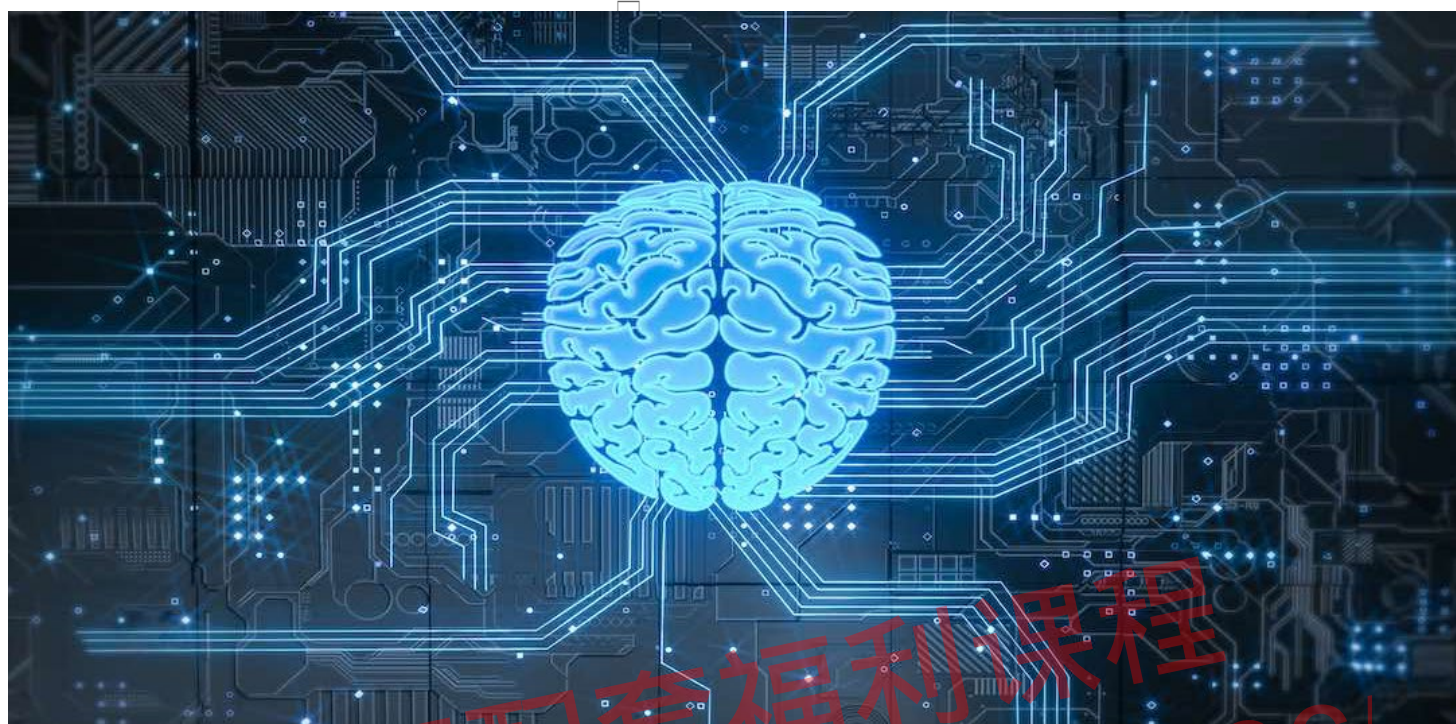




加微信：1716143665，领取配套福利课程



041 | 芯片5：摩尔定律失效后的出路在哪？



卓克·科技参考2（年度日更）

02-17

041 | 芯片5：摩尔定律失效后的出路在哪？

12分35秒

| 卓克亲述 |

众筹新课联系微信：**1716143665**，你好。

欢迎回到《科技参考》，我是卓克。

在昨天的内容里我们说过，当前计算领域存在的几大难题。这些难题可能会导致在 2012 年之后，芯片越来越贵，并且计算性能提升缓慢。

但现在，这些困难已经有了一些突破，那就是类脑芯片的出现。

今天的《科技参考》，我就给你介绍一下类脑芯片是怎么回事。

硬件和软件的双重困境

要说清楚类脑芯片，我们得先回顾一下计算领域的困难。

从硬件上说，因为单个晶体管体积过小，量子效应变得越来越严重。

电子不再按照我们希望的规律流动，于是为了进一步约束量子效应，单个晶体管的结构就从平面走向立体，又从立体走向复杂立体。

而要把几百亿个这样复杂结构的晶体管组合在一起，制造工艺就比从前复杂太多了。今天 5 nm 工艺的芯片，制造步骤是 10 年前的 3 - 4 倍。

硬件上的约束直接导致单个晶体管的成本从过去几十年的每 1 - 2 年下降一半，变成了每年小幅增加。于是，今后的顶级芯片会一代比一代贵。

别看每一代只贵 20%，但按照这个增速，2042 年的顶级旗舰 CPU 就不再是 5000 元了，而是 19 万元。而人们对计算的需求将一直持续下去，那么 50 年、100 年后的成本如何解决呢？

