

**POLSKA
RZECZPOSPOLITA
LUDOWA**



**URZĄD
PATENTOWY
PRL**

OPIS PATENTOWY

132 094

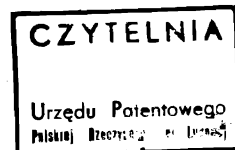
Patent dodatkowy
do patentu _____

Zgłoszono: 79 12 31 (P. 221060)

Pierwszeństwo _____

Zgłoszenie ogłoszono: 81 07 24

Opis patentowy opublikowano: 1987 03 31



Int. Cl.³ G06F 7/48
G06F 7/556

Twórcy wynalazku: Stanisław Majerski, Wojciech Kubera

Uprawniony z patentu: Instytut Maszyn Matematycznych,
Warszawa (Polska)

Cyfrowy układ binarny do obliczania wartości funkcji wykładniczej

Przedmiotem wynalazku jest cyfrowy układ binarny do obliczania wartości funkcji wykładniczej o ustalonej podstawie, przeznaczony do zastosowania w szybkich specjalizowanych procesorach do obliczeń numerycznych, wchodzących w skład komputerów i systemów komputerowych, oraz w innych szybkich urządzeniach cyfrowych obliczających wartości funkcji z dużą dokładnością.

Znane są cyfrowe układy binarne do obliczania wartości funkcji wykładniczej charakteryzujące się tym, że wyznaczaną wartość funkcji wykładniczej otrzymuje się po wykonaniu dodawań lub odejmowań od argumentu funkcji wykładniczej niewielkiej liczby stałych oraz w wyniku dodawań i odejmowań przesuniętych względem siebie odpowiednio dobranych liczb. W skład takich cyfrowych układów wchodzi mały szybki zespół pamięciowy zawierający przeważnie kilkadziesiąt stałych, kilka rejestrów pamiętających przetwarzane liczby i sumator lub kilka sumatorów. Struktura logiczna i działanie takich i podobnych układów opisane są między innymi w następujących publikacjach:

- J.E. Meggit, „Pseudo Division and Pseudo Multiplication Processes”, IBM J. Res. Develop. 6, 210–226 (1962).
- W.H. Specker, „A Class of Algorithms for $\ln x$, $\exp x$, $\sin x$, $\cos x$, $\tan^{-1} x$, and $\cot^{-1} x$ ”, IEEE Trans. Electronic Computers, EC-14, 85–86 (1965).
- Tien Chi Chen, „Automatic Computation of Exponentials, Logarithms, Rations and Square Roots”, IBM J. Res. Develop. No. 4, 380–388, July 1972.
- P.W. Baker, „More Efficient Radix-2 Algorithms for Some Elementary Functions”, IEEE Trans. Computers, C-24, 1049–1054, November 1975.
- A.M. Oranskij, „Apparaturnye metody v cifrovoj vycislitel'noj technike”, 1977, Izdatelstvo BGU im. Lenina, Minsk.

Znany jest również układ cyfrowy dzielenia binarnego przedstawiony w polskim opisie patentowym nr 121732, w którym zamiast kolejnych dodawań lub odejmowań występujących w większości znanych cyfrowych układów dzielenia, zastosowano redukcję równoległą trójek liczb do par liczb. Opisane tam zespoły cyfrowe do wyboru cyfr ilorazu na podstawie par liczb otrzymywanych w wyniku redukcji równoległej adapt-

wane zostały w niniejszym wynalazku jako zespoły cyfrowe wchodzące między innymi w skład cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej.

