Memory Management Units in High-Performance Processors

Ausgewählte Themen in Hardwareentwurf und Optik – Seminar

Universität Mannheim
LS Rechnerarchitektur - Prof. Dr. U. Brüning
WS 2003/2004

Frank Lemke

Inhalt

(i) Grundlagen
(ii) Intel - Itanium2
(iii) AMD - Opteron
(iv) Apple/IBM - G5
(v) Sun - UltraSparc III
(vi) Zusammenfassung

(i) Was ist eine Memory Management Unit?

Die Memory Management Unit (MMU) ist eine funktionale Einheit zur Durchführung der virtuellen/physikalischen Abbildung.
Jeder virtuellen Adresse wird die passende physikalische Adresse zugeordnet.

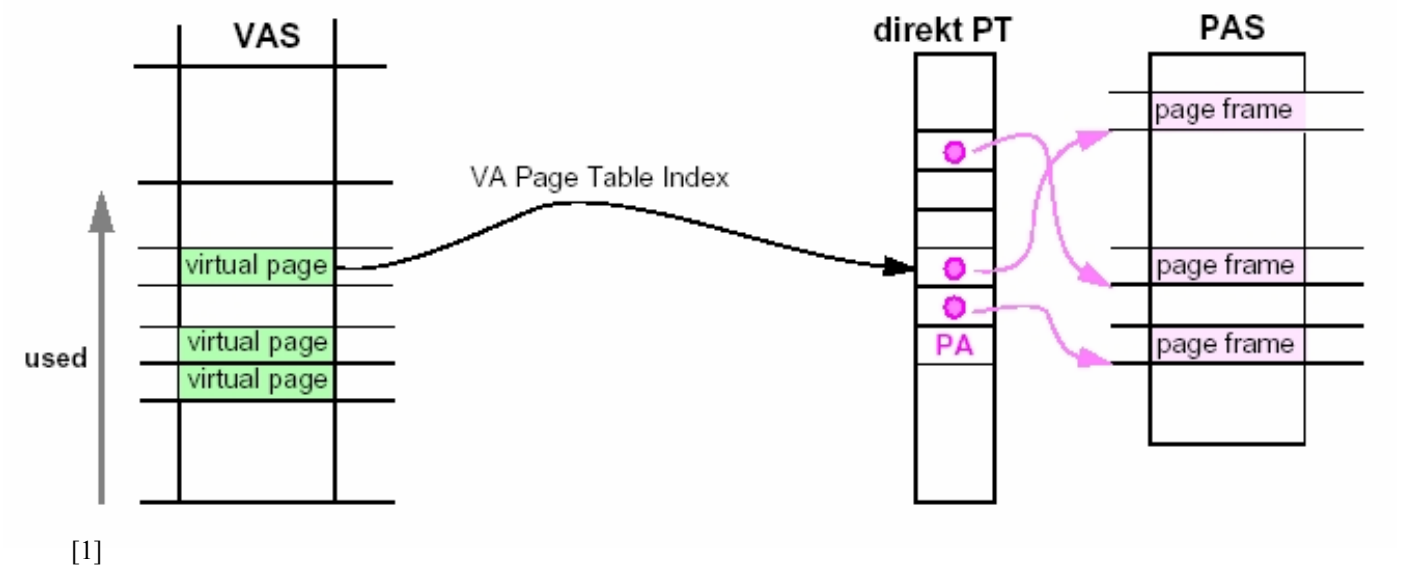
(i) Funktionale Teile der MMU (1)

Hierarchie der Adresstransformation

a. Page Address Translation
b. Segment Address Translation
c. Block Address Translation

(i) Funktionale Teile der MMU (2)

Direct Page Table:

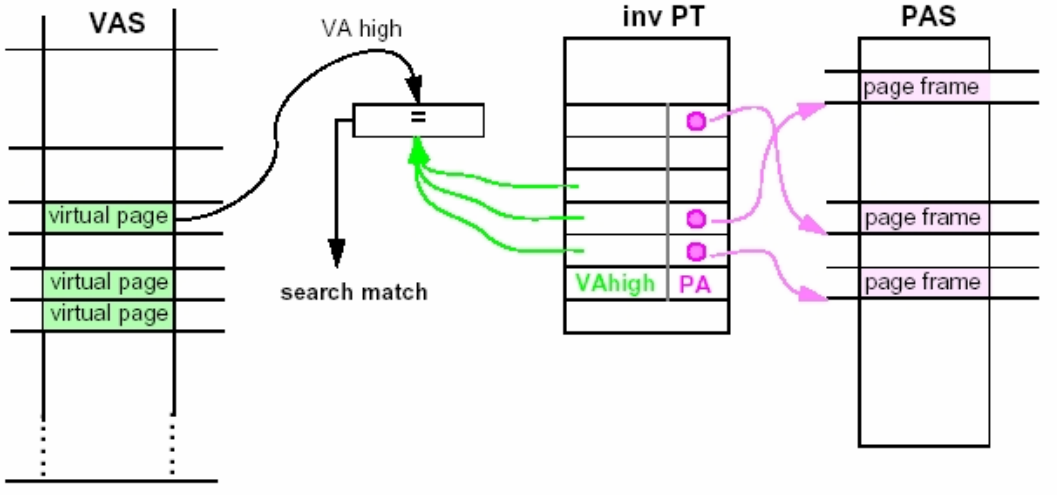


[1]

- Page Table (PT) wird mit der virtuellen Adresse (VA) indiziert und direkt in die physikalische Adresse (PA) umgesetzt
- je VA existiert ein PT-Eintrag

(i) Funktionale Teile der MMU (3)

Inverted Page Table:

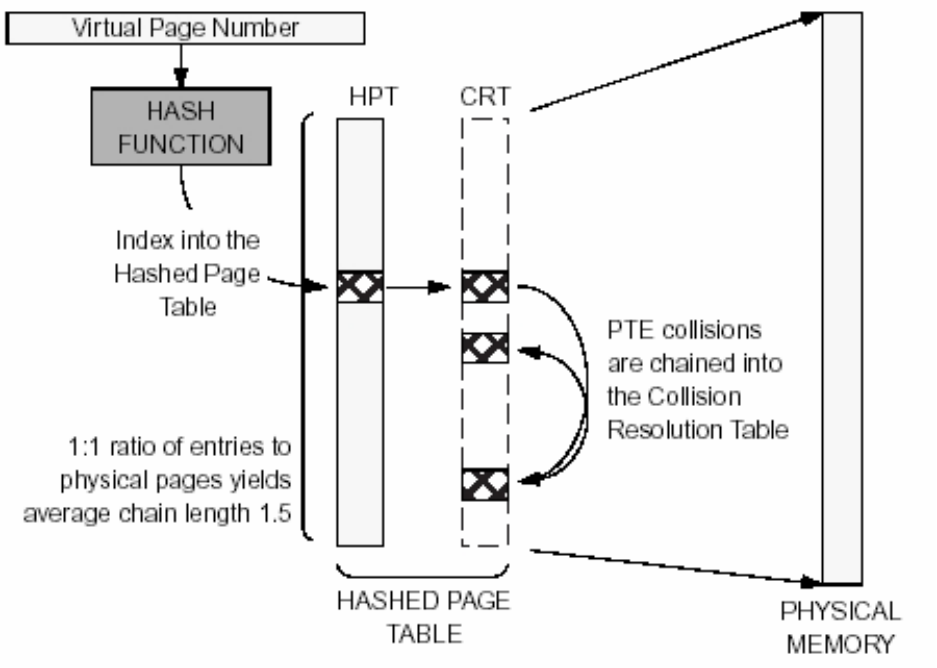


[1]

- Anzahl der PT-Einträge entspricht Anzahl der PAs
- mit linearer Suche oder assoziativem Vergleich wird der passende Eintrag gefunden

(i) Funktionale Teile der MMU (4)

Hashing:

