Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 195 (2004) 743–772
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Studies on the accuracy of time-integration methods for the radiation–diffusion equations ☆

Curtis C. Ober [a], John N. Shadid [b,*]

[a] Computational Science Department, Sandia National Laboratories, MS 0316, P.O. Box 5800, Albuquerque, NM 87185-0316, USA
[b] Computational Science Department, Sandia National Laboratories, MS 1111, P.O. Box 5800, Albuquerque, NM 87185-1111, USA

Received 27 November 2002; received in revised form 15 September 2003; accepted 2 October 2003

Abstract

The governing equations for the radiation–diffusion approximation to radiative transport are a system of highly nonlinear, multiple time-scale, partial-differential equations. The numerical solution of these equations for very large-scale simulations is most often carried out using semi-implicit linearization or operator-splitting techniques. These techniques do not fully converge the nonlinearities of the system so as to reduce the cost and complexity of the transient solution at each time step. For a given time-step size, this process exchanges temporal accuracy for computational efficiency. This study considers the temporal-accuracy issue by presenting detailed numerical-convergence studies for problems related to radiation–diffusion simulations. In this context a particular spatial discretization based on a Galerkin finite-element technique is used. The time-integration methods that we consider include: fully implicit, semi-implicit, and operator-splitting techniques. Results are presented for the relative accuracy and the asymptotic order of accuracy of the various methods. The results demonstrate both first-order and second-order asymptotic order of accuracy for the fully implicit, semi-implicit, and the operator-splitting schemes. Additionally a second-order operator-splitting linearized-diffusion method is also presented.

1. Introduction

Recently, there has been renewed interest in the development of robust, accurate and efficient solutions to the non-equilibrium radiation–diffusion equations. These equations form a coupled highly nonlinear system of reaction–diffusion equations that exhibit solutions with wave-like propagation characteristics and multiple time and length scales [1,2,11]. For this reason the robust and accurate solution of these problems is very challenging. For these systems time integration is most often carried out by using some

☆ This work was partially supported by the ASCI program and the DOE Office of Science MICS program at Sandia National Laboratories, a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.
* Corresponding author. Tel.: +1-505-845-7876; fax: +1-505-845-7442.
E-mail addresses: ccober@sandia.gov (C.C. Ober), jnshadi@sandia.gov (J.N. Shadid).

0021-9991/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2003.10.036

type of low-order expansion for time-step linearization or operator-splitting techniques [1]. These techniques do not fully converge the nonlinearities of the system and attempt to reduce the cost and complexity of the transient solution at each time step. Linearization techniques produce semi-implicit schemes which can be solved in one step with a single linear solve and therefore by-pass the need for nonlinear solvers. Operator-splitting methods simplify the nonlinear system by separating the physics operators in each equation (or system of equations) into individual equations. These individual equations are then time integrated with solvers tailored for the particular characteristics of the physics operators. The combined effect of the split physics is then approximated by fractional-step methods [15,28], or a composition of sequential operators [18]. The advantage of these techniques is a reduction in the complexity of the component solves. For a given time-step size, all these methods clearly exchange temporal accuracy for computational efficiency. Recently there has been an interest in both re-assessing this strategy by considering the relative accuracy of these simulations, as well as reducing the cost of carrying out fully implicit solutions.

In this context recent studies have been reported that compare the accuracy of low-order-expansion linearized schemes (as well as one-step Newton methods, also known as Rosenbrock methods) with fully implicit schemes that converge the nonlinearity of the solution at each time step [12–14]. These studies show that in the absence of spatial-discretization errors, the first-order linearization schemes exhibit a significant degradation of accuracy relative to the fully implicit methods. From these types of studies, the relative numerical efficiency of these methods for a prescribed error tolerance can be studied [12]. In [14], it is further demonstrated that second-order linearization techniques can be used with an analysis of both space and time discretization errors to produce techniques that exhibit similar error bounds as a fully converged fully implicit method. Our own results obtained concurrently to [14] demonstrate this as well. However these methods are essentially an application of one-step of a general Newton scheme, and require a sufficiently accurate representation of the full Jacobian matrix and a highly accurate solve of the linear system to obtain second-order accuracy. Therefore we do not pursue these methods as a separate class of time-integration techniques to be compared in the present study although we do present results in a few specific cases.

As described above, recent comparisons of semi-implicit methods and fully implicit methods have demonstrated accuracy advantages for fully implicit methods. These studies illustrate that for a given accuracy larger time steps can be taken by the fully implicit methods. This advantage, however, can only be realized if efficient solution methods for the underlying nonlinear and linear systems can be developed. This aspect of the analysis of efficient time-integration methods using Newton–Krylov methods for radiation–diffusion problems has been studied in [4,12].

Our work extends this current literature by including additional comparisons of formally first-order and second-order operator-splitting methods as well as one-step Newton methods for relative accuracy and asymptotic order of accuracy. These comparisons are carried out in the context of Galerkin finite-element methods in contrast to finite-volume schemes of the previous studies [4,12,14]. In this study, we have investigated several aspects of the approximations to two nonlinear partial-differential equations (PDEs) related to radiative transfer. The first test problem is a thermal wave [10] which has a smooth analytic solution. Like the radiation–diffusion problem, the thermal-wave problem is a diffusion/reaction problem without an advection operator that has characteristics of a hyperbolic problem (i.e., wave propagation). Thus the wave is maintained by a balance between the diffusion and reaction operators [8]. Unlike the radiation–diffusion problem, the thermal-wave problem has a constant diffusion coefficient and therefore produces a linear component solve for the diffusion operator in an operator-splitting method. Additional details of the thermal-wave problem are given in Section 4.1.

The second test problem is the non-equilibrium radiation–diffusion problem and is the focus of this study. This problem consists of two coupled equations for the radiation-energy density and material temperature. Unlike the thermal-wave equation the radiation–diffusion problem has a nonlinear diffusion

operator in the radiation-energy density equation. Further details on this problem can be found in the literature [12,13,26], and specific details related to this study are given in Section 4.2.

The remainder of the paper is structured as follows. In Section 2, we present the Galerkin finite-element (FE) formulation in a generalized form for diffusion/reaction equations. The time-integration schemes investigated in this study are presented in Section 3 which include fully implicit, semi-implicit, and operator-splitting schemes. The specific implementations for the two test problems are discussed in Section 4. In Section 5, we present the details related to the nonlinear, linear and ODE solvers used in this study. And finally in Section 6, we present and discuss the results from the numerical experiments. We then present a number of conclusions in Section 7.

2. Galerkin FE formulation

In this section a generic notation and representation of a diffusion/reaction PDE system is presented. This representation is used to develop the Galerkin weak form of the equations as well as the discrete finite-element formulation. This generalization also provides a consistent notation for the application of the various time-integration strategies to the two numerical test problems described in Section 4.

The system of N diffusion/reaction equations is solved in a bounded open region Ω in $\Re^d$, with $d = 1, 2, 3$ and Lipschitz continuous boundary $\partial\Omega = \Gamma_m \cup \Gamma_g$ over the time interval $(0, \Im]$. The diffusion/reaction equation is then

$$\frac{\partial \chi_i}{\partial t} + \nabla \cdot \mathbf{j}_i - S_i(\chi) = 0 \quad \text{in } \Omega \times (0, \Im] \tag{1}$$

for each unknown, χ_i, with the flux vector, $\mathbf{j}_i$, defined as

$$\mathbf{j}_i = -D_i \nabla \chi_i \tag{2}$$

and $S_i(\chi)$ is some generic source term. The initial conditions are given by

$$\chi_i(\mathbf{x}, 0) = \chi_{0,i}(\mathbf{x}) \quad \text{in } \Omega. \tag{3}$$

Dirichlet boundary conditions are defined on a portion of the boundary Γ_g,

$$\chi_i = g(\mathbf{x}) \quad \text{on } \Gamma_g \tag{4}$$

as well as mixed or Robin boundary conditions on Γ_m for the remainder of $\partial\Omega$,

$$A\chi_i + B\mathbf{j}_i \cdot \mathbf{n} = C \quad \text{on } \Gamma_m. \tag{5}$$

The formal Galerkin weak form of the governing diffusion/reaction system is derived as follows. First we multiply Eq. (1) by a test function ϕ from the space $H^1_g(\Omega) = \{\psi \in H^1(\Omega) | \psi = g \text{ on } \Gamma_g\}$

$$\int_\Omega \phi \frac{\partial \chi_i}{\partial t} \mathrm{d}\Omega + \int_\Omega \phi \nabla \cdot \mathbf{j}_i \, \mathrm{d}\Omega - \int_\Omega \phi S_i(\chi) \mathrm{d}\Omega = 0. \tag{6}$$

Using the divergence theorem and flux definition from above, we obtain the resulting weak form

$$\int_\Omega \phi \frac{\partial \chi_i}{\partial t} \mathrm{d}\Omega + \int_\Omega D_i \nabla \chi_i \cdot \nabla \phi \, \mathrm{d}\Omega - \int_\Omega \phi S_i(\chi) \mathrm{d}\Omega = -\int_{\partial\Omega} \phi \mathbf{j}_i \cdot \mathbf{n} \, \mathrm{d}\Gamma \quad \forall \phi \in H^1_0(\Omega), \tag{7}$$

which can be summarized as follows. The weak form of the diffusion/reaction system is to seek $\chi \in H^1(\Omega) \times (0, \Im]$ such that Eq. (4) is satisfied and

$$\mathscr{M}_i(\dot{\chi},\phi)+\mathscr{D}_i(\chi,\phi)+\mathscr{S}_i(\chi,\phi)+\mathscr{F}_i(\phi)=0 \quad \forall\phi\in H_0^1(\Omega), \tag{8}$$

where the forms $\mathscr{M}(\cdot,\cdot)$, $\mathscr{D}(\cdot,\cdot)$, $\mathscr{S}(\cdot,\cdot)$, and the functional $\mathscr{F}(\cdot)$ are the transient (or mass) operator,

$$\mathscr{M}_i(\dot{\chi},\phi)=\int_\Omega \phi\frac{\partial\chi_i}{\partial t}\,\mathrm{d}\Omega, \tag{9}$$

the diffusion operator,

$$\mathscr{D}_i(\chi,\phi)=\int_\Omega D_i\nabla\chi_i\cdot\nabla\phi\,\mathrm{d}\Omega, \tag{10}$$

the reaction source-term operator,

$$\mathscr{S}_i(\chi,\phi)=-\int_\Omega \phi S_i(\chi)\mathrm{d}\Omega, \tag{11}$$

and the boundary-flux operator,

$$\mathscr{F}_i(\phi)=\int_{\partial\Omega}\phi\mathbf{j}_i\cdot\mathbf{n}\,\mathrm{d}\Gamma. \tag{12}$$

A Galerkin finite-element (FE) method for the generic diffusion/reaction equation restricts Eqs. (9)–(12) to a finite-element space $\Phi^h\subset H^1(\Omega)$. The discrete problem then seeks $\chi^h\subset\Phi^h$ such that $\chi^h=g$ on Γ_g and

$$\mathscr{M}_i(\dot{\chi}^h,\phi^h)+\mathscr{D}_i(\chi^h,\phi^h)+\mathscr{S}_i(\chi^h,\phi^h)+\mathscr{F}_i(\phi^h)=0 \quad \forall\phi^h\in\Phi_0^h. \tag{13}$$

In the discussion of the time-integration schemes that follow, we further simplify the notation of Eq. (13) by dropping the explicit reference to the weighting function ϕ in the operators and suppressing the use of the superscript h.

3. Overview of time-integration/nonlinear schemes

3.1. Fully implicit schemes

A fully implicit time integration of Eq. (13) evaluates all time-dependent terms at the next time level, $n+1$. The resulting equation in operator form is given by

$$\mathscr{M}_i(\dot{\chi}^{n+1})+\mathscr{D}_i^{n+1}(\chi^{n+1})+\mathscr{S}_i^{n+1}(\chi^{n+1})+\mathscr{F}_i^{n+1}=0, \tag{14}$$

where the notation for the diffusion, the source-term, and the boundary-term operator have been generalized to indicate that all the solution-dependent parameters in these operators are consistently evaluated at the new time, $n+1$. This method is termed implicitly balanced (following the terminology of Knoll et al. [11]) since the physical mechanisms (terms $\mathscr{D}_i$, $\mathscr{S}_i$, and $\mathscr{F}_i$) are all evaluated at a consistent time, in this case, t^{n+1}. As an example, the diffusion operator would be evaluated as

$$\mathscr{D}_i^{n+1}(\chi^{n+1})=\int_\Omega \mathscr{D}_i^{n+1}\nabla\chi^{n+1}\cdot\nabla\phi\,\mathrm{d}\Omega.$$

After full discretization of the system by a FE approximation, the fully implicit nonlinear system of equations is solved by an inexact Newton method and the subsequent linear systems are solved by a preconditioned GMRES method as described in Section 5.1.2. The two particular fully implicit methods

that are considered in this study are the fully implicit first-order (FI 1st) method based on a backward-Euler approximation to $\dot{\chi}^{n+1}$,

$$\dot{\chi}^{n+1} = \frac{\chi^{n+1} - \chi^n}{\Delta t}, \tag{15}$$

and the second-order time-integration technique (FI 2nd) based on the trapezoidal method,

$$\chi^{n+1} = \chi^n + \frac{\Delta t}{2}\left(\dot{\chi}^{n+1} + \dot{\chi}^n\right),$$

where one can obtain a second-order representation for $\dot{\chi}^{n+1}$,

$$\dot{\chi}^{n+1} = \frac{\chi^{n+1} - \chi^n}{\Delta t/2} - \dot{\chi}^n. \tag{16}$$

Further details on specific formulations of the finite-element approximation to the transient (or mass) operator are discussed in Sections 3.4 and 3.5. It should be noted that an alternate formulation of the trapezoidal rule (or Crank–Nicholson method) averages all terms, with the exception of the time-derivative term, at both the old time, t^n, and the new time, t^{n+1}, and then employs the backward-Euler approximation, Eq. (15), to estimate the time-derivative operator. These methods would be equivalent formulations in the absence of numerical round-off error if initiated with similar start-up strategies. In Section 5 we describe a particular starting strategy for Eq. (16).

