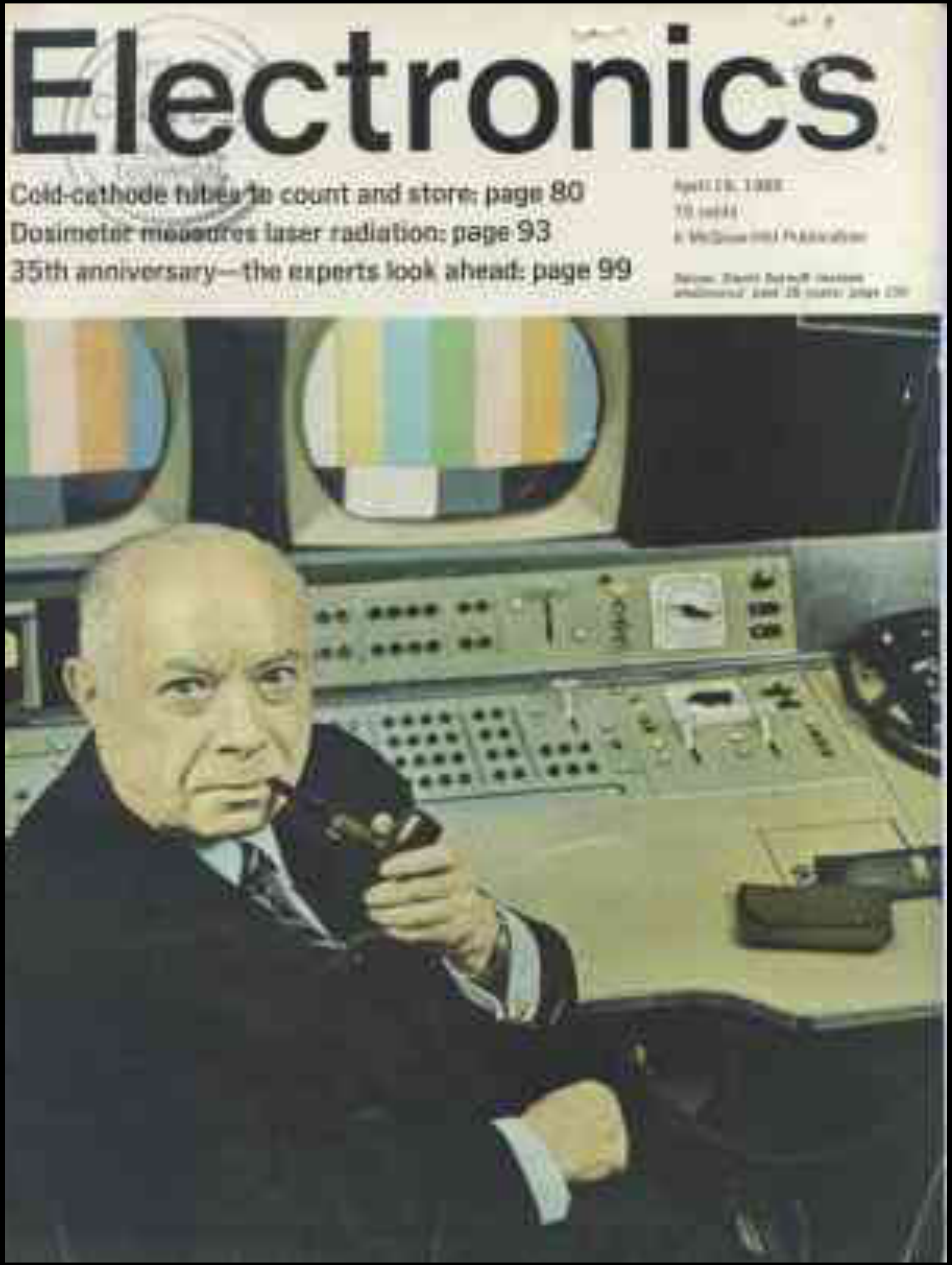
DATA SCIENCE & HEP

Maria Spiropulu
California Institute of Technology

smaria@caltech.edu

MOORE (1965)



Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore
Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

The future of integrated electronics is the future of electronics itself. The advantages of integration will bring about a proliferation of electronics, pushing this science into many new areas.

Integrated circuits will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment. The electronic wrist-

machine instead of being concentrated in a central unit. In addition, the improved reliability made possible by integrated circuits will allow the construction of larger processing units. Machines similar to those in existence today will be built at lower costs and with faster turn-around.

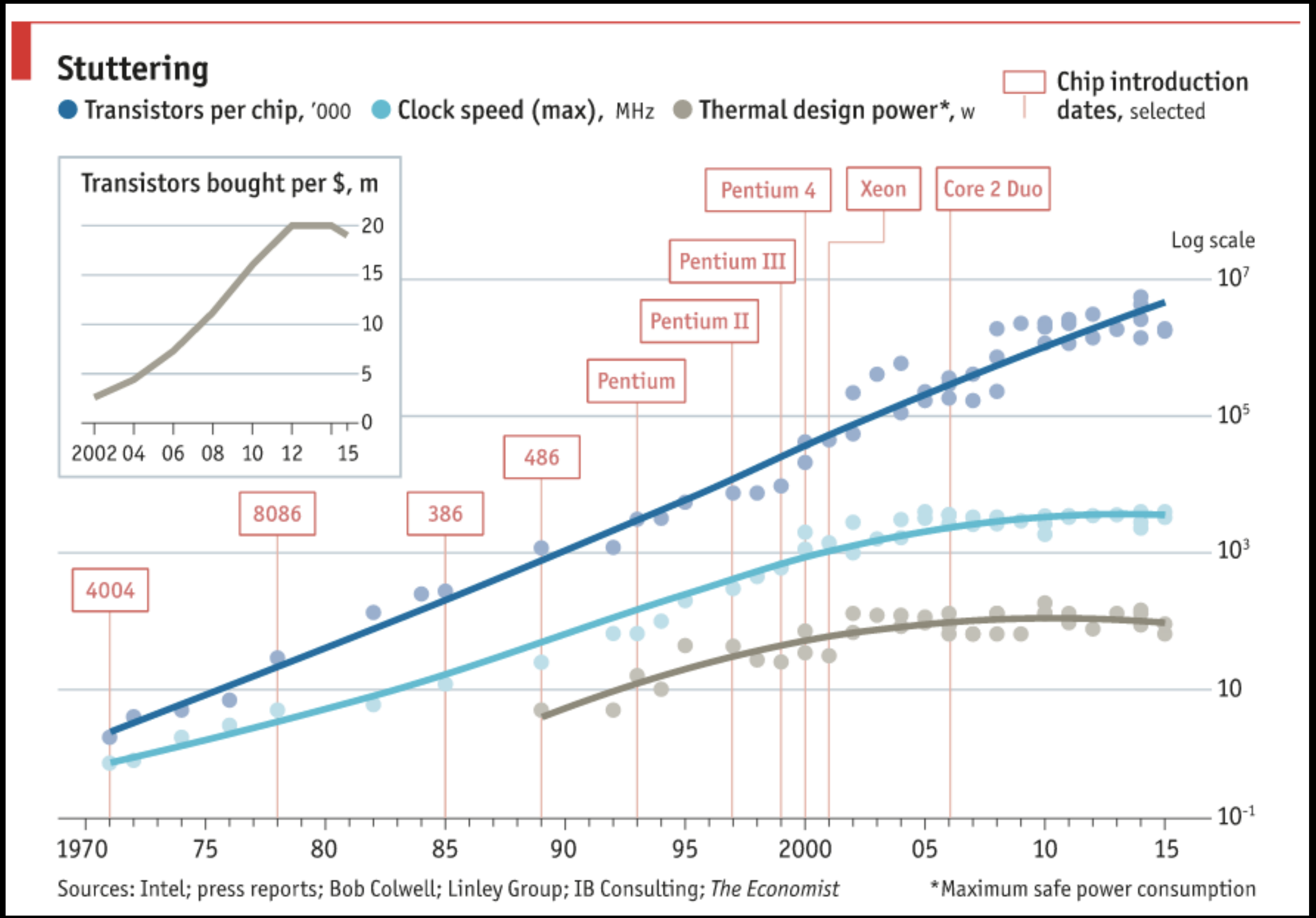
Present and future

By integrated electronics, I mean all the various tech-

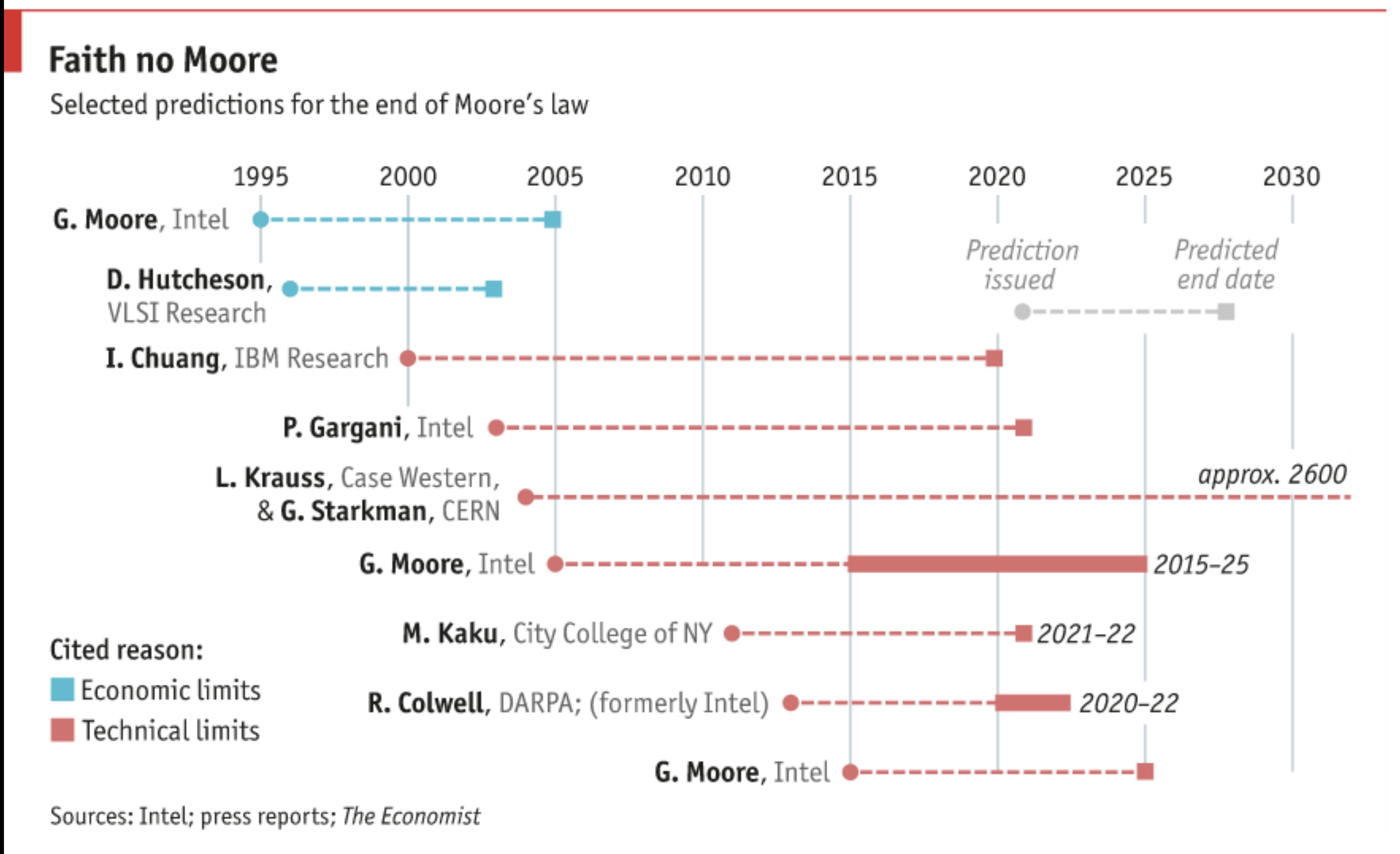
The components are approaching a fundamental limit of smallness: the atom

AFTER MOORE

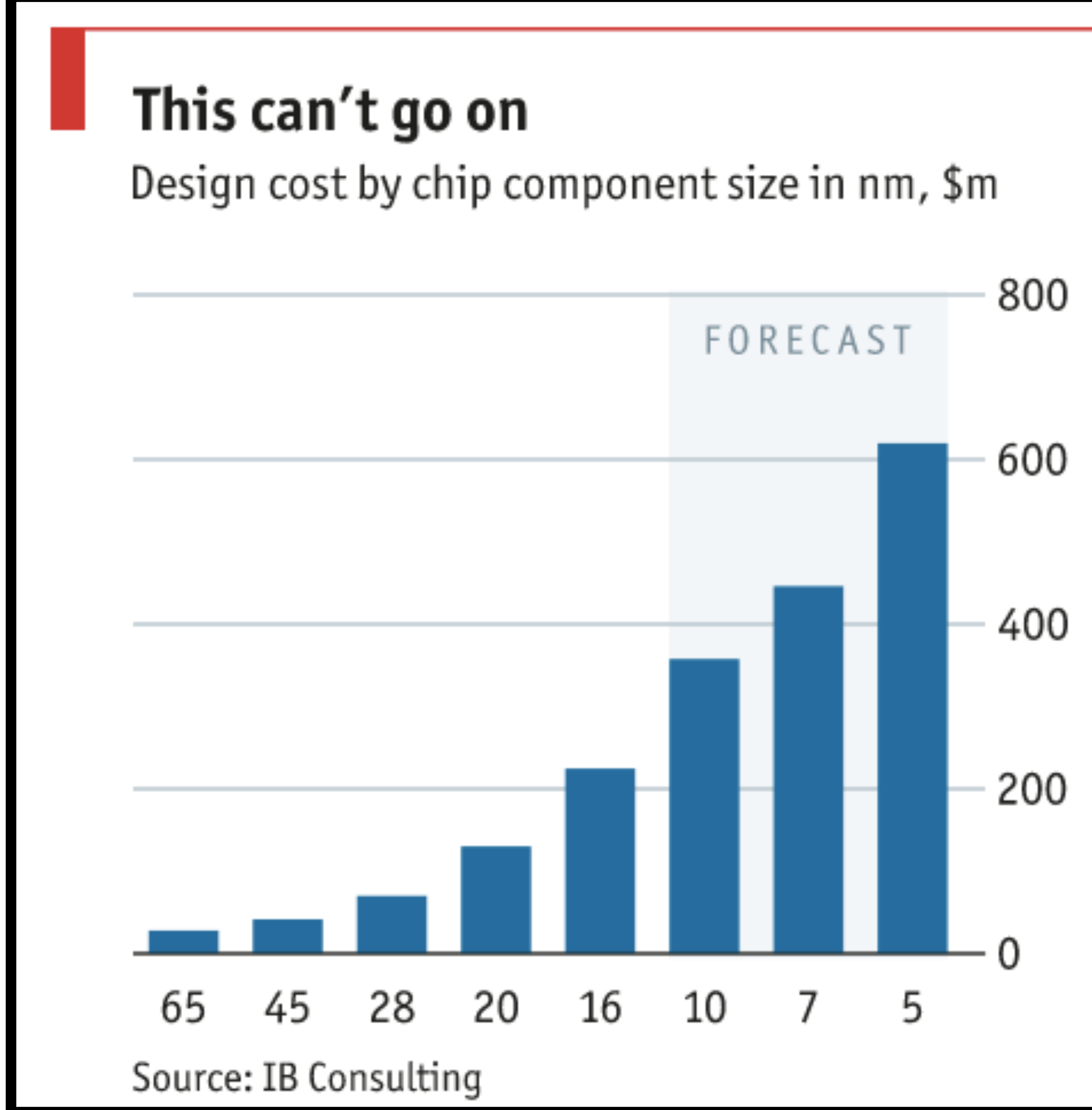
http://www.economist.com/technology-quarterly/2016-03-12/after-moores-law



"The number of people predicting the death of Moore's Law doubles every two years" Peter Lee (a VP at Microsoft)



Faith no Moore
Selected predictions for the end of Moore's law
1995
2000
2005
2010
2015
2020
2025
2030
G. Moore, Intel
D. Hutcheson, VLSI Research
Prediction issued
Predicted end date
I. Chuang, IBM Research
P. Gargani, Intel
L. Krauss, Case Western, & G. Starkman, CERN
approx. 2600
G. Moore, Intel
2015–25
M. Kaku, City College of NY
2021–22
R. Colwell, DARPA; (formerly Intel)
2020–22
G. Moore, Intel
Cited reason:
Economic limits
Technical limits
Sources: Intel; press reports; The Economist



This can't go on
Design cost by chip component size in nm, $m
800
600
400
200
0
FORECAST
65
45
28
20
16
10
7
5
Source: IB Consulting

A pipeline of new technologies to prolong Moore's magic

THE world's IT firms spend huge amounts on research and development. In 2015 they occupied three of the top five places in the list of biggest R&D spenders compiled by PricewaterhouseCoopers, a consultancy. Samsung, Intel and Microsoft, the three largest, alone shelled out $37 billion between them. Many of the companies are working on projects to replace the magic of Moore's law. Here are a few promising ideas.

Optical communication: the use of light instead of electricity to communicate between computers, and even within chips. This should cut energy use and boost performance *Hewlett-Packard, Massachusetts Institute of Technology*.

Better memory technologies: building new kinds of fast, dense, cheap memory to ease one bottleneck in computer performance *Intel, Micron*.

Quantum-well transistors: the use of quantum phenomena to alter the behaviour of electrical-charge carriers in a transistor to boost its performance, enabling extra iterations of Moore's law, increased speed and lower power consumption *Intel*.

Developing new chips and new software to automate the writing of code for machines built from clusters of specialised chips. This has proved especially difficult *Soft Machines*.

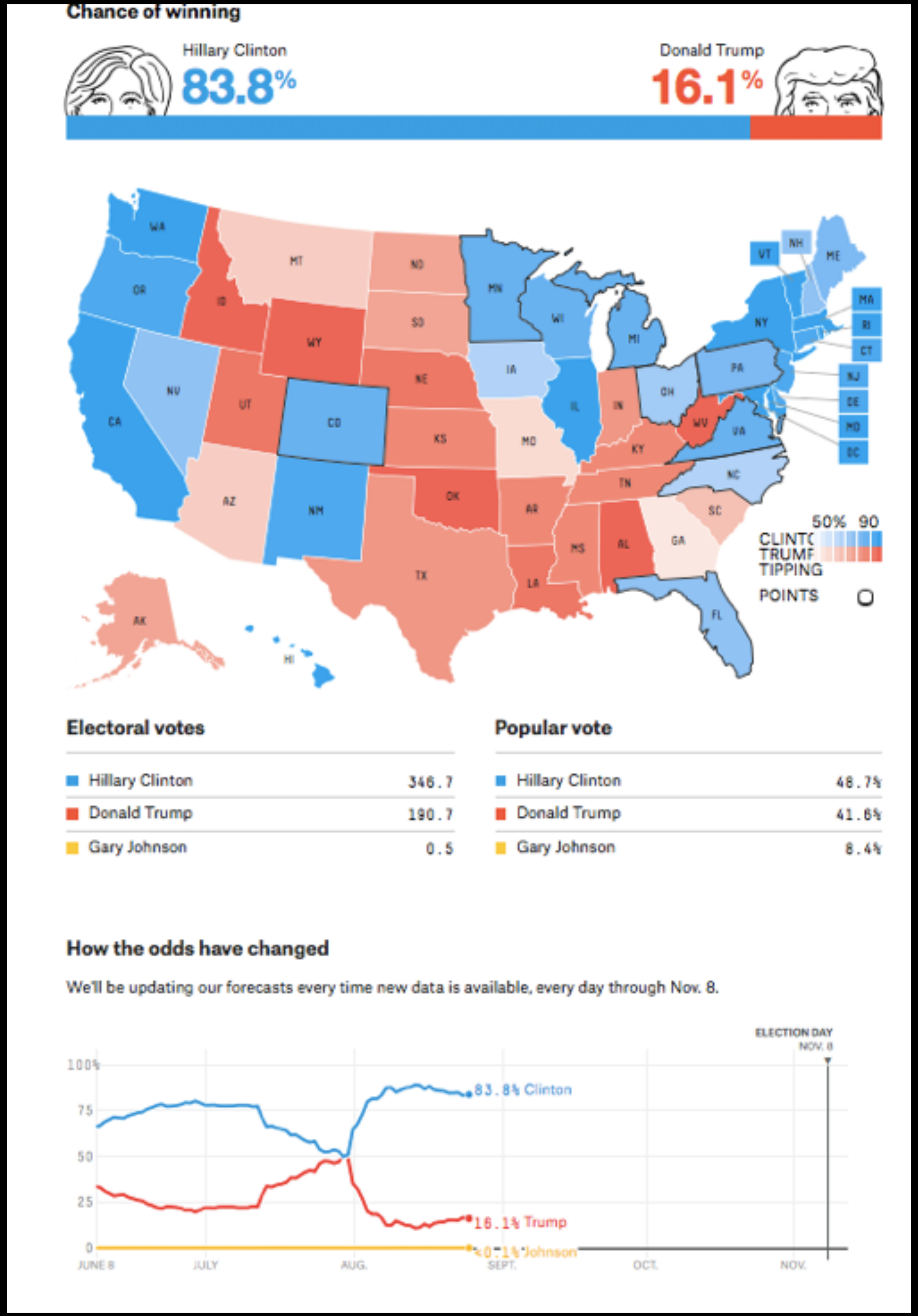
Approximate computing: making computers' internal representation of numbers less precise to reduce the numbers of bits per calculation and thus save energy; and allowing computers to make random small mistakes in calculations that cancel each other out over time, which will also save energy *University of Washington, Microsoft*.