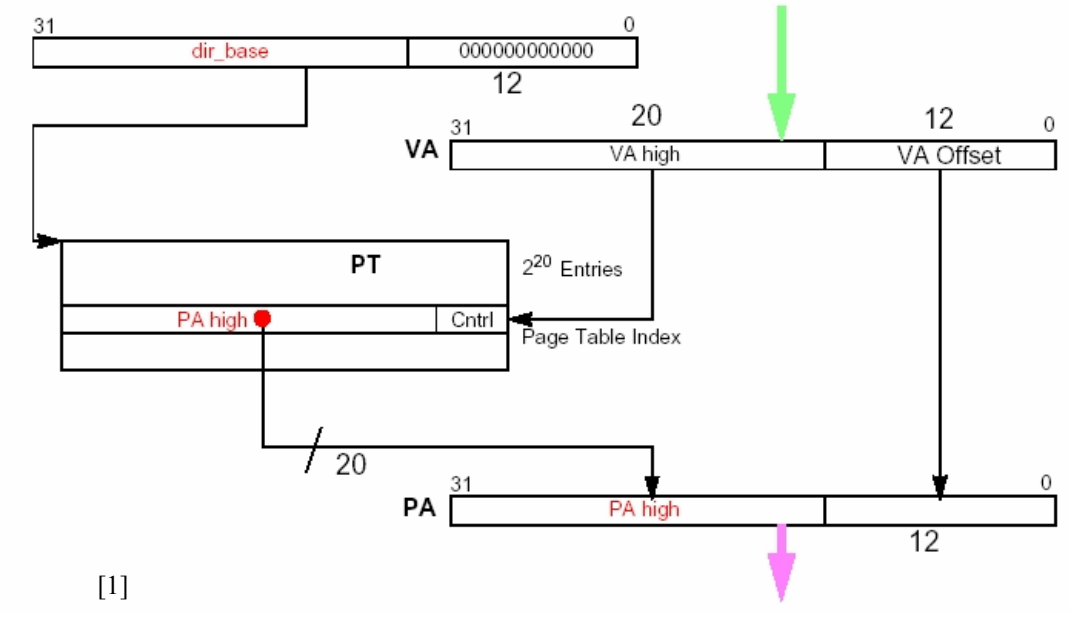


- aus VA und einer Hashfunktion ergibt sich der Index des PT
- durch den Index erhält man den Anfang der Kollisionskett
- Suchen des passenden PA-Eintrags

[2]

(i) Funktionale Teile der MMU (5)

Einstufiges Paging:

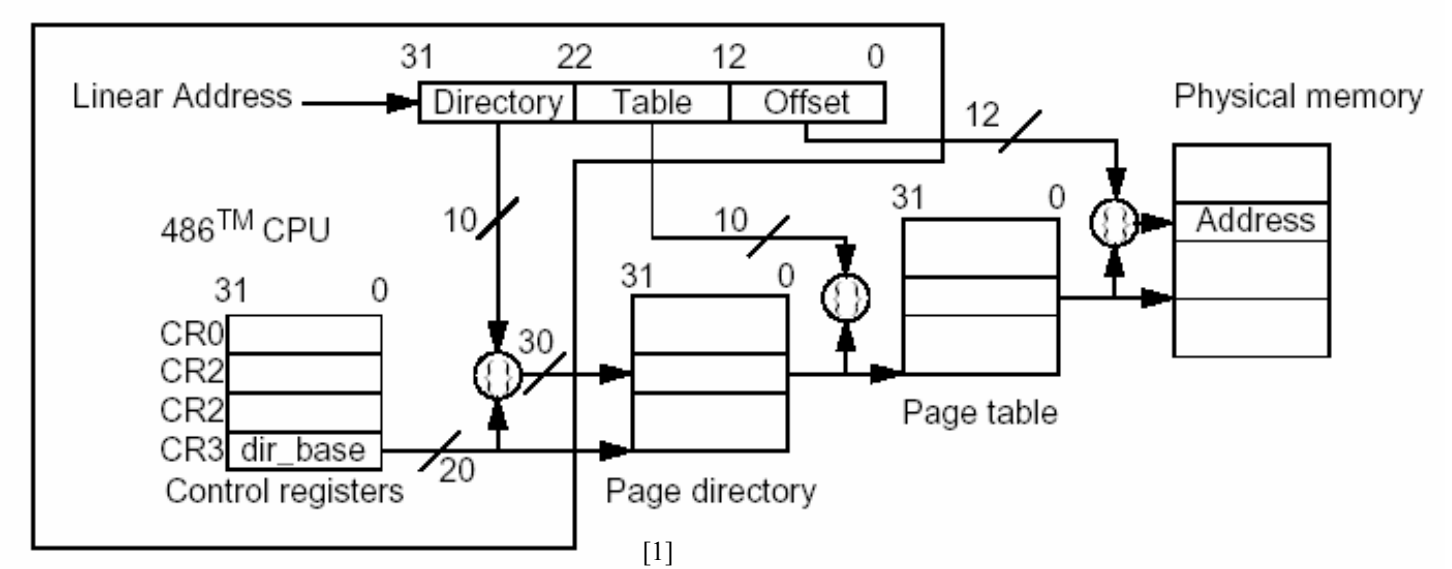


[1]

- VA Aufteilung in Offset und den höherwertigen Teil
- PT wird mit höherwertigen Teil indiziert
- je PT-Eintrag der höherwertigen Teil der PA und den Kontrollbits

(i) Funktionale Teile der MMU (6)

Mehrstufiges Paging:



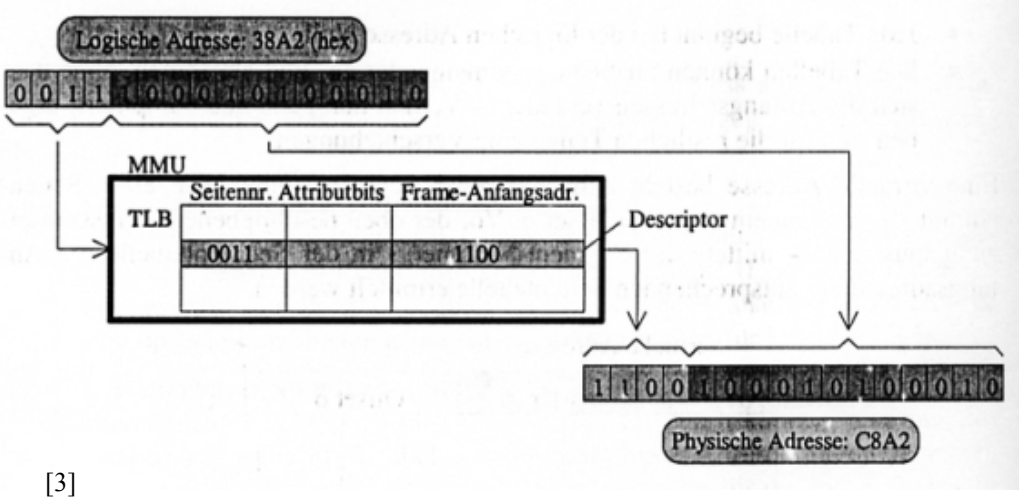
[1]

- VA Aufteilung in Offset den Directory- und den Table-Teil
- durch den Tablewalk erhält man PA

(i) Funktionale Teile der MMU (7)

Translation Lookaside Buffer (TLB):

- TLB besteht aus Descriptoren
- jeder Descriptor enthält die VA, Attributbits und die PA



[3]

(ii) Intel - Itanium2 (1)

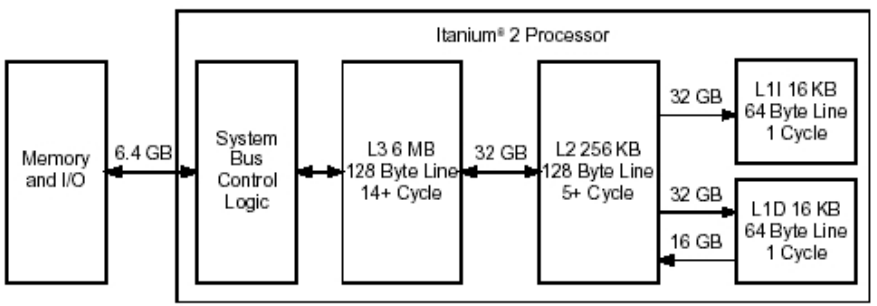
Intel Itanium2 Prozessor: - 900 MHz mit 1,5MB L3-Cache
- 1,0 oder 1,3 GHz mit 3MB L3-Cache
- 1,4 GHz mit 4MB L3-Cache
- 1,5 GHz mit 6MB L3-Cache

Features:
- Explicitly Parallel Instruction Computing (EPIC)
- Bi-endian
- predication und speculation Unterstützung

[4]

(ii) Intel - Itanium2 (2)

- Integrierter L1, L2 und L3 Cache (4-, 8- bzw. 12-way set ass.)
- 128 general und 128 floating-point Register
  (unterstützt Registerrotation, register stack engine)



[4]

- 128-Bit Datenbus
- 6,4 GB/s Bandbreite
- bis zu 4 Prozessoren an einem 400 MHz Datenbus

(ii) Intel - Itanium2 (3)

64-Bit virtuelle Adressierung, 50-Bit physikalische Adressierung

Unterstützte Seitengrößen: 4KB, 8KB, 16KB, 64KB, 256KB, 1MB,
4MB, 16MB, 64MB, 256MB, 1GB und 4G

