Self-consistent implementation of meta-GGA exchange-correlation functionals within the ONETEP linear-scaling DFT code

James C. Womack, Chris-Kriton Skylaris
Chemistry, Faculty of Natural & Environmental Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

RSC Theoretical Chemistry Group Conference, Nottingham, 2016
Overview

Background & Motivation
- Linear-scaling DFT
- Density functional approximations
- Why?

Theory
- ONETEP formalism
- Meta-GGA functionals
- Approaches to self-consistency

Implementation
- General meta-GGA framework in ONETEP
- Specific functionals implemented

Results & Conclusions
- Numerical validation
- Demonstration of linear-scaling
Linear-scaling density functional theory

- Orthogonalize KS orbitals extending over entire system
- Cubically scaling computational cost \( \mathcal{O}(N^3) \)

Conventional Kohn-Sham DFT

\[
E_{\text{KS}}[n] = T_s[n] + E_{\text{ext}}[n] + E_{\text{Hartree}}[n] + E_{\text{xc}}[n]
\]

\[
\left(-\frac{1}{2} \nabla^2 + V_{\text{KS}}(\mathbf{r})\right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r})
\]

Linear-scaling density matrix DFT

\[
\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}')
\]

\[
\rho(\mathbf{r}, \mathbf{r}) = n(\mathbf{r})
\]

\[
T_s[n] = \int \mathrm{d}\mathbf{r} \left[-\frac{1}{2} \nabla_r^2 \rho(\mathbf{r}, \mathbf{r}')\right]_{\mathbf{r}=\mathbf{r}'}
\]

\( \mathcal{O}(N) \)

- The density matrix is “nearsighted”
- For insulators:
  \[
  \rho(\mathbf{r}, \mathbf{r}') \sim e^{-\gamma |\mathbf{r} - \mathbf{r}'|}
  \]
- Exploit this to obtain linear scaling computational cost

*see e.g. S. Ismail-Beigi and T.A. Arias, Phys. Rev. Lett. 82, 2127 (1999)
Approximate exchange-correlation functionals

- No practical exact form is known
  - Approximations are required
- Ideally, these should be
  - Accurate
  - Computationally efficient
- “Jacob's ladder” of approximations
  - Hierarchical series
  - New ingredients on each “rung”
  - Higher rungs are more flexible
- Which is the best tool for the job?

\[ E_{xc}[n] = \int \text{d}r \; \epsilon_{xc}(n(r), \nabla n(r), \tau(r)) \]

Motivation: The meta-GGA “Goldilocks zone”

- Non-local ingredients (e.g. exact exchange) pose a challenge to linear scaling*
- Meta-GGAs sit in a “Goldilocks zone”
  - Well-balanced cost vs. accuracy
  - Simple semi-local form
- Add kinetic energy density, $\tau$
  - More flexible functional forms
  - Greater constraint satisfaction**
- Modern meta-GGAs offer impressive accuracy
  - e.g. B97M-V, M15-L, SCAN...

* Linear scaling exact exchange has been achieved in ONETEP but with a relatively large prefactor:

** For more on constraint satisfaction:
Order-N Electronic Total Energy Package

\[ \rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \varphi_\alpha(\mathbf{r}) K_{\alpha\beta} \varphi_\beta^*(\mathbf{r}') \]

\[ \psi_i(\mathbf{r}) = \sum_\alpha \varphi_\alpha(\mathbf{r}) M_i^\alpha \]

\[ \varphi_\alpha(\mathbf{r}) = \sum_m D(\mathbf{r} - \mathbf{r}_m) c_{m\alpha} \]

\[ K_{\alpha\beta} = \sum_i^N M_i^\alpha f_i M_i^\dagger \beta \]

NGWFs
“Non-orthogonal generalized Wannier functions”
- Strictly localized orbitals
- Self-consistently optimized
- Periodic cardinal sinc “psinc” basis

