

Exploiting sparsity in large scale quantum chemical computations - implementation of LT-AO-MP2 formalism

M. Makowski, G. Mazur, K. Kowalczyk, J. Sumera

Faculty of Chemistry, Jagiellonian University



KU KDM 2009
Zakopane

Introduction

- Chemists, physicists, molecular biologists want to get accurate description of bigger and bigger molecular systems
- Theoreticians keep providing better and better models

this is not enough to get results for real-world systems

what we really need are computational methods which scale well
with the size of the system

Introduction

- Chemists, physicists, molecular biologists want to get accurate description of bigger and bigger molecular systems
- Theoreticians keep providing better and better models

this is not enough to get results for real-world systems

**what we really need are computational methods which scale well
with the size of the system**

Problem solved?

Even at the Hartree-Fock level of theory there were a few bottlenecks

- evaluation of two-electron integrals ($\mathcal{O}(N^2)$)
- diagonalization ($\mathcal{O}(N^3)$)

The cost can be reduced to linear employing

- fast multipole method
- efficient sorting and prescreening techniques
- direct optimization instead of diagonalization

Unfortunately, Hartree-Fock lacks

- correlation
- dispersion interaction
- ...

Problem solved?

Even at the Hartree-Fock level of theory there were a few bottlenecks

- evaluation of two-electron integrals ($\mathcal{O}(N^2)$)
- diagonalization ($\mathcal{O}(N^3)$)

The cost can be reduced to linear employing

- fast multipole method
- efficient sorting and prescreening techniques
- direct optimization instead of diagonalization

Unfortunately, Hartree-Fock lacks

- correlation
- dispersion interaction
- ...

Problem solved?

Even at the Hartree-Fock level of theory there were a few bottlenecks

- evaluation of two-electron integrals ($\mathcal{O}(N^2)$)
- diagonalization ($\mathcal{O}(N^3)$)

The cost can be reduced to linear employing

- fast multipole method
- efficient sorting and prescreening techniques
- direct optimization instead of diagonalization

Unfortunately, Hartree-Fock lacks

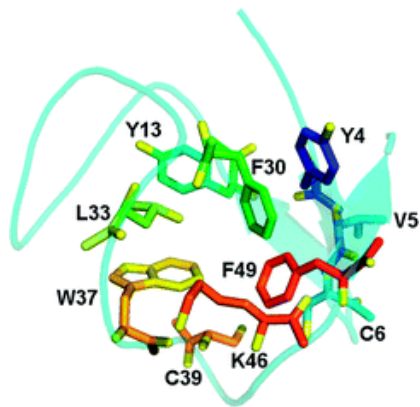
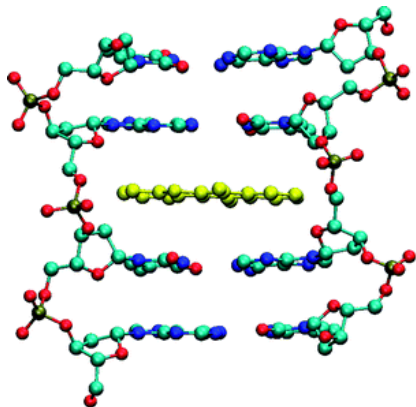
- correlation
- dispersion interaction
- ...

DFT to the rescue?

- Density Functional Theory
 - inherits Hartree-Fock efficiency
 - ships with correlation included
- Still, this is not enough - DFT fails at
 - charge transfer processes
 - description of dispersion interaction

Does it really matter?

Dispersion effects may seem negligible, but Nature begs to differ



MP2

- second order perturbation theory providing correction to HF energy and wavefunction
- the first step in the ladder of post-HF methods
- reproduces most of electron correlation
- includes dispersion effects

Canonical MP2

- energy correction

$$E_2 = - \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

- crucial step

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma} C_{\mu i} C_{\nu a} C_{\lambda j} C_{\sigma b} (\mu\nu|\lambda\sigma)$$

- AO integrals tensor is sparse ($\mathcal{O}(N^2)$)
- C matrix and MO integrals tensor are dense
- time complexity: $\mathcal{O}(N^5)$, memory complexity: $\mathcal{O}(N^4)$
- calculations for large systems are practically impossible

Canonical MP2

- energy correction

$$E_2 = - \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

- crucial step

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma} C_{\mu i} C_{\nu a} C_{\lambda j} C_{\sigma b} (\mu\nu|\lambda\sigma)$$

