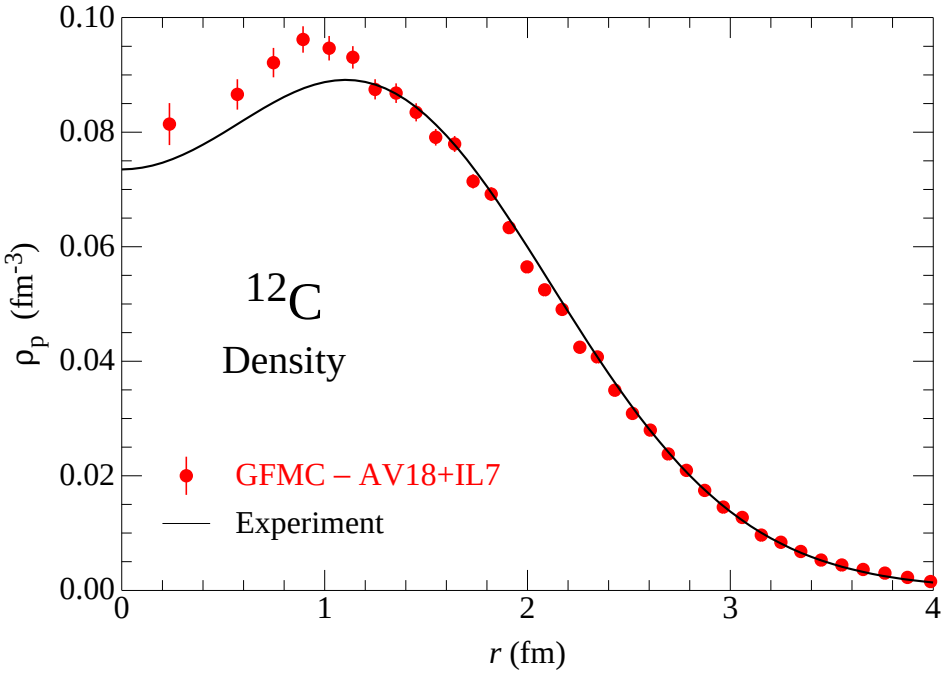
Tuning Green's Function Monte Carlo for Mira

Steven C. Pieper, Physics Division, Argonne National Laboratory

Partners in crime

Ralph Butler (Middle Tennessee State) Joseph Carlson (Los Alamos)
Stefano Gandolfi (Los Alamos) Alessandro Lovato (ANL)
E. L. (Rusty) Lusk (ANL) Saori Pastore (ANL & South Carolina)
Rocco Schiavilla (JLab & ODU) Robert B. Wiringa (Argonne)



U.S. DEPARTMENT OF ENERGY
Office of Science

Office of Nuclear Physics
Exploring Nuclear Matter - Quarks to Stars

Argonne NATIONAL LABORATORY
Physics Division

U.S. DEPARTMENT OF ENERGY
INCITE
LEADERSHIP COMPUTING

SciDAC
Scientific Discovery through Advanced Computing

FUNDING AND SOURCES OF COMPUTER TIME

- DOE INCITE grants of time on Argonne's Blue Gene computers

  The nuclear structure and reactions community has had joint INCITE awards during the last 6 years for time on Oak Ridge Crays and Argonne Blue Genes – currently James Vary is PI.

  2013: 15M core hours on BG/P, 50M on BG/Q

- DOE Early Science grant on Argonne's BG/Q

  Specifically for GFMC calculations of ^{12}C neutrino scattering – SCP is PI.

  2013: 110M core hours on BG/Q

  11/2012 – 11/2014 - Full support of postdoc: Alessandro Lovato

- Argonne LCRC (Fusion) - Many years; 300K+ hours in 2013

- SciDAC-II (UNEDF) & SciDAC-III (NUCLEI)

  Nation-wide collaborations to enable advanced computing solutions for nuclear structure and reactions. funds physicists, applied mathematicians, and computer scientists

  Joe Carlson is PI

  PHY currently gets $110K/yr

- Base program in nuclear theory

Ab Initio Few-Nucleon Calculations

Goal: a microscopic description of nuclear structure and reactions from bare NN & $3N$ forces.

There are two problems that must be solved to obtain this goal

(I) What is the Hamiltonian (i.e. the nuclear forces)?

- NN force controlled by NN scattering – lots of data available
  - Argonne v_{ij}
- $3N$ force determined from properties of light nuclei
  - Recent Illinois models with 2π & 3π rings

(II) Given H, solve the Schrödinger equation for A nucleons accurately.

- Essential for comparisons of models to data
- Quantum Monte Carlo has made much progress for $A \leq 12$
- Nuclei go up to A=238 and beyond!
  – less accurate approximations are used beyond 12

Without (II) comparison to experiment says nothing about (I).

The Many-body Problem

Need to solve

$$\mathcal{H}\Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_A; s_1, s_2, \cdots, s_A; t_1, t_2, \cdots, t_A)$$
$$= E\Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_A; s_1, s_2, \cdots, s_A; t_1, t_2, \cdots, t_A)$$

s_i are nucleon spins: $\pm\frac{1}{2}$

t_i are nucleon isospins (proton or neutron): $\pm\frac{1}{2}$

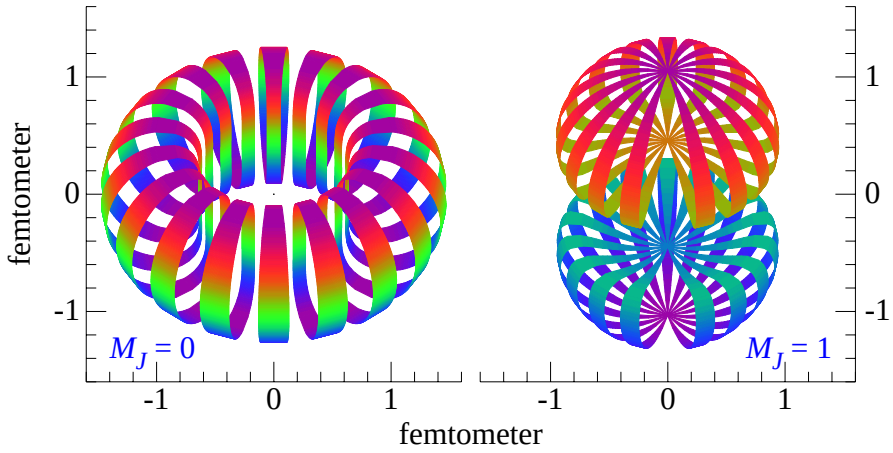
$2^A \times \binom{A}{Z}$ complex coupled 2^{nd} order eqn in $3A$ variables

(number of isospin states can be reduced)

^{12}C: 270,336 coupled equations in 36 variables

Coupling is strong:

- $\langle v_{\text{Tensor}} \rangle$ is $\sim 60\%$ of total $\langle v_{ij} \rangle$
- $\langle v_{\text{Tensor}} \rangle = 0$ if no tensor correlations



GREEN'S FUNCTION MONTE CARLO (GFMC)

