Cyborg Journalism:
Reporting from the Front Lines of Scientific Computing

Kavli Institute for Theoretical Physics
Santa Barbara, CA, USA

June 26, 2013

Aaron Dubrow: Science Writer at the
Texas Advanced Computing Center

Who I am and how I got here

Science journalists come in many forms and play many different functions in the knowledge ecosystem.

I am *not* a trained scientist. No PhD. or lab experience. Last formal training as an undergrad.

I *am* a journalist who finds science really interesting as a subject for articles, documentaries and other creative endeavors.

Who I am and how I got here

Not only science.

Also the *people, processes and creativity* that underlie the pursuit and progress of science.

In that sense, I'm much like my readers…

hungry for STORIES!

Who I am and how I got here

This interest took me:

to West Texas to document efforts to protect prairie dogs;

into Newark high schools to study innovative approaches to science-learning;

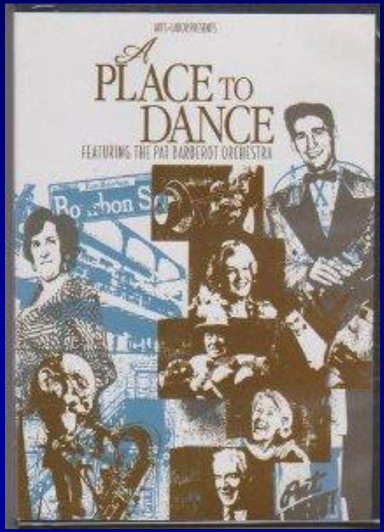
to Lake Okeechobee to report on the draining of the everglades;

and to experimental forests where brain-eating flies were being introduced to control fire ant populations.

(In other words, pretty cool places.)

Who I am and how I got here

During this time, I also co-produced and edited feature-length documentaries about a homeless family, post-Katrina dancing senior citizens and inner-city boys for PBS.



Fairly diverse stuff.

TACC

Who I am and how I got here

Flashback to 2007….

Science Writer position listed at the Texas Advanced Computing Center, based at The University of Texas at Austin

Knowing almost nothing about supercomputers, I apply and get the job.

Who I am and how I got here

What I didn't know then but know now….

The National Science Foundation has insisted on the position as an add-on component to the "*Ranger*" supercomputer.

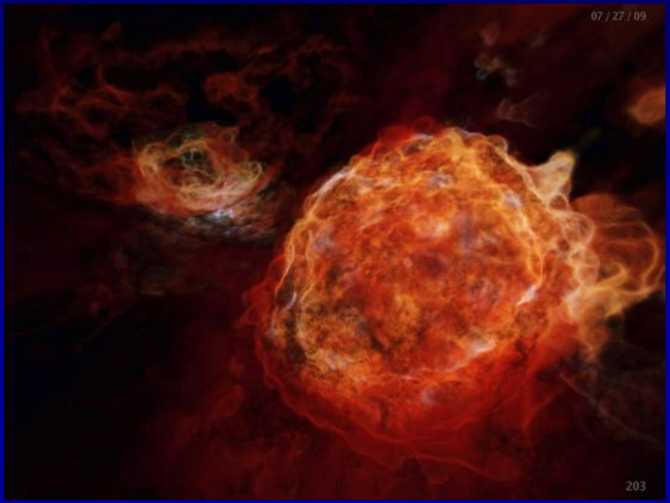
Ranger was a $50 million system that our center had just competed for and won.

Who I am and how I got here

Ranger (and supercomputers in general) needed a publicist!

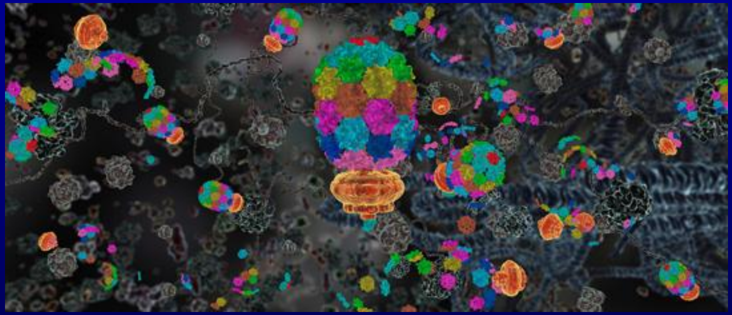
Who I am and how I got here

For the last 6 years, I've worked as an embedded reporter at TACC



Worked to represent the uses and significance of high performance computing (HPC) in 21st century science

Probably have written as many articles about scientific computing as anyone else during that time.



My role in the ecosystem

Many names for what I do:

Science Writer
Public Relations Coordinator
Public Information Officer
Press Officer
PR flack….

Science grunt/
Technical translator/
Knowledge infantryman

TACC

My role in the ecosystem

Many names for what I do:

Press Officer
Science Writer
Public Relations Coordinator
Public Information Officer
External Relations Manager
PR flack….

Or:
Science grunt/
Technical translator/
Knowledge infantryman

My role in the ecosystem

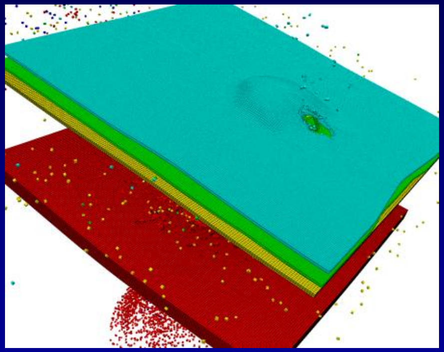
Stigma associated with PIO's among journalists.

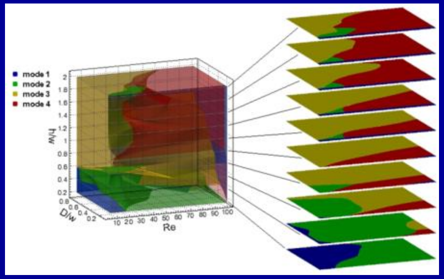
"Not real reporting"
"Hype" (True? Fair?)
"Uninformed"

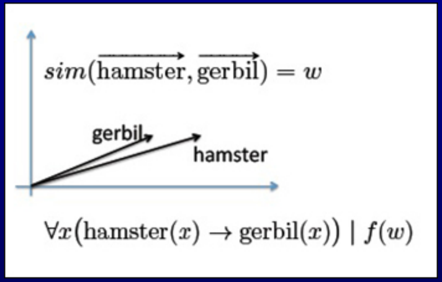
Yet… science journalism is in retreat among newspapers, magazines, TV stations.

Increasingly, news is drawn directly from PIOs. Growing sense of importance to health of science.

A diverse sampling of science







A few of my current and recent articles:

“Shields to Maximum, Mr. Scott”
Researchers use supercomputers to simulate orbital debris impacts on spacecraft and fragment impacts on body armor

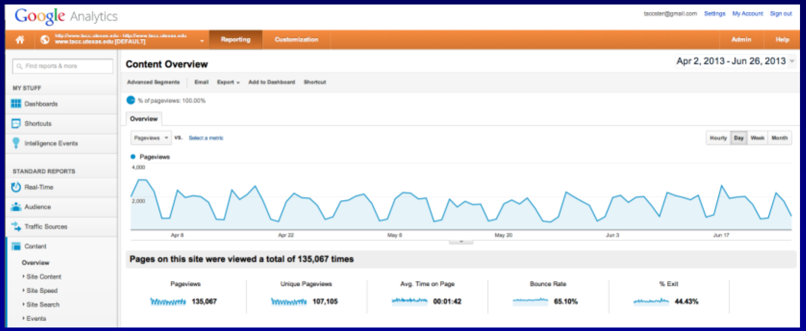
Programming Flow
Supercomputers help microfluidics researchers make waves at the microscopic level