Neuromorphic computing: developing devices loosely modelled on the tangled, densely linked bundles of neurons that process information in animal brains. This may cut energy use and prove useful for pattern recognition and other AI-related tasks *IBM, Qualcomm*.

Carbon nanotube transistors: these rolled-up sheets of graphene promise low power consumption and high speed, as graphene does. Unike graphene, they can also be switched off easily. But they have proved difficult to mass-produce *IBM, Stanford University*.

DATA SCIENCE

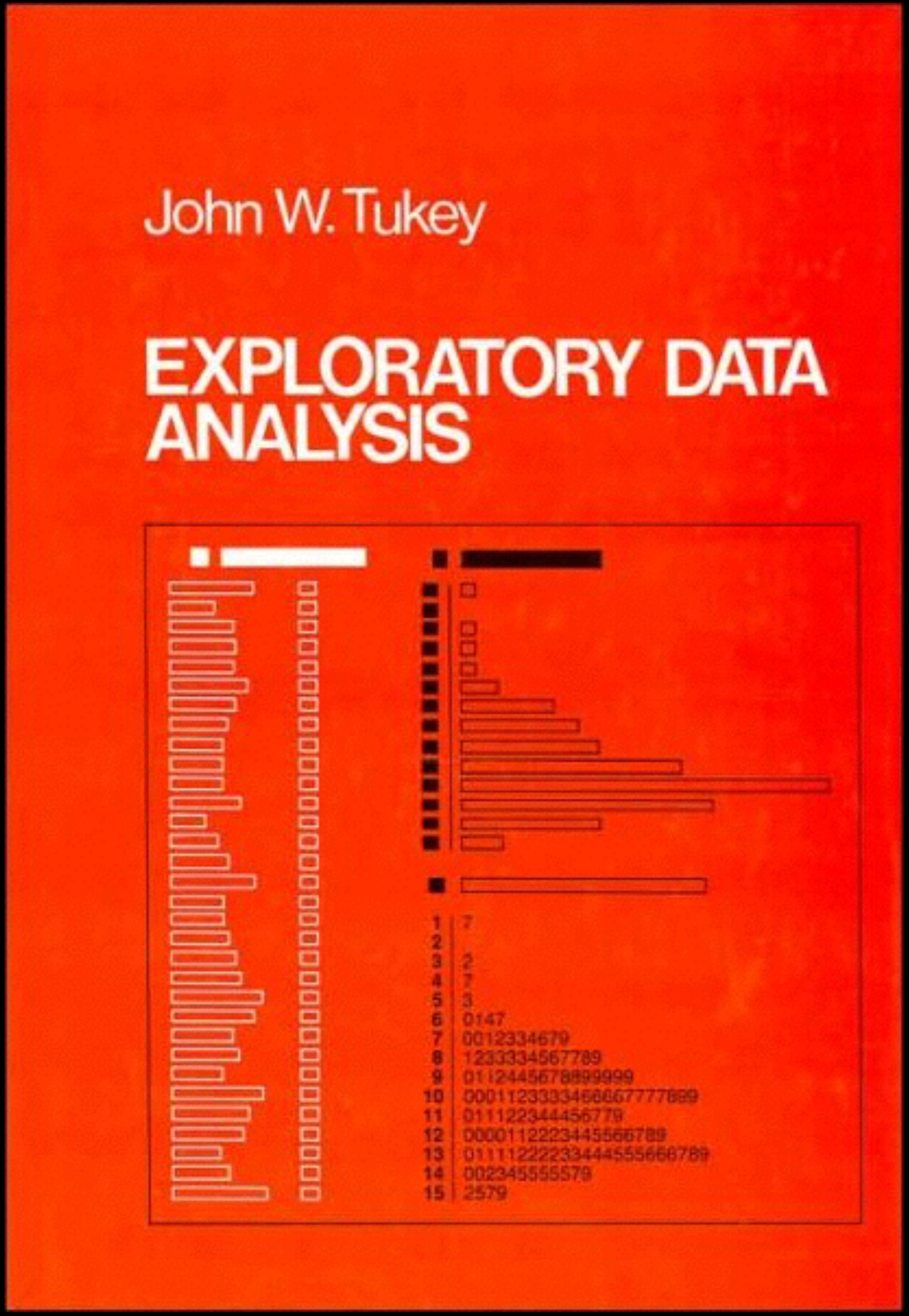
"I think data-scientist is a sexed up term for a statistician," Nate Silver told an audience of statisticians in 2013 at a Joint Statistical Meeting.



This timeline that follows was published in WhatsTheBigData.com
See also A Very Short History of Big Data and A Very Short History of Information Technology (Gil Press)

DATA SCIENCE VS STATISTICS

- 1962 John W. Tukey writes in “The Future of Data Analysis”: “For a long time I thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and doubt… I have come to feel that my central interest is in data analysis… Data analysis, and the parts of statistics which adhere to it, must… take on the characteristics of science rather than those of mathematics… data analysis is intrinsically an empirical science… How vital and how important… is the rise of the stored-program electronic computer? In many instances the answer may surprise many by being ‘important but not vital,’ although in others there is no doubt but what the computer has been ‘vital.’” In 1947, Tukey coined the term “bit” which Claude Shannon used in his 1948 paper “A Mathematical Theory of



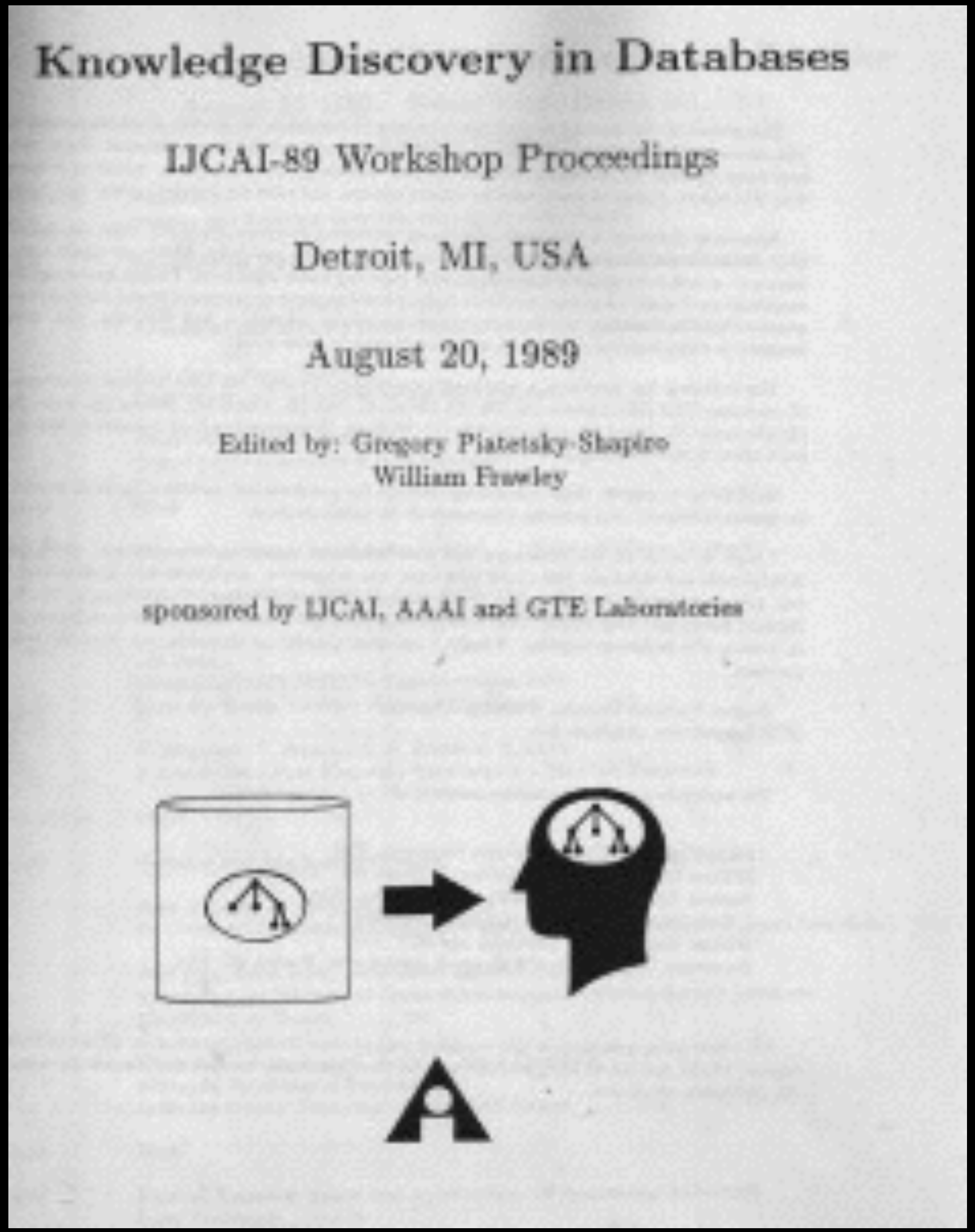
DATA SCIENCE VS DOMAIN SCIENCE

– 1974 Peter Naur publishes Concise Survey of Computer Methods in Sweden and the United States. The book is a survey of contemporary data processing methods that are used in a wide range of applications. It is organized around the concept of data as defined in the IFIP Guide to Concepts and Terms in Data Processing: “[Data is] a representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process.“ The Preface to the book tells the reader that a course plan was presented at the IFIP Congress in 1968, titled “Datalogy, the science of data and of data processes and its place in education,“ and that in the text of the book, ”the term ‘data science’ has been used freely.” Naur offers the following definition of data science: “The science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences.”

DATA SCIENCE VS COMPUTER SCIENCE

- 1977 The International Association for Statistical Computing (IASC) is established as a Section of the ISI. “It is the mission of the IASC to link traditional statistical methodology, modern computer technology, and the knowledge of domain experts in order to convert data into information and knowledge.”

- 1989 Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop. In 1995, it became the annual ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).



THE 90'S

- September 1994 BusinessWeek publishes a cover story on "Database Marketing"
- 1996 Members of the International Federation of Classification Societies (IFCS) meet in Kobe, Japan, for their biennial conference. For the first time, the term "data science" is included in the title of the conference ("Data science, classification, and related methods")
- 1996 Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth publish "From Data Mining to Knowledge Discovery in Databases."
- 1997 In his inaugural lecture for the H. C. Carver Chair in Statistics at the University of Michigan, Professor C. F. Jeff Wu (currently at the Georgia Institute of Technology), calls for statistics to be renamed data science and statisticians to be renamed data

THE 21ST CENTURY

– 1997 The journal Data Mining and Knowledge Discovery is launched

– December 1999 Jacob Zahavi is quoted in "Mining Data for Nuggets of Knowledge" in Knowledge@Wharton: "Conventional statistical methods work well with small data sets. Today's databases, however, can involve millions of rows and scores of columns of data… Scalability is a huge issue in data mining…."

– 2001 William S. Cleveland publishes "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." It is a plan "to enlarge the major areas of technical work of the field of statistics. Because the plan is ambitious and implies substantial change, the altered field will be called 'data science.'"

– April 2002 Launch of Data Science Journal, publishing papers on "the management of data and databases in Science and Technology.

– January 2003 Launch of Journal of Data Science: "By 'Data Science' we mean almost everything that has something to do with data: Collecting, analyzing, modeling…… yet the most important part is its applications–all sorts of applications. This journal is devoted to applications of statistical methods at large…

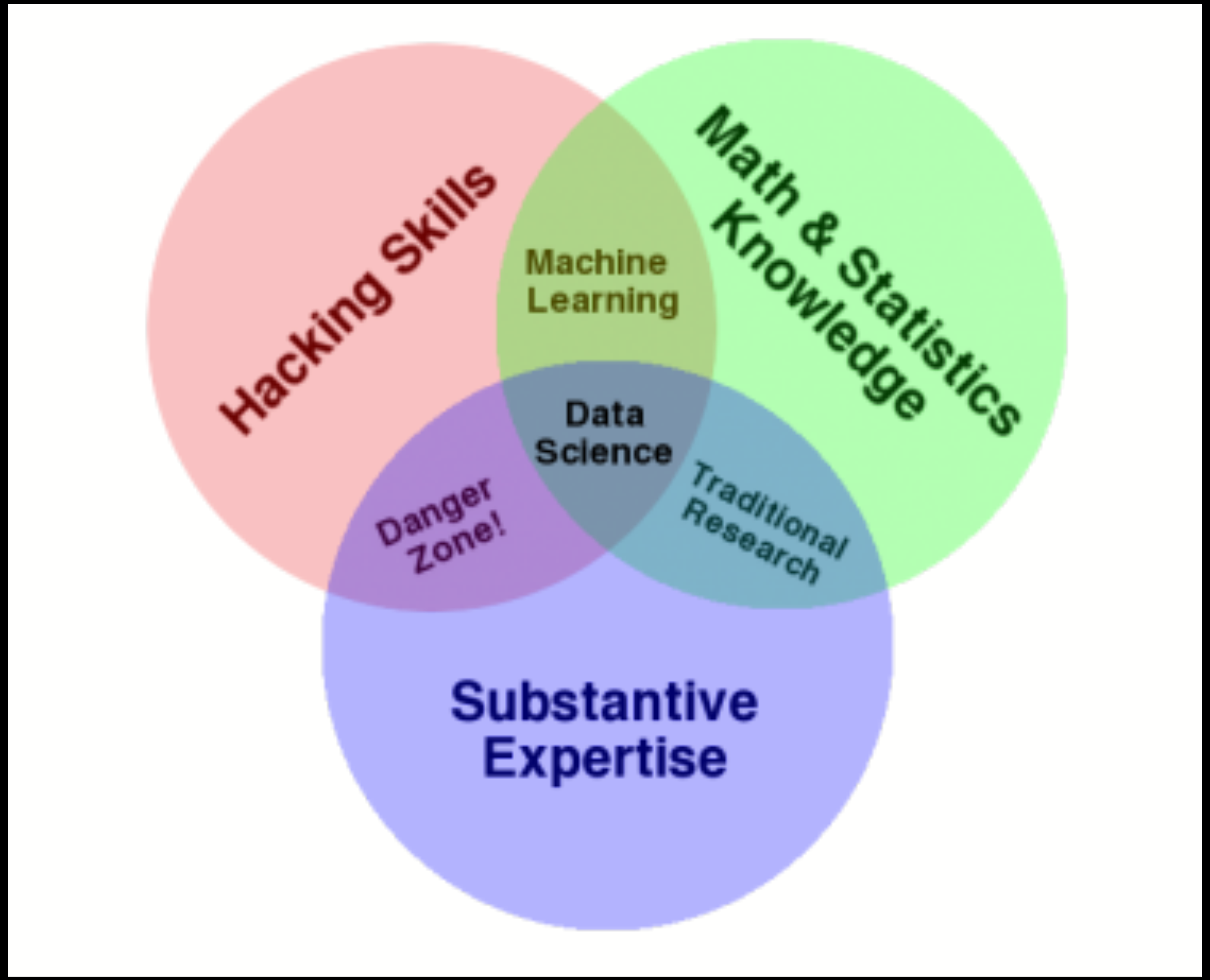
NSF

- September 2005 The National Science Board publishes "Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century." One of the recommendations of the report reads: "The NSF, working in partnership with collection managers and the community at large, should act to develop and mature the career path for data scientists and to ensure that the research enterprise includes a sufficient number of high-quality data scientists." The report defines data scientists as "the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection."

HACKING

- March 2009 Kirk D. Borne and other astrophysicists submit to the Astro2010 Decadal Survey a paper titled “The Revolution in Astronomy Education: Data Science for the Masses “

- September 2010 Drew Conway writes in “The Data Science Venn Diagram”: “…I present the Data Science Venn Diagram… hacking skills, math and stats knowledge, and substantive expertise.”

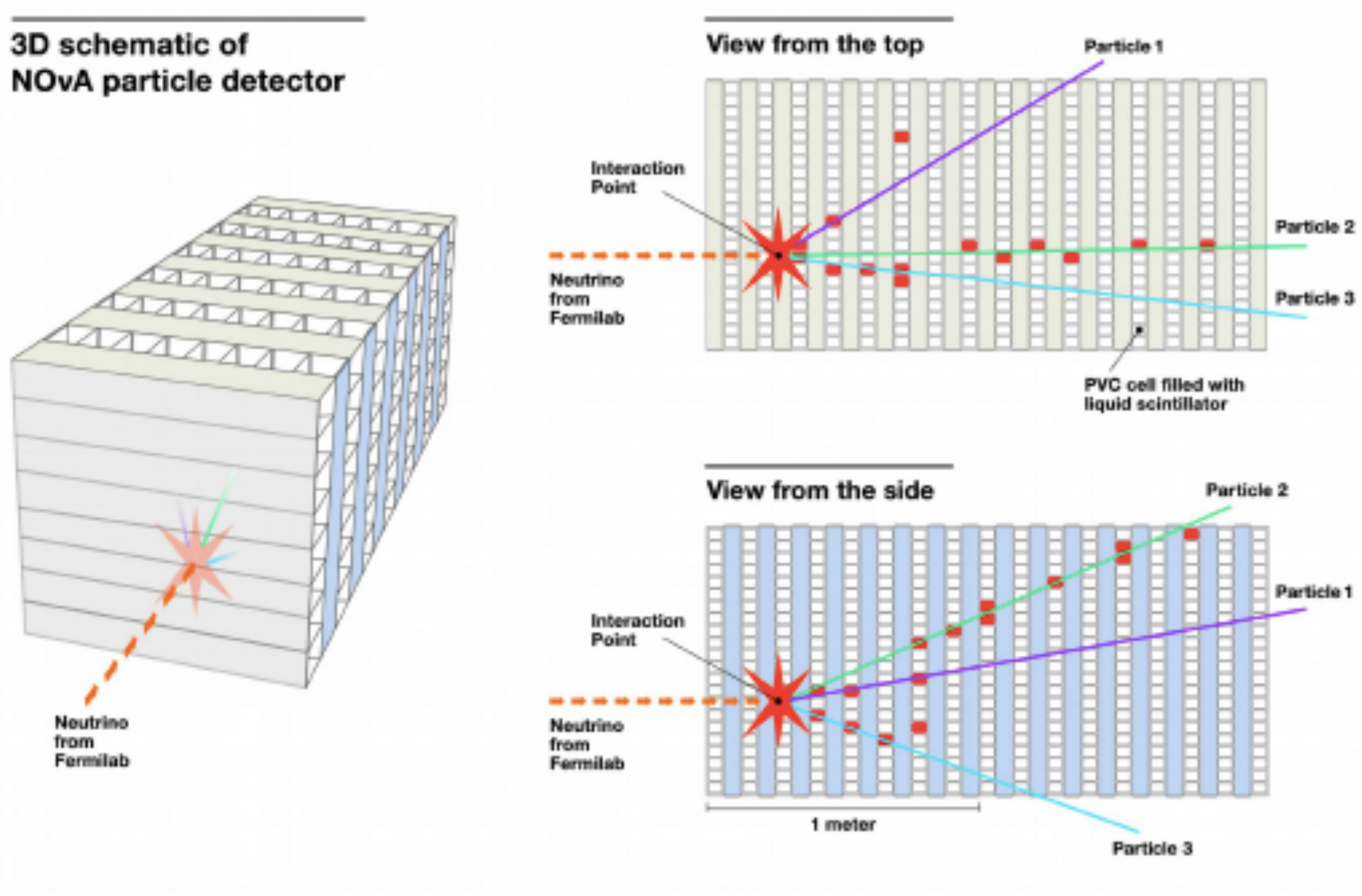


- May 2011 David Smith writes in “’Data Science’: What’s in a name?”: …. I think ‘Data Science’ better describes what we actually do: a combination of computer hacking, data analysis, and problem solving.”

