COMM. MATH. SCI. © 2003 International Press
Vol. 1, No. 3, pp. 449–470

AN IMEX FINITE VOLUME SCHEME FOR REACTIVE EULER EQUATIONS ARISING FROM KINETIC THEORY[*]

MARIA GROPPI[†] AND MICOL PENNACCHIO[‡]

Abstract. A class of reactive Euler-type equations derived from the kinetic theory of chemical reactions is presented and a finite–volume scheme for such problem is developed. The proposed method is based on a flux–vector splitting approach and it is second–order in space and time. The final non–linear problem coming from the discretization has a characteristic block diagonal structure that allows a decoupling in smaller subproblems. Finally, a set of numerical tests shows interesting behaviors in the evolution of the space-dependent fluid-dynamic fields driven by chemical reactions, not present in previous space homogeneous simulations.

Key words. Boltzmann equation; chemical reactions; finite volumes; semi-implicit schemes.

AMS subject classifications. 82C40, 76M12

1. Introduction

We focus here on macroscopic balance equations with source terms arising from a kinetic model for chemically reacting gas mixtures proposed quite recently in [17]. Kinetic descriptions of chemical reactions, based on Boltzmann-like equations, represent a fundamental point to derive macroscopic laws for non-conservative phenomena starting from the mesoscopic level (see the survey paper [16] and the references therein). A closed set of balance equations for the main macroscopic fields of physical relevance can be obtained in the physical situations when elastic scattering is the dominant mechanism in the process [9]. The procedure amounts to a sort of Euler closure and gives a good description of the chemical kinetics at a hydrodynamic level, after the short initial layer in which elastic equilibrium is rapidly approached. This reactive Euler approximation turns out to be robust, in the sense that the main features of the kinetic equations (equilibria, H-theorem, mass action law, conservation laws) are correctly reproduced.

Preliminary numerical simulations in space homogeneous conditions have been presented in [9]. In this work we develop a numerical strategy to simulate the reactive Euler equations in more general space-dependence conditions.

The proposed numerical approximation is not new in the literature for the non–reactive Euler equations but, at least to the authors' knowledge, it has never been used before for this kind of problem arising from kinetic theory. It starts from a splitting of the physical flux vector into a *convective* and a *non–convective* part. Then, for the reactive terms, following the ideas of [5] a suitable reaction matrix is introduced, whose structure and properties facilitated the construction of an efficient numerical solution algorithm. The numerical scheme is derived from a special semi-implicit evaluation of both the numerical flux chosen and the reactive terms; as a result we obtain a method belonging to the class of the IMEX–RK methods.

A global second–order in space and time accuracy is obtained using a linear reconstruction in space [14, 12] and the forward–backward Midpoint time marching

[*] Received: November 9, 2002; accepted (in revised version): May 26, 2003. Communicated by Shi Jin.

[†] Dipartimento di Matematica, Università degli Studi di Parma, via M. D'Azeglio 85A, Parma, Italy, 43100 (maria.groppi@unipr.it).

[‡] Istituto di Matematica Applicata e Tecnologie Informatiche - CNR via Ferrata 1, Pavia, Italy, 27100 (micol@imati.cnr.it).

scheme. The final discretized problem is non–linear due to the semi–implicit treatment of the chemical reaction source terms. This approach yields a time evolution matrix operator with a typical block structure that allows a decomposition of the global non–linear problem into three separate smaller problems, solved iteratively via a block Gauss–Seidel– like algorithm. Concerning the non–linearity, our numerical experience shows that convergence of the iterative fixed point scheme is achieved by a small number of iterations.

Finally, exploiting the structure of the algebraic system we verify that each single block coefficient matrix is an M–matrix. This property guarantees the positivity of the species mass densities at each time step under suitable constraints. Moreover, all the uncoupled systems are solved without the use of Jacobian matrices or their approximations. Indeed we solve different systems characterized by M–matrices; this procedure is both easy and fast and considerably simplifies and accelerates the implementation of the scheme.

The robustness and the accuracy of the method are tested first on the classical Euler equations of gas dynamics, then the role of the chemical source terms is investigated. The promising results reported here show interesting effects due to the chemical reaction, especially on the spatial distribution of the compound, not yet available in literature for this class of reactive equations arising from kinetic theory. These preliminary numerical tests can be regarded as a first step towards effective simulation of industrial reactors; moreover they also provide qualitative informations about the solutions of the Boltzmann-like equations for chemical reactions, at a competitive cost if compared to semi–continuous schemes applied to the original kinetic equations [10].

The paper is organized as follows. In Section 2 we present the mathematical model, in Section 3 we derive the numerical scheme, the algebraic decomposition and its iterative solution procedure. Finally, in Section 4 we report some numerical simulations, testing the proposed method and showing the main features of the model and in Section 5 we summarize our conclusions.

2. The Mathematical Model

The kinetic model of the chemically reacting gas that we consider in this work is composed of a mixture of four different species indicated by the symbol $\mathcal{S}_i$, where the integer index i is ranging from 1 to 4. We assume that – besides the elastic collisions and considering only translational degrees of freedom – these species can also interact according to the bi-molecular reaction

$$\mathcal{S}_1 + \mathcal{S}_2 \rightleftharpoons \mathcal{S}_3 + \mathcal{S}_4. \tag{2.1}$$

First we briefly recall here the Boltzmann–like equations governing the evolution of this chemically reacting mixture, that were deduced in [17] and presented there in a different formulation. Then we summarize the main steps of the closure strategy that allow to obtain the reactive Euler equations starting from the kinetic model (see [9] for details). In what follows, $\mathbf{v}$ and $\mathbf{w}$ stand for velocity vectors, and f_i denotes the i-th one-particle distribution function, with $\underline{f}$ for the vector (f_1, f_2, f_3, f_4). Explicit dependence on position $\mathbf{x}$ and time t will be omitted unless necessary. The relative velocity $\mathbf{v} - \mathbf{w}$ will be cast as $g\mathbf{n}$, with $g = |\mathbf{v} - \mathbf{w}|$ and $|\mathbf{n}| = 1$. If a (i, j) collision results in generation of a pair (h, k) at velocities $\mathbf{v}'$ and $\mathbf{w}'$, the differential scattering cross section is labeled by σ_{ij}^{hk}, and depends only on g and on $\mathbf{n} \cdot \mathbf{n}'$, where $\mathbf{n}' = (\mathbf{v}' - \mathbf{w}')/g'$. In the case of elastic collisions, since $(h, k) = (i, j)$ necessarily, notation will be simplified to σ_{ij} only. The symbols m_i and E_i stand for particle mass and internal energy of chemical link, $\Delta E_{ij}^{hk} = E_h + E_k - E_i - E_j$ denotes the total variation of internal energy, and index ordering may always be chosen in such a way that $\Delta E_{12}^{34} \triangleq \Delta E \geq 0$ (endothermic direct reaction in (2.1)). Moreover, mass conservation in (2.1) means that $m_1 + m_2 = m_3 + m_4 = M$.

The extended Boltzmann equations for the evolution of the one-particle distribution functions f_i are

$$\frac{\partial f_i}{\partial t} + \mathbf{v} \cdot \frac{\partial f_i}{\partial \mathbf{x}} = I_i[\underline{f}] + J_i[\underline{f}], \qquad i = 1, 2, 3, 4 \tag{2.2}$$

where I_i and J_i are the elastic and chemical integral collision terms.

The term I_i takes the usual form [7]

$$I_i[\underline{f}] = \sum_{j=1}^{4} \int_{\mathbf{R}^3} \int_{S^2} g\, \sigma_{ij}(g, \mathbf{n} \cdot \mathbf{n}') \Big[f_i(\mathbf{v}_{ij}^{ij}) f_j(\mathbf{w}_{ij}^{ij}) - f_i(\mathbf{v}) f_j(\mathbf{w}) \Big]\, d_3\mathbf{w}\, d_2\mathbf{n}', \tag{2.3}$$

where the two-dimensional unit sphere S^2 is the domain of integration for the unit vector $\mathbf{n}'$. The post-collision velocities in the general interaction $(i, j) - (h, k)$ are given by

$$\mathbf{v}' = \mathbf{v}_{ij}^{hk} = \alpha_{ij}\mathbf{v} + \alpha_{ij}\mathbf{w} + \alpha_{kh} g_{ij}^{hk}\, \mathbf{n}', \qquad \mathbf{w}' = \mathbf{w}_{ij}^{hk} = \alpha_{ij}\mathbf{v} + \alpha_{ij}\mathbf{w} - \alpha_{hk} g_{ij}^{hk}\, \mathbf{n}', \tag{2.4}$$

where $\alpha_{ij} = m_i/(m_i + m_j)$ are the mass ratios and the outgoing relative speeds are given by

$$g_{ij}^{hk} = \left[\frac{\mu_{ij}}{\mu_{hk}} \left(g^2 - \frac{2\, \Delta E_{ij}^{hk}}{\mu_{ij}} \right) \right]^{\frac{1}{2}}, \tag{2.5}$$

depending also on the reduced masses $\mu_{ij} = \alpha_{ij} m_j$; of course we have $g_{ij}^{ij} = g$.

Each of the chemical integral collision terms J_i is given by a single contribution, because there is a unique chemical reaction in which species i is gained or lost. Upon invoking microreversibility, the first integral term J_1 can be written as [9]

$$J_1[\underline{f}] = \int_{\mathbf{R}^3} \int_{S^2} \Theta\, g\, \sigma_{12}^{34}(g, \mathbf{n} \cdot \mathbf{n}') \left[\left(\frac{\mu_{12}}{\mu_{34}} \right)^3 f_3(\mathbf{v}_{12}^{34})\, f_4(\mathbf{w}_{12}^{34}) - f_1(\mathbf{v})\, f_2(\mathbf{w}) \right] d_3\, \mathbf{w}\, d_2\mathbf{n}', \tag{2.6}$$

and the expressions relevant to $i = 2, 3, 4$ are obtained from (2.6) by a cyclic permutation of the indices.

The unit step function $\Theta = \Theta\Big(g^2 - 2\, \frac{\Delta E}{\mu_{12}}\Big)$ is relevant to the fact that the direct reaction in (2.1) occurs when the kinetic energy of the relative motion overcomes the endothermic threshold ΔE.

The major macroscopic moments of physical relevance are $(i = 1, \ldots, 4)$

- the number densities $n_i = \displaystyle\int_{\mathbf{R}^3} f_i\, d_3\mathbf{v}$ and the total number density $n = \displaystyle\sum_{i=1}^{4} n_i$;
- the mass densities $\rho_i = m_i n_i$ and the total mass density $\rho = \displaystyle\sum_{i=1}^{4} \rho_i$;

- the drift velocities $\mathbf{u}_i = \dfrac{1}{n_i}\displaystyle\int_{\mathbf{R}^3} \mathbf{v}\, f_i \, d_3\mathbf{v}$ and the mass velocity $\mathbf{u} = \dfrac{1}{\rho}\displaystyle\sum_{i=1}^{4} \rho_i \mathbf{u}_i$;
- the pressure tensor $\mathbf{P} = \displaystyle\sum_{i=1}^{4} m_i \int_{\mathbf{R}^3} (\mathbf{v}-\mathbf{u}) \otimes (\mathbf{v}-\mathbf{u})\, f_i \, d_3\mathbf{v}$, whose trace is linked to the gas pressure by the relation $p = \dfrac{1}{3}\mathrm{tr}(\mathbf{P})$;
- the energy density $E = \dfrac{3}{2}\dfrac{p}{\rho}$ and the excitation energy $\mathcal{E}_{\mathrm{ch}} = \displaystyle\sum_{i=1}^{4} E_i n_i$.

