# Analytical PAW Projector Functions for Reduced Bandwidth Requirements

Paul F. Baumeister
Jülich Supercomputing Centre
Forschungszentrum Jülich
Jülich, Germany
p.baumeister@fz-juelich.de

Shigeru Tsukamoto
Peter-Grünberg Institut
Forschungszentrum Jülich
Jülich, Germany

## ABSTRACT

Large scale electronic structure calculations require modern high performance computing (HPC) resources and, as important, mature HPC applications that can make efficient use of those. Real-space grid-based applications of Density Functional Theory (DFT) using the Projector Augmented Wave method (PAW) can give the same accuracy as DFT codes relying on a plane wave basis set but exhibit an improved scalability on distributed memory machines. The projection operations of the PAW Hamiltonian are known to be the performance critical part due to their limitation by the available memory bandwidth. We investigate on the utility of a 3D factorizable basis of Hermite functions for the localized PAW projector functions which allows to reduce the bandwidth requirements for the grid representation of the projector functions in projection operations. Additional on-the-fly sampling of the 1D basis functions eliminates the memory transfer almost entirely. For an quantitative assessment of the expected memory bandwidth savings we show performance results of a first implementation on GPUs. Finally, we suggest a PAW generation scheme adjusted to the analytically given projector functions.

## KEYWORDS

Density Functional Theory, Projector Augmented Wave method, Real-Space grid, Cartesian grid, Radial grid, Many-core, GPUs

**ACM Reference Format:**
Paul F. Baumeister and Shigeru Tsukamoto. 2019. Analytical PAW Projector Functions for Reduced Bandwidth Requirements. In *Proceedings of the Platform for Advanced Scientific Computing Conference (PASC '19), June 12–14, 2019, Zurich, Switzerland.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3324989.3325717

## 1 INTRODUCTION

Uniform real-space grids have proven to be equivalent to plane wave (PW) representations for electronic structure calculations in the framework of Density Functional Theory (DFT) [8, 12]. The singularity of the nuclear potential can be represented accurately on radial grids centered at the atomic site. Furthermore, rapid oscillations of the quantum mechanical wave functions caused by the deep attractive potential can be resolved. The information transfer from 3D Cartesian grids to radial grids and back is a key ingredient of electronic structure calculations. Besides other approaches, an accurate treatment of the singularities of the electronic potential at the positions of the nuclei is enabled by the Projector Augmented Wave (PAW) method [2].

The real-space grid approach, pioneered by Briggs *et al.* [4], is the preferred approach for large scale DFT problems on supercomputers as it leads to very sparse matrix representations of the Kohn-Sham Hamiltonian and is free of global Fourier transforms. These properties allow an improved scalability on parallel computers compared to calculations involving a PW basis. Most implementations of real-space DFT sample the projector functions on a real-space grid and store these in memory, as implemented for frozen-core PAW [17, 19] or using pseudopotentials [9, 16], a limit case of PAW.

Then, the calculation of inner products between localized projector functions and fully extended Kohn-Sham wave functions requires projection operations which are typically bounded by the memory bandwidth (BW) of the computing device. Comparing to the peak achievable rates of floating-point operations, memory BW has become a scarce resource on modern high-performance computing systems.
In this work, we suggest analytical shapes for the PAW projectors which can lower the memory BW requirements for projection and expansion operations in implementations of the PAW and pseudopotential methods.
In this paper we make the following contributions:

- We introduce the Spherical Harmonic Oscillator (SHO) as a natural link between 3D Cartesian grids and radial grids.
- We assess the potential to accelerate electronic structure calculations using SHO eigenstates.
- We analyse the quality of existing PAW projectors when represented in a SHO basis.
- We suggest a new PAW generation method adjusted to the analytical projector shapes.

The rest of this paper is structured as follows. In sec. 2 we introduce general properties of the SHO basis and basics of DFT calculations. In sec. 3 we outline the application of a SHO basis for PAW projector functions under aspects of high-performance computing. In sec. 4 the quality of a SHO basis expansion for existing PAW data sets is assessed. We suggest a modification to the PAW generation scheme in sec. 5. In sec. 6 expectable performance improvements on graphics processors (GPUs) are assessed. Finally, sections 7, 8


*PASC '19, June 12–14, 2019, Zurich, Switzerland*

ACM ISBN 978-1-4503-6770-7/19/06…$15.00
https://doi.org/10.1145/3324989.3325717

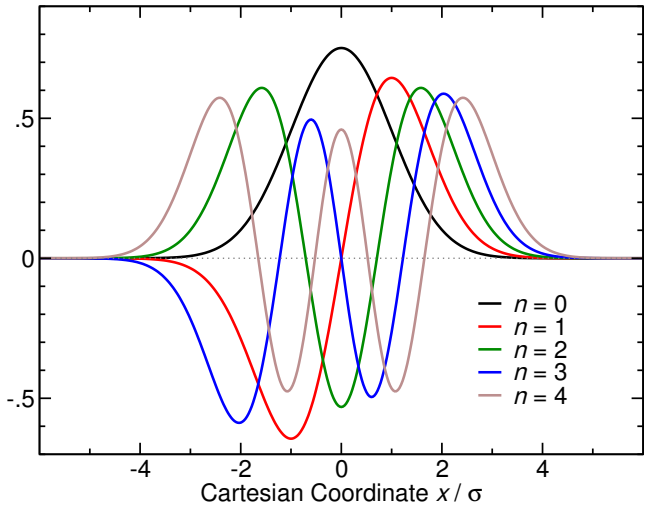


**Figure 1: The eigenstates of the 1D harmonic oscillator are Hermite polynomials times a Gaussian, see eq. (3).**

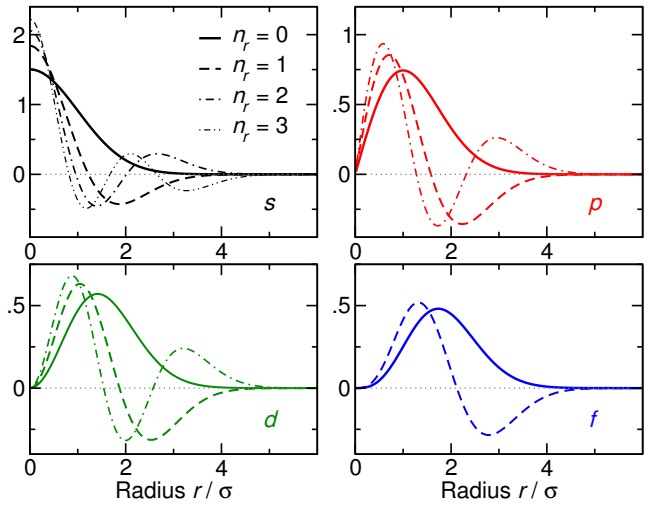


**Figure 2: Radial part of the lowest eigenstates of the spherical harmonic oscillator up to $\nu_{\max} = 6$ for $\ell \in [0, 3]$.**

and 9 contain a discussion of this approach, related works and a summary, respectively.

Throughout the paper we use atomic Rydberg units: $m_e = 1/2$, $\hbar = 1$ and $e^2 = 2$, i.e. length is measured in Bohr (≡ .529 Å) and the unit of energy is Rydberg (≡ 13.6 eV).

## 2 THEORY

### 2.1 Spherical Harmonic Oscillator

The quantum-mechanical Spherical Harmonic Oscillator (SHO)

$$\hat{H}_{\text{SHO}} = -\vec{\nabla}^2 + \vec{r}^2\sigma^{-4} \tag{1}$$

is a well-understood problem, often referred to as three-dimensional isotropic harmonic oscillator. Here, $\sigma$ is a length scale parameter in Bohr units which also defines the energy scale. The 3D partial differential equation (PDE) can be solved using separation of variables, either in the set of Cartesian variables $(x, y, z) \equiv \vec{r}$ or, due to the spherical symmetry of the potential term, using spherical coordinates $r$, $\vartheta$ and $\varphi$.

### 2.2 Cartesian Coordinates

For the Cartesian approach it is sufficient to find the solutions of the 1D harmonic oscillator Hamiltonian

$$\hat{H}_{\text{1D}} = -\frac{\partial^2}{\partial x^2} + x^2\sigma^{-4} \tag{2}$$

which leads to the eigenenergy $E_{\text{1D}} = (2n{+}1)\sigma^{-2}$ with the quantum number $n \in \mathbb{N}_0$. The corresponding eigenstates $|\psi_n\rangle$ are given by Hermite polynomials times a Gaussian envelope function:

$$\psi_n(x) \propto H_n\left(\frac{x}{\sigma}\right)\, \exp(-\frac{x^2}{2\sigma^2}) \tag{3}$$

where $H_n(x)$ are Hermite polynomials. Proper normalizing factors are omitted here for simplicity. The lowest five normalized Hermite functions are shown in fig. 1.

