

第二届光电子集成芯片立强论坛  
暨硅光技术与应用研讨会

2021年5月 | 重庆



# 向量矩阵乘法的光学实现方式

焦述铭  
鹏城实验室 (广东深圳)  
2021年5月15日  
jiaoshm@pcl.ac.cn

## Outline

- 向量矩阵相乘
- 向量矩阵乘法的光学实现
  - 级联MZI网络
  - 波分复用与微环阵列
  - 光衍射神经网络
  - 单像素成像框架
  - 结构可调的光传播介质
  - 多芯光纤
- 特例：卷积形式相乘
- 非线性函数的光学实现
- 总结

## 向量矩阵相乘

- 输入：长度N的列向量X
- 与M×N矩阵U相乘
- 输出：长度M的列向量Y

$$Y = UX$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{M-1} \\ y_M \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & \dots & u_{1N} \\ u_{21} & u_{22} & \dots & \dots & u_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ u_{M1} & u_{M2} & \dots & \dots & u_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{N-1} \\ x_N \end{bmatrix}$$

## 向量矩阵相乘

- 向量矩阵相乘作为最基本的线性数学运算，有着非常广泛的应用
  - 图像处理与计算机视觉
  - 数值计算与方程求解
  - 运筹学和最优化
  - 机器人控制
  - 信息与通信工程
  - 数据挖掘
  - 机器学习
  - 经典及量子物理
  - .....



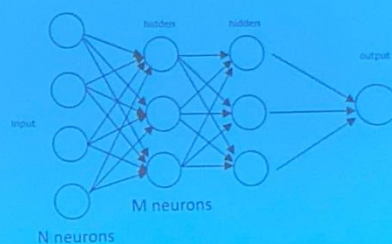
## 向量矩阵相乘

- 线性分类器(Linear classifier)
- X向量:图片中像素的值,声音信号,或者物体特征
- y向量:表示分类结果

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{M-1} \\ y_M \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & \dots & \dots & u_{1N} \\ u_{21} & u_{22} & \dots & \dots & \dots & u_{2N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u_{M1} & u_{M2} & \dots & \dots & \dots & u_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{N-1} \\ x_N \end{bmatrix}$$
$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{M-1} \\ y_M \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \dots \\ 0 \\ 0 \end{bmatrix}$$

## 向量矩阵相乘

- 在深度学习网络中, 相邻两层神经元之间运算:
  - 向量矩阵相乘(包含连接权重)
  - 非线性激活函数

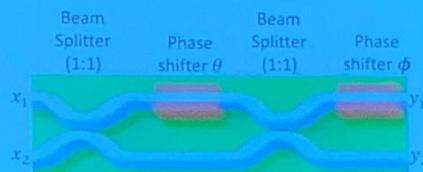


## 向量矩阵相乘

- 使用光学器件取代数字器件完成向量矩阵相乘
  - 高计算速度
  - 高并行性
  - 低能耗
- 近年来研究者也提出了各种不同的光学系统和装置来完成向量矩阵相乘的运算

## 级联MZI网络

- Mach-Zehnder interferometer(MZI)
- 基本单元：马赫-曾德尔干涉仪(片上波导)



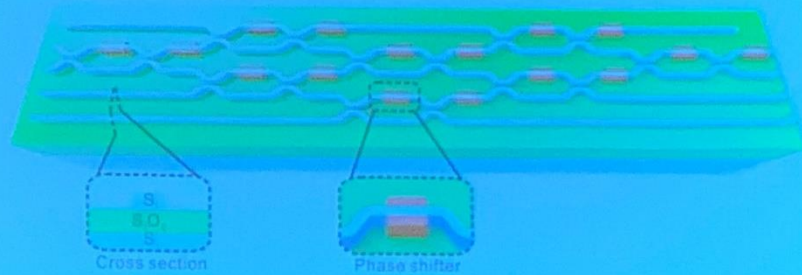
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = i e^{i\frac{\theta}{2}} \begin{bmatrix} e^{i\phi} \sin(\frac{\theta}{2}) & e^{i\phi} \cos(\frac{\theta}{2}) \\ \cos(\frac{\theta}{2}) & -\sin(\frac{\theta}{2}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- Michael V. A. Tang, Sankarsh Maripatnam, Łukasz Wierzbinski, Amir Khourramkhani, and Michael R. DeWeese, "Design of optical neural networks with compact representations," Opt. Express 27(10), 14009-14029 (2019).
- David A. B. Miller, "Scaling a neural network of optical components (Journals)," Photon. Res. 1(1), 1-19 (2013).



