

METHODS FOR EXTENDING HIGH-RESOLUTION SCHEMES TO NON-LINEAR SYSTEMS OF HYPERBOLIC CONSERVATION LAWS

WILLIAM J. RIDER

Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

SUMMARY

In extending high-resolution methods from the scalar case to systems of equations there are a number of options available. These options include working with either conservative or primitive variables, characteristic decomposition, two-step methods, or component-wise extension. In this paper, several of these options are presented and compared in terms of economy and solution accuracy. The characteristic extension with either conservative or primitive variables produces excellent results with all the problems solved. Using primitive variables, the two-step formulation produces high-quality results in a more economical manner. This method can also be extended to multiple dimensions without resorting to dimensional splitting. Proper selection of limiters is also important in non-characteristic extension to systems.

KEY WORDS Euler equations Conservative variables Primitive variables High-resolution

1. INTRODUCTION

In recent years, there has been an abundance of work deriving high-resolution schemes for hyperbolic conservation laws. Most of the development is made with scalar equations and generalized in some fashion to non-linear equations or systems of equations. Typically, the extension to systems of equations takes on great importance as is the case with the solution of the Euler equations of compressible flow. Much of the development of high-resolution methods is devoted to the solution of systems of equations as their primary practical use.

Among the methods developed in recent years, several stand out as seminal works. Perhaps the canonical work in the field was done by Godunov.^{1,2} This work was the spring-board for most of what is considered to be modern upwind methods. This started with the work of Boris and Book³ and their FCT method. At about the same time, van Leer^{4–6} had begun his search for the ‘ultimate conservative difference scheme’ in a series of classic papers. This work represented a complete extension of the ideas of Godunov to higher-order accuracy.

Godunov’s method was truly ingenious in nature, but is only first-order-accurate. This inaccuracy manifests itself as numerical viscosity that results in smeared solutions. The beauty of this scheme is that the physics of the local exact solution to the compressible flow equations is embedded in the differencing scheme. It is this characteristic that makes Godunov’s method so important. Van Leer⁶ takes the differencing scheme developed in Reference 5 and constructs a high-order Godunov (HOG) method. The differencing scheme used is unique in that it gives higher-order accuracy without dispersive ripples. This is accomplished by making the difference scheme itself non-linear and meeting monotonicity constraints.

These methods have been at the genesis of a wider development of methods in the past decade. The notion of Total Variation Diminishing (TVD) schemes as introduced by Harten^{7,8} has given a mathematical theory to back the use of these methods. Other schemes such as those developed by Colella and Woodward,⁹ Colella¹⁰ and Osher¹¹ are close to van Leer's work, but in each case enrich and extend the overall methodology. The contribution of Essentially Non-Oscillatory (ENO) schemes¹²⁻¹⁴ is worth noting. This class of methods extends the HOG methodology to arbitrarily high orders of accuracy.

A system of hyperbolic conservation laws can be written as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{U})}{\partial x} = 0. \quad (1)$$

A system of equations is strictly hyperbolic if its associated signal speeds are real in value and distinct. This condition applies to the eigenvalues of the flux Jacobian, $A \equiv \partial \mathbf{F} / \partial \mathbf{U}$ (i.e. that they be real), but also that they be distinct. Given the eigenvectors of A , \mathbf{r}_k , an eigenvalue, η_k , is defined as being genuinely non-linear if

$$\frac{\partial \eta_k}{\partial \mathbf{U}} \cdot \mathbf{r}_k \neq 0 \quad (2a)$$

and linearly degenerate if

$$\frac{\partial \eta_k}{\partial \mathbf{U}} \cdot \mathbf{r}_k = 0, \quad (2b)$$

as given by Lax.¹⁵ Examples of genuinely non-linear eigenvalues are the characteristic speeds associated with sound waves in the Euler equations. Shocks and rarefactions are associated with this sort of eigenvalue. A linearly degenerate eigenvalue is associated with the eigenvalue(s) associated simply with the material velocity in the Euler equations. Contact discontinuities are associated with this type of eigenvalue. This theory is of some consequence when considering what limiters to apply to a scheme.

Non-strictly hyperbolic systems are frequently solved by the methods discussed in this paper. In a number of cases (such as extension to multidimensional problems), this is not a severe problem. In other cases where the coincidence of eigenvalues is not trivial in nature, the consequences on the solution procedure are more severe. An example of the type of steps taken to deal with such cases is given in Reference 16.

This paper does not cover all the possible methods for extending high-resolution schemes to systems of equations. It also does not cover all the high-resolution schemes. Rather, this paper describes, and discusses several methods for extension with regard to one high-resolution scheme. This is done in an effort to remain as impartial and objective as possible. Further extensions can follow from this work in a logical fashion.

This paper is divided into five sections. Section 2 introduces the methods used for a scalar wave equation. In Section 3, each of these methods is extended to systems of equations. Section 4 presents and discusses results found using these methods for the Euler equations. Finally, concluding remarks are found in Section 5. An appendix describes the characteristic decomposition for both conserved and primitive variables.

2. PRELIMINARIES

In this paper, we will concentrate our efforts on one specific method and its extension to systems of equations. This method is a standard second-order HOG method augmented with TVD

limiters.^{17,18} As noted in References 12 and 19 the process of solving a problem with a Godunov-type method can be divided into two basic steps: reconstruction and evolution. The evolution step involves the use of some sort of exact or approximate Riemann solvers (see, for example, References 20 and 21). The issue at hand here is the method of reconstruction for systems of equations.

The reconstruction step requires that a piecewise polynomial (or some functional representation) be defined for each cell of the system to reconstruct the variables distribution in space to some level of desired accuracy. In this paper, the following form will be used for this polynomial:

$$P_j(x) = u_j + \widetilde{\Delta_j u} \frac{(x - x_j)}{\Delta_j x}, \quad x \in [x_{j-1/2}, x_{j+1/2}], \tag{3a}$$

where

$$\widetilde{\Delta_j u} = Q(r) \Delta_{j-1/2} u, \tag{3b}$$

with

$$\Delta_{j-1/2} u = u_j - u_{j-1}. \tag{3c}$$

The mesh spacing is $\Delta_j x = x_{j+1/2} - x_{j-1/2}$, $x_j = (x_{j+1/2} + x_{j-1/2})/2$ and $r = \Delta_{j+1/2} u / \Delta_{j-1/2} u$. The function $Q(r)$ is a limiter. These schemes are second-order-accurate. For higher spatial order, the degree of the polynomial can be increased.^{9,19,12}

Two common limiters will be used in this study. Figure 1 shows the limiters behaviour for a range of r . The limiters are the centred limiter⁶

$$Q_c(r) = \max[0, \min(2, 2r, \frac{1}{2}(1+r))] \tag{4a}$$

and the superbee limiter²²

$$Q_{SB}(r) = \max[0, \min(2, r), \min(1, 2r)]. \tag{4b}$$

The methodology chosen for extending the method derived for the scalar wave equation to systems can impact the choice of limiters. As will be seen in Section 4, the choice of limiters can have profound impact on the quality of the solutions.

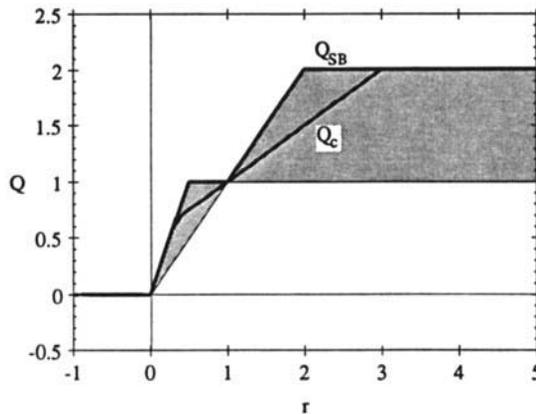


Figure 1. The two argument limiters used in this study are shown. The shaded area shows the second-order TVD region for the scheme used here. Both Q_c and Q_{SB} are second-order limiters, with Q_{SB} defining the upper boundary of the second-order TVD region

The polynomial is then used to define left and right states of the variables at each cell edge, u_L and u_R . These quantities are then used to determine a cell-edge numerical flux \hat{f}_{LR} via a Riemann solver. In the cases (except one as explained in Section 4.3) considered in this paper, Roe's approximate Riemann solver²³ will be used. The basics of this method for systems of equations is given in the following section. For a scalar wave equation, Roe's method can be written as

$$\hat{f}_{LR} = \frac{1}{2} [f(u_L) + f(u_R) - |a_{LR}|(u_R - u_L)], \quad (5a)$$

with

$$a_{LR} = \begin{cases} \frac{f(u_R) - f(u_L)}{u_R - u_L} & \text{if } u_L \neq u_R, \\ a_{LR} = a_L = a_R & \text{if } u_L = u_R. \end{cases} \quad (5b)$$

This gives an overall conservative numerical scheme of

$$u_j^{n+1} = u_j^n - \lambda (\hat{f}_{j+1/2, LR} - \hat{f}_{j-1/2, LR}), \quad (6a)$$

with

$$\hat{f}_{j+1/2, LR} = \frac{1}{\Delta t} \int_t^{t+\Delta t} f(u(x_{j+1/2}, \tau)) d\tau, \quad (6b)$$

where $\lambda = \Delta t / \Delta x$. For second-order temporal accuracy, the interface values for u should be time-centred estimates. For extension to systems not using a characteristic decomposition, it is likely that other approximate Riemann solvers will be used.

