SAND2022-6744C

Sandia National Laboratories

Verification and Validation Using the Null Hypothesis as a Philosophical Grounding

Bill Rider, May 26,2022

AMSE VVUQ 2022, College Statio, TX

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

SAND2022-???? C

SAND2022-???? C

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- The Null Hypothesis and its generalization
- Foundations of verification and Validation
- Applying the null hypothesis to V&V
- An example from verification of shocks

Bottom Line: Working from an assumption that the code/model is “wrong” leads to better evidence more convincing that it is correct.

How to best build a strong case that a code is correct?

The classical Null hypothesis

- The null hypothesis is a mainstay of science using statistical studies and often proved via P values
- The basic premise is looking at the evidence that the central hypothesis of the study can be untrue, or no effect
- This is a direct assault on the study and whether the evidence could say the hypothesis in utterly false
- For much of science, the null hypothesis is an expected premise to be directly addressed in a study
- There is the less common and complementary idea of the alternative hypothesis

Verification and validation are essential to the quality of simulation.

Complementary

- Verification ≈ Solving the equations correctly
  - Mathematics/Computer Science issue
  - Applies to both codes and calculations
- Validation ≈ Solving the correct equations
  - Physics/Engineering (i.e., modeling) issue
  - Applies to both codes and calculations

- **Code Verification** is used to determine the correctness of the code, the rate of convergence is primary, error is secondary.

- **Solution Verification** is error/uncertainty estimation, the error is primary and the rate of convergence is secondary

Generalizing the Null Hypothesis to V&V

How can these concepts apply to verification and validation?

- The first notion is associated with mindset: do we approach V&V as if the code or model is correct? Or incorrect?
- The correct mindset leads to biasing analysis of evidence toward affirmative views of the code or model.
- The incorrect mindset looks at evidence with a bias toward being wrong, and if that evidence is unconvincing the case for being correct is stronger.
- The null hypothesis as a philosophy makes you ask harder questions of the code or model.

The Advantages of the Null Hypothesis as a Philosophy

- This avoids tendencies to approach V&V as a “box checking” exercise where V&V is simply a prelude to useful work.
- This avoids tendencies to reject information that reflects poorly on the code or model
- The null hypothesis orients the study toward harder questions, and more convincing evidence.
- The harder questions will spur code developers or analysts to build better codes or models. It is much more likely to find problems and reflect these back to ”stakeholders”
- Probably more necessary in code verification than any of the other activities in V&V

An example of the Null Hypothesis in Action: Verification of a shock physics code

- Shock physics codes are essential for solving a broad class of problems in defense science
- These methods and codes are mature. Testing has been done for decades, but there are long standing shortcomings
- There is a well developed mathematical theory and this theory is your friend for verification work
- There are common accepted practices that show standard verification practices and their shortcomings
- These practices undermine the value of code verification for solution verification and validation

Shock physics methods and codes

- The computation of shock physics was transformed by high resolution methods during the 1970's and 1980's.
- These high resolution methods allow the merger of high-order approximations for discontinuous flows associated with shocks
- Powerful mathematical theory guides the construction of the codes and the understanding of the results of the analysis
- Shock waves also have serious limitations on their accuracy with solutions relegated to first order accuracy or less

Relevant mathematical theory: a summary

For all verification work mathematical theory is your best friend!

- The Lax-Richtmyer equivalence theorem: the foundation of verification
- The Lax-Wendroff theorem demanding conservation form for weak solutions
- Entropy conditions needed to choose the physically relevant weak solution
- The Majda-Osher theorem, first-order accuracy for discontinuous solutions
- The Hou-LeFloch theorem, loss of convergence to correct solutions without conservation form

Verification and numerical analysis are intimately and completely linked.

The results that verification must produce are defined by the formal analysis of the methods being verified.

The numerical analysis results are typically (always) defined in the asymptotic range of convergence for a method.

- This range is reached as the discretization parameter (mesh, time step, angle, etc.) becomes "small" i.e., asymptotically "*close to zero*".

Practically, the asymptotic range is rarely achieved by verification practitioners in meaningful simulations.

Hence verification is not generally practiced where it is formally valid!

One can being testing the full Euler equations with a smooth wave (that will form a shock). It is smooth until the shock forms and can expose formal order of accuracy.

Introduced by Andy Cook and Bill Cabot (LLNL)

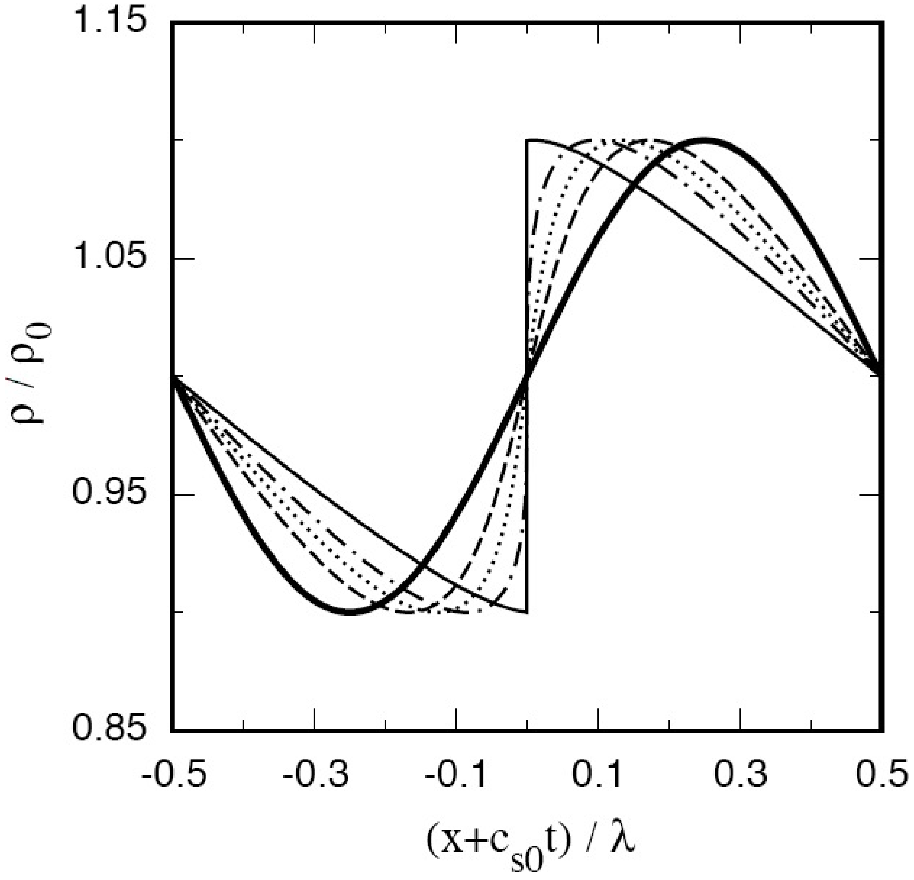
$$\rho / \rho_0 = 1 + \varepsilon \sin\left(2\pi x/\lambda\right)$$

$$p / p_0 = \left(\rho / \rho_0\right)^{\gamma}$$

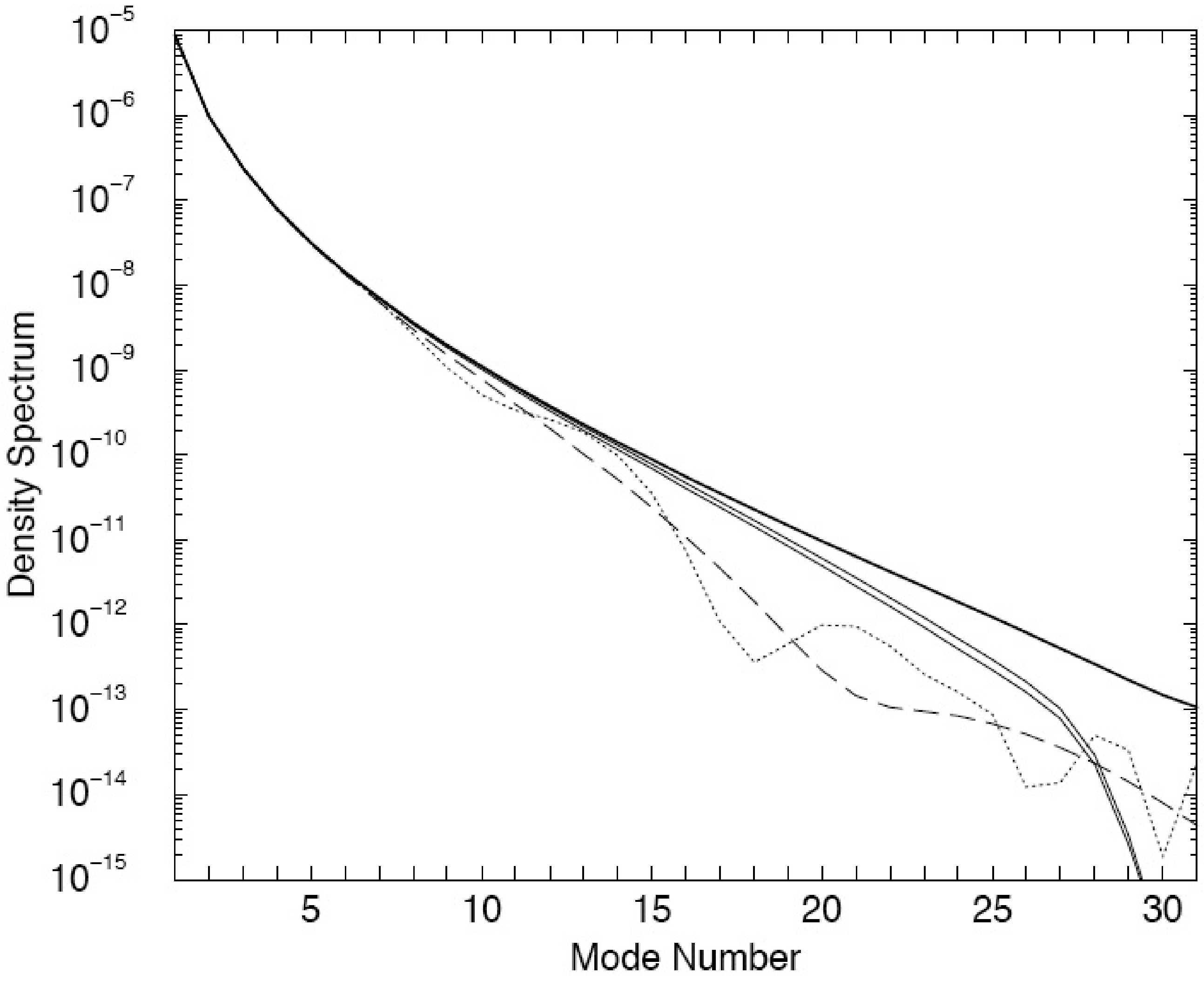
$$c / c_0 = \left(\rho / \rho_0\right)^{(\gamma - 1)/2}$$

