

Pamięć operacyjna K-202 i Mery-400

Andrzej Ziemkiewicz

Ostatnio można było zobaczyć w Internecie wiele dyskusji dotyczących komputera K-202, a w szczególności jego pamięci operacyjnej. Różni ludzie stawiali pytanie czy rzeczywiście pamięć mogła mieć 4MB i jak była zorganizowana.

Znalazłem nawet blog, w którym autor twierdził, że w ogóle to jest jakaś machlojka i oszustwo, bo gdyby nawet sprasować te wszystkie kostki pamięci (układy scalone) za pomocą walca drogowego, to nie mogłyby się zmieścić w obudowie komputera. Pisał też, że on tu zaraz przeprowadzi dochodzenie, które wykryje prawdę i odsłoni cały szwindel.

Wypowiedź taka świadczyła głównie o braku wiedzy i wyobraźni autora. O braku wiedzy, bo nie wiedział, że w tym czasie (rok 1970) kostki pamięci jeszcze nie były dostępne i pamięć operacyjna była budowana na rdzeniach ferrytowych. A o braku wyobraźni, ponieważ uważał że cała pamięć musi mieścić się w obudowie procesora, a nie potrafił sobie wyobrazić, że można umieścić pamięć w zewnętrznych modułach, dołączonych do procesora kablem.

Całe to dochodzenie skończyło się, gdy ktoś mu odpisał, że ma jeszcze w swoim laboratorium zachowany egzemplarz maszyny K-202 z kilkoma dodatkowymi blokami pamięci i zapraszał do obejrzenia. W tym momencie autor blogu zamilkł i okazało się, że nie chodziło wcale o prawdę historyczną, tylko o sensację. A jak nie ma sensacji, to nie ma o czym mówić.

Ponieważ wzbudza to zainteresowanie, więc wydaje się celowe napisać jak się rzecz miała.

1 Pamięć operacyjna (RAM) – uwagi ogólne

Parę słów o ówczesnym stanie teorii i techniki pamięci komputerowej. Większość obecnych na rynku komputerów była jednoprogramowa, z pamięcią adresowaną linearnie. Pojęcia dotyczące architektury pamięci, takie jak 'segmentacja' lub 'stronicowanie' nie były zbyt rozpowszechnione. Nie mniej istniały już i warte są wspomnienia. W 1960 roku firma Borroughs rozpoczęła projekt komputera B5000, a w 1964 roku jego ulepszonej wersji B5500. Była to maszyna o niezwykle pięknej architekturze. Firma ambitnie starała się wstawić do niej wszystkie najnowsze osiągnięcia teoretyczne. Tak więc pamięć RAM była segmentowana. Segment była to część pamięci udostępniająca pewną przestrzeń adresową procesowi. Proces widział tę przestrzeń adresową jako swoją pamięć, a nie widział segmentów przypisanych do innych procesów. Pamięć posiadała też mechanizm stronicowania, tj. małe porcje pamięci zwane stronami (zwykle o pojemności 4K) mogły być dodawane do wskazanego segmentu (lub usuwane). Oba mechanizmy (segmentacji i stronicowania) były dynamiczne, tj. były używane w razie potrzeby w czasie działania procesu. W dodatku wszystko to było realizowane całkowicie sprzętowo. Ogromne osiągnięcie teoretyczne i techniczne !!! Po zakończeniu projektu została wydana książka „Borroughs B5000/B5500”, w której mechanizmy te zostały dokładnie opisane. Książka ta była dostępna w bibliotece IMM (gdzie przeczytałem ją z zapartym tchem). Prawdopodobnie można ją gdzieś jeszcze znaleźć w Internecie.

Dużo informacji o maszynach B5000/B5500 można znaleźć na stronie internetowej w Wikipedia. Autor strony nie pisał jednak o mechanizmach segmentacji i stronicowania, a zachwycał się raczej mechanizmami *meta-bits* (*dirty-bit* i in.), *tags* oraz *descriptors* (przewidziane do wykonywania programów napisanych w językach zorientowanych obiektowo). Bardzo ciekawa lektura. Niektóre z tych pomysłów tak wyprzedzały swoją epokę, że nie ma ich nawet w obecnie budowanych, nowoczesnych komputerach.

Drugim projektem który podjął idee segmentacji i stronicowania pamięci był Multics, wspólne przedsięwzięcie MIT, Bel Labs i General Electric. Był to projekt całości systemu komputerowego, obejmującego system operacyjny i specjalnie zbudowany sprzęt. Pierwszy opis projektu był opublikowany w 1963 roku. Mechanizm segmentacji i stronicowania był realizowany sprzętowo. Prace trwały wiele lat i sprzęt okazał się zbyt skomplikowany, aby uruchomić i konserwować go. W końcu projekt został zaniechany w 1968/69 roku.

