

RESEARCH ARTICLE | AUGUST 06 2026

Variational adaptive Gaussian decomposition: Scalable quadrature-free time-sliced thawed Gaussian dynamics

Rahul Sharma  ; Amartya Bose  



J. Chem. Phys. 165, 054112 (2026)

<https://doi.org/10.1063/5.0333546>



Articles You May Be Interested In

Single-Hessian thawed Gaussian approximation

J. Chem. Phys. (April 2019)

Family of Gaussian wavepacket dynamics methods from the perspective of a nonlinear Schrödinger equation

J. Chem. Phys. (July 2023)

Internal conversion rates from the extended thawed Gaussian approximation: Theory and validation

J. Chem. Phys. (January 2023)

19 September 2026 09:00:53



The Journal of Chemical Physics

Special Topics Open for Submissions

[Learn More](#)

Variational adaptive Gaussian decomposition: Scalable quadrature-free time-sliced thawed Gaussian dynamics

Cite as: J. Chem. Phys. 165, 054112 (2026); doi: 10.1063/5.0333546

Submitted: 12 March 2026 • Accepted: 15 July 2026 •

Published Online: 6 August 2026



Rahul Sharma and Amartya Bose

AFFILIATIONS

Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai 400005, India

^{a)} Author to whom correspondence should be addressed: amartya.bose@tifr.res.in and amartya.bose@gmail.com

ABSTRACT

Time-slicing has emerged as a strategy for incorporating semiclassical propagation into real-time path integral formulation and recovering full quantum dynamics. A central step is the decomposition of a time-evolved wave function into a superposition of Gaussian wave packets (GWPs). Here, we introduce a quadrature-free variational framework for GWP decomposition, reformulating it as an optimization problem in which the GWP parameters are chosen to maximize the overlap with the time-evolving wave function. An autoencoder–decoder neural network is used for this optimization, with the representation being adaptively reoptimized during propagation. Each wave packet in this decomposition represents a localized patch of the underlying semiclassical manifold, while retaining full correlations between all degrees of freedom. This variational adaptive Gaussian decomposition (VAGD) approach yields a compact Gaussian expansion, providing a scalable route to time-sliced semiclassical quantum dynamics. While general, applying VAGD to facilitate time-slicing of thawed Gaussian dynamics allows a route to improving the semiclassical treatment to the full quantum mechanical result in a systematic manner.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0333546>

I. INTRODUCTION

The dynamics of molecular systems is governed by the time-dependent Schrödinger equation. However, its direct numerical solution becomes intractable as the number of degrees of freedom increases, as encountered in grid-based approaches such as the split-operator Fourier transform (SOFT) method,^{1,2} owing to the exponential growth of the Hilbert space. Formally exact wavefunction propagation methods, including the family of multiconfigurational time-dependent Hartree [(ML-)MCTDH] approaches,^{3,4} extend the accessible regime to systems with several tens of degrees of freedom but ultimately remain limited by the scaling of the underlying Hilbert space and specific forms of potential. A different class of approaches that has emerged includes the semiclassical methods, which exploits the classical structure to provide accurate and computationally tractable approximations to quantum dynamics.⁵

Among these, two broad classes can be identified: initial value representations based on stationary-phase approximations to the path integral, such as the van Vleck–Gutzwiller propagator and the Herman–Kluk formulation,^{6–10} and Gaussian wavepacket (GWP)

dynamics, such as the thawed Gaussian approximation (TGA).^{11–15} In TGA, a single Gaussian wavepacket is propagated with time-dependent center, width, and phase, enabling efficient on-the-fly *ab initio* simulations.^{16–20} However, the accuracy of such local semiclassical approximations typically deteriorates at long times due to the breakdown of the underlying harmonic expansion.

A natural route to improved accuracy is to represent the wavefunction as a superposition of Gaussian wavepackets. This idea underlies a wide range of approaches, including variational multiconfigurational Gaussian methods (vMCG) and Gaussian-based MCTDH formulations,^{21,22} which determine Gaussian parameters through the Dirac–Frenkel time-dependent variational principle and have been successfully applied to system–bath problems.^{23–25} Adaptive basis-growth strategies, such as multiple spawning methods, introduce new Gaussians dynamically to capture non-adiabatic effects,^{26,27} while trajectory-guided Gaussian bases and related approaches employ classical or semiclassical trajectories to construct efficient representations.²⁸ Matching-pursuit-based decompositions combined with SOFT propagation (MP/SOFT) further exploit Gaussian expansions to obtain efficient representations

of quantum dynamics of system–bath decomposed problems.^{29–31} Related approaches also include coupled coherent Gaussian (CCG) methods^{32,33} and trajectory-driven Gaussian basis techniques, which propagate Gaussian ensembles along classical or Ehrenfest-like trajectories²⁸ or guided by quantum trajectories.^{34–36} More recently, Gaussian dynamics has also been formulated variationally using Rothe’s method of propagation.^{37–39} While these approaches differ in their exact formulation, they share a common objective of constructing flexible, potentially time-dependent Gaussian bases capable of capturing quantum dynamics at a manageable cost.

An alternative strategy for improving long-time accuracy is time-slicing,^{40–42} in which the propagation is divided into short intervals over which semiclassical approximations remain valid. At the end of each interval, the wavefunction is re-expressed as a superposition of Gaussian wavepackets, which serve as initial conditions for subsequent propagation. In the limit of sufficiently fine slicing, this procedure converges to exact quantum dynamics. However, practical implementations face two key bottlenecks: (i) the need for high-dimensional quadrature to perform Gaussian decomposition and (ii) the rapid growth in the number of Gaussian components required for accurate representation. The former is particularly severe for oscillatory quantum wavefunctions, leading to the well-known Monte Carlo sign problem,^{43–45} which worsens dynamically as complex phases develop during time evolution. The second affects performance by increasing the number of classical trajectories that require simulation.

Time-sliced thawed Gaussian dynamics (TSTG) has demonstrated that highly accurate results can be obtained using quadrature-based Gaussian wavepacket decompositions, with each GWP being propagated using TGA.⁴² However, the exponential scaling of the number of Gaussian components with dimensionality remains a fundamental limitation for high-dimensional and *ab initio* applications. Unlike TSTG, related approaches, such as the basis expansion leaping multiconfiguration Gaussian (BEL-MCG) method,^{46–48} combine periodic re-expansion with evolving coefficients but do not evolve the Gaussian basis over time. The semiclassical evolution of the basis is often an efficient way of increasing the length of the segments in between the re-expansion steps without loss of accuracy.

In this work, we introduce a variational adaptive Gaussian decomposition (VAGD) framework that addresses these limitations by reformulating the Gaussian decomposition step in time-sliced dynamics as a global optimization problem. In particular, the parameters of the output Gaussian ensemble are determined variationally to maximize the overlap with the input wavefunction using a minimal number of wavepackets resulting without losing accuracy. This replaces the conventional quadrature-based decomposition with a nonlinear optimization over the space of Gaussian parameters, effectively converting time-sliced semiclassical dynamics into a quadrature-free, optimization based framework. The resulting problem is highly non-convex but can be efficiently handled in practice using an autoencoder–decoder neural network that maps an input Gaussian ensemble to a compact reparameterized representation. The systematic route of going from semiclassical to fully quantum mechanical results provided by a combination of VAGD-enabled time-sliced dynamics makes it attractive for complex multidimensional problems.

This paper is organized as follows: Time-sliced thawed Gaussian dynamics is reviewed in Sec. II A. The VAGD framework is introduced in Sec. II B, followed by numerical demonstrations in Sec. IV. We conclude the paper in Sec. V.

II. METHODS

A. Time-sliced thawed Gaussian dynamics

The time-evolution of an initial wave-function $|\psi_0\rangle$ of a d -dimensional system, under a Hamiltonian, $\hat{H} = \frac{1}{2} \sum_j \frac{\hat{p}_j^2}{M_j} + V(\{\hat{x}_j\})$, is given by $|\psi(t)\rangle = \exp(-\frac{i}{\hbar} \hat{H}t) |\psi_0\rangle$. Depending upon the particular semiclassical approximation used, the propagator, $\exp(-\frac{i}{\hbar} \hat{H}t)$, is represented in different ways using classical trajectories. Under the thawed Gaussian approximation (TGA),¹¹ the time-dependent wave function is presumed to be approximated as a GWP,

$$\langle \vec{x} | \psi(t) \rangle = \exp \left(\frac{i}{2\hbar} (\vec{x} - \vec{q}_t)^T A_t (\vec{x} - \vec{q}_t) + \frac{i}{\hbar} \vec{p}_t^T (\vec{x} - \vec{q}_t) + \frac{i}{\hbar} \gamma_t \right), \quad (1)$$

where under the local harmonic approximation, the real “position” and “momenta” coordinates, $\vec{q}_t$ and $\vec{p}_t$, respectively, satisfy the classical equations of motion,

$$\dot{\vec{q}}_t = M^{-1} \vec{p}_t, \quad (2)$$

$$\dot{\vec{p}}_t = -\vec{\nabla} V(\{\vec{q}_t\}), \quad (3)$$

where M is the diagonal mass matrix. The “width” matrix, A_t , and the generalized phase, γ_t , satisfy the following equations:

$$\dot{A}_t = -A_t M^{-1} A_t - \nabla^2 V(\{\vec{q}_t\}), \quad (4)$$

$$\dot{\gamma}_t = \frac{i\hbar}{2} \text{Tr}(M^{-1} A_t) + \frac{1}{2} \vec{p}_t^T M^{-1} \vec{p}_t - V(\{\vec{q}_t\}). \quad (5)$$

While exact for harmonic systems, owing to the impact of the non-zero higher-order derivatives of the anharmonic potentials, the quality of traditional TGA deteriorates with time even when the initial state is a Gaussian. Several improvements have been done on single GWP dynamics over the years motivated by using time-dependent variational principle^{12,15,49,50} and also local expansions beyond the harmonic approximation.⁵¹ Although these serve to increase the accuracy of the single GWP dynamics for anharmonic systems at greater computational costs, all of them are still approximations and cannot guarantee formal or numerical exactness in general. Even the improved accuracy of these methods eventually fails, although at longer times. As outlined originally in Ref. 40, this deterioration can be alleviated by recasting the semiclassical propagator in a time-sliced form,^{41,42} which can be viewed as a discrete time approximation to the full quantum time-evolution operator. The “short-time” propagators are approximated semiclassically.

The essential idea is to divide the total propagation time into short time segments, over which the local harmonic approximation is valid. Within each segment, TGA, or any other semiclassical method, provides an accurate approximation to the dynamics. At the end of these segments, the resulting wave function is re-expressed

as a superposition of Gaussians, which then serve as initial conditions for the semiclassical propagation in the next time segment. Thus, through an interleaving of the semiclassical propagation and decomposition into GWPs, one can approximate the fully quantum dynamics at increasing levels of accuracy. Ultimately, in the limit of infinitesimally short propagation segments between successive decompositions ($\tau_{\text{seg}} = N_{\text{seg}} \Delta t$), this time-sliced semiclassical dynamics method approaches exact quantum dynamics.

Methods such as vMCG also represent the wavefunction as a superposition of Gaussian wave packets, with both coefficients and basis parameters evolving continuously in time.^{21–25} In contrast, time-sliced semiclassical dynamics provides an alternative framework in which the individual wavepackets are propagated over short time segments using semiclassical dynamics, followed by a reconstruction step at the segment boundaries, where the wavefunction is re-expressed as a potentially different superposition of Gaussians, resulting in a simultaneously updated set of coefficients and a redefined Gaussian basis for the subsequent propagation. Within each propagation segment, the equations of motion, Eqs. (2)–(5), do not change the expansion coefficients. The time dependence within the segment enters through the independent thawed Gaussian propagation of the Gaussian parameters. However, because the Gaussian basis is nonorthogonal, the overlap matrix between different GWPs changes continuously during propagation. Consequently, even though no coupled equations of motion are solved for the expansion coefficients within a segment, the norm of the resulting superposition is generally not conserved under the independent propagation of the individual GWPs. A normalization step is, therefore, applied before the evaluation of observables. This is equivalent to a common time-dependent rescaling of all coefficients, while leaving their relative amplitudes and phases unchanged. Avoiding explicit coefficient equations of motion preserves the simplicity of the propagation scheme and allows the use of well-established thawed-Gaussian trajectory algorithms. Thus, unlike vMCG-type formulations, we do not solve coupled equations of motion for the expansion coefficients within a segment. This corresponds to a piecewise approximation to the full quantum propagator. The accuracy of this approximation is controlled by the duration of the propagation segments and the quality of the reconstruction step.

