

光学学报

光学非线性激活函数器件的原理与应用

吕青鸿¹, 马睿¹, 肖莘宇², 俞维嘉³, 刘知非³, 胡小永^{1,4*}, 龚旗煌^{1,4}

¹北京大学物理学院人工微结构和介观物理国家重点实验室, 北京 100871;

²南开大学物理科学学院, 天津 300071;

³北京工业大学理学部信息光子技术研究所, 北京 100124;

⁴山西大学极端光学协同创新中心, 山西 太原 030006

摘要 神经网络中的非线性激活层可以改变多层网络数据间的线性变换关系,使神经网络得以进行更复杂的学习。为实现处理速度更快,能耗更低的运算,近年来光子领域的神经网络逐渐受到重视,一系列光学非线性激活函数器件应运而生。本文综述了近年来在光学神经网络中引入非线性激活函数的工作,从光学非线性函数的物理机制及其在光学神经网络中的应用出发,对该领域的工作进行了回顾;总结并讨论了光学神经网络中光学非线性激活函数器件发展所面临的挑战及变化趋势,并基于此展望了其发展前景。

关键词 非线性光学; 光学神经网络; 非线性激活函数

中图分类号 O436 **文献标志码** A

DOI: 10.3788/AOS230903

1 引言

非线性函数是一种常用的将输入转换为输出的数学函数,其特征在于输出与输入并不是线性关系,而是在函数图像上呈曲线或折线。在自然科学的范畴内,描述万物变化的函数是多种多样的,不一而足,而他们往往是非线性的。基于这一特性,非线性函数可用于描述大量不同的系统,例如电力系统、光学系统、经济系统、生物系统、计算机网络和通信系统。利用非线性函数这样一种强大的数学工具,能够描述不同类型的复杂系统,并对这些系统进行建模,从而更好地了解它们的运行方式并加以控制,进而针对系统进行有效的优化。

1961年,在红宝石激光器问世不久, Franken等^[1]将红宝石激光器输出的波长为 694 nm 的激光穿过一个石英晶体时,产生波长为 374 nm 的紫外光。这是最早观察到的光学二次谐波(SHG)现象,标志着非线性光学的诞生。在经典光学盛行之时,科学家研究光在介质中的传播都属于弱光束范畴,在这时的研究中,干涉、衍射、偏振等现象均满足线性叠加原理,这样的光学研究被称为线性光学。随着激光的出现,强光与介质的相互作用产生了多种新现象,而这些现象往往需要用非线性函数描述,我们称之为非线性光学。随着光学神经网络的发展,人们发现在许多研究中都需要引入光学中的非线性关系,这赋予了光学非线性新的

意义。现在,由于光学神经网络中对于非线性激活函数的迫切需要,在实际的物理体系中得到这种非线性激活函数的研究受到了广泛关注。

在光神经网络或光计算中引入光学非线性主要有 4 种方式^[2]:第一种是利用非线性材料,例如利用非线性材料中激光的饱和和吸收函数。当激光入射到介质时,介质的吸收系数随介质中光强的增加而减小;当输入光波强度超过阈值时,介质吸收特性达到饱和。根据激光增益饱和和效应这一特性,利用光衰减放大器、半导体光放大器、饱和吸收体等光电器件,可以获得光学非线性激活函数。第二种是利用光参量经过系统后输入与输出的非线性关系,例如光学双稳态,即一个输入光强可以输出两个相互转换的稳定光强,输入与输出光强的关系为类似磁滞回线的非线性激活函数。利用非线性效应、反馈系统、入射光能三者的关系,可以制造出光学双稳器件(OBD)。第三种是利用高阶光学非线性效应,如克尔效应^[3-4],即光进入非线性介质,将引起折射率变化,进而产生相位偏移。由三阶非线性效应所引入的非线性函数可以用于量子光学计算等领域。第四种是利用光学系统与其他物理系统的转换引入非线性函数,例如可以在相变材料中^[5]利用材料相变过程吸收或释放大量热,从而改变材料的光学性质,引入光学非线性。电磁诱导^[6]、电光调制^[7]等研究中利用光电信息转换引入光学非线性。

光学神经网络的非线性激活函数有 3 个重要指

收稿日期: 2023-05-04; 修回日期: 2023-06-27; 录用日期: 2023-06-27; 网络首发日期: 2023-07-20

通信作者: *xiaoyonghu@pku.edu.cn

标,分别是响应速度、阈值以及损耗。响应速度是指非线性激活函数对输入信号的响应时间。阈值决定了神经元的激活状态,阈值过高,意味着光经过多层神经网络后,其强度将会持续减弱,而阈值数值会根据具体的应用需求和网络结构进行调整。损耗是指输入信号在经过非线性激活函数时的能量损失。在设计网络时,需要考虑损耗对网络性能的影响,并通过优化光学器件的结构和材料来降低损耗。

基于以上背景,本文综述了近年来在光学神经网络中引入非线性激活函数的方法及其应用,旨在整理现有的光学非线性激活函数,以促进光学神经网络的发展。首先,总结了常用的光学非线性线型;其次,介绍了在光-电-光(O-E-O)以及全光体系中引入的各种光学非线性物理机制。对于O-E-O体系,主要介绍电吸收调制器(EAM)、电光调制器(EOM)等引入光学非线性激活函数的方法;对于全光体系,主要介绍定制

化材料、半导体光放大器(SOA)以及光学微环谐振器(MRR)引入光学非线性神经网络的原理。然后,介绍了光学非线性在模式识别、分类、可重构器件,以及实时响应与高效信息处理等方面的应用。最后,总结并讨论了光学神经网络中光学非线性激活函数相关器件发展所面临的挑战及变化趋势,并基于此展望了其发展前景。

2 非线性激活函数的种类

非线性激活函数作为神经网络中不可或缺的模块,已经发展出很多适用于不同情形的线型。图1展示了一些常用的线型,包括Sigmoid、ReLU、Elu、leaky ReLU、Radial Basis、Softplus等。其中,Sigmoid可以被用在Hopfield递归网络中,ReLU可以被用在卷积神经网络中,Radial Basis可以被用在支持向量机中,而Softplus则可以被用在深度学习中^[8]。

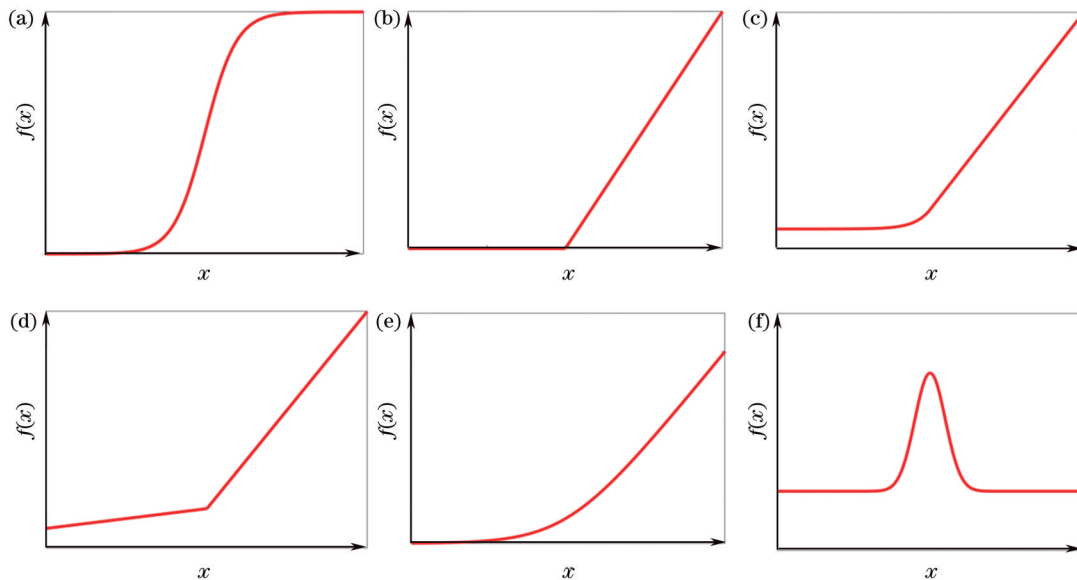


图1 常用的非线性激活函数。(a) Sigmoid;(b) ReLU;(c) Elu;(d) leaky ReLU;(e) Softplus;(f) Radial Basis

Fig. 1 Common nonlinear activation functions. (a) Sigmoid; (b) ReLU; (c) Elu; (d) leaky ReLU; (e) Softplus; (f) Radial Basis

上述非线性激活函数在中央处理器(CPU)这类电学系统中非常容易实现,但是在光学体系中实现这些非线性函数仍面临挑战。进一步地,以更低的阈值、更低的能耗,以及更快的响应速率实现这些非线性激活函数,成为光学神经网络发展中面临的巨大挑战。

3 光学非线性激活函数的实现

实现非线性激活函数的光学体系主要分为两类:一类是O-E-O体系;另一类是全光体系。在介绍了O-E-O体系后,本节将对全光非线性激活函数器件的设计原理进行介绍。该类器件的实现途径可分为定制化材料、SOA,以及光学微环谐振器(MRR)等。

3.1 O-E-O体系的光学非线性激活函数

在O-E-O调制系统中,电吸收调制器(EAM)可

以用来直接调控信号的振幅,而电光调制器(EOM)则通过马赫-曾德尔干涉仪(MZI)或者MRR中的相移来调制振幅^[9]。上述两个体系都可通过不同的方式实现非线性激活函数^[2],且各有优缺点:EOM是由场驱动而不是由载流子驱动的,所以其响应速度比EAM快;EAM由于不需要考虑干涉器件,故具有更小的体积,从而容易具有更高的集成度。

以EAM为例,该器件对光的吸收程度取决于施加在器件的偏压大小,比如:当偏压为0时,该器件可吸收更多的光;随着偏压增大,所吸收的光将变得更少^[9]。如图2(a)所示,不同的EAM体系所产生的线型也不一样。通过把出射的光信号输入光电二极管,便可进一步将该非线性响应关系转换为电压的非线性激活函数,如图2(b)所示。类似EAM器件还可以用

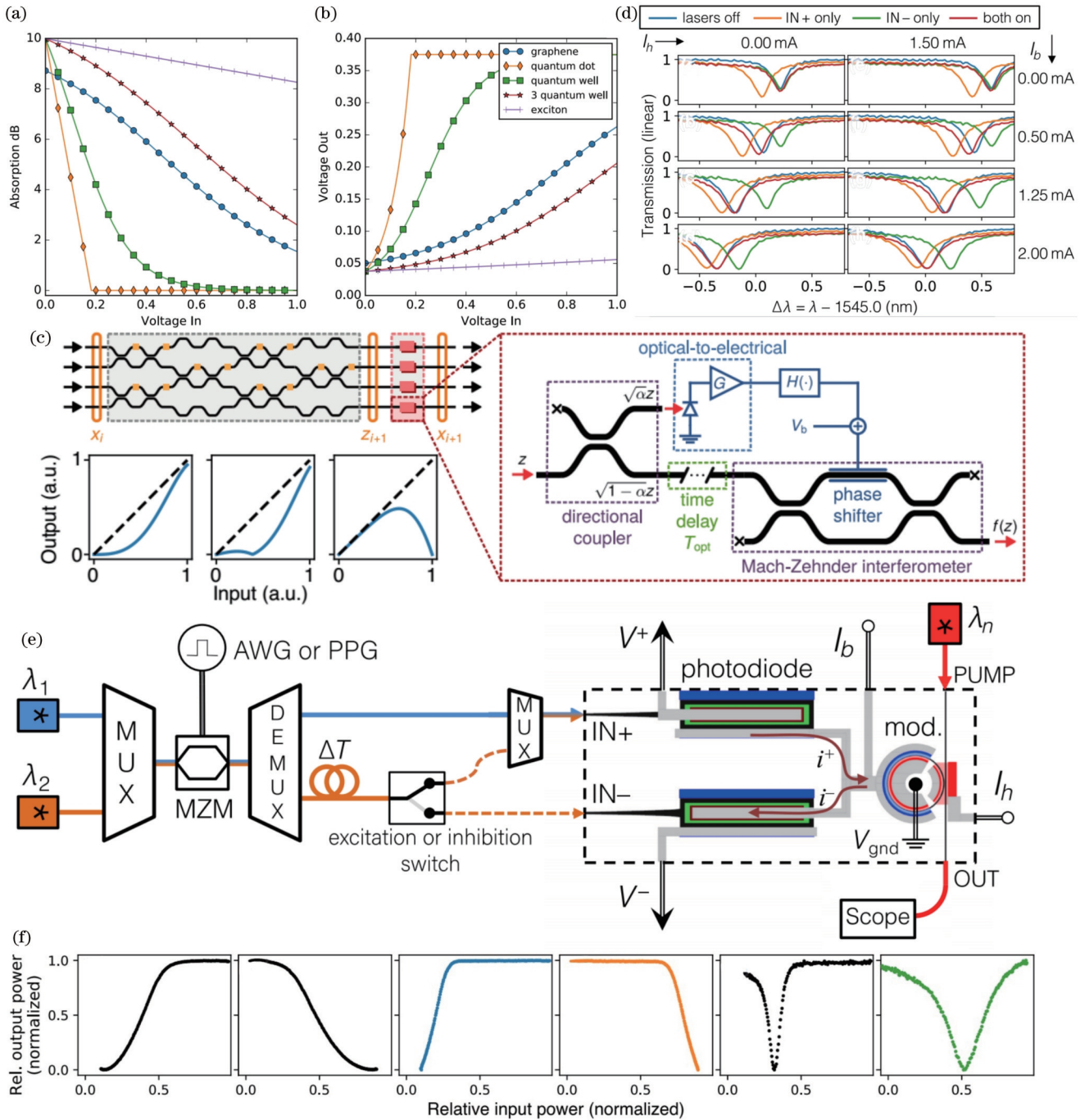


图2 O-E-O体系的非线性激活函数。(a)不同体系的吸收率随电压的变化关系^[9]；(b)接入光电二极管后的电压非线性激活函数^[10]；(c)EOM体系中实现非线性激活函数的方案与可形成的非线性函数^[11]；(d)MRR谐振峰随 I_h 、 I_b 和光电流的响应^[12]；(e)MRR调制器^[12]；(f)通过调节失谐量得到的非线性激活函数线型^[12]

