Toward petascale shock/turbulence computations

Johan Larsson, Eric Johnsen, Sanjiva K. Lele, and Parviz Moin
Center for Turbulence Research, Stanford University, Stanford, CA 94305
E-mail: jola@stanford.edu (J. Larsson)

Abstract. Progress toward highly resolved simulations of shock/turbulence interaction is described, with a focus on the scientific reasons for studying this problem as well as the computational issues involved. We illustrate the importance of computing shock/turbulence interaction at Reynolds numbers of at least one order of magnitude larger than previously done. We also illustrate the need to study the effect of turbulent Mach number on the interaction process. Scaling data for the code is used to estimate the computational resources needed for these novel simulations on petascale systems.

1. Introduction
Most fluid flows of scientific and engineering interest involve and are strongly affected by turbulence. For example, turbulence has a great effect on the process of mixing, which in turn has a major effect on flow separation, combustion, and other phenomena. When interacting with a shock wave, the state of turbulence changes dramatically. Clearly, then, the interaction of turbulence with a shock wave is a key feature in many high-speed engineering systems. The most fundamental problem in this area is that of isotropic turbulence passing through a nominally normal shock. This problem has received considerable attention in past theoretical [1], experimental [2], and computational (e.g., [3, 4]) studies. The linear theory by Ribner [1] predicts that the velocity and vorticity variances are amplified during the shock interaction and that there is an inviscid adjustment region behind the shock where turbulence adjusts to its postshock state. This theory is formally valid in the limits of infinite Reynolds number and zero turbulent Mach number, but has thus far [3, 4] been tested only at Reynolds numbers of $Re_\lambda \approx 20$. At these low Reynolds numbers, the viscous decay behind the shock is comparable to or larger than the inviscid adjustment, which makes direct comparisons to the linear theory difficult. Furthermore, the spectrum of turbulence is not truly broadband at such Reynolds numbers. The ultimate objective of the present study is to use direct numerical simulation (DNS) to produce a set of highly resolved databases at much higher Reynolds number, to be used in examination of the validity of the linear theory as well as in a priori and a posteriori studies of how to model shock/turbulence interaction in the context of large eddy simulation (LES) and engineering turbulence modeling.

2. Numerical method
The Hybrid code (see [5, 6]) solves the compressible Navier-Stokes equations for a perfect gas using solution-adaptive finite-difference schemes on Cartesian (but stretched) grids. Near shock waves or other discontinuities, a fifth-order accurate weighted essentially non-oscillatory (WENO) scheme is used [7]. In these regions, the equations are discretized in divergence (or
conservative) form, thereby ensuring convergence to the correct weak solution. Away from shock waves, a sixth-order accurate central difference scheme is used. This scheme has nominally zero numerical dissipation, which is an important attribute for the prediction of broadband turbulence (see [5]). To aid the nonlinear stability, the convective terms are discretized in the split form

$$\frac{\partial pf}{\partial x_j} \rightarrow \frac{1}{2} \left( \frac{\partial pf}{\partial x_j} + f \frac{\partial \rho}{\partial x_j} + \rho \frac{\partial f}{\partial x_j} \right)$$

for \( f = u_j, \rho u_i, \rho h_0 \) in the conservation equations for mass, momentum, and total energy, respectively. Here \( \rho \) is the density, \( u_i \) is the velocity vector, \( h_0 = e_0 + p/\rho \) is the total specific enthalpy, and \( e_0 \) is the total specific energy. The coupling of different discretizations globally conserves mass, momentum, and energy and was shown in [8] to be linearly stable. The semi-discrete system is integrated in time by using classic fourth-order, fully explicit Runge-Kutta. The method is implemented by using C++ and is parallelized through domain decomposition, with communication handled by MPI. Being fully explicit, the method is expected to scale well. However, one should note that the solution-adaptive “hybrid” nature of the algorithm has implications for the computational cost. The WENO scheme is 5–10 times more expensive than the central difference scheme but is used only in the vicinity of shocks. Since shocks are two-dimensional surfaces, the number of grid points treated by WENO scales as \( N_d^2 \), where \( N_d \) is the number of points in each direction. Thus the fraction of points treated by WENO becomes ever smaller under grid refinement, which raises the possibility of “superlinear” speedup under grid refinement. However, it also raises the possibility of load-balancing problems because (with domain decomposition) a small subset of processors is likely to handle the more expensive WENO regions.

