Ab initio calculations of the structure and reactions of light nuclei with MFDn on current HPC platforms

Pieter Maris

Dept. of Physics and Astronomy
Iowa State University
Ames, IA 50011

supported by US DOE under grant DE-SC0023495 (SciDAC5-NUCLEI)

NTSE2024, December 1-6, 2024
Busan, South Korea

Outline

- What do we mean by Supercomputing Era ?
- Many Fermion Dynamics for nuclear physics
- Recent physics results for p-shell nuclei
- Conclusing remarks

Scientific Computing in the USA

DOE Computer facilities for scientific computing

- National Energy Research Scientific Computing center (NERSC)
  - located at LBNL
  - about 9,000 users
  - many different small and medium-size projects
  - most computing time allocated to DOE Office of Science
    - 12.5% Nuclear Physics (2022)
    - 74 different Nuclear Physics projects (2022)
- Oakridge Leadership Computing Facility (OLCF)
  - relatively small user base
  - limited number of large projects
- Argonne Leadership Computing Facility (ALCF)
  - relatively small user base
  - limited number of large projects

also NSF Computer facilities,
as well as local university computing resources

Supercomputing Era from a personal perspective

- NERSC
  - Seaborg (2002-2008)
    IBM SP, 416 nodes, 6,656 cores, 6,656 GB
  - Franklin (2009-2012)
    Cray XT4, 9,572 nodes, 38,288 cores, 76,576 GB
  - Hopper (2010-2015)
    Cray XE6, 6,384 nodes, 153,216 cores, 216,832 GB
  - Edison (2013-2019)
    Cray XC30, 5,200 nodes, 124,800 cores, 332,800 GB
  - Cori-KNL (2016-2023)
    Cray XC40, 9,688 nodes, 658,784 cores, 893 TB
  - Perlmutter-GPU (2022-), NVIDIA GPUs
    HPE Cray EX, 1,792 GPU nodes, 768 TB
  - Perlmutter (2022-):
    HPE Cray EX, 3,072 CPU nodes, 1,536 TB
  - NERSC-10 planned for 2026

OLCF and ALCF systems

- OLCF
  - Jaguar (2008-2012) Cray XT5, 37,538 nodes, 150,152 cores
    - first supercomputer to run scientific application at petaflop rate
    - nuclear theory early science project: ^{14}C$\rightarrow$ ^{14}N β decay
  - ...
  - Frontier (2023-), AMD GPUs, 9,402 nodes, 4.5 PB
- ALCF
  - Intrepid (2009-2013), IBM Blue Gene/P
  - Mira (2013-2019), IBM Blue Gene/Q
    48 racks, 49,152 nodes, 786,432 cores, 768 TB
  - Theta (2017-2023), Cray XC40
    4.2k nodes, 270k cores, 874 TB
  - Aurora (2024-), 63,744 Intel Data Center GPUs, $>$ 7 PB

HPC is entering the Exascale computing era !

Is Nuclear Theory ready for Exascale computing ?

- Need to expose several levels of parallelism
  - between nodes
  - within CPU nodes between cores (multi-threading)
  - on GPUs: many threads, more than one level parallelism
- Need to consider
  - load balancing
  - memory constraints
  - communication overhead
  - fault tolerance
- Portability NVIDIA GPUs, AMD GPUs, and Intel GPUs
  - including performance portability
- Inhomogeneous systems
  - can we make efficient use of both CPU and GPU resources?

Is the NCSM ready for Exascale computing ?

No-Core Configuration Interaction calculations

Barrett, Navrátil, Vary, *Ab initio no-core shell model*, PPNP69, 131 (2013)

- Expand A-body wave function in basis functions

$$\Psi(r_1, \ldots, r_A) = \sum a_i \Phi_i(r_1, \ldots, r_A)$$

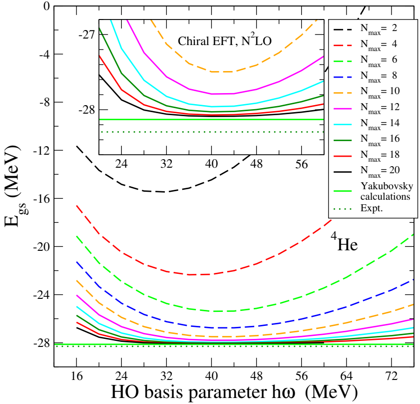
- Use basis of single Slater Determinants of single-particle states
- Use Harmonic Oscillator (HO) radial wavefunctions
  in combination with truncation on total number of HO quanta

