

Algoritmy a struktury neuropočítačů

ASN - P14

Neuropočítače

- speciální výpočetní prostředky
- pro urychlení výpočtů neuronových sítí
- implementace zjednodušených algoritmů
- obvykle celočíselná aritmetika v kombinaci s normováním vstupních vektorů

Rozdělení neuropočítačů

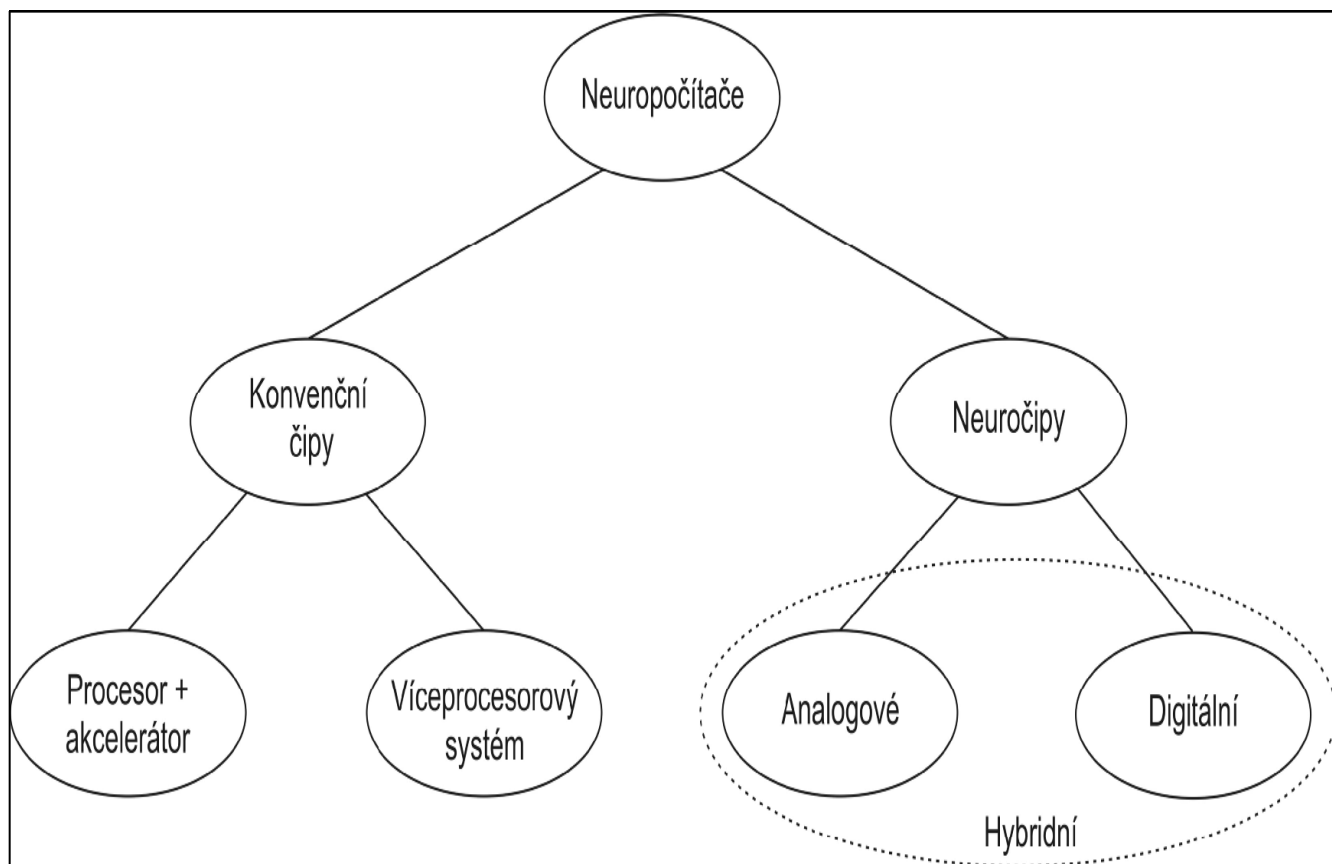
- konstrukce neuropočítačů - konvenční procesory doplněné o speciální instrukce a periferie
- umožňují významně urychlit výpočty neuronových sítí
- zvýšení výkonů zařízení - techniky paralelního zpracování
- jiný přístup - neuročipy

základní stavební prvek – aproximace neuronu
digitální implementace
čipy s analogovými
čipy s hybridními neurony

