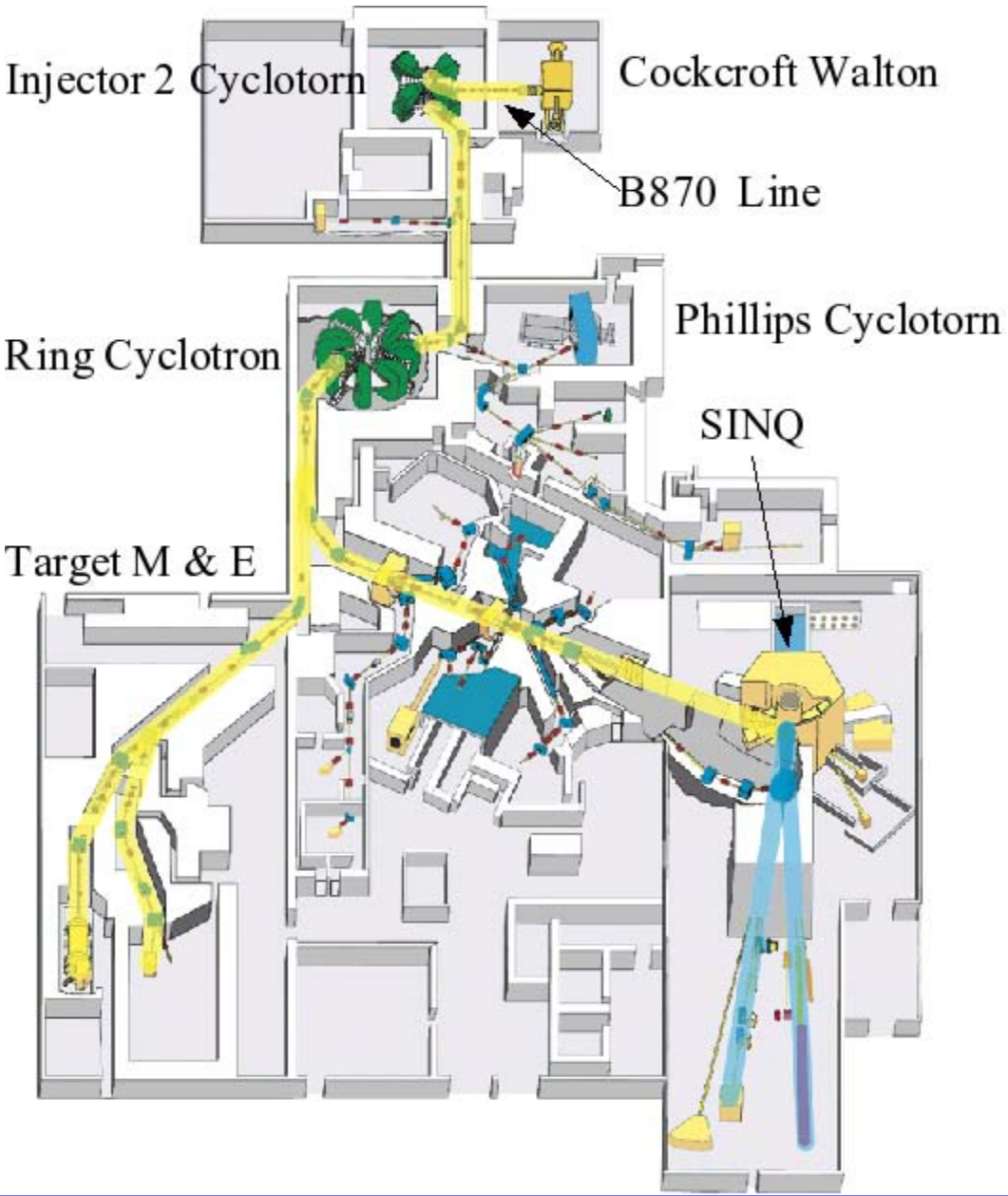
H5Part: A Portable High Performance Parallel Data Interface for Electromagnetics Simulations

A. Adelmann, A. Gsell, B. Oswald, T. Schietinger (PSI)
W. Bethel, J.M. Shalf, C. Siegerist, (LBNL/NERSC, Berkeley)

5. October 2006

1. Motivation
2. H5Part
3. H5Part and Data Analysis
4. Summary and Outlook

Motivation



One Modeling Challenge

Total controlled loss prediction in the order 10^{-3} and fractional uncontrolled losses in the order of 10^{-6} per meter.

Consequence

We need Precise Beam Dynamic Simulations in order to obtain quantitative answers.
Massive parallel simulations are indispensable.

Massive Parallel Particle Tracking (MAD9p/IPPL)

- production runs:
  - Tracking 10^7 particles
  - 3D Space Charge FFT on a 256^3 grid
  - 2D domain decomposition

