

HW AGY, FPGA ALAPÚ MESTERSÉGES INTELLIGENCIA

NAGY BALÁZS



M Ű E G Y E T E M 1 7 8 2



SmartComLab



AITIA

MI AZ A FPGA?

- Programozható hardver
- A CPU előre meghatározott műveleteket képes végre hajtani, a FPGA-ban a tervező határozza meg a műveleteket, logikát
- VHDL vagy Verilog nyelven komplex digitális áramköröket lehet leírni, amik áramkör szinten valósulnak meg a FPGA chipben
- A FPGA fejlesztés hardver tervezés, NEM programozás
- Legnagyobb gyártók: Xilinx—Altera, AMD - Intel



FPGA ELŐNYEI ÉS HÁTRÁNYAI

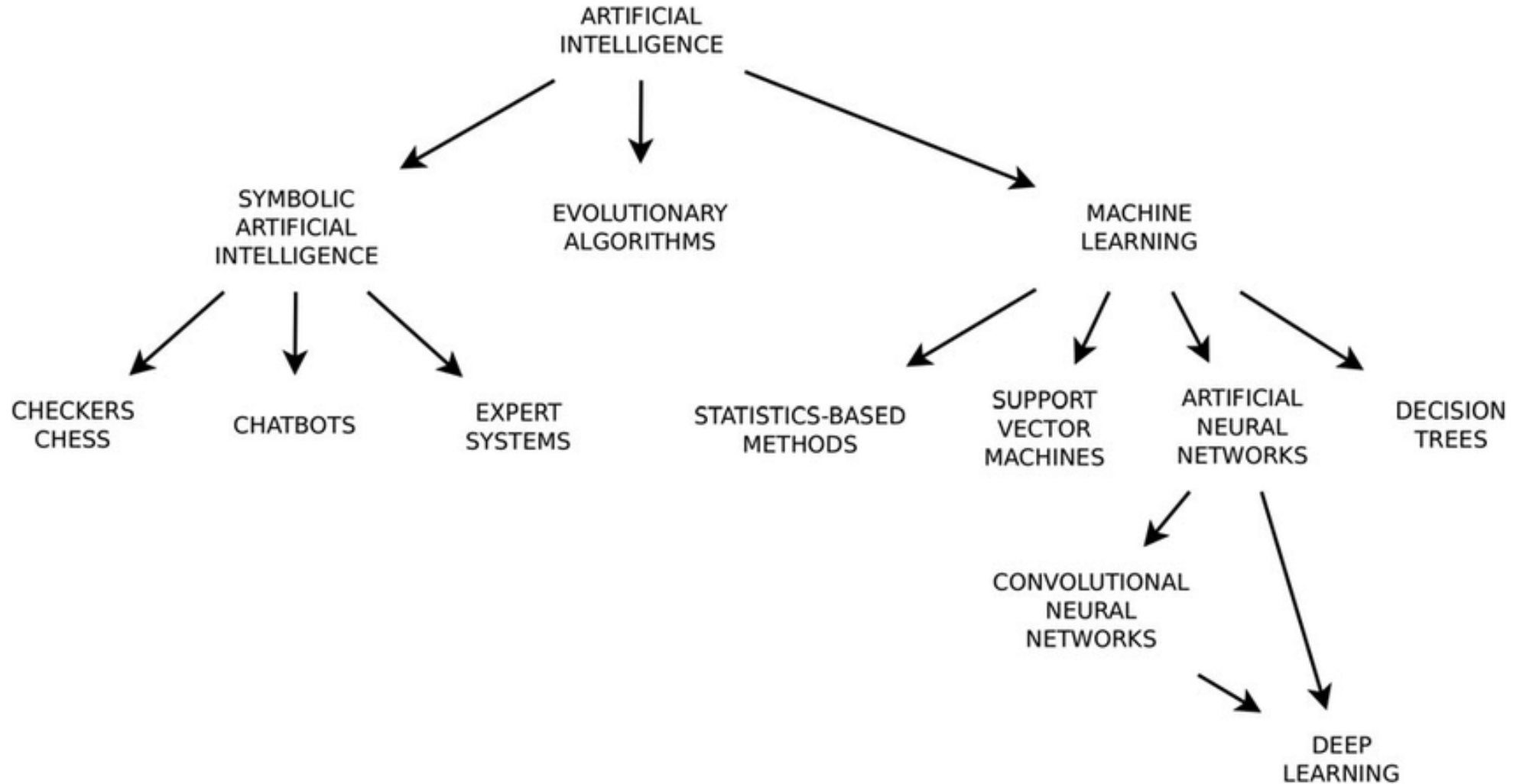
Előnyök:

- Hatalmas paralel számítási kapacitás
- Legnagyobb elérhető szabadság
- Teljes, valós idejű adatfeldolgozás
- Órajel szinten determinisztikus viselkedés

Hátrányok:

- Komoly szaktudás szükséges
- Nehezebb a hiba keresés
- Nagyobb befektetés kell a fejlesztéshez
- Lassabb a fejlesztés
- Nincsenek előre megírt library-k?

Az MI Technológiák

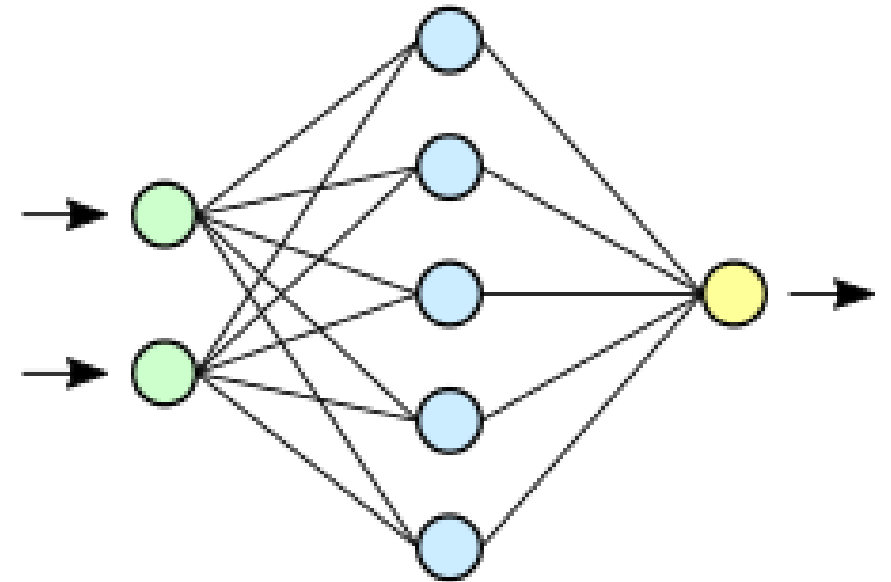


FPGA ÉS MI - MIT ÉRDEMES FPGA-RA IMPLEMENTÁLNI?

- Minden, ami FPGA-n tud működni*, az CPU-n vagy GPU-n is működhet; a teljesítményben, maximálisan elérhető komplexitásban és az energia fogyasztásban van különbség
- A FPGA implementáció különösen időigényes, ezért komolyan meg kell gondolni, belevágunk-e
 - Egy egyszerű neuront Python-ban 30 perc alatt el lehet készíteni
 - Ugyanazt a neuront ~8 óra FPGA-n implementálni
- *Real-time és biztonság kritikus alkalmazások
- Teljesítmény kritikus alkalmazások
- MI területen elsősorban **mesterséges neurális hálózatok** pl. Tesla

