Scaling to 150K cores: recent algorithm and performance engineering developments enabling XGC1 to run at scale

M.F. Adams
Columbia University

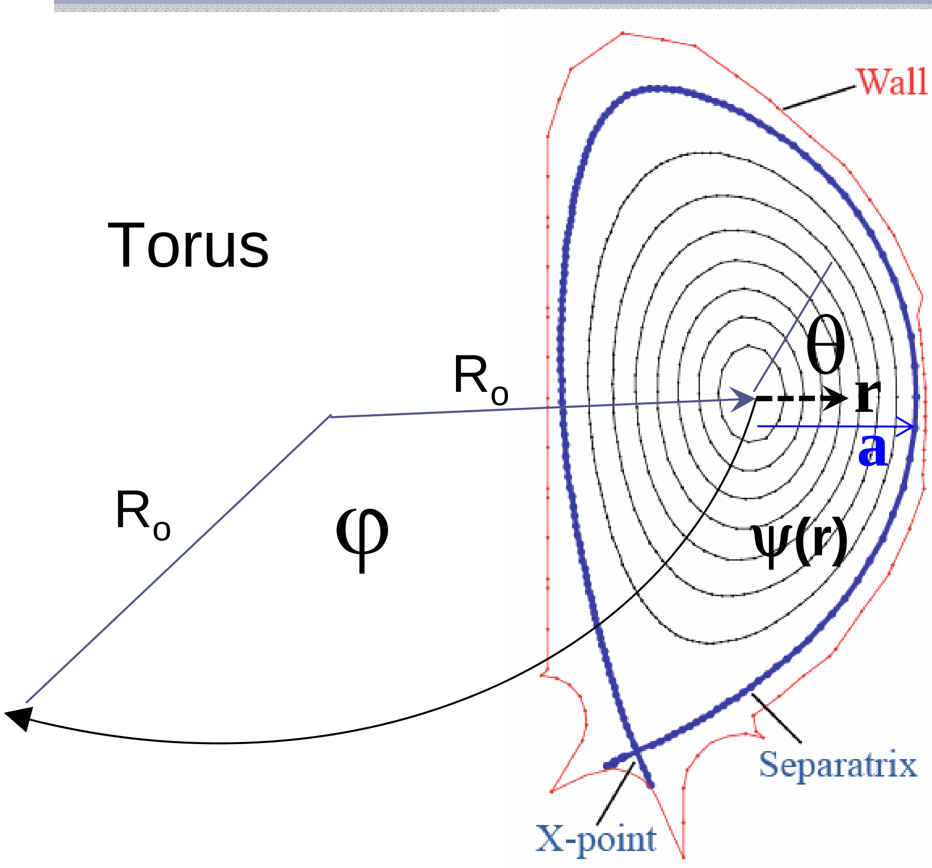
S. Ku, J. Cummings, P. Worley, E. D'Azevedo, C.S. Chang

and the CPES team
(SciDAC Center for Plasma Edge Simulation)

