



HAL
open science

A Quasi Time-Reversible scheme based on density matrix extrapolation on the Grassmann manifold for Born-Oppenheimer Molecular Dynamics

Federica Pes, Étienne Polack, Patrizia Mazzeo, Geneviève Dusson, Benjamin Stamm, Filippo Lipparini

► **To cite this version:**

Federica Pes, Étienne Polack, Patrizia Mazzeo, Geneviève Dusson, Benjamin Stamm, et al.. A Quasi Time-Reversible scheme based on density matrix extrapolation on the Grassmann manifold for Born-Oppenheimer Molecular Dynamics. *Journal of Physical Chemistry Letters*, 2023, 14 (43), pp.9720-9726. 10.1021/acs.jpcllett.3c02098 . hal-04161077

HAL Id: hal-04161077

<https://hal.science/hal-04161077v1>

Submitted on 5 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Quasi Time-Reversible Scheme Based on Density Matrix Extrapolation on the Grassmann Manifold for Born–Oppenheimer Molecular Dynamics

Federica Pes, Étienne Polack, Patrizia Mazzeo, Geneviève Dusson, Benjamin Stamm, and Filippo Lipparini*



Cite This: *J. Phys. Chem. Lett.* 2023, 14, 9720–9726



Read Online

ACCESS |



Metrics & More

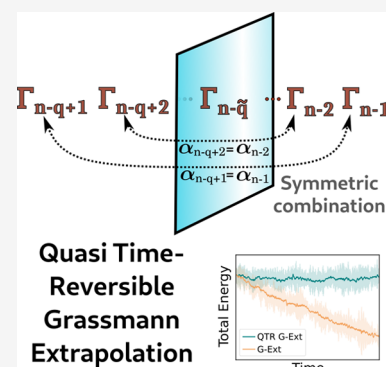


Article Recommendations



Supporting Information

ABSTRACT: This Letter introduces the so-called Quasi Time-Reversible scheme based on Grassmann extrapolation (QTR G-Ext) of density matrices for an accurate calculation of initial guesses in Born–Oppenheimer Molecular Dynamics (BOMD) simulations. The method shows excellent results on four large molecular systems that are representative of real-life production applications, ranging from 21 to 94 atoms simulated with Kohn–Sham (KS) density functional theory surrounded with a classical environment with 6k to 16k atoms. Namely, it clearly reduces the number of self-consistent field iterations while at the same time achieving energy-conserving simulations, resulting in a considerable speed-up of BOMD simulations even when tight convergence of the KS equations is required.



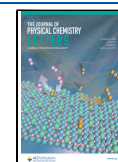
Ab initio Born–Oppenheimer Molecular Dynamics (BOMD) is a very powerful and versatile tool to simulate molecular processes in which the quantum nature of the system is not negligible. Unfortunately, this comes at a high computational price, which stems from the necessity of solving quantum mechanical (QM) equations, typically Kohn–Sham Density Functional Theory (KS-DFT) equations, to compute the energy and forces at every time step. Such equations are nonlinear and are solved using a fixed-point iterative method known as Self-Consistent Field¹ (SCF). Achieving SCF convergence typically requires, in a standard single-point run, up to 20 iterations, making the MD simulation very expensive, as in turn, the SCF has to be performed tens of thousands of times. Two main families of methods have been developed to address such a limitation. In extended Lagrangian methods, such as Car–Parrinello Molecular Dynamics (CPMD)² or Atom-centered Density Matrix Propagation (ADMP),³ the electronic degrees of freedom are propagated, thus avoiding the need of solving the SCF problem. This requires one to endow the electronic degrees of freedom with a fictitious mass that needs to be small enough to keep the trajectory close to its Born–Oppenheimer counterpart. As a consequence, rather short time steps need to be used. A different strategy relies on developing extrapolation techniques^{4–15} for BOMD that allow one to converge the SCF in a limited number of iterations. In this work, we choose the second strategy, which is particularly effective for calculations using localized basis sets, e.g., Gaussian-type orbitals. The extrapolation techniques used in BOMD use converged solutions from previous MD steps to

compute an accurate guess for the SCF, thus limiting the number of iterations required to achieve convergence. A significant contribution to this field was given by Niklasson and co-workers in 2006 with their work on the time-reversible extrapolation for Born–Oppenheimer Molecular Dynamics.¹² The core concept involves generating a guess density matrix by combining the density matrices from previous steps in a symmetric and time-reversible manner. However, numerical applications showed that enforcing an *exact* time-reversibility can lead to errors accumulating in long-time simulations, thus spoiling the convergence properties of the algorithm in the long run. This led to the development of the Extended Lagrangian Born–Oppenheimer approach (XLBO) in 2008.^{13–17} In this particular case, the time-reversible extrapolation is augmented by the inclusion of a dissipative term, which serves to reduce the numerical fluctuations. XLBO can be seen as an intermediate strategy between Car–Parrinello like approaches and extrapolation techniques for BOMD, as it indeed propagates an auxiliary density matrix that can either be used directly in a CPMD spirit,^{18,19} possibly after refining the density using an approximate SCF solver, or be

Received: July 28, 2023

Accepted: October 11, 2023

Published: October 25, 2023



used as a guess for the SCF.¹³ Here, we focus on the latter approach.

