2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Interpreting Restricted Boltzmann Machines From Optics Theory Perspectives*

Ping Guo[1]

Abstract— **Currently, lack of interpretability (or explainability) is one of the major drawbacks for artificial intelligence (AI) models. When we intend to build a physical artificial intelligence (PAI) systems, the model interpretability (MI) becomes a crucial problem. To tackle MI problem, we give the explanation of restricted Boltzmann machines (RBM) from optics theory perspectives in this work. Furthermore, we present a discussion about how to implement optical learning neural network with our developed Optics Theory and Design Methods – OTDMs. With OTDMs, we can better understand the principle behind the good performance of deep neural networks. OTDMs not only give us an alternative explanation of RBM with optics theories, but also provide the guidance on designing a reliable PAI system also. Consequently, OTDMs pave the road to PAI systems, and make it to become possible for realizing all-optical learning neural network.**

I. Introduction

Currently, in the research field of artificial general intelligence (AGI) generative pre-trained transformer (GPT) model has attracted many researchers' attention. While the fundamental backbone of GPT is deep neural networks (DNNs), which is studied extensively in past decade[1][2]. As the focus domain research of artificial intelligence (AI), deep learning has progressed very rapidly and with many successful applications. With an increase in deep learning-based methods, interpretability (or explainability) of the AI systems has recently become one of problems that most scientists concerned.

Why is such a problem arising? using other words, why do we need an eXplainable Artificial Intelligence (XAI)? The reason is that in some high-stakes decision making areas such as medical diagnosis, financial market, we should know how the decision is made, is it reliable, trusted or responsible? So before we trust AI, we first need to understand how AI learns and makes decision [3]. We know that currently highly successful AI models existed are usually applied in a black box manner, no information is provided about what exactly makes them how do such predictions. That is, a major drawback of AI models is lack of transparency. It is expected that AI models can be enable full transparency, and why each decision is the right one morally, socially and financially with XAI [4]. Most scientists also believe that XAI is no doubt the next step for AI, it will improve trust, confidence and transparency.

Because interpretability on its own is a broad, poorly defined concept, why an interpretation is requested and how it is delivered may differ depending on users. In data science and applied statistics, to interpret data means to extract information [5], a set of methods may refer to designing an initial experiment, or visualizing final results. Instead of general interpretability, Murdoch *et al* focus on the use of interpretations to produce insight from AI models as part of the larger data science life cycle [5]. They define XAI as the extraction of relevant knowledge from an AI model concerning relationships either contained in data or learned by the model. They think that these insights are used to guide communication, actions, and discovery.

We believe that XAI is a very important aspect, especially for the physical artificial intelligence (PAI) [6]. We intend to build the Synergetic Learning Systems (SLS)[7], which is a kind of PAI systems. In order to provide the PAI systems design and realization methods, we will focus on AI model interpretability problem. To understand model work principle can help us to arise the model interpretability. With strong model interpretability, we can better understand the principle behind the good performance of DNNs, and guide us to design a better PAI system.

As we known, deep belief network (DBN) is a generative graphical model[8]. DBNs can be regarded as a composition of simple networks such as autoencoders [9] or restricted Boltzmann machines (RBMs) [1]. RBM is also explained as an undirected, generative energy-based model and no connections within layers. RBM can be considered as the basic building block for DBN, and we regard it as one of the simplest SLSs also [7][10]. RBM is rooted from statistical physics, in the previous work, we explain the neural network model with statistical physics approach [11]. While in this work, we mainly focus on the deep learning based model interpretability, and take the RBM as an example to demonstrate how RBM is interpreted from optics theory perspectives. In addition, we also present to implement RBM with **O**ptics **T**echniques as well as **D**esign **M**ethods of optical learning neural networks, named as OTDMs. With alternative perspective and OTDMs, it can help us to design a PAI system, more specific, an optical learning neural network (OLNN) system. And based on the optics theory for an OLNN system with which the RBM is possible implemented with optical hardware, it is no doubt the performance of PAI systems can be speeded up greatly.

II. Background

A. Brief Review of RBM

Initial, RBMs are a variant of Boltzmann machines, their neurons must form a bipartite graph. And RBMs are also

*This work was fully supported by the National Key Research and Development Program of China (No. 2018AAA0100203).

