FROM VISUALISATION TO DATA MINING WITH LARGE DATA SETS *

A. Adelmann, PSI, Villigen, Switzerland
R.D. Ryne, LBNL/AFR, Berkeley, California, USA
J.M. Shalf, C.Siegerist, LBNL/NERSC, Berkeley, California, USA

Abstract

In 3D particle simulations, the generated 6D phase space data can be very large due to the need for accurate statistics, sufficient noise attenuation in the field solver and tracking of many turns in ring machines or accelerators. There is a need for distributed applications that allow users to peruse these extremely large remotely located datasets with the same ease as locally downloaded data. This paper will show concepts and a prototype tool to extract useful physical information out of 6D raw phase space data. PartView allows the user to project 6D data into 3D space by selecting which dimensions are represented spatially and which dimensions are represented as particle attributes, and the construction of complex transfer functions for representing the particle attributes. It also allows management of time-series data. An HDF5-based parallel-I/O library, with C++, C and Fortran bindings simplifies the interface with a variety of codes. A number of hooks in PartView will allow it to connect with a parallel back-end that is able to provide remote file access, progressive streaming, and even parallel rendering of particle sets in excess of 1Billion particles.

MOTIVATION FOR 3D VISUALIZATION METHODS FOR UNDERSTANDING LARGE DATASETS

With particle accelerator simulations moving from *qualitative* to *quantitative* predictions, the dimensionality of the models suddenly changes to three spacial dimensions. For particle simulations this implies a 6 dimensional phase space in the continuum and depending on the model at least a 3 dimensional discrete space to solve, for example, a Poisson problem in the course of space charge dominated beam simulations [1].

On the other hand next generation light sources such as X-ray Free Electron Laser (XFEL), will require start to end simulations including several challenging problems such as: effects of coherent synchrotron radiation, detailed modeling of various misalignment effects and detailed and accurate predictions of emittance growth. In these type of problems a large number of macro particles is required, large scales have to be resolved and the ability to scan a huge parameter space leads to larger datasets.

* This work was partially supported by the Director, Office of Science, Office of Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

3D VISUALIZATION WITH AVS/EXPRESS

The most common method used to inspect the simulation data is to create scatterplots of a subset of the 6 dimensions (eg. plot one of the phase components against the z-coordinate for each particle in the simulation). The visualization in 3D requires the use of specialized visualization software and the development of domain specific applications. Among the commercial visualization packages AVS/Express[2] [3] provides a good environment to quickly develop custom applications and to explore different visualization techniques. Its modular design makes it easy to add custom library modules to the already good library provided with the application. Its 64-bit version supports fairly big datasets with reasonable interactivity. We developed a custom reader that supports the HDF5 data model described in [4] and applications to investigate the time varying 6 dimensional data sets. The number of particles for these applications are around 100,000.

As particle beams are accelerated and focused they interact with their environment. These interactions can produce beam halos, beam quality degradation and can result in beam instabilities. To provide insight into the behaviour of the particles along their trajectories we developed a particle tracking module that allows the user to interactively select particles in a region of interest and follow them in time. Once the particles in the region are identified the user can restrict their selection even more to avoid cluttering of the resulting trajectories. The particles and their trajectories are colored by the magnitude of the momentum. Figure 1 shows a group of particles selected in the halo area in the final distribution of the simulation and their position in the initial step. Another application allows the visualization of the time series selecting which of the 6 dimensions are considered as coordinates and which ones as node data. Figure 2 shows a typical view of the z-pz plane for the same dataset of Figure 1.

To visualize 3D electron cloud instabilities we added more than one species of particles to the application and rendered them as independent fields. The interactive selection of electrons and the visualization of their trajectories allows us to see their complex 3 dimensional behavior. See Figure 3 for an example of electron trajectories.