DL MAJOR ++

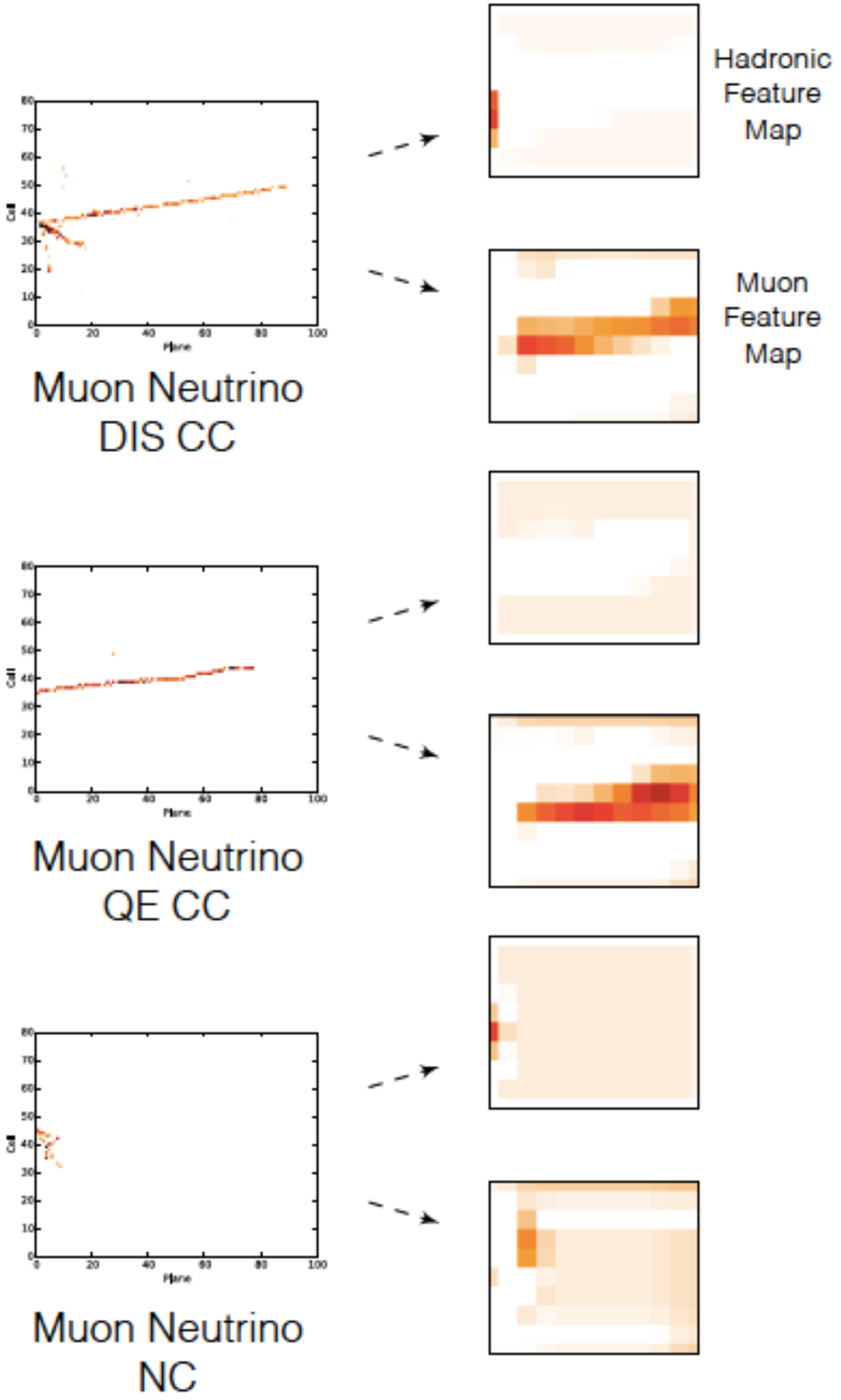
- DNN can encapsulate expensive computations (MEM or Simulations)
- DNN can be faster than traditional algorithm (after training)
- DNN already parallelized and optimized for GPU/HPCs
  - boost from industry building optimized chips, HPC systems, clouds etc
- DL solution to HEP data (e.g. HL-LHC) outpacing Moore's law
  - cannot assume we will get 10x computing power for same $ in 1 years
- DL could replace reconstruction difficulties (e.g. in LAr TPCs)

The NOvA Experiment



	CVN Selection Value	ν_e sig	Tot bkg	NC	ν_μ CC	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	88.4	509.0	344.8	132.1	32.1	–	14.8%
$s/\sqrt{b}$ opt	0.94	43.4	6.7	2.1	0.4	4.3	49.1%	86.6%
$s/\sqrt{s+b}$ opt	0.72	58.8	18.6	10.3	2.1	6.1	66.4%	76.0%

	CVN Selection Value	ν_μ sig	Tot bkg	NC	Appeared ν_e	Beam ν_e	Signal Efficiency	Purity
Contained Events	–	355.5	1269.8	1099.7	135.7	34.4	–	21.9%
$s/\sqrt{b}$ opt	0.99	61.8	0.1	0.1	0.0	0.0	17.4%	99.9%
$s/\sqrt{s+b}$ opt	0.45	206.8	7.6	6.8	0.7	0.1	58.2%	96.4%



- 40% Better Electron Efficiency for same background.

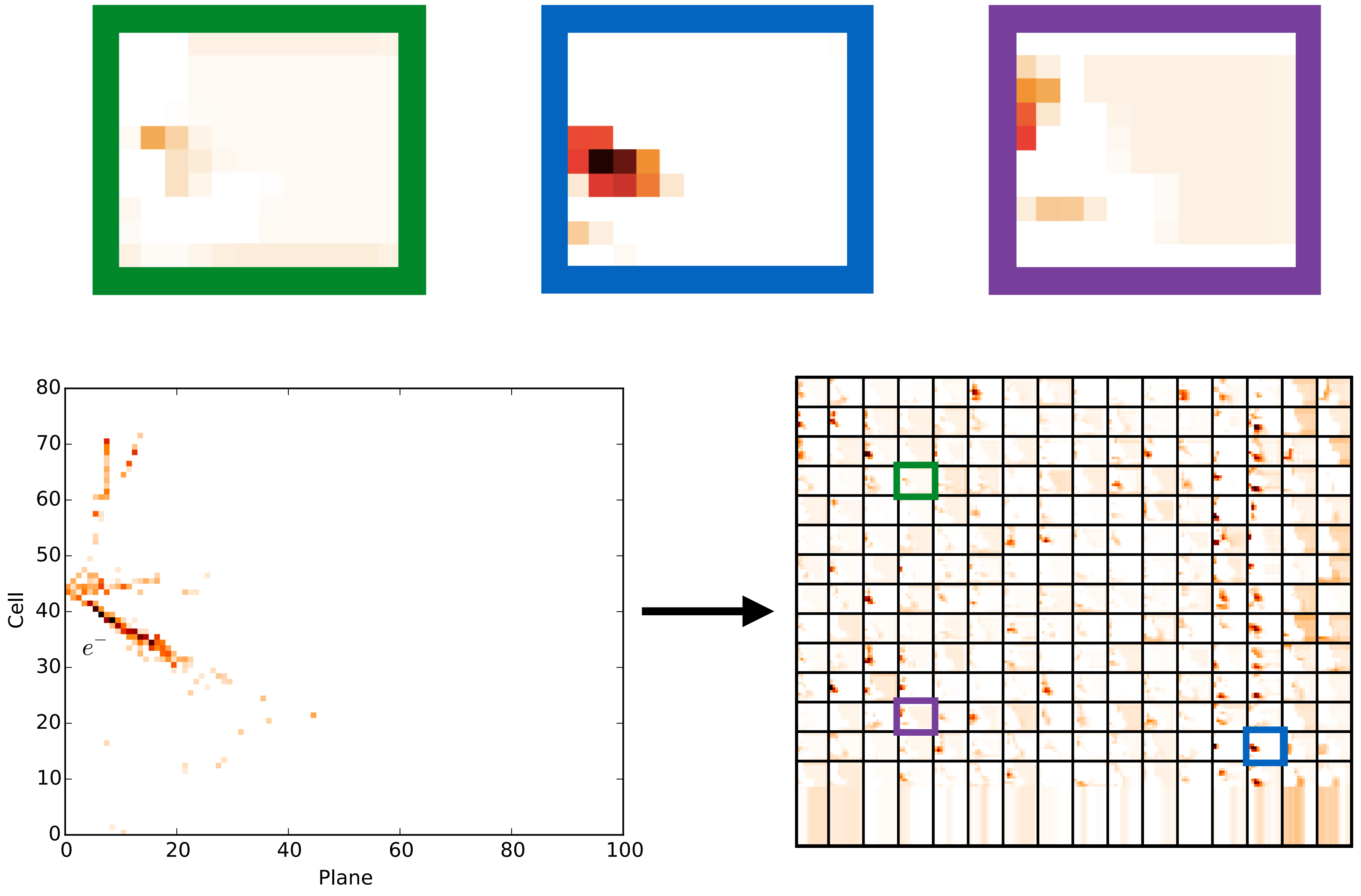
http://arxiv.org/pdf/1604.01444.pdf

DL MADE IT IN DIRECTOR'S SLIDES

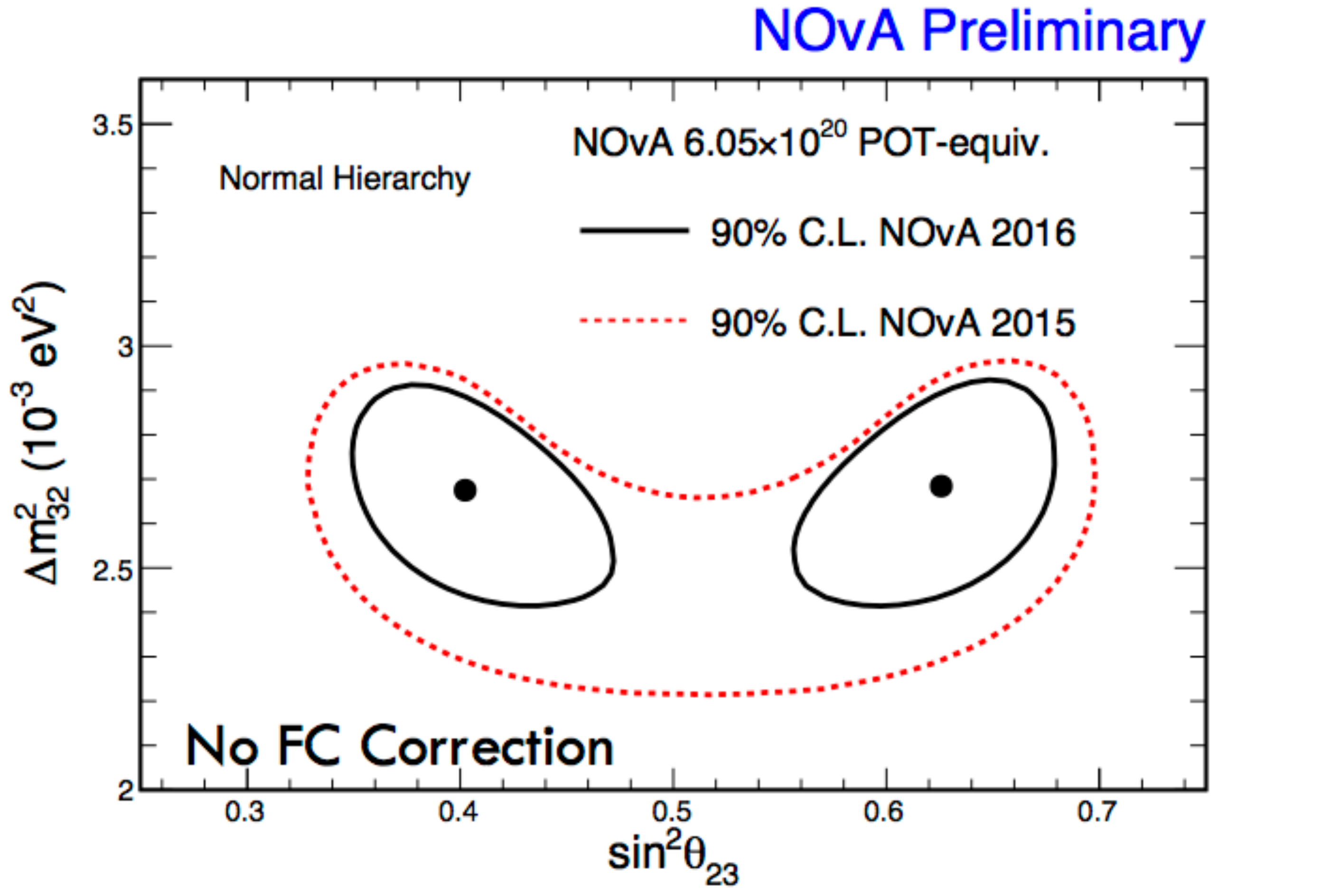
Improved Event Selection

- This analysis features a new event selection technique based on ideas from computer vision and deep learning

- Calibrated hit maps are inputs to Convolutional Visual Network (CVN)
- Series of image processing transformations applied to extract abstract features
- Extracted features used as inputs to a conventional neural network to classify the event



Improvement in sensitivity from CVN equivalent to 30% more exposure



Best Fit (in NH):

$$|\Delta m^2_{32}| = 2.67 \pm 0.12 \times 10^{-3}\mathrm{eV}^2$$

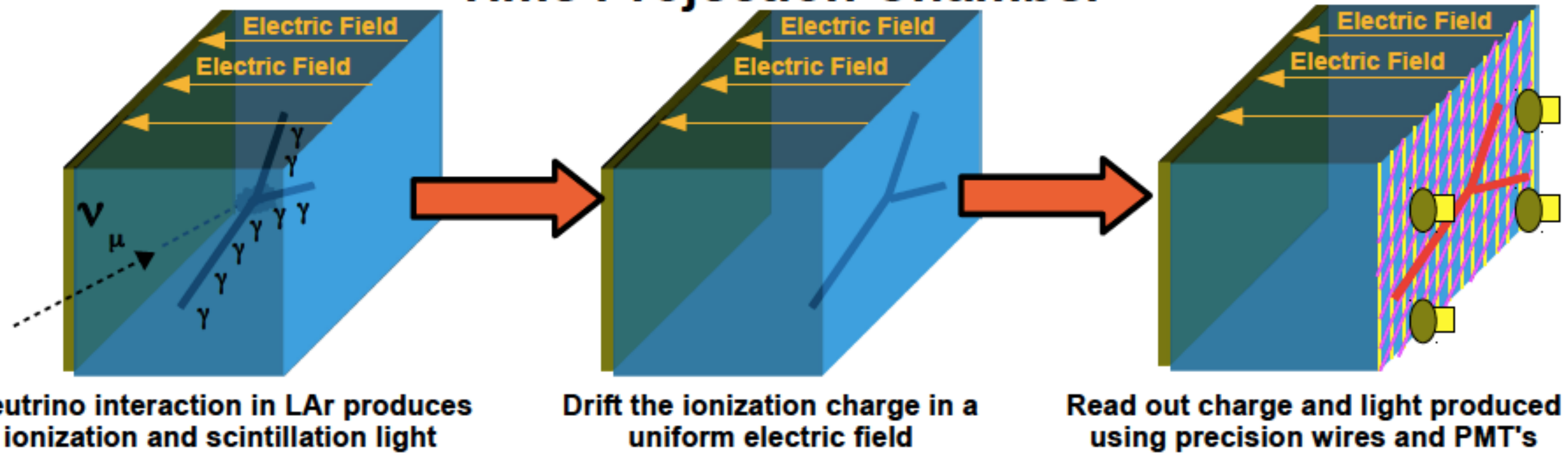
$$\sin^2\theta_{23} = 0.40^{+0.03}_{-0.02}(0.63^{+0.02}_{-0.03})$$

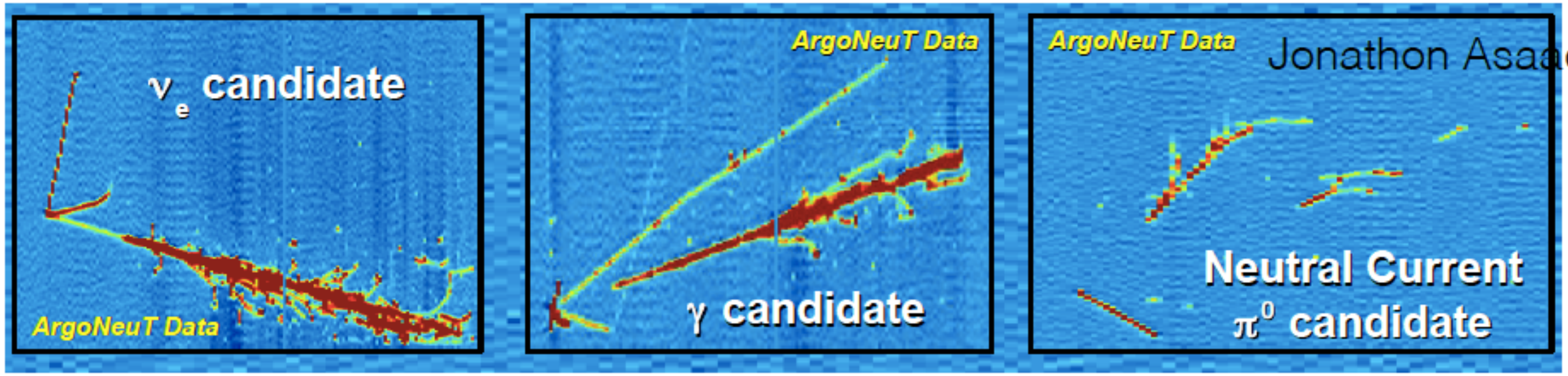
maximal mixing excluded at 2.5 σ

ArgoNeuT: Mini LArTPC Exposure to Fermilab's NuMI Beam

LArTPC

Time Projection Chamber



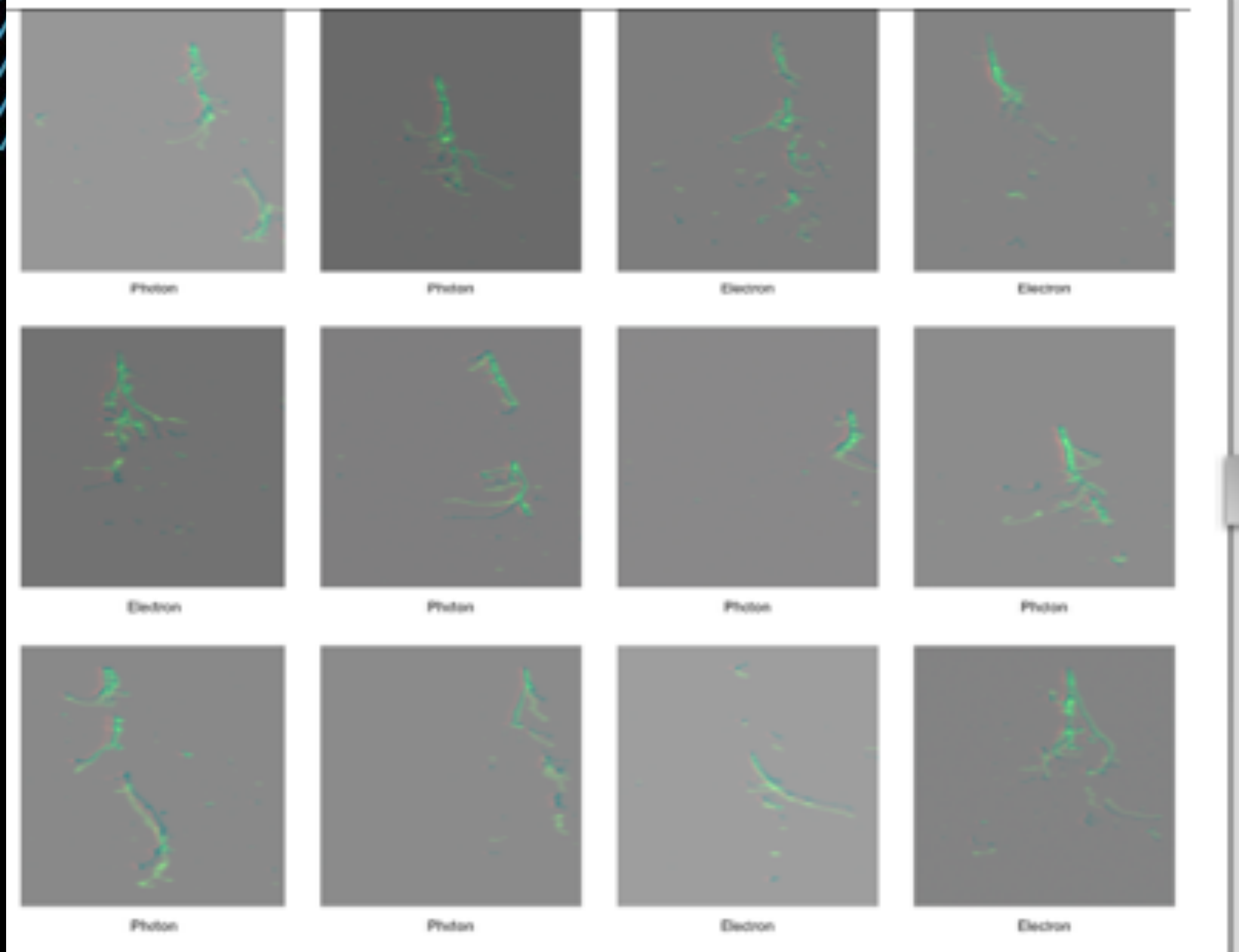


Jonathon Asaadi

Tracking, Calorimetry, and Particle ID in same detector.
Goal ~80% Neutrino Efficiency.
All you need for Physics is neutrino flavor and energy.