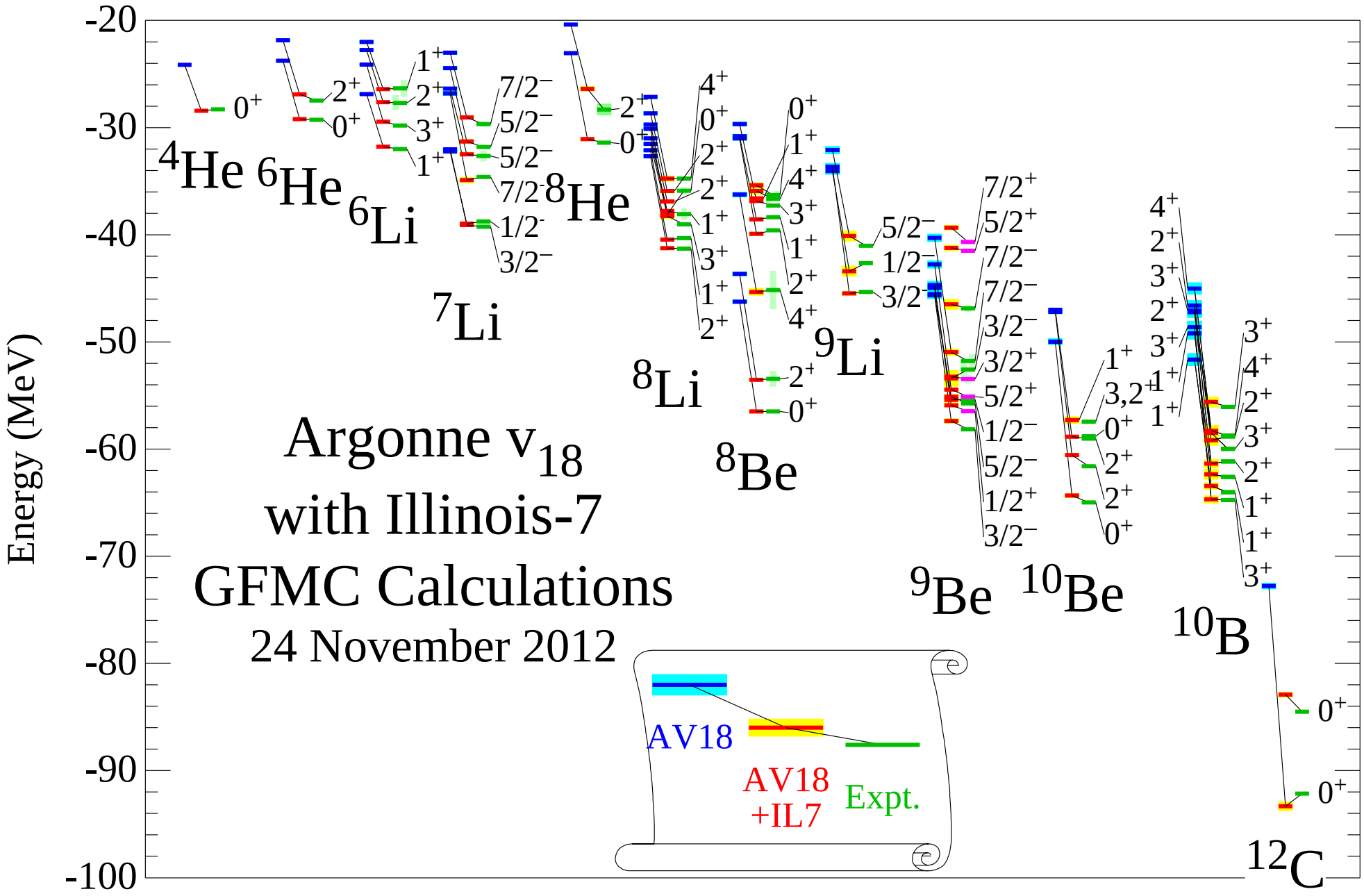
- Starts with an approximate wave function (Ψ_T) and evolves it to the exact Ψ for the given nuclear interaction (Hamiltonian)
- Evolution is done as a sequence of imaginary time steps
- Each time step is a 3×(number-of-nucleons) integral
- For ^{12}C, a ∼70,000-dimensional integral done by Monte Carlo
- Needs grow as $2^A \times \binom{A}{Z}$ for A nucleons with Z protons
- Works directly in coordinate space – no basis expansion
- Three-nucleon forces are not a significant complication
- Can find bound-state (exponentially decaying for $r \to \infty$) or scattering (asymptotically oscillating) solutions

OUTLINE OF GFMC CALCULATION

- A group of "walker" MPI ranks controls propagation of configurations
- These use the Asynchronous Dynamic Load Balancing (ADLB) library to get other ranks to do propagation steps and compute local energies.
- Wave functions or single propagation steps are done on single ranks
  – Most of the CPU time is for these operations
- Wave function vector has $\sim 2^A \times \begin{pmatrix} A \\ Z \end{pmatrix}$ complex numbers.
- CPU time dominated by complex sparse matrix $\times$ vector ops
  A) Matrix has Noncontiguous 4×4 blocks
  B) Structures of matrices same throughout calculation –
  multiply done by specialized table-driven code

Scaling of Ψ_T Calculation Time With Nucleus

	Pairs	Spin×Isospin	$\prod(/^8\mathrm{Be})$
^{4}He	6	8×2	0.002
^{6}Li	15	32×5	0.048
^{7}Li	21	128×14	0.75
^{8}Be	28	128×14	1.
^{8}Li	28	128×28	2.
^{9}Be	36	512×42	15.
^{10}B	45	512×42	19.
^{10}Be	45	512×90	41.
^{11}Li	55	2048×110	247.
^{12}C	66	2048×132	356. $\rightarrow$ 500.
^{16}O	120	32768×1430	112,065.
^{40}Ca	780	$3.6\times10^{21} \times 6.6\times10^{9}$	5.6×10^{19}


Energy (MeV)
-20
-30
-40
-50
-60
-70
-80
-90
-100
^{4}He
^{6}He
^{6}Li
^{7}Li
^{8}He
^{8}Li
^{8}Be
^{9}Li
^{9}Be
^{10}Be
^{10}B
^{12}C
Argonne v$_{18}$
with Illinois-7
GFMC Calculations
24 November 2012
AV18
AV18
+IL7
Expt.

MAKING GFMC WORK ON 131,072 PROCESSORS OF BG/P

GFMC needed to be redone for leadership class computers

- Old program did several Monte Carlo samples per processor
- Branching can kill samples – need enough not to fluctuate to zero
- ^{12}C has only $\sim$15,000 Monte Carlo samples
- Leadership class computers have many 10,000's processors
- Need to split one sample over many processors

Argonne's IBM Blue Gene/P

Automatic Dynamic Load Balancing (ADLB) for sharing work between nodes

- A general-purpose library to help application codes dynamically share work
- Developed by Rusty Lusk and Ralph Butler under UNEDF SciDAC
- GFMC was principal needs driver and test bed
- Good efficiency on 32,768 nodes (4 rows, 32 racks, 131,072 processors) of BG/P

OpenMP allows the 4 cores on one node to work together on one piece of work

- Full memory of node is used for just 1, not 4, tasks
- Efficiency is very good – 4 cores are $\sim$3.8 $\times$ faster than 1 core

ADLB is a general purpose library; give it a try! – http://www.cs.mtsu.edu/$\sim$rbutler/adlb

