Quantum Embedding Methods: GPU Acceleration & Embracing Multiple Frameworks

Corey Melnick (cmelnick@bnl.gov)

Brookhaven National Laboratory

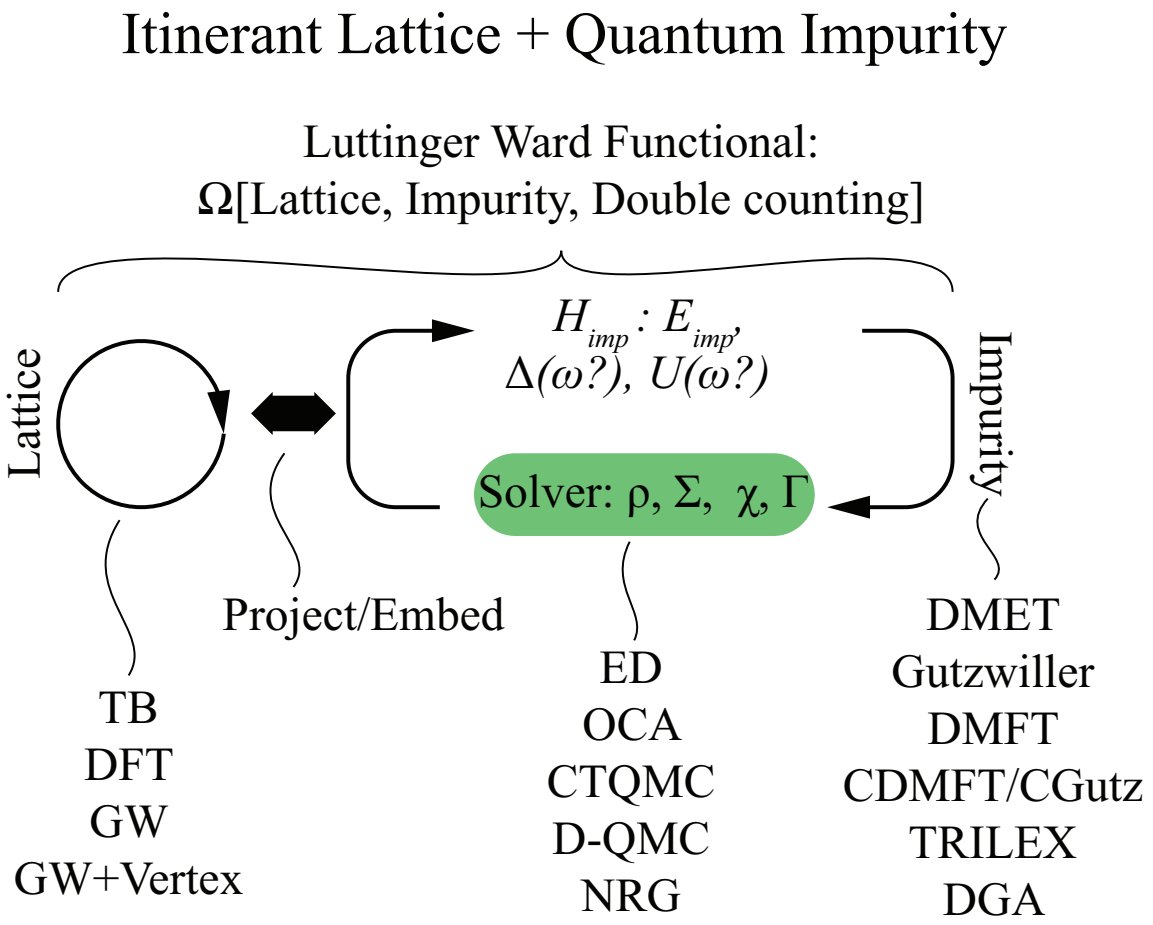
COMSCOPE
from codes to spectroscopies

Outline

1. Context: Quantum embedding methods & CTQMC
2. The computational bottleneck in CTQMC (CT-HYB)
3. GPU's and the acceleration of CTQMC
4. Plutonium: A primer & what GPU acceleration buys us
5. Thermal contraction in Plutonium
6. Crystal fields & symmetry adapted bases
7. Leveraging seamless DFT+DMFT & DFT+G:

1. Context: Quantum Embedding Methods

- Quantum embedding methods share a general framework:
  1. Treat the crystal with some itinerant model
     - (DFT, GW, …)
  2. Project onto a local, correlated subspace
  3. Treat subspace with local model
     - (Gutzwiller, DMFT, …)
  4. Solve impurity problem of local model
     - (ED, CTQMC, NRG, …)
  5. Embed local quantities into itinerant space and solve new itinerant problem
     - (density, self-energy, polarizability)
  6. Repeat until charge self-consist



1. Why do methods proliferate?

- Each method has its particular evolutionary niche with unique capabilities and limitations
  - Lattice:
    - DFT
      - Great at weakly correlated materials, fast, forces, phonons
    - G_0W_0, Quasiparticle-GW, self-consistent GW
      - Band gaps, excitations, and long-range correlation
  - Correlated Subspace:
    - DMFT
      - $T \neq 0$, lifetimes, great at capturing local (non-quasiparticle, Mott) physics
    - Gutzwiller (RISB)
      - T=0, Fast, full crystal fields, great at capturing correlated quasiparticles
    - CDMFT/CGutz/DCA
      - Long range correlations

1. Quantum impurity solvers

Quantum Impurity Solver Proliferation

Exact	Approximate
Numerical renormalization group (NRG)	One-crossing approximation (OCA)
Exact Diagonalization (ED)	Hubbard-one
CTQMC	

Why choose CTQMC?

- Only exact quantum impurity solver which currently handles full d- and f-shell materials
- Extremely and easily parallelizable (near ideal weak scaling)

Why ComCTQMC?

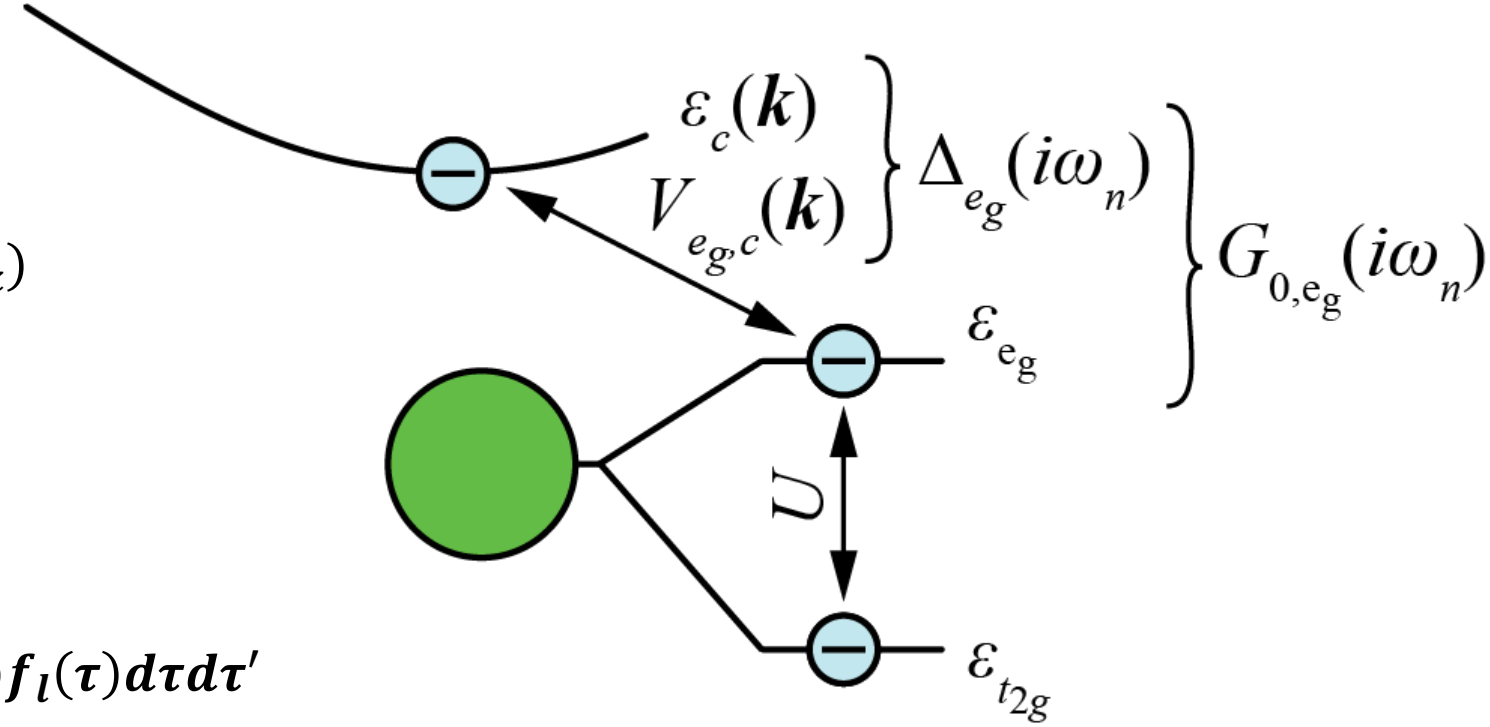
- GPU accelerated (1 - 225x depending on the number of orbitals & symmetries)
- Worm Sampling (Full susceptibility tensor, atomic limit)
- Vertex asymptotics (Build two-particle vertices at any frequency)

1. The quantum impurity problem

Anderson Impurity Model

- Hamiltonian:

$$\boldsymbol{H_{AIM}} = \sum_k \boldsymbol{\epsilon_k c_k^\dagger c_k} + \boldsymbol{\epsilon_f f^\dagger f} + \boldsymbol{U n_\uparrow n_\downarrow} + \sum_k (\boldsymbol{V_k c_k^\dagger f} + \boldsymbol{V_k^* f^\dagger c_k})$$



- Action:

$$\boldsymbol{S} = \sum_{ij} \iint \boldsymbol{f_i^\dagger(\tau) \mathcal{G}_{0,ij}^{-1}(\tau - \tau') f_j(\tau') d\tau d\tau'} + \sum_{ijkl} \iint \boldsymbol{f_i^\dagger(\tau) f_j^\dagger(\tau') \mathcal{U}_{ijkl}(\tau - \tau') f_k(\tau') f_l(\tau) d\tau d\tau'}$$

➢ A dynamical mean-field and a (dynamical) interaction

2. Computational Bottleneck

- Three main computations in CTQMC
  - $w(C) = w_{hyb}(C)\, w_{dyn}(C)\, w_{imp}(C)$

- Local impurity trace

$$w_{\mathrm{imp}}(\mathrm{C}) = \mathrm{Tr}\, e^{-\beta H_{loc}} T_\tau \prod_r^k c_{i'_r}(\tau'_r) c^\dagger_{i_r}(\tau_r) = \mathrm{Tr}\, P_{\beta-\tau_k} F^\dagger_{i_k} P_{\tau_k-\tau'_k} F_{i'_k} \dots F_{i_1} P_{\tau_1-\tau'_1} F^\dagger_{i_1} P_{\tau'_1}$$

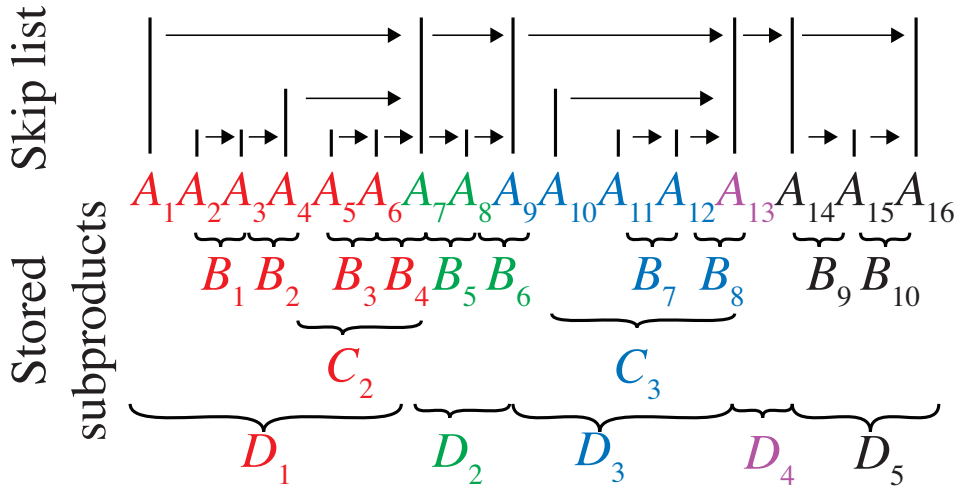
where $(F_i)_{mn} = \langle m|c_i|n\rangle, \quad P_\tau = e^{-\tau H_{loc}}$

