

# Solving Eigenvalue and Linear System Problems for 3D Photonic Device Simulations

---

**Weichung Wang**

Institute of Applied Mathematical Sciences  
National Taiwan University



A Celebration in Honor of Dianne P. O'Leary on the Occasion of her Retirement  
SIAM Conference on Applied Linear Algebra, Atlanta, USA, 2015/10/29

## Collaborators

---

**Wen-Wei Lin, Han-En Hsieh**

National Chiao-Tung University

**Tsung-Ming Huang**

National Taiwan Normal University

**Cheng-Han Du, Jiahong Chen,  
Jhihming Huang, Po-Chiuan Wang**

National Taiwan University

---

# FAME

Fast Algorithms for Maxwell's Equations

# Electromagnetic Waves



- Maxwell's equations

$$\nabla \times E = i\omega B$$

$$\nabla \times H = -i\omega D$$

$$\nabla \cdot B = 0$$

$$\nabla \cdot D = 0$$

$E$ : the electric field

$H$ : magnetic field

$\omega$ : frequency

$B$ : magnetic flux density

$D$ : electric flux density

- Constitutive relations

$$B = \mu H + \zeta E$$

$$D = \varepsilon E + \xi H$$

$\mu$ : magnetic permeability

$\varepsilon$ : electric permittivity

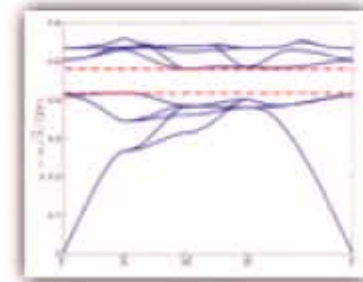
$\zeta$  and  $\xi$ : magnetoelectric parameters

# Dielectric Materials (Photonic Crystal)

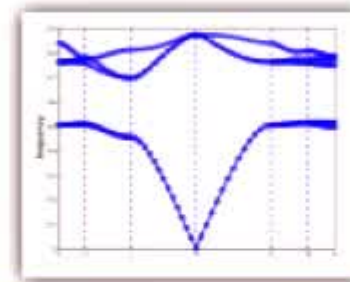
$$\nabla \times \nabla \times E = \mu_0 \omega^2 \epsilon(\mathbf{x}) E$$

$$\epsilon(\mathbf{x}) = \begin{cases} \epsilon_1 & \text{in material 1} \\ \epsilon_2 & \text{in material 2} \end{cases}$$

$$C^* C x = \lambda B x, \quad \lambda = \omega^2$$

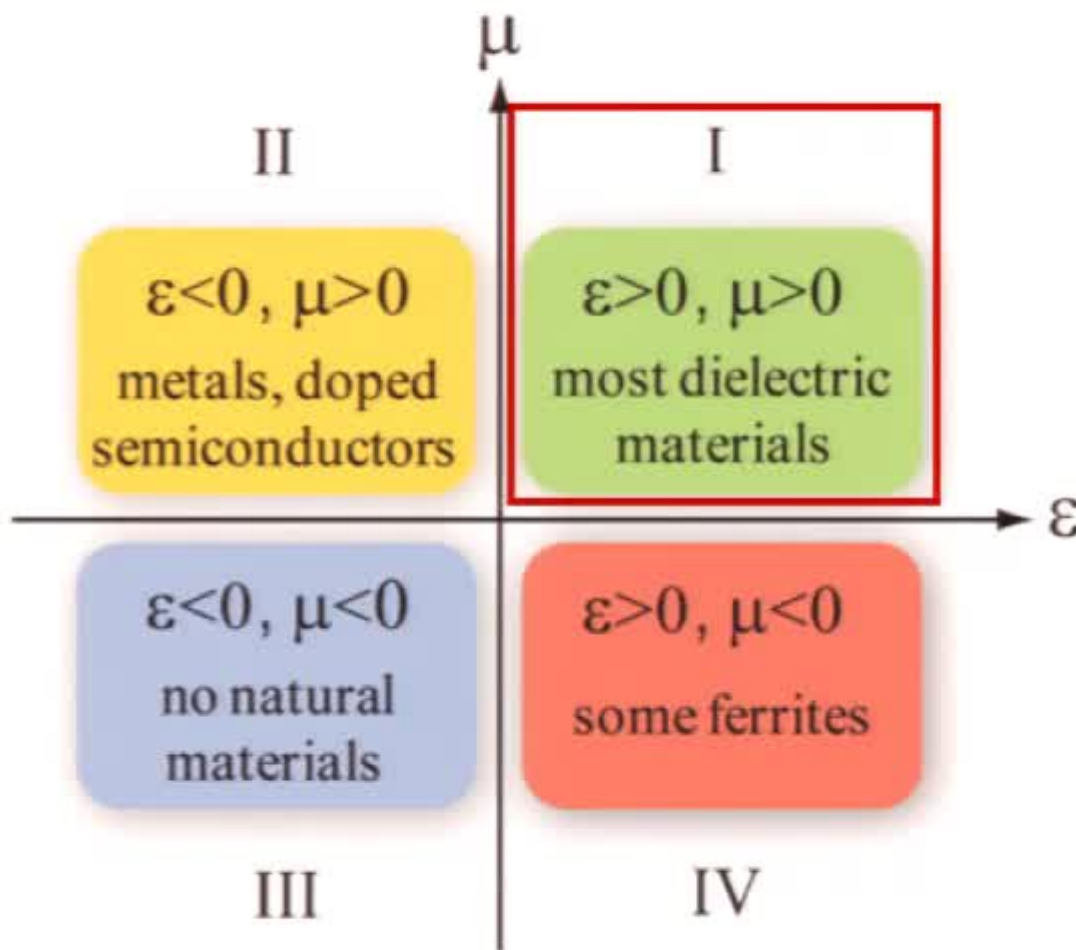


Huang/Chang/Huang/Lin/Wang/W (JCP 2010)  
Huang/Kuo/W (JSC 2013)



Eigendecomposition of discrete double-curl

Huang/Hsieh/Lin/W (SIMAX, 2013)





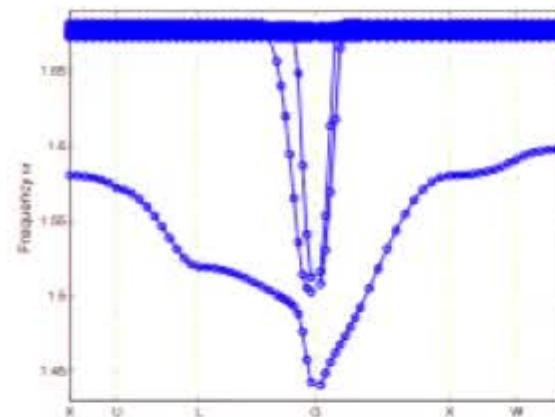
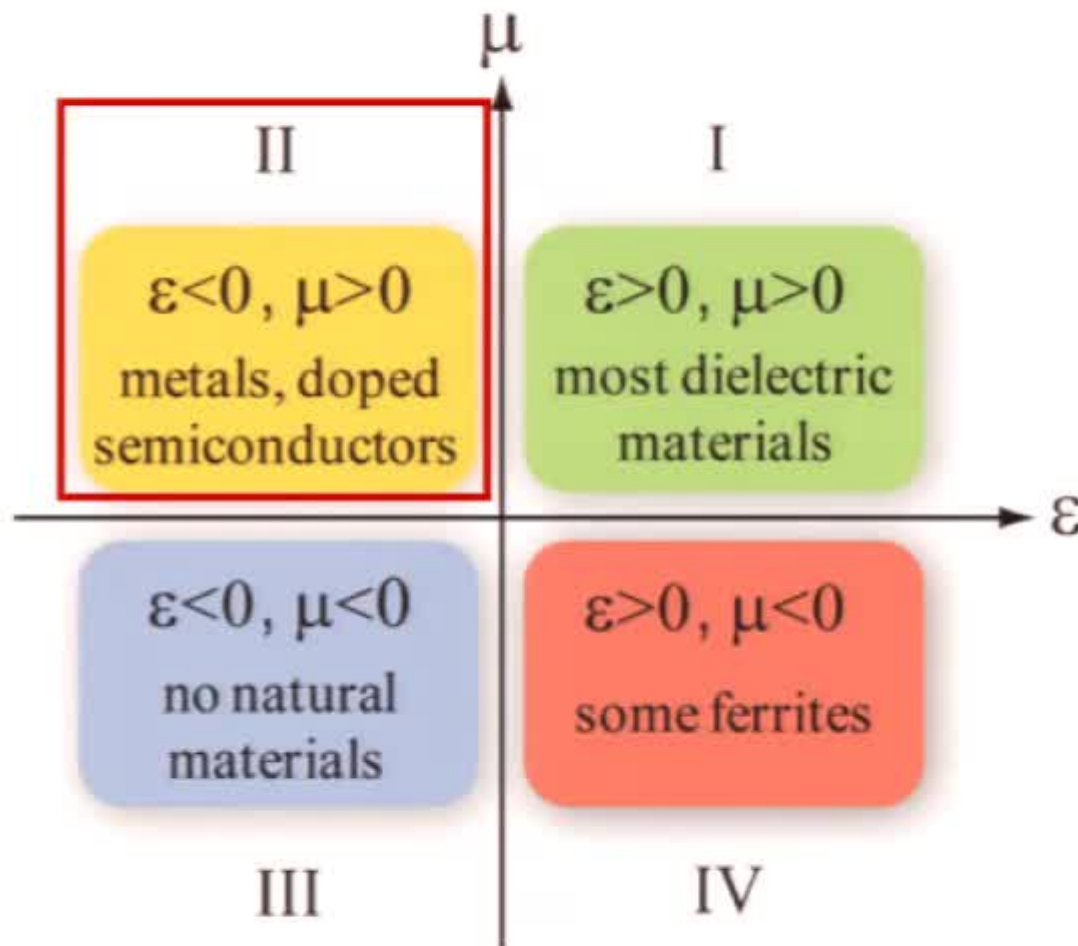
# Dispersive Metallic Materials



$$\nabla \times \nabla \times E = \mu_0 \omega^2 \varepsilon(\mathbf{x}, \omega) E$$

$$\varepsilon(\mathbf{x}, \omega) = \begin{cases} \varepsilon_1(\omega) & \text{in material 1} \\ \varepsilon_2 & \text{in material 2} \end{cases}$$

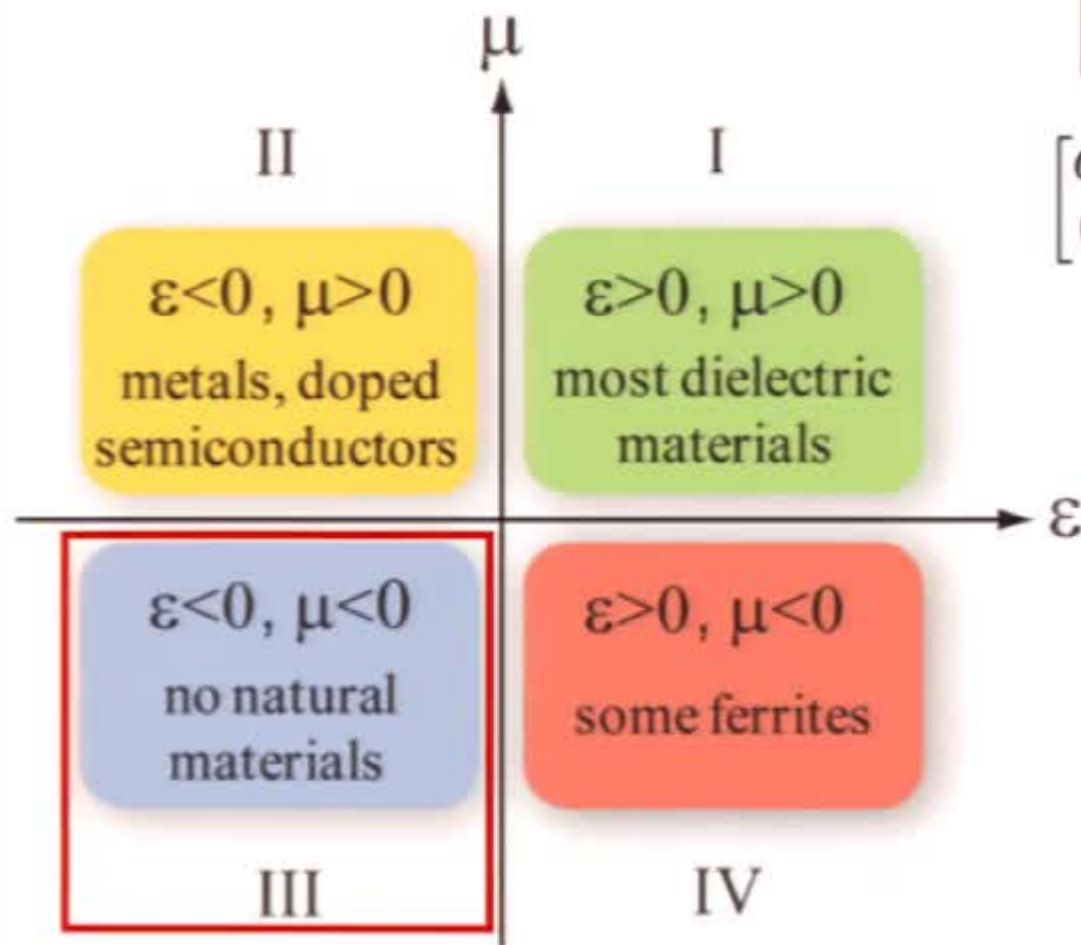
$$(C^* C)x = \omega^2 B(\omega)x$$



Eigensolver for cluster eigenvalues

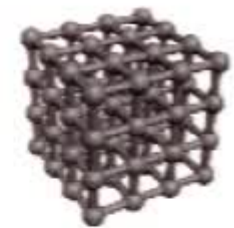
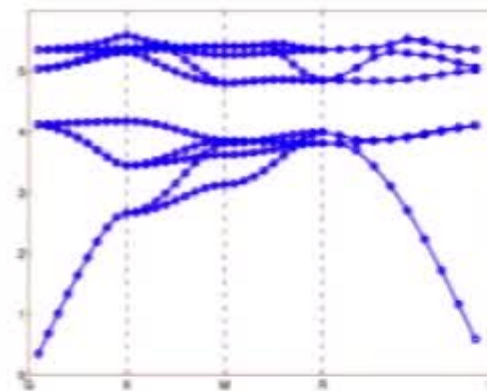
Huang/Lin/W (Preprint, 2015)

# Artificial Complex Media



$$\begin{bmatrix} \nabla \times & 0 \\ 0 & \nabla \times \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix} = \omega \begin{bmatrix} \zeta & \mu \\ -\varepsilon & -\xi \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix}$$

$$\begin{bmatrix} C & 0 \\ 0 & C^* \end{bmatrix} \begin{bmatrix} E \\ H \end{bmatrix} = \omega \left( \imath \begin{bmatrix} \zeta_d & \mu_d \\ -\varepsilon_d & -\xi_d \end{bmatrix} \right) \begin{bmatrix} E \\ H \end{bmatrix}$$



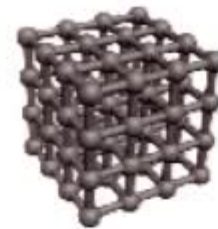
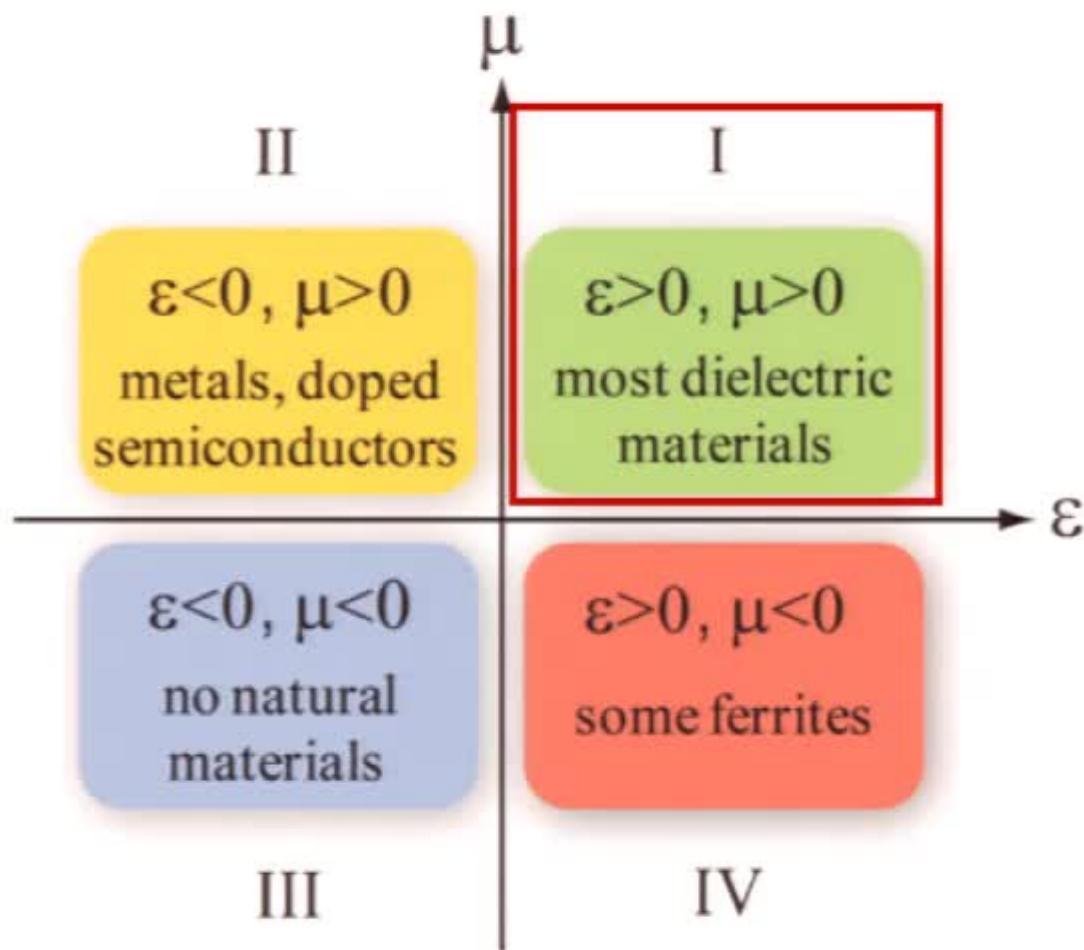
**SVD of discrete single-curl**

