



Open Science Grid

# Science on the OSG in 2014

Bo Jayatilaka  
*Fermilab*

**OSG All-Hands Meeting**  
Northwestern University  
March 25, 2015





# Goals of the OSG

---

- The OSG **enables** science via access to **DHTC**
- The **foundation** of the OSG is the computing infrastructure of **experiments** (e.g. ATLAS and CMS) in collaboration with IT organizations at **universities and national labs**
- The OSG continues to meet the needs of these (largest) stakeholders
  - Record years of utilization by LHC experiments, even during shutdown
- A wide variety of users and projects can benefit from **opportunistic access** to the OSG
  - Take advantage of unused cycles on the 140+ compute elements on the OSG



# What this talk is about

---

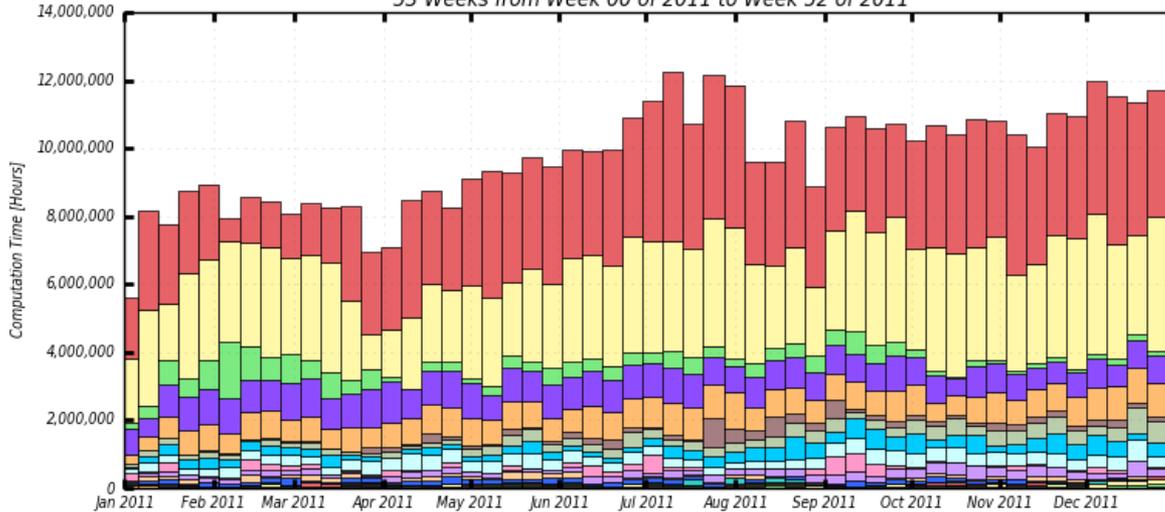
- **Opportunistic computing** at the OSG in 2014
  - How much, where, and for whom
- **Highlights of science** accomplished using opportunistic OSG resources
  - Disclaimer #1: highlights only
  - Disclaimer #2: emphasis on opportunistic (not HEP)



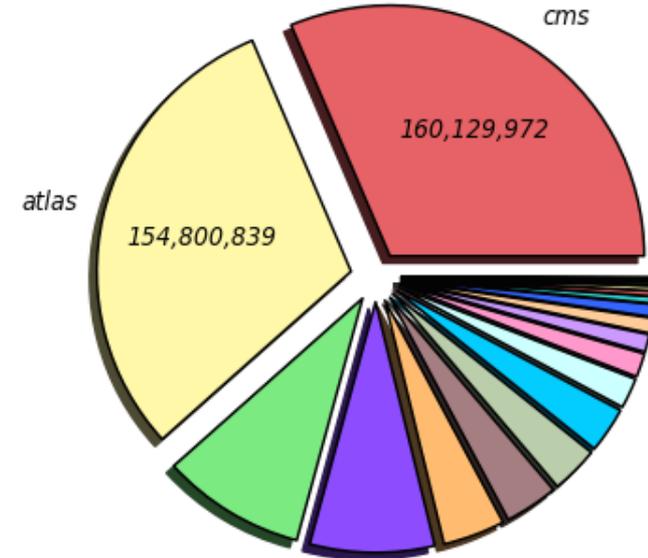
# A brief look back: OSG in 2011

### Hours Spent on Jobs By VO

53 Weeks from Week 00 of 2011 to Week 52 of 2011



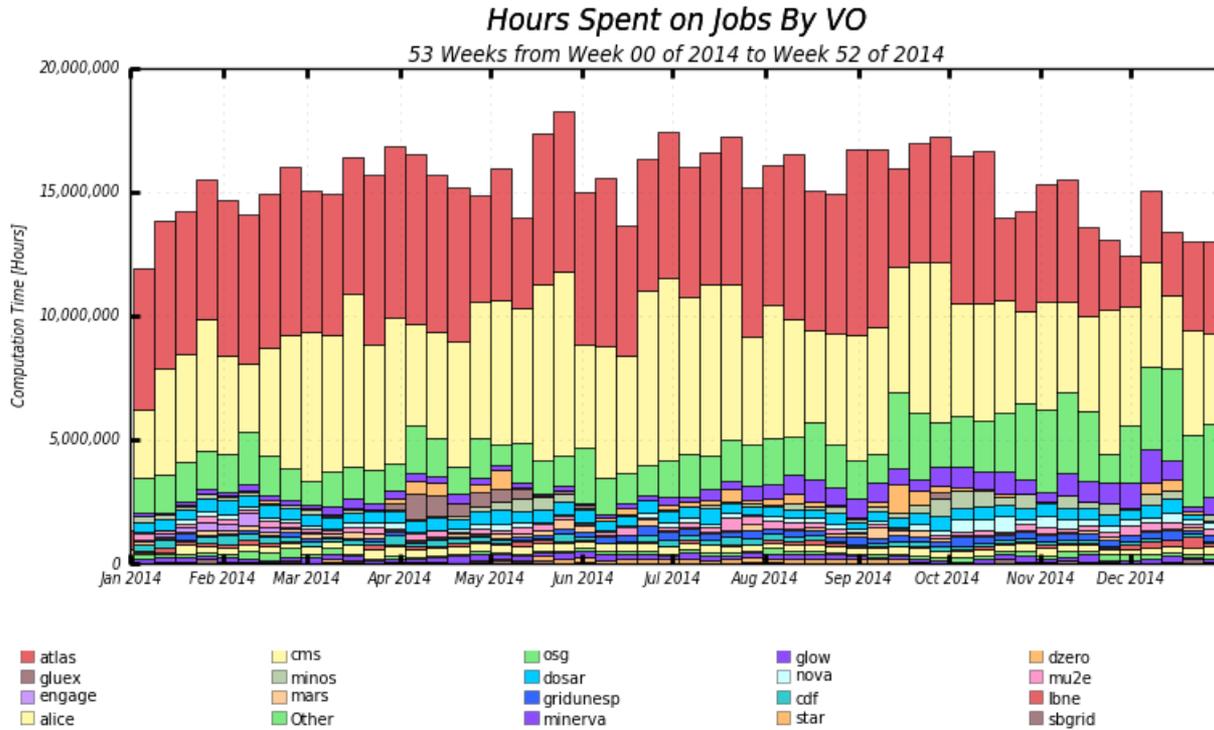
Maximum: 12,250,215 Hours, Minimum: 4,021,849 Hours, Average: 9,635,310 Hours, Current: 4,021,849 Hours



