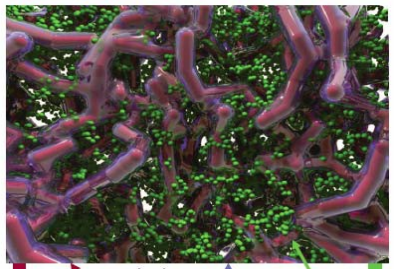
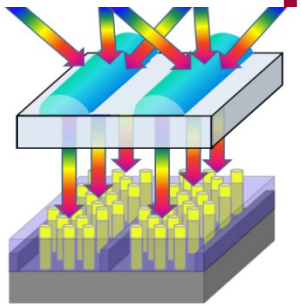
Integrated, multiscale theory/simulation & experiment are key to materials discovery

Modeling and measurement of structure, dynamics, transport must span entire range of time and length scales

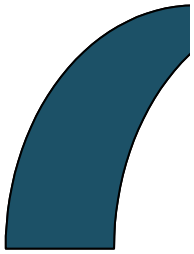




Micro-/Macro-scale

device structure & performance

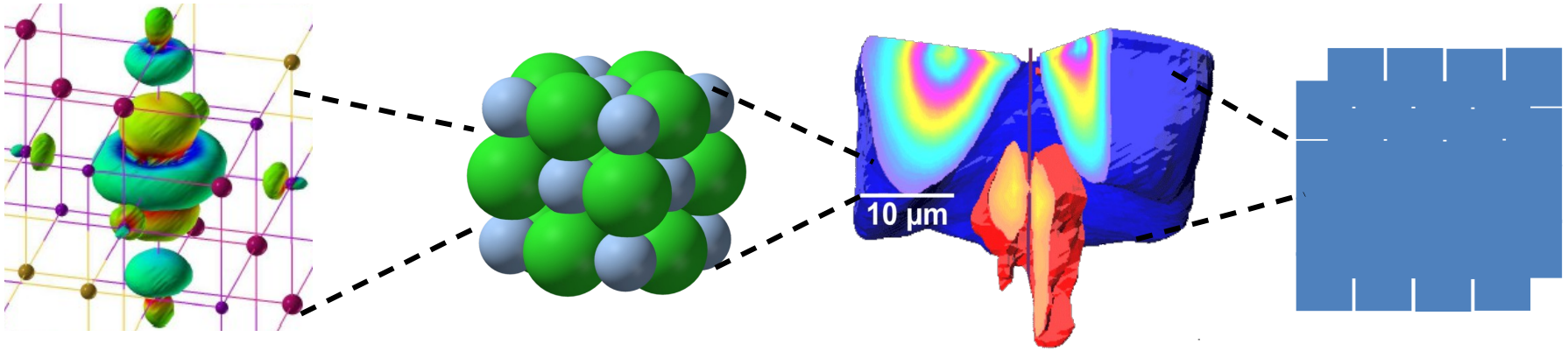
Mori-Tanaka, Halpin-Tsai, Lattice Spring, Finite Element