In Niklasson's XLBO scheme, the guess density is propagated over time, subject to a potential that forces it to be close to the converged density. The result is a guess density that is accurate enough to achieve reasonable SCF convergence (e.g., 10^{-5} RMS norm of the density matrix change) in as little as four iterations: Niklasson's pioneering work has therefore been crucial in extending the applicability of BOMD. However, the XLBO method suffers from a few shortcomings. First, the guess density obtained with XLBO is not exactly idempotent,¹³ unless it is postprocessed using, e.g., McWeeny purification.^{20,21} Second, its performance degrades if a tightly converged SCF solution is required, as is the case when a post-Hartree–Fock method is used to compute the energy and forces (e.g., in a time-dependent DFT excited-state simulation).

Recently, we proposed a different strategy to compute a guess density by using linear extrapolation. This is nontrivial, because in general a linear combination of density matrices does not preserve idempotency or, in other words, density matrices belong to a differentiable manifold called the Grassmann manifold and not to a vector space. Our approach uses tools from differential geometry to map the Grassmann manifold onto its tangent space, which is a vector space. It then performs a linear extrapolation on the tangent space and then maps back the extrapolated density to the manifold. We named such a method Grassmann extrapolation (G-Ext).^{21,22} G-Ext is an accurate and efficient strategy for ab initio MD simulations that has been shown to outperform XLBO, especially if a tight SCF convergence is required.²¹ It has been successfully adopted in the Pisa-group for both ground- and excited-state SCF-based BOMD simulations in a polarizable multiscale framework.^{23–26} Unfortunately, G-Ext suffers from a serious shortcoming. Numerical experiments have shown that the extrapolation introduces a bias causing a drift in the total energy for NVE simulations.²¹ Such an energy drift is modest in absolute terms (few kcal/mol in 10 ps to be compared with total energies of hundreds of thousands kcal/mol) but large if compared with the energetics of typical chemical processes. While using a tight convergence criterion for the SCF solves the problem,²¹ this is not an option for expensive, production simulations, thus limiting the gains introduced with the overall approach.

In this contribution, we not only address such a limitation by introducing a new strategy to perform the extrapolation but also further improve the performance of the method. We name the new strategy the Quasi Time-Reversible Grassmann extrapolation method (QTR G-Ext). This approach leverages the principles of differential geometry, similar to the previous method, but offers enhanced accuracy, improved performances, and excellent energy conservation properties. Given a N -dimensional atomic orbital (AO) basis, the SCF solves the following nonlinear eigenvalue problem, which consists of finding a matrix C and a diagonal matrix E such that

$$\begin{cases} F(D)C = SCE \\ C^T SC = I_N \\ D = CC^T \end{cases}$$

where $C \in \mathbb{R}^{N \times N}$ contains the N coefficients of the N occupied molecular orbitals, $D \in \mathbb{R}^{N \times N}$ is the density matrix,

$E \in \mathbb{R}^{N \times N}$ is a diagonal matrix in which the entries are the energy levels, F denotes the DFT operator, $S \in \mathbb{R}^{N \times N}$ is the overlap matrix, and I_N denotes the identity matrix of order N .

We assume that the density matrix is orthogonal. In any case, it can be transformed into such matrix by considering the Löwdin factorization of the overlap matrix S and consequently the modified coefficient matrix $\tilde{C} = S^{1/2}C$. Then, the normalized density matrix $\tilde{D} = \tilde{C}\tilde{C}^T = S^{1/2}DS^{1/2}$ belongs to the manifold

$$\mathcal{Gr}(N, N) = \{D \in \mathbb{R}^{N \times N} | D^2 = D = D^T, \text{Tr}(D) = N\}$$

which is isomorphic to the so-called “Grassmann manifold”; therefore, we identify \mathcal{Gr} by this name. From now on, we assume that the density matrix has been orthonormalized, and we denote it by D .

