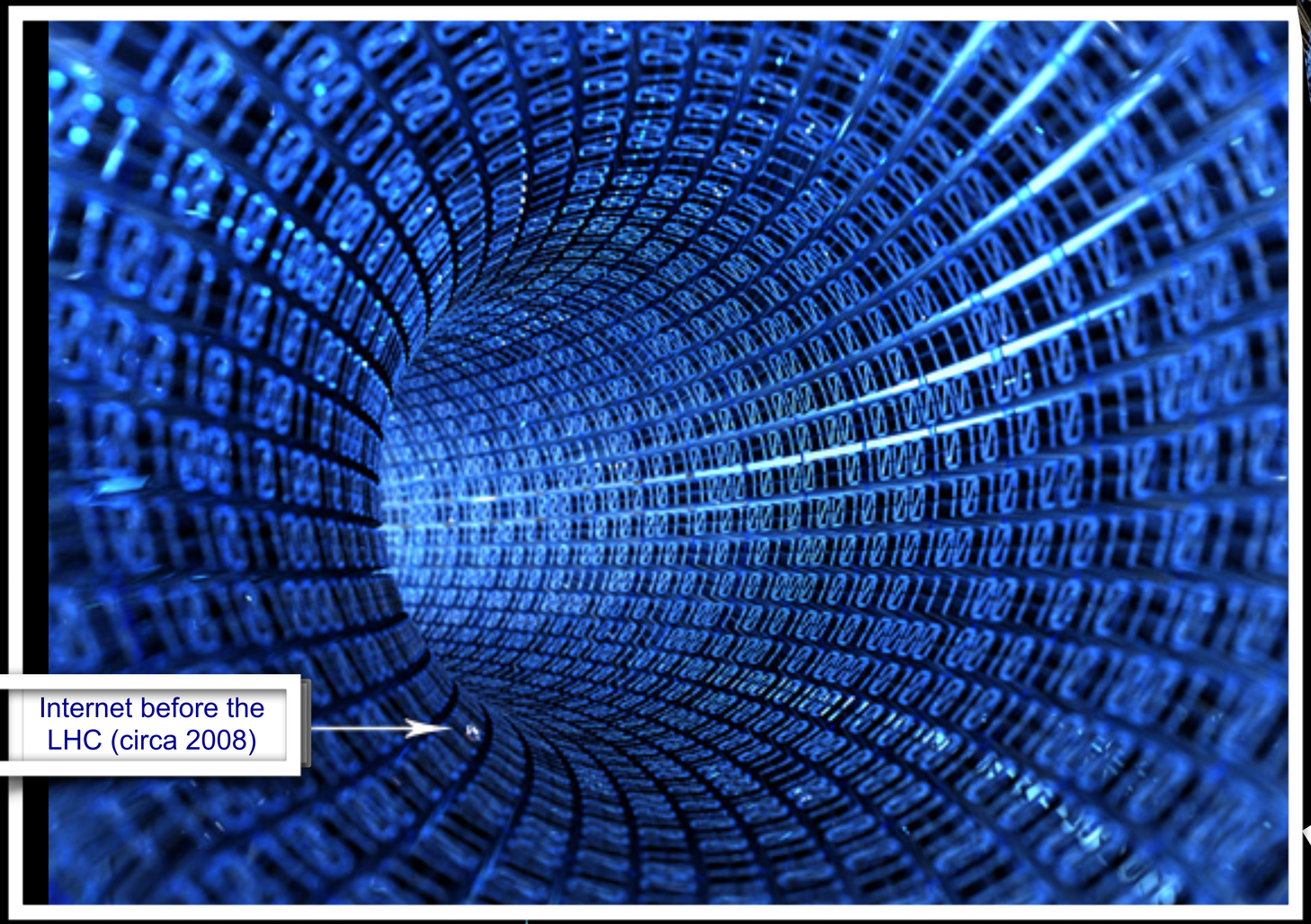
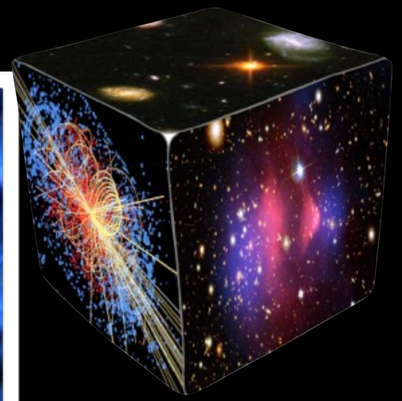
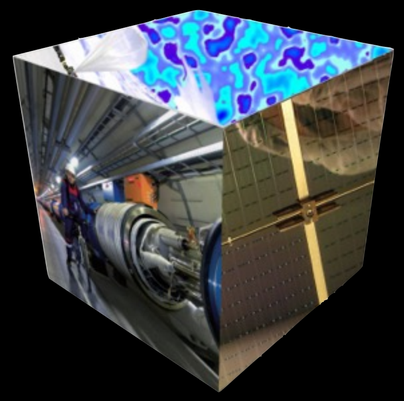
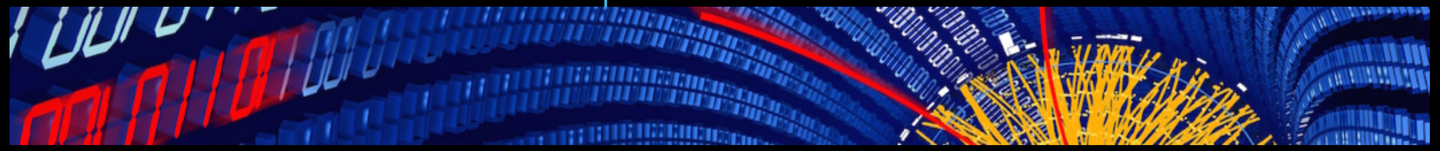
DS@LHC2015 Remarks





