HASLERSTIFTUNG

Math of AI

Volkan Cevher, Associate Professor EPFL

lions@epfl

https://lions.epfl.ch

EPFL

Preface

My research:

- Machine Learning (ML)
- Optimization
- Signal Processing
- Information Theory
- Statistics

My courses (2019-20):

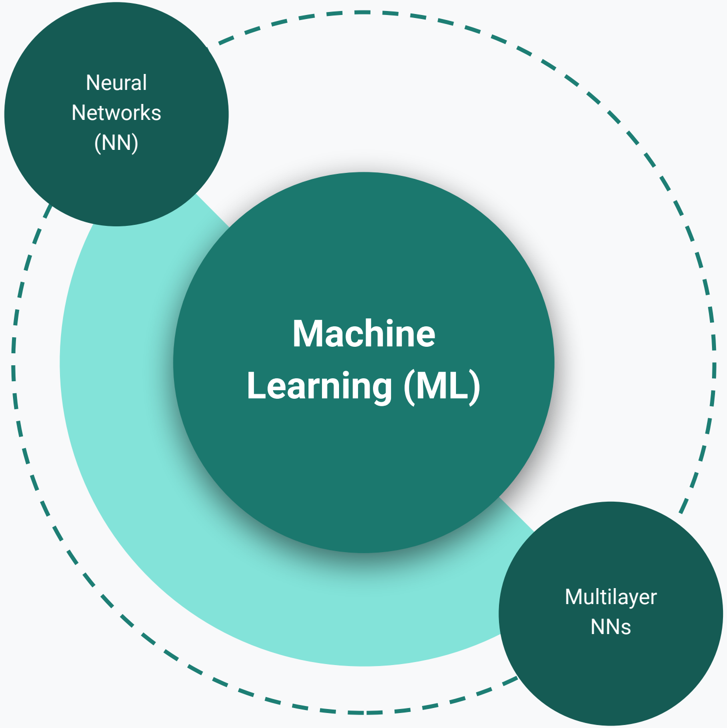
- Mathematics of Data
- Reinforcement Learning
- Advanced Topics in ML

+

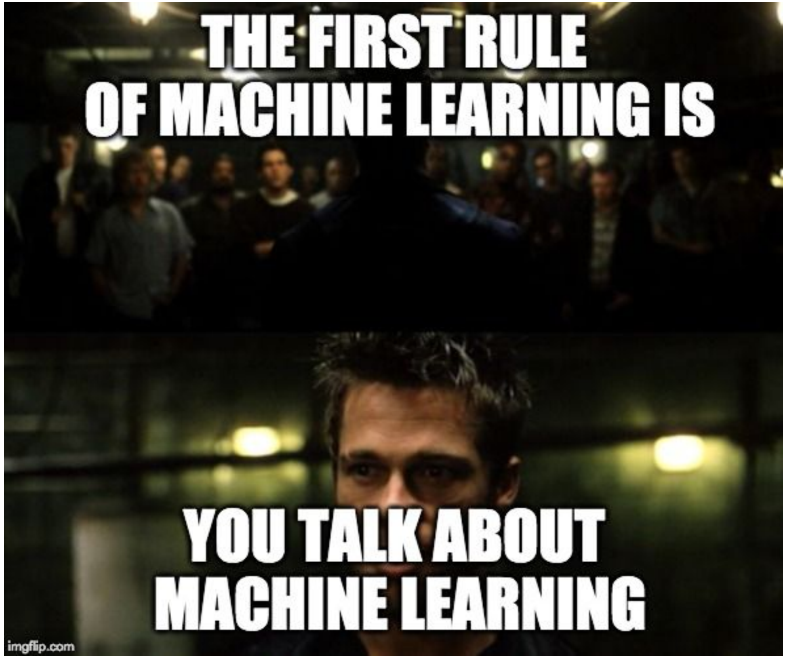
= **This talk**

Strengths

A SWOT Analysis



Machine Learning (ML)



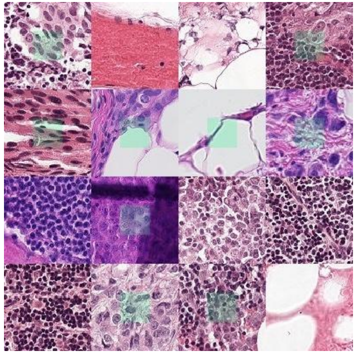
- ML is an interdisciplinary study of algorithms, statistical models, and error functions jointly with computer systems to perform specific tasks

  "Only a fool learns from his own mistakes. The wise man learns from the mistakes of others" - Otto von Bismarck

- ML makes you wiser

The ingredients via a simplified supervised learning example

Retinopathy



Lymph node cancer

- ML is an interdisciplinary study of algorithms, statistical models, and error functions jointly with computer systems to perform specific tasks

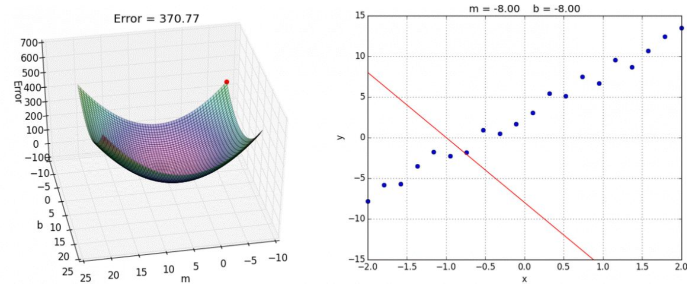
Task: Learn a mapping from image to disease

$$\mathbf{y} = \text{function}_{\mathbf{x}}(\mathbf{a}) = \underbrace{f(\mathbf{a}'\mathbf{x})}_{\text{model}}$$

The ingredients via a simplified supervised learning example

- ML is an interdisciplinary study of algorithms, statistical models, and error functions jointly with computer systems to perform specific tasks

Gradient Descent Algorithm



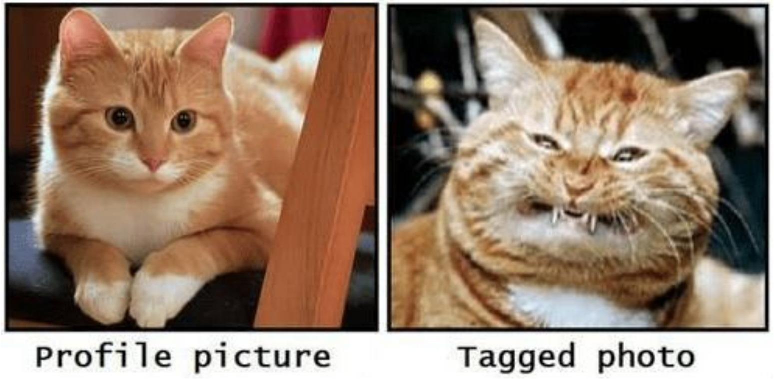
Supervised ML: Use algorithms to learn "model"

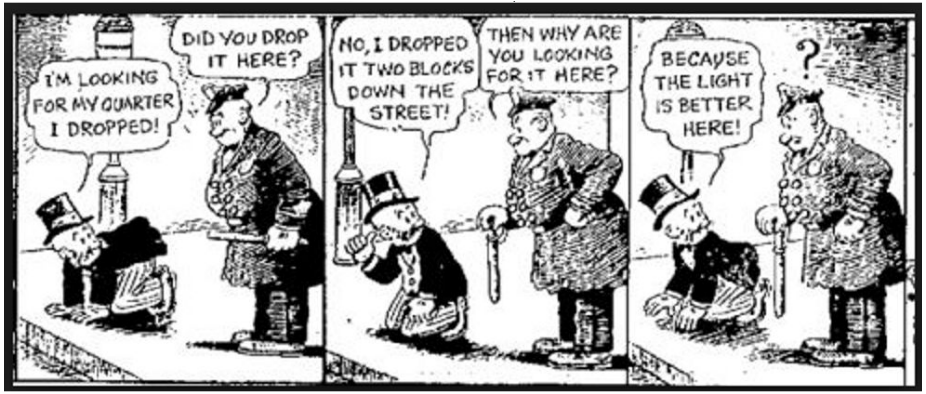
$$\min_{\mathbf{x}} \mathrm{Error}\left(\mathbf{y}, f(\mathbf{a}'\mathbf{x})\right)$$

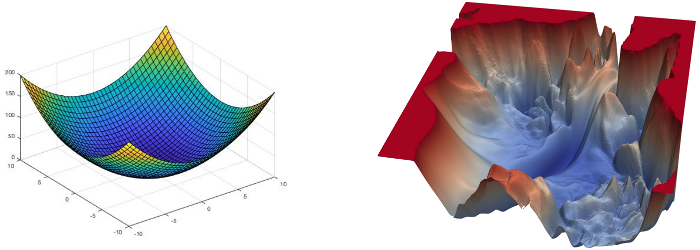
Academic theory vs industrial practice

Conventional wisdom in ML until 2010:

Simple models + simple errors

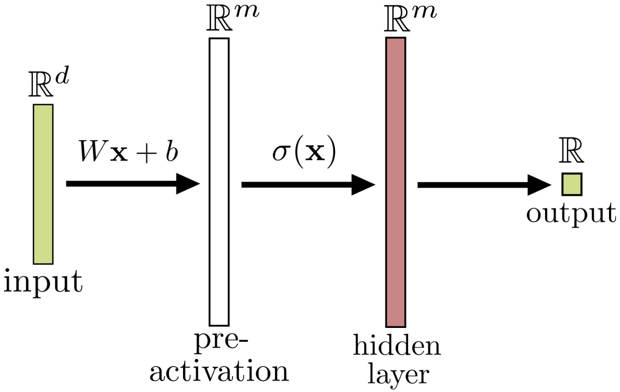




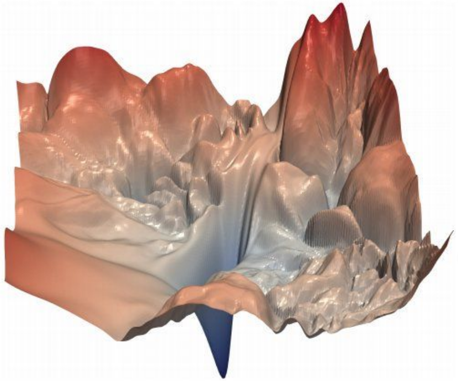


optimization landscapes

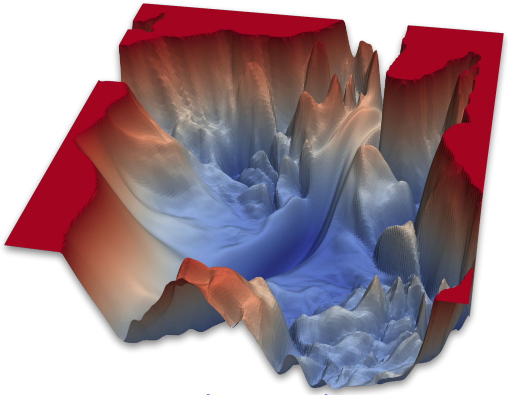
Enter neural networks: Universal approximation



$$f(\mathbf{x}; \beta, W, b) = \beta^T \sigma(W\mathbf{x} + b)$$



real function



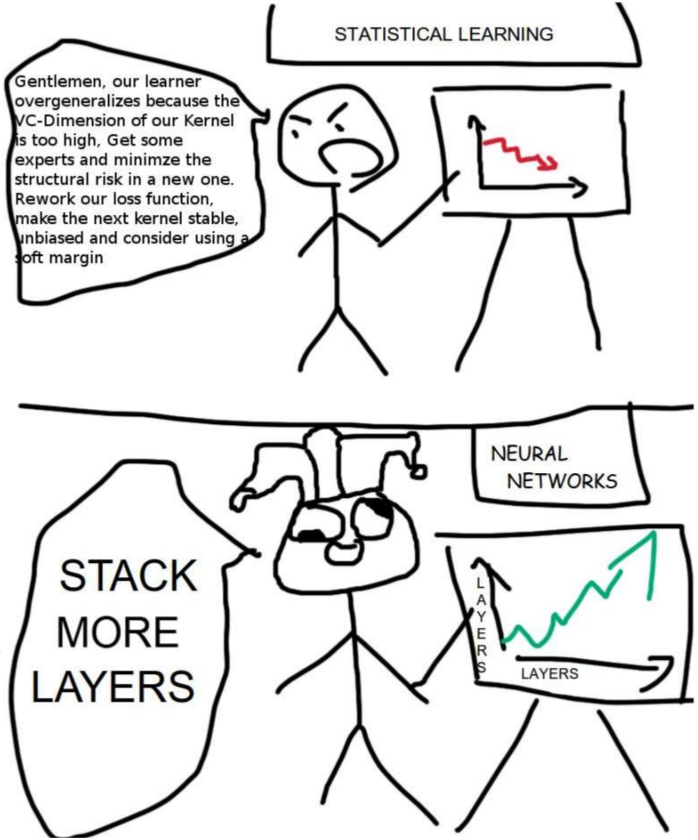
optimization landscape

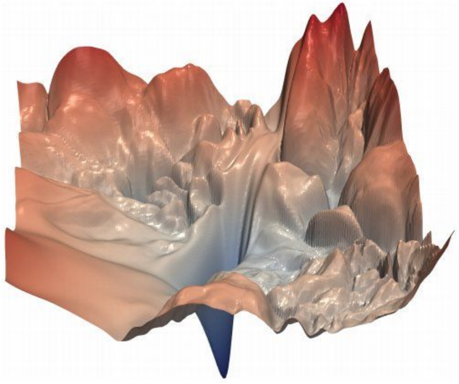
on a hypercube

Challenges:

