

SPIN: A Data-Driven Model to Reduce Large Chemical Reaction Networks

Mayank Baranwal^{a,*}, Jacob C. Saldinger^{b,*}, Doohyun Kim^c, Paolo Elvati^d, Alfred O. Hero^{e,**}, Angela Violi^{b,d,e,f,**}

^aDepartment of Systems & Control Engineering, Indian Institute of Technology, Bombay, 400076 India

^bDepartment of Chemical Engineering, University of Michigan, Ann Arbor, MI, 48109 USA

^cDepartment of Mechanical Engineering, University of Michigan-Dearborn, Dearborn, MI, 48128

^dDepartment of Mechanical Engineering, University of Michigan, Ann Arbor, MI, 48109 USA

^eDepartment of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48109 USA

^fBiophysics Program, University of Michigan, Ann Arbor, MI, 48109 USA

Abstract

The behavior and chemical reactivity of a group of molecules are typically illustrated using intricate and extensive chemical reaction networks. These networks consist of a group of species and a series of reactions that detail their evolution. Although recent years have seen a surge in efforts to create numerical algorithms that generate dense chemical reaction networks with thousands of reactions and molecules, the simulation of these networks is a computationally demanding task, even for reaction mechanisms that describe the combustion of hydrocarbons. This paper introduces an innovative and unbiased approach to data-driven model reduction of extensive reaction networks called the SParse IdeNtification (SPIN) algorithm. SPIN combines tools from different domains to identify a set of crucial reactions using species concentrations and reaction rates, all while maintaining minimal computational costs and without requiring extra data or simulations. SPIN is successfully tested for large combustion networks of propane and *n*-heptane. The study demonstrates that, despite containing only one-fifth of the reactions found in the full mechanism, the SPIN reduced mechanism for *n*-heptane combustion serves as a highly accurate approximation of the original mechanism, with an average deviation of only 8.4% in ignition delay. Notably, this outstanding performance is achieved without bias towards any particular target property, such as ignition delay, as the reduction and model parameters are optimized to obtain the best possible results. We demonstrate that SPIN can operate as a standalone method or be hybridized with existing species-based reduction methods to further enhance its ability to identify the most significant reactions. This capability is particularly beneficial in comprehending the intricate mechanisms of combustion.

Keywords: Mechanism Reduction, Sparse Learning, Mixed Integer Optimization, Wiener Filter

1. Introduction

At the heart of energy lies a fundamental principle: its continual transformation. From the radiant glow of stars to the controlled combustion of renewable fuels, energy undergoes intricate exchanges governed by the delicate dance of molecules. This intricate network of possibilities, where molecules collide, split, and rearrange, is what we call reaction mechanisms. Understanding these mechanisms is crucial for unlocking energy secrets, from optimizing fuel production and engine performance to designing sustainable solutions for a cleaner future [1].

The sheer number and intricate dependencies of elementary reactions in fuel combustion, influenced by varying temperature, pressure, and species concentrations, present a formidable obstacle to fully understanding and optimizing these networks. However, deciphering this chemical dance is key to unlocking these energy sources' full potential and efficient utilization. Numerical modeling has proven to be a valuable tool in analyzing

reactive systems. For instance, computational fluid dynamics (CFD) simulations [2] that combine mass, momentum, and energy equations with a chemical mechanism have become essential for designing combustion systems and have facilitated improvements in efficiency and emissions. Despite the potential benefits of these computational techniques, the excessive computational expense caused by the complexity of chemical mechanisms remains a challenge. While the chemical mechanism with more species and reactions is generally better able to capture kinetic behaviors across a wider range of conditions, it comes at the cost of increased computational burden. As a result, the detailed version of chemical mechanisms for large hydrocarbons relevant to real transportation fuels is practically not applicable to CFD simulations of practical combustion systems. Numerous algorithms have been proposed to address this issue, creating skeletal or reduced mechanisms from a parent detailed mechanism.

These algorithms aim to replicate the desired properties of the reaction network while allowing for a tolerable level of deviation, using only the most significant species and reactions [3]. Previous methods, such as quasi-steady state analysis [4] and sensitivity analysis [5, 6], have been employed to reduce mechanisms. However, these techniques either require prior knowl-

*Equal contribution

**Corresponding authors:

Email addresses: hero@umich.edu (Alfred O. Hero),
avioli@umich.edu (Angela Violi)

edge of the system or are computationally expensive and do not scale well with the size of the system. Alternative reduction techniques, such as the method proposed by Bhattacharjee *et al.* [7], have approached the system as an integer linear programming problem. However, this method requires many reduced models, which can be computationally cumbersome. Direct relational graph (DRG) methods [8] view the reaction network as a graph and evaluate the influence of each species on a set of target species (*e.g.*, major products and reactants). While DRG and related techniques, such as DRG with error propagation [9], have been successful reduction tools, their reliance on prior knowledge of important species poses a challenge for some reaction networks that are not well-understood outside combustion chemistry. Graph theoretic methods based purely on the graph structure have also been applied as a highly efficient reduction tool. However, these methods cannot capture condition-specific phenomena such as negative temperature coefficient chemistry [10]. Recently, a class of data-driven techniques based on machine learning have emerged as tools to model reduced chemistry [11, 12]. These include methods to cluster reactions and species with principal component analyses [13], to optimize reduced mechanisms with genetic algorithms [14, 15], and to represent non-linear chemical relationships using artificial neural networks and a small sub-set of species [16]. Challenges still exist for this class of methods as machine learning methods typically generalize poorly on data that falls outside the training distribution, and not all methods necessarily enforce physically meaningful solutions such as adherence to conservation laws [17].

Our group has recently presented a novel approach called the sparse-learning (SL) [18] to identify important reactions in combustion networks. This approach has a significant advantage over other methods as it is entirely data-driven and requires no prior knowledge of the reaction system. The proposed algorithm is an optimization problem to minimize the number of reactions while ensuring that the concentration deviation remains within a predefined threshold. The discrete form of the chemical rate equation was utilized to solve for the concentrations and identify the error in concentrations. As this technique is applied to larger, more realistic fuels, the non-linear nature and noise in the reaction system become increasingly problematic. The authors of [18] noted that for the reaction mechanisms describing the combustion of *n*-heptane or complex hydrocarbons, a forward-Euler discretization of the continuous-time mass-action kinetics was significantly imprecise in regions with sudden changes in concentration.

This paper presents a new SParse IdeNtification (SPIN) algorithm to identify influential reactions in a combustion network and create a smaller representative network. The SPIN algorithm improves upon existing approaches in three ways. First, it employs a Wiener filter-based system identification technique to obtain an accurate approximation of the discretized mass-action kinetics. The Wiener filter is a linear least-squares filter, ensuring that the identified system is also linear. This is important as it was observed in [18] that a simple discretized mass-action kinetic equation may not accurately approximate reaction kinetics in regions near ignition, with sharp changes

in chemical concentrations. By making appropriate modifications to the discretized mass-action kinetics, the Wiener filter addresses this issue while retaining its linearity. Although our work focuses on demonstrating the benefits of the Wiener filter in our sparse-learning reduction technique, we anticipate that this system identification approach can be applied to other computational strategies, such as multi-scale fluid simulations [19], which require a numerical representation of a complex evolution mechanism. Second, SPIN significantly reduces complex chemical reactions by removing many non-influential reactions. The automatic elimination of non-influential reactions during optimization results from using a Wiener filter. The filter coefficients tend to suppress inactive or moderately active modes in the dynamical system, reducing the number of non-influential reactions. Lastly, we demonstrate that our sparse-learning method can be used in conjunction with directed relational graph (DRG) techniques [9] to achieve an additional level of reduction in skeletal mechanisms for numerical simulations.

Our reduction technique provides a completely data-driven approach to analyzing reaction networks while upholding the underlying chemical relationships. The results demonstrate that we can accurately reconstruct the underlying reaction network for various fuels with different complexities without prior assumptions about the significance of species or reactions. Therefore, our sparse-learning method can effectively reduce a chemical mechanism while maintaining reasonable accuracy. This can be useful in chemical exploration and analysis, especially in cases where expert knowledge is not readily available.

2. The SParse IdeNtification (SPIN) Method

2.1. Preliminaries

The proposed SPIN algorithm leverages tools from multiple domains. Below, we review some useful definitions and properties of Wiener filters, mixed-integer linear programs and their linear programming (LP) relaxations, and fundamental differences between the SPIN algorithm and other network reduction methods.

