

NOWE TRENDY ROZWOJU MIKROPROCESORÓW

Marcin LORENC, Krzysztof CEGIELSKI

Streszczenie: Celem artykułu jest zaprezentowanie kierunków rozwoju mikrokontrolerów. Przedstawiono krótką historię procesorów, pojęcie GPU oraz prawo Moore'a. Omówiono techniczne ograniczenia wraz ze sposobami ich rozwiązania, które są innowacyjnym elementem wprowadzającym nowy schemat programowania układów cyfrowych.

Słowa kluczowe: procesor, GPU, prawo Moore'a, Intel, SCC.

1. Historia rozwoju procesorów

Kilkadziesiąt lat wstecz, w pierwszej połowie dwudziestego wieku, pojęcie komputer kojarzyło się z pomieszczeniem, w którym znajduje się duża liczba sprzętu wykonująca automatycznie zadane operacje. W latach siedemdziesiątych XX wieku, a dokładniej w 1971 r. firma Intel rozpoczęła seryjną produkcję pierwszego na świecie komercyjnego mikrokontrolera o kodowej nazwie 4004 wykonanego w technologii 10 mikronów [1,2]. Wydarzenie to zapoczątkowało dynamiczny rozwój jednostek obliczeniowych pod względem wydajności jak również mające na celu dopracowanie/zmniejszenie procesu technologicznego produkcji układów. Kolejnym ważnym etapem było wprowadzenie pierwszego procesora 16-bitowego wykonanego w technologii 3 mikronów w czerwcu 1978 r. [1,2]. Jednostki te były jednorodniowe, a proces wykonywania przez nie operacji był sekwencyjny.

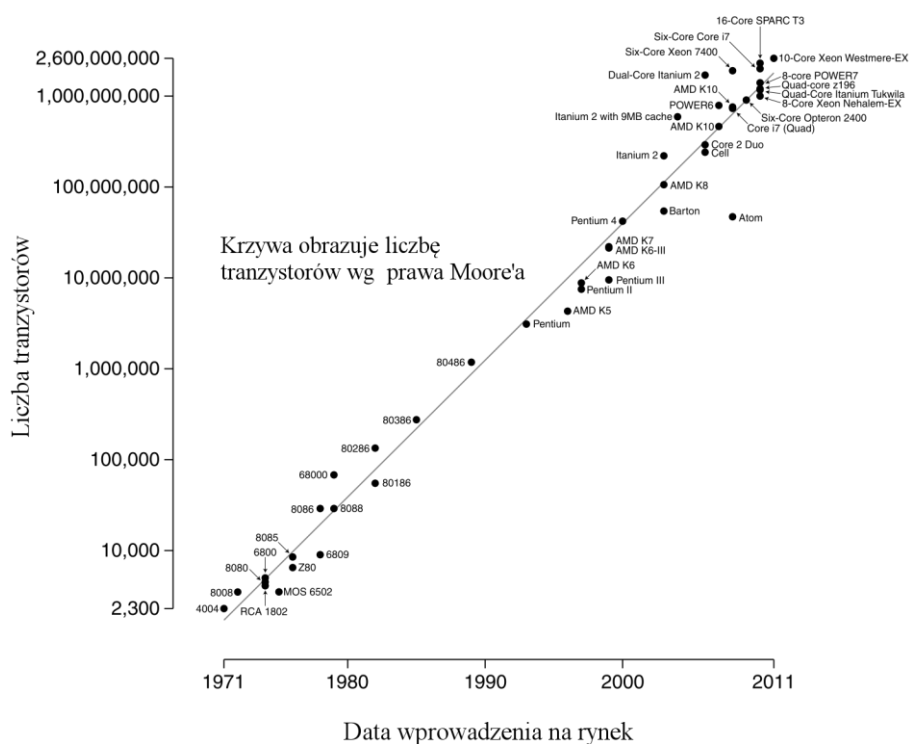
Producenci mikroprocesorów zmierzali w jednym kierunku - zwiększenia wydajności. W przeszłości uzyskiwanie lepszych osiągnięć było stosunkowo proste. Zwiększano ilość pamięci operacyjnej oraz podnoszono zegar procesora. Dziś zakup najnowszej jednostki nie gwarantuje wzrostu wydajności dla konkretnej aplikacji. Można to osiągnąć za pomocą innych, nowych technologii do których zalicza się: wielowątkowość współbieżną (ang. simultaneous multithreading) oraz struktury wielordzeniowe. Korzystanie z nich wymaga zrozumienia w jaki sposób technologie te wpływają na sposób pracy i jak będą one wchodzić w interakcje z aplikacjami w przyszłości.