Chern/Hsieh/Huang/Lin/W (SIMAX, 2015)

# Photonic Crystals



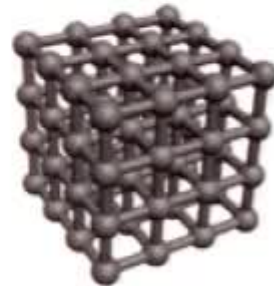
# Dielectric Materials (Photonic Crystal)



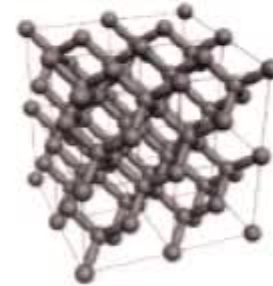
# Photonics Crystal



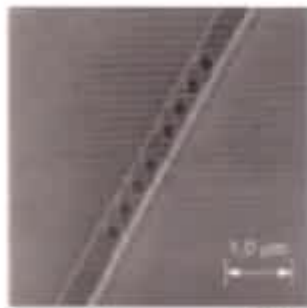
Periodic lattice composed of dielectric or metallic materials



Simple cubic

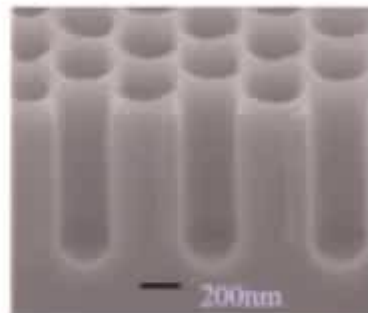


Face centered cubic



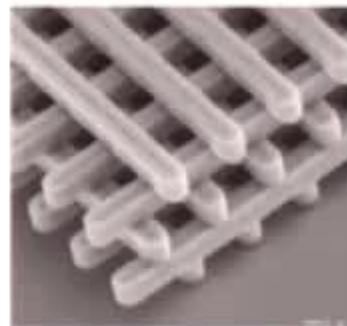
(MIT)

**1D hole-array built on A SOI substrate.**



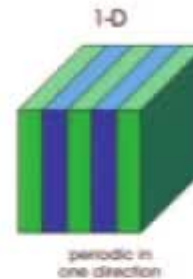
(Sandia)

**2D hole array built from GaAs on Al-oxide.**



(Sandia)

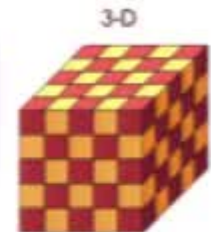
**3D diamond lattice built on a Si substrate.**



periodic in one direction



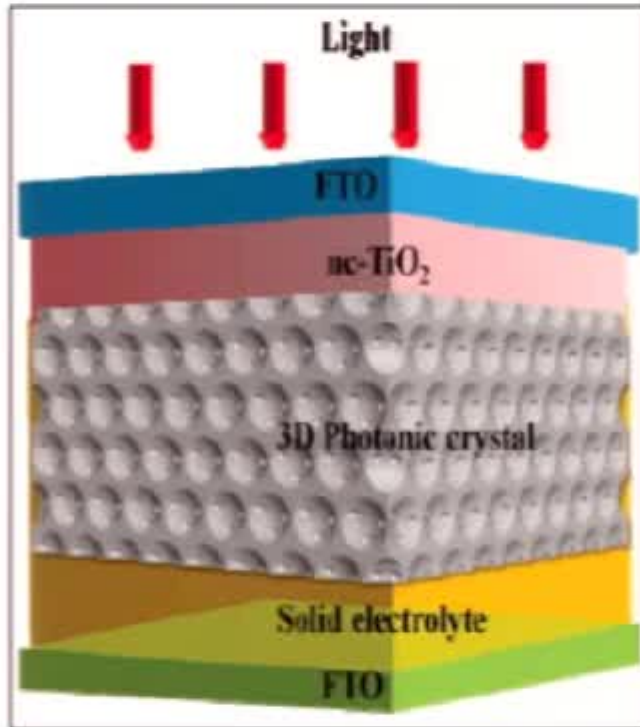
periodic in two directions



periodic in three directions

Slides by Shawn-Yu Lin, <http://goo.gl/zUaFDC>

# Control of Lights

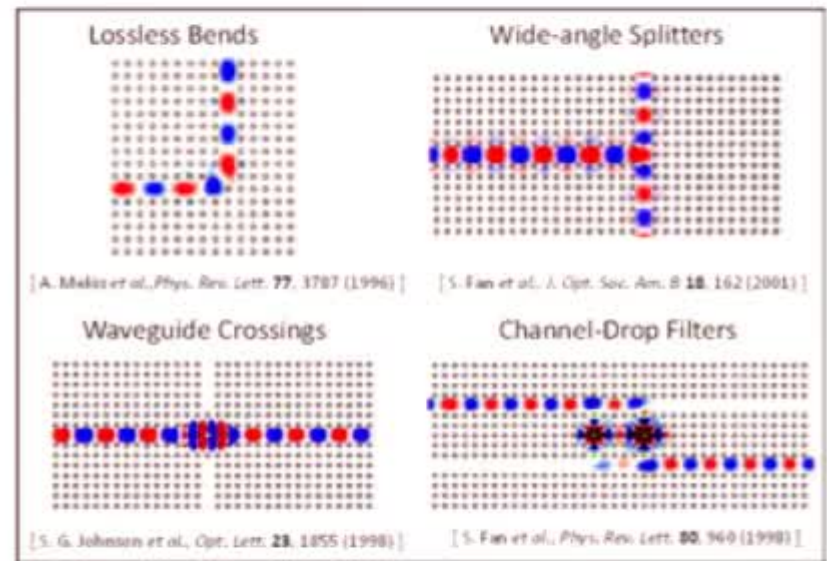
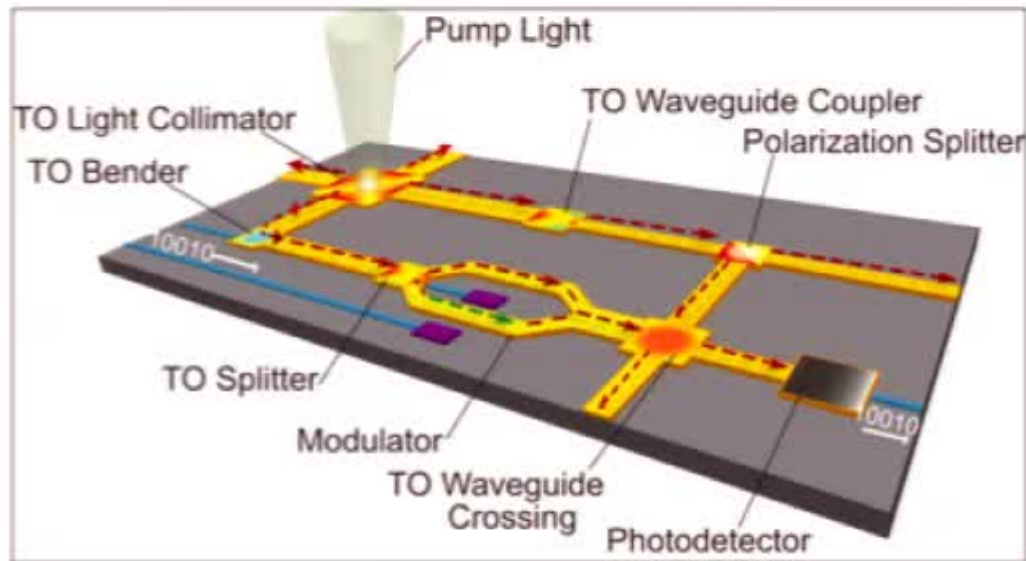


**Solar Cells**, Hwang et al,  
RSC Adv. (2013)



Photonic crystal **light twisters** at Sandia.  
<http://goo.gl/b3xtHk>

# Control of Lights



A schematic **photonic integrated circuit** for imaging, communications, computing, and sensing. (2012) <http://goo.gl/bUeR0F>



# Bandgap

---

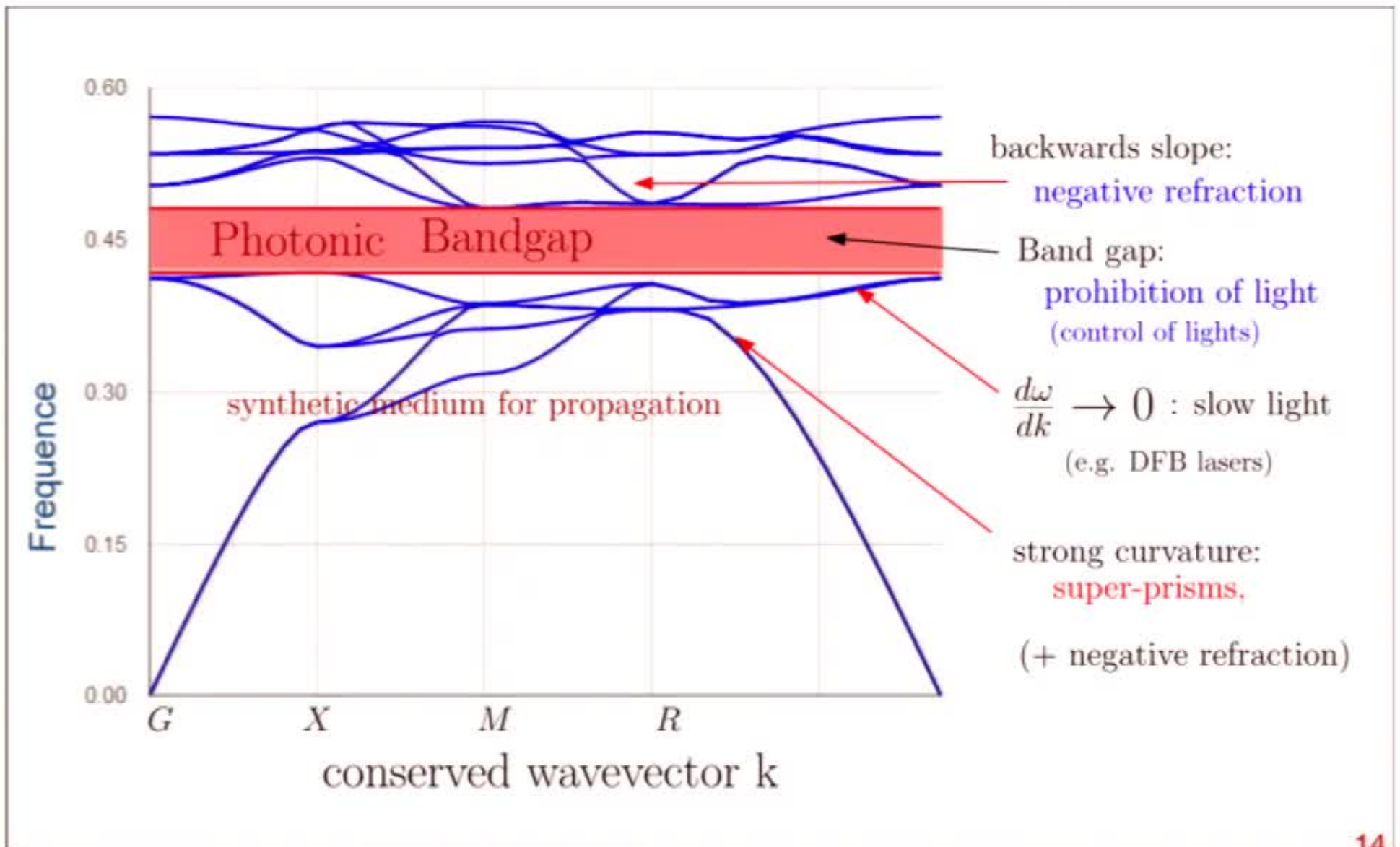


“If only it were possible to make dielectric materials in which electromagnetic waves cannot propagate at certain frequencies all kinds of almost-magical things would be possible.”

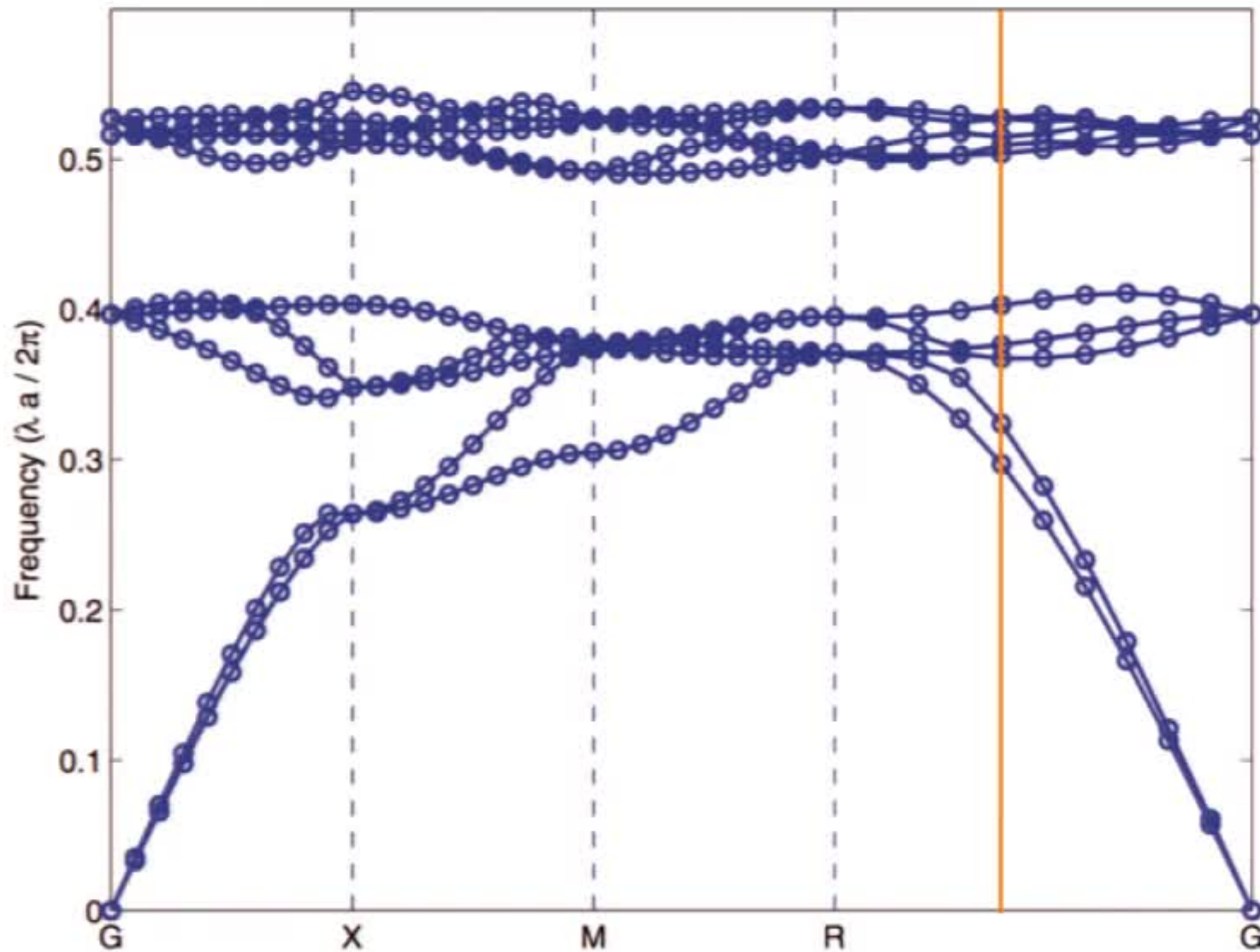
John Maddox, Nature, 348 (1990)



# Band Structure Diagram



# Many Large-scale Interior EVPs



# Numerical Challenges



- Yee's scheme discretizes the equation

$$\nabla \times \nabla \times \tilde{E}(\mathbf{x}) = \mu_0 \omega^2 \varepsilon(\mathbf{x}) \tilde{E}(\mathbf{x})$$

to get the generalized eigenvalue problem (GEVP)

$$A\mathbf{x} = \lambda B\mathbf{x}.$$

Huang/Hsieh/Lin/W (Math Comp Model 2013)

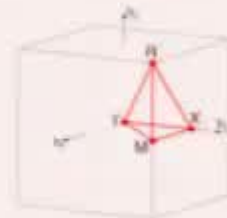
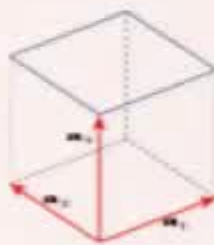
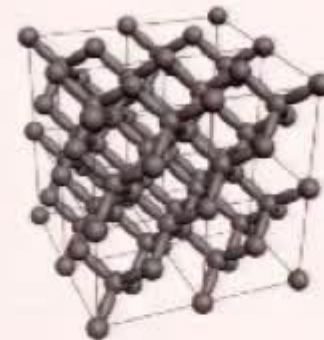
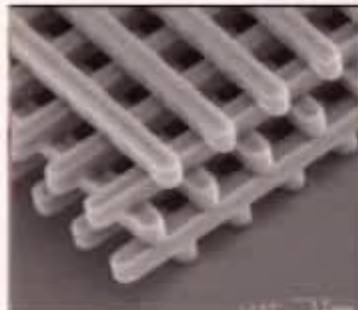
- $A$ : complex Hermitian positive semi-definite
- $B$ : positive diagonal (containing magnetic constant, frequency, material dependent permittivity)
- Dimension:  $n=3n_1n_2n_3$
- Need a few of smallest interior positive eigenvalues
- $A$  has big  $(n/3)$  null space

# Quasi-Periodic Conditions

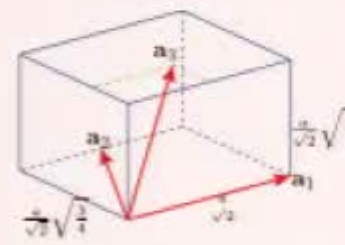


For a Bloch wave vector  $2\pi\mathbf{k}$  in the *irreducible Brillouin zone*, the Bloch eigenfunctions  $E$  and  $H$  satisfy the quasi-periodic condition along the translation vectors  $\mathbf{a}_\ell$ .

$$E(\mathbf{x} + \mathbf{a}_\ell) = e^{i2\pi\mathbf{k}\cdot\mathbf{a}_\ell} E(\mathbf{x}), \quad H(\mathbf{x} + \mathbf{a}_\ell) = e^{i2\pi\mathbf{k}\cdot\mathbf{a}_\ell} H(\mathbf{x}).$$



CCB path:  $\Gamma$ -X-M- $\Gamma$ -X-M- $\Gamma$



FCC path:  $\Gamma$ -X-W-K- $\Gamma$ -X-W-K- $\Gamma$



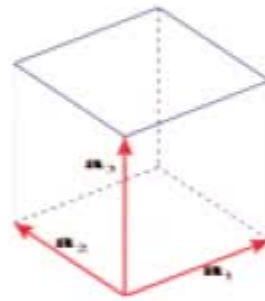
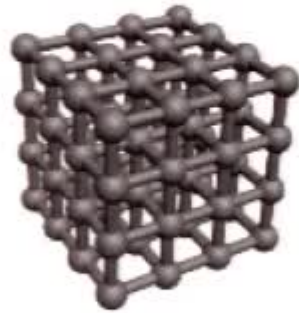
# Simple Cubic PC

---

FFT based Preconditioner



# Simple Cubic



$$Ax = \lambda Bx$$

$$(A - \tau B)z = b$$

Is there a way to find a  
preconditioner to cluster  
the eigenvalues?