The kinetic temperature is defined as usual by $T = \dfrac{p}{n\, K_{\mathrm{B}}}$, with K_{B} Boltzmann constant.

Among all these macroscopic moments of the distribution functions, there exist seven quantities which are conserved under the whole collision process and correspond to seven independent collision invariants [17]: they are three independent partial mass densities, like $\rho_1+\rho_3$, $\rho_1+\rho_4$, $\rho_2+\rho_4$, momentum $\rho\mathbf{u}$ and total energy $\dfrac{1}{2}\rho\, u^2 + \rho\, E + \mathcal{E}_{\mathrm{ch}}$. There correspondingly are seven macroscopic conservation equations [9, 17] obtained by taking the weak form of the kinetic equations (2.2) relevant to the collision invariants. Such equations constitute an exact but not closed set of coupled differential equations for macroscopic observables, expressing conservation of mass, momentum and total energy. Moreover, a detailed balance principle allows then to derive completely and explicitly collision equilibria for (2.2) from collision invariants. They result in the seven-parameter family of Maxwellian distributions [17]

$$\mathcal{M}_i(\mathbf{v}) = n_i \left(\frac{m_i}{2\pi K_{\mathrm{B}} T}\right)^{\frac{3}{2}} \exp\left[-\frac{m_i}{2K_{\mathrm{B}}T}(\mathbf{v}-\mathbf{u})^2\right], \quad \text{for } i=1,\ldots,4 \tag{2.7}$$

with number densities linked by the mass action law

$$\frac{n_1 n_2}{n_3 n_4} = \left(\frac{\mu_{12}}{\mu_{34}}\right)^{\frac{3}{2}} \exp\left(\frac{\Delta E}{K_{\mathrm{B}}T}\right). \tag{2.8}$$

For practical applications, an appropriate closure of the set of macroscopic conservation equations can be obtained from the assumption that the gas mixture is in mechanical equilibrium, but still far from the chemical one. This assumption is correct in the frequent cases in which the time scale of the elastic relaxation processes is very small with respect to the one of the chemical reaction process. The relevant closure strategy consists first in taking moments of the kinetic equations (2.2) relevant to the elastic collision invariants, which constitute now an 8-dimensional linear space, generated for instance by the mass density of each species, momentum and kinetic energy. Then, in the resulting set of 8 equations, one makes use of the corresponding elastic collision equilibria in the evaluation of all unknown moments, either of the distribution functions or of the chemical collision terms. Elastic collision equilibria are essentially an 8–parameter family of local Maxwellian distributions as in (2.7), with uncorrelated number densities.

Following [9], it is possible to show that under the above approximations the set of reactive Euler-type equations that results from this closure is

$$\begin{aligned}
\frac{\partial}{\partial t}(\rho_i) + \frac{\partial}{\partial \mathbf{x}} \cdot (\rho_i \mathbf{u}) &= \mathcal{C}_i \qquad \text{for } i = 1, \dots, 4 \\
\frac{\partial}{\partial t}(\rho \mathbf{u}) + \frac{\partial}{\partial \mathbf{x}} \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p\mathbf{I}) &= \mathbf{0} \\
\frac{\partial}{\partial t}\left(\tfrac{1}{2}\rho u^2 + \rho E\right) + \frac{\partial}{\partial \mathbf{x}} \cdot \left[\left(\tfrac{1}{2}\rho u^2 + \rho E + p\right)\mathbf{u}\right] &= \mathcal{C}_T
\end{aligned} \tag{2.9}$$

where, due to the chemical non-equilibrium, there are collision–like terms $\mathcal{C}_i, \mathcal{C}_T$ different from zero. Such collision source terms are given by $\mathcal{C}_i = m_i \bar{\mathcal{C}_i}$ in the first four density equations of (2.9), with

$$\bar{\mathcal{C}_3} = \bar{\mathcal{C}_4} = -\bar{\mathcal{C}_1} = -\bar{\mathcal{C}_2} = \mathcal{C}_{\text{chem}} \tag{2.10}$$

and the collision source term in the final energy equation of (2.9) is

$$\mathcal{C}_T = -\Delta E \mathcal{C}_{\text{chem}}, \tag{2.11}$$

where

$$\mathcal{C}_{\text{chem}} = \frac{\gamma_T}{m_3 m_4}\left[\rho_1 \rho_2 \left(\frac{\mu_{34}}{\mu_{12}}\right)^{5/2} \exp\left(-\Delta E/(K_{\text{B}}T)\right) - \rho_3 \rho_4\right]. \tag{2.12}$$

The term γ_T in (2.12) is an average microscopic collision frequency with Gaussian weight functions and by using, as we will, Maxwell molecules assumption for the exothermic reaction $3 + 4 \to 1 + 2$, we have $\gamma_T = \text{const}$ [9].

The set of reactive Euler-type equations (2.9) inherits the conservation properties from the original kinetic model (see [9] for details) and in particular the same combinations of mass densities stated above are conserved; it can be easily deduced by rearranging Eqs. (2.9) and by making use of (2.10).

It is worth noticing that for $\gamma_T \to 0$ the set of equations (2.9) tends to the set of the well-known Euler equations of inviscid gas dynamics. On the other hand, the larger γ_T, the stronger is the role played by the chemical reaction in the evolution. In the limiting situation $\gamma_T \to +\infty$ the equations at the Euler level would imply equating to zero the square brackets in (2.12), which leads to the mass action law (2.8).

The physical model previously introduced can be rewritten in vector (conservative) form as

$$\frac{\partial}{\partial t}\mathbf{U} + \nabla \cdot \mathbf{F}(\mathbf{U}) = \mathbf{S}(\mathbf{U}) \tag{2.13}$$

where

$$\mathbf{U} = \begin{bmatrix} \boldsymbol{\rho} \\ \rho \mathbf{u} \\ \mathcal{E} \end{bmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} \boldsymbol{\rho} \otimes \mathbf{u} \\ \rho \mathbf{u} \otimes \mathbf{u} + p\mathbf{I} \\ (\mathcal{E} + p)\mathbf{u} \end{bmatrix}, \quad \mathbf{S}(\mathbf{U}) = \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{0} \\ \omega \end{bmatrix}, \tag{2.14}$$

$\mathbf{U}$ is the conservative solution vector whose six components are collected in the species mass density vector $\boldsymbol{\rho} = (\rho_1, \dots, \rho_4)^T$, in $\rho\mathbf{u}$ total momentum and in $\mathcal{E} = \frac{1}{2}\rho u^2 + \rho E$ total energy (note that $\rho = \sum_{i=1}^4 \rho_i$ is the total mass density); $\mathbf{F}(\mathbf{U})$ is the non-linear flux function.

The reactive source term $\mathbf{S}(\mathbf{U})$ takes into account chemical reactions in a mixture of thermally perfect gases and is expressed in terms of $\boldsymbol{\omega}$, depending on the mass densities $\boldsymbol{\rho}$ and on the thermal state of the gas mixture. In order to treat easily the reactive term we write $\mathbf{S}(\mathbf{U})$ in a more general way, by introducing an $n_s \times n_s$ matrix ($n_s = 4$) with continuous entries $C_{ij}(\boldsymbol{\rho}, T)$ such that

$$\boldsymbol{\omega} = D_m \, \bar{\boldsymbol{\omega}}, \qquad \omega = -\mathbf{1}_{n_s}^T D_E \, \bar{\boldsymbol{\omega}}$$

with $\bar{\boldsymbol{\omega}} = \mathbf{C}(\boldsymbol{\rho}, T)\boldsymbol{\rho}$, $D_m = \mathrm{diag}(m_1, \ldots, m_{n_s})$, $D_E = \mathrm{diag}(E_1, \ldots, E_{n_s})$ and $\mathbf{1}_{n_s}$ vector in $\mathbf{R}^{n_s}$ whose components are all equal to unity. Thus our chemical reaction model can be included in a general formulation proposed in [5] to deal with reactive hypersonic flows, simply defining the entries of this reaction matrix $\mathbf{C}(\boldsymbol{\rho}, T)$. The explicit form of the matrix $\mathbf{C}(\boldsymbol{\rho}, T)$ used in this work is

$$\mathbf{C}(\boldsymbol{\rho}, T) = \begin{pmatrix} -\nu \; \rho_2 & 0 & \dfrac{\beta}{2}\rho_4 & \dfrac{\beta}{2}\rho_3 \\ 0 & -\nu\rho_1 & \dfrac{\beta}{2}\rho_4 & \dfrac{\beta}{2}\rho_3 \\ \dfrac{\nu}{2}\rho_2 & \dfrac{\nu}{2}\rho_1 & -\beta \; \rho_4 & 0 \\ \dfrac{\nu}{2} \; \rho_2 & \dfrac{\nu}{2}\rho_1 & 0 & -\beta \; \rho_3 \end{pmatrix} \tag{2.15}$$

with

$$\nu = \frac{\gamma_T \; k(T)}{m_3 m_4} \qquad \beta = \frac{\gamma_T}{m_3 m_4}$$
$$k(T) = \left(\frac{\mu_{34}}{\mu_{12}}\right)^{5/2} \exp(-\Delta E / K_{\mathrm{B}} T).$$

The structure of the source term written in this way and the properties of the matrix $\mathbf{C}(\boldsymbol{\rho}, T)$ will be useful in the following construction of the numerical approximation algorithm.

3. Numerical Approximation

This section is devoted to the numerical approximation of system (2.13) by means of a finite volume method. We focus on the physical situation where the distribution functions depend only on one spatial coordinate and cylindrical symmetry in molecular velocity space is assumed. In this case, we have that $f_s = f_s(x, v_r, v_z, t)$ and it can be easily checked that the reactive Euler system (2.9) reduces to a set of PDEs in only one space variable (Euler equations in slab symmetry).

The numerical approximation was inspired by the scheme proposed in [5] and here we describe in detail how the procedure can be adapted to our particular model. We start with a decomposition of the physical flux vector function into the sum of a *convective* and a *non-convective* part, i.e. we write the flux vector $\mathbf{F}(\mathbf{U})$ as

$$\mathbf{F}(\mathbf{U}) = a(\mathbf{U}) \; \mathbf{U} + \mathbf{g}(\mathbf{U}) \tag{3.1}$$

with $a(\mathbf{U}) = u$ and $\mathbf{g}(\mathbf{U}) = (\mathbf{0}_{n_s}, p, pu)^T$. The numerical scheme used, that belongs to the class of the IMEX–RK schemes [2, 1, 15], discretizes semi–implicitly the convective part $a(\mathbf{U}) \; \mathbf{U}$ together with the source term $\mathbf{S}(\mathbf{U})$, whereas the non-convective part $\mathbf{g}(\mathbf{U})$ of the flux (3.1) is discretized explicitly.

3.1. The Semi-discrete FV scheme. The Finite Volume (FV) method has been used considering a uniform mesh. Cells are conventionally labeled by an integer identifier ranging from 1 to N, Δx and Δt are spatial and time steps (scales) respectively. We will adopt the following notations: $x_j := j\ \Delta x$, $x_{j\pm 1/2} := (j \pm 1/2)\ \Delta x$, $t^n := n\ \Delta t$, $\mathbf{U}_i$ i-th cell-averaged solution; $\underline{\mathbf{U}}$ is the global collection of N cell-averaged data, i.e. $\underline{\mathbf{U}}$ is the $N \times 6$-size block vector $\underline{\mathbf{U}}^T = (\mathbf{U}_1^T, \mathbf{U}_2^T, \dots, \mathbf{U}_N^T)$, whose i-th block is the vector $\mathbf{U}_i$ with length 6.

