

Article

An Integrated Photorefractive Analog Matrix-Vector Multiplier for Machine Learning †

Elger A. Vlieg ^{1,*}, Lucas Talandier ², Roger Dangel ^{1,†}, Folkert Horst ^{1,†} and Bert J. Offrein ¹

¹ IBM Research—Zurich, 8803 Rüschlikon, Switzerland; rogerf.dangel@gmail.com (R.D.); fho@zurich.ibm.com (F.H.); ofb@zurich.ibm.com (B.J.O.)

² IFISC, 07122 Palma de Mallorca, Spain; lucastalandier@ifisc.uib-csic.es

* Correspondence: VLI@zurich.ibm.com

† This paper is an extended version of our paper published in OSA Technical Digest.

‡ These authors contributed equally to this work.

Abstract: AI is fueling explosive growth in compute demand that traditional digital chip architectures cannot keep up with. Analog crossbar arrays enable power efficient synaptic signal processing with linear scaling on neural network size. We present a photonic photorefractive crossbar array for neural network training and inference on local analog memory. We discuss the concept and present results based on the first prototype hardware.

Keywords: optical computing; in-memory computing; optical neural networks; integrated photonics; holographic memory



Citation: Vlieg, E.A.; Talandier, L.; Dangel, R.; Horst, F.; Offrein, B.J. An Integrated Photorefractive Analog Matrix-Vector Multiplier for Machine Learning. *Appl. Sci.* **2022**, *12*, 4226. <https://doi.org/10.3390/app12094226>

Academic Editors: Ripalta (Patty) Stabile and Christos (Chris) Vagionas

Received: 30 March 2022

Accepted: 19 April 2022

Published: 22 April 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have become key drivers for a wide range of applications. The Artificial Neural Network (ANN) is a cornerstone of ML (Figure 1a). Computationally, the synaptic signal propagation between its neuron layers with sizes N and M involves a matrix-vector multiplication (MVM) with an $N \times M$ matrix and thus involves $N \times M$ multiply-accumulate (MAC) operations. Synaptic signal propagation is required for both inference and training. AI applications are growing in complexity and the associated ANNs are growing in size [1]. Consequently, the compute demand in AI is growing at an exponential rate that traditional digital compute architectures cannot keep up with [2,3].

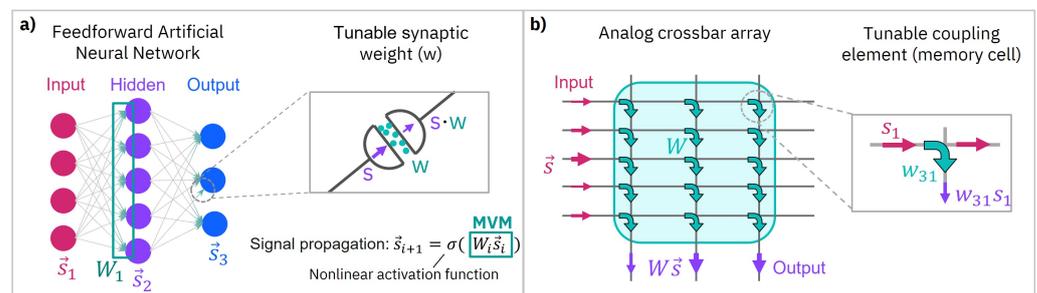


Figure 1. (a) The feedforward Artificial Neural Network (ANN) is a cornerstone of Artificial Intelligence. (b) The analog crossbar array is a chip architecture to efficiently accelerate the synaptic signal transfer between the neuron layers. Conceptually, it directly emulates the synaptic interconnect layer.

Analog crossbar arrays can efficiently accelerate and scale MVM operations (Figure 1b) [4–6]. The crossbar array consists of input and output channels. At each channel intersection, a tunable analog coupling element is placed, representing a matrix coefficient. Once the matrix is programmed, MVM operations are executed directly by setting the analog amplitudes of the input vector. The vector-matrix product is retrieved by

the readout of the output channels. Then, the compute effort of the ANN signal propagation scales only linearly with network size.

In principle, the nonlinear activation operation is not performed by the crossbar array (Figure 1a), and matrix operations are targeted only. Analog nonlinear elements could be added to the outputs of the array to include this functionality. However, the nonlinear activation can be performed digitally, while maintaining linear compute scaling with network size. The acceleration of vector operations yields lower-order efficiency gains.

In digital computing, memory transfers are particularly energy costly, and meaningful efficiency improvements can only be achieved by limiting them [7]. The analog crossbar array prevents repeated $N \times M$ memory transfers of the synaptic weights by storing them locally in the tunable coupling elements (Figure 1b).