The code has been run on several machines. Here we present scaling data from the Franklin machine at the National Energy Researcher Scientific Computing Center. A first test of the scalability is the weak scaling of the nearly incompressible Taylor-Green vortex in a periodic box. This problem does not generate shocks, and thus the WENO scheme is never used, implying a (theoretically) perfect load balance. Each full time step is timed in the code, as well as the cumulative time spent in communication (including waiting). The results are shown in figure 1 for \( n = 1, 2, 4, \ldots, 4096 \), where \( n \) is the number of processors. We note that the code is algorithmically different for \( n = 1 \); for example, the periodic boundary data is copied directly rather than through temporary communication arrays. For this reason the scaling efficiency is defined with respect to \( n = 2 \) as

$$\eta_{weak}(n, N) = \frac{t(2, 2 * N)}{t(n, n * N)},$$

where \( t(n, N) \) is the CPU time for one step on \( n \) processors on a grid with \( N \) points. The efficiency is above 86% and 90% up to 4,096 processors with \( N = 32^3 = 32,768 \) and \( N = 64^3 = 262,144 \) points per processor, respectively. Representing a factor of 2000 increase in \( n \), these efficiencies must be considered satisfactory.

3. Canonical shock/turbulence interaction

The most fundamental shock/turbulence interaction problem is that of isotropic turbulence passing through a nominally normal shock wave (or, equivalently, a normal shock passing through isotropic turbulence). The essence of the problem is shown in figure 2. The incoming turbulence is isotropic, as evidenced by the random orientation of the vortex cores. The shock compresses the turbulence in the \( x \) direction, increasing the vorticity and making the postshock turbulence axisymmetric with vortex cores predominantly oriented in the \( y - z \) plane.
Figure 1. Weak scaling with perfect load balancing, with base grid $N = 32^3$ (thin, lower curves in both plots) and $N = 64^3$ (thick, upper curves).

Figure 2. Snapshot of shock/turbulence interaction at $M = 2$, $M_t = 0.15$, and $Re\lambda \approx 60$. The flow is from left to right, with the shock visualized by transparent isosurfaces of compression. Vortex cores are visualized by isosurfaces of the second invariant of the velocity gradient tensor, colored by the vorticity magnitude.

In order to minimize errors around the shock, a stretched grid is used in the streamwise direction with finer mesh spacing in the vicinity of the shock. The inflow turbulence is taken from a precomputed database of isotropic turbulence. At the outflow, a sponge region is used to gently drive the flow toward a laminar state. We note that the simulations presented here are underresolved and should be characterized as large eddy simulations rather than direct numerical simulations.
The variances of velocity and vorticity fluctuations are shown in figure 3 for a sequence of cases. The transverse vorticity is directly amplified at the shock as a result of the compression, while the streamwise vorticity is initially unchanged. Behind the shock, the out-of-equilibrium turbulence adjusts toward an isotropic state, although the Reynolds stresses never reach isotropy in these runs. One outstanding question in shock/turbulence interaction is whether the turbulence truly returns to isotropy. The low Reynolds number of the present (and all previous) runs implies a quick viscous decay behind the shock, which overwhelms any return to isotropy. Large-scale calculations at higher Reynolds numbers are needed to illuminate this issue.

At high enough turbulent Mach number $M_t$, the turbulent pressure fluctuations become comparable to the pressure jump associated with the shock, which significantly alters the instantaneous shock profile. Some instantaneous profiles of the compression (negative dilatation) and density along the $x$ axis are shown in figure 4 for such a case. The instantaneous structure of the shock varies wildly, from being twice as strong as average, to being replaced by a smooth compression wave, to being replaced by two weaker shocks. This intermittency is largely absent at lower values of $M_t$ and represents a second outstanding fundamental question in shock/turbulence interaction.

### 3.1. Scalability with imperfect load balancing

The solution-adaptive algorithm uses the WENO scheme only around the shock, which, together with the local sponge region, implies a load imbalance. While the instantaneous extent of the WENO region is solution-dependent (and unknown), one can predict the average extent based on prior coarse-grid runs.

An example of the load imbalance is shown in figure 5(a). While the code timings should be averaged over longer times, clearly both the WENO scheme (processors 48-63) and the sponge region (processors 224-255) still induce higher costs. A simple ad hoc fix is to factor in the higher costs of the WENO scheme and the sponge region, along with the known locations of these regions, in the domain decomposition. As shown in figure 5(a), this leads to a better load balance and a small reduction in cost. We point out that the algorithm consists of several steps,
only one of which (the most expensive) incurs the higher cost of WENO and the sponge. Hence the ad hoc decomposition induces load imbalance in the remaining algorithmic steps, which explains why all processors spend significant time waiting for communication.

The weak scaling efficiency is above 92% for both $32^3$ and $64^3$ points per processor, going from $n = 4$ to $n = 2048$.