[1]Ping Guo is with International Academic Center of Complex Systems, Beijing Normal University, Zhuhai 519087, China. pguo@bnu.edu.cn

978-0-7381-4408-5/23/$31.00 ©2023 IEEE

considered as a special case of Boltzmann machines and Markov random fields, their graphical model corresponds to that of factor analysis [12].

The energy function of RBM has the same form with Boltzmann Machine, it is analogous to that of a Hopfield network (also Ising model). The energy function in matrix notation is,

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{a}^{\mathrm{T}}\mathbf{v} - \mathbf{b}^{\mathrm{T}}\mathbf{h} - \mathbf{v}^{\mathrm{T}}\mathbf{W}\mathbf{h}. \tag{1}$$

For the visible and hidden vectors $\mathbf{v}$ and $\mathbf{h}$, the joint probability distribution is defined in terms of the energy function,

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})}, \tag{2}$$

where Z is defined as the sum of $e^{-E(\mathbf{v},\mathbf{h})}$ over all possible configurations, it is a normalizing constant called a partition function in statistical physics. The sum of $P(\mathbf{v}, \mathbf{h})$ over all possible hidden layer configurations gives the marginal probability of a visible vector,

$$P(\mathbf{v}) = \frac{1}{Z} \sum_{\{\mathbf{h}\}} e^{-E(\mathbf{v},\mathbf{h})}, \tag{3}$$

and vice versa.

1) Discrete Value RBM: Since the underlying graph structure of the RBM is bipartite, for m visible units and n hidden units, the conditional probability of a configuration of the visible units $\mathbf{v}$, given a configuration of the hidden units $\mathbf{h}$, is

$$P(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{m} P(v_i|\mathbf{h}). \tag{4}$$

Conversely, the conditional probability of $\mathbf{h}$ given $\mathbf{v}$ is

$$P(\mathbf{h}|\mathbf{v}) = \prod_{j=1}^{n} P(h_j|\mathbf{v}). \tag{5}$$

RBM can be assumed as not only Gaussian visible and Bernoulli hidden units distribution, but also as Gaussian visible and Gaussian hidden units distribution [13].

2) Gaussian–Gaussian RBM: The probability distribution of a Gaussian–Gaussian RBM is defined as follows [14]:

$$p(\mathbf{v}, \mathbf{h}) = \exp\left\{ -\sum_{i=1}^{M} \frac{(h_i - c_i)^2}{2s_i^2} - \sum_{j=1}^{N} \frac{(v_j - b_j)^2}{2\sigma_j^2} + \sum_{i,j} W_{ij} \frac{h_i}{s_i} \frac{v_j}{\sigma_j} - \psi \right\}, \tag{6}$$

where both hidden h_i and visible v_j take continuous values and obey Gaussian distributions characterized by variances s_i^2 $(i = 1, \ldots, M)$ and σ_j^2 $(j = 1, \ldots, N)$.

B. RBM Learning Algorithms

1) CD-k algorithm: Firstly, let us recall the classical algorithm for RBM Learning. In the training RBM research direction, the well known algorithm is the standard contrastive divergence (CD) algorithm proposed by Hinton *et al*[14][15].

RBM is also considered as an energy based neural network model, the $CD - k$ algorithm (k is the number of iteration), a stochastic learning algorithm, is used to train RBM.

$CD - k$ algorithm not only can be applied to above Gaussian visible and Bernoulli hidden units, but also to Gaussian visible and Gaussian hidden units distribution [13].

2) Maximum Likelihood Estimate: The maximum likelihood (ML) estimate of $\mathbf{W}$ is derived by minimizing the Kullback–Leibler (KL) divergence between the input distribution and the model distribution[15]. In Gaussian–Gaussian RBM, the ML Estimate becomes [13]:

$$\tau \frac{d\mathbf{W}}{dt} = \mathbf{W}\Sigma^{-1}C\Sigma^{-1} - \mathbf{W}(\mathbf{I}_N - \mathbf{W}^T\mathbf{W})^{-1}, \tag{7}$$

where τ is a learning constant.

