Fast Algorithms for Electronic Structure Analysis

Lin Lin

Department of Mathematics, UC Berkeley;

Computational Research Division, LBNL

PACM Colloquium,

Princeton, April, 2015

Kohn-Sham density functional theory

$$H[\rho]\psi_i(x) = \left(-\frac{1}{2}\Delta + \int dx' \frac{m(x') + \rho(x')}{|x - x'|} + V_{xc}[\rho]\right)\psi_i(x) = \varepsilon_i \psi_i(x)$$

$$\rho(x) = 2\sum_{i=1}^{N/2} |\psi_i(x)|^2, \qquad \int dx\, \psi_i^*(x)\psi_j(x) = \delta_{ij}, \qquad \varepsilon_1 \le \varepsilon_2 \le \cdots$$

- Efficient: Always solve an equation in R^3, regardless of the number of electrons N.
- Accurate: Exact ground state energy for exact $V_{xc}[\rho]$, [Hohenberg-Kohn,1964], [Kohn-Sham, 1965]
- Best compromise between efficiency and accuracy. Most widely used electronic structure theory for condensed matter systems and molecules
- Nobel Prize in Chemistry, 1998

Density functional theory

Table 1. *Physical Review* Articles with more than 1000 Citations Through June 2003

Publication	# cites	Av. age	Title	Author(s)
PR **140**, A1133 (1965)	3227	26.7	Self-Consistent Equations Including Exchange and Correlation Effects	W. Kohn, L. J. Sham
PR **136**, B864 (1964)	2460	28.7	Inhomogeneous Electron Gas	P. Hohenberg, W. Kohn
PRB **23**, 5048 (1981)	2079	14.4	Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems	J. P. Perdew, A. Zunger
PRL **45**, 566 (1980)	1781	15.4	Ground State of the Electron Gas by a Stochastic Method	D. M. Ceperley, B. J. Alder
PR **108**, 1175 (1957)	1364	20.2	Theory of Superconductivity	J. Bardeen, L. N. Cooper, J. R. Schrieffer
PRL **19**, 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
PRB **12**, 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Anderson
PR **124**, 1866 (1961)	1178	28.0	Effects of Configuration Interaction of Intensities and Phase Shifts	U. Fano
RMP **57**, 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee, T. V. Ramakrishnan
RMP **54**, 437 (1982)	1045	10.8	Electronic Properties of Two-Dimensional Systems	T. Ando, A. B. Fowler, F. Stern
PRB **13**, 5188 (1976)	1023	20.8	Special Points for Brillouin-Zone Integrations	H. J. Monkhorst, J. D. Pack

PR, Physical Review; PRB, Physical Review B; PRL, Physical Review Letters; RMP, Reviews of Modern Physics.

[S. Redner, Citation Statistics from 110 Years of Physical Review]

Density functional theory

Table 1. *Physical Review* Articles with more than 1000 Citations Through June 2003

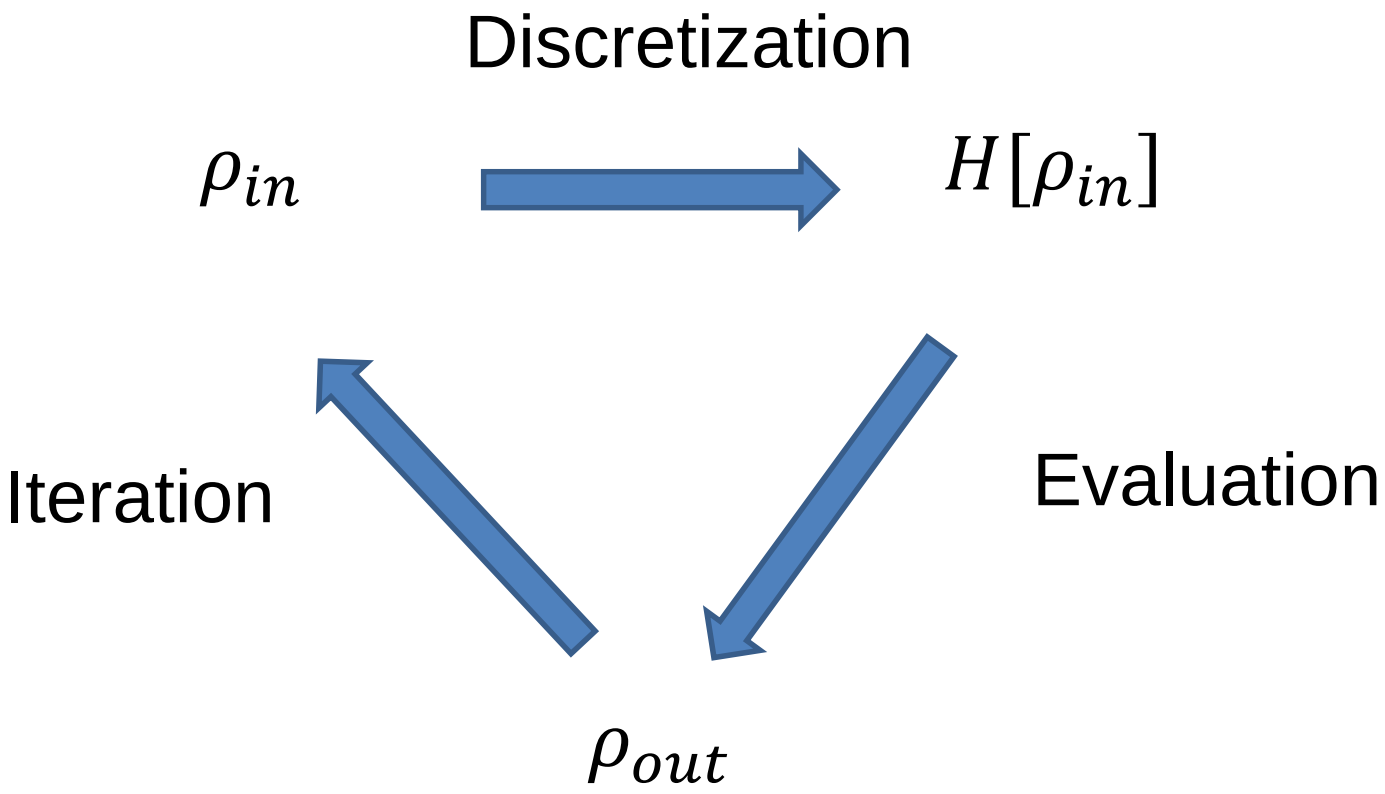
Publication	# cites	Av. age	Title	Author(s)
PR **140**, A1133 (1965)	3227	26.7	Self-Consistent Equations Including Exchange and Correlation Effects	W. Kohn, L. J. Sham
PR **136**, B864 (1964)	2460	28.7	Inhomogeneous Electron Gas	P. Hohenberg, W. Kohn
PRB **23**, 5048 (1981)	2079	14.4	Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems	J. P. Perdew, A. Zunger
PRL **45**, 566 (1980)	1781	15.4	Ground State of the Electron Gas by a Stochastic Method	D. M. Ceperley, B. J. Alder
PR **108**, 1175 (1957)	1364	20.2	Theory of Superconductivity	J. Bardeen, L. N. Cooper, J. R. Schrieffer
PRL **19**, 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
PRB **12**, 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Anderson
PR **124**, 1866 (1961)	1178	28.0	Effects of Configuration Interaction of Intensities and Phase Shifts	U. Fano
RMP **57**, 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee, T. V. Ramakrishnan
RMP **54**, 437 (1982)	1045	10.8	Electronic Properties of Two-Dimensional Systems	T. Ando, A. B. Fowler, F. Stern
PRB **13**, 5188 (1976)	1023	20.8	Special Points for Brillouin-Zone Integrations	H. J. Monkhorst, J. D. Pack

PR, Physical Review; PRB, Physical Review B; PRL, Physical Review Letters; RMP, Reviews of Modern Physics.

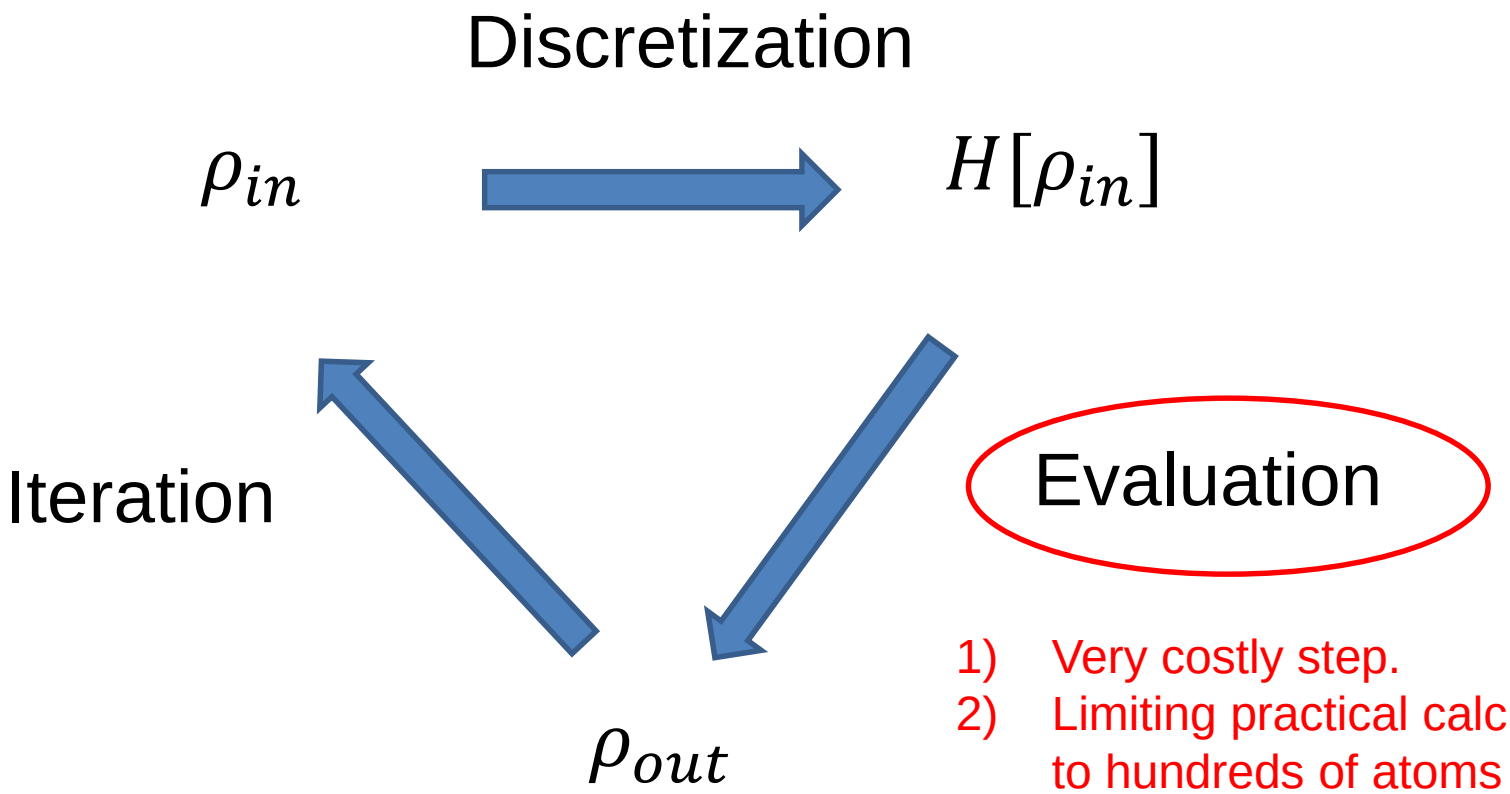
[S. Redner, Citation Statistics from 110 Years of Physical Review]

