Solving large scale eigenvalue problems in electronic structure calculations

C. Bekas

Comp. Science & Engineering Dept.

University of Minnesota, Twin Cities

*** Work supported by NSF under grants NSF/ITR-0082094, NSF/ACI-0305120 and by the Minnesota Supercomputing Institute**

Introduction and Motivation

Computational Materials Science Target Problem

…predict the properties of materials…

How?

- *AB INITIO calculations:* Simulate the behavior of materials at the atomic level, by applying the basic laws of physics: Quantum Mechanics

What do we (hope to) achieve?

- Explain the *experimentally established* properties of materials
- Engineer new materials with *desired properties*

Applications:…numerous (some include)

- Semiconductors, synthetic light weight materials
- Drug discovery, protein structure prediction
- Energy: alternative fuels

In this talk

- ➢ Introduction to the mathematical formulation of *ab initio* calculations…
- ➢ in particular…the Density Functional Theory (DFT)…formulation
- ➢ Identify the computationally intensive "spots"… *eigenvalue calculations*

Large scale eigenvalue problems are central…

- ➢ Symmetric/Hermitian problems
- ➢ very large number of eigenvalues/vectors required…so
- ➢ reorthogonalization (Gram-Schmidt) and synchronization (barrier/join) costs dominate…
- ➢ limiting the feasible size of molecules under study

Alternative…Automated Multilevel Substructuring (AMLS)

- ➢ Significantly limits reorthogonalization costs…
- ➢ can attack very large problems…when O(1000) eigs. are required…

Mathematical Modelling: The Wave Function

We seek to find the *steady state* of the *electron distribution*

- Each electron e_i is described by a corresponding *wave function* ψ_i …
- ψ_i is a function of space (r)…in particular it is determined by
  - The position r_k of all particles (including nuclei and electrons)
  - It is normalized in such a way that

$$\int |\psi_i(r)|^2 dr = 1$$

  - Max Born's probabilistic interpretation: Considering a region *D*, then

$$Pr(\psi_i, D) = \int_D |\psi_i(r)|^2 dr$$

…describes the probability of electron e_i being in region *D.* Thus: the distribution of electrons e_i in space is defined by the wave function ψ_i

Mathematical Modelling: The Hamiltonian

Steady state of the *electron distribution:*

- *it is such that it minimizes the total energy of the molecular system…(energy due to dynamic interaction of all the particles involved because of the forces that act upon them)*

Hamiltonian *H* of the molecular system:

- *Operator* that governs the interaction of the involved particles…
- Considering all forces between nuclei and electrons we have…

$$H = H_{nucl} + H_e + U_{nucl} + V_{ext} + U_{ee}$$

H_{nucl} *Kinetic energy of the nuclei*

H_e *Kinetic energy of electrons*

U_{nucl} *Interaction energy of nuclei (Coulombic repulsion)*

V_{ext} *Nuclei electrostatic potential with which electrons interact*

U_{ee} *Electrostatic repulsion between electrons*

Mathematical Modelling: Schrödinger's Equation

Let the columns of Ψ*:*

$$\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$$

hold the wave functions corresponding the electrons…Then it holds that

$$H\Psi = \epsilon\Psi$$

- This is an eigenvalue problem…that becomes a usual…
- "algebraic" eigenvalue problem when we discretize ψ_i w.r.t. space (r)
- Extremely complex and nonlinear problem…since
- Hamiltonian and wave functions depend upon all particles…
- We can very rarely (only for trivial cases) solve it exactly…

Variational Principle (in simple terms!)

Minimal energy and the corresponding electron distribution amounts to calculating the smallest eigenvalue/eigenvector of the Schrödinger equation

Schrödinger's Equation: Basic Approximations

Multiple interactions of all particles…result to extremely complex Hamiltonian…which typically becomes huge when we discretize

Thus…a number of reasonable approximations/simplifications have been considered…with negligible effects on the accuracy of the modeling:

- **Born-Oppenheimer:** ***Separate the movement of nuclei and electrons…the latter depends on the positions of the nuclei in a parametric way…(essentially neglect the kinetic energy of the nuclei)***

- **Pseudopotential approximation:** ***Nucleus and surrounding core electrons are treated as one entity***

- **Local Density Approximation:** ***If electron density does not change rapidly w.r.t. sparse (r)…then electrostatic repulsion U_{ee} is approximated by assuming that density is locally uniform***

Density Functional Theory

High complexity is mainly due to the many-electron formulation of *ab initio* calculations…is there a way to come up with an one-electron formulation?