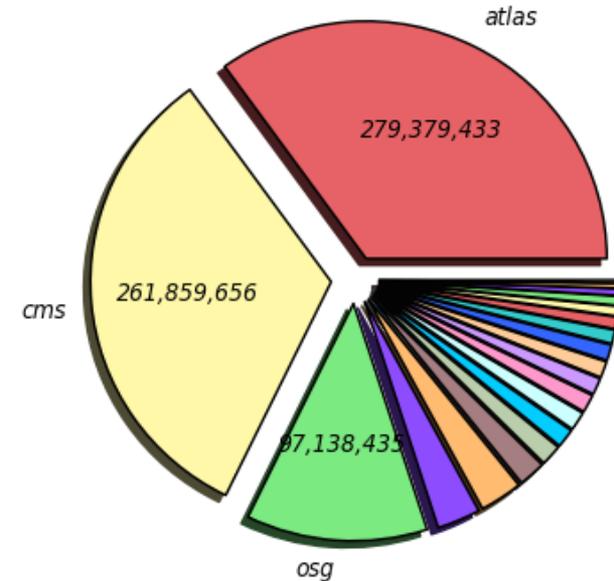
- 510M wall hours
  - 430M wall hours (~85%) from large HEP experiments (ATLAS, CMS, CDF, D0) running on their own sites



# 2014 on the OSG in review



Maximum: 18,264,151 Hours, Minimum: 1,349,098 Hours, Average: 15,066,136 Hours, Current: 12,993,468 Hours



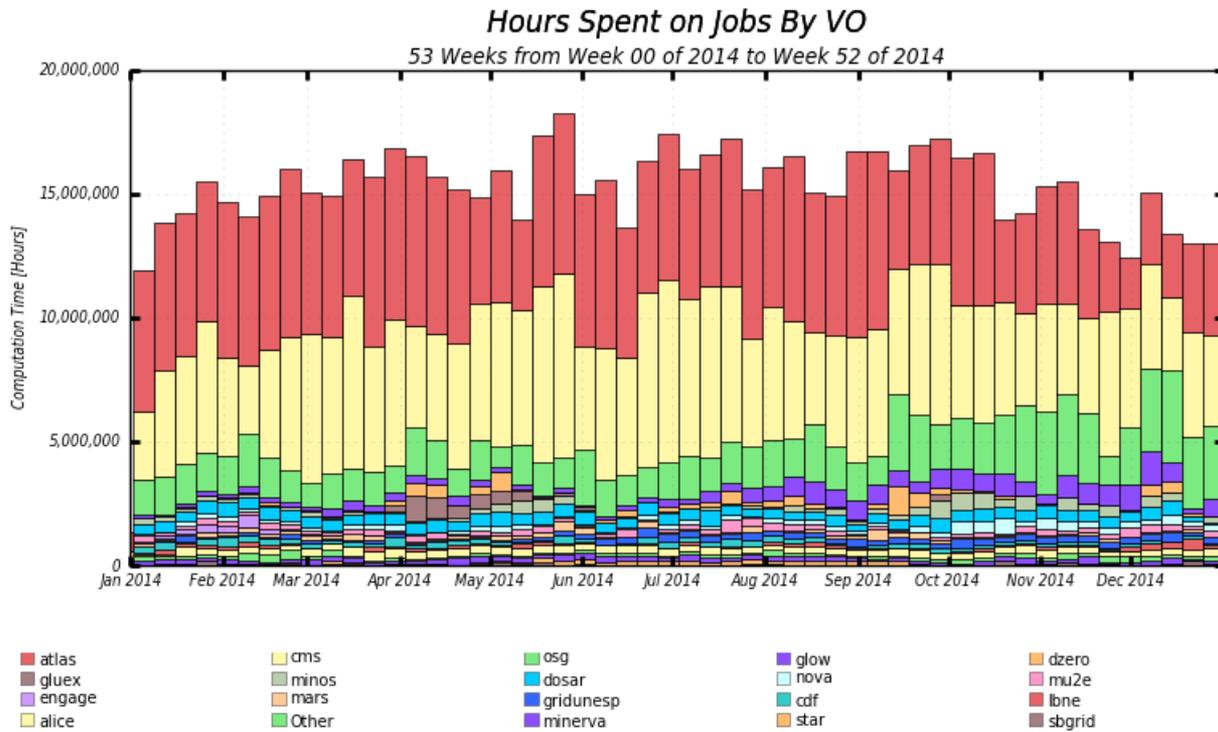
- **~800M wall hours**

- 541M from ATLAS and CMS (~67%)

- Adding all other HEP experiments brings the total to 74%

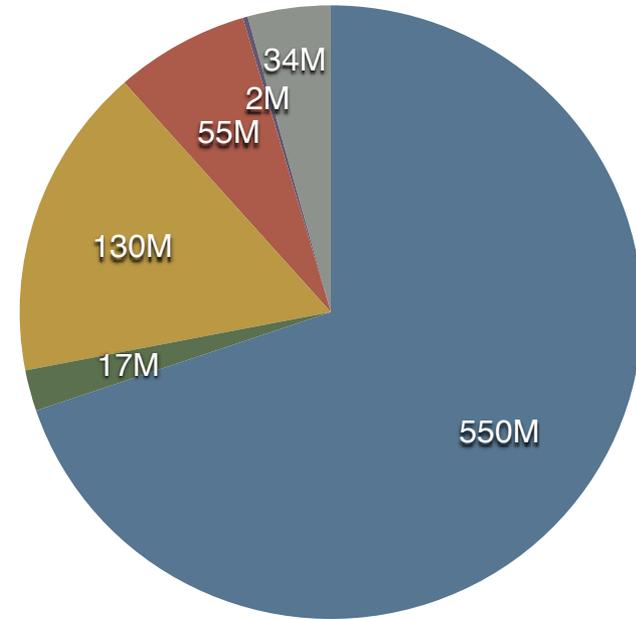


# 2014 on the OSG in review



Maximum: 18,264,151 Hours, Minimum: 1,349,098 Hours, Average: 15,066,136 Hours, Current: 12,993,468 Hours