2.1. Lax–Wendroff-type differencing

Another issue easily addressed with simple model problems is time accuracy. For a second-order-accurate scheme spatially, it is often important to attain second-order accuracy temporally. A common practice is to use a Lax–Wendroff approach to time accuracy. From one point of view, this reduces to characteristic tracing at the cell edges to get a time-centred estimate of the cell-edge state. For our numerical scheme this yields the following form for cell-edge states:

$$u_{j+1/2, L}^{n+1/2} = u_j + \frac{1}{2} \widetilde{\Delta}_j u (1 - \lambda a_{LR}) \quad (7a)$$

and

$$u_{j+1/2, R}^{n+1/2} = u_{j+1} - \frac{1}{2} \widetilde{\Delta}_{j+1} u (1 + \lambda a_{LR}). \quad (7b)$$

This can also be viewed as evaluating in the integral in (6b) by a midpoint or trapezoidal rule. This comparison is shown in Figure 2.

2.2. Two-Step formulation

This procedure becomes more difficult when systems of equations are considered. To combat this difficulty, a procedure in the spirit of the two-step Lax–Wendroff scheme,^{24, 25} has been used.^{26, 27} The left and right states are computed from the reconstructive polynomial and then used to produce time-centred estimates for the cell-edge states. Given the cell-edge states, $u_{j+1/2, L}^n$ and $u_{j+1/2, R}^n$, computed with a high-order method, the time-centred estimates are

$$u_{j+1/2, L}^{n+1/2} = u_{j+1/2, L}^n - \frac{\lambda}{2} [f(u_{j+1/2, L}^n) - f(u_{j-1/2, R}^n)] \quad (8a)$$

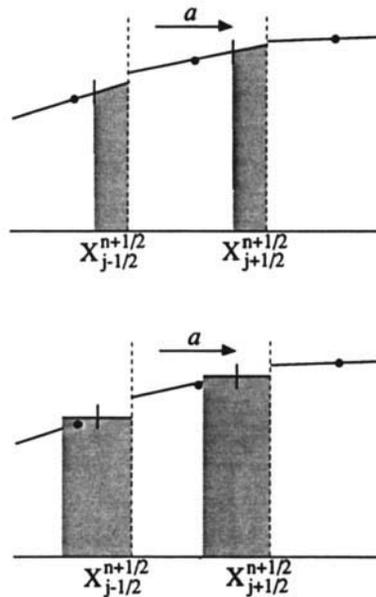


Figure 2. Two views of time-accurate computation of cell-edge values. The first view (top) corresponds to characteristic tracing to find the cell-edge time-centred values given by (7a) and (7b). The second view (bottom) corresponds to finding the cell-edge time-averaged values through temporal integration of (6b) using the midpoint rule. The top view also corresponds to evaluating the integral (6b) through the trapezoidal rule

and

$$u_{j+1/2,R}^{n+1/2} = u_{j+1/2,R}^n - \frac{\lambda}{2} [f(u_{j+3/2,L}^n) - f(u_{j+1/2,R}^n)]. \tag{8b}$$

This gives second-order temporal accuracy and is equivalent to the Lax–Wendroff-type procedure for scalar equations.

Remark 1. Davis²⁸ presents an alternate two-step method that is similar. In that method, the first step is

$$u_j^{n+1/2} = u_j^n - \frac{\lambda}{2} (f(u_{j+1/2,L}^n) - f(u_{j-1/2,R}^n)) \tag{9a}$$

and a second step of

$$u_{j+1/2,L}^{n+1/2} = u_j^{n+1/2} + \frac{1}{2} \widetilde{\Delta}_j u \tag{9b}$$

and

$$u_{j+1/2,R}^{n+1/2} = u_{j+1}^{n+1/2} - \frac{1}{2} \widetilde{\Delta}_{j+1} u. \tag{9c}$$

2.3. Component-wise extension

A third approach is also available. This approach involves the separate limiting of the flux vector and the solution variable. It has been used by Reference 29 with a high-order

Lax–Friedrichs solver.* This method makes use of an identity, $f(u) = au$, which implies that

$$\frac{\partial f}{\partial x} = a \frac{\partial u}{\partial x}, \tag{10}$$

where $a = \partial f(u)/\partial u$, which gives an equivalent form to that used above with a Lax–Wendroff approach. This approach has the limitation of being only correct at the point where $f(u)$ is evaluated if the equation is non-linear. Thus, it has the effect of ‘freezing’ the Jacobian at the point where it is evaluated. Specifically, this can be written

$$u_{j+1/2,L}^{n+1/2} = u_j + \frac{1}{2} \widetilde{\Delta_j} u - \frac{1}{2} \lambda \widetilde{\Delta_j} f_j \tag{11a}$$

and

$$u_{j+1/2,R}^{n+1/2} = u_{j+1} - \frac{1}{2} \widetilde{\Delta_{j+1}} u - \frac{1}{2} \lambda \widetilde{\Delta_{j+1}} f, \tag{11b}$$

where

$$\widetilde{\Delta_j} f = Q(r) \Delta_{j-1/2} f. \tag{11c}$$

Similar to the approach taken with the interpolation of the dependent variables, $r = \Delta_{j+1/2} f / \Delta_{j-1/2} f$ and $\Delta_{j-1/2} f = f_j - f_{j-1}$. Again for the scalar wave equation, this is equivalent to the Lax–Wendroff-type of time differencing.

3. METHOD FOR EXTENSION TO SYSTEMS

This section will concern itself with the subject of extending the methods described in the previous section to systems of equations. We will deal with the specific case of the Euler equations for the conservation of mass, momentum and total energy:

$$\frac{\partial \rho}{\partial t} + \frac{\partial m}{\partial x} = 0, \tag{12a}$$

$$\frac{\partial m}{\partial t} + \frac{\partial}{\partial x} \left(\frac{m^2}{\rho} + p \right) = 0 \tag{12b}$$

and

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x} \left(\frac{m}{\rho} (E + p) \right) = 0, \tag{12c}$$

where $E = \rho e + \frac{1}{2} m^2 / \rho$ with an equation of state $p = f(\rho, e)$ taken for an ideal gas is

$$p = \rho e (\gamma - 1), \tag{12d}$$

where γ is the ratio of specific heats. This equation set can be put in a convenient vector form, i.e. (1) with

$$U = \begin{bmatrix} \rho \\ m \\ E \end{bmatrix} \quad \text{and} \quad F(U) = \begin{bmatrix} m \\ \frac{m^2}{\rho} + p \\ \frac{m}{\rho} (E + P) \end{bmatrix}.$$

* The approach taken here uses Roe’s Riemann solver rather than a Lax–Friedrichs Riemann solver. This is done in order to put the different methods on equal footing.

The above system of equations can be written in the so-called primitive variable form. It has been suggested that this system of variable should be used to determine cell-edge states.^{30,9} In the above form the variables are conserved quantities $(\rho, m, E)^T$, but in the form given below the variables are $(\rho, u, e)^T$, the density, velocity and internal energy. This set of equations is

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial x} = 0, \quad (13a)$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0 \quad (13b)$$

and

$$\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} + \frac{p}{\rho} \frac{\partial u}{\partial x} = 0. \quad (13c)$$

Of the methods available for extending the scheme outlined in the previous section, the characteristic decomposition due to Roe²³ is the most common. In this method, a similarity transform takes the variable from the conservative form to a characteristic form. Each variable can then be computed at the cell edges from its characteristic contributions. This methodology can also be applied to the primitive variables in a similar manner.

The system of equations given above is hyperbolic if the eigenvalues of the flux Jacobian, $\partial \mathbf{F} / \partial \mathbf{U}$, are real and distinct.¹⁵ The Jacobian of the flux function is used to derive a characteristic decomposition of the system of equations; thus, in general,

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 \quad \Rightarrow \quad \frac{\partial \mathbf{U}}{\partial t} + A \frac{\partial \mathbf{U}}{\partial x} = 0, \quad (14a)$$

where $A = \partial \mathbf{F} / \partial \mathbf{U}$ is the flux Jacobian matrix. This is a local linearization of the system. The relation is exact for a linear system, but not for a non-linear system (away from the point where the Jacobian is evaluated). For a non-linear system, this linearization has the effect of 'freezing' the Jacobian locally. If we define the decomposition as

$$A = R \Lambda R^{-1},$$

Λ is a diagonal matrix with the eigenvalues of A , η^k , on the diagonal, R is the matrix of right eigenvectors (columns), and R^{-1} is the matrix left eigenvectors (rows). For linear systems, the characteristic equations are then defined as

$$\frac{\partial \alpha}{\partial t} + \Lambda \frac{\partial \alpha}{\partial x} = 0, \quad (14b)$$

where $\alpha = R^{-1} \mathbf{U}$. These equations can be solved with upwind biased methods to get physically correct propagation of information for data associated with each separate wave. This procedure will be used for non-linear systems to approximate wave propagation locally.

Thus, each characteristic is limited separately in defining the new cell-edge value of \mathbf{U} . For this purpose, we define

$$\widetilde{\Delta}_j u = \sum_{k=1}^3 r^k \widetilde{\Delta}_j \alpha^k, \quad (14c)$$

where

$$\widetilde{\Delta}_j \alpha = Q(r) \Delta_{j-1/2} \alpha \tag{14d}$$

for each component of U where $r = \Delta_{j+1/2} \alpha / \Delta_{j-1/2} \alpha$.

The characteristic approach must also be integrated into the attainment of temporal accuracy. Each wave in the above decomposition travels at different speeds and they can also travel in different directions. For this reason, the cell-edge quantities are computed from the following formulas:

$$U_{j+1/2,L} = U_j + \frac{1}{2} \sum_{k=1}^3 r^k (1 - \eta^k \lambda) \widetilde{\Delta}_j \alpha^k \tag{15a}$$

and

$$U_{j+1/2,R} = U_{j+1} - \frac{1}{2} \sum_{k=1}^3 r^k (1 + \eta^k \lambda) \widetilde{\Delta}_{j+1} \alpha^k. \tag{15b}$$

This decomposition can be done for both systems of equations above and is detailed in the appendix.

