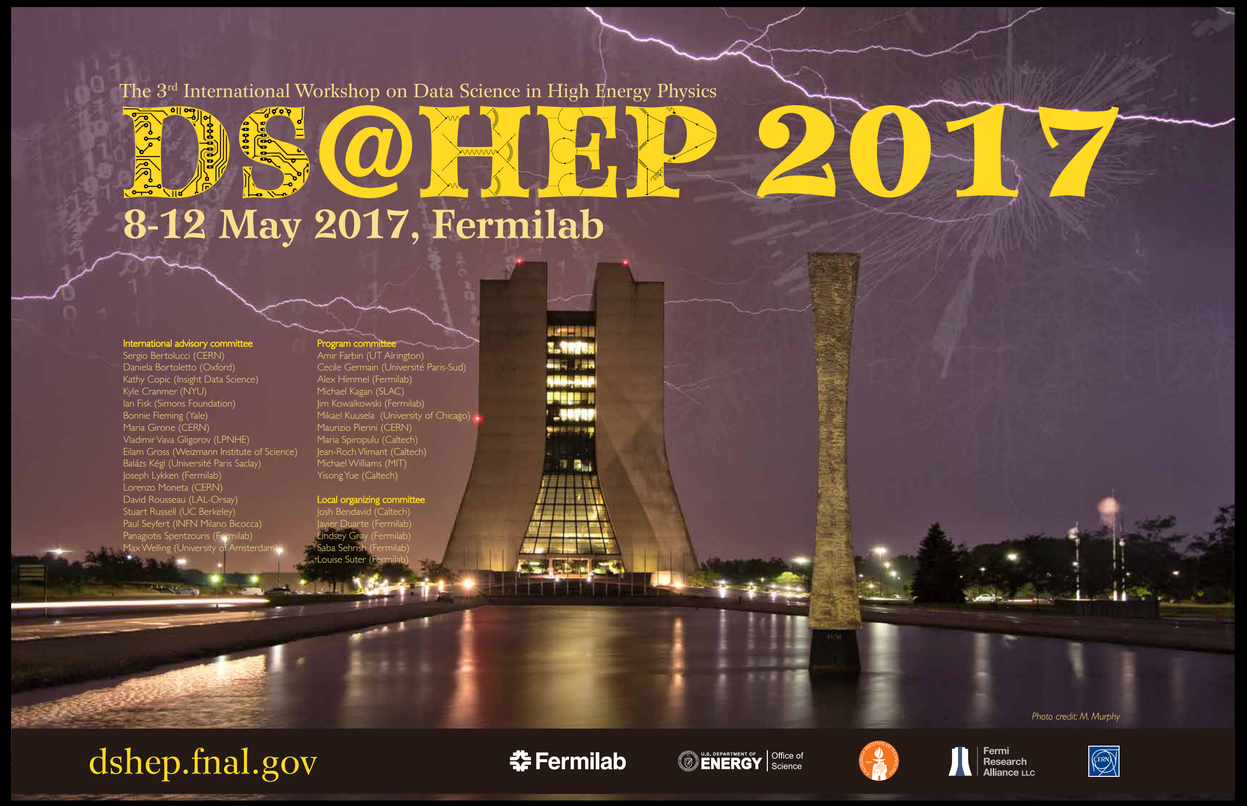
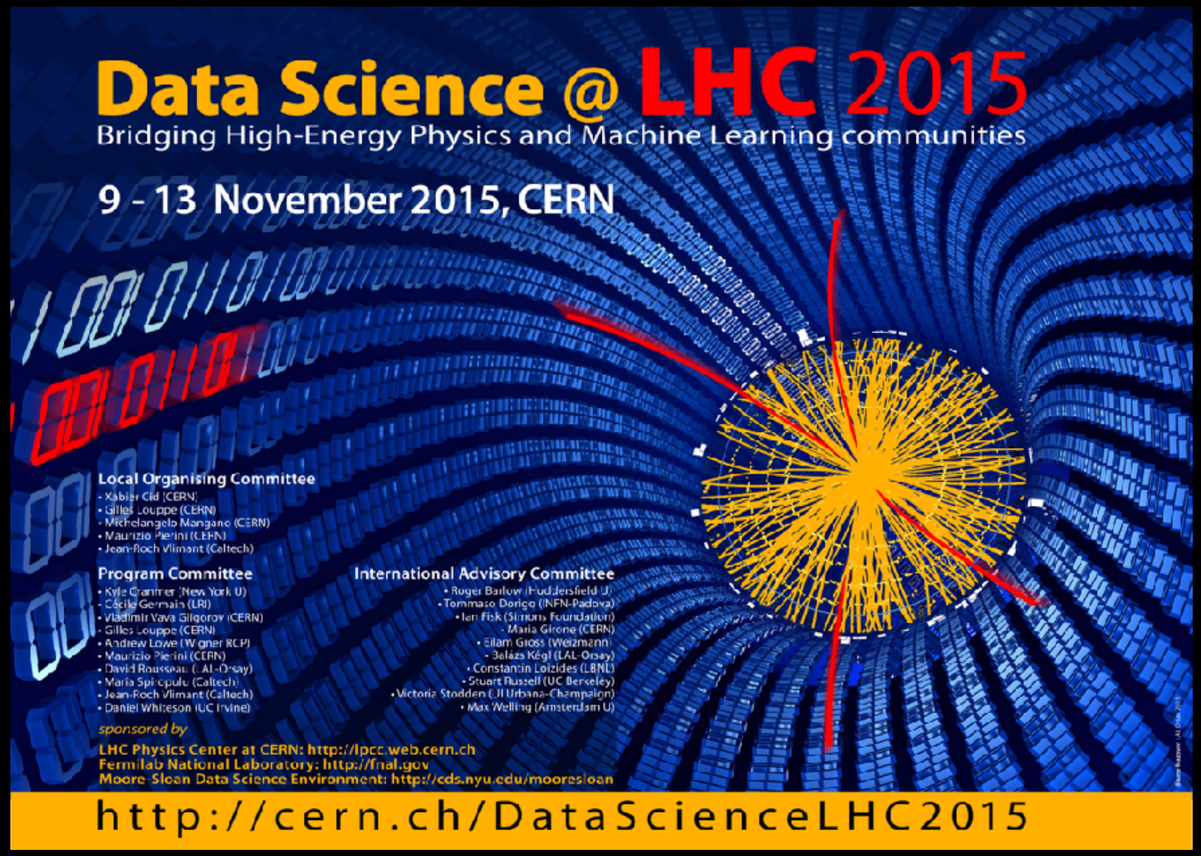


Maria Spiropulu
California Institute of Technology
smaria@caltech.edu



smaria@caltech.edu

Home | Poster

Program | Committees | Registration | Participants

Home

With the parallel progress in pattern recognition algorithms and microelectronic technology, the design and performance of tracking detector is rooted in the solid interplay of hardware and software : sensors, readout and trigger electronics, online and offline reconstruction software. The main focus of the workshop is on pattern recognition and machine learning algorithms devoted to the reconstruction of particle tracks or jets in high energy physics experiments, and the hardware developments that enable them.

