

# Big Data in the Big City

Paul Muzio, Director  
[paul.muzio@csi.cuny.edu](mailto:paul.muzio@csi.cuny.edu)

- City University of New York
- Research examples
- Challenges
- Future plans

The CUNY HPC Center acknowledges the following support

- NSF Grant 0855217
- NSF Grant 0958379
- NSF Grant 1126113
- NYC Council through the efforts of Councilman James Oddo
- Borough President of Staten Island – James P. Molinaro
- New York State Regional Economic Development Grant

- Support CUNY's "Decade of Science" Initiative and the Integrated University Concept of Operation.
- Support the University's research and educational activities by making state-of-the-art HPC resources and expert technical assistance available to faculty and students.
- With CUNY faculty and researchers, develop proposals for external funding.
- Support National, New York State, and New York City initiatives in economic development.
- Support National and New York State initiatives to promote the sharing of HPC resources and technical knowledge.
- Support educational outreach programs designed to encourage intermediary and high school students to pursue higher education and careers in science and technology.



## Senior Colleges

- City College
- Hunter College
- Baruch College
- Brooklyn College
- Queens College
- New York City College of Technology
- College of Staten Island
- John Jay College of Criminal Justice
- York College
- Lehman College
- Medgar Evers College

- **Community Colleges**

- Bronx Community College
- Queensborough Community College
- Borough of Manhattan Community College
- Kingsborough Community College
- LaGuardia Community College
- Hostos Community College
- New Community College

- **Graduate and Professional Schools**

- CUNY Graduate Center
- Sophie Davis School of Biomedical Education
- School of Law
- William E. Macaulay Honors College
- Graduate School of Journalism
- School of Professional Studies
- School of Public Health

- Largest urban university in the United States
  - 271,000 degree program students.
  - 270,000 adult, continuing, and professional education students.
- Gender
  - 61% female
  - 39% male
- Ethnicity
  - 0.2% American Indian/Native Alaskan
  - 15.8% Asian/Pacific Islander
  - 27.1% Afro-American
  - 25.7% Hispanic
  - 31.2% White
- 72% attended New York City public high schools
- 42% first time freshman
- 32% work for pay more than 20 hours per week
- 189 different languages spoken as first language

- Nobel prize winners

- Julius Axelrod
- Kenneth Arrow
- Robert Aumman
- Stanley Cohen
- Gertrude Elion
- Herbert Hauptman
- Robert Hofstadter
- Jerome Karle
- *Henry Kissinger*
- Arthur Kornberg
- Leon Lederman
- Arno Penzias
- *Julian Schwinger*
- Rosalyn Sussman Yalow

- Fields Medal winners

- Jesse Douglas
- Paul Cohen

- Felix Frankfurter, Supreme Court Justice
- Celina Sotomayer, mother of Supreme Court Justice Sonia Sotomayer
- The parents of Michael Dell
- The parents of David Bader
- Dr. Jonas Salk, Polio Vaccine
- Robert Kahn, Co-developer of TCP/IP
- Andy Grove, co-founder and former CEO, Intel Corp.

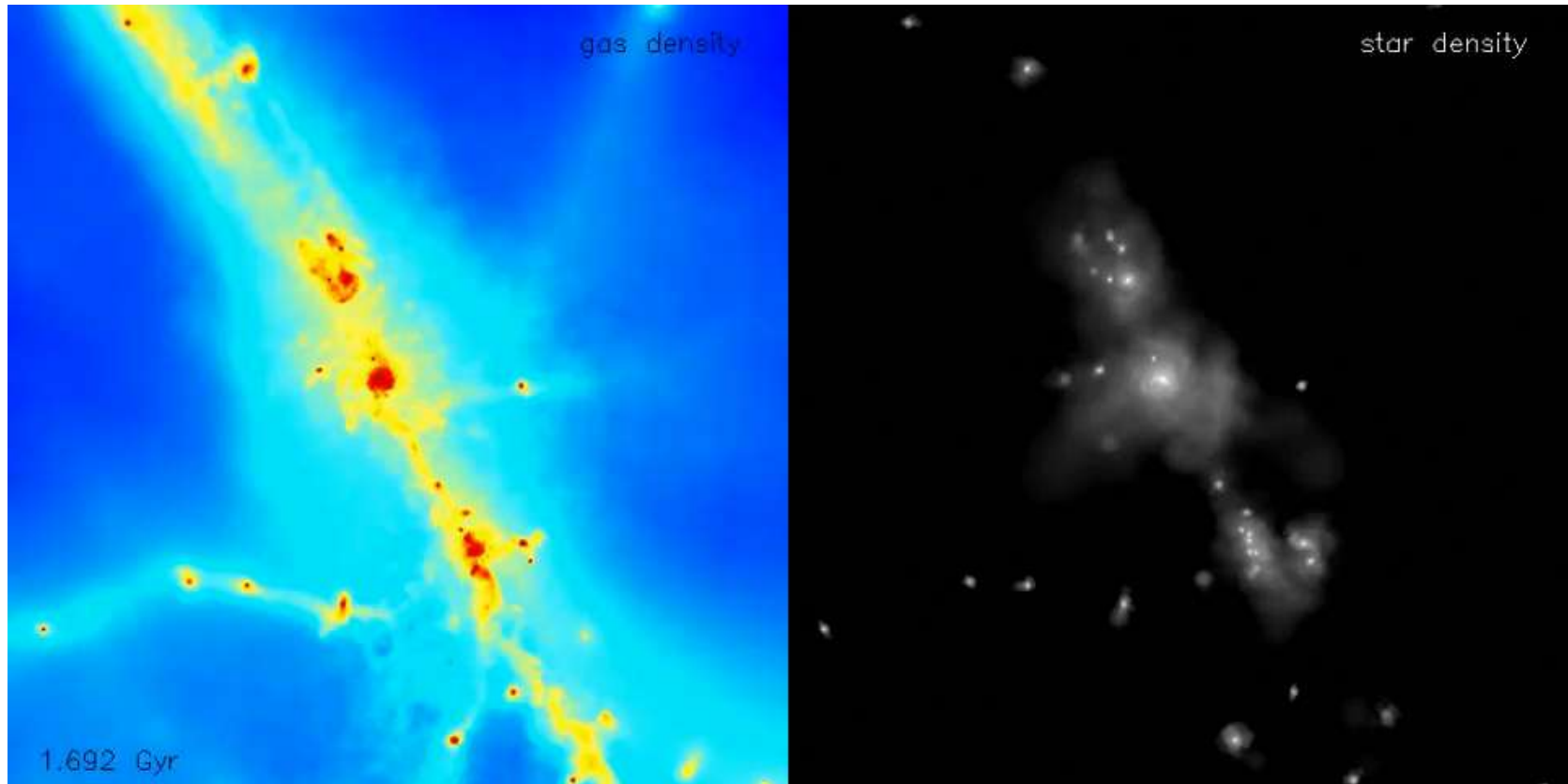
# Located at the College of Staten Island

Welcome to the College of Staten Island,  
City University of New York.