**3.2. Cost estimates at higher Reynolds number**

Well-resolved direct numerical simulation (DNS) requires adequate resolution of the viscous dissipation, which in isotropic turbulence amounts to having the maximum wavenumber $k_{\text{max}}$ =
Table 1. Estimates of total CPU time and Reynolds number upstream of the shock for a well-resolved DNS of $(M, M_t) = (2.0, 0.15)$ shock/turbulence interaction using $k_0 = 4$. $n_{1\text{Gb}}$ and $n_{24\text{h}}$ are the number of processors for 1 GB of memory and 24-hour turnaround, respectively.

| $N_y$ | $Re_\lambda$ | CPU-hours | $n_{1\text{Gb}}$ | $n_{24\text{h}}$ |
|-------|---------------|-----------|------------------|-----------------|
| 256   | 24            | $2.6 \times 10^4$ | 32               | 110             |
| 384   | 54            | $1.4 \times 10^4$ | 110              | 570             |
| 512   | 96            | $4.3 \times 10^4$ | 250              | 1800            |
| 768   | 220           | $2.2 \times 10^5$ | 850              | 9100            |
| 1024  | 390           | $6.9 \times 10^5$ | 2000             | 29000           |

\[ \pi N_y/L_y \gtrsim 1.5/\eta_k, \] where $N_y$ is the number of points in the transverse directions, $L_y$ is the transverse domain size, and $\eta_k$ is the Kolmogorov length scale. Consider turbulence at $M = 2$, $M_t = 0.15$, and some $Re_\lambda$ immediately upstream of the shock. The shock compression modifies the Kolmogorov scale by about a factor of $(\rho_1/\rho_2)^{11/8}(p_2/p_1)^{3/8} \leq 1$, and hence the critical resolution is downstream of the shock. The upstream isotropic turbulence approximately satisfies $\eta_k \approx 0.5/\sqrt{Re_\lambda}$, where $\lambda$ is the Taylor length scale. We need a relation between the wavenumber of peak energy $k_0$ and $\lambda$ upstream of the shock. This is $Re_\lambda$-dependent, but a very crude estimate is to use the present preliminary data at $Re_\lambda \approx 60$, which gives $\lambda \sim L_y/(2\pi k_0)$. We then get, for a given transverse resolution $N_y$,

\[ Re_\lambda \lesssim 0.006(N_y/k_0)^2 \]

immediately upstream of the shock. This is a more stringent criterion than many previous studies but is necessary to adequately resolve the postshock dissipative motions.

Now consider the code performance. Using the data in figure 5, the code needs $1.6 \times 10^{-5}$ seconds per grid point and step on 256 processors or more. Furthermore, the code uses about 1 GB of memory for $1.6 \times 10^6$ points per processor. Requirements on the domain size and the shock-normal grid spacing imply $3N_y$ points in the streamwise direction, and the preliminary runs needed about $47N_y$ time steps for statistical convergence. The final estimates, listed in table 3.2, make for sobering reading.

4. Summary and future work

The progress toward large-scale calculations of the canonical shock/turbulence interaction discussed here is part of a larger effort to developing improved computational methods for complex shock/turbulence interaction problems. The canonical shock/turbulence problem described here serves two purposes in this larger effort. First, it is a useful problem for computational efficiency assessment and improvement because it includes all the relevant algorithmic complexities. The Hybrid code has been shown to scale well up to 4096 processors, which is promising for future production runs. Work on profiling the code to identify single-processor bottlenecks is under way in collaboration with the SciDAC Performance Engineering Research Institute. Second, the shock/turbulence calculations will produce large databases at different combinations of turbulent and mean flow Mach numbers, at higher Reynolds numbers than previously possible. These databases will be of use both in theoretical developments, for example, for assessment of the validity of Ribner’s linear theory at higher levels of compressibility, and in model development, for example, for a priori studies or a posteriori comparisons in subgrid scale modeling.
Acknowledgments
Financial support has been provided by the DOE Scientific Discovery through Advanced Computing (SciDAC) program, with additional support from the Natural Sciences and Engineering Research Council of Canada.

References
[1] Ribner H S 1953 Convection of a pattern of vorticity through a shock wave NACA Report 1164
[2] Agui J H, Briassulis G and Andreopoulos Y 2005 J. Fluid Mech. 524 143–195
[3] Lee S, Lele S K and Moin P 1993 J. Fluid Mech. 251 533–562
[4] Lee S, Lele S K and Moin P 1997 J. Fluid Mech. 340 225–247
[5] Larsson J, Lele S K and Moin P 2007 Annual Research Briefs (Center for Turbulence Research) pp 47–57
[6] Larsson J, Cook A, Lele S K, Moin P, Cabot W, Sjögren B, Yee H and Zhong X 2007 J. Phys.: Conf. Ser. 78 012014
[7] Jiang G S and Shu C W 1996 J. Comput. Phys. 126 202–228
[8] Larsson J and Gustafsson B 2008 J. Comput. Phys. 227 2886–2898