以上是硬件部分遇到的挑战，而软件上的挑战其实也不小。

虽然在大约 10 年前，出现了深度学习、神经网络这样的新算法，让机器在分辨图片、理解语义、自动翻译、下棋游戏等方面的表现大幅提高。但这些提高是不均衡的。

比如，当前机器在棋类游戏上的水平已经远远超过了人类，但在自动驾驶、理解语义上的表现，只能说接近人类，远比不上这方面的人类专家。

怎么才能让计算机在这些薄弱环节上的表现也超过人类专家呢？

现在看来，改进算法这条路好像突破空间不大。因为就在不久之前，这条路刚刚出现过重大突破。于是今天，不论技术专家还是资本巨头都在另外一条路，也就是在堆设备、堆算力上下功夫。

有钱当然可以这么堆下去，但软件上遇到的问题是：想让性能提升 10 倍，通过堆算力这条路付出的成本可不是增加 10 倍，而是可能需要增加 10 亿倍。

让机器拥有当前性能所付出的成本，已经是数以亿计的美元了，如果再增加 10 亿倍，那就完全超出了人们的承载能力。

而且，更硬的限制来自于能源消耗。实际上，利用当前算法把软件性能提升 10 倍，那个能耗可能是地球上几十年来能源消耗的总和。且不说有没有这么多的能量，就算有，消耗过程中附带的杂质排放、碳排放，也足够让生存环境彻底崩溃。

出路：跳出冯·诺依曼结构

在硬件和软件都卡住的时候，出路在哪呢？

这就需要我们跳出当前芯片的结构，也就是从上世纪 40 年代后期就一直延续至今的冯·诺依曼结构。

对于这个结构，如果你上计算机课或者查百科，得到的结果是这样的：首先，会告诉你计算机分为运算器、控制器、存储器、输入输出设备；接着，定性的给你介绍这几个模块都是做什么的。

但如果我也这么说，你就很难理解从这样的结构里跳出来有什么好处。所以，我要根据类脑芯片的特点，重新说一下冯·诺依曼结构。

冯·诺依曼结构依托于天才数学家图灵的证明，最后被冯·诺依曼在现实中实现了。

当年，图灵发明了一种操作，可以模拟一切逻辑运算。这种操作由控制器、读写头和一条无限长的纸带组成的机器来完成。

纸带用来储存信息；读写头可以读取纸带上的信息，还可以把运算结果写在纸带上；控制器则用来左右移动纸带或者擦除当前读写头上的数据。于是，只要左右拉动纸带的动作和读取纸带信息的动作配合得足够合理，中间过程不需要人来插手，就能在有限次操作后得到逻辑运算的数值结果。

此后，一些计算机科学家的的工作就在于，思考要完成加法运算，应该如何安排拉动纸带和保存中间结果的动作顺序呢？甚至，要完成双曲余切函数的计算，要如何安排动作顺序呢？这些计算机科学家，逐渐演化成了今天的算法工程师和程序员。

而另外一批计算机科学家的的工作，是研究如何把纸带加读写头的小单元缩小再缩小，让 200

亿台这样的设备同时工作，这样效率不就高得多了吗？这些计算机科学家，逐渐演化成了今天芯片设计制造的科学家与工程师。

人类在 70 年前遇到的计算任务，用笔和纸顶多配上一些计算尺就能完成。于是当冯·诺依曼结构出现后，随着算法的丰富、计算模块的缩小，70 年时间里的前 60 年，都维持了稳定高速的发展。

但后 10 年，速度却降下来了，因为软件和硬件上都遇到了我一开始提到的瓶颈。

在软件上的问题，可以叫作瓶颈，因为人们或许还能发现一些再把运算性能提升几千上万倍的方法。

但硬件上的问题，无论如何都不能叫瓶颈了，而是撞上了天花板——芯片工艺不能无限缩小下去，1 nm 往下要怎么做，没人说得清，而现在，我们已经走了 3 nm。

在众多解决方法中，类脑芯片是最有潜力的。这个思想早在 1990 年就出现了，突破不在硬件，也不在软件，而在基础计算结构的大换血。也就是说，不再使用冯·诺依曼结构了，于是冯·诺依曼结构注定会遇到的软硬件天花板就不存在了。

类脑芯片的灵感源自人脑。我们知道，人脑肯定在做大量运算，而且运算复杂度远超人造计算设备，运算性能在很多方面也大幅超越人造计算设备。

所以，如果能从人脑上借鉴一些新结构到人造计算设备中，也许可以有所突破。

这方面的研究分为两派：

一派认为，要做新的计算结构，首先要了解人脑的工作原理。这一派，到今天都还困在泥潭里。因为人类研究了一百多年，至今对人脑的基本原理还是两眼一抹黑。

但这一派坚信这条路是对的，因为冯·诺依曼结构当初就是这样自上而下出现的——先由图灵从数学上证明了操作的正确性，再由冯·诺依曼从实施上找到办法。然后计算机行业就像核弹爆炸一样爆发了 70 年。

