Overview of ESSEX-II
- Equipping Sparse Solvers for Exascale -

Tetsuya Sakurai
Center for Artificial Intelligence Research (C-AIR)
University of Tsukuba / JST CREST

Contents

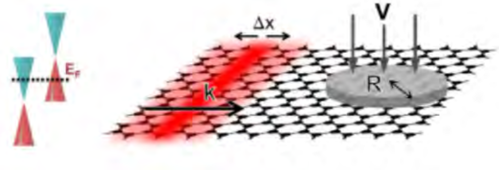
- Overview of ESSEX-II
  - ESSEX Motivation
  - Programming & Computational Algorithms
  - Applications
- Development of Highly Scalable Eigenvalue Solvers
  - Simulation and Data Analysis
  - Applications
  - Numerical Examples
- Conclusions

ESSEX II –
Equipping Sparse Solvers for Exascale

Gerhard Wellein	Computer Science, University Erlangen
Bruno Lang	Applied Computer Science, University Wuppertal
Achim Basermann	Simulation & SW Technology, German Aerospace
Holger Fehske	Institute for Physics, University Greifswald
Georg Hager	Erlangen Regional Computing Center
Tetsuya Sakurai	Applied Mathematics, University of Tsukuba
Kengo Nakajima	Computer Science, University of Tokyo

ESSEX Motivation

Quantum physics/information applications



Large, Sparse

$$i\hbar \frac{\partial}{\partial t}\psi(\vec{r},t) = H\psi(\vec{r},t)$$

and beyond….