The renormalization step potentially can bring in uncontrolled error. In the present work, convergence of the result is assessed through systematic refinement of the simulation. Repeating simulations with shorter propagation segments, higher re-expansion frequency, larger numbers of GWPs, and thinner basis functions till the observables stop changing gives us a rigorous control on the error and a systematic approach to reducing it. It is, however, also possible to use the drift of the norm from unity as a diagnostic tool of the error accumulated in the semiclassical propagation with independent trajectories and could be useful in designing adaptive implementations using it as a trigger for earlier re-expansion.

One possible implementation of this re-expansion can be in terms of the over-complete coherent state basis. However, an immediate challenge that ensues is that the resultant multidimensional integrals are not amenable to Monte Carlo simulations owing to the accumulation of pure phase with time-propagation. A Husimi transformed expansion-based implementation of this time-sliced thawed Gaussian has recently been demonstrated.⁴² Despite its success, the Husimi-based TSTG remains quadrature-dependent and

scales poorly with system dimensionality, motivating the need for a quadrature-free, variational approach to GWP decomposition.

B. Variational adaptive Gaussian decomposition (VAGD)

We introduce a variational approach for constructing an optimized Gaussian representation of an arbitrary input wave function. In the context of VAGD with some Gaussian wavepacket dynamics, the normalized input wave function is written as a sum of N_{inp} Gaussians, $|\psi\rangle = \sum_{j=1}^{N_{\text{inp}}} c_j |\phi_j\rangle$. The goal is to find a new representation in terms of N_{out} Gaussians, $|\Psi\rangle = \sum_{j=1}^{N_{\text{out}}} C_j |\Phi_j\rangle$. Here, N_{out} is capped at a certain user-defined maximum number, K . An exact decomposition requires that the magnitude of the overlap of the normalized input and output wave functions, called the fidelity, $F = \frac{|\langle\psi|\Psi\rangle|}{\sqrt{\langle\psi|\psi\rangle\langle\Psi|\Psi\rangle}}$, be unity. We, therefore, develop a variational approach for maximizing this fidelity to a threshold value, F_{thresh} . The exact N_{out} that is used is adjusted adaptively, with the *optimal* representation being defined as the one that achieves the target fidelity using a minimum N_{out} .

The optimized representation is obtained through a variational procedure that minimizes a loss defined in terms of the fidelity,

$$\mathcal{L} = -\log(F) + (1 - F), \quad (6)$$

which is equivalent to maximizing the fidelity between an input wave function and its Gaussian expansion.

The particular form of the loss has been chosen to accelerate the optimization. The $-\log(F)$ term provides a sharp gradient when F is small, leading to faster optimization and faster convergence in the early stages of optimization when the initial representation may be far from optimal. At these stages, the $1 - F$ term does not contribute significantly to the gradient of the loss. On the other hand, for the larger values of F obtained toward the end of the process, both the $-\log(F)$ and the $1 - F$ terms contribute substantially to the gradient, yielding faster convergence. The optimization is implemented using an autoencoder-decoder neural network. Alternate approaches to optimization are possible, but were found to be significantly less efficient and stable in practice owing to the highly non-convex nature of the problem.⁵² (The optimization procedure requires several evaluations of fidelity, which is obtained as a result of multiple multi-dimensional integrals. However, because the wave functions are written as a superposition of GWPs, well-known closed-form analytical results can be derived, which are summarized in Appendix A.) Methods such as MP/SMART^{29–31} also perform similar optimizations in a sequential manner. However, VAGD leverages an autoencoder-decoder network to provide a general framework for performing a global optimization of the basis.

Figure 1 shows a schematic of the architecture, which maps the input wave function to the set of parameters defining the output representation via a low-dimensional “encoded” layer. The network is optimized such that the loss, which depends upon the input and the output representations, is minimized, ensuring the resulting parameters encode a faithful Gaussian decomposition of the input wave function.

It is important to emphasize that, unlike typical uses of autoencoder-decoder architectures^{53,54} in which the low-dimensional latent “encoded” space representation⁵⁵ is of interest,

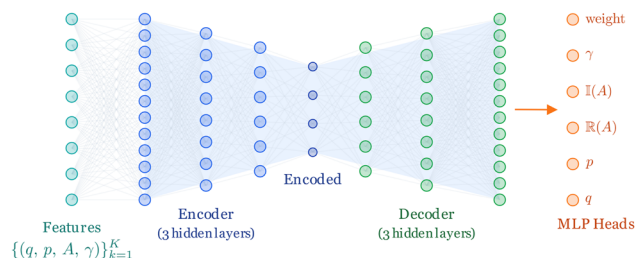


FIG. 1. Schematic of the autoencoder-decoder setup that acts as the basis transformation.

here the focus is on the optimized output of the decoder. This output, $|\Psi\rangle$, encodes a different but faithful representation of the input wave function, $|\psi\rangle$. In this framework, the neural network serves merely as a numerical optimizer for obtaining the parameters for the Gaussian decomposition. With this perspective, it is not necessary to train a generically applicable neural network. Instead, the network is optimized specifically for each wave function under consideration. A key point of this decomposition is that it is only dependent on the wave function and not on the underlying potential energy surface. In addition, because the overlap and the fidelity are constructed on a GWP-basis, we can get analytical expressions for them, making the entire procedure quadrature-free. (The expressions are given in Appendix A.)

An optimal representation is defined by the minimum number of Gaussians used (N_{out}) to achieve F_{thresh} , essential for reducing the number of classical trajectories. However, as end-users, we only set K , which is an upper limit to N_{out} . The algorithm starts with an initial guess of $N_{\text{out}} = N_{\text{start}}$. If upon optimization, the threshold fidelity, F_{thresh} , is not achieved, then the value of N_{out} is increased by a user-specified step size ΔN and the network is reoptimized. The process is repeated till either the threshold fidelity is successfully achieved or the upper limit of K is reached. We refer to this procedure as an adaptive basis expansion technique, which ensures a near optimal Gaussian decomposition. (Some important aspects of the encoding of the GWPs and the techniques employed for efficiency and numerical stability are outlined in Secs. II B 1 and II B 2, respectively. Full details of the particular version of the algorithm used in Sec. IV are provided in the [supplementary material](#).)

1. Encoding of parameters

To construct the input to the neural network and reconstruct the wave function from the neural network, it is necessary to specify the parameters. A generic d -dimensional GWP, Eq. (1), is parameterized by two d -dimensional real “phase space” vectors, $(\vec{q}, \vec{p})$, the $d \times d$ complex A matrix, and the single complex number, γ , representing the phase. Considering that the A matrix has to be symmetric, $\frac{d(d+1)}{2}$ complex numbers can uniquely encode it. Therefore, $d^2 + 3d + 2$ real numbers can be used to represent a single GWP. So, the input to the autoencoder-decoder has $N_{\text{inp}} \times (d^2 + 3d + 2)$ input parameters.

Now, the interpretation of the output is largely similar, with the $(\vec{q}, \vec{p})$ and γ retained. The main difference comes in the A matrix: the real part of A , A^{R} , must be symmetric, but the imaginary part

of A , A^{I} , which encodes the wave packet width, must additionally be positive definite. Directly optimizing A^{I} during training does not guarantee this property. To ensure positive definiteness by construction, we optimize a lower triangular matrix, L , with positive diagonal elements for each GWP. The width matrix of the k th Gaussian is finally reconstructed using a Cholesky decomposition,⁵⁶ $A_k^{\text{I}} = L_k L_k^T$. This parameterization makes the required constraints explicit. Since $(L_k L_k^T)^T = L_k L_k^T$, the resulting matrix A_k^{I} is symmetric. Furthermore, the diagonal entries of L_k are constrained to be strictly positive, so L_k is nonsingular. Therefore, for any nonzero vector x , $x^T A_k^{\text{I}} x = x^T L_k L_k^T x = |L_k^T x|^2 > 0$, which guarantees that A_k^{I} is positive definite by construction.

The input layer to the encoder, therefore, encodes the input wave function, and the output layer of the decoder encodes the output wave function. Because the autoencoder-decoder network optimizes the output in a manner that minimizes the loss between the input and output,^{53,54} in our case, we effectively get a new representation of the input wave function. The output of the encoder part of the neural network is an abstract low-dimensional latent space description of the wave function. The general principle works irrespective of the particulars of the neural network, and consequently, the network itself can also be treated as a “convergence” object.

For more difficult decompositions, the autoencoder-decoder neural-network output can also be used as the starting point for an additional first-order refinement. In this step, the decoded GWP parameters are optimized directly with respect to the same fidelity-based loss, while retaining the same parameterization and constraints. Thus, positive definiteness of the width matrix and any imposed width restrictions are preserved by construction. This optional second step of direct optimization provides a local refinement of the neural network output without altering the adaptive structure of the algorithm.

This completes the discussion on the basic version of the VAGD algorithm. There are some numerical considerations that help with optimizing the network faster. We discuss these subtleties in Sec. II B 2.

2. Numerical stabilization strategies

VAGD, as defined so far, allows for arbitrarily broad Gaussians to be incorporated in the decomposition, if required. While broad Gaussians may help minimize the number of trajectories, the local harmonic approximation is violated sooner for them than for the narrow wave packets, necessitating more frequent decompositions to maintain accuracy. This issue can be mitigated by putting a dimension-specific, user-defined cap on the width of each Gaussian, which effectively sets a lower limit on the diagonal values of imaginary parts of A_k , such that $(A_k^{\text{I}})_{j,j} \geq (W_{\text{min}})_j$ for all dimension, j . In our current implementation, the value of W_{min} is set empirically and held a constant for all times. This is not optimal because regions of lower anharmonicity should allow for a larger width than regions of greater anharmonicity. The algorithm can, therefore, be improved in the future by making the parameters W_{min} dependent on some measure of local anharmonicity.

A direct implementation of a VAGD decomposition step as described till now starts from a randomly initialized network, which increases the number of optimization cycles, or epochs, required for convergence to F_{thresh} . In addition, the position and the width of

the wave function can fluctuate significantly during time-evolution, which makes the numerical optimization of the neural network difficult. We will see that these two difficulties are closely related.

A natural way to address the randomly initialized starting network problem is to use a “warm-start” ansatz. The fundamental intuition is that between consecutive decomposition steps, the wave function may not have significantly changed in shape. Instead of starting from a random network, one can, therefore, reuse the optimized network from the previous step, potentially reducing the number of epochs of training required.

However, while the intuition holds for the shape of the wave function, its position can still change quite substantially between decomposition steps. As a result, a direct warm-start ansatz may still lead to unstable or ineffective optimizations. To stabilize the numerics and also maximize the effectiveness of the warm-start ansatz, the wave functions are “regularized” to some common scale.

This regularization consists of two preprocessing steps. First, the wave function is recentered to remove its average position. Second, the coordinate axes are rescaled so that the wave function has variances comparable with the initial state. The neural network is trained on this preprocessed wave function. After optimization, the regularization transformation is undone to recover the physical representation.

III. VAGD AS A ROUTE TO TIME-SLICED THAWED GAUSSIAN DYNAMICS

VAGD allows an accurate representation of a given wave function as a superposition of adaptive, optimally chosen GWPs. Many Gaussian wave packet-based methods require such a description. Here, we present the details of using VAGD to enable time-slicing of thawed Gaussian dynamics,⁴² a combination that we call VAGD-TG. The idea is to propagate an initial wave function, $|\psi(0)\rangle$, under a known Hamiltonian, H .

In addition to the time step for the classical trajectories (Δt) used in the thawed Gaussian simulations, a VAGD-TG simulation is also characterized by the following:

- The number of time steps between two decomposition steps (N_{seg}).
- The maximum number of Gaussians allowed for any of the decomposition step (K).
- The maximum allowed width of the individual GWPs making up the superposition. This is specified by the W_{min} parameter, which provides a lower bound on the imaginary part of the width matrix.
- The F_{thresh} , which is the minimum fidelity considered acceptable.