Key Theory

DFT: Density Functional Theory (Hohenberg,Kohn,Sham)

✓ **The total ground energy of a molecular system is a functional of the electronic density…(number of electrons in a cubic unit)**

✓**The energy of a system of electrons is at a minimum if it is an exact density of the ground state!**

- **This is an existence theorem…the density functional always exists**
- **…but the theorem does not prescribe a way to compute it…**

- **This energy functional is highly complicated…**
- **Thus approximations are considered…concerning:**
  - ➢ **Kinetic energy and**
  - ➢ **Exchange-Correlation energies of the system of electrons**

Density Functional Theory: Formulation (1/2)

Equivalent eigenproblem:

$$[H_{e_i} + V_{tot}(r;\rho)]\psi_i(r) \quad = \quad \epsilon_i\psi_i(r)$$

ψ_i One electron wave function

ϵ_i Energy of the i-th state of the system

H_{e_i} Kinetic energy of electron e_i

$V_{tot}(r;\rho)$ Total potential that acts on e_i at position r

$\rho(r) = \sum_{occ} |\psi_i(r)|^2$ Charge density at position r

Density Functional Theory: Formulation (2/2)

Furthermore:

$$V_{tot}(r;\rho) = V_{ion} + V_H + Vxc$$

V_{ion} **Potential due to nuclei and core electrons**

V_H **Coulomb potential from valence electrons** $\nabla^2 V_H = -4\pi\rho$

V_{xc} **Exchange-Correlation potential…function of the charge density ρ**

Non-linearity: ***The new Hamiltonian depends upon the charge density ρ while ρ itself depends upon the wave functions (eigenvectors) ψ_i***

Some short of iteration is required until convergence is achieved!

Self Consistent Iteration in PARSEC

Start with initial guess for $V_{tot}(r, \rho_0)$ and iterate

1. Solve $[H_{e_i} + V_{tot}(r; \rho_{k-1})]\psi_i(r) \; = \; \epsilon_i \psi_i(r)$
2. Update charge density $\rho_k(r) = \sum_{occ} \psi_i(r)^* \psi_i(r)$
3. Solve $\nabla^2 V_H = -4\pi\rho_k(r)$
4. Update $V_{tot}(r, \rho_k(r)) = V_{ion}(r) + V_H(r) + V_{xc}(\rho_k)$
5. Has V_{tot} converged? If yes then `break`
6. Update $V_{tot}(r, \rho_k) = \mathcal{F}(V_{tot}(r, \rho_k), V_{tot}(r, \rho_{k-1}))$

Calculate total energy, forces etc...

S.C.I in PARSEC: Computational Considerations

Start with initial guess for $V_{tot}(r, \rho_0)$ and iterate

1. Solve $[H_{e_i} + V_{tot}(r; \rho_{k-1})]\psi_i(r) \;=\; \epsilon_i \psi_i(r)$
2. Update charge density $\rho_k(r) = \sum_{occ} \psi_i(r)^* \psi_i(r)$
3. Solve $\nabla^2 V_H = -4\pi\rho_k(r)$
4. Update $V_{tot}(r, \rho_k(r)) = V_{ion}(r) + V_H(r) + V_{xc}(\rho_k)$
5. Has V_{tot} converged? If yes then `break`
6. Update $V_{tot}(r, \rho_k) = \mathcal{F}(V_{tot}(r, \rho_k), V_{tot}(r, \rho_{k-1}))$

Calculate total energy, forces etc...

S.C.Iin PARSEC: Computational Considerations

Start with initial guess for $V_{tot}(r, \rho_0)$ and iterate

1. Solve $[H_{e_i} + V_{tot}(r; \rho_{k-1})]\psi_i(r) \;=\; \epsilon_i \psi_i(r)$
2. Update charge density $\rho_k(r) = \sum_{occ} \psi_i(r)^* \psi_i(r)$

Conventional approach:

- ➢ ***Solve the eigenvalue problem (1)…and compute the charge densities…***
- ➢ ***This is a tough problem…many of the smallest eigenvalues…deep into the spectrum are required! Thus…***
- ➢ ***efficient eigensolvers have a significant impact on electronic structure calculations!***

Alternative approach:

- ➢ ***The eigenvectors ψ_i are required only to compute $\rho_k(r)$***
- ➢***Can we instead approximate charge densities without eigenvectors…?***
- ➢ ***Yes…!***

Computational Considerations in Applying Eigensolvers for Electronic Structure Calculations

The Eigenproblem

Hamiltonian Characteristics

- *Discretization: High-order finite difference scheme…leads to*
- *Large Hamiltonians!…typically N>100K…with significant…*
- *number of nonzero elements (NNZ)>5M…*
- *Hamiltonian is Symmetric/Hermitian…thus the eigenvalues are real numbers…some smallest and some larger than zero.*

Eigenproblem Characteristics (why is this a tough case?)

- *We need the algebraically smallest (leftmost) eigenvalues (and vectors)*
- *How many? Typically a large number of them. Depending upon the molecular system under study:*
    - *for standard spin-less calculations* → Si_xH_y: $(4x + y)/2$
    - *i.e. for the small molecule* $Si_{34}H_{36}$ *we need the 86 smallest eigenvalues…*
- *For large molecules, x,y>500 (or for exotic entities…quantum dots) thousands of the smallest eigenvalues are required….*
- *Using current state of the-art-methods we need thousands of CPU hours on DOE supercomputers…and we have to do that many times!*

Methods for Eigenvalues (basics only!)