Observations

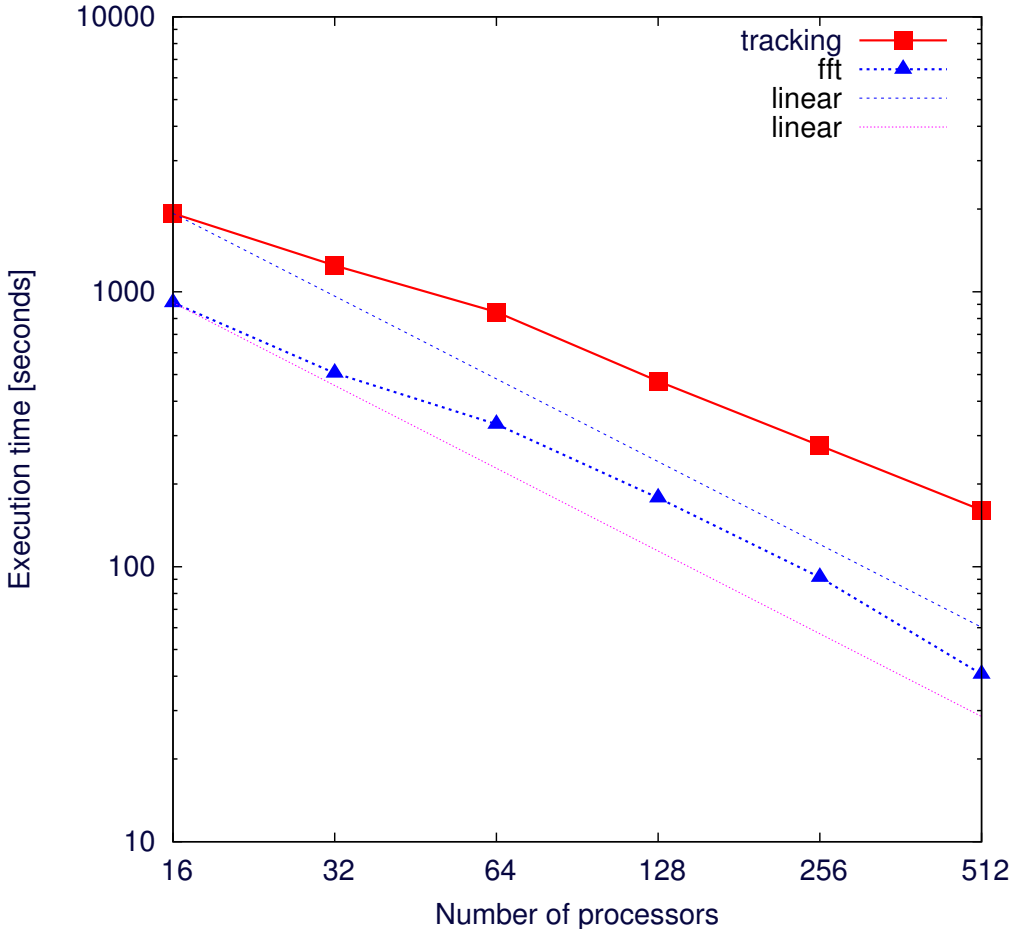
- Max so far: P=1024, $N = 10^9$, $M = 4096^3$, 8 TBytes phase space data

Massive Parallel Particle Tracking (MAD9p/IPPL)

- production runs:
  - Tracking 10^7 particles
  - 3D Space Charge FFT on a 256^3 grid
  - 2D domain decomposition

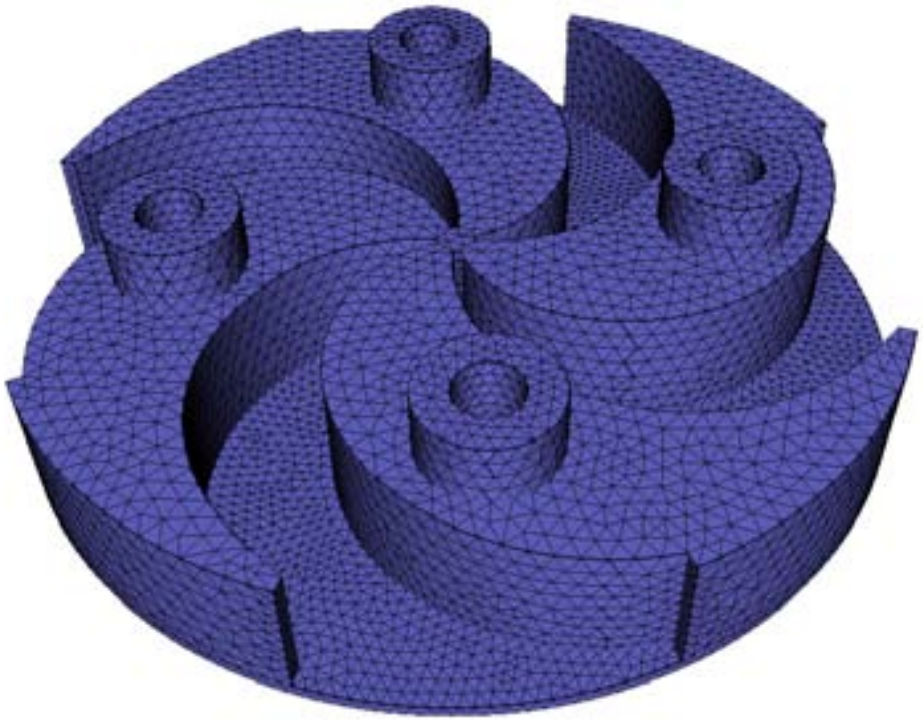
Observations

- The code scales well (no output)



Eigenmode Calculations (FemaXX)

- 5 lowest eigenvalues with eigenvectors of the COMET cavity
- 1.4 million DOFs
- Post-processing not included