Tak więc na przełomie lat 1960/70 wiedza teoretyczna była już dostatecznie zaawansowana i była nam znana, praktyka produkcyjna zaś pozostawała w tyle.

2 Realizacja fizyczna pamięci K-202

Pamięć RAM była zbudowana na rdzeniach ferrytowych o średnicy 0.5 mm, co dawało czas cyklu 1 μ sec. Dla porównania, pamięci budowane do tego czasu, np. w maszynach ZAM, miały rdzenie o średnicy 2 mm, co dawało czas cyklu 10 μ sec, tak więc postęp był tu bardzo znaczny.

Gotowe ramki z rdzeniami były sprowadzane z firmy *Data Product* (Irlandia) i montowane na płytkach drukowanych. Zespołem realizującym pamięć kierował mgr inż. Jerzy Cewe. Aby uzyskać więcej szczegółów należałoby go odnaleźć (może w Internecie).

3 Architektura pamięci K-202

Pamięć była zorganizowana w bloki o maksymalnej pojemności 64 K słów 16-bitowych każdy. Adres miejsca pamięci był konkatencją numeru bloku i adresu wewnątrz bloku. Numer bloku był ustalany dla każdego fizycznego bloku w momencie montażu maszyny dla klienta (za pomocą zworek zakładanych na styki w płycie drukowanej). Adresy wewnątrz bloku były ciągłe, od 0 do maksymalnego adresu w bloku. Tak więc adresacja pamięci była w zasadzie fizyczna. Każdy blok pamięci był dołączany do szyny pamięciowej komputera za pomocą dwóch elementów: jednego odbiornika i jednego nadajnika.

Tak więc w świetle wyżej wspomnianej terminologii pamięć K-202 nie posiadała stronicowania. Ani pamięć fizyczna nie miała odpowiedniego mechanizmu, ani lista rozkazów nie miała rozkazów koniecznych do zarządzania takowym (Żadnego *memory-mapping* nie było). Pamięć była natomiast segmentowana, ponieważ używana przez nas nazwa 'blok' odpowiadała segmentowi. Segmentacja ta nie była jednak dynamiczna, tylko statyczna, ustalona przez montaż, fizycznego bloku.

Nie było to rozwiązanie odpowiadające szczytowym osiągnięciom teoretycznym, było jednak i tak lepsze niż przeciętne. Poza tym było przez nas traktowane jako tymczasowa pierwsza wersja.

Trzeba tu pamiętać, że w projektowaniu i konstrukcji K-202 panował pewien pośpiech. Jako polsko-angielskie przedsięwzięcie maszyna musiała być konkurencyjna na rynkach zachodnich. Konieczne było przedstawienie jej na najbliższej wystawie Olympia w Londynie, inaczej inne firmy zajęłyby rynek. Różne udoskonalenia odkładane były do następnej wersji.

Prace nad drugą wersją K-202, zawierającą dynamiczną segmentację i stronicowanie pamięci oraz inne ulepszenia były już rozpoczęte, ale przedsięwzięcie zostało utrupione.

4 Dyskusja problemów

Podstawowe pytanie wszystkich dyskutantów było czy rzeczywiście maszyna mogła mieć 4MB pamięci. Nie wiem dlaczego wszyscy oni tak się uwzięli na ten parametr. Jest to tylko teoretyczne maksimum przestrzeni adresowej, wynikające z ilości bitów adresu. Istnieje dużo innych czynników bardziej ograniczających pojemność pamięci, z których kilka wymienię poniżej:

1. Wydajność prądowa nadajników

Każdy moduł dołączony do szyny pamięciowej (procesor, kanał pamięciowy lub blok pamięci)

był dołączony za pomocą jednego odbiornika (pobierającego 1 umowną jednostkę prądową) oraz jednego nadajnika (dostarczającego 30 umownych jednostek prądowych). Wynika stąd, że nie można było mieć więcej niż 30 dołączonych modułów. Oczywiście że gdyby zaszła potrzeba, można byłoby zdwoić nadajniki łącząc je kolektorami i uzyskać 60 jednostek prądowych na wyjściu, ale nie było takiego zamówienia.

2. **Długość szyny pamięciowej**

Moduły K-202 były montowane w obudowach o standardowej szerokości 19". Tak więc każdy dołączony moduł wydłużał szynę (wiązkę kabla interfejsu) o ok. 0.5 metra. Przy wielu dołączonych modułach szyna miała więc wiele metrów. Przy szybkości używanych sygnałów była to już *linia długa*, gdzie dawały się zauważyć zjawiska mikrofalowe, takie jak odbicia sygnału od końca linii, dopasowanie falowe (lub niedopasowanie) itp. Można było poprawić nieco sytuację montując moduły nie jeden obok drugiego, ale jeden nad drugim w 19-calowym stojaku (wysokość modułu była o połowę mniejsza niż szerokość), ale nie zmieniało to w zasadniczy sposób problemu.

