高速缓存类型	访问时间（周期）	高速缓存大小（C）	相联度（E）	块大小（B）	组数（S）
L1 i-cache	4	32KB	8	64B	64
L1 d-cache	4	32KB	8	64B	64
L2统一的高速缓存	10	256KB	8	64B	512
L3统一的高速缓存	40~75	8MB	16	64B	8192

图 6-39　Core i7 高速缓存层次结构的特性

6.4.7　高速缓存参数的性能影响

有许多指标来衡量高速缓存的性能：

- 不命中率(miss rate)。在一个程序执行或程序的一部分执行期间，内存引用不命中的比率。它是这样计算的：不命中数量/引用数量。
- 命中率(hit rate)。命中的内存引用比率。它等于 1－不命中率。
- 命中时间(hit time)。从高速缓存传送一个字到 CPU 所需的时间，包括组选择、行确认和字选择的时间。对于 L1 高速缓存来说，命中时间的数量级是几个时钟周期。
- 不命中处罚(miss penalty)。由于不命中所需要的额外的时间。L1 不命中需要从 L2 得到服务的处罚，通常是数 10 个周期；从 L3 得到服务的处罚，50 个周期；，从主存得到的服务的处罚，200 个周期。

优化高速缓存的成本和性能的折中是一项很精细的工作，它需要在现实的基准程序代码上进行大量的模拟，因此超出了我们讨论的范围。不过，还是可以认识一些定性的折中考量的。

1. 高速缓存大小的影响

一方面，较大的高速缓存可能会提高命中率。另一方面，使大存储器运行得更快总是要难一些的。结果，较大的高速缓存可能会增加命中时间。这解释了为什么 L1 高速缓存比 L2 高速缓存小，以及为什么 L2 高速缓存比 L3 高速缓存小。

2. 块大小的影响

大的块有利有弊。一方面，较大的块能利用程序中可能存在的空间局部性，帮助提高命中率。不过，对于给定的高速缓存大小，块越大就意味着高速缓存行数越少，这会损害时间局部性比空间局部性更好的程序中的命中率。较大的块对不命中处罚也有负面影响，因为块越大，传送时间就越长。现代系统(如 Core i7)会折中使高速缓存块包含 64 个字节。

3. 相联度的影响

这里的问题是参数 E 选择的影响，E 是每个组中高速缓存行数。较高的相联度(也就是 E 的值较大)的优点是降低了高速缓存由于冲突不命中出现抖动的可能性。不过，较高的相联度会造成较高的成本。较高的相联度实现起来很昂贵，而且很难使之速度变快。每一行需要更多的标记位，每一行需要额外的 LRU 状态位和额外的控制逻辑。较高的相联度会增加命中时间，因为复杂性增加了，另外，还会增加不命中处罚，因为选择牺牲行的复杂性也增加了。

相联度的选择最终变成了命中时间和不命中处罚之间的折中。传统上，努力争取时钟频率的高性能系统会为 L1 高速缓存选择较低的相联度(这里的不命中处罚只是几个周期)，而在不命中处罚比较高的较低层上使用比较小的相联度。例如，Intel Core i7 系统中，L1 和 L2 高速缓存是 8 路组相联的，而 L3 高速缓存是 16 路组相联的。

4. 写策略的影响

直写高速缓存比较容易实现，而且能使用独立于高速缓存的写缓冲区(write buffer)，用来更新内存。此外，读不命中开销没这么大，因为它们不会触发内存写。另一方面，写