Data- and Compute-Driven Transformation of Modern Science

Edward Seidel

Assistant Director, Mathematical and Physical Sciences, NSF

Jim Ulvestad

Division Director, Astronomical Sciences, NSF

Profound Transformation of Science

Collision of Two Black Holes

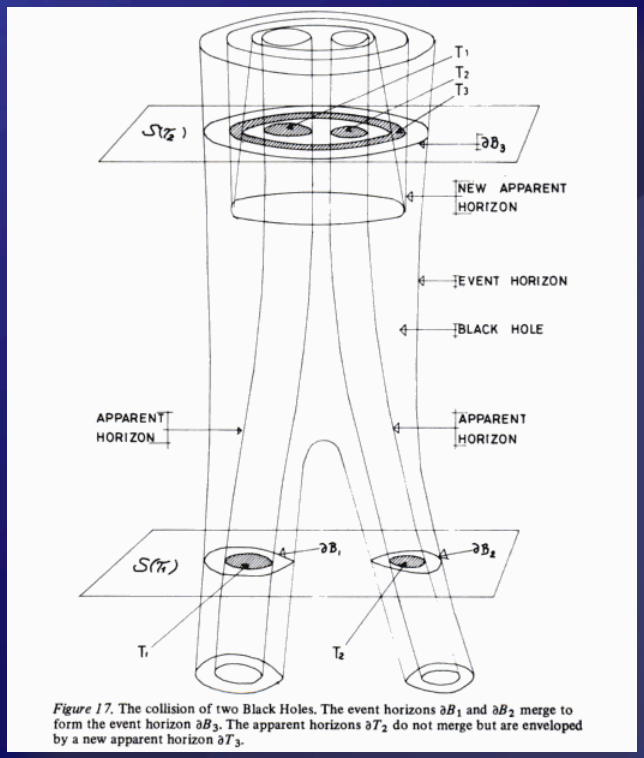
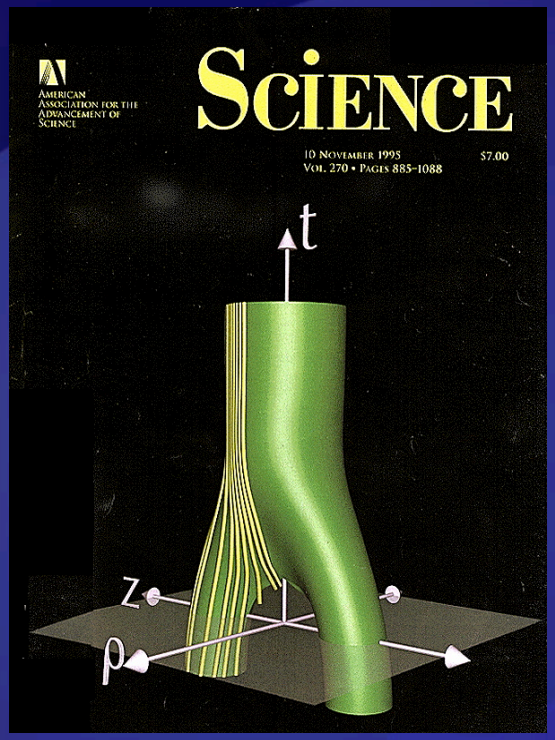
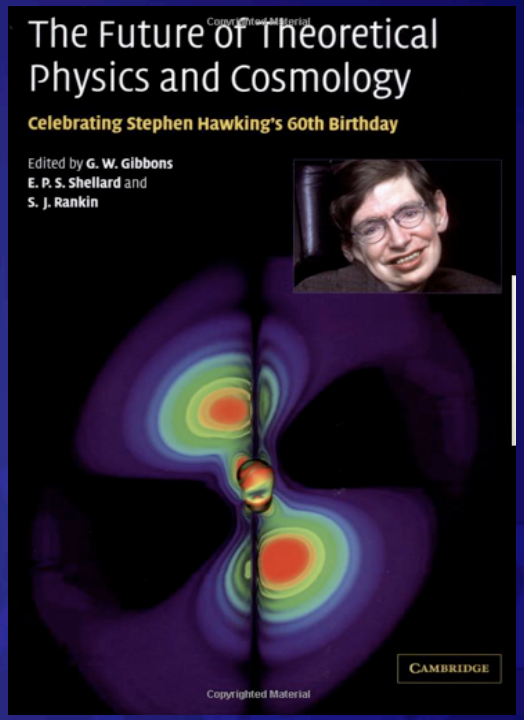