For the 3D problem, three independent quantum numbers $(n_x, n_y, n_z) \in \mathbb{N}_0^3$ and the eigenenergy

$$E_{\text{SHO}}\,\sigma^2 = 2(n_x + n_y + n_z) + 3 := 2\nu + 3 \tag{4}$$

are found. The eigenstates factorize in the three spatial dimensions:

$$\Psi_{n_x n_y n_z}(x, y, z) = \psi_{n_x}(x)\,\psi_{n_y}(y)\,\psi_{n_z}(z) \tag{5}$$

$$= H_{n_x}\left(\frac{x}{\sigma}\right)\, H_{n_y}\left(\frac{y}{\sigma}\right)\, H_{n_z}\left(\frac{z}{\sigma}\right)\, \exp(-\frac{\vec{r}^2}{2\sigma^2}) \tag{6}$$

### 2.3 Spherical Coordinates

The PDE in spherical coordinates $(r, \vartheta, \varphi)$ leads to the radial quantum number $n_r \in \mathbb{N}_0$, the angular momentum character $\ell \in \mathbb{N}_0$ and the magnetic quantum number $m \in [-\ell, \ell] \cap \mathbb{Z}$:

$$\Psi_{n_r \ell m}(r, \vartheta, \varphi) = R_{n_r \ell}\left(\frac{r}{\sigma}\right)\, Y_{\ell m}(\vartheta, \varphi) \tag{7}$$

Here, $Y_{\ell m}$ denotes spherical harmonic functions. The eigenstates of the differential equation for the radial coordinate $r$ are

$$R_{n_r \ell}(r) \propto r^\ell L_{n_r}^{(\ell+\frac{1}{2})}(r^2)\, \exp(-\frac{1}{2}r^2) \tag{8}$$

where $L_n^{(\alpha)}$ are associated Laguerre polynomials. See fig. 2 for the set of normalized radial functions $R_{n_r \ell}(r)$. Due to the spherical symmetry of $\hat{H}_{\text{SHO}}$ the eigenenergy does not depend on $m$:

$$E_{\text{SHO}}\,\sigma^2 = 2(\ell + 2n_r) + 3 := 2\nu + 3 \tag{9}$$

An energy diagram for $n_r$, $\ell$ and $\nu$ is shown in fig. 3.

### 2.4 Connection of Subspaces

In eq. (9) and eq. (4), the energy quantum number $\nu \in \mathbb{N}_0$ was introduced. As both approaches solve the same problem the $\frac{1}{2}(\nu + 2)(\nu + 1)$ energy-degenerate solutions in each subspace indexed by $\nu$ must span the same functions space. In fact, it is possible to find

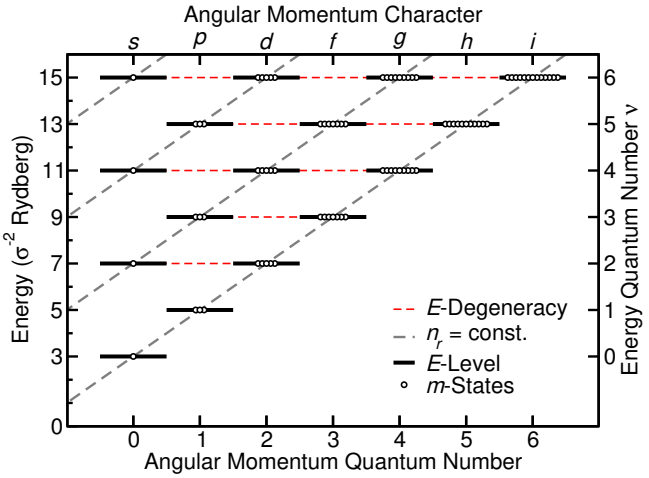


**Figure 3: Energy levels of the spherical harmonic oscillator. The dashed gray diagonals correspond to a constant number of radial nodes, $n_r$. The lowest diagonal ($n_r = 0$) corresponds to Gaussian type orbitals. The red dashed horizontal lines indicate energy-degenerate subspaces indexed by $\nu$.**

a unitary transform $\hat{U}$ such that

$$\Psi_{n_x n_y n_z}(x,y,z) \equiv \sum_{n_r \ell m} U^{n_r \ell m}_{n_x n_y n_z}\, \Psi_{n_r \ell m}(r,\vartheta,\varphi). \tag{10}$$

Matrix elements of $U^{n_r \ell m}_{n_x n_y n_z}$ are only non-zero if

$$\ell + 2n_r = \nu = n_x + n_y + n_z. \tag{11}$$

Furthermore, even within each $\nu$-block many entries of $\hat{U}$ vanish due to other symmetries.

## 2.5 SHO Basis Set

The full set of SHO eigenstates forms a basis of smooth functions defined on $\mathbb{R}^3 \mapsto \mathbb{C}$. Therefore, for both representations, radial and Cartesian, the completeness relations

$$1 = \sum_{n_r \ell m}^{\infty} |n_r \ell m\rangle \langle n_r \ell m| \qquad \text{and} \tag{12}$$

$$1 = \sum_{n_x n_y n_z}^{\infty} \left|n_x n_y n_z\right\rangle \left\langle n_x n_y n_z\right| \tag{13}$$

must hold. Later, we will exploit that

$$1 = \sum_{n_r \ell m}^{\infty} |n_r \ell m\rangle \langle n_r \ell m| \sum_{n_x n_y n_z}^{\infty} \left|n_x n_y n_z\right\rangle \left\langle n_x n_y n_z\right| \tag{14}$$

$$= \sum_{n_r \ell m}^{\infty} |n_r \ell m\rangle \sum_{n_x n_y n_z}^{\infty} \left\langle n_r \ell m | n_x n_y n_z\right\rangle \left\langle n_x n_y n_z\right| \tag{15}$$

$$= \sum_{n_r \ell m}^{\infty} \sum_{n_x n_y n_z}^{\infty} |n_r \ell m\rangle\, U^{n_r \ell m}_{n_x n_y n_z} \left\langle n_x n_y n_z\right| := \hat{\mathcal{B}}_{\infty}. \tag{16}$$

Truncating the series at a given maximal energy $E_{\max} = (2\nu_{\max}+3)\sigma^{-2}$ will break the completeness relation, $\hat{\mathcal{B}}_{\nu_{\max}} \neq 1$. However, through the energy cut-off criterion it is ensured that

$$\sum_{n_r \ell m}^{2n_r+\ell \leq \nu_{\max}} |n_r \ell m\rangle \langle n_r \ell m| = \sum_{n_x n_y n_z}^{n_x+n_y+n_z \leq \nu_{\max}} \left|n_x n_y n_z\right\rangle \left\langle n_x n_y n_z\right| \tag{17}$$

are both projection operators describing the same subspace of functions. The size of a finite SHO basis is

$$N_{\text{SHO}} = \frac{1}{6}(\nu_{\max}+3)(\nu_{\max}+2)(\nu_{\max}+1). \tag{18}$$

## 2.6 Electronic Structure Methods

In the effective potential landscape arising from DFT two major regions can be identified: core and interstitial. Close to an atomic *core* the attractive Coulomb potential dominates. Here, spherical coordinates $(r,\vartheta,\varphi)$ are best to solve the Kohn-Sham equation. Radial grids densely sample the radial coordinate $r$, typically in a logarithmic fashion. All angular dependencies are expanded in spherical harmonics $Y_{\ell m}(\vartheta,\varphi)$. In the *interstitial* region the potential is strongly determined by the geometry of neighboring atoms and their ionic character. Here, we have to solve a 3D PDE in order to find valence states that describe chemical bonding.

There is a large variety of DFT methods that differ either by the basis set used for the 3D problem or by the way these basis functions are connected to the radial grids of each atom.

DFT methods based on a 3D plane wave (PW) basis set can make use of

$$\exp(\imath\vec{k}\cdot\vec{r}) = 4\pi \sum_{\ell=0}^{\infty} \imath^{\ell} \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\hat{\vec{k}})\, Y_{\ell m}(\hat{\vec{r}})\, j_{\ell}(|\vec{k}|\cdot|\vec{r}|) \tag{19}$$

which yields a natural transition from Cartesian space $\vec{r} \in \mathbb{R}^3$ to radial coordinates $r = |\vec{r}|$ via spherical Bessel functions $j_\ell(x)$ and spherical harmonics $Y_{\ell m}(\vartheta,\varphi)$. Methods exploiting this equation directly are referred to as Augmented Plane Wave (APW) methods and implementations of e.g. the (full-potential) linearized APW methods are often considered the most accurate electronic structure codes [14, 24]. However, a basis consisting of $N_{\text{PW}}$ plane waves leads to a dense matrix representation of the Kohn-Sham Hamiltonian requiring $O(N^2_{\text{PW}})$ memory and $O(N^3_{\text{PW}})$ operations to diagonalize them. As this easily exceeds the system capacities of a single compute node, in particular memory capacity, we have to consider the efficiency of parallel eigensolvers for dense problems which typically require intensive all-to-all communication.