  - F_i is a matrix of rank 2^n
    - n is the number of orbitals
    - Computation is of order $O[k\,(2^n)^3]$
      - Even with GPUs, this is prohibitive for most real materials

P. Sémon, C.-H. Yee, K. Haule, and A.-M. S. Tremblay, Phys. Rev. B, vol. 90, p. 75149, 2014.

2. Algorithmic Improvements

- Store sub-products (binary tree, skip-list):
  - $O(k\,(2^n)^3) \rightarrow O(\log k\,(2^n)^3)$
    - Less sensitive to expansion order (low temperature, strong hybridization)



- Evaluate the bounds of the calculation (Lazy-trace):
  - Multiplication of the matrix norms establishes bounds
  - Many proposed moves can be easily discarded w/o full matrix multiplication
    - Norm has cost $O(\log k\,(2^n)^2)$

P. Sémon, C.-H. Yee, K. Haule, and A.-M. S. Tremblay, Phys. Rev. B, vol. 90, p. 75149, 2014.

2. Decomposing the Hilbert space

- **Decompose the Hilbert space** $\mathscr{H}$ according to symmetries of the atomic Hamiltonian

$$\mathscr{H} = \bigoplus_{q=1}^{N} \mathscr{H}(q)$$

  - Each **sector** q has unique set of quantum numbers that all Fock states in the sector share:
    - N, S_z, S^2, …
    - N, J_z, J^2, …

  - **New basis**:

$$[F_i(q,j)]_{mn} = \langle m(q_{j+1})|c_i|n(q_j)\rangle$$

  - The product $P_{\beta-\tau_{2k}} F_{\alpha_{2k}} \dots F_{\alpha_1} P_{\tau_1}$ traverses sectors of $\mathscr{H}$

$$q_0 \to q_1 \to \cdots \to q_{2k} = q_0$$

    - We can store this map and discard moves that pass through the unphysical sectors (prevented by pauli, e.g.)

Sectored Hilbert Space

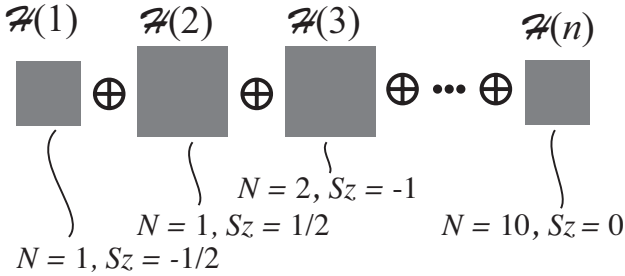
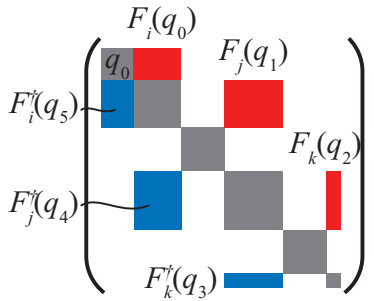
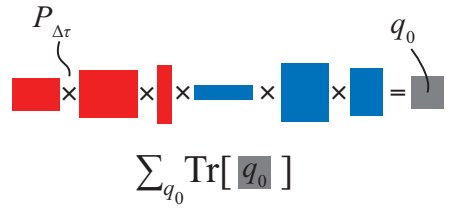


Illustration of Block Diagonalized Hilbert Space & Operators $F_\alpha(q)$



Local Trace



Brookhaven National Laboratory

COMSCOPE from codes to spectroscopies

2. Symmetries and Computational Cost

$$w_{\mathrm{imp}}(\mathrm{C}) = \mathrm{Tr}\ P_{\beta-\tau_{2k}} F_{\alpha_{2k}} \dots F_{\alpha_1} P_{\tau_1} \rightarrow \sum_{q_0} \mathrm{Tr}\ P_{\beta-\tau_{2k}}(q_{2k}) F_{\alpha_{2k}}(q_{2k-1}) \dots F_{\alpha_1}(q_0) P_{\tau_1}(q_0)$$

- $2k$ matrix multiplications means the burden scales as

$$O\left((2^n)^3 k\right) \rightarrow O\left(n_{sectors}(2^{n/n_{sectors}})^3 k\right)$$

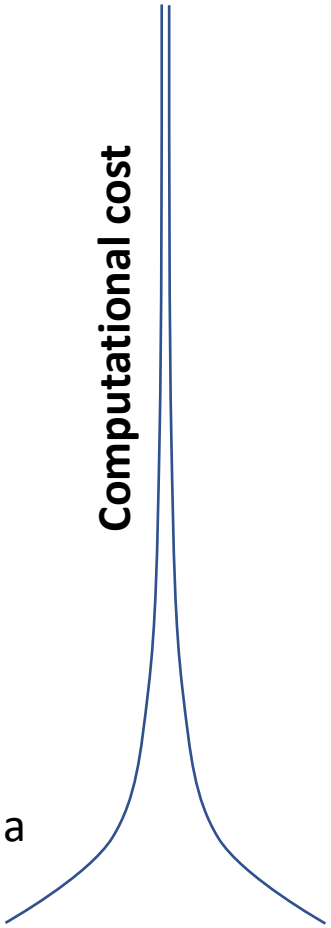
- **The more symmetries (good quantum numbers) there are,** the more sectors there are
  - The more sectors there are, the smaller the sectors are
  - The smaller the sectors are, the easier the local trace becomes
  - The easier the local trace becomes, **the faster the CTQMC is (time-per-step)**

- **Degeneracy within the sectors has no effect on the computational burden of the local trace**, although such degeneracy can be used to combine measurements and reduce the number of steps required for convergence
  - **Time-per-step unaffected, but time-to-solution still improved**

2. What Problems Can and Can't We Solve?

- From easy to impossible
  - 2 orbitals (extremely easy)
    - Cuprates (d_{x2-y2} subspace)
  - 4 orbitals (extremely easy)
    - Nickelates (e_g subspace)
  - 6 orbitals (easy)
    - $SrVO_3$ and other early transition-metal oxides (t_{2g} subspace)
  - 8 orbitals (moderate)
    - Cuprate CDMFT (2x2x1)
  - 10 orbitals (moderate to difficult)
    - Iron, iron pnictides/chalcogens, ruthenates (full d-shell)
  - 14 orbitals (very-difficult to impossible)
    - Actinides & lanthanides
  - 16 orbitals (impossible)
    - Cuprate CDMFT (4x4x1), Nickelate CDMFT (2x2x1)

GPU Acceleration!
Makes the impossible possible and the difficult a whole lot easier!



Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

2. How much should we care?

- What is to be gained from understanding actinide systems?
  - They are much more complex – while d-shell materials often can be described as Hund metals, Mott insulators, etc. actinides support a more diverse combination of physics
    - E.g., Kondo + Mott + Hund + spin-orbit + crystal field effects are all important in Plutonium & UGe_2
    - As such, they offer insight into how these phenomenon interact in a material
  - Practical interest
    - Heavy fermion systems / heavy fermion superconductors
    - Uranium as nuclear energy material

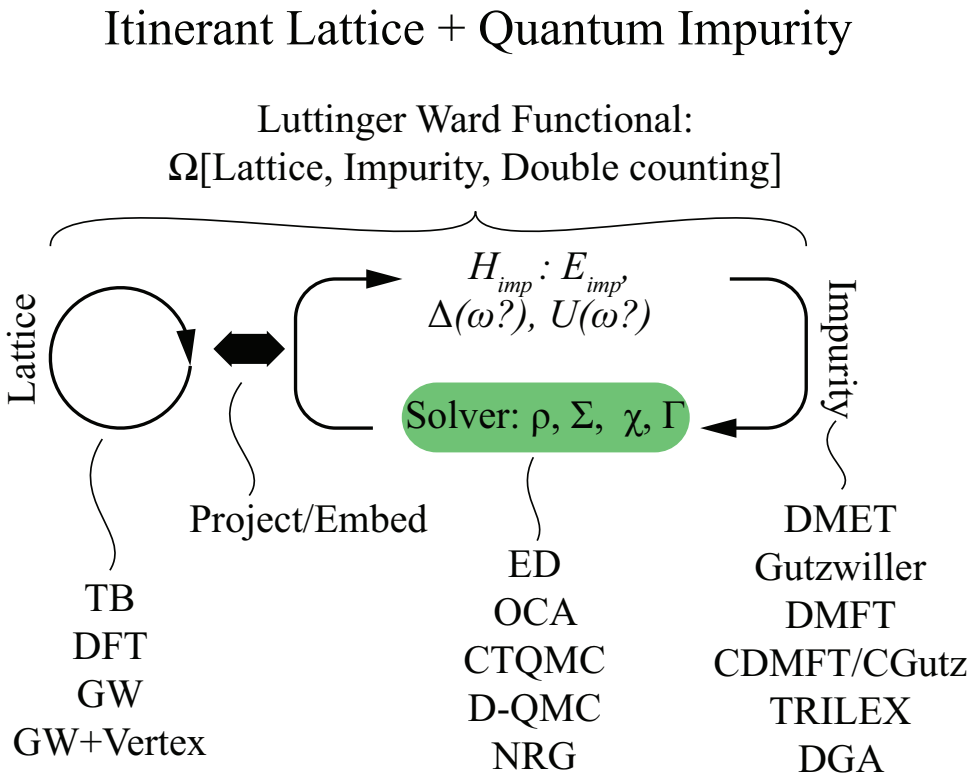
- What is to be gained from simulating clusters?
  - Long-range interactions!
    - Fermi arcs in cuprates
    - Better Mott transitions
    - ???

2. What is to be done?

- Improve or change quantum impurity solver
  - Use a solver without the sign problem
    - (if that's the problem)
  - GPU acceleration

- Change Quantum Embedding Theory
  - Use a faster method
  - Use a method that does not care about the basis
  - Use a method that can reach T=0

  - DFT+G
    - (No finite temperatures, yet)
    - (Bad mott physics?)