$$\boldsymbol{H}\,\boldsymbol{x} = \boldsymbol{\lambda}\,\boldsymbol{x}$$

"Few" (1,…,100s) of eigenpairs

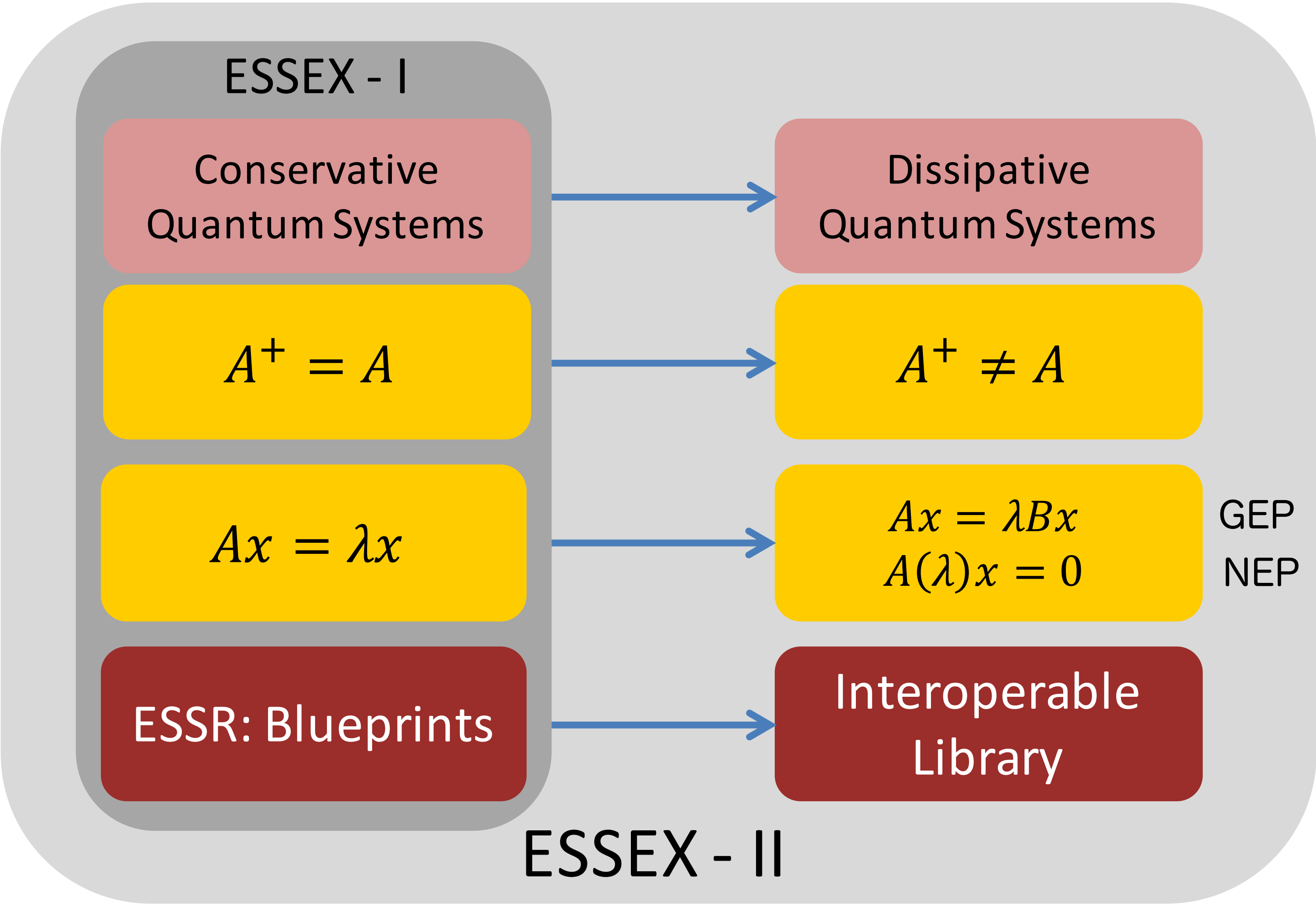
"Bulk" (100s,…,1000s) eigenpairs

$$\{\lambda_1, \lambda_2, \ldots, \ldots, \ldots, \ldots, \lambda_k, \ldots, \ldots, \ldots, \ldots, \lambda_{n-1}, \lambda_n\}$$

Good approximation to full spectrum (e.g. Density of States)

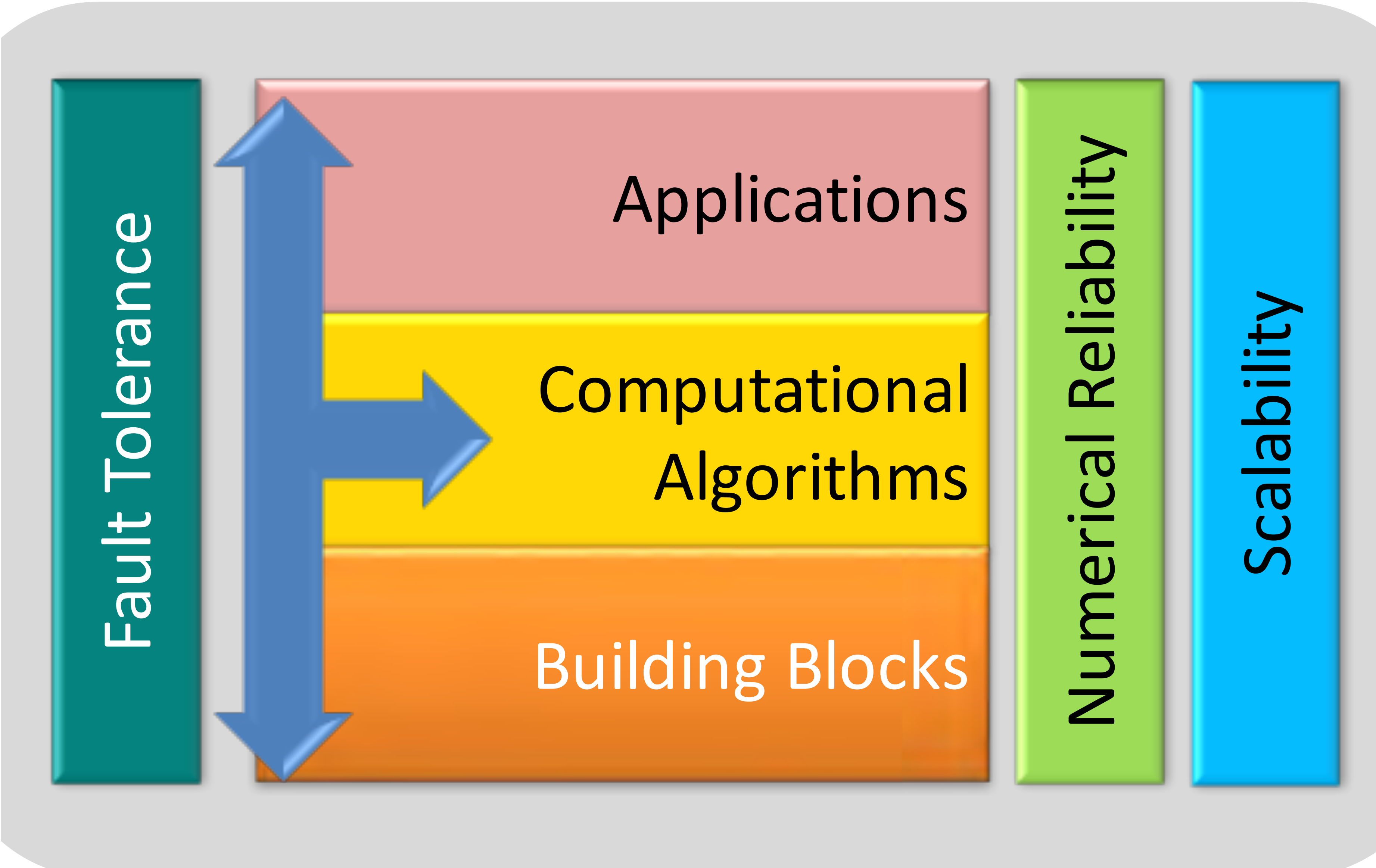
→ Sparse eigenvalue solvers of broad applicability