# Simple Cubic: Preconditioner



$$(A - \tau B)\mathbf{z} = \mathbf{b}$$

$$(A - \tau \varepsilon_0 I)\mathbf{z} = \mathbf{b} \quad \text{by averaging } \text{diag}(\mathbf{B})$$

$$\Leftrightarrow (e^{-i\mathbf{k}\cdot\mathbf{x}} \nabla \times \nabla \times e^{i\mathbf{k}\cdot\mathbf{x}} - \varepsilon_0 \tau)\mathbf{z} = \mathbf{b}$$

$$\nabla^2 \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla \times (\nabla \times \mathbf{F})$$

$$\nabla \cdot (\nabla \times \mathbf{v}) = 0$$

$$\Leftrightarrow (-\nabla^2 - \varepsilon_0 \tau) e^{i\mathbf{k}\cdot\mathbf{x}} \mathbf{z} = e^{i\mathbf{k}\cdot\mathbf{x}} \mathbf{b} + \frac{1}{\varepsilon_0 \tau} \nabla(\nabla \cdot e^{i\mathbf{k}\cdot\mathbf{x}} \mathbf{b})$$

$$\Leftrightarrow \boxed{(\tilde{A} - \tau \varepsilon_0 I)} \tilde{\mathbf{z}} = \tilde{\mathbf{b}}$$

FFT based preconditioned

# Simple Cubic: Preconditioner

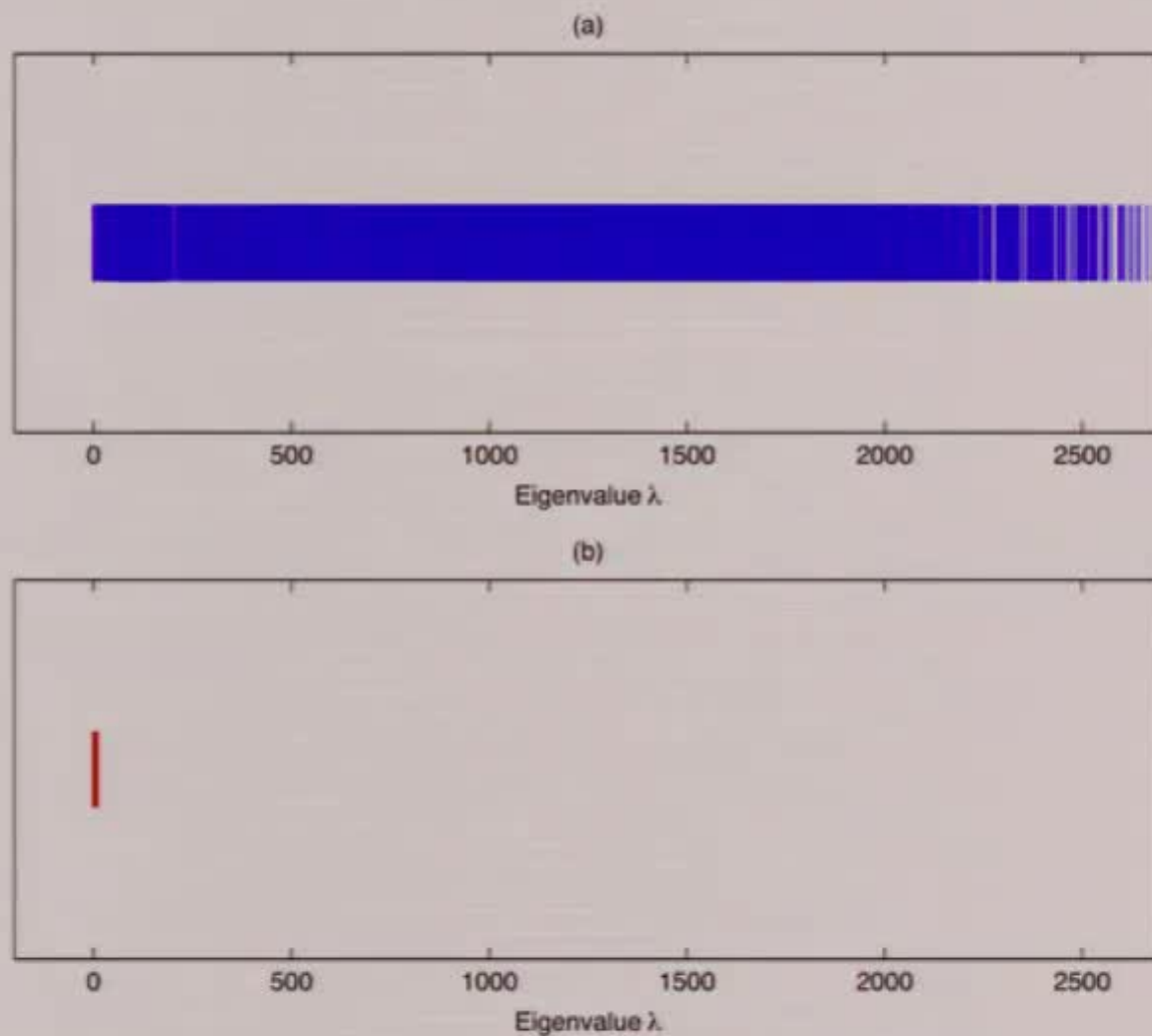


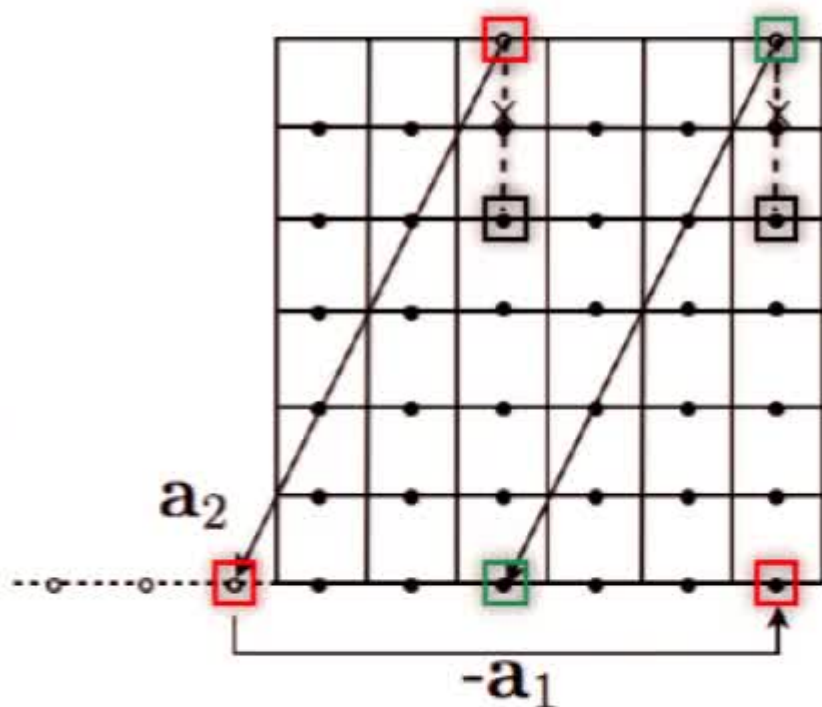
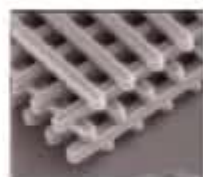
Fig. 4. Spectrum of (a) the matrix  $(A - \tau B)$  and (b) the matrix  $(A - \tau_0 J)^{-1}(A - \tau B)$ . The matrices sizes are equal to  $10,125 (3 \times 15^3)$ ,  $\mathbf{k} = (\pi, \pi, 0)$ , and  $\tau = 0.01$ .

# Face-Centered Cubic PC

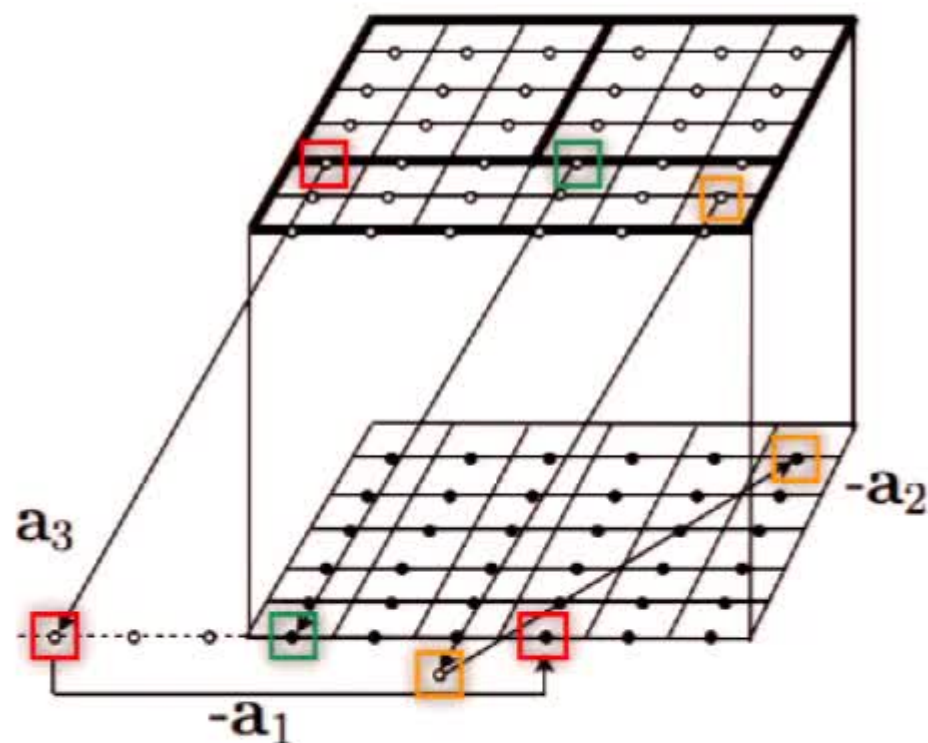
---

Eigendecomposition of Double Curl

# Preconditioner for 3D Coupling A?



$$\frac{\partial E_1}{\partial y}$$



$$\frac{\partial E_1}{\partial z}$$



# Main Result: Eigendecomposition of $A$

$$A = Q_r \Lambda_r Q_r^*.$$

$$Q_r = \left[ Q_1 (3\Lambda_q^2 - \Lambda_q \Lambda_p)^{-\frac{1}{2}} \quad Q_2 (3\Lambda_q - \Lambda_p)^{-\frac{1}{2}} \right] \equiv (I_3 \otimes T) \Lambda.$$

Null space

Range space

Define

$$Q = \begin{bmatrix} Q_0 & Q_1 & Q_2 \end{bmatrix} \text{diag} \left( \Lambda_q^{-\frac{1}{2}}, (3\Lambda_q^2 - \Lambda_q \Lambda_p)^{-\frac{1}{2}}, (3\Lambda_q - \Lambda_p)^{-\frac{1}{2}} \right).$$

Then  $Q$  is unitary. Furthermore,

$$Q^* A Q = \text{diag} (0, \Lambda_q, \Lambda_q)$$

# The Standard Eigenvalue Problem



$$\text{span} \left\{ B^{-1} Q_r \Lambda^{1/2} \right\} = \{ \mathbf{x} \mid A\mathbf{x} = \lambda B\mathbf{x}, \lambda > 0 \}$$

$$A\mathbf{x} = \lambda B\mathbf{x}$$

GEVP (3n-by-3n)

$$\left( \Lambda_r^{-\frac{1}{2}} Q_r^* \right) A \left( B^{-1} Q_r \Lambda_r^{\frac{1}{2}} \mathbf{y} \right) = \lambda \left( \Lambda_r^{-\frac{1}{2}} Q_r^* \right) B \left( B^{-1} Q_r \Lambda_r^{\frac{1}{2}} \mathbf{y} \right)$$

$$\left( \Lambda_r^{\frac{1}{2}} Q_r^* B^{-1} Q_r \Lambda_r^{\frac{1}{2}} \right) \mathbf{y} = \lambda \mathbf{y}$$

$$A_r \mathbf{y} = \lambda \mathbf{y}$$

SEVP (2n-by-2n)

Since  $A = Q_r \Lambda_r Q_r^*$ ,

$$\Rightarrow \Lambda_r^{-\frac{1}{2}} (Q_r^* A) = \Lambda_r^{-\frac{1}{2}} (\Lambda_r Q_r^*) = \Lambda_r^{\frac{1}{2}} Q_r^*$$

# To Compute $T^*p$ for $Q_r^*$



$$A_r \mathbf{y} = \lambda \mathbf{y}$$

$$A_r = \Lambda_r^{\frac{1}{2}} Q_r^* B^{-1} Q_r \Lambda_r^{\frac{1}{2}}$$

$$Q_r = \begin{bmatrix} Q_1 (3\Lambda_q^2 - \Lambda_q \Lambda_p)^{-\frac{1}{2}} & Q_2 (3\Lambda_q - \Lambda_p)^{-\frac{1}{2}} \end{bmatrix} \equiv (I_3 \otimes T) \Lambda.$$

- Rewrite  $T$  as **periodic** part (by FFT) and **diagonal** part

$$T_{i,j}^* \mathbf{p} = U_{\mathbf{z}}^* E_{\mathbf{z},i+j}^* P^\top (\overline{\mathbf{y}_{i,j} \otimes \mathbf{x}_i})$$

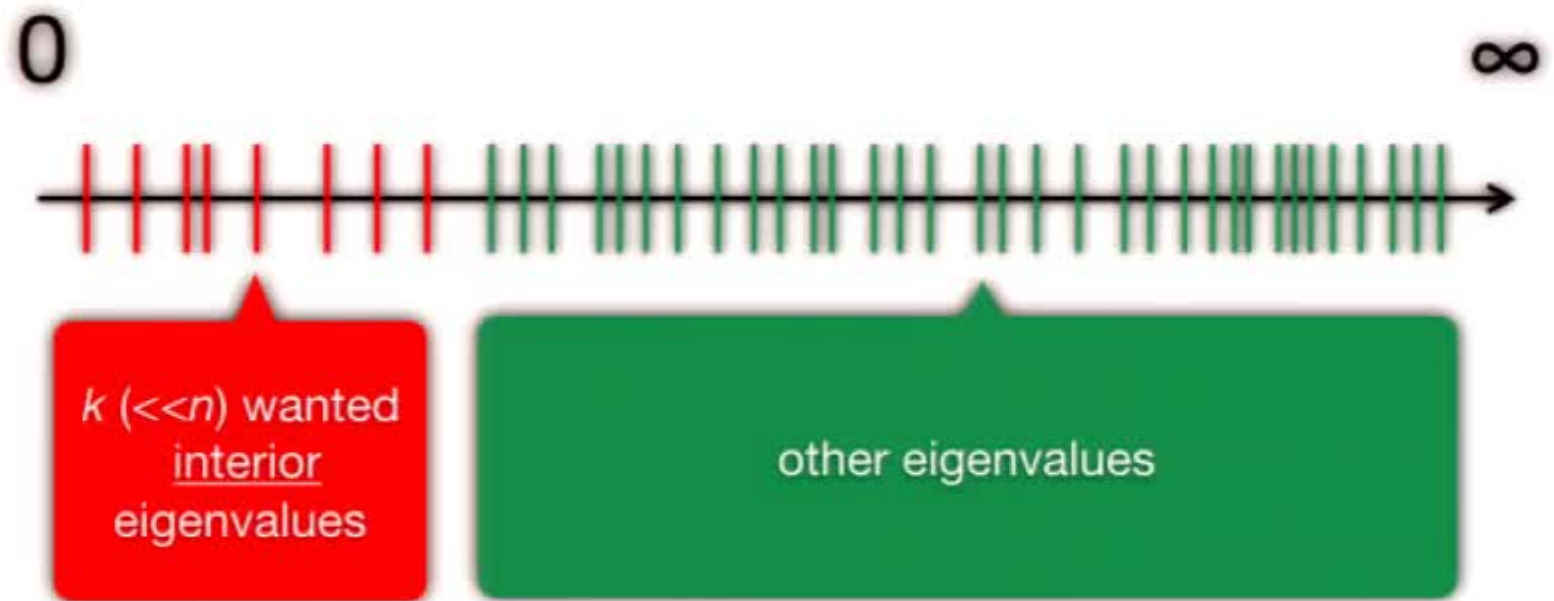
via FFT

diagonal

matrix-vector multp.

