

# AI-driven decision automation in physics research and enterprises

Prof. Dr. Michael Feindt  
Professor of Physics, KIT  
Founder and Chief Scientist, Blue Yonder GmbH

LSDMA Symposium and GRIDKA School at KIT, Sept 2017





1982-1991 DESY (PLUTO, CELLO at PETRA)

1991-1997 CERN (DELPHI at LEP)

since 1997 Professor at Univ. Karlsruhe (now KIT)

since 1997 CDF II at Fermilab / DELPHI at LEP / (CMS at LHC)

since 2008 Belle, Belle II at KEK

1999/2000 invention of NeuroBayes algorithm

2002 foundation of Phi-T

2008 foundation of Blue Yonder,

with offices in Karlsruhe, Hamburg, London, Dallas

**<phi-t>**<sup>®</sup>  
Physics Information Technologies

**blueyonder**  
Forward looking. Forward thinking.





**Combination of Big Data with Data Science,  
Machine-Learning and Artificial  
Intelligence, to reinvent and revolutionalize  
business processes. Basis: CERN**



Blue Yonder =  
Digital innovation / disruption  
from Germany, originating from HEP



Digitalisation:  
Value through data and (scientific) software



A close-up portrait of Marc Andreessen, a bald man with light skin and brown eyes, wearing a dark suit jacket over a white shirt. The background is a solid dark grey. The image is framed by magenta vertical bars on the left and right sides.

„Software Eats the World“

- Marc Andreessen, Silicon Valley Wunderkind

now: AI Eats the World



Mehr als nur Google

# Die Meute ist unterwegs







“Google, Facebook, and Amazon have each applied Machine Learning to effectively eviscerate entire industries. This is just the beginning, so the wait-and-see option is looking pretty scary.”

— Geoffrey Moore



# What makes machines intelligent?

## 2 main branches in AI: „brute force silicon“

Machines learn, what humans can do easily (see, understand, drive a car). Moore's law + deep neural networks

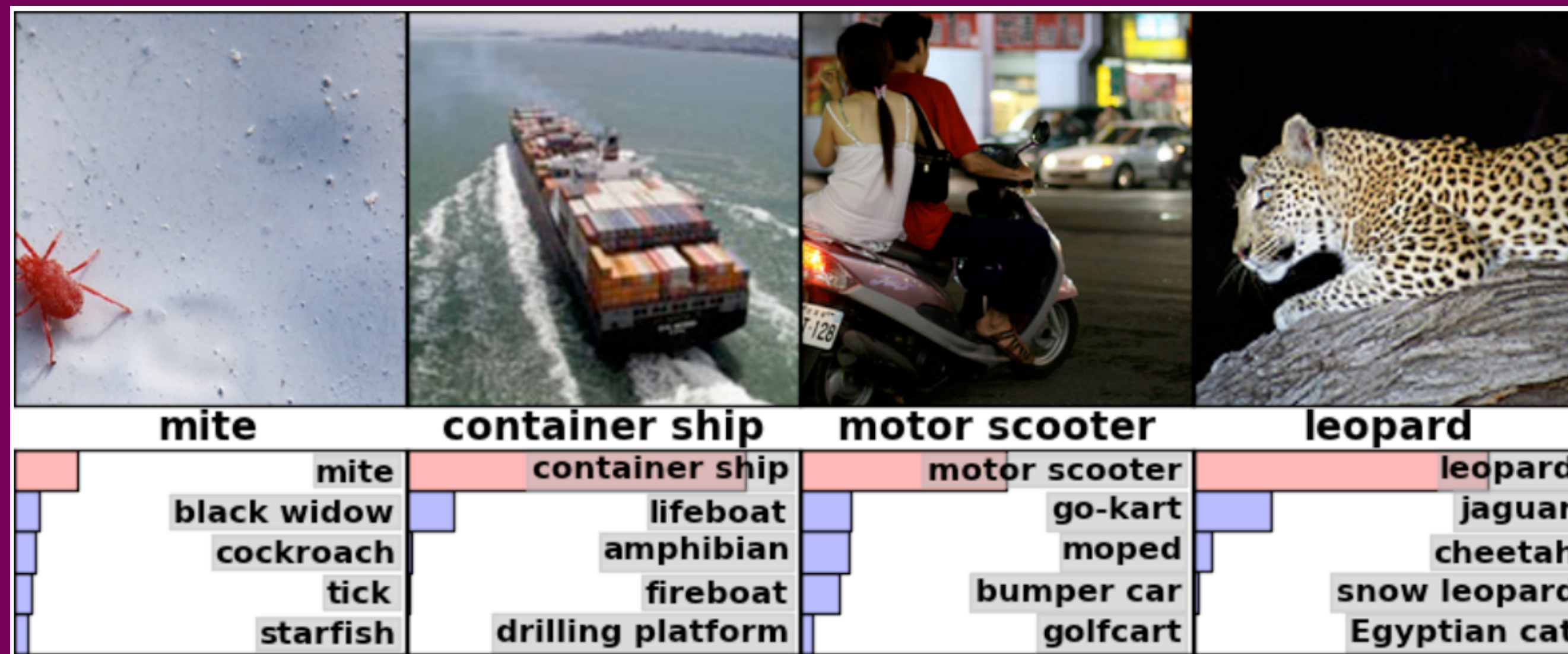
## intelligent algorithms

Construction applying domain knowledge for concrete problem settings, optimized and adapted by Machine Learning on the basis of observed or simulated data. Machines become better than the best human experts



# Artificial Intelligence I

Machines learn, what humans can do easily: e.g. image recognition



helps to improve human-machine communication and automation.

Deep neural networks (deep learning)

Google invested > 1 bn \$ CPU time into training of such neuronal networks.



# Artificial Intelligence II:

Invent intelligent algorithms that are able to optimize very complex action chains.

Deep neural networks + reinforcement learning:

alpha-go:

beats the „go“- world champion,  
learned (also) by playing against itself  
to develop superhuman performance.  
Develops sort of „gut feeling“, cannot calculate all possibilities.





Important personal and professional  
decisions:  
“gut feeling”

00%



Operational decisions e.g. in retail:  
Automatic data driven decisions by  
Artificial Intelligence

990%



# Repeated decisions

Order ?

Price ?

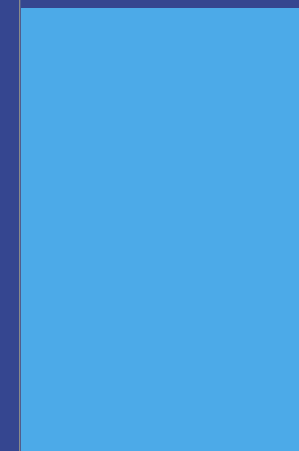
Catalog?



do nothing



business rules



think





Human brain:

(1) fast, intuitive

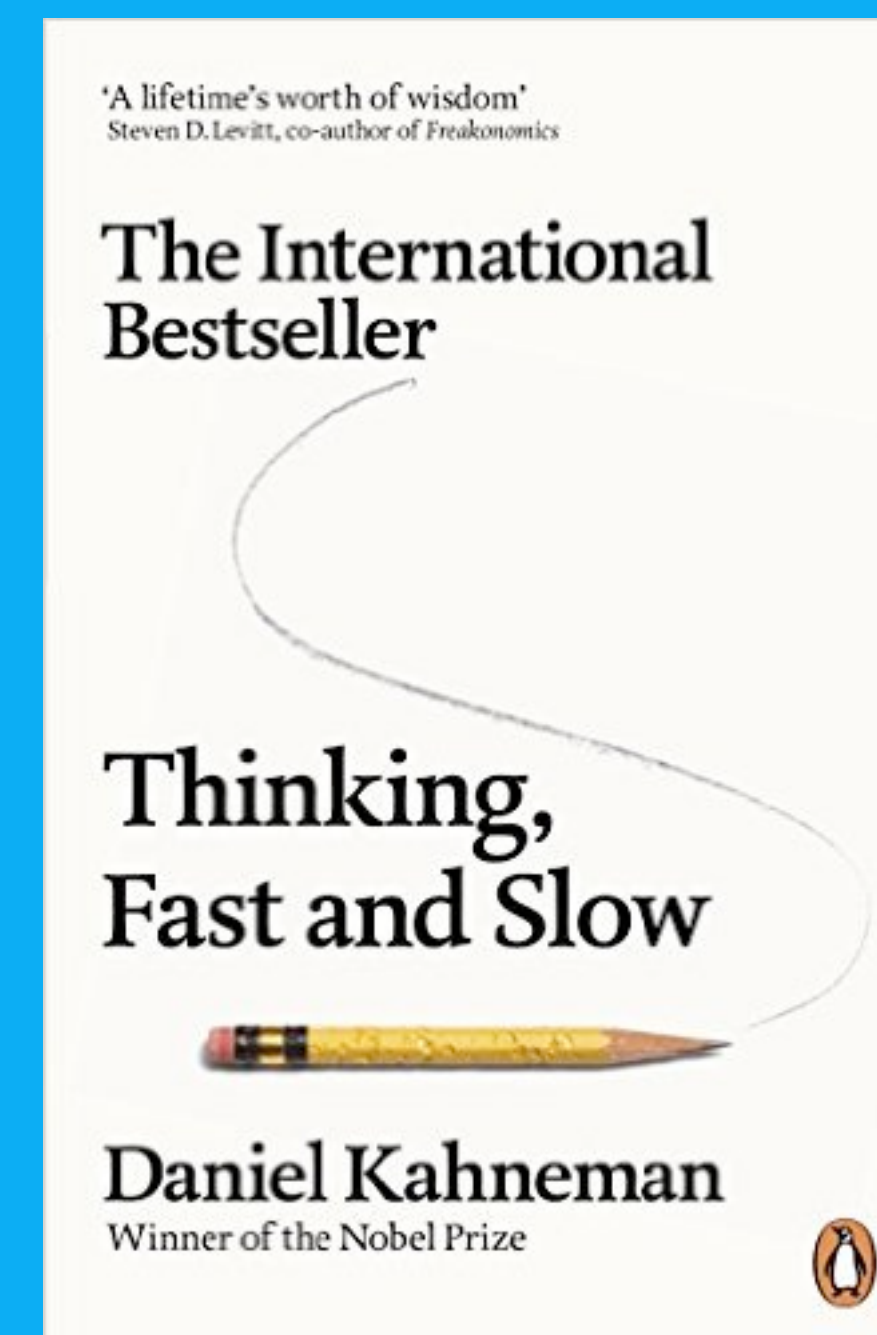
(2) slow, rational





# Human decision making:

# biases, biases





**Predictive Analytics**

**Prescriptive Analytics**

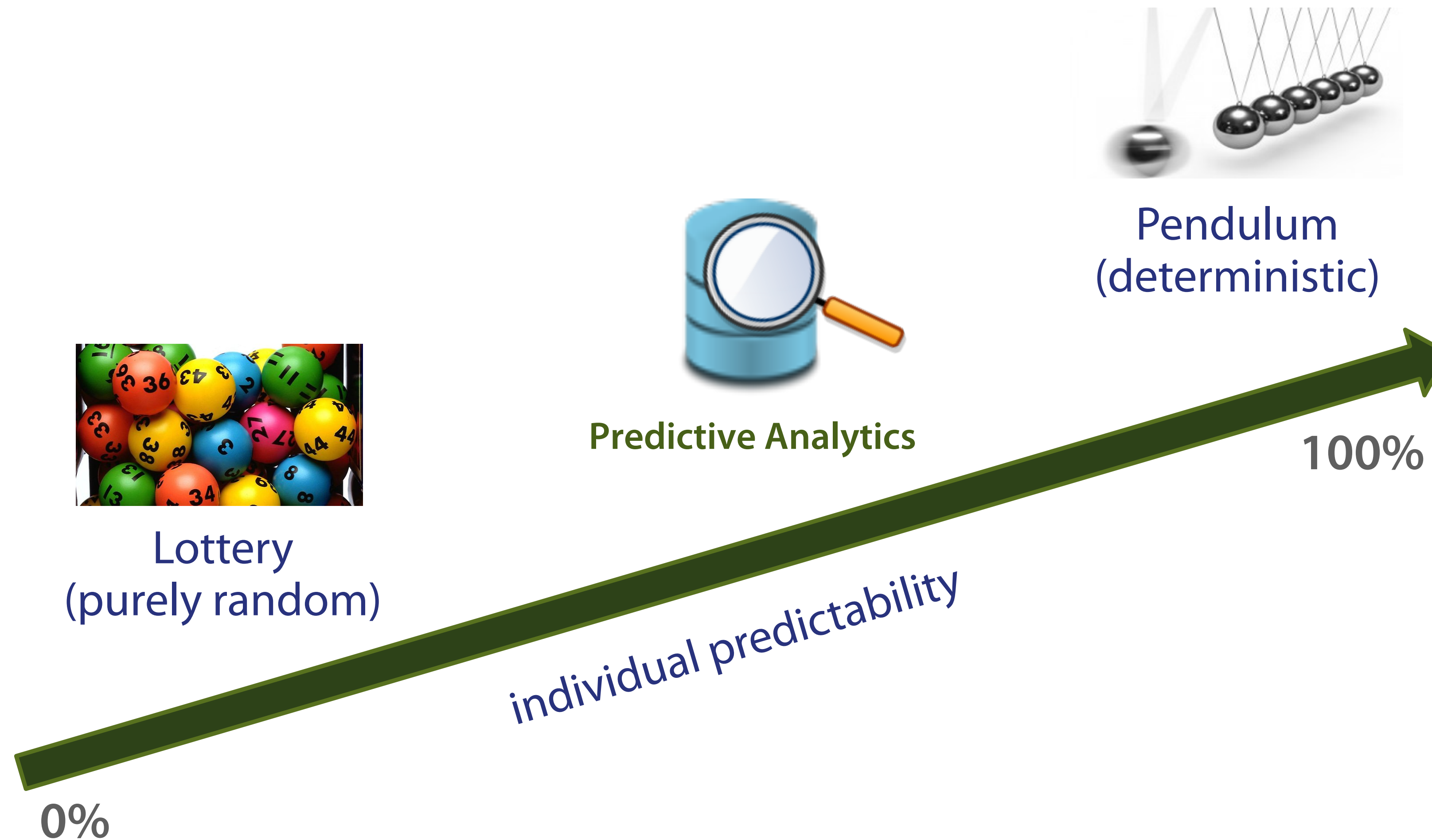
**= Disciplines of  
machine learning ML /  
artificial intelligence AI**



