

Y36SAP-11

Paměťová hierarchie
Virtualizace paměti
Vyrovnávací (cache) paměť

24.5.2007

Y36SAP-Paměťový subsystém

1

Organizace paměťového systému počítače

- Paměťová hierarchie ... několikaúrovňové uspořádání pamětí různých typů (kapacit a rychlostí) s cílem dosáhnout optimálního poměru výkon x cena
- Cena paměti je cca přímo úměrná kapacitě nepřímo dobře přístupu (rychlosti)

24.5.2007

Y36SAP-Paměťový subsystém

2

Paměťová hierarchie

typ paměti	Typická realizace	Doba přístupu	Kapacita
registry	klopné obvody	jednotky ns	desítky-stovky B
vyrovnávací paměť	statická RAM	10-20ns	stovky kB-jednotky MB
hlavní paměť	dynamická RAM	50-70ns	desítky-stovky MB
vnější paměť	pevný disk	5-15ms	Jednotky-desítky GB
záložní paměť	optický disk mag. páska ZIP drive	stovky ms- Stovky s	Stovky GB- jednotky TB

CENA !!!

24.5.2007

Y36SAP-Paměťový subsystém

3

Virtuální paměť

- systém několika pamětí s různými parametry (kapacita, rychlost), řízený tak, aby vytvářel paměťové prostory potřebné velikosti pro program a data
- umožňuje realizaci jednoho nebo několika logických (virtuálních) adresových prostorů, kde každý může být větší než skutečná kapacita hlavní paměti
- hlavní paměť ... fyzický paměťový prostor
- logické adresové prostory jsou ve skutečnosti ve vnější paměti

24.5.2007

Y36SAP-Paměťový subsystém

4

Virtuální paměť

- části programů a data jsou přesouvány do hlavní paměti, požaduje-li k nim procesor přístup
- v hlavní paměti jsou jen ty programy a data, s kterými procesor právě pracuje
- lze přemístit části programu bez nutnosti je znovu překládat
- zajištění ochrany dat před neoprávněným přístupem a modifikací
- práce s logickými adresami ne absolutními
- hlavní paměť se adresuje fyzickými adresami
- překlad logických adres na fyzické zajišťuje mechanismus virtuální paměti

Stránkování

Logický adresový prostor je rozdělen na úseky pevné délky - **stránky** (logické stránky)

Fyzický adresový prostor je rozdělen na stejně velké úseky - **stránkové rámce** (fyzické stránky)

Logický adresový prostor je realizován ve vnější paměti. Data (úseky programu) se přesouvají do hlavní paměti po jednotlivých stránkách, jsou-li v průběhu výpočtu požadována a pokud se příslušná stránka již v paměti nenachází.

Překlad (určení kam se stránka do hlavní paměti přesune) používá datovou strukturu ... **tabulku stránek**

Tabulka stránek

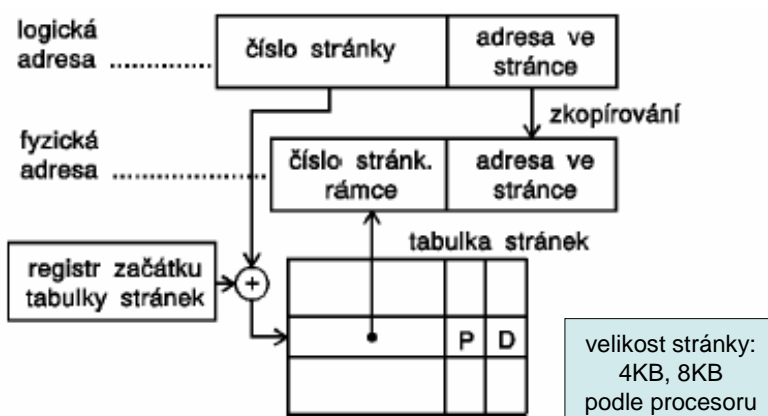
- je uložena v hlavní paměti
- obsahuje pro každou logickou stránku jednu položku
- položka obsahuje informaci, zda se daná stránka nachází v hlavní paměti a pokud ano, tak kde (v kterém stránkovém rámci)

24.5.2007

Y36SAP-Paměťový subsystém

7

Stránkovací mechanismus



24.5.2007

Y36SAP-Paměťový subsystém

8

Realita, problémy

- tabulka stránek musí obsahovat 1 položku pro každou stránku v logickém adresovém prostoru, i když není použita
- při mnohem větším logickém prostoru proti velikosti hlavní paměti, může tabulka stránek zabírat velkou část hlavní paměti