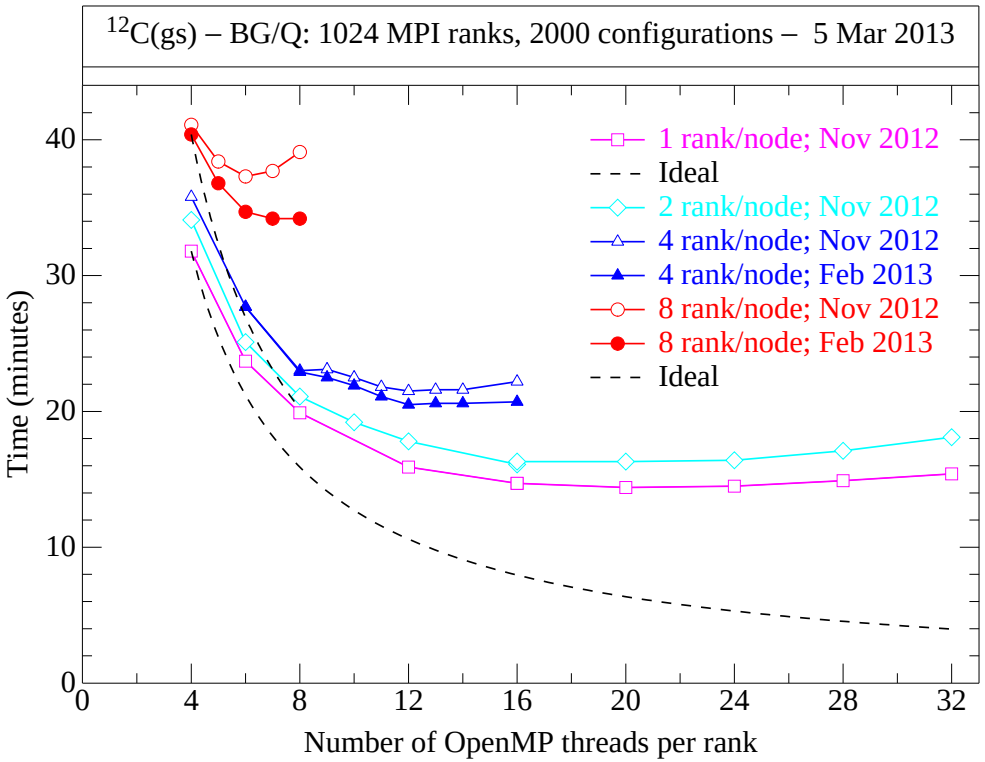
ADVANCING FROM THE IBM BG/P TO THE BG/Q

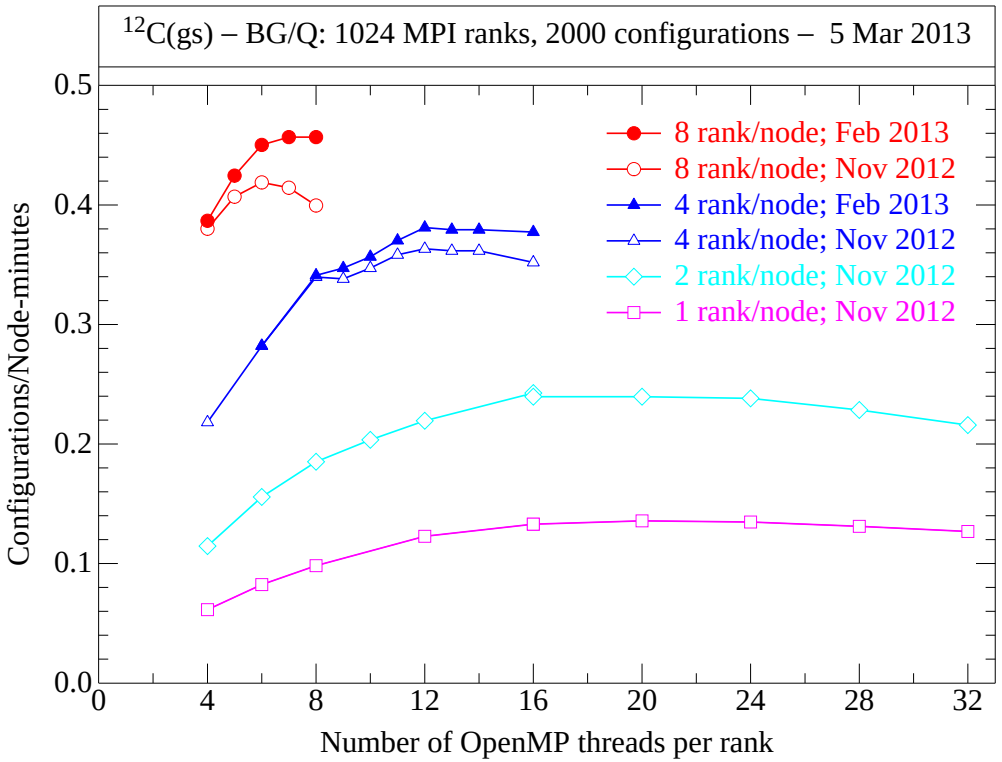
- BG/Q offers new possibilities and challenges
  - 16 Gbytes, 16 cores (each 4 threads) per node
  - 48 $\times$ 1024 nodes
  - ^{12}C(0^+): 8 ranks/node (8 threads each) or 4, 2, or 1 (64 threads)
  - Other ^{12}C states need much more memory/rank (T=1: 14 Gbytes)
- Conversion went very well
  - ADLB performance even better on BG/Q with no modifications!
  - OpenMP scales well to more threads

GFMC OpenMP Strong Scaling

$^{12}C(0^+)$ – 2000 configs for 40 time steps (2 energies) on 1024 ranks

- Time increases with more ranks/node
- For few ranks/node, optimal speed obtained with fewer than max. possible threads
  – Vectors too small for >20 threads
- Best overall use of nodes obtained with most ranks/node
- 8 ranks/node, 8 threads/rank: 6.4 GFLOPS/node = 3.1% peak

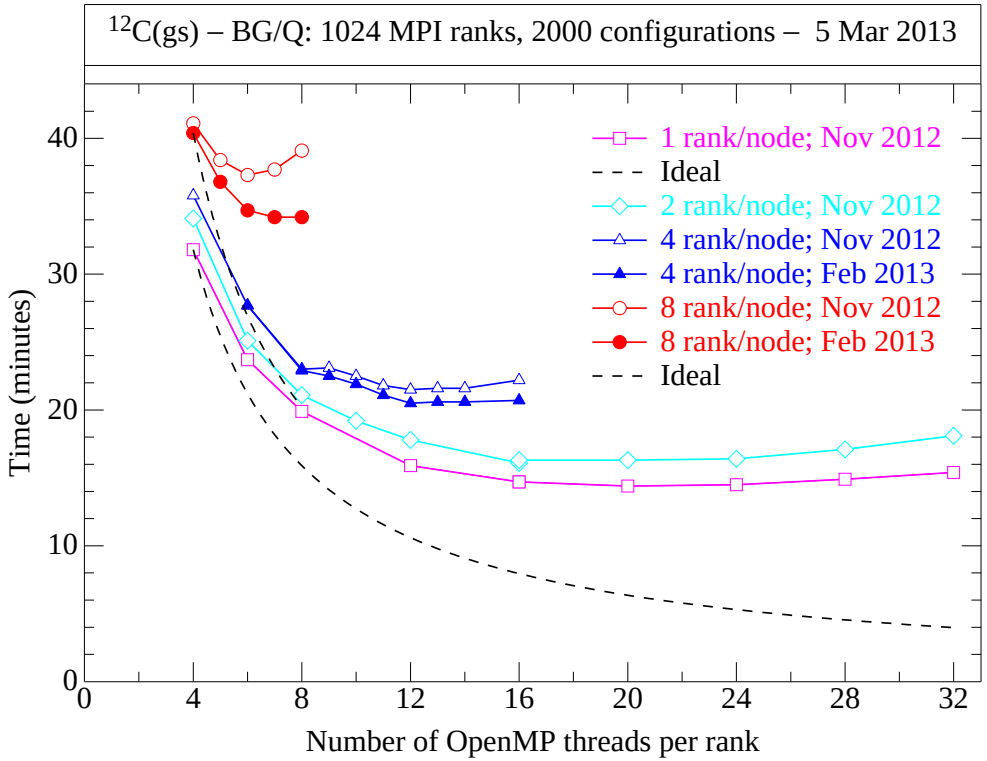


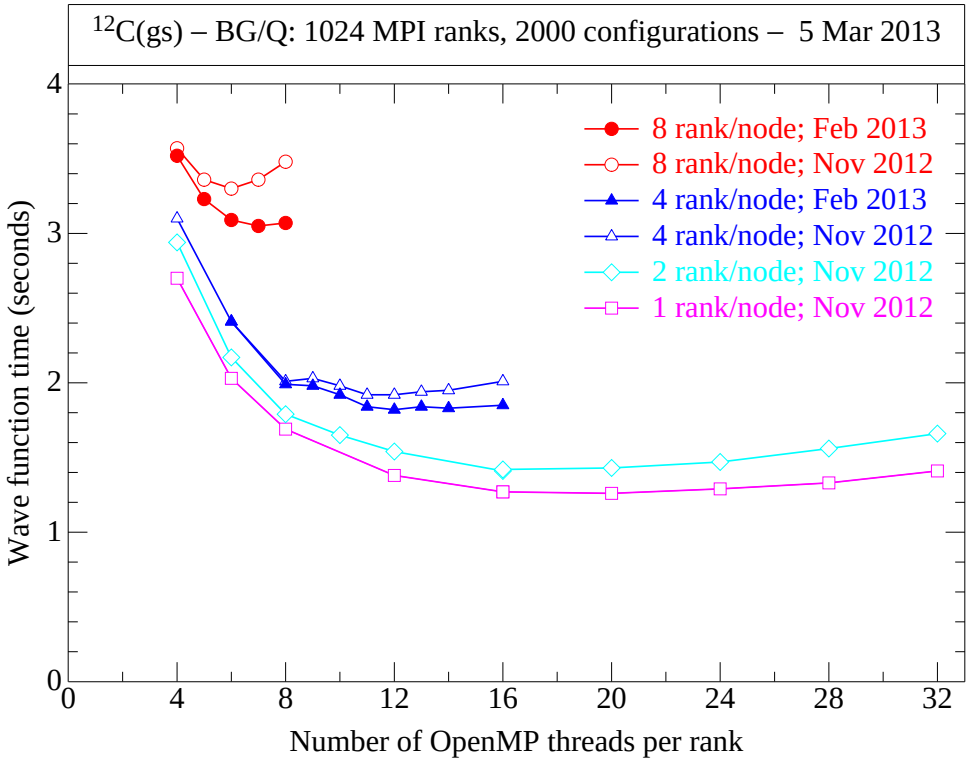


GFMC OpenMP Strong Scaling

$^{12}C(0^+)$ – 2000 configs for 40 time steps (2 energies) on 1024 ranks

- Timing of wave function subroutine shows improved performance is single-rank improvement – Not a MPI or ADLB effect