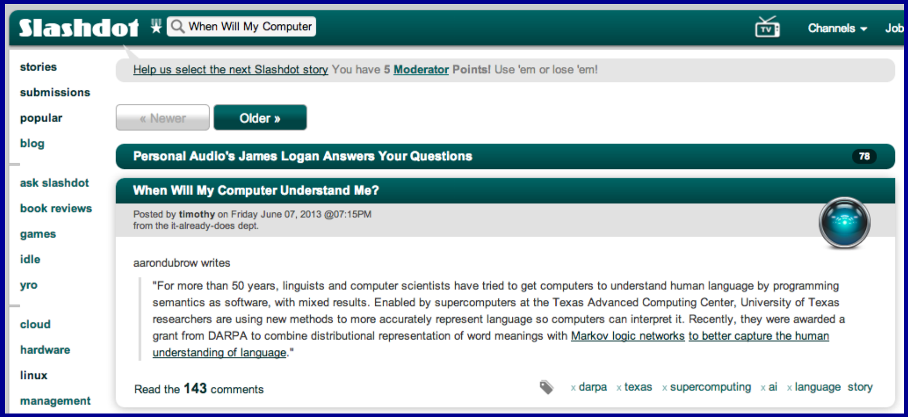
Investigating the Dark Matter of Life
Metagenomic researchers explore ecosystems in oceans and microbes in the human esophagus

When Will My Computer Understand Me?
Linguists and computer scientists use TACC supercomputers to improve natural language processing

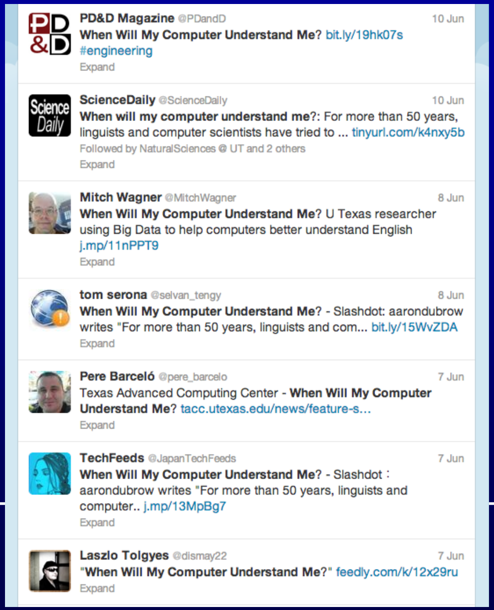
Surprising Audiences

Amazingly, people read them…





Discuss them…

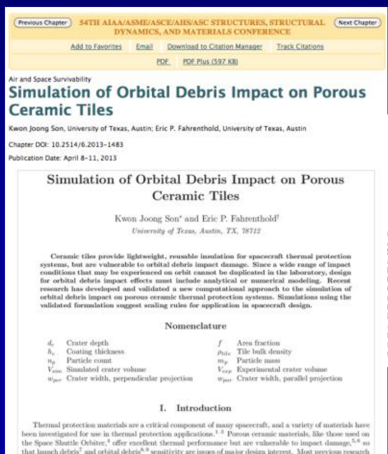


Tweet about them….

(Does this matter? We'll get to that later.)

Case study (from this morning)

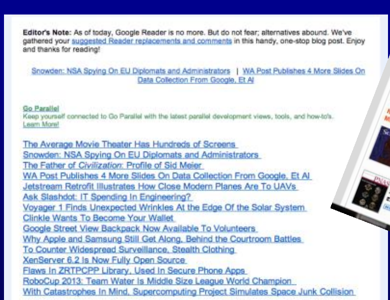
Original papers



TACC article

AAAS newsire

Wider distribution



(Does this matter? We'll get to that later.)

What's next?

In Fall 2013, I start at NSF in Washington, D.C.

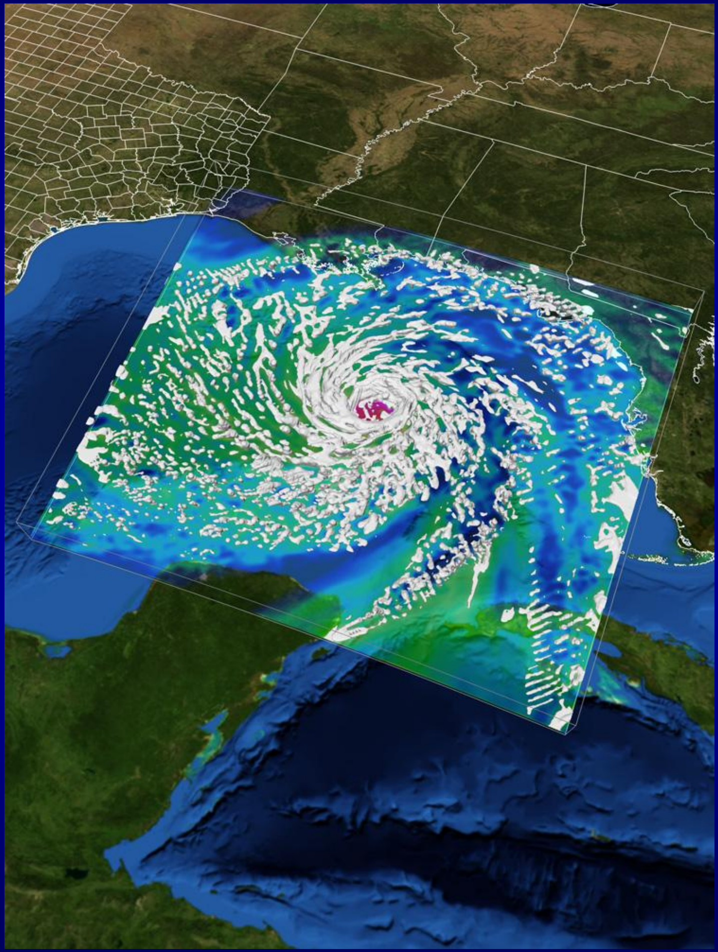
Will serve 1-3 years as a rotator in the Office of Legislative and Public Affairs

Working primarily with the Directorate for Computer and Information Science

Cyborg Journalism? Really?

No…. but there is something unique about my position:

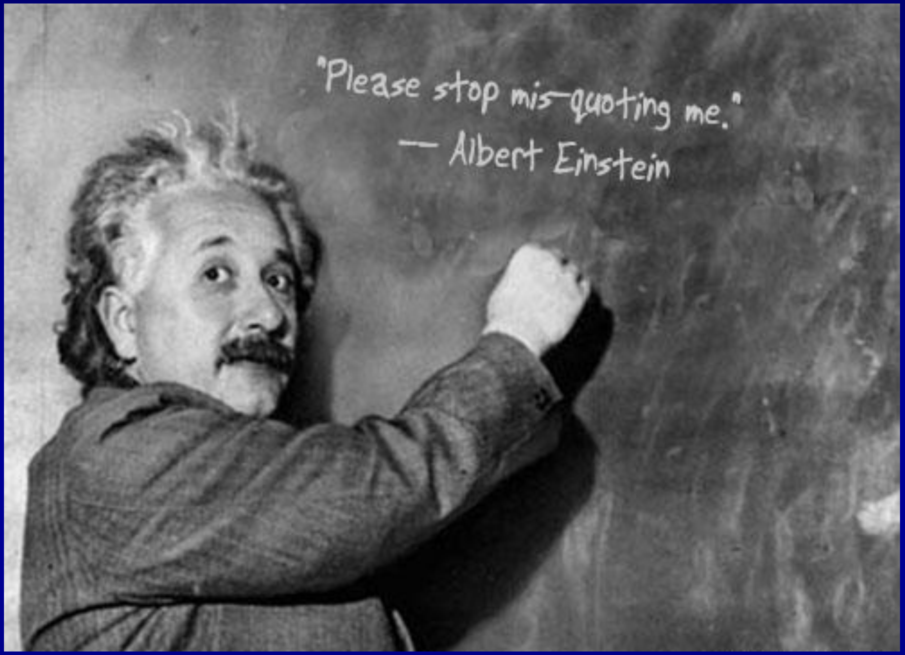
I, and the researchers that I report on, are partners, on a deep level, with machines – ones that can do unfathomable, superhuman actions.



