Using Context Sensitive Parameters in Semiempirical Models

Matteus Tanha[1], Shiva Kaul[2], Alex Cappiello[1], Craig Barretto[1], Geoffrey J. Gordon[2] and David Yaron[1]

[1]Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA

[2]Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA

Goal

There are two aspects of molecular structure that can be used to reduce the cost of electronic structure calculations.

- **Nearsightedness:** whereby interactions become simpler at long range. This is the basis for linear scaling methods.
- **Molecular similarity:** whereby molecular fragments behave similarly in similar environments. Our goal is to develop new means to take advantage of similarity to lower computational costs.

Our strategy embeds parameters in a low-level (LL) low-cost theory and adjust these to obtain agreement with a high-level (HL) theory. Such semiempirical parameters can be expected to work over a limited range of molecules. Our goal is to develop models of molecular fragments that have a wider range of applicability through the use of parameters that are sensitive to the current molecular context. Here, the context is captured by the atomic charges and the bond orders.

Differences from other efforts include:

➢ Parameters are embedded into a LL *ab initio* theory, rather than through a standard semiempirical form. The complexity of the LL method can thus be increased if needed.

➢ Rather than ignore certain classes of integrals, we retain all integrals and modify certain subclasses.

➢ Agreement is sought with expectation values of each available operator, instead of the total energy.

➢ We do not modify nuclear-nuclear interactions, and instead seek a fully electronic model.

Structure of the Model

Parameters are embedded in a minimal basis (STO-3G) *ab initio* model and adjusted to obtain agreement with a split-valance (6-31G) model.