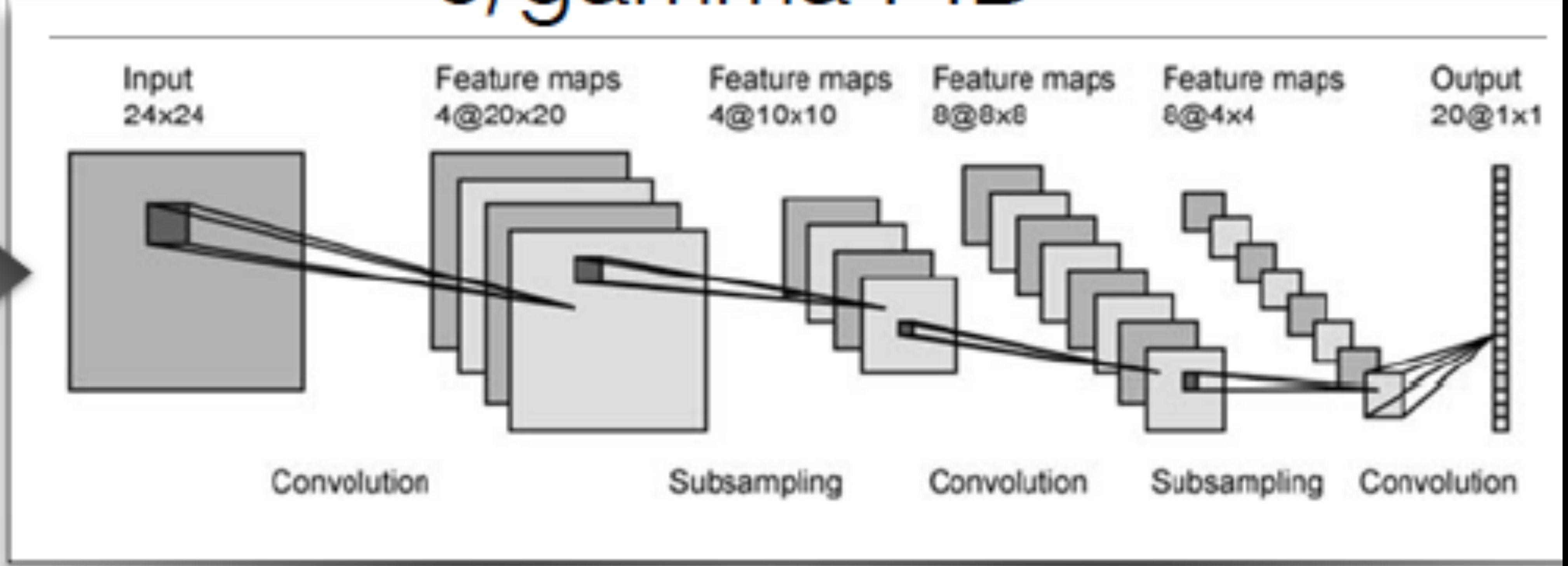
LArIAT

e/gamma PID



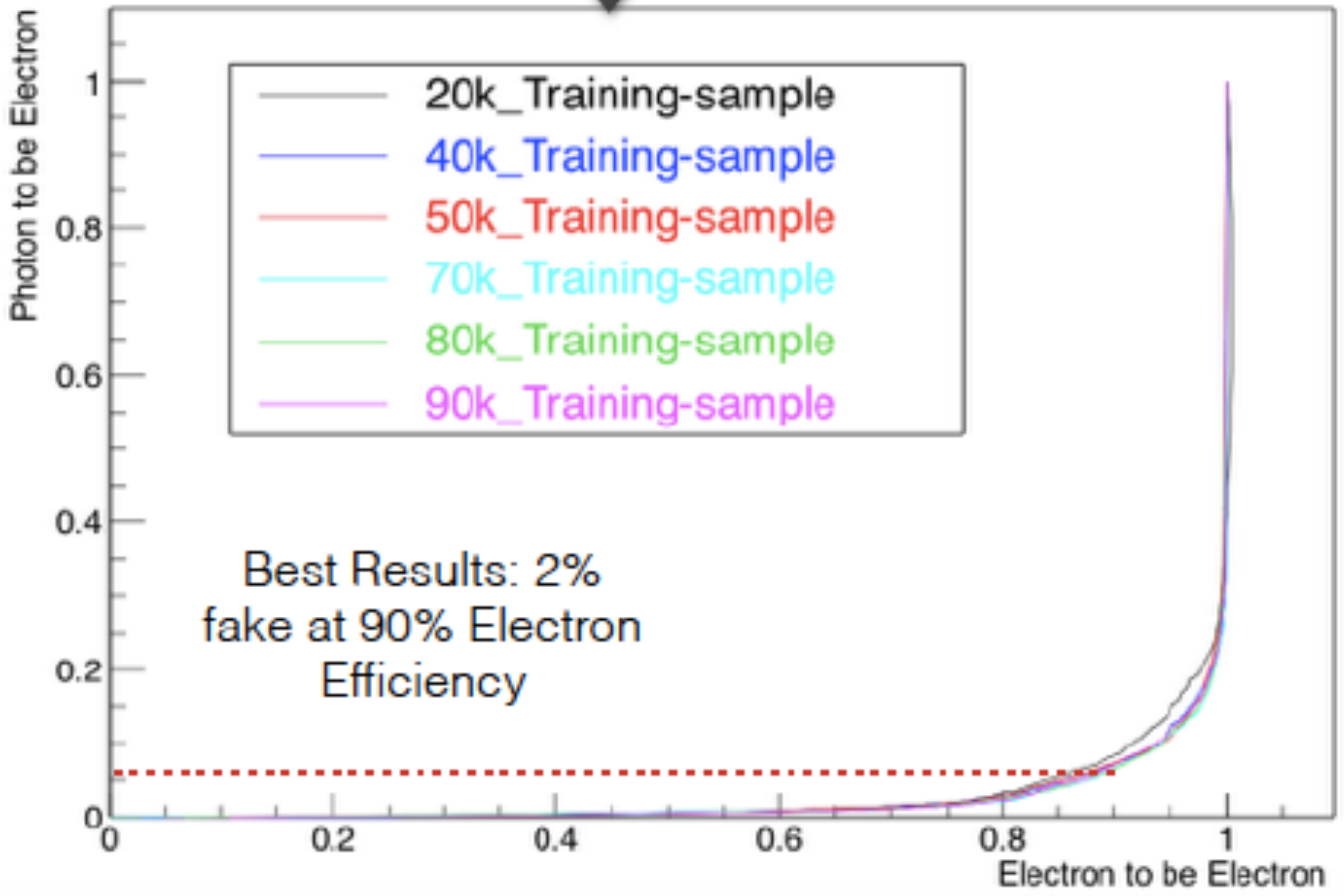
Raw Data: Wire ADC vs Time x Planes
(LArIAT Simulation)

Train



Deep Convolutional Neutral Network
(GoogLeNet)

Out of the box *Feasibility Study* with No attempt at optimization.



- First results with neutrinos:
  - 5% NC at 80% CC
  - 15% Muon CC at 80% Electron CC
- Regression working on Neutrino Energy
- DL efforts present also in other LArTPC experiments (not yet public).
- May be easy and ideal tool for Detector Optimization.

ArgoNeuT

MicroBooNE-NOTE-1019-PUB
Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber

MicroBooNE Collaboration

July 4, 2016

Abstract

We present several studies of convolution neural networks applied to data coming from the MicroBooNE detector, a liquid argon time projection chamber (LArTPC). The algorithms studied include the classification of single particle images, the localization of single particle and neutrino interactions in an image, and the detection of a simulated neutrino event overlaid with cosmic ray backgrounds taken from real detector data. The purpose of these studies was to demonstrate the potential of these networks for particle identification or event detection with simulated neutrino interactions and to address a number of technical issues that arise when applying this technique on data from a large LArTPC located near the surface. The results of these studies can be used to guide similar applications on detector neutrino data. We developed and validated techniques and approaches that demonstrate successful application of these networks for particle identification or event detection on simulated data and can be used to guide similar application on detector data.

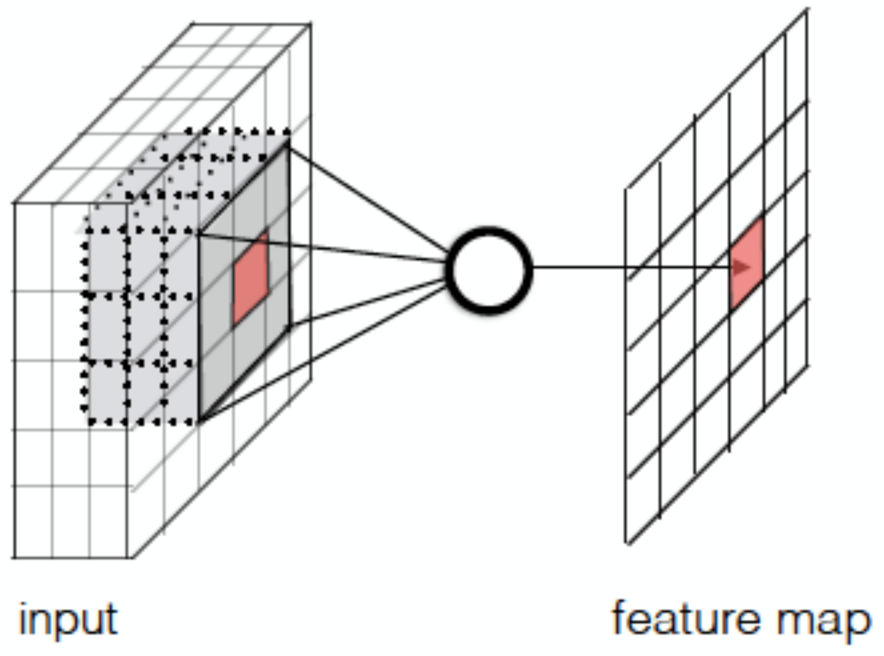
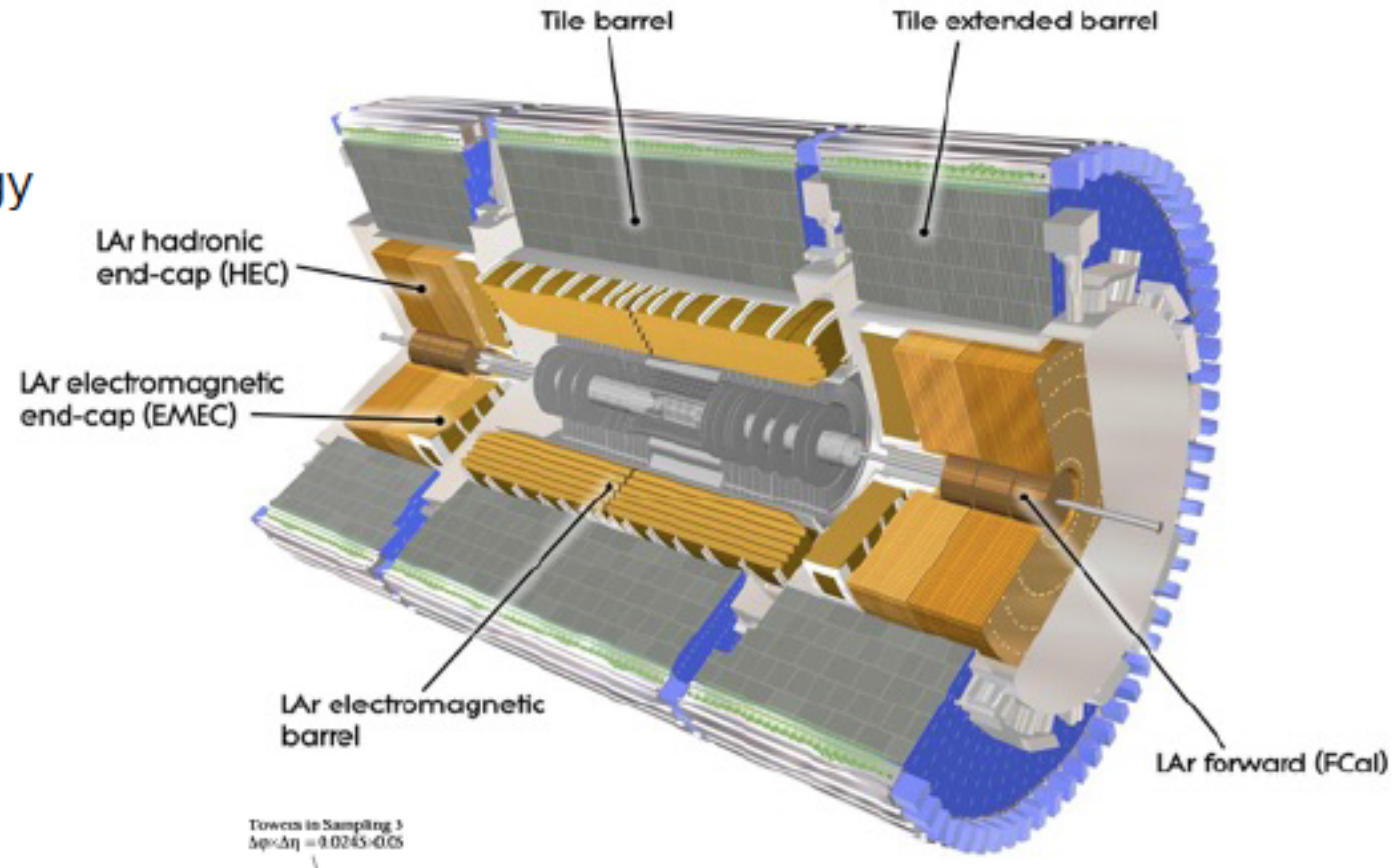
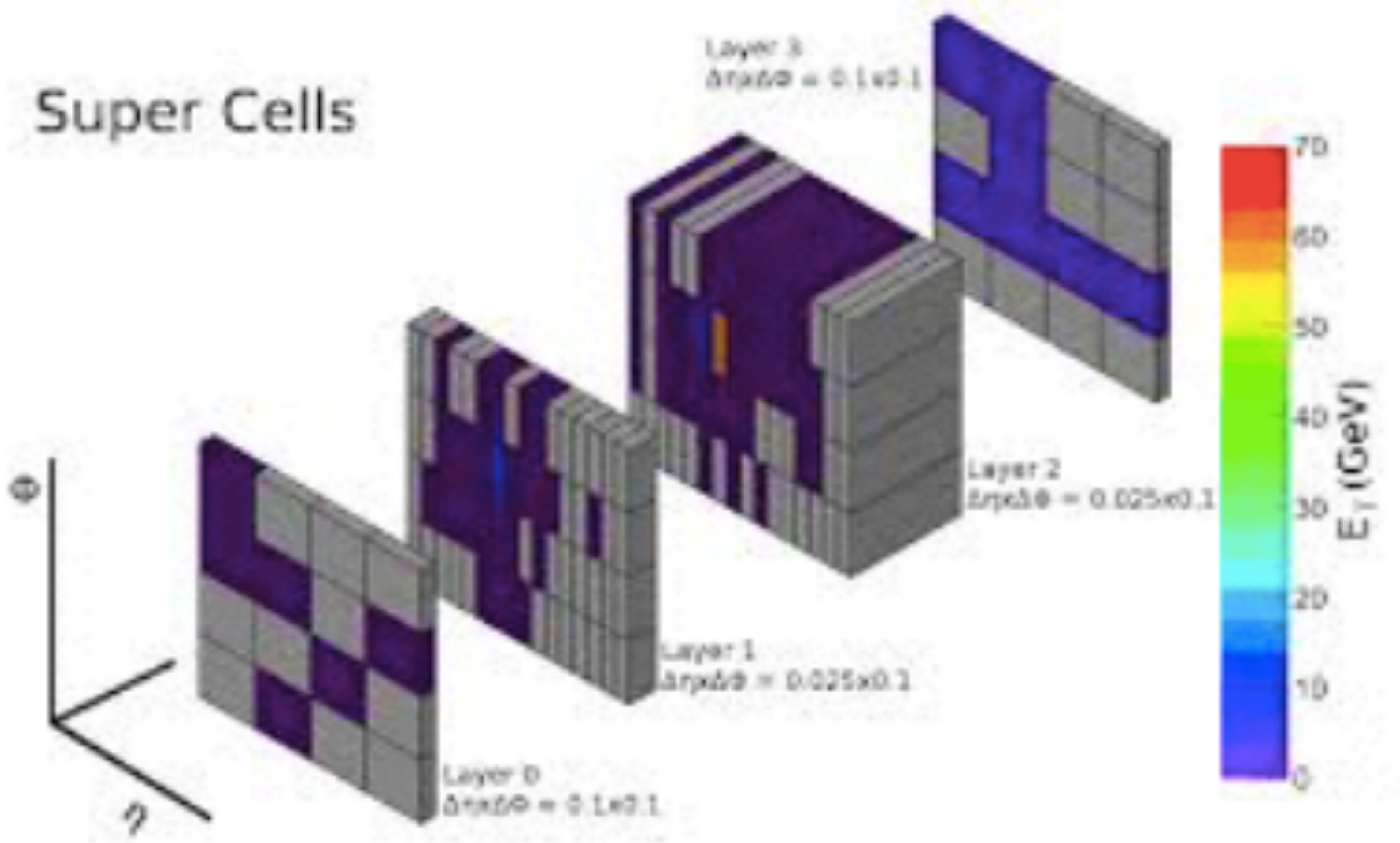


