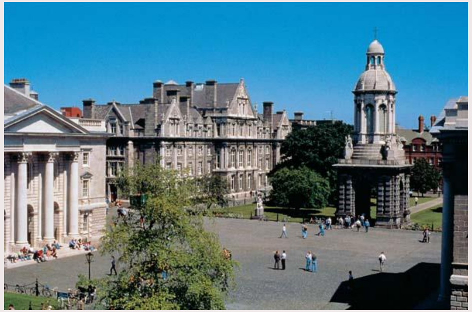
Mathematics: key enabling technology for scientific machine learning

IRISH NUMERICAL ANALYSIS FORUM SEMINAR
LIMERICK
FEBRUARY 23, 2023

Wil Schilders, TU Eindhoven & Hans Fischer senior fellow TUM-IAS
President-elect ICIAM

Ireland – second homeland for me



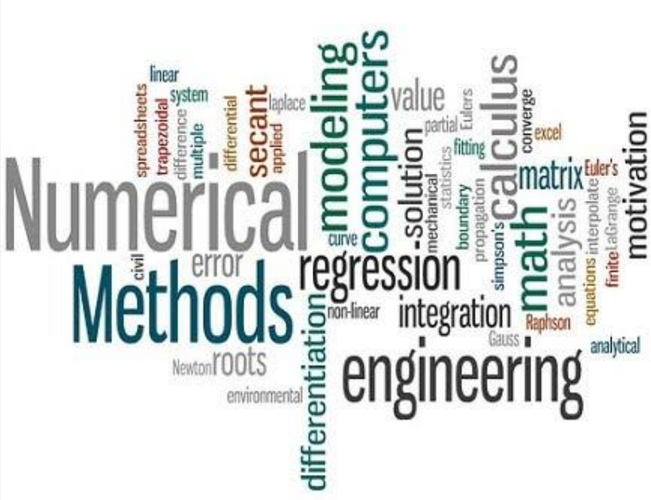
Contents

SOME OPENING THOUGHTS

A few years ago……..

…….I was thinking:

- Is numerical mathematics nearly finished?
- Do we see any new research directions, or is all research just an ''epsilon improvement'' of existing theories?
- Of course, much research was still carried out on interesting topics
  - We worked on model order reduction, the solution of indefinite linear systems and mimetic methods, with some new ideas; nice research, but not revolutionary (probably more evolutionary)
  - Also, new application areas required adaptation of existing methods, and sometimes entirely new techniques
  - Computational Science and Engineering meant working in interdisciplinary teams for mathematicians, adding a new dimension



TU/e

But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore's Law
    - Numerical methods needed to be made parallelizable

TU/e

But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore's Law
   - Numerical methods needed to be made parallelizable
   - ICCG, for example, shows a very bad performance on current supercomputers

HPCG Benchmark June 2019

Rank	Site	Computer	Cores	HPL Rmax (Pflop/s)	TOP500 Rank	HPCG (Pflop/s)	Fraction of Peak
1	DOE/SC/ORNL USA	**Summit**, AC922, IBM POWER9 22C 3.7GHz, Dual-rail Mellanox FDR, NVIDIA Volta V100, IBM	2,397,824	148.60	1	2.926	1.5%
2	DOE/NNSA/LLNL USA	**Sierra**, S922LC, IBM POWER9 20C 3.1 GHz, Mellanox EDR, NVIDIA Volta V100, IBM	1,572,480	94.64	2	1.796	1.4%
3	RIKEN Advanced Institute for Computational Science Japan	**K computer**, SPARC64 VIIIfx 2.0GHz, Tofu interconnect, Fujitsu	705,024	10.51	18	0.603	5.3%
4	DOE/NNSA/LANL/SNL USA	**Trinity**, Cray XC40, Intel Xeon E5-2698 v3 16C 2.3GHz, Aries, Cray	979,072	20.16	6	0.546	1.3%
5	Natl. Inst. Adv. Industrial Sci. and Tech. (AIST) Japan	**ABCI**, PRIMERGY CX2570M4, Intel Xeon Gold 6148 20C 2.4GHz, Infiniband EDR, NVIDIA Tesla V100, Fujitsu	368,640	16.86	10	0.509	1.7%
6	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint**, Cray XC50, Intel Xeon E5-2690v3 12C 2.6GHz, Cray Aries, NVIDIA Tesla P100 16GB, Cray	387,872	21.23	5	0.497	1.8%
7	National Supercomputing Center in Wuxi China	**Sunway** TaihuLight, Sunway MPP, SW26010 260C 1.45GHz, Sunway, NRCPC	10,649,600	93.02	3	0.481	0.4%
8	Korea Institute of Science and Technology Information Republic of Korea	**Nurion**, CS500, Intel Xeon Phi 7250 68C 563584C 1.4GHz, Intel Omni-Path, Intel Xeon Phi 7250, Cray	570,020	13.93	13	0.391	1.5%
9	Joint Center for Advanced High Performance Computing Japan	**Oakforest-PACS**, PRIMERGY CX600 M1, Intel Xeon Phi Processor 7250 68C 1.4GHz, Intel Omni-Path Architecture, Fujitsu	556,104	13.55	14	0.385	1.5%
10	DOE/SC/LBNL/NERSC USA	**Cori**, XC40, Intel Xeon Phi 7250 68C 1.4GHz, Cray Aries, Cray	622,336	14.02	12	0.355	1.3%

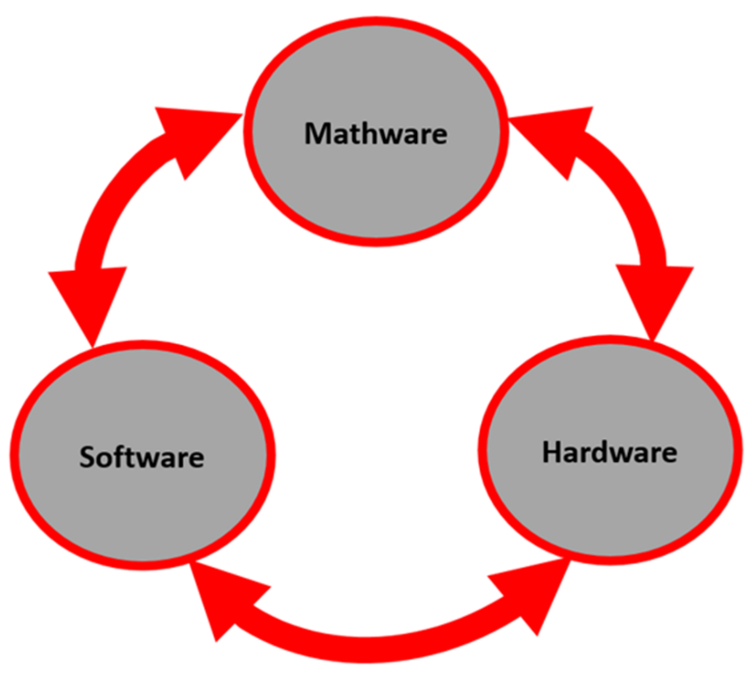
But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore's Law
   - Numerical methods needed to be made parallelizable
   - ICCG, for example, shows a very bad performance on current supercomputers
   - Hence, for the solution of sparse linear systems, entirely new methods need to be developed

REVOLUTIONARY NEW IDEAS NEEDED!

TU/e

Mathematical method development for HPC

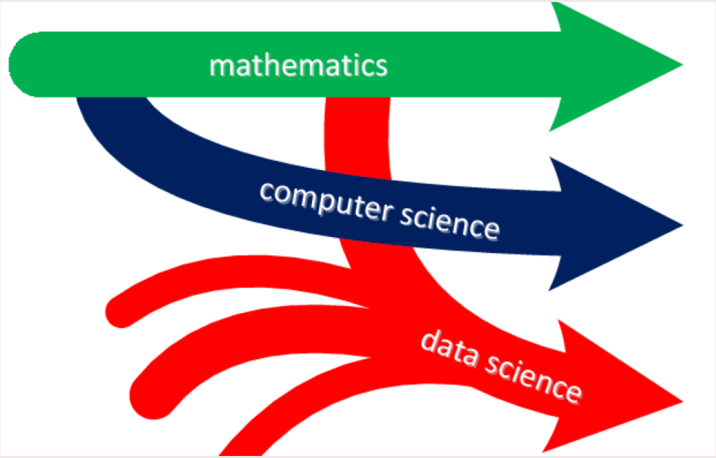


- Mathematical method development must be distinguished from software and hardware
- **Mathware** researchers must engage in discussions with software and hardware colleagues to achieve optimal results
- <u>Example:</u> ease transformations between 16, 32 and 64 bit representations (using FPGA?)

