

# ChemGen: Chemistry Reaction Code Generation for Computational Physics

Ryan F. Johnson<sup>a</sup>, Eric J. Ching<sup>a</sup>, Ethan S. Genter<sup>b</sup>, Joshua E. Lipman<sup>b</sup>, Andrew D. Kercher<sup>a</sup>, Jay Arcities<sup>c</sup>, Hai Wang<sup>b</sup>

<sup>a</sup>*Laboratories for Computational Physics and Fluid Dynamics, U.S. Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC 20375*

<sup>b</sup>*Department of Mechanical Engineering, Stanford University, Stanford, CA 94305*

<sup>c</sup>*Department of Aerospace Engineering, San Diego State University, San Diego, CA 92182*

---

## Abstract

This paper introduces ChemGen, a software package that uses code generation to integrate multispecies thermodynamics and chemical kinetics into C++-based computational physics codes. ChemGen aims to make chemical kinetics more accessible in existing simulation frameworks and help bridge the gap between combustion modeling and computational physics. The package employs the concept of decorators which enable flexible C++ code generation to target established software ecosystems. ChemGen generates code to evaluate thermodynamic properties, chemical source terms, and their analytical derivatives for Jacobian calculations. Also included are a variety of implicit time integration schemes, linear solvers, and preconditioners. The various components of ChemGen are verified by demonstrating agreement with Cantera and/or theoretical convergence rates. Finally, we integrate ChemGen into OpenFOAM and achieve a speedup over its native chemistry solver by approximately four times. ChemGen is an ongoing project released under the NRL Open License, a source-available license provided by the U.S. Naval Research Laboratory.

**Keywords:** Chemistry, Combustion, Computational Physics, Computational Fluid Dynamics, Code Generation

---

## 1. Introduction

Modeling of chemical reactions is critical to many areas of computational physics. One of the most common examples of this coupling is in Computational Fluid Dynamics (CFD) [1, 2, 3], which is used to simulate fluid behavior in a wide variety of applications. The governing equations, typically the Navier-Stokes equations, are discretized in space using approaches such as finite volume [4, 5, 6], finite difference [7], and finite element schemes [8, 9]. CFD relies on large volumes of data to describe the fluid; in particular, the fluid state variables stored at all grid points in the computational domain are known as the degrees of freedom (DoF).

Developing robust CFD software is difficult due to computational challenges such as efficient DoF communication in parallel, adaptive physical modeling, and managing complex geometries. Additionally, there are significant numerical issues, such as preserving stability [10, 11, 12, 13, 14, 15, 16, 17] which is coupled with the choice of formulation of the conservation laws and time integration strategy. There is no single CFD software that simultaneously addresses all the above difficulties for all fluid dynamics applications, necessitating continued research into accelerating simulations while achieving stable and accurate fluid modeling. As a result, a broad family of CFD codes has emerged, ranging from experimental in-house tools used

---

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

for demonstrating new technologies to established commercial software suites employed in production-level engineering.

A typical three-dimensional simulation of compressible, non-reacting flow involves five conserved variables: energy, mass, and three momentum components, such that  $n_y = 5$ , where  $n_y$  is the total state size. In many cases, existing non-reacting CFD software built around these five variables can be extended to chemically reacting flows by expanding the state to include multiple chemical species and incorporating  $n_r$  chemical reactions. The state size then becomes  $n_y = n_s + 4$ , where  $n_s$  represents the number of chemical species<sup>0</sup>. Detailed chemical models may involve hundreds of species and thousands of reactions, often with  $n_r \approx 5n_s$  [18]. The number of species increases with the complexity of the fuel; for instance, the hydrogen model in [19, 17] consists of 10 chemical species and 17 reactions, while the FFCM-2 model for simple hydrocarbons (up to four carbons per chemical species) comprises 96 species and 1054 reactions [20].

In order to extend an existing single-species CFD software to basic chemically reacting flows, three main features must be added to the formulation:

1. **Multicomponent equation of state:** Most single-component formulations assume a calorically perfect gas, where pressure is linear with respect to internal energy. However, this assumption does not hold for multicomponent reacting flow, requiring careful adjustments to ensure consistent computations between among energy, temperature, and pressure.
2. **Stiff, nonlinear chemical source terms:** Other than external forcing or simple buoyancy forces, the simplest CFD Navier-Stokes models typically do not include a source term. In most models, momentum, energy, and mass have no sources or sinks to add or remove them from the domain; instead, the conserved variables are transported and diffused. Furthermore, chemical source terms can be extremely stiff, introducing additional numerical challenges.
3. **Multicomponent transport:** Transport quantities, such as viscosity and thermal conductivity, would need to be adapted to multicomponent mixture models. In addition, the diffusion of each species needs to be considered with a transport method.

There are several approaches to integrate these features. The most common are:

1. Adapt existing legacy software, such as Chemkin [21], for the target CFD code.
2. Use the more widely adopted, open-source, and actively maintained Cantera [22], which can serve as a third-party library for codes written in C, C++, and Fortran, as well as in interpreted languages such as Python and MATLAB.
3. Manually code chemical source term evaluations, thermodynamic fits, transport properties, and other physical models.
4. Use code generation tools to read in the chemical model and produce adaptable source code in the desired format for the CFD software.

Approaches 1 and 2 are reasonable choices but present certain challenges. With approach 1, the existing legacy routines may not easily adapt to the specific requirements of the targeted CFD software, limiting scalability and complicating code maintenance. Regarding approach 2, although the package is excellent for property calculations, it can only act as a third-party library, which may lead to limited control over software functionality. Additionally, relying on a third-party library (e.g. Cantera and Mutation++ [23]) may hinder compatibility with targeted architectures that developers cannot directly access. Using approach 3 can also introduce challenges; it often requires extensive work by developers, as the necessary models would need to be built from scratch (further described in Section 2). This work belongs to approach 4 and delivers a tool to generate source code that is compatible with the existing ecosystem of a target software package.

Code generation is a well-established concept, with a history in chemistry applications. There are two publicly available code generators focused on gas-phase combustion: PyJac [24] and Pyrometheus [25]. PyJac specializes in generating analytical source term Jacobians, which can be used to accelerate the computation

---

<sup>0</sup>The original five state variables are reduced by one since total density can be determined from the species concentrations (or partial densities).

of chemical species evolution. Pyrometheus generates either C++ header files or Python files for gas-phase combustion calculations, incorporating several on-the-fly optimizations to enhance scalability for high-performance computing in CFD. Additionally, KinetiX [26] is a software toolkit that generates chemical source terms and thermodynamic and mixture-averaged transport properties tailored to specific hardware targets, such as CPU or GPU architectures.

In this work, we introduce ChemGen, a software package that leverages code generation to integrate multispecies thermodynamics and chemical kinetics into C++ based codes. ChemGen introduces decorators, discussed in detail in Section (4), to enable flexible C++ code generation tailored to existing software ecosystems. ChemGen currently focuses on thermodynamic properties and chemical source terms, as well as their analytical derivatives. The analytical derivatives facilitate Jacobian assembly for implicit time integration strategies appropriate for CFD. In addition, a number of time integrators, linear solvers, and preconditioners is provided for convenience.

This paper is organized as follows. We begin with an overview of the chemically reacting flow conservation equations used in CFD, highlighting where chemistry fits within these formulations, including details on multi-species thermodynamics and chemical reaction rates. Next, we discuss time integration strategies, focusing on operator splitting approaches that reduce the problem to a system of ordinary differential equations for chemical state evolution, as well as fully coupled formulations. We then introduce the code generation process, emphasizing how malleability is achieved through the use of decorators. Finally, we present results verifying the source term evaluation, Jacobian computation, and time integration. We demonstrate ChemGen’s performance by replacing OpenFOAM’s existing ODE-based chemical integrator in a detonation simulation and achieve faster time-to-solution due to ChemGen’s specialized, generated source code.

## 2. Governing equations and thermodynamics

In this work, we focus on a compressible CFD formulation without viscous terms for simplicity, as we only seek to contextualize chemistry in the framework of fluid dynamics. Here, we primarily highlight the interaction between the fluid dynamic state,  $y_f$ , and the chemical state,  $y_c$ .

The multicomponent chemically reacting compressible Euler equations are given as

$$\frac{\partial y_f}{\partial t} + \nabla \cdot \mathcal{F}(y_f) - \mathcal{S}(y_f) = 0 \quad (1)$$

where  $t$  is time,  $y_f$  is the conservative state vector,  $\mathcal{F}(y_f)$  is the convective flux, and  $\mathcal{S}(y_f)$  is the CFD source term that includes the chemical reactions. The fluid state is

$$y_f = (\rho v_1, \dots, \rho v_d, \rho e_t, C_1, \dots, C_{n_s}), \quad (2)$$

where  $d$  is the spatial dimension,  $n_s$  is the number of species (which yields a state size  $n_y = d + n_s + 1$ ),  $\rho$  is density,  $v = (v_1, \dots, v_d)$  is the velocity vector,  $e_t$  is the mass-specific total energy, and  $C = (C_1, \dots, C_{n_s})$  is the vector of molar species concentrations. The density is computed from the species concentrations as

$$\rho = \sum_{i=1}^{n_s} \rho_i = \sum_{i=1}^{n_s} W_i C_i,$$

where  $\rho_i$  is the partial density and  $W_i$  is the molecular mass of the  $i$ th species. The mass fraction of the  $i$ th species is defined as

$$Y_i = \frac{W_i C_i}{\rho} = \frac{\rho_i}{\rho}.$$

The  $k$ th spatial convective flux component is written as

$$\mathcal{F}_k(y) = (\rho v_k v_1 + p \delta_{k1}, \dots, \rho v_k v_d + p \delta_{kd}, v_k (\rho e_t + p), v_k C_1, \dots, v_k C_{n_s}), \quad (3)$$

where  $p$  is the pressure, and  $\delta_{ij}$  is the Kronecker delta. The mass-specific total energy is the sum of the internal and kinetic energies, given by

$$e_t = u + \frac{1}{2} \sum_{k=1}^d v_k v_k,$$

where the (mixture-averaged) mass-specific internal energy,  $u$ , is:

$$u = \sum_{i=1}^{n_s} Y_i u_i.$$

The chemical state is

$$y_c = (T, C_1, \dots, C_{n_s}) \quad (4)$$

which is the complete derived state needed to calculate the components of  $\mathcal{S}(y_f)$ . The species concentrations,  $C_i$ , are all conserved quantities and can be extracted readily from the fluid state. The pressure from the ideal gas law is

$$p = RT \sum_{i=1}^{n_s} C_i,$$

with  $R = 8314.4626 \text{ J} \cdot \text{kmol}^{-1} \text{K}^{-1}$  denoting the universal gas constant. An additional parameter used in optimizations of reactions is  $\ln T$ , which is common to utilize along with the chemical state to avoid repeated logarithm computations. Finally the mixture concentration,

$$M_c = \sum_{i=1}^{n_s} C_i = \frac{p}{RT},$$

is utilized in the evaluation of many reaction rates, such as third body reactions with no specified efficiencies. In the following sections, we assume that the elements of  $y_c$ ,  $\ln T$ ,  $p$ , and  $M_c$  are accessible at all levels of ChemGen. As such, functional dependencies are omitted for brevity.

### 2.1. Thermodynamics

This work assumes a thermally perfect gas, with  $u_i$  given by [27]

$$u_i = h_i - R_i T = h_{\text{ref},i} + \int_{T_{\text{ref}}}^T c_{p,i}(\tau) d\tau - R_i T, \quad (5)$$

where  $h_i$  is the mass-specific enthalpy of the  $i$ th specie,  $R_i = R/W_i$ ,  $T_{\text{ref}}$  is the reference temperature (298 K),  $h_{\text{ref},i}$  is the reference-state enthalpy of formation of specie  $i$  evaluated as  $h_{\text{ref},i} = h_i(T_{\text{ref}})$ , and  $c_{p,i}$  is the mass-specific heat capacity at constant pressure of the  $i$ th specie.  $c_{p,i}$  is computed from an  $n_p$ -order polynomial as

$$c_{p,i} = \sum_{k=0}^{n_p} a_{ik} T^k, \quad (6)$$

which is based on but not equivalent to the NASA n-coefficient polynomial parametrization [28, 29]. For differentiability and computational efficiency, ChemGen fits the polynomial for  $c_{p,i}$  to arbitrary order. Equation (5) yields

$$h_i = h_{\text{ref},i} + \sum_{k=0}^{n_p} \frac{a_{ik}}{k+1} T^{k+1} = \sum_{k=0}^{n_p+1} b_{ik} T^k \quad (7)$$

and

$$u_i = h_{\text{ref},i} + \sum_{k=0}^{n_p} \frac{a_{ik}}{k+1} T^{k+1} - RT = \sum_{k=0}^{n_p+1} c_{ik} T^k. \quad (8)$$

Here, we use  $h_i(T = 298 \text{ K})$  from the existing NASA coefficients,

$$h_{\text{ref},i} = h_i(T = 298 \text{ K}) - \sum_{k=0}^{n_p} \frac{a_{ik}}{k+1} (298 \text{ K})^{k+1}, \quad (9)$$

which gives

$$b_{ik} = \begin{cases} \frac{a_{i,k-1}}{k}, & k \geq 1 \\ h_{\text{ref},i}, & k = 0 \end{cases} \quad (10)$$

and

$$c_{ik} = \begin{cases} \frac{a_{i,k-1}}{k}, & k > 1 \\ a_{i,0} - R, & k = 1 \\ h_{\text{ref},i}, & k = 0. \end{cases} \quad (11)$$

This integration requires two vectors for temperature: a monomial basis for polynomials of degree  $n_p$ ,  $T_{n_p} = (1, T, T^2, \dots, T^{n_p})$  and a monomial basis for polynomials of degree  $n_p + 1$ ,  $T_{n_p+1} = (1, T, T^2, \dots, T^{n_p+1})$ . These vectors can then be contracted with a vector of coefficients; for example,  $u_i = \langle A_i, T_{n_p+1} \rangle$  where  $A_i = (c_{i,0}, c_{i,1}, c_{i,2}, \dots, c_{i,n_p+1})$ .