If we reformulate equation (2.13) in an integral form for each cell of the mesh, apply the Gauss divergence theorem and introduce suitable numerical flux functions to discretize the physical flux, then we obtain the following semi-discrete FV numerical scheme

$$\frac{d\mathbf{U}_i}{dt} + \frac{1}{\Delta x}\left(\mathsf{H}_{i,i+1}(\underline{\mathbf{U}}) - \mathsf{H}_{i-1,i}(\underline{\mathbf{U}})\right) = \mathbf{S}_i, \qquad i = 1, 2, ..., N \tag{3.2}$$

where $\mathbf{S}_i = \mathbf{S}(\mathbf{U}_i)$ is the source term computed on the i-th cell given by

$$\mathbf{S}_i = \mathbf{S}(\boldsymbol{\rho}_i, T_i) = \begin{pmatrix} D_m \mathbf{C}(\boldsymbol{\rho}_i, T_i)\boldsymbol{\rho}_i \\ 0 \\ -\mathbf{1}_{n_s}^T D_E \mathbf{C}(\boldsymbol{\rho}_i, T_i)\boldsymbol{\rho}_i \end{pmatrix} \qquad i = 1, \dots, N. \tag{3.3}$$

The term $\mathsf{H}_{i,j}(\underline{\mathbf{U}})$ denotes the flux estimated by using the cell-averaged solutions $\mathbf{U}_i$ and $\mathbf{U}_j$ within adjacent cells.

3.2. Construction of the numerical flux. The FV scheme (3.2) is defined in terms of a numerical flux, denoted by H, depending on the left and right solution states $\mathbf{U}^L$ and $\mathbf{U}^R$, i.e. $\mathsf{H} = \mathsf{H}(\mathbf{U}^L, \mathbf{U}^R)$. More specifically, the numerical flux used in this paper can be written as [19]

$$\mathsf{H}(\mathbf{U}^L, \mathbf{U}^R) = \mathbf{F}^+(\mathbf{U}^L, \mathbf{U}^R) - \mathbf{F}^-(\mathbf{U}^L, \mathbf{U}^R) \tag{3.4}$$

where $\mathbf{F}^\pm$ are defined following the splitting (3.1) of $\mathbf{F}$, that is

$$\begin{aligned} \mathbf{F}^+(\mathbf{U}^L, \mathbf{U}^R) &= \mathsf{a}^+(\mathbf{U}^L, \mathbf{U}^R)\ \mathbf{U}^L + \mathsf{G}^+(\mathbf{U}^L, \mathbf{U}^R) \\ \mathbf{F}^-(\mathbf{U}^L, \mathbf{U}^R) &= \mathsf{a}^-(\mathbf{U}^L, \mathbf{U}^R)\ \mathbf{U}^R + \mathsf{G}^-(\mathbf{U}^L, \mathbf{U}^R) \end{aligned} \tag{3.5}$$

with

$$\begin{aligned} \mathsf{a}^+(\mathbf{U}^L, \mathbf{U}^R) &= \frac{s_R}{s_R - s_L}(u_L - s_L) \\ \mathsf{a}^-(\mathbf{U}^L, \mathbf{U}^R) &= \frac{s_L}{s_R - s_L}(u_R - s_R) \end{aligned} \tag{3.6}$$

$$\mathsf{G}^+(\mathbf{U}^L, \mathbf{U}^R) = \frac{s_R}{s_R - s_L}\mathbf{g}(\mathbf{U}^L) \qquad \mathsf{G}^-(\mathbf{U}^L, \mathbf{U}^R) = \frac{s_L}{s_R - s_L}\mathbf{g}(\mathbf{U}^R) \tag{3.7}$$

$$s_R = \max\left\{\lambda_L^{max},\ \lambda_R^{max},\ 0\right\} \qquad s_L = \min\left\{\lambda_L^{min},\ \lambda_R^{min},\ 0\right\}$$

and $\lambda_{L,R}^{min}, \lambda_{L,R}^{max}$ minimum and maximum eigenvalues of the Jacobian matrix of the flux vector function $\mathbf{F}$ computed on $\mathbf{U}^L$ and $\mathbf{U}^R$ respectively. This choice of s_R and s_L yields the so-called *local Lax-Friedrichs* numerical flux (see [4, 8]).

Let us define, for $i = 1, \dots, N$, the vectors $\mathbf{g}_i = \mathbf{g}(\mathbf{U}_{i-1}, \mathbf{U}_i, \mathbf{U}_{i+1}) \in \mathbf{R}^6$

$$\mathbf{g}_i = \mathsf{G}^+(\mathbf{U}_i, \mathbf{U}_{i+1}) + \mathsf{G}^-(\mathbf{U}_{i-1}, \mathbf{U}_i) - \mathsf{G}^-(\mathbf{U}_i, \mathbf{U}_{i+1}) - \mathsf{G}^+(\mathbf{U}_{i-1}, \mathbf{U}_i), \tag{3.8}$$

the $N \times N$ matrix $\mathbf{A} = \mathbf{A}(\underline{\mathbf{U}})$ with elements $a_{i,j}$

$$a_{i,j} = \begin{cases} -\mathsf{a}^-(\mathbf{U}_i, \mathbf{U}_{i+1}) & j = i+1 \\ -\mathsf{a}^+(\mathbf{U}_{i-1}, \mathbf{U}_i) & j = i-1 \\ \mathsf{a}^+(\mathbf{U}_i, \mathbf{U}_{i+1}) + \mathsf{a}^-(\mathbf{U}_{i-1}, \mathbf{U}_i) & i = j \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

and the $6N \times 6N$ matrix $\mathcal{A} = \mathcal{A}(\underline{\mathbf{U}})$

$$\mathcal{A} = (\mathbf{A}(\underline{\mathbf{U}}) \otimes \mathbf{I}_6)$$

where $\otimes$ denotes the tensor product, i.e. given two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ of order $m \times n$ and $p \times q$, then $\mathbf{A}_1 \otimes \mathbf{A}_2$ is the block-matrix of order $mp \times nq$ whose block ij is $(\mathbf{A}_1 \otimes \mathbf{A}_2)_{i,j} = (\mathbf{A}_1)_{i,j}\mathbf{A}_2$.

If we introduce the numerical flux (3.4) into (3.2) we then obtain the final form of the semi-discrete FV scheme

$$\frac{d\mathbf{U}_i}{dt} + \frac{1}{\Delta x}\left(a_{i,i}\mathbf{U}_i + a_{i,i+1}\mathbf{U}_{i+1} + a_{i,i-1}\mathbf{U}_{i-1}\right) + \frac{1}{\Delta x}\mathbf{g}_i = \mathbf{S}_i \tag{3.10}$$

with $a_{i,j}$ elements of the matrix $\mathbf{A}$.

3.3. Second–order in space accuracy. A second–order in space accuracy can be obtained using a linear reconstruction; let q_i be anyone of the components of $\mathbf{U}_i$ for the i-th cell, then the reconstructed value $\tilde{q}_i$ is defined by

$$\tilde{q}_i(x,t) = q_i + (x - x_i)\frac{\sigma_i}{\Delta x} \qquad \text{on the cell } [x_{i-1/2}, x_{i+1/2}] \tag{3.11}$$

with σ_i slope of the i-th cell based on the data $\mathbf{U}_i$ (see [12, 14]). Our choice of slopes is the so-called *minmod slope*[1] $\sigma_i = \text{minmod}(\delta_i^+, \delta_i^-)$, with $\delta_i^+ = q_{i+1} - q_i$ and $\delta_i^- = q_i - q_{i-1}$, which is the simplest one among the limiters proposed in [14]. In order to avoid possible loss of accuracy near local extrema, more effective limiters could be chosen, such as for example the UNO limiter (see [13, 14]). At the interface $x_{i+1/2}$ we have values on the left and right from the two linear approximations in each of the neighboring cells. The left and right cell values for cells i and $i+1$ are obtained as in [12]

$$\tilde{q}_i^L = q_i + \frac{1}{2}\sigma_i \qquad \tilde{q}_{i+1}^R = q_{i+1} - \frac{1}{2}\sigma_{i+1}.$$

Thus when a spatial reconstruction is taken into account, the final form of the semi-discrete FV scheme becomes

$$\frac{d\mathbf{U}_i}{dt} + \frac{1}{\Delta x}\left(a_{i,i}\mathbf{U}_i + a_{i,i+1}\mathbf{U}_{i+1} + a_{i,i-1}\mathbf{U}_{i-1}\right) + \frac{1}{\Delta x}\mathbf{g}_i + \frac{1}{\Delta x}\mathbf{f}_i = \mathbf{S}_i \tag{3.12}$$

with

$$\begin{aligned} \mathbf{f}_i = & \left[\mathsf{a}^+(\tilde{\mathbf{U}}_i^L, \tilde{\mathbf{U}}_{i+1}^R)\left(\tilde{\mathbf{U}}_i^L - \mathbf{U}_i\right) + \mathsf{a}^-(\tilde{\mathbf{U}}_{i-1}^L, \tilde{\mathbf{U}}_i^R)\left(\tilde{\mathbf{U}}_i^R - \mathbf{U}_i\right)\right] \\ & - \left[\mathsf{a}^-(\tilde{\mathbf{U}}_i^L, \tilde{\mathbf{U}}_{i+1}^R)\left(\tilde{\mathbf{U}}_{i+1}^R - \mathbf{U}_{i+1}\right) + \mathsf{a}^+(\tilde{\mathbf{U}}_{i-1}^L, \tilde{\mathbf{U}}_i^R)\left(\tilde{\mathbf{U}}_{i-1}^L - \mathbf{U}_{i-1}\right)\right]. \end{aligned}$$

[1] Here minmod stands for the usual function: $\text{minmod}(a,b) = \frac{1}{2}(sgn(a) + sgn(b))\min(|a|,|b|)$

Now the scalar functions $a_{i,j}$ and the vectors $\mathbf{g}_i$ come from the evaluation of $\mathsf{a}^{\pm}(\tilde{\mathbf{U}}_i^L, \tilde{\mathbf{U}}_j^R)$ and $\mathsf{G}^{\pm}(\tilde{\mathbf{U}}_i^L, \tilde{\mathbf{U}}_j^R)$ using the reconstructed values $\tilde{\mathbf{U}}_i^L, \tilde{\mathbf{U}}_j^R$, whereas $\mathbf{f}_i$ takes into account how much the reconstructed values differ from the cell-averaged solutions.

The semi–discrete formulations (3.10) or (3.12) yield large systems of ordinary differential equations in the unknown $\underline{\mathbf{U}}$. In order to write these systems in a compact way, let us introduce the block vector

$$\underline{\mathbf{b}}(\underline{\mathbf{U}}) = -\underline{\mathbf{g}}(\underline{\mathbf{U}}) - \underline{\mathbf{f}}(\underline{\mathbf{U}}) \tag{3.13}$$

with $\underline{\mathbf{g}}(\underline{\mathbf{U}}) = (\mathbf{g}_1, \dots, \mathbf{g}_N)$ and $\underline{\mathbf{f}}(\underline{\mathbf{U}}) = (\mathbf{f}_1 \dots, \mathbf{f}_N)$. Note that the vector $\underline{\mathbf{g}}(\underline{\mathbf{U}})$ takes into account the contribution given by the non–convective part of the flux $\boldsymbol{\mathsf{g}}(\mathbf{U})$, whereas $\underline{\mathbf{f}}(\underline{\mathbf{U}})$ is generated by the spatial reconstruction used. Hence the components of $\underline{\mathbf{g}}$ related to the species mass density vector $\boldsymbol{\rho}$ are zero, whereas $\underline{\mathbf{f}}$ vanishes when no reconstruction is considered.

