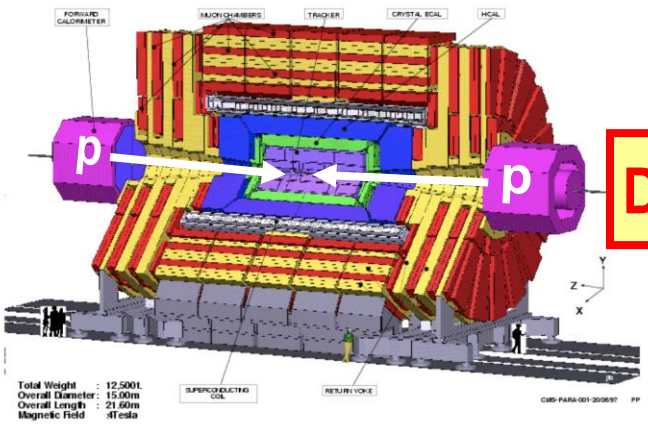
Computing in HEP and Its Future

KAIST-KAIX Workshop for Future Particle Accelerators
July 8 - 19, 2019

Jaehoon Yu
University of Texas at Arlington

What does Computing do in HEP?

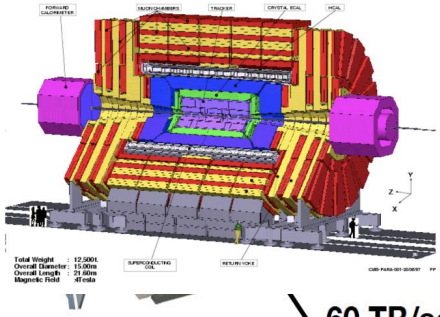
- High Level Event Triggering
- Data Recording
- Data Reconstruction and Processing
  - PID and pattern recognition
- Data Storage and Access
  - Selection and streaming
- Data Analysis
- Simulations



Digital data

Data Reconstruction

Chain of HEP Data



60 TB/sec, 40M “events”/sec

Data

GENERATION

SIMULATION

DIGITIZATION

RECONSTRUCTION

DERIVATION

Monte Carlo

TRIGGER

1.5 GB/sec, 1000 “events”/sec

RECONSTRUCTION

DERIVATION

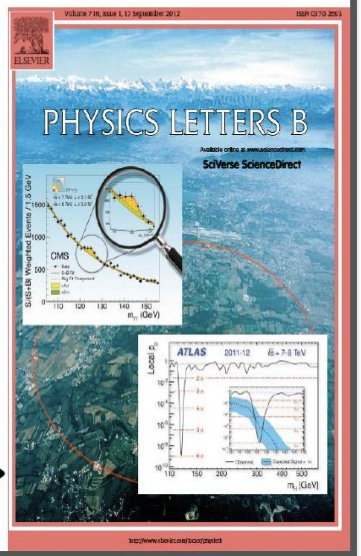
ANALYSIS

Online

Tier-0

Grid

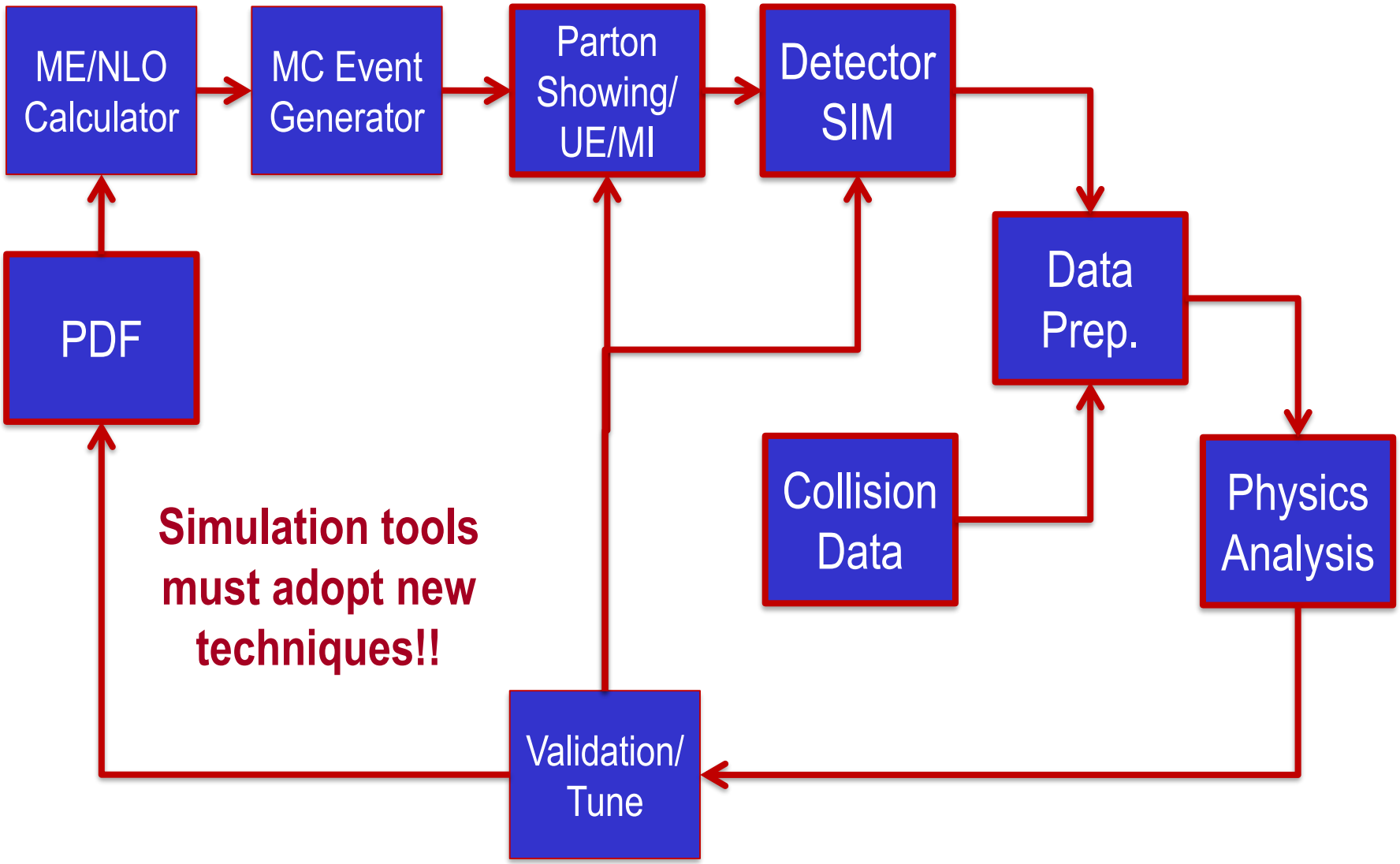
Local



Factors for Computing Needs

- Computing needs: CPU, Network bandwidths, Storage space
- Data size ➔ Network bandwidths, storage space, CPU
  - Accelerator capabilities ➔ CMS energy and luminosity
  - Detector capabilities ➔ number of channels, data zero suppression, trigger reduction factor, etc ➔ Event size
  - Data output rates
    - Manner of triggering (triggered or continuous readout)
- Data reconstruction and analysis software ➔ Storage and CPU
- Offline streaming and data reduction ➔ Storage
- Data analysis activities ➔ Network bandwidths, storage, CPU
- Amount of simulated data ➔ Storage, CPU
  - Detector design & performance studies
  - Background studies
  - Signal simulation and phase space scanning