For continuous value neurons cases, the learning algorithms are non-linear differential equations of the weight matrix $\mathbf{W}$ and are difficult to solve analytically. When we assume that the variances of the visible and hidden units are homogeneous, and by setting $d\mathbf{W}/dt = 0$ in the Eq. (7), the equation of the equilibrium state can be obtained,

$$\mathbf{WC} = \sigma^2\mathbf{W}(\mathbf{I}_N - \mathbf{W}^T\mathbf{W})^{-1}. \tag{8}$$

In [16], Decelle *et al* proposed to train RBM by the singular value decomposition (SVD) spectrum of the weight matrix $\mathbf{W}$ by$\mathbf{W} = \mathbf{UAV}$, where $\mathbf{U}$ is an $M \times M$ orthogonal matrix, $\mathbf{A}$ is an $M \times N$ diagonal matrix, and $\mathbf{V}$ is an $N \times N$ orthogonal matrix. This allows to write a deterministic learning equation leaving aside the fluctuations of RBM learning.

RBMs can be reinterpreted as deterministic feed-forward neural networks also [17], then can be further trained by standard supervised learning algorithms.

3) TAP Algorithm: Training RBMs via the Thouless-Anderson-Palmer (TAP) free energy was proposed by Gabrié *et al* in 2015 [18].

TAP Algorithm is a deterministic learning algorithm for non-parametric learning of lifted RBMs, and adopts the gradient ascent update to approximate weight W with deterministic iteration.

C. Neural Partial Differential Equations

We intend to interpret the RBM by constructing a thin film cavity model, and the mathematics of the model is reviewed as follows.

In previous work, based on the first principle of AI [19], we proposed neural partial differential equations (NPDE) under quasi-linear approximation[20]:

$$\begin{aligned} \frac{\partial \Psi}{\partial t} &= \mathcal{O}_L \Psi, \qquad \text{where} \\ \mathcal{O}_L \Psi &= \nabla \cdot (A(\Psi)\nabla\Psi) + B(\Psi)\nabla\Psi + C(\Psi). \end{aligned} \tag{9}$$

Where Ψ is used to stand for a system state, $\mathcal{O}_L$ is an elliptic operator, and ∇ is the nabla operator.

In wave optics, complex wave function has the form:

$$\Psi(\mathbf{r}, t) = A(\mathbf{r}) \exp[\mathtt{i}\varphi(\mathbf{r})] \exp(\mathtt{i}2\pi\nu t), \tag{10}$$

where $\mathtt{i}$ is imaginary index. Under slowly varying envelope approximation (SVEA), $A(\mathbf{r})$ is adopted to express a complex amplitude,

$$\Psi(\mathbf{r}) = A(\mathbf{r}) \exp(\mathtt{i}kz). \tag{11}$$

A PDE for the complex envelope $A(\mathbf{r})$ is

$$\nabla_T^2 A - \mathtt{i}2k\frac{\partial A}{\partial z} = 0, \tag{12}$$

where $\nabla_T^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse Laplacian operator. This is the SVEA of the Helmholtz equation. It is simply called paraxial Helmholtz equation [21]. It bears some similarity to the Schrödinger equation of quantum physics. This equation is a hyperbolic PDE, which also related heat equation by Wick rotation.

When using the finite differential method (along with finite elements method) to discretize above PDEs, we can the energy configuration of the RBM [20].

From above, we can see that current RBM learning algorithms only are suitable for computer implementation. In this work, we intend to provide the interpretation about RBM from optics theory perspective, which may benefit the implementation of the optical neural network – a specific class of the PAI systems. Following we discuss the basic optical element – thin-film filter.

III. Interpreting RBM from Optics Theory Respective

Our proposed model can be explained as a cavity-like model, which consists of layered thin-film filters.

A. Thin-film optics

In Fig. (1), we model the system as wave transmitting through a nonlinear dispersive medium with electric permittivity κ_i. The following notations are used for our Fabry-Perot (FP) cavity model. The φ_i^+ and φ_i^- stand for forward and backward wave amplitude at layer i, respectively. This configuration is of a FP optical cavity (also called multiple beam interference filter) type, and input/output layers are equivalent to mirrors.

A wave function with harmonic time dependence is used to represent a monochromatic wave.

An optical wave is described mathematically by a real function of time t and position $\mathbf{r} = (x, y, z)$, $u(r, t)$ is known as the wave function.

$$u(\mathbf{r}, t) = \varphi(\mathbf{r}) cos[2\pi\nu t + \psi(\mathbf{r})],$$

where $\varphi(\mathbf{r})$ is amplitude, $\psi(\mathbf{r})$ is phase, and ν is frequency.