$$u = 2\left(c_0 - c\right)/\left(\gamma - 1\right)$$

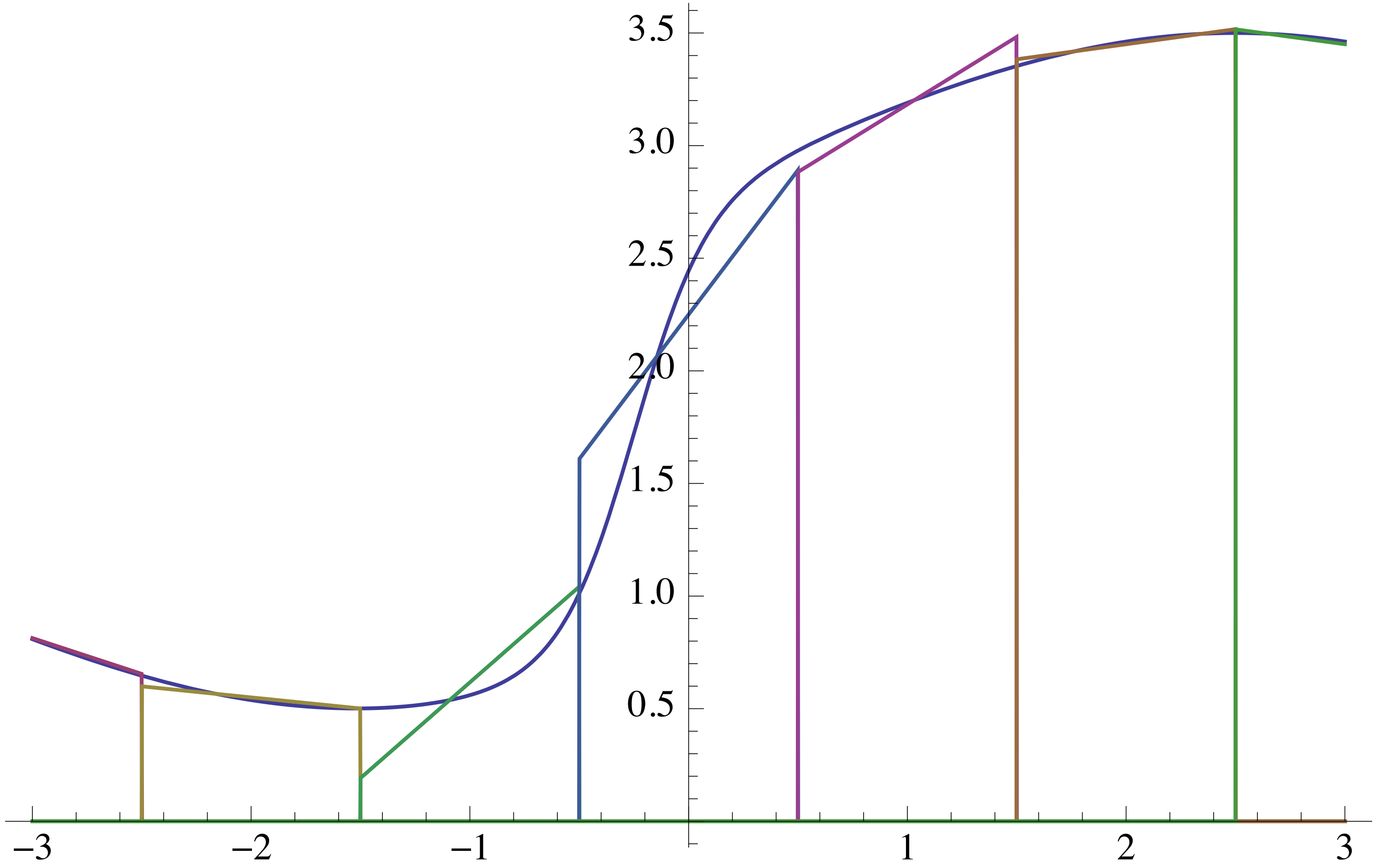
$$\rho_0 = 10^{-3}; p_0 = 10^{6}; \gamma = 5/3$$



Before the shock forms the results are analytical and a spectral plot is useful. Gives interesting information about methods.



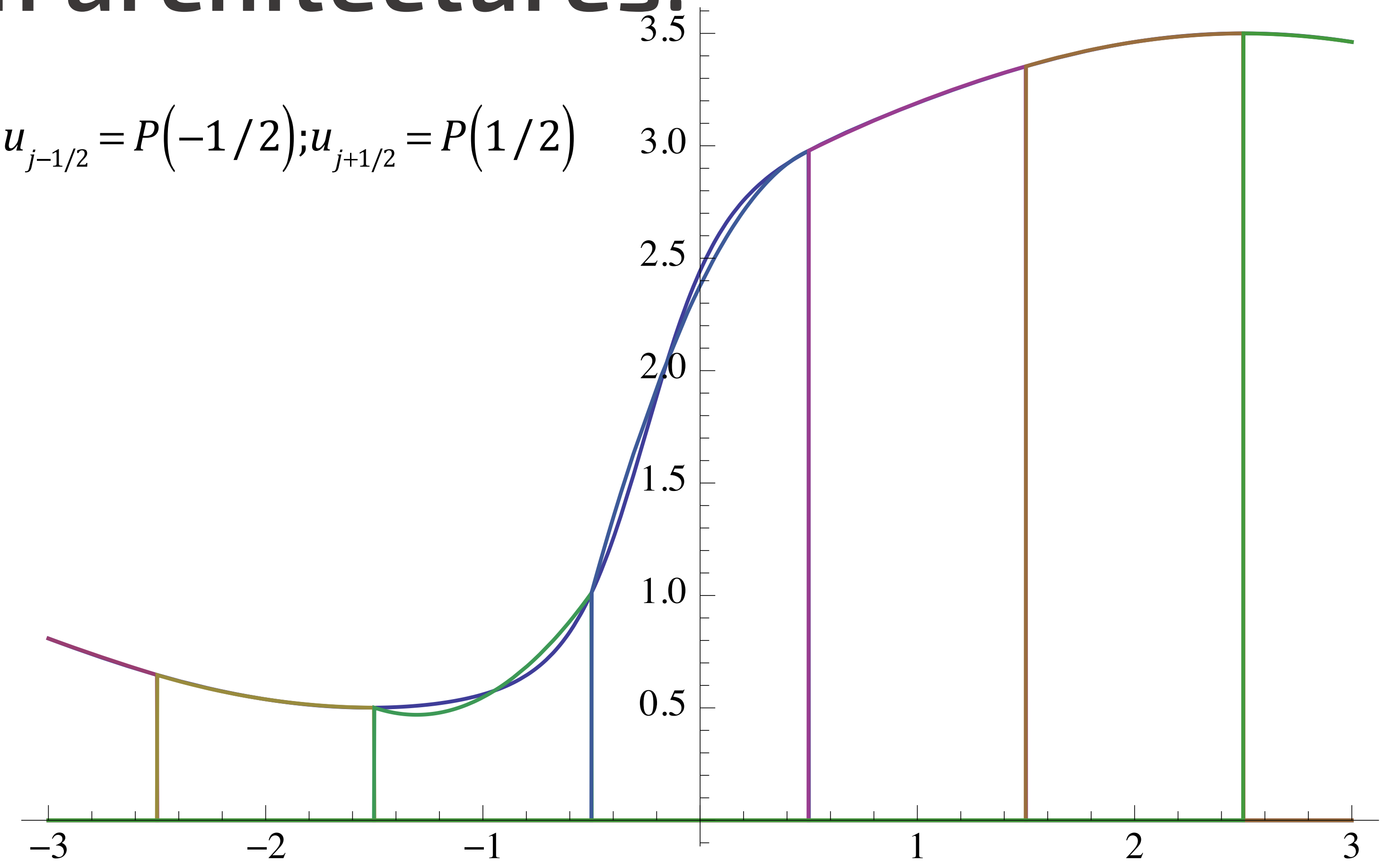
The piecewise linear method is the basis of the second-order methods in many shock codes