Since \mathcal{Gr} is a differential manifold, given a point $D_0 \in \mathcal{Gr}$, there exists a tangent space $\mathcal{T}_{D_0} \subset \mathbb{R}^{N \times N}$, such that tangent vectors $\Gamma(D) \in \mathcal{T}_{D_0}$ can be associated with nearby points $D \in \mathcal{Gr}$.

In MD, $t \rightarrow R(t)$ represents the trajectory of the nuclei. The transformation of the electronic structure can be interpreted as a trajectory denoted by $t \rightarrow D_{R(t)}$ on the manifold. In order not to burden the notation, we simply indicate D in place of $D_{R(t)}$. The objective is to determine a suitable approximation for the density matrix in the next step of the molecular dynamics trajectory by extrapolating the densities from previous steps. Since the tangent space \mathcal{T}_{D_0} is a vector space, we approximate the density matrix on \mathcal{T}_{D_0} . In order to solve the extrapolation problem, we decompose the mapping $R \rightarrow D$ as a composition of several maps

$$\begin{aligned} \mathbb{R}^{3M} &\rightarrow \mathcal{D} \rightarrow \mathcal{T}_{D_0} \rightarrow \mathcal{Gr} \\ R &\mapsto d \mapsto \Gamma \mapsto D \end{aligned} \quad (1)$$

where the first function $R \rightarrow d$ is a map from atomic positions to molecular descriptors. Here, as a descriptor, we use the Coulomb matrix²⁷ $d \in \mathbb{R}^{N_{\text{QM}} \times N_{\text{QM}}}$,

$$(d)_{kl} = \begin{cases} 0.5z_k^{2.4} & k = l \\ \frac{z_k z_l}{\|R_k - R_l\|} & k \neq l \end{cases} \quad (2)$$

where N_{QM} is the number of atoms treated quantum mechanically and z_k and R_k denote the nuclear charge and the position of the k th atom, respectively. Note that other descriptors can also be considered. We will detail the crucial mapping of $d \rightarrow \Gamma$ below. The mapping of $\Gamma \rightarrow \text{Exp}(\Gamma) = D$ is the so-called Grassmann exponential which maps tangent vectors on \mathcal{T}_{D_0} to \mathcal{Gr} , and it is a locally bijective function in a neighborhood of D_0 . Its inverse $D \rightarrow \text{Log}(D) = \Gamma(D)$ is the Grassmann logarithm. These mappings are computed by means of singular value decomposition (SVD). For mathematical details, the interested reader is referred to refs 22, 28, and 29. In our method, during MD, we use a fixed reference point D_0 to construct the tangent space \mathcal{T}_{D_0} .

Let n be the current time step of the MD. Given previous q snapshots $\Gamma_{n-i} = \text{Log}(D_{n-i})$, for $i = 1, \dots, q$, the approximation of the density matrix representative on the tangent space is written as

$$\tilde{\Gamma}_n = -\Gamma_{n-q} + \sum_{i=1}^{\tilde{q}} \alpha_i (\Gamma_{n-i} + \Gamma_{n-q+i}) \quad (3)$$

where $\tilde{q} = q/2$ if q is even, while $\tilde{q} = (q-1)/2$ if q is odd. We remark that, if in eq 3, the term Γ_{n-q} is substituted by $\tilde{\Gamma}_{n-q}$, a “fully” time-reversible approach (instead of quasi time-reversible) is obtained. Numerical experiments with the fully time-reversible approach, which are reported in the Supporting Information (SI), showed good behavior for total energy conservation but, unfortunately, a strong increase in the number of performed SCF iterations. This is consistent with what has been observed by Niklasson and co-workers,¹⁴ who remark that exact time-reversibility under noisy conditions (e.g., not fully converged SCF) can lead to error accumulations and significantly worse SCF convergence.

The descriptors are involved in the computation of the coefficients $\alpha = [\alpha_1, \dots, \alpha_{\tilde{q}}]^T$ appearing in eq 3. Indeed, they are computed by solving the least-squares problem with Tikhonov regularization

$$\min_{\alpha \in \mathbb{R}^{\tilde{q}}} \left\{ \left\| d_n + d_{n-q} - \sum_{i=1}^{\tilde{q}} \alpha_i (d_{n-i} + d_{n-q+i}) \right\|^2 + \varepsilon^2 \|\alpha\|^2 \right\} \quad (4)$$

where $\|\cdot\|$ denotes the l^2 -norm and $\varepsilon > 0$ is the regularization parameter. Since the Coulomb matrix (2) is symmetric, in the above formula, d_i represents the vectorized Coulomb matrix considering the lower triangle. In matrix form, it corresponds to solving the following least-squares problem