VAGD is applied to $|\psi(0)\rangle$ to represent it as a sum of N_1 GWPs, $|\psi(0)\rangle \approx \sum_{j=1}^{N_1} c_j^{(1)} |\phi_j^{(1)}\rangle$. Even if $|\psi(0)\rangle$ is itself a GWP, it might get broken down into several parts, especially if its width is larger than the threshold. For this step of decomposition, the autoencoder-decoder network has to be started from random values because there is no reference to be leveraged for a warm-start. For each of the N_1 GWPs obtained from the first decomposition, $\phi_j^{(1)}$, separate and independent thawed Gaussian trajectories are launched following Eqs. (2)–(5) for N_{seg} time steps.

Now, for the j th decomposition ($j \geq 2$), we collect and regularize all the propagated GWPs and utilize the network from the $j - 1$ th step as a warm-start ansatz, as described in Sec. II B 2. If the required fidelity cannot be reached with N_{j-1} GWPs, then it is increased in user-defined steps of ΔN till the threshold fidelity is crossed upon fitting. (Currently, we rerun the entire VAGD optimization using $N + \Delta N$ GWPs. However, one can easily think of doing a matching pursuit²⁹ on top of the initial VAGD result to incrementally improve it.)

If the previous decomposition step ($j - 1$) required an increase in the number of GWPs from the $j - 2$ th step, then we start with N_{j-2} to ensure the jump was not an optimization artifact. From the resulting decomposition, we start the independent thawed Gaussian simulations for N_{seg} time steps. Depending on the system under study, one can implement several very simple improvements. One possible avenue that has not been incorporated for simplicity is to also have an avenue to decrease the number of GWPs in the representation. In case the target fidelity is reached immediately, one could try lower number of GWPs. This would help in minimizing the number of trajectories. These interleaved thawed Gaussian propagation and VAGD decomposition steps continue till the final time of simulation is reached. (The full algorithm is summarized in Algorithm 1.)

The accuracy of the resultant dynamics obtained from the thawed Gaussian simulations is high in short time-scales and for thinner GWPs because the local harmonic condition is more valid across their length-scales. Therefore, by decreasing the maximum width of the GWPs and shortening the length of the trajectory segments, we can improve the accuracy of the semiclassical propagation for each GWP. However, if the wavepackets are thin, then a larger number of them would be required to describe the same wave function adequately, increasing the number of trajectories. Therefore, in the limit of $K \rightarrow \infty$, $W_{\text{min}} \rightarrow \infty$ (corresponding to arbitrarily thin Gaussians), and the time between two decompositions going to zero (or an infinite frequency of decomposition), VAGD-TG is formally guaranteed to converge to the true quantum result. Obviously, this increases the cost of computation and the goal is to get to this limit at the minimum computational complexity. The variational optimization ensures, for fixed K , the representation is optimal in the sense of maximizing fidelity.

Three factors go into considerations of the cost of a time-sliced thawed Gaussian dynamics simulation: (1) the cost of simulation of individual classical trajectories and Hessians for a system, (2) the number of classical trajectories required, and (3) the cost and number of decomposition steps required. As long as the system is the same, the cost of simulations of individual classical trajectories and Hessians should remain roughly constant across different methods of simulation. Then, the cost of a simulation is governed by the number of trajectories and the cost of the decomposition. As we shall demonstrate in Sec. IV, VAGD requires significantly (often orders of magnitude) fewer trajectories than the previous implementations of the time-sliced idea^{41,42} by leveraging a basis of GWPs with arbitrary phases and arbitrary widths within a maximal bound. Resolutions into coherent state basis typically use GWPs of constant width and phase as the bases functions. The matching pursuit algorithm initially used such coherent state bases.^{29–31} The last consideration is one of the cost of every decomposition. In case

ALGORITHM 1. The VAGD-TG Algorithm

```

Data:  $|\psi(0)\rangle, N_{\text{start}}, N_{\text{seg}}, K, F_{\text{thresh}}, t_{\text{final}}, \Delta t, \Delta N$ 
1  $n \leftarrow 0$ ; // Decomposition number
2  $t_{\text{prev}} \leftarrow 0$ ; // Segment begin time
3  $t_{\text{next}} \leftarrow 0$ ; // Segment end time
4  $N_{\text{traj}}^n \leftarrow 0 \forall n$ ; // Number of Gaussians required at the  $n$ th decomposition
5  $|\psi_{\text{in}}^{(0)}\rangle \xleftarrow{\text{normalize}} |\psi(0)\rangle$ ; // All fidelities, diagnostics, and observables use normalized wave functions
6 while  $t_{\text{prev}} \leq t_{\text{final}}$  do
7   if  $n = 0$  then
8      $N_{\text{out}} \leftarrow N_{\text{start}}$ ;
9   else
10    if  $\text{jump} == \text{false}$  then
11       $N_{\text{out}} \leftarrow N_{\text{traj}}^{(n-1)}$ ; // No jump, so use previous step's number
12    else
13      Take care of the possibility that the jump was a numerical optimization artifact:  $N_{\text{out}} \leftarrow N_{\text{traj}}^{(n-2)}$ ;
14      // Previous step had a jump. Use the number of GWPs in the  $n-2$ th step
15    end
16  end
17   $|\psi_{\text{in}}^{(n)}\rangle \xrightarrow[N_{\text{out}}]{\text{VAGD}} |\Psi_N^{(n)}\rangle = \sum_{j=1}^{N_{\text{out}}} c_j^{(n)} |\phi_j^{(n)}\rangle$ ; Compute the reconstruction fidelity  $F_n$ ; Warm start if  $n \geq 0$ ;
18  if  $F_n < F_{\text{thresh}}$  and  $n \geq 2$  then
19     $\text{jump} \leftarrow \text{true}$ ;
20  else
21     $\text{jump} \leftarrow \text{false}$ ;
22  end
23  while  $F_n < F_{\text{thresh}}$  and  $N_{\text{out}} < K$  do
24     $N_{\text{out}} \leftarrow \min(N_{\text{out}} + \Delta N, K)$ ;
25     $|\psi_{\text{in}}^{(n)}\rangle \xrightarrow[N_{\text{out}}]{\text{VAGD}} |\Psi_N^{(n)}\rangle = \sum_{j=1}^{N_{\text{out}}} c_j^{(n)} |\phi_j^{(n)}\rangle$ ; Compute the reconstruction fidelity  $F_n$ ;
26  end
27  Accept  $|\Psi_{\text{VAGD}}^{(n)}\rangle \leftarrow |\Psi_N^{(n)}\rangle$  as the compact basis,  $N_{\text{traj}}^{(n)} \leftarrow N_{\text{out}}$ ;
28   $t_{\text{next}} \leftarrow \min(t_{\text{final}}, t_{\text{prev}} + N_{\text{seg}}\Delta t)$ ;
29  TGA propagation of each accepted  $|\phi_j^{(n)}\rangle$  on  $[t_{\text{prev}}, t_{\text{next}}]$  and normalization is done while calculating observables;
30   $|\psi_{\text{in}}^{(n+1)}\rangle \xleftarrow{\text{Normalize}} |\psi_{\text{out}}^{(n)}\rangle$ ,  $n \leftarrow n + 1$ ,  $t_{\text{prev}} \leftarrow t_{\text{next}}$ ;
31 end

```

of VAGD, this would be governed by the complexity of the neural network involved. For simple, especially one-dimensional, cases, where the cost of the classical trajectories and estimation of the Hessians are inexpensive, the cost of the neural network becomes a significant proportion of the simulation. However, it is important to point out that the cost of calculating the Hessian is quite complicated. Especially for large dimensional problems, these estimations scale quadratically with the dimensionality of the problem. Consequently, the cost of the semiclassical propagation segments is expected to dominate over the cost of the decomposition for a complex system with large dimensionality. The decrease in the number of trajectories that the VAGD optimization enables becomes even more important for larger dimensional systems. Natural couplings of VAGD-enabled time-slicing with more “practical” approaches

such as single-Hessian TGA¹⁸ should make the simulations much faster. One can also envisage extensions to the single-Hessian idea, which would use a limited number of Hessian evaluations at carefully chosen points at the decomposition step. In addition, a time-sliced approach brings a new perspective to the nature of these approximations. While a single-Hessian TGA is an objectively less accurate approximation when compared to standard TGA, in the context of a time-sliced approach, it simply means a higher frequency of decompositions because the accuracy degrades faster. In principle, one should still be able to converge to fully quantum mechanical results. Evaluation of accuracy of such approaches will be evaluated in the future. Further work aimed toward optimizing the autoencoder-decoder setup is ongoing and would be instrumental in reducing the cost of this step. (A simple empirical

demonstration of the cost of our current implementation is shown in [Appendix C.](#))

IV. NUMERICAL ILLUSTRATIONS

We demonstrate the utility of VAGD in the context of VAGD-TG simulations on a variety of anharmonic systems and deep tunneling systems. The split-operator Fourier transform (SOFT)^{1,2} method is used to generate the numerically exact dynamics that will be the point of comparison for the numerical examples shown here. However, in the absence of exact reference data, VAGD-TG also provides internal convergence criteria. At each decomposition step, the fidelity threshold F_{thresh} ensures the wavefunction is accurately represented within the chosen Gaussian basis, thereby controlling the projection error. Convergence of the full dynamics can then be assessed by systematically increasing the control parameters, K and W_{min} , decreasing N_{seg} , and monitoring the stability of observables of interest. Once these observables become insensitive to further increases in computational effort, convergence is established. This two-tiered approach, fidelity-based control of the representation and observable-based convergence of the dynamics, provides a practical and systematic route to convergence in the absence of exact solutions.

The specific details for the particular neural network used in the examples are provided in the [supplementary material](#).

A. Morse potential

We start our exploration with the dynamics of wave packets in Morse potentials. In [Sec. IV A 1](#), we explore the convergence of VAGD-TG to the full quantum results under different levels of anharmonicity. Subsequently, in [Sec. IV A 2](#), we study the scaling of the algorithm by looking at the dynamics under multidimensional independent Morse oscillators.

1. 1D explorations

First, consider a 1D Morse oscillator,

$$V(x) = D_e [1 - \exp(-a(x - x_{\text{eq}}))]^2, \quad (7)$$

where D_e is the dissociation energy, a is the width of the well, and x_{eq} is the equilibrium bond distance. It is often more convenient to think in terms of the harmonic frequency at the bottom of the well, $\omega = a\sqrt{2D_e}$, and an anharmonicity factor, $\chi = \frac{\omega}{4D_e}$,¹⁸ which is zero for harmonic potentials. The parameters D_e and a can be simply obtained by inverting the relations.

In [Fig. 2](#), we show the overlap of the wave function evolved using the traditional TGA with the numerically exact SOFT simulations for different values of χ . Following the examples in [Ref. 18](#), the equilibrium bond length $x_{\text{eq}} = 20.95 a_0$ and $\omega = 900 \text{ cm}^{-1}$. The initial wave packet is taken to be a real Gaussian centered at $q_0 = 0$, with $A = \frac{i\omega_0}{\hbar}$, where $\omega_0 = 1000 \text{ cm}^{-1}$. The simulations were run with a time step of $\Delta t = 0.053 \text{ fs}$. The reference dynamics for this one-dimensional model was obtained from a converged SOFT calculation on a uniform grid over $x \in [-200, 200]$, with grid spacing $\Delta x = 0.025 \text{ a.u.}$ Consistent with our expectations,

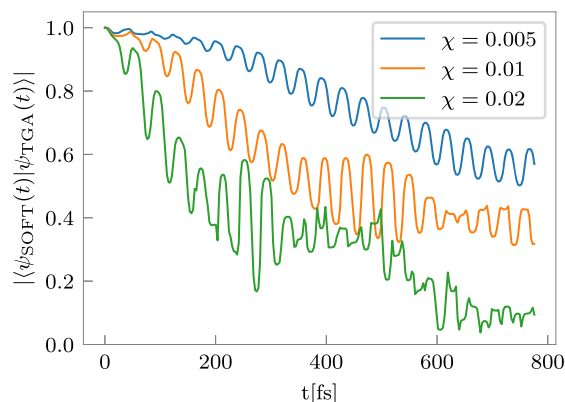


FIG. 2. Breakdown of TGA with anharmonicity of 1D Morse potential.

the quality of TGA deteriorates with time and this deterioration is faster for larger values of χ .