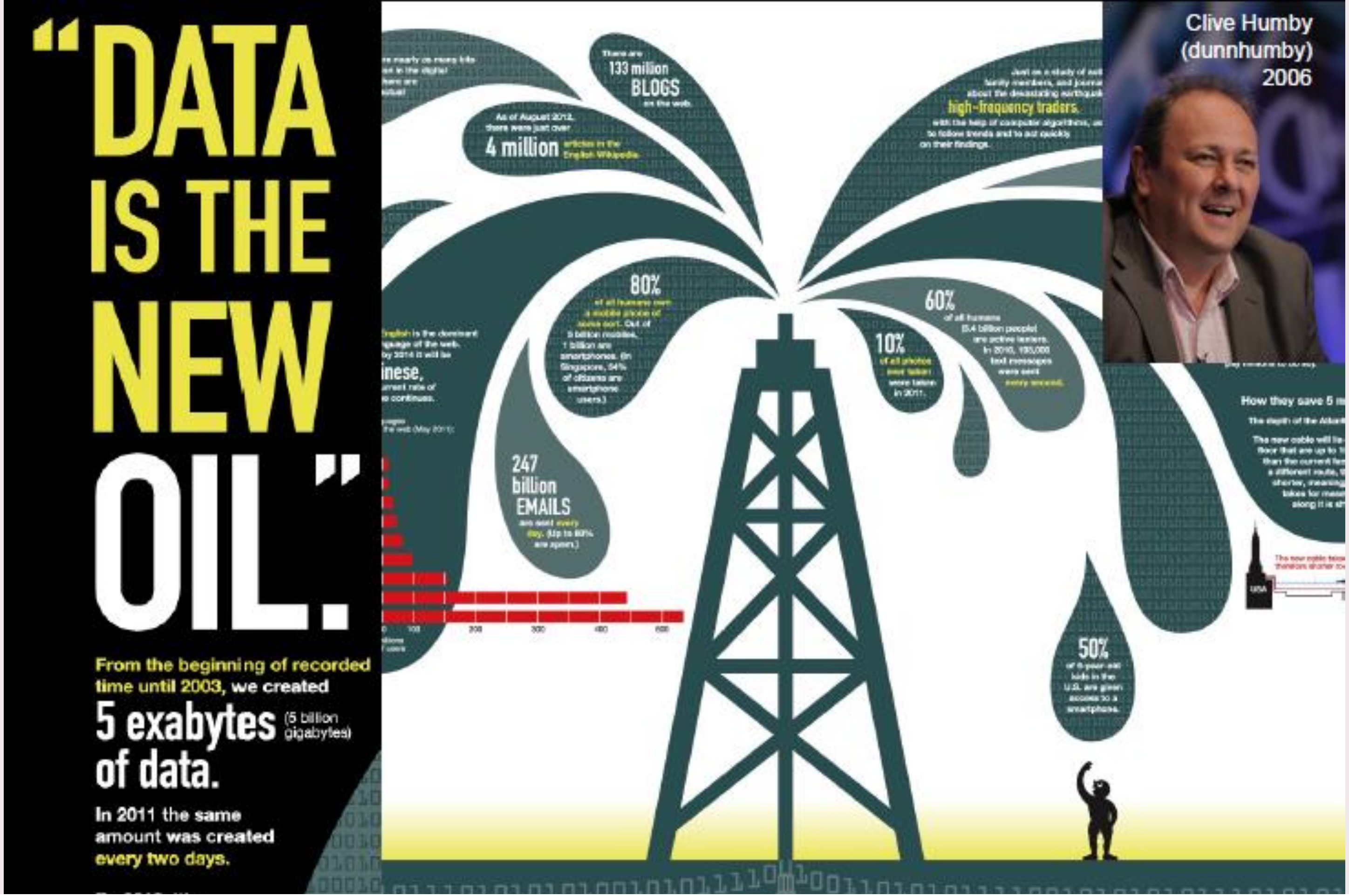
TU/e

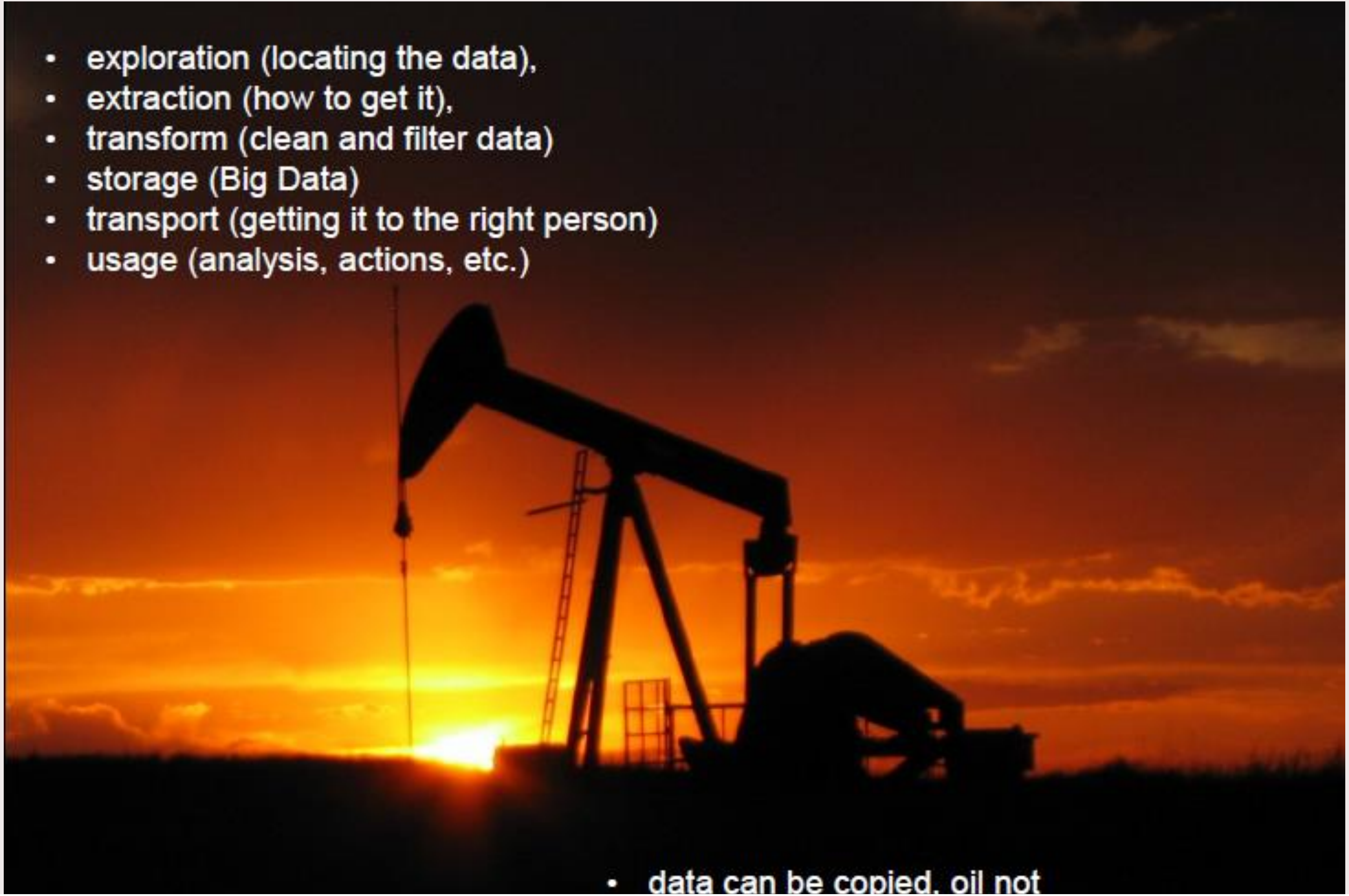
But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore’s Law
2. Data Science emerged as a discipline, and quickly became part of the curriculum at universities



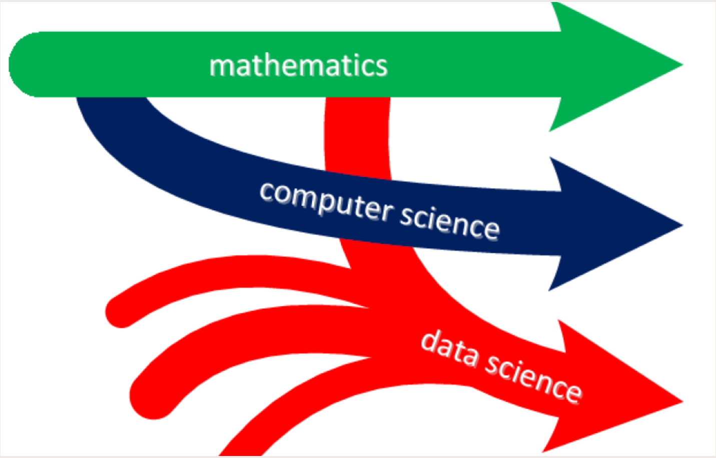
TU/e


"DATA IS THE NEW OIL."
From the beginning of recorded time until 2003, we created
5 exabytes (5 billion gigabytes) of data.
In 2011 the same amount was created every two days.
Clive Humby (dunnhumby) 2006
There are 133 million BLOGS on the web.
As of August 2012, there were just over 4 million articles in the English Wikipedia.
high-frequency traders.
80% of all humans own a mobile phone of some sort. Out of 5 billion mobiles, 1 billion are smartphones. (In Singapore, 54% of citizens are smartphone users.)
247 billion EMAILS are sent every day. (Up to 80% are spam.)
10% of all photos ever taken were taken in 2011.
60% of all humans (5.4 billion people) are active texters. In 2010, 193,000 text messages were sent every second.
50% of 8-year-old kids in the U.S. are given access to a smartphone.


• exploration (locating the data),
• extraction (how to get it),
• transform (clean and filter data)
• storage (Big Data)
• transport (getting it to the right person)
• usage (analysis, actions, etc.)
• data can be copied, oil not

But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore's Law
2. Data Science emerged as a discipline, and quickly became part of the curriculum at universities
   - It is an emerging discipline on the crossroads of multiple existing disciplines
   - David Donohue (Stanford): ''50 years of Data Science''



REVOLUTIONARY NEW IDEAS NEEDED!

But then…….

1. High Performance Computing started (again) to become important, and in fact inevitable due to the ending of Moore’s Law
2. Data Science emerged as a discipline, and quickly became part of the curriculum at universities
3. Artificial Intelligence became extremely popular, with techniques for deep learning, in combination with big data

MANY NEW CHALLENGES AHEAD!

TU/e

Quoting Karen Willcox (Oden, Texas)

"It is such an exciting time to be a computational scientist. The field is in the midst of a tremendous convergence of technologies that generate unprecedented system data and enable automation, algorithms that let users process massive amounts of data and run predictive simulations that drive key decisions, and the computing power that makes these algorithms feasible at scale for complex systems and in real-time or in situ settings."

We will concentrate on the third topic:

Combining methods from the fields of Computational Science and Engineering (CSE) and Artificial Intelligence (AI)

ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND NEURAL NETWORKS

Artificial Intelligence (AI)

- The origins of AI can be traced back to the desire to build thinking machines, or electronic brains.
- In 1958, Frank Rosenblatt created the first artificial neuron that could learn by iteratively strengthening the weights of the most relevant inputs and decreasing others to achieve a desired output.

Brain-inspired AI

- Computation in brains and the creation of intelligent systems have been studied in a symbiotic fashion for many decades.
- Europe has become a hotspot of brain-inspired computing research, the progress being accelerated by the FET flagship ''Human Brain Project''.
- In technology roadmaps, brain-inspired computing is commonly seen as a future key enabler for AI on the edge.
- Researchers at INRIA have presented an interdisciplinary approach towards transferring neuroscientific findings to new models of AI. Quoting them: *"Major algorithms from artificial intelligence (AI) lack higher cognitive functions such as problem solving and reasoning.''*

Machine Learning (ML)

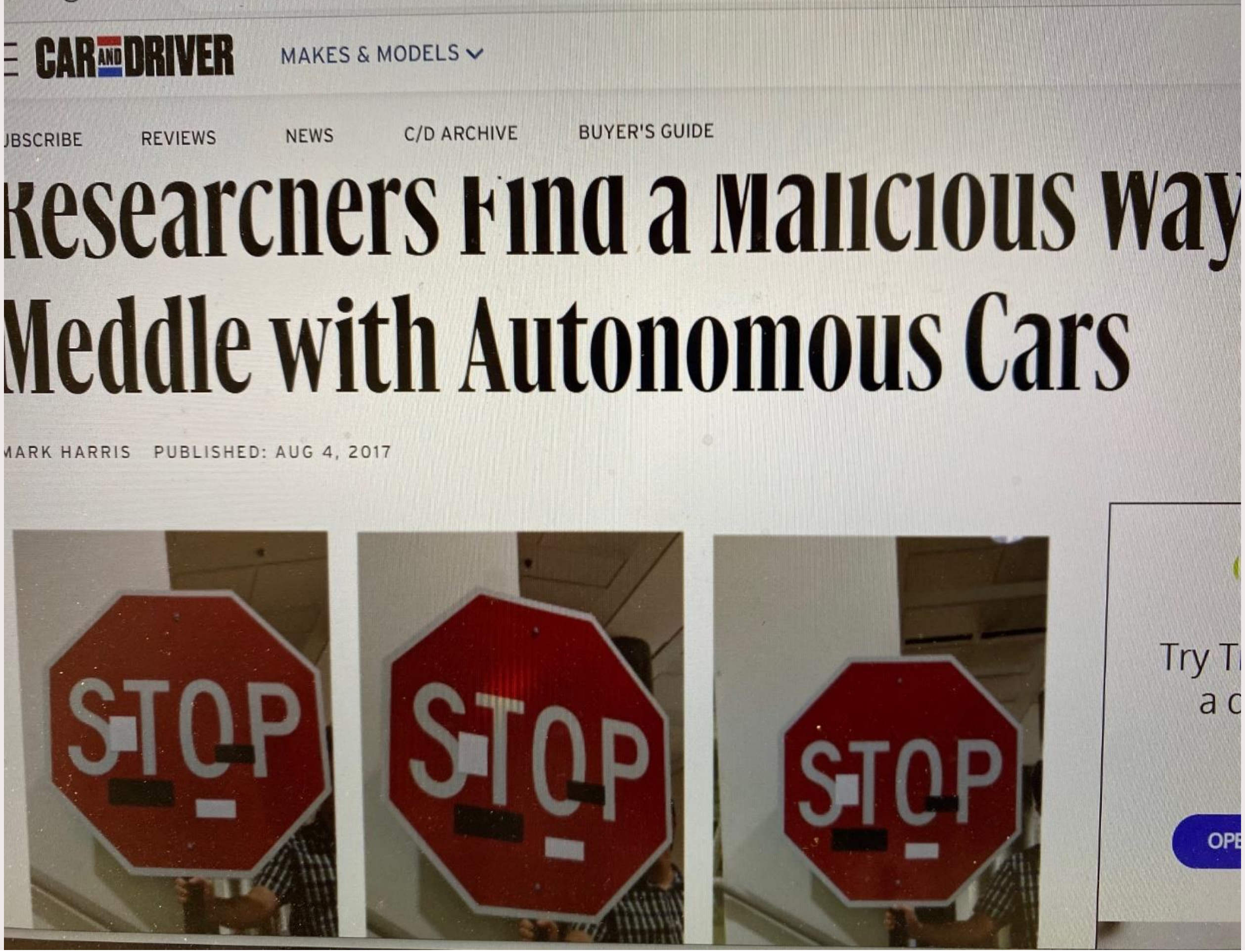
- The discipline of machine learning is often conflated with the general field of AI, but machine learning specifically is concerned with the question of how to develop algorithms and program computers to automatically recognise complex patterns and make intelligent decisions based on data.
- It involves probability theory, logic, combinatorial optimization, statistics, reinforcement learning and control theory.
- Applications are ubiquitous, ranging from vision to language processing, forecasting, pattern recognition, games, data mining, expert systems and robotics.