ANNs are trained before they can be deployed for inference. The training involves the tuning of the synaptic weights in order to realize the correct mapping of the input state to the desired output state. This is achieved by iteratively feeding training data into the network, assessing the output error compared to the desired target, and reducing the error by tuning the synaptic weights. The optimal tuning direction of the weights is along the gradient of the error function, which can be resolved by the back propagation algorithm (Figure 2) [8]. The training is completed once the iterative process has converged and thus when the error function has been minimized. ANN training is computationally intensive and is an important target for acceleration [4].

A crossbar array to accelerate ANN inference needs to support the forward propagating MVM operation solely (Figure 2a). A crossbar array to accelerate ANN training by back propagation needs to support the transposed matrix and scaled outer product operation too (Figure 2b). The transposed matrix operation requires the possibility to swap the roles of the input and output channels (Figure 1b). The outer product operation requires weights that can be updated linearly by the vector signals. If the weight update response is nonlinear, the update is inaccurate and not along the gradient descent direction. Then, the convergence of the error function will either require more training iterations or is hindered altogether. Evidently, an analog crossbar array for ANN training and inference has more elaborate hardware design requirements compared to a strictly forward-propagating array for inference only.

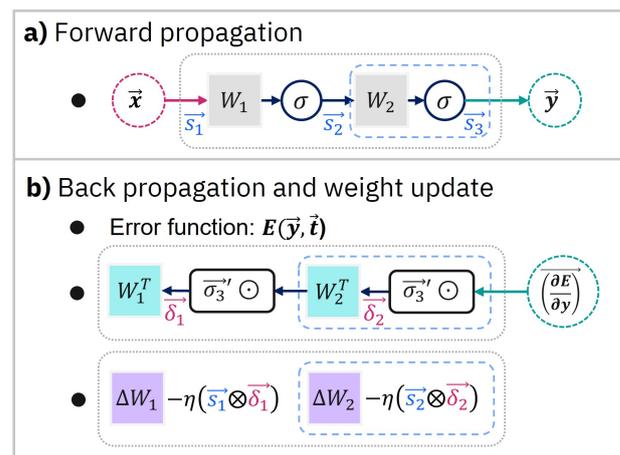


Figure 2. A schematic representation of the data flow in a 3-layer ANN (Figure 1a) [9]. The colored squares denote matrix operations and are candidates for acceleration by a crossbar array. Additional layers can be included trivially, and the contribution of a single layer is highlighted by the blue striped boxes.

The optimal hardware implementation of a crossbar array is a topic of scientific investigation. In this study, a photonic photorefractive crossbar array is discussed, and the status of its development is outlined. What makes this design particularly interesting is the support for ANN training.

In the Materials and Methods section, the concept is discussed. The photorefractive effect is exploited to induce refractive index gratings that diffract light between photonic waveguides. The synaptic interconnect layer is thereby realized in hardware. We explain how ANN training is enabled without intermediate memory access to the synaptic weight values.

In the results section, experimental data is presented based on prototype hardware. Previously, free-space grating writing, beam shaping, and input stages have been demonstrated [4]. The fabrication of integrated photonic circuits based on photorefractive GaAs thin-films and the characterization thereof is new.

2. Materials and Methods

2.1. Photorefractive Effect

2.1.1. Background

The photorefractive effect is a nonlinear optical effect that converts an optical intensity variation to a refractive index variation (Figure 3) [10]. The underlying mechanism is an electrical charge redistribution inside the photorefractive crystal. A photorefractive crystal hosts deep charge traps, from which charge carriers are removed via photoexcitation. Photoexcitation occurs favorably in areas of high optical intensity, and recombination occurs favorably in dark areas. The resulting spatial charge redistribution gives rise to a static local electric field profile, which yields an index of refraction variation via the linear electro-optic effect. In the absence of light, the generated space-charge modification will persist at a time scale depending on the material’s dark conductivity.

At optical frequencies, intensity variations with a non-zero time average can be generated by the interference of two monochromatic beams (Figure 3a). The resulting intensity pattern is sinusoidal with a wave vector equal to the wave vector difference of the interfering beams (Figure 3b). Inside a photorefractive crystal, consequently a sinusoidal refractive index grating is formed that acts as a Bragg reflector (Figure 3e,f). The Bragg condition is met precisely along the respective intersecting beam direction.

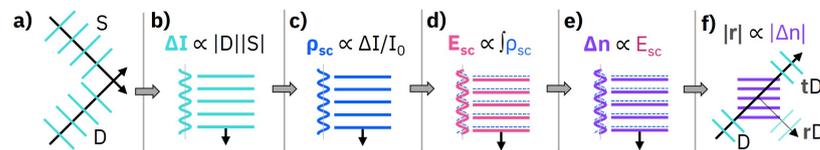


Figure 3. Schematic representation of the photorefractive effect and refractive index grating formation [10]. (a,b) Two interfering monochromatic beams with amplitudes S & D generate a sinusoidal intensity variation ΔI . (c) The fringe visibility $\Delta I / I_0$ relative to the background intensity I_0 causes an electric space-charge redistribution ρ_{sc} . (d) ρ_{sc} gives rise to a local electrical space charge field E_{sc} that is $\pi/2$ phase shifted with respect to the optical fringe pattern. (e) E_{sc} yields an refractive index grating Δn via the linear electro-optic effect, (f) which acts as a Bragg reflector.