DFT, QMC, MD, MC, DMFT

Available experimental and theoretical tools cover all spatial ranges

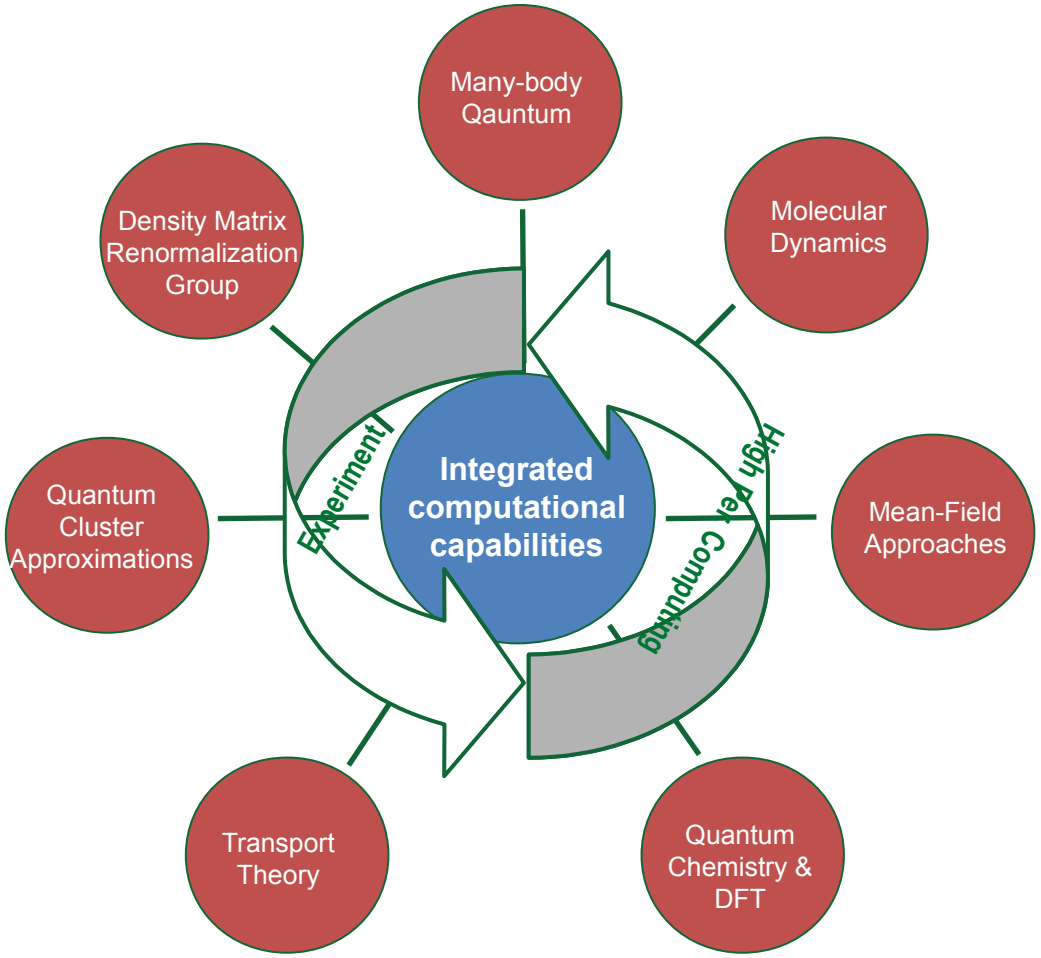
Local probes Aberration-corrected quantitative EM	X-ray /neutron Bragg/ diffuse scattering Electron microscopy Atom probe Neutron inelastic	3D x-ray crystal microscopy Neutron/x-ray texture/strain/width methods	Properties/mec hanical testing Neutron/x-ray residual stress



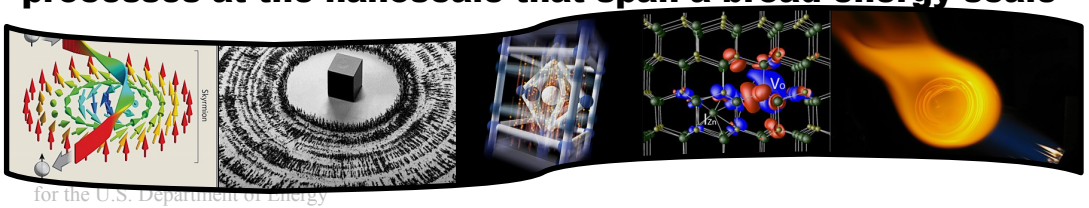
Electronic <0.5 nm	Atomic 0.5-100 nm	Mesoscale 10nm-100 m	Continuum >100 m
Quantum Monte-Carlo Density functional	Molecular dynamics	Plasticity modeling	Finite element

Needs in theory and computation

Requires a close partnership with leadership class computing capabilities and experiments



Must be able to account for coupled physical and chemical processes at the nanoscale that span a broad energy scale