- data
- predictions
- cost/utility
- optimisation
- automation



# When predictive analytics?





# Many influencing factors...





# on many individual events

Items



Stores



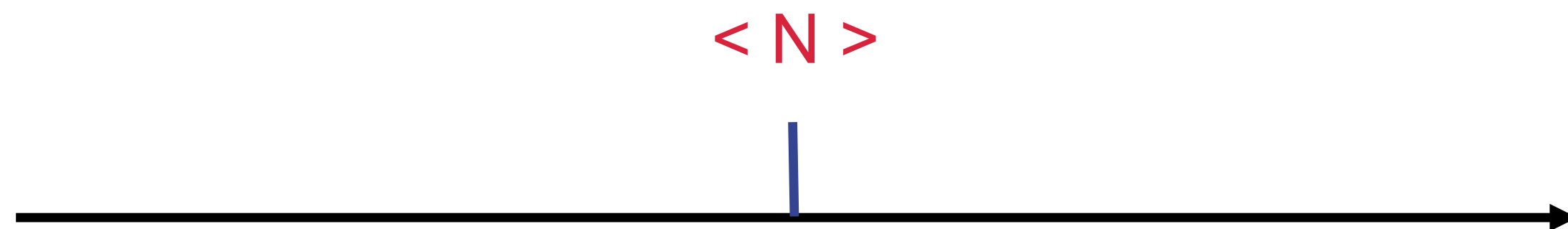
Days





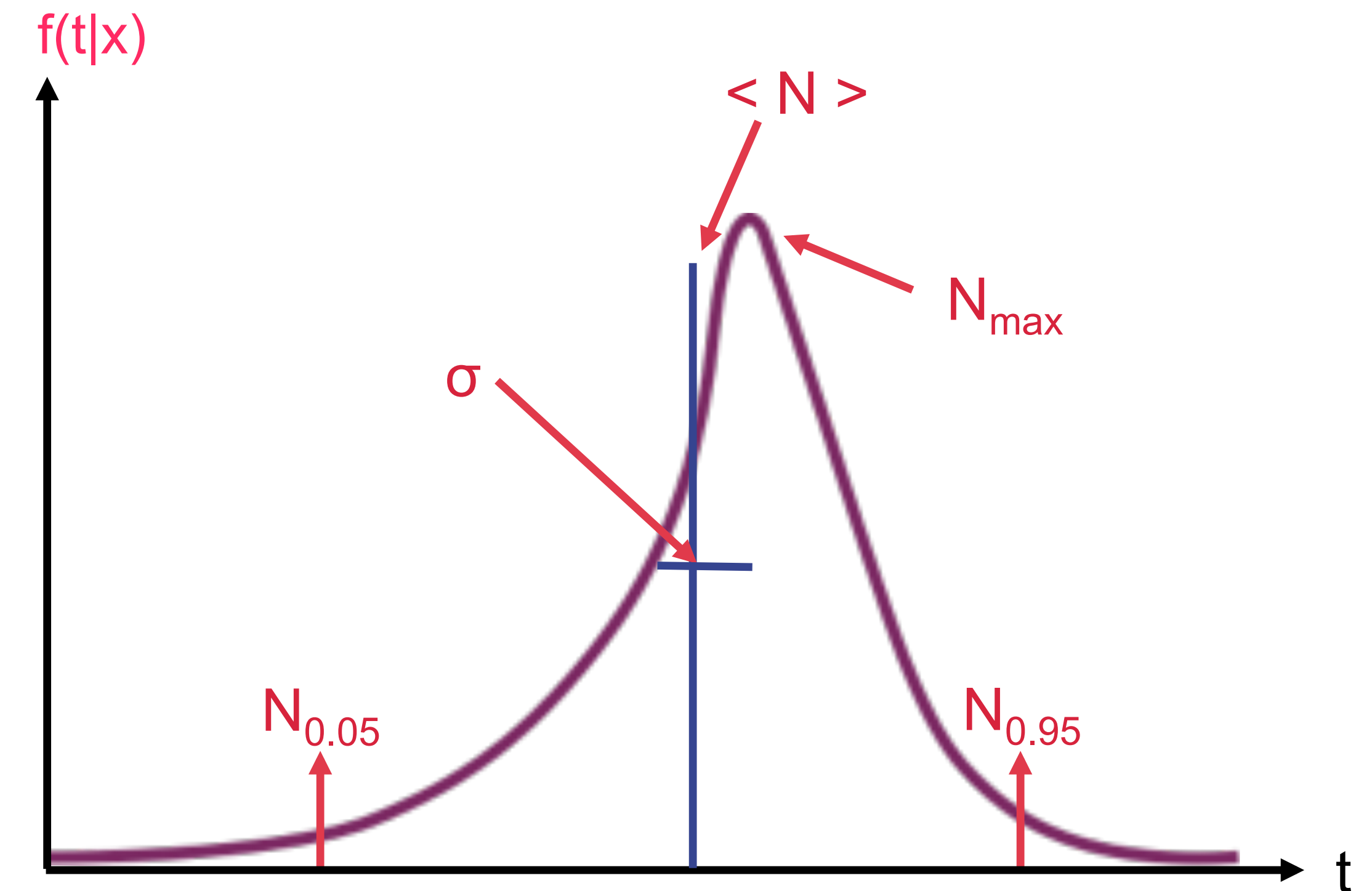
# Predictive Analytics

## STANDARD / LEGACY APPROACH



Prediction is a number (no uncertainty measure)

## BLUE YONDER APPROACH





















Prediction is complete conditional probability density function pdf  
allows risk management  
contains all information



# Bundesliga — Saison 2017/2018

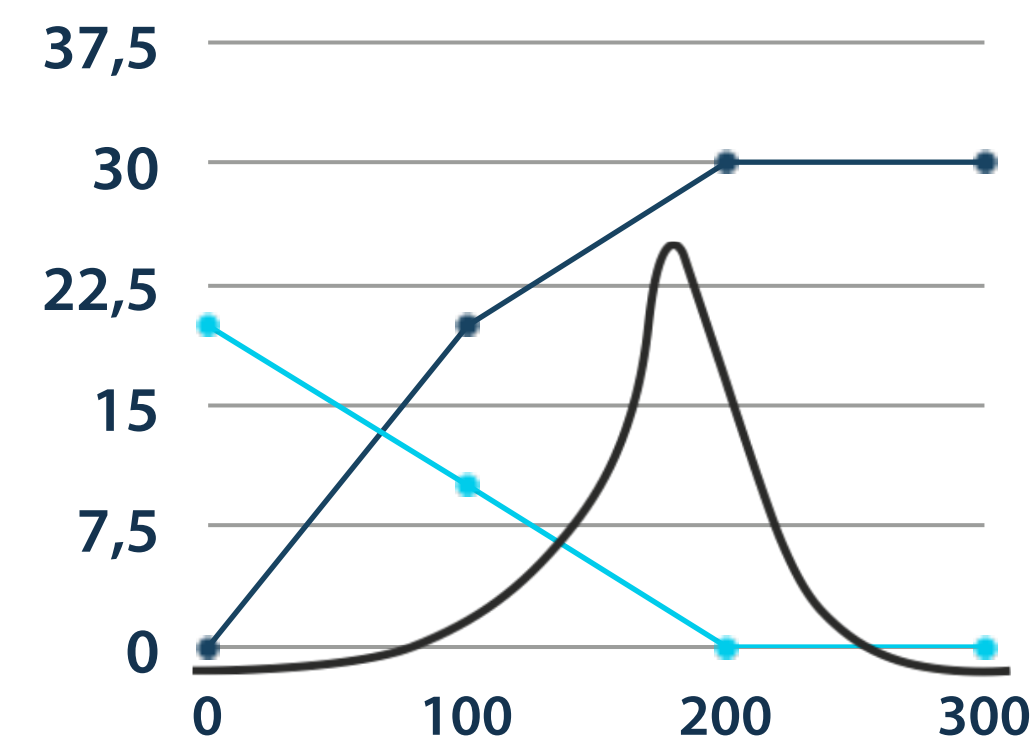
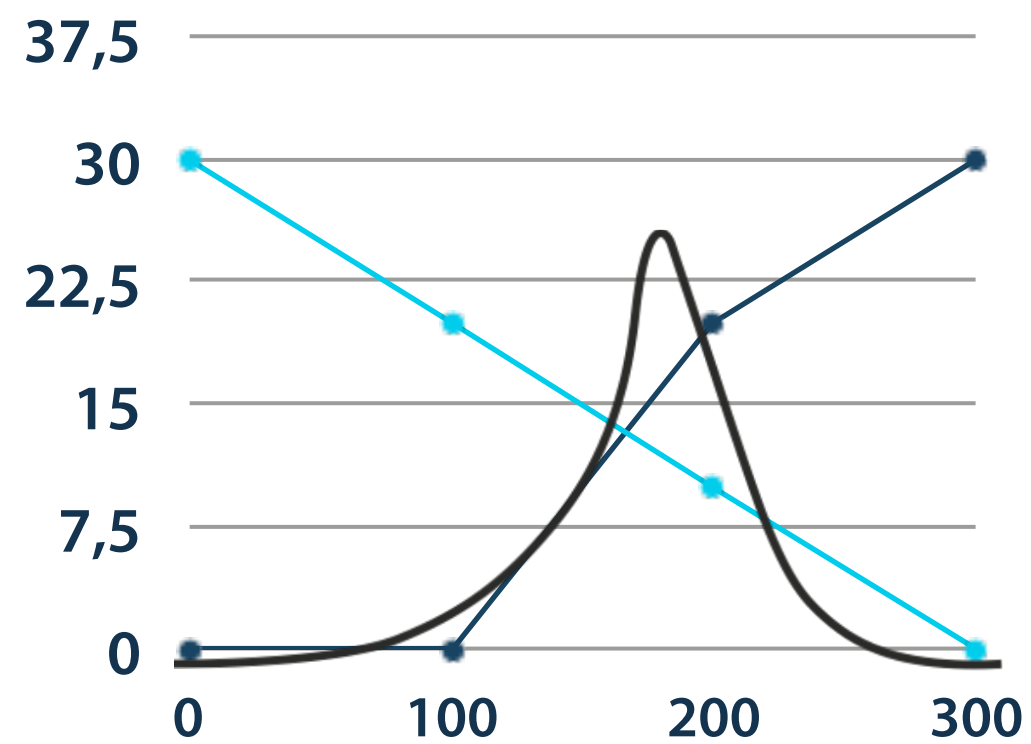
## Prognose für den 3. Spieltag



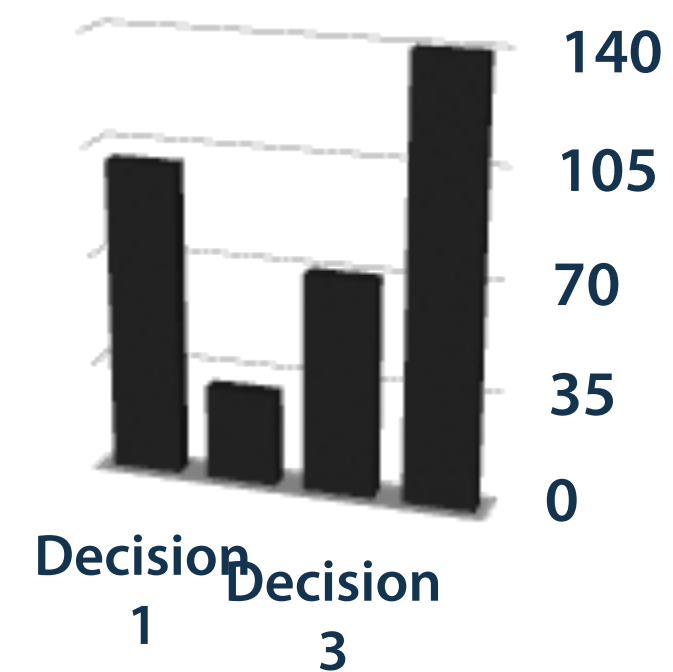
	Hamburger SV	32.3%	35.2%	32.5%	RB Leipzig	
	SC Freiburg	19.0%	26.5%	54.5%	Borussia Dortmund	
	Bor. M'gladbach	52.4%	32.4%	15.1%	Eintracht Frankfurt	
	FC Augsburg	46.1%	21.7%	32.3%	1.FC Köln	
	1.FC Mainz 05	39.3%	24.3%	36.4%	Bayer 04 Leverkusen	
	VfL Wolfsburg	55.9%	23.8%	20.3%	Hannover 96	
	1899 Hoffenheim	24.2%	22.2%	53.6%	FC Bayern München	
	Hertha BSC Berlin	54.2%	22.0%	23.8%	SV Werder Bremen	
	1.FC Schalke 04	64.6%	22.1%	13.3%	VfB Stuttgart	