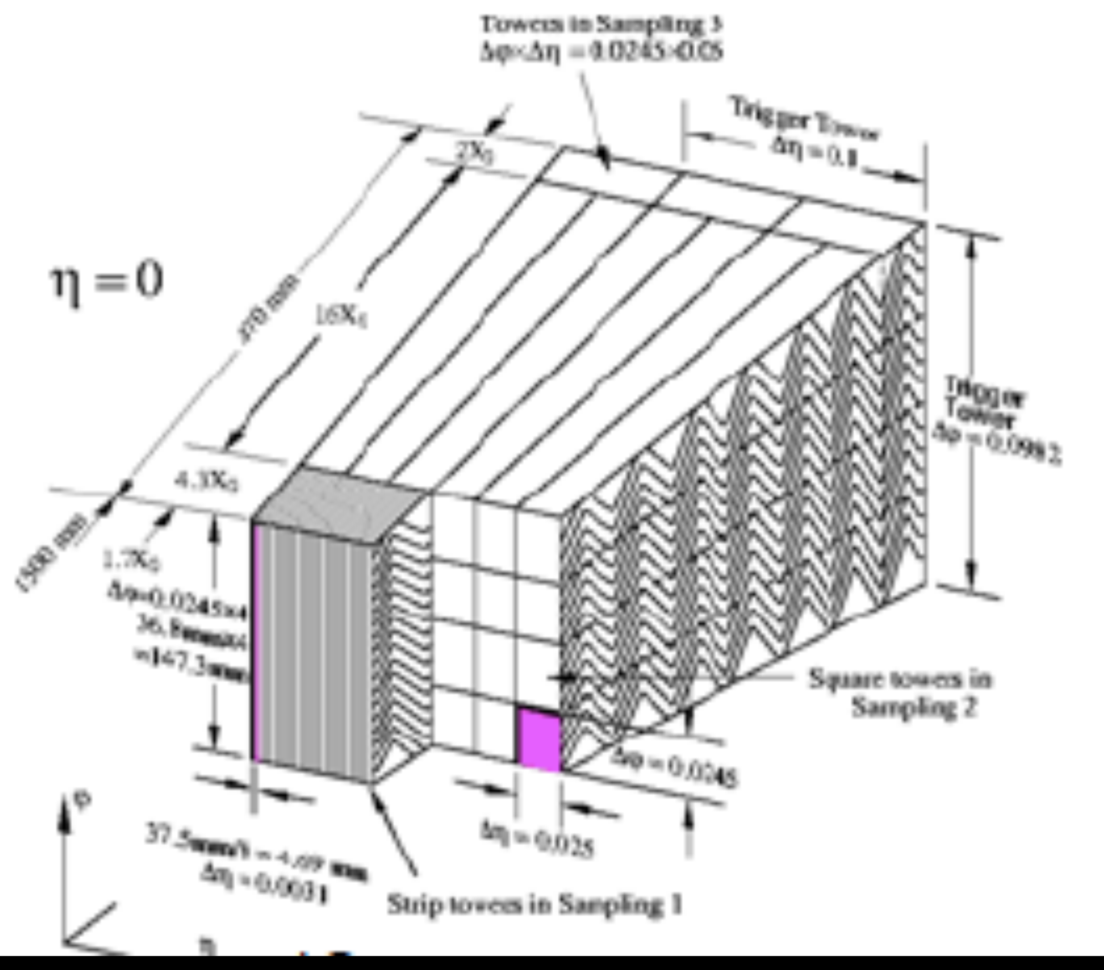
Figure 3: The input and output connections of one neuron in a convolutional neural network (CNN). The neuron looks at a local volume of pixel values arranged in an $M \times N \times L$ tensor. The input consists of a $M \times N$ image with L color channels, represented by the 3D grid on the left. In a CNN, one such neuron acts only on one sub-region at a time but, in turn, it acts on all sub-regions of the input as it is scanned over the height and width dimensions. By arranging the outputs of the neuron as it is being scanned across into a 2D grid, one forms a feature map, represented by the right most grid.

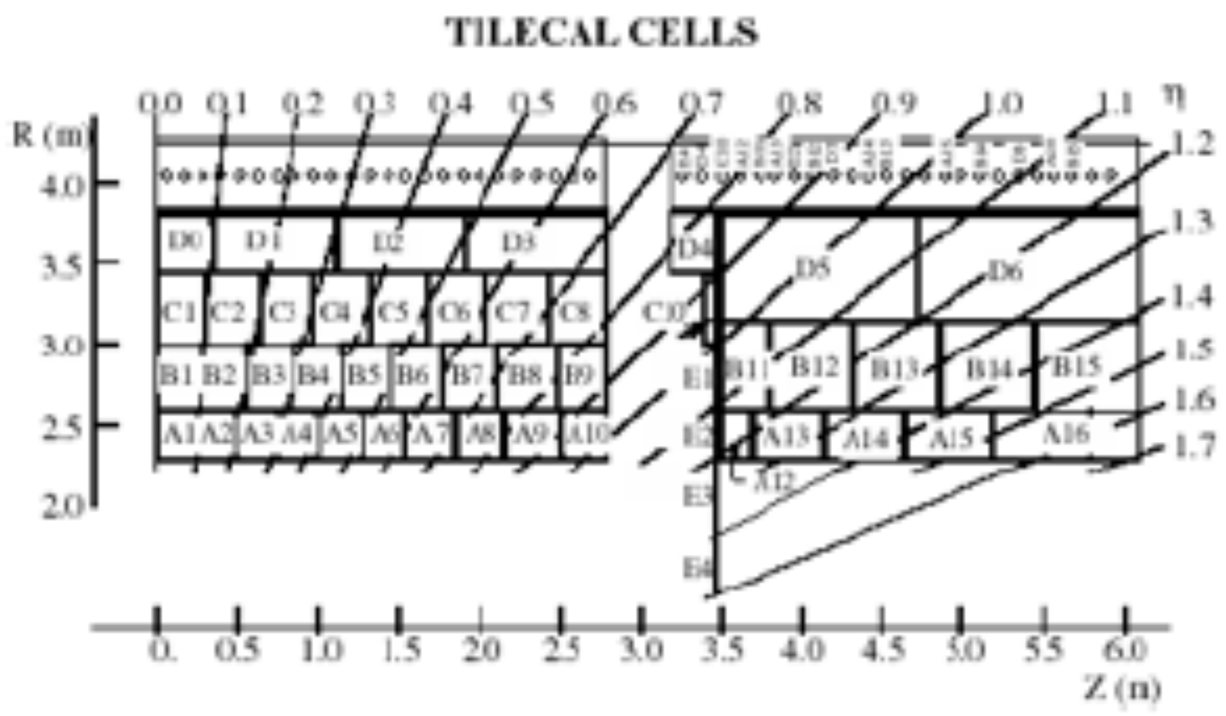
ATLAS CALORIMETER

- Ideally suited for “imaging”
  - Electromagnetic- Highly transverse and longitudinal segmented.
  - Hadronic- Longitudinal sampling
  - 200K Calorimeter cells measure energy deposits.
  - ~ 64 x 36 x 7 3D Image
  - Interesting Challenges: non-uniform granularity, cylindrical geometry.

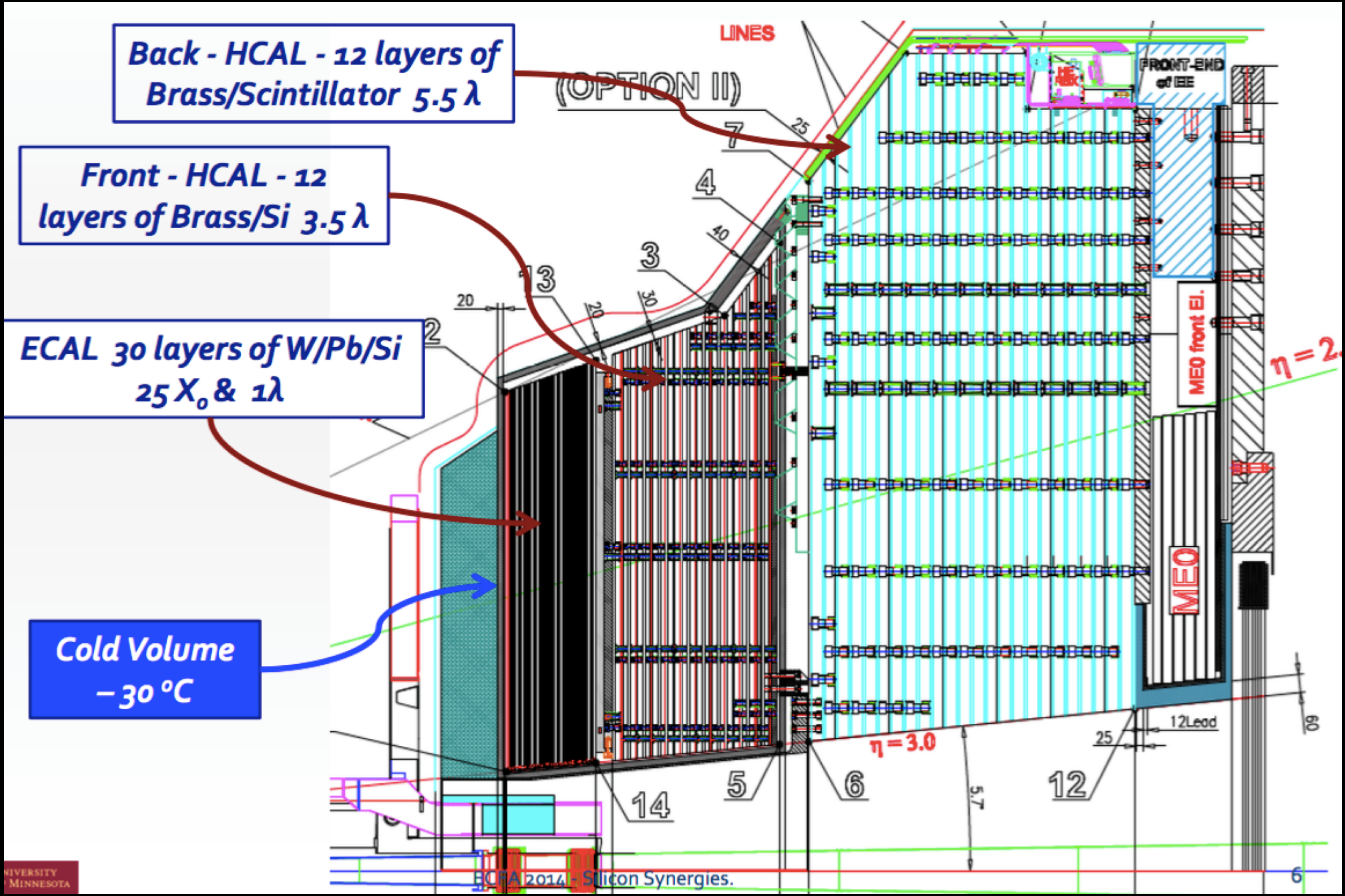






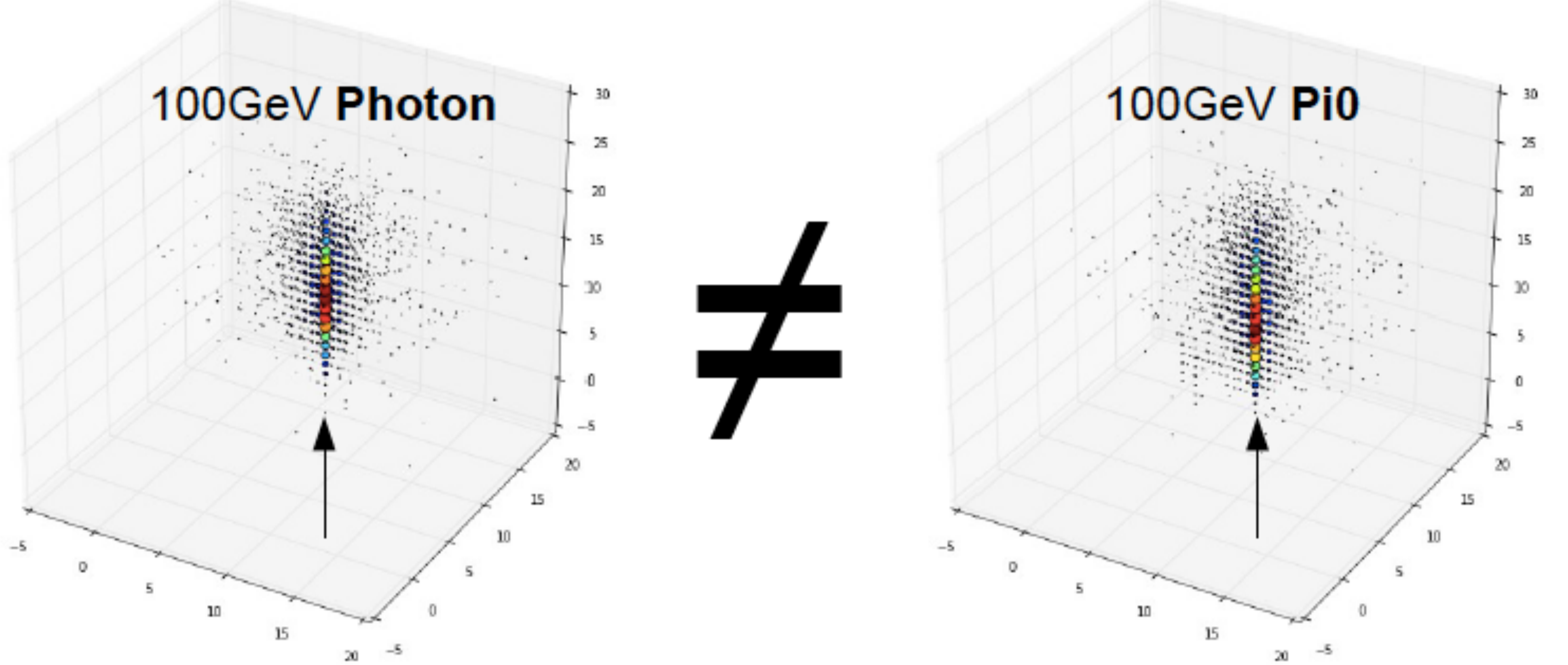


CMS HGC CALORIMETER (UPGRADE)

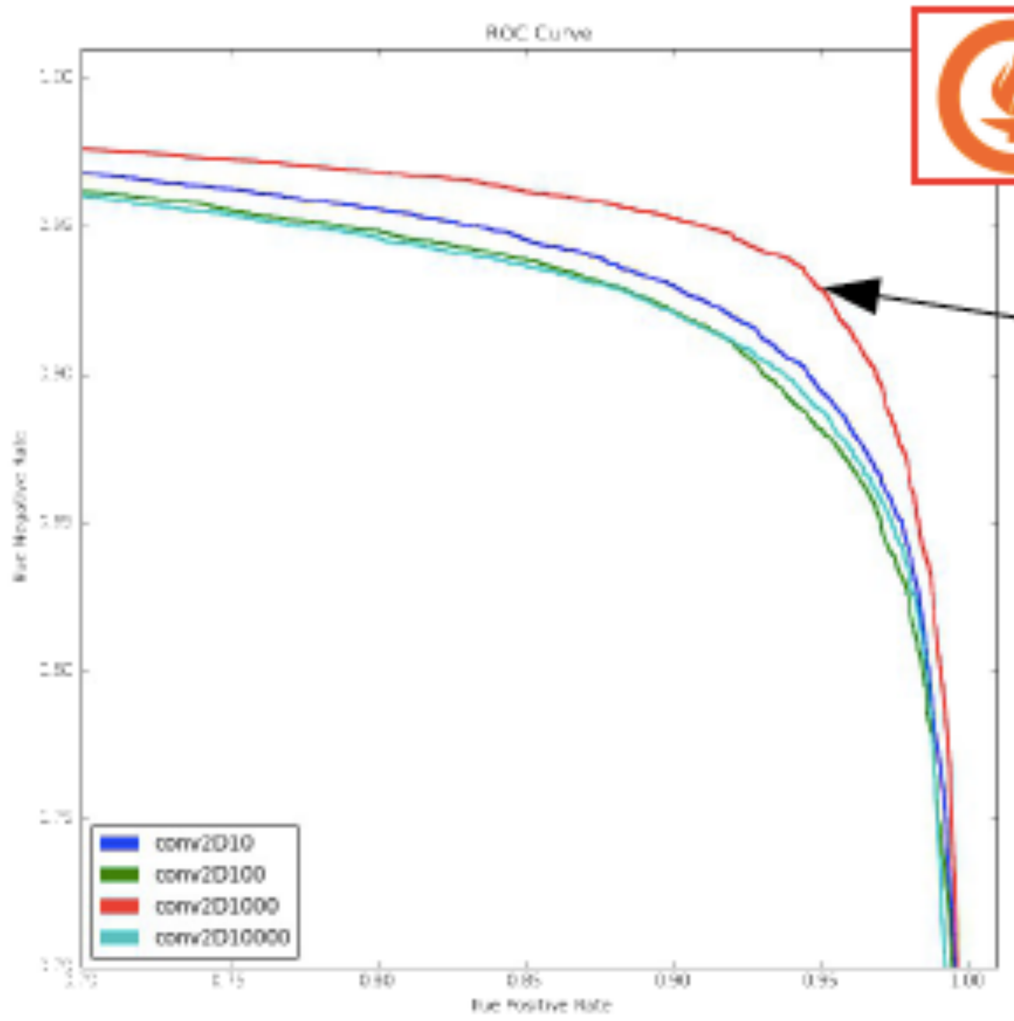


LCD CALORIMETER PILOT (SURF)

3D Calorimeter Imaging



LCD Calorimeter configuration
http://lcd.web.cern.ch
5x5 mm Pixel calorimeter
28 layer deep
Photon and pion particle gun

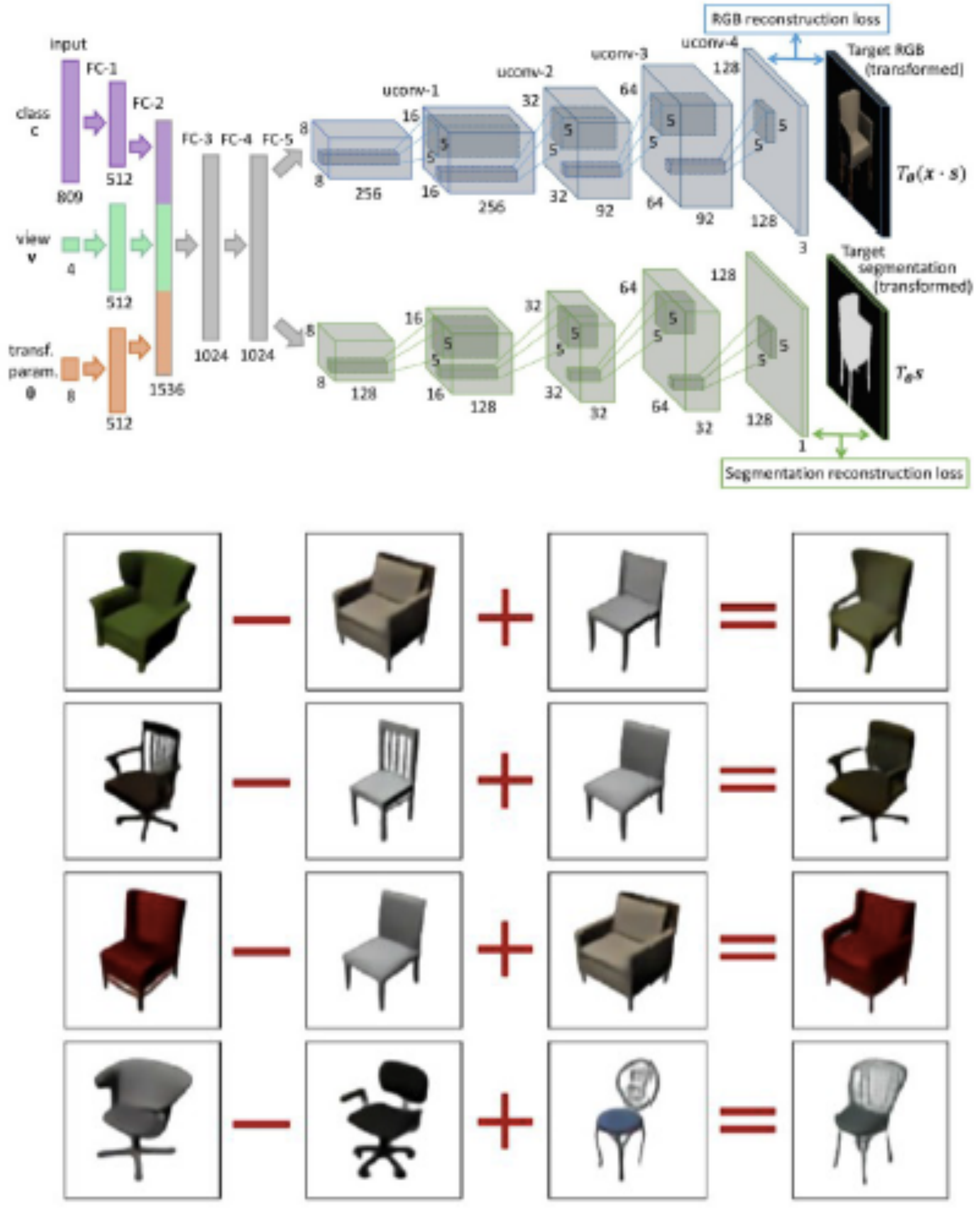


2D x 25 channels Convolutional NN
+ dense layers
Limited dataset
95% efficiency
5% fake