2.1.1. Mass-action kinetic equations & Wiener filter

Mass-action kinetic equations are often used to model the time evolution of chemical reaction networks using a set of coupled ordinary differential equations (ODEs) [20, 21, 22, 23]. A simple forward-Euler discretization of these coupled ODEs can be expressed as:

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \mathbf{M}\mathbf{r}_t\Delta_t + \boldsymbol{\omega}_t, \quad \text{for all } t = 1, 2, \dots, T-1, \quad (1)$$

where \mathbf{x}_t is an $N_s \times 1$ vector representing chemical concentrations of all N_s species at any time t . The quantity \mathbf{M} is an $N_s \times N_r$ matrix of stoichiometric coefficients of the associated reactions with reaction rates \mathbf{r}_t , an $N_r \times 1$ vector. Here, Δ_t represents the sampling time used for discretization, while $\boldsymbol{\omega}_t$ captures modeling error and process noise, which is typically assumed to be small. The reaction rate associated with the i^{th} -reaction is obtained as:

$$\mathbf{r}_t(i) = k_i \prod_{j=1}^{N_s} \mathbf{x}_t(j)^{\nu_{ij}},$$

149 where k_i is the rate constant for the i^{th} reaction and ν_{ij} is the
 150 stoichiometric coefficient of the j^{th} species involved in the i^{th}
 151 reaction and zero otherwise. Typical linear discretized models,
 152 for mass-action kinetics tend to ignore the modeling and pro-
 153 cess noise term ω_t under assumptions on small modeling and
 154 discretization errors, as is the case with our earlier work [18].
 155 However, these assumptions are grossly inaccurate if the sam-
 156 pling time Δ_t is sufficiently large or in the regions near the igni-
 157 tion that are characterized by sudden changes in concentrations,
 158 of species. Consequently, it was observed that model-based re-
 159 duction methods that consider (1) as the underlying discretized,
 160 model do not scale well for more complex mechanisms. In this,
 161 paper, we alleviate this drawback by considering generalized,
 162 mass-action kinetic equations expressed as:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{M}\mathbf{r}_t\Delta_t + \tilde{\omega}_t, \text{ for all } t = 1, 2, \dots, T-1, \quad (2)$$

163 where \mathbf{A}, \mathbf{B} are $N_s \times N_s$ matrices obtained using *Wiener fil-*
 164 *ters*. Wiener theory was formulated by Norbert Wiener [24],
 165 and it forms the foundation of data-dependent system identifi-
 166 cation using linear least square error filters. A finite-duration
 167 impulse response (FIR) Wiener filter uses multiple linear re-
 168 gression to predict the output from multiple time bins of the
 169 input. In particular, for a block of N samples of the input
 170 signal $[\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}]$ and the desired output signal
 171 $[\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_{N-1}]$, the corresponding Wiener filter input-output
 172 relationship is described as:

$$\begin{pmatrix} \hat{\mathbf{y}}_0 \\ \hat{\mathbf{y}}_1 \\ \vdots \\ \hat{\mathbf{y}}_{N-1} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{u}}_0 & \hat{\mathbf{u}}_{-1} & \cdots & \hat{\mathbf{u}}_{1-P} \\ \hat{\mathbf{u}}_1 & \hat{\mathbf{u}}_0 & \cdots & \hat{\mathbf{u}}_{2-P} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{u}}_{N-1} & \hat{\mathbf{u}}_{N-2} & \cdots & \hat{\mathbf{u}}_{N-P} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_{P-1} \end{pmatrix}, \quad (3)$$

173 which can be compactly written as

$$\hat{\mathbf{Y}} = \mathbf{U}\mathbf{w}.$$

174 The Wiener filter error is the difference between the desired
 175 signal \mathbf{Y} and the filter output $\hat{\mathbf{Y}}$ given by

$$\begin{aligned} \mathbf{E} &= \mathbf{Y} - \hat{\mathbf{Y}}, \\ &= \mathbf{Y} - \mathbf{U}\mathbf{w}. \end{aligned}$$

176 The Wiener filter minimizes the energy or the sum of the
 177 squared elements of the error matrix \mathbf{E} . The coefficients \mathbf{w} are
 178 obtained by setting the gradient of the energy function to zero,
 179 resulting in

$$\mathbf{w} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{Y}, \quad (5)$$

180 where \top denotes the matrix-transpose. In the context of the
 181 proposed modification of the discretized mass-action kinetics
 182 described in (2), the Wiener filter coefficients (\mathbf{A}, \mathbf{B}) are ob-
 183 tained by considering the equivalent block-diagonal formula-
 184 tion as shown in (3), with $P \leftarrow 1$, $\mathbf{y}_t \leftarrow \mathbf{x}_{t+1}$, and the two input
 185 signals \mathbf{x}_t and $\mathbf{M}\mathbf{r}_t\Delta_t$.

2.1.2. Integer programs and LP relaxation

The primary focus of this manuscript is the sparse identifica-
 tion of chemical reaction networks by finding influential reac-
 tions that preserve the dominant modes of the underlying full
 mechanism. We prescribe a quantification of the mismatch be-
 tween the full reaction mechanism and the reduced mechanism
 with sparsely selected reactions. The aim is to find the sparsest
 set of reactions that keeps the mismatch between the two mech-
 anisms below a user-defined threshold. At each time instant,
 SPIN seeks to select or reject a reaction in the network. The
 integer programming decision variables corresponding to the
 selection or rejection of reactions are binary decision variables,
 where a value of '1' indicates the selection of a reaction and '0'
 otherwise. On the other hand, the reduction error (mismatch)
 is a real variable. Such constrained optimization problems with
 linear objective functions and consisting of binary (integer) de-
 cision variables are referred to as *integer linear program* (ILP)
 in the literature [25].

To this end, we define a selection/weight vector $\mathbf{w}_t =$
 $[w_{1,t}, w_{2,t}, \dots, w_{N_r,t}]^T$ of size $N_r \times 1$ which represents the binary
 weights of all reactions at time t . The quantity \mathbf{w}_t prescribes the
 evolution of reduced mechanism at time t as:

$$\hat{\mathbf{x}}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{M}(\mathbf{r}_t \odot \mathbf{w}_t)\Delta_t, \text{ for all } t = 1, 2, \dots, T-1, \quad (6)$$

where \odot denotes the element-wise product. Thus, only reac-
 tions that are selected at any instant prescribe the evolution
 of the reduced (approximate) mechanism. Correspondingly,
 we quantify the mismatch between the original dynamical sys-
 tem (2) and the reduced dynamical system (6) as:

$$\begin{aligned} \mathcal{E}_{j,t}(\mathbf{w}_t) &= |\mathbf{x}_{t+1}(j) - \hat{\mathbf{x}}_{t+1}(j)|, \\ &= |(\mathbf{B}\mathbf{M})_j\mathbf{r}_t - (\mathbf{B}\mathbf{M})_j(\mathbf{r}_t \odot \mathbf{w}_t)|, \quad \forall t, \forall j \in \{1, 2, \dots, N_s\}, \end{aligned} \quad (7)$$

213 where $\mathcal{E}_{j,t}$ is the error introduced in concentration of the j^{th}
 214 species at time t , and $(\mathbf{B}\mathbf{M})_j$ denotes the j^{th} row of the matrix-
 215 product $\mathbf{B}\mathbf{M}$. We want to constrain this concentration error to
 216 be small. However, it must be noted that a reaction is more
 217 dominant only at specific intervals, and thus, the magnitude of
 218 concentration error varies significantly with time, even if the
 219 selection vector \mathbf{w}_t is kept fixed at all times. Given this limita-
 220 tion, we impose a constraint on the concentration error of each
 221 species at all times through appropriate normalization:

$$\mathcal{E}_{j,t}(\mathbf{w}_t) \leq \epsilon \mathcal{N}_t(j), \quad \forall t, \forall j \in \{1, 2, \dots, N_s\}, \quad (8)$$

$$\mathcal{N}_t(j) = |(\mathbf{B}\mathbf{M})_j\mathbf{r}_t\Delta_t|, \quad \forall t, \forall j \in \{1, 2, \dots, N_s\}, \quad (9)$$

222 where $|\cdot|$ is applied element-wise. The quantity $\mathcal{N}_t(j)$ is the as-
 223 sociated normalization constant that enforces relative change in
 224 concentration constraint, while $\epsilon > 0$ is a tuning parameter that
 225 indicates the acceptable error tolerance for the normalization
 226 constant. For example, a value of 0.10 enforces a maximum of
 10% error.