- AO integrals tensor is sparse ($\mathcal{O}(N^2)$)
- C matrix and MO integrals tensor are dense
- time complexity: $\mathcal{O}(N^5)$, memory complexity: $\mathcal{O}(N^4)$
- **calculations for large systems are practically impossible**

How to get better computational methods

Correlation in non-metallic systems is extremely local. We need to exploit the locality to get a lean and fast MP2

- Explicit division of the system into subsystems
 - treats each subsystem accurately
 - approximates interactions between subsystems
 - calculations are fast, but the division might be ambiguous
 - error is difficult to control
- Exploiting inherent sparsity
 - reformulation of the formalism required to uncover sparsity
 - no additional physical approximations are made
 - high accuracy and easy to control error

How to get better computational methods

Correlation in non-metallic systems is extremely local. We need to exploit the locality to get a lean and fast MP2

- Explicit division of the system into subsystems
 - treats each subsystem accurately
 - approximates interactions between subsystems
 - calculations are fast, but the division might be ambiguous
 - error is difficult to control
- Exploiting inherent sparsity
 - reformulation of the formalism required to uncover sparsity
 - no additional physical approximations are made
 - high accuracy and easy to control error

How to get better computational methods

Correlation in non-metallic systems is extremely local. We need to exploit the locality to get a lean and fast MP2

- Explicit division of the system into subsystems
 - treats each subsystem accurately
 - approximates interactions between subsystems
 - calculations are fast, but the division might be ambiguous
 - error is difficult to control
- Exploiting inherent sparsity
 - reformulation of the formalism required to uncover sparsity
 - no additional physical approximations are made
 - high accuracy and easy to control error

Laplace transformation

- definition

$$\mathcal{L}[f](x) = \int_0^{\infty} f(t)e^{-tx} dt$$

- transformation of constant function

$$\mathcal{L}[1](x) = \int_0^{\infty} e^{-tx} dt = \frac{1}{x}$$

- the transformation is applied to

$$E_2 = - \sum_{ij}^{\text{occ}} \sum_{ab}^{\text{virt}} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

LT-MP2

Energy correction

$$E_2 = - \int_0^\infty \sum_{iajb} (ia|jb)[2(ia|jb) - (ib|ja)] e^{-(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)t} dt$$

denoting

$$e_2(t) = \sum_{iajb} (ia|jb)[2(ia|jb) - (ib|ja)] e^{-(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)t}$$

we get

$$E_2 = - \int_0^\infty e_2(t) dt$$

LT-AO-MP2

Defining transformed AO orbitals as

$$\underline{\mu} = \sum_{\nu} X_{\mu\nu} \nu \quad \bar{\mu} = \sum_{\nu} Y_{\mu\nu} \nu$$

where X and Y are orbital-energy-weighted density matrices

$$X_{\mu\nu} = \sum_i^{\text{occ}} C_{\mu i} C_{\nu i} e^{\epsilon_i t} \quad Y_{\mu\nu} = \sum_a^{\text{virt}} C_{\mu a} C_{\nu a} e^{-\epsilon_a t}$$

yields the workhorse of the method

$$e_2 = - \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma}) [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)]$$

What's in it for us?

- transformation seems to be as expensive as canonical one
- there is an additional cost due to numerical integration
- where are the gains?

- X and Y matrices are sparse for large molecules
- transformed integrals tensor shows sparse structure as well, its non-zero elements can be identified efficiently
- both time and memory complexity can be reduced to $\mathcal{O}(N^2)$

What's in it for us?

- transformation seems to be as expensive as canonical one
- there is an additional cost due to numerical integration
- where are the gains?

- X and Y matrices are sparse for large molecules
- transformed integrals tensor shows sparse structure as well, its non-zero elements can be identified efficiently
- both time and memory complexity can be reduced to $\mathcal{O}(N^2)$

Computation flow

For each quadrature point

- calculate prescreening data
- execute four-stage transformation

$$(\underline{\mu\nu}|\underline{\lambda\sigma}) = \sum_{\kappa} X_{\mu\kappa}(\kappa\nu|\underline{\lambda\sigma})$$

$$(\underline{\mu\bar{\nu}}|\underline{\lambda\sigma}) = \sum_{\kappa} Y_{\nu\kappa}(\underline{\mu\kappa}|\underline{\lambda\sigma})$$

$$(\underline{\mu\bar{\nu}}|\underline{\lambda\sigma}) = \sum_{\kappa} X_{\lambda\kappa}(\underline{\mu\bar{\nu}}|\kappa\sigma)$$

$$(\underline{\mu\bar{\nu}}|\underline{\lambda\bar{\sigma}}) = \sum_{\kappa} Y_{\sigma\kappa}(\underline{\mu\bar{\nu}}|\underline{\lambda\kappa})$$