One-electron operators

KE: Kinetic Energy

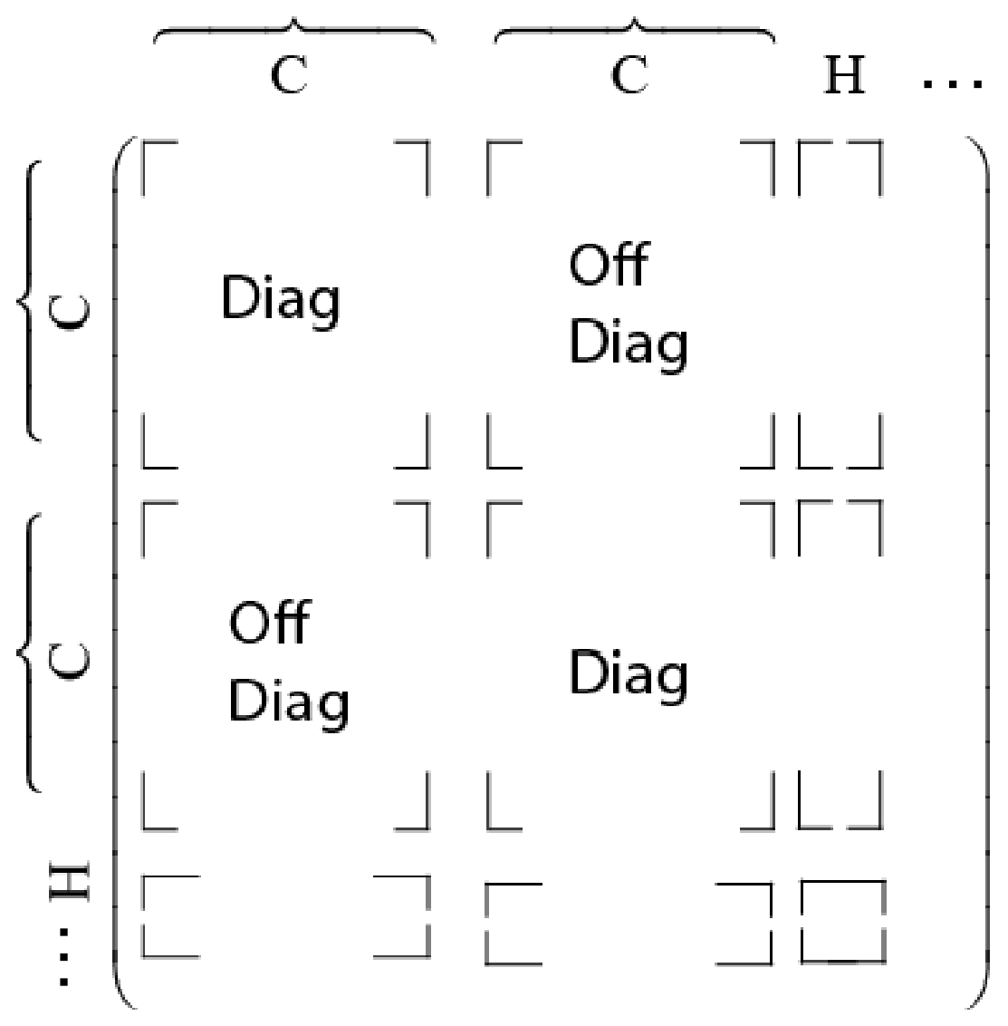
EN: Electron-nuclear interactions

Number of x parameters:

KE (4): Diag C and H
OffDiag CC and CH

EN-C(3): Diag C
OffDiag CC and CH

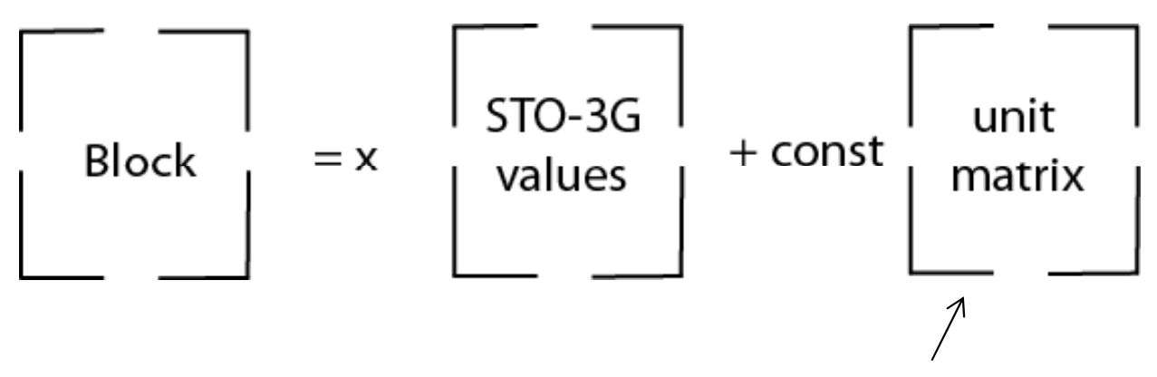
EN-H(2): Diag H
OffDiag HC



Some categories of STO-3G matrix elements are replaced with parameterized versions. The remainder are left as is.

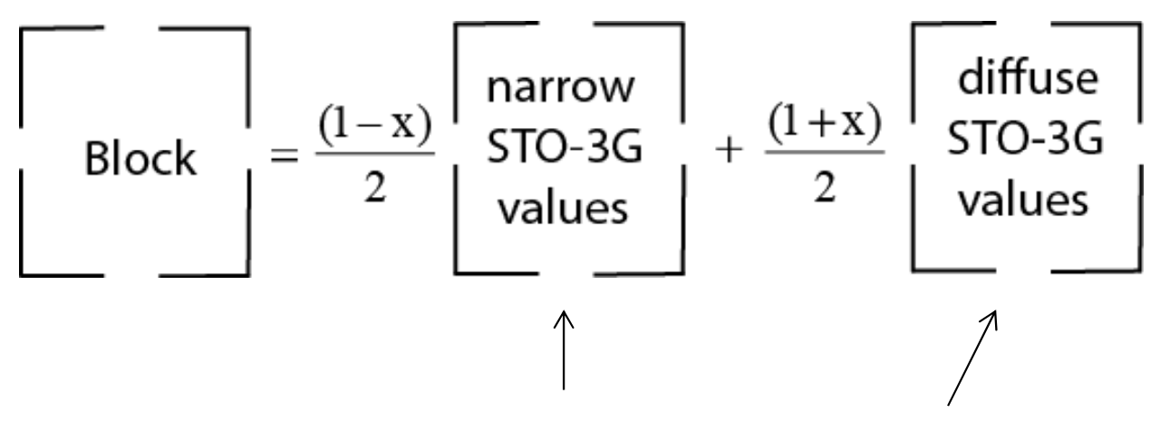
Two different parameterizations are tried

Scaling



Included only for diagonal blocks of KE and EN

Interpolation



Integrals obtained with Slater exponent increased by 5% or decreased by 10%. (X typicall lies outside the range -1..1, so extrapolations are occuring).