The relative error tolerance constraint (8) effectively limits
 the *instantaneous* mismatch between the full and the approxi-
 mate systems. However, the constraint is not ineffective in lim-
 iting constant drifts in species concentrations. The effect of not

selecting a reaction at any time instant t is visible at time t and at all subsequent times. However, this effect weakens as time evolves. In this work, we account for limiting error propagation by instead considering error in concentration of the j^{th} species over a time horizon of length H as:

$$\begin{aligned} \tilde{\mathcal{E}}_{j,t}(\mathbf{w}_t, \dots, \mathbf{w}_{t+H-1}) &= |\mathbf{x}_{t+H}(j) - \hat{\mathbf{x}}_{t+H}(j)|, \\ &= \left| \mathbf{x}_{t+H}(j) - \mathbf{A}\mathbf{x}_t(j) - \sum_{k=t}^{t+H-1} (\mathbf{B}\mathbf{M})_j(\mathbf{w}_k \odot \mathbf{r}_k) \Delta_{\tau} \right|. \end{aligned} \quad (10)$$

Thus, the constraint on the error term in (8) is suitably modified using the new definition of mismatch error (10) as:

$$\tilde{\mathcal{E}}_{j,t}(\mathbf{w}_t, \dots, \mathbf{w}_{t+H-1}) \leq \epsilon \sum_{k=t}^{t+H-1} \mathcal{N}_k(j), \quad \forall t, \forall j \in \{1, 2, \dots, N_s\}, \quad (11)$$

where the normalization constant \mathcal{N}_k is defined as before.

Another important objective of the reduction process is to ensure that if a reaction is selected at any time instant t , it must remain selected for all future times since the reduced mechanism would simulate all the selected reactions in the final mechanism. To address this, we impose another constraint on the selection vector \mathbf{w}_t as

$$\mathbf{w}_t \leq \mathbf{w}_{t+1}, \quad \forall t \in \{1, 2, \dots, T-2\}, \quad (12)$$

where T is the total number of discretization time points. Note that the above set of linear inequality constraints ensures that if $\mathbf{w}_t(i) = 1$ for some $i \in \{1, \dots, N_r\}$, then $\mathbf{w}_{t'}(i) = 1$ for all $t' \in \{t+1, t+2, \dots, T-1\}$.

Finally, we add another set of constraints to ensure that the reduction process does not eliminate many reactions from the complete mechanism for scenarios where significant concentration changes are not observed at sampled time instants. We address this constraint using the following linear inequality:

$$\sum_{i=1}^{N_r} \mathbf{w}_{T-1}(i) \geq \beta N_r, \quad (13)$$

where $0 \leq \beta \leq 1$ is a user-specified parameter. For instance, $\beta = 0.2$ indicates that the reduced mechanism must include at least 20% reactions in the final (reduced) set. Thus, the resulting ILP for each horizon window of length H can be described as:

$$\text{minimize}_{\{\mathbf{w}_k\}_{k=t}^{t+H-1}} \sum_{k=t}^{t+H-1} \sum_{i=1}^{N_r} \mathbf{w}_k(i) \quad (\text{P}_{\text{ILP}})$$

$$\text{s.t. } \mathbf{w}_k(i) \in \{0, 1\}, \quad \forall k \in \{t, \dots, t+H-1\}, \forall i \in \{1, \dots, N_r\} \quad (\text{P}_{\text{ILP-1}})$$

$$\tilde{\mathcal{E}}_{j,t} \leq \epsilon \sum_{k=t}^{t+H-1} \mathcal{N}_k(j), \quad \forall j \in \{1, 2, \dots, N_s\} \quad (\text{P}_{\text{ILP-2}})$$

$$\mathbf{w}_k \leq \mathbf{w}_{k+1}, \quad \forall k \in \{t-1, t, \dots, t+H-2\} \quad (\text{P}_{\text{ILP-3}})$$

$$\sum_{i=1}^{N_r} \mathbf{w}_{T-1}(i) \geq \beta N_r, \quad \text{if } t \geq T-H. \quad (\text{P}_{\text{ILP-4}})$$

The primary objective of the proposed ILP is to minimize the total number of reactions in the reduced mechanism while ensuring that the corresponding constraints (P_{ILP-1})-(P_{ILP-4}) are met. The binary (or integer) aspect of decision variables makes solving an ILP combinatorially complex. For an ILP, the worst-case computational complexity is exponential in the number of decision variables. This makes the proposed ILP for mechanism reduction computationally prohibitive to handle. However, the solution to the above ILP can be closely approximated using *linear programming (LP) relaxation* [26]. Here, we replace the integer constraint (P_{ILP-2}) with its convex relaxation $0 \leq \mathbf{w}_k(i) \leq 1$. LP relaxations to ILPs are computationally tractable, and the worst-case time complexity is still polynomial in the number of decision variables. The exact formulation of the relaxed LP and recovering the original weight selection vector \mathbf{w}_t is described in detail in Sec. 2.2.

2.1.3. Species vs reactions selection

The proposed SPIN algorithm for mechanism reduction works primarily by finding influential reactions in the mechanism. If all reactions corresponding to a given species are eliminated in the reduction process, that particular species is also removed from the final reduced mechanism. Thus, a reaction-based reduction methodology, such as SPIN, targets the elimination of reactions. This is illustrated in Fig. 1. On the other hand, state-of-the-art species reduction approaches, such as directed relation graph (DRG) [27] or directed relation graph with error propagation (DRGEP) [28] are efficient at reducing the number of species in the skeletal mechanism.

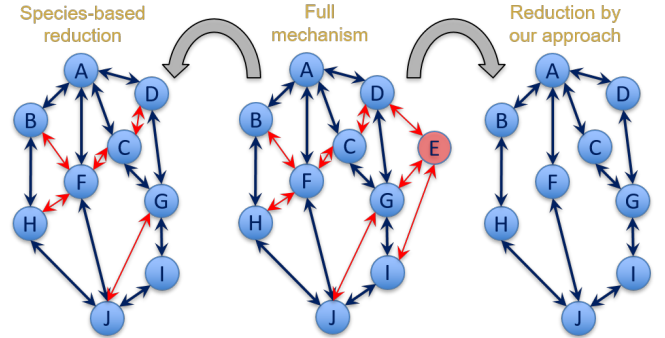


Figure 1: An illustration of working methodology of species-based reduction vs reaction-based reduction. The red arrows indicate weak reaction coupling, while the red node indicates unimportant species. Since species E is connected with the rest of the combustion network through weak coupling, a reaction-based reduction eliminates species E and all other weak reactions. On the other hand, species-based reduction only targets species while retaining all the reactions associated with an important species.

One of the significant advantages of the proposed SPIN algorithm is its lower runtime relative to other reaction reduction methods, which allows hybridization with existing species-based approaches to further reduce the skeletal mechanism to even fewer species and reactions while ensuring a tight approximation to the original full mechanism. Hybridization of SPIN with other algorithms is discussed in detail in Section 3.

2.2. Proposed SPIN algorithm

The SParse IdeNtification (SPIN) algorithm proposed in this manuscript combines tools from multiple domains for principled reduction of the full mechanism. It first leverages a linear least-squared filter (Wiener filter) for data-based identification of sampled linear system that describes the evolution of concentrations of species in a chemical reaction network. The identified generalized mass-action kinetic equations are described in (2). The filter coefficients (\mathbf{A} , \mathbf{B}) are obtained using (5).