# Prescriptive Analytics



**optimisation**

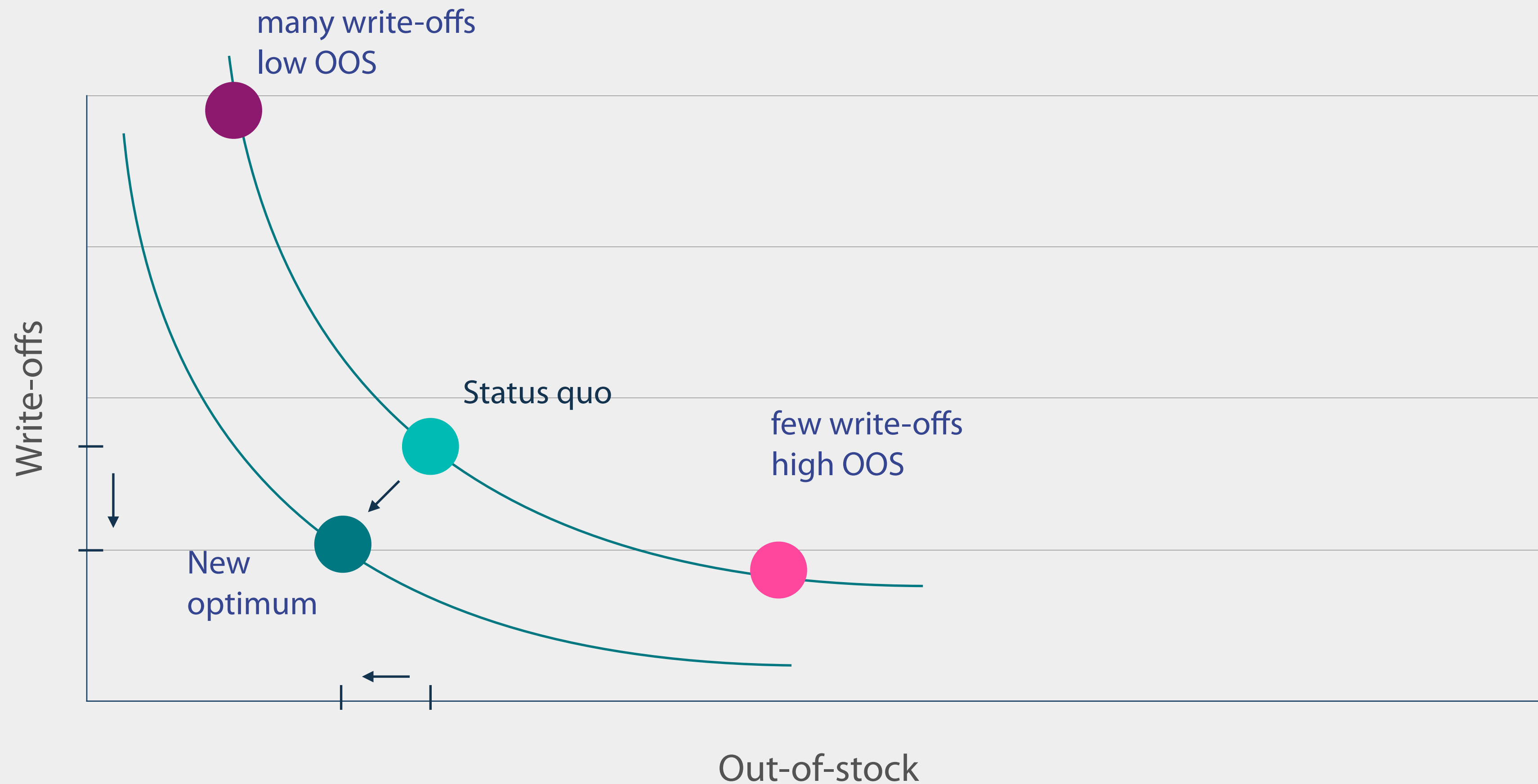


decide optimally, given KPI

Recipe  
order 8

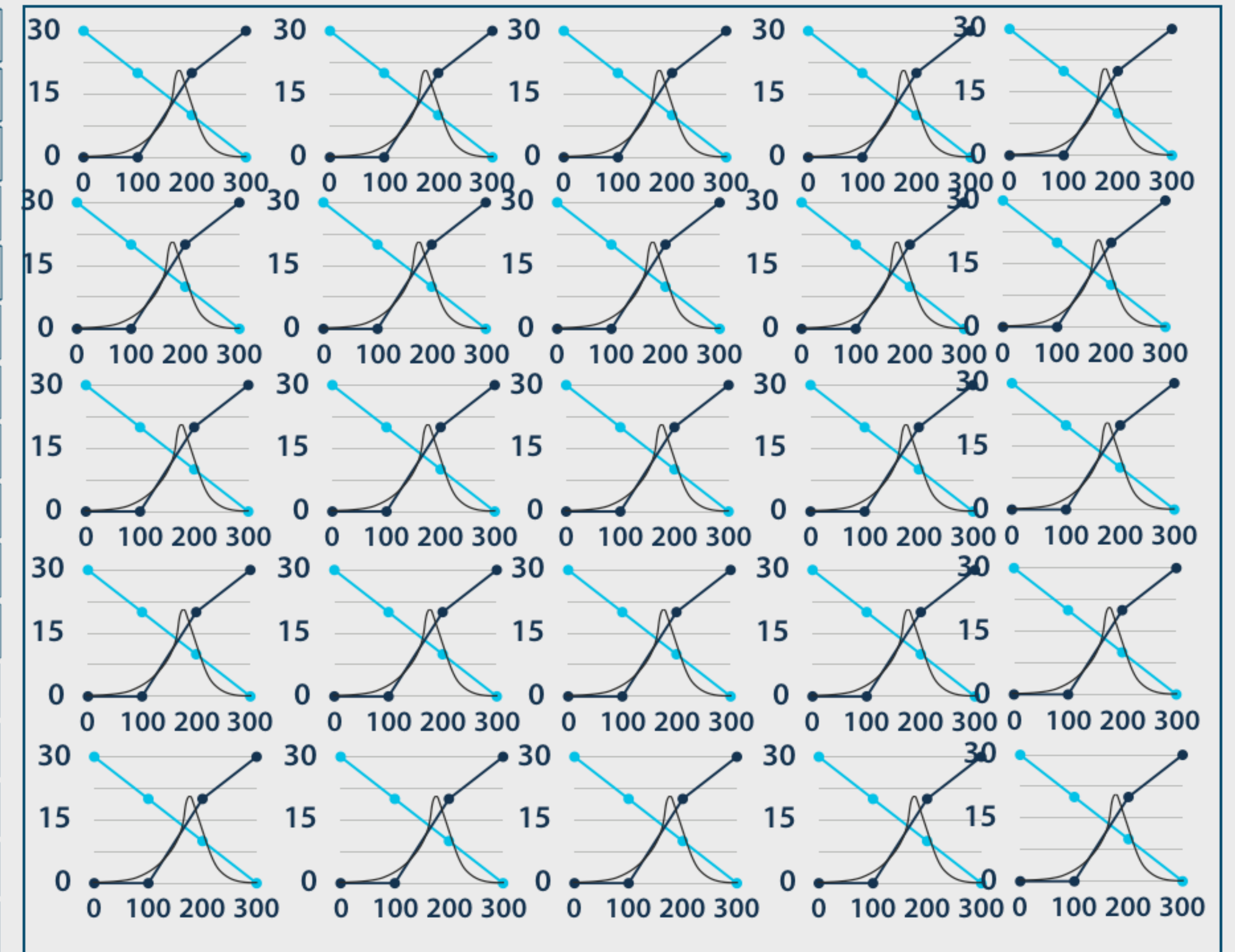
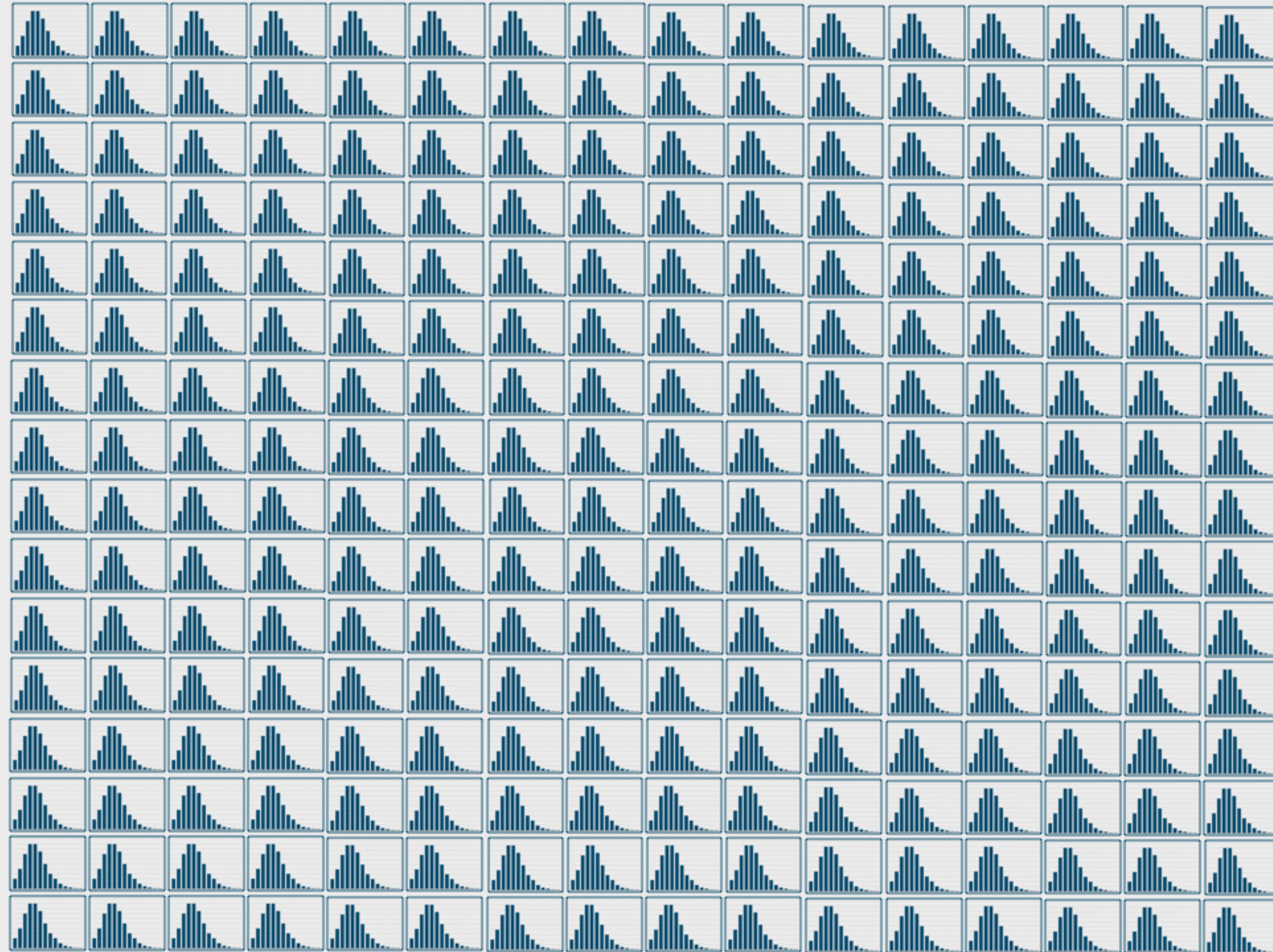


# We can improve mutually exclusive global goals simultaneously (portfolio theory)

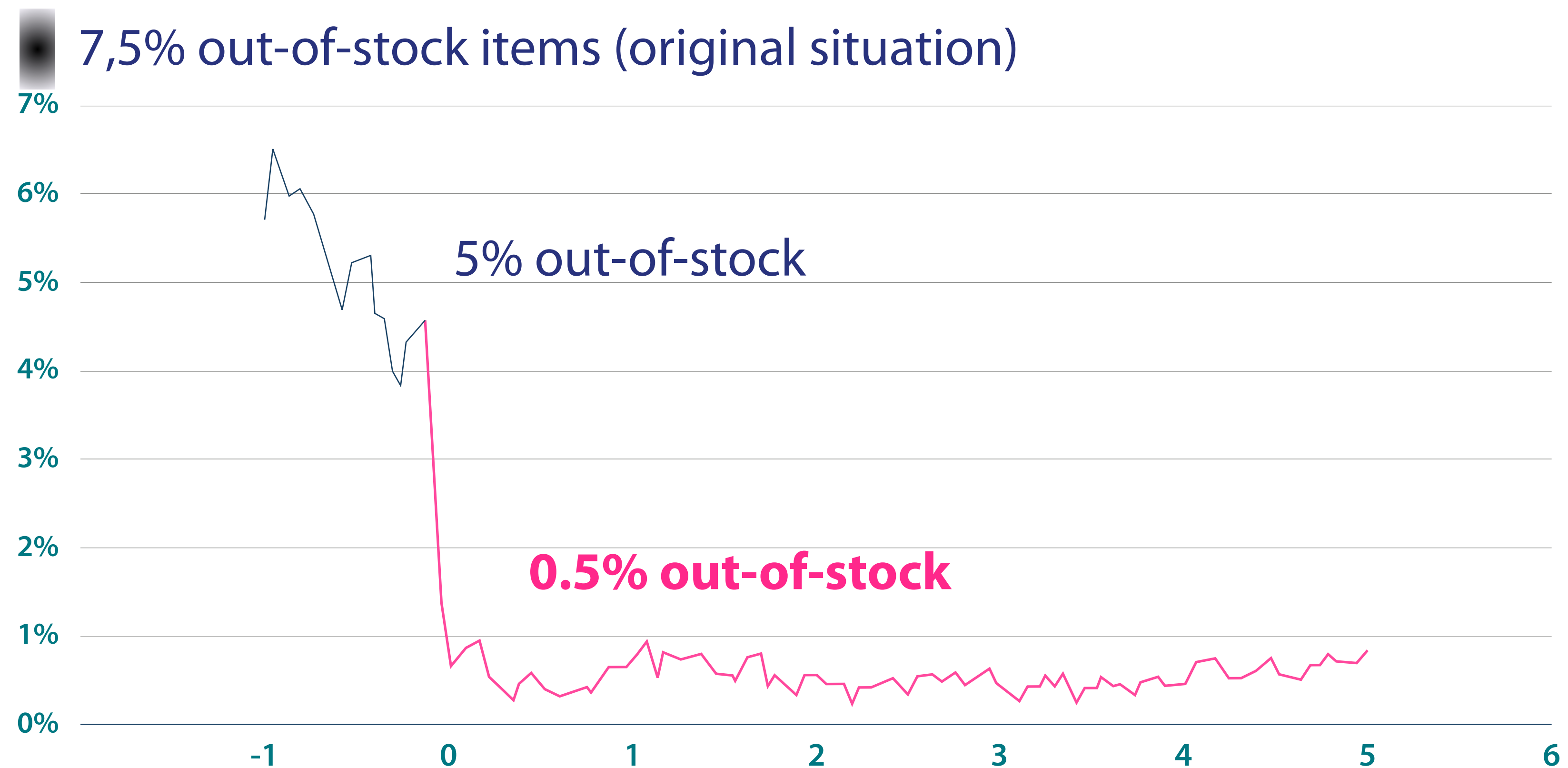




# Many predictions / optimisations / recipes —> automate!



# Influence of automation



OOS rate in German supermarket chain  
at constant overall stock level and waste rate





# AI-Supply Chain in Retail

99% • Automation



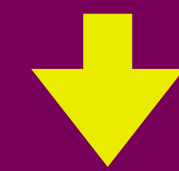
• Write-offs / Waste



• Freshness



• Capital



• out-of-stock



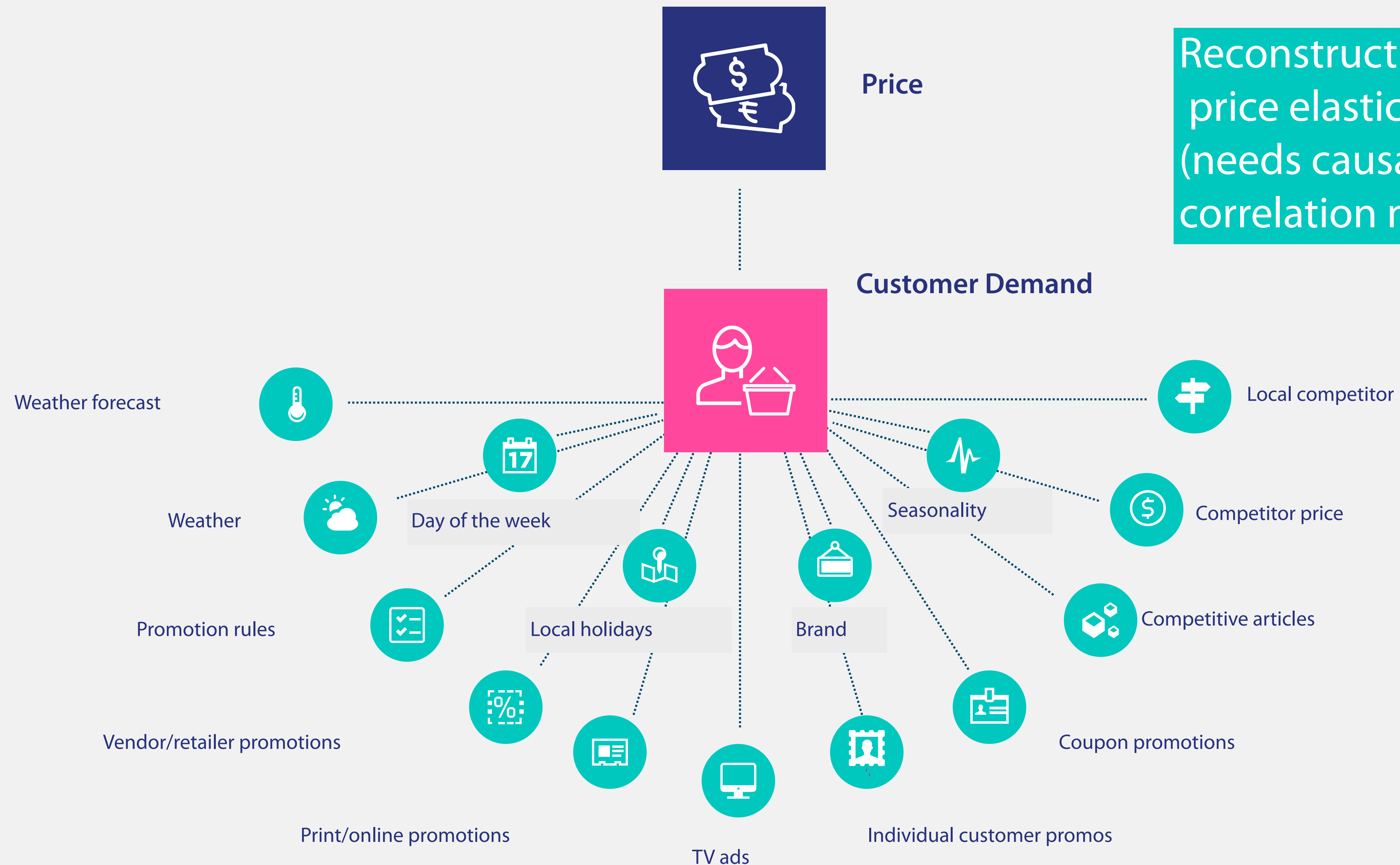
• Turnover



• Efficiency



# Effect of price on demand



Reconstruct conditional price elasticity curve  $D(p|X)$  (needs causality, correlation not enough)