This method is aesthetically pleasing because the coupled non-linear system is locally reduced to a set of decoupled scalar equations. Because of this, the theory developed and applied to simpler model problems carries over without interference to systems. On the other hand, the expense associated with procedure (especially when multidimensional or more complex systems are considered) makes them less attractive than other alternatives. A modification of this method that is touted as increasing the robustness of the reconstruction is given in Reference 30. This method takes into account the direction of wave carrying information and only allows physically meaningful reconstructions to occur.

The other options described in Section 2 are somewhat more straightforward to implement for systems of equations. The two-step method is simply applied in a vector fashion, i.e.

$$U_{j+1/2,L}^{n+1/2} = U_{j+1/2,L}^n - \frac{\lambda}{2} [F(U_{j+1/2,L}^n) - F(U_{j-1/2,R}^n)] \tag{16a}$$

and

$$U_{j+1/2,R}^{n+1/2} = U_{j+1/2,R}^n - \frac{\lambda}{2} [F(U_{j+3/2,L}^n) - F(U_{j+1/2,R}^n)]. \tag{16b}$$

The cell-edge values at time level n can be computed in a component-wise or characteristic fashion. For the component-wise fashion the values are

$$U_{j+1/2,L}^n = U_j^n + \frac{1}{2} \widetilde{\Delta}_j U \tag{17a}$$

and

$$U_{j+1/2,R}^n = U_{j+1}^n - \frac{1}{2} \widetilde{\Delta}_{j+1} U, \tag{17b}$$

and for the characteristic extension

$$U_{j+1/2,L}^n = U_j^n + \frac{1}{2} \sum_{k=1}^3 r^k \widetilde{\Delta}_j \alpha^k \tag{17c}$$

and

$$U_{j+1/2,R}^n = U_{j+1}^n - \frac{1}{2} \sum_{k=1}^3 r^k \widetilde{\Delta}_{j+1} \alpha^k. \tag{17d}$$

Similarly the component-wise extension method can be extended by using limited values of the flux function for each of a system's equations. Thus, the method can be written as

$$\mathbf{U}_{j+1/2,L}^{n+1/2} = \mathbf{U}_j + \frac{1}{2}(\widetilde{\Delta}_j \mathbf{U} - \lambda \widetilde{\Delta}_j \mathbf{F}) \quad (18a)$$

and

$$\mathbf{U}_{j+1/2,R}^{n+1/2} = \mathbf{U}_{j+1} - \frac{1}{2}(\widetilde{\Delta}_{j+1} \mathbf{U} - \lambda \widetilde{\Delta}_{j+1} \mathbf{F}). \quad (18b)$$

For both of these methods, the computation of the cell-edge value could be done in either conservative, primitive or characteristic variables. The advantage of the two-step or the component-wise extension methods can only be obtained if the interpolation is done in either the conservative or primitive variables because of the relative simplicity of each formulation.

Another issue of some importance is the application of limiters in computing the piecewise polynomials. It is common practice to use a compressive limiter such as superbee on the field that produces the contact discontinuity. The compression given by the limiter maintains the sharpness of the interface. The same limiter when applied to shocks or rarefactions can produce entropy violating solutions. For the characteristic decomposition the implementation of this is quite clear. For other methods not involving characteristic decomposition it is usual practice to apply the compressive limiter to the computation of the density profile.⁹

In calculating the results given in the following section, the criteria given above was used in choosing the limiters. For characteristic decompositions of the equations, the centred limiter was used on the non-linear fields, and the superbee limiter was used on the linearly degenerate field. For methods employing the two-step or component-wise extension to systems, the superbee limiter was used on the density reconstruction, and the centred limiter was used on the remaining variables.

4. COMPARISON OF METHODS

In the following section, we will compare the performance of the methods for several standard test problems for the Euler equations in one space dimension. The results of this discussion should provide guidance for more complex systems of equations as well as guidance in a route to take in extending these methods to multidimensional problems. In the interests of saving space, Table I lists the abbreviations used in this section to describe the methods.

4.1. Sod's problem

The problem used by Sod³¹ to test a number of methods for solving the equations of compressible flow has become a standard test problem. The initial condition for this problem consists of two semi-infinite states separated at $t=0$, and the left and right states are set to the

Table I. Abbreviations for the methods used in this study

Scheme	Abbreviation	Section
Characteristic-conservative variables	CC	2.1
Characteristic-primitive variables	PC	2.1
Two-step-conservative variables	CR	2.2
Two-step-primitive variables	PR	2.2
Component-wise-conservative variables	CF	2.3
Component-wise-primitive variables	PF	2.3

following conditions:

for $X < 50.0$

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1.0 \end{bmatrix},$$

for $X \geq 50.0$

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.0 \\ 0.1 \end{bmatrix},$$

with $\gamma = 1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x = 1.0$) and the CFL number is set to 0.9. The solutions are shown at $t = 20$.

The solutions to Sod's problem can be seen in Figures 3–8. In general, the solutions are quite good and exhibit the qualities one would expect with a high-resolution numerical solution.

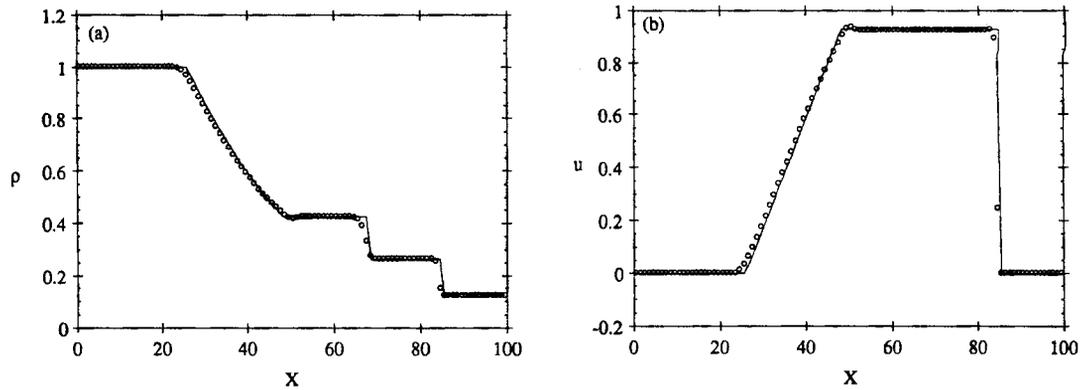


Figure 3. Sod's problem computed with the characteristic formulation with conservative variables. In these figures, the solid line denotes the exact solution, whereas the circles denote the approximate numerical solution

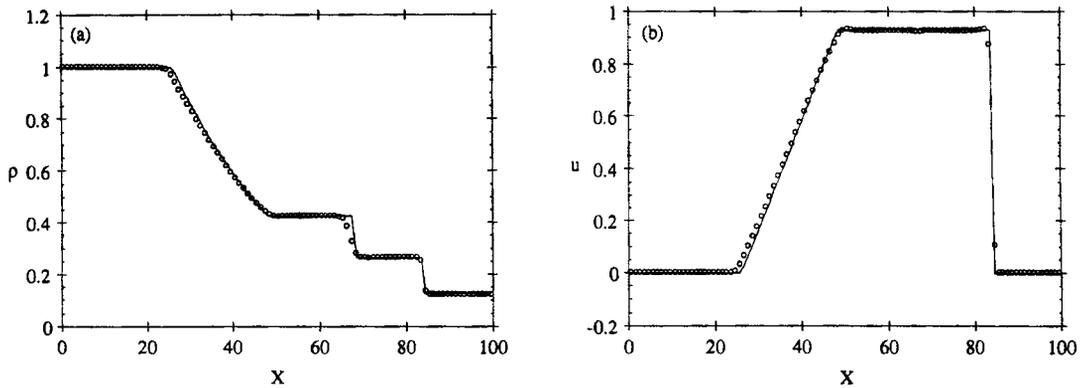


Figure 4. Sod's problem computed with the characteristic formulation with primitive variables

The solutions found with the CC method are seen in Figure 3. They are qualitatively quite good, with the only problem being the glitch in the velocity at the end of the rarefaction wave. With the PC method the velocity glitch is gone, but a small rise is before the shock. As can be seen in Figure 4, the density profile is nearly identical to that found with the CC method.

With the two-step formulation, the solutions are again quite good as can be seen in Figures 5 and 6. The major problems can be seen with the velocity profiles where small problems exist with at the end of the rarefaction wave and in the post-shock region of the flow. These problems are not major in nature. Major features of the flow field such as the shock, contact discontinuity and rarefaction wave are resolved well.

The component-wise extension of the schemes has a few more problems. In Figures 7 and 8 the solutions are shown. The shock wave is exceptionally sharp, improved over the other methods, but in both the conservative and primitive variable formulation there are a number of small oscillations in the velocity solution between the rarefaction and shock waves. In this case, these oscillations are not destructive, but detract from the overall quality of the solutions.

