



národní
úložiště
šedé
literatury

Studie realizovatelnosti neuronové sítě daných vlastností pomocí analogových elektrických obvodů

Hakl, František
2010

Dostupný z <http://www.nusl.cz/ntk/nusl-41393>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 29.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .



Institute of Computer Science
Academy of Sciences of the Czech Republic

Studie realizovatelnosti neuronové sítě daných vlastností pomocí analogových elektrických obvodů

(Fast Analog Circuit Implementation of Switching Neural Network - a
case study)

František Hakl, Miroslav Mrazík

Technical report No. 1068

April 2010



Institute of Computer Science
Academy of Sciences of the Czech Republic

Studie realizovatelnosti neuronové sítě daných vlastností pomocí analogových elektrických obvodů

(Fast Analog Circuit Implementation of Switching Neural Network - a
case study)

František Hakl, Miroslav Mrazík

Technical report No. 1068

April 2010

Abstract:

Neural network with switching units is applicable to diverse data separation problems. Applicability of this universal separation tool was intensively studied in the area of high energy physics on LHC CERN data sets. Potential of the network with switching units among others lies in very fast processing of input data if separating algorithm is implemented as an analog circuit. We show in this report that an implementation based on nowadays best electronic components is able to provide reachable speed of data separation about 25 mega samples per second.

Keywords:

neural network, piecewise linear discrimination, fast data processing, analog circuit, separating algorithm

Summary

NNSU – example of hardware solution

Basic background

Let us suppose the system of 16 sensors which generates signals at high speed. The signal of each sensor is digitized and normalized as 9-bit digital number (8 bit + sign). Sampling rate of the signal is in the order of MSPS and all sensors have the same sampling rate. Required resolution of sensor's data channel is given by the result of the numerical analysis as minimal and concurrently accepted value. The data of all sensors is synchronized. It means there is synchronization signal, which edge has meaning 'data of all sensors is valid'.

The goal is to build hardware system with functionality of NNSU. The main accent is the highest possible data throughput rate and the possibility of changing of NNSU coefficients to modify its properties.

NNSU layout

NNSU can be divided into layers. The data flow goes through individual layers. In every layer the data flow is modified and analyzed to get the set of parameters which will be used in the following layer. The layers work independently and all layers are the same (by hardware, not by parameters). The throughpass of signal data is clocked (synchronized) and every layer has to process its data in shorter time then the clock pulse spacing (edge spacing). Such system can be called pipeline. The data throughput rate depends on the time needed for signal processing in one layer and doesn't depend on number of layers. If the synchronization (clock) pulse starts processing in some layer, the result of this process has not to depend on behaving of previous layer. It has to depend on the state of previous layer at the synchronizing edge only.

The layer has to have the following functionality:

- Input data x_i of each of 16 data channels has to be multiplied with the weight coefficient w_i and the result has to be put at the output of the layer. Weight coefficients w_i are selected from the programmable table according to the result of comparing block of previous layer and actual number of channel. The content of the table is determined by required NNSU properties.

- Output data of every channel (that's product $x_i \times w_i$) has to be added together and the sum has to be compared to the set of values p_j . These values divide full range of possible result of the sum into $(j+1)$ zones. The result of comparison designates one zone which is the sum result lies in. The zone designates the set of weights w of the following layer. The number of zones (that's the number of levels p_j) and levels p_j depends on required properties of NNSU and can be different in different layers.

According to the requirements above we can say the layer consists of the set of equal channel multipliers with programmable weights w , one summation unit and one comparison unit with programmable number of zones and its levels.

Multiplier

The channel multiplier can be designed in a number of ways. The two main ways are analog and digital method of signal processing. The advantage of analog method (that's four-quadrant digital-to-analog converter/multiplier with track&hold at the output) is simplicity of design. The channel value is represented as voltage (or current) level, which can be simultaneously used as the output signal of the layer and the source signal for the summing unit. It is easy to reach required resolution, as multiplying D/A converters are currently used with accuracy of 12 bits and better. The problem is to design four-quadrant multiplier. Operating amplifiers which are

necessary for its design are much slower (at required accuracy) than D/A converter/multiplier itself. Any error in signal path goes through all layers and affects all following processing. The main sources of errors are the input offset voltage (and its temperature drift) and the interference from digital part of system. The signal goes through three operating amplifiers in every layer – two amplifiers in four-quadrant multiplier and one amplifier inside track&hold circuit. Relatively small error on one amplifier damages the signal throughpassing several layers (that's their built-in amplifiers). We have tested one channel of such system including adding unit and comparator. Maximum data rate of such system is in the order of 100 KSPS.

Digital method of signal processing allows to construct system with no error of offset voltage type. Hardware system which is able to multiply two numbers has defined function. But such system is slow. NNSU does not require multiplication of any two numbers. It requires multiplication of random number x_i and one number w_i which is selected from known and small set of numbers. Content of the set is known before starting operation of system. Let us suppose that the count of numbers in the set is 16. It is given by count of zones of comparison unit of previous layer and 16 is reasonable count. Taking required resolution of channel data (9 bit) and the count of weight coefficients (2^4), it is possible to describe all states of multiplier using the table of size $2^{(9+4)}$ numbers. This table can be represented by high speed SRAM which can be downloaded before NNSU starts its work. It allows us to modify required parameters of NNSU. More, SRAM is much faster then ROM (FLASH, EEPROM, ...). It is necessary to use 'true' RAM. Some RAMs seem to be very fast but the speed (access time) is calculated from the block operation. It is not possible to use such type of access in NNSU multiplier. Multiplier needs to address one cell of RAM (resp. two RAMs in parallel for 9-bit data bus). The address is combined using the content of input data x_i and the address of selected weight w_i . The result of multiplication comes from data bus of the memory. Critical time is described as address-to-data access time (t_{AA}) and can be very small. The high speed SRAMs with capacity of $2^{16} - 2^{17}$ bytes can have $t_{AA} = 8$ ns. This is no problem to extend the resolution of the system or the number of weight coefficients. Required capacity of RAM is 2^{13} and available chip capacity is 2^{16} B.

The memory containing of the product table has to be separated using address bus transparent latches. The reason is to separate layers after the synchronization clock edge comes to layer. Before the system starts its work it is necessary to download contents of memories. This function can be realized using common microcomputers (there is no time limit for setting up the system). During setting up both address and data buses of the memories have to be driven by microcomputers. That is why address buses have to have transparent latches with ability of high-Z state to allow access of either microcomputer or channel data. The data bus of RAM doesn't need to have the latch as during main function the data bus is switched to output state while during setting up the bus is switched to reading state. Transparent latches/bus drivers with Hi-Z state of type LVHT373 adds delay of 5 ns.

Summing and comparison unit

The situation in summing unit differs with multiplier. An error rising in the summing unit cannot leave the layer as the output of the unit is assort to zone. The output of the layer is the zone number and small error on output of the summing unit is forgotten. Only in case of comparison on the margin of adjacent zones there can be mistake in zone selecting. According to the numerical model the system is not much sensitive on such type of errors. This allows to use analog summing and comparison unit. Digital method of summing means to transport data bus of each of 16 multipliers to summing unit ($16 \times 9 = 144$ signals), to add all data using PLAs or similar system, to compare sum to zone levels and to put the zone number on the

output of the layer. Such system should be very complex and sensitive on interference (long buses). In case of need of different number of channels it is difficult to change the system.

Analog solution of summing unit is much easier. It needs digital-to-analog converter in every channel of the layer. As there is 16 channels and their analog outputs are added together, an error in one channel shows itself as 1/16 of self value. It means the resolution of D/A converter can be $(2^9/2^4)$ 5 bits. Possible D/A converter has 6 ns conversion time at accuracy of 7 bit (HI5660).

Analog output of each channel can be connected to the summing unit using coaxial 50 Ohm cable, which allows proper shaping to meet placement of channel multipliers. The length of the cable should be the same to get the same propagation delay in each channel. Summing unit can be designed as passive resistor network. The number of channels of NNSU can be easily modify by modifying of this resistor network.

Comparison unit can be designed as the set of comparators which compare output of summing unit to desired analog values. This solution is suitable in case of low number of comparators. Every comparator needs to have its own D/A converter to define the comparison level. More, every comparator has input capacity which loads the output of the summing unit and increases settling time of comparison unit. Changing of the number of comparison levels is difficult.

Better and more universal solution is to use fast A/D converter and the programmable table. The output of the summing unit is digitized and the data is used to address the table of zone numbers. Every possible output of A/D converter has assigned the correspondent zone number which is sent to the output of the layer. Possible A/D converter has resolution of 12 bit and conversion time 7 ns (ADS5463). Its analog input (balanced) has to be driven with appropriate amplifier-driver. Suitable amplifier is THS4509. It can be connected directly to resistor network. Unpleasant property of the A/D convertor is different logic level compare to the rest of the system. LVDS level can be converted to LV CMOS using 65LVDT388 driver (3 ns delay). Conversion table has 12-bit address bus. It is built on high speed SRAM the same way as multiplier table. It has to have three-state driver at the address bus (65LVDT388 driver has this property) and the latch/bus driver at data bus as it has to drive 16 multipliers in the following layer. The comparison unit has to be equipped with microcomputer for setting up the content of the table.

Construction

NNSU can be built as modular system. The smallest element of system is channel multiplier. It can be designed as a single unit (PCB) with 9-bit channel input bus, 9-bit channel output bus, 4-bit (or more) input bus for weight selection, analog output (50 Ohm) and supporting signals (supplying, serial bus for setting up, clock signal). Net can be completed with required number of channels (16 in this example) and required number of layers.

The second element is summing and comparison unit. The number of analog inputs (50 Ohm) depends on number of channels. It has to be designed for maximally expected number of channels and unused inputs can be grounded. The sensitivity of the summing unit can be manually set by changing the gain of amplifier to fit full range. The output of comparison unit has to be equipped with bus driver to drive input bus for weight selection of the next layer (4-bit or more). Supporting signals could be the same as in the case of multipliers.

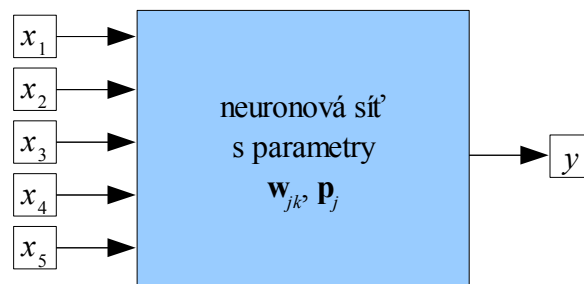
According to the datasheets of common chips mentioned above it is possible to build NNSU with throughput rate of approximately 20 MSPS. This is theoretical value which was not experimentally verified. During finding the appropriate components it was analyzed only small amount of available datasheets thus presented choice is not optimized.

Studie realizovatelnosti neuronové sítě daných vlastností pomocí analogových elektrických obvodů – část 1.

Popis vlastností požadované neuronové sítě

Síť pracuje se vstupním vektorem $\mathbf{x} = (x_1, x_2, \dots, x_N)$, kde x_i jsou digitální data přicházející do sítě v jednom okamžiku. Způsob synchronizace dat vstupního vektoru není řešen, konkrétní řešení závisí na vlastnostech zdroje dat. Data, která vstoupila do sítě, projdou sítí a vyvolají na jejím výstupu nastavení binární hodnoty y v závislosti na vložených parametrech sítě. Hodnota y rozhoduje o tom, jestli vstupní vektor \mathbf{x} patří do jisté množiny (např. vhodné k dalšímu zpracování) nebo do této množiny nepatří.

Situaci je možno znázornit pro $N = 5$ následujícím obrázkem:



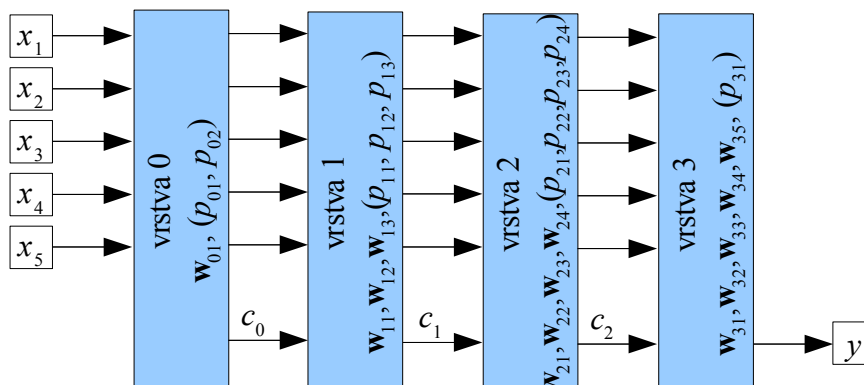
Hodnoty parametrů \mathbf{w}_{jk} a \mathbf{p}_j se do sítě zadají před aplikací vektorů \mathbf{x} a vzhledem k nim je možno parametry \mathbf{w}_{jk} a \mathbf{p}_j považovat za konstanty. Hodnoty těchto parametrů je však třeba operativně a na dálku (tj. elektronickou cestou) měnit podle požadavků na chování neuronové sítě. Význam parametrů \mathbf{w}_{jk} a \mathbf{p}_j vyplývá z algoritmu sítě.

Cílem studie je zjistit, zda je možno realizovat síť pomocí analogových obvodů tak, aby síť byla schopna zpracovat vstupní vektory \mathbf{x} rychlostí aspoň 10^5 vektorů za sekundu s dostatečnou přesností a specifikovat součástky, se kterými by bylo takovou sítí možno realizovat.

Algoritmus sítě

Neuronovou sítí je možno vzhledem k průchodu signálu rozdělit do J vrstev. Jednotlivé vrstvy mají velmi podobnou funkci. Signál (vstupní vektor \mathbf{x}) postupně prochází vrstvami s indexy $j = 0$ až $j = J - 1$, přičemž každá vrstva je charakterizována parametry \mathbf{w}_{jk} a \mathbf{p}_j . Vektory $\mathbf{w}_{jk} = (w_{jk1}, w_{jk2}, \dots, w_{jkN})$ mají všechny stejnou dimenzi N . Dimenze vektorů \mathbf{p}_j se liší pro různé indexy vrstev. Vrstvy 0 až $(j-2)$ mají dimenzi volitelnou mezi čísly 2 a 10 (pro účely této studie). Poslední vrstva má dimenzi 1. Index k se mění od jedné do $(\text{dimenze } \mathbf{p}_{j-1}) + 1$, přičemž pro $j = 0$ je $k = 1$.