# Null-space Free Lanczos Method (NFLM)

- Dim. of GEVP is  $3d$  and SVEP is  $2n$
- GEVP and SEVP has same  $2n$  positive eigenvalues
- SEVP has NO zero eigenvalues





# Null-space Free Lanczos Method (NFLM)



- Hermitian positive definite linear system in NFLM

$$A_r \mathbf{y} = \lambda \mathbf{y}; \quad A_r = \Lambda_r^{\frac{1}{2}} (Q_r^* B_\varepsilon^{-1} Q_r) \Lambda_r^{\frac{1}{2}}.$$

$$(Q_r^* B_\varepsilon^{-1} Q_r) \mathbf{u} = \mathbf{c}$$

- Well-conditioned

$$\kappa(Q_r^* B_\varepsilon^{-1} Q_r) \leq \kappa(B_\varepsilon^{-1})$$

$B_\varepsilon = \text{diag}(\dots, 1, \dots, 13, \dots)$   
electric permittivity of Si=13, Air=1

- CG convergence ratio

$$\gamma = \frac{\sqrt{\kappa(Q_r^* B^{-1} Q_r) - 1}}{\sqrt{\kappa(Q_r^* B^{-1} Q_r) + 1}} \leq \gamma_B = \frac{\sqrt{13} - 1}{\sqrt{13} + 1} \approx 0.5657$$

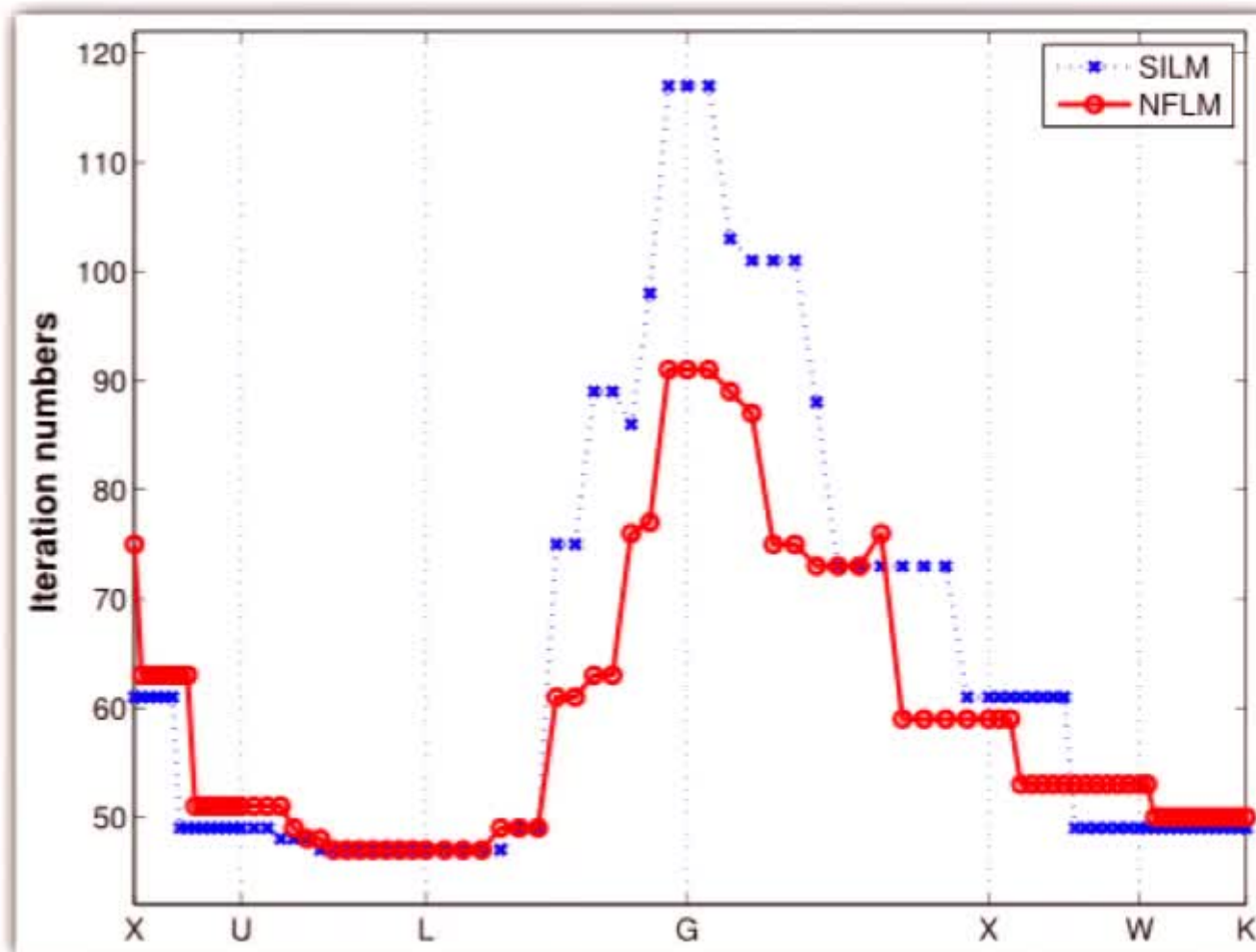
$$(\gamma_B)^{40} \approx 1.27 \times 10^{-10}$$



# Small Iteration Numbers for SILM & NFLM



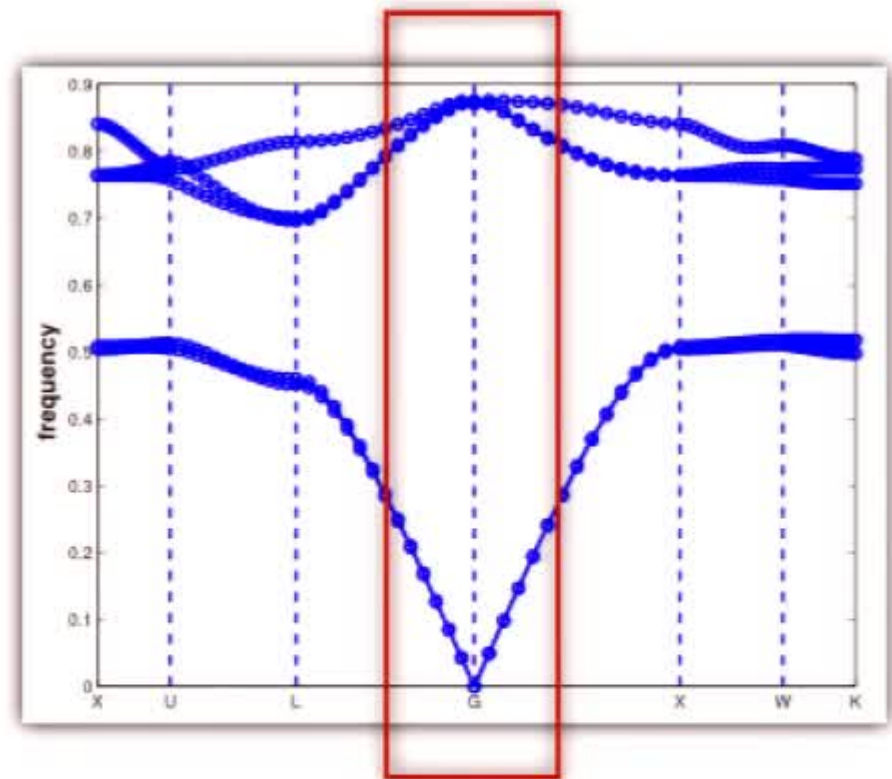
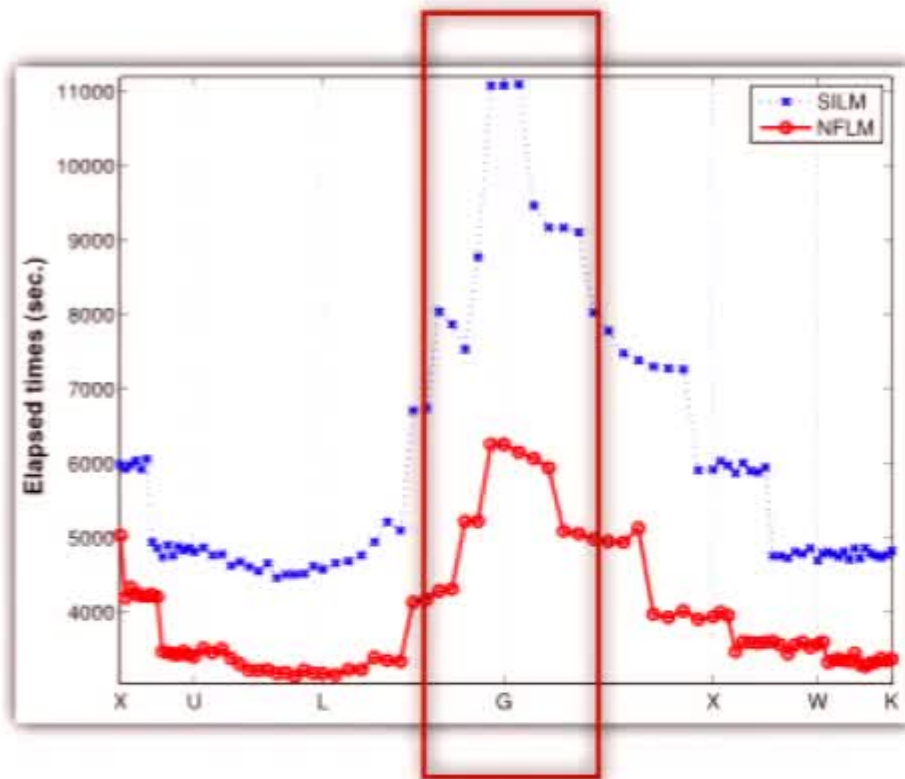
- SILM: 47-115 ite.;  $D= 5,184,000$ ;
- NFLM: 47-91 iter;  $D= 3,456,000$ ;



# Time for SILM and NFLM



- SILM: 47-115 ite.;  $D=5,184,000$ ; 2 MVmult in CGS
- NFLM: 47-91 iter;  $D=3,456,000$ ; 1 MVmult in CG



## Maxwell's Equation for 3D Photonic Crystal

### Eigendecomposition of A

GEVP  $Ax = \lambda Bx$   
 $3n \times 3n$  ( $n$  zero e.v.)

Shift-and-Invert  
Lanczos Method

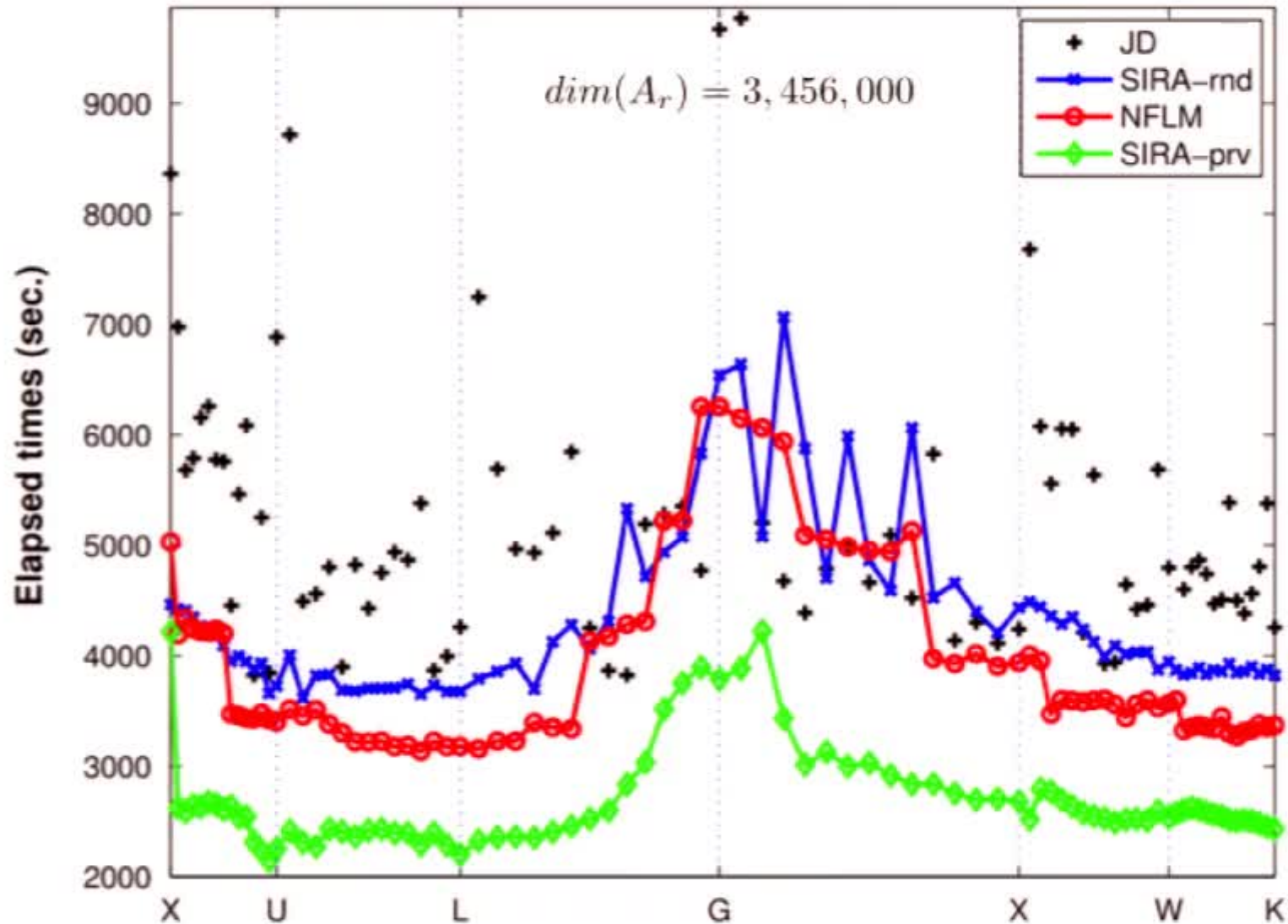
SEVP  $Ax = \lambda x$   
 $2n \times 2n$  (no zero e.v.)

Inv. Lanczos, Jacobi-  
Davidson, SI Res Arnoldi

Iterative Linear System Solver and Preconditioning

FFT-based Matrix Vector Multiplication

# Comparison of Eigensolvers



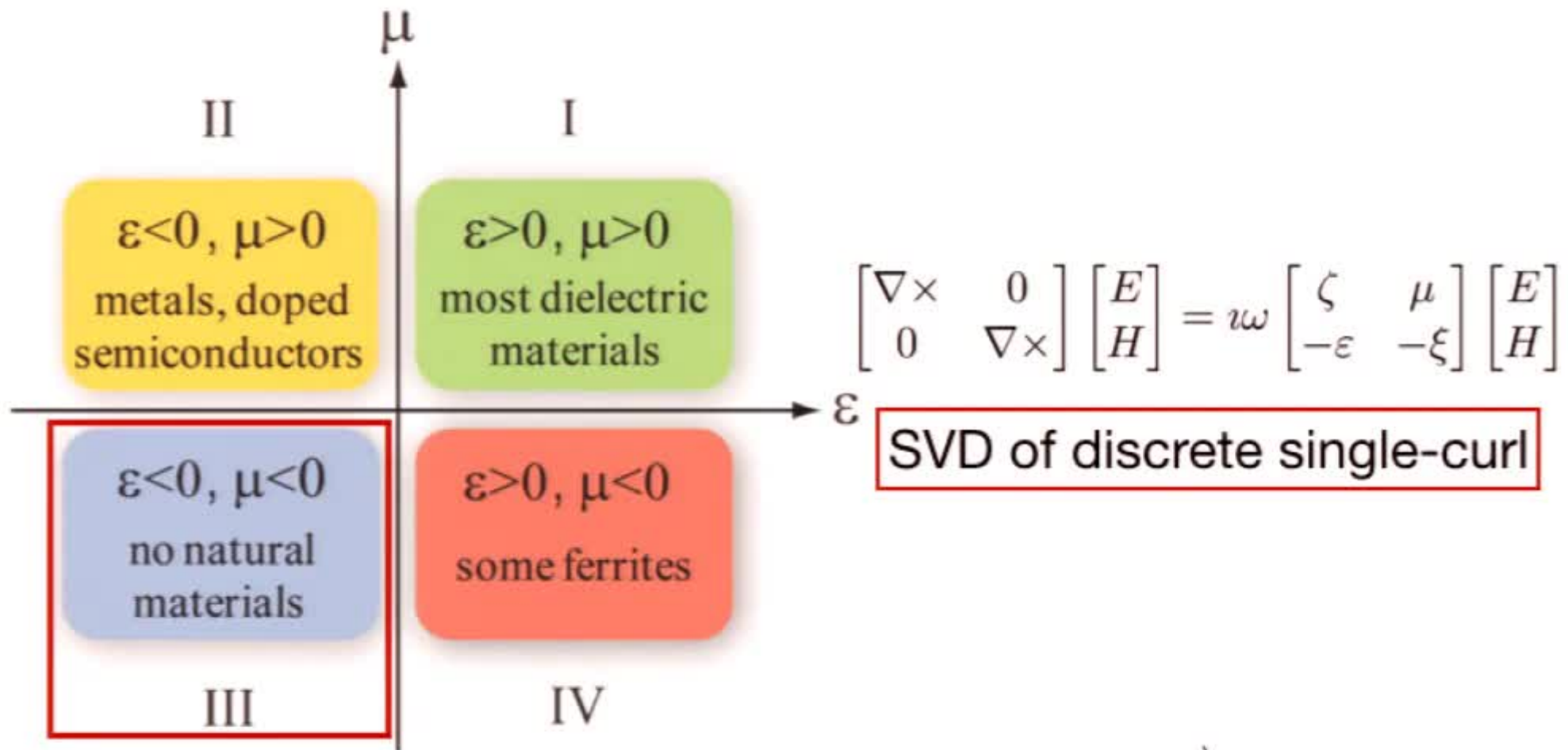
Huang/Hsieh/Lin/W (JCAM 2014)

# Complex Media

SVD of Single Curl



# Artificial Complex Media





# Complex Media



Singular Value Decomposition of Single Curl

$$C = P \operatorname{diag} \left( \Lambda_q^{1/2}, \Lambda_q^{1/2}, 0 \right) Q^*$$