$$\min_{\alpha \in \mathbb{R}^{\tilde{q}}} \left\| \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} A \\ \varepsilon I_{\tilde{q}} \end{bmatrix} \alpha \right\|^2 \quad (5)$$

where the vector $\mathbf{b} = d_n + d_{n-q}$ is padded with \tilde{q} zeroes, $A \in \mathbb{R}^{N_d \times \tilde{q}}$ is the matrix in which columns are defined as $A_{:,i} = d_{n-i} + d_{n-q+i}$, and $I_{\tilde{q}}$ is the identity matrix of order \tilde{q} . Then, the initial guess for the density matrix is obtained as the composition of the three maps in (1), where the second map $d \rightarrow \Gamma$ is given by (3). Note that if this second map denoted by f was linear, then the guess would be close to exact, namely

$$\begin{aligned} \Gamma_n &= f(d_n) \approx f\left(-d_{n-q} + \sum_{i=1}^{\tilde{q}} \alpha_i (d_{n-i} + d_{n-q+i})\right) \\ &= -f(d_{n-q}) + \sum_{i=1}^{\tilde{q}} \alpha_i [f(d_{n-i}) + f(d_{n-q+i})] \\ &= -\Gamma_{n-q} + \sum_{i=1}^{\tilde{q}} \alpha_i (\Gamma_{n-i} + \Gamma_{n-q+i}) = \tilde{\Gamma}_n \end{aligned} \quad (6)$$

After computing the coefficients α_i by solving (4) and the tangent vector $\tilde{\Gamma}_n$ by eq 3, we obtain the sought guess density matrix for the SCF iterative method as $\tilde{D}_n = \text{Exp}(\tilde{\Gamma}_n)$.

The number q of density matrices taken at previous steps and the value of the regularization parameter ε are chosen in a heuristic manner: we computed the error $\|\Gamma_n - \tilde{\Gamma}_n\|$ for different values of q and ε , specifically $q = 3, 4, \dots, 20$ and $\varepsilon = 0.001, 0.002, 0.005, 0.01, 0.02, 0.05$, and we selected the

combination (q, ε) corresponding to the minimal error. When the SCF convergence threshold is 10^{-5} , we found that good values are $q = 5$ and $\varepsilon = 0.005$, while if it is fixed at 10^{-7} , we found $q = 4$ and $\varepsilon = 0.001, 0.002$. Additional details on the selection of q and ε values can be found in Section S1. The computational cost to compute the extrapolation coefficients α is negligible compared to the time for a single MD step. Thanks to the symmetric property of the coefficients, the size of the system (5) is $(N_d + \tilde{q}) \times \tilde{q}$, and \tilde{q} is a small number (in our simulations $\tilde{q} = 2$, as $q = 4$ or $q = 5$).

The QTR G-Ext approach is tested on four different systems. The first system is dimethylaminobenzonitrile (DMABN) in methanol. The second system is 3-hydroxyflavone (3HF) in acetonitrile. The last two systems (OCP and AppA) are chromophores embedded in a biological matrix, namely, a carotenoid in the orange carotenoid protein (OCP) and a flavin in the AppA Blue-Light Using Flavin photoreceptor.^{23,24,30} Some information about the systems is reported in Table 1.

Table 1. Summary of the Systems' Size: Number of QM atoms N_{QM} , Number of MM Atoms N_{MM} , Number of QM Basis Functions \mathcal{N} , Number of Occupied Orbitals N , and Size of Descriptors N_d

system	N_{QM}	N_{MM}	\mathcal{N}	N	N_d
DMABN	21	6843	185	39	234
3HF	28	15046	290	62	409
AppA	31	16449	309	67	468
OCP	94	6058	734	154	4468

KS-DFT has been adopted to describe the QM subsystem with the B3LYP hybrid functional³¹ and the 6-31G(d) Pople's basis set.³² This is coupled with a polarizable description of the environment using the AMOEBA force field.³³ For each system, we performed a QM/AMOEBA geometry optimization until a root-mean-square norm on the forces of 4 kcal/mol/Å was found, and finally, a 2 ps QM/AMOEBA NVT equilibration was performed to obtain the starting point of the simulations presented in this work.