Density kernel
- Sparse matrix
- Truncated above cut-off length, i.e.
\[ R_{\alpha\beta} > r_{\text{cut}} \implies K_{\alpha\beta} = 0 \]

Total energy minimization
- Via two nested loops, i.e.

\[ E_{\text{min}} = \min_{\{\varphi_\alpha\}} \left( \min_K E[K, \{\varphi_\alpha\}] \right) \]

Initial guess
- Optimize \( K \)
- Converged?
  - Yes: Optimize NGWFs
  - No
    - Converged?
      - Yes: Energy minimized
      - No

Psinc basis
- Orthogonal
- Bandwidth-limited delta functions on a regular real-space grid
- Finite sum over plane waves, i.e.

\[ D_m(r) = \frac{1}{N_{\text{grid}}} \sum_{G_p} e^{iG_p \cdot (r - r_m)} \]
Self-consistent meta-GGA evaluation

\[ V_{xc}(r) = \frac{\delta E_{xc}}{\delta n(r)} = \frac{\delta}{\delta n(r)} \int dr \, \epsilon_{xc}(n, \nabla n, \tau[n]) \rightarrow \frac{\delta \tau[n]}{\delta n(r)} = ? \]

- Unknown dependence of \( \tau \) on the density prevents direct functional derivative evaluation
- There are **three** main approaches for overcoming this:

1. **Give up!**
   - Perform meta-GGA calculations non-self-consistently
   - Use orbitals optimized using a non-\( \tau \)-dependent functional (LDA, GGA)
   - Yields surprisingly good results

2. **Optimized effective potential**
   - Obtain a local, multiplicative potential which satisfies
     \[ \frac{\delta E_{xc}}{\delta V_{xc}} = 0 \]
     by solving the OEP integral equation
   - Formidable mathematical machinery (even with approximations)

Example:

Example:
3. FDO approximation

Functional Derivatives (of $\tau$-dependent functionals) with respect to the Orbitals*

Use the functional derivative chain rule to obtain

$$
\frac{\delta E_{xc}^{mGGA}}{\delta \psi_i(r)} = \int dr' \frac{\delta E_{xc}^{mGGA}}{\delta n(r')} \frac{\delta n(r')}{\delta \psi_i(r)}
$$

The functional derivative

$$
\frac{\delta n(r')}{\delta \psi_i(r)} \approx 2\psi_i(r)\delta(r' - r)
$$

allows us to derive the following relationship

$$
\frac{1}{2} \frac{\delta E_{xc}^{mGGA}}{\delta \psi_i(r)} \approx \frac{\delta E_{xc}^{mGGA}}{\delta n(r)} \psi_i(r)
$$

This functional derivative is approximate, since it neglects the dependence of the KS orbitals on each other via the density, i.e.

$$
n(r) = \sum_{i=1}^{N} |\psi_i(r)|^2
$$

* AKA “NNH” approach after the creators:

For a detailed account, see:
We know the functional dependence of $\tau$ on the KS orbitals

$$\tau(\mathbf{r}) = \frac{1}{2} \sum_{i} N_{\text{occ}} |\nabla \psi_i(\mathbf{r})|^2$$

and can therefore evaluate

$$\frac{\delta E_{\text{xc}}^{\text{mGGA}}}{\delta \psi_i(\mathbf{r})}$$

and obtain our FDO XC potential using

$$1 \frac{\delta E_{\text{xc}}^{\text{mGGA}}}{2 \delta \psi_i(\mathbf{r})} \approx V_{\text{xc}} \psi_i(\mathbf{r})$$

$$E_{\text{xc}}[n] = \int \text{d}\mathbf{r} \epsilon_{\text{xc}}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r}))$$

