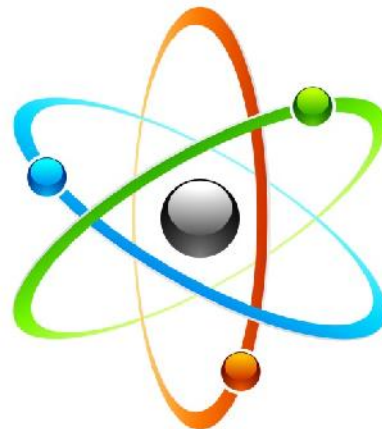


NUCLEAR PHYSICS THE NEXT 5 YEARS

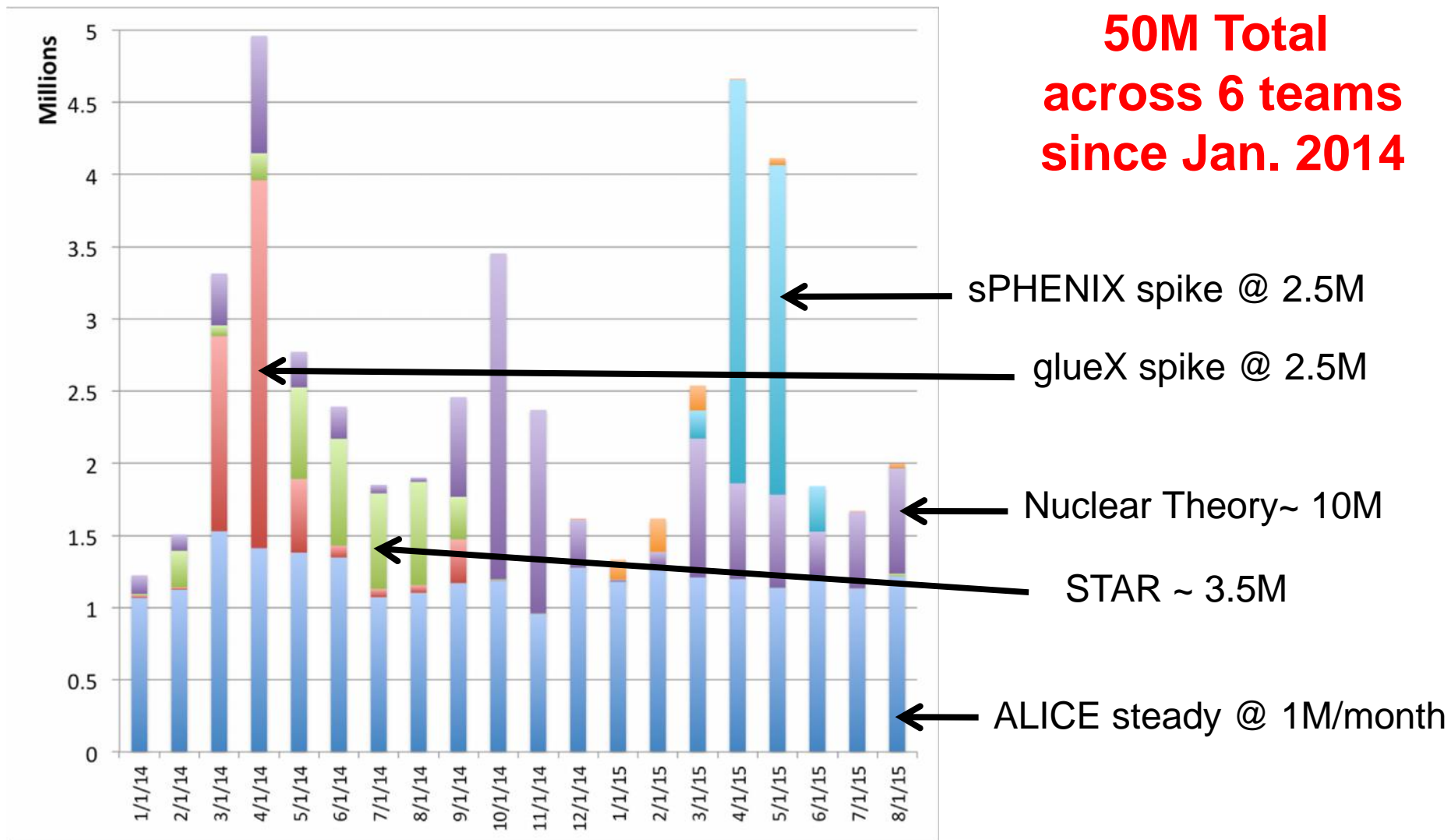
Disclaimer #1: not the whole NP community ☺