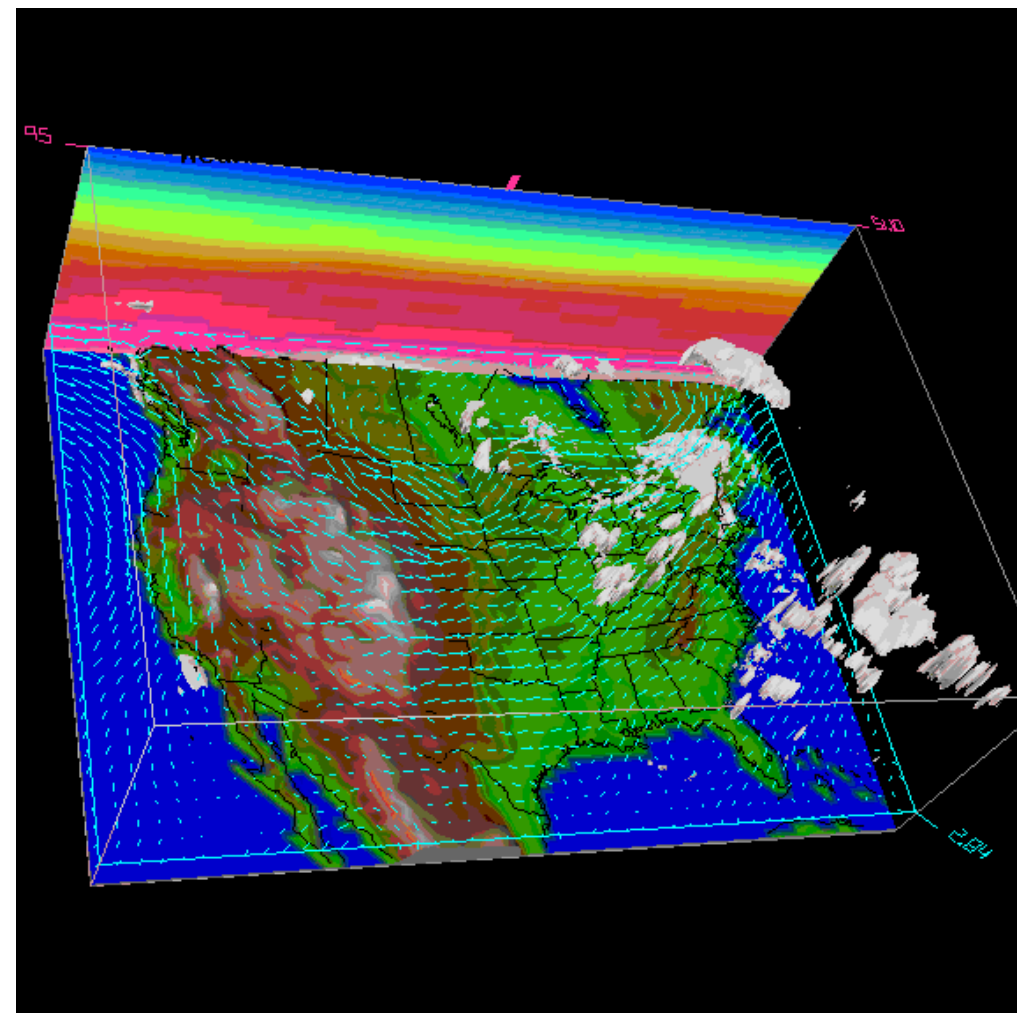
$10^{25}$  ????



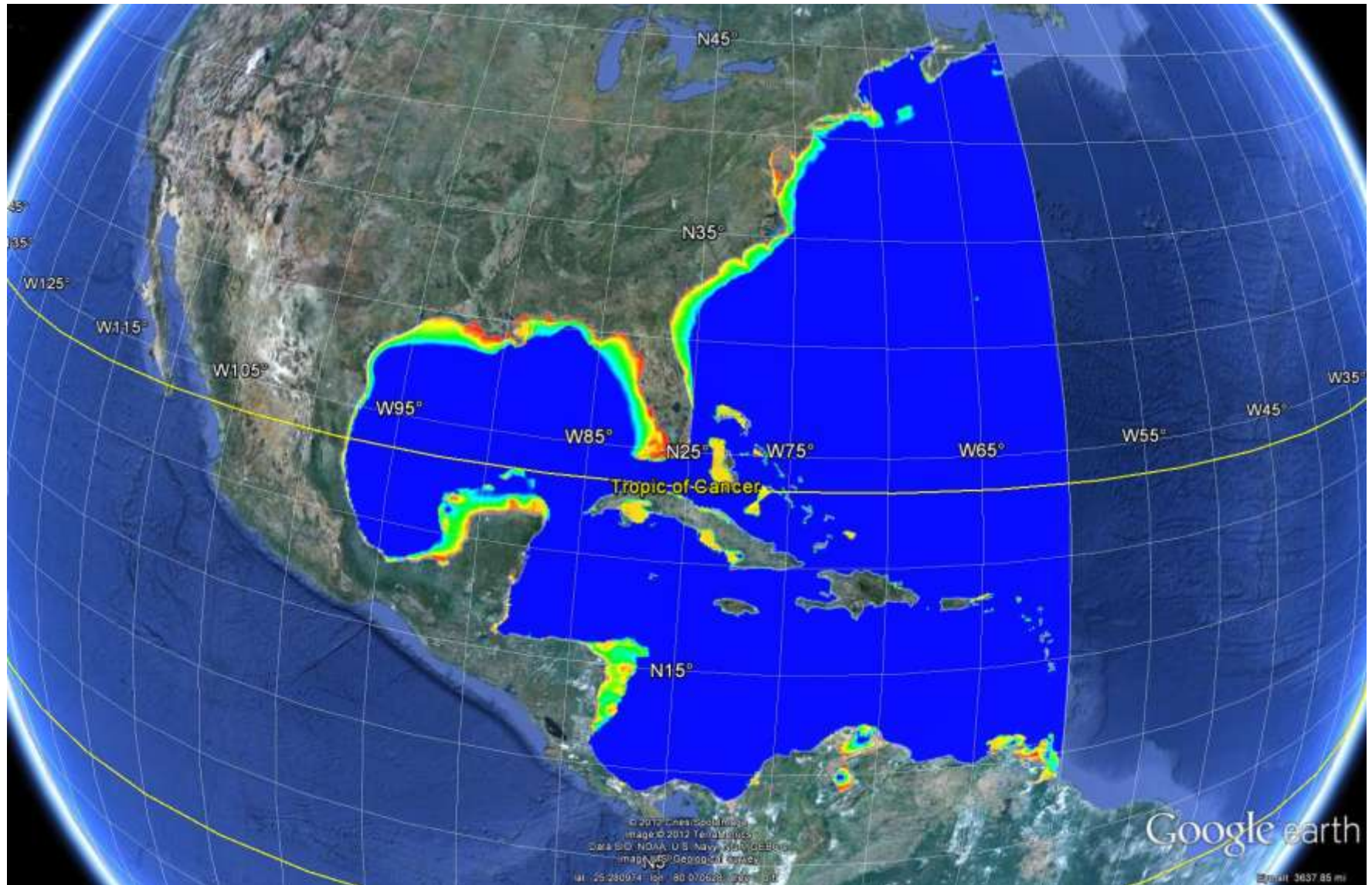
\* Computational cosmology: The evolution of the galaxy  
Courtesy: Dr. Ari Maller, New York College of Technology/CUNY

