

Petrov-Galerkin model reduction for thermochemical nonequilibrium gas mixtures

Ivan Zanardi ¹, Alberto Padovan, Daniel J. Bodony, Marco Panesi *

Center for Hypersonics and Entry Systems Studies, Department of Aerospace Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:

Reduced-order modeling
Thermochemistry
Nonequilibrium
Hypersonics

ABSTRACT

State-specific thermochemical collisional models are crucial to accurately describe the physics of systems involving nonequilibrium plasmas, but they are also computationally expensive and impractical for large-scale, multi-dimensional simulations. Historically, computational cost has been mitigated by using empirical and physics-based arguments to reduce the complexity of the governing equations. However, the resulting models are often inaccurate and they fail to capture the important features of the original physics. Additionally, the construction of these models is often impractical, as it requires extensive user supervision and time-consuming parameter tuning. In this paper, we address these issues through an easily implementable and computationally efficient model reduction pipeline based on the Petrov-Galerkin projection of the nonlinear kinetic equations onto a low-dimensional subspace. Our approach is justified by the observation that kinetic systems in thermal nonequilibrium tend to exhibit low-rank dynamics that rapidly drive the state towards a low-dimensional subspace that can be exploited for reduced-order modeling. Furthermore, despite the nonlinear nature of the governing equations, we observe that the dynamics of these systems evolve on subspaces that can be accurately identified using the linearized equations about thermochemical equilibrium steady states, and we shall see that this allows us to significantly reduce the cost associated with the construction of the model. The approach is demonstrated on two distinct thermochemical systems: a rovibrational collisional model for the O_2 -O system, and a vibrational collisional model for the combined O_2 -O and O_2 - O_2 systems. Our method achieves high accuracy, with relative errors of less than 1% for macroscopic quantities (i.e., moments) and 10% for microscopic quantities (i.e., energy levels population), while also delivering excellent compression rates and speedups, outperforming existing state-of-the-art techniques.

1. Introduction

Nonequilibrium flows are encountered in many engineering and science disciplines, such as hypersonic atmospheric entry [1–4] and material processing and manufacturing with low-temperature plasmas [5,6]. In these applications, the thermodynamic state of the flow frequently exhibits significant deviations of the internal energy distribution of the gaseous mixture from its Maxwell-Boltzmann equilibrium value. Such deviations can influence key macroscopic properties, including diffusion, viscous stress, heat conduction,

* Corresponding author.

E-mail addresses: zanardi3@illinois.edu (I. Zanardi), padovan3@illinois.edu (A. Padovan), bodony@illinois.edu (D.J. Bodony), mpanesi@illinois.edu (M. Panesi).

<https://doi.org/10.1016/j.jcp.2025.113999>

Received 22 November 2024; Received in revised form 5 April 2025; Accepted 7 April 2025

Available online 10 April 2025

0021-9991/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

internal energy transfer, chemical reactions, and thermal radiation. To describe and understand these complex behaviors and their impact, increasingly advanced and extensive mathematical models have been developed [7–11], encompassing multiple physical phenomena across a wide range of spatio-temporal scales.

The exact description of thermochemical and radiative processes can be achieved at the microscopic level by considering state-to-state (StS) collisions among atoms, molecules, and electrons, as well as the emission and absorption of photons. This approach is mathematically formalized through the StS master equations [9–15]. By leveraging quantum chemistry models based on *ab initio* theories [16–21], these equations provide an unprecedented level of accuracy in describing the nonequilibrium state of gases by capturing the microscopic interactions between colliding particles in diverse chemical systems [10,11,22,23]. While these models provide exceptional accuracy, they are impractical in large-scale multi-dimensional simulations due to the significant increase in the number of degrees of freedom (i.e., the molecules' and atoms' energy levels). As a result, detailed StS calculations are typically employed to develop more efficient reduced-order models to be used in computational fluid dynamics (CFD) simulations.

Reduced-order models (ROMs) aim to capture the essential features of the system dynamics while significantly reducing the computational cost associated with the high-fidelity full-order model (FOM). The earliest ROM for thermochemical nonequilibrium, introduced by Landau and Teller [24], was based on several physical assumptions applicable only to simple harmonic oscillator molecules, resulting in a linear relaxation model. Building upon this formulation, more sophisticated models have been developed over the years to account for the coupling between vibration and dissociation [25–28]. However, many of these models, known as multi-temperature (MT) models [2,29–32], rely on crude, semi-empirical assumptions. They are typically constructed based on a rigid separation of internal energy modes (translational, rotational, vibrational, and electronic), without any rigorous derivation from fundamental kinetic equations nor consideration for physical principles and constraints. Given their interpolative nature, these conventional methods are fundamentally inadequate for capturing the behavior of gas particles under strong nonequilibrium conditions, where the distribution of internal energy levels exhibits complex features [10,11,33]. By contrast, state-of-the-art coarse-grained (CG) ROMs [34–38] provide a more effective solution compared to the traditional MT models. These coarse-grained approaches work by grouping individual energy levels into a smaller number of macroscopic bins. State populations are reconstructed using bin-wise distribution functions that maximize entropy and adhere to constraints based on bin properties (e.g., population, energy, or higher moments). The governing equations for the reduced-order system are derived by summing moments of the StS master equations. Although these models have demonstrated sufficient accuracy, they have a few limitations. First, while the principle of maximum entropy used to locally reconstruct microscopic scales within each bin perfectly adheres to physical principles, it may not fully capture the complex features of the distribution function. Second, the accuracy of the coarse-grained model is highly dependent on the effectiveness of the clustering. Significant effort, in the form of time-consuming optimization or physics-based analysis of the chemical system, is often required to achieve optimal clustering [39–45].

This work aims to advance the state of the art of model reduction for thermochemical kinetics by introducing a computationally-efficient pipeline that offers several key improvements over the existing reduced-complexity modeling formulations described above. First, the proposed method bypasses the need for empirical or physical assumptions, enhancing both accuracy and flexibility across a wide range of nonequilibrium conditions. Second, it avoids the need for time-consuming system analysis or optimization to identify the optimal reduced-order representation of the original dynamics. The fundamental component of the model reduction pipeline discussed herein is the construction of a Petrov-Galerkin model obtained by projecting the full-order dynamics (i.e., the StS master equations) onto a low-dimensional linear subspace of the state-space. Obtaining an accurate Petrov-Galerkin model requires the careful design of a linear projection operator of the form $\mathbb{P} = \Phi(\Psi^T\Phi)^{-1}\Psi^T$, where Φ and Ψ are tall rectangular matrices whose spans uniquely define \mathbb{P} . While the most simplistic approach to obtain a projection-based model is by *orthogonally* projecting onto the span of Proper Orthogonal Decomposition (POD) modes of high-fidelity data (see, e.g., [46,47] in the context of compressible flows and [48–50] in the context of thermochemical kinetics), it is well-known that physical systems that exhibit transient growth and high sensitivity to low-energy disturbances require modeling approaches based on *oblique* projections [51–54]. If the full-order system is linear, or if the nonlinear dynamics evolve in the proximity of an equilibrium, appropriate oblique projection operators can be constructed using balanced truncation and its variants [55–59], where a Petrov-Galerkin ROM is obtained by balancing the controllability and observability Gramians associated with the underlying linear system. For nonlinear systems evolving far away from equilibria, Otto et al. [53] recently introduced “Covariance Balancing Reduction using Adjoint Snapshots” (CoBRAS), which is a formulation that determines (quasi) optimal oblique projection operators by balancing the state and gradient covariance matrices associated with the full-order solution map. Given the non-normal nature of the StS master equations, and their high sensitivity to low-energy disturbances, it is essential to seek appropriate oblique projection operators to obtain accurate Petrov-Galerkin models and we therefore choose to adopt the aforementioned CoBRAS formulation. More specifically, we tailor CoBRAS to the StS master equations, and we assemble the state and gradient covariance matrices using linear trajectories governed by the linearization of the StS equations about thermochemical equilibria. Using the linearized equations allows us to generate the required forward and adjoint trajectories using the exact solution of linear ordinary differential equations, as opposed to numerically integrating a stiff set of nonlinear equations, and this results in computational savings of several orders of magnitude. Additionally, despite the fact that these trajectories are a first-order approximation of the nonlinear solution of the StS equations, we shall see that the loss of accuracy is minimal, and we can still capture the relevant subspaces that are necessary to construct an appropriate oblique projector \mathbb{P} for Petrov-Galerkin modeling. The performance of the proposed models is assessed on two thermochemical systems involving a pure oxygen mixture: a rovibrational collisional (RVC) model and a vibrational collisional (VC) model. On both systems, our formulation significantly outperforms state-of-the-art model reduction methods for chemical kinetics and is capable of providing accurate estimates of the moments of the distribution function and of the evolution of the internal energy level population.

The paper is organized as follows. Section 2 introduces the microscopic master equations that describe gas mixtures undergoing internal excitation, dissociation, or ionization within a 0D isothermal chemical reactor. Section 3 provides a detailed description of the proposed model reduction approach. Section 4 presents a brief overview of existing model reduction techniques for nonequilibrium chemical kinetics, and section 5 outlines the numerical results of the ROMs applied to the RVC and VC thermochemical models. Finally, section 6 offers concluding remarks and explores potential avenues for future research.

2. Physical modeling

Our objective is to investigate the behavior of a general mixture under sudden heating in an ideal chemical reactor. For this purpose, we make the following assumptions:

- i. The 0D reactor is plunged into a thermal bath maintained at constant temperature T .
- ii. The translational energy mode of the atoms and molecules is assumed to follow a Maxwell-Boltzmann distribution at the temperature T of the thermal bath.
- iii. At the beginning of the numerical experiment, the population of the rovibrational energy levels is assumed to follow a Boltzmann distribution at the initial equilibrium temperature T_0 .
- iv. The volume of the chemical reactor is kept constant during the experiment, and the thermodynamic system is closed (no mass exchange with the surrounding environment).

We consider both state-to-state (StS) collisional and radiative processes involving atoms or molecules, denoted by the symbol A , which may undergo internal excitation, de-excitation, recombination, ionization, and/or dissociation through interactions with other particles, B , which could include electrons, or through photon absorption and emission, also represented by B . The internal excitation and de-excitation processes can encompass electronic, rotational, and vibrational transitions. Any ionization or dissociation of A results in the formation of particles C and D . It should be noted that A , B , C , and D are generic placeholders and may represent the same type of particles. Superscripts i , j , k , l , p , and q are used to indicate internal states of the species involved. A generic form of collisional and radiative processes can be expressed as

- i. internal excitation and de-excitation:



- ii. ionization, dissociation, and recombination:



We also introduce n_s^i , g_s^i , and ϵ_s^i , representing the population (or number density), degeneracy, and energy level of state $i \in \mathcal{I}_s$ for species $s \in \mathcal{S}$, respectively. Here, \mathcal{I}_s represents the set of energy levels for species s , and \mathcal{S} denotes the set of chemical species in the mixture. All energy levels ϵ_s^i are referenced to a common zero-point energy. Therefore, for each species in the gas mixture, the dissociation or ionization energy is accounted for within ϵ_s^i . Each energy level is treated as an individual species, and their production rates can be computed using the zeroth-order reaction rate theory, leading to the formulation of the master equations. For instance, the governing equation for the population density of state i for species A , accounting for the processes described by (1) and (2), can be expressed as

$$\frac{dn_A^i}{dt} = \sum_{j,k,l} \left(-k_{ij kl}^e n_A^i n_B^j + k_{kl ij}^e n_A^k n_B^l \right) + \sum_{j,l,p,q} \left(-k_{ij pq}^d n_A^i n_B^j + k_{pq li}^r n_C^p n_D^q n_B^l \right). \quad (3)$$

Here, $k_{ij kl}^e$ and $k_{kl ij}^e$ represent the Maxwellian-distribution-based state-to-state excitation and de-excitation rate coefficients for process (1), while $k_{ij pq}^d$ and $k_{pq li}^r$ represent the ionization/dissociation and recombination rate coefficients for process (2). These microscopic state-to-state rate coefficients obey the principles of detailed balance (or microreversibility) and depend on the translational temperature T . The forward (exothermic) rate coefficients were calculated using the quasi-classical trajectory (QCT) method [60]. Equivalent master equations can be written for the internal states of species B , C , and D using a similar form to equation (3). Further details on notation and specific equations will be provided in section 5 for each thermochemical system investigated.