A vast number of DFT calculations are performed on PW methods using Blöchl's PAW method [2], e.g. [6, 13, 27]. PW-PAW can also make use of eq. (19) for the calculation of the overlap between the PWs and PAW projector functions localized in real-space. As this also leads to a large dense Hamiltonian matrix, the sparse representation of the PAW Hamiltonian in real-space is preferred in the limit of large problems.

Switching between a real-space grid and a PW representation of the wave functions demands for the application of global 3D Fourier Transforms (FFTs) which are strongly memory-bound operations and limit the scalability on distributed memory machines.

Real-space grid DFT is, by construction, free of FFTs and can be combined with the PAW method equally well. Therefore, it is the approach most suited for supercomputing.

# 3 SAVING BANDWIDTH

## 3.1 High-Performance Computing

The expected availability of Exascale supercomputers gives hope for the feasibility of large-scale calculations for real materials using DFT and related methods. However, conventional representations and solver algorithms exhibit all-to-all communication patterns which do not scale well on distributed memory architectures. Real-space grid based representations have shown to scale efficiently on large allocations of HPC machines. Using domain decomposition, mainly nearest-neighbor inter-process communication in a periodic 3D topology is required. This allows to parallelize work with large numbers of tasks efficiently. However, the evolution of HPC compute nodes undergoes a strong change from scalar CPUs towards many-core architectures featuring wide vector units. In particular, modern GPUs offer very high floating-point operation performances compared to their power consumption and can, therefore, be found in many of the worlds largest HPC machines[1]. The increase of parallelism due to more cores, wider vectors, increased instruction level parallelism and extended capabilities for simultaneous multithreading lets the processing power in terms of arithmetic operations grow faster over years than technology permits for the memory BW. Despite the advent of new memory technologies, BW has become one of the most costly resources for many applications in the field of computational science that cannot leverage usual cache optimizations. This leads to an urgent need for new bandwidth-saving algorithms.

## 3.2 Hamiltonian Action

For the real-space grid-based PAW approach, the performance of iterative eigensolver methods, like e.g. ChASE [28], relies on a computationally efficient implementation of the action of the Hamiltonian operator onto wave functions. The PAW Hamiltonian takes the form

$$\hat{H}_{\mathrm{PAW}} = \hat{T} + \tilde{V}_{\mathrm{eff}} + \sum_{aij} \left|\tilde{p}_i^a\right\rangle D_{ij}^a \left\langle\tilde{p}_j^a\right| \tag{20}$$

with the kinetic energy operator $\hat{T}$ and a smooth local effective potential $\tilde{V}_{\mathrm{eff}}(\vec{r})$. The largest fraction of compute time for the action of $\hat{H}_{\mathrm{PAW}}$ onto (preliminary) eigenvectors goes to the sum over the dyadic expressions $\left|\tilde{p}_i\right\rangle\left\langle\tilde{p}_j\right|$ for each atom $a$. When applying $\hat{H}_{\mathrm{PAW}}$ to a given set of independent wave functions, $|\Psi_k\rangle$, the application of the so-called *non-local* potential part is performed in three steps:

(1) Projection: $c_{jk}^a = \left\langle \tilde{p}_j^a \middle| \Psi_k \right\rangle$
(2) Multiplication: $d_{ik}^a = \sum_j D_{ij}^a \, c_{jk}^a$
(3) Expansion: add $\sum_{ai} d_{ik}^a \left|\tilde{p}_i^a\right\rangle$ to $(\hat{T} + \tilde{V}_{\mathrm{eff}})\,|\Psi_k\rangle$

In real-space PAW implementations, the projector functions need to be represented on the real-space grid. Following the original PAW scheme, these functions are given as

$$\tilde{p}_i^a(\vec{r}) = P_{n\ell}^a(|\vec{r}_a|) Y_{\ell m}(\hat{\vec{r}}_a) \tag{21}$$

where $\vec{r}_a = \vec{r} - \vec{R}_a$ for each atom $a$ and the index $i$ encodes $(n, \ell, m)$. In contrast to the spherical harmonics $Y_{\ell m}$ the radial parts $P_{n\ell}^a(r)$ are given numerically and usually depend only on the atom species of $a$. In the spirit of linearized APW methods, typical projector set sizes are limited to two projectors per $\ell$ for $s$-, $p$-, and $d$-projectors [10]. Hence, a typical setup used for calculations involving transition metal atoms comprises six different radial functions $P_{n\ell}^a(r)$ which translates to 18 projector functions $\tilde{p}_i^a(\vec{r})$ per atom. The latter number defines the index range for $i$ and $j$ in eq. (20). Light elements typically feature less projectors. For rare earth elements, adding $f$-projectors is mandatory.

[1] top500.org

## 3.3 Projector Filtering

Following the original PAW methodology [2], projector functions are supposed to be strictly confined to the inside of a sphere of radius $R_{\mathrm{aug}}$ around the atomic position $\vec{R}_a$. For clarity of the notation we suppress the atom index $a$ in the rest of sec. 3. The atomic spheres are constructed to be non-overlapping so that $R_{\mathrm{aug}}$ is usually chosen as half the nearest-neighbor distance (touching spheres). This hard localization constraint leads to an unphysical aliasing effect (*egg box effect*). A spurious dependency of the classical DFT observables (total energy and atomic forces) on the relative position of an atom w.r.t. the positions of grid points can be measured. Removing high frequency components from $\tilde{p}_i(\vec{r})$ mitigates this effect and various methods have been proposed to filter the projector functions in reciprocal space. Soler and Anglada [1, 25] summarize the problem of finding the best localization in both, real-space and reciprocal space.

Many implementations of real-space grid-based DFT make use of the filtering recipe by Tafipolsky and Schmid [26] where a Gaussian-shaped mask function restores the localization in real-space after suppressing high frequency components in the Bessel-transformed representation of the radial part of a projector function.

Only the *double-grid technique* by Ono *et al.* [20] applies a filter to the product $P_{n\ell}(|\vec{r}|)Y_{\ell m}(\hat{\vec{r}})$ in 3D space which implies that also high frequency components that stem from the spherical harmonics $Y_{\ell m}$ are successfully suppressed.

Nevertheless, one effect of filtering is an increased projection radius, $R_{\mathrm{prj}} > R_{\mathrm{aug}}$, around the atomic position where the representation of the projector functions on the grid is non-zero. This strongly adds to the cost of the Hamiltonian action proportional to $(R_{\mathrm{prj}}/R_{\mathrm{aug}})^3$.

To summarize: for reasons of costs it is essential to preserve the localization of PAW projector functions in real-space. For reasons of accuracy also localization in reciprocal space is mandatory.

## 3.4 Data Compression

One goal is to make the Hamiltonian action perform as efficient as possible on current HPC compute systems, i.e. vectorized many-core architectures. The projection operation (1) and its counterpart, the expansion (3), can be written as matrix-matrix multiplications. Here, one operand is a sparse matrix since the projector functions are non-zero only close to the atom. The sparsity leads to a strong limitation of the performance by the memory BW.

A major strategy for the reduction of BW requirements is to decrease the communication volume, i.e. by data compression. A projector functions represented on a uniform Cartesian real-space grid with grid points at $\vec{j} \in \mathbb{N}^3$ can be viewed as rank-3 tensor $P_{j_x j_y j_z}$. Here, tensor compression methods as studied intensively by Khoromskaia and Khoromskij [11] try to approximate high-rank

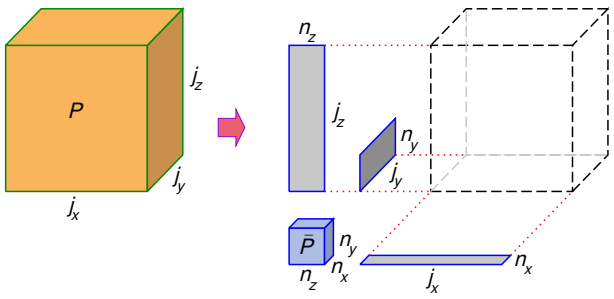


**Figure 4: Rank-3 Tucker tensor decomposition.** $P_{j_x j_y j_z}$ **is compressed into** $v^{[x]}_{n_x j_x}$, $v^{[y]}_{n_y j_y}$, $v^{[z]}_{n_z j_z}$ **and** $\bar{P}_{n_x n_y n_z}$, **see eq. (22).**

tensors by sums of dyadic products. In this particular case we seek a representation

$$P_{j_x j_y j_z} \approx \sum_{n_x n_y n_z} \bar{P}_{n_x n_y n_z}\, v^{[x]}_{n_x j_x}\, v^{[y]}_{n_y j_y}\, v^{[z]}_{n_z j_z} \tag{22}$$

as depicted in fig. 4. If it is possible to find such an approximation with reduced ranges of the indices $(n_x, n_y, n_z)$ compared to the ranges of $(j_x, j_y, j_z)$ then $\bar{P}$ is a rank-3 tensor of much lower volume. Furthermore, the total amount of memory needed for loading the rectangular matrices $v^{[x|y|z]}$ in addition to loading $\bar{P}$ can be lower than the total volume of $P$, depending on the original ranges and its compressibility. In particular if we consider a set of projectors, as usual in PAW, that can make shared use of the same 1D function sets $v^{[x|y|z]}$, we expect an overall good compression ratio.