Figure 17. The collision of two Black Holes. The event horizons ∂B_1 and ∂B_2 merge to form the event horizon ∂B_3. The apparent horizons ∂T_2 do not merge but are enveloped by a new apparent horizon ∂T_3.





1972: Hawking. 1 person, no computer 50 KB

1995: 10 people, large computer, 50MB

1998: 3D! 15 people, larger computer, 50GB

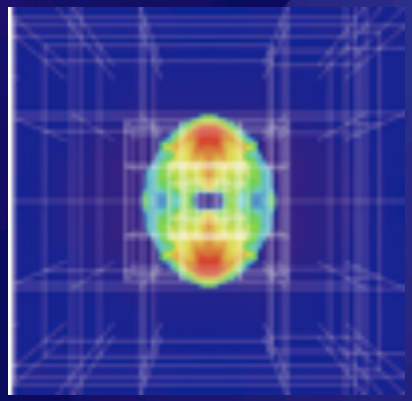
Transient & Data-int

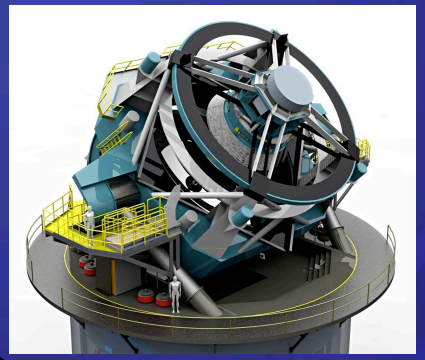
- New era: seeing events as they occur
  - (Almost) here now
    - ALMA, EVLA in radio
    - Ice Cube neutrinos
  - On horizon
    - 24-42m optical?
    - LIG
- Si
phys

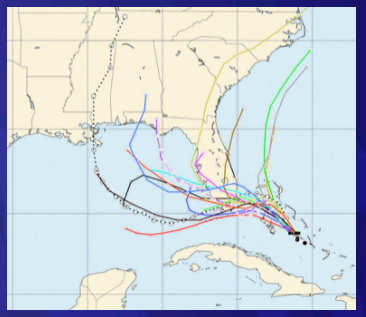
Will require integration across disciplines, end-to-end

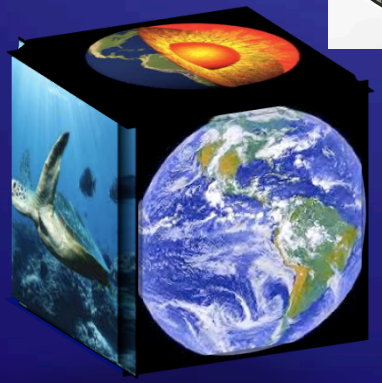
Communities need to share data, software, knowledge, in real time

Scenarios like this in all fields
"*Heroic Age of Digital Observation*"