3.2. Semi-implicit schemes

The semi-implicit schemes that we consider use low-order expansions to linearize the system at each time step to avoid the nonlinear iteration that is required by a fully implicit scheme. The price for this simplification is that the resulting schemes are most often not implicitly balanced methods (i.e., all terms are not evaluated at the same time level). As with the fully implicit schemes, the linear systems are solved by a preconditioned GMRES method which is described in Section 5.1.2.

3.2.1. Lagging (SI Lagged)

A common first-order linearization technique is to lag and evaluate the nonlinear coefficients and source terms with values of the dependent variables at the last time step. The resulting system of equations can be described as

$$\mathscr{M}_i(\dot{\chi}^{n+1}) + \mathscr{D}_i^n(\chi^{n+1}) + \mathscr{S}_i^n(\chi^n) + \mathscr{F}_i^{n+1} = 0. \tag{17}$$

As defined above the SI Lagged scheme is not an implicitly balanced method.

3.2.2. Linearized (SI Linearized)

Here we chose a specific linearization of the source terms and the diffusion operator

$$\mathscr{M}_i(\dot{\chi}^{n+1}) + \mathscr{L}_{\mathscr{D}}\left\{\mathscr{D}_i^{n+1}(\chi^{n+1})\right\} + \mathscr{L}_{\mathscr{S}}\left\{\mathscr{S}_i^{n+1}(\chi^{n+1})\right\} + \mathscr{F}_i^{n+1} = 0 \tag{18}$$

and convert the nonlinear system into a linear system of equations that is solved at each time step. As an example for this case, a simple fixed-point linearization of the diffusion operator uses values at the last time step, n, to evaluate the diffusion coefficient, such as

$$\mathscr{L}_{\mathscr{D}}\left\{\mathscr{D}_i^{n+1}(\chi^{n+1})\right\} = \mathscr{D}_i^n(\chi^{n+1}) = \int_\Omega D_i^n \nabla\chi_i^{n+1} \cdot \nabla\phi \, \mathrm{d}\Omega. \tag{19}$$

In general, linearized methods will not be implicitly balanced, and only first-order accurate unless, $\mathscr{D}_i$, $\mathscr{S}_i$, and $\mathscr{F}_i$ have simplified forms or are fully expanded to second-order with exact derivatives.

3.3. Operator-splitting schemes

Operator-splitting schemes split the operators in the governing equations and time integrate each separately and sequentially to advance to the next time step [15,28]. By construction these methods are not implicitly balanced because the different operators are evaluated at different effective time levels. To motivate these methods and define a notation for further discussion, we first present a common first-order splitting method.

3.3.1. First-order splitting (FS)

The classic two-step splitting method splits the diffusion and reaction terms and uses a sequential solution method. We consider the case when the reaction terms are integrated first followed by the diffusion terms. Advancing the solution from a solution at time level, n, to time level $n+1$ with a time-step size of Δt takes the form

$$\text{Step 1}: \quad \mathscr{M}_i(\dot{\chi}^*) + \mathscr{S}_i^*(\chi^*) = 0, \;\; \chi^*(\mathbf{x},0) = \chi^n(\mathbf{x},t) \quad \text{on } [0,\Delta t], \tag{20}$$

$$\text{Step 2}: \quad \mathscr{M}_i(\dot{\chi}^{**}) + \mathscr{D}_i^{**}(\chi^{**}) + \mathscr{F}_i^{**} = 0, \;\; \chi^{**}(\mathbf{x},0) = \chi^*(\mathbf{x},\Delta t) \quad \text{on } [0,\Delta t], \tag{21}$$

where the next time-step value is $\chi^{n+1} = \chi^{**}(\mathbf{x},\Delta t)$. In the discussion that follows we denote the solution of the split-reaction step, Eq. (20), as $\chi^* = \tilde{S}_{\Delta t}\chi^n$ and the solution of the split-diffusion step, Eq. (21), as $\chi^{**} = \tilde{D}_{\Delta t}\chi^*$ and formally represent the first-order splitting method as

$$\chi^{n+1} = \tilde{D}_{\Delta t}\tilde{S}_{\Delta t}\chi^n. \tag{22}$$

We should note that this splitting method can be second-order accurate if the operators, $\mathscr{D}_i$ and $\mathscr{S}_i$, commute (i.e., $\mathscr{D}_i\mathscr{S}_i = \mathscr{S}_i\mathscr{D}_i$) and if the solutions of Eqs. (20) and (21) are also second-order accurate [15]. Since we consider nonlinear equations and use first-order time-integration for this splitting, we expect to obtain first-order asymptotic convergence rates. For completeness we also note that the first-order splitting can be reversed (FrS) as

$$\chi^{n+1} = \tilde{S}_{\Delta t}\tilde{D}_{\Delta t}\chi^n,$$

but it is not studied here. Sportisee [24] suggests that for stiff problems the stiff operator should be evaluated last in the splitting sequence, which may reduce the leading coefficient in the error.

3.3.2. Strang splitting (SS)

Strang's operator-splitting [25] is a formally second-order scheme which applies half of the reaction physics, the full-diffusion physics, and finally another half of the reaction physics. This three-step splitting method is

$$\begin{aligned}
&\text{Step 1}: \quad \mathscr{M}_i(\dot{\chi}^*) + \mathscr{S}_i^*(\chi^*) = 0, \;\; \chi^*(\mathbf{x},0) = \chi^n(\mathbf{x},t) \quad \text{on } [0,\Delta t/2],\\
&\text{Step 2}: \quad \mathscr{M}_i(\dot{\chi}^{**}) + \mathscr{D}_i^{**}(\chi^{**}) + \mathscr{F}_i^{**} = 0, \;\; \chi^{**}(\mathbf{x},0) = \chi^*(\mathbf{x},\Delta t/2) \quad \text{on } [0,\Delta t],\\
&\text{Step 3}: \quad \mathscr{M}_i(\dot{\chi}^{***}) + \mathscr{S}_i^{***}(\chi^{***}) = 0, \;\; \chi^{***}(\mathbf{x},0) = \chi^{**}(\mathbf{x},\Delta t) \quad \text{on } [0,\Delta t/2],
\end{aligned}$$

where the next time-step value is $\chi^{n+1} = \chi^{***}(\mathbf{x},\Delta t/2)$. Thus over the time step, Δt, all the diffusion and reaction physics have been integrated. Using the operator notation introduced earlier, we can write

$$\chi^{n+1} = \tilde{S}_{\Delta t/2}\tilde{D}_{\Delta t}\tilde{S}_{\Delta t/2}\chi^n. \tag{23}$$

The operators of the Strang splitting can be reversed (SrS) as

$$\chi^{n+1} = \tilde{D}_{\Delta t/2}\tilde{S}_{\Delta t}\tilde{D}_{\Delta t/2}\chi^n \tag{24}$$

to reduce the size of the diffusion time step from Δt to $\Delta t/2$. It should be noted that for formal second-order accuracy the component solves must be at least second-order accurate.

3.3.3. Marchuk splitting (MS)

Marchuk splitting [15, p. 478] is similar to the first-order splitting, but the order of the operators is alternated every time step:

$$\chi^{n+2} = (\tilde{S}_{\Delta t}\tilde{D}_{\Delta t})(\tilde{D}_{\Delta t}\tilde{S}_{\Delta t})\chi^n. \tag{25}$$

This alternating of the order of the operators produces a second-order scheme. For this study even time steps perform $\tilde{S}_{\Delta t}\tilde{D}_{\Delta t}$ and odd time steps perform $\tilde{D}_{\Delta t}\tilde{S}_{\Delta t}$. Marchuk operator-splitting is similar to Strang splitting for two operators but it is applied over $2\Delta t$. Marchuk splitting can be extended to multiple operators over Δt, as in the case of three operators, $\tilde{C}$, $\tilde{D}$, $\tilde{S}$:

$$\chi^{n+1} = \tilde{C}_{\Delta t/2}\tilde{D}_{\Delta t/2}\tilde{S}_{\Delta t}\tilde{D}_{\Delta t/2}\tilde{C}_{\Delta t/2}\chi^n.$$

This situation of three or more operator splittings is not investigated in this study.

3.3.4. Romero splitting (RS)

The Romero-splitting technique [19] is a formally second-order method that is similar to Strang-reversed splitting. However this technique uses first-order time integration for the diffusion solves. In addition the Romero splitting solves the diffusion terms with an alternating explicit and implicit solve (i.e., forward and backward Euler)

$$\chi^{n+1} = \tilde{D}^{\text{imp}}_{\Delta t/2}\tilde{S}_{\Delta t}\tilde{D}^{\text{exp}}_{\Delta t/2}\chi^n. \tag{26}$$

Over two successive time steps, the first-order implicit solve is followed by the first-order explicit solve. Intuitively, by combining these two solves sequentially one can produce a classic second-order central-difference approximation if the time-step sizes are the same. In general, Romero splitting assumes that the reaction has a fast-time scale compared to the diffusion, and the reaction can be solved exactly. However in practice, non-exact numerical solutions to the reaction operator still produce second-order solutions in our studies. Romero splitting is unconditionally stable for linear problems such as the 2D heat equation, despite the explicit operator. This characteristic is similar to ADI schemes, where the individual steps are only conditionally stable but together produce an unconditionally stable scheme [20].

3.4. Efficient FE splitting methods

In order to motivate the development of efficient FE splitting methods we consider the reaction step of a generic operator-splitting method. In the reaction step, the system

$$\mathscr{M}_i(\dot{\chi}^*) + \mathscr{S}^*_i(\chi^*) = 0 \tag{27}$$

must be solved with suitable initial conditions over the appropriate time interval. In the case of a FE method, this system would produce a large sparse nonlinear system of equations of dimension $N = N_{\text{unknowns}} \times N_{\text{nodes}}$. In general this would be expensive to solve but could be accomplished with a Newton–Krylov method. However this methodology would appear to be somewhat inconsistent with the standard operator-splitting philosophy of efficient component solves for each of the respective operators. It

can be easily shown that an alternate approach, which employs a group FE expansion [7] for this system, reduces Eq. (27) to a system of ODEs of dimension $N = N_{\text{unknowns}}$ to be solved at each node of the FE mesh. In practice we implement the solution of this local ODE system with a stiff ODE solver (see Section 5.1.3) and employ very strict error tolerances to minimize error accumulation over the intermediate reaction time step. This procedure which localizes the reaction system solve to a local FE node has an additional benefit. In the results that we present in Section 6, it is demonstrated that these methods control oscillations at the wavefront. This characteristic behavior is similar to the behavior of diagonalized operators in a fully implicit technique as described next.

3.5. Diagonalized operators for control of oscillations

In the results that we present in Section 6 on the solutions for the radiation–diffusion system, the fully implicit and semi-implicit methods exhibit oscillatory behavior at the wavefront. To control this oscillation we employ diagonalized mass and source operators. A particular feature of diagonalizing (or lumping) the mass and source-term operators for a nonlinear parabolic equation of the form of Eq. (13) is that for a diffusion coefficient, $D > 0$, and a non-negative source term, $S \geqslant 0$, there is a positivity result that holds [17,27]. In addition for the case $S = 0$ these methods posses a discrete maximum principle [17,27]. This is in contrast to the consistent-mass and consistent-source-term-operator methods that do not exhibit these properties. The diagonalization process can be obtained equivalently from a lumping procedure or by integrating with one-point quadrature at each FE node [17,27]. This process explicitly decouples the mass and source-term operators spatially from the other FE nodes in the mesh. For our system of nonlinear parabolic equations, this formulation was found to eliminate oscillations at the wavefront as well. This ability to produce an oscillation-free solution is not only advantageous in coupled simulations with complex equations of state but it also allows us to perform direct convergence-study comparisons between the operator-splitting methods and the fully implicit and semi-implicit methods.

4. Description of test problems

In this section we present a description of the two test problems that are used to numerically evaluate the relative accuracy and asymptotic order of accuracy of the time-integration techniques described above. In addition to the description of the test problems, initial conditions and boundary conditions, we also present the operator-specific linearizations that are used in the semi-implicit linearized (SI Linearized) time-integration technique.