PARTVIEW

AVS/Express offers considerable flexibility for constructing sophisticated visualizations, but often users find

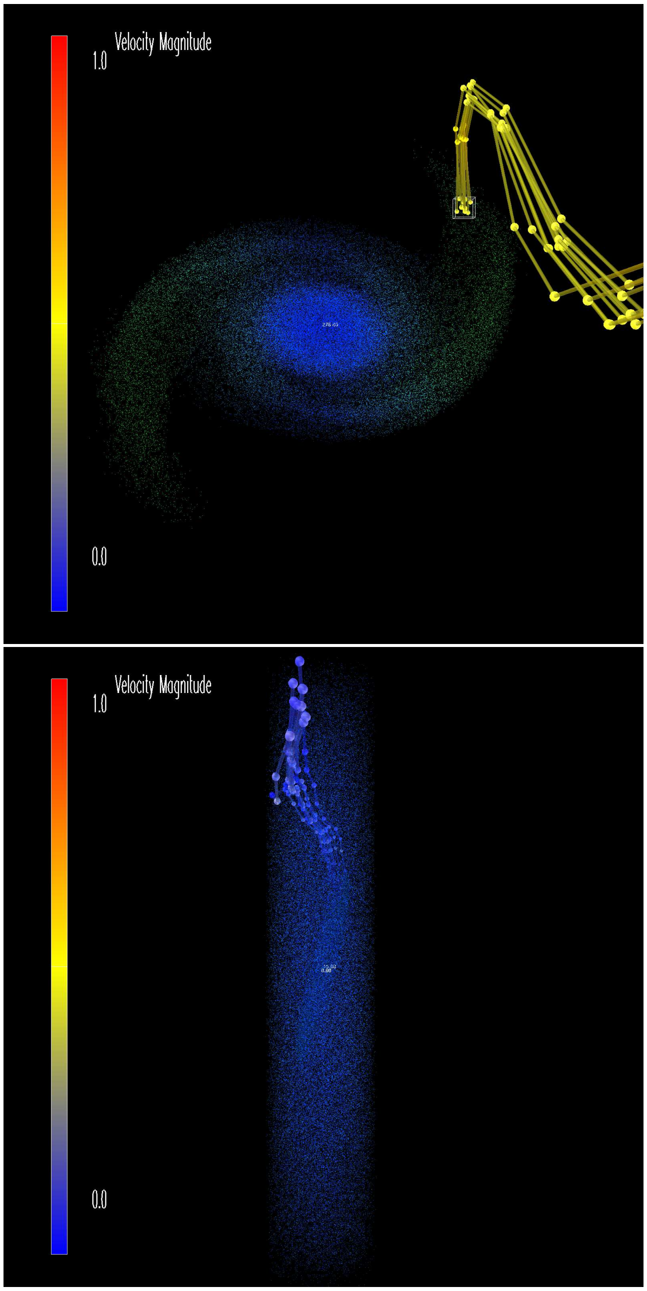

Figure 1: color: Tracking of proton trajectories as the simulation progresses.

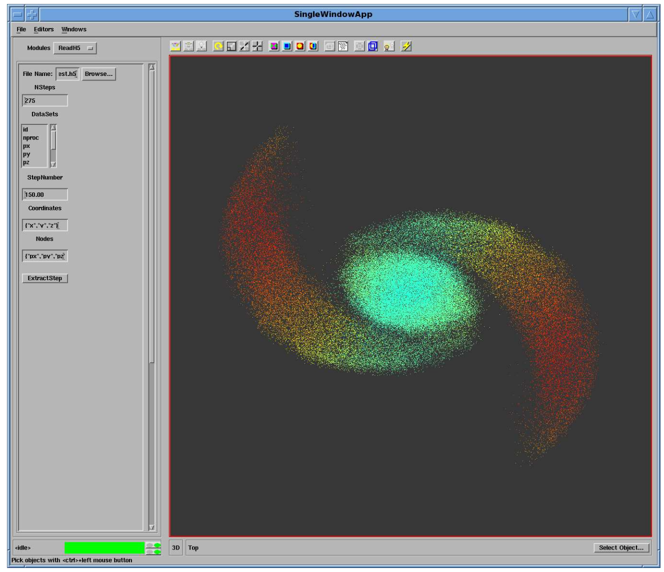

Figure 2: color: AVS/Express interface.