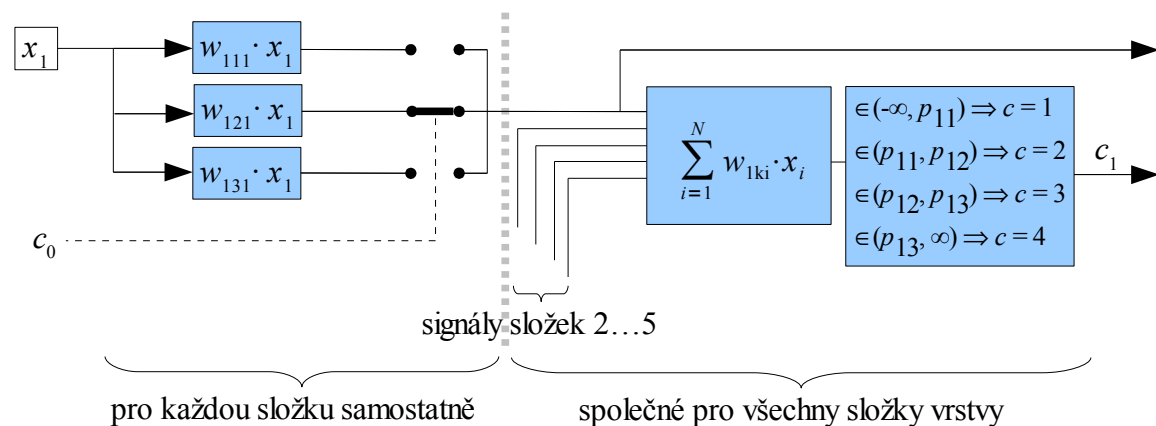
Síť pro $N = 5$, $J = 4$ a postupně rostoucí dimenzi \mathbf{p}_j je možno znázornit takto:



Při průchodu signálu jednotlivými vrstvami síť provádí následující operace (platí pro vnitřní vrstvy):

1. Každá složka vektoru, který vstupuje do vrstvy j , se násobí odpovídající vahou w_{jki} , postupuje na výstup vrstvy a stává se vstupní složkou vrstvy $j + 1$. Vektor \mathbf{x} vstupující do vrstvy j se tedy transformuje podle vztahu $\mathbf{W}_{jk} \cdot \mathbf{x}$, kde $k = c_{j-1}$ a \mathbf{W}_{jk} je diagonální matice dimenze N s diagonálními prvky w_{jki} , $i = 1 \dots N$. O hodnotě indexu k se rozhoduje v předchozí vrstvě podle následujícího pravidla.
2. Parametry \mathbf{p}_j příslušející vrstvě j se seřadí podle velikosti a vytvoří se $\dim(\mathbf{p}_j) + 1$ intervalů:
 $(-\infty, p_{j1}), (p_{j1}, p_{j2}), (p_{j2}, p_{j3}), \dots, (p_{ji}, \infty)$, kde $i = \dim(\mathbf{p}_j)$.
3. Vypočítá se hodnota $s_j = \sum_{i=1}^N w_{jki} \cdot x_{ji}$, kde \mathbf{x}_j je vektor vstupující do vrstvy a $k = c_{j-1}$ je číslo, které bylo zjištěno v předchozí vrstvě (pro $j = 0$ je $k = 1$). Pro $j = 0$ je $\mathbf{x}_0 = \mathbf{x}$, tedy vektor vstupující do celé sítě, a zároveň $w_{01i} \equiv 1$.
4. Hodnota s_j se porovná s intervaly vytvořenými v bodě (2.) a číslu c_j se přiřadí pořadové číslo intervalu, ve kterém s_j leží. Rovnost s_j a krajního bodu intervalu není třeba uvažovat, neboť v analogové elektrické soustavě se s ohledem na šum nemohou dva signály identicky rovnat.
5. Nalezená hodnota c_j se předá spolu s vektorem \mathbf{x}_j vynásobeným vahou \mathbf{w}_{jk} do další vrstvy.

Popsaný algoritmus je možno pro jednu složku vstupního vektoru (zvolme složku x_1) a jednu vrstvu (zvolme druhou vrstvu, $j = 1$) znázornit následujícím schématem:



Složku vektoru vstupujícího do vrstvy je možno označit x_1 , neboť při průchodu nulovou vrstvou se všechny složky vstupního vektoru násobí koeficientem 1, čili procházejí beze změny. Složka x_1 se násobí každým váhovým koeficientem v dané vrstvě a na základě hodnoty c_0 je jeden ze součinů vybrán pro zpracování další vrstvou. Kromě toho vede tento součin do sčítačky, kde se sečtou vybrané součiny všech složek procházejícího vektoru. Suma se porovná s množinou intervalů a zjistí se hodnota c , která bude vybírat součiny s váhami v následující vrstvě.

Všechny koeficienty w a p ve schématu jsou skaláry a pro proud vstupních vektorů jsou konstantní. Jejich zadávání je požadováno v digitální podobě, nejlépe sériovou datovou linkou pomocí PC. Protože výše uvedené schéma se opakuje pro každou složku vstupního vektoru a pro každou vrstvu (vrstvy se liší počty koeficientů w a p), stačí se zabývat pouze realizací matematických operací uvedených ve výše uvedeném schématu – násobení signálu parametrem, součet signálů a zjištění velikosti signálu srovnáním s parametricky danými intervaly hodnot.

Normování signálů

Vstupní vektory i parametry neuronové sítě mohou být obecně reálná čísla. Protože složky vstupních vektorů mají v praxi svůj původ ve snímačích fyzikálních dějů, je jejich rozsah omezen na rozsah snímačů stejně jako jejich rozlišení. Vzhledem k rychlosti generování vstupních vektorů (ne pomaleji než 10^5 vektorů \cdot s⁻¹) lze předpokládat, že jejich reálné rozlišení nebude lepší než 0,1%. Analogová síť by se tedy měla svým rozlišením pohybovat nejméně na této úrovni. Protože signál (složka vstupního vektoru) prochází postupně několika vrstvami sítě a v každé vrstvě dochází ke ztrátě přesnosti na analogovém multiplikátoru, budou se chyby procházejícího signálu produkované každou vrstvou sčítat. V každé vrstvě dojde navíc k přidání chyby sumačního členu a nepřesnosti komparátorů. Tato chyba je v dané vrstvě lokální (nesčítá se při průchodu do další vrstvy), ale pokud se výsledek sumace přiblíží některému krajnímu bodu komparačního intervalu, může dojít k nesprávnému přiřazení intervalu a tím k nesprávné volbě vah v následující vrstvě. V [1] bylo na matematickém modelu prokázáno, že síť je poměrně tolerantní k tomuto druhu chyb a chyby na úrovni několika desetin procenta nezpůsobí ztrátu rozhodovací schopnosti sítě. Pro malý počet vrstev (cca do pěti vrstev) by měla být přesnost v okolí 0,1% pro každý výpočetní prvek (multiplikátor, sčítačka a komparátor) dostačující.

Chyby parametrů vstupují do výpočtu stejným dílem jako chyba procházejícího signálu, neboť chyba součinu je stejnou měrou ovlivněna chybou kteréhokoliv činitele, sčítačka neobsahuje parametr a komparátor je obvod, který pracuje s rozdílem hodnot, čili chyba signálu se projeví stejně jako chyba komparační úrovně. Z těchto důvodů je požadavek na přesnost parametrů w a p stejný jako na signál procházející sítí.

Pro normování signálů je třeba nejprve zvolit typ a rozsah signálu, se kterým bude síť pracovat. Signál může být reprezentován buď velikostí napětí nebo velikostí proudu. Vzhledem k dostupné součástkové základně je rozumné zvolit za nositele signálu napětí, i když, jak vyplývá z dalšího, konverzi signálu na velikost proudu lze v některých místech soustavy s výhodou použít.

Složky vstupního vektoru i parametry mohou nabývat kladných i záporných hodnot. Analogová soustava tedy musí pracovat se signálem obou polarit. Každý aktivní prvek soustavy zanáší do signálové cesty ofsetové napětí, které je dáno výrobními tolerancemi jednotlivých součástek, ale i parazitními termoelektrickými články, kterým se prakticky nelze vyhnout. Toto chybové napětí se může pohybovat i v řádu mV (parametr je v katalogových listech součástek nazýván vstupní napěťová nesymetrie). Každá součástka (včetně pasivních) zanáší do soustavy šum. Jeho analýza by byla vzhledem ke složitosti soustavy velmi náročná a lze říci, že vhodnou volbou součástkové základny bude úroveň šumu pod rozlišovací schopností soustavy.

S ohledem na požadovanou přesnost je třeba zvolit co největší rozsah napětí reprezentujícího signál. Za rozumné maximum je možno zvolit úroveň ± 10 V. S vyšším signálovým napětím nejsou schopny pracovat některé prvky, které jsou potřebné pro vytvoření požadované funkce. Nižší napětí by snížilo poměr signálu k chybovým napětím a tím přesnost realizace matematických funkcí. Ke snížení rozkmitu signálového napětí by se muselo přistoupit při zvýšení rychlosti soustavy, neboť rychlejší součástky pracují s nižší úrovní signálu. To je obecná vlastnost součástek daná konečnými zdroji proudu, které nabíjejí parazitní kapacity polovodičových přechodů nebo hradel tranzistorů FET (ze stejného důvodu výrobci stále snižují napájecí napětí jader číslicových procesorů). Rychlejší součástky mají navíc větší chybová napětí a menší rezervu v zesílení (bude popsáno dále), čímž by se přesnost výpočtů radikálně zhoršila.

Jak složky vstupního vektoru, tak jednotlivé parametry sítě vstupují do sítě v digitální podobě. Před jejich zpracováním je třeba digitální data převést na analogový signál pomocí D/A převodníků. Porovnáním požadované přesnosti analogových signálů a možností součástkové základny v oblasti D/A převodníků s ohledem na rychlost a dostupnost je optimální šířka datového slova 12 bitů. Převodníky s vyšším rozlišením jsou o něco pomalejší, vyšší přesnost nelze využít kvůli nepřesnostem navázaných součástek.

Data pro složky vstupního vektoru i data pro parametry w a p je třeba před aplikací do sítě normovat. Vztah mezi daty a plným rozsahem signálu je popsán lineárním zobrazením

$$\langle 0x000, 0xFFFF \rangle \sim \langle -10 \text{ V}, 10 \text{ V} \rangle.$$

Nejedná se tedy o číslo typu integer se zápornými čísly ve tvaru druhého dvojkového doplňku. Důvod pro tuto reprezentaci je základní vlastnost D/A převodníku vyplývající z jeho vnitřní struktury – převodník umí pracovat pouze s kladnými čísly a bipolární signál se musí vytvořit umělým ofsetem vytvořeným externím obvodem. Nulovému signálu (napětí 0 V) odpovídá digitální údaj 0x800.

Význam digitálního údaje pro váhové koeficienty w je odlišný. Je zřejmé, že analogový multiplikátor se vstupy A a B nemůže mít funkci $Y = A \cdot B$. Kdyby se na vstupech nacházelo napětí např. $A = B = 8 \text{ V}$, na jeho výstupu Y by muselo být napětí 64 V, což není obvodově možné. Proto jsou všechny druhy multiplikátorů konstruovány tak, aby vyhovovaly funkci

$$Y = \frac{A \cdot B}{U_0}, \text{ kde } U_0 \text{ je maximální napětí zpracovávaného signálu, v tomto případě } 10 \text{ V. Pro}$$

účely neuronové sítě je možno realizovanou funkci upravit na $Y = A \cdot \left(\frac{B}{U_0}\right) = A \cdot w$. Váhový

koeficient w tedy nabývá pro rozkmit analogových signálů $\pm 10 \text{ V}$ hodnot z intervalu $\langle -1, 1 \rangle$. Při volbě (výpočtu) váhových koeficientů je nutno zajistit, aby hodnoty w nevybočily z rozsahu ± 1 . Zároveň je tím zajištěno, že v žádné vrstvě se na trase složky vstupního vektoru nevyskytne signál, který by vybočil z limitu $\pm 10 \text{ V}$. Vztah mezi digitálními daty a váhovým koeficientem je tedy popsán lineárním zobrazením

$$\langle 0x000, 0xFFFF \rangle \sim \langle -1, 1 \rangle = \langle w_{min}, w_{max} \rangle.$$

Jiná situace nastává uvnitř obvodu sčítačky. Na každém z N vstupů sčítačky může být teoreticky napětí na hranici signálového rozpětí, tedy např. $+10 \text{ V}$. Protože výstup sčítačky je porovnáván s parametry p , musí být jak tyto parametry, tak výstup sčítačky v rozsahu napětí $\pm 10 \text{ V}$. Z toho vyplývá, že parametry p musejí být normovány stejně jako složky vstupního vektoru, tedy

$$\langle 0x000, 0xFFFF \rangle \sim \langle -10 \text{ V}, 10 \text{ V} \rangle = \langle p_{min}, p_{max} \rangle.$$

Sčítačka musí signál na každém ze svých vstupů lineárně (bez limitace) přičíst k ostatním a výstup musí normovat do rozsahu $\pm 10 \text{ V}$. Musí mít tedy přenosovou funkci

$$Y = \frac{1}{N} \sum_{i=1}^N A_i, \text{ kde } A_i \text{ jsou signály na jednotlivých vstupech a } N \text{ je dimenze procházejícího}$$

vektoru (tj. počet vstupů sčítačky). Zde je možno s výhodou použít převod napětí na proud, sečíst proudy v uzlu o virtuálním napětí 0 V a výsledný proud převést na napětí s požadovaným koeficientem $1/N$. Tím je automaticky zajištěna linearita v rozsahu signálů dokonce větším, než je požadováno.

Normování veličin vstupujících do sítě zobrazuje tabulka:

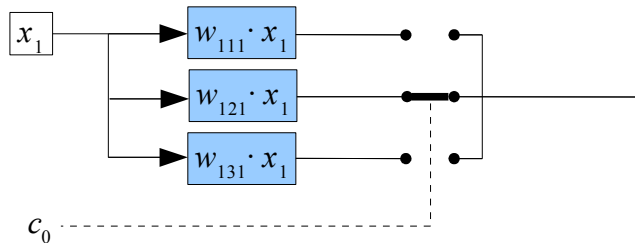
Typ údaje	Datový rozsah	Význam mezí datového rozsahu	Interpretace v síti
x (složka vst. vektoru)	$\langle 0x000, 0xFFFF \rangle$	$\langle x_{min}, x_{max} \rangle, x_{min} = -x_{max}$	$\langle -10 \text{ V}, 10 \text{ V} \rangle$
w (váhový koeficient)	$\langle 0x000, 0xFFFF \rangle$	$\langle w_{min}, w_{max} \rangle, w_{min} = -w_{max}$	$\langle -10 \text{ V}, 10 \text{ V} \rangle$ *)
p (komparační úroveň)	$\langle 0x000, 0xFFFF \rangle$	$\langle p_{min}, p_{max} \rangle, p_{min} = -p_{max}$	$\langle -10 \text{ V}, 10 \text{ V} \rangle$

*) Převod na napětí není nutný, vhodné násobičky dokáží interpretovat digitální data přímo jako váhový koeficient v rozsahu $\langle -1, 1 \rangle$.