Usage

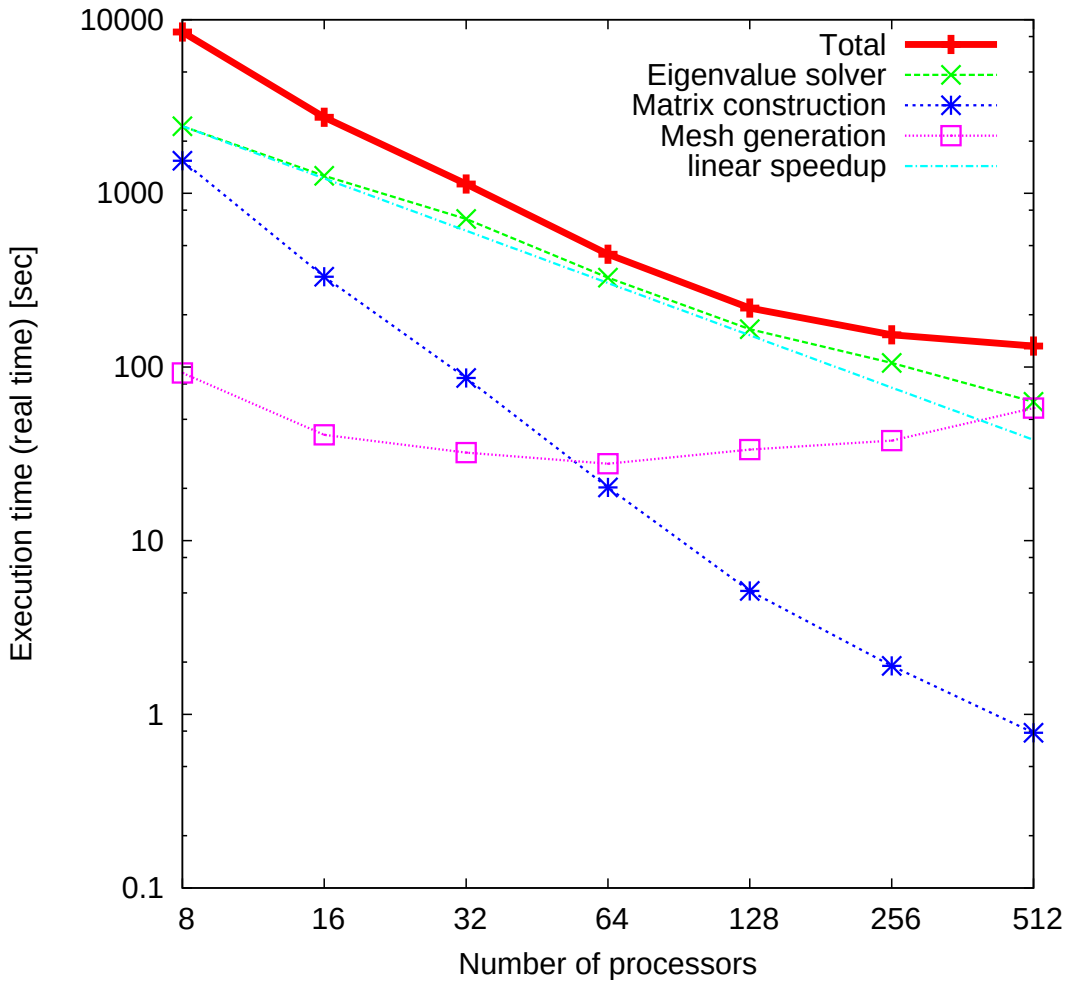
- Integrated into the RF-Design cycle at PSI
- validated against HFSS, ANSIS and MAFIA on real world problems!

Eigenmode Calculations (FemaXX)

- 5 lowest eigenvalues with eigenvectors of the COMET cavity
- 1.4 million DOFs
- Post-processing not included

Observations

- The code scales well to large number of cpus
- Computation takes only 4 minutes on 512 cpus



A honest I/O Story

During December, 2005 - January, 2006 FLASH team members executed the largest homogeneous, isotropic compressible turbulence run with Lagrangian tracers on BG/L at LLNL using FLASH 3 (Data from Katie Antypas)

- Parallel I/O libraries were unavailable
- Implement simple direct I/O where each processor writes its own file to disk – 32,000 files for one dump
- Post processing nightmare.... Linux tools where broken
  - 154 TB of disk capacity in 74 million files

A honest I/O Story

During December, 2005 - January, 2006 FLASH team members executed the largest homogeneous, isotropic compressible turbulence run with Lagrangian tracers on BG/L at LLNL using FLASH 3 (Data from Katie Antypas)

- Parallel I/O libraries were unavailable
- Implement simple direct I/O where each processor writes its own file to disk – 32,000 files for one dump
- Post processing nightmare.... Linux tools where broken
  - 154 TB of disk capacity in 74 million files

1 week wall clock on 65k processors results in 6 months of data post processing to get even minimal results out of the data.

In consequence:

- parallel I/O is a non trivial matter
- parallel I/O will be of utmost importance on PETAFLOP machines

H5Part

Design Philosophy of H5Part

For success i.e. real scalable codes we have to use existing frameworks. In consequence:

- H5Part is a simple API on top of HDF5
- Provides a data parallel look and feel
- Hides necessary meta data from the user

Hence H5Part inherits many of the desired features from HDF5:

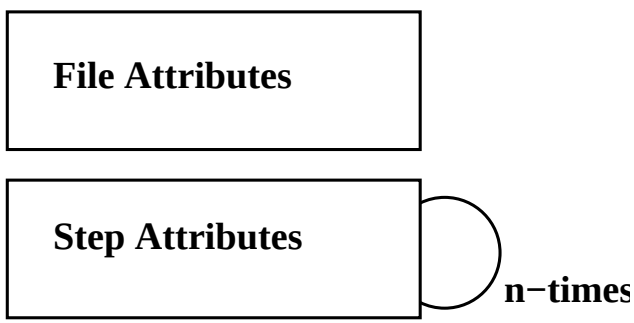
- self describing
- platform independent and platform tuned
- parallel (if needed)
- all HDF5 tools (h5dump etc.)
- C/C++, Fortran77 and Fortran90, Python (soon)

Building Blocks of H5Part

- Particles
- Block structured data (Block)
- Topology (Topo) not yet implemented

