

# Verbundprojekt ELPA-AEO

<http://elpa-aeo.mpcdf.mpg.de>

## Eigenwert-Löser für Petaflop-Anwendungen – Algorithmische Erweiterungen und Optimierungen

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

BMBF Projekt 01IH15001  
Feb 2016 - Jan 2019

## Projekt-Partner



Lehrstuhl für Angewandte Informatik  
Prof. B. Lang (BUW)



**Fritz-Haber-Institut  
der  
Max-Planck-Gesellschaft**

Prof. M. Scheffler, Dr. Ch. Carbogno (FHI)



Dr. H. Lederer, Dr. A. Marek (MPCDF)



Lehrstuhl für Informatik mit Schwerpunkt  
Wissenschaftliches Rechnen (TUM-SCCS)  
Prof. H.-J. Bungartz, Prof. Th. Huckle

Lehrstuhl für Theoretische Chemie  
Prof. K. Reuter, Dr. Ch. Scheurer (TUM-CH)

# Vorhabensziele

## **Eigenwert-Löser für Petaflop-Anwendungen – Algorithmische Erweiterungen und Optimierungen**

- Steigerung der Effizienz von Supercomputer-Simulationen, für die die Lösung des Eigenwertproblems für dichte und Band-strukturierte symmetrische Matrizen zu einem entscheidenden Beitrag wird.
- Adressierung noch größerer Probleme als bisher
- Verringerung des mit der Simulation verbundenen Rechenaufwands
- Reduktion von Ressourceneinsatz und Energieverbrauch für vorgegebene Genauigkeit bei weiterhin hoher Software-Skalierbarkeit

## Zu förderpolitischen Zielen und Relevanz

**Förderprogramm** „IKT 2020 – Forschung für Innovationen.“

**Gebiet** „Anwendungsorientierte HPC-Software für das Hoch- und Höchstleistungsrechnen in Wissenschaft und Wirtschaft“

**Themenfeld** „Entwicklung innovativer paralleler Algorithmen und Methoden“.

**Zusammenarbeit von HPC-Experten und Anwendern** bzgl. Entwicklung innovativer paralleler Algorithmen und Methoden mit sehr guter Skalierbarkeit für wichtige Anwendungen insbesondere aus der Materialforschung

### **Relevante Anwendungsgebiete**

**Materialforschung** - Elektronenstrukturberechnungen

**Biomolekulare Forschung** - Struktur-Analysen

**Strukturdynamik** - Baustatik

## Komplexität der Problemstellung

Adressiert durch **enge interdisziplinäre Zusammenarbeit** von Experten aus

- Elektronenstrukturtheorie
- Numerischer Mathematik
- Informatik
- Hochleistungsrechnerspezialisten

**Verzahnung** von

- Anwendungsdomänen
- Simulationswerkzeugen
- Algorithmenentwicklern

# Ausgangspunkt: ELPA-Bibliothek

ELPA: Entwickelt im BMBF-Verbundvorhaben 01IH08007  
aus dem 2. HPC Call (2008-2011)

ELPA-Bibliothek Open Source seit Mai 2011

Seit Projektende Nov. 2011 Weiterpflege in Eigenregie

ELPA-Bedeutung:

*state-of-the-art* Eigenlöser direkter Art für symmetrische dichte und Band-  
strukturierte Matrizen auf massiv-parallelen Rechnern

- Nutzung von ELPA an größten wiss. Rechenzentren  
(wie FZJ Jülich, ORNL, ANL (USA))
- Einsatz weltweit in wichtigen Simulationspaketen für Elektronen-  
strukturberechnungen für die Materialforschung
- Mitverteilt in Linux-Distributionen von OpenSUSE, Debian und Fedora



ELSEVIER

Contents lists available at ScienceDirect

Parallel Computing

journal homepage: [www.elsevier.com/locate/parco](http://www.elsevier.com/locate/parco)