The optical intensity $I(\mathbf{r}, t)$, is proportional to the average of the squared wave function

$$I(\mathbf{r}, t) = 2\left\langle u^2(r, t)\right\rangle$$

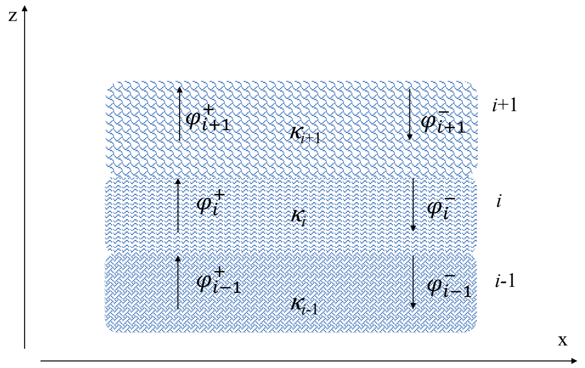

Fig. 1: A schematic diagram for wave propagation in a cavity notations. Fabry-Perot resonator type structure is designed, $(i-1)$th and $(i+1)$th layers are equivalent to two mirrors, respectively.

The $\langle\cdot\rangle$ means average.

Complex wave function has the form:

$$\varphi(\mathbf{r}, t) = U(\mathbf{r}) \exp[\mathtt{i}\psi(r)] \exp(\mathtt{i}2\pi\nu t), \tag{13}$$

so that

$$u(\mathbf{r}, t) = \text{Re}\{\varphi(\mathbf{r}, t)\} = \frac{1}{2}\left[\varphi(\mathbf{r}, t) + \varphi^*(\mathbf{r}, t)\right] \tag{14}$$

The wave is expressed as complex amplitude the time-independent factor $\varphi(\mathbf{r}) = U(\mathbf{r}) \exp[\mathtt{i}\psi(r)]$ referred to as the complex amplitude of the wave.

When wave propagates along z direction passing through the medium, the phase change δ_i is

$$\delta_i = (2\pi/\lambda) N_i \Delta z, \tag{15}$$

where N_i is the ith layer complex refractive index, $\kappa_i = N_i^2$. $N_i = n_i + \mathtt{i}f_i$, n_i is the real part of the refractive index, f_i is the extinction coefficient, with $f_i = \lambda\alpha_i/(4\pi n_i)$. α_i is absorption coefficient, and λ stands for wavelength.

We use k to express the wave number, $k = 2\pi/\lambda$, t_i and r_i stands for transmittance and reflectance coefficient at two media interface, respectively.

$$t_i = \frac{2n_{i-1}}{n_{i-1} + n_i}, \quad \text{and} \quad r_i = \frac{n_{i-1} - n_i}{n_{i-1} + n_i}.$$

In our two-dimensional medium model, the optical intensity is

$$I(x, z_i) = \|\varphi_i^+ + \varphi_i^-\|^2.$$

The amplitude of the electric field in the medium layers can be calculated by using the following formula (a time dependence of $\exp(\mathtt{i}\omega t)$ is assumed, $\omega = 2\pi\nu$ is angle frequency) [22]:

$$\begin{pmatrix} \varphi_i^+ \\ \varphi_i^- \end{pmatrix} = \frac{1}{t_{i+1}} \begin{pmatrix} e^{\mathtt{i}\delta_{i+1}} & r_{i+1}e^{-\mathtt{i}\delta_{i+1}} \\ r_{i+1}e^{\mathtt{i}\delta_{i+1}} & e^{-\mathtt{i}\delta_{i+1}} \end{pmatrix} \begin{pmatrix} \varphi_{i+1}^+ \\ \varphi_{i+1}^- \end{pmatrix}. \tag{16}$$

As we stated in the previous paragraph, φ_i^+, φ_i^- represent the amplitude of the electric field at point i propagating in the positive and negative z direction respectively, $\delta_i = (2\pi/\lambda_0)N_i\Delta z$, Δz is the increment along the z axis. N_i is a complex refractive index. And

$$E(x, z_i) \propto I(x, z_i) = \|\varphi_i^+ + \varphi_i^-\|^2,$$

is optical energy in the medium position (x, z_i).