- LHC
- Opportunistic
- Cosmic Frontier
- Tevatron
- Intensity Frontier
- Other



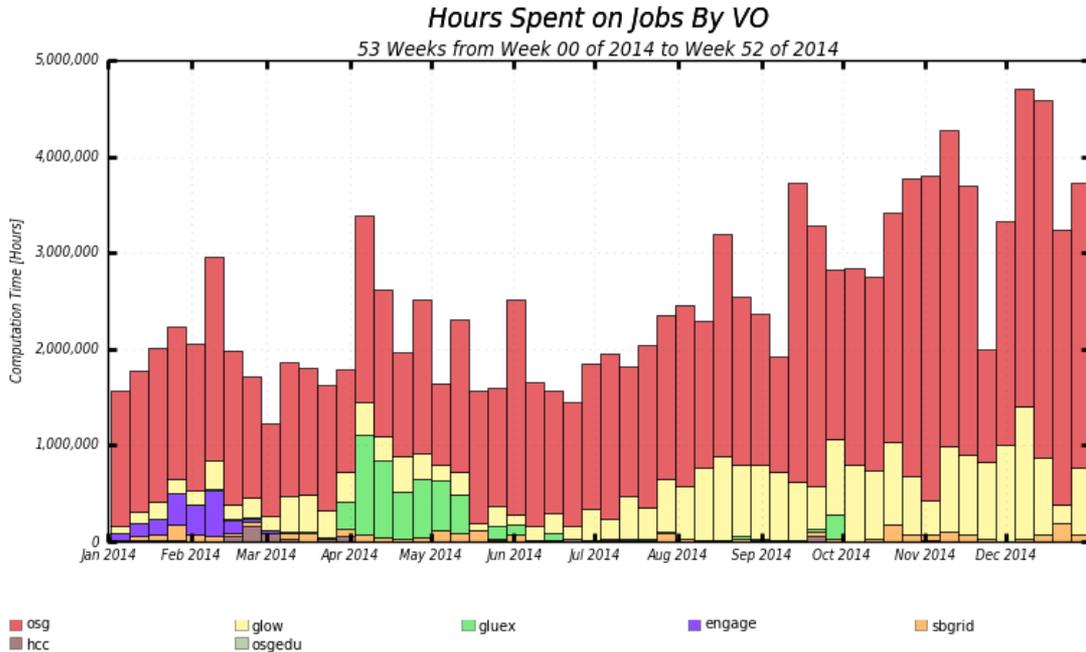
- **~800M wall hours**

- 541M from ATLAS and CMS (~67%)

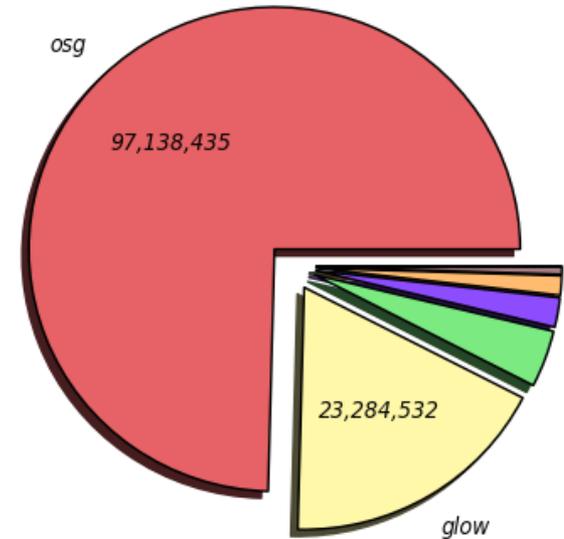
- Adding all other HEP experiments brings the total to 74%