2.1. State-space representation

For the scope of this work, it is convenient to express the governing equations (3) in compact state-space form as follows

$$\frac{d}{dt} \mathbf{n}(t) = \mathbf{h}(\mathbf{n}(t); T), \quad \mathbf{n}(0) = \mathbf{n}_0(\zeta), \quad (4)$$

where the state vector $\mathbf{n}(t) \in \mathbb{R}^N$ represents the number densities associated with each state $i \in \mathcal{I}_s$ for all species $s \in \mathcal{S}$, so that $N = \sum_{s \in \mathcal{S}} |\mathcal{I}_s|$ and $|\mathcal{I}_s|$ denotes the cardinality of the set \mathcal{I}_s . The initial condition $\mathbf{n}(0) \in \mathbb{R}^N$ is defined by the function $\mathbf{n}_0 : \mathbb{R}^p \rightarrow \mathbb{R}^N$, which takes a set of parameters $\zeta \in \mathbb{R}^p$ as input. These parameters include the density ρ , initial equilibrium temperature T_0 , and the initial species mass fractions, w_{s_0} , for each $s \in \mathcal{S}$. For isothermal chemical reactors where $dT/dt = d\rho/dt = 0$, the dynamics in (4) admit thermochemical equilibrium solutions $\bar{\mathbf{n}}(\rho, T)$ that are functions of density and temperature. To emphasize the importance of these two parameters and to adopt a more standard dynamical-systems notation, we pass the parametric density dependence of the solution $\mathbf{n}(t)$ from the initial condition to the dynamics. Thus, using the linear change of coordinates

$$\mathbf{n}(t) = \rho \operatorname{diag}(\mathbf{m})^{-1} \mathbf{w}(t), \quad (5)$$

where $\mathbf{w}(t) \in \mathbb{R}^N$ is the vector of mass fractions and $\mathbf{m} \in \mathbb{R}^N$ is the (constant) vector of masses, equation (4) becomes

$$\frac{d}{dt} \mathbf{w}(t) = \mathbf{f}(\mathbf{w}(t); \theta), \quad \mathbf{w}(0) = \mathbf{w}_0(\boldsymbol{\mu}), \quad (6)$$

where $\theta = (T, \rho)$ and $\boldsymbol{\mu}$ contains the initial equilibrium temperature T_0 and the initial species mass fractions. It is also useful to define a vector of outputs \mathbf{y}

$$\mathbf{y}(t) = \mathbf{C}\mathbf{w}(t), \quad (7)$$

which, in general, contains physical quantities that we are interested in. In the present work, we let $\mathbf{y}(t) \in \mathbb{R}^{(m+1)|\mathcal{S}|}$ represent the moments of orders 0 through m of the distribution function for each species in the mixture. More specifically, the j -th moment associated with species s is defined through the output matrix \mathbf{C} as follows

$$\mathbf{y}_{s,j} = (\mathbf{C}\mathbf{w})_{s,j} = \frac{1}{j!} \sum_{i \in \mathcal{I}_s} \frac{1}{M_s} (\epsilon_s^i)^j w_s^i, \quad (8)$$

where M_s is the species molar mass and the species energy levels ϵ_s^i have units of eV.

3. Projection-based model reduction

The numerical solution of the nonlinear governing equation (6) is often computationally prohibitive due to the large number of degrees of freedom (i.e., the molecules' and atoms' energy levels), and to the presence of extremely fast time scales that impose severe numerical restrictions on temporal integration schemes. These issues can be mitigated by computing reduced-order models that can be integrated in time at a fraction of the computational cost of the full-order model (6). In this work, these models are computed using a Petrov-Galerkin procedure, where the dynamics are (obliquely) projected onto a low-dimensional subspace.

3.1. Petrov-Galerkin model reduction

In (linear) Petrov-Galerkin model reduction, the state $\mathbf{w}(t) \in \mathbb{R}^N$ is constrained to a r -dimensional subspace of \mathbb{R}^N defined by the range of the rank- r linear projection operator \mathbb{P} . That is, $\hat{\mathbf{w}}(t) = \mathbb{P}\mathbf{w}(t) \in \mathbb{R}^N$, where $\hat{\mathbf{w}}(t)$ is the approximation of the state $\mathbf{w}(t)$ on the span of \mathbb{P} . Injecting this ansatz into (6), it is easy to see that the dynamics of $\hat{\mathbf{w}}$ are governed by

$$\begin{aligned} \frac{d}{dt} \hat{\mathbf{w}}(t) &= \mathbb{P}\mathbf{f}(\mathbb{P}\hat{\mathbf{w}}(t); \theta), \quad \hat{\mathbf{w}}(0) = \mathbb{P}\mathbf{w}_0(\boldsymbol{\mu}), \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\mathbb{P}\hat{\mathbf{w}}(t). \end{aligned} \quad (9)$$

Equation (9) is known as a Petrov-Galerkin model of (6). While the state $\hat{\mathbf{w}}(t)$ is N -dimensional (i.e., the same dimension as the original state $\mathbf{w}(t)$), an equivalent reduced-order representation of the dynamics in (9) can be written in terms of an r -dimensional vector of coefficients $\hat{\mathbf{z}}(t)$. In particular, since any rank- r linear projector \mathbb{P} can be written as $\mathbb{P} = \boldsymbol{\Phi}(\boldsymbol{\Psi}^\top \boldsymbol{\Phi})^{-1} \boldsymbol{\Psi}^\top$, where $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are rank- r matrices of size $N \times r$, defining $\hat{\mathbf{z}}(t) = \boldsymbol{\Psi}^\top \hat{\mathbf{w}}(t)$ and left-multiplying the first equation in (9) by $\boldsymbol{\Psi}^\top$ we obtain

$$\begin{aligned} \frac{d}{dt} \hat{\mathbf{z}}(t) &= \boldsymbol{\Psi}^\top \mathbf{f}(\boldsymbol{\Phi}(\boldsymbol{\Psi}^\top \boldsymbol{\Phi})^{-1} \hat{\mathbf{z}}(t); \theta), \quad \hat{\mathbf{z}}(0) = \boldsymbol{\Psi}^\top \mathbf{w}_0(\boldsymbol{\mu}), \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\boldsymbol{\Phi}(\boldsymbol{\Psi}^\top \boldsymbol{\Phi})^{-1} \hat{\mathbf{z}}(t), \end{aligned} \quad (10)$$

which is the Petrov-Galerkin reduced-order model that we sought. The accuracy of this model depends exclusively on the choice of trial and test subspaces spanned by $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, respectively, so it is essential to choose these subspaces appropriately. In this work, we define these subspaces using the recently introduced CoBRAS formulation [53], which we describe in the next subsection.

3.2. Covariance Balancing Reduction using Adjoint Snapshots (CoBRAS)

Let us consider the following map

$$F : \mathbb{R}^N \rightarrow \mathbb{R}^q : \mathbf{w}(t_0; \theta, \boldsymbol{\mu}) \mapsto (\mathbf{y}(t_0; \theta, \boldsymbol{\mu}), \mathbf{y}(t_1; \theta, \boldsymbol{\mu}), \dots, \mathbf{y}(t_{L-1}; \theta, \boldsymbol{\mu})), \quad (11)$$

which sends a state $\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu})$ along a trajectory of (6) to a sequence of L measurements $\mathbf{y}(t_k; \boldsymbol{\theta}, \boldsymbol{\mu})$ collected at times $t_k \geq t_0$. (Notice that t_0 is not necessarily equal to 0.) A similar map may be defined for the Petrov-Galerkin dynamics (9)

$$\hat{F} : \mathbb{R}^N \rightarrow \mathbb{R}^q : \mathbb{P}\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu}) \mapsto (\hat{\mathbf{y}}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu}), \hat{\mathbf{y}}(t_1; \boldsymbol{\theta}, \boldsymbol{\mu}), \dots, \hat{\mathbf{y}}(t_{L-1}; \boldsymbol{\theta}, \boldsymbol{\mu})) . \quad (12)$$

Within CoBRAS, we measure the accuracy of the Petrov-Galerkin model in the mean square sense, and we therefore seek the projection operator \mathbb{P} that minimizes the mean square error $\mathbb{E} [F(\mathbf{w}) - \hat{F}(\mathbb{P}\mathbf{w})]$. Direct optimization of this cost function using, e.g., gradient descent, would be computationally expensive, so we proceed by leveraging a useful result in [61]. In particular, under the assumption that the states $\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu})$ are normally distributed with covariance W_s , Proposition 2.5 in [61] shows that the mean square approximation error is upper-bounded by

$$\text{trace} (W_g (I - \mathbb{P}) W_s (I - \mathbb{P})^\top) , \quad (13)$$

where $W_g = \mathbb{E} [\nabla F \nabla F^\top]$ is the gradient covariance and $\nabla F = D_{\mathbf{w}} F(\mathbf{w})^\top \in \mathbb{R}^{N \times L \dim(\mathbf{y})}$, with $D_{\mathbf{w}}$ being the differential (or Jacobian) operator with respect to \mathbf{w} . Minimizing equation (13) is a much simpler task, and Theorem 2.3 in [53] provides a closed-form expression for the optimal \mathbb{P} . Specifically, given the following factorizations $W_s = X X^\top$ and $W_g = Y Y^\top$ of the state and gradient covariances, the optimal rank- r projector may be expressed as $\mathbb{P} = \Phi \Psi^\top$ with

$$\Phi = X V_r \Sigma_r^{-1/2} , \quad \Psi = Y U_r \Sigma_r^{-1/2} , \quad (14)$$

where U_r , V_r , and Σ_r are the leading r left and right singular vectors and singular values of the product $Y^\top X$. Notice that $\Psi^\top \Phi$ is equal to the identity matrix by construction, so that \mathbb{P} is indeed a projection. In the next subsection, we describe an efficient numerical procedure tailored specifically for the dynamical system in (6) to compute (an approximation of) the state and gradient covariance matrices W_s and W_g .

3.3. Computing the covariance matrices

In this section, we present a computationally-efficient approach to estimate the covariance factors X and Y that are needed to compute the CoBRAS projection (14). The state \mathbf{w} along trajectories of (6) is a function of the temporal variable t_0 and of the physical parameters $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$. Thus, if we draw t_0 , $\boldsymbol{\theta}$, and $\boldsymbol{\mu}$ from continuous mutually-independent distributions with probability density functions f_{T_0} , $f_{\boldsymbol{\theta}}$, and $f_{\boldsymbol{\mu}}$, the state covariance is defined as

$$W_s := \mathbb{E} [\mathbf{w}\mathbf{w}^\top] = \int \mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu}) \mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu})^\top f_{T_0}(t_0) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) f_{\boldsymbol{\mu}}(\boldsymbol{\mu}) dt_0 d\boldsymbol{\theta} d\boldsymbol{\mu} \approx X X^\top , \quad (15)$$

where X is a numerical-quadrature factor that is defined as in equations (20) and (21) of [57]. The computational burden associated with the numerical approximation of (15) lies in computing $\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu})$ along trajectories of (6), which requires integrating the numerically-stiff governing equations (6) from time 0 to time t_0 . Similar considerations hold for the estimation of the gradient covariance matrix, which may be written as

$$W_g := \mathbb{E} [\nabla F \nabla F^\top] = \int \nabla F(\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu})) \nabla F(\mathbf{w}(t_0; \boldsymbol{\theta}, \boldsymbol{\mu}))^\top f_{T_0}(t_0) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) f_{\boldsymbol{\mu}}(\boldsymbol{\mu}) dt_0 d\boldsymbol{\theta} d\boldsymbol{\mu} \approx Y Y^\top , \quad (16)$$

where Y is a numerical-quadrature factor defined similarly to X in (15), and a closed-form expression for ∇F is provided in the proposition below.

Proposition 1. Let $\mathbf{g}_j^{(k)} := (D_{\mathbf{w}} y_j(t_k; \boldsymbol{\theta}, \boldsymbol{\mu}))^\top$ denote the j -th column of the k -th block of $\nabla F \in \mathbb{R}^{N \times L \dim(\mathbf{y})}$, with $j \in \{1, 2, \dots, \dim(\mathbf{y})\}$ and $k \in \{0, 1, \dots, L - 1\}$. Then $\mathbf{g}_j^{(k)} = \boldsymbol{\xi}(t_0)$, where $\boldsymbol{\xi}(t)$ satisfies the (backward-in-time) adjoint equation

$$-\frac{d}{dt} \boldsymbol{\xi}(t) = D\mathbf{f}(\mathbf{w}(t; \boldsymbol{\theta}, \boldsymbol{\mu}))^\top \boldsymbol{\xi}(t), \quad \boldsymbol{\xi}(t_k) = \mathbf{C}^\top \mathbf{e}_j, \quad t \in [t_0, t_k], \quad (17)$$

and \mathbf{e}_j is the j -th unit vector of the standard basis of $\mathbb{R}^{\dim(\mathbf{y})}$.

Proof. The proof is given in Appendix A. \square

The adjoint equation (17) shares the same numerical stiffness as the forward nonlinear dynamics (6), and this can make its numerical integration expensive. In order to reduce the overall computational cost required to estimate the state and gradient covariance matrices, we propose to work with the *linearized* dynamics of (6). This will let us leverage the exact solution of linear ordinary differential equations, thereby bypassing the need for numerical integration of the high-dimensional stiff governing equations. Additionally, we shall see momentarily that the (approximate) gradient ∇F will become a function of $\boldsymbol{\theta}$ only, so that the (approximate) definition of W_g in (16) will no longer require integrating over t_0 and $\boldsymbol{\mu}$.

As previously mentioned, for a given value of θ , the dynamics in (6) exhibit a linearly stable thermochemical equilibrium solution, which we shall denote $\bar{\mathbf{w}}(\theta)$. Then, the linear dynamics of about $\bar{\mathbf{w}}(\theta)$ are governed by

$$\frac{d}{dt} \mathbf{w}'_{\theta}(t) = \underbrace{D\mathbf{f}(\bar{\mathbf{w}}(\theta))}_{:=\mathbf{A}_{\theta}} \mathbf{w}'_{\theta}(t), \quad \mathbf{w}'_{\mu,\theta}(0) = \mathbf{w}_0(\mu) - \bar{\mathbf{w}}(\theta), \quad (18)$$

where \mathbf{A}_{θ} is the state Jacobian evaluated about the equilibrium solution. Using the variation of constants formula, we can then approximate $\mathbf{w}(t_0; \theta, \mu)$ in (6) as follows

$$\mathbf{w}(t_0; \theta, \mu) \approx \bar{\mathbf{w}}(\theta) + e^{\mathbf{A}_{\theta} t_0} \mathbf{w}'_{\mu,\theta}(0). \quad (19)$$

Similarly, we can approximate $\mathbf{g}_j^{(k)}$ in Proposition 1 using the time-invariant adjoint equation

$$-\frac{d}{dt} \boldsymbol{\xi}(t) = \mathbf{A}_{\theta}^{\top} \boldsymbol{\xi}(t), \quad \boldsymbol{\xi}(t_k) = \mathbf{C}^{\top} \mathbf{e}_j, \quad t \in [t_0, t_k], \quad (20)$$

where \mathbf{A}_{θ} is precisely the zeroth-order approximation of $D\mathbf{f}(\mathbf{w}(t; \theta, \mu))$. Once again, using the variation of constants formula, we have

$$\mathbf{g}_j^{(k)} = e^{\mathbf{A}_{\theta}^{\top}(t_k - t_0)} \mathbf{C}^{\top} \mathbf{e}_j. \quad (21)$$

Here, we see that $\mathbf{g}_j^{(k)}$ is a parametric function of θ only. More specifically, we remark that $\mathbf{g}_j^{(k)}$ is not a function of t_0 , but rather of $t_k - t_0$, and this is consistent with the time-invariant nature of equation (20). It follows that the (approximate) gradient covariance matrix in (16) can be written as

$$W_g \approx \int \nabla F(\theta) \nabla F(\theta)^{\top} f_{\Theta}(\theta) d\theta. \quad (22)$$

In equations (19) and (21) the matrix exponential can be computed efficiently using the eigendecomposition of \mathbf{A}_{θ} . The overall procedure to obtain the projection factors Φ and Ψ is illustrated in Algorithm 1.

