TD-SLDA Software Status: Extension to Nuclear Systems

A. Bulgac, A. Luo, P. Magierski, K. Roche, I. Stetcu, S. Yoon

Space-Time Lattice Extension to DFT - evolve time

- explicit time evolution of quasi-particle wavefunctions via Adams-Bashforth-Milne O(h^5)

- action is explicitly re-constructed each time step on the lattice

$$[\frac{\delta E(\rho, j, ...)}{\delta \rho} + \frac{\delta E(\rho, j, ...)}{\delta j} + ... + V_{ext}(\vec{x}, \sigma)]\psi_n(\vec{x}, \sigma) = E_n(\rho, j, ...)\psi_n(\vec{x}, \sigma)$$

$$i\hbar\partial_t \begin{pmatrix} u_n(\vec{x}, t) \\ v_n(\vec{x}, t) \end{pmatrix} = \begin{pmatrix} \hat{h}(\vec{x}, t) + \hat{V}_{ext}(\vec{x}, t) & \hat{\Delta}(\vec{x}, t) + \hat{\Delta}_{ext}(\vec{x}, t) \\ \hat{\Delta}^\dagger(\vec{x}, t) + \hat{\Delta}^\dagger_{ext}(\vec{x}, t) & -\hat{h}^*(\vec{x}, t) - \hat{V}_{ext}(\vec{x}, t) \end{pmatrix} \begin{pmatrix} u_n(\vec{x}, t) \\ v_n(\vec{x}, t) \end{pmatrix}$$

$|n| \sim O(N_x N_y N_z) \subset [40^3, 100^3]$ •each time step and $[10^3, 10^5]$ steps

•observables extracted via Fourier frequency analysis on density data

$$Q(\omega) = \sum_\sigma \int Q(\vec{x}, \sigma, t)\rho(\vec{x}, \sigma, t)e^{i\omega t}d^3x dt$$

•Computation is parallelized over index label n over quasi-particle wave functions

•Particle number, various densities, etc. conserved

•Partial densities are computed locally, reduced and broadcast

Cray XT4

QuadCore AMD Opteron (TM)	2.1e9 Hz clock	4 FP_OPs / cycle / core 128 bit registers
PEs	7832 nodes	31328 cpu-cores (processes)
Memory	• 8 GB / node • 2 MB shared L3 • 512 KB L2 / core • 64 KB D,I L1 / core	single socket nodes (667) 800 MHz DDR2 DIMM 3.2 GBps / core memory bw
Network	AMD HT SeaStar2+	3D torus topology 6 switch ports / SeaStar2+ chip 9.6 GBps interconnect bw / port
Operating Systems	Compute Node Linux	SuSE Linux on service / io nodes

Cray XT5

QuadCore AMD Opteron (TM)	2.3e9 Hz clock	4 FP_OPs / cycle / core 128 bit registers
PEs	19,200 (18,688)nodes	153,600 (149,504)cpu-cores
Memory	• 16 GB / node • 2 MB shared L3 / chip • 512 KB L2 / core • 64 KB D,I L1 / core	dual socket nodes 800 MHz DDR2 DIMM 3.2 GBps / core memory bw (25.6)
Network	AMD HT SeaStar2+	3D torus topology 6 switch ports / SeaStar2+ chip 9.6 GBps interconnect bw / port
Operating Systems	variant of Linux (xt-os2.1.50HD)	SuSE Linux on service / io nodes

Aggregated Features	XT4	XT5
Cycle Rate	65.7888 THz	343.85 THz
Memory	61.1875 TB	321.057 TB
Peak FLOP	263 TFLOPs	1.375 PFLOPs

Observation : a terascale / petascale supercomputer

- Aggregated Cycle rate :
  2.3e9 cycles / second / cpu-core * 8 cpu-core / node * 18,688 nodes ~ 343.85 THz
- Aggregated Memory :
  18,688 nodes * 16 * 2^30 BYTES ~ 321.057 TB
- Peak FLOP rate :
  343.8592 THz * 4 FP_OP / cycle ~ 1.375 PFLOPs !