# Opportunistic VOs in 2014

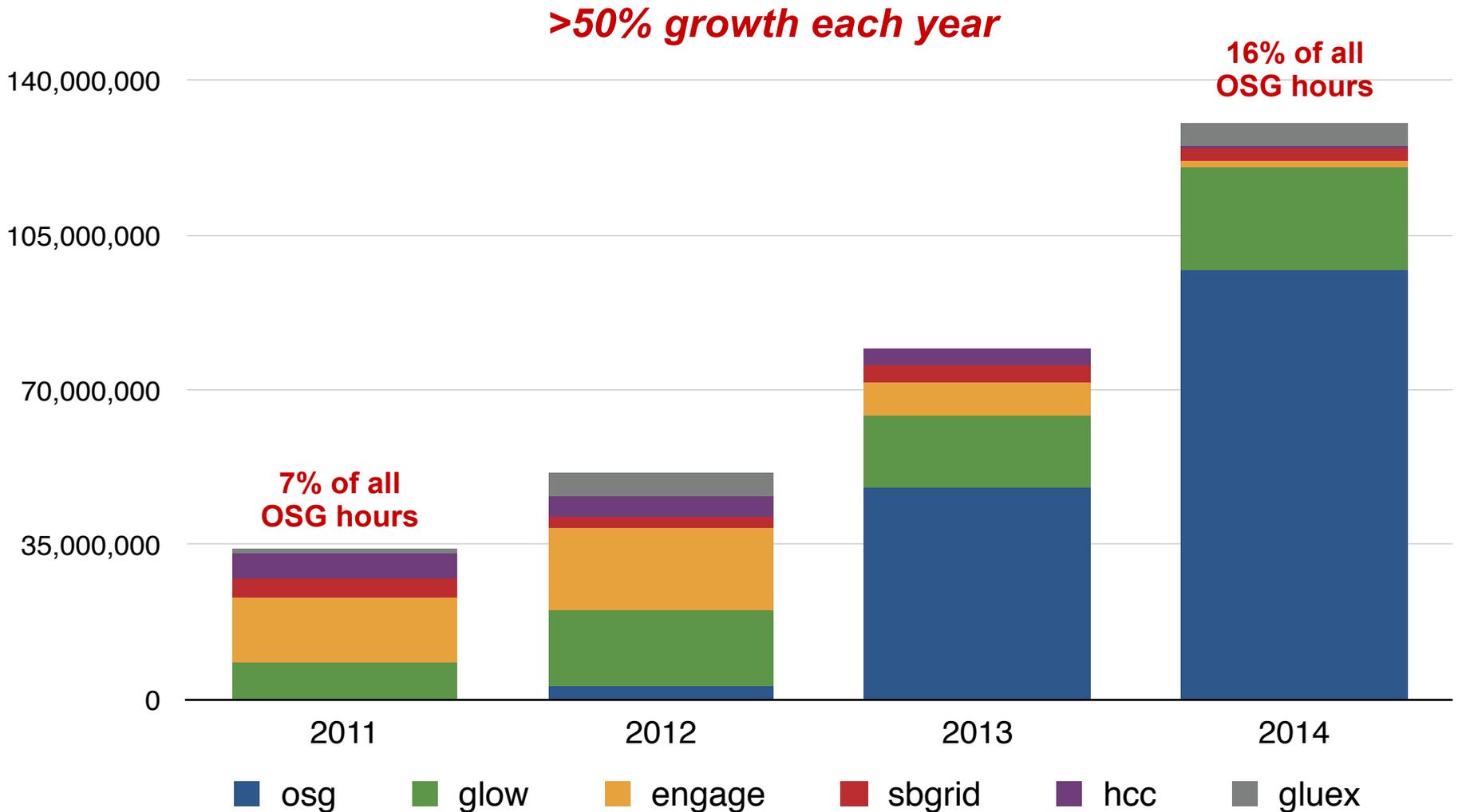


Maximum: 4,699,677 Hours, Minimum: 45,027 Hours, Average: 2,457,438 Hours, Current: 3,728,689 Hours



- Primarily opportunistic VOs (osg, glow, gluex, engage, sbgrid, hcc) received **130M wall hours** in 2014
  - 16% of all OSG hours

# Growth of opportunistic VOs





# Where do these hours come from?

## Top 20 Opportunistic Sites on the OSG in 2014

Site	engage	glow	hcc	osg	sbgrid	Total Opportunistic	Total hours	Percent opportunistic
CIT_CMS_T2		3,416,415	29,763	9,102,039	511,981	13,060,198	28,180,174	46%
FNAL_FERMIGRID	16,874		108,624	12,711,968		12,837,466	82,519,460	16%
UCSDT2	485,006	1,253,139	13,972	10,713,436	252,019	12,717,572	25,256,162	50%
USCMS-FNAL-WC1	136	822,270	5,268	11,073,515	281,014	12,182,203	91,395,332	13%
Tusker	284,971	3,369,581	24,104	7,490,094		11,168,750	14,751,109	76%
Nebraska	35,966	3,333,533	91,792	6,680,426		10,141,717	35,277,858	29%
MWT2	454,547	2,282,363	36,151	6,219,981		8,993,042	61,095,144	15%
MIT_CMS		57	78,228	4,585,284	809,369	5,472,938	24,356,265	22%
Sandhills	23,112	1,748,946	21,573	2,034,164		3,827,795	4,232,466	90%
Crane	66,077	1,492,771	11,247	1,873,012		3,443,107	3,977,168	87%
Purdue-Hadoop		644,142	8,559	2,385,959	26,429	3,065,089	10,888,174	28%
UConn-OSG	76,977	1,081,891	43,159	1,608,189		2,810,216	3,174,455	89%
AGLT2	18,473	1,033,200	10,168	1,327,397	219,954	2,609,192	35,204,179	7%
GridUNESP_CENTRAL	5,587	474,813	5,062	1,974,711	61,271	2,521,444	12,784,909	20%
GLOW		109,332	14,679	2,233,313		2,357,324	27,170,835	9%
NWICG_NDCMS				2,281,689		2,281,689	2,728,851	84%

- From the OSG bylaws: *[OSG] Consortium members recognize that the OSG is a sharing eco-system and **strive to maximize the sharing of computing resources, software, and other assets to enable science.***



# Sharing is good!

## Open Science Grid

Opportunistic Resources provided by the top 10 OSG Sites for the OSG Open Facility (2014-09-01 - 2014-9-30).

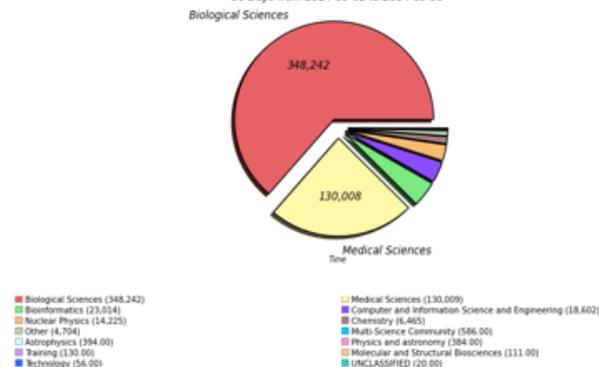
SITE	INSTITUTION (ORG)	RECENT PERIOD RANK	RECENT PERIOD HOURS	PRIOR PERIOD RANK	PRIOR PERIOD HOURS
USCMS-FNAL-WC1	Fermi National Accelerator Laboratory (USCMS Tier1)	1	1,295,720	7	395,587
MWT2	University of Chicago, Indiana University, University of Illinois, Urbana-Campaign (USATLAS Midwest Tier2)	2	1,291,273	4	725,769
UCSDT2	University of California San Diego (USCMS Tier2)	3	1,050,015	1	1,091,183
FNAL_FERMIGRID	Fermi National Accelerator Laboratory (Campus Grid gateway)	4	832,573	3	853,682
MIT_CMS	Massachusetts Institute of Technology (USCMS Tier2)	5	546,943	8	351,981
CIT_CMS_T2	California Institute of Technology (USCMS Tier2)	6	532,943	2	920,799
UFlorida-HPC	University of Florida (USCMS Tier2)	7	509,245	63	0
Tusker	The Holland Computing Center at the University of Nebraska-Omaha (HCC Campus Grid)	8	488,855	6	520,391
Nebraska	The Holland Computing Center at the University of Nebraska-Lincoln (USCMS Tier2)	9	479,262	5	697,329
UConn-OSG	University of Connecticut at Storrs (GLUEX)	10	191,622	11	158,809

## Open Science Grid

SITE MIT\_CMS (2014-09-01 - 2014-9-30):

MIT\_CMS has provided 546,917 wall hours to the OSG Open Facility during 2014-09-01 - 2014-9-30. Last period this site provided 351,981 wall hours.