而另外一派认为，不必对原理了解透彻。先用当前已有的生物学知识，比照人脑的基础单元做出一些结构用起来，看能做什么再说。

这一派也是有先例的。比如，莱特兄弟第一次飞上天之后 20 多年，冯·卡门的空气动力学理

论才完整解释了飞机能飞起来的原理。

再比如，一战、二战里用得火热的无线电，其实是靠地球大气层边缘的电离层多次反射电磁波，才能让几千公里外的人接收到信息。这个原理是在无线电被发明出来 70 多年后，才被解释清楚的。但之前，人们没有因为不透彻了解原理而拒绝使用，都是先用着再说。

所以，第二派依照神经元的基础结构，给类脑芯片做了一些数学描述，也搭建了模型，并且做出了不少可以运行的芯片。于是，今天第二派暂时大胜第一派。

相关的数学描述，在一些和 AI 芯片相关的书里都可以查到，我就不具体说了。大致描述是这样的：

芯片的一个基础单元就类似于人脑中的一个神经元，它是一个小圆球。有几根长长的线和远处的神经元相连。

在脑科学中，这个长长的连线叫作“轴突”。此外，小圆球表面还有一些稀疏的毛毛。它们不一定都有用，但当一些条件满足的时候，其中几根毛毛会和其他小球伸过来的连线握上手。这些毛毛在脑科学中叫作“树突”。

而握手也不是传统电路里导线连通的概念，而是类似于脑科学中“突触”的概念。也就是说，伸过来的轴突和本地的树突并没有实际握上手，而是存在一个间隙，大约几微米。

在人脑里，信号在单个神经元中的传递，是通过放电的形式进行的；但跨越多个神经元的信号传输，就要靠突触上分泌的化学物质，也就是神经递质。

只有神经递质的浓度足够高的时候，才会导致握手的另一端的神经元被高度激活，开始放电；当神经递质的浓度不够高的时候，就不激活后续的神经元。

之所以利用神经递质，而不用电信号直接连接，是因为电信号在生物组织中传输的损耗太大了。如果距离超过几十厘米，电信号就衰减到几乎没有了。这是生物演化中出现的特征。

而放在类脑芯片的数学描述里，“突触”要换成数学中“权重”的概念，“神经元是否能被激活”要换成数学中“阈值”的概念，而神经元被数学中的激活函数描述。

类脑芯片和传统结构的差异

原理部分先到这里。可能有点儿烧脑，但不理解也没关系，我们只要知道它是模仿人类大脑

搭建的就可以了。

我们的问题是，这些新模型和传统的冯·诺依曼结构有什么不同呢？

首先是，存算一体结构。

这是一个什么东西呢？在这里，我只简单的把它形容成一个和冯·诺依曼结构平行的另外的结构。它的读取和存储计算是在同一个单元中同时完成的，这是最基本的结构上的改变。至于具体是怎么回事，我们放在明天的内容里单独讲。

其次是，单元与单元间的连接变了。

一个输入是否会让多个单元产生动作，很像神经元是否会被激活。当神经递质不够浓，没有足够强的放电时，那个神经元就不活跃，或者说没有达到阈值。这样的特性被称为“事件驱动”。

今天，这样的芯片已经应用在动态视觉分析上了。比如，在自动驾驶的过程中，有些判断并不需要全景图的数据，只要知道当前有哪些物体正在运动就可以了。

传统的芯片会分析整个视频画面，而类脑芯片中，只有那些活动的像素才会激活处理过程。只分析这部分活动的内容，计算量自然大大降低，于是芯片功耗大大降低。这个特性，有点像青蛙的视觉处理过程，它们也只是对动态物体更敏感。

功耗大约会降低多少呢？

我们可以参考当年李世石大战 Alpha Go。那一年，他们的胜负在伯仲之间，但李世石的大脑功耗大约是 20 瓦，Alpha Go 大约是 100 千瓦。

尽管今天 Alpha Go 实现同样的性能，功耗已经降低到几百瓦了，看起来和人脑的 20 瓦相差不多。但实际上，人脑那 20 瓦功耗并不只是在处理围棋的走子策略，还在处理视觉、听觉、平衡、维持情绪的平稳等比围棋复杂得多的任务。

所以，人脑依然比传统计算机高效得多。类脑芯片就有可能在很多类似的计算中，大幅减少功耗和计算量。

类脑芯片是一种新结构，一旦普及，就将造就另外一次“计算机时代”。

关于类脑芯片，今天咱们就介绍这么多。明天，我详细说说它的一个重要变革——存算一

我是卓克，我们明天再见。

著名科普作者
卓克

众筹新课联系微信: 1716143665

5

9

642945106 “ ” “2”