Then we finally can write the following system of ordinary differential equations in the unknown $\underline{\mathbf{U}}$

$$\frac{d\underline{\mathbf{U}}}{dt} + \frac{1}{\Delta x}\mathcal{A}(\underline{\mathbf{U}})\ \underline{\mathbf{U}} = \frac{1}{\Delta x}\underline{\mathbf{b}}(\underline{\mathbf{U}}) + \mathbf{S}(\underline{\mathbf{U}}). \tag{3.14}$$

3.4. Time discretization. A semi–implicit scheme is used for the time discretization for both the flux and the reactive source term. This choice is motivated by the fact that the source term becomes stiff for large values of γ_T (see (2.12)). Explicit schemes are not convenient since, because of the stiffness, severe restrictions on the time step can appear. On the other hand, fully implicit discretizations are in general computationally too expensive. Thus we need a numerical scheme that can be always applied, also in the stiff cases, and with acceptable computational costs.

As for the flux, in the framework of IMEX–RK schemes we consider here a semi–implicit scheme where only the convective part is discretized implicitly. More precisely, the term $\mathcal{A}(\underline{\mathbf{U}})(\underline{\mathbf{U}})$ is treated implicitly in $\underline{\mathbf{U}}$ and explicitly in $\mathcal{A}(\underline{\mathbf{U}})$. The first right hand side term $\underline{\mathbf{b}}(\underline{\mathbf{U}})$, including the non–convective part of the flux $\underline{\mathbf{g}}(\underline{\mathbf{U}})$, is discretized explicitly. The resulting semi-implicit scheme, which does not require any numerical evaluation of the Jacobian matrix, produces a system of algebraic equations with a very peculiar block structure that can be easily solved, thus ensuring the effectiveness of the numerical approach.

Concerning the reactive source term $\mathbf{S}(\mathbf{U})$, an implicit discretization is appropriate to handle the possible stiffness introduced by large values of γ_T. However, a fully implicit evaluation of $\mathbf{S}(\mathbf{U})$, i.e. of both $\boldsymbol{\rho}$ and T, yields a strong non–linear system which requires non–linear iterative techniques. The resulting algorithm may be too expensive, then a convenient semi-implicit discretization has been taken into account. Indeed, the particular structure of the chemical source terms $\mathbf{S}(\mathbf{U}_i)$ given in (3.3) suggests to treat implicitly the dependence on $\boldsymbol{\rho}$ and explicitly the one on T.

The first and simpler time-marching scheme considered is a semi–implicit version of the Euler method for ODEs, that is

$$\underline{\mathbf{U}}^{n+1} + \lambda\ \mathcal{A}(\underline{\mathbf{U}}^n)\ \underline{\mathbf{U}}^{n+1} - \Delta t\ \mathbf{S}(\underline{\boldsymbol{\rho}}^{n+1}, \underline{\mathbf{T}}^n) = \underline{\mathbf{U}}^n + \lambda\ \underline{\mathbf{b}}(\underline{\mathbf{U}}^n) \tag{3.15}$$

with $\lambda = \Delta t / \Delta x$, $\mathbf{U}_i^n, \mathbf{U}_i^{n+1}$ cell–averaged solutions at time level $t = t^n$ and $t = t^{n+1}$ respectively.

To obtain a method which is second order accurate in time as well as in space, we can discretize the ODEs (3.14) using a forward–backward Midpoint scheme. The first step evaluates a preliminary solution at the intermediate time $t^{n+1/2}$ while the second step advances the solution from t^n to t^{n+1} by using the approximate solution at $t^{n+1/2}$. More specifically the scheme takes the form

First step.

$$\underline{\mathbf{U}}^{n+1/2} + \frac{\lambda}{2}\,\mathcal{A}(\underline{\mathbf{U}}^n)\;\underline{\mathbf{U}}^{n+1/2} - \frac{\Delta t}{2}\mathbf{S}(\underline{\boldsymbol{\rho}}^{n+1/2}, \underline{\mathbf{T}}^n) = \underline{\mathbf{U}}^n + \frac{\lambda}{2}\,\underline{\mathbf{b}}(\underline{\mathbf{U}}^n), \tag{3.16}$$

Second step.

$$\underline{\mathbf{U}}^{n+1} = \underline{\mathbf{U}}^n + \lambda\left[\underline{\mathbf{b}}(\underline{\mathbf{U}}^{n+1/2}) - \mathcal{A}(\underline{\mathbf{U}}^{n+1/2})\;\underline{\mathbf{U}}^{n+1/2}\right] + \Delta t\;\mathbf{S}(\underline{\boldsymbol{\rho}}^{n+1/2}, \underline{\mathbf{T}}^n). \tag{3.17}$$

Other higher order IMEX–RK scheme can be found in literature (see for instance [1]).

3.5. Solution algorithm. To simplify the presentation of the solution algorithm and to highlight its properties we prefer reordering the vector $\underline{\mathbf{U}}$ as $\underline{\mathbf{U}}^T = \left(\underline{\boldsymbol{\rho}}^T, \underline{\mathbf{q}}^T, \underline{\boldsymbol{\mathcal{E}}}^T\right)$ where $\underline{\boldsymbol{\rho}}$, $\underline{\mathbf{q}}$, $\underline{\boldsymbol{\mathcal{E}}}$ collect all the cell components related to the mass density vector, the momentum and the energy respectively. Then with this notation the Euler method (3.15) and the first step of the Midpoint scheme (3.16) can be written in the following unified matrix form

$$\mathcal{L}_\alpha^n(\underline{\boldsymbol{\rho}}^{n+\alpha})\underline{\mathbf{U}}^{n+\alpha} + \mathcal{C}_\alpha^{n+\alpha} = \mathcal{B}_\alpha^n \tag{3.18}$$

where α is a constant, $\mathcal{L}_\alpha^n(\underline{\boldsymbol{\rho}})$ is the block-diagonal matrix operator

$$\mathcal{L}_\alpha^n(\underline{\boldsymbol{\rho}}^{n+\alpha}) = \begin{pmatrix} \mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha} + \mathbf{M}_\alpha^n \otimes \mathbf{I}_{n_s} & 0 & 0 \\ 0 & \mathbf{M}_\alpha^n & 0 \\ 0 & 0 & \mathbf{M}_\alpha^n \end{pmatrix} \tag{3.19}$$

with

$$\mathbf{M}_\alpha^n = \mathbf{I}_N + \alpha\lambda\mathbf{A}(\underline{\mathbf{U}}^n), \tag{3.20}$$

$\mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha}$ is a block–diagonal non–negative matrix whose i-th block is the $n_s \times n_s$ matrix $\mathbf{D}_{\boldsymbol{\rho},ii}^{n+\alpha}$ defined as

$$\mathbf{D}_{\boldsymbol{\rho},ii}^{n+\alpha} = -\alpha\,\Delta t D_m \mathbf{C}(\boldsymbol{\rho}_i^{n+\alpha}, T_i^n) \tag{3.21}$$

and $\mathcal{C}_\alpha^{n+\alpha}$, $\mathcal{B}_\alpha^n$ are the following vectors

$$\mathcal{C}_\alpha^{n+\alpha} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{b}_{\mathbf{C}}^{n+\alpha} \end{pmatrix} \qquad \mathcal{B}_\alpha^n = \begin{pmatrix} \mathbf{b}_{\boldsymbol{\rho}}^n \\ \mathbf{b}_{\mathbf{q}}^n \\ \mathbf{b}_{\boldsymbol{\mathcal{E}}}^n \end{pmatrix} + \begin{pmatrix} \underline{\boldsymbol{\rho}}^n \\ \underline{\mathbf{q}}^n \\ \underline{\boldsymbol{\mathcal{E}}}^n \end{pmatrix}$$

with $\mathbf{b}_{\boldsymbol{\rho}}^n, \mathbf{b}_{\mathbf{q}}^n, \mathbf{b}_{\boldsymbol{\mathcal{E}}}^n$ components of $\alpha\lambda\underline{\mathbf{b}}(\underline{\mathbf{U}})$ related to $\underline{\boldsymbol{\rho}}$, $\underline{\mathbf{q}}$ and $\underline{\boldsymbol{\mathcal{E}}}$ respectively and $\mathbf{b}_{\mathbf{C}}^n$ the vector whose i-th component is given by

$$\mathbf{b}_{\mathbf{C},\,i}^{n+\alpha} = \alpha\,\Delta t\;\mathbf{1}_{n_s}^T D_E \mathbf{C}(\boldsymbol{\rho}_i^{n+\alpha}, T_i^n)\boldsymbol{\rho}_i^n.$$

Equation (3.15) can be obtained with $\alpha = 1$ whereas to get (3.16) $\alpha = 1/2$.

The typical structure of (3.19) suggests to decompose the global non–linear problem (3.18) into three separate smaller problems, which are solved sequentially via a block Gauss-Seidel–like algorithm. The solution algorithm proceeds as follows:

(i) Solve the non–linear system for species mass densities $\underline{\boldsymbol{\rho}}$

$$\left(\mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha} + \mathbf{M}_{\alpha}^{n} \otimes \mathbf{I}_{n_s}\right) \underline{\boldsymbol{\rho}}^{n+\alpha} = \mathbf{b}_{\boldsymbol{\rho}}^{n} + \underline{\boldsymbol{\rho}}^{n}$$

(ii) Solve the system for momenta $\underline{\mathbf{q}}$

$$\mathbf{M}_{\alpha}^{n} \underline{\mathbf{q}}^{n+\alpha} = \mathbf{b}_{\mathbf{q}}^{n} + \underline{\mathbf{q}}^{n}$$

(iii) Solve the system for total energies $\underline{\boldsymbol{\mathcal{E}}}$

$$\mathbf{M}_{\alpha}^{n} \underline{\boldsymbol{\mathcal{E}}}^{n+\alpha} = \mathbf{b}_{\boldsymbol{\mathcal{E}}}^{n} + \underline{\boldsymbol{\mathcal{E}}}^{n} - \mathbf{b}_{\mathbf{C}}^{n+\alpha}.$$

Note that the term $\mathbf{b}_{\boldsymbol{\rho}}^{n}$ is different from zero only when the spatial reconstruction is considered (see definitions (3.13) and (3.7)).

Step (i) requires the solution of a non–linear problem; this non–linearity comes from the semi–implicit treatment of the chemical source term. If we introduce the map $\Phi(\cdot) : \mathbf{R}^{N \times n_s} \to \mathbf{R}^{N \times n_s}$ defined as follows

$$\Phi(\underline{\boldsymbol{\rho}}^{n+\alpha}) := \left(\mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha} + \mathbf{M}_{\alpha}^{n} \otimes \mathbf{I}_{n_s}\right)^{-1} \left(\mathbf{b}_{\boldsymbol{\rho}}^{n} + \underline{\boldsymbol{\rho}}^{n}\right)$$

then the solution of (i) can be seen as the fixed point $\underline{\boldsymbol{\rho}}^{n+\alpha} = \Phi(\underline{\boldsymbol{\rho}}^{n+\alpha})$. Numerical experiments show that, if the time step Δt is sufficiently small, the map is contractive hence convergence of the iterative fixed point scheme is guaranteed. A detailed study on the existence and uniqueness of the solution can be found in [4]. Finally, steps (ii) and (iii) are linearized by using the upgraded values of $\underline{\boldsymbol{\rho}}^{n+\alpha}$.