Příklad:

virt. paměť	4GB	32b
hlavní paměť	4MB	22b
velikost stránky	4kB	12b

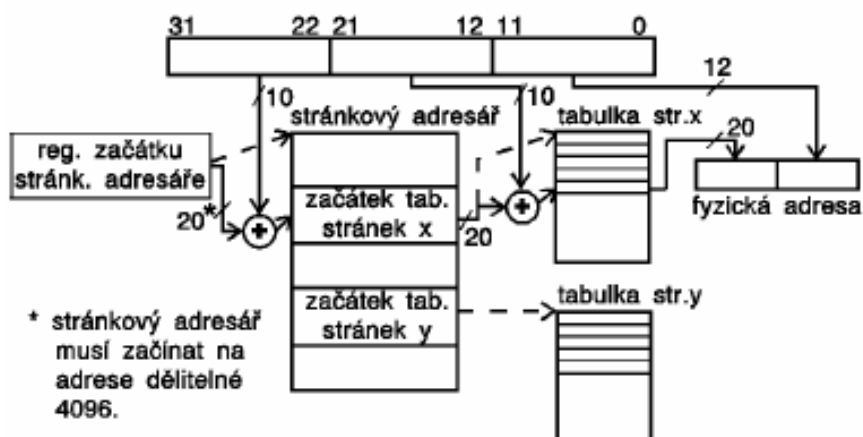
LA	20	12
FA	10	12

Tabulka stránek má 2^{20} položek

1 položka ... 10 bitů	horní část fyzické adresy	} 2B pro 1 položku
1 bit	platnost	

⇒ tabulka stránek zabírá 2MB - nepřipustné.

Dvouúrovňová organizace tabulky stránek



24.5.2007

Y36SAP-Paměťový subsystém

10

Obsah tabulky stránek

- horní část fyzické adresy
- příznak přítomnosti stránky v hlavní paměti
- příznak změny dat ve stránce (zda byla po dobu přítomnosti v HP do stránky zapisováno) ... **Dirty bit**
- další bity, např. určující, zda je vhodné stránku přepsat (vyhodit z hlavní paměti, podle toho jak a kdy byla použita)
-

24.5.2007

Y36SAP-Paměťový subsystém

11

Stránkovací mechanismus

- je-li stránka přítomna v hlavní paměti, přeloží se logická adresa na fyzickou (příznak přítomnosti stránky v HP)
- není-li stránka přítomna, vyvolá se přerušení. Přerušovací mechanismus vyvolá načtení stránky z vnější paměti.
- pokud při načítání není volný žádný stránkový rámec v HP, je třeba nějaký uvolnit ... přesun vhodné stránky do vnější paměti (např. nejdéle nepoužité, viz další bity, sl. 11 ...)
- nebylo-li do stránky zapisováno (Dirty bit), nemusí se přesouvat do vnější paměti, stačí ji jen přepsat.

24.5.2007

Y36SAP-Paměťový subsystém

12

Segmentace

- segmenty jsou funkčně samostatné části programu **proměnné délky**, které lze do hlavní paměti zavádět v případě potřeby
- adresy v segmentu jsou relativní vůči začátku (tzv. bázi) segmentu ... umožňuje přemístitelnost segmentů v HP
- logická adresa se skládá z báze segmentu a offsetu (posunutí)
- báze segmentu je
 - v segmentovém registru
 - tabulce popisovačů segmentů