Algoritmy a struktury neuropočítačů

ASN - P14

Rozdělení neuropočítačů (podle [HEE95])



Využití konvenčních procesorů

- **klasické procesory MPU + speciální periferie pro větší výkonnost**
- **moderní procesory pro osobní počítače obsahují instrukční rozšíření pro práci s multimédií - lze jej použít s jistými omezeními také pro trénování Kohonenových map**

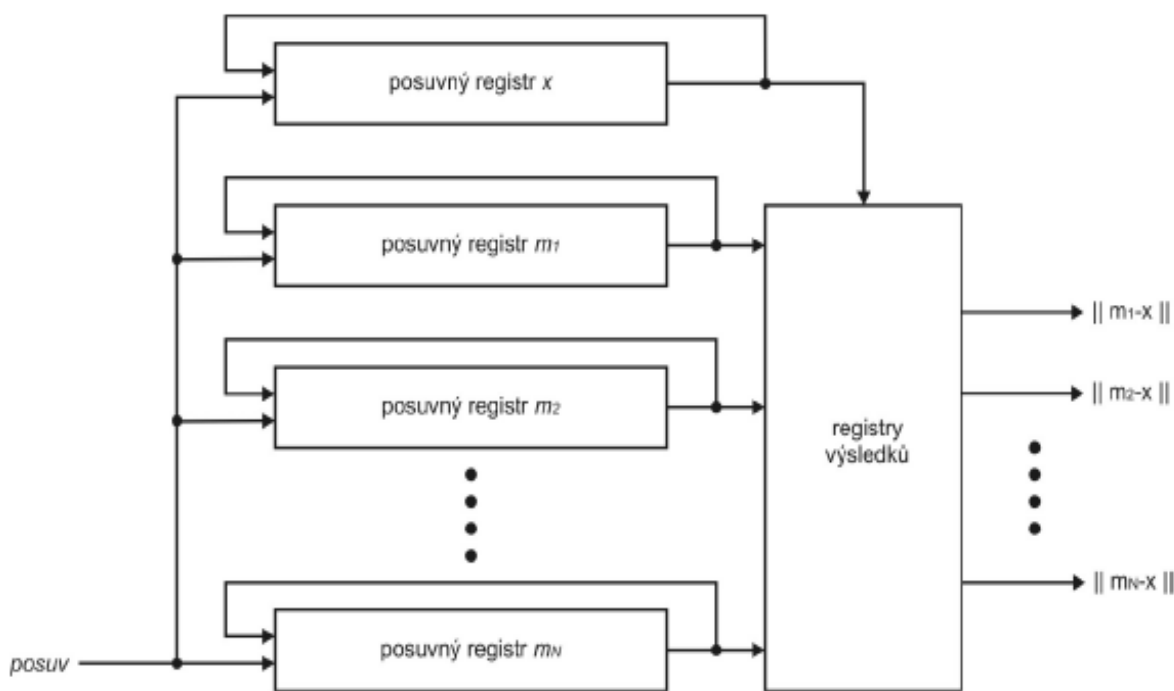
Algoritmy a struktury neuropočítačů

ASN - P14

Paralelní (víceprocesorové) systémy

- pro zvýšení výkonu - několika procesorů současně
 - ↳ vhodné např. pro KSOM Samotný algoritmus KSOM

SIMD – *Single Instruction Multiple Data*



Obrázek 7: Náčrtek architektury SIMD pro výpočet vítězného neuronu (volně podle [KOH01])

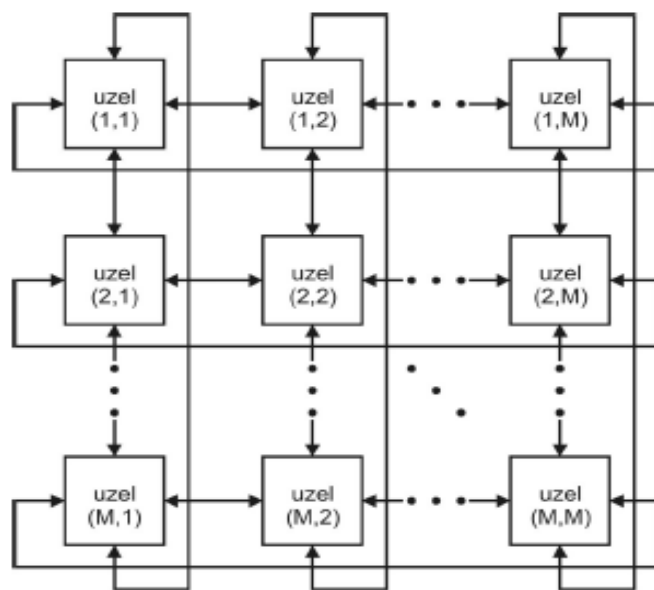
- na jednotlivých procesorech se provádí stejný program
- zpracovávají se různá data
- porovnání bit po bitu, není třeba sekvenčně hledat minimum