The sampled linear system is then subjected to an integral optimization problem over a moving horizon of length H , as shown in (P_{ILP}), to obtain an optimal weight selection vector $\{\mathbf{w}_t\}$ that best approximates the full mechanism up to a user-defined error tolerance. However, integer constraints make the optimization problem computationally prohibitive to handle. Hence, a linear programming (LP) relaxation is suggested, and the following equivalent optimization problem is considered over each horizon of length H :

$$\{\mathbf{w}_k^*\}_{k=t}^{t+H-1} := \arg \min_{\{\mathbf{w}_k\}_{k=t}^{t+H-1}} \sum_{k=t}^{t+H-1} \sum_{i=1}^{N_r} \mathbf{w}_k(i) \quad (\text{P}_{\text{SPIN}})$$

$$\text{s.t. } 0 \leq \mathbf{w}_k(i) \leq 1, \quad \forall k \in \{t, \dots, t+H-2\}, \forall i \in \{1, \dots, N_r\} \quad (\text{P}_{\text{SPIN-1}})$$

$$\tilde{\mathcal{E}}_{j,t} \leq \epsilon \sum_{k=t}^{t+H-1} N_k(j), \quad \forall j \in \{1, 2, \dots, N_s\} \quad (\text{P}_{\text{SPIN-2}})$$

$$\mathbf{w}_k \leq \mathbf{w}_{k+1}, \quad \forall k \in \{t-1, t, \dots, t+H-2\} \quad (\text{P}_{\text{SPIN-3}})$$

$$\sum_{i=1}^{N_r} \mathbf{w}_{T-1}(i) \geq \beta N_r, \quad \text{if } t \geq T-H. \quad (\text{P}_{\text{SPIN-4}})$$

The optimal solution $\{\mathbf{w}_k^*\}_{k=t}^{t+H-1}$ to (P_{SPIN}) is not necessarily binary. However, including or excluding a reaction in the reduced mechanism is a binary decision. Consequently, we perform LP rounding of the optimal solution vector based on a pre-specified threshold $0 \leq \alpha < 1$ as:

$$\tilde{\mathbf{w}}_t^*(i) = \begin{cases} 0, & \text{if } \mathbf{w}_t^*(i) \leq \alpha \\ 1, & \text{if } \mathbf{w}_t^*(i) > \alpha \end{cases} \quad (16)$$

Algorithm 1 describes the proposed SPIN methodology for reducing the full mechanism for specified reaction conditions, namely temperature \tilde{T} , equivalence ratio ϕ , and pressure \tilde{P} (*i.e.*, $\theta = [\tilde{T}; \phi; \tilde{P}]$). The computational complexity of SPIN and the choice of various hyperparameters is discussed in the next section. The output of the algorithm is a binary weight selection matrix θ of size $N_r \times T-1$, with columns indicating the selected reactions at each time instance under specific initial conditions. The algorithm is executed once for each reaction condition θ , and thus, the final reduced mechanism consists of the union-sum of selected reactions at each condition and the species involved in those reactions, *i.e.*,

$$\mathbf{w}_{\text{red}} = \bigcup_{\theta} \mathbf{W}_{T-1}(\theta), \quad (17)$$

where \mathbf{w}_{red} is the vector consisting of all reactions in the reduced mechanism, and $\mathbf{W}_{T-1}(\theta)$ denotes the last column of the matrix $\mathbf{W}(\theta)$.

Algorithm 1 The SParse IdeNtification (SPIN) Algorithm

Input: $\{\mathbf{x}_t\}_{t=1}^T, \{\mathbf{r}_t\}_{t=1}^T, \{\Delta_t\}_{t=1}^T, \epsilon, \alpha, \beta, \mathbf{M}, H, \theta \leftarrow [\tilde{T}; \phi; \tilde{P}]$
Output: Optimal weight selection vector $\mathbf{W}(\theta) := \{\tilde{\mathbf{w}}_t^*(\theta)\}_{t=1}^{T-1}$
Initialization: $t \leftarrow 1; \mathbf{w}_0(i) \leftarrow 0, \quad \forall i \in \{1, \dots, N_r\}$
1: Find Wiener filter coefficients (\mathbf{A} , \mathbf{B}) using (5)
2: **while** $t \leq T-1$ **do**
3: $W \leftarrow \min(H, T-t-1)$
4: Set the time horizon to $\{t, t+1, \dots, t+W\}$
5: Solve the relaxed optimization problem (P_{SPIN}) to obtain $\mathbf{w}_t^*(\theta), t \in \{t, t+1, \dots, t+W-2\}$
6: Binarize $\mathbf{w}_t^*(\theta)$ using (16)
7: Assign $\tilde{\mathbf{w}}_t^*, t \in \{t, t+1, \dots, t+W-1\}$ to columns $\{t, t+1, \dots, t+W-1\}$ of $\mathbf{W}(\theta)$
8: $t \leftarrow t+W$
9: **end while**
10: **return** $\mathbf{W}(\theta)$

2.3. Computational Complexity of SPIN

The original problem (P_{ILP}) is an ILP whose worst-case computational complexity is exponential in the number of integer decision variables. For a horizon of length H , the number of binary decision variables is $N_r H$. While efficient ILP solvers that involve heuristics, such as branch and bound [29], and branch and cut [30], are empirically shown to exhibit polynomial-time convergence, the worst case complexity may still be exponential. For each condition θ with total number of samples T , the worst-case computational complexity of the original ILP (P_{ILP}) is $O(\lceil T/H \rceil 2^{N_r H})$.

The choice of H determines the trade-off between the quality of solutions and computational complexity. Note that the worst-case complexity is exponential in H . Thus, a smaller H is preferred because it involves fewer decision variables. On the other hand, a larger value of H enforces error toleration bound for a longer time. While it may result in a relatively large number of reactions in the reduced mechanism, the effect of excluding a reaction at a given time instant is minimal at future instants. Thus, the resulting reduced mechanism is a tight approximation of the original mechanism. It is observed empirically that a value of $H = 10$ balances this trade-off.

The proposed SPIN algorithm is computationally efficient since it further relaxes the original ILP into an equivalent LP (P_{SPIN}) with convex constraints. Unlike ILPs, LPs are polynomial-time solvable in the number of decision variables. For the LP described in (P_{SPIN}), the worst-case computational complexity for δ -approximation is $O(\lceil \frac{T}{H} \rceil n^\omega \log(n/\delta))$, where $\omega = 2.38$ is the constant of matrix multiplication and n is the number of decision variables. In the context of (P_{SPIN}), the number of decision variables is $2N_r H$ since the convex-relaxation introduces additional $N_r H$ slack variables.

3. Primary Results

3.1. Analysis of Propane Combustion

In this section, we aim to evaluate the performance of SPIN in reducing a propane combustion mechanism and showcase

Table 1: Comparison of different propane mechanisms. Size is described as the number of species N_{spe} and reactions N_{rea} , and for each reduced mechanism, accuracy is measured as the average, ID_{avg} , and maximum, ID_{max} , % error in the ignition delay, and as the mean absolute deviations in peak concentrations for three species. Error is measured relatively to the full mechanism for all the 216 conditions.

	N_{spe}	N_{rea}	ID_{avg} , %	ID_{max} , %	C_2H_4 , %	H_2 , %	CO , %
Full Mechanism	117	1270	-	-	-	-	-
SPIN	111	484	3.3	37.1	4.3	3.1	0.5
DRGEP	80	885	7.2	22.4	2.7	7.0	5.8
DRGEP + SPIN	79	501	7.5	21.4	3.7	6.8	5.7
SPIN + DRGEP	76	342	5.5	25.7	8.4	10.7	6.8

its effectiveness. The propane mechanism adopted in our study [31] consists of 117 species and 1270 reactions. We utilized the time history of species concentration derived from 0-D homogeneous reactor [32] simulations (with constant volume and internal energy) as the input for SPIN. We chose a 0-D homogeneous reactor for our simulation environment since autoignition behavior captured by the 0-D system provides the most influential combustion characteristics governed by oxidation chemistry.

A comprehensive assessment was performed on 216 cases that spanned temperatures from 700 to 1500 K, pressures between 1 and 50 atm, and equivalence ratios from 0.5 to 2. To contrast SPIN with prevalent reduction techniques, we tested the DRGEP method [9], as implemented in CHEMKIN [32]. The thermodynamic conditions (temperature, pressure, equivalence ratio) utilized for the DRGEP execution were identical to those used for SPIN. The selection threshold for DRGEP reduction was the maximum error in ignition delay of no more than 40%, which is roughly equivalent to the maximum error of the SPIN-reduced mechanism. Moreover, we declared significant reactants and products, including C_3H_8 , O_2 , N_2 , CO_2 , and H_2O as species that should not be removed by the DRGEP algorithm. For a consistent comparison, we discarded from all mechanisms any inert gases except for N_2 , specifically He and Ar. The reduction levels for SPIN and DRGEP are presented in Table 1.

Although species reduction is typically prioritized in mechanism reductions [33], DRGEP also includes a method for removing reactions [9]. For consistency in error, we only present the results of DRGEP species reduction.