the user interface daunting. The PartView tool is designed to provide a straightforward interface that covers the simpler 80% of the kinds of 3D analysis and inspection that the accelerator modeling researchers need in order to understand their dataset. It is consistent with recommendations from our users who indicated a preference for tools that had a user interface that was simpler and customized to their problem over powerful/complex general-purpose user interfaces.

In addition, the particle simulations are scaling up very rapidly and may soon exceed the ability of serial tools like Express to manage the resulting datasets. There is a need for lightweight front-end applications that can tap into remote I/O and rendering services that are located at the site where the datasets are produced. This very same architecture can also be used to connect to running applications for visual inspection of simulations that are in progress. This kind of distributed application approach can offer the interactivity of a desktop tool, but allow users to peruse extremely large remotely located datasets with the same ease as locally downloaded data.

To address these concerns, we created PartView. PartView was constructed using FLTK[5], OpenGL[6], and HDF5[7] so that it can easily be compiled on a number of platforms. The tool allows the user to project 6 dimensional data into a 3 dimensional space by selecting which dimensions are represented spatially and which dimensions are represented as particle attributes (color and transparency currently), and the construction of complex transfer functions for representing the particle attributes. It also allows management of time-series data and two different file formats for storing the particle accelerator data. Figure 4 shows PartView's GUI with the colormap editors (gaussian and solid).

An HDF5-based parallel-I/O library was written in conjunction with this application that allows more than one group to use this same tool to understand their data. The HDF5 file format will enable the simulations to handle the I/O requirements of simulations that are an order of magnitude larger than is currently reasonable and also support some interoperability between groups with similar needs. The API for the file format has C++, C and Fortran bindings in order to interface with a variety of different codes employed in this area. See [4] for more details.

PartView has a number of hooks that will allow it to connect with a parallel back-end that is able to provide remote file access, progressive streaming of the largest datasets,