ESSEX-I → ESSEX-II



→ Sparse eigenvalue solvers of broad applicability

Enabling Exascale through software codesign

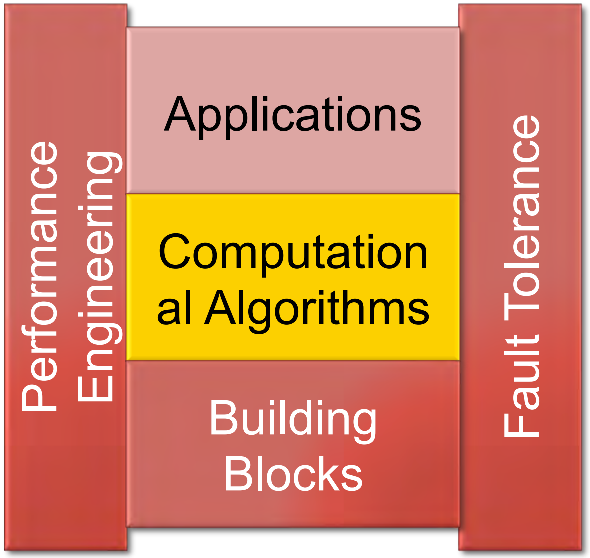


Programming

Extended

Building Blocks, Parallelization, and Performance Engineering

- Holistic performance and power engineering
- Advanced building blocks engineering



Extended

Fault Tolerance

- From prototype to application software
  - Asynchronous checkpointing & I/O
  - Automatically fault-tolerant applications

NEW

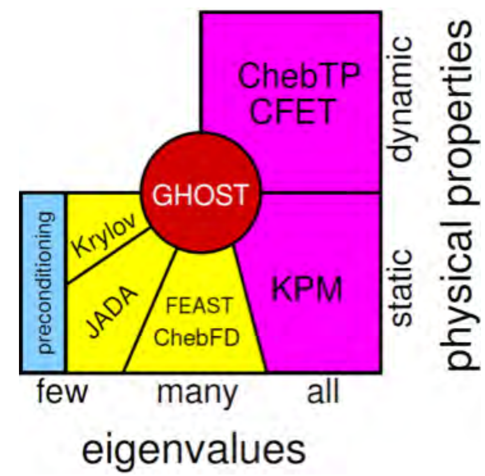
Numerical Reliability

- Performance aspects
  - Silent data corruption / skeptical programming
  - High-precision reduction operations

Computational Algorithms

NEW

- **Non-Hermitian**: ChebTP / CFET / JaDa
  - Extreme-scale simulations for dissipative quantum systems
  - Numerical range computation & matrix balancing



Extended

- **C**heby**s**hev**F**ilter**D**iagonalization:
  - >10^3 interior eigenvalues of >10^9 matrix dim.
  - Simple, HW-efficient & low synchronization cost

NEW

- **Preconditioning** & **Communication Hiding**
  - Asynchronous JaDa: "pipelining" & preconditioning
  - AMG preconditioning for blocked JaDa & FEAST

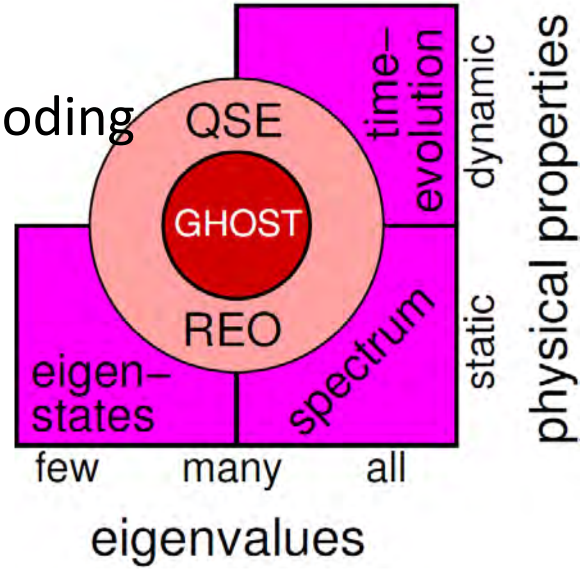
NEW

- Leveraging FEAST techniques + GHOST
  → Nonlinear Sakurai-Sugiura Method (NSSM)

Applications

Extended

- **Quantum State Encoding (QSE)**
  - Complex (non-stencil) matrix structure encoding
  - Dissipative systems: Sparse Dense



New

- **Matrix Reordering Strategies (REO)**
  - Application-specific
  - General techniques, e.g. **PMRSB**

Extended

- **Quantum Physics/Information Applications**
  - **Topological materials**
    Graphene & topological insulators
  - **Dissipative quantum systems**
    Light-harvesting molecules & optomechanics
  - Rich collection of quantum physics problems