In Eqs. (16), $U(r) = \varphi^+ + \varphi^-$.

$$\begin{aligned} E(z_i) &= U(z)^*U(z) = (\varphi_i^+ + \varphi_i^-)^*(\varphi_i^+ + \varphi_i^-) \\ &= (\varphi_i^+)^*(\varphi_i^+) + (\varphi_i^+)^*(\varphi_i^-) \\ &+ (\varphi_i^-)^*(\varphi_i^+) + (\varphi_i^-)^*(\varphi_i^-) \end{aligned} \tag{17}$$

After some mathematics work, we can write energy as

$$\begin{aligned} E(z_i) &= \left(\frac{r_i+1}{t_{i+1}}\right)^2 \left[(\varphi_i^+ - r_i\varphi_i^-)^2 + (r_i\varphi_i^+ - \varphi_i^-)^2\right. \\ &+ \left.\left((1+r_i^2)\varphi_i^+\varphi_i^- - r_i\left[(\varphi_i^+)^2 + (\varphi_i^-)^2\right]\right)\cos(2\delta_{i+1})\right] \end{aligned} \tag{18}$$

Using small phase variance approximation,

$$\cos(2\delta) \approx 1 + 2\delta \tag{19}$$

Eq. (18) can be written in the following form with Eq.(19) approximation:

$$\begin{aligned} E(z_i) &= \left(\frac{r_i+1}{t_{i+1}}\right)^2 \left(\left[1 - r_i - 2r_i\delta_{i+1} + r_i^2\right](\varphi_i^+)^2\right. \\ &+ \left[1 - r_i - 2r_i\delta_{i+1} + r_i^2\right](\varphi_i^-)^2 \\ &+ \left.\left[(1+r_i^2) - 4r_i + 2\delta_{i+1}(1+r_i^2)\right]\varphi_i^+\varphi_i^-\right) \end{aligned} \tag{20}$$

We can write Eq. (20) in the form

$$E(z_i) = a\varphi_i^+ + b\varphi_i^- + \varphi_i^+ W \varphi_i^-, \tag{21}$$

where

$$a = \left(\frac{r_i+1}{t_{i+1}}\right)^2 \left[1 - r_i - 2r_i\delta_{i+1} + r_i^2\right]\varphi_i^+, \tag{22}$$

$$b = \left(\frac{r_i+1}{t_{i+1}}\right)^2 \left[1 - r_i - 2r_i\delta_{i+1} + r_i^2\right]\varphi_i^-, \tag{23}$$

and

$$W = \left(\frac{r_i+1}{t_{i+1}}\right)^2 \left[(1+r_i^2) - 4r_i + 2\delta_{i+1}(1+r_i^2)\right]. \tag{24}$$

Please note that above energy equation is derived under plane wave condition, when we assume the medium is nonlinear, and optical intensity (energy distribution) depends on (x, y). Then energy in this i layer should be

$$\begin{aligned} E_i = \int &\left[a_i(x,y)\varphi_i^+(x,y) + b_i(x,y)\varphi_i^-(x,y)\right. \\ &\left.+\varphi_i^+(x,y)W(x,y)\varphi_i^-(x,y)\right] dxdy. \end{aligned} \tag{25}$$

With two dimensional approximation, and quantifying integral into summation, omit the subscript i we have

$$E = \sum_j a_j\varphi_j^+ + \sum_k b_k\varphi_k^- + \sum_{j,k}\varphi_j^+ W_{jk}\varphi_k^-. \tag{26}$$

In the above equation, the first term is named forward energy, the second term is named backward energy, and the third term we call it as cross energy, it is generated by optical wave interference.

From above, we can know that Eq. (26) has the same form as energy expression for RBM.

Based on the classical wave optics theory, we only derive the optical energy in the cavity have the form with that of RBM, following we discuss the energy distribution from quantum optics perspective.

B. The Quantum Theory of Light

In classical optics theory, as $\varphi(\mathbf{r})$ in Eq. (13) can have any magnitude, the field energy given by Eq. (25) can have any positive value. However, when we suppose the field $U(\mathbf{r}, t)$ satisfies the harmonic oscillator equation, its energy takes only the discrete values if this oscillator is treated quantum mechanically rather than classically,

$$E(n) = \left(n + \frac{1}{2}\right)\hbar\omega, \quad n = 0, 1, 2, 3, \ldots \tag{27}$$

where $\hbar$ is reduced Planck constant, ω is the angular frequency of electromagnetic wave, and $(E(0) = 1/2\hbar\omega)$ is ground-state energy.