Algoritmy a struktury neuropočítačů

ASN - P14

MIMD transputer – *Multiple Instruction Multiple Data*

každý procesor – vlastní program, vlastní data
procesory uspořádané do n-dimenzionálních sítí
méně vhodné pro KSOM – algoritmus nelze rozdělit
na nezávislé paralelní procesy



Obrázek 8: Dvoudimenzionální MIMD architektura (podle [KOH01])

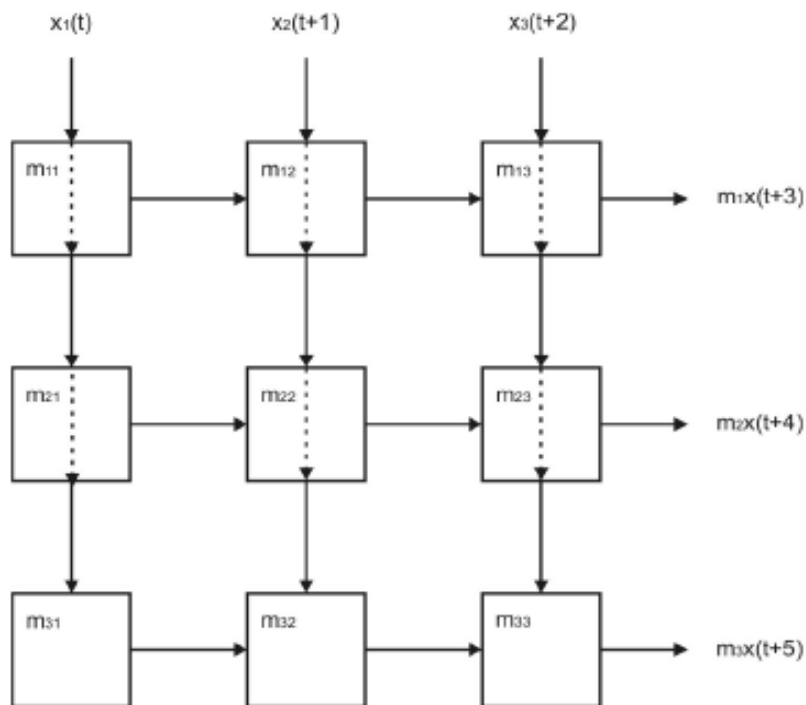
Implementace v systolickém poli

využití většího množství jednotek uspořádaných maticově
data se šíří tímto polem (maticí) jedním směrem (např.dolů)
mezivýsledky se šíří ortogonálně (vpravo)
výpočet je iterativní

Na obrázku je náčrt funkce výpočtu vzdáleností mezi vstupním vektorem x a vektory neuronů $m_1 - m_3$. Složky vstupního vektoru jsou distribuovány vertikálně, součiny se sčítají horizontálně. Při úpravě vah se složky vektoru a koeficient učení šíří vertikálně a vektory neuronů m jsou adaptovány lokálně.

Algoritmy a struktury neuropočítačů

ASN - P14



Obrázek 9: Náčrtek implementace v systolickém poli

snížené nároky na hardware

[IEN93b]

Paralelní stroj složený z klasických osobních počítačů – cluster

- varianta MIMD stroje
- jednotlivé procesory nejsou umístěny v jednom čipu ani na jedné základní desce
- procesory jsou v jednotlivých osobních počítačích spojených pomocí standardizovaného rozhraní (Ethernet)