This 2017 edition is a merger of the Connecting The Dot series (see CTD2015 Berkeley, CTD2016 Vienna) with the Workshop on Intelligent Tracker series (see WIT2010 Berkeley, WIT2012 Pisa, WIT2014 Penn).

contact us | Lodging | Travel to Orsay | Social events

Connecting The Dots 2016

22-24 February 2016
HEPHY Vienna
Europe/Vienna timezone

- Overview
- Scientific Programme
- Call for Abstracts
  - View my Abstracts
  - Submit Abstract
- Timetable
- Contribution List
- Paper Reviewing
- Registration
  - Registration Form
- Participant List
- Organization
- Venue / Hotels / Restaurants
- Travel Information
- Workshop Dinner (updated Feb 7)
- Pictures
- Proceedings

This is a workshop on track reconstruction and other problems in pattern recognition in sparsely sampled data. The workshop is intended to be inclusive across other disciplines wherever similar problems arise. The main focus will be on pattern recognition and machine learning problems that arise e.g. in the reconstruction of particle tracks or jets in high energy physics experiments. Both hardware and software aspects will be addressed.

Talks are by invitation only, please contact the organizers (ctd2016@oeaw.ac.at) or a member of the Scientific Advisory Committee for further details.

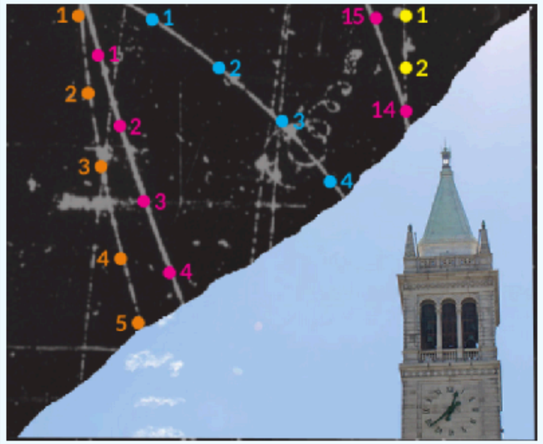
The first workshop on "Connecting the Dots" took place at LBNL in February 2015, see here.

The workshop will be broadcast on vidyo -- the room PIN is 1050.

Starts 22 Feb 2016 08:00
Ends 24 Feb 2016 18:00
Europe/Vienna

HEPHY Vienna
Seminarraum 1,2,3
Wohllebengasse 12-14
1040 Wien
Austria

Pictures
latex template



Connecting The Dots 2015

9-11 February 2015
UC_Berkeley/LBNL
US/Pacific timezone

- Overview
- Scientific Programme
- Timetable
- Contribution List
- Registration
  - Registration Form
- Participant List
- Organization
- Practical Information
- Direct to INDICO style timetable view

Scientific Programme

Algorithms and theoretical analysis

Mathematical evaluation of pattern recognition problems, fitting, effect of noise, treatment of multiple scattering, theoretical limits, etc.

Parallel and/or discrete pattern recognition

Includes Hough transform approaches, look-up tables, associative memory.

Neural networks, machine learning, and neuromporhic approaches

Includes both software/firmware implementations and exploration of neuromorphic hardware

Applications and performance evaluation

Examples of implemented pattern recognition problems and solutions with emphasis on new challenges and limits of scaling existing approaches.

smaria@caltech.edu

BIG CHANGE

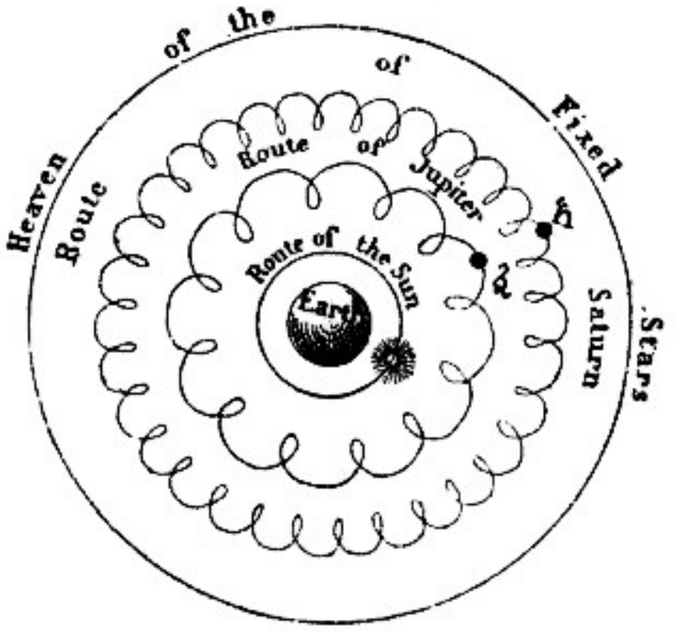
The scientific method (and knowledge-base) has been transformed by massive computation (including networking/sharing, data & data analytics) in all science (and other) domains of research & study.