- Numerical Weather Prediction/Climate Modeling

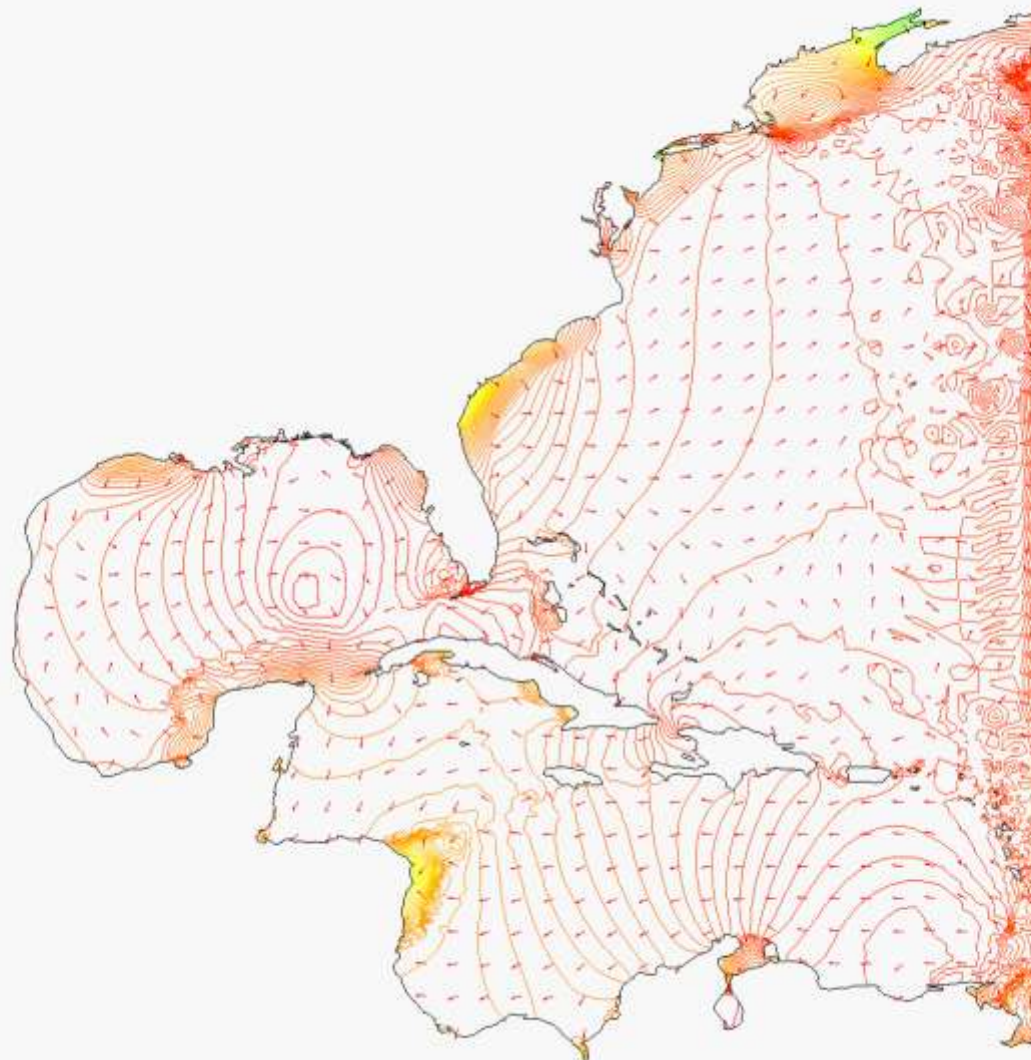
- Ocean models
- Terrain
- Vegetation
- Chemicals
- Particulates
- Hydrology
- Human impact



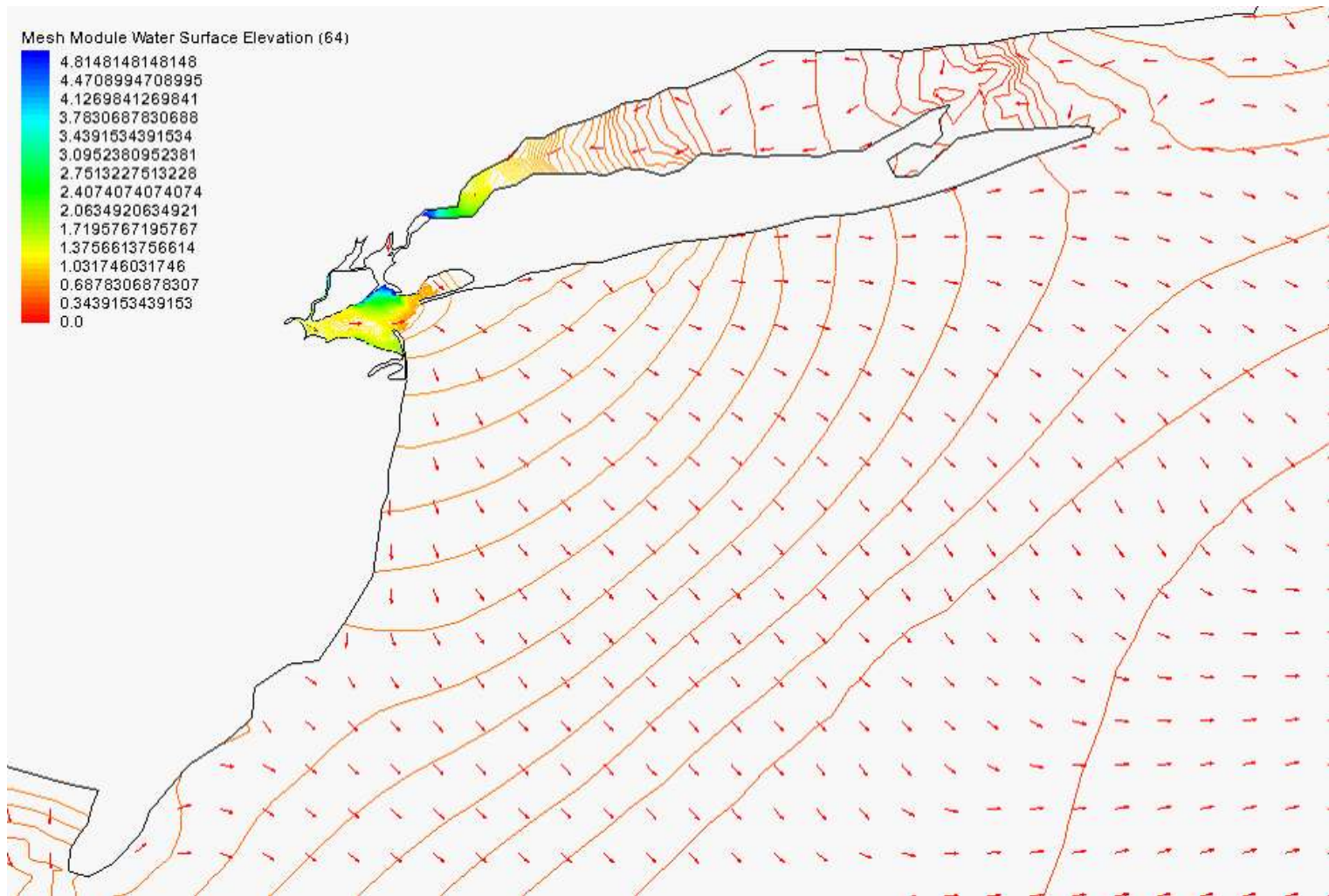
CUNY Collaborative Remote Sensing Systems Technology Institute  
& the CUNY Environmental CrossRoads Initiative



Mesh Module Water Surface Elevation (64)



\* Courtesy: Dr. M. Kress, et al, College of Staten Island

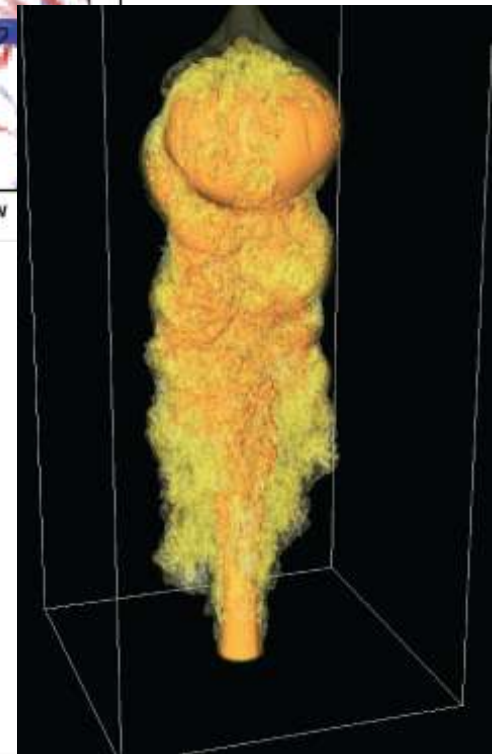
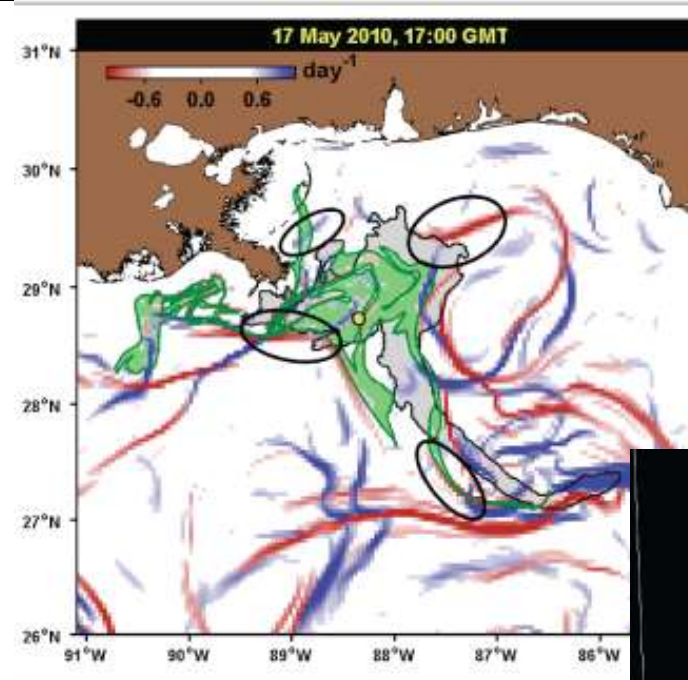