Fig. 2 Nonlinear activation function of O-E-O system. (a) Absorptivity of different systems varied with voltage^[9]; (b) voltage nonlinear activation function after the photodiode is connected^[10]; (c) schematic of implementing a nonlinear activation function scheme in EOM system and the nonlinear activation functions that can be formed^[11]; (d) MRR resonance peak varied with I_h , I_b , and photocurrent under different permutations of illumination^[12]; (e) schematic of MRR^[12]; (f) nonlinear line type under different bias conditions^[12]

ITO实现^[10]。

在EOM体系中,非线性激活函数可以通过先将一小部分入射光转换为电信号,再用该电信号调制剩下光信号的强度来实现。其中,在对剩余光信号的调制过程中需要使用上述MZI机制来实现光强调节,其

示意图如图2(c)所示。此外,该体系可以实现类似于ReLU的非线性激活函数^[11]。该方案可以在最低功率阈值为0.1 mW的条件下工作。

除了上述方法外,还可以通过对MRR进行掺杂,使其共振峰对光电流、调制偏压 I_b 和加热偏压 I_h 敏感,

从而实现 O-E-O 调制^[12]。具体来说,如图 2(d)所示,当施加 I_h 后, MRR 谐振峰将会以强而缓慢的方式整体红移,而通过施加 I_b 可以将 MRR 谐振峰以弱而快的方式整体蓝移,同时还伴随着谐振峰深度增加和线型展宽。在此基础上,通过引入正向或者负向光电流,谐振峰还将相应蓝移或红移,如果同时引入正向和负向光电流,则不会产生明显移动。对应器件的形貌如图 2(e)所示。该体系所能实现的线型将与泵浦光和 MRR 谐振峰之间的失谐直接相关。通过调节失谐量,Prucnal 课题组^[12]得到了 Sigmoid、ReLU、Radial Basis 等线型,如图 2(f)所示。该器件的插入损耗可达 18 dB,器件的阈值则由二极管的阈值决定。

3.2 全光非线性激活函数的实现

本小节将从定制化材料、SOA、MRR 这 3 个方面介绍全光非线性激活函数的实现原理,并简述其效果。

3.2.1 定制化材料

通常情况下,由电子元件和光学元件组成的混合系统,可以实现非线性激活函数,但光信号和电信号之间的转换会导致信号失真,从而降低计算性能^[13]。定制化材料的全光非线性激活器是一个可行的解决方案。

典型的定制化材料有 5 种^[14],其核心原理是利用材料对波导的非线性行为和低耗散特性。当光信号通

过这种材料时,材料会产生非线性光学效应,从而可以使光信号的幅度或相位发生变化。这种变化可以用来模拟神经元中的激活函数,例如 ReLU、Sigmoid 等。

3.2.1.1 饱和吸收体

动态系统蓄水池计算机对信息的处理^[15]可描述为

$$x_i(n+1) = F_{NL} \left[\alpha \sum_{j=1}^N A_{ij} x_j(n) + \beta m_i u(n+1) \right], \quad (1)$$

式中: n 为当前节点序号; N 为系统内部节点数量; F_{NL} 为非线性函数; α 为反馈增益; β 为输入增益; A 为互连矩阵, A_{ij} 为互连矩阵的元素; m 为节点相关系数。每个节点 x_i 通过处理输入 u 进行迭代,最后通过复权 W_i 得到输出 y ,如图 3(d)所示。

非线性函数 F_{NL} 的功能可以通过半导体饱和吸收镜^[15](SESAM)实现,当光信号经过 SESAM 时,反射率发生变化,使得光信号具有非线性特性。SESAM 的结构如图 3(a)所示,其中 InGaAs 层是嵌入在微谐振器中的利用 MOCVD 方法生长的 InGaAs/InP 量子阱^[14,16-17],底部镜通常采用宽带高反射率金属基镜^[14]。SESAM 的非线性行为如图 3(b)所示,Dejonckheere 等^[15]通过对比 SOA 与 SESAM 的输入-输出关系的差异,说明 SESAM 不同于传统光电蓄水池的计算过程^[18-19]。

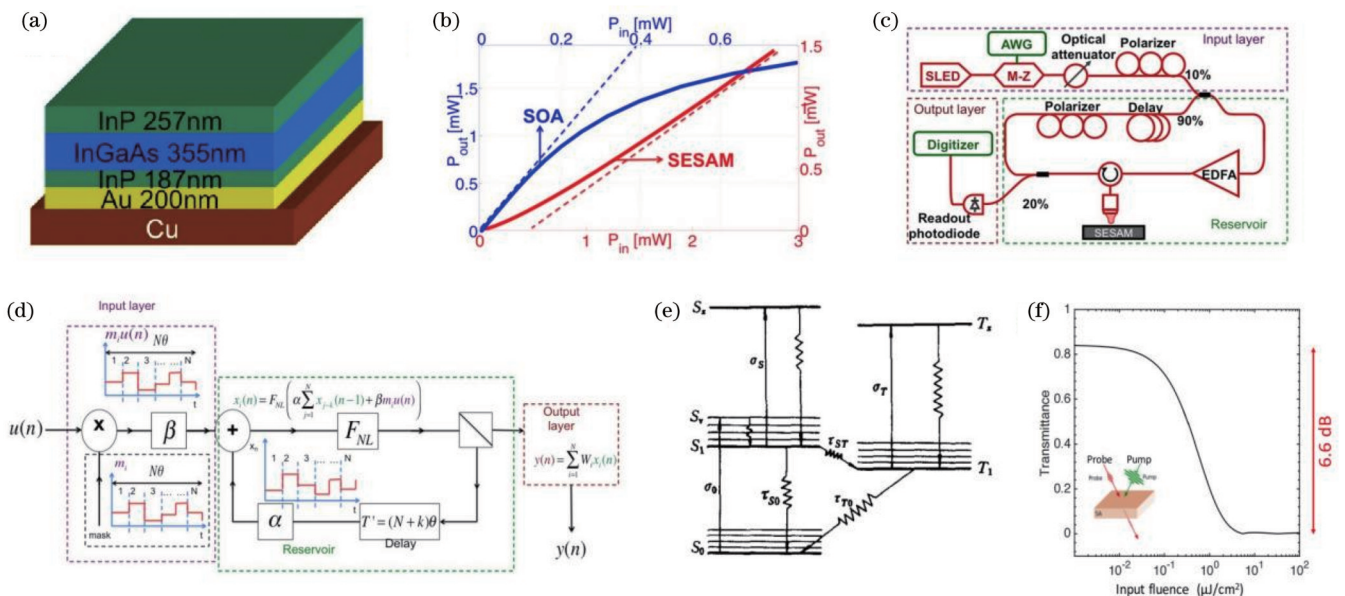


图 3 饱和吸收与反饱和吸收材料。(a) SESAM 的层状结构示意图^[15]; (b) SESAM 和 SOA 的非线性激活函数, SESAM 的非线性特征主要出现在低输入功率下,而 SOA 的非线性特征主要出现在高输入功率下^[15]; (c) 全光蓄水池系统示意图^[15]; (d) 动态系统蓄水池计算机原理图,其中每个输入信号的持续时间为 $T=N\theta$ ^[15]; (e) C_{60} 能级结构示意图^[20]; (f) 高浓度 C_{60} 作为反饱和吸收体在 PVA 主体薄膜中的光强密度与透过率关系^[21]

Fig. 3 Saturated absorption and reverse saturated absorption materials. (a) Schematic of layered structure of SESAM^[15]; (b) nonlinear activation function of SESAM and SOA, the nonlinear characteristics of SESAM are mainly reflected at low input power, while the nonlinear characteristics of SOA are mainly reflected at high input power^[15]; (c) schematic of all-optical reservoir system^[15]; (d) schematic of dynamic reservoir system, the duration of each input signal is $T=N\theta$ ^[15]; (e) schematic of C_{60} energy level structure^[20]; (f) relationship between light intensity density and transmittance of high concentration C_{60} as a desaturated absorber in PVA host film^[21]

图 3(c)为文献[15]中介绍的全光蓄水池系统的示意图,该系统由光输入信号驱动,经过 SLED 产生 40 nm 的宽频谱光,并将其输入到集成的光电调制器中,产生时间相关的输入光信号。输入信号注入环路,该环路中有掺铒光纤放大器(EDFA)、分向器、SESAM 和延迟线。20% 的信号被耦合器从回路中提取出来,用于记录蓄水池内状态的演变,并用数字转换器进行处理。同时,极化控制器可用于匹配不同信号的偏振状态,而放大器被放置在线性区域用于补偿环路中的光强损耗。

文献[16]提出一种内置两个吸收器的全光 2R 再生装置,该装置可对光信号进行再放大和再整形处理,利用两个吸收器在 1548 nm 处进行泵浦-探测,其响应时间分别为 1.5 ps 和 3.4 ps。该再生装置对非线性相位噪声的耐受性提高了 3 dB,当比特误码率为 10^{-4} 时,距离改进因子为 1.3。

3.2.1.2 反饱和吸收体

反饱和吸收(RSA)是指吸收特性随光强的增加而变强,一些有机化合物^[15]、有机金属配合物^[22-23]、金属纳米颗粒^[24]都具有这种特性。另外,一些金属簇的光学限制行为可能是由反饱和吸收机制引起的^[20]。将这种材料放置在光学传输线中,可以利用它的非线性特性将输入信号转换为非线性输出信号。

C_{60} 是由 60 个 π 电子组成的三维共轭体系,具有更强的三阶非线性,可以在较大的波段范围上表现出反饱和和吸收特性^[21]。图 3(e)为 C_{60} 的分子能级结构简图,反饱和吸收材料一般都具有单线态和三重态的能级结构^[25]。文献[21]采用高浓度的 C_{60} 反饱和吸收体在聚乙烯醇(PVA)薄膜中传导非线性激活函数,其效果如图 3(f)所示,与输入功率相关的非线性调制幅度约为 7 dB。

文献[21]探讨了两种独立的方法来实现基于纳米光学结构的光学感知器非线性激活函数,这些结构具有诱导透明性和反饱和和吸收特性。结果表明,对于所考虑的两个系统,全光非线性分别提供大约 3 dB 和 7 dB 的消光比,并得到 97% 和接近 100% 的示例 MNIST 任务的分类准确率。

3.2.1.3 基于电磁诱导透明的腔体

电磁诱导透明(EIT)是利用能级结构和电磁场的相互作用来消除某些频率的光的吸收,使得这些光通过介质时透明度增加^[6]。

图 4(a)所示为没有外加耦合下的能级排布,其中 $|1\rangle$ 为基态, $|2\rangle$ 为亚稳态, $|3\rangle$ 为激发态。图 4(b)为在外加调控激光下 $|2\rangle$ 与 $|3\rangle$ 耦合为二重态 $|a^+\rangle$ 的能级图^[26]。图 4(c)展现了实验测得的外加耦合对透过率的影响。

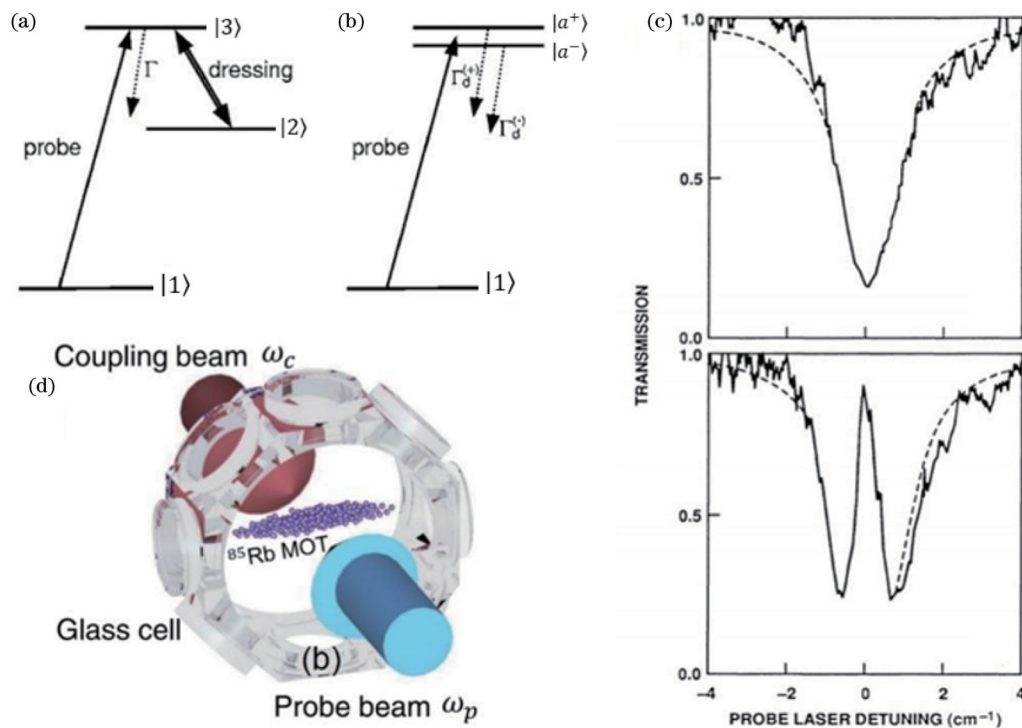


图 4 EIT 诱导的非线性激活函数。(a)不添加耦合场时 EIT 材料的能级排布^[26]; (b)添加耦合场时 EIT 材料的能级排布^[26]; (c)实验测得的 Sr 蒸气的 EIT 现象^[27]; (d)EIT 组件构造^[6]

