

Beszámoló

Zeffer Tamás

2024 június 19

Anyagtudományok és Technológiák Doktori Iskola (VI. félév)

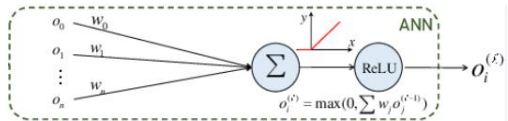
Belső konzulens: Csikósné Pap Andrea

Külső konzulens: Volk János

HUN-REN EK MFA, Nanoérzékelők Laboratórium (2022. októbertől)

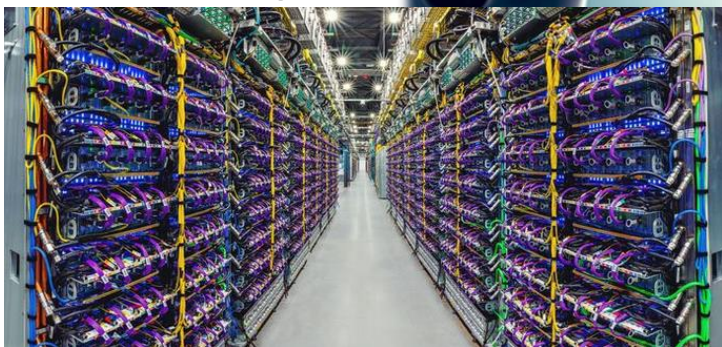
PhD téma: Vékonyréteg technológiával megvalósítható intelligens érzékelés

Motiváció: újszerű, neuromorf érzékelő-hardver

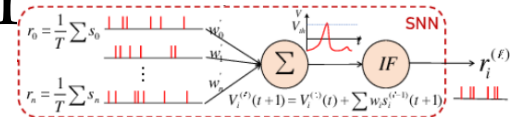


ANN (Artificial Neural Network)

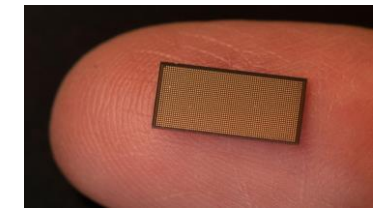
Az AI szomjas és falja az energiát
Gépi Tanulás



**zöldebb SNN⁽¹⁾ architektúrák
a cégek R&D fókuszban**



⁽¹⁾ SNN (Spiking Neural Network)
tüskés neurális hálózat



A nagy nyelvi modellek (LLM⁽²⁾) ökológiai lábnyoma

jelentős⁽³⁾

például a ChatGPT fogyasztása:

- betanítás: 700 000 liter tiszta édesvizet, 190 000 kWh energiát⁽⁴⁾
- használat közben: válaszonként 10...50 ml vizet
- a prognózis 2027-re: az AI globális fogyasztása Anglia éves vízfogyasztásának fele lesz

Emberi agy működését másoló hardverek:

- A biológiai agy mindössze 20 Wattot fogyaszt
- Néhány cég feltörekvő SNN hardvert kutat és tesztel: pl.: Intel Loihi 2
 - max. neuron szám 1 millió (31 mm² die area, bővíthetők)
 - [Intel Hala Point system](#): 1152 darab Loihi Chip 2600 Wattot fogyaszt
 - 2024 ápr. Intel beszámoló : Egy századnyi energiát fogyaszt és 50 x gyorsabbak vs. GPU és CPU szerverek

⁽²⁾ LLM (Large Language Models): AI-nagy nyelvi modellek

⁽³⁾ „arXiv:2023, Making AI Less ‘Thirsty’: Uncovering and Addressing the Secret Water Footprint of AI Models.”

