

# BLUE WATERS

SUSTAINED PETASCALE COMPUTING

## *Exscale – when will it happen?*

William Kramer

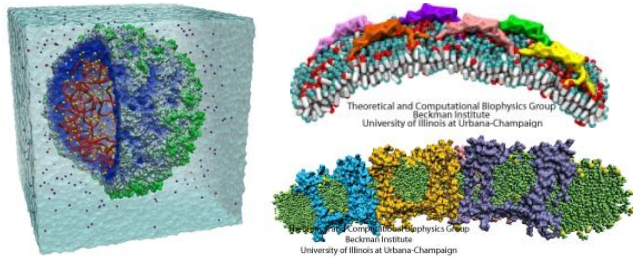
National Center for Supercomputing Applications



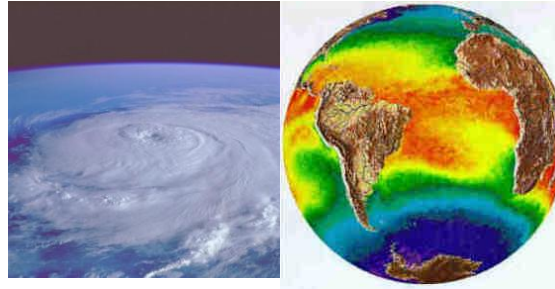
GREAT LAKES CONSORTIUM  
FOR PETASCALE COMPUTATION

## *Sustained Petascale computing will enable advances in a broad range of science and engineering disciplines:*

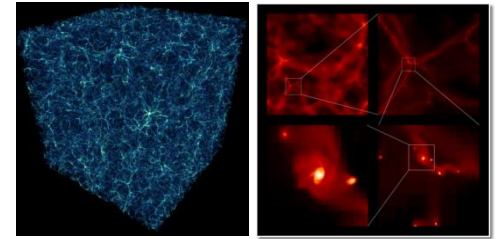
### Molecular Science



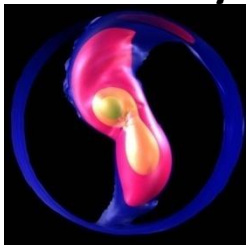
### Weather & Climate Forecasting



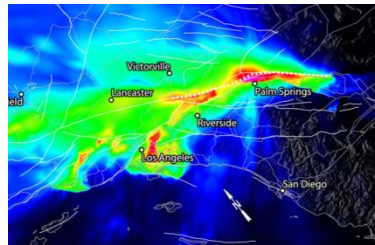
### Astrophysics



### Astronomy



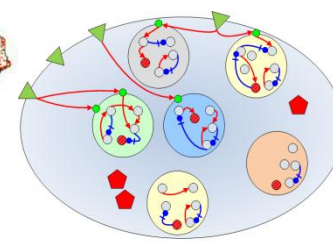
### Earth Science



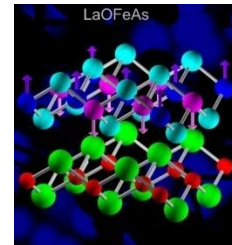
### Health



### Life Science



### Materials



# Blue Waters Computing System

System Attribute	Ranger	Blue Waters
Vendor	Sun	IBM
Processor	AMD Barcelona	IBM Power7
Peak Performance (PF)	0.579	<b>17</b> >10
Sustained Performance (PF)	<0.05	<b>&gt;20</b> >1
Number of Cores/Chip	4	<b>2</b> 8
Number of Processor Cores	62,976	<b>~3.5</b> >300,000
Amount of Memory (TB)	123	<b>10</b> >1.2
Interconnect Bisection BW (TB/s)	~4	<b>&gt;&gt;10</b>
Amount of Disk Storage (PB)	1.73	<b>&gt;10</b> 18
I/O Aggregate BW (TB/s)	?	1.5
Amount of Archival Storage (PB)	2.5 (20)	<b>&gt;200</b> >500
External Bandwidth (Gbps)	10	<b>&gt;10</b> 100-400

# Observations

- Real sustained PF is ~2 years after peak PF
- Sustained Petascale took O(\$1B) for the first system
  - O(\$1.7B) for the first two systems
- In the best case, Peak EF is probably O(\$2-3B) time frame for the first system
  - O(\$4B) for the first two systems
- There are very real HW and SW challenges for computer design that have to be solved regardless of Exascale
  - Scale (HW, System SW, Application SW)
  - Reliability
  - Balance
  - Communications
  - Middle level SW .....
- With a very few notable exceptions such as Charm++ based codes, as system scale increases, the application base and science team able to use the largest systems decreases proportionally

# Key Issues for Exascale – beyond the technical

- If Exascale is not started soon, development expertise will have to be recreated - adding cost and delay
- Is the best value for Exascale in one system? Is the cost of one system worth it
  - Is Exascale in aggregate worth it (and how big of a building block?, data movement, etc)
  - Should we provide (~5-10) 100-200 PF systems linked in a data infrastructure to do Exascale science problems?
    - More familiar and evolutionary
- The energy debate needs to be put into context with everything else – use true TCO, rather than one component
- It is not clear that people will use an EF system for real science if we build it the way some expect.
  - Need clear – really meaningful metrics for sustain (aka time to solution) performance based on needed use rather than easiest use
    - Not Peak Flops nor Linpack Flops
  - Holistic –  $\sum\{\text{Performance, Effectiveness, Resiliency, Consistency, Usability}\}/\text{TCO}$
- Most discussions on Exascale and even Petascale talk about the crisis in the data deluge, but that is little discussion in the HPC community. So, why should the community be pushing "Exaflops" rather than "Yottabytes" in order to improve science productivity and quality?

## Conclusion

- Essentially whether Exascale can be done in the time frame people are talking about depends how one defines success and on how much money someone spends
- If there is a public will - and it appears there is within parts of the administration but not sure about congress - EF can be done.
  - Not clear there is fundamental motivation for the public needed
  - Not clear industrial policy is sufficient without critical S&E driver(s)
- The schedule (2018), cost (~\$3B) and scope (EF for 20MW that is usable by many) is compromised
- There will be Exascale systems
  - In some time period but not in 2018-2020 unless unprecedented public funding occurs immediately