Machine learning for the design of magnonic computing devices - and some hints on experimental realization

Ádám Papp[1], Wolfgang Porod[2], Levente Maucha[1], György Csaba[1]

Markus Becherer[3], Martina Kiechle[3]

[1]Pazmany Peter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary

[2]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA

[3]Technische Universität München, Munich, Germany

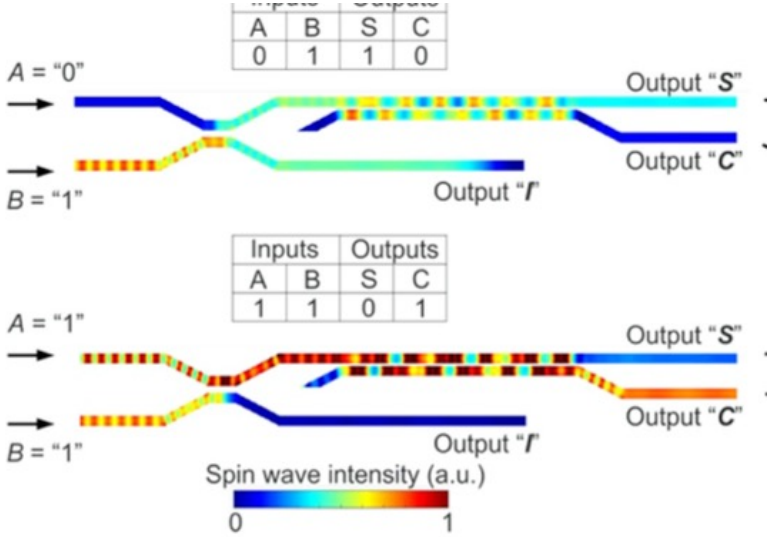
Outline

- **Optics-inspired spin wave devices**
- **The methodology of inverse design**
  - Machine learning integrated with micromagnetics
- **Design of neural networks**
- **Design of complex reservoirs**

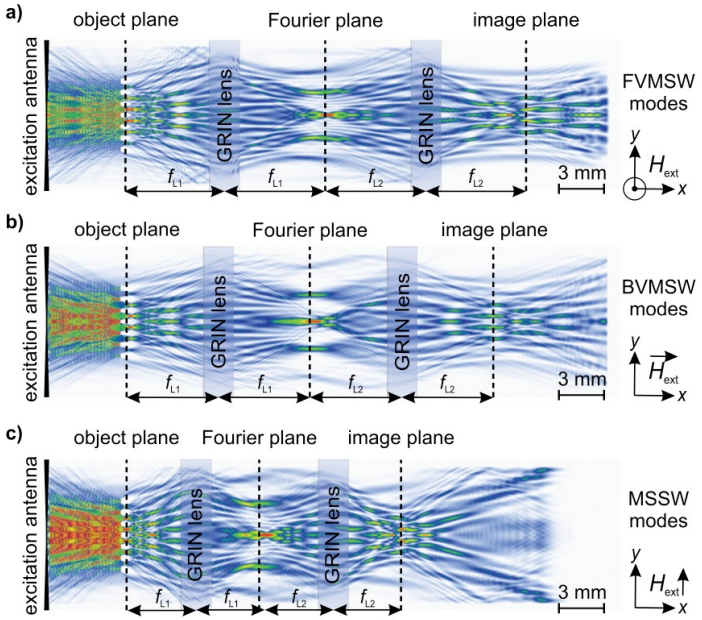
- **Using FIB to to tune magnetic properties**
- **Optically-inspired devices using FIB**

Devices based on spin-wave-optics

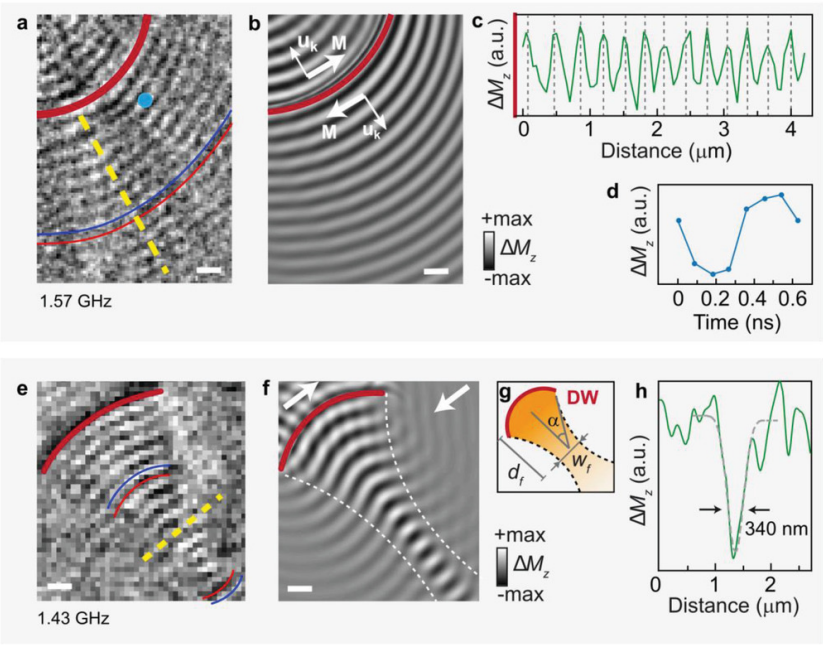
- *Complex, wave-based computing devices became an experimental reality over the last few years*
- *Most of the ideas are inspired by optics*



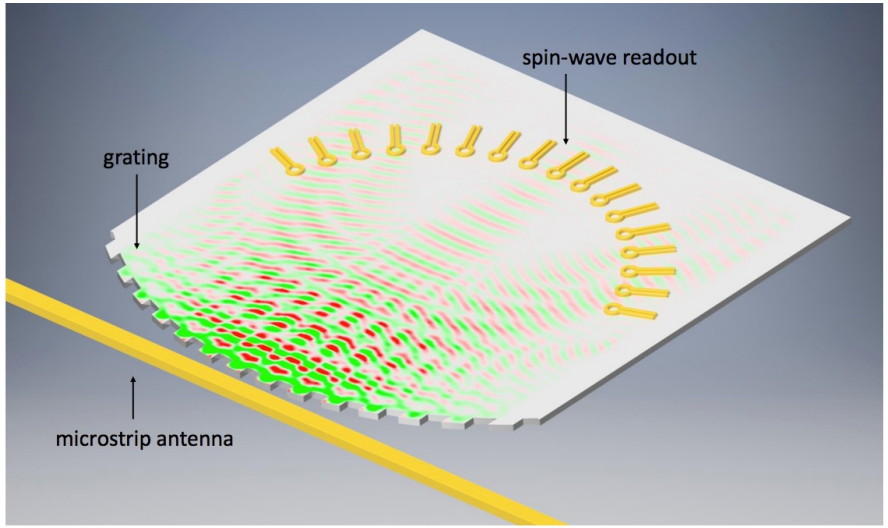
Q. Wang, et. al. Nat. Electron (2020), also Mahmoud, Abdulqader, et al. "Introduction to spin wave computing." Journal of Applied Physics 128, no. 16 (2020): 161101.



Vogel, M., P. Pirro, B. Hillebrands, and G. Von Freymann. "Optical elements for anisotropic spin-wave propagation." Applied Physics Letters 116, no. 26 (2020): 262404.



Albisetti, et al. "Optically inspired nanomagnonics with nonreciprocal spin waves in synthetic antiferromagnets." *Advanced Materials* 32, no. 9 (2020): 1906439.



Papp, Ádám, Wolfgang Porod, Árpád I. Csurgay, and György Csaba *Scientific Reports* 7 (2017).

Inverse design in optics

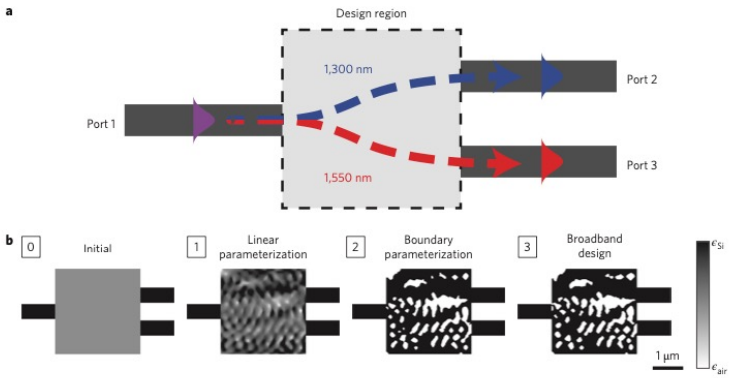


Figure 1 | Overview of the inverse design process. a, The device functionality is defined for the inverse design algorithm by specifying the surrounding structure, the design region and the coupling between a set of input and output modes. For the compact wavelength demultiplexer demonstrated in this work, the structure consists of one input waveguide, two output waveguides and a 2.8 × 2.8 μm^2 design region. The 1,300 nm band light is coupled into the fundamental TE mode of port 2, and 1,550 nm band light is coupled into the fundamental TE mode of port 3. All three waveguides are identical, with a width of 500 nm. **b**, Intermediate structures generated by the inverse design process. In the first stage the structure is optimized while allowing the permittivity ϵ to vary continuously (linear parameterization). In the next stage we convert to a boundary parameterization and optimize the structure for operation at only 1,300 nm and 1,550 nm. In the final stage, broadband optimization is performed to generate a robust device.

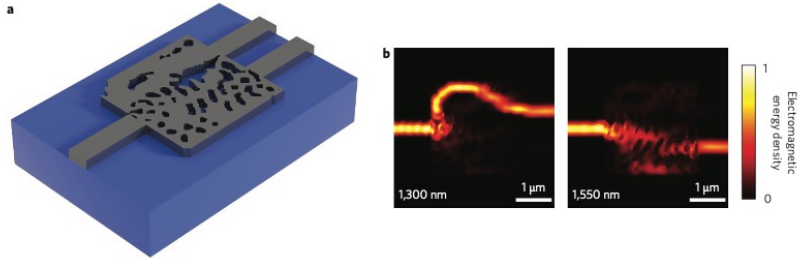
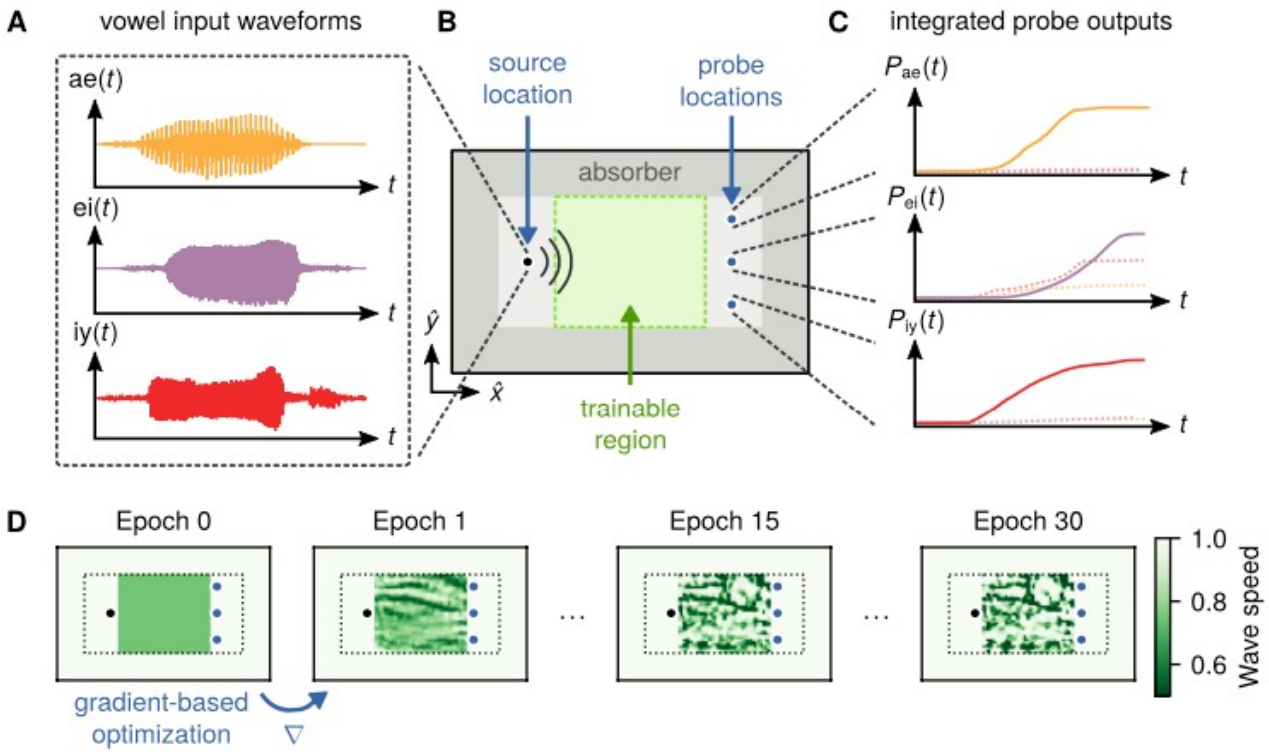


Figure 2 | The compact wavelength demultiplexer designed by the inverse design algorithm. a, A three-dimensional rendering of the structure. Silicon is

Piggott, Alexander Y., Jesse Lu, Konstantinos G. Lagoudakis, Jan Petykiewicz, Thomas M. Babinec, and Jelena Vučković. "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer." *Nature Photonics* 9, no. 6 (2015): 374-377.