Transferring this scheme to the factorizable SHO basis, eq. (5), we are seeking an approximation of the projector set $\tilde{p}_i(\vec{r})$ of each atom as

$$\tilde{p}_i(x,y,z) \approx \sum_{n_x n_y n_z} \bar{p}_{i\, n_x n_y n_z}\, \psi_{n_x}(x)\, \psi_{n_y}(y)\, \psi_{n_z}(z) \tag{23}$$

with small ranges for $(n_x, n_y, n_z)$.

Dealing with compressed representations, an additional decompression operation is necessary. However, executing a task that is limited by the memory BW usually implies idle arithmetic units in the processor, so that the decompression phase is expected not add to the total execution time.

## 3.5 SHO Basis

Consider injecting the SHO basis projector $\hat{\mathcal{B}}_{\nu_{\max}}$ as defined in eq. (17), but for a finite $\nu_{\max}$ and given $\sigma$, into the projection operation: $\left\langle \tilde{p}_j \middle| \hat{\mathcal{B}}_{\nu_{\max}} \middle| \Psi_k \right\rangle$. For simplicity of the notation we suppress the atom index $a$ here. Then, we can evaluate the inner products

$$\left\langle \tilde{p}_j \middle| n_r \ell m \right\rangle \coloneqq F_{j n_r \ell m} \tag{24}$$

on radial grids during the initialization phase of the DFT calculation. The operation left to be performed every time is a projection onto the Cartesian factorizable SHO basis introduced in sec. 2.1:

$$C_{n_x n_y n_z k} = \left\langle n_x n_y n_z \middle| \Psi_k \right\rangle \tag{25}$$

$$= \iiint_{\mathbb{R}^3} \mathrm{d}^3\, \vec{r}\, \psi_{n_x}(x)\, \psi_{n_y}(y)\, \psi_{n_z}(z)\, \Psi_k(x,y,z) \tag{26}$$

In order to retrieve the original projection coefficients $c_{jk}$ we can transform the new coefficients $C_{n_x n_y n_z k}$ with $\left\langle \tilde{p}_j \middle| n_r \ell m \right\rangle \equiv \hat{F}$ and $\left\langle n_r \ell m \middle| n_x n_y n_z \right\rangle \equiv \hat{U}$. However, there is no need to retrieve these coefficients as these are temporary quantities existing only for the duration of projection and expansion operations. In fact, we can inject $\hat{\mathcal{B}}^{\dagger}_{\nu_{\max}}$ also into the expansion operation. Then, we collect all terms such that the original dyadic operator is fully transformed:

$$\left|\tilde{p}_i\right\rangle D_{ij} \left\langle \tilde{p}_j \right| \overset{\text{SHO}}{\longrightarrow} \hat{\mathcal{B}}^{\dagger}_{\nu_{\max}} \left|\tilde{p}_i\right\rangle D_{ij} \left\langle \tilde{p}_j \right| \hat{\mathcal{B}}_{\nu_{\max}} \tag{27}$$

$$= \left|n_x n_y n_z\right\rangle U^{n_x n_y n_z}_{n_r \ell m} F_{i n_r \ell m} D_{ij} F_{j n'_r \ell' m'} U^{n'_r \ell' m'}_{n'_x n'_y n'_z} \left\langle n'_x n'_y n'_z \right|$$

$$\coloneqq \left|n_x n_y n_z\right\rangle \mathcal{D}^{n_x n_y n_z}_{n'_x n'_y n'_z} \left\langle n'_x n'_y n'_z \right| \tag{28}$$

with contraction over adjacent indices (Einstein notation).
The atomic Hamiltonian term $\hat{D}$ is now replaced by $\hat{\mathcal{D}}$ which can be computed as $\hat{U}^{\dagger}\hat{F}^{\dagger}\hat{D}\hat{F}\hat{U}$ at initialization of each SCF cycle. $\hat{\mathcal{D}}$ represents a real-valued matrix of dimension $N_{\text{SHO}}$. See eq. (18) for the definition of $N_{\text{SHO}}$.

## 3.6 Smoothness

Another advantage of the SHO basis for projectors is that it does not require any filtering in reciprocal space before it can be represented on a uniform Cartesian grid, c.f. sec. 3.3.
In fact, the highest kinetic energy in a 1D Hermite function is given by its eigenenergy $E_{1D} = (2n+1)\,\sigma^{-2}$. According to the sampling theorem, $k_{\max} \cdot h_{\max} = \pi$, it is possible to sample the 1D Hermite functions on a uniform real-space grid with maximum grid spacing

$$h_{\max} = \frac{\pi\,\sigma}{\sqrt{2\nu_{\max}+1}}. \tag{29}$$

Since Hermite functions are eigenfunctions of the Fourier transform their representation in reciprocal space reads

$$\mathcal{F}\left\{H_n(x)\, \exp(-\frac{x^2}{2})\right\}(k) \propto (-1)^n\, H_n(k)\, \exp(-\frac{k^2}{2}). \tag{30}$$

The Gaussian suppresses high Fourier components in reciprocal space and, thus, controls the smoothness of the function set. We show in the appendix sec. B that the radial SHO eigenstates $R_{n_r \ell}(r)$ are eigenfunctions of the Fourier-Bessel transform. Hence, the radial SHO eigenstates also feature a Gaussian decay in Fourier space. This shows that SHO states are the simultaneously best localized functions in both spaces.

## 3.7 On-the-fly Basis

Using the recursion relation for Hermite polynomials

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_{n+1}(x) = x\,H_n(x) - \frac{n}{2}\,H_{n-1}(x) \tag{31}$$

it is even possible to cheaply evaluate the grid-sampled 1D Hermite functions during the decompression phase. This reduces the total transferred memory volume to a few Bytes. Hence, it effectively eliminates the memory BW requirement for the data list of the sparse matrix that results from the projector functions. The missing factors ensuring normalization of the Hermite functions can be absorbed into the transformed atomic matrices $\hat{\mathcal{D}}$.

We might experience that in some situations the SHO basis needs to be considerably larger than the original set of numerically

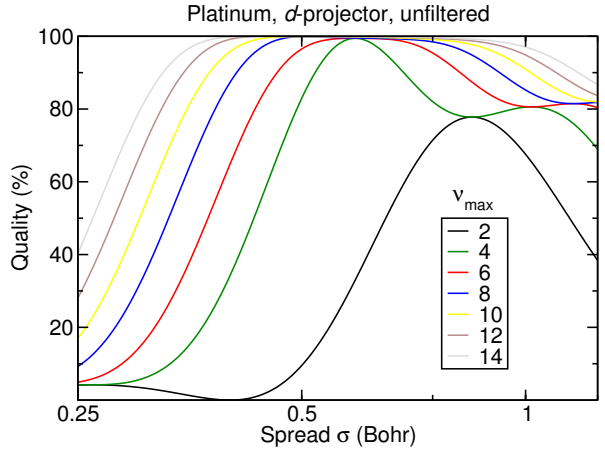


**Figure 5: Convergence of the representation quality for the $d$-projector of platinum. Up to seven radial SHO basis functions were used ($\nu_{\max} = 14$). With respect to the size of the SHO basis, the strongest increase is between $\nu_{\max} = 2$ and 4.**

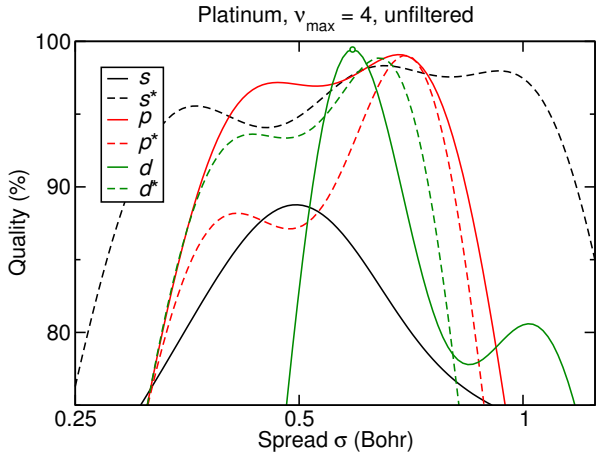


**Figure 6: Best representation quality as a function of $\sigma$ for all projectors of platinum with $\nu_{\max} = 4$. Colors are chosen according to $\ell$-character, solid and dashed lines show the first and second (*) projector, respectively. We select $\sigma = 0.59$ Bohr (green dot). The solid green line again shows the quality of the Pt-$d$-projector as above.**

given projectors which increases the BW requirements for the intermediate coefficients $C_{n_x n_y n_z k}$ for each atom that have to be stored after projection and loaded before expansion. Therefore, we investigate on the necessary minimum size of a SHO basis that still yields an acceptable representation quality of the PAW projector functions.