Simulation Tool Iteration Process



The Problem

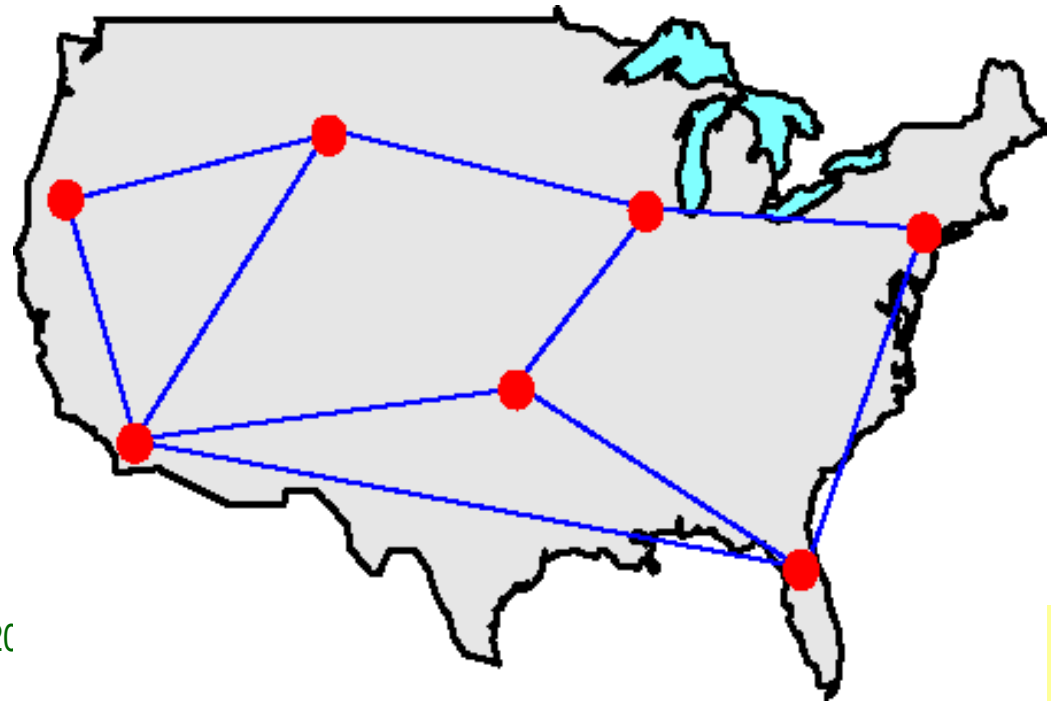
- Detectors are complicated and large ➔ Need large number of collaborators
  - They are scattered all over the world!
  - How do we get them communicate quickly and efficiently?
  - How do we leverage collaborators' capabilities?
  - How do we get all the compute resources?
- Data size is large, expected to be EB's
  - Where and how to store the large amount of data?
  - How do we allow collaborators scattered all over the world to access data in an efficient fashion?

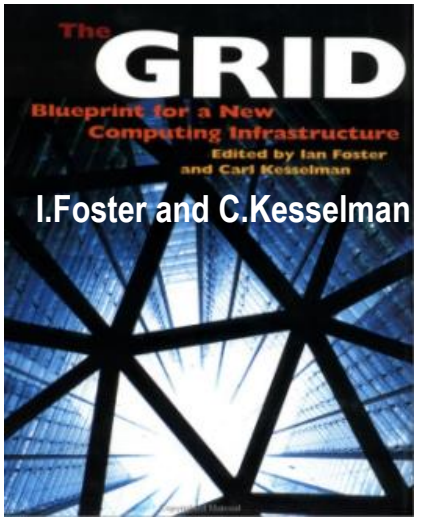
The Problem, cont'd

- How do we allow people's analysis jobs to access data and make progress rapidly and securely?
  - What is the most efficient way to get jobs' requirements matched with resources?
  - Should jobs go to data or data go to jobs?
  - What level of security should there be?
- How do we allow experiments to reconstruct data and generate the large amount of simulated events quickly?
  - How do we garner the necessary compute and storage resources?
  - What network capabilities do we need in the world?
- How do we get people to analyze at their desktops?

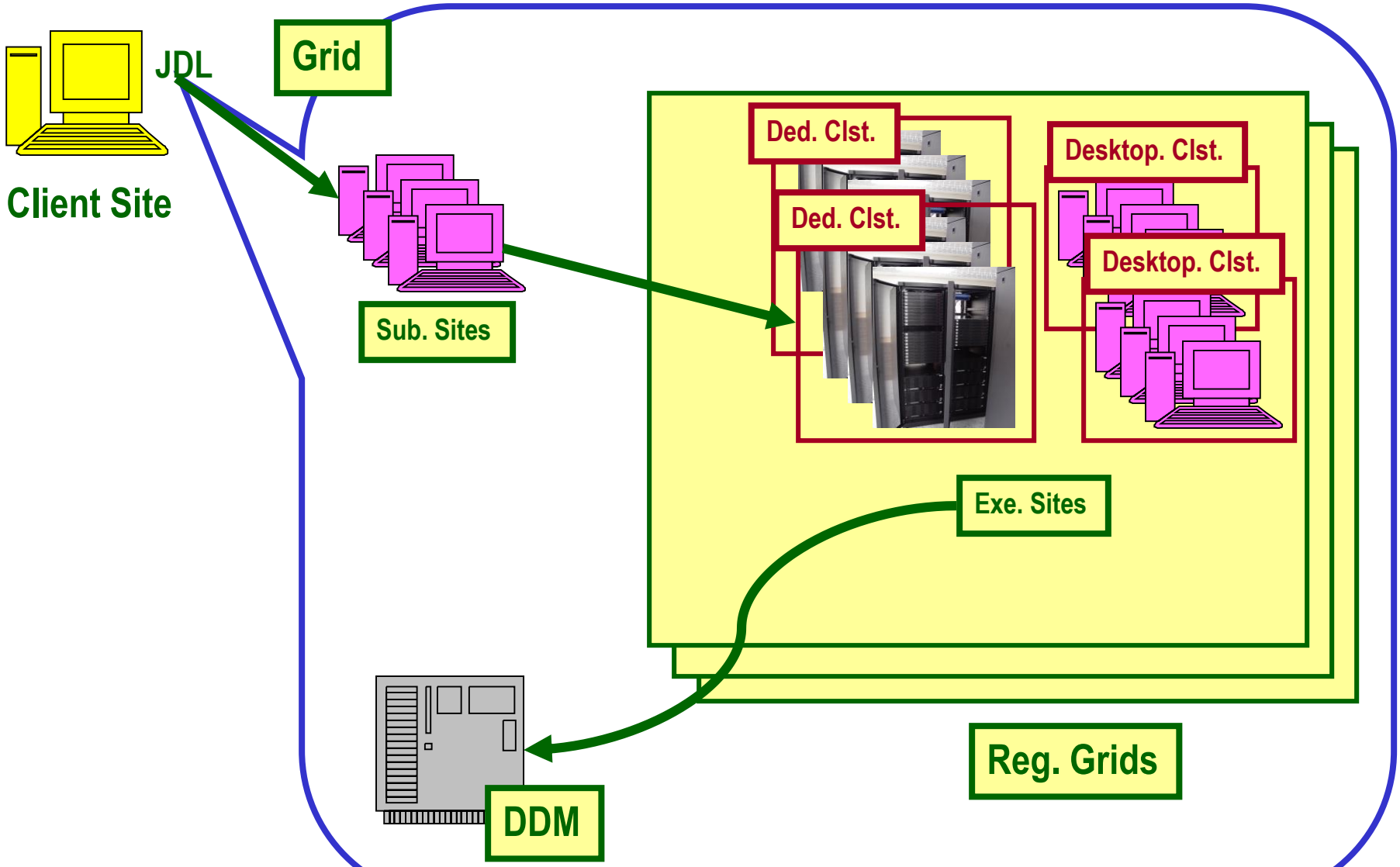
What is the Computing Grid?

- Grid: Geographically distributed computing resources configured for coordinated use
- Physical resources & good network provide hardware capability
- The "Middleware" software ties it together ➔ data distribution, job managements, security, etc
- HEP drove the initial development and implementations