2.1.2. Crossbar Array Architecture

In this subsection, it is discussed how an analog crossbar array can be realized by photorefractive index gratings. Optical beams will represent the synaptic signals. Refractive index gratings will constitute synaptic coupling elements.

Inside a photorefractive crystal, a source and destination laser beam intersect with field amplitudes $S_1^{(w)}$ and $D_1^{(w)}$, respectively (Figure 4a). A refractive index grating is written, and reciprocal coupling is established between the source and destination channel.

The coupling strength w of the grating is tunable and governed by its reflection coefficient r (Figure 3f). When the write beams are enabled, the electric space charge distribution ρ_{sc} (Figure 3c) builds up approximately linearly until saturation comes into effect. Within limits, the forming rate $\frac{\partial \rho_{sc}}{\partial t}$ is proportional to the optical intensity variation ΔI (Figure 3b) [10]. Thus, the coupling strength w is programmed by the electric field amplitudes $S_1^{(w)}$ and $D_1^{(w)}$ and the write time t_{write} (Figure 4b).

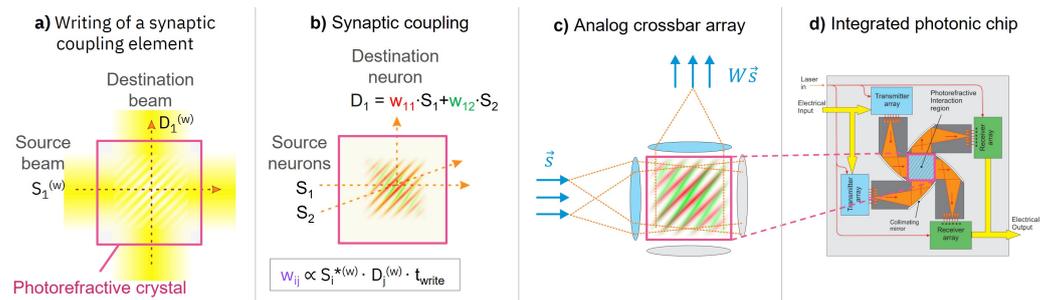


Figure 4. (a–c) A photorefractive scheme for realizing an analog crossbar array in hardware. Photorefractive index gratings constitute synaptic coupling elements, and the full synaptic interconnect layer is realized by superimposing them. (d) An integrated photonic circuit offers performance benefits in terms of stability, power, and write times. All coefficients are complex, and phase stability is of importance.

After the writing phase, the diffracted source signal is given by the product of the source amplitude S and the grating coefficient w (Figure 4b). Thus, the analog multiplication of the input signal and grating amplitude is performed.

Negative, or in general complex, grating and beam coefficients are possible (Figure 4b). The optical signals can simply be phase shifted to rotate the respective coefficient in the complex plane.

In order to perform the MVM, an array of coupling elements needs to be realized and the multiplied signals need to be accumulated (Figure 1b). In the photorefractive crossbar array, refractive index gratings are superimposed to add matrix elements (Figure 4b) [11]. For instance, another input signal coming from a different direction can be coupled to the same destination channel by repeating the above grating writing routine. Since the grating profile is sinusoidal, only one diffraction order is possible and only along the direction of the write beams. Therefore, the gratings do not generate channel crosstalk as long as the angular channel separation is sufficiently large compared to the inverse grating size.

Thus far, signal channels have been discussed as angularly separated collimated beams. However, data enters and leaves the crossbar array via spatially separated signal paths, each carrying a vector coefficient that represents a synaptic signal (Figure 1b). Collimating optics are deployed to transform between the spatial and angular domain (Figure 4c).

The complete synaptic weight matrix W is programmed in the photorefractive crossbar array by applying write signals to the respective input and output vector channel pairs. Then, when an input vector is applied, the superimposed diffracted signal consists of the multiplied and accumulated source amplitudes. The output signal thus represents the matrix-vector product (Figure 4c) and can be retrieved by photodetection in the output channels. We envision a coherent detection scheme in order to retrieve the complex output vector coefficients linearly and to enhance sensitivity compared to direct photodetection.

In order to realize a practical device, the photorefractive crossbar array processor is designed as an integrated photonic chip (Figure 4d). The photorefractive coupling requires an angular degree of freedom, which is provided by a 2D slab waveguide interaction area. Collimating mirrors are realized by a curved waveguide sidewall. The input vector is set by an electronically controlled optical input stage that maps a source beam to the input channels. A detector stage converts the optical output back to electrical signals. Input and detection stages are included for both the source and destination branch.