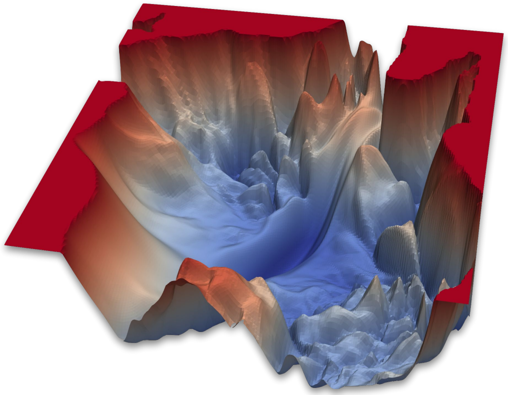
1. too big to optimize!
2. did not have enough data
3. could not find the optimum via algorithms

Multilayer neural networks: Tractable & nearly universal

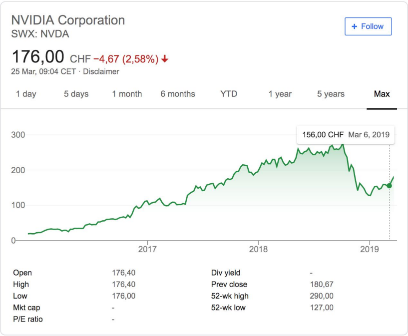


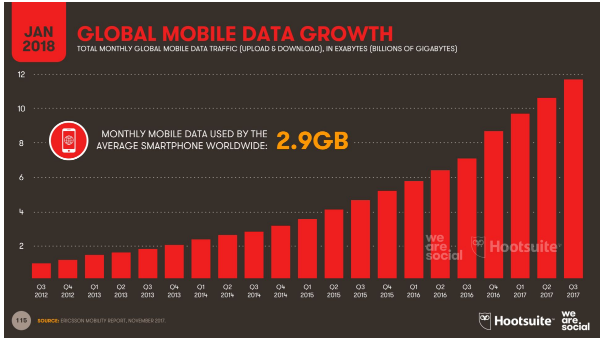


real function



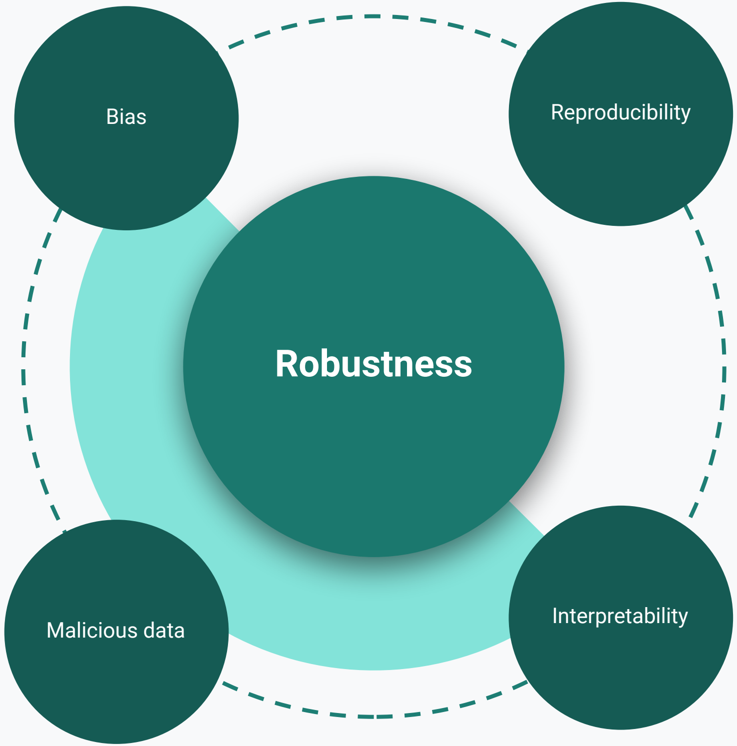
optimization landscape





Weaknesses

A SWOT Analysis




S
Strengths
W
Weaknesses
O
Opportunities
T
Threats

Robustness


rifle
shield, buck
revolver, si

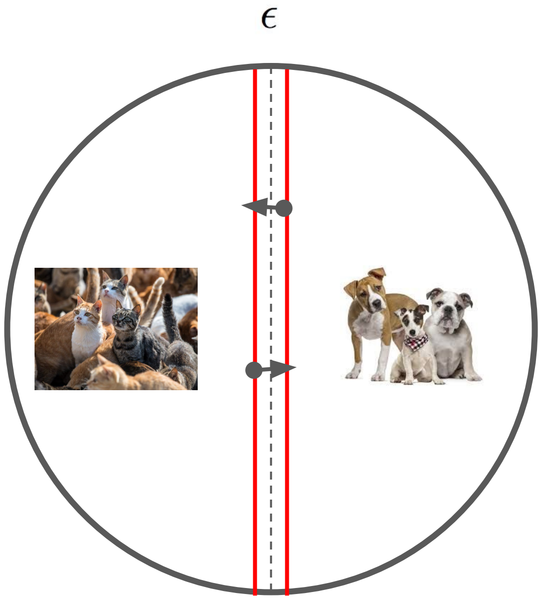

THIS IS FINE.

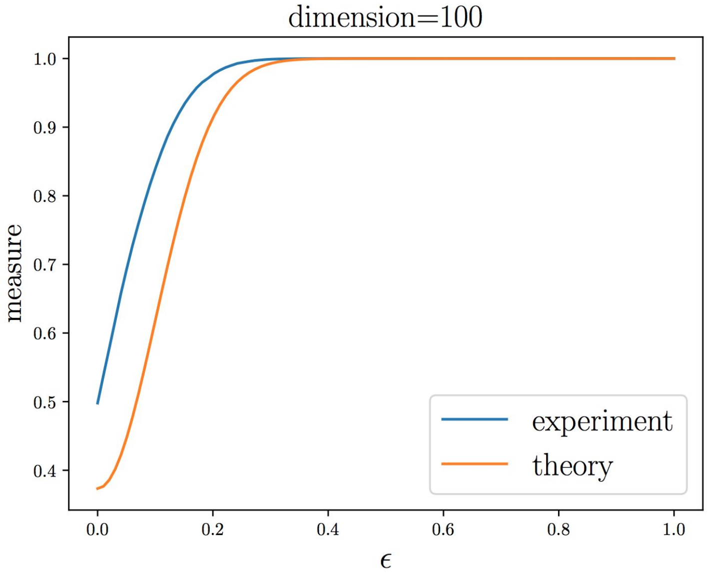

Self-Driving

Robustness is an active research area

- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep Residual Learning for Image Recognition. arXiv e-prints, page arXiv:1512.03385.
- Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. arXiv e-prints, page arXiv:1608.06993.
- Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. In International Conference on Learning Representations.
- Raghunathan, A., Steinhardt, J., and Liang, P. S. (2018). Semidefinite relaxations for certifying robustness to adversarial examples. Neurips.
- Wong, E. and Kolter, Z. (2018). Provable defenses against adversarial examples via the convex outer adversarial polytope. ICML.
- Madry, Aleksander and Makelov, Aleksandar and Schmidt, Ludwig and Tsipras, Dimitris and Vladu, Adrian. Towards Deep Learning Models Resistant to Adversarial Attacks. ICLR.
- Huang, X., Kwiatkowska, M., Wang, S., and Wu, M. (2017). Safety verification of deep neural networks. Computer Aided Verification.

Adversarial examples are inevitable!



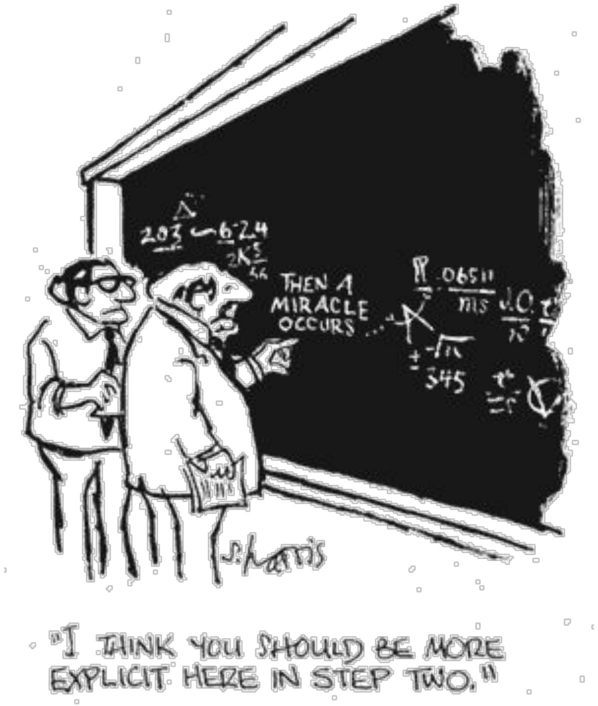


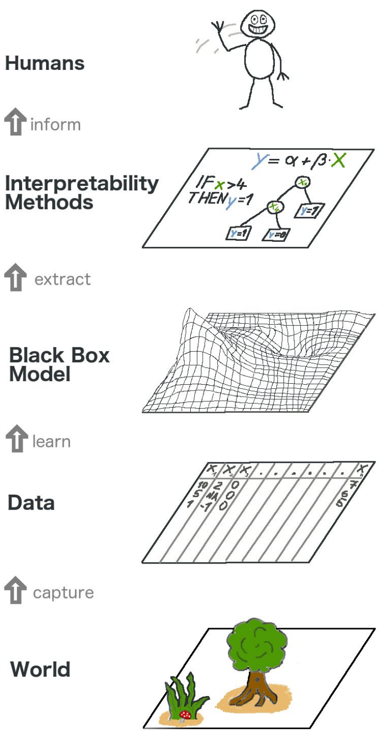
- Concentration-of-measure phenomenon

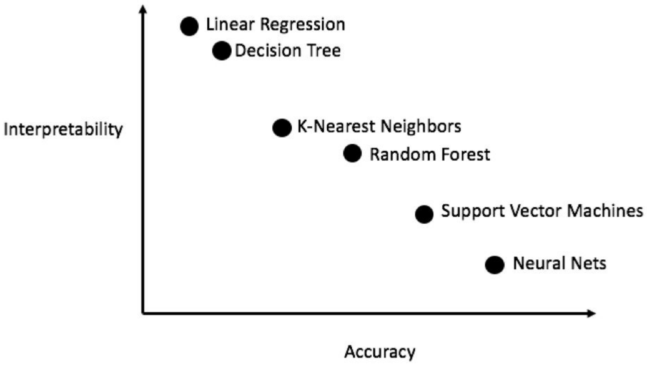
[Shafahi et al. ICLR 2019]

Interpretability

S Strengths | W Weaknesses | O Opportunities | T Threats







Interpretability in ML is an active research field

- Baehrens, David and Schroeter, Timon and Harmeling, Stefan and Kawanabe, Motoaki and Hansen, Katja and Mueller, Klaus-Robert. Simonyan, Karen and Vedaldi, Andrea and Zisserman, Andrew. How to Explain Individual Classification Decisions. JMLR 2010.

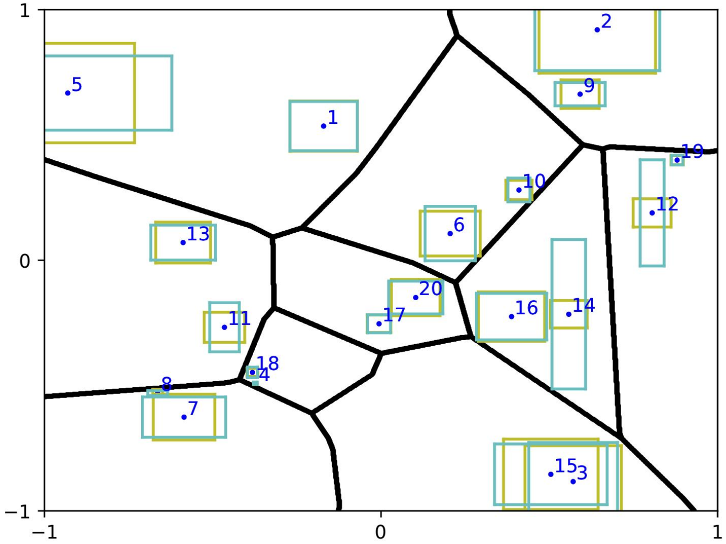
- Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv e-prints. arXiv:1312.6034. 2013.

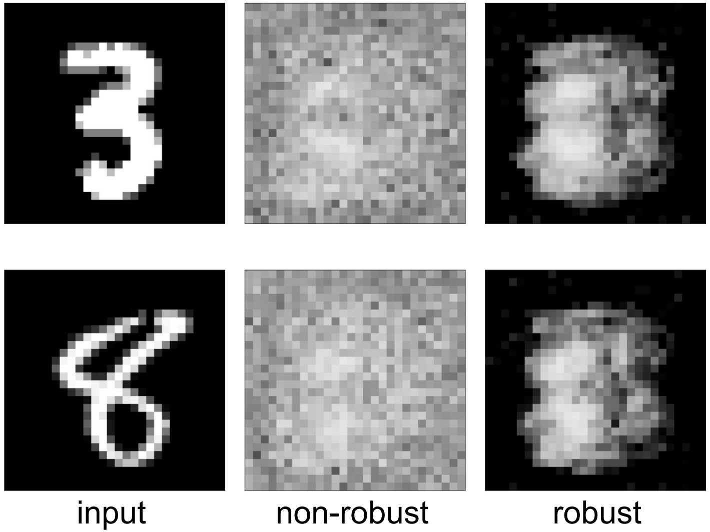
- Ribeiro, Marco and Singh, Sameer and Guestrin, Carlos. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. KDD 2016.

- Sundararajan, Mukund and Taly, Ankur and Yan, Qiqi. Axiomatic Attribution for Deep Networks. ICML'17.

- Shrikumar, Avanti and Greenside, Peyton and Kundaje, Anshul. Learning Important Features Through Propagating Activation Differences. ICML'17.