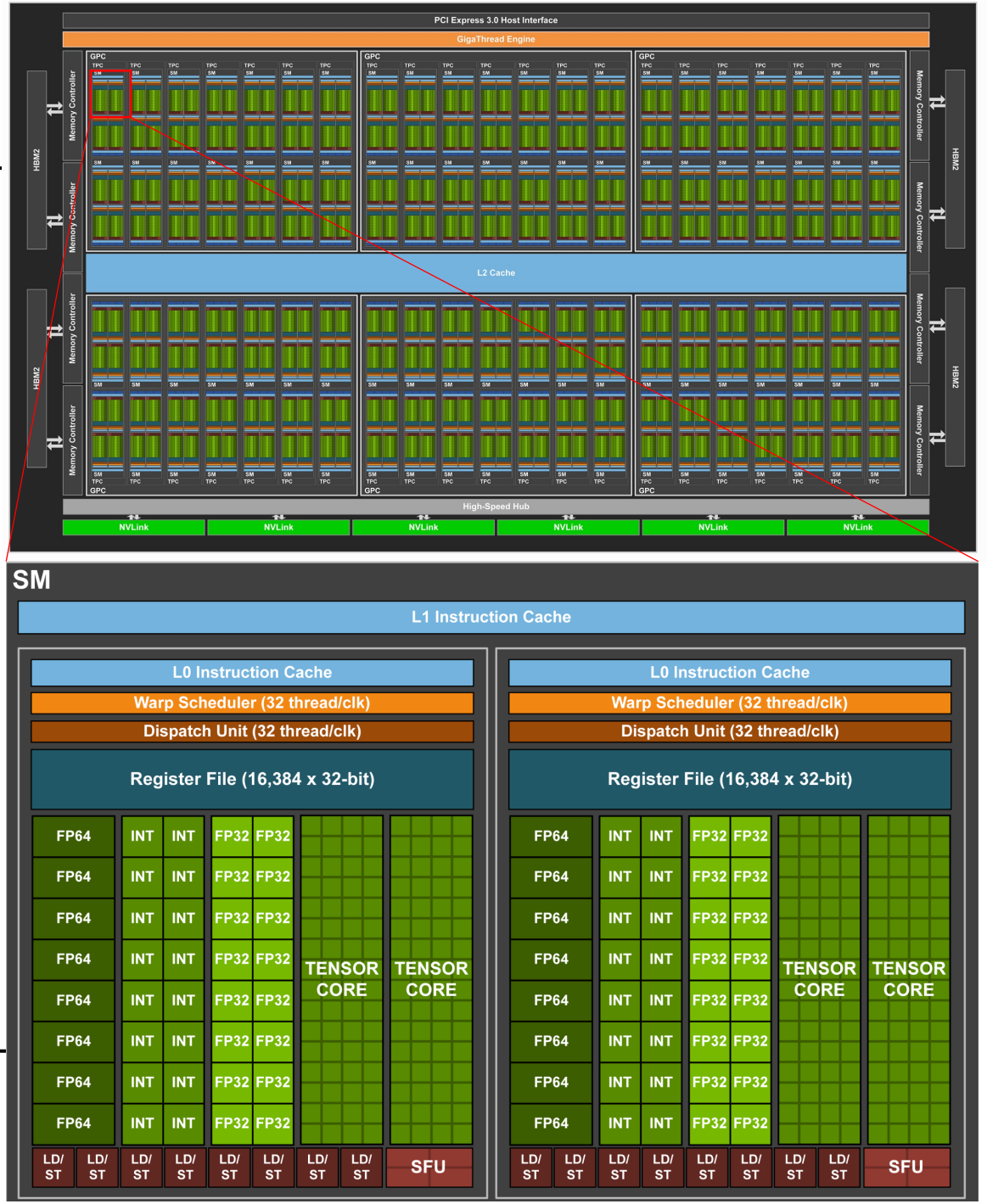


Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

3. GPU Primer

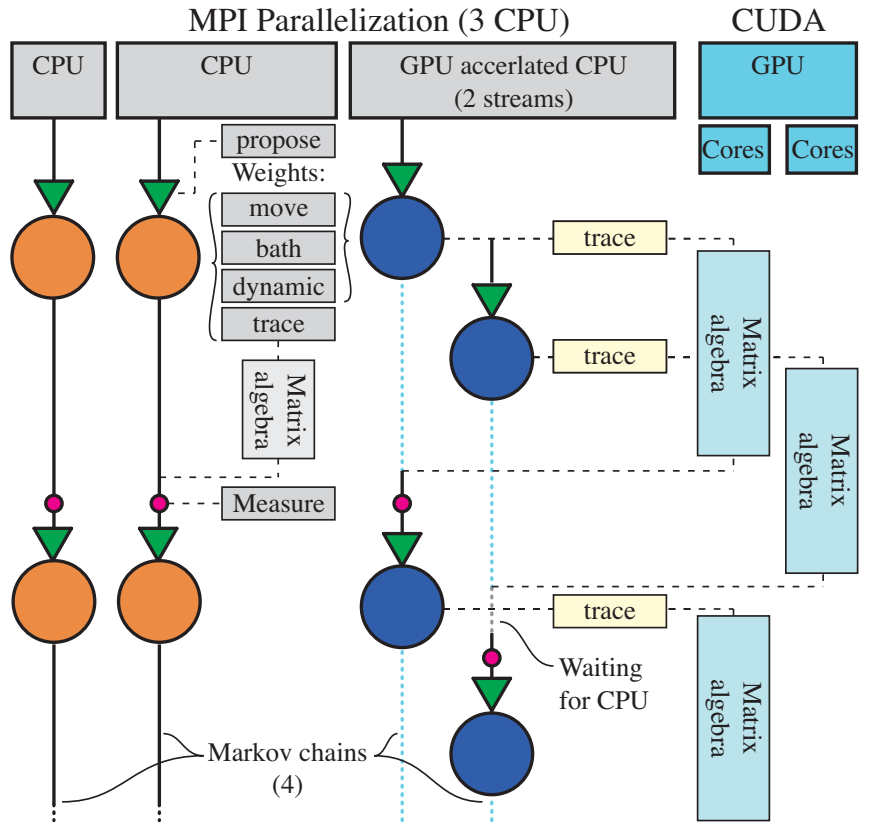
- GPU's excel at large matrix multiplication
  - Impurity trace!
  - O(1000) compute cores
  - Fast memory access
- Limitations:
  - CPU-GPU communication can bottleneck
  - Less memory
  - Slow at branching logic
- Right: A diagram of Summit's GPUs
  - Near exascale!
    docs.olcf.ornl.gov/systems/summit_user_guide.html



Brookhaven National Laboratory

3. Accelerating ComCTQMC

- Handle all linear algebra for local impurity trace on the GPU
  - Matrix multiplication
  - Matrix norm
  - Time evolution operator
  - Trace
- Store on GPU (to limit CPU-GPU communication)
  - Operator Matrices
  - Sub-products in data structure
- Communicate (to limit branching logic)
  - Instructions
    - which matrices to multiply and in which order, but not the contents of these matrices!
  - Results (w_{loc})

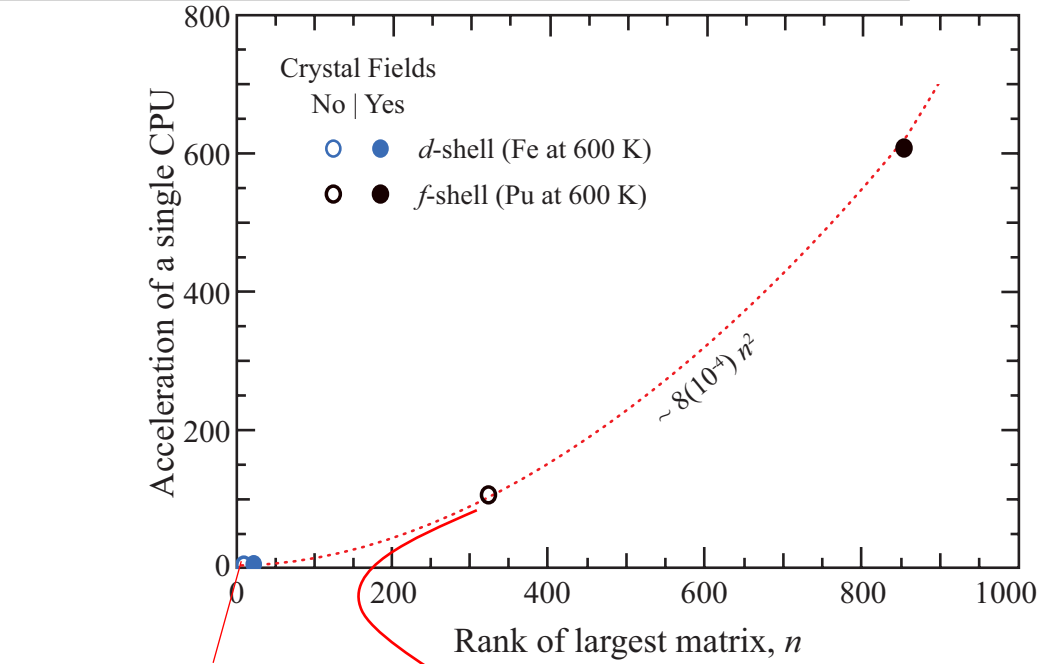


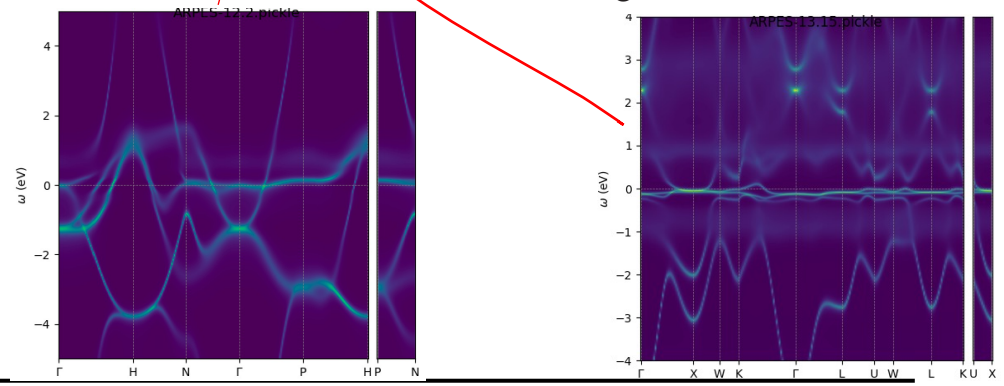
Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

3. GPU acceleration examples

- Density-Density or Ising interactions:
  - 2^n log k matrices of rank 1
  - Deceleration – communication dominates
- Spherically symmetric interaction in d-shell
  - ~50 log k matrices of rank ~20
  - Very little acceleration!
- Spherically symmetric interaction in f-shell
  - ~200 log k matrices of rank ~100
  - 15-20x acceleration of summit node
- Delta-Pu interaction with crystal fields
  - ~50 log k matrices of rank ~500
  - 200x acceleration of summit node
- BUT! Large subspaces DO NOT HELP CTQMC.
  - CTQMC slows to a crawl with $r_{max} > 300$
  - GPUs help alleviate *some* of the associated burden



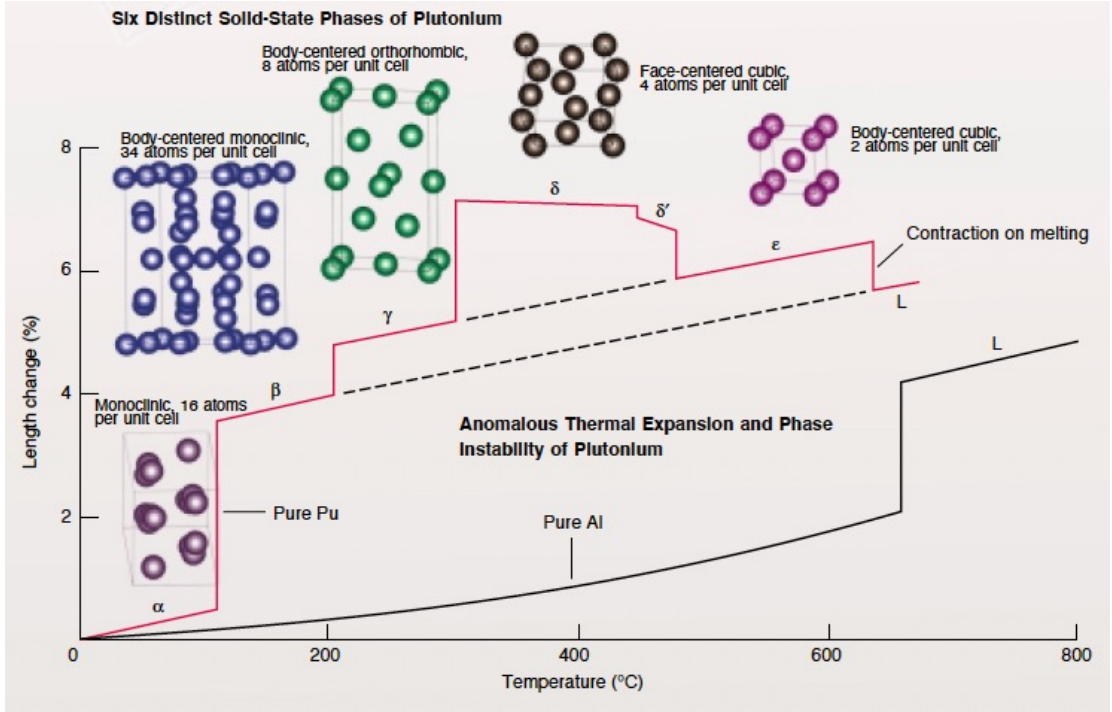


Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

4. Plutonium: "A physicist's dream and a metallurgist's nightmare."

- Plutonium is a fascinating material
  - A lot of phases as we vary temperature
  - Phase changes associated with dramatic volume changes
  - Negative thermal expansion in δ-phase
  - "Nearly magnetic"
  - Why? Correlated 5f electrons!
- Very hard to study
  - Experimentally
    - Toxic & radioactive
    - Controlled heavily
  - Theoretically
    - DFT fails to predict basic behaviors
    - DMFT is extremely costly
    - Gutzwiller cannot (currently) capture T≠0



N. G. Cooper, D. R. Delano, and A. T. Loweree, "Challenges in Plutonium Science," *Los Alamos Sci.*, vol. 26, no. 1 and 2, pp. 1–502, 2000.

