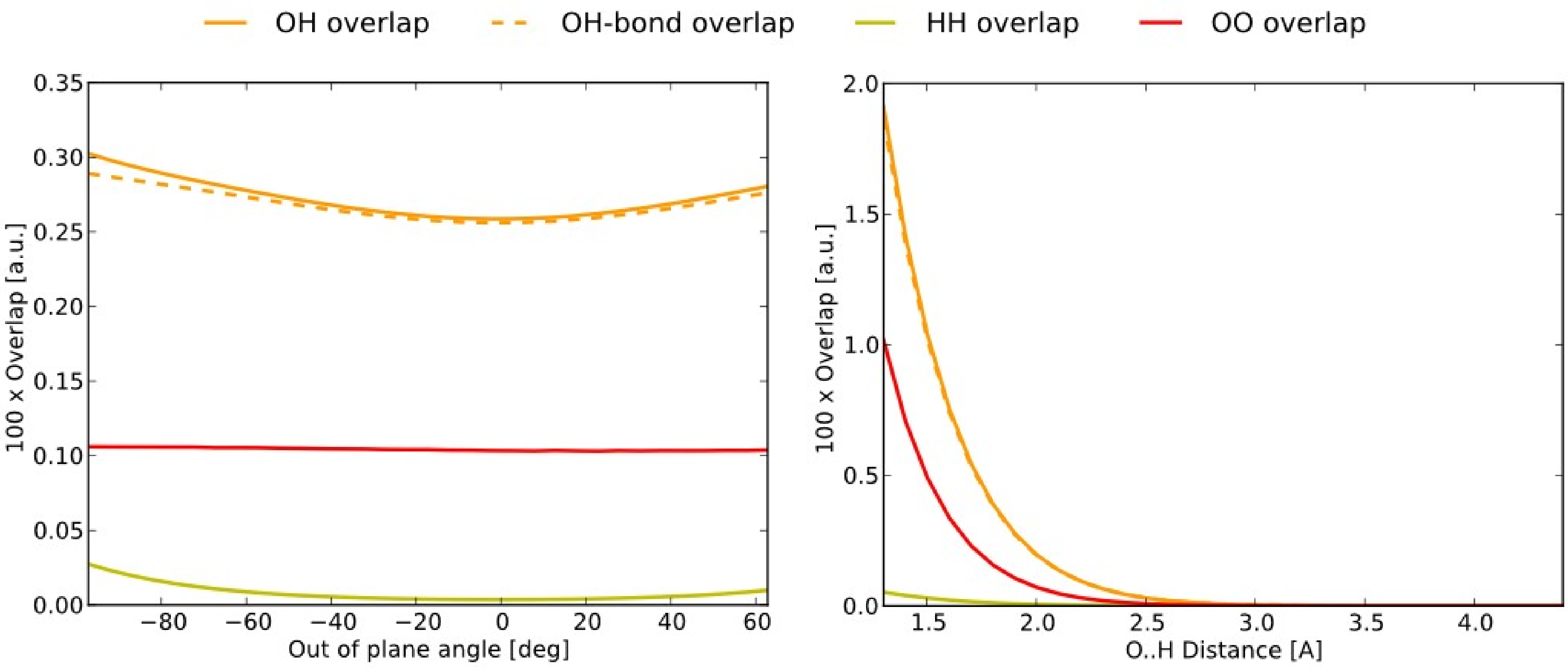
# Pauli Repulsion



$$O_{ij} = \int \rho_i(\mathbf{r})\rho_j(\mathbf{r})d\mathbf{r}$$

XY overlap = sum of overlap integrals over all pairs X and Y
OH-bond overlap = overlap integral for O and H in the HBond only

# Pauli Repulsion

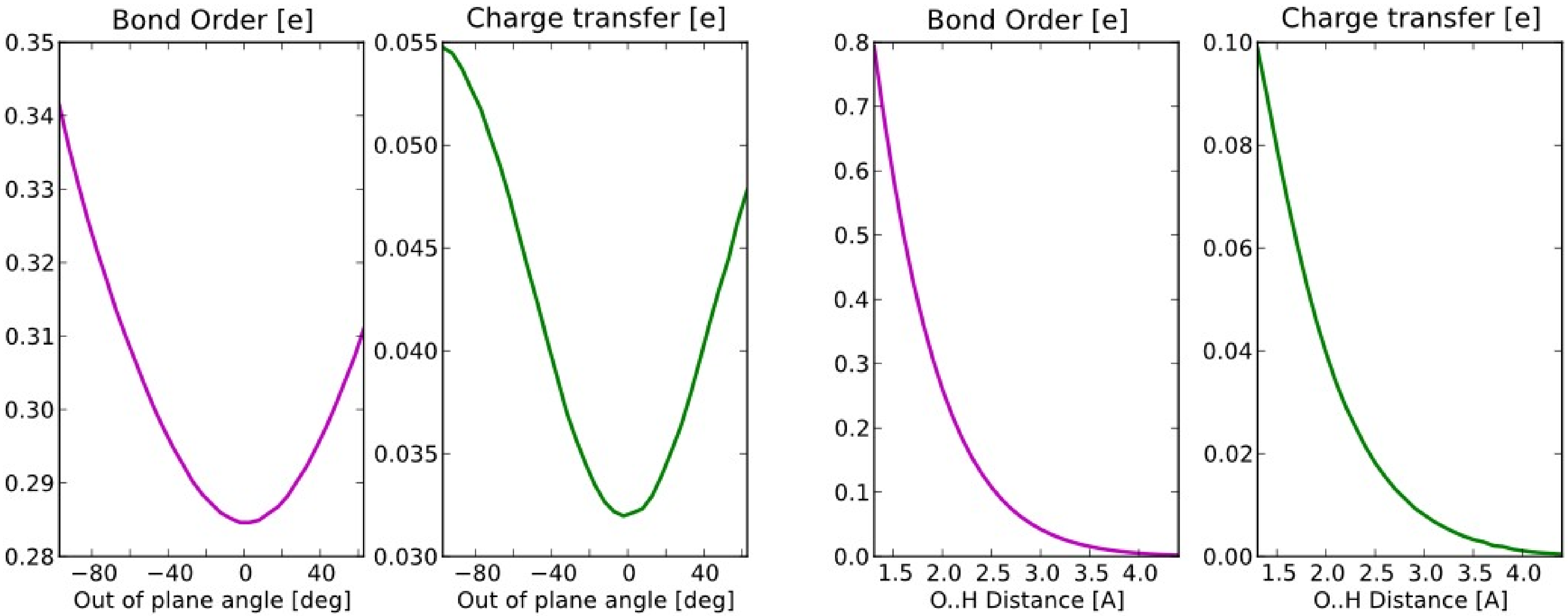


* No effect of lone pairs

* A*exp(-B*R) is a good approximation

* OH pair in hydrogen bond has angular overlap dependence

# Bond order & charge transfer



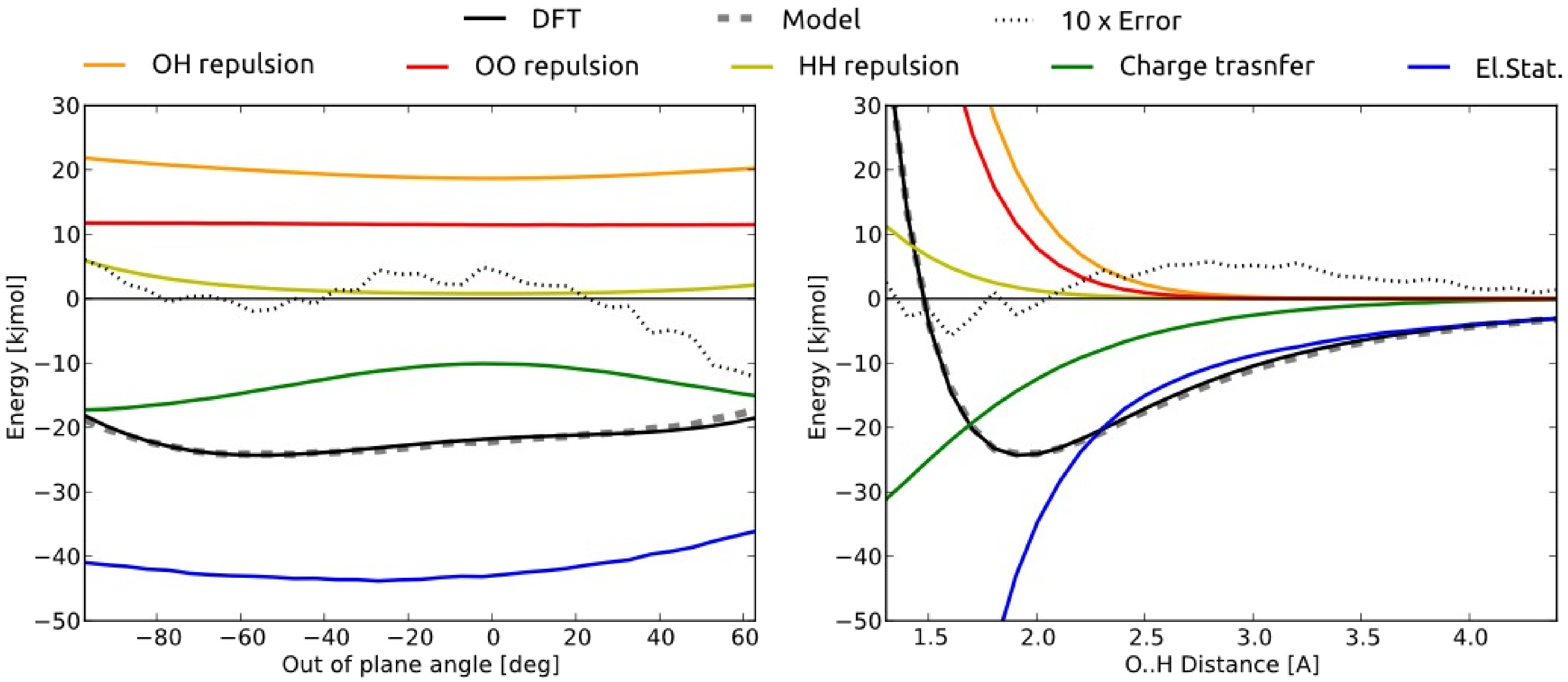
* Electrons go from acceptor $H_2O$ to donor $H_2O$ during HBond formation

* Charge transfer and Bond order are similar but not identical

* OH overlap ~ (Bond order)$^2$

* Large relative fluctuations in charge transfer

# Model for the interaction energy



$$E_{\mathrm{DFT}} \approx A_1 O_{\mathrm{OH}} + A_2 O_{\mathrm{HH}} + A_3 O_{\mathrm{OO}} - A_4 Q_{\mathrm{trans}} + E_{\mathrm{el.stat.}}$$

in atomic units: $A_1 = 2.75$ $A_2 = 8.22$ $A_3 = 4.22$ $A_4 = 0.12$

* Insightful, but still using AIM data from DFT computations.

* Pairwise repulsion parameters are trivial to extract.

* Explicit exchange term (~ bond order) is not helpful.

* Should be combined with EDA (Morokuma) or SAPT

# Overview

| | |
|---|---|
| **Introduction** | Exploring Chemical Compounds<br>Force Field 101<br>Making Force Fields |
| **Atoms in Molecules** | Partitioning of Electron Density<br>Iterative Hirshfeld<br>ESP Fitted charges<br>Benchmarks |
| **The Hydrogen Bond** | AIM Application to FF development<br>Charge Transfer Matters! |
| **Split Charge Equilibration** | Partitioning of Polarization<br>Split Charges<br>Benchmarks |
| **Harmonic Models** | Force Constants & Rest Values |
| **Conclusion** | |

# Dipole Polarizability

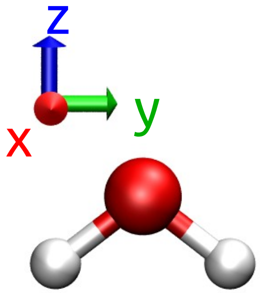
**Definition:**

$$P_{\alpha\beta} = \frac{\partial d_\alpha^{(mol)}}{\partial E_\beta}$$

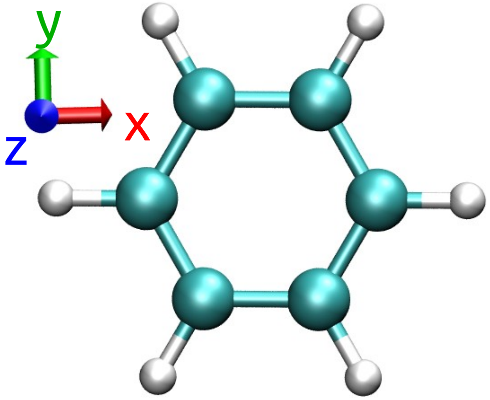
A few examples computed with PBE/DGTZVP:

- water
- benzene
- alaninedipeptide

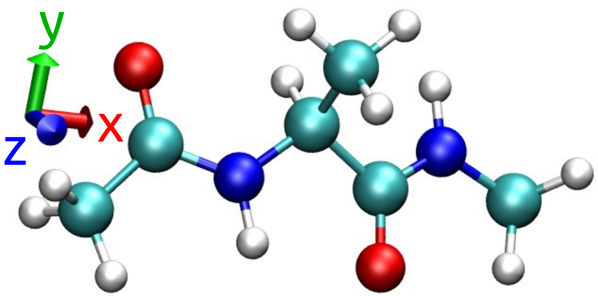
(values in a.u.)



```
[   4.28    0        0    ]
[   0       8.19     0    ]
[   0       0        6.78]
```



```
[ 77.38    0        0    ]
[  0       77.38    0    ]
[  0       0       34.15]
```



```
[ 122.8       4.19    -1.41]
[    4.19    89.57    -5.27]
[   -1.41    -5.27    73.80]
```

