April 1 - 4, 2002 Insitute for Pure and Applied Math (IPAM), UCLA

Linear Scaling Electronic-Structure Methods



Green's Function O(N) DFT Methods (Use and Accuracy for Materials Science Problems)



Duane D. Johnson Dept. of Materials Science and Engineering & Frederick Seitz Materials Research Lab University of Illinois, Urbana-Champaign

Acknowledgements: Materials Research Lab, UIUC (DOE grant DEFG02-96ER-45439) Materials Computation Center (NSF grant DMR-9976550)

Outline

• Motivations and Definitions

- Metals (w/ and w/o defects) require 0.1 mRy relative accuracy.
- bandgap materials able to achieve this with O(N) via localization.
- "Tutorial" on Green's Function Multiple-Scattering Theory

- Korringa-Kohn-Rostoker (KKR) Electronic-Structure Method

- Linear-Scaling in number of atoms per cell, N. (Metals too?)
 - Linear Scaling Green's Fct. (LSGF) Method
 - Linear Scaling Multiple Scattering(LSMS) Method
 - "tight-binding" Hybrid k-space vs. real-space
- Accuracy of real-space versus k-space methods.
- Temperature-broadening for Kinetic Energy
 - Does it help with real-space? NO, not for improving accuracy!
- Some Applications (e.g. fcc, bcc Cu; bcc Mo, hcp Co ...)

Many Approaches

Red= this work

- Plane Wave Pseudo-potentials
- k-space: FLAPW, LMTO, KKR*, LCAO, LASTO
- Real-space
 - divide-and-conquer
 - recursion
 - density matrix
 - grid-based (finite-element, finite difference,)
 - wavelets
 - KKR (LSMS of Stocks et al. ORNL)
 - Screened "tight-binding-like" KKR
- Hybrid (screened) k-space and real-space KKR
- Kinetic Energy Functional Approaches

*Green's function approach allows disorder also to be considered by configurational averagin, e.g., via coherent-potential approximation.

From high to low T: Where do atoms go and why?

characterization

Processing \Rightarrow Structure \Rightarrow Properties \Rightarrow Performance

Most properties can be significantly affected by STATE OF ORDER.



Synposis of our Motivation

- Fast O(N) if possible,
- Accurate (~ 0.1 mRy or 1 meV) (at least for energy difference)
- Can incorporate disorder effects (state of order)
- Checked reliability carefully (both band-gap and metals)
- Present honest appraisal of results

Example includes failure of T=0K at finite-T or non-perfect order

Fe-Al Phase Diagram: Disordered and Partially Ordered Phases



(for Ternaries = $L2_1$ or "Heusler" Phase)

KKR-CPA Energy vs. LRO Parameters ($0 \le \eta \le 1$)



Experiment vs. T=0 K DFT: e.g. fcc Ni₃V $\Delta E(L1_2-DO_{22})$

• <u>T=0 K (LDA) DFT calculation</u>

 $\Delta E(L1_2 - DO_{22}) = 111 \text{ meV} \qquad (DO_{22} \text{ lower})$

* present KKR-ASA results agree with all previous calculations,
see Xu et al. PRB 35, 6940 (1987); Pei et al. PRB 39, 5767 (1989);
Lin et al., PRB 45, 10863 (1992), Wolverton and Zunger PRB 52, 8813 (1995).

• <u>Diffuse scattering measurements</u> $(T_{expt} = 1373 \text{ K and } T_c = 1318 \text{ K})$

Finel et al. NATO-ASI vol. 256 (1993) and Barrachin et al. PRB 50, 12980 (1994)

 $\Delta E(L1_2 - DO_{22}) = 10\pm 5 \text{ meV}$ from $\Delta \alpha^{-1}$ data.

• (LDA) DFT KKR-CPA Short-range Order Calculation via Linear-Response

 $\Delta E(L1_2 - DO_{22}) = 8.3 \text{ meV}$

meV (or less) control ordering, temperature scale, and defect properties in metals

Consider the Effects of Partial Order



* As done in experimental analysis, make $\eta \equiv \eta_{001} = \eta_{01/21}$ (1 LRO parameter) * $\eta(T)$ has T-dependence that must be determined from statistical average.

