注意，虽然 SRAM 的性能滞后于 CPU 的性能，但还是在保持增长。不过，DRAM 和磁盘性能与 CPU 性能之间的差距实际上是在加大的。直到 2003 年左右多核处理器的出现，这个性能差距都是延迟的函数，DRAM 和磁盘的访问时间比单个处理器的周期时间提高得更慢。不过，随着多核的出现，这个性能越来越成为了吞吐量的函数，多个处理器核并发地向 DRAM 和磁盘发请求。

图 6-16 清楚地表明了各种趋势，以半对数为比例(semi-log scale)，画出了图 6-15 中的访问时间和周期时间。

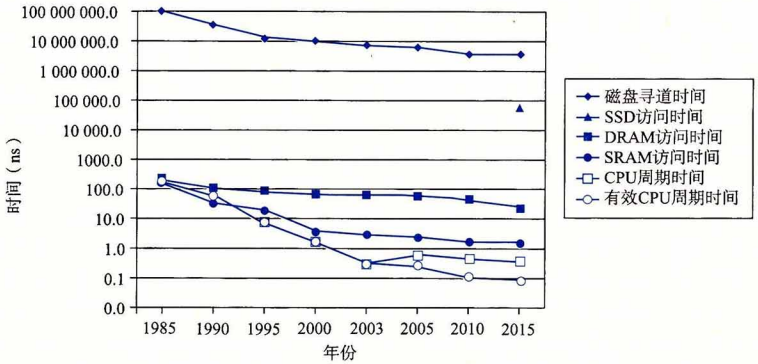

图 6-16　磁盘、DRAM 和 CPU 速度之间逐渐增大的差距

正如我们将在 6.4 节中看到的那样，现代计算机频繁地使用基于 SRAM 的高速缓存，试图弥补处理器-内存之间的差距。这种方法行之有效是因为应用程序的一个称为*局部性*(locality)的基本属性，接下来我们就讨论这个问题。

练习题 6.6　使用图 6-15c 中从 2005 年到 2015 年的数据，估计到哪一年你可以以 $500 的价格买到一个 1PB($10^{15}$ 字节)的旋转磁盘。假设美元价值不变(没有通货膨胀)。

旁注　当周期时间保持不变：多核处理器的到来

计算机历史是由一些在工业界和整个世界产生深远变化的单个事件标记出来的。有趣的是，这些变化点趋向于每十年发生一次：20 世纪 50 年代 Fortran 的提出，20 世纪 60 年代早期 IBM 360 的出现，20 世纪 70 年代早期 Internet 的曙光(当时称为 APRANET)，20 世纪 80 年代早期 IBM PC 的出现，以及 20 世纪 90 年代万维网(World Wide Web)的出现。

最近这样的事件出现在 21 世纪初，当计算机制造商迎头撞上了所谓的“能量墙(power wall)”，发现他们无法再像以前一样迅速地增加 CPU 的时钟频率了，因为如果那样芯片的功耗会太大。解决方法是用多个小处理器核(core)取代单个大处理器，从而提高性能，每个完整的处理器能够独立地、与其他核并行地执行程序。这种多核(multi-core)方法部分有效，因为一个处理器的功耗正比于 $P=fCv^2$，这里 f 是时钟频率，C 是电容，而 v 是电压。电容 C 大致上正比于面积，所以只要所有核的总面积不变，多核造成的能耗就能保持不变。只要特征尺寸继续按照摩尔定律指数性地下降，每个处理器中的核数，以及每个处理器的有效性能，都会继续增加。

从这个时间点以后，计算机越来越快，不是因为时钟频率的增加，而是因为每个处理器中核数的增加，也因为体系结构上的创新提高了在这些核上运行程序的效率。我们可以从图 6-16 中很清楚地看到这个趋势。CPU 周期时间在 2003 年达到最低点，然后实际上是又开始上