Two-electron integrals

(ij | kl)

i j k l all on same atom:
treated same as diagonal one-electron integrals
2 x-parameters: C and H

i j on one atom, k l all another atom:
treated same as off - diagonal one-electron integrals
3 x-parameters: CC, CH and HH

All other integrals retain their STO-3G values

Use of hybrid orbitals

For one-electron operator matrix elements that are between atoms (off-diagaonal), we use hybrid orbitals:

i) Rotate to appropriate (sp^3 sp^2 ..) hybrid orbitals pointing along the bond

ii) Scale matrix elements between these rotated orbitals

iii) Rotate back to original basis

Fits are done to operator expectation values

<KE> Kinetic energy of entire molecule

$\langle EN_A \rangle$ Electron-nuclear attraction to each nucleus A

$\langle E_2 \rangle$ Two electron energy of entire molecule

Data

A library of data is generated for a set of target molecules by varying both the molecular geometry and electrostatic environment. Here, we are working towards a model of hydrocarbons that will be applicable for situations where electron donors/acceptors substantially modify the charge on the atoms.

Methane

25 randomly generated geometries:
bond lengths 1.12 ± 0.15Å
bond angles 115° ± 6°
dihedrals 120° ± 7°

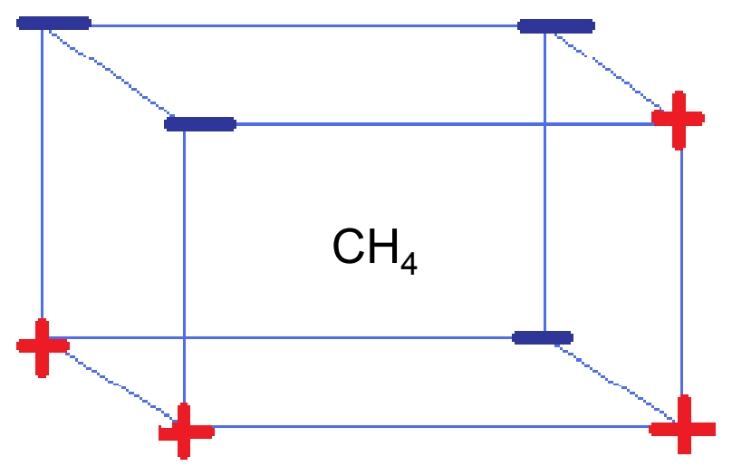
Training set is 10 geometries
Test set is 15 (methane only) or
10 (multi fits)

Ethane

	R_{CH}	R_{CC}	ϕ
4 Training	1.54	1.12	60
	1.54	1.12	0
	1.69	1.12	60
	1.54	1.27	60
3 Test	1.54	1.12	30
	1.39	1.12	60
	1.54	0.97	60

Electrostatic environment

Each molecular geometry is surrounded by a cube of random point charges that are meant to perturb the electronic density in a manner similar to what the molecular fragments will experience in large systems. This includes both inductive effects from electronic acceptors/donors and polarization effects from the surroundings.

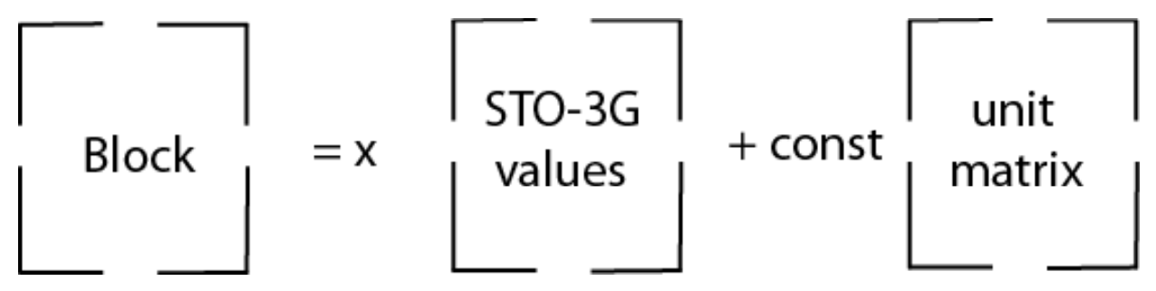


100 environments were generated, from which 6 train and 6 test environments were selected based on spread of induced effects.