A robustness & interpretability result from my own work



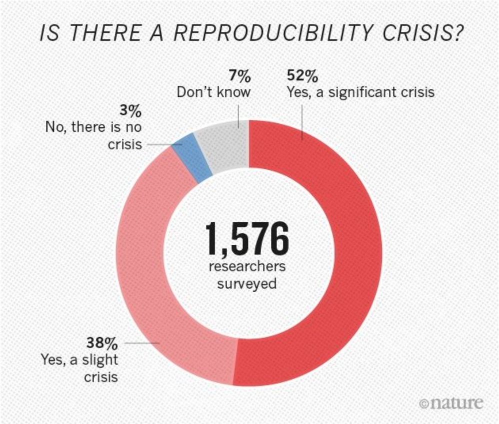


On Certifying Non-Uniform Bounds against Adversarial Attacks.
Liu, Chen and Tomioka, Ryota and Cevher, Volkan. ICML'19.

Many other weaknesses

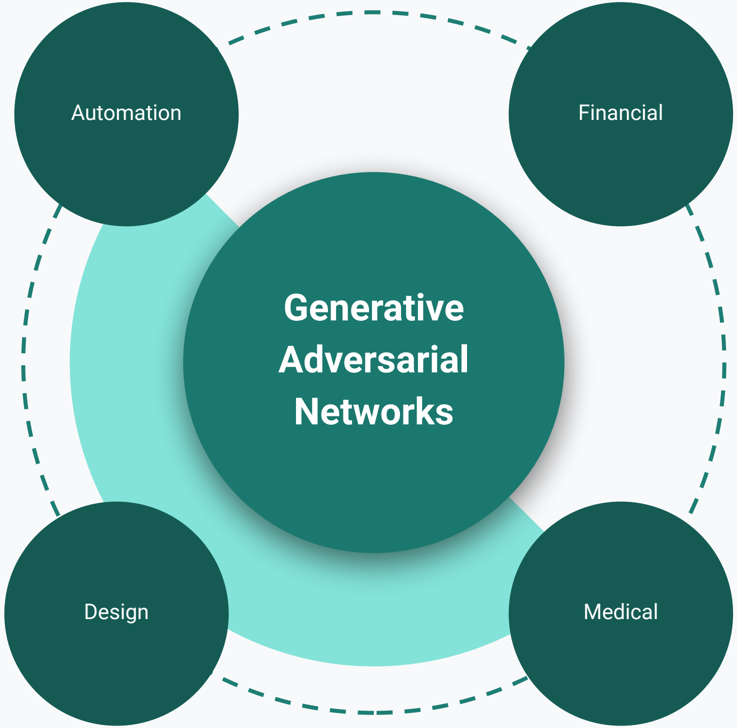
1. Bias
2. Malicious data
3. Reproducibility

I am Tay



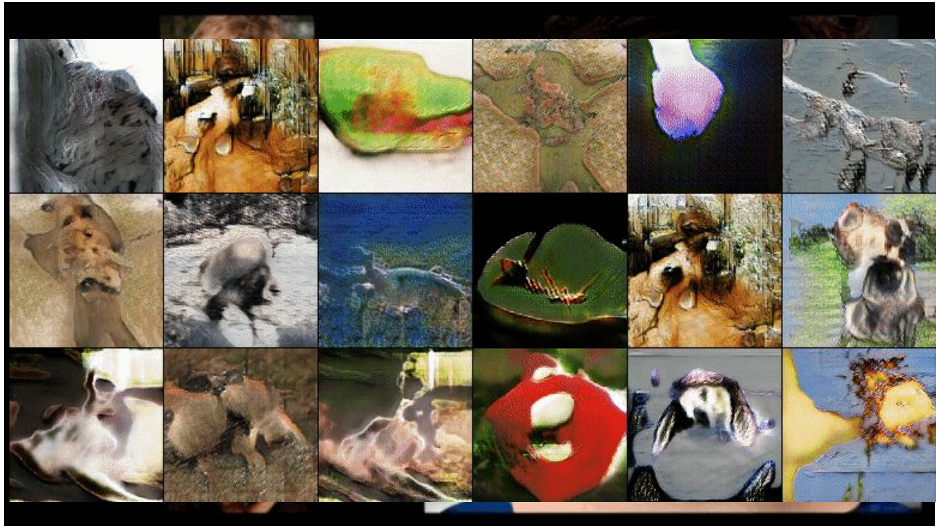
Opportunities

A SWOT Analysis



Generative Adversarial Networks

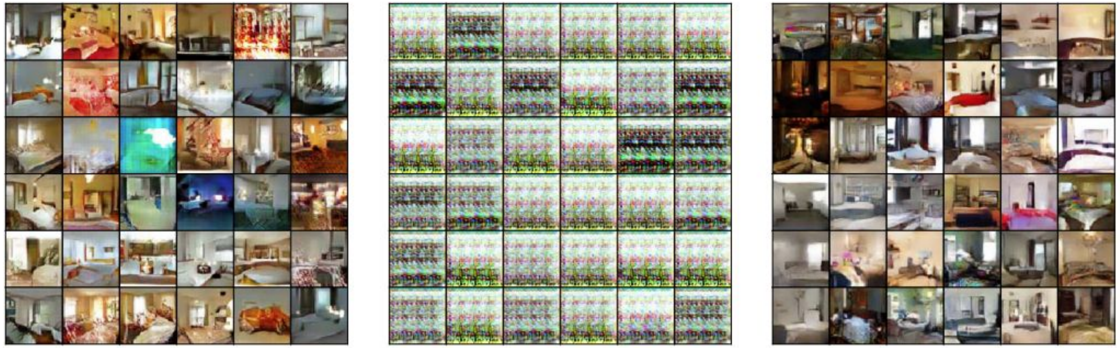
Progressive Growing of GANs for Improved Quality, Stability, and Variation
Karras et al. [ICLR 2018]



A style-based generator architecture for GANs
Karras et al. [2018]

High-Fidelity Image Generation With Fewer Labels
Lucic M, Tschannen M*, Ritter M*, Zhai X, Bachem O,* Sylvain S [2019]

Applications + Highlights from my own work

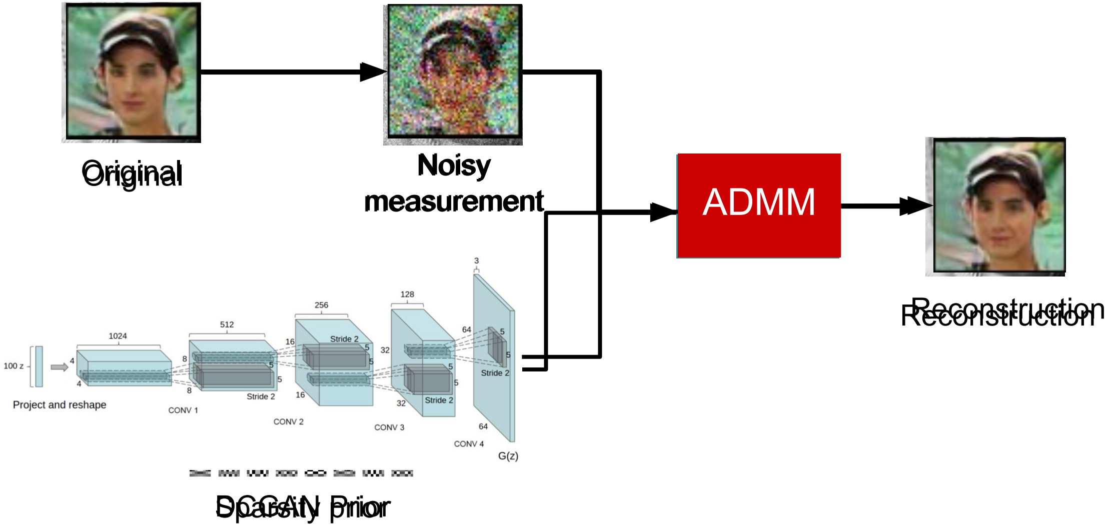


(a) RMSProp (b) Adam (c) Mirror-GAN

Finding Mixed Nash Equilibria of Generative Adversarial Networks
Hsieh et al. [ICML 2019]

+ Uncertainty quantification extension with Schlumberger (Boston)

Applications + Highlights from my own work

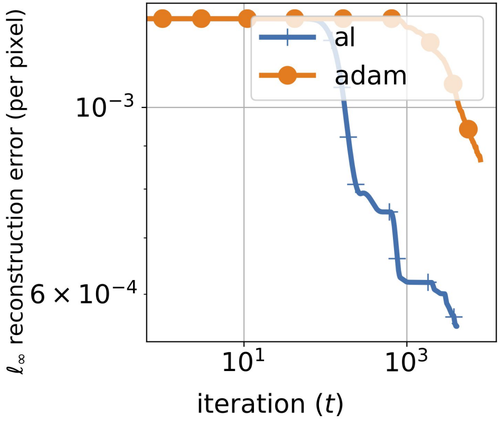


Fast and Provable ADMM for learning with generative priors.
Latorre, F. et Al. [NeurIPS 2019]

HASLERSTIFTUNG

Applications + Highlights from my own work

From clustering to adversarial robustness…

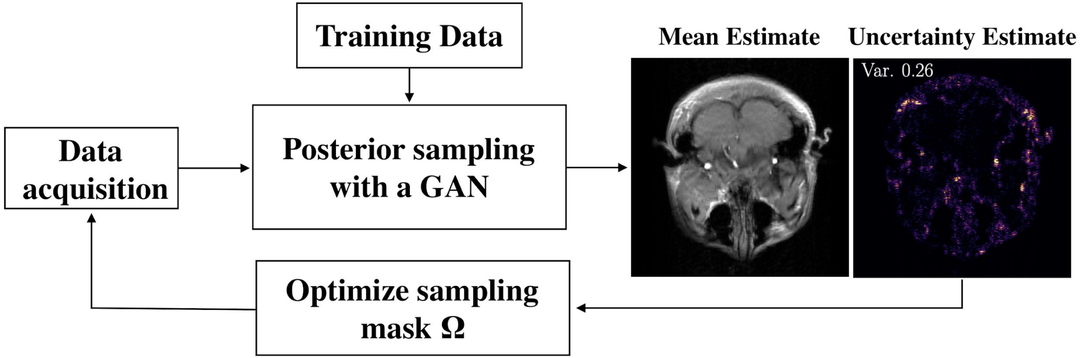


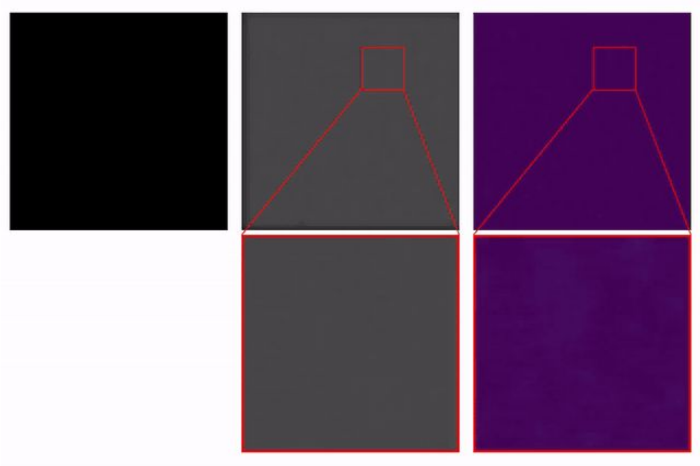
Decision Intelligence via semidefinite programming

An Inexact Augmented Lagrangian Framework for Nonconvex Optimization with Nonlinear Constraints. *Sahin M. F. et. al. [NeurIPS 2019]*

HASLERSTIFTUNG

Applications + Highlights from my own work





HASLERSTIFTUNG

Closed loop deep Bayesian inversion: Uncertainty driven acquisition for fast MRI. Sanchez et al. Under review.