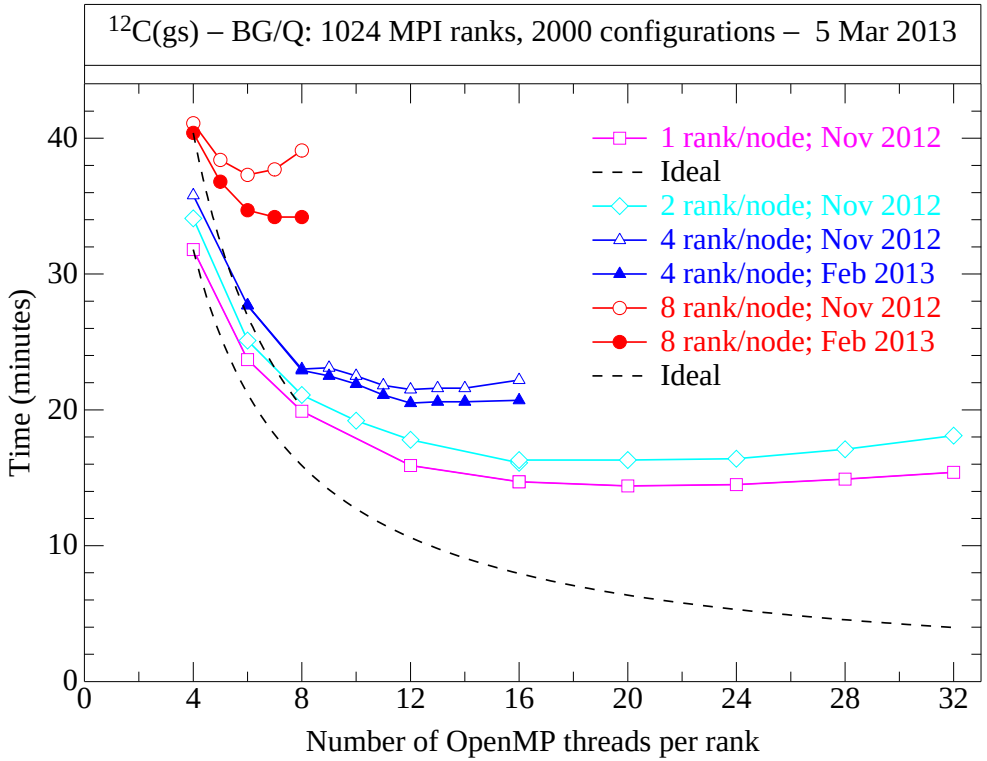


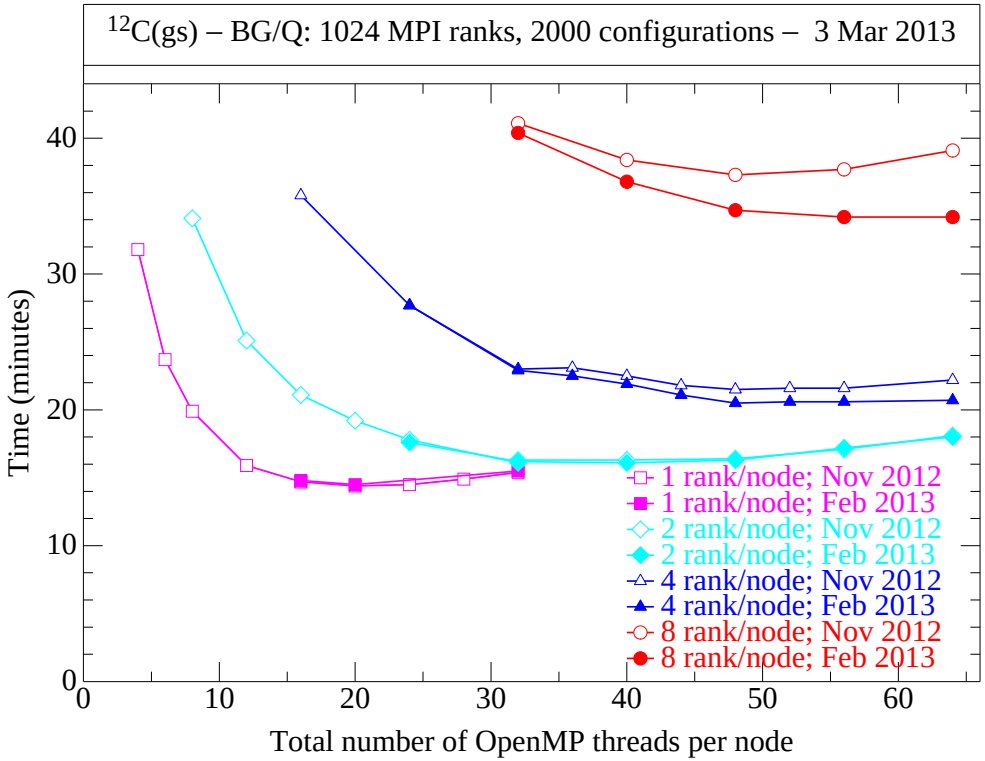


GFMC OpenMP Strong Scaling

$^{12}C(0^+)$ – 2000 configs for 40 time steps (2 energies) on 1024 ranks

- February, 2013, driver changes give better OpenMP performance
  – Not a compiler change; still same November, 2012 compiler
- Performance improvement only when many ranks have many threads



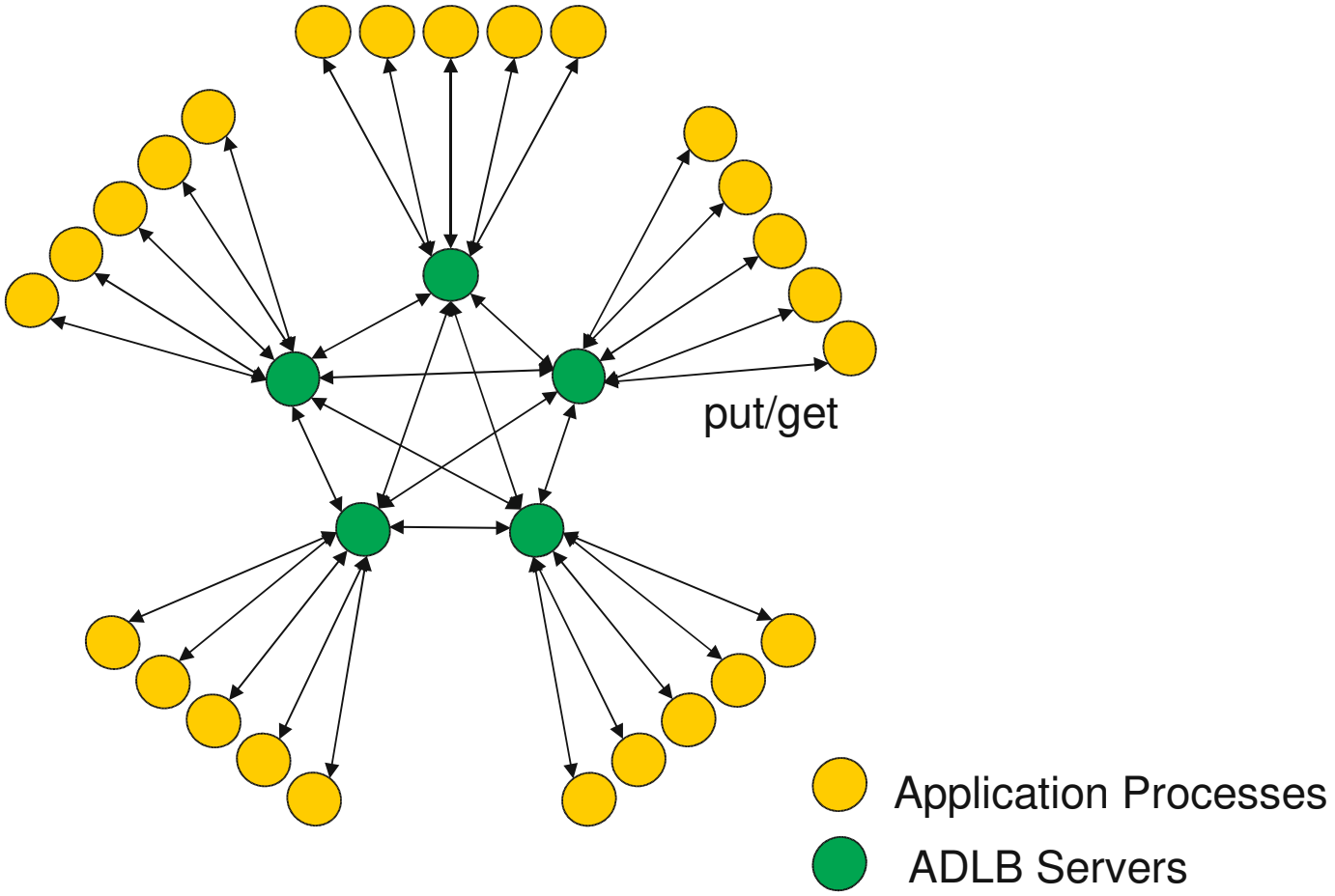


AUTOMATIC DYNAMIC LOAD BALANCING – THE VISION

Developed by Rusty Lusk and Ralph Butler

- Explicit master not needed:
  - Slaves make calls to ADLB library to off-load or get work
  - ADLB accesses local and remote data structures (remote ones via MPI)
- Simple Put/Get interface for application code hides most MPI calls
  - Advantage: multiple applications may benefit
  - Wrinkle: variable-size work units introduce some complexity in memory management
- Proactive load balancing in background
  - Advantage: application never delayed by search for work from other slaves
  - Wrinkle: scalable work-stealing algorithms not obvious
- Some nodes ($\sim$3% for GFMC) are ADLB servers – do no calculating

