Article 25fa pilot End User Agreement

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with explicit consent by the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) ‘Article 25fa implementation’ pilot project. In this pilot research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and/or copyrights owner(s) of this work. Any use of the publication other than authorised under this licence or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the Library through email: copyright@ubn.ru.nl, or send a letter to:

University Library
Radboud University
Copyright Information Point
PO Box 9100
6500 HA Nijmegen

You will be contacted as soon as possible.
Supervised learning of an opto-magnetic neural network with ultrashort laser pulses

Cite as: Appl. Phys. Lett. 114, 192407 (2019); https://doi.org/10.1063/1.5087648
Submitted: 03 January 2019 . Accepted: 29 April 2019 . Published Online: 16 May 2019

A. Chakravarty, J. H. Mentink, C. S. Davies, K. T. Yamada, A. V. Kimel, and Th. Rasing

ARTICLES YOU MAY BE INTERESTED IN

Light confinement and high current density in UVB laser diode structure using Al composition-graded p-AlGaN cladding layer
Applied Physics Letters 114, 191103 (2019); https://doi.org/10.1063/1.5095149

Superfast domain wall motion in lithium niobate single crystals. Analogy with crystal growth
Applied Physics Letters 114, 192902 (2019); https://doi.org/10.1063/1.5094688

Sparse neuromorphic computing based on spin-torque diodes
Applied Physics Letters 114, 192402 (2019); https://doi.org/10.1063/1.5090566
Supervised learning of an opto-magnetic neural network with ultrashort laser pulses

The explosive growth of data and its related energy consumption is pushing the need to develop energy-efficient brain-inspired schemes and materials for data processing and storage. Here, we demonstrate experimentally that Co/Pt films can be used as artificial synapses by manipulating their magnetization state using circularly polarized ultrashort optical pulses at room temperature. We also show an efficient implementation of supervised perceptron learning on an opto-magnetic neural network, built from such magnetic synapses. Importantly, we demonstrate that the optimization of synaptic weights can be achieved using a global feedback mechanism, such that the learning does not rely on external storage or additional optimization schemes. These results suggest that there is high potential for realizing artificial neural networks using optically controlled magnetization in technologically relevant materials, which can learn not only fast but also energy-efficient.

The perceptron is an elementary model of neural computation. An example of the simplest single-layer perceptron is shown in Fig. 1(a). It has inputs $x_1, \ldots, x_p$ that are connected to adaptable synaptic weights $w_1, \ldots, w_p$. The output $O$ is a nonlinear function of the weighted inputs $y_l = w_l x_l$ and the threshold $b$: $O = \text{sign}(\sum y_l - b)$, where we introduced the sign function as a simple nonlinear function. In this form, the perceptron can be interpreted as a neuron that fires when the sum of the weighted inputs is larger than a threshold. We focus on the simple perceptron as an example of a neural network that can learn (by adapting synaptic weights) a given function, also known as supervised learning. To this end, several input patterns $x_1^{\mu}, \ldots, x_p^{\mu}$, $\mu = 1, 2, \ldots, p$ with desired outputs $O^{\mu}$ are provided. The perceptron learning rule can then be defined as an iterative procedure where one
cycles over the input patterns. At each pattern, the actual output \( O'' \) is evaluated and the weights are changed according to a global error \( E = \text{sign}(O'' - O') \),

\[
w_i = w_i + \Delta w_i, \quad \Delta w_i = \eta x_i E. \tag{2}
\]

Here, \( \eta \ll 1 \) is the learning rate. Learning stops when for all patterns the desired output is realized \( (E = 0) \). Note that, although a global feedback is used, weights are changed locally due to the dependence on \( x_i \) in Eq. (2). Importantly, this iterative learning procedure does not necessitate additional external storage or external optimization. Instead, it can exploit the nonvolatile local storage of synaptic weights in the material itself. Implementing a perceptron in a physical system therefore requires continuous adaptable and nonvolatile synaptic weights as well as multiplication of the weights with given inputs. Below, we describe how this is realized using a magneto-optical setup and subsequently show how learning is achieved.