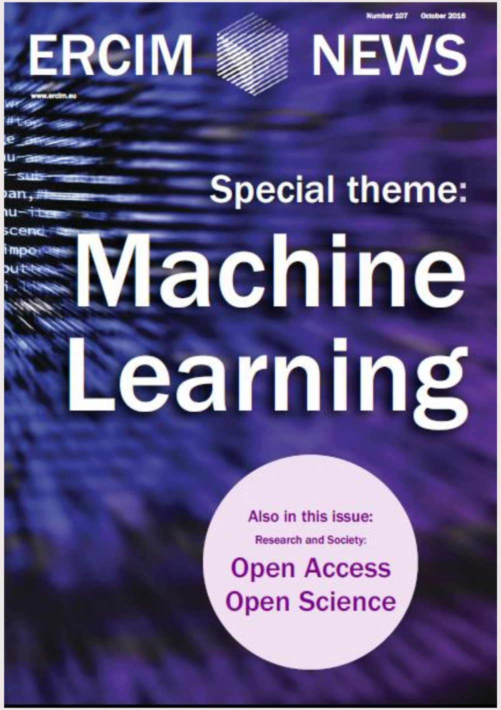
History of Machine Learning

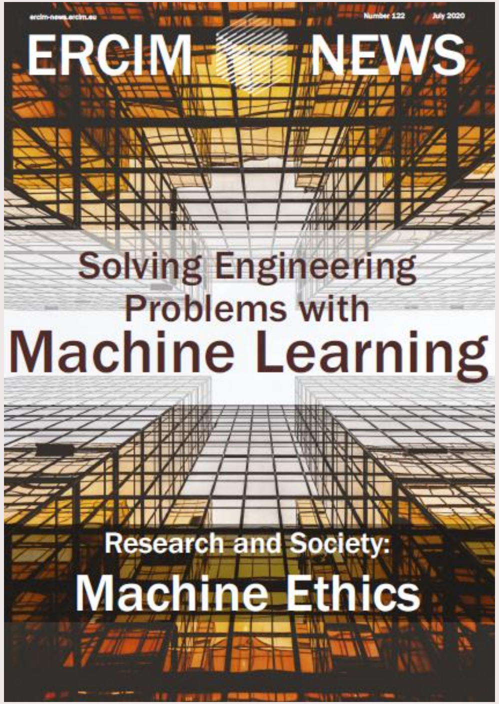
- Arthur Samuels popularized the term ‘’machine learning’’ in 1959; he built a checkers-playing program alongside efforts to understand the computational principles underlying human learning, in the developing field of **neural networks**.
- In the ‘90s, statistical AI emerged, formulating machine learning problems in terms of probability measures.
- Since then, the emphasis has vacillated between statistical and probabilistic learning and progressively more competitive neural network approaches.

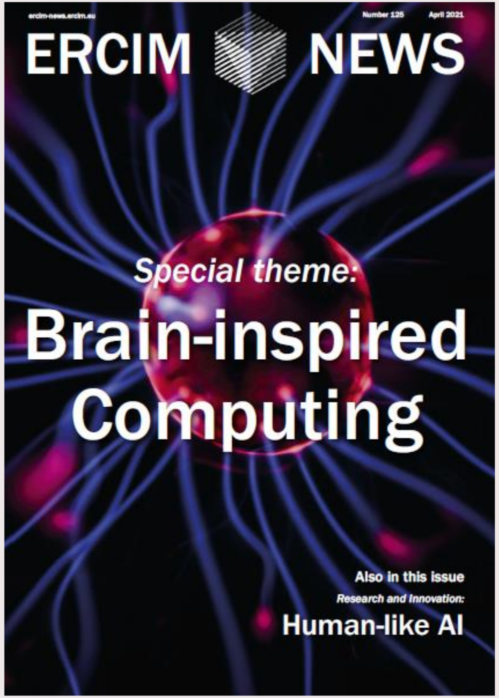
Breakthrough in Machine Learning

- The breakthrough work by Krizhevsky, Sutskever & Hinton in 2012 has been a catalyst for AI research. They used a **deep neural network** trained exhaustively on GPUs.
- Similar advances were then quickly reported for speech recognition and later for machine translation and natural language processing.
- Companies like Google, Microsoft and Baidu established large machine learning groups.
- Since then, with the combination of big data and big computers, rapid advances have been reported, including the use of machine learning for self-driving cars, and consumer-grade real-time speech-to-speech translation.


CAR AND DRIVER
MAKES & MODELS
REVIEWS
NEWS
C/D ARCHIVE
BUYER'S GUIDE
Researchers Find a Malicious Way
Meddle with Autonomous Cars
PUBLISHED: AUG 4, 2017
STOP
STOP
STOP


Number 107 October 2016
ERCIM NEWS
www.ercim.eu
Special theme:
Machine Learning
Also in this issue:
Research and Society:
Open Access
Open Science


ercim-news.ercim.eu
Number 122 July 2020
ERCIM NEWS
Solving Engineering Problems with
Machine Learning
Research and Society:
Machine Ethics


ercim-news.ercim.eu
Number 125 April 2021
ERCIM NEWS
Special theme:
Brain-inspired Computing
Also in this issue
Research and Innovation:
Human-like AI

MACHINE LEARNING

TRANSFORMING OUR WORLD

Tackling Climate Change with Machine Learning

David Rolnick[1]*, Priya L. Donti[2], Lynn H. Kaack[3], Kelly Kochanski[4], Alexandre Lacoste[5], Kris Sankaran[6,7], Andrew Slavin Ross[9], Nikola Milojevic-Dupont[10,11], Natasha Jaques[12], Anna Waldman-Brown[12], Alexandra Luccioni[6,7], Tegan Maharaj[6,8], Evan D. Sherwin[2], S. Karthik Mukkavilli[6,7], Konrad P. Körding[1], Carla Gomes[13], Andrew Y. Ng[14], Demis Hassabis[15], John C. Platt[16], Felix Creutzig[10,11], Jennifer Chayes[17], Yoshua Bengio[6,7]

[1]University of Pennsylvania, [2]Carnegie Mellon University, [3]ETH Zürich, [4]University of Colorado Boulder, [5]Element AI, [6]Mila, [7]Université de Montréal, [8]École Polytechnique de Montréal, [9]Harvard University, [10]Mercator Research Institute on Global Commons and Climate Change, [11]Technische Universität Berlin, [12]Massachusetts Institute of Technology, [13]Cornell University, [14]Stanford University, [15]DeepMind, [16]Google AI, [17]Microsoft Research

Abstract

Climate change is one of the greatest challenges facing humanity, and we, as machine learning experts, may wonder how we can help. Here we describe how machine learning can be a powerful tool in reducing greenhouse gas emissions and helping society adapt to a changing climate. From smart grids to disaster management, we identify high impact problems where existing gaps can be filled by machine learning, in collaboration with other fields. Our recommendations encompass exciting research questions as well as promising business opportunities. We call on the machine learning community to join the global effort against climate change.

Google AI Blog

The latest news from Google AI

Using Machine Learning to "Nowcast" Precipitation in High Resolution

Monday, January 13, 2020

Posted by Jason Hickey, Senior Software Engineer, Google Research

The weather can affect a person's daily routine in both mundane and serious ways, and the precision of forecasting can strongly influence how they deal with it. Weather predictions can inform people about whether they should take a different route to work, if they should reschedule the picnic planned for the weekend, or even if they need to evacuate their homes due to an approaching storm. But making accurate weather predictions can be particularly challenging for localized storms or events that evolve on hourly timescales, such as thunderstorms.

In "Machine Learning for Precipitation Nowcasting from Radar Images," we are presenting new research into the development of machine learning models for precipitation forecasting that addresses this challenge by making highly localized "physics-free" predictions that apply to the immediate future. A significant advantage of machine learning is that inference is computationally cheap given an already-trained model, allowing forecasts that are nearly instantaneous and in the native high resolution of the input data. This *precipitation nowcasting,* which focuses on 0-6 hour forecasts, can generate forecasts that have a 1km resolution with a total latency of just 5-10

Criticism is growing…

- The much-glorified deep learning approaches all rely on the availability of massive amounts of data, often needing millions of correctly labelled examples.
- Many domains, however, including some important areas such as health care, will never have such massive labelled datasets.
- Similarly, robots cannot be trained for millions of trials, simply because they wear out long before.
- **The question is thus how to learn more with less. Here, statistics and prior knowledge will likely play a big role.**

Criticism is growing…

There are serious limitations to current methods, as well as to our understanding of the success of machine learning techniques such as deep neural networks.

Professor Robbert Dijkgraaf* compares machine learning with 16th century alchemy, based on an accumulation of tricks topped with a good shot of credulity rather than on a systematic analysis.

He also quotes Ali Rahimi, a well-known researcher at Google, who last year accused the subject artificial intelligence of magical thinking.

*: Former president of Dutch Royal Academy of Sciences, former director of Princeton Institute of Advanced Studies, since a few months our new minister for Science and Education

TU/e

Criticism is growing…

The New York Times [12] goes even further, claiming that today's AI needs to do something completely different:

- "We need to stop building computer systems that merely get better and better at detecting statistical patterns in data sets – often using an approach as deep learning – and start building computer systems that from the moment of their assembly innately grasp three basic concepts: **time, space and causality**. Today's AI systems know surprisingly little about any of these concepts….. Few people working in AI are even trying to build such background assumptions into their machines."

Criticism is growing…

KEYWORDS: CHRISTOPHER MIMS

Why Artificial Intelligence Isn't Intelligent

Some experts in AI think its name fuels confusion and hype of the sort that led to past 'AI winters' of disappointment

By

Christopher Mims

July 31, 2021 12:00 am ET

A funny thing happens among engineers and researchers who build artificial intelligence once they attain a deep level of expertise in their field. Some of them—especially those who understand what actual, biological intelligences are capable of—conclude that there's nothing "intelligent" about AI at all.

Wall Street Journal, August 4, 2021

Deep Neural Nets are shortsighted

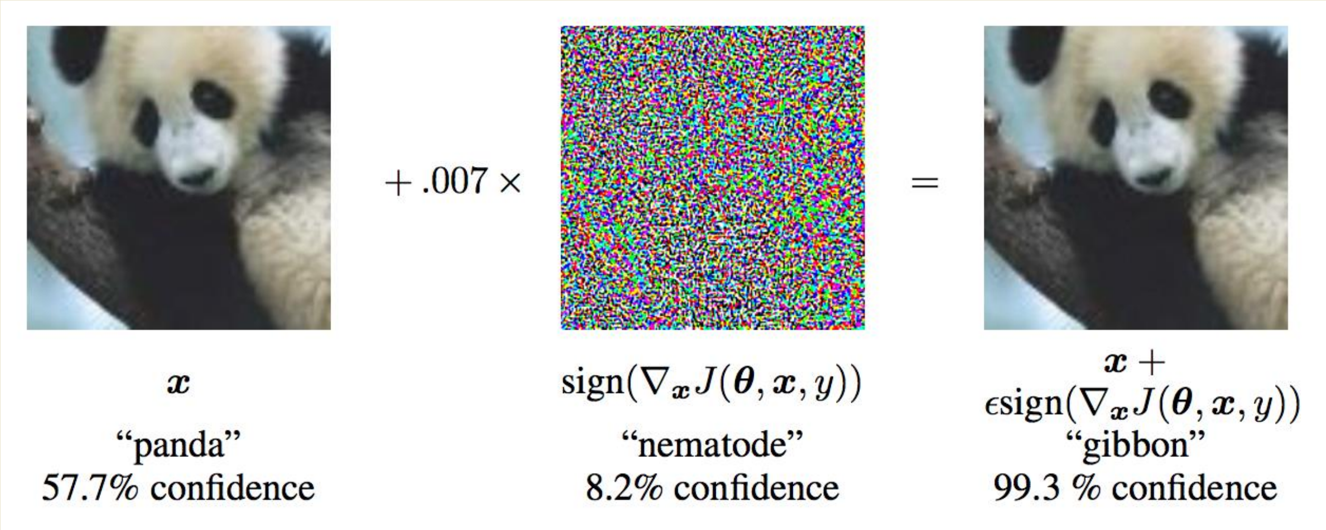
(a) Texture image
81.4% **Indian elephant**
10.3% indri
8.2% black swan

(b) Content image
71.1% **tabby cat**
17.3% grey fox
3.3% Siamese cat

(c) Texture-shape cue conflict
63.9% **Indian elephant**
26.4% indri
9.6% black swan

ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, Geirhos et al. 2019

Deep Neural Nets are shortsighted

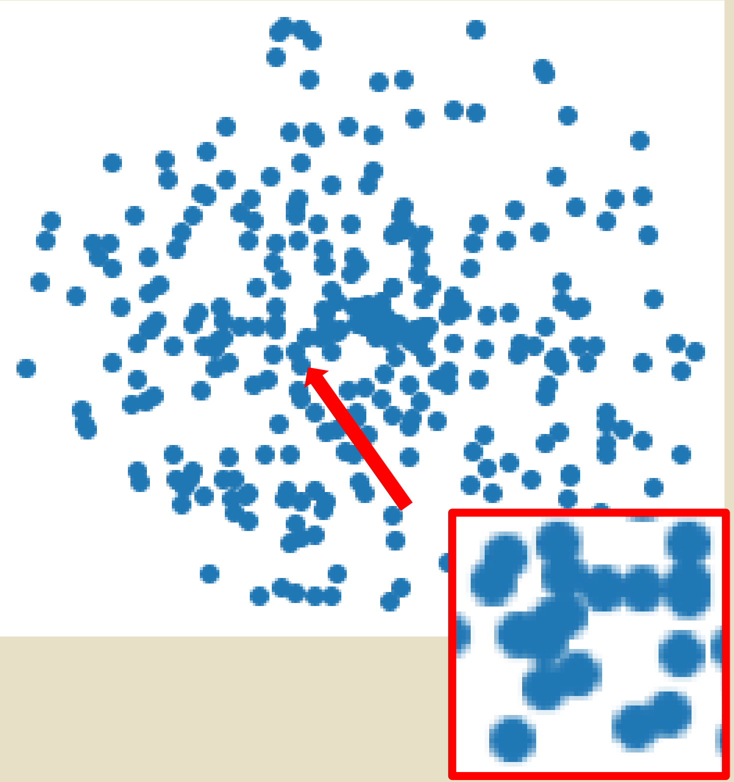


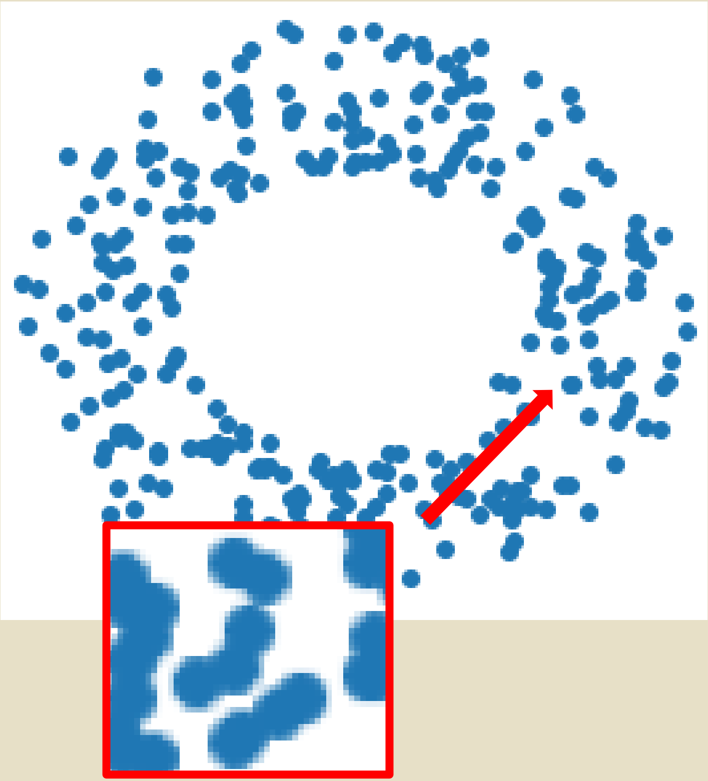
Explaining and Harnessing Adversarial Examples, Goodfellow et al. 2014

Deep Nets are too sensitive to local information.

Why? Because convolution is a local operation.

=> Use Topology to capture global characteristic





They look similar locally,

but apparently different if we zoom out

c.f. Manifolds are locally all Euclidean space and homology distinguishes the global topology of them.

Human is good at

- Rough estimation
- Panoramic view
- Discovering rules/**invariance** from a small number of examples
- **Explaining** the reason

DL is good at

- Precise observation
- Memorising/imitating examples
- Processing huge data
- Accurate operation

Deep Learning (DL)

Data-driven
local

Topological Data Analysis (TDA)

Maths-based
global

complementary

Background

- DL achieves high performance but has some weakness
- TDA has succeeded in capturing data features that conventional techniques have missed

Conclusion on AI and machine learning

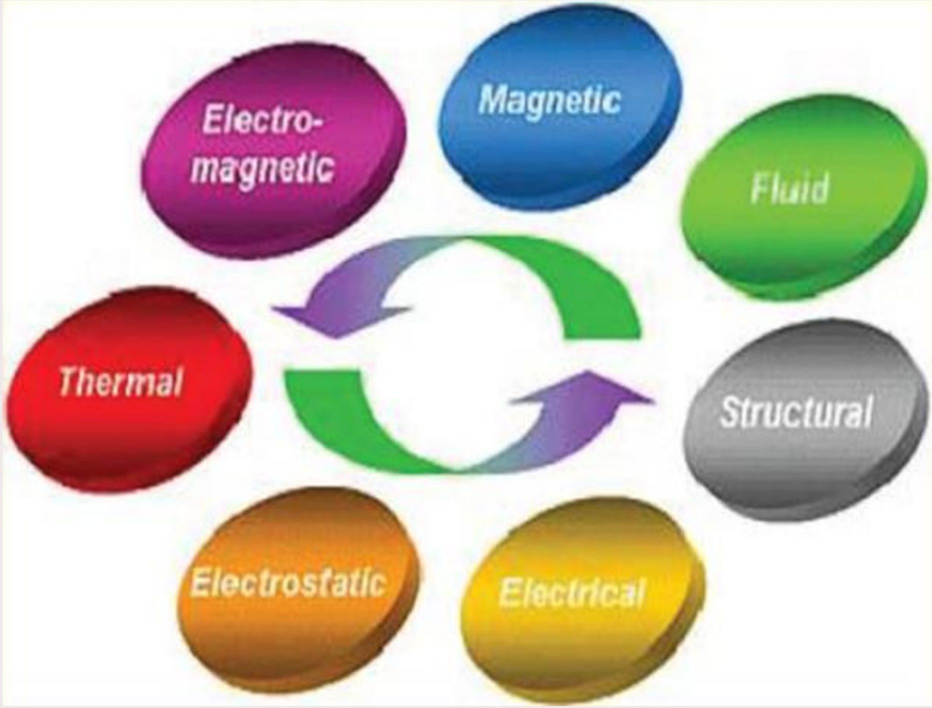
There is a lot of work ahead for mathematicians in the areas of artificial intelligence, machine learning and artificial neural networks (ANN)

- Understanding why methods work or do not work
- Understand the actions of the neurons (new ones?)
- Understanding on what grounds AI systems take decisions
    - In image recognition, use is made of the pixels; mathematics can provide much better methods
- How to select a good set of training data
- Using less data and prior knowledge
- Reducing the size and density of neural networks
- Predicting the topology of ANN
- ………

HYBRID METHODS: COMBINING CSE AND AI

Using AI within CSE

- In recent years, researchers in the field of Computational Science and Engineering realized that they could benefit from AI methods.
- Much more accurate models and simulations, needed for example in the creation of **Digital Twins**, require much more detailed models and coupled simulations.
- Neural networks can be used for accurate models of parameters



Going back in time: semiconductor device simulation

$$\vec{\nabla} \cdot \left(\varepsilon_{rel} \vec{\nabla} \phi \right) = -\frac{e}{\varepsilon_0} (p - n) ,$$

$$\vec{J}_n = -D_n \vec{\nabla} n + \mu_n n \vec{\nabla} \phi,$$

$$\vec{J}_p = -D_p \vec{\nabla} p - \mu_p p \vec{\nabla} \phi,$$

$$\frac{\partial n}{\partial t} = G - R - \vec{\nabla} \cdot \vec{J}_n,$$

$$\frac{\partial p}{\partial t} = G - R - \vec{\nabla} \cdot \vec{J}_p.$$

- Every year new models are constructed for mobility (and recombination), based upon many simulations and measurements, then using physical insight and curve-fitting
  - *Engineers and phycisists provided their neural networks*
- Why not use **artificial neural networks**, based upon the abundantly available measurement and simulation data?

TU/e

Problem in this context

- Mathematicians derived conditions that mobility models must satisfy
- Peter Markowich proved that a monotonicity condition, with respect to the quasi-Fermilevel gradients, must hold
- Once the engineers at Philips presented a model that did not satisfy this condition; simulations failed at some point. They then corrected the model, satisfying the mathematical constraint
- **Obviously, models generated with neural networks should also satisfy the constraint**
- How can we achieve this???

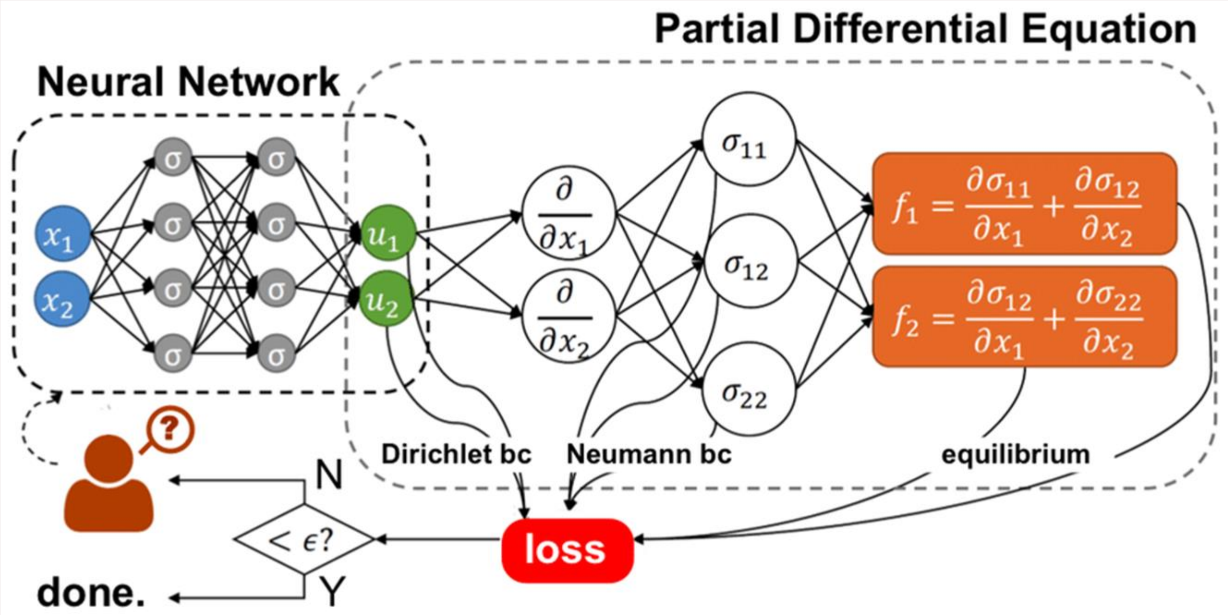
TU/e

“

“The future needs Computational Science and Engineering, blending data driven and physics-based perspectives”

Karen Willcox, director Oden Institute for Computational Engineering and Sciences

Physics Informed Neural Networks (PINNs)

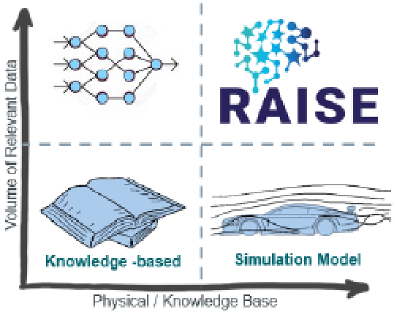


(George Karniadakis, Brown University, USA)

I am not sure that loss functions are the way to go, it leads to many problems

I prefer methods where physical properties are ***hard-coded*** *into the network*

Combining physics based and data-based science and engineering



Combining the best out of two worlds

	Physical Models	RAISE	Data Driven Models
Data Reliance	Small data	R	Big data
Domain Reliance	Deep domain expertise	A	Little domain knowledge
Fidelity & Robustness	Highly non-linear and complex relations	I	Limited to data and model complexity
Interpretability	Consistent physical meaningful	S	Physics agnostics surrogate
Adaptability & Deployability	Complex and time consuming setup	E	Rapidly adapt to a specific problem
Tools & Data	Models & software are USPs		Data is USP; software is OS

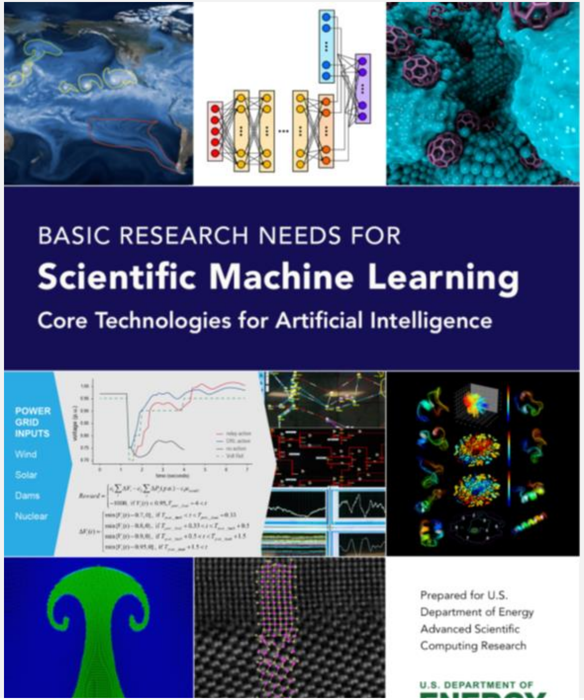
Figure 3: The two "orthog… oaches of Modelling and Simulation and Machine Learning methods based on Data-Driven Models complement each other. MSODE targets the interface of the two.