Eigenvalue Approximation from a Subspace

Consider the standard eigenvalue problem: $Ax = \lambda x$

and let *V* be a thin *N* x *k* (N>>k) matrix…then

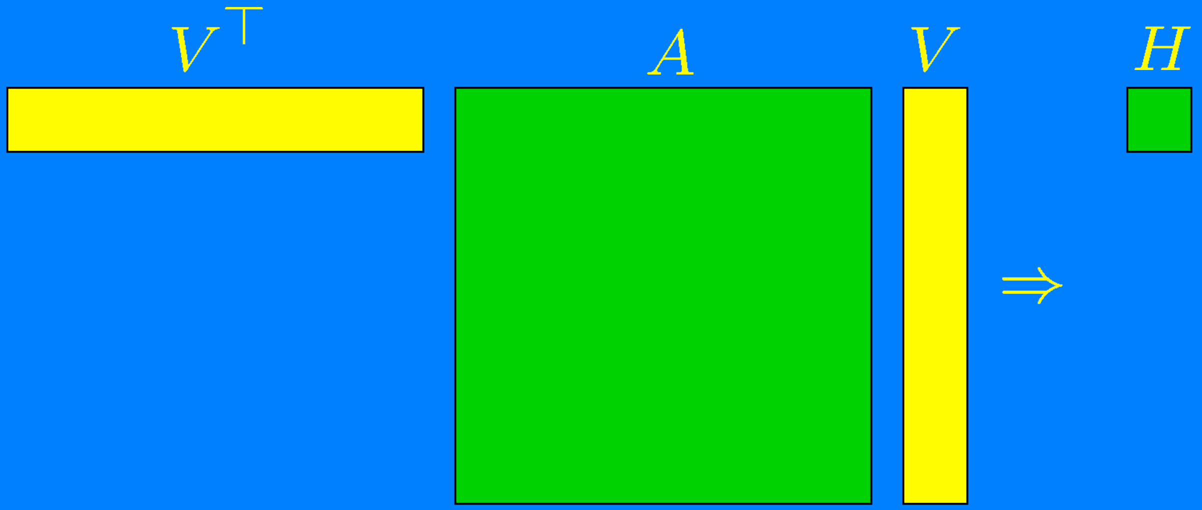
approximate the original problem with: $V^\top A V y = \theta V^\top V y$

Observe that:

- *Selecting V to have orthogonal columns …$V^TV = I$ … but it is expensive to come up with an orthogonal V*
- *Set $H = V^TAV$… then H is k x k …much smaller than N x N*

$$Hy = \theta y, \quad \lambda \approx \theta, \quad x \approx Vy$$

Subspace Methods



H is much smaller than A…use **LAPACK** *for H*

How to compute orthogonal *V* ?

- *For a non-symmetric matrix A…use Gramm-Schmidt (Arnoldi)…*
- *For a symmetric matrix A (our case) use Lanczos…*
- *Other approaches available (i.e. Jacobi-Davidson) but still some sort of Gramm-Schmidt is required…*

REMINDER

We need many eigenvalues/vectors O(1000)…thus V may not be that thin!!!

Symmetric Problems: Lanczos

Basic property

Theoretically…(assuming no round-off errors)…Lanczos can build a very large orthogonal basis V requiring in memory only 3 columns of V at each step!

Lanczos

```
1. Input: Matrix $A$, unit norm starting vector $v_0$, $\beta_0 = 0$, # $k$
2. For j = 1,2,…,k Do
    3. $w_j = Av_j$                         MATRIX VECTOR     NO SYNC.
    4. $w_j = w_j - \beta_j v_{j-1}$        DAXPY             NO SYNC.
    5. $\alpha_j = (w_j, v_j)$              DOT PRODUCT       SYNC. -BCAST
    6. $w_j = w_j - \alpha_j v_j$           DAXPY             NO SYNC.
    7. $\beta_{j+1} = \|w_j\|_2$.           DOT PRODUCT       SYNC. - BCAST
    8. If $\beta_{j+1} = 0$ then STOP
    9. $v_{j+1} = w_j / \beta_{j+1}$        DSCAL             NO SYNC.
10. EndDO
```

Lanczos in Finite Arithmetic…

Round-off errors

- ➢ *Lanczos vectors v_i quickly loose orthogonality…so that*
- ➢ *V^TV is no longer orthogonal…thus*
- ➢ *We need to check if v_j is ⊥ to previous vectors 0,1,…,j-1*
- ➢*If NOT … reorthogonalize it against previous vectors (Gramm-Schmidt)*

Lanczos

```
1. Input: Matrix $A$, unit norm starting vector $v_0$, $\beta_0 = 0$, # $k$
2. For j = 1,2,…,k Do
    3. $w_j = Av_j$
    4. $w_j = w_j - \beta_j v_{j-1}$
    5. $\alpha_j = (w_j, v_j)$
    6. $w_j = w_j - \alpha_j v_j$
    7. $\beta_{j+1} = ||w_j||_2.$
    8. If $\beta_{j+1} = 0$ then STOP
    9. $v_{j+1} = w_j / \beta_{j+1}$
10. EndDO
```

ORTHOGONALITY IS LOST HERE…SO THESE STEPS ARE REPEATED AGAINST *ALL* PREVIOUS VECTORS… SELECTIVE REORTH IS ALSO POSIBLE (SIMON, LARSEN)