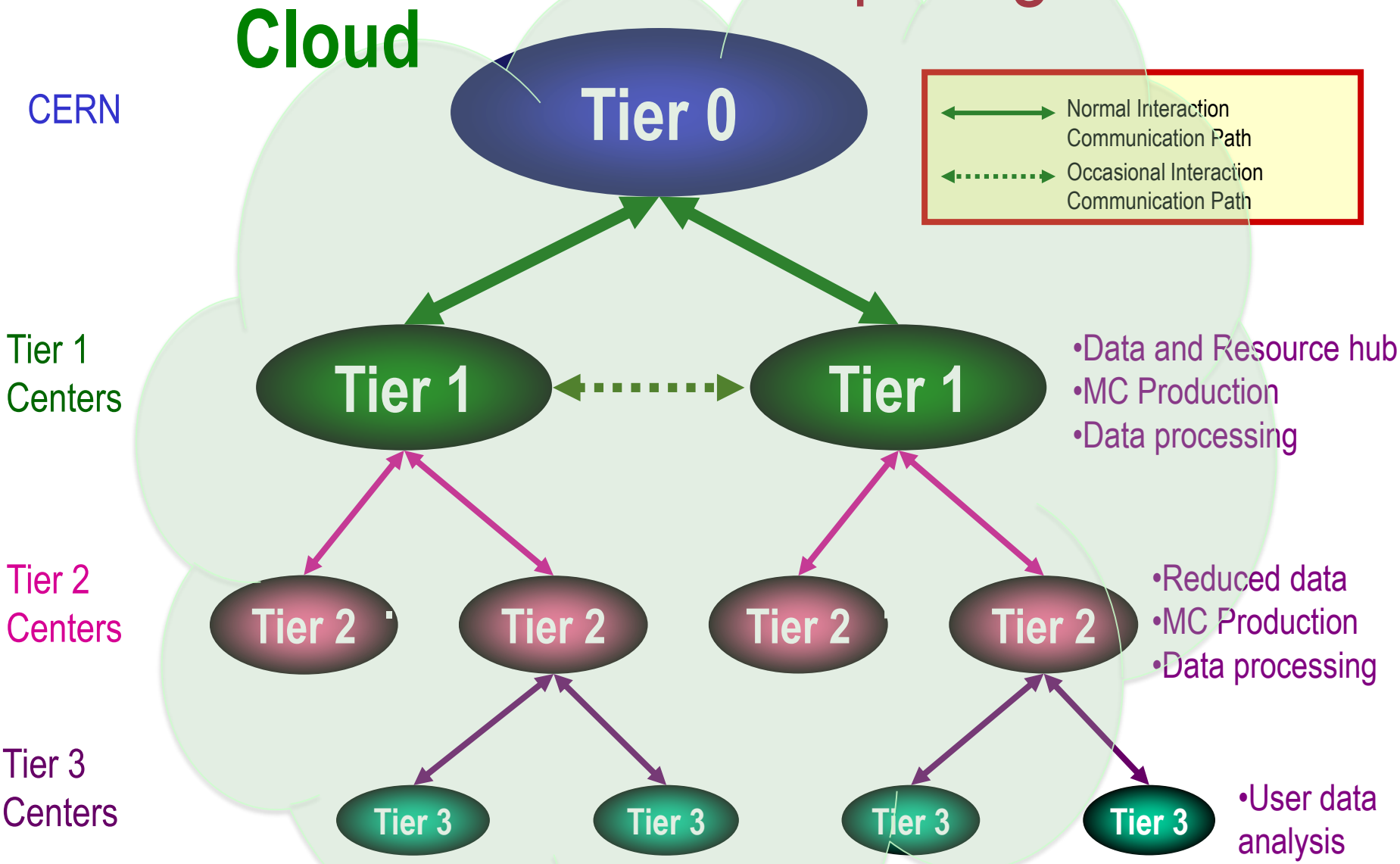




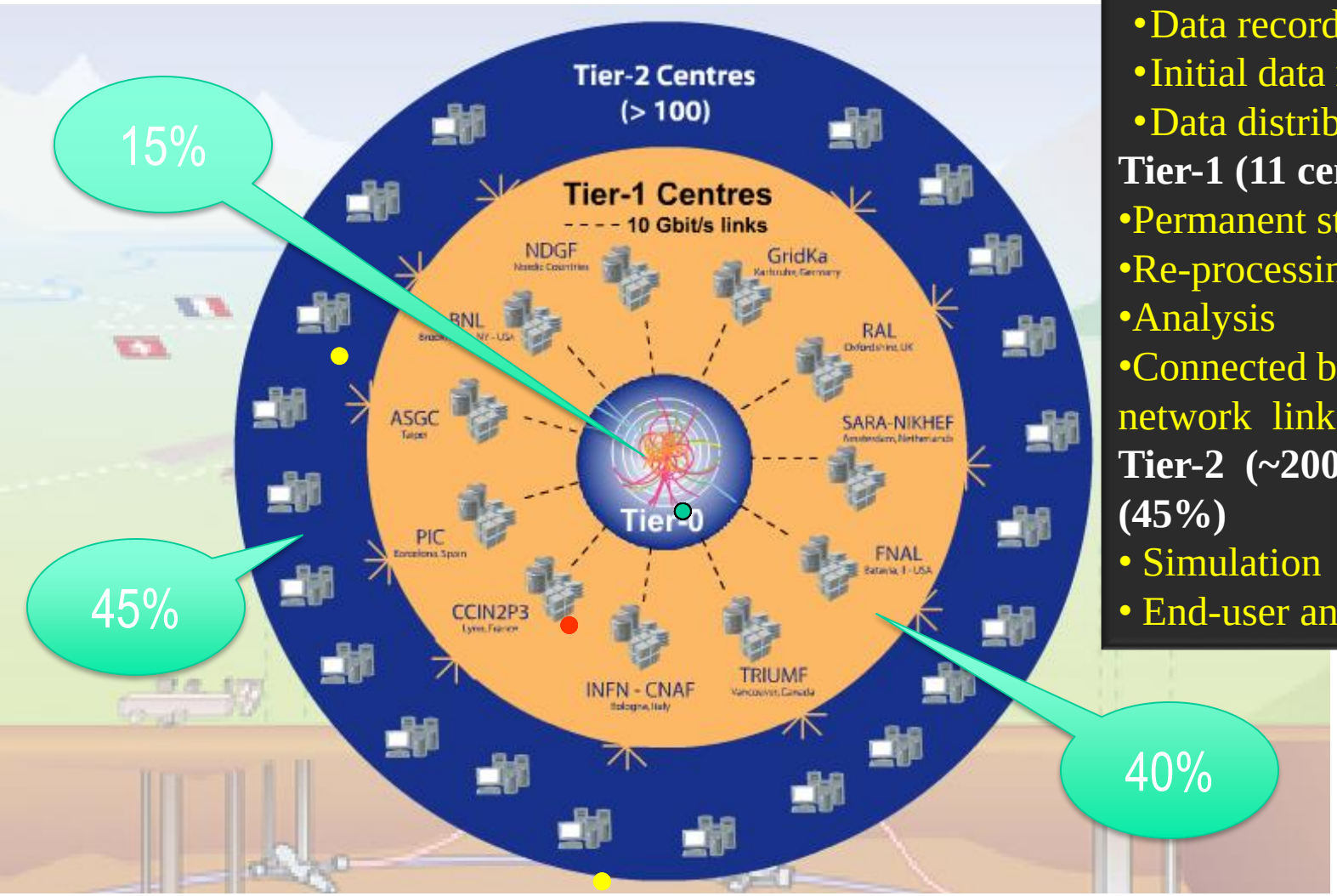
How does a computing Grid work?



Initial Idea of HEP Computing Model



Implemented LHC Grid Structure



Tier-0 (CERN): (15%)
- Data recording
- Initial data reconstruction
- Data distribution

Tier-1 (11 centres): (40%)
- Permanent storage
- Re-processing
- Analysis
- Connected by direct 10 Gb/s network links