Cyborg Journalism? Really?

"Computers are incredibly fast, accurate and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination."

- often attributed to Albert Einstein (probably said by Leo Cherne)



This gets at the hybrid nature of the endeavor.

Cyborg Journalism? Really?

Humans have always been tool-using creatures

Still, computational science seem profoundly weird and new

After centuries of doing science through theory and experimentation, a new avenue to insight has opened up

Involves offloading aspects of our "intelligence" to 1,000s of interconnected computer processors

Cyborg Journalism? Really?

All tools extend our abilities, but this tool creates imaginary worlds of our devising to replicate the dynamics of the universe.

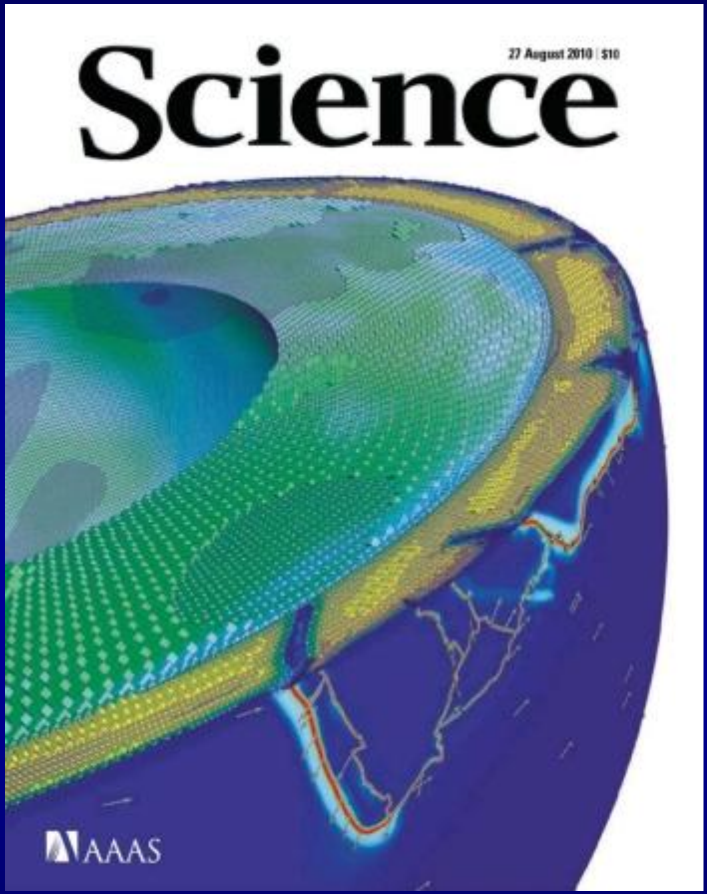
In the process, fundamental aspects of our world that could not be explored can now be probed through numerical simulation, modeling and analysis.

The work is almost entirely immaterial — virtual — yet, when compared to experimental results, it can be remarkably accurate….and useful.

Recent computational findings from Klaus Schulten on the cover of *Nature*.

Cyborg Journalism? Really?

Even after 6 years, I find this aspect of the research utterly fascinating.



Mantle convection simulations by from Omar Ghattas on the cover of *Science*.

(great examples)

Cyborg Journalism? Really?

The articles I write are not really about supercomputers.

About the creative design, deployment and ***use*** of these advanced technologies by scientists.

And yet… they aim to clarify the role that instruments (esp. HPC) play in discovery.

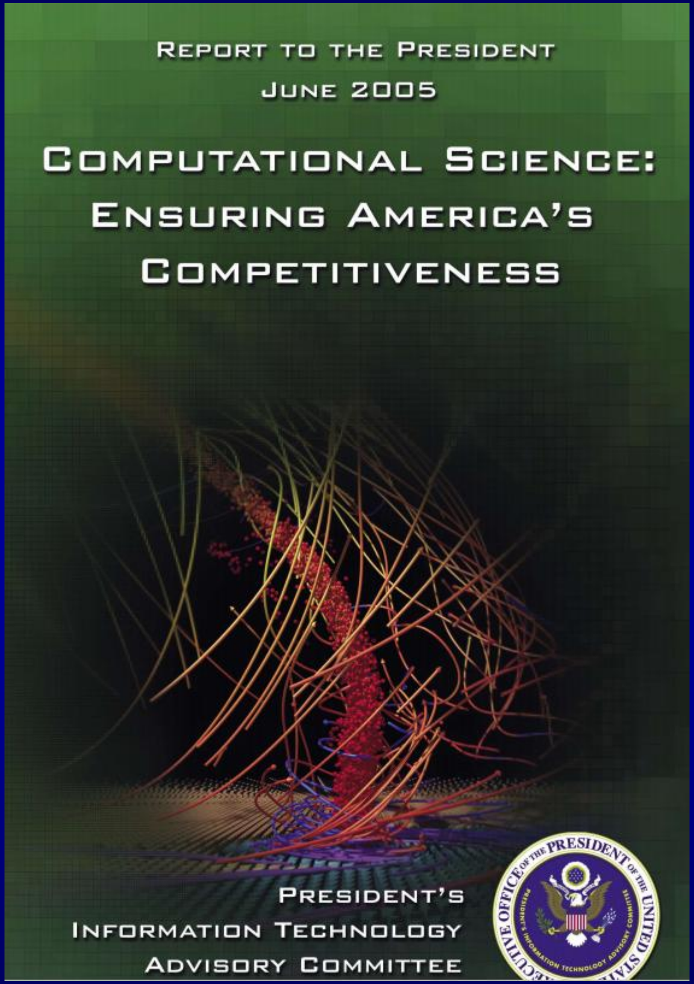
(Beyond a terse note in the acknowledgments section of the journal article.)

TACC

"Third Pillar of Science"

From the executive summary of the President's Information Technology Advisory Committee in 2005:

"This country has not yet awakened to the central role played by computational science and high-end computing in advanced scientific, social science, biomedical, and engineering research; defense and national security; and industrial innovation. *Together with theory and experimentation, computational science now constitutes the "third pillar" of scientific inquiry,* enabling researchers to build and test models of complex phenomena – such as multi-century climate shifts, multidimensional flight stresses on aircraft and stellar explosions – that cannot be replicated in the laboratory, and to manage huge volumes of data rapidly and economically."



TACC

Awareness Test - KITP

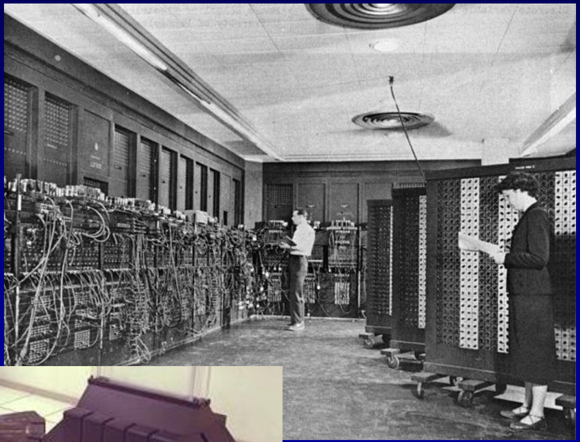
I think it's fair to say, we've awakened.

Or have we?

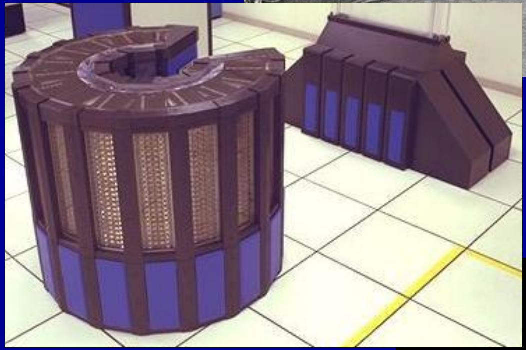
Quick Quiz:

Who has ever heard of a supercomputer?