Functional Materials Design: Developing a fundamental understanding of active/reactive control over physical and chemical properties to build responsive matter

- Developing and advancing the ability for understanding and rational tuning of transport (electron, spin, ion, molecule), reactivity and electronic structure
- Developing methodologies for theoretical and computational nanoscience to establish new capabilities and to enhance links with experiment
- Building the scientific foundation (software, algorithms) to study and design functional correlated electronic materials (such as superconductors)
- Advancing soft matter theory and simulation for understanding morphology, stability, dynamics, and properties of topologically complex multiblock and charged copolymers, brushes, composites and blends

Essential components for a materials innovation infrastructure

Computational tools

- Create user-friendly, validated models of materials behavior
- Implement a framework for code development, maintenance, and deployment across academia and industry
- Apply tools to screen materials and develop materials data sets

Experimental capabilities

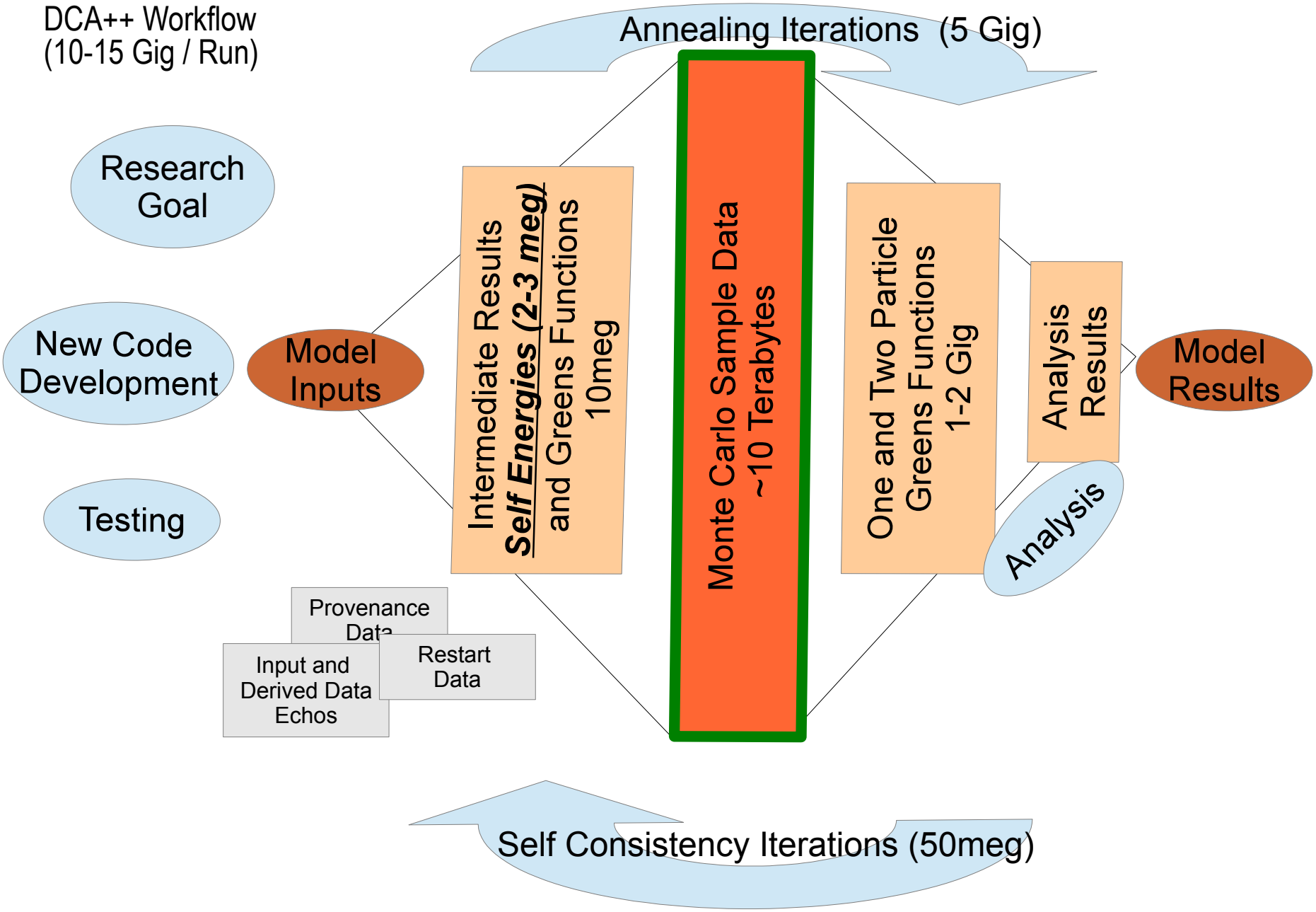
- Integrate experimental results with modeling to validate computer models and accelerate materials screening and development
- Leverage experimental capabilities and facilities to develop/qualify data sets that bridge gaps in understanding and support advances in modeling