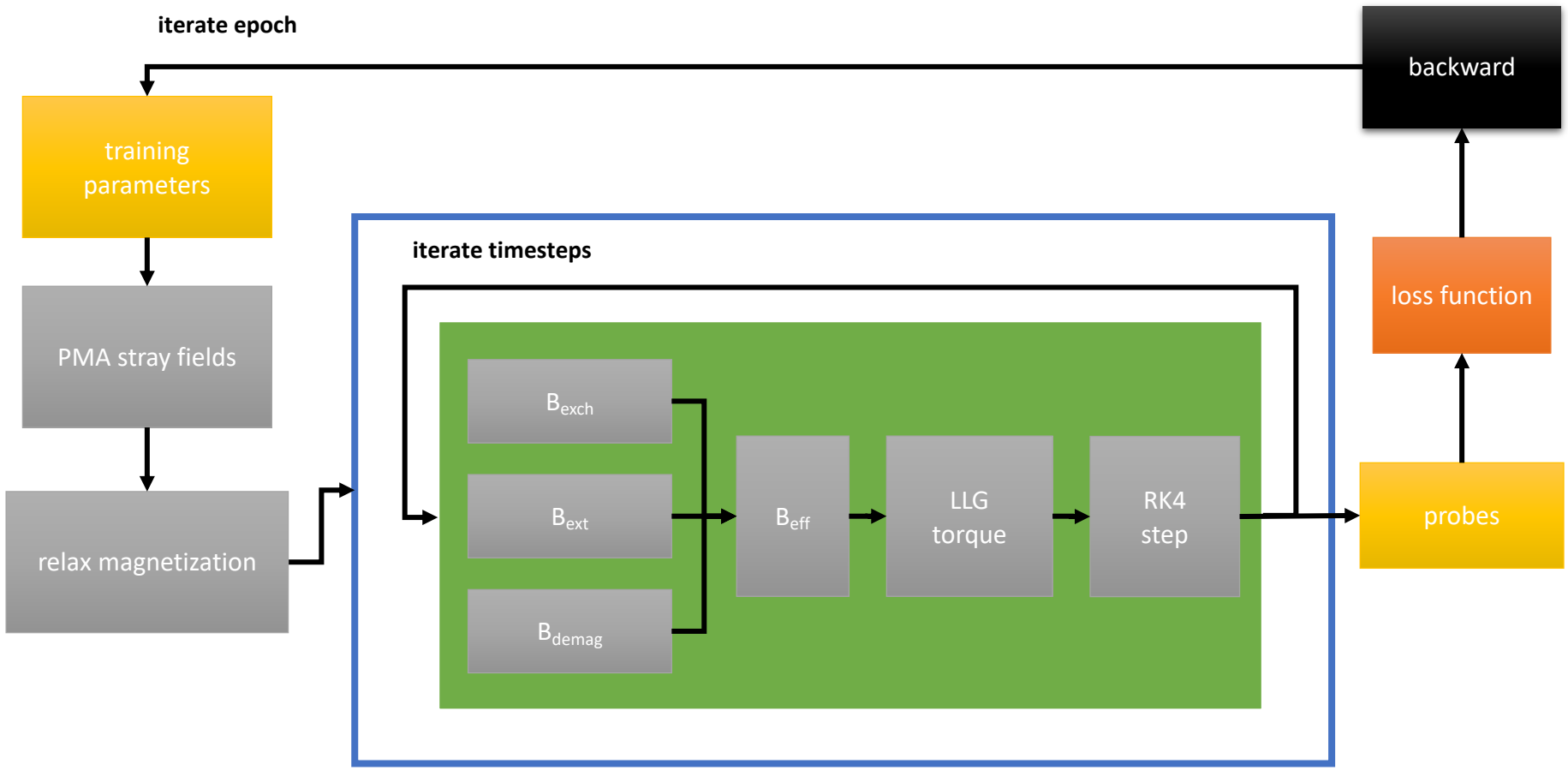
Inverse design for metamaterials / nanophotonic devices

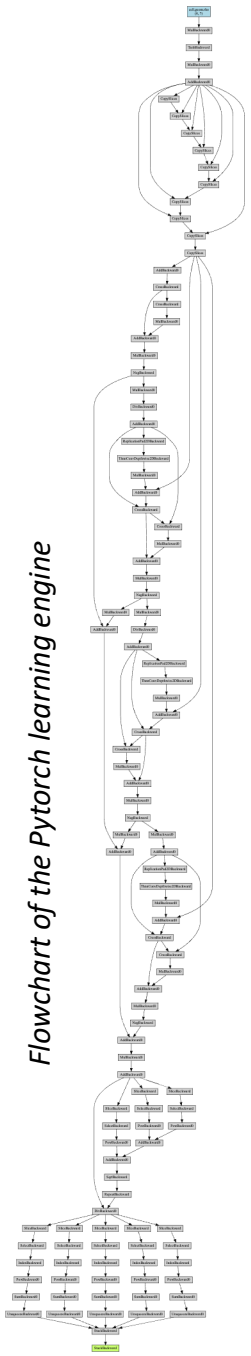


Hughes, Tyler W., Ian AD Williamson, Momchil Minkov, and Shanhui Fan. "Wave physics as an analog recurrent neural network." *Science advances* 5, no. 12 (2019): eaay6946.

Inverse design for neuromorphic computing functions

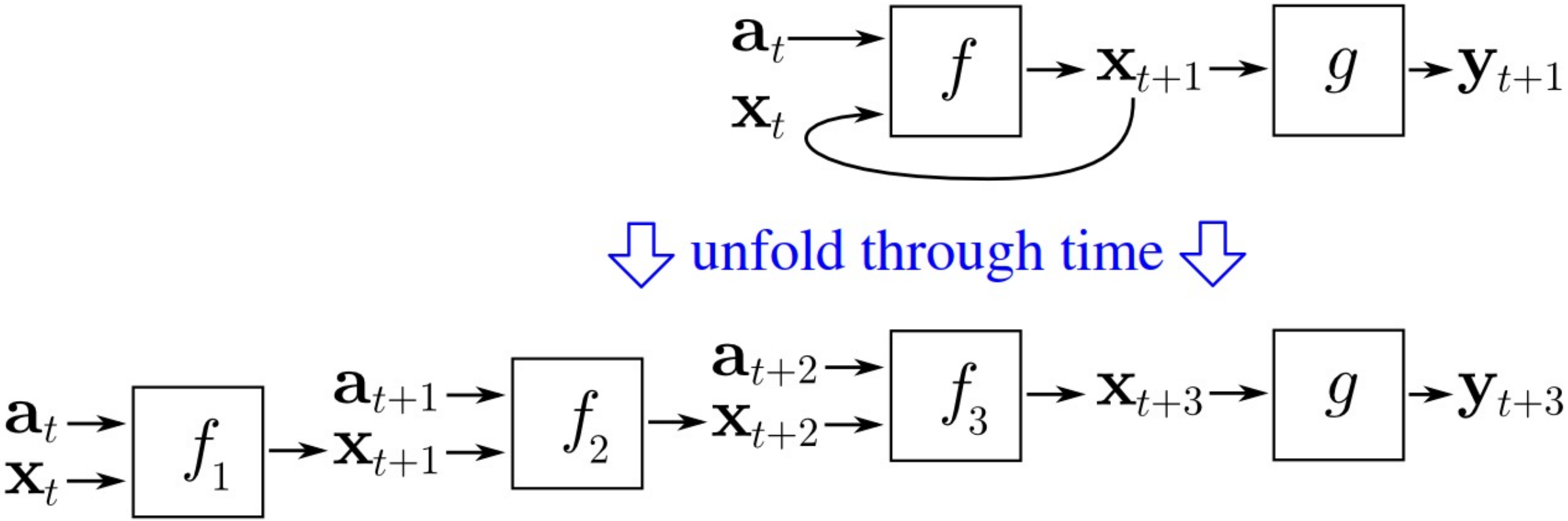
Spintorch: machine learning – neural network style





- *The learning engine performs inverse design without knowledge of the physics (black box) – so it should work for active medium as well.*
- *Algorithm is well-established for machine learning the parameters of hidden weights in neural networks*

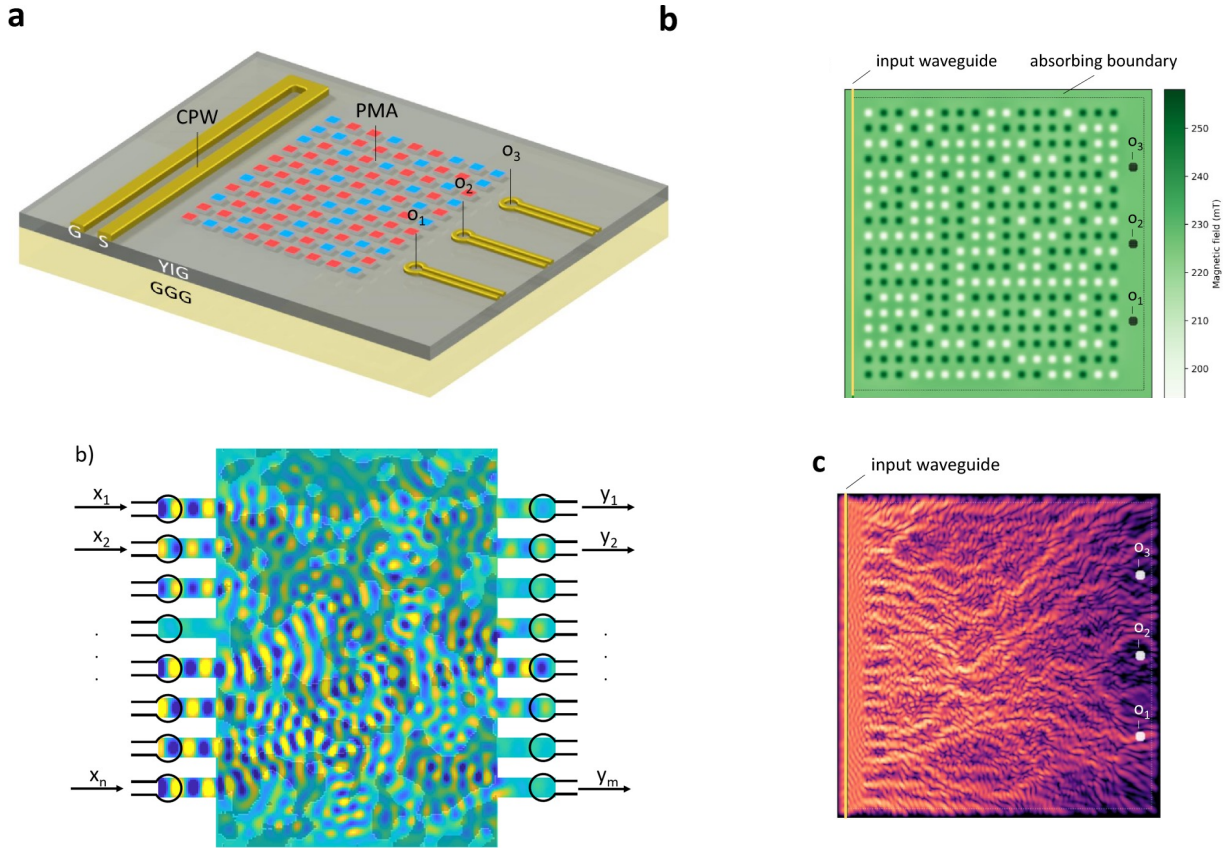
Back Propagation Through Time (BPTT) – a bit of background on machine learning



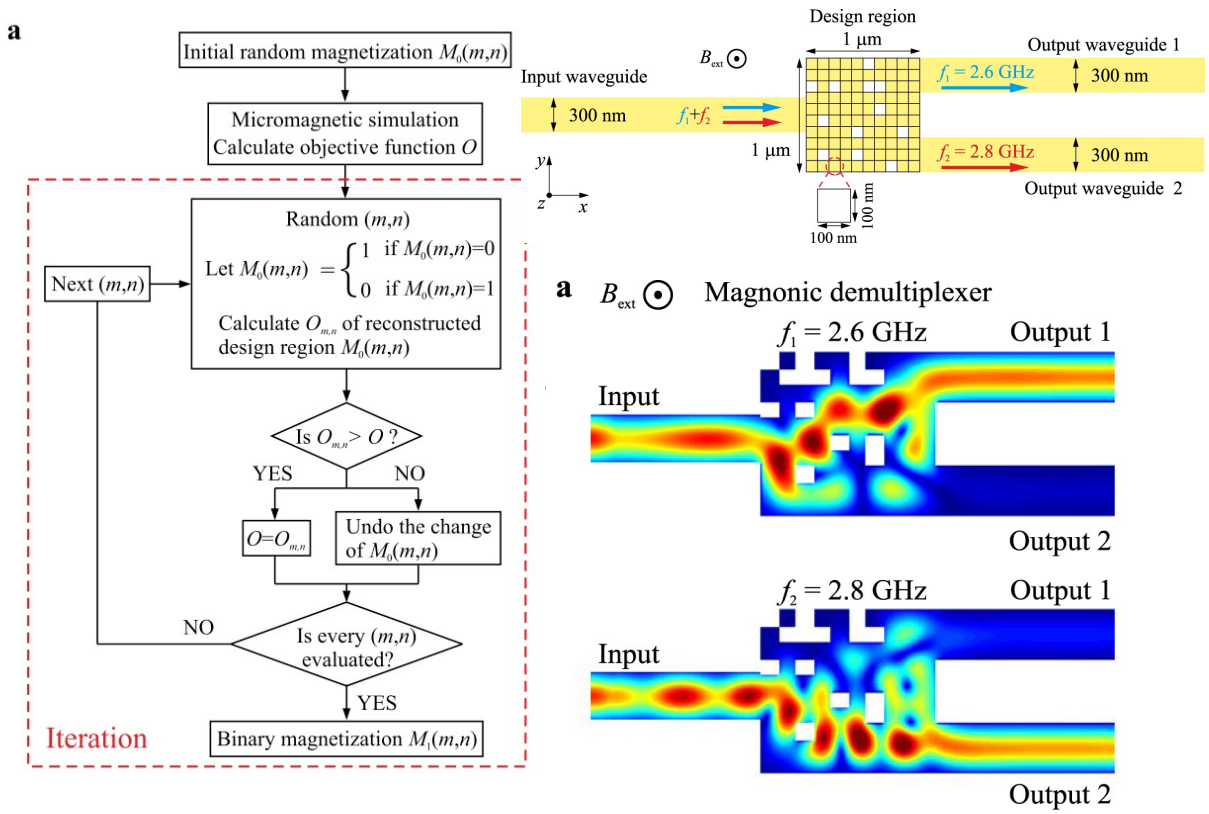
Source: wikipedia

If you write your code in Pytorch the unfolding is done automatically

Machine learning to design magnetic nanostructures – two approaches

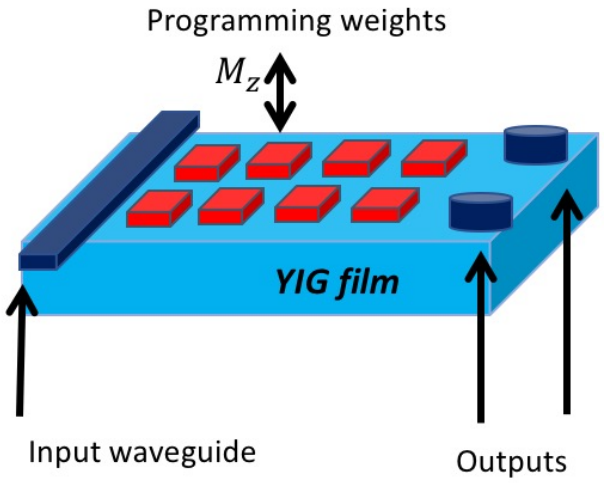


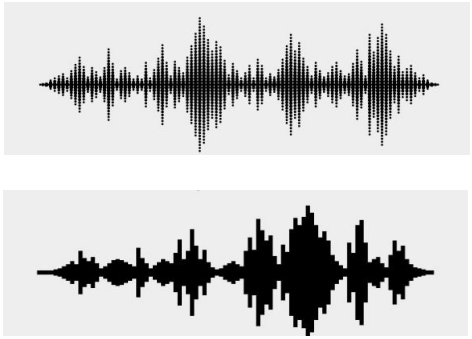
Papp, Á., Porod, W. & Csaba, G. Nanoscale neural network using non-linear spin-wave interference. *Nat Commun* **12,** 6422 (2021). https://doi.org/10.1038/s41467-021-26711-z