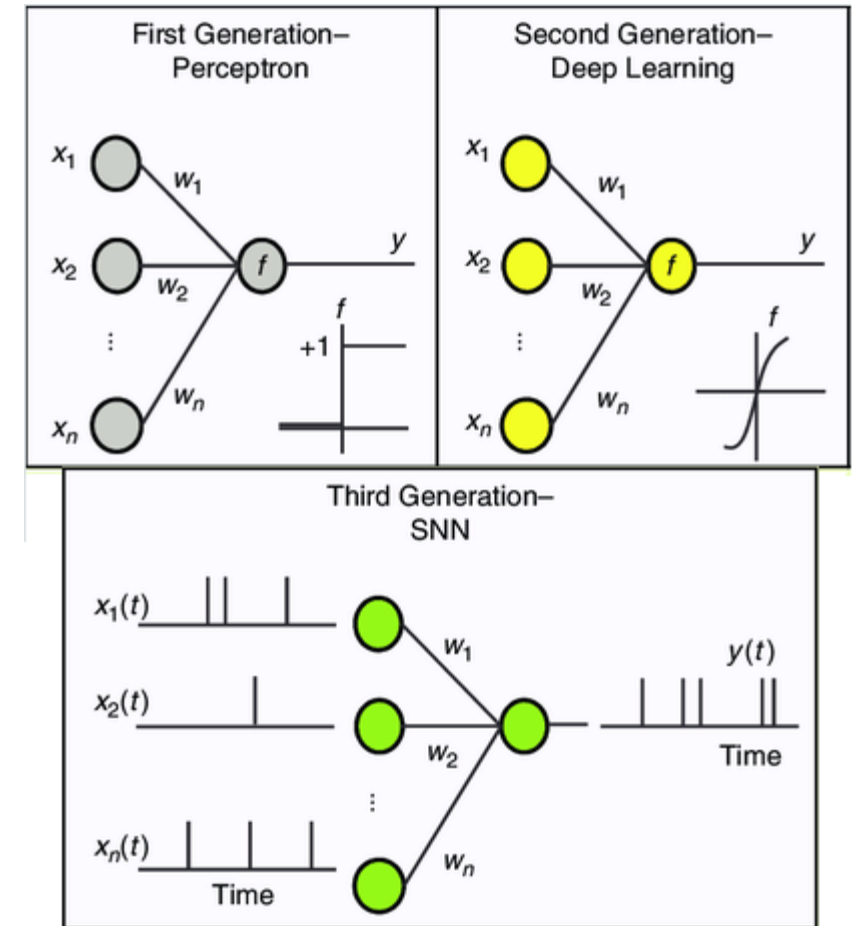
MESTERSÉGES NEURÁLIS HÁLÓZATOK

- Az emberi agy által inspirált számítási struktúra
- Az alap eleme a neuron
- A neuronok rétegekbe vannak szervezve, a rétegek más-más feladatot látnak el
- Példákon keresztül tanul a rendszer
- Probléma megoldáshoz nem kell modellt alkotni