CScADS Workshop, Lake Tahoe, 27 - 30 July 2009

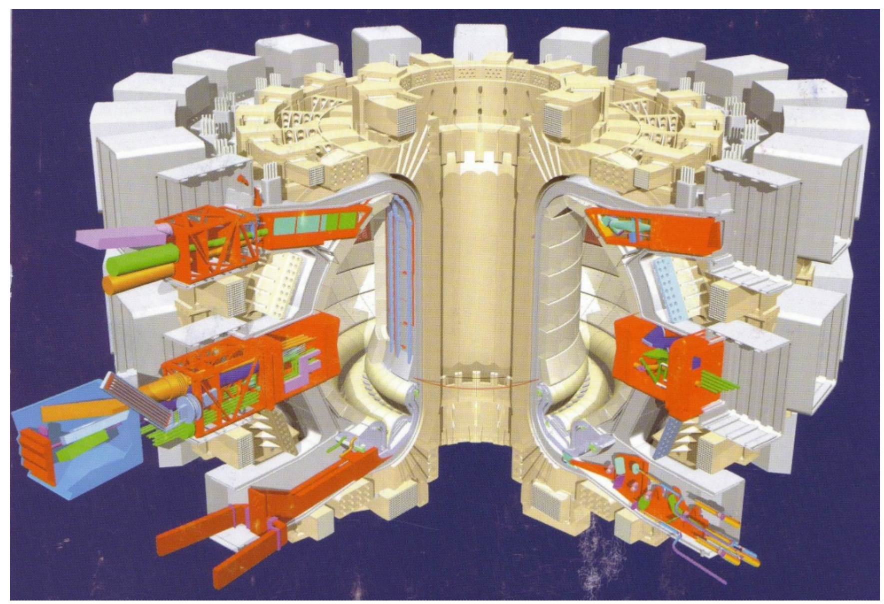
Center for Plasma Edge Simulation

Tokamak geometry

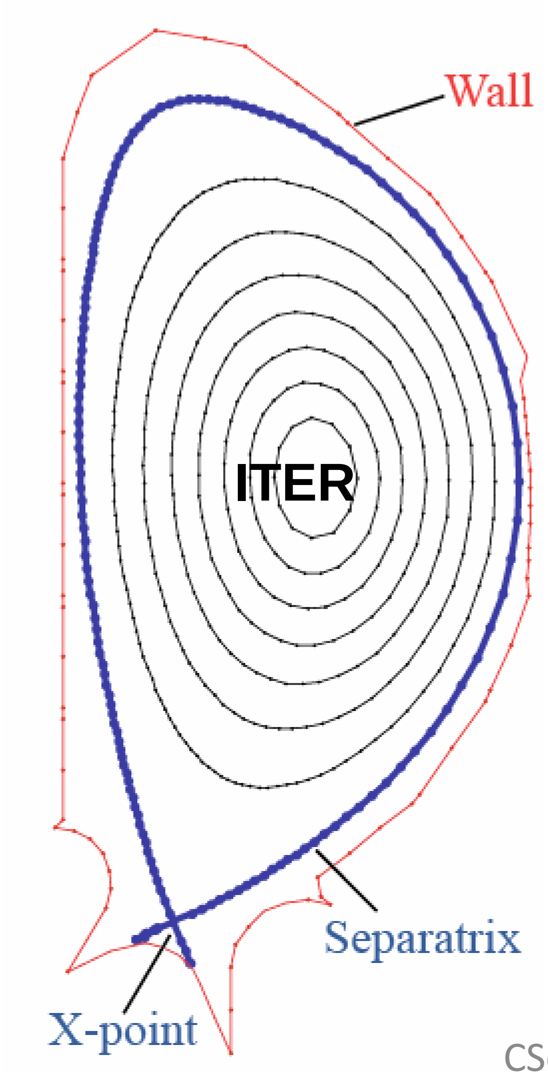


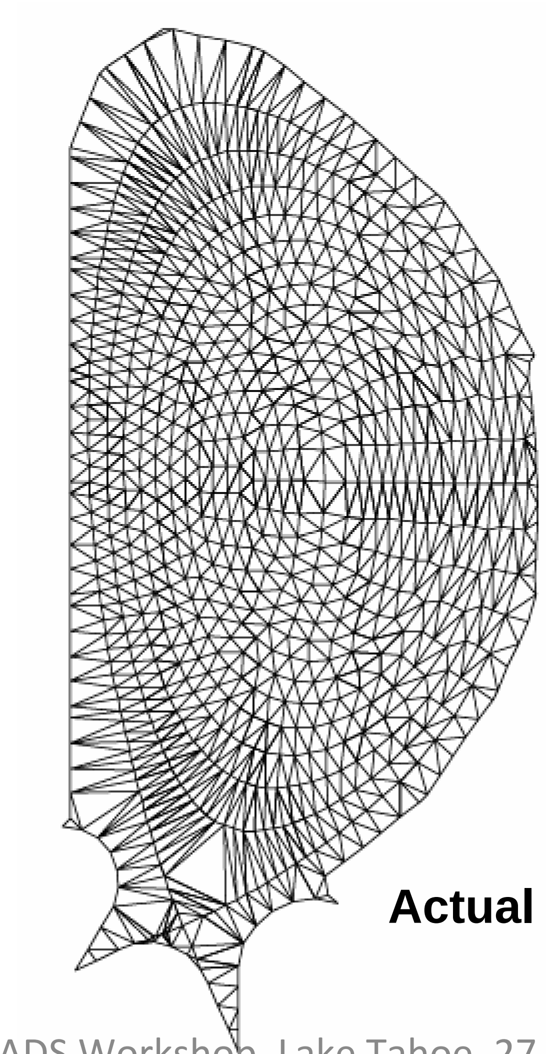
Poloidal cross-section (poloidal plane) at a constant toroidal angle