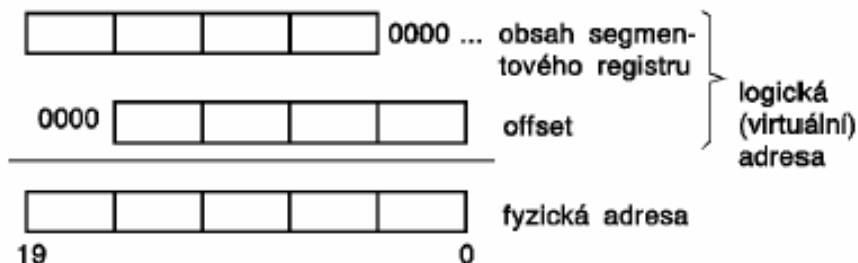
24.5.2007

Y36SAP-Paměťový subsystém

13

Segmentové registry

Příklad z Intel 80x86 16 bitové registry, 20 bitová adresová sběrnice



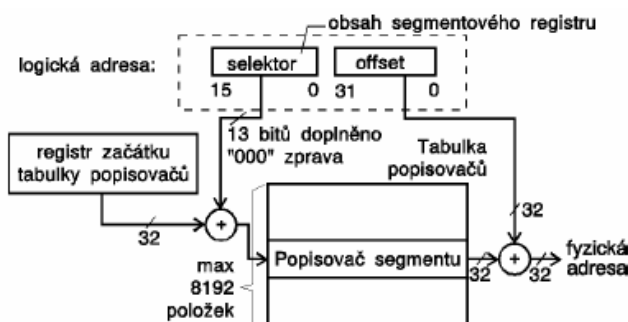
24.5.2007

Y36SAP-Paměťový subsystém

14

Tabulky popisovačů segmentů

Příklad z Intel 80386 "chráněný režim"



24.5.2007

Y36SAP-Paměťový subsystém

15

Popisovač segmentu obsahuje (u 80386 a vyšších, délka 8B)

- 32-bitovou bázi segmentu
- 20-bitový limit segmentu (max. délka)
- tzv. bit G (Granularity), udávající zda limit segmentu je vyjádřen ve slabikách nebo v blocích à 4096 slabik (4kB)
- přístupová práva

Výhody řešení ad b)

- fyzický adresový prostor (velikost hlavní paměti) až 4GB
- logický adresový prostor max 8192 segmentů à max 4GB
- umožňuje přiřadit segmentům přístupová práva - vzájemná ochrana jednotlivých běžících programů a dat
- hodnota selektoru není v přímém vztahu k fyzické adrese. Slouží jako index do předem připravené tabulky popisovačů jednotlivých segmentů.

24.5.2007

Y36SAP-Paměťový subsystém

16

Princip rychlé vyrovnávací paměti

- "malá" rychlá paměť zařazená mezi procesor a hlavní paměť
- využití asociativního přístupu k položkám
- obsahuje kopie nejčastěji používaných položek hlavní paměti
- realizována SRAM
- synonyma cache, buffer memory, skrytá paměť (SP), vyrovnávací paměť, ...

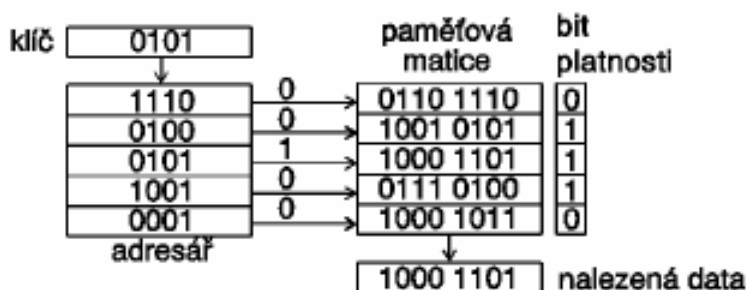
24.5.2007

Y36SAP-Paměťový subsystém

17

Princip plně asociativní paměti

- adresuje se částí datové položky, která se má vyhledat, tzv. **klíčem**
- na rozdíl od adresovatelné paměti např. SRAM neobsahuje adresový dekodér, ale adresář:

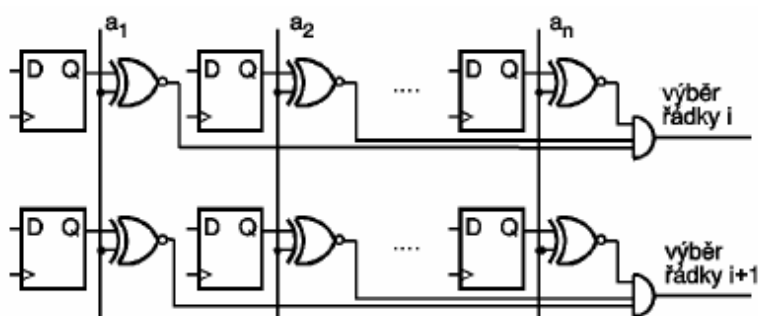


24.5.2007

Y36SAP-Paměťový subsystém

18

Struktura adresáře plně asociativní paměti



Každá položka adresáře obsahuje logické obvody umožňující najednou prohledat všechny položky adresáře

24.5.2007

Y36SAP-Paměťový subsystém

19

Využití asociativní paměti v počítači

- data jsou kopie často používaných položek dat z hlavní paměti
- klíčem je adresa položky
- např. tabulka stránek (nebo její část)

čtení ... zahájení cyklu čtení současně z cache i z hlavní paměti. Pokud se položka v cache nalezne, cyklus hlavní paměti se nedokončí.

