Internet before the LHC (circa 2008)

smaria@caltech.edu

DS@LHC2015

smaria@caltech.edu

DS@LHC2015
FOLLOW-UP

Maurizio Pierini (CERN), Jean-Roch Vlimant (California Institute of Technology (US)), Maria Spiropulu (California Institute of Technology (US)), Kyle Stuart Cranmer (New York University (US)), David Rousseau (LAL-Orsay, FR), Gilles Louppe (CERN), Andrew Lowe (Hungarian Academy of Sciences (HU)), Cecile Germain-Renaud (LRI), Daniel Whiteson (University of California Irvine (US)), Vladimir Gligorov (CERN)

The PC will discuss the collection of ideas or short reports based on what we saw/learned/discussed at the workshop.

e.g. “Use MEM as the dark knowledge to train shallow net…”

“DL: the rawer the data the better”
Couple of points about MEM:

* There is a huge technological and computation barrier to getting MEMs to the hands of average physicists.
  * Lots of very skilled people who are working on MEMs are reinventing the wheel.

* We (Tobias Golling and I) are going to try to get all people interested in MEM together to make an assessment of the tools and the optimization tricks. Talk to ATLAS PC first for advice on how to organize it. Would be good for CMS to do the same.
  * People write me if they are interested so we make sure this is inclusive.

* Then we should try to do sync ATLAS, CMS, and MadGraph/MadWeight community via Experiment Interchange thing (forgot the name).

* The goal is to get some general, efficient, and experimentalist friendly MEM calculating tool, which:
  * Runs on CPU/GPUs.
  * Factorizes different steps, but can all work together. For example:
    * MEM simplification- Oliver is going to develop a symbolic output from ALOHA/MadGraph.
    * Change of Variables- This is one of the main features of MadWeight that hasn’t been reproduced as generally by anyone else. Again Oliver try to factorize this out of MadWeight.
    * Integration- I need to talk to Mike Sokoloff about his grant to do this on GPUs.
  * Response function library
  * Interface to GRID and Supercomputers

* One of the keys to making MEMs is having lots of GPUs.
  * There is a huge interest in US to use HPCs for LHC.
  * ATLAS can land jobs there, but we have no software that uses GPUs (except FTK sim). So it’s kind of a hollow achievement.
  * I hope that I can convince the MadGraph community to work on Event Generation on GPUs on HPCs.
  * Help with NLO, NNLO, and MSSM exploration.
  * It’s been demonstrated, needs to be integrated into MadGraph5.
  * MEMs on HPC would be a meaningful way of using HPCs which could have a huge impact.

* Based on many conversation in this workshop, I believe we have an idea of how to use Dark Knowledge technique from DNN to significantly speed up things.

* It became clear in this workshop that MEM+DNN classifier may be the absolute best method.

* Kyle’s parameterized classifier solution addresses a big set of issues with MEMs relating to systematics and application of MEM to New Physics searches.

* We should keep the momentum going on this stuff.
CS vs THE REST

a short story
DATA (OBSERVED OR SIMULATED)

It's a goal particle physicists have been working on for decades.
DATAFICATION OF EVERYTHING

- 2015: Total IP traffic/year reaches ZB threshold ($10^{21}$)
- 2015: Internet traffic from wireless devices > wired
- 2014: Internet video = 50% of internet consumer traffic
- 2012: 1M households generate > 1 TB/mo
- 2012: Internet video = 50% of internet consumer traffic
- 2010: Internet video > P2P
- 2003: Consumer internet > business internet
Software is eating the world

... because it is more productive!

Technology

Knowledge

Marc Andreessen

on.wsj.com/1w2FbVs
# Machine Intelligence Landscape

## Core Technologies
- **ARTIFICIAL INTELLIGENCE**
  - IBM Watson
  - Amazon Alexa
  - Google Assistant
  - Microsoft Cortana
- **DEEP LEARNING**
  - Facebook
  - Apple
  - Microsoft
- **MACHINE LEARNING**
  - Netflix
  - Spotify
  - Airbnb
- **NLP PLATFORMS**
  - Google Cloud NLP
  - Amazon Comprehend
  - Microsoft Azure Cognitive Services
- **PREDICTIVE API**
  - IBM Predictive Analytics
  - Microsoft PowerBI
  - Salesforce Einstein
- **IMAGE RECOGNITION**
  - Amazon Rekognition
  - Microsoft Azure Computer Vision
  - Google Cloud Vision
- **SPEECH RECOGNITION**
  - IBM Watson Speech to Text
  - Microsoft Azure Speech
  - Google Cloud Speech

## Rethinking Enterprise
### Sales
- **AVISO**
  - NGRDATA
  - FRAMING
  - Infiniti
  - Canada
d
### Security / Authentication
- **Guardian**
  - NESTA
  - Yale Secure
  - LG Ultra-Door
  - Samsung