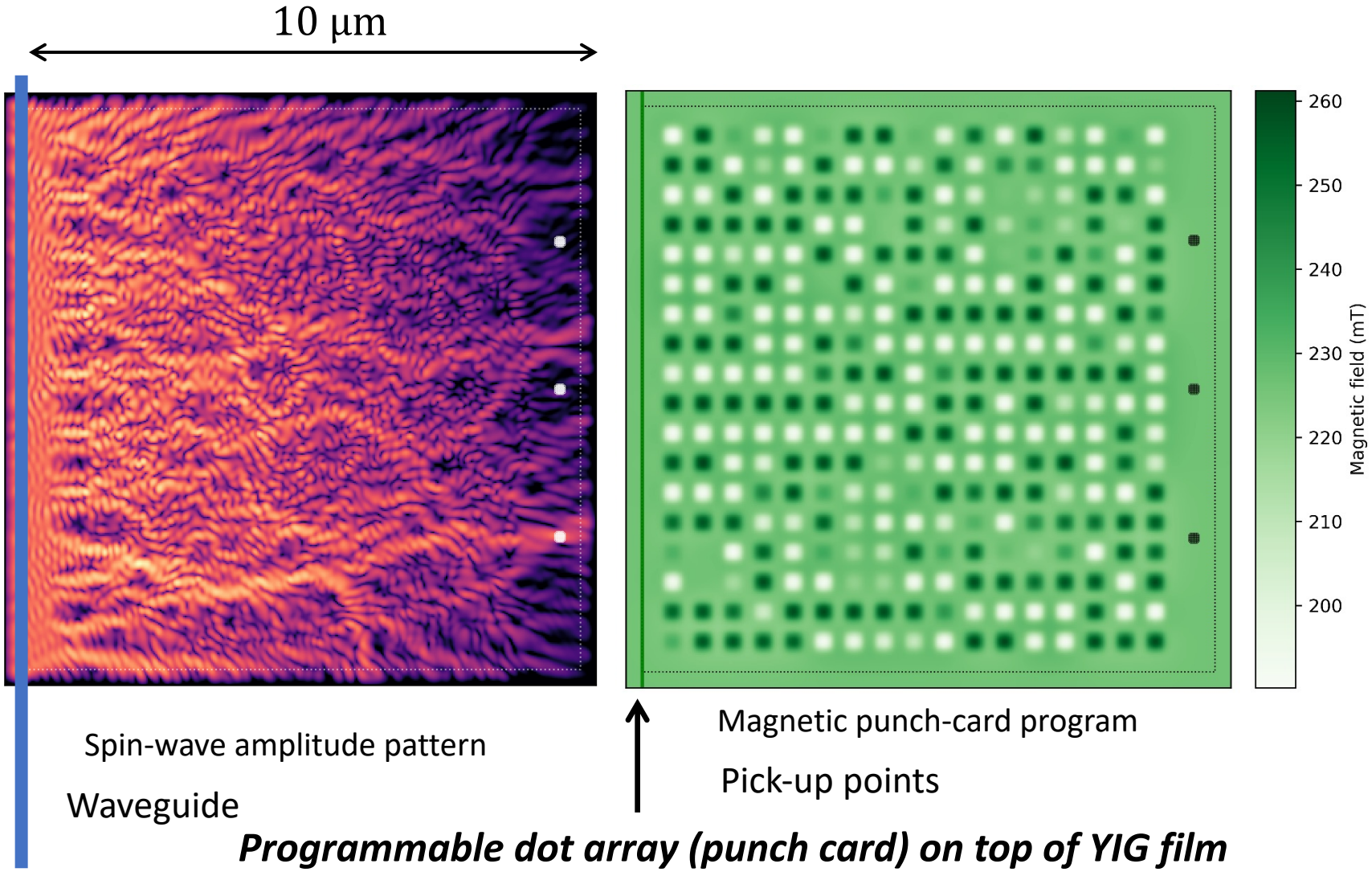


Wang, Q., Chumak, A.V. & Pirro, P. Inverse-design magnonic devices. *Nat Commun* **12,** 2636 (2021). https://doi.org/10.1038/s41467-021-22897-4

Physical structure: programming magnets on top of YIG film





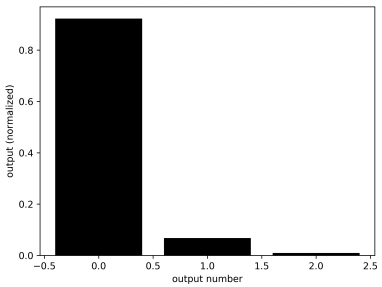


Programmable dot array (punch card) on top of YIG film

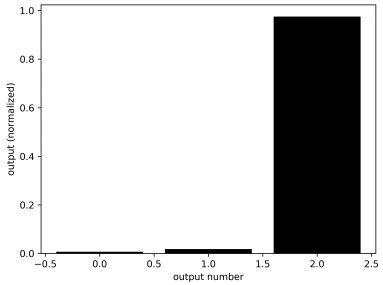
Single waveguide input for information coded in waveform

This is not the easiest-to-realize way of doing it – stay tuned till the end of the talk

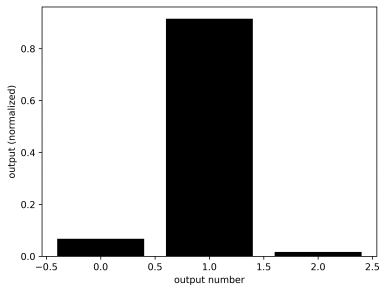
One application: vowel recognition



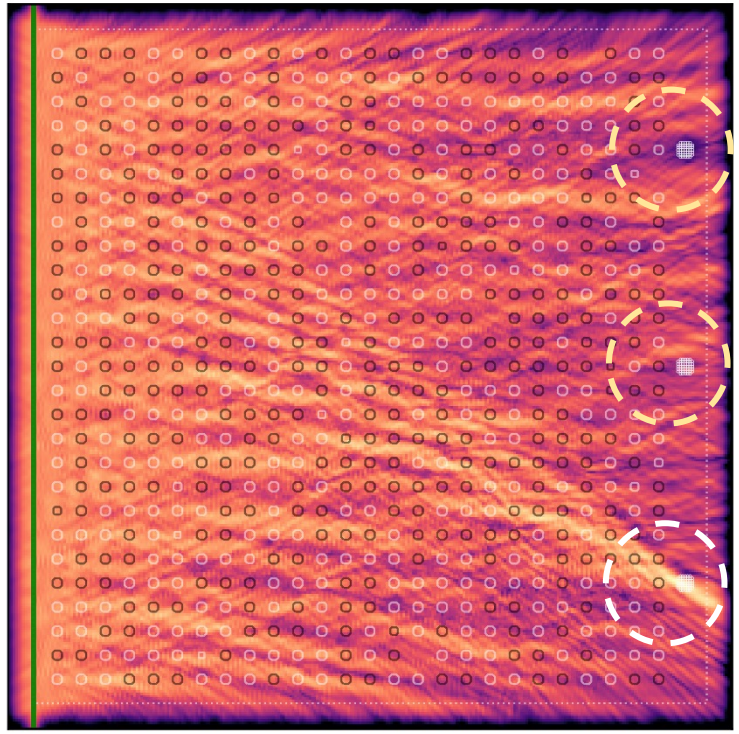
ei

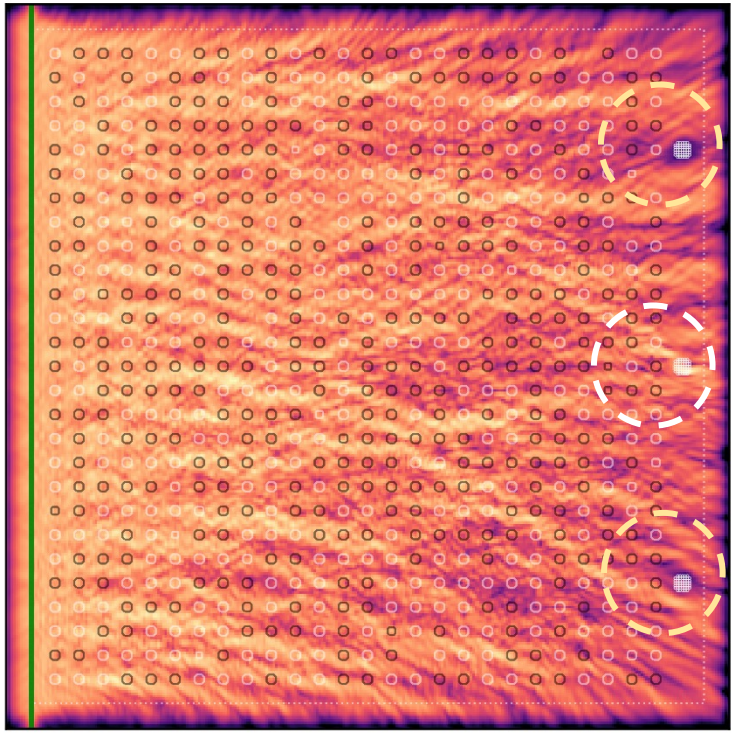


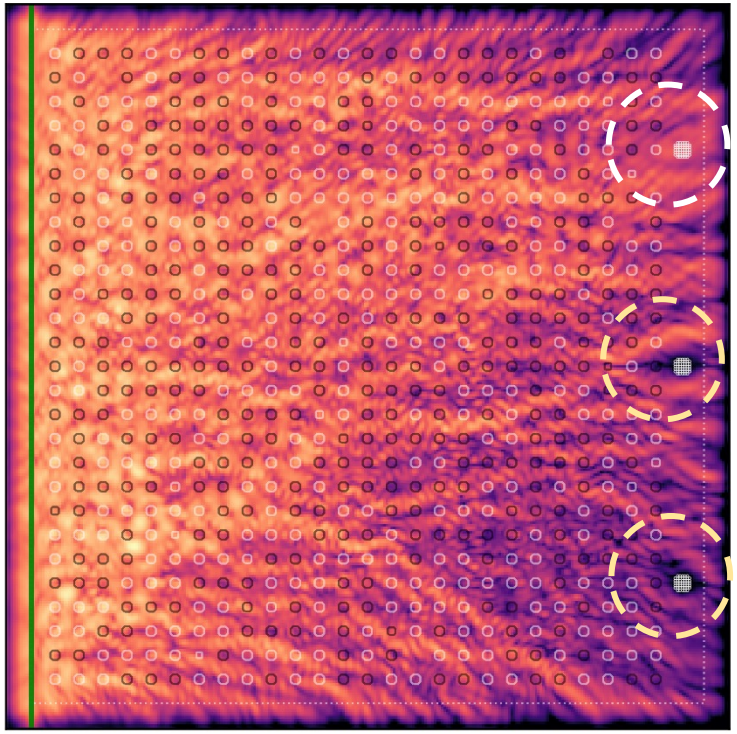
iy



oa



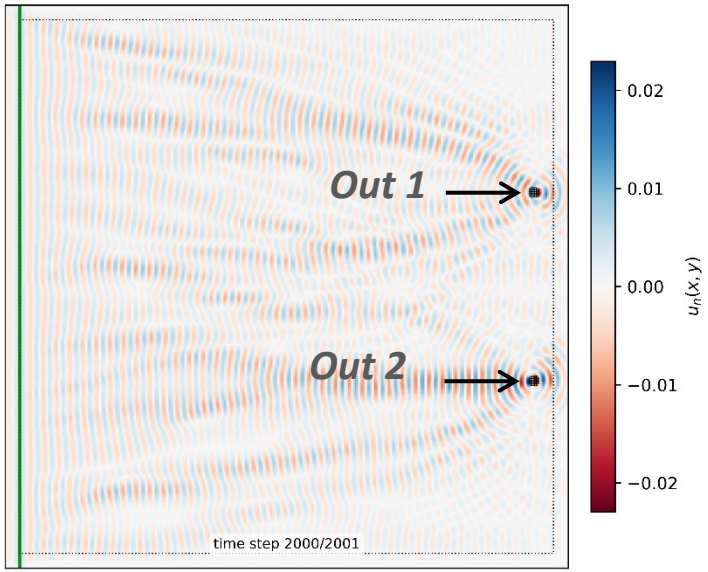




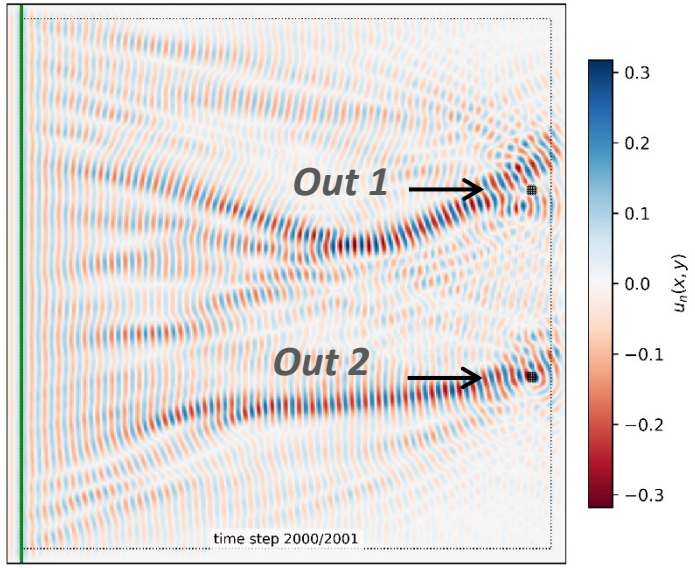
Vowels waveforms are scaled up to microwave frequencies and applied on the waveguide
Different vowels will lead to intensity maxima at different pick-up points