& SUPERFAST SIMULATION

Generative Models

Arxiv:1411:5928, Dosovitskiy, Springenberg, Tatarchenko, Brox

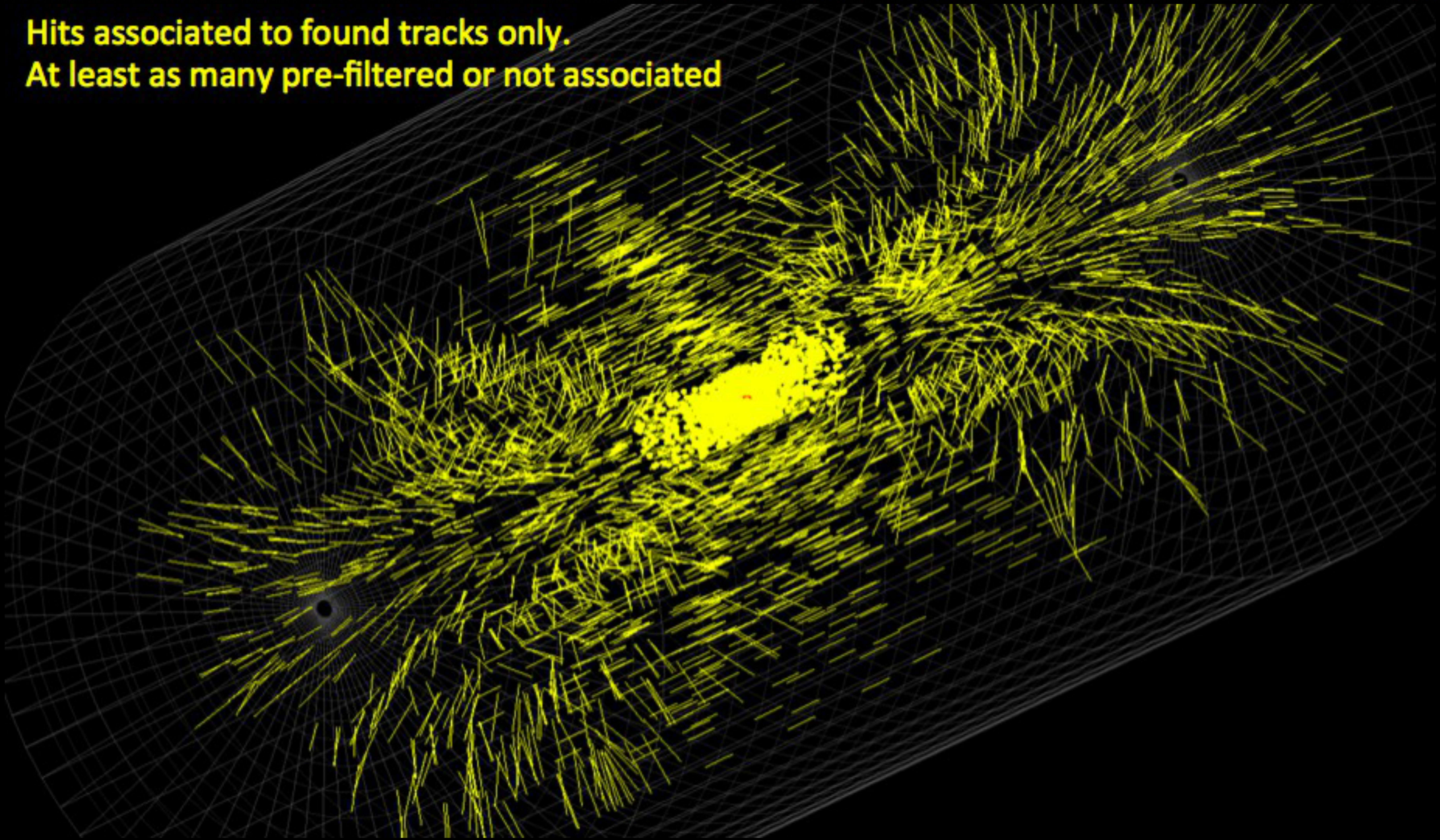


- Computer science can generate images, text, sound
- Performing image arithmetic
- Simulation of collision events is very **computation intensive**
- **Faster simulation** with such generative models
- Address computing bottleneck
- Enable science program

DL RECO IN IMAGING CALORIMETERS

- Improved classification/regression with Convolutional NNs
- Fast Showers with Generative models
- Feature (particle) extraction with Regional NN and semantic segmentation
- Full event classification (RN, RT)
- ++Detector optimization !!

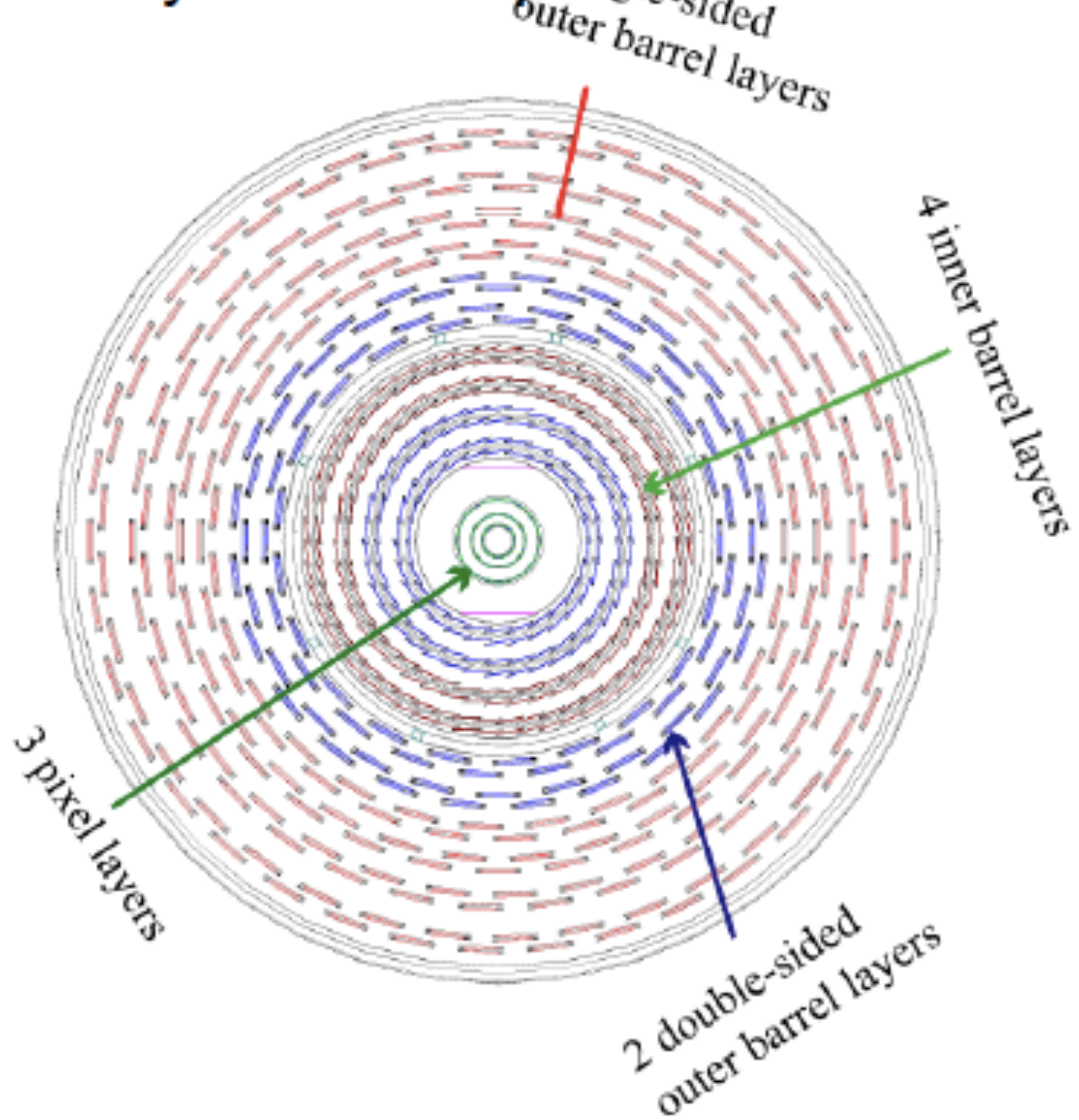
TRACKING/VERTEXING

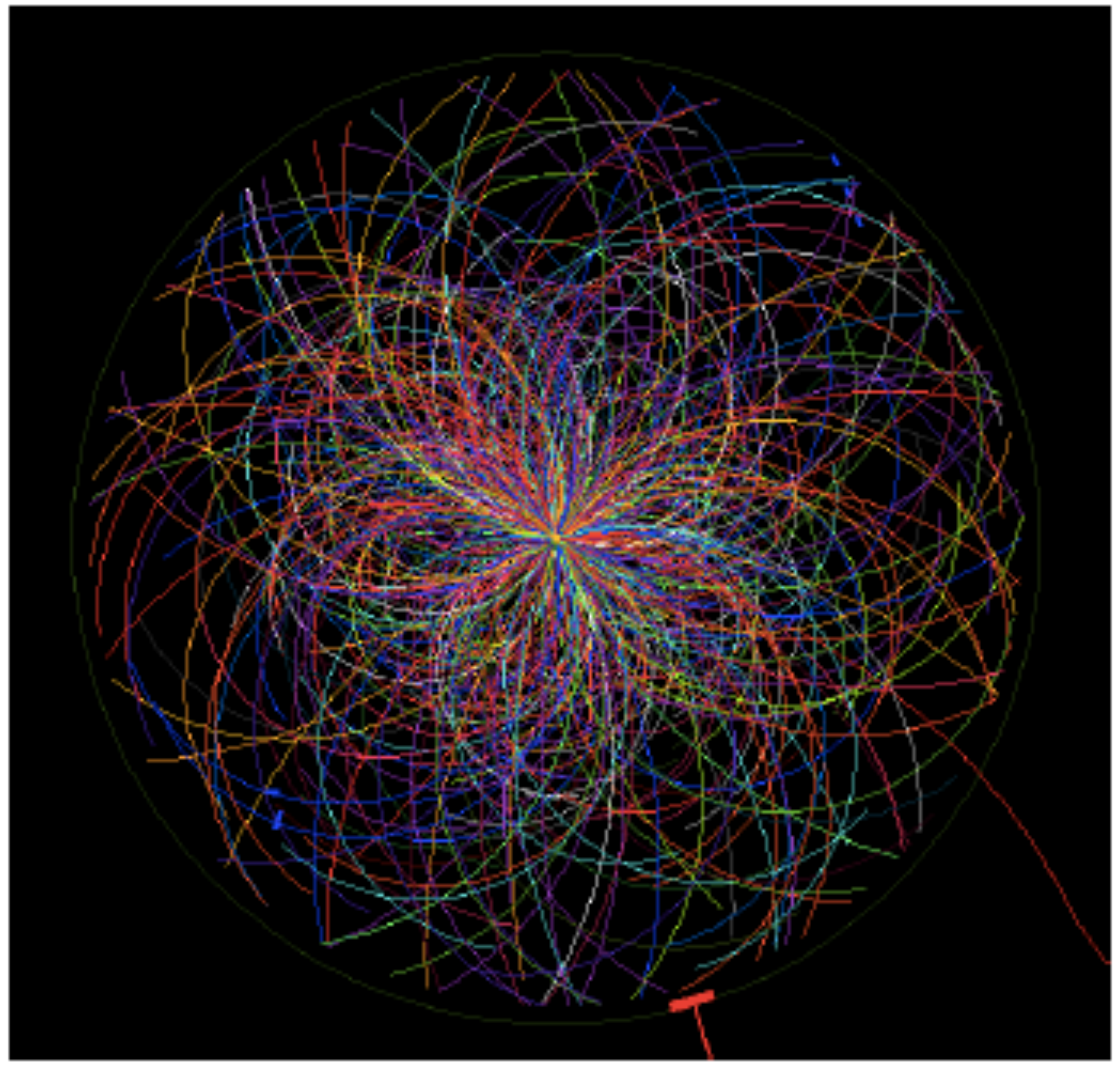


TRACKING

Tracks Pattern Recognition

- From sparse 2D/3D points reconstruct the path of a charged particle
- Iterative process using combinatorics, Kalman Fitting and Filtering
- Most CPU intensive part of the event reconstruction (~10s /event)
- Computation time scales ~quadratically with number of interactions
- Any fraction of patterns that can identified faster will make a difference





EXAMPLE OF ML APPROACH

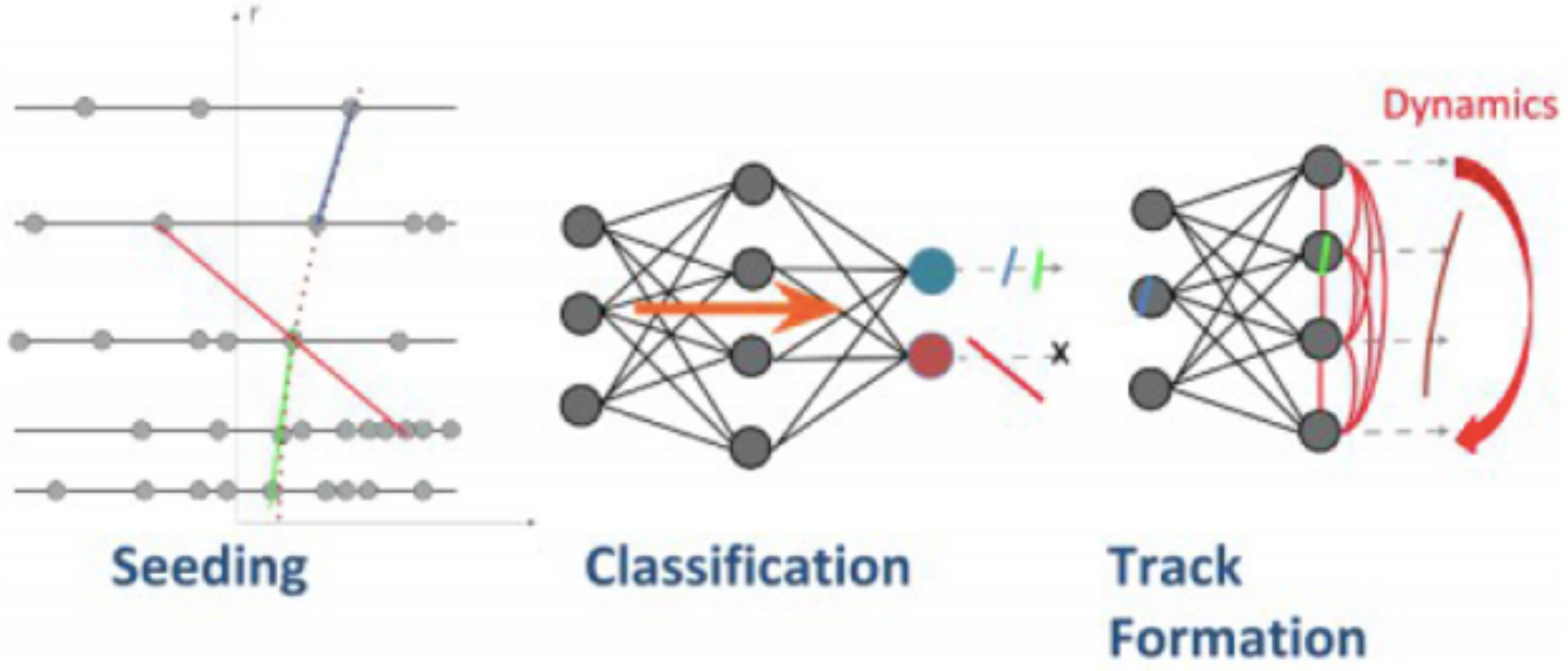
Scene Labelling

Farabet et al. ICML 2012, PAMI 2013

- **Group and classify what each pixel belongs to**
- **Real-time video processing with deep**

- **Attribute each Xtal to a cluster**
- **Attribute tracker hit to a charged particle**