$A^{+} = A$

$A^{+} \neq A$

ESSEX@PASC2017 26 – 28 June, 2017

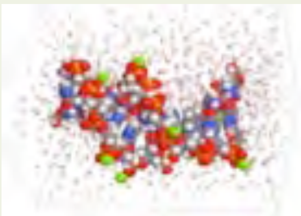
- Computing bulks of inner eigenpairs of large sparse matrices: from applications and algorithms to performance and software engineering (I + II)
  - Takeo Hoshi (Tottori University, Japan)
  - Tetsuya Sakurai (University of Tsukuba, Japan)
  - Yousef Saad (University of Minnesota, USA)
  - Andreas Alvermann (Universität Greifswald, Germany)
  - Kengo Nakajima (The University of Tokyo, Japan)
  - Hartwig Anzt (University of Tennessee, USA)
  - Jonas Thies (German Aerospace Center, Germany)
  - Mike Heroux (Sandia National Laboratories, USA)

  https://pasc17.pasc-conference.org/program/minisymposia/

Development of
Highly Scalable Eigenvalue Solvers

Eigenvalue Problems in Simulation and Data Analysis

Electronic State Analysis | Modeling | Eigenvalue Problem



Nano Materials

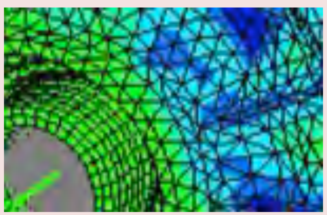
Basis Function Expansion

$$\psi_i(\boldsymbol{r}) = \sum_j \phi_j(\boldsymbol{r})$$

Interaction of Electrons

$$H\boldsymbol{x} = \lambda S\boldsymbol{x}$$

Structural Analysis



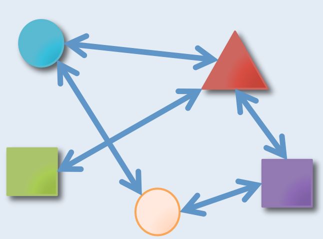
Car body design, etc.