All simulations have been performed using the Gaussian–Tinker interface.^{34–37} We implemented the QTR G-Ext extrapolation approach in Tinker.^{38,39}

To assess the quality of the guess density obtained by the QTR G-Ext extrapolation, we performed 10 ps BOMD simulations with a 0.5 fs time step in the NVE ensemble, using the velocity Verlet integrator.⁴⁰ All systems were tested with an SCF convergence threshold fixed to 10^{-5} and 10^{-7} with respect to the RMS variation of density. We compare our approach in terms of energy stability and number of iterations required to reach convergence with two other extrapolation schemes, which are the G-Ext scheme²¹

$$\tilde{\Gamma}_n = \sum_{i=1}^q \alpha_i \Gamma_{n-i}, \quad q = 6$$

where α_i is computed by solving

$$\min_{\alpha \in \mathbb{R}^q} \left\{ \left\| d_n - \sum_{i=1}^q \alpha_i d_{n-i} \right\|^2 + \varepsilon^2 \|\alpha\|^2 \right\}$$

where $\varepsilon = 0.01$, and XLBO^{13,14}

$$\tilde{D}_n = 2\tilde{D}_{n-1} - \tilde{D}_{n-2} + \kappa(D_{n-1} - \tilde{D}_{n-1}) + c \sum_{i=1}^8 \alpha_i \tilde{D}_{n-i}$$

with fixed parameters $\kappa = 1.86$, $c = 0.0016$, and $\alpha = (-36, 99, -88, 11, 32, -25, 8, -1)$.

Figure 1 provides the plot of the total energy along the DMABN simulation with a 10^{-5} SCF convergence threshold.

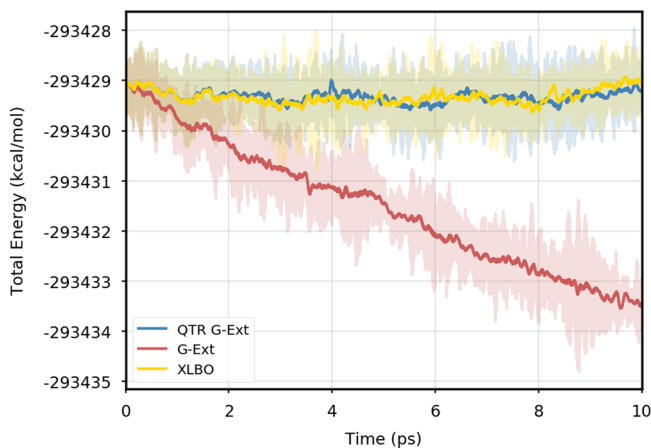


Figure 1. Total energy as a function of simulation time for DMABN using a 10^{-5} convergence threshold for the SCF.

Despite the nonfully time-reversible formulation of our newly implemented approach, we observe a great improvement with respect to the G-Ext scheme. In particular, the results obtained with the QTR G-Ext method resemble the ones obtained the XLBO scheme. The same behavior is almost imperceptible when the SCF convergence is set to 10^{-7} (Figure 2), since the

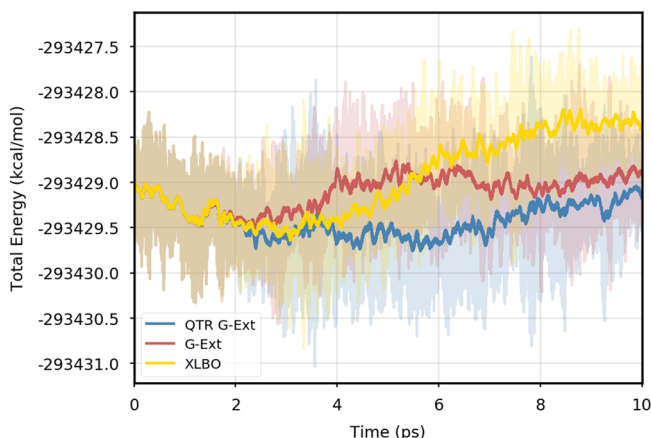


Figure 2. Total energy as a function of simulation time for DMABN using a convergence threshold for the SCF of 10^{-7} .

accumulation of errors that generates the energy drift when G-Ext is used is lower, so we can appreciate the same trend with all the extrapolation schemes. Analogous figures are reported in Section S2 for all tested systems. To better evaluate the energy stability, we consider the average short-time fluctuation (STF) of the energy, which is computed by getting the RMS of the energy fluctuation every 50 fs and averaging over the trajectory, and the long-time drift (LTD) for a long-time analysis, that is, the slope of the linear regression line of the energy. Tables 2 and 3 disclose STF and LTD for convergence thresholds of 10^{-5} and 10^{-7} , respectively. QTR G-Ext, G-Ext, and XLBO show comparable STF, which is specific for the system and is related to the time step for the integration. On the other hand, the absolute value of LTD is in general higher for 10^{-5} simulations, in particular for G-Ext. We can state that the QTR G-Ext method solves the energy-drift issue of G-Ext, showing an LTD that is always similar to the XLBO one, suggesting again a good time-reversible behavior.