Integrated Error = 0.448

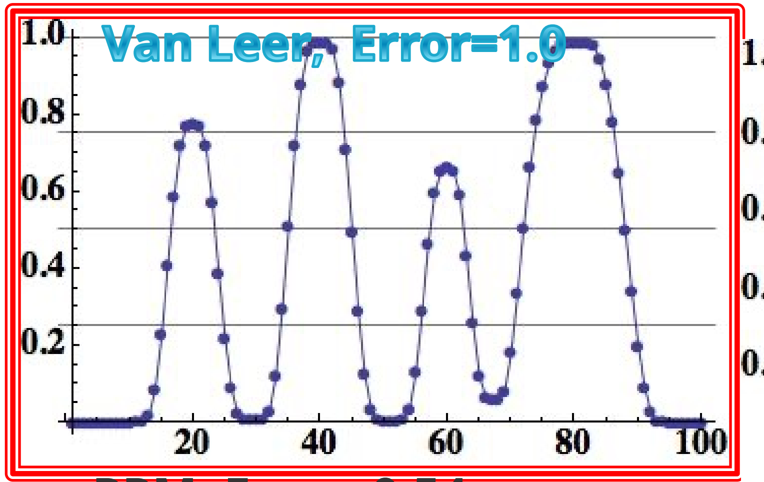
A parabolic reconstruction is much better and the method is very compact, i.e., great for modern architectures.

$$u_j = \int_{j-1/2}^{j+1/2} P(\theta) d\theta; u_{j-1/2} = P(-1/2); u_{j+1/2} = P(1/2)$$

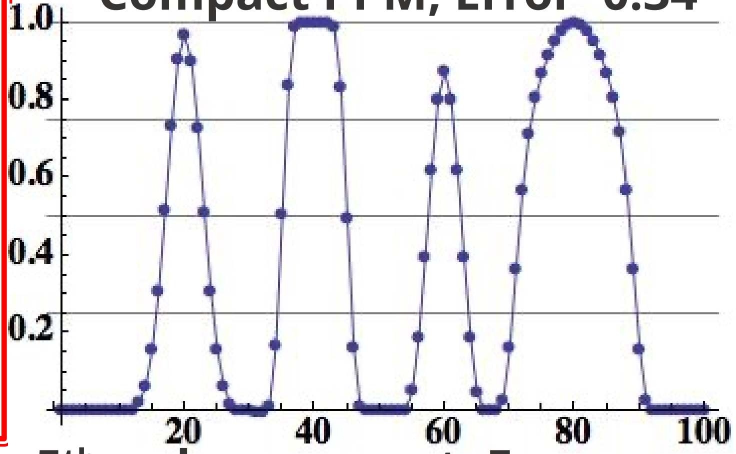


Integrated Error = 0.080, ~6 times better!

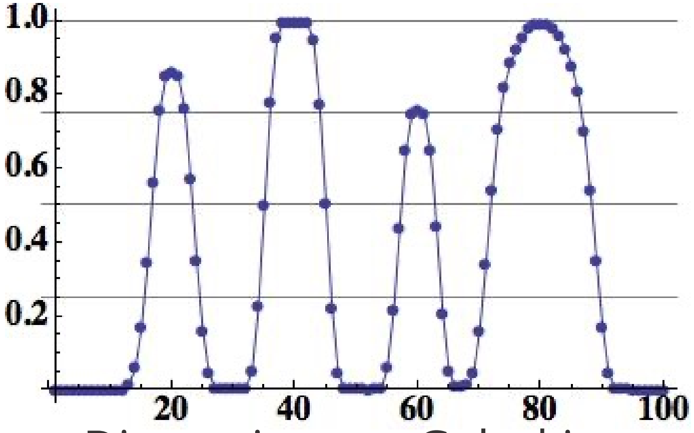
This is a common wave form to test methods used for shocks. It has a Gaussian, square, triangle, and sin squared. The square wave means it is limited to first-order accuracy or less



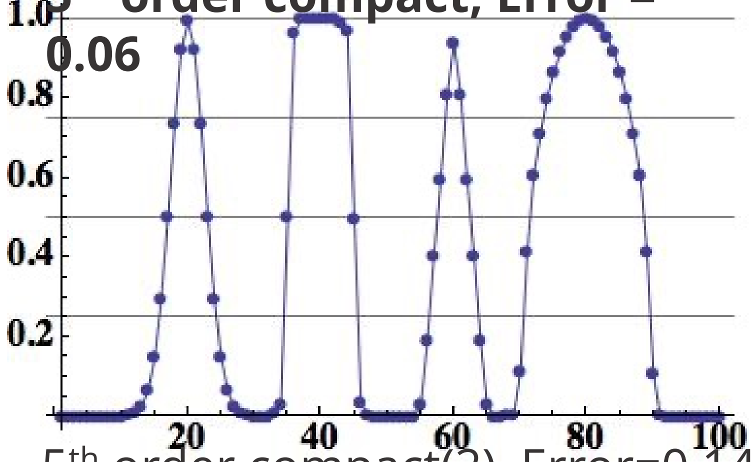
Compact PPM, Error=0.34



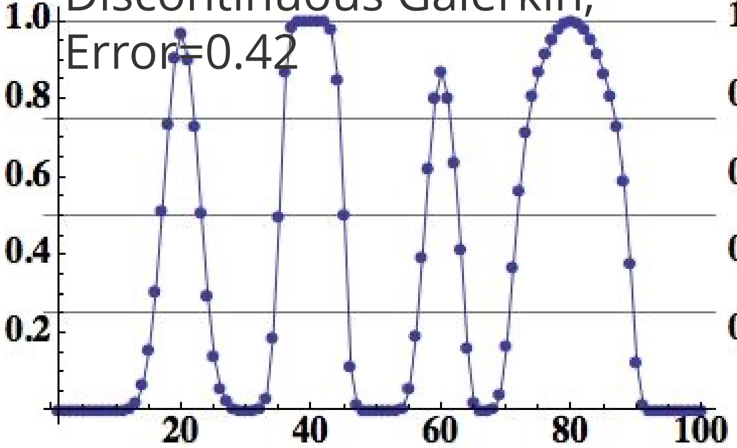
PPM, Error=0.54



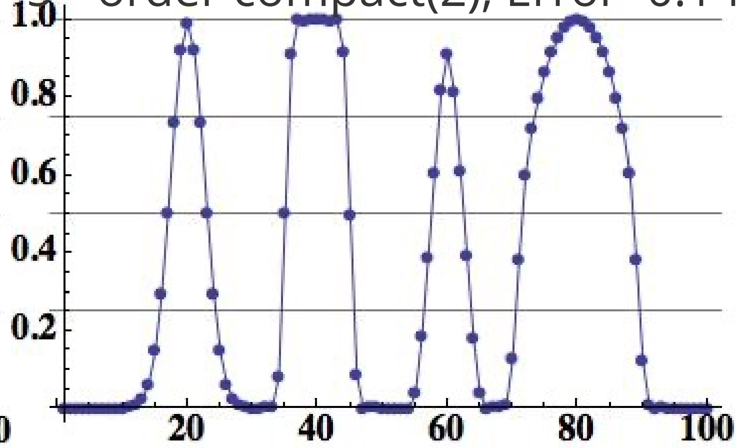
5th order compact, Error = 0.06



Discontinuous Galerkin, Error=0.42



5th order compact(2), Error=0.14



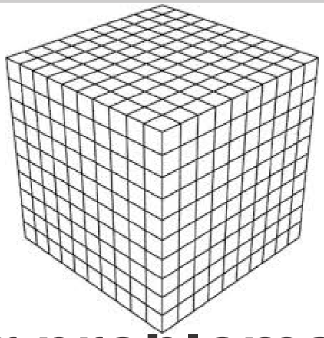
Methods and algorithms will improve the performance of a code on every single platform: laptop, to desktop, to cluster to capacity machine to capability machine

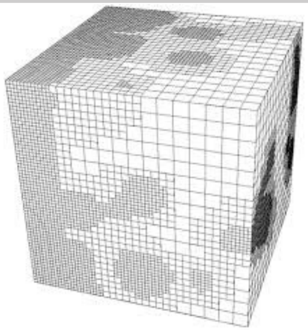
Accuracy can produce the same quality answer with much less effort

For shock-hydro problems we can estimate the impact

First-order accuracy (convergence), for problems containing shocks, and if cost is *effectively* equal.

Error Ratio	3-D N^4	AMR N^3
1	1	1
1/2 (PPM)	16x	8x
1/3 (CPPM)	81x	27x
1/6 (5^{th} O)	1296x	216x
1/16	65,536x	4096x





The advantage is larger for other problems, where high-order methods have a slightly higher rate of convergence

Let's look at the presentation of shock problems in detail.