Zgodnie z wynalazkiem w skład cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej wchodzi zespół pamięciowy stałych, zespół rejestrów równoległych pamiętających liczby przetwarzane w trakcie wyznaczania wartości funkcji wykładniczej, zespół przetwarzający równoległe liczby pobierane ze wspomnianego zespołu rejestrów oraz zespół generujący cyfry, od których zależy wybór liczb z zespołu rejestrów, które są przetwarzane we wspomnianym zespole przetwarzania. W szczególności wspomniane zespoły przetwarzający i generujący mogą tworzyć łącznie jeden zespół wykonujący na przemian funkcje obu tych zespołów. Zespół pamięciowy stałych połączony jest z zespołem generującym cyfry i z zespołem rejestrów równoległych, w celu równoległego przesyłania wybranych stałych do zespołu generującego i zespołu rejestrów. Zespół rejestrów równoległych połączony jest z zespołem generującym cyfry, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole generującym. Zespół rejestrów równoległych połączony jest również z zespołem przetwarzającym, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole przetwarzającym. Zespół przetwarzający połączony jest także z zespołem generującym cyfry, w celu przesyłania tych cyfr do zespołu przetwarzającego, gdzie są one wykorzystywane do sterowania przetwarzaniem informacji. Cyfrowy układ binarny według wynalazku charakteryzuje się tym, że jego zespół generujący cyfry zawiera podzespół redukujący i podzespół dekodujący. Podzespół redukujący dostosowany jest do równoległego redukowania, w kolejnych krokach wyznaczania wartości funkcji wykładniczej, trójek liczb do par liczb z zachowaniem ich sum. Podzespół redukujący połączony jest z podzespołem dekodującym, w celu przesyłania do podzespołu dekodującego grup bitów pobieranych z pięciu, czterech, lub trzech bardziej znaczących pozycji binarnych z liczb redukowanych w podzespole redukującym. Te grupy bitów dekodowane są następnie w podzespole dekodującym na cyfry 0,1,2 lub cyfry -1,0,+1, które sterują przetwarzaniem informacji w zespole przetwarzającym.

Dzięki zastosowaniu omówionego cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej, w którym przez zastosowanie równoległej redukcji trójek liczb i czwórek liczb do par liczb unika się wielu dodawań lub odejmowań, czasochłonnych ze względu na propagację przeniesień, uzyskuje się znaczne skrócenie czasu wyznaczania wartości funkcji wykładniczej, zwłaszcza w tych przypadkach gdy wymagana jest bardzo wysoka dokładność obliczeń.

W szczególności cyfrowym układem według wynalazku jest opisany układ, w którym w skład podzespołu redukującego wchodzi szereg jednopozycyjnych sumatorów binarnych, z których każdy redukuje trójkę bitów do pary bitów, przy czym sumatory te nie mają między sobą połączeń służących do przesyłania informacji.

Cyfrowym układem według wynalazku jest również niżej opisany cyfrowy układ binarny do obliczania wartości funkcji wykładniczej w skład którego wchodzi zespół pamięciowy stałych, zespół rejestrów równoległych pamiętających liczby przetwarzane w trakcie wyznaczania wartości funkcji wykładniczej, zespół przetwarzający równoległe liczby pobierane ze wspomnianego zespołu rejestrów oraz zespół generujący cyfry, od których zależy wybór liczb z zespołu rejestrów, które są przetwarzane we wspomnianym zespole przetwarzania. Zespół pamięciowy stałych połączony jest z zespołem generującym cyfry i z zespołem rejestrów równoległych, w celu równoległego przesyłania wybranych stałych do zespołu generującego i zespołu rejestrów. Zespół rejestrów równoległych połączony jest z zespołem generującym cyfry, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole generującym. Zespół rejestrów równoległych połączony jest również z zespołem przetwarzającym, w celu równoległego dwukierunkowego przesyłania informacji przetwarzanych w zespole przetwarzającym. Zespół przetwarzający połączony jest także z zespołem generującym cyfry, w celu przesyłania tych cyfr do zespołu przetwarzającego, gdzie są one wykorzystywane do sterowania przetwarzaniem informacji.

Cyfrowy układ binarny według wynalazku charakteryzuje się tym, że wspomniany zespół przetwarzający mający strukturę logiczną dostosowaną do równoległego redukowania czwórek liczb do par liczb z zachowaniem ich sum jest połączony równoległe i dwukierunkowo z zespołem rejestrów w celu równoczesnego przesyłania równoległego czwórek liczb lub par liczb między tymi zespołami. Dzięki zastosowaniu omówionego cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej, w którym przez zastosowanie równoległej redukcji trójek liczb i czwórek liczb do par liczb unika się wielu dodawań lub odejmowań czasochłonnych ze względu na propagację przeniesień, uzyskuje się znaczne skrócenie czasu wyznaczania wartości funkcji wykładniczej, zwłaszcza w tych przypadkach gdy wymagana jest bardzo wysoka dokładność obliczeń.

W szczególności cyfrowym układem według wynalazku jest opisany ostatnio układ, w którym w skład zespołu przetwarzającego wchodzi dwa szeregi jednopozycyjnych sumatorów binarnych redukujących trójki bitów do par bitów, przy czym sumatory należące do tego samego szeregu nie mają między sobą połączeń służących do przesyłania informacji.

Strukturę cyfrowego układu według wynalazku wyjaśnia dokładniej następujące rozumowanie.