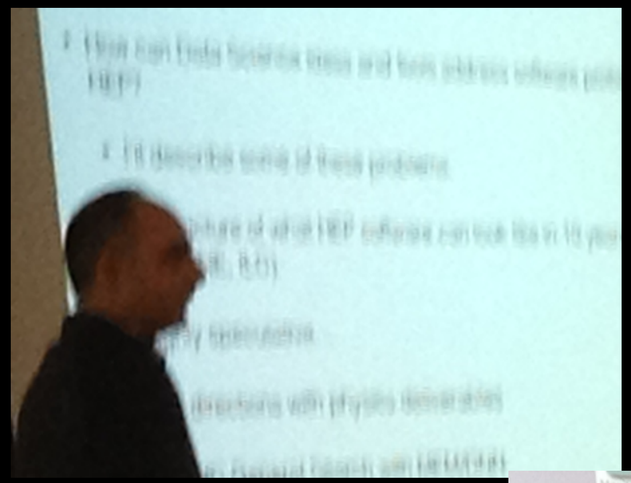


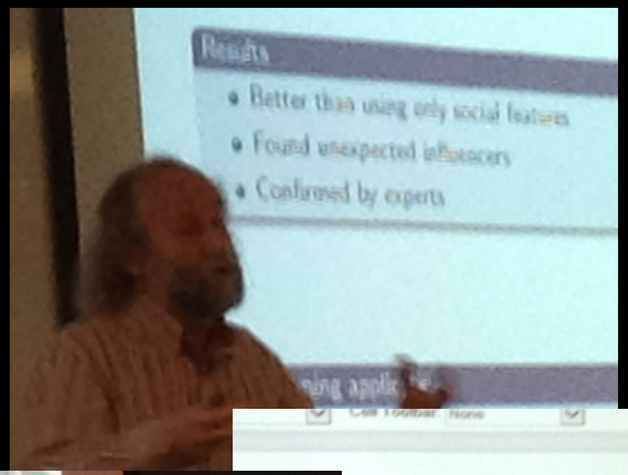
DS@LHC2015

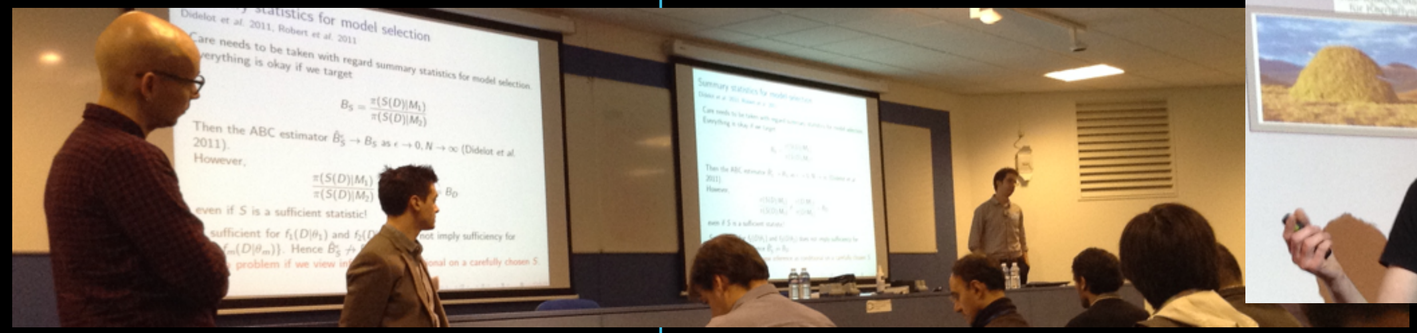


smaria@caltech.edu

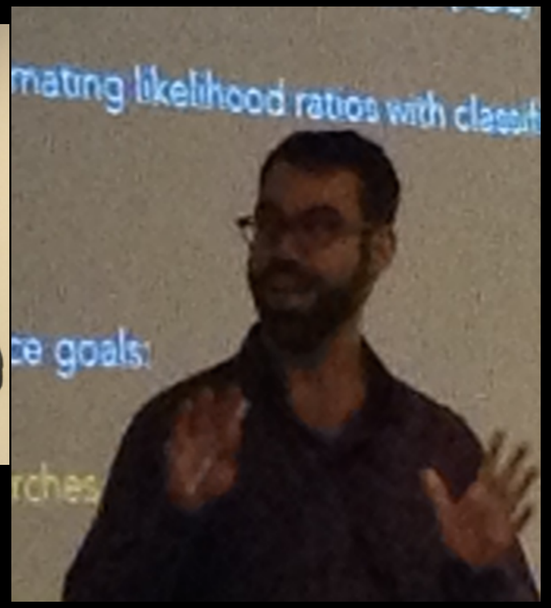

rk City as a laboratory
nique public-private research center that uses
aboratory and classroom to help cities around
come more productive, livable, equitable, and
es, analyzes, and models NYC to optimize
totype new solutions, formalize new tools
, and develop new expertise (and experts!).
Disciplines = Urban Informatics




Results
Better than using only social features
Found unexpected influencers
Confirmed by experts


statistics for model selection
Didelot et al. 2011; Robert et al. 2011