massive computation builds new intuition

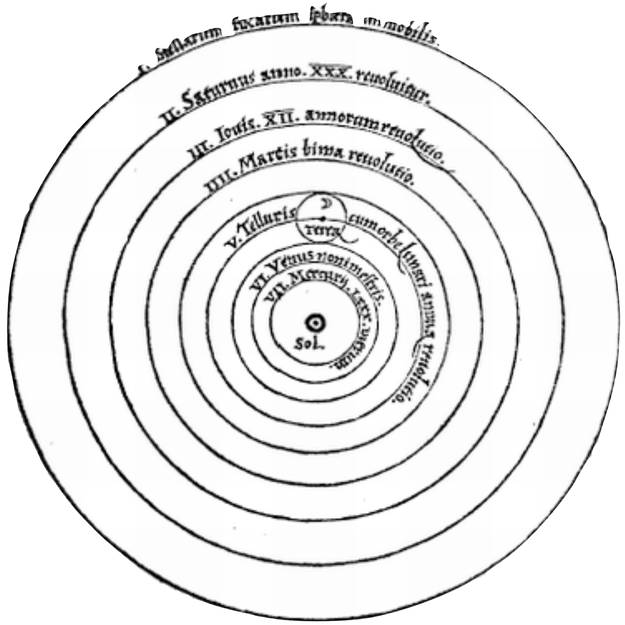
DOMAIN SCIENCES & COMPUTATION

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

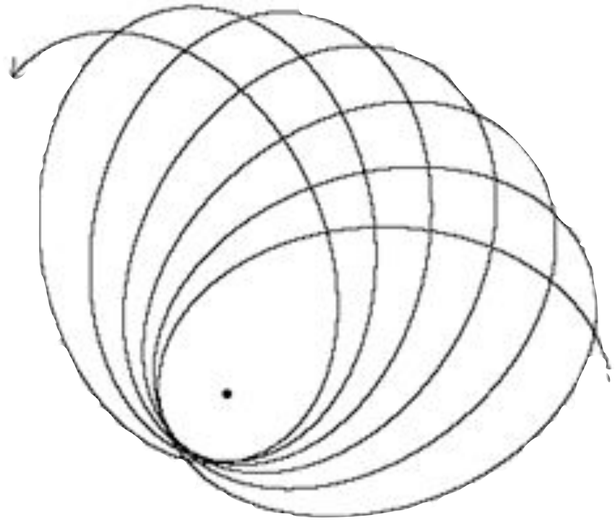
The task is to push more (big) data through intelligent & networked machines such that we get to ask bigger questions and challenge our understanding of the world if the data indicates so



Epicyclic Orbits

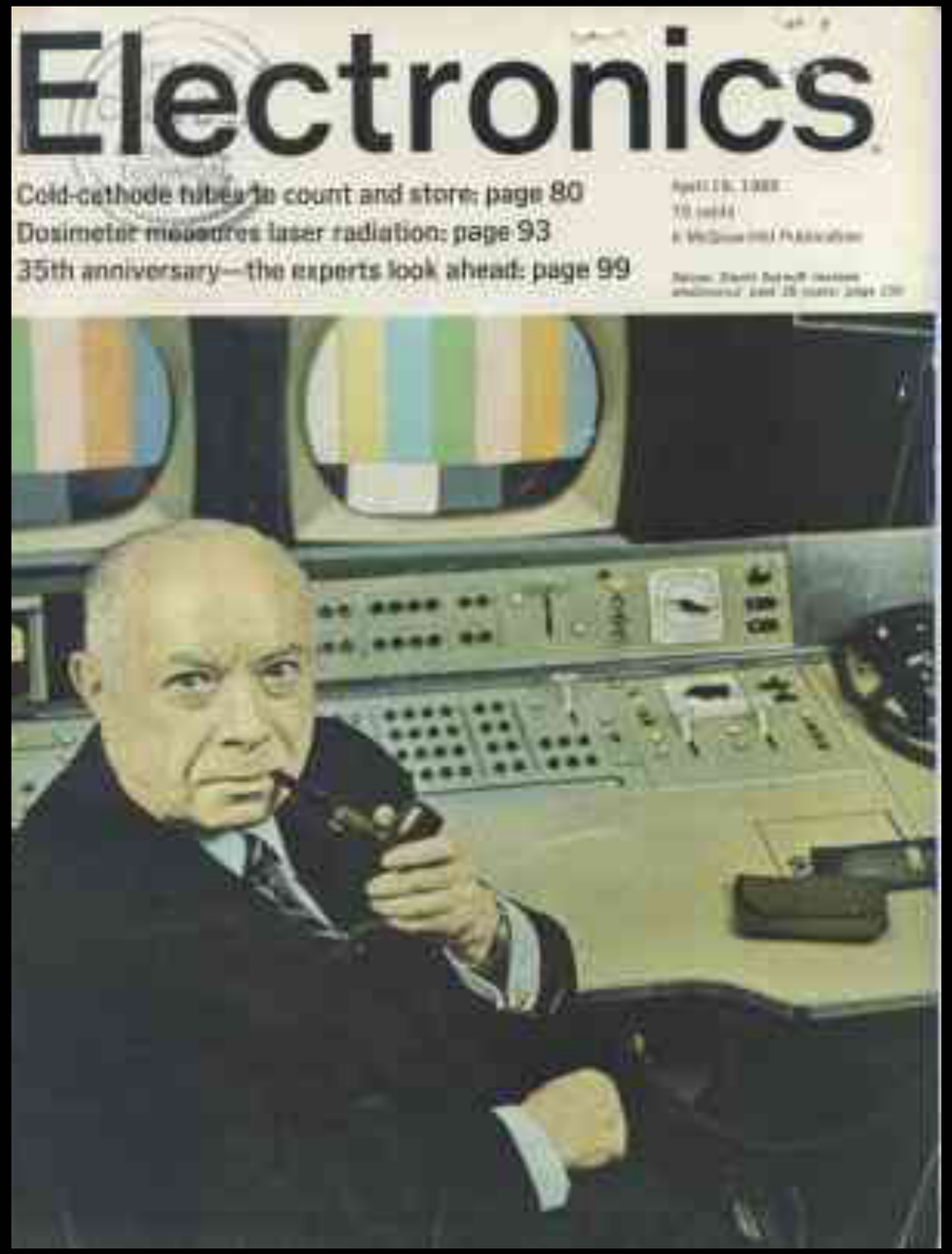


Keplerian Orbits



General Relativistic Orbits

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.

