

# **KONCEPCJA PAMIĘCI PODRĘCZNEJ (CACHE)**

- **Architektura komputera z pamięcią cache**
- **Elementy systemu pamięci cache**
- **Organizacja pamięci cache**

# KONCEPCJA PAMIĘCI PODRĘCZNEJ (CACHE)

Koncepcja ta jest podobna do koncepcji pamięci wirtualnej.

Przypomnijmy:

- pamięci statyczne są szybsze, droższe, pobierają więcej energii i są trudniejsze do scalania
- pamięci dynamiczne są wolniejsze, tańsze, pobierają mniej energii i są łatwiejsze do scalania

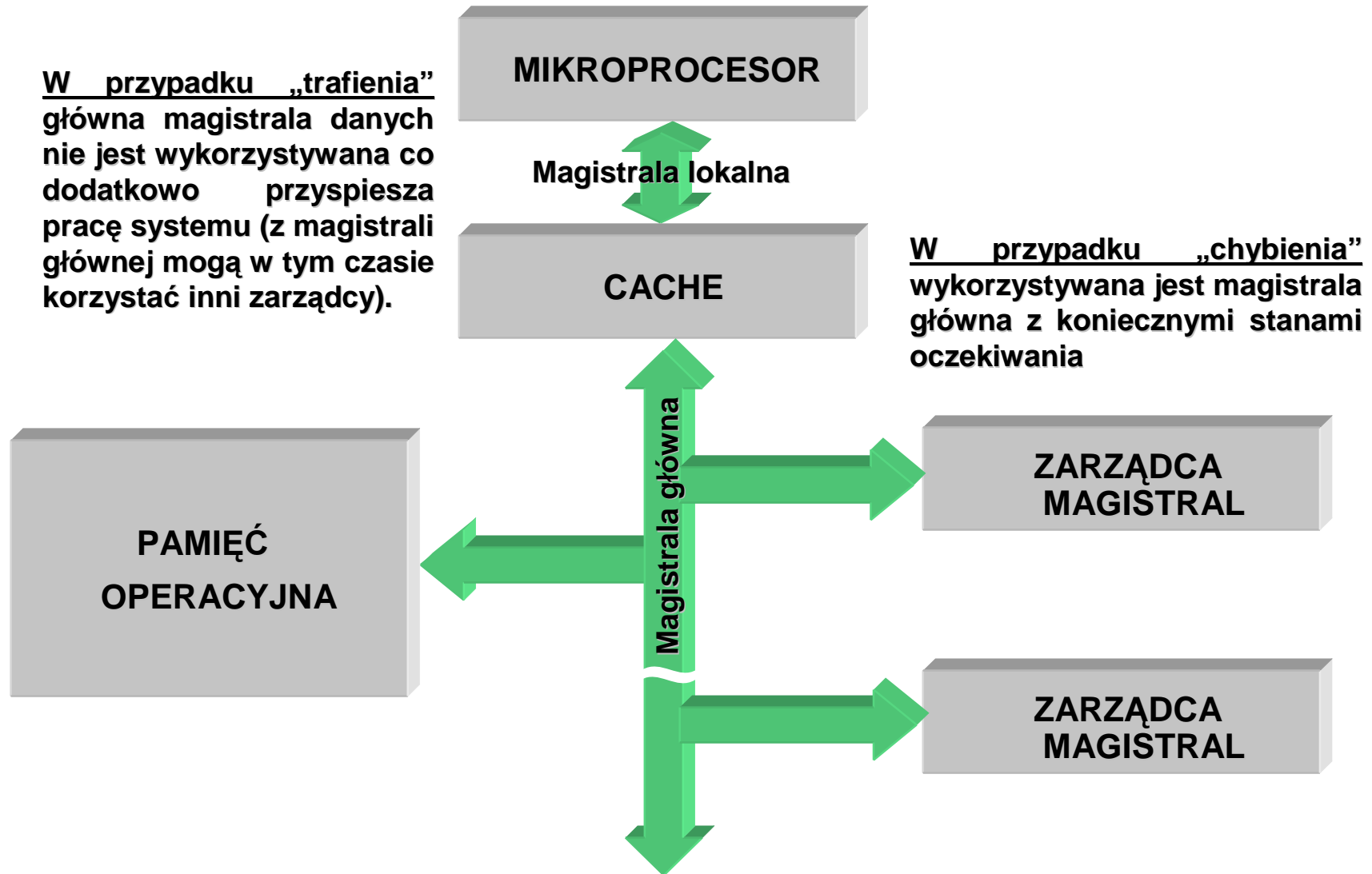
**Wniosek 1:** nie jest możliwe zbudowanie całej pamięci operacyjnej z pamięci statycznych, a pamięci dynamiczne są za wolne dla współczesnych procesorów i częstotliwości zegara