AUTOMATIC DYNAMIC LOAD BALANCING – WORK FLOW



ASYNCHRONOUS DYNAMIC LOAD BALANCING – THE API

- Startup and termination
  - ADLB_Init(num_servers, am_server, app_communicator)
  - ADLB_Server()
  - ADLB_Set_No_More_Work()
  - ADLB_Finalize()
- Putting work or answers
  - ADLB_Begin_Batch_Put(common_buffer, length) – optional
  - ADLB_Put(type, priority, length, buffer, answer_destination)
  - ADLB_End_Batch_Put() – optional
- Getting work or answers
  - ADLB_Reserve(req_types, work_handle, length, type, priority, answer_destination)
  - or ADLB_Ireserve($\cdots$)
  - ADLB_Get_Reserved(work_handle, buffer)

ADLB – CURRENT GFMC IMPLEMENTATION

Old GFMC

Each slave gets several configurations

Slave

propagates configurations

(few w.f. evaluations)

replicates or kills configs (branching)

$\rightarrow$ periodic global redistribution

computes energies

(many w.f. evaluations)

Need $\sim$10 configs per slave

^{12}C will have only $\sim$10,000 configs.

Can't do on more than 2000 processors

Configurations cannot be unit of parallelization

With ADLB

A few "boss" slaves manage the propagation:

- Generate propagation work packages
  – Answers used to make 0,1,2,$\cdots$ new propagation packages (branching)
  – Number of prop. packages fluctuates
  – Global redistribution may be avoided
- Generate energy packages – No answers

When propagation done, become worker slaves

Most slaves ask ADLB for work packages:

- Propagation package
  – Makes w.f. and $3N$ potential packages
- Energy package
  – Makes many w.f. packages
  – Makes $3N$ potential packages
  – Result sent to Master for averaging
- Wave Function or $3N$ potential package
  – Result sent to requester

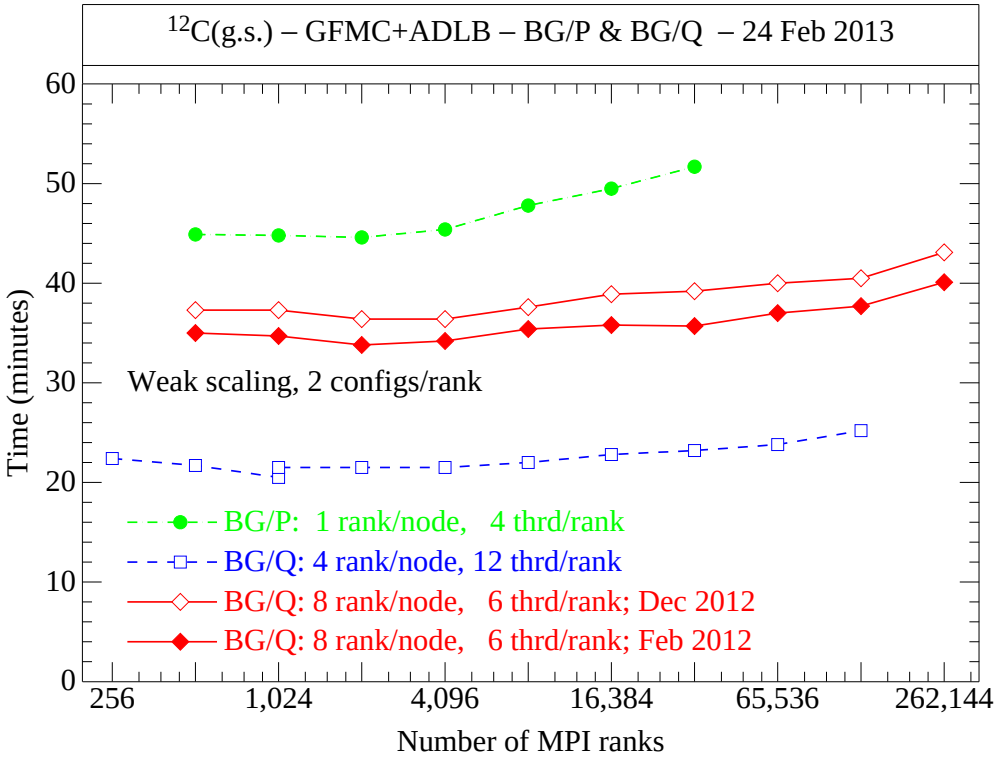
Wave function is parallelization unit

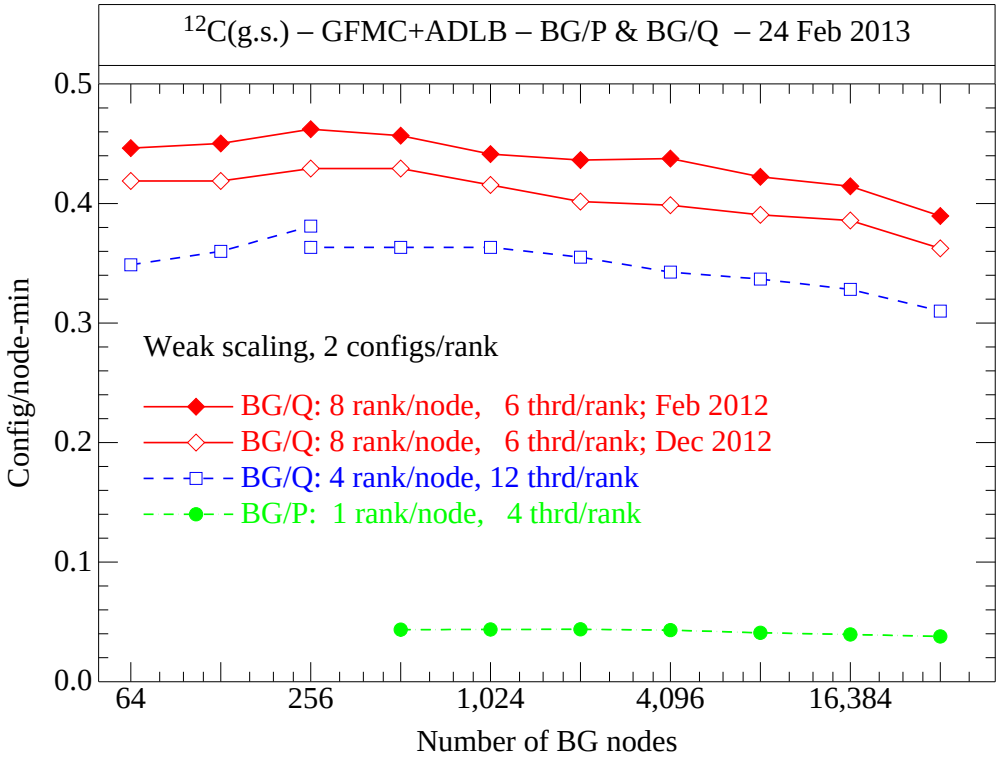
Can have many more nodes than configurations

GFMC ADLB Weak Scaling

$^{12}C(0^+)$ – 2 configs/rank for 40 time steps (2 energies)