Tier-2 (~200 centres): (45%)
- Simulation
- End-user analysis

NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

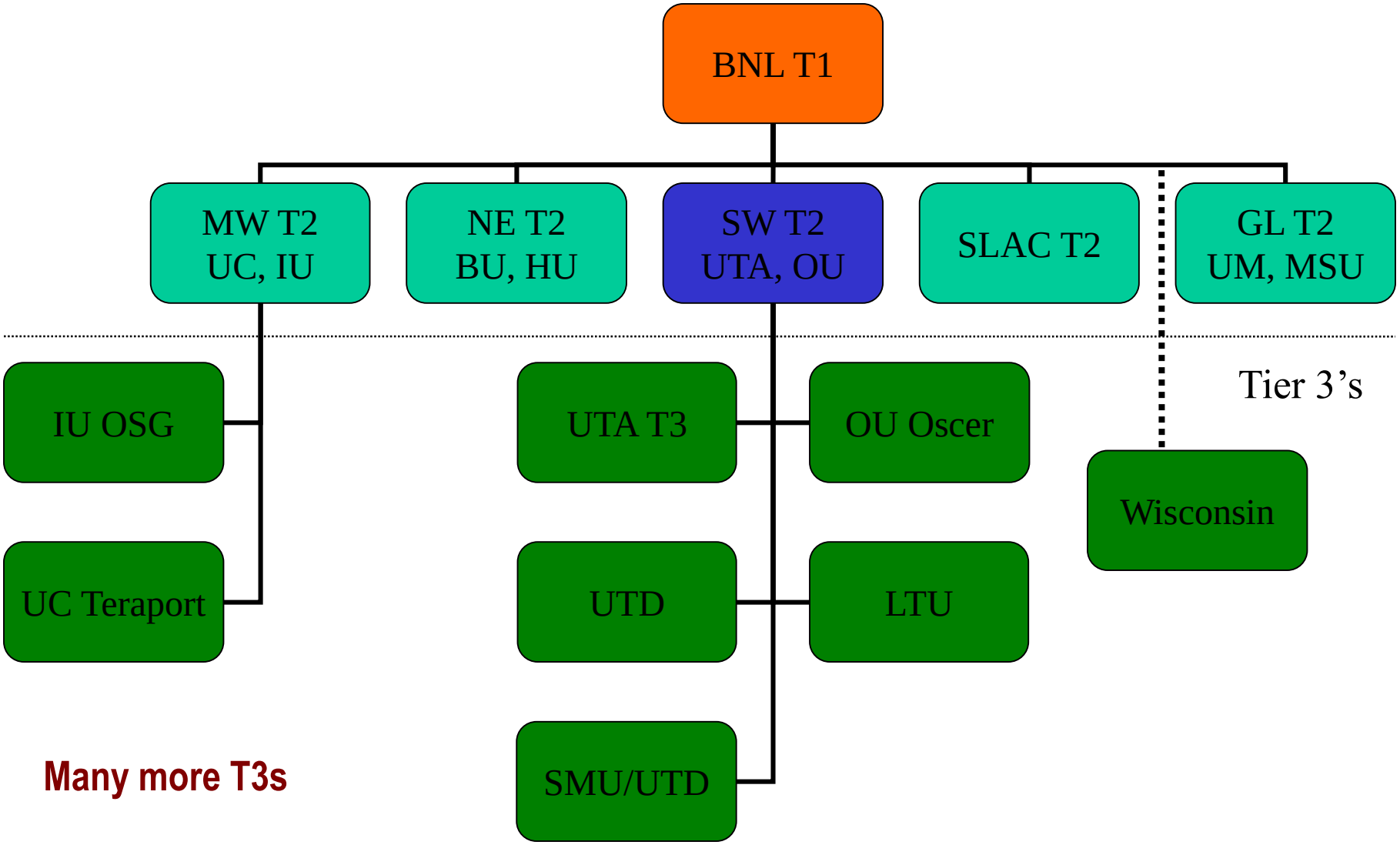
Open Science Grid

LCG

Worldwide L

egi

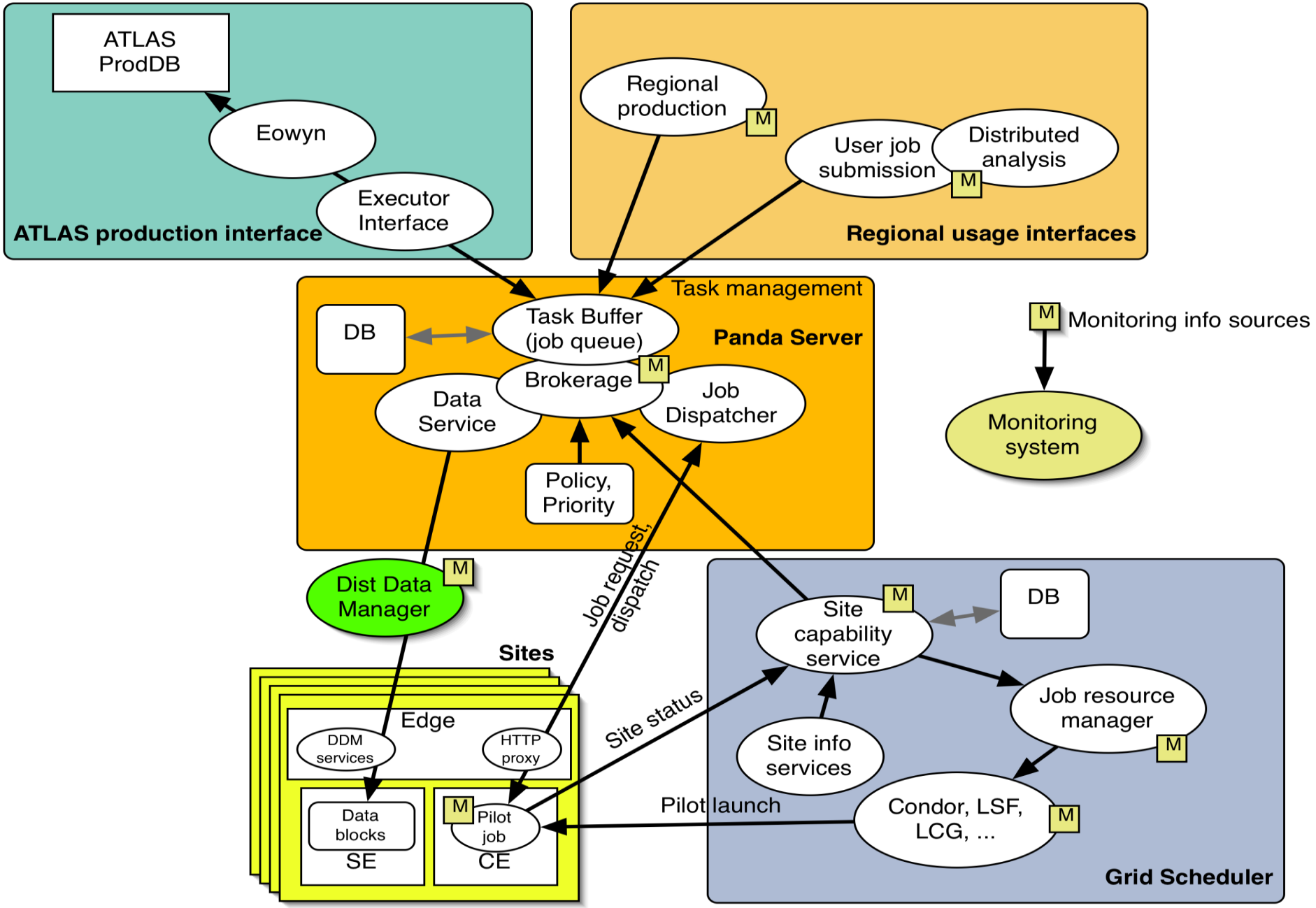
Tiered Example – US Cloud



ATLAS Production and Distributed Analysis System, Panda

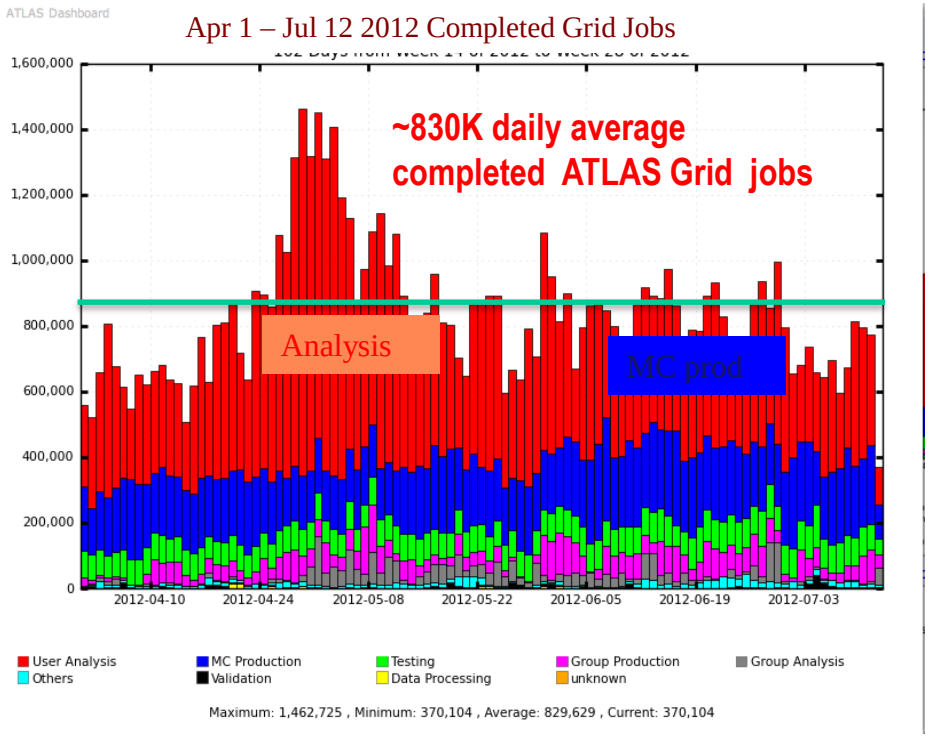
- Designed for analysis as well as production
- Work both with OSG and EGEE/LCG
- A single task queue and pilots
  - Apache-based Central Server
  - Pilots retrieve jobs from the server as soon as CPU is available low latency
- Highly automated, has an integrated monitoring system, and requires low operation manpower
- Integrated with ATLAS Distributed Data Management (DDM) system
- Not exclusively ATLAS and HEP but other disciplines use, too