The photorefractive crossbar array can accelerate both ANN inference and training via back propagation (Figure 5). As discussed in the introduction, the latter requires the transposed matrix and output product operation (Figure 2b). Since the gratings couple reciprocally, the transposed matrix is simply accessed by swapping the input and output branch (Figure 5a,b).

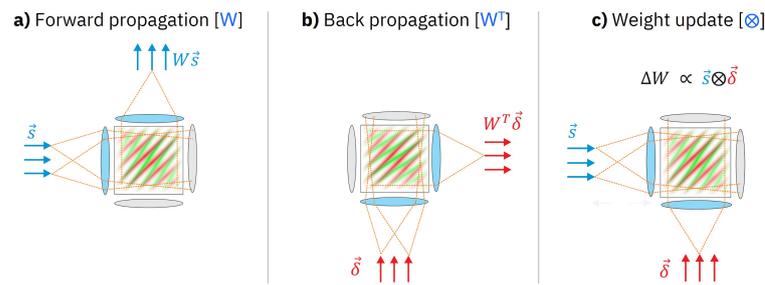


Figure 5. The photorefractive crossbar array supports all matrix operations for ANN inference and training via back propagation (Figure 2). None of these operations require direct access to the matrix coefficients. During training, the crossbar array can be operated as a black box once initialized. The coefficients are only read out once the error function has been minimized.

The weight update by the outer product is performed by applying the respective vector signals simultaneously via the input stages (Figure 5c). The grating write behavior yields the desired operation precisely (Figure 4b). Physically, decreasing a weight value is possible because a phase-shifted write-beam interference pattern may act as an erasing force on an existing grating.

2.1.3. Two-Wave Mixing

Experimentally, the photorefractive index writing as shown in Figure 4a can be demonstrated via two-wave mixing [10,12]. In this experiment, the source and destination write beams are not switched off, and the diffracted grating signal is measured by its interference with the propagating beam.

Two-wave mixing is an attractive experimental approach, since the interference signal can be orders of magnitude stronger than the diffracted grating signal alone. Namely, a linear intensity interference term is possible since the photorefractive grating is phase shifted with respect to the optical interference pattern (Figure 3), thereby, breaking symmetry between the source and destination channel. Within the limits of perturbation theory, the interference signal then scales linearly with the weak coupling strength of the grating ($\propto \gamma l$), rather than quadratically [10]

$$\Delta I_D^{(s)} = \frac{2\kappa\gamma l}{1 + \beta}. \tag{1}$$

Here, $\Delta I_D^{(s)}$ is the steady-state relative intensity change at the detector, $\kappa = \pm 1$ keeps track of constructive versus destructive interference, β is the incoming power ratio of the destination versus the source beam, l is the interaction length, and γ is the coupling constant per unit length. κ is governed by the crystal orientation and light polarization and is mirrored between the source and destination signal path. A detailed expression for γ in terms of material and experimental parameters can be found in literature [10].

If all experimental conditions are fixed, the relative intensity modification due to photorefractive two-wave mixing ΔI_D is directly proportional to Δn , and two-wave mixing is suitable to probe relative modifications to the index grating amplitude.

The grating diffraction signal ΔI_D can be robustly measured by repeatedly triggering the write sequence and observing the intensity modification in time (Figure 6). Write sequences are triggered in an experimentally attractive manner by square wave modulating the phase of the source beam between 0 and π . The equilibrium position of the grating will phase shift by π . Momentarily, the diffracted signal will reverse between constructive and destructive interference until the new grating is formed. This yields up to double the signal modification, and no intermediate grating erase sequence is required.

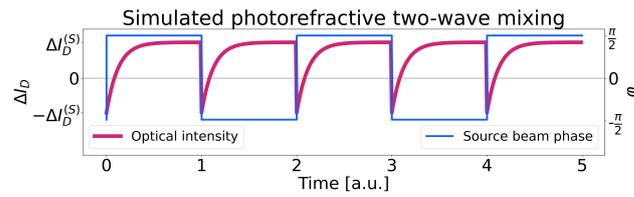


Figure 6. Simulated relative intensity change at the destination channel ΔI_D during photorefractive two-wave mixing with a square wave phase-modulated source beam (Figure 4a).

The photorefractive grating forms via an exponential relaxation process (Figure 6) [10]. For $0 \leftrightarrow \pi$ square wave phase modulation of the source beam at a frequency much below the inverse photorefractive response time τ^{-1} , the relative intensity modification after a phase-flip is then given by

$$\Delta I_D(t) = \Delta I_D^{(S)} \left(1 - 2e^{-t/\tau} \right). \tag{2}$$

A detailed expression for τ can be found in literature [10].