Usage and thoughts in the community ...



Motivations

(plot and starting point taken from Frank's talk at BNL on Sept 24th)




Motivations

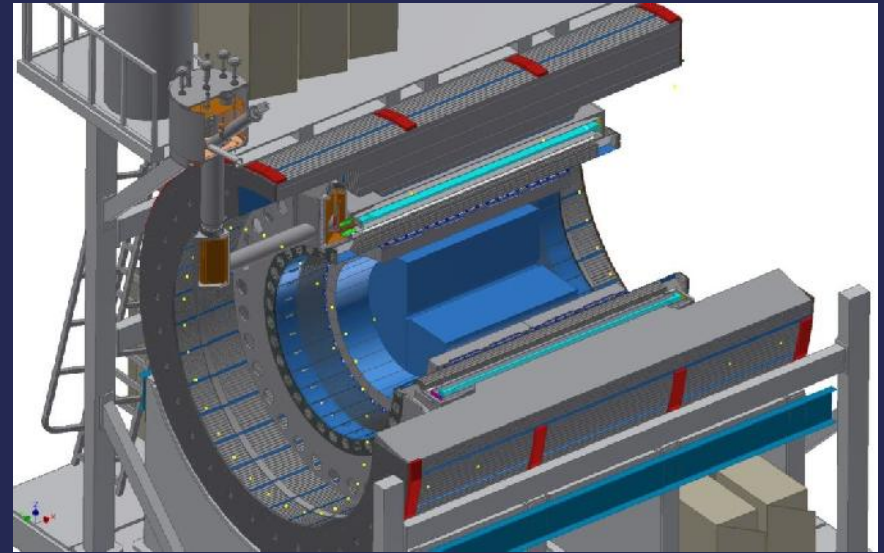
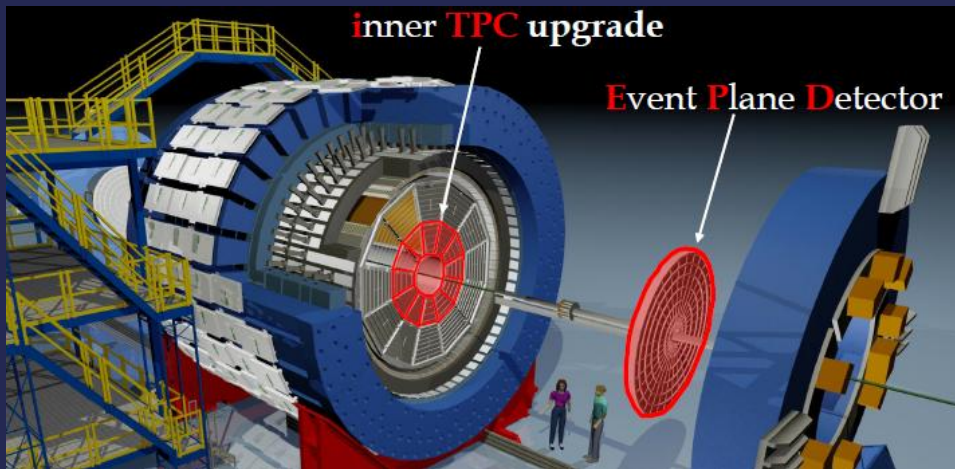
- Recent activities on the OSG showed a burst of NP activities
 - 10 M hours from Nuclear theory – Steffen Bass (Duke)
 - 6 M hours from detector R&D (sPHENIX) – Martin Purschke (BNL)
 - Continuous activities for months from STAR @ 3.5 M hours for ½ of 2015
 - GlueX at 2.5 M hours
 - ALICE at 1 M hours / month
 - EIC simulations at ~ 150 k hours
- 50 M hours and growing ... Is there a trend?
 - OSG provides ~ 92 Million core hours / month – for now, usage from NP is “easy” to absorb but ... **pattern of having individuals, not VO, a possible concern**
 - **A few 10 Million hours every month “may” require attention if time constrained**

