Abstract for an Invited Paper
for the DFD07 Meeting of
The American Physical Society

Analysis of Turbulence Datasets using a Database Cluster: Requirements, Design, and Sample Applications[1]

CHARLES MENEVEAU, Johns Hopkins University

The massive datasets now generated by Direct Numerical Simulations (DNS) of turbulent flows create serious new challenges. During a simulation, DNS provides only a few time steps at any instant, owing to storage limitations within the computational cluster. Therefore, traditional numerical experiments done during the simulation examine each time slice only a few times before discarding it. Conversely, if a few large datasets from high-resolution simulations are stored, they are practically inaccessible to most in the turbulence research community, who lack the cyber resources to handle the massive amounts of data. Even those who can compute at that scale must run simulations again forward in time in order to answer new questions about the dynamics, duplicating computational effort. The result is that most turbulence datasets are vastly underutilized and not available as they should be for creative experimentation. In this presentation, we discuss the desired features and requirements of a turbulence database that will enable its widest access to the research community. The guiding principle of large databases is "move the program to the data" (Szalay et al. "Designing and mining multi-terabyte Astronomy archives: the Sloan Digital Sky Survey," in ACM SIGMOD, 2000). However, in the case of turbulence research, the questions and analysis techniques are highly specific to the client and vary widely from one client to another. This poses particularly hard challenges in the design of database analysis tools. We propose a minimal set of such tools that are of general utility across various applications. And, we describe a new approach based on a Web services interface that allows a client to access the data in a user-friendly fashion while allowing maximum flexibility to execute desired analysis tasks. Sample applications will be discussed. This work is performed by the interdisciplinary ITR group, consisting of the author and Yi Li(1), Eric Perlman(2), Minping Wan(1), Yunke Yang(1), Randal Burns(2), Shiyi Chen(1), Gregory Eyink (3) Alex Szalay (4) with the following departmental affiliations: (1) Mechanical Engineering, (2) Computer Science (3) Applied Mathematics & Statistics, (4) Physics and Astronomy.

[1]This research is funded by a National Science Foundation ITR/ASE grant.