Practical Eigensolvers and Limitations

ARPACK (Sorensen-Lehoucq-Yang): Restarted Lanczos

- *Remember that O(1000) eigenvalues/vectors are required…thus*
- *we need a very long basis V…k = twice the number of eigenvalues which*
- *will result in a large number of reorthogonalizations…*
- *Synchronization costs – Reorthogonalization costs – and Memory costs become intractable for large problems of interest…*

Shift-Invert Lanczos (Grimes et all) – Rational Krylov (Ruhe)

- *work with matrix $(A-\sigma_i I)^{-1}$ instead…*
- *compute some of the eigenvalues close to σ_i each time…thus a smaller basis is required each time…BUT*
- *many shifts σ_i are required…*
- *cost of working with the different "inverses" $(A-\sigma_i I)^{-1}$ becomes prohibitive for (practically) large Hamiltonians…*

We need alternative methods that can build large projection bases without the reorthogonalization-synchronization costs

Automated Multilevel Substructuring

Component Mode Synthesis (CMS) (Hurty ’60, Graig-Bampton ’68)

Well known alternative to Lanczos type methods. Used for many years in Structural Engineering. *But it too suffers from limitations due to problem size…*

AMLS, (Bennighof, Lehoucq, Kaplan and collaborators)

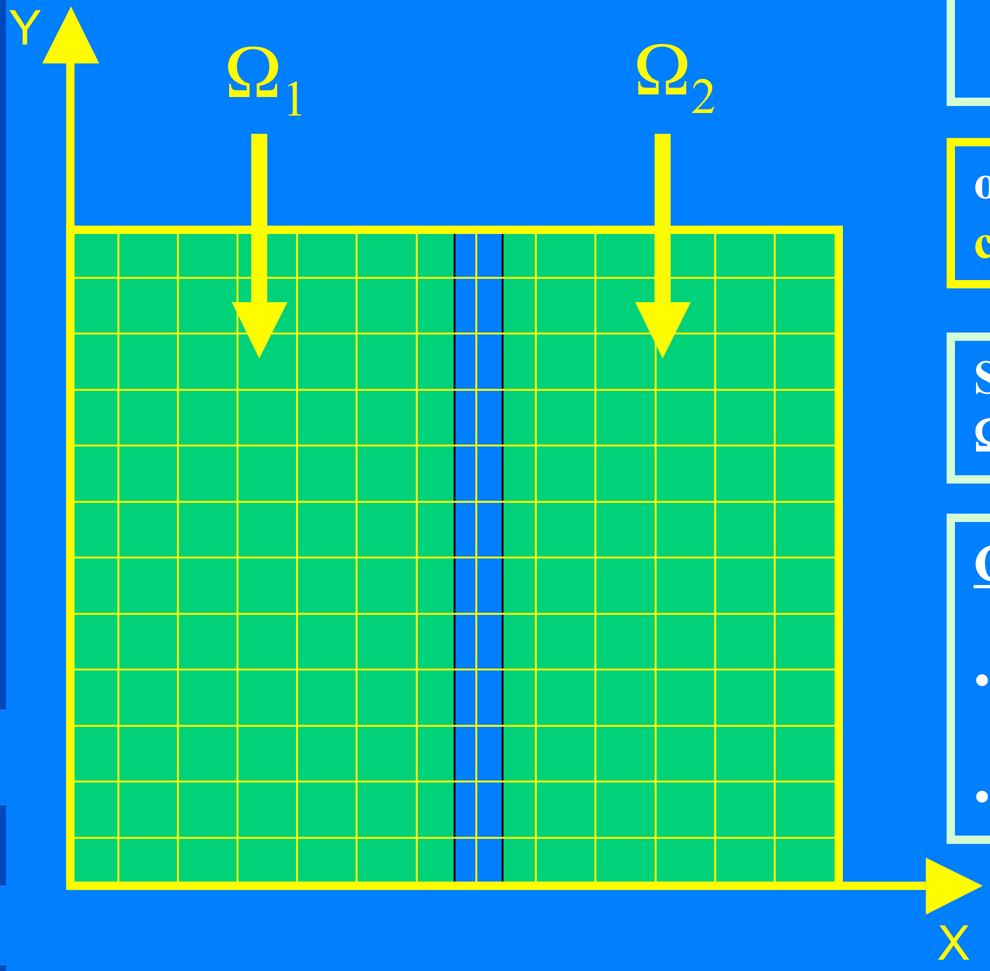
- Multilevel CMS method (solves the dimensionality problem)
- Automatic computation of substructures (easy application)
- Approximation: Truncated Congruence Transformation
- Builds very large projection basis *without reorthogonalization*
- Successful in computing thousands of eigenvalues in vibro-acoustic analysis ($N>10^7$) in a few hours on workstations (Kropp–Heiserer, 02)

Spectral Schur Complements (Bekas, Saad)

✓Significantly improves AMLS accuracy…suitable for electronic structure calculations