4.1. Thermal wave

The first test problem that we will describe is associated with the solution to the time-dependent heat equation with a nonlinear source term. This test problem of Knio et al. [10] provides a numerical example with a smooth analytic solution in the form of a propagating wave. The nonlinear diffusion/reaction equation is

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + \frac{8}{\delta^2} T^2 (1 - T), \tag{28}$$

where the boundary conditions are $T(x = -\infty, t) = 1$ and $T(x = \infty, t) = 0$. The parameter, $\delta > 0$, can be freely selected but does not change the ratio of time scales between the diffusion and the reaction terms. The analytic solution is

$$T(x,t) = \frac{1}{2}\left(1 - \tanh\left[\frac{x - 2t/\delta}{\delta}\right]\right), \tag{29}$$

where we have selected $\delta = 1$.

This one-dimensional problem is modeled with linear elements in the FE mesh, covering the domain $-10 \leqslant x \leqslant 10$. The simulation was integrated to $t_{\text{final}} = 1.024$ with four different mesh spacings, $\Delta x = 0.04$, 0.02, 0.01, and 0.005. We can relate the thermal-wave problem with the general diffusion/reaction formulation (i.e., Eqs. (8)–(12)) as follows:

$$\chi = T,$$

$$\mathbf{j}_i = -D_i \nabla \chi_i = -\nabla T, \tag{30}$$

$$\mathscr{D}_i^{n+1}(\chi^{n+1}) = \int_\Omega \nabla T^{n+1} \cdot \nabla \phi \, \mathrm{d}\Omega, \tag{31}$$

$$\mathscr{S}_i^{n+1}(\chi^{n+1}) = -\int_\Omega \frac{8}{\delta^2}(T^{n+1})^2(1 - T^{n+1})\phi \, \mathrm{d}\Omega. \tag{32}$$

For the SI Linearized runs, the source-term linearization is a simple fixed-point linearization of $\mathscr{S}(T)$ such that

$$\mathscr{L}_{\mathscr{S}}\left\{\mathscr{S}_i^{n+1}(\chi^{n+1})\right\} = -\int_\Omega \frac{8}{\delta^2}(T^n)^2(1 - T^{n+1})\phi \, \mathrm{d}\Omega. \tag{33}$$

Since the diffusion coefficient is constant, no linearization of the diffusion operator is required.

4.2. Non-equilibrium radiation diffusion

As described earlier, we are primarily interested in studying the temporal accuracy of time-integration methods applied to the non-equilibrium radiation–diffusion problem. This problem is defined by a nonlinear coupled set of equations with a solution that contains a very strong gradient at the wavefront. The governing equations for the radiation–diffusion approximation are given by

$$\frac{\partial E}{\partial t} - \frac{\partial}{\partial x}\left(cD_{\mathrm{r}}\frac{\partial E}{\partial x}\right) = c\sigma_{\mathrm{a}}\left(aT^4 - E\right) \quad \text{in } \Omega, \tag{34}$$

$$\frac{\partial (C_v T)}{\partial t} = -c\sigma_{\mathrm{a}}(aT^4 - E) \quad \text{in } \Omega, \tag{35}$$

and are solved in the domain $\Omega = [0,1]$ for $t \in (0,1]$, where E is the radiation-energy density, T is the material temperature, c is the speed of light, D_{r} is the radiation–diffusion coefficient, σ_{a} is the inverse absorption mean free path, $a = 4\sigma_{\mathrm{SB}}$ where σ_{SB} is the Stefan–Boltzmann constant, and C_v is the material heat capacity. Following [12] a system of units has been selected so that $C_v = c = a = 1$. For the computational problem under consideration, the domain is $0 \leqslant x \leqslant 1$, and the initial conditions are

$$E(x,0) = E_0(x) \quad \text{in } \Omega, \tag{36}$$

$$T(x,0) = T_0(x) \quad \text{in } \Omega, \tag{37}$$

where $E_0(x) = 10^{-5}$ and $T_0(x) = E_0(x)^{1/4}$. The boundary conditions are of mixed type and can be written as

$$AE + B(\mathbf{j}\cdot\mathbf{n}) = C \quad \text{on } \partial\Omega\ \forall t, \tag{38}$$

where $A = 1/4$, $B = -1/2$, and $C = 1$ for $x = 0$; and $A = 1/4$, $B = 1/2$, and $C = 0$ for $x = 1$. In the numerical results that are presented, we stabilize the material-temperature equation, Eq. (35), by adding a very small amount of diffusion, $k = 5.0 \times 10^{-7}$. Spatial and temporal convergence studies were then carried out to verify that the added level of diffusion was negligible by varying k from 10^{-3} to 10^{-7}. Therefore the thermal diffusion term was added for all the presented results with a value of $k = 5.0 \times 10^{-7}$, which is approximately 1% of D_r at the leading edge of the wavefront and orders of magnitude smaller behind the wave.

The absorption cross-section is modeled by $\sigma_a = T^{-3}$, and the radiation–diffusion coefficient is $D_r = 1/(3\sigma_a)$ from simple isotropic theory. However in regions of strong gradients, the theory can fail and allow the flux of energy to move faster than the speed of light. A flux-limiting can be applied to D_r to prevent this unphysical behavior, and we use [3,12,13,16]

$$D_r = \frac{1}{3\sigma_a + \frac{1}{E}\left|\frac{\partial E}{\partial x}\right|}. \tag{39}$$

The identification of the radiation–diffusion problem with the general diffusion/reaction form of Eqs. (8)–(12) is given as

$$\chi = [E, T]^T,$$

$$\mathbf{j}_i = -D_i\nabla\chi_i = \begin{cases} -D_r\nabla E & \text{for } i = E, \\ 0 & \text{for } i = T, \end{cases} \tag{40}$$

$$\mathscr{D}_i^{n+1}(\chi^{n+1}) = \begin{cases} \int_\Omega D_r^{n+1}\nabla E^{n+1}\cdot\nabla\phi\,d\Omega & \text{for } i = E, \\ 0 & \text{for } i = T, \end{cases} \tag{41}$$

$$\mathscr{S}_i^{n+1}(\chi^{n+1}) = \begin{cases} -\int_\Omega \phi\sigma_a^{n+1}[(T^{n+1})^4 - E^{n+1}]d\Omega & \text{for } i = E, \\ \int_\Omega \phi\sigma_a^{n+1}[(T^{n+1})^4 - E^{n+1}]d\Omega & \text{for } i = T. \end{cases} \tag{42}$$

The linearization of the diffusion and source-term operators for SI Linearized are defined as

$$\mathscr{L}_{\mathscr{D}}\{\mathscr{D}_i^{n+1}(\chi^{n+1})\} = \begin{cases} \int_\Omega D_r^n\nabla E^{n+1}\cdot\nabla\phi\,d\Omega & \text{for } i = E, \\ 0 & \text{for } i = T, \end{cases} \tag{43}$$

$$\mathscr{L}_{\mathscr{S}}\{\mathscr{S}_i^{n+1}(\chi^{n+1})\} = \begin{cases} -\int_\Omega \phi\sigma_a^n[(T^n)^3T^{n+1} - E^{n+1}]d\Omega & \text{for } i = E, \\ \int_\Omega \phi\sigma_a^n[(T^n)^3T^{n+1} - E^{n+1}]d\Omega & \text{for } i = T. \end{cases} \tag{44}$$

This first-order linearization of the diffusion and source-term operators is non-unique and follows [11,13]. Alternative first-order accurate linearizations are of course possible. In our numerical experiments that follow, we have presented results only for the linearization described above. However we have also carried out computations with alternate linearizations of the source-term operator, $\mathscr{S}$, that include a more accurate second-order Taylor series expansion of the emission term, T^4, as well as a second-order Taylor series expansion of the entire source term. These alternate linearizations are still first-order since the dif-

fusion-operator linearization in Eq. (43) is only first-order accurate (second-order linearization of D_r in Eq. (39) would be very difficult). In addition, these alternate linearizations do not change the relative accuracy of the SI Linearized scheme in any appreciable way.

5. Numerical solutions

In this section we give an overview of the implementation of the nonlinear, linear and ODE solvers, the time-level advancement, and the selection of reference solutions. All the tolerances in this study were selected to be very restrictive to eliminate uncontrolled sources of errors from entering into the comparisons of the relative accuracy and asymptotic order-of-accuracy studies. In practice the tolerances are unnecessarily restrictive for normal computational simulations, however we used them as a means of obtaining careful estimates of the particular time-integration errors of interest. As remarked earlier, studies that consider the efficiency of solving these systems of equations can be found in [4,12]. Similar to these studies, we also employ Newton–Krylov methods as the nonlinear/linear solution algorithms and thus our results can be considered directly applicable to assessing the accuracy of the fully implicit time-integration techniques as described in [4,12].

5.1. Solution methods

5.1.1. Nonlinear solver

To solve the nonlinear system of equations, we employ an inexact Newton method described by Shadid et al. [23]. We provide a brief overview here for reference purposes. The nonlinear systems, which occur in our studies (i.e., Eqs. (14) and (21)), can be written as

$$F(\chi) = 0.$$

Given an initial approximate solution, χ_0, the standard Newton's method will solve the Newton equation

$$J(\chi^k)\Delta\chi^k = -F(\chi^k)$$

to determine the update to the solution using $\chi^{k+1} = \chi^k + \Delta\chi^k$, where $J(\chi^k)$ is the Jacobian and k is the Newton iteration index.

In an inexact Newton method [6], the Newton equation is solved to a tolerance set by an inexact Newton condition

$$\|F(\chi^k) + J(\chi^k)\Delta\chi^k\| \leqslant \eta^k \|F(\chi^k)\|$$

for some $\eta^k \in [0,1)$, where $\|\cdot\|$ is a norm of choice. This formulation naturally allows the use of an iterative linear-algebra method. η^k is selected to force the residual of the Newton equation to be small. For the runs in this study, the linear subproblems are solved with η^k chosen as a constant and equal to the linear-solver normalized-residual tolerance: 10^{-10} for the thermal-wave problem and 10^{-12} for the radiation–diffusion problem.

The convergence of the nonlinear Newton iteration is determined by two criteria. The first is a sufficient reduction in the relative residual, $\|F^k\|/\|F^o\|$, to below 10^{-2}. This requirement, in general, is easily satisfied. The main controlling convergence criteria is then based on a sufficient decrease in a weighted norm of the update vector from Newton's methods. This criteria requires that the correction, $\Delta\chi_i^k$, for any variable, χ_i, is "small" compared to its magnitude, $|\chi_i^k|$, and is given by

$$\sqrt{\frac{1}{N_u}\sum_{i=1}^{N_u}\left[\frac{|\Delta\chi_i|}{\varepsilon_r|\chi_i| + \varepsilon_a}\right]^2} < 1,$$

where N_u is the number of unknowns, ε_r is the relative error tolerance between the variable correction and its magnitude, and ε_a is the absolute error tolerance of the variable correction. In this criteria ε_a essentially sets the magnitude of components that are to be considered to be numerically zero. In this study, the relative-error and absolute-error tolerance are 10^{-8} and 10^{-10}, respectively.

5.1.2. Linear solver

The linear systems generated by the semi-implicit and Newton iteration of the fully implicit schemes are iteratively solved using preconditioned Krylov methods. In this study the Aztec Library [9] of parallel Krylov methods is used. The specific Krylov method is the restarted generalized minimum residual (GMRES) method. For preconditioning, the additive Schwarz domain decomposition technique with incomplete LU factorization sub-domain solvers and row-sum scaling is used. The number of Krylov vectors used before restarting is 200, and as noted earlier the linear-solver normalized-residual tolerance is 10^{-10} for the thermal-wave problem and 10^{-12} for the radiation–diffusion problem.

5.1.3. ODE solver

During the solution of the operator-splitting schemes, the equations for the source-term operator, $\tilde{S}_{\Delta t}$, contain no spatial derivatives for the test problems investigated here. The ODE equations at each FE node is time integrated with the CVODE library [5]. In CVODE the following options are used: (1) backward differentiation formula (BDF) for the time advancement, (2) Newton iteration for the nonlinear solves, (3) a direct method with a dense treatment of the Jacobian where the Jacobian is formed by a difference-quotient routine (since we are dealing with small systems, one or two equations at each node, a direct method is not costly), (4) normal mode where smaller time steps can be taken (i.e., subcycling) to obtain the solution at the next time step, $n+1$, and (5) the relative and absolute error tolerances of 10^{-10}.

5.1.4. Time-level advancement

For each run, the time-step size was kept constant throughout the run except for two situations. The first situation is during the startup phase of second-order time-integration schemes. To obtain second-order accuracy, solution and time-derivative information at two time levels are required. However at the first time step there is only one time level of information and therefore a first-order time integration is used to start-up the algorithm. To reduce the error of the start-up phase, the following ramp-up of the time-step size is used to integrate over the first s time steps:

$$\Delta t^n = \begin{cases} \Delta t_d/2^s & \text{for } n = 0, 1, \\ 2\Delta t^{n-1} & \text{for } n \leqslant s, \\ \Delta t_d & \text{for } n > s, \end{cases}$$

where Δt_d is the desired constant time-step size, and s is the smallest integer which makes the inequality, $\Delta t_d/2^s < (\Delta t_d)^2$, true. This inequality makes the error from the start-up phase the same order as the error from the constant time steps that follow in the remainder of the time-integration process.

The second situation is the occasion when a time step fails to converge by either the nonlinear or linear solver. When this does happen, it is usually for a small number of time steps during the run. To handle this and allow the solution to continue, we reduce the time-step size by a factor of 2 and solve at that time-step size. The next time-step size is doubled to return to Δt_d.