Care needs to be taken with regard summary statistics for model selection.
Everything is okay if we target
$B_S = \frac{\pi(S(D)|M_1)}{\pi(S(D)|M_2)}$
Then the ABC estimator $\hat{B}_S^\epsilon \to B_S$ as $\epsilon \to 0, N \to \infty$ (Didelot et al. 2011).
However,
$\frac{\pi(S(D)|M_1)}{\pi(S(D)|M_2)} \neq B_D$
even if S is a sufficient statistic!


ROOT
Data Analysis Framework
TMVA


mating likelihood ratios with classif
ce goals:
rches

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"

FOLLOW-UP (V. AMIR)

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.

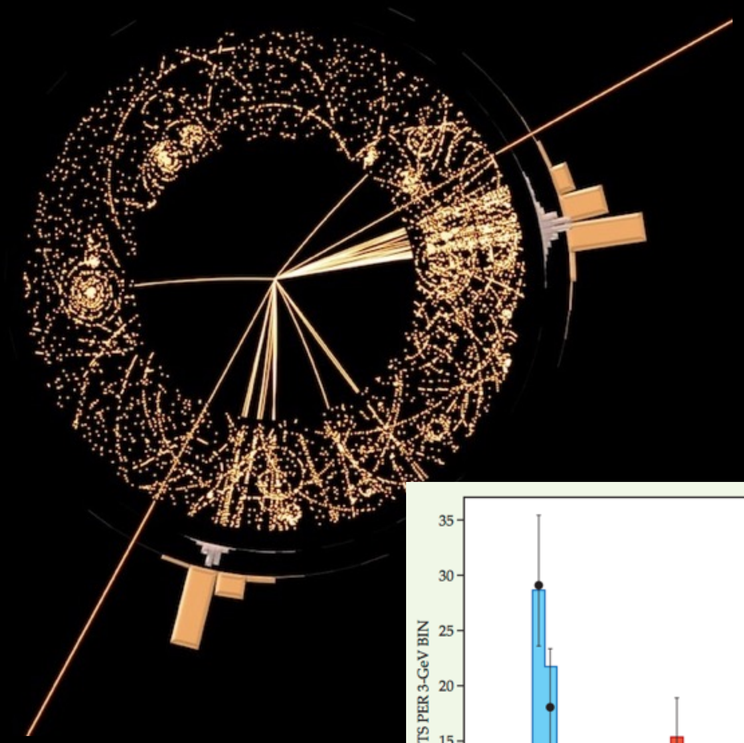
* We should keep the momentum going on this stuff.