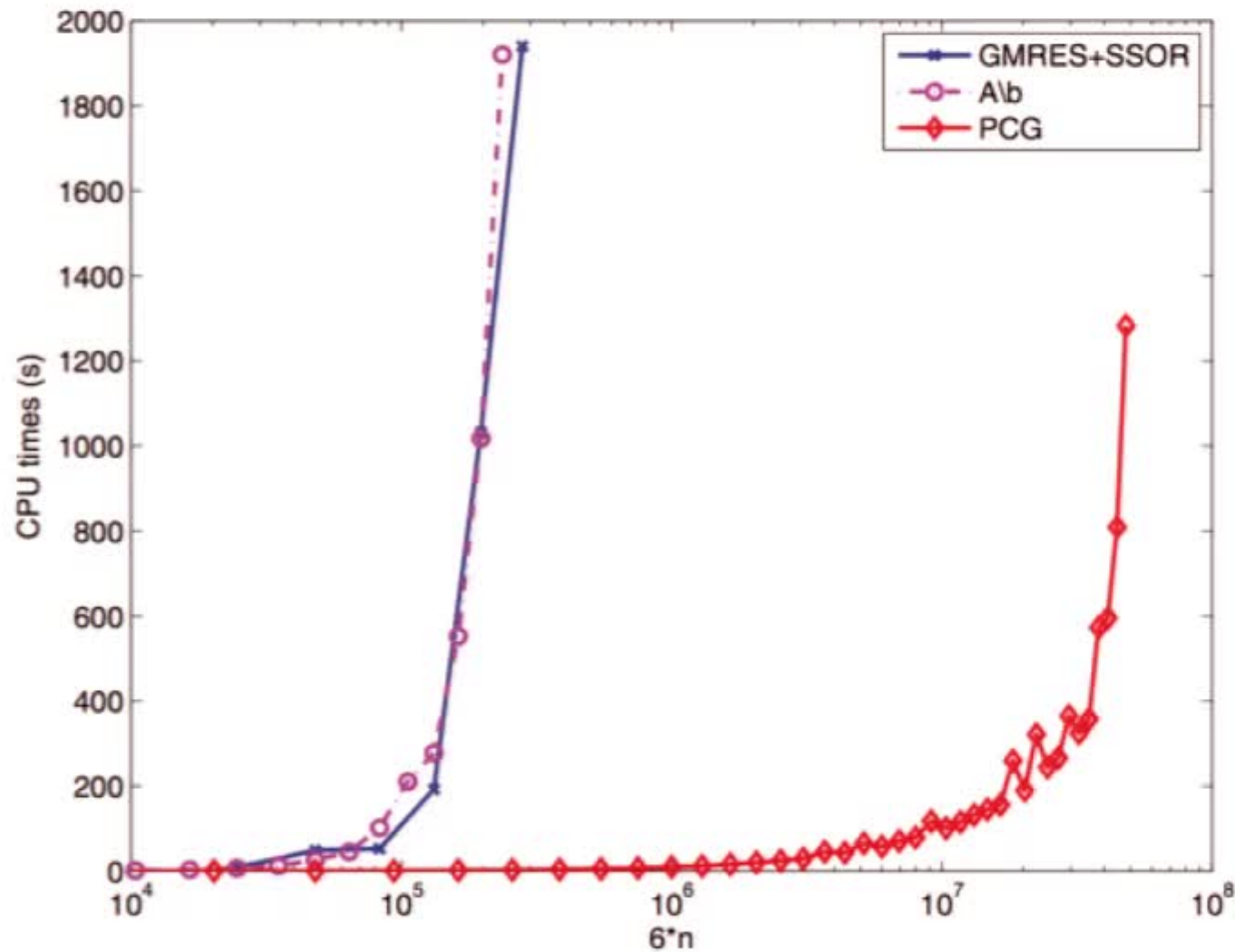
| Eigenvalue problem     | Generalized non-Hermitian  | Generalized Hermitian & HPD |
|------------------------|--|-----------------------------|
| Eigenvalue             | Complex  | Real                        |
| Dimension              | 6n x 6n  | 4n x 4n                     |
| # of zero eigenvalue   | 2n   | 0 (null space free)         |
| Eigensolver            | S.I. Arnoldi   | S.I. Lanczos                |
| Shift                  | Hard to choose   | 0                           |
| Linear system solver   | LU or GMRES (not efficient)  | CG w/ FFT mtx-vec mult      |
| Preconditioner         | Hard to find   | SVD+FFT (well-cond.)        |
| Applications           | Complex media  | Chiral/Pseudochiral         |
| Embedded linear system | $\left( \begin{bmatrix} C & 0 \\ 0 & C^* \end{bmatrix} - \sigma \begin{bmatrix} \zeta_d & I_{3n} \\ -\varepsilon_d & -\xi_d \end{bmatrix} \right) y = b$ | See below                   |

$$A_r u \equiv \begin{bmatrix} P_r^* & \\ & Q_r^* \end{bmatrix} \begin{bmatrix} \zeta_d & -I_{3n} \\ I_{3n} & 0 \end{bmatrix} \begin{bmatrix} \Phi^{-1} & 0 \\ 0 & I_{3n} \end{bmatrix} \begin{bmatrix} \zeta_d^* & I_{3n} \\ -I_{3n} & 0 \end{bmatrix} \begin{bmatrix} P_r & \\ & Q_r \end{bmatrix} u = b$$

# Time for Embedded Linear Systems



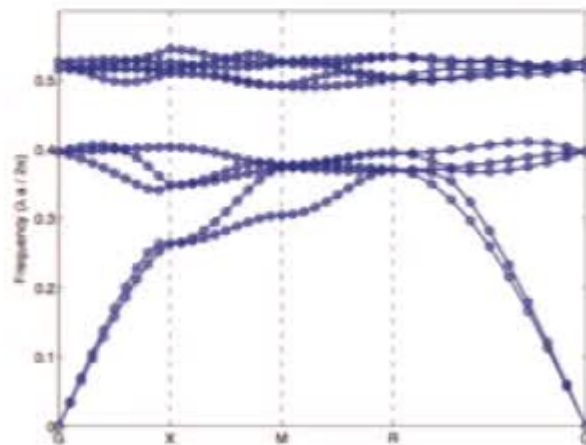
Quad-Core Xeon X5687 3.6GHz CPUs, 48GB, MATLAB 2014a



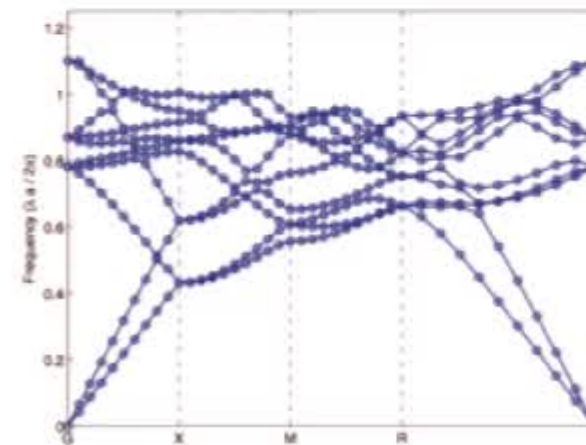
# Iteration Number



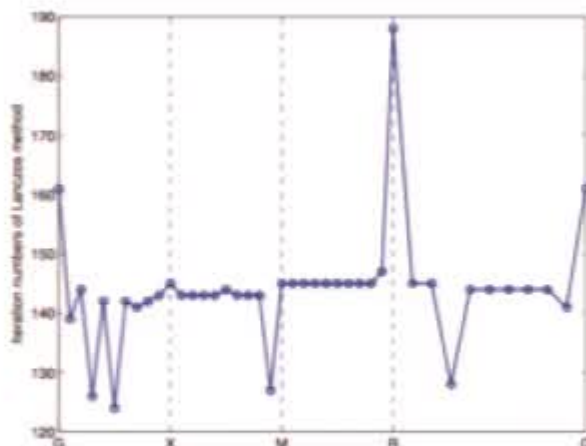
Dim:  $4 \times 128 \times 128 \times 128 = 8,388,608$   
Quad-Core Xeon X5687 3.6GHz CPUs, 48GB, MATLAB 2014a



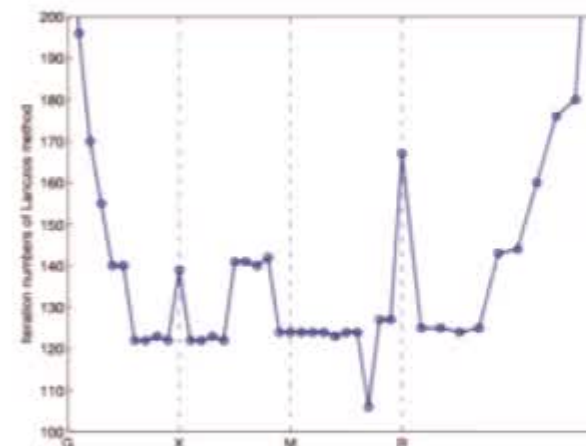
(a) Band structure for  $(\epsilon_i, \epsilon_o, \gamma) = (13, 1, 0.5)$



(b) Band structure for  $(\epsilon_i, \epsilon_o, \gamma) = (1, 1, 0.8)$



(c) Iteration numbers ranging from 120 to 190 with average 143 for  $(\epsilon_i, \epsilon_o, \gamma) = (13, 1, 0.5)$



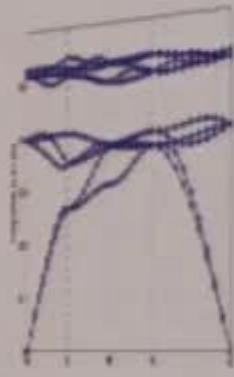
(d) Iteration numbers ranging from 100 to 200 with average 136 for  $(\epsilon_i, \epsilon_o, \gamma) = (1, 1, 0.8)$

multiplicity of eigenvalues

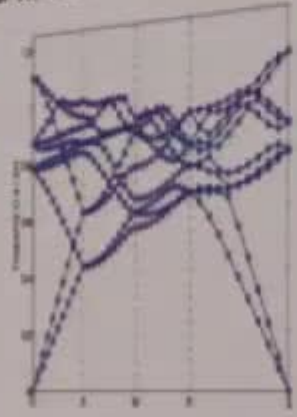
# Iteration Number



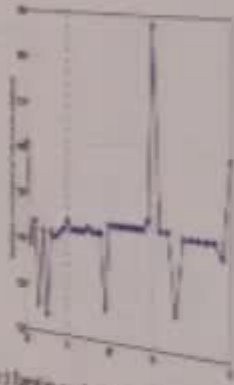
Dim:  $4 \times 128 \times 128 \times 128 = 8,388,608$   
Quad-Core Xeon X5567 3.86GHz CPUs, 48GB, MATLAB 2014a



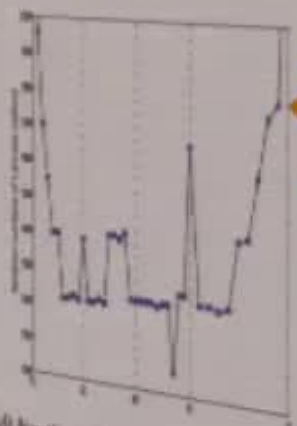
(a) Eigenvalues for  $(\alpha, \beta, \gamma) = (1, 1, 0.5)$



(b) Band structure for  $(\alpha, \beta, \gamma) = (1, 1, 0.8)$



(c) Iteration numbers ranging from 120 to 180 with average 140 for  $(\alpha, \beta, \gamma) = (1, 1, 0.5)$



(d) Iteration numbers ranging from 100 to 200 with average 136 for  $(\alpha, \beta, \gamma) = (1, 1, 0.8)$

multiplicity of eigenvalues

# Software



# FAME.MATLAB




- Various applications  
(e.g. photonic crystal, metallic materials, complex material)
- GUI for photonic structure design
- Choices of eigensolvers and linear system solvers
- Built-in functions for efficient algorithms
- Visualization of results
- Extendable to new algorithms and applications



# FAME.MATLAB



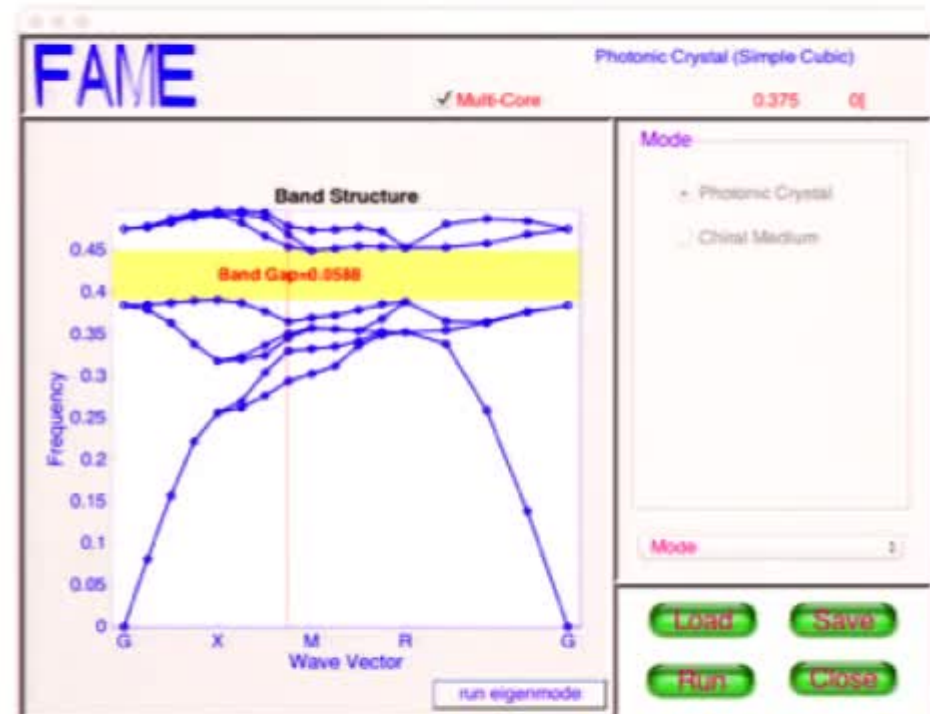
**FAME** Photonic Crystal (Simple Cubic)  Multi-Core

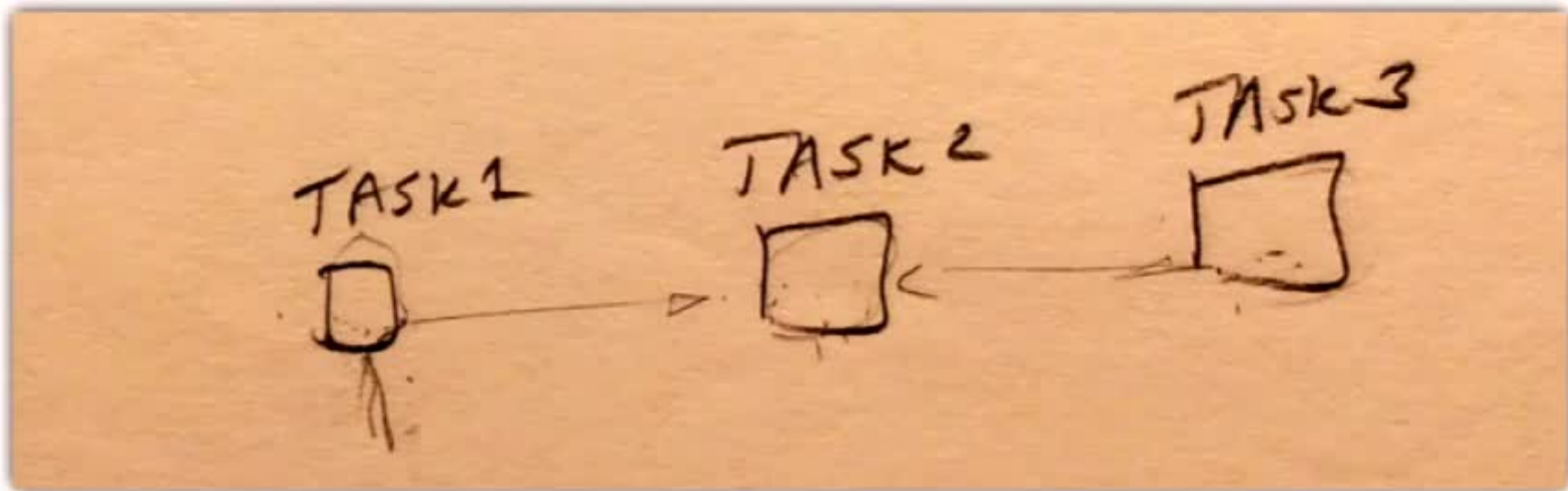
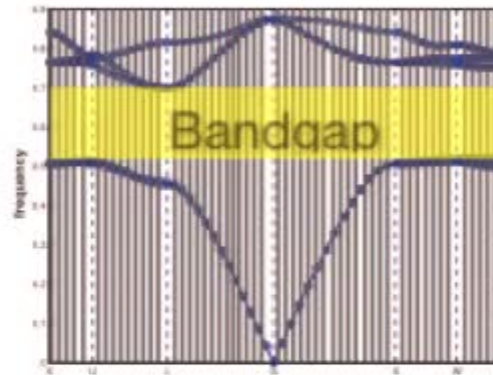


**Domain**  
grid numbers  
x: 16 y: 16 z: 16  
domain length  
x: 1 y: 1 z: 1  
shape  
sphere radius: 0.38  
cylinder radius: 0.1  
 Show 3D structure  
Sample cube

Domain

Load Save  
Run Close





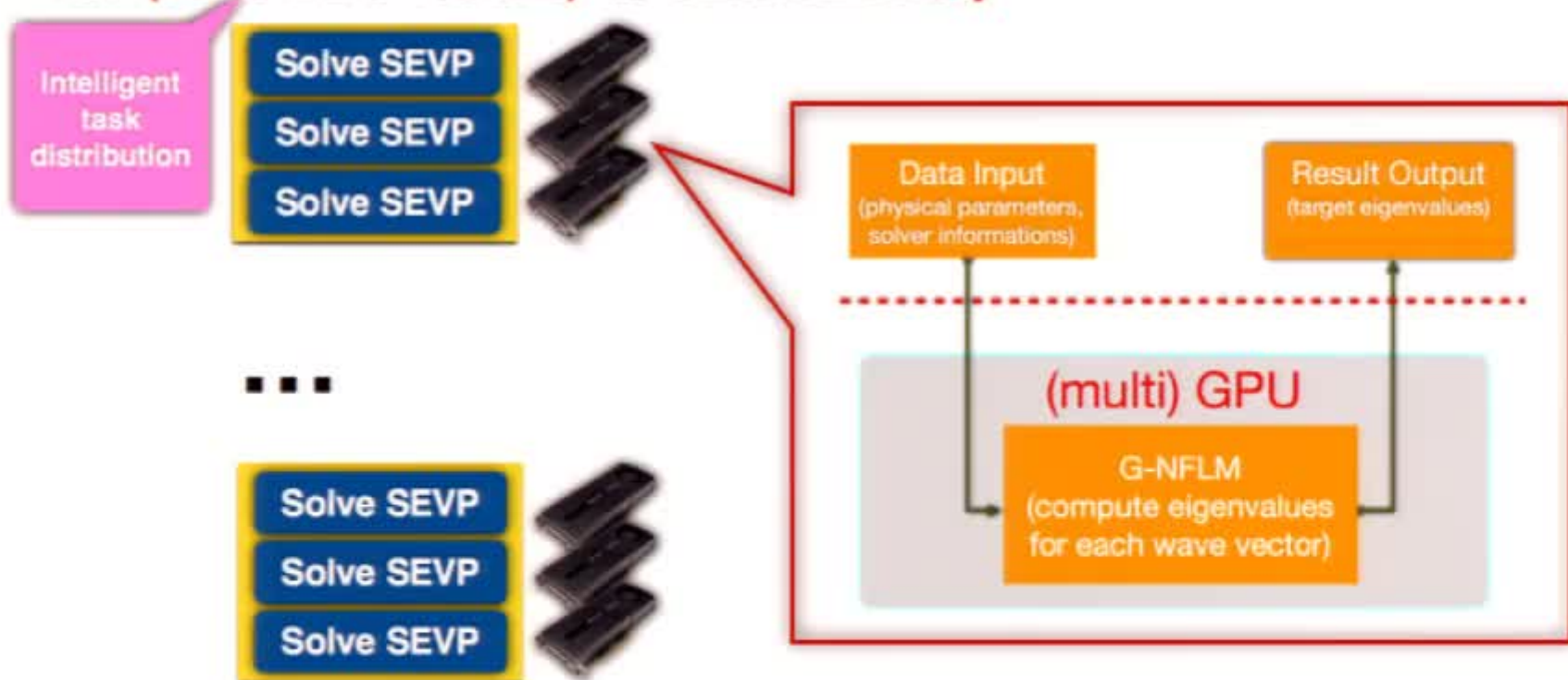
M P I

# FAME.GPU



- Cluster with 16 MPI nodes (each has 3 NVIDIA Tesla M2070)
- Minimal communications between CPU and GPU