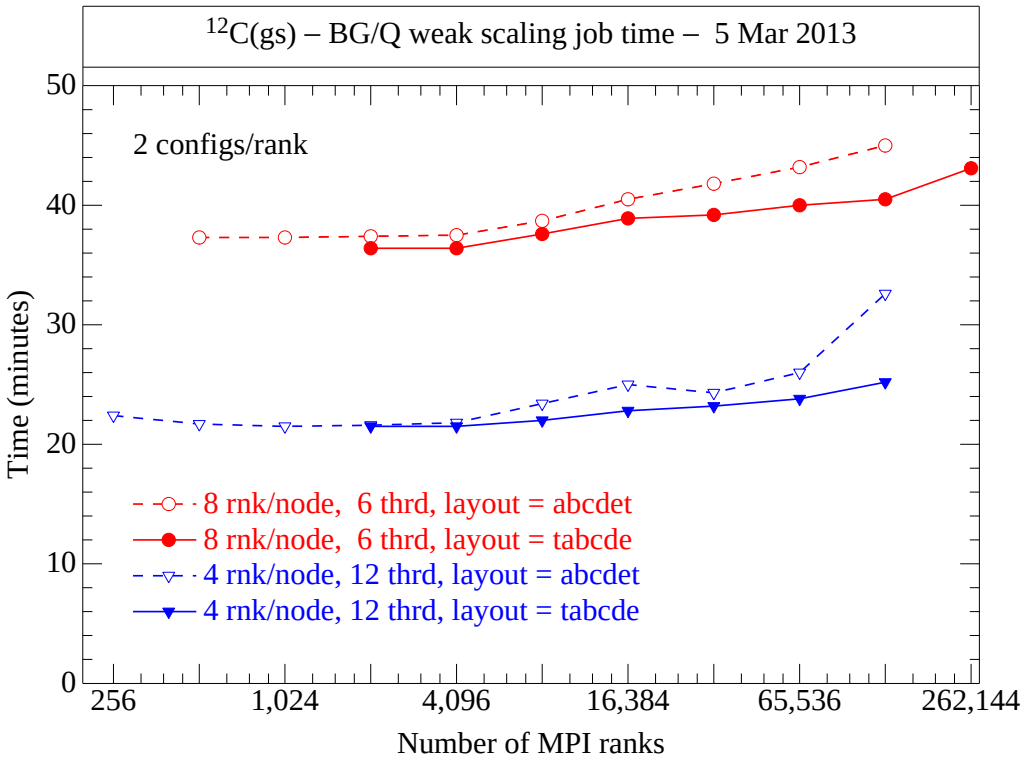
- Best overall use of nodes obtained with most ranks/node
- Good scaling to 262,000 ranks – 524,288 cores – 1,572,864 threads!
- BG/Q node performance 10 × BG/P node





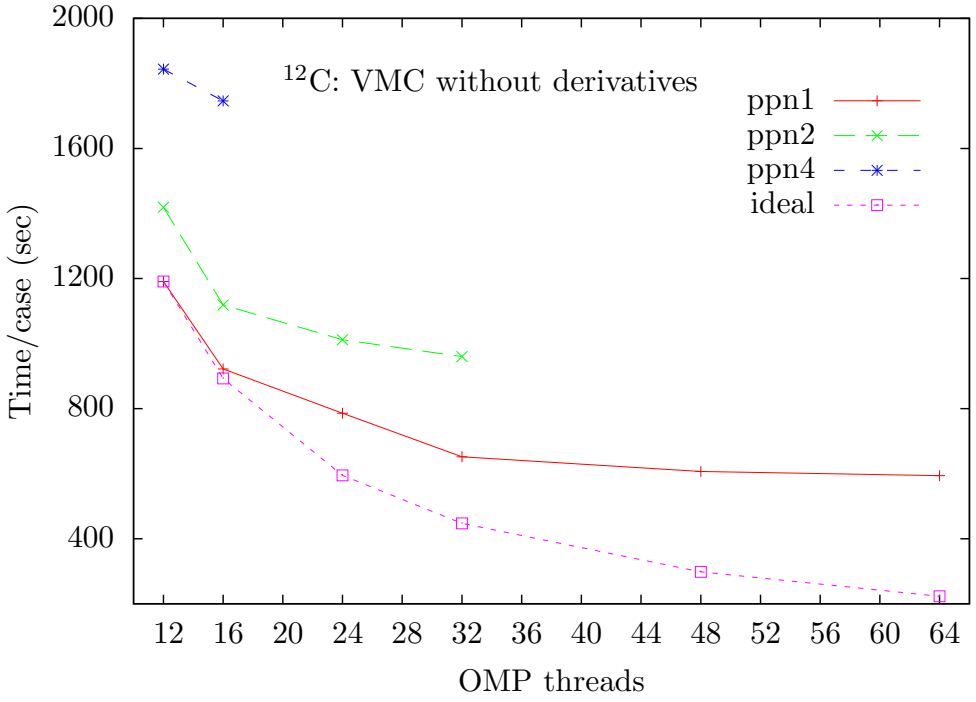
GFMC ADLB Weak Scaling - Layout of Ranks on Nodes

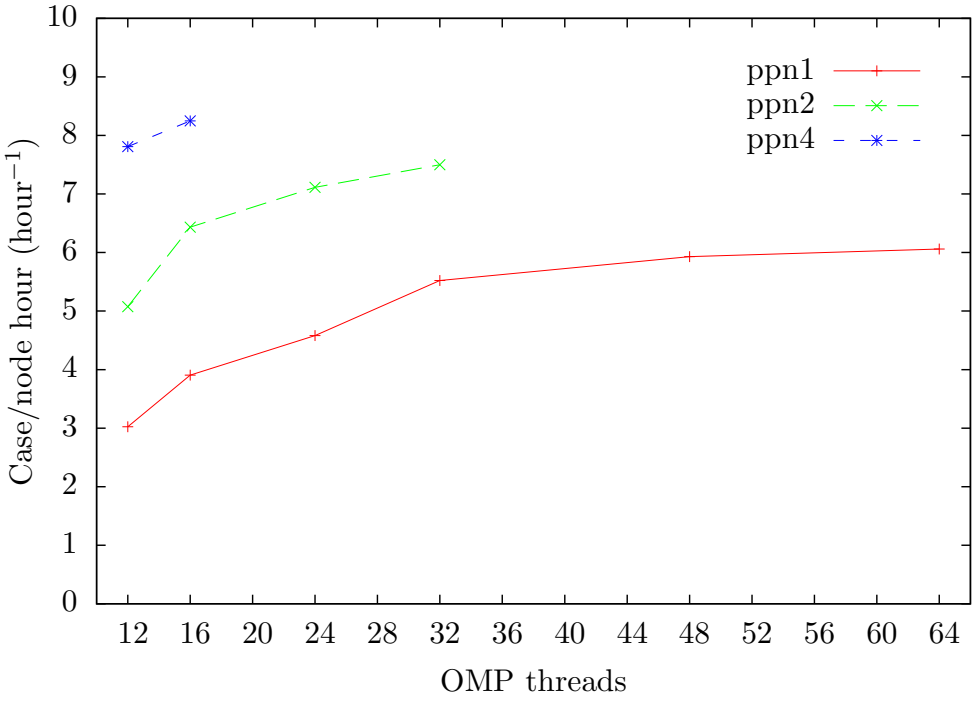
- GFMC “walker” ranks are the lowest numbered ranks
- ADLB servers are the highest numbered ranks
- Having several of these on one node puts heavy communication on that node
- tabcde layout is significantly better than default abcdet



