

Discussion on the Future Development of Quantum Computing and AI Hardware

Xiaoyin Wang

Teradyne, Shanghai, 201206, China

Abstract

With the rapid development of artificial intelligence (AI) technology, especially the increasing scale and complexity of deep learning models, modern society has put forward extremely high requirements for computing power. Modern AI models (such as GPT-4, BERT, etc.) have billions or even hundreds of billions of parameters, and need to process massive amounts of data and perform complex computing tasks. Traditional CPUs are less efficient in handling these tasks and struggle to meet the needs of training and inference. However, quantum computing, as a cutting-edge technology with disruptive potential, has brought new possibilities to the field of AI hardware with its unique computing model. Quantum computing is different from classical computing architectures in that it uses features such as quantum superposition and quantum entanglement to provide exponential performance gains in specific tasks, which provides a new direction for solving the current computing bottlenecks faced by AI. This paper aims to provide researchers and engineers with cutting-edge analysis on the synergistic development of quantum computing and AI hardware.

Keywords

quantum computing; artificial intelligence; future trends

论量子计算与 AI 硬件的发展未来

王晓寅

泰瑞达（上海）有限公司，中国·上海 201206

摘要

随着人工智能（AI）技术的快速发展，特别是深度学习模型的规模和复杂度不断提升，现代社会对计算能力提出了极高的要求。现代的AI模型（如GPT-4、BERT等）拥有数十亿甚至数千亿个参数，需要处理海量数据和执行复杂的计算任务。传统的CPU在处理这些任务时效率较低，难以满足训练和推理的需求。然而，量子计算作为一种具有颠覆性潜力的前沿技术，以其独特的计算模式为AI硬件领域带来了新的可能性。量子计算与经典的计算架构不同，它利用量子叠加和量子纠缠等特性，能够在特定任务中提供指数级的性能提升，这为解决当前AI面临的计算瓶颈提供了新的方向。论文旨在为研究人员和工程师提供有关量子计算与AI硬件协同发展的前沿性分析。

关键词

量子计算；人工智能；未来趋势

1 引言

随着深度学习模型的复杂度不断提升，AI对计算能力的需求急剧加大。现有的AI硬件，包括图形处理单元（GPU）和专用集成电路（ASIC），在处理庞大数据集和复杂模型时展现出了超凡的性能。然而，在面对更大规模的模型与更高的计算需求时，这些硬件的发挥的作用逐渐遇到瓶颈。

量子计算，指一种运用量子力学现象，如量子叠加、量子干涉与量子纠缠进行计算的方式，是过去的数十年中兴起的一种全新计算方式，在近几年中也是一个在各类科普中高频出现的热点词汇。量子计算和量子算法能够加速许多经

典计算机难以处理的问题。目前，量子计算领域依然在快速发展之中。随着量子计算技术的发展，越来越多的研究开始探讨其在AI硬件中的应用前景。

论文首先介绍AI硬件的发展现状及量子计算的基本概念，然后比较量子计算与经典AI硬件（GPU和ASIC）的差异，最后深入探讨量子计算在AI训练与推理中的应用场景，并预测其未来的发展方向^[1]。

2 AI 硬件的发展现状

2.1 GPU 和 AI 硬件的历史回顾

对于许多AI硬件公司来说，最近几年似乎是AI硬件发展的黄金时代；过去三年英伟达股价暴涨约+500%，超越英特尔成为全球市值最高的芯片公司。其他创业公司似乎同样火爆，在过去几年中，他们已花费数十亿美元资助AI硬件初创公司，以挑战英伟达的AI领导地位。

【作者简介】 王晓寅（1986-），男，中国江苏南京人，硕士，工程师，从事半导体研究。

在人工智能，特别是深度学习领域，图形处理单元（GPU）是最常用的硬件加速器。自2000年代末以来，GPU 凭借其高度并行的架构，在神经网络的训练和推理任务中表现出显著优势。相比于传统的中央处理器（CPU），GPU 能够同时处理成千上万个小型计算任务，这使得其在处理矩阵运算和卷积操作时表现出色。因此，GPU 迅速成为深度学习模型训练的核心硬件^[2]。

2.2 专用集成电路（ASIC）与 TPU 的发展

TPU 是谷歌在2015年6月的I/O开发者大会上推出的一款专为计算神经网络设计的芯片，主要是为了优化谷歌的TensorFlow机器学习框架。TPU 广泛用于AlphaGo系统以及谷歌地图、谷歌相册、谷歌翻译等应用，负责处理搜索、图像、语音等任务。与GPU不同，谷歌的TPU属于一种ASIC（应用专用集成电路）芯片。ASIC芯片是为特定应用而定制的，但一般开发周期长且成本较高。尽管数据中心对于AI任务的高计算需求日益增加，许多企业仍倾向于使用现有的GPU集群或GPU与CPU异构计算方案，而不愿轻易在ASIC领域进行尝试。