## Parallel solution of partial symmetric eigenvalue problems from electronic structure calculations <sup>☆</sup>

T. Auckenthaler<sup>a</sup>, V. Blum<sup>b</sup>, H.-J. Bungartz<sup>a</sup>, T. Huckle<sup>a</sup>, R. Johanni<sup>c</sup>, L. Krämer<sup>d</sup>, B. Lang<sup>d,\*</sup>, H. Lederer<sup>c</sup>, P.R. Willems<sup>d</sup>

<sup>a</sup>Fakultät für Informatik, Technische Universität München, D-85748 Garching, Germany

<sup>b</sup>Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany

<sup>c</sup>Rechenzentrum Garching der Max-Planck-Gesellschaft am Max-Planck-Institut für Plasmaphysik, D-85748 Garching, Germany

<sup>d</sup>Fachbereich C, Bergische Universität Wuppertal, D-42097 Wuppertal, Germany

### ARTICLE INFO

#### Article history:

Available online 14 May 2011

#### Keywords:

Electronic structure calculations  
Eigenvalue and eigenvector computation  
Blocked Householder transformations  
Divide-and-conquer tridiagonal eigensolver  
Parallelization

### ABSTRACT

The computation of selected eigenvalues and eigenvectors of a symmetric (Hermitian) matrix is an important subtask in many contexts, for example in electronic structure calculations. If a significant portion of the eigensystem is required then typically direct eigensolvers are used. The central three steps are: reduce the matrix to tridiagonal form, compute the eigenpairs of the tridiagonal matrix, and transform the eigenvectors back. To better utilize memory hierarchies, the reduction may be effected in two stages: full to banded, and banded to tridiagonal. Then the back transformation of the eigenvectors also involves two stages. For large problems, the eigensystem calculations can be the computational bottleneck, in particular with large numbers of processors. In this paper we discuss variants of the tridiagonal-to-banded back transformation, improving the parallel efficiency for large numbers of processors as well as the per-processor utilization. We also modify the divide-and-conquer algorithm for symmetric tridiagonal matrices such that it can compute a subset of the eigenpairs at reduced cost. The effectiveness of our modifications is demonstrated with numerical experiments.

© 2011 Elsevier B.V. All rights reserved.

> 100 Zitate

## Topical Review

# The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science

A Marek<sup>1</sup>, V Blum<sup>2,3</sup>, R Johanni<sup>1,2</sup>, V Havu<sup>4</sup>, B Lang<sup>5</sup>, T Auckenthaler<sup>6</sup>,  
A Heinecke<sup>6</sup>, H-J Bungartz<sup>6</sup> and H Lederer<sup>1</sup>

<sup>1</sup> Rechenzentrum Garching der Max-Planck-Gesellschaft am Max-Planck-Institut für Plasmaphysik, D-85748 Garching, Germany

<sup>2</sup> Fritz Haber Institute of the Max Planck Society, D-14195 Berlin, Germany

<sup>3</sup> Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA

<sup>4</sup> COMP, Department of Applied Physics, Aalto University, FI-00076 Aalto, Finland

<sup>5</sup> Fachbereich C, Bergische Universität Wuppertal, D-42097 Wuppertal, Germany

<sup>6</sup> Fakultät für Informatik, Technische Universität München, D-85748 Garching, Germany

E-mail: [amarek@rzg.mpg.de](mailto:amarek@rzg.mpg.de) and [volker.blum@duke.edu](mailto:volker.blum@duke.edu)