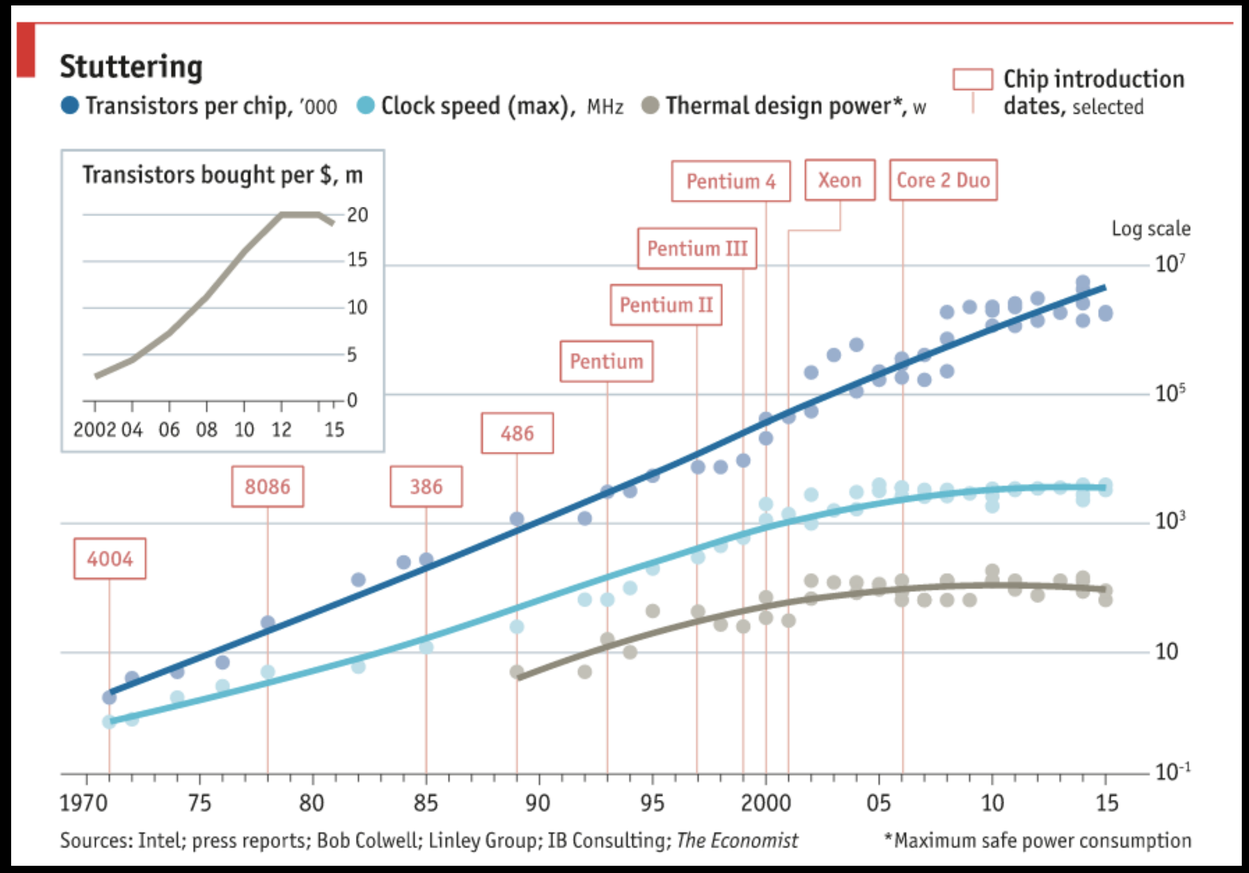
Present and future

By integrated electronics, I mean all the various tech-

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

AFTER MOORE

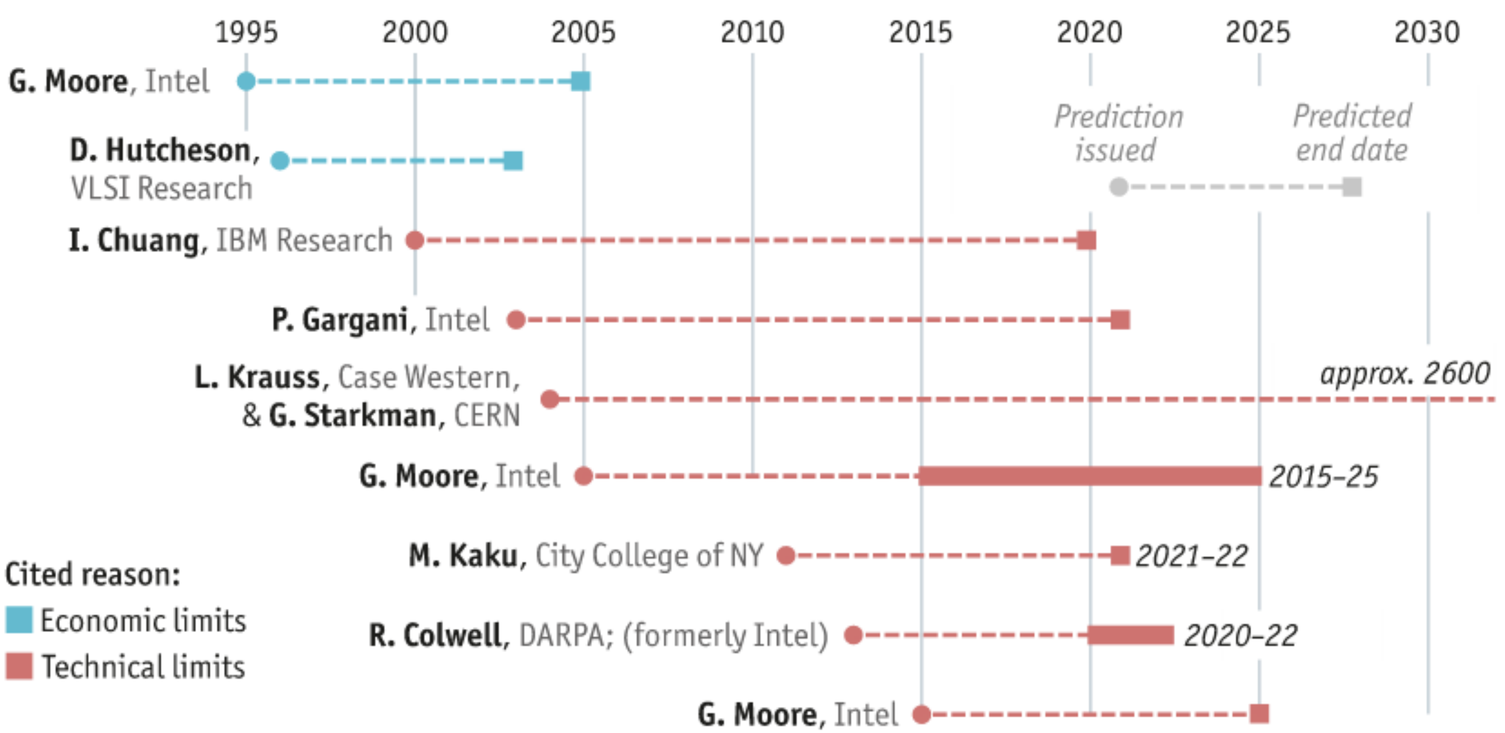
http://www.economist.com/technology-quarterly/2016-03-12/after-moores-law