公开

仅限群内使用！严禁商业！

默认

最新

只看作者回复



常觉不住
02-17 编辑

关注



时下所谓的 AI 芯片大致分为两大类：深度学习 AI 芯片和类脑芯片 (Neuromorphic chip)，虽然它们都源于人工神经网络，但是它们涵盖了两种神经网络计算的概念，前者是用深度神经网络 (Deep Neural Network, DNN)，后者是使用脉冲神经网络 (Spike Neural Network, SNN) 概念。

与 DNN 相比，SNN 有诸多优势。例如，由于 SNN 中的信号强度不是由脉冲幅度来表示的，而是以恒定幅度的信号的时间宽度或时间间隔表示，另外 SNN 的功耗很大程度上取决于激发和编码策略的统计信息，如果使用有效的编码策略，由于脉冲的稀疏性，整个系统的功耗会相对低得多。

SNN 的训练方法是无监督学习，也不需要反向传播，只需要单个数据样本，还可以通过少量数据进行学习，并可以进行在线学习和终身学习等。

IBM 于 2014 年在《科学》杂志上发表了 SyNAPSE 类脑芯片，SyNAPSE 具有 256 个神经元和大约 260,000 个突触。后来，IBM 将 4096 个 SyNAPSE 这样的类脑芯片作为内核并行集成在一个芯片中，制成了被称为 TrueNorth 的第二代数字芯片，目标是实现极低功耗的大型网络处理。

英特尔于 2017 年 9 月宣布了一种专门为 AI 设计的神经形态芯片 Loihi，为类脑芯片迈出了重要的一步。Loihi 芯片具有 1024 个人工神经元或 130,000 个模拟神经元，具有 1.3 亿个可能的突触连接，并结合了脉冲时序相关的突触可塑性模型，适合处理大规模的 SNN 数据。

英特尔又于 2019 年推出了类脑芯片系统 Pohoiki Beach，它包含 4 个 128 核心、14 纳米制程的 Loihi 神经形态芯片，是一款模拟 800 万神经元的 64 芯片

计算机。2020 年 3 月，英特尔再度宣布完成最新的神经形态系统 **Pohoiki Springs**，可提供 1 亿个神经元的计算能力。**Pohoiki Springs** 是一个数据中心机架式系统，是英特尔迄今为止开发的最大的神经形态计算系统。它将 768 个 **Loihi** 类脑芯片集成到一个具有 5 台标准服务器大小的机箱中。**Pohoiki...**

Springs 具有解决各种计算难题的潜力。
总部在澳大利亚悉尼的 **Brainchip** 于 2019 年 10 月发布了 **AkidaNSoC** 类脑芯片，每个 **AkidaNSoC** 芯片具有 120 万个神经元和 100 亿个突触，该芯片使用台积电的 28nm 和脉冲转换器技术，可以根据音频、图像、激光雷达、压力、温度和其他传感器数据，以及互联网数据包和多元时间序列数据生成脉冲信号。

BrainChip 于 2022 年一月十八日宣布 **Akida AKD1000** 芯片的商业化，**BrainChip** 对外声称它们是世界上第一家将类脑芯片应用在物联网 (IoT) 上的产商，在今年早些时候的 **CES** 展示会上，宾士的 **EQXX** 概念车，就使用了 **Akida AKD1000** 的芯片来控制车内语音系统，和传统芯片相比，**BrainChip** 的类脑芯片的功耗是传统芯片的十分之一。

展开

- ☐ 11
- ☐ 10
- ☐ 182
- ☐ 分享



传统上，计算机芯片都是依循冯·诺依曼架构而设计的，存储与计算在空间上分离，计算机每次进行运算时需要在 **CPU** 和内存这两个区域往复调用，可以说 99% 的计算时间和能耗都花在了数据的传输上，真正用于计算的只有 1%。此外，芯片在工作时，大部分的电能将转化为热能，导致功耗提高。

☐ 关注 ☐

这个模式跟我们大脑的运算完全不一样。大脑的运算模式叫做「存算一体」，都是在神经元里面进行的。类脑芯片的研究就是基于微电子技术和新型神经形态器件的结合，希望突破传统计算架构，实现存储与计算的深度融合，大幅提升计算性能、提高集成度、降低能耗。

2011 年 **IBM** 发布了「真北芯片 **True North**」，这也是人类用电路模拟神经行为学的开端。2014 年 **True North** 更新了第二代，功耗达到了平方厘米消耗 20 毫瓦，印证了类脑芯片的低功耗价值，也在一些 **AI** 任务上印证了类脑芯片的实际工作能力。

紧随其后的是英特尔，2017 年发布了类脑芯片 Loihi，其拥有 13 万个人造突触。除此之外，业界比较出名的类脑芯片还有高通的 Zeroth。

而中国浙大的「达尔文」芯片、清华的「天机」芯片都已经在路上。但类脑芯片距离真正确立产业价值，从实验室步入现实世界，还有很长的路要走。

展开

- ☐ 6
- ☐ 8
- ☐ 140
- ☐ 分享



钰箭风

02-17 编辑

类脑芯片就是把神经网络实现其中的芯片，目前的人工智能多数就是通过神经网络算法实现

☐ 关注



“人工智能的神经网络最初就像一个拥有“超级大脑”的婴儿，它可以通过学习外部信息，不断调整“神经元”的连接方式、优化参数，自己实现进化。”