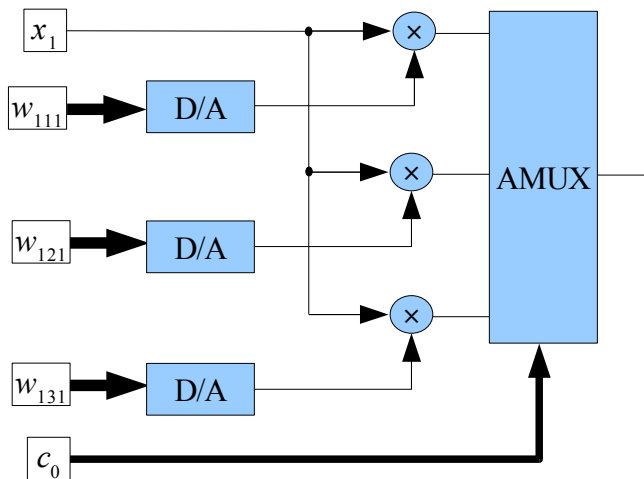
Konstrukce analogové násobičky

Minimalizace bloku násobiček

Vezměme v úvahu blok násobiček popsany výše:

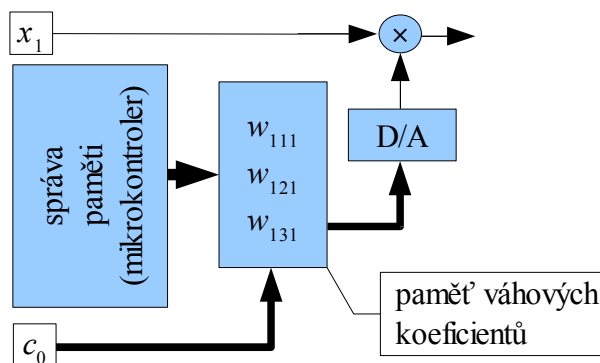


Tento blok je možno rozkreslit následujícím způsobem:



Na realizaci bloku jsou potřeba 3 D/A převodníky, tři násobičky a analogový multiplexer, který vybírá jeden ze tří součinů pro další zpracování. Pro zpracování konkrétní hodnoty x_1 je v reálném čase však zapotřebí pouze jedna násobička a jeden D/A převodník. Výstupní signál ze zbývajících násobiček končí na neaktivních vstupech analogového multiplexeru a není nijak využit.

Protože smysl větvení spočívá v možnosti volby různých váhových koeficientů podle hodnoty c_0 a váhy jsou primárně dány digitálním údajem, nabízí se změna struktury násobičky podle následujícího schématu:



Zapojení je funkčně ekvivalentní a obsahuje pouze jeden D/A převodník a jednu násobičku. Výběr váhového koeficientu je realizován na digitální úrovni výběrem z paměti. Šířka datového slova paměti je 12 bitů, velikost paměti je dána maximálním požadovaným počtem váhových

koeficientů a způsobem jejich adresace. Při číselném způsobu adresování stačí paměť s kapacitou rovnou počtu váhových koeficientů. Protože hodnota c_0 je generována výstupy komparátorů, které tvoří lineární řadu, bylo by nutno signály z výstupů komparátorů překódovat do binárního kódu. Druhá varianta je použit pro adresaci přímo výstupy komparátorů metodou lineární adresace. Pak by kapacita paměti musela být 2^K slov, kde K je počet komparátorů v předchozí vrstvě. Vzhledem ke kapacitám současných pamětí může být např. $K = 16$ pro dostupné paměti s velmi krátkou přístupovou dobou (řádově 50 ns). Paměť musí být typu SRAM, dynamické paměti (např. operační paměti PC) jsou pro tento účel nevhodné. Porovnání obou způsobů adresace pro 6 komparátorů uvádí následující tabulka:

Možné stavy výstupů šestice komparátorů	Lineární adresa A5...A0	Binární adresa A2 A1 A0	Hodnota c
000000	000000 (0x00)	000 (0)	1
000001	000001 (0x01)	001 (1)	2
000011	000011 (0x03)	010 (2)	3
000111	000111 (0x07)	011 (3)	4
001111	001111 (0x0F)	100 (4)	5
011111	011111 (0x1F)	101 (5)	6
111111	111111 (0x3F)	110 (6)	7

Překódování mezi druhým a třetím sloupcem tabulky je možno provést pevně naprogramovanou pamětí (ROM, FLASH), u které se přístupová doba (doba od ustálení adresy do zaručeně platných dat na výstupu) pohybuje kolem 70 ns, nebo pomocí hradlového pole, kde se může doba konverze pohybovat v řádu jednotek nanosekund, avšak za cenu velkého proudového odběru. Jako výhodnější se jeví lineární adresování, které nevyžaduje konverzi adres pro SRAM. Jestliže se použijí paralelně dvě paměti o šířce slova 8 bitů (standardní paměť), je možno zbývající 4 datové bity využít k pomocným účelům, např. pro signalizaci nepřístupného stavu, kdy se uvnitř řady jedniček na výstupech komparátorů objeví nula (vadný komparátor, špatně seřazené komparační úrovně p , neustálené výstupy komparátorů apod.). Tyto pomocné bity mohou napomoci při diagnostice realizované neuronové sítě.

Při zapnutí neuronové sítě nebo při změně jejích parametrů je třeba naplnit paměť SRAM požadovanými hodnotami váhových koeficientů, případně pomocných signalizačních bitů. Pro toto naprogramování slouží mikrokontrolér, který převezme adresovou i datovou sběrnici paměti SRAM a změní její obsah požadovaným způsobem. Mikrokontrolér si může poslední sadu koeficientů pamatovat ve vnitřní paměti FLASH, takže po zapnutí sítě se nahrání paměti SRAM provede autonomně. Při požadavku na změnu parametrů se řídicí PC spojí s mikrokontrolérem pomocí sériové linky a zadá požadovanou změnu, kterou mikrokontrolér samostatně provede.

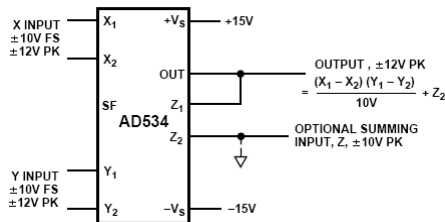
Mikrokontrolér může současně s váhovými koeficienty w udržovat i informaci o komparačních úrovních p v dané vrstvě a starat se pomocí dalšího D/A převodníku o jejich nastavení na odpovídajících vstupech komparátorů.

Čistě analogová varianta násobičky

Násobení váhovým koeficientem je možno provádět dosud uváděnou soustavou násobičky dvou analogových signálů s převodní funkcí $Y = \frac{A \cdot B}{U_0}$ a D/A převodníku s šířkou datového slova 12 bitů a výstupním signálem v rozsahu ± 10 V. Násobiček požadovaných vlastností (přenosová funkce, rozsah signálů, přesnost, rychlost, dostupnost) je na trhu velmi málo a jako nejvhodnější se

jeví obvod AD 534L firmy Analog Devices. Jeho vlastnosti jsou uvedeny v [2] a dostupnost na trhu byla ověřena. Jeho základní vlastnosti, které rozhodují o vhodnosti pro danou aplikaci jsou tyto (hodnoty jsou uváděny při teplotě 25 °C, teplotní součinitele parametrů jsou vyhovující):

- Čtyřkvadrantové násobení dvou analogových signálů v rozsahu ± 10 V.
- Přesnost je lepší než $\pm 0,25\%$, to je $\pm 2,5$ mV.
- Nelinearita přenosu každého z násobících vstupů je lepší než 0,12%.
- Doba ustálení při maximální změně napětí na výstupu je pro chybu 1% typicky 2 μ s.
- Výstupní ofsetové napětí je menší než ± 10 mV, typicky ± 2 mV.



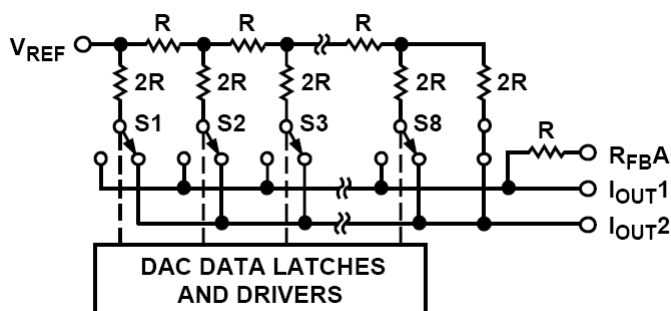
Závislost chyby výstupního napětí na době ustalování není uváděna, na základě analogických údajů pro operační zesilovače lze očekávat, že k ustálení na napětí, které se od statické hodnoty liší o méně než 0,1% dojde za cca 5 μ s.

Na základě uvedených údajů lze konstatovat, že tato součástka je na hranici použitelnosti pro danou aplikaci. Přesnost přenosové funkce vyhoví jen pro velmi malý počet vrstev (chyby se sčítají). Ofsetové napětí je možno vykompenzovat. Doba ustálení výstupu je opět na hranici použitelnosti, prakticky by se mezi jednotlivé vrstvy sítě musely do signálových drah složek vstupního vektoru vložit obvody T/H (track & hold), které by podržely vstupní signál j -té vrstvě, zatímco vrstva $j-1$ by již zpracovávala signál následujícího vstupního vektoru. Tak by se dalo dosáhnout požadované průchodnosti sítě 10^5 vektorů za sekundu.

Byly prostudovány katalogové listy několika výrobců analogových obvodů a výše uvedený typ měl z hlediska požadované aplikace nejlepší vlastnosti. Přesto ani tato analogová násobička není pro danou aplikaci optimální. Jako lepší se jeví řešení násobičky pomocí vhodně zvoleného násobícího D/A převodníku.

Analogově – číslicová násobička

K násobení analogového signálu digitální konstantou je možno použít D/A převodník pracující na principu přepínané žebříčkové sítě $R/2R$. Váhový koeficient udává, jaký podíl vstupního proudu se dostane na výstup převodníku. To znamená, že výstupní proud je přímo úměrný vstupnímu proudu s váhovým koeficientem z intervalu $\langle 0, 1 \rangle$. Následující obrázek ukazuje princip funkce takového převodníku pro 8 bitů (obrázek převzat z [3]):

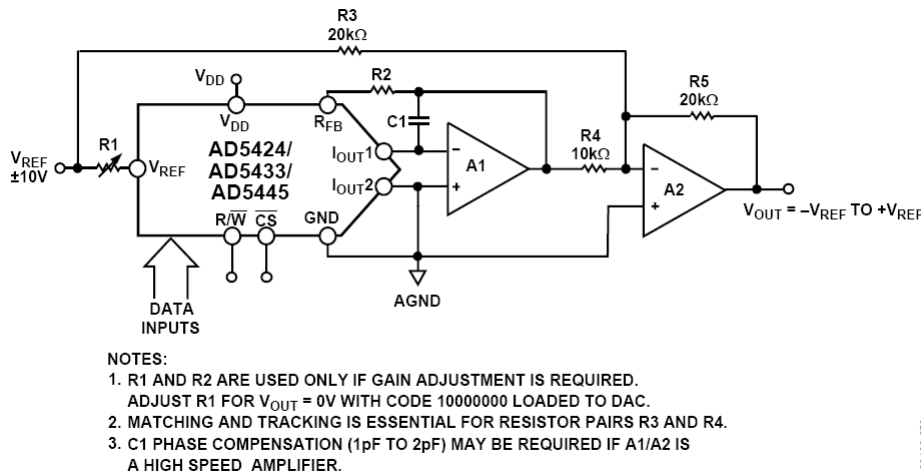


Základní myšlenka převodníku je pomocí přepínačů S1 až S8 rozdělit proud tekoucí ze zdroje V_{REF} do odporové sítě do dvou výstupů. Mají-li oba výstupy I_{OUT1} i I_{OUT2} stejný potenciál, není proud tekoucí do odporové sítě závislý na stavu přepínačů S_i , ale pouze na rozdílu potenciálů V_{REF} a I_{OUT} . Tento proud je roven $-(I_{OUT1} + I_{OUT2})$. Znaménko označuje, že při kladném V_{REF} proud vytéká z obvodu. Je-li potenciál výstupů (fyzicky nebo virtuálně) nulový, potom proud I_{OUT1} , resp. proud tekoucí vývodem R_{FBA} je možno vyjádřit vztahem $-I_{OUT1} = \frac{D}{2^8} \cdot \frac{V_{REF}}{R}$, kde D je binární hodnota $0x00\dots0xFF$ daná stavem $S1\dots S8$ ($S1$ odpovídá MSB).

Teče-li výstupní proud odporem R připojeným k vývodu R_{FBA} , je možno vztah pro výstupní proud přepsat do tvaru $-U_{OUT} = -R \cdot I_{OUT1} = \frac{D}{2^8} \cdot V_{REF} = w \cdot V_{REF}$, tedy do tvaru požadované funkce. Změna znaménka signálu není díky symetrii a bipolaritě neuronové sítě na závadu.

Jsou-li přepínače S_i konstruovány tak, že jimi může téci proud oběma směry (což technologie CMOS umožňuje), může být referenční napětí proměnné a bipolární, takže vhodnou volbou typu převodníku dostaneme přípustné hodnoty referenčního napětí v rozsahu signálů v neuronové síti, tedy ± 10 V.

Hodnota váhového koeficientu $w = \frac{D}{2^8}$ nebo v případě 12-ti bitového převodníku $w = \frac{D}{2^{12}}$ je však pouze kladné číslo z intervalu $\langle 0, 1 \rangle$, takže výsledná přenosová funkce je dvoukvadrantová. Abychom dostali funkci ve všech čtyřech kvadrantech, je nutno zavést umělý ofset do výstupního signálu přičtením vstupního signálového napětí k dvojnásobku výstupního napětí samotného D/A převodníku (obrázek převzat z [3]):

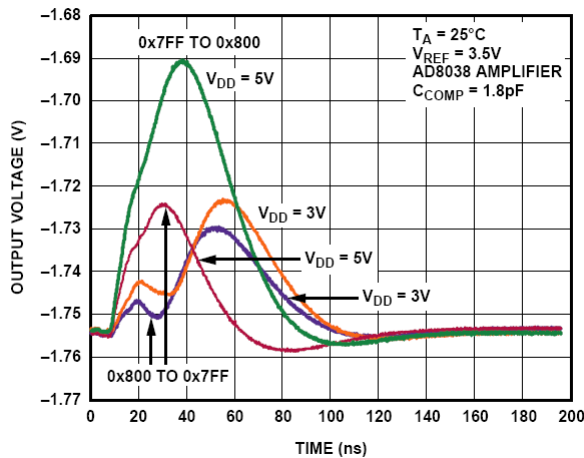


Násobek vstupního napětí je dán poměrem $-R5/R3$, násobek výstupního napětí je dán poměrem $-R5/R4$. Přenosovou funkci je možno pro obvod AD 5445 (12ti bitová varianta převodníku) popsat rovnicí $V_{OUT} = (V_{REF} \cdot \frac{D}{2^{11}}) - V_{REF} = V_{REF} \cdot (\frac{D}{2^{11}} - 1)$. Váhový koeficient $w = \frac{D}{2^{11}} - 1$ nabývá pro data v rozsahu $0x000\dots0xFFF$ hodnot z intervalu $\langle -1, 1 \rangle$.