# Polarizable Force Fields

**Definition**

Empirical model for the linear response
(of the e⁻ density to a change in the external field)

**Basic mathematical form is quadratic**

q = vector with N atomic charges

d = vector with 3 N atomic dipole components

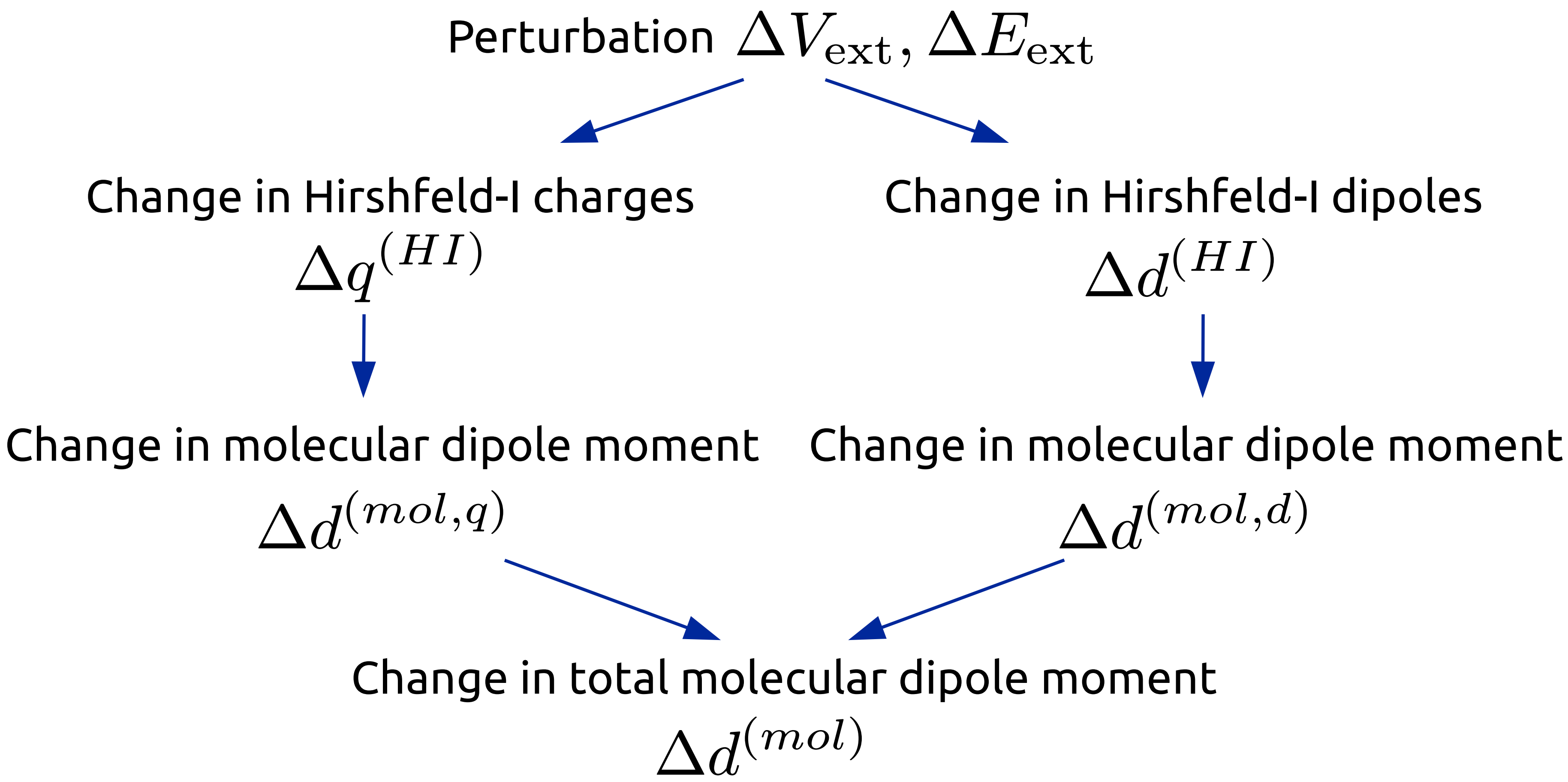
$$E = \tfrac{1}{2} q^T J^{(\mathrm{qq})} q + q^T J^{(\mathrm{qd})} d + \tfrac{1}{2} d^T J^{(\mathrm{dd})} d$$

$$+ q^T x^{(q)} + d^T x^{(d)}$$

$$+ q^T \Delta V_{\mathrm{ext}} + d^T \Delta E_{\mathrm{ext}}$$

e.g. $J_{ij}^{(qq)} = \dfrac{1}{r_{ij}}$

# Partitioning of Induced Dipole Moment



Perturbation $\Delta V_{\text{ext}}, \Delta E_{\text{ext}}$

Change in Hirshfeld-I charges

$\Delta q^{(HI)}$

Change in Hirshfeld-I dipoles

$\Delta d^{(HI)}$

Change in molecular dipole moment

$\Delta d^{(mol,q)}$

Change in molecular dipole moment

$\Delta d^{(mol,d)}$

Change in total molecular dipole moment

$\Delta d^{(mol)}$

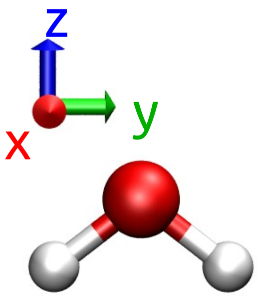
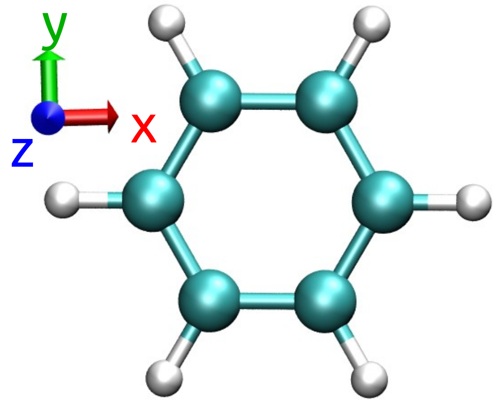
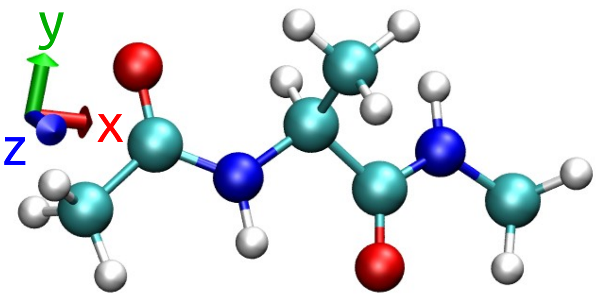
**Do this with uniform X, Y and Z fields:**

=> atomic charge contribution to $P_{\alpha\beta}$

=> atomic dipole contribution to $P_{\alpha\beta}$

Krishtal et al, *JCP* **2006**, 125, p034312.

# Partitioning of Induced Dipole Moment

|  | Charge contribution to $P_{\alpha\beta}$ | Dipole contribution to $P_{\alpha\beta}$ |
|---|---|---|
|  | $\begin{bmatrix} \mathbf{0} & 0 & 0 \\ 0 & \mathbf{6.8} & 0 \\ 0 & 0 & \mathbf{4.92} \end{bmatrix}$ | $\begin{bmatrix} \mathbf{4.28} & 0 & 0 \\ 0 & \mathbf{1.39} & 0 \\ 0 & 0 & \mathbf{1.86} \end{bmatrix}$ |
|  | $\begin{bmatrix} \mathbf{70.13} & -0.03 & 0 \\ 0.34 & \mathbf{70.38} & 0 \\ 0.05 & 0.51 & \mathbf{0} \end{bmatrix}$ | $\begin{bmatrix} \mathbf{7.14} & 0.01 & 0.03 \\ -0.45 & \mathbf{7.1} & -0.05 \\ -0.03 & -0.54 & \mathbf{34.16} \end{bmatrix}$ |
|  | $\begin{bmatrix} \mathbf{101.7} & 2.63 & -0.14 \\ 2.63 & \mathbf{69.05} & -12.97 \\ -0.17 & -12.9 & \mathbf{45.07} \end{bmatrix}$ | $\begin{bmatrix} \mathbf{21.68} & 1.61 & -1.22 \\ 0.97 & \mathbf{20.58} & 7.73 \\ -1.04 & 7.62 & \mathbf{28.78} \end{bmatrix}$ |

# Electronegativity Equalization Method (EEM)

## Mathematical form for the 'Electronic energy'

$$E_{\mathrm{EEM}} = \left(\sum_i \chi_i q_i + \frac{1}{2}\eta_i q_i^2\right) + \left(\frac{1}{2}\sum_{i,j\neq i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}\right) + \left(\sum_i V_{\mathrm{ext}}(\mathbf{r}_i) q_i\right)$$

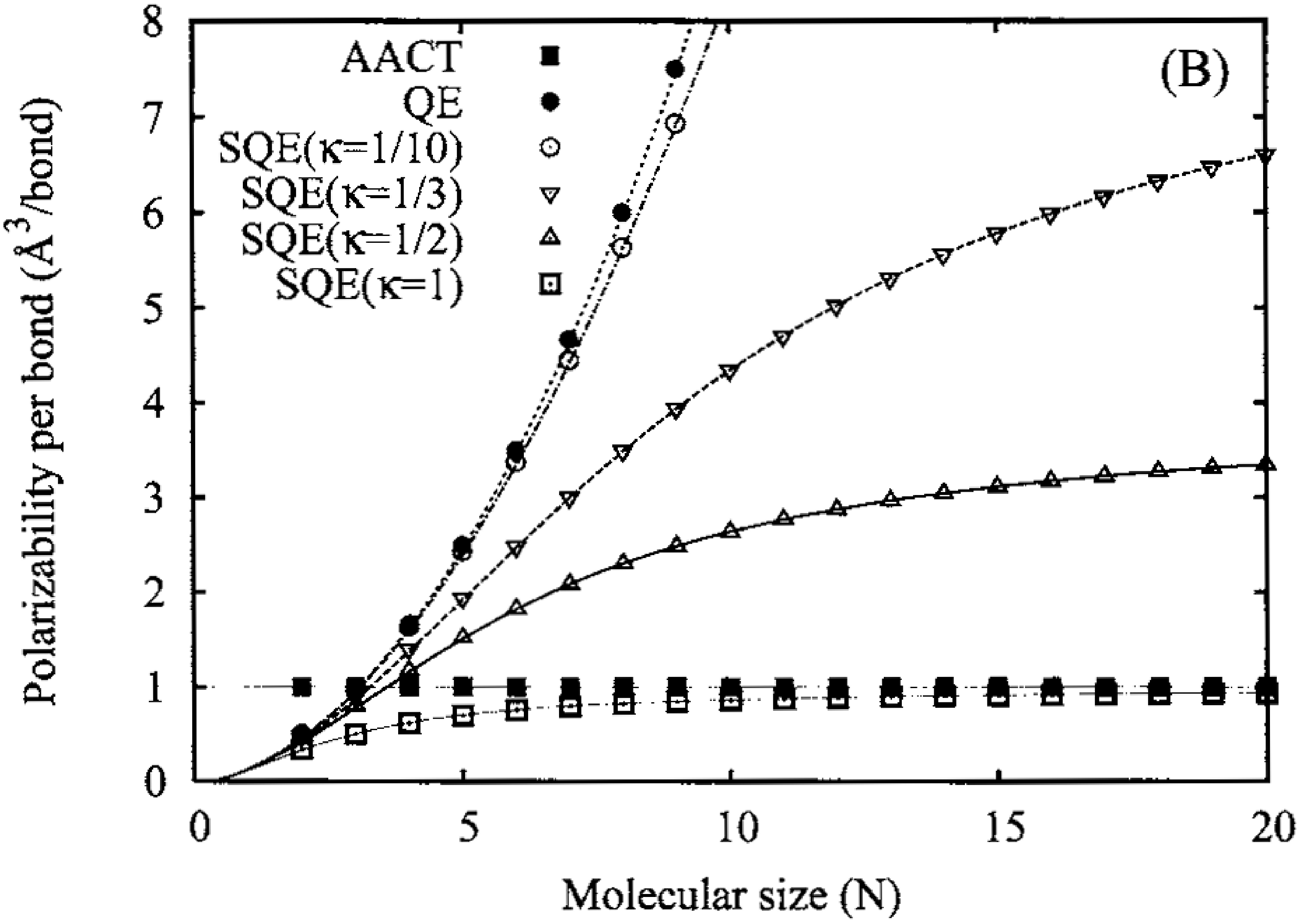
Variables: $q_i$
Parameters: $\chi_i, \eta_i, \bar{r}_i$

## Charges minimize $E_{\mathrm{EEM}}$ (with total charge constraint)

$$\frac{\partial E_{\mathrm{EEM}}}{\partial q_i} = \sum_i \left(\chi_i + V_{\mathrm{ext}}(\mathbf{r}) + \eta_i q_i + \sum_{j\neq i} \frac{q_j}{|\mathbf{r}_i - \mathbf{r}_j|}\right) = \chi_{\mathrm{mol}}$$