For the hydrocarbons studied here, the variance of the randomly generated point charges was chosen to induce variation in the Mulliken charges that are similar to the charges induced by OH and F groups (0.2 amu).

Use of context dependent parameters

Split-valence basis sets allow the electronic density to expand and contract as the molecular structure and environment change. Making the scaling (or mixing) parameters a function of the context of the atom may be able to capture these effect for a specific molecular fragment.



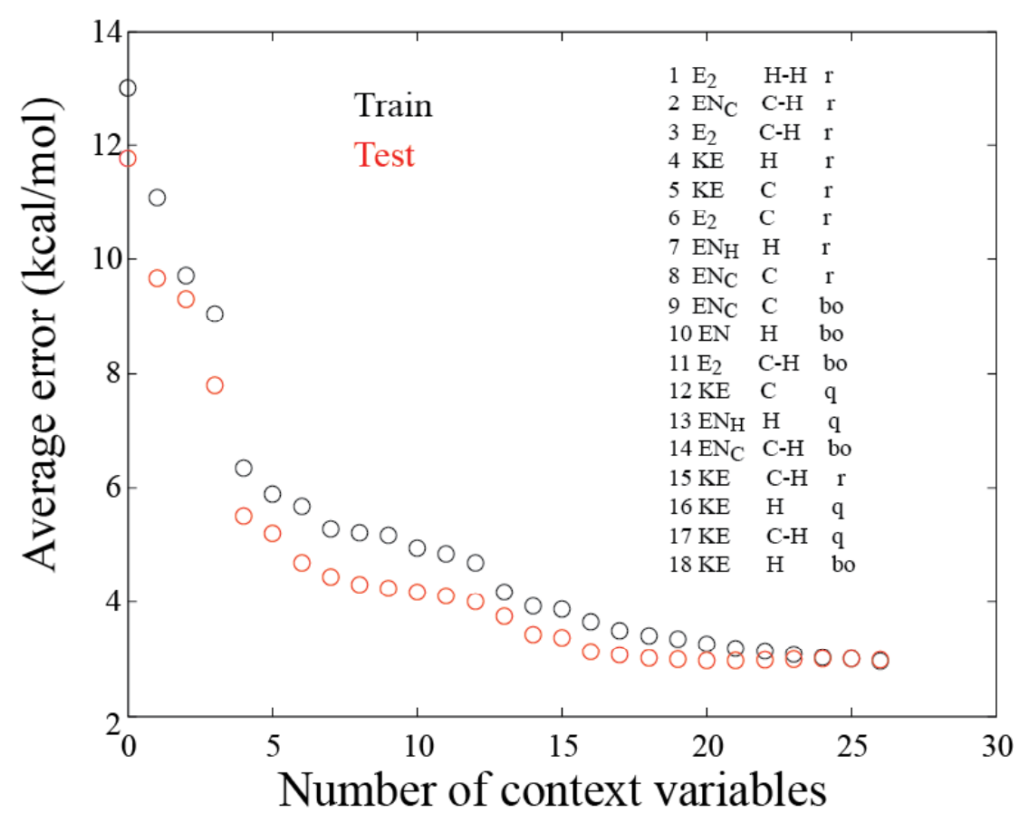
For **diagonal terms**, x depends on:
q: atomic charge
r: average bond length
bo: average bond order

For **off-diagonal terms**, x depends on
r: bond length
bo: bond order
q: polarity (charge difference)