1.1. Prawo Moore'a

W 1965 r. jeden z założycieli firmy Intel, Gordon Moore zaobserwował że ekonomicznie optymalna liczba tranzystorów instalowana w procesorze w kolejnych latach ma trend wykładniczy (rys.1), co w konsekwencji oznacza ich podwojenie co 24 miesiące [3,4]. Prawo to na dzień dzisiejszy jest nadal aktualne.

Dotychczas podstawowym wyznacznikiem szybkości pracy aplikacji jest liczba stanowiąca stosunek pomiędzy prędkością taktowania zegara i wydajnością procesora. Czym wartość ta jest wyższa tym aplikacja będzie szybciej pracowała. W rezultacie twórcy oprogramowania zakładali kontynuację takiego kierunku rozwoju wydajności procesorów. Tendencje związane ze zwiększeniem gęstości tranzystorów nadal jest zachowana, jednak taktowanie zegara zaczęło systematycznie zwalniać od 2003 r. W tym czasie wynosiło ono

3GHz [1]. Zgodnie z prawem Moore'a obecnie powinny być dostępne procesory o taktowaniu 10GHz[3,4]. Jednak aktualnie najszybszy fabryczny procesor, który jest dostępny w seryjnej produkcji, taktowany jest zegarem o częstotliwości 3,8GHz [1].



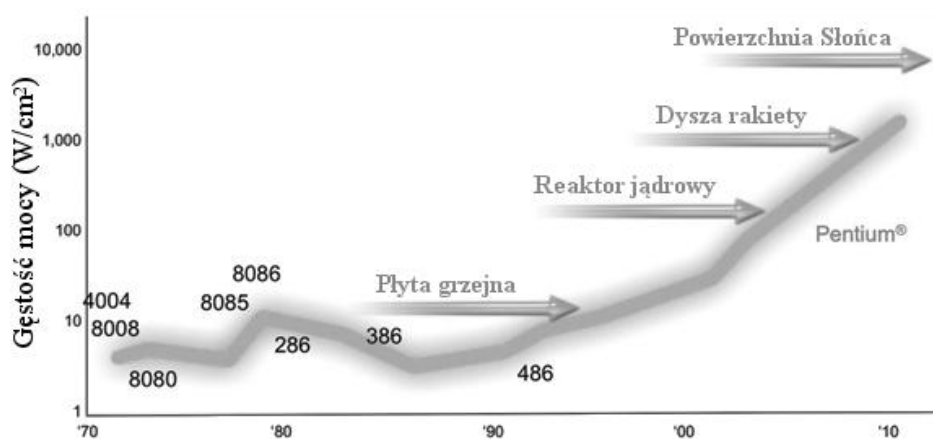
Rys. 1. Liczba tranzystorów w mikroprocesorach w latach 1971-2011 [3]

Osiągnięcie częstotliwości taktowania zegara na poziomie 10GHz nie jest możliwe z prostego powodu. Dzisiejsze procesory są bardzo energochłonne a tym samym wydzielają duże ilości ciepła. Przykładowo procesor o taktowaniu 3,8GHz w czasie pracy ma zapotrzebowanie mocy na poziomie 100W [5]. W 2004 r., Pat Gelsinger, dyrektor ds. technologii w firmie Intel, stwierdził, że ciepło emitowane z nowoczesnych procesorów, mierzone gęstością mocy wyrażoną w W/m^2 , jest porównywalne z ciepłem emitowanym przez rdzeń reaktora jądowego [6]. W dniu dzisiejszym ilość ciepła przypadająca na jednostkę powierzchni wydzielana przez procesor zbliża się do wartości występujących na powierzchni Słońca (rys.2) co stanowi poważny problem wymagający opracowania nowych rozwiązań i technologii.