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

Faith no Moore

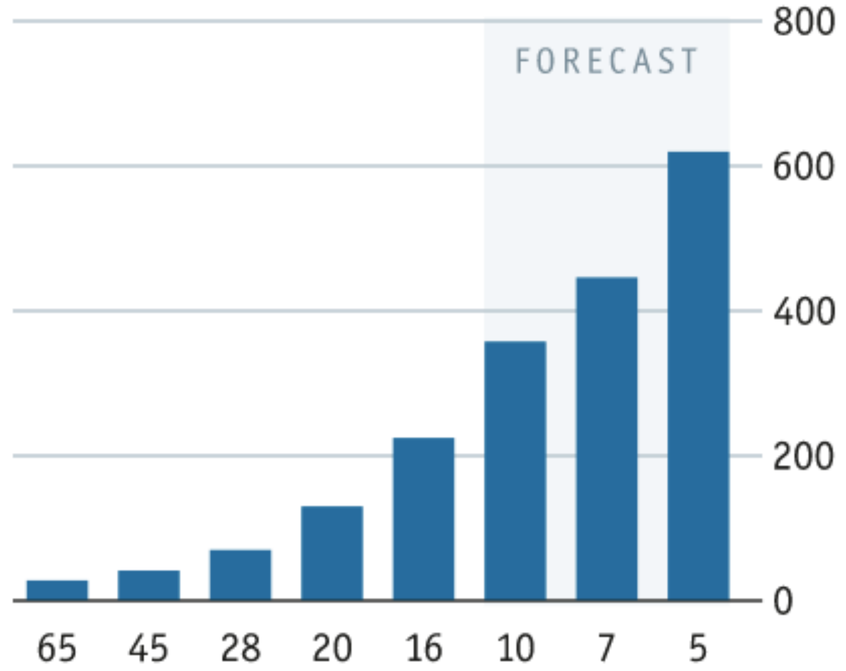
Selected predictions for the end of Moore's law



Sources: Intel; press reports; *The Economist*

This can't go on

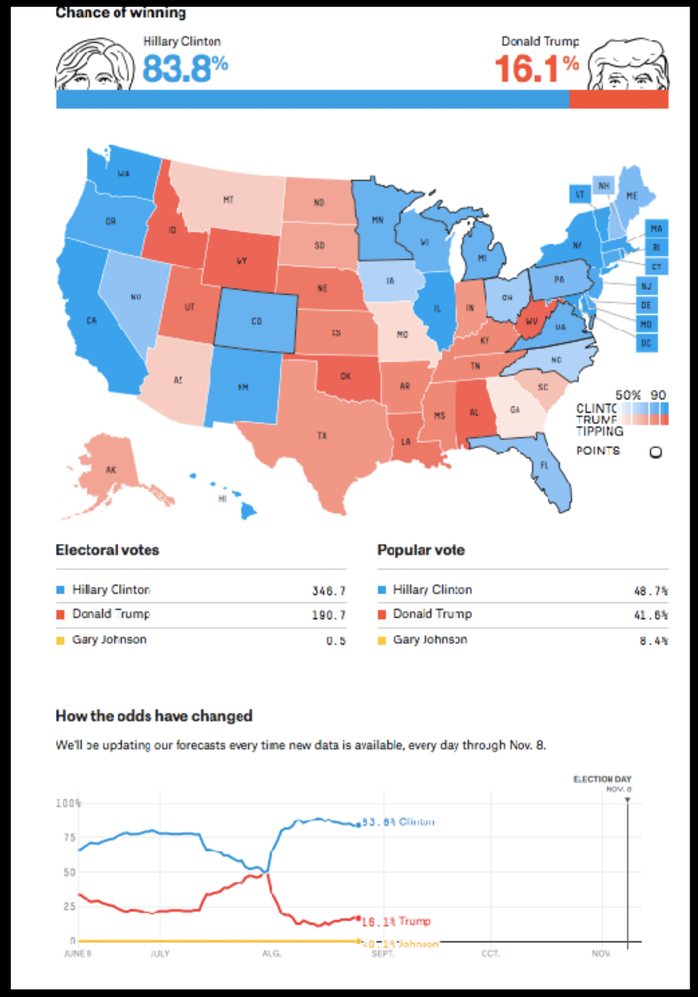
Design cost by chip component size in nm, $m



Source: IB Consulting

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)

THE 21ST CENTURY

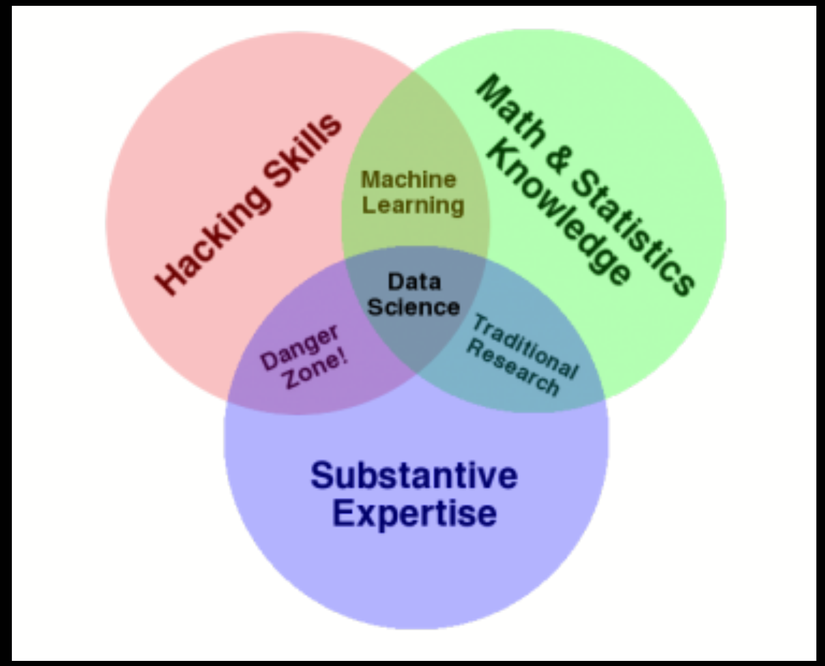
- 1997 The journal Data Mining and Knowledge Discovery is launched
- December 1999 Jacob Zahavi is quoted in "Mining Data for Nuggets of Knowledge" in Knowledge@Wharton: "**Conventional statistical methods work well with small data sets**. Today's databases, however, can involve millions of rows and 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.'"
- **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

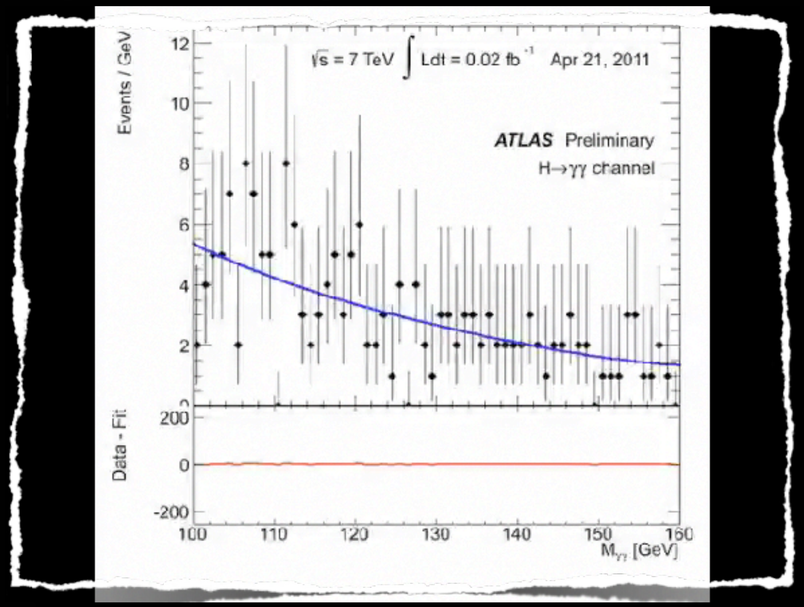
- March 2009 Kirk D. Borne and other astrophysicists submit to the Astro2010 Decadal Survey a paper titled “The Revolution in Astronomy Education: Data Science for the Masses “
- September 2010 Drew Conway writes in “The Data Science Venn Diagram”: “…I present the Data Science Venn Diagram… hacking skills, math and stats knowledge, and substantive expertise.”