5.2. Solution comparison

To compare the accuracy of the various methods, we use a componentwise relative L_2 norm of the error, defined as summing over all the unknowns, v, in the governing equations and over all the nodal points, j, in the domain

$$\left\|\chi-\chi^{\mathrm{ref}}\right\|=\sum_{v}\left[\left\{\frac{1}{N}\sum_{j}^{N}\left(\chi_{v}-\chi_{v}^{\mathrm{ref}}\right)^{2}\right\}^{1/2}\Bigg/\left\{\frac{1}{N}\sum_{j}^{N}\left(\chi_{v}^{\mathrm{ref}}\right)^{2}\right\}^{1/2}\right].$$

By normalizing the L_2 norm of the error, the relative differences in scaling (magnitude) between the unknowns, v, is accounted for and the error of one unknown does not dominate. This componentwise relative error norm can be shown to be equivalent to a standard weighted L_2 norm.

5.3. Reference solutions

Any discrete numerical approximation of a time-dependent PDE will produce both spatial and temporal discretization errors in the numerical solution. To determine the relative magnitude of these errors and to estimate the order of accuracy of the individual methods, it is possible to carry out numerical-convergence studies. In this section we state the error models that form the basis for these numerical studies. From finite-element theory for transient nonlinear parabolic problems [17,27], we expect the global error between the numerical solution and the exact solution to be characterized by

$$\|\chi-\chi^{\mathrm{exact}}\|=\mathrm{o}(\Delta x^{p})+\mathrm{o}(\Delta t^{q}),$$

where p is the asymptotic spatial order of accuracy, and q is the asymptotic temporal order of accuracy. We formally assume that we can represent the global error as

$$e=\|\chi-\chi^{\mathrm{exact}}\|=C_{x}\Delta x^{p}+C_{t}\Delta t^{q}, \tag{45}$$

where in the asymptotic region of convergence both, C_x and C_t, are independent of Δx and Δt. We further assume, that to the lowest order, a pointwise asymptotic error expansion of the type

$$\chi=\chi^{\mathrm{exact}}+\tilde{C}_{x}\Delta x^{p}+\tilde{C}_{t}\Delta t^{q}+R(\chi) \tag{46}$$

holds for a sufficiently regular exact solution where $R(\chi)$ represents the higher-order terms, and the expansion coefficients are again assumed independent of the temporal and spatial-mesh sizes. The numerical studies that follow will support these assumptions.

An estimate of the discretization errors can be determined by comparing against a suitable reference solution, $\tilde{e}=\|\chi-\chi^{\mathrm{ref}}\|$. The present study considers three reference solutions to estimate errors: (1) a discrete evaluation of the exact solution (if available), (2) the best-resolution (finest-resolution) solution, and (3) an extrapolated solution based on Richardson extrapolation. If the exact solution is used as the reference solution, the error can be defined by Eq. (45). A refinement study can be used to determine an estimate of q, termed $\tilde{q}$, by evaluating the error for two time-refinement levels, e_i and e_j, where i and j indicate the level. The estimated temporal order of accuracy can then be found by

$$\tilde{q}=\frac{\ln e_{i}-\ln e_{j}}{\ln\Delta t_{i}-\ln\Delta t_{j}}=\frac{\ln\epsilon_{ij}}{\ln r_{ij}},$$

where $\epsilon_{ij}=e_i/e_j$, $r_{ij}=\Delta t_i/\Delta t_j$ is the temporal refinement ratio, and where it is assumed that $C_x\Delta x^p\ll C_t\Delta t^q$. However the errors, e_i and e_j, contain both spatial and temporal contributions, and as the time-step size, Δt, is reduced the spatial error, $C_x\Delta x^p$, will become dominant causing the error to plateau (see for example Figs. 4 and 5).

For many problems the exact solution is not available. If the best resolution is used as the reference solution and can be expressed by the expansion of Eq. (46), we can write

$$\chi_{i}-\chi_{m}^{\mathrm{best}}=(\chi_{i}-\chi^{\mathrm{exact}})-(\chi^{\mathrm{best}}-\chi^{\mathrm{exact}})=\tilde{C}_{t}\Delta t_{i}^{q}-\tilde{C}_{t}\Delta t_{m}^{q},$$

where m is the best resolution and the spatial discretization is constant during the temporal refinement. Here the spatial errors cancel each other and the local error, $\chi_i - \chi_m^{\text{best}}$, is only dependent on Δt, and will not plateau as Δt is refined. If the global-error norm is based on the best-resolution solution at Δt_m, then we can approximate the error with

$$\tilde{e}_i = \|\chi_i - \chi_m^{\text{best}}\| = \|\tilde{C}_t\|\left[\Delta t_i^q - \Delta t_m^q\right] = \tilde{\tilde{C}}_t\left[\Delta t_i^q - \Delta t_m^q\right].$$

If we neglect the higher-order terms and assume the leading coefficient, $\tilde{\tilde{C}}_t$, is constant during refinement, the error ratio becomes

$$\epsilon_{ij} = \tilde{e}_i/\tilde{e}_j = \frac{\Delta t_i^q - \Delta t_m^q}{\Delta t_j^q - \Delta t_m^q}$$

and the apparent temporal order of accuracy, $\tilde{q}$, becomes

$$\tilde{q} = \left\{ \ln \frac{\Delta t_i^q - \Delta t_m^q}{\Delta t_j^q - \Delta t_m^q} \right\} \Big/ \left\{ \ln \frac{\Delta t_i}{\Delta t_j} \right\},$$

which is a good approximation if Δt_i and Δt_j are not near Δt_m.

However we point out that many studies have used best-resolution solutions which are near their refinement studies, because of the cost associated with computing solutions on very fine meshes and/or time-step sizes. As a numerical example of the inaccuracies associated with the choice of using a too coarse best-resolution solution, let us consider $\Delta t_i = 2\Delta t_j = 4\Delta t_m$. For a first-order discretization ($q = 1$), the apparent order would be $\tilde{q} = 1.585$ (i.e., 58.5% error), and for a second-order discretization ($q = 2$), the apparent order would be $\tilde{q} = 2.322$ (i.e., 16.1% error). If the best resolution is a factor of 10 smaller ($\Delta t_i = 2\Delta t_j = 20\Delta t_m$), the apparent order is much better: $\tilde{q} = 1.078$ for first-order and $\tilde{q} = 2.011$ for second-order. This effect is described in the studies of the thermal-wave test problem.

An extrapolated solution can be formed from a simple Richardson-extrapolated procedure [21] which relies on the pointwise expansion described above. Using two solutions at different resolutions (subscripts indicate resolution levels), and neglecting higher-order terms,

$$\chi_i = \chi^{\text{exact}} + \tilde{C}_x \Delta x^p + \tilde{C}_t \Delta t_i^q,$$

$$\chi_j = \chi^{\text{exact}} + \tilde{C}_x \Delta x^p + \tilde{C}_t \Delta t_j^q,$$

and defining the temporal-extrapolated solution as the exact solution plus the spatial error,

$$\chi^{\text{extrap}} = \chi^{\text{exact}} + \tilde{C}_x \Delta x^p,$$

we can solve for the extrapolated solution,

$$\chi_j^{\text{extrap}} = \chi_j + (\chi_j - \chi_i)/(r_{ij}^q - 1). \tag{47}$$

If the extrapolated solution of Eq. (47) is used as the reference solution, the global error is

$$\tilde{e}_i = \|\chi_i - \chi_j^{\text{extrap}}\| = \|\tilde{C}_t\| \Delta t_i^q = \tilde{\tilde{C}}_t \Delta t_i^q.$$

Again the spatial errors have canceled out, leaving only the temporal errors, $\tilde{\tilde{C}}_t \Delta t_i^q$ and higher-order terms. However the temporal order of accuracy, q, in Eq. (47) must be selected in order to obtain the extrapolated solution (i.e., $q = 1$ for first-order or $q = 2$ for second-order time integration). If q is unknown, an iteration

is required between the estimated $\tilde{q}$ and the generation of χ^{extrap} until the estimated $\tilde{q}$ matches the actual asymptotic order q. An alternative approach is to compute an apparent order of accuracy, $\tilde{q}$, which can be obtained using three numerical solutions, (χ_i, χ_j, χ_k) [21].

6. Numerical experiments

The numerical results that are presented in this section, for both the thermal-wave problem and the radiation–diffusion problem, were obtained by the solution methods described in Section 5 as implemented in the MPSalsa finite-element transport/reaction code [22]. The computational runs were carried out on a modest number of processors, (10–64) of the Sandia-Intel Tflop machine or a single processor of an SGI Octane. The required CPU times for these runs varied from a few seconds to about 24 hours, for the coarsest to finest resolution simulations. As described above overly restrictive convergence criteria in the nonlinear and linear solvers were enforced to eliminate extraneous numerical errors from influencing the estimation of the time-integration errors that are reported. For this reason no relative CPU performance data is presented. In the process of this investigation over 1000 solutions of the thermal-wave problem, and approximately 500 solutions of the radiation–diffusion problem were carried out.

6.1. Thermal wave

As a means of exploring the usefulness of the various reference solutions described above for error estimation, we begin with a study of the thermal-wave problem which has a smooth analytical solution. To assess the utility and accuracy of the best-resolution and extrapolated reference solutions, we will generate the exact error, e, using the analytic solution evaluated at the appropriate discrete points corresponding to the numerical solution. The approximate error, $\tilde{e}$, will be generated from other reference solutions (i.e., best resolution and extrapolated) and will be compared to the exact error.

6.1.1. Solution profiles

In Fig. 1, temperature profiles for several time-integrations methods are shown. Profiles are shown for two times, $t = 0.512$ and $t = 1.024$, and the spatial and temporal resolutions for these solutions are respectively, $\Delta x = 0.04$ and $\Delta t = 0.064$. For all the thermal-wave runs, the consistent-mass matrix was used unless otherwise noted. By $t = 1.024$, the thermal wave has traveled approximately half its width in the positive x-direction. As the simulation progresses an accumulation of phase error is apparent as the various solutions deviate from the exact solution. An additional modification of the wavefront profile is also apparent for the first-order methods which indicates the existence of different dissipative effects for the various methods.

Fig. 2(a) presents the relative position of the first-order schemes for the solutions in Fig. 1. The FI 1st solution precedes the exact solution while the FS, the SI Lagged and the SI Linearized solutions, respectively, lag the exact solution. The second-order schemes (FI 2nd and SS) in Fig. 2(a) have too small an error to be visible at this scale.

In Fig. 2(b), the FI 2nd solution is shown along with FI 2nd solutions which lump the mass (FI 2nd Lmpd M) and lump the mass and the source-term operators (FI 2nd Lmpd M&S). The FI 2nd solutions all lag the exact solution, and the effect of lumping the mass and source-term operators slightly increases the error by further lagging the exact solution. For FI 1st solutions (not shown), the effect of lumping the mass and source-term operators is also to increase the error but by further preceding the exact solution.

In Fig. 2(c), the operator-splitting methods are shown for the thermal-wave problem. As noted earlier, the FS scheme lags the exact solution in contrast to the FI 1st scheme which proceeds the exact solution. The relative position of the second-order splitting methods indicates that the SS proceeds the exact solution

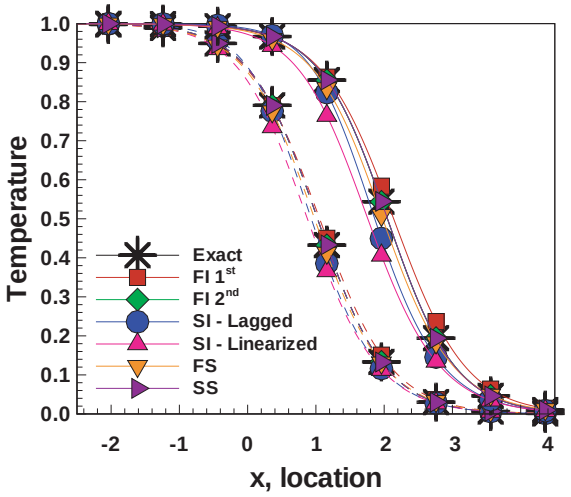

Fig. 1. Thermal-wave profiles at $t = 0.512$ (dashed lines) and $t = 1.024$ (solid lines), using $\Delta x = 0.04$ and $\Delta t = 0.064$. All methods shown use consistent-mass matrix, and every 20th data point has been shown with a symbol.