Brookhaven National Laboratory

COMSCOPE from codes to spectroscopies

4. Plutonium: Pioneering Works

- Pioneering DMFT w/CTQMC studies

Electronic Coherence in δ-Pu: A Dynamical Mean-Field Theory Study

Authors	CA Marianetti, K Haule, G Kotliar, MJ Fluss
Publication date	2008/8/1
Journal	Physical review letters

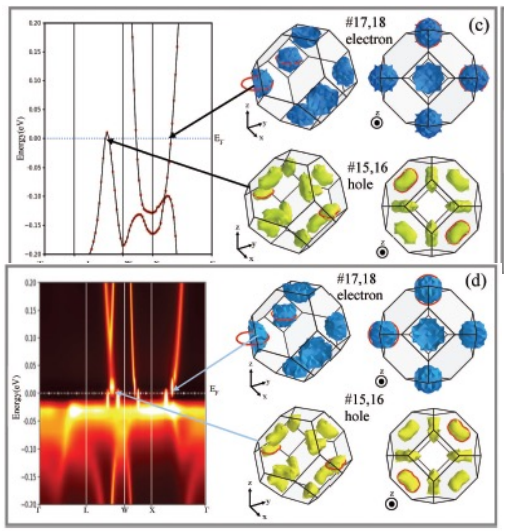
Site-selective electronic correlation in α-plutonium metal

Authors	Jian-Xin Zhu, RC Albers, K Haule, G Kotliar, JM Wills
Publication date	2013/10/18
Journal	Nature communications

- DFT/DFT+U/DFT+G/DFT+DMFT comparison

Electronic correlation induced expansion of Fermi pockets in δ-plutonium

Roxanne M. Tutchton, Wei-ting Chiu, R. C. Albers, G. Kotliar, and Jian-Xin Zhu
Phys. Rev. B **101**, 245156 – Published 23 June 2020



- Crystal fields in early actinides DMFT w/ OCA & CTQMC

Arrested Kondo effect and hidden order in URu_2Si_2

Kristjan Haule & Gabriel Kotliar

Nature Physics **5**, 796–799 (2009) | Cite this article

Chirality density wave of the "hidden order" phase in URu_2Si_2

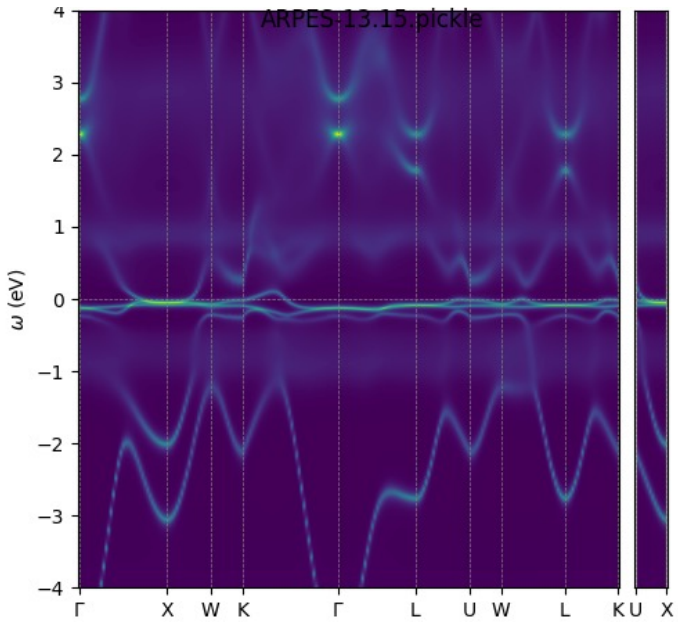
H.-H. KUNG, R. E. BAUMBACH, E. D. BAUER, V. K. THORSMØLLE, W.-L. ZHANG, K. HAULE, J. A. MYDOSH, AND G. BLUMBERG Authors Info & Affiliations

Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

4. GPU acceleration of Plutonium

- For a CPU, the f-shell of δ-Plutonium at 600 K with a spherically symmetric interaction is near the limit of what one reasonably wants to solve with CTQMC
  - 500-1000 *node* hours per impurity problem solution
  - 10,000-20,000 node hours per DFT+DMFT!
    - nearly a million cpu-hours
    - K. Haule's very optimized CTQMC allowed for the first successful DMFT w/ CTQMC of Pu & other late actinides



- With GPU acceleration, it becomes (relatively) easy!
  - 25-50 node hours per impurity problem
  - 500-1000 node hours per DFT+DMFT
- Interactions which consider more crystal field effects become possible
  - 20,000,0000 → 10,000 node hours per impurity problem
  - (DFT+G remains easy – we can see whether it is worth it!)

Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

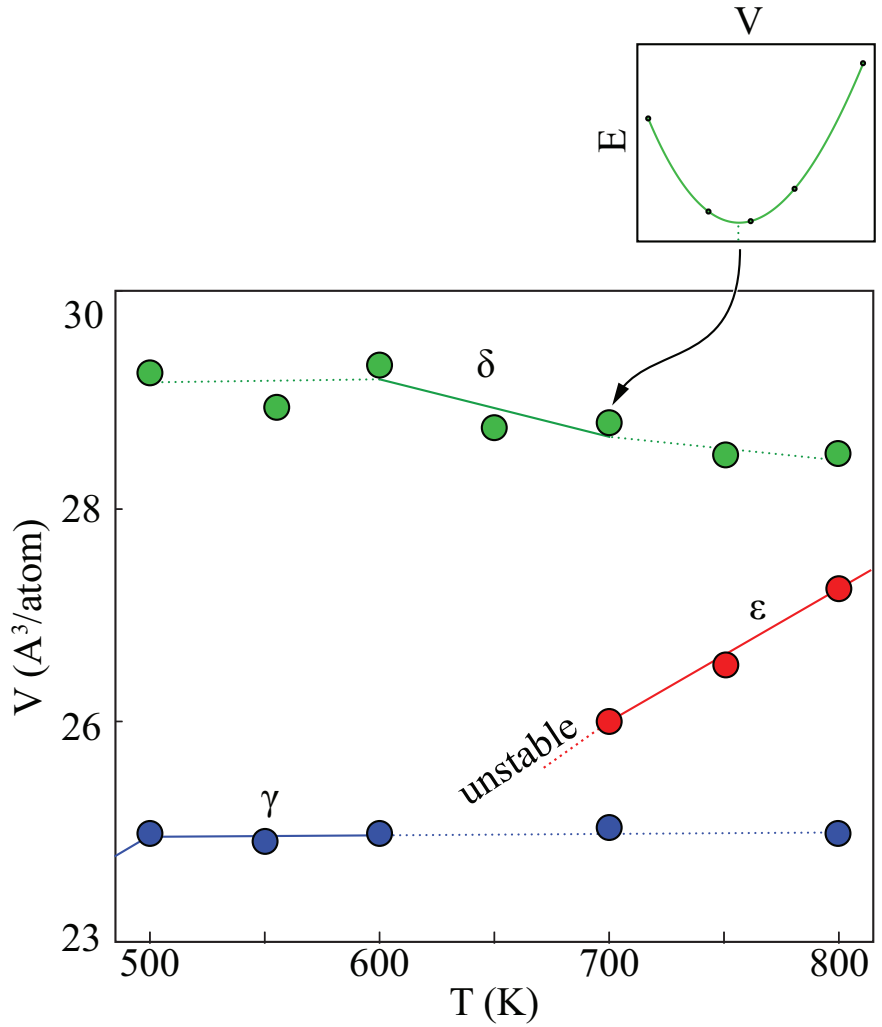
4. What can we do with accelerated DFT+DMFT?

- Explore parameterization
  - The effects of spin-orbit coupling
  - Valence fluctuation
- Structural relaxation
  - Energy vs. volume curves – no forces in DFT+DMFT yet

$$\Omega_{\mathrm{DFT+DMFT}} = \Omega_x + \frac{1}{\beta}\sum_k \sum_{i\omega} \left[trln(-G^{-1}G_x) - \Sigma^c G \right] + \Phi_{\mathrm{CTQMC}}[G_{loc}] - E_{DC}$$

$$\Omega_x = -\frac{1}{\beta}\sum_k \log\left(1 + e^{-\beta\tilde{\epsilon}_k}\right) + E_{coul}[\rho] - tr(V^H G) - \frac{1}{2}tr(\Sigma^x G)$$

- Thermal expansion and contraction!
  - Energy volume curves as a function of temperature
    - Each data point is 5+ runs
      - immense effort without GPU

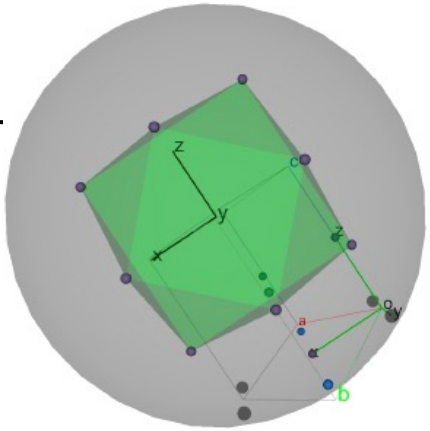


Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

COMSCOPE from codes to spectroscopies

6. Crystal field effects



1. Analyze the local symmetry group (what is the atom's local environment?)
2. The symmetry group enforces
   A. Which hopping elements in the one-body Hamiltonian vanish
   B. Which elements are members of same irreducible representation (irrep)

- Crystal fields (the local symmetry group) bring us from a spherically symmetric solution to one with some number of hopping elements, and some splitting of elements on the diagonal

```
[[1 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 1 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 1 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 1 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 1 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 2 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 2 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 2 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 2 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 2 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 2 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 2 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 2]]
```

Crystal fields

In Jm_z basis

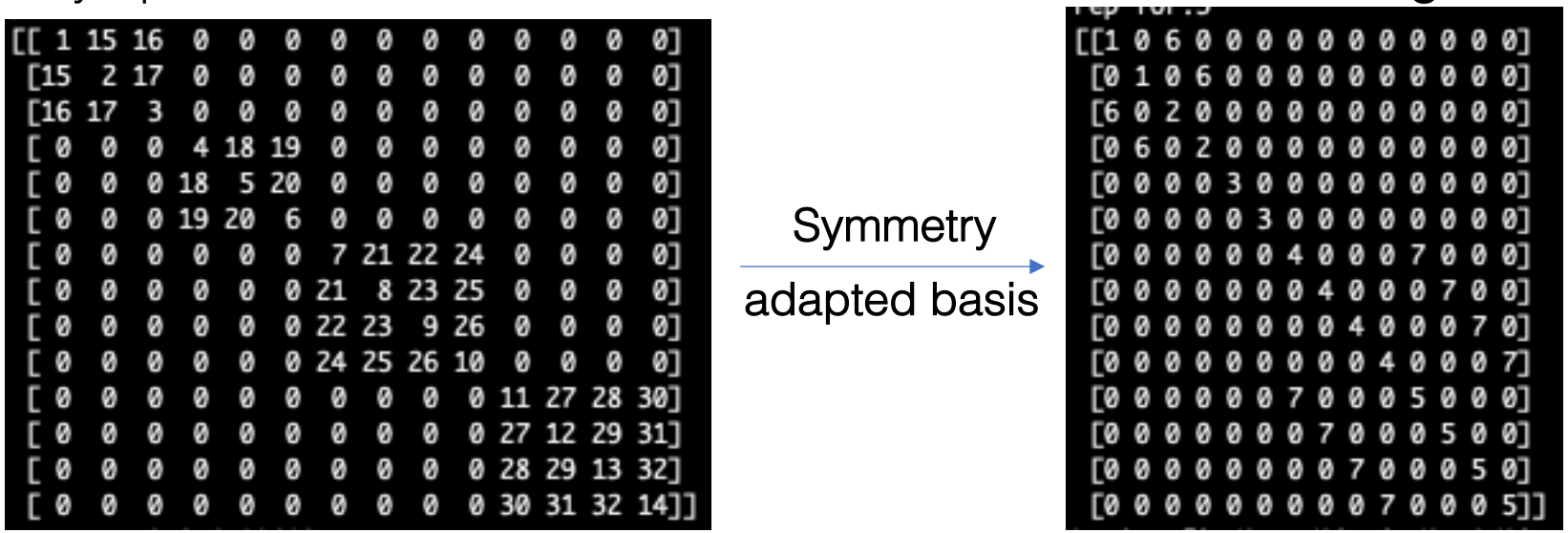
```
[[ 1 15 16  0  0  0  0  0  0  0  0  0  0  0]
 [15  2 17  0  0  0  0  0  0  0  0  0  0  0]
 [16 17  3  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  4 18 19  0  0  0  0  0  0  0  0]
 [ 0  0  0 18  5 20  0  0  0  0  0  0  0  0]
 [ 0  0  0 19 20  6  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  7 21 22 24  0  0  0  0]
 [ 0  0  0  0  0  0 21  8 23 25  0  0  0  0]
 [ 0  0  0  0  0  0 22 23  9 26  0  0  0  0]
 [ 0  0  0  0  0  0 24 25 26 10  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 11 27 28 30]
 [ 0  0  0  0  0  0  0  0  0  0 27 12 29 31]
 [ 0  0  0  0  0  0  0  0  0  0 28 29 13 32]
 [ 0  0  0  0  0  0  0  0  0  0 30 31 32 14]]
```