The mass-specific thermodynamic entropy of the mixture is defined as

$$s = \sum_{i=1}^{n_s} Y_i s_i, \quad (12)$$

with  $s_i$  given by

$$s_i = s_i^o - R \log \frac{p_i}{p_{\text{ref}}}, \quad s_i^o = s_{\text{ref},i}^o + \int_{T_{\text{ref}}}^T \frac{c_{p,i}(\tau)}{\tau} d\tau, \quad (13)$$

where  $s_{\text{ref},i}^o$  is the species entropy of formation from the NASA n-coefficient polynomial parametrization evaluated at  $T = 298 \text{ K}$ , and  $p_{\text{ref}}$  is the reference pressure (usually  $p_{\text{ref}} = 1 \text{ atm}$ ),  $s_i^o$  denotes the species entropy at the reference pressure, and  $p_i = C_i RT$  is the partial pressure. The  $R \log \frac{p_i}{p_{\text{ref}}}$  term accounts for the effect of pressure on entropy and does not depend on thermodynamic polynomial fits. Using Equation (6),  $s_i^o$  is computed as

$$s_i^o = s_{\text{ref},i}^o + a_{i,0} \ln T + \sum_{k=1}^{n_p} \frac{a_{ik}}{k} T^k = a_{i,o} \ln T + \sum_{k=0}^{n_p} d_{ik} T^k, \quad (14)$$

where

$$d_{ik} = \begin{cases} \frac{a_{i,k}}{k}, & k \geq 1 \\ s_{\text{ref},i}^o, & k = 0. \end{cases} \quad (15)$$

Finally, the Gibbs free energy of the  $i$ th species is

$$g_i^o = h_i - T s_i^o = b_{i,0} - a_{i,0} T \ln T + \sum_{k=1}^{n_p} (b_{ik} - d_{i,k-1}) T^k + b_{i,n_p+1} T^{n_p+1}, \quad (16)$$

which can be represented as

Thermodynamic Quantity	Fit	Coefficient Description
Species specific heat at constant pressure, $c_{p,i}$	$c_{p,i} = \sum_{k=0}^{n_p} a_{ik} T^k$	Re-fit from existing data
Species specific enthalpy, $h_i$	$h_i = \sum_{k=0}^{n_p+1} b_{ik} T^k$	$b_{ik} = \begin{cases} \frac{a_{i,k-1}}{k}, & k \geq 1 \\ h_{\text{ref},i}, & k = 0 \end{cases};$ $h_{\text{ref},i} = h_i(T_{\text{ref}}) - \sum_{k=0}^{n_p} \frac{a_{ik}}{k+1} (298)^{k+1}$
Species specific internal energy, $u_i$	$u_i = \sum_{k=0}^{n_p+1} c_{ik} T^k$	$c_{ik} = \begin{cases} \frac{a_{i,k-1}}{k}, & k > 1 \\ a_{i,0} - R, & k = 1 \\ h_{\text{ref},i}, & k = 0. \end{cases}$
Species specific standard entropy, $s_i^o$	$s_i^o = a_{i,0} \ln T + \sum_{k=0}^{n_p} d_{ik} T^k$	$d_{ik} = \begin{cases} \frac{a_{i,k}}{k}, & k \geq 1 \\ s_{\text{ref},i}^o, & k = 0. \end{cases}$
Species specific standard Gibbs free energy, $g_i^o$	$g_i^o = -a_{i,0} T \ln T + \sum_{k=0}^{n_p+1} f_{ik} T^k$	$f_{ik} = \begin{cases} \left(\frac{1}{k} - \frac{1}{k-1}\right) a_{i,k-1} & n_p + 1 > k > 1 \\ a_{i,0} - s_{\text{ref},i}^o, & k = 1 \\ h_{\text{ref},i}, & k = 0. \end{cases}$

Table 1: Summary of polynomial coefficients in ChemGen for thermally perfect species specific relationships.

$$g_i^o = -a_{i,0} T \ln T + \sum_{k=0}^{n_p+1} f_{ik} T^k. \quad (17)$$

Note, the Gibbs free energy, like entropy, has a pressure dependency that is calculated independent of these thermodynamic fits.

This gives two new temperature vectors:  $T_{n_p, \ln T} = (1, T, T^2, \dots, T^{n_p}, \ln T)$  and  $T_{n_p+1, T \ln T} = (1, T, T^2, \dots, T^{n_p+1}, -T \ln T)$ . These vectors can then be contracted with a vector of coefficients,  $s_i = \langle A_i, T_{n_p, \ln T} \rangle$  where  $A_i = (d_{i,0}, d_{i,1}, d_{i,2}, \dots, d_{i,n_p}, a_{i,0})$  and  $g_i = \langle A_i, T_{n_p+1, T \ln T} \rangle$  where  $A_i = (f_{i,0}, f_{i,1}, f_{i,2}, \dots, f_{i,n_p+1}, a_{i,0})$ .

The thermodynamic relationships used by ChemGen are summarized in Table (1).

ChemGen supplies a Newton's method implementation to extract temperature from internal energy in the case where temperature is needed from  $y_f$ . For example, internal energy is typically part of a conserved state. The total energy in  $y_f$  is given as

$$\rho e_t = \rho u + \frac{1}{2} \sum_{k=1}^d \rho v_k v_k, \quad (18)$$

where the term  $\frac{1}{2} \sum_{k=1}^d \rho v_k v_k$  can be computed from other conserved variables. Thus one can directly compute the internal energy at any time as

$$\rho u = \frac{1}{2} \sum_{k=1}^d \frac{(\rho v_k)(\rho v_k)}{\sum_i^n W_i C_i} - \rho e_t. \quad (19)$$

The temperature is determined by a root finding algorithm which iterates via Netwon's method. The temperature at the  $(k+1)$ th Newton iteration is given by

$$T_{k+1} = T_k - \frac{\rho u(T_k)}{\frac{\partial \rho u(T_k)}{\partial T}}. \quad (20)$$

where  $T_k$  is the temperature at the  $k$ th iteration. The numerator and denominator are given by

$$\rho u(C, T_k) = \sum_{i=1}^{n_s} W_i C_i u_i(T_k) = \sum_{i=1}^{n_s} W_i C_i \langle A_i, T_{k,n_p+1} \rangle \quad (21)$$

and

$$\frac{\partial \rho u(C, T_k)}{\partial T} = \rho c_v(C, T_k) = \sum_{i=1}^{n_s} W_i C_i \frac{\partial u_i(T_k)}{\partial T} = \sum_{i=1}^{n_s} W_i C_i \left\langle A_i, \frac{\partial T_{k,n_p+1}}{\partial T} \right\rangle, \quad (22)$$

respectively, where

$$A_i = (c_{i,0}, c_{i,1}, c_{i,2}, \dots, c_{i,n_p+1}) \quad (23)$$

and

$$\frac{\partial T_{k,n_p+1}}{\partial T} = (0, 1, 2T_k, \dots, (n_p + 1) T_k^{n_p}). \quad (24)$$

$c_v$  in Equation 22 is the specific heat at constant volume, which is related to  $c_p$  as

$$c_v = c_p - R. \quad (25)$$

In ChemGen, Newton's method for computing temperature terminates after five iterations, which in our experience is sufficient to be within  $1 \times 10^{-8}$  relative error for a variety of models at different conditions.

## 2.2. Chemical reaction rates and their ChemGen representations

The source term in Equation (1) is a function of the fluid state, written as [30]

$$\mathcal{S}(y_f) = (0, \dots, 0, 0, \omega_1, \dots, \omega_{n_s}), \quad (26)$$

where  $\omega_i$  is the production rate of the  $i$ th specie, which satisfies mass conservation:

$$\sum_{i=1}^{n_s} W_i \omega_i = 0. \quad (27)$$

The production rate is computed as

$$\omega_i = \sum_{j=1}^{n_r} \nu_{ij} q_j,$$

where  $n_r$  is the number of reactions,  $\nu_{ij} = \nu_{ij}^r - \nu_{ij}^f$  is the difference between the reverse ( $\nu_{ij}^r$ ) and the forward stoichiometric coefficients ( $\nu_{ij}^f$ ), and  $q_j$  is the rate of progress of the  $j$ th reaction, computed as

$$q_j = k_j^f \prod_{i=1}^{n_s} C_i^{\nu_{ij}^f} - k_j^r \prod_{i=1}^{n_s} C_i^{\nu_{ij}^r}, \quad (28)$$

where  $k_j^f$  and  $k_j^r$  are the forward and reverse rate constants, respectively, of the  $j$ th reaction. The forward and reverse rate constants are related via the equilibrium constant

$$K_j = \frac{k_j^f}{k_j^r}, \quad k_j^r = \frac{k_j^f}{K_j}, \quad (29)$$

where

$$K_j = \exp \left( -\frac{\Delta G'_j}{RT} \right) \left( \frac{p_{\text{ref}}}{RT} \right)^{\sum_i \nu_{ij}}, \quad (30)$$

with  $\Delta G'_j$  denoting the change in reference-state Gibbs free energy for the  $j$ th reaction, given as

$$\Delta G'_j = \sum_{i=1}^{n_s} \nu_{ij} g_i^o.$$

For the term  $(\frac{p_{ref}}{RT})^{\sum_i \nu_{ij}}$  in (30), ChemGen inspects  $\sum_i \nu_{ij}$  and finds the specific power. If the power is a positive integer  $(\frac{p_{ref}}{RT})$  is expressed as a series of multiplications, e.g.  $\sum_i \nu_{ij} = 3$  gives  $(\frac{p_{ref}}{RT}) \cdot (\frac{p_{ref}}{RT}) \cdot (\frac{p_{ref}}{RT})$  with  $1/RT$  precomputed. Similarly, if  $\sum_i \nu_{ij}$  is a negative integer, ChemGen precomputes  $1/p_{ref}$  and then calculates the corresponding power of  $RT$ , e.g.,  $\sum_i \nu_{ij} = -3$  gives  $(1/p_{ref})^3 (RT)^3$ . Without code generation, these types of expressions require expensive exponentiation functions and possibly if-statements.

There exist various models for approximating the forward rate expressions,  $k_j^f$ , in Equation (28). The reaction models currently supported by ChemGen are discussed below.

### 2.2.1. Arrhenius reactions

The Arrhenius form is the most common model for approximating reaction rates. The forward rate constants are computed as

$$k_j^f = k_{A,j}^f = A_j T^{b_j} \exp\left(-\frac{E_j}{RT}\right),$$

where  $A_j$  is the prefactor,  $b_j$  is the temperature exponent, and  $E_j$  is the activation energy [27, 30]. In ChemGen, an optimization is made to again avoid the power function. Generally,  $b_j$  is not an integer, however, since the exponential function is unavoidable, we can utilize the precomputed  $\ln T$  and insert an additional term in the exponent

$$k_{A,j}^f = A_j \exp\left(b_j \ln T - \frac{E_j}{RT}\right).$$

### 2.2.2. Three-body reactions

Third-body reactions require the inclusion of the mixture concentration, which represents the collective effect of all species acting as a third body. Dissociation and recombination reactions are often of this type. The rate of progress is scaled by a prefactor as [30]

$$q_j = \left( \sum_{i=1}^{n_s} \alpha_{ij} C_i \right) \left( k_j^f \prod_{i=1}^{n_s} C_i^{\nu_{ij}^f} - k_j^r \prod_{i=1}^{n_s} C_i^{\nu_{ij}^r} \right), \quad (31)$$

where  $\alpha_{ij}$  are the third-body efficiencies. In ChemGen, the rate of progress is treated to always resemble Equation (28); therefore, the prefactor for third-body reactions is incorporated into the Arrhenius expression as

$$k_j^f = k_{TB,j}^f = \left( \sum_{i=1}^{n_s} \alpha_{ij} C_i \right) k_{A,j}^f = \left( M_c + \sum_{i=1}^{n_s} (\alpha_{ij} - 1) C_i \right) k_{A,j}^f \quad (32)$$

which is still attributed to the second term in Equation (31) via the relationship of  $k_j^f$  to  $k_j^r$  from Equation (29). An additional optimization is made during the code generation process for the  $\alpha_{ij}$  terms. The established convention is that the efficiencies always default  $\alpha_{ij} = 1$ . Therefore, for any species without an explicitly prescribed efficiency, the  $(\alpha_{ij} - 1) C_i$  term is removed during code generation. This avoids possibly repetitive work to determine the efficiency from a sum of products of all species to a sum of products of all species with  $\alpha_{ij} \neq 1$  by utilizing the precomputed  $M_c$ .

### 2.2.3. Fall-off reactions

Unimolecular/recombination fall-off reactions incorporate a dependence on pressure. In general, this model predicts an increase in the reaction rate with increasing pressure. Given Arrhenius-type low-pressure and high-pressure limits for the rate functions ( $k_{A,0,j}$  and  $k_{A,\infty,j}$ , respectively with their own  $A_j$ ,  $b_j$ , and  $E_j$  constants),  $k_j^f$  is computed as

$$k_j^f = k_{FO,j}^f = k_{A,\infty,j} \left( \frac{P_r}{1 + P_r} \right) F, \quad (33)$$

where  $P_r$  is the reduced pressure, defined as

$$P_r = \frac{k_{A,0,j}}{k_{A,\infty,j}} \left( M_c + \sum_{i=1}^{n_s} (\alpha_{ij} - 1) C_i \right).$$

The same optimization for efficiencies can be made as in the third-body prefactor. Various falloff reactions exist where the only difference is in the function  $F$ , known as the broadening term, , which we outline below, the simplest being  $F = 1$  from Lindemann [31].

#### Troe broadening

For the Troe fall-off reaction the broadening term,  $F$ , in Equation (33) is computed using the following relations:

$$\begin{aligned} \log_{10} F &= \frac{\log_{10} F_{cent}}{1 + f_1^2} \\ F_{cent} &= (1 - \alpha_j) \exp(-T/T_{j,3}) + \alpha \exp(-T/T_{j,1}) + \exp(-T_{j,2}/T) \\ f_1 &= (\log_{10} P_r + c)/(n - 0.14(\log_{10} P_r + c)) \\ c &= -0.4 - 0.67 \log_{10} F_{cent} \\ n &= 0.75 - 1.27 \log_{10} F_{cent} \end{aligned} \quad (34)$$

The origin of these fits is given in [32, 33, 34, 35].

#### SRI broadening

In the SRI [36] broadening approach,  $F$  is represented by

$$F = d[a \exp(-b/T) + \exp(-c/T)]^{1/(1+\log_{10} P_r)} T^e,$$

where coefficients  $a$  through  $e$  are specified for each reaction.

### 2.2.4. Pressure-dependent Arrhenius rate expressions

The previous fall-off reaction rates seek to fit between a low-pressure reaction rate and a high-pressure reaction rate. However, some rates have several pressure dependent zones that cannot be accurately represented by a single or two Arrhenius expressions. The pressure-dependent Arrhenius rate expression, known as the P-log reaction rate, was introduced to utilize pressure-dependent reaction rates by logarithmically interpolating between Arrhenius rate expressions at various pressures [37]. Given two Arrhenius reaction rates at two specific pressures,

$$\begin{aligned} p_i : k_i &= A_i T^{b_i} \exp(-E_i/RT) \text{ and} \\ p_{i+1} : k_{i+1} &= A_{i+1} T^{b_{i+1}} \exp(-E_{i+1}/RT), \end{aligned} \quad (35)$$

the rate at an intermediate pressure is computed as

$$\ln k_{plog} = \ln k_i + (\ln k_{i+1} - \ln k_i) \frac{\ln p - \ln p_i}{\ln p_{i+1} - \ln p_i}. \quad (36)$$

An additional complexity is that multiple rate expressions may be given at the same pressure, in which case the rate used in the interpolation formula is the sum of all the rates given at that pressure. For pressures outside the given range, the rate expression at the nearest bounded pressure is used. In ChemGen, the  $\ln p_i$  terms are calculated during code generation.

To further elaborate on this rate, consider, for example, the 81st reaction in FFCM2 [20],

```
- equation: CH2(S) + H2O <=> CH2O + H2 # Reaction 81
  type: pressure-dependent-Arrhenius
  rate-constants:
    - {P: 0.1 atm, A: 3.36e+22, b: -3.33, Ea: 3950.0}
    - {P: 1.0 atm, A: 4.8e+23, b: -3.63, Ea: 5220.0}
    - {P: 3.0 atm, A: 6.85e+23, b: -3.66, Ea: 5820.0}
    - {P: 10.0 atm, A: 1.53e+24, b: -3.73, Ea: 6820.0}
    - {P: 30.0 atm, A: 8.75e+23, b: -3.62, Ea: 7655.0}
    - {P: 100.0 atm, A: 1.27e+22, b: -3.06, Ea: 7950.0},
```

which requires seven boolean conditionals to be checked,  $p \leq 0.1$ ,  $0.1 < p \leq 1.0$ ,  $1.0 < p \leq 3.0$ ,  $3.0 < p \leq 10.0$ ,  $10.0 < p \leq 30.0$ ,  $30.0 < p \leq 100.0$ ,  $p > 100.0$ , where  $p$  is in atm. ChemGen utilizes an optimization where  $\ln p$  is pre-calculated before traversing over the booleans, where its value is checked against pre-calculated  $\ln p_i$  values.