H5Part File Layout

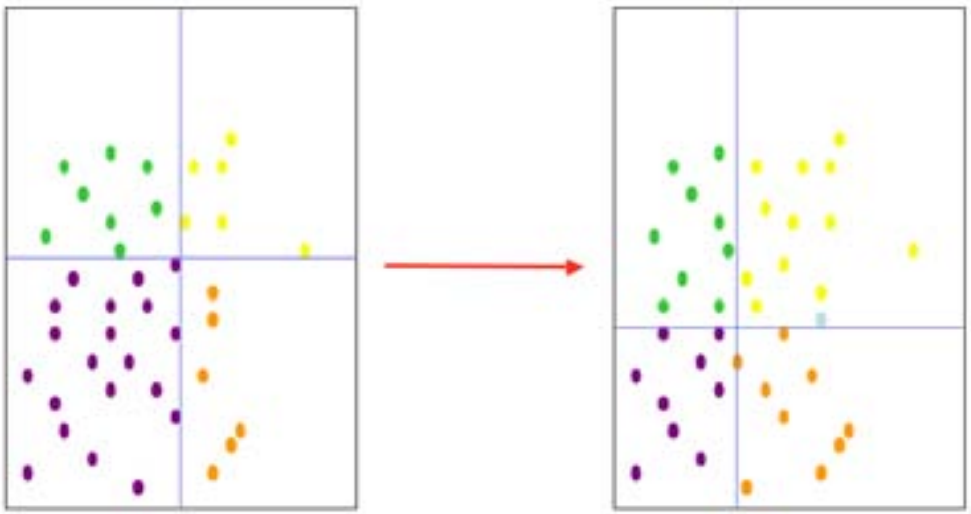
- File Attributes: Units, title, lattice file
- Step Attributes: Fields, particles, statistical quantities



H5Part Basic Attribute Types

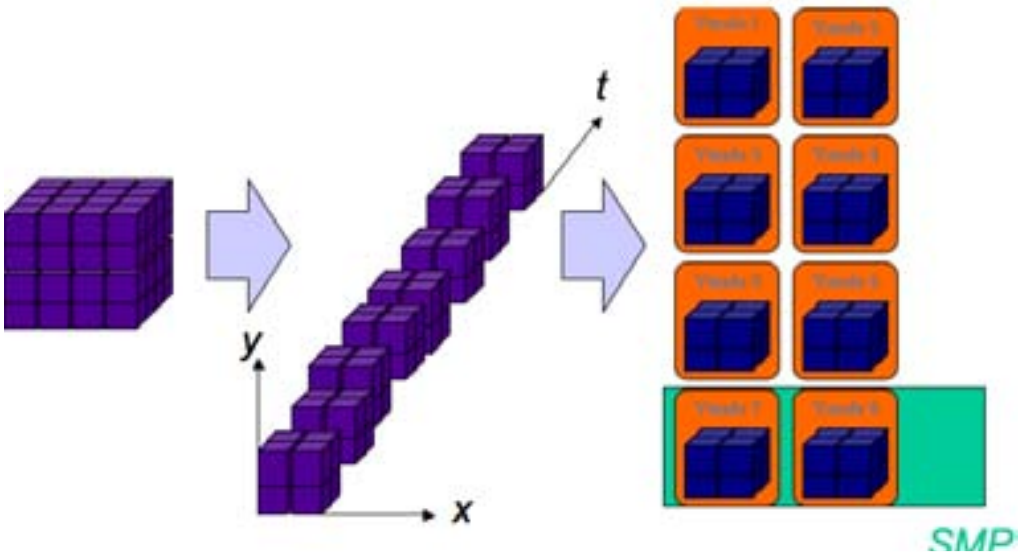
- String
- Array of 64bit floating point values
- Array of 64bit integer values

```
if(parallel);
  h5fp = h5pt_openwpar("IMPACTT.h5",commMPI)
else
  h5fp = h5pt_openw("IMPACTT.h5")
rc = h5pt_setnpoints(h5fp,h5partint)
rc = h5pt_writefileattrib_string(h5fp,"rmsUnit","m")
loop(step = 1,NSteps);
  rc = h5pt_setstep(h5fp,step)
  rc = h5pt_writestepattrib_r8(h5fp,"gamma",gammar8,1)
  rc = h5pt_writestepattrib_r8(h5fp,"rms",rmsr8,3)
  rc = h5pt_writedata_r8(h5fp,"x",xpts(1:))
     do more stuff ...
end loop
rc = h5pt_close(h5fp)
```



Selected H5Part Features

- Provides a convenient layer to save restore particle data
- One file for all data (time steps)
- Follows domain decomposition



Selected H5Block Features

- Inherits all particle features
- Any number of fields per time-step
- Fields in time-step may have different dimensions
- Fields are stored in Fortran convention