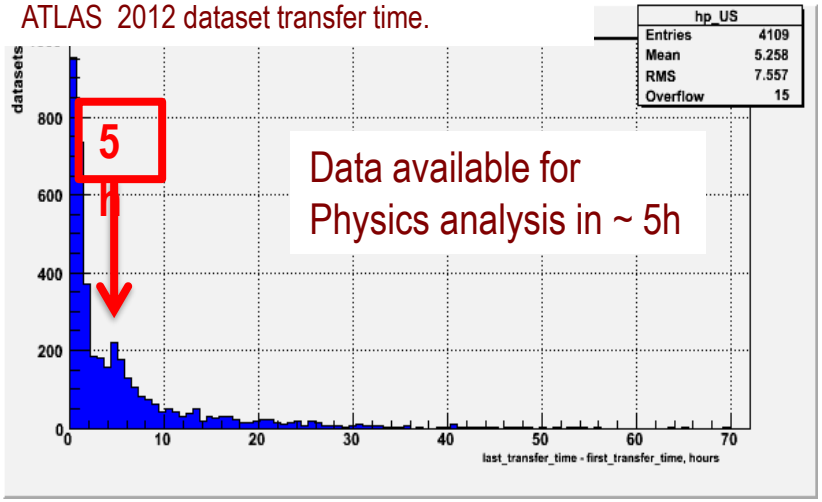
ATLAS Panda Architecture

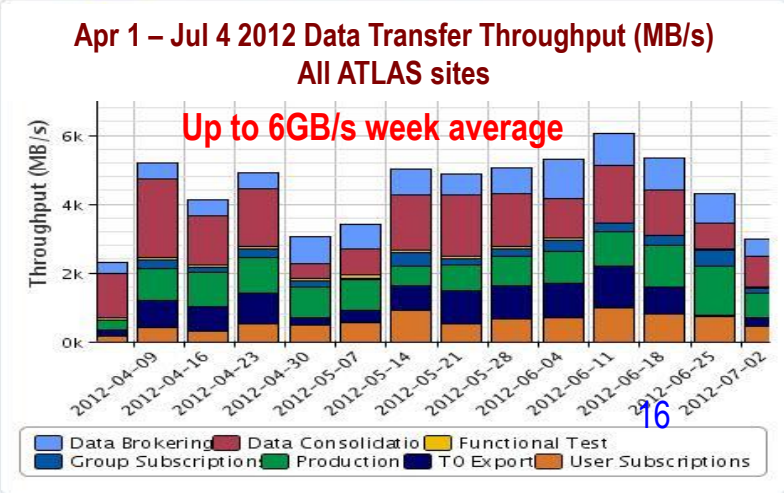


The Little Grid that could…

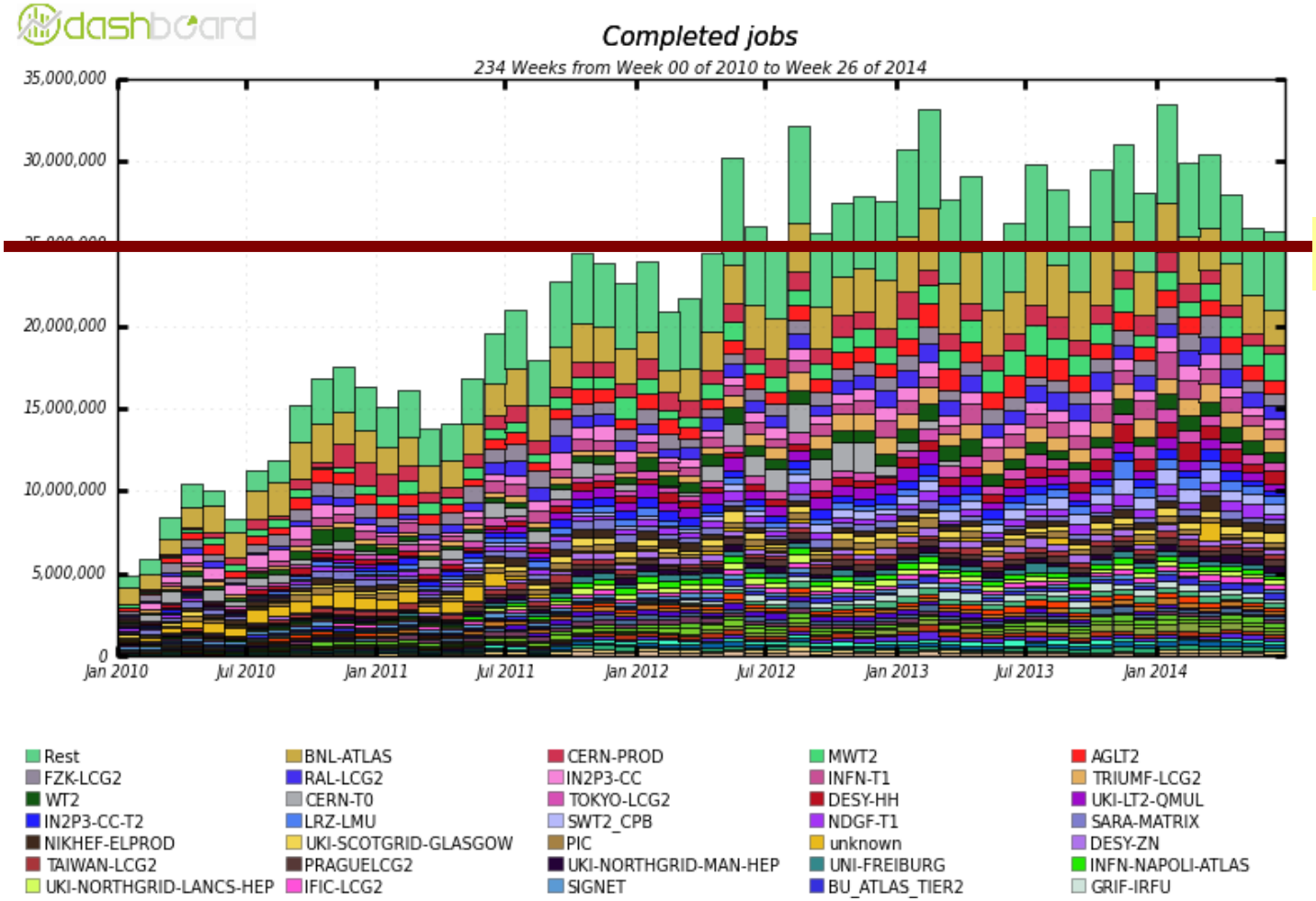
- ATLAS Distributed Computing on the Grid : 10 Tier-1s + CERN + ~70 Tier-2s +…(more than 80 Production sites)
- High volume, high throughput process through fast network!!







PanDA Performance



25M Jobs!!

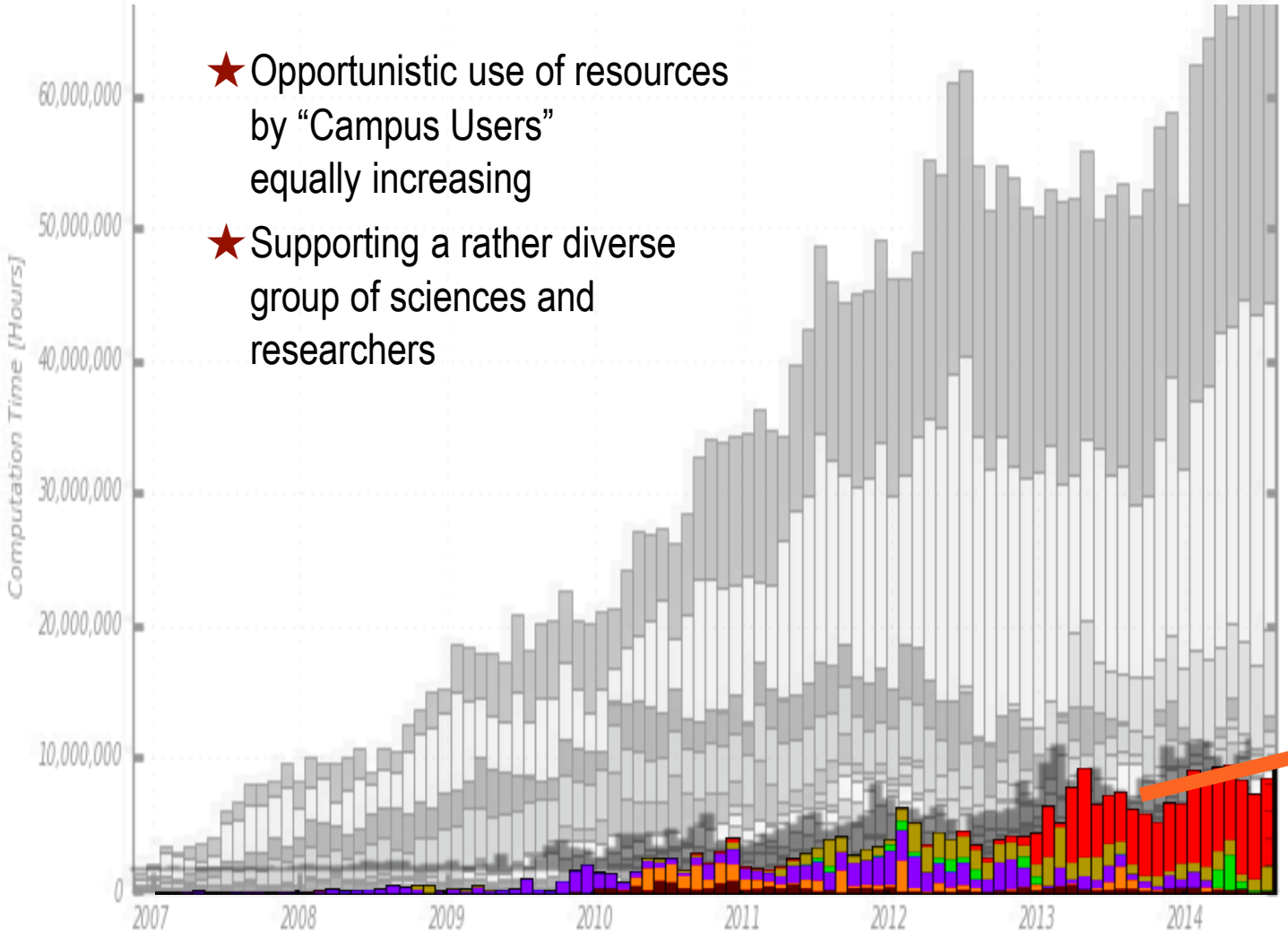
Current scale – 25M jobs completed every month at >hundred sites