Self-consistent equations including exchange and correlation effects
W **Kohn**, LJ Sham - Physical Review, 1965 - APS
Abstract From a theory of Hohenberg and **Kohn**, approximation methods for treating an inhomogeneous system of interacting electrons are developed. These methods are exact for systems of slowly varying or high density. For the ground state, they lead to **self-consistent ...**
Cited by 35282 Related articles All 23 versions Web of Science: 24040 Import into BibTeX

Self Consistent Field Iteration



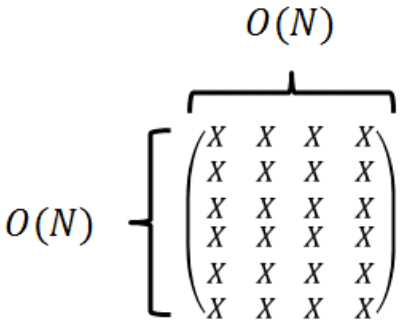
Self Consistent Field Iteration



1) Very costly step.
2) Limiting practical calculations to hundreds of atoms
3) Almost always treated by solving a linear eigenvalue problem

Cubic scaling of KSDFT

- KS orbitals (eigenfunctions) are delocalized in the global domain.

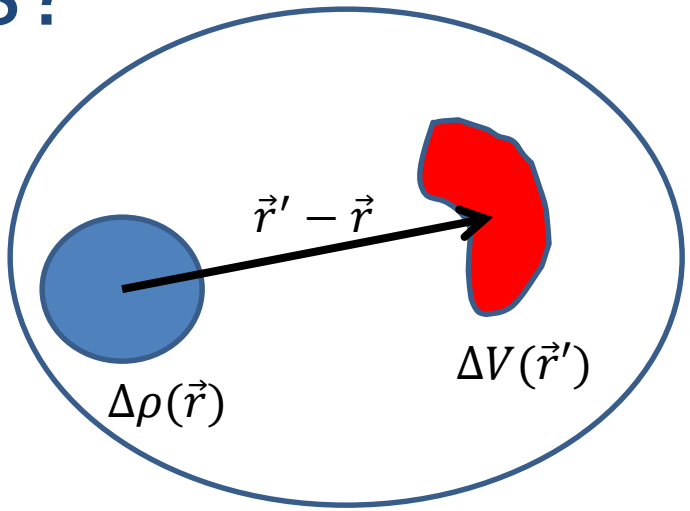


- N atoms. $O(N)$ grid points. $O(N)$ KS orbitals.

- Orthogonalization of an $O(N) \times O(N)$ matrix $\Rightarrow$ $O(N^3)$ scaling, regardless of what eigensolver is being used.

- Conclusion: DO NOT directly treat KS orbitals that are delocalized in the global domain.

Evaluation: Alternatives?



- Linear scaling algorithms
  - Near-sightedness [Kohn, PRL 1996] [Benzi-Boito-Razouk, SIAM Rev. 2013]
  - Truncation based algorithm:

    hard to balance efficiency and accuracy
  - Only applicable to insulators.

[Bowler and Miyazaki, Rep. Prog. Phys 2012]

"…The second challenge is that of metallic systems: there is no clear route to linear-scaling solution for systems with low or zero gaps and extended electronic structure…"

- Difficult task:
  - Accurate and efficient
  - Applicable to insulators, semiconductors and metals

Alterative solution?

Linear scaling methods

- Truncation (KS orbital, 1-dm). Near-sightedness.
- Costly for metals (large preconstant)
- Complicated user-interface (select truncation region)

[Yang, 1991], [Kohn, 1996].
Review: [Goedecker, 1999].
[Bowler-Miyazaki, 2012].

What we propose

- No truncation. Not based on near-sightedness.
- Applicable to insulator, semiconductor and metal.
- Black-box user-inteface.
- Scales at most $O(N^2)$.
- Localized basis and relatively small number of basis functions per atom.

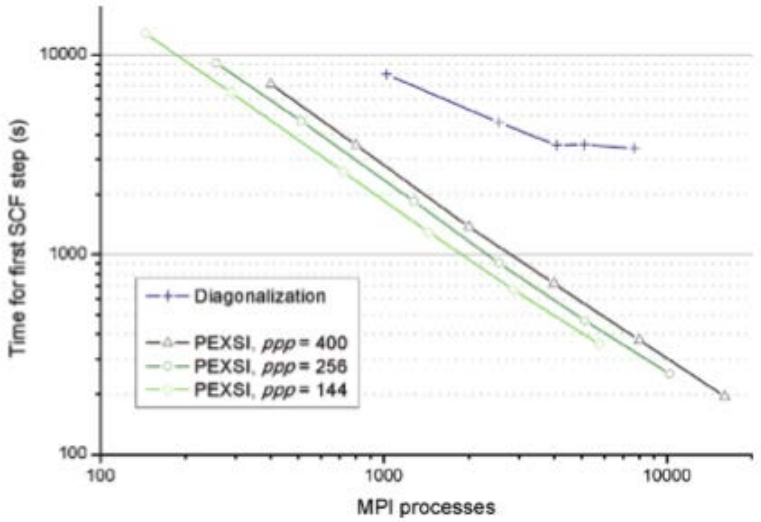
Outline

PEXSI: Pole EXpansion and Selected Inversion

- Pole Expansion
- Selected Inversion
- From model problem to practice

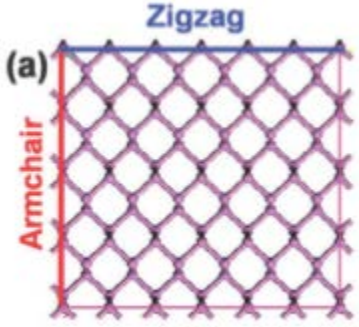
PEXSI at work

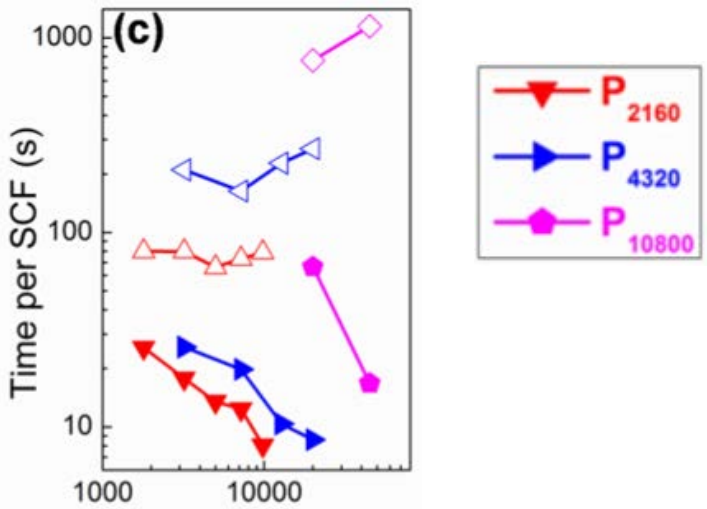
C-BN Bilayer (12770 atoms)



SIESTA-PEXSI. [LL-Garcia-Huhs-Yang, 2014]

Phospherene nanoribbon (10800 atoms)





DGDFT-PEXSI. [Hu-LL-Yang, 2015]

KSDFT: Matrix point of view

$$\rho(x) = 2\sum_{i=1}^{N/2} |\psi_i(x)|^2$$

$$= (\psi_1(x) \quad \dots \quad \psi_{N_t}(x)) \begin{pmatrix} \chi(\varepsilon_1 - \mu) & & \\ & \ddots & \\ & & \chi(\varepsilon_{N_t} - \mu) \end{pmatrix} \begin{pmatrix} \psi_1(x) \\ \vdots \\ \psi_{N_t}(x) \end{pmatrix}$$

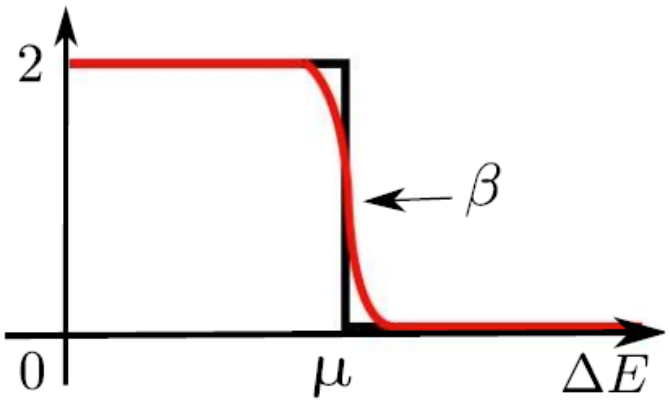
$$= \{\chi(H[\rho] - \mu I)\}_{x,x}$$

- μ : Chemical potential such that $\#\{\sigma(H) \le \mu\} = N/2$
- χ : Heaviside function satisfying $\chi(x) = \begin{cases} 2, & x \le 0, \\ 0, & x > 0 \end{cases}$

$$\rho = \operatorname{diag} \chi(H[\rho] - \mu I)$$

Finite temperature: Fermi operator

$$\rho = \mathrm{diag}\frac{2}{1+e^{\beta(H[\rho]-\mu I)}}$$



- $\beta = 1/k_B T$: inverse temperature
- μ: Chemical potential

- Finite temperature, Fermi-Dirac
- Zero temperature, Heaviside

Fermi operator expansion

$$\rho = \mathrm{diag}\frac{2}{1+e^{\beta(H[\rho]-\mu I)}} = \mathrm{diag}\frac{2}{1+e^{\beta\Delta E\frac{H[\rho]-\mu I}{\Delta E}}}$$

$$\approx \mathrm{diag}\left\{\sum_{l=1}^{P} c_l\left(\frac{H[\rho]-\mu I}{\Delta E}\right)^l + \sum_{l=1}^{Q}\frac{\omega_l}{\left(z_l I-\frac{H[\rho]-\mu I}{\Delta E}\right)^{q_l}}\right\}$$

- $\Delta E = \sigma(H-\mu I)$.
- Fermi operator expansion: solving KSDFT without diagonalization

- [Goedecker, 1993], $P \sim O(\beta\Delta E)$
- [Ceriotti et al, 2008], Q $\sim O\left(\sqrt{\beta\Delta E}\right)$; other work