WEAK SCALING EXAMPLE

increase number of processes to compute increased amount of work in same time - fix the amount of work per process

N^3	30^3	40^3	50^3
quasiparticles	28288	66796	130528
PEs	168	942	3626
[s] / 10 time steps	297.31	296.15	319.03
Total INS	1.41538E+14	7.87635E+14	3.13385E+15
Total FLOP	3.3787E+13	1.84227E+14	8.22772E+14
Total BYTES	5.37701E+11	3.00957E+12	1.14865E+13

t(50^3)/t(40^3) = 1.077	PE/PE=3.849	INS/INS=3.97	FLOP/FLOP=4.46
t(50^3)/t(30^3) = 1.073	PE/PE=21.583	INS/INS=22.14	FLOP/FLOP=24.35
t(40^3)/t(30^3) = .996	PE/PE=5.607	INS/INS=5.564	FLOP/FLOP=5.45

STRONG SCALING EXAMPLE :

increase number of processes to compute same amount of work in less time (perfect strong scaling reported in red)

$Nx^3 = 30^3$

PEs	576	1152	1728	2304
<NWF / PE>	48	24	16	12
[s]/ts	56.2	28.8	19.3	14.92 (14.05)

PEs (jaguar)	942 (x4)	1884	2826	3768
<NWF/PE>	70	36	24	17
[s] / 10 time steps	296.15	153.31	103.17	77.02 (74.03)
Total INS	7.87635E+14	8.15353E+14	8.15732E+14	8.10577E+14
Total FLOP	1.84227E+14	1.84997E+14	1.85766E+14	1.86536E+14

40^3

Total Memory Demands

Unitary(N^3)	Quasiparticles	BYTEs	BYTEs (OOC)
30^3	28288	5.37701E+11	44280000
40^3	66796	3.00957E+12	104960000
50^3	130528	1.14865E+13	205000000
60^3	226156	3.43902E+13	354240000
70^3	359056	8.6702E+13	562520000
80^3	535516	1.93026E+14	839680000
90^3	763824	3.92007E+14	1195560000
100^3	1046604	7.36809E+14	1640000000
110^3	1393008	1.30528E+15	2182840000
120^3	1808172	2.19966E+15	2833920000
130^3	2299056	3.55592E+15	3603080000

Nuclear (N^3)	QPs
40^3	267184
50^3	522112

E.g. 80^3 enables 535516 wfns on 13062 Cray XT5 NUMA nodes

Given Peta-scale Hardware Resources, Significant Problems are NOW Computable

```
#PBS -V
#PBS -l walltime=00:30:00,size=48000
#PBS -A csc053
#PBS -N papi-slda-nx70
#PBS -j oe
cd ${PBS_O_WORKDIR}
time aprun -n 6000 -N 1 -d 8 ./xslda 70
```

INIT RESULTS

PAPI_TOT_INS : Tot[470417708353027] Rt[80429863165]
PAPI_FP_INS : Tot[121453160131640] Rt[20013866944]
PAPI_L2_DCM : Tot[2757886476081] Rt[460049232]
PAPI_real_cyc = 97904168597 PAPI_real_usec = 42567030

TIME EVOLUTION OPERATOR RESULTS

PAPI_TOT_INS : Tot[7312266697827541] Rt[1279747745842]
PAPI_FP_INS : Tot[2878470206355248] Rt[481009289789]
PAPI_L2_DCM : Tot[15893496294521] Rt[2628441482]
PAPI_real_cyc = 1290061462680 PAPI_real_usec = 560896288
PAPI_user_cyc = 1290116000000 PAPI_user_usec = 560920000