From Sod's classical 1978 paper (*J. Comp. Phys*. **27**)
(i.e., where Sod's problem comes from) *"Hello World"* for Shocks

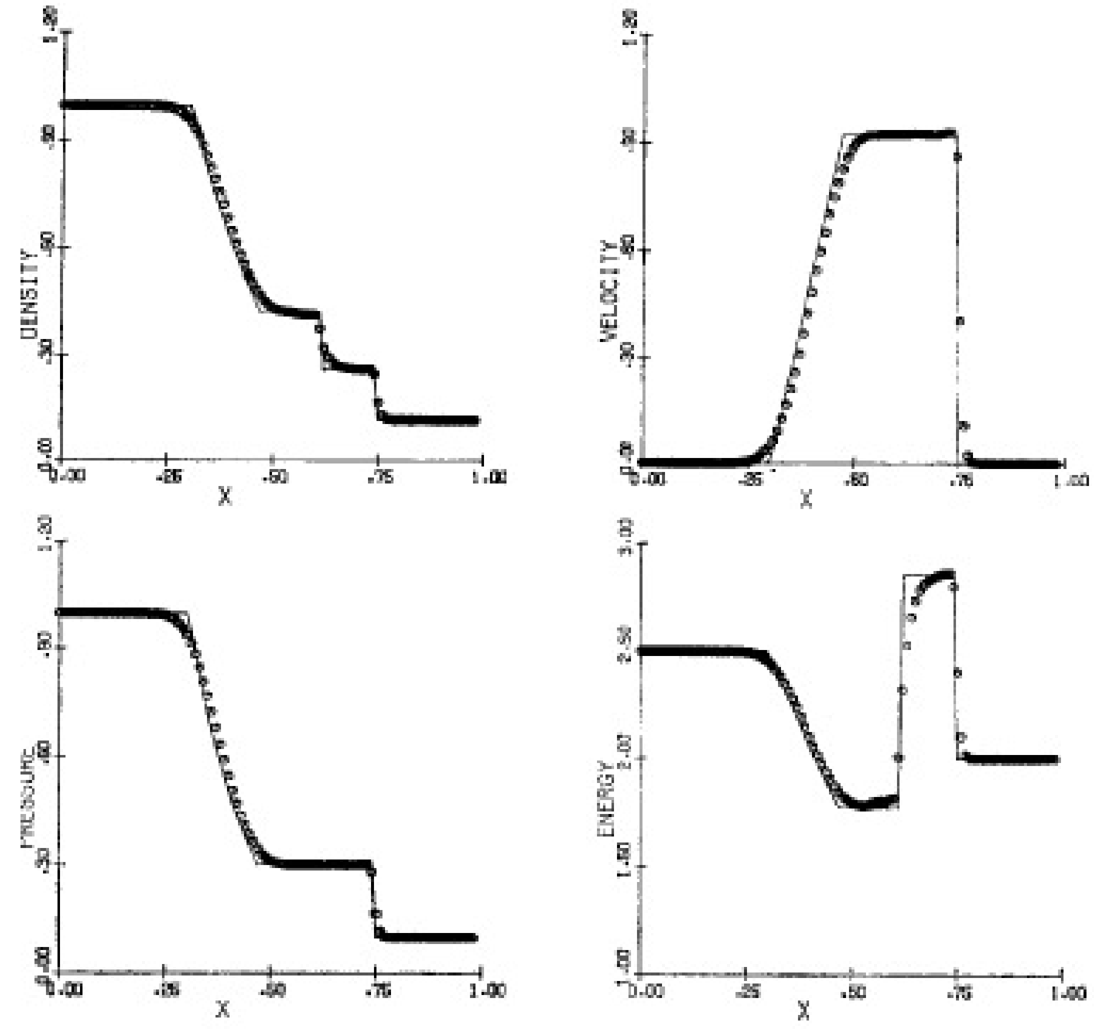


FIG. 14. Hybrid method with *ACM*.

No error or convergence rates discussed anywhere in the paper. Run time on a computer is given.

Move forward to Harten's paper introducing TVD methods

From Harten's classical 1983 paper (*J. Comp. Phys*. **49**)
(i.e., where TVD methods are introduced)

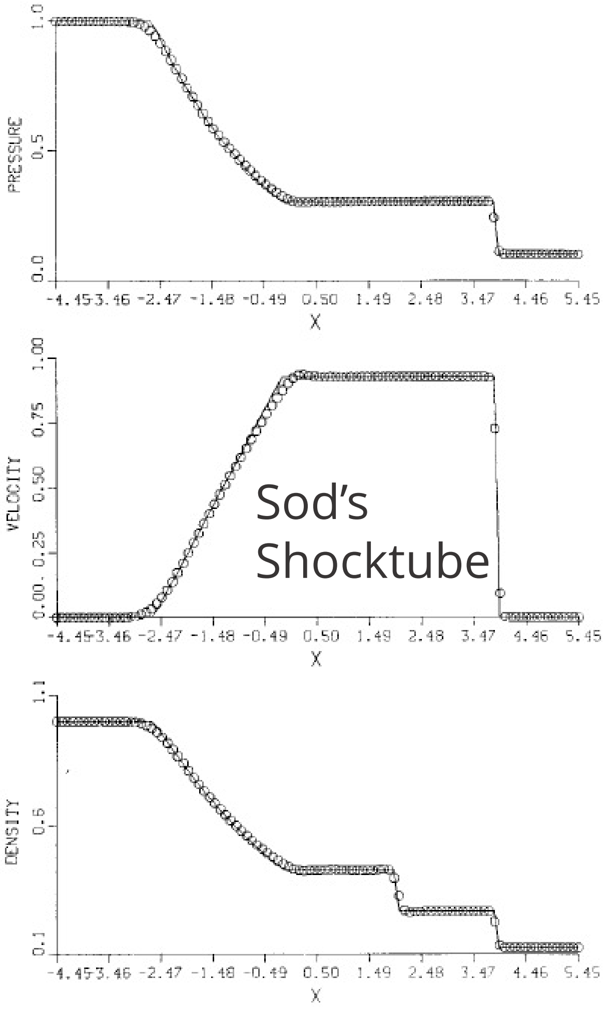
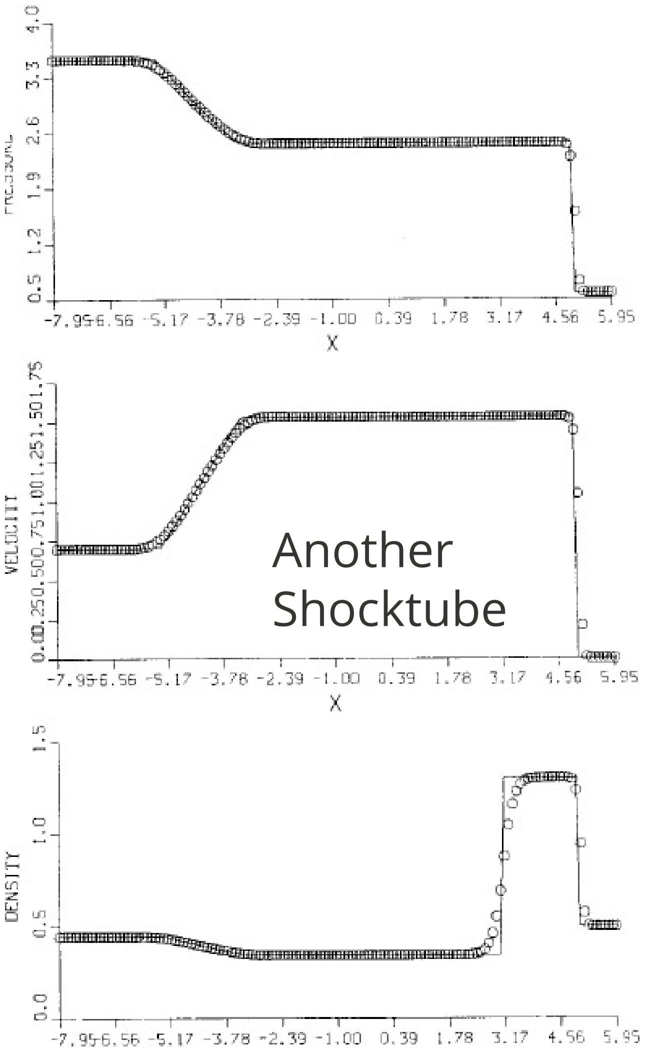


FIG. 3. ULT1C scheme for (7.8).

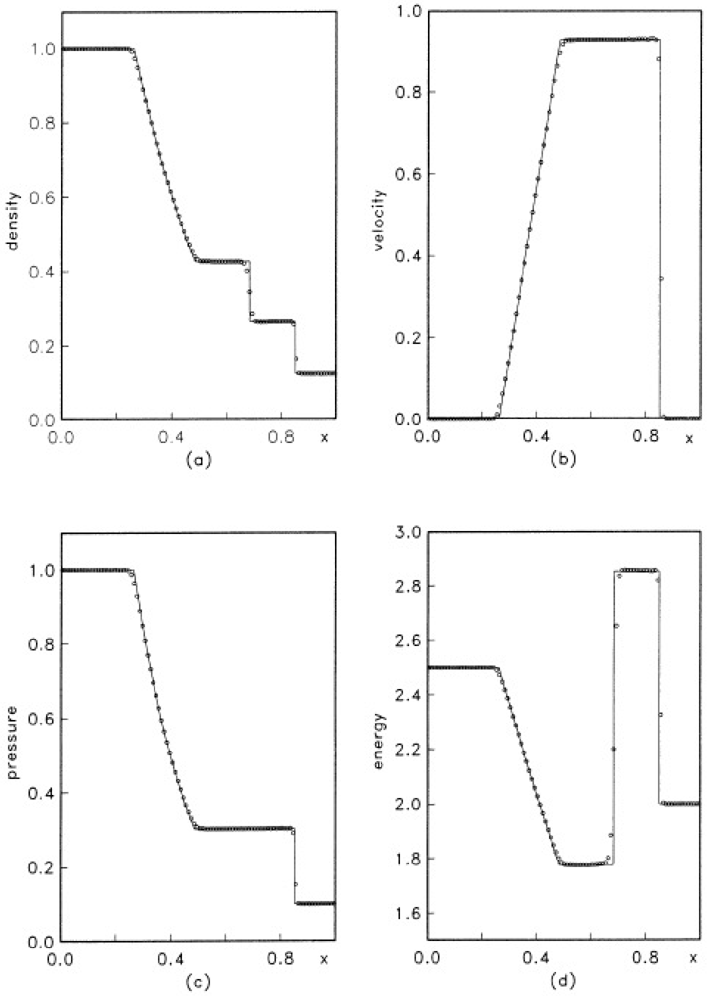


No error or convergence rates discussed anywhere in the paper.