Pole expansion

- [LL-Lu-Ying-E, 2009] $Q \sim O(\log(\beta \Delta E))$

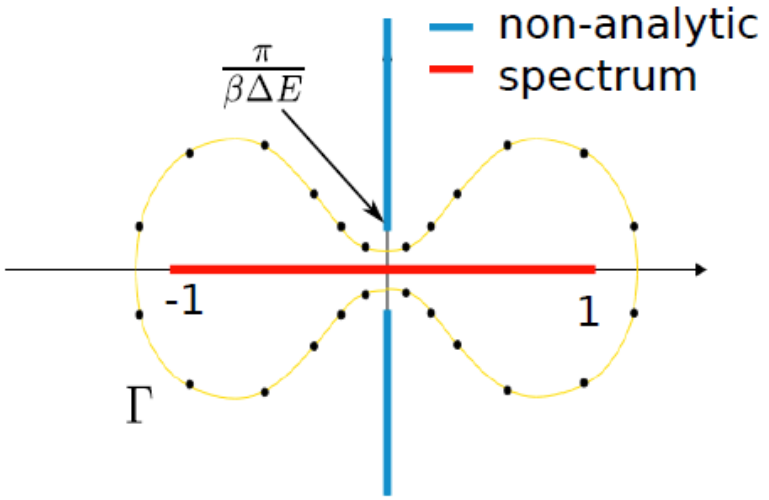
$$\rho \approx \mathrm{diag} \sum_{i=1}^{Q} \frac{\omega_i}{H - z_i I}$$

- $z_i, \omega_i \in \mathbb{C}$ are complex shifts and complex weights

Contour integral technique

Fermi-Dirac

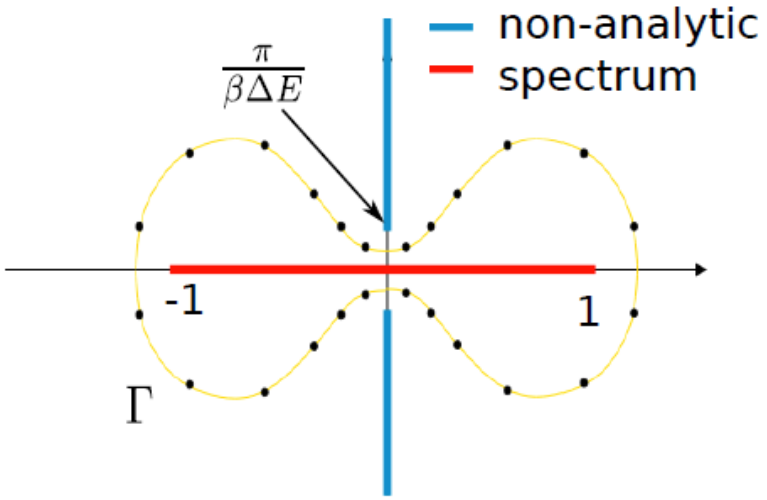
$$\rho(\xi) = \frac{1}{2\pi i}\oint_{\Gamma}\frac{\rho(z)}{z-\xi}dz \approx \frac{1}{2\pi i}\sum_{i=1}^{Q}\frac{\rho(z_i)w_i}{z_i-\xi}$$



Contour integral technique

Fermi-Dirac

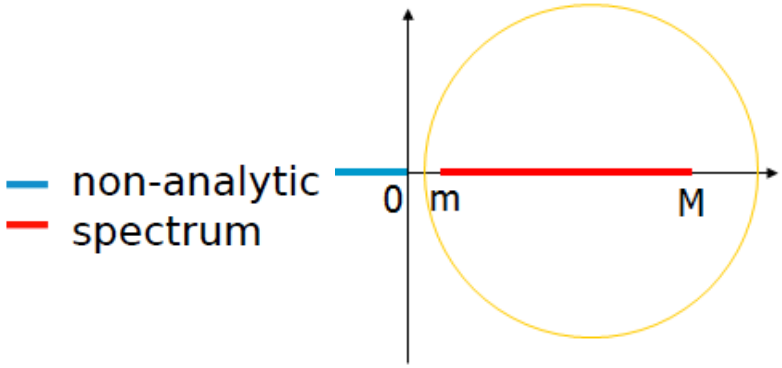
$$\rho(\xi) = \frac{1}{2\pi i}\oint_{\Gamma}\frac{\rho(z)}{z-\xi}dz \approx \frac{1}{2\pi i}\sum_{i=1}^{Q}\frac{\rho(z_i)w_i}{z_i-\xi}$$



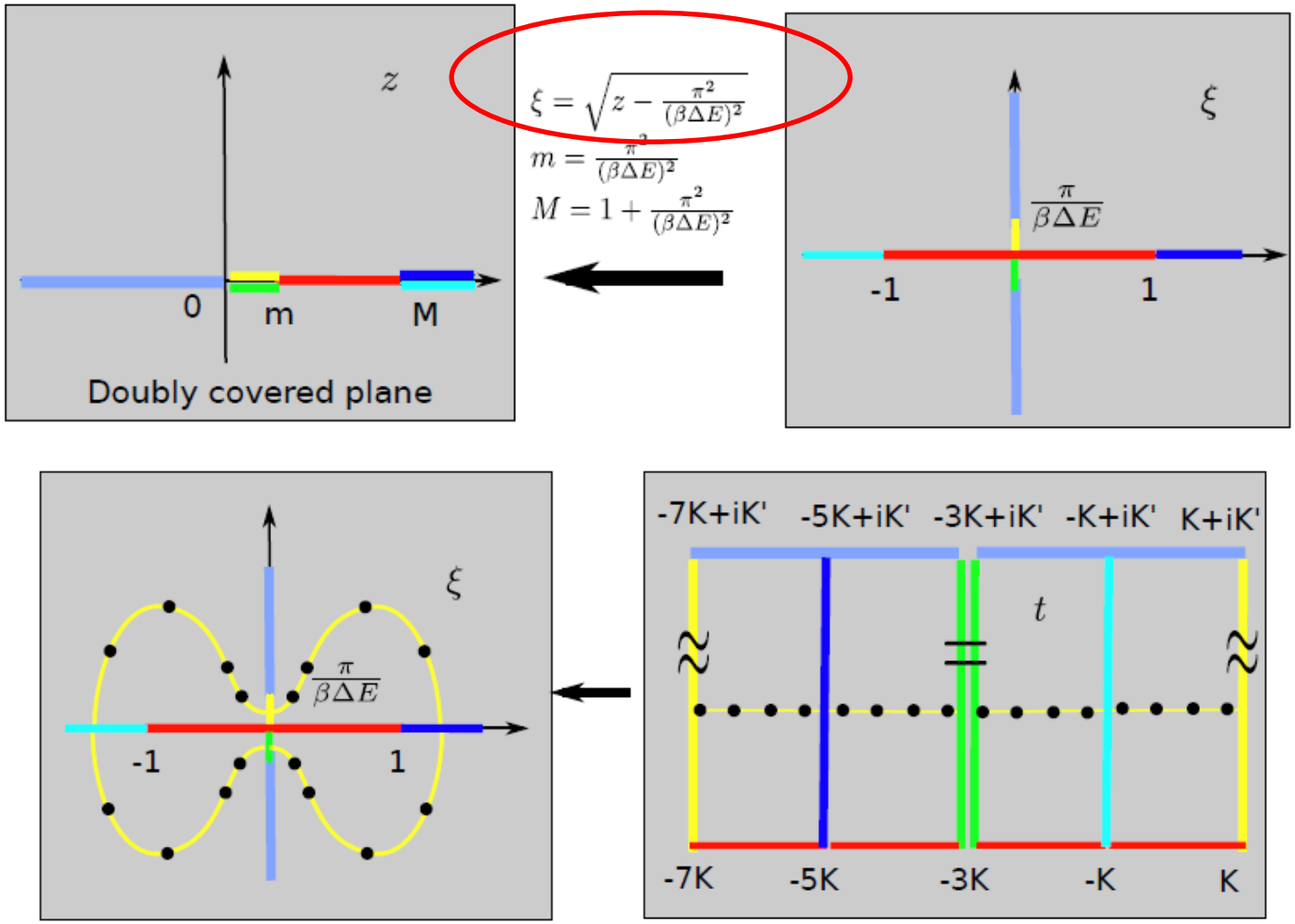
Simpler problem

[Hale, Higham and Trefethen, 2008]

$$\|\rho(\xi) - \rho_Q(\xi)\| \sim O(e^{-CQ/\log(M/m)})$$

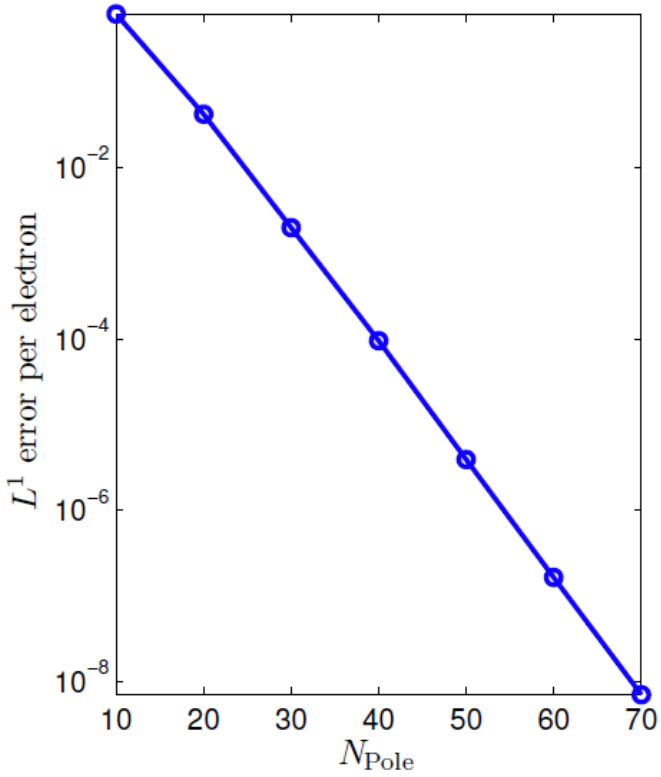


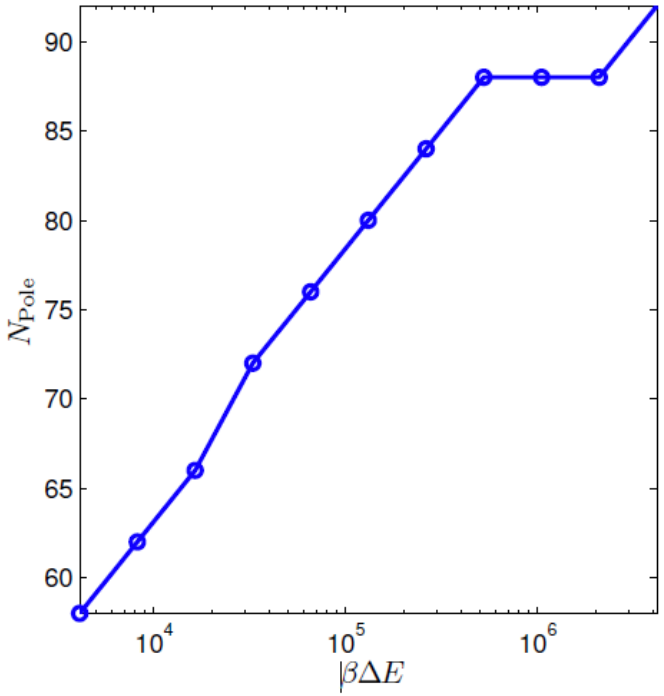
Pole expansion



Numerical result

Model tight binding problem





Outline

PEXSI: Pole EXpansion and Selected Inversion

- Pole Expansion
- Selected Inversion
- From model problem to practice

Selected inversion

$$\rho \approx \mathrm{diag} \sum_{i=1}^{Q} \frac{\omega_i}{H - z_i I}$$

- All the diagonal elements of an inverse matrix.
- H is a sparse matrix, but $(H - z_i I)^{-1}$ is a full matrix.
- Naïve approach: $O(N^3)$.
- Need selected inversion.