In Table II, the L_1 norm errors using these methods are shown. In these terms the best solution is the PC method with both of the two-step methods of slightly lower quality. The PF method is

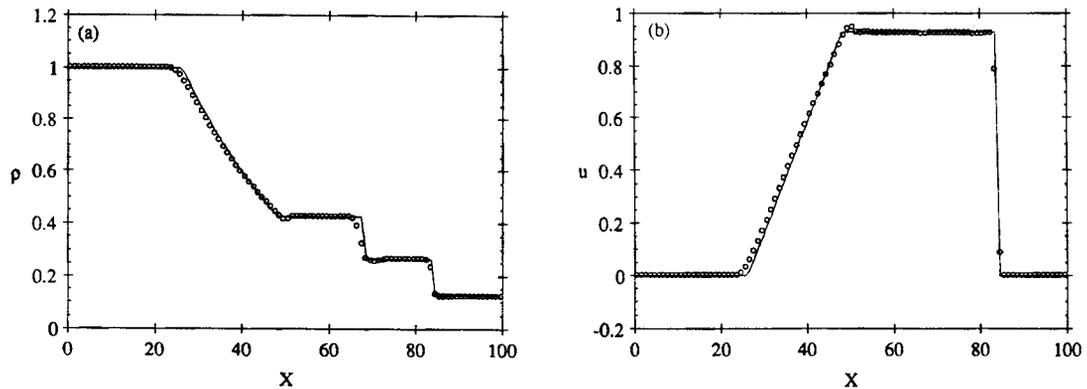


Figure 5. Sod's problem computed with the two-step formulation with conservative variables

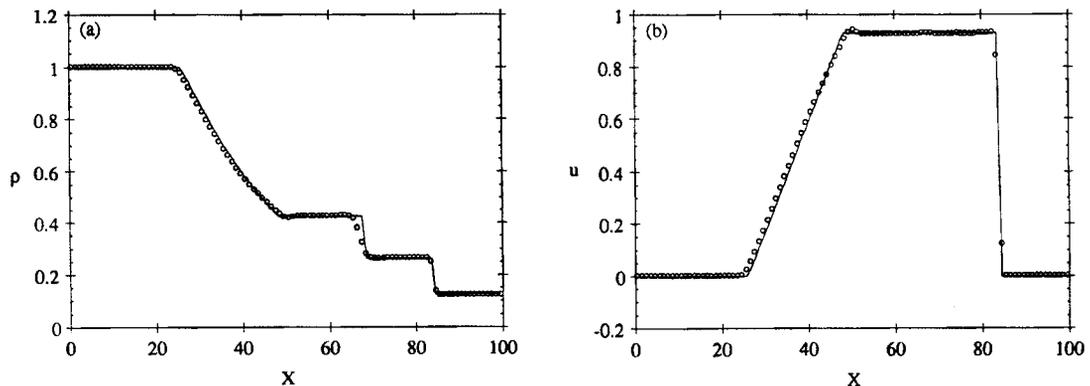


Figure 6. Sod's problem computed with the two-step formulation with primitive variables. Note the small spikes at the end of the rarefaction waves and the post-shock spike in the velocity solution

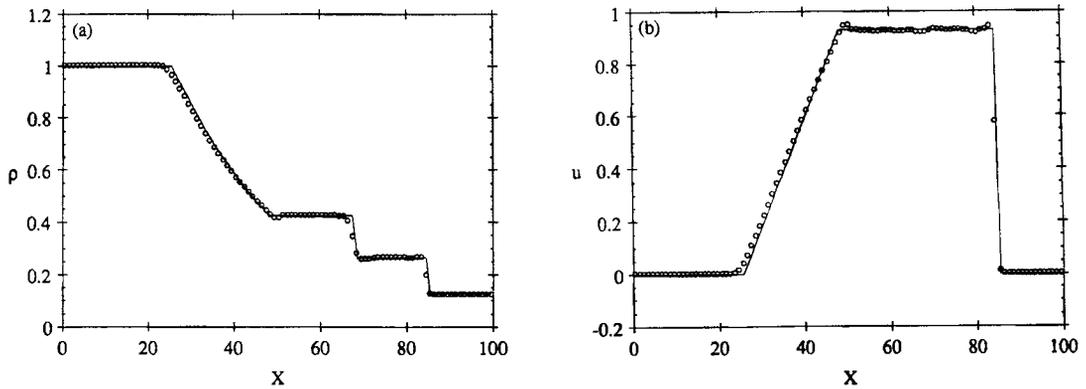


Figure 7. Sod's problem computed with the component-wise formulation with conservative variables. Note the small oscillations in the velocity solution between the rarefaction and shock waves

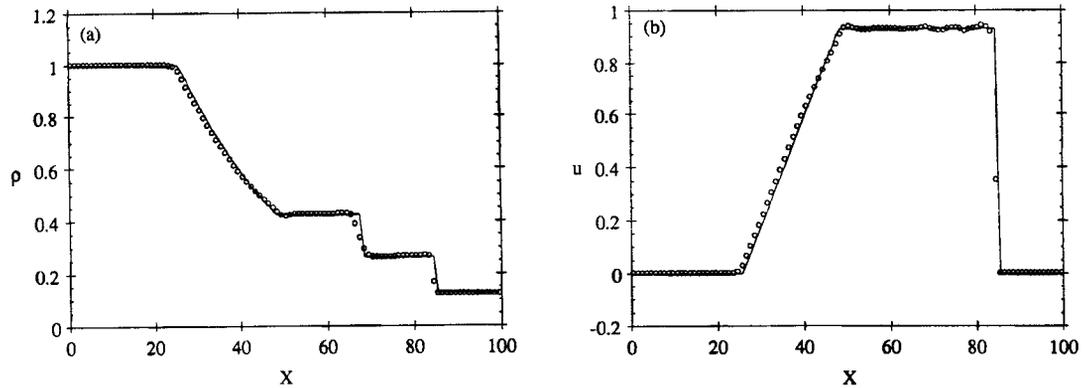


Figure 8. Sod's problem computed with the component-wise formulation with conservative variables. Note the small oscillations in the velocity solution between the rarefaction and shock waves

the worst, with the CC formulation slightly better. However, the better qualitative appearance of the CC makes it much superior to the PF method.

4.2. Lax's problem

Lax's problem is a shock tube problem similar to Sod's, but with one of the two semi-infinite states used as initial conditions not being at rest. The initial condition for this problem consists of two semi-infinite states separated at $t=0$, the left and right states are set to the following conditions:

for $X < 50.0$

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 0.445 \\ 0.698 \\ 3.528 \end{bmatrix},$$

Table II. The L_1 error norms for each scheme on Sod's problem

Scheme	Density	Velocity
CC	5.86×10^{-3}	1.19×10^{-2}
PC	4.90×10^{-3}	6.14×10^{-3}
CR	5.26×10^{-3}	7.27×10^{-3}
PR	5.45×10^{-3}	7.58×10^{-3}
CF	5.34×10^{-3}	9.33×10^{-3}
PF	6.20×10^{-3}	1.22×10^{-2}

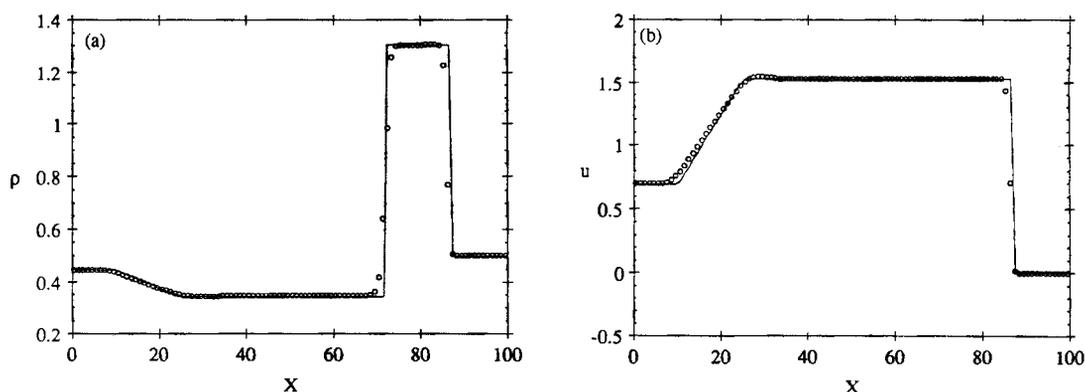


Figure 9. Lax's problem computed with the characteristic formulation with conservative variables. With the exception of this solution, all the solutions to Lax's problem have small spikes or oscillations associated with the contact discontinuity. This is indicative of the overcompressive nature of the limiter placed on the density. The conservative characteristic formulation guards against this problem

for $X \geq 50.0$

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.0 \\ 0.571 \end{bmatrix},$$

with $\gamma=1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x=1.0$) and the CFL number is set to 0.9. The solutions are shown at $t=15$.

The solutions to this problem by the methods discussed in this paper are shown in Figures 9–14. Again the solutions are quite good across the board, but problems with the methods show more strongly in the density profiles. The region between the shock wave and the contact discontinuity is sensitive to the limiter used, and in the non-characteristic methods, problems show up.

Figures 9 and 10 show the CC and PC solutions to Lax's problem, respectively. The only problem with these solutions is evident in the PC velocity solution where a small dip in the velocity is present coincident with the contact discontinuity. This problem also shows up with all the other methods. This is an artifact of the compressive superbee limiter used on the linearly degenerate wave. When other less compressive limiters are used, the problems associated with the contact discontinuity are removed from the solutions.