H5Block Field Types

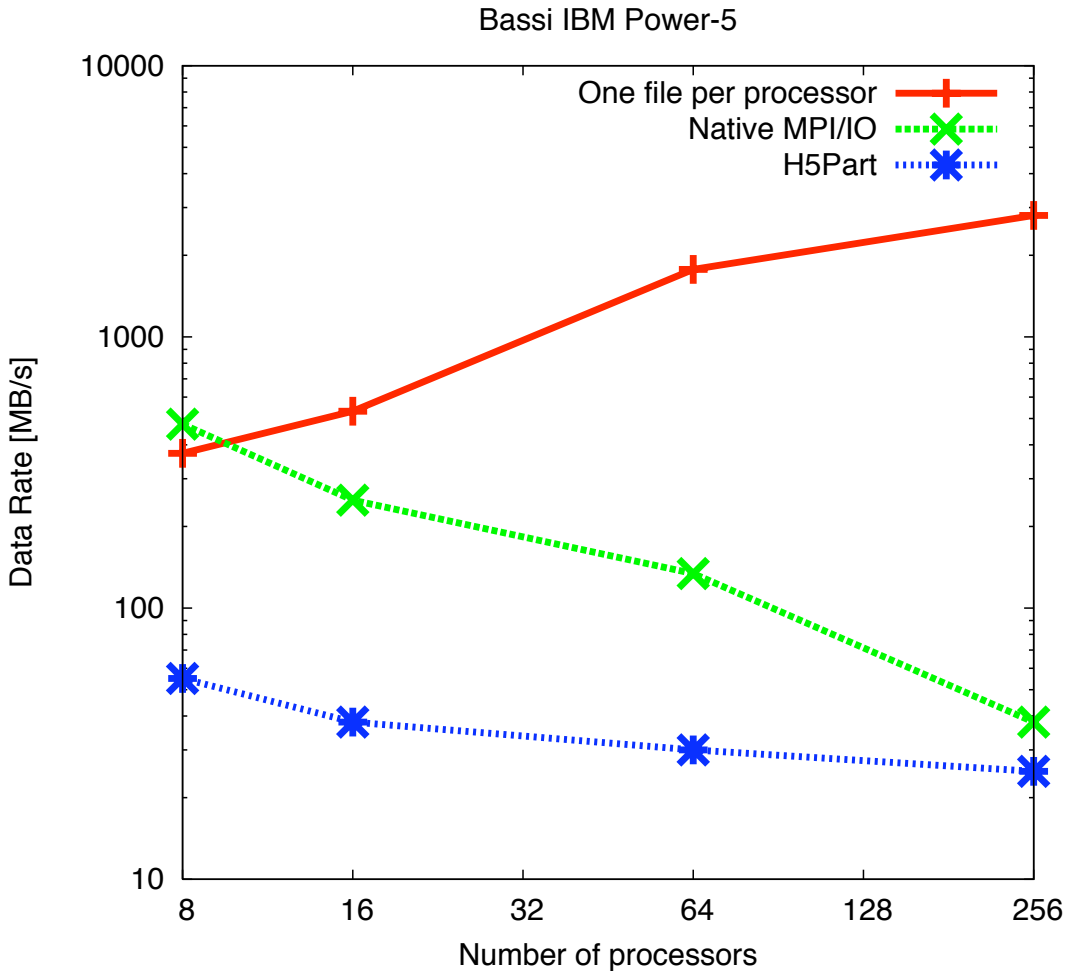
- 3d fields, 64bit floating point values:
  - Scalar
  - 3d vector
- Attributes can be asigned to fields to save additional information
- API functions for common attributes like “Origin” and “Spacing”
- Automatic handling of ghost-zones for writing to prevent concurrent writes

Performance

- save 10^7 particles (6D + ID)
- 10 time steps
- average times are shown
- 2D domain decomposition

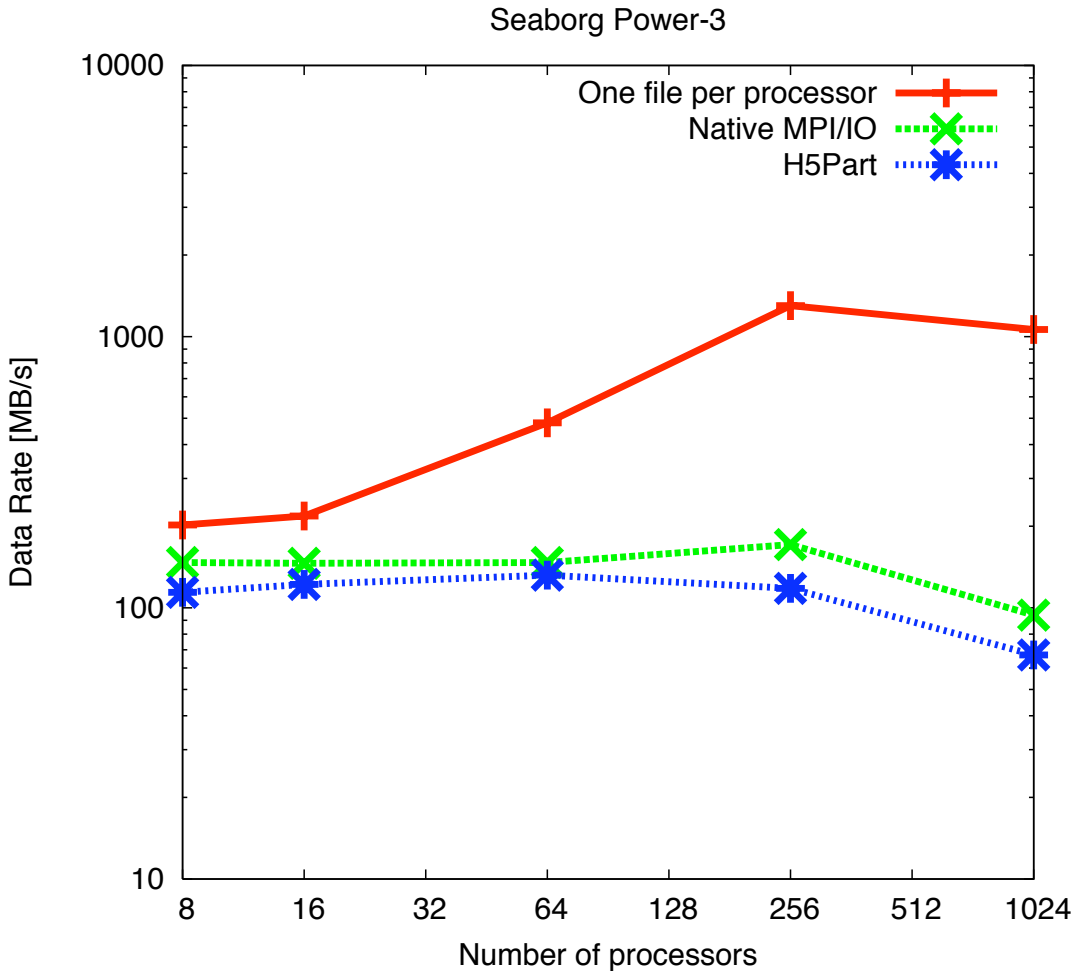
I/O Methods

- One file pre processor
- MPI/IO
- H5Part



Performance

- save 10^7 particles (6D + ID)
- 10 time steps
- average times are shown
- 2D domain decomposition



H5Part and Data Analysis

We have to cover from "gnuplot" type of simple but important data analysis to more sophisticated data mining task. Parallel data post processing will be important if we start to ask more complicated queries on our data.