Idea was to gather feedback on possible usage and demands for the next 5 years - **NP community interest may generate a NP office positive response (science case is strong)**

The players

(who was contacted so far)

- STAR 
- sPHENIX
 - Martin Purschke
- EIC
 - [BNL](#): [Thomas Ullrich](#) & [Elke-Caroline Aschenauer](#)
 - [JLab](#): [Rolf Ent](#), Markus Diefenthaler, [Amber Boehnlein](#), Graham Heyes
- Nuclear Theory (HI)
 - [Scott Pratt](#) (MSU)
 - Rel HI collisions, QCD, Hydro Models and Data Analysis Initiative (MADAI, NSF initiative) + RHIC
 - [Steffen Bass](#) (Duke)
 - RHI, QCD, Hydro
 - [Derek Teaney](#) (SUNY-SB)
 - QGP at RHIC, Hydro
 - [Raju Venugopalan](#) (BNL)
 - QCD/CGC, QGP at RHIC & LHC
 - [Berndt Mueller](#) (BNL)
 - ...



STAR & SPHENIX

Timelines

(as I know it, depends on budget outcomes and BUR)

2014--2016	2017	2019--2020		2021--2022	2023--2024	2025--2027	2027--	Year
Heavy-Flavor	W+-	BES II		Jets/DY/		eRHIC	eRHIC	Collider
HFT/MTD/FMS		iTPC/EPD		Forward-ion (west) upgrade	Forward-e construction	1) re-use STAR mid-rapidity	Fully constructed EIC2	STAR
		CDR		EIC2 construction (mid-rapidity)*		2) EIC2		EIC
VTX/FVTX/MP C-ex		sPHENIX construction		sPHENIX operation	fsPHENIX	EIC1 (*)	upgrade	PHENIX

- STAR running up to 2020 (BES-II) with a downtime in 2018
- PHENIX sPHENIX constructions to run in 2021 & STAR FWD upgrades phased in sPHENIX & STAR+ run until 2022
- Transition to eRHIC afterward – no runs in 2023-2024?
- EIC on the horizon... when? where? [JLAB | BNL]

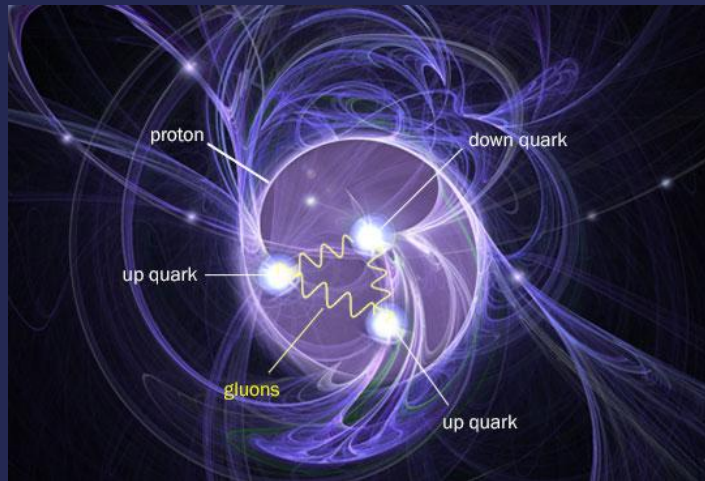
For the next 5 years, known running scenario at RHIC
 Beyond, strong science case but some unknown (transitional period)
Difficult to predict needs beyond 5 years