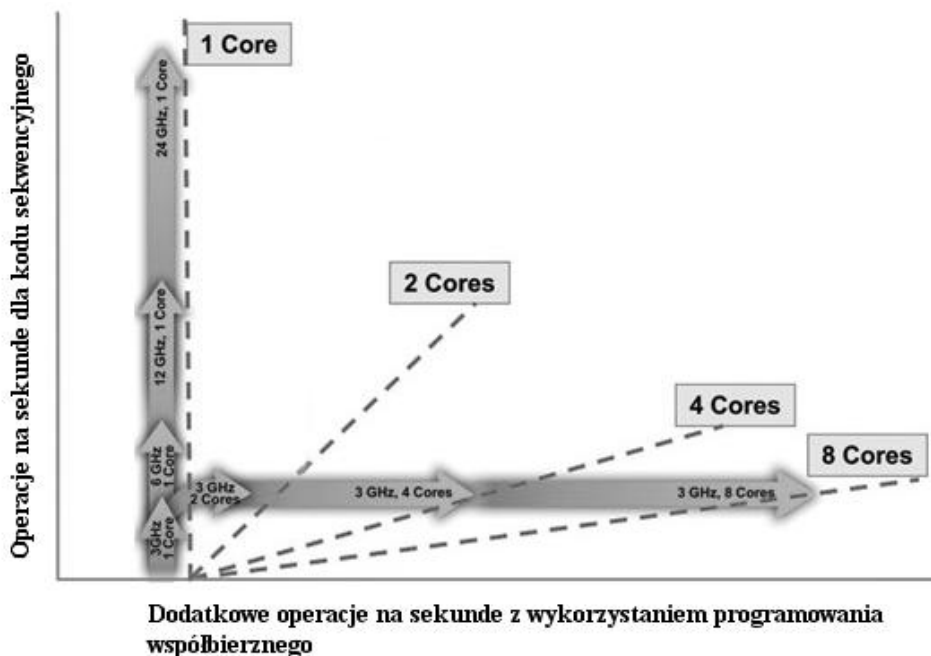
1.2. Wielordzeniowe układy

Prace inżynierów w walce z ograniczeniem wydzielanego ciepła przynoszą efekty w postaci nowej generacji procesorów, pobierających mniejszą ilość energii, oferujących jednocześnie zwiększoną wydajność dzięki zastosowaniu dwóch, czterech i większej ilości

rdzeni. Zakładając, że producenci mikroprocesorów mogą podwoić liczbę rdzeni w 24 miesiące to stosując prawo Moore'a na zasadzie analogi, w 2018 r. można spodziewać się 128-rdzeniowych jednostek. Większa liczba rdzeni w tym przypadku nie musi oznaczać wyższej wydajności pojedynczej aplikacji, gdyż na ogół układy wielordzeniowe taktowane są niższym zegarem niż procesory z pojedynczym rdzeniem. Z punktu widzenia użytkowników wielordzeniowość ma radykalnie poprawić ich interakcję z komputerem co będzie miało swoje przełożenie np. na szybsze przełączenie pomiędzy aplikacjami czy też na działanie aplikacji multimedialnych, dostarczając treści w wysokiej rozdzielczości.



Rys. 2. Ciepło wydzielane przez procesory na przestrzeni 40 lat [6]



Rys. 3. Wydajność aplikacji wielordzeniowych [6]

Niektóre programy do edycji zdjęć i obrazów działają zauważalnie szybciej na wielordzeniowych jednostkach. Obecnie istnieje na rynku duża liczba aplikacji, które nie wykorzystują możliwości wielordzeniowych układów, co w konsekwencji powoduje, że technologia ta nie daje użytkownikom żadnych wymiernych korzyści.

Na wykorzystanie wielu rdzeni CPU (ang. Central Processing Unit) pozwala oprogramowanie posiadające możliwość jednoczesnej obsługi wielu wątków lub zadań bądź też inteligentnego ich podziału pomiędzy rdzenie aby zapewnić równowagę między wydajnością a interaktywnością. Szybkość działania aplikacji napisanej z obsługą wielu rdzeni będzie taka sama na procesorze jednorodzeniowym o taktowaniu zegara dwukrotnie wyższym niż dwurdzeniowej jednostki (rys.3). W związku z tym, firmy produkujące zaawansowane oprogramowanie zatrudniają programistów specjalizujących się w programowaniu równoległym oraz wielowątkowym.

