Core i7 Page Table Translation

CR3 Physical address of L1 PT

VPN 1
L1 PT Page global directory
L1 PTE
512 GB region per entry

VPN 2
L2 PT Page upper directory
L2 PTE
1 GB region per entry

VPN 3
L3 PT Page middle directory
L3 PTE
2 MB region per entry

VPN 4
L4 PT Page table
L4 PTE
4 KB region per entry

VPO

VPN 1
VPN 2
VPN 3
VPN 4

L1 PT
L2 PT
L3 PT
L4 PT

PPO
PPN

Physical address

Offset into physical and virtual page

Physical address of page

Virtual address

512 GB region per entry
1 GB region per entry
2 MB region per entry
4 KB region per entry
Review of Symbols

■ Basic Parameters
  ▪ \( N = 2^n \): Number of addresses in virtual address space
  ▪ \( M = 2^m \): Number of addresses in physical address space
  ▪ \( P = 2^p \): Page size (bytes)

■ Components of the virtual address (VA)
  ▪ TLBI: TLB index
  ▪ TLBT: TLB tag
  ▪ VPO: Virtual page offset
  ▪ VPN: Virtual page number

■ Components of the physical address (PA)
  ▪ PPO: Physical page offset (same as VPO)
  ▪ PPN: Physical page number
  ▪ CO: Byte offset within cache line
  ▪ CI: Cache index
  ▪ CT: Cache tag
### Core i7 Level 1-3 Page Table Entries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>12</th>
<th>11</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>XD</td>
<td>Unused</td>
<td>Page table physical base address</td>
<td>Unused</td>
<td>G</td>
<td>PS</td>
<td>A</td>
<td>CD</td>
<td>WT</td>
<td>U/S</td>
<td>R/W</td>
<td>P=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Available for OS

P=0

Each entry references a 4K child page table. Significant fields:

- **P**: Child page table present in physical memory (1) or not (0).
- **R/W**: Read-only or read-write access access permission for all reachable pages.
- **U/S**: user or supervisor (kernel) mode access permission for all reachable pages.
- **WT**: Write-through or write-back cache policy for the child page table.
- **A**: Reference bit (set by MMU on reads and writes, cleared by software).
- **PS**: Page size: if bit set, we have 2 MB or 1 GB pages (bit can be set in Level 2 and 3 PTEs only).

**Page table physical base address**: 40 most significant bits of physical page table address (forces page tables to be 4KB aligned)

**XD**: Disable or enable instruction fetches from all pages reachable from this PTE.
Core i7 Level 4 Page Table Entries

Each entry references a 4K child page. Significant fields:

- **P**: Child page is present in memory (1) or not (0)
- **R/W**: Read-only or read-write access permission for this page
- **U/S**: User or supervisor mode access
- **WT**: Write-through or write-back cache policy for this page
- **A**: Reference bit (set by MMU on reads and writes, cleared by software)
- **D**: Dirty bit (set by MMU on writes, cleared by software)

**Page physical base address**: 40 most significant bits of physical page address (forces pages to be 4KB aligned)

**XD**: Disable or enable instruction fetches from this page.

### Core i7 Level 4 Page Table Entries

<table>
<thead>
<tr>
<th>63</th>
<th>62</th>
<th>52</th>
<th>51</th>
<th>12</th>
<th>11</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>XD</td>
<td>Unused</td>
<td>Page physical base address</td>
<td>Unused</td>
<td>G</td>
<td>D</td>
<td>A</td>
<td>CD</td>
<td>WT</td>
<td>U/S</td>
<td>R/W</td>
<td>P=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Available for OS (for example, if page location on disk) P=0
End-to-end Core i7 Address Translation

Virtual address (VA) → VPN → VPO

TLB

TLB hit

L1 TLB (16 sets, 4 entries/set)

VPN1, VPN2, VPN3, VPN4

Page tables

CR3

PTE

Physical address (PA)

40

VPN

L1 d-cache

(64 sets, 8 lines/set)

PPO

L1 hit

L1 miss

32/64

Result

Physical address (PA)

L2, L3, and main memory

CPU

Page tables

VPN

VPO

TLBT

TLBI

36

12

32

4

9

9

9

9

40

12

CT

CI

CO

PTE

PTE

PTE

PTE

PTE

PTE
Cute Trick for Speeding Up L1 Access

- Bits that determine CI identical in virtual and physical address
- Can index into cache while address translation taking place
- Cache carefully sized to make this possible: 64 sets, 64-byte cache blocks
- Means 6 bits for cache index, 6 for cache offset
- That’s 12 bits; matches VPO, PPO → One reason pages are $2^{12}$ bits = 4 KB
Virtual Address Space of a Linux Process

- **Process-specific data structs** (ptables, task and mm structs, kernel stack)
- Physical memory
- Kernel code and data
- **User stack**
- **Memory mapped region** for shared libraries
- **Runtime heap** (malloc)
- Uninitialized data (.bss)
- Initialized data (.bss)
- Program text (.text)

**Identical for each process**

**Kernel virtual memory**

**Process virtual memory**

- Processor virtual memory
- Identical for each process
- Process-specific data structs (ptables, task and mm structs, kernel stack)
- Physical memory
- Kernel code and data
- User stack
- Memory mapped region for shared libraries
- Runtime heap (malloc)
- Uninitialized data (.bss)
- Initialized data (.bss)
- Program text (.text)

**Process virtual memory**
| P  | 0 | The fault was caused by a non-present page. |
|    | 1 | The fault was caused by a page-level protection violation. |
| W/R| 0 | The access causing the fault was a read. |
|    | 1 | The access causing the fault was a write. |
| U/S| 0 | A supervisor-mode access caused the fault. |
|    | 1 | A user-mode access caused the fault. |
| RSVD| 0 | The fault was not caused by reserved bit violation. |
|    | 1 | The fault was caused by a reserved bit set to 1 in some paging-structure entry. |
| I/D| 0 | The fault was not caused by an instruction fetch. |
|    | 1 | The fault was caused by an instruction fetch. |
| PK | 0 | The fault was not caused by protection keys. |
|    | 1 | There was a protection-key violation. |
| SGX| 0 | The fault is not related to SGX. |
|    | 1 | The fault resulted from violation of SGX-specific access-control requirements. |

**Figure 4-12. Page-Fault Error Code**