Partially-Ordered (PO) States Relative to Disordered State

DDJ et al., Phys. Rev. B 62, RC11917 (2000)

- T=0 K DFT get $\Delta E \sim 100 \text{meV}$, but not relevant to experiment.
- ΔE depends on configuration, i.e., on LRO parameter, for L1₂.
- Configuration dependence arises from DOS changes in $L1_2$ below ε_F .

• Extrapolation from high-T PO states yields $\Delta E \sim 7-12$ meV, whereas direct ASRO calculations get $\Delta E \sim 8.3$ meV.



Partially ordered states must be considered in general



- Ni_2Cr SRO and LRO highly depended on local state of order.
- Estimated to change T_c by factor of 2 from linear-response type calculation.

• c-axis Layered $(L1_0)$ CoPt ordering well described by linear extrapolation to T=0 K.

Structural defects in L1₀

- •L1 $_0$ forms from A1 during annealing via nucleation and growth.
- •Within individual grains of A1, several $L1_0$ variants form antiphase boundaries (APB) and chemical domain boundaries (DB).
- •Slight contraction of lattice along tetragonal c axes in $L1_0$ causes formation of twins (not pictured).



Classification of APB's: Diffuse have disordered

- APB can be either layer- or interlayer-symmetric.
- Mean-field theory predicts that APB's become more diffuse with decrease in η .







Multiple-scattering Electronic-Structure Calculations for Complex Applications

Korringa-Kohn-Rostoker Method using Screened structure constants (sparsity) Sparse matrix techniques (memory & size) Iterative inversion technique (size) Hybrid k-space/r-space (accuracy) r-space only (speed) Coherent Potential Approx. (disorder)



M= atoms in screening cluster (2-6 n.n. shells) L=maximum angular momentum $(L \sim 3)$

Definitions

N = number of atoms in system (possibly in periodic cell) M = number of atoms in small region around some central atom. L = maximum index for spherical harmonics (angular momentum) $R_I - R_J =$ vector from atom I to J.



If only local information needed, real-space give O(N*M³)

As e⁻ scatter through collection, what is range of decay length?

Motivation: A code that

- Addresses large-scale, defected materials
 - (e.g. GB, APB with concentration gradients, ...)
- Uses a <u>non-spectral</u> Green's function

(e.g., avoids eigenvalue problem and use of unoccupied states ... transport, FS)

- Handles many inequivalent atoms in arbitrary structures
- Can obtain accurate spectral properties and energetics (i.e. physical DOS, Fermi surfaces, bands not just by T-broadening)
- Can be straightforwardly parallelized (E, k, inverses)
- Is fast on workstations, faster on clusters, fastest on parallel machine
- Uses a mixed real-space and k-space representation (i.e. to optimize speed and accuracy)
- Accurate for metallic systems

KKR Multiple-Scattering Theory Method

$$\left[-\nabla^2 + V(r) - E\right]G(r,r';E) = \delta(r-r')$$
 E can be complex



•"free electron" wave scatters off site with a phase-shift, $\delta_1(E)$, and then is free again.

• wavefunction expanded on site as $\psi_{\mathbf{k},\nu}(\mathbf{r}) = \sum_{L=(l,m)} c_{L,\nu}(\mathbf{k}) R_{l\nu}(r;E) Y_{L}(\hat{r})$

• interstitial
$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi\sum_{L} i^{l}j_{l}(kr)Y_{L}^{*}(\hat{k})Y_{L}(\hat{r})$$

For $l\approx 3$ or 4, $\delta_l(E)=0$, hence rather compact basis. Matrices have rank = N*(L+1)² ≈ N*16 or N*25.

Note: G(E) is complex and non-Hermitian.

