Invited: SHWD-I-02

Extensible, reusable, and reproducible computing: a case study of PySPH

Prabhu Ramachandran
Department of Aerospace Engineering, IIT Bombay, Powai, Mumbai 400076

Reproducibility and the ability to incorporate/extend the research of others has always been a cornerstone of scientific research. However, even a cursory sampling of most modern publications in computational research will show that this does not hold true in practice. Research papers are often hard to extend, reuse or reproduce. Sometimes results are not reproducible because either some parameters are mistyped or not specified. The authors cannot be blamed entirely as it is easy to overlook some details of a particular computation. There is little interest or incentive in making one's results reusable and reproducible. There are several strategies to overcome these but they typically require a fair bit of effort and knowledge of software engineering.

Many researchers are not experts in software engineering, yet high performance code is also desirable. It would be ideal to have an environment where the user of a tool can write high-level code that is numerically well behaved, mathematically consistent, and physically relevant without worrying about high-performance or requiring a degree in software engineering.

In this talk, I will consider the Smoothed Particle Hydrodynamics (SPH) technique as an example of a typical computational research area. Over the last several years my group has been working on an open source framework for SPH called PySPH(http://pysph.bitbucket.org). PySPH is implemented in Python which is a powerful, easy to learn programming language. We automatically generate high performance code from high-level Python code. We use good software development practices for testing our code on a regular basis on multiple platforms. Finally, we use simple tools to automate computational tasks. This can be used to facilitate easy reproduction of results by others. This does require additional effort but is very useful to both the researchers themselves and others. The collection of practices makes it easy to extend, reuse, and reproduce any results produced from the framework without sacrificing performance.

We discuss the general techniques and approaches we have used to achieve this so as to make this useful to a general audience.