To implement the multiplication, we exploit the concept of a polarizing microscope, that is, illustrated in Fig. 1(b). The magnetic sample that features magnetization parallel or antiparallel to the optical wave vector is placed between two polarization filters. Due to the magneto-optical Faraday effect, the axis of optical polarization is rotated clockwise (anticlockwise) via the magneto-optical Faraday effect. The relative angle between the polarizers is tuned such that minimum (maximum) light passes through the analyzer when the light transmits through magnetic domains pointing parallel (antiparallel) to the light’s direction, i.e., pointing up or down.

From analyzing the light power \( P \propto I_{out} \), that is, locally detected by the CCD camera we obtain, for small Faraday rotations as applicable here, a linear relation with the incoming illumination \( I_{in} \),

\[
I_{out} = K M I_{in}, \tag{3}
\]

Here, \( K \) depends on the material-specific Verdet constant, the thickness \( d \), and magnetization \( M \) of the material, while \( M = d(S_1 - S_2)m \) is the net magnetic moment, where \( S_1 \{ S_2 \} \) denotes the area of up (down) domains. By identifying \( I_{out} \) with the weighted input \( y_i \), “KM” with the synaptic weight \( w_i \) and \( I_{in} \) with the synaptic input \( x_i \), we directly obtain the desired multiplication.

To realize adaptable and nonvolatile synaptic weights, we optically control the magnetization \( M \). Therefore, a magnetic material with strong out-of-plane magnetic anisotropy and a strong opto-magnetic response is needed. Moreover, instead of exploiting single-shot binary switching seen in ferrimagnetic alloys and multilayers, \(^{16-18}\) reversible gradual changes of magnetization in response to laser pulses are required. To this end, the multishot helicity-dependent control of magnetization in ferromagnetic Co/Pt thin films is an attractive candidate. \(^{14,15,18}\) In this study, we used a sputter-grown Co/Pt multilayered thin film structure, mounted on a synthetic quartz glass substrate, with the nanometer thickness composition of Glass/Ta(3)/Pt(3)/Co(0.6)/Pt(3)/MgO(2)/Ta(1). The structure has a coercive field of 20 mT. The magneto-optical setup in which both the multiplication and optical control of synaptic weights are implemented is schematically shown in Fig. 1(c). Gaussian pump pulses of 800 nm central wavelength and 4 ps duration were selected from a 1 kHz repetition rate Ti:sapphire amplified laser system. The pump pulse was targeted to hit the sample at about 10° incidence angle and focused to a spot with a Gaussian waist of 38 \( \mu \)m. The measured incident optical fluence was maintained at 1.3 mJ/cm². A rotatable quarter wave plate was used to select left/right circularly polarized pump pulses to either decrease or increase the net magnetic moment of the sample and thereby adapt the synaptic weight. For ease of demonstration, a continuous wave laser of 532 nm wavelength was used for probing.

In Fig. 2, we study the response of the magnetization in the Co/Pt thin film to laser pulses. Without laser irradiation, the selected area of the sample shows a nearly uniform light gray contrast. Two positions of the sample are selected for the definition of synaptic weights \( w_1, w_2 \). These are seen in Fig. 2(a) as dark regions with opposite

![FIG. 1.](image-url)  
(a) Schematic illustration of the global feedback mechanism for a single-layer perceptron with adaptable synapses \( w_i \), threshold \( b \), and inputs \( x_i \). (b) Implementation of an artificial synapse using the remnant magnetization of a magnetic medium placed between two nearly crossed polarizers, P and A. (c) The weight is continuously adapted by optical control of the magnetization using pump pulses. Translating the sample (T-Stage) allows selecting the positions of distinct synaptic weights. Computer-controlled rotation of the quarter wave plate (\( \lambda/4 \)) is used to determine the sign of pump-induced magnetization changes. Binary inputs are implemented by moving a shutter (T-Shutter) in front of the CCD camera.