## 4 QUALITY OF REPRESENTATION

We investigate how well commonly used projector functions $P_{n\ell}(r)$ can be represented in the radial SHO basis $R_{n_r\ell}(r/\sigma)$ as a function of the spread $\sigma$. We define the representation quality $Q$ as

$$Q_\nu(\sigma) = \sum_{n_r=0}^{\lfloor \frac{1}{2}(\nu-\ell)\rfloor} \left| \langle P_{n\ell}(r) | R_{n_r\ell}(r/\sigma) \rangle \right|^2 . \tag{32}$$

Applying this metric to all PAW data sets (`*.PBE`) publicly available in the package `gpaw-setups-0.9.9672` from GPAW [7] shows that we can represent most unfiltered projector functions with at least 90 % quality. Even higher values are seen if the projector functions are mask-filtered [26] before.

A typical convergence of $Q$ w.r.t. $\nu_{\max}$ can be observed at the example of the platinum (Pt) $d$-projector, see fig. 5. For $\nu_{\max} = 2$, the radial SHO basis in the $d$-channel consists of a single function that features no radial nodes. The Pt-$d$-projector function from the GPAW data base can only be represented up to $Q = 77$ % at $\sigma = 0.84$ Bohr (solid black line). Already for two basis function ($\nu_{\max} = 4$), we can reach up to $Q = 99.4$ % at $\sigma = 0.59$ Bohr (solid green line). Adding more basis functions does not increase $Q$ much but makes the maxima wider, i.e. a larger basis is capable of accurately representing the numerically given projectors with some flexibility on $\sigma$. We can see that the best increase in quality is reached with $\nu_{\max} = 4$. For higher $\nu_{\max}$, the gain in $Q$ is small compared to the increase in costs since the SHO basis size grows $\propto \nu_{\max}^3$, c.f. eq. (18).

The simplicity of the SHO basis is an advantage and a drawback at the same time. Only a single $\sigma$ can be selected for all projector functions of one atom. As visible in fig. 6 the quality functions for the $s$, $s^*$, $p$, $p^*$, $d$ and $d^*$-projectors have their maxima at slightly different values of $\sigma$. Therefore, we selected the peak of the sharpest curve which for the case of platinum is the $d$-projector. As discussed above, the best quality for $\nu_{\max} = 4$ is found at $\sigma = 0.59$ Bohr which is marked with a dot in fig. 6.

With respect to costs, we would like to keep the SHO basis as small as possible. The minimal number of radial basis functions can be derived from the number of projectors in each $\ell$-channel. The minimal $\nu_{\max}$ for all PAW data sets analysed can be found in the appendix sec. A. This is, as a rule of thumb, $\nu_{\max} = 4$ for transition metals, $\nu_{\max} = 3$ for non-metals and $\nu_{\max} = 2$ for light elements, $Z < 5$. Further, we suggest $\nu_{\max} = 5$ for the treatment of rare earth elements in order to host two projectors in the $f$-channel. For some elements with the minimal $\nu_{\max}$ it is, however, difficult to find a single value $\sigma$ that produces good representation qualities for all projectors. As it would be desirable in terms of costs to be able to use the minimal $\nu_{\max}$ we suggest a modified PAW dataset generation procedure described in the following section.

## 5 MODIFIED PAW DATA GENERATION

For the generation of PAW data sets several different recipes can be found in the literature, see e.g. [10] or [13] and references therein. This starts at the pseudization of the true local potential $V_{\text{eff}}(r) \rightarrow \tilde{V}_{\text{eff}}(r)$ i.e. there are different shapes for a smooth continuation of the potential inside the augmentation sphere. For the generation of true partial waves $\phi_{n\ell}(r)$, a set of energy parameters $\varepsilon_{n\ell}$ needs to be chosen. Assuming a spherically symmetric true potential $V_{\text{eff}}(r)$, a true partial wave $\phi_{n\ell}(r)$ is found by outwards integration of the radial ordinary differential equation (ODE) for a given energy $\varepsilon_{n\ell}$. This can be either the scalar relativistic equation or the Schrödinger equation. Furthermore, the generation of smooth partial waves

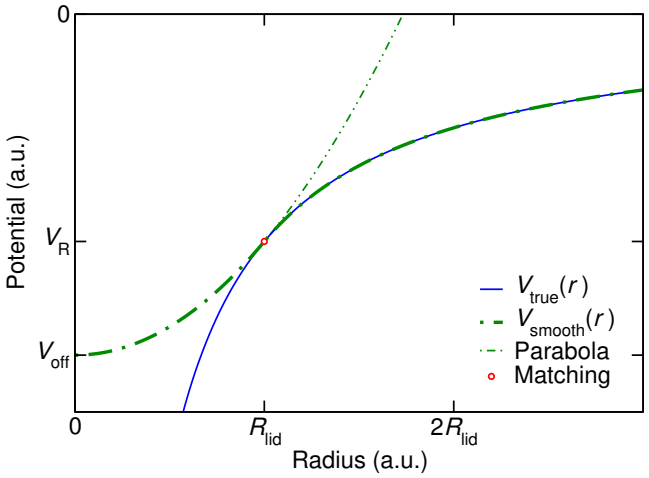


**Figure 7: Pseudization of the local potential according to the *potential lid technique*. A parabola of given curvature closes the true deep core potential like a lid closes a pot.**

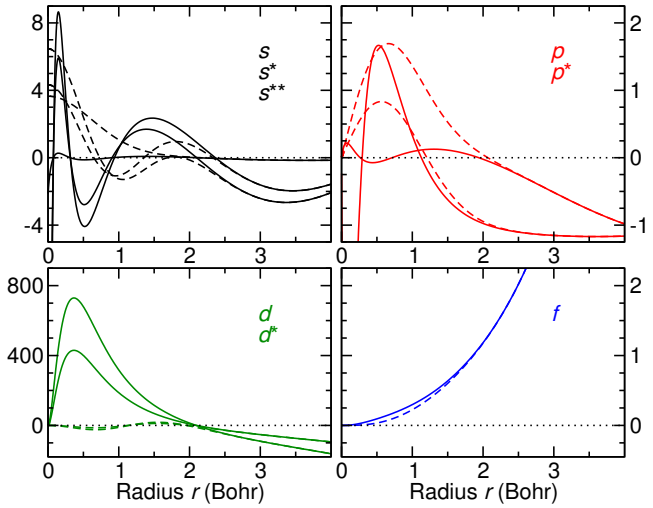


**Figure 8: Iron true partial waves (solid lines) and smooth partial waves (dashed lines) generated by the inverted PAW generation scheme.**

$\tilde{\phi}_{n\ell}(r)$ and projector functions $\tilde{p}_{n\ell}(r)$ depends on the recipe one follows. The simplest prescription is to use $r^\ell$ times a low-order polynomial in $r^2$ for the construction of the smooth partial waves, $\tilde{\phi}_{n\ell}(r)$. This is matched with value and derivative to the true partial wave $\phi_{n\ell}(r)$ at $r$=$R_{\text{aug}}$ [21]. Once obtained $\tilde{\phi}_{n\ell}(r)$, a first guess for the shape of the radial part of the projector function comes from

$$|\tilde{p}_{n\ell}\rangle = \left(\hat{T} + \tilde{V}_{\text{eff}} - \varepsilon_{n\ell}\right)\left|\tilde{\phi}_{n\ell}\right\rangle \tag{33}$$

as suggested by Blöchl in his original work on PAW [2]. In order to keep the SHO basis small (minimum $\nu_{\max}$, if possible) we suggest to invert this scheme. We regard eq. (33) as inhomogeneous ODE and solve it for $\tilde{\phi}_{n\ell}(r)$ using the radial SHO eigenstates $R_{n_r\ell}(r/\sigma)$ with $n_r$=$n$ as projector functions:

$$\left|\tilde{\phi}_{n\ell}\right\rangle = \left(\hat{T} + \tilde{V}_{\text{eff}} - \varepsilon_{n\ell}\right)^{-1}|R_{n\ell}\rangle \tag{34}$$

This removes a lot of free parameters from the PAW generation scheme as we are only left with $\nu_{\max}$ and $\sigma$. The cut-off radius $R_{\text{aug}}$ is only required for the construction of $\tilde{V}_{\text{eff}}(r)$. We can even try to eliminate this dependency: GPAW constructs $\tilde{V}_{\text{eff}}(r)$ as parabola for $r < R_{\text{aug}}$ which matches the true potential at $R_{\text{aug}}$ in value and first derivative. If we constrain the curvature of this parabola to a fixed function of $\sigma$, e.g. $(2\sigma)^{-4}$, the parabola and the true potential touch at a radius $R_{\text{lid}}$ which comes out of the procedure instead of being an input. We call this the *potential lid technique* since the parabola's offset is lowered until it touches the true potential, much like a lid that closes a pot. See fig. 7 for a schematic picture. The factor of two rescaling $\sigma$ has been chosen to produce good results for iron. However, we suspect it to work well for a wider range of elements which would allow to eliminate it from the list of tunable parameters.