Fig. 4 Nonlinear activation function caused by EIT. (a) Energy level arrangement of EIT material without coupling field^[26]; (b) energy level arrangement of EIT material with coupling field^[26]; (c) experimentally measured EIT phenomenon of Sr vapor^[27]; (d) EIT component construction^[6]

如图 4(d)所示,将 EIT 材料嵌入磁光阱(MOT),通过 MOT 上的探针输入耦合光束,则输出光强和输入光强的关系^[27]为

$$I_{p,out} = I_{p,in} \exp\left(-d_{OD} \frac{4\gamma_{12}\gamma_{13}}{\Omega_c^2 + 4\gamma_{12}\gamma_{13}}\right), \quad (2)$$

式中: $I_{p,out}$ 和 $I_{p,in}$ 分别为输出光强和输入光强; d_{OD} 为 $|1\rangle$ 到 $|3\rangle$ 的深度; γ_{ij} 为状态 $|i\rangle$ 到 $|j\rangle$ 的失相速率; Ω_c 为耦合场的 Rabi 频率。

3.2.1.4 基于光-物质作用的纳米材料

利用光-物质作用(LMI)设计光子集成电路,可以

解决光信号和电信号之间转换而产生的失真和延迟问题。文献[25]报道了一个由一对相距 10 nm 的金纳米棒组成的混合系统,这 10 nm 间隙中存在一个 CdSe 量子点,如图 5(a)所示。

如图 5(b)所示,量子组件的能量耗散与输入功率呈非线性关系,因此可以通过波导的透过率与输入功率建立非线性函数。此外,文献[15]还报道了将量子组件置于波导顶部和波导中部,可以得到不同的曲线,如图 5(c)所示。该量子组件可以通过两种表面活性剂分别附着金纳米棒和纳米球,通过表面活性剂间的选择性进行端部连接^[28],该过程如图 5(d)、(e)所示。

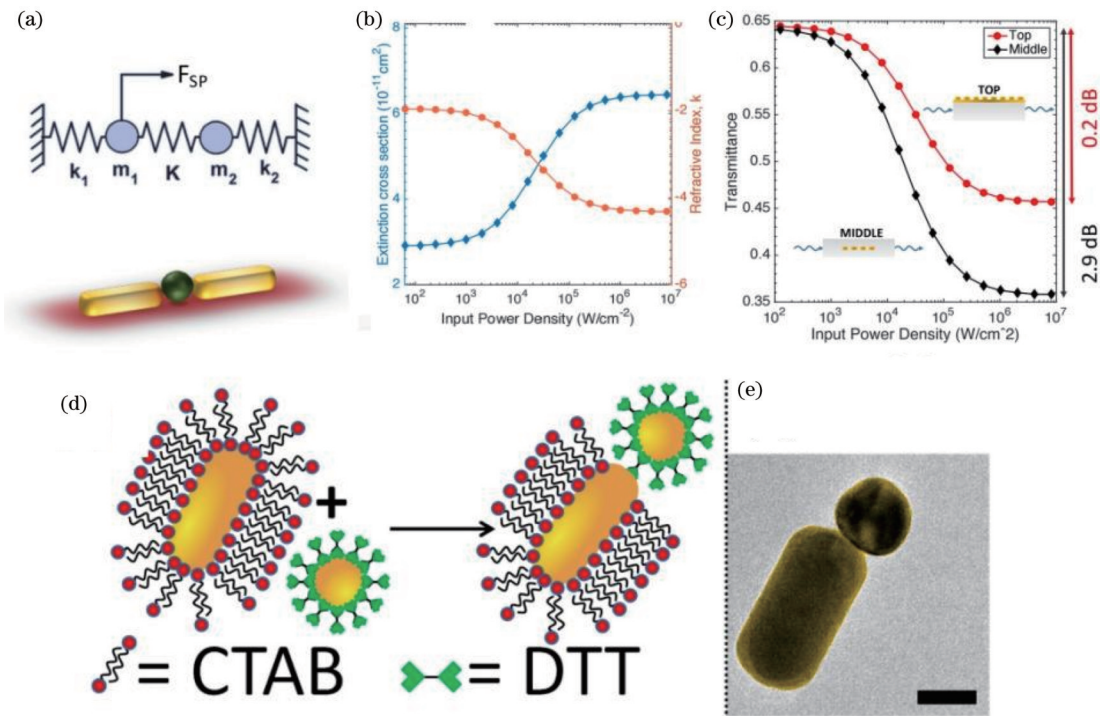


图 5 LMI 材料诱导的非线性激活函数。(a) LMI 材料结构示意图^[25]; (b) 纳米棒和量子点系统输入功率密度发生变化时,该系统的消光截面(左纵轴)和折射率的虚部(右纵轴)也会发生变化^[25]; (c) 在波导不同位置放置量子组件时,透射率与输入功率密度的非线性函数关系^[15]; (d) 表面活性剂对金纳米材料末端选择性的原理示意图; (e) 显微镜下纳米棒与纳米球二聚体图像^[28]

Fig. 5 Nonlinear activation function caused by LMI materials. (a) Schematic of LMI material structure^[25]; (b) when the input power density of nanorod and quantum dot systems changes, the extinction cross-section (left y-axis) and imaginary part of the refractive index (right y-axis) of the system also change^[25]; (c) nonlinear function of transmittance versus input power density when quantum components are placed at different locations on the waveguide^[15]; (d) schematic of end-selectivity of surfactants to gold nanomaterials; (e) microscopic image of nanorods and nanospheres dimer^[28]

3.2.1.5 相变材料

相变材料(PCM)具有可逆的相变特性,通常是晶态和非晶态的转变^[29],或者是金属相和介电相的转变。PCM 受到入射光的作用时,会发生相变并且改变其光学特性,这些光学特性的变化可以用来实现非线性光学功能^[30],目前主要采用的 PCM 是 VO₂ 和 Ge₂Sb₂Te₅ (GST)^[29-30]。此类 PCM 在类比光学神经网络上具有良好的应用前景。如图 6(a)所示,通过将 GST 和氮化硅波导与该全光回路相结合^[23],能以全光的形式实现神经突触。为了实现全光非线性激活函数,Feldmann

等^[5]直接将 PMC 引入图 6(c)所示的光学神经网络。如图 6(b)所示,在不同能量的脉冲激励下,嵌有 GST 的环形谐振腔能以 0.23 dB 的损耗和 430 pJ 左右的阈值呈现出 ReLU 非线性激活函数。这种现象可以理解当积累的脉冲能量超过某一阈值时,便会改变 GST 材料的性质,并引起激射。通过波分复用技术,Feldmann 等^[5]将光子神经网络继续扩展到了更多的节点,从而提高了其功能和效率。该网络利用波分复用技术将多个信息通道连接到同一光纤上,既实现了高速传输,又降低了光纤使用的成本。

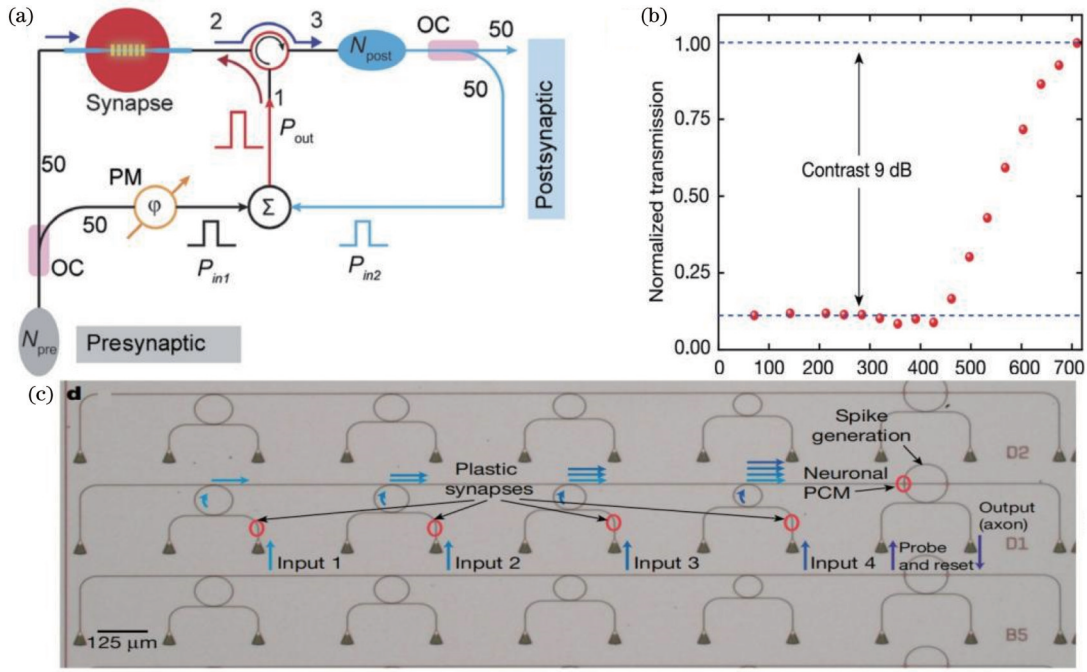


图 6 PCM 诱导的非线性激活函数。(a)通过波导上的 PCM 实现光学突触前后的非线性转化^[29];(b)由 PCM 实现的光学非线性激活函数^[5];(c)通过 PCM 对多组输入脉冲的非线性处理而实现的神经网络^[5]

Fig. 6 Nonlinear activation function caused by PCM. (a) Nonlinear conversion before and after optical synapses is achieved through PCM on the waveguide^[29]; (b) optical nonlinear activation function implemented by PCM^[5]; (c) neural network achieved by nonlinear processing of multiple sets of input pulses utilizing PCM^[5]

3.2.2 SOA 简介

与半导体激光器类似,SOA 以半导体为增益介质,但由于缺乏振荡腔结构,只能对一定功率的光信号进行放大。SOA 的非线性光学效应大致可分为三类:交互相位调制(XPM)、交叉增益调制(XGM)以及四波混频(FWM)^[31]。本文主要介绍 XPM 和 XGM 的原理及其在光计算领域的应用。

对于 XPM 效应,可将 SOA 的 z 位置瞬时增益系数 $g(z, t)$ 以及控制信号的功率 $P_C(z, t)$ 和相位漂移 $\varphi_C(z, t)$ 简化为

$$\frac{\partial g(z, t)}{\partial t} = \frac{g_0 - g(z, t)}{\tau_c} - \frac{g(z, t)}{1 + \epsilon P_C(z, t)} \frac{P_C(z, t)}{E_{SAT}}, \quad (3)$$

$$\frac{\partial P_C(z, t)}{\partial z} = \frac{g(z, t)}{1 + P_C(z, t)} P_C(z, t), \quad (4)$$

$$\frac{\partial \varphi_C(z, t)}{\partial z} = -\frac{1}{2} \alpha \frac{g(z, t)}{1 + E_C(z, t)}, \quad (5)$$

式中: g_0 为小信号增益; τ_c 为自发载流子寿命; α 为线宽增幅因子; ϵ 为由载流子发热和光谱烧孔效应造成的非线性增益压缩因子; E_{SAT} 为饱和能量。

$$E_{SAT} = P_{SAT} \cdot \tau_c = \frac{\hbar \omega_0 \tau_c}{a_N} \cdot \frac{dW}{\Gamma}, \quad (6)$$

式中: ω_0 为控制信号的中心波长; a_N 为增益因子; d 和 W 分别为 SOA 激活区的深度和宽度; Γ 为限制因子; \hbar 为约化普朗克常量。瞬时增益系数为

$$h(t) = \int_{-\infty}^{\infty} d\tau \int_{-L/2}^{L/2} g(z, \tau) \delta(\tau - t \pm zn_{SOA}/c) dz. \quad (7)$$

SOA 的增益 $G(t)$ 以及相位漂移 $\Delta\varphi(t)$ 可表示为

$$\begin{cases} G(t) = \exp[h(t)] \\ \Delta\varphi(t) = -ah(t)/2 \end{cases} \quad (8)$$

利用该关系,可使 SOA 对光信号进行类似于 MZI 的相位调制。图 7(a) 为 2013 年 Singh 课题组^[32] 利用 SOA-MZI 实现逻辑“与”门运算的实验结构图。在 B_1 和 B_2 端口分别引入相同的控制信号,同时在 B_2 端口引入时域延迟 $\Delta\tau$, 此时两个 SOA 的增益和相位分别为

$$\begin{cases} G_1(t) = \exp[h(t)] \\ G_2(t) = \exp[h(t - \Delta\tau)] \\ \Delta\varphi_1(t) = -ah(t)/2 \\ \Delta\varphi_2(t) = -ah(t - \Delta\tau)/2 \end{cases} \quad (9)$$