神经网络是目前人工智能的主流，他需要大数据的投喂，也就是所谓的学习培训过程。

培训完成后，神经网络的连接记忆完成了，就可以用来干活啦。

神经网络算法目前分为 3 大类

★1、全连接神经网络 (FNN)

每一层是全连接层 — 即每一层的每个神经元与上一层所有神经元都有连接；

场景，所有的神经网络都可以使用

缺点是，全连接的神经节点需要足够大的连接关系，每增加一层都是指数级的连接增量。...

★2、卷积神经网络 (CNN)

不过神经网络只能对特定训练产生结论，普适性目前来看还比较差，难以成为使用最普遍

场景：人脸识别、图片识别

能下围棋赢了世界冠军的 AlphaGo Zero，不会动弹，落子需要人协助。

★3、循环神经网络 (RNN)

场景智能用以挺弄填瓷，有时序列工蔬含视羽等髫能数据

每是梯神经网络种类的数据在能能酶场对应的髫能能用的应用选择不同的网络

也可能同时用到三种网络搭建更复杂的网络。

类脑芯片就是把神经元固化其中，所有连接权重由计算人物训练而生。因此可

以统一智能硬件，但目前还难以使用单一芯片实现多种任务需求。

展开

- ☐ 6
- ☐ 7
- ☐ 104
- ☐ 分享



类脑芯片参考了人脑能耗低，和传统芯片底层架构不同，可以解决传统芯片遭遇天花板的问题。

☐ 关注 ☐

虽然人还搞不懂大脑原理，但是可以模仿利用。对比我深表赞同，像我小时候做题都是先看看参考答案，虽然不是很懂为什么，但是参考着可以做出来，做多了就慢慢懂啦。

感觉芯片中的量子效应是距离我生活最近的。在微观层面，比如在纳米级别...

- ☐ 转发
- ☐ 1
- ☐ 30
- ☐ 分享



目前人工智能的主流研究思路有两种。一种是从计算机科学的角度出发，借鉴大脑的分层处理机制与学习训练特性，基于冯·诺依曼架构的计算机和人工神经网络与深度学习方法；另一种是从脑科学研究出发，采用纳米级别的器件模拟生物神经元和突触的信息处理特性，采用非冯·诺依曼架构的神经形态芯片和脉冲神经网络。两个方法各有优缺点，将二者融合是目前公认的最佳发展路径之一。...

☐ 关注 ☐

展开

我爱问卓克

- ☐ 2
- ☐ 1



Wei

02-17

专题讲到这里太让人兴奋了，我自己就是卓老板讲到的工程师中的一类，但是从未从这个高度思考这个算力效率问题。搬好小板凳认真听讲。

☐ 关注



感觉信息传递的准确性和功耗是一对核心矛盾，不知道怎样的新架构可以上一个层次，做到更优的平和组合，在需要准确的时候花高能耗，只需定性时又降下来功耗。

另外特别想知道类脑芯片结构是用什么制造工艺怎么做的。

作者 回复：

初期肯定还是希望基于现有半导体硅基芯片技术优化来制造



Shan

02-17

图灵和冯诺依曼的架构从今天内容的视角看就是存算分离，时钟驱动，总线连接。请问卓老板，类脑芯片是不是要打破了单机里全局时钟，靠输入输出设备的反馈驱动芯片运算？

☐ 关注



当今计算机擅长的领域普遍和环境互动少，例如下棋，科学计算等工作，这些工作都可以用数学描述。自动驾驶，理解语义都和环境或者上下文相关，冯诺依曼架构下，需要把环境和上下文，全部表达成二进制，然后用存算分离的...

作者 回复：

还要看后面的算法研发，现在的应用面很窄



毕昊峰

02-17

对类脑芯片原理，“存算一体”从字面上理解了，但单元与单元之间的连接却没有太明白。但课中所举的案例：莱特兄弟第一次飞上天之后 20 多年，冯·卡门的空气动力学理论才完整解释了飞机能飞起来的原理，倒给了我一些经济方面的启迪，那就是市场永远是正确的，如果对一些经济现象无法理解，暂时不用去想过多的因果关系，待云淡风轻时，再去尝试解读背后的经济学原理。

关注



我爱问卓克

转发

评论

19

分享



Joe

02-17

感謝卓老師對於創新晶片架構的簡介，另外對於金屬 3D 打印的進展，不知老師覺得是否有實用性跟商業規模化的價值，富士康郭董曾經說若 3D 打印能商業化，他的姓可以倒過來寫，但我看對於航天工程或汽車產業已經有越來越多的應用，美國這兩年 spac 上市的打印公司也不少，譬如 DM,MKFG,FATH,RKLB,VLD 等，新創公司 relativity space 甚至在下個月底即將試射 3D 打印的火箭 Terran 1, 想問問老師的看法，謝謝