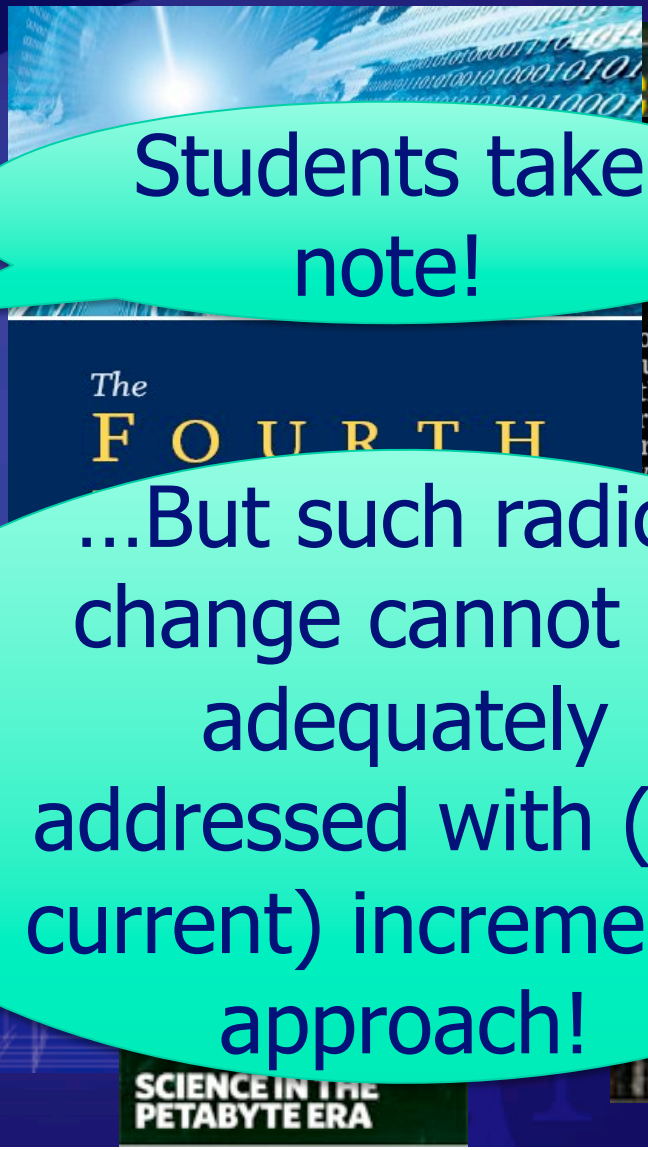
Framing the Challenge:
Society Transformed by Data

- Inte ative, multiscale
  - 4 centuries of constancy, 4 decades 10^{9-12} change!
- Multi-disciplinar Collaborations
  - Individuals, groups teams, communiti
- Sea of Data
  - Heroic Age of Digital Observation

We still think like this…

Students take note!

…But such radical change cannot be adequately addressed with (our current) incremental approach!



The Shift Towards a "Sea of Data"
Implications

- Science & society are now data-dominated
  - Experiment, computation, theory
  - Publication
- Totally new methodologies
  - Algorithms, mathematics, culture
- Data become the medium for
  - Multidisciplinarity, communication, publication…science

Fundamental questions become focused around data: How to remove boundaries? How to incentivize sharing?

How do we attribute credit for this new publication form? How are data peer reviewed? What is a publication in the modern data-rich world?

Changes Coming at NSF for Data!

- Long-standing NSF Policy on Data
  - "*Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the p… gathered in the c…*
- NSF now requi…
  - DMP will be ≤2…
  - DMP subject to pe…
  - It will not be possible to su… …out DMP
  - Customization by discipline, … …n necessary
- Developing unifying data f… …ework for science
  - Should connect globally; dis…ussions underway with EU
- National Science Board beginning to examine policy for access and openness of data and publication

Sharing data, software will be needed for both interdisciplinary work and reproducibility

Recommendation of NSF Advisory Committee on Cyberinfrastructure ACCI

"The National Science Foundation should create a program in Computational and Data-Enabled Science and Engineering (CDS&E), based in and coordinated by the NSF Office of Cyberinfrastructure. The new program should be collaborative with relevant disciplinary programs in other NSF directorates and offices."

NSF can make a strong statement that will lead the Foundation, researchers it funds, and US universities and colleges generally, by recognizing Computational and Data-Enabled Science and Engineering as the distinct discipline it has clearly become.

Approved
Arden L. Bement, Jr.
Director
National Science Foundation

05/27/2010

Date

Critical Lessons to Take Home

- Science and society profoundly changing
- Comprehensive approach needed to address complex problems of 21st century
  - All elements must be addressed, not just a few; can't even start to address problems without all
  - Many exponentials: data, compute, collaborate
- Data-intensive science increasingly dominant
  - Modern data-driven CI presents numerous crises, opportunities
- Academia and Agencies must address
  - Rethinking Academic Structures, Curriculum, P&T
  - NSF Responding through CIF21, changes in implementation of data policy, new programs