STAR

- Challenging data sets, computing resources exceeds initial projections
 - **Exciting new science opportunities appeared** ... but larger datasets not helping procurement cycle. Storage will be taken care by BNL/Tier0 center for RHIC
 - Computational resources a challenge – **data production cycles spanning over 2-3 years with (2015) 15k CPU at 95%+ efficiency, trend continues to 2019**
- Model shift
 - Simulation production on Grid or Cloud remains marginal (those workflows moved “there” ages ago)
 - Remote facility software stack tested (from BNL) using Grid
 - Opportunistic User analysis (local resources) to make room for the core data production (sparse unused resources from other experiments)
 - Moved to data production on Grid on dedicated facilities (KISTI/Korea ~ 10% boost) – proof of principle run smoothly (99% efficiency achieved) ... for real!
- Needs
 - STAR will come with additional resources from Tier1 centers – resources will be federated as much as possible (ex: resources being tested in Russia would not). ~ 0.3 FTE from within for general operation and production on Grid support (STAR need mostly + testing new services).
 - **Sparse resource aggregation will happen and join the OSG (Online compute resources, Tier-X, ...)** – some may not be opened to the community (resources at the experiment ⇔ security enclave)
 - **Possible interest for smaller usage (simulation, R&D) in OSG-Connect or Cloud resources**
 - **Expecting to leverage OASIS services to (a) make our middleware stack broadly available at remote sites (b) help leverage opportunistic cycles if possible**
 - HTC mostly within the next 5 years – spec planned at BNL (HS06/100)
 - 2015: 35148, 2016: 46908, 2017: 62808, 2018+: 70010
 - **3.5 Million hours in 2015, assume scaling with Spec for a 10% recovery**
 - **Best case – recover as much as possible to fallback onto 1 year production cycles (20, 25, 30, 35 M h)**

sPHENIX

- Science case endorsed through Department of Energy review
- First constitutional collaboration meeting December 10-12, 2015 at Rutgers University, New Jersey, USA
- OSG usage reported in “[*Using Open Science Grid to prepare for ‘the next big thing’ at Brookhaven*](#)”
- Detector simulation was run on the OSG
 - Helped the sPHENIX detector design – 5 Trillion collision events simulated, 99% efficiency reported, **2.5 M hours**
 - **Offloading computational resources for R&D** while BNL CPUs are busy with PHENIX data production
- Needs & findings
 - Leveraged CVMFS, DB access done through a DB on the BNL/Science DMZ
 - **Finding 32/64 bits consistency (many 64 bits nodes not able to run 32 bits execs) – matchmaking?** Now using pure 64 bits but memory footprint changes (3GB)
 - Transfer of the output back was not easy nor optimal – used Condor mechanism at the end. **Need more transparent /efficient mechanism. [TBD]**
 - **Jobs often requires > 4 GB of memory (not common on OSG) – more “classads”**
- Future use
 - The big pass is already done. Ease of use of OSG resources brought attention from within
 - **sPHENIX considering migrating their simulation workflows and needs to the Grid (OSG). Usage level at ~ 0.5 M CPU hours / year.**



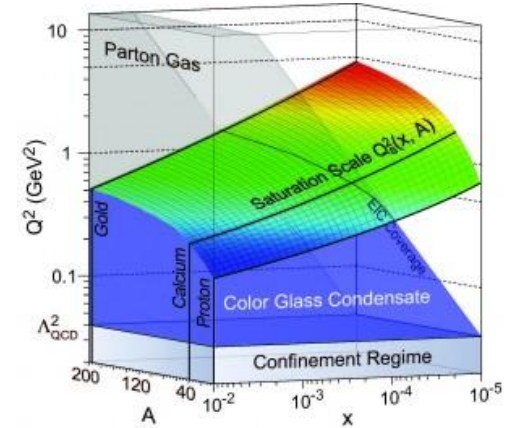
EIC

Note: one community, two sites working on creating a future frontier experiment for NP – **EIC user group (just) forming**