Now, we explore our VAGD method. We perform the variational decomposition with $N_{\text{seg}} = 50$, resulting in around 300 decompositions. At each decomposition boundary, the reconstructed state is required to satisfy $F_{\text{thresh}} = 0.9995$ with the propagated wavefunction. The width parameter, W_{min} , was set to 5.

Consider the case of $\chi = 0.005$ as a first example. In [Fig. 3](#), we examine the improvement of VAGD with increasing K . There is a clear and systematic improvement of the overlap with the SOFT result with the maximum number of GWPs allowed, K . This improvement is most clearly observed at longer times, as the quality of the semiclassical approximation gets worse. By $K \geq 6$, the VAGD-TG becomes practically identical to SOFT. For this example, we push the limits and claim convergence at $K = 8$. In addition, we demonstrate the effect of decreasing the length of the semiclassical segments and for thinner constituent GWPs shown in [Fig. 4](#) and how these parameters help achieve improved agreement with the exact quantum results in a systematic manner. On increasing the W_{min} parameter from 5 to 10, it was seen that just 8 GWPs cannot

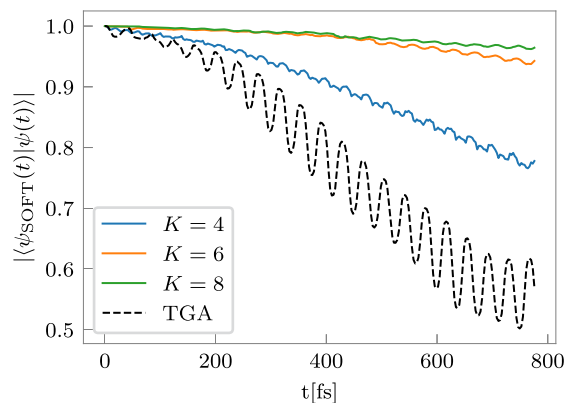


FIG. 3. Overlap of the VAGD-TG method with the SOFT wave function for different maximum number of Gaussians.

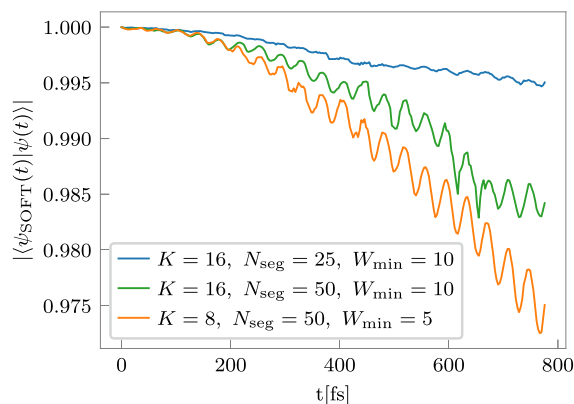


FIG. 4. Variation of accuracy with respect to width of clusters and segment lengths.

adequately describe the wave function over all times. Thus, it becomes necessary to increase K . Then, we see an improvement as a combination of these two effects. On decreasing the N_{seg} value from 50 to 25, we see a further improvement of the overlap. We explore how the VAGD algorithm places the GWPs at different time-points for this problem in Appendix B.

On increasing anharmonicity, while $\chi = 0.01$ converged with the same N_{seg} , $\chi = 0.02$ required more frequent decompositions, with $N_{\text{seg}} = 25$ amounting to around 600 decomposition steps. The number of decomposition steps can be decreased by using more number of thinner Gaussians or making the cap of width depend on the local curvature and departure from harmonic behavior. We show the overlap and the number of trajectories for the converged simulation in Fig. 5. The convergence gets more difficult with anharmonicity. One way to think about this is greater the anharmonicity, quicker is the breakdown of the local harmonic approximation for GWPs of the same width. Therefore, we need to perform the VAGD steps more frequently.

The number of trajectories, $N_{\text{traj}}(t)$, also shown in Fig. 5, grows with time, demonstrating the build-up of the effects of

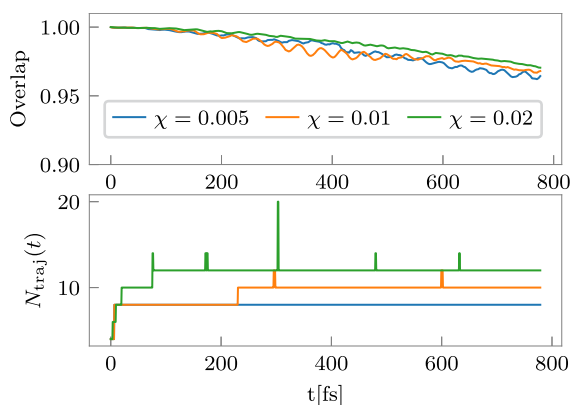


FIG. 5. Overlap of the VAGD-TG wave function with SOFT solution and number of VAGD trajectories of 1D-Morse oscillator for different values of χ .

anharmonicity and deviation from the basic thawed Gaussian ansatz. However, it is noteworthy that the number of trajectories required is relatively small in all the cases. This becomes especially important when running *ab initio* dynamics where trading the cost of a better decomposition for a smaller number of trajectories can prove to be extremely lucrative. For some of the runs, there are spikes in the $N_{\text{traj}}(t)$ values. These sudden jumps come from the optimizer getting stuck in local minima of the loss function. Step 14 in Algorithm 1 takes care of this possibility and uses the number of steps from the penultimate step to try the fitting. Moreover, depending on the physics of the problem at hand, the actual number of trajectories may be substantially less than the upper limit K provided. This is another benefit of the adaptive algorithm. We emphasize that checking the overlap with the numerically exact SOFT results is by far the most stringent test of the correctness of the method. It requires the approximate state to reproduce not only the spatial distribution but also the detailed phase structure of the exact quantum wave function. Therefore, even a moderate reduction in the overlap does not necessarily imply a comparable loss in the accuracy of physically relevant observables, which would converge with far fewer trajectories. This is demonstrated numerically in the supplementary material for the current case, where the auto-correlation spectrum for the $\chi = 0.005$ system is already converged at $K = 6$ despite the overlap not being tightly converged.

2. Multidimensional explorations

Pushing the semiclassical simulations to the full quantum levels of accuracy gets increasingly difficult with dimensionality. The time-slicing procedure, if done using quadrature, can very quickly become intractable. To study the scaling of VAGD with dimensions, consider the simple case of d uncoupled Morse oscillators as an example,

$$V(\vec{x}) = D \sum_{i=1}^d (1 - \exp[-a(x_i - x_{0,i})])^2, \quad (8)$$

where all the centers at origin ($x_{0,i} = 0.0$) and $\omega = 900 \text{ cm}^{-1}$ and $\chi = 0.005$ for all dimensions. While the scaling with dimensionality would of course be worse for cases where the different dimensions are correlated, this independent dimensional model allows us to study the scaling in a relatively simplified setting. The VAGD-TG simulation does not “know” that the dimensions can be separated and hence does not use that information explicitly. We will explore correlated multidimensional models in terms of the double-well potentials.

For the multidimensional Morse calculations, the initial wavefunction is taken as

$$\Psi(\mathbf{x}, 0) = N \exp \left[-\frac{\omega}{2\hbar} \sum_{i=1}^d \left(x_i + \sqrt{\frac{8}{d}} \right)^2 \right], \quad (9)$$

with zero initial momentum. We start by demonstrating the deterioration of the quality of the standard TGA simulation for different dimensionality of the system shown in Fig. 6. A time step of $\Delta t = 0.053 \text{ fs}$ was used for all the calculations. Because the dimensions are independent and the displacement of the initial wave packet along each dimension decreases with d , the errors in the single GWP TGA is less pronounced for larger d .

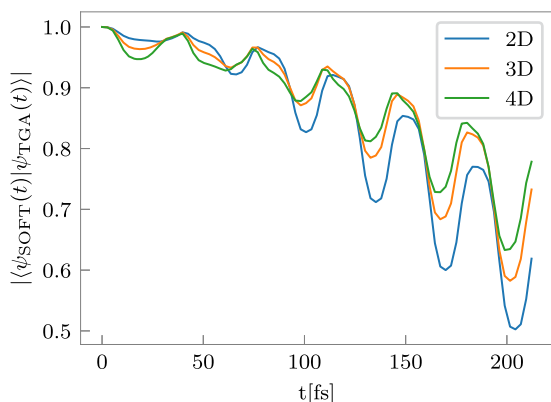


FIG. 6. Overlap of TGA solutions with the SOFT solution for multidimensional Morse models as a function of dimensionality.

Next, we perform the VAGD procedure every 25 steps. At each decomposition boundary, the VAGD-reconstructed state is required to satisfy $F_{\text{thresh}} = 0.9995$ with the propagated, and the W_{min} for individual Gaussian wave packets was set to 5.0 along each dimension. The results shown in Fig. 7 demonstrate the recovery of full quantum mechanical results. In all these cases, the number of trajectories saturated to the maximum limit allowed, K . However, there are a couple of things to be noted. First, the number of trajectories does not immediately converge to K . Depending upon the exact system setup, it can use less trajectories as well. This is the adaptive nature of the algorithm. The threshold K that is required is also a function of the duration of simulation. The longer a wave packet propagates in an anharmonic potential, the more non-Gaussian character it develops and the higher is the number of trajectories that are required.

Finally, in Fig. 8, we summarize the number of trajectories required to obtain the converged results at a final time of $t_{\text{final}} = 212$ fs. It is important, however, to interpret Fig. 8 only as an empirical summary of the present numerical experiments. Therefore, we do not make any broader claim regarding a universal

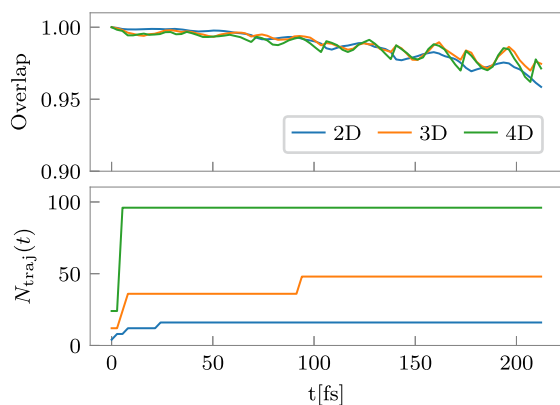


FIG. 7. Overlap of the VAGD-TG wave function with SOFT solution and number of trajectories for multidimensional Morse models as a function of dimensionality.

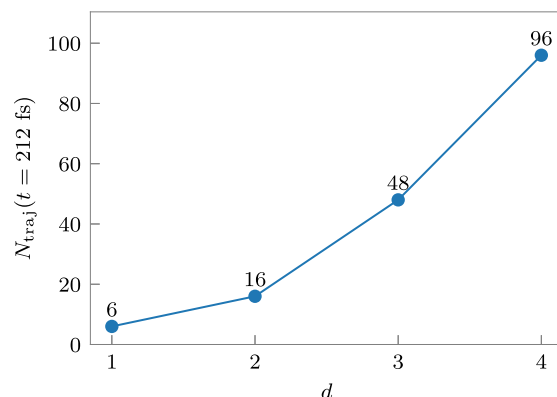


FIG. 8. Number of trajectories required for converged results at $t = 212$ fs for different numbers of dimensions.

scaling law from these data alone. Nevertheless, the trend provides a useful indication of how the complexity of the adaptive Gaussian representation grows for these multidimensional Morse calculations.

B. Double well potentials

As a next class of more complicated examples, we demonstrate the VAGD method on the double-well potential. These systems are dynamically strongly correlated in time, in the sense that no single classical trajectory can capture the true dynamics. The deep tunneling observed in these systems cannot be described by simple single trajectory methods and, therefore, single Gaussian ansatz fails. Time-sliced TGA has been shown to be able to account for the tunneling event.⁴² We demonstrate how VAGD is able to optimally represent the wave-function in one- and two-dimensions.