Vibration Analysis

Interaction of Finite Elements

$$K\boldsymbol{x} = \lambda M\boldsymbol{x}$$

Data Analysis



Data

Similarity of data elements

$$d_i = \sum_j A_{i,j} \quad L = D - A$$

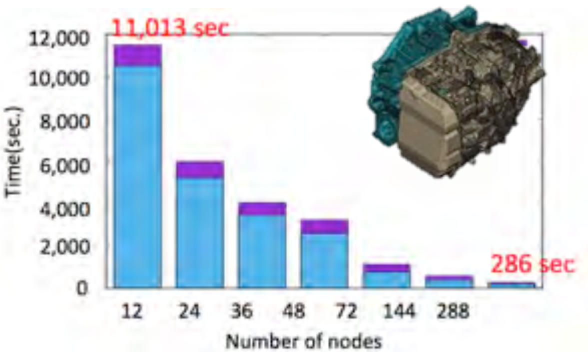
Interaction of Data Elements

$$L\boldsymbol{x} = \lambda D\boldsymbol{x}$$

Applications

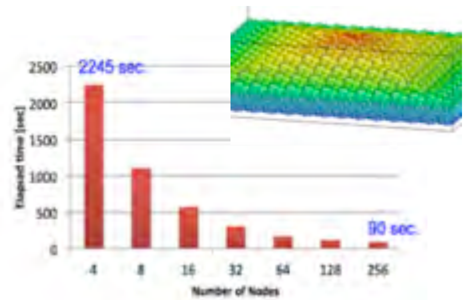
Simulation

Structural analysis

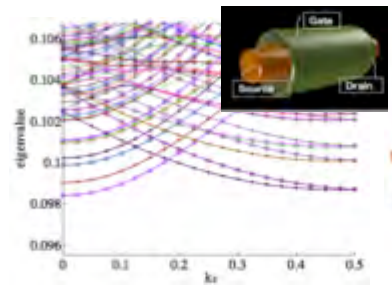


Finite element models

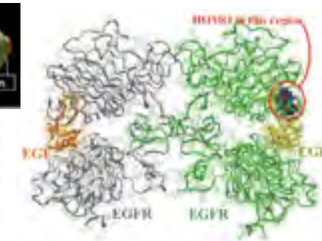
Nano materials



Quantum dots

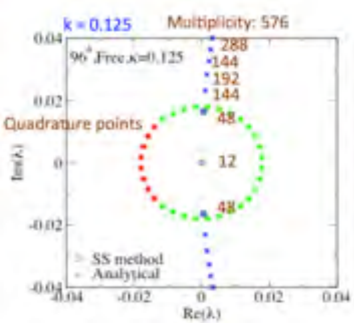


Nano device

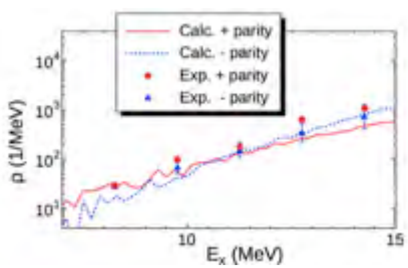


Biomolecule

Elementary physics

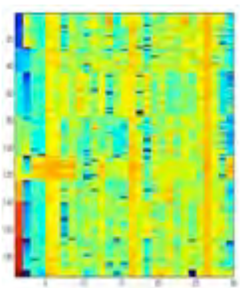


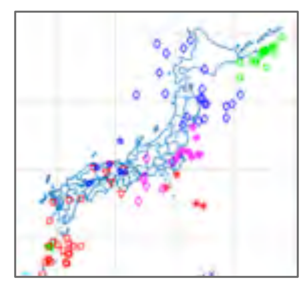
Lattice-QCD



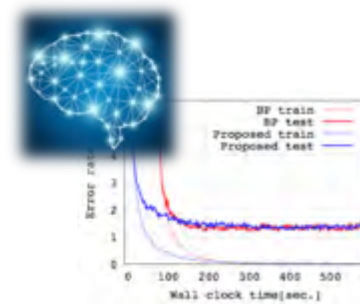
Nuclei Physics

Data analysis, Deep learning





Gene data,
earthquake data, etc.

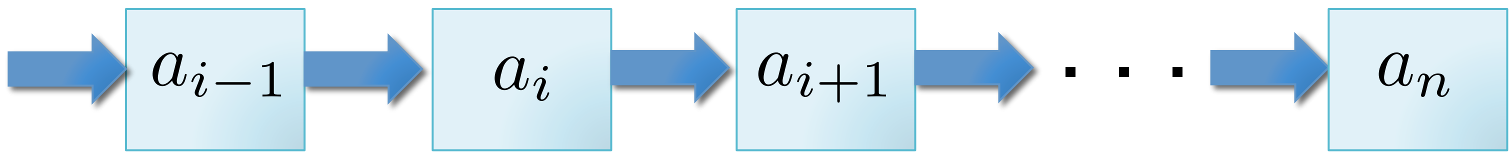


Neural
networks

For Parallel Scalability

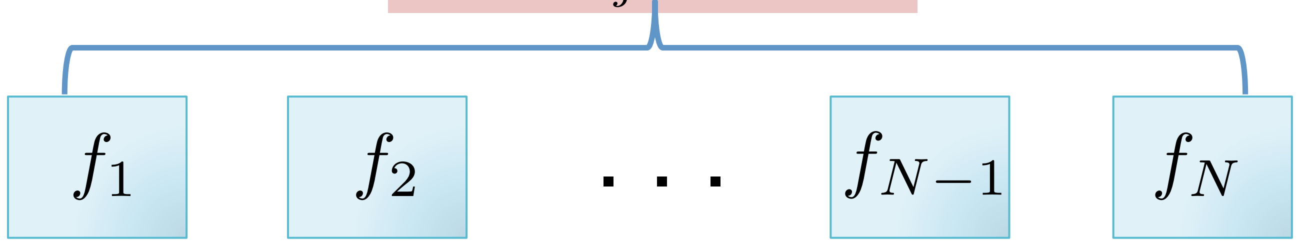
- Avoid recurrence calculations in eigenvalue computation
  - Algorithms described by recurrence relations:
    (ex. Krylov methods):

$$a_{i+1} = b_i\, a_i + c_i\, a_{i-1}$$



  - Algorithms without recurrence calculations:
    (ex. Numerical quadrature):