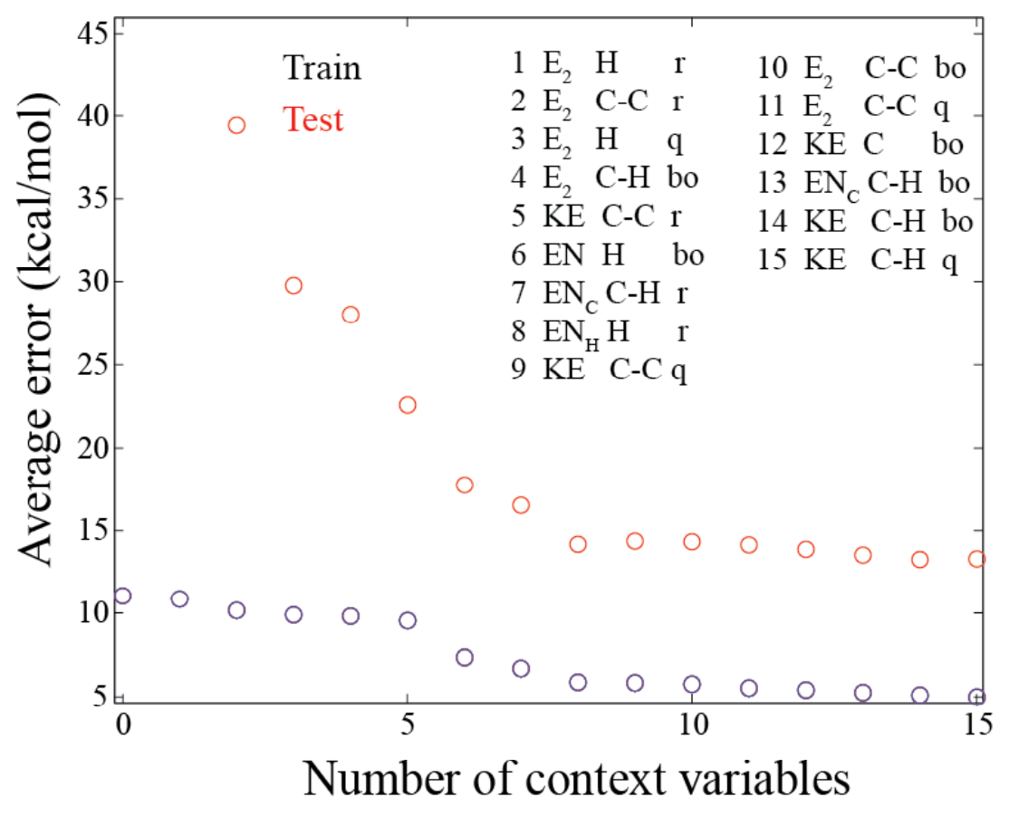
Forward selection of context variables:

```
Loop over all unused context variables:
  - Fit model, with this one additional context variable, to the training data
  - Determine residual when resulting model is applied to the test data
Select the context variable that leads to the smallest residual, and add to the model
```