# AI-pricing in retail

- 99% • Automation
- ↑ • market share
  - ↑ • turnover
  - ↑ • raw profit
  - ↑ • customers / new customers
  - ↓ • returns
  - 0 • complaints
  - ↓ • rests at end of season







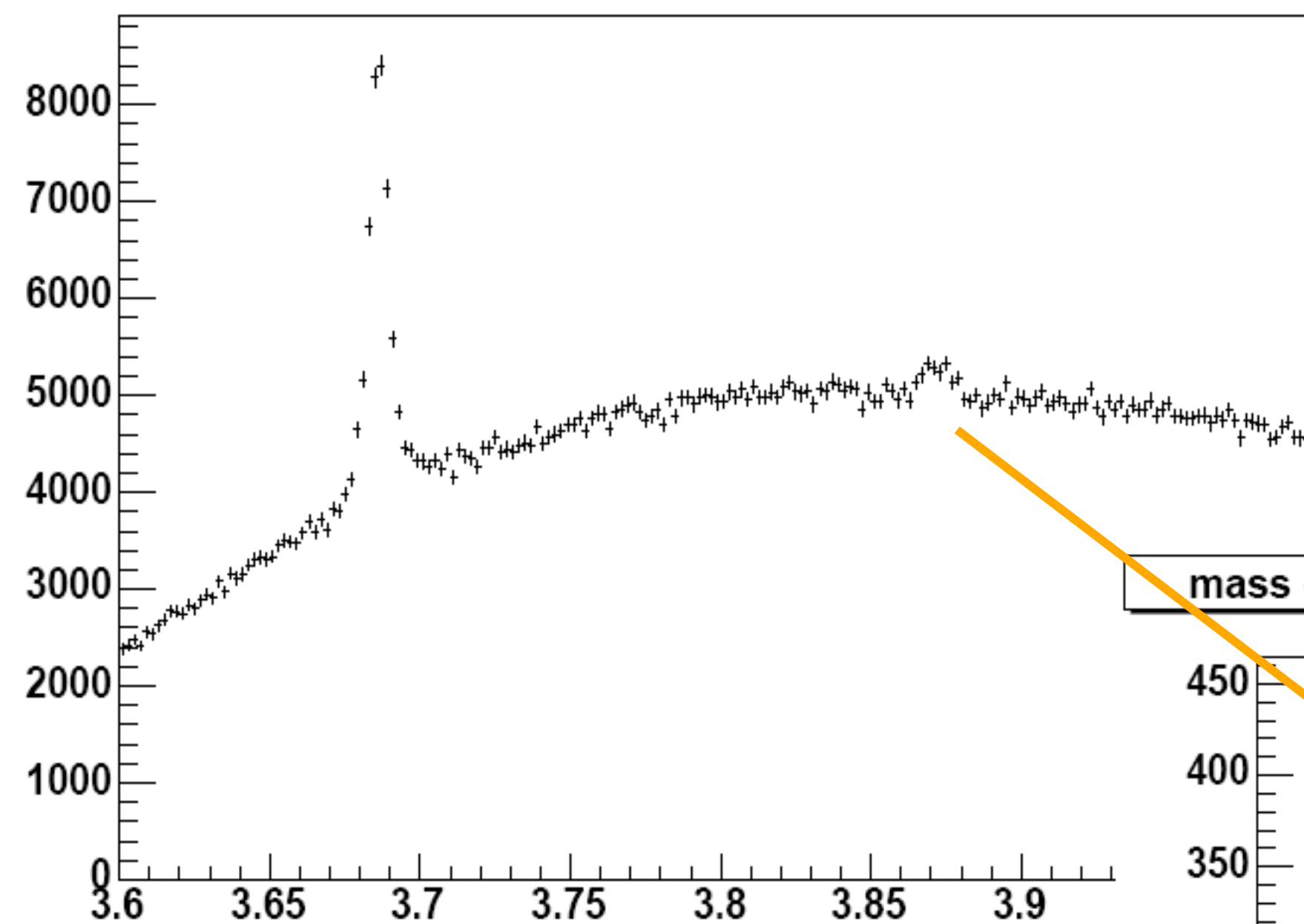
»For every complex problem  
there is an answer that is clear,  
simple, and wrong.«

H.L. Mencken

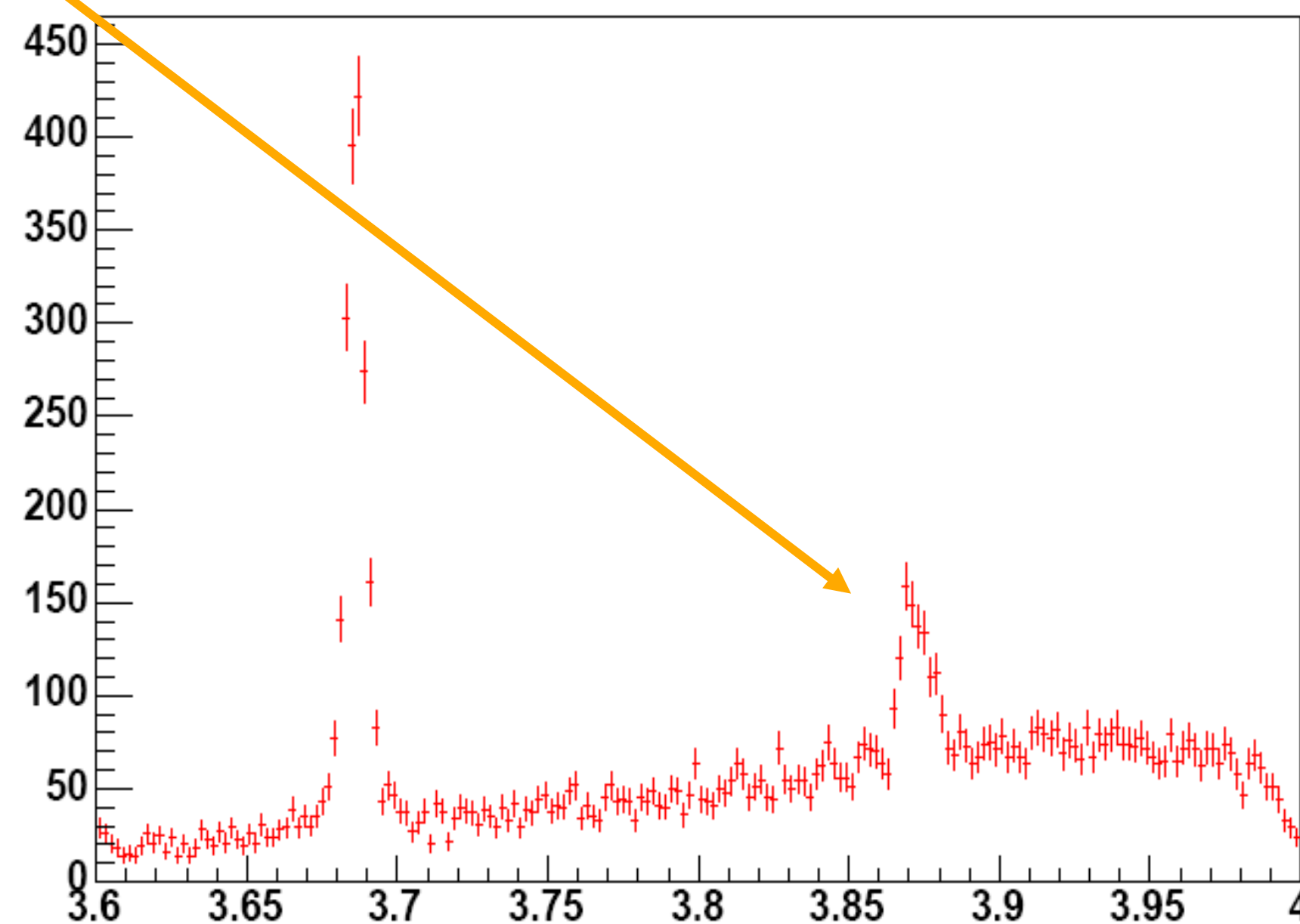
**CAUSALITY and CORRELATION**



mass (J/Psi Pi Pi)