H5Part Data Analysis Tools

- VisIt plugins

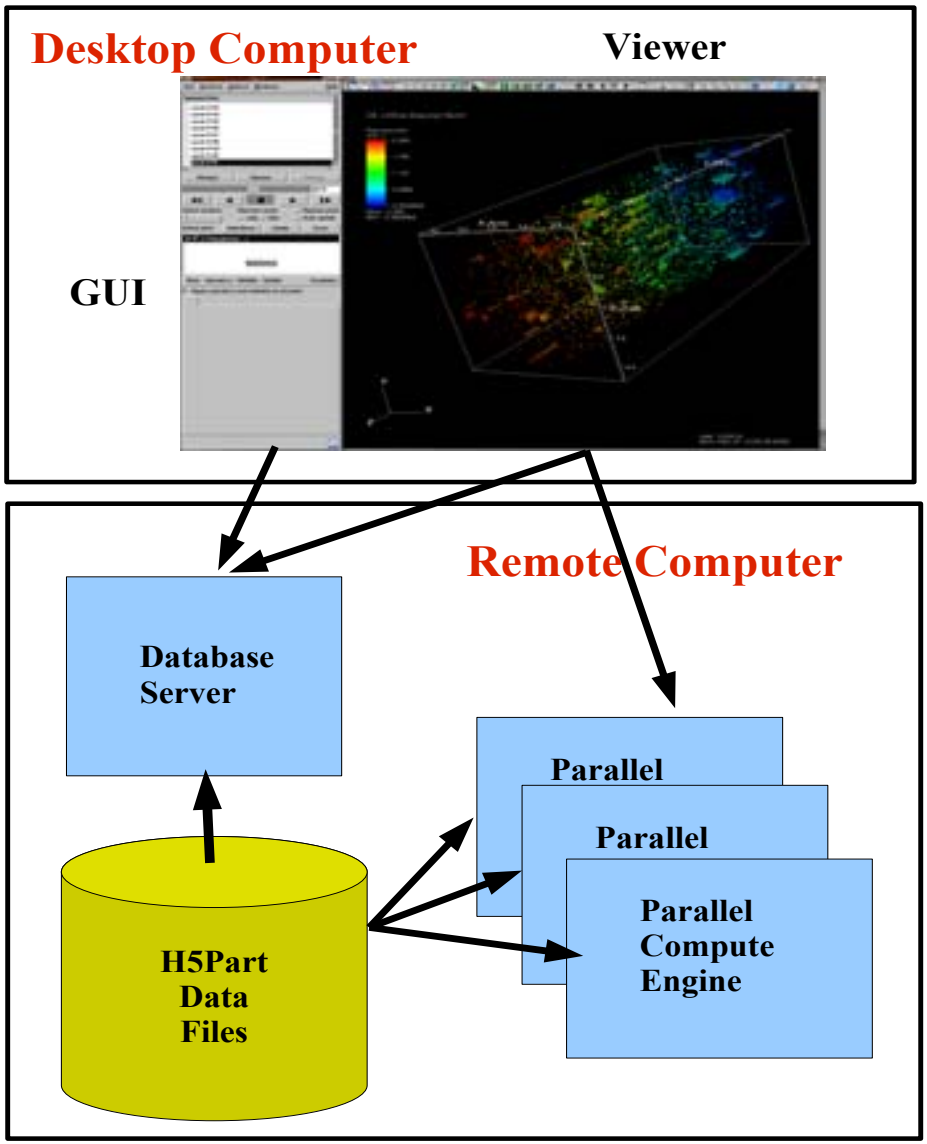
VisIt (LLNL) is a point-and-click 3D scientific visualization and analysis application that supports most of the standard visualization techniques (isocontouring, slicing, resampling, projection, volume rendering, movie generation) on time varying structured and unstructured grids.

- H5PartRoot

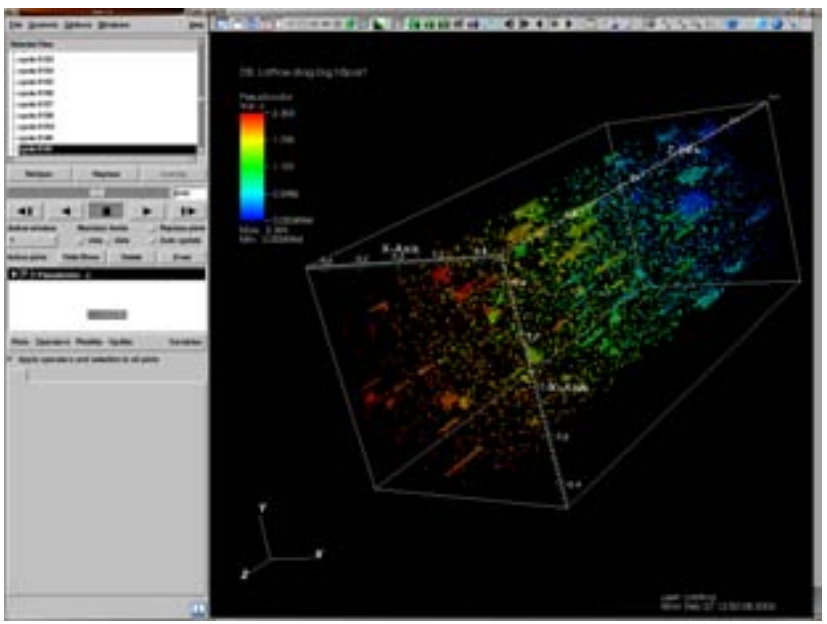
H5PartRoot is based on ROOT which is the main LHC data post processing toolbox.