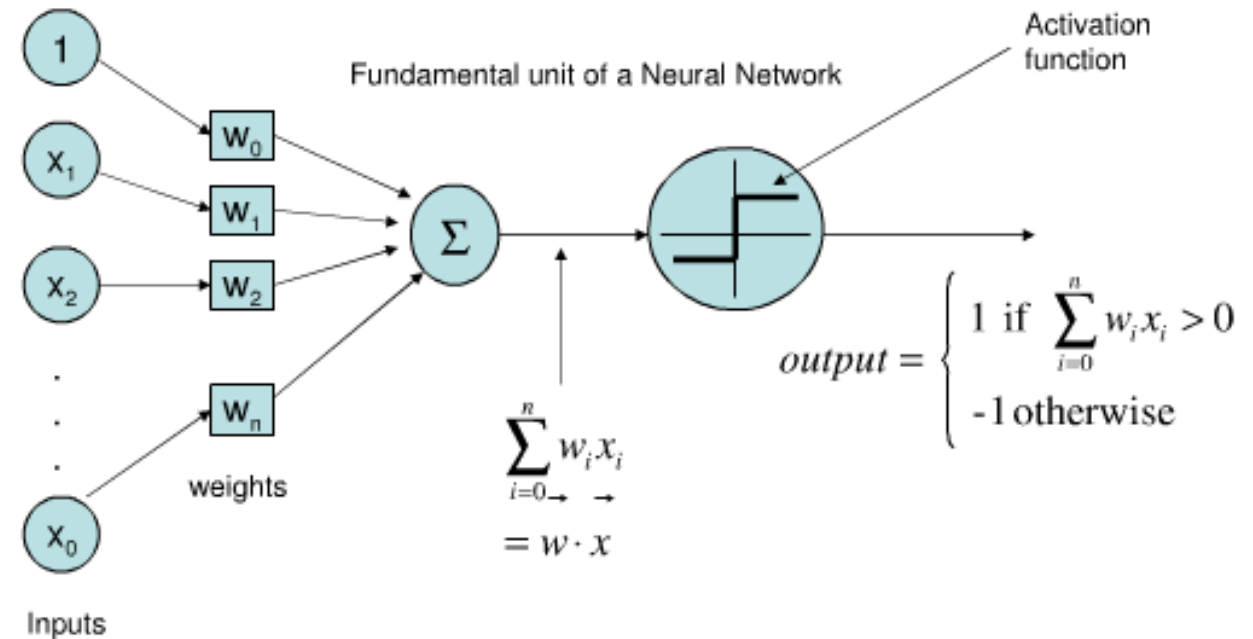
NEURÁLIS HÁLÓK 3 GENERÁCIÓJA

- *Perceptron*
 - Bináris osztályozó
- *Deep Learning*
 - Többrétegű fuzzy osztályozó
- *Spiking neural network*
 - Biológia ihletésű modell



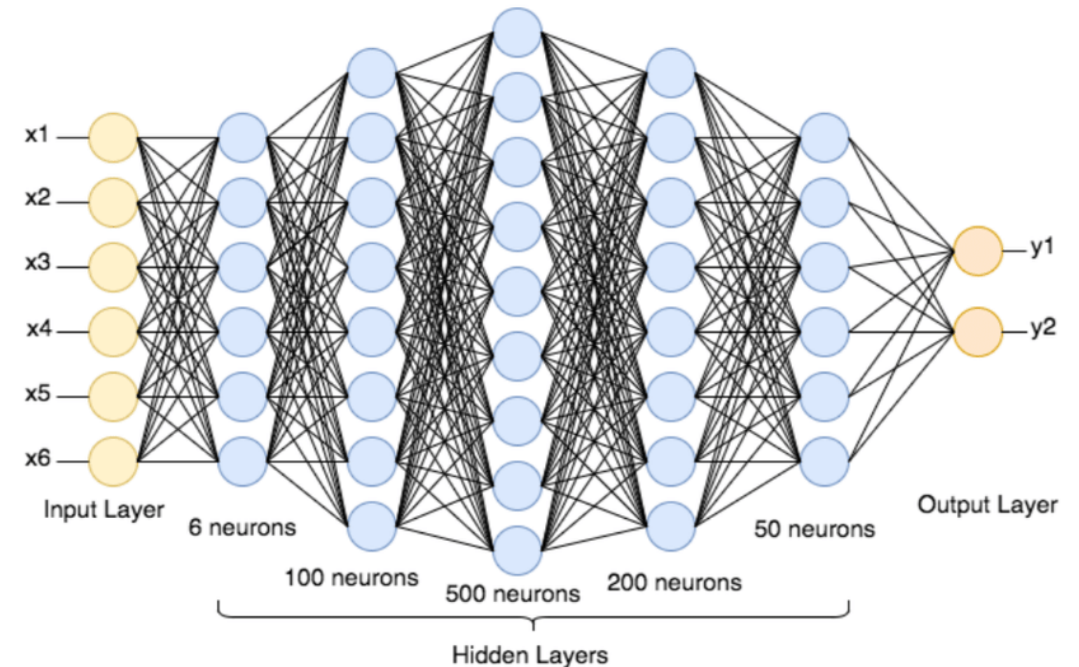
PERCEPTRON

- Bináris osztályozó
- A bemeneti vektorról eldönti, az osztály tagja-e
- A vektor elemeit súlyozva összegzi
- Egy réteg



DEEP LEARNING

- Neuronok hálózatba kapcsolása
- A perceptronnal ellentétben nem bináris a kimenet
- Egészen komplex problémák megoldására lehet használni
- Minél több réteg annál magasabb szintű feldolgozás lehetséges



DEEP LEARNING FPGA-N

- FPGA-n a neurális hálózatokat le lehet implementálni - minden neuron fizikailag létezni fog a chipen
 - Real-time applikációk, garantált válaszidővel
- Kompetitív teljesítmény/ár [\[Xilinx\]](#)
- Nagyon kedvező energia fogyasztás [\[Microsoft\]](#)
- Kevesen használják a HW tervezői tudás igénye miatt
- Programozás alapú HW leírás lehetséges: [LeFlow](#), [OpenVINO](#), [Vivado HLS](#)

