Address translation overhead

Average Memory Access Time =

Hit time (no page fault) + 

Miss rate (page fault rate) x Miss penalty

Examples of VM performance

Hit time = 50 ns.

Page fault rate = 0.001%

Miss penalty = 2 ms

\[ T_{av} = 50 + 10^{-5} \times 2 \times 10^6 \text{ ns} = 70 \text{ ns}. \]
Improving the Performance of Virtual Memory

1. Hit time involves one extra table lookup. Hit time can be reduced using a TLB (TLB = Translation Lookaside Buffer).

2. Miss rate can be reduced by allocating enough memory to hold the working set. Otherwise, thrashing is a possibility.

3. Miss penalty can be reduced by using disk cache
Page Replacement policy

Determines which page needs to be discarded to accommodate an incoming page. Common policies are

- Least Recently Used (LRU)
- Least Frequently Used (LFU)
- Random

Writing into VM

Write-through is possible if a write buffer is used. But write-back makes more sense. The page table must keep track of dirty pages. There is no overhead to discard a clean page, but to discard dirty pages, they must be written back to the disk.
Working Set

Consider a page reference string

0, 1, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2, … 100,000 references

The size of the working set is 2 pages.

Always allocate enough memory to hold the working set of a program (Working Set Principle)

Disk cache

Modern computers allocate up a large fraction of the main memory as file cache. Similar principles apply to disk cache that drastically reduces the miss penalty.
**Address Translation Using TLB**

TLB is a **set-associative cache** that holds a partial page table. In case of a TLB hit, the block number is obtained from the TLB (fast mode). Otherwise (i.e. for TLB miss), the block number is obtained from the direct map of the page table in the main memory, and the TLB is updated.
Multi-level Address Translation

Example 1: The old story of VAX 11/780

30-bit virtual address (1 GB) per user
Page size = 512 bytes = $2^9$
Maximum number of pages = $2^{21}$ i.e. 2 million
Needs 8 MB to store the page table. Too big!

Solution?

Store the page table in Virtual Memory.

Thus, page table is also paged!
Two-level address translation

Virtual address space

<table>
<thead>
<tr>
<th>Directory Entry</th>
<th>Offset in page table</th>
<th>Offset in page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

Virtual Address
Exception Handling

It is a very important issue in computer organization. Exception (or interrupt or trap) is an unanticipated event that needs attention. Devices send interrupt signal to the processor when they are ready for data transfer. Software events like divide-by-zero or page fault interrupt the processor.

At the end of each instruction cycle, the processor checks for the interrupt flag. If it is set, then the processor determines the cause of the interrupt, and control is transferred to the operating system that services the interrupt, and then returns to the original program.