Wall Hours by Field of Science (Sum: 546,943 Hours)  
30 Days from 2014-09-01 to 2014-09-30



- In our experience site owners like to see their resources **being used for science**
  - Site reports show, in detail, how opportunistic resources were used at a site

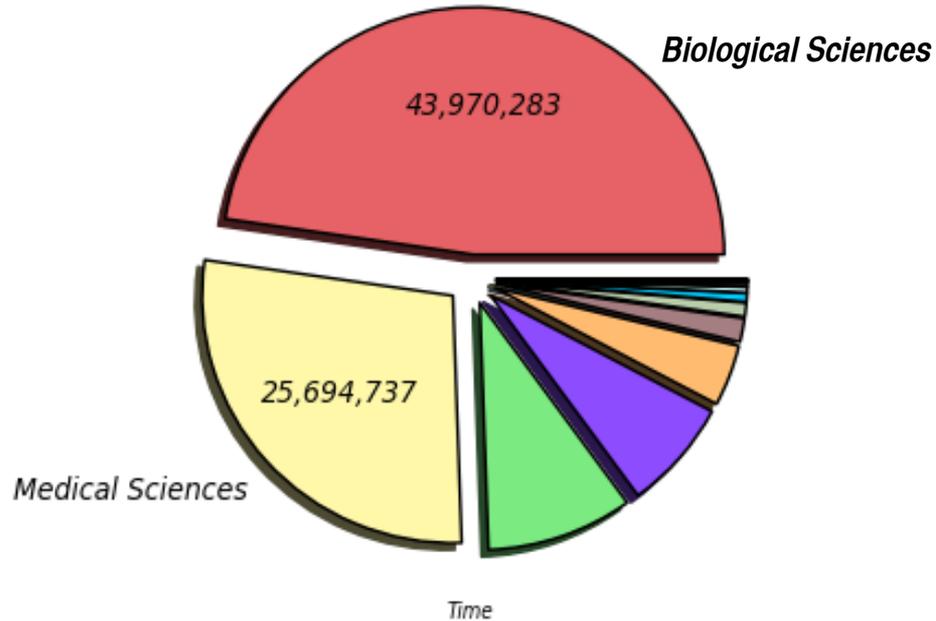
PROJECT NAME	FIELD OF SCIENCE	PI	VO	SUBMISSION HOST	WALL HOURS
<a href="#">TG-IBN130001</a>	Biological Sciences	Donald Krieger	osg	XD-LOGIN	348,242
<a href="#">SPLINTER</a>	Medicine	Robert Quick	osg	XD-LOGIN	109,152
<a href="#">ProtFolding</a>	Bioinformatics	Jinbo Xu	OSG-Connect	OSGCONNECT	22,599
<a href="#">DetectorDesign</a>	Medical Imaging	John Strogas	osg	XD-LOGIN	20,857
<a href="#">Orbiter</a>	Mathematics	Anton Betten	OSG-Connect	OSGCONNECT	17,632
<a href="#">Duke-QGP</a>	Nuclear Physics	Steffen A. Bass	osg	XD-LOGIN	14,199
<a href="#">AIGDock</a>	Chemistry	David Minh	OSG-Connect	OSGCONNECT	6,465
<a href="#">KnowledgeLab</a>	Other	James Evans	OSG-Connect	OSGCONNECT	4,704



# Who uses these hours?

Wall Hours by Field of Science (Sum: 92,202,816 Hours)

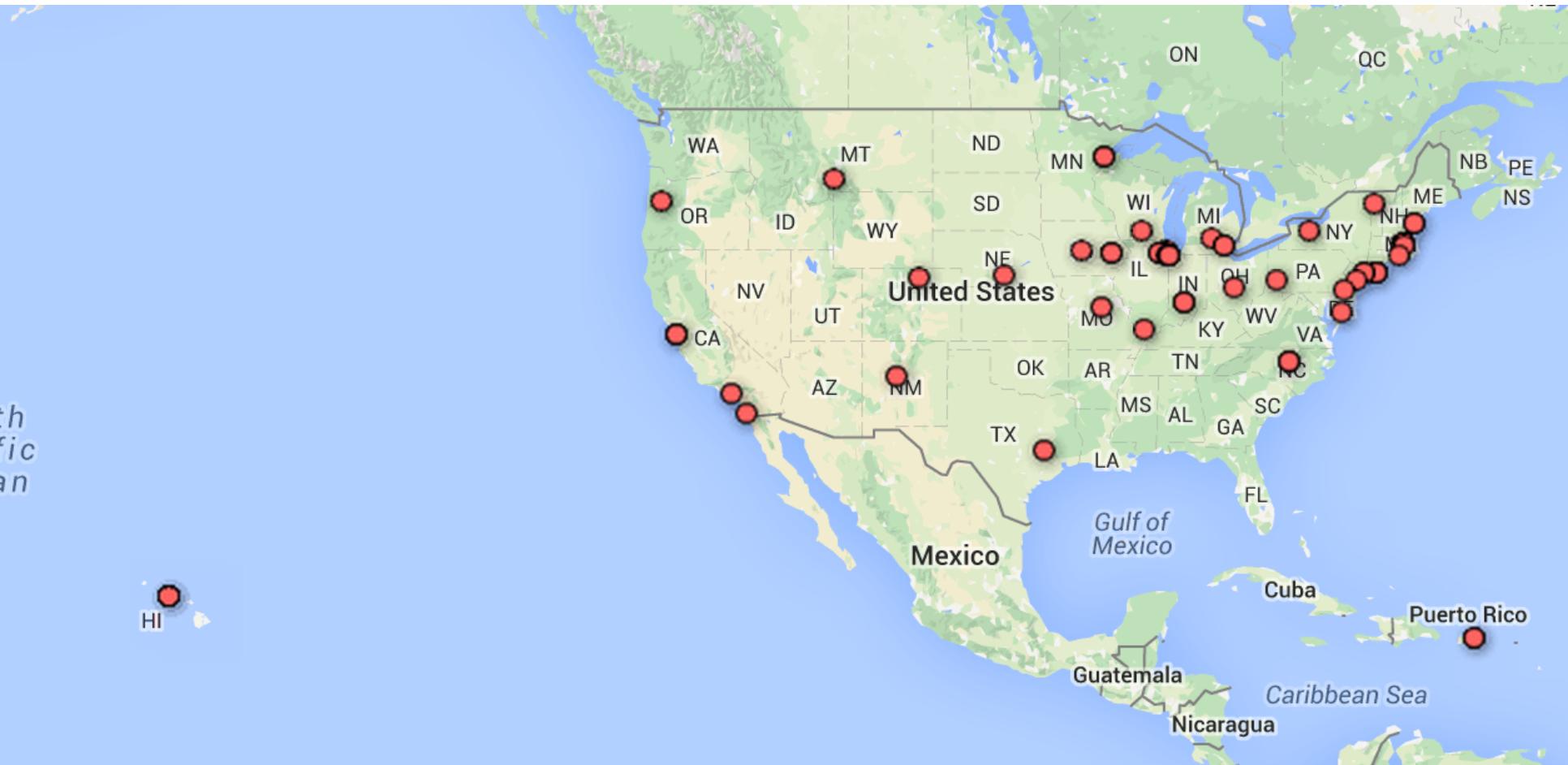
365 Days from Week 00 of 2014 to Week 52 of 2014