Poloidal magnetic flux label
$\psi(r)$: 1 at $r/a=1$, 0 at $r/a=0$



Field-line following unstructured triangular mesh

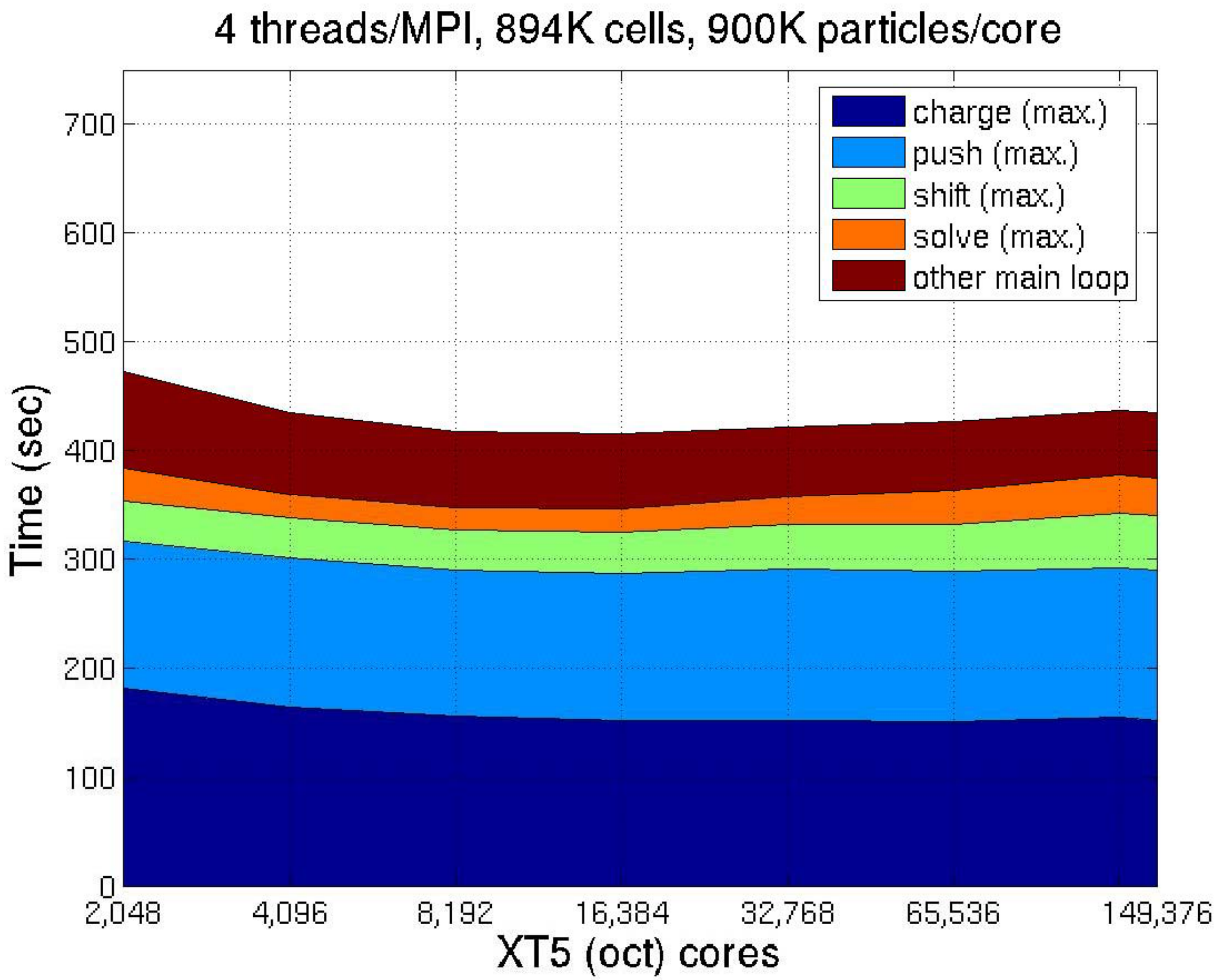




Unstructured triangular mesh on numerical B data → Extra difficulty in particle sorting and interpolation

Actual mesh is much denser.

Scaled speedup, ITER mesh, 4 thrds

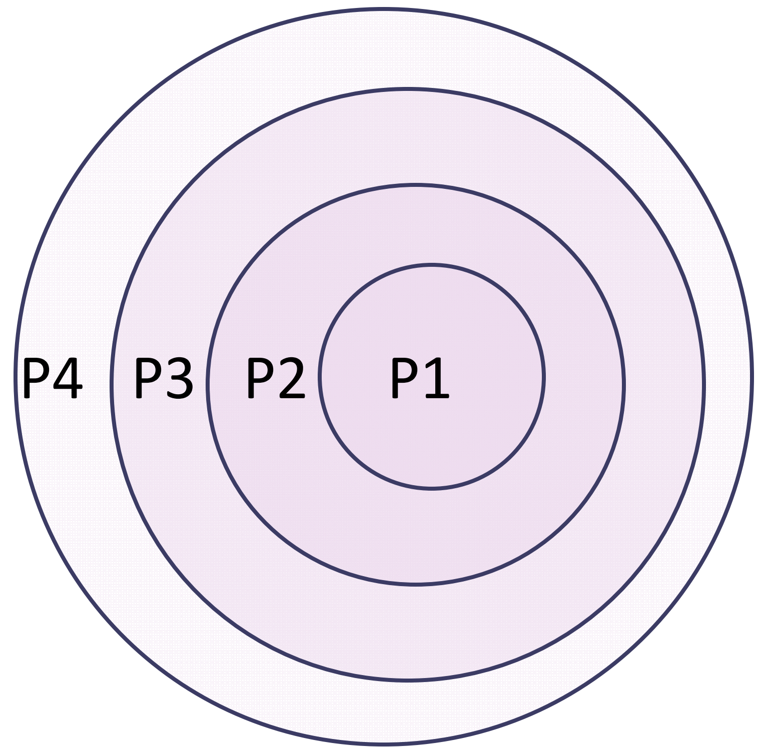


TODO

- Codes:
  - XGC1: Edge gyrokinetic PIC, unstructured
  - GTS: Core gyrokinetic PIC, semi-structured
- Both codes are similar (same community)
  - running well on Cray XTs
  - Fortran + MPI + OpenMP codes
  - Three processes to optimize:
    - Particles: simple array processing
    - Grid work (cells): (semi) unstructured processing
    - Interaction between particles and grid: most changing perf. issues
- Action items – BG/P
  - Finish porting ☹
  - Measure performance to scale ☺
  - ??? 😐

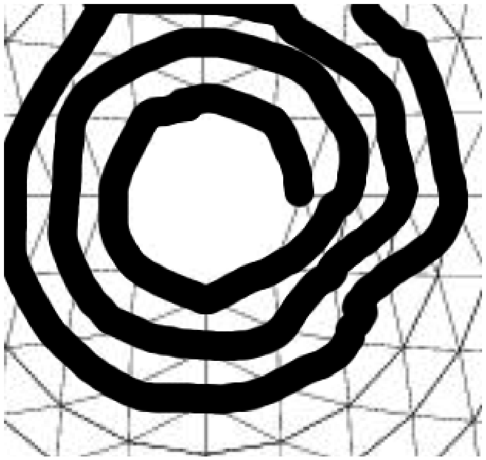
Dynamic Load balancing (M. Adams)