Distinguishing linear training from nonlinear training

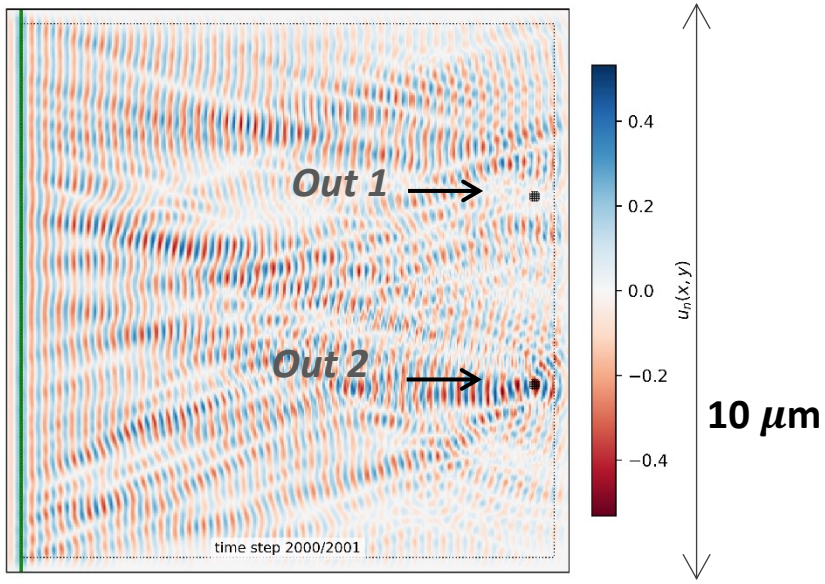
1 mT excitation



20 mT excitation



50 mT excitation



Linear regime → Moderate nonlinearity → Strong nonlinearity

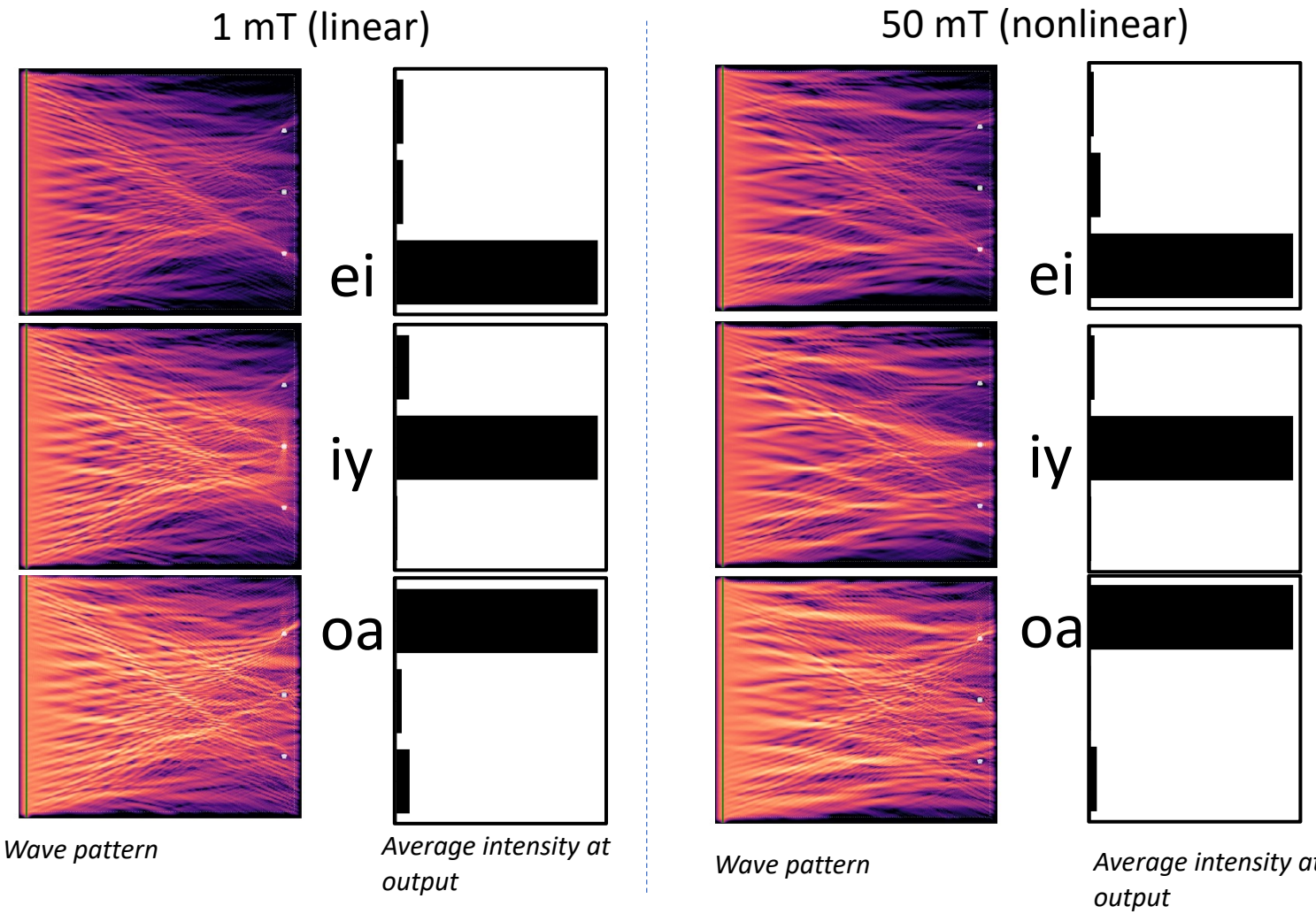
Interference patterns if a 3 GHz and a 4 GHz input is simultaneously applied and the following training function is defined:

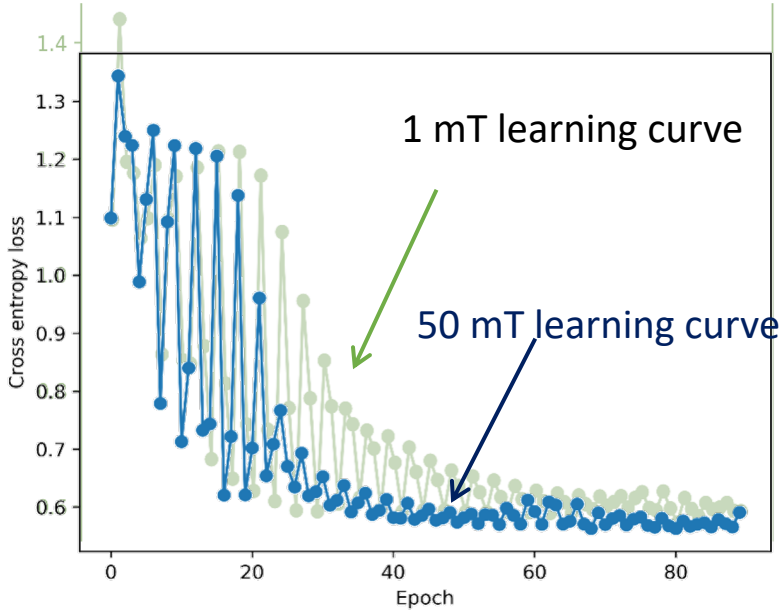
Training goal:

- Focus waves on **Out 1 at 3 GHz** input frequency
- Focus waves on **Out 1 at 4 GHz** input frequency
- Focus waves on **Out 2 if and only if 3 and 4 GHz** are simultaneously present

The network learns poorly at small amplitudes but gets much better at high amplitudes
The task cannot be learned if the superposition principle holds for the waves.

Recognizing vowels that were part of the training set



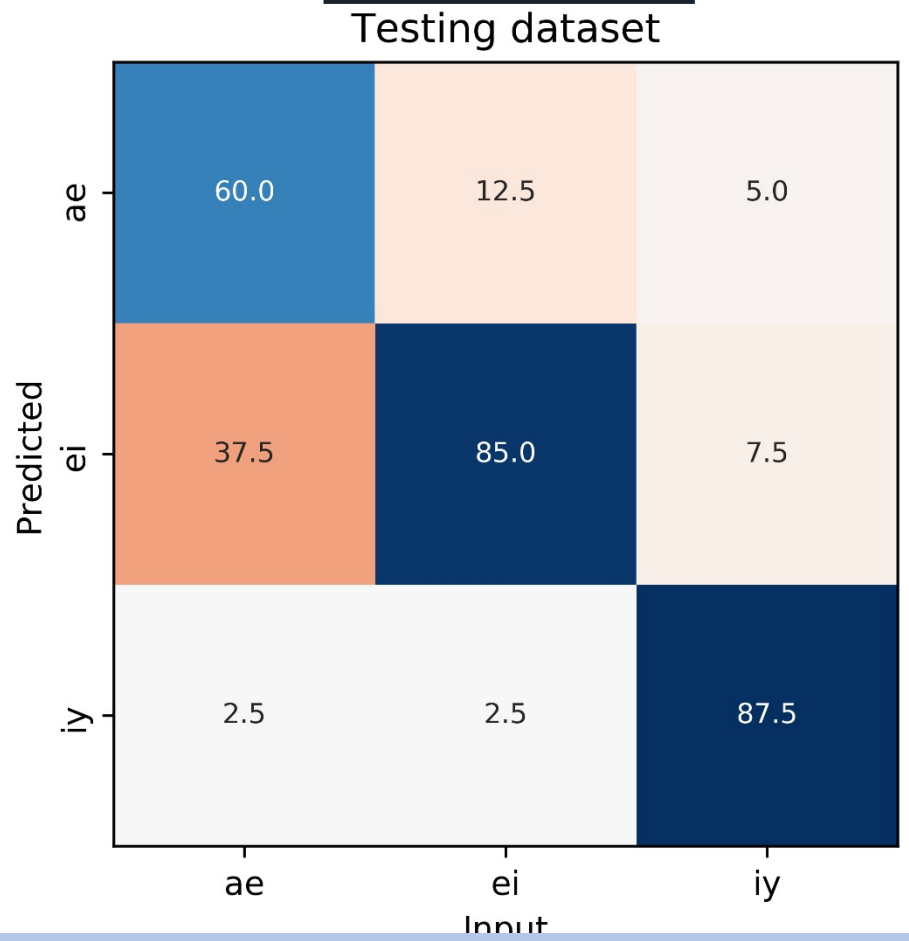


Network performance as a function of training steps (epochs)

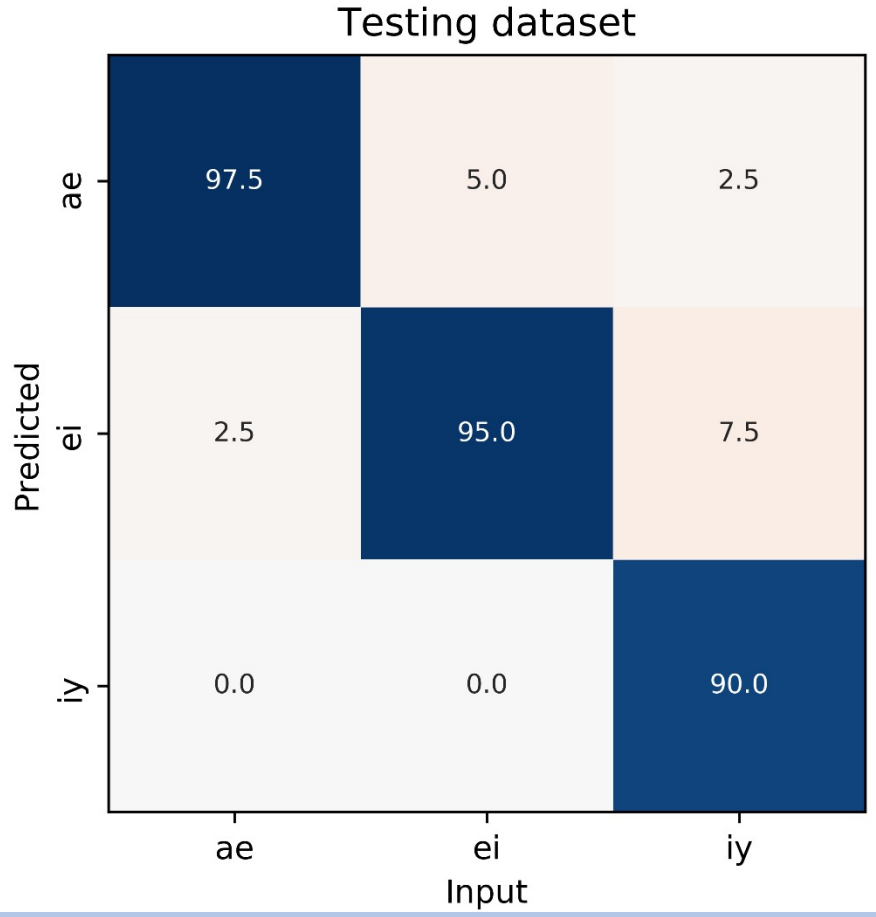
- Recognizing the dominant spectral component does not require nonlinearity – so for this task there is only a slight improvement with nonlinearity and learning becomes faster.
- Recognizing the elements of the training set does not yet demonstrate the superiority of nonlinear behavior.