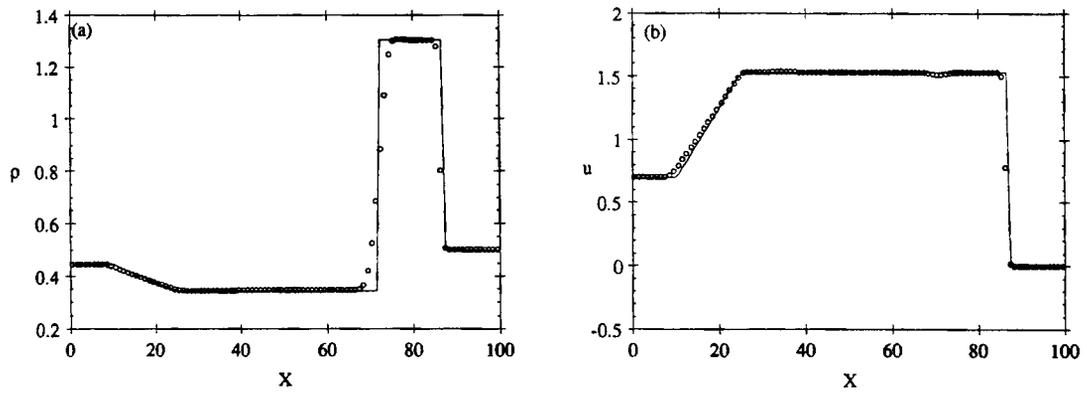


Figure 10. Lax's problem computed with the characteristic formulation with primitive variables. Despite using a characteristic formulation, a small oscillation is present with the contact discontinuity

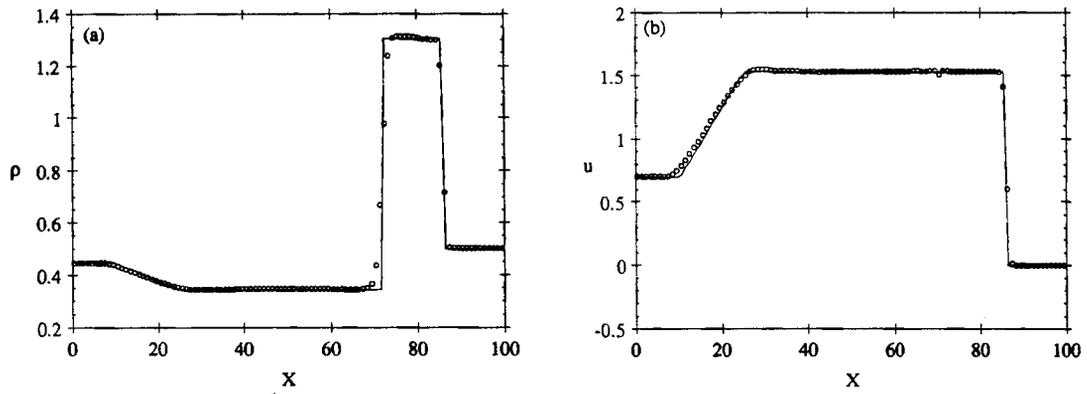


Figure 11. Lax's problem computed with the two-step formulation with conservative variables

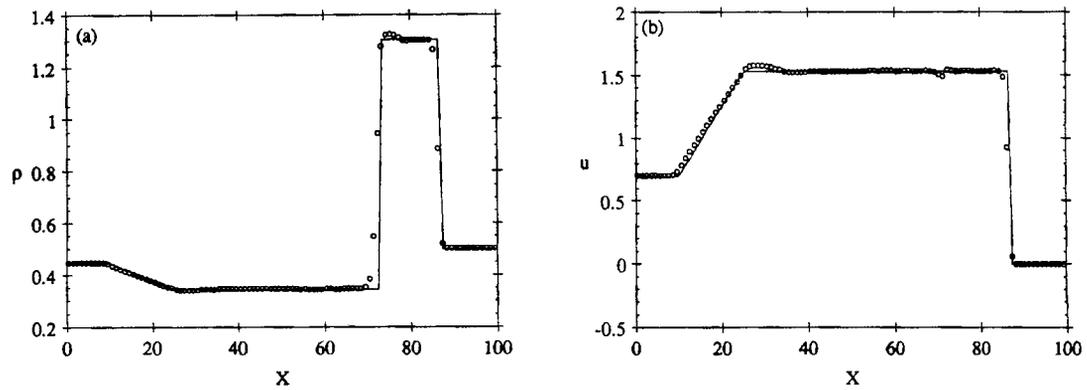


Figure 12. Lax's problem computed with the two-step formulation with primitive variables

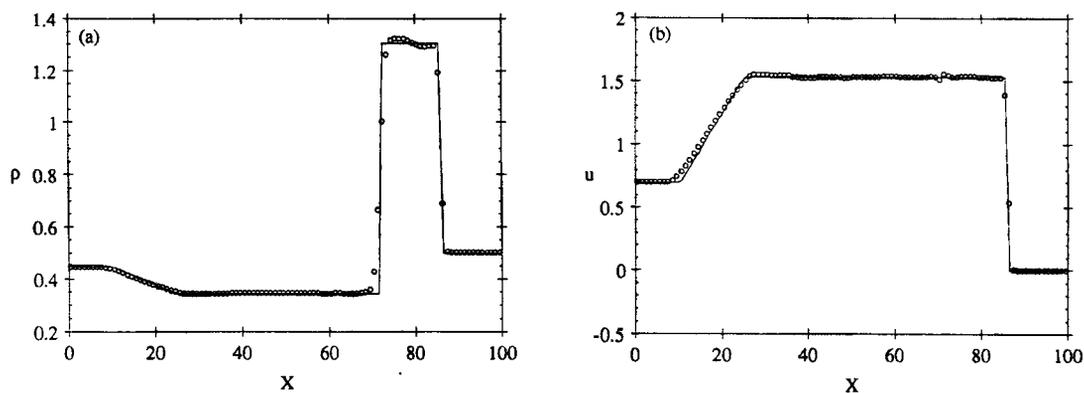


Figure 13. Lax's problem computed with the component-wise formulation with conservative variables

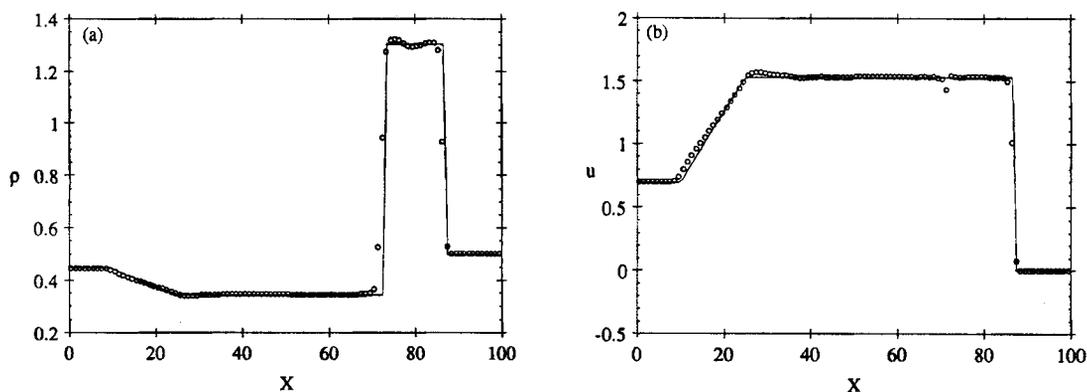


Figure 14. Lax's problem computed with the component-wise formulation with conservative variables

Table III. The L_1 error norms for each scheme on Lax's problem

Scheme	Density	Velocity
CC	1.46×10^{-2}	1.61×10^{-2}
PC	1.92×10^{-2}	1.42×10^{-2}
CR	1.30×10^{-2}	1.53×10^{-2}
PR	1.52×10^{-2}	1.61×10^{-2}
CF	1.29×10^{-2}	1.54×10^{-2}
PF	1.44×10^{-2}	1.62×10^{-2}

Figures 11–14 show the solutions found with other methods. These solutions all share common characteristics. The contact discontinuity causes oscillations in the solutions as evident in both the density and velocity profiles. These oscillations are more severe in the primitive variable formulations. These oscillations can be controlled through another choice of a limiter to apply to the density interpolation.

In terms of L_1 error (see Table III), the conclusions that are drawn are somewhat different to those found with Sod's problem. The velocity errors are very close in magnitude and no real conclusions can be drawn from them. The density errors seem to favour the conservative formulations, but for the two-step or component-wise formulations the differences are not profound.

4.3. Vacuum problem

As noted in Section 2, one case in this study does not use Roe's approximate Riemann solver. The case of the vacuum problem considered below cannot use Roe's solver as explained in Reference 32. For this case, a more diffusive scheme is used to maintain physical solutions. This is the HLLC Riemann solver,^{20,32,33} which is briefly described below.

This method has several desirable properties: its simplicity, ease of implementation and satisfaction of entropy inequalities. The general form of a flux function with this solver is

$$f_{LR} = \frac{b_{LR}^+ f(u_L) - b_{LR}^- f(u_R)}{b_{LR}^+ - b_{LR}^-} + \frac{b_{LR}^+ b_{LR}^-}{b_{LR}^+ - b_{LR}^-} (u_R - u_L), \quad (19a)$$

where $b_{LR}^+ = \max(0, b_{LR}^+)$ and $b_{LR}^- = \min(0, b_{LR}^+)$. The signal speeds b_{LR}^+ and b_{LR}^- are upper and lower bounds on the signal velocities, respectively. Reference 32 makes the suggestion for the computation of b_{LR}^+ and b_{LR}^- . The formulas are

$$b_{LR}^+ = \max(a_{R, \max}, a_{LR, \max}) \quad (19b)$$

and

$$b_{LR}^- = \min(a_{L, \min}, a_{LR, \min}), \quad (19c)$$

where max and min refer to the maximum and minimum characteristic speeds at the respective locations. The values for a_{LR} come from the Roe linearization that will be discussed below.