## 级联MZI网络

- 数学上可以证明，对于任意unitary matrix，可以用级联的MZI网络来表示，所有 $\theta$ 和 $\phi$ 的值要优化设计



- Tian Zhang, Xu Wang, Yihang Dai, Yuxiang Cao, Fan Dai, Xu Han, Xueqian Sun, and Kun Xu, "Efficient training and design of photonic neural network through neuroevolution," Opt. Express 27(26), 32150-32163 (2019).
- Reck M, Collins A, Bertoni H, Borri P. Experimental realization of any discrete unitary operator. Phys Rev Lett. 1994;73(1):58.

## 级联MZI网络

- 对于任意一个矩阵，可以进行奇异值分解(Singular Value Decomposition, SVD)，获得特征值和特征向量，分解成三个矩阵相乘
- 分别都可以用级联MZI网络实现

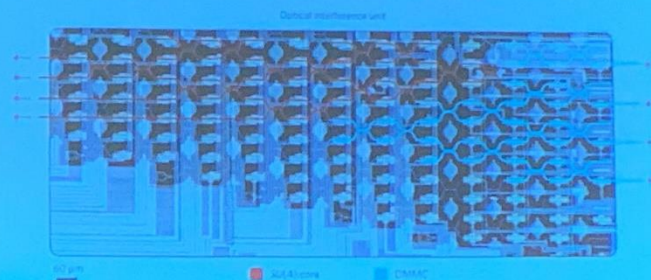
$$\begin{array}{c}
 \mathbf{A} \qquad \qquad \mathbf{Q} \qquad \qquad \mathbf{\Lambda} \qquad \qquad \mathbf{Q}^{-1} \\
 \left[ \begin{array}{ccc} & & \\ & & \\ & & \end{array} \right] = \left[ \begin{array}{ccc} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{array} \right] \left[ \begin{array}{ccc} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{array} \right] \left[ \begin{array}{ccc} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{array} \right]^* \\
 \text{Eigen vectors} \quad \text{Eigen values} \quad \text{Eigen vectors} \\
 \text{of} \quad \text{of} \quad \text{of} \\
 \mathbf{A} \quad \mathbf{A} \quad \mathbf{A} \\
 \text{Unitary} \quad \quad \quad \text{Unitary} \\
 \text{matrix} \quad \quad \quad \text{matrix}
 \end{array}$$

## 级联MZI网络

- 对于一个 $N \times N$  unitary matrix, 网络通常需要至少  $N(N-1)/2$  个MZI基本单元
- 给定要表示的矩阵, 怎样优化网络中各个MZI的参数(训练过程)?
  - 最大梯度下降(SGD)
  - Adjoint variable method
  - 遗传算法
  - 粒子群算法
  - .....

## 级联MZI网络

- 代表性工作: 2016年Nature Photonics上的论文
- 构建了一个包含了56个MZI的网络
- 完成四种不同语音的简单识别任务

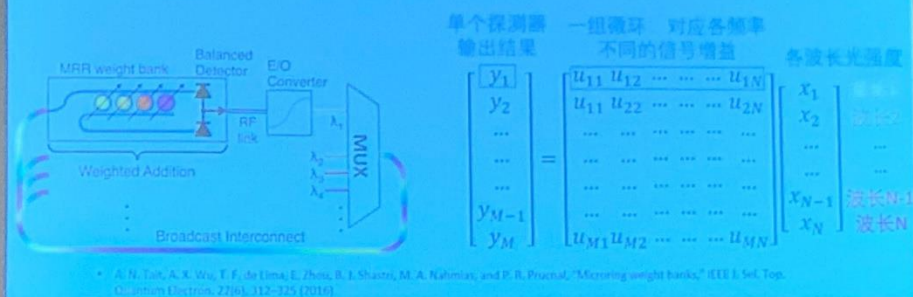


• Shao Y, Harris NG, Shafik S, Prabhu M, Bache-Jones T, Hochberg M, et al. Deep learning with coherent photonic circuits. Nat Photonics. 2017;11(7):441.