Obliczanie wartości $z = a^y$ dla ustalonej dodatniej liczby a i dodatniej liczby y mającej w zapisie binarnym postać $y = x.xxx...$, gdzie x oznacza bit 0 lub 1, można przeprowadzić następująco. Mając dany ciąg stałych $\log_2(1+2^{-i})$, dla kolejnych liczb całkowitych i można wybrać z nich taki ciąg liczb $\log_2(1+q_i 2^{-i})$, gdzie cyfry

$q_i \in \{0,1,2\}$, aby liczba y była sumą wyrazów tego ciągu. Wybór taki odbywa się przez próby odejmowań kolejnych stałych $\log_a(1+2^{-i})$ i wyznaczenie na tej podstawie wartości q_i dla kolejnych liczb całkowitych i , w sposób analogiczny do prób odejmowania lub dodawania przesuwanego dzielnika w trakcie wykonywania dzielenia i wyznaczania na tej podstawie kolejnych cyfr ilorazu. Na podstawie otrzymanej z wzorów

$$z = a^y \quad y = \sum_i \log_a(1+q_i 2^{-i}) \quad (1)$$

równości

$$\log_a z = \sum_i \log_a(1+q_i 2^{-i}) \quad (2)$$

można, z wyznaczonego ciągu cyfr q_i wyliczyć szukaną wartość

$$z = \prod_i (1+q_i 2^{-i}) \quad (3)$$

Każde z wynikających z ostatniego wzoru mnożeń przez kolejne czynniki $1+q_i 2^{-i}$, gdzie $q_i \in \{0,1,2\}$, sprowadza się do przesunięcia i dodania jednej, poprzednio otrzymanej liczby.

Ciąg kolejnych odejmowań lub dodawań prowadzący do wyznaczania cyfr q_i można zastąpić przez ciąg kolejnych redukcji równoległych trójek liczb do par liczb z zachowaniem ich sum. Cyfry q_i wyznacza się wówczas na podstawie grup bitów wziętych z trzech lub czterech pozycji binarnych ze zredukowanych par liczb. Analogicznie ciąg kolejnych dodawań, wymaganych dla wykonania ciągu kolejnych mnożeń przez czynniki $1+q_i 2^{-i}$, można zastąpić przez ciąg kolejnych redukcji równoległych czwórek liczb do par liczb z zachowaniem ich sum. Suma liczb ostatniej otrzymanej w ten sposób pary zredukowanych liczb stanowi szukaną wartość funkcji wykładniczej.

Analogiczne rozumowanie można przeprowadzić również dla czynników postaci $1-q_i 2^{-i}$ gdzie $q_i \in \{0,1,2\}$ i odpowiadających im stałych $\log_a(1-2^{-i})$, jak również przy założeniu że $q_i \in \{-1,0,+1\}$. W ostatnim przypadku potrzebne są zarówno stałe postaci $\log_a(1+2^{-i})$ jak i stałe postaci $\log_a(1-2^{-i})$.

W cyfrowym układzie binarnym według wynalazku, o strukturze logicznej dla której przyjęto, że $q_i \in \{0,1,2\}$ stałymi zapisanymi w zespole pamięciowym są liczby $-\log_a(1+2^{-i})$ lub liczby $-\log_a(1-2^{-i})$ dla ustalonej wartości a , oraz $i=1,2,\dots,m$, gdzie m nie przekracza liczby bitów wyznaczanej liczby a^y .

Struktura logiczna cyfrowego układu binarnego według wynalazku może być również dostosowana do cyfr $q_i \in \{-1,0,+1\}$ zamiast cyfr $q_i \in \{0,1,2\}$. W przypadku takim przeprowadzone wyżej rozumowanie ulega tylko niewielkim modyfikacjom. O wyborze kolejnej cyfry q_i decyduje wtedy grupa bitów branych z trzech lub czterech pozycji binarnych z dwóch składników otrzymanych w wyniku redukcji. Dla cyfr $q_i \in \{-1,0,+1\}$ stałymi zapisanymi w zespole pamięciowym są zarówno liczby $-\log_a(1+2^{-i})$ jak i liczby $-\log_a(1-2^{-i})$ dla ustalonej wartości a oraz $i=1,2,\dots,m$, gdzie m nie przekracza liczby bitów wyznaczanej liczby a^y .

Liczba pozycji binarnych, z których są brane grupy bitów dekodowanych na cyfry q_i może być zwiększona do pięciu, w przypadkach gdy bity te są pobierane nie z liczb zredukowanych, ale z liczb redukowanych. Stosowanie odpowiadających takiemu przypadkowi rozwiązań układu może skrócić czas wyznaczania wartości funkcji wykładniczej kosztem rozbudowy układu.

