

## IMPLEMENTACIJA DIGITALNOG VEŠTAČKOG NEURONA

Miljan S. Nikolić i Dejan M. Maksimović

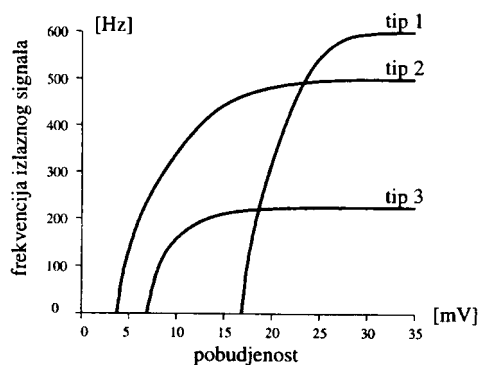
Univerzitet u Nišu, Elektronski fakultet, Beogradska 14, 18000 Niš

**Sadržaj:** Neuronske mreže uspešno se koriste za rešavanje problema kod kojih standardne metode ne daju zadovoljavajuće rezultate. Sastoje se od više međusobno povezanih neurona. Efikasna hardverska implementacija veštačkog neurona omogućuje hardversku realizaciju neuronskih mreža. U ovom radu predložene su dve implementacije veštačkog digitalnog neurona na FPGA čipu.

### 1. UVOD

Težnja čoveka da stvori inteligentne mašine stara je više decenija. Pri tome se kao uzor koristi nervni sistem živih bića, pre svega čoveka, zbog njegove superiornosti čak i u poređenju sa najsavremenijim super-računarima u rešavanju slabo definisanih i nejasnih problema. Nervni sistem čoveka sastoji se od velikog broja (reda  $10^{11}$ ) neurona koji su grupisani u stotinak grupa. Svaka grupa zadužena je za određenu vrstu poslova. Neuroni u svakoj grupi međusobno su jako povezani. Grupe neurona međusobno su takodje povezane čineći jedan kompleksan sistem za obradu informacija.

Osnovni element ovog sistema je neuron, odnosno nervna ćelija. Neuron se sastoji od tela, dendrita i aksona. Preko dendrita neuron prima signale od drugih neurona, u telu se ulazni signali integrale i generiše se izlazni signal koji se preko aksona dalje vodi na ulaze drugih neurona ili na aktuatorne organe kao što su na primer mišići. Spojevi medju neuronima nazivaju se sinapsama. Dovodjenje pobudnog signala na neku od sinapsi izaziva hemijsko-fizičku reakciju koja menja potencijal neurona. Zavisno od toga da li je sinapsa ekscitatorna ili inhibitorna, potencijal tela neurona se respektivno povećava ili smanjuje. Kad se prekorači aktivacioni potencijal (prag), dolazi do izlazne aktivnosti. Informacija o stanju neurona sadržana je u frekvenciji njegovog izlaznog signala, kao što je prikazano za nekoliko tipova neurona na slici 1 [1]. Pod *pobudjenošću* se podrazumeva razlika izmedju potencijala neurona i njegovog normalnog potencijala. Može se uočiti da postoji prag ispod koga ne postoji izlazni signal. Po prekoračenju praga, frekvencija izlaznog signala raste do zasićenja.



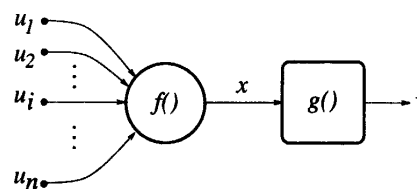
Sli 1. Prenosna (aktivaciona) karakteristika neurona

Neuron se može modelovati softverski ili hardverski. Softverski modeli neurona koriste se za rešavanje numerički slabo definisanih problema kakvi su recimo prepoznavanje slike i glasa, dijagnostika u medicini i elektronici, odlučivanje i drugo. Hardverski se veštački neuron realizuje kao analogno ili digitalno električno kolo sa prenosnom karakteristikom sličnom prenosnoj karakteristikom prirodnog neurona datoj na slici 1 [2, 3]. U ovom radu opisane su dve praktične implementacije digitalnog veštačkog neurona. Digitalni neuron je implementiran na Xilinx-ovom FPGA čipu XC4003E sa ciljem da se omogući razvoj neuronskih mreža za detekciju i dijagnostiku defekata u elektronskim kolima [4].