Selected inversion: basic idea

- LDL^T factorization

$$A = \begin{pmatrix} A_{11} & A_{21}^T \\ A_{21} & \hat{A}_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ L_{21} & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} 1 & L_{21}^T \\ 0 & I \end{pmatrix}$$

$$L_{21} = A_{21}A_{11}^{-1}, \qquad S_{22} = \hat{A}_{22} - A_{21}L_{21}^T$$

- Inversion

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + L_{21}^T S_{22}^{-1} L_{21} & -L_{21}^T S_{22}^{-1} \\ -S_{22}^{-1} L_{21} & S_{22}^{-1} \end{pmatrix}$$

Observation:

If L_{21} is sparse, $L_{21}^T S_{22}^{-1} L_{21}$ only require rows and columns of S_{22}^{-1} corresponding to the sparsity pattern of L_{21}.

Recursive relation

$$S_{22} = \begin{pmatrix} A_{22} & A_{32}^T \\ A_{32} & \hat{A}_{33} \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 \\ L_{21} & I \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & L_{32} & I \end{pmatrix} \begin{pmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & \hat{A}_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & L_{32}^T \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} 1 & L_{21}^T \\ 0 & I \end{pmatrix}$$

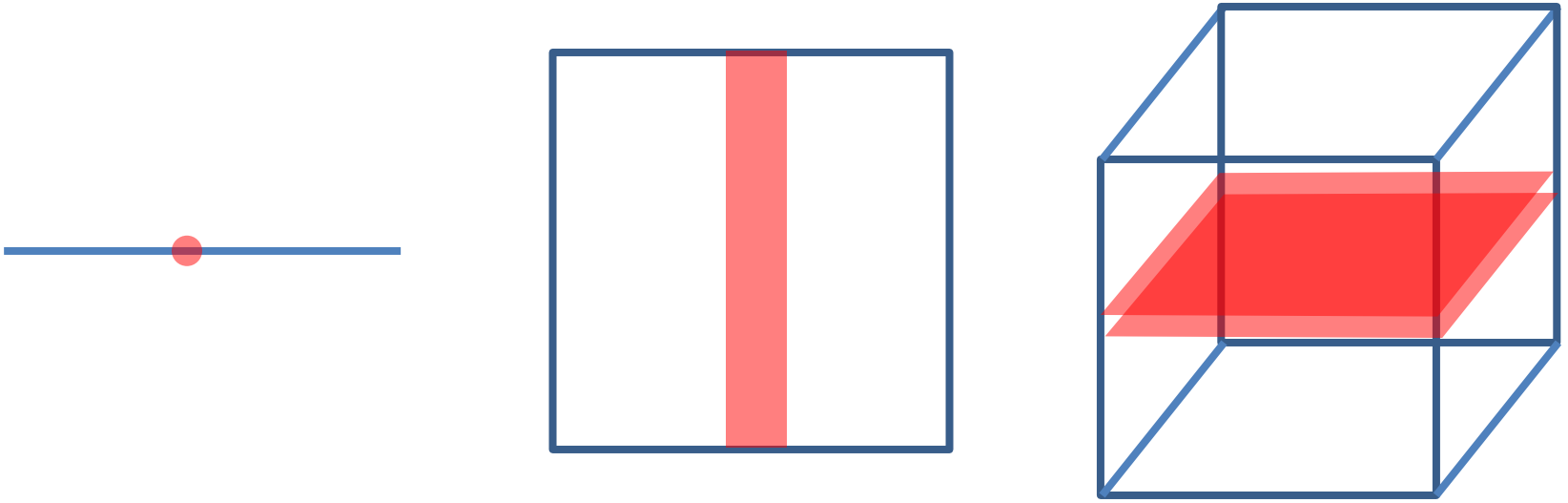
$$A^{-1} = \begin{pmatrix} A_{11}^{-1} + L_{21}^T S_{22}^{-1} L_{21} & -L_{21}^T S_{22}^{-1} \\ -S_{22}^{-1} L_{21} & \begin{matrix} A_{22}^{-1} + L_{32}^T S_{33}^{-1} L_{32} & -L_{32}^T S_{33}^{-1} \\ -S_{33}^{-1} L_{32} & S_{33}^{-1} \end{matrix} \end{pmatrix}$$

Selected inversion

- $A = LDL^T$: A^{-1} restricted to the non-zero pattern of L is "self-contained": Only compute $\left\{A_{ij}^{-1} \middle| L_{ij} \neq 0 \; or \; L_{ji} \neq 0\right\}$.
- Exact method with exact arithmetic.
- For KS Hamiltonian discretized by local basis set, the cost of selected inversion is $O(N)$ for 1D systems, $O(N^{1.5})$ for 2D systems, and $O(N^2)$ for 3D systems.
- Combined with pole expansion: At most $O(N^2)$ scaling for solving Kohn-Sham problem.
- Idea of selected inversion dates back to [Erisman and Tinney, 1975], [Takakashi et al 1973]; For electronic structure [LL-Lu-Ying-Car-E, 2009]; For quantum transport [Li, Darve et al, 2008, 2012]

Complexity of selected inversion

- For $d > 1$, asymptotic cost usually dominated by the inversion of largest Schur complement.



1D: $O(N)$

2D: $O(N^{0.5})^3 = O(N^{1.5})$

3D: $O\left(N^{2/3}\right)^3 = O(N^2)$

SelInv: Numerical results

SelInv: a selected inversion package for general sparse symmetric matrix written in FORTRAN.
[LL-Yang-Meza-Lu-Ying-E, 2011]

Problems from Harwell-Boeing Test Collection and the University of Florida Matrix Collection.

problem	n	selected inversion time	direct inversion time	speedup
bcsstk14	1,806	0.01 sec	0.13 sec	13
bcsstk24	3,562	0.02 sec	0.58 sec	29
bcsstk28	4,410	0.02 sec	0.88 sec	44
bcsstk18	11,948	0.24 sec	5.73 sec	24
bodyy6	0.19,366	09 sec	5.37 sec	60
crystm03	24,696	0.78 sec	26.89 sec	34
wathen120	36,441	0.34 sec	48.34 sec	142
thermal1	82,654	0.44 sec	95.06 sec	216
shipsec1	140,874	17.66 sec	3346 sec	192
pwtk	217,918	14.55 sec	5135 sec	353
parabolic_fem	525,825	20.06 sec	7054 sec	352
tmt_sym	726,713	13.98 sec	$>$ 3 hours	$>$ 772
ecology2	999,999	16.04 sec	$>$ 3 hours	$>$ 673
G3_circuit	1,585,478	218.7 sec	$>$ 3 hours	$>$ 49

Outline

PEXSI: Pole EXpansion and Selected Inversion

- Pole Expansion
- Selected Inversion
- From model problem to practice

Challenges from Physicists/Chemists/Materials Scientists

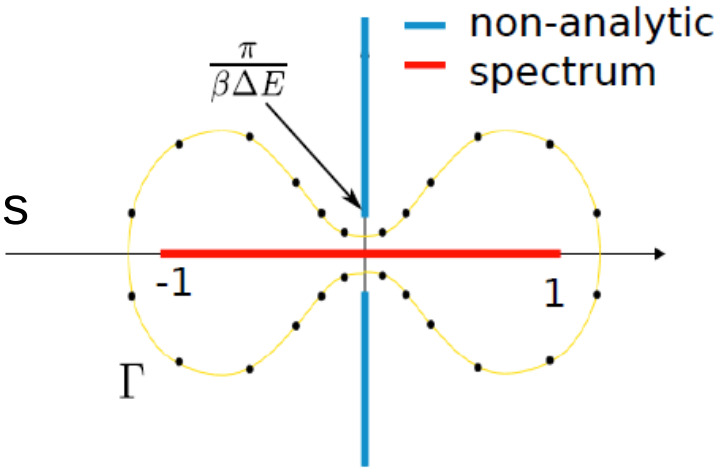
- Nonorthogonal basis 
- Various physical quantities 
- Parallel scalability 

Force

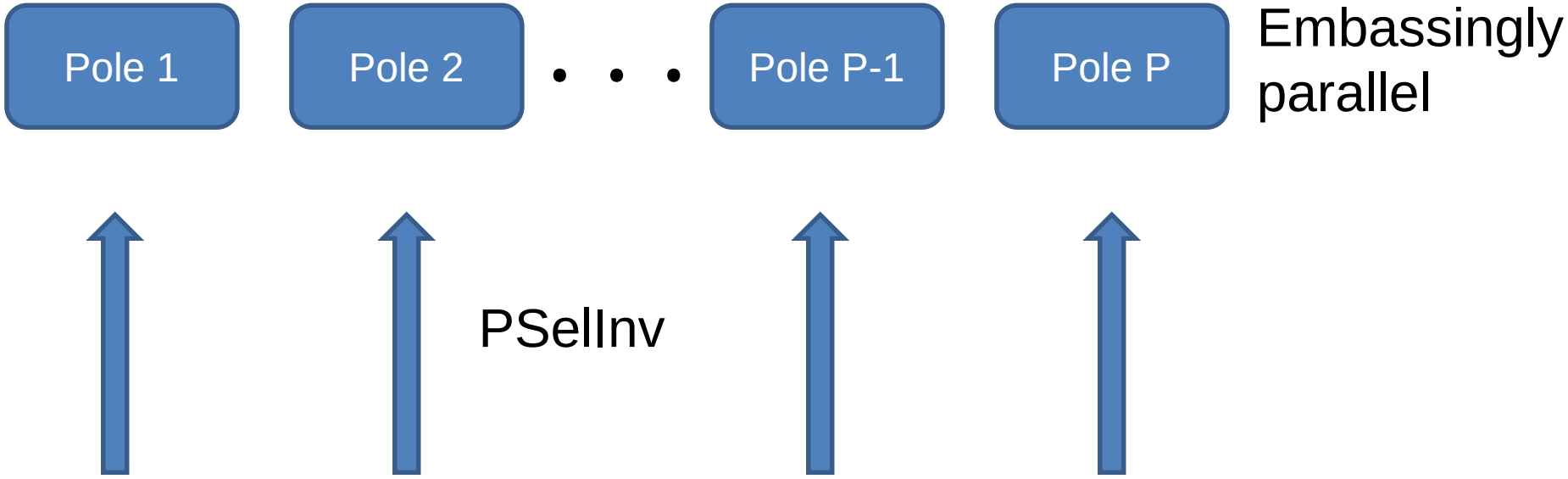
$$F_I = -Tr\left[\gamma \frac{\partial H}{\partial R_I}\right] + Tr\left[\gamma^E \frac{\partial S}{\partial R_I}\right]$$