Some advantages of non-spectral (non-wave-fct) Green's Function

- Density of States: $n(E) = \int dk G(k;E)$
- Charge Density: $\rho(r) = \oint dE \ TrG(r;E)$ (contour integration)

• G(k;E) provides direct spectral information of any E without knowledge of any other (un-)occupied states, e.g. Fermi surface of metal can be obtained directly.



• Chemical disorder can be handled by averaging G.

Sum of the scattering

The positions, R_i , of the scatterers (i.e. lattice structure) are given by free-space Green's function, $G_0(R_{ij};E)$.

$$G_{0}^{ij}(E) = -\frac{1}{4\pi} \sum_{i \neq j} \frac{e^{i\kappa|R_{i} - R_{j}|}}{|R_{i} - R_{j}|} \qquad \kappa = \sqrt{E}$$

$$\begin{split} \text{Dyson's Eq.} \\ G &= G_0 + G_0 \text{ t } G = G_0 + G_0 \text{ t } G_0 + G_0 \text{ t } G_0 \text{ t } G_0 + \dots \\ &= G_0 \ (1 - tG_0)^{-1} & \text{t-matrices} \\ &\quad t(E) = t^i_{\text{LL}}, (E) \delta_{ij} \end{split}$$

potential dependent

If $G_0(E)$ decays, then G is localized.

(Origin of "Tight-binding" LMTO)

<u>Contour integration for SCF Total E</u>: $G = G_0 + G_0 \Delta V G$



For a metal, as you approach real E axis, G has a range which grows algebraically, hence M>>N, rather than M<<N.

For material with a gap, G decays exponentially always!

Linear-Scaling Multiple-Scattering Methods

Real-space LSMS scales as O(N*M³). Yang et al., PRL 75, 2867 (1995) Visit N sites and invert an M site cluster

Advantages

- With M fixed (≈6 n.n. shells), LSMS is O(N). Use "localization" produced by contour integration keeping ImE > 15 mRy, and hope errors in Fermi energy, charges, total energy are small.
- Can be easily parallized, N sites on N nodes.

Disadvantages

- Accurate charge density (esp. T = 0 K) cannot be obtained.
- Density of states cannot be obtained.
- Spectral properties cannot be obtained.
- Large memory storage for G₀.

Ab initio "tight-binding-like" KKR or LSMS Methods

Can produce a "tight-binding-like" version of KKR or LSMS using:

(1) Reference Green's Function screened KKR Zeller et al. PRB 52, 8807 1995

*Two step solution of Green's function: use reference G^{ref} to produce localization.

$$G = G^{ref} [I - (t - t^{ref})G^{ref}]^{-1}$$
$$G^{ref} = G_0 [I - t^{ref}G_0]^{-1}$$

*Basically, the mathematical origin from KKR perspective of "Tight-binding" LMTO.

Disadvantage:

• Accuracy (in some metallic case).

Ab initio "tight-binding-like" KKR or LSMS Methods

(2) For alloys only, coherent potential approximation LSGF method Abrikosov et al PRL 76 4203 (1996)

$$G = G^{ref} [I - (t - t^{ref})G^{ref}]^{-1}$$

$$G^{ref} = G_0 [I - t^{ref}G_0]^{-1}$$

Let $G^{ref} \equiv G^{cpa}$ CPA has (strong)decay scattering, so that G can be solved as perturbation of G^{cpa} in real-space.

With perfect lattice, the G^{cpa} can be solved in k-space for a metal to obtain the correct long-range behavior of the Green's fct (algebraic decay).

Then, using a small cluster about the I-th atom, you can solve the local perturbation from the actual local environment to get the correct G.

Disadvantage:

- G^{cpa} must be obtain in k-space for 1-atom per cell
- G^{cpa}_{IJ} obtained by FFT in real-space to get local correction.
- CPA is for rigid (perfect) lattice

Localized (tight-binding) representation

Zeller et al., Phys. Rev. 52, 8807 (1995) and Phys. Rev. 55, 9400(1997)

Choose a reference system to solve Dyson's Equation, then solve for true system.

reference G: $G^r = G_0 + G_0 t^r G^r = G_0 (1 - t^r G_0)^{-1}$



• <u>scattering decay</u>

For k-space, can obtain $G^{r}(\mathbf{k})$ with a few n.n. shells of R's only advantage => sparse storage of G^{r} .