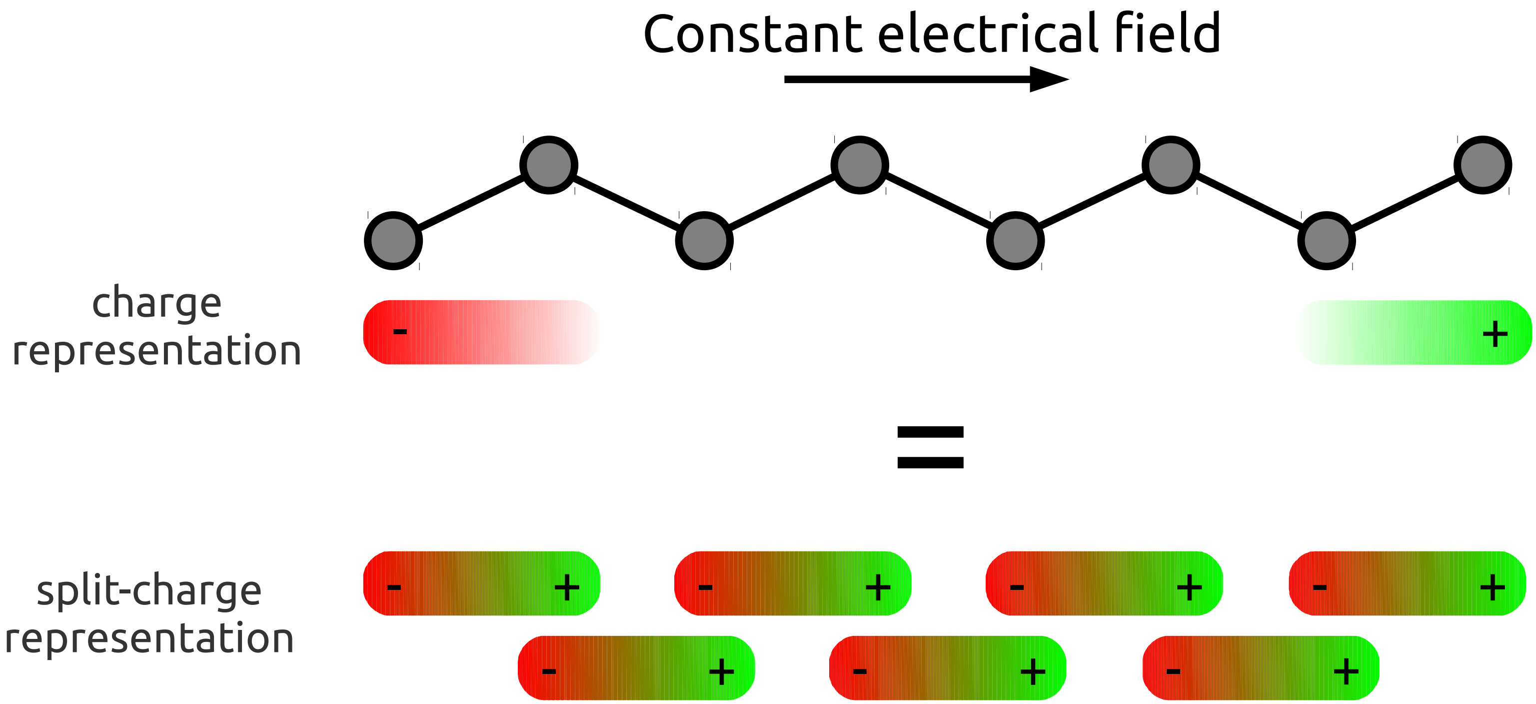
Mortier, W.; Ghosh, S.; Shankar, S. *JACS* **1986**, 108, 4315-4320.

# Superlinear Scaling of Polarizability



Warren et al, *JCP* **2008**, 128, p144110

# Split Charge Equilibration



Constant electrical field

charge representation

split-charge representation

$$\begin{bmatrix} q_1 \\ \vdots \\ q_{N^a} \end{bmatrix} = \begin{bmatrix} +1 & 0 & \cdots \\ -1 & +1 & \\ 0 & -1 & \\ \vdots & & \ddots \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_{N^b} \end{bmatrix}$$

$$q = Tp$$

$$(q = Tp + q_0)$$

# Split Charge Equilibration

## Electronic energy (EEM)

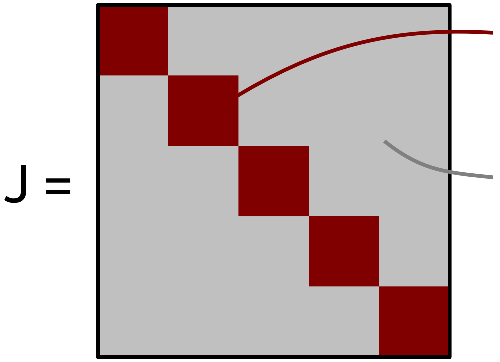
$$E_{\mathrm{EEM}} = \frac{1}{2} q^T J q + x^T q \quad \text{with} \quad \sum_i q_i = 0$$

$$J_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad \text{when} \quad i \neq j$$

$$J_{ii} = \eta_i$$

$$x_i = \chi_i + V_{\mathrm{ext}}(\mathbf{r}_i)$$

## Second order contributions



hardness parameters

coulomb interaction
(e.g. 1/$r_{ij}$, or smeared charges)

# Split Charge Equilibration

## Electronic energy (EEM)

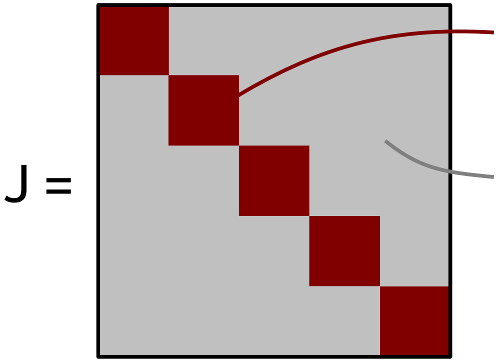
$$E_{\mathrm{EEM}} = \frac{1}{2} p^T T^T J T p + x^T T p \quad \text{with} \quad q = Tp$$

$$J_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad \text{when} \quad i \neq j$$

$$J_{ii} = \eta_i$$

$$x_i = \chi_i + V_{\mathrm{ext}}(\mathbf{r}_i)$$

## Second order contributions



hardness parameters

coulomb interaction (e.g. $1/r_{ij}$, or smeared charges)

## Electronic energy (SQE)

$$E_{\text{SQE}} = \frac{1}{2} p^T T^T J T p + \frac{1}{2} p^T J' p + x^T T p \quad \text{with} \quad q = Tp$$

$$J_{ij} = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad \text{when} \quad i \neq j \qquad J'_{kl} = \delta_{kl} \kappa_k$$

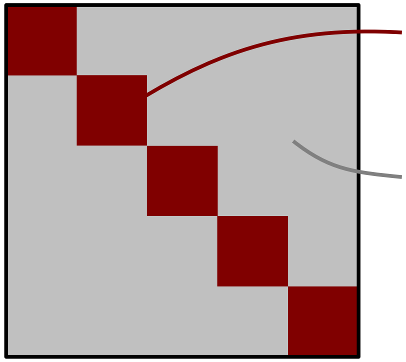
bond hardness

$$J_{ii} = \eta_i$$

$$x_i = \chi_i + V_{\text{ext}}(\mathbf{r}_i)$$

## Second order contributions

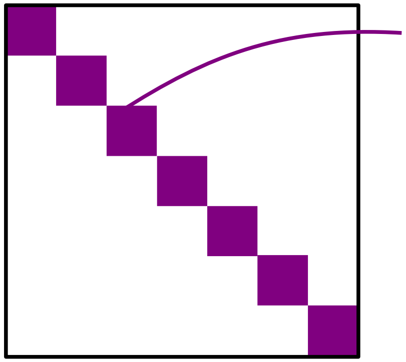
J =



hardness parameters

coulomb interaction

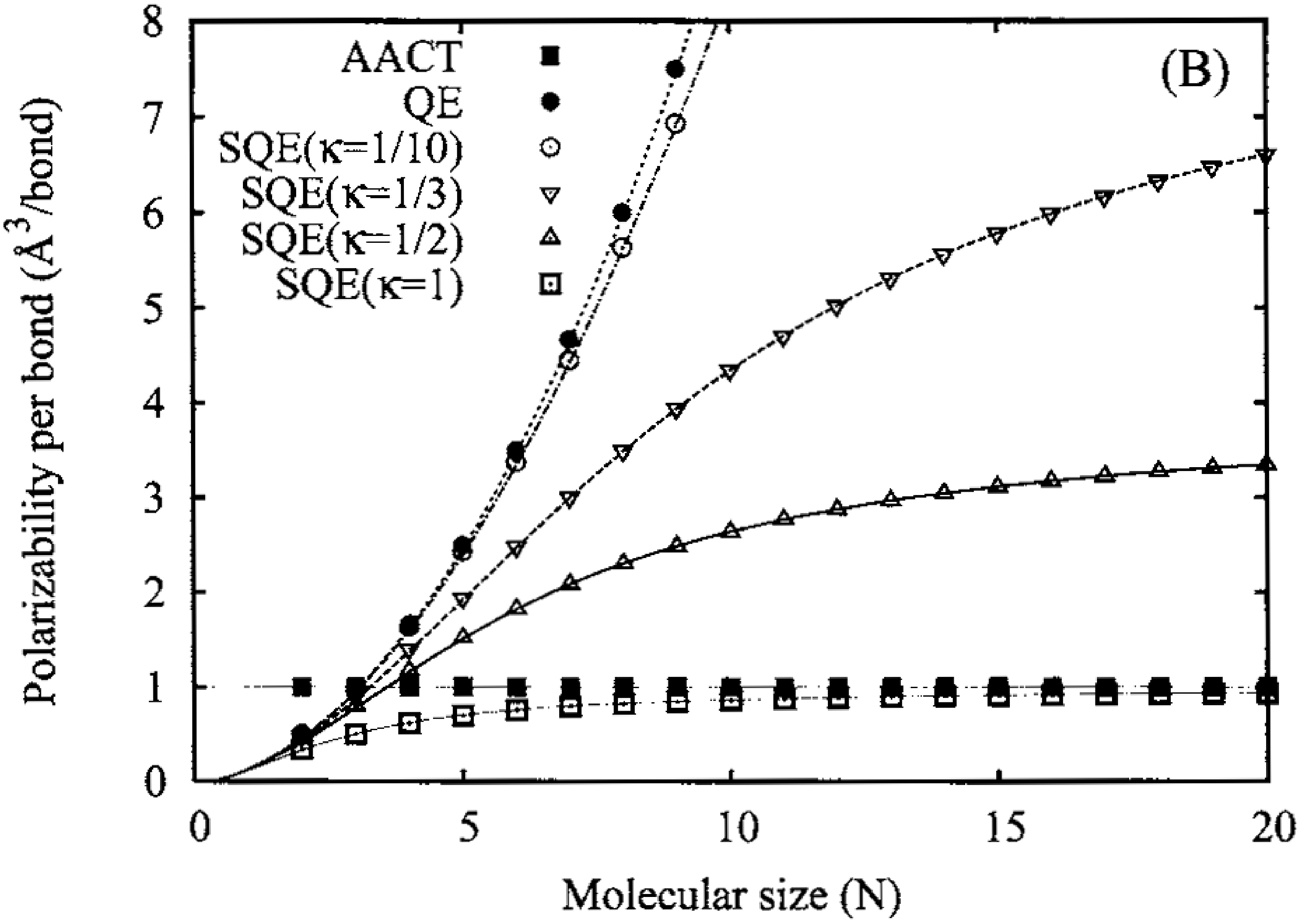
J' =



bond hardness parameters

Nistor et al, *JCP* **2006**, 125, p094108

# Superlinear Scaling of Polarizability



Warren et al, *JCP* **2008**, 128, p144110

# SQE/EEM calibration for organic systems

**Training set**

- * 500 small organic molecules
- * Algorithmic selection from Pubchem
  - - maximized diversity
  - - constraints on atom types and number of atoms

**Calibration of EEM/SQE in 12 different ways**

- * 3 charge schemes: Mulliken, Natural, Hirshfeld-I
- * 2 atom types: elements & elements+number of bonds
- * 2 cost functions: with and without linear response data