Non-multiplicative, orbital-specific potential

$$V_{\text{xc}}^{\text{GGA}} \psi_i(\mathbf{r}) \approx V_{\text{xc}}^{\text{GGA}} \psi_i(\mathbf{r}) + \left\{ \hat{V}_{\text{xc}}^\tau \psi_i(\mathbf{r}) \right\}$$

$$V_{\text{xc}}^{\text{GGA}} = \left\{ \frac{\partial \epsilon_{\text{xc}}}{\partial n} - \nabla \cdot \left( \frac{\partial \epsilon_{\text{xc}}}{\partial \nabla n} \right) \right\}$$

$$\left\{ \hat{V}_{\text{xc}}^\tau \psi_i(\mathbf{r}) \right\} = -\frac{1}{2} \nabla \cdot \left( \frac{\partial \epsilon_{\text{xc}}}{\partial \tau} \nabla \psi_i(\mathbf{r}) \right)$$

This approach has the advantages of

- Theoretical simplicity
- Computational simplicity
- Wide use in other code (i.e. well-tested)

Implementation overview

\[ \tau(\mathbf{r}) = \frac{1}{2} \sum_{i}^{N_{\text{oocc}}} |\nabla \psi_i(\mathbf{r})|^2 \]

Kinetic energy density

\[ V_{xc}^{\text{GGA}}(\mathbf{r}) \psi_i(\mathbf{r}) \]

Exchange-correlation potential

Self-consistent meta-GGA functionals in ONETEP

\[ \langle \psi_i | \hat{V}_{xc}^\tau | \psi_j \rangle \]

Integral over \( \tau \)-dependent potential

\[ E_{xc}[n] \]

Functional forms

\[ \frac{\partial E_{xc}}{\partial K \eta \theta} \frac{\delta E_{xc}}{\delta \varphi^*_\alpha(\mathbf{r})} \]

Energy gradients
Kinetic energy density

We require $\tau$ in terms of the NGWFs and density kernel, i.e.

$$\tau(r) = \frac{1}{2} \sum_{i}^{N_{occ}} |\nabla \psi_i(r)|^2 \quad \rightarrow \quad \tau(r) = \frac{1}{2} \sum_{\alpha \beta} (\nabla \varphi_{\alpha}(r)) \cdot (K^{\alpha \beta} \nabla \varphi_{\beta}(r))$$

In practice, we evaluate $\tau$ via a multi-step process:

1. $t_{\alpha}(r) = \nabla \sum_{\beta} K^{\alpha \beta} \varphi_{\beta}(r)$
   Evaluate intermediate quantity, summing over NGWFs which overlap with $\alpha$

2. $\tau(r; \alpha) = (\nabla \varphi_{\alpha}(r)) \cdot t_{\alpha}(r)$
   Form part of $\tau$ local to each $\alpha$ NGWF

3. $\tau(r) = \frac{1}{2} \sum_{\alpha} \tau(r; \alpha)$
   Deposit each $\alpha$-component in simulation cell

Based on the approach used to evaluate the charge density in ONETEP:
We require an XC potential in terms of NGWFs, not KS orbitals:

\[ V_{xc}^{GGA} = \frac{\partial \epsilon_{xc}}{\partial n} - \nabla \cdot \left( \frac{\partial \epsilon_{xc}}{\partial \nabla n} \right) \]

\[ \left\{ \hat{V}_{xc}^{\tau} \varphi_{\alpha}(\mathbf{r}) \right\} = -\frac{1}{2} \nabla \cdot \left( \frac{\partial \epsilon_{xc}}{\partial \tau} \nabla \varphi_{\alpha}(\mathbf{r}) \right) \]

\[ \sum_{\alpha} \left[ \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \varphi_{\alpha}(\mathbf{r}) \right] M_{\alpha}^{i, \text{FDO}} \approx \sum_{\alpha} \left[ V_{xc}^{GGA} \varphi_{\alpha}(\mathbf{r}) + \left\{ \hat{V}_{xc}^{\tau} \varphi_{\alpha}(\mathbf{r}) \right\} \right] M_{\alpha}^{i} \]