✓(unlike AMLS) …framework for the iterative refinement of the approximations

Component Mode Synthesis: a model problem



Consider the model problem:

$$-\nabla^2 u = \lambda u$$

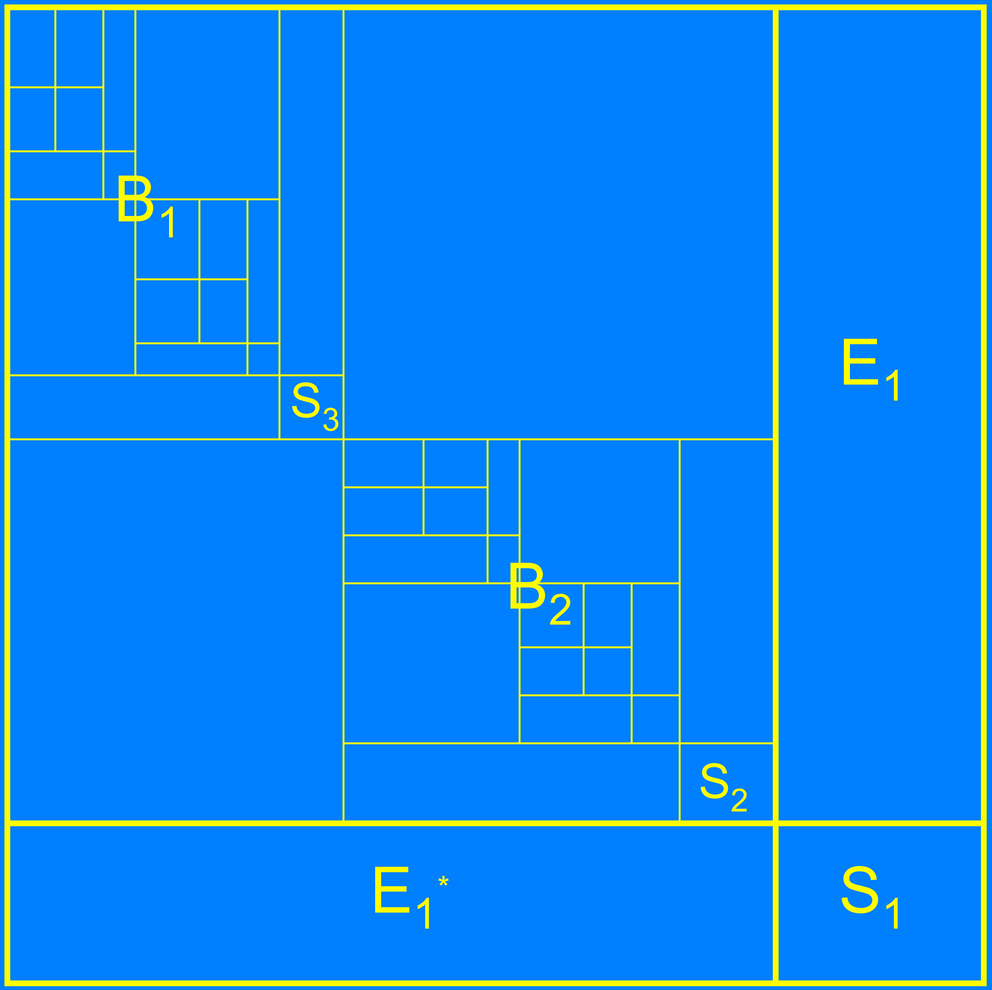
on the unit square Ω. **We wish to compute smallest eigenvalues.**

Subdivide Ω into 2 subdomains: Ω_1 and Ω_2

Component Mode Synthesis

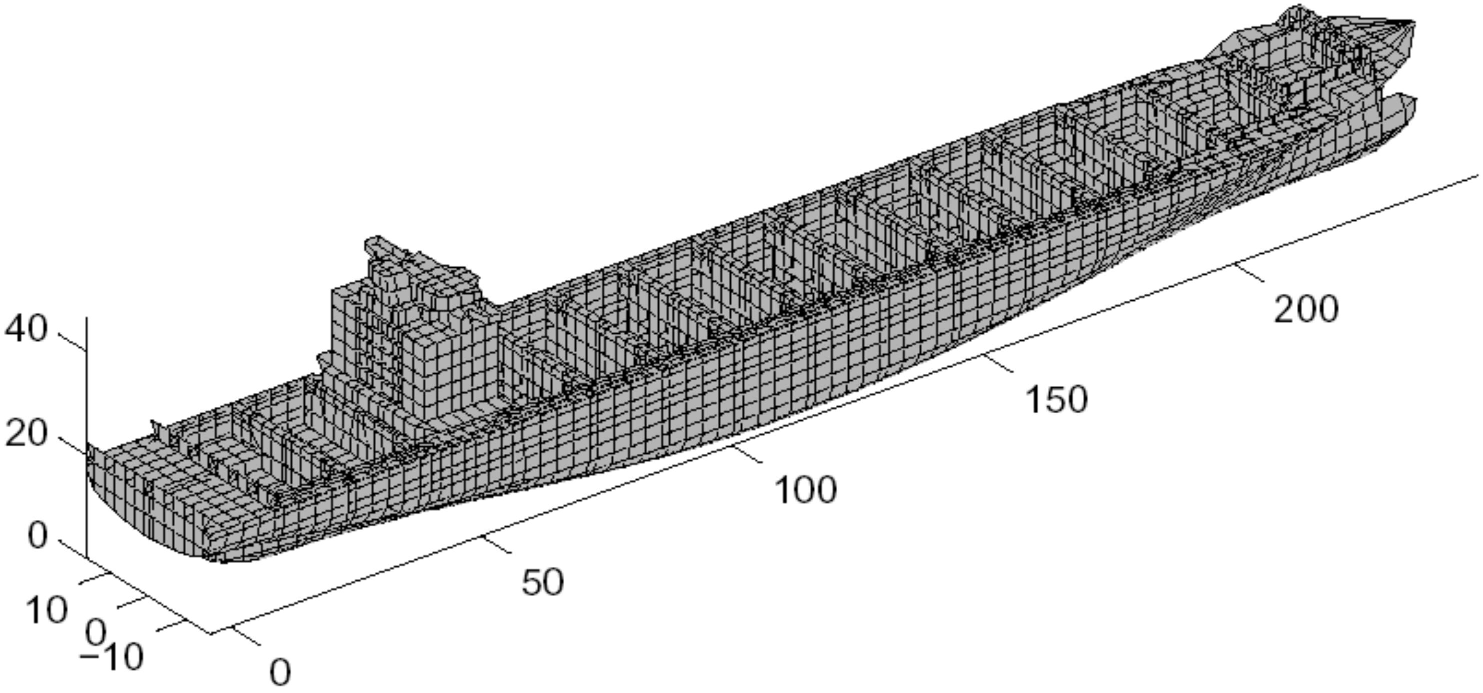
- Solve problem on each Ω_i
- "Combine" partial solutions