Disclaimer: no plan written-in-stone yet (current activities & thoughts) – community is thinking and serious design will happen by 2020-2025

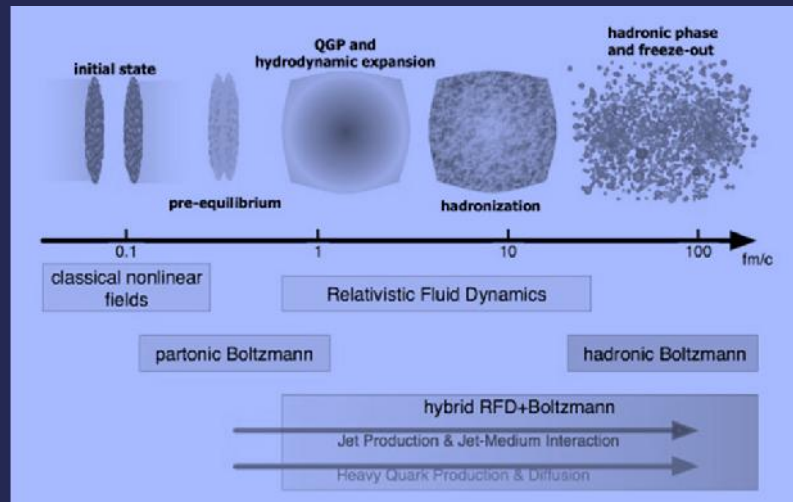
EIC – BNL activities

- Detector design simulation will not require more resources (local resources sufficient)
- Modeling simulation use – “[*Using the OSG to plan for an electron-ion collider at Brookhaven*](#)”
 - QCD Phase space study (The saturation scale as a function of momentum fraction x , resolution Q^2 and nuclear mass number A – “walk the phase space” and pre-calculate values)
 - Physics case study for the EIC – [SARTRE](#) simulator
 - OSG allowed reasonable time-scales and fast turn-around as resources are intensive (but spotty) – 50k hours x 3 calculations so far (150 k hours with Liang & Tobias)
 - OSG allowed ease of access with very limited knowledge from end-user – **opened access / ease of access a “+” for R&D work**
- Projection: **at least another 150 k hours in the next year, possibly up to ½ M hours within 2 years (simulations)**
- Note that 10 years of EIC may be \Leftrightarrow 1 year of today’s LHC data (so, low all considering). Computing model will need to take this into account.
- Needs now
 - Workflows can be converted to embarrassingly // (HTC) well suited to run on the OSG
 - Possible interests in multi-core / multi-threaded approach – would open to a thorough / fine grain phase space calculation



JLAB comments (EIC + beyond EIC)

- Very hard to plan / predict computing model for 2025-2030 (EIC)
- Assumption: Following the “12 GeV model” – Use multi-threaded software on a local farm
 - Largely shared resources with LQCD efforts
 - Most EIC work is on the accelerator side, OSG not optimized for this work
- Note: JLAB Activities on OSG (beyond EIC) and model assuming Grid (Hall D) and Cloud (Hall B)
 - Use of OSG from GlueX (modest all considering). CLAS12 and GLUEX each of which will generate 200-300 MByte/s of raw data likely candidates (*will revisit*) + SOLID in a few years.
 - Cost of maintaining a local infrastructure versus networking cost investigated (privilege local resources) – OSG a model for the future? ⇔ Rapidly changing landscape and cost ...
 - Hybrid model in the thinking (reduced data moved on OSG only) ...
- Implication/Needs: multi-core / multi-threaded application is a requirement (and plays an important role) in the software design
 - Interest in supplementing the existing computing infrastructure with shared resources on a super computer
 - (Markus) opened to investigate further in the distributed computing direction



NUCLEAR THEORY (HI)