1. One-dimensional problem

Following Refs. 41 and 42, the 1D double well potential is defined to be

$$V_{\eta}(x) = \frac{x^4}{16\eta} - \frac{x^2}{2}, \quad \eta > 0. \quad (10)$$

There are two minima at $x_{\pm} = \pm 2\sqrt{\eta}$ separated by a barrier at $x = 0$. Consistent with the previous literature, we use $\eta = 1.3544$, so that the minima are located at $x_{\pm} \approx \pm 2.329$. We use an initial GWP localized in the left well,

$$\Psi(x, 0) = \pi^{-1/4} \exp\left(-\frac{1}{2}(x + 2\sqrt{\eta})^2\right), \quad (11)$$

with zero momentum. The potential and the probability distribution corresponding to $\Psi(x, 0)$ are shown in Fig. 9. The initial wave packet is completely localized in the left well with almost no penetration into the right well. In fact, there is negligible initial population on $x > 0$, with the probability of finding the particle at $x > 0$ being 4.97×10^{-4} . This setup can, therefore, be treated as a case of deep tunneling. It is also known that in the absence of time-slicing standard semiclassical techniques such as Herman–Kluk, we cannot capture

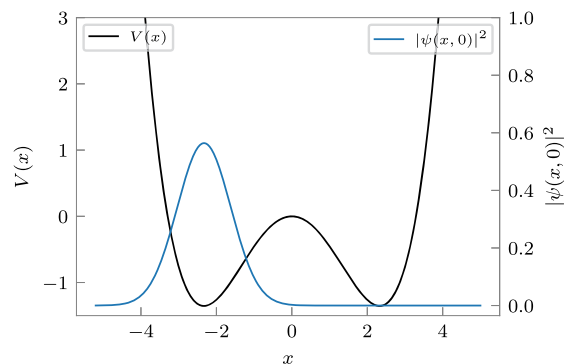


FIG. 9. Potential and initial state for the one-dimensional double-well.

the tunneling physics.⁴¹ Although some higher-order or variational single GWP dynamics methods^{49–51} may be able to qualitatively describe deep tunneling phenomena in some cases, in this particular case of deep tunneling, the variational Gaussian dynamics⁵⁰ fails. In contrast, both time-sliced Herman–Kluk⁴¹ and TGA⁴² can capture the physics in a numerically exact manner.

For the one-dimensional double-well calculations, the TGA trajectories were propagated with a time step of $\Delta t = 0.01$ a.u. The VAGD reconstruction was carried out every 20 time steps, corresponding to 750 decompositions, using $F_{\text{thresh}} = 0.9999$ and $W_{\text{min}} = 7.5$. The SOFT reference calculation used the same time step on a uniform grid $x \in [-8, 8]$ with spacing $\Delta x = 0.025$ a.u.

The shorter segment length, compared with the Morse examples, is required because the local harmonic approximation deteriorates more rapidly in the strongly anharmonic double-well potential. More frequent reconstruction is, therefore, needed to prevent the semiclassical error from accumulating between decomposition steps.

In addition to the overlap with SOFT results, following Ref. 42, we correlate the mirror-image of the initial wave function with the time-propagated one,

$$C(t) = \int dx \Psi^*(-x, 0) \Psi(x, t), \quad (12)$$

measuring the extent of tunneling captured by the various levels of approximation.

In Fig. 10, we show the overlap of the semiclassical solutions with the SOFT wave function [Fig. 10(a)] and the absolute value of the mirror correlation function, Eq. (12) [Fig. 10(b)]. The VAGD procedure used $W_{\text{min}} = 7.5$. From the overlaps, we see that the TGA result falters immediately, giving an overlap of almost 0.0 at $t \approx 75$ a.u., at which time the SOFT wavefunction has completely tunneled from the left to the right well. This first tunneling event corresponds to the maximum of $|C(t)|$ at around $t \approx 75$ a.u.. The subsequent lowering of the correlation function corresponds to the wave function tunneling back to the initial well.

Notice from Fig. 10(b) that the TGA wave function does not tunnel at all. This is expected because of its purely classical nature. It is interesting that in our VAGD-TG simulations, upto $K = 8$, the wave function does not really tunnel. As we move from $K = 8$ to $K = 9$, there is a jump in the amount of tunneling captured.

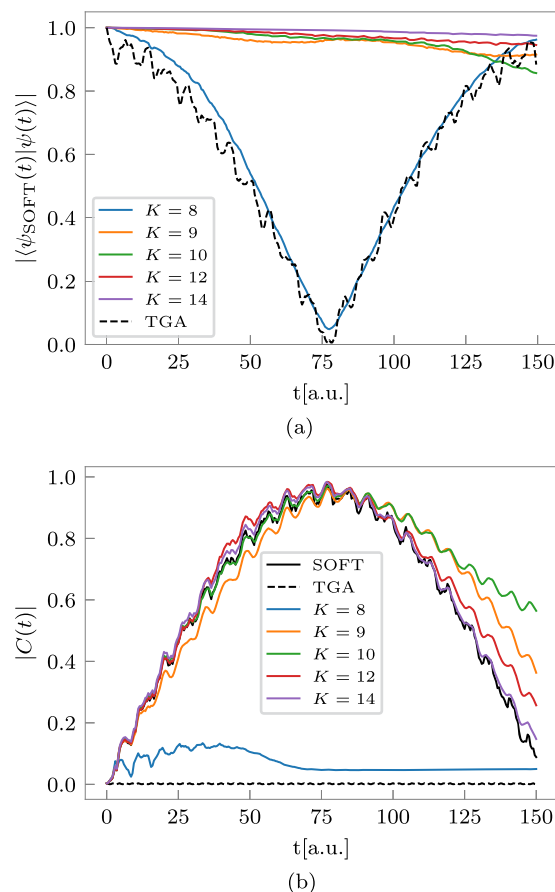


FIG. 10. Accuracy of the VAGD-TG scheme for the 1D double well. (a) Overlap of TGA and VAGD-TG wave functions with SOFT as a function of time. (b) Comparison of $|C(t)|$ at different levels of approximation.

Qualitatively, all the essential features of $|C(t)|$ are captured by a $K = 9$ calculation, although the quantitative agreement is still missing. It is the second tunneling event back to the original well that proves to be even more challenging. Using a maximum of $K = 14$ trajectories, we start to get quantitative agreement of the correlation function upto a maximum time of $t = 150$. This is a significant improvement over the original TSTG method,⁴² which used 2048 trajectories (without filtering), and the TS Herman–Kluk method,⁴¹ which used around 525 trajectories. VAGD-TG, in comparison, requires tens of trajectories to get the numerically exact result.

To explore the dramatic improvement observed between 8 and 9 trajectories, we try to understand how well the decomposition is happening. The fidelity achieved by the neural network tells us the accuracy of the output wave function as a representation of the input wave function. We plot this as a function of time in Fig. 11. With eight trajectories, we see that the wave function has a poor representation right from the first few steps of decomposition. Any error during the fitting procedure would have a cumulative effect in that the dynamics of the approximate wave function will over time deviate from the true result. So, the relatively low fidelity of

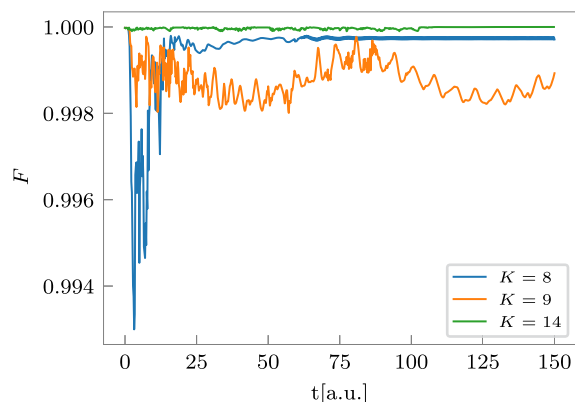


FIG. 11. Fidelity of fitting at different stages of dynamics.

the eight-trajectory run, along with the sustained period of this low fidelity, means that the physics is not going to be faithfully captured. We see that for the nine-trajectory run, the fidelities achieved are higher, leading to a better approximation of the tunneling dynamics. For the quantitatively correct case of 14-trajectories, the fidelities are practically unity throughout the run. Consequently, there is no drift from the true result. The physical intuition is that the exponentially decaying tail, which governs tunneling dynamics, is not captured properly at low numbers of GWPs. Once enough GWPs are used to

faithfully represent the tail, the tunneling is also reproduced properly. (It is also noticed that when the high target fidelities are not achieved, there is a much greater variance in the output dynamics with the starting guess of the neural network parameters. This variability goes away when the number of Gaussian wavepackets is large enough to give a faithful representation of the input wave function. This is demonstrated in the [supplementary material](#).)

A different perspective on this qualitative change can be obtained by looking at the agreement of the SOFT and the VAGD-TG wave functions at various time points. [Figure 12](#) shows these data for $K = 8, 9$, and 14 . Even at a time as short as $t = 3.40$ or 6.00 , we already see that $K = 8$ is unable to describe the smaller peaks along the right tail with proper accuracy. Now, tunneling is mediated through these tail features that extend into the classically forbidden regions. Consequently, the ability to describe this is crucial for capturing deep tunneling. As we proceed to longer times, we see that this and later representational error in $K = 8$ grow to limit its ability to show tunneling. $K = 9$ qualitatively gets all the features, while at $K = 14$, consistent with the overlaps, the dynamics is nearly quantitatively reproduced. The real and imaginary parts are also separately shown in the [supplementary material](#) (Figs. 7–9). The current discussion was all limited to the single-step variational optimization with the autoencoder–decoder neural network. When the two step process is used, direct optimization can significantly improve the quality of the representation even at low numbers of GWPs used. We show these results in the [supplementary material](#). However, even with this improvement, the essential physics and

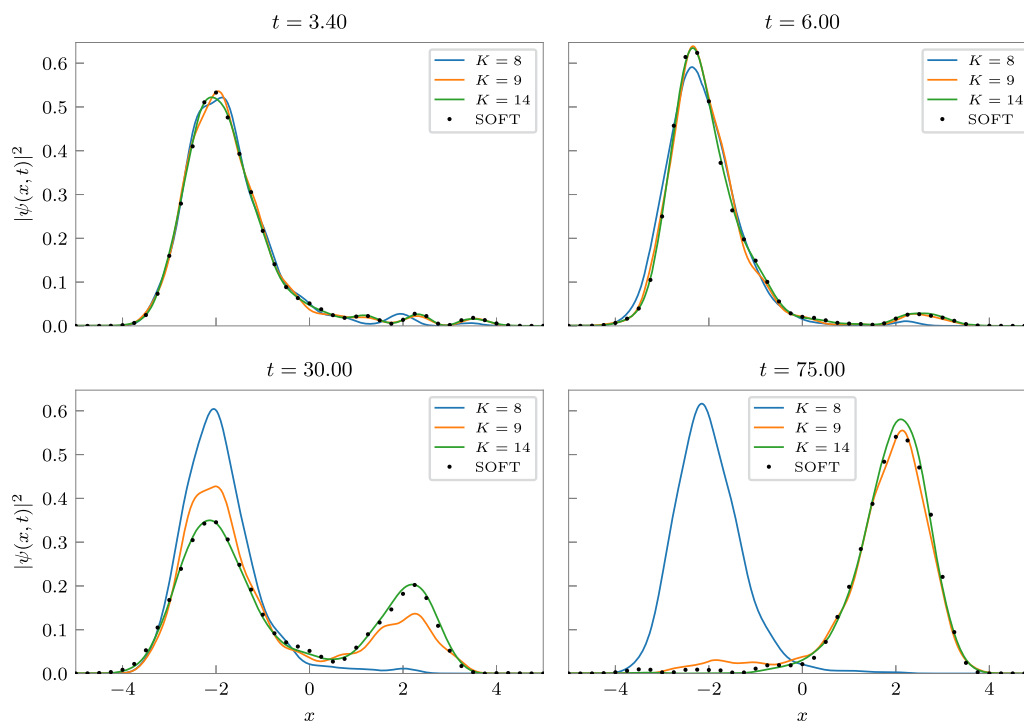


FIG. 12. Density snapshots, $|\psi(x, t)|^2$, for the 1D double-well system at representative times. The SOFT reference is shown using black markers, while VAGD-TG results are shown for $K = 8$, $K = 9$, and $K = 14$.