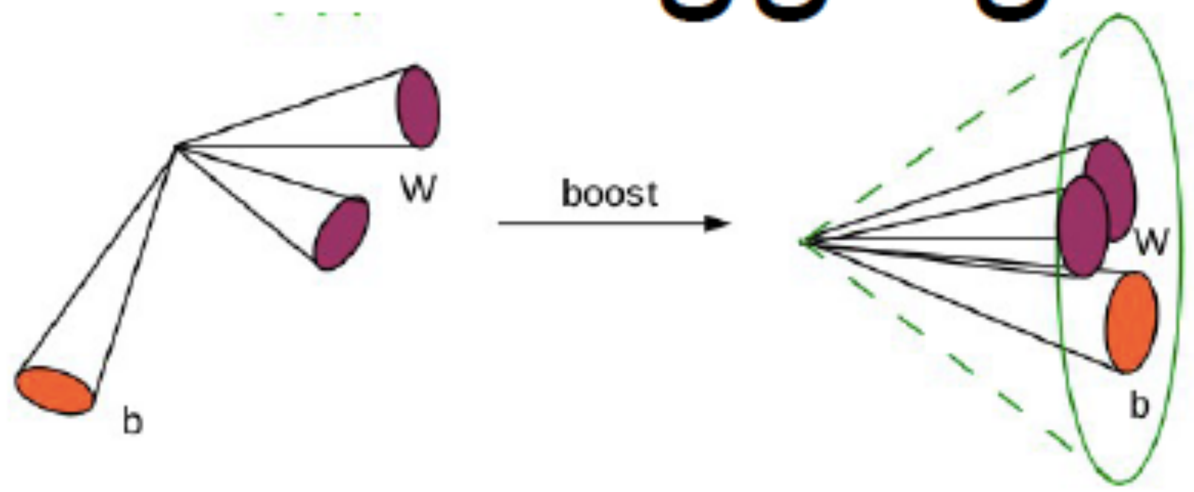
Advanced Tracking Algorithms



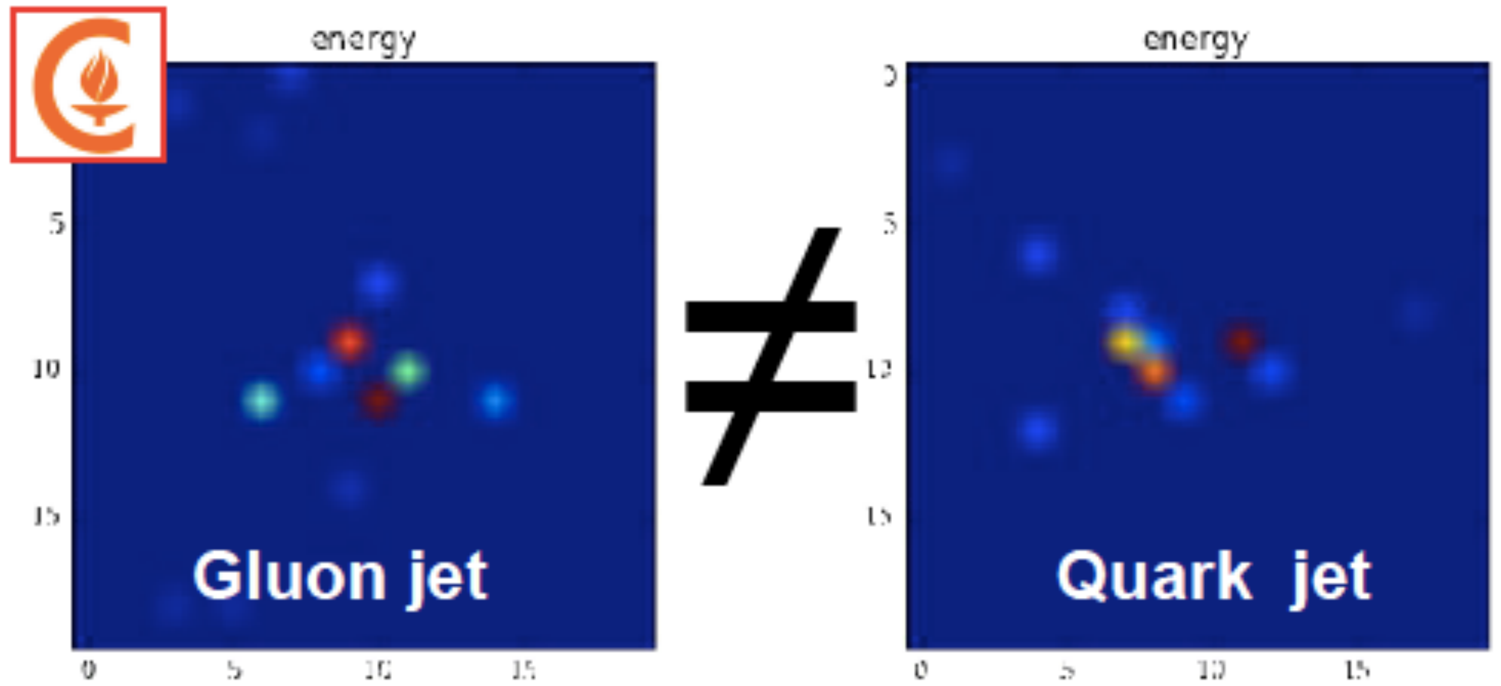
- **Investigation may involve**
  - Application of scene labeling to seed formation
  - Application of object detection to track assembling
  - More reference algorithms
- **Medium/High risk, very high reward problem**
  - Exploratory phase on the model definition

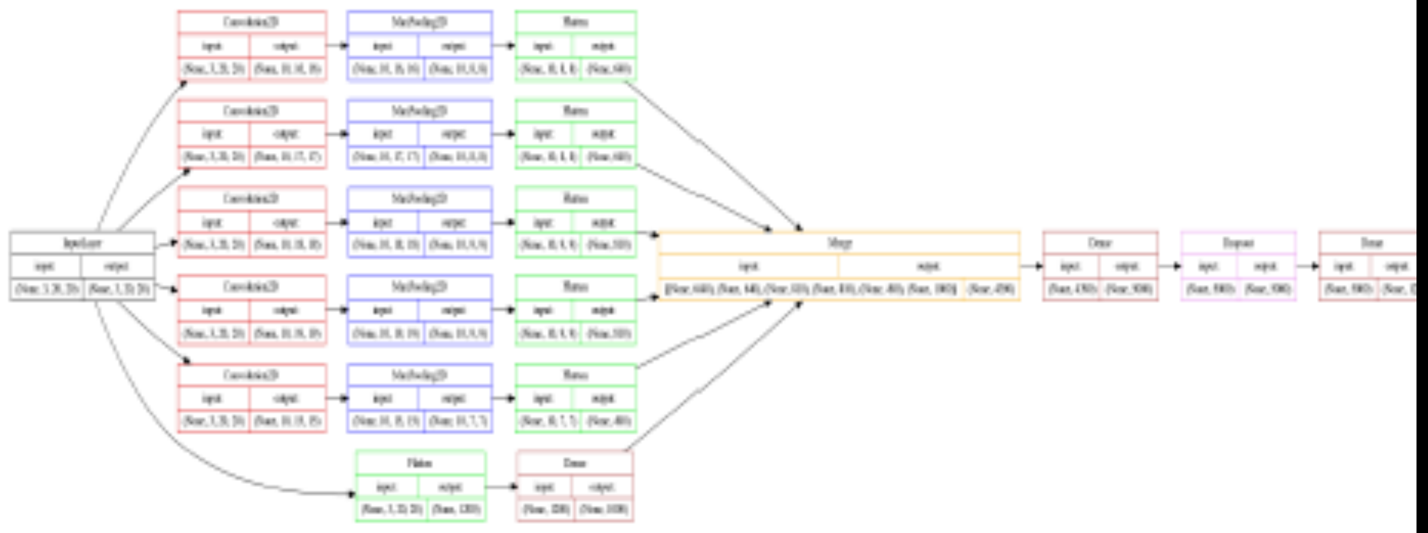
EXAMPLE OF ML APPROACH

Jet Tagging



- Hadronic activity results in bundle of collimated particles
- The more energetic, the more collimated : W-jet
- With even higher energy, even mother particles are collimated top-jet, Higgs-jet
- Available discriminators are performing well. Not taking advantage of the full substructure of the jets
- Image processing methods are natural candidates to perform the classification

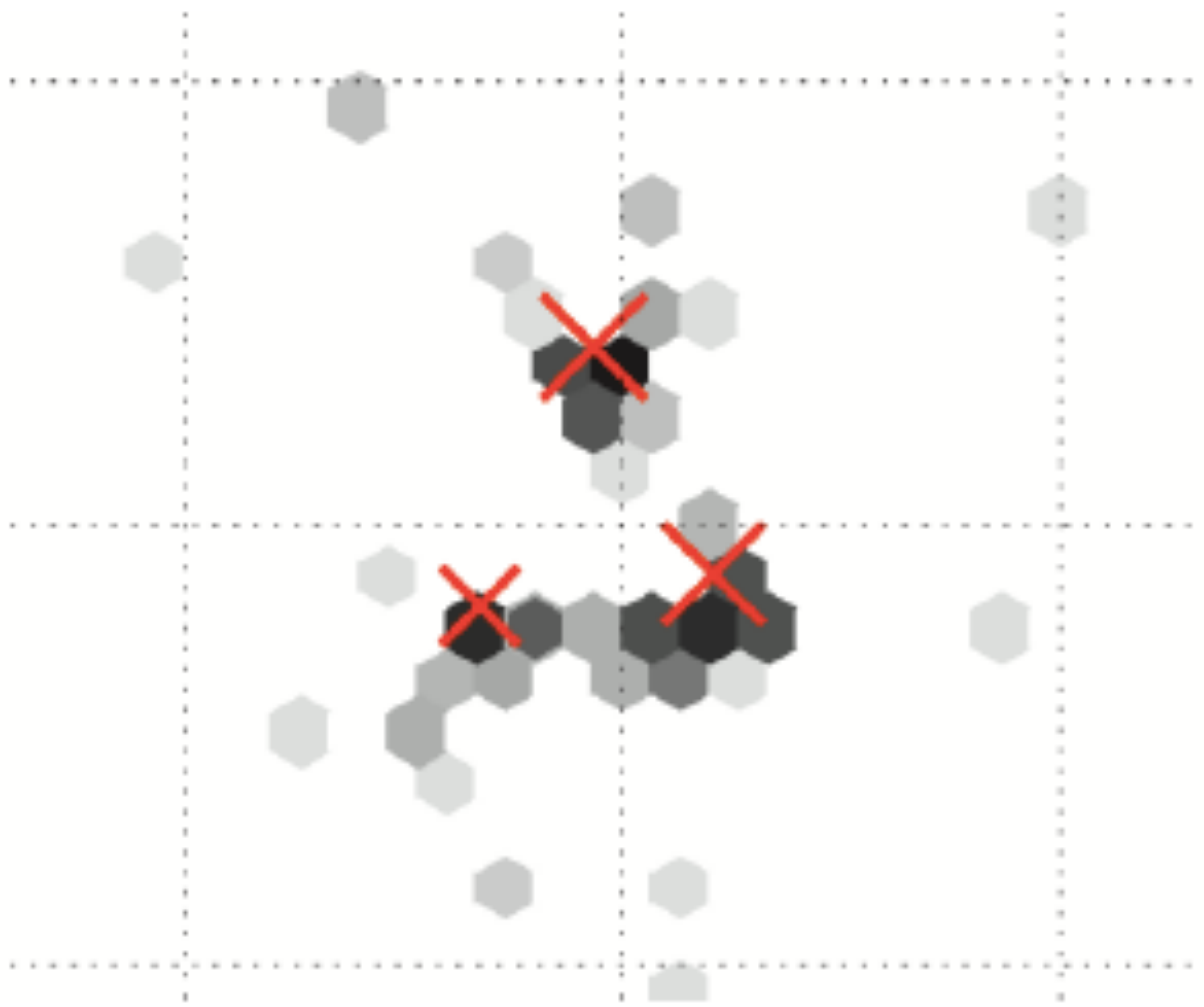




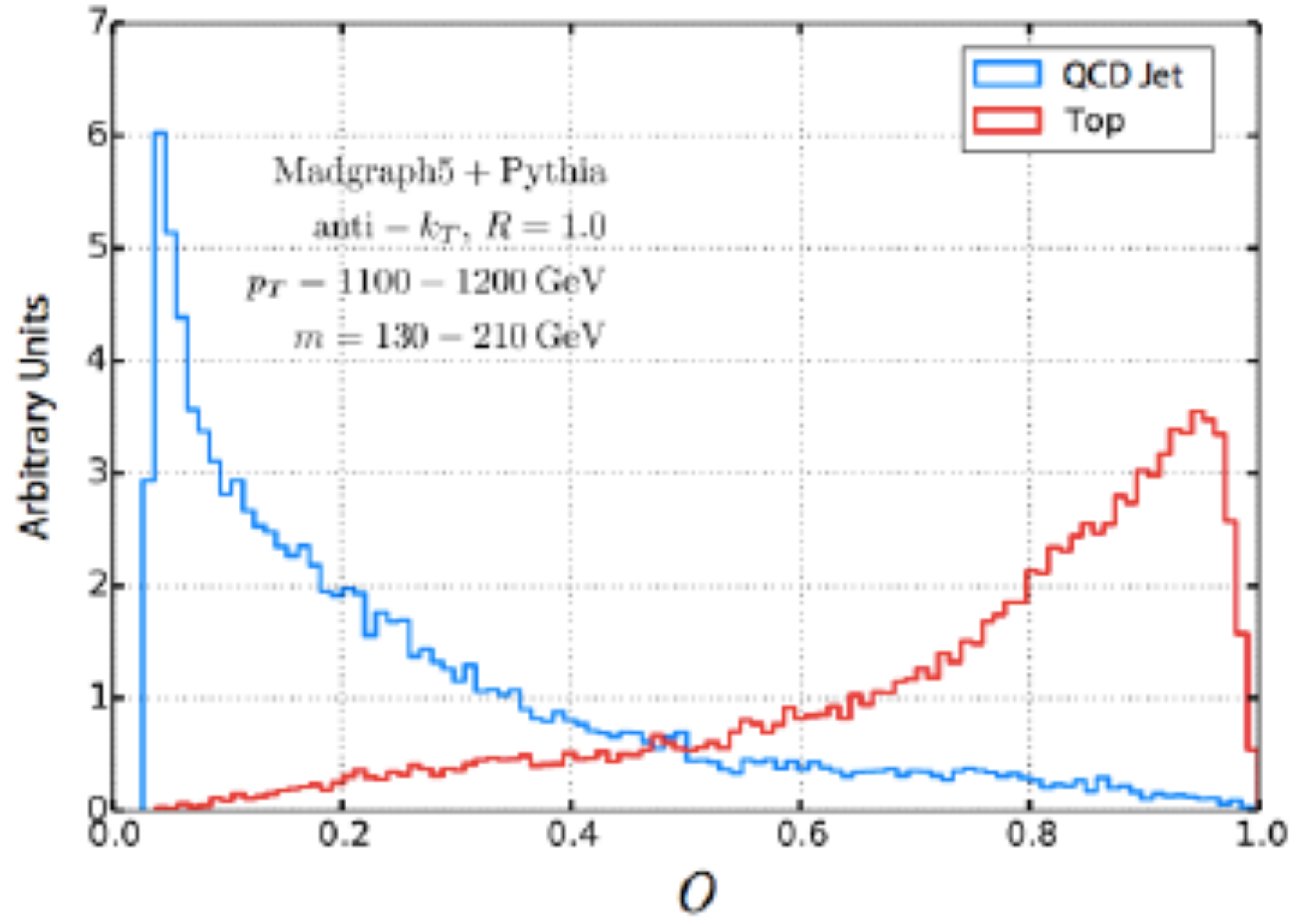
Small dataset, 11 categories
60% accuracy on gluon versus any quark. pre-preliminary

PHYSICS OBJECTS AND ANALYSIS EXAMPLES

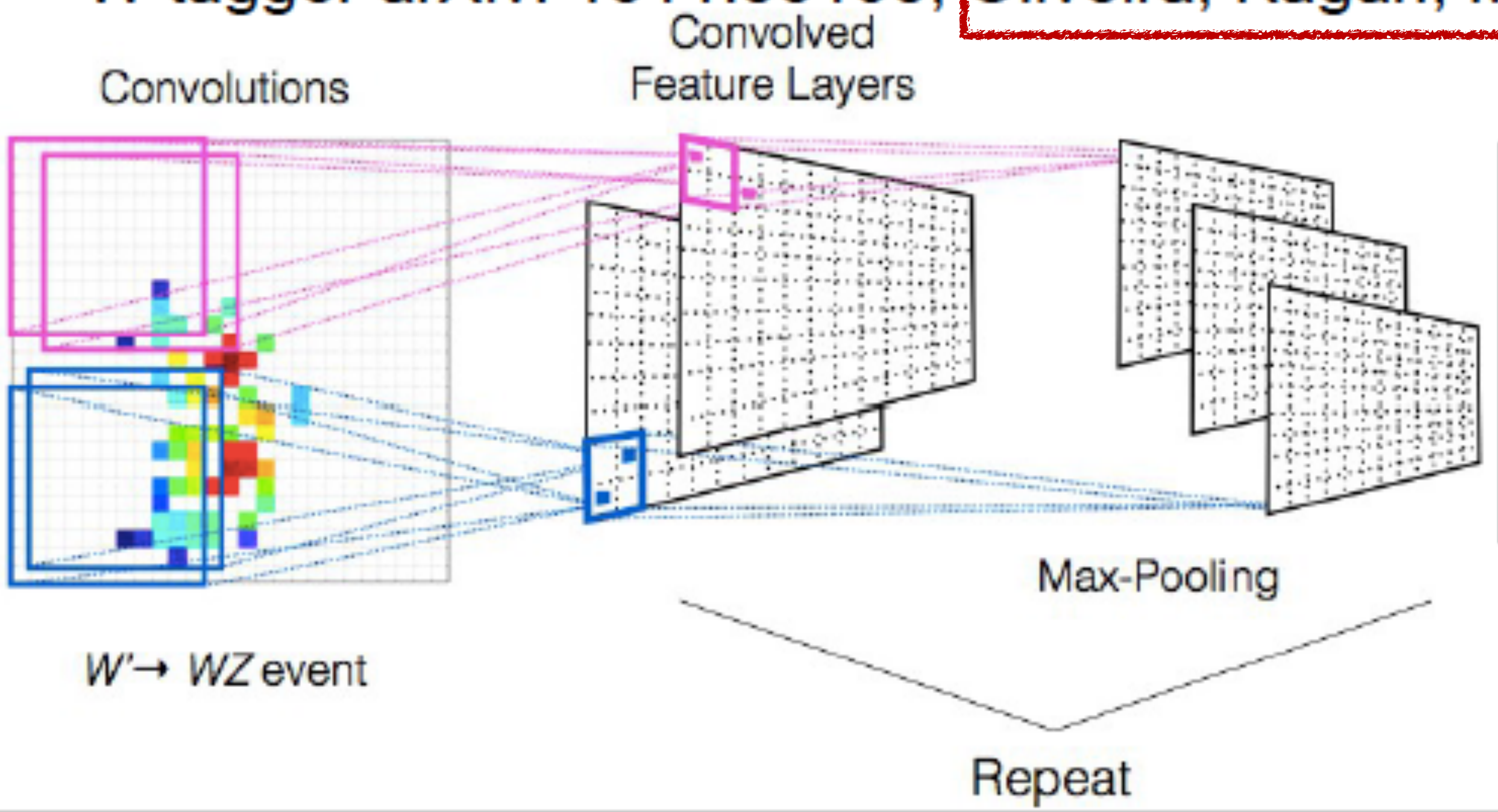
Top Tagger arXiv: 1501.05968 Almeida, Backovic, Cliche, Lee, Perelstein