$$\sum_{k=1}^{A} \left(2\, n_k + l_k\right) \quad \leq \quad N_0 + N_{\max}$$

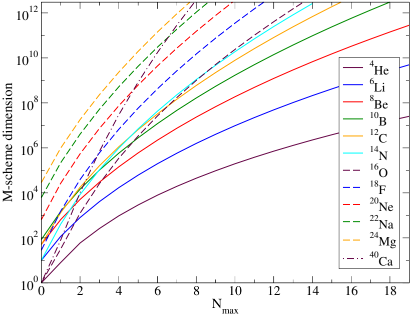
  to obtain exact factorization of Center-of-Mass motion
- Solve the eigenvalue problem $\hat{\mathbf{H}}\, \Psi(r_1, \ldots, r_A) = E_b\, \Psi(r_1, \ldots, r_A)$
- Converges to exact results in the limit $N_{\max} \to \infty$
- Computational task
  - obtain lowest eigenvalues of large sparse matrix (iteratively)
  - most expensive step: repeated Sparse Matrix-Vector multiplication

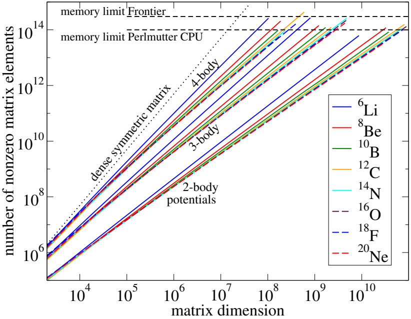
Convergence of basis expansion

- Variational: for any finite truncation of the infinitely-large basis, eigenvalue is an upper bound for the ground state energy
- Smooth approach to asymptotic value with increasing basis space
- Convergence: independence of both truncation paramere $N_{\rm max}$ and HO basis parameter $\hbar\omega$
  - different methods using the same interaction give same results within numerical uncertainties



Computational Challenge





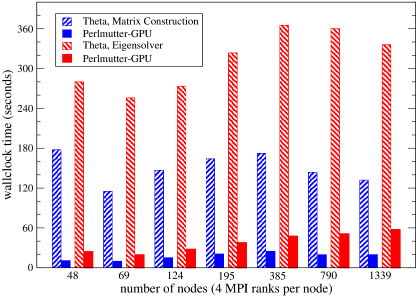
- Increase of basis space dimension with increasing A and $N_{\max}$
  - need calculations up to at least $N_{\max} = 8$, preferably $N_{\max} = 10$ or 12 for meaningful extrapolation and numerical error estimates
- Largest calculation with NN-only potentials: ^{12}Be at $N_{\max} = 12$
  - dimension of 35 billion, with 60 trillion nonzero matrix elements
  - needed all of Theta at ALCF (or all of Cori-KNL at NERSC)
  - in queue on Perlmutter: ^{13}B at $N_{\max} = 13$, with 83 trillion nonzeros

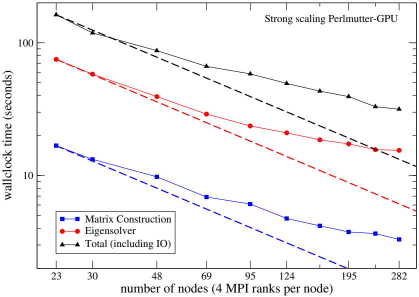
Many-Fermion Dynamics for Nuclear Physics

MFDn: No-Core Configuration Interaction code
for nuclear structure calculations

- Fortran legacy code
  - distributed memory parallel F90/95 code using MPI +
  - construct and stores many-body matrix H_{ij} from input 2NFs (plus 3NFs)
  - obtain lowest eigenpairs using LOBPCG or Lanczos algorithm
  - use eigenvectors to calculate additional observables
- Current status production code
  - hybrid MPI/OpenMP for CPUs
  - hybrid MPI/OpenACC for GPUs (Perlmutter @ NERSC)
  - mixed precision: single precision vectors and matrix elements but dot-products (orthonormalization) in double precision
  - optional: double precision vectors
- Under development
  - performance portability (OpenMP target offload)
  - on-the-fly implementation targeting GPUs

Current Performance on Perlmutter-GPU





- Factor of 6× to 16× speedup
  on Perlmutter-GPU node compared to Theta-KNL node
- Weak scaling
  - local work load kept approximately fixed in terms of number of nonzero matrix elements per node for 7 different problem sizes on 48 nodes up to 1339 nodes
- Reasonable strong scaling
  - runs on 23 and 30 nodes on large-memory nodes