As we want the pseudo Hamiltonian to produce the same energy ordering in the valence range as the true Hamiltonian, we demand that that $\varepsilon_{0\ell} < \varepsilon_{1\ell} < \cdots < \varepsilon_{n_{\max,\ell}\ell}$ where $n_{\max,\ell} = \left\lfloor \frac{1}{2}(\nu_{\max} - \ell) \right\rfloor$.

## 5.1 Results

As an example, we will discuss the results obtained for iron ($Z = 26$). Here, we find a lid radius $R_{\text{lid}}$ of 2.08 Bohr which is very close to the radius used in the GPAW data set. A spread parameter of $\sigma$ = 0.65 Bohr for the analytical SHO projectors was selected to yield the simultaneously best representation quality of all projectors with $\nu_{\max} = 4$. The selected reference energies are −0.4 (Fe-4*s*), −0.11 (Fe-4*p*) and -0.57 Rydberg (Fe-3*d*). The latter is also used for $\ell > 2$. Additional true partial waves are generated at $\frac{n}{\ell+1}$ Rydberg above their reference energies. The resulting true and smooth partial waves are shown in fig. 8.

We verify the scattering properties of the new PAW dataset for iron generated according to eq. (34) by comparing its logarithmic derivative $L_\ell(E)$ to that of the true potential, see fig. 9. Except for a 4 mRydberg shift in the position of the *s*-resonance, the (solid) lines of the true $L_\ell(E)$ lie on top of the pseudo $L_\ell(E)$ (dashed lines) which is a reasonably good agreement. The Fe-3*d*-states are strongly localized which makes them appear as a narrow resonance at -0.57 Rydberg, see inset in fig. 9. Since the spread $\sigma$ and the cutoff $\nu_{\max}$ are tunable input parameters to the generation procedure we have sufficient freedom to produce a transferable PAW dataset.

## 5.2 Rescaling of partial waves

Blöchl's PAW generation scheme [2] allows to rescale partial waves and projectors according to

$$\left(\phi_{n\ell}, \tilde{\phi}_{n\ell}, \tilde{p}_{n\ell}\right) \to \left(s\,\phi_{n\ell}, s\,\tilde{\phi}_{n\ell}, s^{-1}\tilde{p}_{n\ell}\right), \quad s \in \mathbb{R} \setminus \{0\} \tag{35}$$

independently for each $n$ and $\ell$. This is because after a successful PAW generation procedure, the dual orthogonality

$$\left\langle \tilde{p}_{n\ell} \middle| \tilde{\phi}_{n'\ell} \right\rangle = \delta_{nn'} \tag{36}$$

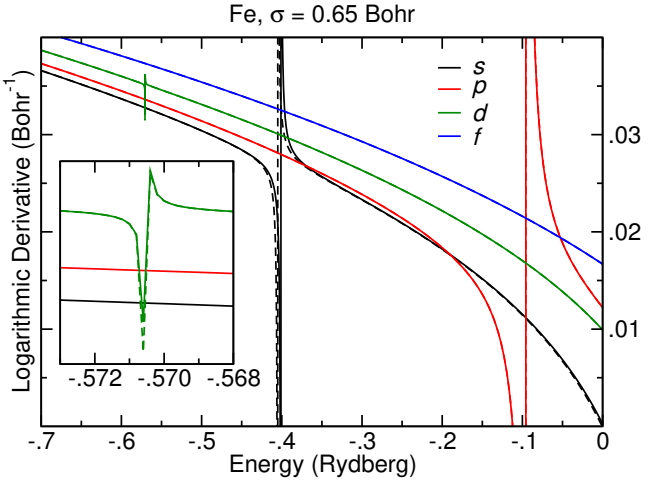


**Figure 9: Logarithmic derivative (at $r$=7.1 Bohr) generated by the inverted PAW generation scheme for iron. The strongest deviation inside the valence energy range is a 4 mRydberg shift of the resonance in the $s$-channel. All other dashed lines (pseudo logarithmic derivatives) are covered by the solid lines (true logarithmic derivatives).**

must hold. Typically, this orthogonality is established by rescaling the lowest projector and the lowest partial waves and orthogonalizing the higher ones to the lowest. This procedure is equivalent to an $LU$-decomposition.

With the SHO projectors, rescaling is not allowed if we aim to exploit the 3D factorization, c.f. eq. (10). It is rather important that the $R_{n\ell}(r/\sigma)$ enter eq. (34) with proper normalization. Therefore, eq. (36) is enforced by applying the inverse of $\left\langle \tilde{p}_{n\ell} | \tilde{\phi}_{n'\ell} \right\rangle$ only to the preliminary pairs of true and smooth partial waves.

## 6 PERFORMANCE MEASUREMENTS

For an impression of the expectable savings we assess the performance of projection (`prj`) and expansion (`add`) operations as defined in sec. 3.2 by (1) and (3), respectively, for a first implementation on NVIDIA GPUs. In order to provide a meaningful comparison between the runtime of `SHO` projectors which are generated on-the-fly and the reference (`USU`) using precomputed projector values, we work in the same framework and exchange only the kernel. The test system is a cubic domain of edge length 16 Å with a grid spacing of 0.25 Å, i.e. $64 \times 64 \times 64$ grid points. Assuming a face-centered cubic crystal of atoms with a lattice constant of 4.08 Å (e.g. silver or gold) lets us find 241 atom positions inside the domain. With a projection radius of 3.55 Å exactly 665 atoms contribute with a non-zero overlap of their projection sphere and the cubic domain, see fig. 10. Mind that there is also a non-zero overlap between projection spheres of neighboring atoms. Unlike the augmentation spheres, the projection spheres may overlap.

The Hamiltonian is applied to 1024 wave functions at a time. This index is used for vectorization over warps of 32 GPU-threads. The performance results are converged w.r.t. this number.

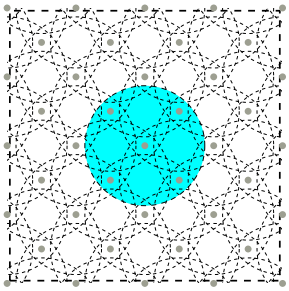

**Figure 10: 2D sketch of the overlapping projection spheres inside a cubic domain of edge length 16 Å. This setup is used for the GPU performance benchmarks in `double` precision. In average, the wave function values of each grid point are involved into eleven projection operations. The radius of each projection sphere is 3.55 Å while the nearest neighbor distance of atomic nuclei is 2.9 Å.**

For stable performance results the kernel execution is repeated 10 times. After a warm-up run which is not considered, the median of 15 runs is reported in tab. 1.

The memory consumption for the precomputed projectors is 76 MByte per projection coefficient. This results in an additional GPU memory request of up to 4.2 GByte.

We executed on an IBM PowerNV 8335-GTB with NVIDIA Tesla P100-SXM2 (Pascal) GPUs connected by `NVlink`. Runtime and compilers are `CUDA 9.2.148`, `GCC 7.2.0`. Furthermore, we benchmarked the latest NVIDIA Tesla V100 (Volta) GPUs mounted on an Intel system (Xeon Gold 6148) connected via PCIe. Here, `CUDA 9.2.148` and `GCC 7.3.0` toolchains were used.

Estimating the upper limit of floating point operations executed we can verify that all kernels are performance-limited by the device memory BW. Hence, an improvement in the timings directly relates to savings in the BW.

