29-09-2023

D2.6 – Report on Final Release of TREX Platform with Interoperable Flagship Codes

Version 1.0

GA no 952165

Dissemination Level

☒ PU: Public
☐ PP: Restricted to other programme participants (including the Commission)
☐ RE: Restricted to a group specified by the consortium (including the Commission)
☐ CO: Confidential, only for members of the consortium (including the Commission)

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Document Information

Project Title	Targeting Real Chemical Accuracy at the EXascale
Project Acronym	TREX
Grant Agreement No	952165
Instrument	Call: H2020-INFRAEDI-2019-1
Topic	INFRAEDI-05-2020 Centres of Excellence in EXascale computing
Start Date of Project	01-10-2020
Duration of Project	36 Months
Project Website	https://trex-coe.eu/
Deliverable Number	D2.6
Deliverable title	Report on pre-release of an integrated platform with inter-operable Open Source flagship codes, including the I/O and, gradually, the QMCkl libraries
Due Date	M36 – 31-09-2023
Actual Submission Date	29-09-2023
Work Package	WP2 – Code modularization and interfacing
Lead Author (Org)	Anthony Scemama (Centre National de la Recherche Scientifique (CNRS))
Contributing Author(s) (Org)	Claudia Filippi (UT), Ravindra Shinde (UT), Edgar Landinez Borda (UT), Kousuke Nakano (SIS), Oto Kohulak (CNRS), Michal Hapka (TUL), Katarzyna Pernal (TUL), Ali Alavi (MPI), Pablo Lopez Rios (MPI)
Reviewers (Org)	Katarzyna Pernal (TUL), William Jalby (UVSQ)
Version	1.0
Dissemination level	Public
Nature	Report
Draft / final	Final
No. of pages including cover	vii (front matter), 15 (main text), VIII (back matter)

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Disclaimer

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

Versioning

Version	Date	Authors	Notes
1.0	29-09-2023	Anthony Scemama (CNRS)	First Official Release

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Abbreviations

AO	Atomic Orbital
CCSD	Coupled Cluster with Single and Double excitations
CI	Configuration Interaction
CIPSI	Configuration Interaction using a Perturbative Selection made Iteratively
CNRS	Centre National de la Recherche Scientifique
CoE	Center of Excellence
DMRG	Density Matrix Renormalization Group
FCIQMC	Full Configuration Interaction Quantum Monte Carlo
HDF5	Hierarchical Data Format version 5
HF	Hartree-Fock
I/O	Input/Output
MO	molecular orbital
MPI	Message Passing Interface
QMC	Quantum Monte Carlo
TREX	Targeting REal chemical accuracy at the eXascale
TREXIO	TREX Input/Output
SCF	Self-Consistent Field
WP	Work Package
VMC	Variational Monte Carlo

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Table of Contents

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

List of Figures

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

1 Executive summary

The TREX project is a collaborative initiative aimed at enhancing Quantum Monte Carlo (QMC) simulations by developing and optimizing flagship codes tailored for exascale computing. This report provides a comprehensive overview of the advancements and achievements under Work Package (WP)2, *Code Modularization and Interfacing*, and highlights its essential role in accomplishing the TREX project's primary objectives. The TREX project focuses on three key goals:

1. Enhancement of accuracy and efficiency in QMC simulations.
2. Facilitation of seamless integration between different QMC codes and libraries.
3. Ensuring interoperability and data exchange capabilities across these codes.

WP2 further narrows down these objectives into four specific milestones:

1. Standardization of Input/Output (I/O) data format and libraries.
2. Modularization of flagship codes for integration with the QMCkl library.
3. Release of optimized versions of flagship codes.
4. Generation of a comprehensive set of reference test cases.

Key Contributions

TREXIO: A standardized I/O data format and library specifically designed for quantum chemistry applications. It is an open-source platform supporting multiple programming languages and offers flexibility in data storage via binary and text-based backends.

Quantum Package: Marked improvements in computational efficiency via GPU accelerators, with the capability to export data in TREXIO format, facilitating cross-code integration.

QMC=Chem: Currently in its second version, focuses on methodological advancements and leverages the QMCkl library for computation and the TREXIO library for reading wave function parameters.

TurboRVB: TurboRVB has evolved significantly, incorporating GPU acceleration, becoming open-source, and integrating with libraries like TREXIO and QMCkl for enhanced performance and interoperability. The TurboGenius Python package further augments its capabilities, enabling high-throughput quantum Monte Carlo calculations and demonstrating consistency with other quantum chemistry packages like PySCF.

CHAMP: Modernized using current Fortran standards and managed by CMake, includes native integration of the TREXIO and QMCkl libraries and brings significant performance improvements.

NECI: The TCHInt library, developed as part of the NECI project, enables the pre-computation, on-the-fly evaluation, and access to transcorrelated Hamiltonian matrix elements for accurate quantum chemical simulations, and has been integrated with TREXIO for robust interoperability, supporting not just FCIQMC but also other quantum chemical methods like CCSD and DMRG.

GammCor: GammCor has undergone significant optimization, enhancing computational speed, scalability, and interoperability. The revamped code now efficiently handles molecules with tens of

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

atoms, more than 50 active electrons, and over 100 active orbitals, and is compatible with popular quantum chemistry software like Molpro, Dalton, and ORCA, as well as the TREXIO library.

Benchmarking and Validation

Task 2.3 generated a rigorous set of reference test cases aimed at optimizing existing computational codes and validating new releases. These test cases are constructed from well-established molecular systems and validated using PySCF Hartree-Fock calculations and Quantum Package. Their application in validating TurboRVB and other codes showcases their importance in maintaining the project's scientific integrity.

Impacts and Benefits

The successful completion of WP 2 promises substantial advancements in the field of quantum chemistry. By promoting data standardization and interoperability, it will enable greater collaboration among researchers, ensure code compatibility, and establish reliable benchmarks for future research.

Academic and Community Engagement

TREXIO, as a pivotal component of WP 2, has garnered significant academic attention, including a high Research Interest Score on ResearchGate and a dedicated publication in the Journal of Chemical Physics.

Conclusion

Work Package 2 serves as a linchpin in achieving the TREX project's overarching objectives by focusing on code modularization, data standardization, and computational efficiency. Its advancements lay the groundwork for a more collaborative and effective research landscape in quantum chemistry.

2 Introduction

The Targeting REal chemical accuracy at the eXascale (TREX) project is a collaborative effort aimed at advancing the field of Quantum Monte Carlo (QMC) simulations. The project brings together experts from various disciplines to develop and optimize flagship codes for exascale computing environments. The overarching objectives of the TREX project include enhancing the accuracy and efficiency of QMC simulations, facilitating the integration of different QMC codes and libraries, and enabling seamless data exchange and interoperability among flagship codes.

WP 2, titled "Code Modularization and Interfacing," focuses on the design and implementation of a platform that integrates the unique capabilities of flagship codes within a modular and interoperable framework. This work package addresses several key aspects, including the development of a common data format and Input/Output (I/O) library, and the modularization of codes to leverage the I/O library and the QMCkl libraries, and the creation of a set of reference test cases.

The specific objectives of WP 2 are as follows:

1. Define a common input/output data format for wave function data and deliver an I/O library that enables easy migration and compatibility between codes, ensuring interoperability and archiving capabilities.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

2. Refactor and modularize the flagship codes to incrementally leverage the functionalities of the QMCkl library from WPs 1 and 3, as well as the common I/O library.

3. Deliver a final release of the flagship codes, enhanced through refactoring and modularization, to take full advantage of the well-structured functionalities provided by the QMCkl library.

4. Generate a comprehensive set of reference test cases to optimize and ensure reproducibility of the main numerical tasks performed by the flagship codes. These domain-specific benchmarks will be utilized to validate new releases of higher-level specific algorithms within each individual code.

The achievement of interoperable flagship codes within the TREX platform brings several significant benefits. First, it enables seamless collaboration and knowledge sharing among researchers and developers working on different codes, fostering the advancement of QMC simulations. Second, the integration of flagship codes within a modular framework allows users to harness the combined capabilities of various codes, resulting in improved accuracy and efficiency in molecular property predictions. Third, the interoperability ensures compatibility between different releases, facilitating the adoption of new features and advancements across the TREX community. Finally, the availability of reference test cases enhances optimization and reproducibility, providing a standardized benchmarking suite for assessing the performance of new algorithms and code versions.

In the subsequent sections of this report, we delve into the specific achievements and outcomes of Work Package 2, including the development of a common data format and I/O library, the modularization of flagship codes, the enhanced release of flagship codes, and the generation of reference test cases. We will present detailed insights into each of these aspects, highlighting the contributions made towards achieving the goals of the TREX project.

3 TREXIO: Common Input/Output Data Format and I/O Library

One of the primary objectives of WP 2 was to establish a common I/O data format and develop an accompanying I/O library. The creation of a standardized data format was crucial to enable seamless data migration and compatibility among different flagship codes within the TREX platform.