当控制信号 B 为逻辑 0 时,输出端口 C 中无输出信号;当控制信号 B 为逻辑 1 时,若探测信号 A 与控制信号 B 在 SOA 中同向传播,瞬时增益系数 $g(z, \tau)$ 变为与位置 z 无关的函数。瞬时增益系数积分为

$$h(t) = g(t)L. \quad (10)$$

若探测信号 B 与信号 A 在 SOA 中异向传播,瞬时增益系数积分为

$$h(t) = \frac{1}{L} \int_{-L/2}^{L/2} \ln G(t - 2zn_{SOA}/c) dz. \quad (11)$$

两路 SOA 输出信号在 C_2 内发生干涉,并从端口 3

输出,实现 A、B 信号的逻辑“与”运算。

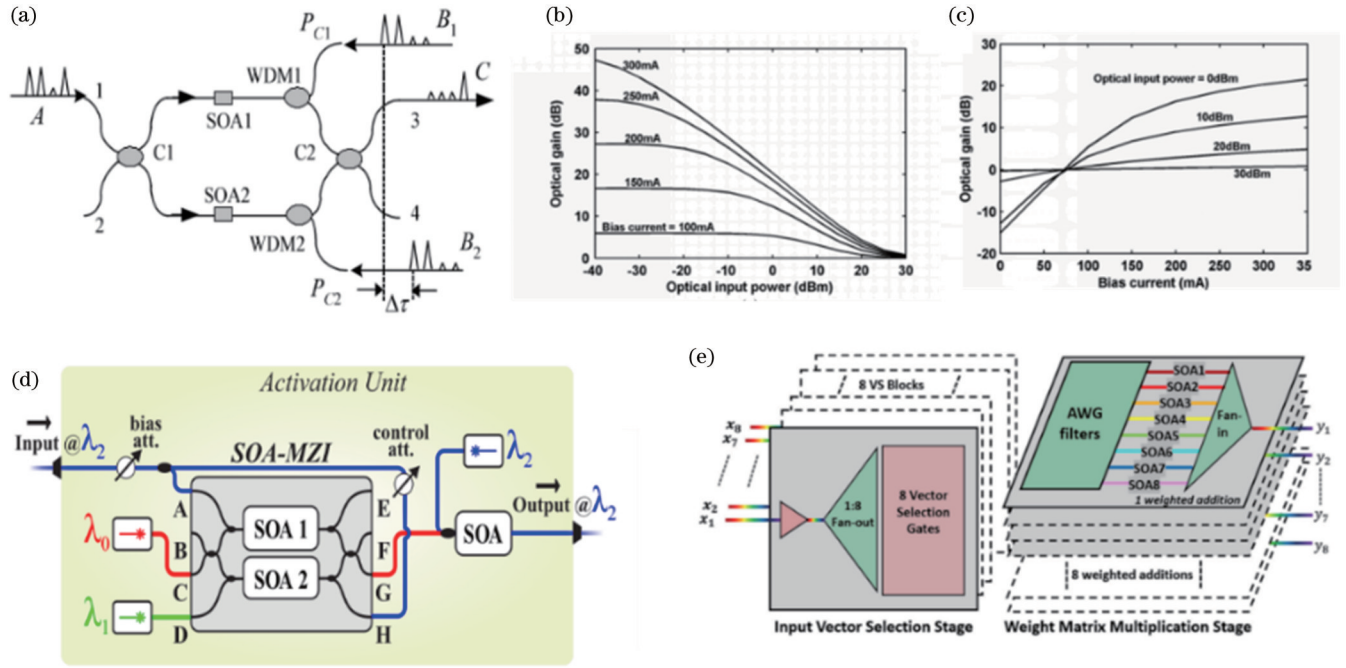


图 7 SOA 非线性激活函数。(a) 基于 SOA 的 MZI 实现逻辑“与”门^[32]; (b)(c) SOA 的增益与输入光功率和偏置电流的非线性关系^[35]; (d) 结合 XPM 和 XGM 的 Sigmoid 触发函数神经网络^[36]; (e) 片上 SOA 阵列神经网络^[33]

Fig. 7 SOA nonlinear activation function. (a) SOA-based MZI implements a logic AND gate^[32]; (b)(c) nonlinear relationship among SOA gain, input optical power, and bias current^[35]; (d) Sigmoid trigger function neural network combined with XPM and XGM^[36]; (e) on-chip SOA array neural network^[33]

对于 XGM 效应,可将 SOA 内载流子密度 $N(z, t)$ 和光波幅值 $A_{L,A}(z, t)$ 表示为

$$\begin{cases} \frac{\partial N}{\partial \tau} = \frac{I}{qwdL} - \frac{N}{\tau_c} - \frac{\Gamma a(N - N_0)}{\hbar\omega_0\omega b} \\ \frac{\partial A}{\partial z} = \frac{i\omega_0}{2\bar{n}c} \chi A - \frac{1}{2} \alpha_{\text{int}} A \end{cases}, \quad (12)$$

式中: I 为注入电流; q 为电子电荷; F 为限制因子; $\hbar\omega_0$ 为光子能量; a 为增益系数; N_0 为诱导透明时载流子密度; α_{int} 为内部损耗; n 为有效折射率; ω 、 b 、 L 为 SOA 的几何参数。系数 χ 可表达为

$$\chi(N) = -\frac{\bar{n}c}{\omega_0} (\alpha' + i) a (N - N_0), \quad (13)$$

式中: α' 为线宽增幅因子; i 为虚数单位。若 $A_{L,A} = \sqrt{P} \exp(i\phi)$, 其中 P 和 ϕ 分别为光功率和相位, 则式(12)可改写为

$$\begin{cases} \frac{\partial P}{\partial z} = (g - \alpha_{\text{int}}) P \\ \frac{\partial \phi}{\partial z} = -\frac{1}{2} \alpha g \\ \frac{\partial N}{\partial \tau} = \frac{I}{qwdL} - \frac{N}{\tau_c} - \frac{\Gamma a(N - N_0) P}{\hbar\omega_0\omega d} \end{cases}, \quad (14)$$

式中: g 为 SOA 的增益, 且

$$g = \Gamma a (N - N_0). \quad (15)$$

若输入光场 λ_1 和 λ_2 , 可得

$$\begin{cases} \frac{\partial P_i}{\partial z} = \Gamma a_i (N - N_{0,i}) P_i, \quad i = 1, 2 \\ \frac{\partial N}{\partial \tau} = -\frac{N}{\tau_c} - a_1 (N - N_{0,1}) \frac{P_1}{\hbar\omega_1\omega d} - a_2 (N - N_{0,2}) \frac{P_2}{\hbar\omega_2\omega d} + \frac{I}{qwdL} \end{cases}, \quad (16)$$

积分后表达式为

$$\begin{cases} P_i(\tau, L) = P_i(\tau, 0) \exp\{\Gamma a_i [y(\tau, L) - N_{0,i} L]\}, \quad i = 1, 2 \\ y(\tau, L) = \int_0^L N(\tau, z') dz' \end{cases}. \quad (17)$$

由式(17)可得到 SOA 的输出功率^[34]。SOA 的增益与输入光功率和偏置电流的非线性关系如图 7(b)、

(c) 所示。2009 年, Tang 课题组^[35]发现 SOA 的增益与偏置电流正相关, 且与输入光功率呈负相关关系, 当光

功率在 10 dBm 以下、偏置电流在 100 mA 以上时, SOA 的增益和偏置电流呈现强烈的非线性关系, 其总体变化趋势接近对数函数关系。

综上所述, XPM 和 XGM 使得 SOA 能被用于光学神经网络运算。在本质上, 神经网络是一种对输入参量进行权重分配后求和, 再导入触发函数进行运算的逻辑结构, 而光波具有幅值、波长、偏振态等可分配权重的自由度。与此同时, 利用 SOA 的非线性相位调制和非线性增益调制可实现类似于干涉仪、波长转换器、逻辑门的功能。Pleros 课题组^[36]提出一种结合 XPM 和 XGM 效应的 Sigmoid 激发 WDM 信号加权器, 如图 7(d) 所示。该结构分为由两个 SOA 并联形成的 SOA-MZI 和由单个 SOA 形成的 XGM 波长转换器 (XGM-WC) 两个部分。将输入信号的波长视为权重值, 输入两个峰值不同的脉冲信号, SOA-MZI 将把信号的 π 相位差取 0, 最终输出强度符号相反、最低值相同的信号序列。XGM-WC 将该信号转换回原波长的同时将其再次取反, 整个系统将原本峰值不同的两个脉冲取平。由于 SOA-MZI 和 XGM-WC 具有非线性特性, 该系统具有 Sigmoid 触发函数的特性, 并且 Sigmoid 函数的斜率可通过改变系统参量进行调节。该信号加权器可将输入信号的消光比提高 3.76 dB, 且信号分辨能力相较于传统光学神经网络激发系统提高了 100%。若将多个类似系统集合甚至集成至片上, 可实现运算量更大、功率较低的神经网络。如图 7(e) 所示, Calabretta 课题组^[33]于 2020 年通过波导光栅阵列将光波分为波长不同的多频道, SOA 阵列为不同的波长分配不同的权重, 再使用光栅阵列对输出信号进行求和, 由此实现标准输出功率为 -25.5 dBm 的片上全光神经网络运算。该神经网络共设置三层神经元, 每层神经元的数量分别为 4、4、3。使用鸢尾花分类模型测试时, 预测准确度为 85.8%, 比理论上可达到的最高准确度低 9.2%。在第三层神经元测得的累计正规化均方根误

差 (NRMSE) 为 0.2。网络中每一个 SOA 的功耗为 42 mW, 每层各有一个输入端的 8 个神经元总功耗为 2.7 W。考虑到光子运算中功耗不伴随波特率线性增加, 若假设传输速率为 10 GSa/s, 每一次神经网络运算共消耗的能量仅为 4.2 pJ。

3.2.3 基于 MRR 的全光非线性激活函数

通过利用光学 MRR 中的载流子色散效应^[8,37-39], 或者通过将热光效应^[40-41]与 MRR 相结合, 可以在 SOI (silicon-on-insulator) 平台上实现可编程的片上全光非线性激活函数。同时, 在氮化硅 (Si_3N_4) 平台中, 可以利用克尔效应来实现可编程的片上全光非线性激活函数^[42]。这些工作不仅提升了全光神经网络的性能, 还为未来超紧凑的片上可调全光非线性神经网络提供了发展思路。

3.2.3.1 基于载流子色散效应的非线性激活函数

2020 年, 普林斯顿大学的 Prucnal 课题组^[8]首次提出一种基于 MRR 的片上可调全光非线性激活函数。该体系通过调节入射信号光波长与 MRR 共振波长之间的失谐量, 可以实现 4 种 (ReLU、Sigmoid、Radial Basis、Softplus) 线型。他们在实验中验证了上述线型, 并利用该非线性函数分别以 100% 和 94% 的准确率完成了对 XOR 以及对 MNIST 手写数字的分类任务。

该工作采用的构型如图 8(a) 所示。整个器件在马赫-曾德尔耦合器 (MZC) 后面级联一个带有 MRR 的 MZI, 实验装置如图 8(b) 所示。MZC 的作用是引入额外的调制自由度, 而该器件的非线性主要来自硅波导中自由载流子色散 (FCD) 效应。FCD 效应在 MRR 的积累下将会转变为受光功率影响的非线性相位响应, 而这种相位响应又会在 MZI 的作用下转变为具有非线性的透过率响应。上述 3 个器件都装载了热光调制器, 用于引入主动调控能力。

利用非线性耦合模理论得到式 (18) 来描述 MRR 中的非线性效应, 即

$$\begin{cases} \delta a / \delta t = \sqrt{\gamma_p P_{in}(t)} + i(\delta\omega - n_{Kerr}|a|^2 + \sigma_{fcd}\alpha_{TPA}n)a - (1 + \alpha_{TPA}|a|^2 + \gamma_{FCA}\alpha_{TPA}n)a \\ \delta n / \delta t = |a|^4 - n/\tau \end{cases}, \quad (18)$$

式中: a 为 MRR 腔中归一化的光场复振幅; n 为自由载流子数量; $\delta\omega$ 为入射信号光频率和 MRR 共振频率的失谐量; t 为归一化时间; n_{Kerr} 、 σ_{fcd} 、 α_{TPA} 、 γ_{FCA} 、 γ_p 分别表示克尔效应、FCD 效应、双光子吸收 (TPA)、自由载流子吸收 (FCA) 和品质因子的贡献。其中, FCD 效应在上述非线性相位响应中占主导地位^[37-38], 并导致 MRR 共振峰随着入射光功率的增加而蓝移。随着共振峰的蓝移, 不同频率的入射光将会以不同的方式调制其耦合到 MRR 的大小, 并进一步影响 FCD 产生的相位响应, 从而形成反馈机制。在此机制下, 不同频率的入射光将会呈现出不同的

非线性响应。通过实验优化 MZI 上施加的电压, 便可得到如图 8(c) 所示的 4 种线型, 此时的阈值可以达到 3.16 mW。

2022 年, 该课题组继续利用上述构型制备了图 8(d) 所示的样品, 并基于 MRR 中的 FCD 效应, 实现了图 8(e) 所示的 Sigmoid 型非线性激活函数^[38]。基于该出发点, 该课题组继续探讨通过 MZI 调节光学双稳态的方法, 从而实现高达 100 dB 的切换对比, 同时利用该方案的可编程性制备出将信号对比度提升 40 倍的全光阈值器件, 以及将长脉冲信号转换为短脉冲信号的光脉冲整形器件。通过利用反向偏压来降低自由载

流子的寿命,该器件的响应速度可达到 2.5 Gb/s。

2022 年,该课题组在 MRR 上引入石墨烯,进一步增强了非线性^[39],如图 8(f)所示。这是因为石墨烯中的克尔效应加强了硅中的 FCD 效应,而且石墨烯还具

有饱和吸收特性,从而导致腔的品质因子受到光强的影响,在增强非线性的同时降低了光学双稳态的阈值。将这一影响加入式(18)中,可得到对该体系的进一步表述,即

$$\begin{cases} \delta a / \delta t = \sqrt{P} + i(\delta \omega - n_{\text{Kerr}} |a|^2) a + i(n + \sigma_{\text{fcd}} n^{0.8}) a - (1 + \alpha_{\text{TPA}} |a|^2 + \gamma_{\text{FCA}} n) a - 1 / \left(1 + \frac{|a|^2}{W_{\text{sat}}} \right), \\ \delta n / \delta t = |a|^4 - n / \tau \end{cases} \quad (19)$$