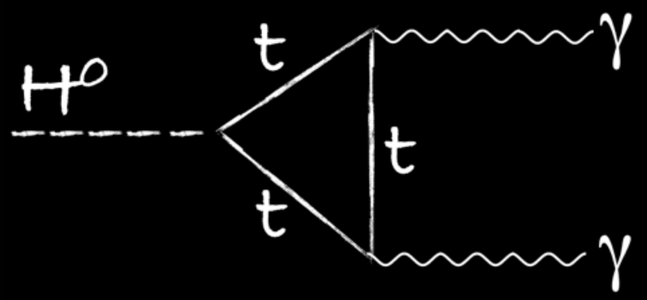


- May 2011 David Smith writes in “’Data Science’: What’s in a name?”: …. I think **‘Data Science’ better describes what we actually do: a combination of computer hacking, data analysis, and problem solving.”**

DO WE REALLY WANT/NEED ML/AI IN SCIENCE?

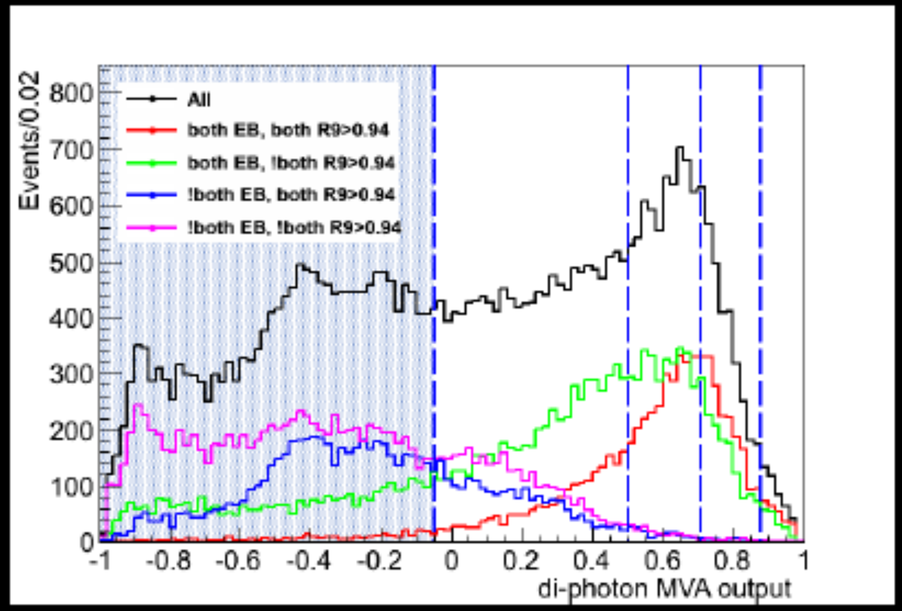
- We used it already to discover the Higgs in the form of MVAs and BDTs !
- We would have discovered it with traditional cut & count methods with about 2x the data (time & $$)
- The problem with the techniques is the “black box” aspect/ interpetability - the treatments or errors and correlations and more…





MVAs

- State of the art at the LHC : MVAs
  - The H(γγ) uses 6 BDTs (outputs feed into each other)
- Single classifier
- Complexity challenges
  - adapt to changing conditions
  - understand systematic errors
  - interprete deviations