关注



1

1

17

分享

作者 回复：

我觉得他那种说法并不是对行业的严肃解读，只是表达一个意思，能够量产的产品离使用3D打印技术还很远



水镜
02-17

所有的系统在本系统内都会碰到一个在质量，效率，成本 这三者之间写入一个悖论。如果追求每个要素极致的好的话，就是一个戏谑称之为“滚”模型。那么两两交集：质量与效率的交集：就是贵。效率和成本的交集：就是差；成本和质量的交集：就是慢。如果取三者共同的交集：就叫做“滚”。

也就是说在本层次本领域内，不存在万事大全的解决方案，哥德尔不完备定律，也说本领域内一定有一个问题是在本领域无法证明和解决的。如要解决...

展开

关注



- 1
- 1
- 16
- 分享



金戈铁马
02-17

学习了今天的课程，有一个问题想要请教卓克老师。如果说类脑芯片是根据人类当前已经掌握的关于大脑的生物学原理，对芯片的基础架构做出改造；那么，基于我们当前对于大脑的有限认知，是否意味着类脑芯片目前也只能应用于少数的特定领域呢？而类脑芯片当前的研究方向，是否也只是集中在如何减少计算量，大幅降低功耗的方面呢？而类脑芯片想要实现突破性的进展，是否还是得取决于我们对于人类大脑机理的研究能够取得多大进展呢？想听听老师...

关注



- 转发
- 1
- 16
- 分享

作者 回复：

目前就是集中在我这期说的，向量乘法



佛祖门徒

去年 9 月的 **Nature** 子刊上提到了三星联合哈佛大学就提出的一种设计类脑芯片“简单粗暴”的方法：复制粘贴。

[关注](#)

具体而言，就是通过将大脑神经元连接图，“复制粘贴”到固态存储器的高密度 3D 存储网络上。

如此以来，就有望实现低功耗、易学习、适应环境，甚至是具备自主性和认知能力等特性的芯片。为了实现这个目标，研究人员通过使用纳米电极阵列和存储芯片对神经元连接进行“复制、粘贴”。这里的纳米电极阵列采用的是此前该团队在哈佛大学的研究成果：CMOS 纳米电极阵列（CMOS nanoelectrode... array, CNEA）。实验中，首先通过电镜下 CMOS 纳米电极阵列模仿的大鼠神经元，然后通过计算机辅助分析程序从该 CNEA 上提取出的突触连接图。然后，“粘贴”到存储网络中，通过对每个记忆进行编程合成记忆网络，使存储器能通过电导（conductance）反应每个神经元连接的强度来实现。或者采用直接下载连接图到存储芯片上的方案，即用记录的信号直接驱动一个 $N \times N$ 的存储器交叉杆阵列，进行突触连接图的物理压印。

展开

2

评论

16

分享



乌尚书

02-18

关于关于芯片的极限，之前听铁夫老师和孙亚飞老师的课程，知道在尺寸上继续缩小达到极限后，也在通过改变芯片材料或者结构来实现芯片性能的进一步扩大。

[关注](#)

卓老板这次讲的内容就是一个更宏大的前景，直接改变了底层结构，用新的结构来构建新一代的计算机，从而也就需要新的芯片和软件了。

一说起类脑芯片，就发现在电脑出现以后，人类一直恐惧着被电脑取代或者...

展开

转发

评论

15

分享



卷心菜

02-17

1 - 今天的内容有点烧脑，听得很吃力。

关注



为了提高计算机芯片的性能 需要将芯片的架构进行革命性的改进 打破以前的冯诺伊曼架构

2 - 学习人脑的思维和判断方式，以求达到更高效又更节能

3 - 有一个小问题：人脑虽然思维方式都相同，然而有的人天赋就很高 有的人愿意经过刻意练习而能达到一个高度，人和人最终的结果差异是非常大的。

4 - 那么计算机模仿人脑的思维判断方式 他是朝着各个方面都学习，成为通才计算机 还是只能专注于某个领域，成为某个领域专业的计算机呢？

5 - 如果未来的计算机是否会越分越细？而面向大众的计算机会是相当于什么人类的大脑水平呢？

6 - 它们会不会模仿人类，从而反过来改变人类普遍的思维方式呢？

展开

我爱问卓克

1

评论

14

分享



沛文沛语

02-17

冯诺依曼结构的计算机的基础元器件是三极管，从电子管逐步发展到晶体管，但本质上都还是三极管。类脑计算机的基础架构不是冯诺依曼架构，其基本元器件也不是三极管，目前比较热门的一个方向是忆阻器作为类脑计算机的基础元器件。1971 年在美国伯克利加州大学任教的华裔科学家蔡少堂就首次从理论上预言：除了电阻、电容和电感之外，电子电路还应该存在第四种基本元件——忆阻。直到 2008 年，忆阻器才被惠普公司实际制作出来，比晶体管的...