### 3. Time integration strategies

The stability of the CFD time integration may be compromised by explicitly applying the source term of Equation (2). Different approaches can be utilized by time-splitting the evolution of the chemical state and retain the stability of the fluid dynamic algorithms [38, 39, 13, 15]. These approaches typically do not involve the flux from either convection or diffusion. By removing any spatial fluxes, the conservation Equation (1) reduces to an ordinary differential equation,

$$\frac{dy_f}{dt} = \mathcal{S}(y_f). \quad (37)$$

By removing the spatial fluxes, the fluid state from Equation (2) is reduced to

$$y_f = (0, \dots, 0, \rho u, C_1, \dots, C_{n_s}) \quad (38)$$

and by inspecting the source term,

$$\mathcal{S}(y_f) = (0, \dots, 0, 0, \omega_1, \dots, \omega_{n_s}), \quad (39)$$

and analyzing the time rate of change of the internal energy,

$$\frac{\partial \rho u}{\partial t} = \frac{\partial \rho u}{\partial T} \frac{\partial T}{\partial t} + \sum_{i=1}^{n_s} \frac{\partial \rho u}{\partial C_i} \frac{\partial C_i}{\partial t} = 0, \quad (40)$$

the temperature source term is realized as

$$\frac{\partial T}{\partial t} = \omega_T = - \frac{\sum_{i=1}^{n_s} W_i u_i \omega_i}{c_v}. \quad (41)$$

This yields the chemical evolution system solved by ChemGen

$$\frac{dy_c}{dt} = S(y_c), \quad y_c = (T, C_1, \dots, C_{n_s}), \quad S(y_c) = (\omega_T, \omega_1, \dots, \omega_{n_s}). \quad (42)$$

For brevity we now refer to  $y$  as the chemicals state,  $y = y_c$ . In this section we will cover some techniques implemented in ChemGen to integrate this state in time. Isolating the source term as a system of ODEs allows the stiffness of the problem, which comes from disparate time scales between slow and fast chemical reactions. However, these approaches can introduce errors due to time-splitting.

The formulation presented in Equation (42). is derived using a constant internal energy assumption and thus allows for pressure to change. Some systems require a constant pressure assumption. Appendix (Appendix B) covers how a constant pressure formulation can be extracted using an assumption that uses a global  $\rho h$ .

### 3.1. Explicit Runge-Kutta time integration method

ChemGen provides a Runge-Kutta 4th-order method (RK4) for integrating the chemical state from  $y^n$  to  $y^{n+1}$  over a time interval of  $\Delta t$  from Equation (42). This explicit time integration scheme approximates the solution by evaluating the source term at four intermediate stages within each time step [40]. Given state  $y_n$  at time  $t_n$ , the RK4 method updates to the next state,  $y^{n+1}$  at time  $t^{n+1}$ . The algorithm is

$$\begin{aligned}\Gamma_1 &= S(y^n) \\ \Gamma_2 &= S\left(y^n + \frac{\Delta t}{2}\Gamma_1\right) \\ \Gamma_3 &= S\left(y^n + \frac{\Delta t}{2}\Gamma_2\right) \\ \Gamma_4 &= S(y^n + \Delta t\Gamma_3) \\ y^{n+1} &= y^n + \frac{\Delta t}{6}(\Gamma_1 + 2\Gamma_2 + 2\Gamma_3 + \Gamma_4)\end{aligned}\tag{43}$$

This method achieves fourth-order accuracy in time, meaning the local truncation error scales as  $\mathcal{O}(\Delta t^5)$  and the global error as  $\mathcal{O}(\Delta t^4)$ . RK4 is effective for non-stiff problems where high accuracy is desired without the overhead of the implicit solvers, which, as detailed below, require potentially expensive linear solvers.

### 3.2. Implicit time integration and Jacobians

Although explicit methods like RK4 offer high accuracy, they can suffer from stability issues, particularly for stiff problems. In contrast, implicit time integration methods provide improved stability at the cost of increased computational overhead. Consider first the backward Euler implicit time integration method, which seeks  $y^{n+1}$  that satisfies

$$\frac{y^{n+1} - y^n}{\Delta t} = S(y^{n+1}).\tag{44}$$

Newton's method can be used to find  $y^{n+1}$  via iteration by solving

$$\mathcal{G}(y_k^{n+1})(y_{k+1}^{n+1} - y_k^{n+1}) = -f(y_k^{n+1}), \quad f(y_k^{n+1}) = \frac{y^{n+1} - y^n}{\Delta t} - S(y^{n+1}),\tag{45}$$

for  $y_{k+1}^{n+1}$ , the state at the  $(k+1)$ th Newton iteration.  $\mathcal{G}(y_k^{n+1})$  is the Jacobian for the time integration given by

$$\mathcal{G}(y_k^{n+1})_{ij} = \frac{\partial f(y_{k,i}^{n+1})}{\partial y_j} = \mathcal{D}_{ij} - \mathcal{J}_{ij}, \quad \mathcal{J}_{ij} = \frac{\partial S_i(y_k^{n+1})}{\partial y_j},\tag{46}$$

where  $\mathcal{D}_{ij}$  depends on the time integration scheme ( $\mathcal{D}_{ij} = \delta_{ij}/\Delta t$  for the backward Euler method) and  $\mathcal{J}_{ij}$  is the Jacobian of the chemical source term with  $y_j$  as the  $j$ th element of  $y$  and  $S_i$  is the  $i$ th element of  $S$ .

ChemGen employs analytical differentiation to evaluate the chemical source term Jacobian,  $\mathcal{J}_{ij}$ . The code generation is optimized to compute only necessary components for efficient matrix assimilation. The elements of the species concentration source term Jacobian are

$$\frac{\partial \omega_k(y)}{\partial C_j} = \frac{\partial \omega_k}{\partial C_j} + \frac{\partial \omega_k}{\partial T} \frac{\partial T}{\partial C_j},\tag{47}$$

Table 2: Implicit time integration strategies provided by ChemGen.  $n_{\max}$  is the maximum number of Newton iterations.

Strategy	Order	Stages	# Linear Solves	Section/Appendix
SDIRK-2	2nd	2	up to $n_{\max}$ per stage	(Appendix A.1)
SDIRK-4	4th	5	up to $n_{\max}$ per stage	(Appendix A.2)
Rosenbrock	2nd	2	one per stage	(Appendix A.3)
YASS	1st	1	one	(Appendix A.4)
Backward Euler	1st	1	up to $n_{\max}$	(3.2)

where

$$\frac{\partial T}{\partial C_j} = - \left( \frac{\partial \rho u}{\partial T} \right)^{-1} \frac{\partial \rho u}{\partial C_j} = - \frac{1}{\rho c_v} W_j u_j = - \frac{W_j u_j}{\rho c_v}. \quad (48)$$

With these derivatives known an  $n_s + 1$  by  $n_s + 1$  matrix is formed,

$$\mathcal{J}_{ij} = \begin{cases} \frac{\partial \omega_T}{\partial T} & j = 1 \\ \frac{\partial \omega_T}{\partial C_{j-1}} & i = 1; \quad j > 1 \\ \frac{\partial \omega_{i-1}}{\partial T} & i > 1; \quad j = 1 \\ \frac{\partial \omega_{i-1}}{\partial C_{j-1}} & i > 1; j > 1 \end{cases}, \quad (49)$$

where

$$\frac{\partial \omega_T}{\partial T} = - \frac{\sum_{i=1}^{n_s} W_i \frac{\partial u_i}{\partial T} \omega_i}{c_v} + \frac{\sum_{i=1}^{n_s} W_i \frac{\partial u_i}{\partial T} \omega_i}{c_v^2} \frac{\partial c_v}{\partial T} \quad (50)$$

and

$$\frac{\partial \omega_T}{\partial C_{j-1}} = - \frac{\sum_{i=1}^{n_s} W_i \frac{\partial u_i}{\partial T} \frac{\partial \omega_i}{\partial C_{j-1}}}{c_v} + \frac{\sum_{i=1}^{n_s} W_i u_i \omega_i}{c_v^2} \frac{\partial c_v}{\partial C_{j-1}}. \quad (51)$$

### 3.3. Additional time integrators

While the backward Euler time integration is inherently stable, it may exhibit poor accuracy or convergence for certain chemical models and time step sizes. Therefore, ChemGen offers additional implicit integrators, as listed in Table 2 with formulation details in Appendix (Appendix A). The appropriate choice of time integrator is problem dependent. For example, in low Mach number solvers where acoustic time scales are removed, chemical integration may proceed on the order of microseconds. In contrast, fully compressible solvers may require time steps as small as nanoseconds or even less, in which case a lower-order but computationally faster time integrator may be more suitable.

Other time integration strategies not yet implemented in ChemGen can be added using the same framework as the existing methods. These include, but are not limited to, multi-step implicit methods such as backward differentiation formulas, which utilize information from previous time steps, and higher-order Rosenbrock methods [41], which require additional stages. Integration methods such as DGODE [15] and CHEMEQ2 [42] may be more challenging to implement. DGODE is based on finite element numerics, while CHEMEQ2 requires the source term to be separated into distinct production and consumption components, introducing additional complexities not currently addressed by ChemGen.

### 3.4. Linear solvers

The aforementioned implicit time integrators require the solution to the linear system (45), recast here as

$$Ax = b. \quad (52)$$

It is well-established that employing a direct solver requires  $\mathcal{O}(n^3)$  operations for an  $n \times n$  square matrix [43]. Although ChemGen supplies a direct solver, to address the  $\mathcal{O}(n^3)$  cost, its default linear solver is the Generalized Minimal Residual (GMRES) method [44], which is implemented as a standalone function. GMRES is an iterative Krylov subspace method that builds an orthonormal basis of the Krylov subspace and minimizes the residual over this subspace at each iteration.

The number of GMRES iterations can be reduced by applying a preconditioner,  $M$ , as

$$M^{-1}Ax = M^{-1}b. \quad (53)$$

ChemGen provides two standard options: Jacobi preconditioning and Gauss-Seidel preconditioning. In the former,  $M$  simply consists of the diagonal of  $A$ , which can be efficiently inverted by computing the reciprocal of each diagonal entry. In the latter,  $M$  consists of the lower triangular portion of  $A$  (including the diagonal), which can be easily applied with forward substitution. For more advanced preconditioning techniques used in chemical kinetics, see, for example, [45].

ChemGen is not intended to serve as a comprehensive linear-solver library. For more advanced or scalable linear solver capabilities, ChemGen can be interfaced with external libraries such as Eigen [46] or PETSc [47], which offer efficient preconditioned Krylov solvers and direct sparse solvers for large systems. Additionally, while ChemGen includes several implicit time integration strategies (e.g., backward Euler, SDIRK methods), offloading the integration of stiff ODE systems to established libraries like CVODES [48] is also a reasonable approach. Such packages allow user-supplied source terms and Jacobians that can then be used in time integration strategies.

## 4. Code generation

### 4.1. The concept of decorators

C++ is a malleable language that allows developers to abstract underlying data structures through type aliases (`using scalar = float;`) and operator overloading. As a result, variables may appear to be simple scalars while in fact representing complex types with custom behavior. This flexibility enables expressive, domain-specific syntax, but this means that there is no single set of generated code that can be embedded into all existing C++ software.

For instance, it may be more suitable for the function

```
float sqr (float a){return a * a;} \\example function 1
```

to be declared as a `const` in a `struct` or `class`. In addition, the scalar `float` could also be `double`, and the user may wish to pass the function parameters by `const` reference, i.e., `float a → const float& a`. A function annotation may also be desirable, such as a preprocessor macro provided the Kokkos library [49]. This can turn the above example to

```
KOKKOS_INLINE_FUNCTION
float sqr (const float& a) const {return a * a;} \\example function 2
```

ChemGen utilizes *decorators* to assist in the flexibility of generating desirable C++ code. The above example can be changed to a Python formattable Python string as

```
{device_function}
{scalar_function} sqr ({scalar_parameter} a) {const_option} {{return a * a;}} \\Python formattable
```

ChemGen utilizes these decorators throughout the code generation process. When the code generation takes place, a `configuration.yaml` file is utilized that replaces the decorators with the appropriate text. So for example function 1 is

```
decorators:
  scalar_function: "float"
  scalar_parameter: "float"
  device_option: ""
  const_option: ""
```

and for example function 2 is

```
decorators:
    scalar_function: "float"
    scalar_parameter: "const float&"
    device_option: "KOKKOS_INLINE_FUNCTION"
    const_option: "const"
```

The configuration file also includes the ability to switch from `std::array` to `std::vector` and allows other data types such as `Views` from Kokkos or `std::tuple`-derived types.

#### 4.2. Derivatives

ChemGen employs a mechanical differentiation approach. It supplies a differential codebase that contains derivatives of common functions and their chain rules. For instance, the Arrhenius function as generated using the standard configuration in ChemGen is

```
double
arrhenius(const double& A, const double& B, const double& E, const double&
          temperature)
{
    double pow_term = pow_gen(temperature, B);
    double exp_term = exp_gen(divide(-E, universal_gas_constant() *
                                      temperature));

    return multiply(A,
                    multiply(pow_term,
                            exp_term));
}
```

ChemGen supplies the derivatives for the mathematical operations contained in the `arrhenius` function. For example, the derivatives of `pow_gen`, given that it has two parameters, would yield two derivatives and one chain rule,

```
double dpow_da(const double& a, const double& b)
{return b * std::pow(a, b - double(1));}
double dpow_db(const double& a, const double& b)
{return std::pow(a, b) * std::log(a);}

double
pow_gen_chain(const double& a,
              const double& a_perturbation,
              const double& b,
              const double& b_perturbation)
{
    return dpow_da(a, b) * a_perturbation + dpow_db(a, b) * b_perturbation;
}
```

For all existing lower-level functions, the partial derivatives with respect to each parameter, as well as the associated chain rule contractions with perturbations, can be assumed accessible when developing higher-level functions in ChemGen. With this capability, the derivative of the Arrhenius function with respect to temperature is immediately available,

```
double
darrhenius_dtemperature(const double& A,
                        const double& B,
                        const double& E,
                        const double& temperature)
```

```

    double pow_term = pow_gen(temperature, B);
    double dpow_term_dtemperature = dpow_da(temperature, B);

    double exp_term =
        exp_gen(divide(-E,
                        universal_gas_constant() * temperature));
    double dexp_term_dtemperature =
        exp_chain(divide(-E,
                          universal_gas_constant() * temperature),
                  ddivide_db(-E,
                              universal_gas_constant() * temperature));

    return multiply(A,
                    multiply_chain(pow_term,
                                   dpow_term_dtemperature,
                                   exp_term,
                                   dexp_term_dtemperature));
}

```

The mechanical differentiation approach is both efficient and more readable than symbolic differentiation using tools such as SymPy [50]. While symbolic differentiation can be valuable, the mechanical approach is better suited for organizing and maintaining ChemGen’s code generation process. Regardless, since the functions are available in C++, packages such as autodiff [51] can be implemented to use a wider variety of differentiable functionality. For a mathematical overview of the derivatives needed for chemically reacting flow, see [24].