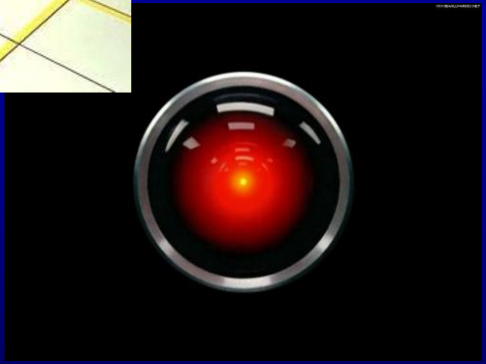
- *How many people have heard of TACC? TeraGrid/XSEDE?*
- *How many use supercomputers in your research?*
- *How about other researchers at your university or in your field?*



ENIAC



Cray-2



HAL 9000

Which is real? Which accurately represents current HPC technology?

Awareness Test – KITP II

Survey cont.

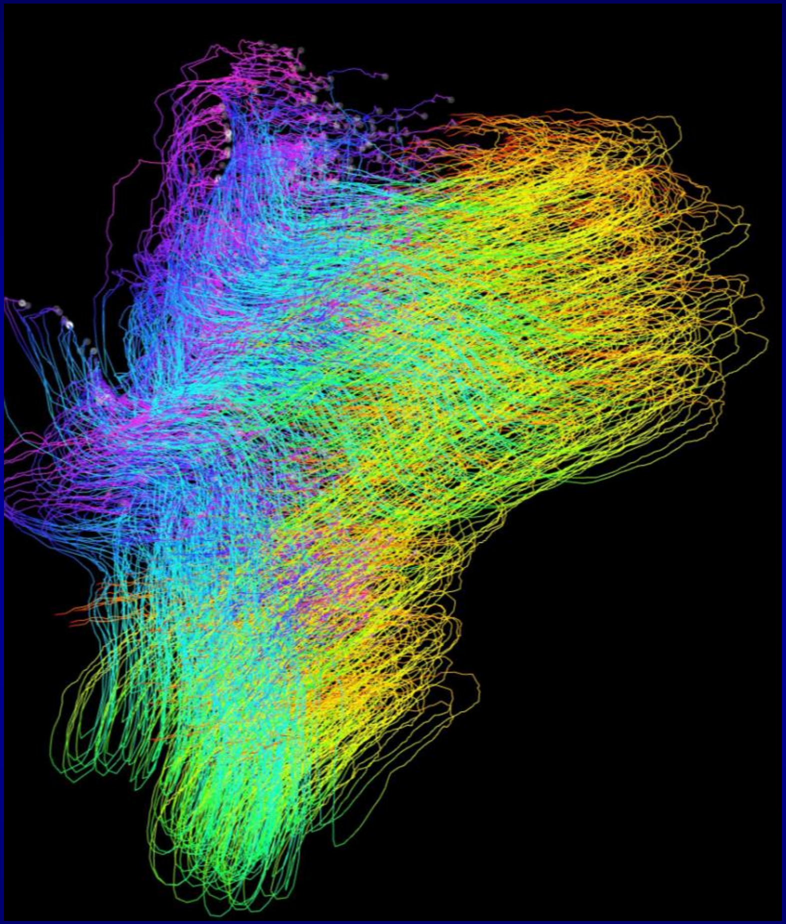
- Who considers themself primarily a computer scientist or programmer?
- (Who thinks in a generation there will be any distinction?)
- Who doesn’t use supercomputers because they had a bad experience?
- Who would use supercomputers more if they were easier to use?
- What if you could just do a simulation, trust the results, share them, and try something else, without huge amounts of effort?

I want to know what *you* think

For those of use who use computational methods in your research, I'm very interested in how you use and access these resources.

For those of use who *do not* use computational methods in your work, I'm interested to know why not and what you think about the role HPC plays in science generally?

Please pull me aside at any time and tell me. I have a long way to go in understanding my core audience - scientists.



A brief introduction to my world

XSEDE/TeraGrid

The Extreme Science and Engineering Discovery Environment (XSEDE) is the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. It is a single virtual system that scientists can use to interactively share computing resources, data, and expertise.

The five-year, $121-million project is supported by the National Science Foundation. It replaces and expands on the NSF TeraGrid project. More than 10,000 scientists used the TeraGrid to complete thousands of research projects, at no cost to the scientists.

A brief introduction to my world

XSEDE/TeraGrid

XSEDE is led by the University of Illinois' National Center for Supercomputing Applications. The partnership includes:

- Cornell University Center for Advanced Computing
- Indiana University
- Jülich Supercomputing Centre
- National Center for Atmospheric Research
- National Center for Supercomputing Applications - University of Illinois at Urbana-Champaign
- National Institute for Computational Sciences - University of Tennessee Knoxville/Oak Ridge National Laboratory
- Ohio Supercomputer Center - The Ohio State University
- Pittsburgh Supercomputing Center - Carnegie Mellon University/University of Pittsburgh
- Purdue University
- Rice University
- San Diego Supercomputer Center - University of California San Diego
- Shodor Education Foundation
- Southeastern Universities Research Association
- Texas Advanced Computing Center - The University of Texas at Austin
- University of California Berkeley
- University of Chicago
- University of Virginia

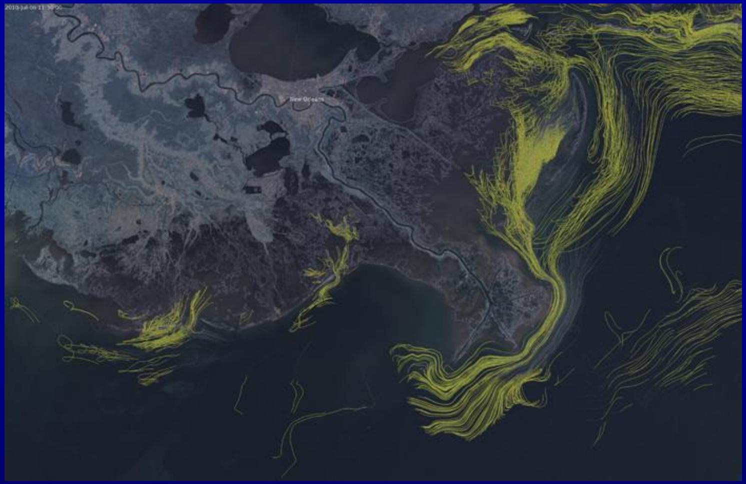
← Here we are.

(It's a large virtual organization.)

A brief introduction to my world

Texas Advanced Computing Center

Powering Discoveries that Change the World



Rapid simulations of the Gulf Oil Spill helped to direct emergency efforts.

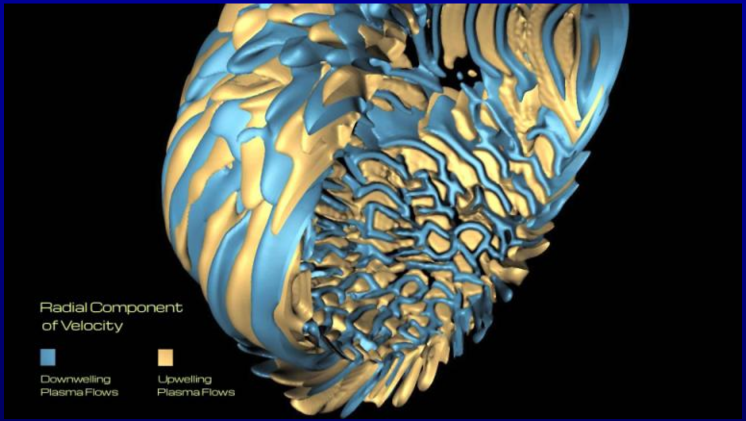
"The center's mission is to enable discoveries that advance science and society through the application of advanced computing technologies."

(It's nice to work for an organization with a mission.)

A brief introduction to my world

More about TACC

- TACC is a relatively new organization. (We just celebrated our 12^{th} birthday last month.)
- Claim to fame: Developed some of the first Linux clusters for the scientific community based on commodity components
- Upstart: won string of NSF solicitations for big systems against bigger, more established competition
- Start-up mentality: rapid growth, agile strategies
- Not just simulation and modeling; aspire to leadership in visualization, data analysis and storage technologies



TACC

A brief introduction to my world

TACC

An interesting organization to work for.