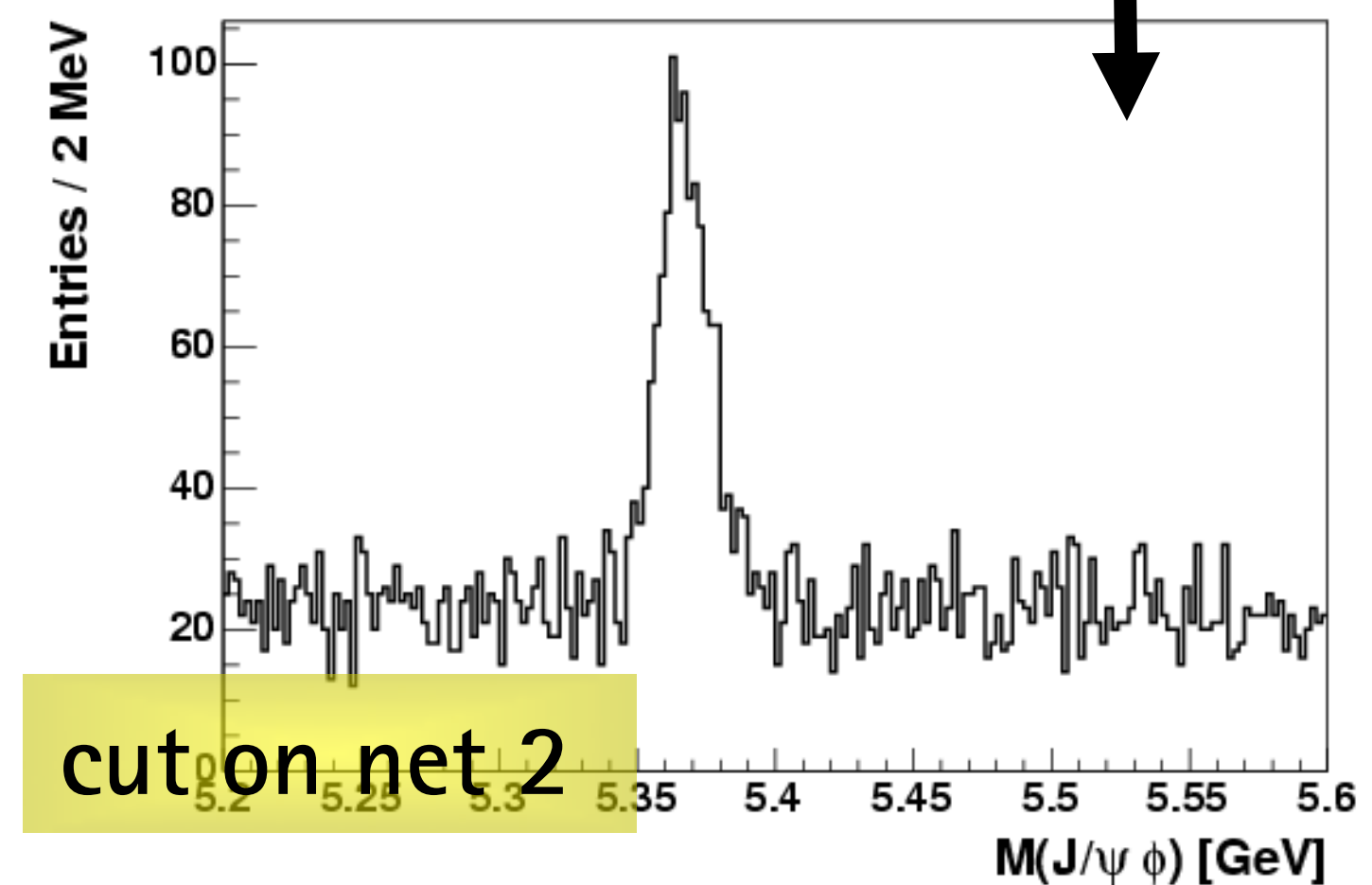
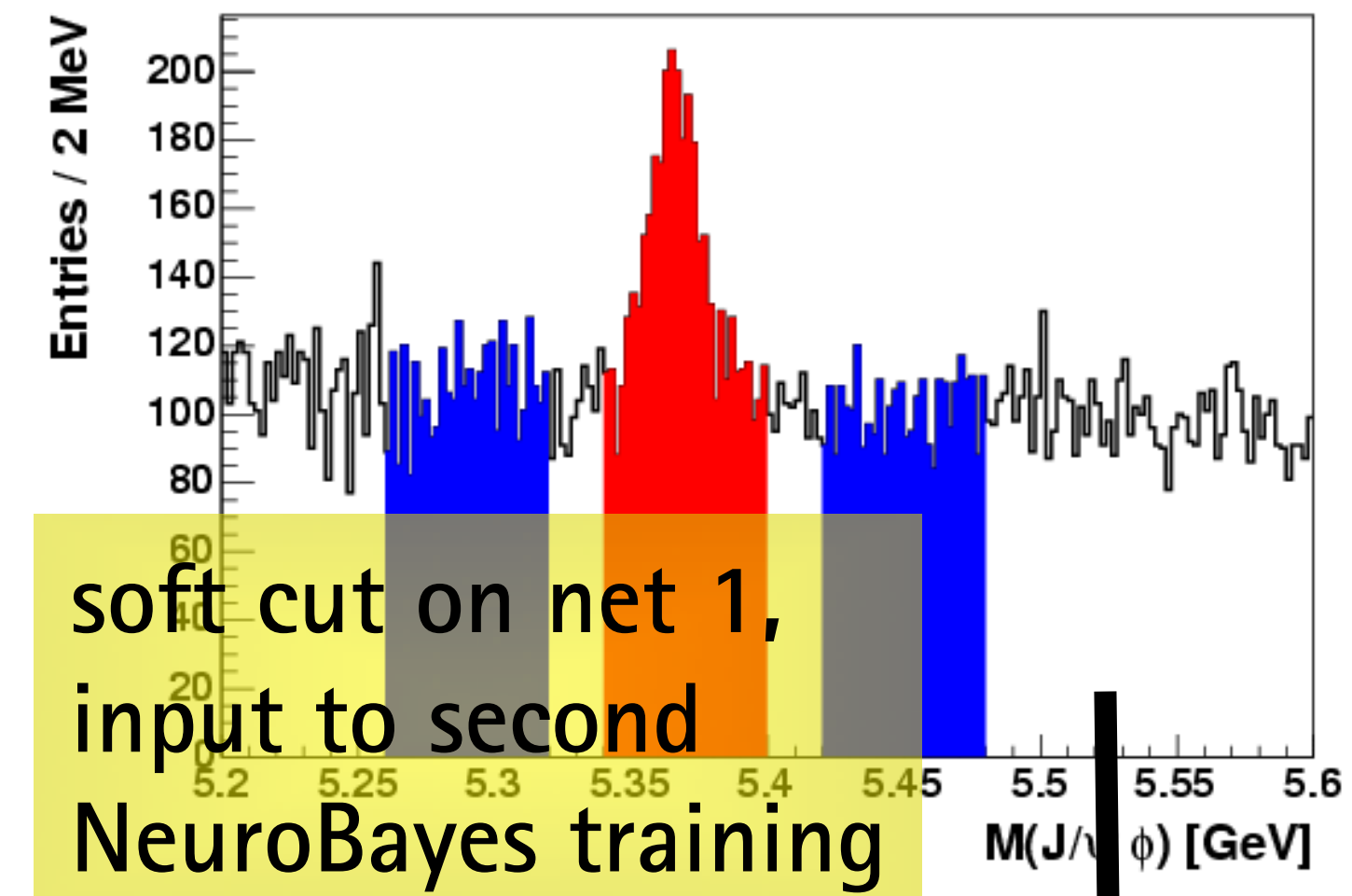
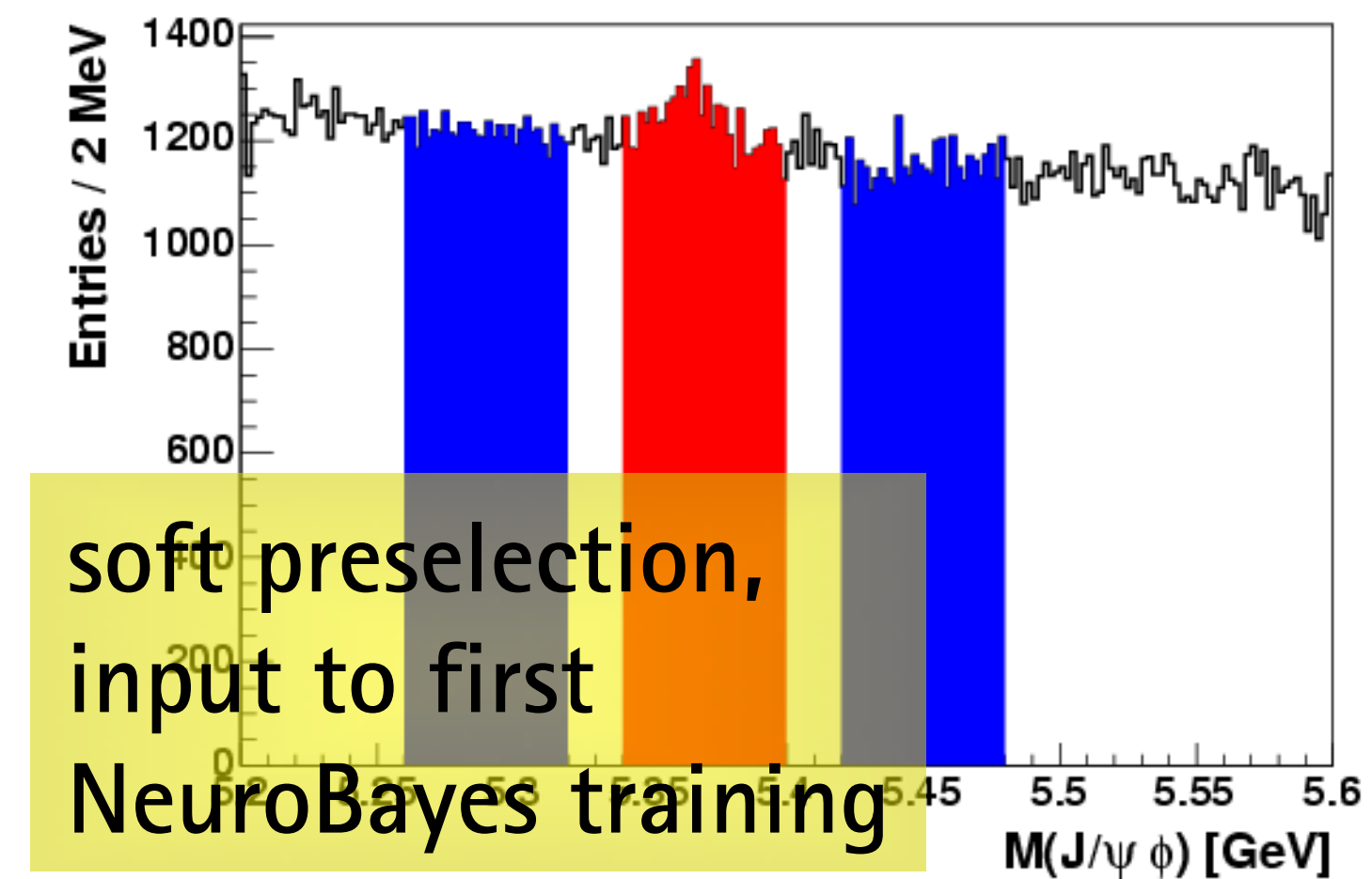
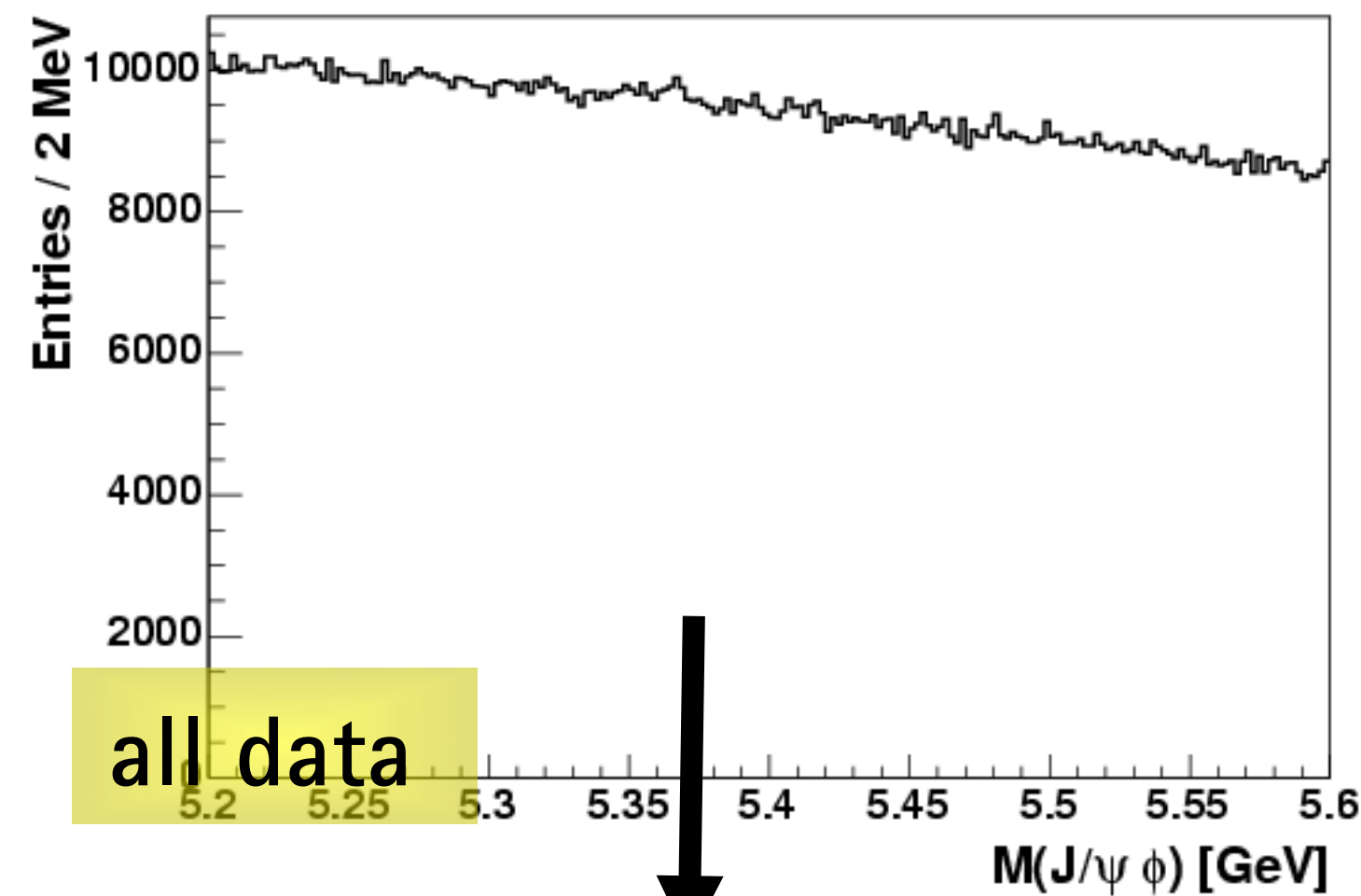
mass (J/Psi Pi Pi) &amp;&amp; nnOut &lt; -0.8



Very clear signal enrichment of the X(3872) (discovered 2004) by NeuroBayes



# NeuroBayes $B_s$ to $J/\psi \Phi$ selection without MC (2 stage background subtraction training process)



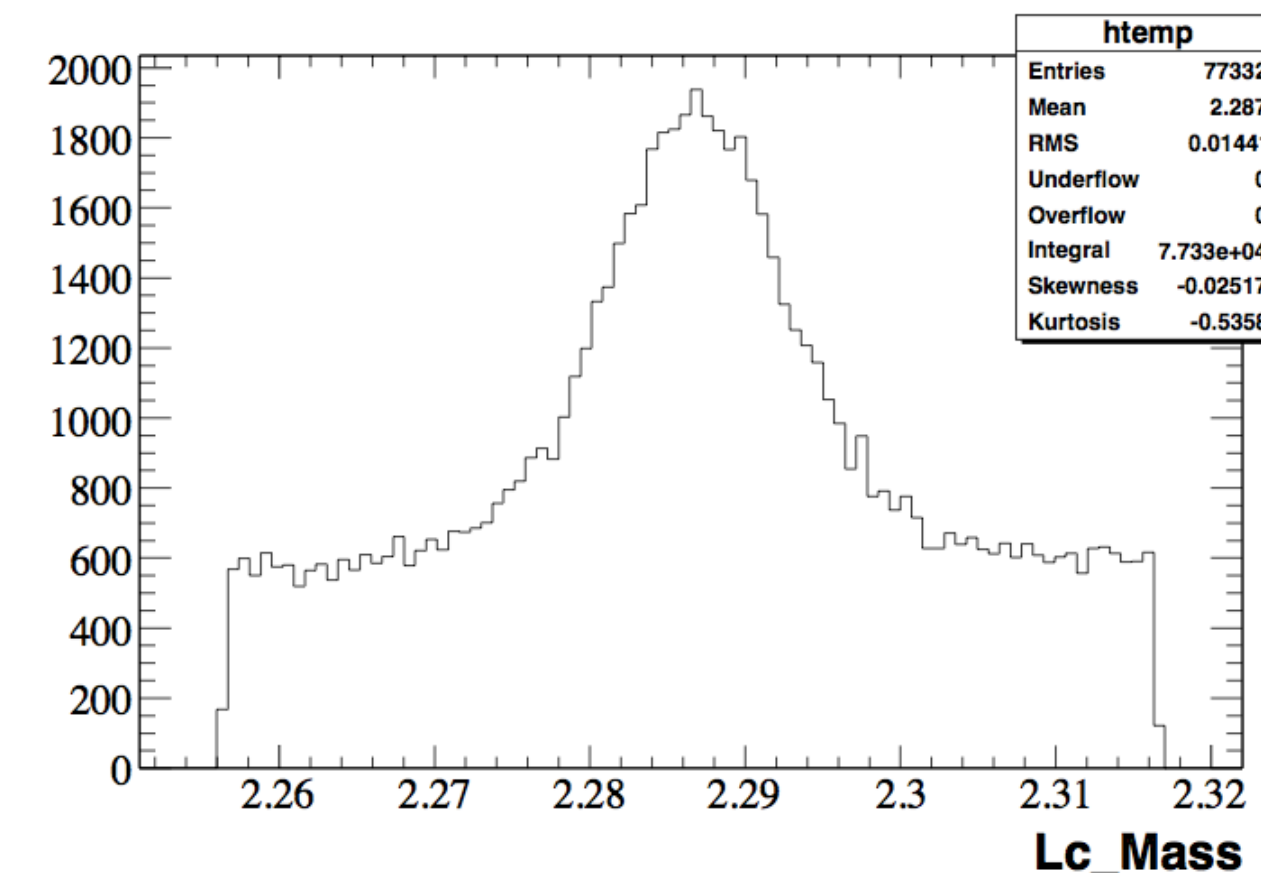
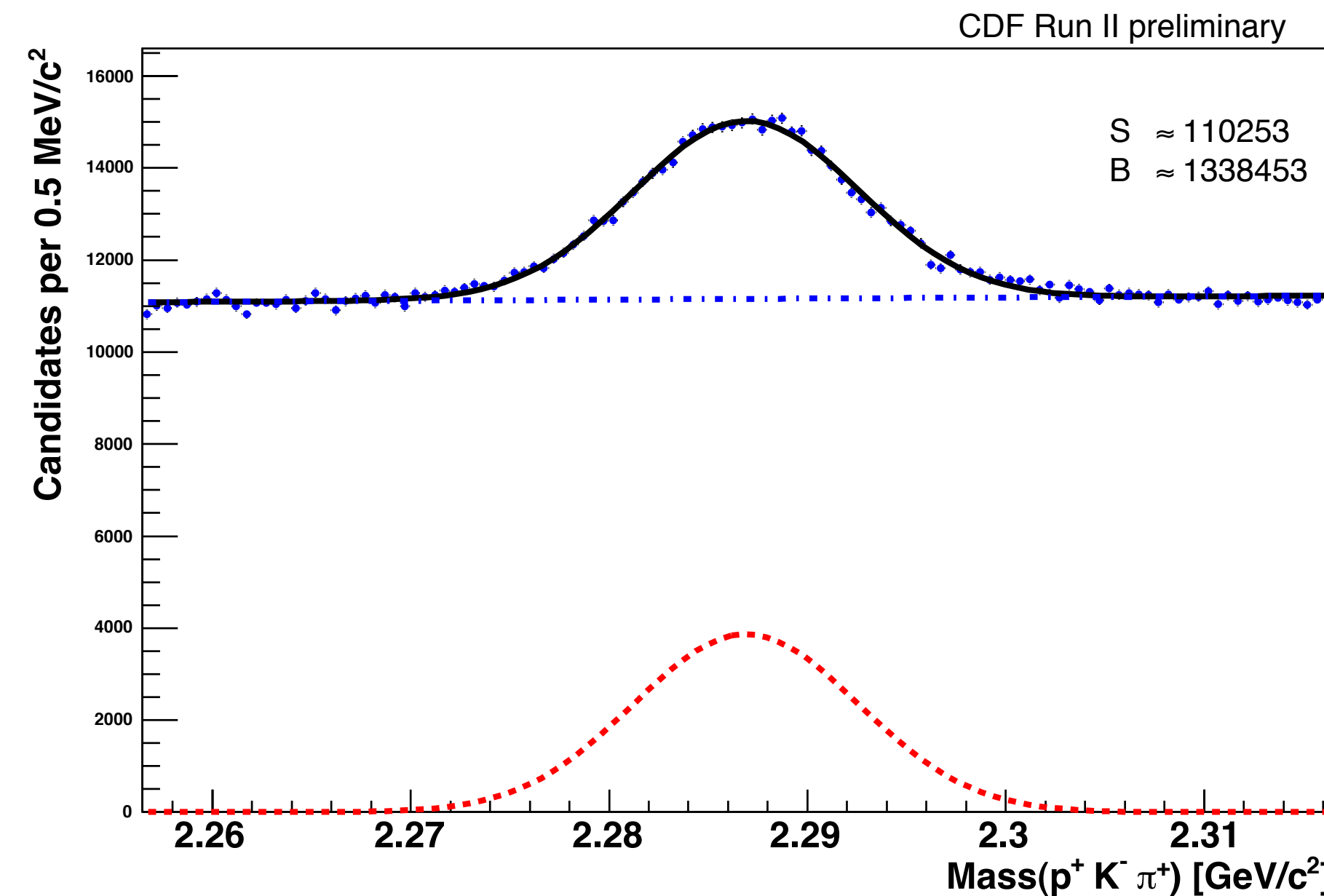


# Exploiting S/B information more efficiently : The sPlot-method

Fit data signal and background in one distribution (e.g. mass). Compute sPlot weights  $w_s$  for signal (may be  $<0$  or  $>1$ ) as function of mass from fit.