#### 4.3. Generation process

In the ChemGen software, there are three main file types. The file type consists of Python files that manipulate chemistry data and existing static code to create compilation-ready source code. The second file type comprises static code files that are ingested by the aforementioned Python code and reformatted to the desired target software using decorators. The final type is a mixture of Python code used in tutorials, most of which are covered in Section (5), and `yaml` files that contain the chemical model data.

The code generation process in ChemGen is designed to be as streamlined as possible and is depicted in Figure 1. ChemGen relies on Cantera to interpret chemical model files. Cantera performs several key tasks: parsing the chemistry data, converting units to the International System of Units (SI), and providing thermodynamic relationships that are refit into the desired polynomial form. As a well-maintained and widely used tool in the combustion community, Cantera provides a dependable foundation for ChemGen. Nevertheless, ChemGen could also be extended to handle data parsing, unit conversion, and thermodynamic relationships for refitting internally.

After the Cantera step, ChemGen begins to generate source code. The state size is hardcoded with several automatically generated variables, `Species`, `ChemicalState`, `Jacobian`. The stoichiometric coefficients ( $\nu_{ij}$ ,  $\nu_{ij}^f$ , and  $\nu_{ij}^r$ ) and reaction coefficients, which depend on the reaction type, are then accrued. Each reaction rate is generated into `areactions.h` header file; for example, if the first reaction of a Cantera-based `yaml` file is

```
- equation: H + O2 <=> O + OH
rate-constant: {A: 1.0399e+11, b: 0.0, Ea: 6.405704e+07}
```

then the following code is generated with standard configuration settings,

```
double call_forward_reaction_0(const double& temperature,
                               const double& log_temperature)
{
    return arrhenius(103990000000.0, 64057040.0, temperature, log_temperature)
;
}
```

the units here are in SI. Once all reactions are generated, a chemical source term file is created. The source term file begins with generated code that calculates  $M_c$ ,  $\ln T$ , and  $p$ , as well as  $\Delta G'_j$  for all reactions, `gibbs_reactions`. ChemGen then loops through all reactions, and generates the forward rates and the corresponding equilibrium constants for reversible reactions. In the above example, a reverse reaction is necessary, so ChemGen generates

```
double forward_reaction_0 = call_forward_reaction_0(temperature, log_temperature);
double equilibrium_constant_0 = exp_gen(-gibbs_reactions[0]);
double rate_of_progress_0 =
species[0] * species[3] * forward_reaction_0
- species[2] * species[4] * forward_reaction_0/equilibrium_constant_0;
```

While many options can be specified during the generation process, the simplest approach requires only the following terminal command to generate source code using ChemGen:

```
chemgen.py mech.yaml .
```

This command, as long as `mech.yaml` is a Cantera supported file, generates C++ code in the current directory in a `src` folder using the standard settings.

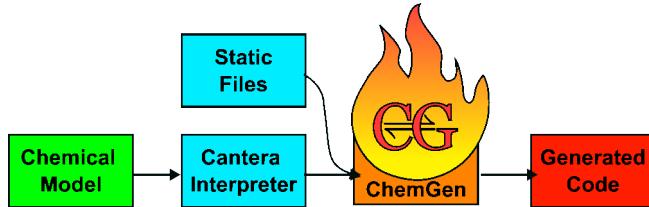


Figure 1: ChemGen flowchart

## 5. Results and demonstration cases

This section presents standalone test cases to confirm that the generated source code is behaving as intended. We then embed the generated code in `detonationEulerFoam` [52] based on the OpenFOAM package [53] to demonstrate its performance in a CFD simulation of a moving detonation wave.

### 5.1. Source term accuracy analysis

To test the source term accuracy, we randomly generate a chemical state as

$$\begin{aligned} T &= 1000 + r_T 1500 \text{ K} \\ p &= 0.1 + 9.9r_p \text{ atm} \\ X_i &= r_i, \quad \text{for } i = 1, \dots, n_s, \end{aligned} \tag{54}$$

where  $r_T, r_p, r_1, \dots, r_{n_s} \sim \mathcal{U}(0, 1)$ , which denotes a uniform distribution.  $X_i$  is the mole fraction of the  $i$ th species, defined as

$$X_i = \frac{C_i}{\sum_{i=1}^{n_s} C_i}, \tag{55}$$

which is renormalized after all  $n_s$  species are randomly generated. We then calculate the concentrations as  $C_i = X_i p / (RT)$ , forming the complete chemical state for ChemGen,  $y_c$ . We consider five different chemical models with increasing size that are common to simulations of chemically reacting flows and summarized in Table 3. For this analysis we generated 10,000 random chemical states generated for each model.

We use Cantera's source term prediction, `net_production_rates`, for comparison in this analysis. The average relative error per chemical state is

Model	$n_s$	$n_r$	Reference
Ó Connaire	10	40	[54]
Burke	13	27	[55]
GRI-Mech 3.0	53	325	[56]
UCSD	63	268	[57]
FFCM-2	96	1054	[20]

Table 3: Summary of chemical models analyzed

$$\epsilon = \frac{1}{n_s} \sum_{i=1}^{n_s} \left( \begin{cases} \left| \frac{S_{cg,i} - S_{ct,i}}{S_{ct,i}} \right| & \text{if } |S_{ct,i}| > 10^{-10} \\ 0 & \text{otherwise} \end{cases} \right), \quad (56)$$

where  $S_{ct,i}$  denotes the Cantera-predicted source term for the  $i$ th species and  $S_{cg,i}$  denotes the corresponding ChemGen-predicted source term. If the magnitude of the Cantera source term is less than  $10^{-10}$ , the error is considered to be zero.

The distribution of errors for the randomly generated states and the chemical models analyzed are shown in Figure (2). Agreement with Cantera predictions varies with model complexity. Among the models tested, the UCSD model exhibits the largest discrepancy. While the ChemGen implementation is designed to be consistent with Cantera’s formulation, differences arise primarily due to the thermodynamic polynomial representations. Cantera supports multiple polynomial formats for thermodynamic property calculations, whereas ChemGen refits all species thermodynamic properties to a user-specified polynomial order during code generation. The ChemGen refits for the data in Figure (2) are based on  $n_p = 7$ . To investigate the effect of polynomial degree, we repeat the error analysis with different choices of  $n_p$  which confirms an inverse relationship between the mean relative error and the polynomial degree. The details from this further analysis are shown in Appendix (Appendix C.1).

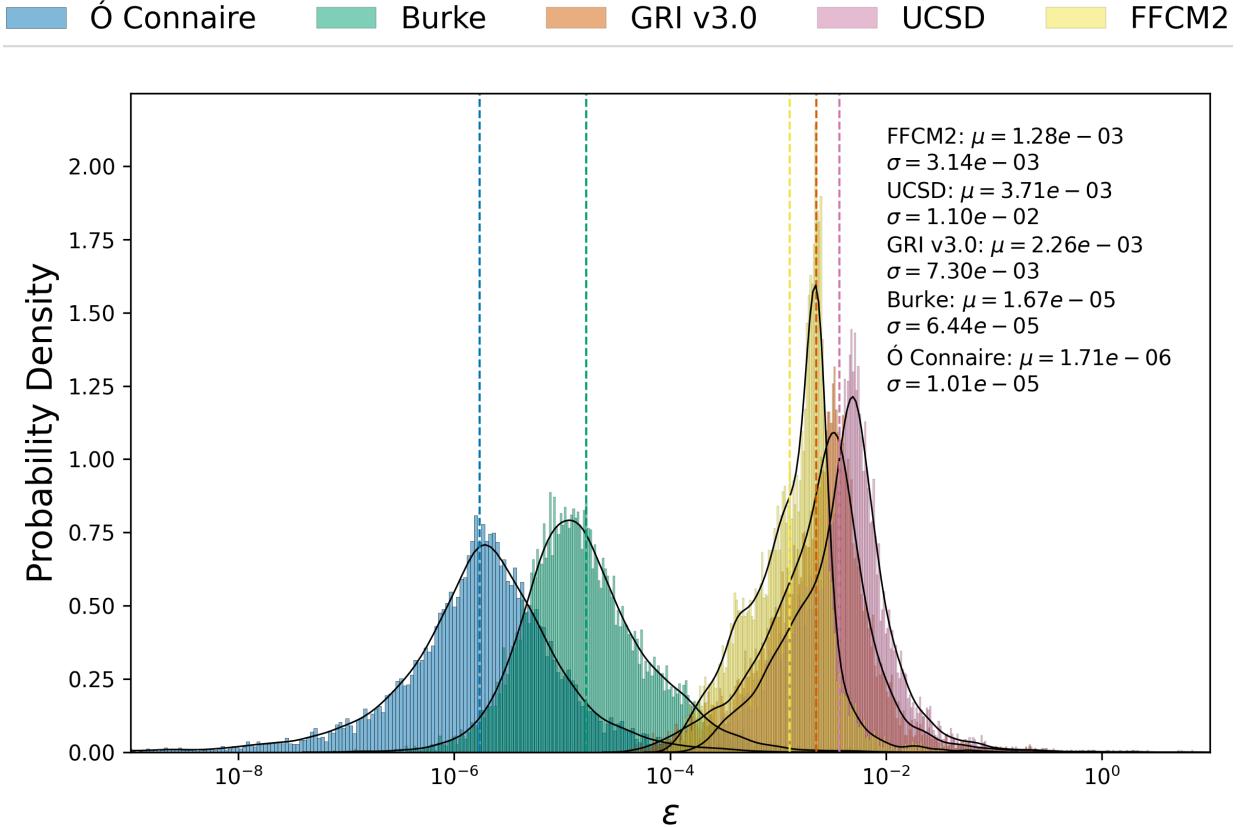


Figure 2: Distribution of error,  $\epsilon$ , as defined in Equation (56), for ChemGen source calculations in comparison to Cantera source calculations for five tested chemical models from Table 3. The vertical dash line represents the mean of the distribution and is colored according to the corresponding model. The distribution mean,  $\mu$ , and corresponding standard deviation,  $\sigma$ , are also reported as text.

### 5.2. Explicit time integration

With confidence in the source term calculation, we proceed to apply ChemGen’s explicit RK4 time integrator to temporally evolve the following initial chemical state:

$$\begin{aligned} T &= 1800 \\ p &= 101325.0 \\ (X_{H_2}, X_{O_2}, X_{N_2}) &= (0.2, 0.2, 0.6). \end{aligned} \quad (57)$$

We use a hydrogen submodel extracted from FFCM-2 [20], which is created by removing carbon containing species, electronically excited states, such as  $OH^*$ , and inert species  $Ar$  and  $He$ . The resulting submodel consists of 9 species and 25 reactions. We calculate the concentrations,  $C_i$ , in the same manner as in Section (5.1). We employ a time-step size of  $\Delta t = 1 \times 10^{-8}$  seconds. With larger time-step sizes, the RK4 time integrator becomes unstable. The resulting temperature and species concentration profiles are shown in Figure (3) , where the ChemGen results are represented by dashed black lines and Cantera results are represented by solid colored lines. We produce the Cantera results using the ReactorNet class.

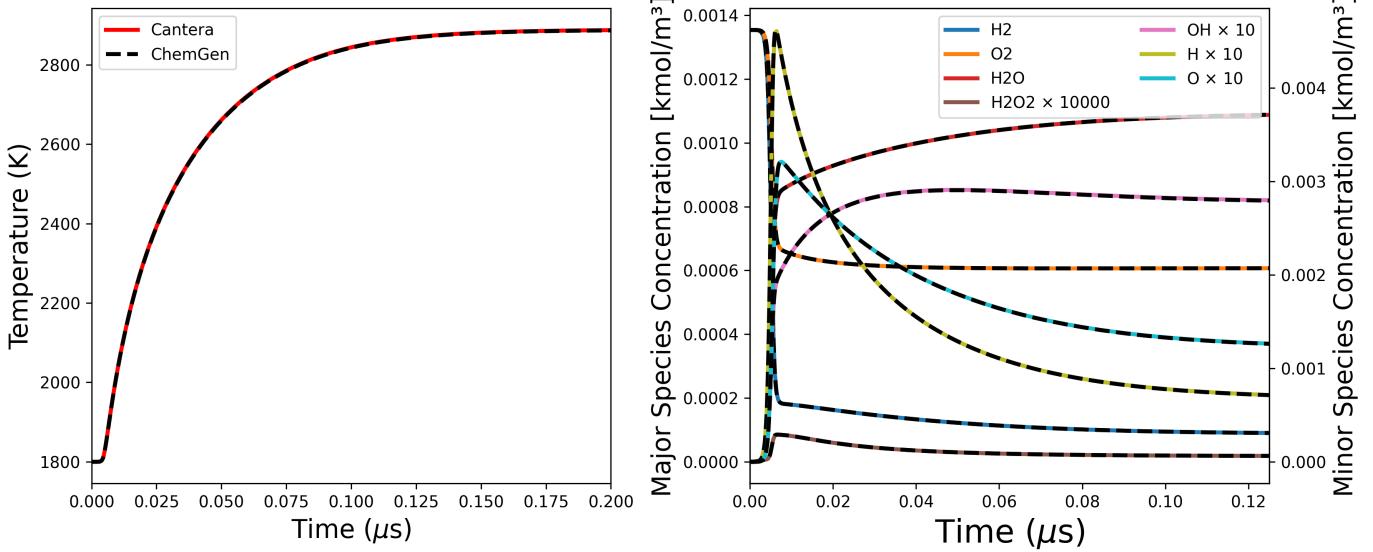


Figure 3: Temperature (left) and species (right) evolution for a constant volume reactor using an explicit RK4 scheme and initial conditions given in Equation (57). These results are produced using a hydrogen submodel extracted from FFCM2 [20]. The solid lines represent the Cantera solution, and the dashed black lines represent the ChemGen solution.

Following a brief induction period, the chemical state undergoes a sudden temperature rise accompanied by rapid species changes, where the reactants,  $H_2$  and  $O_2$ , are consumed and the product,  $H_2O$ , along with intermediate species,  $H_2O_2$ ,  $OH$ ,  $O$ , and  $H$ , are created. This is followed by a more gradual evolution toward a steady state. The main reactants and products are referred to as major species, which for this model are  $H_2$ ,  $O_2$ , and  $H_2O$ . The intermediate and small remnant species are referred to as minor species and usually exist in trace quantities ( $X_i < 1 \times 10^{-3}$ ). Therefore, scaling multipliers are applied to minor species in Figure (3) in order to easily visualize their profiles on the same graph.