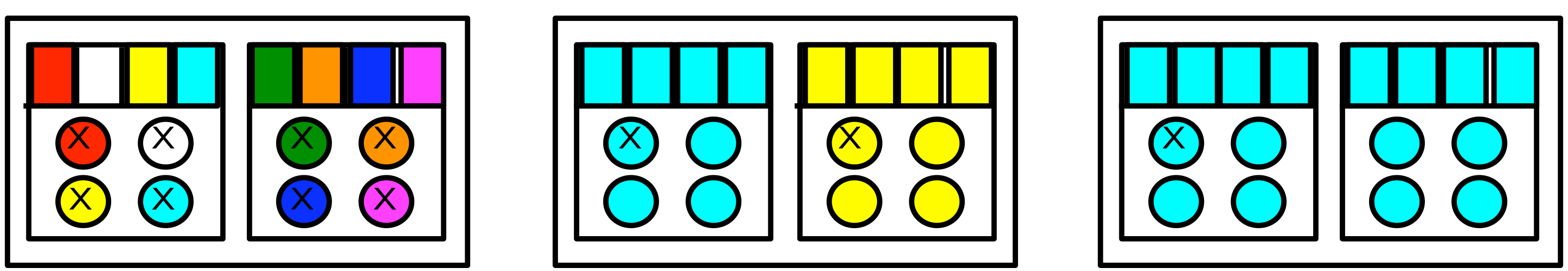
Real 10m14.046s

NUMA Node Structure of XT5 --> Hybrid Programming Model

- MPI processes spawn lightweight processes
- OpenMP threads, #include <omp.h> , omp_set_num_threads();
- POSIX threads, #include <pthread.h> , pthread_create();

-lsize=8	MPI	LWP	DRAM
aprun -n <1-8>	1 - 8	1	2 * 2^30
aprun -n 2 -sn 2 -S 1 -d 4	2	1 - 4	4 * 2^30
aprun -n 1 -N 1 -d 8	1	1 - 8	16 * 2^30

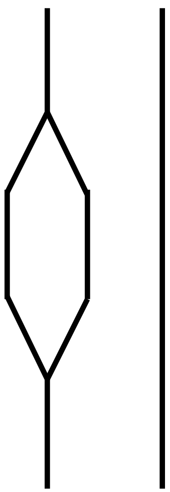
<-S> * <-d> cannot exceed the maximum number of CPUs per NUMA node



Works but no NUMA for user. ⟶ Memory Affinity

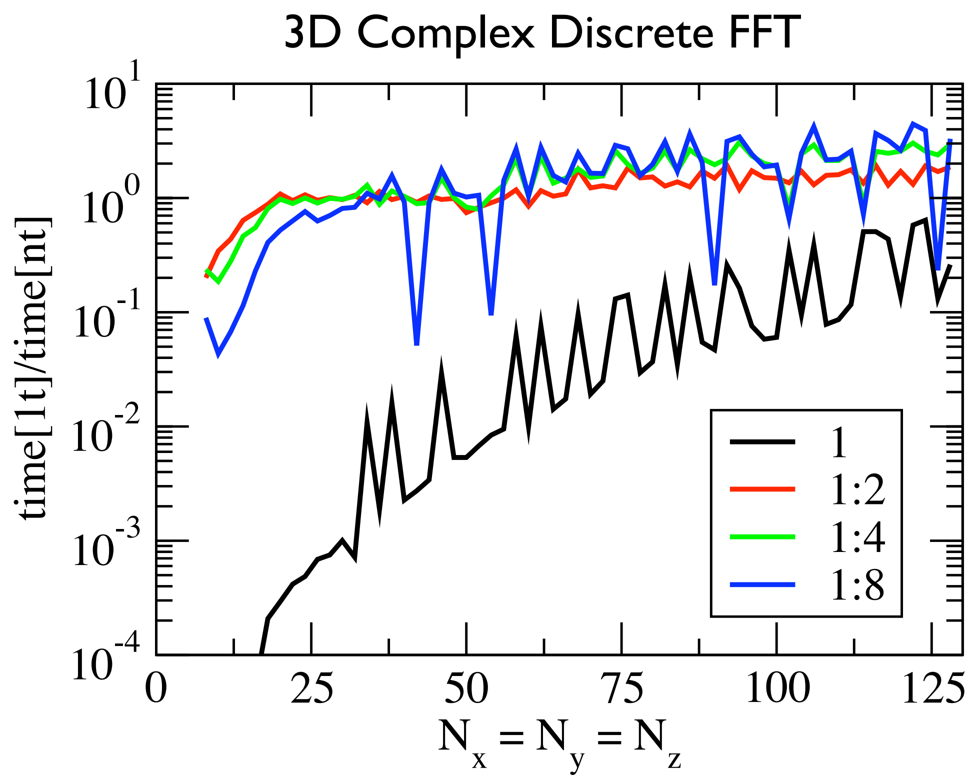
Exploiting Lightweight Processes on Multi-Core NUMA Nodes for SLDA