Additionally, we explore the potential for further reduction by combining SPIN and DRGEP methodologies, which has been successful in previous reduction frameworks [9, 34, 35]. To achieve higher levels of reduction, we implement a two-stage reduction approach, where we first reduce the propane mechanism with SPIN, prioritizing the removal of reactions over species. We then apply DRGEP for species reduction, targeting major products and reactants, CO_2 , O_2 , C_3H_8 , N_2 , and H_2O , with a 40% error tolerance from the reduced SPIN mechanism. Both SPIN + DRGEP and DRGEP + SPIN reduction orders are considered, and the results of these combined reduction methods are presented in Table 1. By utilizing these two-stage reduction frameworks, we can obtain a mechanism with fewer reactions and species than the individual reduction methodologies.

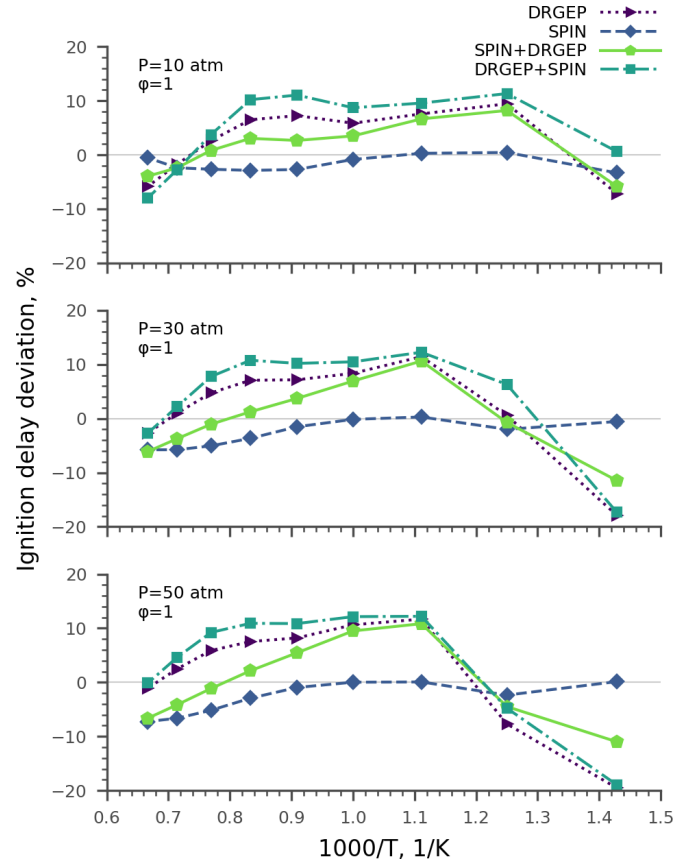


Figure 2: Comparison of ignition delay for propane mechanisms as % deviation from the full mechanism.

To begin with, we examine how accurately the reduced mechanism can replicate the combustion behavior of the full mechanism. Specifically, we calculate the ignition delay, which is defined as the time when the temperature increases by 400 K, for each of the 216 conditions and compare it with that of the full system. The results for the performance of the reduction methodologies in predicting the ignition delay of the full mechanism are presented in Fig. 2. We observe that, for most conditions, the SPIN and SPIN + DRGEP mechanisms have an ignition delay within 10% of the full mechanism. Only in six of the 216 conditions does SPIN have an error of greater than 10%. While the combined SPIN + DRGEP mechanism increases the mean deviation compared to SPIN, its mean error is still lower than that of DRGEP, and there are some conditions where it

427 performs similarly or better than DRGEP and SPIN. This in-452
 428 dicates that it may be useful when a high degree of reduction453
 429 is required. Despite having fewer species and reactions in the454
 430 mechanism, SPIN + DRGEP also outperforms DRGEP + SPIN455
 431 regarding mean absolute deviation. This suggests that remov-456
 432 ing unimportant reactions with SPIN before removing species457
 433 may result in a more effective reduction strategy. Interestingly,458
 434 the mechanism reduced with DRGEP + SPIN has similar er-459
 435 rors to the mechanism reduced with DRGEP with an identi-460
 436 cal number of species and significantly fewer reactions. This461
 437 demonstrates that SPIN can identify additional unimportant re-462
 438 actions that were not identified in the first DRGEP step, result-463
 439 ing in further reduction. All three methodologies, SPIN, SPIN464
 440 + DRGEP, and DRGEP + SPIN, outperform the published per-465
 441 formance of SL [18] in both accuracy and level of reduction,466
 442 highlighting the advantages of this new methodology.

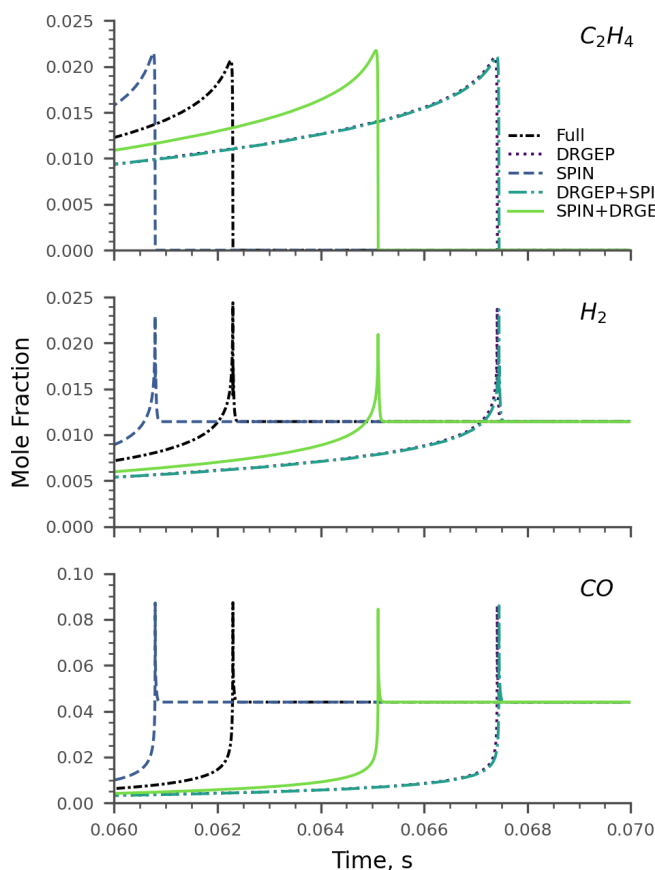


Figure 3: Time evolution of C_2H_4 , H_2 , and CO mole fractions for propane combustion at 1 atm, 1000 K and equivalence ratio of 1.

443 We also evaluate how SPIN reproduces important species500
 444 concentrations. Here, we consider the concentrations of C_2H_4 ,501
 445 H_2 , and CO . We select these intermediates as they have been
 446 considered in previous reduction studies [28] and are impor-502
 447 tant parts of the propane combustion chemistry [36, 37] and the
 448 formation of other chemical byproducts [38]. The concentra-503
 449 tion profiles of these species for a single reaction condition are504
 450 shown in Fig. 3. The concentrations of these molecules in the505
 451 full mechanism increase rapidly and reach a peak concentration506

during ignition. While overall, the reduced mechanisms follow
 this behavior, deviations are observed in the time in which this
 concentration increases, the concentration at the endpoint of the
 simulation, and the magnitude of the increase. The difference
 in time when the increase in concentration occurs can largely be
 characterized by the ignition delay discussed above. When consid-
 ering the endpoint concentration, all mechanisms reproduce
 this value well. Except for C_2H_4 , where the endpoint concentra-
 tion is approximately zero, the mean deviation of all other
 endpoint concentrations is below 0.1%, showing that the chemi-
 cal equilibrium of the mechanism is still intact. To evaluate
 how well these reduction models reproduce the ideal gas peak
 concentrations of species across all conditions, we assess the
 maximum concentration observed during combustion relative
 to the full propane mechanism. We calculate the mean absolute
 deviation in maximum concentration between each condi-
 tion's full and reduced mechanisms, presented in Table 1. The
 data shows that, except for C_2H_4 , SPIN performs better than the
 other mechanisms in reproducing the peak concentrations. This
 is likely because SPIN optimizes the full combustion chem-
 istry within an error tolerance, while the graph-based methods
 used in our study identify species importance relative to target
 species and set a tolerance to maintain the ignition delay below
 a specific threshold, which does not necessarily correspond to
 accurately reproducing the concentrations of other combustion
 intermediates.