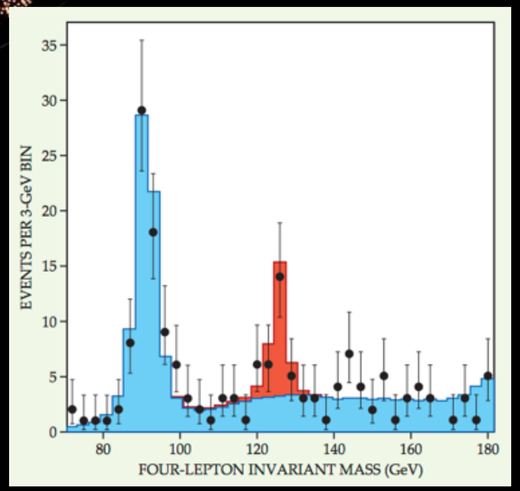
CS vs THE REST

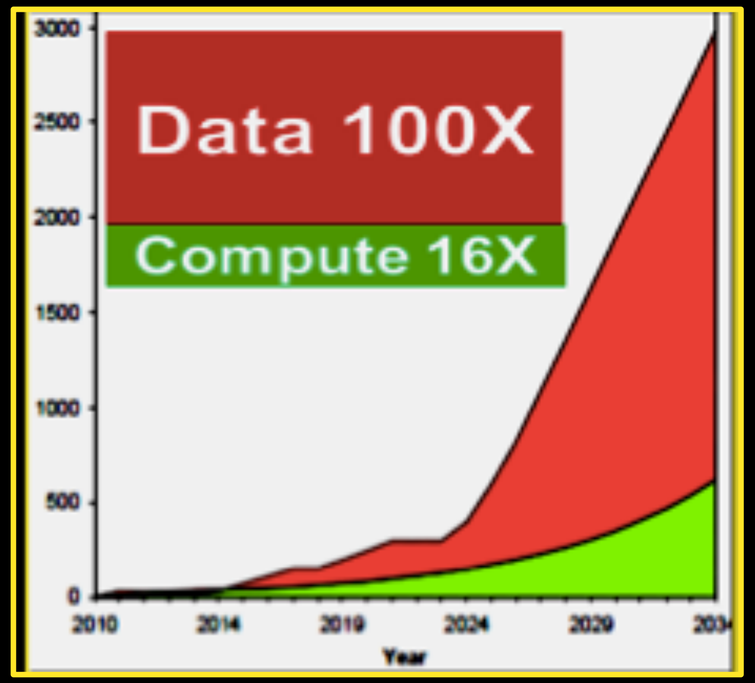
a short story

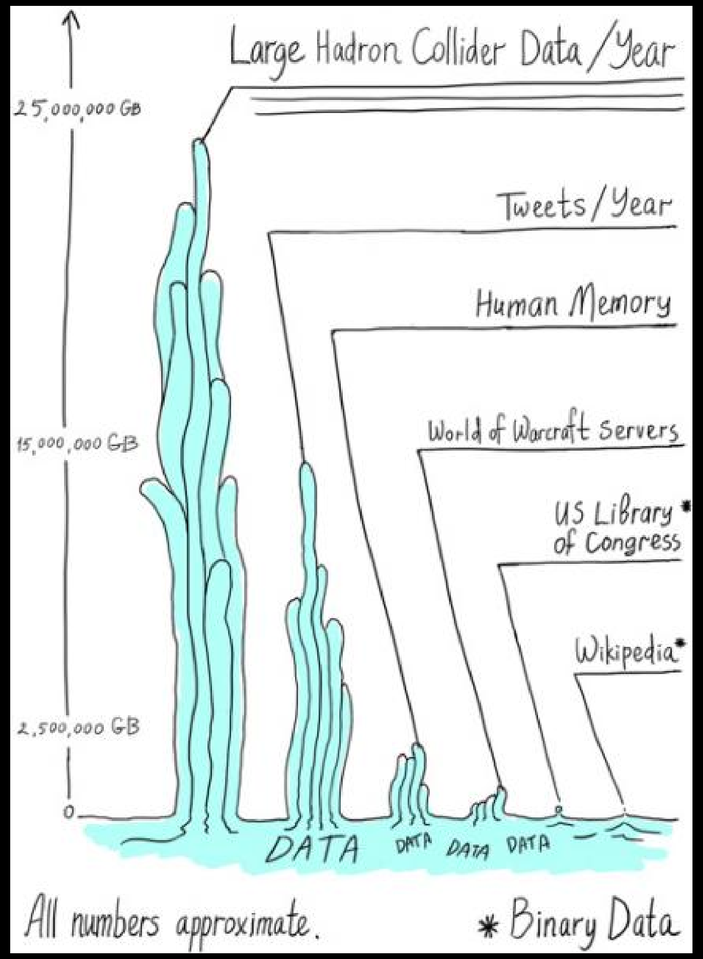
DATA (OBSERVED OR SIMULATED)