- Communication and I/O (MPI_THREAD_Init())
- Partial densities
- Discrete Fourier Transforms (3D complex)
- COSTS: mutex, condition variables, semaphores

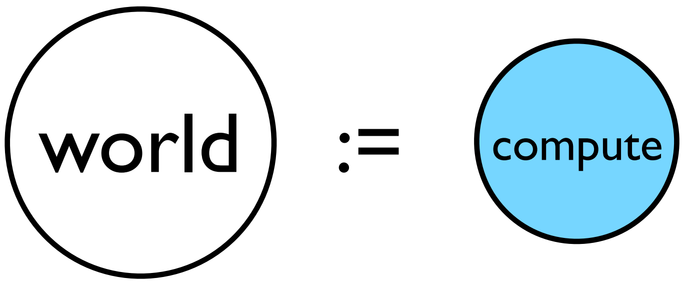


E.g. fork, mutex , join

NT	Cycles	L2DCM
1	1959379	69
2	2020818	81
4	2289393	122
6	2366367	146
8	2499159	239

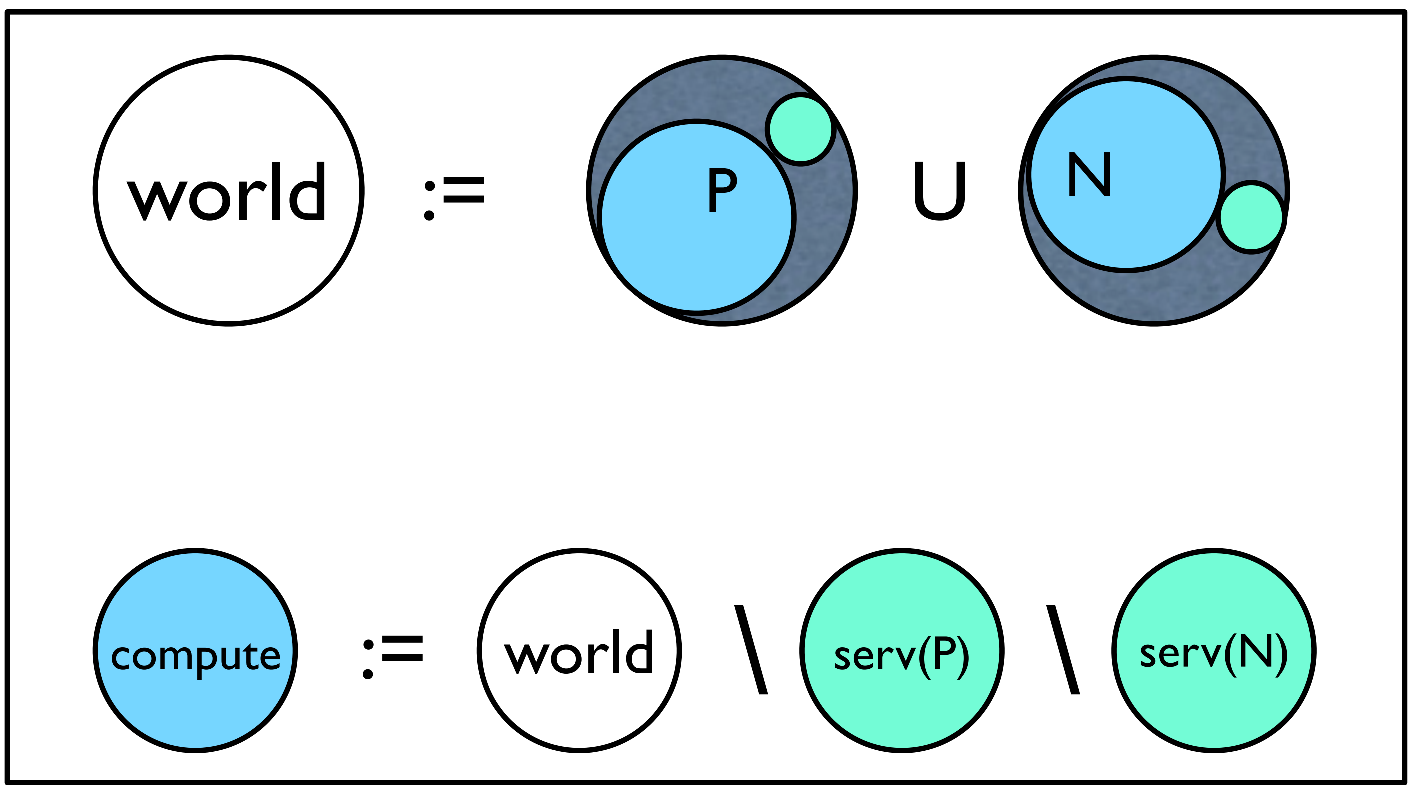


XT5, nts 5, 40^3	lsize=2048, -n 2048 := 2GB	lsize=2048, -n 512 -S1 -sn 2 := 8GB	lsize=2048, -n 256 -N 1 -d8 := 16GB
