## A Selfenergy Cookbook - State-of-the-Art Computing for the NEGF Key Ingredient

Niclas Schlünzen
in collaboration with:
Jan-Philip Joost and Michael Bonitz

| C A | U |
| :--- | :--- | :--- |
| Christian-Albrechts-Universität zu Kiel |  |
| Mathematisch- <br> Naturwissenschaftliche Fakultät |  |

KBEt ${ }^{2}$, Kiel
March 11, 2019

$$
\begin{array}{cc}
\sum_{l}\left[\mathrm{i} \hbar \frac{\mathrm{~d}}{\mathrm{~d} z_{1}} \delta_{i l}-h_{i l}\left(z_{1}\right)\right] G_{l j}\left(z_{1}, z_{2}\right) & G_{i j}\left(z_{1}, z_{2}\right)=G_{i j}^{(0)}\left(z_{1}, z_{2}\right)+ \\
=\delta_{\mathcal{C}}\left(z_{1}, z_{2}\right) \delta_{i j}+\sum_{l} \int_{z_{3}} \Sigma_{i l}\left(z_{1}, z_{3}\right) G_{l j}\left(z_{3}, z_{2}\right) & +\int_{\mathcal{C}} \mathrm{d} z_{3} \mathrm{~d} z_{4} \sum_{m n} G_{i m}^{(0)}\left(z_{1}, z_{3}\right) \Sigma_{m n}\left(z_{3}, z_{4}\right) G_{n j}\left(z_{4}, z_{2}\right) \\
\sum_{l} G_{i l}\left(z_{1}, z_{2}\right)\left[-\mathrm{i} \hbar \frac{\overleftarrow{\mathrm{~d}}}{\mathrm{~d} z_{2}} \delta_{l j}-h_{l j}\left(z_{2}\right)\right] \\
=\delta_{\mathcal{C}}\left(z_{1}, z_{2}\right) \delta_{i j}+\sum_{l} \int_{z_{3}} G_{i l}\left(z_{1}, z_{3}\right) \Sigma_{l j}\left(z_{3}, z_{2}\right) & W_{i j k l}^{\mathrm{ns}}\left(z_{1}, z_{2}\right)=\sum_{m n} w_{i m n l}\left(z_{1}\right) \int_{\mathcal{C}} \mathrm{d} z_{3} \sum_{p q} P_{n q p m}\left(z_{1}, z_{3}\right) W_{p j k q}\left(z_{3}, z_{2}\right) \\
\sum_{i j}^{\mathrm{xc}}\left(z_{1}, z_{2}\right)=\mathrm{i} \hbar \sum_{m p q} w_{i p q m}\left(z_{1}\right) \int_{\mathcal{C}} \mathrm{d} z_{3} \sum_{n} G_{m n}\left(z_{1}, z_{3}\right) \Lambda_{n q p j}\left(z_{3}, z_{2}, z_{1}\right) & \sum_{i j}^{\mathrm{xc}}\left(z_{1}, z_{2}\right)=\mathrm{i} \hbar \int_{\mathcal{C}} \mathrm{d} z_{3} \sum_{m p q} W_{i p q m}\left(z_{1}, z_{3}\right) \times \\
& \int_{\mathcal{C}} \mathrm{d} z_{4} \sum_{n} G_{m n}\left(z_{1}, z_{4}\right) \Gamma_{n q p j}\left(z_{4}, z_{2}, z_{3}\right)
\end{array}
$$