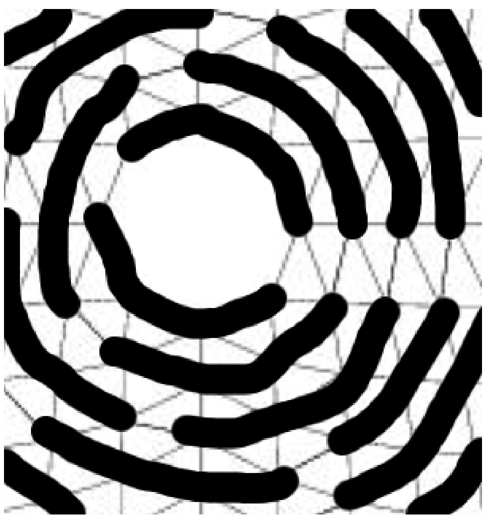
- Multi dimensional decomposition of particle data has been implemented.
- Multi dimensional decomposition raises load balancing issue.
- Dynamic load balancing is imposed.



multi-dimensional decomposition of field data of n = 4, [# of Proc] =16 case
Poloidal plain of tokamak

2D XGC decomposition





- Use given "space filling curve" of node ordering
- Simple "chopping" of grid data points
  - Uses space filling curve property for **data locality**
- Use nearest grid point method (Voronoi diagram) to partition domain

Performance improvements

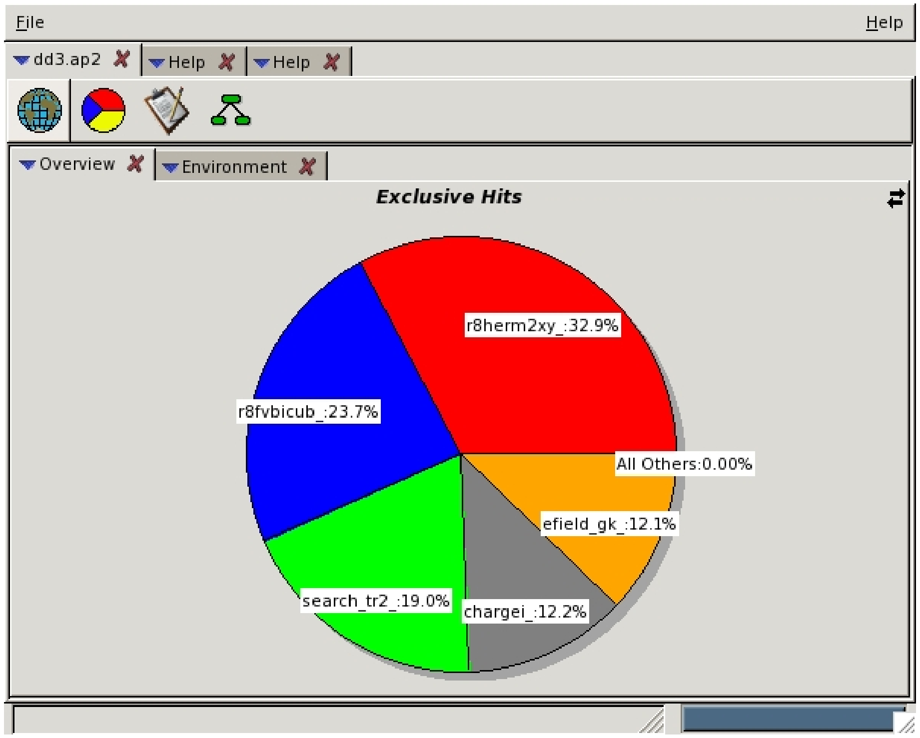
Cyclone test case -

~60,000 grid points

32 processors per plane

~50% of Time in spline routines

	Main loop time	Charge deposition time	Shift
No decomp.	199	118	5
Decomp.	159	93	14
% change	20	21	<280>



Excellent XGC1 scalability and rapid HPC development enables the formidable study

MPI + OpenMP on ITER grid