First exascale system in HEP – 1.2 Exabytes processed already in 2013

Kaushik De

Growing Use of "Owned" and of "Opportunistic" Resources

- ★ Opportunistic use of resources by "Campus Users" equally increasing
- ★ Supporting a rather diverse group of sciences and researchers

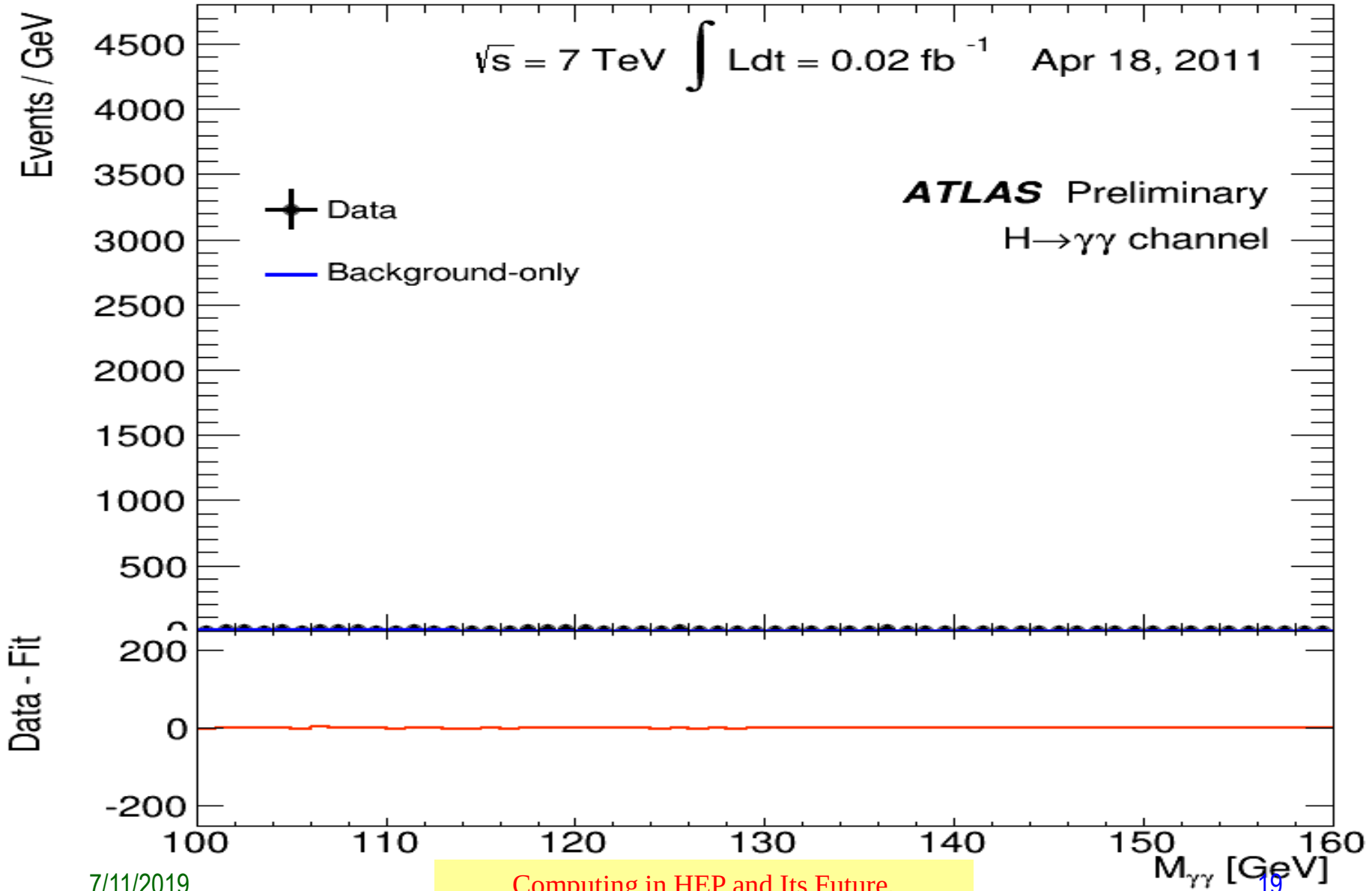


Opportunistic Use?

>90M CPU hours opportunistic use past 12 Months

Lotha

What did grid do for Higgs?



Now the industry picked up..

Early 90's

2004

facebook

1996

amazon

hadoop

1998

Google

2006

twitter

Amazon
EC2

Dropbox

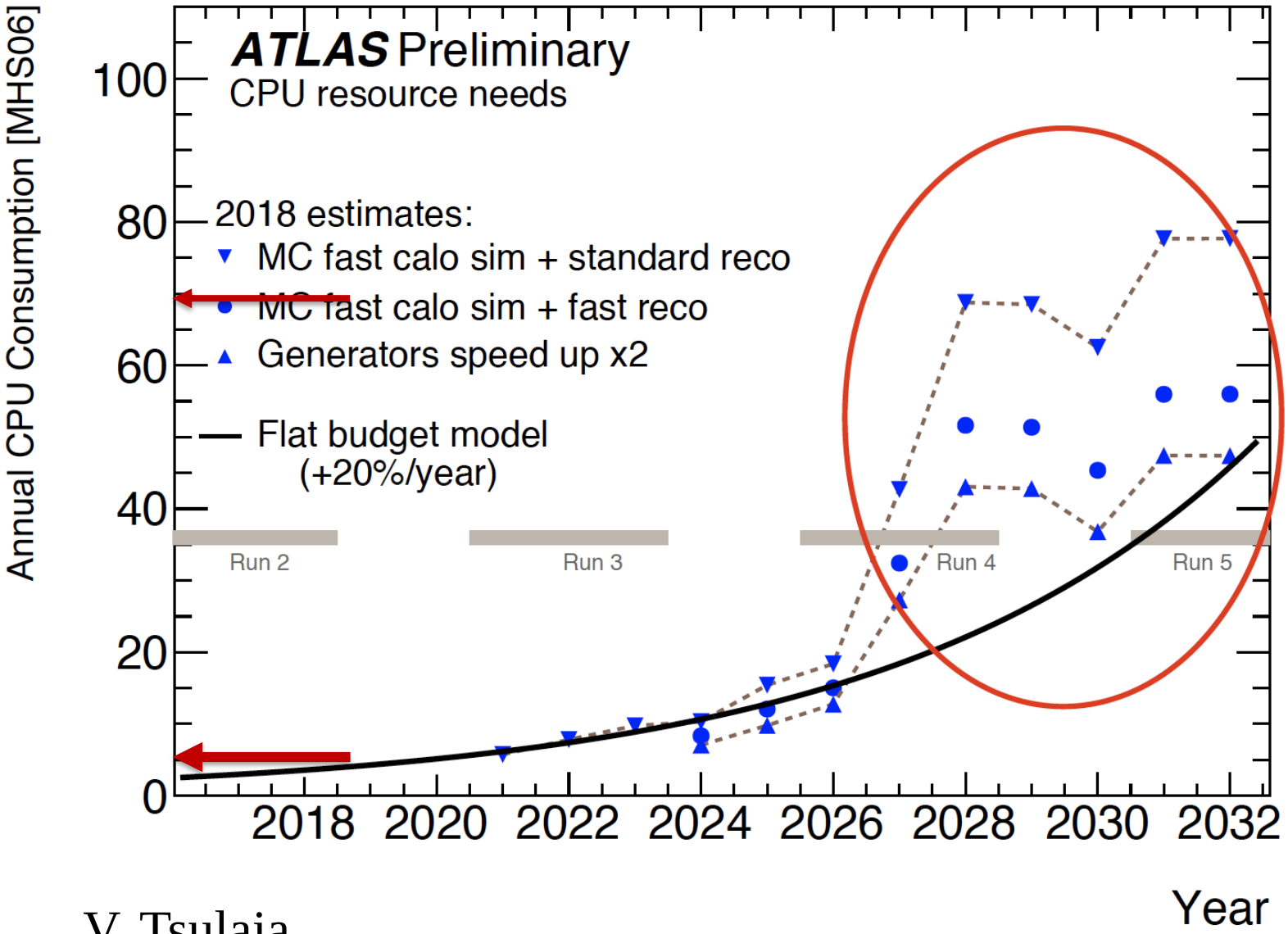
Many private entities fully utilized the internet communication we've developed to multi-trillion dollar venture!!