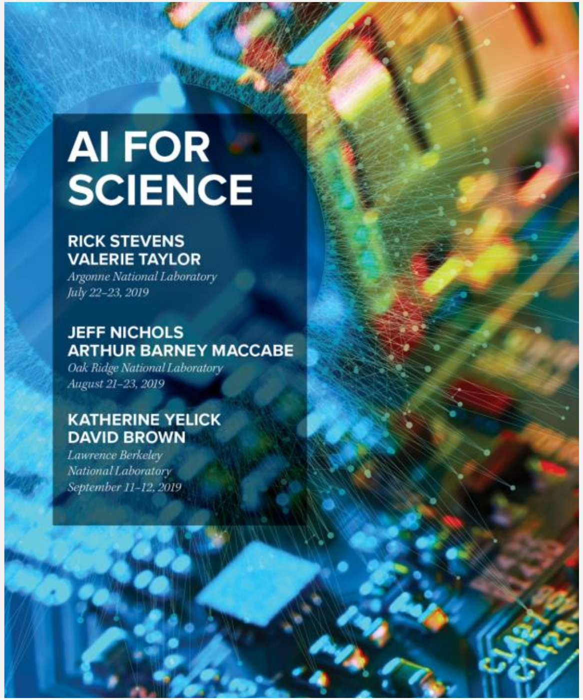
Courtesy Dirk Hartmann, Siemens

Richard Feynman:
"People who wish to analyse nature without using mathematics must settle for a reduced understanding."

TU/e

USA is front runner

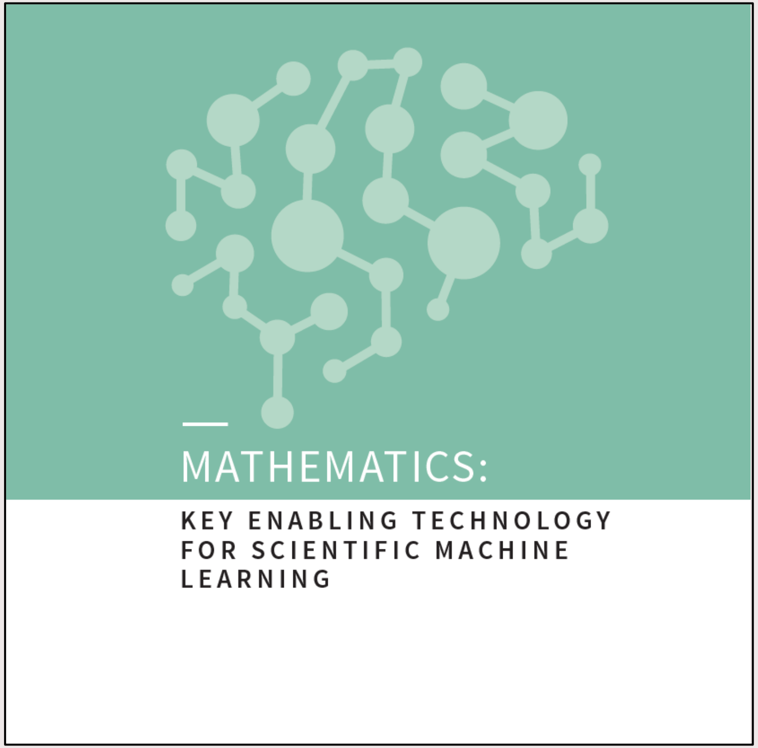




Workshop Lorentz Center (Leiden), November 1-5, 2021

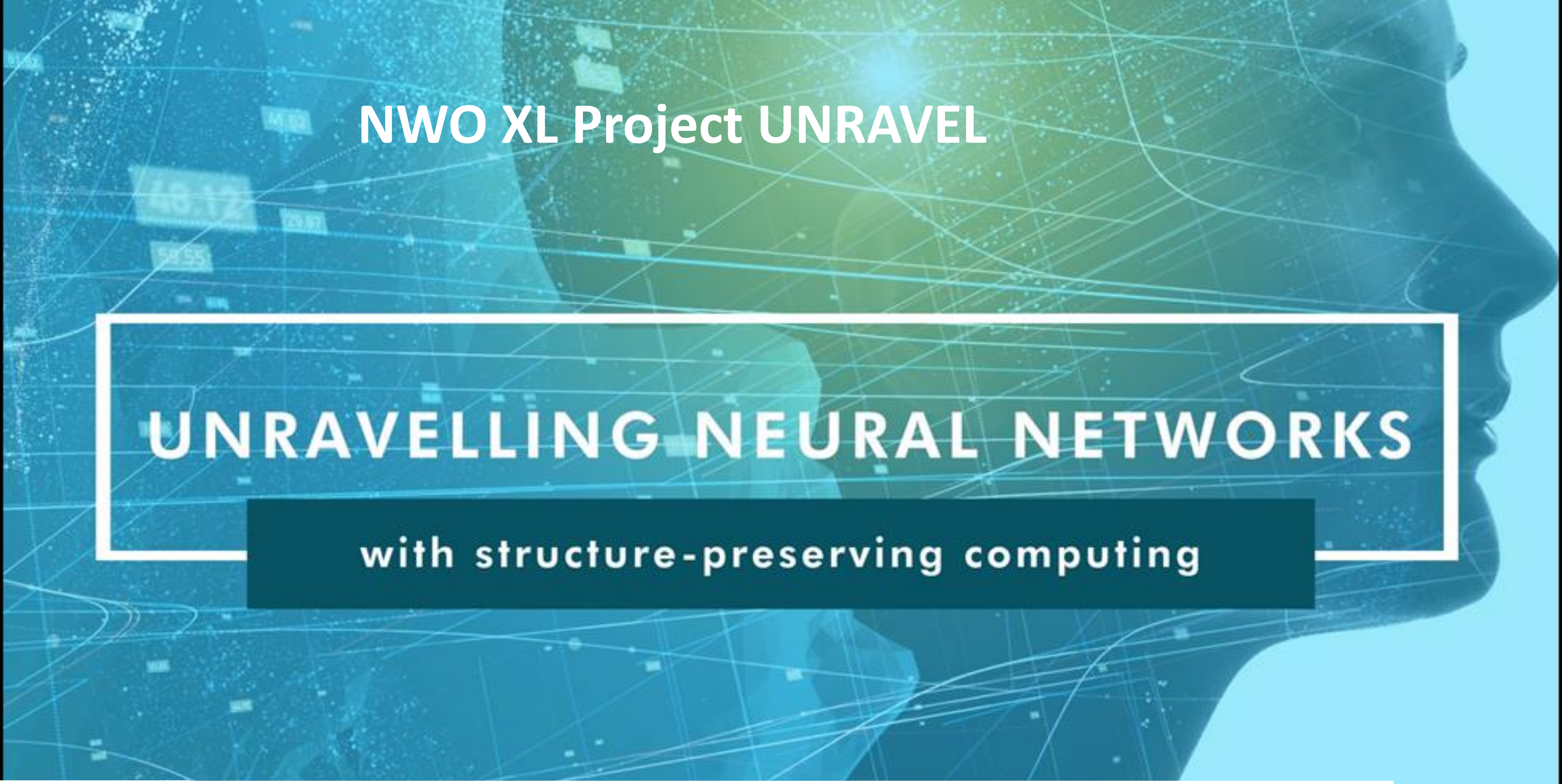
- **''Computational mathematics and machine learning''**
- Keynote speakers:
  - George Karniadakis
  - Weinan E
  - Petros Koumoutsakos
  - Carola Schönlieb
  - Stéphanie Allasonnière
  - Karen Willcox
  - Stephan Wojtowytsch
  - Paris Perdikaris
  - Erik Bekkers