Network performance for generalization (recognizing unseen samples)

1 mT - linear

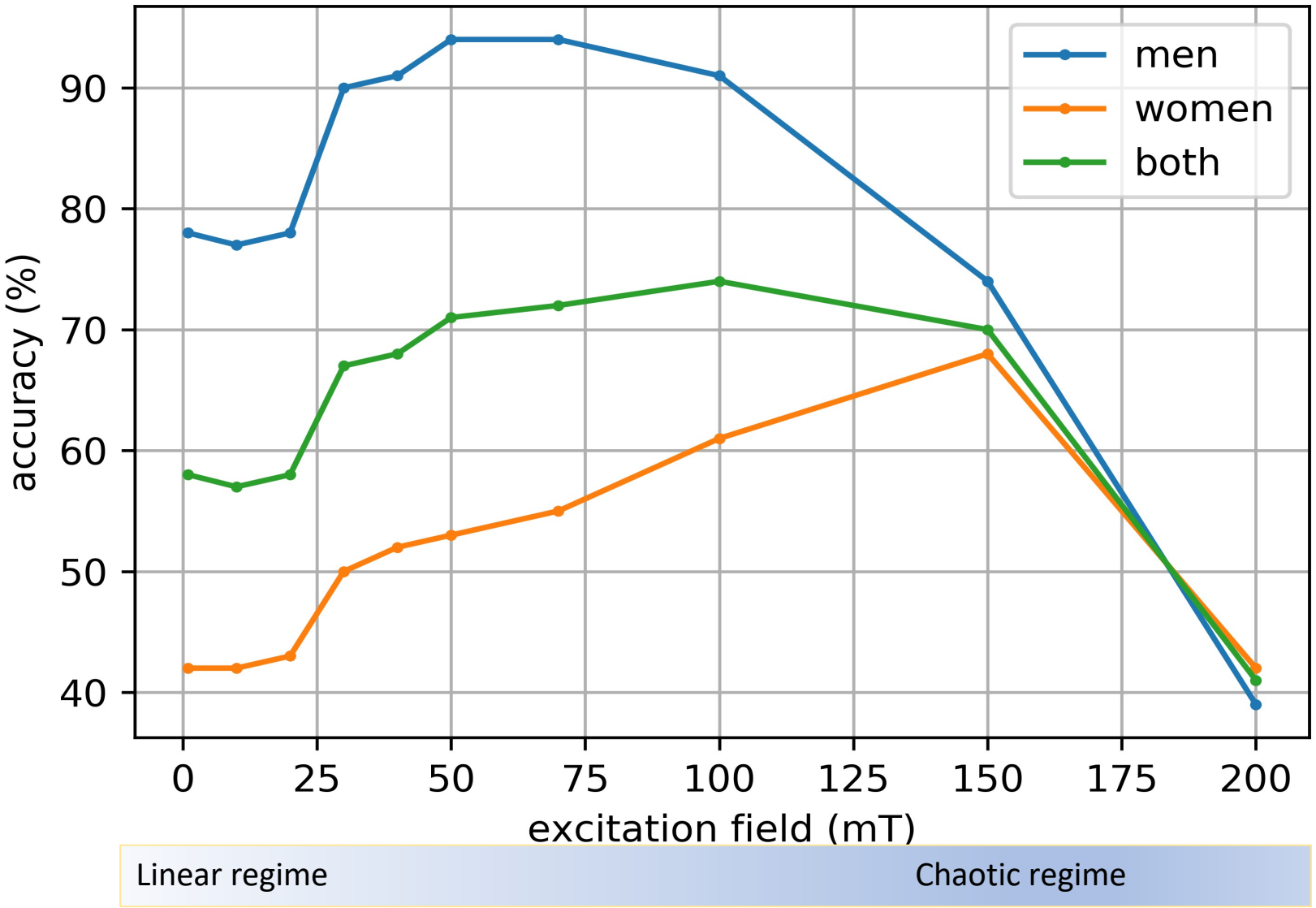


50 mT – highly nonlinear



- The network is trained on a small set of vowels, and then a new set of vowels (that was not present in the training set) is used for testing. Confusion matrix should look diagonal for perfect recognition.
- **Nonlinear regime is significantly better in recognizing unseen samples – this really is a neural network that learns.**

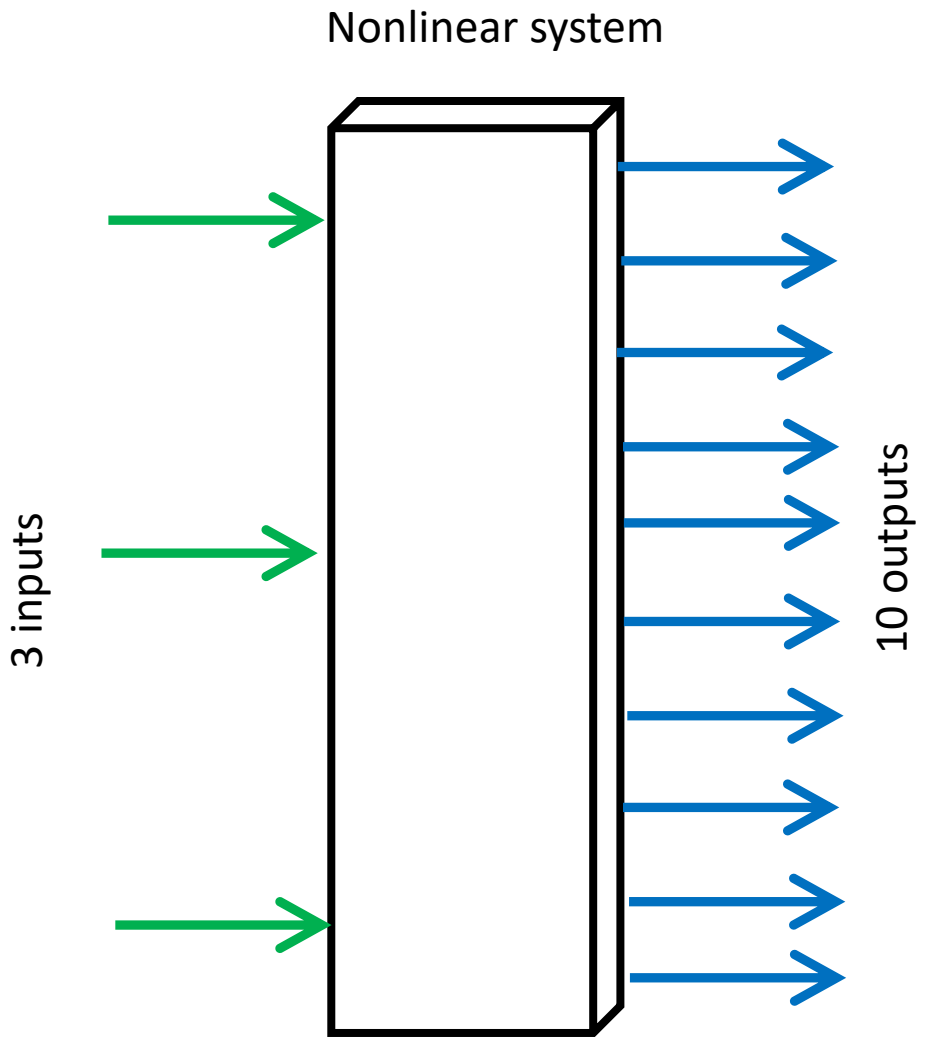
Excitation amplitude vs. network performance



We designed and trained networks at different excitation amplitudes. Clearly, there is an optimum amplitude, where the nonlinearity is sufficiently high but no chaotic behavior occurs

Design for complexity: kernel rank and generalization rank

- **Kernel rank**
  - „KR is a measure of the reservoir's ability to separate distinct input patterns."
  - Measures how much the reservoir can **increase the parameter space**.
- **Generalization rank**
  - „GR is a measure of the reservoir's capability to generalize given similar input streams."
  - Low GR means a robust ability to **map similar inputs to similar reservoir states** (i.e. exclude noise, avoid chaos).
  - To some extent it also addresses fragility / sensitivity of the system

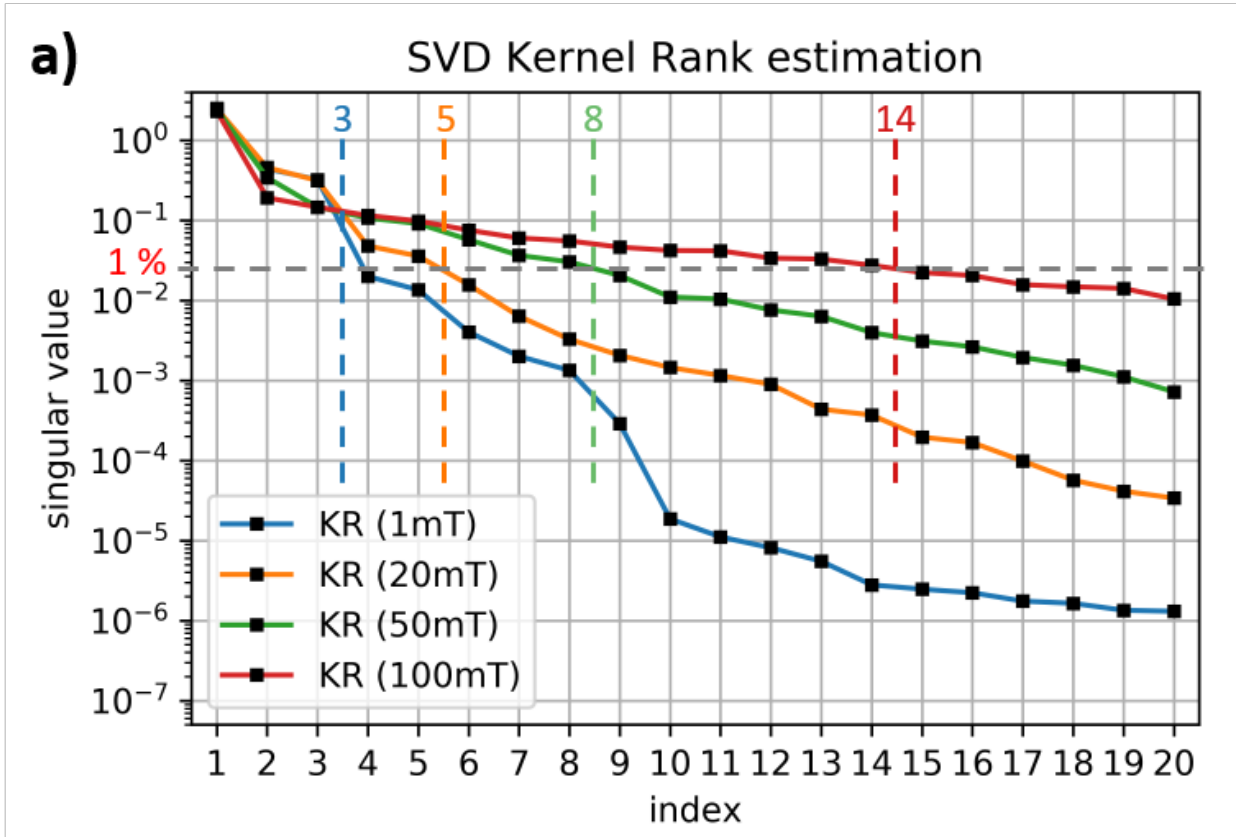


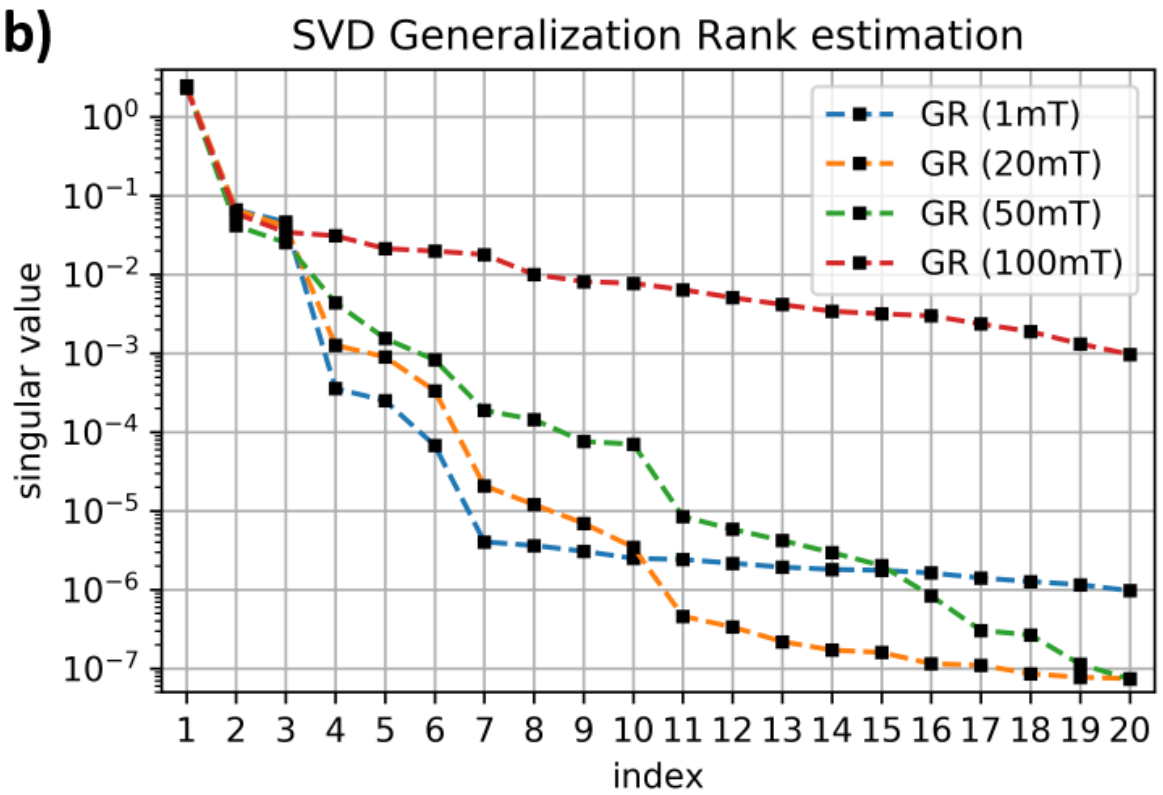
A 'good' nonlinear system can make 10 linearly independent outputs from 3 linearly independent inputs (high kernel rank)

It should do it in a robust way, and map similar (small Euclidean distance) inputs to similar (small Euclidean distance) outputs → generalization rank