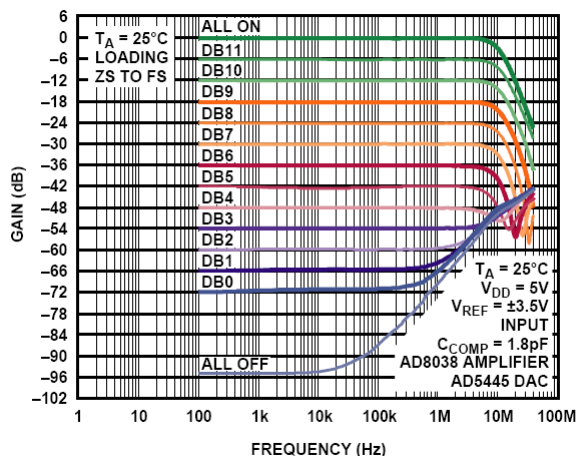
Jak je zřejmé z předchozích řádků, uvedený obvod AD 5445 firmy Analog Devices splňuje všechny požadavky pro konstrukci čtyřkvadrantové násobičky v požadovaném rozsahu procházejícího signálu (tj. ± 10 V) s váhovým koeficientem s rozlišením 12 bitů. Další parametry důležité z hlediska požadované aplikace jsou vypsány z [3]:

- Relativní přesnost ± 1 LSB.
- Diferenciální nelinearita $-1/+2$ LSB.

- Chyba zesílení (odchylka od teoretického napětí pro 0xFFF) max. ± 10 mV.
- Doba ustálení výstupu s přesností 1 mV (0x000 \leftrightarrow 0xFFF) je kratší než 120 ns.



Reakce výstupu na skokovou změnu napětí na vstupu V_{REF} není v [3] přímo popsána, je na ni však možno usuzovat z grafu pro šířku pásma signálu přivedeného na referenční vstup:



Z uvedených grafů lze vyčíst, že do frekvence 4 MHz jsou čáry vodorovné, čili přenos není citelně ovlivněn vnitřními kapacitami. Signál se mění z $-3,5$ V na $+3,5$ V s půlperiodou 125 ns, takže při plném rozkmitu by se měl obvod ustálit za 360 ns. Protože v katalogovém listu nejsou popsány podmínky měření (zejména tvar signálu) a hodnota je lineárně extrapolována, jedná se pouze o hrubý odhad očekávané hodnoty. V obou uvedených grafech jsou křivky závislé na vlastnostech použitého operačního zesilovače. Uvedený typ (AD8038) nelze v dané aplikaci použít, takže grafy slouží jen pro řádovou orientaci.

Prakticky není doba reakce na skokovou změnu napětí na vstupu podstatná, pokud nepřekročí určitou mez. Od doby ustálení napětí na vstupu musí ještě uplynout doba ustálení sčítačky a poté komparátorů předcházející vrstvy, aby bylo možno vybrat správný váhový koeficient z paměti. Tato doba bude určitě delší než 360 ns odhadnutých v předchozím odstavci.

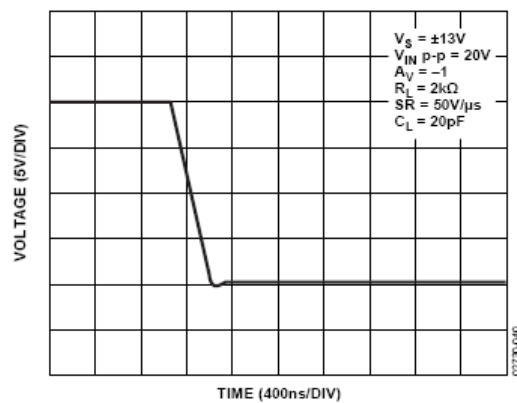
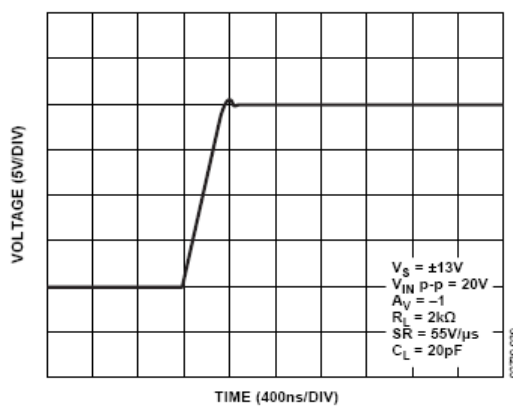
Výběr operačních zesilovačů

Vlastnosti operačních zesilovačů použitých v násobičce (a také ve sčítačce) určují vlastnosti celé neuronové sítě. Kromě základních požadavků na rozsah signálu pro lineární funkci ± 10 V a co nejkratší čas ustálení pro danou přesnost zde vyvstává požadavek na minimální výstupní a maximální vstupní odpor zesilovače. Zatímco minimální výstupní odpor není problém zaručit a

navíc se snižuje účinkem záporné zpětné vazby, vysoký vstupní odpor bývá v rozporu s malou vstupní napěťovou nesymetrií. Proud reprezentující 1 LSB má velikost 250 nA, takže operační zesilovač zajišťující převod proudu na napětí nesmí velikost proudu ovlivnit více než na úrovni desítek nA. Tuto vlastnost mají bez problémů splněny operační zesilovače se vstupy JFET, bohužel bývá vysoký vstupní odpor (nízký vstupní proud) zaplacen větší vstupní napěťovou nesymetrií, která má navíc poměrně velký teplotní součinitel.

Studiem katalogových listů operačních zesilovačů předních výrobců analogových obvodů se podařilo nalézt typ, který splní všechny na něj kladené podmínky. Jde o typ AD 8610 firmy Analog Devices. Vyrábí se ve dvou variantách označených A a B, varianta B je výběr z produkce s ohledem na důležité parametry. Obě varianty jsou dostupné na českém trhu, varianta B je cca 3× dražší. Ve výčtu nejdůležitějších parametrů je uváděna hodnota parametrů varianty B, hodnota levnější varianty A je uvedena v závorce, pokud se liší. Hodnoty platí pro 25 °C a jsou převzaty z [4]:

- Vstupní napěťová nesymetrie lepší než 100 μV (250 μV).
- Teplotní součinitel vstupní napěťové nesymetrie lepší než 1 $\mu\text{V/K}$ (3,5 $\mu\text{V/K}$).
- Vstupní proud menší než ± 10 pA.
- Napěťové zesílení pro velký signál je minimálně 100 V/mV.
- Doba ustálení na 0,01% po skoku výstupního napětí o 10 V je typicky 600 ns.



Z obrázků přeběhu výstupu při skokové změně signálu a zisku $G = -1$ (případ zapojení zesilovačů v násobičce i ve sčítačce) vyplývá, že k ustálení s přesností 1 LSB dojde za cca 500 ns.

Napěťová reprezentace signálu o velikosti 1 LSB je cca 5 mV. Vlastní vstupní napěťová nesymetrie je menší než 250 μV i pro horší z obou vyráběných variant, což je 20× méně než napěťová reprezentace 1 LSB. Není tedy potřeba nesymetrii kompenzovat vnějšími součástkami (úspora nákladů). Vstupní proud je o 4 řády menší, než je požadované maximum. Napěťové zesílení 10^5 je také více než dostatečné.

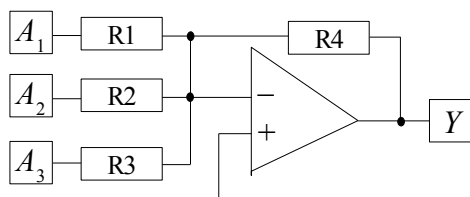
Parametr, který nejvíce omezuje výsledné vlastnosti konstrukce násobičky, je doba ustálení po velkém skoku v signálu nebo váze. Jestliže je doba ustálení D/A převodníku odhadnuta na 380 ns, je doba ustálení dvou navazujících operačních zesilovačů o velikosti 500 ns limitující. Výsledná doba ustálení bude kratší než součet jednotlivých časů. Oba operační zesilovače se ustalují již v době ustalování napětí na výstupech předchozích článků řetězu, takže výsledná doba ustálení celé čtyřkvadrantové násobičky po skokové změně informace na vstupech (x , w) se může pohybovat velmi hrubým odhadem kolem 1 μs , možná o něco méně.

Konstrukce analogové sčítačky

Sčítačka musí signál na každém ze svých vstupů lineárně (bez limitace) přičíst k ostatním a výstup musí normovat do rozsahu ± 10 V. Musí mít tedy přenosovou funkci

$$Y = \frac{1}{N} \sum_{i=1}^N A_i, \text{ kde } A_i \text{ jsou signály na jednotlivých vstupech a } N \text{ je dimenze procházejícího}$$

vektoru (tj. počet vstupů sčítačky). Funkční schéma takové sčítačky pro tři vstupní signály je na následujícím obrázku:



Operační zesilovač v lineárním režimu udržuje na svém invertujícím vstupu napětí 0 V (virtuální zem). Proto je proud tekoucí ze vstupu A_1 do uzlu spojeného s invertujícím vstupem roven

$$I_1 = \frac{A_1}{R_1}. \text{ Totéž lze napsat pro další vstupy. Celkový proud tekoucí do uzlu je } I = \sum_{i=1}^3 \frac{A_i}{R_i}. \text{ Tento}$$

proud musí z uzlu vytéci a protože do vstupu operačního zesilovače téci nemůže, odtéká odporem

R_4 do výstupu Y . Lze tedy psát $Y = -R_4 \cdot I = -R_4 \cdot \sum_{i=1}^3 \frac{A_i}{R_i}$, neboli pro $R_1 = R_2 = R_3 = R$

$$Y = -\frac{R_4}{R} \cdot \sum_{i=1}^3 A_i. \text{ Je-li } R_4 = R / 3, \text{ dostáváme požadovaný vztah } Y = -\frac{1}{3} \cdot \sum_{i=1}^3 A_i. \text{ Záporná}$$

hodnota koeficientu nevadí, protože součet je použit jako signál pro komparátory a otočit znaménko koeficientů p není na digitální úrovni problém (může se o to postarat např. mikrokontrolér vrstvy).

Vhodným kandidátem na typ operačního zesilovače je opět AD 8610A. Pro výběr platí stejné požadavky jako na operační zesilovače pro násobičku, které tento obvod splňuje. Zbývá zvolit vhodnou velikost odporů, aby obvod pracoval s podobnými proudy jako v násobičce. Odpory ve vstupech sčítačky by se měly pohybovat kolem 10 k Ω , zpětnovazební odpor se vypočte podle přenosové funkce.

Konstrukce komparátorů

Výběr komparátorů podléhá stejným kritériím jako výběr operačních zesilovačů, tedy zejména splnění požadavku na rozkmit signálových napětí $\pm 10V$ a co nejkratší dobu ustálení na požadovanou přesnost. Nepodstatné jsou vstupní proudy komparátoru, protože jeden vstup je připojen k výstupu operačního zesilovače sčítačky (výstupní odpor téměř 0 Ω) a druhý ke zdroji stejnosměrného napětí o velikosti p , který bude také tvořen operačním zesilovačem. Málo podstatná je i vstupní napěťová nesymetrie, neboť tu je možno kompenzovat korekcí komparační úrovně p ještě na digitální trase (mikrokontrolér vrstvy). Hlavní důraz je kladen na rychlost při splnění požadavku rozkmitu signálů a přesnosti.

Vhodní kandidáti jsou obvody HA-4905 firmy Intersil a AD 790 firmy Analog Devices. Obvod HA-4905 je čtyřnásobný komparátor, takže by se oproti jednotlivým pouzdrům ušetřila zástavbová plocha. Má však horší parametry než obvod AD 790. Oba obvody jsou dostupné na českém trhu a cena za jeden komparátor je srovnatelná. Základní parametry obou obvodů při teplotě 25 $^{\circ}C$ jsou popsány v tabulce (podle [5] a [6]):

Parametr	HA-4905	AD 790
Vstupní napěťová nesymetrie [mV]	< 7,5	< 0,5
Doba průchodu signálu ze vstupů na výstup [ns]	< 215	< 45

Pokud není třeba uspořít místo na desce s plošnými spoji, je výhodnější zvolit komparátor typu AD 790, u kterého není třeba kompenzovat vstupní napěťovou nesymetrii a je navíc rychlejší.

Zdroj komparační úrovně p

Zadávání komparačních úrovní je možno realizovat různými způsoby. Nejjednodušší je sada D/A převodníků s rozlišením 12 bitů a rozkmitem výstupního napětí ± 10 V, jejichž výstupy jsou přivedeny na vstupy komparátorů. Na každou hodnotu parametru p je třeba jeden převodník. Protože se jedná o statické napětí, lze použít pomalý, levný typ. Z hlediska plošného spoje je nejvýhodnější, má-li převodník sériovou komunikaci. Tím se uspoří i potřebný počet pinů mikrokontroléru.

Je možná i varianta dynamického vytváření komparačních úrovní. Pak je potřeba jeden D/A převodník, analogový demultiplexer a sada paměťových kondenzátorů, pro každou komparační úroveň jeden. Mikrokontrolér nastaví D/A převodník na požadovanou hodnotu a z jeho výstupu nabije pomocí demultiplexeru odpovídající kondenzátor. Stejnou operaci provede i s ostatními kondenzátory a celý cyklus opakuje s takovou rychlostí, aby se na kondenzátorech nezměnilo napětí o více než např. 0,5 LSB. Tato varianta sice vyjde pro větší počet parametrů p ve vrstvě levněji, v soustavě se však objeví asynchronní napětí o velikosti ± 10 V, které může pronikat do signálové cesty složek vstupního vektoru. Další možností je generovat potřebná napětí mimo prostor vrstev sítě a do vrstev je přivádět zvenku.