Zwei Arten von TLBs: - Data Translation Lookaside Buffer (DTLB)
- Instruction Translation Lookaside Buffer (ITLB)

Der DTLB ist unterteilt in: - first Level DTLB (DTLB1)
- second Level DTLB (DTLB2)

Der ITLB ist unterteilt in: - first Level ITLB (ITLB1)
- second Level ITLB (ITLB2)

(ii) Intel - Itanium2 (4)

DTLB1: 32 Einträge, fully associative, unterstützt 4KB Pages oder auch größere Sets, die in 4KB Teile aufgespalten werden, 2 read und 1 write Port

DTLB2: 128 Einträge (64 als Translation Register (TR)), fully associative, unterstützt 4KB bis 4GB große Pages, 4 ports

ITLB1: 32 Einträge, fully associative, nur 4 KB Pages, dual ported

ITLB2: 128 Einträge (64 als Translation Register (TR)), fully associative, unterstützt 4KB bis 4GB große Pages, single ported

(ii) Intel - Itanium2 (5)

- bei einem TLB miss in ITLB oder DTLB wird mit Hilfe des *Hardware Page Walker* (HPW) auf eine 8B und 32B *Virtual Hash Page Table* (VHPT) zugegriffen

- VHPT wird nur im L2 und L3 Cache gehalten (nicht im L1)

- HPW greift auf L2, L3 oder den Speicher zu, um den Page entry zu erhalten

- wenn der HPW die Page nicht findet, wird ein Softwarehandler aufgerufen, um die Umsetzung zu vollenden

(ii) Intel - Itanium2 (6)

- HPW akzeptiert neue Anfragen, während er läuft

- mehrere Anfragen werden immer serialisiert

- HPW kann jederzeit unterbrochen werden

Event	Penalty in Cycles
Hit in L2	25
Miss in L2, hit in L3	31
Miss in both L2 and L3	20 + Main memory Latency (System dependent)

[4]

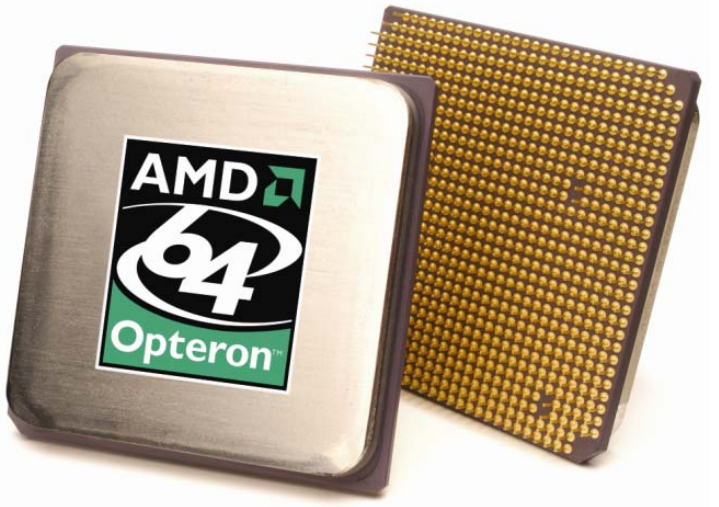
Event	Penalty in Cycles
HPW Abort	OS trap/abort
HPW mapping failed	OS trap/abort

[4]

(iii) AMD - Opteron (1)

AMD Opteron Prozessor: 800MHz bis 2600MHz (0,13µ)

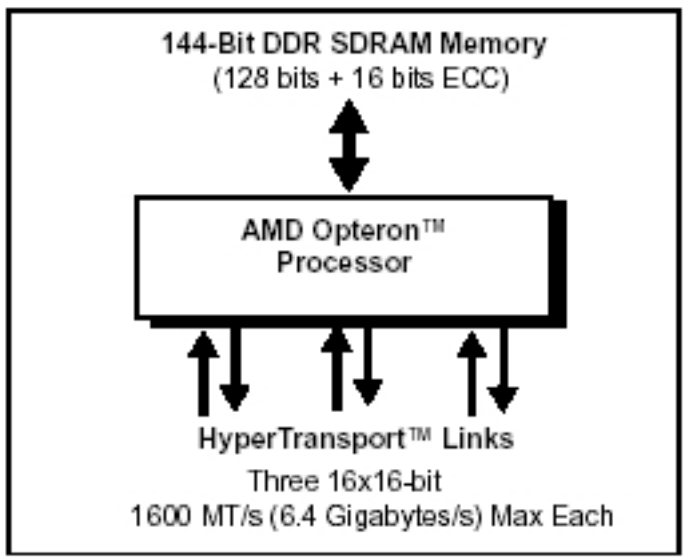
Features:
- advanced branch prediction
- 16-Bit und 32-Bit Kompatibel
- kompatibel mit der SSE2 Technologie (SIMD)



[5]

(iii) AMD - Opteron (2)

- L1-Cache: 64KB D-Cache und 64KB I-Cache (2-way ass., ECC)
- L2-Cache: 1MB on-chip ECC-Protected (16-way ass.)



[5]

- 3 Hyper Transport Links
- 19,2 GB/s
  (3,2 GB/s je Richtung,
  16x16-Bit)
- Integrierter Memory Controller
  mit 6,4 GB/s

(iii) AMD - Opteron (3)

64-Bit virtuelle Adressierung, 52-Bit physikalische Adressierung

Unterstützte Seitengrößen: 4KB, 2MB und 4MB

Unterstützt Segmentierung

(iii) AMD - Opteron (4)

Operating Mode		Operating System Required	Application Recompile Required	Defaults		Register Extensions	Typical
				Address Size (bits)	Operand Size (bits)		GPR Width (bits)
Long Mode	64-Bit Mode	New 64-bit OS	yes	64	32	yes	64
	Compatibility Mode		no	32		no	32
				16	16		16
Legacy Mode	Protected Mode	Legacy 32-bit OS	no	32	32	no	32
				16	16		
	Virtual-8086 Mode			16	16		16
	Real Mode	Legacy 16-bit OS					

[5]

(iii) AMD - Opteron (5)

- Real Mode Segmentation (Virtual-8086 Mode Segmentation):
    - 64KB Segmente (CS, DS, ES, FS, GS, SS)
    - Umsetzung in 20-Bit lineare Adresse

- Protected Mode Segmented-Memory Models:
    - Multi-Segmented Model:
        - Größe von 1B bis 4GB
        - unique base address
    - Flat-Memory Model:
        - die base address aller Segmente ist 0

(iii) AMD - Opteron (6)

Page Translation:

- Hierarchical Page Tables softwaregesteuert
- die virtuelle Adresse ist in Felder unterteilt, die als offset für einen Tablewalk verwendet werden
- es gibt vier Kontrollbits:
  - Page-Translation Enable (CR0.PG)
  - Physical-Address Extensions (CR4.PAE)
  - Page-Size Extensions (CR4.PSE)
  - Long-Mode Active (EFER.LMA)

  mit denen das geeignete Verfahren gewählt wird

(iii) AMD - Opteron (7)

Legacy Mode Page Translation:

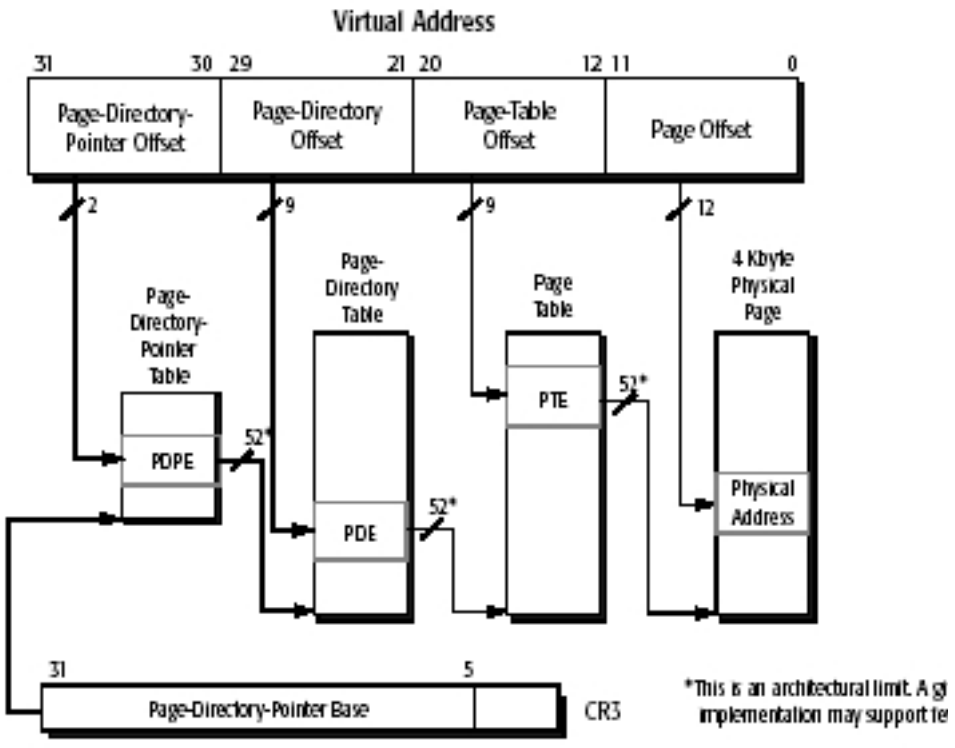
- Normal Paging (CR4.PAE=0):
  - Page Translation Table Enty 32-Bit
  - 32-Bit VA => 40-Bit PA
- PAE Paging (CR4.PAE=1):
  - Page Translation Table Enty 64-Bit
  - 32-Bit VA => 52-Bit PA