Run times are given.

Move forward another decade to Huynh's excellent paper in SIAM J. Num. Anal.

From Huynh's 1995 paper (*SIAM J Num. Anal.* **49**)
(i.e., where a fantastic overview of methods is provided)

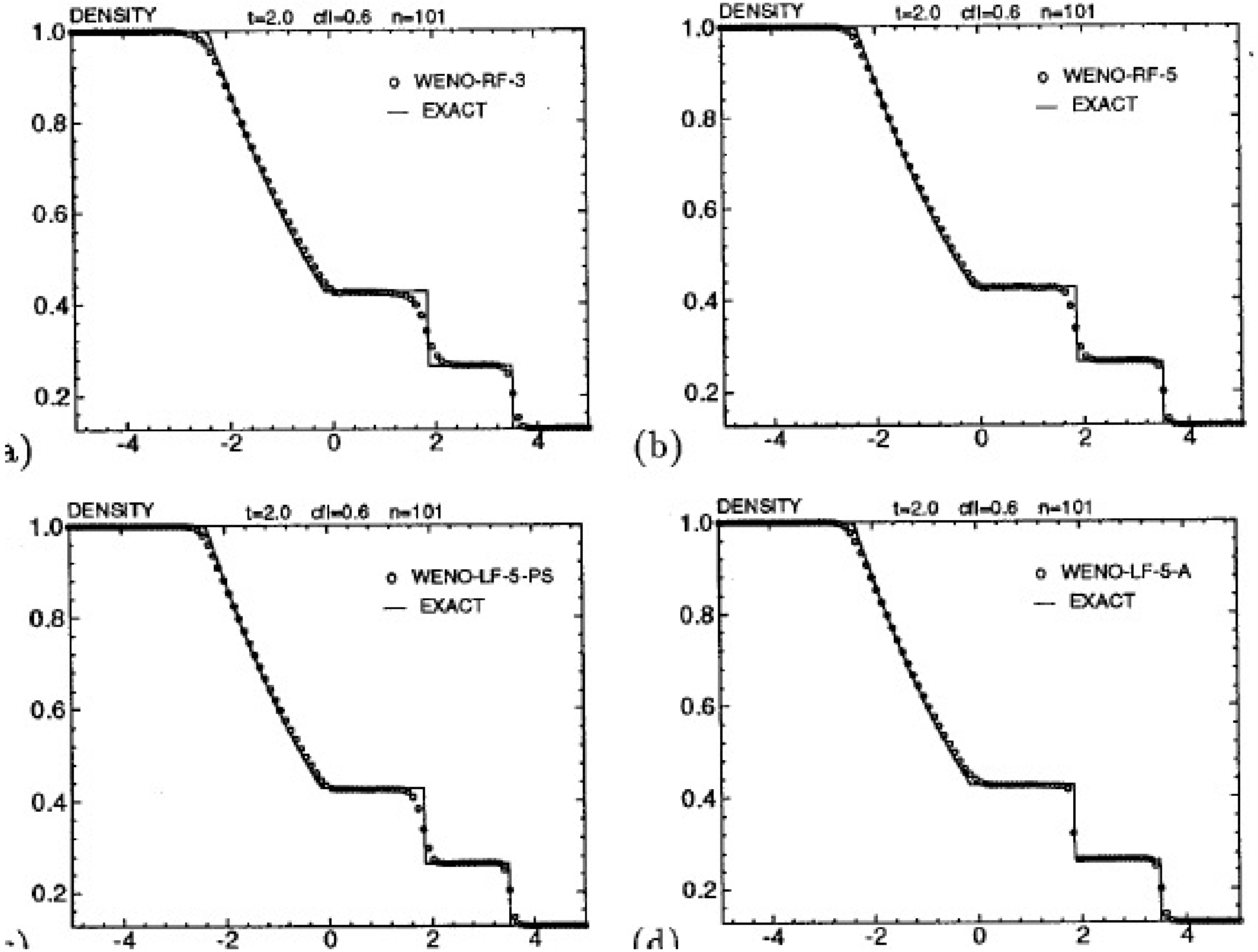


No error or convergence rates discussed anywhere in the paper.

Run times are given.

Staying in this era, but returning to J. Comp. Phys.

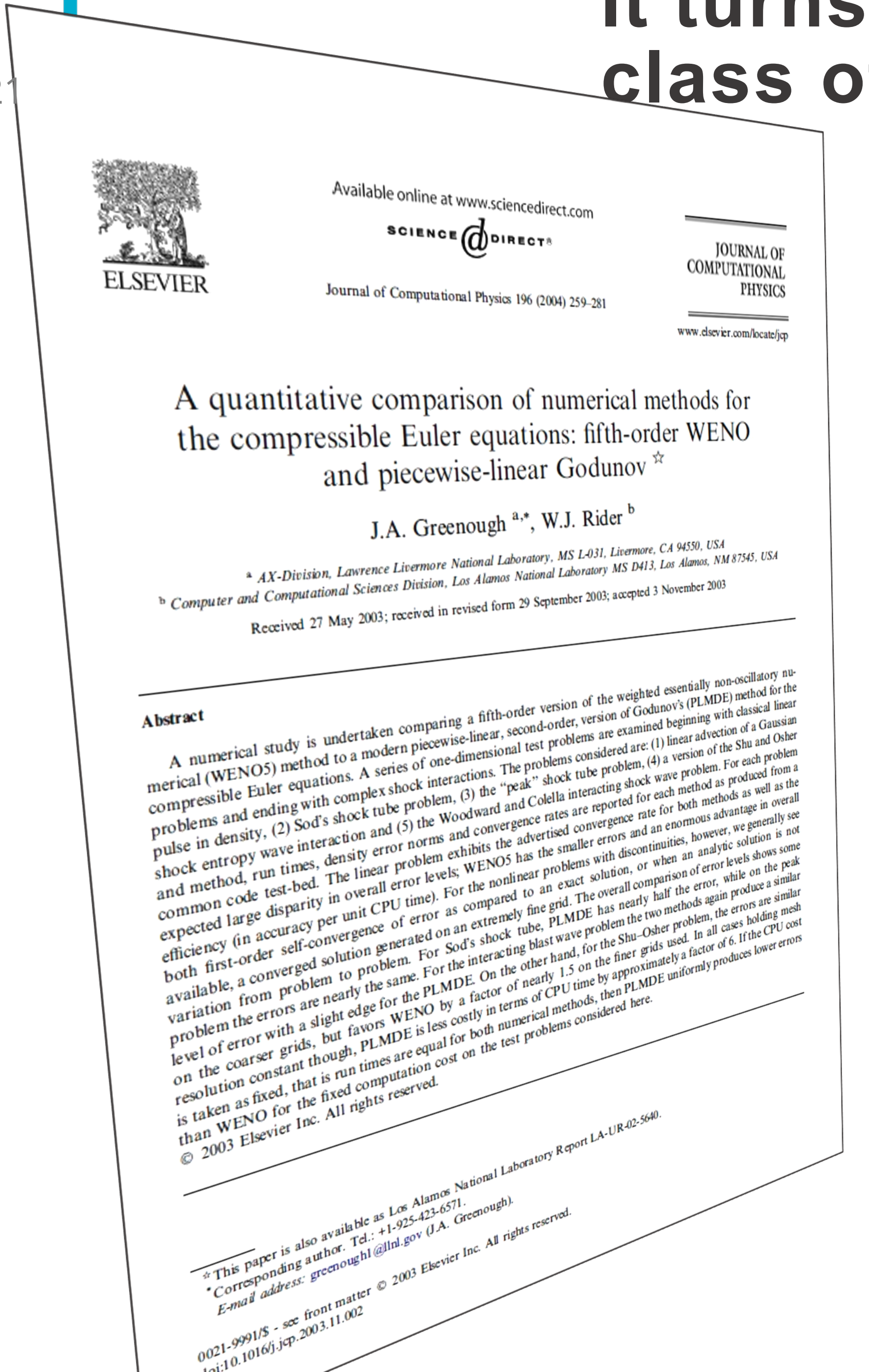
From Jiang and Shu's WENO paper (*J. Comp. Phys*. **126** - introduced 5th order WENO)



No error or convergence rates discussed anywhere in the paper.

Run times are given.

It turns out that there is much more to this class of problems as I documented.


Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
ELSEVIER
Journal of Computational Physics 196 (2004) 259–281
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A quantitative comparison of numerical methods for the compressible Euler equations: fifth-order WENO and piecewise-linear Godunov ☆

J.A. Greenough [a,*], W.J. Rider [b]
[a] AX-Division, Lawrence Livermore National Laboratory, MS L-031, Livermore, CA 94550, USA
[b] Computer and Computational Sciences Division, Los Alamos National Laboratory MS D413, Los Alamos, NM 87545, USA

Received 27 May 2003; received in revised form 29 September 2003; accepted 3 November 2003

Abstract

A numerical study is undertaken comparing a fifth-order version of the weighted essentially non-oscillatory numerical (WENO5) method to a modern piecewise-linear, second-order, version of Godunov's (PLMDE) method for the compressible Euler equations. A series of one-dimensional test problems are examined beginning with classical linear problems and ending with complex shock interactions. The problems considered are: (1) linear advection of a Gaussian pulse in density, (2) Sod's shock tube problem, (3) the "peak" shock tube problem, (4) a version of the Shu and Osher shock entropy wave interaction and (5) the Woodward and Colella interacting shock wave problem. For each problem and method, run times, density error norms and convergence rates are reported for each method as produced from a common code test-bed. The linear problem exhibits the advertised convergence rate for both methods as well as the expected large disparity in overall error levels; WENO5 has the smaller errors and an enormous advantage in overall efficiency (in accuracy per unit CPU time). For the nonlinear problems with discontinuities, however, we generally see both first-order self-convergence of error as compared to an exact solution, or when an analytic solution is not available, a converged solution generated on an extremely fine grid. The overall comparison of error levels shows some variation from problem to problem. For Sod's shock tube, PLMDE has nearly half the error, while on the peak problem the errors are nearly the same. For the interacting blast wave problem the two methods again produce a similar level of error with a slight edge for the PLMDE. On the other hand, for the Shu–Osher problem, the errors are similar on the coarser grids, but favors WENO by a factor of nearly 1.5 on the finer grids used. In all cases holding mesh resolution constant though, PLMDE is less costly in terms of CPU time by approximately a factor of 6. If the CPU cost is taken as fixed, that is run times are equal for both numerical methods, then PLMDE uniformly produces lower errors than WENO for the fixed computation cost on the test problems considered here.

☆ This paper is also available as Los Alamos National Laboratory Report LA-UR-02-5640.
* Corresponding author. Tel.: +1-925-423-6571.
E-mail address: greenough1@llnl.gov (J.A. Greenough).

0021-9991/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2003.11.002

WENO5 (5^{th} order accurate) is much more efficient for linear problems

PLMDE (2nd order accurate) is more efficient than WENO5 on all nonlinear problems (with discontinuities)

The advantage is unambiguous for Sod's shock tube and the Interacting Blast Waves

The advantage is less clear-cut for the "peak" problem

At a given mesh spacing WENO5 gives better answers for the Shu-Osher problem, but worse than PLMDE at fixed computational expense

Greenough & Rider, *J. Comp. Phys.*, **194**, 2004 (GR2004).

Greenough & Rider (2004) provided quantitative errors for these problems.

Table 5

E_{L_1} and E_{L_∞} errors and convergence rates for PLMDE on Sod's shock tube at different grid resolutions

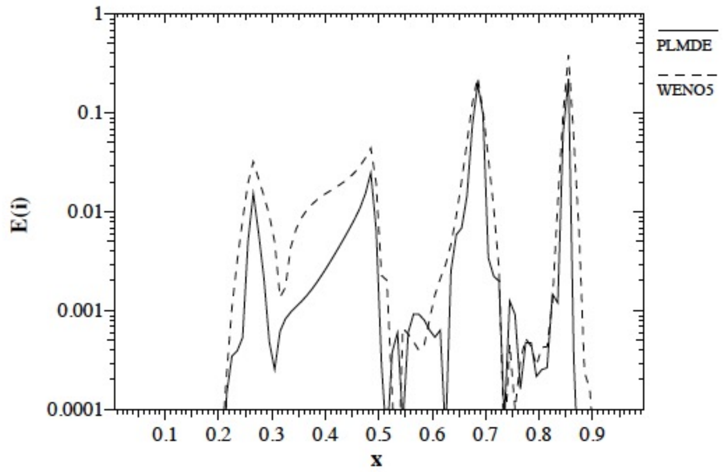
N	E_{L_1}	L_1 rate	E_{L_∞}	L_∞ rate
100	8.22e − 03	–	0.22e − 00	–
200	4.48e − 03	0.88	0.25e − 00	−0.20
400	2.62e − 03	0.77	0.33e − 00	−0.37

Table 6

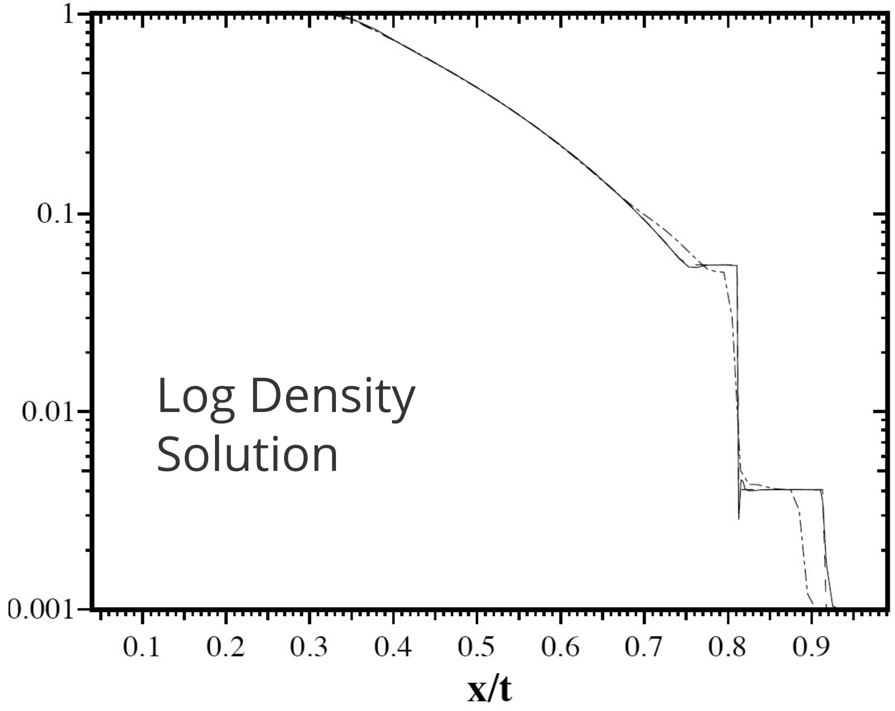
E_{L_1} and E_{L_∞} errors and convergence rates for the WENO method for Sod's shock tube at different grid resolutions

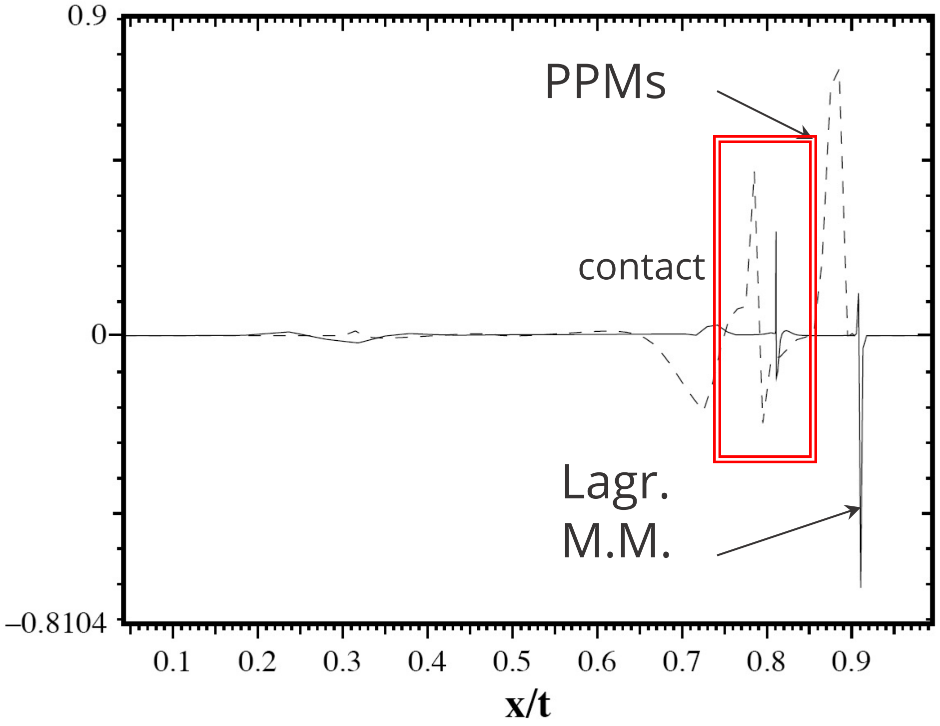
N	E_{L_1}	L_1 rate	E_{L_∞}	L_∞ rate
100	1.58e − 02	–	0.37e − 00	–
200	8.24e − 03	0.93	0.40e − 00	−0.01
400	4.47e − 03	0.88	0.46e − 00	−0.18

- We plotted the errors as a function of position too. WENO is worse than PLMDE almost everywhere, but for a much greater computational expense of six times greater.
- Asymptotic rates of 4/5 and 2/3 respectively
- For this problem theoretically WENO would win at a ridiculous resolution ~2,000,000 cells



LeBlanc's shock tube - Density Error
Computed with 100 cells is dominated by error in the shock location.
This problem used to be a survival test for codes, but now is a verification problem.