Brookhaven National Laboratory

COMSCOPE from codes to spectroscopies

6. Crystal field effects: What can DMFT do?

- DMFT (with CTQMC as an impurity solver) cannot typically handle off-diagonal elements
  - Sign problem (unless we severely truncate the interaction tensor)
  - We can just throw away all off-diagonals or…
- Symmetry operations can **rotate basis to minimize number of off-diagonals**



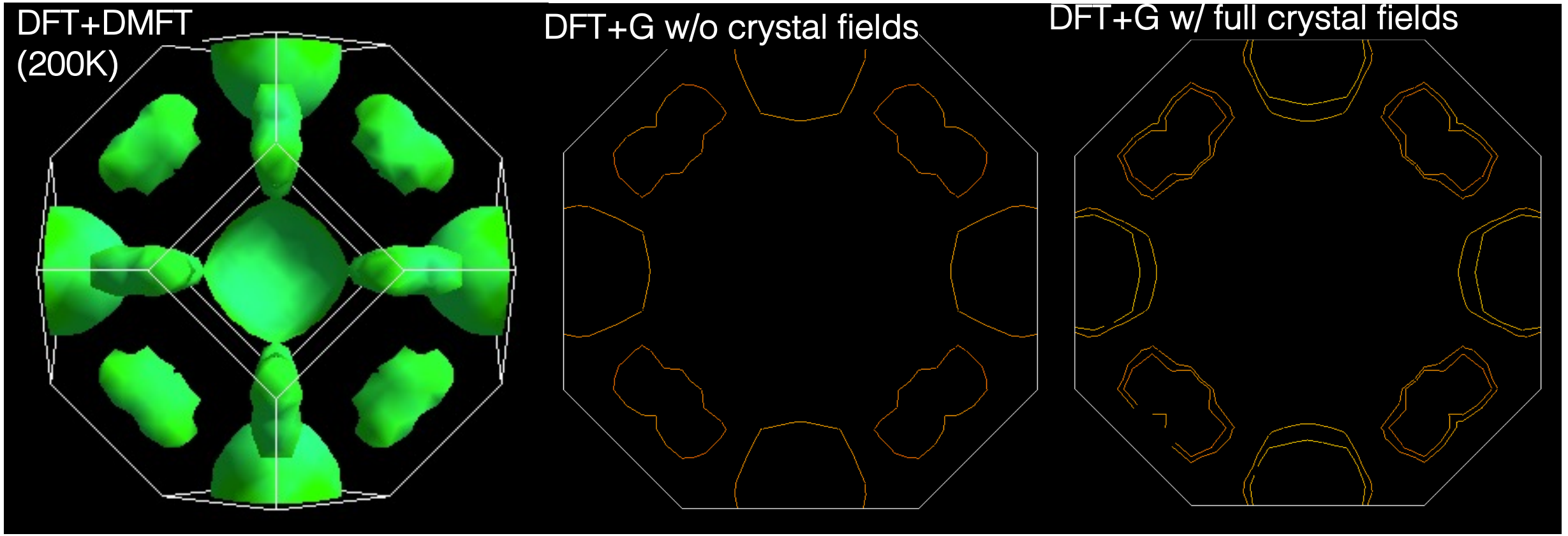
  - Then we still throw the remaining off-diagonals away
- This brings crystal field into the interaction tensor, reducing its symmetries!
  - Much harder problem, but tractable with GPU acceleration

Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

6. Crystal field effects: Fermi surface

- Even on the scale of the FS, the crystal field has little affect on delta-Pu
  - (Color represents quasiparticle weight)



Brookhaven National Laboratory

Electronic correlation induced expansion of Fermi pockets in δ-plutonium

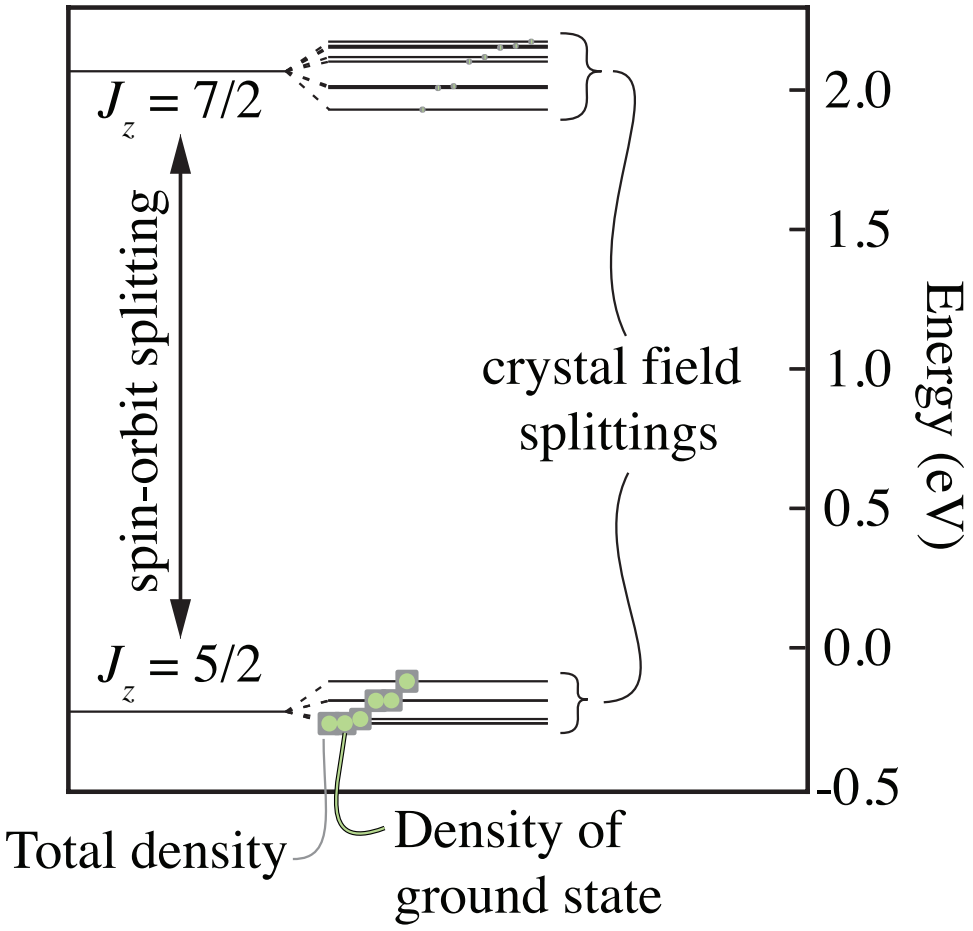
Roxanne M. Tutchton, Wei-ting Chiu, R. C. Albers, G. Kotliar, and Jian-Xin Zhu
Phys. Rev. B **101**, 245156 – Published 23 June 2020

COMSCOPE
from codes to spectroscopies

6. Crystal field splitting dominated by spin-orbit splitting

- Why is the crystal field unimportant in δ-Pu?
  - Crystal field splitting ≪ Spin-orbit splitting
    - (right: delta-Pu DFT+G with full crystal fields)

- What about the less symmetric phases?
  - Lower symmetry = stronger crystal field effects
    - Much more mixing between 5/2 and 7/2 states
    - Less preference for the fully-filled j=5/2 shell
    - A fluctuating magnetism!
      - DFT+G oscillates between a magnetic and non-magnetic state in gamma-Pu
      - (It agrees with DMFT that delta-Pu is non-magnetic)



6. Symmetry adapted bases: a primer

- Generate the local symmetry group: $G = \{g\}$
  - An analysis of the local environment surrounding an atom (or molecule for cluster-X)
- Group theory
  - From any basis, L, we gen generate representations (reps) of that group,
    - E.g., $L = \psi(r)$ is the basis of cubic spherical harmonics, or $L = \mathbb{R}^3$
    - $\hat{D}(g)$ describes the transformation of $\psi(r)$ or [x,y,z] by g
  - Representations can be decomposed into
    - irreducible reps (irreps), $D_k^{(i)}$, with invariant subspaces, $L_k^{(i)}$
      - $\hat{D}(g) = \sum_k D_k^{(i)}(g),\ L = \sum_k L_k^{(i)},\ L_k^{(i)} = D_k^{(i)\dagger} L_k^{(i)} D_k^{(i)}$
  - Symmetry adapted basis has irreps which (block)-diagonalize Hamiltonians!
    - Wigner's theorem:
      - $H_{i\mu jv}^{\alpha\beta} = \int dr \left(\phi_{i\mu}^{\alpha}(r)\right)^* \hat{H} \phi_{jv}^{\beta}(r) = H_{i\alpha\mu j\beta v}\delta_{ij}\delta_{\alpha\beta}$

R. A. Evarestov V. P. Smirnov

Site Symmetry in Crystals

Theory and Applications

Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

6. Generating different representations

1. Get local symmetry group, G
   - Assume quantization axis is along z: the l,Lz or j,Jz basis
   - Get “Wigner D Matrices”
     - Describe the transformation of spherical harmonics under the symmetry operations
   - Only include transformations that commute with magnetization axis (if there is one)
   - This is our rep, $D^{native}(G)$
     - (This is the most common basis choice – tends to work quite well in
2. Generate possible quantization axes for the symmetry group
   - Possible axes come from [x,y,z], [crystal axes], [local symmetry group rotations]
     - Associated transformation of our rep is some unitary matrix U
3. Characterize associated reps & check that they’re symmetry adapted
   - $D^{adapt}(G) = U^{\dagger} D^{native}(G) U$
   - Use group theory to decide if it is symmetry adapted

7. Towards Predictive DFT+ DMFT

- DFT+DMFT w/ CTQMC as “exact”:
  - CTQMC has strong “feelings” about the impurity problem generated by DMFT
    - If it doesn’t “like” the problem (and basis): Time-to-solution becomes near infinite
    - Otherwise, it happily converges to the exact solution
  - Tends to like bases that capture the most important quantum numbers
    - E.g.: Magnetized along x? M_x should be a diagonal operator

- How do we gain confidence in our predictions after truncating the Hamiltonian?
  - Experiment (not a predictive approach)
  - DFT+G?
    - No truncation
    - Find best bases quickly
    - Create benchmarks for DFT+DMFT without truncation of the basis
    - (It also lets us investigate T=0 behavior without a sign problem)

7. Supplementing with DFT+G (& more?)

- On the COMSCOPE roadmap is the ability to *seamlessly* and easily simulate materials within DFT+G and DFT+DMFT
  - Seamless:
    - Same user-friendly interface

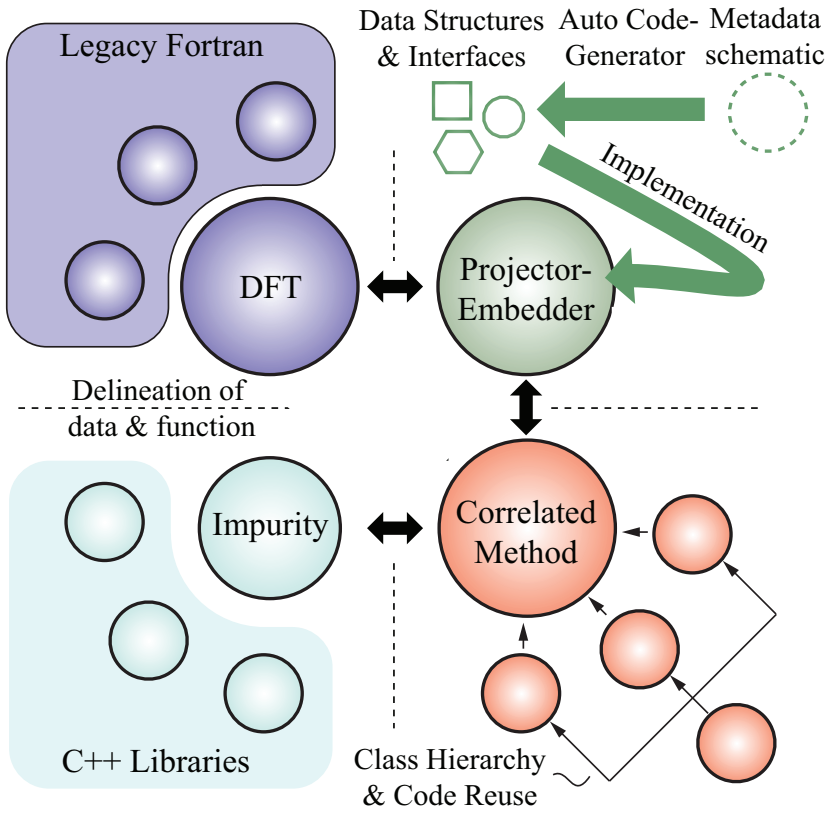
```
dft+g Pu.cif -T10 -U4.5 -J0.512 -N5
dft+dmft Pu.cif -T200 -U4.5 -J0.512 -N5
```