Performance Challenges

- Communication overhead during distributed SpMV in eigensolver in particular when running on GPUs
  Optimizing Nuclear Configuration Interaction Calculations on GPUs: A Comparative Performance Study of Programming Models,
  Nan Ding, Alperen et al, submitted to IPDPS2025
- Load-balancing of inner-loops on GPUs
  - 'irregular' sparsity structure
  - evaluation of nonzero matrix elements can involve loops over $(A-1)$, or even $(A-1)(A-2)$, spectator nucleons
  - many-body orbitals vary in j and m_j, leading to decoupling loops of reduced TBMEs of different length

  Note: not an issue for CPU implementation, where we use OpenMP dynamic scheduling for multiple threads
- Atomic updates in inner-loops on GPUs in order to avoid race-conditions
  Note: not an issue for CPU implementation, where we use OpenMP private arrays on multiple threads

Memory Challenge

Goal: Reduce numerical uncertainties by increasing HO basis size

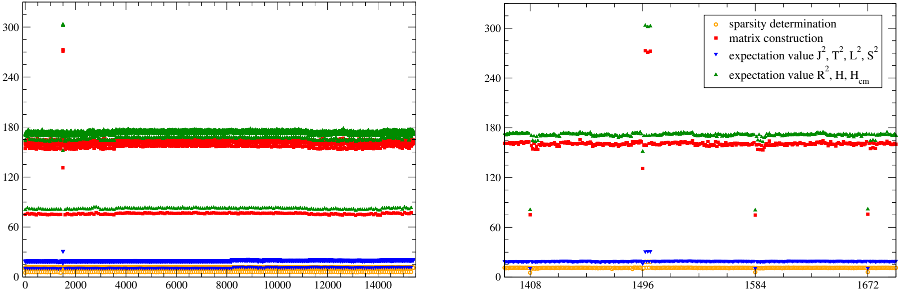
- Largest calculations are limited by available aggregate memory
- Explored benefits of additional remote memory
  - benefits for MFDn limited by memory bandwidth

  Evaluating the potential of disaggregated memory systems for HPC applications, Nan Ding et al, Concurrency Computat Pract Exper. 36(19):e8147 (2024)
- 'On-the-fly' implementation
  - functional, but performance-limited
    ^{12}C at $N_{\max} = 12$ (D = 81 billion) on Perlmutter-GPU: 1 iteration/hr
  - may be feasible if we can improve load-balancing of inner-loops and/or reduce (or eliminate) need for atomics in inner-loops
- Eigenvector perturbation method

  Accelerating Eigenvalue Computation for Nuclear Structure Calculations via Perturbative Corrections, Dong Min Roh et al, arXiv: 2407.08901, under review, Journal of Computational Physics

Hardware Challenges

- Memory usage near-perfectly distributed over all MPI ranks
  - OOM error if a single node (or CPU or GPU) is missing memory
- Local compute load near-perfectly distributed over all MPI ranks
  - poor performance if a single node (or CPU or GPU) underperforms



- Run on 15,400 MPI ranks, 4 ranks per CPU,
  with one CPU underperforming by a factor of two

Convergence Challenge

Achieve numerical convergence for No-Core CI calculations using a finite amount of CPU time on current HPC systems

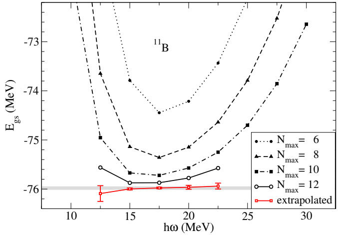
- Perform a series of calculations with increasing N_{max} truncation
- Extrapolate to infinite model space
  $\longrightarrow$ exact results with uncertainties
- Empirical: fit binding energy using exponential in N_{max}

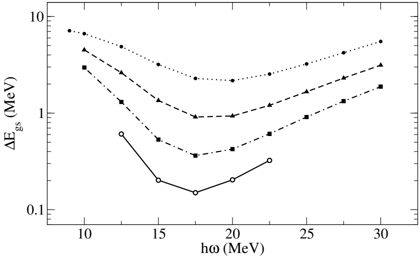
$$E^{\hbar\omega}(N) = E_{\infty}^{\hbar\omega} + a_1 \exp(-a_2 N)$$

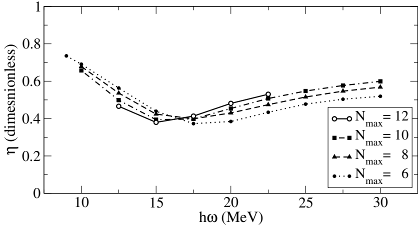
PM, Shirokov, Vary, PRC79, 014308 (2009)