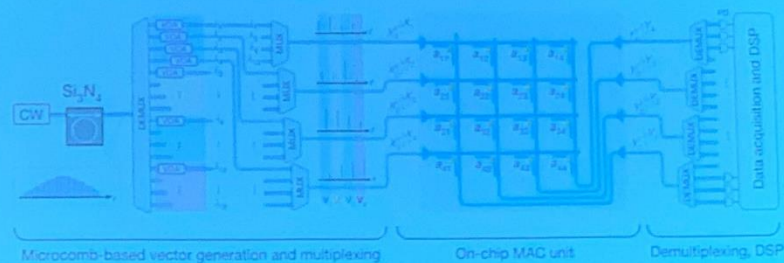
## 波分复用与微环阵列

- 输入信号包含多个不同波长的光，每个波长成分的强度分别对应输入向量中各元素值
- 每一个微环(Microring, MRR)相当于对应频率的滤波器，不同微环的信号增益对应矩阵中一个权重参数
- 每一组微环滤波后的各频率信号总能量用一个宽谱探测器接收，相当于矩阵中一行与输入列向量加权求和



## 波分复用与微环阵列

- 波分复用光计算新研究结果
  - 微频率梳(Microcomb)
  - 可变光衰减器(variable optical attenuator, VOA)
  - 相变材料(PCM)网格

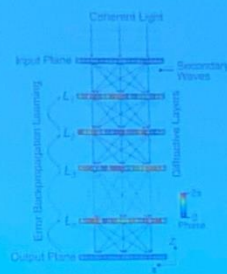


- Taldumnong, I., Youngblood, N., Karpov, M., Gehring, M., Li, X., Mappers, M., Laflamme, M., Fu, X., Lukashchuk, A., Raju, A., Gu, J., Wright, C.D., Sebastian, A., Kippelen, B., Perovic, W.P., Bhaskaran, H. Parallel convolutional processing using an integrated photonic tensor core. *Nature* 2021, 599, 53-58.
- Xu, X., Tan, M., Carpenter, B., Wu, J., Bore, A., Nguyen, T., Chu, S.T., Little, B.E., Hake, D.G., Muralidharan, R., Mitchell, A., Moss, D.J. 14 TOPS photonic convolutional accelerator for artificial neural networks. *Nature* 2023, 619, 44-51.

# 光衍射神经网络

- X: 输入平面光场强度空间分布
- Y: 输出平面光场强度空间分布
- 具有一定空间间隔的多个相位板每个像素都可以优化编码，模拟与一个包含了目标权重参数的矩阵相乘

空间中光场自由传播Rayleigh-Sommerfeld diffraction:

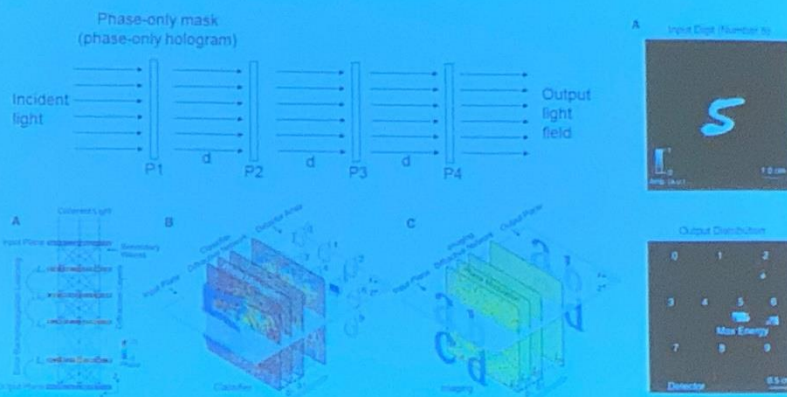


$$w_l^f(x, y, z) = \frac{e^{-jkz}}{z^2} \left( \frac{1}{2\pi r} + \frac{1}{jk} \right) \exp\left(\frac{j2\pi r}{\lambda}\right)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M-1} \\ y_M \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & \dots & u_{1N} \\ u_{21} & u_{22} & \dots & \dots & u_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{M1} & u_{M2} & \dots & \dots & u_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix}$$

# 光衍射神经网络

- 光衍射神经网络进行图像分类



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, 361(6406), 1004-1008 (2018).