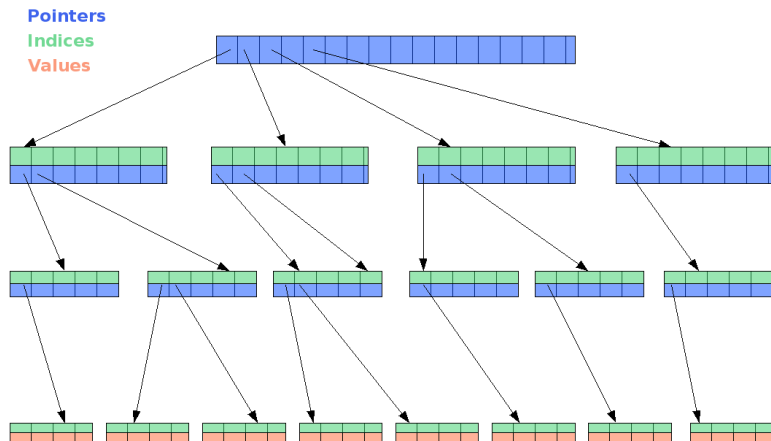
- calculate and accumulate contribution to the energy correction

Data structures

The tensors we deal with are extremely sparse

- We need data structures which exploit the sparsity
- To make things more complicated we require
 - efficient parallelization
 - fast browsing and manipulation
 - ability to perform single contraction in dense memory

Data structures



Data structures

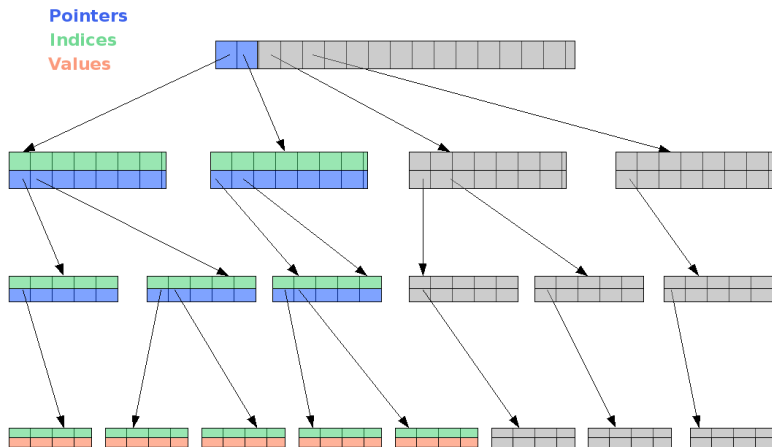
- separate trees used at every quarter of four-stage transformation
- sequence of indices is chosen for each transformation differently to maximize performance
- dynamic rewriting to small dense matrices is used to speed-up algebra operations

Multi-pass/parallel algorithm

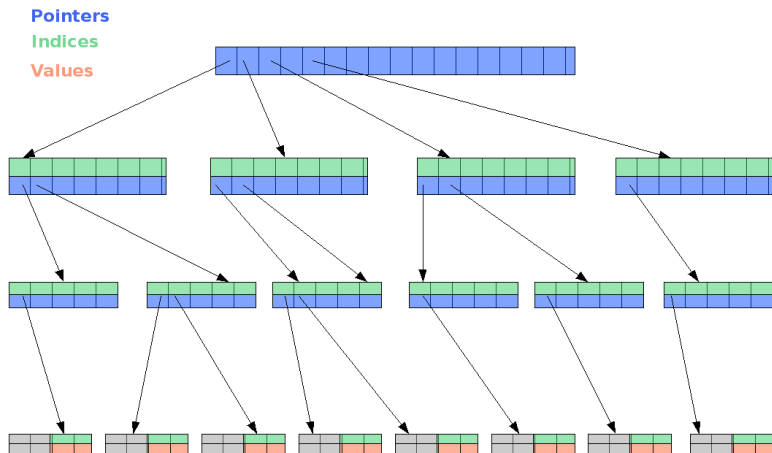
$$\begin{aligned}
 e_2 &= - \sum_{\mu\nu\lambda\sigma} (\underline{\mu\nu} | \underline{\lambda\sigma}) [2(\mu\nu | \lambda\sigma) - (\mu\sigma | \lambda\nu)] \\
 &= - \sum_{\mathcal{M}} \sum_{\mu \in \mathcal{M}} \left(\sum_{\nu\lambda\sigma} (\underline{\mu\nu} | \underline{\lambda\sigma}) [2(\mu\nu | \lambda\sigma) - (\mu\sigma | \lambda\nu)] \right)
 \end{aligned}$$