    - Same analysis tools

```
dos; fermi_surface; spectral_functions; band_structure; crystal_fields
```

    - Same software platform
    - Abstracted implementation of framework
      - Projection / Embedding as the cornerstone

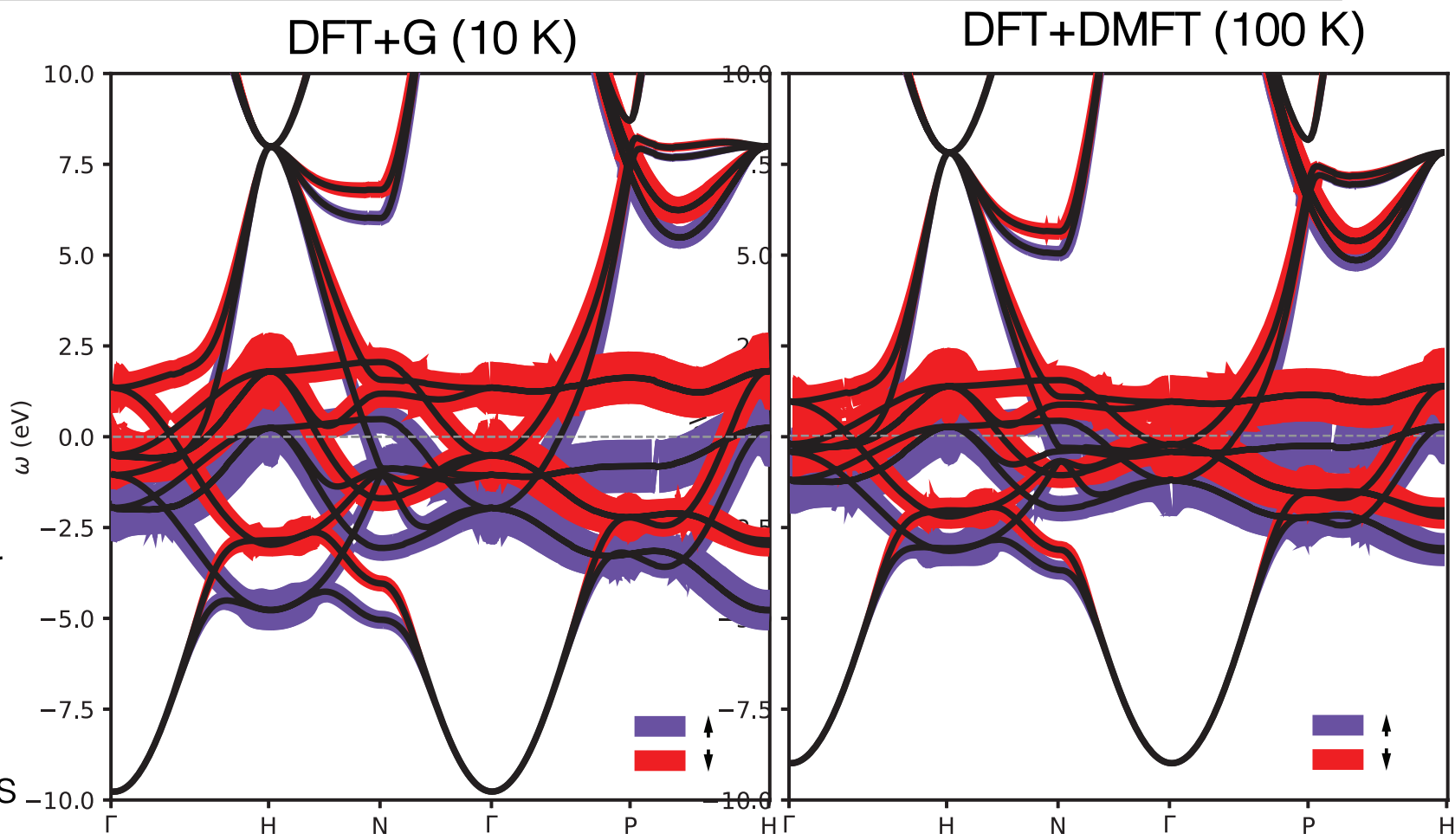
- **Ultimate goal**: Framework for *general* quantum embedding problems and seamless simulation / analysis



7. Comparing DFT+G and DFT+DMFT: Ferromagnetic Fe

- Ferromagnetic Fe
  - no spin-orbit coupling
  - DMFT requires no truncation – but we use ising interaction for convenience
  - U=5, J=0.7, N_0=5.1

- (right) Quasiparticle band structures with spin character
- Similar
  - Magnetization
    - 2.6 vs 2.2
  - Bands / spectral-functions
  - Spin character

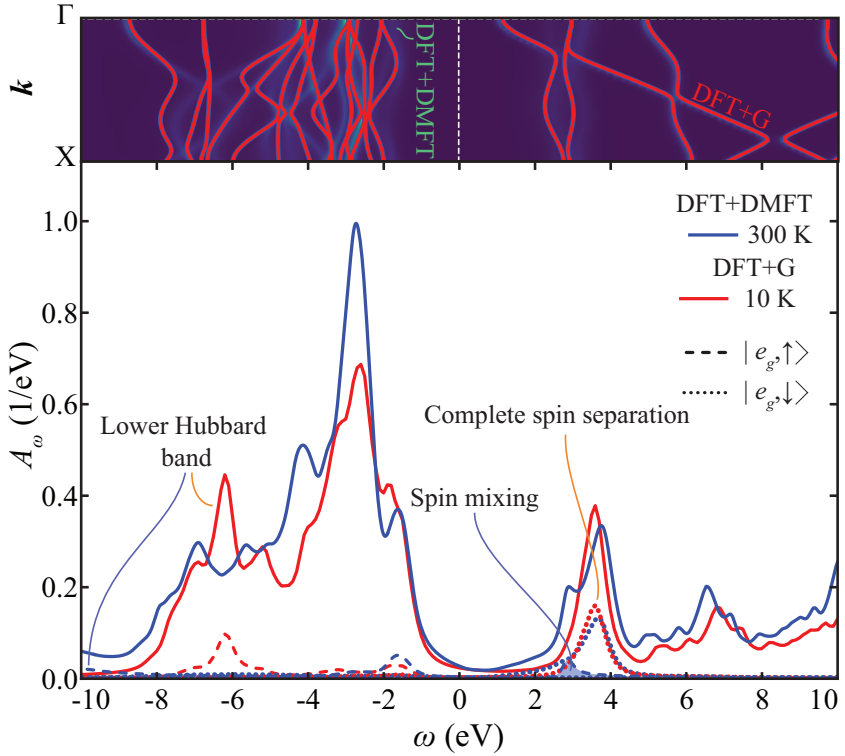


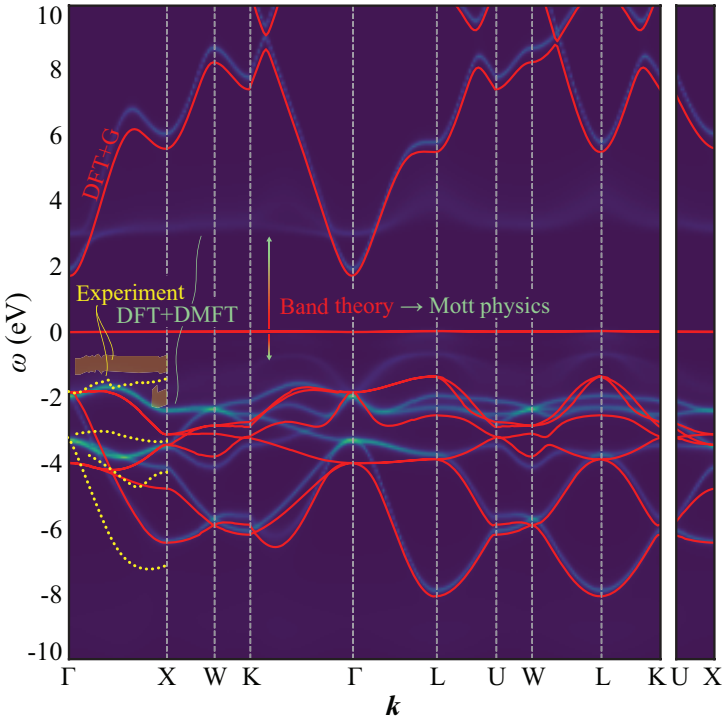
Brookhaven National Laboratory

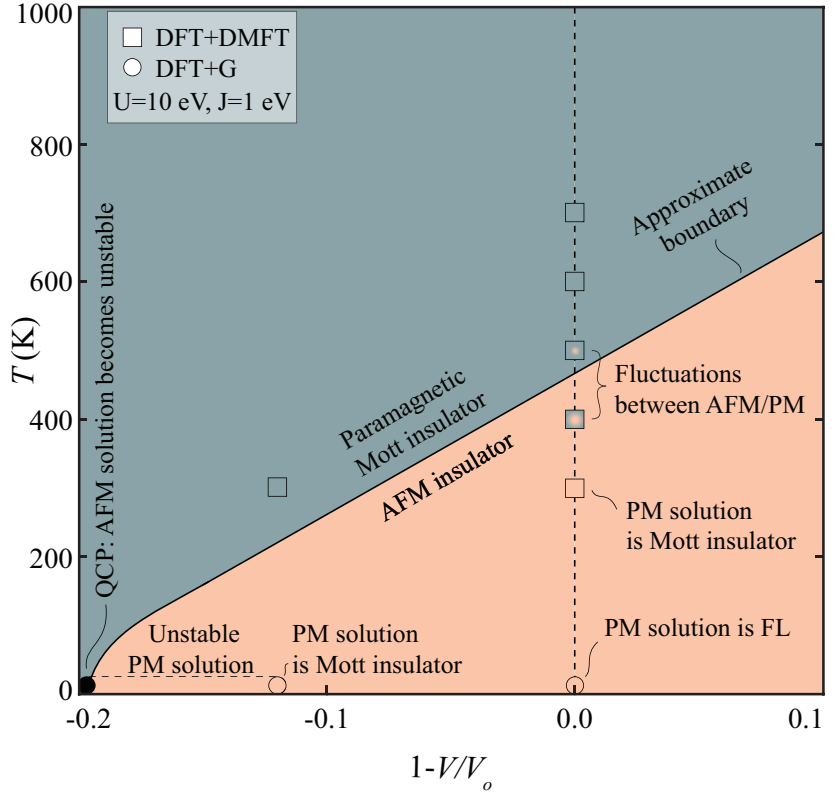
comscope
from codes to spectroscopies

7. Comparing DFT+G and DFT+DMFT: AFM & PM NiO

- DFT+G provides a look at the T=0 phase diagram
  - Extremely accurate AFM prediction compared against DFT+DMFT
- DFT+DMFT provide a look at the temperature driven transitions, mott physics