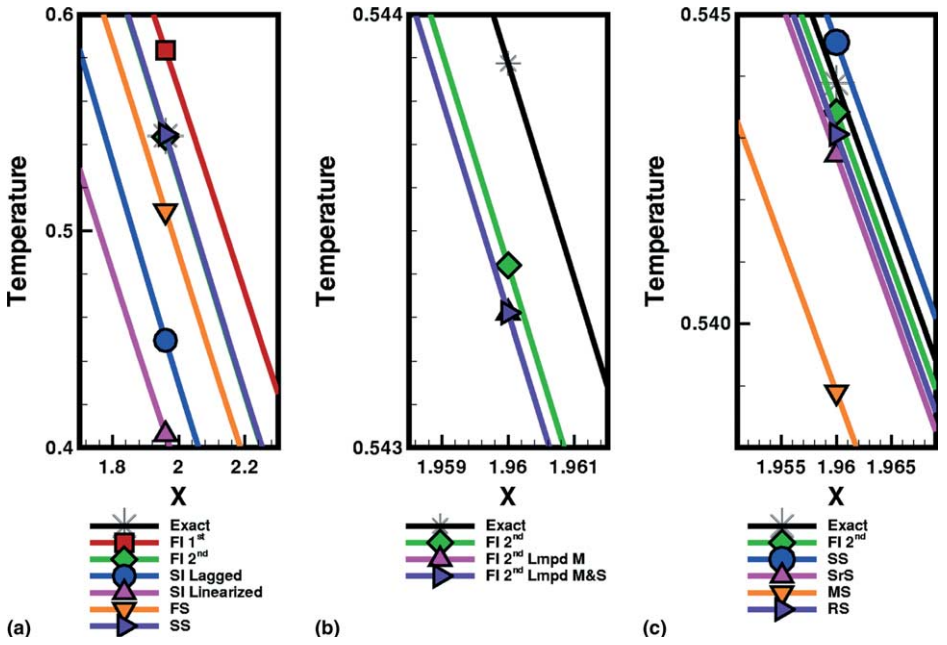

Fig. 2. Blow up of thermal-wave profiles at $t = 1.024$ for (a) the solutions shown in Fig. 1, (b) lumped mass and source-term operators, and (c) operator-splitting methods using $\Delta x = 0.04$ and $\Delta t = 0.064$. Every 20th data point has been shown with a symbol.

and RS, SrS, and MS lag the exact solution. It should be noted that the MS solution shown in Fig. 2(c) finished on an odd time step and thus did not complete its odd–even cycle which is required for formal second-order accuracy.

6.1.2. Spatial accuracy

To ensure that the temporal-convergence studies are not complicated by the spatial errors, a mesh-convergence study was completed to determine if the spatial errors where in the asymptotic range. In Fig. 3, the L_2 norms of the error are shown for various mesh spacings using both linear and quadratic elements. Despite the wide range of time-integration methods, second-order accuracy is obtained by all the methods. When quadratic elements are used, fourth-order spatial accuracy is obtained until the temporal error dominates near 4×10^{-8}. The source of this super-convergent behavior in this discrete nodal L_2 norm is not yet apparent. However the FE methods are clearly demonstrated to at least provide the expected spatial order of accuracy for this smooth solution. From these results, it is also clear that the lumped mass and source operators achieve nearly second-order spatial order of accuracy with only a slight offset in the relative error of these methods. The temporal convergence studies that follow are verified to be in the asymptotic region by these studies.

6.1.3. Temporal accuracy

To investigate the predictive characteristics of the other reference solutions, we will use FI 1st and FI 2nd solutions as test solutions and compare them against the exact solution (Eq. (29)); an extrapolated solution (Eq. (47)) using $\Delta t_i = 0.002$ and $\Delta t_j = 0.001$; and the best resolution of each solution against itself using $\Delta t_m = 0.001$. As a means of visually confirming the order of accuracy of a method, we include in the remainder of the paper a dashed line in the figure as a first-order reference slope, and a dotted line as a second-order reference slope in appropriate figures.

In Fig. 4, the L_2 norms of the error are shown for FI 1st and FI 2nd against the exact solution. FI 1st scheme shows first-order accuracy over the entire time-step range, but the FI 2nd scheme only shows second-order accuracy for larger Δt. For smaller Δt, FI 2nd plateaus due to spatial error related to the coarse spatial discretization of $\Delta x = 0.04$.

If the extrapolated solution is used as the reference, the spatial error is eliminated as described earlier. This can be seen for the FI 2nd solution in Fig. 4 as the second-order accuracy is exhibited for the full range of Δt. Thus the extrapolated solution has an advantage as a numerical reference solution when investigating temporal error which is smaller than the spatial error. As an illustration of the relative magnitudes of the two error components, we also present results for the first-order fully implicit method. In this case since the

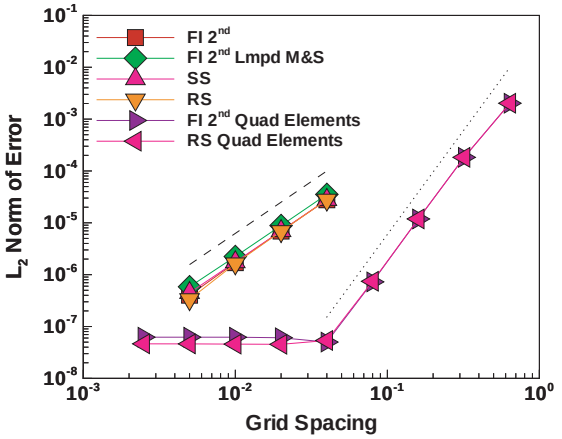

Fig. 3. Thermal-wave second-order and fourth-order spatial accuracy with $\Delta t = 0.001$ using the exact solution as the reference solution. The dashed line is a second-order reference slope and the dotted line is a fourth-order reference slope.

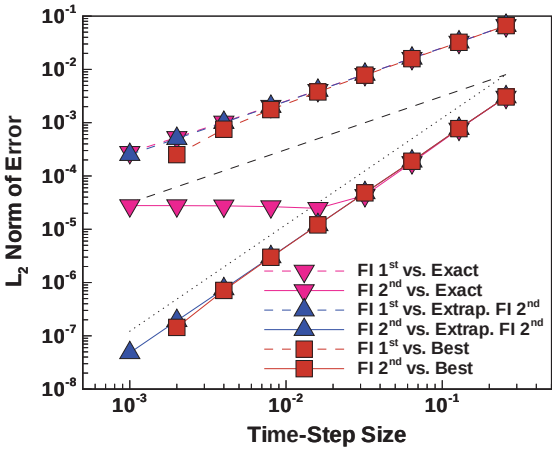

Fig. 4. Thermal-wave L_2 norm of the error with $\Delta x = 0.04$ for FI 1st and FI 2nd solutions compared against several references. The dashed line is a first-order reference slope and the dotted line is a second-order reference slope.

temporal error for the FI 1st solution is still larger than the spatial error; no plateau is yet present over the given interval.

We next consider using the best-resolution solution as the reference. If the best resolution ($\Delta t_m = 0.001$) is used as the reference, the spatial error again can be subtracted out to remove the plateau in the L_2 norm of the error. However as the time-step size decreases and approaches Δt_m, the order-of-accuracy increases (i.e., the slope increases). As discussed in Section 5.3, this slope is misleading and suggests that the best-resolution reference should use at least an order of magnitude smaller Δt than the lower bound of the time-step range of interest. Therefore, if possible, an extrapolated solution should be used in the absence of an exact solution.

Next a comparison of the consistent and lumped operators is presented. In Fig. 5 the FI 2nd with consistent mass and source-term solution (FI 2nd) and the FI 2nd with lumped mass and source-term

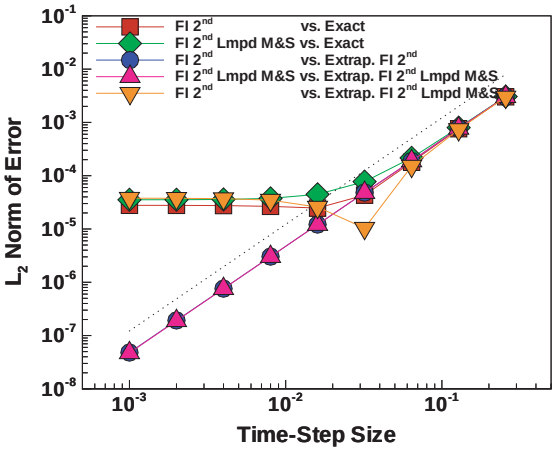

Fig. 5. Thermal-wave L_2 norm of the error with $\Delta x = 0.04$ for FI 2nd with both consistent and lumped mass and source-term operators compared against several references.

solution (FI 2nd Lmpd M&S) obtain second-order accuracy for large Δt. As Δt is decreased the spatial error begins to dominate and the exact error norm plateaus for smaller Δt. Additionally the FI 2nd Lmpd M&S solution is demonstrated to have a slightly larger spatial error. When these solution are compared against their own extrapolated solution, the spatial error is removed and they both show second-order accuracy for the entire time-step range.

For a cross-comparison of the FI 2nd solution against the FI 2nd Lmpd M&S extrapolated solution, a plateau returns to the L_2 norm curve. At large Δt, a second-order slope is obtained, but at small Δt, the spatial error caused by the different spatial discretizations (i.e., consistent versus lumping) for the mass and source-term operators can be easily seen. The dip at the mid-range of the time-step sizes is likely due to favorable cancellation of the spatial and temporal errors which are approximately the same magnitude in this range.

In Fig. 6(a), the L_2 norm of the error is shown for several time-integration schemes. The first-order methods (FI 1st, SI Lagged, SI Linearized and FS) all show first-order accuracy for the time-step range studied, but each has a different leading coefficient in the error, C_t. Despite that the SI Linearized includes information from the $n+1$ time step, it has a larger leading coefficient than the SI Lagged. The FI 1st method performs better than the semi-implicit schemes, but surprisingly the FS scheme has the smallest leading coefficient of the first-order methods. This increased accuracy can most likely be attributed to the smaller time steps that are used in the sub-cycling of the CVODE solution for the reaction source term and the existence of a linear diffusion subproblem. Apparently in this case, the splitting errors are smaller in magnitude than the reduction in error obtained by the sub-cycling of the nonlinear terms. For a given time-step size, Δt, the first-order solutions have errors that vary over about a factor of 6. This, of course, can be translated into a factor of 6 in Δt to obtain the same level of error. Thus there is a possible factor of 6 in time-step size when using FS versus SI Linearized. It should be noted that the SI Lagged and the SI Linearized solutions show a decrease in slope at larger Δt beginning at $\Delta t \approx 0.1$. This behavior is similar to the results presented in [11] for a simplified PDE system as well. In [11], the authors suggest that a change in one of the component time scales might occur at an intermediate time-step size. Also in Fig. 6, FI 2nd and SS schemes methods are demonstrated to obtain second-order accuracy and plateau near an L_2 norm of the error of 3×10^{-7} due to the spatial error associated with the mesh spacing of $\Delta x = 0.005$.

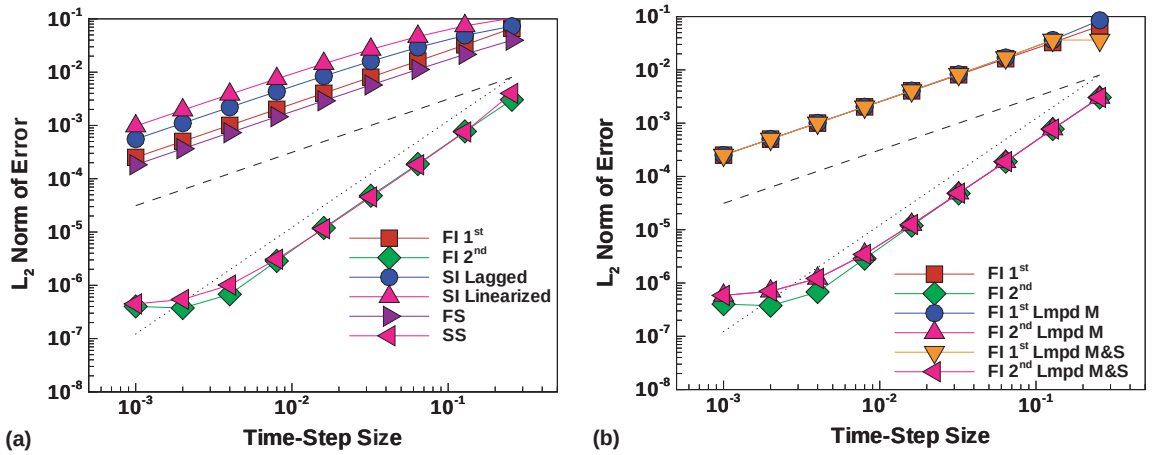

Fig. 6. Thermal-wave L_2 norm of the error with $\Delta x = 0.005$ using the exact solution as reference against FI 1st and FI 2nd with (a) consistent and (b) lumped mass and source-term operators.

Fig. 6(b) compares the effect of using consistent and lumped versions of the mass and source operators for the thermal-wave problem. For the first-order schemes, there is no apparent effect of lumping on the temporal accuracy; all the L_2 norms of the error fall nearly on top of one another. For the second-order schemes, however, the spatial error associated with lumped operators can be seen as Δt decreases and the spatial error dominates the L_2 norm of the error. In Fig. 6(b), lumping does not appear to effect the temporal relative error or the order of accuracy, but does effect the effective spatial error. From Fig. 3, we can see that this spatial error primarily effects the leading coefficient of the spatial error. There is very little difference between a lumped mass operator and lumped mass and source-term operators. This suggests that the diagonalization of the mass operator accounts for most of the variation in this problem.