$$
\begin{aligned}
& \sum_{l}\left[\mathrm{i} \hbar \frac{\mathrm{~d}}{\mathrm{~d} z_{1}} \delta_{i l}-h_{i l}\left(z_{1}\right)\right] G_{l j}\left(z_{1}, z_{2}\right) \\
& =\delta_{\mathcal{C}}\left(z_{1}, z_{2}\right) \delta_{i j}+\sum_{l} \int_{z_{3}} \sum_{i l}\left(z_{1}, z_{3}\right) \sum_{l j}\left(z_{3}, z_{2}\right) \\
& \sum_{l} G_{i l}\left(z_{1}, z_{2}\right)\left[-\mathrm{i} \hbar \frac{\overleftarrow{\mathrm{~d}}}{\mathrm{~d} z_{2}} \delta_{l j}-h_{l j}\left(z_{2}\right)\right] \\
& =\delta_{\mathcal{C}}\left(z_{1}, z_{2}\right) \delta_{i j}+\sum_{l} \int_{z_{3}} G_{i l}\left(z_{1},\right. \\
& \sum_{i j}^{\mathrm{xc}}\left(z_{1}, z_{2}\right)=\frac{\mathrm{i} \hbar \sum_{\text {mpq }} w_{\text {ipqm }}\left(z_{1}\right) \int_{\mathcal{C}} \mathrm{d} z_{3} \sum_{n} G \quad \text { Central Quantity: }}{} \\
& G_{i j}\left(z_{1}, z_{2}\right)=G_{i j}^{(0)}\left(z_{1}, z_{2}\right)+ \\
& \begin{array}{l}
+\int_{\mathcal{C}} \mathrm{d} z_{3} \mathrm{~d} z_{4} \sum_{m n} G_{i m}^{(0)}\left(z_{1}, \Sigma_{m n}\left(z_{3}, z_{4}\right) S_{n j}\left(z_{4}, z_{2}\right)\right. \\
=\sum_{m n} w_{i m n l}\left(z_{1}\right) \int_{\mathcal{C}} \mathrm{d} z \sum_{p q} P_{n q p m}\left(z_{1}, z_{3}\right) W_{p j k q}\left(z_{3}, z_{2}\right)
\end{array} \\
& \sum_{i j}^{\mathrm{xc}}\left(z_{1}, z_{2}\right)=\mathrm{i} \hbar / \int_{\mathcal{C}} \mathrm{d} z_{3} \sum_{m p q} W_{i p q m}\left(z_{1}, z_{3}\right) \times \\
& \int_{\mathcal{C}} \mathrm{d} z \sum_{n} G_{m n}\left(z_{1}, z_{4}\right) \Gamma_{n q p j}\left(z_{4}, z_{2}, z_{3}\right)
\end{aligned}
$$

$$
\begin{aligned}
& \Lambda_{i j k l}\left(z_{1}, z_{2}, z_{3}\right)=\delta_{c}(,) \delta_{\delta}\left(z_{3}, z_{2}\right) \quad \text { Selfenergy } \boldsymbol{\Sigma} \\
& \text { Central Quantity: }
\end{aligned}
$$

$$
\begin{aligned}
& \int_{\mathcal{C}} \mathrm{d} z_{7} \sum_{q} G_{q n}\left(z_{7}, z_{5}\right) \Gamma_{p j k q}\left(z_{6}, z_{7}, z_{3}\right)
\end{aligned}
$$

Going beyond Second Born - Selfenergy Overview

Computational Remarks

The Hubbard model

Performance Tests

Going beyond Second Born - Selfenergy Overview

Computational Remarks

The Hubbard model

Performance Tests

- Use Feynman diagrams to visualize Green functions and interactions
- general vs. diagonal basis:

$$
G_{i j}\left(z, z^{\prime}\right)=i, z \text {, }
$$

...

- Hartree-Fock selfenergy becomes:

$$
\left.\begin{array}{rl}
\sum_{i j}^{\mathrm{HF}}\left(z, z^{\prime}\right) & = \pm \mathrm{i} \hbar \delta_{\mathcal{C}}\left(z, z^{\prime}\right) \sum_{k, l} \int_{\mathcal{C}} \mathrm{d} \bar{z} w_{i k l j}(z, \bar{z}) G_{l k}\left(\bar{z}, \bar{z}^{+}\right) \\
& +\mathrm{i} \hbar \sum_{k, l} w_{i k j l}\left(z, z^{\prime}\right) G_{l k}\left(z, z^{\prime+}\right)
\end{array} \rightarrow \rightarrow \rightarrow \pm \mathrm{i} \hbar\right\}+\mathrm{i} \hbar
$$

selfenergy approximations that solve the Martin-Schwinger hierarchy can be derived from a closed set of equations
two equivalent, formally exact approaches based on:

- the screened interaction/ vertex:
- the bare interaction/ vertex:

$$
\begin{aligned}
& G(1,2)=G^{(0)}(1,2) \\
& \quad+G^{(0)}(1,3) \Sigma(3,4) G(4,2) \\
& \Sigma(1,2)= \pm \mathrm{i} \hbar \delta(1,2) w(1,3) G\left(3,3^{+}\right) \\
& \quad+\mathrm{i} \hbar W(1,3) G(1,4) \Gamma(4,2,3)
\end{aligned}
$$

$$
\begin{aligned}
& W(1,2)=w(1,2) \\
& \quad+w(1,3) P(3,4) W(4,2)
\end{aligned}
$$

$$
P(1,2)= \pm \mathrm{i} \hbar G(1,3) G(4,1) \Gamma(3,4,2)
$$



$$
\Gamma(1,2,3)=\delta\left(1,2^{+}\right) \delta(3,2)
$$

$$
+\frac{\delta \Sigma^{\times ¢}(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \Gamma(6,7,3)
$$


$G(1,2)=G^{(0)}(1,2)$

$$
+G^{(0)}(1,3) \Sigma(3,4) G(4,2)
$$

$$
\begin{aligned}
& \Sigma(1,2)= \pm \mathrm{i} \hbar \delta(1,2) w(1,3) G\left(3,3^{+}\right) \\
& \quad+\mathrm{i} \hbar w(1,3) G(1,4) \Lambda(4,2,3)
\end{aligned}
$$

$$
\Lambda(1,2,3)=\delta\left(1,2^{+}\right) \delta(3,2)
$$

$$
+\frac{\delta \Sigma(1,2)}{\delta G(4,5)} G(4,6) G(7,5) \Lambda(6,7,3)
$$


quick \& easy
SOA, 2B
Second-Order (Born) Approximation

- all diagrams up to second order in the interaction
- easiest way to include correlation effects
- combination with the GKBA $\rightarrow N_{t}^{2}$-scaling
- applicable to a wide range of systems


## Computation

## simple matrix multiplication

Numerical Scaling

- general basis: $\mathcal{O}\left(N_{t}^{2} N_{b}^{5}\right)$
- diagonal basis: $\mathcal{O}\left(N_{t}^{2} N_{b}^{4}\right)$
from screened approach:

ancon




$$
\begin{aligned}
& \Sigma^{(2)}(1,2) \\
& = \pm(i \hbar)^{2} w(1,3) G(3,4) G(4,3) w(4,2) G(1,2) \\
& \quad+(i \hbar)^{2} w(1,3) G(1,4) w(4,2) G(4,3) G(3,2)
\end{aligned}
$$

## Third-Order Diagrams

Derivation of third-order terms from Hedin's equations (screened approach):

N. Schlünzen, S. Hermanns, M. Scharnke, and M. Bonitz, submitted, arXiv:1902.07038 (2019).

## the nitpicker <br> TOA

Third-Order Approximation

- all diagrams up to third order in the interaction
- more involved calculation
- applicability range restricted to moderate basis size


## Computation

matrix multiplication +
numerical integration

Numerical Scaling

Accuracy Range

- moderate to
strong coupling
- no filling dependence


$$
\Sigma^{(3)}(1,2)
$$

$=(\mathrm{it})^{3} G(1,2) w(1,3) G(3,4) G(4,3) w(4,5) G(5,6) G(6,5) w(6,2)$
$\pm(\mathrm{it})^{3} w(1,3) G(3,5) G(6,3) w(5,6) G(5,4) G(4,6) w(4,2) G(1,2)$
$\pm(i \hbar)^{3} w(1,5) G(5,6) G(6,5) w(6,3) G(1,4) w(4,2) G(4,3) G(3,2)$
$+(i \hbar)^{3} w(1,3) G(1,4) w(4,2) G(4,5) G(6,2) w(5,6) G(5,3) G(3,6)$
$\pm(i \hbar)^{3} w(1,3) G(1,4) w(4,6) G(6,7) G(7,6) w(7,2) G(4,3) G(3,2)$
$+(i \hbar)^{3} w(1,3) G(1,4) w(4,6) w(5,2) G(5,6) G(6,2) G(4,3) G(3,5)$
$\pm(i \hbar)^{3} w(1,3) G(1,4) w(4,6) G(6,5) w(5,2) G(4,2) G(5,3) G(3,6)$
$+(i t)^{3} w(1,3) G(1,4) w(4,6) w(5,2) G(5,6) G(6,2) G(4,3) G(3,5)$
$+(\text { it })^{3} w(1,3) G(1,4) w(4,6) G(4,5) w(5,2) G(6,2) G(5,3) G(3,6)$
$+(i \hbar)^{3} w(1,3) G(1,4) w(4,5) G(4,6) w(6,2) G(6,5) G(5,3) G(3,2)$