TREXIO is an open-source file format and library specifically developed for the storage and manipulation of data generated by quantum chemistry calculations. It serves as a reliable and efficient method for storing and exchanging wave function parameters and matrix elements, making it an essential tool for researchers in the field of quantum chemistry, beyond the TREX Center of Excellence (CoE). The progress on TREXIO was reported in deliverables *D2.1 – Report on a first alpha release of the I/O library, ready for WP4* and *D2.2 – Report on the final release of the I/O library*.[1, 2]

The development of TREXIO has been motivated by the need for a standardized and versatile format that can handle complicated quantum chemistry data (wave function parameters, integrals and others) across different platforms and software packages. To achieve this, the TREXIO library consists of a front-end implemented in the C programming language, along with two back-end options: a binary back-end based on the Hierarchical Data Format version 5 (HDF5) library enabling fast read and write operations, and a text back-end acting as a fall-back when the HDF5 back-end is not

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

available. The TREXIO library is designed to be platform-independent and offers interfaces for other programming languages such as Fortran, Python, Julia and OCaml. This allows researchers to seamlessly integrate TREXIO into their existing computational workflows and take advantage of its functionalities in their preferred programming language.

In addition to the core library, a suite of accompanying tools has been developed to facilitate the utilization of the TREXIO format and library. These tools include converters for popular quantum chemistry codes, enabling the seamless translation of data between different file formats and the TREXIO format. Furthermore, utilities have been developed to validate and manipulate data stored in TREXIO files, ensuring data integrity and providing researchers with the flexibility to analyze and extract relevant information from their quantum chemistry calculations.

The simplicity and ease of use of TREXIO make it an invaluable resource for researchers working with quantum chemistry data. The format's straightforward structure allows for intuitive data storage and retrieval, enabling efficient access to wave function parameters and matrix elements. The versatility of TREXIO facilitates collaboration and data sharing among researchers utilizing different quantum chemistry software packages, enabling seamless exchange of results and fostering advancements in the field.

TREXIO has been showcased and shared with the scientific community through multiple conferences and workshops, where it has been presented in the form of oral presentations[3, 4, 5, 6, 7] or poster presentations.[8, 9, 10] In addition, a paper on TREXIO,[11] describing its design, features, and implementation details, has been published in the Journal of Chemical Physics as part of the special collection titled *High Performance Computing in Chemical Physics*. This publication serves as a comprehensive reference for researchers interested in adopting TREXIO and provides detailed insights into the format's capabilities and advantages. Three months after its publication, this paper reached a *Research Interest Score* of 8.0 on the ResearchGate social media, a score higher than 98% of research items published in 2023.

A web interface is also being developed (hosted at `https://trexio.thescience.dev`) enabling users to convert outputs of quantum chemical calculations with the GAMESS-US package to the TREXIO (HDF5) format. This web application will be extended to enable the conversion also from other quantum chemical codes.

In summary, TREXIO offers a common data format and library specifically tailored for quantum chemistry calculations. Its simplicity, versatility, and compatibility across platforms and programming languages make it an indispensable tool for storing, exchanging, and manipulating quantum chemistry data. The accompanying suite of tools further enhances the usability of TREXIO, providing researchers with a comprehensive set of utilities to streamline their data processing and analysis workflows.

4 Enhancement of Flagship Codes

Task 2.2 focused on the modularization and improvement of the flagship codes within the TREX platform. The objective was to refactor and restructure the codes to incrementally leverage the functionalities provided by the QMCkl library from Work Packages 1 and 3, as well as the TREX Input/Output (TREXIO) library developed in Task 2.1. The modularization process aimed to enhance

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

code modularity, improve code readability and maintainability, and enable seamless integration with the underlying libraries.

The integration of the QMCkl library and the TREXIO library into the flagship codes further enhanced their capabilities. The QMCkl library provided advanced numerical algorithms and techniques that significantly improved the efficiency of the simulations. The TREXIO library ensured seamless data exchange and compatibility among the codes, eliminating potential data format inconsistencies and simplifying the overall data management process, enhancing collaboration and promoting interoperability within the TREX platform..

As a result of the modularization process, the flagship codes within the TREX platform have undergone significant enhancements, and demonstrated substantial improvements in various aspects. The code optimizations and parallelization techniques implemented in collaboration with WP 3 have led to significant performance gains, enabling more efficient simulations and reducing the time required for calculations. The enhanced codes also exhibit improved scalability, allowing for simulations on larger systems and the exploration of more complex chemical and physical phenomena.

Within WP 6, we organized a training on Autotools and CMake with presentations of different packaging possibilities (Spack, Conda, Guix, ...) During this training, all the TREX flagship codes were present to introduce or improve their build systems, facilitating the installation process. In addition, it is a necessary step for preparing packages for packaging systems.

4.1 NECI

The TCHInt library [12, 13, 14] has been developed as part of the NECI project [15, 16] in order to enable pre-computation, on-the-fly evaluation, and access to transcorrelated Hamiltonian matrix elements, which enable a highly-accurate description of dynamic electronic correlations in quantum chemical settings. The TCHInt library allows the use of transcorrelated Hamiltonians not only in the Full Configuration Interaction Quantum Monte Carlo (FCIQMC) method, as implemented in NECI, but in other quantum chemical methods such as Coupled Cluster with Single and Double excitations (CCSD) [17], or Density Matrix Renormalization Group (DMRG) [18].

While the functionality offered by QMCkl is primarily relevant to first-quantized methods, implemented in other TREX flagship codes, TREXIO provides a path towards robust interoperability with other quantum chemistry codes, and it is within the context of TCHInt that TREXIO support has been added to the NECI project. NECI and TCHInt are now capable of directly reading zero-, one-, and two-body matrix elements of the non-transcorrelated Hamiltonian produced by other codes, and TCHInt can write zero-, one-, and two-body matrix elements of the transcorrelated Hamiltonian, as an alternative to using the venerable FCIDUMP format.

4.2 Quantum Package

To simplify the installation, OpamPack[19] was developed to enable the installation of the OCaml compiler and libraries on x86 and ARM systems without the need for access to the internet, which is sometimes blocked for security motivations. In addition, a repository was created containing the source files of most of the external dependencies, simplifying the setup of the environment required to compile Quantum Package. In addition, a recipe to build singularity containers was written to help

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

users install Quantum Package in unfriendly environments.

Algorithmic and methodological improvements were also introduced enabling calculations on larger systems:

- a new algorithm for the Davidson diagonalization reducing the memory footprint,[20]
- the Cholesky decomposition of the electron repulsion integrals tensor and of two-particle reduced density matrices,
- the transcorrelated Self-Consistent Field (SCF) in a bi-orthonormal basis,[21]
- the transcorrelated Configuration Interaction using a Perturbative Selection made Iteratively (CIPSI) in a bi-orthonormal basis,[22, 23]
- integrals involved in transcorrelated calculations were re-expressed in terms of matrix multiplications to increase the computational efficiency,
- a multi-GPU module for the CCSD algorithm, providing a significant acceleration on systems equipped with multiple GPU accelerators,
- a stochastic algorithm for the computation of the perturbative triples correction in coupled cluster calculations (article in preparation).

Exporting integrals and wavefunctions in TREXIO format enables Quantum Package to create the necessary quantities needed by all the other TREX codes. Beyond the TREX CoE, thanks to TREXIO Quantum Package was interfaced with the Auxiliary-Field QMC code *Ipie*, and with two other codes developed in different groups: one for the computation of quantum dots, and the other one for performing selected Coupled Cluster calculations.

GPU acceleration posed a significant challenge for Quantum Package, primarily due to the algorithmic intricacies of CIPSI, which is characterized by low arithmetic intensity. In the first phase of the TREX project, an attempt was made during a GPU Hackathon hosted by Nvidia and the Jülich Supercomputing Center to re-engineer the computational bottleneck in CIPSI. The aim was to formulate it as a series of small matrix multiplications, which could be offloaded to the GPU for execution. Despite considerable effort, the performance gain was marginal, yielding only a 1.1x speedup relative to the CPU-based implementation. This underperformance was attributed to the insufficient matrix size, which failed to offset the data transfer overheads. Additionally, the GPU-accelerated version of the code required the use of Nvidia's proprietary compiler. This compiler lacks support for certain intrinsic Fortran functions outlined in the Fortran2008 standard, such as `trailz` and `leadz`. Consequently, these elementary functions had to be re-implemented, leading to a noticeable performance degradation in the CPU-based code. Furthermore, the code had to be restructured to incorporate conditional statements, ensuring compatibility across both GPU-accelerated and CPU-only platforms. Given these constraints, and considering the minor performance gains, it was determined that the drawbacks, in terms of both code readability and ease of installation, outweighed the benefits. As a result, the decision was made to discontinue this particular development branch of the code.

Very recently, a fresh approach was employed to implement GPU acceleration for the CCSD module within Quantum Package. Unlike the previous attempts, the CCSD algorithm possesses

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Table 1: Comparison of the timings (in seconds) obtained for the calculation of the benzene molecule in CCSD/cc-pVTZ using the CPU or the GPU implementation in Quantum Package, and comparison with the theoretical peak performance.