The real Green's function

Obtained G^r , get $G^r(\mathbf{k})$ if needed, now solve Dyson's Equation for true system in k-space or r-space depending on scattering behavior of electrons.

true G:
$$G = G^r + G^r \Delta t G = G^r (1 - \Delta t G^r)^{-1}$$
, $\Delta t = t - t^r$
= $\Delta t^{-1} (1 - \Delta t G^r)^{-1} - \Delta t^{-1}$

- Is G sparse? <u>Generally not.</u> But Δt^{-1} is block diagonal and we only need $\rho = Tr G$
- Recall, for any matrix A: $AA^{-1}=1$
- So 1 $\Delta t G^r$ is sparse (but not banded) and (1 - $\Delta t G^r$)⁻¹ (usually) is dense!

Matrices and k-space near real E axis



Calculation Details of Example

- Screened G₀ in real-space for sparse storage
- Screening cluster size $M \approx 27$ (3 n.n. shells)
- Real-space + k-space to get proper structure of G
- With sparse matrices, use <u>iterative methods</u> to get $(1 \Delta t G^r)^{-1}$
- All-electron, relativistic core, scalar-relativistic valence
- k-point methods $N_k^{(1)}N^{(1)} = N_k^{(2)}N^{(2)} = \text{constant}$

Important for scaling and maintain same accuracy from one size cell to another

• Memory $N^*M^*(L+1)^2$ (complex double precision)

$$+ \alpha \ N^2 \qquad \left[\text{sparse map (integer)} + \text{Madelung (real)} + a_{ij} \text{ (real)} \right] \\ \text{calculate (time)} \qquad <--> \qquad \text{store (memory)} \\ \end{array}$$

That depends on what you want. You decide.

Conclusion for scaling behavior in metals

(with same level of accuracy for all N)

- For SCF Total Energy: $O(N^{1+\epsilon})$ $\epsilon_1 \approx 0.0-0.5$
- For DOS calculations: $O(N^{1+\epsilon})$ $\epsilon_2 \approx 0.5-1.0$
- For Bands: $O(N^{2+\epsilon}) \qquad \epsilon_3(k) \approx 0.5-1.0$

<u>Memory scales as O(N) N*M*(L+1)²</u>

M= atoms in screening cluster (2-6 n.n. shells) L=maximum angular momentum (2-4)

Scaling is O(N) for real-space and k-space. Speed vs. accuracy vs. scaling.

<u>Time per E point (single Unix node)</u>

A.V. Smirnov and D.D. Johnson, Phys. Rev. B 64, 235129 (2001)

432 atom/cell (27 atom screening cluster)



~39% non-zero elements N=54 N=432

- ~5% non-zero elements
- N=2000 ~1% non-zero elements



 $O(N^{1+\epsilon})$ scaling for SCF in Metals ($\epsilon \approx 0-1$)

A.V. Smirnov and D.D. Johnson, Phys. Rev. B 64, 235129 (2001)



Errors for k-space vs. r-space





* Scr-LSMS more Reliable than LSMS.



open points refer to fcc, bcc Cu, bcc Mo, hcp Co: full real-space not accurate



In metals, error are related to Thomas-Fermi error (DOS error $\rightarrow E_f$ error) from G decay.

Energy Scaling and Timings: r-space vs. k-space not independent from accuracy requirements



• For a given small cluster size, r-space is O(N) and always faster! But not as accurate.

• For accurate total energies, hybrid r-space/ k-space is necessary.

Time and scaling then controlled by k-space for a few energies near real-E axis.

• Direct inversion (e.g. SuperLU) restricts size and scaling but gets all atoms at once.

•For fixed accuracy, iterative solves scale energy integration as O(N), but per atom so effectively $O(N^2)$.