  - use $\hbar\omega$ and N_{max} dependence to estimate numerical uncertainties
- Extrapolate using exponentials in $\sqrt{\hbar\omega/N}$ and $\sqrt{\hbar\omega N}$ motivated by IR and UV behavior of wavefunction, based on s.p. asymptotics
  Coon *et al*, PRC86, 054002 (2012); Furnstahl *et al*, PRC86, 031301(R) (2012); More *et al*, PRC87, 044326 (2013); Wendt *et al*, PRC91, 061301 (2015); Gazda *et al*, PRC106, 054001 (2022);...
- Use Machine Learning/Artificial Neural Networks
  Negoita *et al*, PRC99, 054308 (2019); Jiang it at al, PRC100, 054326 (2019); Knöll *et al*, PLB839, 137781 (2023); ...

Empirical exponential extrapolation of binding energies





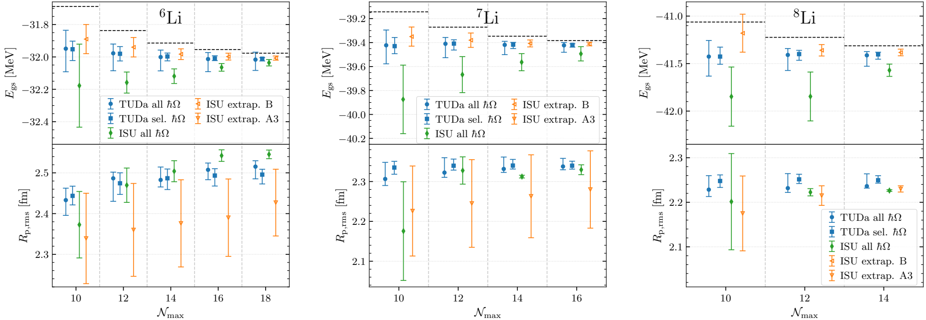


- Use 3 consecutive $N_{\rm max}$ values to determine $E_\infty^{\hbar\omega}$
- Optimal $\hbar\omega$ where $|E_\infty^{\hbar\omega} - E^{\hbar\omega}(N_{\rm max})|$ is minimal (near var. min.)
- Uncertainty estimate based on
  - difference in $E_\infty^{\hbar\omega}$ for two successive extrapolations
  - half the variation in $E_\infty^{\hbar\omega}$ over 6 ∼ 8 MeV $\hbar\omega$-interval

$$\eta(\hbar\omega, N_{\rm max}) \equiv \frac{\Delta E_{\rm gs}(\hbar\omega, N_{\rm max})}{\Delta E_{\rm gs}(\hbar\omega, N_{\rm max}-2)}$$

Benchmarking ML/ANN extrapolations

in preparation, Knöll, Lockner, Maris, McCarthy, Roth, Vary, Wolfgruber

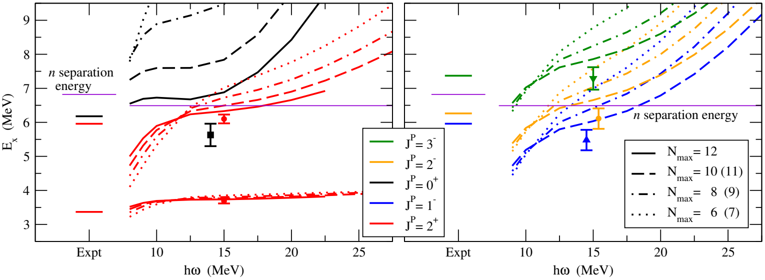


- Based on ML/ANN extrapolations
  - TUDarmstadt: Wolfgruber, Knöll, Roth, PRC 110, 014327 (2024)
  - ISU: Negoita *et al*, PRC99, 054308 (2019); McCarthy, Lockner, PM, Vary, in preparation
- Binding energy: ISU B, empirical exponential extrapolation
- RMS proton radius: ISU A3, based on IR and UV analysis

Shin et al, J. of Physics G: Nuclear and Particle Physics, 44, 075103 (2017)

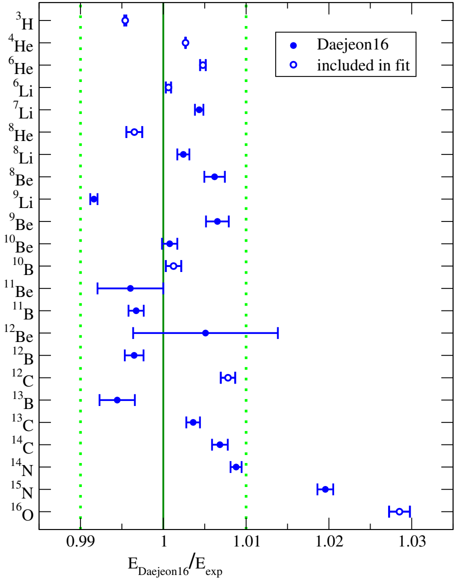
Excitation energies as well as threshold and separation energies