DEEP LEARNING: FPGA vs GPU

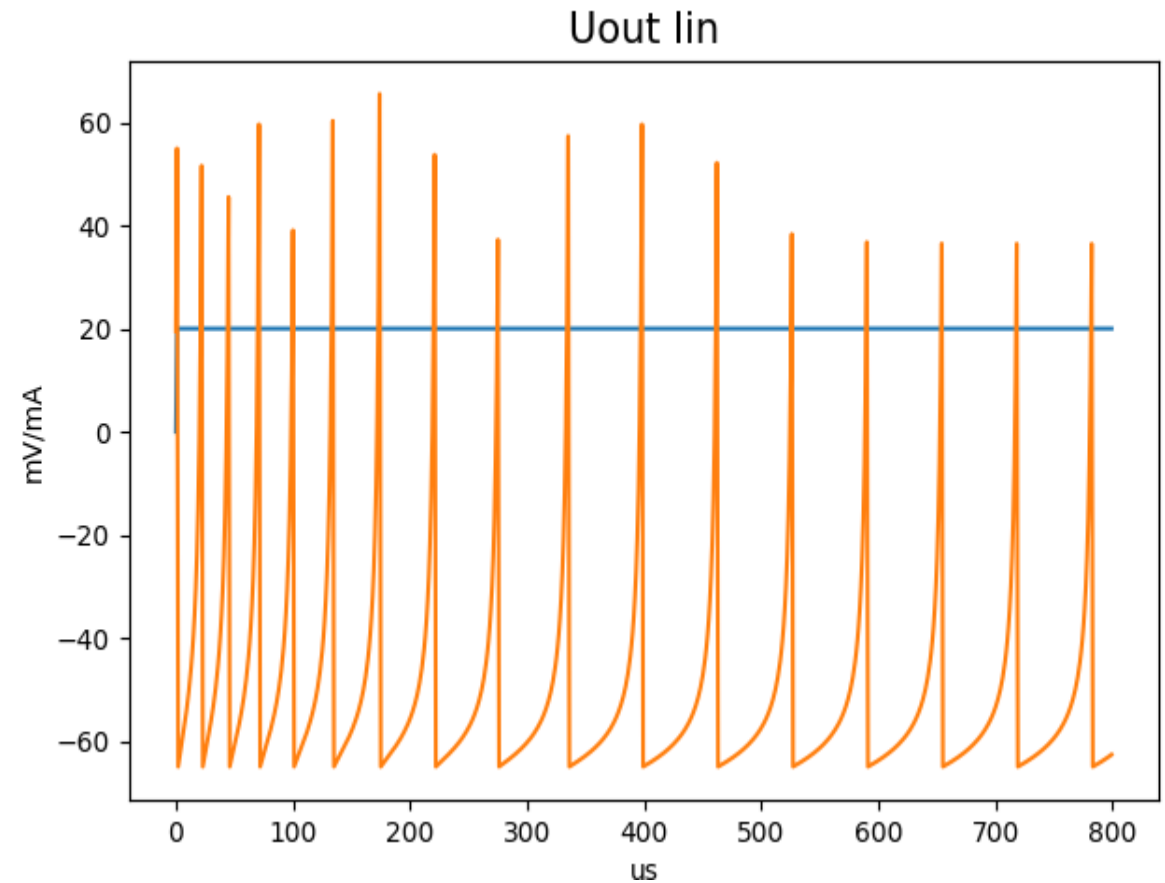
	FPGA	GPU
Ár	=	
Maximális komplexitás	>=	
Áram fogyasztás	<	
Real-time rendszerek	>	
Biztonság kritikus rendszerek	>>	
Fejlesztési idő	>	
Rendelkezésre álló munkaerő	<<	

SPIKING NEURAL NETWORK 1.