Hardware	# CPUs	# GPUs	Time (s)	GFlops/s	% Peak
2 Intel Xeon Gold 6130 CPUs @ 2.1 GHz	32	0	693	1.243	57.8
2 AMD Zen2 Epyc 7402 CPUs @ 2.8 GHz	48	0	695	1.239	57.6
2 Intel Xeon Gold 6140 CPUs @ 2.3 GHz	36	0	703	1.225	52.0
with 4 Nvidia V100 GPUs	36	1	211	4.082	52.3
	36	2	114	7.556	48.4
	36	3	82	10.379	44.3
	36	4	67	12.857	41.2
1 ARM Ampere Altra Q80-30 CPU @ 3 GHz	80	0	767	1.123	58.5
with 2 Nvidia A100 GPUs	80	1	111	7.761	39.8
	80	2	60	14.357	36.8

high arithmetic intensity and lends itself naturally to an efficient GPU implementation. This is due to the algorithmic design, which can be articulated as a sequence of large matrix multiplications, interspersed with row or column substitutions in rank-four tensors. Such a computational profile is highly compatible with GPU architectures, which excel at handling these types of tasks via the user of *tensor cores*. To circumvent the limitations imposed by Nvidia's proprietary Fortran compiler, a novel strategy was deployed. The entire module was recast in C, deliberately omitting any OpenMP directives geared towards accelerators, such as `omp target`. Instead, external libraries were utilized to manage all data transfers and matrix multiplications. For multi-GPU support, different cuBLAS calls were run in separate CPU threads, which were spawned by OpenMP sections. This design decision made it possible to compile Quantum Package in its standard manner. The sole modification was the inclusion of `libcudart.so`, `libcublas.so`, and `libcublasLt.so` in the linking phase. This strategy has yielded considerable success. Not only did it have a minimal impact on the original source code and installation procedures for Quantum Package, but it also produced a level of acceleration (Table 1) sufficient to merit the integration of this GPU-enabled code into the main development branch.

Drawing from this positive outcome, we have outlined a comprehensive strategy for extending GPU acceleration to other modules within Quantum Package. Instead of making direct calls to cuBLAS within a C function invoked by Quantum Package, we are in the process of developing a standalone library. This library will serve as a wrapper around various BLAS implementations such as cuBLAS and RocBLAS. The matrix multiplication operations will be divided into tasks that are scheduled by StarPU[24], ensuring optimal task distribution between CPUs and GPUs. This also facilitates a more effective overlap between data transfers and computational tasks. Furthermore, the `configure` script

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

associated with this new library will automatically determine the optimal configuration for the host machine, thereby obviating the need for such considerations within the Quantum Package codebase.

4.3 GammCor

GammCor is a quantum chemistry code for Multiconfigurational Symmetry Adapted Perturbation Theory, SAPT(MC)[25], and multireference adiabatic connection (AC)[26] electron correlation energy calculations. The unique feature of these two quantum chemistry methods is that they rely only on the low-order reduced functions: one- and two-electron reduced density matrices, which enables efficient calculations. GammCor can be interfaced with any other quantum chemistry code which provides density matrices from wavefunction or wavefunction-free calculations.

In due course of realizing the TREX project, GammCor has been substantially rewritten in order to achieve speed up in calculations and to extend the applicability of both SAPT(MC) and AC to larger molecular systems. Currently, the code is able to treat molecules comprising a few tens of atoms and basis sets exceeding 1000 one-electron functions. The latest version of GammCor is capable of handling more than 50 active electrons in more than 100 active orbitals. Progress in GammCor's computational abilities has been realized through:

1. Improved handling of two electron integrals, concerning both memory requirements and scaling of the computational cost with the increase of the system size.

2. Development and implementation of novel computational algorithms for extended random phase approximation - the engine for SAPT(MC) and AC.

3. Improved vectorization and parallelization of the code.

4. Improved interoperability of GammCor with other codes.

Ad.(1)

Until recently, GammCor required access to external electron integrals. As the first step in code optimization, the handling of two-electron integrals was considerably improved. Disk space requirements for sorting of atomic integrals were reduced by more than a factor of two due to the implementation of a one-step sorting algorithm based on multiple passes through raw binary files. Although the two-step algorithm requires only a single pass through the integral file, the new one-step/multiple-passes approach is as efficient in terms of the CPU time. This is possible by better handling of RAM allocations and avoiding writing large, intermediate files on disk, so that the overall amount of I/O operations is significantly reduced in the one-step approach compared to the two-step procedure. The second stage of code optimization and a milestone in the GammCor development was the introduction of a separate module for Cholesky decomposition of two-electron integrals. Representation of four-index integrals via three-index Cholesky vectors enables reduction of the computational cost with the system size by roughly one order of magnitude. This was achieved not only by using an efficient, 3-index integral transformation parallelized at the MKL level, but also by developing new quantum mechanical algorithms. Finally, an independent library for on-the-fly calculation and Cholesky decomposition of atomic integrals has been incorporated in GammCor.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

This development allows one to avoid dumping, storing and transferring to GammCor large files with two-electron integrals generated by an external code.

Ad.(2)

The algorithmic advancement has been achieved by replacing a costly diagonalization problem in Extended Random Phase Approximation by a recursive solution of an order-consistent linear equation. In combination with the projection on the Cholesky vector space, the new algorithm affords reduction of the computational cost by one order of magnitude. As a result, the bottleneck steps in SAPT(MC) and AC calculations now scale with the fifth power with the size of the system.[27, 28]

Ad.(3)

The performance of GammCor has been examined with MAQAO since the beginning of the TREX project. This has led to improved vectorization and parallelization of the code. Most of the threaded code relies on parallel MKL library routines such as DGEMM, while parts of the program responsible for handling of two-electron integrals use OpenMP.

Ad.(4)

GammCor has been interfaced with widely used popular quantum chemistry codes: Molpro, Dalton, ORCA. Recently, GammCor has been also interfaced with the TREXIO library to obtain atomic integrals and reduced density matrices from external software. This new capability is used within the Quantum Package $\rightarrow$ TREXIO $\rightarrow$ GammCor workflow. As a result, noncovalent interactions SAPT(MC) calculations with nearly exact wave functions (CIPSI) are possible. GammCor can be also used to recover out-of-CIPSI electron correlation energy, which should allow to push the limits of CIPSI and achieve chemical accuracy for systems comprising dozens of active electrons in hundreds of active orbitals.

The installation of GammCor using the CMake tool is now possible. A set of macros checks for the necessary external libraries and compiles, if necessary. The installation macros have been designed to facilitate customization for different architectures and processor instruction flags.

4.4 QMC=Chem

A second version of QMC=Chem is currently under active development, and it comes with significant structural and functional modifications. One of the most notable changes is the complete removal of internal functions initially responsible for the computation of wave functions. Instead, all computational tasks related to basis functions and Molecular Orbitals (molecular orbitals (MOs)) are now outsourced to the QMCkl library. Moreover, parameters defining the wave function are exclusively read from the TREXIO file format, thereby streamlining the process and enhancing interoperability with other computational tools.

This new version also introduces a range of methodological enhancements, which include:

- The incorporation of a specialized algorithm designed for the efficient optimization of exceptionally large Configuration Interaction (CI) expansions, as detailed in a recent publication[29].
- The formulation of a Jastrow factor that is uniquely adapted for transcorrelated calculations[22, 23].
- The ability to compute the variational energy for right-wavefunctions that are generated using transcorrelated methods in Quantum Package.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

- The introduction of a feature that enables the computation of the transcorrelated energy using both left- and right-wave functions.

Given that these advancements have resulted in a substantial reduction in the code's size and complexity, future plans entail the full integration of QMC=Chem as a dedicated module within Quantum Package. This strategic move is aimed at leveraging Quantum Package's already established user base to accelerate the adoption of QMC methods.

4.5 TurboRVB

At the beginning of the TREX project, TurboRVB was already able to harness the computational power of accelerators such as GPUs. However, the initial implementation underwent substantial revisions, with a particular emphasis on the utilization of mathematical libraries, notably cuBLAS and cuSOLVER. The incorporation of these libraries has been instrumental in facilitating significant enhancements in computational speed. A range of computational tasks can leverage the parallel processing capabilities of GPUs. Specifically, these tasks include but are not limited to, the transformation from atomic orbitals to molecular orbitals, stochastic reconfiguration, and wavefunction updates. The efficacy of these GPU enhancements is illustrated in Fig. 1, where speed-up factors are presented for various periodic systems.

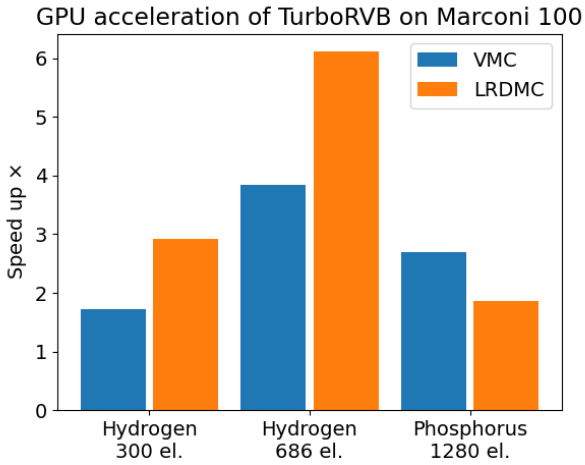

Figure 1: Comparison of TurboRVB's GPU-accelerated performance with its optimal CPU-only counterpart, as benchmarked on the Marconi 100 supercomputer. The figure covers three distinct scenarios involving Variational Monte Carlo (VMC) and Lattice-Regularized Diffusion Monte Carlo (LRDMC) simulations, with different numbers of electrons (300, 686 and 1280). Given TurboRVB's inherent capacity for parallelization as a Monte Carlo code, we recommend the concurrent execution of multiple Markov chains for maximum efficiency. Utilization of NVIDIA's Multi Process Service is crucial for achieving optimal GPU performance, as evidenced by the results presented.

As of April 2023, TurboRVB has transitioned into an open-source project governed by the GNU General Public License v3.0 (https://github.com/sissaschool/turborvb). The journey towards open sourcing required a comprehensive code refactoring process. Specifically, we had to remove

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

or modify segments of the code that were proprietary or otherwise non-compliant with the GPL v3.0 guidelines to ensure complete adherence to the license's stipulations. In addition to this, the user experience was a significant concern; therefore, we produced extensive new documentation to facilitate easier use of the software (https://sissaschool.github.io/turborvb/). To further support users, we initiated a dedicated repository containing illustrative examples (https://github.com/kousuke-nakano/turbotutorials). As part of our effort to streamline the installation and deployment process, we also overhauled the existing build system. Now, it is based on `CMake`, a widely-used, cross-platform build system, thereby enhancing its usability and adaptability across different computational environments.