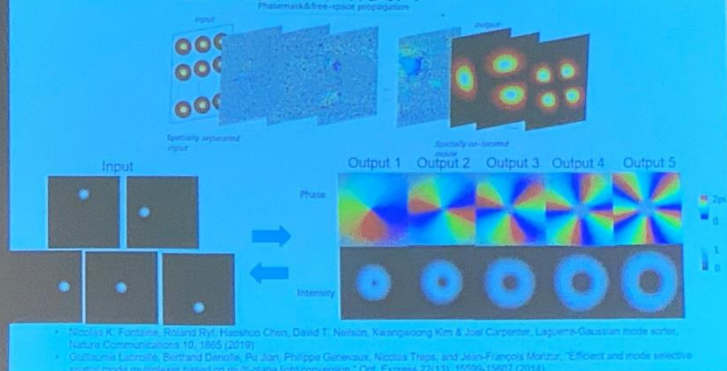


## 光衍射神经网络

- 后续的改进和拓展：
  - 优化算法（误差反传，梯度下降）的改进
  - 不同频率相干光（太赫兹，近红外，可见光）
  - 多波长复用
  - 与透镜结合
  - 多种视觉任务：图像分割，图像显著性检测，图像超分辨，人体姿态识别
  - 其他任务：模分复用光通信，波束整形，激光雷达光束调控

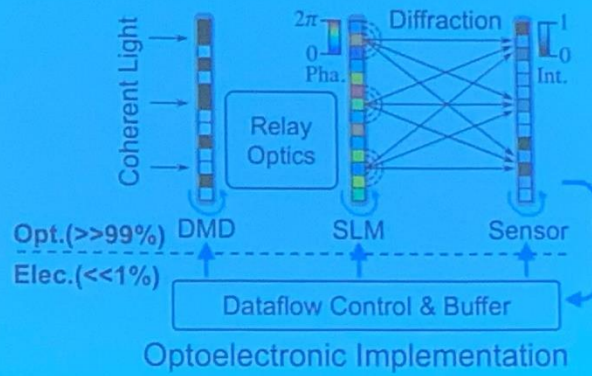
## 光衍射神经网络

- 应用：光模式复用与解复用



# 光衍射神经网络

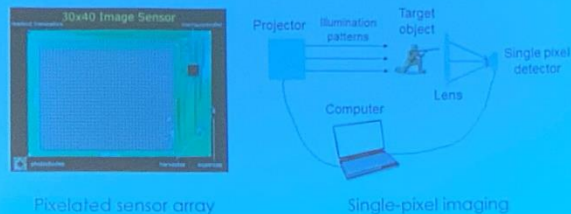
- 光电结合可编程的衍射神经网络系统



\* Zhou, T., Liu, K., Wu, Y. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* 15, 367–371 (2021).

## 单像素成像框架

- 普通相机：面阵传感器
- 单像素成像：传感器只有一个像素，收集物体场景总体光强度



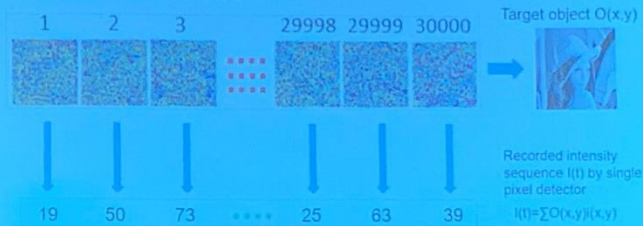


## 单像素成像框架

- 单像素成像框架下的向量矩阵相乘:

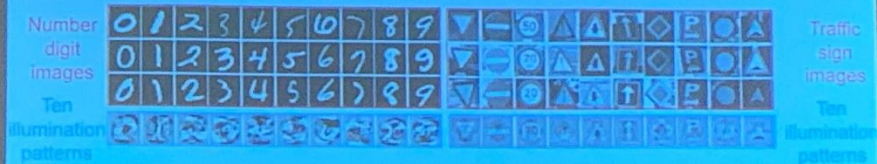
- 物体图像: 输入向量  $X$
- 每个投影图案: 权重矩阵中的一行
- 单像素值: 输出向量中的一个值