The resulting linear algebraic systems may be solved by a standard Krylov solver, such as the preconditioned Bi-CGSTAB method. In the next section we will show that each diagonal block of (3.19) is an M–matrix. This property can help to solve the algebraic systems in a simpler and more efficient way (see [4, 5]).

3.6. Properties of the scheme. Positivity of the mass densities $\boldsymbol{\rho}$ can be obtained using some properties of matrices $\mathbf{M}_{\alpha}^{n}$ and $\mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha}$.

First let us focus on the matrix $\mathbf{C}$ previously introduced for the source term $\mathbf{S}$. It can be shown that the entries $C_{ij}(\boldsymbol{\rho}, T)$ of the $n_s \times n_s$ matrix $\mathbf{C}(\boldsymbol{\rho}, T)$ satisfies

1. $C_{ii}(\boldsymbol{\rho}, T) \leq 0 \; i = 1, \ldots, n_s$
2. $C_{ij}(\boldsymbol{\rho}, T) \geq 0 \; i \neq j$
3. $\sum_{i=1}^{n_s} C_{ij}(\boldsymbol{\rho}, T) = 0 \; j = 1, \ldots, n_s$

For reader's convenience we report the definition of M–matrix and a Lemma that will be used in the sequel (see [3]).

Definition 3.1. *A real square matrix* $\mathbf{A}$ *is called an* M*–matrix if and only if* A *can be written in the form*

$$\mathbf{A} = s\mathbf{I} - \mathbf{B}, \quad s > 0, \quad \mathbf{B} \textit{ non–negative matrix}$$

where $s \geq \rho(\mathbf{B})$ *and* $\rho(\mathbf{B})$ *is the spectral radius of* $\mathbf{B}$*. A non–singular* M*–matrix requires* $s > \rho(\mathbf{B})$.

Lemma 3.2. *The following statements are equivalent:*
i) $\mathbf{A}$ *is a non–singular* M*–matrix*
ii) $\mathbf{A}^T$ *is a non–singular* M*–matrix*
iii) there exists a positive vector $\mathbf{x}$ *such that* $\mathbf{A}\mathbf{x}$ *is also a positive vector*[2]

[2] A positive vector is a vector whose entries are all positive

iv) $\mathbf{A}^{-1}$ *is a non–negative matrix*
Following [4, 5] we can state that

Proposition 3.3. *The three diagonal blocks defined in (3.19) are M–matrices.*

Proof.

1. $\mathbf{M}_\alpha^n$ is a non–singular M–matrix.
   In fact by using the definition of $\mathbf{M}_\alpha$ and from (3.9) we have

$$\left(\mathbf{1}_N^T \mathbf{M}_\alpha^n\right)_j = \sum_{i=1}^{N} M_{\alpha,ij}^n = M_{\alpha,jj}^n + \sum_{i \neq j} M_{\alpha,ij}^n = 1 > 0$$

   Thus from Lemma 3.2 with $\mathbf{x} = \mathbf{1}_N$, we obtain that the second and the third diagonal block in (3.19) are M–matrices.
2. $\mathbf{M}_\alpha^n \otimes \mathbf{I}_{n_s}$ is a non–singular M–matrix.
   Taking $\mathbf{x} = \mathbf{1}_N \otimes \mathbf{1}_{n_s}$, by using the definition of dyadic product and the property $\mathbf{1}_N^T \mathbf{M}_\alpha^n > 0$ previously obtained, we have

$$\begin{aligned} \mathbf{x}^T \left(\mathbf{M}_\alpha^n \otimes \mathbf{I}_{n_s}\right) &= \left(\mathbf{1}_N \otimes \mathbf{1}_{n_s}\right)^T \left(\mathbf{M}_\alpha^n \otimes \mathbf{I}_{n_s}\right) = \\ &= \mathbf{1}_N^T \mathbf{M}_\alpha^n \otimes \mathbf{1}_{n_s}^T \mathbf{I}_{n_s} = \\ &= \mathbf{1}_N^T \mathbf{M}_\alpha^n \otimes \mathbf{1}_{n_s}^T > 0. \end{aligned}$$

3. From the properties of the matrix $\mathbf{C}$ and definition (3.21) we get $\mathbf{x}^T \mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha} = 0$ with $\mathbf{x} = \mathbf{1}_N \otimes \mathbf{1}_{n_s}$. Hence, the first diagonal block in (3.19) $\mathbf{D}_{\boldsymbol{\rho}}^{n+\alpha} + \mathbf{M}_\alpha^n \otimes \mathbf{I}_{n_s}$ is a non–singular M–matrix.

□

Proposition 3.4. *Let us assume to use the time–marching scheme (3.15) without any space reconstruction, hence a globally first–order scheme. If* $\underline{\boldsymbol{\rho}}^0 \geq 0$ *then the scheme is unconditionally non–negative, that is* $\underline{\boldsymbol{\rho}}^{n+1} \geq 0$ $\forall n$ *without any condition on* Δt*.*

Proof. The first diagonal block of equations related to $\underline{\boldsymbol{\rho}}$ (with $\alpha = 1$) is

$$\left(\mathbf{D}_{\boldsymbol{\rho}}^{n+1} + \mathbf{M}_1^n \otimes \mathbf{I}_{n_s}\right) \underline{\boldsymbol{\rho}}^{n+1} = \mathbf{b}_{\boldsymbol{\rho}}^n + \underline{\boldsymbol{\rho}}^n$$

If no reconstruction is devised, $\mathbf{b}_\rho^n$ reduces to zero. Moreover, since $\mathbf{D}_{\boldsymbol{\rho}}^{n+1} + \mathbf{M}_1^n \otimes \mathbf{I}_{n_s}$ is a non–singular M–matrix, from Lemma 3.2*(iv)* if $\underline{\boldsymbol{\rho}}^n \geq 0$ we get unconditionally $\underline{\boldsymbol{\rho}}^{n+1} \geq 0$, i.e. this result is independent of any constraint on Δt. □

Remark 1. When a spatial reconstruction is taken into account $\mathbf{b}_\rho^n$ does not vanish. By using the non–negativity of the inverse of an M–matrix (Lemma 3.2*(iv)*) we obtain the non–negativity of $\underline{\boldsymbol{\rho}}^{n+1}$ under the condition $\mathbf{b}_{\boldsymbol{\rho}}^n + \underline{\boldsymbol{\rho}}^n \geq 0$.

Numerical experience shows that this condition is not too strict and it can be achieved for a quite large range of values λ. However when a second–order in time and space scheme is used, a general discussion is more difficult (see for instance [4, 11]). In this case numerical simulations show that we still have positivity of the species mass densities and also of the total energy, i.e. the scheme prevents the non–physical quantities from appear during the solution process.

4. Numerical Results

In this section we perform some numerical tests both to validate our method, assessing some properties of the scheme used, and to investigate the effects of the chemical reaction on the evolution of the fluid–dynamic quantities characterizing the mixture. All the numerical results presented are obtained using the Midpoint scheme (3.16)-(3.17) and a linear reconstruction in space.

4.1. Application to Euler equations of gas dynamics. We now present an application of the scheme to the classic evolution equations for a mixture of monoatomic gases. Our first test, chosen to validate our numerical scheme, is the Riemann problem for the Euler system of inviscid gas dynamics (system (2.9) in the limit $\gamma_T \to 0$), similar to the one proposed by Sod [18] which consists of the following initial data

$$\rho = 1 \qquad u = 0 \qquad p = \frac{5}{3} \qquad \text{for x } < 0.5$$

$$\rho = \frac{1}{8} \qquad u = 0 \qquad p = \frac{1}{6} \qquad \text{for x } > 0.5.$$

For comparison purposes, we will use the same parameters as in [13] considering the case of four equal gases; hence, we will plot the total mass density ρ. A fixed mesh ratio $\Delta t = \Delta x/9$ as in [13] has been used.

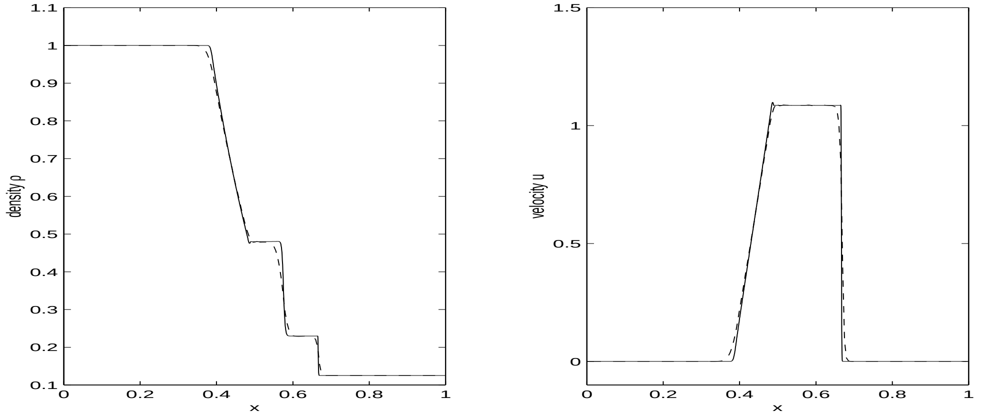


FIG. 4.1. *Solution at time instant $t = 0.07$ of the Euler equations of gas dynamics for a mixture of four equal monoatomic gases. Density (left) and velocity (right) computed with 200 and 1600 grid points (dashed and continuous line respectively).*

We consider the equations in the interval $[0, 1]$ and in Fig. 4.1 we present numerical results at time instant $t = 0.07$. Fig. 4.1 shows the behavior of density and velocity, computed using 200 and 1600 grid points. It is worth noticing that our results well reproduce those obtained in [13] by using generalized Nessyahu-Tadmor schemes.

Then we can extend our scheme to more general cases and we start by dealing with a mixture of four gases with different mass densities, i.e. we consider the following initial data

$$\rho_1 = \frac{1}{10} \quad \rho_2 = \frac{2}{10} \quad \rho_3 = \frac{3}{10} \quad \rho_4 = \frac{4}{10} \qquad u = 0 \qquad p = \frac{5}{3} \qquad \text{for x } < 0.5$$

$$\rho_1 = \frac{1}{80} \quad \rho_2 = \frac{2}{80} \quad \rho_3 = \frac{3}{80} \quad \rho_4 = \frac{4}{80} \qquad u = 0 \qquad p = \frac{1}{6} \qquad \text{for x } > 0.5.$$

The given numerical values have to be considered as dimensionless and corresponding to arbitrary scales; they have been chosen for illustrative purposes only. As before, in Fig. 4.2 we display the solutions at time instant $t = 0.07$ for the different mass densities ρ_i, for global mass density ρ, for velocity u and for pressure p. As expected, the same trend characterizes the plots of mass densities and we obtain results very similar to the ones of Fig. 4.1.