GWA
GW Approximation

- easiest way to decribe dynamical-screening effects
- sums up polarization-bubble diagram series
- computationally demanding, but scaling advantage for diagonal basis sets


## Computation

matrix multiplication + numerical integration, solution by iteration or inversion

Accuracy Range

- moderate to
strong coupling
- around half filling
screened approach for

$\mapsto$
0

xc $\mapsto+\mathrm{i} \hbar$


$$
\begin{aligned}
& W(1,2) \\
& =w(1,2) \pm i \hbar w(1,3) G(3,4) G(4,3) W(4,2) \\
& \Sigma^{G W}(1,2)=\Sigma^{H}(1,2)+\mathrm{i} \hbar G(1,2) W(1,2)
\end{aligned}
$$

Truncation of the bare-vertex recursion by


The selfenergy becomes:


The selfenergy derivative starts off three diagram series:

N. Schlünzen, S. Hermanns, M. Scharnke, and M. Bonitz, submitted, arXiv:1902.07038 (2019).

Truncation of the bare-vertex recursion by


The selfenergy becomes:


The selfenergy derivative starts off three diagram series:


N. Schlünzen, S. Hermanns, M. Scharnke, and M. Bonitz, submitted, arXiv:1902.07038 (2019).

## climbing the ladder I

TPP

## Particle-Particle

$\boldsymbol{T}$-Matrix Approximation

- sums up the diagrams of the Born series
- computationally expensive $\rightarrow$ applicable only to moderate basis size
- becomes exact in the limit of low (large) density


## Computation

matrix multiplication + numerical integration, solution by iteration or inversion

Numerical Scaling

- general basis:
$\mathcal{O}\left(N_{t}^{3} N_{b}^{6}\right)$
- diagonal basis:
$\mathcal{O}\left(N_{t}^{3} N_{b}^{6}\right)$

Accuracy Range

- moderate to
strong coupling
- low/large density

$+i \hbar$ o~~~



$+(i \hbar)^{3}$


$$
\begin{aligned}
T^{\mathrm{pp}}(1,2)= & w(1) G(1,2) G(1,2) w^{ \pm}(2) \\
& +w(1) G(1,3) G(1,3) T^{\mathrm{pp}}(3,2) \\
\Sigma^{\mathrm{TPP}}(1,2)= & \mathrm{i} \hbar T^{\mathrm{pp}}(1,2) G(2,1)
\end{aligned}
$$

climbing the ladder II

## TPH, TEH

Particle (Electron)-Hole
T-Matrix Approximation

- sums up a series of particle-hole diagrams
- computationally demand and reach similar to TPP
- specifically designed to describe systems around half half filling


## Computation

matrix multiplication + numerical integration, solution by iteration or inversion

Numerical Scaling

- general basis:
$\mathcal{O}\left(N_{t}^{3} N_{b}^{6}\right)$
- diagonal basis:

Accuracy Range

- moderate to strong coupling

$\pm(i \hbar)^{3}$


$$
\begin{aligned}
T^{\mathrm{ph}}(1,2)= & w(1) G(1,2) G(2,1) w^{ \pm}(2) \\
& +w(1) G(1,3) G(3,1) T^{\mathrm{ph}}(3,2) \\
\Sigma^{\mathrm{TPH}}(1,2) & =\mathrm{i} \hbar T^{\mathrm{ph}}(1,2) G(2,1)
\end{aligned}
$$