**for (some wave vectors) do simultaneously**

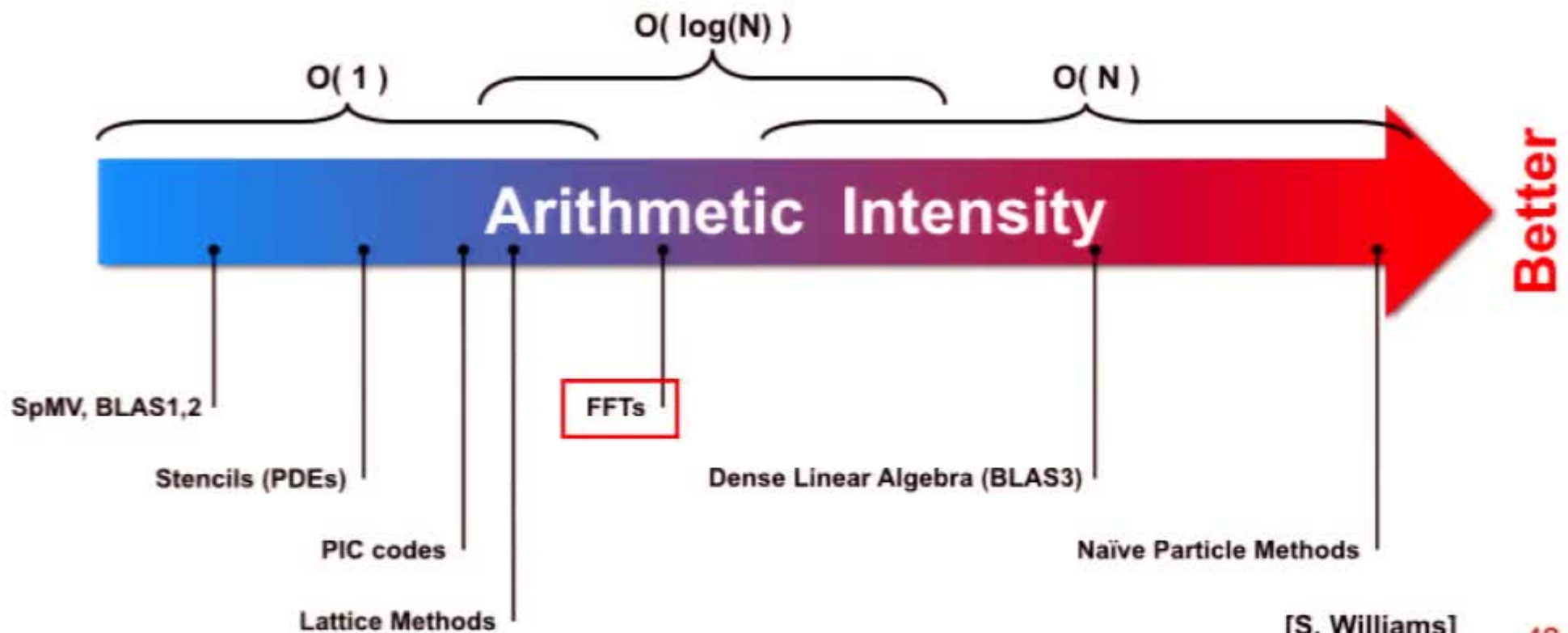


**end for**

# FAME.GPU

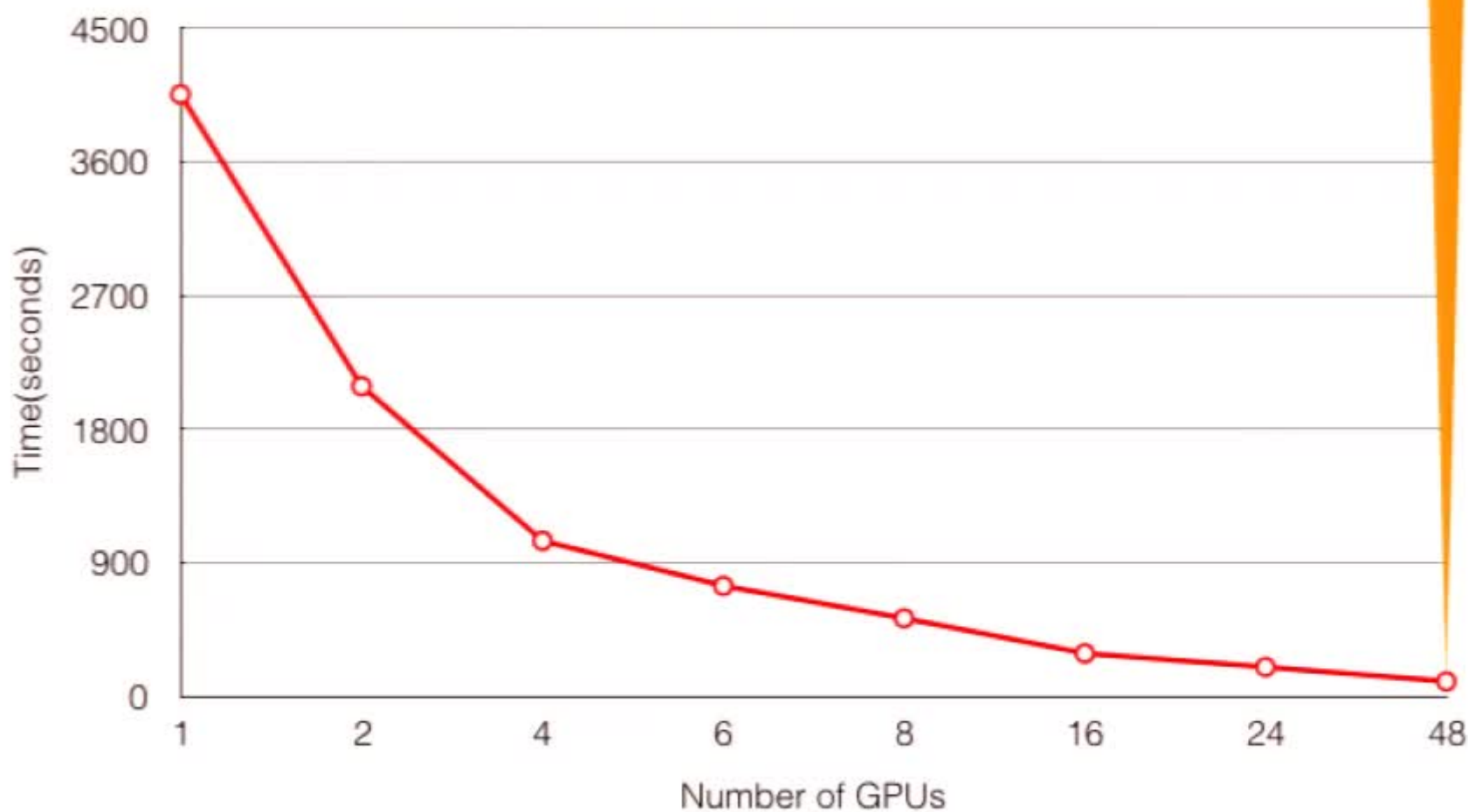


- Algorithm:  
Inverse Lanczos (w/ restart) + Nullspace free + CG (well-cond.)
- Parallelism:  
Embarrassing parallel in terms of wave vectors w/ intelligent task distribution



# Scalability on Multiple GPU (6.29M)

~1.5 minute for a complete band structure diagram

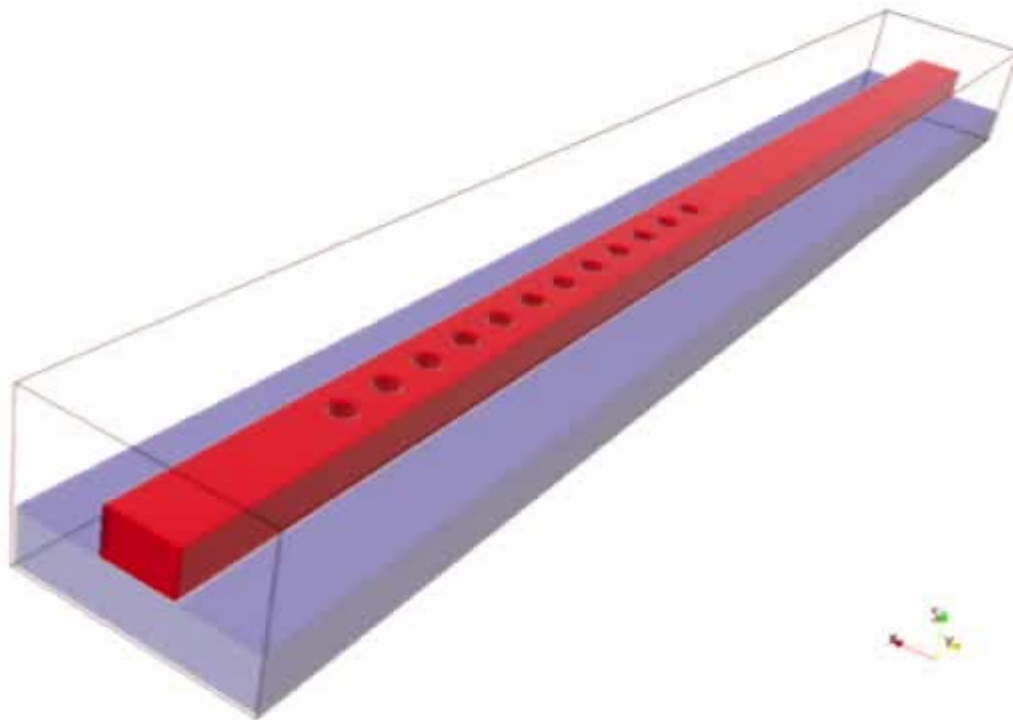




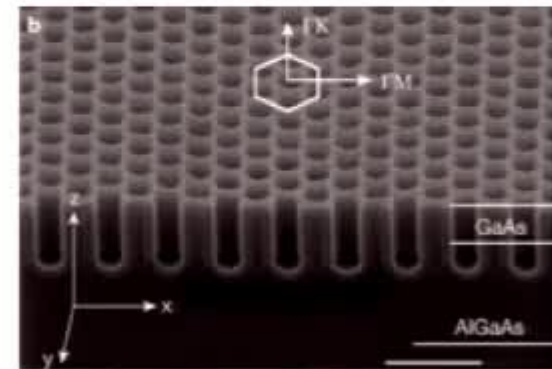
# Compressed Hierarchical Schur

Memory Saving BLAS3 Direct Solver

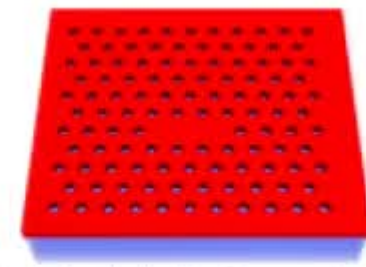
# Photonic Devices



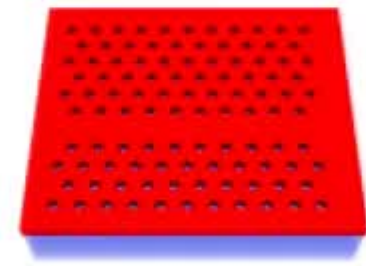
Periodic airhole array on a CdTe ridge waveguide



Ref: Chow et al., "Three-dimensional control of light in a two-dimensional photonic crystal slab," Nature, vol. 407, 983-986 (26 October 2000)



Point defect photonic crystal slab

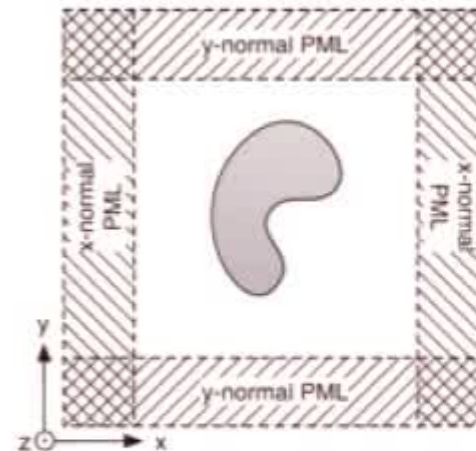
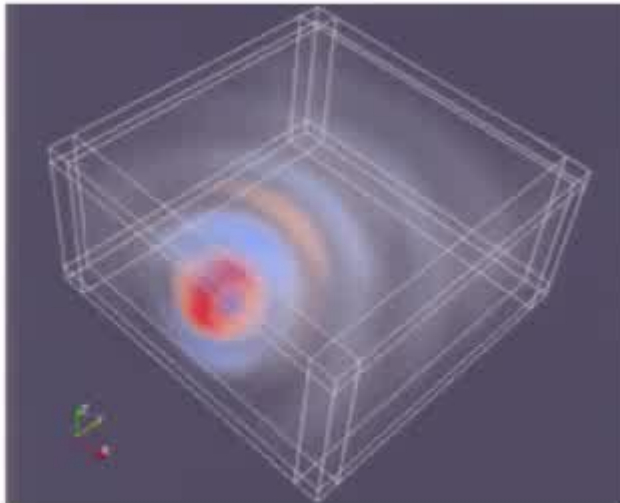


Linear defect photonic crystal slab

# Perfectly Matched Layer (PML)



- $\nabla_s = \hat{x} \frac{1}{s_x} \frac{\partial}{\partial x} + \hat{y} \frac{1}{s_y} \frac{\partial}{\partial y} + \hat{z} \frac{1}{s_z} \frac{\partial}{\partial z}$  on complex stretching coordinate
- $w$ -normal scale stretching factors:  $s_w(l) = 1 + \frac{j\sigma_w(l)}{\omega\epsilon_0}$   
( $w = x, y, z$  and  $l$ : depth from PML interface)
- Artificial absorptive layer that absorb outward wave and reduce reflection from boundary



Ref: Shin and Fan, JCP, vol. 231, pp. 3406, 2012.

# Linear System Problems



- Nonmagnetic vector wave equation ( $\mu_r = 1$ )

- $$-\nabla_s \times \nabla_s \times \vec{E} + k_0^2 \epsilon_r \vec{E} = j\omega \mu_0 \vec{J}$$

- Linear problem:

For  $k_0 = \omega \sqrt{\mu_0 \epsilon_0} = \frac{2\pi}{\lambda_0}$ , given  $\vec{J}$ ,  $\lambda_0$ ,  $\epsilon_r$ , solve  $\vec{E}$

- Photonic band analysis equation without source

- $$\frac{1}{\epsilon_r} \nabla_s \times \nabla_s \times \vec{E} = k_0^2 \vec{E}$$

- Eigenvalue problem:

Given  $\epsilon_r, \vec{k}$ , solve  $k_0^2$  and band field pattern  $\vec{E}$

- Shift-and-invert method with given shift  $k_s^2$

- Solve 
$$-\frac{1}{\epsilon_r} (-\nabla_s \times \nabla_s \times \vec{\psi}_{i+1} + k_s^2 \epsilon_r \vec{\psi}_{i+1}) = \vec{\psi}_i$$