Przedmiot wynalazku został pokazany w przykładzie wykonania na rysunku przedstawiającym schemat blokowy cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej.

Cyfrowy układ binarny do obliczania wartości funkcji wykładniczej obejmuje następujące zespoły cyfrowe: zespół pamięciowy stałych M zawierający kilkadziesiąt stałych $-\log_2(1+2^{-i})$ oraz niewielką liczbę innych stałych; zespół rejestrów R obejmujący siedem rejestrów równoległych; zespół przetwarzający P zawierający dwa szeregi jednopozycyjnych sumatorów binarnych nie połączonych ze sobą w ramach tego samego szeregu; zespół generujący G , który generuje cyfry $q_i \in \{0,1,2\}$, zawierający podzespół redukujący S , który redukuje trójki liczb do par liczb, oraz podzespół dekodujący D , który dekoduje grupy bitów wziętych z czterech pozycji binarnych z par liczb zredukowanych w podzespole S na kodowane binarnie cyfry 0,1,2. Podzespół S obejmuje szereg nie połączonych ze sobą jednopozycyjnych sumatorów binarnych, z których każdy redukuje trójkę bitów do pary bitów. Podzespół dekodujący D zawiera szybki czteropozycyjny binarny sumator równoległy dodający modulo 2^4 dwie czterobitowe liczby oraz zawiera dekodery, dekodujące czwórki bitów otrzymane w wyniku tego dodawania na zakodowane binarnie cyfry 0,1,2, przy czym czwórkom bitów 00xx, 01xx, 1xxx przyporządkowuje się odpowiednio cyfry 0,1,2, gdzie x oznacza dowolny bit. Poszczególne zespoły i podzespoły cyfrowego układu połączone są ze sobą w sposób następujący: wyjście 1 zespołu M ma połączenie z wejściem 2 podzespołu S i z wejściem 3 zespołu R służące do równoległego przesyłania stałych $-\log_2(1+2^{-i})$ dla kolejnych wartości i ; wyjście 4 zespołu R ma połączenie z wejściem 5 podzespołu S służące do równoczesnego równoległego przesyłania trzech liczb; wyjście 6 podzespołu S ma połączenie z wejściem 7 zespołu R służące do równoczesnego równoległego przesyłania dwóch liczb; wyjście 8 podzespołu S ma połączenie z wejściem 9 podzespołu D służące do równoległego przesyłania dwóch czwórek bitów; wyjście 10 podzespołu D ma połączenie z wejściem 11 podzespołu S i z wejściem 12 zespołu P służące do przesyłania zakodowanych binarnie cyfr 0,1,2; wyjście 13 zespołu R ma połączenie z wejściem 14 zespołu P służące do równoczesnego równoległego przesyłania czterech liczb; wyjście 15 zespołu P ma połączenie z wejściem 16 zespołu R służące do równoczesnego równoległego przesyłania dwóch liczb.

Działanie omawianego cyfrowego układu binarnego do obliczania wartości funkcji wykładniczej obejmuje wykonanie niżej opisanych czynności. Dodatnia liczba y z zakresu $1 \leq y < 2$ o postaci binarnej $1,xxx\dots$ zostaje wprowadzona do jednego z rejestrów zespołu R, przy czym do drugiego i trzeciego rejestru zespołu R zostaje wprowadzona jedna ze stałych z zespołu M lub zero, w zależności od pierwszych dwóch bitów po przecinku liczby y . Od tych dwóch bitów liczby y zależą również pierwsze cyfry q_i i pierwszy czynnik postaci $1+2^{-i}$ liczby $z=a^y$, który wprowadzony zostaje do czwartego i szóstego rejestru zespołu R. Na tym zakończony zostaje wstępny krok wyznaczania liczby $z=a^y$. Każdy z dalszych kroków obejmuje: redukcję równoległą w podzespolu S trójki liczb, wprowadzonych z trzech pierwszych rejestrów zespołu R i/lub z zespołu M, do pary liczb; wprowadzenie pary zredukowanych w podzespolu S liczb do drugiego i trzeciego rejestru zespołu R; redukcję równoległą w zespole P czwórki liczb wprowadzonych z rejestrów zespołu R do pary liczb; wprowadzenie pary zredukowanych w zespole P liczb do czwartego i piątego rejestru zespołu R i tej samej ale odpowiednio przesuniętej pary liczb do szóstego i siódmego rejestru zespołu R; dekodowanie w podzespolu D dwóch czwórek bitów, wybranych z pary zredukowanych liczb w podzespolu S, na dwubitowo kodowaną cyfrę 0,1 lub 2, która steruje wyborem z zespołów R i M trójki liczb redukowanych w podzespolu S, oraz wyborem czwórki liczb z czwartego, piątego, szóstego i siódmego rejestru zespołu R lub zer, które redukowane są do pary liczb w zespole P. Po ostatnim takim kroku otrzymuje się w zespole P dwa składniki końcowego wyniku, które mogą być zsumowane bądź w tym samym zespole P, bądź w dowolnym sumatorze znajdującym się poza omówionym cyfrowym układem binarnym do wyznaczania wartości funkcji wykładniczej.