Received 18 January 2014, revised 24 February 2014

Accepted for publication 25 February 2014

Published 2 May 2014

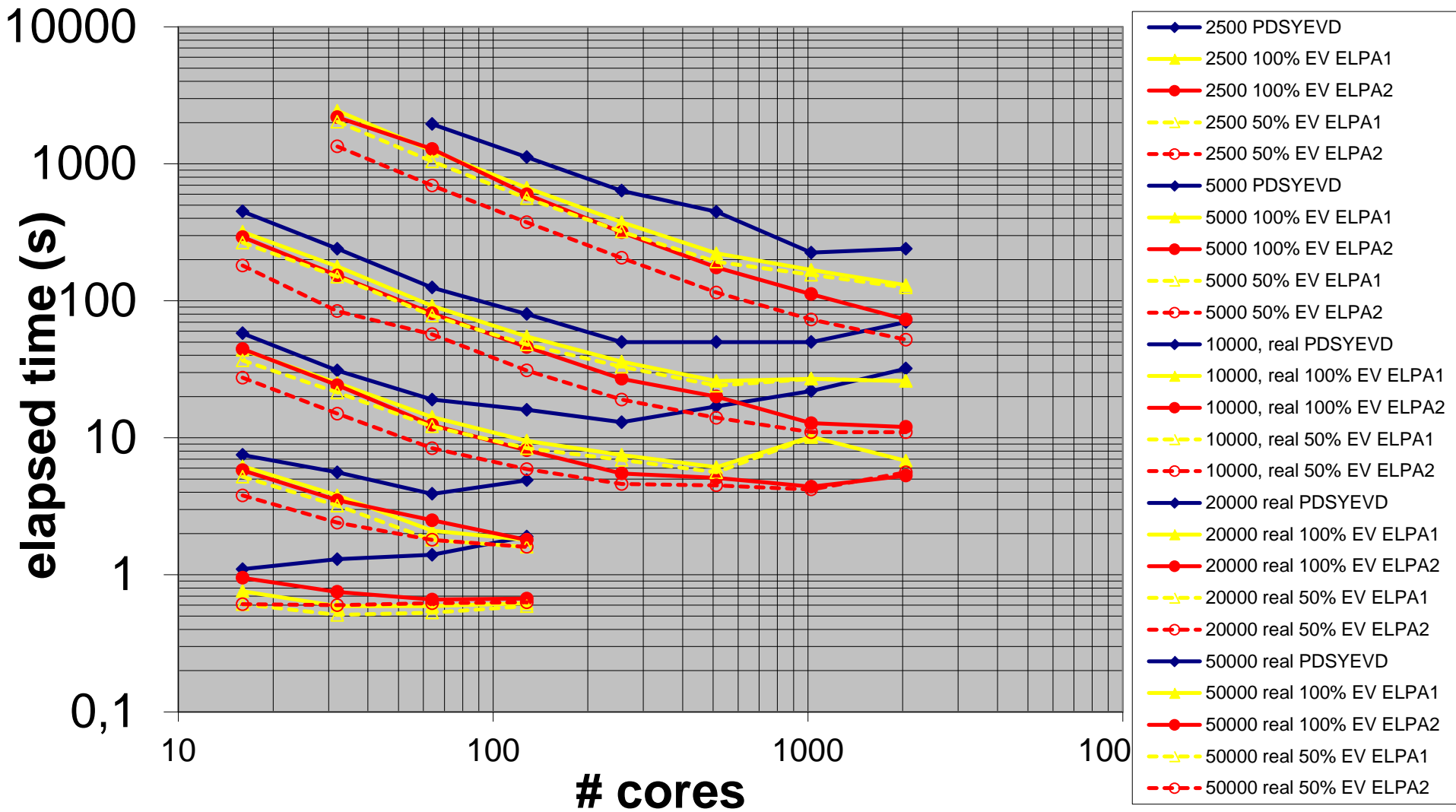
### Abstract

Obtaining the eigenvalues and eigenvectors of large matrices is a key problem in electronic structure theory and many other areas of computational science. The computational effort formally scales as  $O(N^3)$  with the size of the investigated problem.  $N$  (e.g. the electron count in electronic

> 25 Zitate



# Benchmarks zu ELPA1, ELPA2 (2012)



## Wichtige neue Aspekte

Vergrößerung eines Hochleistungsrechners führt in der Regel zur Steigerung des Energieverbrauchs

Gegenmaßnahme: Verbesserung der Effizienz eingesetzter Software sowohl in Richtung „Time-to-solution“ wie auch „Energy-to-solution“

# Inhaltliche Abgrenzung zum ELPA-Vorhaben (2008-2011)

## Wesentliche Gesichtspunkte:

Leistungssteigerung für bereits vorhandene Funktionalität

- u.a. für die Reduktion des verallgemeinerten auf das Standard-Eigenwertproblem
- für weitere besonders rechenintensive Schritte durch Optimierung auf inzwischen oder bald verfügbare Architekturen.