Finally in Fig. 7, we present the L_2 norm of the error for various splitting methods as applied to the thermal-wave problem. The FS obtains first-order accuracy, and the second-order splitting methods all obtain second-order accuracy, however some new behavior can be noted. The SS follows the FI 2nd nearly identically except for smaller Δt near the plateau of the spatial error associated with $\Delta x = 0.005$. These results indicate SS has a smaller leading coefficient than SrS which ends with the diffusion operator. The RS solution has a leading coefficient very similar to SrS which may not be surprising since they have the same sequence of split operators. For large Δt, the RS solution is more accurate than the FI 2nd solution but then transitions to the same error as SrS (again this occurs at $\Delta t \approx 0.1$).

From Fig. 7, it is apparent that the MS solution has a stair-step behavior which overall obtains second-order accuracy. The valleys of the stairs correlate to simulations which ended with even numbered time steps. This is a diffusion solve followed by a reaction solve and completes the odd–even cycle to obtain second-order accuracy. The peaks of the stairs correlate to simulations which ended with odd-numbered time steps. This time step is a reaction solve followed by a diffusion solve. Of course, the final time step could be halved, if the solution would end on an odd number of time steps, to ensure that the odd–even cycle is completed. However this was not done in this study to demonstrate the importance of applying the correct sequence of operators.

6.2. Non-equilibrium radiation diffusion

In this section we consider the relative accuracy and the asymptotic order of accuracy for the fully implicit, semi-implicit and operator-splitting methods on the radiation–diffusion problem of Section 4.2.

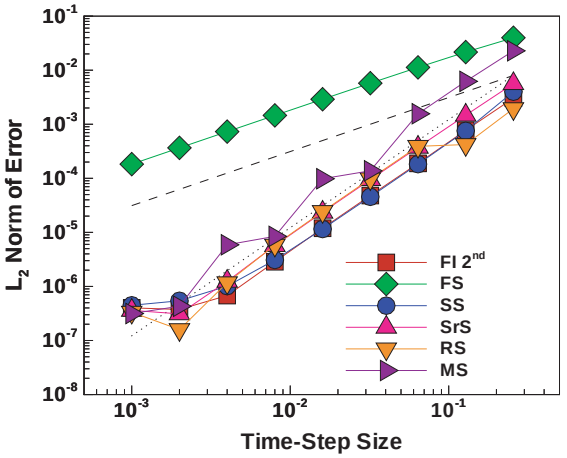

Fig. 7. Thermal-wave L_2 norm of the error with $\Delta x = 0.005$ using the exact solution as reference against FI 2nd and the operator-splitting methods.

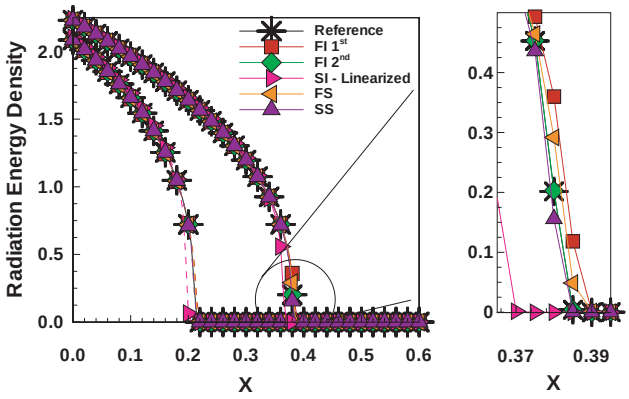

Fig. 8. Radiation–diffusion profiles at $t = 0.5$ (dashed lines) and $t = 1.0$ (solid lines), using $\Delta t = 0.0025$ and $\Delta x = 0.005$. Every 10th mesh point is indicated by symbol. The inset figure details the relative position of the wavefront at time $t = 1.0$ for the various techniques. The reference solution is the FI 2nd methods with $\Delta t = 1.953125 \times 10^{-5}$.

This problem differs from the thermal-wave problem of Section 4.1 in three major aspects. First there is not an available analytic solution to use as a reference solution and second the diffusion–reaction system that defines the problem is a coupled system of two nonlinear PDEs. The third distinguishing characteristic of this problem is the discontinuous derivative of the solution at the propagating wavefront.

6.2.1. Solution profiles

Representative solution profiles for the radiation-energy density are shown in Fig. 8. This figure includes solution profiles at times $t = 0.5$ and $t = 1.0$ for the three broad classes of time-integration methods of our study (the fully implicit, semi-implicit and operator-split techniques) along with our reference solution. In this case we use a highly resolved FI 2nd method as a reference solution ($\Delta t = 1.953125 \times 10^{-5}$) for a comparison of these qualitative aspects of the simulation. In general the radiation energy density increases along the inflow boundary ($x = 0.0$) with time and E varies smoothly near this boundary. From these results and the results detailed in Fig. 9, qualitative information on the relative wave speed (or phase errors) and relative magnitude of numerical oscillations in the vicinity of the wavefront for the various formulations can be obtained. For example, Figs. 8 and 9 indicate the relative spreading of the solution profiles due to temporal errors. As time progresses in the simulation, a spreading of the wavefront locations for the various methods takes place. These results indicate that the SI schemes lag the reference solution. [1] In addition, the FI 1st method and the FS method lead the reference solution. Clearly the FI 2nd and SS schemes are nearly indistinguishable from the location of the reference solution in this case. The results for the FI and SI methods are similar to those of Knoll et al. [12,13].

Next we consider the wavefront for the radiation energy density. Fig. 9 exhibits very clearly the oscillations in the vicinity of the wavefront for a subset of the methods that employ consistent source-term operators as described in Section 3. For the results of the radiation–diffusion problem, all the methods use a lumped mass operator and either a consistent or lumped source operator. In this regard, Fig. 9(a) exhibits the oscillatory behavior of the FI and SI schemes that do not use source-operator lumping, relative to the operator-splitting methods which implicitly decouple the nonlinear source terms spatially. This decoupling is a result of the splitting process and the group FE expansion as discussed previously.

[1] The results for the SI Lagged solution are not shown here because of the relatively large error in the wave speed and restriction of the scale in Fig. 9(a) to resolve the numerical oscillations.

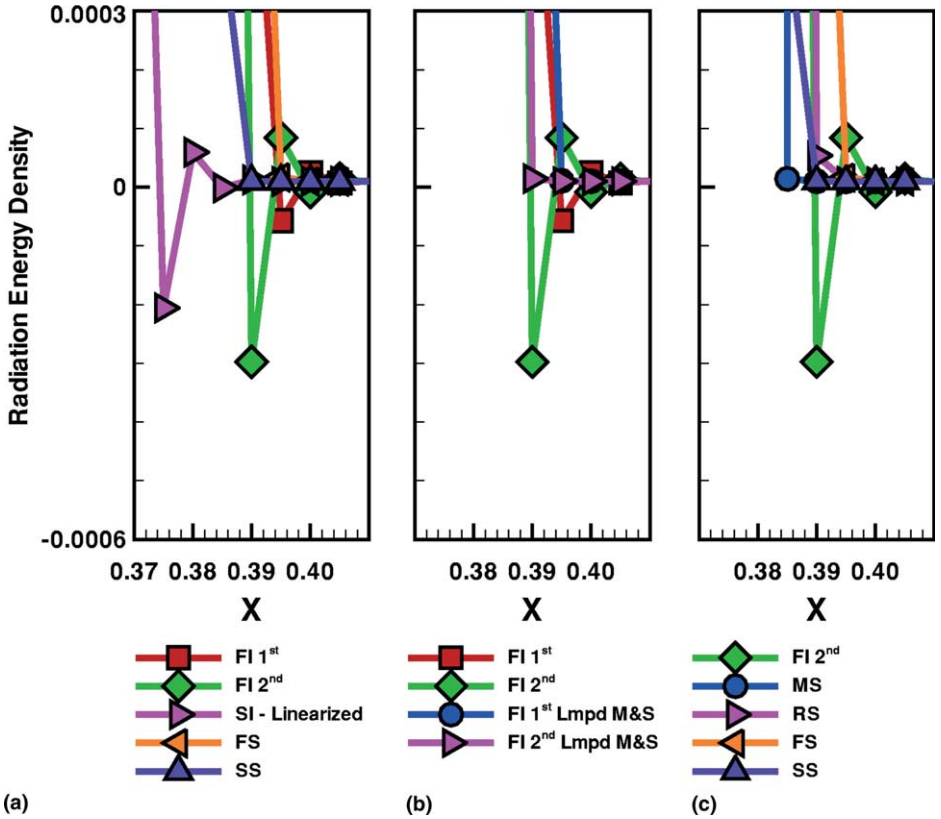

Fig. 9. Radiation–diffusion profiles at the wavefront for the various solution methods. Lumped-mass operators are used for all of the methods. (a) Fully implicit, semi-implicit, and operator-split methods, (b) lumped source-term comparison, and (c) operator-split methods are shown for the same conditions as Fig. 8.

The wavefront detail of Fig. 9(b) demonstrates the effectiveness of the lumping (or diagonalization) procedure for the suppression of oscillations in the vicinity of the wavefront. In this figure, the lumped-source operator variants of the methods remove the oscillations at the wavefront and still resolve the location of the wavefront accurately. This tendency for under-resolved Galerkin FE methods to oscillate in the vicinity of steep gradients and discontinuities has been well demonstrated. Commonly these oscillations are removed by post-processing methods, stabilization techniques or methods as we employ here. It is clear however, from our results, that while the oscillations exist they do not produce inaccuracies in the phase speed of the methods relative to the reference solution (this claim will be further substantiated in the asymptotic order-of-accuracy studies that follow).

For completeness we include Fig. 9(c) which demonstrates that all the splitting methods do not exhibit oscillatory behavior in the vicinity of the wavefront. It is interesting to note that with the exception of the Marchuk splitting method all the second-order split schemes essentially have the same wavefront location. The discrepancy for the Marchuk scheme is that the solution at time $t = 1.0$ ended on an odd time step and did not complete the odd–even cycle.

6.2.2. Spatial accuracy

In this section we briefly describe the results of a spatial-convergence study using linear finite elements for the radiation–diffusion problem. The study considers a cross comparison of the solution methods that

implicitly have an effective difference in the spatial discretization scheme that are used in the weak form of the system as described in Section 2. As described in Section 3.4, these two methods can be considered to differ in the quadrature method that is used to evaluate the operators: full quadrature for the consistent method (in this case two-point Gaussian quadrature in one dimension) and a one-point nodal quadrature for the lumped (or diagonalized) mass and source-term operators. For this reason it is important to verify that the two discretization techniques are approximately achieving the expected order of accuracy for the radiation–diffusion problem solution. This study of a solution with discontinuous derivatives for the radiation–diffusion problem, coupled with the convergence study of the smooth thermal-wave example, will demonstrate the effective order of accuracy for the consistent and lumped FE implementations.

Fig. 10(a) presents mesh-convergence results for the first-order fully implicit methods with consistent and lumped source-term operators along with the operator-splitting FS method. The second-order methods are presented in Fig. 10(b). In this study, a moderately small time step ($\Delta t = 0.0003125$) is used. This time-step size is mid-range in the following temporal-convergence studies of the next section and is clearly in the range of asymptotic order of accuracy for the various methods. These results use the best-resolution solution, for each of the various methods, as a reference solution for the mesh-resolution study. From this figure it is clear that there are at least three stages to the spatial convergence of these methods. For the first-order methods, the two initial stages appear to have a pre-asymptotic range where the methods at first begin to converge at a super-linear rate and then relax to a sub-linear rate and almost begin to plateau. After sufficient resolution of the wavefront is reached, a third stage with an asymptotic order of accuracy of about 1.7 is obtained. Similarly Fig. 10(b) presents the spatial convergence for the second-order methods. In this case there are also three stages of convergence as the second-order in-time methods begin to resolve the wavefront. In contrast to the first-order methods, the second-order techniques do not appear to fully obtain an asymptotic order of accuracy in the third stage. In fact if a linear regression fit to the third stage is used, these methods converge at a rate of about 1.3. We conjecture that the higher-order spatial convergence of the first-order in-time methods might be due to the beneficial effect of the larger numerical dissipation at the wavefront. In this case the numerical dissipation would tend to smooth the discontinuity and allow effective resolution of the large gradients at relatively larger mesh spacings. For the less dissipative second-order in-time methods, it would appear that finer resolution of the wavefront is necessary before asymptotic order of

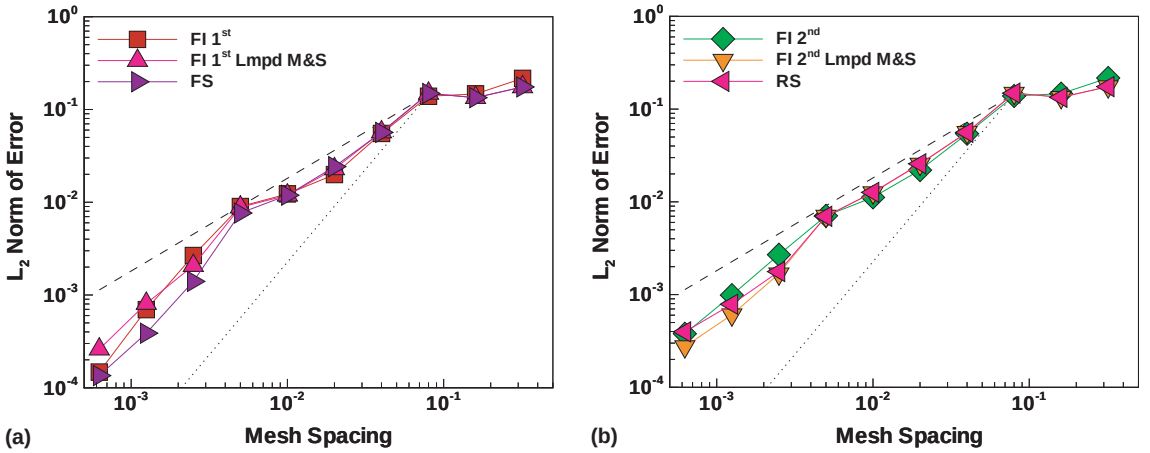

Fig. 10. Spatial accuracy with $\Delta t = 0.0003125$ using (a) first-order temporal and (b) second-order temporal methods. Reference lines for a first-order (dashed) and a second-order (dotted) order of accuracy are shown.

accuracy will be reached. For this reason we present only the best-resolution reference results for the spatial-convergence studies of this section.