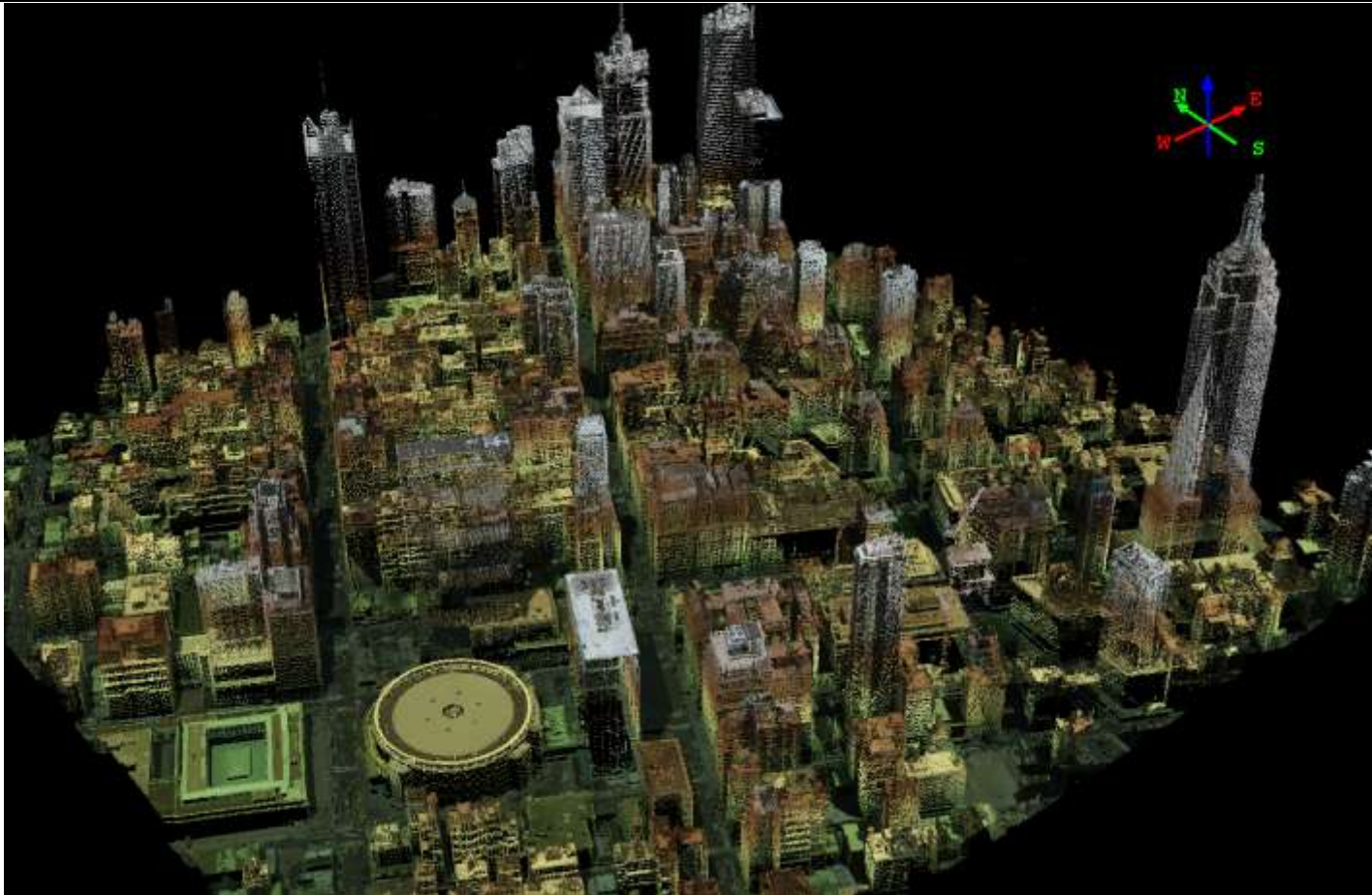
## Hydrocarbon Transport

- Huge range of scales
  - 1 U surface layer
  - O(10 m) plume
  - O(100 km) eddies
  - Tidal & Surf zones
  - Multi-scale turbulence
  
- Evolving Physio-Chemistry
  - Complex chemistry
  - Biology
  - Photo-chemistry
  - Air-sea interactions
  
- Deep Ocean Processes

\* Courtesy: D. Poje, CSI/CUNY







\* Courtesy: Center for Advanced Research in Spatial Information, Sean Ahearn, et al, Hunter/CUNY



**Technique 2:** Adaptive triangulation with smoothing filter by 1 bin size radius and maximum  $g$  tolerance 3 meter, spike and well removal