Zastrzeżenia patentowe

1. Cyfrowy układ binarny do obliczania wartości funkcji wykładniczej, w którym zespół pamięciowy stałych połączony jest z zespołem generującym cyfry i z zespołem rejestrów równoległych w celu równoległego przesyłania wybranych stałych do zespołu generującego i zespołu rejestrów, oraz w którym zespół rejestrów równoległych połączony jest z zespołem generującym cyfry, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole generującym, jak również w którym zespół rejestrów równoległych połączony jest z zespołem przetwarzającym, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole przetwarzającym, przy czym zespół przetwarzający połączony jest także z zespołem generującym cyfry, w celu przesyłania tych cyfr do zespołu przetwarzającego, gdzie są one wykorzystywane do sterowania przetwarzaniem informacji, przy czym w szczególności zespoły przetwarzający i generujący mogą stanowić łącznie również jeden zespół wykonujący na przemian funkcje obu tych zespołów, z n a m i e n n y t y m, że podzespół redukujący (S) zespołu generującego (G), dostosowany do równoległego bezprzeniesieniowego redukowania trójek liczb do par liczb z zachowaniem ich sum, połączony jest z podzespołem dekodującym (D) zespołu generującego (G), w celu przesyłania grup bitów pobieranych z pięciu, czterech, lub trzech bardziej znaczących pozycji binarnych z liczb redukowanych w podzespolu redukującym (S), dla dekodowania tych grup bitów w podzespolu dekodującym (D) na cyfry 0,1,2 lub cyfry $-1,0,+1$, które sterują przetwarzaniem informacji w zespole przetwarzającym (P).

2. Cyfrowy układ binarny według zastrz. 1, z n a m i e n n y t y m, że w skład podzespołu redukującego (S) wchodzi szereg jednopozycyjnych sumatorów binarnych, z których każdy redukuje trójkę bitów do pary bitów, przy czym sumatory te nie mają między sobą połączeń służących do przesyłania informacji.

3. Cyfrowy układ binarny do obliczania wartości funkcji wykładniczej, w którym zespół pamięciowy stałych połączony jest z zespołem generującym cyfry i z zespołem rejestrów równoległych w celu równoległego przesyłania wybranych stałych do zespołu generującego i zespołu rejestrów, oraz w którym zespół rejestrów równoległych połączony jest z zespołem generującym cyfry, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole generującym, jak również w którym zespół rejestrów równoległych połączony jest z zespołem przetwarzającym, w celu równoległego dwukierunkowego przesyłania informacji, przetwarzanych w zespole przetwarzającym, przy czym zespół przetwarzający połączony jest także z zespołem generującym cyfry, w celu przesyłania tych cyfr do zespołu przetwarzającego, gdzie są one wykorzystywane do sterowania przetwarzaniem informacji, przy czym w szczególności zespoły przetwarzający i generujący mogą stanowić łącznie również jeden zespół wykonujący na przemian funkcje obu tych zespołów, z n a m i e n n y t y m, że zespół przetwarzający (P) mający strukturę logiczną dostosowaną do równoległego redukowania czwórek liczb do par liczb z zachowaniem ich sum, jest połączony równolegle i dwukierunkowo z zespołem rejestrów (R) w celu równoczesnego przesyłania równoległego czwórek liczb lub par liczb między tymi zespołami.

4. Cyfrowy układ binarny według zastrz. 3, z n a m i e n n y t y m, że w skład zespołu przetwarzającego (P) wchodzi dwa szeregi jednopozycyjnych sumatorów binarnych, redukujących trójki bitów, przy czym sumatory należące do tego samego szeregu nie mają między sobą połączeń służących do przesyłania informacji.