Somewhere between a brain trust (like KITP) and a public utility plant.

1/3 PhDs (mostly domain scientists)

1/3 hackers/IT whizzes

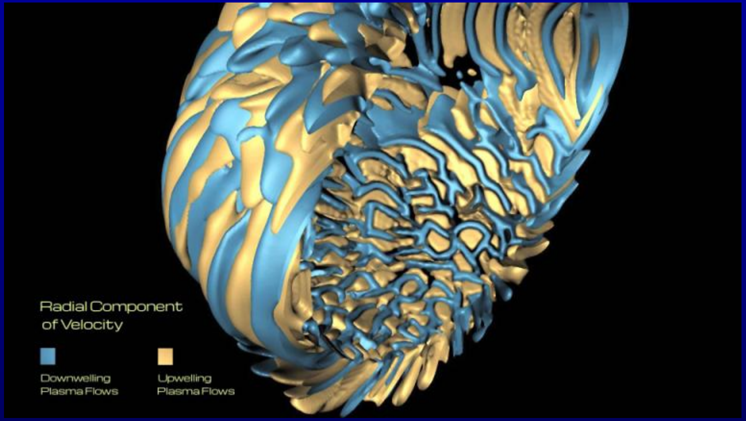
1/3 regular folks like me

Primarily “we” design, build, and keep advanced computing systems running so researchers ***<u>don’t have to</u>***

TACC

Ranger (2008-2013) R.I.P.

- Ranger came online as the 5th most powerful system in the world
- Served as the workhorse for the open scientific computing world for 4+ years
- Researchers from over 359 institutions across the country used Ranger in support of 2,244 research projects
- Over 4,000 scientists completed 2.69 million jobs, consuming 2.1 billion processor core hours, in support of these research projects
- (To accomplish the same amount of science on a PC would take more than 200,000 years)



TACC

Stampede (2013-?)

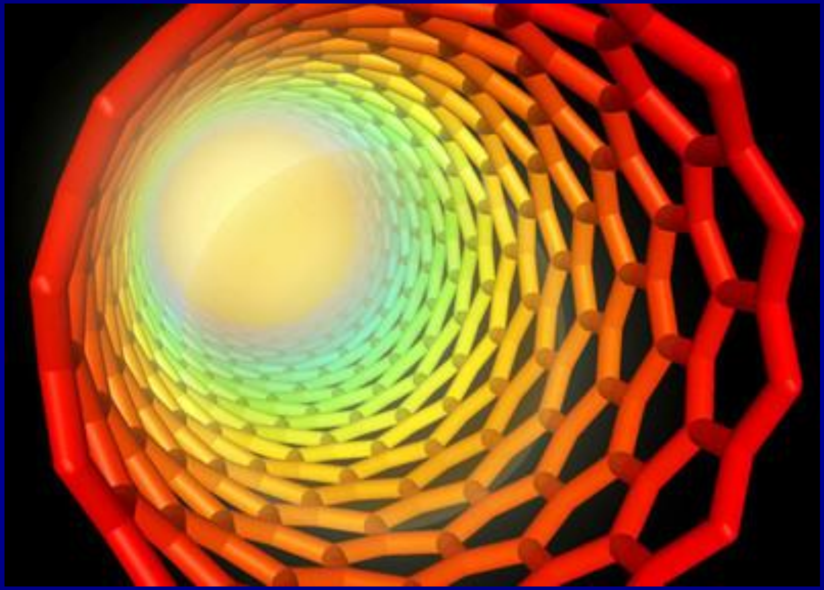
- Replaced Ranger
- Peak performance near 10 Petaflops
- Another $50+M investment from NSF
- Fills an 11,000 sq. ft. machine room
- Currently the 6th most powerful supercomputer in the world (according to the June 2013 Top500 list)
- Will serve as the workhorse for the open scientific computing world for the next 4+ years
- Free to use for any U.S. researcher at a university
- Came online in January; already serves more than 6,000 researchers

Stampede Early Science Highlights

CO_2 Capture and Conversion

- Alexie Kolpak of MIT is using *Stampede* to explore a new class of tunable nanocatalysts that can capture CO2 from an exhaust stream and convert it into *cyclic carbonate*, a valuable substance used in industrial applications.

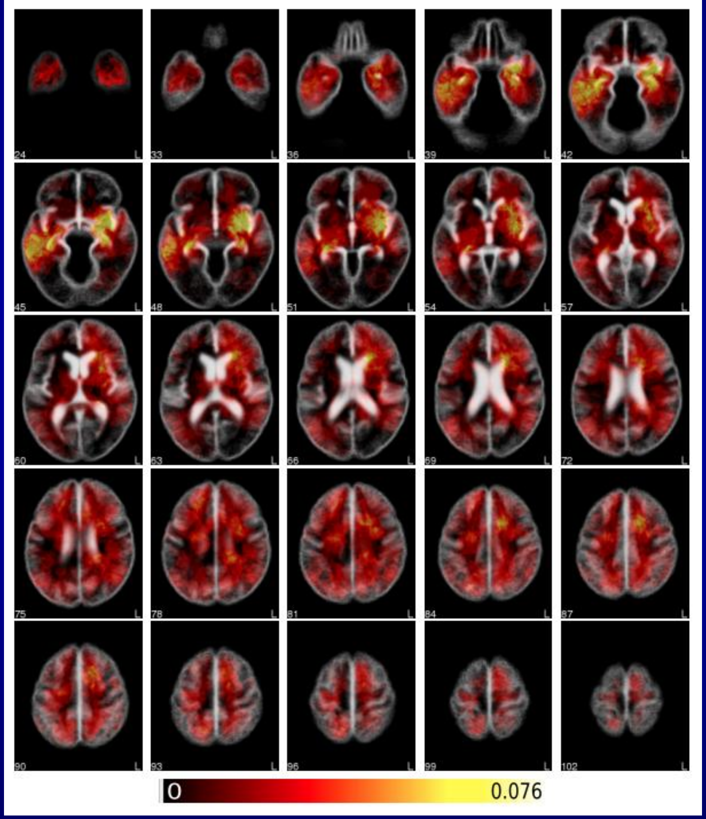
- The reactions involved occur at the atomic level, beyond the reach of any microscope, and can only be investigated dynamically via computer simulations.



Stampede Early Science Highlights

Improving Brain Tumor Imaging

- Surgeons want to know how aggressive a tumor is, and its infiltration into the surrounding tissue.
- Using Stampede, researchers can assimilate massive amounts of data from MRI scans and other imaging modalities and combine these with biophysical models that represent tumor growth.
- The addition of biophysical tumor models increases the accuracy and effectiveness of the interpretation of images
- Involves large amounts of complex computations that must be accomplished quickly.



Big Science (1940-?)

The Era of "Big Science"

- Large-scale scientific research funded by a national government or group of governments
- Complex funding infrastructure
- Large, interdisciplinary groups
- Massive instruments needed to push boundaries are big, powerful and expensive.

HPC is a part of this trend; also differs:

'A General-Purpose Instrument'

Try to play a democratizing role.

The huge scale of scientific research in the second half of the twentieth century is hard to ignore. Large-scale, coordinated science occupies whole regions of the United States—Silicon Valley, Route 128, the Research Triangle; it consumes a substantial fraction of all federal expenditures on research. Defense laboratories sprawl over thousands of acres and employ tens of thousands of workers. Some particle physics accelerators encircle whole towns. Seen from the outside, big science has, by its very scope, entered the realm of public debate. Questions are being asked about the implication of the huge scientific/technological projects—such as the Superconducting Supercollider (SSC), controlled fusion, and Star Wars—for the rest of science and for their effect on society.