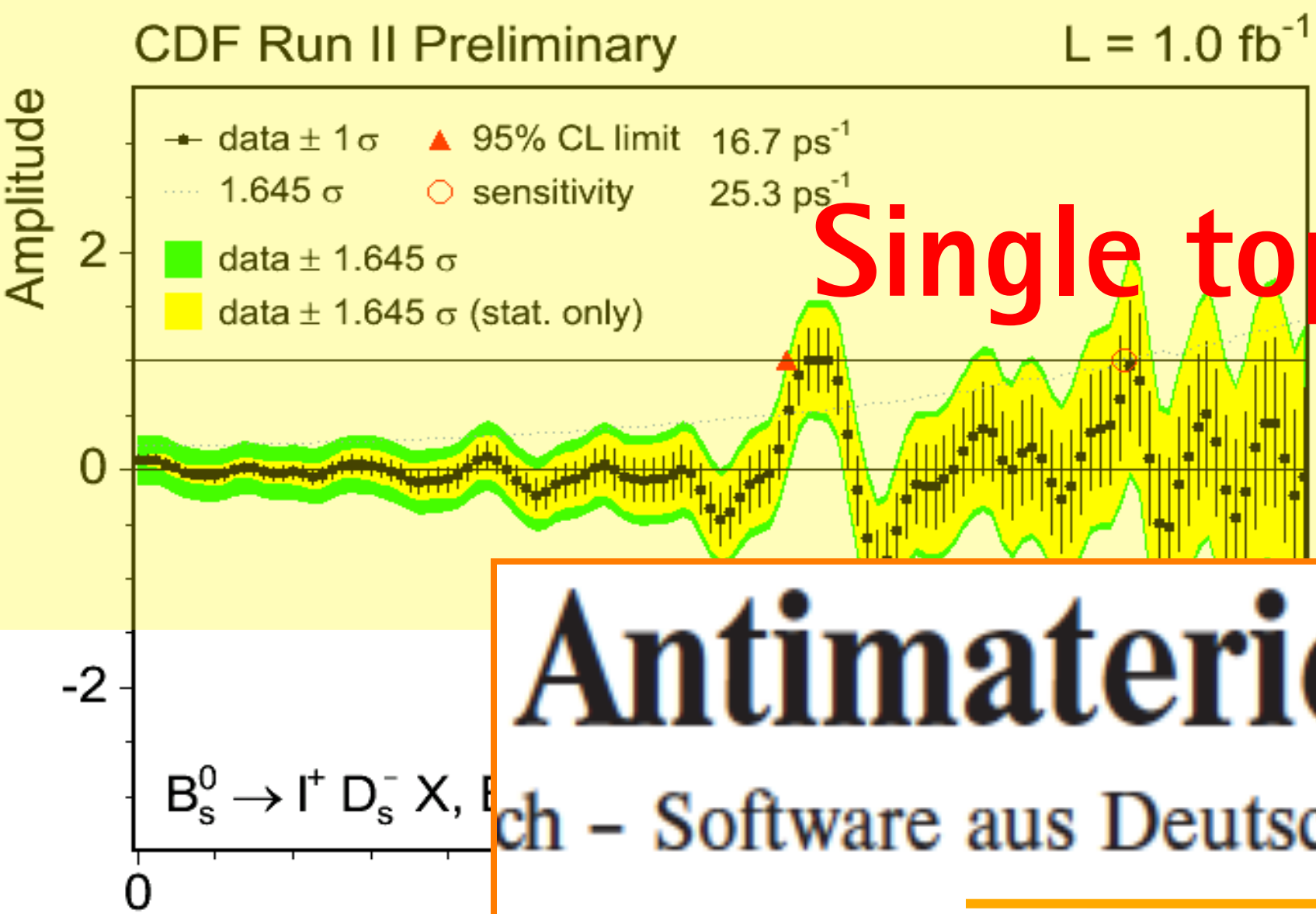
Train NeuroBayes network with each event treated both as signal with signal weight  $w_s$  and as background with weight  $1-w_s$ .

Soft cut on output enriches S/B considerably:  
Make sure network cannot learn mass!





Some NeuroBayes highlights: Bs oscillations  
Discovery of excited Bs states  
X(3872) properties  
Single top quark production discovery  
High mass Higgs exclusion



## Antimaterie wandeln

ch – Software aus Deutschland ermöglichte die Messungen

Sekunde 2,8 Bil-  
-B-Mesonen um  
ede Sekunde et-

Tempo dieses Tanzes gemessen“,  
sagt Jacobo Königsberg, Sprecher  
der CDF-Kollaboration.

kenntnisse über die Eigenschaften  
der Elementarteilchen, sondern  
auch über die Entwicklung des

von Medikamenten zu befragen. Der Arzt schrieb auf, was helfen sollte, und das besorgte man sich im „Krüdhüs“, für das sich schon im 13. Jahrhundert die Bezeichnung Apotheke einbürgerte. Das aus dem Griechischen entlehnte Wort bedeutet „Magazin“, ein „Ort, wo man etwas aufbewahrt“. In der Überlieferung der Familiennamen geschieht aber Seltsames. Während in den alten Urkunden die Berufsbezeichnung Apothekarius, eingedeutscht zu Apentegger und Aptecker, auch als Familienname oft vorkommt, gibt es ihn heute praktisch nicht mehr. Dafür ist an deutschen Apothekernamen kein Mangel: Apotheker waren die Vorfahren des Komponisten Peter Kreuder. Dazu gehört eine riesige Namenfamilie: Kräuter, Kreuter, Kreutler, Krütler, Kreudler, Krautner, Krütner, Krude, Kröder und Krüdener. Nicht zu vergessen die Würzler, Würzler, Würzner, Wurzer, Wurtz und Wurtz. Das bekannteste heimische heilkräftige Kraut, das Origani vulgare, auf deutsch: Dost, schenkte die Namen Dostler, Dostler und Dostmann. Der Salbenhersteller darf nicht fehlen: Schmer Schneider. Hans Markus Thomsen

Das unvollständige Skelett der Ur-Schlange Najash rionegrina wurde in Argentinien entdeckt. Es bedeutet für Paläobiologen eine Sensation. Die Wissenschaftler fanden bei dem sehr ursprünglichen Tier Knochen von kräftigen Beinen, ein Kreuzbein und einen Beckengürtel. FOTO: AP

Riesenschlangen gibt es Reste von Becken und Oberschenkel. DW

seit vielen Jahren bekannt, doch erst jetzt entdeckten die Forscher, daß es bei der Aktivierung von genetischen Programmen im Zellkern eine zentrale Rolle spielt und folglich auch an der Entstehung von Fehlfunktionen und Krebs beteiligt sein kann. AP

**Glückshormon gut für die Leber**  
Das Glückshormon Serotonin fördert die Regeneration von verletztem Lebergewebe. Dies beobachteten Forscher des Max-Planck-Instituts für molekulare Genetik und des Max-Deibäck-Centrums in Berlin. N.L.

**Musik lindert Schmerzen**  
Patienten, die während einer Operation in örtlicher Betäubung Musik hören, benötigen weniger Schmerzmittel, berichtet die Ärztezeitschrift „Praxis-Depesche“. Is.

Das Ressort Wissenschaft erreichen Sie unter:  
Telefon: 030 25 91 - 7 19 68  
Fax: 030 25 91 - 7 19 67  
E-Mail: [wissenschaft@welt.de](mailto:wissenschaft@welt.de)  
Internet: [www.welt.de/wissenschaft](http://www.welt.de/wissenschaft)

## Materie kann sich in Antimaterie wandeln

Amerikanische Elementarteilchenphysiker melden Durchbruch – Software aus Deutschland ermöglichte die Messungen

VON CHRISTIAN MEIER

**Chicago** – Elementarteilchenphysiker melden eine Sensation: Sie haben erstmals die Umwandlungen zwischen Materie und Antimaterie direkt beobachtet. Das seit vielen Jahren existierende Standardmodell der Teilchenphysik – also das vorherrschende Modell für die kleinsten Teilchen und der Kräfte zwischen ihnen – sagt voraus, daß so genannten B-Mesonen die einzigartige Fähigkeit besitzen, sich spontan in ihr Antiteilchen umwandeln zu können – und umgekehrt. Jetzt ist es US-Physikern am Fermilab bei Chicago gelungen, die extrem schnelle Umwandlung zeitlich aufgelöst zu beobachten und damit die theoretische Vorher-

sage experimentell zu bestätigen. Als einzige deutsche Institution war die Universität Karlsruhe maßgeblich an dem Experiment beteiligt. Zwanzig Physiker um Thomas Müller und Michael Feldt haben die komplexe Software für eine gezielte Auswertung der Rohdaten geliefert. Das Team gehört zu der Kollaboration „Collider Detector at Fermilab“ (CDF), an der etwa 700 Physiker von 60 Institutionen beteiligt sind. Im Fermilab, dem leistungsfähigsten Teilchenbeschleuniger der Welt, werden Protonen und Antiprotonen auf nahezu Lichtgeschwindigkeit beschleunigt und dann aufeinander geschossen. Die dabei neben vielen anderen Teilchen entstehenden B-Mesonen

wandeln sich pro Sekunde 2,8 Billionen Mal in Anti-B-Mesonen um und zurück, also jede Sekunde etwa 500-mal, wie viele Menschen auf der Erde leben. „Dieser Wert liegt

Tempo dieses Tanzes gemessen“, sagt Jacobo Königsberg, Sprecher der CDF-Kollaboration. B-Mesonen existieren im heutigen Kosmos nicht mehr, waren

kenntnisse über die Eigenschaften der Elementarteilchen, sondern auch über die Entwicklung des frühen Universums gewinnen. Seit 1995 arbeiteten die Karlsruher an Software, die aus dem Gewirr elektronischer Teilchenspuren im CDF-Detektor rekonstruieren kann, ob ein B-Meson bei seiner Entstehung Teilchen oder Antiteilchen war. Dies gelang mit komplexen statistischen Verfahren. Zusammen mit der Messung der Lebensdauer des B-Mesons (rund eine Millionstel Millionstel Sekunde) und der relativ einfach zu gewinnenden Information, ob es beim seinem Zerfall Teilchen oder Antiteilchen war, kann auf die Anzahl der Umwandlungen pro Sekunde geschlossen werden.

„Wenn man sich Materie vorstellt, die mit Antimaterie tanzt, dann haben wir das unglaubliche Tempo dieses Tanzes gemessen“

Jacobo Königsberg, Sprecher der Teilchenphysiker-Gruppe

Im Bereich, den das Standardmodell vorhersagt“, erläutert Müller. „Wenn man sich Materie vorstellt, die mit Antimaterie tanzt, dann haben wir das unglaubliche

aber im jungen Universum kurz nach dem Urknall vorhanden. Physiker können sie nur in großen Teilchenbeschleunigern untersuchen. Sie wollen so nicht nur Er-

**More than 200 Ph.D. theses and many publications ...**

from experiments DELPHI, CDF II, AMS, CMS ATLAS, LHCb and Belle used NeuroBayes<sup>®</sup> or predecessors very successfully.

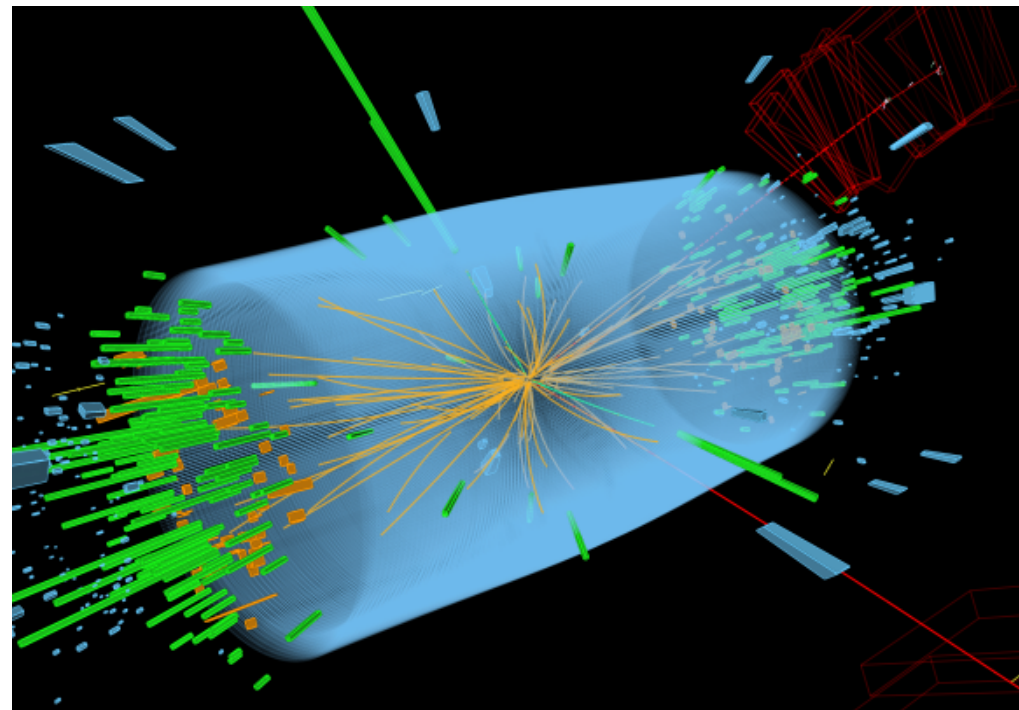
Many of these can be found at  
[www.neurobayes.de](http://www.neurobayes.de)

Talks about NeuroBayes<sup>®</sup> and applications:  
[www-ekp.physik.uni-karlsruhe.de/~feindt](http://www-ekp.physik.uni-karlsruhe.de/~feindt) → Forschung