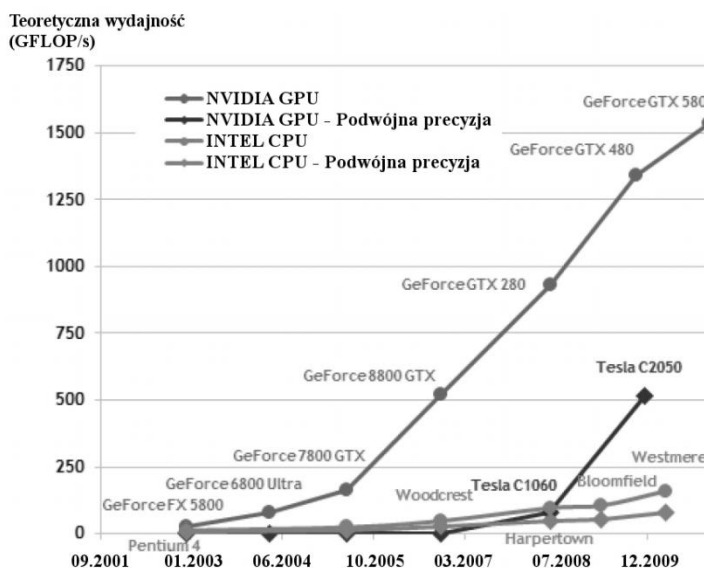
2. Trendy rozwoju

Naukowcy i inżynierowie wielu firm związanych z produkcją mikrokontrolerów intensywnie pracują nad rozwojem układów wielordzeniowych. Opracowywane technologie opierają się na założeniu, że wielordzeniowe procesory przyszłości będą wykorzystywać dziesiątki lub setki bardzo prostych, pracujących równolegle rdzeni. Wśród tych konstrukcji obecnie wyraźnie zarysowują się dwie tendencje bazujące na doświadczeniach zdobytych przy konstrukcji układów graficznych oraz z budowy klastrów komputerowych.

2.1. Wykorzystanie mocy obliczeniowej kart graficznych

Możliwości kart graficznych są bardzo duże jednak w dalszym ciągu pozostają nie wykorzystane.

Wydajność GPU (ang. Graphics Processing Unit) zwiększa się sukcesywnie. Aktualnie dostępna na rynku karta graficzna firmy NVIDIA, model GeForce GTX580, potrafi wykonać 1581 miliardów operacji zmiennoprzecinkowych na sekundę [6]. Dla przykładu najmocniejszy mikroprocesor firmy Intel, Core i7 980XE, wykona w tej samej jednostce czasu 218



Rys. 4. Porównanie wydajności mikroukładów [7]

miliardów operacji tego samego typu [6] (rys.4).

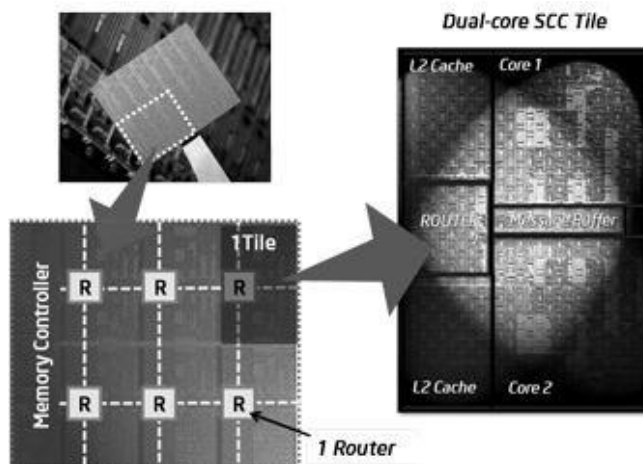
Podstawową zaletą GPU jest wydajność oraz bardzo duża przepustowość danych. Prace nad rozwojem tej technologii oraz zmiany w technikach programowania dają efekt w postaci możliwości częściowego przekierowania obciążenia CPU na GPU. W rezultacie możemy zaobserwować wzrost wydajności. Jest to trudne zadanie i wymaga stworzenia nowego schematu selekcji zadań, które będą mogły być przetworzone przez GPU.