谷歌从2006年就开始酝酿研发一款专用于神经网络的芯片，尤其在2013年这种需求变得更加紧迫。当时，谷歌的图像搜索、谷歌照片、云视觉API和谷歌翻译等产品，依赖于深度神经网络的计算能力。庞大的计算需求让他们的数据中心需要成倍扩展，而单靠GPU和CPU已经难以满足性能和成本的要求。在这种背景下，谷歌正式启动了TPU的研发。经过15个月的设计、验证和建造，TPU于2014年正式完成，并首先部署在谷歌内部的数据中心。

尽管GPU和ASIC在提升AI计算能力方面取得了巨大成功，但随着AI模型的规模和复杂度不断增加，这些硬件面临的挑战也日益浮现：功耗和热量问题：深度学习模型的训练需要大量计算，伴随着显著的功耗和热量产生。这在大规模模型训练过程中尤为突出。硬件扩展性问题：当前的AI硬件体系结构在扩展性上存在局限，难以应对未来数倍甚至数十倍增加的计算需求。模型训练时间长：即使有了GPU和TPU，训练大规模深度学习模型的时间依然较长，尤其是在涉及大量参数和数据的情况下。这些瓶颈促使研究人员开始探索新的计算模式，其中量子计算被认为是一种有潜力突破现有硬件瓶颈的技术。

3 量子计算的基础与特性

3.1 量子计算的基本原理

量子计算利用专门技术（包括计算机硬件和利用量子力学的算法）来解决传统计算机或超级计算机无法解决或无法快速解决的复杂问题。量子计算是基于量子力学的基本原理，它与传统的二进制计算不同。量子计算利用量子比特（qubits）进行计算。量子比特不仅可以处于0或1的状态，还可以处于0和1的叠加态，显著增加了信息的表达能力。所以量子计算是一种不同于经典计算的新型计算范式。更为

有效的是，量子纠缠与量子干涉等现象使得量子计算能够同时处理大量并行任务，从而在某些特定问题上展示出指数级的加速潜力。

3.2 量子计算的优势

传统计算机在处理某些复杂任务时表现出色，如对大型分子数据库的分类。然而，当面临更复杂的问题，如模拟这些分子的行为时，传统计算机的能力显得有限。传统的超级计算机或许会通过大量处理器的并行计算来模拟分子行为，尝试暴力破解每个分子部分的可能性。然而，当问题超出简单分子的范畴时，超级计算机的处理能力便难以为继。现有的计算机存储能力不足以应对所有可能的分子行为组合排列。

目前，科学家若想了解一种分子的行为，通常需要合成该分子并在实际环境中进行实验。如果想探究细微调整对其行为的影响，往往需要重新合成新版本的分子，再次进行实验。这种方式既昂贵又耗费时间，阻碍了医学、半导体设计等领域的进步。传统计算机难以创建这种多维计算空间，因此在没有量子计算的情况下用途有限。量子算法则采用了一种全新的方法来应对这些复杂问题，即创建多维的计算空间。这种方法在化学模拟等复杂问题中，表现出了极高的效率。工业化学家已经开始研究如何将量子方法应用于他们的工作中。工程、金融、全球物流等领域的公司也在探索量子计算解决其行业内关键问题的潜力。量子计算的研究与开发正在带来显著进展，随着量子硬件的提升与算法的进步，许多重大问题，如分子模拟，或将找到有效的解决途径^[2]。

3.3 量子计算的局限性

然而，量子计算目前也面临着若干挑战，限制了其在实际应用中的广泛推广。

量子计算机的限制在于，它依赖于量子比特进行计算，而这种计算过程如果交由传统计算机处理，可能需要无限的时间，甚至根本无法完成。量子计算机能够通过八个量子比特同时表示从0~255之间的每个数值，这类类似于计算中的并行处理，大幅提升了计算速度。然而，量子计算机主要是为处理穷举法问题设计的专用设备，并不会完全取代经典计算机。未来，量子计算机很可能会像显卡一样，作为“量子加速器”被集成到主板硬件中，而非取代传统计算机。自Feynman等提出量子计算的原始思想至今已有40多年，相应进展令人瞩目；但不可否认的是，依然没有实现量子计算的任何实际应用。关于量子计算应用，学术界普遍认为：量子计算并不能全面替代经典计算；量子计算本身具有很多的限制；原则上经典计算可以替代量子计算，二者区别主要在于效率的高低^[3]。