式中: P 表示由总线波导和 MRR 的耦合系数以及腔寿命、硅的 TPA 系数、TPA 体积以及非线性载流子体积所决定的系数; n_{Kerr} 表示硅与石墨烯的克尔系数相叠加而引入的折射率; $(1 + \alpha_{\text{TPA}} |a|^2 + \gamma_{\text{FCA}} n) a - 1 / \left(1 + \frac{|a|^2}{W_{\text{sat}}} \right)$ 表示由 FCA、TPA 以及石墨烯饱和吸收带来的损耗;式(19)第二个式子描述了 MRR 内归一化载流子密度随时间的变化,其中 τ 表示腔中归一化

的自由载流子寿命,而 $|a|^4$ 的存在意味着自由载流子的生成与 TPA 过程紧密联系。

基于该体系,他们通过类比共振-点火机制证明了其可以用于脉冲神经网络,并指出在 FCD 以及 MRR 中光强共同作用下可以实现自脉冲。这种自脉冲的激发阈值将带来图 8(g)所示的 leaky ReLU 线型。除此以外,该体系还能以更低的等效能耗(0.7 pJ)实现更快的响应速度(40 GHz),并以 90% 的准确率识别 MNIST 手写数字数据集。

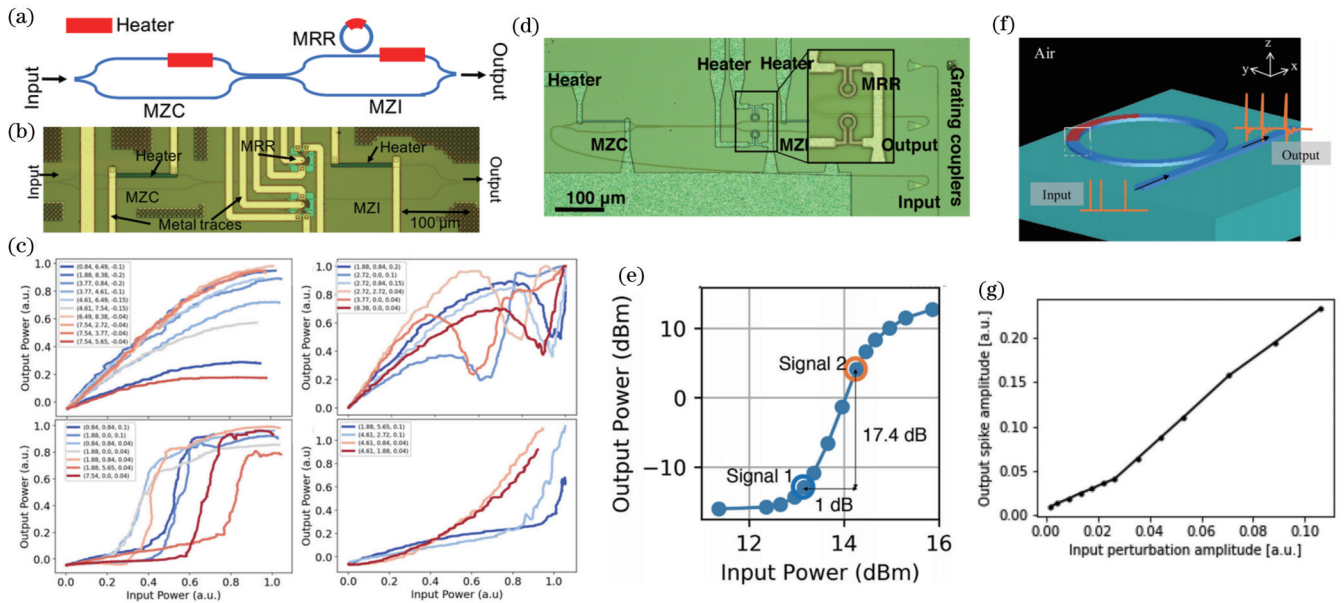


图 8 FCD 诱导的非线性激活函数。(a)(b) MZC 器件与带有 MRR 的 MZI 耦合的示意图与实验装置图^[8]; (c) 该方案能实现的 4 种线型; (d) 实验装置图^[38]; (e) 体系生成的 Sigmoid 函数; (f)(g) 加入石墨烯后的示意图以及可实现的线型^[39]

Fig. 8 FCD-induced nonlinear activation function. (a)(b) Schematic and experimental setup of MZC device coupling with MZI with MRR^[8]; (c) four line types that can be achieved by the scheme; (d) diagram of the experimental apparatus^[38]; (e) system-generated Sigmoid function; (f)(g) schematic after adding graphene and achievable line patterns^[39]

以上工作基于 MRR 的 FCD 效应实现了不同的全光非线性响应线型,在可编程的同时还可以利用光学双稳态以及自脉冲效应实现全光阈值器件、光脉冲整形器件的制备以及脉冲神经网络的搭建,为在片上全光神经网络中引入非线性激活函数提供了新的方案。

3.2.3.2 基于热光效应的非线性激活函数

由于硅波导中的 FCD 效应较弱,如果只利用 FCD 效应,将意味着只有当输入功率足够高时才可激

发非线性。为了解决这个问题,2022 年华南理工大学张新亮团队^[40]将 Ge 与 Si 波导相结合,实现了 0.74 mW 的低阈值全光非线性激活函数。整体结构以及实验样品如图 9(a)所示。通过调节入射光的波长,该体系可以实现 Radial Basis、ReLU、Elu 等非线性线型,如图 9(c)所示。

实现非线性响应的原理如图 9(b)所示,Ge 会吸收 1550 nm 左右的人射光并发热,这会让 Ge 的吸收峰红

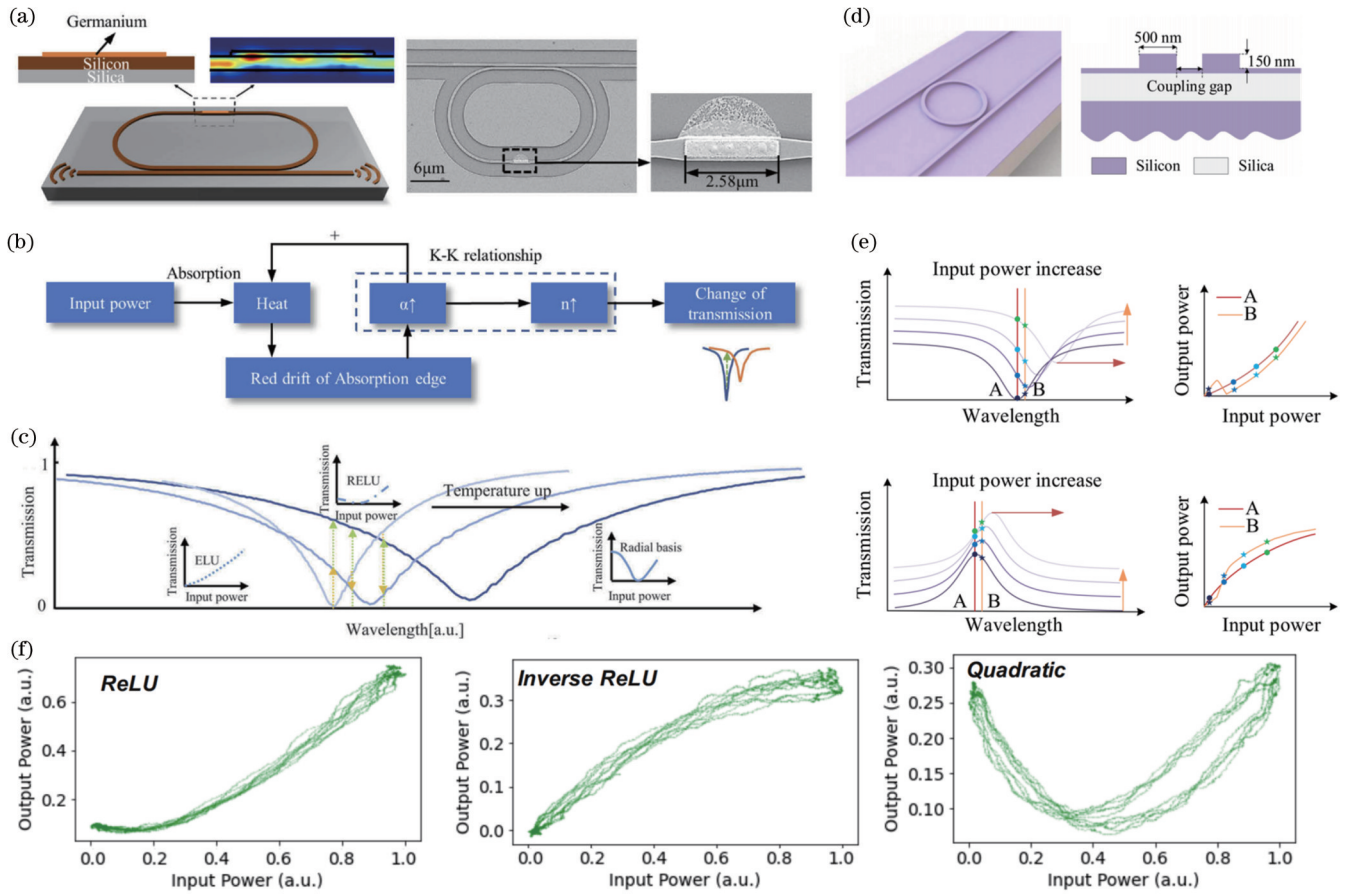


图9 Ge/Si复合材料全光非线性激活函数^[40-41]。(a)Ge/Si复合跑道型MRR示意图与实验样品;(b)实现非线性响应的原理图;(c)不同波长实现不同线型的示意图;(d)MRR器件及其横截面;(e)在through端和drop端随入射光功率而变化的透射谱以及该过程对应的非线性激活函数;(f)Si₃N₄平台中MRR器件生成的非线性激活函数

Fig. 9 All-optical nonlinear activation functions of Ge/Si composites^[40-41]. (a) Schematic and experimental samples of Ge/Si composite runway MRR; (b) schematic for achieving a nonlinear response; (c) schematic of different wavelengths achieving different line types; (d) schematic of MRR and its cross-sectional view; (e) schematic of the transmission spectrum and the corresponding nonlinear activation function at the through and drop ports as the input optical power increases; (f) nonlinear activation functions achieved in the Si₃N₄ platform

移,从而进一步增大Ge的吸收系数 α 。吸收系数的增大会进一步促进发热,从而形成正反馈机制。根据Kramers-Kronig (K-K)关系,吸收系数的增大又会引起折射率的直接变化,从而导致MRR的透射谱红移,同时伴随着更浅的共振谷,如图9(c)中不同颜色的透射谱线所示。为了对该机制进行数学表达,假设吸收系数和折射率随温度的变化关系为

$$\begin{cases} \Delta\alpha = k_1 e^{\Delta T} \\ \Delta n = k_2 \Delta T \end{cases} \quad (20)$$

可以推出功率的输入-输出关系为

$$P_{\text{out}} = \left| \frac{t - \exp(i\varphi - \alpha)}{1 - t \exp(i\varphi - \alpha)} \right|^2 P_{\text{in}}, \quad (21)$$

式中: φ 和 α 分别表示当光环绕MRR一圈时积累的相位(和折射率直接相关)和受到的损耗。当改变入射光功率 P_{in} 从而引入温度变化时,MRR的折射率和吸收系数会依据式(19)变化,从而在 P_{out} 中引入非线性。当

改变入射信号的频率时,可以得到3种非线性线型,如图9(c)所示。将该体系产生的Radial Basis和ReLU非线性函数用到MNIST手写数字数据集的训练中,可将准确率从没有加非线性的91.8%提高到94.8%。该过程对应的最快响应时间为7 μs ,对应的响应阈值为0.74 mW,应用到深度神经网络时,每一层的损耗为1 dB。

2022年,Yu等^[41]同样利用热光效应结合MRR在SOI平台中实现了拥有更小占地面积($10^2 \mu\text{m}^2$)和更低阈值(0.08 mW)的可调谐全光非线性激活函数器件。光入射到图9(d)所示的器件后,会发生TPA过程,在改变波导折射率(克尔效应)的同时激发自由载流子,从而引入FCA与FCD等现象,而FCA和TPA过程产生的载流子的带间和带内弛豫效应又会对结构进行加热,从而进一步改变折射率。在这些非线性效应中,效果最显著的是热光效应。在这种效应的作用下,随着入射光功率的增加,MRR的共振峰将会红移。因此在

through 端:当入射光波长对应于 MRR 初始共振峰波长时,随着入射光功率的增加,出射光功率将会先以线性方式递增,并在开始引起共振峰红移的某一阈值开始快速上升,形成 Softplus 型非线性激活函数,如图 9(e)所示;当入射光设置在 MRR 初始共振波长右侧时,将会得到 Radial Basis 型非线性激活函数。在 drop 端:当入射光设置为 MRR 初始共振波长时,将会得到 Clamped ReLU 型非线性激活函数;当入射光设置在 MRR 初始共振波长右侧时,将会得到 Sigmoid 型非线性激活函数。值得注意的是,这种器件之所以拥有很低的阈值,是因为其充分利用了 MRR 能将光局域在很小的体积内的能力,当 MRR 的 Q 值进一步增大,该器件的非线性阈值将会进一步降低^[41]。利用该非线性激活函数,能以 12 dB 的插入损耗将手写数字数据集的识别准确率提升至 98%。该器件的相应速率可以达到 12 μ s,对应于 80 kb/s 的传输速率。

3.2.3.3 基于 Si_3N_4 平台的非线性激活函数

由于热光效应受到散热时间的内在限制,其响应时间被限制在 10 μ s 左右^[40-41]。为了提高响应速率,需要利用其他非线性效应,比如 FCD 效应和克尔效应等。自由载流子在腔中的寿命大约为 1 ns,使得 FCD 效应的响应速率无法突破 ns 量级^[42]。为了将响应时间进一步缩短,可以结合 Si_3N_4 平台。由于 Si_3N_4 具有宽禁带,FCD 可被成功阻止,从而更加容易实现更快的克尔效应。2021 年, Jha 等^[42]在 Si_3N_4 平台中将设计好的 MRR 以类似于图 8(a)的方式加载在 MZI 的一个臂上,并将该 MZI 级联在 MZC 之后,实现了高达 10 Gb/s 的响应速率,并实现了 5 dB 的插入损耗,其响应阈值可以通过调节偏差电压进行调控。通过进一步调节 MRR 上的加热器,该平台可以用不同的阈值实现 ReLU、inverse-ReLU 和 Quadratic 型非线性激活函数,如图 9(f)所示。