# Solar Insolation Calculation



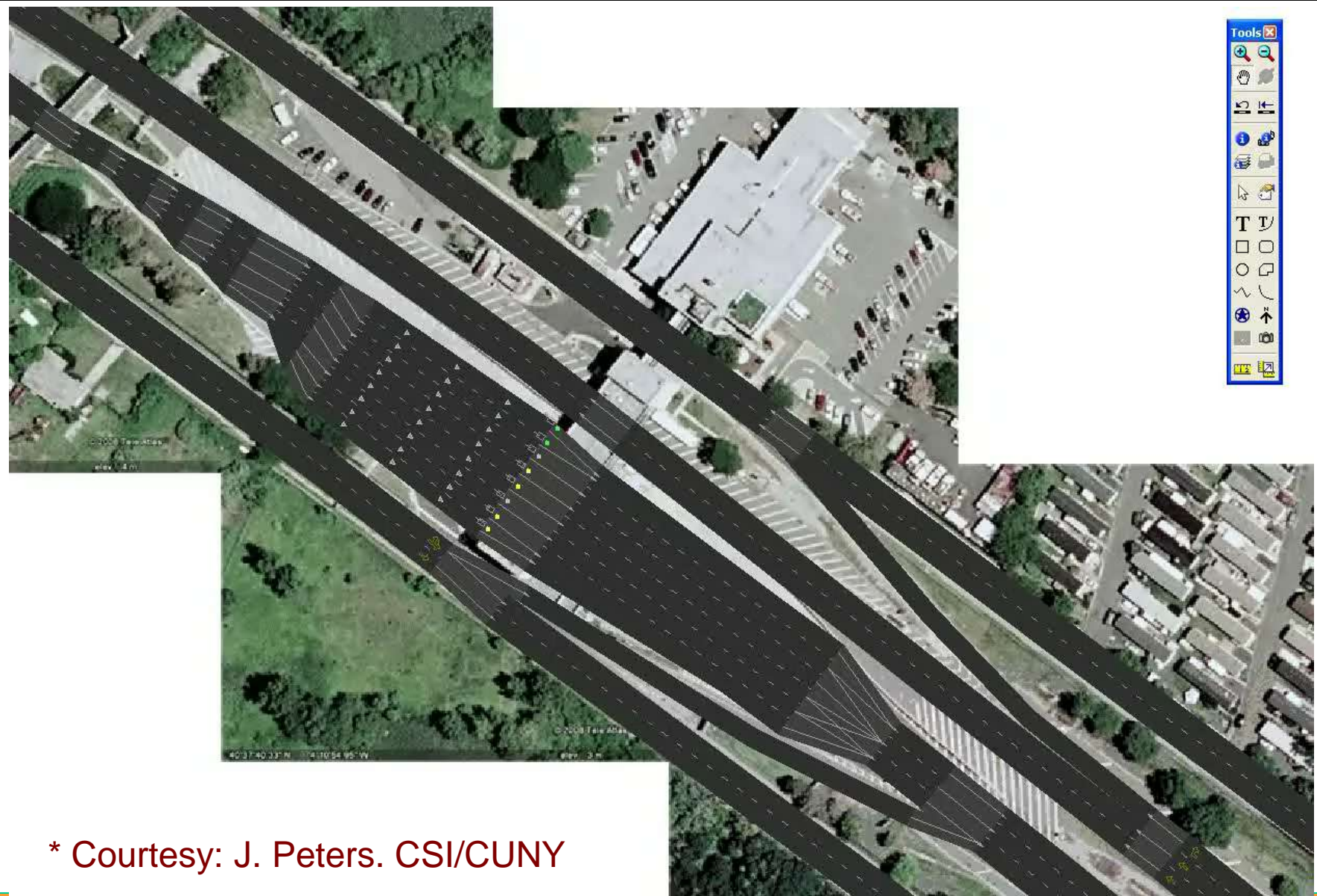
The screenshot shows the NYC Solar Map interface. At the top, there is a navigation bar with the title "NYC SOLAR MAP" and a search bar containing "365 5th Ave, Manhattan, NY 10". The interface includes several interactive buttons: "Zoom to Borough", "Find Address", "Select Building", "Draw Solar", "Clear Selection", "Legend", and "About". On the right side, there are "Street Map" and "Satellite" view toggles. The main area displays a satellite view of a city block in Manhattan, with a blue arrow pointing to a specific building. An information popup for "355 5 AVENUE" is overlaid on the map, providing the following summary:

**Summary**

- Annual electricity bill savings up to: **\$25,174.**
- You can install up to **111.34** kilowatts of solar here.
- Reduce your annual carbon emissions by up to **92,872 lbs/yr.**
- That's the same as planting **248 trees!**

Note: estimates only, actual values may vary. [Click here](#) to learn how this building's solar potential was estimated.

[Get Site Details >](#)

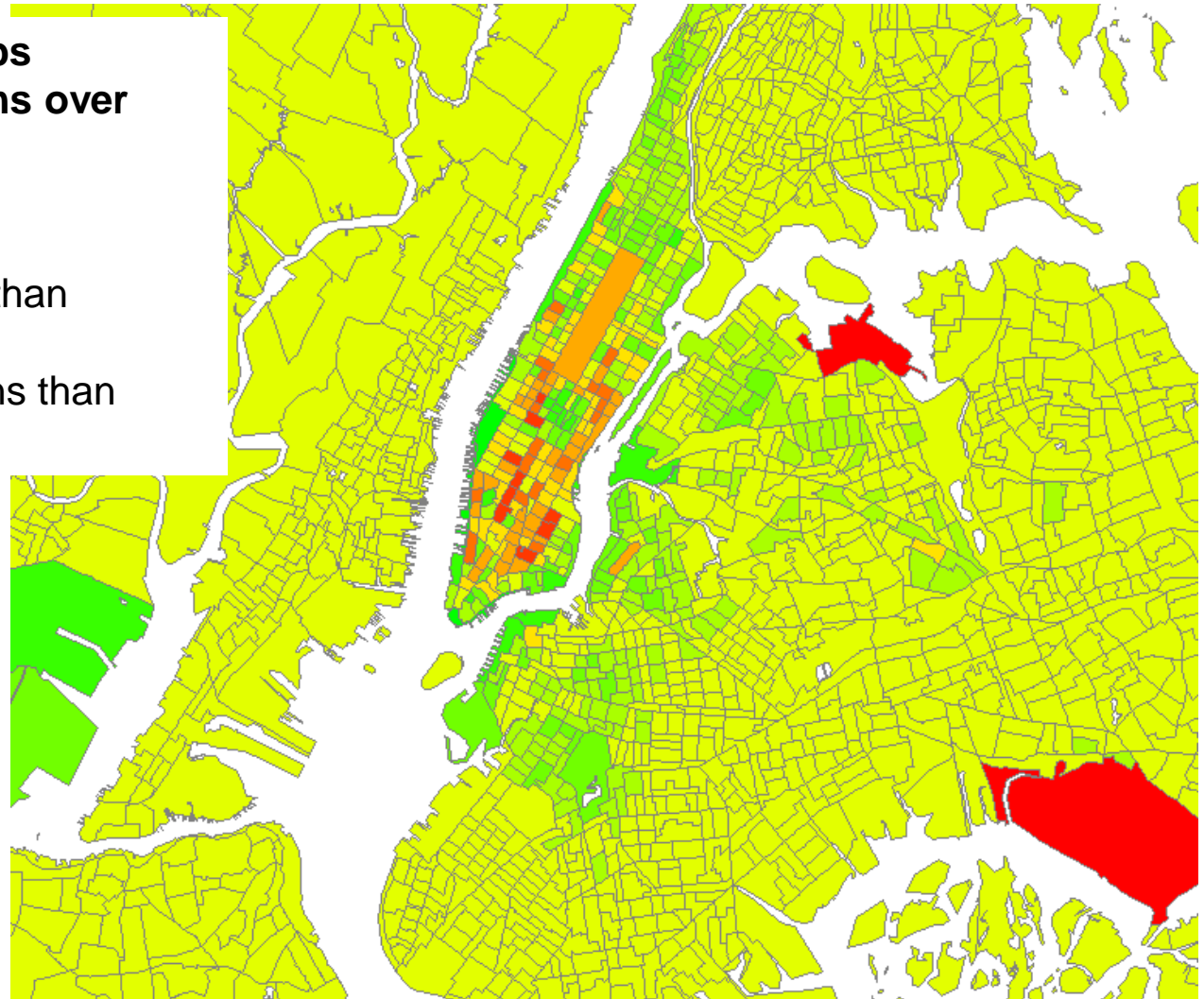


\* Courtesy: J. Peters. CSI/CUNY

**GPS tracking of taxi trips  
378,532,118 observations over  
18 month period**

**Red** = More originations than destinations

**Green** = More destinations than originations



\* Courtesy:

D. King, Columbia U  
J. Peters, CSI/CUNY  
C. Gordon, U of Canberra

# NEW YORK CITY TAXI ACTIVITY

## ORIGIN AND DESTINATION DENSITIES

**Red** = More originations than destinations  
**Green** = More destinations than originations

\* Courtesy  
D. King, Columbia U  
J. Peters, CSI/CUNY  
C. Gordon, U of Canberra