- Non-Hermitian complex ill-conditioned linear systems

$\vec{E}$ : electric field

$\vec{H}$ : magnetic field

$\omega$ : angular frequency

$\mu$ : permeability

$\mu_0$ : vacuum permeability

$\mu_r$ : relative permeability

$\epsilon$ : permittivity

$\epsilon_0$ : vacuum permittivity

$\epsilon_r$ : relative permittivity

$k_0$ : vacuum wavenumber

$\lambda_0$ : vacuum wavelength



# Challenges

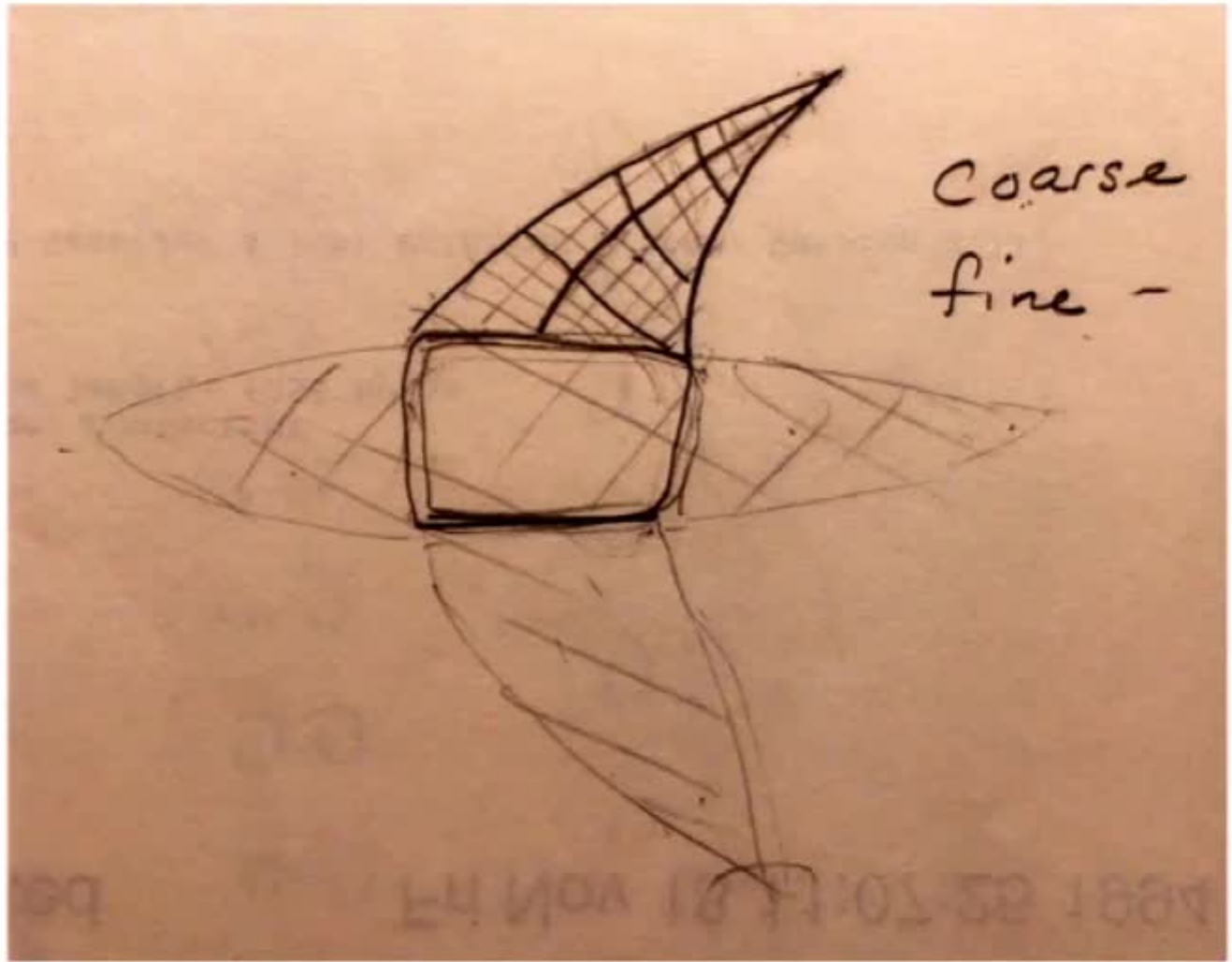
---



- Non-Hermitian ill-conditioned linear systems
- Without eigendecomposition of the coefficient matrices
- Hard to find efficient preconditioner



DIRECT



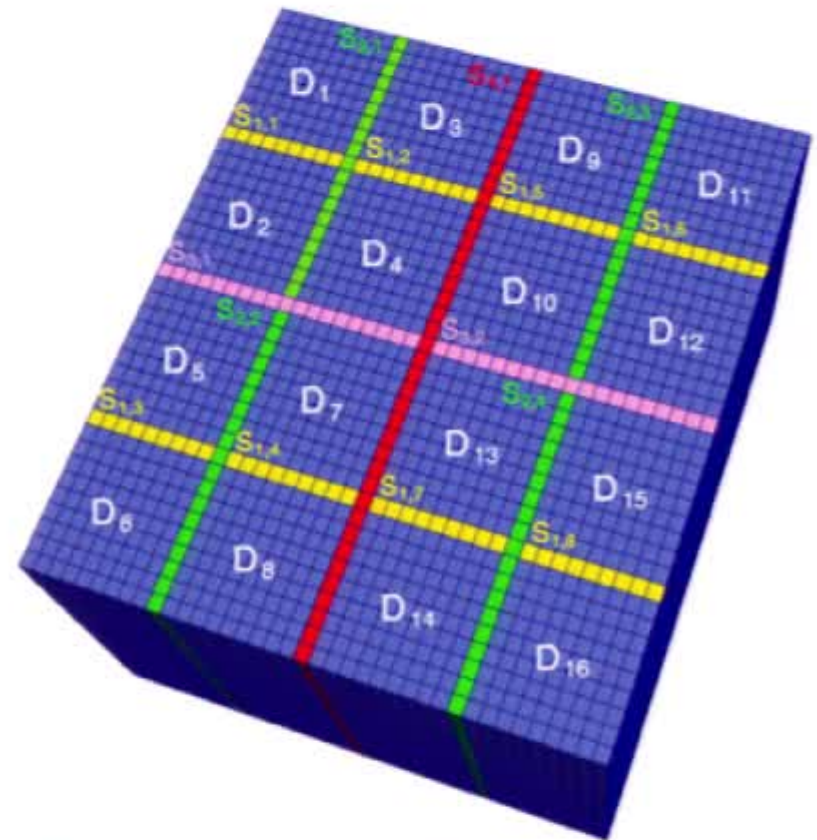
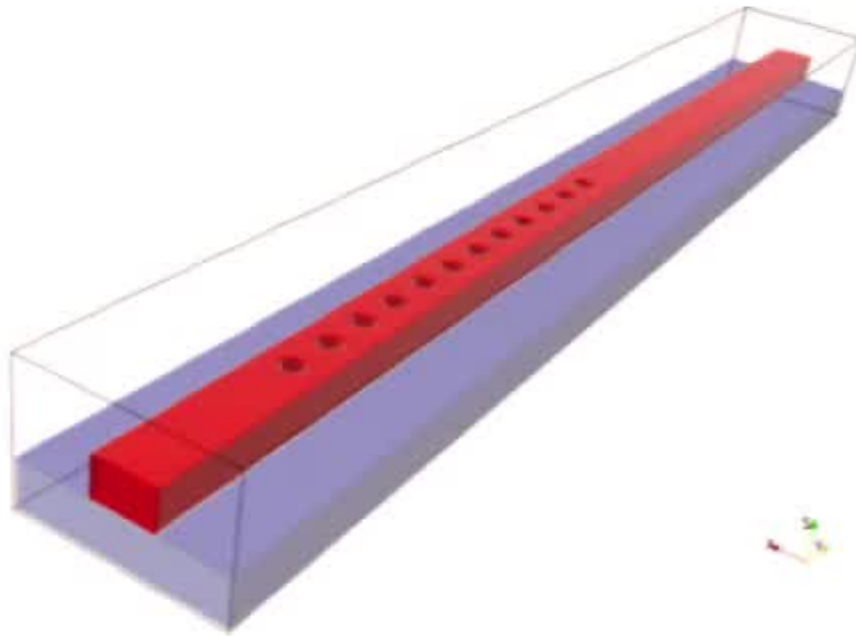
# Approaches

---



- Hierarchical Schur direct solver
- Memory saving
- Multilevel parallelism
- BLAS3 operations

# Physical and Computational Domains

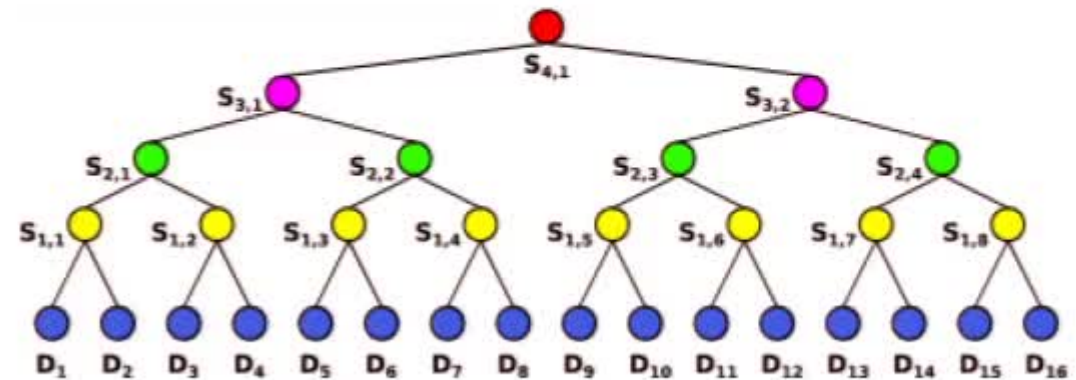
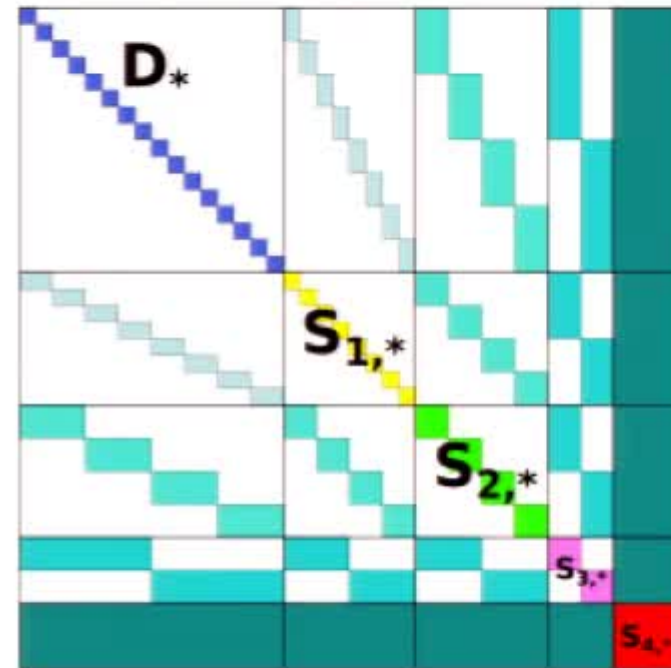
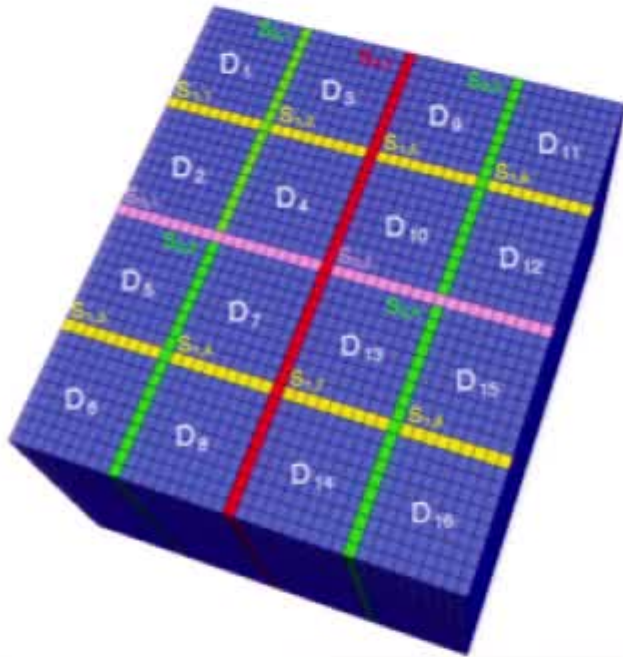


Subdomains, interfaces, separators  
in domain decomposition

# Elimination Tree and Matrix



- Customized nested dissection ordering





# Subdomain Level



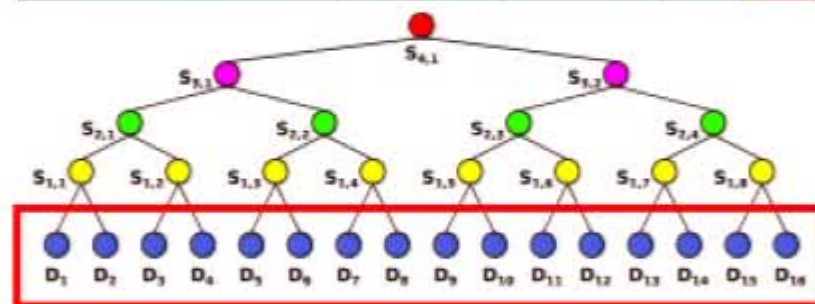
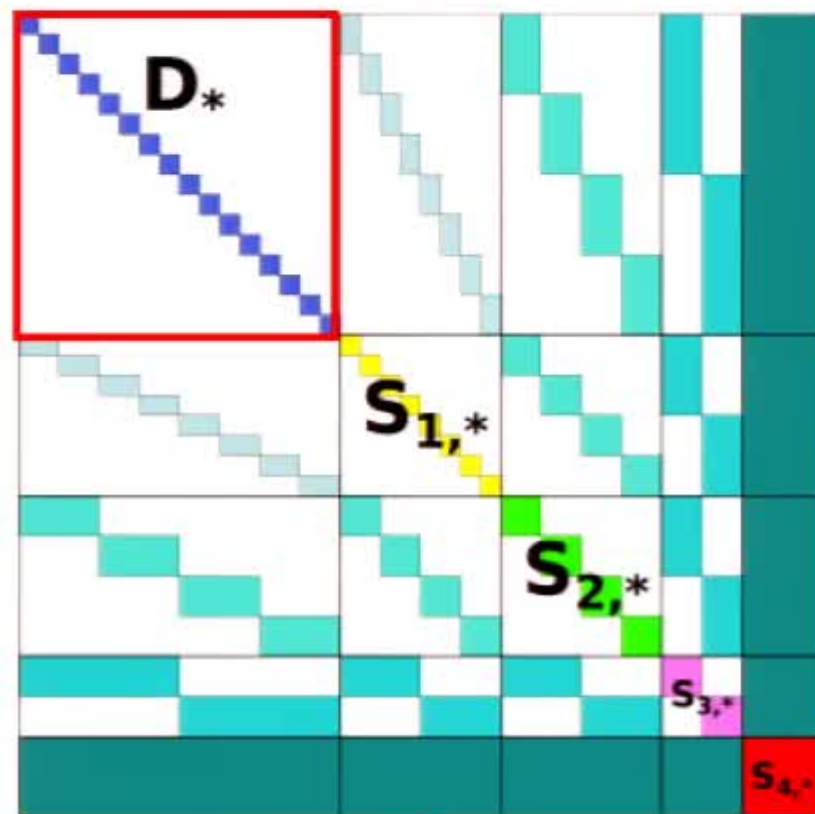
- Operations

- Factor sparse diagonal blocks
- Scale sparse interface blocks
- Update Schur submatrices

$$S_{1,1} = A_{17,17} - A_{17,1}A_{1,1}^{-1}A_{1,17} - A_{18,2}A_{2,2}^{-1}A_{2,18}$$

- Parallelisms

- One node for each  $D_i$
- Multi-threads for each  $D_i$





# Separator Levels

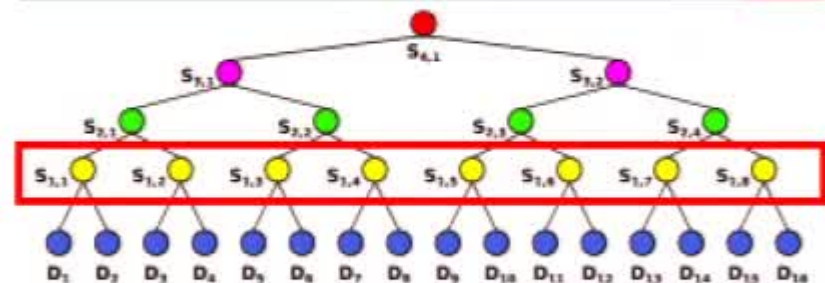
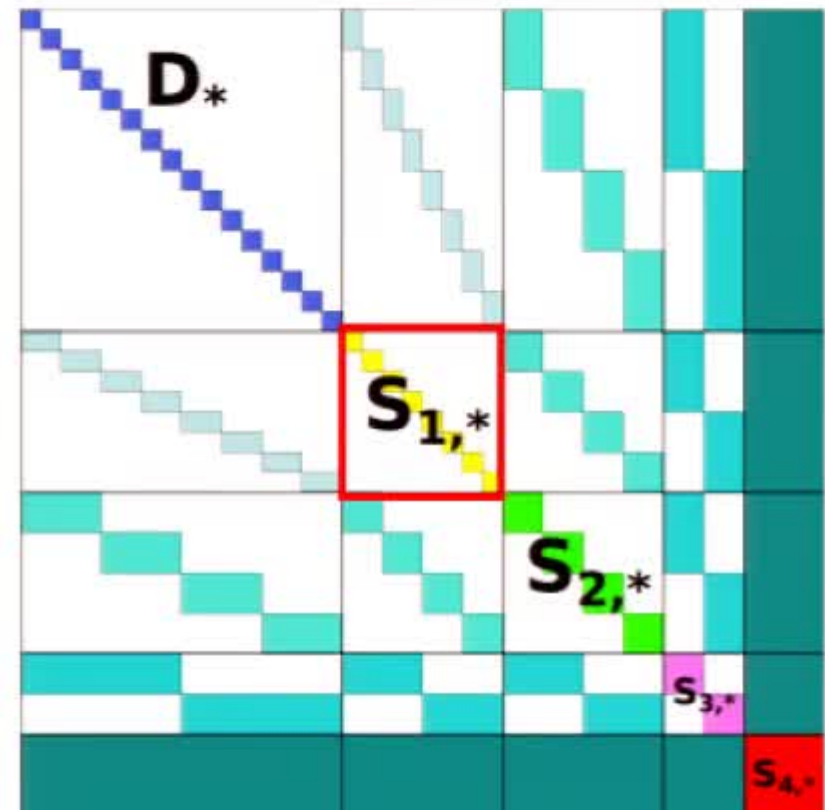