Nevýhoda: relativně nižší výkonnost

Výhoda: příznivější cena

Algoritmy a struktury neuropočítačů

ASN - P14

Měření výkonnosti neuropočítačů

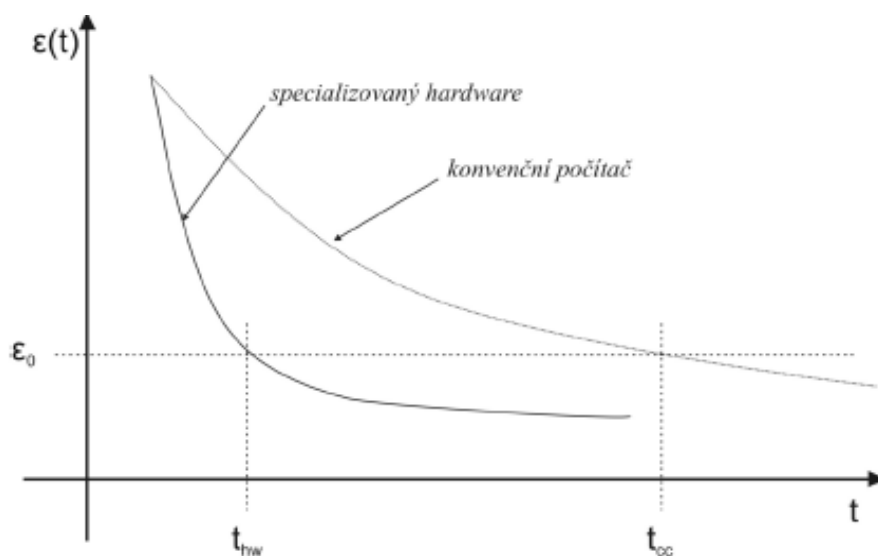
ukazatel pro neuronové sítě :

MCPS (Million Connections Per Second) - vyjadřuje výkon neuropočítače ve fázi využívání natrénované sítě

MCUPS (Million Connections Update Per Second) - pro výkon ve fázi učení

Ukazatele vyjadřují výpočetní sílu počítače, neberou v úvahu vliv konkrétní implementace na konvergenci algoritmu.

Nelze porovnávat neuropočítače různých konstrukcí.

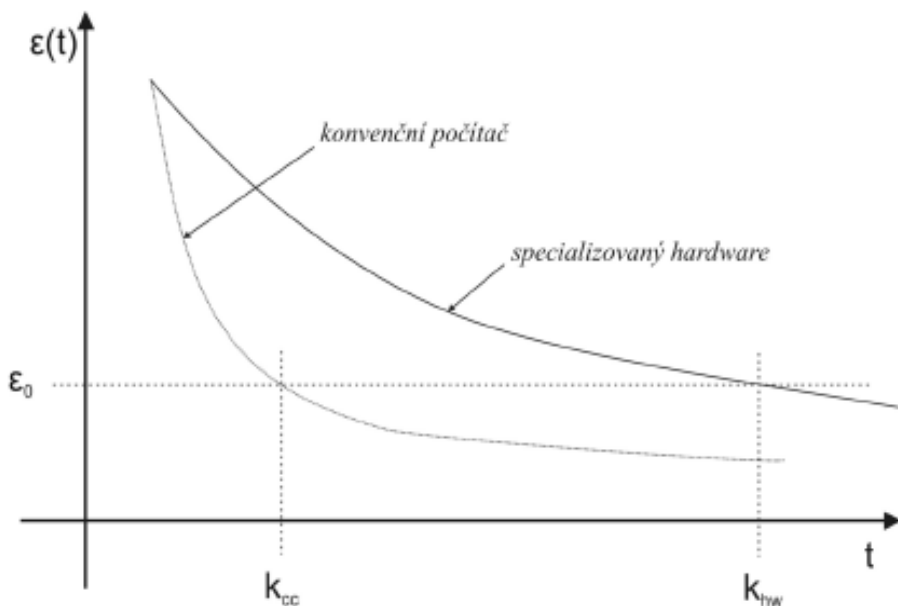


Obrázek 10: Porovnání rychlosti konvergence (volně podle [COR94a])

z porovnání doby potřebné k dosažení určitého stupně uspořádání sítě

Algoritmy a struktury neuropočítačů

ASN - P14



Obrázek 11: Porovnání počtu iterací (volně podle [COR94a])

MANTRA I - neuropočítač vyvíjený v EPFL v Lausanne [COR94c], [VIR93a], [VIR93b]

- návrh s ohledem na implementaci jednovrstvých i vícevrstevných neuronových sítí, Hopfieldových sítí a Kohonenových map
- složení - ze zákaznických obvodů GENES IV
- v maximální konfiguraci obsahuje matici 40×40 těchto obvodů
- výkon 200 – 400 MCPS a 100-200 MCUPS (v závislosti na typu NN)