- Biológiai neuron model
- 1958: Hodgkin–Huxley neuron [modell](#) 4 tagú differenciál egyenlet rendszer
- 2003: Izhikevich neuron [modell](#) 2 tagú redukált diff. egyenlet rendszer

$$v' = 0.04v^2 + 5v + 140 - u + I$$

$$u' = a(bv - u)$$

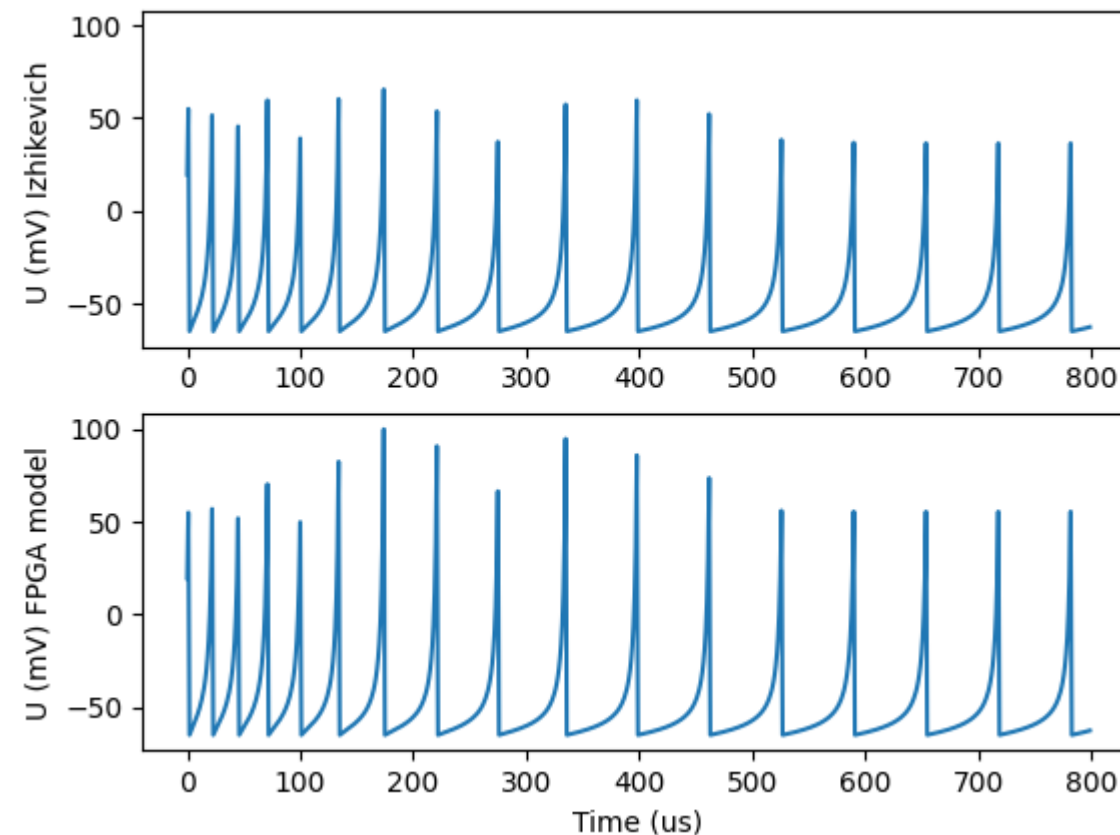


SPIKING NEURAL NETWORK 2.

- Jelenleg kutatási fázisban lévő technológia
- Memóriával rendelkező rendszer, natív stream feldolgozásra alkalmas
- Energia takarékos modell, nyugalmi állapotban alig fogyaszt
- Számítási kapacitása lényegesen magasabb, mint a második generációs hálózatoknak
- A tanítás nem megoldott
- A bemeneti adatok “tüske konverziója” nagyon kezdetleges

SPIKING NEURAL NETWORK FPGA-N

- Kutatási fázis
- Izhikevich model aproximációja minimális hibával, tanítás
- Tökéletesen illik a FPGA-ra
 - Natív órajel szintű determinizmus
 - Egyedi, konvolúcióra nem visszavezethető műveletek
- A FPGA hatalmas előnyben a GPU-val szemben; több nagy cég mixed-signal ASIC-eket gyárt



ÖSSZEFOGLALÁS

- FPGA-k az egyik legjobb lehetőséget biztosítják a nagy teljesítményű gépi tanulás implementációjára
- A 3. generációs neurális hálózatok és real-time rendszerek esetében ez különösen igaz
- Adatközponti környezetben - ahol az energia költség nagyobb, mint a HW költség – a legköltséghatékonyabb megoldás
- Legnagyobb kihívás a szakember hiány

KÖSZÖNÖM A FIGYELMET!

This work is licensed under the Creative Commons Nevezd meg! - Így add tovább! 4.0 Nemzetközi License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.