TurboRVB has its own format for storing wavefunction information, specifically contained within a file designated as `fort.10`. The software can work with a broad spectrum of types of atomic basis sets, offering considerable flexibility. However, it should be noted that in the majority of production-level computations, Gaussian-type orbitals (GTOs) are predominantly utilized. In a significant enhancement to its features, TurboRVB can now be compiled with support for the TREXIO library. This addition has particular relevance when GTO-based wavefunctions are employed. Utilizing the TREXIO file format for wavefunction storage not only enhances the efficiency of the software but also markedly increases its interoperability with other computational codes. This integration with TREXIO is an essential step toward creating a more cohesive and flexible computational environment, making TurboRVB a more versatile tool for quantum chemical calculations.

In a recent enhancement to its feature set, TurboRVB can now be compiled in conjunction with the QMCkl library. This integration represents a considerable advancement in the computational efficiency of TurboRVB, particularly in the evaluation of the determinant component of the wavefunction, a crucial element in QMC calculations. While TurboRVB possesses its own internal wavefunction evaluator, integrating the computation of the Atomic Orbitals (AOs) with the QMCkl library has led to a 20% computational speed-up.

We have recently developed TurboGenius, an open-source Python package tailored to provide comprehensive management of QMC computations through Python scripts. The design philosophy behind TurboGenius centers on enabling high-throughput calculations with TurboRVB. Implemented in Python 3, the package capitalizes on Python's inherent advantages for workflow orchestration, including its compatibility with major scientific workflow frameworks, to facilitate seamless integration.

Employing TurboGenius, we conducted an extensive assessment to scrutinize the inter-package consistency between TurboRVB and other well-established quantum chemistry software suites. Specifically, we carried out comparisons involving Hartree-Fock (HF) energies as determined by PySCF against Variational Monte Carlo (VMC) energies as calculated by TurboRVB, using the same HF wavefunctions for an array of molecular and crystalline systems. The consistency observed in these computational outcomes serves as a validation for both TurboRVB and TurboGenius, affirming the accuracy and reliability of their implementations. These findings are documented in a manuscript currently under preparation, authored by K. Nakano *et al*, and expected to be published in 2023.

4.6 CHAMP

In March 2023, the CHAMP code (https://github.com/filippi-claudia/champ) was made publicly available under the GPL-3.0 license. During the TREX project, significant work of refactoring,

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

modularization, and performance improvement has been performed on CHAMP to make it user- and developer-friendly as well as efficient. We further elaborate below on various aspects already outlined in deliverable D2.5 (28-10-2022) and on some recently introduced features.

CHAMP was migrated from Fortran77 to the modern Fortran standard with the removal of common blocks, implicit declarations, *etc*. Furthermore, a modern object-oriented Fortran programming approach has been adopted to ensure a modular structure of the code, also in view of future developments. Many subroutines have been revised with the modern failsafe interfaces for the external procedures and are wrapped as individual modules for better argument checking.

The cross-compiling tool CMake has been introduced in CHAMP to build the code and extensively checked. Several short macros (`FindSIMD`, `FindTREXIO`, and `FindQMCKL`) have been developed to detect the computer architecture, processor instruction flags, and preinstalled libraries. For code developers, macros (`FindGitInfo` and `FindSystemInfo`) have been introduced to print version control information while compiling; automatic testing and packaging utilities are now also available. Automatic detection of processor instruction flags in the CMake building tool has been used to maximize the processor capabilities and to improve the cache access or vectorization of the code via data aligning, inlining, and fused-multiply add flags.

Three different Continuous Integration (CI) workflows (two for testing CHAMP and one for testing the TREXIO Python interface) have been incorporated to run the tests after every commit in the code on GitHub servers as well as the self-hosted server at UTwente. A Continuous Deployment (CD) workflow has been incorporated to generate documentation. We have started introducing Fortran-based unit testing of the code.

CHAMP user manual is currently being completed (in particular the complete description of keywords) and hosted at https://trex-coe.github.io/champ-user-manual/.

A new (user- and developer-friendly) input file parser has been designed for CHAMP. The parser is derived from the `libfdf` library which was developed within the E-CAM CoE. Our implementation can handle multiple data types and file formats. The parser is coupled with an improved Message Passing Interface (MPI) implementation that significantly reduces I/O access to input data files leading to performance gain. The code has adapted the input data to the standards set by the TREXIO library for atomic and molecular orbital ordering. The code output printout has been revised. Several elapsed-time printouts have been introduced to keep track of code performance.

The TREXIO library has been implemented natively in the code. This allows us to import only one data file to perform a calculation. Additionally, a fallback text converter that translates the binary TREXIO format into text has been written to provide more flexibility to the users and developers. This has led to a more compact and modular input file for the CHAMP code, a significant reduction in I/O access, and improved interoperability among the TREX flagship codes as well as with other quantum chemical codes outside the consortium.

The subroutines that store and manipulate molecular orbital information and multi-determinant wave functions have undergone large restructuring to improve the bottlenecks in memory as well as the performance of the code. We relied on the MAQAO profiling tools (UVSQ) to identify specific hot spots in the code, where recasting the loops and arrays has improved cache accessing and eliminating the dependency on conditional statements has helped in vectorization. All of this has given a significant improvement on scalability and speed performance by 50% compared to the CHAMP version of January 2022.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

The QMCkl library developed within TREX has been implemented natively in the code for the computation of the molecular orbitals. We obtained an additional gain of 40-45% in speed performance as illustrated in Fig. 2 for the computation of the energy of the benzene molecule with a single determinant. Further improvements are expected once the QMCkl library is also called in the computation of the Jastrow factor. We also note that, so far, we have been testing only the human-friendly (and yet optimized) version of the QMCkl library. We are currently finalizing the integration and testing of the GPU-enabled QMCkl.

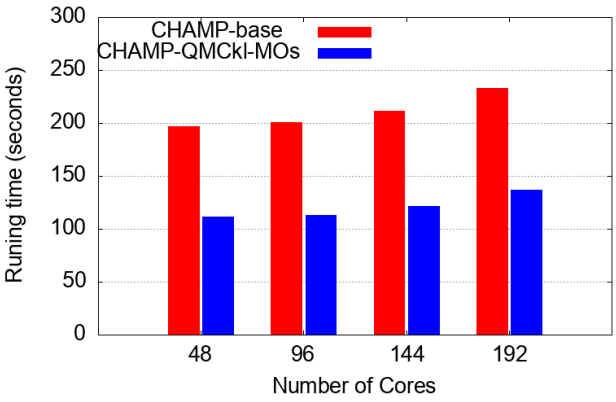

Figure 2: Performance improvement of CHAMP thanks to the integration of (human-readable version of) the QMCkl library for computation of molecular orbitals. The run is a QMC energy computation for a benzene molecule (42 electrons) on JUWELS Skylake CPU.

Two major new features have been implemented to significantly extend the capabilities of the code:

a) The state-specific QMC optimization of Jastrow-Slater wave functions for multiple states imposing orthogonality constraints. This enables optimizing all wave function (linear, orbital, and Jastrow) parameters for ground and excited states via the stochastic reconfiguration method. The method has been successfully employed in the study of double excitations [30], a collaborative work of the UT, CNRS/Toulouse, and SISSA.

b) The ability to perform QMC calculations on periodic systems using Gaussian-based atomic orbitals (currently, restricted to real molecular orbitals). This part of the code has been extensively tested and its performance is being actively investigated. The QMCkl library has been integrated in the periodic part of the code for the computation of the atomic orbitals of the multiple periodic images. The periodic calculations can be easily set up using TREXIO after running a self-consistent PySCF calculation.

5 Reference Test Cases

Task 2.3 is an integral component of our ongoing work, centering on the generation of a comprehensive set of reference test cases. These serve as critical benchmarks designed for the meticulous

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

optimization and ensuring reproducibility of various numerical tasks executed by the flagship computational codes that constitute the TREX platform. These reference test cases are multi-faceted in their utility within the TREX ecosystem. In the first instance, they serve as an invaluable tool for optimizing these flagship codes by establishing a standardized measure. Benchmarking and performance comparison against this curated reference data empowers developers to pinpoint specific areas that require enhancement, thereby fine-tuning algorithms and significantly elevating the overall computational efficiency of the codes.

Secondly, these test cases act as the bedrock for the validation of new releases, especially when dealing with higher-level specific algorithms within each discrete computational code. This ensures that any technological advancements or updates are incorporated without sacrificing the code's existing levels of accuracy and reliability. In doing so, these reference test cases serve not merely as static benchmarks but as dynamic tools for quality assurance. By instituting a universally applicable suite of standardized tests, researchers are furnished with the capability to independently corroborate the accuracy and validity of their scientific results. This indispensable feature substantially augments transparency and instills a heightened level of trust within the broader scientific community, thereby facilitating research outcomes that are both more reliable and reproducible.