Remark. In this paper, we consider systems that exhibit one single equilibrium for a given value of θ . In systems where multiple equilibria coexist, the same approximations illustrated in equations (19) and (21) hold around any one of the coexisting steady states and the user may choose freely which steady state to use for the linearization.

Algorithm 1 Computing the factor Φ and Ψ for the linear projection operator $\mathbb{P} = \Phi\Psi^{\top}$.

Require: Set of quadrature points for temperatures and densities (\mathcal{P}), initial condition parameters (\mathcal{M}), and initial times (\mathcal{T}_0), along with their associated quadrature weights, to approximate the integrals in equations (15) and (16).

- 1: Initialize dynamic set \mathcal{W}_0 .
 - 2: **for** $\mu \in \mathcal{M}$ **do**
 - 3: Compute the parametrized initial condition $\mathbf{w}_0(\mu)$.
 - 4: Add \mathbf{w}_0 to the set \mathcal{W}_0 .
 - 5: **end for**
 - 6: Initialize dynamic arrays X and Y .
 - 7: **for** $\theta \in \mathcal{P}$ **do**
 - 8: Compute the equilibrium steady-state $\bar{\mathbf{w}}(\theta)$.
 - 9: Construct matrix \mathbf{A}_{θ} and compute its eigendecomposition $\mathbf{A}_{\theta} = \mathbf{V}\Lambda\mathbf{V}^{-1}$.
 - 10: Computing state covariance
 - 11: **for** $\mathbf{w}_0 \in \mathcal{W}_0$ **do**
 - 12: **for** $t_0 \in \mathcal{T}_0$ **do**
 - 13: Compute $\mathbf{w}(t_0; \theta, \mu)$ using equation (19) and $e^{\mathbf{A}_{\theta} t_0} = \mathbf{V}e^{\Lambda t_0}\mathbf{V}^{-1}$.
 - 14: Append $\sqrt{\delta_s} \mathbf{w}(t_0; \theta, \mu)$ to X , where $\delta_s = f_{T_0}(t_0) f_{\Theta}(\theta) f_{\mathcal{M}}(\mu) w_{t_0} w_{\theta} w_{\mu}$, and w_{t_0} , w_{θ} , and w_{μ} are the quadrature weights.
 - 15: **end for**
 - 16: **end for**
 - 17: Computing gradient covariance
 - 18: **for** $j \in \{1, 2, \dots, \dim(\mathbf{y})\}$ **do**
 - 19: **for** $k \in \{0, 1, \dots, L-1\}$ **do**
 - 20: Compute $\mathbf{g}_j(t_k; \theta)$ using equation (21) and $e^{\mathbf{A}_{\theta}^{\top}(t_k - t_0)} = \mathbf{V}^{-1} e^{\Lambda(t_k - t_0)} \mathbf{V}^{\top}$.
 - 21: Append $\sqrt{\delta_g} \mathbf{g}_j(t_k; \theta)$ to Y , where $\delta_g = f_{\Theta}(\theta) w_{\theta}$.
 - 22: **end for**
 - 23: **end for**
 - 24: **end for**
 - 25: Perform SVD $Y^{\top} X = U \Sigma V^{\top}$.
 - 26: Compute matrices Φ and Ψ using equation (14).
-

4. Existing model reduction formulations

The conventionally used model reduction techniques primarily focus on predicting macroscopic quantities, such as total species mass and internal energy, which are essential in flow measurements and simulations. These quantities can be modeled using additional conservation equations, beyond the standard hydrodynamic equations, to account for variations in chemical composition and the nonequilibrium relaxation of energy modes. In the most general formulation, this additional set of governing equations can be derived by taking successive moments of the master equations (3), as follows

$$\sum_{i \in I_g \subseteq I_s} (\epsilon_s^i)^m \frac{dn_s^i}{dt} = \Omega_{s,m}^g, \quad \forall g \in \{1, \dots, N_g\}. \quad (23)$$

In this context, g represents a pseudo-species, which refers to a particular species internal degree of freedom treated as a state variable. The parameter m indicates the moment order, while $\Omega_{s,m}^g$ denotes the reactive source term of order m for the pseudo-species g . Based on the assumptions used to define the subset $I_g \subseteq I_s$, two distinct simplified models can be derived.

4.1. Coarse-grained models

If the subset I_g represents a group of states, the approach is named coarse-grained (CG) modeling [35,38–43,62]. The construction of such a model involves a two-step procedure: (i) grouping energy states into N_g macroscopic bins (or groups) according to a specified strategy, and (ii) defining a bin-wise distribution function to represent the population within each group, subject to a series of moment constraints. Various clustering techniques have been proposed for step (i) [34,41,42,45] and in this work, we adopt the advanced spectral clustering technique proposed by Sahai et al. [41]. For step (ii), Liu et al. [38] applied the maximum entropy principle to derive a log-linear representation of the bin-wise distribution function, yielding a thermalized local Boltzmann distribution for each bin. This distribution is expressed as follows

$$\ln \frac{g_s^i}{n_s^i} = \alpha_s^g + \beta_s^g \epsilon_s^i, \quad i \in I_g, \quad (24)$$

where the bin-specific coefficients α_s^g and β_s^g are expressed as a function of the macroscopic group constraints (i.e., number density and energy). The populations and internal energies of the different bins serve as the unknowns in the reduced-order system, leading to a total of $2 \times N_g$ additional conservation equations. For further details, please refer to [38].

This model reduction technique effectively maintains positivity and guarantees positive entropy production by design. However, these advantages come with practical challenges due to the use of a nonlinear function of exponential nature. Specifically, coarse-grained rates must be interpolated across group energies (or temperatures), and this can increase computational costs and introduce potential errors and instabilities, as it necessitates prior knowledge of the relevant temperature ranges.

4.2. Multi-temperature models

In the specific case of binning one group per internal energy mode, which is a particular case of the CG approach, multi-temperature (MT) models are employed [2,30,32]. This framework adds an additional conservation equation for the species total number density, along with individual conservation equations for each energy mode (vibrational, rotational, and electronic). While this method is still commonly used in the CFD community, many of these models are based on simplistic, semi-empirical assumptions. They often rely on a rigid separation of internal energy modes without rigorous derivation from fundamental kinetic equations or consideration of physical principles. In this study, we adopt a more rigorous approach, utilizing the recent quasi-steady-state (QSS) method [10], which computes its kinetic database directly from state-to-state calculations.

5. Numerical experiments

In this section, we evaluate the performance of the proposed CoBRAS model on two distinct thermochemical systems. Both of these involve a pure oxygen mixture, denoted as $S = \{O, O_2\}$, with both species in their electronic ground state.

- i. The first system is a rovibrational collisional (RVC) model, which considers only molecule-atom collisions and has been extensively studied in prior works [10,38,41,42]. This model serves as a benchmark to compare CoBRAS with traditional approaches, including coarse-graining (CG), multi-temperature models (MT), and the POD-Galerkin approach [48,49].
- ii. The second system is a vibrational collisional (VC) model, which includes both molecule-molecule and molecule-atom collisions. The resulting equations are quadratic in the O_2 energy level population. Despite the nonlinearity, we demonstrate that the linearized dynamics around thermochemical equilibrium remain representative of the subspaces on which the nonlinear dynamics evolve, and this justifies the use of the procedure described in section 3.3 to approximate the covariance matrices. Here, we compare CoBRAS against POD-Galerkin only since we have no access to validated solvers that implement the CG approach on this system.

Table 1

Sampled parameter space for the RVC model. Sampling bounds and distributions for each parameter used to generate the training and testing trajectories.

	T [K]	ρ [kg/m ³]	w_{O_0}	T_0 [K]
Minimum	5 000	10 ⁻⁴	0.01	500
Maximum	15 000	1	1	5 000
Distribution	Uniform	Log-uniform	Uniform	Log-uniform

In all (Petrov-)Galerkin models implemented in this section, we seek a reduced-order representation of the O₂ energy levels, while the O dynamics are accounted for exactly (see sections S.1.2 and S.2.2 in the Supplementary Material). All test cases were implemented in Python, employing the SciPy `solve_ivp` function with the LSODA method for time integration. The code used for our experiments can be found at <https://github.com/ivanZanardi/ronek>. All the CG and MT solutions were computed using the PLATO (PLAsma in Thermodynamic nOnequilibrium) library [63].

5.1. O₂-O rovibrational collisional model

In our first numerical experiment, we investigate the rovibrational excitation and dissociation of an O₂ molecule colliding with an O atom. The O₂ molecule has 6 115 rovibrational energy levels, resulting in a total of $N = 6\,116$ degrees of freedom for the system. The kinetic database for the O₂-O system comprises two types of processes:

- i. collisional dissociation,



- ii. collisional excitation, including both inelastic (nonreactive) and exchange processes,



The RVC model here described is governed by the following set of equations

$$\frac{dn_i}{dt} = \sum_{j \in I} \left(-k_{ij}^e n_i n_O + k_{ji}^e n_j n_O \right) - k_i^d n_i n_O + k_i^r n_O^3, \quad (27)$$

$$\frac{dn_O}{dt} = -2 \sum_{i \in I} \frac{dn_i}{dt}. \quad (28)$$

For interested readers, the linearized FOM equations necessary to implement the method described in Sections 3.1 to 3.3, as well as the derivation of the ROM, are available in Supplementary Material sections S.1.1 and S.1.2.

The bounds and distributions used to sample training and testing trajectories for the RVC model are provided in Table 1. Here, w_{O_0} denotes the initial mass fraction of O, with $w_{O_2} = 1 - w_{O_0}$ always applicable. For training, we use Arrhenius-fitted rates derived from quasi-classical trajectory (QCT) calculations, evaluated at 10 different temperatures uniformly sampled within the range of $[5, 15] \times 10^3$ K. This range includes the boundary temperatures, with a step size of $\Delta T = 1\,111.11$ K. For testing, we select 10 randomly chosen temperatures within the same range using LHS. A total of $M = 1\,000$ testing trajectories are sampled, with 100 trajectories for each selected temperature.

5.1.1. Linearized FOM performances

In this section, we assess the performance of the linearized FOM employed in constructing the Petrov-Galerkin ROM. In particular, our goal is to demonstrate that the procedure illustrated in section 3.3 leads to significant computational savings while still capturing the relevant subspaces for Petrov-Galerkin modeling. The computational cost is tabulated in Table 2, where we compare the cost required to generate one (or more) trajectories of the nonlinear FOM in equation (6) and the cost of evaluating (19) using the matrix exponential. In particular, the third column of the table shows that once the eigendecomposition of \mathbf{A}_θ is available, evaluating (19) is four orders of magnitude faster than time-stepping (6). For a more realistic comparison, however, we should also account for the cost associated with computing the eigendecomposition, which scales with the cube of the matrix dimension. This operation is indeed costly, but fortunately, it can be performed once (per equilibrium) and then held in memory (or stored) for future use. The last column in Table 2 shows that, despite the costly eigendecomposition, we still achieve a factor of 400 speed-up when generating 200 trajectories. A comparable speed-up is also observed when solving the adjoint equations, with the linearized form (20) proving significantly more efficient than the nonlinear one (17).

In order to assess the effectiveness of the linearized dynamics in identifying the appropriate subspaces for model reduction, we simply compute the mean relative L^2 error (Fig. 1)

Table 2

Comparison of computational cost: FOM vs. linearized FOM for the RVC Model. This table illustrates the performance differences between the FOM and its linearized counterpart across various numerical operations.