**Wniosek 2:** wprowadza się dużą pamięć operacyjną (rzędu kilka lub kilkaset MB) zbudowaną z pamięci dynamicznych i małą (kilka kilkaset kB) pamięć podręczną - statyczną - znacznie szybszą (wraz ze sterownikiem)

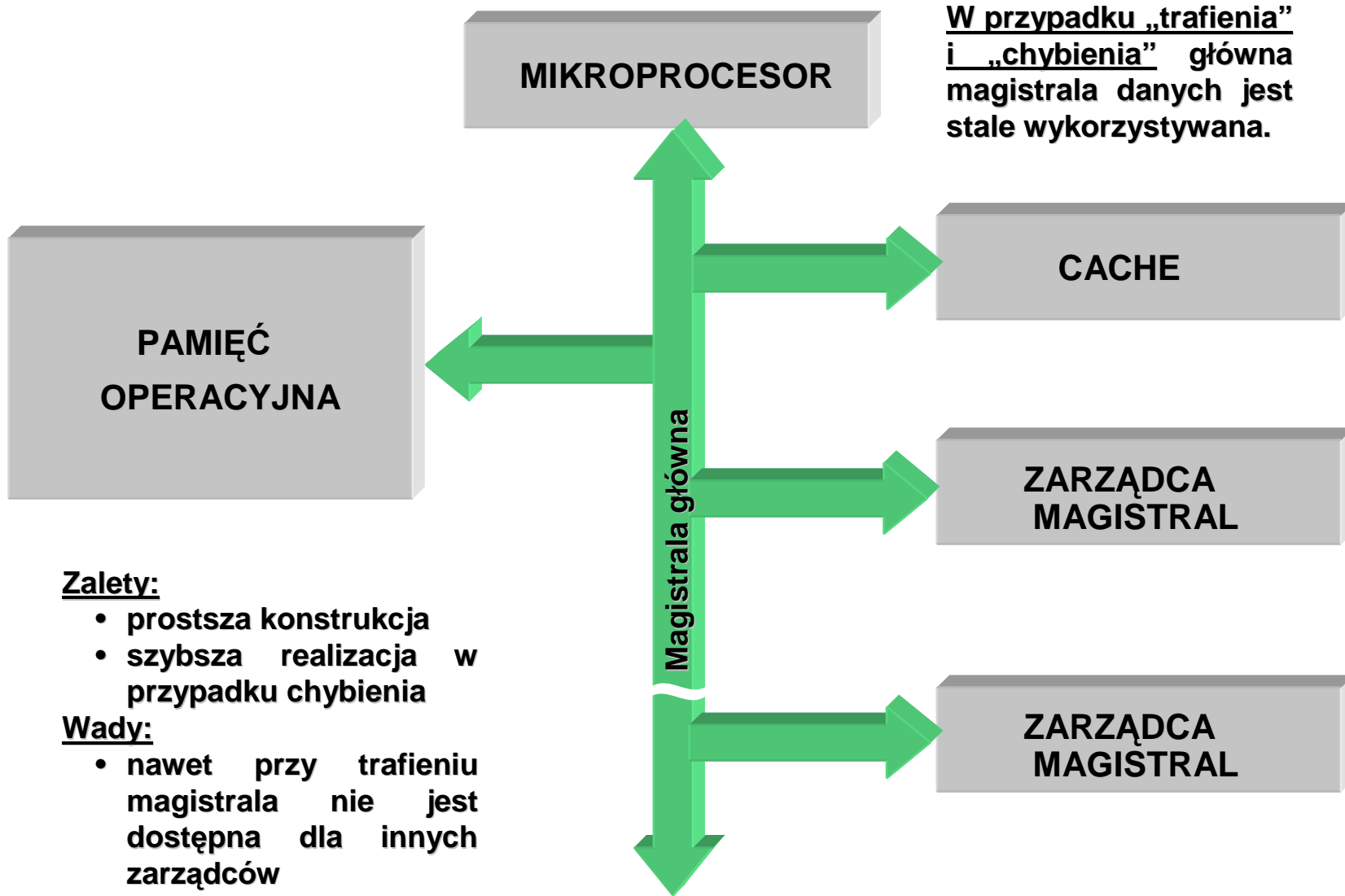
**Wniosek 3:** dwa rodzaje architektury:

- Look-trough
- Look-aside

# ARCHITEKTURA LOOK-THROUGH



# ARCHITEKTURA LOOK-ASIDE



# ELEMENTY SYSTEMU PAMIĘCI CACHE

- **Bank danych pamięci cache** (pamięć danych)  
przechowuje, umożliwia zapisywanie i odczytywanie informacji
- **Katalog pamięci cache** (zwany TAG-RAMem)  
umożliwia szybkie sprawdzanie czy poszukiwana informacja znajduje się w pamięci danych cache (konkretnie: czy dany adres jest odwzorowany w pamięci cache)
- **Sterownik pamięci cache**  
realizuje sprawdzenie adresu w katalogu pamięci cache, organizuje współpracę pamięci cache z systemem i zapewnia zgodność zawartości pamięci cache z pamięcią główną

# ZAPEWNIENIE ZGODNOŚCI PAMIĘCI CACHE

## Przyczyny niezgodności:

- nastąpił zapis do pamięci cache bez zapisu do pamięci głównej (w przypadku trafienia przy zapisie do pamięci)
- nastąpił zapis do pamięci głównej bez zapisu do pamięci cache (gdy z magistrali głównej korzystał inny zarządca magistrali niż procesor np. transmisja DMA do pamięci głównej)

## Strategie utrzymania zgodności:

- Write-through (zapis do pamięci cache powoduje jednoczesny zapis do pamięci głównej - proste ale zmniejsza szybkość, bo zapisuje się informacje do wolnej pamięci głównej)
- Buforowane (opóźnione) write-through (jak wyżej ale zapis do bufora i dopiero później z bufora do pamięci głównej aby nie spowalniać pracy)
- Write-back (zawartości obu pamięci są uzgadniane gdy zajdzie taka potrzeba - operacje bardzo skomplikowane)

Pewne obszary w pamięci głównej nie mogą być odzwierciedlone w pamięci cache, gdyż nie jest możliwe zapewnienie zgodności, np. obszar pamięci RAM karty sieciowej, gdzie zapis może realizować procesor przez magistralę systemową jak i sterownik sieci (nie ma możliwości śledzenia magistrali sieci). W systemie muszą więc istnieć układy NCA (*non-cachable access*) uniemożliwiające niektóre zapisy do pamięci cache.

# PAMIĘĆ CACHE DRUGIEGO POZIOMU

Pamięć cache (zwłaszcza typu Look-through) może występować:

- wewnątrz procesora - pamięć cache pierwszego poziomu (L1- level 1),
- pamięć zewnętrzna - pamięć drugiego poziomu (L2).

Pamięć tego typu pojawiła się po raz pierwszy w procesorach 80486.

Pamięć L2 jest znacznie większa od L1, dzięki czemu jeśli nie nastąpiło „trafienie” w pamięci L1 to istnieje duża szansa trafienia do L2 (oczywiście informacja z pamięci L1 znajduje się w pamięci L2) - przyspiesza to pracę systemu.