- around half filling


## it's a matter of patience

FLEX
Fluctuating-Exchange Approximation

- merges the diagram series of the TPP, the TPH and the GWA
- combines advantages of its ingredients
- highest computational demands of the presented approximations


## Computation

matrix multiplication + numerical integration, solution by iteration or inversion, avoid double counting of mutual terms

Numerical Scaling

- general basis:
$\mathcal{O}\left(N_{t}^{3} N_{b}^{6}\right)$
- diagonal basis:

Accuracy Range

- moderate to strong coupling

$+(\mathrm{i} \hbar)^{3}$

$\pm(\mathrm{i} \hbar)^{3}$

$\pm(\mathrm{i} \hbar)^{3}$


$$
\begin{aligned}
\Sigma^{\text {FLEX }}(1,2) & =\Sigma^{(1)}(1,2)+\Sigma^{G W, \text { corr }}(1,2) \\
& +\Sigma^{\text {TPP,corr }}(1,2)+\Sigma^{\text {TPH,corr }}(1,2) \\
& -2 \Sigma^{(2)}(1,2)
\end{aligned}
$$

- no filling dependence

Going beyond Second Born - Selfenergy Overview

Computational Remarks

The Hubbard model

Performance Tests

- high-performance computing and massive parallelization are essential to outgrow toy models
- NEGF is well-suited, contains high degree of independent calculations in matrix multiplication and numerical integration
- possible with multi-core programming on multiple CPUs
- "cheaper" way $\Rightarrow$ parallelization on GPUs
- con: memory structure is hard to manage


## GPU Motivation (I): Performance Trends

Peak Double Precision FLOPS


Peak Memory Bandwidth



- high-performance computing and massive parallelization are essential to outgrow toy models
- NEGF is well-suited, contains high degree of independent calculations in matrix multiplication and numerical integration
- possible with multi-core programming on multiple CPUs
- "cheaper" way $\Rightarrow$ parallelization on GPUs
- con: memory structure is hard to manage
- salvation: unified memory since NVIDIA Pasca!!


## GPU Motivation (I): Performance Trends

Peak Double Precision FLOPS


Peak Memory Bandwidth



Going beyond Second Born - Selfenergy Overview

Computational Remarks

The Hubbard model

Performance Tests


$$
\hat{H}(t)=J \sum_{i j, \alpha} h_{i j} \hat{c}_{i \alpha}^{\dagger} \hat{c}_{j \alpha}+U \sum_{i} \hat{c}_{i \uparrow}^{\dagger} \hat{c}_{i \uparrow} \hat{c}_{i \downarrow}^{\dagger} \hat{c}_{i \downarrow}+\sum_{i j, \alpha \beta} f_{i j, \alpha \beta}(t) \hat{c}_{i \alpha}^{\dagger} \hat{c}_{j \beta}
$$

$h_{i j}=-\delta_{\langle i, j\rangle}$ and $\delta_{\langle i, j\rangle}=1$, if $(i, j)$ nearest neighbors, $\delta_{\langle i, j\rangle}=0$ otherwise; on-site repulsion $(U>0)$ or attraction $(U<0), U$ favors doublons (correlations)

- $f$ : excitation (1-particle hamiltonian): EM field, quench, particle impact etc.
- finite inhomogeneous system, size and geometry dependence
dispersion of 40 -site Hubbard chain (2B/SOA):

band gap for the infinite 1D chain:


[^0]band structure for the honeycomb lattice:


- $N_{t}$ : number of time steps, $N_{b}$ : basis size
- scaling of propagation scheme
- full KBE: $\mathcal{O}\left(N_{t}^{3}\right), \mathcal{O}\left(N_{b}^{3}\right)$
- HF-GKBA: $\mathcal{O}\left(N_{t}^{2}\right), \mathcal{O}\left(N_{b}^{3}\right)$
- scaling of selfenergy approximations:

|  | HF | 2B | TOA | GW | TPP | TEH | FLEX |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| order | $\sim w^{1}$ | $\sim w^{2}$ | $\sim w^{3}$ | $\rightarrow w^{\infty}$ |  |  |  |  |
| $N_{t}$-scaling | $\mathcal{O}\left(N_{t}^{1}\right)$ | $\mathcal{O}\left(N_{t}^{2}\right)$ | $\mathcal{O}\left(N_{t}^{3}\right)$ |  |  |  |  |  |
| $N_{b}$-scaling | $\mathcal{O}\left(N_{b}^{4}\right)$ | $\mathcal{O}\left(N_{b}^{5}\right)$ |  | $\mathcal{O}\left(N_{b}^{6}\right)$ |  |  |  | $w_{i j k l}$ |
|  | $\mathcal{O}\left(N_{b}^{2}\right)$ | $\mathcal{O}\left(N_{b}^{4}\right)$ |  | $\mathcal{O}\left(N_{b}^{3}\right)$ | $\mathcal{O}\left(N_{b}^{6}\right)$ |  |  | $V_{i j}:=w_{i j i j}$ |
|  | $\mathcal{O}\left(N_{b}^{1}\right)$ | $\mathcal{O}\left(N_{b}^{2}\right)$ | $\mathcal{O}\left(N_{b}^{3}\right)$ |  |  |  |  | $U_{i}:=w_{i i i i}$ |

- lattice models greatly reduce numerical complexity




Going beyond Second Born - Selfenergy Overview

Computational Remarks

The Hubbard model

Performance Tests


- initial state: doubly occupied sites at the center
- Hubbard chain of 65 sites with 34 particles
- non-trivial expansion, $U$-dependent
- mean squared displacement

$$
R^{2}(t)=\frac{1}{N} \sum_{s} n_{s}(t)\left[s-s_{0}\right]^{2}
$$

$\boldsymbol{s}_{0}$ : center of the system

- rescaled cloud diameter $d(t)=\sqrt{R^{2}(t)-R^{2}(0)}$
- expansion velocity $v_{\exp }(t)=\frac{\mathrm{d}}{\mathrm{d} t} d(t)$
- Hartree-Fock misses the slowing-down of the expansion
- trend: two-time propagation results underestimate the slowing-down
- best performance by TPP, TPPEH, FLEX combined with GKBA

$\mathrm{O}-\mathrm{O}-\mathrm{O}-\cdots-\mathrm{O}-\mathrm{O}-\mathrm{O}$

- Hartree-Fock results are not sufficient
- two-time results become steady due to artificial damping
- best performance by GKBA+TOA
- initial state: doubly occupied sites arranged as a charge density wave
- Hubbard chain of 20 sites with 20 particles
- Relaxation dynamics towards homogeneous density distribution
- fast build-up of correlations
- observables: double occupation


Results for a ten-site Hubbard chain for the $U=4 J$ ground state:

low/high density:

- SOA, GWA and TEH slightly off
- excellent agreement for the TPP around half filling:
- TPP and SOA fail and underestimate band gap
- GWA becomes strikingly accurate
- TEH slightly overestimates correlations, precise band gap

Results for a ten-site Hubbard chain for the $U=4 J$ ground state:


low/high density:

- SOA, GWA and TEH slightly off
- excellent agreement for the TPP around half filling:
- TPP and SOA fail and underestimate band gap
- GWA becomes strikingly accurate
- TEH slightly overestimates correlations, precise band gap

FLEX dominated by TEH, no significant improvement
no "allrounders" $\Rightarrow$ physical circumstances prescribe the best choice of $\Sigma$

- in many cases Second Born is not sufficient to describe correlations accurately
- going beyond SOA is not straight forward $\Rightarrow$ no "allrounders"
- controlled choice of selfenergy: dictated by filling and interaction strength, accurate up to $U \simeq$ bandwidth
- best performance by
- low/large filling: TPP
- half filling: GWA, TEH
- mixed nonequilibrium: TOA
- Hubbard basis drastically reduces numerical effort by scaling and diagram number
- parallelization is crucial and can be done on GPUs



[^0]:    J.-P. Joost, N. Schlünzen, and M. Bonitz, phys. stat. sol. (b), doi: 10.1002/pssb.201800498, (2019)