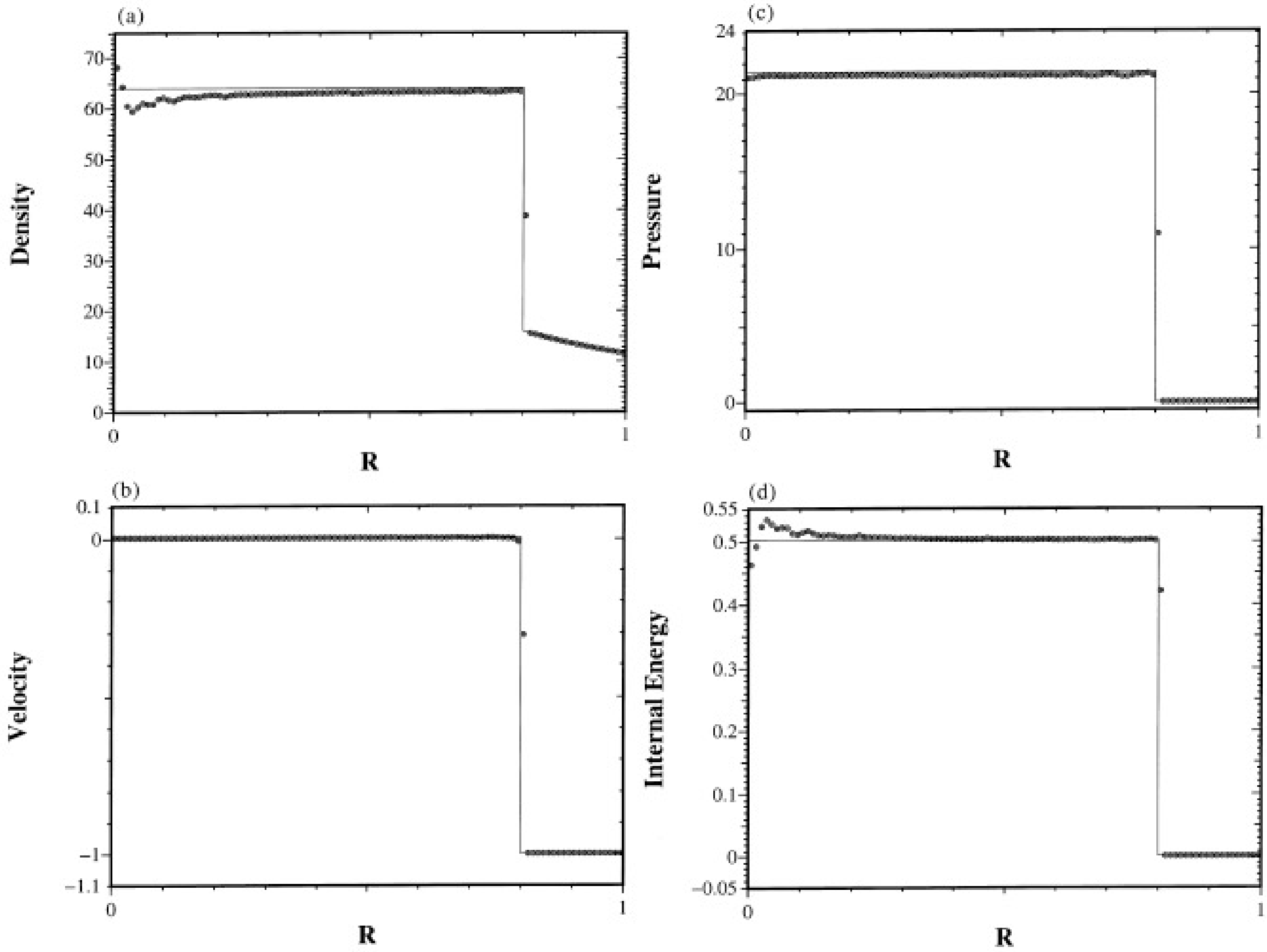




The Lagrangian solution leads the exact solution, whereas the PPMs solution lags the exact.

We can look at Noh's solution in detail.

Noh's problem used to be a survival problem for codes



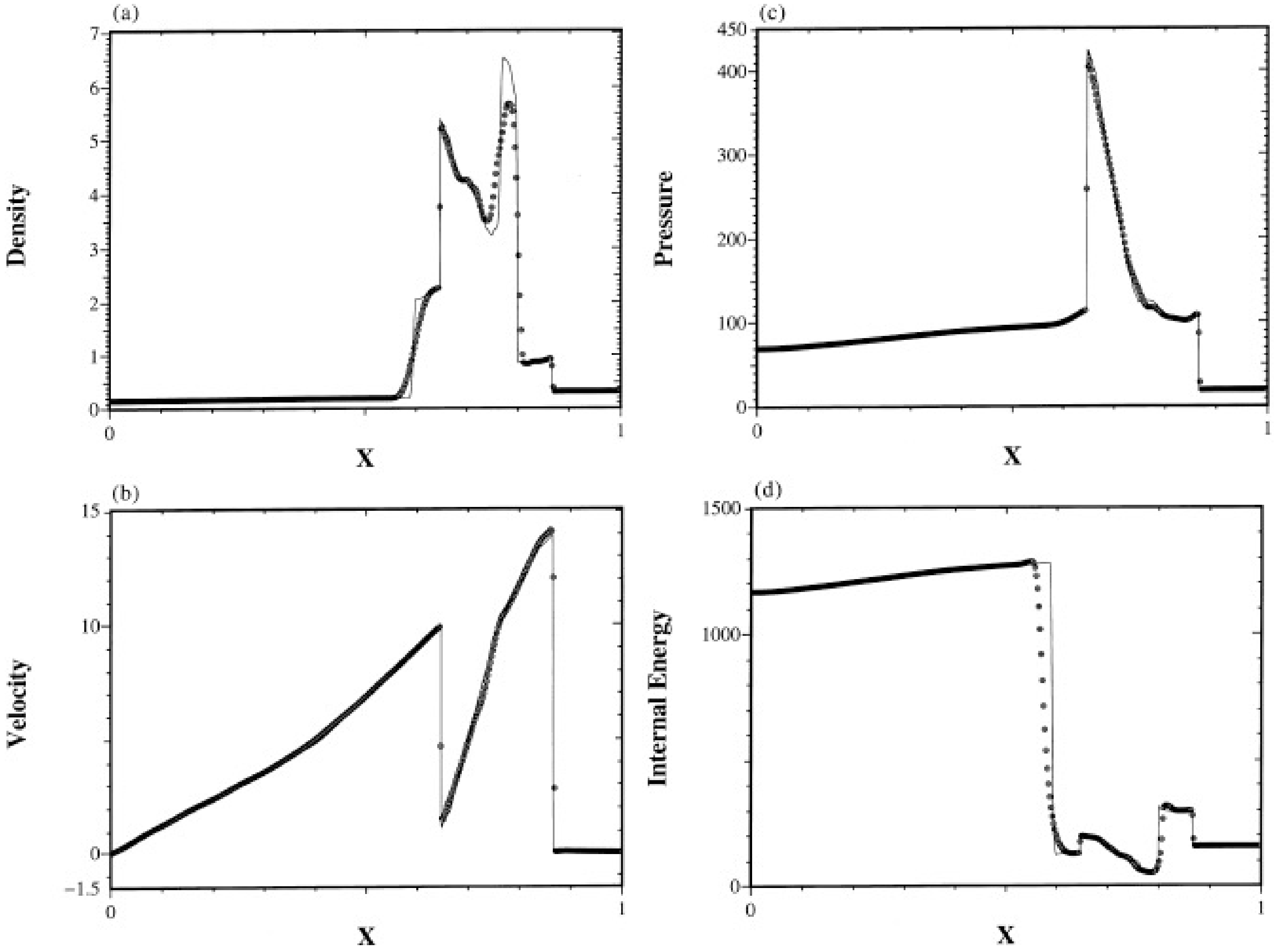
Many problems do not have an analytical solution, but are useful nonetheless.

One of the best of these problems is Woodward and Colella's interacting blast wave problem.

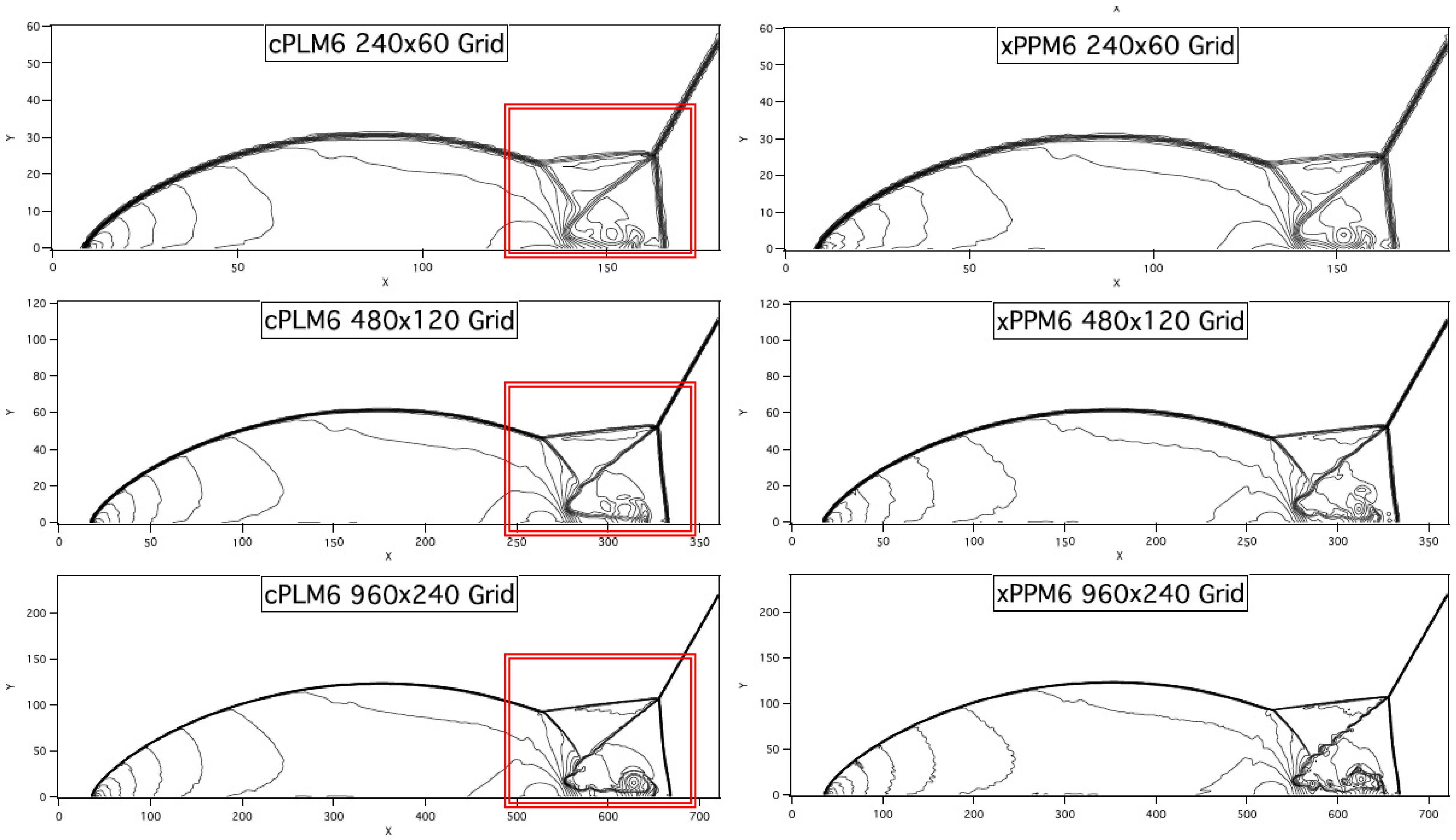
$$\begin{pmatrix} \rho \\ u \\ p \\ x \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1000 \\ 0,0.1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0.01 \\ 0.1,0.9 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 100 \\ 0.9,1 \end{pmatrix}, \gamma = 1.4$$