The sub-optimal spatial convergence rates of the methods is understandable given the discontinuous derivative of the solution. In general the first-order methods do relatively well and the second-order in-time methods, while not yet fully asymptotic, would appear to be close to reaching an asymptotic stage. All the methods appear to achieve a comparable relative and asymptotic order of accuracy which is self-consistent irrespective of the treatment of the source-term operator. As demonstrated above, the lumped (diagonalized) and operator-split methods are characterized by non-oscillatory solutions near the wavefront. This behavior was observed for all mesh spacings (Δx) used in our study. In contrast the consistent source-term-operator methods produced oscillations even for the finest mesh of our study ($\Delta x = 0.0003125$) was used. This result indicates an important issue that makes a cross comparison of these methods at a reasonable mesh spacing (Δx) difficult. From Fig. 10, it is clear that the relative accuracy (L_2 norm of the error) of the methods decreases at a sub-optimal rate ($\sim$1.7 for first-order and $\sim$1.3 for second-order), in addition it is clear that the relative error in the solution for these given mesh spacings is moderately high.

In order to decrease these spatial errors substantially, a much smaller (Δx) would be required. To carry out all the required transient simulations at a finer mesh spacing would be prohibitive in terms of CPU time for our study. Of course, for a practical simulation, this issue would be addressed with local mesh refinement, however in our case this is unacceptable since the resulting effect on temporal convergence studies would be difficult to account for carefully. For this reason we must be satisfied with only moderately well resolved spatial computations ($\Delta x = 0.005$) and look for a self-consistent means to compare the temporal accuracy of the various methods. Since we are interested in temporal-convergence studies over several orders of magnitude in time-step size (e.g., Fig. 11), the relative spatial errors described above are large compared to the relative errors that we present in the temporal studies that follow. The magnitude of these relative errors coupled with the significant differences in the profiles for the consistent (oscillatory) and lumped (non-oscillatory) methods indicate that a careful cross-comparison of these methods for temporal-convergence studies will be difficult. In the next section we demonstrate this problem and propose a solution which compares the consistent-source and lumped-source methods only within the appropriate self-similar categories.

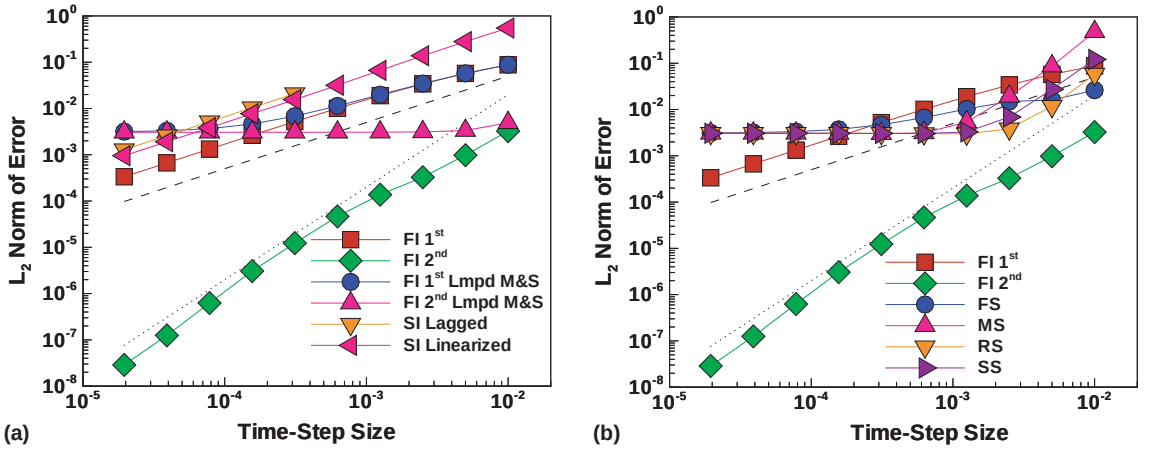

Fig. 11. Radiation–diffusion L_2 norm of the error with $\Delta x = 0.005$ using the extrapolated solution of FI 2nd lumped-mass solution as the reference solution against (a) fully implicit and semi-implicit schemes, and (b) operator-splitting schemes.

6.2.3. Temporal accuracy

In this section we describe a study of the temporal-convergence behavior of the time-integration schemes in the context of the radiation–diffusion problem. The focus of the study is to provide a quantitative comparison of the relative accuracy of the methods as well as establish and compare, with numerical experiments, the asymptotic order of accuracy of the various techniques. We begin the discussion by considering more carefully the issue of choosing an appropriate reference solution for methods which use a different effective spatial discretization. As described above the choice of the reference solution is critical in the case when the effective discretization differs for two solutions that are being compared. This point is illustrated more clearly in Figs. 11 and 12.

In Fig. 11, the reference solution is taken to be the extrapolated fully implicit second-order lumped-mass (FI 2nd) solution. Fig. 11(a) compares the fully implicit and semi-implicit methods and Fig. 11(b) compares the fully implicit and operator-split methods. In Fig. 11(a), the second-order convergence of the FI 2nd method is displayed relative to the first-order and second-order reference slopes. In Fig. 11(a), it is clear that the fully implicit first-order (FI 1st) and the semi-implicit schemes (SI) also display the expected first-order rate of convergence relative to the FI 2nd extrapolated solution. The graphs of these methods display an asymptotic first-order rate of convergence. In contrast, the lumped-source variants of the fully implicit methods display convergence with decreasing time-step size only until a lower limit of time-step size is reached (see for example the FI 1st Lmpd M&S method). This plateau occurs in both lumped source methods, and therefore these techniques do not display either a first-order or second-order asymptotic order of accuracy using the FI 2nd solution as a reference. The plateau or stagnation of the error is associated with the discrepancy between the two classes of solutions at the wavefront. The magnitude of error for which the stagnation occurs is related to the error norm at $\Delta x = 0.005$ exhibited above in the mesh-convergence study summarized in Fig. 10. The corresponding comparison between the fully implicit methods and the operator-splitting methods is shown in Fig. 11(b). Clearly the same quantitative behavior occurs as in the lumped-source-term variants in Fig. 11(a).

To further demonstrate the issue related to choice of reference solution, we include Fig. 12. In this comparison, the reference solution is selected as the FI 2nd Lmpd M&S solution. In this case it is clear that the consistent variants do not display first-order or second-order asymptotic rates of convergence and instead display a similar stagnation in the error. This is in contrast to the operator-split methods which now

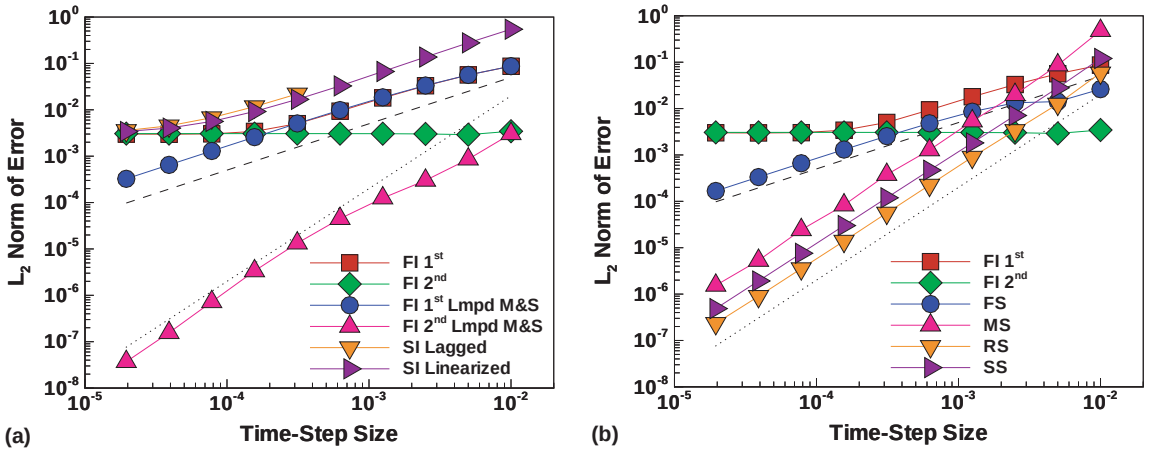

Fig. 12. Radiation–diffusion L_2 norm of the error with $\Delta x = 0.005$ using the extrapolated solution of FI 2nd lumped-mass-and-source solution as the reference solution against (a) fully implicit and semi-implicit schemes, and (b) operator-splitting schemes.

display the appropriate first-order and second-order rate of convergence when compared to the lumped-source reference solution. The relative order of accuracy of the methods is also apparent from this graph as well. In the next set of figures, this comparison of relative errors is more clearly displayed and discussed. To close the discussion of the appropriate choice of a reference solution we indicate that careful consideration of the effective spatial discretization of a method is necessary to provide a valid comparison between methods.

An overall comparison of the temporal convergence of the methods is presented in Fig. 13. This figure summarizes the temporal-accuracy behavior for both the lumped and consistent source-term variants and uses a reference solution based on an extrapolated solution of the appropriate FI 2nd method as described in the discussion above. From this figure, both the relative accuracy (or leading coefficient) and the asymptotic order of accuracy are compared with a first-order and second-order reference slope. In general the FI 2nd methods (lumped and consistent) exhibit similar convergence behavior. The semi-implicit methods are shown to be first-order convergent and the SI Lagged method has the largest relative error. This method also had problems in converging for time steps above $\Delta t = 0.0003125$. For the radiation–diffusion problem, the SI Lagged and SI Linearized methods exhibit a reverse behavior to the results in the thermal-wave comparison. In general for an arbitrary linearization, there is no clear a priori means of assessing the accuracy of the linearization without recourse to some type of Taylor series-expansion and modified-equation analysis [11]. In terms of the first-order consistent-source methods, the fully implicit method is demonstrated to be more accurate than both of the semi-implicit methods.

A comparison of the lumped-source methods reveals that the formally second-order operator-splitting methods all achieve second-order order of accuracy on the radiation–diffusion problem when compared relative to the FI 2nd Lmpd M&S method. The relative accuracies of the methods indicate that the FI 2nd Lmpd M&S method is the most accurate followed by the Romero, Strang and Marchuk splitting method. The formally first-order splitting scheme, FS, after an initial startup phase of sub-linear convergence, exhibits first-order convergence.

For a fixed time-step size there is, in general, about two orders of magnitude relative accuracy difference between the second-order fully implicit method and the least accurate operator-splitting method (Marchuk splitting). Comparing first-order methods there is also a relative accuracy difference of about one order of magnitude between the various methods. Expanding the comparison to both first-order and second-order methods there is in general a difference of about three orders of magnitude between the fully implicit

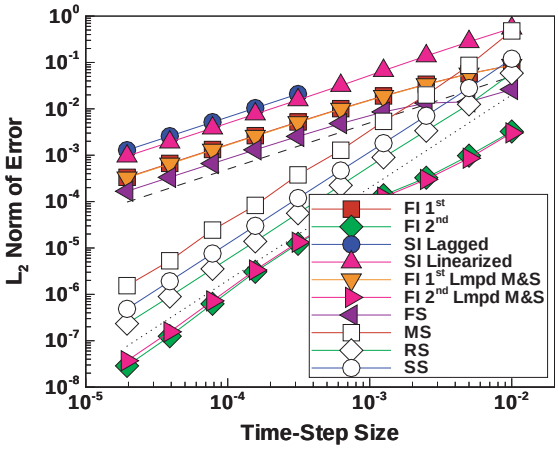

Fig. 13. Radiation–diffusion L_2 norm of the error with $\Delta x = 0.005$ and using the appropriate reference solution for each method.

second-order method and the semi-implicit methods for a fixed time-step size of $\Delta t = 0.0025$. This variation is very significant for well-resolved simulations and is a result of the higher order of accuracy of these schemes.

For a specified acceptable error, other interesting conclusions can be drawn in terms of possible efficiency differences between the various methods. For a fixed error there is, in general, about one order of magnitude difference in time-step size that can be taken with the best first-order method relative to the least accurate first-order (semi-implicit) technique. For the second-order methods this same ratio is also exhibited. Comparing across first-order and second-order methods a two order of magnitude difference in time-step size between the best second-order method and the least accurate semi-implicit (first-order) technique is apparent for L_2 norm of the error near 10^{-3}. This significant discrepancy makes possible the development of more efficient highly accurate methods that employ such fully implicit methods (see [4,11]).