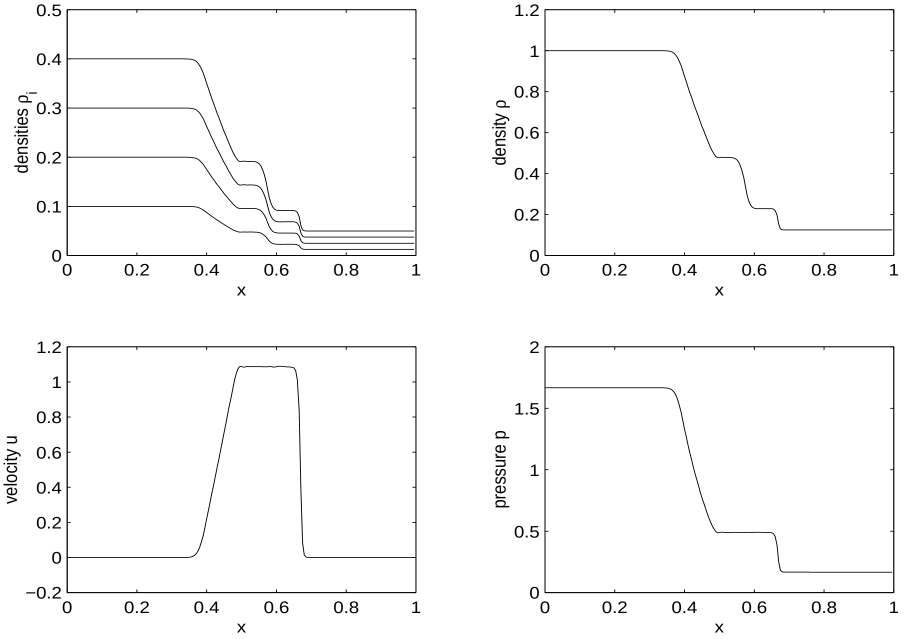


FIG. 4.2. *Solution at time instant $t = 0.07$ of the Euler equations of gas dynamics for a mixture of four gases with different mass densities. Top pictures: densities (left) and total density (right) computed with 200 grids points. Bottom pictures: velocity (left) and pressure (right).*

4.2. Application to the reactive Euler equations. We now perform a few numerical simulations to test our method when applied to the reactive Euler equations. As a first step, we start by considering the case of energy gap $\Delta E = 0$. Consequently, the computed global quantities $\rho, \rho u, \mathcal{E}$ have to be the same as those obtained from the classical Euler system, whereas the mass densities ρ_i may differ from each other; this property can be easily verified by summing equations (2.9).

Our first test is given again by the Sod problem, and in Fig. 4.3 (left) we display the global ρ for the two different cases of no reaction and of reaction with $\Delta E = 0$; we observe complete overlapping of the two solutions, as predicted by the property above mentioned. This figure allows a qualitative comparison; an error analysis based on the computation of the relative errors between the global mass densities arising from the systems (reactive with $\Delta E = 0$ and non–reactive, respectively) was also performed. More precisely, in our test the relative errors between the reactive and the non–reactive cases in L^∞ norm for $\rho, \rho u, \mathcal{E}$ are reported in table 4.1. In these simulations, the chosen values for masses, appearing in the reactive terms, are $m_1 = 58.5$, $m_2 = 18$, $m_3 = 40$, $m_4 = 36.5$.

The difference between the density ρ_1 is shown in Fig. 4.3 (right), where we compare the numerical solutions coming from the reactive and non–reactive equations. The two profiles are clearly different, as expected, due to the effects of the chemical reaction.

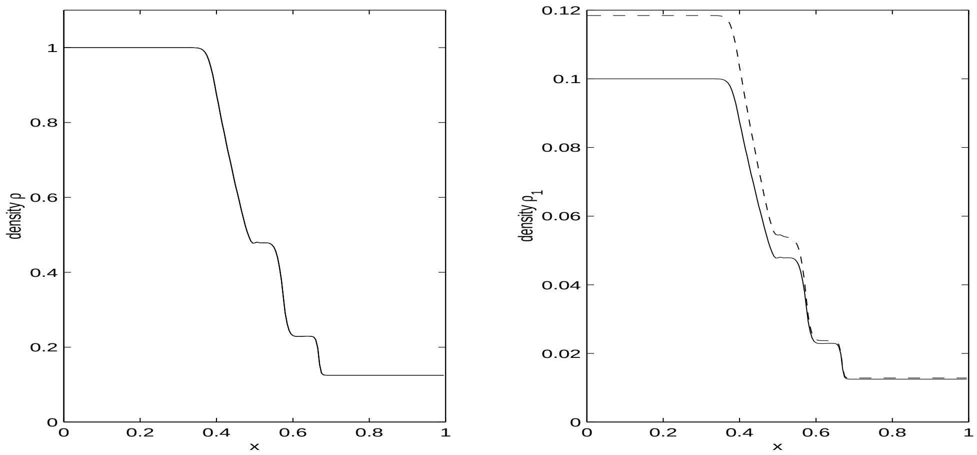


FIG. 4.3. *Comparison of the solutions of the non–reactive and reactive equations for a mixture of four gases with different masses and mass densities. The reactive equations relates to the case of $\Delta E = 0$. Total mass density $\rho = \sum_{i=1}^{4} \rho_i$ (left), mass density ρ_1 (right) for the non–reactive (continuous line) and the reactive case (dashed line).*

TABLE 4.1. *Relative errors in norm $\| \cdot \|_\infty$ for ρ, $\rho\ v$, $\mathcal{E}$ computed solving the non–reactive and the reactive (with $\Delta E = 0$) equations. Note that we used $\gamma_T = 100$.*

Grid points	ρ	$\rho\ u$	$\mathcal{E}$
200	$5.957\ 10^{-5}$	$1.406\ 10^{-4}$	$5.587\ 10^{-5}$
400	$2.761\ 10^{-5}$	$4.600\ 10^{-5}$	$2.373\ 10^{-5}$
800	$2.206\ 10^{-5}$	$4.146\ 10^{-5}$	$1.744\ 10^{-5}$

This figure shows that in the unperturbed regions where $u = 0$ the evolution of the mass density ρ_1 (and the same for all the other ρ_i) is characterized by a space-homogeneous trend [9]. The rate of variation (with respect to the non–reactive case) depends on the strength of the reactive source terms and it is mainly proportional to the values of γ_T and ΔE.

In Fig. 4.4 (left) different spatial distributions of ρ_1 at the same time instant $t = 0.07$ are plotted, each of them corresponding to increasing values of ΔE between 0 and 200. Variations with respect to the non–reactive case are more pronounced in the region ahead the shock (near $x = 0$) and less evident behind it (near $x = 1$), where the reactive source terms are closer to zero; in fact, for the chosen values of parameters and at $t = 0.07$, ahead the initial shock $\max |C_{chem}|$ increases from $5.1\ 10^{-5}\,(\Delta E = 0)$ to $8.1\ 10^{-5}\,(\Delta E = 200)$, whereas behind the shock we have $\max |C_{chem}|$ ranging from $7.7\ 10^{-7}$ $(\Delta E = 0)$ to $1.26\ 10^{-6}$ $(\Delta E = 200)$. Analogous trends can be observed in the spatial distributions of the total energy $\mathcal{E}$ when we compare the non–reactive and the reactive case (Fig. 4.4 (right)).

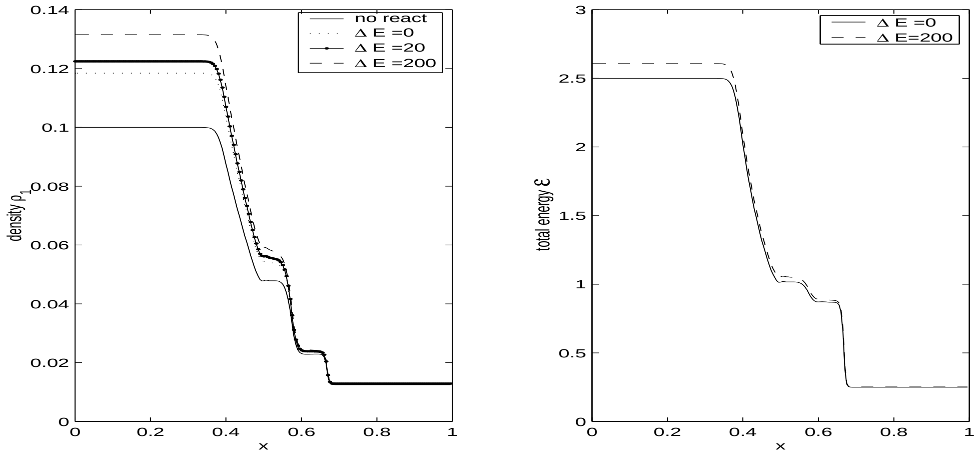


FIG. 4.4. *Reactive case with* $\gamma_T = 100$. Left: *spatial distribution at time instant* $t = 0.07$ *of density* ρ_1 *for different values of* ΔE. Right: *total energy* $\mathcal{E}$ *at the same instant* $t = 0.07$ *when* $\Delta E = 0$ *and* $\Delta E = 200$.

For larger time instants, the shock propagates into the spatial domain with a smoother profile, as it can be seen in Fig. 4.5 (top left), where the spatial distribution of ρ_1 is reported at different time instants. The flat regions for ρ_1 correspond again to the shorter intervals where u is still zero, before the arrival of the shock wave (see Fig. 4.5 top right).

The time evolution of ρ_i at point $x = 0.2$ for large t is shown in Fig. 4.5 (bottom left). It is interesting to notice the trend to a steady situation after the passage of the shock wave, due to the fact that the chemical reactive terms become smaller and smaller. Then the time evolution of the total energy $\mathcal{E}$ is reported in Fig. 4.5 (bottom right), at two points $x = 0.2$ and $x = 0.9$, showing two different trends ahead and behind the initial shock.

REMARK 2. The proposed semi-implicit discretization of the flux was chosen since it may give some benefits in the stiff cases when compared with an explicit treatment of it. As an example, let us focus again on the Sod problem for the reactive Euler equations and let us consider the stiff case in which $\gamma_T = 10000$ and $\Delta E = 200$. A qualitative comparison between the proposed semi-implicit scheme and a second-order explicit scheme for the flux is shown in fig. 4.6, where the resulting velocities are reported; in both cases the stiff reactive term is treated semi-implicitly as discussed before.

This figure shows that the semi-implicit method assures stability for larger time steps than the explicit one: indeed, whereas the two methods are in good agreement when the time step is chosen sufficiently small (left, $\Delta t = \Delta x/9$), the explicit method gives rise to unphysical quantities and unstable trends when the time step becomes larger, contrary to the semi-implicit one which is still accurate (right, $\Delta t = \Delta x/3$).

The quantitative comparison between the two methods was carried out on a SUN Blade 100 by measuring the CPU time: each time step required 12.38 sec. and 13.69 sec. for the explicit and the semi-implicit scheme respectively.

