

# 鲲鹏处理器流水线工作机理

*Hollow Man*

(兰州大学 信息科学与工程学院 兰州 730107)

**摘要** 本文通过作者对计算机体系结构课程的学习和研究,针对华为推出鲲鹏处理器,利用现有常见流水线实现方法相关资料,给出其流水线工作机理。

**关键词** 流水线, 鲲鹏, 体系结构

## Principles of the KungPeng CPU Pipeline

Songlin Jiang

(School of Information Science and Engineering, Lanzhou University, Lanzhou, 410073, China)

**Abstract** Through the author's study and research on the Computer Architecture course, aiming at the KunPeng processor created by Huawei, this paper uses the relevant materials of the existing common pipeline implementation methods, gives pipeline working principle of KunPeng.

**Keywords** Pipeline, KunPeng, Architecture

## 1 引言

流水线是指在程序执行的时候多条指令重叠进行的操作的一种准并行处理实现技术。传统基本处理器的五个执行步骤为取指、译码、执行、访存、写回,对应形成了5级流水线的处理器。流水线主要工作障碍为流水线竞争,其包括结构竞争、数据竞争、控制竞争。同时,由于指令执行所需的时钟周期不同,多周期操作指令的流水实现十分复杂<sup>[1]</sup>。因而如何设计出一种CPU流水线结构,做到在确保竞争问题得到解决的情况下实现更优化的性能,是目前计算机体系结构研究的热点问题,而鲲鹏流水线架构是对其的一次绝佳尝试。

## 2 鲲鹏流水线架构

鲲鹏处理器为超标量(以鲲鹏920为例,其一个时钟周期内可以发射4条指令)乱序发射流水线处理器,其CPU流水线主要阶段分别为:取指(Fetch)、译码(Decode)、分配(Allocation)、发射(Issue)、执行(Execute)、写回(Write Back),提交(Commit)。

在取指(Fetch)阶段,通过分析分支预测(Branch Prediction)、分支目标缓冲区(Branch Target Buffer)、返回地址堆栈(Return Address Stack)预测程序流来调整指令的执行,并分析程序的数据流来选择指令执行的最佳顺序。随后根据时间局部性原理,访问指令缓存,如果该地址最近被访问过则有可能在访问指令缓存中找到,否则访问内存读取指令。

在译码(Decode)阶段,首先将指令流分解为独立指令,并且将X86指令翻译为类似的RISC架构微指令(uOps),然后根据指令的类型(如控制指令,内存读写指令,运算指令)、需要的资源(包括读写需要的寄存器)以及操作类型进行分类,从中选取出指令从而确保要执行的指令是不相关的,确定该时钟周期中需要发射的指令。

在分配(Allocation)阶段,将指令中需要用到的寄存器进行修改重命名,以确保在发射执行时不会产生寄存器使用的冲突。同时,通过安排插入资源停顿(resource stalls)预留出发射的指令执行时需要使用的资源。

在发射(Issue)和执行(Execute)阶段,通过动态调度算法(记分牌策略或Tomasulo算法)将指令分发到相应的执行单元,进行指令的乱序执行。

单独的每条指令乱序执行完成后进行写回（Write Back）阶段。在写结果阶段，指令的结果暂时被存储在寄存器堆（Register File）重排序缓冲区（Reorder Buffer）中。如果其他指令急切需要此结果，那么重排序缓冲区可以直接为其传输所需的数据。

最终，在提交（Commit）阶段，通过重排序缓冲区内的数据重整执行结果次序，如果是推测执行则验证推测执行结果的正确性。最终将多发射的指令执行结果储存在寄存器或主存储器中。

## 参考文献

[1] 林和，计算机体系结构课程幻灯片. 兰州大学，2021

[2] 华为，鲲鹏架构概念介绍. [https://support.huaweicloud.com/ug-spt-kunpengdevps/kunpengsys\\_06\\_0005.html](https://support.huaweicloud.com/ug-spt-kunpengdevps/kunpengsys_06_0005.html), 2021

# “元宇宙”下的新型计算机体系结构

*Hollow Man*

(兰州大学 信息科学与工程学院 兰州 730107)

**摘要** 本文通过作者对计算机体系结构课程的学习和研究取得的成果,探索“元宇宙”时代的计算机体系结构,提出量子以及分布式的计算机架构是未来计算机体系结构的发展方向。

**关键词** 元宇宙, 分布式, 量子计算, 体系结构

## Principles of the KungPeng CPU Pipeline

Songlin Jiang

(School of Information Science and Engineering, Lanzhou University, Lanzhou, 410073, China)

**Abstract** Through the author's study and research on the Computer Architecture course, this paper explores the computer architecture in the era of Metaverse and puts forward that quantum and distributed computing architecture are the future of computer architecture.

**Keywords** Metaverse, Distributed, Quantum Computing, Architecture

## 1 量子计算机

目前,在硬件制造方面,我们都是通过缩小晶体管体积和增加微芯片的晶体管数目来提高计算能力的。摩尔定律过去准确预测了一个芯片上可以容纳的晶体管数目每两年会翻一番的现象<sup>[1]</sup>,但是目前,摩尔定律正在失效。现在,芯片中集成的晶体管体积已经变得非常小,密度较过去得到了极大提升,但是这也导致出现了电流泄漏的问题,能量效率降低,并增加了过热风险。因而目前一心只想要缩小晶体管体积已经是弊大于利了。

在 2019 年,谷歌在量子计算领域的突破<sup>[2]</sup>登 *Nature* 封面,首先实现了量子霸权。据谷歌宣称其实现的用具有 53 个超导量子比特的可编程处理器运行 200 秒所能达到的计算能力相当于目前超算一万年所能达到的。

在传统电子计算机中,一个比特位只能处于 0 或者 1 两种状态中的一种。而量子力学告诉我们,一个量子比特位可以同时拥有这些状态。这意味着,只需要  $N$  个量子比特,就可以存储  $2^N$  个传统比特位对应的数据。同理,如果我们拥有一个  $N$  核量子处理器同时并行计算,相较于传统处理器,则相当于一个  $2^N$  核传统处理器同时并行计算。

量子计算机的指数级性能提升将会极大地提高人类的现有计算能力,同时也会改写现有的计算机体系结构,量子比特将会取代现有的比特,不同的适用于量子时代的数据结构处理方式也将会出现。

## 2 分布式计算架构

当然,目前量子计算机还处于起步阶段,目前技术所能达到的稳定性远远不足以运用。假如后面量子计算机发展遇到了瓶颈,并不能成为“元宇宙”的基石的话,目前已经开始应用的分布式计算架构将成为主导。分布式架构通过网络通信技术,可以将各处的计算资源整合在一起,进行计算。目前的区块链技术即是运用的分布式架构,来实现可信的数据存储。

分布式计算架构下,网络的性能成为限制整体计算性能的瓶颈,同时计算任务的分配问题也十分富有挑战性,与此相关的新的缓存结构以及算法的提出必将见证新的计算机体系结构的诞生。

## 参考文献

- [1] 林和, 计算机体系结构课程幻灯片. 兰州大学, 2021
- [2] Arute, Frank, et al. "Quantum supremacy using a programmable superconducting processor." *Nature* 574.7779 (2019): 505-510.