4 非线性激活函数在智能光子学中的应用

前文主要介绍了非线性激活函数的种类及实现方法,下面将针对非线性激活函数在智能光子学中的应用进行介绍。由于激活函数作为神经网络中非线性元素的引入者,因此神经网络可以应用到更多非线性模型中,而激活函数的应用主要依托于以神经网络为主体的光计算架构。下面以神经网络为例,介绍激活函数在神经网络架构中的应用,包括分类神经网络、实时响应网络、可重构神经网络等。

4.1 分类神经网络及模式识别

分类功能被视为智能光子学中最重要、最基础的功能之一,分类神经网络是一种重要的神经网络。非线性激活函数可以用来解决分类或模式识别问题。总的来说,非线性激活函数可以应用于两种神经网络——全光调制神经网络和电光调制神经网络架构。

在全光调制神经网络中,可以利用波分复用技术来实现光子神经网络的可扩展电路架构,在光学领域演示了模式识别^[5],并通过绘制不同激励脉冲下环形谐振器的传输能量与固定波长下脉冲能量的关系,获得了激活函数。其结构如图 10(a)、(b)所示。该激活函数可被应用到单层脉冲神经网络中,从而实现模式识别,如图 10(b)所示。该工作利用了光学系统固有的高速和大带宽的优点,有望直接处理光学通信和视觉数据。除此之外,全光调制神经网络还可以应用到分类模型中。2019 年 *Optica* 报道了一项利用全光调制神经网络对统计 Ising 模型的有序和无序阶段进行分类的工作,其中线性操作由空间光调制器和傅里叶透镜控制,而非线性光学激活函数在具有电磁感应透明度的激光冷却原子中实现,如图 10(c)、(d)所示。在该系统中,线性矩阵元素和非线性函数都可以独立编程,以实现不同的人工神经网络架构和应用^[6]。

此外,还可以在电光调制系统中构造非线性激活函数来解决分类问题。如使用基于铟锡氧化物(ITO)的电吸收调制器,将该层单片集成到硅光子波导中,其动态响应范围可被用作光子神经元的非线性激活函数。图 11(a)展示了该光子神经网络的架构,网络中的非线性由光电压驱动电吸收调制器,而电吸收调制器模拟激活函数非线性地调制激光功率实现。图 11(b)展示了 Si 和 ITO 中的自由载流子基吸收调制器及吸收率与波导横截面几何形状、截面注入电荷、驱动电压的关系^[10]。

4.2 可重构神经网络体系

近年来,基于可重构体系的光学应用异军突起,在当代光学领域迅速占领了一席之地。相比于传统不可重构的体系,可重构体系往往具有更高的应用灵活性,甚至可以与计算机科学结合,为光子学带来新的研究机遇。以编码超材料为代表的可编程光子器件^[43]和以电光调制实现拓扑相变的可重构非厄密拓扑器件^[44]都是可重构领域的典型代表。可重构非线性激活函数和基于可重构非线性激活函数的智能光计算平台研究都是智能光子学领域的重要方向。目前已有研究表明,可重构的全光非线性函数可以在全光调制神经网络和电光调制神经网络实现。

在全光调制神经网络中,主要通过硅波导中的自由载流子色散实现非线性,其体现为对输入光功率的非线性相位响应^[8]。上述器件主要通过马赫-曾德尔干涉仪的臂上微环谐振器实现非线性,其上的热光调谐加热器允许调谐器件的振幅和相位偏置,从而实现各种非线性传递函数的可编程性。如图 7(a)所示。实验中发现,这样的设置可以实现 ReLU、Sigmoid、Quadratic 和 Softplus 非线性激活函数的重构。除此之外,还可以通过多种不同方式实现激活函数重构,可实现的激活函数也不拘泥于上述 4 种。2022 年, Campo 和 Pérez-López^[45]介绍了基于 MZI 和 MRR 的可重构

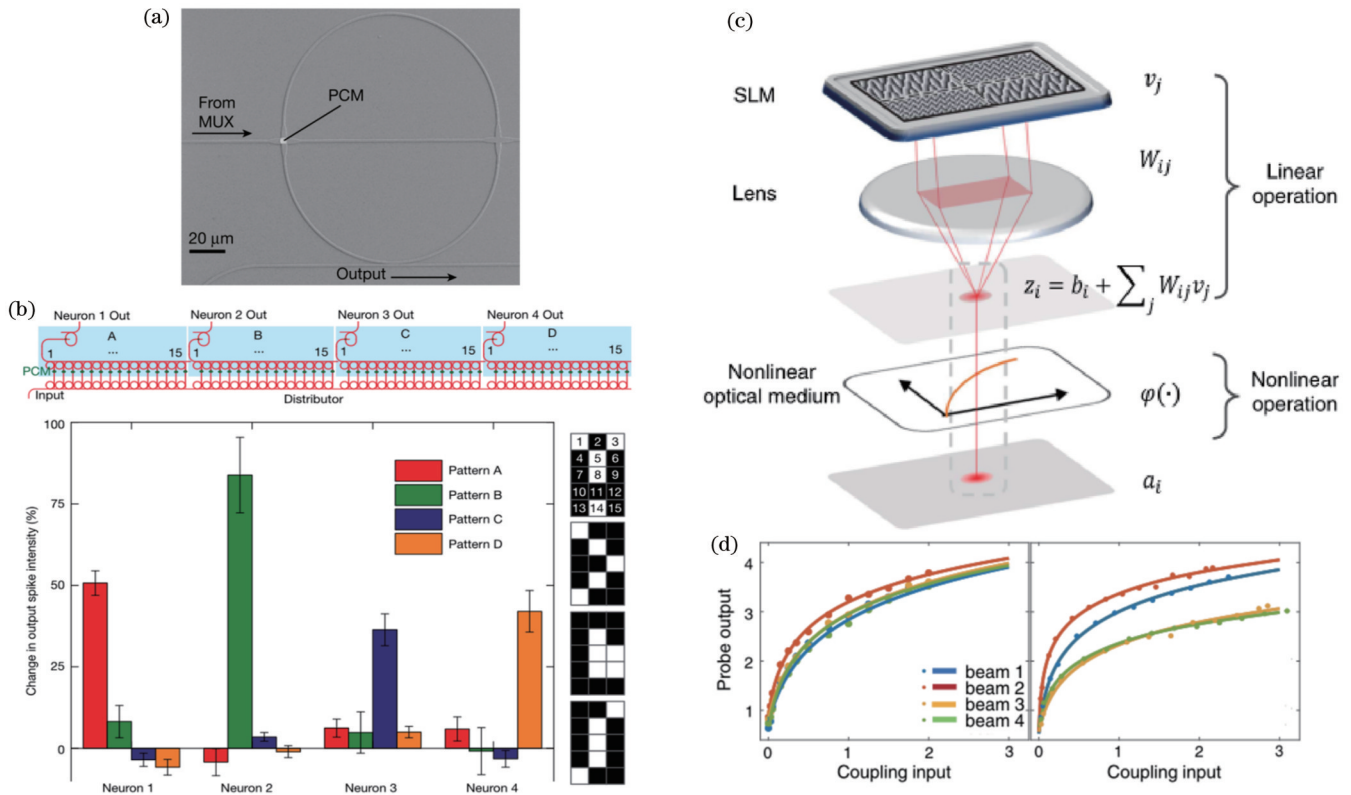


图 10 全光神经网络解决分类问题。(a)实现非线性激活函数的环形谐振器^[5]；(b)应用该激活函数实现模式识别的功能^[5]；(c)实现分类功能的神经网络架构，展示了网络中的线性操作和非线性操作^[6]；(d)EIT腔输出的非线性激活函数^[6]

Fig. 10 All-optical neural networks used to solve classification problems. (a) A ring resonator implementing a nonlinear activation function^[5]; (b) apply the activation function to implement the function of pattern recognition^[5]; (c) a neural network architecture that implements classification functions, demonstrating linear and nonlinear operations in the network^[6]; (d) nonlinear activation function of EIT cavity output^[6]

激活函数体系,其可以生成 Sine、Tanh、Softsign 等 15 种激活函数。这样的可重构激活函数可以应用到集成光学神经网络中,如图 12(a)所示。基于该可重构系统,可以很方便地评估不同形式的非线性激活函数解决相同问题的能力,为解决特定问题找到最优方案,图 12(b)展示了利用不同激活函数在 Iris 数据集中进行训练的结果。

除全光调制神经网络外,在电光调制神经网络中也可以实现可重构激活函数。2020 年,Williamson 等^[11]介绍了一种可调谐的非线性激活函数,该激活函数主要依托于实现光学非线性的电光硬件平台,可以显著提升神经网络的分类性能。图 2(c)展示了一种基于可调谐波导干涉仪网格的单层光学神经网络,该网络中的非线性激活函数主要通过光电调制实现,图 2(c)右端展示了实现该非线性激活函数的电路示意图,其中的偏压 V_b 是可调制的,因此该平台具有可重构的性质,可以实现 ReLU 或 Sigmoid 等激活函数。图 13(a)展示了实现图像分类的神经网络示意图;如图 13(b)所示,利用非线性激活函数可以显著提高分类精度,相比于不激活的情形,其精度提升了 8%。在神经网络中应用可重构的激活函数,可以根据实际需要灵

活编程,用于执行各种神经形态任务,这为光子器件智能化和集成化提供了新的实现方式。

4.3 实时响应神经网络与高效信息处理器

激活函数在信息的高效处理领域具有重要的应用价值,而实现宽带宽、低噪声、低能耗的光子网络体系是当今时代所需。因为集成光学网络中光子的响应时间往往是 ps 量级,且相比于电子体系,光子体系的能量损耗往往更低,故基于光子的神经网络设计具有高吞吐量和低功耗的优点^[9]。基于非线性元件的电光调制是实现信息高效处理的潜在器件。2019 年 Sorger 团队介绍了基于相变或强载流子色散效应的电光器件,认为其很有可能被应用到神经网络的感知器和非线性激活函数中,并且指出具有 ps 响应、低损耗、高调制范围、不会恶化或猝灭的光学非线性器件具有极大的实现全光神经元的潜力。在片上光学神经网络构建中,非线性激活函数的响应速度、阈值、损耗等指标参数尤其重要,图 14(a)比较了不同材料系统电光调制器和吸收调制器应用于感知器和非线性激活函数的关键指标,包括可积性、紧凑性、能量效率、速度,以及消光比和插入损耗之比,这些因素都会影响基于电光调制器的非线性激活函数的指标参数^[46]。Sorger 团队^[47]

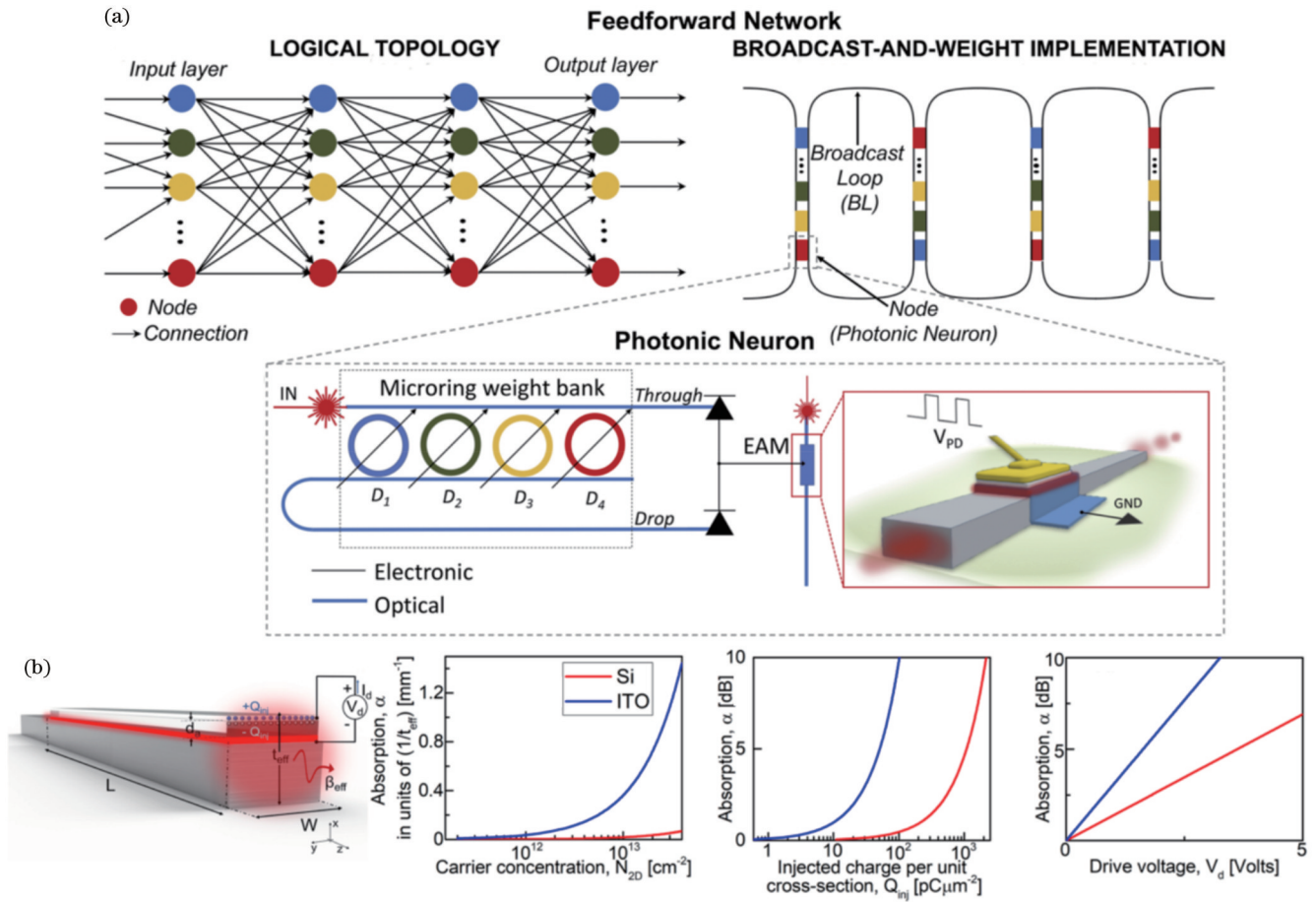


图 11 电光神经网络解决分类问题^[10]。(a)电光神经网络的架构;(b) Si和ITO中的自由载流子基吸收调制器及吸收率
 Fig. 11 Electro-optical neural networks used to solve classification problems^[10]. (a) Architecture of electro-optical neural networks;
 (b) free carrier-based absorption modulators and absorption rates in Si and ITO