### 5.3. Jacobian verification

Next, we verify the Jacobian calculations. Recall the Jacobian supplied by ChemGen is

$$\mathcal{J}_{ij} = \begin{cases} \frac{\partial \omega_T}{\partial T} & j = 1 \\ \frac{\partial \omega_T}{\partial C_{j-1}} & i = 1; \quad j > 1 \\ \frac{\partial \omega_{i-1}}{\partial T} & i > 1; \quad j = 1 \\ \frac{\partial \omega_{i-1}}{\partial C_{j-1}} & i > 1; j > 1 \end{cases}. \quad (58)$$

In this section we present verification of the Jacobians using the finite difference method. To calculate the finite difference approximation of the Jacobians, we utilize the second order finite difference approximation for each element by perturbing each species forward,

$$C_j^+ = (C_1 + \delta_{1j}\delta C, \dots, C_{n_s} + \delta_{n_sj}\delta C), \quad (59)$$

and backwards,

$$C_j^- = (C_1 - \delta_{1j}\delta C, \dots, C_{n_s} - \delta_{n_sj}\delta C). \quad (60)$$

These perturbations are utilized in the finite difference Jacobian,

$$\hat{\mathcal{J}}_{ij} = \begin{cases} \frac{\omega_T(C,T+\delta T) - \omega_T(C,T-\delta T)}{2\delta T} & j = 1 \\ \frac{\omega_T(C_{j-1}^+, T) - \omega_T(C_{j-1}^-, T)}{2\delta C} & i = 1; \quad j > 1 \\ \frac{\omega_{i-1}(C,T+\delta T) - \omega_{i-1}(C,T-\delta T)}{2\delta T} & i > 1; \quad j = 1 \\ \frac{\omega_{i-1}(C_{j-1}^+, T) - \omega_{i-1}(C_{j-1}^-, T)}{2\delta C} & i > 1; \quad j > 1, \end{cases} \quad (61)$$

where  $\delta_{ij}$  is from Equation (46),  $\delta T$  and  $\delta C$  are the temperature and concentration perturbations, respectively, and  $(\rho u, C) \rightarrow T$  is the Newton solve for the temperature. An initial test to verify the Jacobian is to check the order of convergence of the finite difference errors as compared to the analytical Jacobian provided by ChemGen.

We use the  $L_2$ -norm to assess the error of the finite difference approximation,

$$L_2 = \|\mathcal{J}_{ij} - \hat{\mathcal{J}}_{ij}\| = \sqrt{\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} (\mathcal{J}(C_i, T)_{ij} - \hat{\mathcal{J}}(C_i, T)_{ij})^2}, \quad (62)$$

for four refinement levels using consecutively halved values of  $\delta C$  and  $\delta T$ . The initial  $\delta C$  is chosen as a quarter of the smallest non-zero concentration, and the initial temperature perturbation is chosen as  $\delta T = 100$  K. Figure (4) shows the order of convergence for a single random chemical state determined using the UCSD chemical model [57]. The refinement results agree well with the expected second-order accuracy. We repeat this refinement study across all previously mentioned chemical models and observed consistent order of accuracy for both Jacobians (not shown for brevity). In some cases, the initial  $\delta C$  is too small, leading to errors approaching machine precision and preventing further reduction in the  $L_2$ -norm. By adjusting the initial  $\delta C$ , the expected order of accuracy was recovered.

Additionally, we evaluate the computational cost of the analytical Jacobian relative to their corresponding finite difference approximations. Using the same approach to generate random chemical states as described in Equation (54), we compute 1,000 Jacobians for each method and compared the total execution time. For the discrete Jacobians, we use first-order finite difference schemes to minimize computational cost,

$$\hat{\mathcal{J}}_{ij,1} = \begin{cases} \frac{\omega_T(C,T+\delta T) - \omega_T(C,T)}{\delta T} & j = 1 \\ \frac{\omega_T(C_{j-1}^+, T) - \omega_T(C,T)}{\delta C} & i = 1; \quad j > 1 \\ \frac{\omega_{i-1}(C,T+\delta T) - \omega_{i-1}(C,T)}{\delta T} & i > 1; \quad j = 1 \\ \frac{\omega_{i-1}(C_{j-1}^+, T) - \omega_{i-1}(C,T)}{\delta C} & i > 1; \quad j > 1. \end{cases} \quad (63)$$

The results are summarized in Table 4 for a variety of chemical models ranging in size and complexity. Despite using the fastest discrete Jacobian optimization, the analytical Jacobians are consistently faster across all cases. As the chemical mechanism increases in complexity, the reduction in cost of computing the analytical Jacobian relative to the finite different approximation becomes greater.

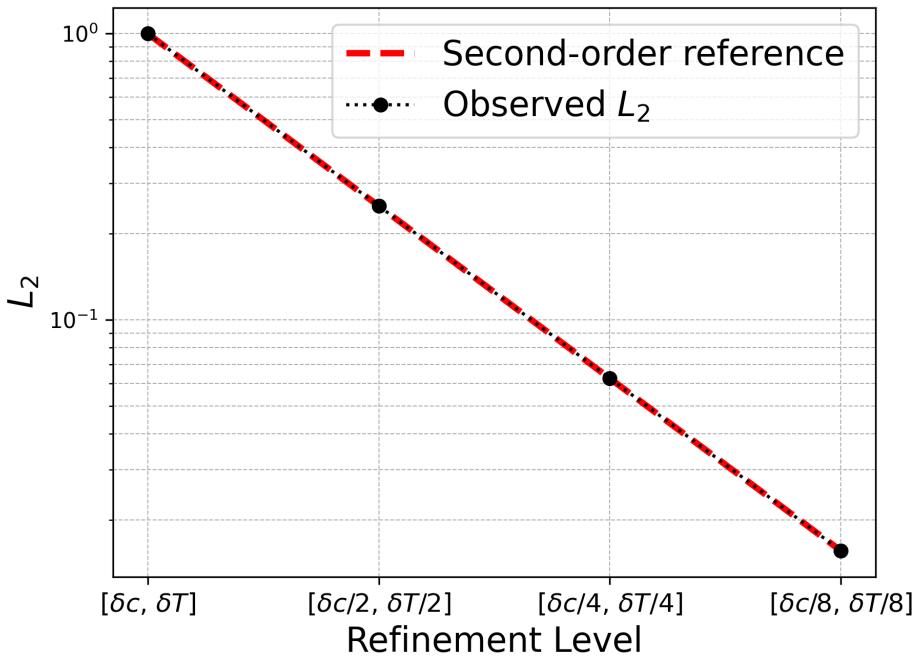


Figure 4: Order of accuracy convergence for  $\hat{\mathcal{J}}_{ij}$  when compared to  $\mathcal{J}_{ij}$ .

Table 4: Comparison of Jacobian Timings. All Jacobians are normalized by their respective finite difference Jacobian

	FFCM-2 (H2)	Ó Connaire	Burke	GRI-Mech 3.0	UCSD	FFCM-2 (Full)
$\mathcal{J}_{ij}/\hat{\mathcal{J}}_{ij,1}$	2.84	3.23	3.03	9.52	9.61	19.28

Finally, Niemeyer et al. [24] reported error metrics that can be used to verify Jacobian accuracy. These metrics are utilized in Appendix (Appendix C.2) to provide additional confidence in the correctness of ChemGen’s Jacobian computations.

#### 5.4. Implicit time integration

This section presents results obtained with the implicit time integrators provided by ChemGen, motivated by the confirmed accuracy of the source term from Section 5.1, successful explicit time integration from Section 5.2, and confidence in the computed source term Jacobian from Section 5.3. We utilize the FFCM-2 chemical model [20] to integrate a chemical state with initial conditions

$$\begin{aligned}
 T &= 1800 \text{ K} \\
 p &= 101325 \text{ Pa} \\
 (X_{O_2}, X_{N_2}, X_{H_2}, X_{C_2H_4}) &= (0.2, 0.4, 0.2, 0.2)
 \end{aligned} \tag{64}$$

Figure 5 compares the performance of ChemGen’s implicit time integrators against Cantera’s ReactorNet class. Among the integrators, the first-order methods, YASS and Backward Euler, exhibit the largest discrepancies with Cantera’s results. The Rosenbrock method shows improved agreement, while SDIRK-2 and SDIRK-4 integrators match Cantera most closely.

Integrator	$\Delta t$	$\tau_{cg}/\tau_{ct}$
Backward Euler	$10^{-7}$ s	1.26
SDIRK-2	$7 \times 10^{-7}$ s	3.19
SDIRK-4	$10^{-6}$ s	3.51
Rosenbroc	$10^{-7}$ s	2.26
YASS	$10^{-7}$ s	3.22
Cantera	$10^{-6}$ s	n/a

Table 5: Implicit time integrator optimal time step and simulation time, normalized by the Cantera simulation time

The time step sizes are chosen (via trial and error) to yield the fastest time-to-solution while maintaining reasonable accuracy. These are summarized in Table (5) with the simulation time,  $\tau_{cg}$ , of the ChemGen simulation normalized by the Cantera simulation time,  $\tau_{ct}$ . The Cantera ReactorNet time step is iterated on to find the fastest time to solution, while utilizing both GMRES and the default adaptive preconditioner.

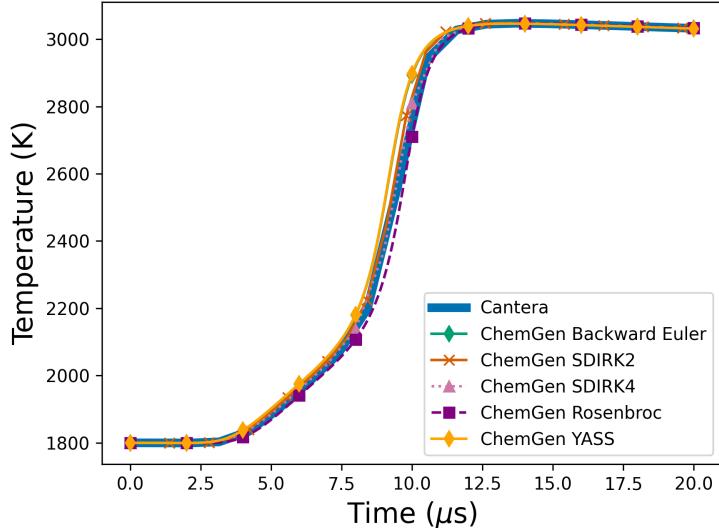


Figure 5: Homogeneous reactor results obtained with ChemGen’s implicit time integrators

To assure the correct implementation of these implicit methods beyond comparisons to Cantera, we assess order of convergence of the YASS, SDIRK-4, and Rosenbrock methods. The implementation of SDIRK-2 is similar to SDIRK-4 and Backward Euler is YASS with more linear solver iterations. Therefore, we omit them from the discussion here. We employ the extracted hydrogen submodel from FFCM-2 in a homogeneous reactor simulation with initial conditions

$$\begin{aligned} T &= 1800 \text{ K} \\ p &= 101325 \text{ Pa} \\ (X_{O_2}, X_{H_2}, X_{N_2}) &= (0.2, 0.2, 0.6). \end{aligned} \quad (65)$$

We utilize the state at time  $t = 4 \mu\text{s}$ , which is the approximate location of maximum  $\frac{\partial T}{\partial t}$ . This location in time represents the approximate location in the solution where the chemical state changes the most. We select  $\Delta t = 2 \times 10^{-6}$  for SDIRK-4 and  $\Delta t = 1 \times 10^{-7}$  for YASS and Rosenbroc. We perform five

refinements that change the integration time step,  $h$ , by a factor of 2,  $h = \Delta t/2^r$  for  $r = 0 \dots 4$ . We use the finest solution, corresponding to  $h = \Delta t/16$ , as the reference solution when computing the error,

$$\epsilon_h = \sum_{i=1}^{n_s} \left( C_{h,i} - C_{\frac{\Delta t}{16},i} \right)^2. \quad (66)$$

The results of the convergence study are summarized in Figure 6. The dashed lines represent the theoretical convergence rates. In all three cases, the expected convergence rates are recovered.

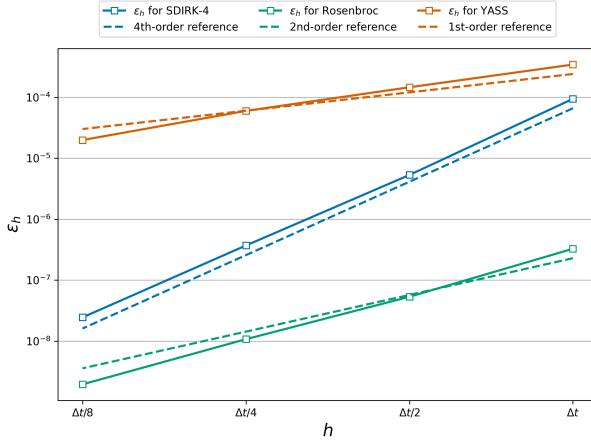


Figure 6: Convergence study for YASS, Rosenbroc, and SDIRK-4 using a hydrogen submodel extracted from FFCM-2 with initial conditions from Equation 65. The error is evaluated at  $t = 4 \mu s$ .

### 5.5. ChemGen in a detonation simulation

To demonstrate the usage of ChemGen in an existing computational framework, we target the chemistry evaluation in detonationFoam [52] which is based on the OpenFOAM package [53]. In particular, we focus on the detonationEulerFoam solver, which does not include viscous effects. The solver treats each individual species conservation equation via a segregated approach by solving

$$\frac{\partial \rho Y_i}{\partial t} + \nabla_{FVM} \cdot \mathcal{F}(\rho Y_i) = \mathcal{R}_i, \quad (67)$$

where  $\nabla_{FVM} \cdot \mathcal{F}(\rho Y_i)$  denotes the finite volume discretization of the species flux and  $\mathcal{R}_i$  is an applied source term of the  $i$ th species. For the complete set of conservation equations and additional details on the computational strategy, see [52]. Here the source term is calculated as

$$\mathcal{R} = \left( \frac{C_i^{n+1} - C_i^n}{\Delta t} \right) W_i \quad (68)$$

and is held constant during each species equation solve. In Equation (68)  $W_i$  is the molecular weight,  $C_i^n$  are the current concentrations calculated from the current thermodynamic state,  $\Delta t$  is the desired fluid time step for integrating Equation 67, and  $C_i^{n+1}$  are the predicted concentrations from solving the species evolution ODEs,  $\frac{\partial y_c}{\partial t} = S(y_c)$ . Therefore,  $\mathcal{R}_i$  is a linearized production term estimated from solving the detailed kinetics.

In detonationFoam [52], the seulex ODE solver [40] supplied by OpenFOAM is used to solve for  $C_i^{n+1}$ . The seulex solver utilizes an extrapolation algorithm based on the implicit Euler method with step size control and order selection. Due to the small time step size  $\Delta t = 1 \times 10^{-9}$  s we replace the seulex solver with ChemGen's YASS integrator combined with Jacobi-preconditioned GMRES.

We consider a two-dimensional detonation test that has been simulated in many previous studies (see Oran et al. [58], Houim and Kuo [39], Lv and Ihme [13], Johnson and Kercher [15], and Ching et al. [17]).

The domain here, which differs slightly from those in the aforementioned works, is 0.5 m long ( $x_1$ -direction) with a channel height of 0.06 m ( $x_2$ -direction). The initial mesh is a uniform quadrilateral mesh with 2000 cells in the  $x_1$ -direction and 240 cells in  $x_2$ -direction. We use the AMR default settings from [52], resulting in three possible refinements, with a minimum mesh size of  $31 \times 10^{-6}$  m. We use the slip wall condition for all boundaries. We again utilize the hydrogen submodel extracted from FFCM-2 [20]. The initial conditions are based on a steady detonation solution from the Shock & Detonation Toolbox [59], where the pre-shock state is given by

$$\begin{aligned} v_1 &= 0 \text{ m/s}, \\ X_{Ar} : X_{O_2} : X_{H_2} &= 7 : 1 : 2 \\ p &= 6670 \text{ Pa} \\ T &= 298 \text{ K}. \end{aligned} \quad (69)$$

In addition, we perturb the detonation front such that it varies in the  $x_2$ -direction. In Sun et al. [52] and in this work a sinusoidal profile is specified to perturb the front location, giving rise to instabilities that then transition to the expected solution.