To construct these reference test cases, molecules and materials were carefully selected from well-established benchmark systems, including but not limited to G2, which covers atomization energies of 55 molecules, S22, featuring interaction energies of small model complexes, DNA base pairs, and amino acid pairs, A24, focusing on non-covalent interactions of 24 complex systems, SCAI, detailing potential interacting amino acid side chain pairs grounded in data from the Atlas of Protein Side-Chain Interactions, and a set of 12 cubic crystals. For each of these molecular systems, a TREXIO file was generated using a PySCF HF calculation. These TREXIO files underwent a rigorous validation process, where the energy of the wave function stored within each file was recomputed using Quantum Package, thereby ensuring its consistency with the original PySCF calculations.

These validated TREXIO files were subsequently employed to affirm the accuracy of TurboRVB. The energies of the wave functions were evaluated utilizing a VMC sampling mechanism, confirming that for all systems tested, the exact energy resided within the statistical error bars calculated with TurboRVB. In a further extension of this validation process, each TREXIO file was augmented with a set of electron coordinates along with the corresponding local energy. This enriched reference data set can be employed to ascertain that the wave function is accurately evaluated at a specific set of electron positions by any QMC code. Moreover, TurboRVB underwent an additional layer of validation through the comparison of chemical properties such as binding energies, equilibrium lattice parameters, equilibrium geometries, and harmonic frequencies against reference data from coupled-cluster calculations. This extensive validation work is thoroughly documented in a dedicated scientific publication,[31] and the corresponding data is given in appendix.

6 Conclusion

The development and subsequent integration of TREXIO into the overarching suite of TREX codes have been transformative in establishing a synergistic ecosystem that allows for seamless interoperability across all flagship codes. This advancement has also facilitated an extraordinarily simplified and streamlined communication interface for the exchange of wave function data between TREX native

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

codes and external computational software. By acting as a centralized repository and communication hub, TREXIO eliminates data inconsistency and redundancy, thereby ensuring that wave function data can be easily shared and manipulated across a myriad of platforms without loss of integrity or fidelity.

Moreover, the integration of the QMCkl library into the existing QMC codes represents another significant milestone in this work package. By doing so, we have achieved a consolidated avenue for High-Performance Computing (HPC) optimization that circumvents the need to customize or optimize each individual code in isolation. Instead of a fragmented approach that requires separate optimization efforts, the inclusion of the QMCkl library allows for a centralized focus where HPC optimization strategies can be deployed in a unified and coordinated manner. This has resulted in appreciable gains in computational performance and efficiency, consequently accelerating the simulation processes and reducing the computational time and resources required. It brings forth the notion of 'write once, optimize once, and run everywhere efficiently,' embodying a more effective strategy for computational resource management.

In summary, the objectives delineated for this particular work package have not only been achieved but have been fulfilled with exceptional efficacy. The creation of a cohesive and interconnected software ecosystem through the integration of TREXIO, and the centralized HPC optimization made possible by incorporating the QMCkl library, are seminal contributions. They represent pivotal strides in the TREX platform's evolution, offering robust solutions for interoperability and computational efficiency that are aligned with the long-term objectives of delivering reliable, high-performance computational tools for scientific research. With these accomplishments, we affirm the successful realization of all aims and objectives set forth in this work package, setting the stage for future innovations and improvements in the TREX platform.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

References

[1] S. Sorella and A. Scemama, "D2.1 – Report on a first alpha release of the I/O library, ready for WP4," Mar. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.7682800

[2] A. Scemama, "D2.2 – Report on the final release of the I/O library." Sep. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.7735045

[3] C. Filippi, A. Alavi, K. Pernal, K. Nakano, E. Posenitskiy, A. Scemama, S. Pittonet, R. Shinde, P. López Ríos, G. Tenti, A. Zen, and S. Shepard, "TREX High Performance Software Solutions for Quantum Mechanical Simulations at the Exascale," Feb. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7624691

[4] A. Scemama, "Library development within trex," Apr. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4708403

[5] A. Scemama, W. Jalby, V. G. Chilkuri, E. Posenitskiy, P. de Oliveira Castro, and C. Valensi, "Libraries developed in the TREX Center of Excellence - poster presented at the CECAM Workshop 2022," Oct. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7180354

[6] A. Scemama, "The trexio file format and library," Apr. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7804737

[7] ——, "The trexio file format and library," Jun. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6624127

[8] A. Auweter, M. Casula, W. Jalby, S. Pittonet, J.-M. Denis, R. Dolbeau, M.-A. Garigue, and C. Prunty, "TREX at TERATEC 2022: Enabling the community codes for stochastic quantum chemical simulations at he exascale," Jul. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6821103

[9] E. L. Borda, G. Abrusci, A. Alavi, V. G. Chilkuri, F. Coppens, C. Filippi, A. Delval, M. Hapka, M. Hoffer, W. Jalby, P. L. Rios, K. Nakano, P. de Oliveira Castro, R. Panades, K. Pernal, E. Posenitskiy, R. Shinde, A. Sokół, S. Sorella, and A. Scemama, "Leveraging stochastic electronic structure methods at the exascale," Sep. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7108146

[10] F. Coppens, W. Jalby, and A. Scemama, "Leveraging stochastic electronic structure methods at the exascale," Mar. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7753407

[11] E. Posenitskiy, V. G. Chilkuri, A. Ammar, M. Hapka, K. Pernal, R. Shinde, E. J. Landinez Borda, C. Filippi, K. Nakano, O. Kohulák, S. Sorella, P. de Oliveira Castro, W. Jalby, P. L. Ríos, A. Alavi, and A. Scemama, "TREXIO: A file format and library for quantum chemistry," *J. Chem. Phys.*, vol. 158, no. 17, May 2023.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

[12] J. P. Haupt, S. M. Hosseini, P. López Ríos, W. Dobrautz, A. Cohen, and A. Alavi, "Optimizing Jastrow factors for the transcorrelated method," *J. Chem. Phys.*, vol. 158, no. 22, p. 224105, Jun. 2023.

[13] E. M. Christlmaier, T. Schraivogel, P. López Ríos, A. Alavi, and D. Kats, "xTC: An efficient treatment of three-body interactions in transcorrelated methods," *J. Chem. Phys.*, vol. 159, no. 1, p. 014113, Jul. 2023.

[14] "TCHInt," 2023, available from the authors upon reasonable request.

[15] K. Guther, R. J. Anderson, N. S. Blunt, N. A. Bogdanov, D. Cleland, N. Dattani, W. Dobrautz, K. Ghanem, P. Jeszenszki, N. Liebermann, G. L. Manni, A. Y. Lozovoi, H. Luo, D. Ma, F. Merz, C. Overy, M. Rampp, P. K. Samanta, L. R. Schwarz, J. J. Shepherd, S. D. Smart, E. Vitale, O. Weser, G. H. Booth, and A. Alavi, "NECI: N-Electron Configuration Interaction with an emphasis on state-of-the-art stochastic methods," *J. Chem. Phys.*, vol. 153, no. 3, p. 034107, Jul. 2020.

[16] "NECI," Sep. 2023. [Online]. Available: https://github.com/ghb24/NECI_STABLE

[17] T. Schraivogel, E. M. Christlmaier, P. López Ríos, A. Alavi, and D. Kats, "Transcorrelated coupled cluster methods. II. Molecular systems," *J. Chem. Phys.*, vol. 158, no. 21, p. 214106, Jun. 2023.

[18] K. Liao, H. Zhai, E. M. Christlmaier, T. Schraivogel, P. López Ríos, D. Kats, and A. Alavi, "Density matrix renormalization group for transcorrelated hamiltonians: Ground and excited states in molecules," *J. Chem. Theory Comput.*, vol. 19, no. 6, pp. 1734–1743, 2023.

[19] "OpamPack," Jul. 2023, [Online; accessed 6. Jul. 2023]. [Online]. Available: https://github.com/scemama/OpamPack

[20] V. G. Chilkuri, T. Applencourt, K. Gasperich, P.-F. Loos, and A. Scemama, "Spin-adapted selected configuration interaction in a determinant basis," in *Advances in Quantum Chemistry*. Cambridge, MA, USA: Academic Press, Jan. 2021, vol. 83, pp. 65–81.

[21] A. Ammar, A. Scemama, and E. Giner, "Biorthonormal Orbital Optimization with a Cheap Core-Electron-Free Three-Body Correlation Factor for Quantum Monte Carlo and Transcorrelation," *J. Chem. Theory Comput.*, vol. 2023, Jun. 2023.

[22] ——, "Extension of selected configuration interaction for transcorrelated methods," *J. Chem. Phys.*, vol. 157, no. 13, Oct. 2022.

[23] ——, "Transcorrelated selected configuration interaction in a bi-orthonormal basis and a cheap three-body correlation factor," *arXiv*, Jun. 2023.

[24] C. Augonnet, S. Thibault, R. Namyst, and P.-A. Wacrenier, "StarPU: A Unified Platform for Task Scheduling on Heterogeneous Multicore Architectures," *CCPE - Concurrency and Computation: Practice and Experience, Special Issue: Euro-Par 2009*, vol. 23, pp. 187–198, Feb. 2011. [Online]. Available: http://hal.inria.fr/inria-00550877

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

[25] M. Hapka, M. Przybytek, and K. Pernal, "Symmetry-adapted perturbation theory based on multiconfigurational wave function description of monomers," *Journal of Chemical Theory and Computation*, vol. 17, no. 9, pp. 5538–5555, 2021.

[26] K. Pernal, "Electron correlation from the adiabatic connection for multireference wave functions," *Physical review letters*, vol. 120, no. 1, p. 013001, 2018.

[27] M. Hapka, A. Krzeminska, M. Modrzejewski, M. Przybytek, and K. Pernal, "Efficient calculation of the dispersion energy for multireference systems with cholesky decomposition: Application to excited-state interactions," *The Journal of Physical Chemistry Letters*, vol. 14, pp. 6895–6903, 2023.