The gain of our new methodology is not only in terms of accuracy (energy stability) but also in terms of the computational time of the simulation. Tables 4 and 5 report the average number of SCF iterations required to achieve convergence as well as the standard deviation for 10^{-5} and 10^{-7} SCF thresholds, respectively. We remark that each strategy requires q previous density matrices; before having them available, a standard SCF is performed. Therefore, for the computation of the average and standard deviation, we discard the first q points. The two tables show that for all the tested systems, the QTR G-Ext method requires the lowest number of SCF iterations for both convergence thresholds. Moving averages of SCF iteration numbers during the simulations for all systems and with both SCF convergence thresholds are reported in Section S2.

The performances of the QTR G-Ext guess are also maintained for larger and smaller time steps. We compared QTR G-Ext and XLBO for time steps of 0.1, 0.25, 0.75, and 1 fs by running MD simulations for the DMABN system with an SCF convergence threshold of 10^{-5} . All the results can be found in Section S3. Both methods show excellent energy conservation for the smaller time steps and afford reasonably stable simulations even for the larger ones, which is remarkable, as such simulations employing a time step that is too large to accurately sample molecular vibrations involving protons and are in general very noisy. For all time steps, QTR G-Ext requires a smaller average number of SCF iterations than XLBO. Finally, we tested the method for a looser SCF convergence of 10^{-4} , again a value that should not be used for production applications, as the error on the SCF solution transfers to the forces, thus affecting the quality of the dynamics. The results are reported in Section S4. Again, good energy conservation is shown for both methods, and QTR G-Ext outperforms XLBO in terms of average number of SCF iterations required.

Table 2. Short- and Long-Time Stability Analysis of the QTR G-Ext, G-Ext, and XLBO Methods^a

	DMABN		3HF		AppA		OCP	
	STF	LTD	STF	LTD	STF	LTD	STF	LTD
QTR G-Ext	0.33	-0.01	0.62	-0.40	0.57	-0.08	0.36	-0.23
G-Ext	0.35	-0.43	0.61	-0.94	0.56	-0.93	0.38	-1.38
XLBO	0.32	0.01	0.57	-0.42	0.59	0.14	0.39	-0.28

^aSCF convergence threshold 10^{-5} .

Table 3. Short- and Long-Time Stability Analysis of the QTR G-Ext, G-Ext, and XLBO Methods^a

	DMABN		3HF		AppA		OCP	
	STF	LTD	STF	LTD	STF	LTD	STF	LTD
QTR G-Ext	0.37	0.01	0.59	-0.30	0.53	0.18	0.38	-0.16
G-Ext	0.33	0.04	0.60	-0.27	0.54	0.06	0.38	-0.20
XLBO	0.32	0.13	0.64	-0.37	0.56	0.06	0.38	-0.08

^aSCF convergence threshold 10^{-7} .Table 4. Performance of the QTR G-Ext Method Compared to the G-Ext Method and XLBO Algorithm^a

	DMABN		3HF		AppA		OCP	
	\bar{k}	σ	\bar{k}	σ	\bar{k}	σ	\bar{k}	σ
QTR G-Ext	3.04	0.22	2.98	0.21	3.00	0.02	2.96	0.31
G-Ext	3.55	0.85	3.16	0.69	3.03	0.54	2.91	0.41
XLBO	4.00	0.05	4.00	0.00	4.00	0.07	4.00	0.01

^aAverage \bar{k} and standard deviation σ of the SCF iterations. Convergence threshold 10^{-5} .Table 5. Performance of the QTR G-Ext Method Compared with the G-Ext Method and XLBO Algorithm^a

	DMABN		3HF		AppA		OCP	
	\bar{k}	σ	\bar{k}	σ	\bar{k}	σ	\bar{k}	σ
QTR G-Ext	5.42	0.69	5.42	0.80	5.37	0.84	4.86	0.83
G-Ext	7.33	0.63	6.96	0.79	6.56	0.75	5.83	0.87
XLBO	7.51	0.65	7.45	0.65	7.43	0.80	7.21	0.75

^aAverage \bar{k} and standard deviation σ of SCF iterations. Convergence threshold 10^{-7} .