6.2.4. Variations of time-integration schemes

Finally we consider variants of the above second-order fully implicit and operator-splitting methods which attempt to reduce the nonlinear solution requirement to a single solve of the spatially coupled large sparse linear system per time step. The two general methods include a one-step Newton variation of the fully implicit method, and a one-step Newton variation along with a linearization technique for the nonlinear diffusion solve in the operator-splitting methods. The one-step Newton variation uses a numerical approximation to the Jacobian matrix for each specific technique. In this method a finite-difference approximation to the Jacobian entries is generated by a two-point finite-difference formula. The subsequent Newton iteration is then limited to just one step or equivalently to one linear solve step. The results of applying this technique to the second-order fully implicit method is shown in Fig. 14(a). For this problem the one-step variant exhibits second-order asymptotic order of accuracy and very similar relative accuracy to the full-Newton method with the use of a single linear solve. This result for the fully implicit method is similar to the results for the LIN2 algorithm of Lowrie [14]. A disadvantage of the one step variant is that it exhibited a degradation of stability and did not converge for the two largest time step sizes of interest in our study.

In these computations, an accurate numerical approximation to the true Jacobian that included the effect of all the terms in the Jacobian was used. In general these methods either rely on an accurate analytic

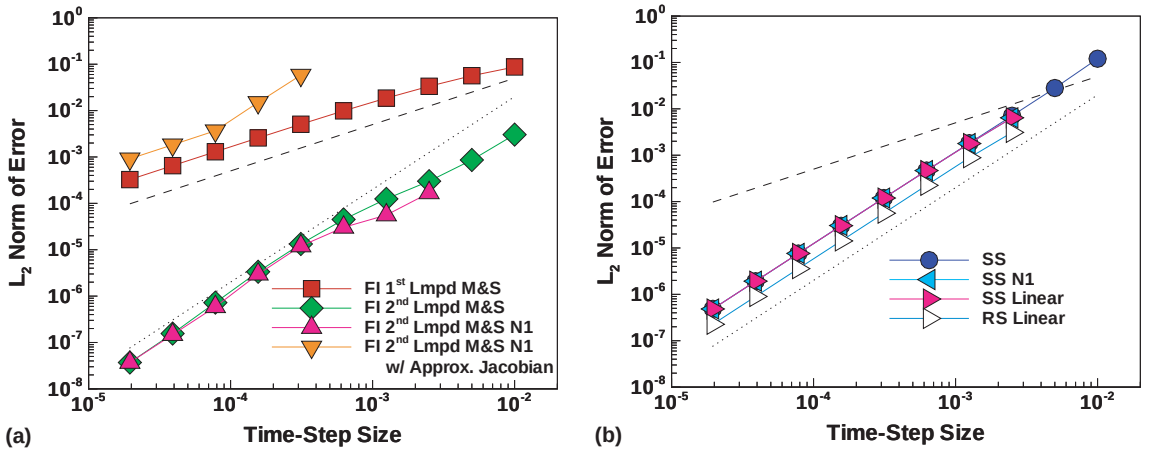

Fig. 14. Radiation–diffusion L_2 norm of the error for (a) one-step Newton variations of FI 2nd Lmpd M&S, and (b) operator splitting methods with linearized diffusion solves with $\Delta x = 0.005$. Both results use the extrapolated FI 2nd lumped mass and source-term solution as the reference solution.

expansion of the nonlinear terms to $o(\Delta t^2)$ accuracy or they require a sufficiently accurate numerical approximation to the Jacobian-vector product with a highly accurate linear solve (please note we leave the definition of sufficient conditions unstated since it is beyond the scope of the current manuscript).

To illustrate the possible loss of stability and accuracy for such a method with an inaccurate Jacobian, we carried out computations with a one-step Newton method with our approximate analytical-Jacobian matrix. This matrix includes only the Jacobian terms associated with the source-term operators and does not include the variation of the nonlinear diffusion coefficient. In these results the method did not converge for time steps larger than $\Delta t = 0.0003125$ and had a relative accuracy that was larger by approximately four orders of magnitude. An important point is that by using the same approximate Jacobian that was used in our fully converged Newton schemes and only allowing one Newton step per time step to be taken, produced these significant degradations in stability and accuracy of the FI 2nd method. Conversely stated, fully converging the Newton sequence produces a FI time-step accuracy independent of the accuracy of the Jacobian. Therefore it is clear that care must be taken when applying one-step methods with approximate Jacobians. The operator-split version of the one-step Newton method for the diffusion subproblem is demonstrated in Fig. 14(b).

The final one-step variant we present is a linearized form of the operator-splitting methods that employs a diffusion coefficient that is defined by the intermediate solution of the preceding source-term step. This diffusion coefficient is then held constant in the subsequent diffusion solve and thus produces a linear diffusion problem. In Fig. 14(b), it is clear that over a large range of time-step sizes these methods (SS linear and RS linear) exhibit a second-order asymptotic order of accuracy with relative temporal-convergence errors that are nearly equivalent to the fully nonlinear schemes (e.g., SS). This behavior, however, is not true at larger time-step sizes where the linearized-diffusion and one-step Newton schemes did not converge.

The results from both the one-step Newton methods and the linearized-diffusion schemes are promising. In context, however, we believe that these methods always need to be carefully applied, and throughly tested with convergence studies relative to fully converged implicitly balanced methods. In addition, the practice of suppressing time-step nonlinearity for a gain in efficiency has been demonstrated to sometimes degrade the accuracy and/or stability. While accuracy issues should be considered relative to the other sources of errors in the simulation, a decrease in stability or robustness is of great concern in complex multi-physics applications. In these systems it is often very difficult to isolate the source of such instabilities.

7. Conclusions

In this paper we have presented a study of the accuracy of time-integration techniques applied to the governing equations for the non-equilibrium radiation–diffusion approximation to radiative transport. As a preliminary step, we have included a numerical study of a related thermal-wave propagation problem that has a smooth analytical solution. This prototype problem allowed us to evaluate and compare the use of exact solutions, best-resolution solutions, and extrapolated solutions as references for estimating errors. As a result of the thermal-wave study we have demonstrated: (1) spatial-convergence studies to verify expected finite element (FE) order of accuracy, (2) the specific implementations of semi-implicit lagged and linearized methods, operator-split and fully implicit time-integration techniques, (3) the expected temporal asymptotic order of accuracy for each method, (4) the usefulness of an extrapolated solution as a reference solution for estimating errors, and finally, (5) a similar relative and asymptotic order of accuracy for the diagonalized (lumped) operator formulation in relation to the consistent operator formulation.

Building on the results of the thermal-wave study have presented a detailed numerical study of a radiation–diffusion problem presented in [13,14]. In our study we have demonstrated: (1) the effectiveness of employing lumped mass and source-term FE operators for controlling oscillations, (2) the relative accuracy and asymptotic order of accuracy of various semi-implicit, operator-splitting and fully implicit methods, (3)

the need to use self-consistent reference solutions when performing numerical temporal-convergence studies between operator-splitting methods and fully implicit and semi-implicit techniques, (4) that the splitting methods can obtain second-order asymptotic order of accuracy for these nonlinear problems, (5) that the second-order implicitly balanced methods (full-Newton and one-step Newton) obtain higher relative accuracies than the second-order splitting methods, (6) that one-step methods are susceptible to stability problems for larger time-step sizes as compared to the fully converged Newton methods, (7) that the one-step methods also require a sufficiently accurate Jacobian if a second-order method is to be obtained, (8) that the second-order operator-split linearized-diffusion methods deserve significant further study, and (9) the fully converged, fully implicit balanced methods are the benchmark solvers by which these techniques are to be evaluated.

Since our study employed Newton–Krylov methods for solution of the nonlinear subsystems of the splitting techniques and the one-step methods, we believe this further underscores the usefulness, robustness and flexibility of Newton–Krylov iterative techniques. Finally it is our belief that careful convergence studies such as these are required to help accurately assess the potential accuracy, cost and implementation complexity of transient nonlinear-solution techniques for predictive simulations of complex nonlinear multiple time-scale systems.

Acknowledgements

The authors thank Vanessa Lopez for her work on the initial study of the radiation–diffusion problem during the summer of 2001 along with Dana Knoll and Rob Lowrie for many helpful and spirited discussions on issues related to implementation and accuracy of these methods. In addition we thank Pavel Bochev for his helpful discussions on the error models for the FE methods.

References

[1] R.L. Bowers, J.R. Wilson (Eds.), Numerical modeling in applied physics and astrophysics, Jones and Bartlett, Boston, 1991.
[2] J. Brackbill, B. Cohen (Eds.), Multiple Time Scales, Academic Press, Orlando, 1985.
[3] P.N. Brown, C.S. Woodward, Preconditioning strategies for fully implicit radiation diffusion with material-energy transfer, Technical Report UCRL-JC-139087, Lawrence Livermore National Laboratory, November 2000.
[4] P.N. Brown, C.S. Woodward, Preconditioning strategies for fully implicit radiation diffusion with material-energy transfer, SIAM Journal on Scientific Computing 23 (2) (2001) 499–516.
[5] S.D. Cohen, A.C. Hindmarsh, CVODE, A Stiff/Nonstiff ODE Solver in C, Computers in Physics 10 (2) (1996) 138–143.
[6] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, SIAM Journal of Optimization 4 (1994) 393–422.
[7] C.A.J. Fletcher (Ed.), Computational Techniques for Fluid Dynamics, vol. 1, Springer, Berlin, 1988.
[8] P. Grindrod, The Theory and Applications of Reaction Diffusion Equations: Patterns and Waves, Oxford University Press, London, 1996.
[9] S.A. Hutchinson, L. Prevost, J.N. Shadid, C. Tong, R.S. Tuminaro. Aztec User's guide, Version 2.0, Technical Report SAND99-8801J, Sandia National Laboratories, Albuquerque, NM, October 1999.
[10] O.M. Knio, H.N. Najm, P.S. Wyckoff, A semi-implicit numerical scheme of reacting flow: II. Stiff, operator-split formulation, Journal of Computational Physics 154 (1999) 428–467.
[11] D.A. Knoll, L. Chacon, L.G. Margolin, V.A. Mousseau, On balanced approximations for time integration of multiple time scale systems, Journal of Computational Physics 185 (2003) 583–611.
[12] D.A. Knoll, W.J. Rider, G.L. Olson, An efficient nonlinear solution methods for non-equilibrium radiation diffusion, Journal of Quantitative Spectroscopy & Radiative Transfer 63 (1999) 15–29.
[13] D.A. Knoll, W.J. Rider, G.L. Olson, Nonlinear convergence, accuracy and time step control in non-equilibrium radiation diffusion, Journal of Quantitative Spectroscopy & Radiative Transfer 65 (2001) 25–36.
[14] R.B. Lowrie, A comparison of time integration methods for nonlinear relaxation and diffusion, Journal of Computational Physics (2004), in press.

[15] G.I. Marchuk, On the theory of the splitting-up method, in: Proceedings of the 2nd Symposium on Numerical Solution of Partial Differential Equations, SVNSPADE, 1970, pp. 469–500.
[16] V.A. Mousseau, D.A. Knoll, W.J. Rider, Physics-based preconditioning and the Newton–Krylov method for non-equilibrium radiation diffusion, Journal of Computational Physics 160 (2000) 743–765.
[17] Y.Y. Nie, V. Thomee, A lumped mass finite element method with quadrature for a non-linear parabolic problem, IMA Journal of Numerical Analysis 5 (1985) 371–396.
[18] E.S. Oran, J.P. Boris, Numerical Simulation of Reactive Flow, Cambridge University Press, Cambridge, MA, 2001.
[19] L.A. Romero, On the accuracy of operator-splitting methods for problems with multiple time scales, Technical Report SAND2002-1448, Sandia National Laboratories, August 2002.
[20] L.A. Romero, D.L. Ropp, J.N. Shadid, Studies of a family of operator-splitting methods (2003), in preparation.
[21] C.J. Roy, Grid convergence error analysis for mixed-order numerical schemes, AIAA Paper 2001-2606, 2001.
[22] J.N. Shadid, A.G. Salinger, R.C. Schmidt, T.M. Smith, S.A. Hutchinson, G.L. Hennigan, K.D. Devine, H.K. Moffat, A finite element computer program for reacting flow problems part 1: theoretical development, Technical Report SAND98-2864, Sandia National Laboratories, January 1999.
[23] J.N. Shadid, R.S. Tuminaro, H.F. Walker, An inexact Newton method for fully coupled solution of the Navier–Stokes equations with heat and mass transport, Journal of Computational Physics 137 (1997) 155–185.
[24] B. Sportisee, An analysis of operator splitting techniques in the stiff case, Journal of Computational Physics 161 (2000) 140–168.
[25] G. Strang, On the construction and comparison of difference schemes, SIAM Journal of Numerical Analysis 5 (3) (1968) 506–517.
[26] R.H. Szilard, G.C. Pomraning, Numerical transport and diffusion methods in radiative transfer, Nuclear Science and Engineering 112 (1992) 256–269.
[27] V. Thomee (Ed.), Galerkin Finite Element Methods for Parabolic Problems, Springer, New York, 1997.
[28] N.N. Yanenko (Ed.), The Method of Fractional Steps, Springer, New York, 1971, Translation Editor M. Holt.