Illumination patterns  $I(x,y;t)$  ( $t=1,2,3,\dots$ )



## 单像素成像框架

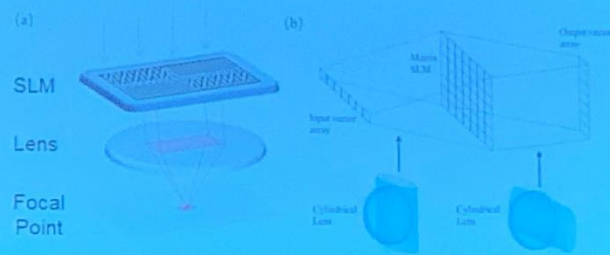
- 单像素成像的线性分类器:



Shuming Jiao, Jun Feng, Yang Gao, Ting Lei, Zhewei Kai and Xiaocang Yuan, "Optical machine learning with incoherent light and a single pixel detector," Opt. Lett. 44(21), 5186-5189 (2019)

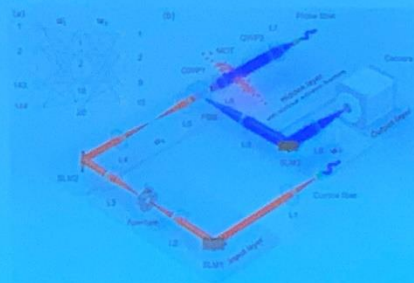
## 单像素成像框架

- 可编程性：物体图像(或输入向量)可以用空间光调制器显示
- 光强度求和时可使用普通透镜，也可使用柱状透镜



## 单像素成像框架

- 相比于之前几种框架
  - 光学器件和系统简单，普通非相干光
  - 系统尺寸的缩小和大规模片上集成难度大
- 但近期工作也展示了基于这一框架可以搭建一个包含了共174个神经元的光学神经网络



• Zuo Y., Tran Y., Chen Y.C., Du S., Xia J. Scalability of all-optical neural networks based on spatial light modulators. *arXiv Preprint arXiv:2102.09722*, 2021.



## 结构可调的光传播介质

- 在玻璃中掺杂气泡（或石墨烯）介质
- 对于掺杂物的形状和分布进行编码
- 光场传播过程中产生不同散射效应



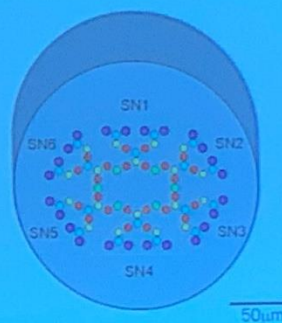
- Khosravi E., Chen A., Liu D., Jing L., Wang Q., Yuan M., et al. Nanophotonic media for artificial neural inference. *Photonics Res.* 2019, 7(8): 823–827.
- Gholami M., El-Ghazali, B. & Feghri, N. A single neuron-designed photonic structure that performs parallel computing. *Nat Commun.* 12, 3456 (2021).

## 多芯光纤

- 利用多个芯之间的信号耦合效应

Network Function

- input cores
- synaptic cores
- neuron cores
- transfer cores
- output cores



- Cohen E., Malka D., Shemer A., Shalmonai A., Zilensky Z., Tandon M. Neural networks within multi-core optic fibers. *Sci Rep.* 2019; 9(1): 1–15.

## 特例：卷积形式相乘

- 卷积：一种特殊的向量矩阵相乘
- 简化：少量参数(卷积核)很多次重复使用
- 可以设计专门针对卷积相乘光学计算框架(不适用于普通向量矩阵相乘)

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{M-1} \\ y_M \end{bmatrix} = \begin{bmatrix} \boxed{1} & \boxed{2} & \boxed{3} & \dots & \dots & \dots \\ \dots & 1 & 2 & 3 & \dots & \dots \\ \dots & \dots & \dots & 1 & 2 & 3 & \dots \\ \dots & \dots & \dots & \dots & 1 & 2 & 3 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{N-1} \\ x_N \end{bmatrix}$$