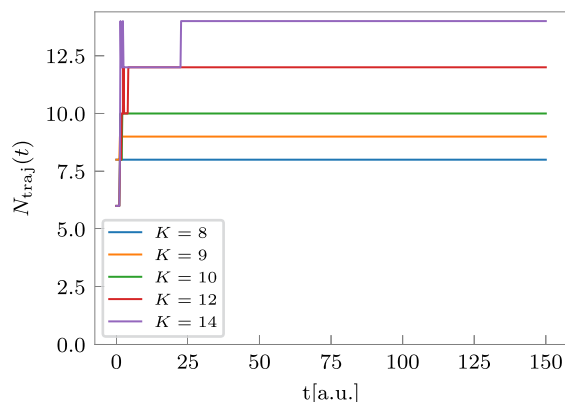


FIG. 13. Number of trajectories for the 1D double well system.

importance of the tail in describing the tunneling remains the same.

In Fig. 13, we show the actual number of trajectories used as a function of time for different values of K . We see a difference from the previous examples. Here, the number of trajectories almost immediately jumps to the maximum value owing to the significantly greater degree of anharmonicity. Even the adaptive algorithm is unable to effectively restrict $N_{\text{traj}}(t)$ to extremely low values. If we had not capped it, the number would have been even larger. However, the perfect overlaps tell us that the simulations are converged with respect to K . Using larger values of K would have resulted in the adaptive algorithm making $N_{\text{traj}}(t)$ significantly higher than what is required.

2. Tunneling in a two-dimensional double-well

Finally, we explore the 2D double-well studied by Kong *et al.*⁴² The model potential couples two strongly correlated degrees of freedom and is given by

$$V_{\eta}(x_1, x_2) = \frac{x_1^4}{16\eta} - \frac{x_1^2}{2} + \frac{x_2^2}{2} + \frac{x_1 x_2}{2}, \quad \eta > 0. \quad (13)$$

As in the one-dimensional case, the quartic term generates a double-well structure along x_1 , while the bilinear coupling term ($x_1 x_2/2$) induces a strong correlation between the coordinates. The initial state is chosen as a single normalized GWP localized in the left well along x_1 ,

$$\Psi(\vec{x}, 0) = \pi^{-1/2} \exp\left(-\frac{1}{2}\left((x_1 + 2\sqrt{\eta})^2 + x_2^2\right)\right). \quad (14)$$

Three VAGD-TG simulations were run with $\Delta t = 0.01$ a.u., and Fig. 14 compares their results with TGA. The numerically exact reference calculations were performed with the SOFT method on a uniform grid with $x_1, x_2 \in [-15, 15]$ (spacing $\Delta x = 0.1$ a.u.) and time step $\Delta t = 0.001$ a.u. (convergence is achieved even at $\Delta t = 0.01$ a.u.) For the present set of parameters, the $K = 60$ and $K = 90$ calculations were carried out with $W_{\min} = 10$ along both dimensions and $N_{\text{seg}} = 10$, while the improved $K = 300$ calculation used $W_{\min} = 15$ and $N_{\text{seg}} = 10$. In all the cases, 2000 decomposi-

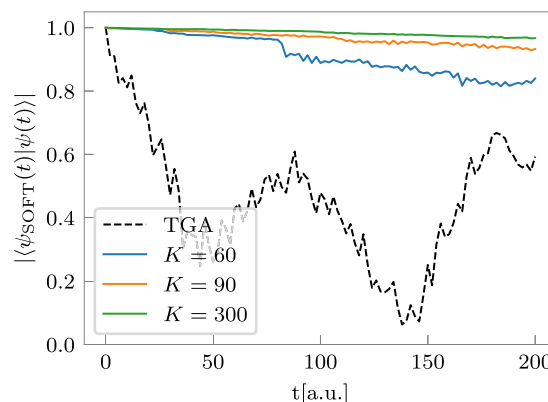


FIG. 14. Overlap of VAGD-TG and TGA method for the 2D double-well potential with SOFT.

tions steps were used. All calculations were run at $F_{\text{thresh}} = 0.99995$. Here, $W_{\min}$ of 15 enforces narrower, more localized Gaussians, so that the local harmonic approximation remains more reliable over each segment. Since this is a more difficult optimization problem, after reaching a minimal fidelity using the neural network, a first-order direct optimization of the GWPs was used to hit the target fidelity.

Figure 14 compares the overlap of the simulations using different number of GWPs with the SOFT result. As expected, TGA fails rapidly for this strongly correlated anharmonic tunneling problem. Its overlap with the exact SOFT result drops quickly at early times and then exhibits large oscillations, indicating that a single thawed Gaussian is unable to represent the evolving correlated structure of the wave function. In contrast, all VAGD-TG calculations retain substantially higher overlap throughout the propagation window. The $K = 60$ result already captures a large part of the wave-function structure, but shows a systematic loss of overlap at later times especially beyond $t \approx 75$ a.u. Increasing the GWP expansion to $K = 90$ gives a clear improvement and keeps the overlap high over the full propagation window, implying that the bulk of the physics has been completely accounted for. The $K = 300$ calculation is nearly converged on the scale of the plot, remaining closest to the SOFT reference throughout the dynamics.

To examine this more directly, Fig. 15 shows the population that has tunneled to the right half-plane,

$$P(x_1 > 0, t) = \int_0^\infty dx_1 \int_{-\infty}^\infty dx_2 |\Psi(x_1, x_2, t)|^2. \quad (15)$$

The $K = 60$ calculation captures the onset of tunneling and the population transfer to the right half-plane. The recurrence is also qualitatively captured. However, it shows visible deviations from SOFT in terms of the timescales involved. With $K = 60$, the entire tunneling dynamics seems to have slowed down.

The VAGD-TG results for $K = 90$ and $K = 300$ track the SOFT calculation rather well throughout the entire simulation. The rise of the population during the first tunneling event, its subsequent decrease as the wave packet returns, and the re-emergence of the population at later times are all captured correctly. Despite the drop in the overlap of the $K = 90$ run, it gets the correct physical

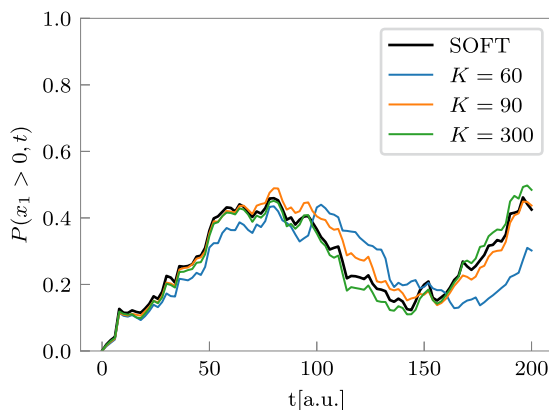


FIG. 15. Evolution of the tunneling population with time for 2D double-well.

picture for the tunneling dynamics. Pushing K to 300 brings quantitative agreement with the SOFT results for both the overlap and the tunneling population. (The evolution of the positions are shown in the [supplementary material](#), and they follow a similar trend to the tunneling population.) In comparison, the TSTG calculation of the same system required more than a million trajectories⁴² in comparison with the current VAGD-TG simulation using around a hundred GWPs. A final comment is due about the flexibility of the GWP representations. In these simulations, we have put the same $W_{\min}$ along both x_1 and x_2 . Given that the anharmonicities intrinsic to the two dimensions are different, it is expected that the x_1 direction would possibly require thinner wave functions for the local harmonic approximation to be valid, while x_2 may be able to accommodate wider wavepackets. Such ideas may be leveraged in the future for decreasing the number of GWPs and increasing the accuracy of the dynamics.

V. CONCLUSION

The time-slicing approach^{40–42} enables semiclassical simulations to achieve full quantum accuracy at the cost of propagating additional classical trajectories. The primary challenge, however, lies in decomposing the multidimensional time-evolved wave function into a basis of Gaussian wave packets. Monte Carlo-based approaches might appear to offer a natural solution to this problem, but, in practice, they are fundamentally limited by the dynamical sign problem.^{43–45} Even when Monte Carlo sampling can be converged, the number of trajectories required becomes prohibitively large, with the statistical error scaling as $N^{-1/2}$ for N samples in the absence of a sign problem. Through our Variational Adaptive Gaussian Decomposition (VAGD) scheme, we demonstrate that the essential physics encoded in the wave function can instead be captured using a relatively small number of carefully selected trajectories, providing a scalable and quadrature-free route to time-sliced thawed Gaussian dynamics.

VAGD provides a variationally optimal decomposition of a given wave function in terms of GWPs. Its main advantage is that the effective “basis” is optimized specifically for the wave function under consideration through an autoencoder–decoder architecture. This decomposition scheme can be utilized in a variety of

time-sliced architectures built on semiclassical or quantum schemes. Most prominently, one can think of combinations with time-sliced versions of Herman–Kluk semiclassical,⁴¹ thawed Gaussian approximation,⁴² or even with Rothe’s method of propagation.^{37–39} In the present work, we demonstrate the combination with time-sliced thawed Gaussian dynamics (VAGD-TG). The method has been tested on the Morse potential and on one- and two-dimensional tunneling problems. Dimensional scaling was further examined using independent Morse oscillators up to four dimensions, where the computational cost was found to grow only mildly with dimensionality. Our results suggest a polynomial scaling behavior, although the precise performance cannot yet be commented on from the numerical explorations. In fact, the scaling is dependent on the correlation between the dimensions and the overall anharmonicity of the system under study. Nevertheless, VAGD promises a substantial improvement over the exponential scaling typically encountered in conventional approaches. In addition, it provides a systematic route to converging to the fully quantum result, giving us an observable-specific handle on accuracy. One can decide upon the trade-off of increasing the number of trajectories to increase the accuracy for the specific problem at hand.

Particularly encouraging is the performance observed in tunneling problems. In the one-dimensional case, the tunneling dynamics becomes visible with as few as ten trajectories, while quantum accuracy is achieved with only 14 trajectories, representing an improvement of roughly two orders of magnitude compared with the $\sim 10^3$ trajectories required by TSTG. In two-dimensional tunneling, similarly accurate results are obtained with ~ 300 trajectories, whereas TSTG required millions. VAGD-TG is extremely effective in these tunneling scenarios, where the wave function acquires interference-driven non-trivial spatial structure, which cannot be captured by a single Gaussian. The adaptive multi-Gaussian representation that VAGD provides thus yields an optimally adequate representation of the complexity.

Beyond its practical efficiency, VAGD additionally offers an appealing conceptual perspective on semiclassical dynamics. The number of Gaussian wave packets required in the decomposition naturally reflects the degree of non-classical structure present in the evolving wave function. Systems exhibiting largely classical behavior can be represented with only a small number of trajectories, whereas strongly quantum phenomena, such as interference and tunneling, require richer Gaussian representations. In this sense, the adaptive Gaussian basis produced by VAGD provides a direct measure of the intrinsic quantum complexity of the dynamics. Future work will probe applications of VAGD to study the quantum dynamics of many-particle systems. In addition, developments will focus on simultaneously decreasing the decomposition time and the number of trajectories, even further making it suitable for use in combination with *ab initio* applications. Coupling with other dynamical methods such as VGD or single-Hessian TGA would also be evaluated. Going forward, VAGD promises to provide a platform for developing and refining methods aimed at scalable simulations of large dimensional systems that systematically converge to the full quantum results.

SUPPLEMENTARY MATERIAL

The accompanying [supplementary material](#) includes a detailed description of the implementation of the neural network and

training protocol, a demonstration of convergence of spectra vs the overlap for the 1D Morse oscillator, a discussion on the variance in 1D double well tunneling case, and graphs showing the placements of the GWPs in this case, dynamics of positions for the 2D double-well.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Rahul Sharma: Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Amartya Bose:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: FIDELITY CALCULATIONS

Optimizing the representation means that the fidelity has to be evaluated multiple times. The overlap calculation is a multidimensional integral and can prove to be a bottle-neck for multidimensional problems. However, in practice, fidelity can be evaluated entirely analytically when both $|\psi\rangle$ and $|\Psi\rangle$ are expressed as superpositions of GWPs. The overlap that is required for the fidelity is given as $\langle\psi|\Psi\rangle = \sum_{i=1}^{N_{\text{inp}}} \sum_{j=1}^{N_{\text{out}}} c_i^* C_j \langle\phi_i|\Phi_j\rangle$. All the overlaps of the pairs of GWPs have closed form expressions. For two d -dimensional GWPs of the form given in Eq. (1), with $|\phi_i\rangle$ parameterized by $(\vec{q}_i, \vec{p}_i, A_i, \gamma_i)$ and $|\Phi_j\rangle$ parameterized by $(\vec{Q}_j, \vec{P}_j, A_j, \Gamma_j)$, we define

$$B_{ij} = A_j - A_i^*, \quad (\text{A1})$$

$$\vec{b}_{ij} = \vec{P}_j - \vec{p}_i - A_j \vec{Q}_j + A_i^* \vec{q}_i. \quad (\text{A2})$$

Then, the elementary overlap is

$$\langle\phi_i|\Phi_j\rangle = \sqrt{\frac{(2\pi i\hbar)^d}{\det B_{ij}}} e^{\frac{i}{\hbar}(\Gamma_j - \gamma_i^*) - \frac{i}{\hbar}(\vec{P}_j^T \vec{Q}_j - \vec{p}_i^T \vec{q}_i)} \times \exp\left(-\frac{i}{2\hbar}\left[\vec{b}_{ij}^T B_{ij}^{-1} \vec{b}_{ij} - (\vec{Q}_j^T A_j \vec{Q}_j - \vec{q}_i^T A_i^* \vec{q}_i)\right]\right). \quad (\text{A3})$$

Thus, the loss defined in Eq. (6) can be efficiently evaluated using these overlap matrices.