Despite being non-analytic it is one of the most discerning of common test problems.

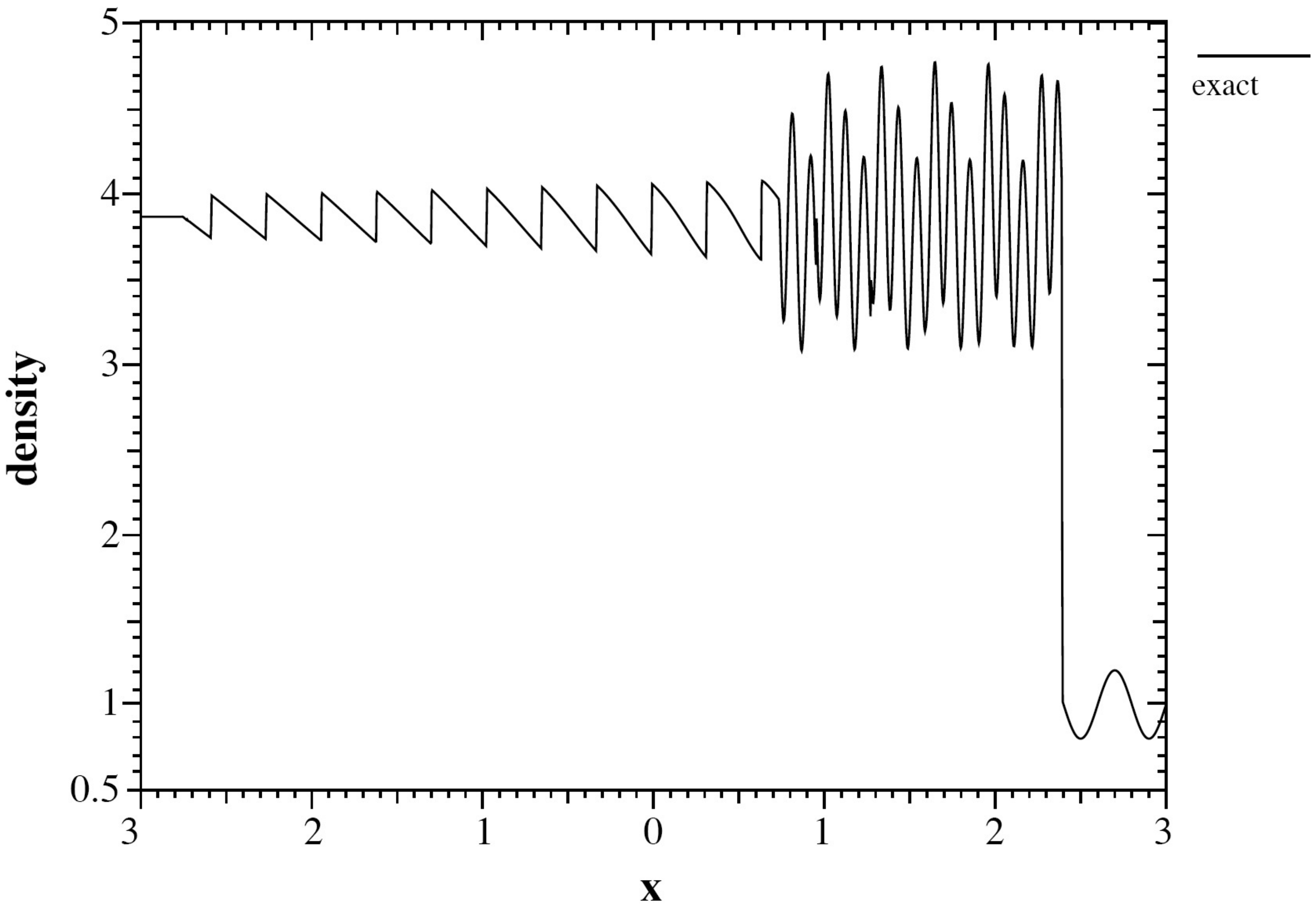
The W-C Blast Wave Problem in detail
The issue is whether one can get the solution to interacting shock waves correct.



The Mach reflection has had a long, fruitful life. It is a qualitative test of methods and behavior under mesh refinement

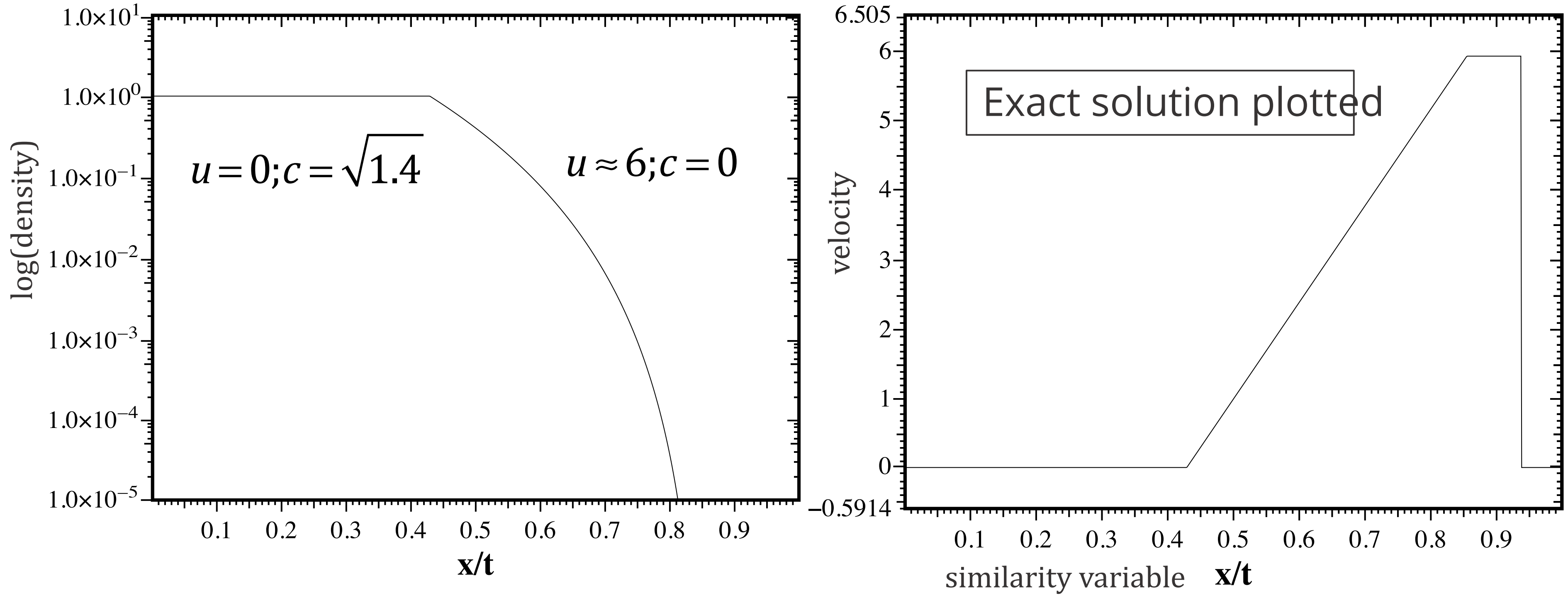


There is also the Shu-Osher entropy wave problem - idealized shock-turbulence



Void expansion results have asked more fundamental questions of the code. Current methods cannot solve this problem properly.

Stability issues associated with the expansion problem highlight a fundamental oversight



The characteristic speed at the free surface is 5 times larger than the initial conditions arising from a change in pressure (Δp)

Summary

- The null hypothesis is a useful way to conduct science and V&V.
- The gist is to provide an adversarial viewpoint that tests a code or model more aggressively. If the code/model can stand up to this aggressive testing, the confidence of correctness is much higher.
- I showed the way this looks for code verification of a shock physics codes. The analysis couples theory to deep quantitative analysis applicable to applications.
- **Bottom Line: Assuming the code is wrong and failing to find evidence is stronger evidence that the code is correct. It is a better way to do verification (and validation).**