### 2. PROJEKTOVANJE DIGITALNOG VEŠTAČKOG NEURONA

Modele neurona moguće je svrstati u dve velike klase. Prvu čine *statički*, a drugu *dinamički* modeli. Statički modeli opisuju neuron algebarskim jednačinama koje povezuju ulazne signale i signal na izlazu. Za razliku od njih, u dinamičkim modelima ulazni signali i izlazni signal neurona povezani su diferencijalnim ili, kod diskretnih modela, diferencnim jednačinama. Razlika je evidentna. U statičkim modelima je zanemarena dinamika pojava u neuronu, to jest smatra se da su ulazni signali sporo promenljivi. U ovom radu je izabran statički model neurona jer ovaj u potpunosti zadovoljava zahteve za primenu u dijagnostici elektronskih kola.

Uopštena blok šema statičkog modela neurona prikazana je na slici 2. Na ulaz neurona dovodi se  $n$  signala  $\mathbf{u}=(u_1 u_2 \dots u_n)^T$  sa ulaznih sinapsi. Interni signal  $x$  (često se naziva i *aktivacioni* signal neurona) izračunava se kao  $x=f(\mathbf{u})$ , a izlazni signal neurona je  $v=g(x)=g(f(\mathbf{u}))$ . Funkcija  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  daje zavisnost aktivacionog signala  $x$  od ulaznih signala i poznata je pod imenom *diskriminanta* ili *funkcija interakcije ulaza*. Funkcija  $g: \mathbf{R} \rightarrow \mathbf{R}$  definiše uticaj aktivacionog signala  $x$  na izlazni signal neurona. Funkcija  $g$  naziva se *aktivaciona* ili *prenosna funkcija* neurona i treba da aproksimira neku od krivih sa slike 1.



Sli 2. Statički model neurona

Funkcija interakcije ulaza  $f$  može se definisati na različite načine. Najčešće se upotrebljava *linearna* funkcija interakcije ulaza koja aktivacioni signal izražava kao linearnu kombinaciju pobudnih signala:

$$f(\mathbf{u}) = \sum_{i=1}^n w_i u_i \quad (1)$$

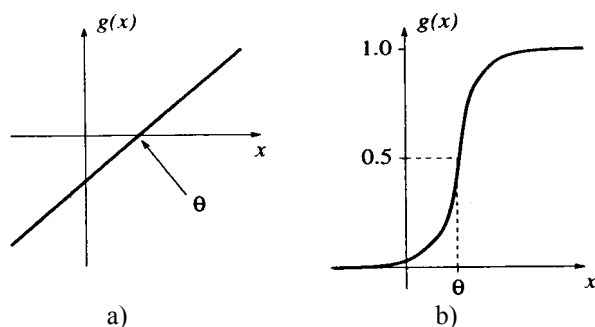
U neuronskoj mreži sastavljenoj od više neurona signali  $u_i$ ,  $i=1, 2, \dots, n$  su izlazni signali drugih neurona, a koeficijenti  $w_i$ ,  $i=1, 2, \dots, n$  predstavljaju težine veza između posmatranog neurona i neurona u mreži koji pobudjuju posmatrani neuron. Po analogiji sa prirodnim neuronom, funkcijom (1) modeluje se osobina nervne ćelije da integriše ulazne signale dok težine  $w_i$  predstavljaju uticaj sinapsi. Ako je  $w_i > 0$ , onda je sinapsa ekscitatorna i obrnuto, ako je  $w_i < 0$ , ona je inhibitorna. U literaturi je moguće sresti i drugačije funkcije interakcije ulaza kod kojih se uvode nelinearnosti, obično polinomskog oblika, da bi se povećala procesna moć neurona [1]. Za implementacije neurona koje će biti opisane u ovom radu korišćena je funkcija interakcije ulaza definisana izrazom (1).

