

Pair cutoff determination based on subsets of simulation data points



Elisa Rebolini^{1,2}, Gustav Baardsen¹, Audun Skau Hansen¹,
Karl Røad Leikanger¹, and Thomas Bondo Pedersen¹.



¹Hylleraas Centre for Quantum Molecular Sciences
University of Oslo, NO-0315 Oslo, Norway

²Institut Laue-Langevin, 38042 Grenoble, France

gustav.baardsen@kjemi.uio.no

1. Physical model

The goal is to model periodic molecular systems, such as a molecule attached to a crystal surface, with high precision.

- Simulations of periodic molecular systems, such as a diamond crystal, for example, using a quantum-mechanical description
- An infinite number of electrons and atomic nuclei involved

A differential equation (Schrödinger equation) must be solved [1]:

$$\sum_{i=0}^{\infty} \nabla_i^2 \Phi(r_1, r_2, \dots) + \sum_{i<j} V(r_1, r_2) \Phi(r_1, r_2, \dots) = E \cdot \Phi(r_1, r_2, \dots), \quad (1)$$

where r_i is the position of particle i , ∇ is the nabla operator, V is a two-body potential, and E is the total energy of the system.

Utilization of periodic symmetries + truncation of long-range interactions
⇒ Finite number of computing operations

This gives

$$\infty \Rightarrow N \text{ in Equation (1)}. \quad (3)$$

2. The many-body equation

The differential equation (1) is rewritten as a discretized integral equation [1].

The total energy is decomposed in two parts; that is,

$$E = E_{HF} + E_C, \quad (4)$$

where the Hartree-Fock energy E_{HF} is often cheaper to compute than E_C , which is called the correlation energy.