2.1.4. Experimental Setup for Two-Wave Mixing

The free-space experimental setup for two-wave mixing is outlined in Figure 7. This setup is used to characterize the photorefractive performance of the thin-film semi-insulating GaAs samples (Section 2.2).

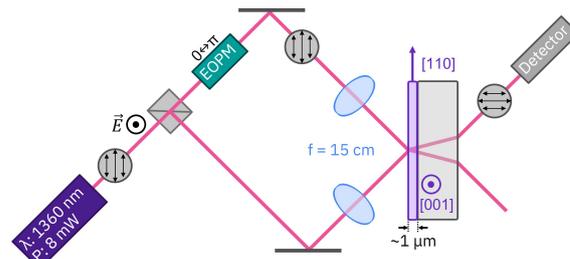


Figure 7. Experimental setup for two-wave mixing. An electro-optic phase modulator (EOPM) is used to trigger write cycles for the photorefractive index grating.

The photorefractive GaAs thin-film sample (purple) faces the beam path perpendicularly. The GaAs crystal axes are aligned in a way so that the incoming s-polarized light generates a p-polarized grating diffraction signal [12]. A linear polarizer in front of the detector is set at an azimuthal angle shifted away from p-polarization by a few degrees. Thereby, the s-polarized transmitted beam is mostly blocked, and interference with the p-polarized diffraction signal is enabled. Partially blocking the propagating beam enables higher detector gain for an improved signal-to-noise ratio.

Two focusing lenses are deployed in front of the sample in order to shrink the laser spot size. The benefit is twofold: relatively small photonic circuit structures can be probed, and the increase in intensity lowers the photorefractive response time. A faster response time enables a faster experimental repetition rate.

2.1.5. Photonic Prototype Chip Design

The design of the prototype photorefractive crossbar processor is outlined in Figure 8a. The device supports multiple input and output channels. A final design will have fully programmable input stages. Here, instead, a single input channel addresses both input branches via a 1×2 multi-mode interferometer (MMI). A thermo-optic phase modulator is included at one of the input branches. Effectively, this mimics the free-space two-wave mixing experimental setup (Figure 7). The prototype chip design allows for the demonstration of the writing of a single photorefractive synaptic coupling element.

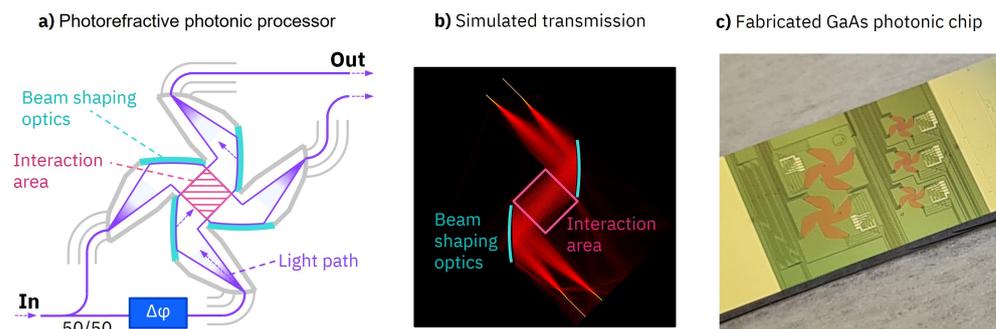


Figure 8. Overview of the prototype photonic photorefractive processor design. The input beams couple at the interaction area. The photonic circuit is patterned in a 1 μm thick photorefractive semi-insulating GaAs film.

2.2. Fabrication

An intermediate goal is to achieve integrated photonics with photorefractive materials. This requires the fabrication of photorefractive thin-film layers for the patterning of photonic circuits.

For the prototype devices, semi-insulating GaAs was selected as the photorefractive material. GaAs is appealing because it is a widely studied photorefractive material.

Photorefractive GaAs thin-films were successfully fabricated by using a top-down approach (Figure 9). Photorefractive semi-insulating crystalline GaAs wafers are commercially readily available. A 2" GaAs wafer is wafer bonded to a 4" silicon wafer containing a 2 μm SiO₂ buffer layer. Next, the GaAs is thinned down consecutively by grinding, wet etching, and chemical mechanical polishing (CMP). Thin-film samples were fabricated with thicknesses ranging between 0.2 and 4 μm .

By utilizing bulk GaAs material, the desired photorefractive characteristics are directly obtained. This implies a crystalline structure and the correct charge trap density. The growth of photorefractive GaAs thin-films by metal organic chemical vapor deposition (MOCVD) or molecular-beam epitaxy (MBE) is comparatively challenging.