- Biological Sciences (43,970,284)
- High Energy Physics (8,710,754)
- Computer and Information Science and Engineering (3,688,696)
- Materials Science (818,120)
- Physics and astronomy (332,932)
- Plant Biology (58,717)
- Molecular and Structural Biosciences (37,793)
- Information, Robotics, and Intelligent Systems (20,401)
- Medical Sciences (25,694,738)
- Nuclear Physics (6,661,506)
- Bioinformatics (1,409,715)
- Microbiology (508,368)
- Chemistry (204,194)
- Ocean Sciences (42,793)
- Multi-Science Community (24,328)
- Evolutionary Sciences (19,479)



# Who uses these hours?



Institutions of users from **OSG**, **GLOW**, and **HCC** VOs in 2014

# Opportunistic VO Publications

---

- Papers published or submitted in 2014 (reported in OSG Annual Report)
  - GLOW: 49
  - HCC: 2
  - SBGrid: 616
  - OSG Open Facility: 23
- Including 33 in *Nature* and 13 in *Science*

# **Science Highlights from 2014**

# Some (more) disclaimers

---

- Focus is more on “individual” researchers using opportunistic VOs (rather than larger VOs/experiments)
- Really meant to be a **sample** of the science done opportunistically using the OSG
  - Much more than time permits
- Despite the best efforts of the researchers who provided me material to present I am **not** an expert (a casual follower at best) in any of these areas
  - See presentations by actual experts and their use of the OSG this afternoon (topics not covered in this talk)



# Human immune response to tuberculosis

- TB infects ~2 billion annually
  - ~1.5 million deaths
- Determined parameters for 2 main TB drugs that contribute to long treatment
  - Opportunity to test more drug regimens and guide clinical evaluation
- Defined a pro/anti inflammation polarization ratio, predictive of disease progression, and suggesting NfκB signaling dynamics as a therapeutic outcome
- Used a novel systems biology approach to identify TB biomarkers
  - Discovered that linking data from blood to sites of infection is key to identifying biomarkers and can accelerate their discovery, aiding development of vaccines and intervention strategies

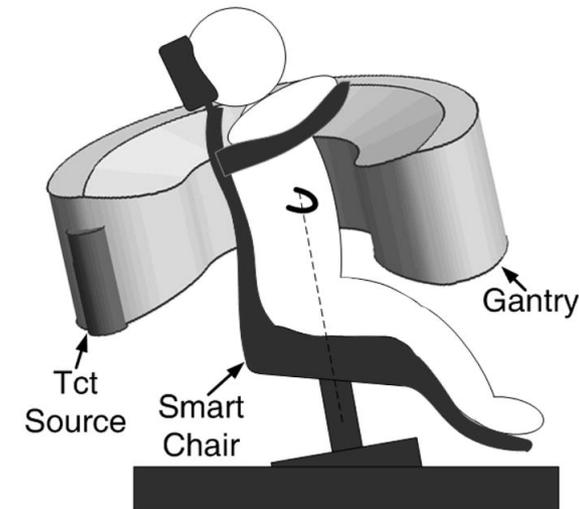
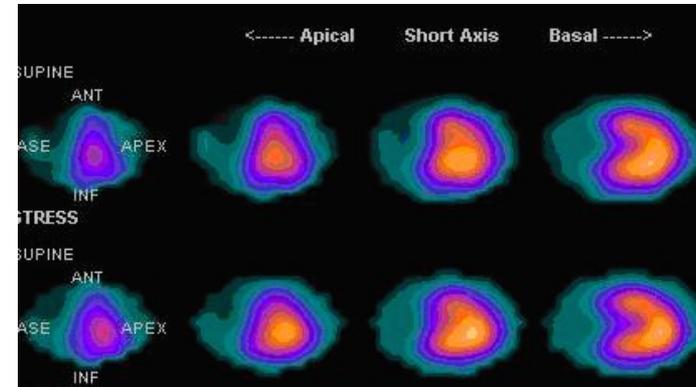
Paul Wolberg  
*University of Michigan*



# Detector Design

- Non-invasive diagnosis of coronary artery disease requires imaging of the myocardium
  - Predominantly with SPECT (single photon emission computed tomography)
  - Radio-pharmaceuticals+collimators and radiation detectors allow 3D imaging of the myocardium
  - Literally see areas not getting enough blood
- Design of new C-SPECT optimized for cardiac imaging (cheaper x2 and more sensitive x2-3)
  - Extensive Monte Carlo simulations utilized in design studies as well as in calibrating prototypes

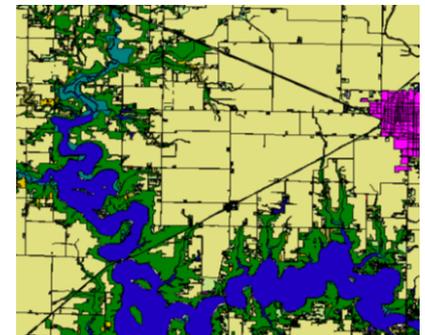
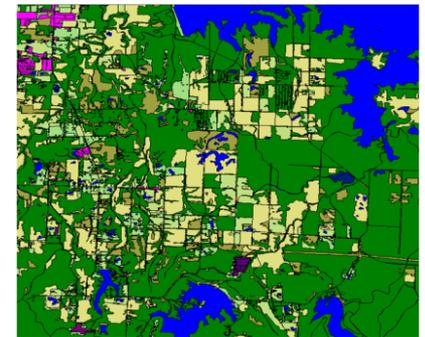
John Strologas, Wei Chang  
*Rush University*



# Agent-based modeling of disease transmission in white-tailed deer

- Chronic wasting disease (CWD) in white-tailed deer
  - Prion disease (like mad cow) causing fatal degeneration of the brain
  - Not much known about the disease and its long-term effects on deer population
- Simulate spread of CWD with an agent-based model (DeerLandscapeDisease)
  - Each deer modeled separately with its own behavioral rules (landscapes, empirically-based movement, deer behavior, disease)
  - OSG used for CPU
- DLD shows transmission may be a mix of direct (contact) and indirect (environmental) transmission
  - Potential for coexistence
  - Landscape and social structure play a major part in CWD transmission