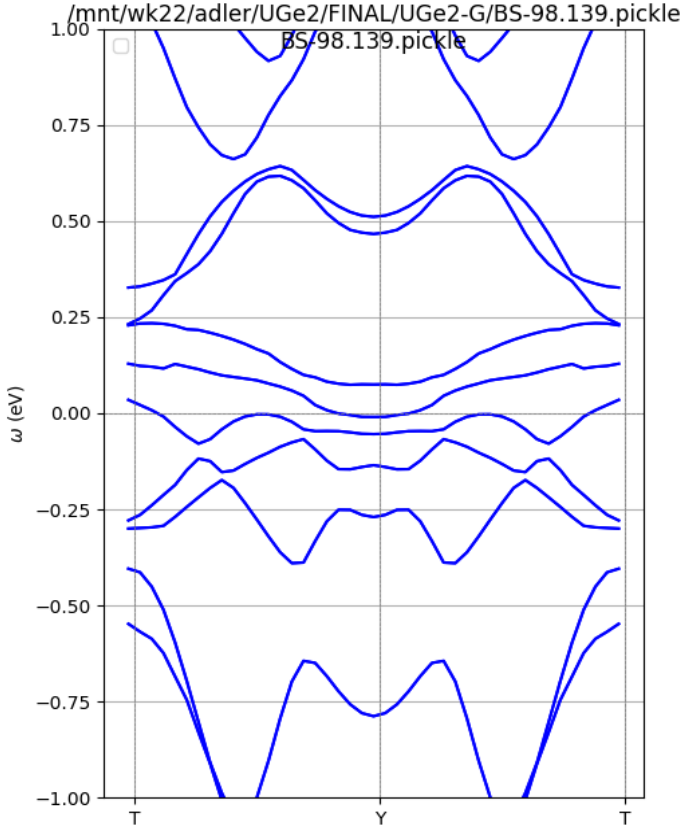


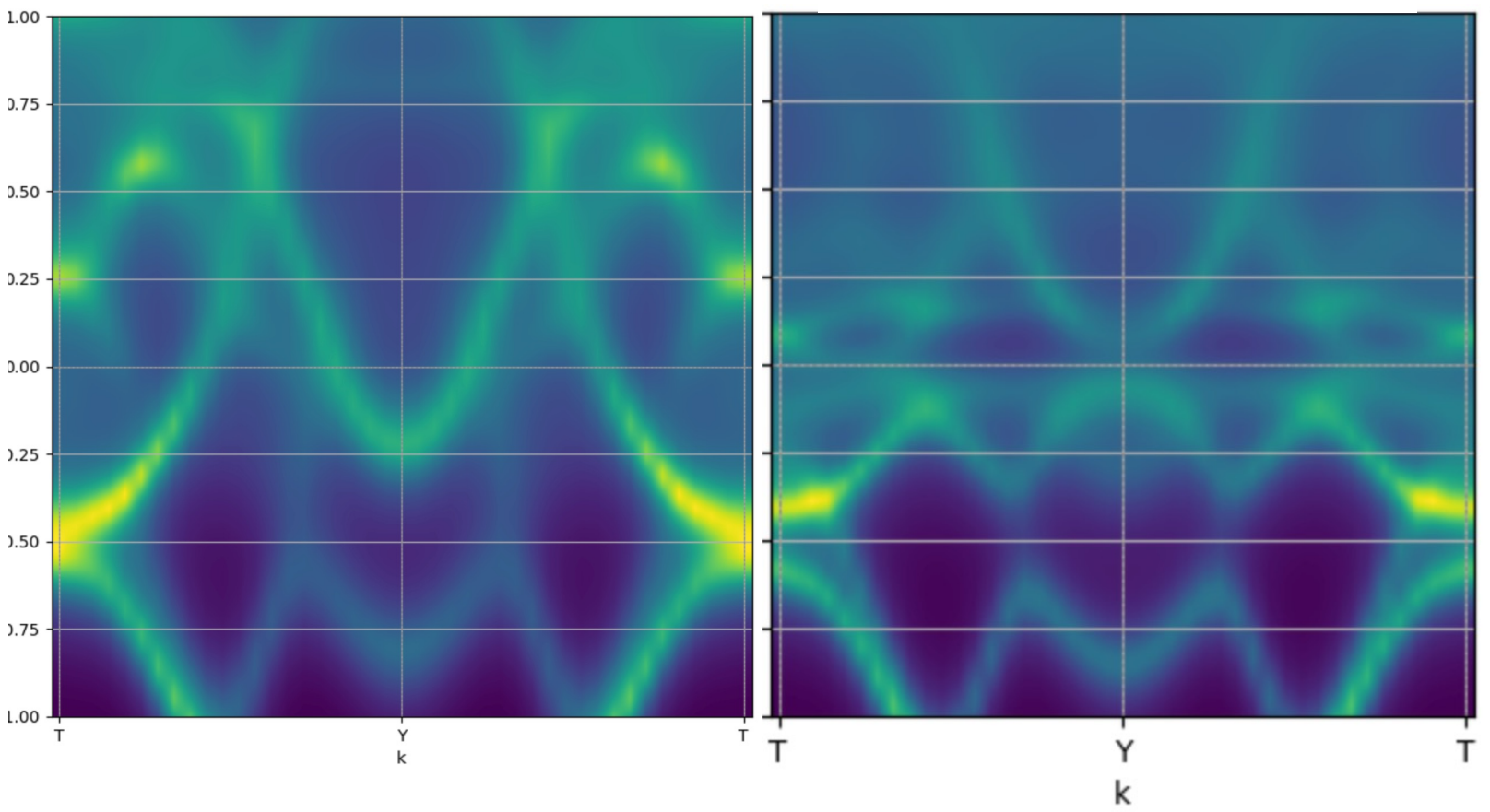
Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

7. UGe2: Finding bases for DMFT in high crystal fields

- Jz basis vs. Mx basis (DFT+DMFT) as compared to the “full” solution (DFT+G)





Brookhaven National Laboratory

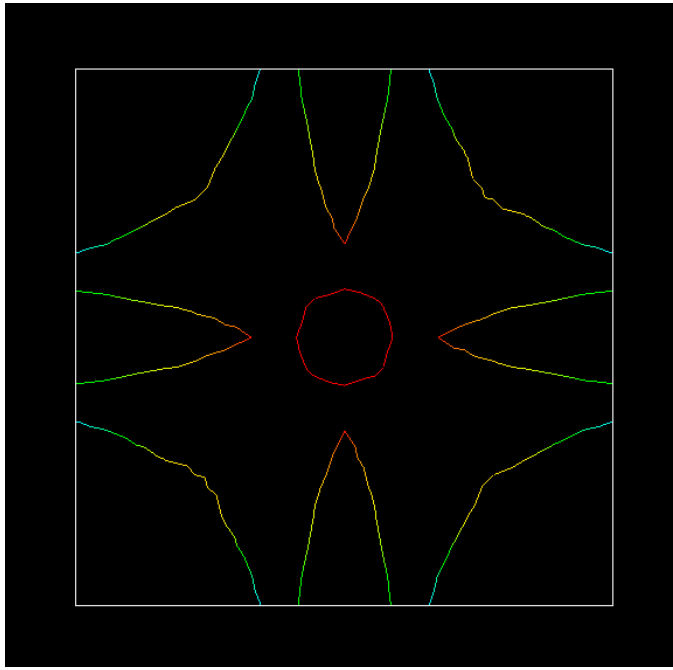
COMSCOPE
from codes to spectroscopies

7. Simultaneous Extension of Methods

- Example one: symmetry adapted bases
  - Simultaneously implemented for all methods!

- Example two: Wanted to implement DFT+CDMFT
  1. Implementation all in the projectors!
  2. This means implemented DFT+Cluster-X
     a) DFT+Cluster-DMFT
     b) DFT+Cluster-Gutzwiller

  - Right: (folded) Quasiparticle weight on Fermi surface in cuprate superconductor within DFT+CGutz



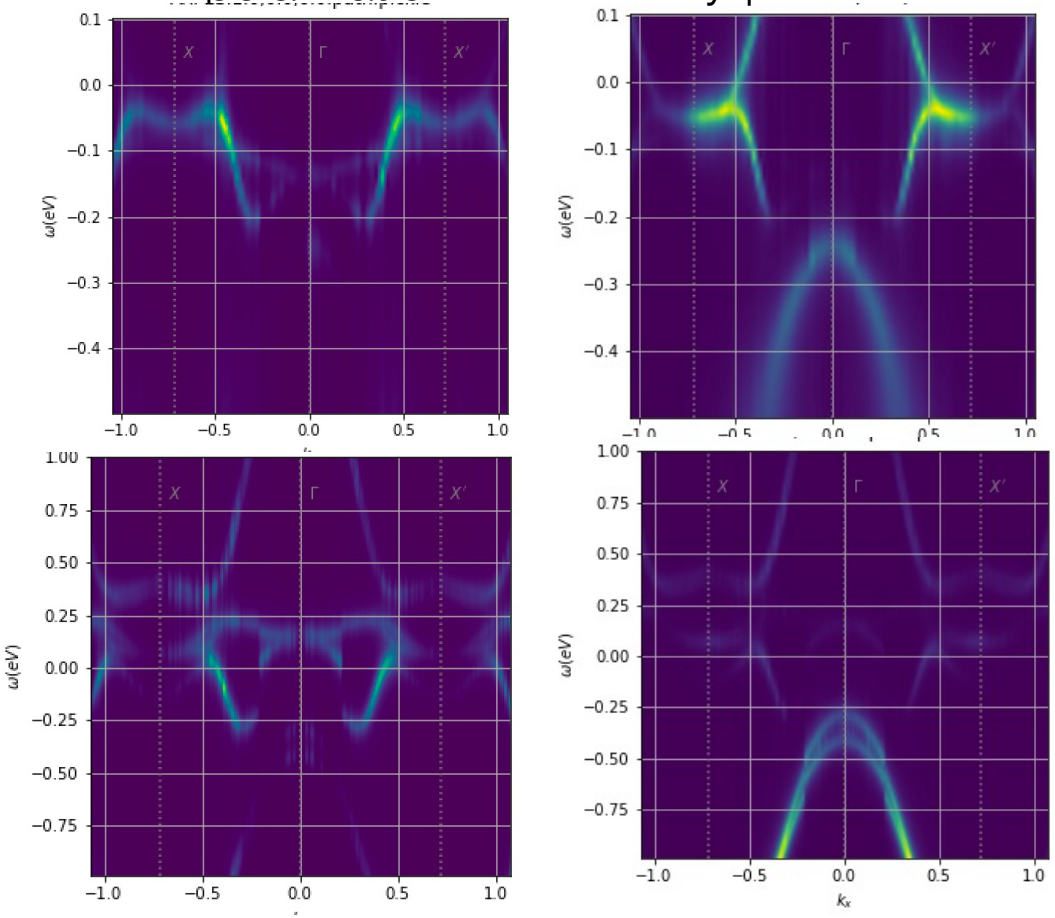
- Example 3: qpGW+Gutz
  - Sangkook Choi will be talking about (qp)GW+DMFT
  - When we bring that into our framework, we will also get qpGW+G, nearly for free

Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

7. Simultaneous Extension of Analyses

- ARPES (spectral functions with with polarized light-electron interaction)
- delta-Pu
  - DFT+DMFT
  - DFT+G
    - (w/ full crystal fields)
- x-polarization
- y-polarization



- Equations
  - ARPES intensity

$$I(\boldsymbol{k},\omega) = \boldsymbol{M}^{\dagger}\boldsymbol{A}(\boldsymbol{k},\omega)\boldsymbol{M}$$

  - Matrix elements

$$\begin{aligned} M_m(\boldsymbol{k}) &= \langle \phi_f^{\boldsymbol{k}_f} | \boldsymbol{\epsilon}\cdot\boldsymbol{p} | \psi_m^{\boldsymbol{k}} \rangle . \\ &= \langle \phi_f^{\boldsymbol{k}_f} | \sum_n |\psi_n^{\boldsymbol{k}}\rangle \boldsymbol{\epsilon}\cdot\langle \psi_n^{\boldsymbol{k}} | i\boldsymbol{\nabla}_k | \psi_m^{\boldsymbol{k}} \rangle \\ &= [\boldsymbol{KV}]_m, \end{aligned}$$

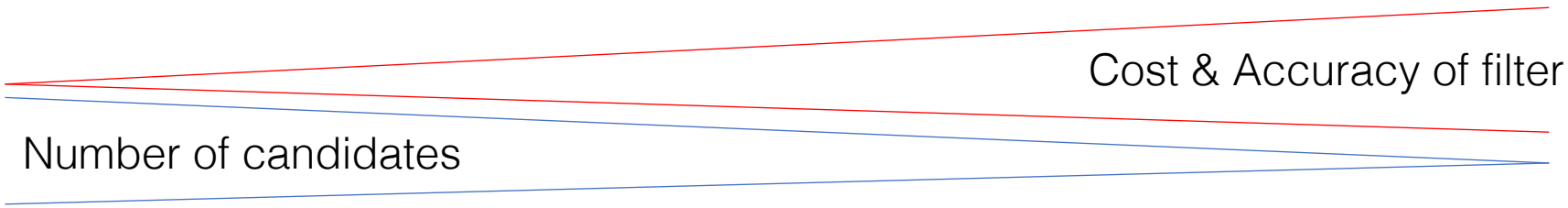
Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

7. Usability in Material Design

- High-throughput & material design require *thousands* of possible combinations are tested
  - Fast method like DFT+G must be used to refine structures & select promising candidates
  - More accurate methods must be used to validate or test at finite temperatures those candidates, narrowing them down
  - The most accurate methods may be used on a final, select few

  - Each method or level of approximation as a filter!