**Conclusions (limited to small molecules)**

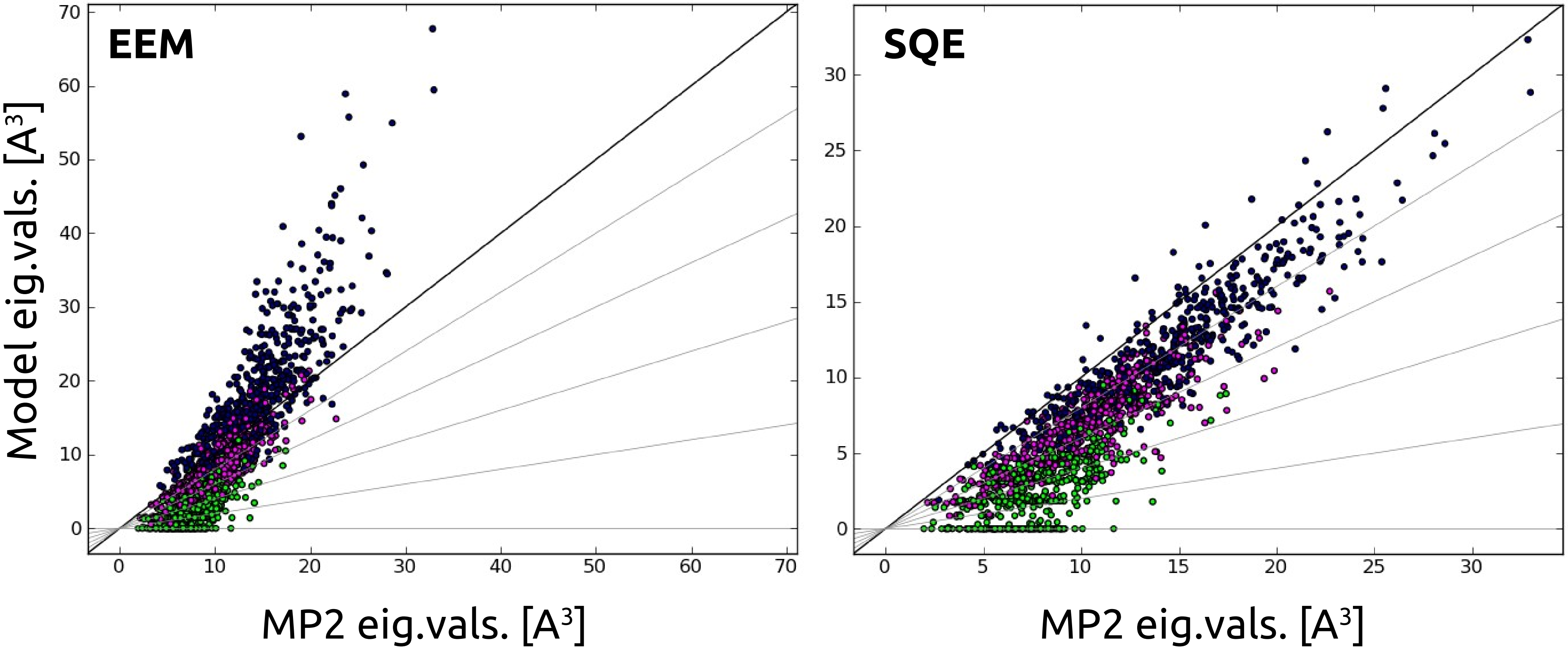
- * SQE outperforms EEM in all tests
- * Cross validation confirms transferability of parameters
- * Hirshfeld-I based calibrations are most useful
- * SQE model gives good dipole polarizabilities

Verstraelen, T et al *JCP* **2009**, 131, p044127

# SQE/EEM calibration for organic systems

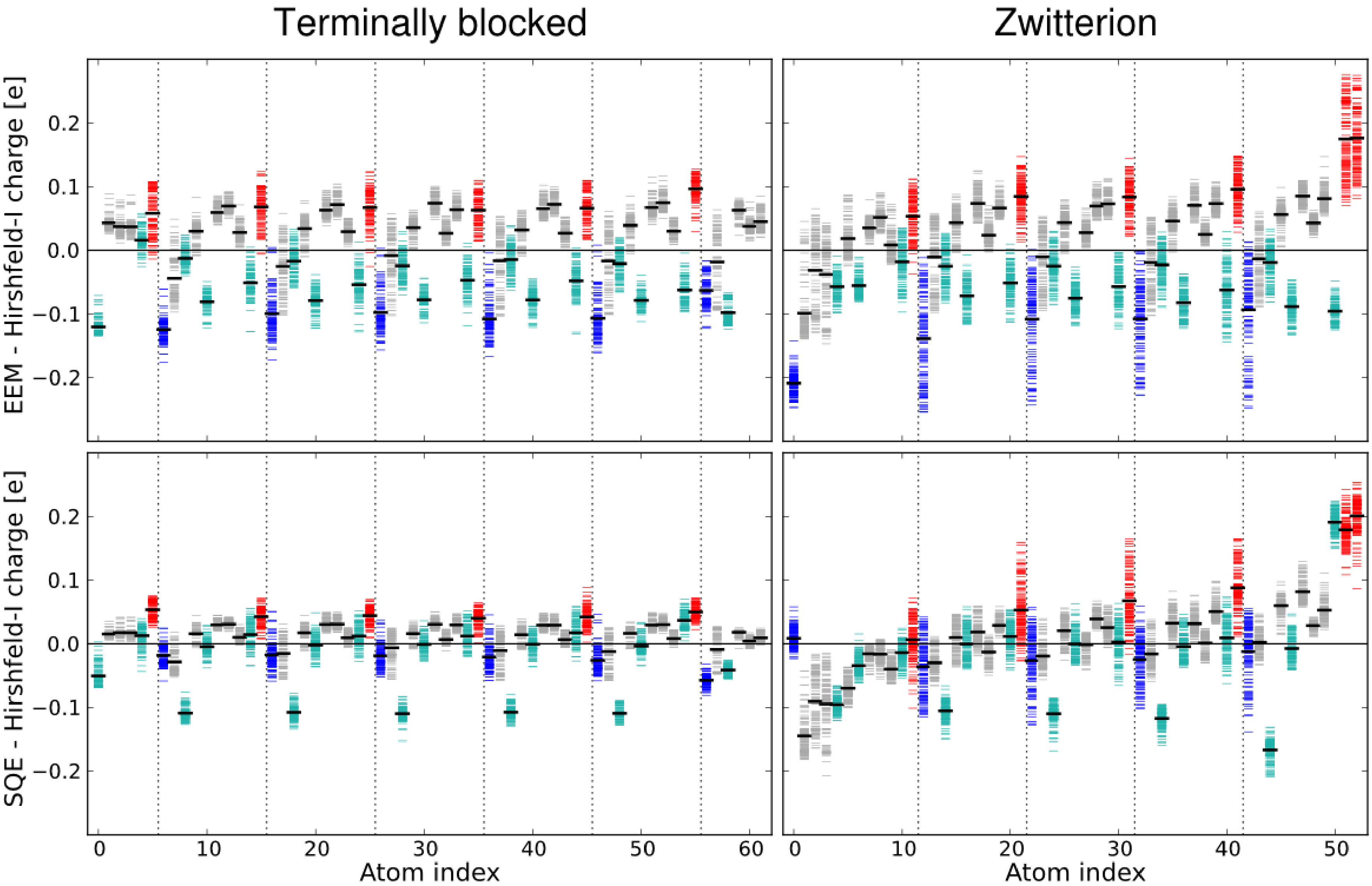
## Comparison of dipole polarizability eigenvalues

(500 small organic molecules)



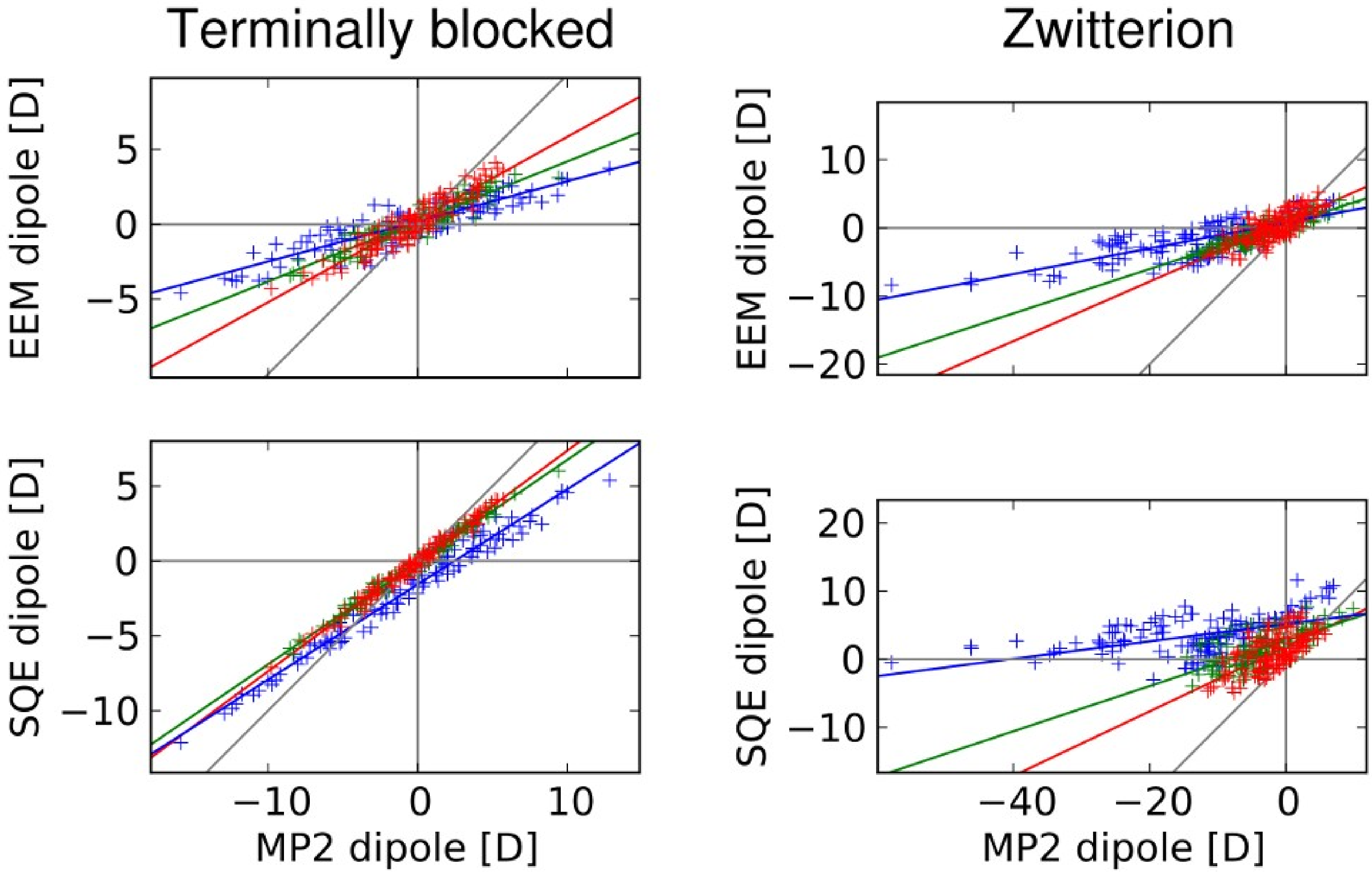
Verstraelen, T et al *JCP* **2009**, 131, p044127

# Penta Alanine Benchmark



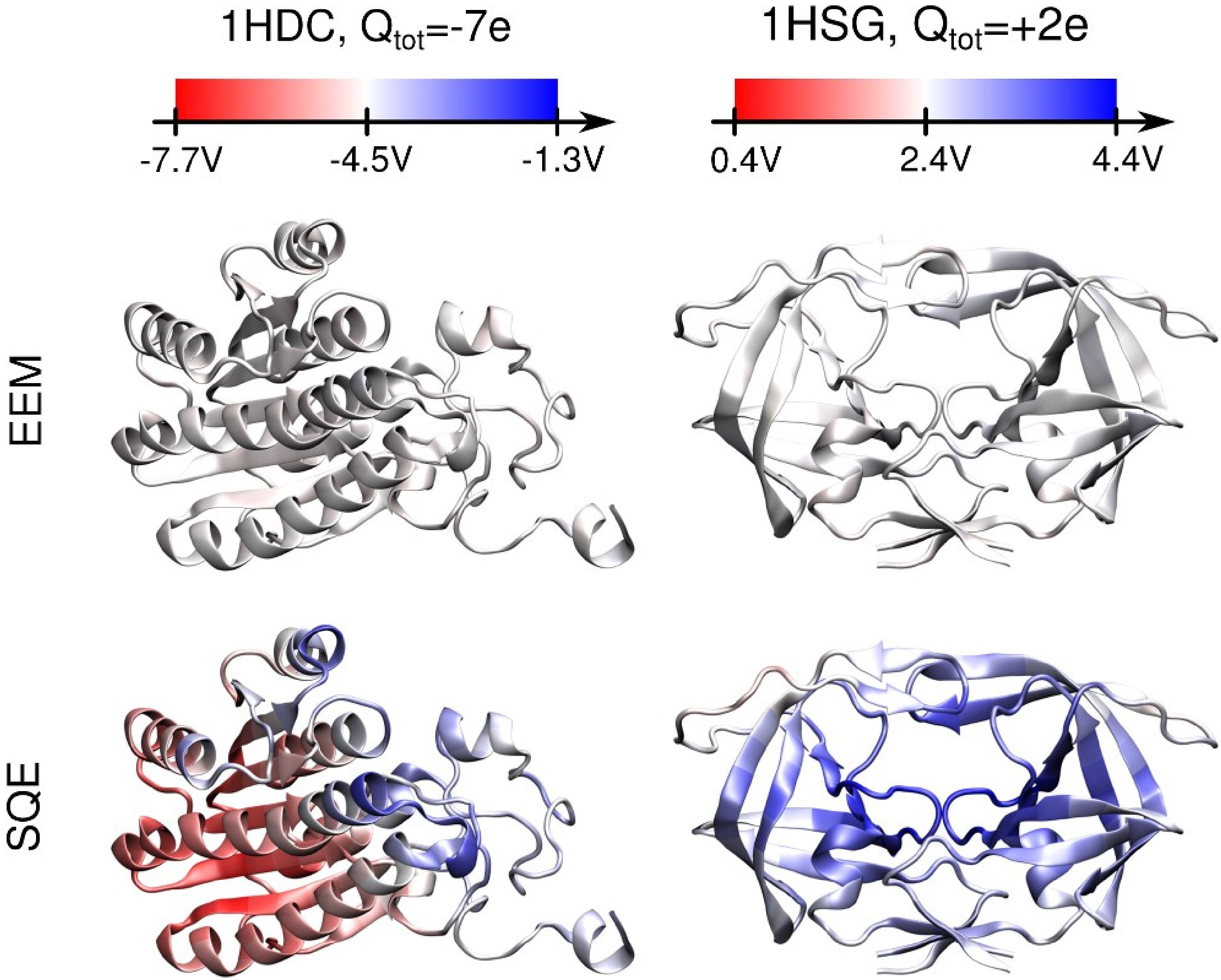
(paper in preparation)