$$I_N = \sum_{j=1}^{N} w_j f_j$$

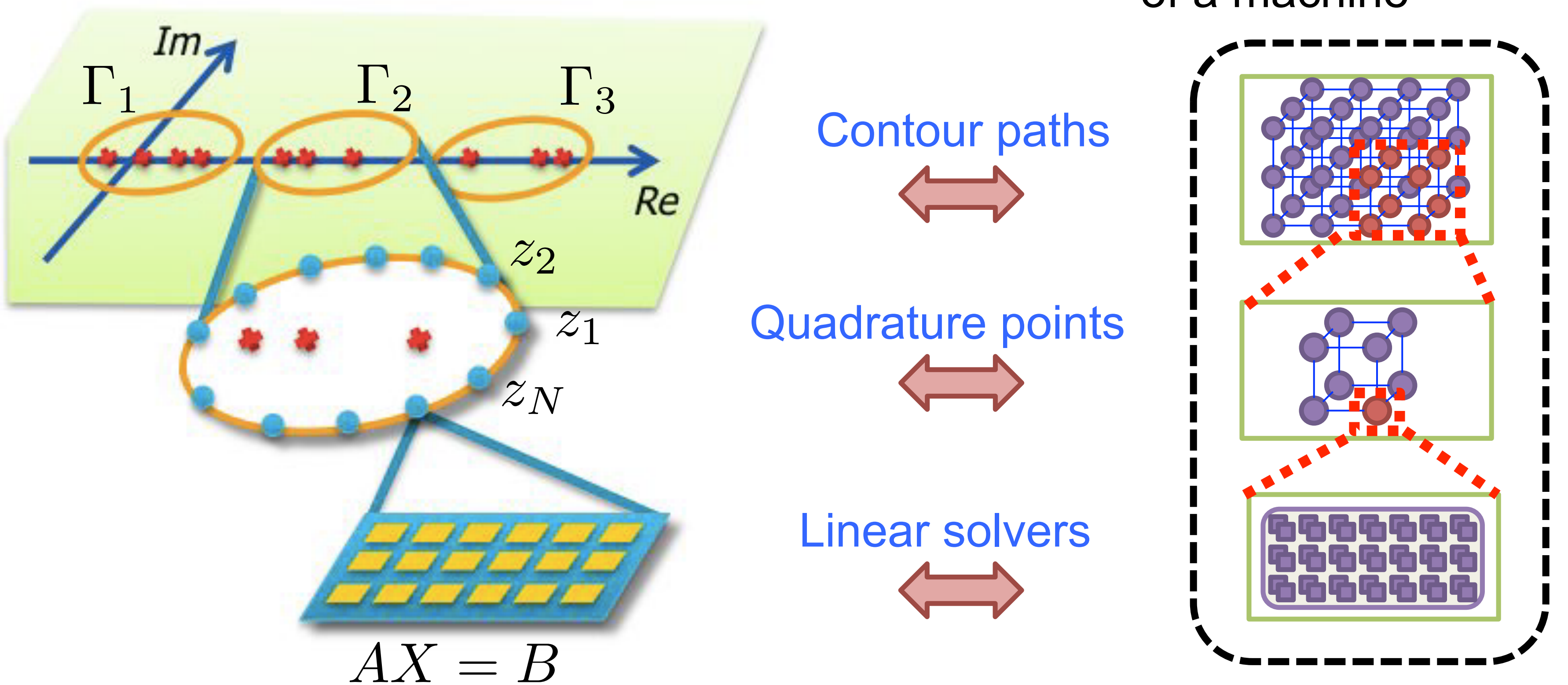


Hierarchical Parallel Structure

- Computing resources are assigned according to a hierarchical structure of the algorithm

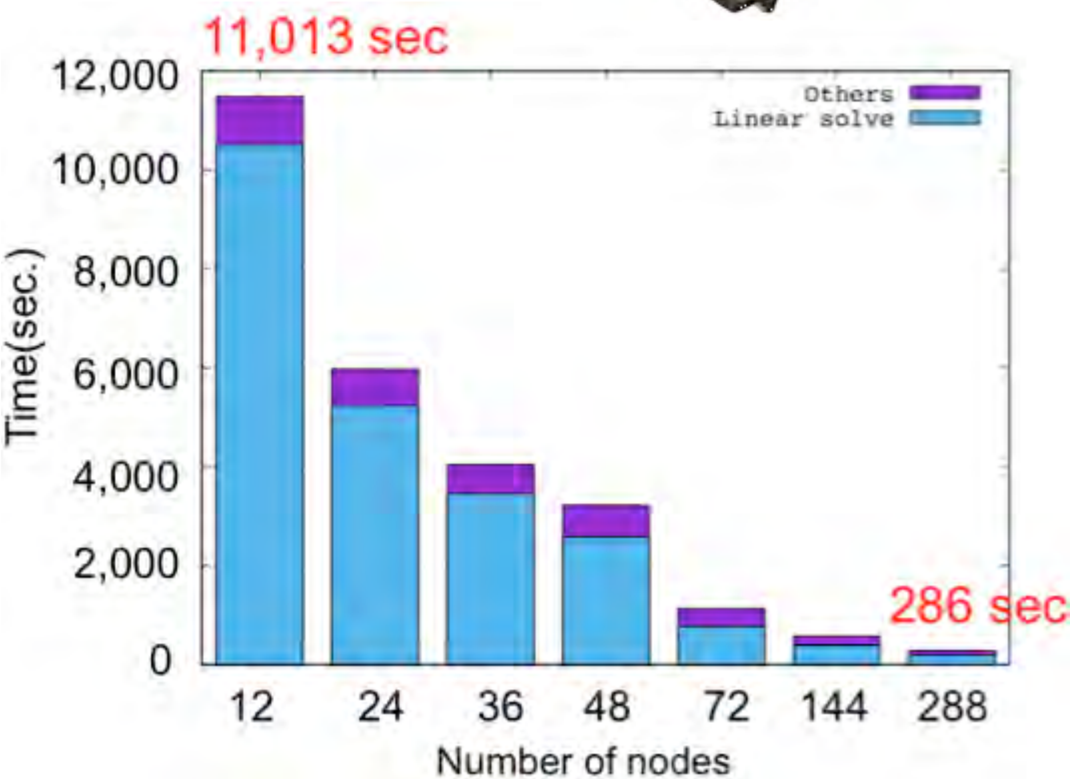
Hierarchical structure of the algorithm

Hierarchical structure of a machine



Numerical Example: Transmission Design

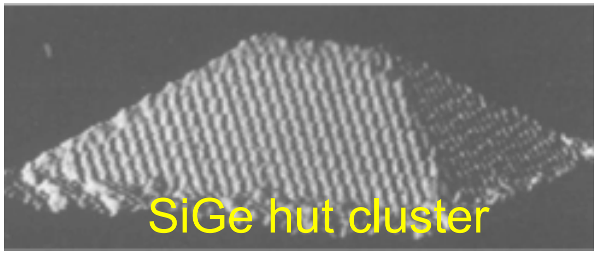
- Application: Automatic transmission design
  - Matrices are derived by FEM
    - Mixture of shell and solid models
    - Frequency range: 0 ~ 6,000 Hz
    - GEP: 16,747,926 DOF (916 eigenpairs)
  - Test environment: COMA @Univ. of Tsukuba
  - Solvers
    - z-Pares (SSM)
      - #contour paths: 4
      - #quadrature points: N = 16
    - Linear solver: MUMPS