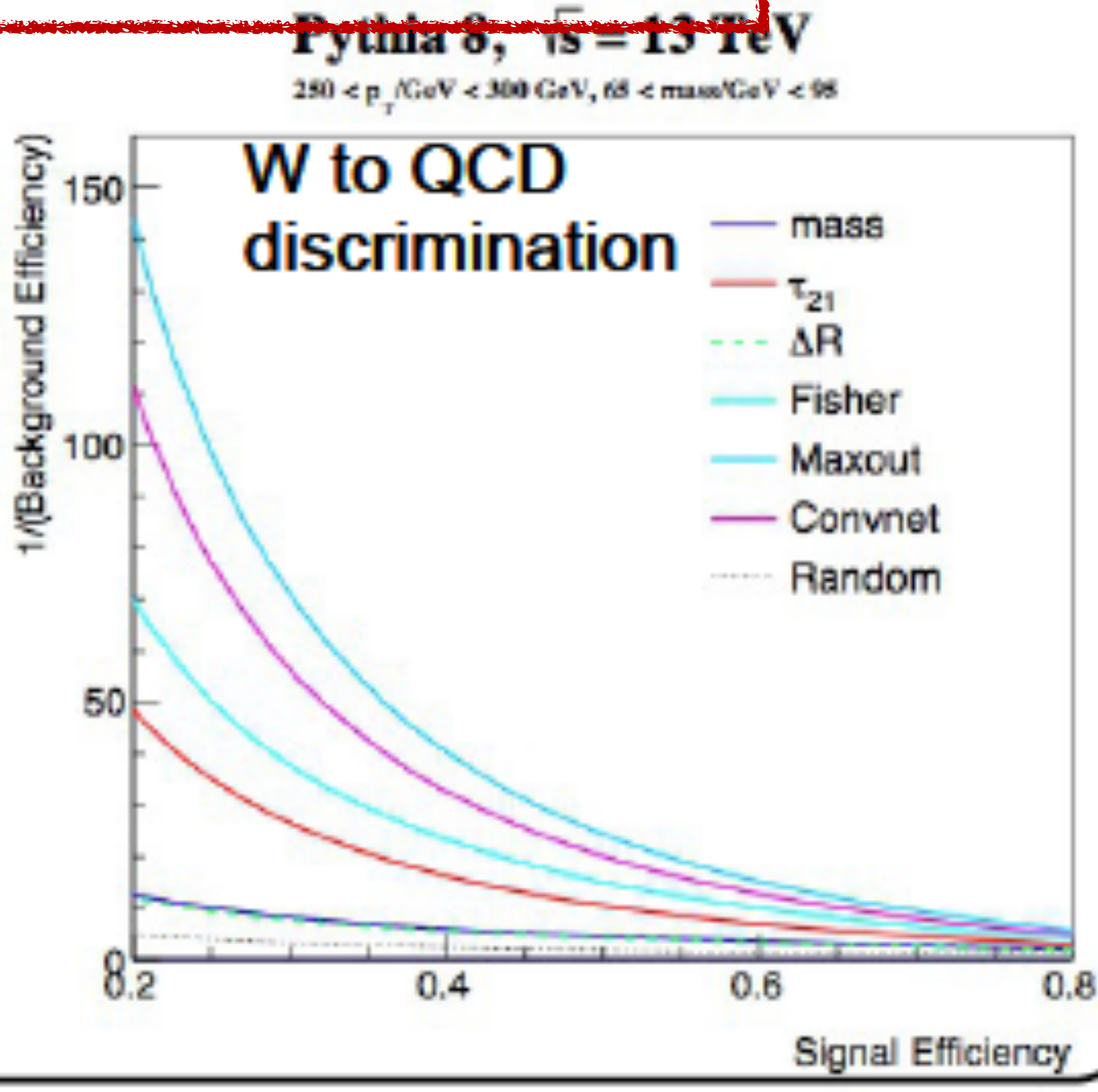
Neural net



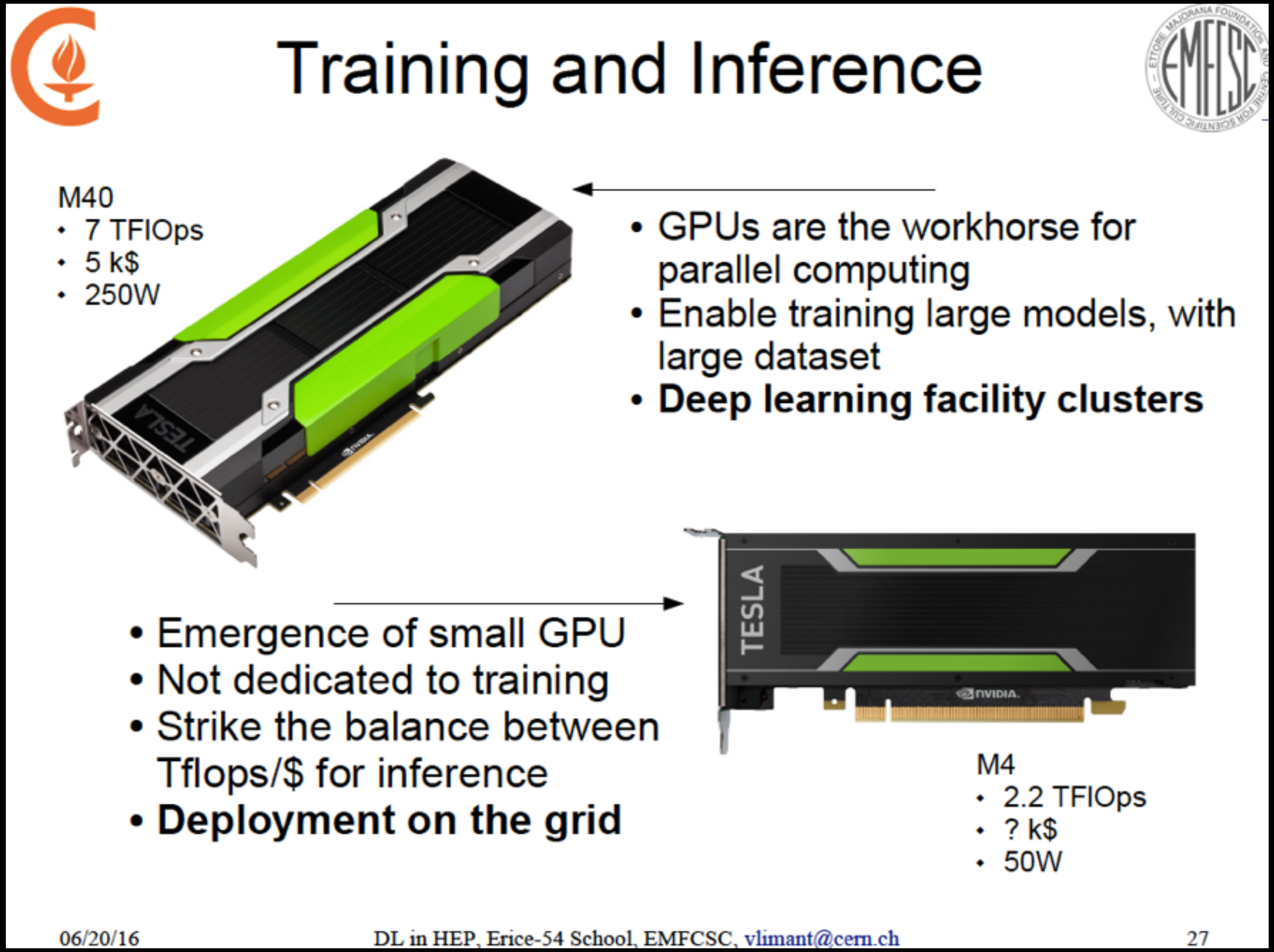
W tagger arXiv: 1511.05190, Oliveira, Kagan, Mackey, Nachman, Schwartzman

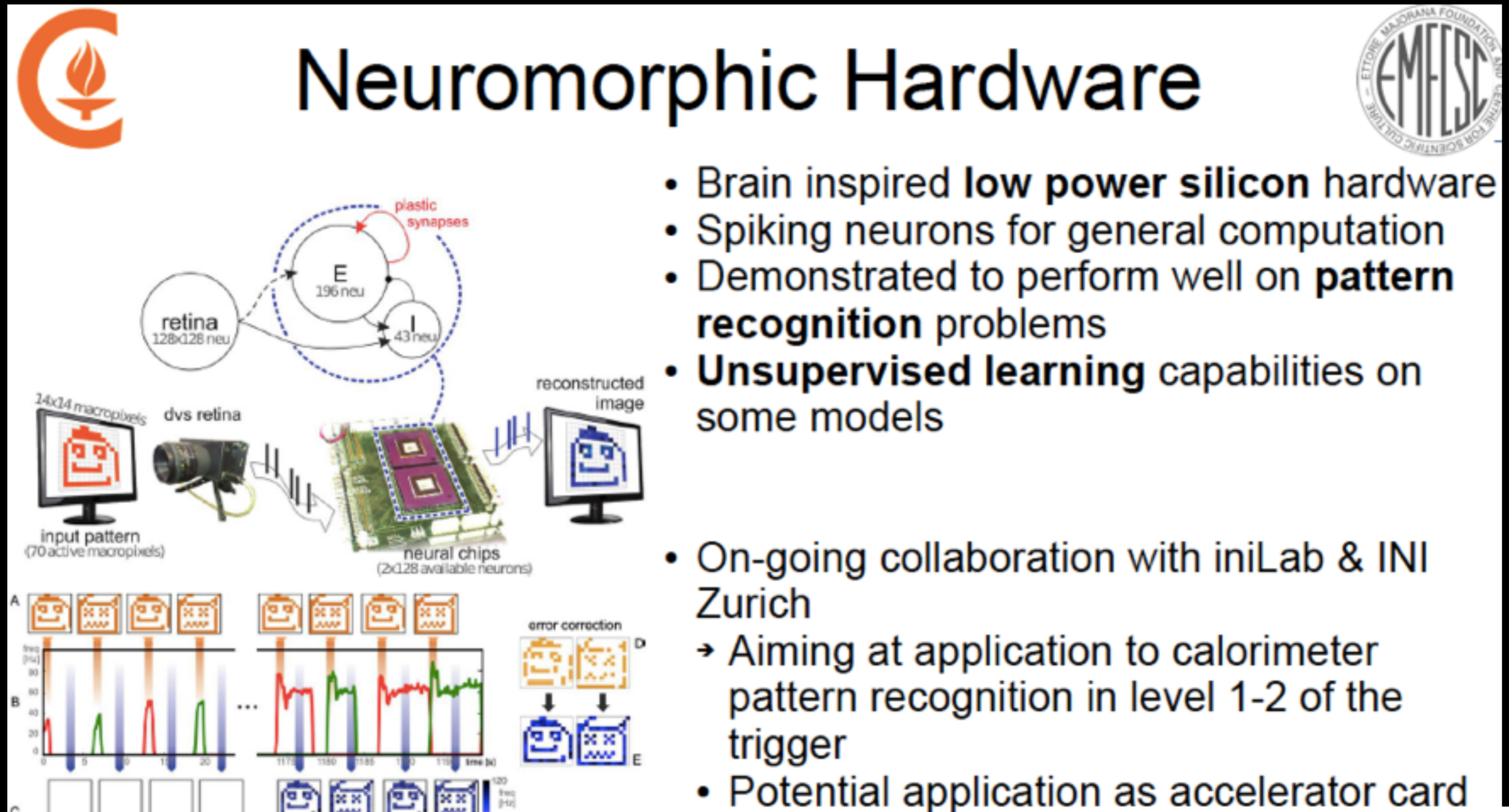


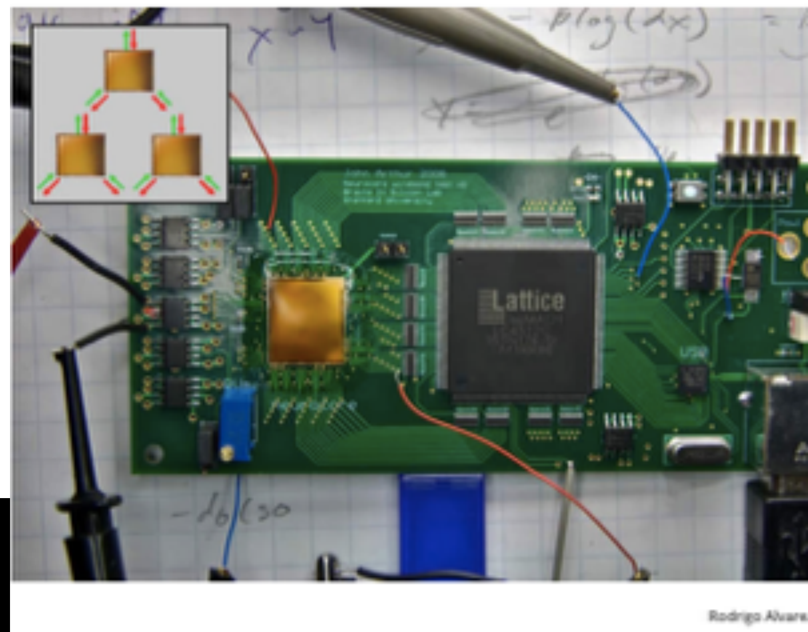
Train



HARDWARE







Neurogrid
65,536 artificial neurons packed onto just one of Neurogrid's chips

++ cognitive computing (e.g. IBM TrueNorth Spiking neuron technology, v low power consumption — application to pattern recognition)

OTHER APPS

Machine Learning Not Only for Data

- Schedule processing jobs on the grid
- Optimize network usage
- Reduce storage utilization
- All of the above ...
- Apply language processing to machine logs for fault prediction
- Data certification
- ...

GOOGLE DEEPMIND BLOG

DEEPMIND AI REDUCES GOOGLE DATA CENTRE COOLING BILL BY 40%

WEDNESDAY, 20TH JULY, 2016

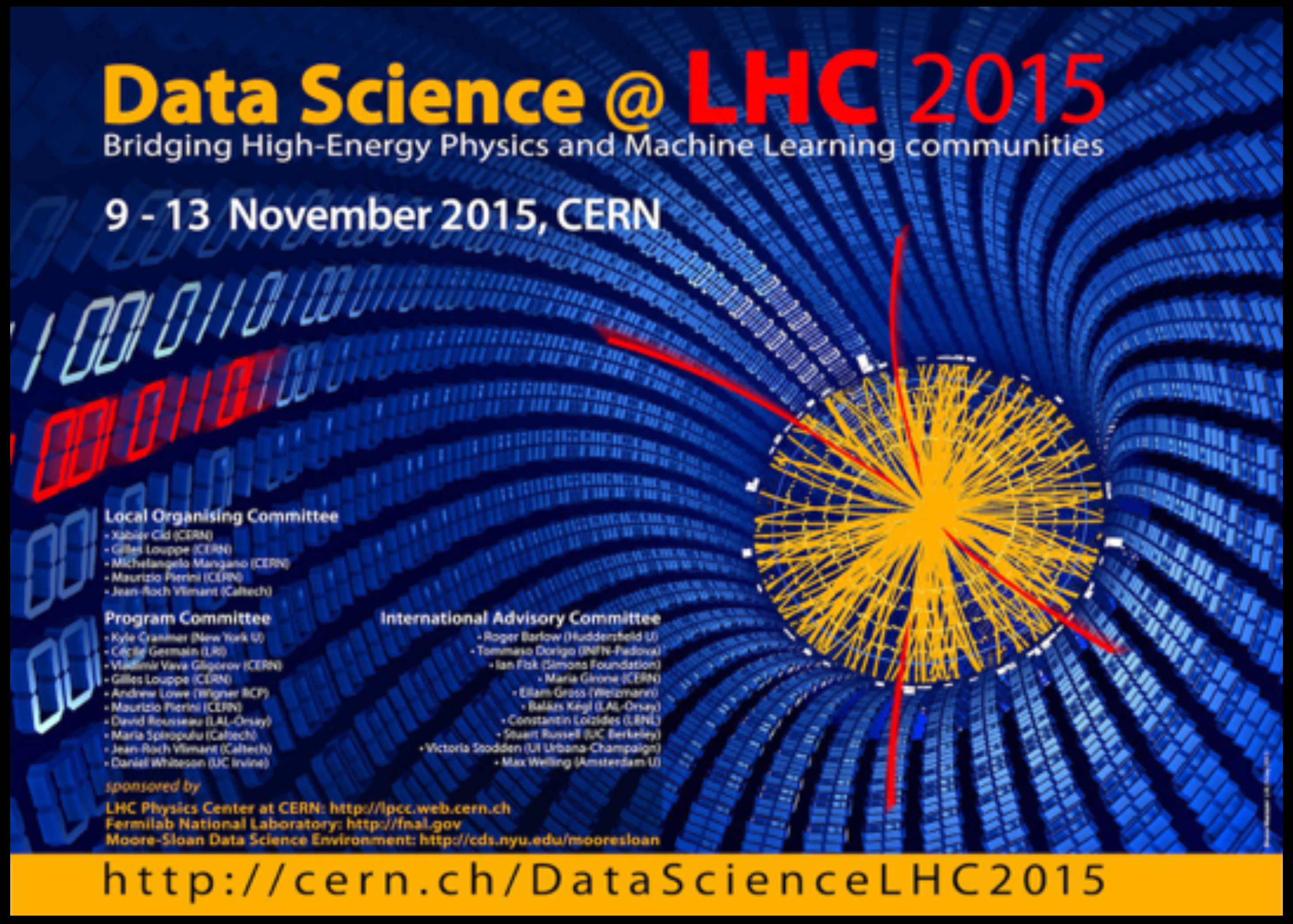
by Rich Evans, Research Engineer, DeepMind and Jim Gao, Data Centre Engineer, Google

From smartphone assistants to image recognition and translation, machine learning already helps us in our everyday lives. But it can also help us to tackle some of the world's most challenging physical problems -- such as energy consumption. Large-scale commercial and industrial systems like data centres consume a lot of energy, and while much has been done to stem the growth of energy use, there remains a lot more to do given the world's increasing need for computing power.

SUMMARY

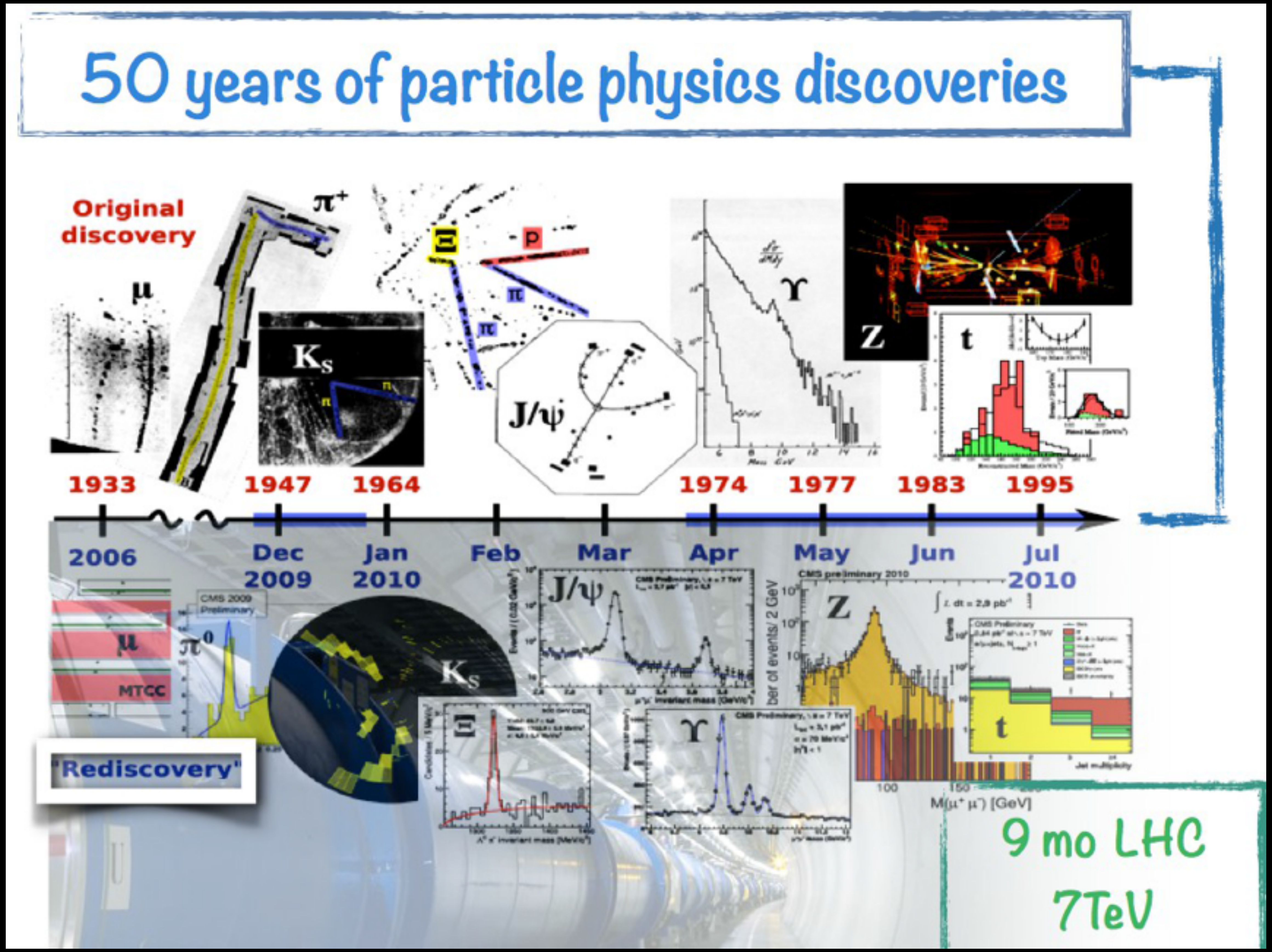
- Deep learning has leaped forward over the last decade
- Machine learning is a potential solution to several HL-LHC computational challenges
- Deep learning can further enable & accelerate scientific return, by tackling complexity
- Engage & collaborate with Data Science Experts on HEP Challenges
  - http://cern.ch/DataScienceLHC2015
  - https://indico.cern.ch/event/514434
  - https://indico.hep.caltech.edu/indico/event/102

stay tuned: NEXT at FNAL March 2017



SIZE OF THE BIG DATA CHALLENGE

- 1000 Hz 80% uptime
- 25B data events / year
- 150M CPUh @ 20s/event
- 40PB @ 1.5MB/event
- Simulation is half but takes more space and requires more CPU
- Analysis lightweight datasets of the order 10TB
- Analysis ran on the cluster / grid mostly I/O bound



UNBLINDING

- Past 48 hours were spent trying to repeat the entire analysis with the photon ID corrected.
- 24-hour analysis (shift) coverage for past 3 days between:
  - CERN (Dustin, Jay, Zhicai)
  - Australia - SUSY2016 (Cristian)
  - US (Si)