AMLS: Multilevel application



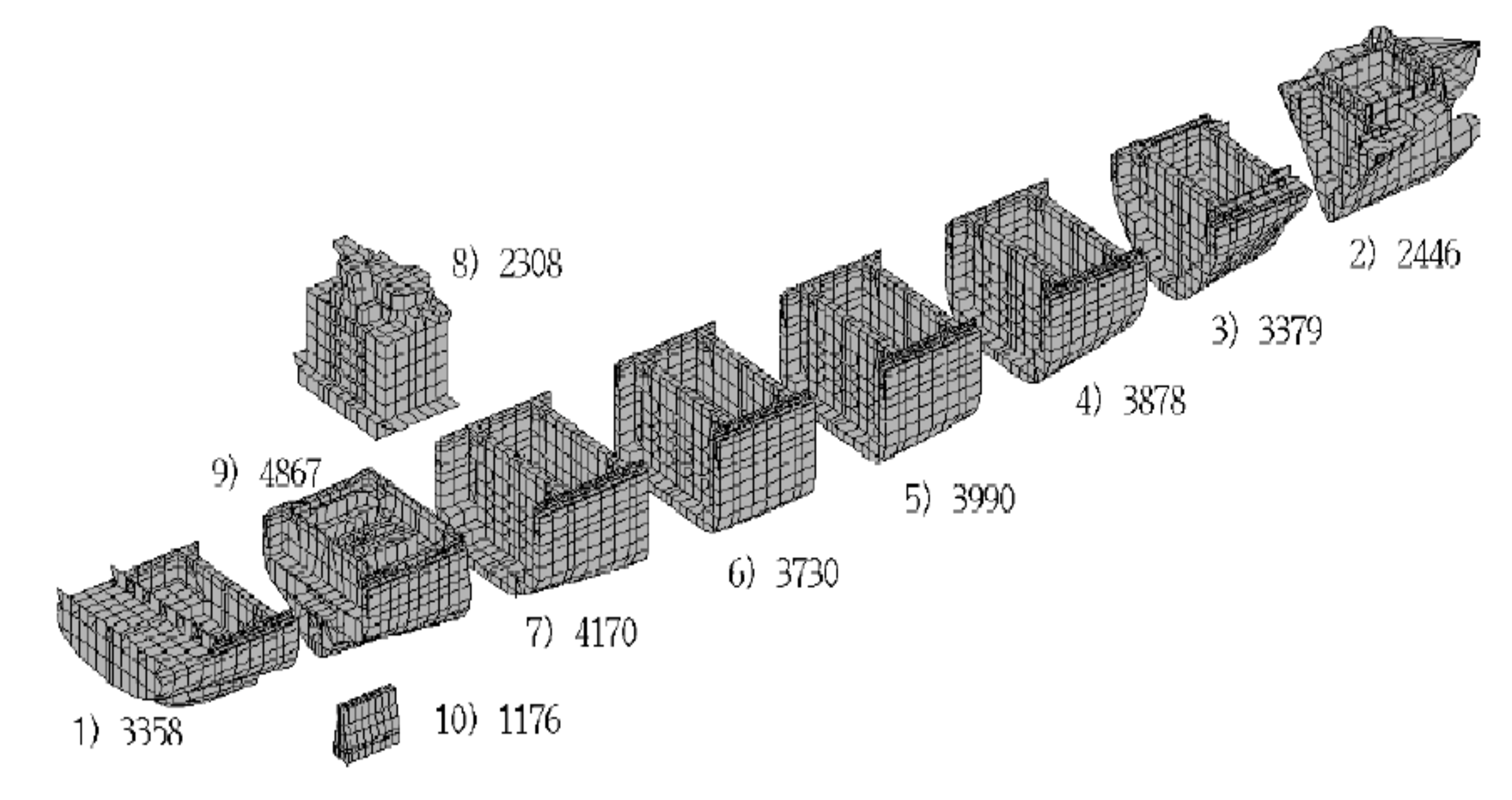
Scheme applied recursively. Resulting to thousands of subdomains. Successful in computing thousands smallest eigenvalues in vibro-acoustic analysis with problem size $N>10^7$ (Kropp – Heiserer BMW)