Konkrétní variantu je třeba vybrat až v okamžiku, kdy bude existovat úplné zadání pro konstrukci sítě (dimenze vstupního vektoru, maximální dimenze p , počet vrstev). Kterákoliv zmíněná varianta je bez problémů realizovatelná.

Paměť váhových koeficientů

Na paměť váhových koeficientů jsou kladeny následující požadavky:

- Typ musí být SRAM nebo NVRAM, obsah se bude operativně měnit.
- Šířka datové sběrnice je minimálně 12 bitů (je možno použít 2×8 bitů).
- Šířka adresové sběrnice je minimálně 10 bitů.
- Co nejkratší přístupová doba.
- Dostupnost.

Tyto vlastnosti splní mnoho pamětí, vzhledem k požadavku na přístupovou dobu je třeba použít paměti, které se obvykle používají jako cache. Jako příklad lze uvést typ AS7C256 firmy Alliance Semiconductor, jehož základní charakteristiky jsou podle [7] tyto:

- Typ CMOS SRAM.
- Datová sběrnice 8 bitů.
- Adresová sběrnice 15 bitů, organizace $32 \text{ kB} \times 8$ bitů.
- Přístupová doba od adresy k platným datům se liší podle podtypu a pohybuje se v rozsahu 10 až 35 ns.
- Dostupnost nebyla ověřována, paměti s podobnými vlastnostmi je na trhu větší množství typů.

Přístupová doba se přímo sečítá s ostatními časy potřebnými k ustálení analogových obvodů. Protože se analogové obvody ustalují řádově 1 μ s, je i maximální přístupová doba uvedeného typu paměti 30× kratší. Bylo by možno použít i běžné paměti CMOS (bez přívlastku „high speed“), jejichž přístupová doba se pohybuje od 70 ns výše, a to bez závažného zhoršení propustnosti sítě. Takové paměti jsou levnější a ještě dostupnější. Jednotlivé typy pamětí bývají i fyzicky záměnné (pinově kompatibilní).

Propojení paměti s výstupy komparátorů na straně adres a se vstupy D/A převodníku na straně

násobičky je třeba realizovat takovým způsobem, aby do obou sběrnic mohl zasahovat mikrokontrolér v době programování paměti. Protože tato technologie je běžně používána v číslicové technice, nemá smysl ji zde podrobněji rozebírat. Za zmínku stojí možnost zpětně mikrokontrolérem číst datový výstup paměti v ustáleném stavu (postačí např. 10 μ s) a tím zjistit konkrétní váhový koeficient pro aktuální vstupní vektor. Toho je možno využít při diagnostice chování sítě.

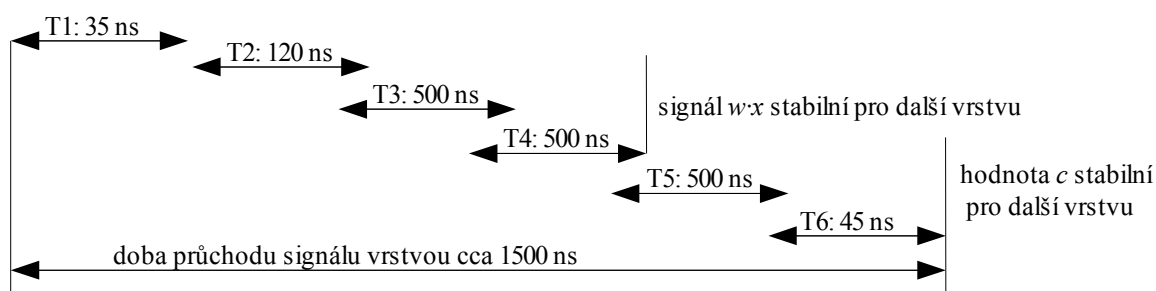
Řídicí mikrokontrolér

Volba typu mikrokontroléru závisí na potřebném počtu ovládaných signálů. Tento počet je závislý na způsobu řízení paměti váhových koeficientů a na způsobu vytváření komparačních úrovní pro komparátory. Mikrokontrolér musí mít sériový komunikační kanál, kterým může přijímat parametry vrstvy z řídicího PC, případně zpětně předávat diagnostické údaje. Přibližně je potřeba 4 nebo 5 osmibitových portů. Rychlost mikrokontroléru není kritická, jeho funkce neovlivňuje průchodnost neuronové sítě.

Protože se jedná o standardní úlohy, pro které jsou mikrokontroléry navrženy, není třeba se volbou typu více zabývat. Volba typu závisí na konkrétní realizaci hardwaru a bude dána spíše zvyklostmi konstruktéra a programátora firmwaru. Příkladem může být produkt firmy Atmel s označením T89C51CC01 (viz. [8]). Pokud by počet jeho přímo přístupných portů nestačil, je možno tento počet snadno rozšířit podle potřeby pomocí externích registrů typu latch.

Časový průběh průchodu vrstvou

Na základě popsaných částí vrstvy neuronové sítě a specifikace použitých součástek je možno nakreslit přibližné časové schéma průchodu signálu vrstvou:



T1	výběr váhového koeficientu z paměti SRAM
T2	ustálení proudového výstupu násobícího D/A převodníku
T3	ustálení převodníku I/U na výstupu násobícího D/A převodníku
T4	ustálení ofsetového zesilovače na výstupu násobícího D/A převodníku
T5	ustálení sčítačky
T6	ustálení komparátorů

Jak již bylo uvedeno, nedá se odhadnout, k jakým vzájemným časovým překryvům dochází mezi T2 a T3, T3 a T4, T4 a T5, T5 a T6. Uvedené časy jsou odečteny z katalogových údajů navržených součástek pro nejhorsí případ, tedy pro maximální skok signálu na vstupu. Ve skutečnosti se signály mění po nastavení váhy v násobičce (začátek úseku T2) spojitě a jednotlivé součástky v signálové trase mají tak více času na ustálení (po ustálení jednoho prvku je následující prvek velmi blízko ustálení, protože po celou dobu ustalování prvního prvku sledoval jeho výstup). Odhad celkového času potřebného na průchod signálu vrstvou je jen hrubý, přesto dává představu o možných vlastnostech takto konstruované sítě.

Závěr 1.části

Rozborem možného zapojení signálové trasy jedné vrstvy neuronové sítě pro zpracování jedné složky vstupního vektoru bylo zjištěno, že lze zkonstruovat elektrický obvod, který analogově-číslicovou cestou řeší výchozí zadání. Pokud se experimentálně potvrdí odhadnuté dynamické vlastnosti, bylo by možno zpracovávat tok 10^5 vektorů za sekundu sítí složenou až ze šesti vrstev.

Studie realizovatelnosti neuronové sítě daných vlastností pomocí elektrických obvodů – část 2.

Hardwarová realizace a její analýza

Na základě rozboru provedeném v 1. části studie byl realizován funkční vzorek buňky sítě NNSU. Použité součástky byly vybrány na základě výše uvedeného rozboru. Vzorek umožňuje zadávání vstupního vektoru, váhového koeficientu a komparační úrovně v digitální podobě. Výstupem buňky je jednobitová informace, která může být odečtena v přednastavené době od změny vstupního vektoru. Taková soustava umožňuje měření odezvy buňky na různě velkou změnu vstupních dat a ověření nejistoty výsledku v hraničních oblastech (hodnota srovnávaná komparátorem je blízká referenční hodnotě). Kromě toho má funkční vzorek nástroje pro vlastní kalibraci.

Realizovaný vzorek byl změřen a naměřené výsledky (zjištěné chyby realizace matematických operací) byly použity v matematických modelech složitějších sítí [9]. Z analýzy provedené v [9] vyplývá, že výsledná (složitější) síť je poměrně odolná vůči chybám jednotlivých buněk, problém je ale kumulace chyb při průchodu signálu větším počtem vrstev. Složitější soustavy (s větším počtem vrstev) dávaly horší výsledky než jednodušší soustavy se stejnou funkcí. Z toho lze usoudit, že snížení přesnosti operací s lokálním charakterem vlivu chyb je přípustné (sumace, komparace), zatímco u operací s kumulativním charakterem vlivu chyb je žádoucí chyby minimalizovat (násobení).

Upravené zadání

Na základě výsledků uvedených v [9] bylo doplněno zadání pro studii s cílem dosáhnout maximální propustnosti sítě při zachování vyhovující přesnosti.

Dimenze vstupního vektoru (tj. počet měřicích kanálů) se může pohybovat v rozsahu 16 až 32. Rozlišení složky vstupního vektoru (rozlišení měřicího kanálu) je 9 bitů (8 bitů + znaménko nebo 9 bitů unipolárně) nebo lepší. Počet komparačních úrovní se může pohybovat v rozsahu 8 až 16. Snadná dostupnost (časová a cenová) není podmínkou, postačuje dostupnost střednědobá – je možno použít i součástky, které nejsou u dodavatelů skladem a jejichž dodací doby se pohybují v řádu několika týdnů až měsíců.

Možné varianty řešení

Jestliže je kladen maximální důraz na propustnost sítě, je z 1. části studie zřejmé, že jednotlivé vrstvy sítě musí pracovat samostatně. V opačném případě by doba ustalování signálů při postupném průchodu vrstvami sítě snižovala propustnost sítě. S rostoucím počtem vrstev by se doba ustalování prodlužovala přímo úměrně. Řešením je oddělit činnost probíhající v jednotlivých vrstvách do samostatných navazujících časových úseků, tedy struktura pipeline. Tak bude rychlost sítě dána rychlostí zpracování signálu v jedné vrstvě bez ohledu na počet vrstev v síti.

Řešení takové sítě metodou zpracování analogového signálu, jak bylo popsáno v 1. části studie, by znamenalo použít na rozhraní vrstev obvody typu sample&hold. Tyto obvody jsou poměrně pomalé a navíc zanášejí do signálu postupujícího vrstvami další chybu (vstupní napětíovou nesymetrii a chybu zesílení).

Násobička

Možností, jak se vyhnout kumulaci chyb při průchodu signálu vrstvami a zajistit strukturu pipeline, je realizace násobičky digitální cestou. Obecně je digitální násobička zařízení, které

realizuje zobrazení $V \times V \rightarrow V$, kde V je množina všech možných čísel z rozsahu, se kterým zařízení pracuje. Pro případ šířky dat 9 bitů se jedná o stavový automat, který má 2^{18} možných stavů. Násobička v buňce NNSU však z této množiny stavů využívá jen některé. Realizuje zobrazení $U \times V \rightarrow V$, kde U je množina všech váhových koeficientů w_{jk} pro danou vrstvu j a daný kanál k . Tato množina může být například šestnáctiprvková, tedy stavový automat bude mít $16 \times 2^9 = 8192$ možných stavů. Takový stavový automat je možno realizovat tabulkou umístěnou v paměti RAM, kde je možno měnit obsah tabulky podle požadovaných hodnot koeficientů w_{jk} . Malý rozsah tabulky dovoluje použít extrémně rychlé asynchronní paměti (high speed SRAM) s přístupovou dobou až 8 ns od zadání adresy do dodání dat (parametr t_{AA}). Kromě rychlosti realizace operace násobení má použití digitální násobičky výhodu v možnosti zapamatovat si výsledek násobení v digitální podobě jednoduchým obvodem typu latch na výstupu vrstvy. Také zvýšení přesnosti (šířka sběrnice) není problém, jde o využití více bitů z šestnáctibitové sběrnice paměti, kterou je nutno použít pro požadovaných 9 bitů jako minimální (šířka sběrnice je násobek 8 bitů).

Omezení v použití paměti SRAM jako rychlé násobičky spočívá v maximálně vyráběných kapacitách. Lze použít pouze jednu paměť nebo dvě paměti paralelně (rozšíření sběrnice). Při pokusu použít více pamětí za účelem zvýšení kapacity (počtu adresovatelných slov) vyvstane požadavek na demultiplex horních adresních vodičů, což snižuje průchodnost paměti (demultiplex a ovládání signálů chip-enable) a zvyšuje složitost řídicích obvodů. Větší zastavěná plocha má za následek větší kapacity sběrnic. Více paralelně zapojených adresních a datových vodičů má za následek větší kapacitu. Nutnost budít tyto kapacity by si vynutila použití přídavných budičů, které opět způsobí snížení propustnosti sítě.

Prostudováním nabídky různých výrobců high speed SRAM je možno omezit použitelnou kapacitu na 128 kB. Větší paměti jsou již realizovány jako synchronní a nelze je tedy pro funkci stavového automatu použít (jejich konstrukce je podřízena práci s datovými balíky – sektory HDD apod.). K dispozici je 17 adresních vodičů, které je třeba rozdělit na část pro vstup složky procházejícího vektoru a na část pro výběr váhy. Pokud se použije pro adresaci váhy metoda lineární adresace (viz. 1. část studie), je možno ve vrstvě použít maximálně $17 - 9 = 8$ komparátorů (9 bitů je šířka procházející složky vektoru x). Pokud bude výstup komparátorů předchozí vrstvy binárně kódován, mohlo by teoreticky být komparátorů až 255 (256 intervalů). Omezení počtu intervalů pro komparátory např. na 32 dovolí rozšířit sběrnici na 12 bitů, čímž lze dále snížit chybu (v tomto případě zaokrouhlovací) složky vstupního vektoru, která prochází jednotlivými vrstvami.

Sčítačka

Řešení sčítačky digitální cestou je pro realizaci NNSU nevhodné. Signály, které se mají sečíst, by se musely přivést ze všech kanálů (v počtu 16 až 32 podle zadání) do jednoho místa na vstupy sčítačky. Znamenalo by to vést napříč vrstvou minimálně $16 \times 9 = 144$, maximálně $32 \times 9 = 288$ datových vodičů. Tyto signály by se musely paralelně zpracovat 16ti až 32ti-kanálovou sčítačkou. Její realizace by byla velmi složitá a muselo by se zřejmě použít několik spolupracujících hradlových polí. Mechanická složitost soustavy by měla za následek větší náchylnost k poruchám a v neposlední řadě by byla náchylná k rušení. Pravděpodobnost rušení vodiče délky několika decimetrů, kterým prochází číslicový signál s frekvencí řádově desítky megahertzů, je poměrně vysoká. Pokud by se jednalo o MSB jednoho z kanálů, mohlo by to znamenat zcela chybné vyhodnocení vstupního vektoru jako celku.

Schůdnější a spolehlivější řešení je použít analogovou sčítačku. V každém kanálu je nejprve nutno převést digitální signál z násobičky na analogovou hodnotu. Zde je třeba uvážit, jakou přesnost musí mít D/A převodník. Analogové signály z minimálně šestnácti kanálů se slučují do jednoho a ten je poté porovnán s množinou intervalů, aby se vybral požadovaný váhový vektor pro další vrstvu. To znamená, že signál přicházející z každého kanálu se uplatňuje v součtu nejvýše jednou šestnáctinou. Při požadovaném rozlišení celé soustavy 9 bitů tedy stačí přesnost D/A převodu každého z kanálů 5 bitů. To velmi usnadňuje výběr D/A převodníku. Je k dispozici D/A

převodník s rozlišením 8 bitů, jehož doba převodu je 6 ns při přesnosti 7 bitů, což je přesnost pro danou aplikaci zcela vyhovující.