[28] D. Drwal, P. Beran, M. Hapka, M. Modrzejewski, A. Sokół, L. Veis, and K. Pernal, "Efficient adiabatic connection approach for strongly correlated systems: Application to singlet–triplet gaps of biradicals," *The Journal of Physical Chemistry Letters*, vol. 13, no. 20, pp. 4570–4578, 2022.

[29] A. Ammar, E. Giner, and A. Scemama, "Optimization of Large Determinant Expansions in Quantum Monte Carlo," *J. Chem. Theory Comput.*, vol. 18, no. 9, pp. 5325–5336, Sep. 2022.

[30] S. Shepard, R. L. Panadés-Barrueta, S. Moroni, A. Scemama, and C. Filippi, "Double excitation energies from quantum monte carlo using state-specific energy optimization," *Journal of Chemical Theory and Computation*, vol. 18, no. 11, pp. 6722–6731, 2022. [Online]. Available: https://doi.org/10.1021/acs.jctc.2c00769

[31] K. Nakano, O. Kohulak, A. Raghav, M. Casula, and S. Sorella, "TurboGenius: Python suite for high-throughput calculations of *ab initio* quantum Monte Carlo methods," *in preparation*, 2023.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

A Reference data for test cases

Table 2: The comparison of the LDA energies obtained by PySCF and those obtained by TurboRVB with the same basis sets for 38 molecules. The TREXIO files used for the sanity check are available from our GitHub repository.

Molecules	Basis set	LDA energy (Ha) [PySCF]	LDA energy (Ha) [TurboRVB]	Difference (Ha)
H_2	ccecp-ccpv5z	-1.052985	-1.052981	+0.000004
LiH	ccecp-ccpvqz	-0.766956	-0.766955	+0.000001
Li_2	ccecp-ccpvqz	-0.424899	-0.424898	+0.000000
CH_4	ccecp-ccpvqz	-7.999578	-7.999577	+0.000002
H_2O	ccecp-ccpvqz	-17.149230	-17.149231	-0.000002
HF	ccecp-ccpvqz	-24.789625	-24.789628	-0.000003
NH_3	ccecp-ccpvqz	-11.649519	-11.649520	-0.000001
LiF	ccecp-ccpvqz	-24.488407	-24.488394	+0.000013
C_2H_2	ccecp-ccpvqz	-12.387838	-12.387840	-0.000002
CO	ccecp-ccpvqz	-21.599002	-21.599006	-0.000004
HCN	ccecp-ccpvqz	-16.080629	-16.080632	-0.000003
N_2	ccecp-ccpvqz	-19.782359	-19.782358	+0.000001
C_2H_4	ccecp-ccpvqz	-13.619256	-13.619257	-0.000001
H_2CO	ccecp-ccpvqz	-22.767856	-22.767858	-0.000002
C_2H_6	ccecp-ccpvqz	-14.834373	-14.834374	-0.000001
F_2	ccecp-ccpvqz	-48.234293	-48.234293	-0.000000
H_2O_2	ccecp-ccpvqz	-33.029733	-33.029733	+0.000001
H_3COH	ccecp-ccpvqz	-23.966120	-23.966118	+0.000001
N_2H_4	ccecp-ccpvqz	-22.090553	-22.090549	+0.000003
CO_2	ccecp-ccpvqz	-37.641550	-37.641554	-0.000004
SiH_2-singlet	ccecp-ccpvqz	-4.949041	-4.949041	+0.000001
H_2S	ccecp-ccpvqz	-11.301548	-11.301539	+0.000009
HCl	ccecp-ccpvqz	-15.489788	-15.489782	+0.000005
PH_3	ccecp-ccpvqz	-8.264345	-8.264338	+0.000007
SiH_4	ccecp-ccpvqz	-6.189380	-6.189385	-0.000005
CS	ccecp-ccpvqz	-15.691692	-15.691695	-0.000003
SiO	ccecp-ccpvqz	-19.851557	-19.851567	-0.000011
CH_3Cl	ccecp-ccpvqz	-22.317235	-22.317234	+0.000001
ClF	ccecp-ccpvqz	-39.027754	-39.027747	+0.000007
H_3CSH	ccecp-ccpvqz	-18.134103	-18.134100	+0.000003
HOCl	ccecp-ccpvqz	-31.405201	-31.405195	+0.000006
SO_2	ccecp-ccpvqz	-42.089335	-42.089337	-0.000002
Na_2	ccecp-ccpvqz	-0.399062	-0.399064	-0.000002
NaCl	ccecp-ccpvqz	-15.171508	-15.171512	-0.000005
P_2	ccecp-ccpvqz	-13.030309	-13.030299	+0.000010
Cl_2	ccecp-ccpvqz	-29.774661	-29.774649	+0.000012
Si_2H_6	ccecp-ccpvqz	-11.238912	-11.238919	-0.000007
Ne_2	ccecp-ccpv6z	-69.728581	-69.728586	-0.000004

Table 3: The comparison of the LDA energies obtained by PySCF and those obtained by TurboRVB with the same basis sets for 8 crystals with twist average ($4 \times 4 \times 4$). The TREXIO files used for the sanity check are available from our GitHub repository.

CODID	Crystals	Crystalsystem	Spacegroup	LDA energy (Ha) [PySCF]	LDA energy (Ha) [TurboRVB]	Difference (Ha)
1011097	alpha-SiO_2	trigonal	152	-108.021418	-108.021413	+0.000005
1526655	Si	cubic	227	-31.213676	-31.213692	-0.000016
2101499	Diamond	cubic	227	-45.398692	-45.398686	+0.000007
9008997	h-BN	hexagonal	194	-25.627859	-25.627856	+0.000003
9011660	w-BN	hexagonal	186	-25.640240	-25.640238	+0.000002
1534043	Si_3N_4	cubic	220	-206.216212	-206.216225	-0.000013
1010954	hp-SiO_2	cubic	227	-287.772846	-287.772842	+0.000005

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

9008667	LiF	cubic	225	-98.395453	-98.395446	+0.000007
1000053	MgO	cubic	225	-67.925608	-67.925614	-0.000006
9008830	AlAs	cubic	216	-33.435721	-33.435736	-0.000015

Table 4: The comparison of the LDA energies obtained by PySCF and those obtained by TurboRVB with the same basis sets for 4 crystals with twist average ($4 \times 4 \times 4$). The TREXIO files used for the sanity check are available from our GitHub repository.

CODID	Crystals	Crystalsystem	Spacegroup	LDA energy (Ha) [PySCF]	LDA energy (Ha) [TurboRVB]	Difference (Ha)
57408	Li-fcc	cubic	225	-29.573978	-29.573972	+0.000006
109012	Li-cl16	cubic	220	-118.109248	-118.109234	+0.000014
8104233	Na	cubic	229	-94.770109	-94.770104	+0.000005
2300202	Fe	cubic	229	-246.587802	-246.587783	+0.000019

Table 5: The comparison of the RHF energies obtained by PySCF and the VMC energies obtained by TurboRVB with the WFs converted from the corresponding PySCF checkpoints files via TREXIO for 100 molecules. The error bars refer to 3σ of the VMC calculations. The TREXIO files used for the sanity check are available from our GitHub repository.

Molecules	Charge	Spin	Basis set	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
H	0	1	ccecp-ccpvqz	-0.500000	-0.499992(18)	+0.000008(18)
He	0	0	ccecp-ccpvqz	-2.861679	-2.861677(39)	+0.000003(39)
Li	0	1	ccecp-ccpv5z	-0.196853	-0.1968528(15)	-0.0000001(15)
Be	0	0	ccecp-ccpv6z	-0.961893	-0.961896(31)	-0.000003(31)
B	0	1	ccecp-ccpv6z	-2.539305	-2.539324(32)	-0.000019(32)
C	0	2	ccecp-ccpv6z	-5.314319	-5.314307(33)	+0.000012(33)
N	0	3	ccecp-ccpv6z	-9.633866	-9.633874(31)	-0.000007(31)
O	0	2	ccecp-ccpv6z	-15.689783	-15.689774(34)	+0.000009(34)
F	0	1	ccecp-ccpv6z	-23.937918	-23.937922(37)	-0.000004(37)
Ne	0	0	ccecp-ccpv6z	-34.708819	-34.708814(28)	+0.000005(28)
Na	0	1	ccecp-ccpvqz	-0.186205	-0.1862031(56)	+0.0000019(56)
Mg	0	0	ccecp-ccpvqz	-0.788392	-0.788372(27)	+0.000020(27)
Al	0	1	ccecp-ccpvqz	-1.877084	-1.877101(27)	-0.000017(27)
Si	0	2	ccecp-ccpvqz	-3.672547	-3.672553(30)	-0.000006(30)
P	0	3	ccecp-ccpvqz	-6.340966	-6.340973(31)	-0.000007(31)
S	0	2	ccecp-ccpvqz	-9.920494	-9.920504(29)	-0.000010(29)
Cl	0	1	ccecp-ccpvqz	-14.691247	-14.691263(34)	-0.000016(34)
Ar	0	0	ccecp-ccpvqz	-20.779663	-20.779668(26)	-0.000005(26)
K	0	1	ccecp-ccpvqz	-27.934622	-27.934614(28)	+0.000009(28)
Ca	0	0	ccecp-ccpvqz	-36.349734	-36.349743(35)	-0.000009(35)
Sc	0	1	ccecp-ccpvtz	-46.122158	-46.122171(31)	-0.000013(31)
Ti	0	2	ccecp-ccpvtz	-57.598094	-57.598110(29)	-0.000016(29)
V	0	3	ccecp-ccpvtz	-70.898419	-70.898434(34)	-0.000015(34)
Cr	0	6	ccecp-ccpvtz	-86.048478	-86.048479(28)	-0.000001(28)
Mn	0	5	ccecp-ccpvdz	-103.120673	-103.120673(35)	-0.000000(35)
Fe	0	4	ccecp-ccpvtz	-122.572205	-122.572210(39)	-0.000005(39)
Co	0	3	ccecp-ccpvqz	-144.264994	-144.264997(55)	-0.000003(55)
Ni	0	2	ccecp-ccpvtz	-168.272851	-168.272814(51)	+0.000038(51)
Cu	0	1	ccecp-ccpvtz	-195.337049	-195.337048(69)	+0.000001(69)
Zn	0	0	ccecp-ccpvtz	-225.275021	-225.275017(70)	+0.000005(70)
Ga	0	1	ccecp-ccpvqz	-1.984210	-1.984208(32)	+0.000001(32)
Ge	0	2	ccecp-ccpvqz	-3.664901	-3.664907(26)	-0.000006(26)
As	0	3	ccecp-ccpvqz	-6.065876	-6.065870(38)	+0.000006(38)
Se	0	2	ccecp-ccpvqz	-9.149929	-9.149927(36)	+0.000002(36)
Br	0	1	ccecp-ccpvqz	-13.121474	-13.121470(29)	+0.000004(29)
Kr	0	0	ccecp-ccpvqz	-18.228060	-18.228079(28)	-0.000019(28)
H-	-1	0	ccecp-ccpvqz	-0.476733	-0.476736(37)	-0.000003(37)
He+	1	1	ccecp-ccpvqz	-1.999787	-1.999809(30)	-0.000022(30)
He-	-1	1	ccecp-ccpvqz	-2.251807	-2.251804(28)	+0.000004(28)