Lene Jung Kjaer  
*Southern Illinois University*

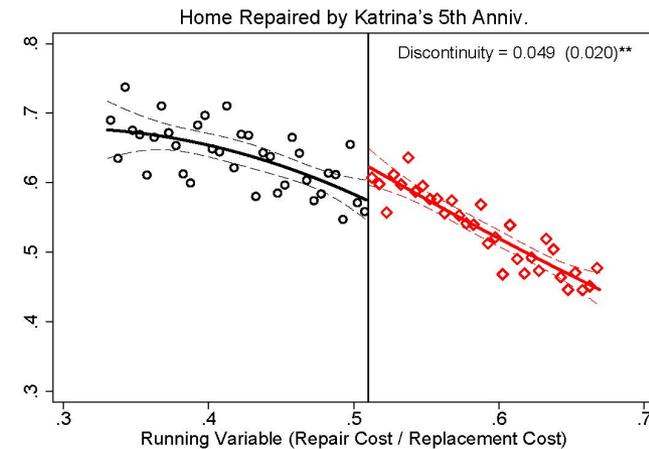
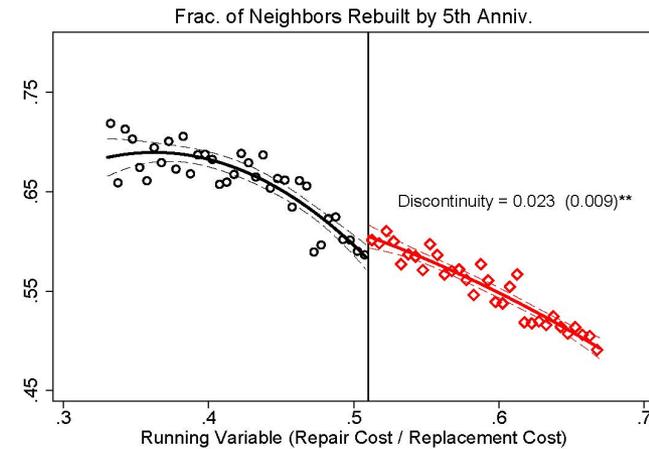




# Evaluating post-Katrina rebuilding grants

- A study of post-disaster rebuilding
  - In what way does policy (grants) influence individual rebuilding/relocation decisions?
- Post-Katrina grants were more generous for homeowners rebuilding in the same location
  - What “spillover” effect is caused by this?
- Simulate rebuilding choices under a large number of parameter variations
  - Model includes preferences for building location as well as other goods consumption
  - Iterate model choices until results match what was observed in Louisiana

Jesse Gregory  
*University of Wisconsin*

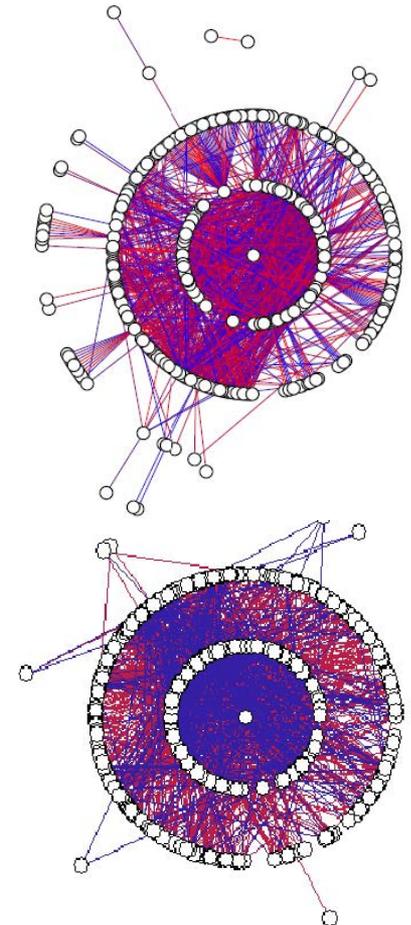




# Modeling of bank networks

- What's the optimal structure for a financial system?
  - Critical question for informing banking regulation policy
  - Is “too interconnected to fail” always a bad thing?
- Simulate financial systems with various degrees of maximum interconnectedness
  - Study of efficiency vs. stability
  - DHTC ideal since parallel “financial systems” don't need to talk to each other
- Result: restricting bank counterparties has (perhaps) unforeseen adverse effects

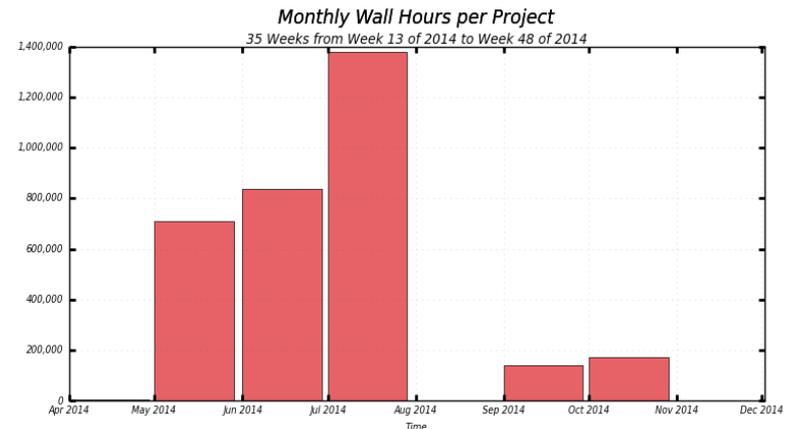
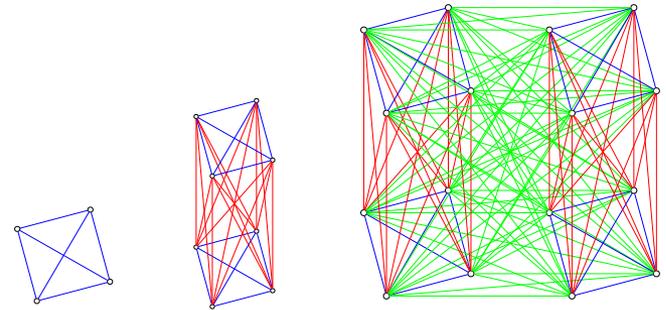
Michael Gofman  
Alexander Dentler  
*University of Wisconsin*



# Enumeration of all $(C_5, C_5, C_5; n)$ Colorings

- A  $(G_1, G_2, \dots, G_k; n)$  coloring is a coloring of the edges of the complete graph on  $n$  vertices with  $k$  colors such that the  $i^{\text{th}}$  color does not contain the graph  $G_i$
- $R_3(C_5)=17$ . How many  $(C_5, C_5, C_5; 16)$  colorings are there?
- Start with smaller  $C_5$  colorings
  - Initial input of 140M  $n=12$  colorings
  - Iterated by adding one vertex at a time
- Result: 1,701,746,176 Ramsey 3-colorings avoiding  $C_5$ , the cycle of length 5, in all 3 colors