A cross-section of the HI Nuclear Theory community

Nuclear Theory (HI)

- Major prospects and interest
 - Steffen's (student) run should be seen as a “*proof of principles*” was the tip of an ice-cube ☺ (hopefully not an iceberg if panned well) & more coming ...
- Typical work mode
 - Run a wave of calculations, assess (tweak analysis, understand results – may take time), run another wave, ... *deeper and deeper knowledge*. Exact needs not known (synergistic relation with experiment, + iterative work) but a best guess attempt possible
 - Community does not typically have dedicated resources – the OSG appears like an *attractive match* for the way the community works.
 - Low threshold and little administrative to get running is a bonus (and an advantage over capacity computing facilities)
 - On-request needs rather than allocation per year is a desired *modus-operandi*
- Community behind - Steffen, Scott, Derek, Berndt, ...
- So what is being run?
 - “Nuclear physics and computer science meet on the OSG” – study of the formation and properties of the QGP
 - No use of multi-core for now, as workflow are trivially parallelizable
 - 2D Hydrodynamic calculations is the major contributor – going to 3D Hydro => x10 the CPU need
 - Systematic and rigorous science investigations (i.e. systematic studies in the BES era of RHIC)

Nuclear Theory (HI)

- Task
 - ~20 beam energies / target / projectile combinations
 - ~10K hydro runs for each combination to cover the range of impact parameters - ~10 hours to run one 3D hydro code + Repeat ~1000 times for each point in parameter space
 - Hydro part requires 2 M hours for a single analysis (skipping the Hydro after-burner calculation, a 10% effect)
 - Full statistical analysis => repeat 1,000 to investigate different model parameters
- ~ B hours *scale* potential, targeting for now 2D Hydro, 100s M hours usage on the OSG
 - Note: **Storage IS an issue – no funded large storage facilities. 200 M events ⇔ 1 PBytes**
 - Being used/investigated: save the after-burner result only temporarily? Throw away events? Archival?
- Next 5 years?
 - Assume **2-3 groups will approach the OSG** (4 max, not 10) as the *community is organizing itself*. The current level of services is judged appropriate (to excellent)
 - **Ramp up from 10 M hours to the 100 M level / year within the next 5 years**
 - 3D viscous hydro codes may require multi-core (also investigating GPU usage) and will require more memory (~ 4 GB)
 - Beyond 5 years, 3D Hydro, 1 B CPU hours / year regime and a few PB of data per year. Is the OSG a realistic platform? Leadership facilities? Hard work to optimize code ongoing ...

Willingness to discuss further as the OSG may provide a good “framework” for the NP/HI-Theory work and initiatives. Willing to also discuss sharing resources – initiatives being pursued (“DOE Topical Collaboration”).

Summary of needs

Team / experiment	CPU hours / year	Notes
STAR	Objective: 2016: 4.5 M, 2017: 6 M, 2018: 7 M Max: 20, 25, 30, 35 M	OSG-Connect for individual users? OASIS Service usage for software distribution 0.3 FTE in-house to help users / production / infrastructure + test new services
sPHENIX	Simulation usage TBC – estimated 0.5 M	Big try at 2.5 M hours done File transfer optimization, fine grain match making (64/32 bits), Mem > 4 GB (more classads)
EIC	0.15 M in 2016 foreseen, possible peak at 0.5 M	Model not crystalized, user community forming (too early). A few interim activities. May need multi-core support
Jlab / non-EIC (?)	(TBC)	Indicated a desire to brainstorm further on use of distributed computing (beyond GlueX)
Nuclear Theory / HI	Ramp-up from 10 M to 100 M within 5 years Later huge demand for full 3 D hydro – B hours level	Major interest – what you saw was only a proof of principle (a precursor and sign of more coming) Collaboration organizing – 2-3 people running (4 max, not 10+) – OSG low entry threshold attractive. Model simulation requires intense computing & possibly storage (not resolved to date) Willingness to discuss further, share what they have and make the program's vision a success