![FIG. 2.](image-url)  
(a) A raw image of two physically separated laser-written weight spots, \( w_1, w_2 \) (dark spots); the bottom shows eight background subtracted images illustrating the evolution of the magnetization change due to the right (top row) and left (bottom row) circularly polarized pump pulses irradiating the Co/Pt sample. (b) Extracted intensity changes as a function of the number of pump pulses, demonstrating continuously controllable weights, shown in black (blue) for \( w_1 \) (\( w_2 \)). Black solid disks correspond to the images in the bottom rows of (a). The inset shows the pulse packets used for learning.
magnetization and are obtained by irradiation with 600 right circularly polarized laser pulses, sufficient to yield saturation. The eight images at the bottom of Fig. 2(a) show experimental results of pumping with laser pulses for right (top row) and left (bottom row) circular polarization. For clarity, the background magnetization of the sample prior to laser-writing was subtracted. We observe that for right circularly polarized light, the intensity, and thus the switched domain area, grows with the number of pulses. Conversely, for left circularly polarized light, the previously large domain area decreases in size, as the number of pulses increases.

To quantify the adaptability, Fig. 2(b) shows the integrated pixel intensity detected by the CCD camera as a function of the number of right handed circularly polarized pump pulses for both synaptic weights. A maximum of \( N = 600 \) pulses, each of 4 ps duration, was used. We find that the magnetic contrast changes approximately linearly with the number of pulses. Deviation from purely linear behavior is attributed to deviations of the polarization state, which were estimated to be about 5% by comparison of initial and final ellipticity. After 600 pulses, the pump helicity was changed to left circular polarization, and a nearly linear decrease in the integrated intensity (ending close to the initial value) was observed after exposure to another 600 pulses. Hence, using the Co/Pt sample, we demonstrate continuous and reversible laser-writing of nonvolatile magnetic synaptic weights.

To implement supervised learning, we focused on the simplest example of a single-layer perceptron with two inputs. This artificial neural network can learn linearly separable problems such as the logical AND and OR functions. For two inputs, there are four input patterns \((x^1, x^2) = (0, 0); (1, 0); (0, 1); (1, 1)\). Weighted inputs \(y_1 = x_1 + b\), \(y_2 = x_2 + b\) respectively, while the third row shows the integrated intensity calculated from the fourth row. These are used to evaluate the normalized integrated pixel intensity signal shown in the bottom row. This integrated intensity is negative for all input patterns prior to learning, as seen in Figs. 3(a) and 3(d).

The goal of the learning scheme is to obtain the desired output for each of the input patterns by adapting the synaptic weights. This is demonstrated in Fig. 3. Figures 3(a) and 3(d), and Figs. 3(c) and 3(f) show the magneto-optical images recorded before and after the learning of the AND (OR) function, respectively. In each of these panels, rows 1–3 show recorded images of weighted inputs \(y_1, y_2\) and threshold \(-b\), respectively. The columns indicate the different input patterns. For each pattern, the sum \(y_1 + y_2 - b\) is shown in the fourth row. These are used to evaluate the normalized integrated pixel intensity signal shown in the bottom row. This integrated intensity is negative for all input patterns prior to learning, as seen in Figs. 3(a) and 3(d).

![Supervised learning in the magneto-optical setup.](image)

(a) Panel (b) shows the evolution of the integrated intensity \(y_1 + y_2\) during learning, which proceeds by cycling through the input patterns. (d)–(f) show the corresponding results for the implementation of the OR function.
In order to learn the AND function, weights are updated according to the perceptron learning rule [Eq. (2)], which uses 1 pulse packet of 5 pulses as the learning rate. Figure 3(b) shows the evolution of the integrated intensity of the measured weighted sum \( y_{1} + y_{2} - b \) as a function of the number of learning steps. For each step, only one pattern \( \mu \) is evaluated and subsequent steps are obtained by sequentially cycling through the input patterns. After 16 learning steps, the AND function is learned. This is directly verified from the integrated pixel intensities shown in the bottom of Fig. 3(c), displaying a positive sign of the integrated pixel intensity only when both inputs are one (\( \mu = 1 \)). Similarly, an OR gate was learned, for which 70 learning steps were needed [Fig. 3(e)]. Again, this is directly verified from the bottom panel of Fig. 3(b), showing positive integrated intensities except when both inputs are zero (\( \mu = 0 \)).