AMLS: Example



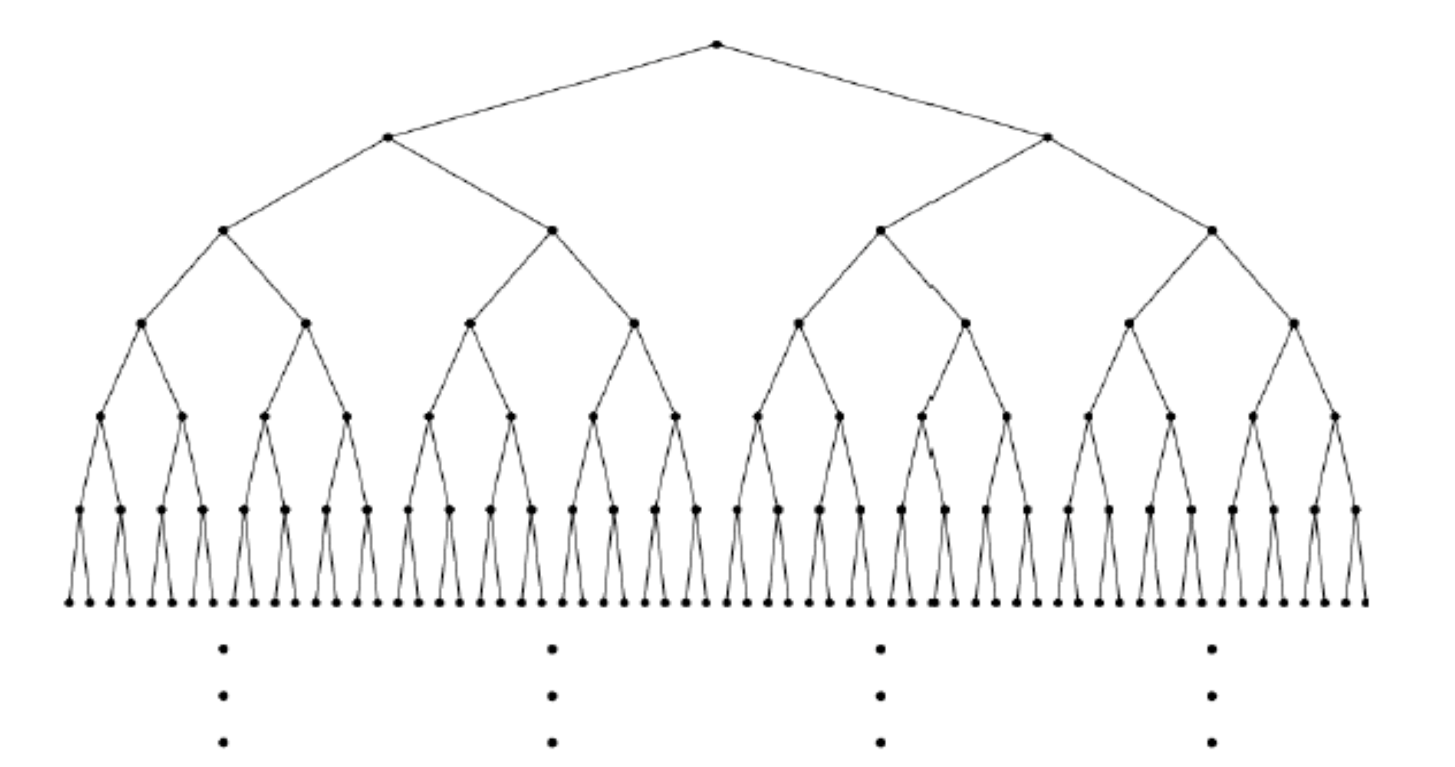
Example: Container ship, 35K degrees of freedom
(Research group of prof. H. Voss, T. U. Hamburg, Germany)

AMLS: Example



Example: Container ship, 35K degrees of freedom (Research group of prof. H. Voss, T. U. Hamburg, Germany)