- If using real-space but want to increase accuracy, N_{inv} increases rapidly and scaling and speed is lost.
- Can there be a pragmatic solution?

Temperature decay and Temperature Broadening

• Consider band energy

$$\begin{split} E_{\text{band}} = \oint dE \ \left(E - \mu\right) n(E) f(E) &\to \sum_{n=1} G(\omega_n) \\ \text{for } f(E) = \text{Fermi function} = (1 + e^{-(E - \mu)/kT})^{-1} \\ \omega_n = (2n+1)\pi k_B T \text{ are Matsubara poles.} \end{split}$$

• Similar to use of temperature in other methods.

But, in contour integration, last point cannot be closer than ImE ~ $\omega_{n=1}$.

• At T \approx 800-1000 K (ImE ~ 15 mRy), integration is easier because temperature has broadened spectral features some. That is, both KKR and LSMS are more localized.

•But is accuracy maintained?

DFT Free Energy is variational not Energy!

Temperature Extrapolation

• Determine F(T) = E(T) - TS(T)

• But need S(T) = f(E)lnf(E) - (1 - f(E)) ln(1 - f(E)), which has a singularity at $\omega_n = (2n+1)\pi k_B T$ with infinite branch-cut. So the farthest off for S(T) calculation is $\pi k_B T/2$.

- Also, easy to show that $E(T=0) \sim E(T) TS(T)/2$.
- But does this speed up calculations and increase accuracy of the real-space O(N) method?

No! Need entropy within first Matsubara pole (small ImE).



Hole states at Γ point control magnetism and stability. Need M $\rightarrow \infty$ for this! That is k-space.

Pre-conditioners for Matrix Inversion of Complex non-Hermitian

• Standard iterative methods (e.g., GMRES and TF-QMR)

For Im E small, G(E) require extremely large N_{iter} , making them slower than direct inversion (if memory were available, which it is usually not for N > 400). GMRES is too memory intensive, while CG methods are much slower than Transpose-Free Quasi-minimal residual (QMR) methods.

•"Cluster inverse" approximation

A physical-based direct inversion of a diagonal block of the sparse (real-space) matrix to get an approximate G_{ii} and has proposed recently to achieve apparent O(N) scaling. "cluster inverse" fails near real E because G does not just depend on the band-diagonal elements.

Modified Block Jacobi

A block Jacobi pre-conditioner was tried that uses the diagonal blocks of G inverse. Use $(G^{-1})^{*T}$ because standard block Jacobi does not work with matrices with complex eigenvalues, λ . The modified method convergences because the eigenvalues $\lambda\lambda^*$ are positive and real. No significant improvement in CPU requirements.

•Incomplete Block LU

Blocks along the diagonal were used in a standard LU factorization to generate a preconditioner. This is fast and efficient, and easily distributed over parallel machines. However, the method once again fails for small Im(Z), essentially for reasons as above.

Inverse-Jacobian (Rank-1) iterative update (Broyden's 2nd Method)
Approximate inverse preconditioners (minimizations)
Embedding in Effective Medium for Approximate Inversion

Synopsis

• Through mathematics, physics, and algorithms, Green's function, KKR method leads to O(N) scaling, where N is really atom/cell.

• O(N) for band-gap materials, but only if "close-packed",

e.g., by adding in empty spheres to cell - $O(N_{atoms} + N_{holes})$.

• O(N) for metals, but must check reliability!

see, Smirnov and Johnson Phys Rev B 64 235129 (2001).

- Sparse matrix method very useful.
- Iterative Schemes must include Physics (CS ↔ science) Need method for complex non-Hermitian case (non-Lanczos type) Investigating methods near Real E.
- Preconditioners must include Physics $(CS \leftrightarrow science)$ Investigating methods near Real E.

see, Smirnov and Johnson, Comp. Phys. Comm (2002).

Future

- Preconditioning of iterative inverse: decrease time and scaling (?)
- Iterative methods are easily parallel (MPI/PVM version)
- Forces
- Applications to large-scale defects
- Embedding into O(N) Discontinuous Galerkin-FEM