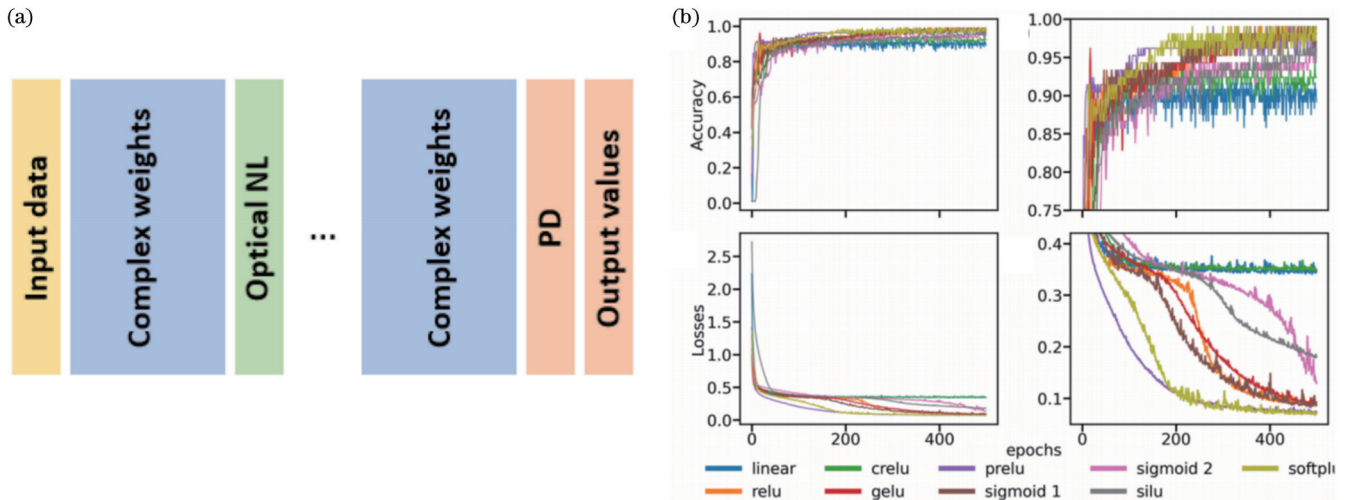


图 12 可重构神经网络体系^[45]。(a)应用可重构激活函数的集成光学神经网络;(b) 利用不同激活函数在 Iris 数据集上进行训练时所得的结果
 Fig. 12 Reconfigurable neural network architecture^[45]. (a) Integrated optical neural networks applying reconfigurable activation functions;
 (b) results of training in the Iris dataset using different activation functions

还在另一项工作中介绍了基于诱导透明和反向饱和吸收的纳米光子结构的光学传感器的非线性激活函数。基于此构建的神经网络可以提供较高的分类精度,且

网络中的噪声可被忽略,如图 14(b)所示。这项工作还展示了不同神经形态架构的神经元(MAC)计算速度和功率效率,对于全光调制神经网络,其计算速度比

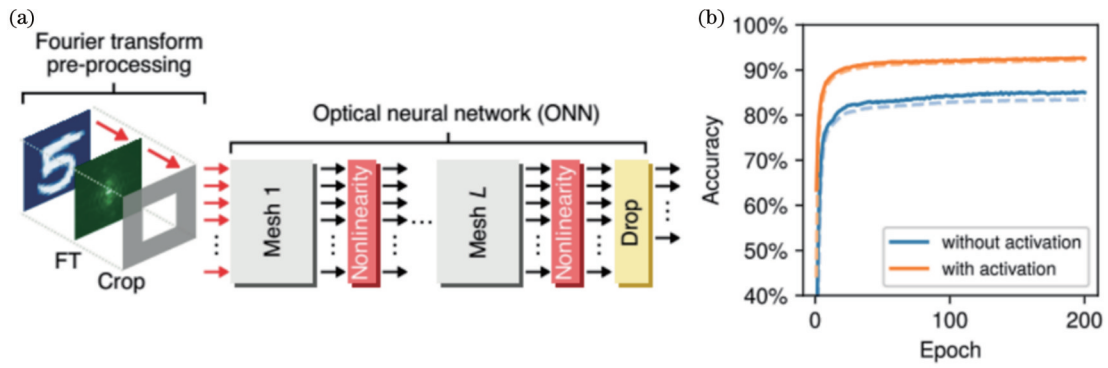


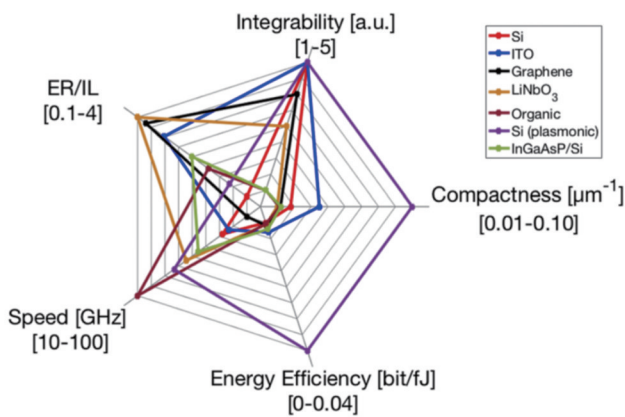
图 13 基于可调谐的非线性激活函数实现分类功能。(a) 实现图像分类的神经网络示意图^[2]; (b) 应用和不应用激活函数的光学神经网络的分类精度比较^[11]

Fig. 13 Classification function based on tunable nonlinear activation function. (a) Schematic of neural network to implement image classification^[2]; (b) comparison of classification accuracy of optical neural network with and without activation function^[11]

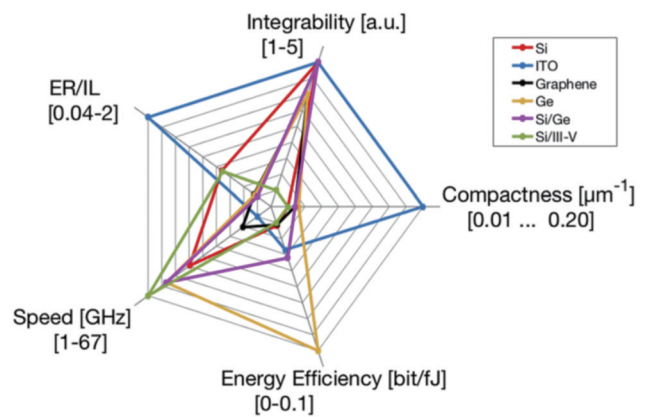
电光调制神经网络高 2 个数量级, 比 NVIDIA GPU 高 3 个数量级, 而计算效率比电光调制神经网络高 4 个数量级, 比 NVIDIA GPU 高 10 个数量级, 这为下一代信息处理硬件实现具有巨大吞吐量且低能耗的纯光子神经网络指明了方向。除此之外, 对于多层神经网络架构,

可以使用有源集成光子器件构建非线性激活函数, 该方法在单片上构建多层神经网络, 消除了由外部转换产生的能量和延迟成本, 为实现信息的高效处理提供了可能^[9]。综上所述, 非线性激活函数在构建实时响应神经网络与高效信息处理器上具有巨大的潜力。

(a) ELECTRO OPTIC MODULATOR (EOM)



ELECTRO ABSORPTION MODULATOR (EAM)



(b)

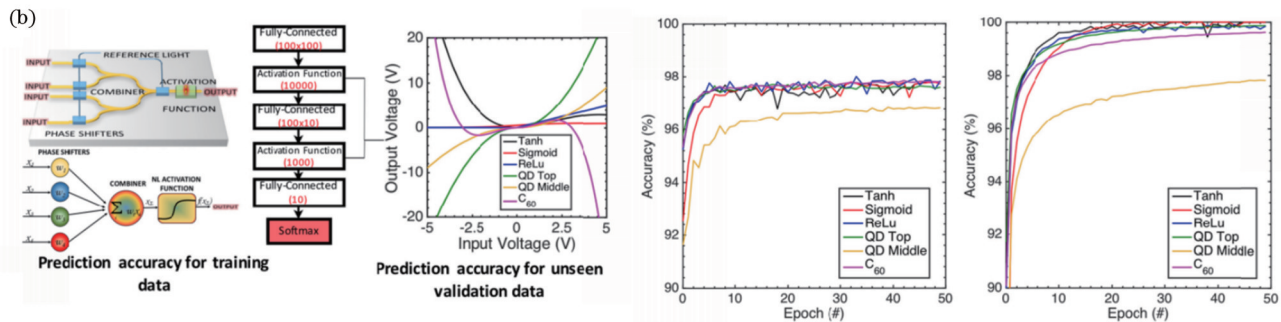


图 14 激活函数在实时响应神经网络与高效信息处理器中的应用。(a) 不同材料系统的电光调制器和吸收调制器的关键指标的评估结果^[45]; (b) 基于诱导透明和反向饱和吸收的纳米光子结构的非线性激活函数原理图与训练结果^[46]

Fig. 14 Application of activation function in real-time response neural network and efficient information processor. (a) Evaluation results of key indicators of electro-optical modulators and absorption modulators in different material systems^[45]; (b) schematic and training results of nonlinear activation function of optical sensors exhibiting nanophotonic structures that induce transparency and reverse saturated absorption^[46]

5 总结与展望

由于非线性函数可以用来描述大量不同的系统,因此引入非线性激活函数后,神经网络可以应用到更多的非线性模型中。随着智能光子学技术的发展,非线性激活函数作为一种强大的数学工具,已经成为光子网络架构中不可缺少的一部分,而由强光与介质的相互作用引起的非线性现象为实现非线性激活函数提供了物理基础。本文对非线性激活函数的实现与应用进行综述。从实现非线性激活函数的物理机制出发,讨论了利用定制化材料、SOA 的非线性光学效应,以及 MRR 体系等构造非线性激活函数的方法,总结了非线性激活函数器件在响应速度、阈值和损耗方面的数据,为在光子网络架构中实现非线性激活函数提供理论基础和实验平台。此外,总结了在光子神经网络架构中应用非线性激活函数的实例,包括分类神经网络架构、可重构神经网络架构、实时响应神经网络与高效信息处理器等。在光子神经网络中引入非线性元素是一个有潜力的、快速发展的研究领域,已经引起了越来越多研究者的关注。非线性激活函数拓宽了光子神经网络的应用范围,并为下一代片上集成光子器件提供了崭新且有潜力的构建途径。

本文总结出一些具有前瞻性的发展方向:1)基于可重构体系的激活函数神经网络。基于电光调制系统和全光系统均可实现对激活函数的重构,可重构系统具有灵活解决多种问题的能力,这将成为未来集成光子器件的重要发展方向之一。2)超快响应的低损耗光子计算器件。相比于电子体系,光子神经网络系统的能量损耗往往更低,且由于光子的运行时间为 ps 量级,基于光子的神经网络设计具有高吞吐量和低功耗的优点,相应的非线性光学器件有望实现全光神经元,其响应速度快、功耗低,在实现超快响应的低损耗光子计算上具有巨大潜力,在未来可以进一步实现 fs 量级的响应时间、 μW 量级的响应阈值,以及低于 1 dB 的低损耗非线性激活函数器件。3)量子系统。光学非线性在量子信息处理中具有重要地位,在光子神经网络中引入非线性有望使神经网络获得处理量子问题的能力。

综上所述,本文介绍了非线性激活函数的种类及其实现方式,从物理原理上分析了在光子神经网络中加入非线性激活函数的物理原理,并介绍了非线性激活函数在智能光子学中的应用。非线性激活函数是现代光子技术领域中有潜力的发展方向,对未来智能光子学的发展具有至关重要的意义。

参 考 文 献

[1] Franken P A, Hill A E, Peters C W, et al. Generation of second harmonic[J]. *Physical Review Letters*, 1961, 7: 118.
 [2] Yang Z, Tan W M, Zhang T J, et al. MXene-based broadband ultrafast nonlinear activator for optical computing[J]. *Advanced*

