

Politechnika Świętokrzyska

Laboratorium
Mikrokontrolerów

Ćwiczenie 3
Operacje stałoprzecinkowe

dr inż. Robert Kazała

Kielce 2015

Cel ćwiczenia

Celem ćwiczenia jest poznanie operacji na liczbach w zapisie stałoprzecinkowym.

Obliczenia stałoprzecinkowe

Procesor ARM Cortex 3M jest procesorem całkowitoliczbowym, jeżeli istnieje potrzeba obliczeń zmiennoprzecinkowych to należy je emulować z wykorzystaniem obliczeń całkowitoliczbowych. Emulacja obliczeń zmiennoprzecinkowych jest czasochłonna co powoduje znaczne zmniejszenie ogólnej wydajności systemu.

Alternatywnym rozwiązaniem jest wykorzystanie obliczeń stałoprzecinkowych, które umożliwiają wykonywanie obliczeń na liczbach rzeczywistych w efektywniejszy obliczeniowo sposób.

W arytmetyce komputerowej liczby mogą być aproksymowane przez wykorzystanie pary liczb całkowitych (n,e), gdzie n-mantysa, e-eksponent. Para ta reprezentuje liczbę, którą można zapisać w postaci

$$n2^{-e}$$

Ponieważ we wzorze występuje potęga liczby 2 to wszystkie operacje zmiany eksponentu polegają na przesuwaniu liczb w prawo przy ujemnym wykładniku i w lewo przy dodatnim. W przypadku zapisu liczby w rejestrze mikrokontrolera eksponent e może być traktowany jako liczba cyfr o jaką należy przesunąć wirtualny przecinek.

Mantysa (n)	Eksponent (e)	Postać binarna	Postać dziesiętna
01100100	-1	011001000.	200
01100100	0	01100100.	100
01100100	1	0110010.0	50
01100100	2	011001.00	25
01100100	3	01100.100	12,5
01100100	7	0.1100100	0,78

Jeżeli e jest wartością zmienną przechowywaną w rejestrze i nieznaną w czasie kompilacji to zapis jest zapisem zmiennoprzecinkowym. Jeżeli e jest znane w czasie kompilacji i nie jest przechowywane w pamięci to mamy do czynienia z zapisem stałoprzecinkowym. W tym wypadku w pamięci przechowywana jest tylko mantysa.

Dla operacji stałoprzecinkowych eksponent e jest najczęściej zapisywany jako q. Jeżeli przyjmiemy, że mamy dwie liczby zapisane w formacie stałoprzecinkowym, $a = n2^{-p}$, $b = m2^{-q}$ to podstawowe operacje matematyczne na tych liczbach mają następujące postacie:

- **zmiana eksponentu**

Operację zmiany eksponentu można opisać następującym wyrażeniem

$$n2^{-p} = n2^{r-p} x 2^{-r}$$

Odpowiada to następującym operacją do wykonania:

$$k = n \ll (r-p) \text{ if } (r \geq p)$$

$$k = n \gg (p-r) \text{ if } (p > r)$$

- **dodawanie i odejmowanie**

W celu wykonania operacji dodawania i odejmowania należy tak przekonwertować liczby aby miały przecinek na tej samej pozycji. Wynik dodawania można zapisać w postaci

$$n2^{-r} + m2^{-r} = (n+m)2^{-r}$$

Po dokonaniu konwersji wykonujemy operacje takie same jak na liczbach całkowitych. Po wykonaniu obliczeń przecinek pozostaje na tej samej pozycji.

- **mnożenie**

Mnożenie liczb $c = ab$ może być wykonane za pomocą jednej operacji mnożenia całkowitoliczbowego. Wynik operacji można przedstawić za pomocą następującej zależności

$$ab = n2^{-p} x m2^{-q} = (nm)2^{-(p+q)}$$

Należy zwrócić uwagę, że eksponent jest sumą liczb p i q. W związku z tym po wykonaniu operacji mnożenia mantys należy odpowiednio przesunąć przecinek, żeby znajdował się na wymaganej pozycji r. Można to zrealizować w następujący sposób

$$k = (n*m) \gg (p+q-r)$$

- **dzielenie**

Dzielenie $c = a/b$ może być także zrealizowane za pomocą jednego dzielenia całkowitoliczbowego. Opisuje to następujące równanie

$$\frac{a}{b} = \frac{n2^{-p}}{m2^{-q}} = \left(\frac{n}{m}\right)2^{q-p} = \left(\frac{n}{m}\right)2^{r+q-p}2^{-r}$$

W celu ograniczenia utraty precyzji mnożenie 2^{r+q-p} powinno być wykonane przed dzieleniem przez m. Można to opisać w następujący sposób zakładając, że $r+q \geq p$

$$k = (n \ll (r+q-p)) / m$$

Przebieg ćwiczenia

1. Zapisać z możliwie największą dokładnością w rejestrach 8-, 16-, 32-bitowych liczby 324051, 0.0153, 367234.456876, w wersji ze znakiem oraz bez znaku oraz -324051, -0.0153, -367234.456876 w wersji ze znakiem. Określić dokładność z jaką można zapisać liczbę, określić ile bitów należy przeznaczyć do zapisania tych liczb z dokładnością 1%.
2. Napisać program wykonujący dodawanie, odejmowanie, mnożenie i dzielenie liczb znajdujących się w tabeli

Argument 1	Argument 2
324051	0.0153
324051	-0.0153
324051	367234.456876
324051	-367234.456876
324051	-324051
324051	324051
0.0153	0.0153
0.0153	-0.0153
367234.456876	367234.456876
367234.456876	-367234.456876
367234.456876	0.0153
367234.456876	-0.0153
-367234.456876	-0.0153

Obliczenia wykonać w wersjach ze znakiem i bez znaku jeśli to jest możliwe.
Określić dokładność z jaką zapisany jest wynik.

3. Utworzyć tablice o wymiarze 10, zawierającą losowe napięcia zebrane z przetworników AC o rozdzielczości 10 bitów i zakresie napięć wejściowych $0 \div 5V$ oraz $-5 \div +5V$. Na wejściu znajduje się wzmacniacz o wzmocnieniu 0,5. Napisać program dokonujący skalowania danych do wartości rzeczywistych i zapisać dane w rejestrach 8-, 16-, 32-bitowych w postaci całkowitej oraz stałoprzecinkowej, dla każdej wartości określić dokładność reprezentacji.
4. Napisać różne wersje programu realizującego filtr uśredniający o długości 10, dane i współczynniki wpisać do odpowiednich tabel. Program napisać w wersji: z minimalną liczbą instrukcji oraz maksymalną wydajnością. Oba programy w wersji z instrukcjami mnożenia z dodawaniem oraz bez tych instrukcji. Określić ilość cykli potrzebnych na wykonanie pętli filtru. Przyjmując częstotliwość pracy zegara procesora na 50MHz i zakładając, że każda instrukcja wykonuje się w jednym cyklu obliczyć maksymalną częstotliwość z jaką może być filtrowany przebieg wejściowy.