Funkcija  $g$  modeluje prenosnu karakteristiku neurona. U literaturi se mogu sresti različite funkcije za aproksimaciju ove karakteristike. Najprostija aproksimacija je linearna. S obzirom da je ova aproksimacija loša, neuroni sa linearnom prenosnom funkcijom koriste se uglavnom u neuronskim mrežama u kojima osim ovakvih neurona ima i neurona sa nelinearnim prenosnim funkcijama. Najčešće se, međutim, za aproksimaciju karakteristika sa slike 1 koristi klasa sigmoidalnih funkcija, od kojih je najvažnija logistička funkcija:

$$g(x) = \frac{1}{1 + e^{-(x-\theta)}} \quad (2)$$

Grafički prikaz pomenutih aktivacionih funkcija dat je na slici 3. Na osnovu pozitivnih iskustava u korišćenju sigmoidne funkcije za softverski model neurona [5, 6], za hardversku realizaciju izabrana je logistička funkcija definisana jednačinom (2). Efikasnoj hardverskoj realizaciji ove funkcije već godinama se u naučnoj javnosti posvećuje velika pažnja [7, 8].

U ovom radu biće opisana dva rešenja za implementaciju digitalnog neurona koji ima  $n=2$  sinapse. Svi izvedeni zaključci, međutim, važe i za neuron sa proizvoljnim brojem sinapsi  $n$ . Svaka sinapsa je  $N$ -tobitna, što znači da neuron ima dva  $N$ -tobitna ulaza. Težine su realizovane kao  $N$ -tobitne konstante. Podrazumevano je da će se obučavanje neuronske mreže izvršiti na softverskom modelu, tako da će težine veza biti poznate pre konačne hardverske implementacije. Kako bi se sa izlaza neurona mogli pobudjivati drugi neuroni u mreži, izlaz neurona je takodje  $N$ -tobitan. Moguće je realizovati i neuron kod koga je broj bitova u izlaznoj reči veći od  $N$ , čime bi se izbeglo prekoračenje u izračunavanjima.

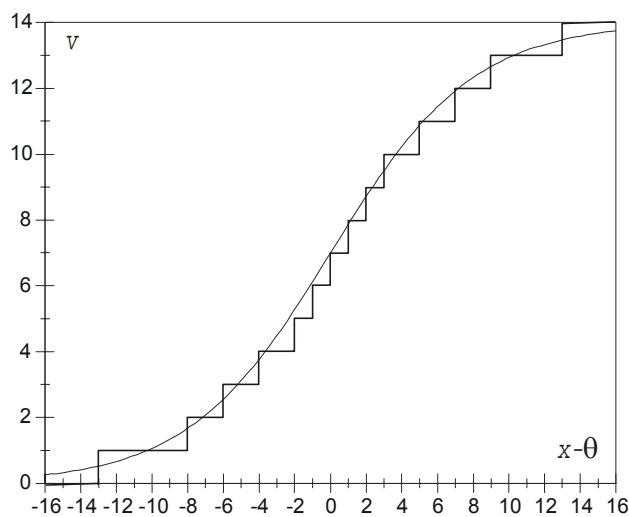


Sl. 3: Aktivacione funkcije: a) linearna i b) logistička

Ulazni signali i težine ulaznih sinapsi implementiranog neurona normalizovani su tako da se mogu predstaviti  $N$ -tobitnim celim nenegativnim brojevima. Neuron ima dve sinapse. Aktivacioni signal  $x$  predstavlja se sa  $2N+1$  bitova. Aktivaciona funkcija implementirana je pomoću kombinacione mreže. Analiza logističke funkcije pokazuje da je za aproksimaciju od interesa interval od  $[\theta-4, \theta+4]$ . Kada je vrednost promenljive  $x$  u ovom opsegu, vrednost  $g(x)$  kreće u granicama od 0.018 do 0.98. Dakle, ako je  $x < \theta-4$ , može se smatrati da je  $g(x)=0$ , a ako je  $x > \theta+4$ , može se smatrati da je  $g(x)=1$ . U cilju poboljšanja rezolucije u ovom opsegu, logistička funkcija je realizovana u normalizovanom obliku koji se opisuje izrazom:

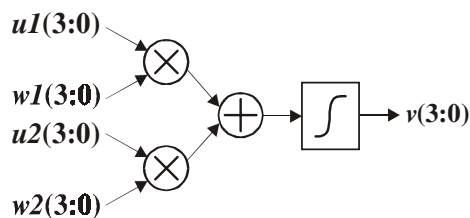
$$g(x) = \frac{k_1}{1 + e^{-\frac{x-\theta}{k_2}}} \quad (3)$$

Ako normalizacione konstante imaju vrednosti  $k_1=14$ ,  $k_2=4$ , dobija se aproksimacija logističke funkcije prikazana na slici 4. Ovakva aproksimacija funkcije  $g$  odgovara realizaciji neurona sa dužinom ulaznih i izlaznih reči  $N=4$ .



Sl. 4: Aproksimacija aktivacione funkcije u hardverskoj implementaciji digitalnog neurona

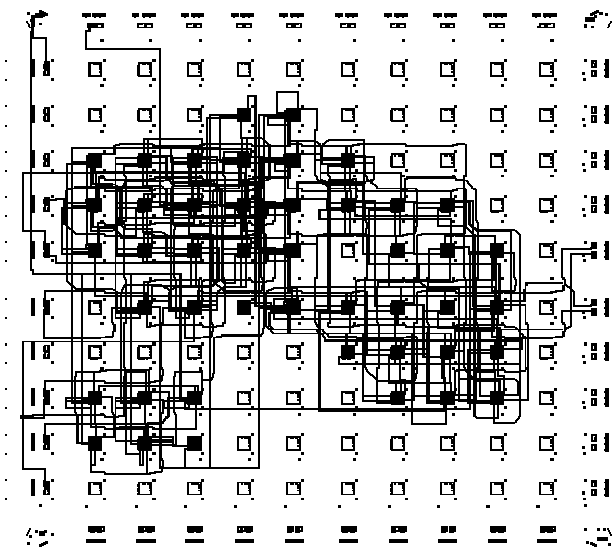
Implementirane su dve strukture veštačkog digitalnog neurona. Prvo rešenje prikazano je blok šemom na slici 5. Za modelovanje svake sinapse koristi se  $N$ -tobitni množač. Kolo je asinhrono i sastavljeno samo od kombinacionih elemenata. Množači formiraju parcijalne proizvode  $u_1 \cdot w_1$  i  $u_2 \cdot w_2$ . Na izlazu sabirača dobija se aktivacioni signal  $x$ . Nad aktivacionim signalom primenjuje se diskretizovana sigmoide funkcija prikazana na slici 4.



Sl. 5: Blok šema rešenja I

Drugo rešenje dato je blok dijagramom na slici 6. Koristi se samo jedan množač, a sabiranje je akumulaciono. Blok šema sa slike 6 implementira se sinhronim sekvencijalnim kolom. Ovo rešenje zahteva sinhronizaciju izračunavanja u





Sl 10: Lejaut rešenja II na čipu XC4003E

#### 4. ZAKLJUČAK

Rešenje I je asinhrono i realizovano kao kombinaciona mreža. Primenjeni električni model je blizak strukturi prirodnog neurona, jer se koristi više paralelnih množača koji odgovaraju pojedinačnim sinapsama. Kašnjenje najdužeg puta od ulaza do izlaza neurona realizovanog korišćenjem rešenja I je 47.7ns, što znači da je maksimalna prihvatljiva frekvencija promene vrednosti ulaznih signala nešto veća od 20MHz.

Rešenje II je sekvencijalno i sinhrono. Njegova upotreba zahteva dobro poznavanje ugrađenog algoritma i kontrolnih signala, jer treba pažljivo sinronizovati postavljanje vrednosti ulaznih signala i težina, kao i interni protok podataka u neuronu. Izlaz se dobija posle devet perioda taktnog signala *clk*. Post-lejaut tajming simulacijom utvrđeno je da je maksimalna prihvatljiva frekvencija signala *clk* nešto veća od 56MHz, što odgovara maksimalnoj frekvenciji promene ulaznih signala od 4MHz. Dakle, rešenje I ima veću radnu frekvenciju, što se i moglo očekivati.