# Transportation Policy

Easypay usage

3 billion rides per year

Potential tracking of individual; user travel patterns

Linked to credit card

Posted Date	Transaction Date & Time	MetroCard Number	Ride Location	Transaction	Amount	Balance	Description
06/26/12	06/25/12 21:06:00	2228065429	R126 CHRISTOPHER ST SUBWAY	Ride: Subway	\$1.10	\$40.10	EasyPay Reduced Fare Plan
06/26/12	06/25/12 18:30:00	2228065429	R232 33 ST SUBWAY	Ride: Subway	\$1.10	\$41.20	EasyPay Reduced Fare Plan
06/26/12	06/25/12 12:06:00	2228065429	A069 CHAMBERS ST SUBWAY	Ride: Subway	\$1.10	\$42.30	EasyPay Reduced Fare Plan
06/22/12	06/22/12 20:52:42	-		Bonus	\$2.45	\$43.40	Reduced Fare
06/22/12	06/22/12 20:52:42	-		Payment Received	\$35.00	\$40.95	Reduced Fare
06/22/12	06/21/12 18:48:00	2228065429	R103 RECTOR ST SUBWAY	Ride: Subway	\$0.00	\$5.95	EasyPay Reduced Fare Plan
06/22/12	06/21/12 18:12:00	2228065429	X10 EXPRESS	Ride: Express Bus	\$5.50	\$5.95	EasyPay Reduced Fare Plan
06/18/12	06/17/12 21:36:00	2228065429	R138 34 ST-PENN STA SUBWAY	Ride: Subway	\$1.10	\$11.45	EasyPay Reduced Fare Plan
06/15/12	06/14/12 22:12:00	2228065429	Q30	Ride: Bus	\$1.10	\$12.55	EasyPay Reduced Fare Plan
06/12/12	06/11/12 18:42:00	2228065429	N083 W 4 ST-WASH SQ SUBWAY	Ride: Subway	\$1.10	\$13.65	EasyPay Reduced Fare Plan
06/12/12	06/11/12 17:54:00	2228065429	R103 RECTOR ST SUBWAY	Ride: Subway	\$0.00	\$14.75	EasyPay Reduced Fare Plan
06/12/12	06/11/12 17:12:00	2228065429	X10 EXPRESS	Ride: Express Bus	\$5.50	\$14.75	EasyPay Reduced Fare Plan
06/07/12	06/06/12 21:06:00	2228065429	A046 CANAL ST SUBWAY	Ride: Subway	\$1.10	\$20.25	EasyPay Reduced Fare Plan
06/07/12	06/06/12 18:30:00	2228065429	A060 WHITEHALL ST SUBWAY	Ride: Subway	\$1.10	\$21.35	EasyPay Reduced Fare Plan
06/05/12	06/04/12 10:38:00	2228065429	X10 EXPRESS	Ride: Express Bus	\$1.65	\$22.45	EasyPay Reduced Fare Plan



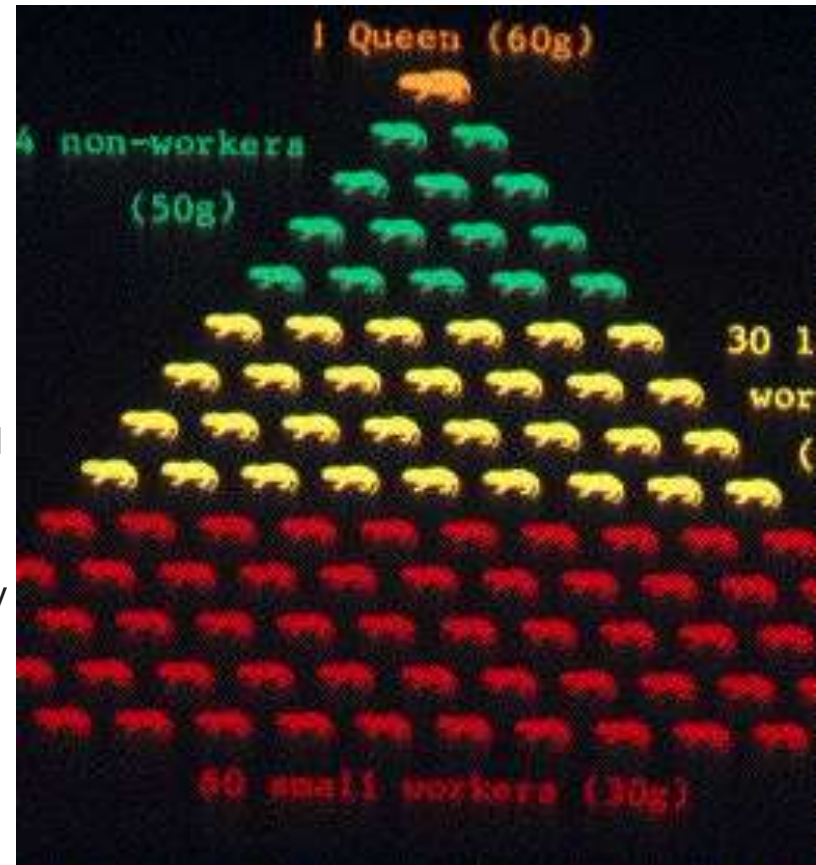


- Multi-level study of a “society” and its individual members
- Naked African mole rats
  - Not a mole
  - Not a rat
  - The only cold blood mammalian species (closest relative is the porcupine)
  - Offspring will likely never reproduce (less than 5% chance)
  - Genetically homogenous population
  - Life expectancy greater than 30 years
  - Skin is not sensitive to pain
  - Not known to ever develop cancer
  - Navigate using magnetic fields
  - Males and females are barely distinguishable
  - Communicate through chirps
  - Have individual recognition