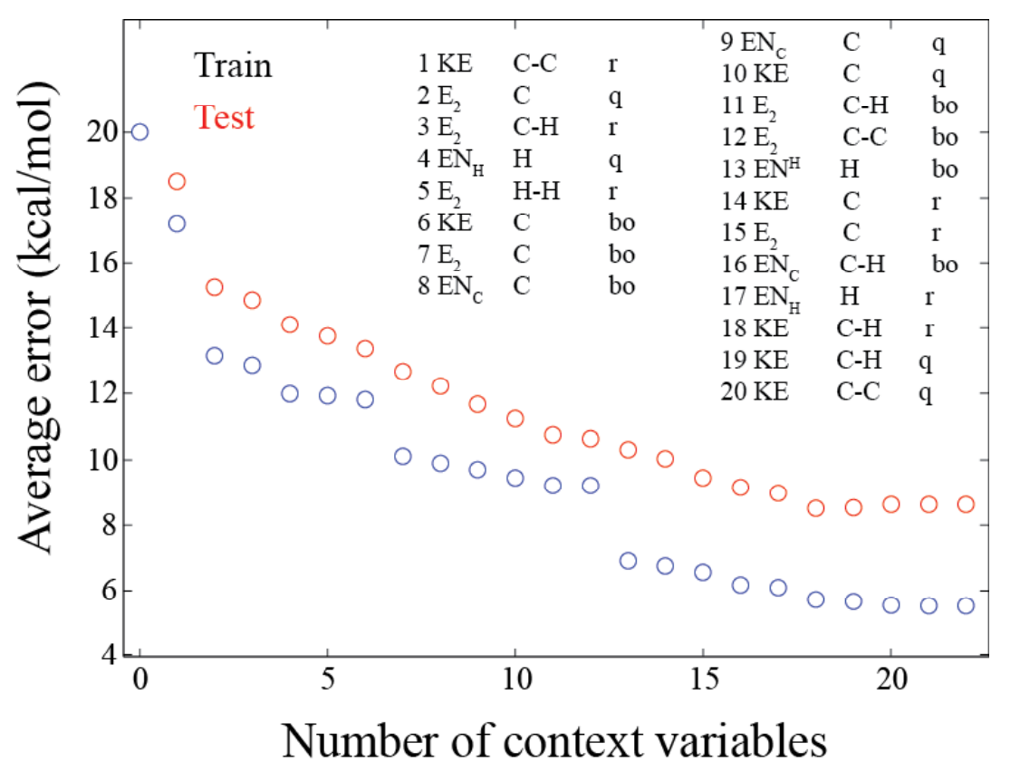
Methane



Ethane



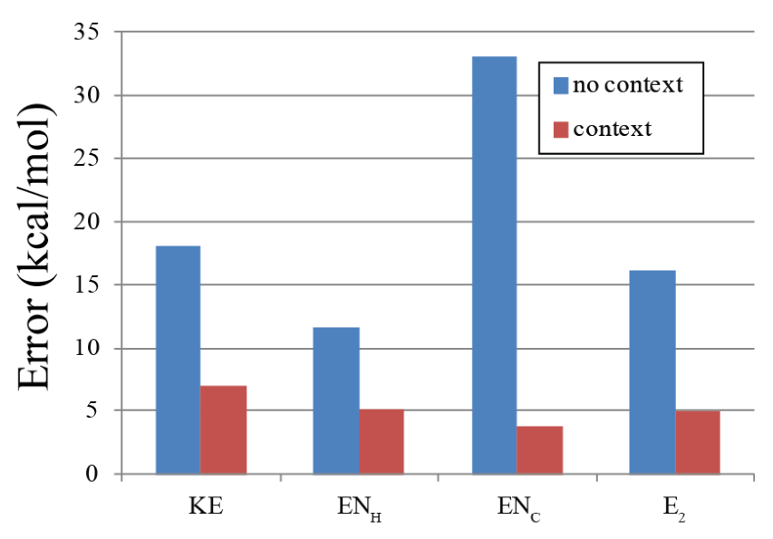
Methane and Ethane



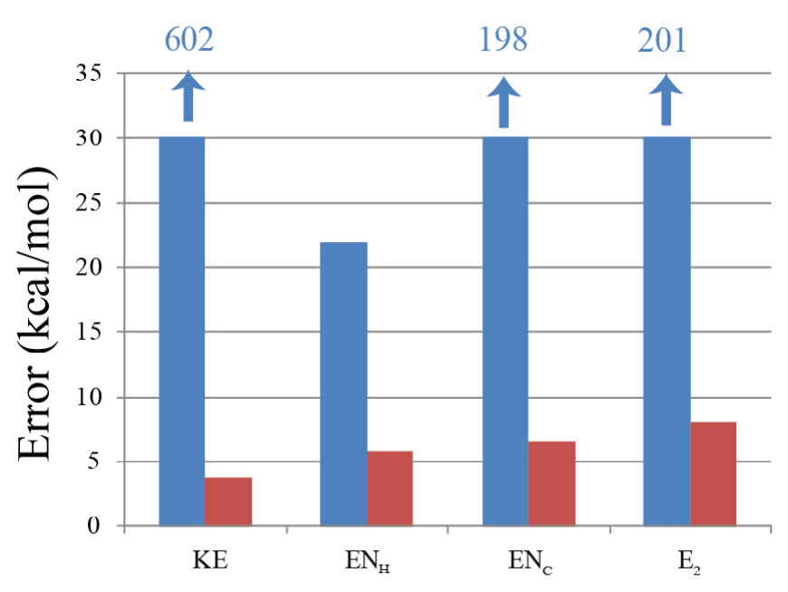
Analysis of results from simultaneous fit of methane and ethane