V opačném případě se přečtou data z hlavní paměti (a obvykle zároveň uloží do cache)

24.5.2007

Y36SAP-Paměťový subsystém

20

zápis ...

- pokud položka v cache není zapíše se zpravidla jen do hlavní paměti
- pokud je, postupuje se dvěma způsoby
 - průběžný zápis (**write through**) ... nová hodnota se zapíše se zároveň do cache i HP
 - odložený zápis (**write back**) ... nová hodnota se zapíše jen do cache a při uvolňování položky z cache se musí její obsah přepsat do HP pokud byla modifikována po dobu přítomnosti v cache

24.5.2007

Y36SAP-Paměťový subsystém

21

Asociativní paměť s omezeným stupněm asociativity

- nevýhoda plně asociativní paměti:
 - adresář je tvořen speciálními obvody
 - při stejné kapacitě cca trojnásobná plocha čipu
- řešení ... omezený stupeň asociativity, tzn. každé položce je určeno místo (nebo několik míst podle stupně asociativity), kde se může nacházet. Toto místo je určené částí adresy položky.

24.5.2007

Y36SAP-Paměťový subsystém

22

... asociativní paměť s omezeným stupněm asociativity

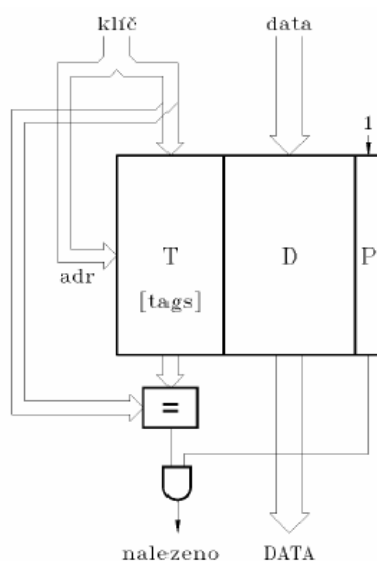
- adresář je možné realizovat běžnou pamětí RAM
- přítomnost položky se zjistí porovnání s klíčem (nebo několika klíči) uloženém v adresáři
- klíčem je část adresy
- stupeň asociativity je počet míst, na kterých se položka může nacházet
- pro zvýšení efektivity jsou data uložena po blocích (řádcích) např. 16 slabik