Booklet presented during Lorentz workshop

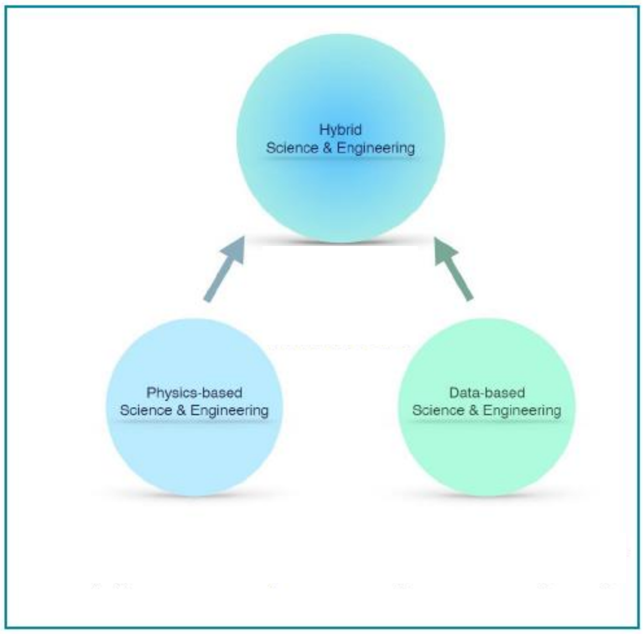


https://platformwiskunde.nl/wp-content/uploads/2021/11/Math_KET_SciML.pdf

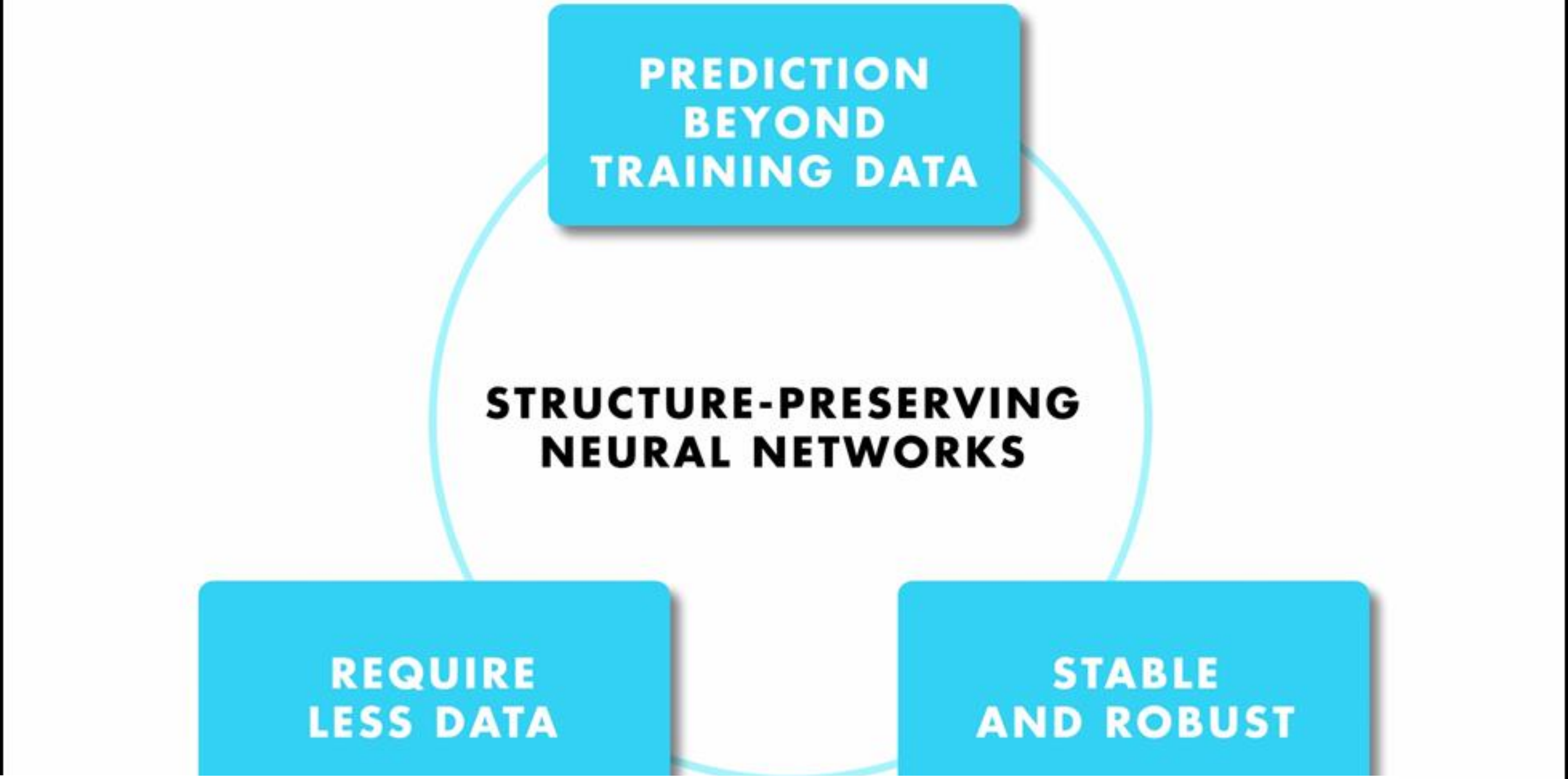
NWO XL Project UNRAVEL

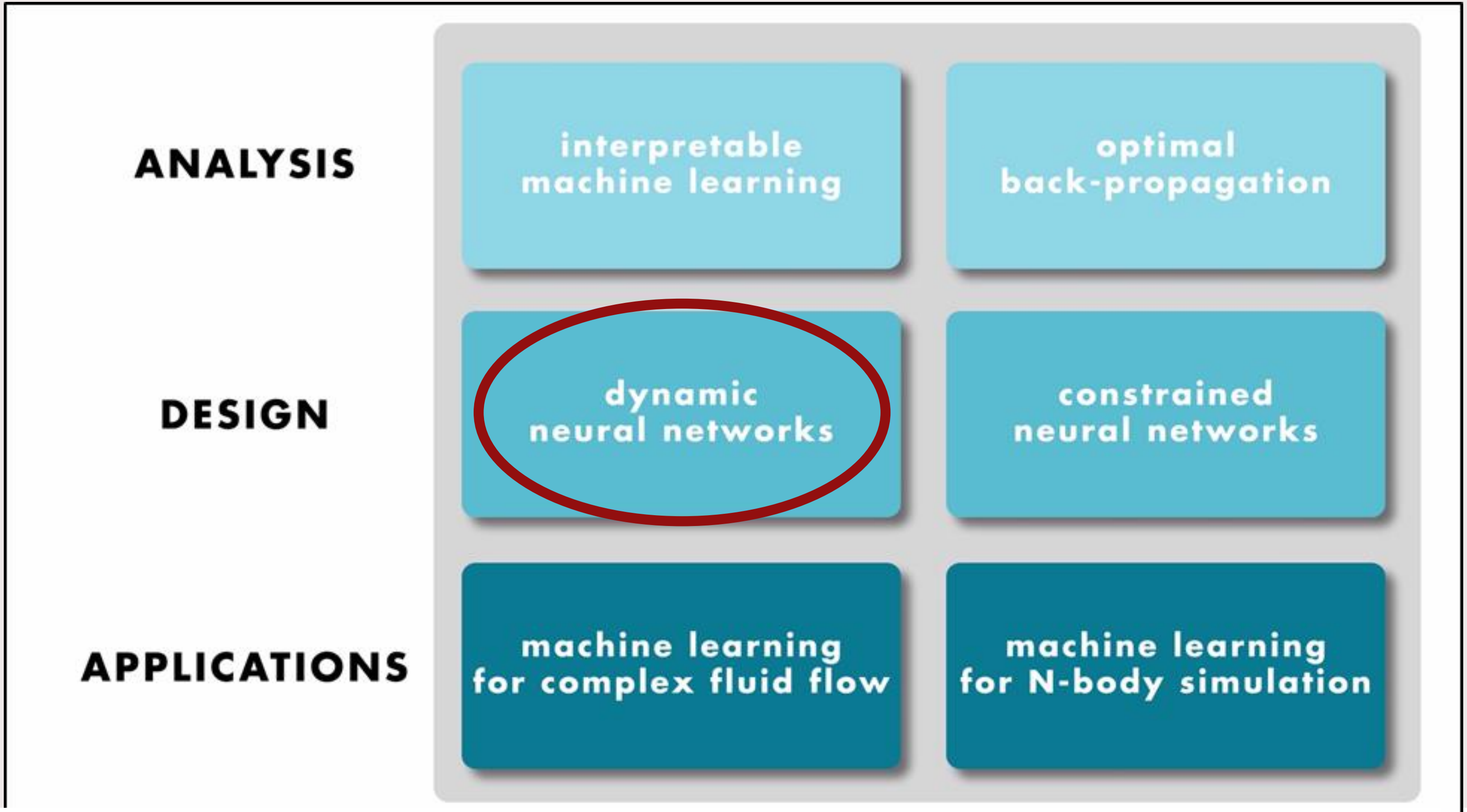


Combining physics based and data-based science and engineering



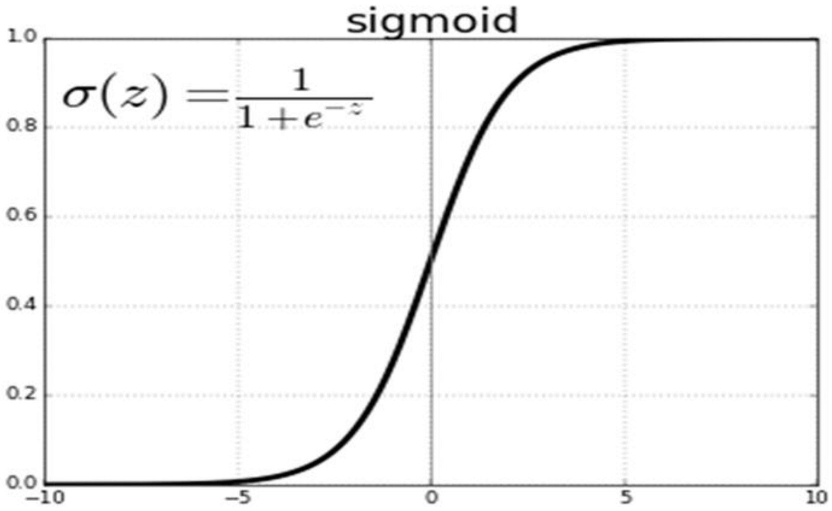
- We aim at using so-called **mimetic methods**, i.e. methods that preserve properties of the underlying system
- How to develop mimetic neural networks or mimetic machine learning methods is an open challenge
- Such methods may need (much) less data, i.e. also work in case of ''little data'' rather than ''big data''

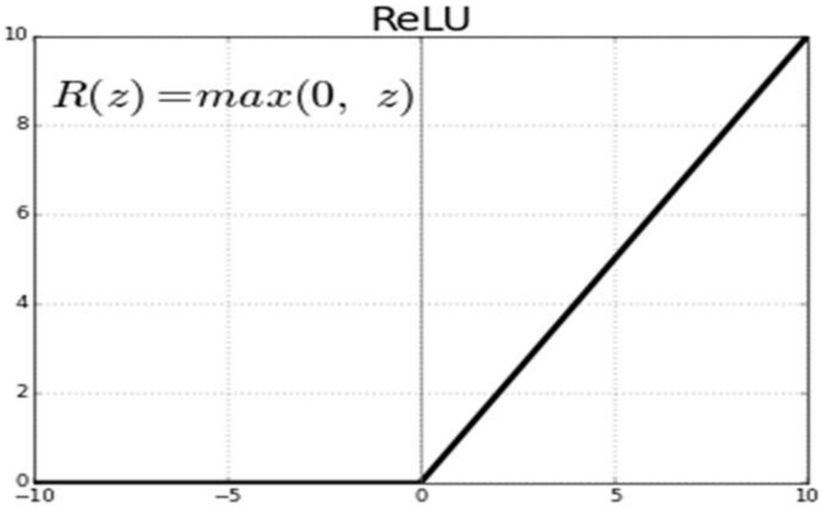

PREDICTION BEYOND TRAINING DATA
STRUCTURE-PRESERVING NEURAL NETWORKS
REQUIRE LESS DATA
STABLE AND ROBUST


ANALYSIS
interpretable machine learning
optimal back-propagation
DESIGN
dynamic neural networks
constrained neural networks
APPLICATIONS
machine learning for complex fluid flow
machine learning for N-body simulation

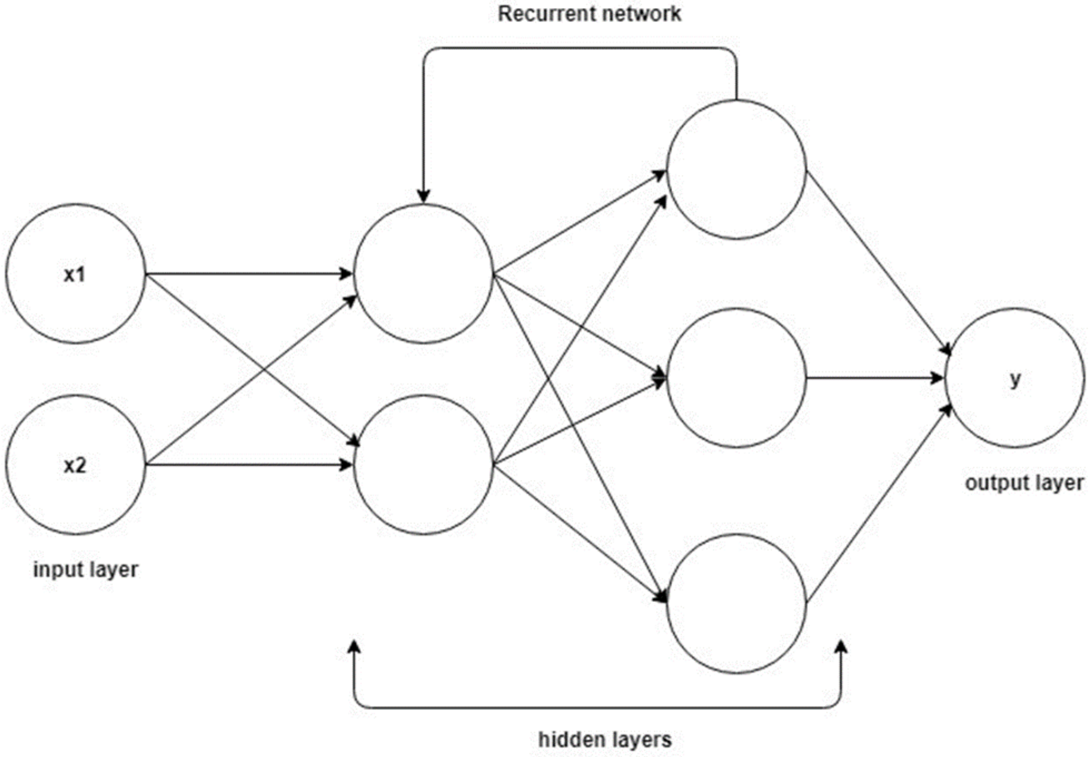
EXAMPLE 1: DYNAMIC NEURAL NETWORKS

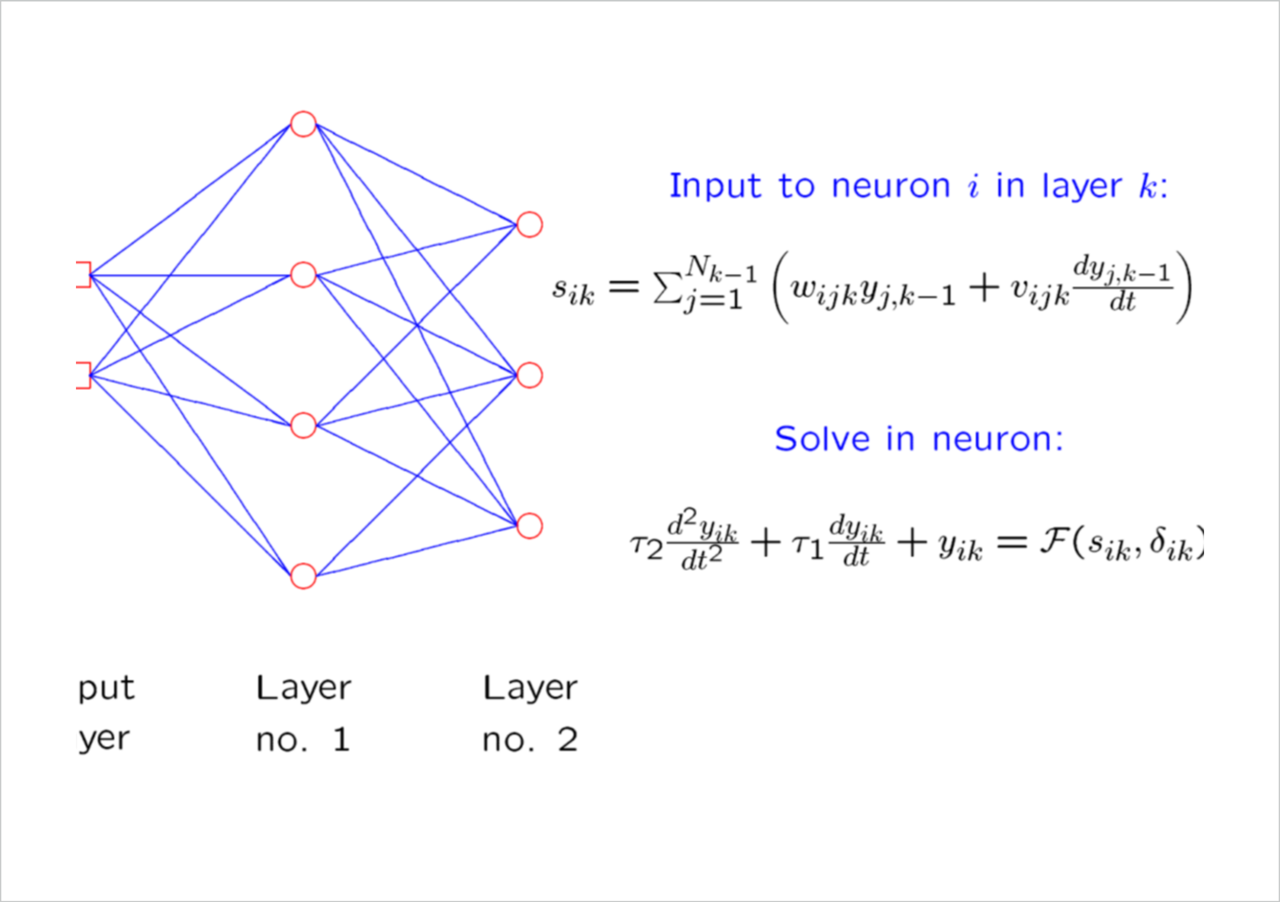
Neural networks are often static, and use the following neuron activation functions