The vacuum problem is a shock tube problem where two identical states are moving away from each other at $t=0$. The states are kinetic-energy-rich, which causes problems for the finite difference schemes. The initial condition for this problem consists of two semi-infinite states separated at $t=0$, the left and right states are set to the following conditions:

for $X < 50.0$

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ -2.0 \\ 1.0 \end{bmatrix},$$

for $X \geq 50.0$

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 1.0 \\ 2.0 \\ 1.0 \end{bmatrix},$$

with $\gamma=1.4$. The domain is discretized into 100 cells of equal lengths ($\Delta x=1.0$) and the CFL number is set to 0.9. The solutions are shown at $t=10$. An additional caveat is that the computation of the stability criteria also involves the condition based on a condition similar to the 'tangling' or 'emptying' conditions in Lagrangian computations, i.e.

$$\Delta t \leq C \frac{\Delta x}{|u_{j+1/2} - u_{j-1/2}|}, \quad (20)$$

where $C \in [0, 1]$.

The solutions found with the CC, PC, PR and PF (Figures 15–20) methods are not worth much discussion. All of them are quite good and appear to be nearly identical in terms of resolution. Table IV shows this as well.

The solutions found with the CR and CF methods do warrant some discussion. The CR solution is shown in Figure 17 and the CF solution in Figure 19. Both solutions are of exceedingly poor quality. In fact, if measure had not been taken to prevent this, the computer code should

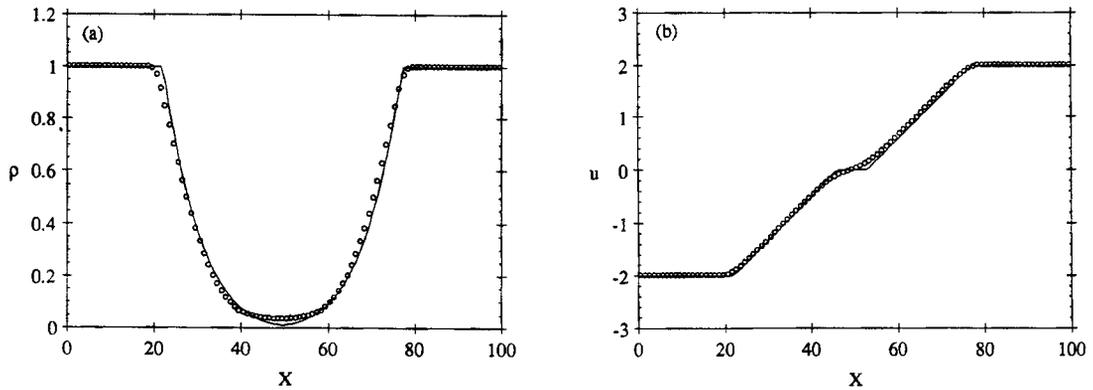


Figure 15. The vacuum problem computed with the characteristic formulation with conservative variables

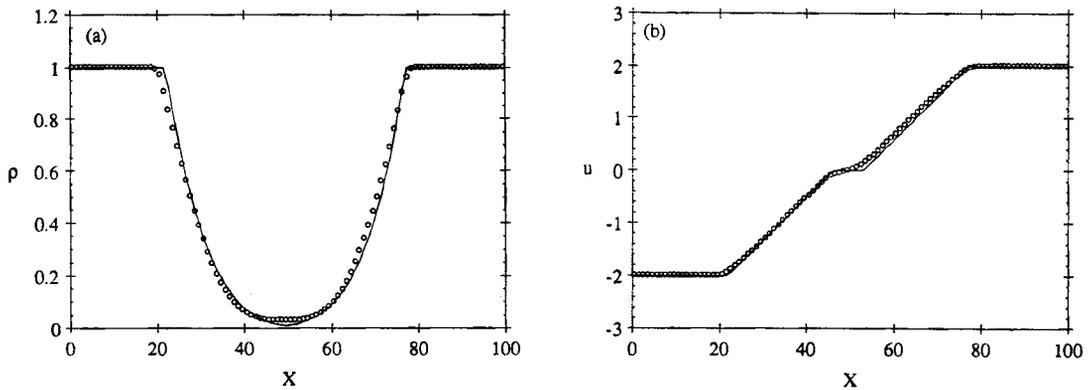


Figure 16. The vacuum problem computed with the characteristic formulation with primitive variables

Table IV. The L_1 error norms for each scheme on the vacuum problem

Scheme	Density	Velocity
CC	1.27×10^{-2}	2.63×10^{-2}
PC	1.24×10^{-2}	2.85×10^{-2}
CR	2.72×10^{-2}	1.00×10^{-1}
PR	1.20×10^{-2}	2.39×10^{-2}
CF	2.81×10^{-2}	5.85×10^{-2}
PF	1.20×10^{-2}	2.40×10^{-2}

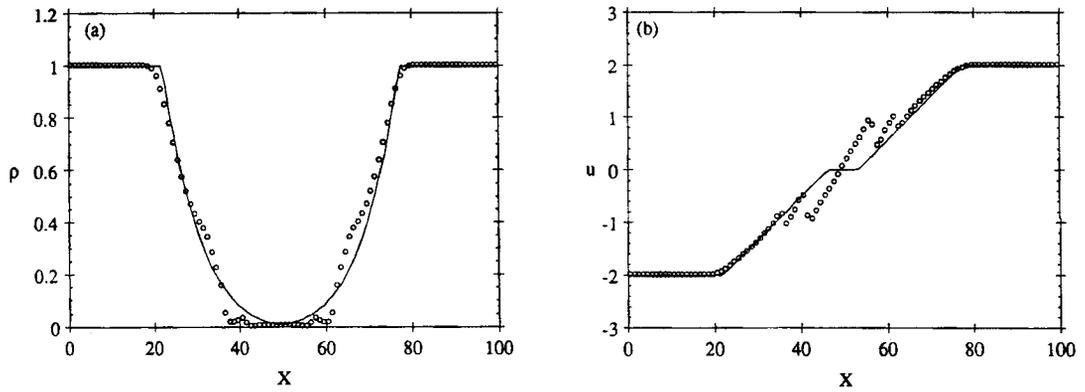


Figure 17. The vacuum problem computed with the two-step formulation with conservative variables. The use of conservative variables with this flow is disastrous. The total energy has become negative in the region around $X=50$

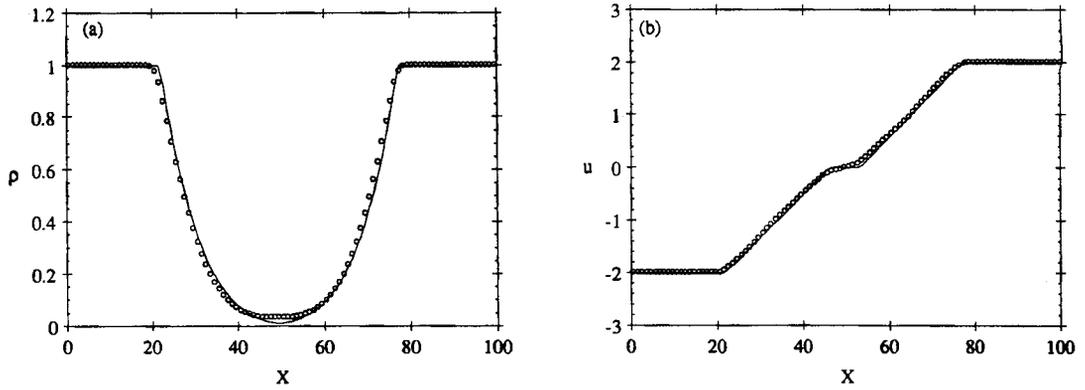


Figure 18. The vacuum problem computed with the two-step formulation with primitive variables

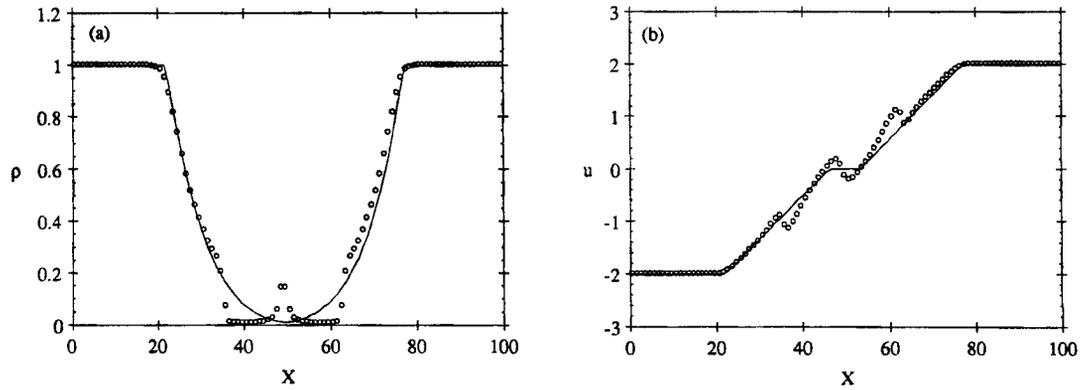


Figure 19. The vacuum problem computed with the component-wise formulation with conservative variables. The conservative variables have not guaranteed that positive-definite quantities (total energy) stay positive-definite

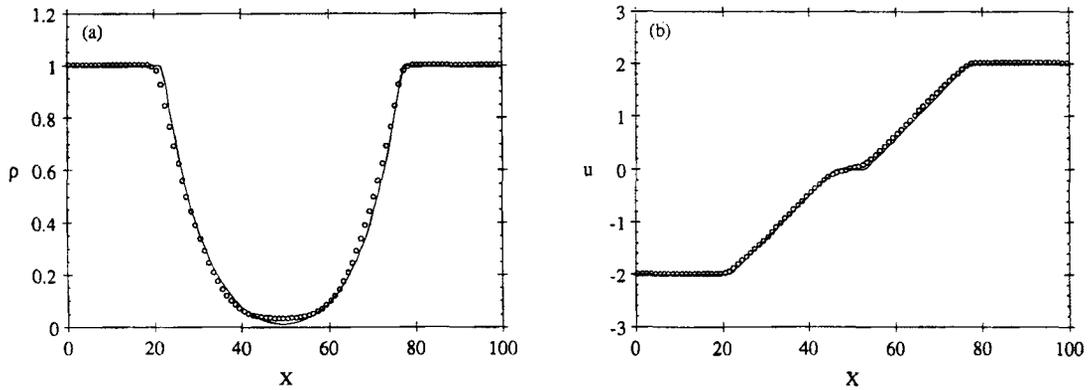


Figure 20. The vacuum problem computed with the component-wise formulation with primitive variables

have blown up early in the solution process. This is because the total energy in the solutions becomes negative in the vicinity of the vacuum in the solution. The use of the conservative variables in a non-characteristic method when the solution is kinetic-energy-rich causes the problem. This is akin to the problems with the Roe linearization studied in Reference 32. The interpolation of the variables creates non-physical states in the total energy. Lowering the compression of the limiters alleviates this problem as does moving to primitive or characteristic variables for the interpolation.