关注



展开

- ☐ 1
- ☐ 1
- ☐ 13
- ☐ 分享



潘飏
02-17

类脑芯片，应该也只会部分领域代替传统计算机吧，毕竟传统计算机在精确计算领域是远胜人脑的

☐ 关注



- ☐ 转发
- ☐ 评论
- ☐ 13
- ☐ 分享



张洋
02-17

类脑芯片是从架构上模仿人脑神经结构的芯片，与当前 AI 芯片普遍作为神经网络算法加速器不同，前者模仿神经结构从底层构建人工智能，后者则模仿神经处理信息的功能流程。IBM 在 10 年前就开始类脑芯片的研究，主要基于脉冲神经网络（Spiking Neural Network, SNN），通过脉冲的频率或者时间在神经元之间传递信息，而不是通过节点之间的权重。

☐ 关注



这种芯片把数字处理器当作神经元，把内存作为突触，跟传统冯诺依曼结构不一样，它的内存、CPU 和通信部件是完全集成在一起。因此信息的处理完全在本地进行，而且由于本地处理的数据量并不大，传统计算机内存与 CPU 之间的瓶颈不复存在了。同时神经元之间可以方便快捷地相互沟通，只要接收到其他神经元发过来的脉冲（动作电位），这些神经元就会同时做动作。

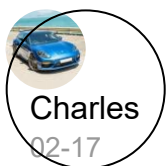
该类芯片优点是能耗非常低，且有较好的感知能力；缺点则是其缺乏高效的深度学习训练和应用方法，脉冲神经网络精度在精度上不能和机器学习类的神经网络相比。

展开

- ☐ 3
- ☐ 1
- 13



分享



Charles

02-17

点燃不同的科技树，就会得到不同的科技成果！

关注



现在的新能源方面就同时在点几种科技树：锂电池、钠电池、镁电池、锌电池及氢燃料电池，这些科技路线的不同，确保了将来发展方向的多样性，之所以出现这么多路径之争，也和每个国家的自然资源、人力资源等有一定的关系。

冯·诺依曼结构经历了 70 年的发展，已经遇到了瓶颈，这件事和人类在陆地上的行进速度不断的更新工具有相似性。两条腿奔跑，不如借助轱辘速度快、驱动轱辘用牛羊不如用煤速度快、用煤不如用石油速度快、和地面接触有摩擦、那就脱离地面用飞行器，速度也同时变快了！

发展遇到瓶颈，改变其中的一个或者多个变量，是有机会突破这个发展瓶颈的！

展开

我爱问卓克

转发

评论

13

分享



山人行者

02-17

处理信号的方式决定处理效率，在不改变方式的情况下，效率总会遇上极限墙的。

关注



二进制是一项伟大的发明，通过二进制一下子让我们对电信号控制能力的潘多拉魔盒被打开了。尤其是在硬件技术相当有限的昨天，能够实现对信号的精准表达与控制，实在了不起。

冯洛伊曼结构是也是在硬件技术条件非常有限的情况下对信息存储、运算非...

展开

- ☐ 转发
- ☐ 评论
- ☐ 11
- ☐ 分享



MakerJA

02-17

类脑芯片成熟后，是不是真的可以模拟脑信号，从而复制人的大脑呢？到那时人是不是在机器上实现了永生？

☐ 关注



- ☐ 转发
- ☐ 1
- ☐ 10
- ☐ 分享

作者 回复：

那要看是怎么成熟的，如果是今天主流的方式长出来的，那和大脑工作的真正原理还差很远



Vincent文申

02-17

去年阿里巴巴达摩院也研究一款存算一体的人工智能芯片，性能比行业均值提升 10 倍，能效提升 300 倍，大量节省了在数据传输上的能耗。

☐ 关注



宅家的欢乐时光

- ☐ 转发
- ☐ 评论
- ☐ 9
- ☐ 分享



湖心亭

02-17

既然 CPU 无法通过无限做小计算单元来提高速率，那么是否可以通过适当扩

☐ 关注



大 CPU 的体积来提升计算单元的总数呢？

- ☐ 转发
- ☐ 1
- ☐ 8
- ☐ 分享

作者 回复：

可以，但今天已经把“适当扩大”中的“适当”充分利用了，再扩大，最难处理的是发热太大



Aming
02-17

仅从家用角度谈谈看法。冯诺伊曼 40 年代就提出了目前计算机架构。但这种架构在中国普及到千家万户应该在起码 2000 年之后。而且现在对大部分个人来说，计算机升级的主要目的只是游戏。只有少部分人将机器用于科研。如果类脑芯片缺乏游戏应用，家庭普及前途漫漫。