478 Finally, we compare the species removed from each of the
 479 mechanisms. SPIN removes six species from the full mecha-
 480 nism. These species consist primarily of dimethyl ether perox-
 481 ides and dimethyl ether peroxy radicals. This suggests that the
 482 mechanism where dimethyl ether adds peroxides and undergoes
 483 β -scission [39] is not relevant under these propane combustion
 484 conditions. Interestingly, this is supported by other works [40],
 485 which suggest that propane mechanisms could be considered
 486 separately from dimethyl ether mechanisms for many condi-
 487 tions as there is minimal cross-over. Further removing species
 488 with DRGEP, SPIN + DRGEP, and DRGEP + SPIN produces
 489 very similar sets of species. When considering the effects of
 490 two-stage reduction (SPIN and DRGEP) on species removal,
 491 DRGEP and DRGEP + SPIN have identical species sets as the
 492 species removal step is performed with an identical methodol-
 493 ogy. Of the 41 species removed by SPIN + DRGEP, all but one
 494 are oxygenated. Many of the species removed are alkyl peroxy
 495 radicals, which are a part of low-temperature propane combus-
 496 tion pathways [37]. Of the 37 compounds DRGEP removes,
 497 there is an overlap with SPIN + DRGEP on all but 3 of these
 498 compounds. This suggests that the SPIN reaction removal step
 499 does not significantly alter the importance coefficients. SPIN +
 DRGEP removes ethanol and four similar alcohol compounds,
 while DRGEP retains these species.

3.2. n-Heptane Combustion

We then move on to analyze the combustion mechanism
 of *n*-heptane [41], which is a larger and more complex fuel
 compared to the previous mechanism. To compare the reduc-
 tion methods, we consider a DRGEP-reduced mechanism, a

Table 2: Comparison of different *n*-heptane mechanisms. Size is described as the number of species N_{spe} and reactions N_{rea} , and for each reduced mechanism, accuracy is measured as the average, ID_{avg} , and maximum, ID_{max} , % error in the ignition delay, and as the mean absolute deviations in peak concentrations for three species. Error is measured relatively to the full mechanism for all 144 conditions.

	N_{spe}	N_{rea}	ID_{avg} , %	ID_{max} , %	C_2H_4 , %	H_2 , %	CO , %
Full Mechanism	629	5478	-	-	-	-	-
SPIN	454	1092	9.5	31.9	4.2	13.2	5.0
DRGEP	236	2283	13.7	32.3	3.8	4.3	1.5
DRGEP + SPIN	234	1814	13.5	28.3	15.2	6.6	4.9
SPIN + DRGEP	222	934	14.4	42.1	7.8	11.1	4.4

SPIN-reduced mechanism, a SPIN+DRGEP-reduced mechanism, and a DRGEP+SPIN-reduced mechanism. Our analysis is based on a total of 144 conditions, with temperatures ranging from 700 K to 1300 K, pressures ranging from 1 atm to 40 atm, and equivalence ratios between 0.5 and 2.

To maintain uniformity, the 40% maximum error tolerance employed for propane is utilized for the DRGEP species reduction and the combined SPIN and DRGEP methods as well. Table 2 displays the reductions, while Fig. 4 depicts the performance of each approach. In most cases, the SPIN-reduced mechanism's ignition delay deviates by less than 20% from the full mechanism, demonstrating its efficiency in reducing the larger mechanism to approximately one-fifth of the original reactions. Using the SPIN + DRGEP approach generates the fewest species and reactions in the mechanism. Implementing DRGEP + SPIN eliminates 469 reactions and 2 species from the DRGEP mechanism, with minimal change in ignition delay deviation. However, performing SPIN and then DRGEP results in a significantly greater level of reduction.

Interestingly, while SPIN did not significantly reduce the number of species in the propane mechanism, many species were eliminated during the reduction of *n*-heptane. Since SPIN only removes a species if it no longer participates in reactions and eliminates reactions, more species are likely to be eliminated in a sparse reaction network. To determine this sparsity, we computed the median number of reactions (separately counting forward and reverse reactions) involving a species for each mechanism. The full *n*-heptane mechanism has a median number of 8 reactions, whereas the median for propane is 28, suggesting that fewer reactions would have to be removed to eliminate species in the *n*-heptane mechanism. Of the 198 species eliminated during the initial SPIN reduction, 186 contain four or more carbons, indicating that this sparsity is due to larger species participating in fewer pathways. These observations suggest that SPIN can eliminate many species for larger mechanisms by removing all the reactions involving those species.

Next, we examine the effectiveness of the mechanisms in reproducing the concentrations of C_2H_4 , H, and CO. Figure 5 displays the concentration profiles, and the mean deviation for peak concentrations is shown in Tab. 2. Compared to propane, SPIN performs slightly worse, on average, in reproducing peak concentrations relative to other reduction methods. This difference can be attributed to the much higher reduction offered by SPIN, which involves roughly 20% of the reac-

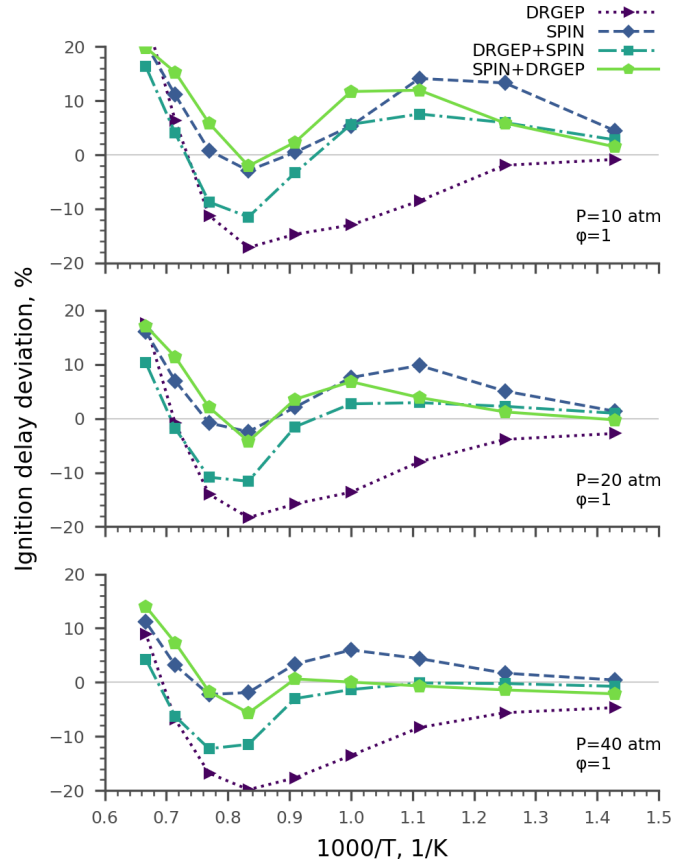


Figure 4: Comparison of ignition delay for propane mechanisms as % deviation from the full mechanism.

tions in the full mechanism and less than 50% of the reactions in DRGEP. Additionally, the data illustrate how significantly SPIN enhances overall computational time when implemented in a two-step algorithm, beginning with SPIN and culminating with DRGEP. This combination not only yields a significant, statistically noteworthy decrease in computational time but also streamlines the reaction network regarding both reactions and species. Explicitly, when compared to the reduction realized through DRGEP, this two-step method diminishes the number of species from 236 to 222 and reactions from 2283 to 934, as detailed in Tab. 2.