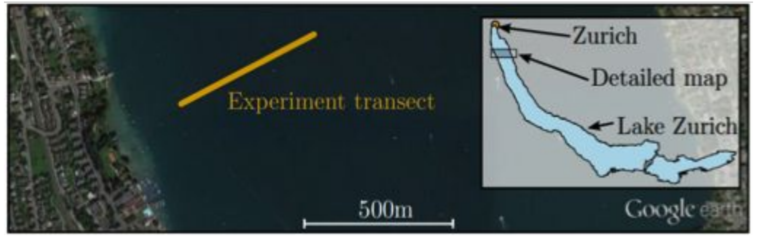
Applications + Highlights from my own work

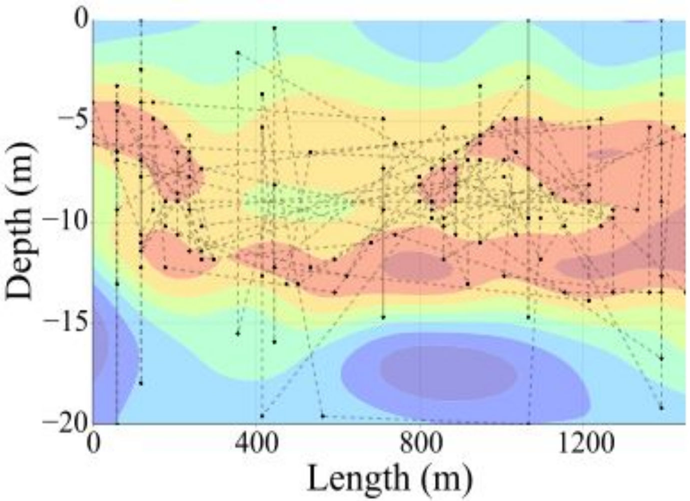
Additional fundamental trade-offs published at leading venues thanks to Hasler:

- *Optimal rates for spectral algorithms with least-squares regression over hilbert spaces.*
  Lin, J. et al. [ATCHA 2018]
- *Optimal Convergence for Distributed Learning with Stochastic Gradient Methods and Spectral Algorithms.*
  Lin, J. and Cevher, V. [ICML 2018]
- *Optimal rates of sketched-regularized algorithms for least-squares regression over Hilbert spaces.*
  Lin, J. and Cevher, V. [ICML 2018]

HASLERSTIFTUNG

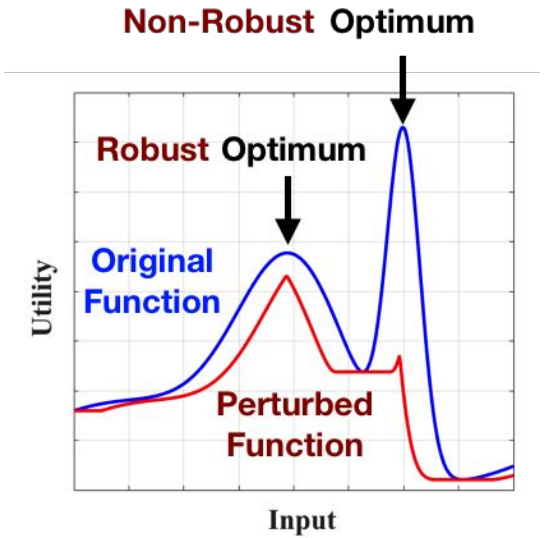
Applications + Highlights from my own work

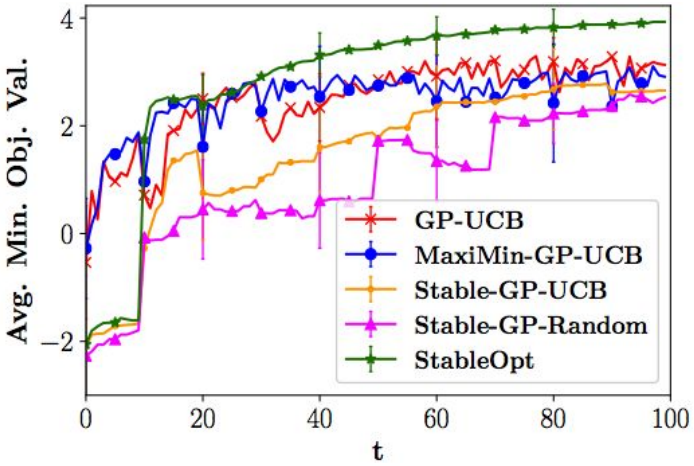




Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation Bogunovic et al. [NIPS 2017]

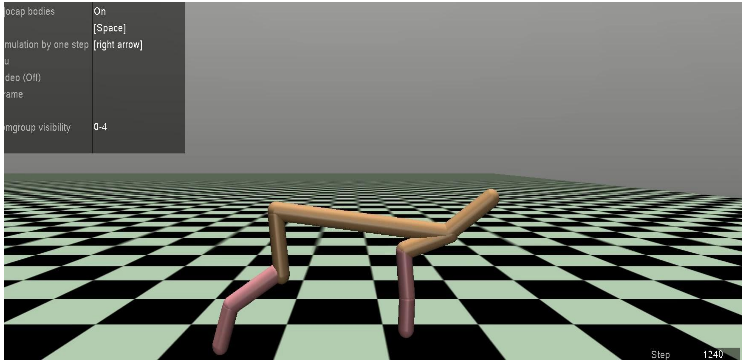
Applications + Highlights from my own work

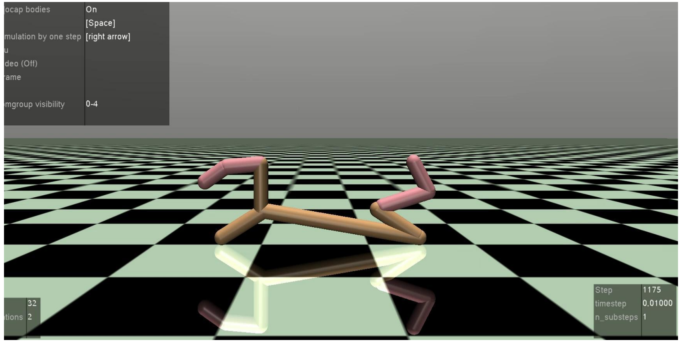


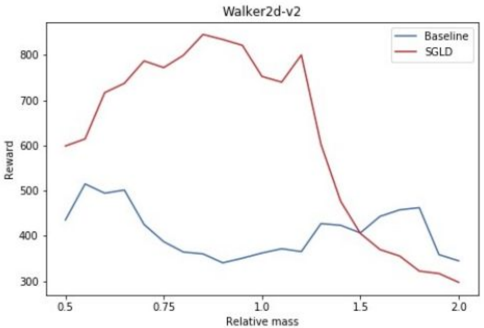


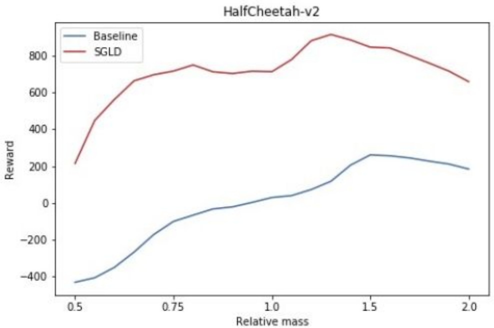
Adversarially robust Gaussian Process Optimization Bogunovic et al. [NeurIPS 2018]

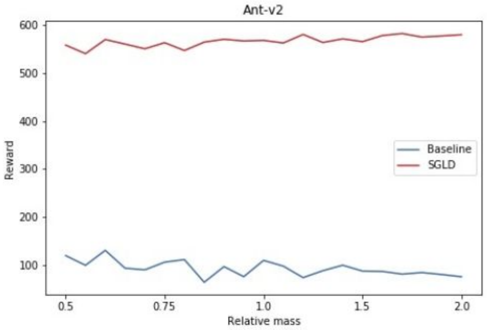
Applications + Highlights from my own work





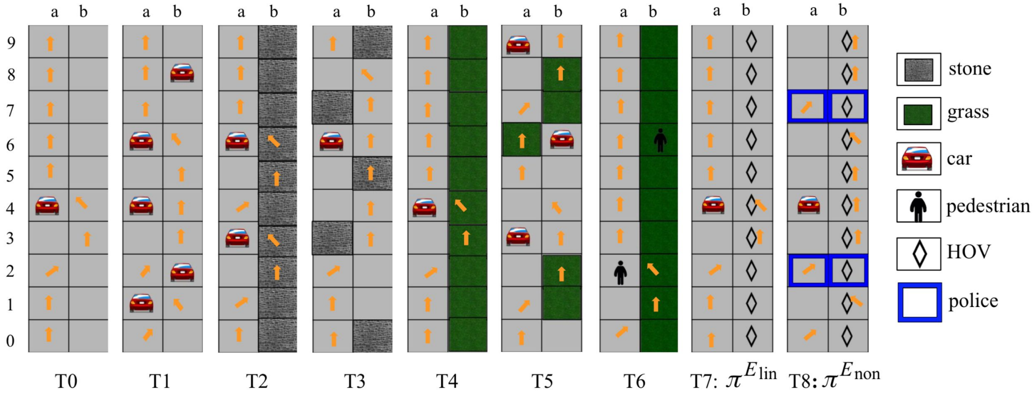


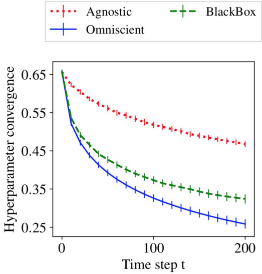


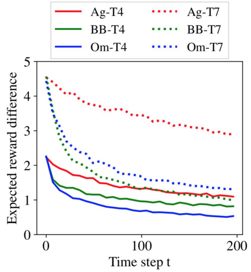


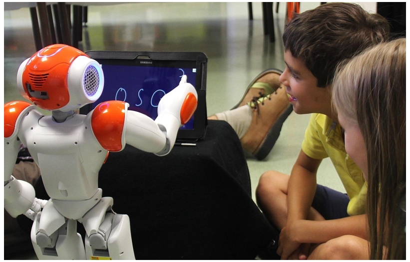
Robust Reinforcement Learning with Langevin Dynamics. Under review.

Applications + Highlights from my own work



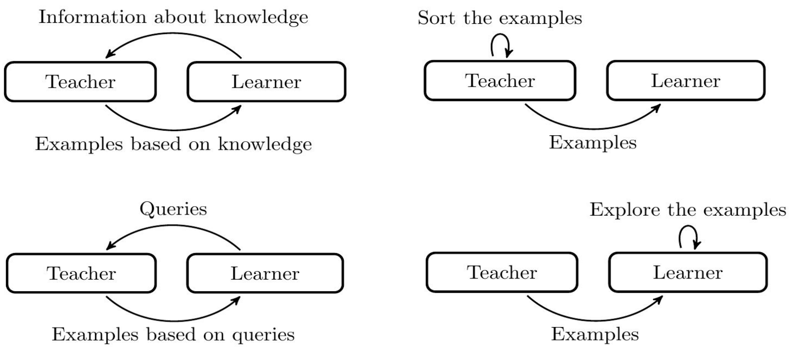


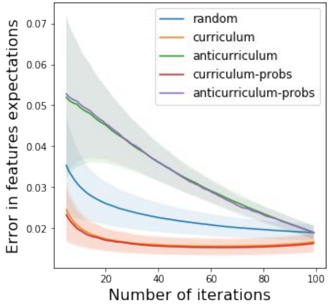


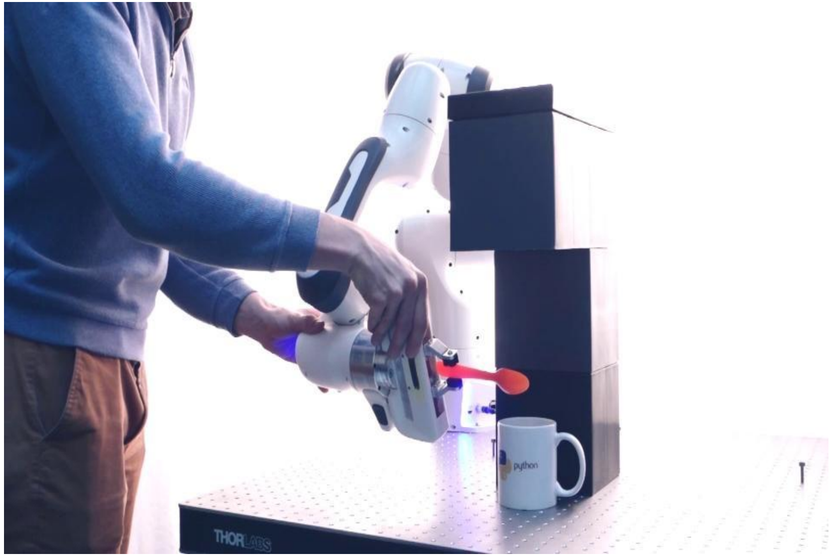


Interactive Teaching Algorithms for Inverse Reinforcement Learning.
Kamalaruban et al. [IJCAI 2019]

Applications + Highlights from my own work



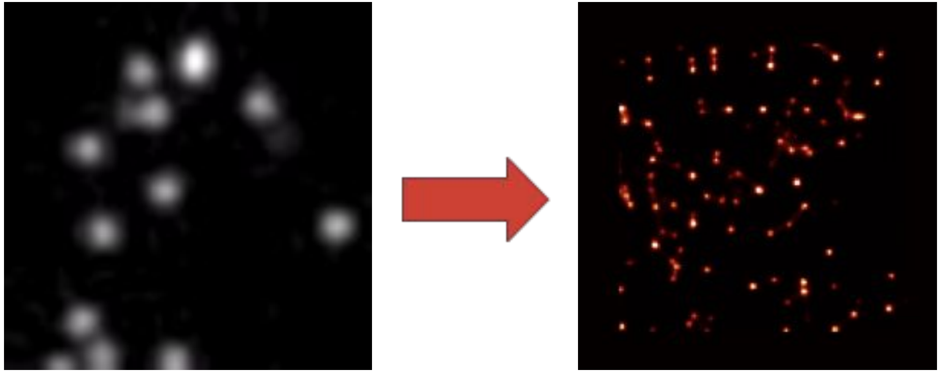




Interaction-limited Inverse Reinforcement Learning. Under review AAAI.