Expand KS orbital in terms of NGWFs

\[ \frac{\delta E_{xc}}{\delta n(\mathbf{r})} \psi_{i}(\mathbf{r}) \approx V_{xc}^{GGA} \varphi_{\alpha}(\mathbf{r}) + \left\{ \hat{V}_{xc}^{\tau} \varphi_{\alpha}(\mathbf{r}) \right\} \]

"Per-NGWF” exchange-correlation potential
Integral over $\tau$-dependent potential

Hamiltonian matrix elements contain integrals over the $\tau$-dependent potential:

$$\langle \psi_i | \hat{V}_{xc}^\tau | \psi_j \rangle = \frac{1}{2} \int \text{d} \mathbf{r} \nabla \psi_i(\mathbf{r}) \cdot \left( \frac{\partial \epsilon_{xc}}{\partial \tau} \nabla \psi_j(\mathbf{r}) \right)$$

Gradient of this function avoided using integration-by-parts

Form originally used by NNH and in later implementations

Form used in ONETEP

$$\langle \varphi_\alpha | \hat{V}_{xc}^\tau | \varphi_\beta \rangle = -\frac{1}{2} \int_V \text{d} \mathbf{r} \varphi_\alpha(\mathbf{r}) \nabla \cdot \left( \frac{\partial \epsilon_{xc}}{\partial \tau} \nabla \varphi_\beta(\mathbf{r}) \right)$$

Apply divergence to vector field in reciprocal space

Integrand is zero outside of localization region of $\alpha$

No problem for ONETEP!
Energy gradients

\[ E_{\text{min}} = \min_{\{\varphi_{\alpha}\}} \left( \min_{K} E[K, \{\varphi_{\alpha}\}] \right) \]

- Gradient expressions are required for ONETEP's energy minimization scheme.
- Only minor changes and additions to the existing expressions are needed.

New integral type
\[ \langle \varphi_{\theta} | \hat{V}^{\tau}_{xc} | \varphi_{\eta} \rangle \]

\[ \frac{\partial E_{xc}}{\partial K_{\eta\theta}} = \int \text{d}r \, \varphi_{\theta}^{*}(r)V_{xc}^{GGA}(r)\varphi_{\eta}(r) + \int \text{d}r \, \varphi_{\theta}^{*}(r)\{\hat{V}^{\tau}_{xc}\varphi_{\eta}(r)\} \]

\[ \frac{\delta E_{xc}}{\delta \varphi_{\alpha}^{*}(r)} = [Q_{V,FDO}]^{\kappa\alpha}_{\kappa}(r) + V_{xc}^{GGA}(r)\varphi_{\kappa}(r)K^{\kappa\alpha} + \{\hat{V}^{\tau}_{xc}\varphi_{\kappa}(r)\}K^{\kappa\alpha} \]

- \( Q \) matrix depends on \( V_{\alpha\beta} = \langle \varphi_{\alpha} | V_{xc} | \varphi_{\beta} \rangle \)
- New \( \tau \)-dependent potential term
# Functionals

## PKZB
- An early meta-GGA (1999)
  - 1 empirical parameter
  - Evolved into TPSS (2003)
- Simple to implement
  - Based on PBE, but with \( \tau \)
- Widely available
  - Stable implementations to compare and test against
- Inconsistent accuracy
  - Superseded by modern meta-GGAs


## B97M-V
- Modern meta-GGA (2015)
  - Combinatorially designed form
  - Empirically fitted to very large data set
- Incorporates non-local VV10 functional
  - Available in ONETEP
- Excellent accuracy
  - Comparable to popular hybrid functionals
  - Broadly applicable
- Code provided by N. Mardirossian