☐ 关注 ☐

- ☐ 转发
- ☐ 评论
- ☐ 8
- ☐ 分享



无量靠谱
02-17

冯诺依曼结构的好处是通用，实现了软件、硬件的分离，代价就是时延和能耗，而这对人工智能极不友好。

☐ 关注 ☐

类脑芯片的解决思路，让我想起吴军老师在《信息论 40 讲》中提起的机器翻译的突破。

关于机器智能的争论由来已久，争论的激烈程度愈演愈烈，图灵的自杀多少与此有关。

在贾里尼克之前，由语言学家、逻辑学家主导的研究认为识别语音是一个智力活动，试图让计算机学会构词法，能够分析语法，理解语义。但是相关研究早

起有一点进展，之后就长期停滞。

贾里尼克认为语音识别是一个通信问题，他基于香农的信息论，用通信的编解码模型，以及有噪音的信道传输模型，构建了语音识别的模型。这个模型需要海量数据才能算出模型参数，于是要做的就是收集数据，训练各种统计模型。

有趣的现象出现了，随着团队语言学家，生物学家不断被裁掉，研究却不断取得突破。在短短几年时间里，由数学家和理论物理学家组成的团队，就将语音识别的规模扩大到 22000 词，错误率降低到 10% 左右。

类脑芯片也是试图用仿生学的思路解决芯片的智能计算问题，问题的关键是不是冯诺依曼结构呢？
一个朴素的认知是，人脑的计算不是二进制的。

展开

- ☐ 1
- ☐ 1
- ☐ 7
- ☐ 分享



按照线性思维的惯性，促进芯片产业的发展，无非就是增加单位面积的晶体管数量，升级软件的运算能级。

☐ 关注 ☐

但世界从来就不是非黑即白的。在日常领域适用的牛顿力学，到了量子领域就失效了。
...

- ☐ 转发
- ☐ 评论
- ☐ 7
- ☐ 分享





不周山

02-17

这期点挺多的；

关注



电脑算法能力发展不均衡，在棋牌语言上发展迅速，视觉识别，行走上进展缓慢；这可不见得是算法发展不平衡，而是我们假设棋牌语言和视觉识别、行走是同等复杂；回想一下，棋牌语言出现了多少年，是不是人生存的必须技能？而视觉识别和行走可是实打实的生死悠关的技能，早在人出现前就由不同种族不同形式不间断的演化上亿年，我们是不是太低看这些演化成果了？ ...

展开

1

1

6

分享



王学敖

02-17

老师，我想知道

关注



- 1、类脑芯片的算法，是不是要建立在完全理解“人脑工作原理之上”？
- 2、真的搞懂“人脑工作原理”是不是会带来很大的伦理和风险问题？
- 3、量子计算机本质上还是冯氏结构吧？

转发

1

6

分享

作者 回复：

- 1 不需要，你看今天的神经卷积也不是等价于神经元的工作机制，只是借用其中一些逻辑
- 2 没有必要在具体东西还没出现时想这些
- 3 不是



娃娃

02-17

您聊到的类脑芯片让我想到了学习外语。非专业语言类学习者不必深挖语形、语义、语用这些专业理论知识。只要依照初步的造句规则，大量重复练习，获得有效反馈，继续重复练习就好了

[关注](#)



- ☐ 1
- ☐ 1
- ☐ 4
- ☐ 分享



牛奶爱玉

02-17

老师您好，请问一下，为什么利用软件让性能提升 10 倍的成本与能耗会暴涨那么多？

[关注](#)



- ☐ 转发
- ☐ 1
- ☐ 3
- ☐ 分享

作者 回复：

这个之后我们有一篇文章专门介绍这个研究



AI架构师易筋

02-17

AI 神经网络的原理，一般分为监督学习，非监督学习和强化学习，大多数属于监督学习。监督学习，在于通过输入数据 X ，和结果数据 Y ，大量喂给机器，机器能够总结出规律姑且认为是模型 **model**。那么就可以预测新数据了，比如输入新数据 x_{new} ，通过 **model**，计算出新结果 y_{new} 。类脑计算机结构，存算一体化，有点违反单一职责原则，因为越简单越清晰会越通用。期待明天的未来计算机畅想。

[关注](#)



- ☐ 转发
- ☐ 评论
- ☐ 3
- ☐ 分享



马踏飞燕

02-17

生物芯片是奔着低功耗去的，为的是节约能源。

☐ 关注



可是，能源问题也可以从节流和开源两个方向来考虑，可控核聚变一旦可以商业化运作，那么能源问题就解决了。

另外，从文中对生物芯片模拟神经元的控制机制来看，跟场效应管（MOSFET）也有异曲同工之妙。递质的作用就跟加在 **Gate** 上的电压一样，达到一定的量就能使 **P-N** 结导通，而这个开关的速度非常快，肯定比化学反应快上 **N** 个数量级的。

类脑芯片既然是模拟神经网络的结构和运行机理，那么这些生物芯片是否需要“睡觉”呢？

☐ 转发

☐ 评论

☐ 3

☐ 分享



轻描淡写

02-17

卓克考试，能不能讲一下碳基芯片？

☐ 关注



☐ 转发

☐ 1

☐ 2

☐ 分享

作者 回复：

之后有碳纳米管晶体管的内容

1

2

3

4

5

6

7

8