The initial front progresses into the unreacted flow region, and the irregular structures from the initialization cause the development of two-dimensional fluid dynamic features eventually leading to an established two-dimensional cellular structure. These structures were visualized using a maximum-pressure history,  $p^{*,n+1}(x) = \max \{p^{n+1}(x), p^{*,n}(x)\}$  shown in Figure 7 at  $t = 200 \mu\text{s}$ . These results were calculated by detonationEulerFoam using ChemGen for the chemistry and demonstrate that ChemGen yields the expected cellular detonation structures.

Figures 8a and 8b show the distribution of temperature and hydroxyl radical mass fraction,  $Y_{OH}$ , at  $t = 200 \mu\text{s}$ . These results are consistent with those produced using the existing seulex ODE integrator, which are not reported here for brevity. Figures 9b and 9b show the production term for O-atoms,  $\mathcal{R}_O = \left( \frac{C_O^{n+1} - C_O^n}{\Delta t} \right) W_O$ , at  $t = 200 \mu\text{s}$  for detonationEulerFoam. The OpenFOAM result of  $\mathcal{R}_O$  using the seulex ODE solver is denoted as  $R_{of}$  and the result using the ChemGen YASS ODE solver is denoted as  $R_{cg}$ . These results demonstrate that an accurate CFD solution can be obtained using ChemGen in the detonationEulerFoam solver. Furthermore, the YASS solver provided by ChemGen is found to be on average 8 times faster than the seulex solver.

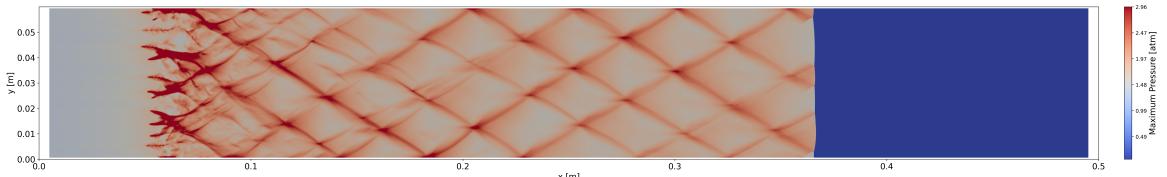


Figure 7: Maximum pressure,  $p^{*,n+1}(x) = \max \{p^{n+1}(x), p^{*,n}(x)\}$ , at  $t = 200 \mu\text{s}$ .

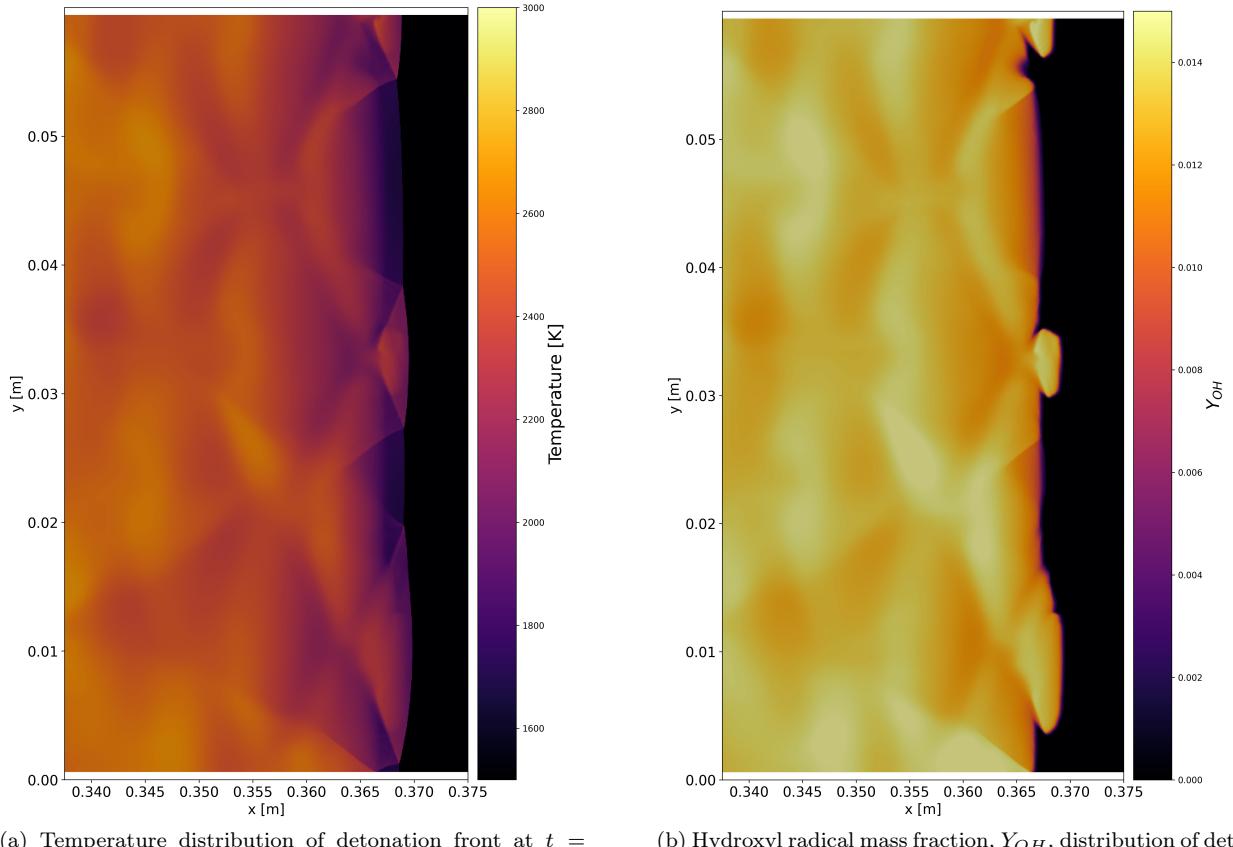


Figure 8: Temperature (left) and hydroxyl radical mass fraction (right) distribution for two dimensional  $\text{Ar}/\text{H}_2/\text{O}_2$  detonation using ChemGen to fully replace the chemistry solve in detonationEulerFoam.

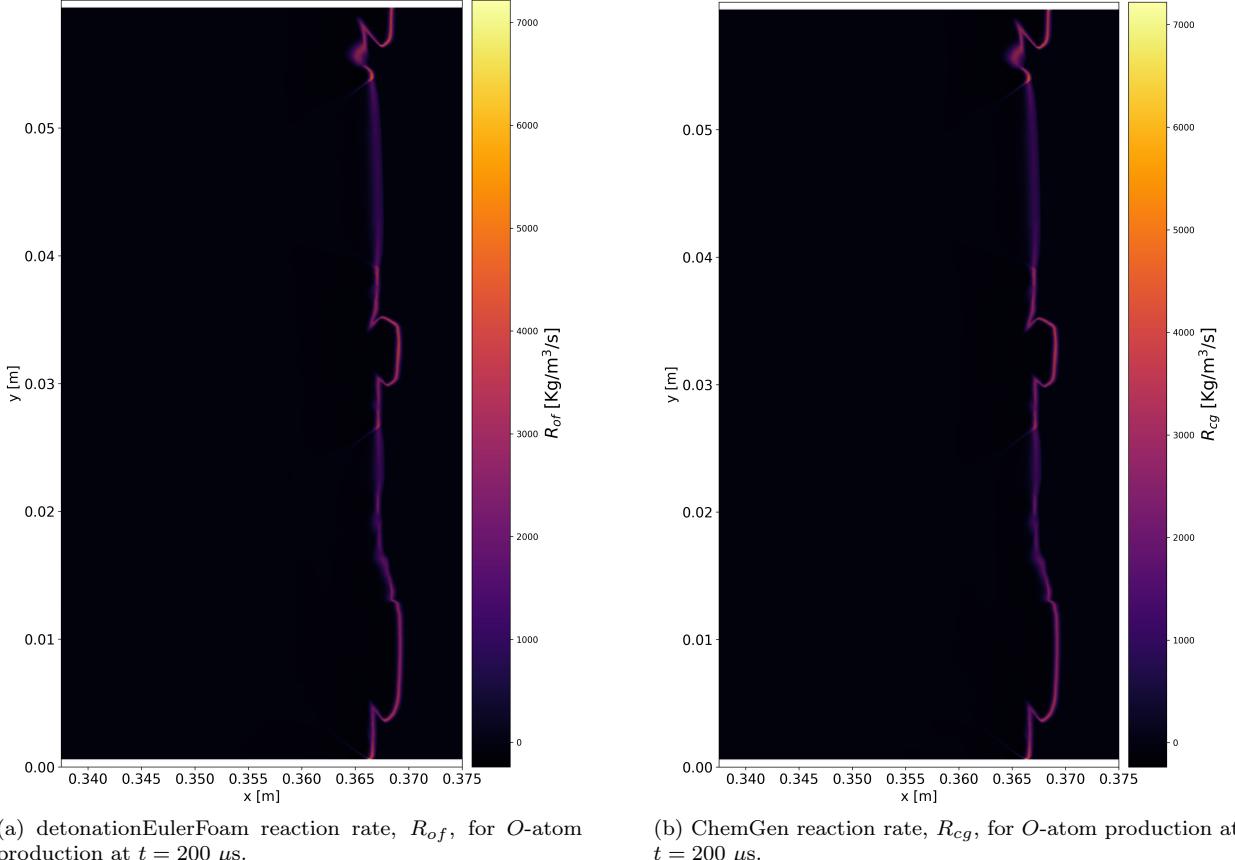


Figure 9: The production term for  $O$ -atoms,  $\mathcal{R} = \left( \frac{C_O^{n+1} - C_O^n}{\Delta t} \right) W_i$ , at  $t = 200 \mu\text{s}$  for a detonationEulerFoam simulation of a two dimensional detonation wave using OpenFOAM seualex ODE solver (left with subscript  $of$ ) and for ChemGen using the YASS ODE solver (right with subscript  $cg$ ).

## 6. Conclusion

In this work, we introduced ChemGen, a software package that leverages code generation to integrate multispecies thermodynamics and chemical kinetics into C++-based computational physics software, with an emphasis on CFD applications. ChemGen employs a decorator-based approach to enable flexible and modular C++ code generation tailored to existing simulation ecosystems. It generates both thermodynamic property approximations and source term expressions, along with their analytical derivatives, facilitating Jacobian assembly for implicit time integration. Additionally, ChemGen includes a variety of time integrators, linear solvers, and preconditioners.

We validated the correctness of the generated code by comparing its results against established Cantera source terms and showed consistency across a range of reaction mechanisms commonly used in CFD. We further verified the accuracy of the Jacobian computations and demonstrated optimal convergence across several supplied implicit integration methods. Finally, we embedded ChemGen into an established CFD code to perform a detonation simulation and reproduced results consistent with those in the literature, illustrating its ability to integrate seamlessly into mature computational frameworks. All data, source code, and supplementary tutorials are publicly available.

ChemGen lays the groundwork for the consistent inclusion of chemical source terms in diverse computational environments. Its flexibility also opens the door for scalable solvers, reduced-order modeling strategies,

and accelerator-aware chemistry integration methods that can advance the frontier of high-fidelity chemical reacting flow simulation.

## References

- [1] J. Anderson, Computational Fluid Dynamics: The Basics with Applications, McGraw-Hill, 1995.
- [2] J. H. Ferziger, M. Perić, Computational Methods for Fluid Dynamics, Springer, 2002.
- [3] C. Hirsch, Numerical Computation of Internal and External Flows: The Fundamentals of Computational Fluid Dynamics, 2nd Edition, Butterworth-Heinemann, Oxford, 2007.
- [4] S. V. Patankar, Numerical Heat Transfer and Fluid Flow, Hemisphere Publishing Corporation, Washington, D.C., 1980.
- [5] H. K. Versteeg, W. Malalasekera, An Introduction to Computational Fluid Dynamics: The Finite Volume Method, 2nd Edition, Pearson Education, Harlow, England, 2007.
- [6] F. Moukalled, L. Mangani, M. Darwish, The Finite Volume Method in Computational Fluid Dynamics: An Advanced Introduction with OpenFOAM and Matlab, Springer, Cham, 2016.
- [7] R. J. LeVeque, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007.
- [8] J. Donea, A. Huerta, Finite Element Methods for Flow Problems, Wiley, Chichester, 2003.
- [9] O. C. Zienkiewicz, R. L. Taylor, The Finite Element Method for Fluid Dynamics, 6th Edition, Elsevier Butterworth-Heinemann, Oxford, 2005.
- [10] S. Karni, Multicomponent flow calculations by a consistent primitive algorithm, *Journal of Computational Physics* 112 (1) (1994) 31 – 43. doi:<https://doi.org/10.1006/jcph.1994.1080>.
- [11] R. Abgrall, How to prevent pressure oscillations in multicomponent flow calculations: A quasi conservative approach, *Journal of Computational Physics* 125 (1) (1996) 150 – 160. doi:<https://doi.org/10.1006/jcph.1996.0085>.
- [12] G. Billet, R. Abgrall, An adaptive shock-capturing algorithm for solving unsteady reactive flows, *Computers and Fluids* 32 (10) (2003) 1473 – 1495. doi:[https://doi.org/10.1016/S0045-7930\(03\)00004-5](https://doi.org/10.1016/S0045-7930(03)00004-5).
- [13] Y. Lv, M. Ihme, Discontinuous Galerkin method for multicomponent chemically reacting flows and combustion, *Journal of Computational Physics* 270 (2014) 105 – 137. doi:<https://doi.org/10.1016/j.jcp.2014.03.029>.
- [14] Y. Lv, M. Ihme, High-order discontinuous Galerkin method for applications to multicomponent and chemically reacting flows, *Acta Mechanica Sinica* 33 (3) (2017) 486–499.
- [15] R. F. Johnson, A. D. Kercher, A conservative discontinuous Galerkin discretization for the chemically reacting Navier-Stokes equations, *Journal of Computational Physics* 423 (2020) 109826. doi:[10.1016/j.jcp.2020.109826](https://doi.org/10.1016/j.jcp.2020.109826).
- [16] E. J. Ching, R. F. Johnson, A. D. Kercher, Positivity-preserving and entropy-bounded discontinuous Galerkin method for the chemically reacting, compressible Euler equations. Part I: The one-dimensional case, *Journal of Computational Physics* (2024) 112881.
- [17] E. J. Ching, R. F. Johnson, A. D. Kercher, Positivity-preserving and entropy-bounded discontinuous Galerkin method for the chemically reacting, compressible Euler equations. Part II: The multidimensional case, *Journal of Computational Physics* (2024) 112878.
- [18] T. Lu, C. K. Law, Toward accommodating realistic fuel chemistry in large-scale computations, *Progress in Energy and Combustion Science* 35 (2) (2009) 192–215. doi:<https://doi.org/10.1016/j.pecs.2008.10.002>. URL <https://www.sciencedirect.com/science/article/pii/S036012850800066X>
- [19] C. K. Westbrook, Chemical kinetics of hydrocarbon oxidation in gaseous detonations, *Combustion and Flame* 46 (1982) 191 – 210. doi:[https://doi.org/10.1016/0010-2180\(82\)90015-3](https://doi.org/10.1016/0010-2180(82)90015-3).
- [20] Y. Zhang, W. Dong, L. Vandewalle, R. Xu, G. Smith, H. Wang, Foundational Fuel Chemistry Model Version 2.0 (FFCM-2) (2023). URL <https://web.stanford.edu/group/haiwanglab/FFCM2>
- [21] R. Kee, F. Rupley, J. Miller, Chemkin-ii: A Fortran chemical kinetics package for the analysis of gas-phase chemical kinetics (9 1989).
- [22] D. G. Goodwin, R. L. Speth, H. K. Moffat, B. W. Weber, Cantera: an object-oriented software toolkit for chemical kinetics, thermodynamics, and transport processes, version 2.4.0 (2018). doi:[10.5281/zenodo.1174508](https://doi.org/10.5281/zenodo.1174508). URL <http://www.cantera.org>
- [23] J. B. Scoggins, V. Leroy, G. Bellas-Chatzigeorgis, B. Dias, T. E. Magin, Mutation++: Multicomponent thermodynamic and transport properties for ionized gases in c++, *SoftwareX* 12 (2020). doi:[10.1016/j.softx.2020.100575](https://doi.org/10.1016/j.softx.2020.100575).
- [24] K. E. Niemeyer, N. J. Curtis, C.-J. Sung, pyjac: Analytical jacobian generator for chemical kinetics, *Computer Physics Communications* 215 (2017) 188 – 203. doi:<https://doi.org/10.1016/j.cpc.2017.02.004>.
- [25] E. Cisneros, M. Anderson, Pyprometheus, <https://github.com/ecisneros8/PyPrometheus>, accessed: 2025-06-03 (2025).
- [26] B. A. Danciu, C. E. Frouzakis, Kinetix: A performance portable code generator for chemical kinetics and transport properties, *Computer Physics Communications* 310 (2025) 109504. doi:[10.1016/j.cpc.2025.109504](https://doi.org/10.1016/j.cpc.2025.109504). URL <https://dx.doi.org/10.1016/j.cpc.2025.109504>
- [27] V. Giovangigli, Multicomponent flow modeling, Birkhauser, Boston, 1999.
- [28] B. J. McBride, S. Gordon, M. A. Reno, Coefficients for calculating thermodynamic and transport properties of individual species (1993).
- [29] B. J. McBride, M. J. Zehe, S. Gordon, NASA Glenn coefficients for calculating thermodynamic properties of individual species (2002).