Page Translation Tables:

- Page Table (PT) zeigt auf 4KB Page
- Page Directory (PD) zeigt auf PT oder 2MB bzw. 4MB Pag
- Page Directory Pointer (PDP) zeigt auf PD

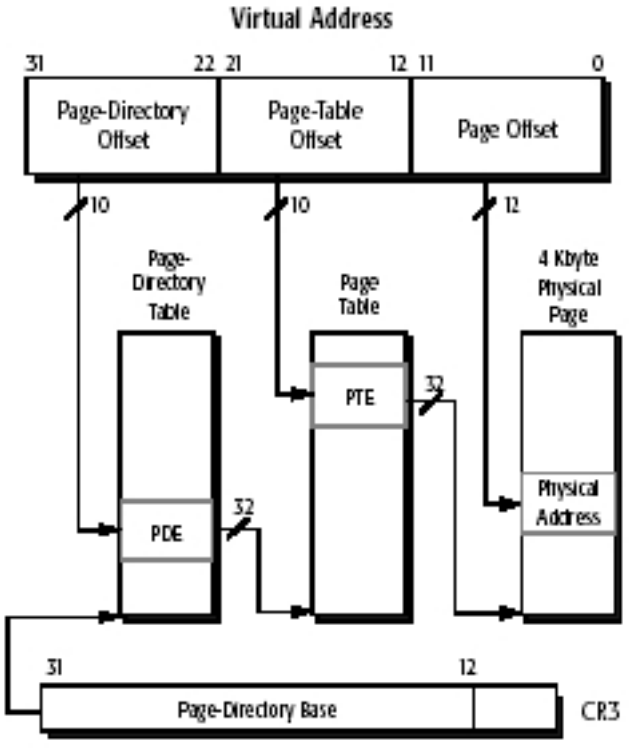
(iii) AMD - Opteron (8)

PAE Paging:



[5]

Normal Paging:

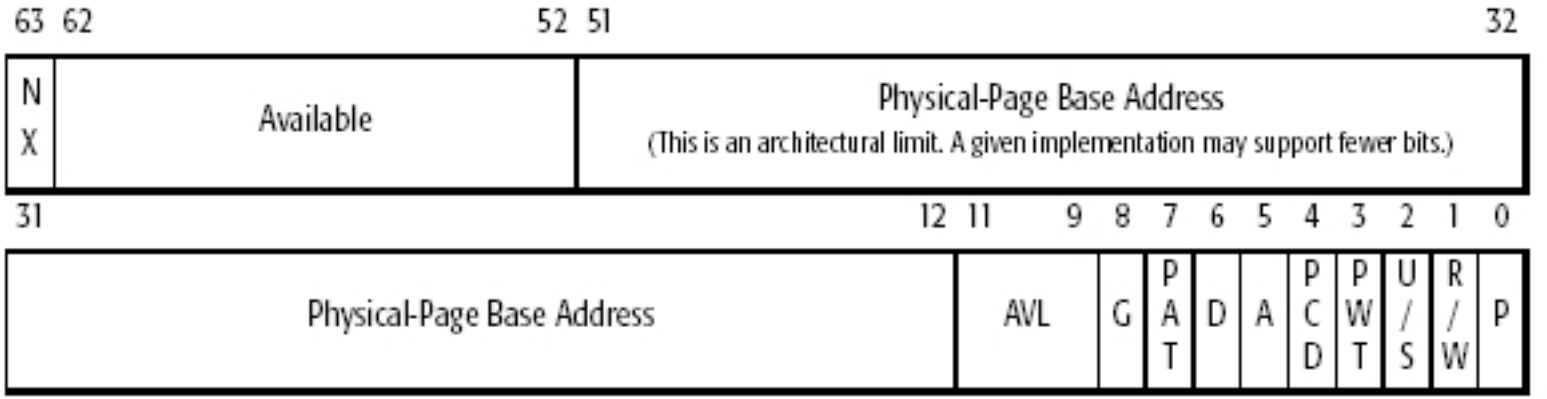


[5]

(iii) AMD - Opteron (9)

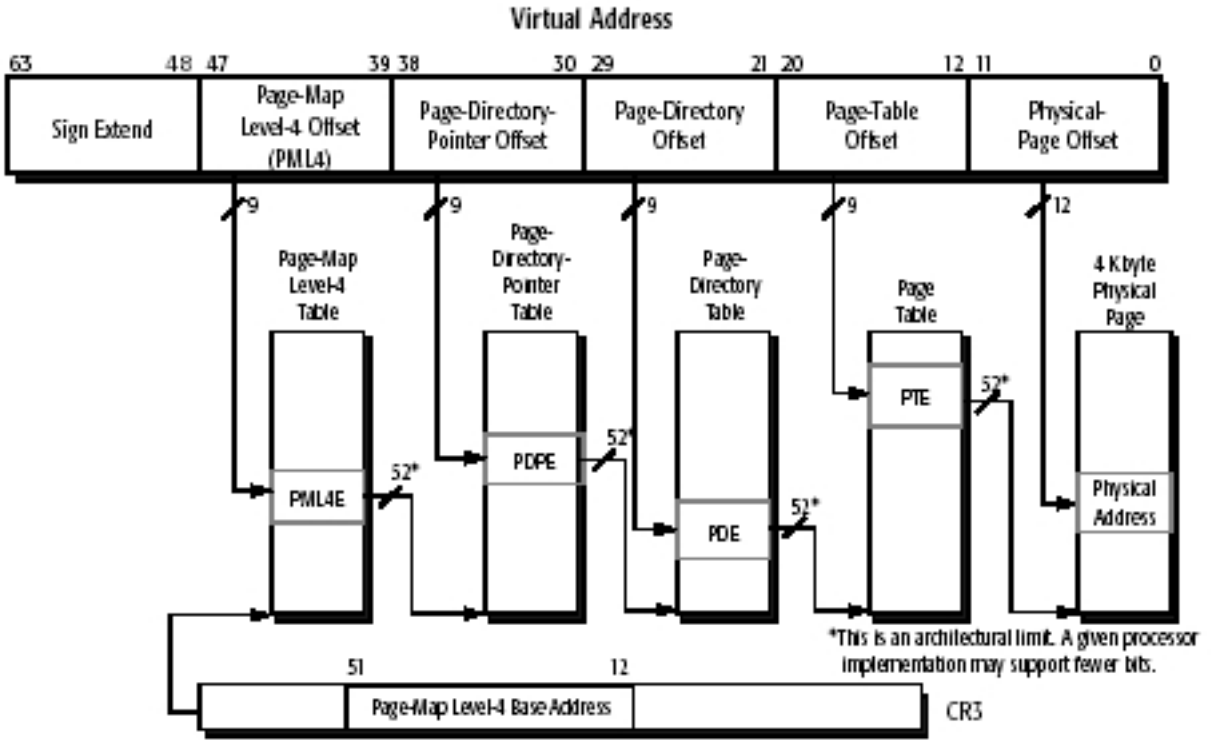
Long Mode Page Translation:
(CR4.PAE=1 muss gesetzt sein bevor EFER.LMA=1)

- ein neuer Table wir an das PAE Paging angefügt, der page-map level-4 (PML4)



[5]

(iii) AMD - Opteron (10)



[5]

(iv) Apple/IBM - G5 (1)