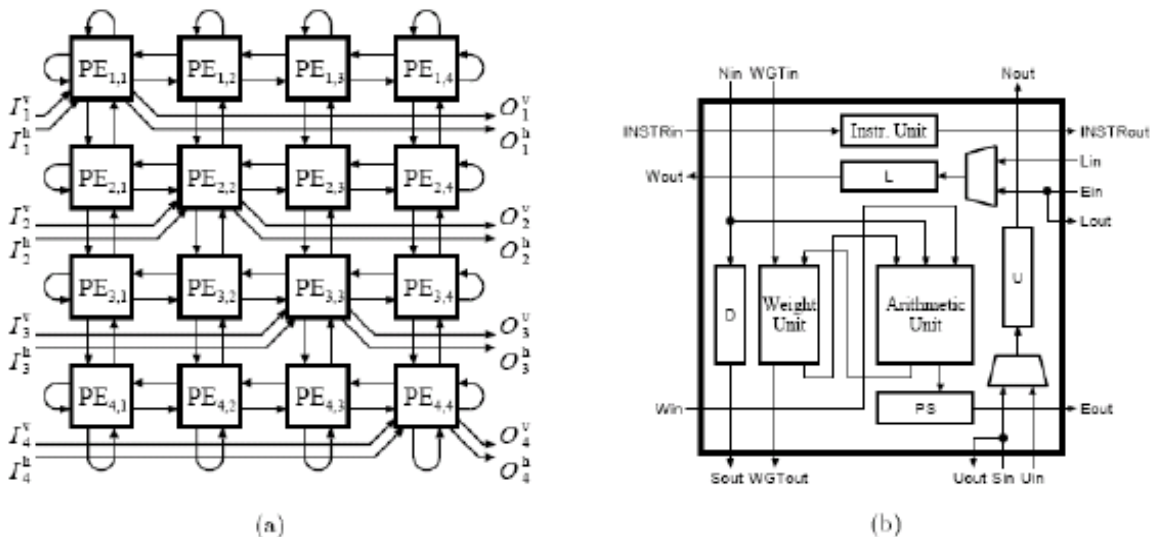
Každý obvod - šestnáct výpočetních jednotek uspořádaných do čtvercové matice

Každá jednotka - šest základních operací:

1. násobení matice vah W vstupním vektorem x
2. výpočet druhé mocniny Eukleidovské vzdálenosti
3. Hebbovský zákon učení
4. Kohonenovské učení – aktualizace vah
5. hledání největšího elementu vektoru
6. hledání nejmenšího elementu vektoru.

Algoritmy a struktury neuropočítačů

ASN - P14



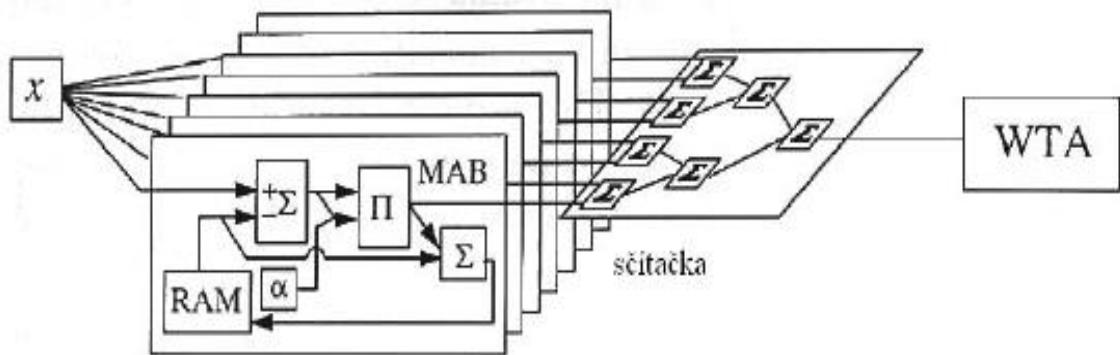
Obrázek 13: Struktura čipu GENES IV (a), vnitřní struktura jednotky (b) (podle [IEN94])

COKOS - COpcessor for KOhonen's Self-organizing map

- čip speciálně navržený pro podporu Kohonenových map
- osm paralelních jednotek nazývaných MAB - Memory Arithmetic Board
- odčítačka + násobička + sčítačka v každé MAP
- pro uložení mezivýsledků paměť RAM v každé MAP
- každá MAB jednotka počítá druhou mocninu vzdálenosti vstupního vektoru x a modelu neuronu mi .
- výsledky jednotlivých jednotek jsou sečteny sčítačkou
- vítěz je nalezen pomocí WTA (Winner Takes All) jednotky obsahující i separátní paměť pro uložení vzdáleností jednotlivých vektorů