HEP working with industry to rent the commercial compute resources for data and simulation processing

What would be the next?

- Grid computing infrastructure has served well thus far
  - 1000's of users process 10s of EBs of data & 10^9 of jobs
- Upcoming experiments will further strain computing ➔ must be much more efficient and speedy

Expected ATLAS Computing Needs



V. Tsulaia

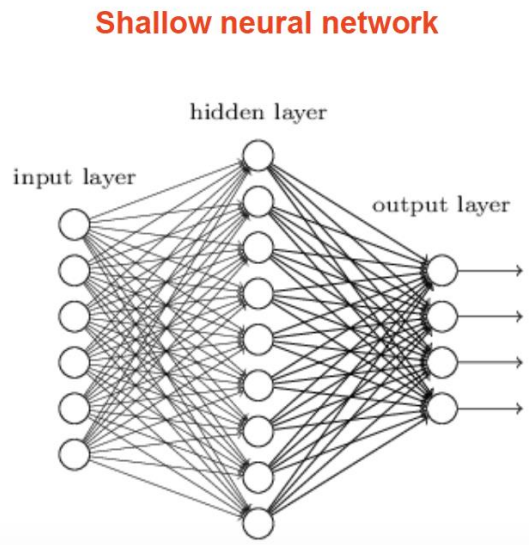
J. Yu

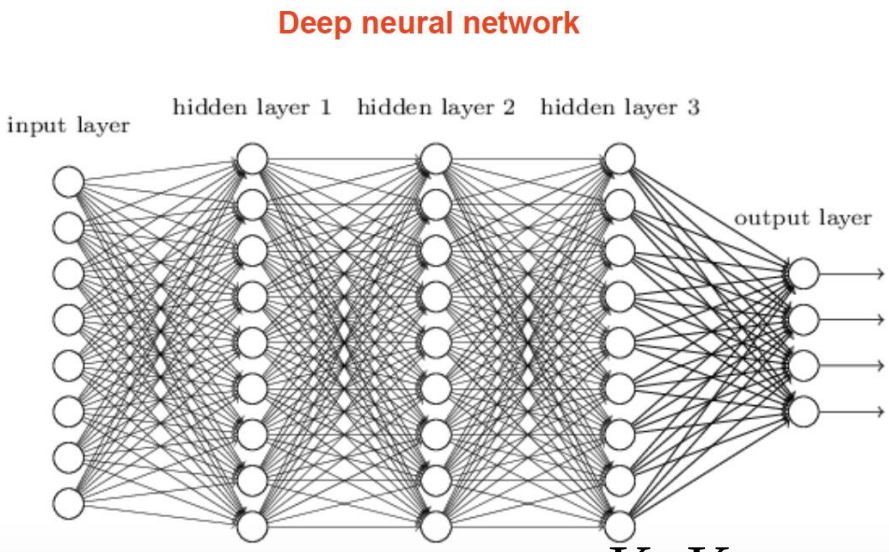
What would be the next?

- Grid computing infrastructure has served well thus far
  - 1000's of users process 10s of EBs of data & 10^9 of jobs
- Upcoming experiments will put further strain computing ➔ must be much more efficient and speedy
- Deep Learning or Machine Learning technology improving fast
- Unlike Computing Grid technology where HEP lead the initial concept and software infra development
- Industry actively leads the ML technology
  - Very different primary goals but the same idea
  - Train machines and let it make a decision by itself
  - Not just speediest but intelligent, accurate, effective & efficient
- Many current and future HEP experiments are actively adopting ML for PID algorithms and HL triggering

What is Deep Learning?

- Artificial Intelligence technology
  - Possible thanks to the dramatic improvements in computing hardware, e.g. GPU and the emergence of Big Data tools
- Enables machines mimicking complicated computations performed by a brain
  - Early 2000's neural network technology could not train big networks
  -





J. Yu

K. Kranmer


ANF
TUNA&SHRIMP
참치새우캔
Net wt. 95g

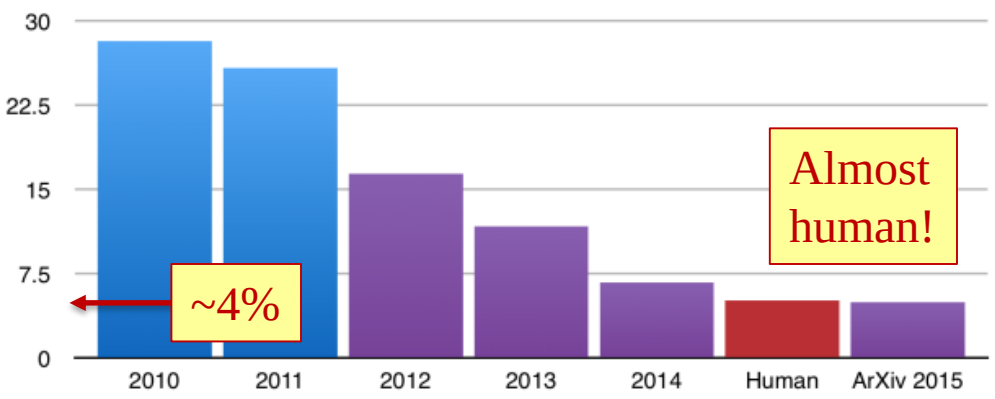
How does an animal brain work?

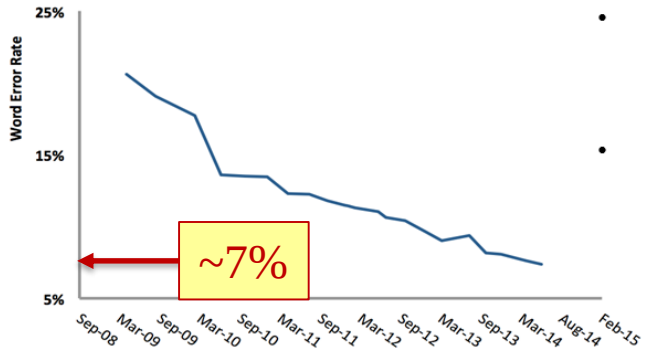
- Brain takes sensory data, build a hierarchical model of the world
  - Cells in visual cortex response specific low level features, such as color contrast or horizontal/vertical lines
  - Other cells combine low level features to identify higher level features such as geometrical shapes
  - Pattern repeats until the recognition of objects, like the chair
- A representation of the input is assembled in the brain
  - Eyes see limited window but scan around to establish the model of the surroundings ➔ These include maps of the environment registered in the cells which light up when we are placed in it
  - These specific cells light up when we imagine specific object or location
- When a decision is made we use these models to predict the outcome of an action

Artificial Neural Network

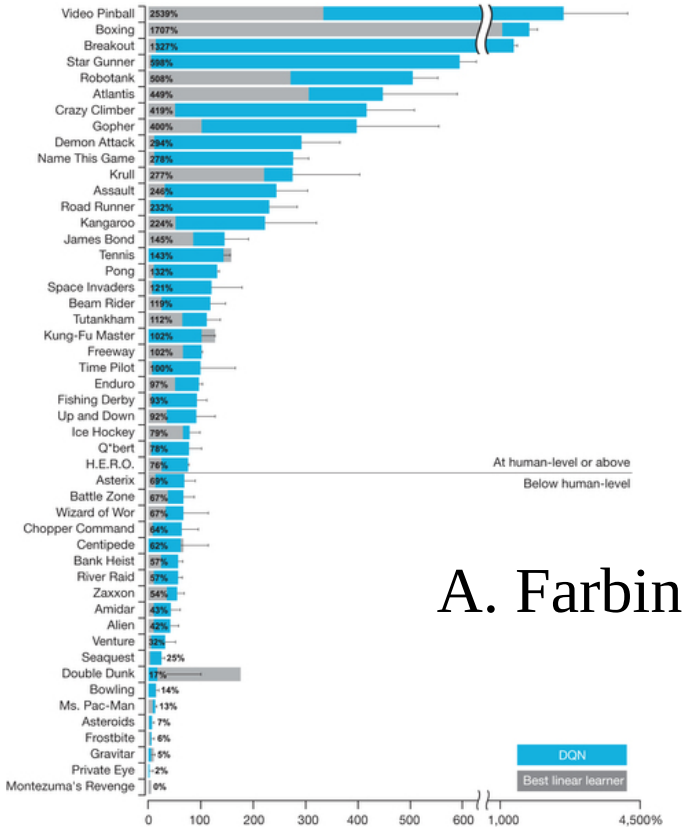
- Biology inspired computation (1st tried in 1943)
  - Probabilistic inference of signal vs background
  - Universal Computational Theorem (1989)
- Multi-layer (or Deep) Neural Networks (DNN)
  - Idea of 1960's impractical to training ➔ new techniques of layer-wide training
  - Feed in a big training data set and intense computations ➔ Utilization of Big Data and GPU
- Deep Learning Renaissance w/ 1st DNN in 2014
  - Amazing success in recognition, captioning and generation of audio/image/video
  - Text analysis, Language translation, video game playing agents
  - Picking up speed in industry