## Operations

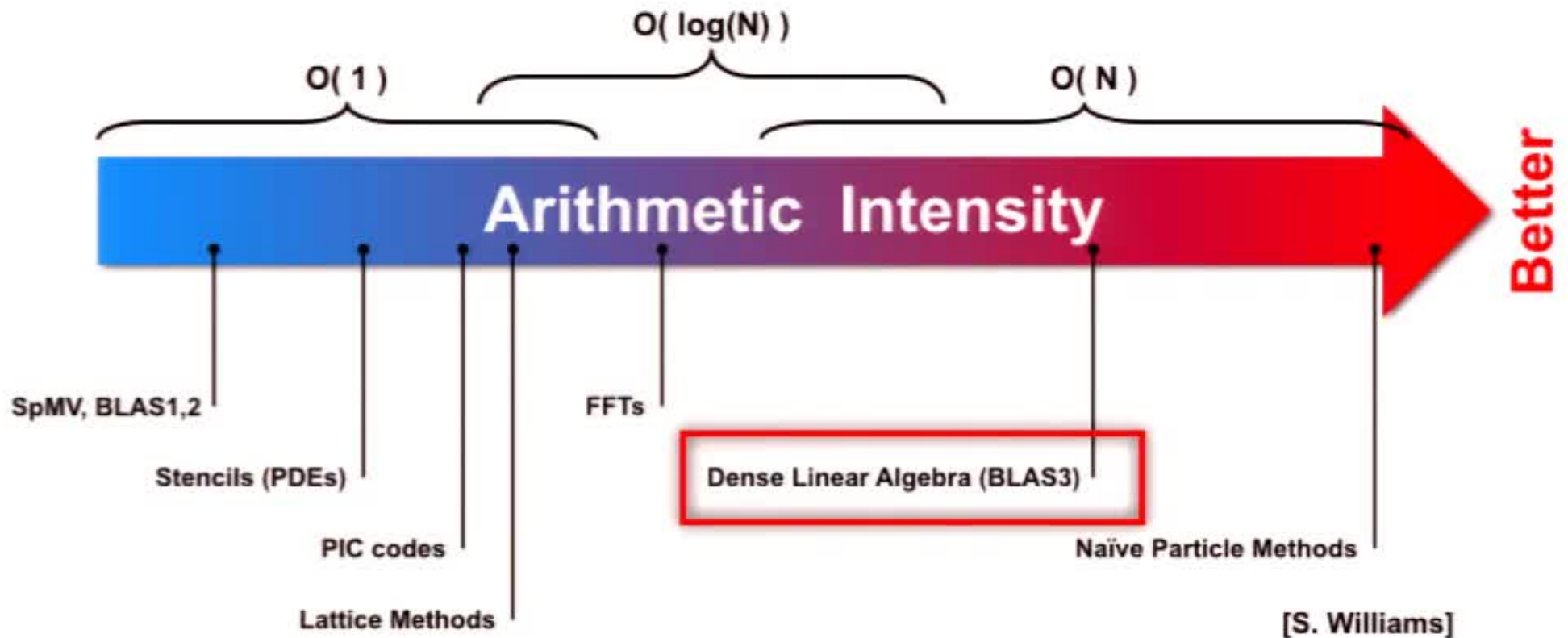
- Factor dense diagonal blocks (ZGETRF)
- Scale dense interface blocks (ZGETRS)
- Update dense Schur submatrices (ZGEMM)

## Parallelism

- Concurrent dense BLAS3 operations
- Fine-grained parallelism via GPU/Xeon Phi
- "Right sizes" dense matrices



# Arithmetic Intensity

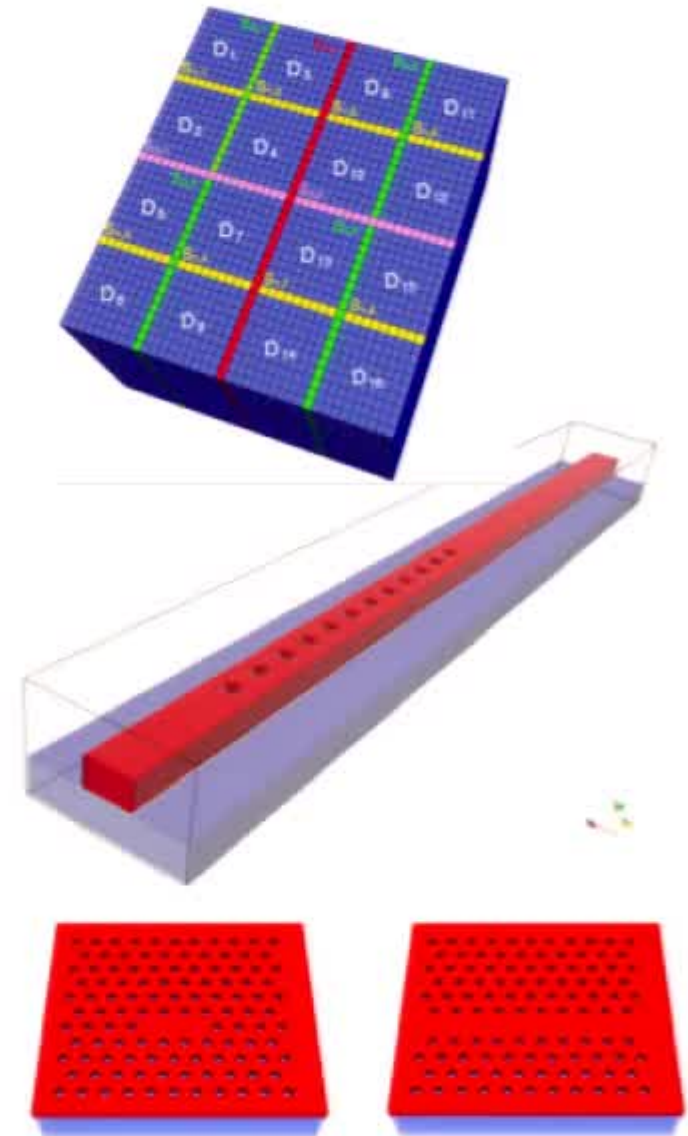


- Arithmetic Intensity  $\sim$  (Total Flops) / (Total DRAM Bytes)
- Example: dense matrix-matrix multiplication:  $(N^3 \text{ flops}) / (N^2 \text{ memory})$
- Higher arithmetic intensity  $\sim$  better locality  $\sim$  higher chance to achieve machine peak

# Memory/Computation Saving



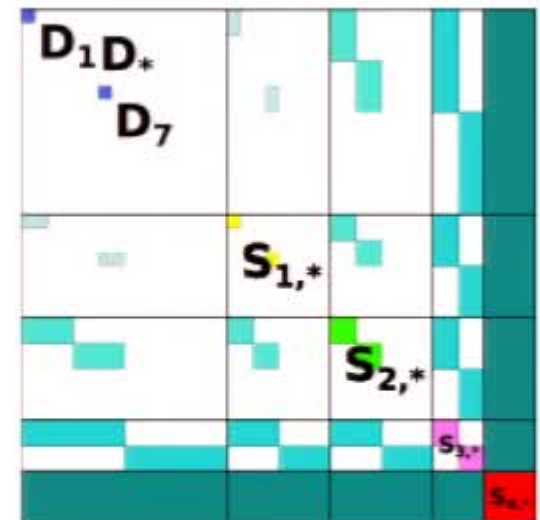
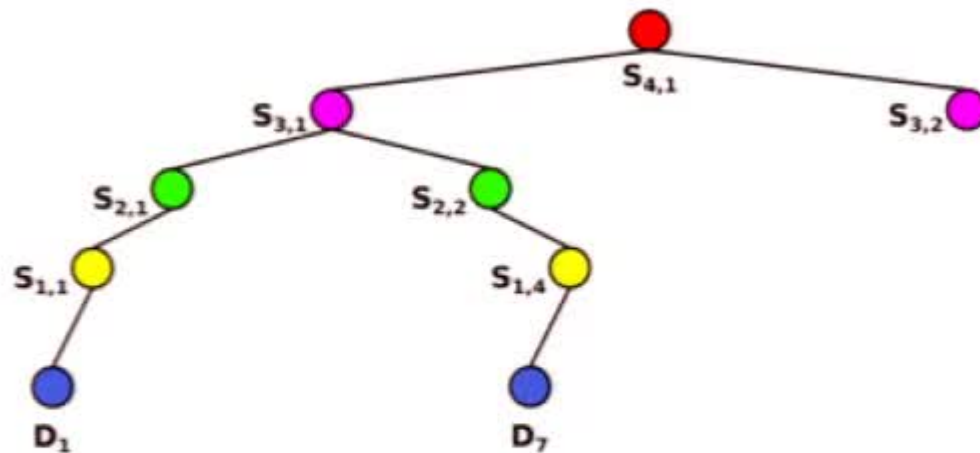
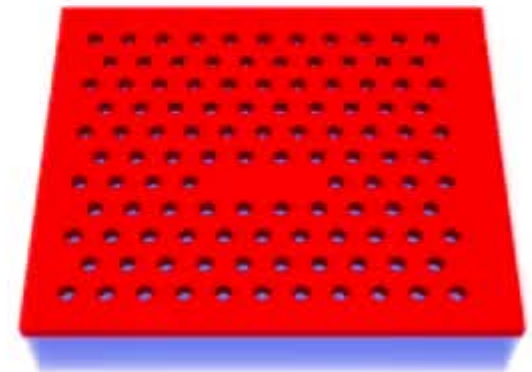
- Criteria of redundancy
  - Identical dimensions
  - Identical physical properties (e.g. periodic structures or homogeneous material)
  - Identical discretization (e.g. No PML and uniform grids)
  - Identical grid index ordering



# Memory/Computation Savings

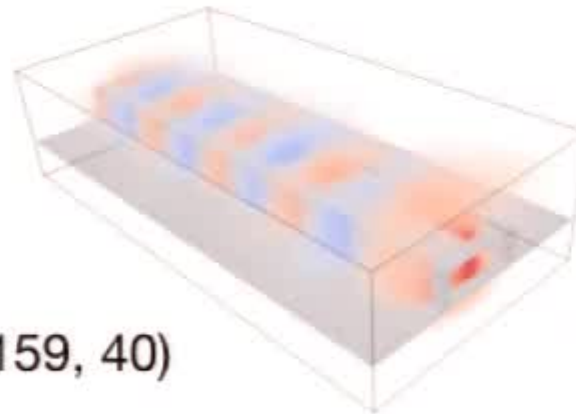


- Identical redundant matrix blocks removal
- For defected parts
  - $D_7$  is different from other subdomains
  - No redundancy to its parent separators





# Memory for Dielectric Waveguide



(79, 159, 40)

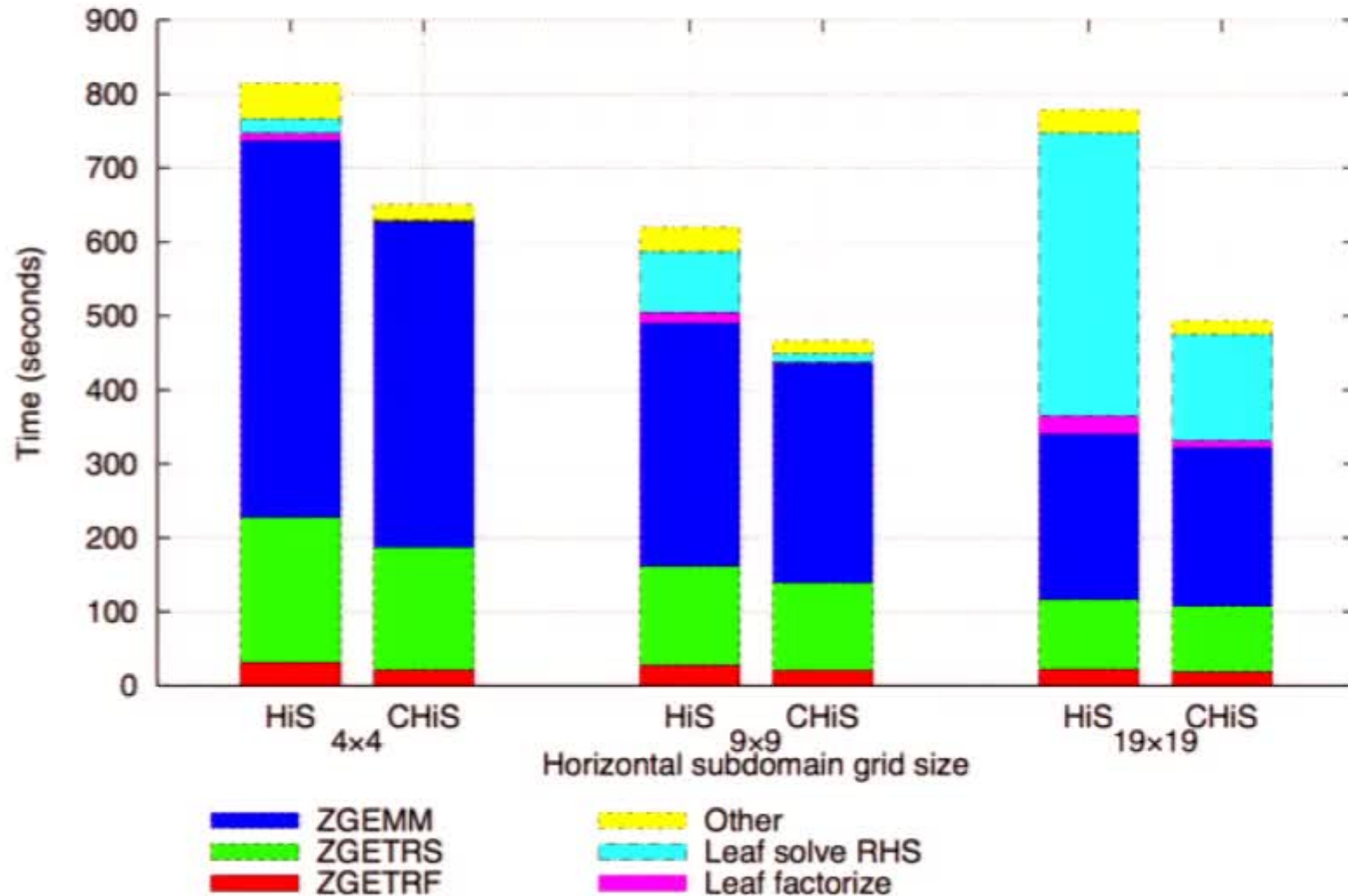
|                       | Peak<br>Memory | Number of non-redundant elements |           |           |           |           |           |           |           |           |           |  |
|-----------------------|----------------|----------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|--|
|                       |                | $D_*$                            | $S_{1,*}$ | $S_{2,*}$ | $S_{3,*}$ | $S_{4,*}$ | $S_{5,*}$ | $S_{6,*}$ | $S_{7,*}$ | $S_{8,*}$ | $S_{9,*}$ |  |
| HiS <sub>4×4</sub>    | 115 GB         | 512                              | 256       | 128       | 64        | 32        | 16        | 8         | 4         | 2         | 1         |  |
| CHiS <sub>4×4</sub>   | 75 GB          | 40                               | 30        | 18        | 12        | 12        | 6         | 6         | 3         | 2         | 1         |  |
| HiS <sub>9×9</sub>    | 89 GB          | 128                              | 64        | 32        | 16        | 8         | 4         | 2         | 1         | 0         | 0         |  |
| CHiS <sub>9×9</sub>   | 64 GB          | 18                               | 12        | 12        | 6         | 6         | 3         | 2         | 1         | 0         | 0         |  |
| HiS <sub>19×19</sub>  | 68 GB          | 32                               | 16        | 8         | 4         | 2         | 1         | 0         | 0         | 0         | 0         |  |
| CHiS <sub>19×19</sub> | 52 GB          | 12                               | 6         | 6         | 3         | 2         | 1         | 0         | 0         | 0         | 0         |  |
| UMFPACK               | 245 GB         |                                  |           |           |           |           |           |           |           |           |           |  |



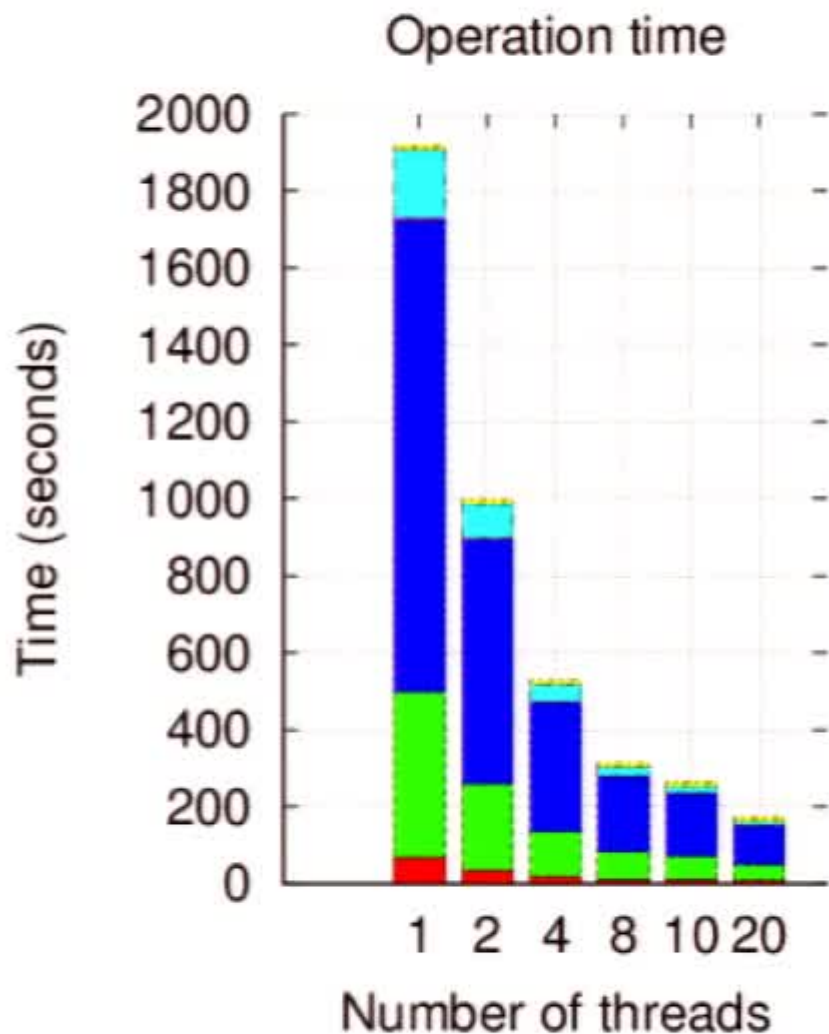
# Timing Breakdown Analysis



Dual Intel E5-2650 v3 CPUs + 256 GB memory



# Factorization Time for PC Slab



- Matrix dimension:  $95 \times 289 \times 20 = 1,362,300$
- Intel E5-2650 v3 (10 cores) x 2
- 256 GB DDR4-2133 main memory
- Intel Parallel Studio XE 2015 Cluster edition with ICC 15.0
- Flags: -O3 -openmp
- OS: CentOS 6.5
- Libraries: MKL 11.2, PARDISO (included in Intel MKL)

# ZGEMM on CPU/GPU for PC Slab



- Projected GPU acceleration performance of dense linear algebra
  - Acceleration on ZGETRF, ZGETRS, and ZGEMM
  - Dual E5-2650 v3  $\approx$  490 GFLOPS
  - Dual NVIDIA K40 only: 1000 ~ 2400 GFLOPS
  - Expect more than 3x speedup with tuning and hybrid computing