According to the quantum theory of light [23], the the electromagnetic radiative energy can be expressed as:

$$E_R(n) = \left(n + \frac{1}{2}\right)\hbar\omega. \tag{28}$$

Here we can see that a quantum harmonic oscillator with each mode of the field is thus the association of the essence of the quantum theory of the radiation field.

1) Quantization of the field energy: The probability $P(n)$ that the mode oscillator is thermally excited to its n-th excited state in thermal equilibrium at temperature T, is given by the usual Boltzmann factor,

$$P(n) = \frac{\exp(-E(n)/k_BT)}{\sum_n \exp(-E(n)/k_BT)}, \tag{29}$$

where k_B is Boltzmann constant.

The mean number $\langle n \rangle$ of photons excited in the field mode at temperature T is therefore

$$\langle n \rangle = \frac{1}{\exp(\hbar\omega/k_BT) - 1}.$$

It is assumed above mean number $\langle n \rangle$ is position dependent, so energy distribution in our model obey Boltzmann distribution.

2) Fluctuations in photon number: RBM is considered as a generative stochastic neural network also, here we explain the randomness of energy with photon number fluctuations in the cavity.

The fluctuations of numbers of photons in each mode of the radiation field in the cavity are caused by the occurrence of photon absorption and emission processes. On a characteristic time scale the fluctuations take place, but without any knowledge of the time scales involved, some average properties of the fluctuations can be deduced. The characteristic time scale of the fluctuations must clearly be much shorter than the period over which the time average is taken. The cavity model is regarded as a quantum well, when photon numbers fluctuate over a certain threshold value, the "switch" of cavity border will open (optical bistability element) and photons will escape from the quantum well. This can be explained at this position, "neuron" is active as that in RBM. When photon numbers fluctuate below a certain threshold value, photons still trap in quantum well, this state is explained as "neuron" is inactive as that in RBM.

When make use of the *ergodic* property of the fluctuations in thermal light [24], we have the probability $P(n)$

$$P(n) = \frac{\langle n \rangle^n}{(1 + \langle n \rangle)^{1+n}}. \tag{30}$$

It is known that photons are Gauge Bosons, while in quantum statistics, Bosons are non-interacting, indistinguishable particles, obeying Bose-Einstein statistics. In the limit of high temperature, the Bose - Einstein distribution approaches the Maxwell - Boltzmann distribution.

In a canonical ensemble, for any state s of the system, the Maxwell - Boltzmann distribution can be expressed as

$$P(s) = \frac{1}{Z} e^{-E(s)/kT}, \quad \text{and} \quad Z = \sum_s e^{-E(s)/kT}, \tag{31}$$

where the index s runs through all micro-states of the system.

The Eq. (3) has the same form with that of Eqs. (29) and (31), in fact, the name of RBM is from Boltzmann distribution in statistical physics.

Here we present the explanation of RBM from optics theory perspectives, this pave the way to realize optical learning neural networks. However, to implement all optical learning neural network is very hard task because of restriction of off-the shelf optical devices. Next section we will present our consideration about optical learning RBM.

IV. Elucidate the Design of OLNN

An optical neural network (ONN) is a physical implementation of an artificial neural network (ANN) with optical components. Early ONN used a photorefractive Volume hologram to interconnect arrays of input neurons to arrays of output, while the synaptic weights are in proportion to the multiplexed hologram's strength.

All-optical deep neural network is very promising for PAI systems, it may find applications in all-optical image analysis, feature detection, and object classification. From the literature, we know that some ANN that have been implemented as optical neural networks, for example, Kohonen self-organizing map with liquid crystal spatial light modulators[25]. Recently, the deep diffractive optical neural network (D2NN) has been implemented by Lin *et al* [26]. In their report, all-optical deep learning framework can perform, at the speed of light, various complex functions that computer-based neural networks can execute. They also admire that the inference and prediction mechanism of the physical network is all optical, but the learning part that leads to its design is done through a computer. They implemented D2NN design using TensorFlow framework. And TensorFlow based design of a D2NN architecture took approximately 8 hours and 10 hours to train for the classifier and the lens networks, respectively [26].