Dale M, Miller JF, Stepney S, Trefzer MA. 2019 A substrate-independent framework to characterize reservoir computers. *Proc. R. Soc. A* **475**: 20180723. http://dx.doi.org/10.1098/rspa.2018.0723

Reservoir metrics for a trained scatterer





- There is a range of excitations where sufficiently strong and sufficiently robust nonlinearity is present. Paper about out findings is submitted to APL.
- *These figures also show that there are approx. 8-10 useful outputs for the scatterer.*

Spin-wave reservoir

Characterization of nonlinear spin-wave interference by reservoir-computing metrics

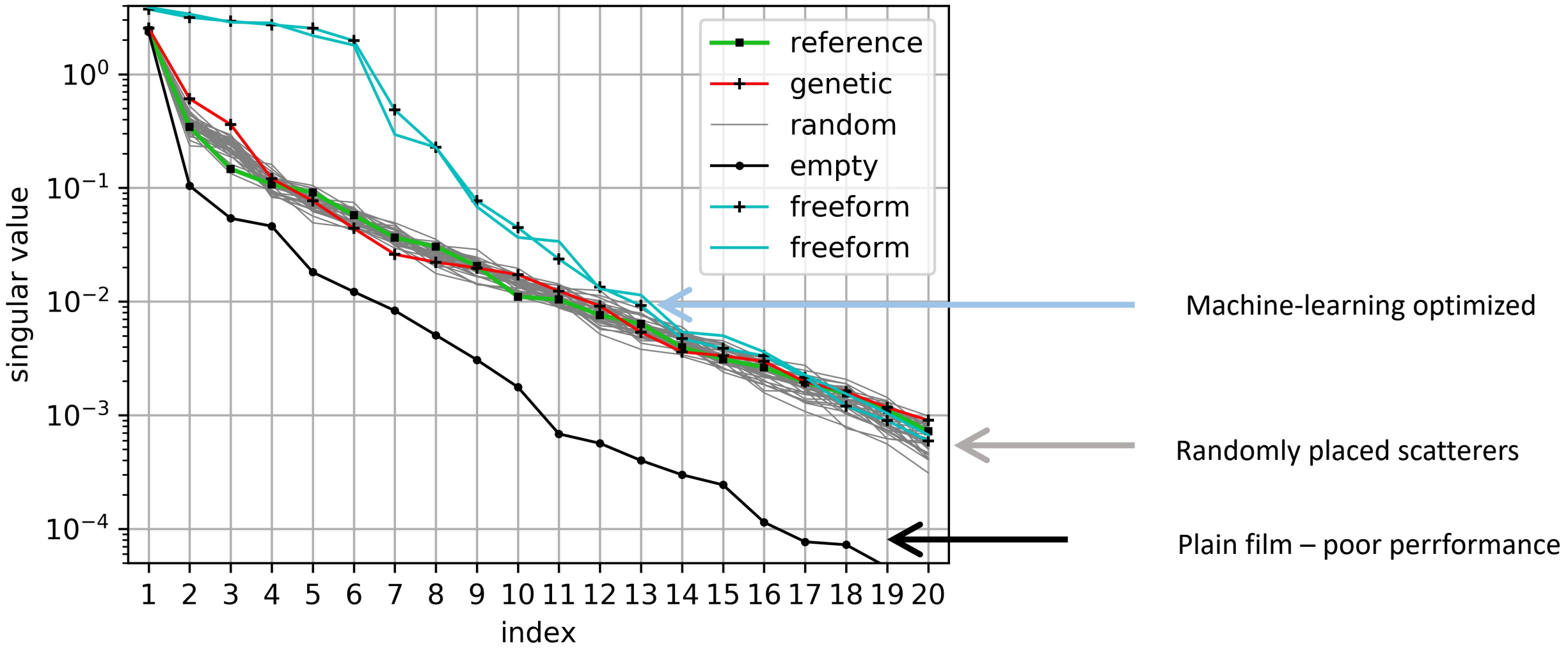
A. Papp,[1, a)] G. Csaba,[1, b)] and W. Porod[2, c)]

[1)] *Faculty for Information Technology and Bionics, Pazmany Peter Catholic University, Budapest*

[2)] *Department of Electrical Engineering, University of Notre Dame*

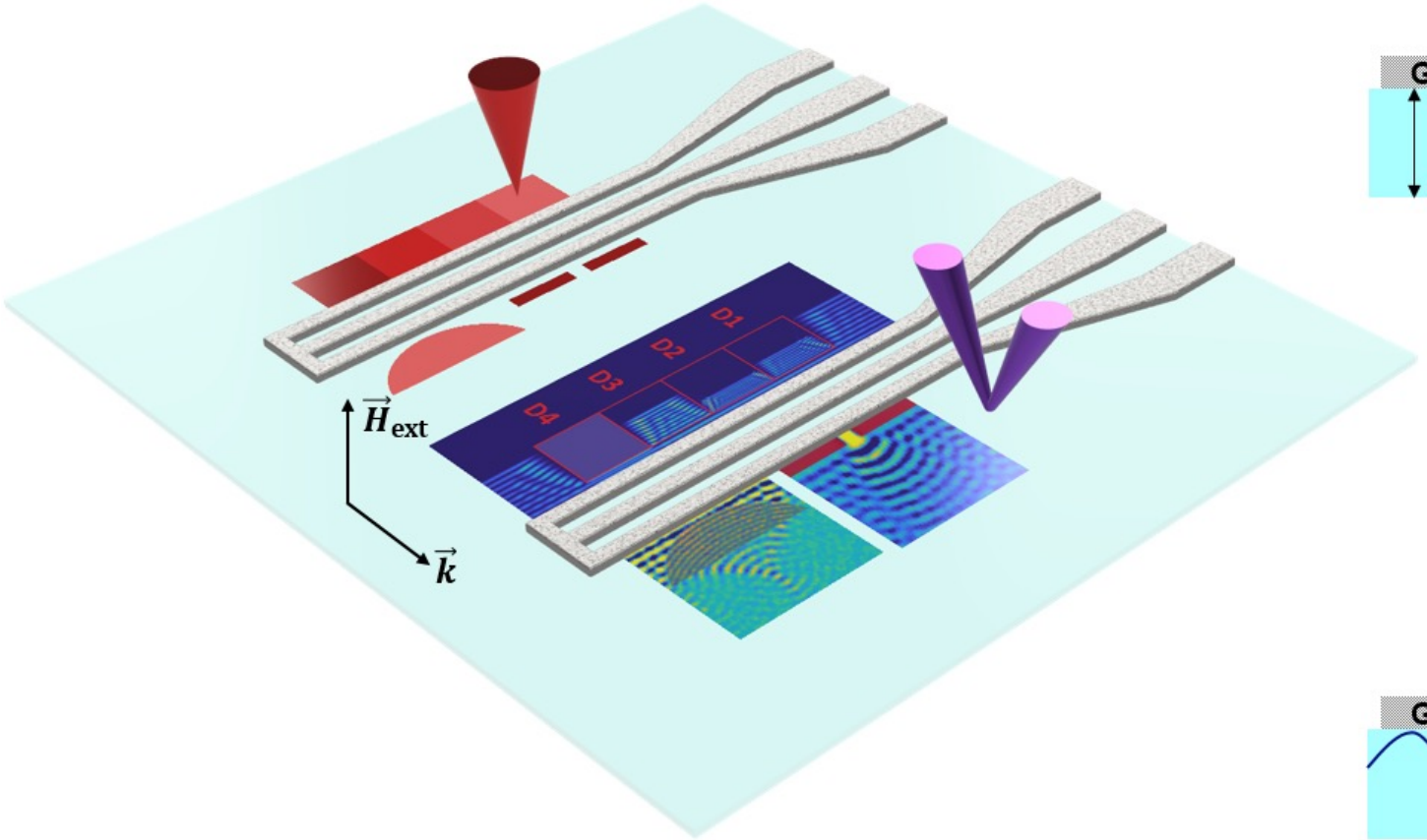
(Dated: 24 March 2021)

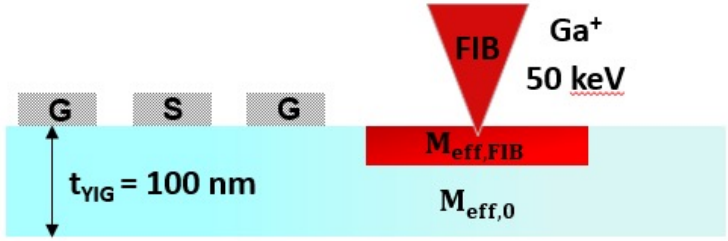
Training for good reservoir performance

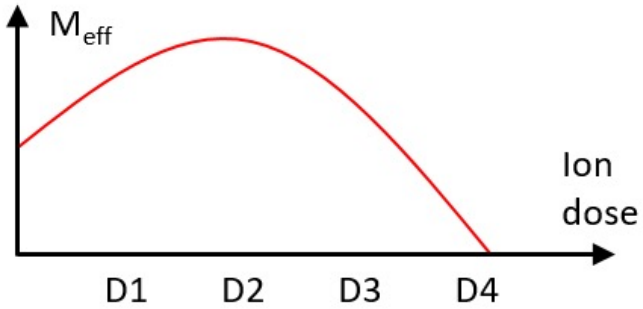


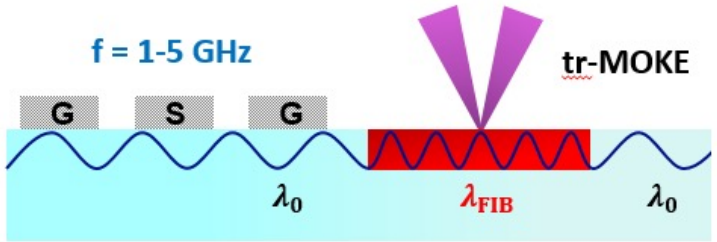
- *From three linearly independent inputs the trained film generates six strong and at least nine reasonably strong linearly independent outputs.*
- *This could also serves as a metric for the computational performance of the scatterer*

Experimental investigations









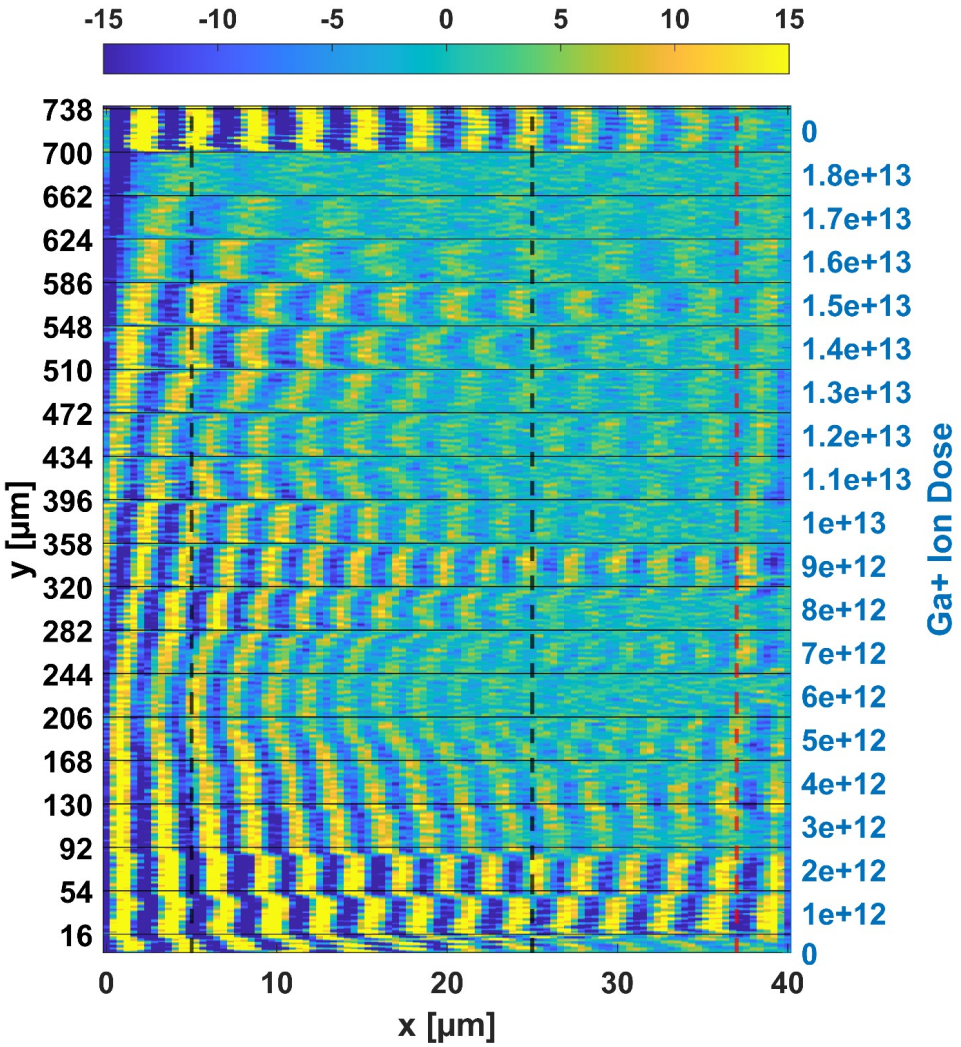
This is work from the group of Markus Becherer TUM, special thanks to Martina Kiechle

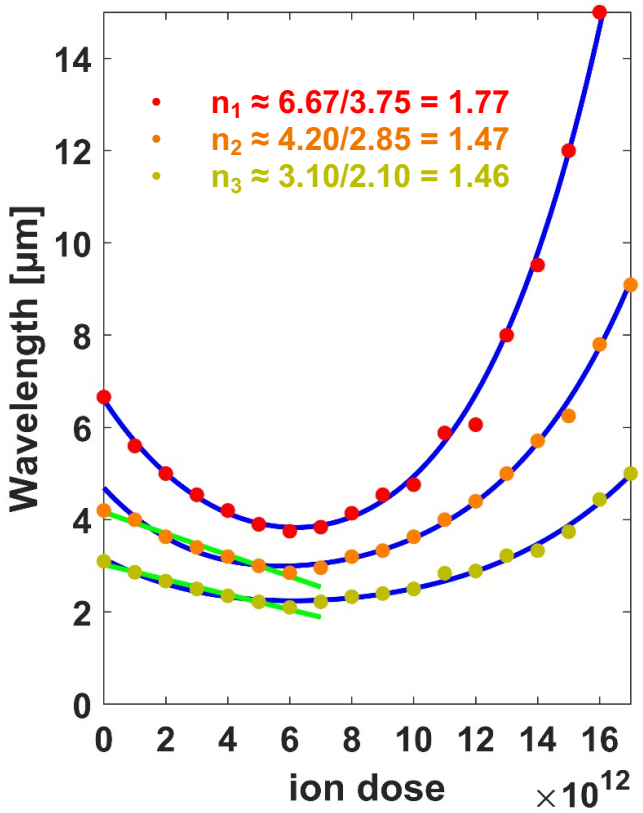
Spin Wave Optics in YIG by Ion Beam Irradiation