To infer how accurately the experimental implementation follows the perceptron learning rule, we compared the evolution of weights seen experimentally with those from numerical simulation. The latter is a software implementation in MATLAB of the supervised learning algorithm described above with the learning rules Eqs. (1) and (2). To facilitate the comparison, we normalized the integrated intensity experimentally determined in Fig. 2. For the AND gate, we obtained \( (w_1, w_2) = (0.0027, 0.27) \) and the threshold \( b = 0.66 \). Using a learning rate of \( \eta = 0.0665 \), the simulation learns the AND function with the same number of learning steps as in the experiment. The evolution of the weights is plotted in Fig. 4(a), which shows experimental and simulated weights as red triangles and black stars, respectively. Experimental error bars take into account intensity fluctuations in the background subtracted CCD images as measured by blocking the pump beam. We observe that unlike the simulated perceptron, in practice the weights do not exhibit a strictly monotonic increase during learning. Deviations from strictly linear weight changes are seen also in Fig. 2. They can be modeled as variations in the learning rate due to changes in the pump polarization but do not give rise to nonmonotonic weight variations.

Instead, we attribute this to light scattering from the shutter, which is present for the patterns \( \mu = 1 \), no weight measurements take place since inputs are blocked by the shutter), leading to smaller measured weight values for pattern \( \mu = 4 \). The influence of such scattering is stronger when more learning steps are needed as we explored for the OR gate in Fig. 4(b). In this case, the normalized threshold is \( b = 0.58 \) and the normalized initial weights are the same as for the AND gate. Without accounting for light scattering, the simulations learn the OR function 36 steps faster than seen in experiment (data not shown). If, instead we account for a systematic offset in the weight measurement for \( \mu = 4 \): \( w_{1} \rightarrow w_{1} + c_{4} \), where \( c_{4} = -0.027 \), we recover qualitatively similar oscillatory weight changes during the learning process. As shown in Fig. 4(b), this also proceeds with the same total number of learning steps as in the experiment, supporting the conjecture that the scattered light is the source of the non-monotonic evolution of weights. Importantly, despite the presence of these experimental nonidealities that slow down the learning, convergence is still reached. This shows that the perceptron learning rule is rather robust against device imperfections, suggesting that upscaling to larger opto-magnetic neural networks might well be feasible.

In summary, we have demonstrated an opto-magnetic implementation of an artificial neural network. As an application, we demonstrated supervised learning of the Boolean AND and OR functions. The synaptic weights are stored nonvolatile and adapted reversibly using ultrashort circularly polarized laser pulses exploiting a global feedback mechanism. These results suggest high potential for improving the energy-efficiency and learning speed of artificial neural networks. In the current setup, the energy absorption per learning step is estimated from the optical fluence and diameter used for the synaptic weights (1.125 \( \mu m \)), yielding 64 pJ/synapse/step. The time between successive learning steps is currently limited by the translation stages, being on the order of seconds. In principle, much faster control of light is feasible using wavefront shaping.\(^{(19,20)}\) Ultimately, the learning speed is only limited by the optical pulse duration, yielding rates as high as 50 GHz for the current setup. It will be very interesting to investigate downsampling the synaptic weight area to reduce the energy cost and time required per learning step, potentially down or even below the femtosecond and picosecond regime for nanoscale synapses.

We would like to acknowledge discussions with Bert Kappen, Misha Katsnelson, Alex Khajetoorians and Riccardo Zecchina, as well as F. Ando and T. Ono for sample growth. This research was in part supported by the ERC Advanced Grant No. 339813 (EXCHANGE), the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), and the Nationale Wetenschaps Agenda (NWA, startimpuls GreenICT). J.H.M. acknowledges funding from NWO by a VENI grant and is part of the Shell-NWO/FOM-initiative “Computational sciences for energy research” of Shell and Chemical Sciences, Earth and Life Sciences, Physical Sciences, FOM and STW. Technical support from Tonnie Toonen, Chris Berkhout, and Sergey Semin is gratefully acknowledged.

**REFERENCES**


