

Analysis of the characteristics of computer memory in program running

Meng Xiao

Quanzhou Economic and Trade Vocational and technical college, Quanzhou

Abstract: This paper discusses all levels of memory involved in the operation of computer programs, their characteristics and relations, analyzes the measures taken to solve the contradiction between speed, capacity and price, and studies the factors affecting the speed of program operation and the contents related to storage and protection. This paper focuses on cache memory, virtual memory and hit rate.

Key words: memory; storage protection; cache memory; virtual memory; hit rate

Received: 2019-10-02; Accepted: 2019-11-10; Published: 2019-11-16

程序运行中计算机存储器特点分析

蒙 绍

泉州经贸职业技术学院，泉州

邮箱: mengshao868@163.com

摘 要: 讨论了计算机程序运行的过程中涉及到的各级存储器及它们的特点和联系等内容, 分析了为了解决速度、容量、价格的矛盾所采取的措施, 研究了影响程序运行速度快慢的因素和对存储保护相关的内容。重点探讨了高速缓冲存储器、虚拟存储器及命中率。

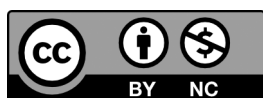
关键词: 存储器; 存储保护; 高速缓冲存储器; 虚拟存储器; 命中率

收稿日期: 2019-10-02; 录用日期: 2019-11-10; 发表日期: 2019-11-16

Copyright © 2019 by author(s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 计算机的存储体系结构

1.1 存储器件

现代计算机的存储体系都是按照美籍匈牙利科学家冯·诺依曼思想设计出来的, 我们经常听到计算机中许多可以存储数据的器件, 例如: 通用寄存器、Cache (即高速缓冲存储器)、内存、虚拟存储器、硬盘、U 盘、光盘等。按照它们在计算机中的位置和作用, 它们形成如下的金字塔结构:

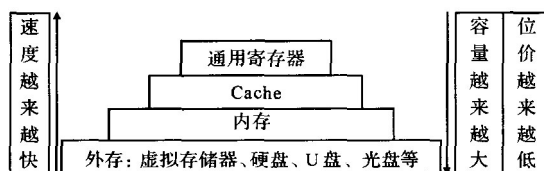


图 1 计算机的存储体系结构

1.2 存储体系的三级结构

根据图 1，当代计算机一般都将计算机系统的存储系统可分为 3 级：高速缓冲存储器、内存、外存。高速缓冲存储器用来改善主存储器与中央处理器的速度匹配问题，内存用于存储当前正在运行的程序和数据，外存用于扩大存储空间。目标就是在一定的成本下，获得尽可能大的存储容量和尽可能高的存取速度以及可靠性。

（1）高速缓冲存储器

高速缓冲存储器的容量一般只有主存储器的几百分之一，但它的存取速度能与中央处理器相匹配。根据程序局部性原理，正在使用的主存储器某一单元邻近的那些单元将被用到的可能性很大。因而，当中央处理器存取主存储器某一单元时，计算机硬件就自动地将包括该单元在内的那一组单元内容调入高速缓冲存储器，中央处理器即将存取的主存储器单元很可能就在刚刚调入到高速缓冲存储器的那一组单元内。于是，中央处理器就可以直接对高速缓冲存储器进行存取。在整个处理过程中，如果中央处理器绝大多数存取主存储器的操作能为存取高速缓冲存储器所代替，计算机系统处理速度就能显著提高。

现代计算机一般又把高速缓冲存储器分成两个层次。一级缓存：即 L1Cache。集成在 CPU 内部中，用于 CPU 在处理数据过程中数据的暂时保存。由于缓存指令和数据与 CPU 同频工作，L1 级高速缓存缓存的容量越大，存储信息越多，可减少 CPU 与内存之间的数据交换次数，提高 CPU 的运算效率。但因高速缓冲存储器均由静态 RAM 组成，结构较复杂，在有限的 CPU 芯片面积上，L1 级高速缓存的容量不可能做得太大。二级缓存：即 L2Cache。由于 L1 级高速缓存容量的限制，为了再次提高 CPU 的运算速度，在 CPU 外部放置一高速存储器，即二级缓存。工作主频比较灵活，可与 CPU 同频，也可不同。CPU 在读取数据时，先在 L1 中寻找，再从 L2 寻找，然后是内存，最后是外存。