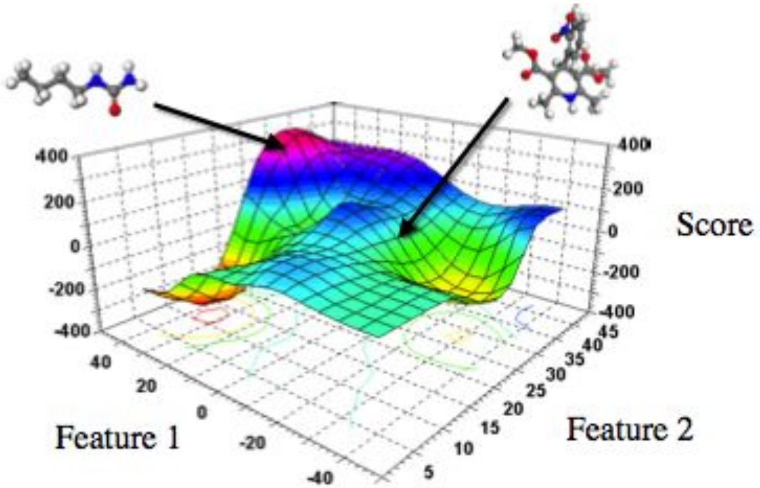
Applications + Highlights from my own work

Single Molecule Localization Microscopy



Strategies for increasing the throughput of super-resolution microscopies
Mahecic et al. [Current Opinion in Chemical Biology]

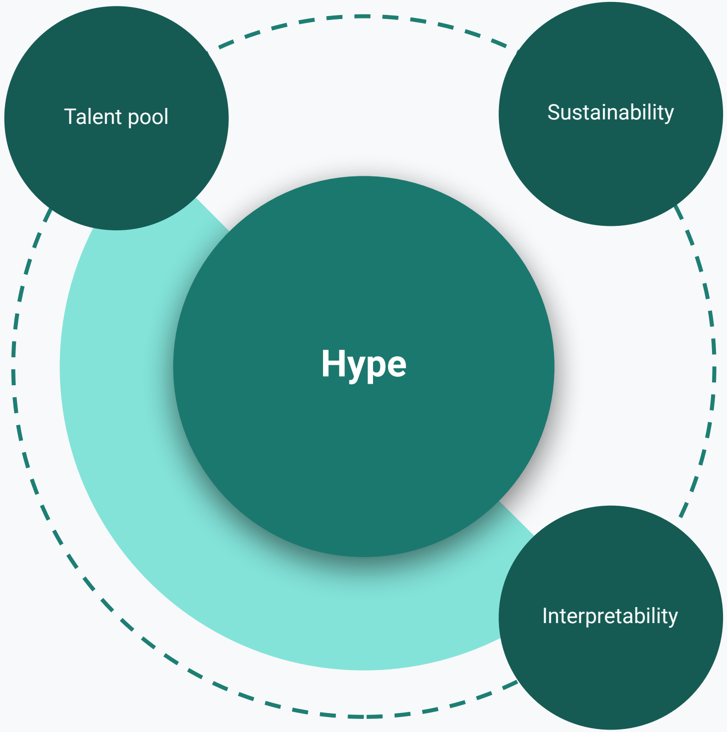
Applications + Highlights from my own work



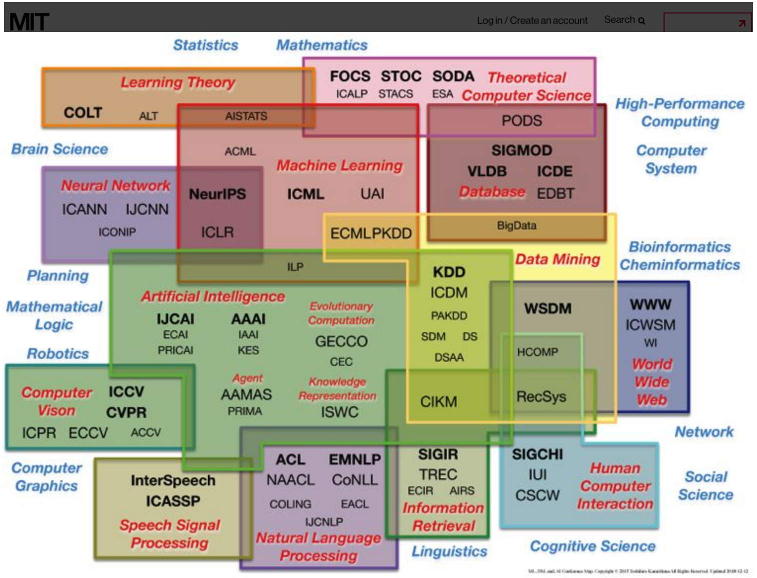
Chemical machine learning with kernels: The impact of loss functions.
Van Nguyen et al. [Quantum Chemistry 2019]

Threats

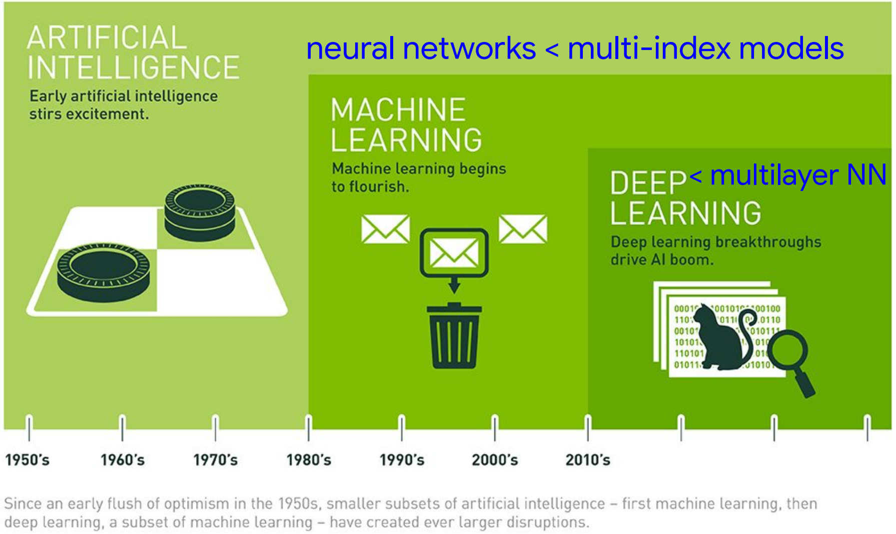
A SWOT Analysis



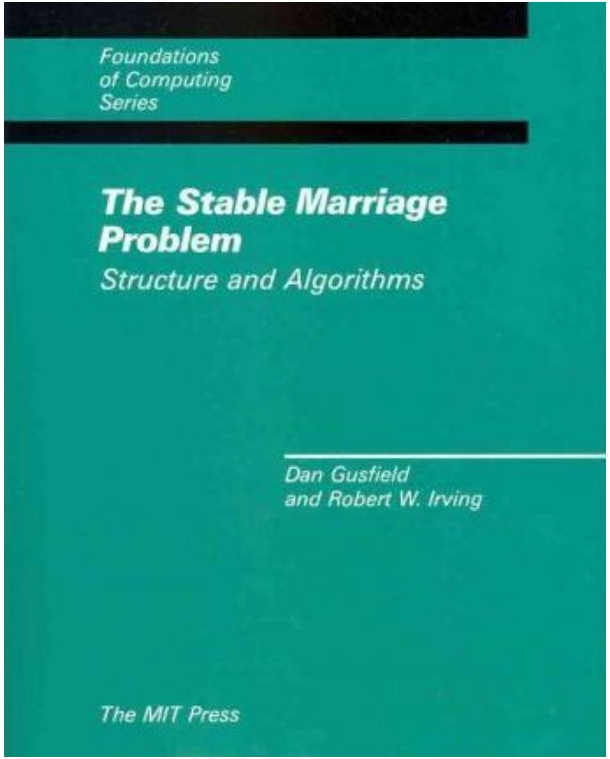
The AI hype vs the ML revolution



A surprising number of firms are jumping on the artificial-intelligence bandwagon—without actually investing in any AI.



Talent pool: Missing the top talent vs the needed talent



S Strengths | W Weaknesses | O Opportunities | T Threats

Sustainability:

The estimated costs of training a model

	Date of original paper	Energy consumption (kWh)	Carbon footprint (lbs of CO2e)	Cloud compute cost (USD)
Transformer (65M parameters)	Jun, 2017	27	26	$41-$140
Transformer (213M parameters)	Jun, 2017	201	192	$289-$981
ELMo	Feb, 2018	275	262	$433-$1,472
BERT (110M parameters)	Oct, 2018	1,507	1,438	$3,751-$12,571
Transformer (213M parameters) w/ neural architecture search	Jan, 2019	656,347	626,155	$942,973-$3,201,722
GPT-2	Feb, 2019	-	-	$12,902-$43,008

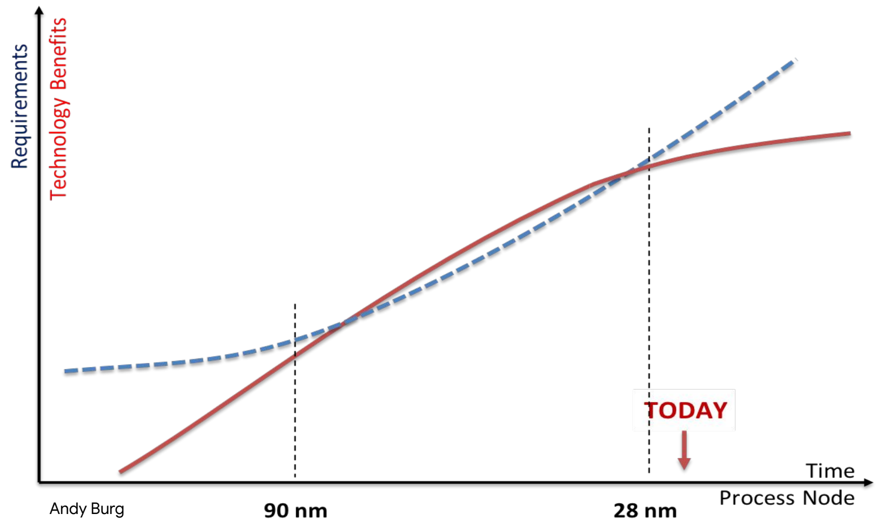
Note: Because of a lack of power draw data on GPT-2's training hardware, the researchers weren't able to calculate its carbon footprint.

Table: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

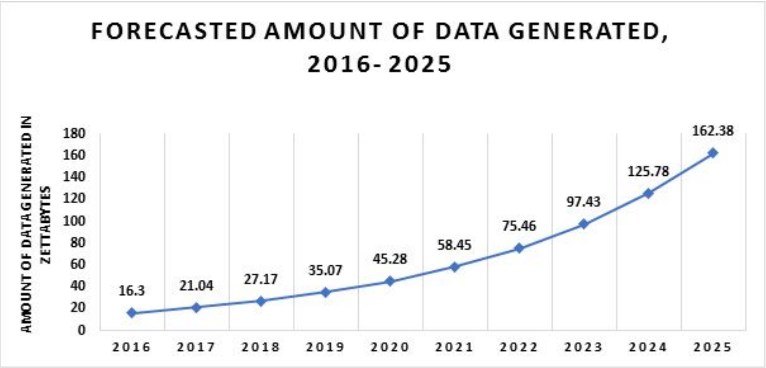
S Strengths W Weaknesses O Opportunities T Threats

Sustainability:

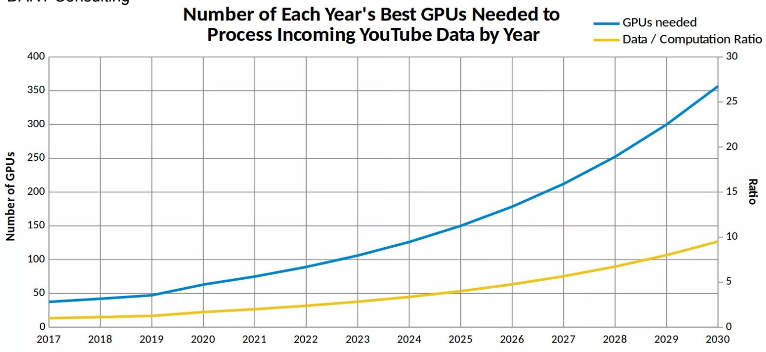
Dennard scaling & Moore's law vs Growth of data



Andy Burg



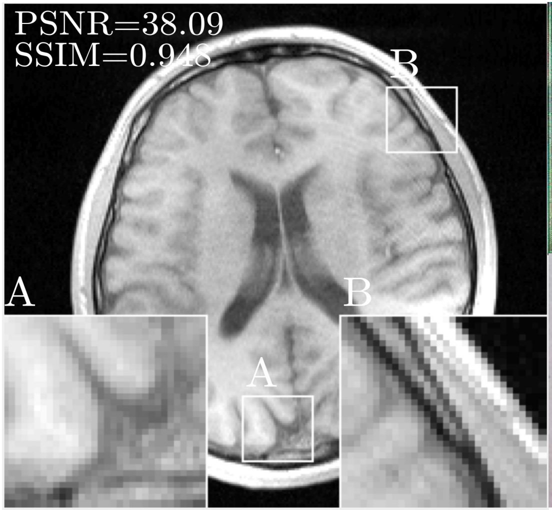
DART Consulting

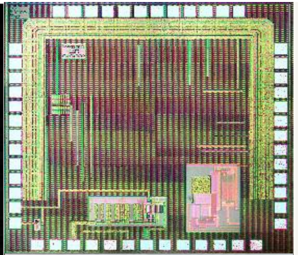


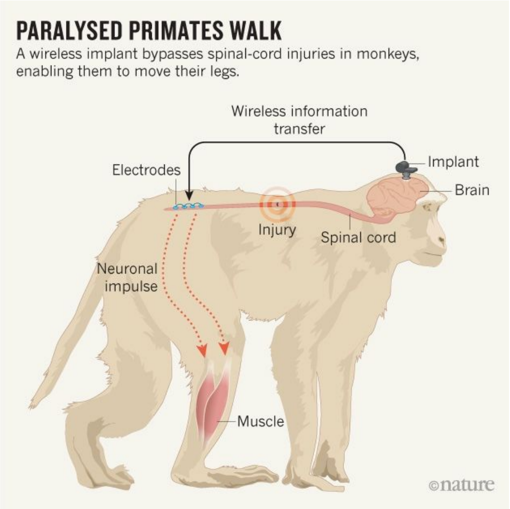
Tim Dettmers

Sustainability:

Energy constraints / Time constraints







Learning-based compressive sensing + hardware design. *Baldassarre et al., Gozcu et al., Aprile et al. [IEEE TMI, IEEE TSP, IEEE CnS, IEEE TCAS]*