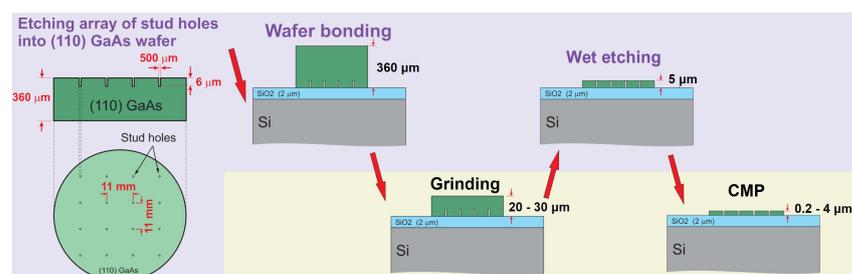


Figure 9. A semi-insulating GaAs wafer is thinned down to fabricate photorefractive thin-films. Stud holes are patterned in the GaAs wafer and enable the intermediate monitoring of the layer thickness by a profilometer during the wet etching process.

The photonic circuit (Figure 8c) was patterned into a 1 μm thick GaAs thin-film sample by using e-beam direct-write lithography. Furthermore, 15–20 nm of SiO₂ was deposited by plasma-enhanced chemical vapor deposition (PECVD) and using an SiH₄ based process. The SiO₂ promotes the adhesion of 150–200 nm hydrogen silsesquioxane (HSQ) e-beam resist. After lithography and development, the SiO₂ adhesion promotor was removed by reactive ion etching (RIE), upon which the GaAs thin-film layer was accessible for inductively-coupled plasma (ICP) etching. The SiO₂ and glass-like HSQ e-beam resist stack were not removed from the GaAs waveguides. A 1 μm SiO₂ upper cladding layer was deposited by PECVD utilizing a tetraethoxysilane (TEOS) process.

Electrical tungsten heaters were manufactured on top of the circuit for thermo-optical phase modulation. Again, e-beam direct-write lithography was used. Furthermore, 100 nm

W was sputtered and etched by RIE after patterning. An aluminum layer was fabricated on top of the tungsten electrical leads, excluding the heating strip on top of the waveguide, to promote localized heating. For the Al layer, laser direct-write lithography was used. Then, 500 nm Al was sputtered and wet etched after patterning. Lastly, the bonding of the W and Al layer was fortified by rapid thermal annealing (RTA) in an H₂-in-Ar forming gas.

2.3. Integrated Photonic Measurement Setup

Laser light was coupled in and out of the cleaved photonic chip (Figure 8c) by lensed fibers. Laser wavelength sweep measurements were performed with an optical component analyzer. Time domain measurements were performed using a photodiode and an oscilloscope.

3. Results

3.1. Characterization of the GaAs-Based Integrated Photonic Platform

The prototype chip (Figure 8c) features test structures to probe the optical and electrical performance of the GaAs photonic circuits. The characterization was done for a wavelength range of 1260 to 1360 nm. The results are shown in Figure 10. Comments on the individual graphs are listed below, respectively:

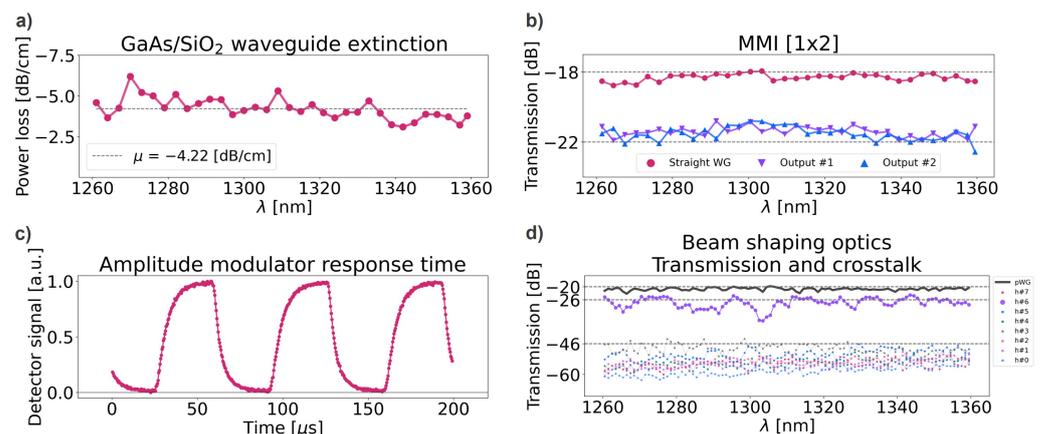


Figure 10. Performance characterization measurements of the GaAs/SiO₂ integrated photonic chip (Figure 8c).