Model	Operation	Device	Time [s] - 1 Run	Time [s] - 200 Runs
FOM	Simulation	CPU - 8 Threads	1.25×10^2	2.50×10^4
Linearized FOM	Eigendecomposition	CPU - 8 Threads	6.17×10^1	6.17×10^1
	Simulation	CPU - 8 Threads	2.19×10^{-2}	4.38

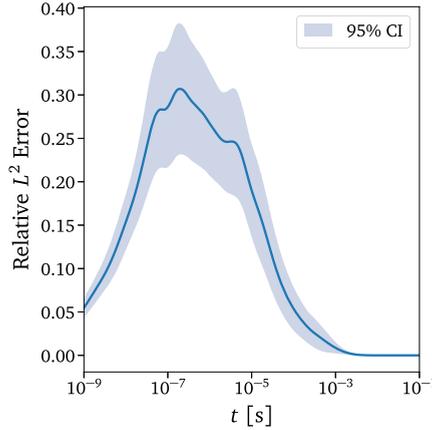


Fig. 1. Relative L^2 error of the linearized FOM for the RVC model. This figure presents the time evolution of the mean relative L^2 error, along with a 95% confidence interval derived from 1 000 testing trajectories, using the linearized FOM.

$$e(t) = \frac{1}{M} \sum_{k=1}^M \frac{\|\mathbf{w}^{(k)}(t) - \mathbf{w}_{\text{lin}}^{(k)}(t)\|_2}{\|\mathbf{w}^{(k)}(t)\|_2 + \varepsilon}, \quad (29)$$

between \mathbf{w}_{lin} given by (19) and the solution \mathbf{w} of (6) over $M = 1\,000$ trajectories. Here, we use $\varepsilon = 10^{-7}$ to avoid division by zero. The relative L^2 error is particularly relevant here because the methodology described in sections 3.1 to 3.3 produces optimal test and trial bases in the L^2 sense. With a 95% confidence interval, the maximum relative L^2 error is approximately 30%. This finding suggests that the linearized equations are sufficiently accurate to identify appropriate subspaces for Petrov-Galerkin projection of the full-order nonlinear equations. (In the interest of clarity, we remark that the error goes to zero at long time because the linear system describes fluctuations around the equilibrium solution.)

5.1.2. Models for macroscopic quantities

In this section, we seek models that accurately predict the time evolution of the system's macroscopic quantities such as mass and energy, and we therefore choose $\mathbf{y} \in \mathbb{R}^2$ in (7) to contain the zeroth and first-order moments (i.e., the mass and energy, respectively) of the distribution. Fig. 2 illustrates the time evolution of the mean relative error between the FOM output $y_j(t)$ and the predicted output $\hat{y}_j(t)$ from CoBRAS (left panels) and POD (right panels) for the $j = 0$ (top row) and $j = 1$ (bottom row) moments of O_2 energy level population. This analysis is based on $M = 1\,000$ different testing trajectories across various reduced system dimensions, r . (The equation for atomic oxygen is not included in r .) The figure is essential for understanding how the model performs over time and across different reduced dimensions. With CoBRAS, the error remains consistently low, under 1%, even for a small reduced dimension of $r = 7$, representing a compression of approximately 870 times. By contrast, the POD method yields predictions with errors consistently exceeding 1%. Similar qualitative performance is observed for the time evolution of the energy (bottom row), where we see that CoBRAS maintains low errors across all times, while POD exhibits errors that are consistently above 10%.

Fig. 3 presents a comparative analysis of various reduced-order modeling approaches for a test case with initially cold gas. The test condition includes a system temperature of $T = 10\,000$ K, an initial pressure of $p_0 = 1\,000$ Pa, initial molar concentration of atomic oxygen $x_{\text{O}_0} = 0.05$, and initial equilibrium temperature $T_0 = 500$ K. In the left panels, we show the time evolution of the zeroth and first-order moments, while the right panels illustrate the relative error. Additional comparisons at various testing temperatures for the same initial condition are provided in figure S1 of the Supplementary Material. These figures are crucial for evaluating the performance of CoBRAS technique against the POD method, the traditional ROMs, and the reference FOM solution. The dimensions of the reduced-order models are as follows: $r = 8$ for CoBRAS and POD ROMs, $r = 24$ for CG (using 12 bins), and $r = 2$ for MT. The left panels demonstrate that for both moments, the CoBRAS-based ROM closely matches the FOM, with a maximum error consistently below 1%. Additionally, CoBRAS outperforms the POD, CG, and MT approaches, each of which fails to accurately predict the system dynamics. These performance differences become even more pronounced in the error evolution depicted in the right panels. For the zeroth-order moment, the point-wise error for POD, CG, and MT approaches 100%, whereas for the first-order moment, it is approximately 10% for CG and over 100% for POD and MT. The high error in MT's energy evolution is due to the separation of

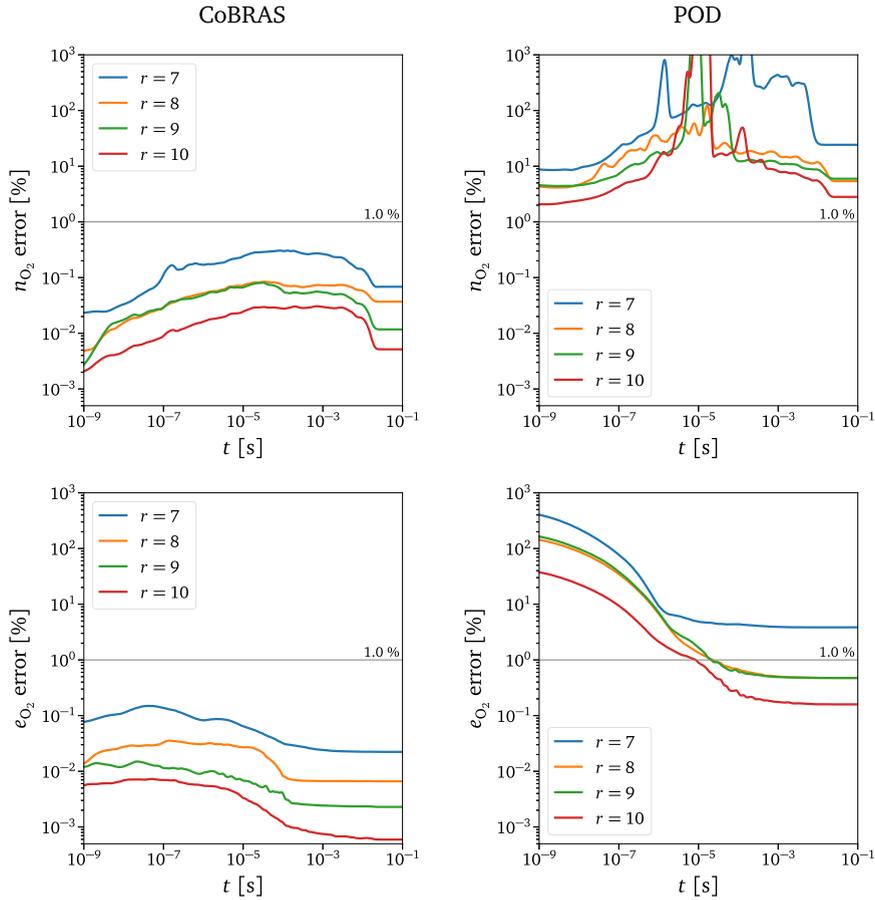


Fig. 2. Mean moments error for RVC model using CoBRAS and POD ROMs. Time evolution of the mean relative errors for the zeroth (top row) and first-order (bottom row) moments of the O_2 energy level population across 1000 testing trajectories. Results are shown for different dimensions r of the CoBRAS (left panels) and POD (right panels) ROM systems.

energy modes, where the rotational mode is in equilibrium with the translational mode at temperature T , while the vibrational energy is modeled with a fictitious temperature T_v . This separation of energy modes is a fundamental assumption distinguishing MT from the other ROMs, rooted in physical considerations not discussed here.

5.1.3. Models for microscopic quantities

Here, we seek models that can provide an accurate description of the *whole* distribution along trajectories. In principle, this can be done by choosing the output \mathbf{y} to be equal to the state vector \mathbf{w} , but the evaluation of the gradient covariance would be rather expensive (given that ∇F scales with $\dim(\mathbf{y})$). Instead, we choose $\mathbf{y} \in \mathbb{R}^{10}$ to contain the first ten moments of the distribution, since user experience suggests that a model that is capable of tracking these moments is likely capable of producing accurate estimates of the distribution itself. Fig. 4 presents the time evolution of the mean relative error between the FOM solution $\mathbf{w}(t)$ and the predicted solutions $\hat{\mathbf{w}}(t)$ obtained from CoBRAS (left panel) and POD (right panel) for the O_2 energy level population. The error is calculated over $M = 1000$ trajectories and for different reduced system dimension r . The figure highlights the model's accuracy over time and across different reduced dimensions, demonstrating its effectiveness in predicting microscopic quantities. The error analysis is consistent with the considerations made for Fig. 2, indicating that the overall error remains below 10% when using CoBRAS with $r \geq 28$, which corresponds to a compression factor of nearly 220. In contrast, the POD technique consistently yields results with higher errors compared to those produced by CoBRAS. It is worth remarking that neither POD nor CoBRAS guarantee positivity of the predicted state $\hat{\mathbf{w}}$, and we have observed that non-physical negative values may appear at very early times $t < 10^{-8}$ s. This is particularly true at low initial temperatures $T_0 \leq 500$ K, where the distribution spans approximately 20 orders of magnitude ($10^3 - 10^{23}$), and the tail of the predicted distribution may take negative values.

Fig. 5 presents four snapshots of the O_2 internal state population taken from the FOM and from CoBRAS, POD, and the CG reduced-order models, each operating at a reduced dimensionality of $r = 24$ (with 12 bins for the CG approach). This figure is central for evaluating the effectiveness of the CoBRAS-based ROM in capturing microscopic scales of the physical system compared to POD and the conventional CG technique. This test case corresponds to the scenario illustrated in Fig. 3. Further comparisons at various testing temperatures for the same initial condition are provided in figure S2 of the Supplementary Material. In qualitative

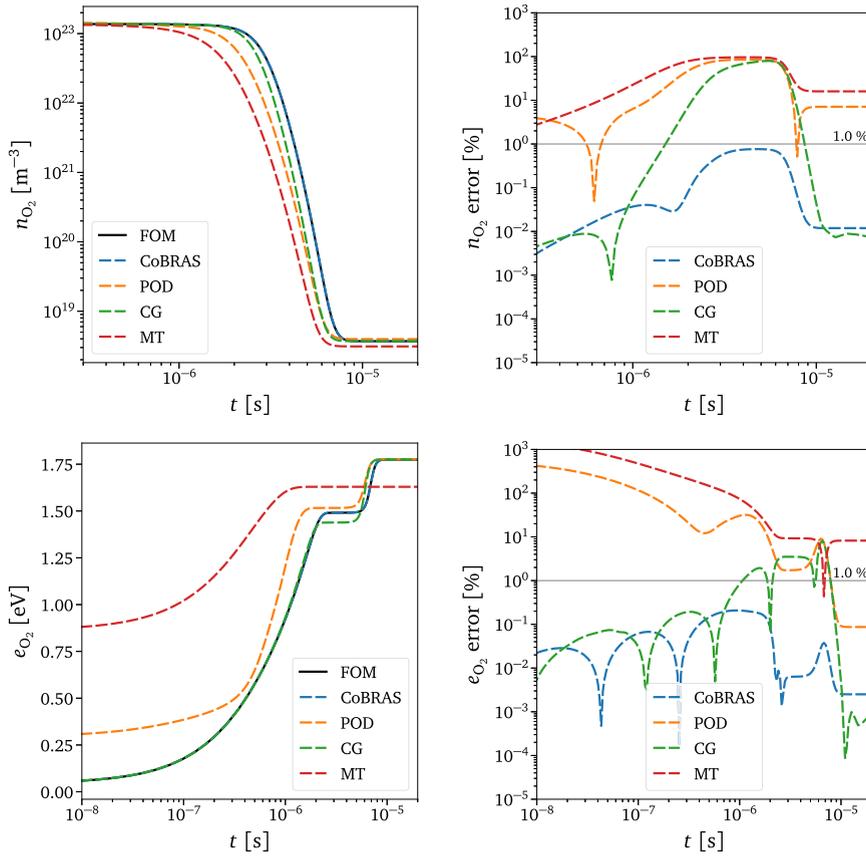


Fig. 3. FOM vs. ROMs for RVC model: moments evolution. Time evolution of the zeroth (top row) and first-order (bottom row) moments of the O₂ energy level population in an initially cold gas with $T = 10\,000$ K, as calculated using the FOM and various ROMs. The left panels display the computed moment evolution, while the right panels show the relative errors of the ROMs compared to the FOM. The CoBRAS and POD ROMs use a reduced model dimension of $r = 8$, the CG ROM uses $r = 24$, and the MT ROM uses $r = 2$.

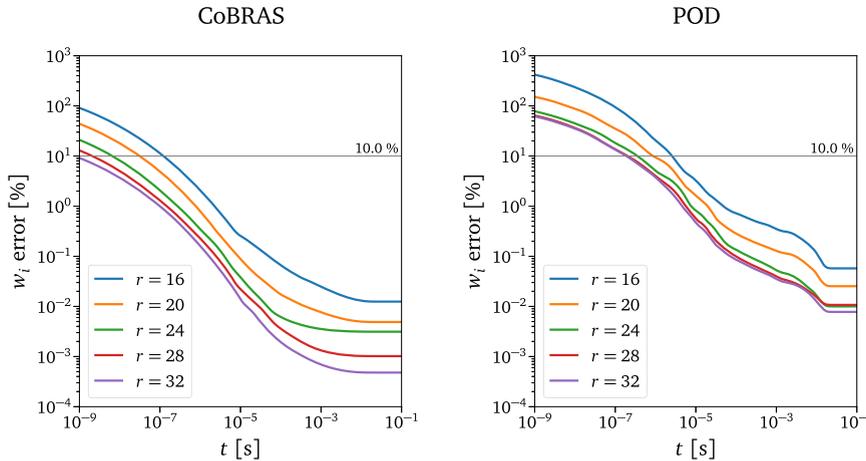


Fig. 4. Mean relative error in the distribution function for RVC model using CoBRAS and POD ROMs. Time evolution of the mean relative error in the O₂ energy level population across 1 000 testing trajectories. Results are shown for different dimensions r of the CoBRAS (left panel) and POD (right panel) ROM systems.

terms, CoBRAS displays consistently higher fidelity to the FOM across all snapshots, effectively capturing the microscopic details of the distribution function with greater accuracy than the other ROMs. Additionally, although the POD model demonstrates improved performance compared to the CG model, it still does not reach the accuracy achieved by the CoBRAS model. The discrepancy between CoBRAS and CG is particularly pronounced at $t = 3 \times 10^{-6}$ s, where the quasi-steady-state (QSS) distribution is present. Here CoBRAS matches the reference solution very accurately, whereas the CG shows a vertical shift indicative of faster dynamics, as previously noted in Fig. 3. Notably, all ROMs successfully reconstruct the equilibrium distribution at $t = 10^{-4}$ s.