continued.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Molecules	Charge	Spin	Basis set	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
Li-	-1	0	ccecp-ccpvqz	-0.187667	-0.187681(29)	-0.000013(29)
LiH	0	0	ccecp-ccpvqz	-0.749666	-0.749652(31)	+0.000015(31)
BeH	0	1	ccecp-ccpvqz	-1.544205	-1.544220(30)	-0.000015(30)
Li_2	0	0	ccecp-ccpvqz	-0.399560	-0.399556(27)	+0.000004(27)
CH	0	1	ccecp-ccpvqz	-5.906189	-5.906203(35)	-0.000014(35)
CH_2-singlet	0	0	ccecp-ccpvqz	-6.523371	-6.523377(29)	-0.000007(29)
CH_2-triplet	0	2	ccecp-ccpvqz	-6.564203	-6.564222(27)	-0.000019(27)
NH	0	2	ccecp-ccpvqz	-10.212125	-10.212108(33)	+0.000017(33)
CH_3	0	1	ccecp-ccpvqz	-7.205901	-7.205875(45)	+0.000026(45)
NH_2	0	1	ccecp-ccpvqz	-10.821853	-10.821870(31)	-0.000017(31)
OH	0	1	ccecp-ccpvqz	-16.300322	-16.300335(30)	-0.000013(30)
CH_4	0	0	ccecp-ccpvqz	-7.847115	-7.847102(34)	+0.000013(34)
H_2O	0	0	ccecp-ccpvqz	-16.944934	-16.944924(32)	+0.000009(32)
HF	0	0	ccecp-ccpvqz	-24.597194	-24.597211(32)	-0.000017(32)
NH_3	0	0	ccecp-ccpvqz	-11.460900	-11.460893(30)	+0.000007(30)
LiF	0	0	ccecp-ccpvqz	-24.281560	-24.281576(25)	-0.000016(25)
CN	0	1	ccecp-ccpvqz	-15.088891	-15.088881(38)	+0.000010(38)
C_2H_2	0	0	ccecp-ccpvqz	-12.115888	-12.115899(27)	-0.000010(27)
CO	0	0	ccecp-ccpvqz	-144.264994	-144.264997(55)	-0.000003(55)
HCN	0	0	ccecp-ccpvqz	-15.781632	-15.781644(30)	-0.000012(30)
N_2	0	0	ccecp-ccpvqz	-19.463663	-19.463643(38)	+0.000020(38)
HCO	0	1	ccecp-ccpvqz	-21.805381	-21.805384(32)	-0.000002(32)
NO	0	1	ccecp-ccpvqz	-25.412957	-25.412974(30)	-0.000018(30)
C_2H_4	0	0	ccecp-ccpvqz	-13.329730	-13.329719(33)	+0.000011(33)
H_2CO	0	0	ccecp-ccpvqz	-22.431967	-22.431958(27)	+0.000009(27)
O_2	0	2	ccecp-ccpvqz	-31.420428	-31.420430(32)	-0.000002(32)
C_2H_6	0	0	ccecp-ccpvqz	-14.526751	-14.526748(29)	+0.000003(29)
F_2	0	0	ccecp-ccpvqz	-47.824583	-47.824585(30)	-0.000002(30)
H_2O_2	0	0	ccecp-ccpvqz	-32.606154	-32.606174(32)	-0.000020(32)
H_3COH	0	0	ccecp-ccpvqz	-23.609361	-23.609348(31)	+0.000012(31)
N_2H_4	0	0	ccecp-ccpvqz	-21.708394	-21.708387(31)	+0.000007(31)
CO_2	0	0	ccecp-ccpvqz	-37.113177	-37.113187(32)	-0.000010(32)
SiH_2-singlet	0	0	ccecp-ccpvqz	-4.852829	-4.852828(33)	+0.000001(33)
SiH_2-triplet	0	2	ccecp-ccpvqz	-4.845542	-4.845530(31)	+0.000012(31)
PH_2	0	1	ccecp-ccpvqz	-7.508716	-7.508714(30)	+0.000001(30)
SiH_3	0	1	ccecp-ccpvqz	-5.463594	-5.463585(31)	+0.000009(31)
H_2S	0	0	ccecp-ccpvqz	-11.132951	-11.132948(32)	+0.000004(32)
HCl	0	0	ccecp-ccpvqz	-15.320450	-15.320478(35)	-0.000028(35)
PH_3	0	0	ccecp-ccpvqz	-8.116063	-8.116043(30)	+0.000020(30)
SiH_4	0	0	ccecp-ccpvqz	-6.086629	-6.086614(38)	+0.000015(38)
CS	0	0	ccecp-ccpvqz	-15.402549	-15.402531(39)	+0.000018(39)
SiO	0	0	ccecp-ccpvqz	-19.550999	-19.551009(28)	-0.000010(28)
SO	0	2	ccecp-ccpvqz	-25.695157	-25.695170(38)	-0.000013(38)
ClO	0	1	ccecp-ccpvqz	-30.394450	-30.394447(38)	+0.000003(38)
CH_3Cl	0	0	ccecp-ccpvqz	-21.994015	-21.994012(33)	+0.000003(33)
ClF	0	0	ccecp-ccpvqz	-38.652034	-38.652008(33)	+0.000027(33)
H_3CSH	0	0	ccecp-ccpvqz	-17.809387	-17.809395(28)	-0.000007(28)
HOCl	0	0	ccecp-ccpvqz	-31.018384	-31.018365(28)	+0.000019(28)
SO_2	0	0	ccecp-ccpvqz	-41.493266	-41.493266(36)	-0.000000(36)
Na_2	0	0	ccecp-ccpvqz	-0.372231	-0.372232(28)	-0.000001(28)
NaCl	0	0	ccecp-ccpvqz	-14.993974	-14.993958(38)	+0.000016(38)
Si_2	0	0	ccecp-ccpvqz	-7.320352	-7.320365(30)	-0.000013(30)
P_2	0	0	ccecp-ccpvqz	-12.742485	-12.742494(23)	-0.000009(23)
S_2	0	2	ccecp-ccpvqz	-19.920105	-19.920101(35)	+0.000005(35)
Cl_2	0	0	ccecp-ccpvqz	-29.423830	-29.423800(37)	+0.000030(37)
Si_2H_6	0	0	ccecp-ccpvqz	-11.021000	-11.021016(31)	-0.000017(31)
K_2	0	0	ccecp-ccpvqz	-55.865096	-55.865107(33)	-0.000011(33)
Ne_2	0	0	ccecp-ccpv6z	-69.417548	-69.417569(55)	-0.000021(55)

continued.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Molecules	Charge	Spin	Basis set	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
O_2+	1	1	ccecp-ccpvqz	-30.940484	-30.940494(26)	-0.000010(26)
O_2++	2	0	ccecp-ccpvqz	-29.978460	-29.978464(27)	-0.000003(27)
O_2-	-1	3	ccecp-ccpvqz	-31.164295	-31.164306(27)	-0.000011(27)

end.

Table 6: The comparison of the UHF energies obtained by PySCF and the VMC energies obtained by TurboRVB with the WFs converted from the corresponding PySCF checkpoints files via TREXIO for 100 molecules. The error bars refer to 3σ of the VMC calculations. The TREXIO files used for the sanity check are available from our GitHub repository.