HASLERSTIFTUNG

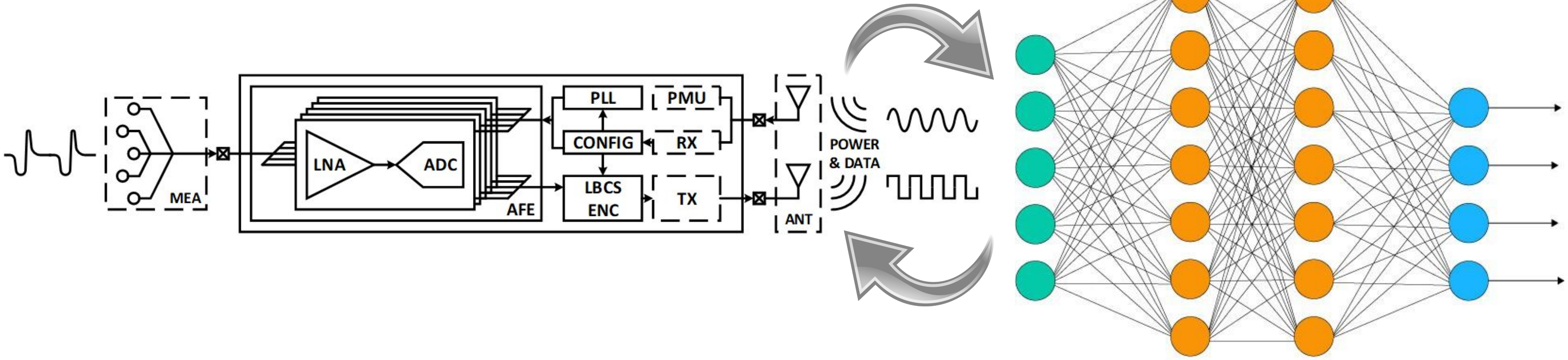
IBM Thesis Award 2019

Sustainability:

Energy constraints of recording neural data

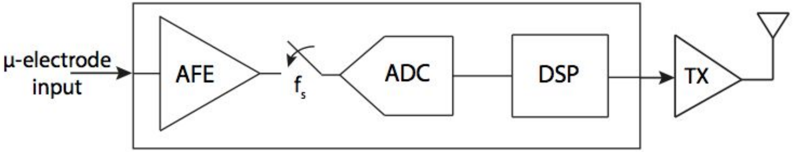
Hardware/software co-design

- **Informed design of hardware** of application needs
- **Informed design of software** of hardware capabilities

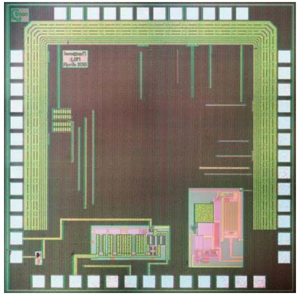


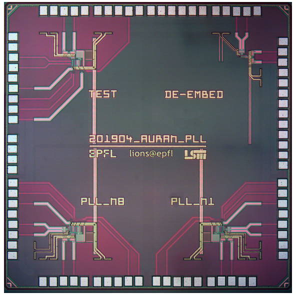
Sustainability:

Energy constraints of recording neural dat

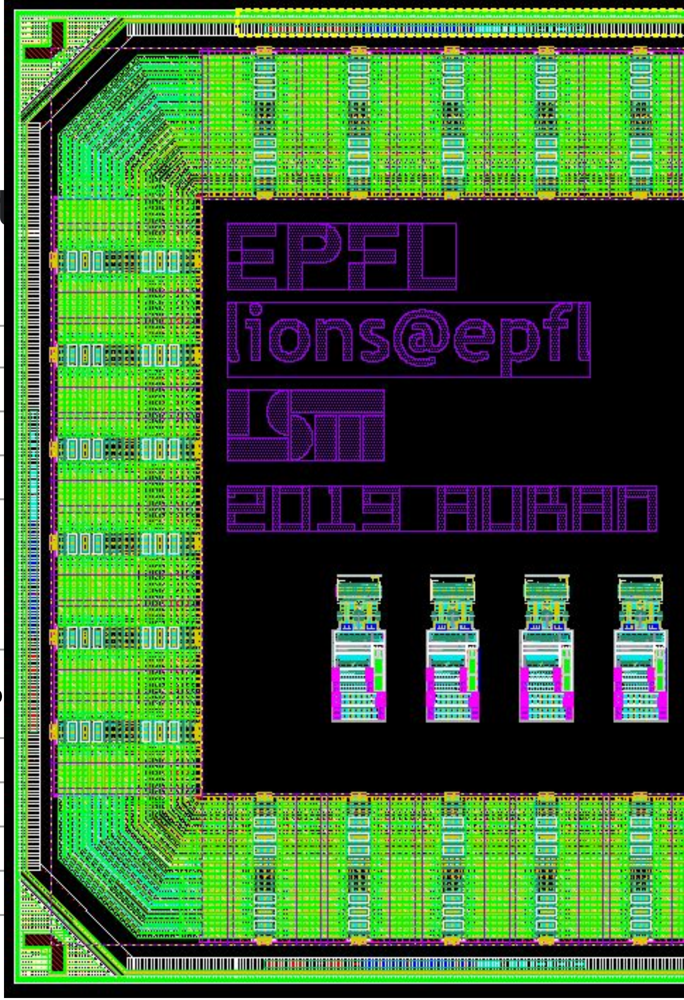


> 30 dB quality
AFE + ADC
DSP
TX





Metho
LBCS
SHS
BERN
MCS

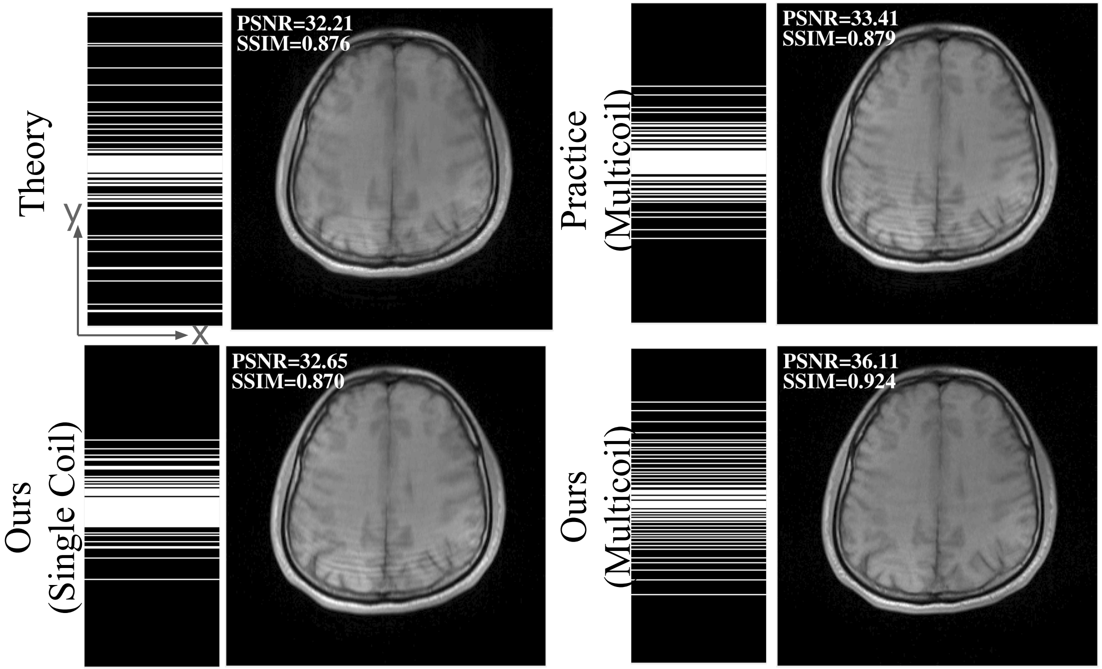


HASLERSTIFTUNG

Sustainability:

Time constraints of MRI

- Accelerate the MRI scan 5 times.
- Pick the most relevant data only for your method.



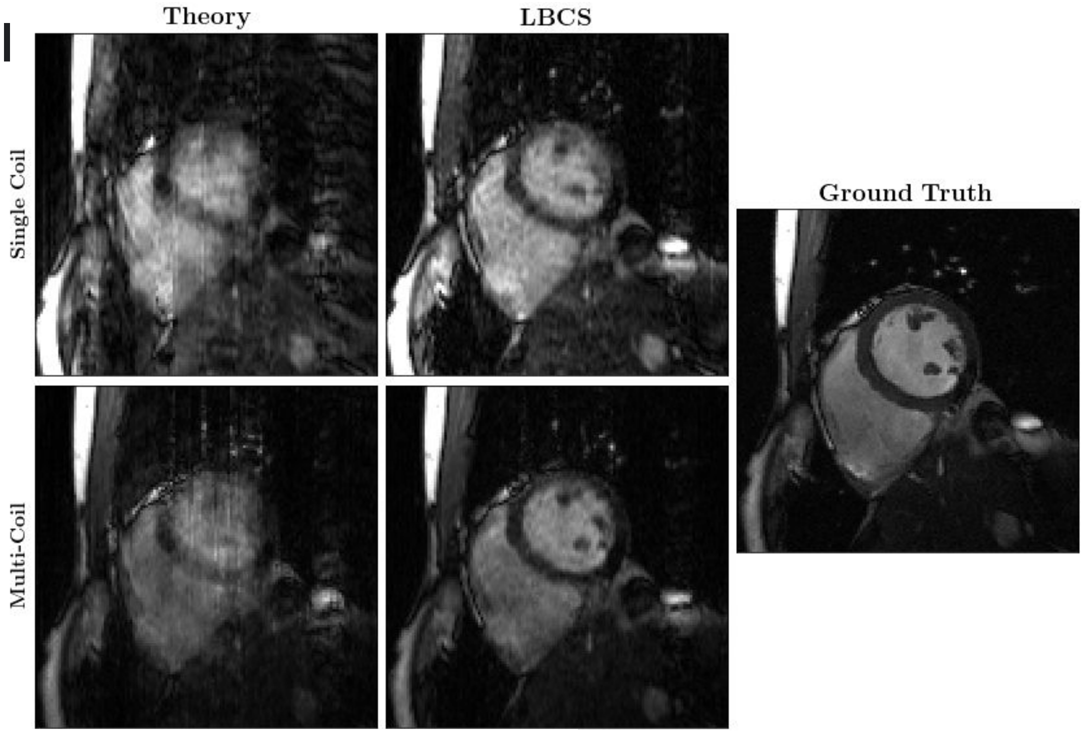
HASLERSTIFTUNG

Learning-based compressive MRI. Gözcü B., et al [IEEE TMI - 2018]
Rethinking Sampling in Parallel MRI: A Data-Driven Approach. Gözcü B. et al. [EUSIPCO 2019]

Sustainability:

Time constraints of MRI

- Time drastically increases the dimensionality of data
- Reduce computations by a factor 200: from a month to 4 hours without losing performance.

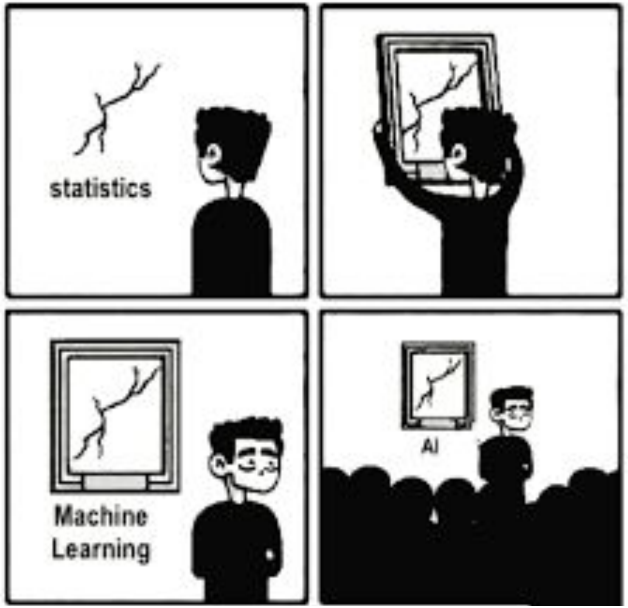


HASLERSTIFTUNG

Scalable learning-based sampling optimization for compressive dynamic MRI.
Sanchez T., et al. Under review.

Conclusions

- Are you wiser?
  - time-data-energy trade offs
- Existential threats = “Opportunities”
  - talk to me offline
- ML partnerships with EPFL & Hasler
  - Hype protection
- Thanks for the support!
  volkan.cevher@epfl.ch
  https://ml.epfl.ch



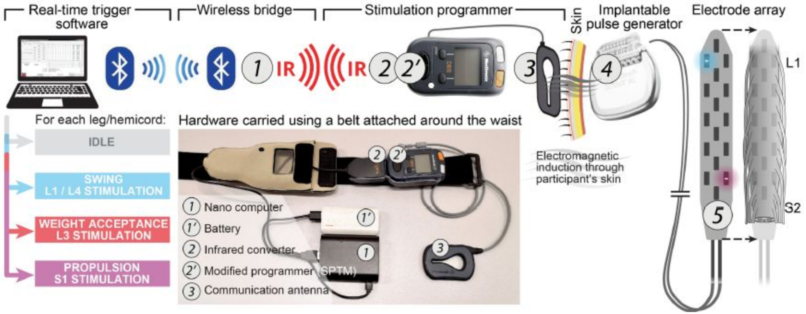
Neural Interfaces for Cyber-Human Systems

Neuroscientific progress in spinal cord injury (SCI) rehabilitation is fast.

Proof of concept on rats to humans in 6 years.

Enabling hardware lags behind.

We are ambitious to fix that.



Hardware challenges in neural interface design

Recording the most of spatiotemporal information

Multichannel, high-resolution acquisition front end

How many channels fit in one chip?