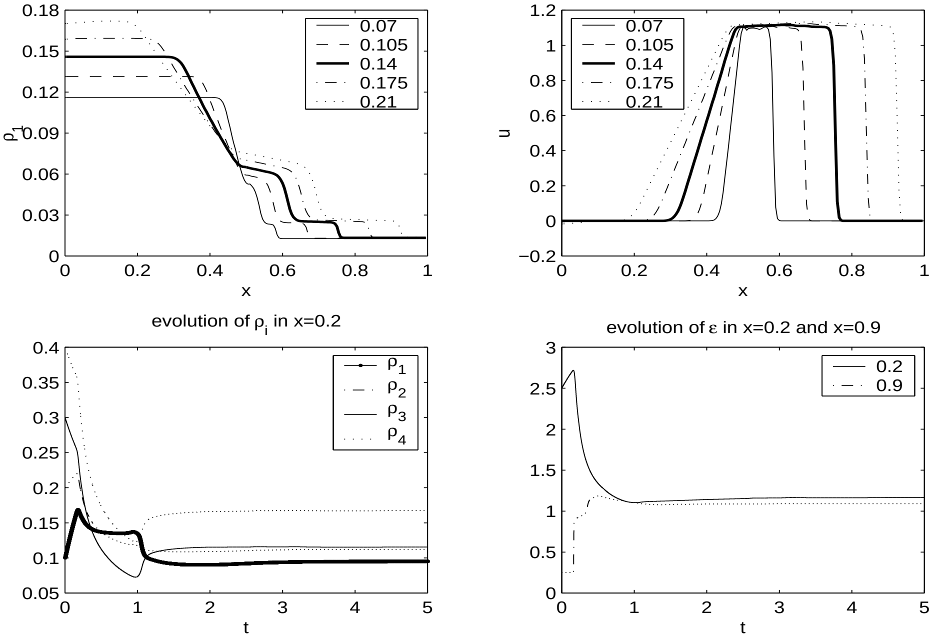


FIG. 4.5. REACTIVE CASE WITH $\Delta E = 200$ AND $\gamma_T = 100$. Top Left: *Density* ρ_1 *for different time instants from* $t = 0.07$ *to* $t = 0.21$ *with a time step of 0.035.* Top Right: *Velocity* u *for different time instants from* $t = 0.07$ *to* $t = 0.21$ *with a time step of 0.035.* Bottom Left: *Evolution of* ρ_i, $i = 1, \dots, 4$ *in* $x = 0.2$. Bottom Right: *Evolution of* $\mathcal{E}$ *in* $x = 0.2$ *and* $x = 0.9$.

Even if each time step of the semi-implicit approach is slightly more costly, the semi-implicit scheme becomes competitive when a run of several time steps is taken into account. More specifically, we considered the time interval $[0, 0.09]$ and we compared the total CPU time required by the explicit scheme with time step $\Delta t = \Delta x/4$ and the semi-implicit scheme with time step $\Delta t = \Delta x/3$. The time step chosen for the explicit case yields a CFL number close to the optimal one, that is the maximum allowed for the stability of the method; the CFL number was defined by CFL$= \max_j \rho(J_F)_j \, \Delta t/\Delta x$, where $\rho(J_F)_j$ is the spectral radius of the Jacobian matrix of the flux J_F in the j-th cell. Although a rigorous stability analysis is beyond the scope of this work, numerical experiments show that instability arises when the CFL number for the explicit scheme (with semi-implicit treatment of the reactive term) is about 1.37. Figure 4.7 shows that again the two solutions are in good agreement as in Fig. 4.6 - Left, where a smaller time step is used for both methods.

Our experiments showed that the explicit scheme required 72 time steps to cover the time interval $[0, 0.09]$ and a total CPU time of 886.60 sec., whereas for the semi-implicit one we had 54 time step and a final CPU time of 668.05 sec.. Thus, a reduction in time of about 24% for the semi-implicit discretization was observed.

Inflow problem. Finally, we present and discuss the results of a more realistic simulation and more precisely we consider the inflow problem in a channel for a four-component mixture, characterized by a positive constant inflow velocity $u(0, t)$. This may be regarded as a sample problem in the simulation of chemical reactions in industrial pipelines.

With reference to the reaction $NaCl + H_2O \rightleftharpoons NaOH + HCl$ the following

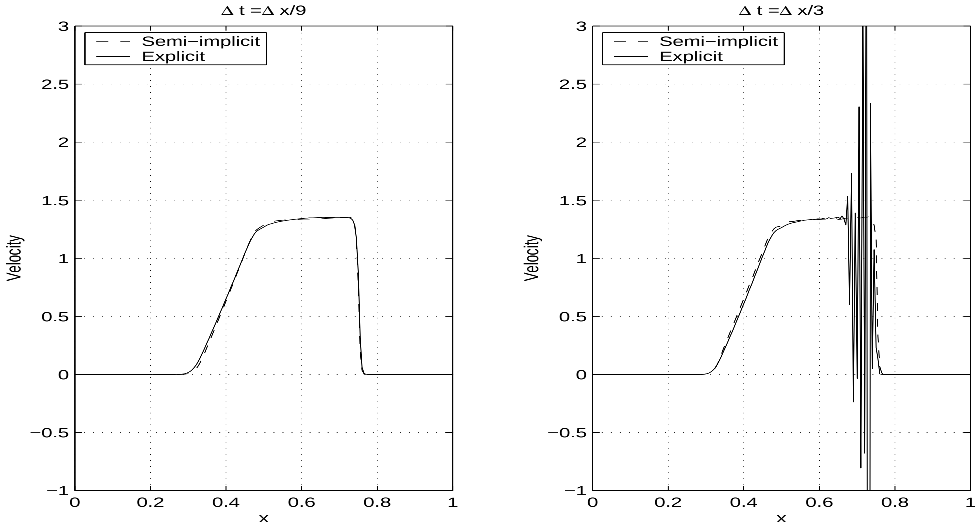


FIG. 4.6. *Velocity for the reactive case at t=0.09:* $\Delta E = 200$, $\gamma = 10000$, $\Delta x = 1/200$. *Left:* $\Delta t = \Delta x/9$. *Right:* $\Delta t = \Delta x/3$, *the explicit scheme can't reach the final time instant t=0.09 because it stops at t=0.0833 producing a negative temperature.*

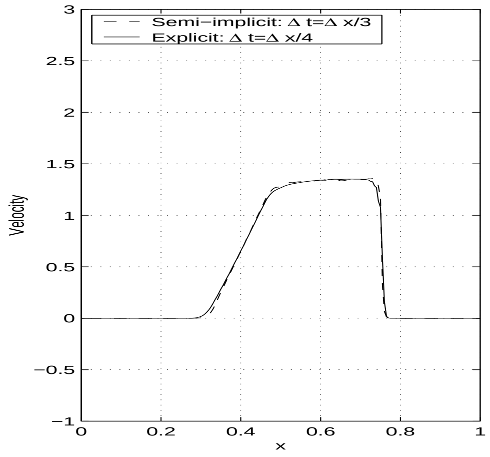


FIG. 4.7. *Velocity at t=0.09 for the reactive case:* $\Delta E = 200$, $\gamma = 10000$, $\Delta x = 1/200$. *The time step chosen for the explicit scheme is* $\Delta t = \Delta x/4$ *whereas for the semi-implicit one* $\Delta t = \Delta x/3$.

values for masses have been chosen

$$m_1 = 58.5 \qquad m_2 = 18 \qquad m_3 = 40 \qquad m_4 = 36.5 \quad (g/mol)$$

We start with the following initial data ($i = 1, \ldots, 4$)

$$\rho_i(x, 0) = m_i n_i^0 - \bar{\rho}\sin(2\pi x/L) \qquad \mathcal{E}(x, 0) = \mathcal{E}^0 + \bar{\mathcal{E}}\sin(2\pi x/L) \qquad u(x, 0) = u^0$$

where

$$n_1^0 = 0.5 \qquad n_2^0 = 0.6 \qquad n_3^0 = 0.7 \quad n_4^0 = 0.654 \quad (mol/m^3)$$
$$u^0 = 0.4\,m/s \quad \mathcal{E}^0 = 8.445\,KJ/mol \quad L = 1m \quad \bar{\rho} = 1g/m^3 \quad \bar{\mathcal{E}} = 1\,KJ/mol.$$

At the inflow boundary $x = 0$ all the unknown fields are kept fixed on the mean values

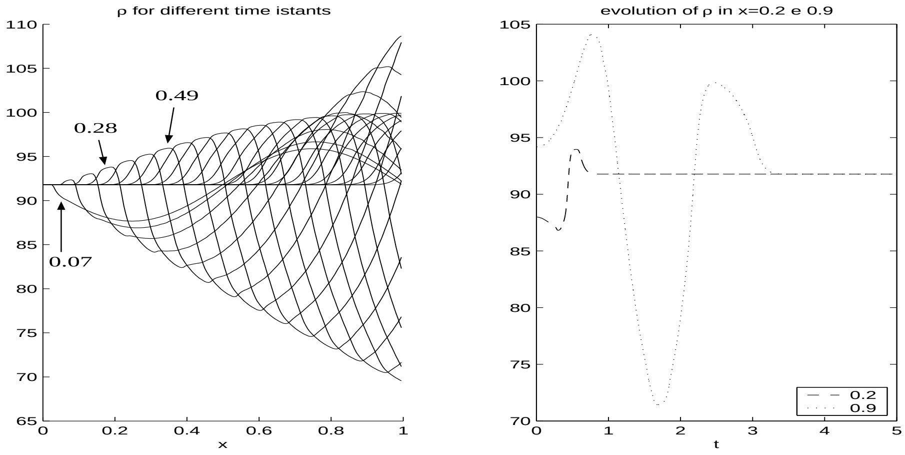


FIG. 4.8. *Non–reactive case. Spatial distribution of ρ for different time instants: from t=0.07 with a time step of 0.07 (Left). Evolution of ρ in $x = 0.2$ e $x = 0.9$ (Right).*