# Penta Alanine Benchmark



(paper in preparation)

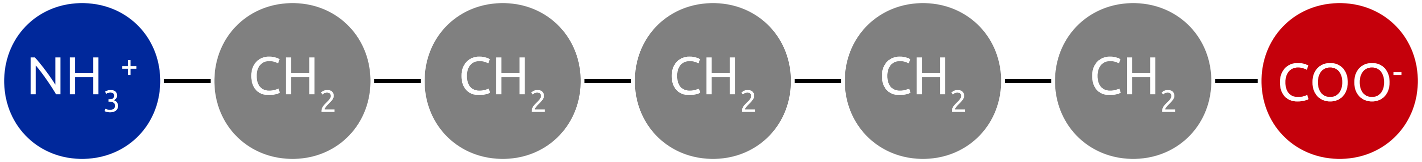
# Penta Alanine Benchmark



(paper in preparation)

# Penta Alanine Benchmark

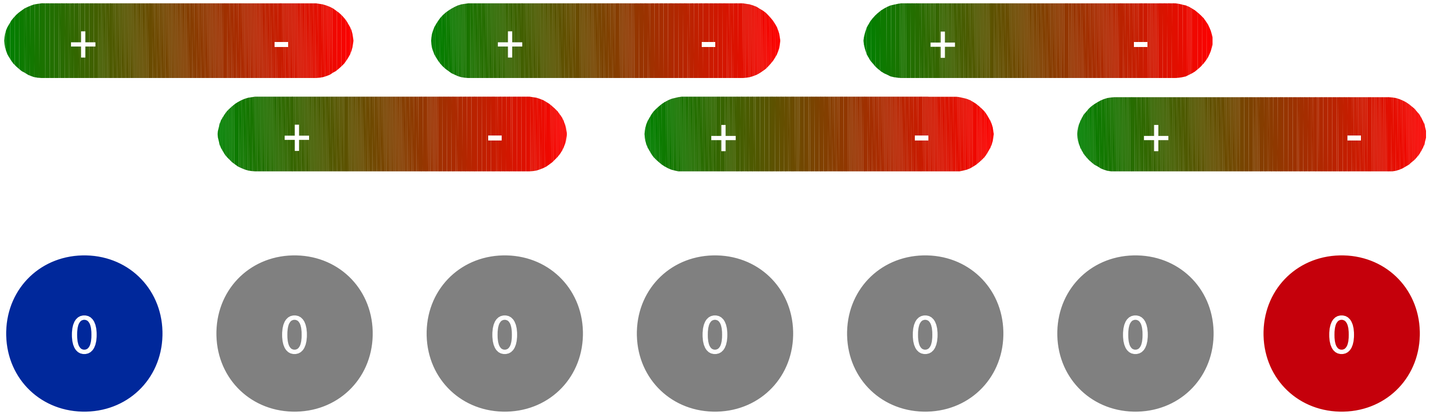
**Charged molecules (and zwitter ions) are problematic.**



SQE model: $q = Tp$

split charges

reference charges

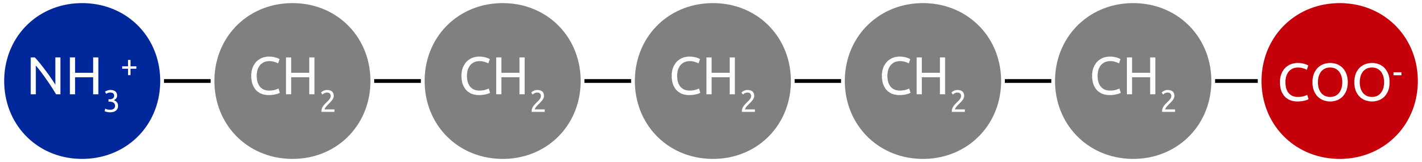


→ electronegativities of endpoints depend on chain length

(paper in preparation)

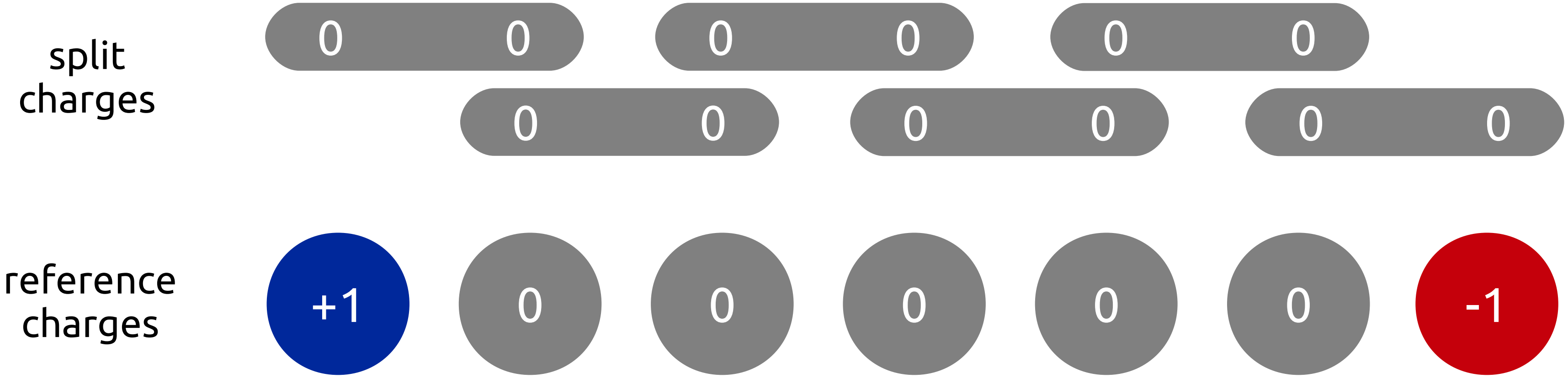
# Penta Alanine Benchmark

**Charged molecules (and zwitter ions) are problematic.**



$NH_3^+$ — $CH_2$ — $CH_2$ — $CH_2$ — $CH_2$ — $CH_2$ — $COO^-$

Better choice (for transferable parameters): $q = Tp + q_0$



split charges

0 0 0 0 0 0

0 0 0 0 0 0

reference charges

+1 0 0 0 0 0 -1

→ More parameters. How to model the reference charges?

(paper in preparation)

# Overview

- **Introduction**
  - Exploring Chemical Compounds
  - Force Field 101
  - Making Force Fields
- **Atoms in Molecules**
  - Partitioning of Electron Density
  - Iterative Hirshfeld
  - ESP Fitted charges
  - Benchmarks
- **The Hydrogen Bond**
  - AIM Application to FF development
  - Charge Transfer Matters!
- **Split Charge Equilibration**
  - Partitioning of Polarization
  - Split Charges
  - Benchmarks
- **Harmonic Models**
  - Force Constants & Rest Values
- **Conclusion**

# The Harmonic Model (1)

## Mathematical form

$$E_{\text{harm}} = \frac{1}{2}\sum_{i=1}^{N_q}\sum_{j=1}^{N_q} K_{ij}(q_i - q_i^{(0)})(q_j - q_j^{(0)})$$

where $q_i$ are a series of internal coordinates (distances, angles, ...)

## Purpose of the harmonic model

- Reproduce geometries and vibrational frequencies
- To be included in FF model, next to electrostatics and others
- Should represent the covalent interactions

## Traditional approach to get parameters

- set $q_i^{(0)}$ to the internal coordinates of a QM/XRD optimized structure
- *Invert* the QM Hessian, i.e. find **a** solution for

$$\frac{\partial^2 E^{(\mathrm{QM})}}{\partial x_k \partial x_l} = H_{kl} = \sum_{i=1}^{N_q}\sum_{j=1}^{N_q} K_{ij}\frac{\partial q_i}{\partial x_k}\frac{\partial q_j}{\partial x_l} \qquad \text{or} \qquad H = J^T K J$$

Ermoshin et al, *Chem. Phys.* **1996**, 202, p53

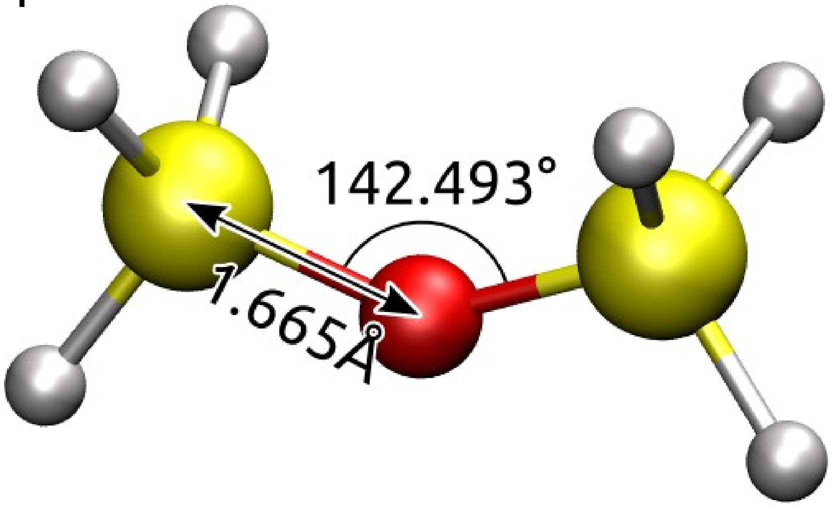
# The Harmonic Model (2)

## Problems with the traditional approach

- **Not optimal:** which Jacobian inverse is best?
- **Not optimal:** it is common to neglect many cross terms
- One assumes that other terms (electrostatics, ...) do not affect geometry
  - ⇒ requires 1-2, 1-3 or 1-4 exclusion rules
  - **However ...**
    - 1-2, 1-3 and 1-4 electrostatics are real
    - For EEM or SQE, all electrostatic terms must be present

## Calibration of valence parameters must be reinvented!

- Turns out to be more complicated than expected
- Demo with a 2T silica cluster
- QM training data
  - PBE/DGTZVP
  - geometry optimization
  - frequency computation



# Bare-bones FF model for 2T

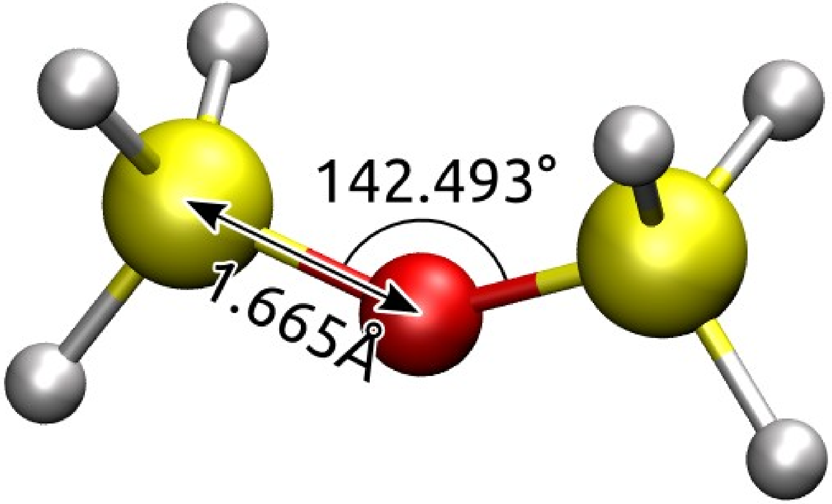
## Valence Part (without cross terms)

Harmonic bond stretch terms: **Si-H**; **Si-O**

$$\frac{1}{2}K_b\left(d - d_b^{(0)}\right)^2$$

Harmonic Cosine terms: **H-Si-H**, **H-Si-O**, **Si-O-Si**

$$\frac{1}{2}K_a\left(\cos(\alpha) - \cos\left(\alpha_a^{(0)}\right)\right)^2$$



## Electrostatics

Fixed charge transfer: **Si-H**, **Si-O**

$$\frac{1}{2}\sum_{i\neq j}\frac{q_i q_j}{r_{ij}}$$

(with or without exclusion rules, both cases will be considered)

Charge transfers are ESP-fitted: $q_{\text{Si->H}}$= -0.118e  $q_{\text{Si->O}}$= -0.229e

# Method A

## Traditional approach

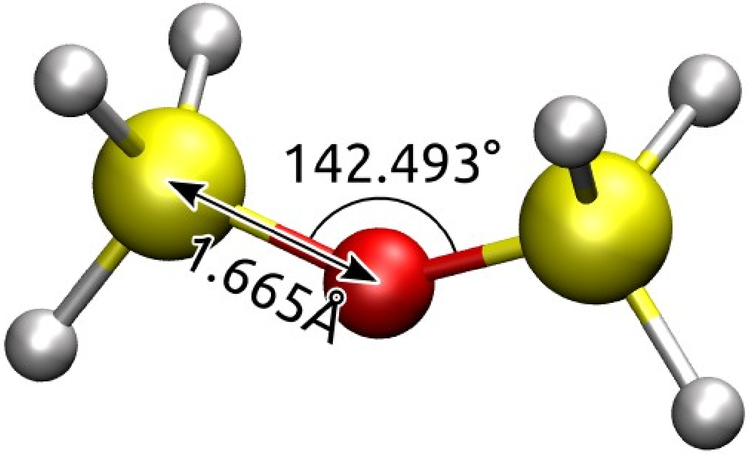
- * Force constants
  - - Use Moore-Penrose inverse of Jacobian.
  - - Neglect cross terms.
- * Rest values
  - - Optimal internal coordinates
- * 1-3 exclusion lists (bonds and bends are excluded)