From the fact described above, it is known that all-optical learning network is very hard to be implemented. How to realize an OLNN all-optical is a very challengeable work, following we will present our consideration.

A. Design of OLNN

By analyzing popular learning algorithms for RBMs, we can know that to implement optical RBM with traditional optical components is very difficult, and ideally requires advanced photonic materials.

With our theory and model, the bias and weights are formed with optical wave interference and diffraction, this gives us a quite different viewpoint for RBM learning. By analyzing Eqs. (22) and (23) for bias, while weights is expressed with Eq. (24), we can implement OLNN with proposed FP resonator cavity model.

When input field vector is encoded with a spatial light modulator (SLM) [27], we obtain input φ_i^+ , which carry on information of visible vector $\mathbf{v}$. When light incite into FP resonator, the diffraction and interference occurred of optical wave. The output layer plays a role of mirror, which will reflect the optical wave back to generate backward field vector φ_i^- (proportional to hidden vector $\mathbf{h}$). At initial stage optical energy distribution is unstable, photons will oscillate in FP resonator, this is considered as learning processing. Finally, photons in FP resonator will approach a steady state of distribution with time elapsed. As long as FP resonator output approach a stead state (it is similar with Laser output), the learning processing is finished. In this time, volume hologram materials filled in the cavity should be used to record bias and wights information.

Following we will present the optical components needed to implement OLNN, while details of optical circuit design for OLNN will presented in another paper.

B. Optical Components

1) Multi-layers thin-film interference filters: The main optical component for OLNN we considered is FP resonator, which may be fabricated by designing a multi-layers thin-film interference filter (MLTFIF).

If we replace mirrors with FP resonators, which is equivalent to a sigmoid type activation function. With this configuration, the functions of RBM learning can be realized.

Furthermore, if we intend to implement DBN, (stacked RBMs), multi-cavity systems may be a candidate. However, multi-cavity systems have very complicated input-output characteristic, even for a nonlinear three-mirror FP multistability element [28], inverse design should be utilized for OLNN.

2) Spatial Light modulators: In optical computing, SLMs have been become most used component, since it is an object that can make some form of spatially varying modulation on a beam of light [27]. We can use SLMs to generate input vectors of an OLNN, and/or used at inference stage of ONN as part of

3) Volume Hologram: Volume holograms can be used for recording OLNN parameters. In the case diffraction of light from the hologram is possible only as Bragg diffraction, after learnt, D2NN/OLNN can performance inference or prediction by diffraction of light through the hologram.

4) Random Phase Plates: Random phase plates (RPP) [29][30] can be utilized to generate random phase when optical wave φ_i^- is reflected from output mirror. They are not necessary for deterministic learning of OLNN, however, if we intend to realize stochastic learning procedure, RPPs can be placed in front of hidden layer mirror to make that OLNN with a random learning configuration becomes possible.

V. Summary

In this work, we interpret the RBM from optics theories, which is an alternative perspective relative to statistical physics perspective. The energy distribution in FP cavity is derived, and relationship connecting with RBM is established. Based on our derived bias and weights formula of RBM, we designed a FP filter configuration, which can provide learning process and make the implementation of optical learning neural network all-optical become possible. In addition, off-the-shelf optical components for constructing an OLNN are presented.

In the future research work, we will develop OTDMs further, especially the Laser resonator theory by solving paraxial Helmholtz equation. And we will investigate inverse design method for materials engineering, explore OLNN implementation for practical applications in image analysis, feature detection, and object classification.

References

[1] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, pp. 504–507, July 2006.

[2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[3] D. Castelvecchi, “Can we open the black box of AI?” *Nature*, vol. 538, no. 7623, pp. 20–23, 2016.

[4] A. B. Arrieta, N. D. Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI,” *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[5] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, “Definitions, methods, and applications in interpretable machine learning,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 44, pp. 22 071–22 080, 2019.

[6] A. Miriyev and M. Kovač, “Skills for physical artificial intelligence,” *Nature Machine Intelligence*, vol. 2, no. 11, pp. 658–660, 2020.

[7] P. Guo and Q. Yin, “Synergetic learning systems (I): Concept, architecture, and algorithms,” presentation at the third China Systems Science Conference (CSSC2019), Changsha, May 18-19, 2019. [arXiv], https://arxiv.org/abs/2006.06367

[8] G. E. Hinton, “Deep belief networks,” *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[9] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” in *Advances in Neural Information Processing Systems 19, 2006*.