- Including both the Hellmann-Feynman force and the Pulay force
- Energy density matrix
  $\gamma^E = C f^E(\Xi - \mu) C^T$
  $f^E(x-\mu) = x f(x-\mu)$
- Pole expansion with the same shift but different weight
- The same selected elements of $(H - z_i S)^{-1}$
- Similar treatment for other physical quantities

[LL-Chen-Yang-He, 2013]



Two level parallelization strategy



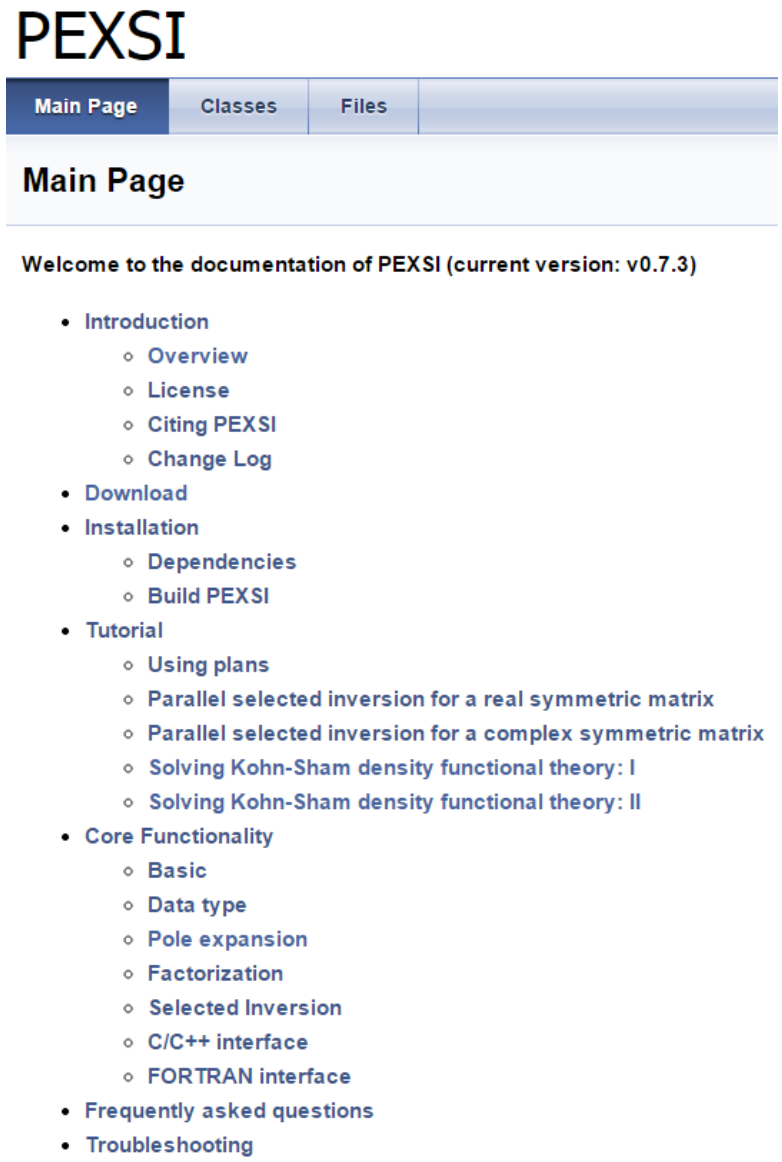
Example:
80 poles × 1024 procs per pole = 81920 procs.

PSelInv: Distributed memory parallel selected inversion for general sparse symmetric matrices

[Jacquelin-LL-Yang, 2014] [Jacquelin-LL-Wichmann-Yang, 2015]

http://www.pexsi.org/

- Work with sparse H, S matrices in distributed compressed sparse column (CSC) format
- Return density matrix etc in the same format
- Allow integration with different electronic structure software packages
- Integrated with SIESTA, CP2K and DGDFT



PEXSI

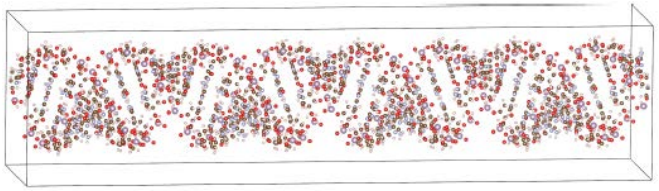
Main Page | Classes | Files

Main Page

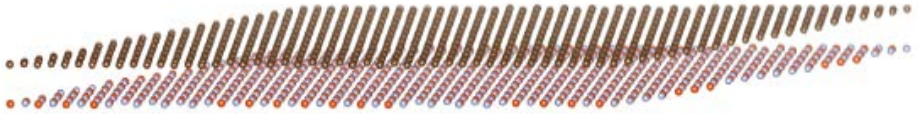
Welcome to the documentation of PEXSI (current version: v0.7.3)

- Introduction
  - Overview
  - License
  - Citing PEXSI
  - Change Log
- Download
- Installation
  - Dependencies
  - Build PEXSI
- Tutorial
  - Using plans
  - Parallel selected inversion for a real symmetric matrix
  - Parallel selected inversion for a complex symmetric matrix
  - Solving Kohn-Sham density functional theory: I
  - Solving Kohn-Sham density functional theory: II
- Core Functionality
  - Basic
  - Data type
  - Pole expansion
  - Factorization
  - Selected Inversion
  - C/C++ interface
  - FORTRAN interface
- Frequently asked questions
- Troubleshooting

Test systems

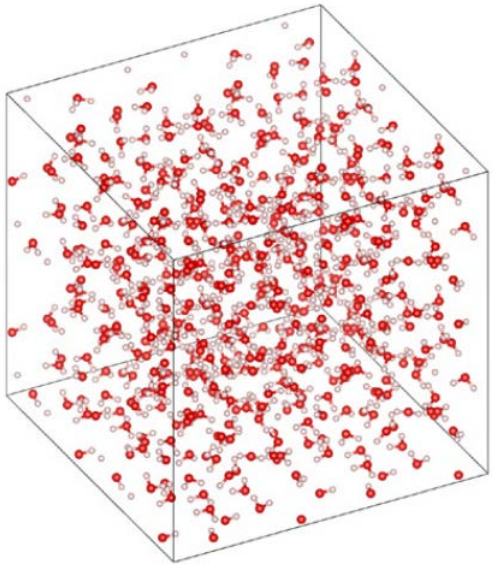


DNA



Graphene-Boron Nitride

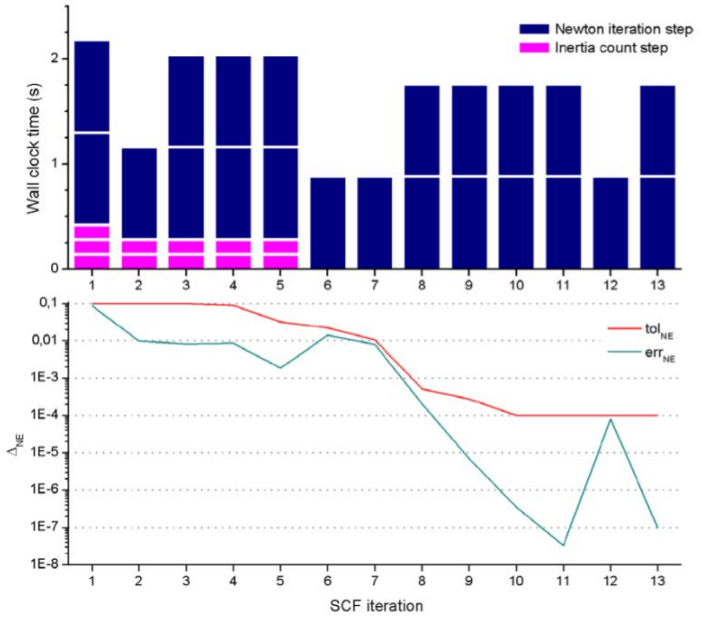
Example	Atoms	N	s_H (%)	s_{LU} (%)	l (nm)
DNA-1	715	7183	6.8	23	3.1
DNA-4	2860	28732	1.7	6.2	12
DNA-9	6435	64647	0.75	2.9	28
DNA-16	11440	114928	0.42	1.7	49
DNA-25	17875	179575	0.27	1.1	76
C-$BN_{2.3}$	1988	25844	5.9	36	5.5
C-$BN_{1.43}$	3874	50362	3.0	24	7.7
C-$BN_{0.57}$	7988	103844	1.5	15	11
C-$BN_{0.00}$	12770	166010	0.91	11	14
H_2O-8	1536	11776	2.3	28	2.5
H_2O-27	5184	39744	0.69	18	3.7
H_2O-64	12288	94208	0.29	12	5.0
H_2O-125	24000	184000	0.15	8.4	6.2
Al	512	6656	36	94	1.6
SiH	65	833	74	97	1.1



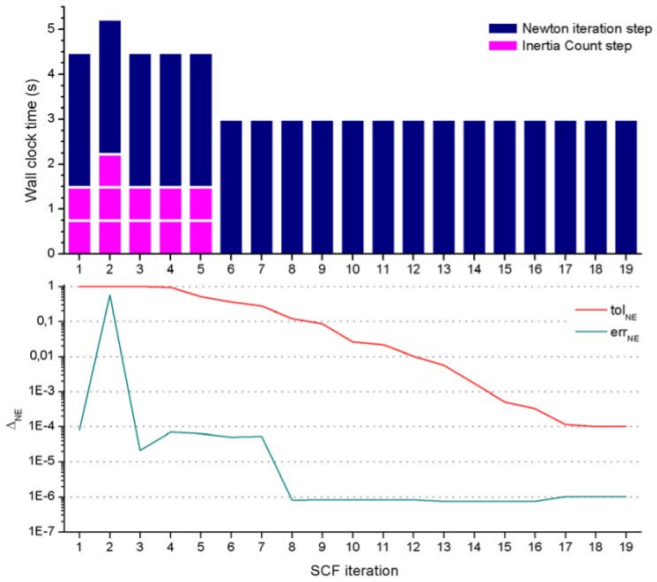
Water

Accuracy of energy and SCF convergence

System	P	tol_{NE}	err_E(eV/atom)	err_F(eV Å^{-1})
DNA-1	50	10^{-4}	2×10^{-8}	2×10^{-6}
C-$BN_{2.3}$	40	10^{-3}	3×10^{-6}	2×10^{-5}
H_2O-8	40	10^{-4}	3×10^{-6}	6×10^{-5}
Al	40	10^{-3}	6×10^{-6}	2×10^{-6}
SiH	40	10^{-3}	10^{-4}	7×10^{-5}
SiH	60	10^{-4}	6×10^{-6}	7×10^{-5}