卷积核

## 特例：卷积形式相乘

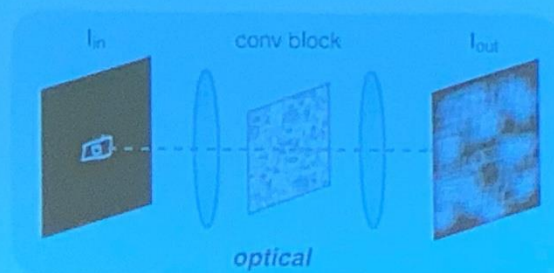
- 深度神经网络中的全连接层和卷积层





## 特例：卷积形式相乘

- 数学上，两信号卷积操作等效于傅里叶域相乘
- A与B卷积： $A*B = \text{IFFT}[\text{FFT}(A) \cdot \text{FFT}(B)]$
- 傅里叶变换和傅里叶逆变换可以通过相干光双透镜4f装置实现

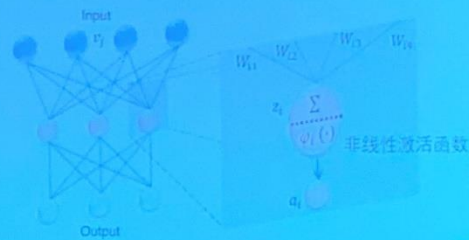
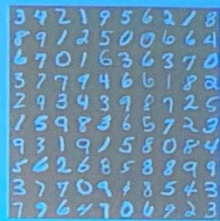


### 特例：卷积形式相乘

- 卷积核信号时域延迟

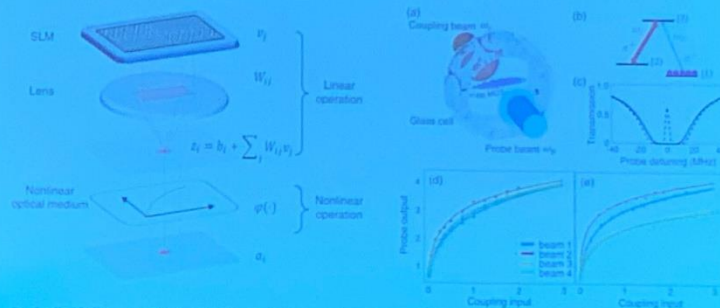
## 非线性函数的光学实现

- 向量矩阵相乘的线性运算虽然可以解决很多问题,但是非线性函数很多时候也必不可少
- MNIST数字图片识别
  - 线性分类器准确率: 91%-92%
  - 包含非线性激活函数的深度学习网络: 接近100%



## 非线性函数的光学实现

- 很多研究中采用光学和电子(或数字)混合的形式
- 全光学非线性激活函数: 电磁诱导透明效应 (electromagnetically induced transparency, EIT)

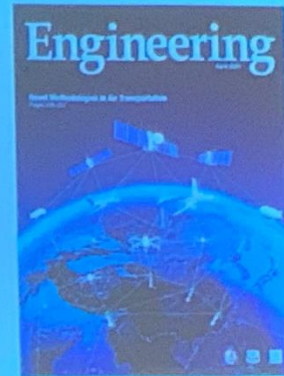


• Zuo Y, Li R, Zhao Y, Jiang X, Chen Y-C, Chou P, et al. All-optical neural network with nonlinear activation functions. *Optica*. 2019;6(9):1132-7.



## 总结

- 即将发表综述论文 Analog Optical Computing for Artificial Intelligence (Engineering 期刊)
- 作者 Jiamin Wu#, Xing Lin#, Yuchen Guo#, Junwei Liu, Lu Fang\*, Shuming Jiao\*, Qionghai Dai\*
- 包括光学向量矩阵相乘、光学非线性激活函数、reservoir computing、脉冲神经网络等内容



## 总结

- 向量矩阵相乘在光学计算和光学神经网络中是一个基本的运算方式，具有广泛的应用
- 近年来很多研究中利用不同的光学系统和原理实现向量矩阵相乘，各有特点
- 简捷的卷积运算具有专门的光学实现方式
- 非线性函数实际可行的光学实现方式相对较少
- 今后研究中光学的向量矩阵相乘相比于电子计算有待表现出更多的优势

Thank you!