To get the full overlap, we need $N_{\text{inp}} \times N_{\text{out}}$ elementary overlap computations, the cost of each of which also grows with the dimensionality of the problem. So, despite having the exact analytical

forms, the computation can get costly. Because each of the computation is independent, they can be trivially parallelized.

The fidelity, F , is defined as $F = \frac{|\langle\psi|\Psi\rangle|^2}{\langle\psi|\psi\rangle\langle\Psi|\Psi\rangle}$, where $|\psi\rangle$ is the candidate output obtained from the optimization procedure and $|\Psi\rangle$ is the target state, both represented as sums of Gaussian wavepackets. The norm of the target state can obviously be cached once per decomposition step. In addition, by judiciously precomputing and storing various portions of Eq. (A3) corresponding to the known and unchanging target state, one can significantly reduce the computational burden of the fidelity calculations.

Further optimizations can be brought in by filtering the overlaps by distance of the centers of the wavepackets. If the centers of two GWPs are far apart in comparison with their combined width, then the overlap will become negligible. One can consequently avoid the matrix operations for such pairs altogether, thereby improving the speed of the overlap computation.

APPENDIX B: NATURE OF GWPs GENERATED: 1D MORSE EXAMPLE

We demonstrate the time-evolution of an initial Gaussian wave packet under a Morse potential and its VAGD decompositions at

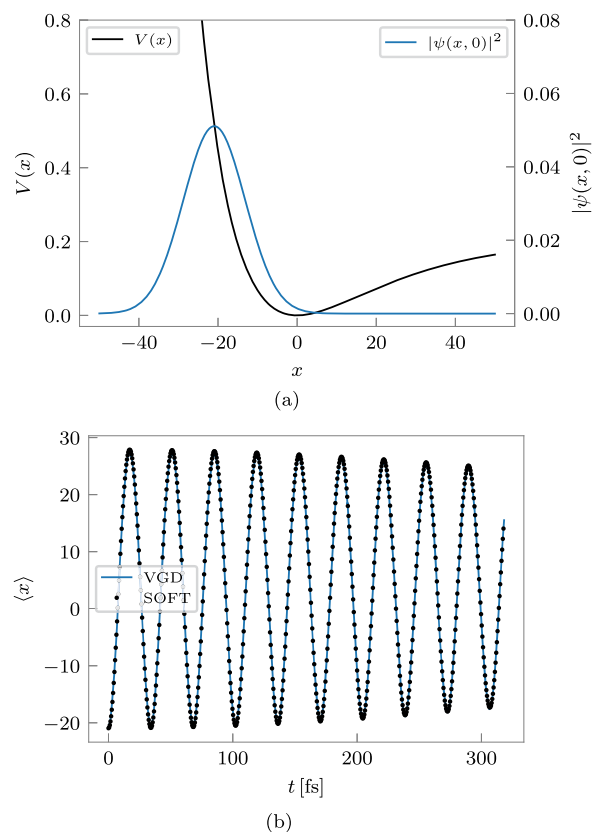


FIG. 16. 1D Morse oscillator ($\chi = 0.005$), initial condition, and dynamics. (a) 1D Morse oscillator and initial state as a Gaussian wave packet. (b) Expectation value of position with time.

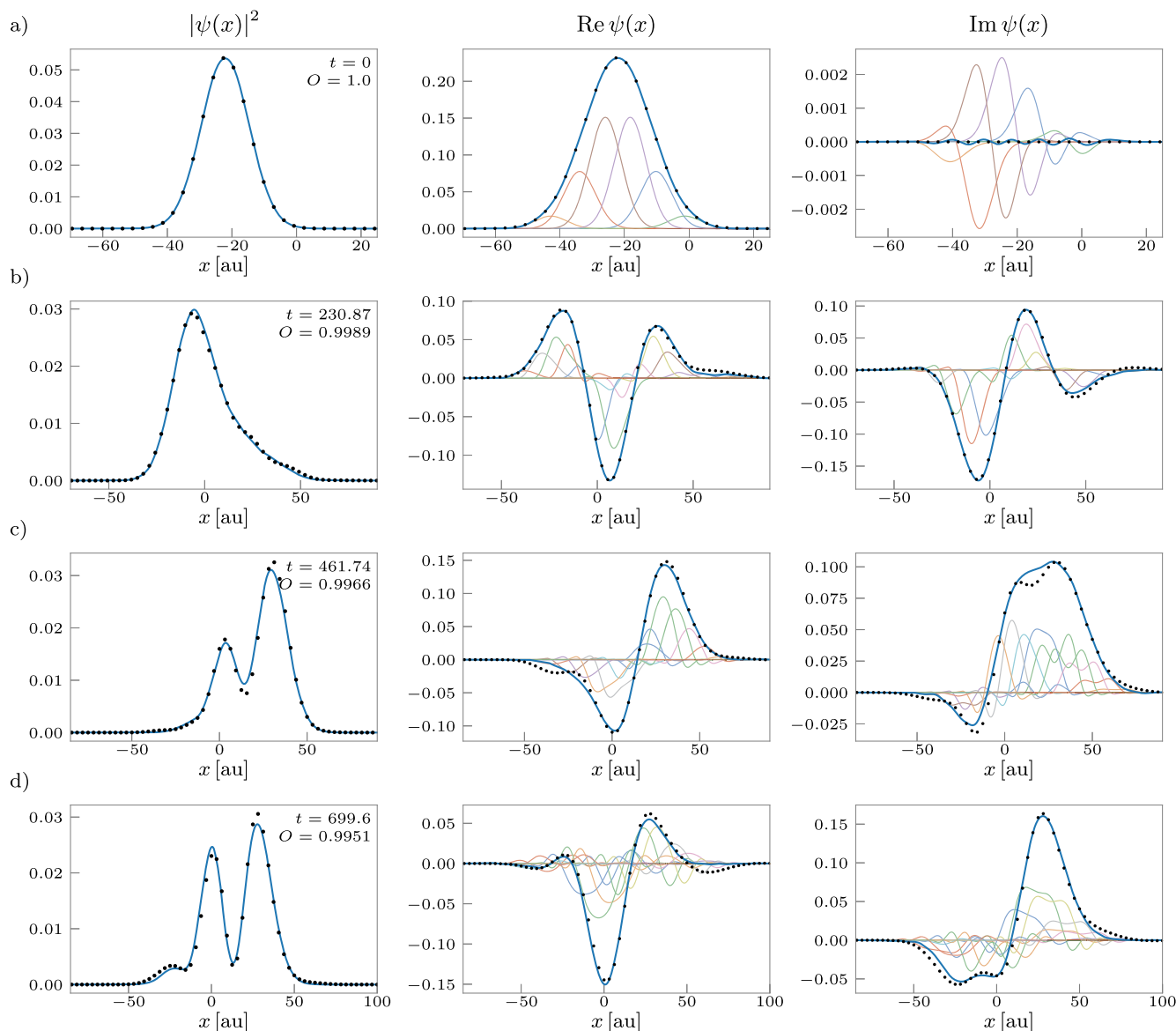


FIG. 17. VAGD decomposition at different times. The thin colored lines indicate individual GWPs. Markers: true SOFT wavefunctions; solid blue line: VAGD-TG wavefunction; O is the overlap of SOFT and VAGD results.

different time snapshots. As an example, we use the $\chi = 0.005$ case from Sec. IV A 1. Figure 16(a) shows the potential and the initial wave function. The evolution of position with time is also shown in Fig. 16(b). At different decomposition steps along the VAGD-TG trajectory, VAGD positions varying numbers of GWPs optimally. In Fig. 17, we show the decomposition of the wave function into GWPs at different times. Notice how even the initial time wave function, despite being a single GWP, is decomposed as a superposition of multiple GWPs owing to the width limit. We see six GWPs being involved in the decomposition at $t = 0$ [Fig. 17(a)]. At $t = 230.87$ in Fig. 17(b) we see a shoulder start to develop, and the number of GWPs required is increased to 14 to capture the greater

complexity. However, by the time we reach the bimodal wave function in Fig. 17(d), 16 GWPs are required for a proper description. Not only is the number of GWPs dependent on the input wave function but also the details of the individual GWPs change from case to case, adapting to the requirements.

APPENDIX C: COST OF VAGD-TG

We present a preliminary empirical exploration of the cost of VAGD training for our particular implementation. It should be noted that the following study is not an optimized benchmark by any measure. While several algorithmic improvements can still be

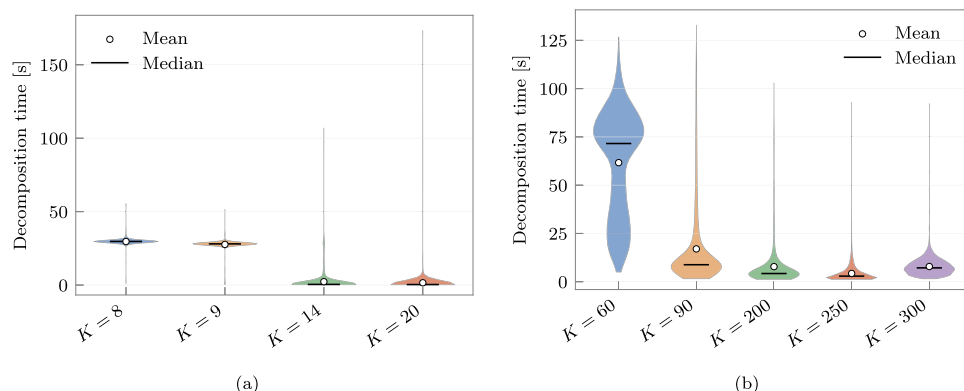


FIG. 18. Distribution of decomposition times along a VAGD-TG simulation for the 1D double-well calculation where a time-evolved K -GWP input state is decomposed into a K -GWP output state reinforcing all the conditions. (a) 1D double well. (b) 2D double well.

incorporated, this provides an intuitive picture into the performance for an initial implementation. We use the 1D and 2D double well systems to study the run-times. Given the extremely small dimensionality of the problem, the actual time spent on the classical trajectories is negligible in both cases and all the run-time comes from the decompositions. All the timing data reported in this section were obtained on a single NVIDIA V100 GPU.

We show the distribution of decomposition times along the VAGD-TG simulation of the one-dimensional double well for changing K in Fig. 18(a). There are some interesting features. First, it might come as a surprise that the $K = 8$ simulations spend the maximum time in the VAGD procedure on average. In hindsight, this makes sense for two reasons: (a) for this simple problem, neural network optimization is an overkill and the startup costs are really not worth it; and more importantly, (b) owing to the low representational power of the $K = 8$ *ansatz*, the code repeatedly fails to achieve the target fidelity threshold and keeps searching for the optimal representation for several further epochs. This latter issue can easily be suppressed by changing meta-parameters of the VAGD algorithm. Similar results are obtained for the $K = 9$ run. Although the *ansatz* is flexible enough to show the first tunneling event qualitatively, it is also not enough to hit the target fidelity for most of the decomposition steps. Next, notice that once the representation becomes flexible enough, at $K = 14$ and $K = 20$, the median time taken drops down to very low values. One notices that the variance, however, is very large. A few decomposition steps take exceptionally long times. The reason is likely to be the failure of the warm-start *ansatz*.