Komplikovanější je vlastní sečtení analogových hodnot. Kdyby se jednalo o „pomalé“ signály (řádově 100 ns a více), stačilo by sečíst proudové výstupy D/A převodníků jednotlivých kanálů v uzlu a výsledný proud pomocí rychlého operačního zesilovače převést na napětí pro komparátory podobně jako v 1. části studie. Zde se však situace komplikuje tím, že pracujeme se signály, které vznikají v různých místech v prostoru (krajní kanály budou od sebe vzdáleny několik decimetrů) a jsou velmi rychlé. Běžný vodič se vzhledem k takovým signálům chová na tuto vzdálenost jako indukčnost, která by způsobovala nepřipustné překmity na výstupech D/A převodníků a možná i jejich rozkmitání. Pro přenos signálu s takovým spektrem se používají vedení s definovanou impedancí a disperzními vlastnostmi. Mohou to být např. pásková vedení realizovaná na plošném spoji nebo koaxiální vedení. Typická charakteristická impedance takových vedení bývá 50 ohmů. Signál se jimi šíří jako elektromagnetická vlna a vzdálenost, na kterou lze signál přenášet, je z hlediska konstrukce NNSU neomezená. Je však třeba zajistit, aby signály ze všech kanálů dorazily do sumačního bodu přibližně ve stejnou dobu (stejná délka vedení od jednotlivých kanálů). Vzhledem k tomu, že rychlost signálu ve vedení je typicky kolem 0,8 c (v závislosti na relativní permitivitě dielektrika), činí zpoždění signálu cca 0,27 ns na každý decimetr vedení. Při „rozumných“ rozměrech soustavy nebude tedy hrát zpoždění roli.

Přítomnost vedení v signálové cestě vylučuje možnost sečítání proudů jednotlivých D/A převodníků v uzlu přímo, jak by se to dalo realizovat u „pomalých“ signálů. Proudový výstup D/A převodníku je nutno nejprve převést na napětí o vnitřní impedanci shodné s charakteristickou impedancí vedení. Signál procházející vedením musí navazovat na sumační síť, která bude vůči každému z vedení od jednotlivých kanálů vystupovat opět jako charakteristická impedance vedení. Tím se zabrání odrazům na vedení a tím překmitům na výstupech D/A převodníků. Zmíněný D/A převodník takové zapojení výstupu umožňuje, poskytuje napěťový signál v rozsahu 0 až 1V. Konstrukce impedance vyvážené sčítačky závisí na konkrétním počtu měřicích kanálů, jedná se o pasivní odporovou síť.

Komparátory

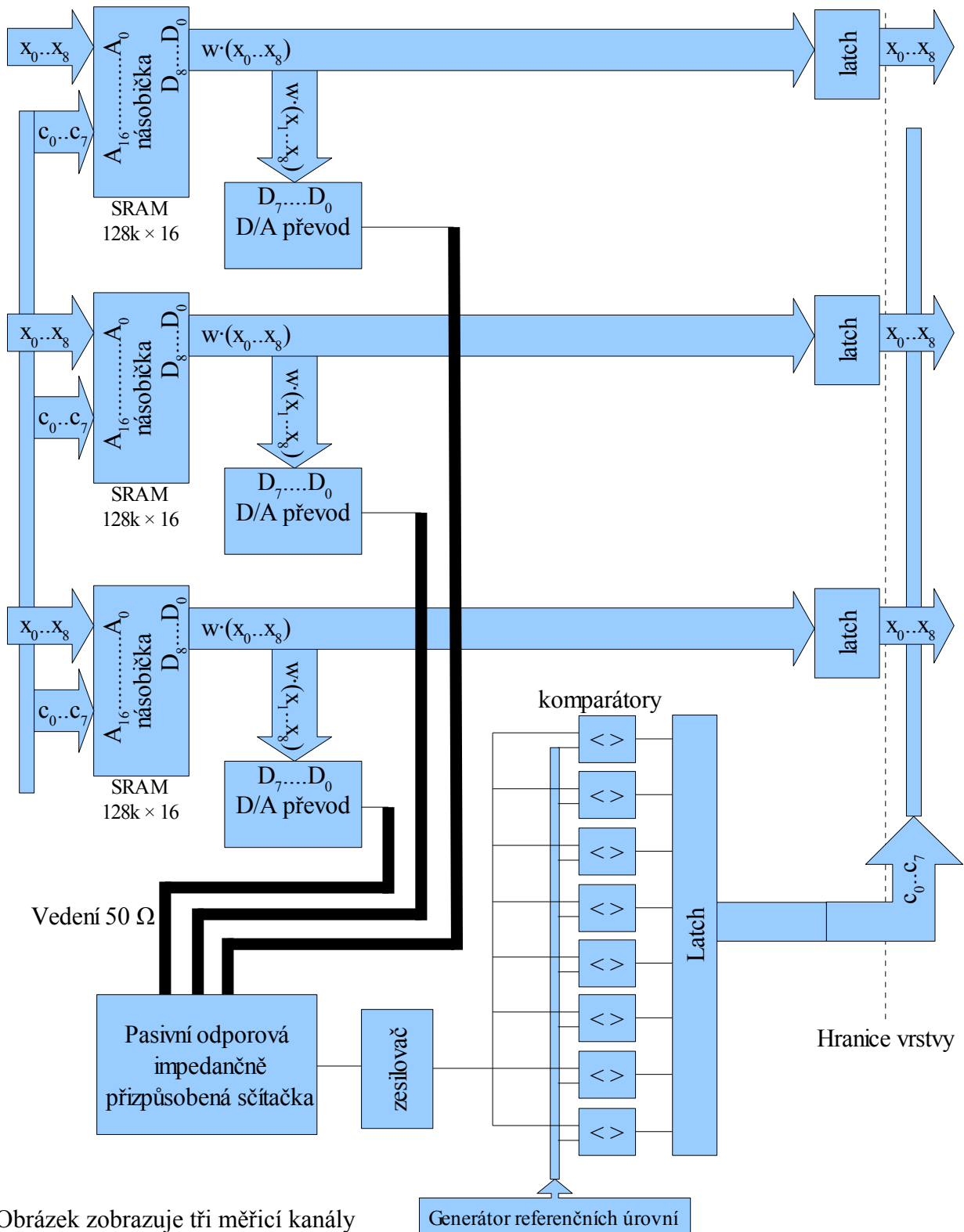
Komparátory pracující s odezvou ve srovnatelné časové oblasti (jednotky až desítky nanosekund) zpracovávají signál o velikosti maximálně jednotky voltů. Vstupní napěťová nesymetrie se pohybuje v jednotkách mV. Aby došlo k ustálení výstupu komparátoru v dostatečně krátké době, musí rozdíl napětí mezi vstupy přesáhnout cca 5 mV. Pro rozlišení cca 9 bitů by musel být signál na vstupu komparátoru zesílen na cca 2,5 V. Nejvhodnější nalezený zesilovač má maximální výstupní napětí 1,4 V, což odpovídá rozlišení komparátoru cca 8 bitů. Vzhledem k tomu, že takto vzniklá chyba má jen lokální charakter (nekumuluje se) a ovlivňuje pouze přesnou polohu hranice intervalu, lze snížení přesnosti z devíti na osm bitů považovat za přijatelné.

Podle počtu komparačních hladin je signál zatížen paralelně zapojenými vstupy několika komparátorů. Při použití metody lineární adresace váhových koeficientů je možno uvažovat 8 paralelně zapojených vstupů. Vstupní kapacita jednoho vstupu komparátoru je 3 pF. Při standardní (výrobcem doporučené) impedanci výstupu zesilovače 50 ohmů prochází signál RC členem s časovou konstantou $RC = 50 \times 24 \times 10^{-12} = 1,2$ ns. Na dosažení úrovně s maximální odchylkou pod rozlišovací schopností komparátoru (0,5 LSB při rozlišení 8 bitů) je třeba 6,2 násobek časové konstanty, tedy cca 7,5 ns.

Výstupy komparátorů musí být od další vrstvy odděleny budičem sběrnice s funkcí latch ze tří důvodů. Je třeba zafixovat po ustálení komparátorů stav jejich výstupů po celou dobu zpracování signálu následující vrstvou, zatímco aktuální vrstva již zpracovává nový vektor. Druhý důvod je nutnost distribuovat stav výstupů komparátorů do všech násobiček následující vrstvy, tedy budit dlouhou sběrnici. Samotné výstupy komparátorů nejsou konstruovány na buzení sběrnice. Třetí důvod je nutnost uvolnit sběrnici (odpojit budiče) v době, kdy dochází k programování obsahu

násobiček následující vrstvy. V této době řídí sběrnici pomocný mikrokontrolér, který na základě váhových koeficientů dodaných z řídicího PC vypočítává tabulky násobiček a plní hodnotami odpovídající paměti.

Blokové schéma vrstvy s použitím komparátorů



Obrázek zobrazuje tři měřicí kanály a jejich napojení na sčítačku a komparátory.

Označení procházejícího vektoru je pro jednoduchost shodné pro všechny tři kanály.

Výběr součástek

Při výběru součástek je třeba nejprve zvolit napájecí soustavu, která určuje úroveň číslicových signálů. Rozborem vlastností velkého množství typů součástek vychází nejlépe napájecí napětí 3,3 V, pro které je možno vytvořit z nabídky největší počet použitelných kombinací a současně sestavit strukturu s maximální rychlostí. Výjimku tvoří komparátor, který je napájen úrovní 5 V (přesto je kompatibilní s úrovní LV CMOS).

Paměť SRAM

Vhodné paměti pro digitální násobičku byly vybírány z produkce firem Amic, Utron, ISSI, TM tech a Winbond. Přístupové doby se pohybují od 8 do 12 ns (vybavení dat po změně adresy, parametr t_{AA}). Bylo nalezeno asi 10 typů s uvedenou přístupovou dobou a napájením 3,3 V. Paměti se liší kapacitou (32 kB, 64 kB, 128kB), šířkou datové sběrnice (8 nebo 16 bitů) a spotřebou. Největší kapacitu má paměť IS61LV12816 firmy ISSI [10]. Organizace paměti je 128k × 16 bitů, přístupová doba t_{AA} je 8 ns, odběr z napájecího zdroje 65 mA. Tato paměť je uvažována v násobičce popsané výše.

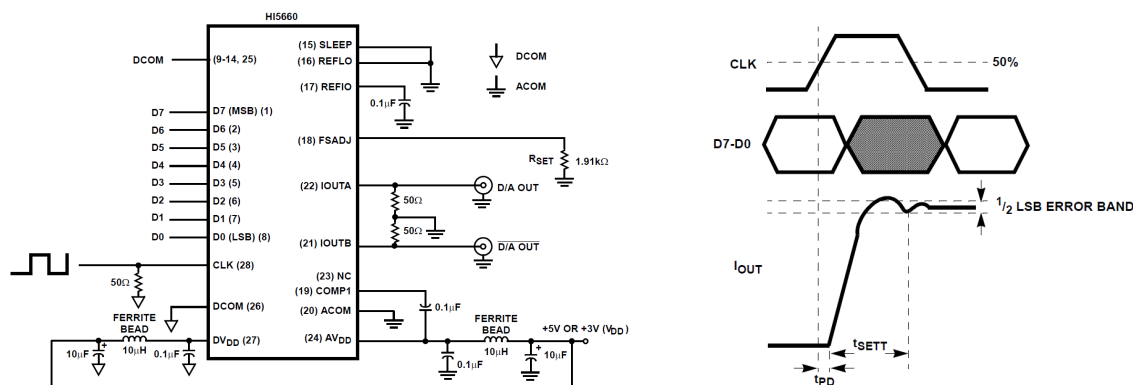
Latch

Požadavky na vlastnosti latche jsou kromě rychlosti a napájení transparentnost, odpojitelnost a schopnost budit sběrnici. Transparentnost znamená, že v průchozím režimu jsou data trvale kopírována na výstup, nedochází k přepsu pouze hranou, jako je tomu u klopných obvodů typu D. Odpojitelnost je schopnost výstupů přejít do stavu s vysokou impedancí. To je potřeba při programování paměti SRAM, kdy je jejich adresová i datová sběrnice ovládána pomocným mikrokontrolérem. Schopnost budit sběrnici je důležitá zejména u latche zachycujícího stav výstupů komparátorů, neboť tento signál je distribuován všem kanálům následující vrstvy.

Vhodné funkční typy jsou obvody 74*573 a 74*373. Na základě prostudování vlastností několika zástupců těchto obvodů od různých výrobců byl jako nejvhodnější vybrán obvod 74LVTH373 firmy Fairchild Semiconductor TM [11]. Ten splňuje všechny požadavky a navíc má interně ošetřeny nezapojené bity, což se s výhodou využije, neboť je pro sběrnici třeba použít dva obvody (2 × 8 bitů), ale obsazeno je pouze 9 bitů. Doba průchodu signálu v transparentním režimu je cca 5 ns (záleží na kapacitě sběrnice).

D/A převodník

U D/A převodníku jsou kladeny požadavky na rychlost při rozlišení nejméně 5 bitů, kompatibilitu řídicích vstupů se zvolenou LV logikou a schopnost dodávat na výstupu napěťový signál bez potřeby zpětnovazebního operačního zesilovače (požadavek na rychlost). Z nabídky známých výrobců nejlépe vyhověl obvod HI5660 firmy Intersil [12]. Jeho rozlišení je 8 bitů a doba ustálení na přesnost 7 bitů je 6 ns. Dovolенý rozkmit napětí na výstupu je 0 až 1,25 V.



Výstup převodníku je tvořen dvěma komplementárními proudovými zdroji, jejichž součet proudů je nastavitelný v rozmezí 2 až 20 mA. Proud výstupu A je úměrný binární hodnotě na

vstupech D0 až D7. Jak je patrné z časového diagramu, stav na vstupech se přepisuje do převodního jádra s nástupnou hranou hodinového signálu. Vhodný čas pro přepis je dán dobou ustálení dat na výstupu násobičky a musí být určen pomocným časovacím obvodem (časovým dispečerem vrstvy), který není v blokovém schématu uveden. Čas ustálení se skládá z fixního času t_{PD} (typicky 1 ns) a času t_{SETT} , který závisí na požadované kvalitě ustálení signálu. Pro chybu ± 1 LSB (rozlišení 7 bitů) je tento čas 5 ns, pro rozlišení $\pm 0,5$ LSB (rozlišení 8 bitů) je tento čas 15 ns. Vzhledem k požadované přesnosti 5 bitů je možno počítat s dobou ustálení celkem 6 ns.