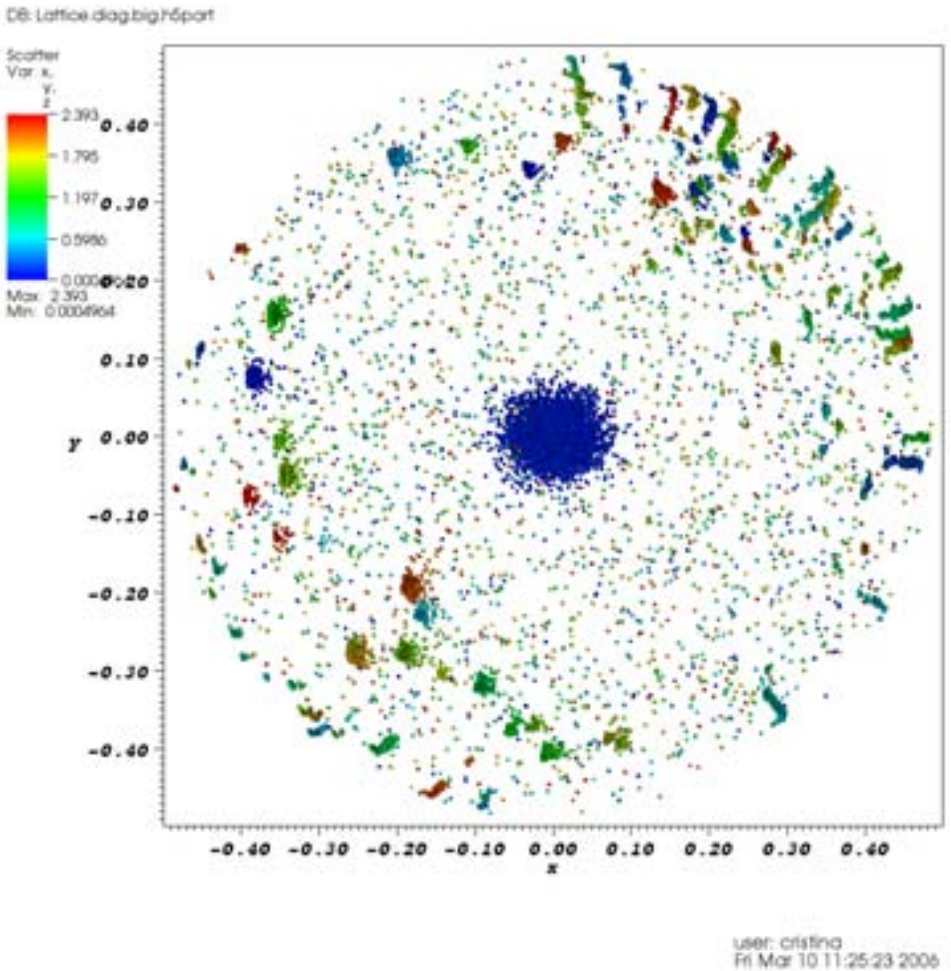
VisIt Plugins for H5Part

- It employs a parallel distributed architecture in order to handle extremely large data sets interactively or in batch
- Extensibility is achieved through the development of plugins
- It has a python interface for scripting frequently used tasks
- It is free and open source



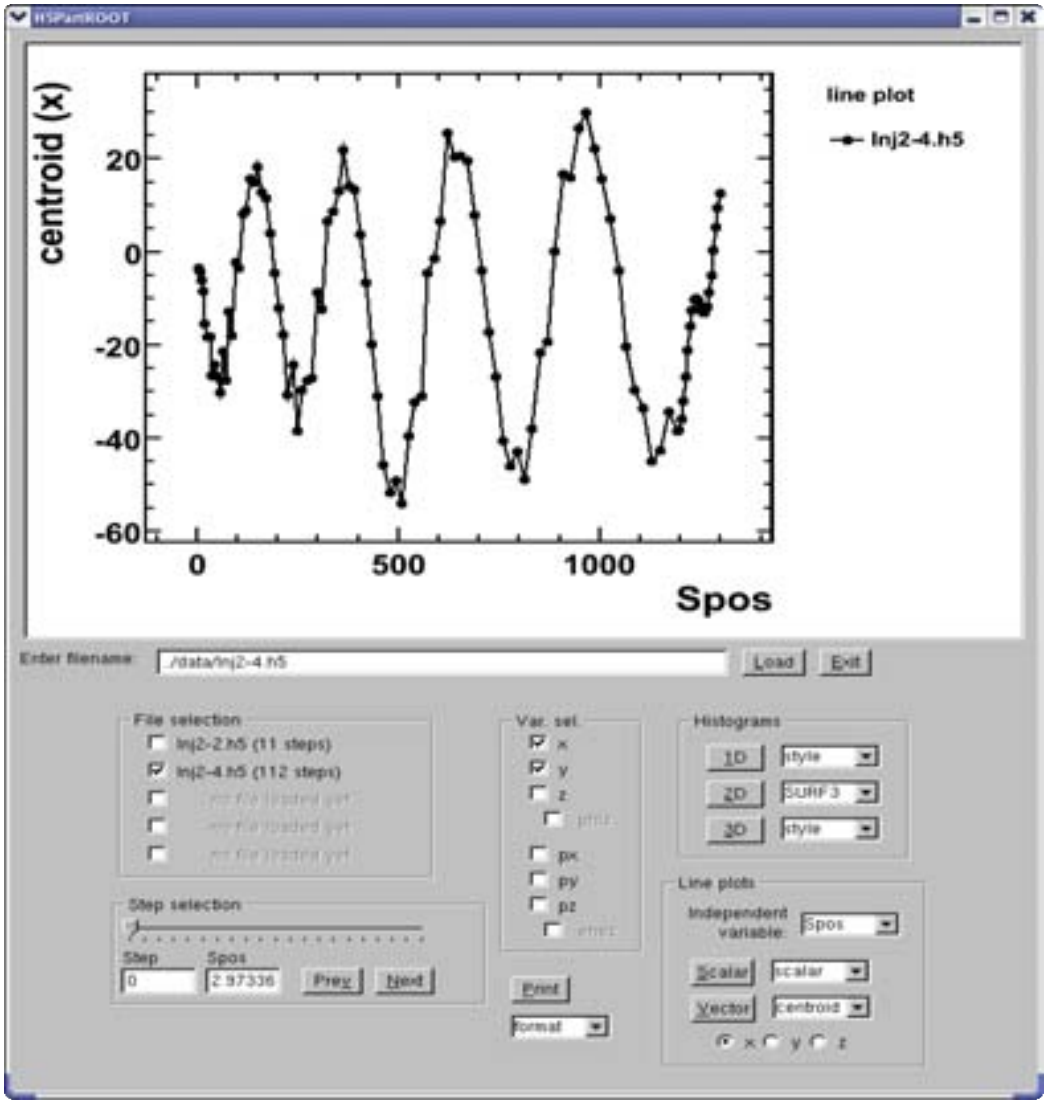
- H5Part reader plugin is developed for for Particle and Field data
- For more information: http://vis.lbl.gov/Research