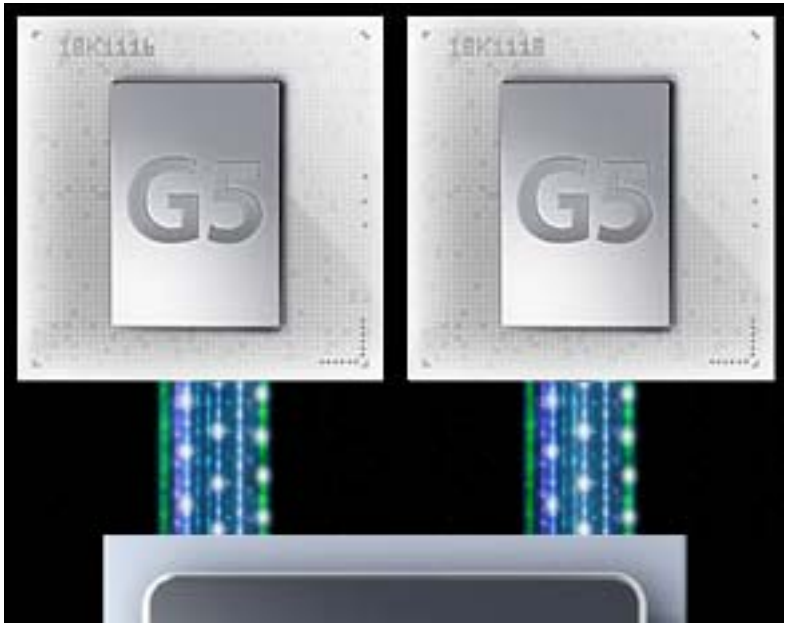
Power PC G5 Prozessor: 1,6GHz, 1,8GHz und dual 2GHz (0,13µ, 8 layer)

Features:
- Dual-pipeline Velocity engine (128-bit SIMD processing)
- Highly accurate dynamic branch prediction
- 32-Bit Kompatibel

[6]

(iv) Apple/IBM - G5 (2)

- L1-Cache: 64KB instruction (direct-mapped) und
  32KB data (2-way set ass.)
- L2-Cache: 512KB on-chip (8-way set ass.)



- 2 unidirectional busses
  (32-Bit read, 32-Bit write)
- bis zu 1GHz 64-Bit DDR
  frontside bus mit 8 GB/s
  (im zwei Prozessorsystem
  einer auf jedem)

(iv) Apple/IBM - G5 (3)

64-Bit effektive Adressierung, 42-Bit physikalische Adressierung

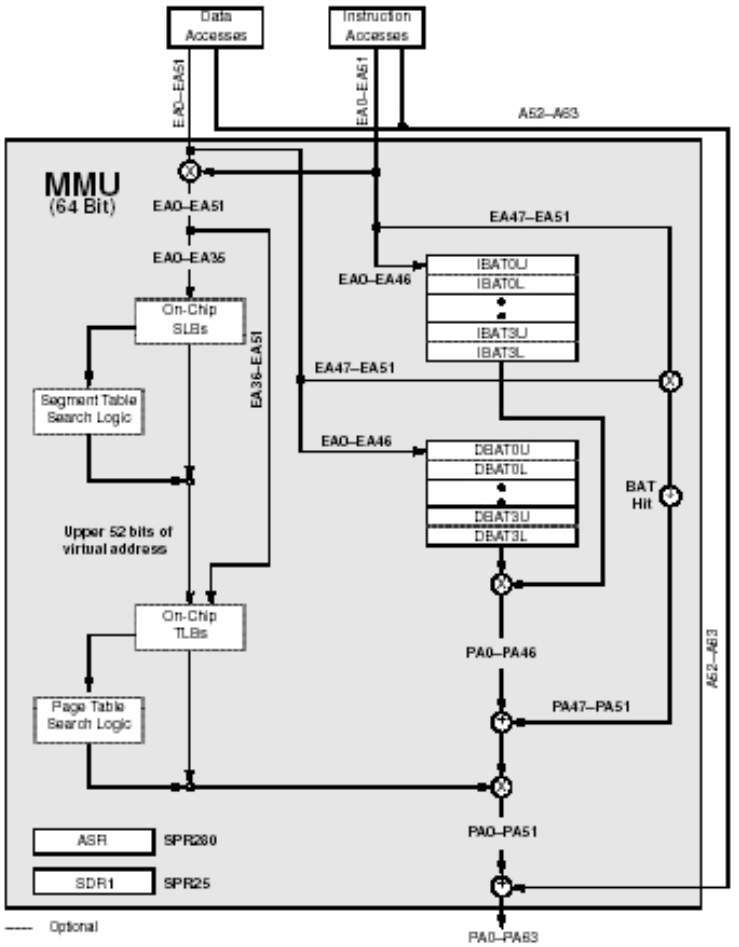
Unterstützte Seitengröße: 4KB

Unterstützte Segmentgröße: 256MB

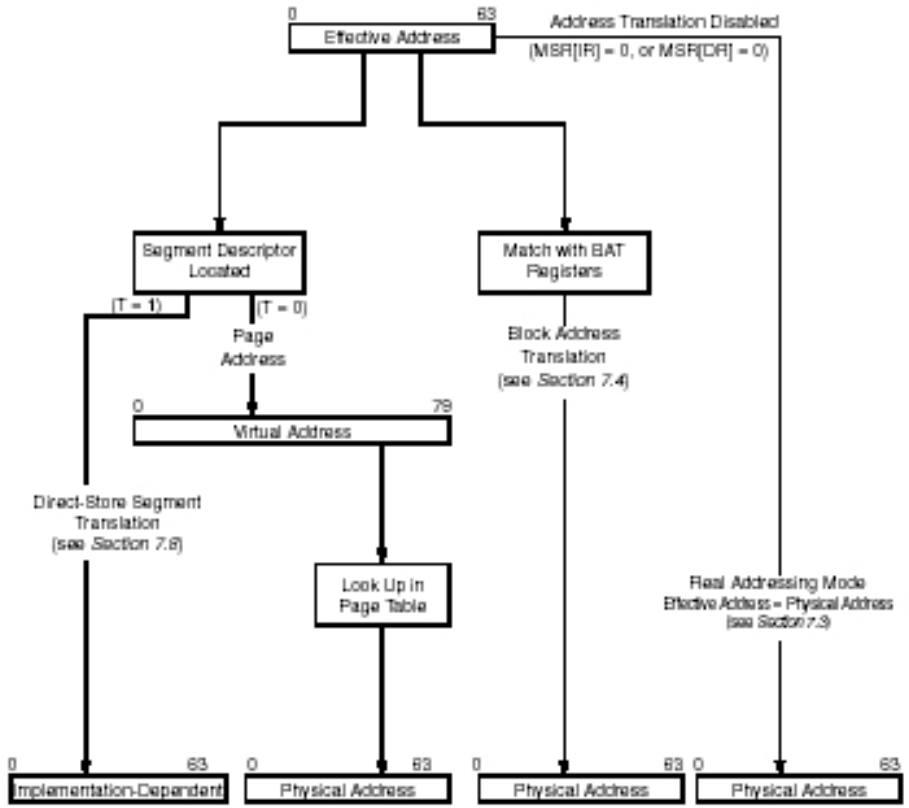
Unterstützte Blockgröße: 128KB bis 256MB

Segment Lookaside Buffer (SLB): 64 Einträge fully ass.
(davon 16 Segment Register)
Translation Lookaside Buffer (TLB): 256x4-way TLB

(iv) Apple/IBM - G5 (4)



[6]



[6]

(iv) Apple/IBM - G5 (5)

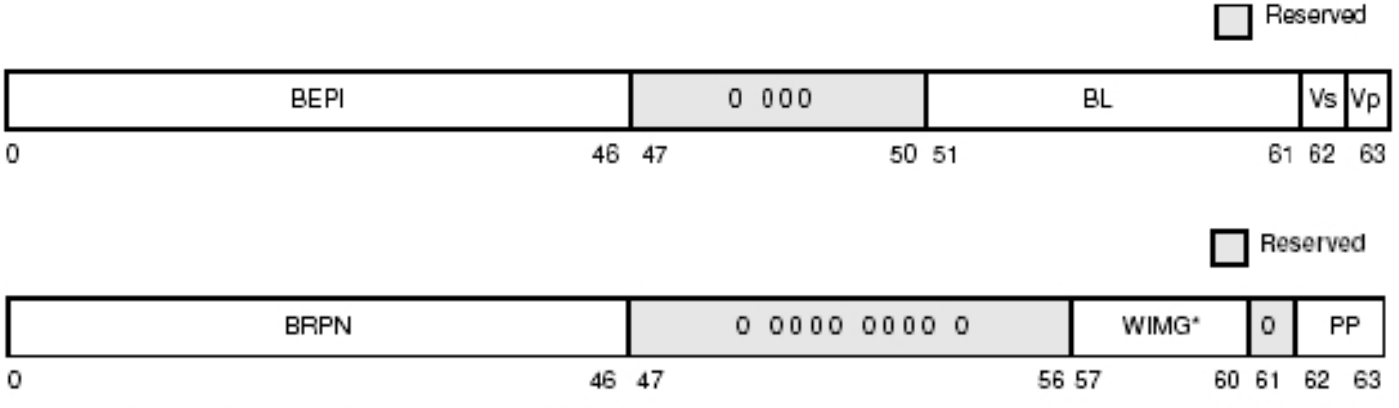
Real Addressing Mode:

- bei ausgeschalteter Adressübersetzung wird die effektive Adresse wie eine physikalische Adresse genutzt
- ist die physikalische Adresse kleiner werden die extra high-order Bits ignoriert

(iv) Apple/IBM - G5 (6)

Block Address Translation (BAT):

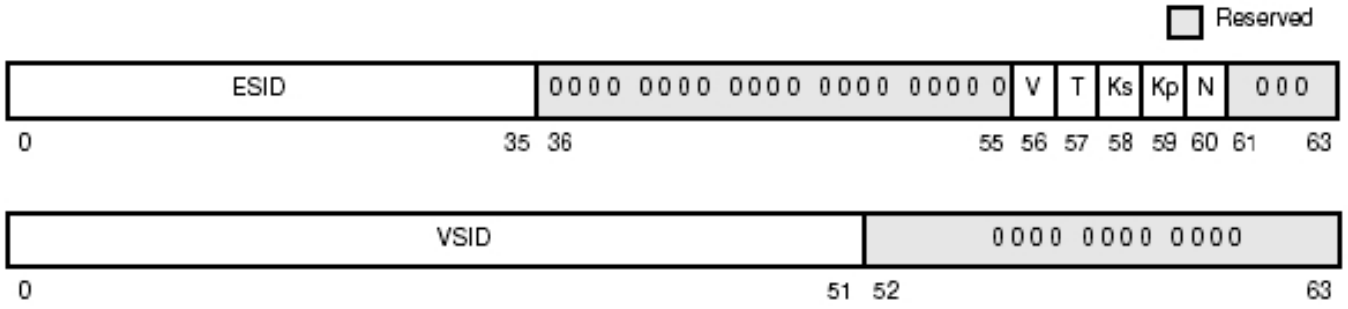
- 16 special-purpose Registers (SPRs)
- je Block ein upper und lower SPR (BAT Register)
- zwei fully associative BAT-arrays mit vier Einträgen
- BAT Register enthalten die Anfangsadresse des Blocks effektiv und physikalisch, sowie die Größe



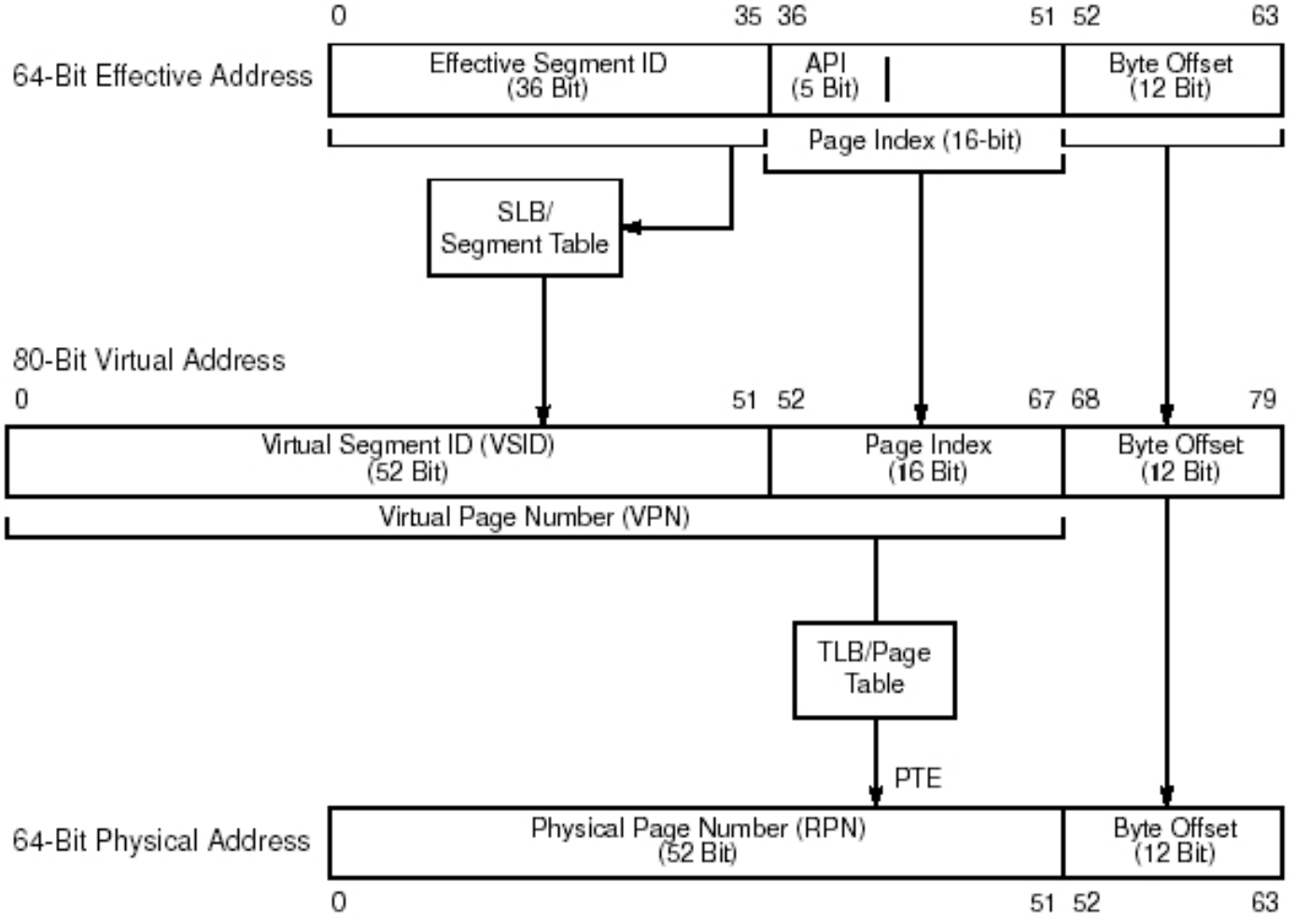
(iv) Apple/IBM - G5 (7)

Memory Segment Model:

- zwei Segmenttypen: - Memory Segment
  - Direct-store Segment
- zwei Stufen: 1. effektive Adresse => virtuelle Adresse
  2. virtuelle Adresse => physikalische Adresse



(iv) Apple/IBM - G5 (8)

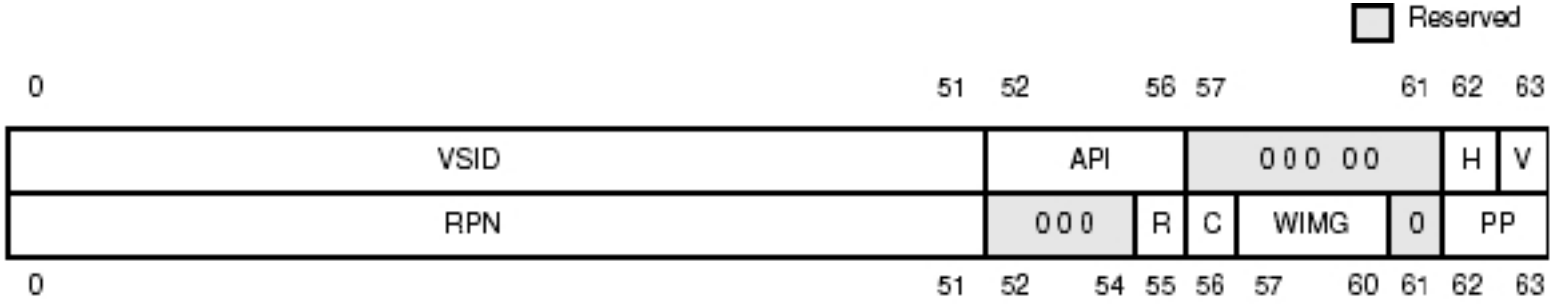


[6]

(iv) Apple/IBM - G5 (9)

Page Table Entry (PTE):