For dynamic situations (ODE, PDE, DAE), often recurrent neural networks are suggested





At Philips Research, we developed truly dynamic neural networks

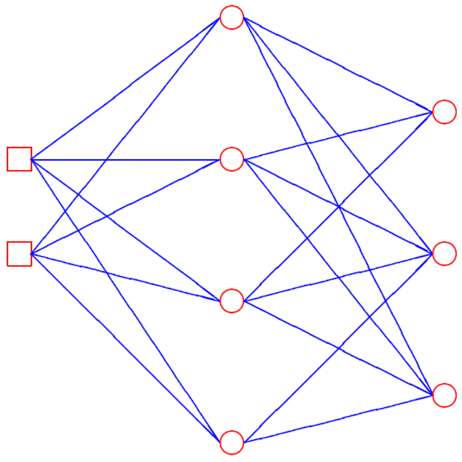
Dynamic neural networks

- We were able to show that there is a 1-1 relation to state space models of the form

$$\frac{d\mathbf{x}(t)}{dt} = \mathcal{A}\mathbf{x}(t) + \mathcal{B}\mathbf{u}(t),$$

$$\mathbf{y}(t) = \mathcal{C}\mathbf{x}(t) + \mathcal{D}\mathbf{u}(t).$$

- Using this relation, the topology of the network can be defined (using the MOESP algorithm):
    - Number of hidden layers related to multiplicity of eigenvalues of A
    - Number of neurons related to number of complex eigenvalues
    - Real eigenvalue → neuron with 1st order ODE
    - Complex eigenvalue(s) → neuron with 2nd order ODE
    - Methodology involves SVD, QR, Bartels-Stewart algorithm, solving Sylvester equations



Input to neuron i in layer k:

$$s_{ik} = \sum_{j=1}^{N_{k-1}} \left(w_{ijk} y_{j,k-1} + v_{ijk} \frac{dy_{j,k-1}}{dt} \right) - \theta_{ik}$$

Solve in neuron:

$$\tau_2 \frac{d^2 y_{ik}}{dt^2} + \tau_1 \frac{dy_{ik}}{dt} + y_{ik} = \mathcal{F}(s_{ik}, \delta_{ik})$$

The action of the first (hidden) layer in the network can be summarized as

$$T_2 x''(t) + T_1 x'(t) + x(t) = W u(t) + V u'(t) - \theta,$$

where T_1, T_2 are diagonal matrices.

The MOESP algorithm results in a system of the form

$$x'(t) = \mathcal{A}x(t) + \mathcal{B}u(t),$$

$$y(t) = \mathcal{C}x(t) + \mathcal{D}u(t).$$

Hence, we need to find $\mathcal{Z}$ such that $\mathcal{Z}^{-1}\mathcal{A}\mathcal{Z} = \mathcal{T}$ is block diagonal (1×1 and 2×2).

For the construction of $\mathcal{Z}$, consider the real Schur decomposition of $\mathcal{A}$:

$$\mathcal{Q}^T\mathcal{A}\mathcal{Q} = \mathcal{R},$$

where

$$\mathcal{R} = \begin{bmatrix} \mathcal{R}_{11} & \mathcal{R}_{12} & \cdots & \mathcal{R}_{1q} \\ & \mathcal{R}_{22} & \cdots & \mathcal{R}_{2q} \\ & & \ddots & \vdots \\ & & & \mathcal{R}_{qq} \end{bmatrix}.$$

The matrices $\mathcal{R}_{ij}$ are either 1×1 or 2×2 blocks, depending on whether or not the corresponding eigenvalue is complex.
The Bartels-Stewart algorithm can be used to find $\mathcal{Y}$ such that

$$\mathcal{Y}^{-1}\mathcal{R}\mathcal{Y} = \mathcal{T} = \mathrm{diag}\left(\mathcal{R}_{11}, \mathcal{R}_{22}, ..., \mathcal{R}_{qq}\right).$$

Hence, we find the desired result:

$$\mathcal{Y}^{-1}\mathcal{Q}^{-1}\mathcal{A}\mathcal{Q}\mathcal{Y} = \mathcal{T}.$$

Having found $\mathcal{Z}$ such that $\mathcal{Z}^{-1}\mathcal{A}\mathcal{Z} = \mathcal{T}$ is block diagonal, we can translate the MOESP linear system into a neural network.

$$x'(t) = \mathcal{A}x(t) + \mathcal{B}u(t)$$

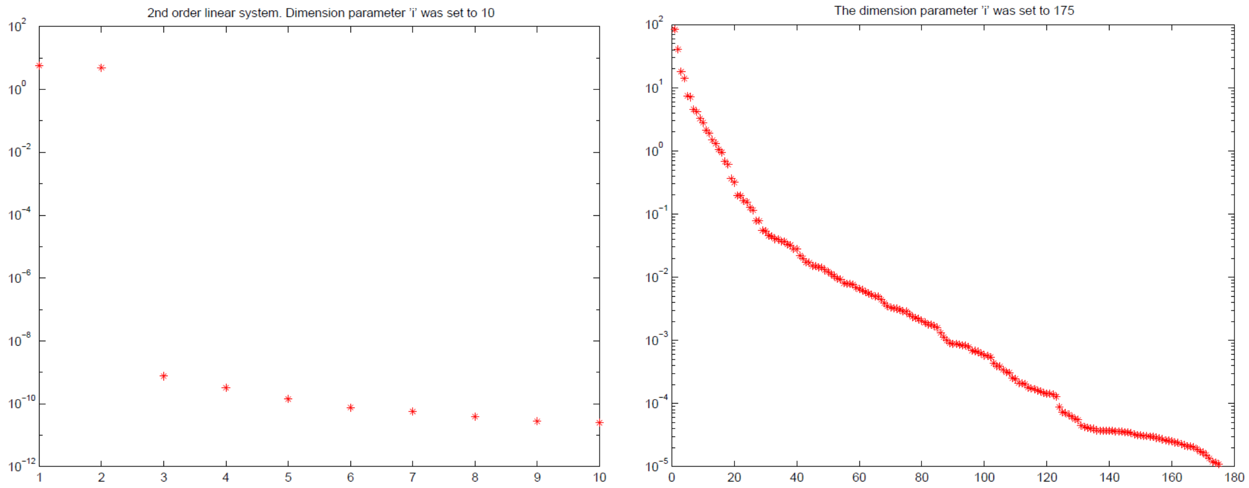
On multiplying by $\mathcal{Z}^{-1}$:

$$\mathcal{Z}^{-1}x'(t) = \mathcal{Z}^{-1}\mathcal{A}x(t) + \mathcal{Z}^{-1}\mathcal{B}u(t).$$

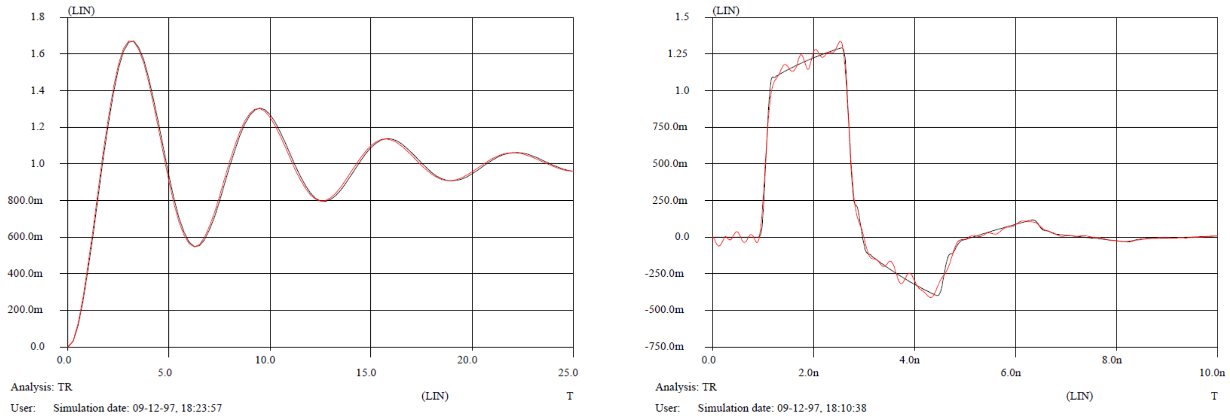
Transform to new variable $\hat{x} = \mathcal{Z}^{-1}x$:

$$\hat{x}'(t) = \mathcal{T}\hat{x}(t) + \mathcal{Z}^{-1}\mathcal{B}u(t).$$

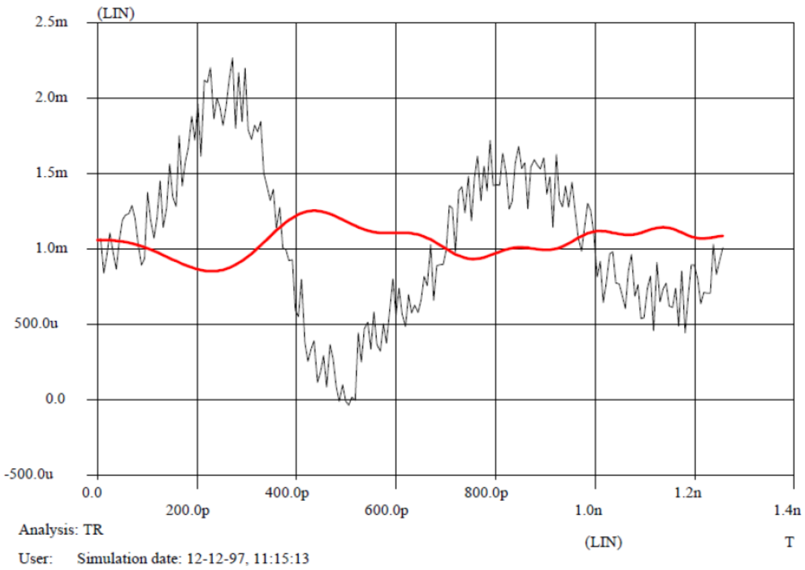
- 1×1 block: 1 neuron, first order ODE
- 2×2 block: 1 neuron, second order ODE