In conclusion, we presented the Quasi Time-Reversible Grassmann Extrapolation scheme, a new extrapolation method for ab initio molecular dynamics that not only allows for energy-conserving simulations but also exhibits overall excellent performances. Our numerical tests, performed on large, complex systems described with a polarizable multiscale strategy and taken from real-life production applications, show that QTR G-Ext is able to provide a guess density to BOMD simulations that allows the convergence of the SCF procedure in about 3 iterations on average for convergence thresholds that are typical of ground-state production runs, which is a 25% gain with respect to the state-of-the-art XLBO method. Tighter convergences, which are required for, e.g., time-dependent DFT excited-state simulations, can also be achieved in as little as 5–6 iterations. Furthermore, our numerical tests show that the new method does not introduce any significant bias in the guess density and thus exhibits very good energy conservation properties. This can be clearly seen by comparing the long-term drift observed in simulations for the two different SCF convergence thresholds used in our tests: while the previous G-Ext method shows a sharp increase in the drift going from 10^{-7} to 10^{-5} SCF convergence threshold, this is not the case for the QTR G-Ext method. We stress here that, due to the cost of BOMD simulations, every gain in performance is important, as it can easily translate into hundreds or thousands of saved CPU hours. The QTR G-Ext method is easy to implement and does not introduce any significant computational overhead and represents therefore an effective strategy to extend the applicability of BOMD simulations to larger and more complex systems.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.3c02098>.

Determination of the optimal parameters, plots of the total energy and number of SCF iterations along the dynamics for all tested systems, and results for the time step dependence and lower SCF convergence threshold (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Filippo Lipparini – Dipartimento di Chimica e Chimica Industriale, Università di Pisa, 56124 Pisa, Italy;
orcid.org/0000-0002-4947-3912;
 Email: filippo.lipparini@unipi.it

Authors

Federica Pes – Dipartimento di Chimica e Chimica Industriale, Università di Pisa, 56124 Pisa, Italy;
orcid.org/0000-0001-9064-7876
 Étienne Polack – CERMICS, 77455 Marne-la-Vallée, France
 Patrizia Mazzeo – Dipartimento di Chimica e Chimica Industriale, Università di Pisa, 56124 Pisa, Italy;
orcid.org/0000-0002-7015-8124
 Geneviève Dussan – Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, Université de Franche-Comté, 25030 Besançon, France
 Benjamin Stamm – Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, 70569 Stuttgart, Germany; orcid.org/0000-0003-3375-483X

Complete contact information is available at:
<https://pubs.acs.org/doi/10.1021/acs.jpcllett.3c02098>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Italian Ministry of University and Research under grant 2020HTSXMA_002 (PSI-MOVIE) and by the French 'Investissements d'Avenir' program, project Agence Nationale de la Recherche (ISITE-BFC) (contract ANR-15-IDEX-0003). É.P. also acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367—project EMC2) as well as from the Simons Targeted Grant Award No. 896630. F.P. is a member of the GNCS group of INdAM. We thank the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for supporting this work by funding - EXC2075 - 390740016 under Germany's Excellence Strategy. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