（2）内存

也称主存，用半导体晶体管做成相应的内存颗粒，再排列成条状，也就是我们通常所说的内存条，用于存放计算机运行期间的大量程序和数据。它既能

为 CPU 直接访问, 又能把最活跃的程序和数据存放到 Cache 中, 提供给 CPU 使用。

(3) 外存

存放当前暂不参与运行的程序和数据, 当 CPU 需要时, 丰耳与主存成批交换信息。外存有磁盘、光盘和 flash 盘等。

1.3 存储体系的两个层次

两个层次如图 2 所示:

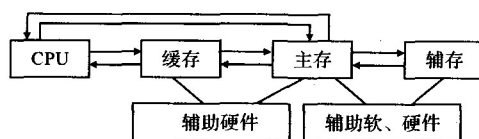


图 2 存储体系主两个层次

“高速缓存—主存”层次: 在 CPU 和主存之间增加一级速度快, 但容量较小。较高的高速缓冲存储器借助于辅助软硬件, 与主存构成一个有机的整体, 以弥补主存速度的不足。这个层次的工作主要由硬件实现。

“主存—辅存”层次: 目的是为了弥补主存容量的不足。它是在主存外面增加一个容量更大, 单位价格更低, 但速度更慢的存储器 (称为辅存, 一般是硬盘)。它们依靠辅助软硬件的作用, 构成一个整体。“主存—辅存”层次常被用来实现虚拟存储器, 向编程人员提供大量的程序空间。

“高速缓存—主存”与“主存—辅存”层次的区别如表 1 所示。

表 1 两个层次之间的区别

	高速缓存	主存—辅存
目的	为了弥补主存速度的不足	为了弥补主存容量的不足
实现	全部由专用硬件实现	系统软件, 辅助硬件
速度比	几分之一	几百分之一
数据块大小	几十个字节	几百到几千个字节
CPU 对下一级的访问	可直接访问	均通过第一级
失效时 CPU 是否切换	不切换	切换到其它进程

1.4 虚拟存储器

虚拟存储器位于“主存—辅存”层次, 在系统软件和辅助硬件的管理下就

像一个单一的、可直接访问的大容量存储器，以透明方式为用户程序提供一个远大于主存容量的存储空间。

在未使用虚拟存储器的机器中，用户程序所需存储空间不能大于机器实际能提供的主存容量，如果用户程序过大，其所需存储空间超出了主存能提供的容量的大小，则要么该程序无法运行，要么由用户自行完成对程序的分块处理和存储分配工作。这对一般用户来讲是非常困难的。而在使用了虚拟存储器的机器中，允许用户程序空间大于机器实际主存空间。当用户程序所需存储空间大于主存当前能够提供的存储容量时，由系统将用户程序的一部分先存放在辅助存储器中，当运行到这部分程序时，再将它们从辅存中调出，并分配给它们合适的主存空间执行。

在实现虚拟存储器管理的机器中，能提供给用户程序运行的是一个比主存大的多的虚拟空间，称为虚存；而真正为用户程序提供实际运行空间的是主存，又称为实存。程序中出现的必须指出的是，当一个程序运行时，计算机内部的指令流和数据流涉及到的整个存取过程都是由硬件和操作系统设计好的流水流程自动执行的，不需要用户干涉。用户也几乎感觉不到相应的任务调度过程的存在。用户所能感觉到的只是程序运行时间的长短而已。我们把这种整个任务调度的自动执行过程称为对用户是“透明”的。程序在存储器中的分配完全在系统的控制下进行，无需用户干预。

2 程序的运行

2.1 运行过程分析

程序运行时，先从外存调入内存（或部分调入内存），再将当前要运行的那一个程序段调入内存，然后等待 CPU 的调度，当 CPU 把运行到本程序段时，把内存中当前要执行的那个任务调入 CPU，当然，如果在高速缓冲存储器里有这个任务就直接从高速缓冲存储器调入，如果内存里没有，则从虚拟存储器里找，再没有就到外存里找，找到后再一级一级地向上一级的存储器里载入。程序执行中间结果一般都存在通用寄存器里，最终结果按相同的方式被写入外存。