- Differences of total energies (= eigenvalues of Hamiltonian)
- Current practice
  - difference of extrapolated total energies
  - numerical uncertainty estimate:
    max. of extrapolation uncertainties of total energies
  - e.g. ^{10}Be spectrum with Daejeon16 (in preparation)



- Need correlations between extrapolation uncertainties

Ground state energies with Daejeon16

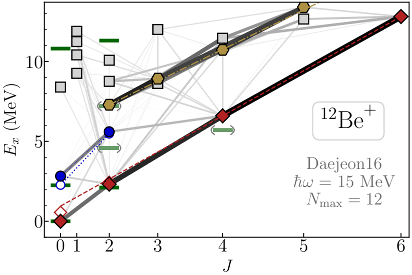


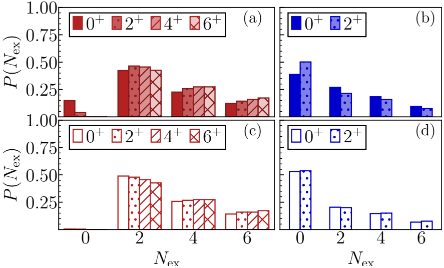
(PM, Shin, Caprio, Vary, in preparation)

- NN interaction, based on SRG evolved chiral EFT, using phase-equivalent transformations to fit select *p*-shell nuclei
- Ground state energies agree with experiment to within 1% for all *p*-shell nuclei up to A=14
- Correct J^{π} for ^{10}B (fitted)
- Parity inversion ^{11}Be
- Correct J^{π} for ^{12}B
- Energies of narrow excited states also agree with data

Intruder band mixing in ^{12}Be

McCoy, Caprio, PM, Fasano, PLB856, 138870 (2024)

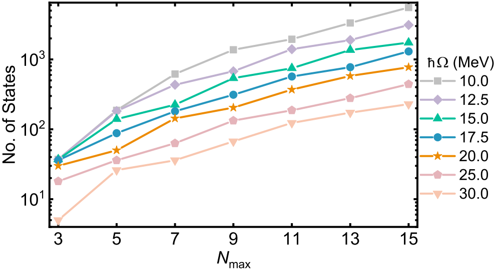


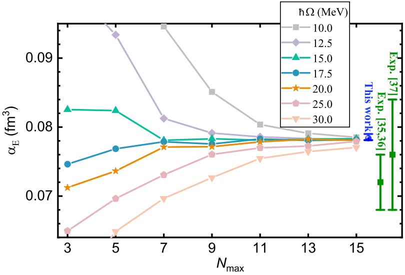


- Intruder and valence-space 0^+ and 2^+ states mix, depending on N_{max} and $\hbar\omega$
- At large N_{max}
  - ground state is dominated by intruder configurations
  - lowest 2^+ state appears to be (almost) pure intruder state
  - N_{ex} decompositions of un-mixed states nearly the same among rotational band members
- In agreement with experimental B(E2) transition strengths

Electric dipole polarizability of ^{4}He (Daejeon16)

Peng Yin, Shirokov, PM, Fasano, Caprio, Li, Zuo, Vary, PLB855, 138857 (2024)

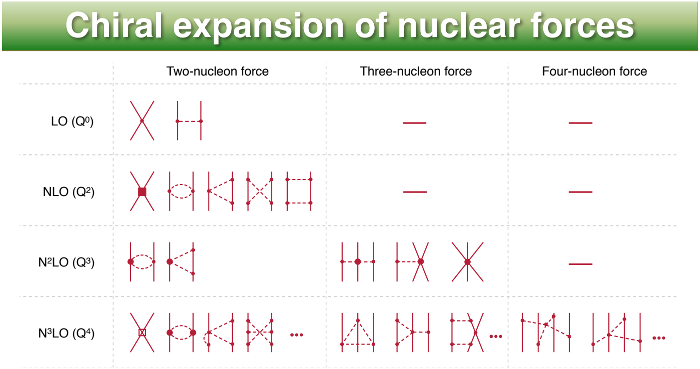




- Obtained result seems to converge and agrees with experiment
- Computational effort
  - need thousands of 1^- eigenstates up to $E_x \simeq 100$ MeV: use $M_j = 1$ and add $\lambda(J^2 - 2)$ to Hamiltonian
  - calc'ns dominated by orthonormalization, most efficient on GPUs
- Similar calculations for ^{6}He in progress
  - for larger nuclei, need to revamp total-J code