From tab. 1 we can see an up to 6.6 times faster execution depending on the kernel, the basis size and the device used. The speedup $S$ compares the timings of the `USU`al projection operations using precomputed projector functions to timings of the `SHO` projection operations. While `USU` kernels can take any number of projectors a SHO basis of size 18 cannot be constructed, therefore, the corresponding entries for `SHO` are empty. However, we included the timing result for `USU18` as it allows to consider a typical calculation setup for transition metal elements. In order to cover $s$, $s^*$, $p$, $p^*$, $d$ and $d^*$ projectors, the SHO basis must be constructed with $\nu_{\max} = 4$, i.e. 35 projection coefficients. Hence, a fair comparison of the two methods is to take `USU18` vs. `SHO35` which translates into an algorithmic speedup given in tab. 2. Executing both kernels after each other, `prj+add`, in `double` precision results in an algorithmic speedup of 2.6 on P100 GPUs and 2.0 on the more recent Volta architecture, V100.

## 7 DISCUSSION

The suggested application of a SHO basis for projection operations or, even going further, analytical SHO projector functions brings

**Table 1: Performance improvement on NVIDIA GPUs. All times are reported in seconds. The speedup $S$ is the unitless ratio of timings USU/SHO for the same # of projectors.**

| NVIDIA P100 Pascal | | | | | | |
|---|---|---|---|---|---|---|
| # | USUprj | SHOprj | $S$ | USUadd | SHOadd | $S$ |
| 1 | 0.480 | 0.453 | 1.1 | 0.274 | 0.187 | 1.5 |
| 4 | 0.834 | 0.471 | 1.8 | 0.803 | 0.363 | 2.2 |
| 10 | 1.259 | 0.430 | 2.9 | 1.820 | 0.520 | 3.5 |
| 18 | 2.080 | | | 3.173 | | |
| 20 | 2.366 | 0.482 | 4.9 | 3.511 | 0.733 | 4.8 |
| 35 | 3.904 | 0.967 | 4.0 | 6.052 | 1.027 | 5.9 |
| 56 | 6.058 | 1.214 | 5.0 | 9.625 | 1.455 | 6.6 |
| NVIDIA V100 Volta | | | | | | |
| # | USUprj | SHOprj | $S$ | USUadd | SHOadd | $S$ |
| 1 | 0.291 | 0.271 | 1.1 | 0.140 | 0.118 | 1.2 |
| 4 | 0.354 | 0.310 | 1.1 | 0.378 | 0.189 | 2.0 |
| 10 | 0.555 | 0.284 | 2.0 | 0.823 | 0.277 | 3.0 |
| 18 | 0.830 | | | 1.411 | | |
| 20 | 1.013 | 0.300 | 3.4 | 1.557 | 0.398 | 3.9 |
| 35 | 2.087 | 0.575 | 3.6 | 2.660 | 0.564 | 4.7 |
| 56 | 1.806 | 0.793 | 2.3 | 4.208 | 0.792 | 5.3 |

**Table 2: Algorithmic speedup comparing USU18 vs. SHO35.**

| | float | | | double | | |
|---|---|---|---|---|---|---|
| GPU | prj | add | both | prj | add | both |
| P100 | 2.3 | 2.9 | 2.7 | 2.2 | 3.1 | 2.6 |
| V100 | 1.4 | 1.7 | 1.6 | 1.4 | 2.5 | 2.0 |

many advantages but also comes with some drawbacks. In this section, we discuss pro and contra.

## 7.1 Advantages

The SHO basis

- factorizes in the Cartesian coordinates $(x, y, z)$.
- can be sampled on uniform grids without filtering.
- has only two parameters, $\sigma$ and $\nu_{\max}$.
- is given analytically.

In particular the last point, the analytical shapes, allows to save memory bandwidth and memory capacity for implementations of grid-based projection operations, as shown in sec. 6.

## 7.2 Drawbacks

The SHO basis does not offer the flexibility to represent any projector function with good quality. However, we have shown that the PAW generation scheme can be adjusted to using radial SHO basis as projectors.

The SHO basis leads to more projection coefficients than usual PAW. For example, two projectors for the $s$, $p$ and $d$-channel makes 18 coefficients. In order to fit a $d^*$-projector into a SHO basis, we need a minimum $\nu_{\max}$ of 4, i.e. 35 coefficients. However, we have to take it as an upside. With $\nu_{\max} = 4$ the SHO basis contains an additional $s^{**}$, seven $f$ and nine $g$-projectors. Adding projectors and partial waves of higher $\ell$ may potentially improve the transferability of the PAW data sets for chemical environments with low symmetry since gradients in the potential scatter into the next higher $\ell$-channel. Compared to commonly used projector sets, SHO projectors treat also higher $\ell$-channels in full-potential accuracy. However, the true transferability remains to be shown in 3D calculations.

# 8 RELATED WORK

Already 1986, Obara and Saika presented the ansatz of 3D products of primitive Cartesian Gaussian (pCG) functions $x^\nu \exp(-\frac{1}{2}x^2)$ in order to evaluate two- and three-center integrals [18]. In fact, there is a strong relation between pCGs and the 1D harmonic oscillator states. The pCG functions form a non-orthogonal basis set. Applying Gram-Schmidt orthogonalization to pCGs leads to 1D Hermite functions. Many works in quantum chemistry, as e.g. Schlegel and Frisch [23], have mentioned the increased basis size $\frac{1}{2}(\ell_{\max}+2)(\ell_{\max}+1)$ of the pCGs compared to $(2\ell_{\max}+1)$ but this has usually been taken as a weakness rather than a strength.

The idea of using the SHO basis has probably not received attention as the related basis set

$$\chi_{n_r \ell m}(\vec{r}) = r^{\ell+2n_r} \exp(-\frac{r^2}{2\sigma^2}) \, Y_{\ell m}(\hat{r}) \tag{37}$$

has been found not to converge as rapidly as a set of only Gaussian-type orbitals (GTOs) contracted over a set of different spread parameters $\sigma$ [5] when used as a basis set for quantum chemistry calculations [22].
Note that the subset of radial SHO states with $n_r = 0$ are GTOs.

# 9 CONCLUSIONS

The eigenstates of the Spherical Harmonic Oscillator (SHO) form a natural link between 3D Cartesian grids and radial representations with sharp quantum numbers for the angular momentum, $(\ell, m)$, without the use of global Fourier transforms. Projection and expansion operations are the heart of the Projector Augmented Wave (PAW) method [2] for electronic structure calculations. Using the SHO basis for these operations allows to exploit its special properties. Similar to compressed tensor representations, the 3D SHO states can be factorized into eigenstates of the 1D harmonic oscillator for each of the three Cartesian coordinates $(x, y, z)$. This offer a great potential for reducing memory bandwidth requirements and memory capacity constraints. In particular since it is cheap to recompute the analytically given 1D basis functions compared to loading them from memory. A first implementation on GPUs shows that we can expect an at least two times faster application of the non-local parts of the PAW Hamiltonian for transition metals. Commonly used PAW projector functions can be represented in a SHO basis, however, the cost-saving minimal basis size is not suitable for representing arbitrary functions. Therefore, we suggest a modified PAW generation scheme in which the projector functions are fixed to the analytically given radial SHO basis functions eliminating the freedom of choosing the shape and scale of the smooth partial waves.

## ACKNOWLEDGEMENTS

The authors thank Thorsten Hater for proofreading and Miriam Hinzen for assistance with the Fourier-Bessel transform.

## REFERENCES

[1] Eduardo Anglada and José M. Soler. 2006. Filtering a distribution simultaneously in real and Fourier space. *Phys. Rev. B* 73 (Mar 2006), 115122. Issue 11. https://doi.org/10.1103/PhysRevB.73.115122

[2] P. E. Blöchl. 1994. Projector augmented-wave method. *Phys. Rev. B* 50 (Dec 1994), 17953–17979. Issue 24. https://doi.org/10.1103/PhysRevB.50.17953

[3] Fred Brackx, Nele De Schepper, and Frank Sommen. 2009. The Fourier-bessel Transform. In *18th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering : July 7 - 9 2009, Bauhaus-University Weimar*, Klaus Gürlebeck and Carsten Könke (Eds.). 18.

[4] E. L. Briggs, D. J. Sullivan, and J. Bernholc. 1996. Real-space multigrid-based approach to large-scale electronic structure calculations. *Phys. Rev. B* 54 (Nov 1996), 14362–14375. Issue 20. https://doi.org/10.1103/PhysRevB.54.14362

[5] Knut Faegri Jr. and Jan Almlof. 1986. Energy-optimized GTO basis sets for LCAO Calculations. A Gradient Approach. *Journal of Computational Chemistry* 7, 4 (1986), 396–405. https://doi.org/10.1002/jcc.540070403 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540070403

[6] Paolo Giannozzi, Stefano Baroni, and et al. 2009. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* 21, 39 (2009), 395502 (19pp). http://www.quantum-espresso.org

[7] GPAW developers. 2017. GPAW: DFT and beyond within the projector-augmented wave method. https://wiki.fysik.dtu.dk/gpaw/. Accessed: 2017-11-13.