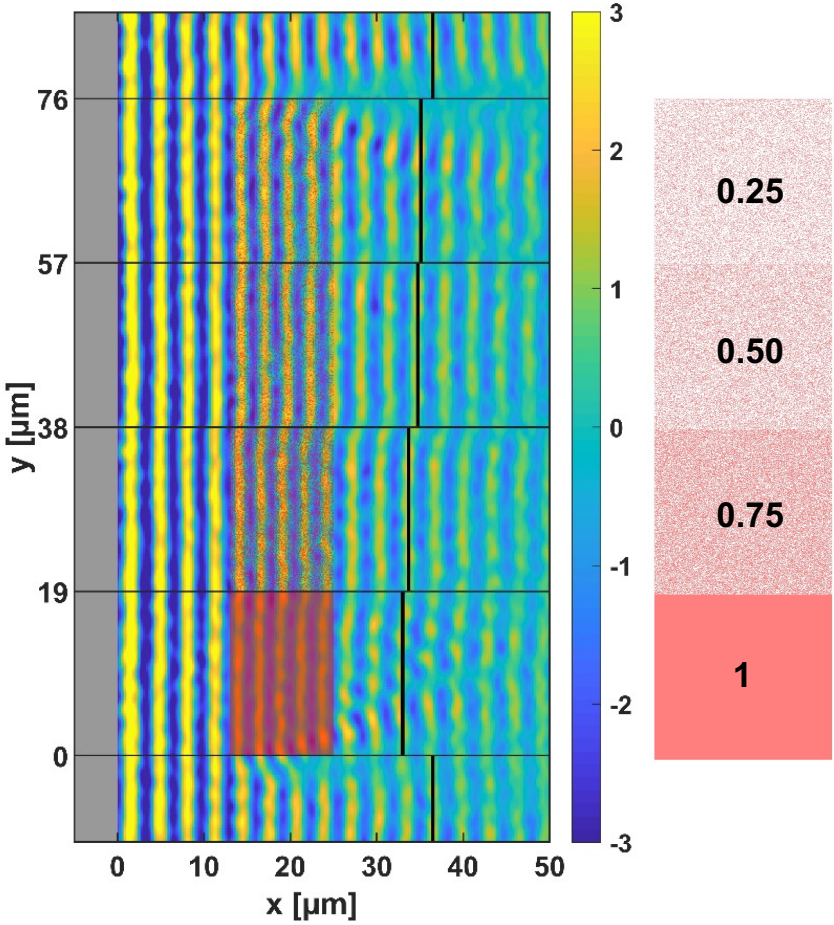
Martina Kiechle[1,*], Adam Papp[2], Simon Mendisch[1], Valentin Ahrens[1], Matthias Golibrzuch[1], Gyorgy Csaba[2], and Markus Becherer[1,+]

[1]Department of Electrical and Computer Engineering, Technical University of Munich, Germany
[2]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary
[*]martina.kiechle@tum.de