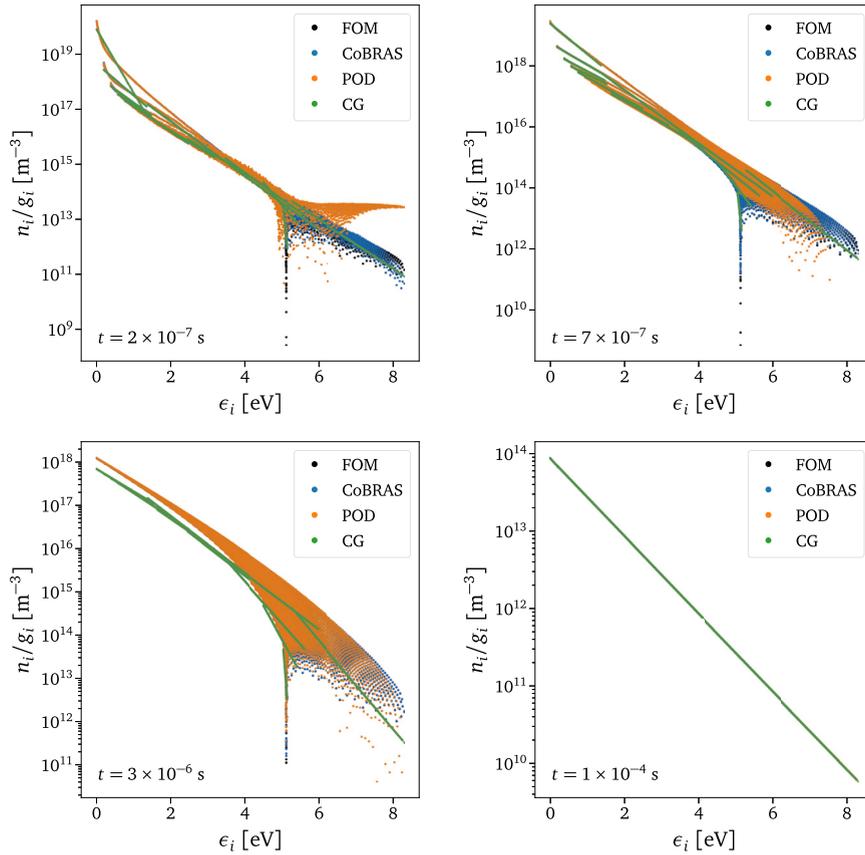


Fig. 5. FOM vs. ROMs for RVC model: distribution snapshots. Time snapshots of the O_2 energy level population in an initially cold gas with $T = 10000$ K, calculated using FOM, CoBRAS, POD, and CG reduced-order models. All ROMs have a dimension of $r = 24$.

Table 3

Computational cost comparison: FOM vs. CoBRAS ROM for the RVC model. The table reports total FLOPs and FLOP/Byte ratios for key numerical operations, including right-hand side (RHS) evaluation and linear system solve (LSS), along with estimates of system stiffness. Results are shown for the FOM and two ROM configurations.

Model	d	FLOPs		FLOP/Byte		Fastest Timescale $1/ \lambda_{\min} $
		RHS - $\mathcal{O}(d^2)$	LSS - $\mathcal{O}(d^3)$	RHS - $\mathcal{O}(1)$	LSS - $\mathcal{O}(d)$	
FOM	6 116	7.480×10^7	1.526×10^{11}	0.250	509.750	1.36×10^{-10}
CoBRAS	25	1.225×10^3	1.167×10^4	0.227	2.160	5.35×10^{-10}
	9	1.530×10^2	6.480×10^2	0.193	0.818	1.29×10^{-9}

5.1.4. Computational performance

In this section, we assess computational performance by examining the total number of floating-point operations (FLOPs) required to solve the FOM and the CoBRAS-based ROM at different dimensions. We also report the *arithmetic intensity* (see Appendix C), defined as the ratio of FLOPs to memory access in bytes (FLOP/Byte), for key numerical operations, along with estimates of system stiffness. While we acknowledge that additional hardware-dependent factors such as memory bandwidth, processor speed, and cache efficiency influence actual runtime performance, the FLOP count and FLOP/Byte ratio provide hardware-independent metrics that serve as useful proxies for comparing algorithmic efficiency.

As shown in Table 3, where d denotes the system dimension ($d = r + 1$ for the ROM), the ROM reduces the total number of FLOPs by over four orders of magnitude for right-hand side (RHS) evaluations, which primarily involve dense matrix-vector products with $\mathcal{O}(d^2)$ complexity (see equations (S.1.2)–(S.1.3) in the Supplementary Material and Appendix C.2), and by over seven orders of magnitude for the linear system solve (LSS) required by the implicit integration scheme, which has $\mathcal{O}(d^3)$ complexity (see Appendix C.1). Additionally, the ROM results in a less stiff system, as indicated by the fastest timescale, quantified as the inverse of the smallest Jacobian eigenvalue, $1/|\lambda_{\min}|$. This reduced stiffness allows for larger time steps during time integration and contributes further to the computational savings.

However, the efficiency of the RHS and LSS operations is ultimately constrained by their *arithmetic intensity*. Both operations exhibit relatively low FLOP/Byte ratios in the ROM setting, indicating that performance is likely to be limited by memory bandwidth rather than computational throughput. In such cases, the actual runtime gains from FLOP reductions may be smaller than expected due to the bottleneck in data transfer between memory and processor.

5.2. O_2 - O_2 and O_2 - O vibrational collisional model

In this experiment, we focus on the vibrational excitation and dissociation of O_2 molecules during collisions with either O atoms or other O_2 molecules. The O_2 molecule has 45 vibrational energy levels, resulting in a total of $N = 46$ degrees of freedom for the system. We opted not to use the RVC model, as it would require managing an impractical kinetic database with billions of possible transitions. Instead, we employed the VC model, which is based on the RVC model but relies on the simplifying assumption of equilibrium (Maxwell-Boltzmann) distribution of the rotational levels,

$$\frac{n_i}{n_v} = \frac{q_i}{Q_v}, \quad v \in \mathcal{V}, \quad i \in \mathcal{I}_v, \quad (30)$$

with

$$q_i = g_i \exp\left(-\frac{\epsilon_i}{k_B T_v}\right) \quad (31)$$

and $Q_v = \sum_{i \in \mathcal{I}_v} q_i$ being the partition function of the vibrational level v . \mathcal{V} denotes the set of all vibrational levels, while \mathcal{I}_v refers to the set of rotational levels with the same vibrational level v . In the VC model, it is further assumed that rotation and translation are in equilibrium, such that $T_v = T$ for all $v \in \mathcal{V}$. Each vibrational level can be characterized by an energy ϵ_v and a fictitious degeneracy g_v , which are determined as follows

$$\epsilon_v = \frac{1}{Q_v} \sum_{i \in \mathcal{I}_v} q_i \epsilon_i, \quad (32)$$

$$g_v = Q_v \exp\left(\frac{\epsilon_v}{k_B T}\right). \quad (33)$$

These values are then used in the computation of backward rates by enforcing detailed balance.

The kinetic database includes O_2 - O collisions as described in section 5.1, properly reduced for the equilibrium assumptions of rotational levels, as well as O_2 - O_2 collisions, which involve three types of processes:

- i. collisional excitation, including both inelastic (nonreactive) and exchange processes,



- ii. combined collisional dissociation and excitation,



- iii. collisional dissociation,



where w , p and q represent different vibrational levels. The VC model here described is governed by the following set of equations:

$$\begin{aligned} \frac{dn_v}{dt} = & \sum_{w,p,q \in \mathcal{V}} \left(-m_k^{c_{vwpq}} n_v n_w + m_k^{c_{pqvw}} n_p n_q \right) \\ & + \sum_{w,p \in \mathcal{V}} \left(-m_k^{ed_{vw}} n_v n_w + m_k^{c_{pvw}} n_p n_O^2 \right) \\ & + \sum_{w \in \mathcal{V}} \left(-m_k^d n_v n_w + m_k^r n_O^4 \right) \\ & + \sum_{w \in \mathcal{V}} \left(-{}^a k_{vw}^c n_v n_O + {}^a k_{vw}^c n_w n_O \right) \\ & - {}^a k_v^d n_v n_O + {}^a k_v^r n_O^3, \end{aligned} \quad (37)$$

$$\frac{dn_O}{dt} = -2 \sum_{v \in \mathcal{V}} \frac{dn_v}{dt}. \quad (38)$$

Table 4

Sampled parameter space for the VC model. Sampling bounds and distributions for each parameter used to generate the training and testing trajectories.

	T [K]	ρ [kg/m ³]	w_{O_0}	T_0 [K]
Minimum	10 000	10^{-4}	0	500
Maximum	10 000	1	1	10 000
Distribution	-	Log-uniform	Uniform	Log-uniform

Table 5

Comparison of computational cost: FOM vs. linearized FOM for the VC Model. This table illustrates the performance differences between the FOM and its linearized counterpart across various numerical operations.

Model	Operation	Device	Time [s] - 1 Run	Time [s] - 200 Runs
FOM	Simulation	CPU - 4 Threads	6.14×10^{-2}	1.23×10^1
Linearized FOM	Eigendecomposition	CPU - 4 Threads	7.36×10^{-4}	7.36×10^{-4}
	Simulation	CPU - 4 Threads	5.16×10^{-5}	1.03×10^{-2}

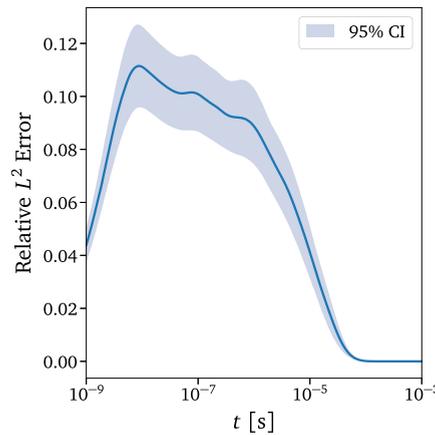


Fig. 6. Relative L^2 error of the linearized FOM for the VC model. This figure presents the time evolution of the mean relative L^2 error, along with a 95% confidence interval derived from 1 000 testing trajectories, using the linearized FOM.

For interested readers, the linearized FOM equations necessary to implement the method described in Sections 3.1 to 3.3, as well as the derivation of the ROM, are available in Supplementary Material sections S.2.1 and S.2.2.

The bounds and distributions used for sampling the training and the $M = 1\,000$ testing trajectories for the VC model are provided in Table 4. For this thermochemical system, we consider a single temperature of $T = 10\,000$ K.

5.2.1. Linearized FOM performances

In line with the analysis in section 5.1.1, we assess the efficiency and accuracy of the linearized FOM to estimate the state and gradient covariance matrices associated with the VC model. As in the previous section, we observe that using the linearized FOM leads to a significant reduction in computational cost (see Table 5) while also providing a very good approximation of the subspaces on which the nonlinear dynamics evolve (see Fig. 6). In this experiment, we use half the number of threads (4 instead of 8) due to the smaller problem dimension compared to the first experiment.

5.2.2. Models for macroscopic quantities

In this section, we seek models that accurately predict the time evolution of the system's macroscopic quantities such as mass and energy, and we therefore choose $\mathbf{y} \in \mathbb{R}^2$ in (7) to contain the zeroth and first-order moments (i.e., the mass and energy, respectively) of the distribution. Fig. 7 presents the time evolution of the mean relative error between the FOM output $y_j(t)$ and the predicted output $\hat{y}_j(t)$ from the CoBRAS (left panels) and POD (right panels) ROMs for the $j = 0$ (top row) and $j = 1$ (bottom row) moments of the O_2 energy level population in the VC model. This analysis is conducted over $M = 1\,000$ test trajectories for various reduced system dimensions, r . (The equation for atomic oxygen is not included in r .) The figure provides insight into the model's accuracy over time and across different reduced dimensions when evaluating macroscopic quantities. For CoBRAS, the overall error remains consistently low, staying below 1% for both moments even with a reduced system size of $r = 3$, achieving a compression rate above 11. By contrast, the POD ROM is significantly more inaccurate and it may even be unstable for very low dimensions (see the $r = 3$ curve in the top right panel). Similar qualitative behavior is observed in the bottom row of Fig. 7, which is the analog of the top row for the first-order moment (i.e., the energy of the O_2 population).