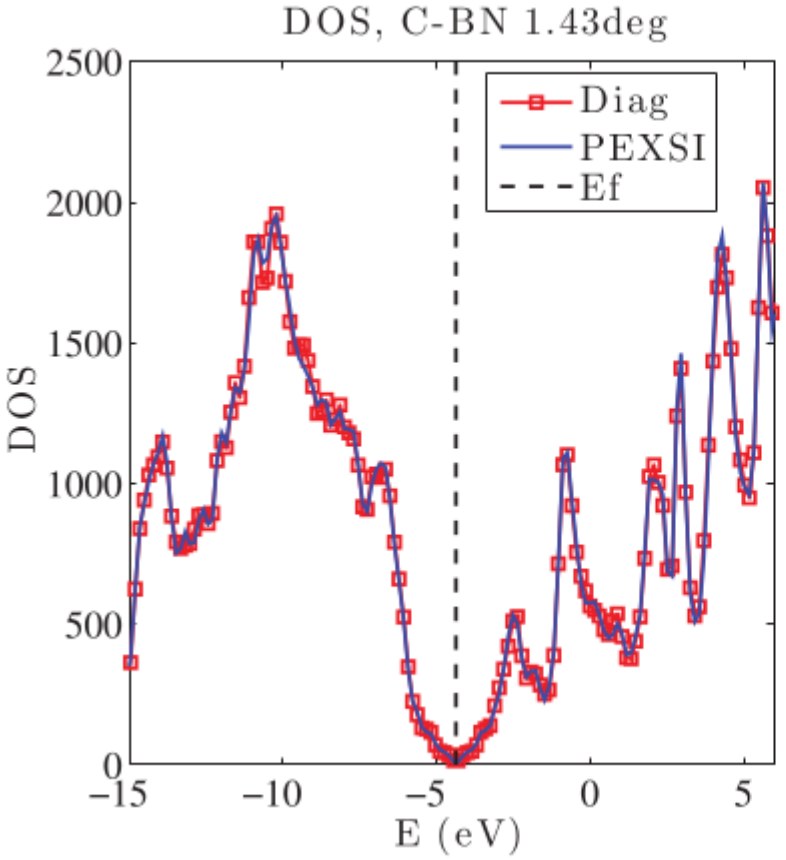


SiH (Metal)

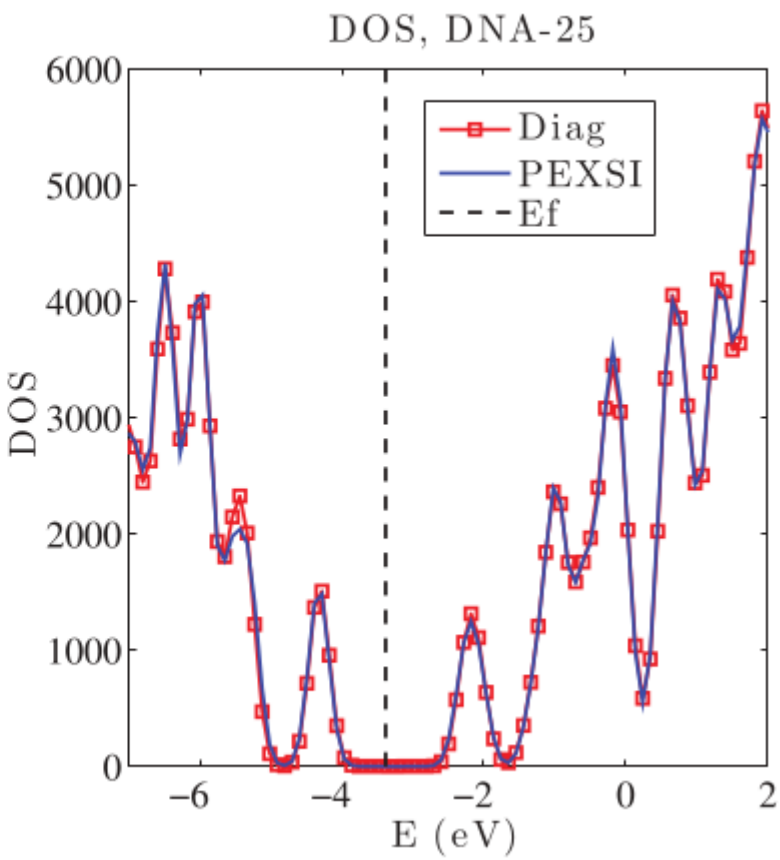


DNA (Insulator)

