

- Aby pobrać tę instrukcję, procesor musi oczywiście znać jej położenie w pamięci głównej (operacyjnej) — informacje o położeniu kolejnej instrukcji są umieszczone w specjalnym rejestrze, nazywanym *wskaźnikiem instrukcji*.
- W architekturze x86 używany jest 32-bitowy wskaźnik instrukcji oznaczony jest symbolem **EIP** (ang. extended instruction pointer).
- W architekturze x86-64 używany jest 64-bitowy rejestr RIP; w innych architekturach omawiany rejestr nazywany jest także *licznikiem rozkazów* lub *licznikiem programu* (PC – ang. program counter).



POLITECHNIKA
GDAŃSKA

Cykl rozkazowy (5)

- Czynności wykonywane przez procesor w trakcie pobierania i wykonywania poszczególnych rozkazów powtarzane są cyklicznie, a cały proces nosi nazwę *cyklu rozkazowego*.
- Cykl rozkazowy jest podstawą działania wszystkich komputerów.





Rozkazy niesterujące (2)

- W procesorach zgodnych z architekturą x86 kolejna zawartość rejestru EIP jest obliczana wg formuły:

$$\text{EIP} \leftarrow \text{EIP} + \langle \text{liczba bajtów aktualnie wykonywanego rozkazu} \rangle$$

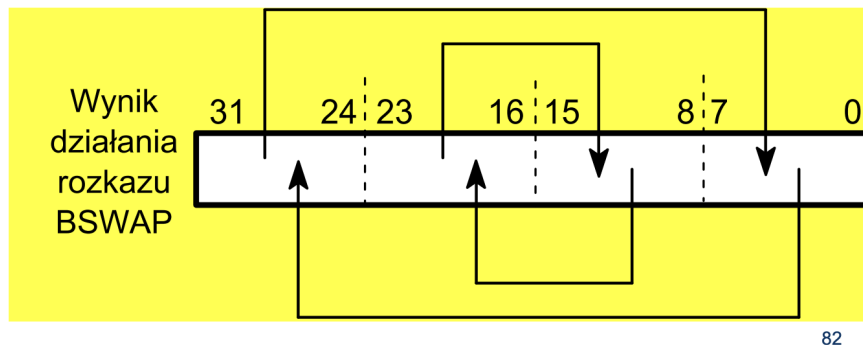


Porządek bajtów (3)

- Format *little endian* stosowany jest m.in. w procesorach rodziny x86/64 (AMD/Intel), VAX, Alpha.
- Format *big endian* stosowany jest m.in. w procesorach Motorola 680x0, SunSPARC i większości procesorów klasy RISC.
- Procesor PowerPC udostępnia oba tryby pracy — w rejestrze MSR (Machine Status Register) wprowadzono dwa bity, z których pierwszy określa stosowaną kolejność bajtów dla procesora działającego w trybie systemu operacyjnego (ang. kernel mode), drugi bit określa aktualną kolejność dla zwykłego programu.

Porządek bajtów (4)

- W procesorach rodziny x86 dostępny jest rozkaz BSWAP, który zamienia liczbę w 32-bitową w formacie mniejsze niżej (little endian) na format mniejsze wyżej (big endian) lub odwrotnie. Działanie rozkazu ilustruje poniższy rysunek.



Porządek bajtów (5)

- Analogiczne działania rozkaz BSWAP wykonuje na rejestrze 64-bitowym.
- Zamianę formatu liczb 16-bitowych wykonuje się za pomocą rozkazu XCHG, który zamienia zawartości obu operandów, np.:

xchg dl, dh

wyównywanie danych w pamięci głównej (operacyjnej)

- Dane liczbowe przechowywane w pamięci komputera mają długość 1, 2, 4, ... bajtów. W przypadku danych o rozmiarze 2, 4, ... bajtów pożądanym jest by znajdowały się one w pamięci pod adresem podzielonym przez ich długość liczoną w bajtach — takie ułożenie danych pozwala na uzyskanie najszybszego dostępu do nich.
- Jeśli adres danej jest podzielny przez jej długość (np. liczba 4-bajtowa została zapisana w pamięci pod adresem 456), to mówimy, że stosowane jest *wyrównanie naturalne*.

ASCII

Najpierw 7 bitów + 1 kontrolny
później brakowało, więc używano 8:

- Istnieje wiele kodów rozszerzonych ASCII; w Polsce najbardziej znane są:
 - Windows 1250 (Microsoft CP 1250)
 - ISO 8859–2
 - Latin 2
 - Mazovia (wyszedł z użycia)

- Wynikiem prac obu instytucji było zdefiniowanie dwóch zestawów znaków:
 - UCS — Universal Character Set (ISO 10646).
 - Unicode (konsorcjum producentów).
- Znaki obu standardów są identyczne, ale obie instytucje wydają odrębne dokumenty, a także występują inne, drobne różnice.
- W dalszej części wykładu omawiany *uniwersalny zestaw znaków* określać będziemy terminem **Unikod** (lub Unicode).
- W tej sytuacji okazało się konieczne wprowadzenie kodowania 16-bitowego — wówczas wyłonił się standard znany jako Unicode
- W odniesieniu do znaków z podstawowego kodu ASCII, w Unikodzie rozszerzono ich kody binarne z 8 do 16 bitów poprzez „dopisanie” 8 zer z lewej strony

- W systemie Unicode każdemu znakowi przypisana jest wartość liczbową określaną jako *punkt kodowy* (ang. code point), przy czym dodatkowo każdemu znakowi przyporządkowana jest także nazwa, nie jest natomiast określony kształt drukowanego znaku.
- Przykładowo:
 - wielka litera **A** ma przypisany kod liczbowy, który zapisywany jest w postaci U+0041 (szesnastkowo), a oficjalna nazwa brzmi „LATIN CAPITAL LETTER A”.
 - litera **ą** ma przypisany kod U+0105, a oficjalna nazwa brzmi „LATIN SMALL LETTER A WITH OGONEK”.

- Punkty kodowe Unikodu zapisywane są w postaci liczb złożonych z 4, 5 lub 6 cyfr w zapisie szesnastkowym.
- Zazwyczaj stosowany jest zapis, w którym wartość liczbową poprzedzona jest znakami U+, co należy traktować jako informację, że jest to wartość punktu kodowego podana w zapisie szesnastkowym.
- Kody kilku początkowych liter alfabetu polskiego zawiera tabela (wartości podane są w kodzie szesnastkowym).

Znak	Kod
a	0061
A	0041
ą	0105
Ą	0104
b	0062
B	0042
c	0063
C	0043
ć	0107
Ć	0106

- Większość powszechnie używanych znaków jest przyporządkowana punktom kodowym o wartościach nie przekraczających 65 535 — zbiór ten, obejmujący kody od 0 do 65535, oznaczany jest skrótem BMP (ang. Basic Multilingual Plane). Zatem wartości punktów kodowych ze zbioru BMP dają się przedstawić w postaci 16-bitowych liczb binarnych.

- Z podanych powodów wprowadzono bardziej efektywne sposoby przechowywania znaków Unikodu w pamięci komputera — najczęściej używane są formaty UTF-8 i UTF-16 (Unicode Transformation Format).