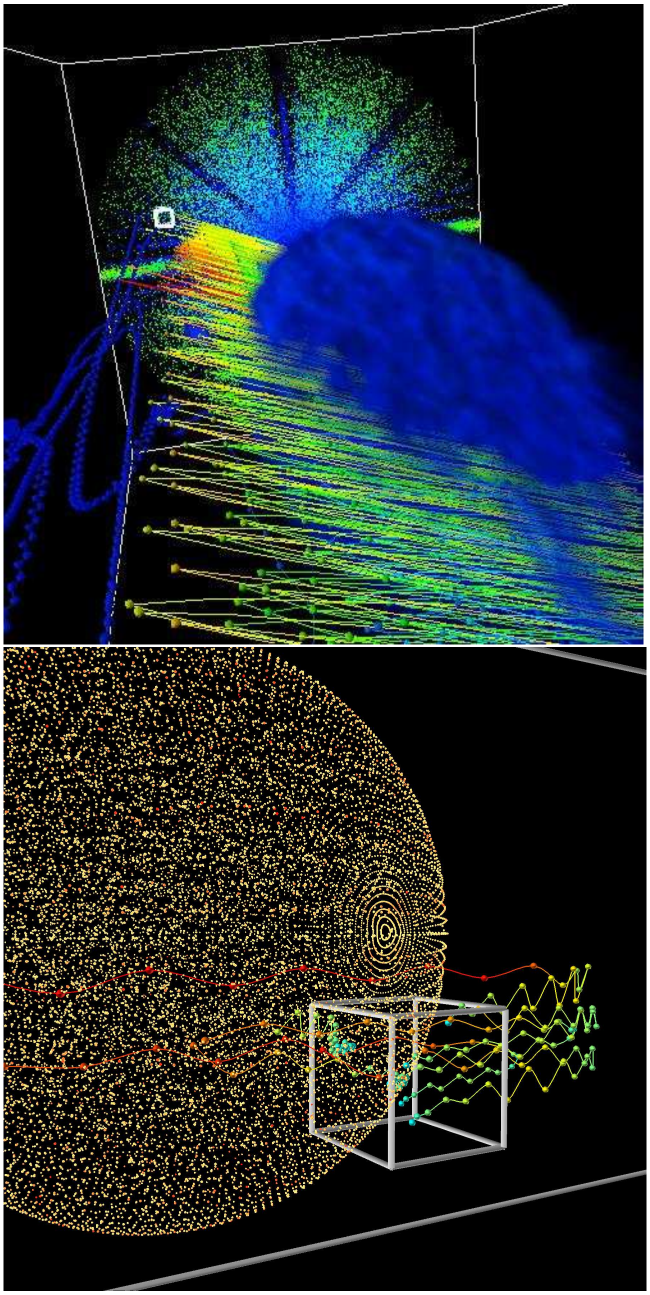

Figure 3: color: Electron cloud visualization: electron trajectories as the simulation progresses.

and even parallel rendering of particle sets in excess of 1Billion particles. This capability is still in development.

CONCLUSIONS

As high performance computing becomes indispensable for the design and test of particle accelerators, the high dimensionality of the problems and the size of the simulations requires the design of clever visualizations that can help understand the underlying physical nature of the problem. As one example, the tracking of particles that are lost in the accelerator can help optimize the starting conditions

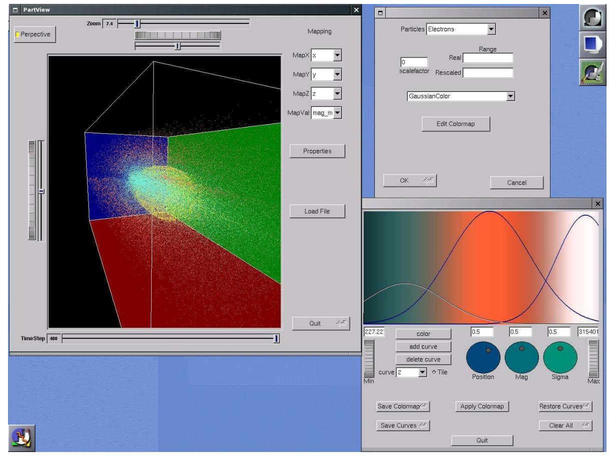

Figure 4: color: PartView's GUI with the colormap editor.

of the particle distributions.

Over the course of the next year, our primary focus will be on increasing the capability and capacity of PartView which is being targeted at simulations that generate 1 billion particles or more.

We will continue to use AVS/Express to produce the more sofisticated visualizations, high resolution images and animations. For the bigger data sets we will need to find an acceptable subsampling mechanism.

ACKNOWLEDGEMENTS

The work reported in this paper (LBNL-57603) was performed under the support from the DOE SciDAC project on ”Advanced Computing for 21st Century Accelerator Science and Technology”

REFERENCES

[1] A. Adelmann et al. ,On Start to End Simulation and Modeling Issues of the Megawatt Proton Beam Facility at PSI , PAC05, submitted.

[2] Advanced Visual Systems, http://www.avs.com.

[3] Jean M. Favre and Mario Valle, AVS and AVS/Express, The Visualization Handbook, p. 655, Elsevier Butterworth Heinnemann, 2005.

[4] A. Adelmann, R.D. Ryne, J.M. Shalf, C. Siegerist, A Portable Parallel High Performance Data Interface for Particle Simulations, PAC05, submitted.

[5] Fast Light Toolkit, http://www.fltk.org.

[6] http://www.opengl.org.

[7] Hierarchical Data Format 5, http://hdf.ncsa.uiuc.edu/HDF5.