- Figure 10a: The GaAs/SiO₂ waveguides display a typical propagation loss of less than 5 dB/cm. The extinction coefficient was obtained by comparing the transmission through different lengths of waveguide.
- Figure 10b: 1 × 2 MMIs display the correct approximate −3 dB attenuation compared to a reference waveguide.
- Figure 10c: The electric tungsten-heater-based thermo-optic phase modulator performs as intended, as demonstrated by its deployment in a Mach–Zehnder amplitude modulator. The response times are in the order of μs (Table 1).
- Figure 10d: This graph displays the power transmission through the photorefractive processor (Figure 8a). The experiment was performed by inserting light via a single input channel. Specifically, the input was not split up by an MMI into both input branches as denoted in Figure 8a. The black line corresponds to a waveguide that follows a similar trajectory as the light inside the processor and acts as a reference. The thick purple line denotes the signal from the designated output channel, whereas the thin lines denote the output from the neighboring output channels. The input channel is correctly transmitted to the designated output channel. The key performance indicators are the contrast between the designated output channel and its neighbors and the attenuation compared to reference. Both are indicative of the correct collimation and refocusing of the input beam and the undistorted propagation

through the GaAs slab waveguide. The contrast between the target output channel and its neighboring output channels is retrieved at around 20 dB. The transmission through the processor yields less than -6 dB power attenuation at wavelength bands along the entire spectral range.

Table 1. Fitted response times (τ) for the electric tungsten-heater-based thermo-optic amplitude modulator in μs (Figure 10c).

	Trace #1	Trace #2	Trace #3	Average
Heating	4.3	4.3	-	4.3 ± 0.1
Cooling	6.0	6.1	5.9	6.0 ± 0.1

3.2. Photorefractive Characterization of Thin-Film GaAs Layers

The photonic chip (Figure 8c) is demonstrated to be photorefractive by two-wave mixing (Figure 11a). These experiments were performed in the free-space setup (Figure 7). The signal scales directly with the amplitude of the refractive index grating (Section 2.1.3).

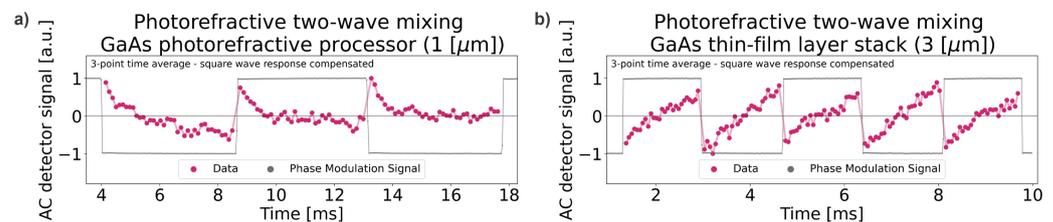


Figure 11. Photorefractive two-wave mixing in thin-film GaAs structures (Section 2.1.3). Graph (a) shows data from the photonic chip and was measured by focusing the laser beams on the interaction area of the photorefractive photonic processor (Figure 8). Graph (b) was measured on a $3 \mu\text{m}$ thin-film sample.

Additionally, photorefractive two-wave mixing is demonstrated for a $3 \mu\text{m}$ GaAs thin-film sample (Figure 11b). This sample was not patterned with a photonic circuit but did host the SiO_2 upper cladding layer. The experimental setup was modified for this measurement: no focusing lenses were deployed, and a wavelength of 1330 nm was used. The electrical charge redistribution inside the crystal does not reach saturation in this experiment due to the chosen phase modulation frequency. This data illustrates that the index grating can be tuned approximately linearly at the expense of dynamic range. Linearity is required to perform the outer product operation (Figure 5).

The interference behavior between the transmitted and diffracted beam is reversed for Experiments (a) and (b) (Figure 11). This behavior can be tuned at will and is of no significance.

4. Discussion

The results presented here constitute important milestones in the development of the integrated photorefractive crossbar array.

The photorefractive effect is well studied in bulk crystals [10,12]. Here, we report that semi-insulating GaAs can be thinned down to $1 \mu\text{m}$ thick films and that photonic circuits can be fabricated therein, while preserving photorefractive behavior (Figure 11a).

The two-wave mixing signal observed in the free-space measurement (Figure 11a) implies that the photorefractive coupling strength in the circuit is adequate for a similar demonstration in the integrated processor (Figure 8a). Both experiments are configured to resolve the intensity cross term between the transmitted and diffracted field components (Section 2.1.3). Then, the relative modification of both fields may be estimated to assess the impact of switching to the integrated experiment. The transmitted beam amplitude is estimated to increase by 15 dB, since it is not blocked by an output polarizer. The diffracted

beam amplitude is expected to increase by 23 dB. Namely, the beam interaction length will increase from 1 μm to 200 μm (Equation (1)). Moreover, the photorefractive coupling constant γ is equal or larger for the integrated circuit configuration [12]. Then, at the same laser power, the two-wave mixing modulation signal (Figure 6) should be around 10 dB larger for the integrated experiment, after 26 dB of coupling and propagation losses (Figure 10d).