### Thermochemical data

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSD (kcal/mol)</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>M06-2X</td>
<td>3.21</td>
<td>✓</td>
</tr>
<tr>
<td>ωB97X-V</td>
<td>3.60</td>
<td>✓</td>
</tr>
<tr>
<td>ωB97X-D</td>
<td>3.61</td>
<td>✓</td>
</tr>
<tr>
<td>B97M-V</td>
<td>3.93</td>
<td></td>
</tr>
<tr>
<td>B97-D2</td>
<td>3.97</td>
<td>✓</td>
</tr>
<tr>
<td>M11</td>
<td>3.97</td>
<td>✓</td>
</tr>
<tr>
<td>M06</td>
<td>4.18</td>
<td>✓</td>
</tr>
<tr>
<td>B3LYP-D3</td>
<td>4.66</td>
<td>✓</td>
</tr>
<tr>
<td>M06-L</td>
<td>5.63</td>
<td></td>
</tr>
<tr>
<td>TPSS-D3</td>
<td>6.45</td>
<td></td>
</tr>
<tr>
<td>VV10</td>
<td>9.81</td>
<td></td>
</tr>
<tr>
<td>PBE-D3</td>
<td>10.10</td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>3836.93</td>
<td></td>
</tr>
</tbody>
</table>

*RMSDs in kcal/mol for thermochemical and non-covalent datasets (combined training and test data). Based on Table V of N. Mardirossian and M. Head-Gordon, J. Chem. Phys. 142, 074111 (2015).*

### Non-covalent data

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSD (kcal/mol)</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>B97M-V</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>ωB97X-V</td>
<td>0.32</td>
<td>✓</td>
</tr>
<tr>
<td>M06-L</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>B97-D2</td>
<td>0.48</td>
<td>✓</td>
</tr>
<tr>
<td>ωB97X-D</td>
<td>0.54</td>
<td>✓</td>
</tr>
<tr>
<td>M11</td>
<td>0.55</td>
<td>✓</td>
</tr>
<tr>
<td>M06</td>
<td>0.57</td>
<td>✓</td>
</tr>
<tr>
<td>M06-2X</td>
<td>0.77</td>
<td>✓</td>
</tr>
<tr>
<td>B3LYP-D3</td>
<td>0.77</td>
<td>✓</td>
</tr>
<tr>
<td>TPSS-D3</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>VV10</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>PBE-D3</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>14.95</td>
<td></td>
</tr>
</tbody>
</table>

*aQZ basis set

NLC is not sufficient for good performance in either set.

Performs well against hybrid functionals!

**Aside: How good is B97M-V?**

2460 data points

**RMSDs in kcal/mol for thermochemical and non-covalent datasets (combined training and test data).**
Results I: Numerical testing

- Comparison of relative energies for B97M-V
  - QChem (Gaussian basis, BSSE corrected)
  - ONETEP (NGWFs/psincs)
  - B97M-rV*

- Complicated by several possible sources of error
  - Pseudopotential vs. all-electron
  - Periodic vs. open boundary conditions

### Water dimer binding energies in kcal/mol.

<table>
<thead>
<tr>
<th>Program</th>
<th>Basis</th>
<th>Program</th>
<th>Basis</th>
<th>Binding energy</th>
<th>E(ONETEP) – E(QChem)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PBE</td>
<td>B97M-rV</td>
<td>PBE</td>
<td>B97M-rV</td>
</tr>
<tr>
<td>QChem</td>
<td>aug-pc4</td>
<td>-4.89560</td>
<td>-4.88540</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONETEP</td>
<td>900eV / 8 bohr</td>
<td>-5.03667</td>
<td>-4.88986</td>
<td>-0.141</td>
<td>0.004</td>
</tr>
<tr>
<td>ONETEP</td>
<td>900eV / 12 bohr</td>
<td>-5.11562</td>
<td>-4.92943</td>
<td>-0.220</td>
<td>0.044</td>
</tr>
<tr>
<td>ONETEP</td>
<td>900eV / 15 bohr</td>
<td>-5.11822</td>
<td>-5.10057</td>
<td>-0.223</td>
<td>0.215</td>
</tr>
<tr>
<td>ONETEP</td>
<td>1200eV / 8 bohr</td>
<td>-5.02495</td>
<td>-4.85846</td>
<td>-0.129</td>
<td>0.027</td>
</tr>
<tr>
<td>ONETEP</td>
<td>1200eV / 12 bohr</td>
<td>-5.11374</td>
<td>-4.90818</td>
<td>-0.218</td>
<td>0.023</td>
</tr>
<tr>
<td>ONETEP</td>
<td>1200eV / 15 bohr</td>
<td>-5.11674</td>
<td>-5.09488</td>
<td>-0.221</td>
<td>0.209</td>
</tr>
</tbody>
</table>