## Fitted parameters

- * Numbers in red are compromises
- * Many significant off-diagonal $K_{ij}$ (mainly H-Si-H)

| | Force constant | Rest value |
|---|---|---|
| Si-H | 1636.7 kJ $mol^{-1}$ $Å^{-2}$ | 1.50 Å |
| Si-O | 2876.1 kJ $mol^{-1}$ $Å^{-2}$ | 1.67 Å |
| H-Si-H | 344.0 kJ $mol^{-1}$ | 109.38 ° |
| H-Si-O | 265.2 kJ $mol^{-1}$ | 109.56 ° |
| Si-O-Si | 159.0 kJ $mol^{-1}$ | 142.49 ° |

# Results for Method A

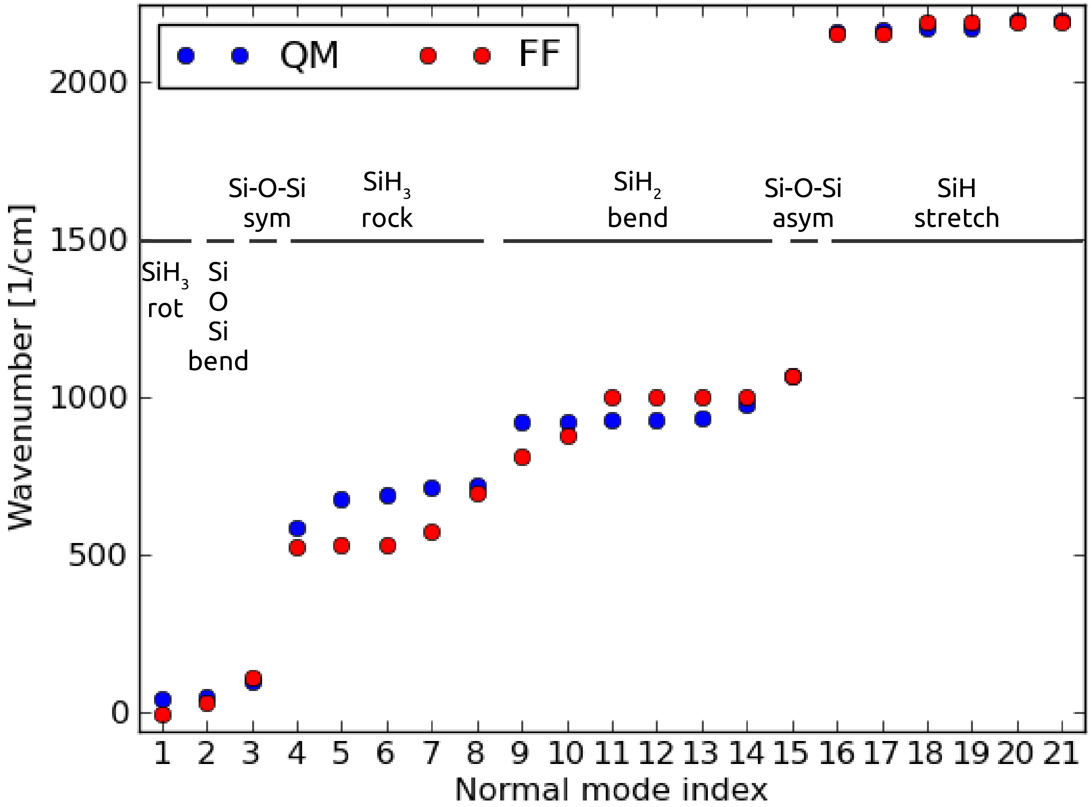


## Optimized internal coordinates

| | | |
|---|---|---|
| Si-O-Si | angle | 125.055 ° |
| Si-O | bond | 1.655 Å |

## Comments

- Si-O-Si angle is wrong due to 1-4 and 1-5 electrostatics
- FF Spectrum is reasonable
- Only way to improve is "trial and error"



# Method B

## Least-squares approach

No Exclusion lists

1) Guess force constants: 'approximate' least-squares fit

$$\frac{\partial^2 E^{(\mathrm{QM})}}{\partial x_k \partial x_l} = \frac{\partial^2 E^{(\mathrm{ES})}}{\partial x_k \partial x_l} + \sum_{i=1}^{N_q} K_{ii} \frac{\partial q_i}{\partial x_k} \frac{\partial q_i}{\partial x_l}$$

2) Rest values: least squares fit of (zero gradient)

$$0 = \frac{\partial E^{(\mathrm{QM})}}{\partial x_k} = \frac{\partial E^{(\mathrm{ES})}}{\partial x_k} + \sum_{i=1}^{N_q} K_{ii}(q_i - q_i^{(o)}) \frac{\partial q_i}{\partial x_k}$$

3) Redo force constants: 'correct' least-squares fit

$$\frac{\partial^2 E^{(\mathrm{QM})}}{\partial x_k \partial x_l} = \frac{\partial^2 E^{(\mathrm{ES})}}{\partial x_k \partial x_l} + \sum_{i=1}^{N_q} K_{ii} \frac{\partial q_i}{\partial x_k} \frac{\partial q_i}{\partial x_l} + \sum_{i=1}^{N_q} K_{ii}(q_i - q_i^{(0)}) \frac{\partial^2 q_i}{\partial x_k \partial x_l}$$

# Parameters for Method B

## Least-squares approach

After step 2:

| | Force constant | Rest value |
|---|---|---|
| Si-H | 1664.3 kJ $mol^{-1}$ $Å^{-2}$ | 1.51 Å |
| Si-O | 2713.8 kJ $mol^{-1}$ $Å^{-2}$ | 1.70 Å |
| H-Si-H | 287.9 kJ $mol^{-1}$ | 140.69 ° |
| H-Si-O | 374.7 kJ $mol^{-1}$ | 131.95 ° |
| Si-O-Si | 24.6 kJ $mol^{-1}$ | 23.12 ° |

| Traditional f.c.'s | |
|---|---|
| Si-O | 1636.7 |
| Si-O | 2876.1 |
| H-Si-H | 344.0 |
| H-Si-O | 265.2 |
| Si-O-Si | 159.0 |

* Poor rest values for angles

* Bending f.c.'s change a lot

* Bond rest values increase to compensate electrostatics

# Parameters for Method B

**Least-squares approach**

After step 3:

| | Force constant | Rest value |
|---|---|---|
| Si-H | 1659.2 kJ mol$^{-1}$ Å$^{-2}$ | 1.51 Å |
| Si-O | 2709.9 kJ mol$^{-1}$ Å$^{-2}$ | 1.70 Å |
| H-Si-H | 172.6 kJ mol$^{-1}$ | 140.69 ° |
| H-Si-O | 250.3 kJ mol$^{-1}$ | 131.95 ° |
| Si-O-Si | 0.0 kJ mol$^{-1}$ | 23.12 ° |

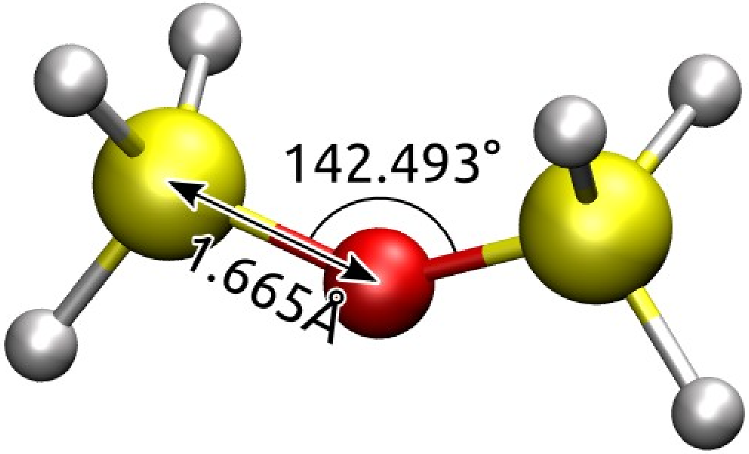
| | Traditional f.c.'s |
|---|---|
| Si-O | 1636.7 |
| Si-O | 2876.1 |
| H-Si-H | 344.0 |
| H-Si-O | 265.2 |
| Si-O-Si | 159.0 |

* This step is supposed to fine-tune the f.c.'s, but...

* Bending f.c.'s values do not become more realistic.

In general: parameters are not encouraging

# Results for Method B(2)

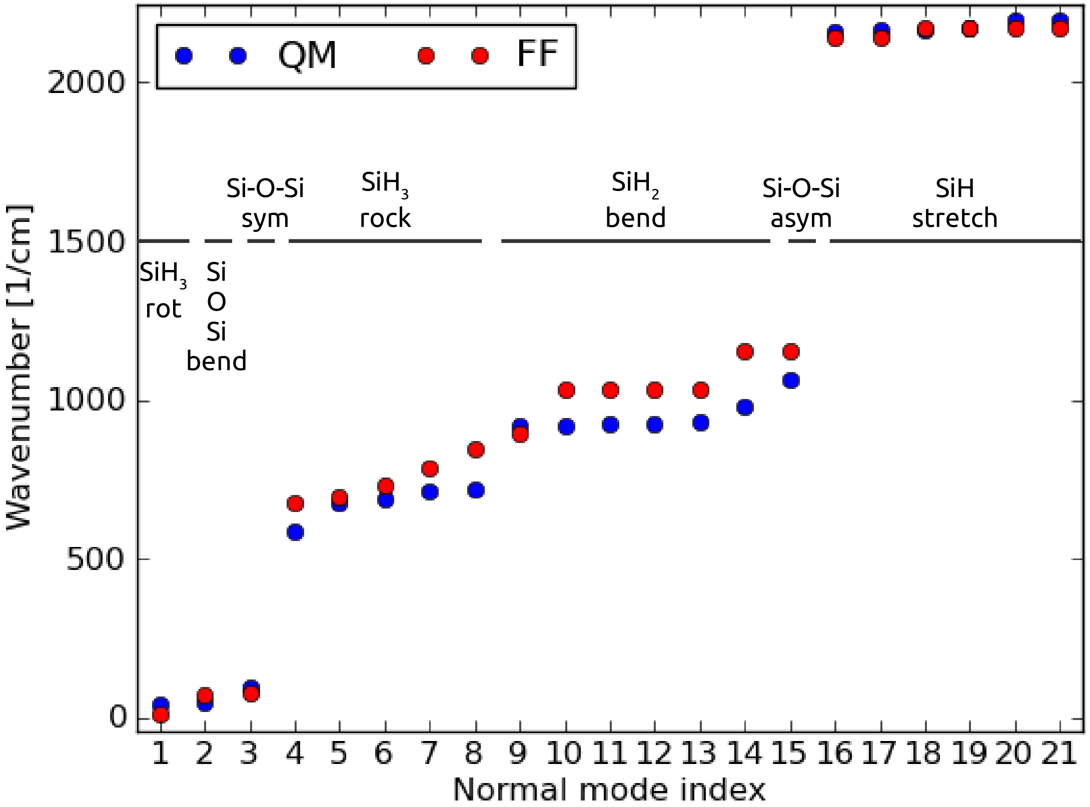


## Optimized internal coordinates

| | | |
|---|---|---|
| Si-O-Si | angle | 108.204 ° |
| Si-O | bond | 1.668 Å |

## Comments

- Si-O-Si angle is wrong because a small error in the gradient along a low modes causes a large error in geometry
- Bond length is better
- FF Spectrum is worse



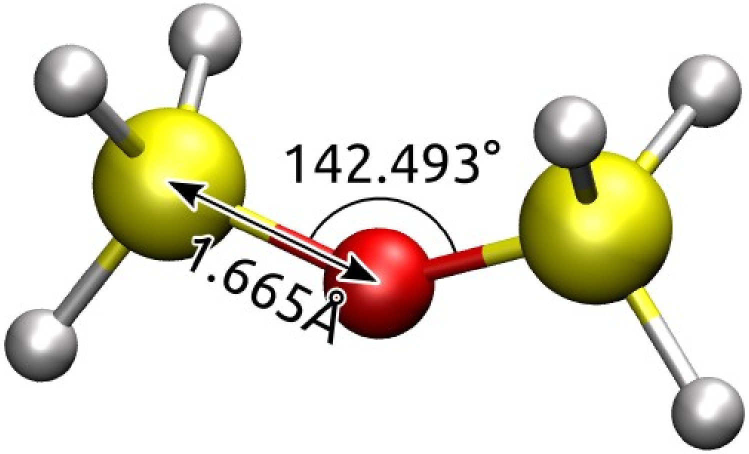
# Results for Method B(3)

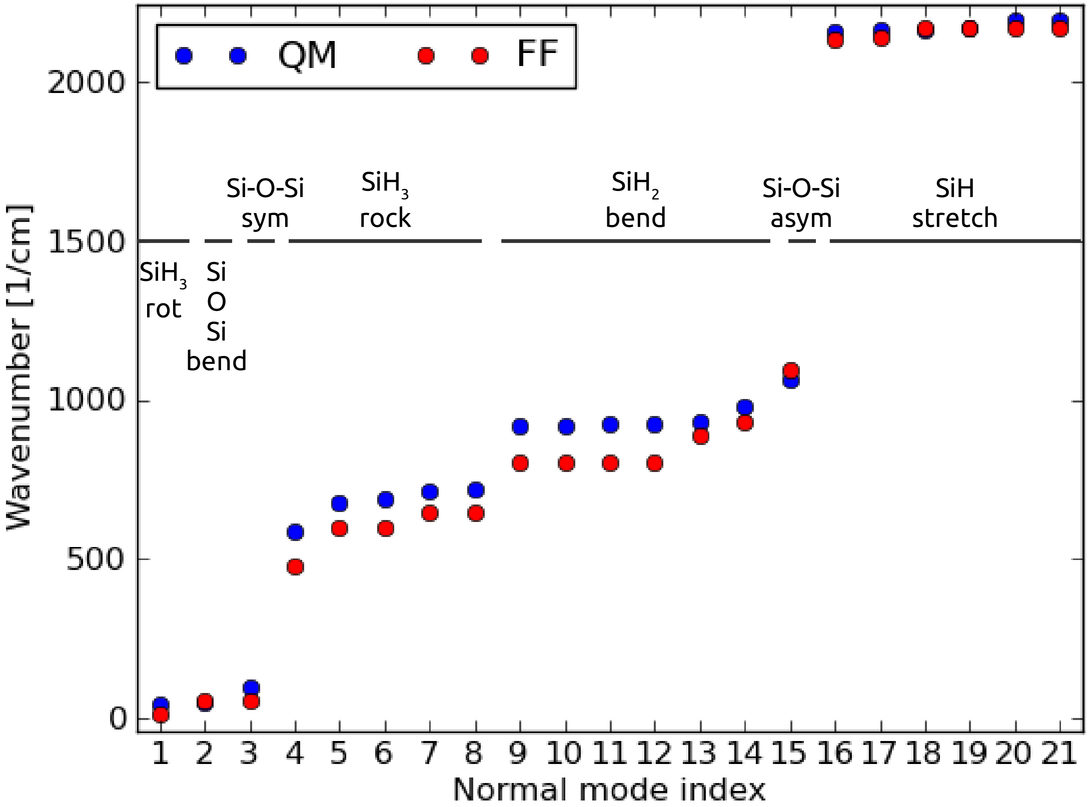
## Optimized internal coordinates

| | | |
|---|---|---|
| Si-O-Si | angle | 180.00 ° |
| Si-O | bond | 1.666 Å |

## Comments

- Pfff...