Nuclear Interactions from chiral EFT

- Controlled power series expansion in $Q = \max(p, M_\pi)/\Lambda_B \sim 0.3$
- Hierarchy for many-body forces $V_{NN} \gg V_{NNN} \gg V_{NNNN}$
  - 3NFs appear at N^{2}LO, 4NFs appear at N^{3}LO

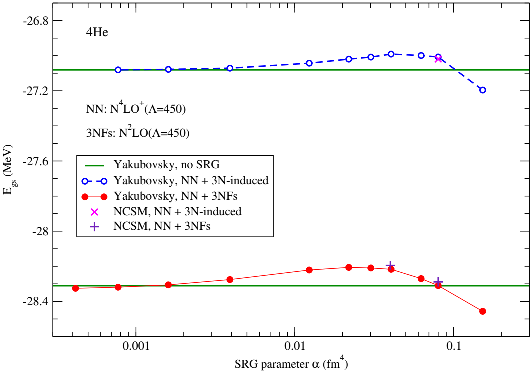


(figure credit: Evgeny Epelbaum)

- Allows for quantification of uncertainties due to the interaction

Similarity Renormalization Group (SRG) evolution

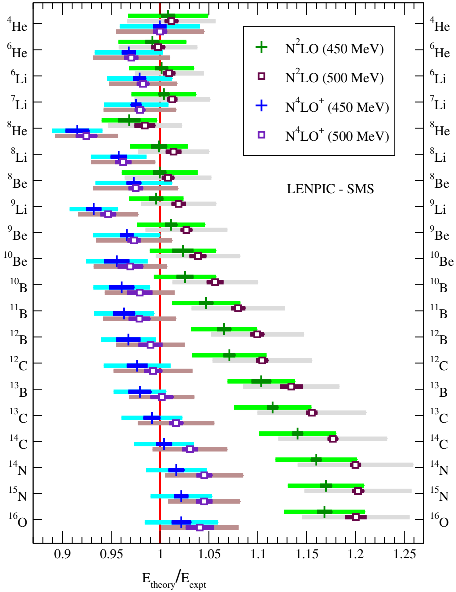
- Apply unitary transformation on Hamiltonian
  to improve numerical convergence in finite HO model spaces



- No change in observables provided that
  - any induced many-body interactions are included
  - same SRG evolution is applied to operators for observables
- Effect of induced 4-body interactions
  - small down to $\alpha < 0.1$ fm^4
  - grow rapidly for $\alpha > 0.2$ fm^4

PM, Le, Nogga, Roth, Vary, Front. Phys. 11 1098262 (2023)

Binding Energies with LENPIC-SMS chiral EFT



PM, Le, Nogga, Roth, Vary,
Front. Phys. 11 1098262 (2023)

- NN potential up to N^{4}LO$^+$
- 3NFs at N^{2}LO
- SRG evolved to $\alpha = 0.08$ fm^4
- LECs fitted to
  - NN scattering data
  - ^{3}H binding energy
  - Nd scattering
- Parameter-free predictions
- Error bars
  - numerical uncertainty
  - chiral EFT uncertainty from Bayesian analysis

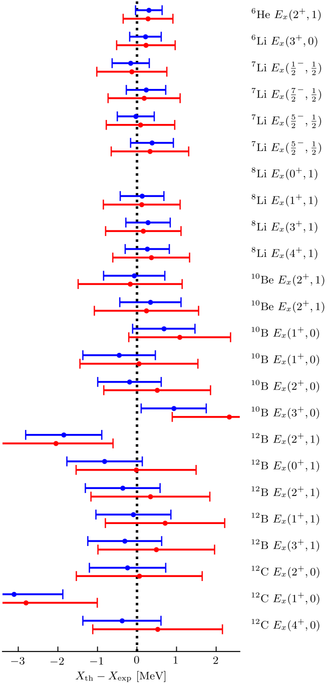
Excitation Energies with UQ

Incorporating correlations
LENPIC collaboration, PRC103, 054001 (2021);
following Melendez *et al* PRC100, 044001 (2019)

- Need to learn variance $\bar{c}$ and Q but only limited orders up to N^{2}LO
- Use just c_3 to find posterior for Q gives $Q \approx 0.3$
- Use c_2 and c_3 to fit empirical covariance matrix