- [30] R. J. Kee, F. M. Rupley, E. Meeks, J. A. Miller, CHEMKIN-III: A FORTRAN chemical kinetics package for the analysis of gas-phase chemical and plasma kinetics, Tech. rep., Sandia National Labs., Livermore, CA (United States) (1996).
- [31] F. A. Lindemann, S. Arrhenius, I. Langmuir, N. Dhar, J. Perrin, W. M. Lewis, Discussion on “the radiation theory of chemical action”, *Transactions of the Faraday Society* 17 (1922) 598–606.
- [32] J. Troe, Theory of thermal unimolecular reactions at low pressures. I. Solutions of the master equation, *The Journal of Chemical Physics* 66 (11) (1977) 4745–4757. doi:[10.1063/1.433837](https://doi.org/10.1063/1.433837).  
URL <https://doi.org/10.1063/1.433837>
- [33] J. Troe, Predictive possibilities of unimolecular rate theory, *The Journal of Physical Chemistry* 83 (1) (1979) 114–126. doi:[10.1021/j100464a019](https://doi.org/10.1021/j100464a019).  
URL <https://doi.org/10.1021/j100464a019>
- [34] R. Gilbert, K. Luther, J. Troe, Theory of thermal unimolecular reactions in the fall-off range. II. Weak collision rate constants, *Berichte der Bunsengesellschaft für physikalische Chemie* 87 (2) (1983) 169–177.
- [35] J. Troe, Albert weller, Berichte der Bunsengesellschaft für physikalische Chemie 91 (4) (1987) 260–262. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/bbpc.19870910403>, doi:<https://doi.org/10.1002/bbpc.19870910403>.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bbpc.19870910403>
- [36] P. Stewart, C. Larson, D. Golden, Pressure and temperature dependence of reactions proceeding via a bound complex. 2. application to  $2\text{ch}_3 \rightarrow \text{c}_2\text{h}_5 + \text{h}$ , *Combustion and Flame* 75 (1) (1989) 25–31. doi:[https://doi.org/10.1016/0010-2180\(89\)90084-9](https://doi.org/10.1016/0010-2180(89)90084-9).  
URL <https://www.sciencedirect.com/science/article/pii/0010218089900849>
- [37] X. Gou, J. A. Miller, W. Sun, Y. Ju, Implementation of plog function in chemkin ii and iii, <https://engine.princeton.edu/model-reduction/>, accessed: 2022-04-25 (2011).
- [38] G. Strang, On the construction and comparison of difference schemes, *SIAM Journal on Numerical Analysis* 5 (3) (1968) 506–517.
- [39] R. Houim, K. Kuo, A low-dissipation and time-accurate method for compressible multi-component flow with variable specific heat ratios, *Journal of Computational Physics* 230 (23) (2011) 8527 – 8553. doi:<https://doi.org/10.1016/j.jcp.2011.07.031>.
- [40] E. Hairer, G. Wanner, Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, Vol. 14, 1996. doi:[10.1007/978-3-662-09947-6](https://doi.org/10.1007/978-3-662-09947-6).
- [41] H. Zhang, A. Sandu, Fatode: A library for forward, adjoint, and tangent linear integration of odes, *SIAM Journal on Scientific Computing* 36 (5) (2014) C504–C523. doi:[10.1137/130912335](https://doi.org/10.1137/130912335).
- [42] D. Mott, E. Oran, Chemeq2: A solver for the stiff ordinary differential equations of chemical kinetics (2001) 67.
- [43] G. H. Golub, C. F. V. Loan, *Matrix Computations*, 4th Edition, Johns Hopkins University Press, 2013.
- [44] Y. Saad, M. H. Schultz, Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM Journal on Scientific and Statistical Computing* 7 (3) (1986) 856–869. doi:[10.1137/0907058](https://doi.org/10.1137/0907058).
- [45] A. S. Walker, R. L. Speth, K. E. Niemeyer, Generalized preconditioning for accelerating simulations with large kinetic models, *Proceedings of the Combustion Institute* 39 (4) (2023) 5395–5403. doi:<https://doi.org/10.1016/j.proci.2022.07.256>.  
URL <https://www.sciencedirect.com/science/article/pii/S1540748922003273>
- [46] G. Guennebaud, B. Jacob, et al., Eigen v3, <https://eigen.tuxfamily.org> (2010).
- [47] S. Balay, W. D. Gropp, L. C. McInnes, B. F. Smith, Efficient management of parallelism in object oriented numerical software libraries, *Modern Software Tools in Scientific Computing* (1997).
- [48] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, C. S. Woodward, Sundials: Suite of nonlinear and differential/algebraic equation solvers, *ACM Transactions on Mathematical Software* 31 (3) (2005) 363–396. doi:[10.1145/1089014.1089020](https://doi.org/10.1145/1089014.1089020).
- [49] C. R. Trott, D. Sunderland, D. Copeland, H. C. Edwards, V. Zinzow-Kramer, N. Ellingwood, Kokkos 3: Programming model extensions for the exascale era, *IEEE Transactions on Parallel and Distributed Systems* 33 (4) (2022) 805–817. doi:[10.1109/TPDS.2021.3097283](https://doi.org/10.1109/TPDS.2021.3097283).
- [50] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, S. Rathnayake, F. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, Sympy: symbolic computing in python, *PeerJ Computer Science* 3 (2017) e103. doi:[10.7717/peerj-cs.103](https://doi.org/10.7717/peerj-cs.103).
- [51] A. M. M. Leal, autodiff, a modern, fast and expressive C++ library for automatic differentiation, <https://autodiff.github.io> (2018).  
URL <https://autodiff.github.io>
- [52] J. Sun, Y. Wang, B. Tian, Z. Chen, detonationfoam: An open-source solver for simulation of gaseous detonation based on openfoam, *Computer Physics Communications* 292 (2023) 108859. doi:<https://doi.org/10.1016/j.cpc.2023.108859>.
- [53] OpenFOAM Foundation Ltd., OpenFOAM – The OpenFOAM Foundation, version 6, retrieved from <https://github.com/OpenFOAM/OpenFOAM-6> (2025).  
URL <https://www.openfoam.org>
- [54] M. Ó Conaire, H. J. Curran, J. M. Simmie, W. J. Pitz, C. K. Westbrook, A comprehensive modeling study of hydrogen oxidation, *International Journal of Chemical Kinetics* 36 (11) (2004) 603–622. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/kin.20036>, doi:[10.1002/kin.20036](https://doi.org/10.1002/kin.20036).
- [55] M. P. Burke, M. Chaos, Y. Ju, F. L. Dryer, S. J. Klippenstein, Comprehensive h<sub>2</sub>/o<sub>2</sub> kinetic model for high-pressure combustion, *International Journal of Chemical Kinetics* 44 (7) (2012) 444–474. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/kin.20603>, doi:<https://doi.org/10.1002/kin.20603>.  
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/kin.20603>

- [56] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, W. C. G. Jr., V. V. Lissianski, Z. Qin, Gri-mech: An optimized detailed chemical reaction mechanism for methane combustion, available at: [http://www.me.berkeley.edu/gri\\_mech/](http://www.me.berkeley.edu/gri_mech/) (2000).
- [57] University of California, San Diego, Chemical-kinetic mechanisms for combustion applications, <http://combustion.ucsd.edu>, san Diego Mechanism, Mechanical and Aerospace Engineering (Combustion Research) (2025).
- [58] E. S. Oran, E. I. Weber J. W. and Stefaniw, M. H. Lefebvre, J. D. Anderson, A numerical study of a two-dimensional h<sub>2</sub>-o<sub>2</sub>-ar detonation using a detailed chemical reaction model, Combustion and Flame 113 (1) (1998) 147 – 163. doi: [https://doi.org/10.1016/S0010-2180\(97\)00218-6](https://doi.org/10.1016/S0010-2180(97)00218-6).
- [59] J. E. Shepherd, Explosion dynamics laboratory: Shock and detonation toolbox - 2018 version, jES 9-19-2018 (2018). URL <http://shepherd.caltech.edu/EDL/PublicResources/sdt/>
- [60] J. Franco, I. Gomez, L. Randez, Sdirk methods for stiff odes with oscillating solutions, Journal of Computational and Applied Mathematics 81 (2) (1997) 197–209. doi:[https://doi.org/10.1016/S0377-0427\(97\)00056-3](https://doi.org/10.1016/S0377-0427(97)00056-3).

## Appendix A. Implicit time integrators

### Appendix A.1. SDIRK-2

The two stage Singly Diagonally Implicit Runge–Kutta (SDIRK-2) method belongs to a class of implicit Runge–Kutta schemes [40, 60]. These methods are stable and exhibit higher order of accuracy than backward Euler, but are more expensive. The SDIRK-2 method is as follows:

$$\begin{aligned}\Gamma_1 &= S(y_n + \gamma\Delta t\Gamma_1) \\ \Gamma_2 &= S(y_n + (1-\gamma)\Delta t\Gamma_1 + \gamma\Delta t\Gamma_2) \\ y_{n+1} &= y_n + \Delta t [(1-\gamma)\Gamma_1 + \gamma\Gamma_2]\end{aligned}\tag{A.1}$$

For the first stage,

$$f(\Gamma_1) = \Gamma_1 - S(y_n + \gamma\Delta t\Gamma_1)\tag{A.2}$$

$$\delta\Gamma_1 = (\Gamma_{1,k+1} - \Gamma_{1,k}) = -\mathcal{G}(\Gamma_{1,k})^{-1} f(\Gamma_{1,k})\tag{A.3}$$

$$\mathcal{G}(\Gamma_{1,k}) \delta\Gamma_1 = -f(\Gamma_{1,k})\tag{A.4}$$

where

$$\mathcal{G}_{ij}(\Gamma_{1,k}) = \delta_{ij} - \gamma\Delta t\mathcal{J}_{ij}(y_n + \gamma\Delta t\Gamma_{1,k})\tag{A.5}$$

Here  $\omega_i$  is evaluated from the initial guess of  $\Gamma_{1,k=1} = \mathcal{S}(y_n)$  with subsequent  $\Gamma_{1,k}$  coming from the iterated for decrement,  $\Gamma_{1,k+1} = \Gamma_{1,k} + \delta\Gamma_1$ . Once  $\Gamma_{1,k}$  is realized by the decrement being less than a prescribed tolerance,  $\|\delta\Gamma_{1,k}\| < \epsilon$ , the next stage can be solved for in a similar manner,

$$f(\Gamma_2) = \Gamma_2 - S(y_n + (1-\gamma)\Delta t\Gamma_1 + \gamma\Delta t\Gamma_2)\tag{A.6}$$

$$\delta\Gamma_2 = (\Gamma_{2,k+1} - \Gamma_{2,k}) = -\mathcal{G}(\Gamma_{2,k})^{-1} f(\Gamma_{2,k})\tag{A.7}$$

$$\mathcal{G}(\Gamma_{2,k}) \delta\Gamma_2 = -f(\Gamma_{2,k})\tag{A.8}$$

Where

$$\mathcal{G}_{ij}(\Gamma_{2,k}) = \delta_{ij} - \gamma\Delta t\mathcal{J}_{ij}(y_n + (1-\gamma)\Delta t\Gamma_1 + \gamma\Delta t\Gamma_{2,k})\tag{A.9}$$

Here  $\omega_i$  is evaluated from the initial guess of  $\Gamma_{2,k=1} = \mathcal{S}(y_n)$  with subsequent  $\Gamma_{2,k}$  coming from the iterated for decrement,  $\Gamma_{2,k+1} = \Gamma_{2,k} + \delta\Gamma_2$ . Once  $\Gamma_{2,k}$  is realized the next stage can be solved for in a similar manner. Chemgen utilizes  $\gamma = 1 - \sqrt{2}/2$  [41].