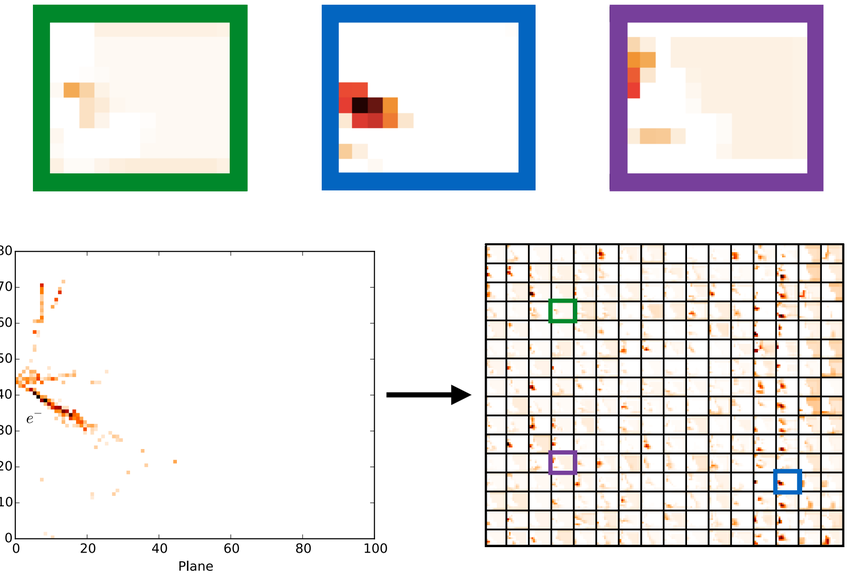


DL MADE IT IN DIRECTOR'S SLIDES

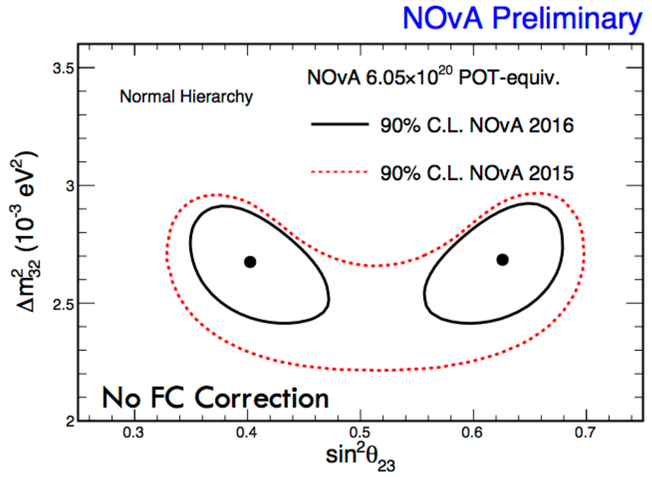
Improved Event Selection

- 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

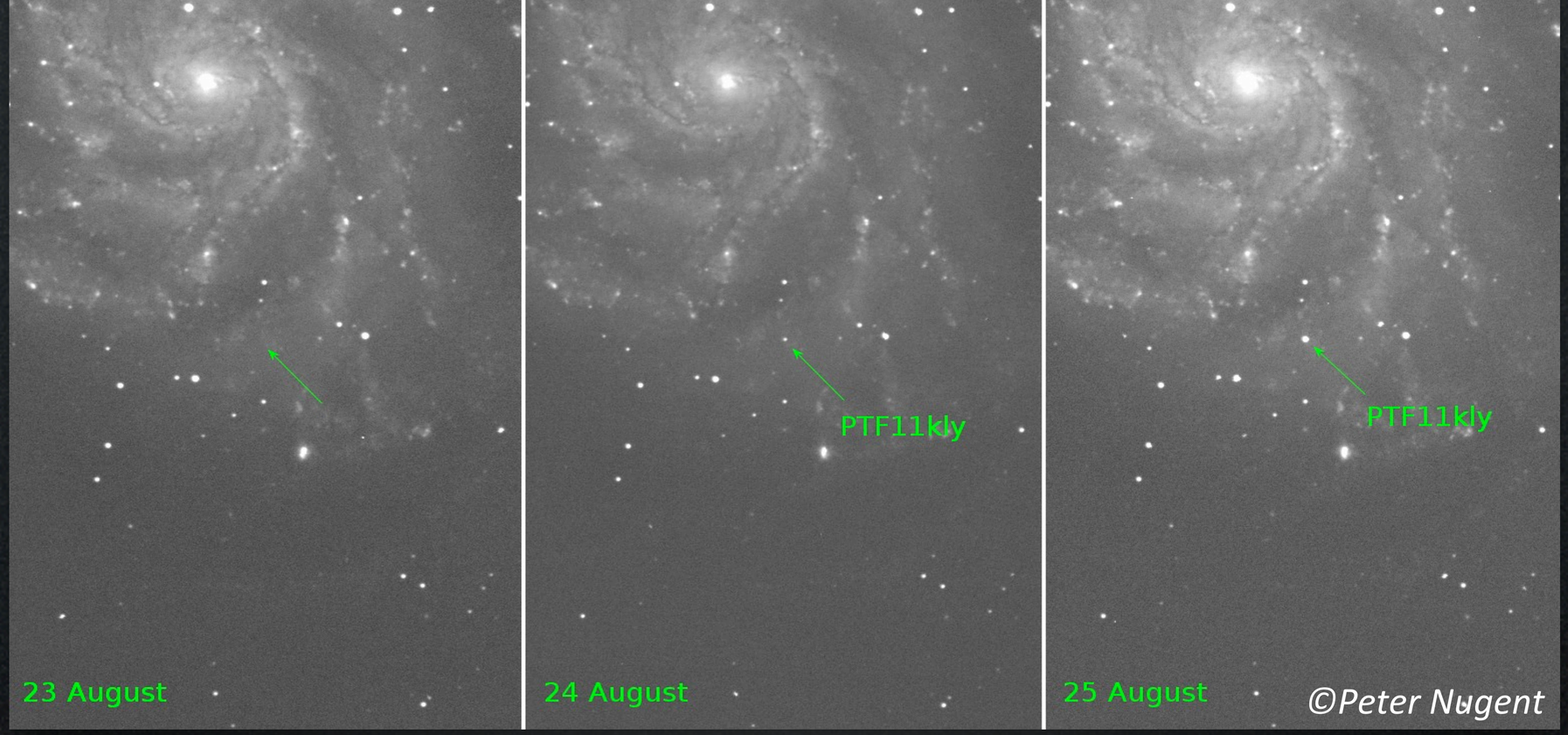


Best Fit (in NH):

$$|\Delta m^2_{32}| = 2.67 \pm 0.12 \times 10^{-3}\text{eV}^2$$
$$\sin^2\theta_{23} = 0.40^{+0.03}_{-0.02}(0.63^{+0.02}_{-0.03})$$

maximal mixing excluded at 2.5 σ

Fermilab



Supernova Discovery in the Pinwheel Galaxy

11 hr after explosion

nearest SN Ia in >3 decades

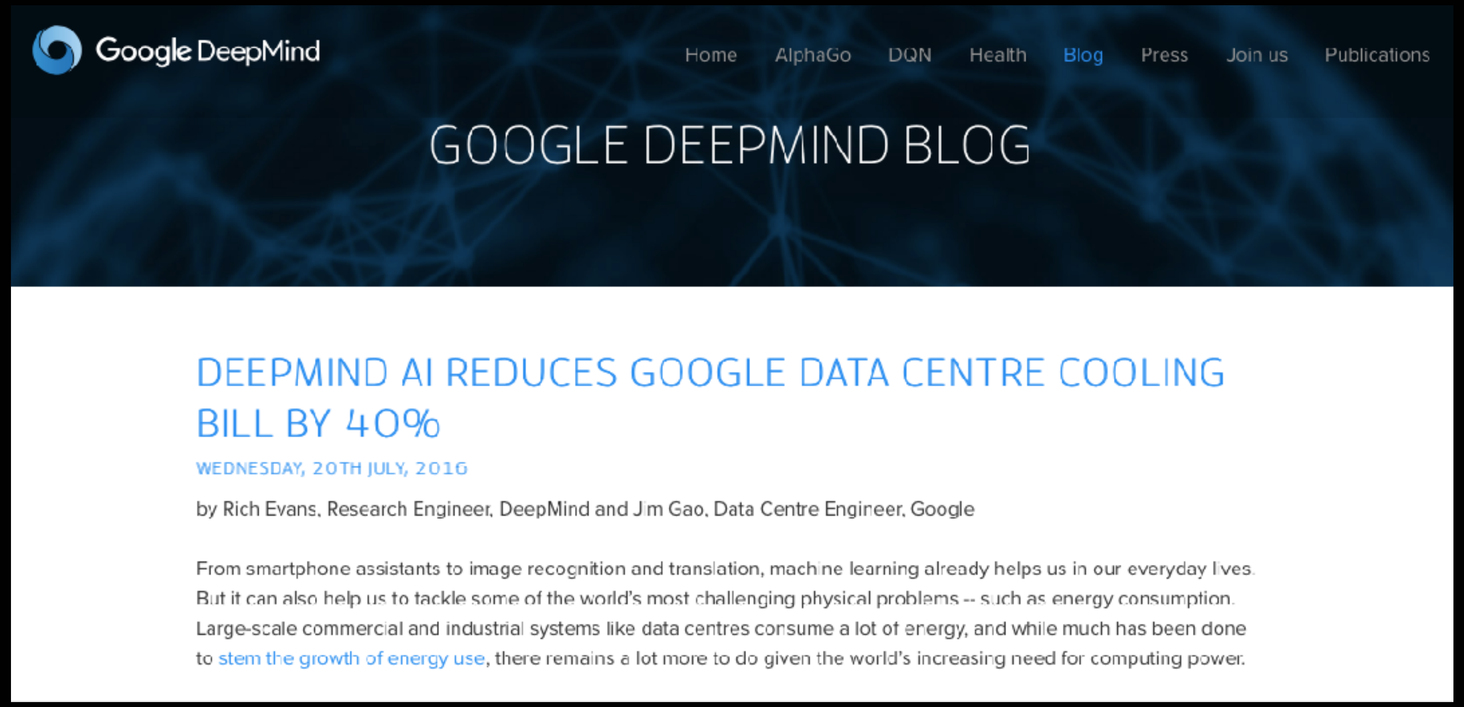
Discovered by our machine-learning framework