- for each quadrature point
 - generate optimal μ ranges \mathcal{M}
 - for each subset \mathcal{M}
 - calculate prescreening data
 - execute partial four-stage transformation
 - calculate and accumulate partial contribution to the energy correction

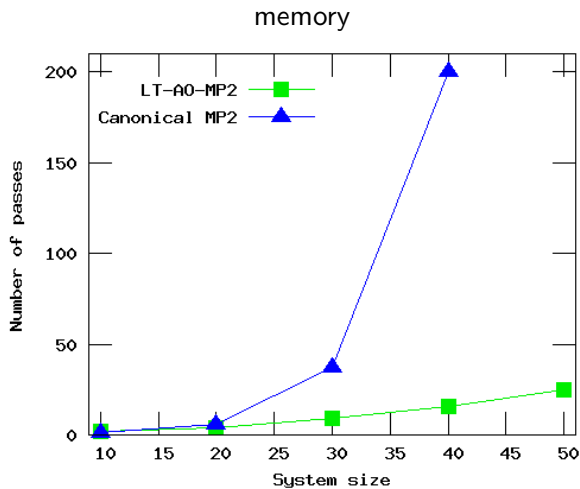
Partial trees



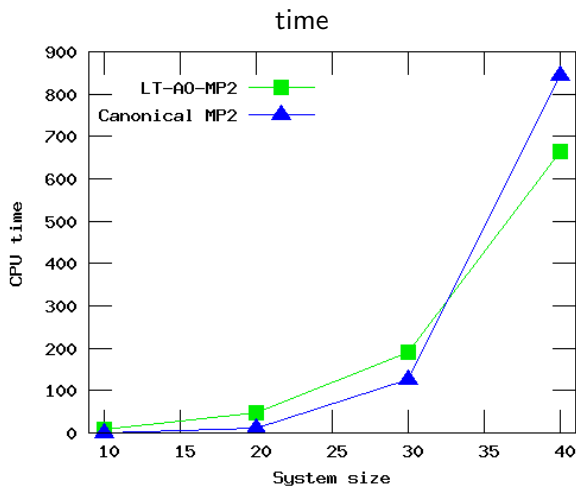
Partial trees



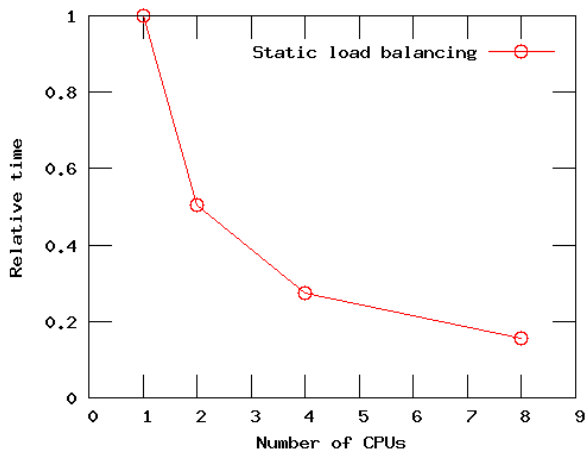
Performance



Performance



Parallelization efficiency



Modest superlinear scaling observed ($\alpha = 0.94$, $\beta = 1.06$, $\gamma = 0.05$)

Conclusions

- our LT-AO-MP2 implementation fulfills the expected performance criteria
- the implementation exploits advanced memory management and dynamic data structures
- the design fits well into computational chemistry framework being developed in our group

Future outlook

- reduction of the complexity to linear
- energy derivatives
- application for
 - nanotubes
 - biochemical systems

Summary

- the vintage post-HF methods are getting boosters
- this way they may become competitive to DFT
- while the ideas are old, the practical implementations started appearing only recently
- it seems that the trend is getting stronger
- and we try to be a part of it