尽管面临这些挑战，量子计算在某些特定领域，如人工智能，有望提供突破性的进展。

4 量子计算与经典AI硬件的对比

4.1 并行计算能力对比

在并行计算方面，经典硬件如GPU和ASIC的并行度

依赖于物理硬件的设计，而量子计算则依赖于量子叠加态，理论上能够在单个计算周期内同时处理更多的数据路径。因此，在处理某些高维数据或需要大规模矩阵运算的任务时，量子计算可能具有更高的并行效率。

4.2 数据处理能力对比

AI 训练和推理过程通常需要处理大量数据，尤其是在深度学习等领域，涉及大规模的数据集和复杂的模型。这些模型的训练依赖于矩阵计算和卷积操作。这就要运用到 GPU 了。经典硬件（特别是 GPU）由于其强大的并行处理能力，非常适合执行这些任务。GPU 可以同时处理大量的计算任务，大大加快了训练和推理速度，因此在当前的 AI 领域中被广泛使用。

但是，当面对一些高度复杂的优化问题时，经典硬件的性能可能会受到阻碍。例如，某些涉及大量变量和非线性约束条件的优化问题，传统的计算方法可能需要极长的时间来找到最佳设计方案。此时，量子计算可能展现出其优势。量子计算能够利用量子叠加和量子纠缠的特性，创建一个多维的计算空间，从而并行处理多个状态。这使得量子计算在解决某些特定的复杂优化问题上，能够比经典计算机更快地找到可行的解决方案。

AI 硬件面临的一个主要瓶颈是能耗问题。随着 AI 模型规模的增加，训练一个深度学习模型的能耗也急剧上升。量子计算有望通过计算方式降低能耗。

4.3 可扩展性对比

在 AI 硬件的扩展性方面，量子计算和经典硬件各有优势。经典硬件的扩展性主要受限于摩尔定律和晶体管的物理尺寸。量子计算则依赖于量子比特的扩展，但由于量子纠缠的特性，增加量子比特的数量并不会像增加晶体管那样显著增加硬件的复杂度。然而，量子计算目前的扩展能力仍受技术瓶颈限制，如量子比特的精度和纠错技术等问题。

5 量子计算在 AI 训练中的应用

量子计算在 AI 训练中的应用具有巨大的潜力，尤其是

在解决复杂的优化问题和处理庞大数据集方面^[4]。量子计算可以在特定问题中比经典计算机更快地处理复杂的计算任务。它通过其独特的并行计算能力和量子叠加态的方法。AI 模型的训练通常依赖于大规模矩阵运算、反向传播和梯度下降，这些过程对计算资源的需求极为庞大。传统的硬件，如 GPU 和 CPU，尽管在并行计算上表现出色，但在处理特别复杂的优化问题时，仍需要消耗大量时间和计算力。

量子计算在模型训练中的某些步骤（如参数调整和权重优化）实现显著的加速通过量子优化算法（如量子近似优化算法，QAOA）。与经典优化算法相比，量子算法能够在大规模、多维参数空间中快速搜索最优的设计方案，减少花费的时间和耗能，从而提高生产效率。

最后，量子计算可以通过量子退火和量子梯度下降等方法，能够更有效地探索全局最佳计算方案，从而避免传统优化算法所遇到的瓶颈。量子退火可以在高维的参数空间中快速探索，找到更加理想的模型权重，从而提高 AI 模型的表现。相比经典方法，量子优化算法能够以指数级的速度提升训练效率，特别适合需要处理高难度维数据集和复杂网络结构的 AI 任务。

6 结论

量子计算在优化复杂模型、加速大规模数据处理、提升生成模型的训练效率等方面具有广阔的运用优势。尽管当前量子计算仍然面临许多的挑战但随着量子计算技术的不断进步，它有望在未来彻底改变 AI 训练的方式和效率。

参考文献

- [1] 曹亚菲. 戴尔科技发布四大科技趋势展望[J]. 软件和集成电路, 2024(1):68-69.
- [2] 吁超华, 黄玉坤, 吴方君, 等. 吸纳量子计算的数据挖掘课程教学探索[J]. 电子元器件与信息技术, 2023, 7(12):101-106.
- [3] 李晓巍, 付祥, 燕飞, 等. 量子计算研究现状与未来发展[J]. 中国工程科学, 2022, 24(4):133-144.
- [4] 廖勤. 下一个前沿技术——量子计算?[N]. 解放日报, 2024-04-14(005).