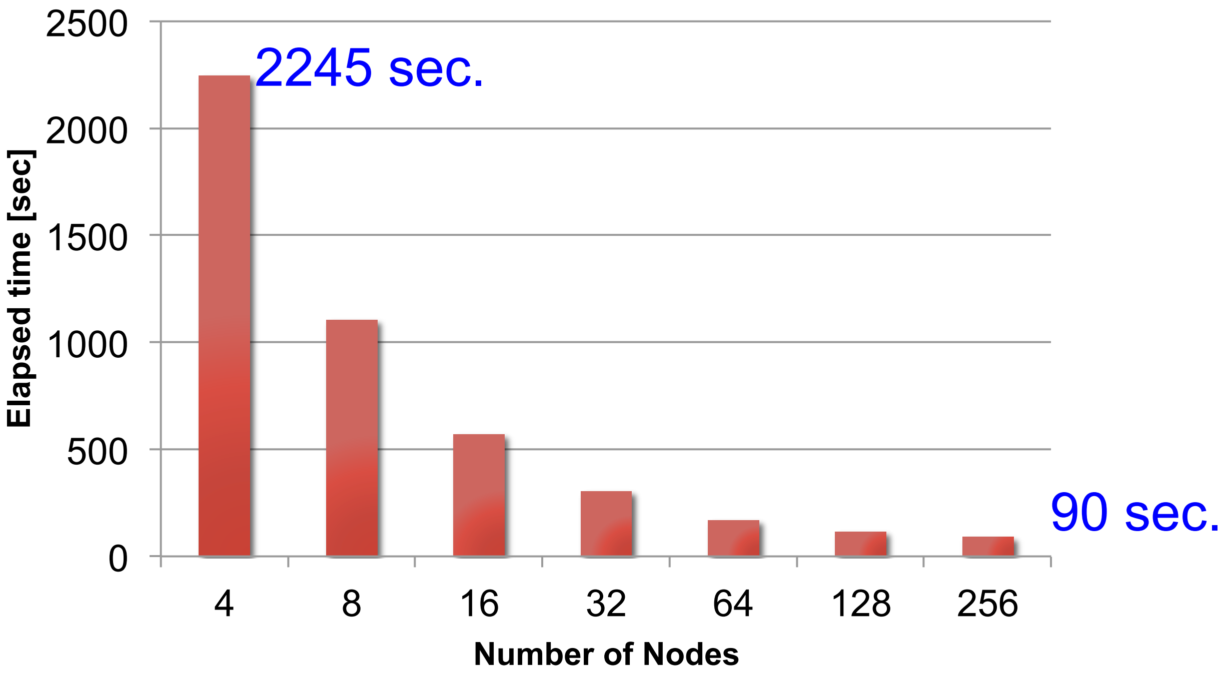


[Ide, Toda, Futamura, S '2016]