- 128-Bit PTE
- dient zur Umsetzung der virtual page number(VPN) in die physical page number (RPN)
- enthält nur die VSID und den abbreviated page index (API) der VPN
- die fehlenden 11-Bits des Page Index werden für eine Hashfunktion genutzt, die PTE-Gruppen erzeugt



[6]

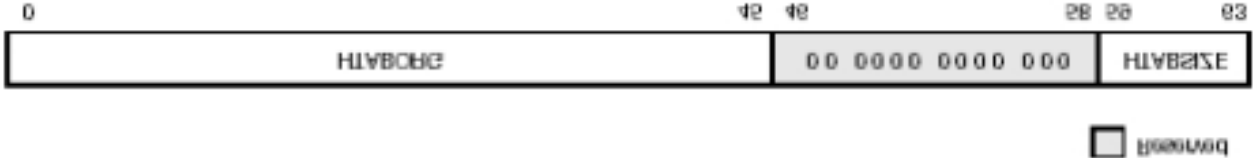
(iv) Apple/IBM - G5 (10)

Hashed Page Table (HPT):

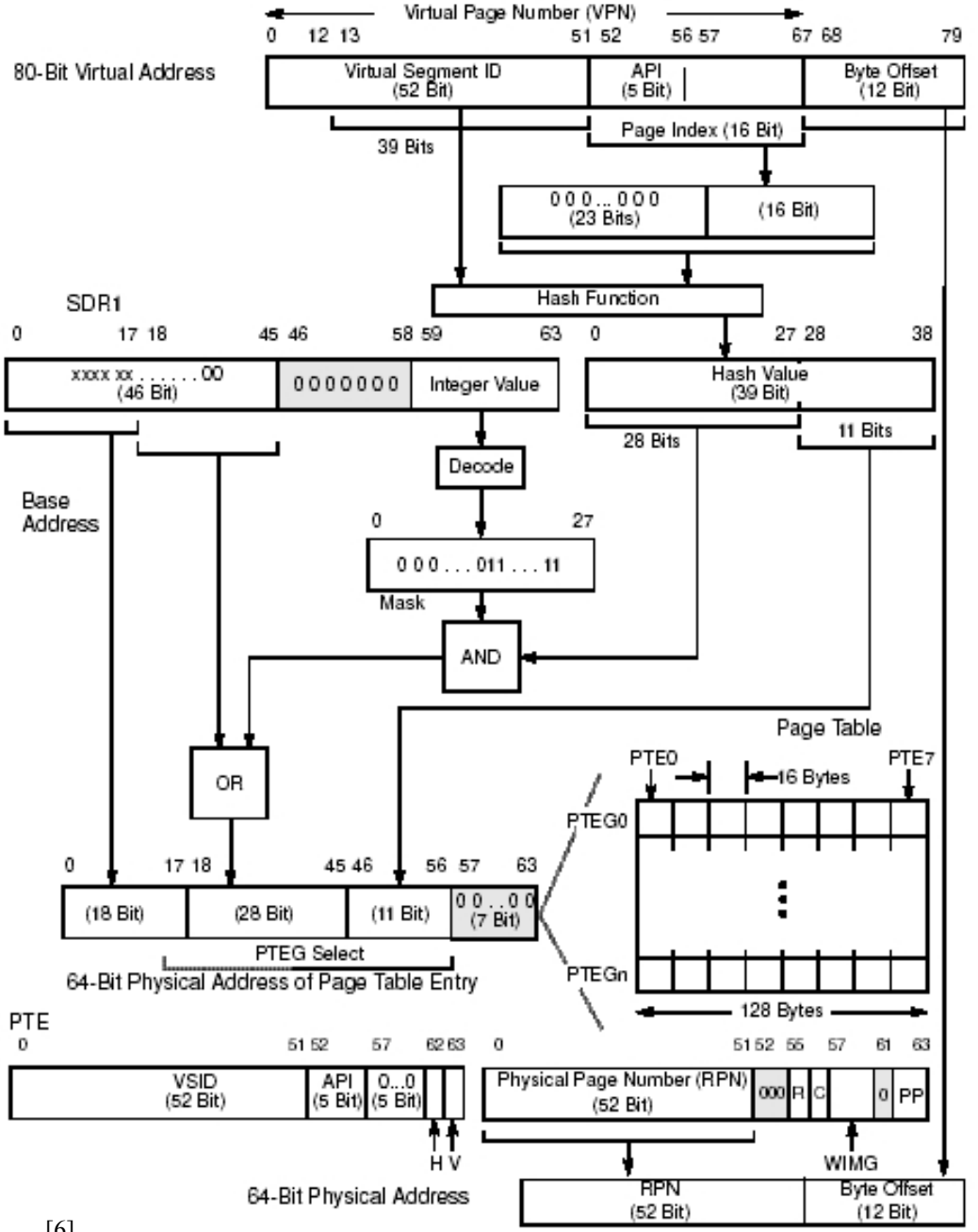
- Datenstruktur variabler Größe (2^{18}-2^{46} Bytes)
- besteht aus page table entry groups (PTEGs, 2^{11}-2^{39})
- PTEG besteht aus acht PTEs je 16-Bytes

HPT Suche: - bei der Suche wird das Ergebnis der primären Hash-funktion mit dem Wert im SDR1 Register verknüpft, man erhält die PA der PTEG

- ist die Suche in der PTEG nicht erfolgreich, wird die sekundäre Hashfunktion genutzt



[6]

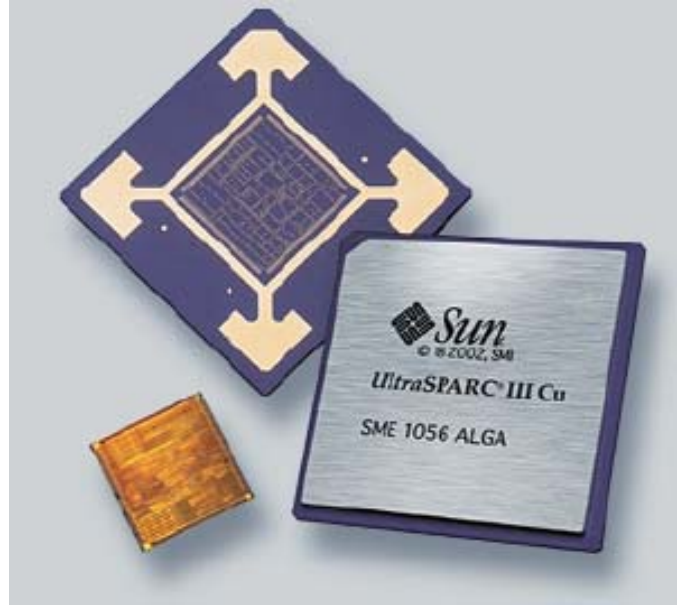

Virtual Page Number (VPN)
0 12 13 51 52 56 57 67 68 79
80-Bit Virtual Address
Virtual Segment ID (52 Bit)
API (5 Bit)
Byte Offset (12 Bit)
Page Index (16 Bit)
39 Bits
0 0 0 ... 0 0 0 (23 Bits)
(16 Bit)
Hash Function
SDR1
0 17 18 45 46 58 59 63
xxxx xx . . . . . 00 (46 Bit)
0 0 0 0 0 0 0
Integer Value
0 27 28 38
Hash Value (39 Bit)
28 Bits
11 Bits
Decode
Base Address
0 27
0 0 0 . . . 0 1 1 . . . 1 1
Mask
AND
OR
Page Table
PTE0
16 Bytes
PTE7
PTEG0
PTEGn
128 Bytes
0 17 18 45 46 56 57 63
(18 Bit)
(28 Bit)
(11 Bit)
0 0 . . 0 0 (7 Bit)
PTEG Select
64-Bit Physical Address of Page Table Entry
PTE
0 51 52 57 62 63
VSID (52 Bit)
API (5 Bit)
0...0 (5 Bit)
H V
0 51 52 55 57 61 63
Physical Page Number (RPN) (52 Bit)
000
R
C
0
PP
WIMG
64-Bit Physical Address
RPN (52 Bit)
Byte Offset (12 Bit)

[6]

(v) Sun - UltraSparc III (1)

UltraSparc III Prozessor: 900 MHz, 1050MHz und 1,2 GHz
UltraSparc IIIi Prozessor: 1GHz
(0,13µ, 7 layer)

Features:
- 4-way superscalar
- 14 stage non-stalling pipeline
- Advanced RAS features
  (IIIi: Robust RAS features)
- Distributed Memory Contol



[7]

(v) Sun - UltraSparc III (2)

- L1-Cache: instruction 32KB, data 64KB, 2KB prefetch Cache
  and 2KB write Cache (4-way set ass.)
- L2-Cache: 2, 4 oder 8MB external (2-way set ass.)
  (IIIi: 1MB on-chip, 4-way set ass.)