Reliable long-term use

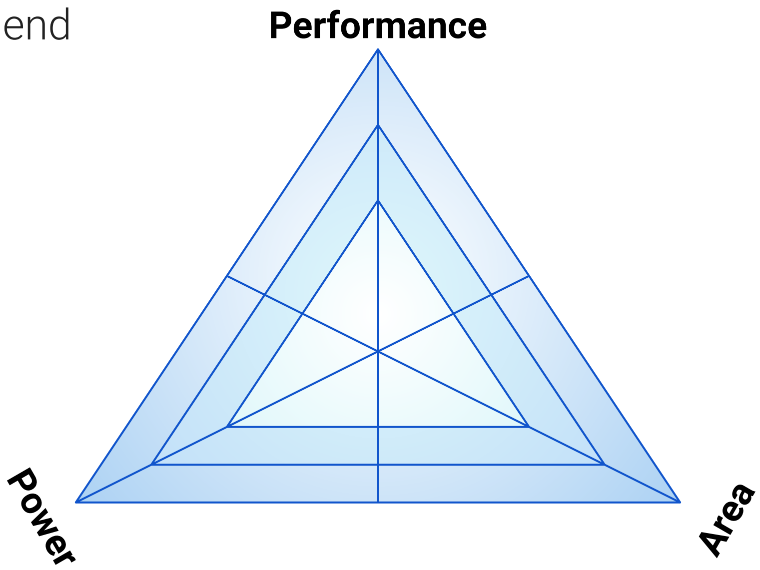
Wireless power and data transfer

How much energy is needed per bit sent?

Processing information

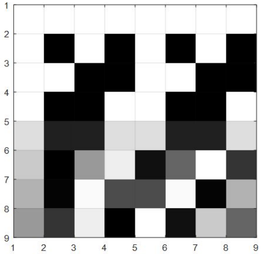
Extract/compress information

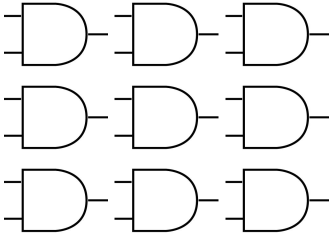
How well is the representation?

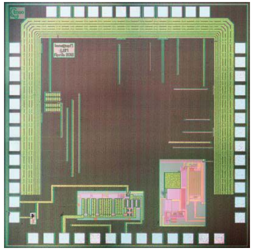


Neural Interface Project @ LIONS

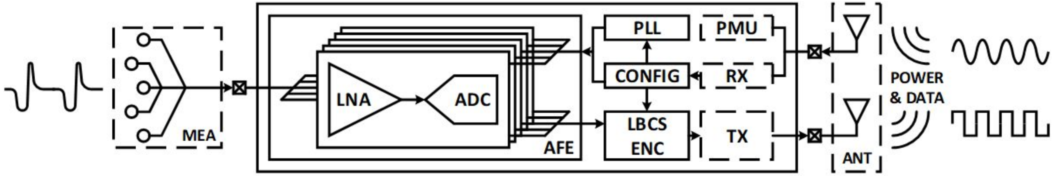
1. Sampling mask design
2. Description to digital hardware
3. ASIC implementation



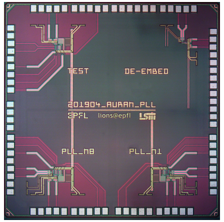


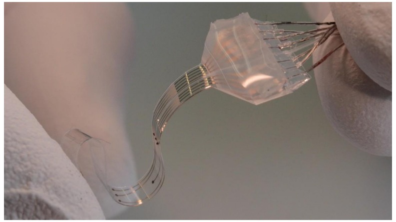


4. Mixed-signal sensor
5. Wireless connectivity*
6. SoC implementation



7. Probe integration*
8. In-vivo validation*





* Collaborative work

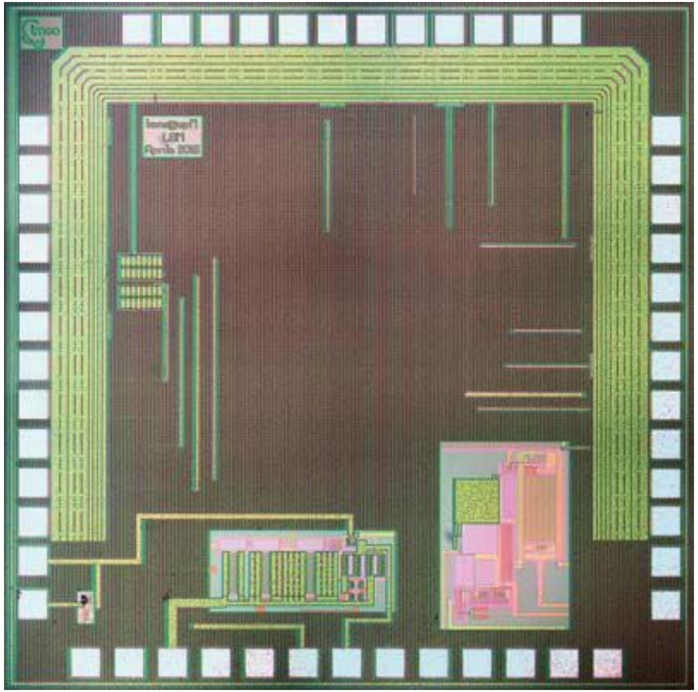
Our online compression solution: LBCS-on-chip

Learning-based Compressive Subsampling

- ✓ Reduces data rate
- ✓ Saves power on wireless transmitter
- ✓ Cost of compression less than transmission
- ✓ **Higher overall energy efficiency**

State-of-the art reconstruction performance

2018, UMC 180nm



C. Aprile et al., "Adaptive Learning-Based Compressive Sampling for Low-power Wireless Implants," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 11, pp. 3929-3941, Nov. 2018.

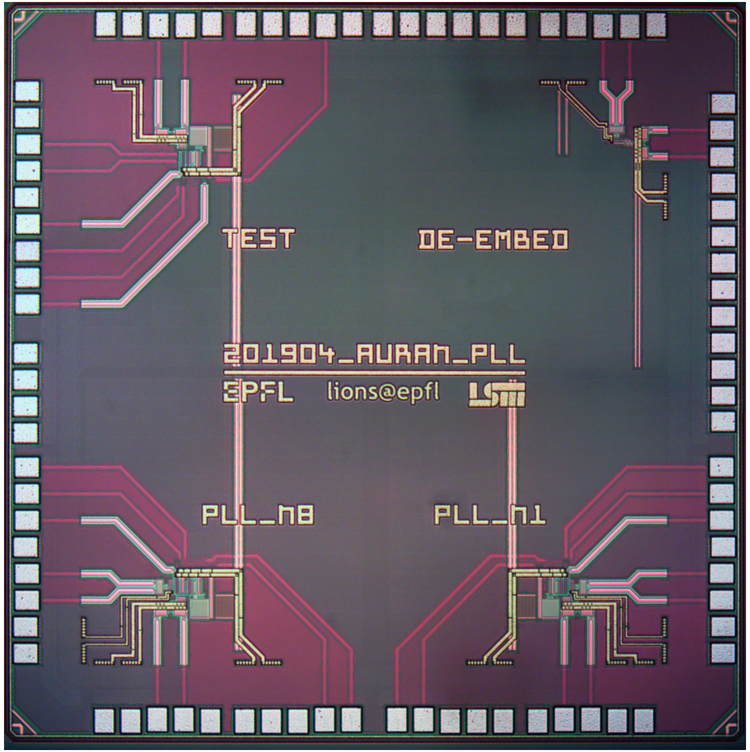
Prototype v2:

Multichannel timing compatibility with LBCS

On-chip clock generation and distribution

Ensures recording time synchrony

2019, TSMC 40nm



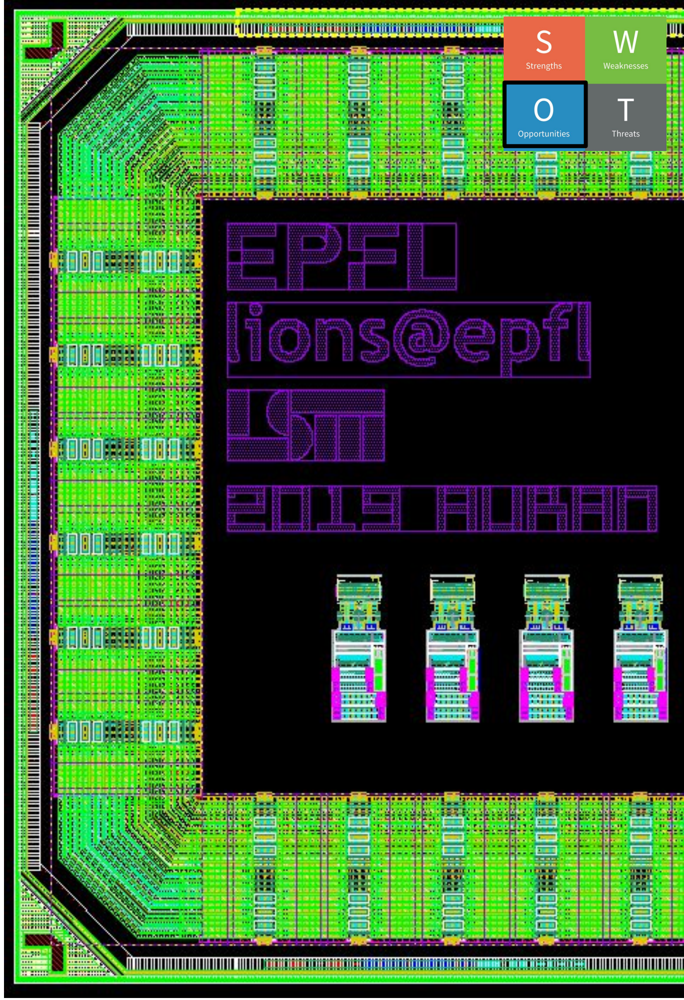
Prototype v3:

Electrode-to-wave signal chain

- Signal conditioning
- Digitization
- Wireless power and data transfer

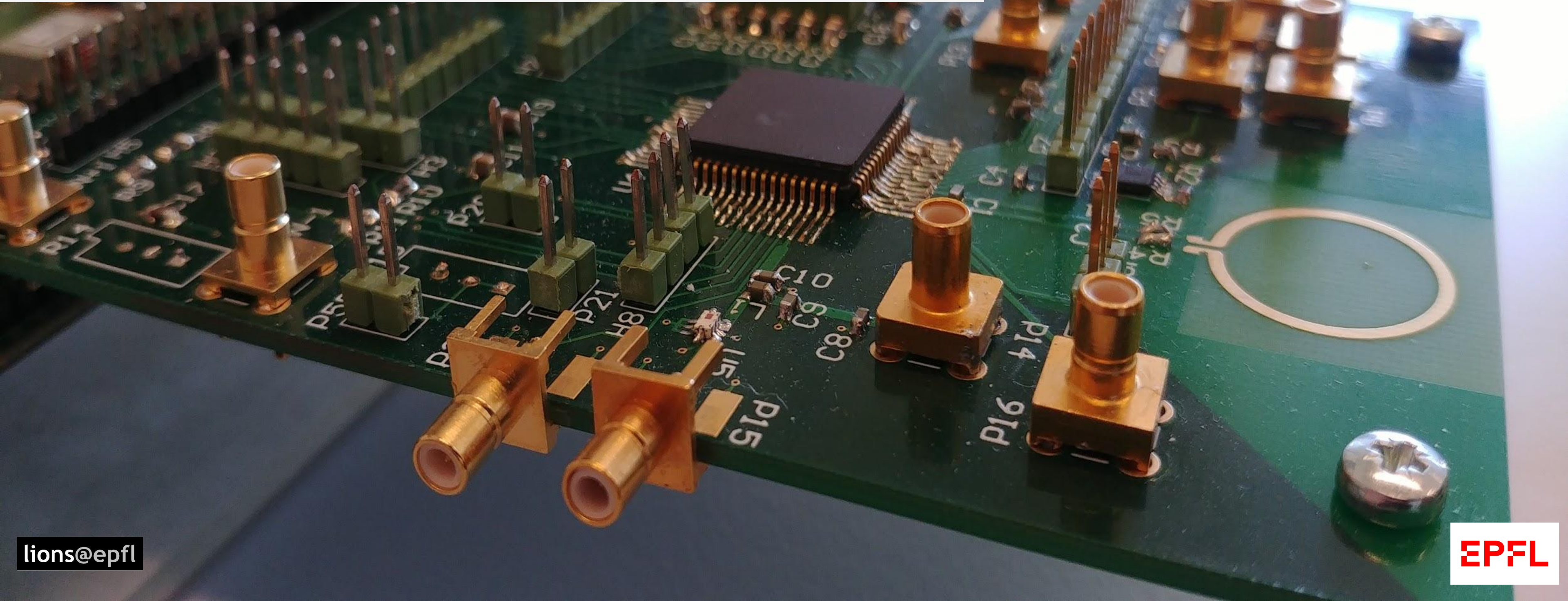
Beats SoA in performance, power and area*