Výstupní proud 0..20 mA teče do zatěžovacího odporu 50Ω , čímž vniká napěťový signál v rozsahu 0..1 V. Signál je propojen se vstupem sčítačky vedením s charakteristickou impedancí 50Ω . Je-li charakteristická impedance vedení stejná jako zatěžovací odpor, dojde k přenosu energie z/do vedení bez zámity napětí. Vedením se eliminuje problém s kapacitami a indukčnostmi fyzického zapojení obvodu – indukčnosti a kapacity vodičů dlouhé signálové trasy jsou u vedení v prostoru rovnoměrně rozloženy a chovají se vůči signálu jako reálný odpor, pouze dojde ke zpoždění signálu při průchodu vedením.

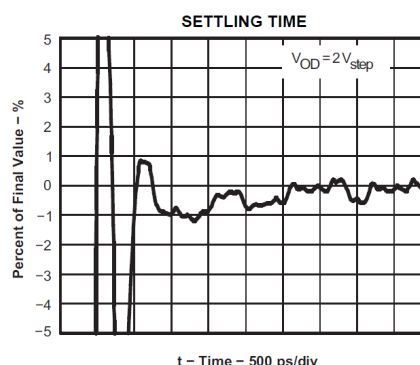
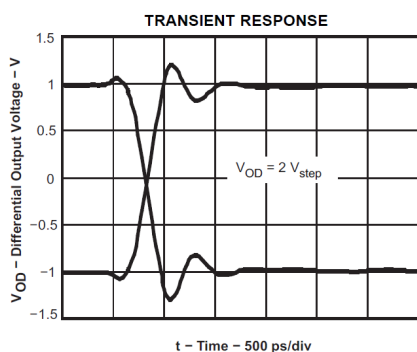
Impedančně vyvážená sčítačka

Sčítačka musí mít vstupní impedanci každého vstupu rovnu 50Ω . Konkrétní hodnoty odporů závisí na počtu vyhodnocovaných kanálů, protože impedance sčítacího uzlu, do kterého vede každý ze sčítacích odporů, je tvořena paralelní kombinací ostatních kanálů se započtením výstupních impedancí odpovídajících zdrojů signálu (zatěžovací impedance proudových výstupů D/A převodníků). Například pro 16 kanálů by hodnota sčítacích odporů byla přibližně 44Ω .

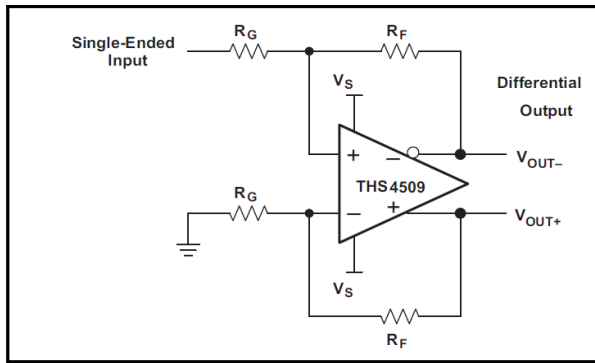
Výstup sčítačky není možno připojit přímo k následujícímu zesilovači. Jak bylo uvedeno výše, potřebné zesílení je 1,4. Protože zvolený zesilovač dovoluje zesílení minimálně 2, je nutno signál nejprve zmenšit, aby by zesilovač pracoval ve svém pásmu stability. Odporová sčítačka bude tedy o něco složitější.

Zesilovač

Zesilovače pracující v požadované frekvenční oblasti jsou konstruovány jako diferenciální. Základním požadavkem při výběru vhodného typu je rychlost ustálení výstupu a vhodný rozsah signálu. Nejvhodnějším typem se jeví výrobek firmy Texas Instruments Inc. s typovým označením THS4509 [13]. Rozsah signálu je typicky $\pm 1,4$ V. Zesilovač vyniká vysokou rychlostí přeběhu 6600 V/ μ s, která umožňuje dobu ustálení na 1% za 1 ns a na 0,1% za 10 ns při napěťovém skoku 2 V. Interpolací s ohledem na grafy lze usoudit, že zesilovač bude ustálen na požadovanou přesnost za cca 5 ns.



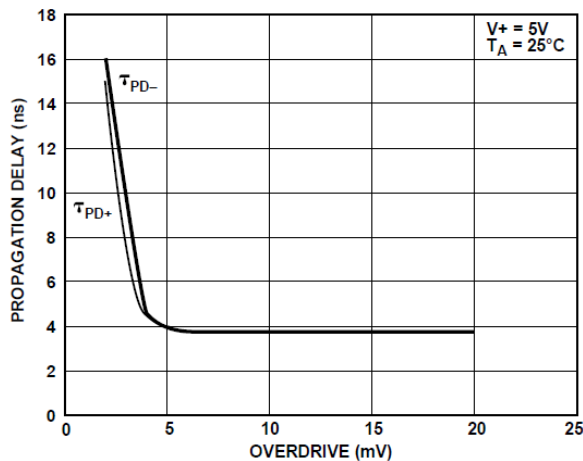
Bohužel nelze vhodným způsobem využít diferenciální signál zesilovače pro zvýšení použitelného rozkmitu. Pro takové zapojení by se musela použít transformátorová vazba, která však neumožňuje přenos DC složky signálu. Komplementární větev zesilovače je nutno vhodně impedančně zakončit a signál procházející touto větví zůstane bez užitku. Zapojení zesilovače pro nesymetrický signál je uvedeno na následujícím obrázku.



Single-Ended Input to Differential Output Amplifier

Komparátory

Při volbě komparátorů je kromě kompatibility výstupů s úrovněmi LV CMOS základním požadavkem rychlost odezvy na malý rozdíl napětí mezi vstupy (rozlišovací schopnost) a celkový pracovní rozsah vstupních napětí. Jako nejvhodnější byl zvolen výrobek firmy Analog Devices AD8611 [14]. Rozsah vstupních napětí je 0..3 V, vstupní kapacita je 3 pF. Typická vstupní napěťová nesymetrie je 1 mV. Odezvu na změnu napětí na vstupech je možno odečíst z grafu.



Při rozvážení vstupů o 5 mV dojde k ustálení za 4 ns. Z grafu je zřejmé, že při menších rozdílech napětí mezi vstupy se doba nastavení výstupu silně prodlužuje. To omezuje rozlišovací schopnost komparátoru na 8 bitů při úrovni signálu 1,4 V, jak bylo uvedeno výše. Výstupy komparátorů je třeba oddělit obvodem typu latch, lze použít stejný typ jako u sběrnice přenášející hodnotu $w \cdot x$. Komparátory se vyrábějí i ve dvojicích v jednom pouzdře pod označením AD 6812 [14].

Časový diagram vrstvy s použitím komparátorů

V tabulce jsou uvedeny časy potřebné na průchod signálu nejdelší trasou.

Funkční blok	Doba průchodu signálu	Komentář
násobička	8 ns	Doba ustálení výstupů paměti obsahující tabulku součinnů vah se všemi možnými kombinacemi vstupního vektoru
D/A převodník	6 ns	Doba ustálení výstupního proudu převodníku od signálu pro provedení převodu
vedení ke sčítačce	1,1 ns	Průchod signálu vedením o délce 4 dm, předpokládá se symetrické umístění sčítačky mezi kanály (odhad vycházející z předpokládané velikosti systému)
sčítačka	0 ns	Pasivní odporová sčítačka nezanese znatelné zpoždění
zesilovač	5 ns	Doba ustálení výstupu zesilovače na požadovanou přesnost.
komparátory	6,2 ns + 4 ns	Doba potřebná na nabití vstupní kapacity bloku komparátorů a doba ustálení komparátorů s požadovaným rozlišením
budič sběrnice	5 ns	Doba průchodu signálu budičem / latchem
distribuce výstupních signálů komparátorů	1,1 ns	Doba průchodu signálu od budiče směrem ke vstupům následující vrstvy v délce 4 dm na obě strany od budiče
rezerva	2 ns	Nejistota odhadnutých časů (hodnota může být poddimenzována)
Celková doba průchodu signálu	38,4 ns	
Průchodnost	26 MSPS	Teoretická rychlost zpracování signálů (megasamples per second)

Varianta s náhradou komparátorů A/D převodníkem

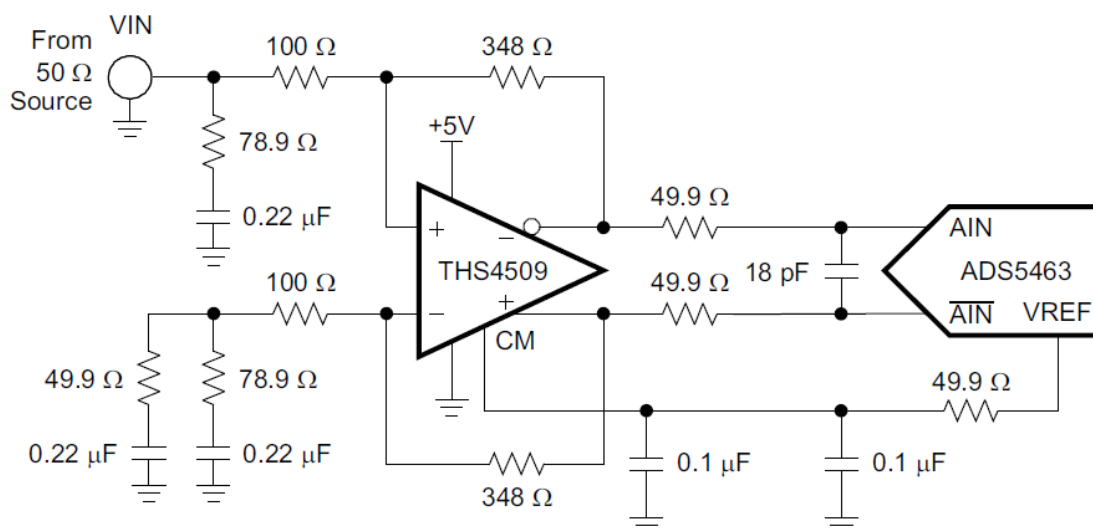
Výše popsané řešení má omezení v maximálním počet komparátorů. Pokud by se měl použít větší počet než osm, znamenalo by to nutnost použít demultiplexer z lineárního adresování na binární. Tato operace se dá realizovat pomocí hradlového pole za několik nanosekund, zvýší se však vstupní kapacita bloku komparátorů a tím se citelně prodlouží doba ustalování napětí na tomto společném vstupu. Kromě vlastní kapacity vstupů komparátorů přibude i kapacita propojovací cesty mezi vstupy. Celkové rozměry bloku pak ještě narostou o další generátory komparačních úrovní řízené mikrokontrolérem.

Od určitého počtu komparátorů se jeví jako výhodnější použít jinou metodu rozdělení signálu podle příslušnosti do daného intervalu. Analogové dělení pomocí bloku komparátorů je možno nahradit digitálním dělením. Výstup sčítačky je možno digitalizovat A/D převodníkem a výstupní data pomocí tabulky převést na informaci stejného druhu, jako je signál na výstupu komparátorů. Zde je přirozené použít binární kódování, neboť nabízí téměř neomezený počet komparačních úrovní (při osmibitové sběrnici až 255 úrovní).

Nejprve je však třeba nalézt dostatečně rychlý A/D převodník, který by spolu s pamětí obsahující

tabulku komparačních úrovní mohl konkurovat bloku komparátorů. Nabízí se použití A/D převodníku typu flash, který provádí A/D převod paralelně. Takový převodník obsahuje v integrované formě také blok komparátorů, pro každou úroveň o velikosti LSB jeden. Z tohoto důvodu se vyrábí A/D převodníky flash s rozlišením maximálně 8 bitů, což je pro danou aplikaci nedostatečné. Existují však převodníky, které provádějí převod v několika krocích a jejich vnitřní struktura pracuje jako pipeline. Převodníky dosahují vysokých rychlostí při velmi dobrém rozlišení, data jsou však proti signálu zpožděna o mnoho hodinových taktů (cca 10 taktů a více). Takové převodníky se hodí zejména pro zpracování spojitého signálu, jako je digitalizace obrazu, analýza radarového příjmu, digitální osciloskopy apod. Pro účely diskontinuálního provozu jako je analýza signálu sčítačky uvnitř buňky NNSU jsou nevhodné, protože nelze využít vysoké rychlosti dosahované díky struktuře pipeline.

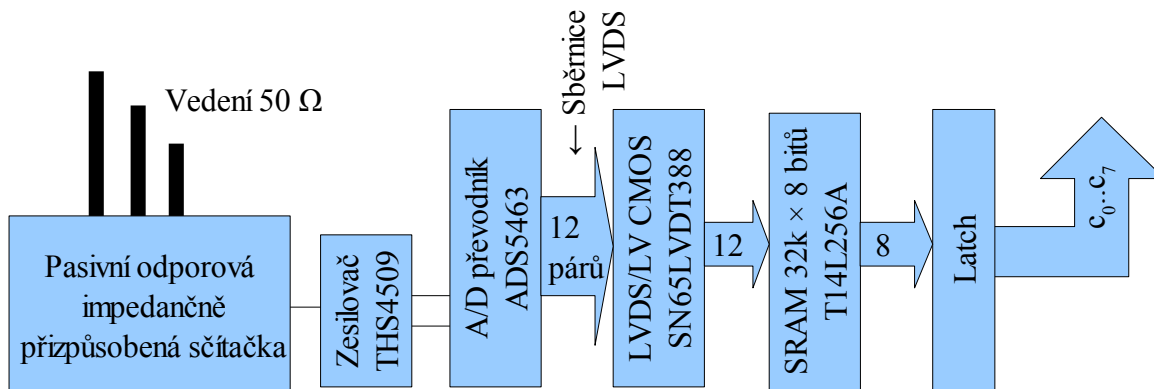
Výjimkou je A/D převodník firmy Texas Instruments Inc. s typovým označením ADS5463 [15], který má dobu latence (doba mezi vzorkem vstupního analogového napětí a jemu odpovídajícími platnými výstupními daty) pouhý 3,5-násobek hodinového taktu, který může mít periodu až 2 ns. Rozlišovací schopnost je 12 bitů. Doba převodu je tedy 7 ns, což je srovnatelné s blokem komparátorů. Převodník má diferenciální vstup, který musí mít střední potenciál na hladině 2,4 V. Takové přizpůsobení signálů je schopen zajistit zesilovač THS4509.