# Method C

## Improved least-squares approach

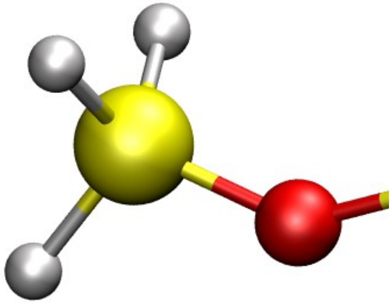
* Still no exclusion lists

* Try to fix all the issues encountered in B

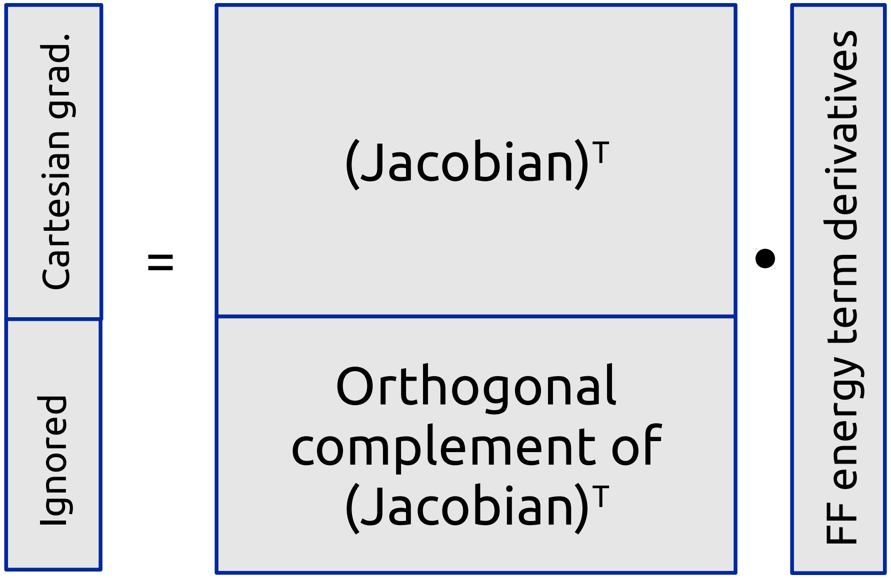
## Problem 1: the computation of the Cartesian gradient is a projection

$N_q$ derivatives of FF energy terms → 3N-6 independent

⇒ rest values for **H-Si-H** and **H-Si-O** are ill-defined



6 angles
5 independent



Cartesian grad.

Ignored

=

$(\text{Jacobian})^T$

Orthogonal complement of $(\text{Jacobian})^T$

•

FF energy term derivatives

(paper in preparation)

# Method C

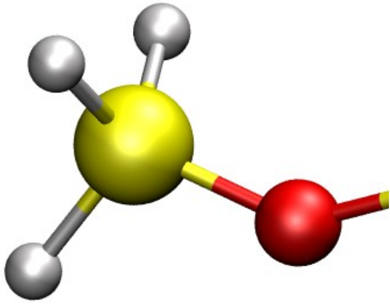
## Improved least-squares approach

- Still no exclusion lists
- Try to fix all the issues encountered in B

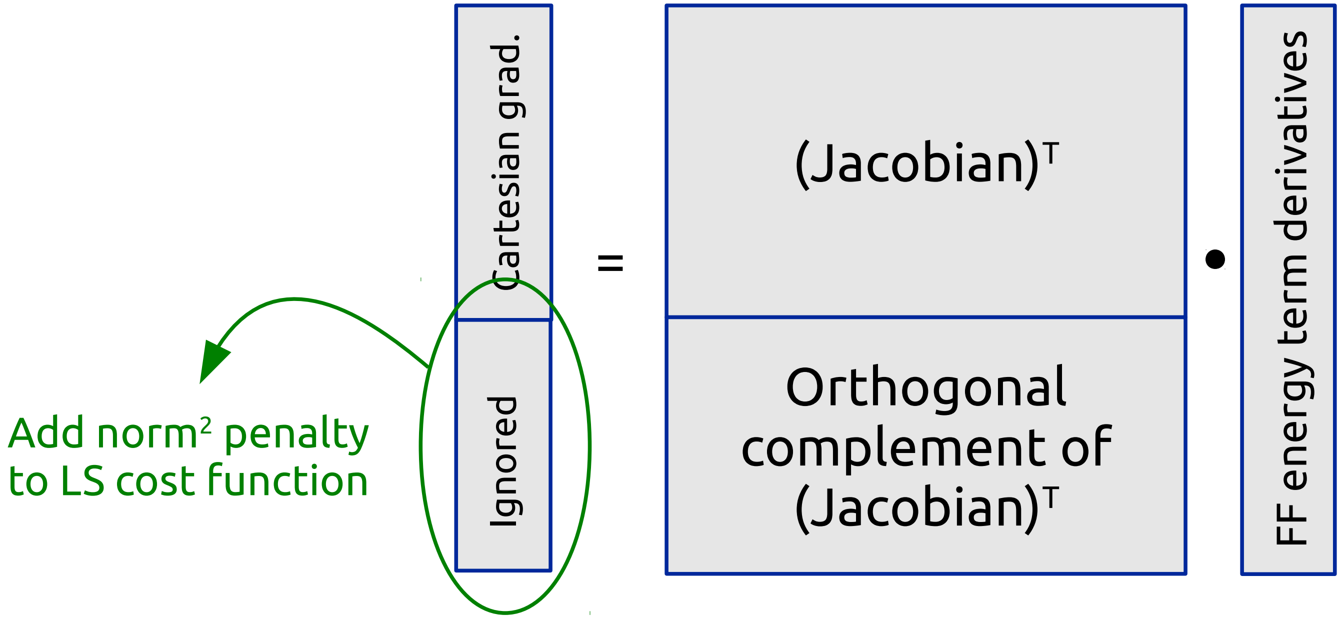
## Problem 1: the computation of the Cartesian gradient is a projection

$N_q$ derivatives of FF energy terms → 3N-6 independent

⇒ rest values for **H-Si-H** and **H-Si-O** are ill-defined



6 angles
5 independent



Cartesian grad.

Ignored

=

(Jacobian)$^T$

Orthogonal complement of (Jacobian)$^T$

•

FF energy term derivatives

Add norm$^2$ penalty
to LS cost function

(paper in preparation)

# Method C

**Improved least-squares approach**

* Still no exclusion lists

* Try to fix all the issues encountered in B

**Problem 2: gradient along low modes should be more accurate**

Solution: use a transformed gradient in LS fit, as follows

$$\Delta X = H_{\mathrm{QM}}^{-1} G$$

Related transformation for Hessian fit:

$$H' = H_{\mathrm{QM}}^{-\frac{1}{2}} H H_{\mathrm{QM}}^{-\frac{1}{2}}$$

(paper in preparation)

# Method C

**Improved least-squares approach**

- * Still no exclusion lists
- * Try to fix all the issues encountered in B

**Problem 3: not entirely optimal**

The last step (3) should optimize all parameters with one total cost function

$$X_{\text{total}} = \frac{X_{\text{grad}} + \lambda_1 X_{\text{hess}}}{1 + \lambda_1} + \lambda_2 X_{\text{penalty}}$$

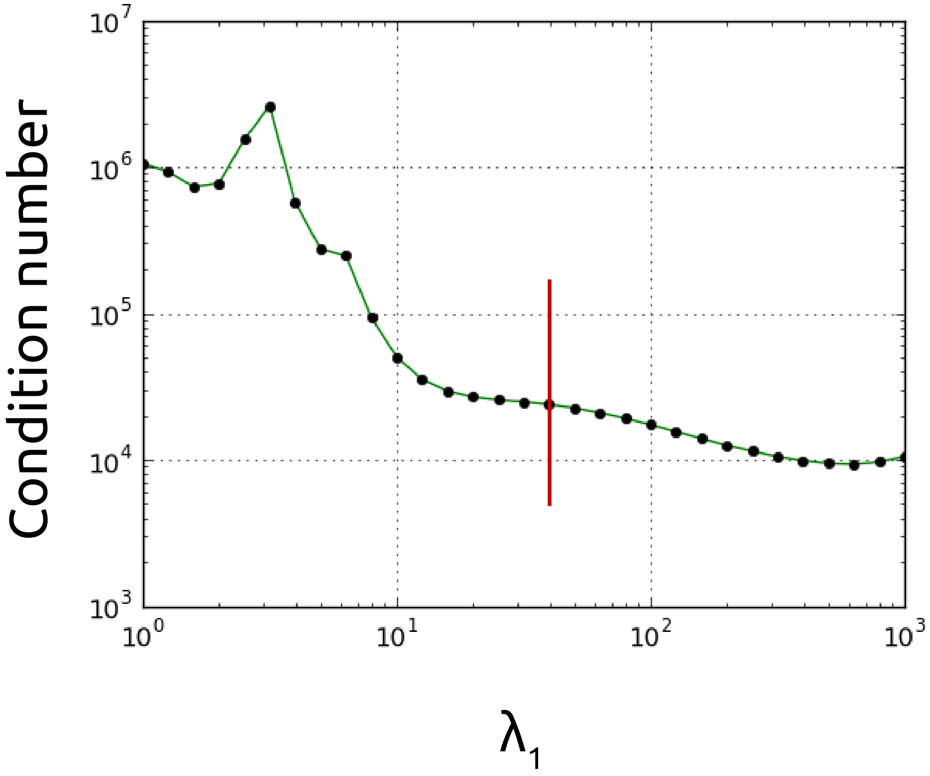
$\lambda_2$ can be set relatively high, does not hurt

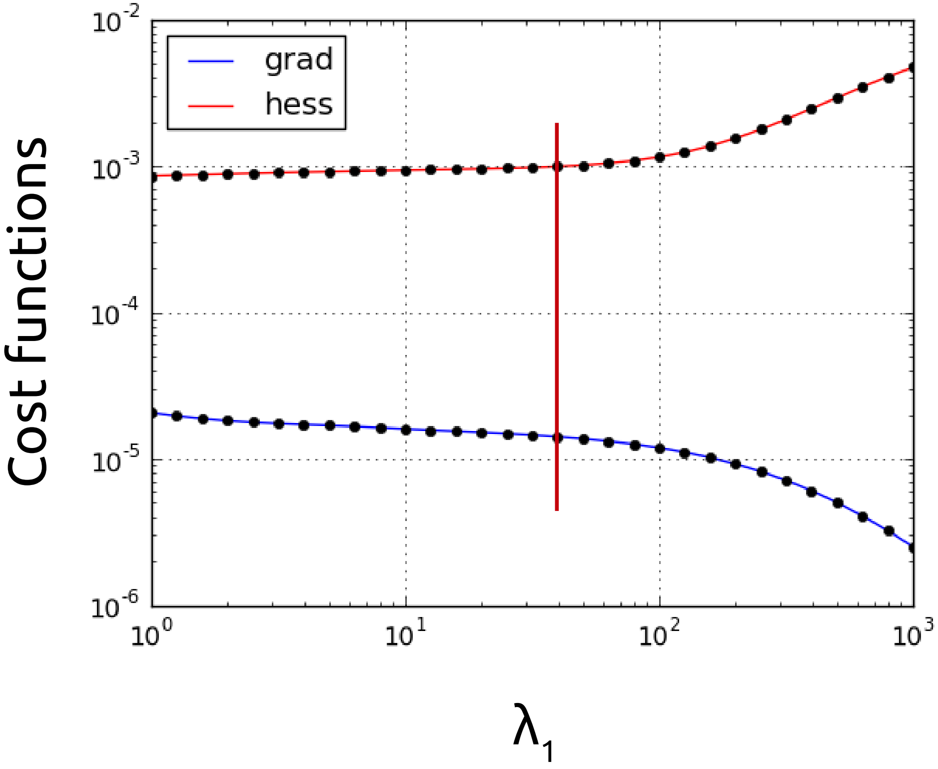
$\lambda_1$ must be scanned over a few orders of magnitude:

- * Good compromise between $X_{\text{grad}}$ and $X_{\text{hess}}$
- * Low condition number

(paper in preparation)

# Tuning λ1 in Method C





(paper in preparation)

# Parameters for Method C

## Improved least-squares approach

After step 3:

differ traditional values

| | Force constant | Rest value |
|---|---|---|
| Si-H | 1668.1 kJ mol$^{-1}$ Å$^{-2}$ | 1.51 Å |
| Si-O | 2468.4 kJ mol$^{-1}$ Å$^{-2}$ | 1.70 Å |
| H-Si-H | 269.9 kJ mol$^{-1}$ | 110.39 ° |
| H-Si-O | 330.1 kJ mol$^{-1}$ | 109.01 ° |
| Si-O-Si | 48.4 kJ mol$^{-1}$ | 122.26 ° |

Penalty

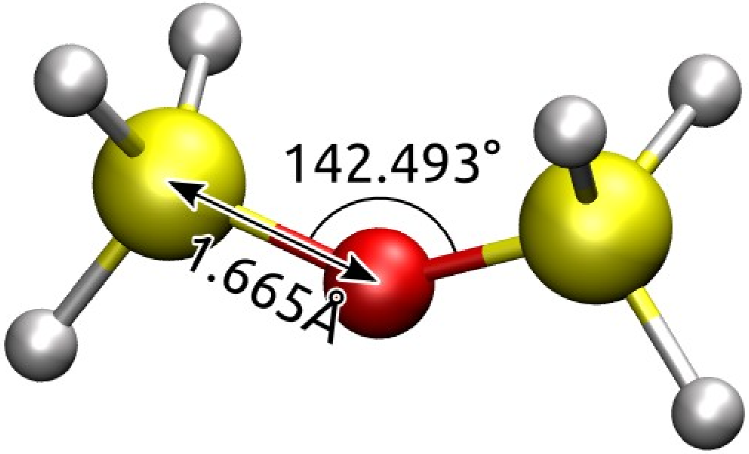
Differs from optimized value

Traditional approach (method A)

| | Force constant | Rest value |
|---|---|---|
| Si-H | 1636.7 kJ mol$^{-1}$ Å$^{-2}$ | 1.50 Å |
| Si-O | 2876.1 kJ mol$^{-1}$ Å$^{-2}$ | 1.67 Å |
| H-Si-H | 344.0 kJ mol$^{-1}$ | 109.38 ° |
| H-Si-O | 265.2 kJ mol$^{-1}$ | 109.56 ° |
| Si-O-Si | 159.0 kJ mol$^{-1}$ | 142.49 ° |

(paper in preparation)

# Results for Method C

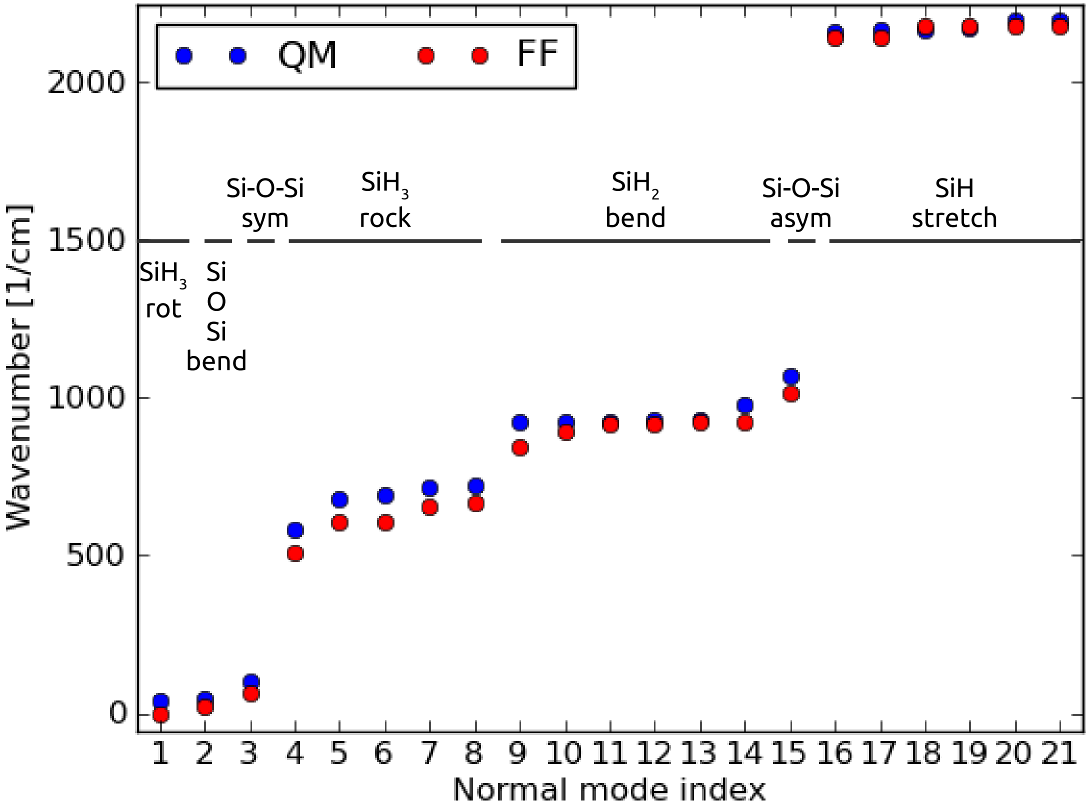


## Optimized internal coordinates

| | | |
|---|---|---|
| Si-O-Si | angle | 143.146 ° |
| Si-O | bond | 1.663 Å |

## Comments

- Best so far
- To improve further:
    1. Cross terms
    2. Improved electrostatics (e.g. atomic dipoles)
    3. Inclusion of repulsion (and dispersion)



(paper in preparation)

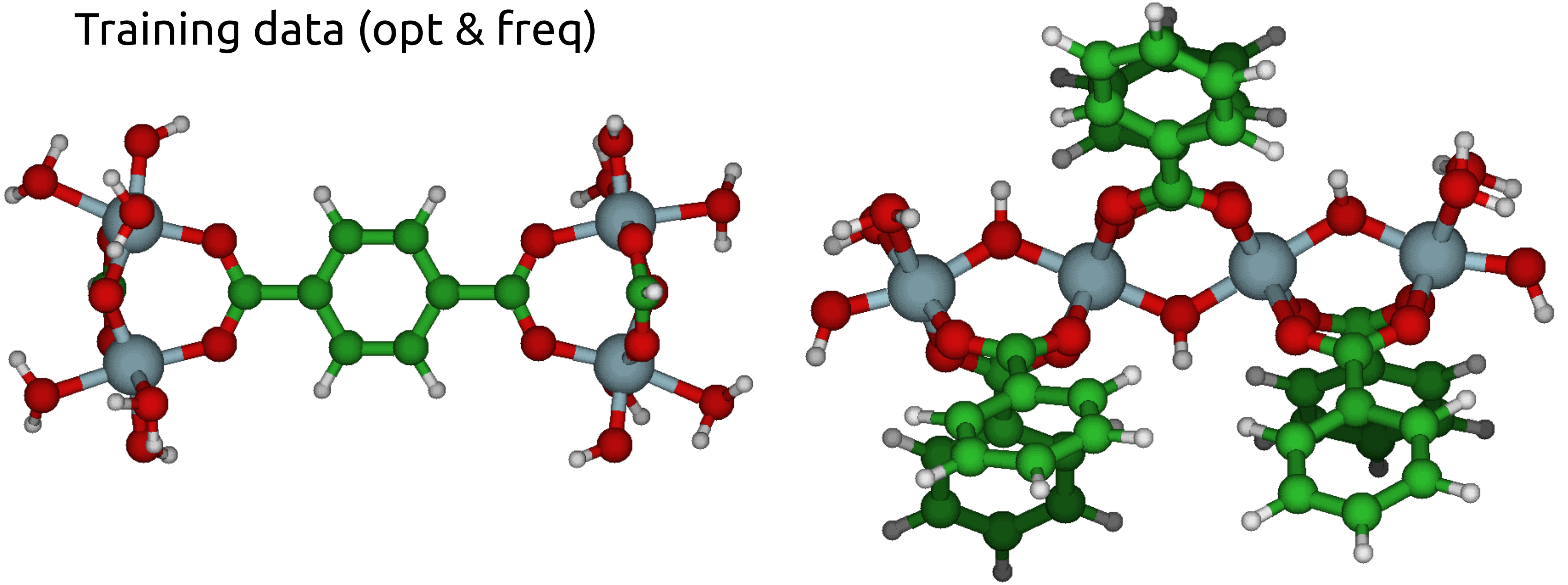
# There is more...

## Implemented (FFit2 Code), but not demonstrated in the 2T example

* Use Hessian & gradient data from a large set of molecules
* Use charge equilibration models for the electrostatics
* Account for repulsion/dispersion terms calibrated earlier
* Possibility to ignore parts of molecules (e.g. weird terminations)

## Calibration of MIL-53

Training data (opt & freq)



(paper in preparation)

# Application to MIL-53

## Cell optimization

* Experimental (HT phase):

Loiseau et al, *Chem. Eur. J.* **2004**, 10, p1373

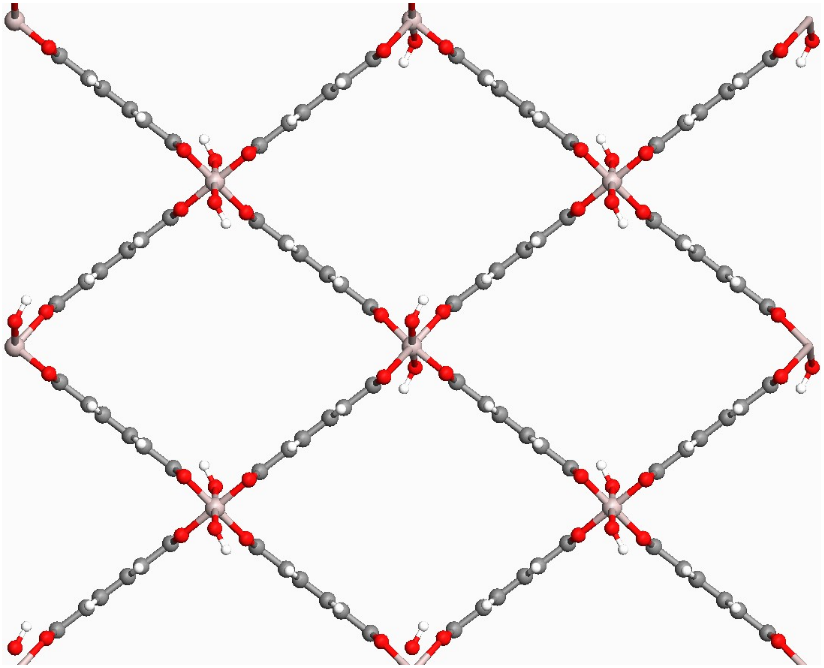
a=16.675 Å b=6.609 Å c=12.813 Å

* Cell optimization with FF

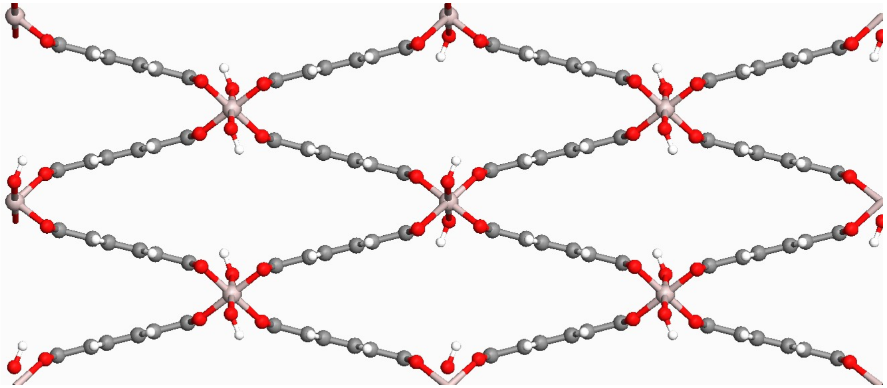
a=16.120 Å b=7.417 Å c=13.093 Å

## Testing sets of dispersion parameters to reproduce breathing

Open form (HT)



Closed form (LT)



(paper in preparation)

# Conclusion

## Major advances in FF's calibration are on the way!

| | |
|---|---|
| **Atoms in Molecules** | Iterative Hirshfeld is a very neat partitioning scheme:<br>- Robust (conformational, basis)<br>- ESP is OK<br>- AIM very similar to isolated atoms |
| **The Hydrogen Bond** | AIM → FF parameters without fitting<br>e.g. Charge Transfer in Hydrogen Bond<br>(cfr. work Morokuma) |
| **Split Charge Equilibration** | e- Linear Response is ruled by charges<br>SQE fixes polarizability scaling in EEM<br>SQE only good for locally neutral systems |
| **Harmonic Models** | Improved calibration scheme<br>→ Accurate models,<br>without manual tweaking |

# Acknowledgements

Fine colleagues at the CMM ...

Software used: Gaussian09, **HiPart**, **FFit2**, **Yaff**, Python, VMD, POVRay

MIL-53:
L. Vanduyfhuys

Funding: Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF) and BELSPO in the frame of IAP/6/27, European Research Council under the European Community's Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483).

Computing facilities: Ghent University (Stevin Supercomputer Infrastructure).