

## *Memoria Cache*

**Cache:** memoria rápida (SRAM) cercana del procesador. Ayuda a mantener instrucciones y datos cercanos a las unidades funcionales, entonces el procesador puede ejecutar más instrucciones más rápidamente.

**Translation-Lookaside Buffer (TLB):** mantiene direcciones de páginas (bloques de memoria) en memoria principal que recientemente han sido accesados (una cache para direcciones de memoria).

### *Beneficios de la Cache.*

- La cache de datos fue diseñada con dos conceptos claves en mente

–Localidad Espacial

- Cuando un elemento es referenciado sus vecinos deberán ser referenciados también
- Las líneas de cache son buscadas juntas
- Trabajar sobre elementos de datos consecutivos en la misma línea de cache

–Localidad Temporal

- Cuando un elemento es referenciado, éste volverá a ser referencia de nuevo muy pronto
- Los datos en la cache son reusados constantemente

### *Correspondencias de la caches:*

*Cache de mapeo directo:* Un bloque en la memoria principal puede ir exactamente a un lugar en la cache. Esto es llamado **correspondencia directa** porque existe un mapeo directo desde cualquier dirección de bloque en memoria a una simple ubicación en la cache.

*Cache Completamente Asociativa:* Un bloque en memoria principal puede ser colocado en cualquier lugar de la cache. Esto es llamado **Completamente asociativa** porque el bloque en memoria principal puede ser asociado con cualquier entrada en la cache.

*Cache asociativa por conjuntos:* El diseño intermedio entre cache de mapeo directo y cache completamente asociativa es llamado **cache asociativa por conjuntos**.

### *Políticas de Escritura:*

Antes de reemplazar un bloque de la cache, es necesario considerar si ha sido o no alterado en cache pero no en memoria principal, si ha sido modificado, esto significa que se ha realizado al menos una operación de escritura sobre una palabra de la línea correspondiente de la cache, y la memoria principal debe actualizarse de acuerdo con ello, por ello se consideran algunas técnicas fácilmente de implementar, las cuales son:

- *Escritura Inmediata:* significa que todas las operaciones de escritura se hacen tanto en cache como en memoria principal, asegurando que el contenido de memoria principal siempre es valido. La principal desventaja esta en el hecho de que bajar la jerarquía de memoria cada vez que hay una actualización podría

- generar un cuello de botella por las diferencias de tiempo en los accesos a memoria.
- *Post-Escritura*: Las actualizaciones se hacen solo en memoria cache, cuando se tiene lugar a una actualización se activa un bit “actualizar” asociado a la línea en cache; después, cuando el bloque es sustituido es escrito en memoria principal si y solo si el bit “actualizar” esta activado (actualizar = 1, si actualizar = 0 es porque no hubo ninguna modificación por lo que el dato no baja a la memoria principal). La principal desventaja es que en ciertos momentos de la ejecución el dato que esta en memoria principal no es un dato valido ya que el dato ha sido modificado y solo guardado en la cache.

### *Estrategias de Reemplazos.*

**Least Recently Used (LRU)**: El bloque de la cache que es el menos recientemente usado es reemplazado con un nuevo bloque.

**Random Replace**: Un bloque de la cache es reemplazado aleatoriamente.