[8] P. Hohenberg and W. Kohn. 1964. Inhomogeneous Electron Gas. *Phys. Rev.* 136 (Nov 1964), B864–B871. Issue 3B. https://doi.org/10.1103/PhysRev.136.B864

[9] Jun-Ichi Iwata, Daisuke Takahashi, Atsushi Oshiyama, Taisuke Boku, Kenji Shiraishi, Susumu Okada, and Kazuhiro Yabana. 2010. A massively-parallel electronic-structure calculations based on real-space density functional theory. *J. Comput. Phys.* 229, 6 (2010), 2339 – 2363. https://doi.org/10.1016/j.jcp.2009.11.038

[10] Francois Jollet, Marc Torrent, and Natalie Holzwarth. 2014. Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format. *Computer Physics Communications* 185, 4 (2014), 1246 – 1254. https://doi.org/10.1016/j.cpc.2013.12.023

[11] Venera Khoromskaia and Boris N. Khoromskij. 2015. Tensor numerical methods in quantum chemistry: from Hartree-Fock to excitation energies. *Phys. Chem. Chem. Phys.* 17 (2015), 31491–31509. Issue 47. https://doi.org/10.1039/C5CP01215E

[12] W. Kohn and L. J. Sham. 1965. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* 140 (Nov 1965), A1133–A1138. Issue 4A. https://doi.org/10.1103/PhysRev.140.A1133

[13] G. Kresse and D. Joubert. 1999. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* 59 (Jan 1999), 1758–1775. Issue 3. https://doi.org/10.1103/PhysRevB.59.1758

[14] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier. 2014. Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals. *Critical Reviews in Solid State and Materials Sciences* 39, 1 (2014), 1–24. https://doi.org/10.1080/10408436.2013.772503 arXiv:https://doi.org/10.1080/10408436.2013.772503

[15] Wilhelm Magnus, Fritz Oberhettinger, and Raj Pal Soni. 1966. *Orthogonal polynomials.* Springer Berlin Heidelberg, Berlin, Heidelberg, 204–262. https://doi.org/10.1007/978-3-662-11761-3_5

[16] Miguel A.L. Marques, Alberto Castro, George F. Bertsch, and Angel Rubio. 2003. octopus: a first-principles tool for excited electron-ion dynamics. *Computer Physics Communications* 151, 1 (2003), 60 – 78. https://doi.org/10.1016/S0010-4655(02)00686-0

[17] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. 2005. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* 71 (Jan 2005), 035109. Issue 3. https://doi.org/10.1103/PhysRevB.71.035109

[18] S. Obara and A. Saika. 1986. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *The Journal of Chemical Physics* 84, 7 (1986), 3963–3974. https://doi.org/10.1063/1.450106 arXiv:https://doi.org/10.1063/1.450106

[19] Tomoya Ono, Marcus Heide, et al. 2010. Real-space electronic structure calculations with full-potential all-electron precision for transition metals. *Phys. Rev. B* 82 (Nov 2010), 205115. Issue 20. https://doi.org/10.1103/PhysRevB.82.205115

[20] Tomoya Ono and Kikuji Hirose. 1999. Timesaving Double-Grid Method for Real-Space Electronic-Structure Calculations. *Phys. Rev. Lett.* 82 (Jun 1999), 5016–5019. Issue 25. https://doi.org/10.1103/PhysRevLett.82.5016

[21] Carsten Rostgaard. 2009. The Projector Augmented-wave Method. (Oct. 2009), 25. https://arxiv.org/abs/0910.1921

[22] Ansgar Schäfer, Hans Horn, and Reinhart Ahlrichs. 1992. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics* 97, 4 (1992), 2571–2577. https://doi.org/10.1063/1.463096 arXiv:https://doi.org/10.1063/1.463096

[23] H. Bernhard Schlegel and Michael J. Frisch. 1995. Transformation between Cartesian and pure spherical harmonic Gaussians. *International Journal of Quantum Chemistry* 54, 2 (1995), 83–87. https://doi.org/10.1002/qua.560540202 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560540202

[24] David J. Singh and Lars Nordstrom. 2006. *Planewaves, Pseudopotentials, and the LAPW Method* (2 ed.). Spinger US, New York, NY, USA. 134 pages. https://doi.org/10.1007/978-0-387-29684-5 arXiv:978-0-387-29684-5

[25] José M. Soler and Eduardo Anglada. 2009. Optimal Fourier filtering of a function that is strictly confined within a sphere. *Computer Physics Communications* 180, 7 (2009), 1134 – 1136. https://doi.org/10.1016/j.cpc.2009.01.017

[26] Maxim Tafipolsky and Rochus Schmid. 2006. A general and efficient pseudopotential Fourier filtering scheme for real space methods using mask functions. *The Journal of Chemical Physics* 124, 17 (2006), 174102. https://doi.org/10.1063/1.2193514 arXiv:https://doi.org/10.1063/1.2193514

[27] Marc Torrent, François Jollet, François Bottin, Gilles Zérah, and Xavier Gonze. 2008. Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure. *Computational Materials Science* 42, 2 (2008), 337 – 351. https://doi.org/10.1016/j.commatsci.2007.07.020

[28] Jan Winkelmann, Paul Springer, and Edoardo Di Napoli. 2018. ChASE: Chebyshev Accelerated Subspace iteration Eigensolver for sequences of Hermitian eigenvalue problems. *ACM Trans. Math. Softw.* (May 2018), 33. https://arxiv.org/abs/1805.10121

## A MINIMUM $\nu_{\max}$ FOR ALL ELEMENTS

We suggest the minimal $\nu_{\max}$ for the GPAW setups in the structure of a periodic table of elements from $_1$H to $_{86}$Rn:

```
2                                                       2
2 2                                       3 3 3 3 3 3
3 3                                       3 3 3 3 3 3
3 3 4         4 4 4 4 4 4 4 4 4 3 3 3 3 3 3
3 3 4         4 4 4 - 4 4 4 4 4 4 4 4 3 3 3
3 4 4 5 ... 5 4 4 4 4 4 4 4 4 4 4 4 4 - - 3
```

Rare earth elements (RE) are not included in the publicly available collection `gpaw-setups-0.9.9672` [7]. We suggest $\nu_{\max} = 5$ for REs in order to have two projectors in the $f$-channel. For $\nu_{\max} = 2$, 3, 4 or 5 the size of a 3D SHO basis is 10, 20, 35 or 56, respectively.

## B EIGENFUNCTIONS OF THE FOURIER-BESSEL TRANSFORM

The radial SHO eigenfunctions shown in fig. 2 are eigenfunctions of the Fourier-Bessel transform ($\mathcal{FB}$) which reads

$$\mathcal{FB}\{R_{n\ell}(r)\}(q) = \sqrt{\frac{2}{\pi}} \int_0^\infty r^2 \mathrm{d}r \, j_\ell(qr) \;\; R_{n\ell}(r) \tag{38}$$

where $j_\ell(x)$ are the spherical Bessel functions of the first kind featuring a regular $x^\ell$ behaviour at the origin. The proof is given by Brackx *et al.* [3] (dimension $m = 3$, degree $k = 0$) exploiting the identity given by Magnus *et al.* [15], page 244:

$$\int_0^\infty \mathrm{d}r \exp(-\frac{r^2}{2}) r^{\alpha+1} L_n^{(\alpha)}(r^2) J_\alpha(r\sqrt{x}) = (-1)^n x^{\alpha/2} \exp(-\frac{x}{2}) L_n^{(\alpha)}(x).$$

Replacing $\alpha = \ell + \frac{1}{2}$ and $x = q^2$ this becomes

$$\int_0^\infty \mathrm{d}r \, J_{\ell+\frac{1}{2}}(qr) \, r^{\ell+\frac{3}{2}} \;\; \exp(-\frac{r^2}{2}) \, L_n^{(\ell+\frac{1}{2})}(r^2)$$
$$= (-1)^n \, q^{\ell+\frac{1}{2}} \;\; \exp(-\frac{q^2}{2}) \, L_n^{(\ell+\frac{1}{2})}(q^2).$$

With the common definition of spherical Bessel functions

$$j_\ell(z) = \sqrt{\frac{\pi}{2z}}\, J_{\ell+\frac{1}{2}}(z) \qquad \text{we arrive at}$$

$$\sqrt{\frac{2}{\pi}} \int_0^\infty r^2 \mathrm{d}r\, j_\ell(qr)\;\; r^\ell \exp(-\frac{r^2}{2})\, L_n^{(\ell+\frac{1}{2})}(r^2)$$

$$= (-1)^n\, q^\ell \exp(-\frac{q^2}{2})\, L_n^{(\ell+\frac{1}{2})}(q^2).$$

This shows that the $\mathcal{FB}$-transform produces the same function with exchanged argument ($r \leftrightarrow q$) and eigenvalue $(-1)^n$.