Numerical Example: Order-*N* DFT

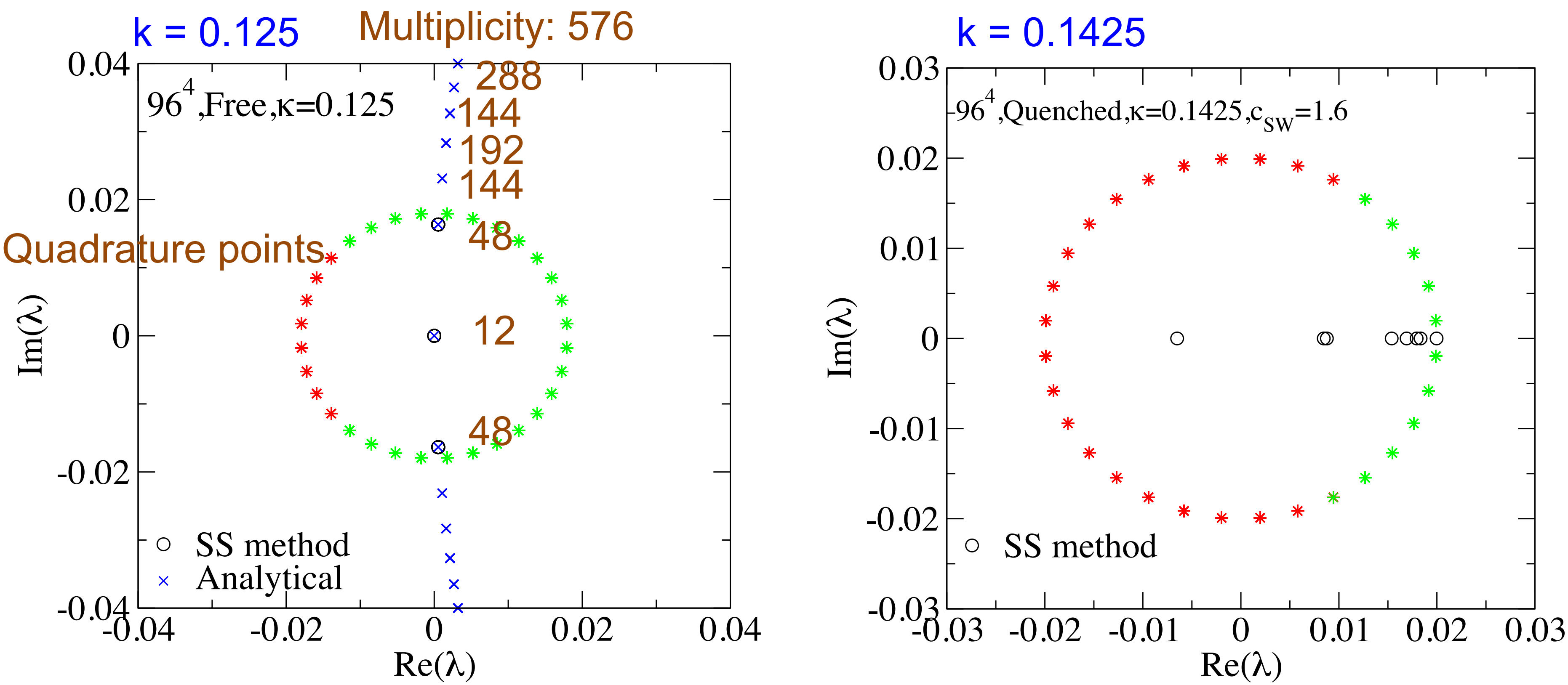
- Order-*N* DFT code CONQUEST
  - SiGe hut cluster with 200,000 atoms
    by Nakata and Miyazaki
    - Matrix size : 778,292, NNZ : 13,247,248
    - GEP: 223 eigenpairs around HOMO-LUMO are computed
    - Linear solver : MUMPS





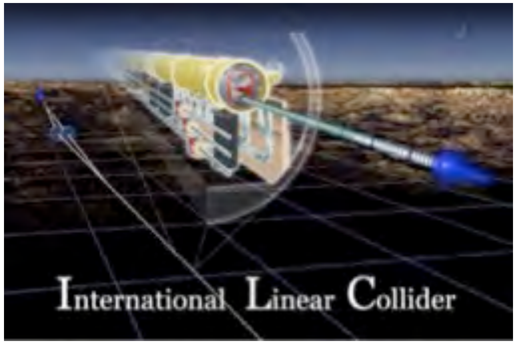
Numerical Example: Lattice QCD

- O(a)-improved Wilson-Dirac operator (Suno, Kuramashi, S, et al.)
  Matrix dim.: 1,019,215,872 (SEP, non-Hermitian)
  Test environment: 16,384 nodes of the K-Computer
  Linear solver: BiCGStab



Numerical Example: Nonlinear Eigenvalue Problem

- Test problem (NEP):
  Simulation of the international linear collider



http://www.linearcollider.org/

$$T(\lambda)\boldsymbol{x} = \mathbf{0}$$

$$T(\lambda) = K - \lambda^2 M + \mathrm{i}\sum_{j=1}^{t} \sqrt{\lambda^2 - \sigma_j^2} W_j,$$

where $t = 1$, $\sigma_1 = 0$.

Matrix size: 2,738,556

Test environment: Cray-XT4 at NERSC @Berkeley

Linear solver: SuperLU_DIST

#cores	256	512	1024	2048
time(sec.)	2513	1273	661	334
speedup	-	1.97	1.93	1.92

[Yamazaki et al. '2013]

Examples on Oakforest-PACS

Oakforest-PACS: KNL Cluster system

(Co-operated with Univ. Tokyo)

- Intel Xeon Phi 7250, 96GB memory, 8208 nodes
  - Peak performance: 25 PFLOPS

- Structural Analysis in Car Body Design
  - Generalized eigenvalue problem (GEP)
  - Shell model, 95 million DOF
  - 512 nodes, 1 hour

- Complex Band Structure Calculation
  - Quadratic eigenvalue problem (NEP)
  - Eigenvalues in ring region on the complex plane
  - CNT with 10240 atoms, 2048 nodes, 2 hours

Conclusions

- Overview of ESSEX-II
  - ESSEX Motivation
  - Programming & Computational Algorithms
  - Applications

- Simulation and Data Analysis
  - Large-scale problems
  - High performance eigenvalue solver

- Developing a scalable parallel eigensolver
  - Scalable eigensolver: Sakurai-Sugiura method (SSM)
  - Software:
    z-Pares in Fortran95, CISS in SLEPc, sseig in MATLAB

    Implementation on GHOST/PHIST in ESSEX-II