2020, TSMC 65nm



*Based on simulation results

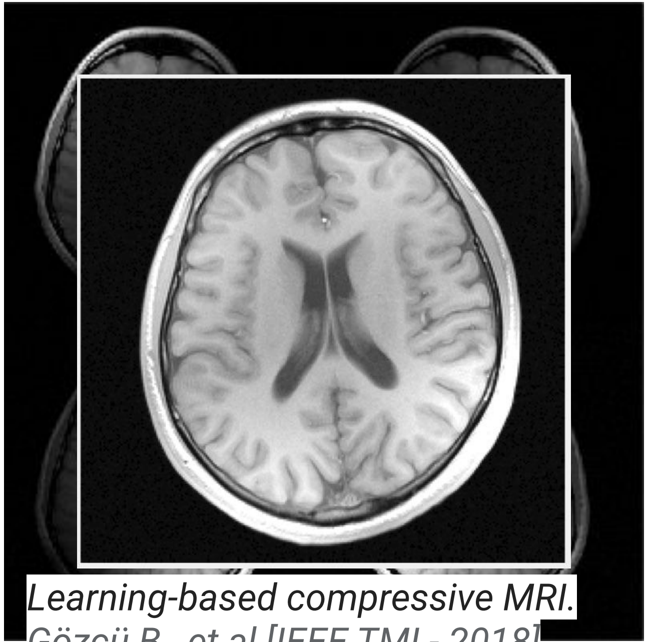
Neural Interface Project @ LIONS



Sustainability

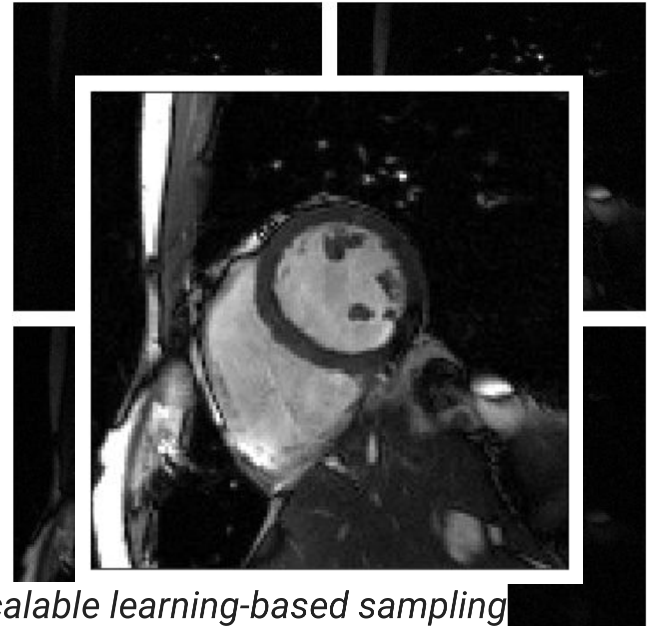
Scaling up LBCS: Use less data

HASLERSTIFTUNG



Learning-based compressive MRI. Gözcü B., et al [IEEE TMI - 2018]

Rethinking Sampling in Parallel MRI: A Data-Driven Approach. Gözcü B. et al. [EUSIPCO 2019]



Scalable learning-based sampling optimization for compressive dynamic MRI. Sanchez T., et al. Under review.

ing optimization for compressive dynamic MRI. Sanchez T. et al. Under review.

Sampling for Imaging

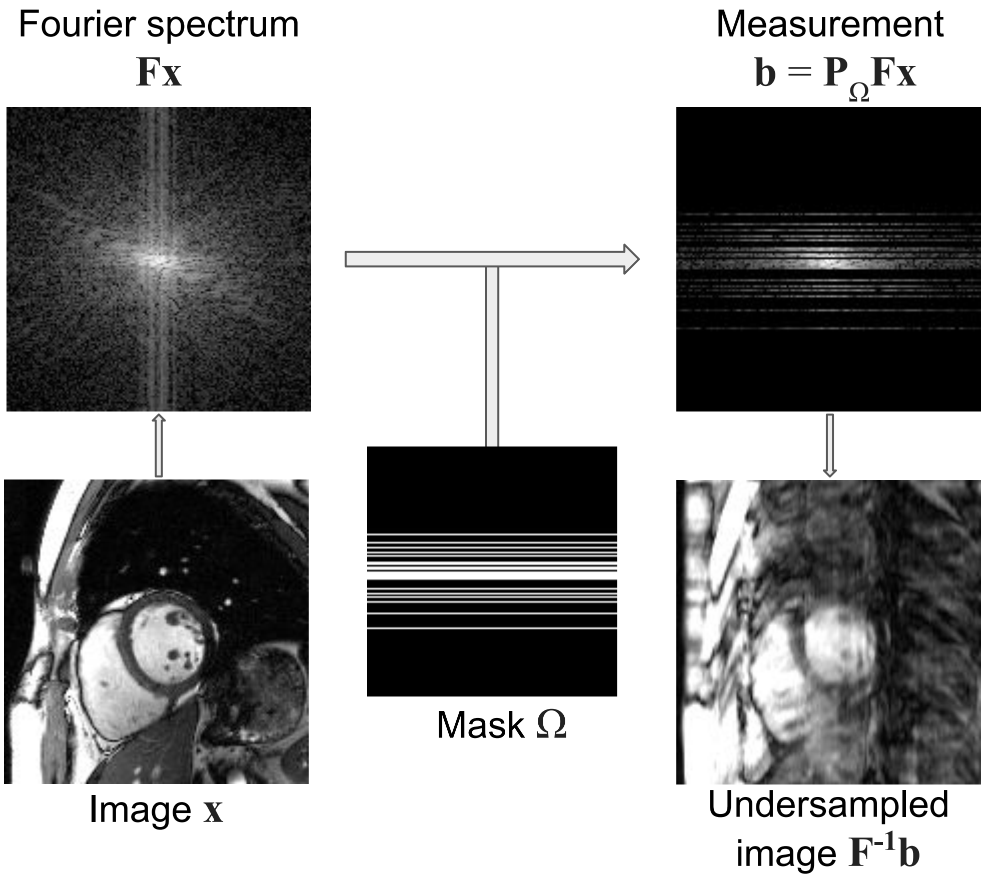
- Acquisition model:

$$\mathbf{b} = \mathbf{P}_{\Omega}\mathbf{F}\mathbf{x}$$

- Use a reconstruction algorithm to form an estimate of $\mathbf{x}$

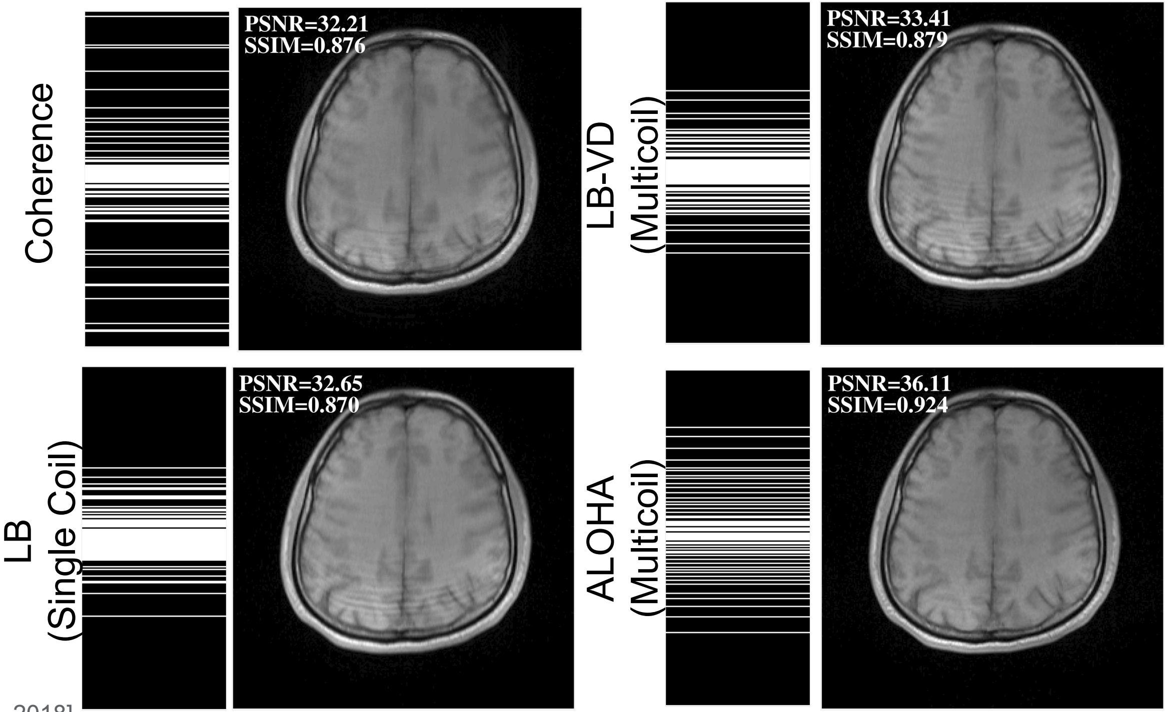
- How can we design a good sampling mask $\boldsymbol{\Omega}$ given a reconstruction method and anatomy ?

Solution: Data-driven approach



Mask design should not be isolated from reconstruction

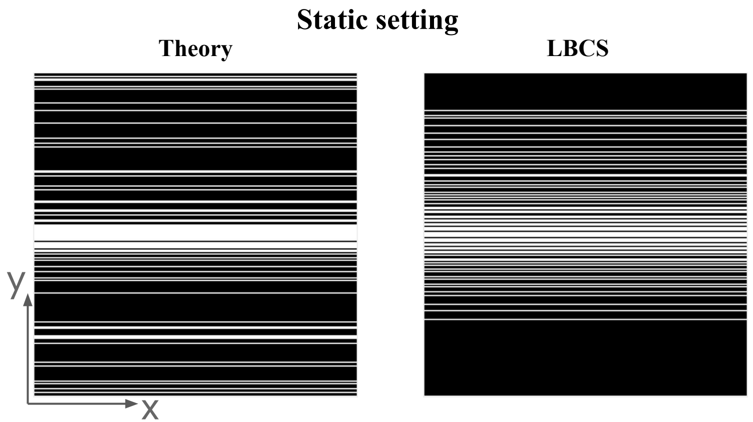
- Adapt to the reconstruction method, anatomy and imaging procedure (single vs multi-coil)
- Model-based approaches (e.g. variable density) are limiting
- **Theorem (informal):** Given a cardinality constraint, the optimal mask sampling distribution has a compact support.

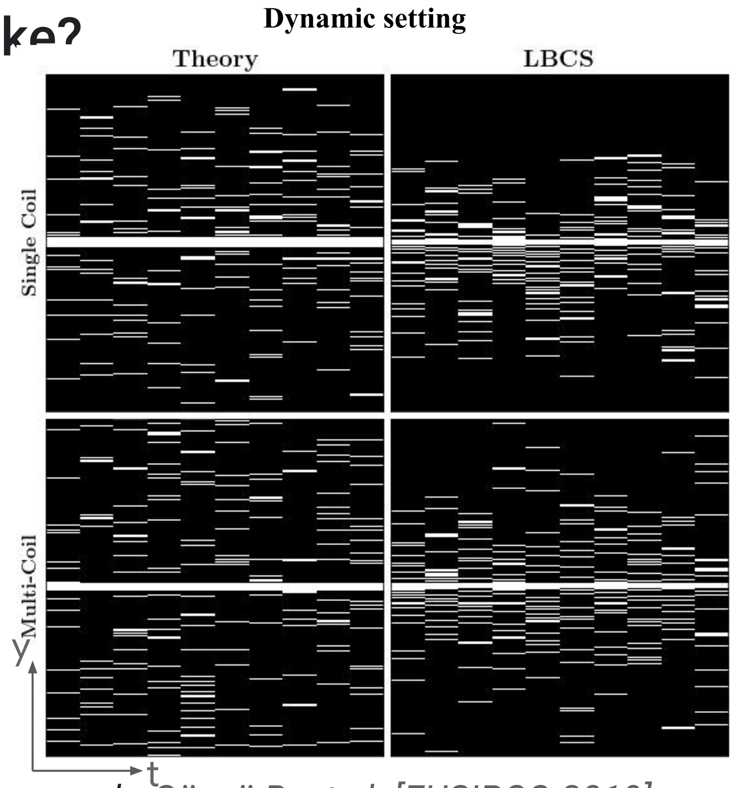


Learning-based compressive MRI. Gözcü B., et al [IEEE TMI - 2018]
Scalable learning-based sampling optimization for compressive dynamic MRI. Sanchez T., et al.(2019).
Rethinking Sampling in Parallel MRI: A Data-Driven Approach. Gözcü B. et al. (2019). [EUSIPCO 2019].

How do the sampling masks look like?

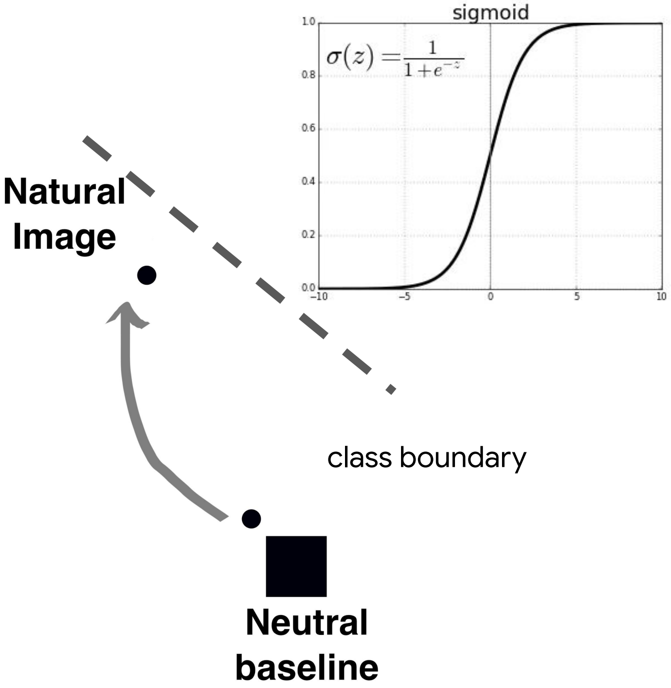
- **Static and dynamic:** our masks achieve structures that VD cannot obtain.





Rethinking Sampling in Parallel MRI: A Data-Driven Approach. *Gözcü B. et al. [EUSIPCO 2019]*
Scalable learning-based sampling optimization for compressive dynamic MRI.
Sanchez T., et al. Under review.

Integrated Gradients



Sundararajan, Mukund and Taly, Ankur and Yan, Qiqi, Axiomatic Attribution for Deep Networks. ICML'17