Aufnahme weiterer Basis-Funktionalität in die ELPA-Bibliothek

- u.a. direkte und iterative Verfahren für das verallgemeinerte Eigenwertproblem mit Bandstruktur
- Varianten mit reduzierter (Rechen-)Genauigkeit.

# Strukturierung des Projektes

**1 – Algorithmische Entwicklungen**

**2 – Effiziente Parallelisierung**

**3 – Implementierung, Portierung und Optimierung**

**4 – Autotuning**

**5 – Anwendungen I**

Kontext: Oberflächen und fest-feste Grenzflächen (z.B. Thermoelektrik)

**6 – Anwendungen II**

Kontext: fest-flüssige und flüssig-flüssige Grenzflächen (z.B. Photo-Elektro-Katalyse)

# 1 – Algorithmische Entwicklungen

- Entwicklung eines direkten Eigenlösers für verallgemeinerte Eigenprobleme mit Bandstruktur
- Untersuchungen zur Eignung (und ggf. Entwicklung) eines iterativen Eigenlösers für verallgemeinerte Eigenprobleme mit Bandstruktur
- Entwicklung von numerischen Techniken zur Reduktion des Aufwandes bei verringerten Genauigkeitsanforderungen und zur Herstellung approximativer Bandstruktur
- Entwicklung von Strategien zur Reduktion des Aufwandes bei Sequenzen von Eigenwertproblemen
- Identifikation hocheffizienter Varianten der Transformation auf das Standard-Eigenproblem
- Anwendungsspezifische Weiterentwicklung von Methoden zur Reduktion der approximativen Bandstruktur (Basissatz-Lokalisierung)
- Entwicklung und Adaption von Prädiktionierungs- und Stabilisierungsalgorithmen für das nicht-lineare Problem (SCF).

## 2 – Effiziente Parallelisierung

- Effiziente parallele Umsetzung der unter 1 entwickelten Algorithmen und Strategien
  - insbesondere Rücktransformation der Eigenvektoren bei Band-strukturierten Problemen und
  - Aufrechterhaltung der Orthogonalität vieler Vektoren im iterativen Eigenlöser.

## 3 – Implementierung, Portierung und Optimierung

- Implementierungen und Optimierungen der Ergebnisse von 2
- Portierung und Optimierung von ELPA-Routinen bzgl. neuer Architekturen
- Berücksichtigung der Rückmeldungen zu Qualität, Stabilität und Gesamteffizienz durch die Anwender (5 und 6)

## 4 – Autotuning

- Konzeptionierung und Umsetzung eines Monitoring-Tools für das Verhalten der verschiedenen eingesetzten Komponenten
- Automatisches ELPA Autotuning
- Definition und Implementierung eines APIs für das Reporting an das rufende Hauptprogramm sowie für manuelle Vorgaben nach Erfordernissen der Anwendungen (5 und 6).
- Identifikation der notwendigen Genauigkeitsgrenzen für die Anwendungen I und II
- Entwicklung semi-manueller Tuningstrategien auf Anwendungsebene

## 5 – Anwendungen I

- Bereitstellung von sehr großen Beispielmatrizen und relevanter Anwendungsprobleme zu **Oberflächen und fest-festen Grenzflächen** (z.B. Thermoelektrik)
- Integration der neuen Algorithmen in die SCF (*Self Consistent Field*) - und ab-initio Molekulardynamik-Methoden im Simulationspaket FHI-aims, gemeinsam durch FHI-TH und TUM-CH
- Monitoring von Qualität, Stabilität und Gesamteffizienz der neu entwickelten Algorithmen im Anwendungskontext von FHI-TH
- Tier-0 Benchmark-Simulation für wissenschaftlichen Impact