Accuracy: Density of States

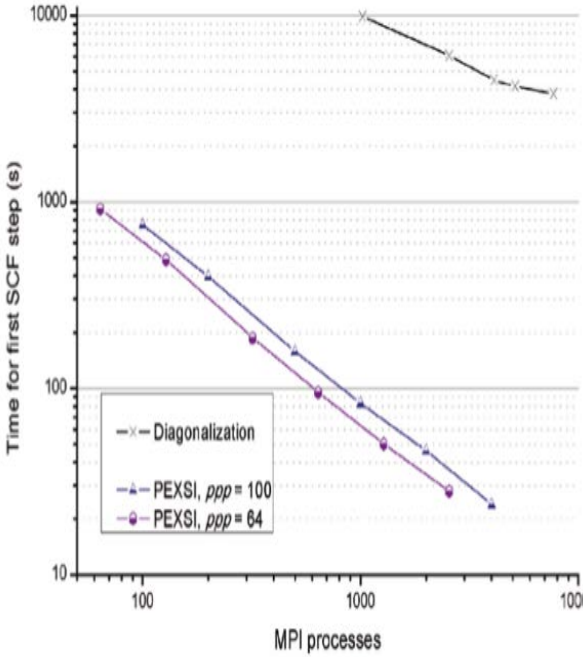


3874 atoms

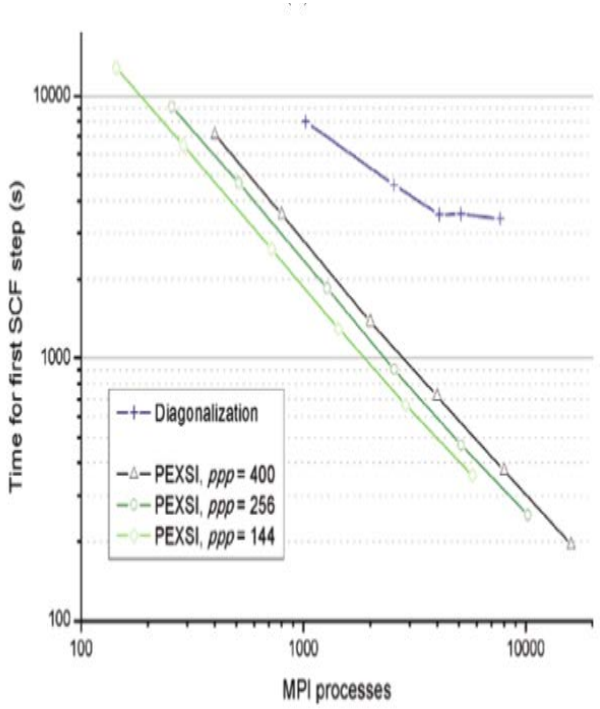


17875 atoms

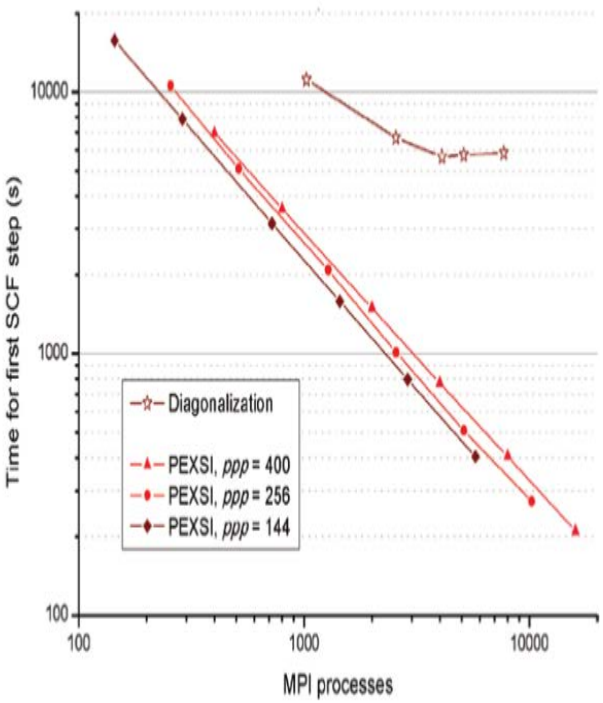
Efficiency: Wall clock time



(a) DNA-25



(b) C-BN$_{0.00}$



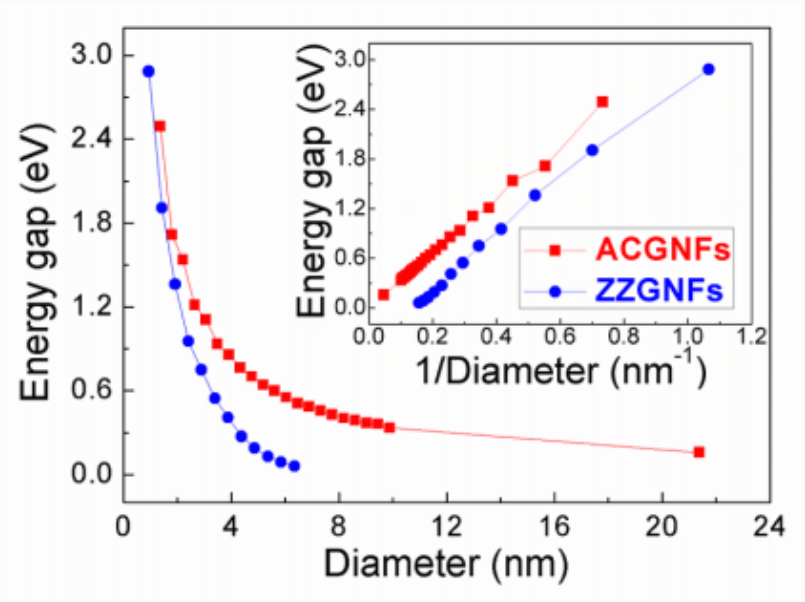
(c) H_2O-125

Speedup (4096 cores): 168

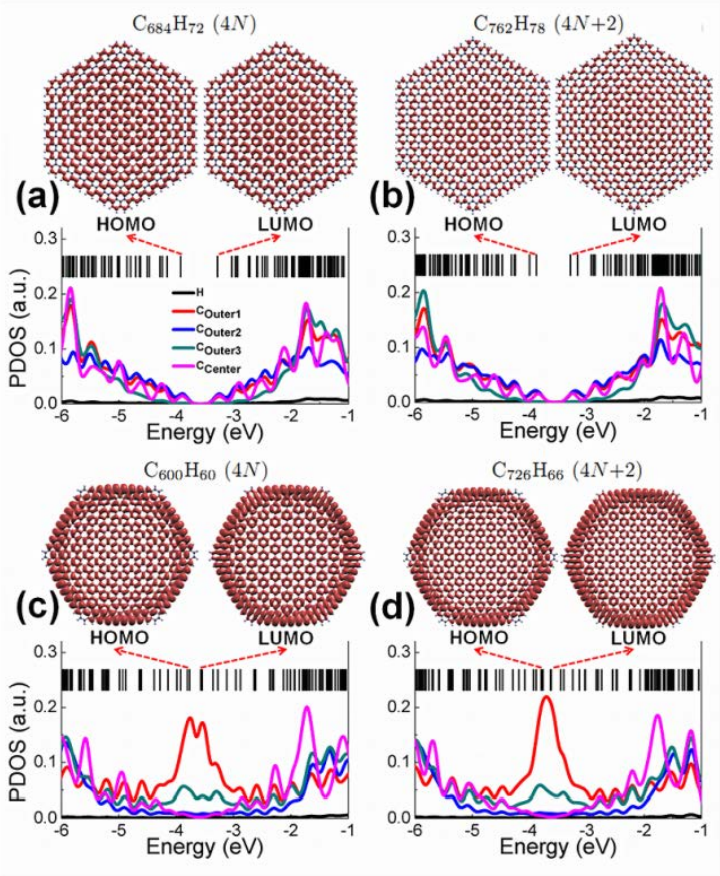
Speedup (4096 cores): 14

Speedup (4096 cores): 18

Electronic structure of large-scale graphene nanoflakes

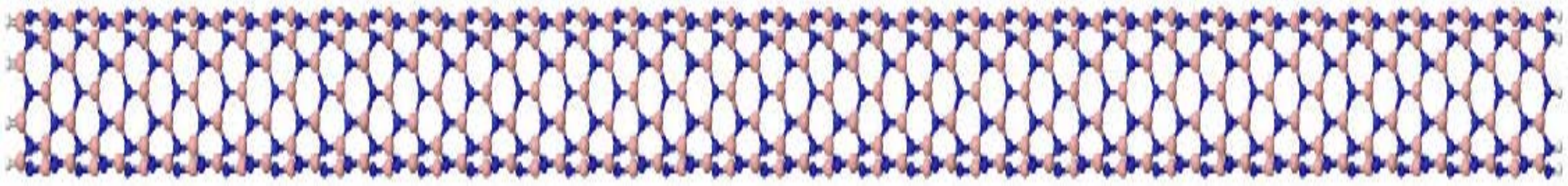


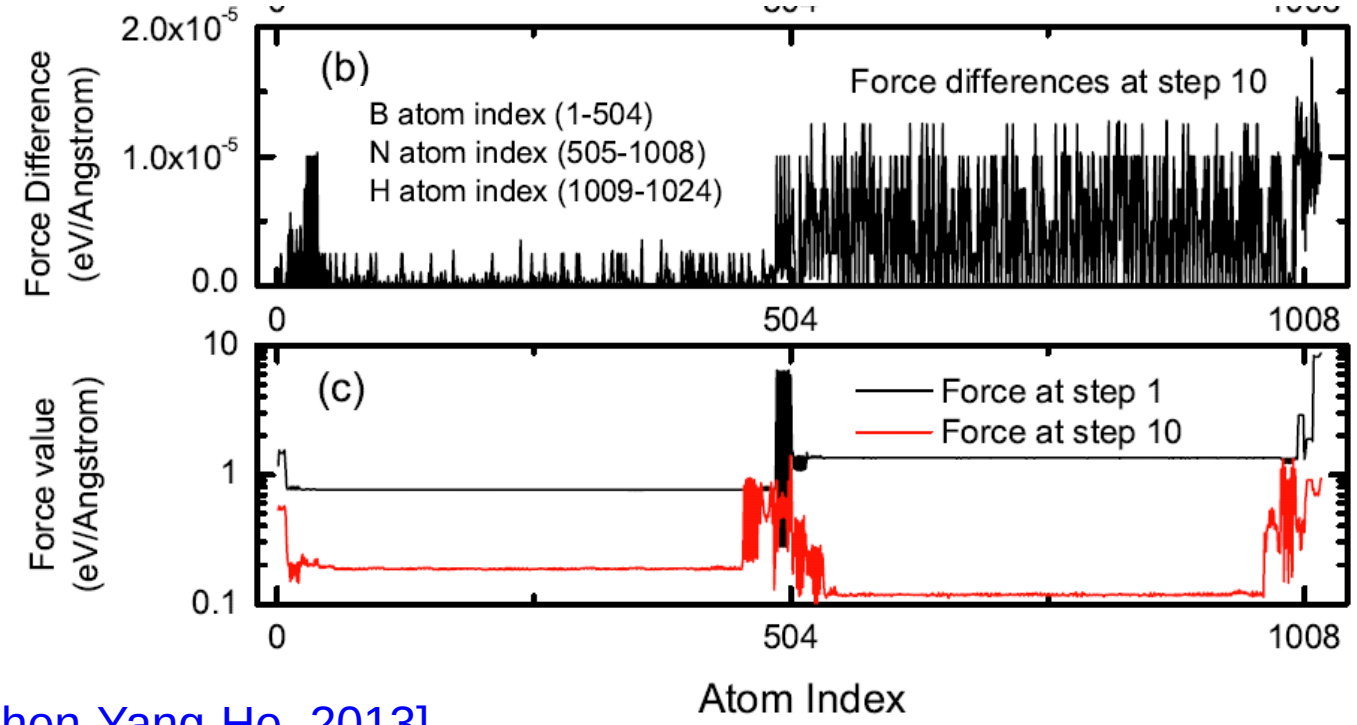
Unprecedented size: 11700 atoms



Geometry optimization: BNNT

Truncated Boron Nitride Nanotube (BNNT). 504 B atoms, 504 N atoms, 16 H atoms



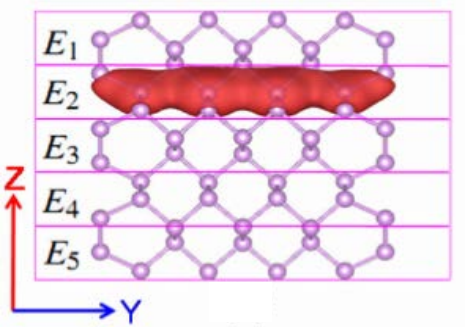


[LL-Chen-Yang-He, 2013]

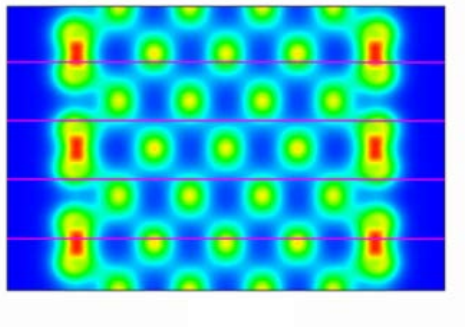
DGDFT-PEXSI

Adaptive local basis functions in a discontinuous Galerkin framework (DGDFT)

[LL-Lu-Ying-E 2012]

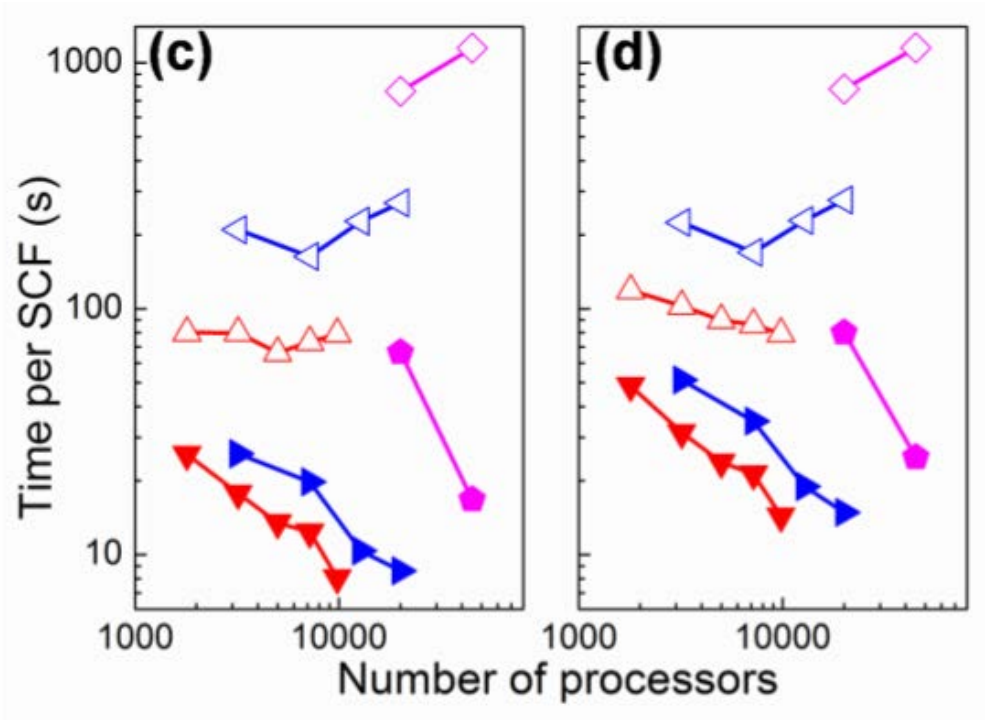


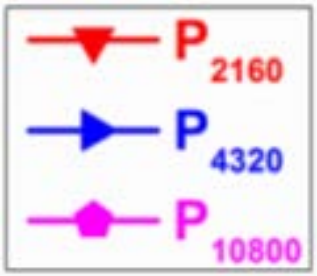
Basis (discontinuous)



Density (nearly continuous)