24.5.2007

Y36SAP-Paměťový subsystém

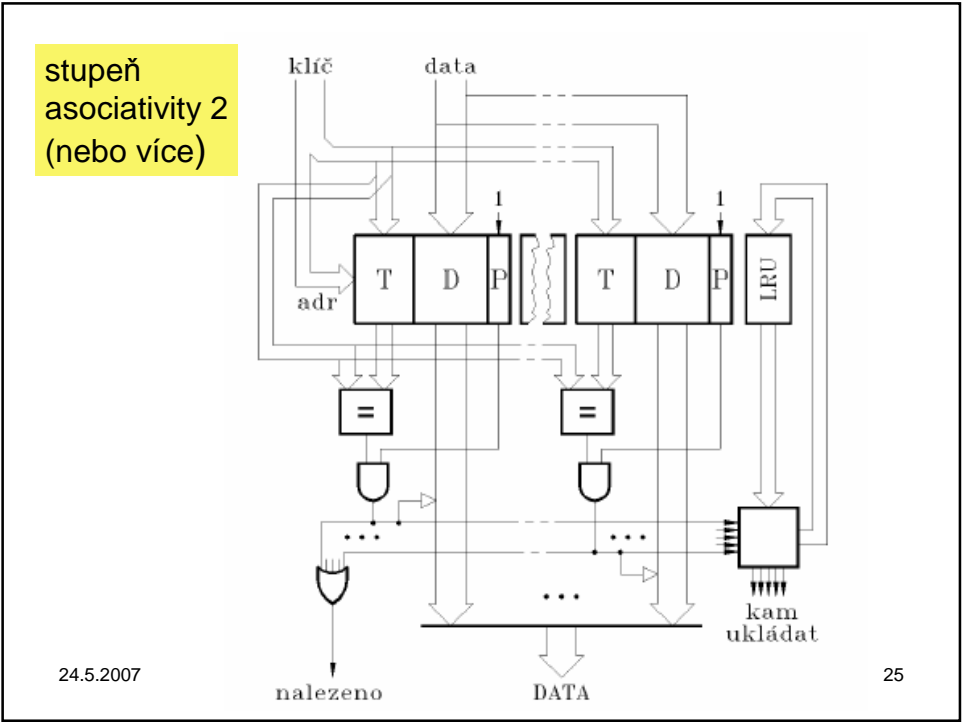
23

stupeň asociativity 1, princip



24.5.2007

24



hlavní paměť

kapacita = 64KB = 2^{16} B

16 bitová adresa

...	...
0FFF	F0 F1 F2 F3 F4 F5 F6 F7 00 01 02 03 04 05 06 07
1000	08 09 0A 0B 0C 0D 0E 0F FFF FE FD FC FE EE ED ED
...	...

cache

4 adresy
délka bloků = 4

... 2 bity

stupeň asociativity = 2
kapacita = 32B = $4 \cdot 4 \cdot 2$ B

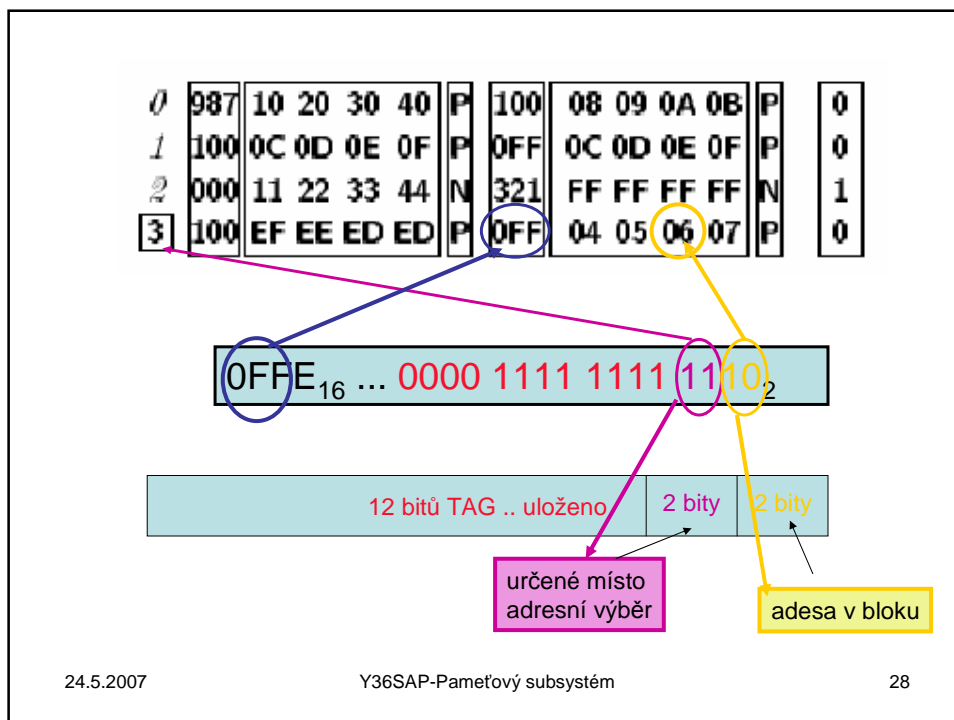
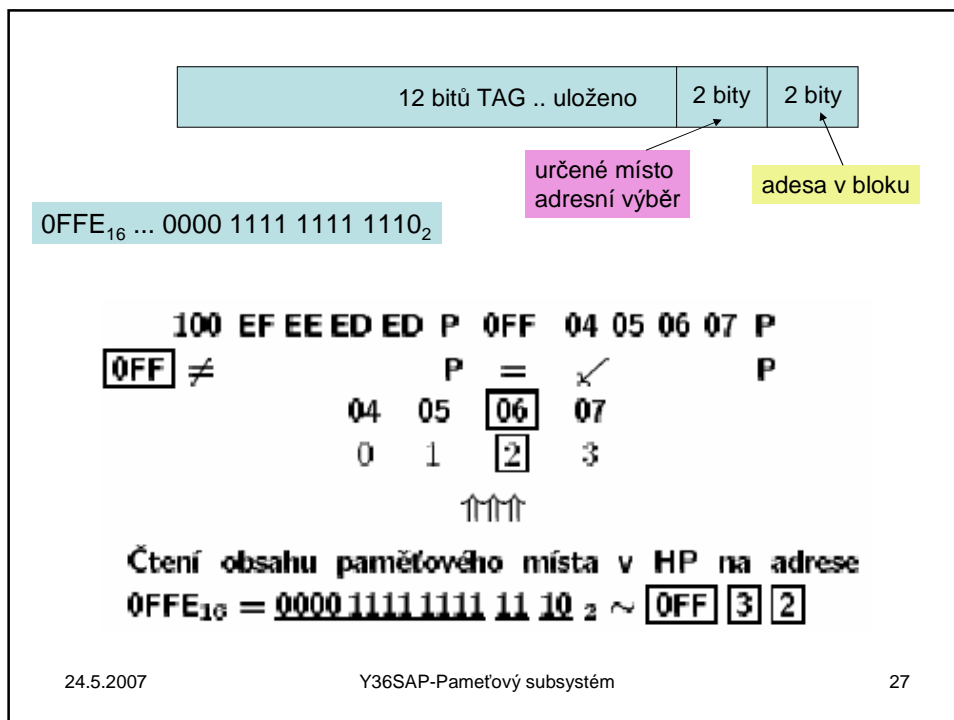
0	987	10 20 30 40	P	100	08 09 0A 0B	P	0
1	100	0C 0D 0E 0F	P	0FF	0C 0D 0E 0F	P	0
2	000	11 22 33 44	N	321	FF FF FF FF	N	1
3	100	EF EE ED ED	P	0FF	04 05 06 07	P	0