Several of these features carry over to the 2D case, as shown in Fig. 18(b). In this case, even the $K = 60$ run shows the high decomposition time characteristic of the representation not being sufficiently flexible. The cost drops down substantially at $K = 90$. Interestingly, one notices a not insignificant decrease in the mean and median decomposition times from $K = 90$ to 200 to 250. The increase in representational power substantially aids the optimization processes in this regime, although the number of calculations required for the fidelity calculations increase. From $K = 250$ to $K = 300$, we notice the increased cost of the fidelity, and optimization of larger set of parameters become dominant, leading to an increase in the total decomposition time.

While this exploration of the decomposition times are preliminary, and several further optimizations can be implemented to make the algorithm faster, a conclusion that stands out is that there is only

a very weak dependence of the run-time with the number of GWPs being used in the representation, once *ansatz* is flexible enough to give a proper description of the wave function. Going from one-dimension to two-dimensions also does not substantially increase the cost of a decomposition step. This holds great promise for systems with larger dimensions and which require numerous GWPs for adequate descriptions.

REFERENCES

- 1 M. D. Feit, J. A. Fleck, Jr., and A. Steiger, "Solution of the Schrödinger equation by a spectral method," *J. Comput. Phys.* **47**, 412–433 (1982).
- 2 D. Kosloff and R. Kosloff, "A Fourier method solution for the time dependent Schrödinger equation as a tool in molecular dynamics," *J. Comput. Phys.* **52**, 35–53 (1983).
- 3 H.-D. Meyer, U. Manthe, and L. S. Cederbaum, "The multi-configurational time-dependent Hartree approach," *Chem. Phys. Lett.* **165**, 73–78 (1990).
- 4 U. Manthe and T. Weike, "On the multi-layer multi-configurational time-dependent Hartree approach for bosons and fermions," *J. Chem. Phys.* **146**, 064117 (2017).
- 5 W. H. Miller, "Semiclassical methods in chemical physics," *Science* **233**, 171–177 (1986).
- 6 J. H. Van Vleck, "The correspondence principle in the statistical interpretation of quantum mechanics," *Proc. Natl. Acad. Sci. U. S. A.* **14**, 178–188 (1928).
- 7 M. C. Gutzwiller, "Phase-integral approximation in momentum space and the bound states of an atom," *J. Math. Phys.* **8**, 1979–2000 (1967).
- 8 M. F. Herman and E. Kluk, "A semiclassical justification for the use of non-spreading wavepackets in dynamics calculations," *Chem. Phys.* **91**, 27–34 (1984).
- 9 W. H. Miller, "An alternate derivation of the Herman—Kluk (coherent state) semiclassical initial value representation of the time evolution operator," *Mol. Phys.* **100**, 397–400 (2002).
- 10 K. G. Kay, "The Herman—Kluk approximation: Derivation and semiclassical corrections," *Chem. Phys.* **322**, 3–12 (2006), a Part of Special Issue: Real-time dynamics in complex quantum systems.
- 11 E. J. Heller, "Time-dependent approach to semiclassical dynamics," *J. Chem. Phys.* **62**, 1544–1555 (1975).
- 12 E. J. Heller, "Time dependent variational approach to semiclassical dynamics," *J. Chem. Phys.* **64**, 63–73 (1976).
- 13 E. J. Heller, "The semiclassical way to molecular spectroscopy," *Acc. Chem. Res.* **14**, 368–375 (1981).
- 14 E. J. Heller, "Cellular dynamics: A new semiclassical approach to time-dependent quantum mechanics," *J. Chem. Phys.* **94**, 2723–2729 (1991).
- 15 J. J. L. Vaníček, "Family of Gaussian wavepacket dynamics methods from the perspective of a nonlinear Schrödinger equation," *J. Chem. Phys.* **159**, 014114 (2023).

- ¹⁶T. Begušić, J. Roulet, and J. Vaniček, "On-the-fly *ab initio* semiclassical evaluation of time-resolved electronic spectra," *J. Chem. Phys.* **149**, 244115 (2018).
- ¹⁷T. Begušić, A. Patoz, M. Šulc, and J. Vaniček, "On-the-fly *ab initio* three thawed Gaussians approximation: A semiclassical approach to Herzberg-Teller spectra," *Chem. Phys.* **515**, 152–163 (2018).
- ¹⁸T. Begušić, M. Cordova, and J. Vaniček, "Single-Hessian thawed Gaussian approximation," *J. Chem. Phys.* **150**, 154117 (2019).
- ¹⁹T. Begušić, E. Tapavicza, and J. Vaniček, "Applicability of the thawed Gaussian wavepacket dynamics to the calculation of vibronic spectra of molecules with double-well potential energy surfaces," *J. Chem. Theory Comput.* **18**, 3065–3074 (2022).
- ²⁰T. Begušić and J. Vaniček, "Finite-temperature, anharmonicity, and Duschinsky effects on the two-dimensional electronic spectra from *ab initio* thermo-field Gaussian wavepacket dynamics," *J. Phys. Chem. Lett.* **12**, 2997–3005 (2021).
- ²¹I. Burghardt, H.-D. Meyer, and L. S. Cederbaum, "Approaches to the approximate treatment of complex molecular systems by the multiconfiguration time-dependent Hartree method," *J. Chem. Phys.* **111**, 2927–2939 (1999).
- ²²F. D. Maiolo, G. A. Worth, and I. Burghardt, "Multi-layer Gaussian-based multi-configuration time-dependent Hartree (ML-GMCTDH) simulations of ultrafast charge separation in a donor-acceptor complex," *J. Chem. Phys.* **154**, 144106 (2021).
- ²³I. Burghardt, M. Nest, and G. A. Worth, "Multiconfigurational system-bath dynamics using Gaussian wave packets: Energy relaxation and decoherence induced by a finite-dimensional bath," *J. Chem. Phys.* **119**, 5364–5378 (2003).
- ²⁴I. Burghardt, K. Giri, and G. A. Worth, "Multimode quantum dynamics using Gaussian wavepackets: The Gaussian-based multiconfiguration time-dependent Hartree (G-MCTDH) method applied to the absorption spectrum of pyrazine," *J. Chem. Phys.* **129**, 174104 (2008).
- ²⁵S. Römer, M. Ruckebauer, and I. Burghardt, "Gaussian-based multiconfiguration time-dependent Hartree: A two-layer approach. I. Theory," *J. Chem. Phys.* **138**, 064106 (2013).
- ²⁶M. Ben-Nun, J. Quenneville, and T. J. Martínez, "Ab initio multiple spawning: Photochemistry from first principles quantum molecular dynamics," *J. Phys. Chem. A* **104**, 5161–5175 (2000).
- ²⁷M. Ben-Nun and T. J. Martínez, "A multiple spawning approach to tunneling dynamics," *J. Chem. Phys.* **112**, 6113–6121 (2000).
- ²⁸M. A. C. Saller and S. Habershon, "Basis set generation for quantum dynamics simulations using simple trajectory-based methods," *J. Chem. Theory Comput.* **11**, 8–16 (2015).
- ²⁹Y. Wu and V. S. Batista, "Matching-pursuit for simulations of quantum processes," *J. Chem. Phys.* **118**, 6720–6724 (2003).
- ³⁰Y. Wu and V. S. Batista, "Quantum tunneling dynamics in multidimensional systems: A matching-pursuit description," *J. Chem. Phys.* **121**, 1676–1680 (2004).
- ³¹Y. Wu, M. F. Herman, and V. S. Batista, "Matching-pursuit/split-operator Fourier-transform simulations of nonadiabatic quantum dynamics," *J. Chem. Phys.* **122**, 114114 (2005).
- ³²D. V. Shalashilin and M. S. Child, "Multidimensional quantum propagation with the help of coupled coherent states," *J. Chem. Phys.* **115**, 5367–5375 (2001).
- ³³D. V. Shalashilin and M. S. Child, "The phase space CCS approach to quantum and semiclassical molecular dynamics for high-dimensional systems," *Chem. Phys.* **304**, 103–120 (2004).
- ³⁴B. Gu and S. Garashchuk, "Quantum dynamics with Gaussian bases defined by the quantum trajectories," *J. Phys. Chem. A* **120**, 3023–3031 (2016).
- ³⁵M. Dutra, S. Wickramasinghe, and S. Garashchuk, "Quantum dynamics with the quantum trajectory-guided adaptable Gaussian bases," *J. Chem. Theory Comput.* **16**, 18–34 (2020).
- ³⁶M. Dutra, S. Wickramasinghe, and S. Garashchuk, "Multidimensional tunneling dynamics employing quantum-trajectory guided adaptable Gaussian bases," *J. Phys. Chem. A* **124**, 9314–9325 (2020).
- ³⁷E. Rothe, "Zweidimensionale parabolische Randwertaufgaben als Grenzfall eindimensionaler Randwertaufgaben," *Math. Ann.* **102**, 650–670 (1930).
- ³⁸S. E. Schrader, H. E. Kristiansen, T. B. Pedersen, and S. Kvaal, "Time evolution as an optimization problem: The hydrogen atom in strong laser fields in a basis of time-dependent Gaussian wave packets," *J. Chem. Phys.* **161**, 044105 (2024).
- ³⁹S. E. Schrader, T. B. Pedersen, and S. Kvaal, "Multidimensional quantum dynamics with explicitly correlated Gaussian wave packets using Rothe's method," *J. Chem. Phys.* **162**, 024109 (2025).
- ⁴⁰E. J. Heller, "Wavepacket path integral formulation of semiclassical dynamics," *Chem. Phys. Lett.* **34**, 321–325 (1975).
- ⁴¹J. C. Burant and V. S. Batista, "Real time path integrals using the Herman-Kluk propagator," *J. Chem. Phys.* **116**, 2748–2756 (2002).
- ⁴²X. Kong, A. Markmann, and V. S. Batista, "Time-sliced thawed Gaussian propagation method for simulations of quantum dynamics," *J. Phys. Chem. A* **120**, 3260–3269 (2016).
- ⁴³N. Makri and W. H. Miller, "Monte Carlo path integration for the real time propagator," *J. Chem. Phys.* **89**, 2170–2177 (1988).
- ⁴⁴N. Makri and W. H. Miller, "Monte Carlo integration with oscillatory integrands: Implications for Feynman path integration in real time," *Chem. Phys. Lett.* **139**, 10–14 (1987).
- ⁴⁵C. H. Mak and D. Chandler, "Solving the sign problem in quantum Monte Carlo dynamics," *Phys. Rev. A* **41**, 5709–5712 (1990).
- ⁴⁶W. Koch and T. J. Frankcombe, "Basis expansion leaping: A new method to solve the time-dependent Schrödinger equation for molecular quantum dynamics," *Phys. Rev. Lett.* **110**, 263202 (2013).
- ⁴⁷T. Murakami and T. J. Frankcombe, "Accurate quantum molecular dynamics for multidimensional systems by the basis expansion leaping multi-configuration Gaussian (BEL MCG) method," *J. Chem. Phys.* **149**, 134113 (2018).
- ⁴⁸T. Murakami and T. J. Frankcombe, "Non-adiabatic quantum molecular dynamics by the basis expansion leaping multi-configuration Gaussian (BEL MCG) method: Multi-set and single-set formalisms," *J. Chem. Phys.* **150**, 144112 (2019).
- ⁴⁹E. Faou and C. Lubich, "A Poisson integrator for Gaussian wavepacket dynamics," *Comput. Visualization Sci.* **9**, 45–55 (2006).
- ⁵⁰R. Moghaddasi Fereidani and J. J. L. Vaniček, "High-order geometric integrators for the variational Gaussian approximation," *J. Chem. Phys.* **159**, 094114 (2023).
- ⁵¹R. Moghaddasi Fereidani and J. J. L. Vaniček, "High-order geometric integrators for the local cubic variational Gaussian wavepacket dynamics," *J. Chem. Phys.* **160**, 044113 (2024).
- ⁵²L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.* **60**, 223–311 (2018).
- ⁵³Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).
- ⁵⁴I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016); <http://www.deeplearningbook.org>.
- ⁵⁵G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**, 504–507 (2006).
- ⁵⁶G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (Johns Hopkins University Press, Baltimore, MD, 2013).