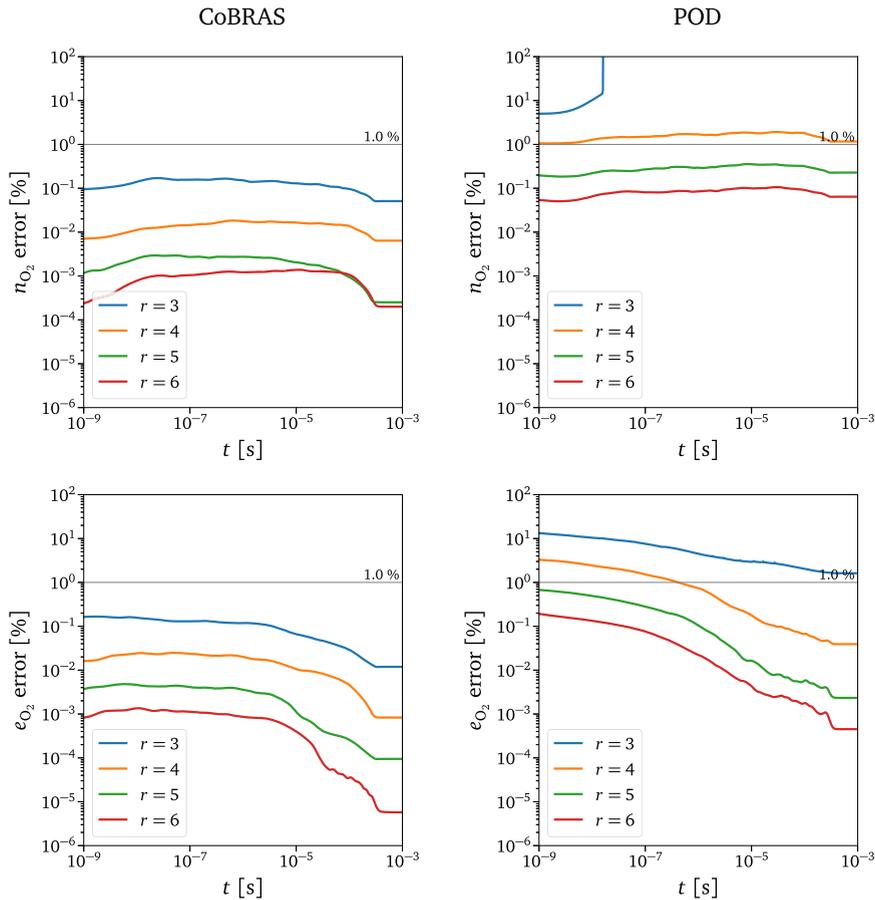


Fig. 7. Mean moments error for VC model using CoBRAS and POD ROMs. Time evolution of the mean relative errors for the zeroth (top row) and first-order (bottom row) moments of the O_2 energy level population across 1000 testing trajectories. Results are shown for different dimensions r of the CoBRAS (left panels) and POD (right panels) ROM systems.

Fig. 8 provides a comparative analysis of different reduced-order modeling techniques for a test case involving an initially cold gas, focusing on the zeroth and first-order moments of the O_2 internal energy distribution function. In the left panels, we show the time evolution of these moments, while the right panels display the relative error. The initial conditions match those in Fig. 3, with the ROM dimension set to $r = 4$. These figures are essential for assessing the performance of the CoBRAS technique in comparison to the POD method and the reference FOM solution for macroscopic quantities. As shown in the left panels, CoBRAS closely matches the FOM results, maintaining a maximum error below 1% over the entire time frame for both moments. This demonstrates the superior accuracy of CoBRAS relative to POD, for which the error exceeds 10%.

5.2.3. Models for microscopic quantities

Here, we seek models that can provide an accurate description of the *whole* distribution along trajectories. We follow the same rationale as in section 5.1.3 and we choose $\mathbf{y} \in \mathbb{R}^{10}$. Fig. 9 illustrates the time evolution of the mean relative error between the FOM solution $\mathbf{w}(t)$ and the predicted solutions $\hat{\mathbf{w}}(t)$ obtained from the CoBRAS (left panel) and POD (right panel) ROMs for the O_2 energy level population. The comparison uses the same set of $M = 1000$ trajectories as in the previous section, evaluating the ROMs for various reduced system dimensions r . This figure emphasizes the accuracy of the models over time and across different reduced dimensions in capturing microscopic quantities. As for the RVC system that we considered in the previous sections, the overall error remains below 10% for most of the dynamics when employing CoBRAS with $r \geq 10$, corresponding to a compression factor of nearly 4.5. By contrast, POD-Galerkin consistently produces results with higher errors. As already noted in the RVC system, neither POD-Galerkin nor CoBRAS guarantee the positivity of the predicted state $\hat{\mathbf{w}}$ and this can lead to non-physical negative values at very early times.

Fig. 10 displays four snapshots of the distribution function calculated using the FOM, along with CoBRAS and POD ROMs, with the reduced model dimension set to $r = 8$. This case corresponds to the one presented in Fig. 8. For all snapshots, CoBRAS closely matches the FOM results, whereas POD exhibits significant discrepancies in the tails (i.e., high energy levels) of the distribution function. Notably, all ROMs successfully reconstruct the equilibrium distribution at $t = 10^{-4}$ s.

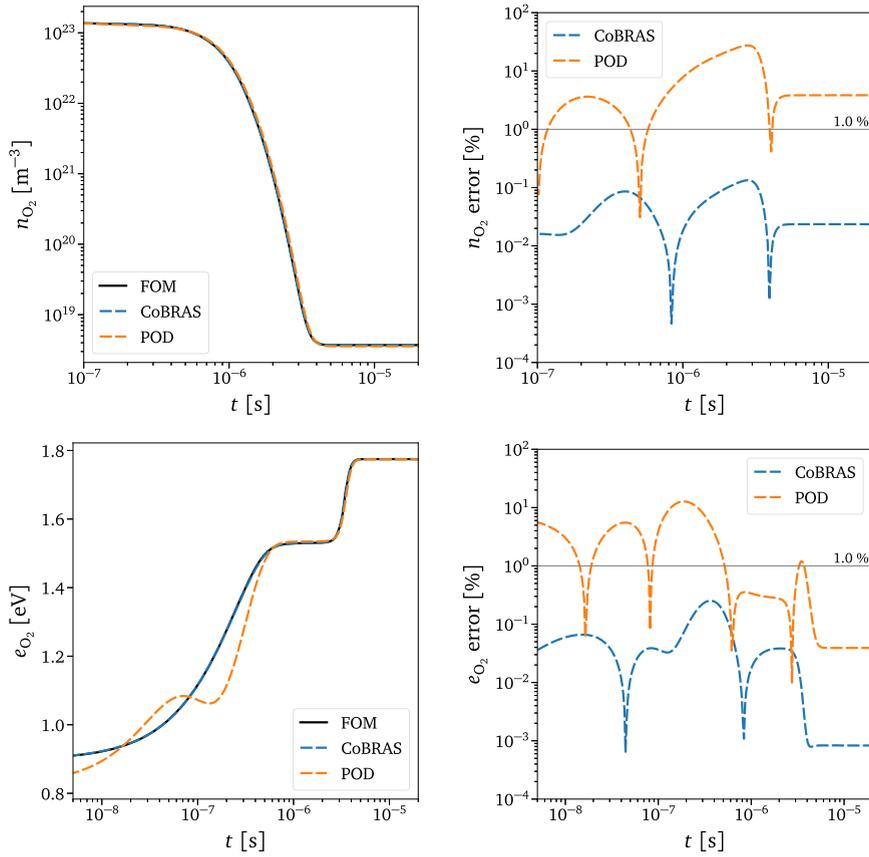


Fig. 8. FOM vs. ROMs for VC model: moments evolution. Time evolution of the zeroth (top row) and first-order (bottom row) moments of the O_2 energy level population in an initially cold gas with $T = 10\,000$ K, as calculated using the FOM and various ROMs. The left panels display the computed moment evolution, while the right panels show the relative errors of the ROMs compared to the FOM. The CoBRAS and POD ROMs utilize a reduced model dimension of $r = 4$.

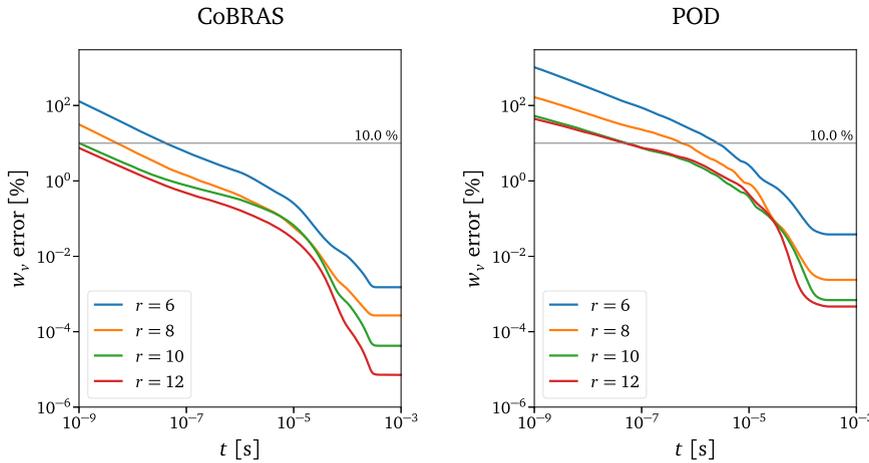


Fig. 9. Mean relative error in the distribution function for VC model using CoBRAS and POD ROMs. Time evolution of the mean relative error in the O_2 energy level population across 1 000 testing trajectories. Results are shown for different dimensions r of the CoBRAS (left panel) and POD (right panel) ROM systems.

5.2.4. Computational performance

Following the same methodology as in section 5.1.4, we assess here the computational performance of the CoBRAS-based ROM in comparison to the FOM for the VC model.

As shown in Table 6, the ROM reduces the total number of FLOPs by more than two orders of magnitude for RHS evaluations, which primarily involve dense 3D tensor double contractions with $\mathcal{O}(d^3)$ complexity (see equations (S.1.2)–(S.1.3) in the Supplementary Material and Appendix C.3), and by a similar amount for the linear system solve (LSS) required by the implicit integration scheme. In

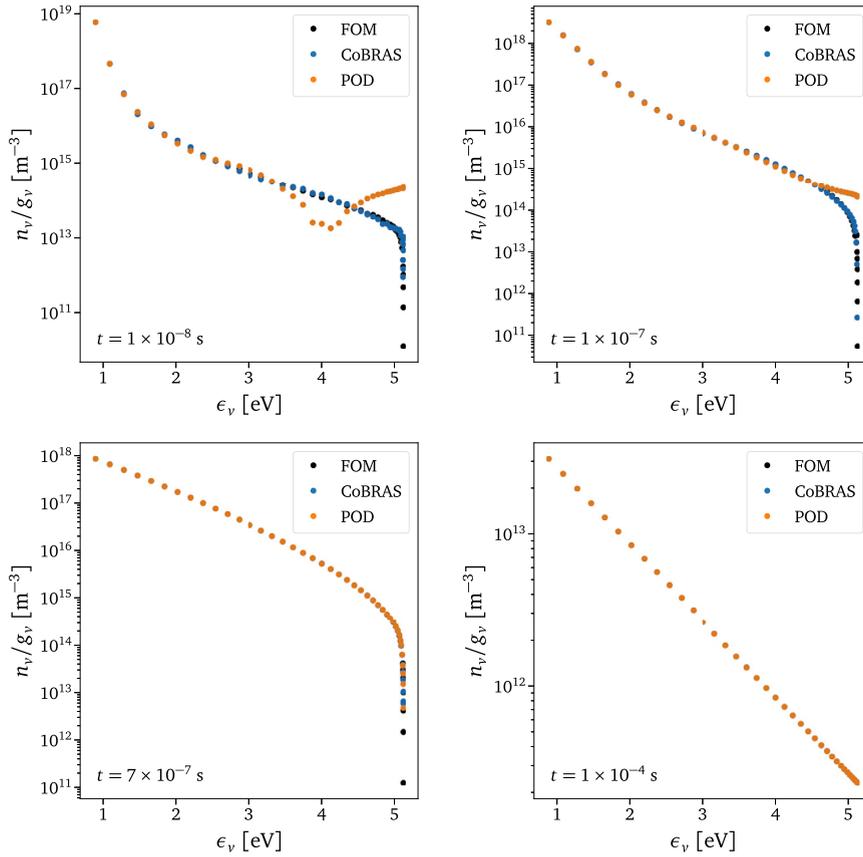


Fig. 10. FOM vs. ROMs for VC model: distribution snapshots. Time snapshots of the O_2 energy level population in an initially cold gas with $T = 10000$ K, calculated using FOM, CoBRAS, and POD reduced-order models. All ROMs have a dimension of $r = 8$.

Table 6

Computational cost comparison: FOM vs. CoBRAS ROM for the VC model. The table reports total FLOPs and FLOP-per-byte ratios for key numerical operations, including right-hand side (RHS) evaluation and linear system solve (LSS), as well as estimates of system stiffness. Results are shown for the FOM and various dimensions of the CoBRAS model.

Model	d	FLOPs		FLOP/Byte		Fastest Timescale
		RHS - $\mathcal{O}(d^3)$	LSS - $\mathcal{O}(d^3)$	RHS - $\mathcal{O}(1)$	LSS - $\mathcal{O}(d)$	$1/ \lambda_{\min} $
FOM	46	2.920×10^5	6.912×10^4	0.375	3.913	1.43×10^{-9}
CoBRAS	9	2.178×10^3	6.480×10^2	0.364	0.818	5.39×10^{-9}
	5	3.700×10^2	1.333×10^2	0.343	0.476	2.27×10^{-8}

addition, the ROM yields a significantly less stiff system, allowing for integration time steps up to ten times larger than those required by the FOM.

As observed previously in Table 3, the efficiency of RHS and LSS operations is ultimately limited by their *arithmetic intensity*. In this case as well, both operations exhibit even lower FLOP/Byte ratios in the ROM setting, indicating that actual runtime gains from FLOP reductions may be limited by memory bandwidth and data transfer.

6. Conclusions

This work presents a data-driven Petrov-Galerkin reduced-order model (ROM) for detailed state-to-state kinetics in thermochemical nonequilibrium, avoiding the conventional assumptions often required in traditional models. In contrast to standard ROMs such as multi-temperature (MT) or coarse-grained (CG) models, which rely on empirical or physical assumptions, we sought a more flexible, assumption-free, and computationally efficient approach.