Seen from the inside—from the scientists' perspective—big science entails a change in the very nature of a life in science. Teamwork and hierarchy increasingly characterize daily work at the big particle accelerators. Indeed, all but the very newest entrants to the field of high-energy physics, for example, have witnessed radical changes in the character of research even over the last few years. Teams of five or six researchers have been replaced by teams of tens; teams of tens in the more recent past now exceed one hundred. With the

Big Science (1940-?)

New XSEDE Goals:

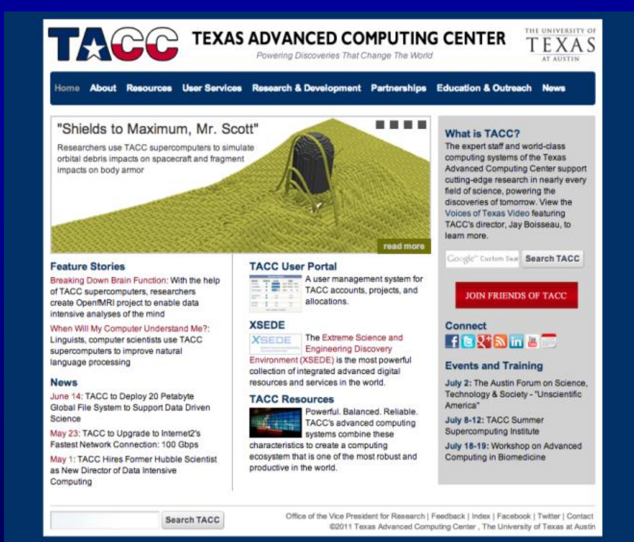
- **Deepen and extend the impact of eScience infrastructure on research and education**; in particular, to reach communities that have not previously made use of it;
- **Prepare the current and next generation of researchers, engineers, and scholars** in the effective use of advanced digital technologies;
- **Collaborate with institutions to ensure a more seamless use of the advanced technology capabilities** in the national eScience infrastructure to enhance the productivity of researchers, engineers, and scholars;
- **Create an open and evolving environment** that facilitates integration and sharing of heterogeneous digital services into a comprehensive national eScience infrastructure.
- **Expand the environment** through the integration of new capabilities and resources such as instruments and data repositories based on the identified needs of the community.
- **Deepen and expand the array of technical expertise and support services** provided to the community to maximize the effectiveness of their use of the eScience infrastructure;
- **Raise awareness** of the value of eScience infrastructure and, in particular, the critical technical expertise and support services.

↔

All aim to broaden use of HPC across science, engineering and the humanities

TACC

Beyond computing



In addition to building/managing systems, we collaborate with researchers to help them get best performance out of the system.

Lead training, outreach, education for all levels of users and learners.

Beyond computing

Communicate our activities to the general public.

This is where I and those like me come in

…and what the rest of my talk will be about.)

TACC

Media outreach

- Why do it?
- What to expect
- Where to get help

Media outreach

Why do it?
1) Because we owe it to taxpayers

Media outreach

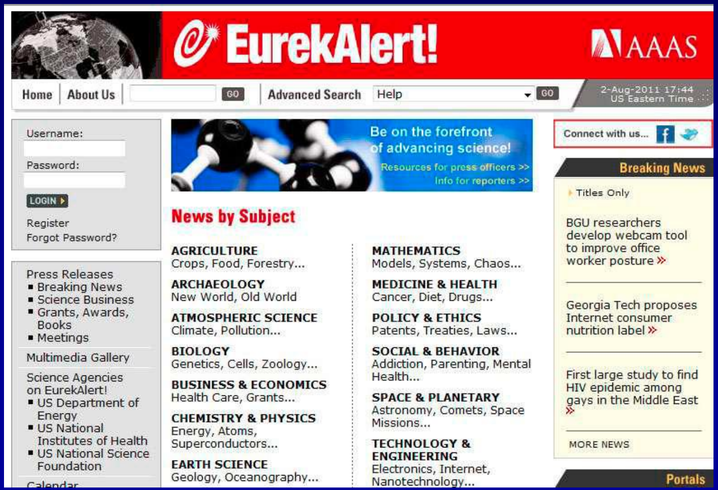
NSF Broader Impacts Criteria

1. Advance discovery and learning
2. Broaden the participation of underrepresented groups
3. Enhance infrastructure for research and education
4. **Disseminate results broadly**
5. Benefit society

The media can help.

Media outreach

Why do it?
2) Because journals ask you to



Media outreach

Why do it?
3) Because the media needs you to

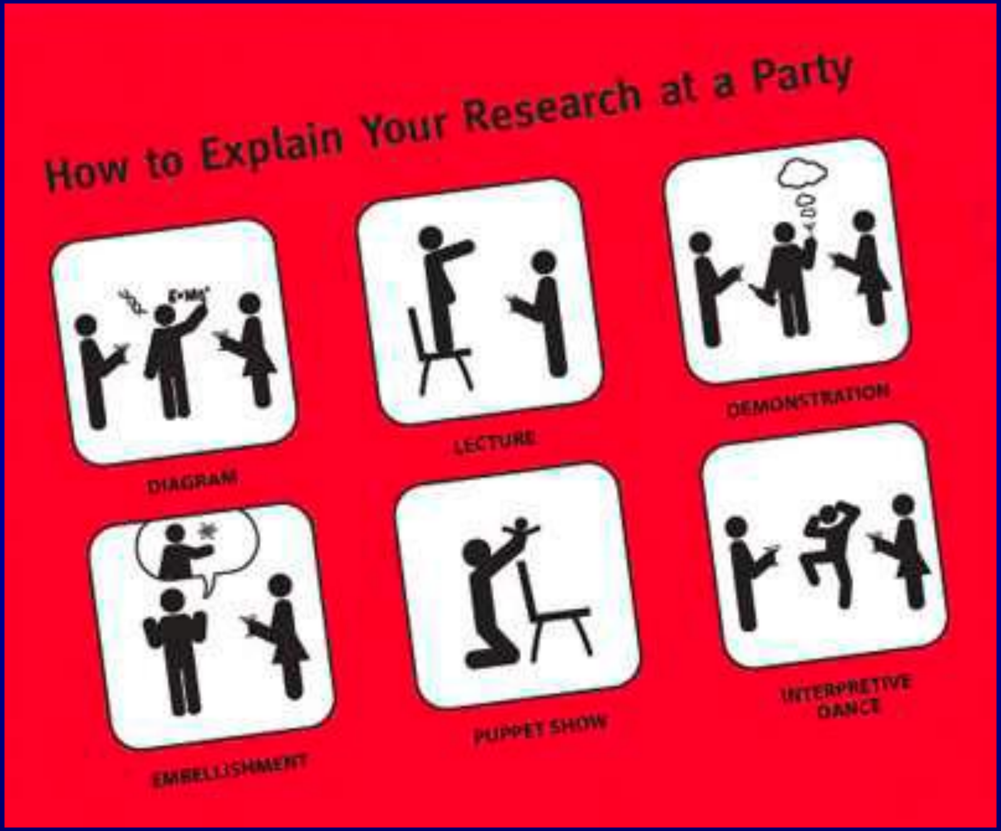
Once, 70 major American newspapers had science sections. Now, few do.

"Science is always the first thing that gets cut when times get tough."
-Ira Flatow, NPR

Media outreach

We've talked about why.

But how?



What's news? What's not?

What the media are looking for in a story… and why your work may be more newsworthy than you think.

What's news? What's not?

Publication of a scientific paper

What’s news? What’s not?

Sometimes the secret is timing

“Pine pollen still potent miles from the tree”

AMERICAN JOURNAL OF BOTANY

CORNELL UNIVERSITY PRESS

Institution: DUKE UNIVERSITY | Sign In via User Name/Password

AJB Advance Access
Published online ahead of print March 26, 2010;
doi:10.3732/ajb.0900255
American Journal of Botany
© 2010 Botanical Society of America, Inc.