Data systems

- Establish materials data storage and sharing systems that respect proprietary information while facilitating incorporation of new data in models and across the materials lifecycle

Key Science Drivers			Network Needs
Science Instruments and Facilities	Process of Science	Data Set Size	WAN/LAN Transfer Time
Current			
• Electronic structure calculations, quantum many-body simulations, molecular dynamics (atomistic and mesoscale), mean field approaches, statistical simulations on large computer clusters and capability computing systems • Quantum transport in nanostructures and virtual STM imaging • Self-assembly of nanostructured materials • Statistical physics in non-equilibrium systems • Emergent behavior in strongly correlated electronic systems	• Users and staff generate output from calculations or experiments. Results based on electronic structure, Monte Carlo, mean field theory, and dynamics calculations can be a significant fraction of the memory available on the computer platform. • Users will generally desire to transfer data files back to their home institutions for future reference, analysis, etc. • Memory ranges from several hundreds of GB on clusters to tens of TB on capability machines.	• Simulations can generate several hundred GB of data • Restart files, production data/results, (10-100) depending on the type of calculation • **Case 1**: *CCSDT/aug-cc-pvtz* for a heteroatom supramolecular complex of 60-100 atoms (100-500 GB); • **Case 2**: *molecular dynamics* for a million atom carbon system with a trajectory dump for the **(p,q)** every 10 fs over a time of 100 ps (100GB-1TB). • **Case 3**: *plane wave periodic electronic structure calculation*s for heteroatom supercells of 100-300 atoms (10-50 GB) • **Case 4**: *DCA++ calculations* (5-10 GB) • **Case 5**: *QMC* for 100 + atom systems (10-100 GB) • **Case 6**: *DMRG* calculations on GMR (1-5 GB) • **Case 7**: *experimental data from scanning probe microscopies*, 5-30 GB • All seven of the case studies will increase in the amount of data produced as faster	1-10 Gbps

Question 1:
DCA++ Workflow
(10-15 Gig / Run)



Question 1:
Work Flow Short Cuts

- Stored self energy file means that we can skip to the end of an annealing run
- Stored two particle Greens functions mean that we can do analysis with out regenerating large datasets.

Question 2:
Data Types and
Generation Rate (Simulation)

- MetaData is integral but a small fraction of the volume
- We generally analyze/process the data as we go
  so we don't need to keep really the really large datasets.
- We usually could use more main memory
  rather than more storage.
- Self Energies could be kept in a database and
  used to kick-start new runs.
- Keeping old runs does add up over time.

Questions 3,4,5:
Where do we keep data,

- Home (not much):
  - codes, provenance, final results …
- Work (90%):
  - Directories containing the output files of Current runs (~10 gig each)
- Archival:
  - Supporting comparision runs for automated testing
  - WaveFunctions (access often due to low bandwidth to hpss)
  - SelfEnergy Databases
    - If we were to use it we would access it once per run to get an initial Self Energy (2-3 meg)