Error by operator type

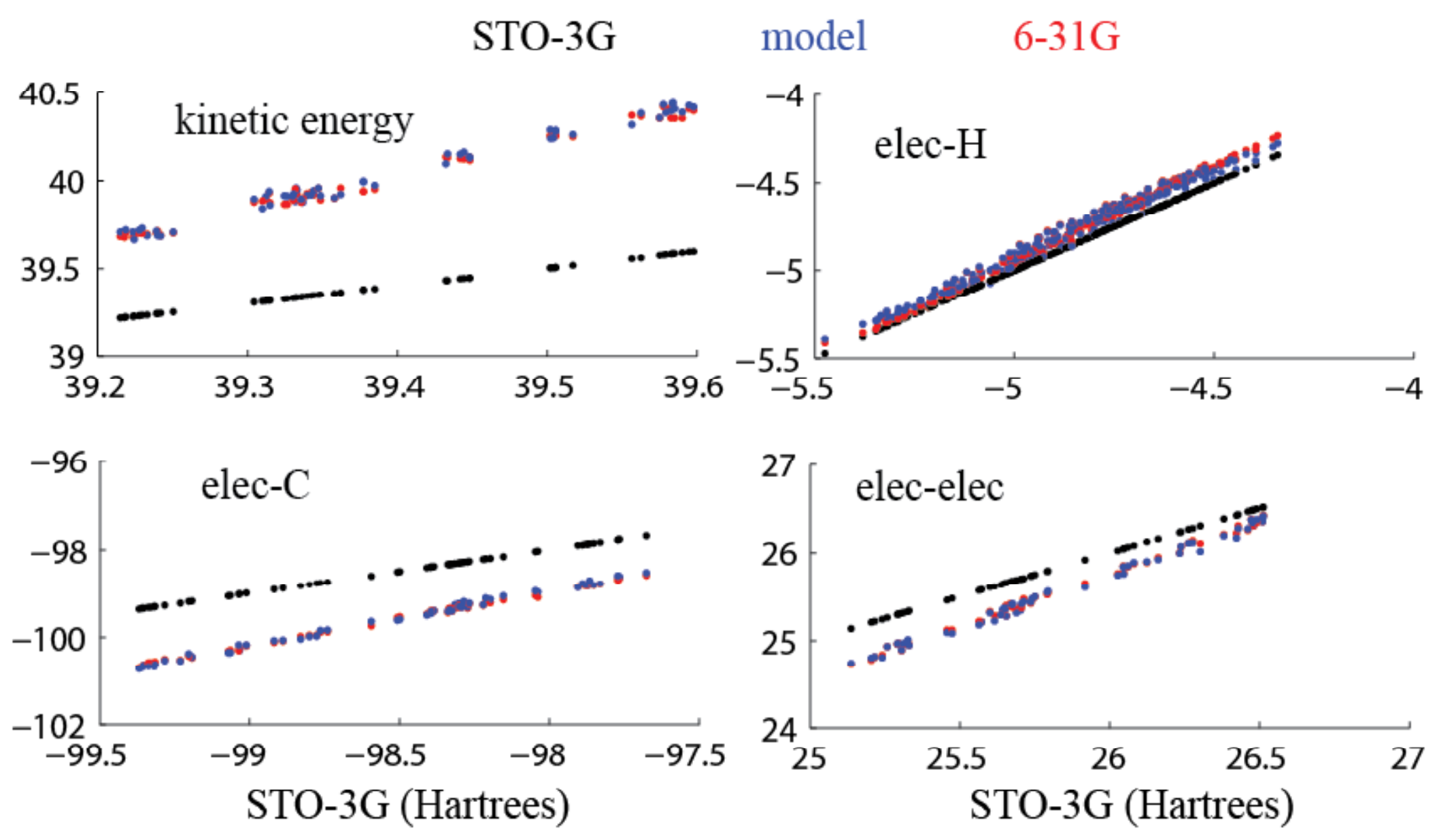
Methane



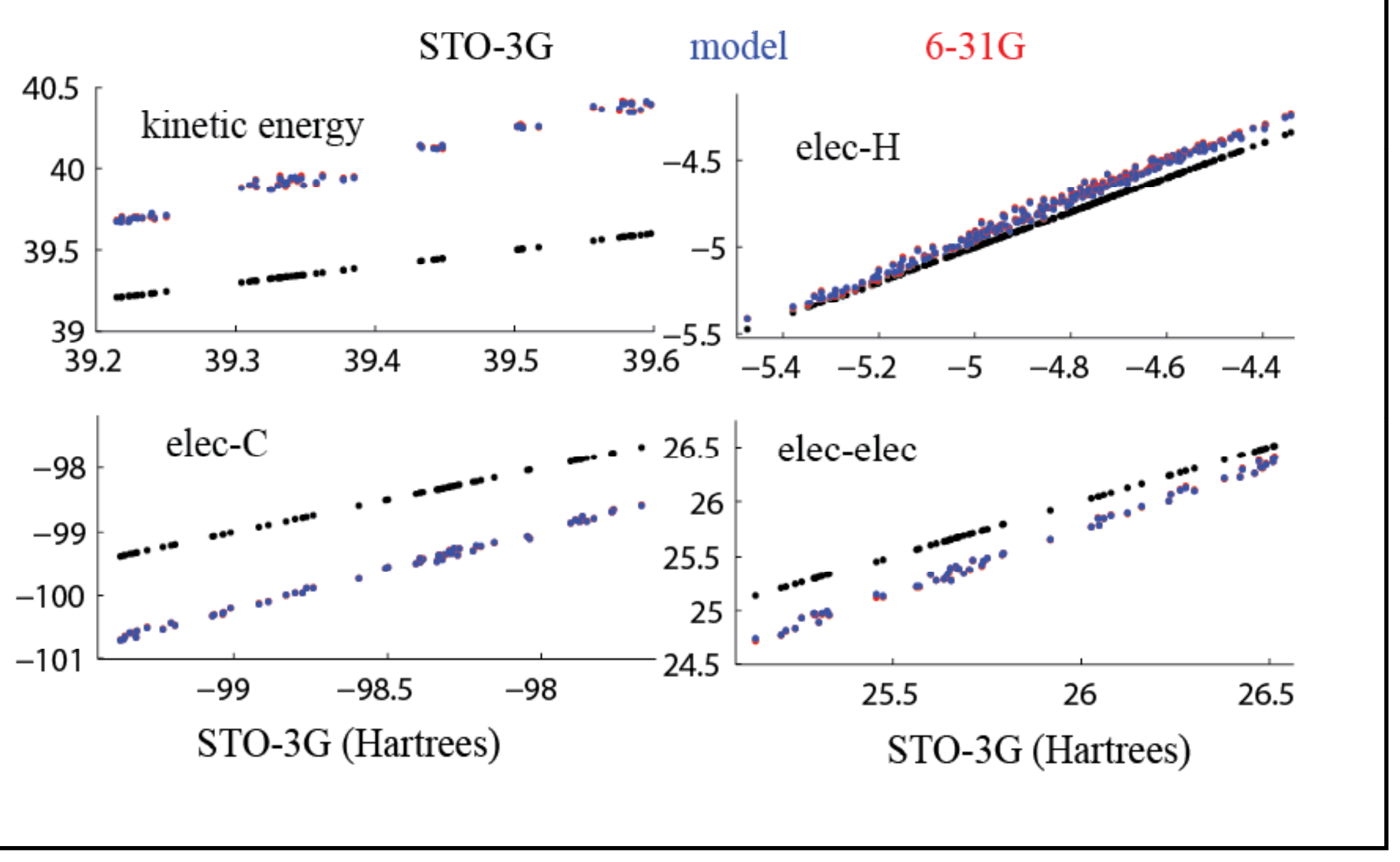
Ethane



Methane without context



Methane with context



Future

➢ Fits of more extensive sets of molecules

➢Extend parameterization to other molecular fragments

➢ Explore improved forms of two-electron parameterizations

➢ Test whether correlated solutions of the parameterized LL model agree with correlated solutions of the HL model. This is part of the rationale for seeking agreement between LL and HL at only the Hartree-Fock level.

Bibliography

(1) Ediz, V.; Monda, A. C.; Brown, R. P.; Yaron, D. J. *Journal of Chemical Theory and Computation* **2009**, *5*, 3175-3184.

(2) Janesko, B. G.; Yaron, D. *The Journal of chemical physics* **2004**, *121*, 5635-45.

(3) Sastry, K.; Johnson, D. D.; Thompson, A. L.; Goldberg, D. E.; Martinez, T. J.; Leiding, J.; Owens, J. *Materials and Manufacturing Processes* **2007**, *22*, 553-561.

(4) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Physical Review B* **1998**, *58*, 7260-7268.

(5) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *The Journal of Physical Chemistry B* **2001**, *105*, 569-585.