[10] P. Guo, “Synergetic learning systems (III): Automatic organization and evolution theory of neural network architecture,” Presentation at the Fourth China Systems Science Conference (CSSC2020), QingDao, September 19-20, 2020. [researchgate.net, September 2020, https://doi.org/10.13140/RG.2.2.23186.07360]

[11] P. Guo, “Synergetic learning systems (II): Interpretable neural network model with statistical physics approach,” Preprint, researchgate.net, May 2019, the Fifth National Statistical Physics & Complex Systems Conference (SPCSC 2019), Hefei, July 26-29, 2019.

[12] M. A. Cueto, J. Morton, and B. Sturmfels, “Geometry of the Restricted Boltzmann Machine,” arXiv:0908.4425 [stat.ML], 2010.

[13] R. Karakida, M. Okada, and S. Amari, “Dynamical analysis of contrastive divergence learning: Restricted Boltzmann machines with Gaussian visible units,” *Neural Networks*, vol. 79, pp. 78–87, 2016.

[14] G. E. Hinton, “A Practical Guide to Training Restricted Boltzmann Machines,” in *Neural Networks: Tricks of the Trade - Second Edition*, ser. Lecture Notes in Computer Science, G. Montavon, G. B. Orr, and K. Müller, Eds. Springer, 2012, vol. 7700, pp. 599–619.

[15] G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[16] A. Decelle, G. Fissore, and C. Furtlehner, “Spectral dynamics of learning in restricted boltzmann machines,” *EPL (Europhysics Letters)*, vol. 119, no. 6, p. 60001, sep 2017.

[17] A. Fischer and C. Igel, “Training restricted Boltzmann machines: An introduction,” *Pattern Recognition*, vol. 47, no. 1, pp. 25–39, 2014.

[18] M. Gabrié, E. W. Tramel, and F. Krzakala, “Training Restricted Boltzmann Machines via the Thouless-Anderson-Palmer Free Energy,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’15. Cambridge, MA, USA: MIT Press, 2015, p. 640–648.

[19] P. Guo, “What is the First Principles for Artificial Intelligence (ver. 2),” *CCCF (Communications of the China Computer Federation)*, vol. 16, no. 10, pp. 53–58, Oct. 2020. (in Chinese).

[20] P. Guo, K. Huang, and Z. Xu, “Partial Differential Equations is All You Need for Generating Neural Architectures,” preprint, researchgate.net, January 2021, [arXiv:2103.08313].

[21] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*, 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc, 2019.

[22] H. A. MacLeod, *Thin-Film Optical Filters*, 4th ed., ser. Series in Optics and Optoelectronics. Los Angeles: CRC Press, 2010.

[23] R. Loudon, *The quantum theory of light*, 3rd ed. Oxford: Oxford University Press, 2000.

[24] J. W. Goodman, *Statistical Optics*. New York: John Wiley & Sons, 2015, wiley Series in Pure and Applied Optics.

[25] J. Duvillier, M. Killinger, K. Heggarty, K. Yao, and J. L. de Bougrenet de la Tocnaye, “All-optical implementation of a self-organizing map: a preliminary approach,” *Applied Optics*, vol. 33, no. 2, pp. 258–266, Jan 1994.

[26] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[27] A. D. Chandra and A. Banerjee, “Rapid phase calibration of a spatial light modulator using novel phase masks and optimization of its efficiency using an iterative algorithm,” *Journal of Modern Optics*, vol. 67, no. 7, pp. 628–637, 2020.

[28] P. Guo, Y.-G. Sun, J. Xiong, W. Wang, and Z. Jiang, *Nonlinear three-mirror Fabry-Perot multistability element*. Singapore: World Scientific, 1987, pp. 83–87.

[29] P. F. Almoro and S. G. Hanson, “Random phase plate for wavefront sensing via phase retrieval and a volume speckle field,” *Appl. Opt.*, vol. 47, no. 16, pp. 2979–2987, Jun 2008.

[30] H. A. Rose and D. F. DuBois, “Statistical properties of laser hot spots produced by a random phase plate,” *Physics of Fluids B: Plasma Physics*, vol. 5, no. 2, pp. 590–596, 02 1993.