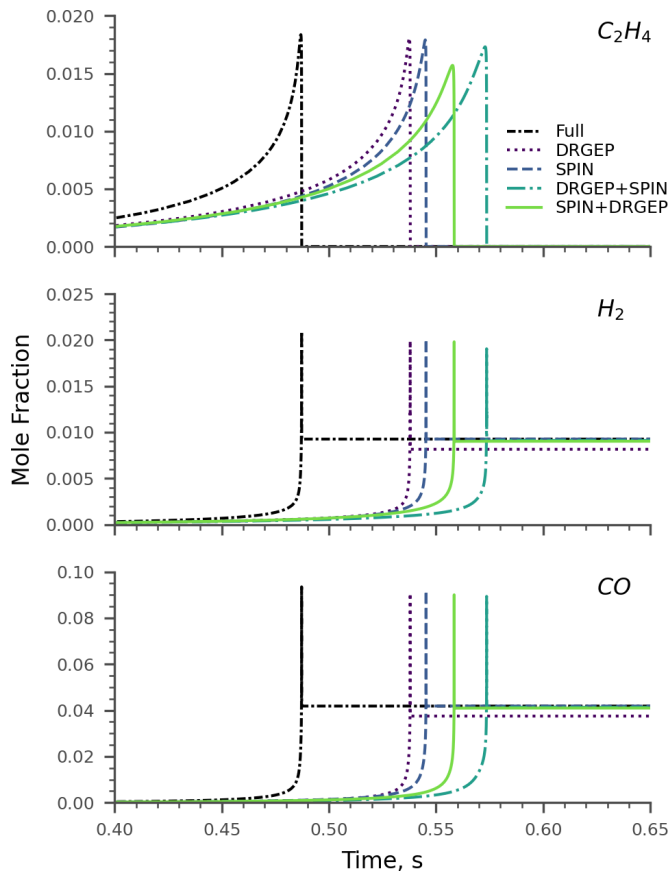


Figure 5: Time evolution of C_2H_4 , H_2 , and CO mole fractions for n -heptane combustion at 1 atm, 1000 K and equivalence ratio of 1.

3.3. Computational Cost

When it comes to the performance gained by removing reactions and species, several factors vary with the computing framework (e.g., instructions per cycle, cache, language optimization), and therefore general statements that hold the test of time are hard to make, beyond the complexity analysis outlined before. For example, we considered two classes of systems to demonstrate the potential reduction in computational cost when using a reduced mechanism. The first was 0-D homogeneous reactor simulations, reflecting the computational cost when combustion chemistry is solved in each cell during 3D CFD simulations. The second class comprised 1-D reactors that are used for premixed burners and opposed flow flames. These flame models are expected to potentially reduce computational costs when CFD simulations involve creating lookup tables based on flame calculations. However, as an example, we considered two classes of systems: 0-D homogeneous reactors, since CFD simulations rely on these model types, and 1-D reactors, as they highlight different dependencies of the execution time.

The results for the 0-D reactors, which span the same conditions used to create the reduced models, are shown in Table 3. The values are reported as the average relative execution time compared to the full mechanism, which is 0.7 s for the propane and 280 s for the heptane, truncating the ratio to the

Table 3: Relative mean time of execution for a 0-D homogeneous reactor model (Propane and Heptane), burner stabilized flame (Burner), and opposite flow flame (Opposed, as the average of 3 conditions) are reported. Additional details for these systems can be found in the text.

	0 D		1 D	
	Propane	Heptane	Burner	Opposed
Full	1.00	1.00	1.00	1.00
SPIN	0.90	0.64	0.31	0.25
DRGEP	0.93	0.53	0.10	0.09
DRGEP + SPIN	0.90	0.52	0.10	0.11
SPIN + DRGEP	0.83	0.45	0.08	0.07

last significant digit. For the 1-D systems, reported in Tab. 2, we considered only heptane combustion since it has the largest mechanism and ran a premixed and three opposed flow flames. We established the inlet condition for the premixed flame to stoichiometric n -heptane/air at 373 K and 1 atm. For the opposed flow flame, we set the fuel-sided stream to n -heptane/Air at 300 K, and the oxidizer stream to air ($O_2 + 3.76 N_2$) at the same temperature for three n -heptane/Air ratio to attain stoichiometric, lean, or rich equivalence ratios. For these simulations, given the smaller sample size, the time should be considered to have an uncertainty of about 15%.

4. Conclusions

In this paper, we present a cutting-edge, data-driven algorithm named SPIN. This algorithm has been designed to discern the most impactful reactions within a combustion network by utilizing species concentrations and reaction rates at a specific moment. This allows SPIN to provide an optimal reaction set that can accurately approximate the concentration of the species, all within a pre-set error-bearing tolerance.

SPIN's distinctive strengths are manifold. Its unique single-shot, non-iterative nature allows for a direct approach to problem-solving. The algorithm also promotes unbiased results towards any predetermined target, not including post-reduction parameter fine-tuning. The performance of SPIN can be accredited to the fusion of tools from multiple domains. Initially, SPIN utilizes Wiener filters for network identification, applying least-squared error minimization and suppression techniques for non-dominant reaction modes. The sparse identification problem is then effectively formulated as a MILP, thus simplifying computational challenges through the use of convex LP relaxation. Lastly, DRGEP's optional application for species reduction further diminishes the overall combustion network size.

We validated SPIN's effectiveness through applications on the reaction networks resulting from propane and n -heptane combustion. The algorithm notably exceeds the performance of our preceding sparse learning approach, especially in reducing propane mechanisms. Impressively, even with the intricacies inherent to n -heptane fuel, the mechanism reduced by SPIN presents a mere 9.5% average deviation in ignition delay, even

628 after more than 80% of the original mechanism's reactions have⁶⁸⁸
 629 been disposed of.⁶⁸⁹
 630 This performance, devoid of biases towards any specific tar⁶⁹⁰
 631 get property, highlights the immense potential of SPIN. Op-⁶⁹¹
 632 timized model parameters and reduction have facilitated the⁶⁹³
 633 best possible outcome, further reinforcing the fact that SPIN⁶⁹⁴
 634 can function as either a standalone method or a supplemental⁶⁹⁵
 635 tool to extant species-based reduction methods. This amplifies⁶⁹⁶
 636 SPIN's strength in identifying crucial reactions and, thus, map-⁶⁹⁸
 637 ping complex combustion processes. In conclusion, the unpar-⁶⁹⁹
 638 alleled potential of this data-driven approach for scrutinizing⁷⁰⁰
 639 intricate reaction networks and facilitating catalytic mechanism⁷⁰²
 640 reduction in a computationally effective manner is clear. The⁷⁰³
 641 results underscore the potential impact SPIN may hold for the⁷⁰⁴
 642 future progress of combustion network analysis.⁷⁰⁵

643 Acknowledgments

644 This research has been funded in part by the US Army Re-⁷¹¹
 645 search Office grants W911NF-14-1-0359 (co-PI A. Violi) and⁷¹²
 646 W911NF-19-1-0269 (PI A.O. Hero) by the U.S. Department⁷¹³
 647 of Transportation, FAA Center of Excellence 13-C-AJFE-GIT-⁷¹⁴
 648 067 (co-PI A. Violi), the University of Michigan BlueSky Ini-⁷¹⁵
 649 tiative (P.I. A. Violi), the National Science Foundation Graduate⁷¹⁷
 650 Research Fellowship (J. Saldinger) under Grant No. 1256260.⁷¹⁸