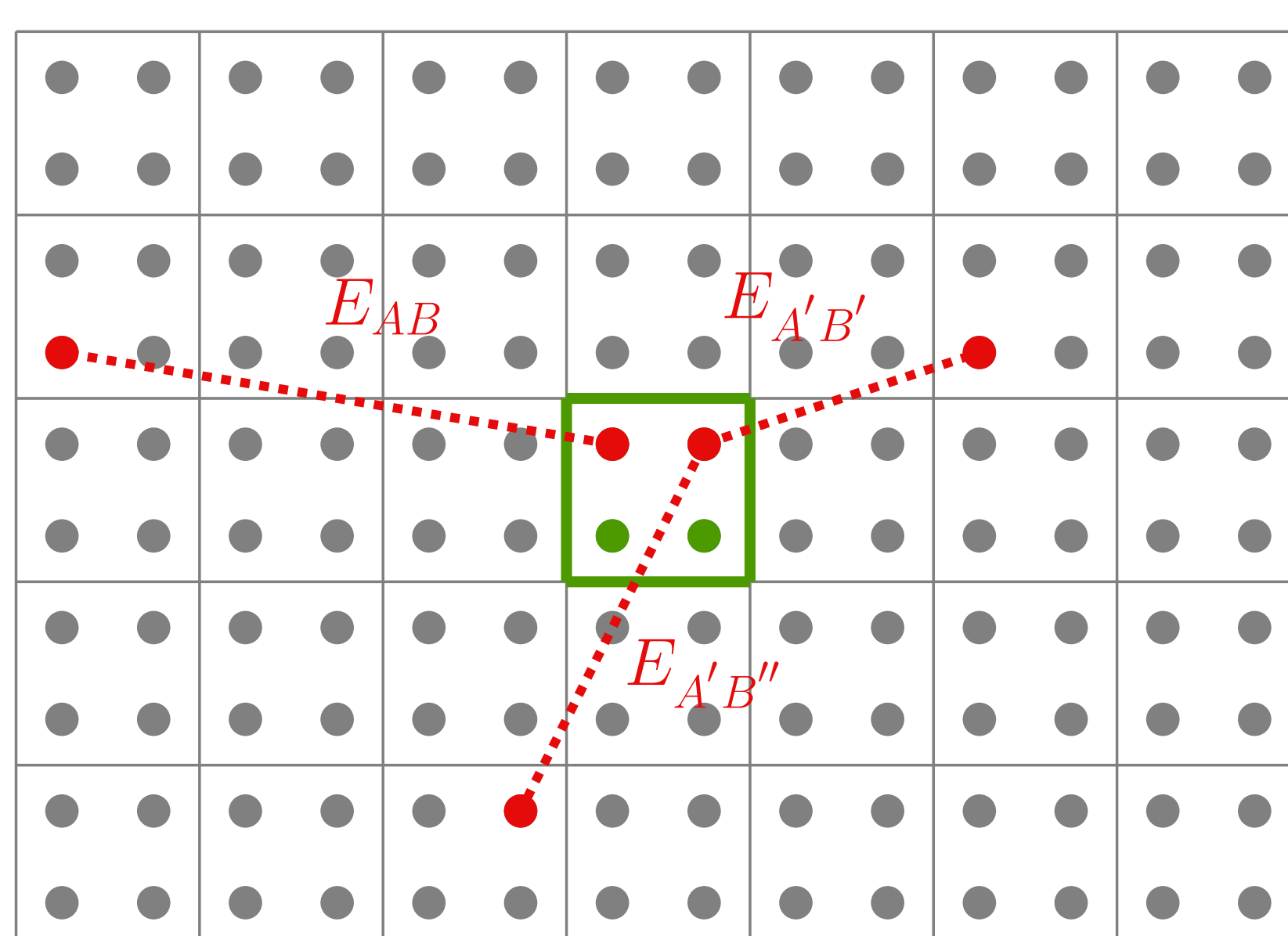


Figure 1: A large number of atomic pair energies E_{PQ} must be computed. Due to periodicity, atom P is always within the unit cell (green box), whereas atom Q may be an atom in any cell.

In the divide-expand-consolidate (DEC) method [2, 3, 4], the correlation energy is computed as

a sum over single-atom terms plus a sum over atom-pair terms:

$$E_C = \sum_P E_P + \underbrace{\sum_{P<Q} E_{PQ}}_{\text{Sum of pair energies}}, \quad (5)$$

where P is an atom within the unit cell (marked by a green box in Figure 1) and Q is an atom in any cell [5, 6].

3. Automatic determination of a pair cutoff

- Theoretically, one can show that [4]

$$E_{PQ}(R_{PQ}) \propto R_{PQ}^{-6} \quad (6)$$

when the pair distance R_{PQ} is sufficiently large.

- A periodic system has an infinite number of pairs, and the pair energies are expensive to compute.
- The assumption is that pairs with a sufficiently large distance R_{PQ} can be neglected [4] or approximated [7].
- In Ref. [6], we suggest an algorithm to determine the pair cutoff R_c by using a smoothed cubic spline [8] as a model, estimating the error associated with different thresholds, and choosing R_c so that the error is estimated to be below a given energy value.