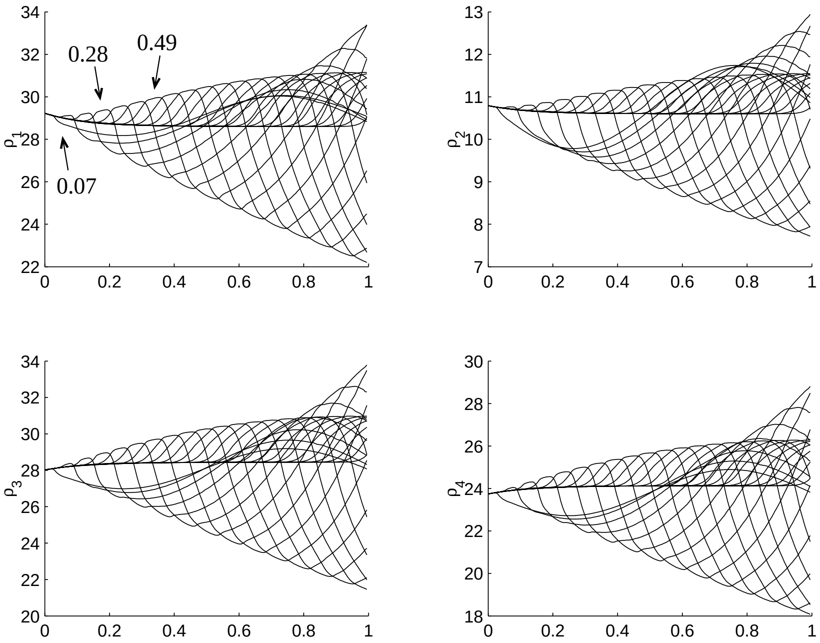
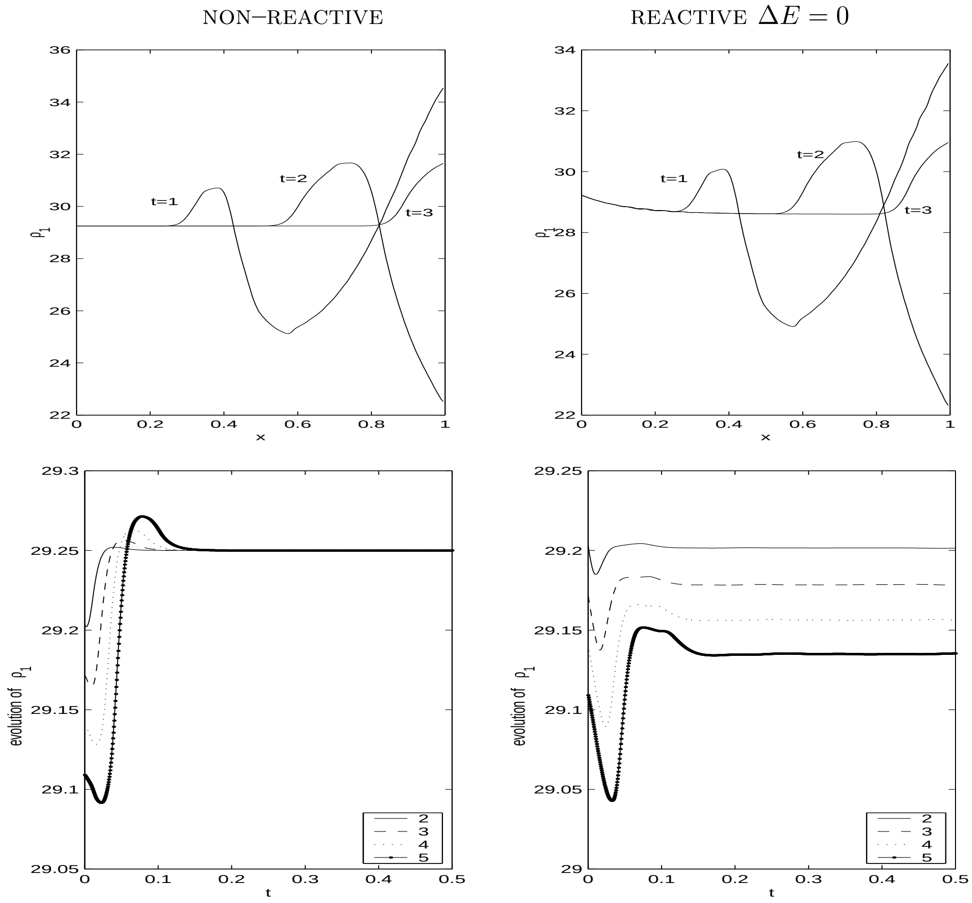


FIG. 4.9. *Reactive case with $\Delta E = 0$. Spatial distribution of ρ_i $i = 1, \ldots, 4$ for different time instants from t=0.07 with a time step of 0.07.*

of the initial profiles.

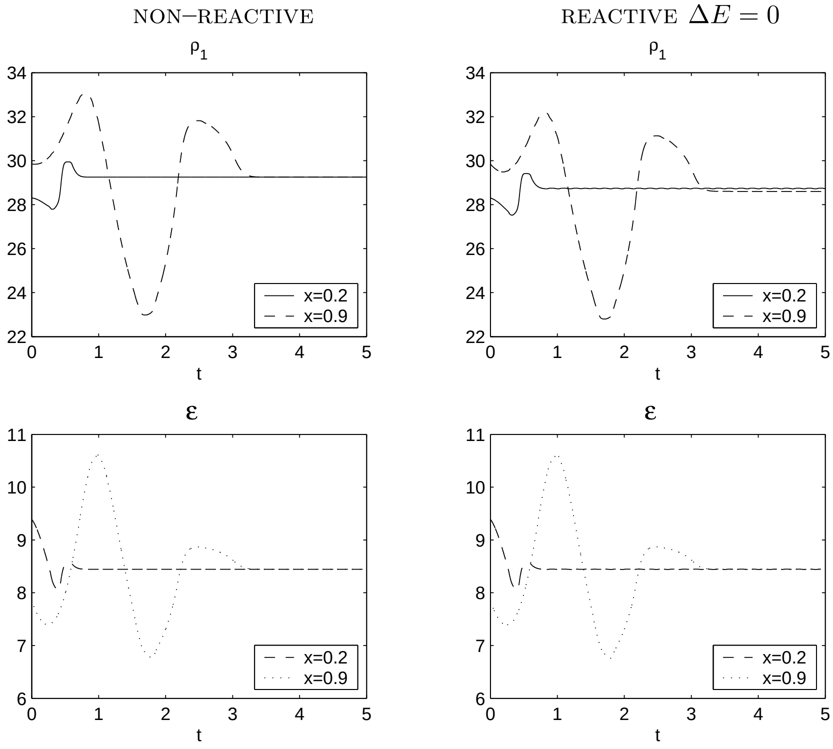
First we consider the classical Euler equations, and we can observe in Fig. 4.8 the relaxation of the total mass density ρ to the constant values that are maintained at the entrance of the channel, after a transient in which initial profiles are pushed ahead. In this figure the spatial distributions of ρ relevant to increasing time instants (left) and temporal evolutions in two internal points (right) are depicted. Analogous behaviour can be shown for all other macroscopic unknowns.

Then we put into the channel the mixture initially in chemical non–equilibrium, assuming $\Delta E = 0$ for comparison purposes. In Fig. 4.9 the spatial distribution of the four mass densities is reported and we can observe that the chemical reactions give rise to variations of the ρ_i close to the entrance of the channel, whereas for large time instants the mass densities tend to steady profiles different from the ones in the non–reactive case (see Fig. 4.10 top). Then in Fig. 4.10 (bottom) we report the temporal evolution of ρ_1 in the first cells and compare again the reactive and the non–reactive cases; the former shows a space-dependent trend, contrary to the latter. The temporal evolutions of ρ_1 and $\mathcal{E}$ in two internal points are reported in Fig. 4.11 and compared with the corresponding solutions of the classical Euler equations. It is worth noticing that equilibrium values for ρ_1 are different in the two points, whereas $\mathcal{E}$ at equilibrium has the same value throughout the interval. In fact, as already pointed out, total energy relevant to chemical processes with $\Delta E = 0$, together with global mass density ρ and velocity u, satisfies also the corresponding non-reactive classical Euler equations.


NON–REACTIVE

REACTIVE $\Delta E = 0$

FIG. 4.10. *Non–reactive case (Left). Reactive case (Right). Spatial distribution of ρ_1 at time instants $t = 1, 2, 3$ (Top). Temporal evolutions of ρ_1 for the first points of the mesh (Bottom).*



NON–REACTIVE

REACTIVE $\Delta E = 0$

FIG. 4.11. *Non–reactive case (Left). Reactive case (Right). Temporal evolution of ρ_1 in $x = 0.2$ and $x = 0.9$ (Top) Temporal evolution of $\mathcal{E}$ in $x = 0.2$ and $x = 0.9$ (Bottom)*

5. Conclusion

In this paper we have considered a class of reactive Euler equations arising from kinetic theory and we have applied a numerical strategy already known in literature but never used for these kinds of problems.

A semi–implicit scheme has been used for the time discretization for both the flux and the reactive source term. This choice is originated from the fact that there exist physical regimes in which the reactive source term becomes stiff. Explicit schemes are not convenient for the restrictions on the time step that can appear because of the stiffness of the problem. The system of algebraic equations that arises from the semi-implicit discretization chosen can be easily solved thus ensuring the effectiveness of the numerical approach.

The promising numerical results illustrate the role played by the chemical source terms and show interesting effects due to the reaction, especially on the spatial distribution of the compound, not yet available in literature for this class of reactive equations. The presented results can be regarded as a preliminary step for extensive 2D and 3D simulations of reactive models arising from kinetic theory. Moreover, further developments can be obtained from the application of this numerical scheme to other more detailed macroscopic models derived again from the kinetic description, such as Grad 13 moment equations [6]. In fact, the scheme used is quite general and can be easily extended to more complex systems in two and three spatial dimensions, where efficient methods are essential.

Acknowledgments. The authors thank Dr. G. Manzini for helpful suggestions and discussions on [5] and Dr. E. Bertolazzi and G. Manzini for making their paper [4] available. We would like to thank Prof. G. Russo for carefully reading the paper and for his helpful remarks. This work was performed in the frame of the activities sponsored by MURST (Project "Mathematical Problems of Kinetic Theories"), by CNR, by GNFM, by the University of Parma (Italy).

REFERENCES

[1] U.M. Ascher, S.J. Ruuth, and R.J. Spiteri, *Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations*, Appl. Numer. Math., 25:151–167, 1997. Special issue on time integration (Amsterdam, 1996).

[2] U.M. Ascher, S.J. Ruuth, and B.T.R. Wetton, *Implicit-explicit methods for time-dependent partial differential equations*, SIAM J. Numer. Anal., 32:797–823, 1995.

[3] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, 1994, (republished in "Classics in Applied Mathematics").

[4] E. Bertolazzi and G. Manzini, *High-order IMEX-RK finite volume methods for multidimensional hyperbolic systems*, Tech. Rep. 1202, IAN, 2000.

[5] ———, *A triangle-based unstructured finite volume method for chemically reactive hypersonic flows*, J. Comput. Phys., 166:84–115, 2001.

[6] M. Bisi, M. Groppi, G. Spiga, *Grad's distribution functions in the kinetic equations for a chemical reaction*, Continuum Mech. Thermodyn., 14:207–222, 2002.

[7] C. Cercignani, *The Boltzmann Equation and its Applications*, Springer, New York, 1988.

[8] B. Einfeldt, *On Godunov-type methods for gas dynamics*, SIAM J. Numer. Anal., 25:294–318, 1988.

[9] M. Groppi and G. Spiga, *Kinetic theory of a chemically reacting gas with inelastic transitions*, Trans. Th. Stat. Phys., 30:305–324, 2001.

[10] W. Koller, *A semi–continuous kinetic model for bimolecular chemical reactions*, J. Phys. A, 33:6081–6094, 2000.

[11] Z. Horvath, *Positivity of Runge-Kutta and diagonally split Runge-Kutta methods*, Appl. Numer. Math., 28:309–326, 1998.

[12] R.J. LeVeque, *Numerical Methods for Conservations Laws*, Birkhäuser Verlag, Basel, 1992.

[13] S.F. Liotta, V. Romano, and G. Russo, *Central schemes for balance laws of relaxation type*, SIAM J. Numer. Anal., 38:1337–1356, 2000.

[14] H. Nessyahu and E. Tadmor, *Non–oscillatory central differencing for hyperbolic conservation laws*, J. Comput. Phys., 87:408–463, 1990.

[15] L. Pareschi and G. Russo, *Implicit-explicit runge-kutta schemes for hyperbolic systems with stiff relaxation*, in Recent Trends in Numerical Analysis, vol. 3, Edited by L.Brugnano and D.Trigiante, 269–289, 2000.

[16] J. Polewczak, *A review of the kinetic modelings for non–reactive and reactive dense fluids*, Riv. Mat. Univ. Parma, (6) 4*:23–55, 2001.

[17] A. Rossani and G. Spiga, *A note on the kinetic theory of chemically reacting gases*, Physica A, 272:563–573, 1999.

[18] G.A. Sod, *A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws*, J. Comput. Phys., 27:1–31, 1978.

[19] J.L. Steger and R.F. Warming, *Flux-vector splitting of the inviscid gas dynamic equations with application to finite-difference methods*, J. Comput. Phys., 40:263–293, 1981.