### Appendix A.2. SDIRK-4

The five stage Singly Diagonally Implicit Runge-Kutta (SDIRK-4) follows the same structure as SDIRK-2 with additional stages requiring a set of coefficients. We utilize the coefficients implemented by FATODE [41] which have shown stability for stiff problems and achieve 4th order accuracy. The SDIRK-4 method is written as

$$\begin{aligned} y_{n+1} &= y_n + \Delta t \sum_{i=1}^5 b_i \Gamma_i \\ \Gamma_p &= S \left( y_n + \sum_{j=1}^i a_{pj} \Delta t \Gamma_j \right) \end{aligned}$$

with coefficients

$$\begin{aligned} a_{ii} &= 0.2\bar{6} & a_{21} &= 0.5 & a_{31} &= 3.541539 \times 10^{-1} \\ a_{32} &= -5.415395 \times 10^{-2} & a_{41} &= 8.515494 \times 10^{-2} & a_{42} &= -6.484332 \times 10^{-2} \\ a_{43} &= 7.915325 \times 10^{-2} & a_{51} &= 2.100115 & a_{53} &= 2.399816 \\ a_{54} &= -2.998818 & b_1 &= 2.100115 & b_2 &= -7.677800 \times 10^{-1} \\ b_3 &= 2.399816 & b_4 &= -2.998818 & b_5 &= 0.2\bar{6} \end{aligned} \quad (\text{A.10})$$

The implicit strategy is similar to SDIRK-2 but written more generally here

$$f(\Gamma_p) = \Gamma_p - S \left( y_n + \sum_{j=1}^p a_{pj} \Delta t \Gamma_j \right) \quad (\text{A.11})$$

$$\delta \Gamma_p = (\Gamma_{p,k+1} - \Gamma_{p,k}) = -\mathcal{G}(\Gamma_{p,k})^{-1} f(\Gamma_{p,k}) \quad (\text{A.12})$$

$$\mathcal{G}(\Gamma_{p,k}) \delta \Gamma_p = -f(\Gamma_{p,k}) \quad (\text{A.13})$$

Where

$$\mathcal{G}_{ij}(\Gamma_{p,k}) = \delta_{ij} - a_{pp} \Delta t \mathcal{J}_{ij} \left( y_n + \sum_{j=1}^p a_{pj} \Delta t (\Gamma_j)_k \right). \quad (\text{A.14})$$

Like SDIRK-2, the subsequent  $\Gamma_{p,k}$  comes from the iterated decrement,  $\Gamma_{p,k+1} = \Gamma_{p,k} + \delta \Gamma_p$ . Once  $\Gamma_{p,k}$  is realized by the decrement being less than a prescribed tolerance,  $\|\delta \Gamma_p\| < \epsilon$ , the next stage can be solved for in a similar manner.

### Appendix A.3. 2nd-order Rosenbrock

Rosenbrock schemes are another attractive time integration strategy but removes the need of multiple Jacobian inversions that are required in a Newton solve per stage as shown in the prior SDIRK based methods.

$$\begin{aligned} \left( \frac{1}{\Delta t \gamma_{11}} \mathcal{I} - \mathcal{J}_n \right) \Gamma_1 &= S(y_n) \\ \left( \frac{1}{\Delta t \gamma_{22}} \mathcal{I} - \mathcal{J}_n \right) \Gamma_2 &= S(y_n + \alpha \Gamma_1) + \frac{\beta}{\Delta t} \Gamma_1 - S(y_n) \\ y_{n+1} &= y_n + m_1 \Gamma_1 + m_2 \Gamma_2 \end{aligned}$$

Here  $\mathcal{J}_n$  is the chemical Jacobian,  $\mathcal{J}_{ij}$ , held constant for both stages, thus removing the need for multiple Jacobian evaluations. Again we utilize the coefficients from FATODE [41],

$$\gamma_{ii} = \gamma = 1 + \frac{\sqrt{2}}{2}, \alpha = 1/\gamma, \beta = \frac{-2}{\gamma} m_1 = \frac{3}{2\gamma} m_2 = \frac{1}{2\gamma}. \quad (\text{A.15})$$

#### Appendix A.4. YASS

Yet another stiff solver (YASS) is an appealing solver that only uses one linear solve to predict the species evolution. YASS is equivalent to one Newton step for backwards Euler that relies on sub-stepping if the change in  $y_n$  is too large. The integration is

$$(\mathcal{I} - \Delta t \mathcal{J}) \delta y_n = \Delta t S(y_n) \quad (\text{A.16})$$

where  $y_{n+1} = y_n + \delta y_n$ . If the norm of the decrement is larger than some scalar value,  $\|\delta y_n\| > \alpha$ , then the YASS integration's time step is reduced to  $\Delta t_{s+1} = \frac{1}{2}\Delta t_s$  where  $\Delta t_s = \Delta t$  in the first integration. This means that there is an appealing chance that the solver is performing a minimum of one linear solve. This necessitates restricting the solver to fail if the time step reaches a minimum value,  $\Delta t_{s+1} < \Delta t_{\min}$ , to where a possible higher order method can be utilized.

## Appendix B. Constant pressure strategy

In compressible CFD, it is typically assumed that the volume of a cell remains constant over time. This assumption influences the allowable time integration strategies and, in compressible flows, permits pressure to rise in response to changes in species. However, in incompressible or low-Mach solvers, pressure is treated differently: it is assumed to adjust instantaneously, effectively decoupling it from the time evolution and thus remain constant at some thermodynamic value. This means pressure is held constant in time, which imposes a constraint on the energy conservation. Specifically, if pressure does not evolve, then the enthalpy,  $\rho h = \sum_{i=1}^{n_s} W_i C_i h_i$ , must also remain constant in time. This constraint becomes clear when examining the thermodynamic relationship between internal energy, enthalpy, and pressure,

$$\rho u = \sum_{i=1}^{n_s} W_i C_i u_i = \sum_{i=1}^{n_s} W_i C_i h_i - R^o T \sum_{i=1}^{n_s} C_i = \sum_{i=1}^{n_s} W_i C_i h_i - p = \rho h - p, \quad (\text{B.1})$$

and applying a time derivative

$$\frac{\partial \rho u}{\partial t} = \frac{\partial \rho h}{\partial t} - \frac{\partial p}{\partial t} = \frac{\partial \rho h}{\partial t} = 0. \quad (\text{B.2})$$

Therefore, the constant pressure assumption can be enforced by treating enthalpy as a globally constant, while leaving the rest of the system formulation unchanged. In the same manner that arrived at Equation (41) a source term for temperature is found under constant enthalpy,

$$\frac{\partial \rho h}{\partial t} = \frac{\partial \rho h}{\partial T} \frac{\partial T}{\partial t} + \sum_{i=1}^{n_s} \frac{\partial \rho h}{\partial C_i} \frac{\partial C_i}{\partial t} = 0 \quad (\text{B.3})$$

which gives

$$\frac{\partial T}{\partial t} = \omega_T = -\frac{\sum_{i=1}^{n_s} W_i h_i \omega_i}{c_p}. \quad (\text{B.4})$$

## Appendix C. Source term and source term Jacobian errors

### Appendix C.1. Source term error due to thermodynamic polynomial degree

To investigate the effect of polynomial order on source term accuracy, we perform the same error analysis of Section (5.1) while varying  $n_p$ . Figure (C.10) presents the results for two alternative choices:  $n_p = 4$  (left) and  $n_p = 15$  (right). The error distributions shift with polynomial order. Specifically, the mean errors  $\mu$  associated with  $n_p = 4$  are approximately an order of magnitude larger than those for  $n_p = 7$ , while the  $n_p = 15$  results exhibit mean errors that are an order of magnitude smaller. The standard deviation  $\sigma$  follows the same trend, indicating that increasing the polynomial degree not only improves the accuracy but also enhances the precision of the predicted source terms.

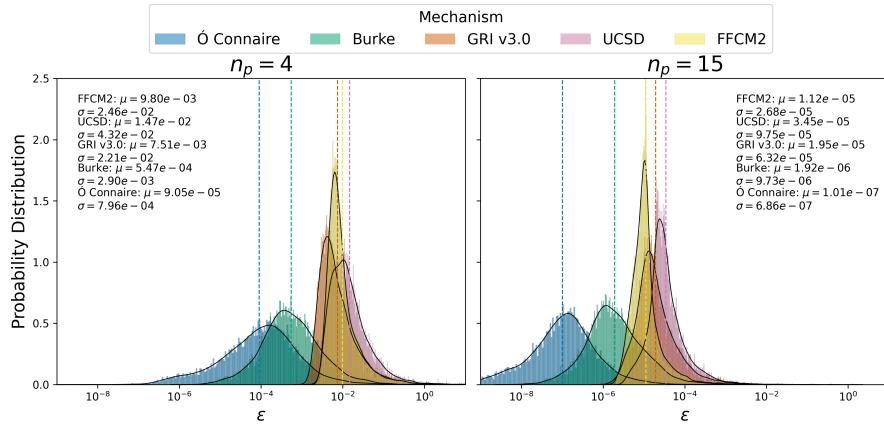


Figure C.10: Error distributions (Equation (56)) for ChemGen source calculations in comparison to Cantera source calculations for five tested chemical models: O'Conaire [54], Burke [55], GRI-Mech 3.0 [56], UCSD [57], and FFCM2 [20]. The vertical dash line represents the mean of the distributions and is colored according to the corresponding model. The distribution mean,  $\mu$ , and corresponding standard deviation,  $\sigma$ , are also reported.

The reduction in error with increasing polynomial degree provides additional confidence in the correctness of the ChemGen implementation. We also found that several parameters in the UCSD model, particularly those used in Troe fall-off reactions, contain extremely large or small values. For instance, in reaction 9, the parameters are specified as T3: `1.0e-30`, T1: `1.0e+30`. The interpretation and handling of such extreme values can vary between software implementations. Examining  $F_{\text{cent}}$ , given by

$$F_{\text{cent}}(T) = (1 - \alpha_j) \exp(-T/T_{j,3}) + \alpha \exp(-T/T_{j,1}) + \exp(-T_{j,2}/T),$$

we note that the very small value of  $T_3$  effectively drives the first term to zero, while the extremely large value of  $T_1$  drives the second exponential term to one. It is assumed, based on the specification, that omitting  $T_2$  entirely implies that the third term is excluded from the calculation, rather than setting  $T_2 = 0$  [22]. Updating ChemGen to address these edge cases and avoid numerical instabilities led to reduced error and recovered the expected trend that higher-order polynomial fits improved accuracy.

### Appendix C.2. Source term Jacobian error

Niemeyer et al. verified the PyJac-generated Jacobian by comparing their analytical Jacobian to a differentiated Jacobian across a wide range of thermochemical states, evaluating between 100,000 and 1 million samples depending on the chemical model [24]. They reported two metrics that could be used per chemical state to evaluate the relative error,

$$E_{\text{rel}} = \left\| \frac{\mathcal{J}_{ij} - \hat{\mathcal{J}}_{ij}}{\hat{\mathcal{J}}_{ij}} \right\|_F = \sqrt{\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \left( \frac{\mathcal{J}(C_i, T)_{ij} - \hat{\mathcal{J}}(C_i, T)_{ij}}{\hat{\mathcal{J}}(C_i, T)_{ij}} \right)^2} \quad \text{and} \quad (\text{C.1})$$

Table C.6:  $E_{norm}$  and  $E_{rel}$  mean and max errors for different chemical models given 10,000 randomly produced states each. Overall, the  $E_{rel}$  errors were generally higher, but never exceeding  $3 \times 10^{-3}$ .

Mechanism	$E_{norm}$ Mean Error	$E_{norm}$ Max Error	$E_{rel}$ Mean Error	$E_{rel}$ Max Error
Ó Connaire	$3.92 \times 10^{-10}$	$2.39 \times 10^{-7}$	$3.38 \times 10^{-7}$	$8.71 \times 10^{-5}$
Burke	$1.23 \times 10^{-10}$	$5.43 \times 10^{-8}$	$4.44 \times 10^{-7}$	$1.01 \times 10^{-4}$
GRI v3.0	$8.46 \times 10^{-11}$	$5.57 \times 10^{-9}$	$8.10 \times 10^{-6}$	$4.07 \times 10^{-4}$
UCSD	$7.55 \times 10^{-12}$	$7.27 \times 10^{-10}$	$9.06 \times 10^{-5}$	$2.53 \times 10^{-3}$
FFCM-2	$3.41 \times 10^{-9}$	$3.35 \times 10^{-5}$	$2.70 \times 10^{-4}$	$1.70 \times 10^{-4}$

$$E_{norm} = \frac{\|\mathcal{J}_{ij} - \hat{\mathcal{J}}_{ij}\|_F}{\|\hat{\mathcal{J}}_{ij}\|_F} = \frac{\sqrt{\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} (\mathcal{J}(C_i, T)_{ij} - \hat{\mathcal{J}}(C_i, T)_{ij})^2}}{\sqrt{\sum_{i=1}^{n_s} \sum_{j=1}^{n_s} (\hat{\mathcal{J}}(C_i, T)_{ij})^2}}. \quad (C.2)$$

In these error metrics, only matrix elements where  $|\mathcal{J}_{ij}| > \|\hat{\mathcal{J}}_b\|_F / 10^{20}$  are considered, where the subscript  $F$  denotes the Frobenius norm, i.e.,  $\|A\|_F = \sum_{j=1}^m \sum_{i=1}^n (A_{ij})^2$  for matrix  $A$  of size  $n \times m$ .

In Niemeyer et al., it was reported that  $E_{rel}$  may be more useful in identifying discrepancies in both larger and smaller matrix elements. Table (C.6) reports the mean and maximum error norms for both  $E_{rel}$  and  $E_{norm}$  using ChemGen for five chemical models with 10,000 randomly generated chemical states. Each analytical Jacobian,  $\mathcal{J}_{ij}$ , is compared to the second order finite difference Jacobian,  $\hat{\mathcal{J}}_{ij}$ , with  $\delta C = 10^{-6}$  Kmol/m<sup>3</sup> and  $\delta T = 10^{-3}$  K. Overall, both error norms suggest that the analytical Jacobian is producing the expected derivatives. In this study,  $E_{rel}$  errors are generally higher, but never exceeding  $3 \times 10^{-3}$ . Figure (C.11) shows the distributions of  $E_{norm}$  (left) and  $E_{rel}$  (right). The  $E_{norm}$  distributions are orders of magnitude smaller than the  $E_{rel}$  distributions. Regardless, these results give confidence in the ability to accurately calculate the analytical Jacobians.

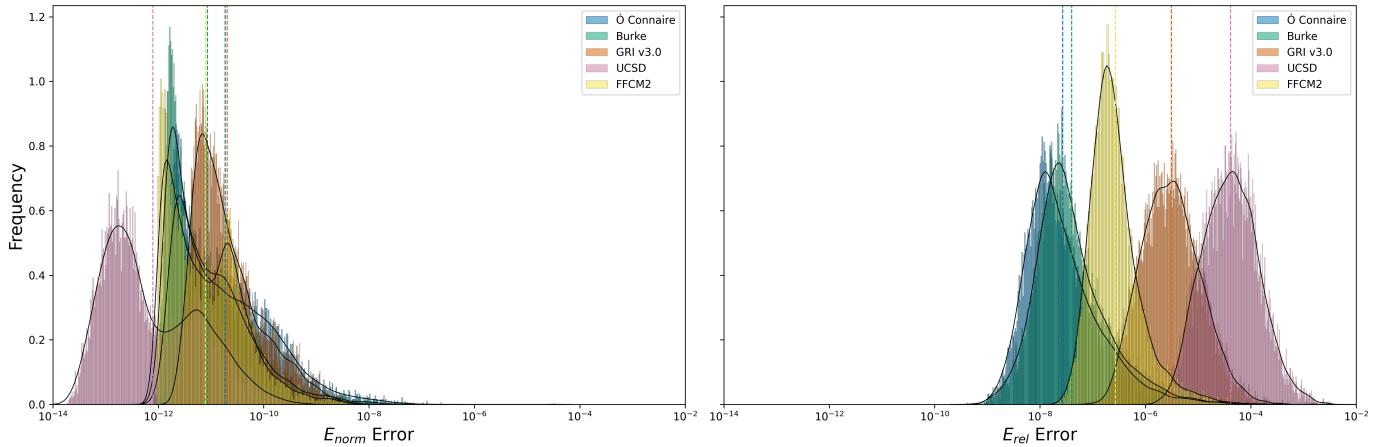


Figure C.11: Distributions of  $E_{norm}$  and  $E_{rel}$  for ChemGen source term Jacobian calculations in comparison to Cantera source calculations for five tested chemical models from Table 3