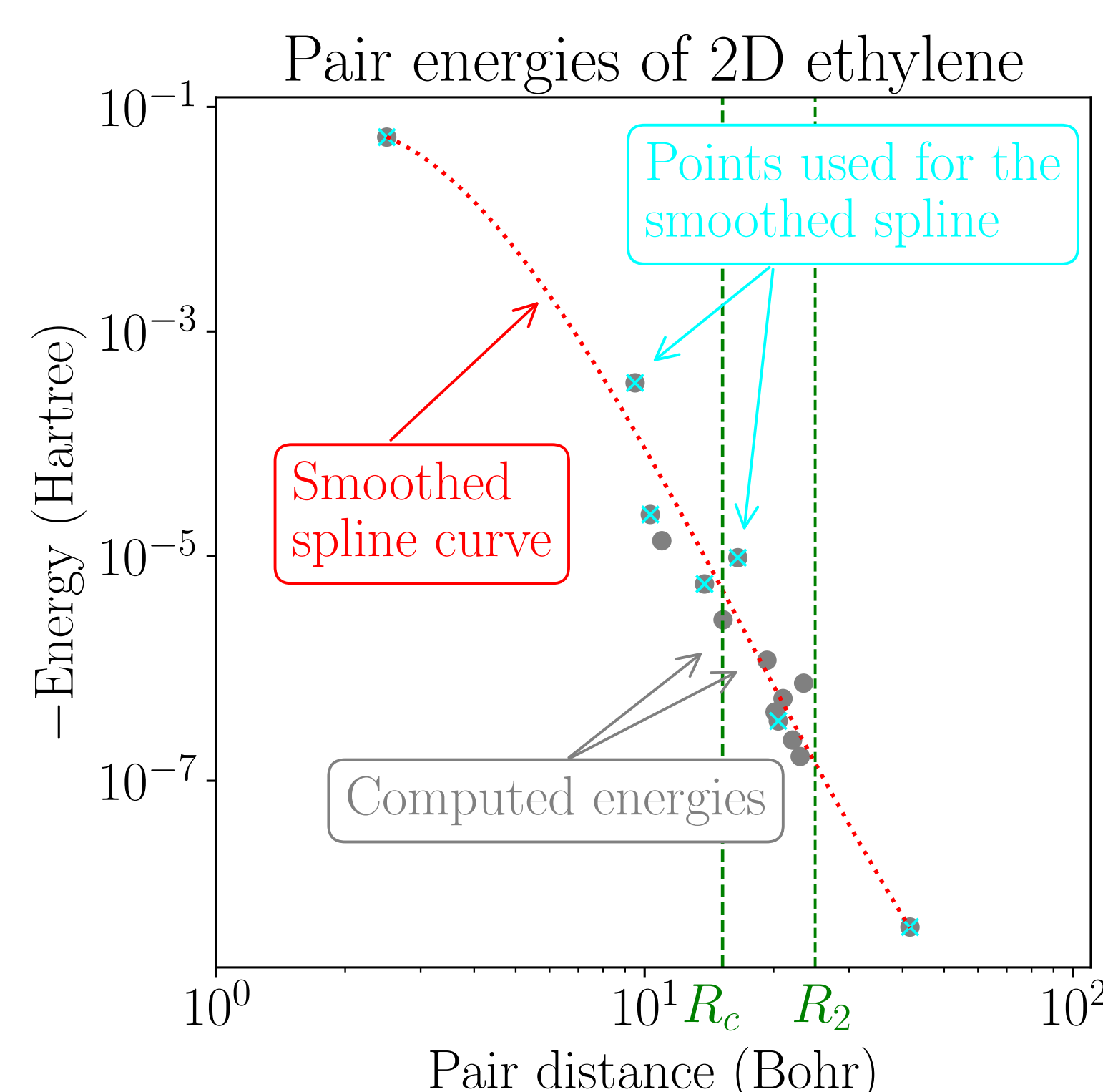


Figure 2: A smoothed cubic spline [8, 9] is created from a subset of the data points chosen in the interval $[0, R_2]$ plus one or more extra points with a large pair distance $R_{PQ} > R_2$. The pair cutoff R_c is chosen such that the sum of all spline-estimated pair energies in the interval $[R_c, R_2]$ is smaller than a given energy threshold. The data is from Ref. [6], and was presented in Figure 5 of that paper.

Algorithm 1: Determination of the pair cutoff distance based on an energy threshold, as we have suggested in Ref. [6].

- 1: Compute the energy E_{PQ} for a few pairs with pair distance $R_{PQ} \in [R_1, R_2]$.
- 2: Using these points, get a smoothed cubic spline function.
- 3: Estimate all pair energies in the interval $R_{PQ} \in [R_1, R_2]$ based on the smoothed spline.
- 4: Choose R_c such that the sum of all spline-estimated pair energies corresponding to pair distances $R_{PQ} \in [R_c, R_2]$ is smaller than a given energy threshold.

Finally, all pair energies corresponding to pair distances $R_{PQ} \leq R_c$ are computed 'exactly'.

The contribution of pair energies beyond R_2 may be estimated using a R^{-6} extrapolation based on linear regression with a least-squares error functional [7, 6].

4. Conclusions

Pros:

- Seems to work well for simple systems [6]
- Provides an error-based threshold

Open questions:

- For systems with a more complex geometry, the points may be more scattered. Possibly, different smoothed spline functions are needed for different space angles.
- Could perhaps a clustering algorithm be used to distinguish points that should belong to different smoothed spline functions?
- With more advanced algorithms, reproducibility and dependency on parameters might become a problem.

5. Acknowledgements

- Thanks to Lorenzo Maschio for useful discussions.
- Research Council of Norway, Grant No. 240698
- Research Council of Norway, Centre of Excellence Grant No. 262695
- The Norwegian Supercomputing Program (NO-TUR), Grant No. NN4654K for computing time.

References

- [1] T. Helgaker and P. Jørgensen and J. Olsen, Molecular Electronic-Structure Theory, John Wiley & Sons, Chichester (2000)
- [2] Ziolkowski *et al.*, J. Chem. Phys. **133**, 014107 (2010)
- [3] Kristensen *et al.*, J. Chem. Theory Comput. **7**, 1677 (2011)
- [4] Høyvik *et al.*, J. Chem. Phys. **136**, 014105 (2012)
- [5] Pisani *et al.*, J. Chem. Phys. **122**, 094113 (2005)
- [6] Rebolini *et al.*, J. Chem. Theory Comput. **14**, 2427 (2018)
- [7] Pisani *et al.*, J. Chem. Theory Comput. **29**, 2113 (2008)
- [8] Dierckx, J. Comput. Appl. Math. **1**, 165 (1975)
- [9] The scipy community, scipy.interpolate.splrep, <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.interpolate.splrep.html>, accessed Jan 5, 2018.