- Systembus mit 150 MHz Clock Frequency
- Architekturdesign für >1000 CPUs/system
- IIIi: bis zu 4 Prozessoren, 183-Bit Datenbus, bis zu 200MHz
  Memory Interface 266 DDR-1, 4,25 GB/s

(v) Sun - UltraSparc III (3)

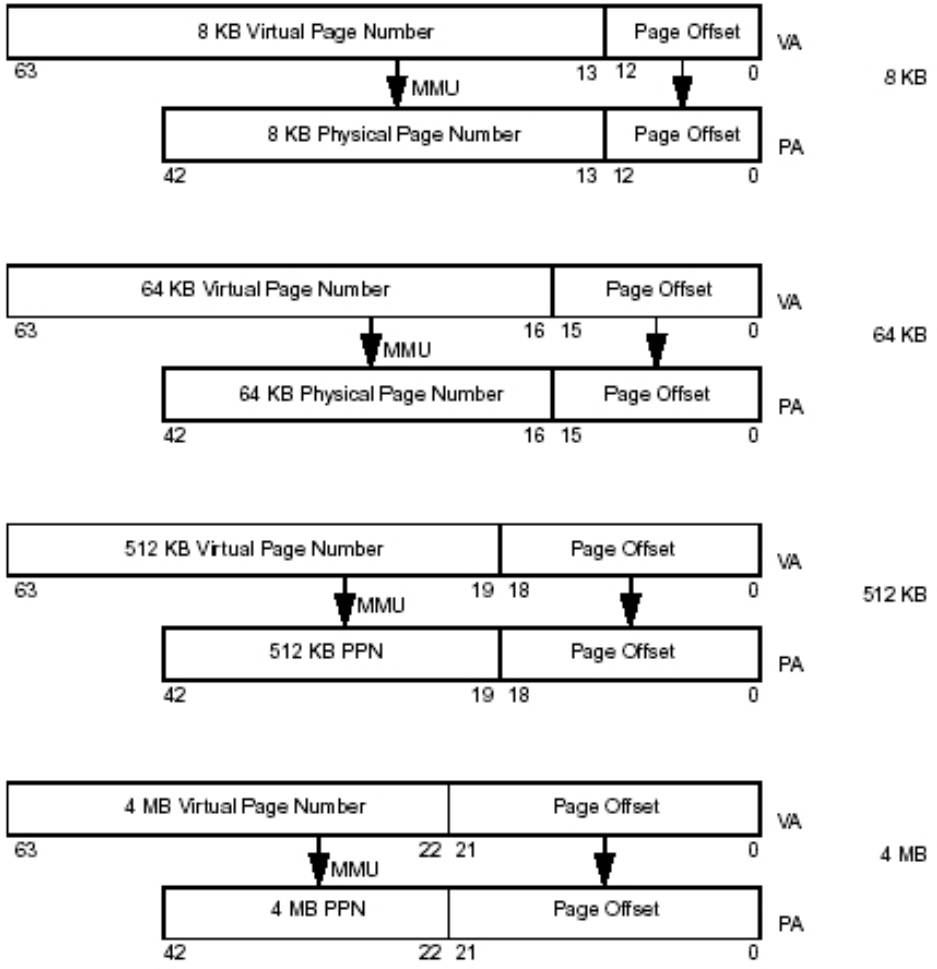
64-Bit virtuelle Adressierung, 43-Bit physikalische Adressierung

Unterstützte Seitengrößen: 8KB, 64KB, 512KB und 4MB

Zwei Arten von MMUs: - Data MMU (D-MMU)
- Instruction MMU (I-MMU)

- MMUs bestehen aus mehreren TLBs
- *Software Translation Table* erledigt die Adressumsetzung

(v) Sun - UltraSparc III (4)



[7]

(v) Sun - UltraSparc III (5)

D-MMU: - 2 TLBs mit 512 Einträgen, 2-way associative,
unterstützt alle Seitengrößen
- 1 TLB mit 16 Einträgen, fully associative,
unterstützt alle Seitengrößen

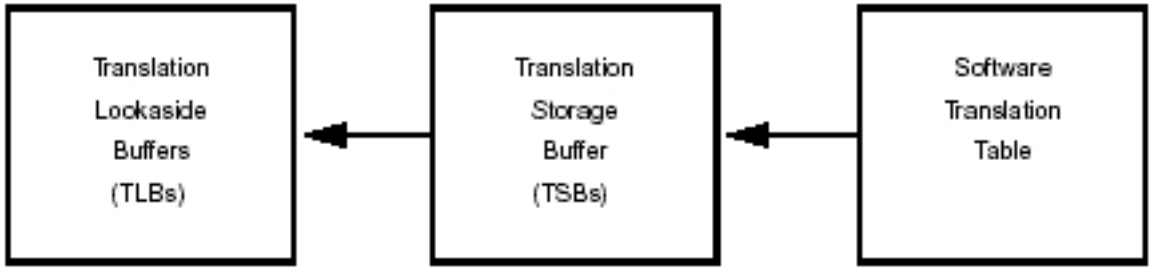
I-MMU: - 1 TLB mit 128 Einträgen, 2-way associative,
nur 8KB Seiten
- 1 TLB mit 16 Einträgen, fully associative,
für 64KB, 512KB und 4MB Seitengrößen
(8KB nur locked)

(v) Sun - UltraSparc III (6)

- TLB miss löst trap aus, die einen Software Miss Handler aktiviert

- Software Miss Handler füllt der TLB aus dem *Software Translation Table* auf

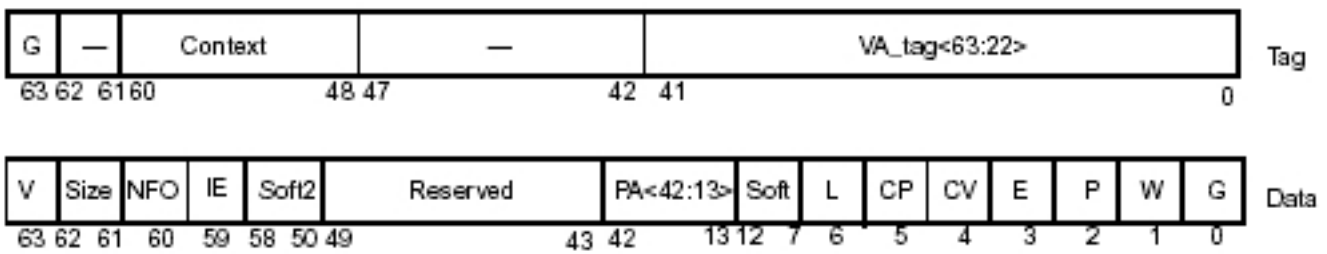
- Translation Storage Buffer (TSB) dient als direct-mapped Cache für den langsamen Software Translation Table



[7]

(v) Sun - UltraSparc III (7)

Translation Table Entry (TTE): - enthält die Informationen für ein single page mapping
- besteht aus 64-Bit Worten für Tag und Daten
- der Tag funktioniert wie im Cache
- die Daten werden mit Hilfe der Software beschafft



[7]

(v) Sun - UltraSparc III (8)

Translation Storage Buffer (TSB):

- ein Feld aus TTEs
- der „Cache“ des *software translation table*
- TLB Einträge müssen nicht im TSB sein
- TSB Pointer (mit Hardwareunterstützung)
- TSB Indexing Support
- TSB Organization

(vi) Zusammenfassung

Prozessor	Block Address Translation	Segment Address Translation	Page Address Translation	Translation Table
Intel - Itanium2	Nein	Nein	4KB - 4GB	Virtual Hash Page Table
AMD - Opteron	Nein	1B - 4GB	4KB, 2MB, 4MB	Hierarchical Page Tables
Apple/IBM - G5	128KB - 256MB	256MB	4KB	Hashed Page Table
Sun – UltraSparc III	Nein	Nein	8KB, 64KB, 512KB, 4MB	Software Translation Table

Literaturliste

[1]: Prof. Dr. U. Brüning, Vorlesungsskript LS Rechnerarchitektur

[2]: Bruce L. Jacob, Trevor N. Mudge
„A Look at Several MemoryManagement Units,
TLB-Refill Mechanisms, and Page Table Organizations“
(ASPLOS-VIII), San Jose CA, Oct. 3-7, 1998

[3]: Malz, „Rechnerarchitektur“, Vieweg Verlag

[4]: http://www.intel.com

[5]: http://www.amd.com

[6]: http://www.ibm.com und http://www.apple.com

[7]: http://www.sun.com