Što se tiče kompleksnosti, rešenje I zauzima nešto manju površinu na čipu, ali razlika nije znatna. Interesantno je razmotriti kako utiče povećanje dužine ulazne/izlazne reči *N* na kompleksnost čipa. Kod rešenja I, sa povećanjem broja bitova na ulazu raste kompleksnost  $n=2$  paralelnih množača. Kada je broj sinapsi *n* dosta veći od 2, kompleksnost rešenja I može vrlo brzo da raste sa povećanjem dužine reči *N*. Sa druge strane, kod rešenja II upotrebljen je samo jedan množač tako da kompleksnost rešenja II ne raste tako brzo sa porastom *N*.

S druge strane, da bismo iskoristili rešenje I za realizaciju veštačke neuronske mreže, treba samo povezati ulaze i izlaze pojedinačnih neurona u mreži. U slučaju da koristimo rešenje II za realizaciju veštačke neuronske mreže, pored povezivanja neurona, potrebno je projektovati i ugraditi kontrolnu logiku za sinhronizaciju rada pojedinih neurona u mreži, što povećava kompleksnost čipa i povećava cenu.

Takođe, treba razmotriti posledice povećanja broja sinapsi (odnosno ulaza) neurona *n*. Kod rešenja I povećanje broja *n* dovodi do povećanja kompleksnosti čipa jer se sa svakim novim ulaznim priključkom dodaje po jedan *N*-bitni množač, a i složenost sabirača raste. Pri tome se, međutim,

ne menja značajno maksimalna radna frekvencija kola. Kod rešenja II porast broja ulaza *n* dovodi do povećanja dužine upravljačke sekvence. Ovo se odražava samo na kompleksnost kontrolne logike i dovodi do neznatnog povećanja kompleksnosti čipa. Međutim, maksimalna frekvencija rada neurona smanjuje se jer se akumulaciono sabiranje realizuje u većem broju ciklusa upravljačkog takta *clk*.

#### LITERATURA

- [1] S.Lj.Milenković, "Veštačke neuronske mreže", *Zadužbina Andrejević*, biblioteka *Dissertatio*, Beograd 1997.
- [2] B.W.Wah, "Special Issue on Artificial Neural Networks - Guest Editor's Introduction", *IEEE Trans. on Computers*, Vol. 40, No. 12, pp. 1317-1319, December 1991.
- [3] A.F.Murray, "Silicon implementations of neural networks", *IEE Proc.-F*, Vol. 138, No. 1, pp. 3-12, February 1991.
- [4] V.B.Litovski, "Projektovanje elektronskih kola: simulacija, optimizacija, testiranje, fizičko projektovanje", *DGIP "Nova Jugoslavija"*, Vranje, 2000.
- [5] V.B.Litovski, J.Radjenović, Ž.Mrčarica, S.Milenković, "MOS transistor modelling using neural network", *Electronics Letters*, Lett. 28(13), pp. 1766-1768, 1992.
- [6] V.B.Litovski, Ž.Mrčarica, T.Ilić, "Simulation of non-linear magnetic circuits modelled using artificial neural network", *Simulation Practice and Theory*, No. 5, pp. 553-570, 1997.
- [7] P.Murtagh, A.C.Tsoi, "Implementation issues of sigmoid function and its derivative for VLSI digital neural networks", *IEE Proc.-E*, Vol. 139, No. 3, May 1991.
- [8] K. Barsterretxea, J.M.Tarela, I.d.Campo, "Digital design of sigmoid approximator for artificial neural networks", *Electronics Letters*, Vol. 38, No. 1, January 2002.
- [9] Active-HDL, ver. 3.6., ALDEC Inc., 1998.
- [10] <http://www.xilinx.com>

**Abstract:** Neural networks are successfully used for solving the problems where standard methods cannot be applied. They consist of a number of mutually highly connected neurons. Efficient hardware implementation of an artificial neuron enables the implementation of artificial neural networks on chip. In this paper two implementations of digital artificial neuron on an FPGA chip are presented.

#### IMPLEMENTATION OF DIGITAL ARTIFICIAL NEURON

Miljan S. Nikolić, Dejan M. Maksimović