100 EF EE ED ED P 0FF 04 05 06 07 P

24.5.2007

Y36SAP-Paměťový subsystém

26



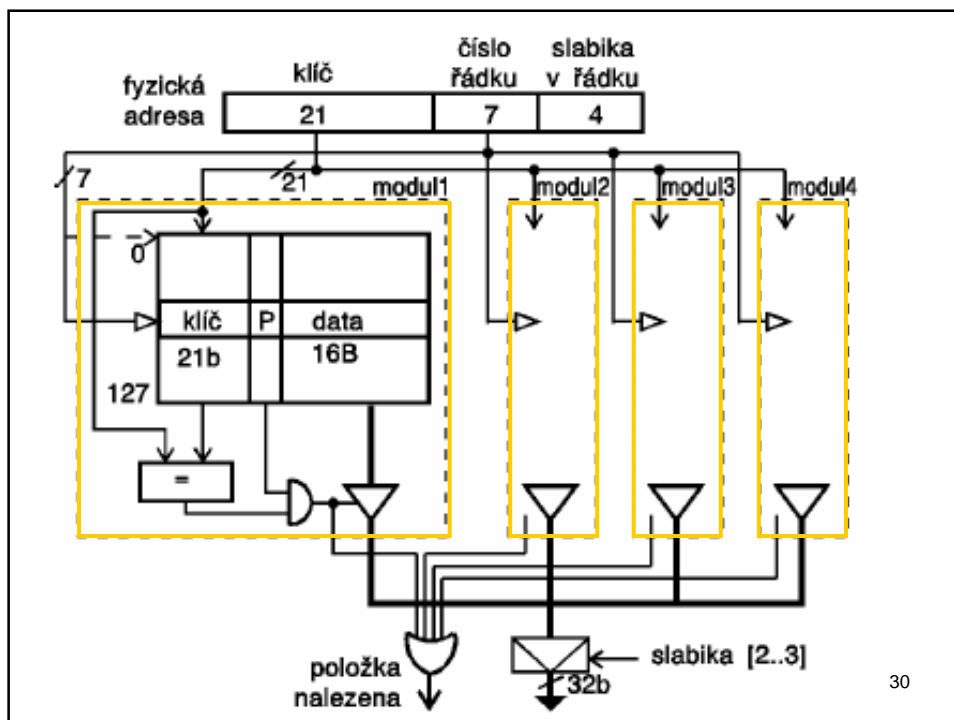
Příklad - cache v procesoru I80487

- 32 bitová fyzická adresa
- 8 KB cache
- stupeň asociativity 4
- bloky dat 16 B (slabik)
- uvolňování položek - různé strategie, např. LRU, náhodný výběr z počtu položek v řádku (stupeň asociativity 4 ... tedy ze 4)

24.5.2007

Y36SAP-Paměťový subsystém

29



30