To leverage the low-rank dynamics often observed in kinetic systems under thermal nonequilibrium, we used the recently developed ‘‘Covariance Balancing Reduction using Adjoint Snapshots’’ (CoBRAS) [53] technique as the foundation of our model reduction pipeline. This formulation allows for the computation of a linear projection operator for Petrov-Galerkin modeling by balancing the

state and gradient covariances associated with the full-order system (the state-to-state kinetic equations in our specific case). Our analysis revealed that these systems generally do not exhibit significant nonlinear behavior, allowing us to apply linearized equations around thermochemical equilibrium during the data-collection phase of CoBRAS. Experimental validation of our test cases confirmed that the linearized equations provided a sufficiently accurate description of the subspace on which the nonlinear system evolves. While the linearization facilitated rapid data generation with minimal information loss, we acknowledge that we could also utilize the fully nonlinear system. However, this would lead to a much more computationally expensive training phase, with the computational cost seeing increases of two to four orders of magnitude. The approach was applied to two distinct thermochemical systems: a pure oxygen rovibrational collisional (RVC) model and a vibrational collisional (VC) model. In both cases, CoBRAS computed models with excellent predictive accuracy. In particular, these models exhibited relative errors of less than 1% for macroscopic quantities (i.e., moments) and under 10% for microscopic quantities (i.e., energy level population) along with excellent compression rates. Additionally, our models significantly outperformed existing MT and CG models, as well as the Galerkin model based on Proper Orthogonal Decomposition (POD).

The proposed data-driven approach offers several advantages over traditional models, such as MT or CG. By relying solely on the structure of the underlying state-to-state governing equations, the proposed method eliminates the need for empirical or physical assumptions, resulting in enhanced accuracy and flexibility across various nonequilibrium conditions. Furthermore, it bypasses the time-consuming system analysis or optimization typically required to determine the optimal reduced representation, enabling the generation of the reduced model within a few minutes. Additionally, the proposed projection-based model avoids the computational costs and potential errors associated with interpolation needed for nonlinear models such as CG. For these reasons, we believe that this work lays the foundation for more efficient computational tools that preserve the fidelity of detailed state-to-state kinetics while making them suitable for real-world applications in fields such as hypersonic flight, plasma dynamics, and combustion.

While the proposed data-driven ROM shows great potential, further research is required to improve its performance. A significant limitation is that the current model does not preserve the positivity of the distribution function by design, which could result in negative values during the initial stages of the dynamics (e.g., $t < 10^{-8}$ s). Addressing this issue and testing the model in adiabatic systems will be priorities in future research to ensure the model's applicability in real-world computational fluid dynamics applications.

CRedit authorship contribution statement

Ivan Zanardi: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Alberto Padovan:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Daniel J. Bodony:** Writing – review & editing, Supervision. **Marco Panesi:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I. Zanardi and M. Panesi acknowledge support from the Vannevar Bush Faculty Fellowship OUSD(RE), under Grant #N00014-21-1-295. A. Padovan was supported by the Center for Hypersonics and Entry Systems Studies through internal funding provided by M. Panesi and K. Stephani via The Grainger College of Engineering at the University of Illinois Urbana-Champaign. The views and conclusions expressed in this work are those of the authors and do not necessarily reflect the official policies or endorsements, either expressed or implied, of the United States Government.

Appendix A. Proof of Proposition 1

Proof. Given the exact solution

$$\mathbf{w}(t) = \mathbf{w}(t_0) + \int_{t_0}^t \mathbf{f}(\mathbf{w}(\tau); \theta) d\tau, \quad (\text{A.1})$$

to the ordinary differential equation (6), it follows that a small perturbation $\delta\mathbf{w}(t_0)$ induces a corresponding change $\delta\mathbf{w}(t)$ according to

$$\delta\mathbf{w}(t) = \delta\mathbf{w}(t_0) + \int_{t_0}^t D\mathbf{f}(\mathbf{w}(\tau); \theta)\delta\mathbf{w}(\tau) d\tau := \mathcal{L}(t, t_0)\delta\mathbf{w}(t_0). \quad (\text{A.2})$$

Here, $\mathcal{L}(t, t_0) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is the state transition matrix, which satisfies the linear time-varying dynamics

$$\frac{d\mathcal{L}(t, t_0)}{dt} = D\mathbf{f}(\mathbf{w}(t); \theta)\mathcal{L}(t, t_0). \quad (\text{A.3})$$

Assuming an Euclidean inner product on \mathbb{R}^N , the adjoint $\mathcal{L}^\top(t_0, t)$ satisfies the adjoint equation

$$-\frac{d\mathcal{L}^\top(t_0, t)}{dt} = D\mathbf{f}(\mathbf{w}(t); \boldsymbol{\theta})^\top \mathcal{L}^\top(t_0, t). \quad (\text{A.4})$$

Now, recalling that $y_j(t) = \mathbf{e}_j^\top \mathbf{C}\mathbf{w}(t)$, where \mathbf{e}_j is the j -th unit vector in the standard basis of $\mathbb{R}^{\dim(\mathbf{y})}$, we have

$$\delta y_j(t_k) = \mathbf{e}_j^\top \mathbf{C} \delta \mathbf{w}(t_k) = \mathbf{e}_j^\top \mathbf{C} \mathcal{L}(t_k, t_0) \delta \mathbf{w}(t_0) := \langle \mathbf{g}_j^{(k)}, \delta \mathbf{w}(t_0) \rangle \quad (\text{A.5})$$

where $\mathbf{g}_j^{(k)}$ is the gradient of $y_j(t_k)$ with respect to $\mathbf{w}(t_0)$, and the last equality follows from the definition of the gradient. Then, it follows immediately that

$$\mathbf{g}_j^{(k)} = \mathcal{L}^\top(t_0, t_k) \mathbf{C}^\top \mathbf{e}_j. \quad (\text{A.6})$$

Using (A.4) we recover (17) and this concludes the proof. \square

Appendix B. Covariance matrices

The probability density functions used to compute the covariance matrices in equations (15) and (16) are defined as follows

$$f_{T_0}(t_0) = \frac{1}{b_{t_0} - a_{t_0}}, \quad (\text{B.1})$$

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(T, \rho) = \frac{1}{b_T - a_T} \frac{1}{b_\rho - a_\rho}, \quad (\text{B.2})$$

$$f_{\mathbf{M}}(\boldsymbol{\mu}) = f_{\mathbf{M}}(w_{O_0}, T_0) = \frac{1}{b_{w_{O_0}} - a_{w_{O_0}}} \frac{1}{T_0 \left[\ln(b_{T_0}) - \ln(a_{T_0}) \right]}. \quad (\text{B.3})$$

Here, $a_{(\cdot)}$ and $b_{(\cdot)}$ represent the distribution bounds provided in Tables 1 and 4 for $f_{\boldsymbol{\theta}}$ and $f_{\mathbf{M}}$, while for f_{T_0} , the bounds are $a_{t_0} = 0$ s and $b_{t_0} = 10^{-2}$ s.

Appendix C. Arithmetic intensity

The *arithmetic intensity* of a computational kernel is defined as the ratio between the number of floating-point operations (FLOPs) and the amount of data transferred to or from memory (in bytes) during the computation. It quantifies how much computation is performed per unit of data movement and is a key metric for understanding whether a given operation is *compute-bound* (limited by processing speed) or *memory-bound* (limited by memory bandwidth).

C.1. Linear system solve

We estimate the *arithmetic intensity* of solving a dense linear system of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (\text{C.1})$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$, and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^d$.

FLOP count For a dense system solved using LU decomposition (no pivoting), the computational cost consists of:

- LU factorization: $\frac{2}{3}d^3$ FLOPs,
- Forward and backward substitution: approximately $2d^2$ FLOPs.

Therefore, we approximate the total FLOPs as:

$$\text{FLOPs} \approx \frac{2}{3}d^3 + 2d^2. \quad (\text{C.2})$$

Memory accesses Assuming double-precision floating-point numbers (8 bytes per value), the memory traffic includes:

- Reading the matrix \mathbf{A} : $8d^2$ bytes,
- Reading the right-hand side vector \mathbf{b} : $8d$ bytes,
- Writing the solution vector \mathbf{x} : $8d$ bytes.

Thus, the total memory access is approximately:

$$\text{Bytes} \approx 8(d^2 + 2d). \quad (\text{C.3})$$

Arithmetic intensity The resulting *arithmetic intensity* is:

$$\text{FLOP/Byte} \approx \frac{d}{12} \frac{d+3}{d+2}. \quad (\text{C.4})$$

This expression shows that *arithmetic intensity* increases linearly with the problem size d .

C.2. Matrix-vector product

We estimate the *arithmetic intensity* of a dense matrix-vector product of the form:

$$\mathbf{y} = \mathbf{A}\mathbf{b}, \quad (\text{C.5})$$

where $\mathbf{A} \in \mathbb{R}^{d \times d}$, and $\mathbf{b}, \mathbf{y} \in \mathbb{R}^d$.

FLOP count In a dense matrix-vector multiplication, each output element involves d multiplications and $d - 1$ additions, yielding:

$$\text{FLOPs} \approx d(2d - 1). \quad (\text{C.6})$$

Memory accesses The memory access pattern is identical to that described in equation (C.3).

Arithmetic intensity The resulting *arithmetic intensity* is:

$$\text{FLOP/Byte} \approx \frac{1}{8} \frac{2d - 1}{d + 2}. \quad (\text{C.7})$$

For large d , the intensity approaches 1/4, indicating that the matrix-vector product is limited by memory access rather than computation.

C.3. Tensor double contraction

We estimate the *arithmetic intensity* of a double contraction between a 3D tensor and two vectors:

$$\mathbf{y} = \mathbf{A} : \mathbf{b}\mathbf{b}^\top, \quad (\text{C.8})$$

where $\mathbf{A} \in \mathbb{R}^{d \times d \times d}$, and $\mathbf{b}, \mathbf{y} \in \mathbb{R}^d$.

FLOP count Each output element involves $2d^2$ multiplications and $d^2 - 1$ additions, for a total of:

$$\text{FLOPs} \approx d(3d^2 - 1). \quad (\text{C.9})$$

Memory accesses The total memory access is approximately:

$$\text{Bytes} \approx 8(d^3 + 2d). \quad (\text{C.10})$$

Arithmetic intensity The *arithmetic intensity* is then:

$$\text{FLOP/Byte} \approx \frac{1}{8} \frac{3d^2 - 1}{d^2 + 2}. \quad (\text{C.11})$$

For large d , this expression approaches 3/8, which is higher than that of a matrix-vector product but still relatively low. This indicates that the operation remains memory-bound, with performance primarily constrained by memory bandwidth rather than computational throughput.

Appendix D. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jcp.2025.113999>.

Data availability

The code used in this work is available online at <https://github.com/ivanZanardi/ronek>.

References

- [1] W.G. Vincenti, C.H. Kruger, *Introduction to Physical Gas Dynamics*, vol. 1, Wiley, New York, 1965.
- [2] C. Park, *Nonequilibrium Hypersonic Aerothermodynamics*, Wiley, New York, 1990.
- [3] I.D. Boyd, T.E. Schwartzentruber, *Nonequilibrium Gas Dynamics and Molecular Simulation*, Cambridge University Press, 2017, <https://experts.umn.edu/en/publications/nonequilibrium-gas-dynamics-and-molecular-simulation>.