Water dimer binding energies in kcal/mol. QChem results provided by N. Mardirossian (UC Berkeley).
Results I: Numerical testing

- The 8 bohr results seem “too good to be true”!
  - Why are the results worse for larger NGWF radii?
- Use ONETEP's “extended” NGWF functionality.
- Possibly a favourable cancellation of errors (unfortunately).

Water dimer

Further ONETEP calculation details:
- 24Å cubic simulation cell
- Norm-conserving pseudopotentials*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E(ONETEP,local) – E(ONETEP,extended)</td>
<td></td>
</tr>
<tr>
<td>PBE</td>
<td>B97M-rV</td>
<td></td>
</tr>
<tr>
<td>900eV / 8.0 bohr</td>
<td>-0.141</td>
<td>-0.004</td>
</tr>
<tr>
<td>900eV / 12.0 bohr</td>
<td>-0.220</td>
<td>-0.044</td>
</tr>
<tr>
<td>900eV / 15.0 bohr</td>
<td>-0.223</td>
<td>-0.215</td>
</tr>
<tr>
<td>900eV / Extended</td>
<td>-0.223</td>
<td><strong>-0.249</strong></td>
</tr>
<tr>
<td>1200eV / 8.0 bohr</td>
<td>-0.129</td>
<td><strong>0.027</strong></td>
</tr>
<tr>
<td>1200eV / 12.0 bohr</td>
<td>-0.218</td>
<td>-0.023</td>
</tr>
<tr>
<td>1200eV / 15.0 bohr</td>
<td>-0.221</td>
<td>-0.209</td>
</tr>
<tr>
<td>1200eV / Extended</td>
<td>-0.217</td>
<td><strong>-0.251</strong></td>
</tr>
</tbody>
</table>

Energy differences in kcal/mol. QChem results provide by N. Mardirossian (UC Berkeley).
*Rappe-Bennett pseudopotential library: http://www.sas.upenn.edu/rappegroup/research/psp.html
Results II: Computational efficiency

Computational cost of DFT calculations on amyloid fibrils of increasing size

Structure provided by authors of J.T. Berryman, S.E. Radford, and S.A. Harris, Biophys. J. 97, 1 (2009)
Conclusions

- Self-consistent meta-GGA functional evaluation in ONETEP
  - Theoretical and computational framework implemented and tested
  - Further functional forms can easily be added
  - Meta-GGAs can be applied to very large systems

- Provisional numerical agreement with other implementations
  - Further testing and validation is ongoing

- Linear-scaling computational cost
  - Self-consistent evaluation of meta-GGA functionals
  - Demonstrated for systems with 1000s of atoms
The future

- Extensions to the framework
  - Support for projector augmented waves (PAW)
  - Support for hybrid meta-GGA functionals using existing exact exchange in ONETEP*

- Additional meta-GGA functional forms
  - TPSS, SCAN, M15-L are of particular interest

- Public availability
  - Self-consistent meta-GGA support will be made available in a future version of ONETEP

Acknowledgements

• People
  – Martin Head-Gordon (Berkeley)
  – Narbe Mardirossian (Berkeley)
  – Jacek Dziedzic (Southampton)

• Compute resources
  – University of Southampton IRIDIS HPC facility

• Funding
  – Engineering and Physical Sciences Research Council

• University of Southampton