Dostęp do GPU przez typowe dla tradycyjnych jednostek (GPGPU) zadania ma swoje ograniczenia. Jednostki te są przeznaczone do przetwarzania równoległego z wykorzystaniem technik programowania o nazwie przetwarzania strumieniowego. Technika ta umożliwia przesyłanie danych równoległe, co radykalnie zwiększa przepustowość danych i w efekcie zwiększa łączną prędkość obliczeniową. Dlatego 600Mhz GPU przewyższa szybkością 3GHz CPU [6]. Przetwarza masowo równoległe operacje w taki sam sposób jak ludzki mózg, w przeciwieństwie do sekwencyjnej pracy CPU. W efekcie programiści podejmują decyzje, które operacje są równoległe, a które GPU oblicza na pożądanym poziomie precyzji.

W chwili obecnej moc obliczeniowa GPU wykorzystywana jest w wielu dziedzinach. Począwszy od wizualizacji aerodynamicznych samolotów, samochodów lub pociągów, poprzez wyliczanie przepływów płynów, dyspersji cząsteczek, efektów kolizji, oraz innych zjawisk fizycznych w oparciu o dane wejściowe. Wszystkie tego typu operacje mogą być wykonywane w czasie rzeczywistym.

2.2. Miniaturowe klastry komputerowe

Odmienne podejście w kwestii rozwoju wydajności układów mają naukowcy i inżynierowie pracujący nad projektem Intel SCC (ang. Single-Chip Cloud Computer). Długofalowym celem jest dodanie zwiększonych możliwości skalowania do przyszłych komputerów co umożliwi opracowanie nowych aplikacji oraz interfejsów człowiek-maszyna. Pierwszym efektem tych prac był prototypowy 48-rdzeniowy procesor zademonstrowany w 2009 r. [8]. Można stwierdzić że był to zmminiaturyzowany klaster komputerowy w obudowie pojedynczego układu scalonego. Składał się on z 24 dwurdzeniowych procesorów Pentium, gdzie każdy z procesorów tworzył jeden węzeł klastra nazywanego przez firmę Intel kafelkiem. Posiadał on własną pamięć podręczną. W kafelku oprócz procesora i pamięci znajdował się również router zapewniający komunikację z pozostałymi dwurdzeniowymi kafelkami, które połączono ze sobą w większe grupy nazwane bankami (rys. 5). W przypadku omawianego układu, banków jest sześć – każdy składa się



Rys. 5. Budowa Single-Chip Cloud Computer [8]

z czterech kafelków. Połączenie wszystkich banków zostało nazwane przez firmę Intel siatką, w której każdy bank jest połączony bezpośrednio z innym bankiem. Rozwiązanie takie zwiększa przepustowość i niezawodność systemu, oraz pozwala w łatwy sposób zarządzać dostarczaną do rdzeni energią elektryczną oraz rozkładem częstotliwości pracy dla poszczególnych układów. Jest to możliwe ponieważ każdy bank jest niezależnie zasilany oraz może być sterowany własnym sygnałem zegarowym.

W układzie SCC zastosowano mechanizm niezależnego taktowania zegarów również wewnątrz banków. Efektem takiej konstrukcji jest pobieranie przez procesor złożony z 48 rdzeni jedynie od 25 do 125W [8,9]. Drugą istotną zaletą takiego rozwiązania jest możliwość realizacji przez każdy kafelek innego, niezależnego zadania, a nawet na każdym z nich można uruchomić inny system operacyjny.

Do budowy układu SCC wybrano proste technologicznie układy Pentium, które pracują z częstotliwością 533MHz. Celem w tym przypadku prowadzonych prac było empiryczne sprawdzenie możliwości współpracy wielu standardowych rdzeni procesorowych. Należy przyjąć że w niedalekiej przyszłości dostępne będą na rynku układy składające się z rdzeni procesorów Core i7 i pomimo że SCC ma 48 rdzeni, architektura ta pozwala na połączenie ze sobą nieograniczonej ilości jednostek.