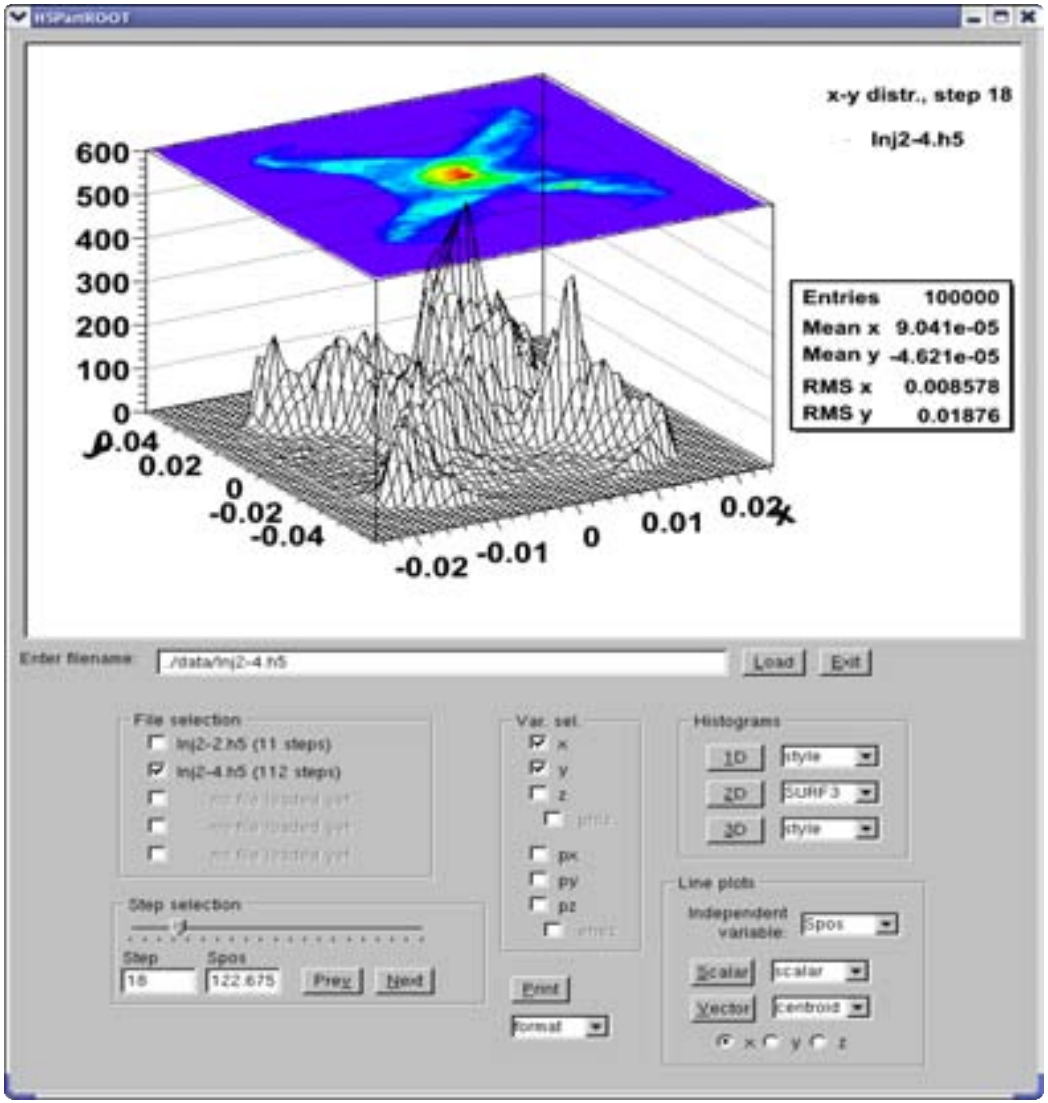
H5PartRoot

- line plots
- histograms
- one-, two-, and three-dimensional particle distributions
- data fitting
- inherits all capabilities provided by ROOT



H5PartRoot

- line plots
- histograms
- one-, two-, and three-dimensional particle distributions
- data fitting
- inherits all capabilities provided by ROOT



Summary and Outlook

Summary

- Parallel I/O is the most overlooked issue in the HPC community
- H5Part provides a convenient and efficient way to **store, retrieve** and **share largest amount of data**
- H5Part runs on the LapTop and the super computer
- H5Part comes with a set of mature tools for analysis
- H5Part is distributed with a LGPL License
- Checkout http://h5part.web.psi.ch/ or *andreas.adelmann@psi.ch*

Outlook

- H5Topo
  - for unstructured (FE) grids (B. Oswald)
- Fast index technology (K. Stockinger LBNL)