Changing YIG properties by FIB



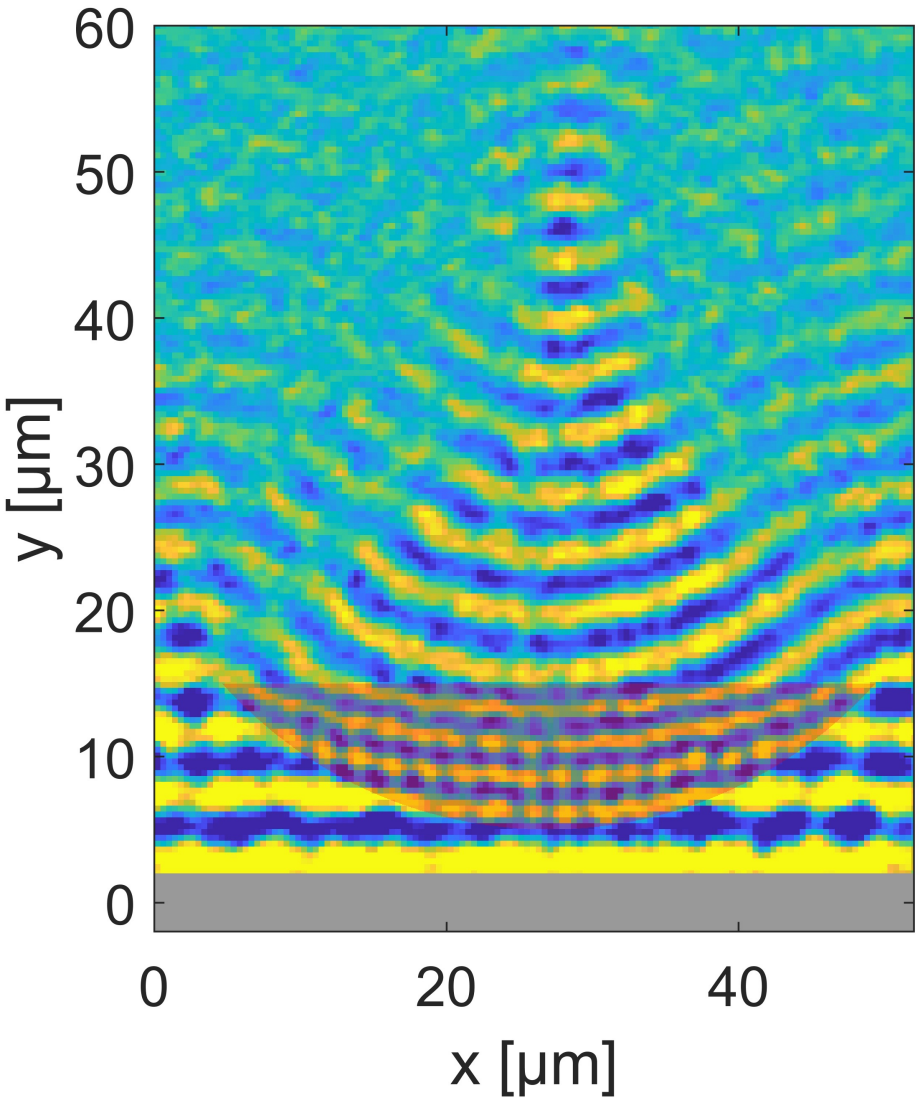


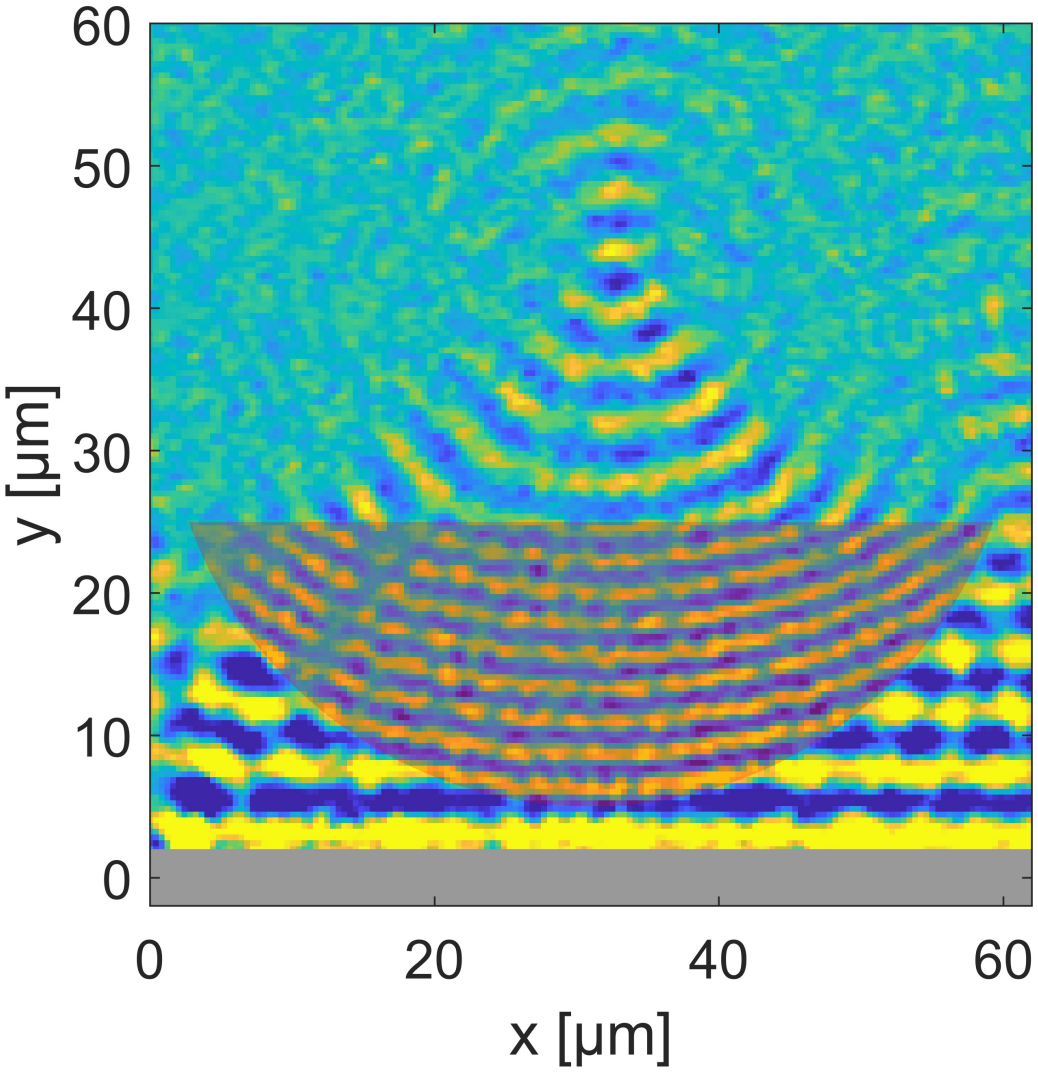


Changing the dose

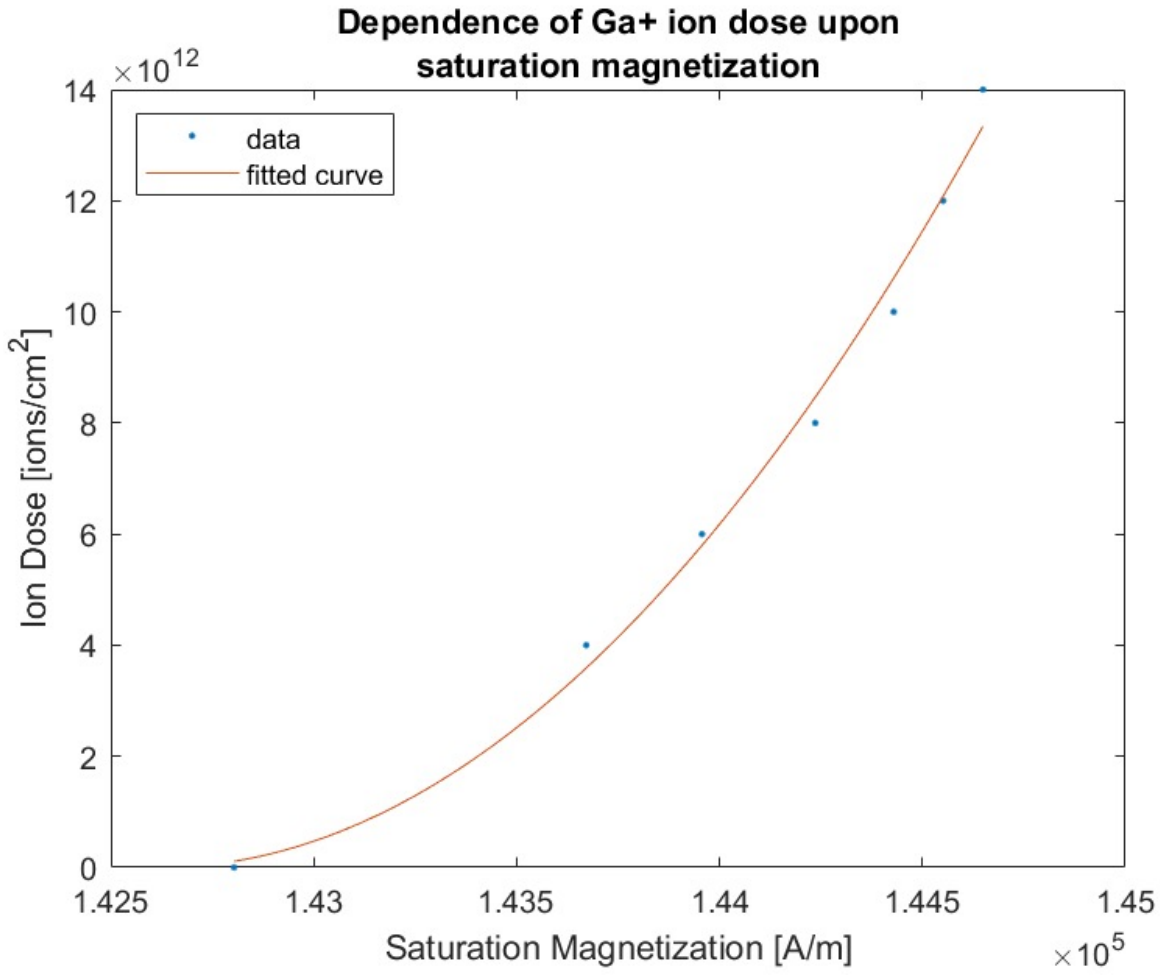
Changing the filling factor

Making lenses using FIB

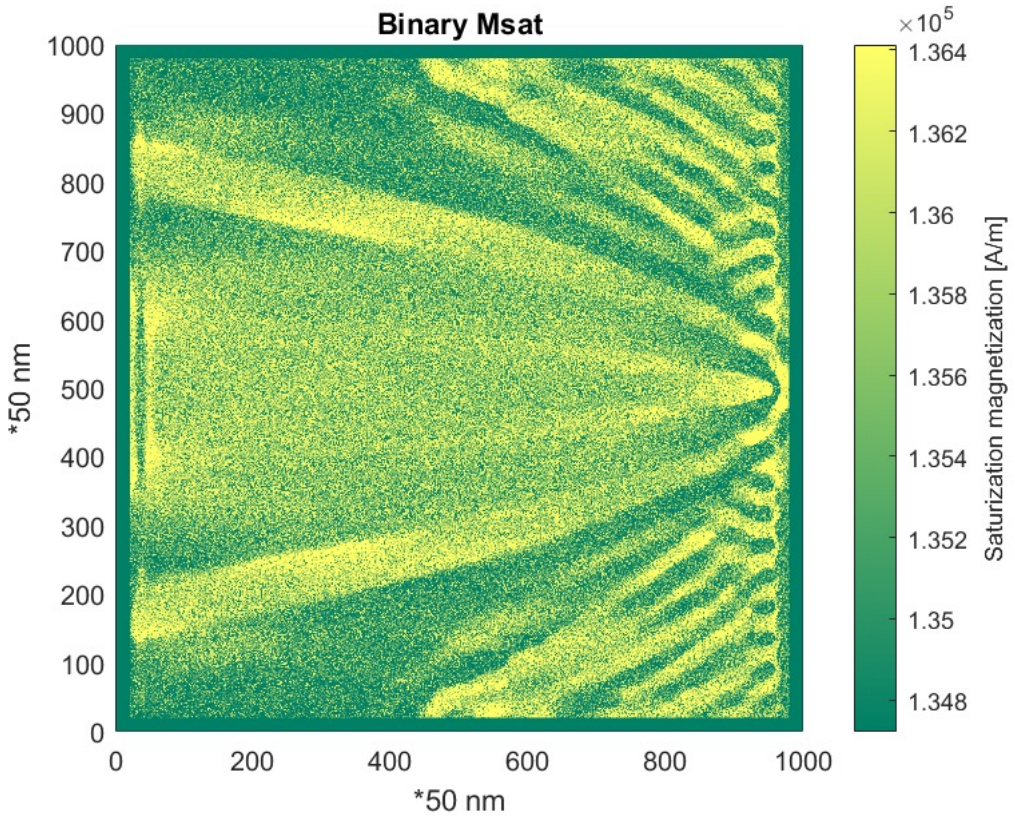


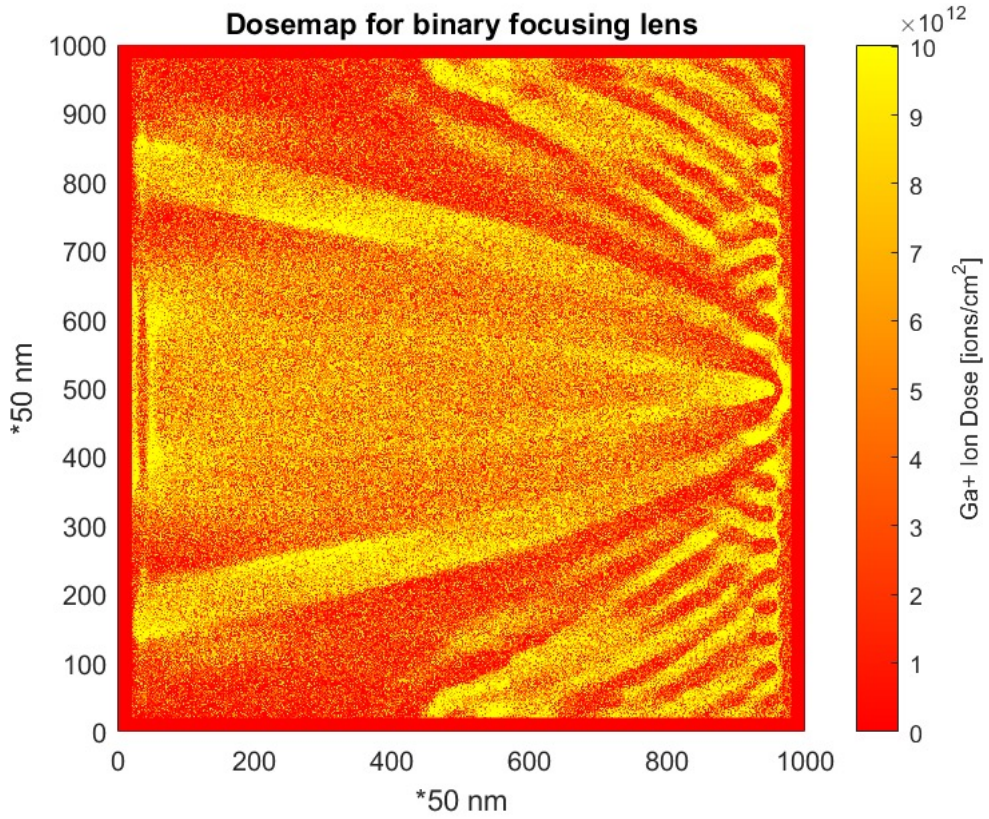


Mapping ion doses to M_s

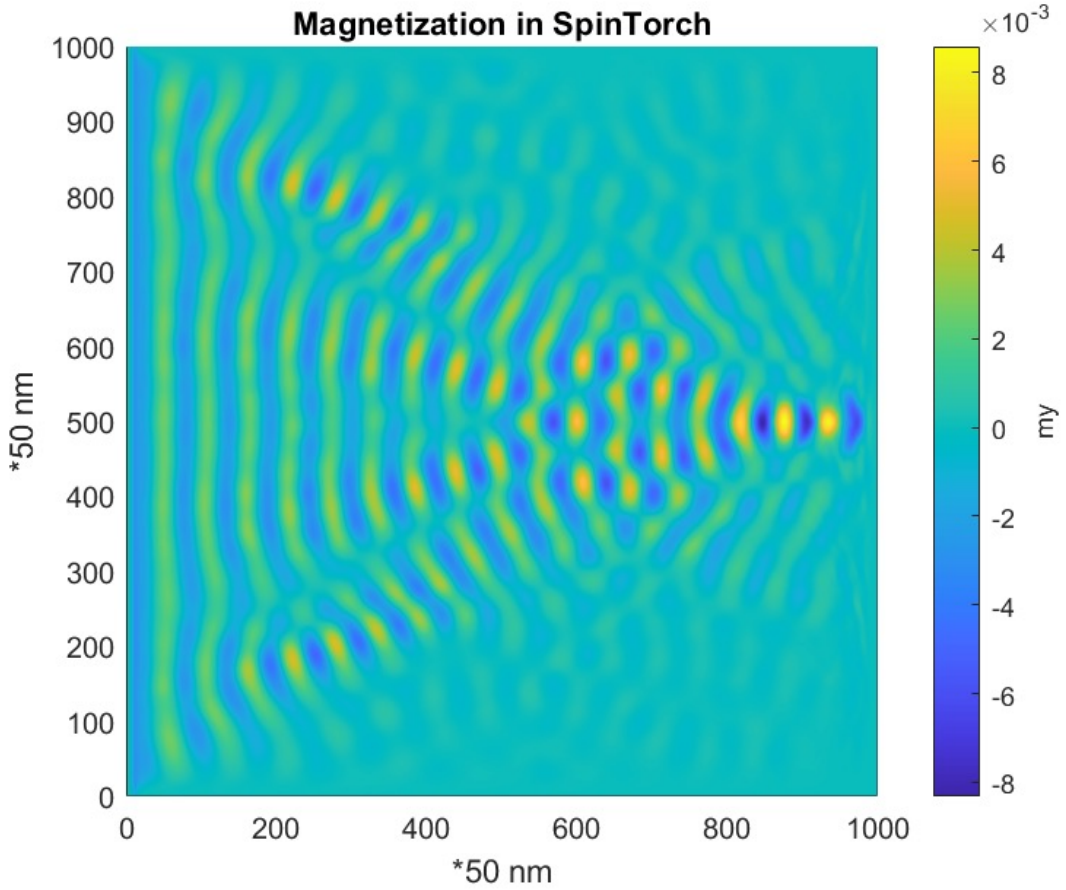


Machine learning M_s and the dose

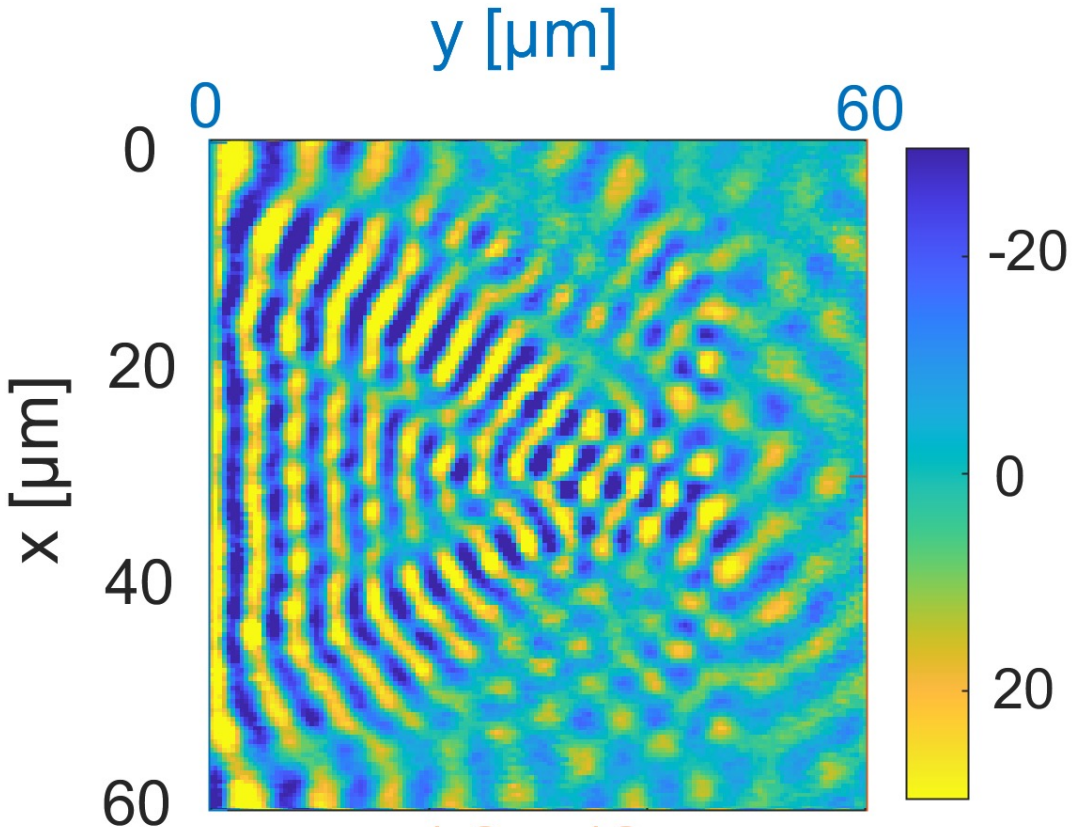




And it works!



Inverse-designed lens - simulation

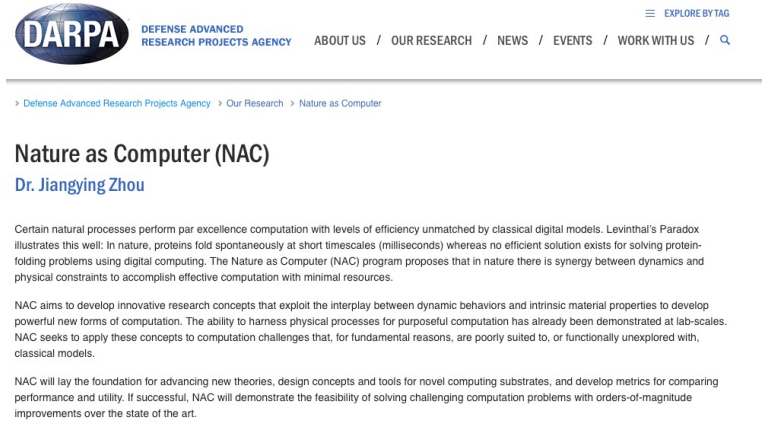


TR-MOKE measurement

Summary

- *Machine learning is a great tool to design magnetic / magnonic nanostructures*
- *Allows to design the system at once, solves the interconnection problem*
- *We hope that it could be used for other types of magnetic systems*
- *Nonlinear spin waves are an excellent media to do wave-based neuromorphic computing*
- *☹ machine learning is very computationally intensive – this limits the complexity of the system that can be simulated*
- *FIB is excellent tool for prototyping spin wave scatterers*

- *Code is in github : https://github.com/a-papp/SpinTorch*


DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

EXPLORE BY TAG

ABOUT US / OUR RESEARCH / NEWS / EVENTS / WORK WITH US /

› Defense Advanced Research Projects Agency › Our Research › Nature as Computer

Nature as Computer (NAC)

Dr. Jiangying Zhou

Certain natural processes perform par excellence computation with levels of efficiency unmatched by classical digital models. Levinthal's Paradox illustrates this well: In nature, proteins fold spontaneously at short timescales (milliseconds) whereas no efficient solution exists for solving protein-folding problems using digital computing. The Nature as Computer (NAC) program proposes that in nature there is synergy between dynamics and physical constraints to accomplish effective computation with minimal resources.

NAC aims to develop innovative research concepts that exploit the interplay between dynamic behaviors and intrinsic material properties to develop powerful new forms of computation. The ability to harness physical processes for purposeful computation has already been demonstrated at lab-scales. NAC seeks to apply these concepts to computation challenges that, for fundamental reasons, are poorly suited to, or functionally unexplored with, classical models.

NAC will lay the foundation for advancing new theories, design concepts and tools for novel computing substrates, and develop metrics for comparing performance and utility. If successful, NAC will demonstrate the feasibility of solving challenging computation problems with orders-of-magnitude improvements over the state of the art.