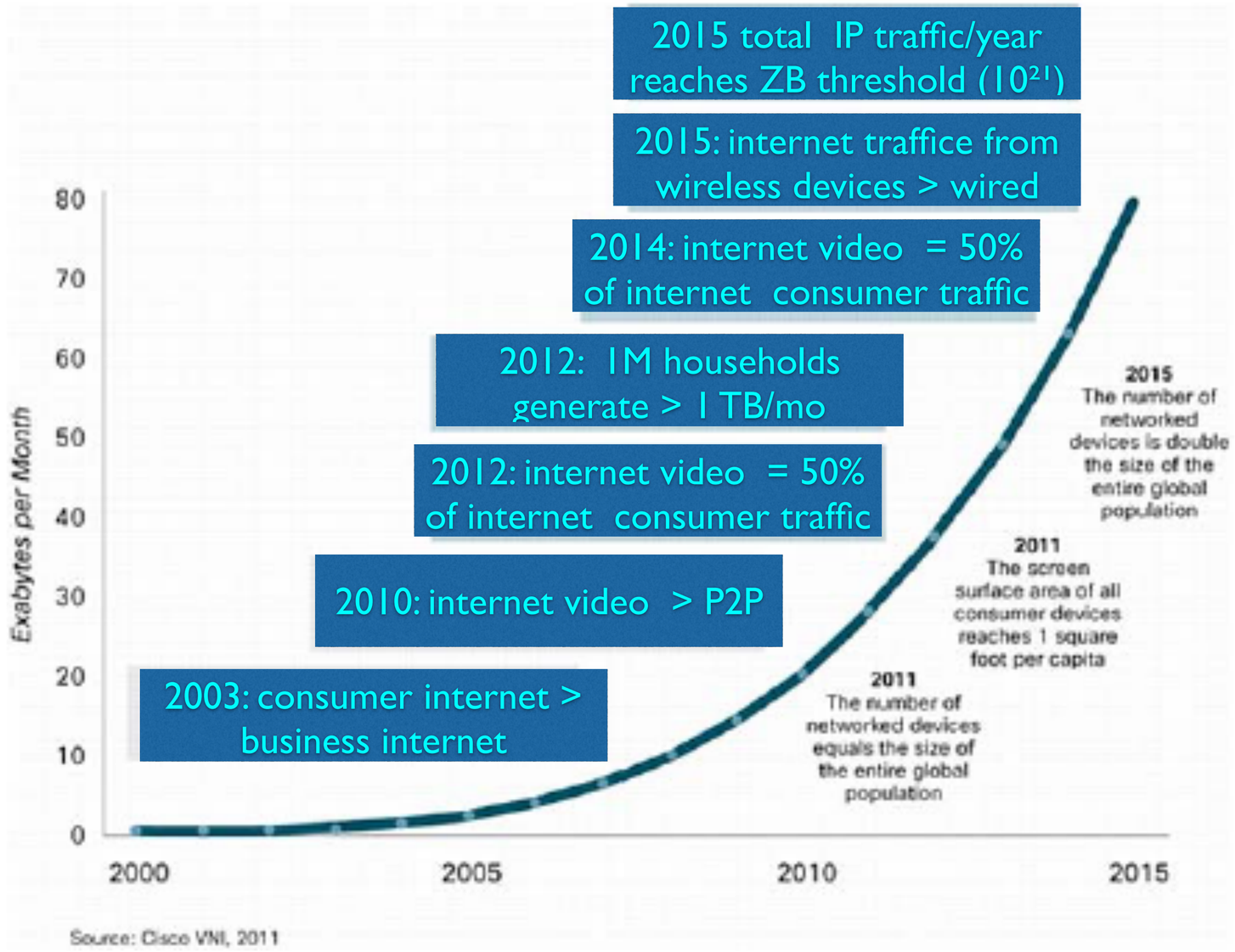


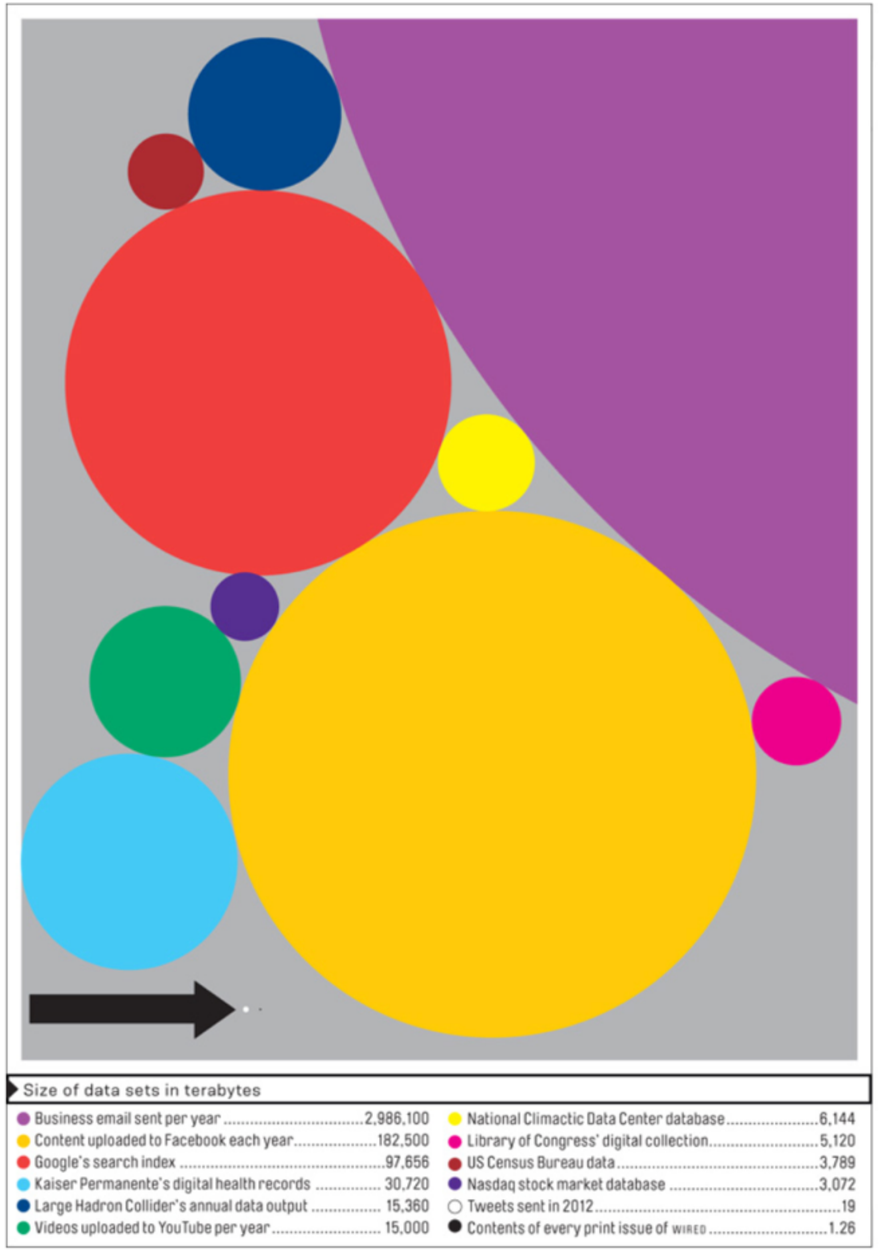






DATAFICATION OF EVERYTHING

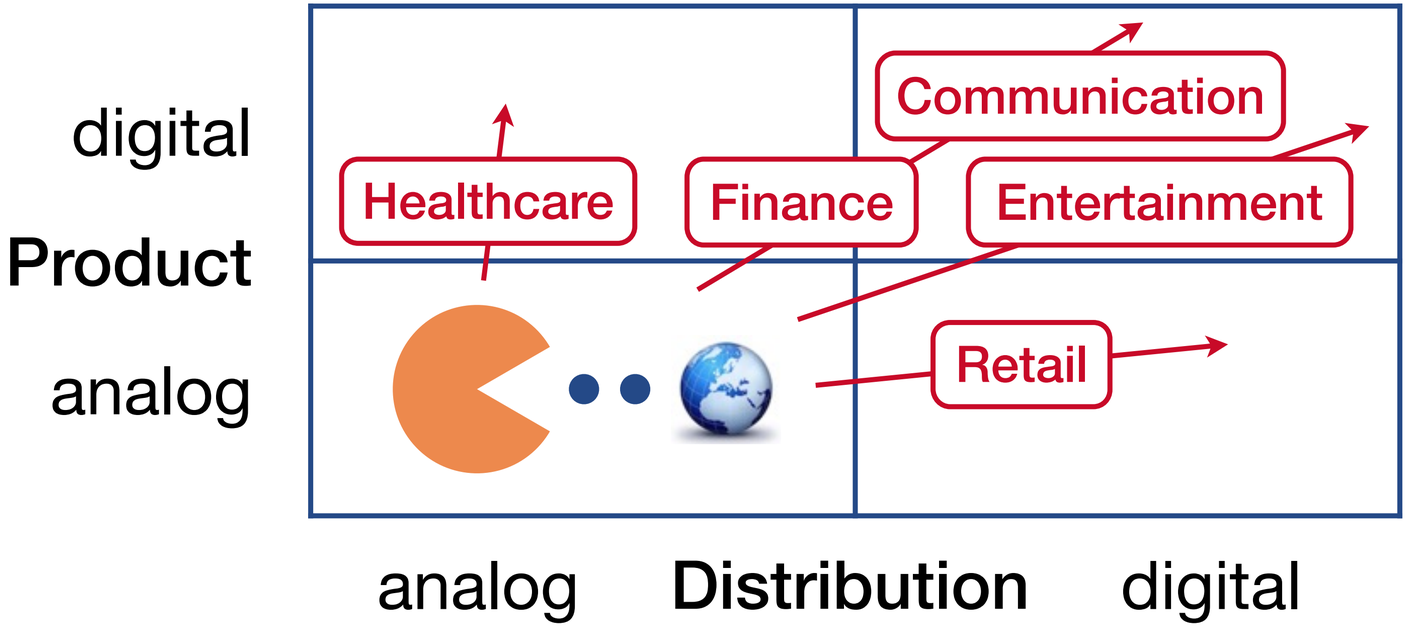


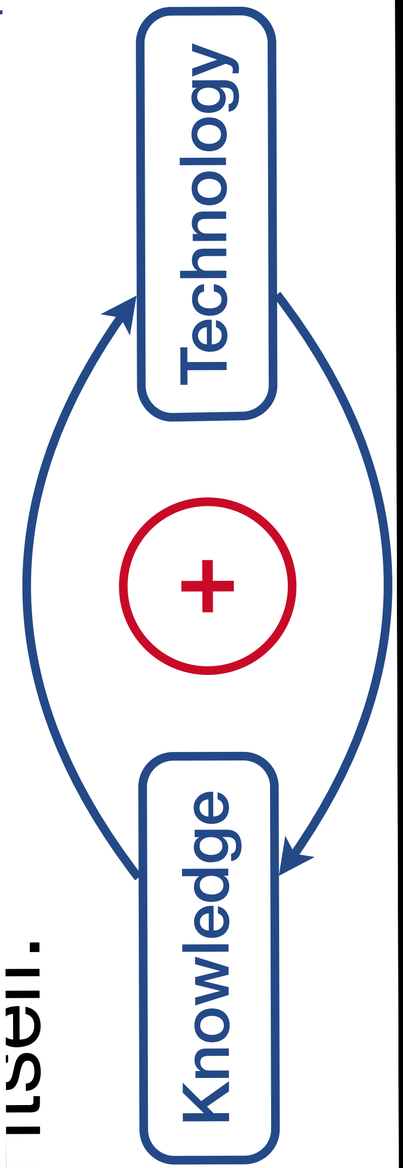


Kaspar

Software is eating the world

… because it is more productive!





Marc Andreessen
on.wsj.com/1w2FbVs

Machine Intelligence LANDSCAPE

CORE TECHNOLOGIES

ARTIFICIAL INTELLIGENCE
IBM WATSON, MetaMind, Numenta, ai-one, Cycorp, Microsoft Research, nara, Reactor, SI SCALED INFERENCE

DEEP LEARNING
vicarious, Vision Factory, facebook, Baidu IDL, Google, ersatz, SKYMIND, SignalSense

MACHINE LEARNING
rapidminer, context relevant, 0xdata H2O, DATARPM, Liftigniter, SPARKBEYOND, Azure ML, yhat, Sense, GraphLab, Alpine, Ionian