Research Article

Long-distance pine pollen still germinates after meso-scale dispersal[1]

Claire G. Williams[2]

Forest History Society and NESCent, William Vickers Avenue, Durham, North Carolina 27701 USA

ABSTRACT

Viability of long-distance pollen links ecological models to the genetic structure of forest tree populations, determining how genetically modified (GM) pine plantations. Addressing this landscape-scale inquiry is feasible when the pollen source, the this end, I measured long-distance pollen germination along a 160-km transect along the North Carolina coastline, include islands. Using this system, I tested three hypotheses: (1) pine pollen germinates after dispersal on meso-scale distances, captured over open saltwater, and (3) viable pine pollen is present at high altitudes before local peak pollen shed. The ex rates of 2 to 57% after dispersal at distances from 3 to 41 km, sodium chloride solutions mildly reduced *P. taeda* poller altitude of 610 m. GM pine plantings thus have a potential to disperse viable pollen at least 41 km from the source. We together exert a powerful influence on the genetic structure of forest tree populations.

Key Words: aerosols • climate change • conifer reproductive biology • genetically modified (GM) forest trees • hete Carolina • Pinaceae • *Pinus taeda* • U.S. Forest Health Initiative

Received for publication 24 August 2009. Accepted for publication 16 February 2010.

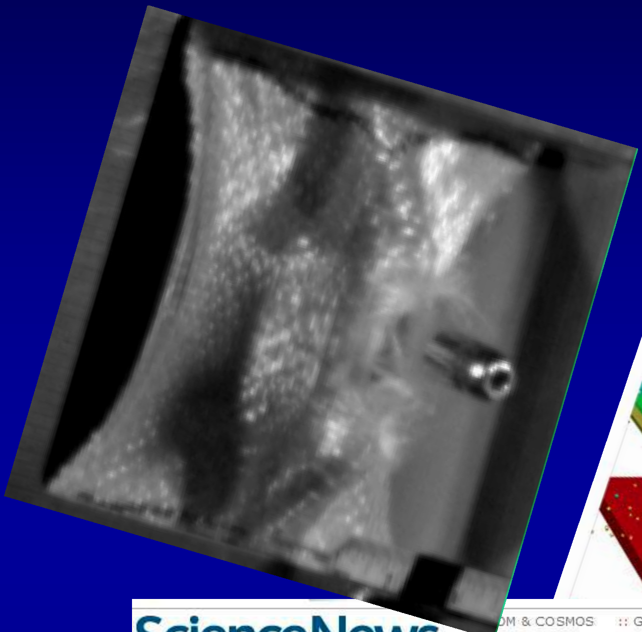
Guess when this came out?

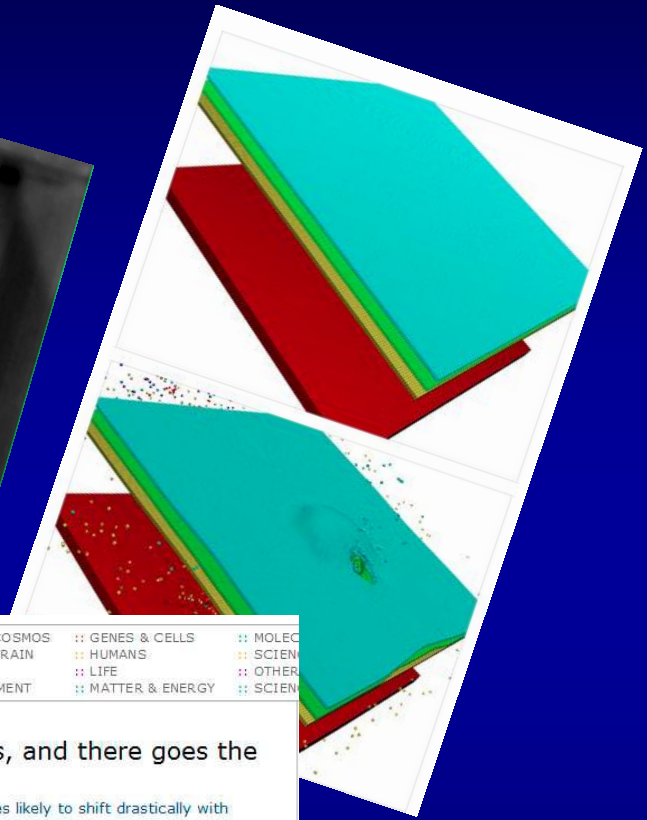
TACC

What's news? What's not?

What is also (sometimes) news?

A talk at a conference or meeting





ScienceNews
MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC

:: ATOM & COSMOS :: BODY & BRAIN :: EARTH :: ENVIRONMENT :: GENES & CELLS :: HUMANS :: LIFE :: MATTER & ENERGY :: MOLEC :: SCIEN :: OTHER :: SCIEN

HOME
NEWS
FEATURES
BLOGS
COLUMNS
DEPARTMENTS
RSS FEEDS
E-MAIL ALERTS
SUBSCRIBE

In the November 6 Issue:

Home / News / Article

Climate changes, and there goes the neighborhood

Ranges of rattlers and voles likely to shift drastically with warming

By Susan Milius
Web edition : Friday, October 15th, 2010

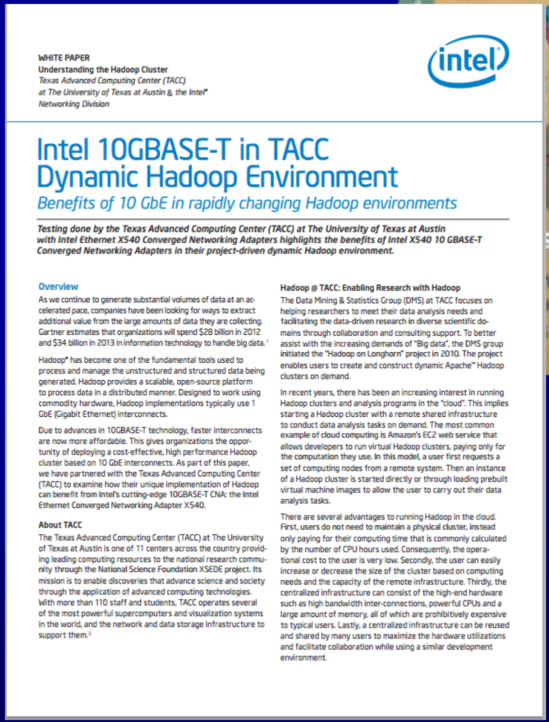
Text Size

PITTSBURGH — Rattlesnakes and voles could be facing real estate meltdowns of their own, as climate change forecloses habitats or shifts livable conditions into new regions at speeds as much as a thousand times faster than prehistoric averages.

What (usually) ISN'T news?

Exhibit A:
review papers

Exhibit B:
white papers /
policy papers



WHITE PAPER
Understanding the Hadoop Cluster
Texas Advanced Computing Center (TACC) at The University of Texas at Austin & the Intel® Networking Division

intel®

Intel 10GBASE-T in TACC Dynamic Hadoop Environment

Benefits of 10 GbE in rapidly changing Hadoop environments

Testing done by the Texas Advanced Computing Center (TACC) at The University of Texas at Austin with Intel Ethernet X540 Converged Networking Adapters highlights the benefits of Intel X540 10 GBASE-T Converged Networking Adapters in their project-driven dynamic Hadoop environment.

Overview

As we continue to generate substantial volumes of data at an accelerated pace, companies have been looking for ways to extract additional value from the large amounts of data they are collecting. Gartner estimates that organizations will spend $28 billion in 2012 and $34 billion in 2013 in information technology to handle big data.[i]

Hadoop* has become one of the fundamental tools used to process and manage the unstructured and structured data being generated. Hadoop provides a scalable, open-source platform to process data in a distributed manner. Designed to work using commodity hardware, Hadoop implementations typically use 1 GbE (Gigabit Ethernet) interconnects.

Due to advances in 10GBASE-T technology, faster interconnects are now more affordable. This gives organizations the opportunity of deploying a cost-effective, high performance Hadoop cluster based on 10 GbE interconnects. As part of this paper, we have partnered with the Texas Advanced Computing Center (TACC) to examine how their unique implementation of Hadoop can benefit from Intel's cutting-edge 10GBASE-T CNA: the Intel Ethernet Converged Networking Adapter X540.