3. **Przesłuchy między-bitowe**

W wiązce przewodów biegnących obok siebie na dużej długości sygnał w jednym przewodzie indukuje zakłócenia w sąsiednich przewodach. Zjawisko to daje się odczuć tym bardziej im dłuższa linia. Sygnały w szynie były przesyłane parami skręconymi, jeden przewód sygnałowy i jeden przewód masy. Przewód masy oprócz tego że dostarczał poziomu odniesienia, pełnił też rolę ekranu. Badania laboratoryjne pozwoliły szybko opracować instrukcję dla montażystów, określającą minimalną ilość skrętów na metr bieżący kabla, tak aby to ekranowanie było dostatecznie skuteczne. Przy naprawdę dużej ilości bloków (a więc długiej szynie) prawdopodobnie trzeba byłoby zastąpić pary skręcone przez kabelki koncentryczne, ale tu również nie było takich zamówień.

4. **Problem pojemności**

Każdy odbiornik i nadajnik posiadał pewną pojemność elektryczną. Wraz z opornością umieszczaną na końcu linii tworzyło to układ RC, w którym odpowiedzią na wysłany sygnał (teoretycznie skok jednostkowy) jest przebieg eksponencjalny, którego narastające zbocze jest tym bardziej położone im większa pojemność. Zjawisko to nie przedstawiało większego problemu, dzięki temu że w maszynie wszędzie było sterowanie asynchroniczne (żadnego centralnego zegara). Wystarczało odpowiednio wydłużyć czas oczekiwania na ustalenie się sygnałów na szynie. Jediną konsekwencją dużej ilości bloków było więc niewielkie zmniejszenie szybkości działania.

5. **Problem zasilania**

Tu sprawa wydaje się prosta. Maszyna posiada zasilacz który dostarcza do układów scalonych napięcia o standardowej wartości 5V i już. Problemy pojawiają się jednak w dynamicznym działaniu. Wyobraźmy sobie początek cyklu pracy pamięci dla zapisu. Procesor wysyła na szynę pamięciową adres i dane do zapisania, w sumie kilkadziesiąt bitów, każdy z nich mnożony przez 30 jednostek prądowych potrzebnych do wysterowania nadajników. Oznacza to, że wymagamy od zasilacza dużego chwilowego 'szarpnięcia prądowego' co wcale nie jest tym samym co wartość średnia prądu, jaki może on dostarczyć. Zasadnicze znaczenie ma tu oporność wewnętrzna źródła prądowego, do której można zaliczyć także oporność ścieżek zasilających na płytkach drukowanych, oporność styków w łączówkach i oporność wewnętrzną kondensatorów elektrolitycznych w filtrze wyjściowym zasilacza. Rozwiązanie polegało na odpowiednim zaprojektowaniu ścieżek zasilających tak aby były możliwie jak najszersze, na zdublowaniu styków zasilających na łączówkach i na dołączeniu (równoległe do kondensatorów elektrolitycznych) dodatkowych kondensatorów o mniejszej pojemności, ale za to o wysokiej jakości i pomijalnej oporności wewnętrznej.

Tak więc na pytanie czy maszyna K-202 mogła mieć 4MB pamięci odpowiedź jest: nie mogła z różnych względów technicznych, ale też wcale nie musiała.

5 Pamięć operacyjna Mery-400

Komputer Mera-400 był zaprojektowany na przełomie lat 1973/74. W porównaniu z K-202 była to już następna epoka, układy scalone RAM były już dostępne... Aby nie rozwódzić się nad szczegółami piszę tylko co było nowe lub różne od K-202.

Pamięć Mery-400 była już w pełni segmentowana i stronicowana dynamicznie. Fizycznie była ona zbudowana jako zbiór stron o pojemności 4K słów. Każda strona była wyposażona w rejestr zawierający numer segmentu (bloku) i *tag* strony (najstarsze bity adresu wewnątrz segmentu).

System operacyjny miał prawo (za pomocą specjalnych rozkazów) wpisać zawartość do dowolnego z tych rejestrów. W ten sposób fizycznej stronie pamięci był przypisywany adres logiczny. Dopiero od tego momentu strona pamięci reagowała na sygnały odczytu/zapisu według adresu logicznego.

W ten sposób system operacyjny mógł dynamicznie budować segment pamięci przypisany do procesu. Adresacja w tym segmencie nie musiała nawet być ciągła. Dla systemu operacyjnego mechanizm ten był wystarczający, aby zarządzać pamięcią w trybie wieloprogramowym.