ML Performance

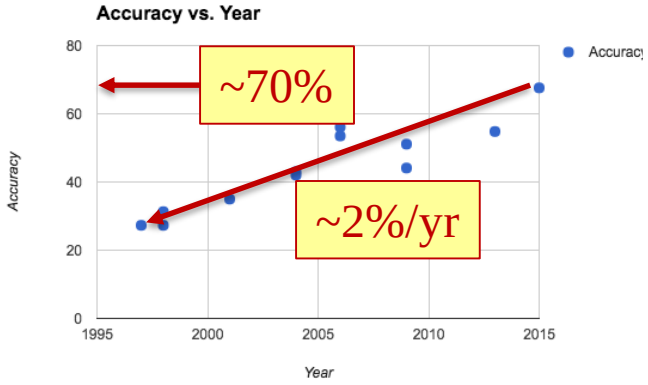


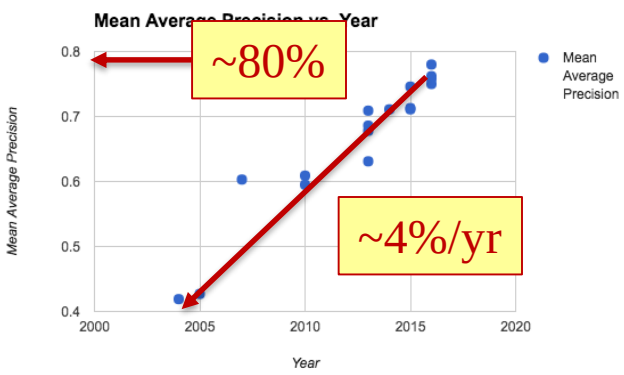


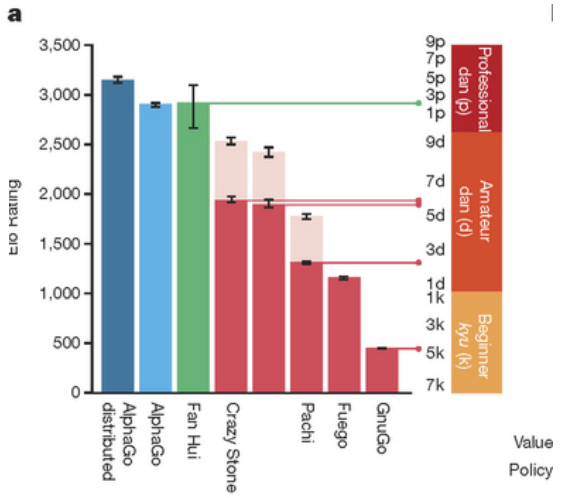
- Continuous server ASR word error rate (WER) reduction ~18% / year: combination of algorithms, data, and computing
- Deep learning (DNNs) is driving recent performance improvements in ASR and meaning extraction



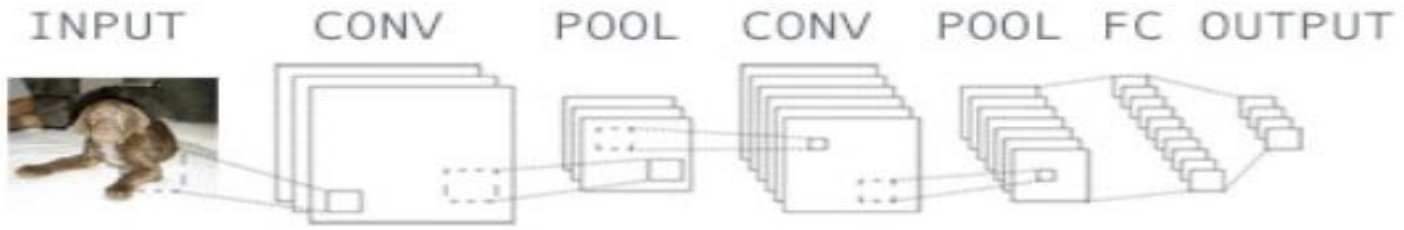
A. Farbin







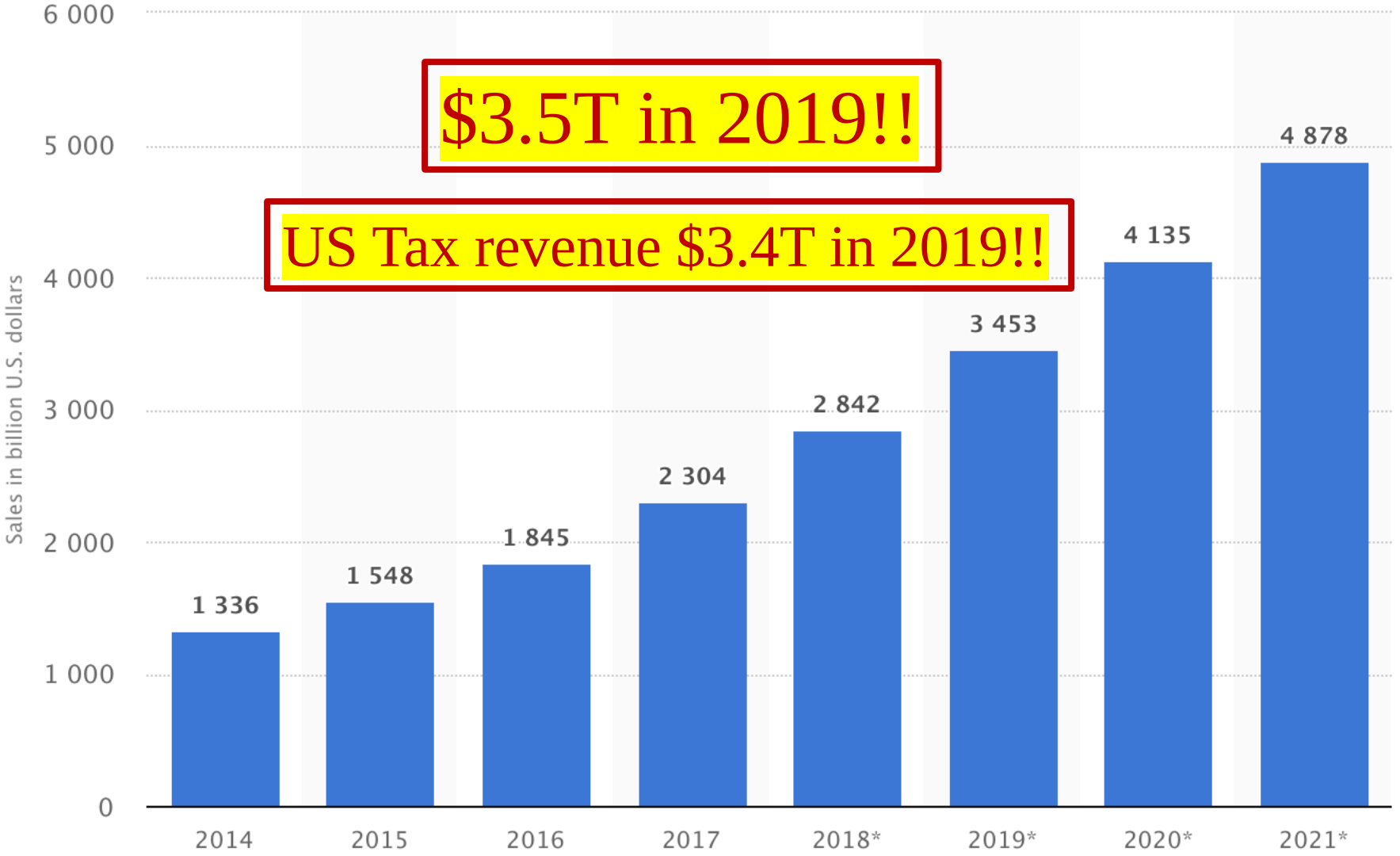
Current Level of ML Performance



Conclusions

- Computing essential for particle physics
- As the amount of data gets bigger, high performance and high throughput computing vital
- HEP driven computing grid technology for hadron collider experiments, delivered necessary performance
  - Software infrastructure established and improving
  - Computing grid now outside of HEP into everyday lives
  - HEP now work together with industry to use their resources
- Powerful computing infra enables highly effective ML
  - Industry drives the development of AI (or ML or DL)
  - Must work adopting ML in preparation for future experiments
- HEP computing helps society virtually immediately

Impact to the World Economy



KISTI Supercomputing Center Tour

9:00 – 9:45am, Friday, July 12

Meet here and depart to the center!