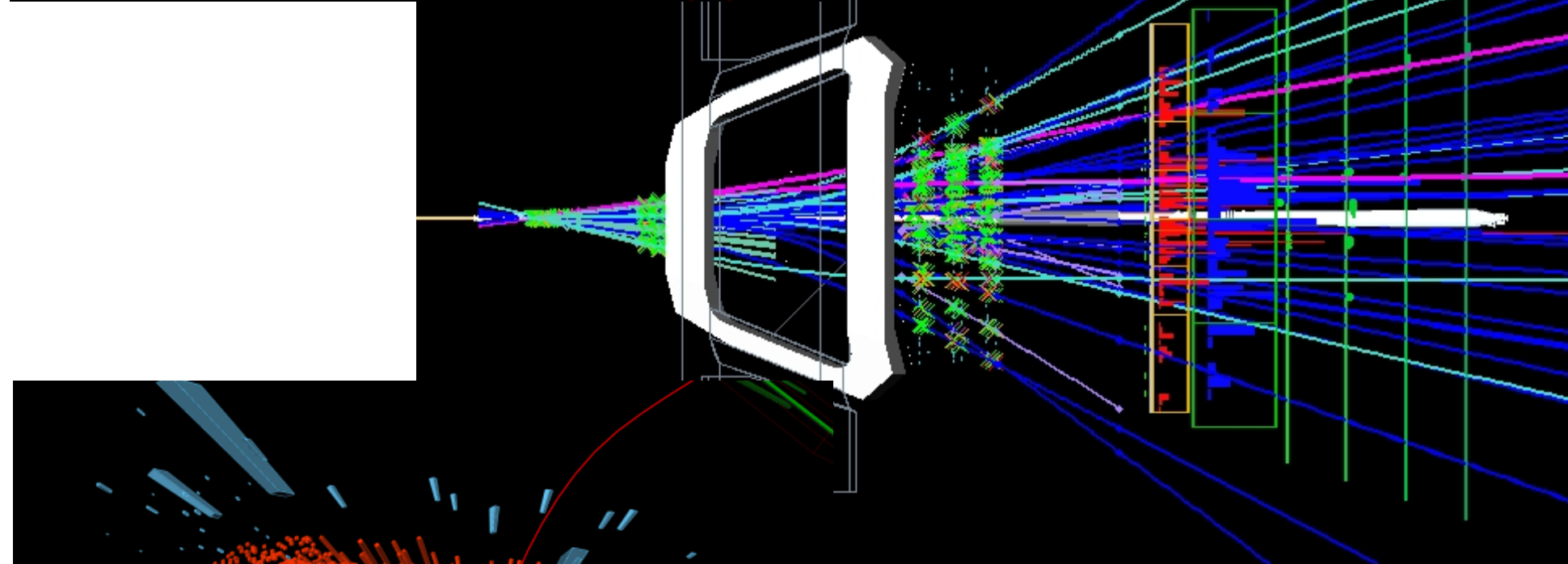


# NeuroBayes example: The LHCb trigger

very fast intelligent decisions with NeuroBayes

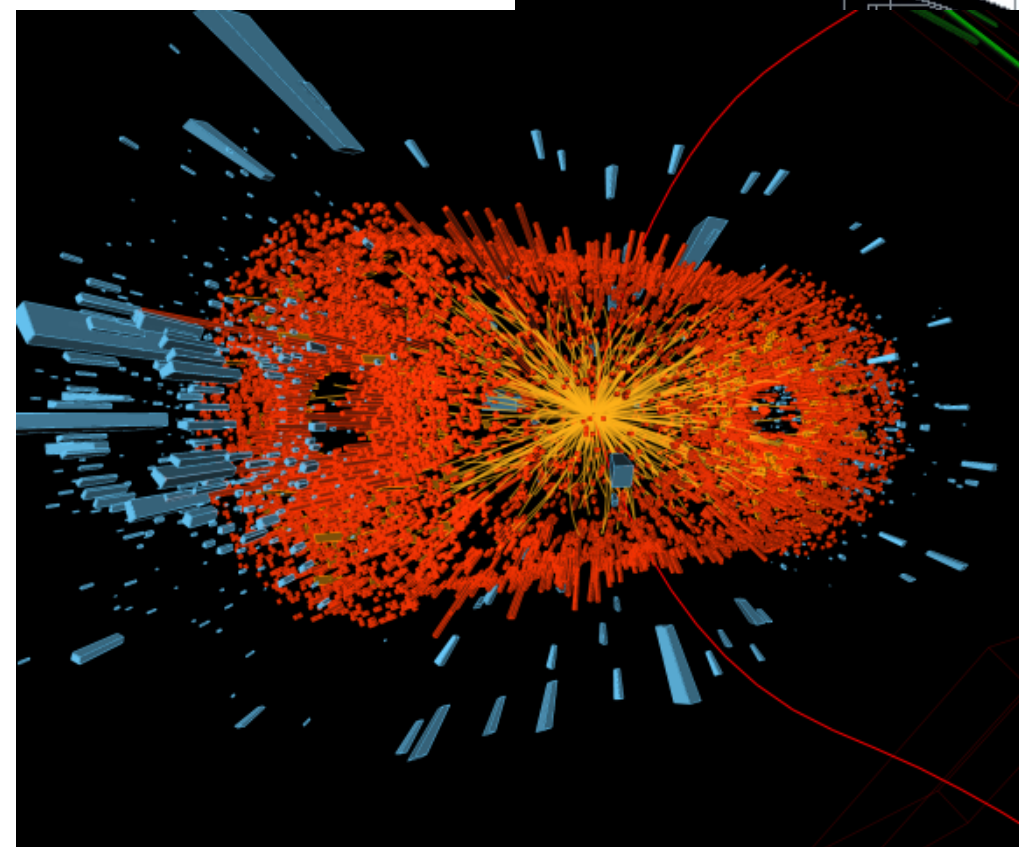


At the LHC (CERN) – per experiment:  
40 000 000 events per second, which translates into  
1 PetaByte (1,000,000,000,000,000 Byte)  
per second raw data



But only 1 PB of interesting data per  
year can be stored.

Need online reduction by  
1 : 10,000,000



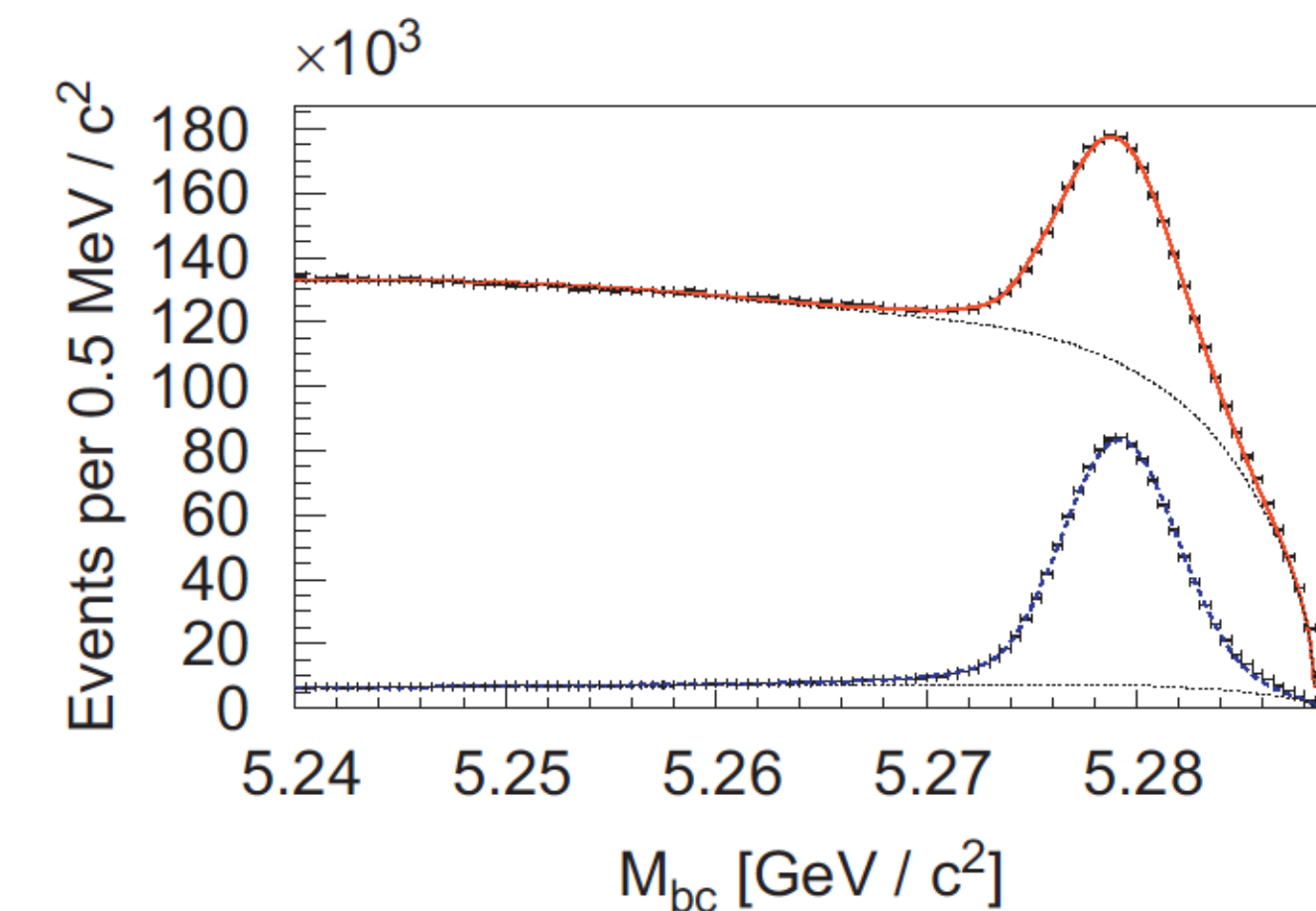
At the LHCb experiment  
30 000 instances of NeuroBayes running  
real-time 24/7 filter out the „interesting“  
events without introducing lifetime bias



# Prescriptive analytics even automates the work of 400 world-class research physicists: improvement $> +100\%$

Meta-Analysis: What does a scientist do in the analysis of data from particle collider experiment?

→ Automatic hierarchical reconstruction system with 72 NeuroBayes-networks  
reconstructs 1100 different reactions **with a factor 2 better efficiency relative to  
all analyses performed during 10 years by world-wide 400 physicists together!**



Nuclear Instruments and Methods in Physics Research A 654 (2011) 432–440

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in  
Physics Research A

journal homepage: [www.elsevier.com/locate/nima](http://www.elsevier.com/locate/nima)

ELSEVIER

A hierarchical NeuroBayes-based algorithm for full reconstruction  
of B mesons at B factories

M. Feindt, F. Keller, M. Kreps<sup>1</sup>, T. Kuhr, S. Neubauer\*, D. Zander, A. Zupanc

Institut für Experimentelle Kernphysik, Karlsruher Institut für Technologie, Campus Süd, Postfach 69 80, 76128 Karlsruhe, Germany

ARTICLE INFO

Article history:  
Received 7 April 2011  
Received in revised form  
3 June 2011  
Accepted 3 June 2011  
Available online 17 June 2011

Keywords:  
Full reconstruction  
B-factory  
Neural networks  
Probability

ABSTRACT

We describe a new B-meson full reconstruction algorithm designed for the Belle experiment at the B-factory KEKB, an asymmetric  $e^+e^-$  collider that collected a data sample of  $771.6 \times 10^6$   $B\bar{B}$  pairs during its running time. To maximize the number of reconstructed B decay channels, it utilizes a hierarchical reconstruction procedure and probabilistic calculus instead of classical selection cuts. The multivariate analysis package NeuroBayes was used extensively to hold the balance between highest possible efficiency, robustness and acceptable consumption of CPU time.

In total, 1104 exclusive decay channels were reconstructed, employing 71 neural networks altogether. Overall, we correctly reconstruct one  $B^\pm$  or  $B^0$  candidate in 0.28% or 0.18% of the  $B\bar{B}$  events, respectively. Compared to the cut-based classical reconstruction algorithm used at the Belle experiment, this is an improvement in efficiency by roughly a factor of 2, depending on the analysis considered.

The new framework also features the ability to choose the desired purity or efficiency of the fully reconstructed sample freely. If the same purity as for the classical full reconstruction code is desired ( $\sim 25\%$ ), the efficiency is still larger by nearly a factor of 2. If, on the other hand, the efficiency is chosen at a similar level as the classical full reconstruction, the purity rises from  $\sim 25\%$  to nearly 90%.

© 2011 Elsevier B.V. All rights reserved.

1. Full B meson reconstruction at B factories

1.1. The experimental setup

One of the biggest advantages of lepton colliders like the KEKB or

2. For the  $B^+B^-$  or  $B^0\bar{B}^0$  pairs produced in this two-body decay, the four-momenta are related by

$$p(B_1) + p(B_2) = p(e^+) + p(e^-). \quad (1)$$

3. The two B mesons are almost at rest in the center of mass

➤ Work by “Artificial Intelligence” and  
3 PhD students

➤ Corresponds to 500 “normal” PhD  
theses

➤ Corresponds to another 10 years of  
data taking (costs 700 M€)

Belle II: factor 4,  
full event interpretation ready to run directly at data taking (Th. Keck et al.)





# Predictive applications on a chip blue yonder & KIT

## Use case: Belle II pixel detector at KEK particle collider

- CERN: about 30.000 computers in parallel reduce data by about 1/100 million
- Next generation particle collider in Japan: so much data that it cannot be read out and distributed on computers any more.

## Solution:

- Implement Blue Yonder NeuroBayes decision algorithm on a chip
- Implement one such chip per sensor hardware module
- Chip decides which part of the detector is read out.
- World record: 8 billion decisions per second achieved.





# Neural networks

The NeuroBayes classification core is based on a simple feed forward neural network.

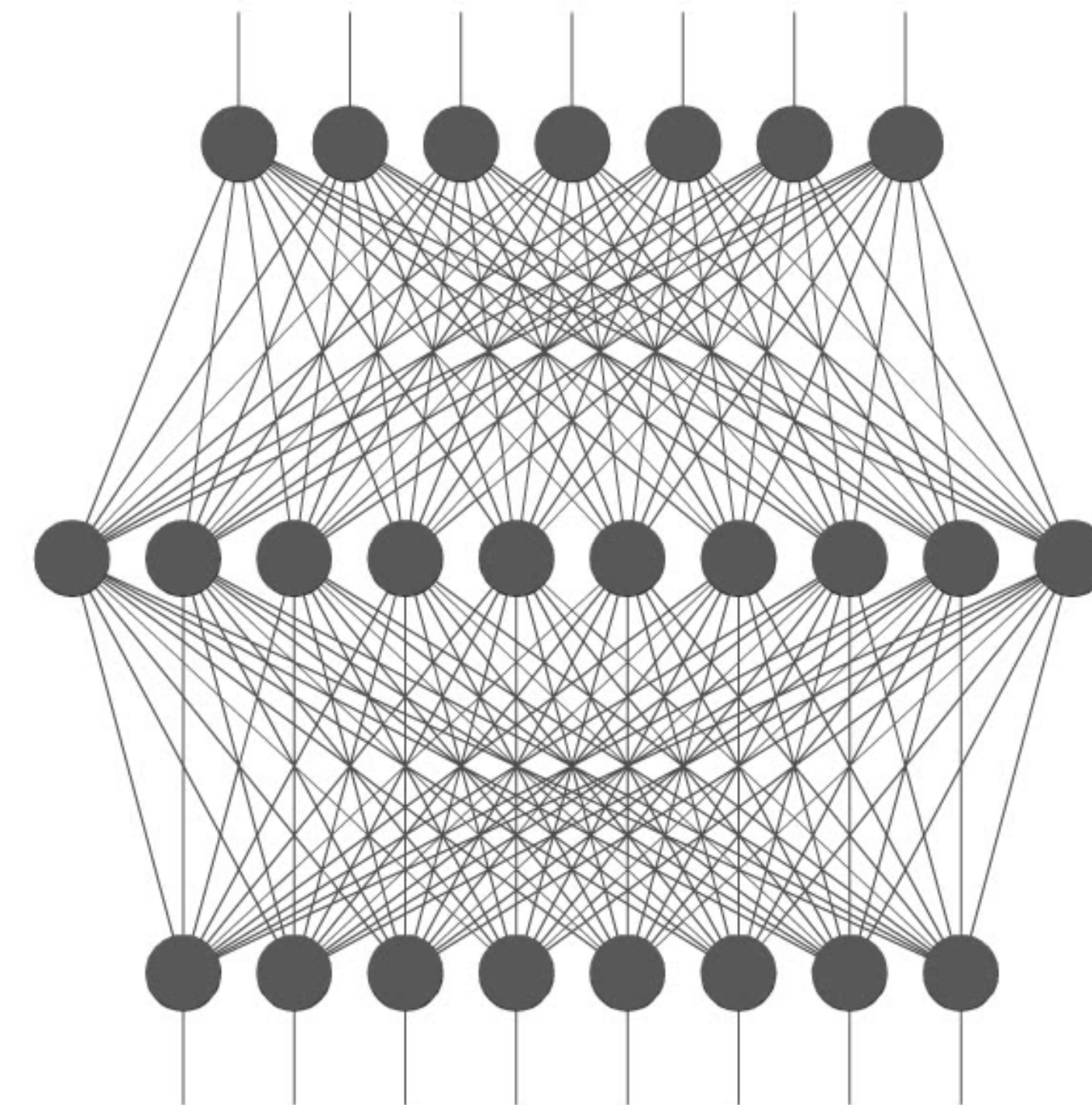
The information (the knowledge, the expertise) is coded in the connections between the neurons.