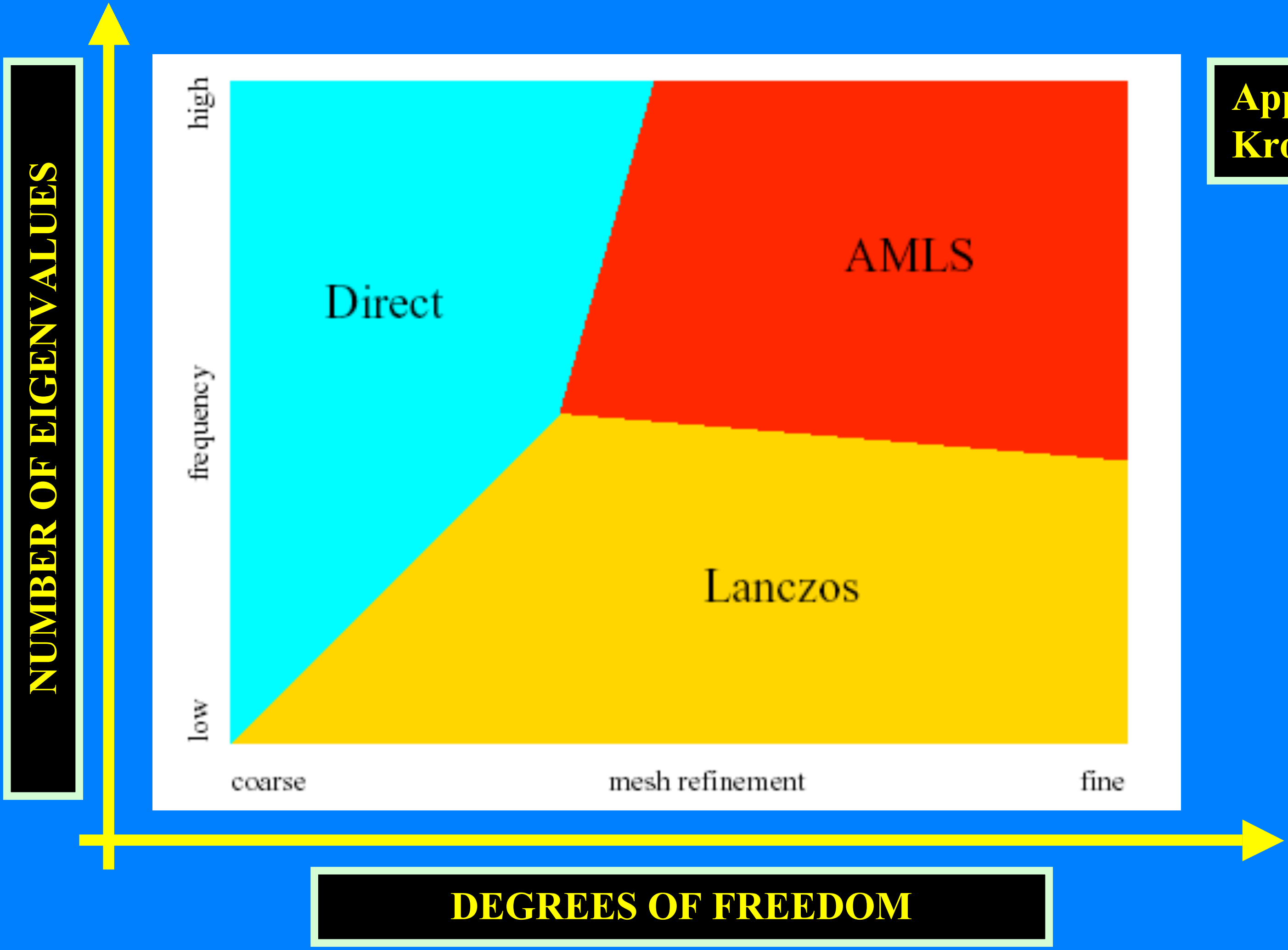
AMLS: Example



AMLS: Substructure tree (Kropp-Heiserer, BMW)

- ➢ Multilevel parallelism…
- ➢ Both Top-Down and Bottom-Up implementations are possible…
- ➢ At each node we need to solve a linear system…
- ➢ Multilevel solution of linear systems…level k depends-benefits from level k+1

Problem Set…AMLS v.s. Standard Methods



NUMBER OF EIGENVALUES

DEGREES OF FREEDOM

Application Domains, Kropp-Heiserer, 02

Implementation Issues – Trilinos

ab initio calculations:…many ingredients required for successful techniques

- Mesh generation…discretization
- Visualization of input data…results…geometry
- Efficient data structures-communicators for parallel computations
- Efficient (parallel) Matrix-Vector and inner products
- Linear system solvers
- State-of-the-art eigensolvers…

A unifying software development environment will prove to be very useful

- ease of use…
- reusability…(object oriented)
- portable…

TRILINOS http://software.sandia.gov/trilinos

- software multi-package…developed at SANDIA (M. Heroux)
- modular…no need to install everything in order to work!
- Capabilities of LAPACK, AZTEC, Chaco, SuperLU, etc…combined
- very active user community…ever evolving!
- ease of use…without sacrificing performance

Conclusions

Large Scale Challenges in Computational Materials Science

- In DFT eigenvalue calculations dominate…
- …many O(1000) eigenvalues/vectors required…
- …easily reaching and exceeding the limits of state-of-the-art traditional solvers
- AMLS appears as an extremely attractive alternative…however accuracy requirements and efficient parallel implementation is still under development

Many open problems in *ab initio* calculations…one of the most active fields of research today!