2.2 影响程序运行快慢的主要因素

2.2.1 命中率

CPU 要读取一个数据时，首先从 Cache 中查找，如果找到就立即读取并送给 CPU 处理，称为命中；如果没有找到，称未命中，就用相对慢的速度从内存中读取并送给 CPU 处理，同时把这个数据所在的数据块调入 Cache 中，可以使得以后对整块数据的读取都从 Cache 中进行，不必再调用内存。正是这样的读取机制使 CPU 读取 Cache 的命中率非常高（大多数 CPU 可达 90% 左右），也就是说 CPU 下一次要读取的数据 90% 都在 Cache 中，只有大约 10% 需要从内存读取，这大大节省了 CPU 直接读取内存的时间。虚拟存储器里的执行过程也是如此，但命中率会低一些。

2.2.2 各级存储器的容量、存取速度

各级存储器的容量反映了它能够存储的最大数据量，同时它的存取速度也直接关系到数据在本级存储器中流动的快慢。

例如：早期的 Pentium 的 Cache 只有 64KB，目前 CPU 的 ~ 级缓存一般是 256KB，二级缓存是 512KB 到 2MB。对大多数 CPU 来说，缓存越大，运行速度越快。Pentium4 和赛扬的区别就是缓存大小不一样。L1、L2 缓存目前基本上都是全速的，传输数据的频率和 CPU 主频一样，但是有延迟。L1 通常延迟几个周期（ ≤ 5 ），L2 通常延迟十几个周期。换句话说，CPU 有多快，L1、L2 就有多快，只不过存取缓存的命令要延迟一会儿才能被执行。

2.2.3 其它因素

（1）并行执行程序的数量：Windows 本身就是一个多任务的操作系统，我们可以明显感觉到，同时运行的程序较多时计算机的速度会比较慢。这里涉及到的一个问题就是任务的调度，而这个任务之间的调度和切换比较频繁时，就需要一定的调度时间，另外，每个不同程序的运行也要占用不等的 CPU 时间片。

（2）CPU 的工作频率：CPU 相当于人的心脏，频率越高，速度越快。

（3）显卡的工作速度：主要指显示速度，与 CPU 的工作速度关系不大，但当送往显示设备的数据量较大。而显示系统又不能及时处理时，会感觉显示

有些滞后，这给人们造成的错觉便是计算机的速度慢，其实有时是显示卡的工作速度较慢造成的。

(4) 带宽：计算机内部的带宽越宽，计算机处理程序的速度也会越快。

2.3 存储保护

现代操作系统都是多任务的，为了使系统能正常工作，应防止由于一个程序出错而破坏其它的程序，还要防止一个程序不合法地访问不是分配给它的主存区域。因此，要为计算机程序的正常运行提供存储保护，通常采用的方法是：存储区域保护和访问方式保护。

(1) 存储区域保护：对于物理寄存器可以采用界限寄存器方式。由系统软件经特权指令设置上、下界寄存器，为每个程序划定存储区域，禁止越界访问。由于程序不能改变上下界的值，所以它如果出现错误，也只能破坏该程序，不会侵犯到其它的程序；对于虚拟寄存器则可以采用页表保护、段表保护和键式保护方式、环保护方式。

(2) 访问方式保护：对主存信息的使用可以有 3 种方式：读 (R)、写 (W) 和执行 (E)，相应的访问方式保护就有 R, W, E 3 种方式形成的逻辑组合。这些访问方式保护通常作为程序状态寄存器的保护位，并且和区域保护结合起来实现。

表 2 访问方式保护的逻辑组合

逻辑组合	含义	逻辑组合	含义
R+W+E	不允许任何访问	(R+E)*W	只能写访问
R+W+E	可进行任何访问	(R+E)*W	不准写访问
(R+W)*E	只能读写，不可执行	R*(W+E)	只能读访问
(R+W)*E	只能执行，不可读写	R*(W+E)	不准读访问

3 束语

随着半导体技术、通讯技术等计算机相关技术的发展，存储系统的发展呈现出了新的层次结构和典型的金字塔模型。有关处理器内部的高速缓存的层次结构的前沿研究。主要集中在最后一级高速缓存的有效利用上。同时，多核技

术带来了计算机性能的飞速提升，为计算机的存储结构带来了许多的机遇和挑战，也为程序的快速运行提供了更多的途径和可能。

参考文献

- [1] 薛胜军. 计算机组成原理 [M]. 武汉: 华中科技大学出版社, 2000.
- [2] 白中英. 计算机组成原理 [M]. 北京: 科学出版社, 2001.
- [3] 尹朝庆. 计算机系统结构 [M]. 武汉: 华中科技大学出版社, 2000.
- [4] 唐朔飞. 计算机组成原理 [M]. 北京: 高等教育出版社, 2000.