TIME	417.20 s	786.13 s	1587.56 s
FLOP	3.78546E+14	3.73964E+14	3.732E+14



Structure of unitary Fermi gas code

Nuclear problem is more complex



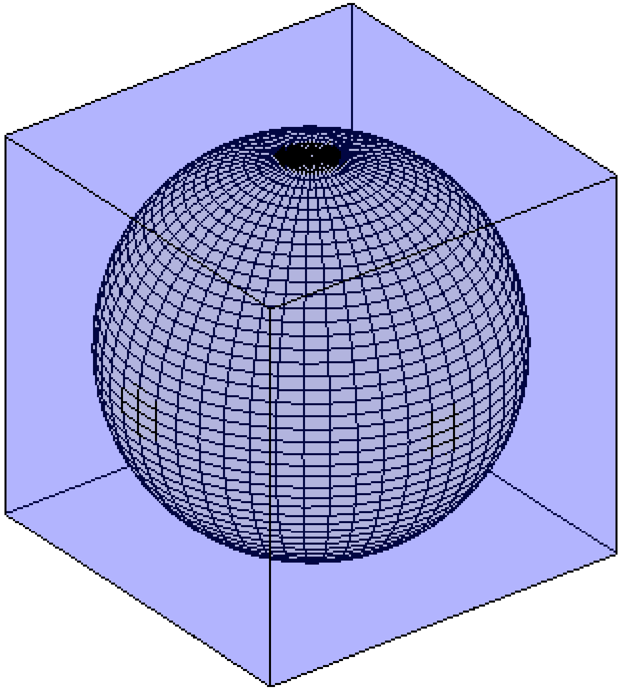
48Cr:
30^3 Lattice
1560 Wavefunctions
13000 Timesteps
1562 Tasks
471m 33s

8,386,386,255,460,901 FP_OP / 13000 TS

~ 6.45e+11 FP_OP / TS

Initial Conditions

- imported from HFBRAD
  - extensively tested
  - mapping issues
- additional terms