SCALING FOR ^{12}C(GS) ELECTROMAGNETIC RESPONSE

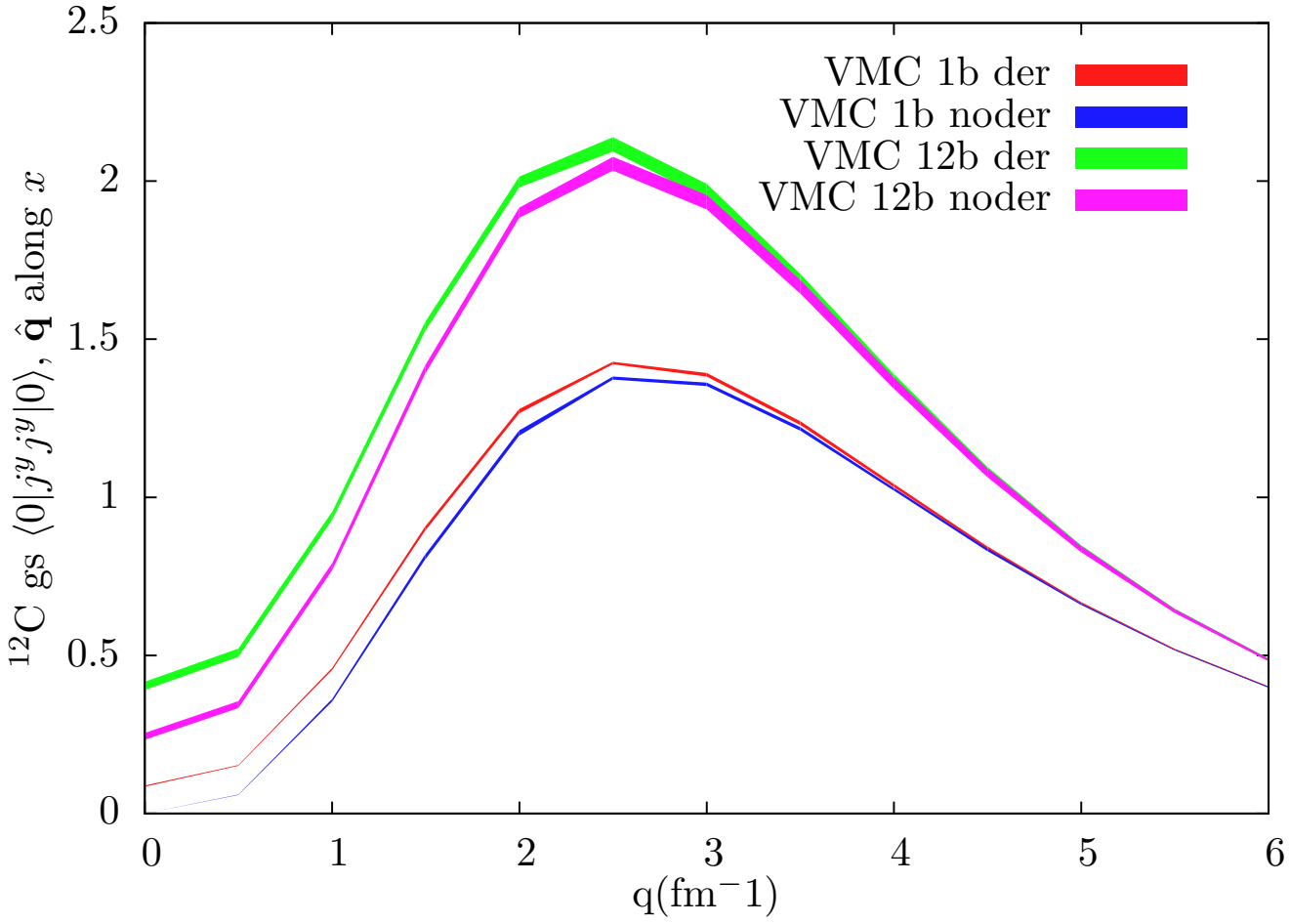
- Large new collection of subroutines previously used only for small nuclei
- Alessandro Lovato has greatly improved OpenMP aspects for ^{12}C – 2× faster
- Two modes – without and with derivatives
  –Can use 4 ranks/node without derivatives but only 1 rank/node with derivatives
- OpenMP keeps improving all the way to 64 threads, but barely
- 4 ranks/node, 16 threads/rank: 12. GFLOPS/node = 5.8% peak
- Impossible calculation on BG/P





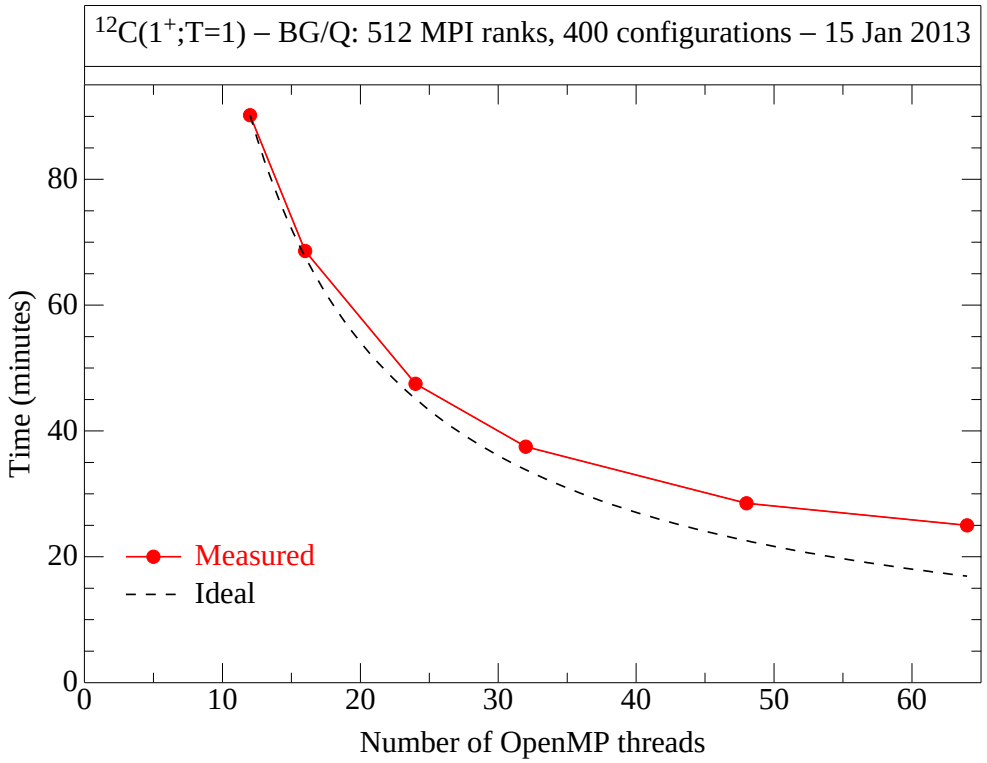
^{12}C(gs) Electromagnetic Response – First results

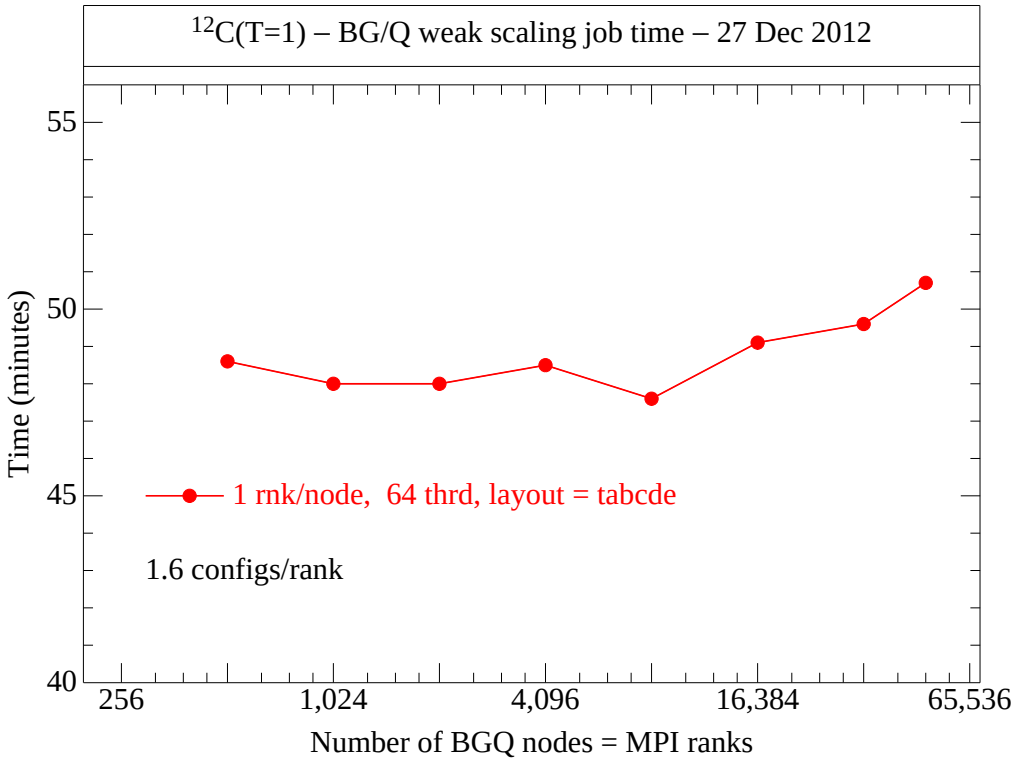
- VMC only; GFMC still to be done
- Two-body currents have large effect!
- Jefferson Lab experiment nearing publication – we want to predict, not postdict, their results



SCALING FOR $^{12}C(1^+;T=1)$

- Wave function calculation is much more involved; size is 2.25× previous
- 14 Gbytes per rank – only one rank/node possible
- OpenMP keeps improving all the way to 64 threads
- Good scaling to full machine (48 × 1024 nodes or ranks)
- Actual performance sorta poor: 64 threads give 1.0 GFLOPS/node = .5% peak
- Impossible calculation on BG/P