## 6 – Anwendungen II

- Bereitstellung von Sequenzen von Beispielmatrizen und relevanter Anwendungsprobleme zu **fest-flüssigen und flüssig-flüssigen Grenzflächen** (z.B. Photo-Elektro-Katalyse)
- Entwicklung einer effizienten Ausgabe der Matrizen entlang von SCF (*Self Consistent Field*)-Zyklen und AIMD (*ab-initio Molekulardynamik*)-Schritten im Simulationspaket FHI-aims
- Monitoring von Qualität, Stabilität und Gesamteffizienz der neu entwickelten Algorithmen im Anwendungskontext von TUM-CH
- Tier-0 Benchmark-Simulation für wissenschaftlichen Impact

# Erste Ergebnisse

## **Prototyp für direkten Eigenlöser für verallgemeinertes Band-Eigenwertproblem ohne Berechnung der Eigenvektoren**

- Neues Verfahren wurde entwickelt (BUW)
- Parallelisierung bereits in Arbeit (TUM-SCCS)

# Erste Ergebnisse

## Portierung und Optimierung von ELPA für eine neue HPC-Architektur:

- NVIDIA Tesla GPGPUs (K20, K40, K80, P100):  
GPU Version wurde entwickelt  
Getestet und einsetzbar auf Intel Xeon + NVIDIA TESLA
- Intel Xeon Phi KNL: ELPA bereits erfolgreich eingesetzt,  
Performance-Optimierung ist in Gange
- OpenPower Architektur:  
Portierung auf Minsky begonnen  
(Minsky-Knoten: 2 CPUs IBM Power8 + 4 GPUs NVIDIA P100)

# Erste Ergebnisse

## Neues ELPA-Release vom 25.11.2016

- Enthält Unterstützung für NVIDIA Tesla GPGPUs
- Enthält zusätzlich 32 Bit Versionen von elpa1 und elpa2 sowohl für GPUs wie auch X86 Architektur
  - > signifikante Performance-Steigerung
  - > insbesondere für jene Anwendungen, die selbst nur in 32 Bit rechnen

# Automatisches ELPA Autotuning

**Ziel** bei vorgegebener Problemstellung und Hardwarekonfiguration:  
Dem Benutzer soll die ansonsten komplizierte Wahl des effizientesten Löser-Verfahrens erleichtert bzw. abgenommen werden.

## Vorgehen bei vorgegebener Problemstellung

(Matrixgröße und Anzahl der zu berechnenden Eigenwerte):

- Vergleich der unterschiedlichen Verfahren
  - einstufiger Löser "elpa1"
  - zweistufiger Löser "elpa2"
- Identifikation des effizienteren Verfahrens  
Beim komplexeren elpa2-Verfahren:
  - Optimierung hinsichtlich der verschiedenen rechenintensiven Kernel
- Falls Zusatzhardware wie GPGPUs vorhanden:
  - Einbeziehung derselben in die Auswahlkriterien
- Erweiterung des Autotuning-Verfahrens auf entstehende Neuerungen durch ELPA-AEO

# ELPA Autotuning - Design

Automatische Ermittlung des effizientesten Löser-Verfahrens bei vorgegebener Problemstellung und Hardwarekonfiguration

Optionen:

- Einstufiger Löser elpa1
- Zweistufiger Löser elpa2; hierfür ca. 15 verschiedene rechenintensive Kernel vorhanden
- Einbeziehung von Zusatzhardware (z.B. GPGPUs)

Main Program

..

Call ELPA (param1, param2, ..., e)

Bisher:  $e \in \{\text{elpa1}, \text{elpa2}\}$  (einzustellen durch Benutzer aufgrund von Einschätzung)

**Neu:**  $e \in \{\text{elpa1}, \text{elpa2}, \text{auto}\}$  (Erweiterung des Interfaces, back-kompatibel)

Geplante Umsetzung falls  $e = \text{auto}$ :

Zuerst werden elpa 1 und elpa 2 (mit Default-Kernel für gegebene Architektur) verwendet und vermessen.

In den nachfolgenden Iterationen wird automatisch die schnellere Variante verwendet.