NLP PLATFORMS
cortical.io, idibon, LUMINOSO, wit.ai, MaluubA

PREDICTIVE APIS
AlchemyAPI, MINDOPS, Google, bigml, indico, ALGORITHMIA, Expect Labs, PredictionIO

IMAGE RECOGNITION
clarifai, MADBITS, DNNresearch, DEXTRO, VISENZE, lookflow

SPEECH RECOGNITION
GRIDSPACE, popUp archive, NUANCE

RETHINKING ENTERPRISE

SALES
Preact, AVISO, RelateIQ, NGDATA, CLARABRIDGE, FRAMED, infer, ATTENSITY, causata

SECURITY / AUTHENTICATION
CROSSMATCH, conjur, EYEVERIFY, BITSIGHT, CYLANCE, AREA 1 SECURITY, bionym

FRAUD DETECTION
sift science, SOCURE, ThreatMetrix, feedzai, Brighterion, VERAFIN

HR / RECRUITING
TalentBin, entelo, predikt, Connectifier, gild, hiQ, CONCEPTINDEX

MARKETING
brightfunnel, bloomreach, CommandIQ, AIRPR, RADIUS, TellApart, people pattern, Freshplum

PERSONAL ASSISTANT
Siri, Google now, Cortana, cleversense, tempo, Robinlabs, KASISTO, fuse|machines, VIV, Tricorder, CLARA LABS

INTELLIGENCE TOOLS
ADATAO, Palantir, Quid, Digital Reasoning, FirstRain

RETHINKING INDUSTRIES

ADTECH
METAMARKETS, dstillery, rocketfuel, YieldMo, ADBRAIN

AGRICULTURE
BLUE RIVER, TerrAvion, ceresimaging, HONEYCOMB, THE CLIMATE CORPORATION, tule, MAVEN

EDUCATION
declara, coursera, KNEWTON, kidaptive

FINANCE
Bloomberg, FinGenius, alphasense, KENSHO, Dataminr, minettabrook, BINATIX, MAVEN

LEGAL
Lex Machina, brightleaf, COUNSELYTICS, RAVEL, JUDICATA, Brevia, DiligenceEngine

MANUFACTURING
SIGHT MACHINE, MICROSCAN, BOULDER IMAGING, IVISYS

MEDICAL
Parzival, transcriptic, Genescient, ZEPHYR HEALTH, grand round table, bina, TUTE GENOMICS

OIL AND GAS
kaggle, AYASDI, TACHYUS, biota, Flutura

MEDIA / CONTENT
outbrain, newsle, ARRIA, SAILTHRU, wavii, Owlin, NarrativeScience, ysoap, Sumnly, Prismatic, ai

CONSUMER FINANCE
affirm, inMenture, zest finance, BILLGUARD, LendUp, LendingClub, Kabbage

PHILANTHROPIES
DataKind, thorn, DATA GUILD

AUTOMOTIVE
Google, Continental, TESLA, MOBILEYE, CRUISE

DIAGNOSTICS
enlitic, 3SCAN, lumiata, ENTOPIS

RETAIL
BAY SENSORS, PRISM SKYLABS, celect, euclid

RETHINKING HUMANS / HCI

AUGMENTED REALITY
wearable intelligence, MAGMAR, APX, blippar, META, layar

GESTURAL COMPUTING
THALMIC LABS, omek, LUTTER, LEAP MOTION, eyeSight, 3Gear Systems, GestureTek, nod

ROBOTICS
intel, LIQUID ROBOTICS, SoftBank, iRobot, Boston Dynamics, jibo, anki, evolution robotics

EMOTIONAL RECOGNITION
affectiva, BEYONDVERBAL, EMOTIENT, BRIGHAM, cogito

SUPPORTING TECHNOLOGIES

HARDWARE
NVIDIA, XILINX, QUALCOMM, NERVANA SYSTEMS, TERADEEP, Artificial Learning, rigetti

DATA PREP
TRIFACTA, Paxata, tamr, Alation

DATA COLLECTION
diffbot, kimono, CrowdFlower, Connotate, WorkFusion, import.io

www.shivonzilis.com/machineintelligence

Bloomberg BETA

DOMAIN SCIENCES & COMPUTATION

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)

DOMAIN SCIENCES & COMPUTATION

Different domain sciences <u>have common or similar challenges</u> 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)

VINCENZO'S FIREWORKS

Valentin Volk

BIG CHANGE

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.

BIG CHANGE

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 bootstap 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/

DOMAIN SCIENCES & COMPUTATION

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