REFERENCES

- (1) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- (2) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- (3) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **2001**, *114*, 9758–9763.
- (4) Fang, J.; Gao, X.; Song, H.; Wang, H. On the Existence of the Optimal Order for Wavefunction Extrapolation in Born-Oppenheimer Molecular Dynamics. *J. Chem. Phys.* **2016**, *144*, 244103.
- (5) Alfè, D. Ab Initio Molecular Dynamics, a Simple Algorithm for Charge Extrapolation. *Comput. Phys. Commun.* **1999**, *118*, 31–33.
- (6) Arias, T. A.; Payne, M. C.; Joannopoulos, J. D. Ab Initio Molecular-Dynamics Techniques Extended to Large-Length-Scale Systems. *Phys. Rev. B* **1992**, *45*, 1538–1549.
- (7) Pulay, P.; Fogarasi, G. Fock Matrix Dynamics. *Chem. Phys. Lett.* **2004**, *386*, 272–278.
- (8) Herbert, J. M.; Head-Gordon, M. Accelerated, Energy-Conserving Born-Oppenheimer Molecular Dynamics via Fock Matrix Extrapolation. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3269–3275.
- (9) Hutter, J.; Parrinello, M.; Vogel, S. Exponential Transformation of Molecular Orbitals. *J. Chem. Phys.* **1994**, *101*, 3862–3865.
- (10) VandeVondele, J.; Hutter, J. An Efficient Orbital Transformation Method for Electronic Structure Calculations. *J. Chem. Phys.* **2003**, *118*, 4365–4369.
- (11) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (12) Niklasson, A. M. N.; Tymczak, C. J.; Challacombe, M. Time-Reversible Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.* **2006**, *97*, 123001.
- (13) Niklasson, A. M. N. Extended Born-Oppenheimer Molecular Dynamics. *Phys. Rev. Lett.* **2008**, *100*, 123004.
- (14) Niklasson, A. M. N.; Steneteg, P.; Odell, A.; Bock, N.; Challacombe, M.; Tymczak, C. J.; Holmström, E.; Zheng, G.; Weber, V. Extended Lagrangian Born-Oppenheimer Molecular Dynamics with Dissipation. *J. Chem. Phys.* **2009**, *130*, 214109.
- (15) Vitale, V.; Dziedzic, J.; Albaugh, A.; Niklasson, A. M. N.; Head-Gordon, T.; Skylaris, C.-K. Performance of extended Lagrangian schemes for molecular dynamics simulations with classical polarizable force fields and density functional theory. *J. Chem. Phys.* **2017**, *146*, 124115.
- (16) Niklasson, A. M. N. Next generation extended Lagrangian first principles molecular dynamics. *J. Chem. Phys.* **2017**, *147*, 054103.
- (17) Niklasson, A. M. N. Density-Matrix Based Extended Lagrangian Born-Oppenheimer Molecular Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 3628–3640.
- (18) Niklasson, A. M. N. Extended Lagrangian Born-Oppenheimer molecular dynamics using a Krylov subspace approximation. *J. Chem. Phys.* **2020**, *152*, 104103.
- (19) Kulichenko, M.; Barros, K.; Lubbers, N.; Fedik, N.; Zhou, G.; Tretiak, S.; Nebgen, B.; Niklasson, A. M. N. Semi-Empirical Shadow Molecular Dynamics: A PyTorch Implementation. *J. Chem. Theory Comput.* **2023**, *19*, 3209–3222.
- (20) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.
- (21) Polack, É.; Dusson, G.; Stamm, B.; Lipparini, F. Grassmann extrapolation of density matrices for Born-Oppenheimer molecular dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 6965–6973.
- (22) Polack, É.; Mikhalev, A.; Dusson, G.; Stamm, B.; Lipparini, F. An Approximation Strategy to Compute Accurate Initial Density Matrices for Repeated Self-Consistent Field Calculations at Different Geometries. *Mol. Phys.* **2020**, *118*, No. e1779834.
- (23) Bondanza, M.; Cupellini, L.; Faccioli, P.; Mennucci, B. Molecular Mechanisms of Activation in the Orange Carotenoid Protein Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 21829–21841.
- (24) Hashem, S.; Macaluso, V.; Nottoli, M.; Lipparini, F.; Cupellini, L.; Mennucci, B. From crystallographic data to the solution structure of photoreceptors: the case of the AppA BLUF domain. *Chem. Sci.* **2021**, *12*, 13331–13342.
- (25) Mazzeo, P.; Hashem, S.; Lipparini, F.; Cupellini, L.; Mennucci, B. Fast Method for Excited-State Dynamics in Complex Systems and Its Application to the Photoactivation of a Blue Light Using Flavin Photoreceptor. *J. Phys. Chem. Lett.* **2023**, *14*, 1222–1229.
- (26) Nottoli, M.; Bondanza, M.; Mazzeo, P.; Cupellini, L.; Curutchet, C.; Loco, D.; Lagardère, L.; Piquemal, J.-P.; Mennucci, B.; Lipparini, F. QM/AMOEBA description of properties and dynamics of embedded molecules. *WIREs Comput. Mol. Sci.* e1674;
- (27) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.
- (28) Zimmermann, R. Manifold Interpolation and Model Reduction. *arXiv* 2019, arXiv:1902.06502; <http://arxiv.org/abs/1902.06502>.
- (29) Edelman, A.; Arias, T. A.; Smith, S. T. The Geometry of Algorithms with Orthogonality Constraints. *SIAM J. Matrix Anal. Appl.* **1998**, *20*, 303–353.
- (30) Bondanza, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. The Multiple Roles of the Protein in the Photoactivation of Orange Carotenoid Protein. *Chem.* **2020**, *6*, 187–203.
- (31) Becke, A. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (32) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (33) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (34) Nottoli, M.; Mennucci, B.; Lipparini, F. Excited State Born-Oppenheimer Molecular Dynamics through a coupling between Time Dependent DFT and AMOEBA. *Phys. Chem. Chem. Phys.* **2020**, *22*, 19532–19541.
- (35) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.
- (36) Loco, D.; Lagardère, L.; Cisneros, G. A.; Scalmani, G.; Frisch, M.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings. *Chem. Sci.* **2019**, *10*, 7200–7211.

(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian Development Version*, Revision J.16; Gaussian, Inc.: Wallingford, CT, 2020.

(38) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(39) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; et al. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.

(40) Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.