DGDFT-PEXSI



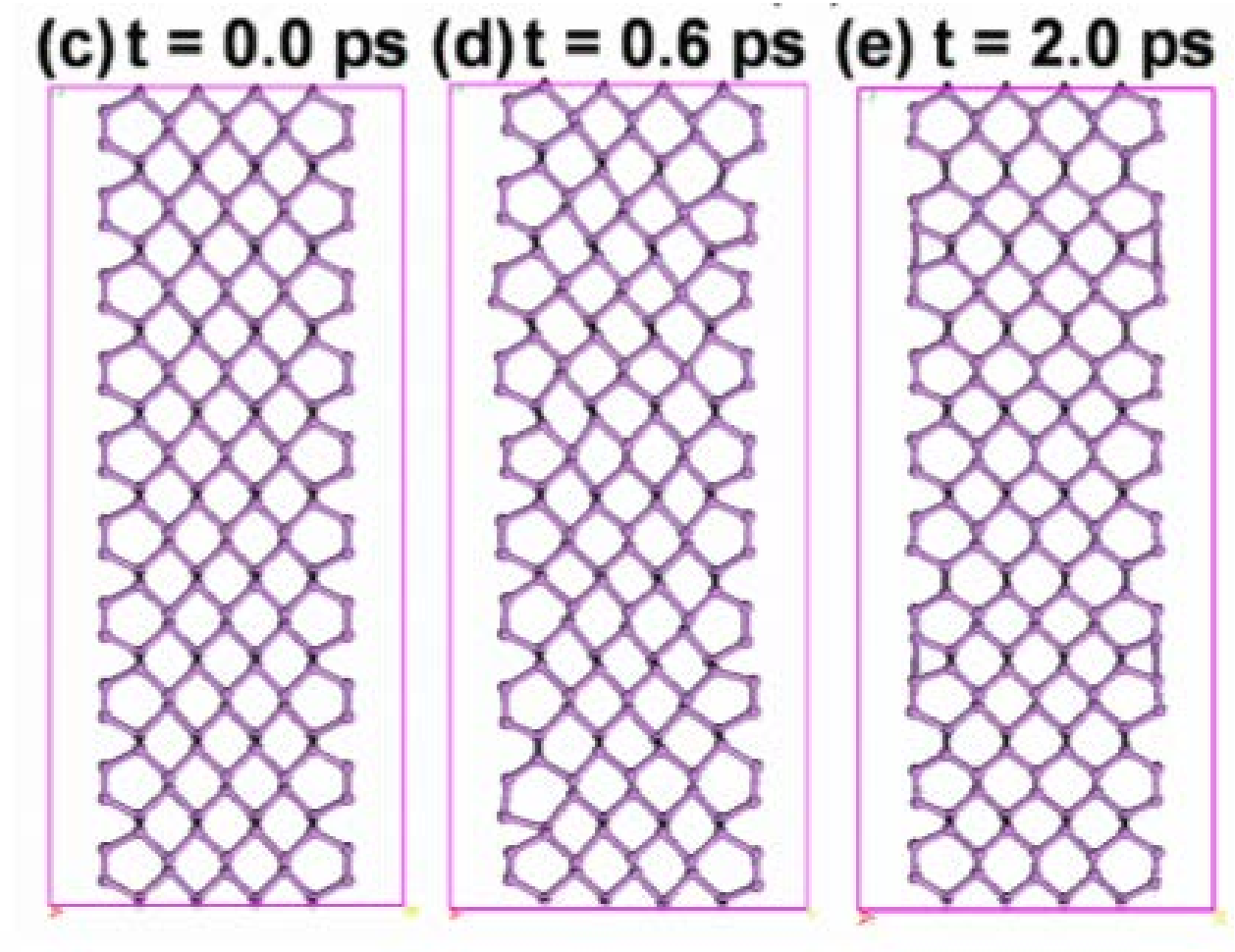


Empty: DIAG
Filled: PEXSI

[Hu-LL-Yang, 2015] [Hu-LL-Yang, in preparation]

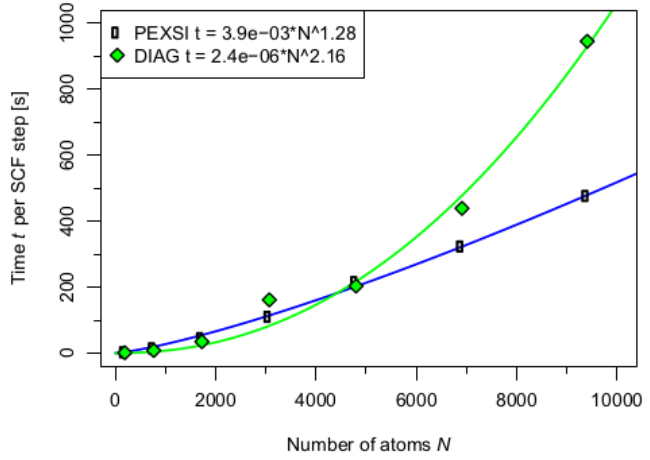
Most recent: scale to > 100,000 cores

Edge reconstruction of phospherene nanoribbon

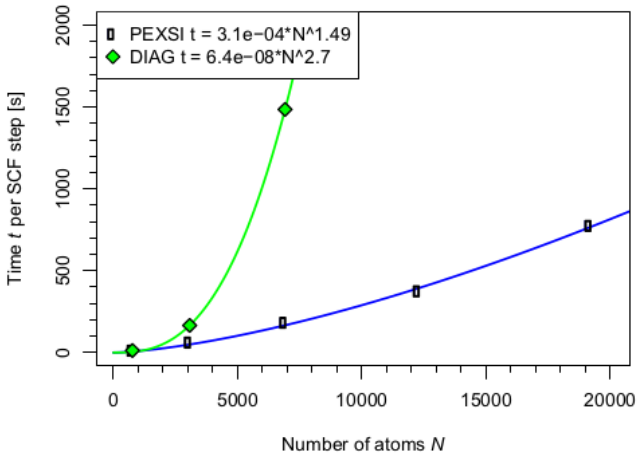


CP2K-PEXSI

CP2K: Gaussian orbital based electronic structure code



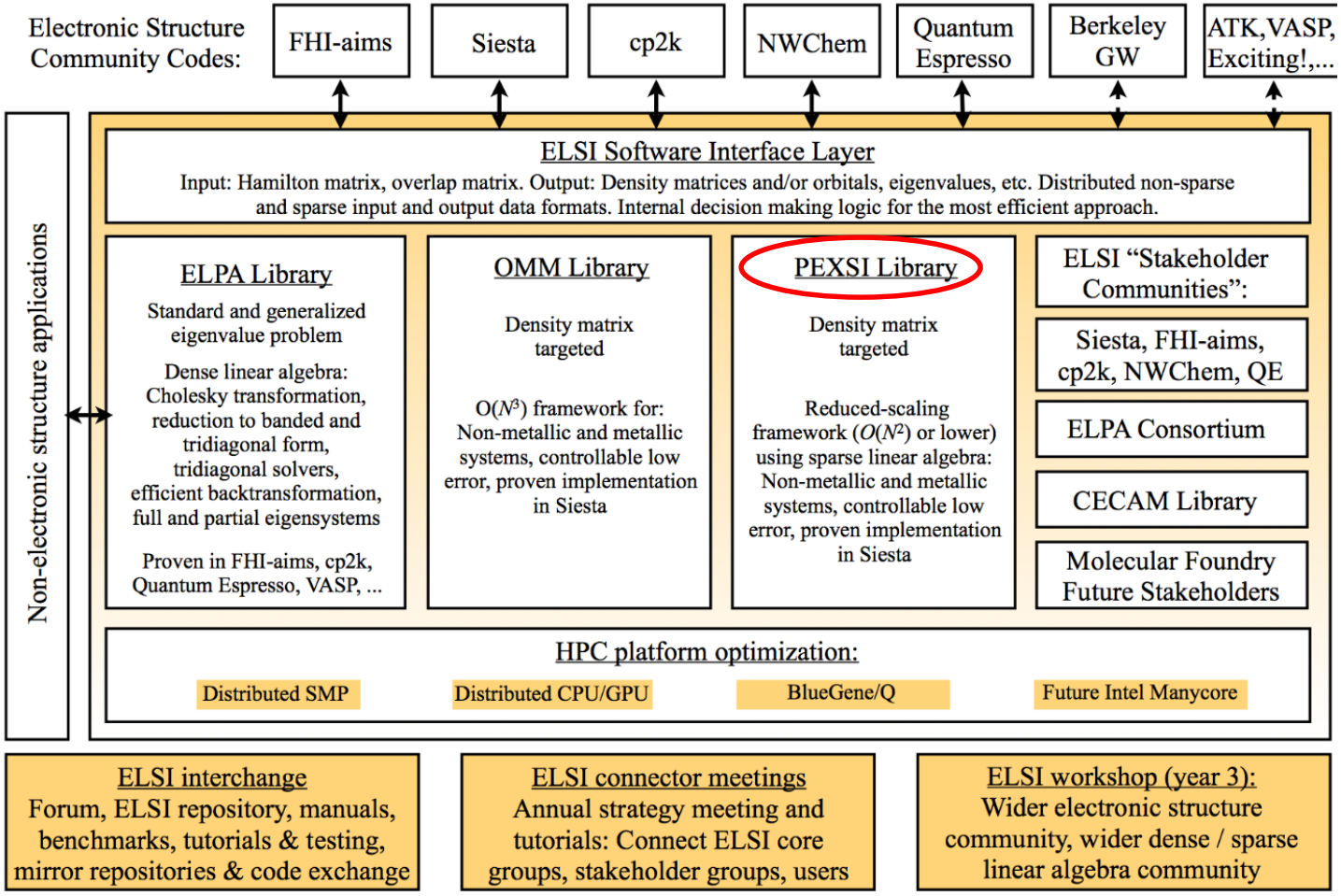
(a) Aluminium surface with 3 atomic layers, DZVP basis



(b) Monolayer graphene, DZVP basis

[Seewald, Master thesis, 2015]

ELSI Software Interface Layer (Planned)



Conclusion

- Pole Expansion and Selected Inversion (PEXSI) method for KSDFT at large scale.
- Based on the sparsity of Hamiltonian and overlap matrix. Require local basis set with small number of basis per atom (such as NAO and GTO, not applicable to PW)
- Accurate calculation of density, total energy, free energy and force (no truncation) for insulating and metallic systems.
- $O(N)$ for quasi-1D system, $O(N^{1.5})$ for quasi-2D system, and $O(N^2)$ for 3D bulk systems.
- Black-box: suitable for all codes localized basis set such atomic orbitals.

Future work

- Quantum transport
  - Parallel selected inversion for asymmetric matrices
  - Heterogeneous computational architecture
- Green's function for handling complex boundary conditions in materials
  - QM/QM coupling
  - QM/MM coupling [Nobel prize 2013]

Acknowledgment

Mathematics

Weinan E, Princeton

Jianfeng Lu, Duke

Chao Yang, LBNL

Lexing Ying, Stanford

Computer Science

Mathias Jacquelin, LBNL

Georg Huhs, BSC

Juan Meza, UC Merced

Nathan Wichmann, Cray Inc

Physics/Chemistry/Materials

Roberto Car, Princeton

Mohan Chen, Princeton

Alberto Garcia, ICMAB

Lixin He, USTC

Wei Hu, LBNL

Joost Vandevondele, ETH Zurich

Jinlong Yang, USTC

Support from DOE Alvarez fellowship, SciDAC, CAMERA funding