3. Wnioski

W przeszłości układy mikroprocesorowe wykonywały operację sekwencyjnie. Ograniczenia technologiczne w produkcji zmusiły inżynierów do zmiany techniki przetwarzania informacji. W rezultacie rozpoczęto prace nad nową formą wykonywania obliczeń, w której nieokreślona liczba instrukcji wykonywana jest jednocześnie. Operację realizowane w sposób równoległy oferują duże korzyści dla osób potrafiących sobie wyobrazić przyszłe potrzeby i zrozumieć jak zmiany te mogą poprawić wydajność w poszczególnych dziedzinach.

Konstrukcja mikroprocesorów wielordzeniowych będzie się w najbliższej przyszłości dynamicznie rozwijać. Wydaje się że szybciej wdrożone zostaną układy wywodzące się z architektury kart graficznych, gdyż technologia ta jest praktycznie gotowa. Producenci sprzętu komputerowego przygotowują swoje produkty do nadchodzącej rewolucji. Jeden z największych producentów układów graficznych, firma Nvidia, stworzyła układy oparte na najnowszej architekturze Fermi (np. karty GeForce GTX580), którą zaprojektowano z myślą o wysokowydajnych obliczeniach, a następnie dodano do nich elementy niezbędne do generowania grafiki [7].

Układy wieloprocessorowe wywodzące się ze schematu budowy klastrów, takie jak omawiany projekt Single-Chip Cloud Computer, pojawią się na pewno później niż rozwiązania oparte o architekturę kart graficznych. Wiąże się to ze stworzeniem technologii która, pozwoli na wyprodukowanie względnie małego układu zawierającego kilkanaście czy też kilkadziesiąt rdzeni przy niewielkim koszcie produkcji. Mimo to firma Intel zamierza przekazać kilkadziesiąt prototypowych układów społeczności akademickiej w celu prowadzenia praktycznych badań nad opracowywaniem nowych aplikacji i modeli programistycznych.

Literatura

1. Intel: The evolution of a revolution. [dostęp 22-12-2011] <http://download.intel.com/pressroom/kits/IntelProcessorHistory.pdf>

2. Metzger P.: Anatomia PC. Wydanie IX, Helion, 09.2004.
3. Gordon E. Moore: Cramming more components onto integrated circuits. Electronics Magazine 38 (8), 19 kwietnia 1965. [dostęp 05 styczeń 2012].
ftp://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf
4. Malczewski Ł.: Kryzys kontrolowany. Personel i Zarządzanie, 5/2003, Warszawa 2003.
5. Intel Pentium D Processor 800 Datasheet. Document Number: 307506-00, luty 2006. [dostęp 22 grudzień 2011] <http://www.intel.com/Assets/PDF/datasheet/307506.pdf>
6. Floyd S.: Workstation Performance: Tomorrow's Possibilities. Cadalyst Magazine, 1 maj 2008. [dostęp 23 grudzień 2011]. <http://www.cadalyst.com/hardware/workstation-performance-tomorrow039s-possibilities-viewpoint-column-6351?>
7. NVIDIA CUDA C - Programming Guide, Version 4.0, 5 maja 2011. [dostęp 23 grudzień 2011].
http://developer.download.nvidia.com/compute/cuda/4_0/toolkit/docs/CUDA_C_Programming_Guide.pdf
8. Intel Labs: The SCC Platform Overview. Revision 0.7, 24 maj 2010. [dostęp 22 grudzień 2011].
http://techresearch.intel.com/spaw2/uploads/files/SCC_Platform_Overview.pdf
9. Held j.: Exploring programming models with the Single-chip Cloud Computer research prototype. 2 grudzień 2009 [dostęp 22 grudnia 2011]
<http://blogs.intel.com/research/2009/12/02/sccloudcomp/>

Dr hab. inż. Marcin LORENC, prof. PO
 Mgr inż. Krzysztof CEGIELSKI
 Instytut Innowacyjności Procesów i Produktów
 Politechnika Opolska
 45-370 Opole, ul. Ozimska 75
 tel./fax.: (0-77) 423 40 44
 e-mail: m.lorenc@po.opole.pl
 k.cegielski@po.opole.pl