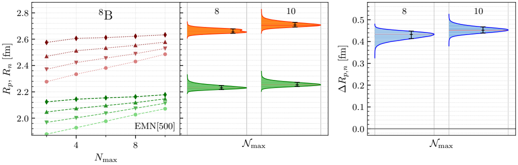
Excitation energies with 95% DoB

- Uncertainties noticably smaller than uncertainties on g.s. energies
- Two outliers: 2^+ in ^{12}B and 1^+ in ^{12}C
- Outliers less correlated with g.s., suggesting different structure

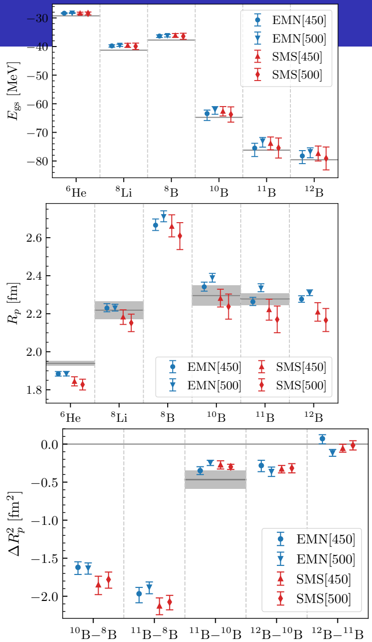


Prediction of charge radius of ^{8}B

preliminary, w. Wolfgruber, Gesser, Knöll, Roth

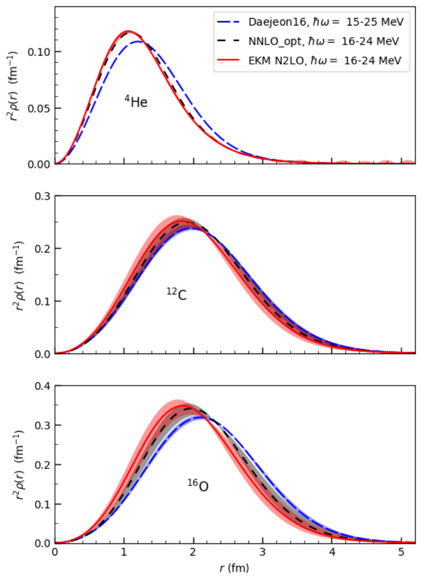


- motivated by planned experiments
- TUDarmstadt ML/ANN energies and radii extrapolations
- NCSM truncation uncertainties folded with chiral uncertainties
- Entem-Machleidt-Nosyk interactions w. N^{3}LO chiral uncertainties
- LENPIC SMS interactions w. N^{2}LO chiral uncertainties



Charge distribution

- Dominated by local proton one-body density
  - center-of-mass motion factorizes in NCSM
  - slow convergence of asymptotic behavior
  - depends on interaction
- Corrections due to
  - electromagnetic current corrections
  - relativistic corrections
- Charge form factor
  - Fourier transform of charge distribution



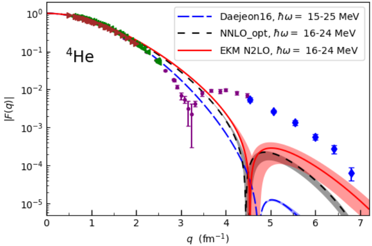
Baker, Burrows, Elster, PM, Popa, Weppner, PRC108, 044617 (2023)

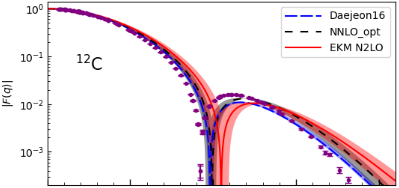
Charge form factor

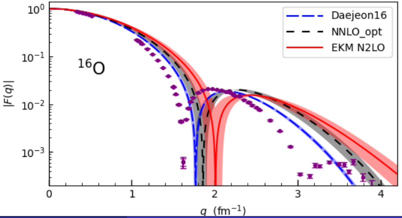
Baker et al, PRC108, 044617 (2023)

$$F_c(q) = \frac{G_p(Q)F_p(q) + (N/Z)G_n(Q)F_n(q)}{\sqrt{1 + Q^2/4m_N^2}}$$

with $G_{p,n}$ proton and neutron form factors







Need to

- include current corrections
- quantify uncertainties

(bands represent spread over 8 MeV window in $\hbar\omega$)

Nucleon-Nucleus Scattering

- Spectator expansion of multiple scattering theory
  - ordering of the scattering process according to the number of active target nucleons interacting directly with the projectile nucleon
- Generally restricted to leading-order term:
  projectile interacts directly with only one target nucleon
  - believed to be applicable in the 50 $\sim$ 200 MeV projectile energies
  - need one-body density of target nucleus
- Note: unclear what the actual expansion parameter is