About TACC

The Texas Advanced Computing Center (TACC) at The University of Texas at Austin is one of 11 centers across the country providing leading computing resources to the national research community through the National Science Foundation XSEDE project. Its mission is to enable discoveries that advance science and society through the application of advanced computing technologies. With more than 110 staff and students, TACC operates several of the most powerful supercomputers and visualization systems in the world, and the network and data storage infrastructure to support them.[ii]

Hadoop @ TACC: Enabling Research with Hadoop

The Data Mining & Statistics Group (DMS) at TACC focuses on helping researchers to meet their data analysis needs and facilitating the data-driven research in diverse scientific domains through collaboration and consulting support. To better assist with the increasing demands of "Big data", the DMS group initiated the "Hadoop on Longhorn" project in 2010. The project enables users to create and construct dynamic Apache™ Hadoop clusters on demand.

In recent years, there has been an increasing interest in running Hadoop clusters and analysis programs in the "cloud". This implies starting a Hadoop cluster with a remote shared infrastructure to conduct data analysis tasks on demand. The most common example of cloud computing is Amazon's EC2 web service that allows developers to run virtual Hadoop clusters, paying only for the computation they use. In this model, a user first requests a set of computing nodes from a remote system. Then an instance of a Hadoop cluster is started directly or through loading prebuilt virtual machine images to allow the user to carry out their data analysis tasks.

There are several advantages to running Hadoop in the cloud. First, users do not need to maintain a physical cluster, instead only paying for their computing time that is commonly calculated by the number of CPU hours used. Consequently, the operational cost to the user is very low. Secondly, the user can easily increase or decrease the size of the cluster based on computing needs and the capacity of the remote infrastructure. Thirdly, the centralized infrastructure can consist of the high-end hardware such as high bandwidth inter-connections, powerful CPUs and a large amount of memory, all of which are prohibitively expensive to typical users. Lastly, a centralized infrastructure can be reused and shared by many users to maximize the hardware utilizations and facilitate collaboration while using a similar development environment.

TACC

What (usually) ISN'T news?

Also (usually) not news:

- Grants, awards and fellowships
- New programs, centers, institutes, etc.
- Building dedications

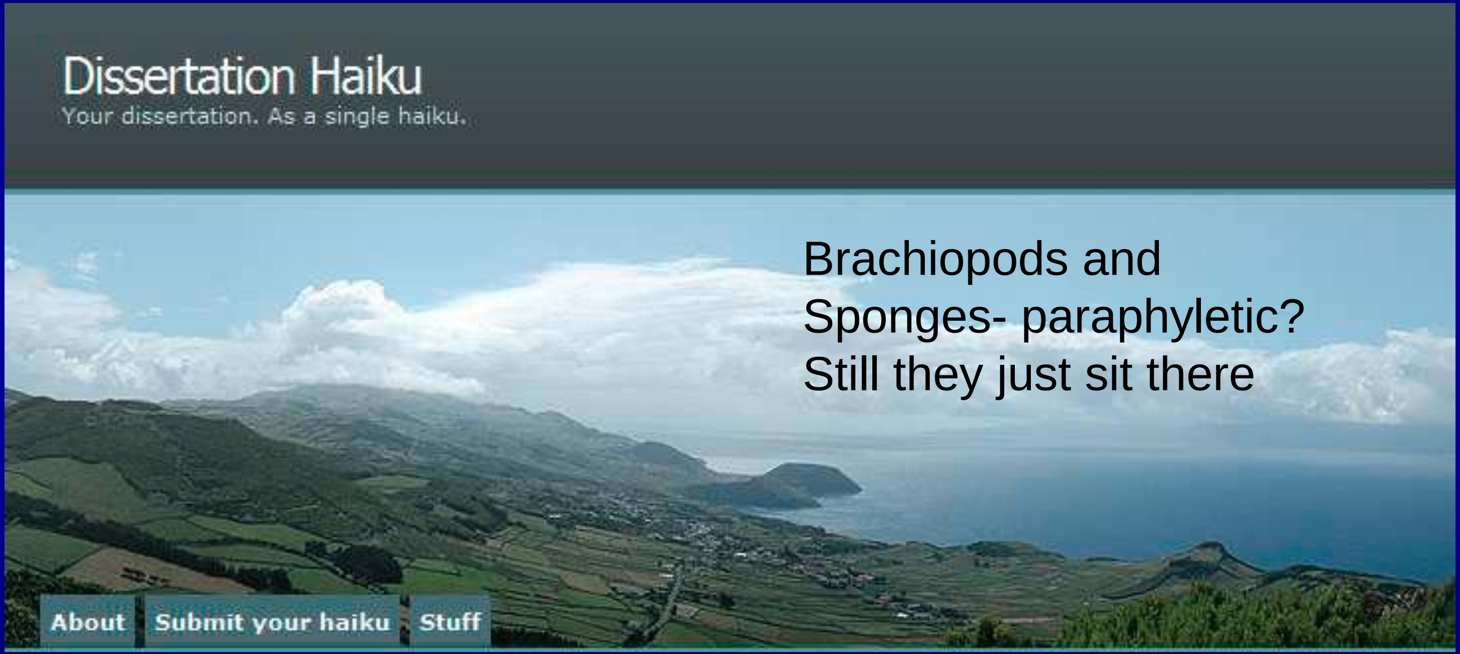
What (usually) ISN'T news?

Also (usually) not news:

- Grants, awards and fellowships
- New programs, centers, institutes, etc.
- Building dedications

Explaining your work in a quick & compelling way: a few tips

Dissertations are long and boring.



By contrast, everybody likes haiku.

http://dissertationhaiku.wordpress.com

When engaging larger audiences brevity is key

Here's why:

TV story = 80 seconds

Radio story = 45 seconds

Newspaper article = 400-600 words

Undergraduate attention span = 10 min

When engaging larger audiences brevity is key

Say it simply

"But this is the simplified version for **the general public."**

More tips

Try not to slide into incomprehensible jargon

Analogies are your friend

It’s ok to repeat yourself

The Index of Banned Words (The Continually Updated Edition)

By Carl Zimmer | November 30, 2009 3:35 pm

Over the summer, I posted a list of words I banned from my science writing class at Shoals Marine Lab. Readers offered some equally abysmal suggestions. And this fall, teaching a seminar at Yale, I came across some others. I suspect that this list is just going to keep growing. So I’m giving it a home here, where I can add in new entries as they arise in assignments in my classes. You can easily direct people to it through this url: http://bit.ly/IndexBanned (caps required).

By assembling this list, I don’t mean to say that no one should *ever* use these words. I am not teaching people how to write scientific papers. What I mean is that anyone who wants to learn how to write about science–***and*** to be read by people who aren’t being paid to read–should work hard to learn how to explain science in plain yet elegant English–not by relying on scientific jargon, code-words, deadening euphemisms, or meaningless cliches.

Here’s a post where I go into more depth about why words matter- …nces, paragraphs, etc.]

Questions you might be asked

- Big picture: So what? Why should we care?
- Can you think of any good analogies?
- Do you have any images we could use?
- What’s next?

More tips

Got a paper in review? Think it's likely to be accepted?

Contact your press officer/public information officer (yes, you have one)

4 things you didn't know about your press officer (and why you should care)

1. Your PO writes for a living

2. His/her job is to tell the media about your work

3. Your PO has media contacts and dissemination tools you don't have

4. This is the stage where you have the most control

More tips

Timing is everything

Alert your press officer to papers in the pipeline. At acceptance works best.

Recap

Why?

- Because NSF/taxpayers want you to
- Because journals ask you to
- Because the media needs you to

How?

- Say it simply
- Don't use jargon
- Do use analogies, similes, metaphors
- Contact your press officer

When?

- Before publication is best

Q&A

Thanks for listening!

If there's time, let's have a discussion about supercomputing, the media and how your research can reach a broader audience (or whether you want it to).