David Narváez,  
Stanislaw Radziszowski  
*Rochester Institute of Technology*

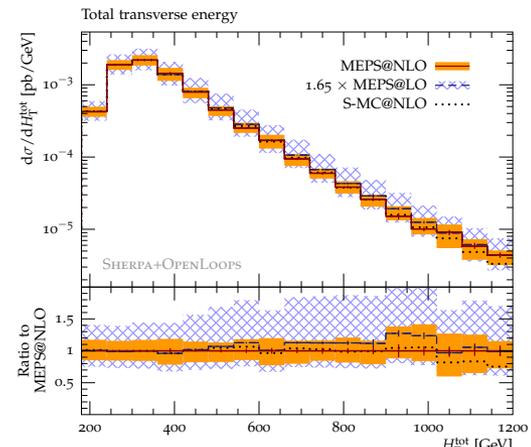
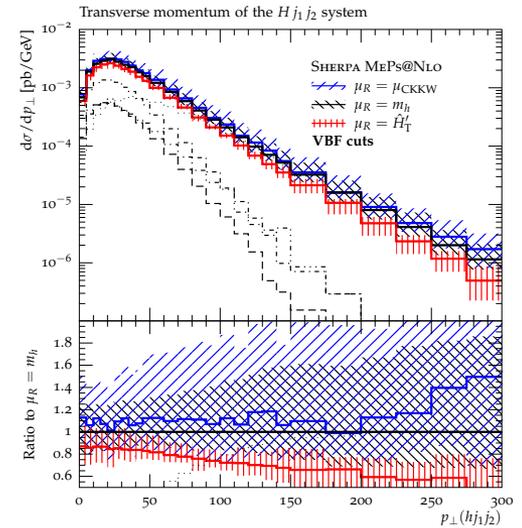




# Collider phenomenology on the grid

- LHC final states are very complex and interesting signals are often in corners of phase space
- Precise theoretical prediction of known processes crucial in finding new physics
- OSG resources used for refining modeling of SM processes and reducing theory uncertainties
  - gluon fusion Higgs+jets production
  - top quark pairs+jets
  - Both cases  $\leq 2$  jets@NLO and 3 jets @LO
- Uses Sherpa+Rivet+MCFM+OpenLoops
- Each process used O(500k) hours over 1-2 weeks

Stefan Höche  
SLAC







# Conclusion

---

- Opportunistic resources and their consumption continue to grow on the OSG
- A vibrant opportunistic user community co-exists with large (site-owning) VOs
  - A wide range of research is conducted (and published) using results obtained using OSG resources
  - More details can be found in the OSG Annual Report
- Always looking for more campus researchers (talk to Rob) and VOs (talk to me) who want to take advantage of these resources
  - And more sites who want to provide these resources!
- A **big thank you** to the researchers who provided material for this talk!



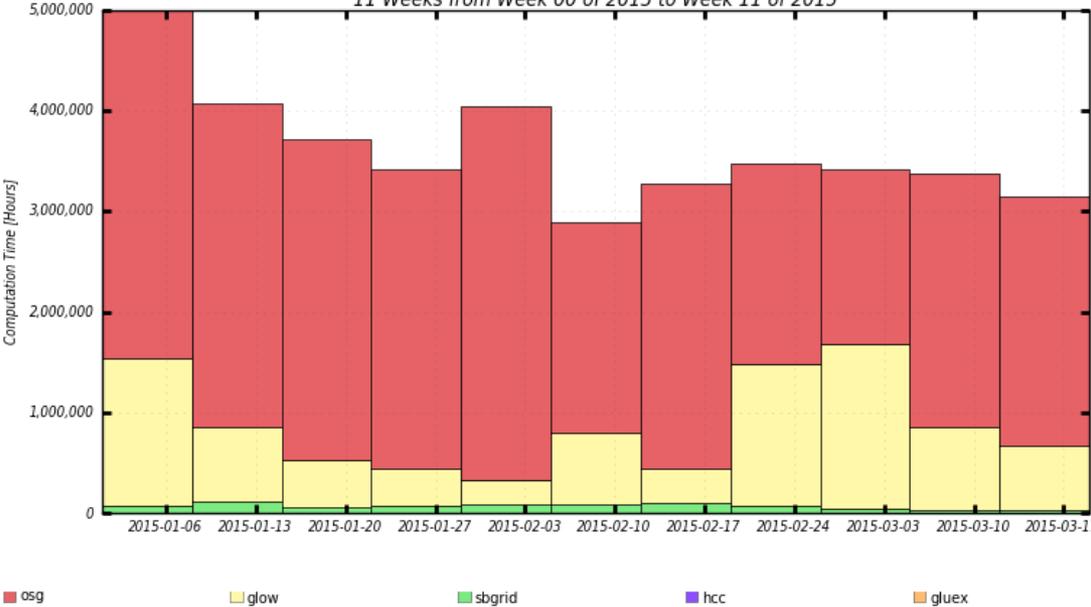
Open Science Grid

# Backup



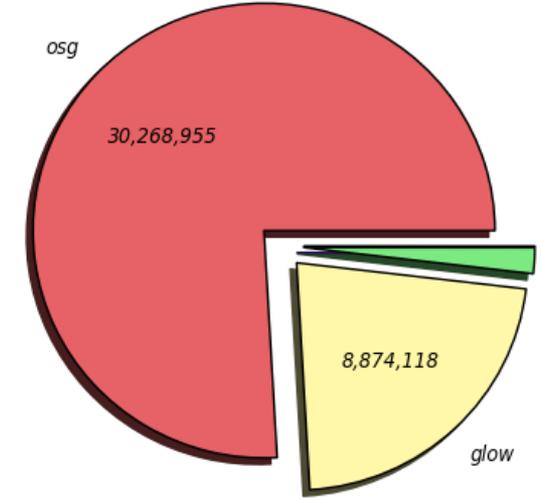
# Opportunistic in 2015 so far

Hours Spent on Jobs By VO  
11 Weeks from Week 00 of 2015 to Week 11 of 2015



Maximum: 4,998,485 Hours, Minimum: 2,894,461 Hours, Average: 3,625,110 Hours, Current: 3,153,405 Hours

Wall Hours by VO (Sum: 39,876,209 Hours)  
11 Weeks from Week 00 of 2015 to Week 11 of 2015



- Total of **39.9M** hours to-date in 2015 for primarily opportunistic VOs
  - ~24% of all OSG hours