Molecules	Charge	Spin	Basis set	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
H	0	1	ccecp-ccpvqz	-0.500000	-0.500011(19)	-0.000011(19)
Li	0	1	ccecp-ccpv5z	-0.196853	-0.1968520(20)	+0.0000008(20)
B	0	1	ccecp-ccpv6z	-2.543510	-2.543520(37)	-0.000010(37)
C	0	2	ccecp-ccpv6z	-5.319856	-5.319848(35)	+0.000009(35)
N	0	3	ccecp-ccpv6z	-9.638600	-9.638594(31)	+0.000005(31)
O	0	2	ccecp-ccpv6z	-15.696937	-15.696941(30)	-0.000004(30)
F	0	1	ccecp-ccpv6z	-23.943069	-23.943065(28)	+0.000004(28)
Na	0	1	ccecp-ccpvqz	-0.186205	-0.1862063(70)	-0.0000013(70)
Al	0	1	ccecp-ccpvqz	-1.881133	-1.881133(29)	+0.000000(29)
Si	0	2	ccecp-ccpvqz	-3.677013	-3.677011(28)	+0.000002(28)
P	0	3	ccecp-ccpvqz	-6.341271	-6.341273(29)	-0.000001(29)
S	0	2	ccecp-ccpvqz	-9.926822	-9.926832(28)	-0.000010(28)
Cl	0	1	ccecp-ccpvqz	-14.697582	-14.697592(26)	-0.000011(26)
K	0	1	ccecp-ccpvqz	-27.934701	-27.934716(33)	-0.000016(33)
Sc	0	1	ccecp-ccpvtz	-46.124357	-46.124365(30)	-0.000008(30)
Ti	0	2	ccecp-ccpvtz	-57.614825	-57.614826(28)	-0.000001(28)
V	0	3	ccecp-ccpvtz	-70.892874	-70.892876(29)	-0.000002(29)
Cr	0	6	ccecp-ccpvtz	-86.048802	-86.048803(31)	-0.000002(31)
Mn	0	5	ccecp-ccpvdz	-101.744686	-101.744694(34)	-0.000008(34)
Fe	0	4	ccecp-ccpvtz	-122.654507	-122.654504(40)	+0.000002(40)
Co	0	3	ccecp-ccpvqz	-144.271266	-144.271273(55)	-0.000007(55)
Ni	0	2	ccecp-ccpvtz	-168.486062	-168.486075(67)	-0.000012(67)
Cu	0	1	ccecp-ccpvtz	-195.337588	-195.337586(65)	+0.000002(65)
Ga	0	1	ccecp-ccpvqz	-1.987373	-1.987371(33)	+0.000002(33)
Ge	0	2	ccecp-ccpvqz	-3.668481	-3.668463(26)	+0.000018(26)
As	0	3	ccecp-ccpvqz	-6.066117	-6.066098(37)	+0.000019(37)
Se	0	2	ccecp-ccpvqz	-9.155003	-9.155017(32)	-0.000014(32)
Br	0	1	ccecp-ccpvqz	-13.126572	-13.126580(28)	-0.000008(28)
He+	1	1	ccecp-ccpvqz	-1.999787	-1.999782(28)	+0.000005(28)
He-	-1	1	ccecp-ccpvqz	-2.253591	-2.253590(29)	+0.000001(29)
BeH	0	1	ccecp-ccpvqz	-1.544551	-1.544556(39)	-0.000005(39)
CH	0	1	ccecp-ccpvqz	-5.910949	-5.910948(28)	+0.000002(28)
CH_2-triplet	0	2	ccecp-ccpvqz	-6.570105	-6.570116(30)	-0.000011(30)
NH	0	2	ccecp-ccpvqz	-10.220726	-10.220740(39)	-0.000014(39)
CH_3	0	1	ccecp-ccpvqz	-7.210633	-7.210627(33)	+0.000006(33)
NH_2	0	1	ccecp-ccpvqz	-10.827123	-10.827132(35)	-0.000009(35)
OH	0	1	ccecp-ccpvqz	-16.305559	-16.305564(28)	-0.000006(28)
CN	0	1	ccecp-ccpvqz	-15.107056	-15.107047(29)	+0.000009(29)
HCO	0	1	ccecp-ccpvqz	-21.811238	-21.811259(39)	-0.000021(39)
NO	0	1	ccecp-ccpvqz	-25.421073	-25.421057(27)	+0.000017(27)
O_2	0	2	ccecp-ccpvqz	-31.444494	-31.444481(27)	+0.000013(27)
SiH_2-triplet	0	2	ccecp-ccpvqz	-4.847604	-4.847610(31)	-0.000006(31)
PH_2	0	1	ccecp-ccpvqz	-7.514646	-7.514636(23)	+0.000010(23)
SiH_3	0	1	ccecp-ccpvqz	-5.465299	-5.465287(30)	+0.000012(30)
SO	0	2	ccecp-ccpvqz	-25.712539	-25.712538(32)	+0.000001(32)
ClO	0	1	ccecp-ccpvqz	-30.402361	-30.402377(30)	-0.000016(30)
S_2	0	2	ccecp-ccpvqz	-19.935972	-19.935957(33)	+0.000015(33)

continued.

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

Molecules	Charge	Spin	Basis set	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
O_2+	1	1	ccecp-ccpvqz	-30.951269	-30.951288(25)	-0.000019(25)
O_2-	-1	3	ccecp-ccpvqz	-31.191031	-31.191034(29)	-0.000003(29)

end.

Table 7: The comparison of the HF energies obtained by PySCF and the VMC energies obtained by TurboRVB with the WFs converted from the corresponding PySCF checkpoints files via TREXIO for 9 crystals at $k = \Gamma$. The error bars refer to 3σ of the VMC calculations. The TREXIO files used for the sanity check are available from our GitHub repository.

CODID	Crystal	Crystalsystem	Spacegroup	Charge	Spin	HF energy (Ha) [pySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
1011097	alpha-SiO_2	trigonal	152	0	0	-106.707420	-106.70747(21)	-0.00005(21)
1526655	Si	cubic	227	0	0	-30.097653	-30.09766(15)	-0.00001(15)
2101499	Diamond	cubic	227	0	0	-43.928111	-43.92810(15)	+0.00001(15)
9008997	h-BN	hexagonal	194	0	0	-24.883084	-24.88315(13)	-0.00007(13)
9011660	w-BN	hexagonal	186	0	0	-24.341137	-24.34112(15)	+0.00002(15)
9008667	LiF	cubic	225	0	0	-97.762562	-97.76255(22)	+0.00001(22)
1000053	MgO	cubic	225	0	0	-66.958279	-66.95830(16)	-0.00003(16)
1528336	O	monoclinic	12	0	4	-62.927947	-62.92795(19)	-0.00001(19)
9008830	AlAs	cubic	216	0	0	-32.418408	-32.41850(15)	-0.00009(15)

Table 8: The comparison of the HF energies obtained by PySCF and the VMC energies obtained by TurboRVB with the WFs converted from the corresponding PySCF checkpoints files via TREXIO for 9 crystals at $k = (0.25, 0.25, 0.25)$. The error bars refer to 3σ of the VMC calculations. The TREXIO files used for the sanity check are available from our GitHub repository.

CODID	Crystal	Crystalsystem	Spacegroup	Charge	Spin	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
1011097	alpha-SiO_2	trigonal	152	0	0	-106.737229	-106.73730(24)	-0.00007(24)
1526655	Si	cubic	227	0	0	-30.474552	-30.47466(19)	-0.00010(19)
2101499	Diamond	cubic	227	0	0	-44.575966	-44.57597(14)	-0.00000(14)
9008997	h-BN	hexagonal	194	0	0	-25.494781	-25.49480(13)	-0.00001(13)
9011660	w-BN	hexagonal	186	0	0	-25.754344	-25.75435(14)	-0.00001(14)
9008667	LiF	cubic	225	0	0	-97.719914	-97.71994(25)	-0.00003(25)
1000053	MgO	cubic	225	0	0	-67.174487	-67.17452(18)	-0.00003(18)
1528336	O	monoclinic	12	0	4	-62.950168	-62.95008(16)	+0.00009(16)
9008830	AlAs	cubic	216	0	0	-32.697346	-32.69733(16)	+0.00002(16)

Table 9: The comparison of the HF energies obtained by PySCF and the VMC energies obtained by TurboRVB with the WFs converted from the corresponding PySCF checkpoints files via TREXIO for 9 crystals with twist average ($4 \times 4 \times 4$). The error bars refer to 3σ of the VMC calculations. The TREXIO files used for the sanity check are available from our GitHub repository.

CODID	Crystal	Crystalsystem	Spacegroup	Charge	Spin	HF energy (Ha) [PySCF]	HF energy (Ha) [TurboRVB]	Difference (Ha)
1011097	alpha-SiO_2	trigonal	152	0	0	-106.730695	-106.73079(34)	-0.00010(34)
1526655	Si	cubic	227	0	0	-30.496624	-30.496662(96)	-0.000038(96)
2101499	Diamond	cubic	227	0	0	-44.593443	-44.59338(19)	+0.00007(19)
9008997	h-BN	hexagonal	194	0	0	-25.372483	-25.372477(92)	+0.000006(92)
9011660	w-BN	hexagonal	186	0	0	-25.436420	-25.436378(98)	+0.000041(98)
9008667	LiF	cubic	225	0	0	-97.719191	-97.71901(36)	+0.00018(36)
1000053	MgO	cubic	225	0	0	-67.178425	-67.17845(26)	-0.00003(26)
1528336	O	monoclinic	12	0	4	-62.958163	-62.95806(14)	+0.00010(14)
9008830	AlAs	cubic	216	0	0	-32.711201	-32.71116(12)	+0.00004(12)

TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.