- [4] E.H. Hirschel, C. Weiland, Selected Aerothermodynamic Design Problems of Hypersonic Flight Vehicles, Springer Berlin Heidelberg, 2009, <https://doi.org/10.1007/978-3-540-89974-7>.
- [5] A. Harpale, M. Panesi, H.B. Chew, Communication: surface-to-bulk diffusion of isolated versus interacting C atoms in Ni(111) and Cu(111) substrates: a first principle investigation, *J. Chem. Phys.* 142 (2015) 061101, <https://doi.org/10.1063/1.4907716>.
- [6] A. Harpale, M. Panesi, H.B. Chew, Plasma-graphene interaction and its effects on nanoscale patterning, *Phys. Rev. B* 93 (2016) 035416, <https://doi.org/10.1103/PhysRevB.93.035416>.
- [7] C. Cercignani, *The Boltzmann Equation and Its Applications*, vol. 67, Springer, 1988.
- [8] J.H. Ferziger, H.G. Kaper, *Mathematical Theory of Transport Processes in Gases*, 4, North-Holland Publishing Company, Amsterdam, 1972, <https://doi.org/10.1017/S0022112073210960>.
- [9] M. Capitelli, R. Celiberto, G. Colonna, F. Esposito, C. Gorse, K. Hassouni, A. Laricchiuta, S. Longo, *Fundamental Aspects of Plasma Chemical Physics*, Springer Series on Atomic, Optical, and Plasma Physics, vol. 85, Springer New York, New York, NY, 2016, <http://link.springer.com/10.1007/978-1-4419-8185-1>.
- [10] M. Panesi, R.L. Jaffe, D.W. Schwenke, T.E. Magin, Rovibrational internal energy transfer and dissociation of $N_2(^1\Sigma_g^+)-N(^4S_u)$ system in hypersonic flows, *J. Chem. Phys.* 138 (2013) 044312, <https://doi.org/10.1063/1.4774412>.
- [11] M. Panesi, A. Munafò, T.E. Magin, R.L. Jaffe, Nonequilibrium shock-heated nitrogen flows using a rovibrational state-to-state method, *Phys. Rev. E* 90 (2014) 013009, <https://doi.org/10.1103/PhysRevE.90.013009>.
- [12] E.A. Nagnibeda, E. Kustova, *Non-Equilibrium Reacting Gas Flows, Heat and Mass Transfer*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, <http://link.springer.com/10.1007/978-3-642-01390-4>.
- [13] M. Panesi, T.E. Magin, A. Bourdon, A. Bultel, O. Chazot, Electronic excitation of atoms and molecules for the FIRE II flight experiment, *J. Thermophys. Heat Transf.* 25 (2011) 361–374, <https://doi.org/10.2514/1.50033>.
- [14] R.L. Macdonald, A. Munafò, C.O. Johnston, M. Panesi, Nonequilibrium radiation and dissociation of CO molecules in shock-heated flows, *Phys. Rev. Fluids* 1 (2016) 043401, <https://doi.org/10.1103/PhysRevFluids.1.043401>.
- [15] R.L. Macdonald, E. Torres, T.E. Schwartztruber, M. Panesi, State-to-state master equation and direct molecular simulation study of energy transfer and dissociation for the N_2-N system, *J. Phys. Chem. A* 124 (2020) 6986–7000, <https://doi.org/10.1021/acs.jpca.0c04029>.
- [16] D. Wang, J.R. Stallcop, W.M. Huo, C.E. Dateo, D.W. Schwenke, H. Partridge, Quantal study of the exchange reaction for $N+N_2$ using an ab initio potential energy surface, *J. Chem. Phys.* 118 (2003) 2186–2189, <https://doi.org/10.1063/1.1534092>.
- [17] F. Esposito, I. Armenise, M. Capitelli, $N-N_2$ state to state vibrational-relaxation and dissociation rates based on quasiclassical calculations, *Chem. Phys.* 331 (2006) 1–8, <https://doi.org/10.1016/j.chemphys.2006.09.035>.
- [18] B.R.L. Galvão, A.J.C. Varandas, Accurate double many-body expansion potential energy surface for $N_3(^4A'')$ from correlation scaled ab initio energies with extrapolation to the complete basis set limit, *J. Phys. Chem. A* 113 (2009) 14424–14430, <https://doi.org/10.1021/jp903719h>.
- [19] R.L. Jaffe, D.W. Schwenke, G. Chaban, Theoretical analysis of N_2 collisional dissociation and rotation-vibration energy transfer, in: 47th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2009, <https://arc.aiaa.org/doi/10.2514/6.2009-1569>.
- [20] S. Venturi, R.L. Jaffe, M. Panesi, Bayesian machine learning approach to the quantification of uncertainties on ab initio potential energy surfaces, *J. Phys. Chem. A* 124 (2020) 5129–5146, <https://doi.org/10.1021/acs.jpca.0c02395>.
- [21] M.S. Priyadarshini, S. Venturi, I. Zanardi, M. Panesi, Efficient quasi-classical trajectory calculations by means of neural operator architectures, *Phys. Chem. Chem. Phys.* 25 (2023) 13902–13912, <https://doi.org/10.1039/D2CP05506F>.
- [22] A. Munafò, A. Lani, A. Bultel, M. Panesi, Modeling of non-equilibrium phenomena in expanding flows by means of a collisional-radiative model, *Phys. Plasmas* 20 (2013) 073501, <https://doi.org/10.1063/1.4810787>.
- [23] E. Kustova, M. Mekhonoshina, Models for bulk viscosity in carbon dioxide, *AIP Conf. Proc.* 2132 (2019) 150006, <https://doi.org/10.1063/1.5119646>.
- [24] L.D. Landau, E. Teller, On the theory of sound dispersion, *Phys. Z. Sowjetunion* 10 (1936) 34, <https://doi.org/10.1016/B978-0-08-010586-4.50027-4>.
- [25] P. Hammerling, J.D. Teare, B. Kivel, Theory of radiation from luminous shock waves in nitrogen, *Phys. Fluids* 2 (1959) 422–426, <https://doi.org/10.1063/1.1724413>.
- [26] S.P. Heims, Moment equations for vibrational relaxation coupled with dissociation, *J. Chem. Phys.* 38 (1963) 603–606, <https://doi.org/10.1063/1.1733712>.
- [27] P.V. Marrone, C.E. Treanor, Chemical relaxation with preferential dissociation from excited vibrational levels, *Phys. Fluids* 6 (1963) 1215–1221, <https://doi.org/10.1063/1.1706888>.
- [28] C.E. Treanor, P.V. Marrone, Effect of dissociation on the rate of vibrational relaxation, *Phys. Fluids* 5 (1962) 1022–1026, <https://doi.org/10.1063/1.1724467>.
- [29] C. Park, Assessment of two-temperature kinetic model for ionizing air, *J. Thermophys. Heat Transf.* 3 (1989) 233–244, <https://doi.org/10.2514/3.28771>.
- [30] C. Park, Review of chemical-kinetic problems of future NASA missions. I - Earth entries, *J. Thermophys. Heat Transf.* 7 (1993) 385–398, <https://doi.org/10.2514/3.431>.
- [31] C. Park, J.T. Howe, R.L. Jaffe, G.V. Candler, Review of chemical-kinetic problems of future NASA missions. II - Mars entries, *J. Thermophys. Heat Transf.* 8 (1994) 9–23, <https://doi.org/10.2514/3.496>.
- [32] C. Park, R.L. Jaffe, H. Partridge, Chemical-kinetic parameters of hyperbolic Earth entry, *J. Thermophys. Heat Transf.* 15 (2001) 76–90, <https://doi.org/10.2514/2.6582>.
- [33] K.L. Heritier, R.L. Jaffe, V. Laporta, M. Panesi, Energy transfer models in nitrogen plasmas: analysis of $N_2(^1\Sigma_g^+)-N(^4S_u)-e^-$ interaction, *J. Chem. Phys.* 141 (2014) 184302, <https://doi.org/10.1063/1.4900508>.
- [34] T.E. Magin, M. Panesi, A. Bourdon, R.L. Jaffe, D.W. Schwenke, Coarse-grain model for internal energy excitation and dissociation of molecular nitrogen, *Chem. Phys.* 398 (2012) 90–95, <https://doi.org/10.1016/J.CHEMPHYS.2011.10.009>.
- [35] M. Panesi, A. Lani, Collisional radiative coarse-grain model for ionization in air, *Phys. Fluids* 25 (2013) 057101, <https://doi.org/10.1063/1.4804388>.
- [36] A. Munafò, M. Panesi, T.E. Magin, Boltzmann rovibrational collisional coarse-grained model for internal energy excitation and dissociation in hypersonic flows, *Phys. Rev. E* 89 (2014) 023001, <https://doi.org/10.1103/PhysRevE.89.023001>.
- [37] A. Munafò, Y. Liu, M. Panesi, Modeling of dissociation and energy transfer in shock-heated nitrogen flows, *Phys. Fluids* 27 (2015) 127101, <https://doi.org/10.1063/1.4935929>.
- [38] Y. Liu, M. Panesi, A. Sahai, M. Vinokur, General multi-group macroscopic modeling for thermo-chemical non-equilibrium gas mixtures, *J. Chem. Phys.* 142 (2015) 134109, <https://doi.org/10.1063/1.4915926>.
- [39] R.L. Macdonald, R.L. Jaffe, D.W. Schwenke, M. Panesi, Construction of a coarse-grain quasi-classical trajectory method. I. Theory and application to N_2-N_2 system, *J. Chem. Phys.* 148 (2018) 054309, <https://doi.org/10.1063/1.5011331>.
- [40] R.L. Macdonald, M.S. Grover, T.E. Schwartztruber, M. Panesi, Construction of a coarse-grain quasi-classical trajectory method. II. Comparison against the direct molecular simulation method, *J. Chem. Phys.* 148 (2018) 054310, <https://doi.org/10.1063/1.5011332>.
- [41] A. Sahai, B. Lopez, C.O. Johnston, M. Panesi, Adaptive coarse graining method for energy transfer and dissociation kinetics of polyatomic species, *J. Chem. Phys.* 147 (2017) 054107, <https://doi.org/10.1063/1.4996654>.
- [42] S. Venturi, M. Sharma Priyadarshini, B. Lopez, M. Panesi, Data-inspired and physics-driven model reduction for dissociation: application to the O_2+O system, *J. Phys. Chem. A* 124 (2020) 8359–8372, <https://doi.org/10.1021/acs.jpca.0c04516>.
- [43] M. Sharma Priyadarshini, Y. Liu, M. Panesi, Coarse-grained modeling of thermochemical nonequilibrium using the multigroup maximum entropy quadratic formulation, *Phys. Rev. E* 101 (2020) 013307, <https://doi.org/10.1103/PhysRevE.101.013307>.

- [44] I. Zanardi, S. Venturi, M. Panesi, Adaptive physics-informed neural operator for coarse-grained non-equilibrium flows, *Sci. Rep.* 13 (2023) 1–22, <https://doi.org/10.1038/s41598-023-41039-y>.
- [45] C. Jacobsen, I. Zanardi, S. Bholra, K. Duraisamy, M. Panesi, Information theoretic clustering for coarse-grained modeling of non-equilibrium gas dynamics, *J. Comput. Phys.* 507 (2024) 112977, <https://doi.org/10.1016/j.jcp.2024.112977>.
- [46] C.W. Rowley, T. Colonius, R.M. Murray, Model reduction for compressible flows using POD and Galerkin projection, *Physica D* 189 (2004) 115–129, <https://doi.org/10.1016/j.physd.2003.03.001>.
- [47] M.F. Barone, I. Kalashnikova, D.J. Segalman, H.K. Thornquist, Stable Galerkin reduced order models for linearized compressible flow, *J. Comput. Phys.* 228 (2009) 1932–1946, <https://doi.org/10.1016/j.jcp.2008.11.015>.
- [48] J.C. Sutherland, A. Parente, Combustion modeling using principal component analysis, *Proc. Combust. Inst.* 32 (2009) 1563–1570, <https://doi.org/10.1016/j.proci.2008.06.147>.
- [49] A. Parente, J.C. Sutherland, Principal component analysis of turbulent combustion data: data pre-processing and manifold sensitivity, *Combust. Flame* 160 (2013) 340–350, <https://doi.org/10.1016/j.combustflame.2012.09.016>.
- [50] A. Bellemans, A. Munafò, T.E. Magin, G. Degrez, A. Parente, Reduction of a collisional-radiative mechanism for argon plasma based on principal component analysis, *Phys. Plasmas* 22 (2015), <https://doi.org/10.1063/1.4922077>.
- [51] C.W. Rowley, S.T. Dawson, Model reduction for flow analysis and control, *Annu. Rev. Fluid Mech.* 49 (2017) 387–417, <https://doi.org/10.1146/annurev-fluid-010816-060042>.
- [52] S.E. Otto, A. Padovan, C.W. Rowley, Optimizing oblique projections for nonlinear systems using trajectories, *SIAM J. Sci. Comput.* 44 (2022) A1681–A1702, <https://doi.org/10.1137/21M1425815>.
- [53] S.E. Otto, A. Padovan, C.W. Rowley, Model reduction for nonlinear systems by balanced truncation of state and gradient covariance, *SIAM J. Sci. Comput.* 45 (2023) A2325–A2355, <https://doi.org/10.1137/22M1513228>.
- [54] A. Padovan, B. Vollmer, D.J. Bodony, Data-driven model reduction via non-intrusive optimization of projection operators and reduced-order dynamics, *SIAM J. Appl. Dyn. Syst.* 23 (2024) 3052–3076, <https://doi.org/10.1137/24M1628414>.
- [55] B.C. Moore, Principal component analysis in linear systems: controllability, observability, and model reduction, *IEEE Trans. Autom. Control* 26 (1981) 17–32, <https://doi.org/10.1109/TAC.1981.1102568>.
- [56] G.E. Dullerud, F. Paganini, *A Course in Robust Control Theory*, Texts in Applied Mathematics, vol. 36, Springer New York, New York, NY, 2000, <http://link.springer.com/10.1007/978-1-4757-3290-0>.
- [57] C.W. Rowley, Model reduction for fluids using balanced proper orthogonal decomposition, *Int. J. Bifurc. Chaos* 15 (2005) 997–1013, <https://doi.org/10.1142/S0218127405012429>.
- [58] A. Varga, Balanced truncation model reduction of periodic systems, *Proc. IEEE Conf. Decis. Control* 3 (2000) 2379–2384, <https://doi.org/10.1109/CDC.2000.914155>.
- [59] A. Padovan, C.W. Rowley, Continuous-time balanced truncation for time-periodic fluid flows using frequential Gramians, *J. Comput. Phys.* 496 (2024) 112597, <https://doi.org/10.1016/j.jcp.2023.112597>.
- [60] R.L. Jaffe, D.W. Schwenke, M. Panesi, First principles calculation of heavy particle rate coefficients, in: *Hypersonic Nonequilibrium Flows: Fundamentals and Recent Advances*, American Institute of Aeronautics and Astronautics, Inc., Reston, VA, 2015, pp. 103–158, <https://arc.aiaa.org/doi/10.2514/5.9781624103292.0103.0158>.
- [61] O. Zahm, P.G. Constantine, C. Prieur, Y.M. Marzouk, Gradient-based dimension reduction of multivariate vector-valued functions, *SIAM J. Sci. Comput.* 42 (2020) A534–A558, <https://doi.org/10.1137/18M1221837>.
- [62] A. Munafò, M. Panesi, R.L. Jaffe, G. Colonna, A. Bourdon, T.E. Magin, QCT-based vibrational collisional models applied to nonequilibrium nozzle flows, *Eur. Phys. J. D* 66 (2012) 188, <https://doi.org/10.1140/epjd/e2012-30079-3>.
- [63] A. Munafò, M. Panesi, Plato: a high-fidelity tool for multi-component plasmas, in: *AIAA SCITECH 2023 Forum*, American Institute of Aeronautics and Astronautics (AIAA), 2023, <https://arc.aiaa.org/doi/10.2514/6.2023-3490>.