Algoritmy a struktury neuropočítačů

ASN - P14



Obrázek 14: Architektura neuropočítače COKOS (podle [KOH01])

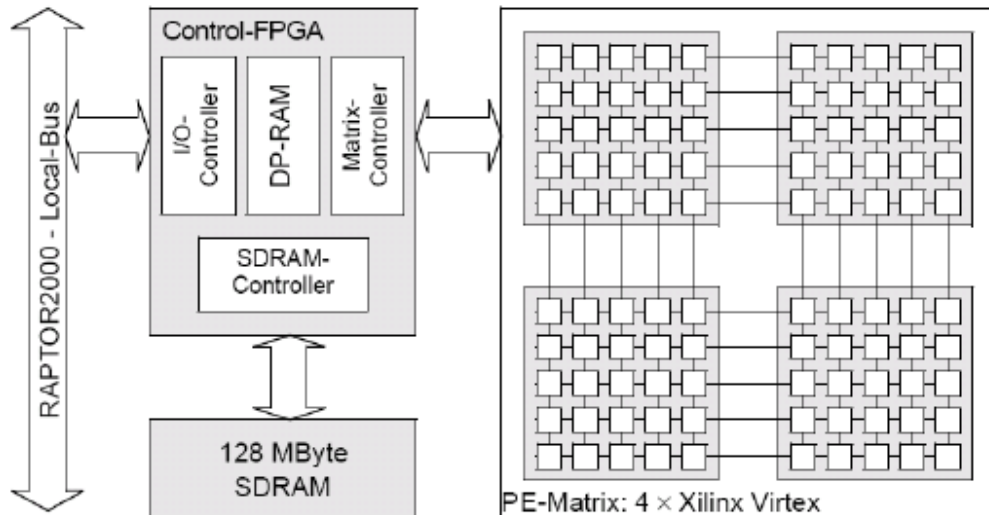
- neuropočítač je propojen s osobním počítačem pomocí asynchronního rozhraní
- součástí systému je software pro práci s neuropočítačem
- výkon: kolem 16MCUPS
- váhy neuronů jsou 16-ti bitové

NBISOM_25 a NBX

- vyvinutý speciálně pro zrychlení výpočtů Kohonenových map
- neuropočítač sestavený z 16-ti zakázkových čipů
- každý obsahuje matici 5×5 speciálních výpočetních jednotek
- každá jednotka je vybavena pamětí pro uložení 64 elementů vektoru modelu m_i neuronu, jednotkou pro výpočet vzdálenosti (se 14-ti bitovým akumulátorem) a s řídicí logikou

Algoritmy a struktury neuropočítačů

ASN - P14



Obrázek 15: Implementace NBX v systému RAPTOR2000 (podle [POR02a])

- upraven pro implementaci v programovatelném hradlovém poli (FPGA) Každá jednotka podporuje mimo
- základní operace pro výpočet Kohonenových map (výpočet vzdálenosti, hledání minima a adaptace vah)
- operace pro výpočet U-matic používaných k zobrazení

Literatura:

[BAR06] Bártů, M.: *Analýza možností implementace Kohonenových map*. Diplomová práce, FEL ČVUT v Praze, 2006

[COR94a] Cornu T., Ienne P.: Performance of Digital Neurocomputers, In *Proceedings of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems*, Turin, Italy, pp 87-93., September 1994.

Algoritmy a struktury neuropočítačů

ASN - P14

[HEE95] Heemskerk J.: Overview of Neural Hardware, disertační práce, Unit of Experimental and Theoretical Psychology, Leiden University, The Netherlands, 1995

[IEN93b] Ienne P.: Quantitative Comparison of Architectures for Digital Neuro-Computers, In *Proceedings of the International Joint Conference on Neural Networks*, Nagoya, Japan, Volume II, pp. 1987-90, October 1993

[IEN94] Ienne P., Viredaz M. A.: Implementation of Kohonen's Self-Organizing Maps on MANTRA I, In *Proceedings of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems*, Turin, Italy, pp. 273-79, September 1994

[KOH01] Kohonen T.: *Self-Organizing Maps*. Springer-Verlag 2001, 3. vydání, ISBN 3-540-67921-9

[POR02a] Porrman M., Witkowski U., Kalte H., Ruckert U.: Implementation of Artificial Neural Networks on a Reconfigurable Hardware Accelerator, In *Proceedings of the 10th Euromicro Workshop on Parallel, Distributed and Network-based Processing*, Grand Canaria Island, Spain, pp. 243 – 250, January 2002