Optical Materials, 2022, 10(17): 2200714.
 [3] Coarer F D L, Sciamanna M, Katumba A, et al. All-optical reservoir computing on a photonic chip using silicon-based ring resonators[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2018, 24(6): 7600108.
 [4] Steinbrecher G R, Olson J P, Englund D, et al. Quantum optical neural networks[J]. *NPJ Quantum Information*, 2019, 5: 60.
 [5] Feldmann J, Youngblood N, Wright C D, et al. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. *Nature*, 2019, 569(7755): 208-214.
 [6] Zuo Y, Li B H, Zhao Y J, et al. All-optical neural network with nonlinear activation functions[J]. *Optica*, 2019, 6(9): 1132-1137.
 [7] Williamson I A D, Hughes T W, Minkov M, et al. Reprogrammable electro-optic nonlinear activation functions for optical neural networks[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2020, 26(1): 7700412.
 [8] Jha A, Huang C R, Prucnal P R. Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics[J]. *Optics Letters*, 2020, 45(17): 4819-4822.
 [9] George J K, Mehrabian A, Amin R, et al. Neuromorphic photonics with electro-absorption modulators[J]. *Optics Express*, 2019, 27(4): 5181-5191.
 [10] Amin R, George J K, Sun S, et al. ITO-based electro-absorption modulator for photonic neural activation function[J]. *APL Materials*, 2019, 7(8): 081112.
 [11] Williamson I A D, Hughes T W, Minkov M, et al. Tunable nonlinear activation functions for optical neural networks[C]// *Conference on Lasers and Electro-Optics*, May 10-15, 2020, Washington, DC. Washington, D.C.: Optica Publishing Group, 2020: SM1E.2.
 [12] Tait A N, de Lima T F, Nahmias M A, et al. Silicon photonic modulator neuron[J]. *Physical Review Applied*, 2019, 11(6): 064043.
 [13] Miscuglio M, Mehrabian A, Hu Z B, et al. All-optical nonlinear activation function for photonic neural networks[EB/OL]. (2018-10-02)[2023-02-03]. <https://arxiv.org/abs/1810.01216>.
 [14] Perry J W, Alvarez D, Choong I, et al. Enhanced reverse saturable absorption and optical limiting in heavy-atom-substituted phthalocyanines[J]. *Optics Letters*, 1994, 19(9): 625-627.
 [15] Dejonckheere A, Duport F, Smerieri A, et al. All-optical reservoir computer based on saturation of absorption[J]. *Optics Express*, 2014, 22(9): 10868-10881.
 [16] Gao Y C, Zhang X R, Li Y L, et al. Saturable absorption and reverse saturable absorption in platinum nanoparticles[J]. *Optics Communications*, 2005, 251(4/5/6): 429-433.
 [17] 李淳飞, 王玉晓, 国风云, 等. C_{60} 介质反饱和和吸收动力学[J]. *物理学报*, 1993, 42(8): 1236-1244.
 Li C F, Wang Y X, Guo F Y, et al. Kinetics of reverse saturation absorption in C_{60} medium[J]. *Acta Physica Sinica*, 1993, 42(8): 1236-1244.
 [18] Wu J M, Lin X, Guo Y C, et al. Analog optical computing for artificial intelligence[J]. *Engineering*, 2022, 10: 133-145.
 [19] 陈蓓, 张肇阳, 戴庭舸, 等. 光学神经网络及其应用[J]. *激光与光电子学进展*, 2023, 60(6): 0600001.
 Chen B, Zhang Z Y, Dai T G, et al. Photonic neural networks and its applications[J]. *Laser & Optoelectronics Progress*, 2023, 60(6): 0600001.
 [20] Bramerie L, Le Q T, Gay M, et al. All-optical 2R regeneration with a vertical microcavity-based saturable absorber[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2012, 18(2): 870-883.
 [21] Massoubre D, Oudar J L, Dion J, et al. Scaling of the saturation energy in microcavity saturable absorber devices[J]. *Applied Physics Letters*, 2006, 88(15): 153513.
 [22] Massoubre D, Oudar J L, Fatome J, et al. All-optical

- extinction-ratio enhancement of a 160 GHz pulse train by a saturable-absorber vertical microcavity[J]. *Optics Letters*, 2006, 31(4): 537-539.
- [23] Paquot Y, Dupont F, Smerieri A, et al. Optoelectronic Reservoir computing[J]. *Scientific Reports*, 2012, 2: 287.
- [24] Su W J, Cooper T M, Brant M C. Investigation of reverse-saturable absorption in brominated porphyrins[J]. *Chemistry of Materials*, 1998, 10(5): 1212-1213.
- [25] Larger L, Soriano M C, Brunner D, et al. Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing[J]. *Optics Express*, 2012, 20(3): 3241-3249.
- [26] Fleischhauer M, Imamoglu A, Marangos J P. Electromagnetically induced transparency: optics in coherent media[J]. *Reviews of Modern Physics*, 2005, 77(2): 633-673.
- [27] Boller K J, Imamoglu A, Harris S E. Observation of electromagnetically induced transparency[J]. *Physical Review Letters*, 1991, 66(20): 2593-2596.
- [28] Leng H X, Szychowski B, Daniel M C, et al. Dramatic modification of coupled-plasmon resonances following exposure to electron beams[J]. *The Journal of Physical Chemistry Letters*, 2017, 8(15): 3607-3612.
- [29] Rios C, Stegmaier M, Hosseini P, et al. Integrated all-photonic non-volatile multi-level memory[J]. *Nature Photonics*, 2015, 9(11): 725-732.
- [30] Cheng Z G, Rios C, Pernice W H P, et al. On-chip photonic synapse[J]. *Science Advances*, 2017, 3(9): e1700160.
- [31] Amiri I S, Rashed A N Z, Mohamed A E N A, et al. Nonlinear effects with semiconductor optical amplifiers[J]. *Journal of Optical Communications*, 2023, 44(1): 11-17.
- [32] Singh V K, Singh Yadav A K, Kumar A, et al. Semiconductor-optical-amplifier Mach-Zehnder interferometer based optical networks[J]. *Telkomnika Indonesian Journal of Electrical Engineering*, 2013, 11(1): 525-528.
- [33] Shi B, Calabretta N, Stabile R. Deep neural network through an InP SOA-based photonic integrated cross-connect[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2019, 26(1): 7701111.
- [34] Fan H L, Wu C M, Dutta N K, et al. Cross gain modulation in semiconductor optical amplifier[J]. *Proceedings of SPIE*, 1999, 3625: 250-256.
- [35] Wei J L, Hamie A, Giddings R P, et al. Semiconductor optical amplifier-enabled intensity modulation of adaptively modulated optical OFDM signals in SMF-based IMDD systems[J]. *Journal of Lightwave Technology*, 2009, 27(16): 3678-3688.
- [36] Mourgias-Alexandris G, Tsakyridis A, Passalis N, et al. An all-optical neuron with sigmoid activation function[J]. *Optics Express*, 2019, 27(7): 9620-9623.
- [37] Leuthold J, Koos C, Freude W. Nonlinear silicon photonics[J]. *Nature Photonics*, 2010, 4(8): 535-544.
- [38] Huang C R, Jha A, de Lima T F, et al. On-chip programmable nonlinear optical signal processor and its applications[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2021, 27(2): 6100211.
- [39] Jha A, Huang C R, Peng H T, et al. Photonic spiking neural networks and graphene-on-silicon spiking neurons[J]. *Journal of Lightwave Technology*, 2022, 40(9): 2901-2914.
- [40] Wu B, Li H K, Tong W Y, et al. Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator[J]. *Optical Materials Express*, 2022, 12(3): 970-980.
- [41] Yu W Z, Zheng S A, Zhao Z Y, et al. Reconfigurable low-threshold all-optical nonlinear activation functions based on an add-drop silicon microring resonator[J]. *IEEE Photonics Journal*, 2022, 14(6): 5559807.
- [42] Jha A, Huang C R, Prucnal P R. Programmable, high-speed all-optical nonlinear activation functions for neuromorphic photonics [C]//*Optical Fiber Communication Conference (OFC) 2021*, June 6-11, 2021, Washington, DC. Washington, D.C.: Optica Publishing Group, 2021: Tu5H.3.
- [43] Wan X, Qi M Q, Chen T Y, et al. Field-programmable beam reconfiguring based on digitally-controlled coding metasurface[J]. *Scientific Reports*, 2016, 6: 20663.
- [44] Leykam D, Mittal S, Hafezi M, et al. Reconfigurable topological phases in next-nearest-neighbor coupled resonator lattices[J]. *Physical Review Letters*, 2018, 121(2): 023901.
- [45] Campo J R R, Pérez-López D. Reconfigurable activation functions in integrated optical neural networks[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2022, 28(4): 8300513.
- [46] Miscuglio M, Adam G C, Kuzum D, et al. Roadmap on material-function mapping for photonic-electronic hybrid neural networks[J]. *APL Materials*, 2019, 7(10): 100903.
- [47] Miscuglio M, Mehrabian A, Hu Z B, et al. All-optical nonlinear activation function for photonic neural networks[J]. *Optical Materials Express*, 2018, 8(12): 3851.

Principles and Applications for Optical Nonlinear Activation Function Devices

Lü Qinghong¹, Ma Rui¹, Xiao Shenyu², Yu Weijia³, Liu Zhifei³, Hu Xiaoyong^{1,4*},
Gong Qihuang^{1,4}

¹*State Key Laboratory of Artificial Microstructure and Mesoscopic Physics, School of Physics, Peking University, Beijing 100871, China;*

²*School of Physics, Nankai University, Tianjin 300071, China;*

³*Institute of Information Photonics Technology, Faculty of Science, Beijing University of Technology, Beijing 100124, China;*

⁴*Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, Shanxi, China*

Abstract

Significance In recent years, with the development of computer technology, artificial intelligence has gradually penetrated many aspects of current human life. As the basic architecture of artificial intelligence, machine learning and neural networks have been given increasingly more attention to by researchers in recent years. At present, neural networks have been applied in matrix calculation, equation solving, data analysis, and many other fields, and have become a research field with great development potential in the 21st century. In conventional neural networks, linear functions are the primary mathematical tool. Nowadays, nonlinear activation functions (NAFs) can be employed to describe a large number of different systems, such as electric power systems, optical systems, economic systems, biological systems, computer networks, and communication systems. Nonlinear functions are a much more powerful mathematical tool than linear functions. Therefore, they are introduced into the neural networks to apply the neural networks to more nonlinear models.

Currently, the application of nonlinear functions in neural networks is mainly realized through the nonlinear activation layers. The nonlinear activation layers of neural networks (NNs) can alter the linear transformation relationship between multi-layer networks, and thus enabling the NNs to solve more complex and advanced learning with flexibility. In pursuit of faster processing speed and lower energy consumption, optical neural networks (ONNs) have caught much attention from researchers in recent years. The response time of photons in ONNs is often picoseconds, and the energy loss of optical systems is often lower than that of electronic systems. Thus, the design of ONNs features high throughput and low power consumption. Therefore, except for electronic systems, photonic systems have shown a broad development prospect in computing. Additionally, the intelligent photonics represented by ONNs has also emerged and become an important development direction of information processing in the future. As an indispensable module in ONNs, a series of optical NAF devices have emerged. The nonlinear phenomena caused by the interaction between strong light and medium provide a powerful theoretical basis for applying NAF in photonic network architecture, and the integrated photonic device is also a feasible experimental platform for realizing NAFs. The NAFs broaden the application range of ONNs and provide a potential way to construct the next generation of integrated photonic devices on chip, which has a very broad development prospect. In this review, we summarize the recent studies that introduce optical NAFs to the systems and discuss their physical mechanisms and application capabilities. Meanwhile, this review also summarizes and discusses the challenges and future trends for the development in the research on optical NAF devices in ONNs, and provides outlooks related to such devices.

Progress Methods and principles for generating various NAFs under optic-electro-optic (O-E-O) and all optical regimes, which are the two major regimes reported so far, are summarized with more emphasis on the later regime. The applications for NAFs are also presented. NAF modulators under the O-E-O regime can be dissected into electro-absorption-modulators (EAMs) and electro-optic-modulators (EOMs), both of which have distinct merits and drawbacks. EOM utilizes a Mach-Zehnder interferometer (MZI) or the phase shifts in a micro-ring resonator (MRR) to modulate amplitude via interference. EAM can directly modulate the light amplitude without the need for interference and thus can be designed with a smaller footprint. However, the carrier-based EAM suffers from lower speed compared with the field-driven EOM. More specific comparison can be seen in the main content (Fig. 14). Besides the aforementioned categories, O-E-O modulation can also be achieved by doping the MRR to make its transmission sensitive to electrical currents and thus can engender different NAFs under different biasing conditions. NAF modulators under the all optical regime can be dissected into three major categories including customized materials, semiconductor optical amplifier (SOA), and MRR.

Customized materials can be further divided into saturated absorption, reverse saturated absorption materials, electromagnetically induced transparency (EIT) materials, phase change materials (PCMs), and light-matter interaction (LMI) material structure. Meanwhile, we provide the mathematical basis of the cross-phase and cross-gain modulation effects in SOA. Additionally, examples of how these nonlinear effects can be utilized to realize optical neural-network devices capable of simple operations are provided by chaining multiple semiconductor optical amplifiers together in certain configurations, such as optical "AND" logic gates and optical signal thresholders. For MRR, by utilizing the free-carrier dispersion (FCD) effect or thermo-optical (TO) effect, NAFs with distinct responding times and threshold can be designed and optimized by incorporating materials like graphene or Ge and platforms like Si_3N_4 . The Kerr effect in graphene can enhance the FCD in silicon, while Ge can be adopted to facilitate the TO process. In addition, the Si_3N_4 platform can be utilized to increase the processing speed by blocking FCD and turning to the Kerr effect. Performance parameters relevant to the threshold, responding time, and loss of these devices are also summarized and compared. The generated NAF can be leveraged to improve the performance in applications like pattern recognition and classification while adding reconfigurability to the NNs and facilitating real-time response NNs and efficient information processors. Finally, the prospects for NAF development, including reconfigurability, better performance, and developments in combination with quantum information processing are put forward.

Conclusions and Prospects NAFs in the optical regime have been realized with various schemes and enhanced the performance of ONNs. In summary, the performance of optical NAFs still needs improvement in terms of faster responding time or lower threshold and loss whether by incorporating new materials or by deliberately designing the SOA or MRR systems. As a result, they can better serve ONN to perform more accurate and complex tasks.

Key words nonlinear optics; optical neural networks; nonlinear activation function