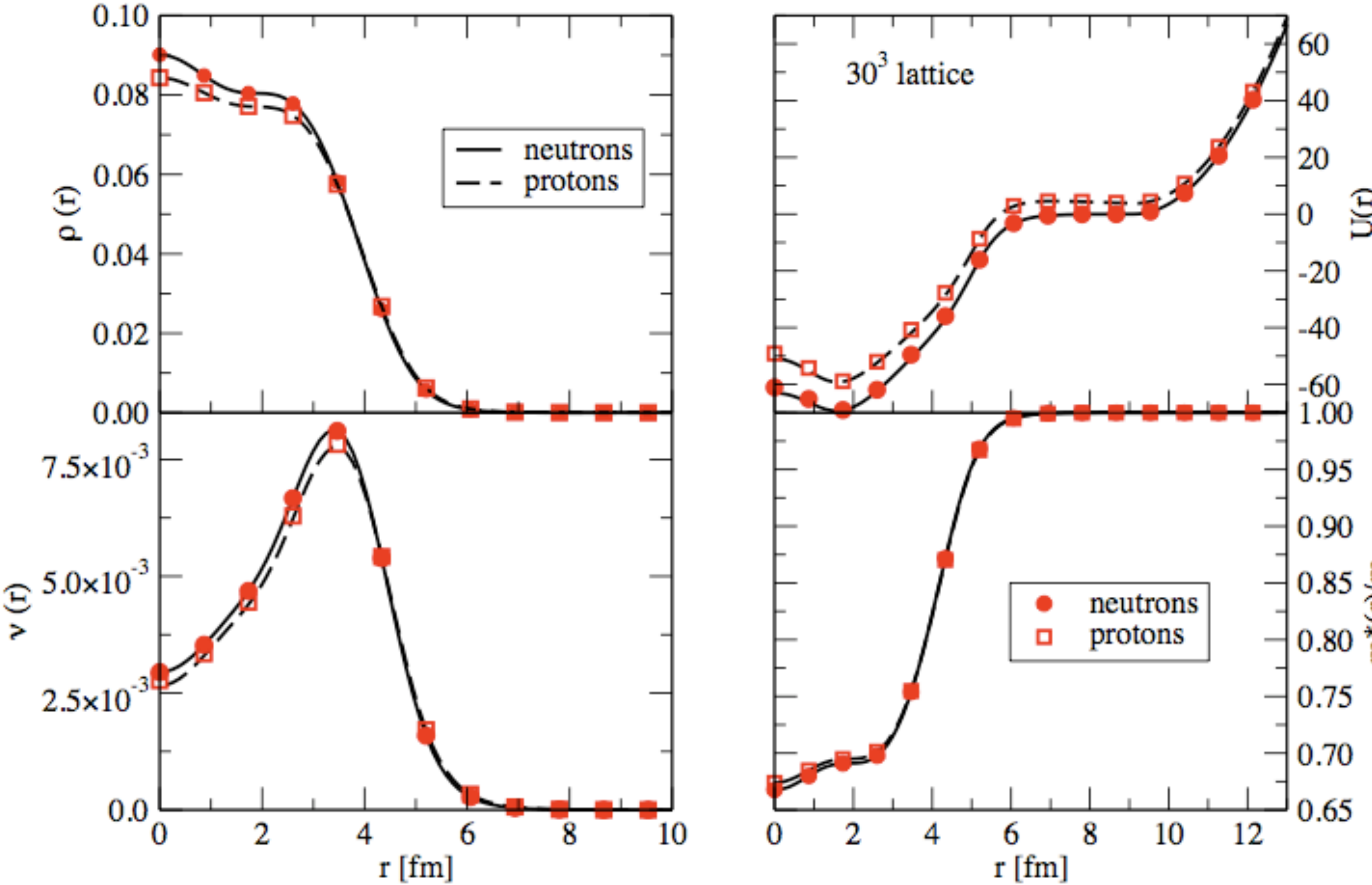
$$\mathcal{E}_T = C_T^{\rho}\rho_T^2 + C_T^{\Delta\rho}\rho_T\Delta\rho_T$$

$$+C_T^{\tau}(\rho_T\tau_T - \vec{j}_T^2) + C_T^{\nabla J}(\rho_T\vec{\nabla}\cdot\vec{J}_T + \vec{s}_T\cdot\nabla\times\vec{j}_T)$$

$$+C_T^{\gamma}\rho_0^{\gamma}\rho_T^2,$$

$$U(\vec{r}) \rightarrow U_0(\vec{r}) + \vec{U}(\vec{r})\cdot\vec{\sigma} - i\vec{V}(\vec{r})\cdot\vec{\nabla} - i\vec{W}(\vec{r})\cdot(\vec{\nabla}\times\vec{\sigma})$$

HFBRAD vs. TD-SLDA



First Application

48Cr

See the movies. http://www.phys.washington.edu/groups/qmbnt/index.html

Current Plan

Remainder of Current Year

- more benchmarking
- optimizations
    - multi-threading
- connection to 3D lattice solver
    - parallel implementation
- Coulomb excitation, simple nuclear reaction(s)

Year 4 + Beyond

- optimizations
- io for (re)start, data analysis
- scaling study
- description of collective excitations
- prepare the software for release

Summary

Fully capable of exploiting current and coming suite of massively parallel supercomputers

-demonstrated scaling behavior of TD code

Software can be leveraged easily by other fields

-cold atomic systems, neutron star crusts, condensed matter

New results

- vortex dynamics in superfluids, non-trivial nuclear process