651 References

652 [1] D. T. Gillespie, Exact stochastic simulation of coupled chemical reac-⁷²⁴
 653 tions, *The Journal of Physical Chemistry* 81 (25) (1977) 2340–2361.⁷²⁵
 654 [2] V. Raman, M. Hassanal, Emerging trends in numerical simulations of⁷²⁶
 655 combustion systems, *Proceedings of the Combustion Institute* 37 (2)⁷²⁷
 656 (2019) 2073–2089.⁷²⁸
 657 [3] T. Lu, C. K. Law, Toward accommodating realistic fuel chemistry in⁷²⁹
 658 large-scale computations, *Progress in Energy and Combustion Science*,⁷³⁰
 659 35 (2) (2009) 192–215.⁷³¹
 660 [4] A. S. Tomlin, M. J. Pilling, T. Turányi, J. H. Merkin, J. Brindley, Mech-⁷³²
 661 anism reduction for the oscillatory oxidation of hydrogen: sensitivity and⁷³³
 662 quasi-steady-state analyses, *Combustion and flame* 91 (2) (1992) 107–⁷³⁴
 663 130.⁷³⁵
 664 [5] A. Stagni, A. Frassoldati, A. Cuoci, T. Faravelli, E. Ranzi, Skeletal mech-⁷³⁶
 665 anism reduction through species-targeted sensitivity analysis, *Combustion*⁷³⁷
 666 and *Flame* 163 (2016) 382–393.⁷³⁸
 667 [6] Y. Chang, M. Jia, B. Niu, M. Xie, C. Zhou, Reduction of Detailed Chemi-⁷³⁹
 668 cal Mechanisms Using Reaction Class-Based Global Sensitivity and Path⁷⁴⁰
 669 Sensitivity Analyses, *Energy & Fuels* 33 (9) (2019) 9289–9301.⁷⁴¹
 670 [7] B. Bhattacharjee, D. A. Schwer, P. I. Barton, W. H. Green, Optimally⁷⁴²
 671 Reduced Kinetic Models: Reaction Elimination in Large-Scale Kinetic⁷⁴³
 672 Mechanisms, *Combustion and Flame* 135 (3) (2003) 191–208.⁷⁴⁴
 673 [8] T. Lu, C. K. Law, A Directed Relation Graph Method for Mechanism Re-⁷⁴⁵
 674 duction, *Proceedings of the Combustion Institute* 30 (2005) 1333–1341.⁷⁴⁶
 675 [9] P. Pepiot-Desjardins, H. Pitsch, An Efficient Error-Propagation-Based Re-⁷⁴⁷
 676 duction Method for Large Chemical Kinetic Mechanisms, *Combustion*⁷⁴⁸
 677 and *Flame* 154 (1-2) (2008) 67–81.⁷⁴⁹
 678 [10] J. Wang, Y. Li, M. Liu, L. Ji, Statistical degree screening method for⁷⁵⁰
 679 combustion mechanism reduction, *Combustion and Flame* 230 (2021),⁷⁵¹
 680 111440.⁷⁵²
 681 [11] K. Wan, C. Barnaud, L. Vervisch, P. Domingo, Chemistry reduction us-⁷⁵³
 682 ing machine learning trained from non-premixed micro-mixing model-⁷⁵⁴
 683 ing: Application to DNS of a syngas turbulent oxy-flame with side-wall⁷⁵⁵
 684 effects, *Combustion and Flame* 220 (2020) 119–129.⁷⁵⁶
 685 [12] M. A. Calicchia, E. Atefi, J. C. Leylegian, Creation of small kinetic mod-⁷⁵⁷
 686 els for CFD applications: a meta-heuristic approach, *Engineering with*⁷⁵⁸
 687 *Computers* (2021) 1–15.

[13] G. D'Alessio, A. Parente, A. Stagni, A. Cuoci, Adaptive chemistry via
 pre-partitioning of composition space and mechanism reduction, *Combustion and Flame* 211 (2020) 68–82.
 [14] N. Jaouen, L. Vervisch, P. Domingo, Auto-thermal reforming (ATR) of
 natural gas: An automated derivation of optimised reduced chemical
 schemes, *Proceedings of the Combustion Institute* 36 (3) (2017) 3321–
 3330.
 [15] M. Kelly, M. Fortune, G. Bourque, S. Dooley, Machine learned compact
 kinetic models for methane combustion, *Combustion and Flame* 253
 (2023) 112755.
 [16] J. An, G. He, K. Luo, F. Qin, B. Liu, Artificial neural network based
 chemical mechanisms for computationally efficient modeling of hydro-
 gen/carbon monoxide/kerosene combustion, *International Journal of Hy-
 drogen Energy* 45 (53) (2020) 29594–29605.
 [17] P. Zhang, S. Liu, D. Lu, R. Sankaran, G. Zhang, An out-of-distribution-
 aware autoencoder model for reduced chemical kinetics, *Discrete and
 Continuous Dynamical Systems - S* 15 (4) (2022) 913–930.
 [18] F. Harirchi, D. Kim, O. Khalil, S. Liu, P. Elvati, M. Baranwal, A. Hero,
 A. Violi, On sparse identification of complex dynamical systems: A study
 on discovering influential reactions in chemical reaction networks, *Fuel*
 279 (2020) 118204.
 [19] M. J. Zimón, R. Prosser, D. Emerson, M. K. Borg, D. Bray, L. Grinberg,
 J. M. Reese, An evaluation of noise reduction algorithms for particle-
 based fluid simulations in multi-scale applications, *Journal of Computa-
 tional Physics* 325 (2016) 380–394.
 [20] M. Feinberg, Chemical reaction network structure and the stability of
 complex isothermal reactors—I. The deficiency zero and deficiency one
 theorems, *Chemical engineering science* 42 (10) (1987) 2229–2268.
 [21] D. T. Gillespie, Stochastic simulation of chemical kinetics, *Annu. Rev.
 Phys. Chem.* 58 (2007) 35–55.
 [22] V. Chellaboina, S. P. Bhat, W. M. Haddad, D. S. Bernstein, Modeling and
 analysis of mass-action kinetics, *IEEE Control Systems Magazine* 29 (4)
 (2009) 60–78.
 [23] D. F. Anderson, T. G. Kurtz, Continuous time Markov chain models for
 chemical reaction networks, in: *Design and analysis of biomolecular cir-
 cuits*, Springer, 3–42, 2011.
 [24] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time
 series*, vol. 2, 1949.
 [25] F. Glover, Improved linear integer programming formulations of nonlin-
 ear integer problems, *Management Science* 22 (4) (1975) 455–460.
 [26] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer, New York,
 NY, USA, 2006.
 [27] T. Lu, C. K. Law, On the applicability of directed relation graphs to the
 reduction of reaction mechanisms, *Combustion and Flame* 146 (3) (2006)
 472–483.
 [28] P. Pepiot, H. Pitsch, Systematic reduction of large chemical mechanisms,
 in: *4th Joint Meeting of the US Sections of the Combustion Institute*, vol.
 2123, Drexel University Pittsburgh, PA, USA, 2005, 2005.
 [29] E. L. Lawler, D. E. Wood, Branch-and-bound methods: A survey, *Opera-
 tions research* 14 (4) (1966) 699–719.
 [30] M. Tawarmalani, N. V. Sahinidis, A polyhedral branch-and-cut approach
 to global optimization, *Mathematical programming* 103 (2) (2005) 225–
 249.
 [31] E. L. Petersen, D. M. Kalitan, S. Simmons, G. Bourque, H. J. Curran,
 J. M. Simmie, Methane/propane oxidation at high pressures: Experimen-
 tal and detailed chemical kinetic modeling, *Proceedings of the combus-
 tion institute* 31 (1) (2007) 447–454.
 [32] CHEMKIN version 18.0, 2016.
 [33] C. K. Law, Combustion at a crossroads: Status and prospects, *Proceedings
 of the Combustion Institute* 31 (1) (2007) 1–29.
 [34] K. E. Niemeyer, C.-J. Sung, M. P. Raju, Skeletal mechanism generation
 for surrogate fuels using directed relation graph with error propagation
 and sensitivity analysis, *Combustion and Flame* 157 (9) (2010) 1760 –
 1770.
 [35] Y. Shi, H.-W. Ge, J. Brakora, R. Reitz, Automatic Chemistry Mechanism
 Reduction of Hydrocarbon Fuels for HCCI Engines Based on DRGEP
 and PCA Methods with Error Control, *Energy Fuels* 24 (3) (2010) 1646
 – 1654.
 [36] E. Simmons, ESR measurements of free radical concentration profiles in
 the propane-oxygen flame, *Journal of Magnetic Resonance* (1969) 16 (1)
 (1974) 130–135.

- 759 [37] S. S. Merchant, C. F. Goldsmith, A. G. Vandeputte, M. P. Burke, S. J.
760 Klippenstein, W. H. Green, Understanding low-temperature first-stage ig-
761 nition delay: Propane, *Combustion and Flame* 162 (10) (2015) 3658–
762 3673.
- 763 [38] H. Wang, M. Frenklach, A detailed kinetic modeling study of aromatics
764 formation in laminar premixed acetylene and ethylene flames, *Combustion
765 and Flame* 110 (1) (1997) 173 – 221.
- 766 [39] H. J. Curran, S. L. Fischer, F. L. Dryer, The reaction kinetics of dimethyl
767 ether. II: Low-temperature oxidation in flow reactors, *International Jour-
768 nal of Chemical Kinetics* 32 (12) (2000) 741–759.
- 769 [40] E. E. Dames, A. S. Rosen, B. W. Weber, C. W. Gao, C.-J. Sung,
770 W. H. Green, A detailed combined experimental and theoretical study on
771 dimethyl ether/propane blended oxidation, *Combustion and Flame* 168
772 (2016) 310–330.
- 773 [41] M. Mehl, W. Pitz, C. Westbrook, H. Curran, Kinetic Modeling of Gas-
774 oline Surrogate Components and Mixtures Under Engine Conditions, *Pro-
775 ceedings of the Combustion Institute* 33 (2011) 193 – 200.