Each neuron performs fuzzy decisions.

A neural network can learn from examples. Supervised machine learning.

Human brain: about 100 billion (  $10^{11}$  ) neurons  
about 100 trillion (  $10^{14}$  ) connections

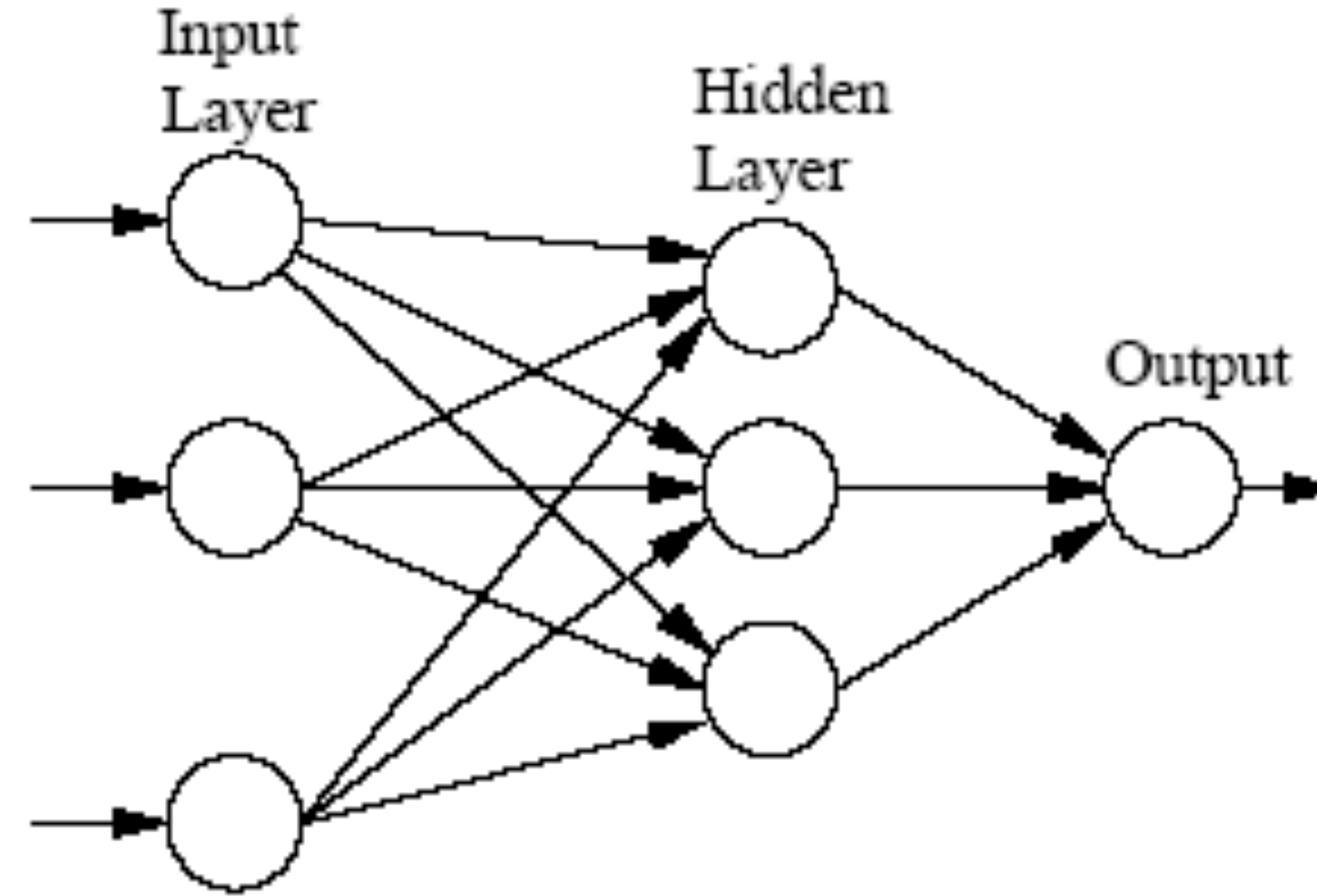
NeuroBayes : 10 to few 100 neurons





# Neural Network

## basic functions



The output of node  $j$  in layer  $n$  is calculated from weighted sum of outputs in layer  $n - 1$ :

$$x_j^{(n)} = f\left(\sum_i w_{i,j}^{(n)} x_i^{(n-1)} + w_{0,j}^{(n)}\right)$$

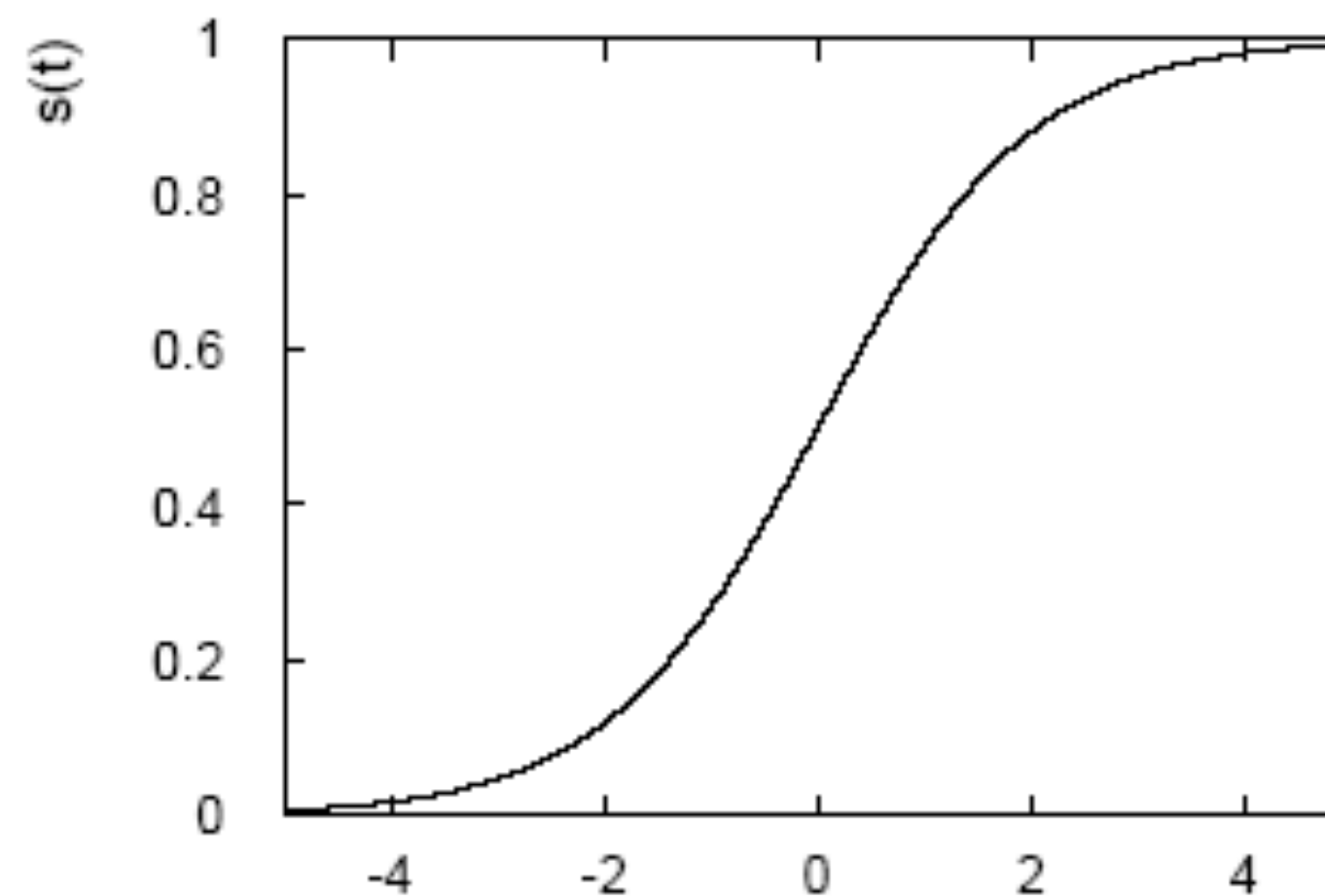
Each connection has associated a weight  $w_{i,j}^{(n)}$ , each node a bias  $w_{0,j}^{(n)}$ .

# Neural network transfer functions

A non-linear monotonuous transfer function  $f(x)$  is applied at the output of each node, e.g. the sigmoid function:

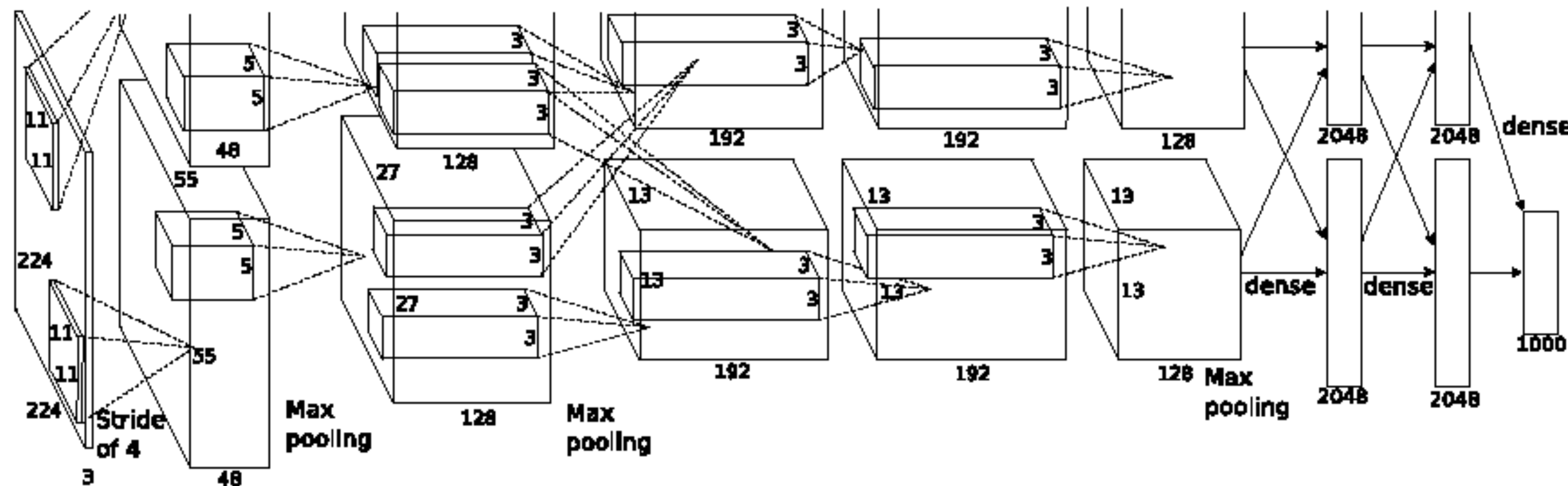
$$f(x) = \frac{1}{1 + \exp(-x)}$$

It maps the intervall  $(-\infty, \infty)$  to the compact  $(0, 1)$ .





Deep (and broad) architecture  
Convolutional layers, shared weights (e.g. imagenet)



GPU speedup typical 1GPU = 16 times faster than 20 CPU  
Still very time consuming. (Bosch example: predict with 19 Hz)

→ more specialized hardware? TPU, FPGA

Parallelisation in gradient calculation.

Synchronous and asynchronous data parallelization possible

Deep neural network:

Hope that it learns feature engineering.

In fact: it does. (It can do if you succeed to train well)

For the price of massively increased training computing time.

Our experience in physics research:

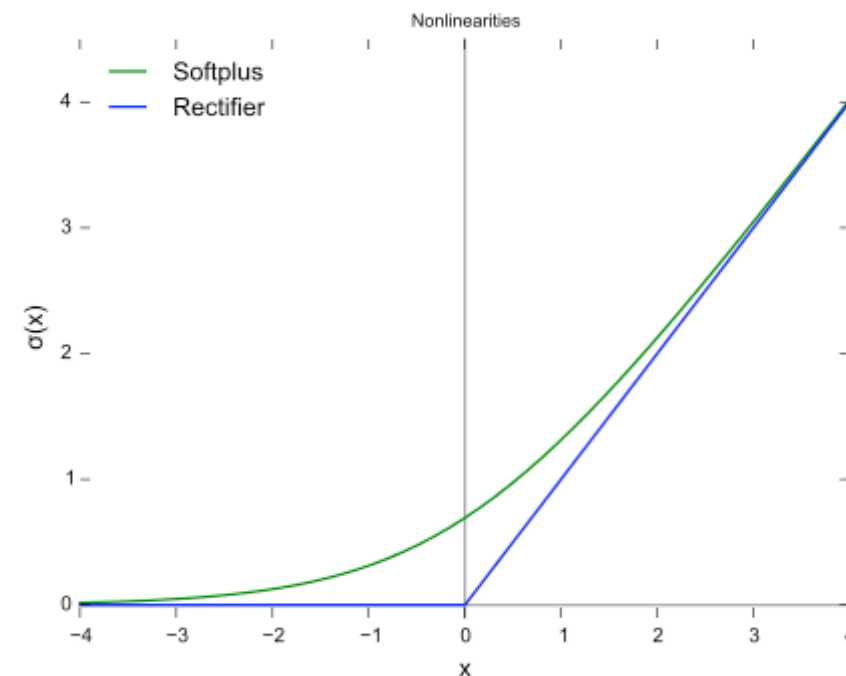
Not better than the features we designed with our large experience. But also not worse.

You exchange one kind of expertise with another.

Neural network learning is usually starting from scratch for each problem . Human: One network for all problems. Transfer learning.

Optimum: combine goodies of both to achieve best performance.





relu nonlinearity instead of sigmoid or RBF  
(or elu, maxout,... )

Minibatch training  
ADAM optimizer

Dropout regularisation,  
L1/L2 weight decay regularisation  
Preprocessing  
batch normalisation (= preprocessing in each layer)

History:  
Deep learning very difficult  
One layer after the other  
Non-linear dimension reduction by  
Boltzmann machines (unsupervised  
training between supervised  
iterations)

Physics:

Learn mainly from Monte Carlo simulation.

Economy:

Learn behavior of complex systems from historical data.

No „Standard Model“ available.

No proof that relations stay constant in time.

Automated science with very high SLA and quality requirements.

Language python (numpy, cython).

Trivial parallelization often possible by clustering, but not optimal.

DASK for parallel and out-of-core computing (lazy programming model).



Prof. Dr. Michael Feindt  
@M\_Feindt  
michael.feindt@blue-yonder.com  
www.blue-yonder.com