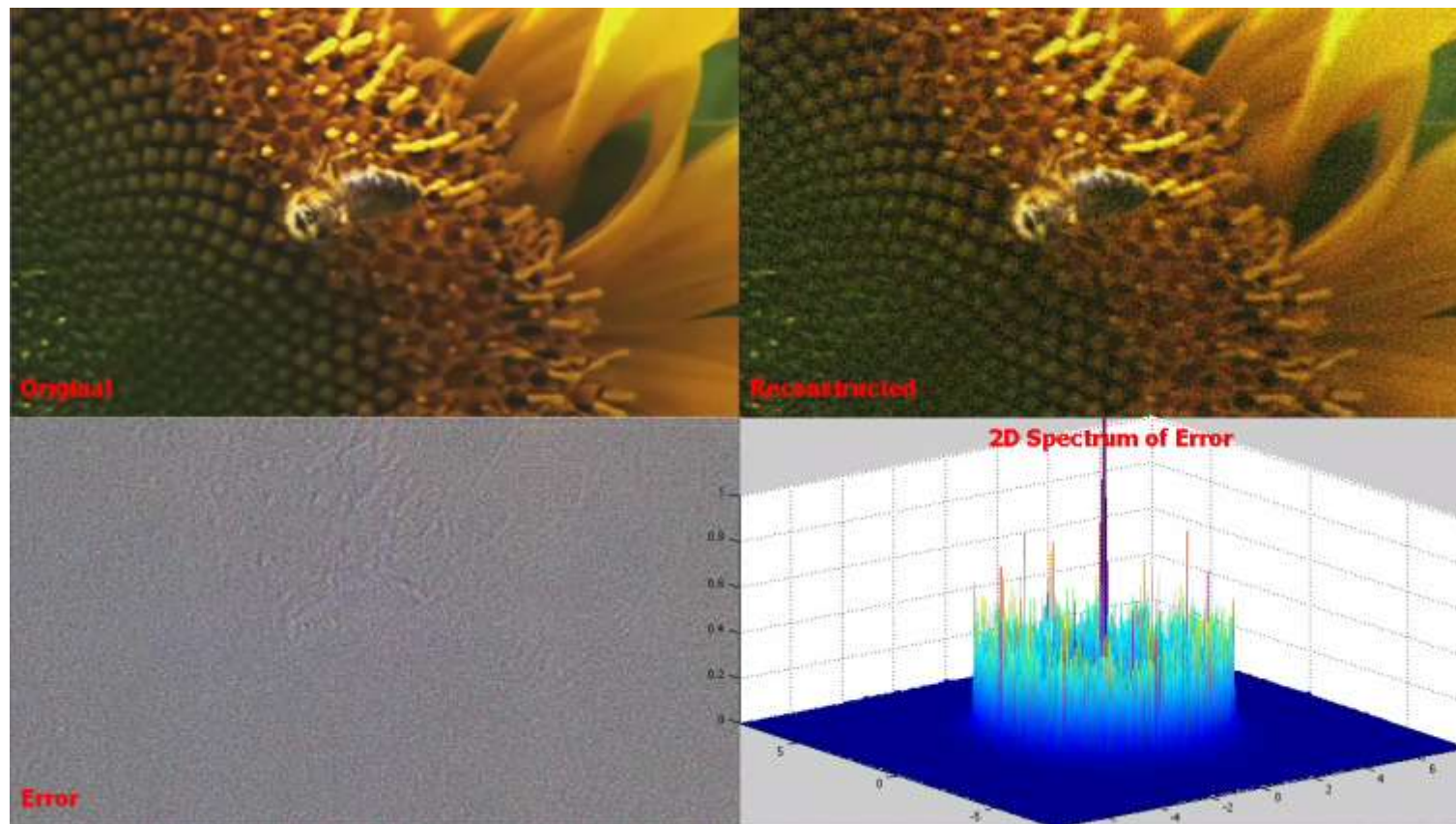


\* Courtesy; Dan McCloskey, College of Staten Island/CUNY

- **Structured society**
  - Only the queen has off-spring
  - Tunnel diggers
  - House keepers
  - Food gatherers
  - Exclusive society
- **Instrumented**
  - RFIDs to capture movement and social interactions
  - Monitor communications
  - Monitor oxytocin levels/blood chemistry
  - Brain structure analysis (MRI and biopsies)
- **Longitudinal studies over time**
  - 100s of millions of data points
  - Social graphs
- **Correlate social interaction/social position with:**
  - Structure of neuronal circuits
  - Oxytocin levels

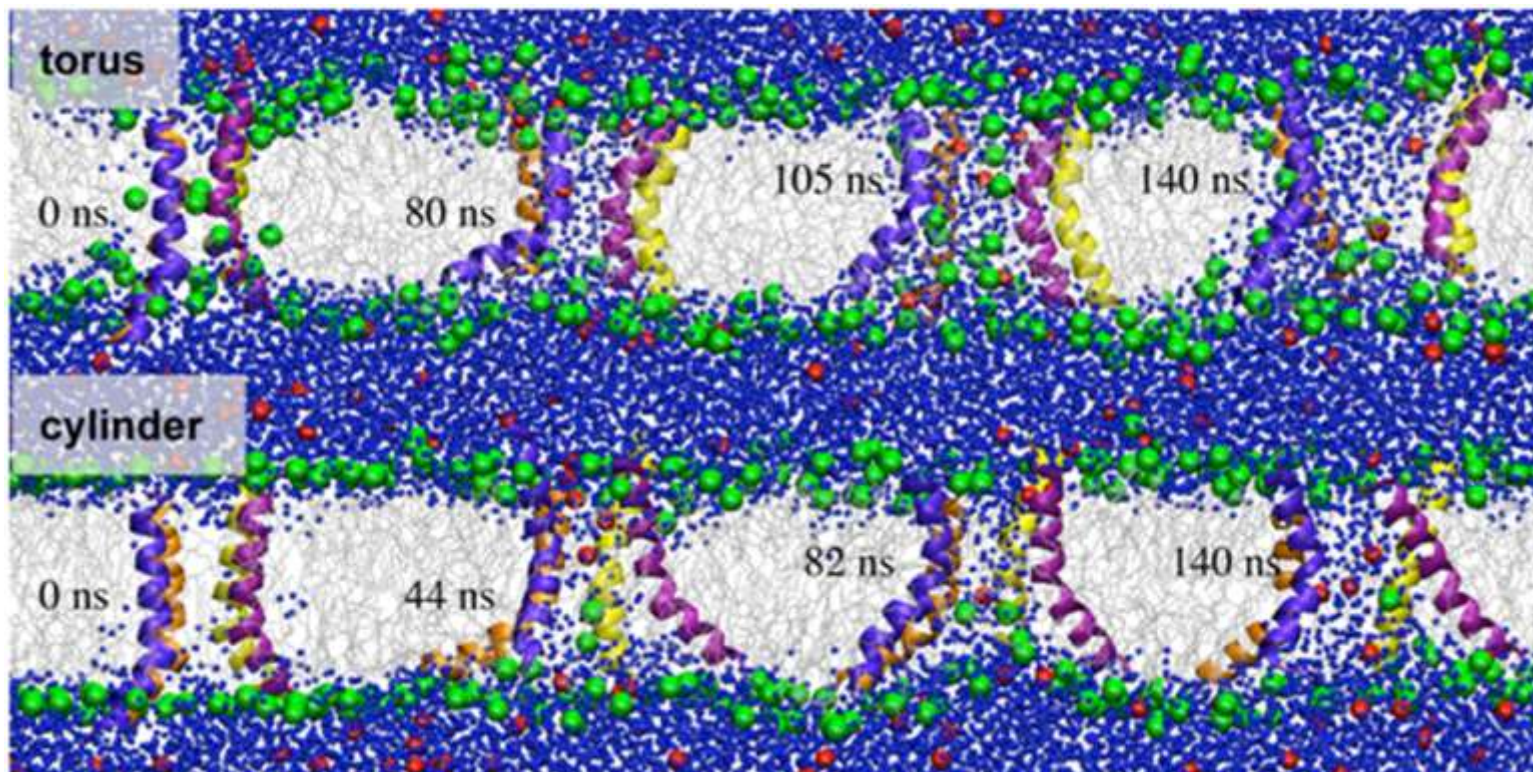


Courtesy; Dan McCloskey,  
College of Staten  
Island/CUNY



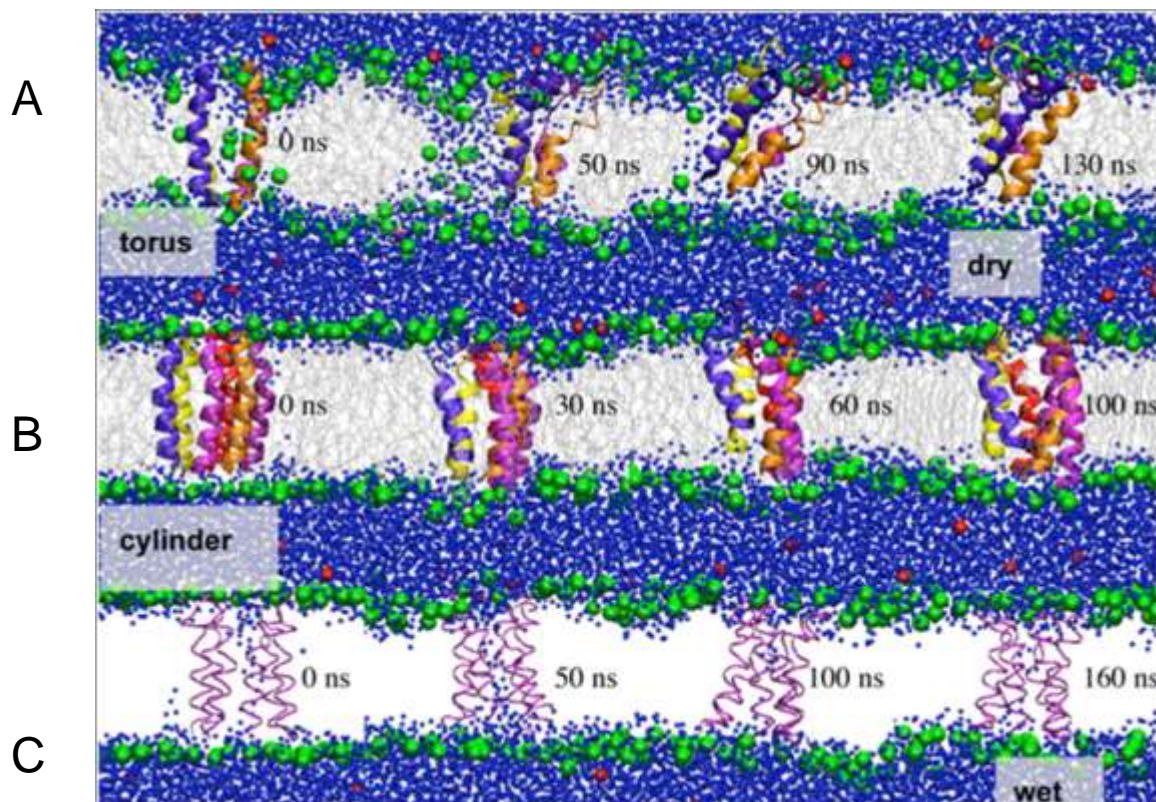
*Emulating the neural encoding and decoding circuits of the early vision system*

\* Courtesy: Aurel Lazar, BioNET Group, Columbia University  
 Computations using 55 GPGPU's on CUNY's Andy HPC system



Through this molecular dynamics calculation, we see that melittin preserved the toroidal pore shape or reshaped an initially barrel-stave pore into a semitoroidal pore. This indicates that the electrostatic interactions are stronger in toroidal shaped pores. In this figure, the green balls represent phosphocholines; the gray lines: lipid tails; red balls: Cl-ions; blue balls: water; cartoons: peptides.