in PTF: >10,000 events in > 0.2 PB of imaging → 50+ journal articles

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
- ...



we will want AI to help us
debug our thinking by using all data from all experiments optimally and “open our eyes” just as AlphaGo opened the eyes of the professional go players and enhance our intuition and creativity and ability to break paradigms and boxes

The 3rd International Workshop on Data Science in High Energy Physics

DS@HEP 2017

8-12 May 2017, Fermilab

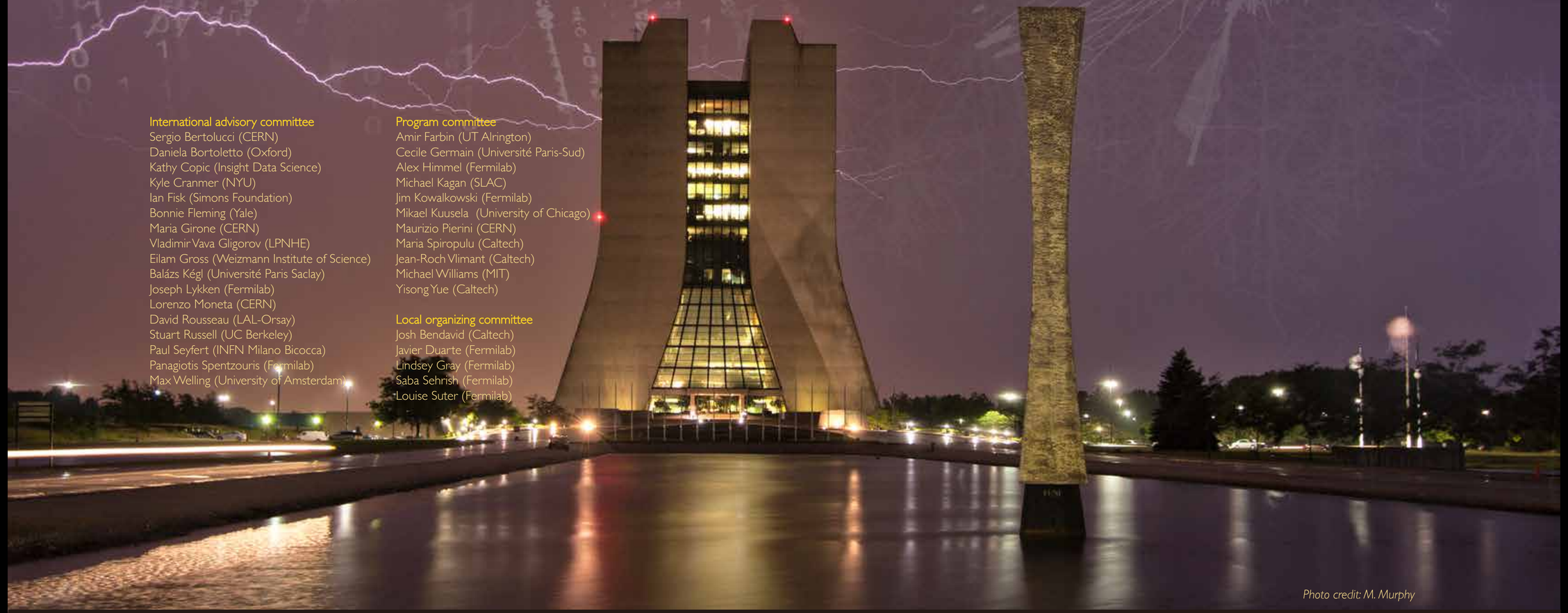


Photo credit: M. Murphy

International advisory committee
Sergio Bertolucci (CERN)
Daniela Bortoletto (Oxford)
Kathy Copic (Insight Data Science)
Kyle Cranmer (NYU)
Ian Fisk (Simons Foundation)
Bonnie Fleming (Yale)
Maria Girone (CERN)
Vladimir Vava Gligorov (LPNHE)
Eilam Gross (Weizmann Institute of Science)
Balázs Kégl (Université Paris Saclay)
Joseph Lykken (Fermilab)
Lorenzo Moneta (CERN)
David Rousseau (LAL-Orsay)
Stuart Russell (UC Berkeley)
Paul Seyfert (INFN Milano Bicocca)
Panagiotis Spentzouris (Fermilab)
Max Welling (University of Amsterdam)

Program committee
Amir Farbin (UT Alrington)
Cecile Germain (Université Paris-Sud)
Alex Himmel (Fermilab)
Michael Kagan (SLAC)
Jim Kowalkowski (Fermilab)
Mikael Kuusela (University of Chicago)
Maurizio Pierini (CERN)
Maria Spiropulu (Caltech)
Jean-Roch Vlimant (Caltech)
Michael Williams (MIT)
Yisong Yue (Caltech)

Local organizing committee
Josh Bendavid (Caltech)
Javier Duarte (Fermilab)
Lindsey Gray (Fermilab)
Saba Sehrish (Fermilab)
Louise Suter (Fermilab)

dshep.fnal.gov

Fermilab | U.S. DEPARTMENT OF ENERGY | Office of Science | Fermi Research Alliance LLC | CERN

It was in 1820 that the great French scientist and mathematician Pierre Laplace staked out the ultimate claim of the physicists for intellectual hegemony over all science — and possibly a lot else —:

> "We ought to regard the present state of the universe as the effect of the antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes...."
>
> — Laplace, 1820

THANK YOU FOR LISTENING