### Fraud Detection
- **Softscience**
  - Sourcefire
  - BrighterSecurity
  - ThreatConnect

### HR / Recruiting
- **TalentEIN**
  - ConnectIq
  - Paycor
  - HiRecruit

### Marketing
- **Genius360**
  - Yonhap
  - RADIUS
  - Talstrin

### Personal Assistant
- **Siri**
  - Google Now
  - Clever Sans

### Intelligence Tools
- **Qumulo**
  - Palantir
  - Tableau
  - FICO

## Rethinking Industries
### AdTech
- **Gravity**
  - Rocket Fuel
  - YieldMo
  - Adbrain

### Agriculture
- **Blue River**
  - American AgCredit
  - Farm Rich

### Education
- **Knewton**
  - Knewton
  - Edgenuity
  - ALEAP

### Finance
- **Bloomberg**
  - M1 Finance
  - SoFi
  - Robinhood

### Legal
- **Lex Machina**
  - Covington
  - Ravel
  - Justia

### Manufacturing
- **MicroScan**
  - 3M
  - R2M
  - Advanced Sensors

### Medical
- **Paravion**
  - Transcriber
  - SkyGenius
  - Mindful MD

### Oil and Gas
- **Kaggle**
  - AYS and
  - Tachyon

### Media / Content
- **Outbrain**
  - NNTV
  - HubSpot
  - Zefr

### Consumer Finance
- **DataKind**
  - thorn
  - Data Guard
  - DataSafe

### Philanthropies
- **Fōllay**
  - Alston
  - MSL

### Automotive
- **Enlight**
  - ASCAN
  - DSC
  - Trim Acta

### Diagnostics
- **Bay Sensors**
  - Praxis Skylabs
  - Cellect

### Retail
- **NA**
  - NA
  - NA

## Rethinking Humans / HCI
### Augmented Reality
- **Magenta**
  - Magic Leap
  - Neuralink
  - Next VR

### Gestural Computing
- **Leap Motion**
  - Intel RealSense
  - Myo

### Robotics
- **Robotics**
  - Boston Dynamics
  - Rethink Robotics

### Emotional Recognition
- **Neural Turing Machines**
  - NeuralDynamics
  - Neural Matters

## Supporting Technologies
### Hardware
- **Intel**
  - NVIDIA
  - ARM

### Data Prep
- **Tableau**
  - KNIME
  - R

### Data Collection
- **Cloudera**
  - Hortonworks
  - MapR

---

[www.shivonzilis.com/machineintelligence](http://www.shivonzilis.com/machineintelligence)

Bloomberg BETA
Different domain sciences are facing different challenges in the computation e.g.

— climate: local predictions (high granularity)
— neuroscience: data structures uniformity
— astronomy: real time classification/discovery, monitoring and trigger
— microscopy: high dimensional data
— biology/genomics: discovering context (systems decomposition, or reduction amidst complexity)
Different domain sciences have common or similar challenges when massive computation and data is involved;

— denoising, errors (local data dynamics)

— data compression, data architectures, data cleaning etc (local data mechanics)

— intelligent data storage, data trafficking (networking), data accessing, sharing, & monitoring for emergent patterns and optimization (global data “organics”)

COMPUTATIONAL REFOCUSING

computer/software becomes part of the imaging system (as has been the case for a long time in HEP)
The scientific method (and knowledgebase) has been transformed by massive computation (includes networking/sharing, data & data analytics) in all science (and other) domains of research & study.
Massive computation:

i) data mining & analytics for subtle/rare patterns in vast databases

ii) massive simulations of a physical system’s complete evolution repeated numerous times, as simulation parameters vary systematically

iii) ...
THE ENHANCED SCIENTIFIC METHOD

A. Deductive/Theory (tools: Math, Logic)
B. Inductive/Empirical (tools: controlled experiments, statistical analysis of the data, machinery for hypothesis testing)

Large scale extrapolation and prediction = knowledge from computation AND tool for enhancing the capacity of both A. and B.

massive computation builds new intuition
HEP & COMPUTATION

HEP has very long experience in computation, data analytics, data architectures and networking because of the science needs and requirements (instruments used + data produced + scales of collaborations) (ability and necessity dwell near each other Pythagoras)

N.B HEP has a theory. this can be used as a big plus to bootstrap on the data without losing track of the fundamental mechanisms
Accelerators

- Faster hardware
- Better algorithms
- Massive datasets

The breakthroughs that have unleashed AI on the world
www.wired.com/2014/10/future-of-artificial-intelligence/

Kaspar Etter
Research is to see what everybody else has seen and to think what nobody else has thought.
Albert Szent-Györgyi

The task will be to use the data in new ways such that we ask **bigger questions** and challenge our understanding of the world if the data indicates so.