\* Courtesy Drs. M. Mihajlovic & T. Lazaridis, City College/CUNY



To have a pore, water has to stay inside the pore region (“wet” pore); otherwise the peptides form an aggregate and the pore dries up (or closes). In our simulations, the latter happened when Q7’s were not located within the pore. Glutamine is a polar residue and thus important for providing the hydrophilic environment necessary to preserve the aqueous pore. These studies showed that alamethicin preserved the barrel-stave shaped pore. Here we see that Glutamine Q7 seems to be important for stabilizing a wet pore. The pore remained wet when the glutamine Q7’s were initially located within the pore, but dried up in the simulation when the Q7’s were not oriented towards the interior of the pore. In this figure, the green balls represent phosphocholines; gray lines: lipid tails; red balls: Cl- ions; blue balls: water; cartoons (in (A) and (B)), ribbons (in (C)): peptides; in (C), the lipid tails are removed for clarity.

- Display excellent scalability
- GPU versions available
  - Amber
  - DESMOND
  - Gromacs
  - NAMD

System		Cores	Chip	Memory/core (GB)	Interconnect
Arno	Kepler ready	1,152	SandyBridge E5-2660	2	QDR
Zeus	Gaussian	64	Harpertown	2	Ethernet
	Gaussian	16	Woodcrest	2	Ethernet
Bob	Gaussian	232	Barcelona	2	Infiniband
Andy	Serial & small parallel	744	Nehalem	3	DDR and QDR
	GPU	96 x 448	GPU - FERMI	3	PClex Gen2
SGI UV-2	Genomics Graphs	512	SandyBridge E5-4600	8 TB	SMP
Salk	Large scalable	2,816	Magny-Cours	2	Custom - Gemini
Karle		24	Westmere	96	SMP

Arno – Arno Penzias, physicist, Nobel Laureate

Andy – Andy Grove, co-founder, Intel

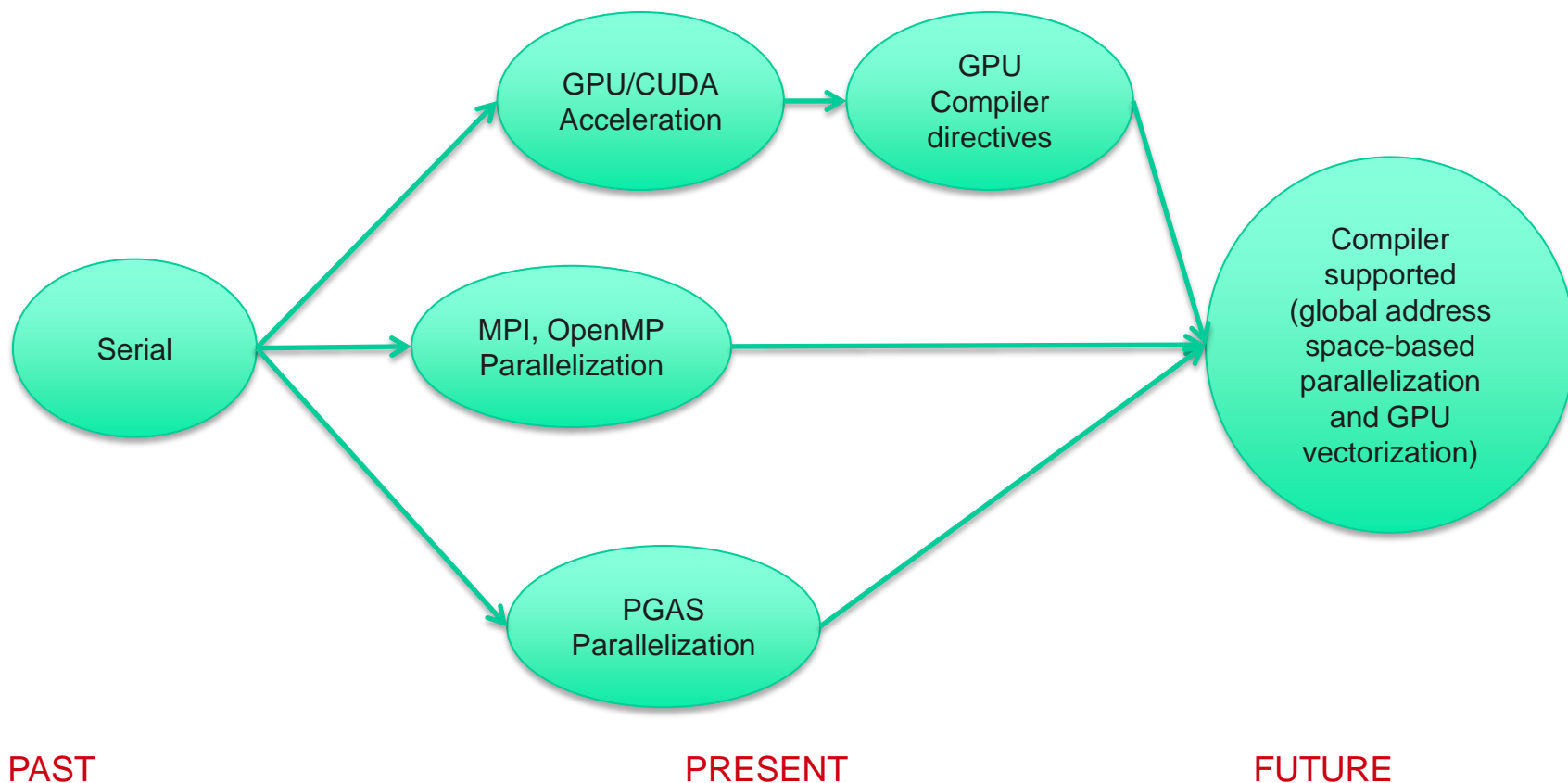
Karle – Jerome Karle, mathematician, Nobel Laureate

Bob – Robert Kahn, co-developer TCP/IP

Salk – Jonas Salk, developer of the polio vaccine



# HPC Philosophy/Vision



The HPCC has the state-of-the-art parallel HPC systems supporting the following programming models: OpenMP, MPI, GPU acceleration, and partitioned global address (Co-array Fortran and Unified Parallel C. This anticipates a future HPC trend leading to unified HPC programming models

- Application development is now more difficult, not less
- One computing architecture does not fit all
- Breakthrough research requires multi-disciplinary skills--universities are not meeting the challenge
  - Departments are stove-piped
  - Academic change is slow, change in technology is rapid
  - Pushing internal change is a key role of a university HPC Center
- New skills are required – big data, semantic data, graph theoretic approaches
- Shortage of skilled people
- Universities lack computational resources
- Energy consumption per \$ spent on hardware has grown by an order of magnitude in 25 years

$10^{25}$  ?

- Telescope
- Microscope
- Time machine



- Augmented with the help of HPC

- New 170,000 sq. ft. facility in planning stages
- 10,000 sq. ft. of raised floor
- Staff offices
- \$6.5 million committed for facility design
- Expected occupancy: 2017