4.4. Blast wave problem

This blast wave problem was used by Woodward and Colella³⁴ to test a variety of high-resolution methods. This test turns out to be an extremely stringent test of numerical methods for solving hyperbolic conservation laws. The initial conditions consist of the following:

for $X \leq 10.0$

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 1000.0 \end{bmatrix},$$

for $10.0 > X > 90.0$

$$\begin{bmatrix} \rho_L \\ u_L \\ p_L \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.01 \end{bmatrix},$$

for $X \geq 90.0$

$$\begin{bmatrix} \rho_R \\ u_R \\ p_R \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.0 \\ 100.0 \end{bmatrix},$$

with $\gamma = 1.4$. The domain is discretized into 400 cells of equal lengths ($\Delta x = 0.25$) and the CFL number is set to 0.95. The boundary conditions play an important role in this problem and are

reflective at both the left ($X = 0$) and right ($X = 100$) walls. The solutions are shown at $t = 3.80$. The solution develops into two strong shock waves that collide. The result of this is a complex set of shock and rarefaction waves as well as contact discontinuities in a small region of space. These interactions are exceedingly difficult to resolve on a fixed Eulerian grid without prior knowledge of the solution so that the grid can be locally refined (certain adaptive meshing procedures can avoid the need for *a priori* knowledge of the solution). The 'exact' solution is found with the CC method using 2000 grid points.

The solutions are in general all quite good. The major features of this complex flow field are all depicted in the plotted density profiles (Figures 21–26). The major differences can be seen in the resolution of the contact discontinuity at $X \approx 60$, the 'well' at $X \approx 75$, and the peak at $X \approx 80$.

In Figure 21, the CC method's major problem is the clipping of the second peak in the solution. Other features are well resolved in comparison to the other methods. The PC method (Figure 22) smears all the features of the flow considerably more than the CC method. The CR method is generally like the CC method with the exception of the contact discontinuity at $X \approx 60$, which is

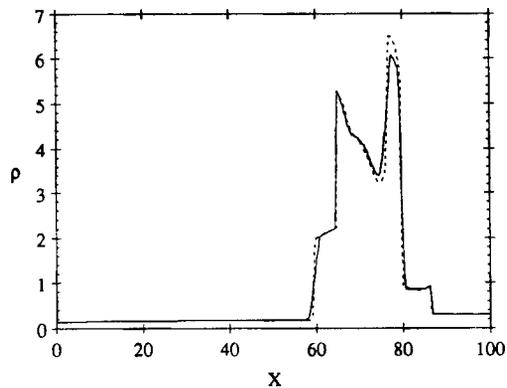


Figure 21. The blast wave problem computed with the characteristic formulation with conservative variables. The first peak is captured very well, but the second is clipped severely. With the blast wave solution, the 'exact' solution is marked by the dashed line and the approximate numerical solution by the solid line

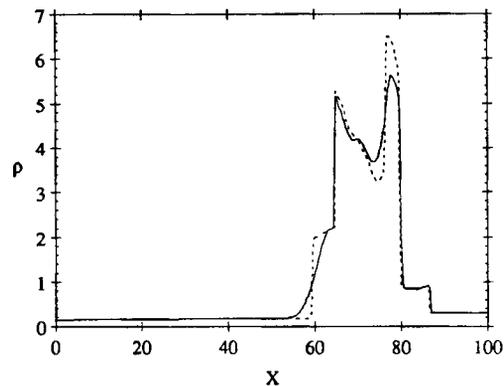


Figure 22. The blast wave problem computed with the characteristic formulation with primitive variables. Both peaks are clipped and the contact discontinuity at $X \approx 60$ is smeared

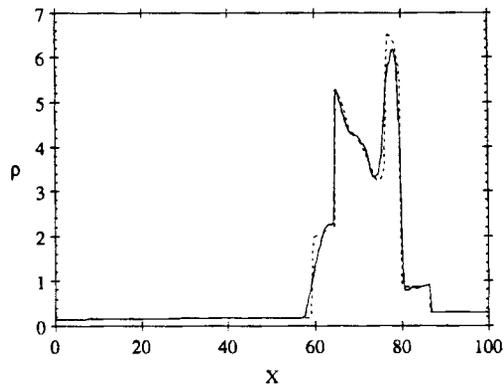


Figure 23. The blast wave problem computed with the two-step formulation with conservative variables. This is similar to Figure 21, but the contact discontinuity at $X \approx 60$ is smeared significantly more

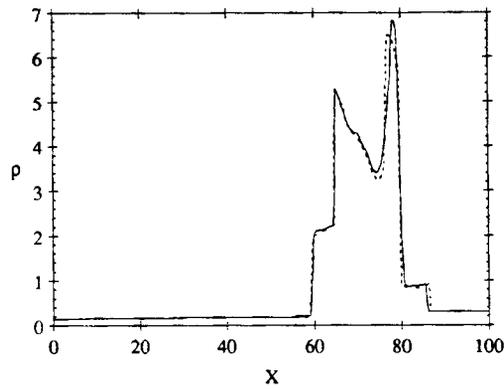


Figure 24. The blast wave problem computed with the two-step formulation with primitive variables. This solution is highly resolved and is of high quality with the exception of the overshoot of the second peak

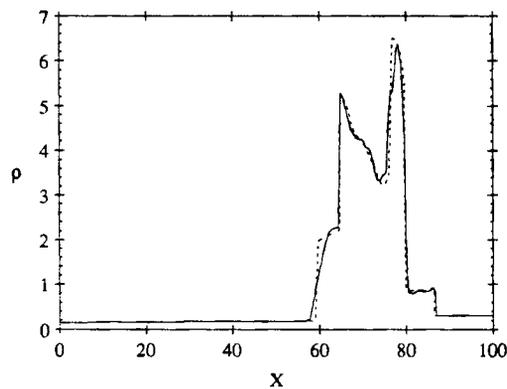


Figure 25. The blast wave problem computed with the component-wise formulation with conservative variables. This solution is fairly well resolved, but is somewhat 'noisier' than other solutions

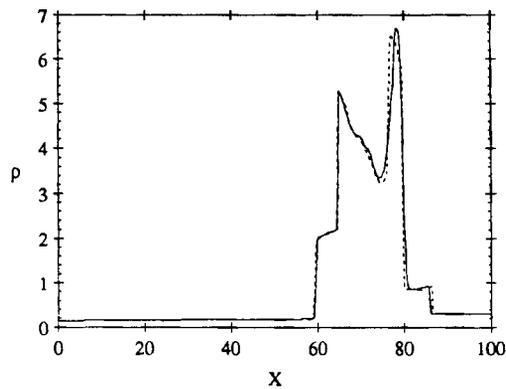


Figure 26. The blast wave problem computed with the component-wise formulation with conservative variables. This solution is very similar to Figure 24

Table V. The times for the blast wave solution computation using each method

Scheme	Total time (s)	Percentage in reconstruction
CC	81.93	49.58
PC	79.41	49.55
CR	82.49	43.12
PR	72.04	42.57
CF	84.22	40.44
PF	69.07	40.54

smearred much more than that by the CC method. The solution is somewhat 'noisier' with over/undershoots in several locations. These characteristics are duplicated in large part by the CF method (cf. Figures 23 and 25).

The PR and PF methods produce nearly same results. Both solutions are remarkably crisp and each feature in the flow field is sharply defined. Figures 24 and 26 also show the major detriment to these solutions. The second peak ($X \approx 80$) significantly overshoots the 'exact' solution. Nevertheless, the solution found by these methods is quite good in all other respects.

5. CONCLUDING REMARKS

Table V shows the total time taken for the blast wave solutions and the percentage of that time taken by the reconstruction of the cell-edge values.† In terms of economy, the PR and PF methods have clear advantages. Taking this into account with the results in mind several conclusions can be drawn. These conclusions are summarized below:

- All the methods described in the paper produce quality results.
- When a non-characteristic extension is used care must be taken in applying limiters (to not over-compress the density).
- For non-characteristic extensions, the primitive variables formulation should be used.
- Non-characteristic formulations using the primitive variables are lower in cost.

† The timings were done on a SPARCStation 2 running SunOS 4.1.1b.

Another point not emphasized here has been extension to multiple dimensional problems. All of these methods can be used with a dimensional splitting method, but the two-step method has clear applicability to a purely multidimensional methods without splitting. This is clearly an advantageous feature. In sum, both of the characteristic approaches (CC and PC) are reliable and produce excellent results in all cases. The two-step primitive variable method (PR) with appropriate selection of limiters is both economical and has applicability to a multidimensional algorithm.