Na obrázku je znázorněna metoda propojení zesilovače a A/D převodníku. Konkrétní hodnoty odporů by musely být přizpůsobeny požadavkům na zesílení soustavy.

Nepříjemnou vlastností A/D převodníku jsou jeho diferenciální logické výstupy s logickou úrovní LVDS, kterou je nutno převést na úroveň LV CMOS, se kterou pracuje zbytek soustavy. Tento převod uskuteční dvojice převodníků SN65LVDT388 za cca 3 ns [16]. Převodníky mají signál pro přechod výstupů do stavu s vysokou impedancí (chip enable), takže dovolují odpojení následujících obvodů. Tato funkce je nezbytná při programování obsahu konverzní RAM.

Výstupy A/D převodníku (po úpravě úrovně) je třeba překonvertovat na informaci o příslušnosti k intervalu pomocí paměti SRAM s konverzní tabulkou. Paměť musí mít aspoň 12 adresních vodičů a 8 datových vodičů. Je možno použít např. paměť firmy TM Tech s typovým označením T14M256A [17] s architekturou $32k \times 8$ bitů s přístupovou dobou (t_{AA}) cca 8 ns. Datový výstup paměti již poskytuje stejný typ informací jako výstup bloku komparátorů a může být přiveden do výstupního budiče (latche). Následující obrázek ukazuje náhradu bloku komparátorů od sčítačky po latch.



Po vstupu zpracovávaného vektoru x do vrstvy je třeba spustit hodinový signál 500 MHz pro A/D převodník. V okamžiku, kdy je signál na výstupu zesilovače ustálen, odpočítá se ještě 7 půlperiod hodinového signálu a hodiny se zastaví. Všechny následující obvody jsou transparentní, takže výsledné zařazení signálu do intervalu se objeví po průchodu obvody na sběrnici údaje c . Po uplynutí doby potřebné na ustálení stavu na sběrnici je možno uzamknout latch a spustit zpracování dalšího vektoru x . Pokud se bude rychlost průchodu vektoru x blížit maximální propustnosti sítě, bude vhodnější synchronizovat chod hodin pro A/D převodník s hranami spouštějícími zpracování vektoru (např. fázovým závěsem). Pokud bude časová rezerva, stačí použít autonomní oscilátor 500 MHz a jeho výstup klíčovat impulzem potřebné šířky spouštěným od hran spouštějících zpracování vektoru x . Toto klíčovací hradlo nemůže být v logice LV CMOS, která nemá takovou mezní frekvenci. Je třeba použít hradlo např. z logiky ECL.

Časový diagram vrstvy s použitím A/D převodníku místo komparátorů

V tabulce jsou uvedeny časy potřebné na průchod signálu nejdelší trasou.

Funkční blok	Doba průchodu signálu	Komentář
násobička	8 ns	Doba ustálení výstupů paměti obsahující tabulku součinitelů vah se všemi možnými kombinacemi vstupního vektoru
D/A převodník	6 ns	Doba ustálení výstupního proudu převodníku od signálu pro provedení převodu
vedení ke sčítačce	1,1 ns	Průchod signálu vedením o délce 4 dm, předpokládá se symetrické umístění sčítačky mezi kanály (odhad vycházející z předpokládané velikosti systému)
sčítačka	0 ns	Pasivní odporová sčítačka nezanese znatelné zpoždění
zesilovač	5 ns	Doba ustálení výstupu zesilovače na požadovanou přesnost.
A/D převodník	7 ns	Doba mezi vzorkem vstupního napětí a umístěním převedené hodnoty tohoto vzorku na sběrnici (doba latence).
konverze sběrnice	3 ns	Převod napěťové úrovně LVDS na LV CMOS
paměť SRAM	8 ns	Paměť komparačních úrovní (rozdělení do intervalů)
budič sběrnice	5 ns	Doba průchodu signálu budičem / latchem
distribuce výstupních signálů komparátorů	1,1 ns	Doba průchodu signálu od budiče směrem ke vstupům následující vrstvy v délce 4 dm na obě strany od budiče

rezerva	2 ns	Nejistota odhadnutých časů (hodnota může být poddimenzována)
Celková doba průchodu signálu	46,2 ns	
Průchodnost	21 MSPS	Teoretická rychlost zpracování signálů (megasamples per second)

Normování signálu

Vzhledem k použití obvodů z kategorie „high speed“ je napěťová úroveň signálu vždy kladná. Nelze použít polaritu jako indikátor znaménka složky vstupního vektoru. Složka vstupního vektoru může nabývat hodnot v rozsahu 0 až 1FFh, je proto přirozené položit signálovou nulu do středu intervalu, tedy 100h. Analogová reprezentace na výstupu D/A převodníku bude mít signálovou nulu položenu na úrovni 0,5 V, čili uprostřed pracovního rozsahu. Analogový výpočet zahrnující sumaci a komparaci bude pracovat s tímto offsetem korektně, podmínkou je stejný offset ve všech kanálech. Vzhledem k identické konstrukci jednotlivých kanálů je tato podmínka splněna. Offset signálu na výstupu diferenciálního zesilovače je u varianty s A/D převodníkem 2,4 V, A/D převodník prezentuje tuto hladinu jako střed pracovního rozsahu s digitalizovanou hodnotou 800h. Tato hodnota tedy odpovídá signálové nule sčítačky, čili napětí 0,5 V. Rozsah výstupu 0 až FFFh A/D převodníku odpovídá plnému rozsahu zpracovávaných bipolárních kanálů.

Pomocné obvody

Správa obsahu paměti

Jak paměť násobičky, tak případně paměť komparačních úrovní může být oddělena od signálových tras buňky NNSU uvedením odpovídajících budičů do stavu s vysokou impedancí. Nad adresovou a datovou sběrnicí převezme řízení mikrokontrolér, který paměť naplní požadovaným obsahem. Lze očekávat, že naplnění paměti může trvat několik desítek sekund (velmi hrubý odhad za předpokladu, že každý kanál každé vrstvy bude mít svůj mikrokontrolér). Po naplnění paměti ponechá mikrokontrolér zápisový signál „write enable“ v neaktivním stavu. Obsah paměti zůstane zachován do doby, než zanikne napájecí napětí. Pokud to bude požadováno, je možno obsah paměti zálohovat pomocným zdrojem.

Správa komparačních úrovní

Při použití bloku komparátorů je třeba pro každý komparátor vytvořit komparační úroveň. To je možno realizovat pomocí D/A převodníku se sériovou sběrnicí (SPI, I2C, ...). Potřebný větší počet převodníků tak zabere stále stejný (a malý) počet výstupů mikrokontroléru. Při vhodném výběru sběrnice ji může mikrokontrolér hardwarově podporovat, takže nastavení komparačních úrovní bude velmi rychlé. K dispozici je velké množství D/A převodníků s požadovanými vlastnostmi, výběr závisí na zvoleném typu sériové sběrnice.

Časování sítě NNSU

Časování sítě závisí na vlastnostech zdrojů dat v jednotlivých kanálech. Pokud by zdroje dat byly asynchronní a nebyl by kladen požadavek na jejich synchronizaci, mohl by být zdroj časovacích impulsů pro NNSU tvořen volnoběžným generátorem s frekvencí odpovídající požadované rychlosti zpracování vstupních vektorů. Od hran tohoto generátoru by se odvodily zpožděné hrany ovládající povel k D/A konverzi, A/D konverzi (v případě použití A/D převodníku) a krátký impuls pro uvedení latchů do transparentního režimu pro přepis dat ze vstupů na výstupy latche. Transparentnost latchů není za běžného provozu využita, význam má pouze při ožívování hardwaru ve výrobě.

Pokud jsou data v jednotlivých kanálech synchronní díky externě řízenému měření nebo díky periodičnosti zkoumaného děje, musí dojít nejprve k časovému sesazení jednotlivých kanálů do jednoho časového bodu, od kterého se odvodí hrana spouštějící síť NNSU. Od této hrany se pak odvodí vnitřní řízení sítě jako v předchozím odstavci. Jednotlivé operace sítě nemusejí být nutně ekvidistantní, podmínkou pro správnou funkci je dodržení minimálního odstupu mezi operacemi daného propustností sítě.

Digitalizace analogových kanálů

Pokud je k dispozici pouze analogový signál, je třeba ho před vstupem do sítě digitalizovat. Jako možný A/D převodník lze použít ADS5463 popsaný jako náhrada bloku komparátorů. Předřazený zesilovač THS4509 může mít zesílení v rozsahu 2 až 10, takže velikost analogového signálu lze před digitalizací upravit. Zesilovač navíc dokáže zpracovat jednoduchý (single-ended) i diferenciální signál. Zajímavou možností je umístit zesilovač a A/D převodník přímo do místa vzniku signálu (minimální rušení signálu) a výstupy přivést k síti NNSU v digitální podobě. Zmíněné výstupy LVDS jsou totiž určeny pro přenos dat na větší vzdálenosti díky své odolnosti vůči rušení. Na vstupu sítě by se signály převedly převodníkem SN65LVDT388.

Kromě těchto obvodů je možno použít relativně velké množství jiných obvodů – rychlostí 20 až 50 MSPS jsou schopny pracovat vytypované obvody většiny firem, které se specializují na analogovou techniku.

Přenos původního vektoru sítě NNSU

Základní funkcí sítě je rozhodovat o vlastnostech vstupního vektoru. Při průchodu sítě se mezi vrstvami převádějí násobky tohoto vektoru po složkách. Na konci sítě je však třeba mít k dispozici původní vstupní vektor, aby mohl na základě výsledku výpočtu uvnitř sítě postoupit k dalšímu zpracování.

Z konstrukčního hlediska se zdá nejvýhodnější vytvořit v každém kanálu ještě paralelní kanál, kterým se bude síť šířit původní složka vstupního vektoru. Tento paralelní kanál by se skládal pouze z latche, který by byl ovládán stejným signálem jako latch pro $w \cdot x$ nebo hodnotu c . Výhodou integrace paralelního kanálu přímo ke kanálu vrstvy je automatická správná funkce po sestavení sítě s libovolným počtem kanálů a vrstev. Původní složky vstupního vektoru by postupovaly sítí synchronně s analyzovanými signály jak v čase, tak v prostoru. Na konci by se odebraly ze sběrnice poslední vrstvy.

Možná je varianta s posuvným registrem FIFO s volitelnou délkou a volitelným počtem kanálů, který by se posouval synchronně se sítí NNSU. Zřejmě by však taková konstrukce byla složitější než rozprostřít registr FIFO rovnoměrně do kanálů a vrstev sítě.

Závěr 2.části

Bylo teoreticky prokázáno, že je možno zkonstruovat síť NNSU s rozlišením 9 bitů s průchodností cca 25, resp. 20 MSPS bez ohledu na počet vrstev. První varianta sítě je o něco rychlejší, omezení spočívá v maximálním počtu komparátorů a jejich nižší rozlišovací schopnosti. Druhá varianta je pomalejší, konstrukčně složitější, ale základní předností je softwarově volitelný počet komparátorů (volbou počtu komparačních hladin) téměř bez omezení počtu a zachování přesnosti i v oblasti rozřazení do intervalů (A/D převod 12 bitů). Každý kanál každé vrstvy by měl identické hardwarové zapojení.

Jestliže se vyjde z pevných čísel 7 bitů pro D/A převod a 16 kanálů ($= 2^4$), lze přejít na rozlišení $7 + 4 = 11$ bitů. Pro váhové koeficienty zbývá $17 - 11 = 6$ bitů. To dovoluje při binárním kódování 63 komparačních úrovní. Průchodnost sítě je cca 20 MSPS.

Jestliže se vyjde z pevných čísel 7 bitů pro D/A převod a 32 kanálů ($= 2^5$), lze přejít na rozlišení $7 + 5 = 12$ bitů. Pro váhové koeficienty zbývá $17 - 12 = 5$ bitů. To dovoluje při binárním kódování 31 komparačních úrovní. Průchodnost sítě je cca 20 MSPS.

Všechny údaje vycházejí z katalogových údajů výrobců součástek, nebyl proveden žádný experiment pro podporu zde prezentovaných tvrzení. Nebyla ani prostudována celá nabídka součástkové základny v této oblasti (je to mimo rozsah této studie). V úvahu byly vzaty běžně nabízené součástky (i když některé na objednávku) od předních firem zabývajících se touto technikou. Nelze však očekávat, že by se s uvedenými principy dala sestavit několikanásobně rychlejší síť ani po detailnějším studiu součástkové základny.

Použitá literatura:

Část 1.

- [1] Müller Jaromír: Studie a simulace optoelektronické implementace NNSU (Výzkumný úkol)
- [2] Internally Trimmed Precision IC Multiplier AD 534, katalogový list Analog Devices 1999
- [3] 8-/10-/12-Bit, High Bandwidth Multiplying DACs with Parallel Interface AD5424/AD5433/AD5445, katalogový list Analog Devices 2005
- [4] Precision, Very Low Noise, Low Input Bias Current, Wide Bandwidth JFET Operational Amplifiers AD8610/AD8620, katalogový list Analog Devices 2008
- [5] HA-4900, HA-4902, HA-4905, katalogový list Intersil č. FN2855.4, 20.10.2005
- [6] Fast, Precision Comparator AD790, katalogový list Analog Devices 2002
- [7] High Performance 32K´8 CMOS SRAM AS7C256, katalogový list Alliance Semiconductor, 5.1996
- [8] Enhanced 8-bit MCU with CAN controller and Flash T89C51CC01, katalogový list Atmel, 17.12.2001

Část 2.

- [9] Müller Jaromír: Studie proveditelnosti hardwarové realizace přepínacích neuronových sítí, diplomová práce, 9.5.2009
- [10] IS61LV12816L, katalogový list Integrated Silicon Solution, Inc., říjen 2005
- [11] 74LVT373 · 74LVTH373 Low Voltage Octal Transparent Latch with 3-STATE Outputs, katalogový list Fairchild Semiconductor™, revize říjen 1999
- [12] HI5660, katalogový list Intersil č. FN4521.7, červenec 2004
- [13] Wideband, Low-noise, Low Distorsion, Fully Differential Amplifier THS4509, katalogový list Texas Instruments, Inc., revize květen 2008
- [14] Ultrafast, 4 ns Single-Supply Comparators AD8611/AD8612, katalogový list Analog Devices, Inc., 2006
- [15] 12-Bit, 500-/550-MSPS Analog-to-Digital Converters ADS5465, ADS54RF63, katalogový list Texas Instruments, Inc., revize červenec 2009
- [16] SN65LVDT388 a další, High-Speed Differential Line Receivers, katalogový list Texas Instruments, Inc., revize prosinec 1999
- [17] T14M256A 32K X 8 High Speed CMOS Static RAM, katalogový list Taiwan Memory Technology, Inc., revize G, září 2001