Future role of DMR in Cyber Infrastructure

D. Ceperley

NCSA, University of Illinois Urbana-Champaign

N.B. All views expressed are my own.

What will happen during the next 5 years?

What will the performance be used for?

- Community is large and diverse
- Many areas
- Large movement to local clusters and away from large centers.
- Impact of new computing trends: consumer devices, grids, data mining,...

- Will increase CPU time by >32
- Memory increases similar
- Some methods (eq. simulations) are not communication/memory bound. They will become more useful.

After Schroedinger, Maxwell, Boltzmann,... we know the model. **A key scientific problem.** We have a difficult computational/mathematical problem.
(attention: reductionist approach! There are other approaches represented within ITR.)

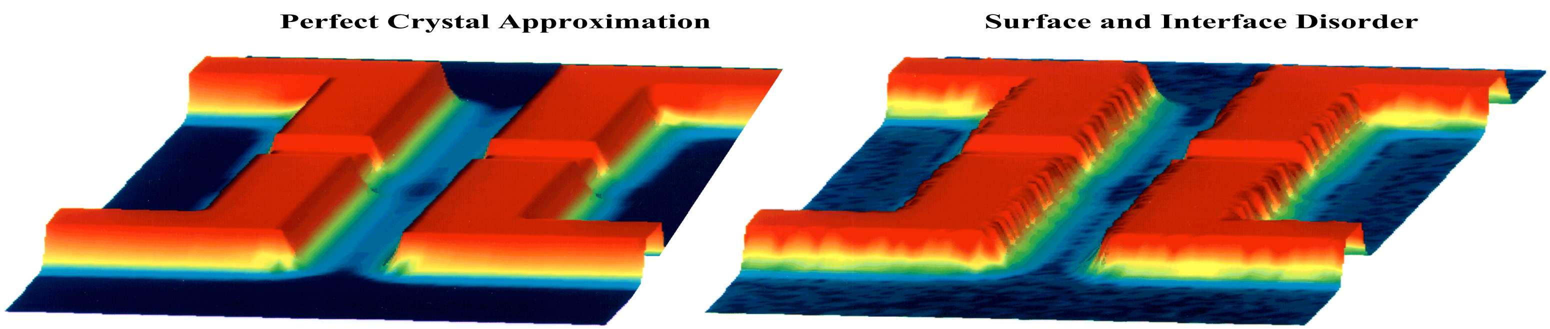
4 main predictable directions

1. Accuracy

- Typical accuracy today (systematic error) is 1000K.
- Accuracy needs to be 100K to predict room temperature phenomena.
- Simulation approach only needs 100x the current resources if systematic errors are under control and efficiency maintained.

2. Larger Systems

- Complexity of simulation methods are similar. Ranging from O(1) to O(N). Only some methods are ready.
- But simulations are really 4d -- both space and time need to be scaled.
- 10^4 increase in CPU means 10-fold increase in length-time scales. Go from 2nm to 20nm.
- But this is interesting-features of molecules, nanowires,.. come into range.



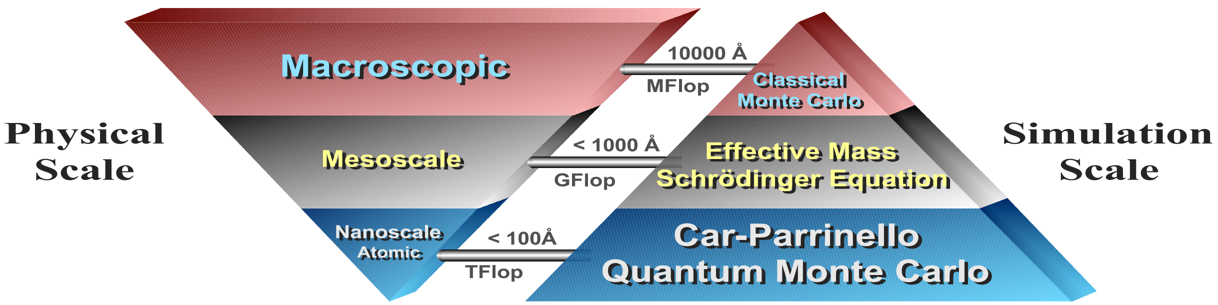
3. More Systems

- What is feasible are parameter studies.
- Typical example: **materials design**.
- Combinatorics leads to a very large number of possible compounds to search. [>92^k]
- But it is starting to take place: Morgan,..
- Needs both accurate QM calculations, statistical mechanics, multiscale methods, easily accessible experimental data,… Interdisciplinary!

4. Multiscale

Challenge is to integrate what is happening on the microscopic quantum level with the mesoscopic classical level.

- How to do it without losing accuracy?
  - QMC /DFT
  - DFT-MD
  - SE-MD
  - FE
- How to make it parallel? (Load balancing with different methods)



Multiscale Hierarchy

Lots of software/interdisciplinary work needed. Important progress reported here.

Computational funding modes

1. Large collaborations (medium ITR's)
   - Needed for multidisciplinary/large projects
2. Algorithmic research (small ITR's)
   - Fits into "scientific/academic" culture.
3. Cycle providers (NSF centers&local clusters)
   - "time machine", for groups not having their own cluster or having special needs
4. Software/infrastructure development
5. Education in CI

} unmet opportunities

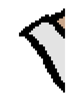
Software/infrastructure development motivation

- Why are some groups more successful than the US materials community?
  - Europeans (VASP ABINIT …), Quantum Chemistry, Lattice Gauge theory, applied math,…
- Does not fit into the professional career path as well.
- Software is expensive
  - We need long term, carefully chosen projects
- Unlike research, the effort is wasted unless the software is, documented, maintained and used.
- Big opportunities: my impression is that the state of software in our field is low. We could be doing research more efficiently.
- Basic condensed matter software needed in education.

Software/infrastructure development

- Support development of tested methodology, including user documentation, training, maintenance (e.g. codes from medium and small ITRs reported here.)
- Yearly competitions for small (1 PDRA) grants
- Standing panel to rank proposals based on expected impact within 5 years.
- Key factors in the review should be communication with actual users of the software and experts in the methodology.

Dan Reed's observations

- **Research**
  - let a thousand flowers bloom, then pick the beautiful ones
    - maximize creative options
  - projects have (relatively) short lifetimes
    - less coordination required among agencies, projects, …
- **Infrastructure**
  
  - invest in "winners" and sustain them at adequate levels
    - maximize community benefit
  - projects have (relatively) long lifetimes
    - strategic, long-term coordination and planning required

Education in Computational Science

- Need for ongoing specialized training:workshops, tutorials, courses
  - Parallel computing, optimization
  - Numerical Libraries ad algorithms
  - Languages, Code development tools
- Develop a computational culture and community
- Meeting place for scientists of different disciplines having similar problems
- Reach a wider world through the web.
- Large payoff for relatively low investment

Databases for materials?

- We need vetted benchmarks with various theoretical and experimental data
- Storage of all the outputs?
  - What is balance between computation and storage? Computed data is perishable in that the cost to regenerate decreases each year and improvements in accuracy mean newer data is more reliable.
  - Useful in connection with published reports in testing codes and methods. Expanding role of journals?
  - need to handle "drinking from firehose." This could be handled by XML based data structure (standards) to store inputs and outputs.