It is implied that response times constitute the main experimental challenge in demonstrating two-wave mixing in the prototype processor (Figure 8a). Compared to the free-space experiment (Figure 7), at the same laser power, the grating is expected to form two orders of magnitude faster. Namely, the optical intensity is increased by the confinement of the field in integrated photonics. Then, the photorefractive response time would be similar to the response time of the thermo-optic phase modulator (Table 1). If the grating responds sufficiently fast to track the phase modulation, the optical intensity will not be modified. The photorefractive response time can be increased by lowering the laser power. However, longer integration times will be required to compensate for the loss in signal. For future chip designs, faster phase modulators should be implemented to avoid this limitation.

Lastly, this work demonstrates that a thin-film of photonic quality can be realized by a top-down fabrication approach and using CMP (Figure 10). Still, bottom-up fabrication approaches should yield an enhanced film thickness homogeneity and offer an opportunity for performance enhancement.

5. Conclusions

The analog crossbar array aims to satisfy a growing compute demand in AI by accelerating the synaptic signal transfer in ANNs. Increasingly advanced applications in AI are thereby enabled. A photorefractive crossbar array is proposed and being developed. This design is particularly attractive because it emulates the synaptic interconnect layer for training in addition to inference.

A prototype photorefractive analog crossbar chip has been fabricated. The photonic components perform according to design, and light is successfully transmitted through the processor. The waveguide material is confirmed to be photorefractive.

Photorefractive coupling in integrated photonics remains to be demonstrated. Future work will include the manufacturing of a fully programmable crossbar array and the demonstration of all critical signal processing operations.

Author Contributions: Conceptualization, F.H.; formal analysis, E.A.V.; investigation, E.A.V., L.T., R.D. and F.H.; writing—original draft, E.A.V.; writing—review and editing, L.T., R.D., F.H. and B.J.O.; visualization, E.A.V., R.D. and F.H.; supervision, R.D., F.H. and B.J.O.; funding acquisition, B.J.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Union's Horizon 2020 research and innovation program, grant numbers 828841 (ChipAI) and 860360 (Post-Digital).

Data Availability Statement: The data that supports the findings of this study is available from the corresponding author, E.A.V., upon reasonable request.

Acknowledgments: We thank the members of the Binnig and Rohrer Nanotechnology Center (BRNC) operations team for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bommasani, R.; Hudson, D.A.; Adeli, E.; Altman, R.; Arora, S.; Arx, S.V.; Bernstein, M.S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. On the Opportunities and Risks of Foundation Models. *arXiv* **2021**, arXiv:2108.07258.
2. Strubell, E.; Ganesh, A.; McCallum, A. Energy and policy considerations for modern deep learning research. In Proceedings of the AAAI 2020—34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13693–13696. [[CrossRef](#)]
3. Amodei, D.; Danny, H. AI and Compute. 2018. Available online: <https://openai.com/blog/ai-and-compute/> (accessed on 18 April 2022).

4. Stark, P.; Horst, F.; Dangel, R.; Weiss, J.; Offrein, B.J. Opportunities for integrated photonic neural networks. *Nanophotonics* **2020**, *9*, 4221–4232. [[CrossRef](#)]
5. Sebastian, A.; Le Gallo, M.; Khaddam-Aljameh, R.; Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **2020**, *15*, 529–544. [[CrossRef](#)] [[PubMed](#)]
6. Burr, G.W.; Shelby, R.M.; Sebastian, A.; Kim, S.; Kim, S.; Sidler, S.; Virwani, K.; Ishii, M.; Narayanan, P.; Fumarola, A.; et al. Neuromorphic computing using non-volatile memory. *Adv. Phys. X* **2017**, *2*, 89–124. [[CrossRef](#)]
7. Han, S.; Liu, X.; Mao, H.; Pu, J.; Pedram, A.; Horowitz, M.A.; Dally, W.J. EIE: Efficient Inference Engine on Compressed Deep Neural Network. In Proceedings of the 2016 43rd International Symposium on Computer Architecture, ISCA, Seoul, South Korea, 18–22 June 2016; Volume 16, pp. 243–254. [[CrossRef](#)]
8. Rumelhart, D.; Hinton, G.; Williams, R. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
9. Kriegeskorte, N.; Golan, T. Neural network models and deep learning. *Curr. Biol.* **2019**, *29*, R231–R236. [[CrossRef](#)] [[PubMed](#)]
10. Cronin-golomb, M.; Klein, M. Photorefractive Materials and Devices. In *Handbook of Optics: Devices, Measurements, and Properties*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1995; Volume 2, Chapter 39, pp. 1–42.
11. Owechko, Y.; Soffer, B.H. Holographic Neurocomputer Utilizing Laser Diode Light Source. *SPIE* **1995**, *2565*, 12–19.
12. Yeh, P. Two-Wave Mixing in Nonlinear Media. *IEEE J. Quantum Electron.* **1989**, *25*, 484–519. [[CrossRef](#)]