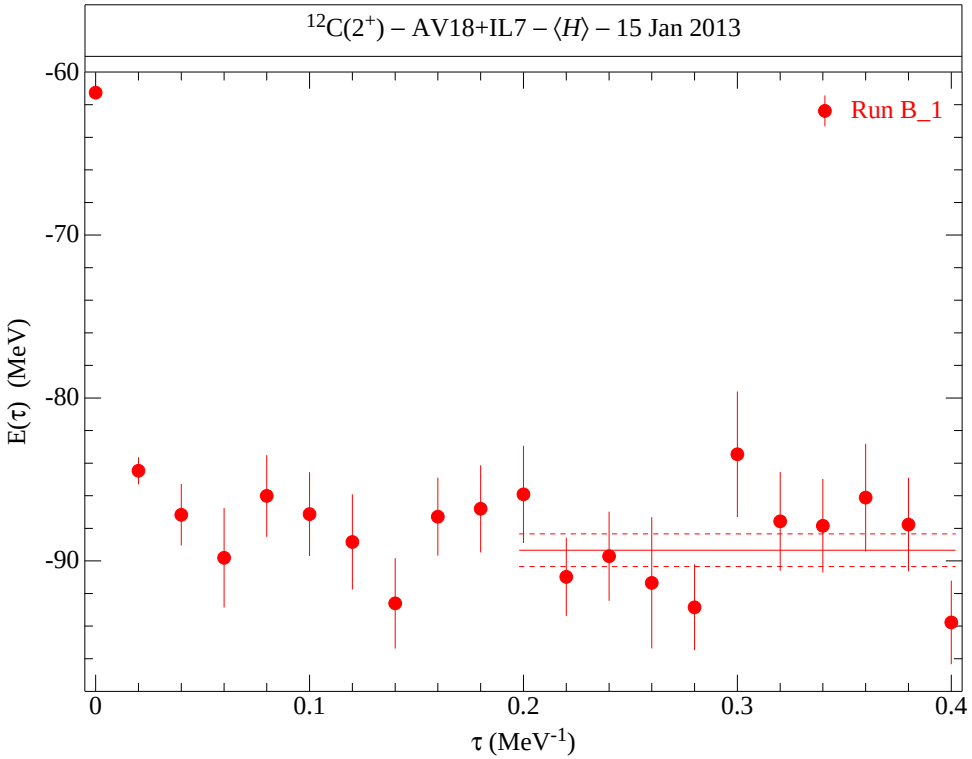


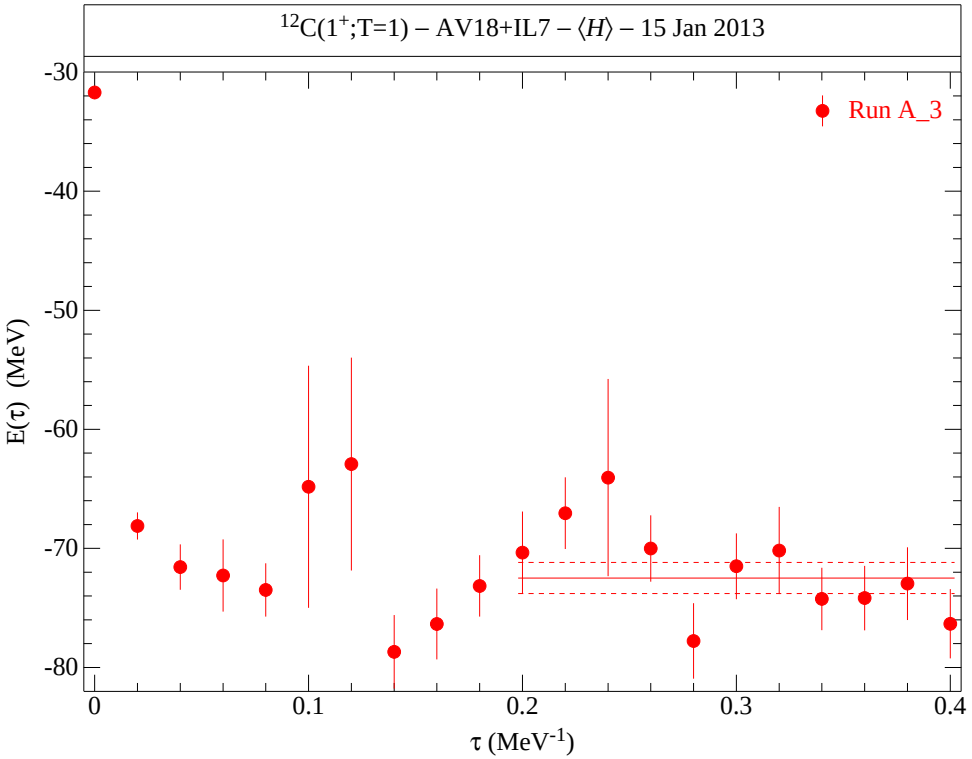


MIRA ENABLES CALCULATION OF ^{12}C STATES WITH SPIN > 0

- Have made initial VMC Ψ_T and GFMC propagations for
  - ^{12}C(2^+), E^* = 3.9(1.0) vs Expt. = 4.44
  - ^{12}C(1^+;1), E^* = 21.6(1.3) vs Expt. = 15.11

- Not possible on BG/P because of large (up to 14 Gbytes/rank) memory needs





CONCLUSIONS & FUTURE

Conversion of GFMC code to BG/Q has been quite successful

- OpenMP performs well and very well to 64 threads when we need it
- ADLB library with OpenMP allows efficient use of $>$ 100,000 processors for GFMC
- Calculations of ^{12}C not possible on BG/P now started

and there is still much to do

- EM response of ^{12}C (recent JLAB expt)
- Neutrino scattering on ^{12}C and weak response
- 2^+ $E2$ form factor
- other ^{12}C states,

ADLB is a general purpose library; give it a try! – http://www.cs.mtsu.edu/∼rbutler/adlb

If sequestration becomes really bad

To Learn More

Pointers to the following are at http://www.phy.anl.gov/theory/staff/SCP.html & RBW.html

- *Nucleon-nucleon interactions*, R. B. Wiringa, in Contemporary Nuclear Shell Models, ed. X.-W. Pan, D. H. Feng, and M. Vallières (Springer-Verlag, Berlin, 1997)
- *Monte Carlo calculations of nuclei*, S. C. Pieper, in Microscopic Quantum Many-Body Theories and Their Applications, ed. J. Navarro and A. Polls, Lecture Notes in Physics **510** (Springer-Verlag, Berlin, 1998)
- *Quantum Monte Carlo Calculations of Light Nuclei*, S. C. Pieper and R. B. Wiringa, Annu. Rev. Nucl. Part. Sci. **51**, 53-90 (2001)
- *Quantum Monte Carlo Calculations of Light Nuclei*, S. C. Pieper, in Proceedings of the "Enrico Fermi" Summer School, Course CLXIX, ed. A. Covello, F. Iachello, and R. A. Ricci (Societ Italiana di Fisica, Bologna, 2008); arXiv:0711.1500 [nucl-th]
- A simplified VMC program and description: *Variational Monte-Carlo Techniques in Nuclear Physics*, J. A. Carlson and R. B. Wiringa, Computational Nuclear Physics 1, ed. K. Langanke, J. A. Maruhn, and S. E. Koonin (Springer-Verlag, Berlin, 1990), Ch. 9
  source & input files available at http://www.phy.anl.gov/theory/research/vmc-demo
- ADLB load-balancing library is at http://www.cs.mtsu.edu/∼rbutler/adlb

BIBLIOGRAPHY, CONTINUED

Detailed descriptions of the potentials

- *Accurate nucleon-nucleon potential with charge-independence breaking*, R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, Phys. Rev. C **51**, 38-51 (1995)
- *Realistic models of pion-exchange three-nucleon interactions* Steven C. Pieper, V. R. Pandharipande, R. B. Wiringa, and J. Carlson, Phys. Rev. C **64**, 014001-1:21 (2001)

Detailed descriptions of VMC and GFMC methods and many results

- *Quantum Monte Carlo calculations of nuclei with* $A \leq 7$, B. S. Pudliner, V. R. Pandharipande, J. Carlson, S. C. Pieper, and R. B. Wiringa, Phys. Rev. C **56**, 1720-1750 (1997)
- *Quantum Monte Carlo calculations of A=8 nuclei*, R. B. Wiringa, Steven C. Pieper, J. Carlson, and V. R. Pandharipande, Phys. Rev. C **62**, 014001-1:23 (2000).
- *Quantum Monte Carlo calculations of A=9,10 nuclei*, Steven C. Pieper, K. Varga, and R. B. Wiringa, Phys. Rev. C **66**, 044310-1:14 (2002).
- *Quantum Monte Carlo Calculations of Neutron-alpha Scattering*, K.M. Nollett, S.C. Pieper, R.B. Wiringa, J. Carlson, G. M. Hale, Phys. Rev. Lett. **99**, 022502 (2007)

SCALING FOR ^{12}C(GS) ELECTROMAGNETIC RESPONSE

- With derivatives, OpenMP keeps improving all the way to 64 threads, but barely
- 64 threads/rank: XXXX GFLOPS/node = XXXX% peak
- Impossible calculation on BG/P