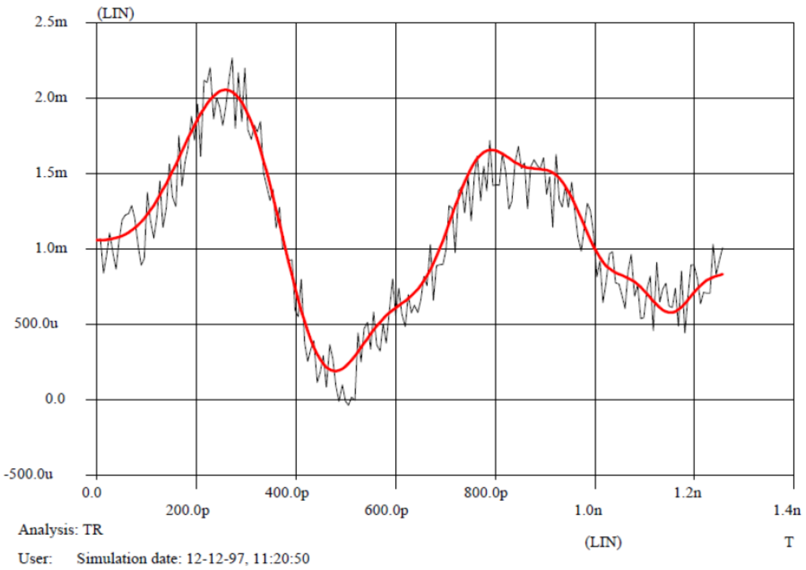


semi-logarithmic plot of singular values



Pstar analog test bench generated by NEUREKA





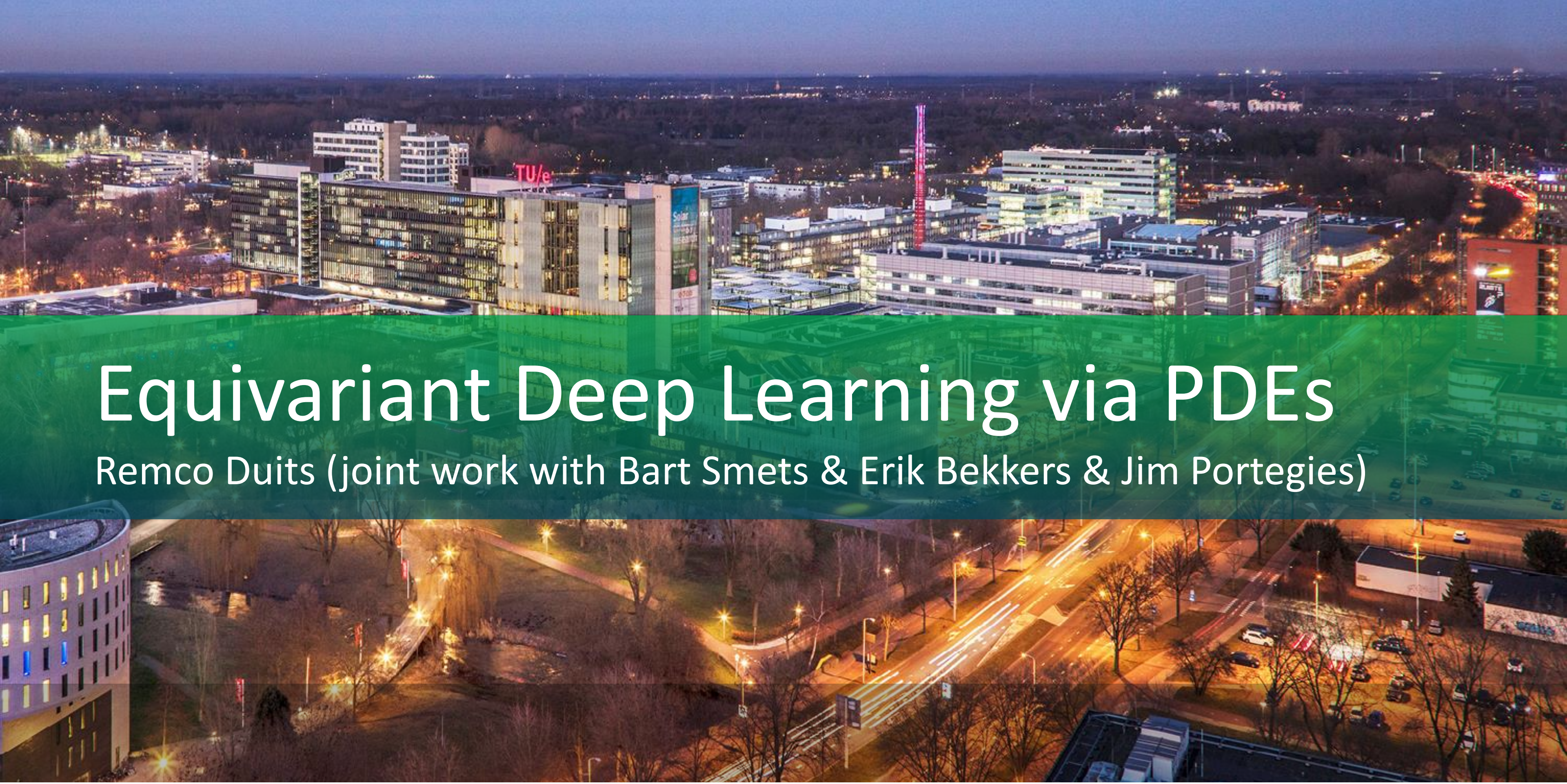
without MOESP preprocessing

with MOESP preprocessing

Potential of dynamic neural networks

- We were able to predict the topology of dynamic neural networks (# hidden layers, # neurons per layer) by establishing a 1-1 correspondence with state space models
- This correspondence also opens up the way to methods for model order reduction of neural networks, translating MOR concepts for state space models
- We are currently also investigating ''pruning of neural networks'', which is related to model order reduction
- Neuron action in these dynamic neural networks can be viewed as so-called high pass or low pass filters in electronics, implying that we are using electronic concepts for the construction of the networks mimicking true behaviour

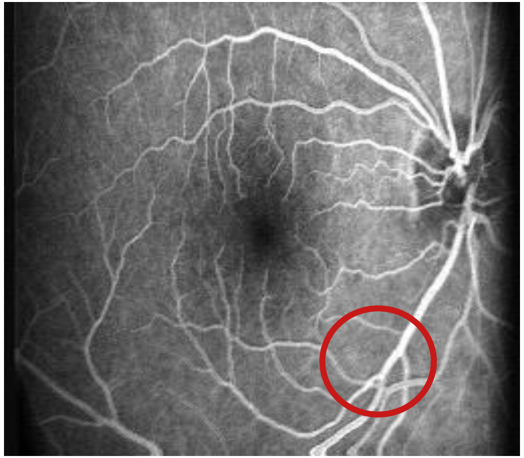
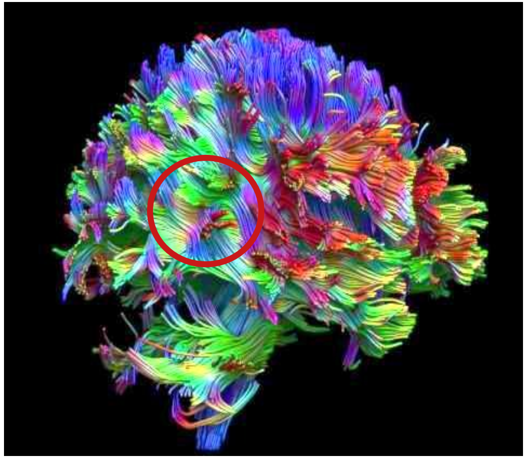
EXAMPLE 2: GEOMETRIC CONCEPTS AND AI



Equivariant Deep Learning via PDEs

Remco Duits (joint work with Bart Smets & Erik Bekkers & Jim Portegies)

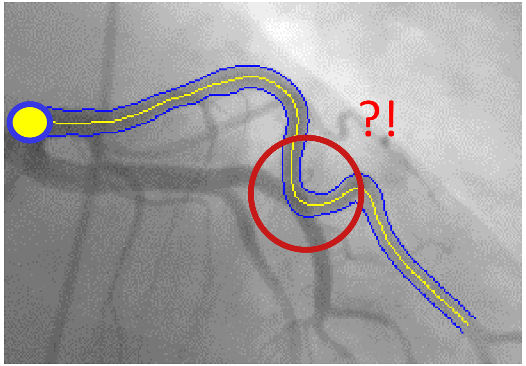
Current image analysis methods fall short

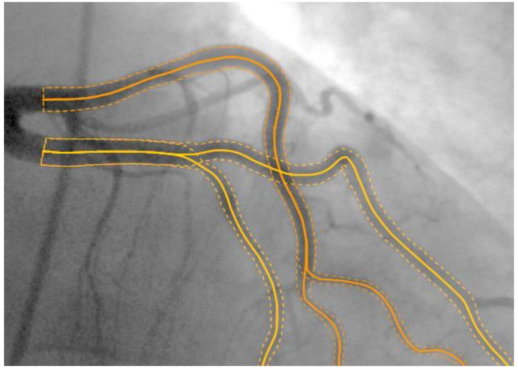

Costly user-input to correct

Original

Problem

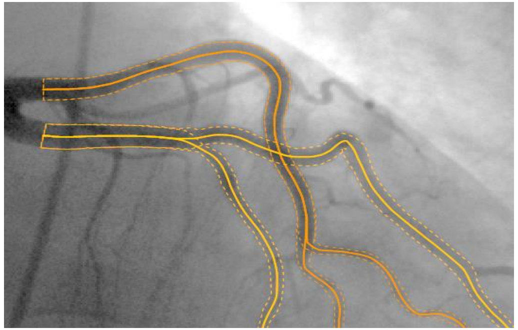


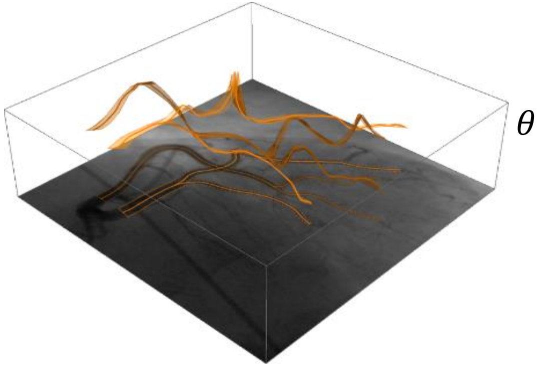
Solution

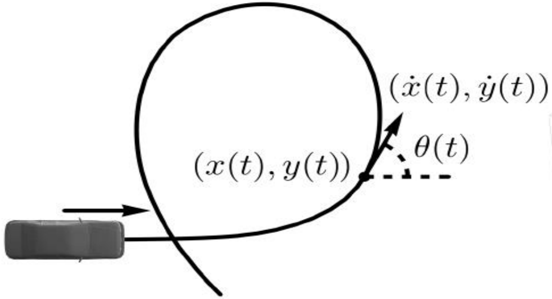


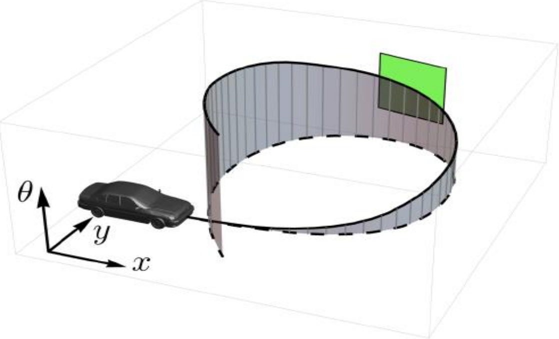
PDE-based geometric learning

New Dimensions









Merge geometry and machine learning

Geometric Image Analysis

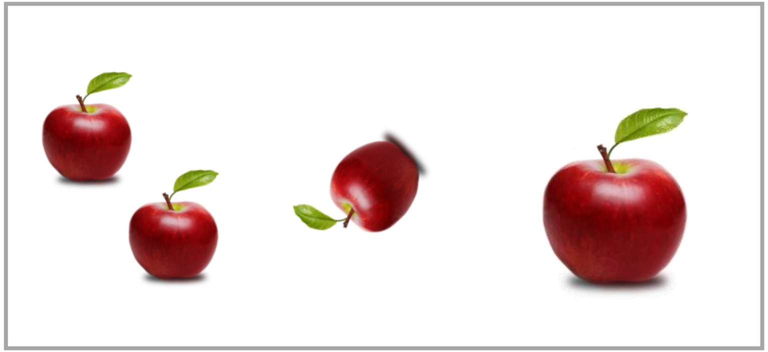
- − Limited performance
- − Limited scope
- − **Hand-crafting**
- + **Geometric Interpretation by PDEs**
- + Low computational load
- + Few parameters
- + Little training-data

Deep Learning

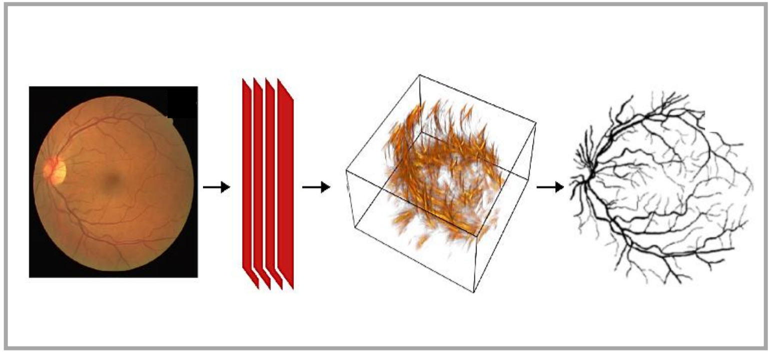
- + High performance
- + Wide scope
- + **Automatic**
- − **No geometric interpretation**
- − High computational load
- − Too many parameters
- − Huge training-data

Geometric PDE-Based neural networks

Reduce neural network by employing symmetry



Learn geometry by PDEs to improve classification



Equivariant Deep Learning via PDEs

- An exciting area of research, improving the performance of convolutional neural networks (CNN) with geometric concepts, leading to the so-called G-CNN networks
- Remco Duits has obtained a very prestigious NWO Vici grant (2.5 MEuro) to carry out this research
- For more information: https://www.win.tue.nl/~rduits/

CONCLUSION

Conclusion

- These are exciting times for researchers in the mathematical sciences, with the advent of high-performance computing, data science and artificial intelligence
- Combining ''traditional'' methods in Computational Science and Engineering with methods from Artificial Intelligence, Machine Learning and Neural Networks is the way forward to increase accuracy of models, as required by e.g. Digital Twinning
- **Using prior knowledge** will be key to improve the performance of neural networks
    - Increased accuracy, less data, more robustness

Conclusion

- Expertise from numerical linear algebra and model order reduction can be used to ‘’prune’’ neural networks: reducing them in size, and improving the sparsity
- Mathematics may aid in predicting the topology of neural networks, avoiding the currently employed guesswork
- The mathematical sciences are indispensable in the new multidisciplinary field of **scientific machine learning**, combining model- and data-based methods

Real intelligence is needed to make artificial intelligence work

(you may quote me on this)

TU/e