  (assuming the spectator expansion has a well-defined expansion parameter . . .)
- Until recently
  - use realistic NN potential
  - purely phenomenological, local, one-body density
- Recent development
  - use one-body density from ab initio nuclear structure calculation
  - use same NN interaction for structure and scattering
  - consistency:
    also restricted to NN-only interactions in structure calculation

Leading-order spectator expansion for NA scattering

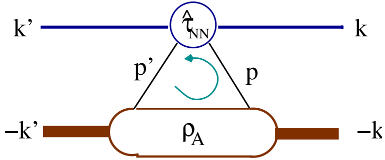
Effective (optical) potential

$$\widehat{U}(q, \mathcal{K}_{NA}, \epsilon) = \sum_{\alpha=n,p} \sum_{K_s} \int d^3\mathcal{K}\, \eta(q, \mathcal{K}, \mathcal{K}_{NA}) \; \hat{\tau}_\alpha^{K_s}\left(q, \frac{1}{2}\left(\frac{A+1}{A}\mathcal{K}_{NA} - \mathcal{K}\right); \varepsilon\right) \rho_\alpha^{K_s}\left(\mathcal{K} - \frac{A-1}{A}\frac{q}{2}, \mathcal{K} + \frac{A-1}{A}\frac{q}{2}\right)$$

with $\eta(\ldots)$ the Moller factor, relating the NN frame to the NA frame
$\hat{\tau}_{n,p}^{K_s}$ the NN amplitude between projectile and target nucleon ($K_s = 0, 1$)
$\rho_{n,p}^{K_s}$ the (nonlocal) one-body density of the target ($K_s = 0, 1$) and

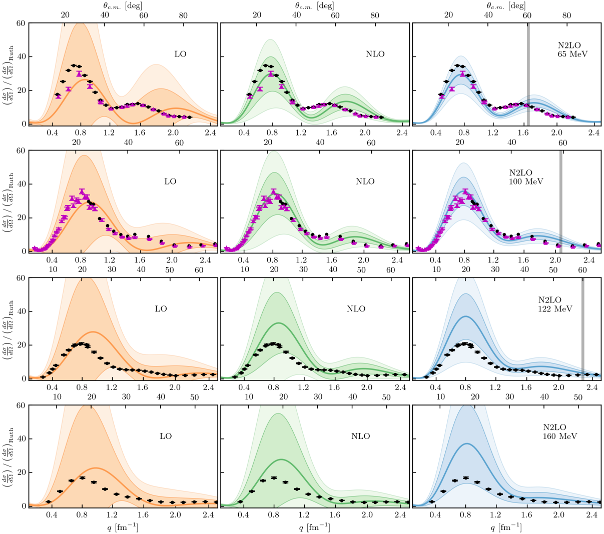
$$q = p' - p \qquad \mathcal{K} = \frac{1}{2}(p' + p)$$

$$\mathcal{K}_{NA} = \frac{A}{A+1}\left[(k' + k) + \frac{1}{2}(p' + p)\right]$$



NA scattering: Differential cross section $^{12}C(p,p)^{12}C$

Baker, Burrows, Elster, Launey, PM, Popa, Weppner, Front.Phys. 10 1071971 (2022)



LENPIC SCS, NN-only

- up to N^{2}LO
- chiral uncertainties

Leading Order spectator expansion

- uncertainties in the spectator expansion ?
- 3NFs in scattering ?

Numerical uncertainties

- negligible compared to N^{2}LO chiral EFT and LO spectator expansion uncertainties

- NCCI approach powerful tool for ab initio nuclear structure
  - Binding and excitation energies agree with experiments
    - Daejeon16 (based on chiral N^3LO NN potential)
  - ML/ANN can help to quantify truncation uncertainties
    - extensions to other observables in progress
  - Long-range observables converge slowly in HO basis
    - ratio's of long-range observables converge significantly better
  - Can quantify EFT truncation uncertainties
    - LENPIC NN + 3NF chiral EFT interactions
  - Need to incorporate chiral EFT corrections for electroweak observables ('Meson Exchange Currents')
    - in particular for magnetic moments and beta-decay
  - Provide densities as input for NA scattering calculations
- Caveats
  - Neither NN (and 3NFs) nor nuclear wave functions are unique
  - Unitary transformations are often used in nuclear CI calculations to improve numerical convergence
    - need to consistently transform all operators
- Yes, we are ready for Exascale computing
  - thanks to ongoing collaborations with applied mathematicians and computer scientists