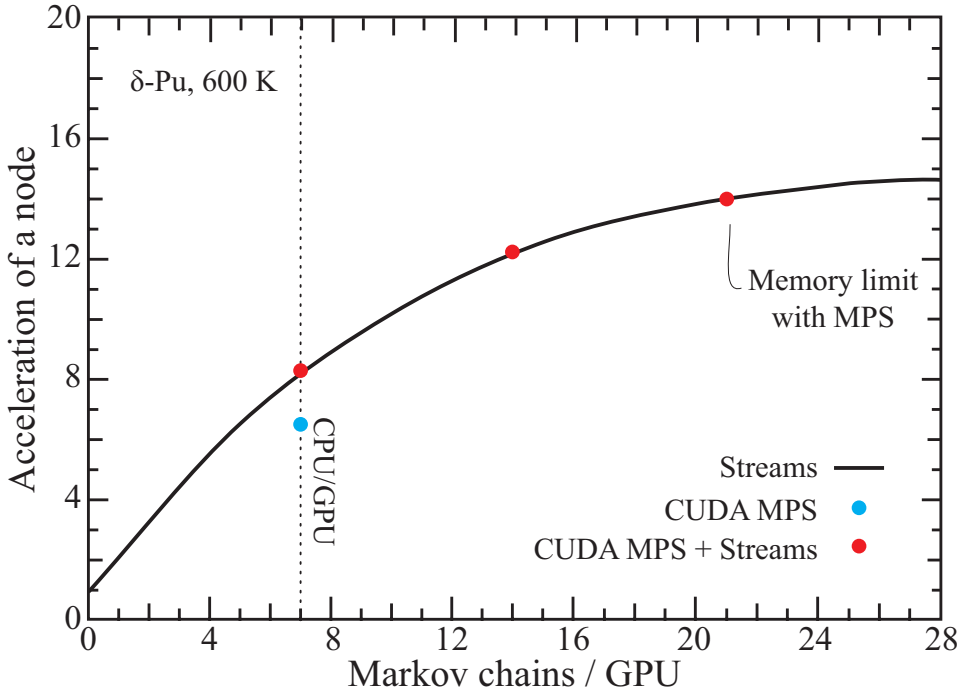
8. Thanks for you time! Questions?

3. Accelerating ComCTQMC: Challenges & Solutions

1. One Markov chain does not fully utilize GPU
   - Matrices are not large enough
   - ➢ Handle multiple Markov chains simultaneously

2. Memory limitations on GPU
   - Multiple CPU's using the same GPU cannot share GPU memory
   - ➢ One CPU simulates multiple Markov chains
   - ➢ CUDA Streams instead of MPS

3. Memory allocation is synchronizing
   - Cannot allocate memory for new products and sub-products on GPU during computation
   - ➢ Allocate whole device on startup
   - ➢ Parcel out memory ourselves



Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

3. Accelerating ComCTQMC: Challenges & solutions

4. Memory restrictions
   - ➢ Avoid multiple CPU's using the same GPU
   - ➢ Use mixed-precision
     - Single (half) precision take ½ (¼) the memory

Mixed Precision

- GPU has dedicated cores for different levels of precision
  - Double, single, and half precision
    - 8 → 16 → 32 or 100+ TFLOP/s per GPU

- Accumulate results with double precision but multiply single or half-precision matrices!
  - (Half-precision does not have the dynamic range to handle real materials, sadly – perhaps when NVIDIA implements a different half-precision float)

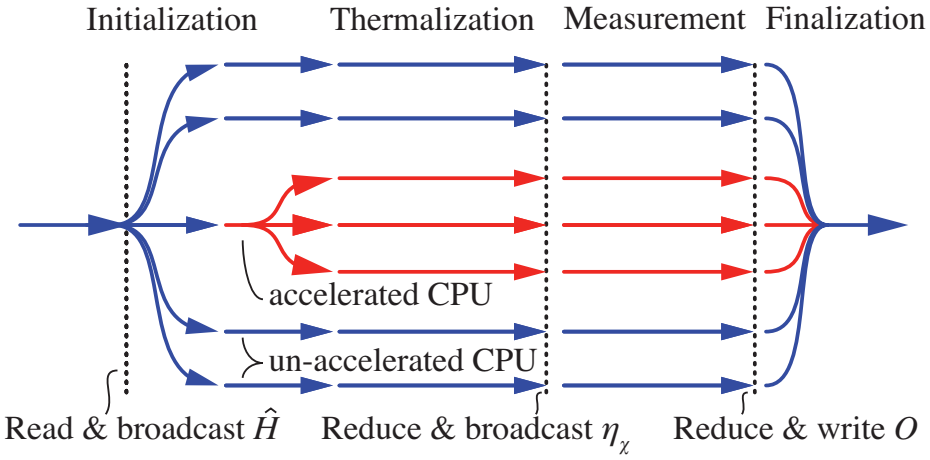
Brookhaven National Laboratory

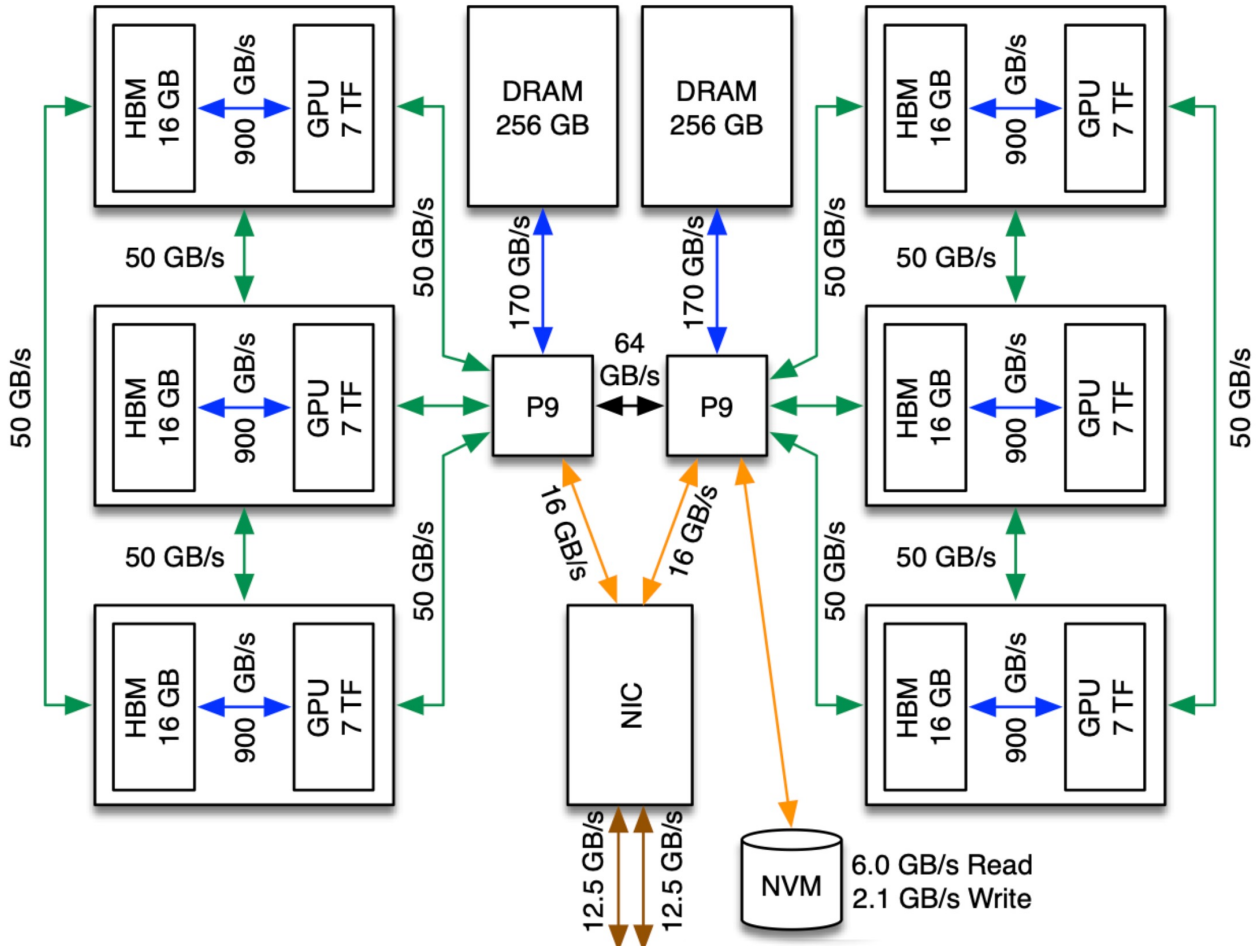
COMSCOPE
from codes to spectroscopies

3. Utilizing HPC

- A lot of CPU's
  - Summit: (2 physical, 42 cores)
- A few GPU's
  - Summit: 6

- Accelerate one cpu / gpu
- Remaining cpu's work as usual, unaccelerated





Brookhaven National Laboratory

COMSCOPE
from codes to spectroscopies

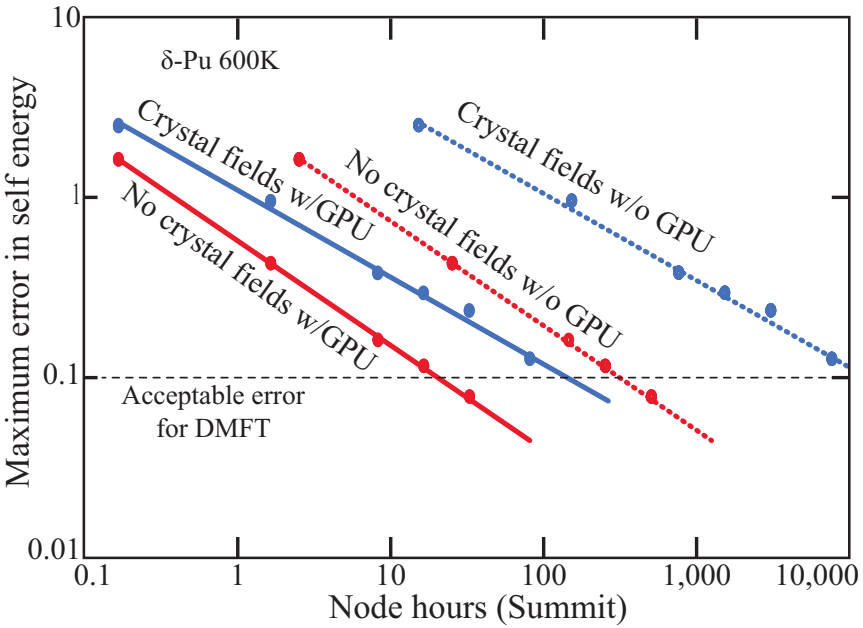
3. Accelerating ComCTQMC:

- GPUs are designed to handle *large* matrix multiplication problems

- In CTQMC
  - Less time is wasted communicating results between CPU and GPU as $r \rightarrow O(1{,}000)$
  - Saturating the GPU becomes easier as $r \rightarrow O(1{,}000)$
  - GPU becomes more efficient as $r \rightarrow O(1{,}000)$

- *The larger the subspaces (and the fewer the quantum numbers), the more the GPU accelerates the CTQMC*

  - BUT! Large subspaces DO NOT HELP CTQMC.
    - CTQMC slows to a crawl with $r_{max} > 300$
    - GPUs help alleviate *some* of the associated burden



BROOKHAVEN NATIONAL LABORATORY

COMSCOPE from codes to spectroscopies

6. The Best Symmetry Adaptation

1. Apply Wigner's theorem
2. Count off-diagonals elements in $H^{\alpha}_{\mu\nu}$ blocks
3. Choose rep with fewest off-diagonals
   - Must discard off-diagonals to solve impurity problem with CTQMC
   - More elements in off-diagonal → harder Gutzwiller
   - (Diagonal of rep always filled with irreps)
4. Choose a rep which (co-)diagonalizes observables of choice
   - We may have physical intuition about the best quantum numbers
     - E.g.
       - Mx in a magnetic material magnetized along x
       - Mz, Lz, Sz in a d-shell material with spin-orbit coupling
         - (In my experience CTQMC *hates* the Jm basis in d-shell problems!)