ACKNOWLEDGEMENT

This work was performed under the auspices of the U.S. Department of Energy. Part of this work is contained in a Ph.D. dissertation written at the University of New Mexico.³⁵

APPENDIX: THE CHARACTERISTIC DECOMPOSITION OF THE EULER EQUATIONS

This appendix is a detailed description of the characteristic decomposition used in obtaining the results given in this paper. For the Euler equations in conservation from the flux Jacobian is

$$A = \begin{bmatrix} 0 & 1 & 0 \\ (\gamma-3)\frac{u^2}{2} & (3-\gamma)u & \gamma-1 \\ (\gamma-1)u^3 - uH & H - 3\frac{(\gamma-1)}{2}u^2 & \gamma u \end{bmatrix}, \tag{21a}$$

where $H = (E + p)/\rho$. The eigenvalues of this matrix are

$$(\eta^1, \eta^2, \eta^3) = (u, u + c, u - c). \tag{21b}$$

The right eigenvectors form a matrix

$$R = (r^1, r^2, r^3) = \begin{bmatrix} 1 & 1 & 1 \\ u & u + c & u - c \\ \frac{1}{2}u^2 & H + uc & H - uc \end{bmatrix} \tag{21c}$$

and by using

$$z_1 = \frac{1}{2}(\gamma - 1)\frac{u^2}{c^2},$$

and

$$z_2 = \frac{\gamma - 1}{c^2},$$

the left eigenvectors form a matrix

$$R^{-1} = \begin{bmatrix} l^1 \\ l^2 \\ l^3 \end{bmatrix} = \begin{bmatrix} 1 - z_1 & z_2 u & -z_2 \\ \frac{1}{2}\left(z_1 - \frac{u}{c}\right) & -\frac{1}{2}\left(z_2 u - \frac{1}{c}\right) & \frac{1}{2}z_2 \\ \frac{1}{2}\left(z_1 + \frac{u}{c}\right) & -\frac{1}{2}\left(z_2 u + \frac{1}{c}\right) & \frac{1}{2}z_2 \end{bmatrix}. \tag{21d}$$

The equations in primitive form give a much simpler system. The equation set does not have a conservation form and can be written as

where
$$\frac{\partial \mathbf{U}}{\partial t} + A \frac{\partial \mathbf{U}}{\partial x} = 0,$$

$$A = \begin{bmatrix} u & \rho & 0 \\ \frac{e(\gamma-1)}{\rho} & u & \gamma-1 \\ 0 & \frac{p}{\rho} & u \end{bmatrix}. \quad (22a)$$

Again, the eigenvalues of this matrix are

$$(\eta^1, \eta^2, \eta^3) = (u, u+c, u-c). \quad (22b)$$

The right eigenvectors form a matrix

$$R = (r^1, r^2, r^3) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -\frac{c}{\rho} & \frac{c}{\rho} \\ \frac{p}{(\gamma-1)\rho^2} & \frac{p}{\rho^2} & \frac{p}{\rho^2} \end{bmatrix} \quad (22c)$$

and by using $z_1 = (\gamma-1)\rho^2$ and $z_2 = 2\gamma p$, the left eigenvectors form a matrix

$$R^{-1} = \begin{bmatrix} l^1 \\ l^2 \\ l^3 \end{bmatrix} = \begin{bmatrix} \frac{\gamma}{\gamma-1} & 0 & -\frac{2z_1}{z_2} \\ \frac{1}{2\gamma} & -\frac{\rho}{2c} & \frac{z_1}{z_2} \\ \frac{1}{2\gamma} & \frac{\rho}{2c} & \frac{z_1}{z_2} \end{bmatrix}. \quad (22d)$$

REFERENCES

1. S. K. Godunov, 'Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics', *Math. Sbornik*, **47**, 271–306 (1959).
2. S. K. Godunov, A. W. Zabrodyn and G. P. Prokopov, 'A computational scheme for two-dimensional nonstationary problems of gas dynamics and calculation of the flow from a shock wave approaching steady-state', *USSR J. Comput. Math. Mathemat. Phys.*, **1**, 1187–1219 (1961).
3. J. P. Boris and D. L. Book, 'Flux-corrected transport I. SHASTA, a fluid transport algorithm that works', *J. Comput. Phys.*, **11**, 38–69 (1973).
4. B. van Leer, 'Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme', *J. Comput. Phys.*, **14**, 361–370 (1974).
5. B. van Leer, 'Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection', *J. Comput. Phys.*, **23**, 276–299 (1977).
6. B. van Leer, 'Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method', *J. Comput. Phys.*, **32**, 101–136 (1979).
7. A. Harten, 'High resolution schemes for hyperbolic conservation laws', *J. Comput. Phys.*, **49**, 357–393 (1983).
8. A. Harten, 'On a class of high resolution total-variation-stable finite-difference schemes', *SIAM J. Numer. Anal.*, **21**, 1–23 (1984).
9. P. Colella and P. Woodward, 'The piecewise parabolic method (PPM) for gas-dynamical simulations', *J. Comput. Phys.*, **54**, 174–201 (1984).

10. P. Colella, 'A direct Eulerian MUSCL scheme for gas dynamics', *SIAM J. Sci. Stat. Comput.*, **6**, 104–117 (1985).
11. S. Osher, 'Convergence of generalized MUSCL schemes', *SIAM J. Numer. Anal.*, **22**, 947–961 (1985).
12. A. Harten, B. Engquist, S. Osher and S. Chakravarthy, 'Uniformly high order accurate essentially non-oscillatory schemes, III', *J. Comput. Phys.*, **71**, 231–303 (1987).
13. C.-W. Shu and S. Osher, 'Efficient implementation of essentially non-oscillatory shock-capturing schemes', *J. Comput. Phys.*, **77**, 439–471 (1988).
14. C.-W. Shu and S. Osher, 'Efficient implementation of essentially non-oscillatory shock-capturing schemes II', *J. Comput. Phys.*, **83**, 32–78 (1989).
15. P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, SIAM, Philadelphia, PA, 1972.
16. J. B. Bell, P. Colella and J. A. Trangenstein, 'High order Godunov methods for general systems of hyperbolic conservation laws', *J. Comput. Phys.*, **82**, 362–397 (1989).
17. W. J. Rider, 'Limiters in the high resolution solution of hyperbolic conservation laws', *Technical Report LA-UR-91-3568*, Los Alamos National Laboratory, 1991.
18. P. K. Sweby, 'High-resolution schemes using flux limiters for hyperbolic conservation laws', *SIAM J. Numer. Anal.*, **21**, 995–1011 (1984).
19. B. van Leer, 'Upwind-difference methods for aerodynamic problems governed by the Euler equations', in B. Engquist *et al.*, (ed.), *Lectures in Applied Mathematics*, Vol. 22, 1985, pp. 327–336.
20. A. Harten, P. D. Lax and B. van Leer, 'On upstream differencing and Godunov-type schemes for hyperbolic conservation laws', *SIAM Rev.*, **25**, 35–61 (1983).
21. W. J. Rider, 'The use of approximate Riemann solvers with Godunov's method in Lagrangian coordinates', *Technical Report LA-UR-91-2555*, Los Alamos National Laboratory, 1991; to be published *Comput. Fluids*.
22. P. L. Roe, 'Some contributions to the modelling of discontinuous flows', in B. Engquist *et al.* (ed.), *Lectures in Applied Mathematics*, Vol. 22, 1985, pp. 163–193.
23. P. L. Roe, 'Approximate Riemann solvers, parameter vectors, and difference schemes', *J. Comput. Phys.*, **43**, 357–372 (1981).
24. S. Z. Burstein, 'Finite-difference calculations for hydrodynamic flows containing discontinuities', *J. Comput. Phys.*, **1**, 198–222 (1966).
25. R. D. Richtmyer, 'A survey of difference methods for non-steady gas dynamics', *Technical Report NCAR Tech Note 63-2*, NCAR, 1963.
26. G. D. van Albada, B. van Leer and W. W. Roberts, 'A comparative study of computational methods in cosmic gas dynamics', *Astron. Astrophys.*, **108**, 76–84 (1982).
27. B. van Leer, 'On the relation between the upwind-differencing schemes of Godunov, Engquist–Osher and Roe', *SIAM J. Sci. Stat. Comput.*, **5**, 1–20 (1984).
28. S. F. Davis, 'Simplified second-order Godunov-type methods', *SIAM J. Sci. Stat. Comput.*, **9**, 445–473 (1988).
29. H. Nessyahu and E. Tadmor, 'Non-oscillatory central differencing for hyperbolic conservation laws', *J. Comput. Phys.*, **87**, 408–463 (1990).
30. P. Colella, 'Multidimensional upwind methods for hyperbolic conservation laws', *J. Comput. Phys.*, **87**, 171–200 (1990).
31. G. Sod, 'A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws', *J. Comput. Phys.*, **27**, 1–31 (1978).
32. B. Einfeldt, C. D. Munz, P. L. Roe and B. Sjögreen, 'On Godunov-type methods near low densities', *J. Comput. Phys.*, **92**, 273–295 (1991).
33. B. Einfeldt, 'On Godunov-type methods for gas dynamics', *SIAM J. Numer. Anal.*, **25**, 294–318 (1988).
34. P. Woodward and P. Colella, 'The numerical simulation of two-dimensional fluid flow with strong shocks', *J. Comput. Phys.*, **54**, 115–173 (1984).
35. W. J. Rider, 'The design of high-resolution upwind shock-capturing methods', *Ph.D. Thesis*, University of New Mexico, 1992.