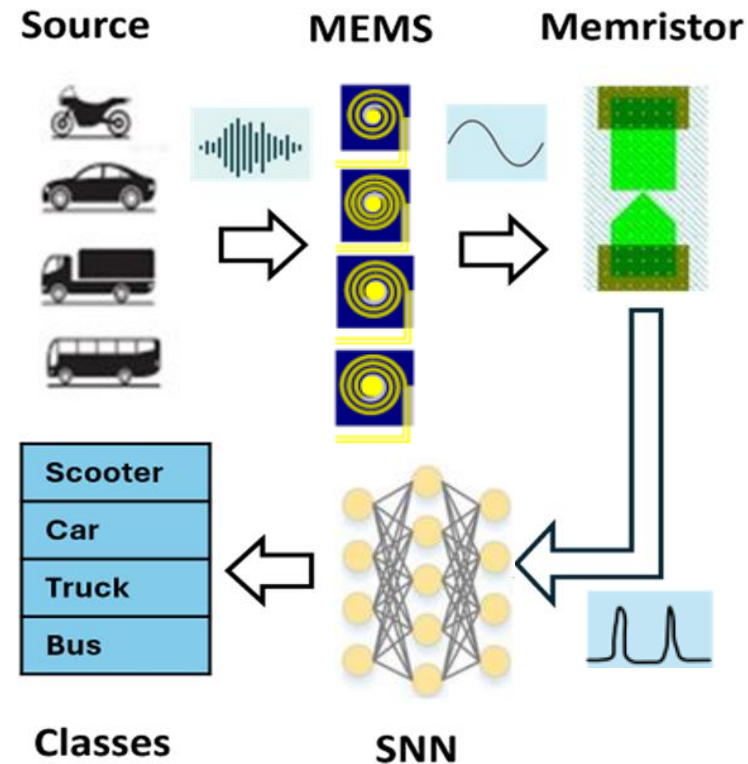
⁽⁴⁾ „IPCCC: 2023, Evaluating the Carbon Impact of Large Language Models at the Inference Stage”

Kutatásom Célja: SNN ökoszisztémába illeszkedő újszerű érzékelő létrehozása

Példa alkalmazás: SNN alapú **intelligens** (peremhálózati) **érzékelő** osztályozási funkcióval

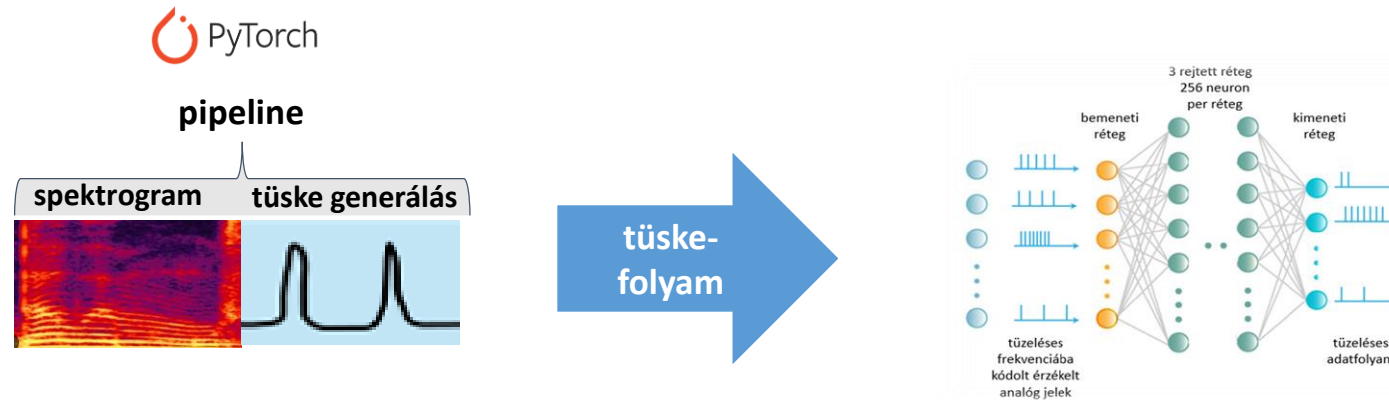
Főbb kutatási feladatok:

- DSP⁽¹⁾ mentes, spektrogram számítás kiváltása hangolt rezgőnyelvsorral
- SNN kompatibilis jeleket előállító memrisztoros interfész
- Az érzékelő-hw tervezéséhez SNN szoftveres analízis szükséges



(1) DSP: Digital Signal Processor (speciális célú processzorok)

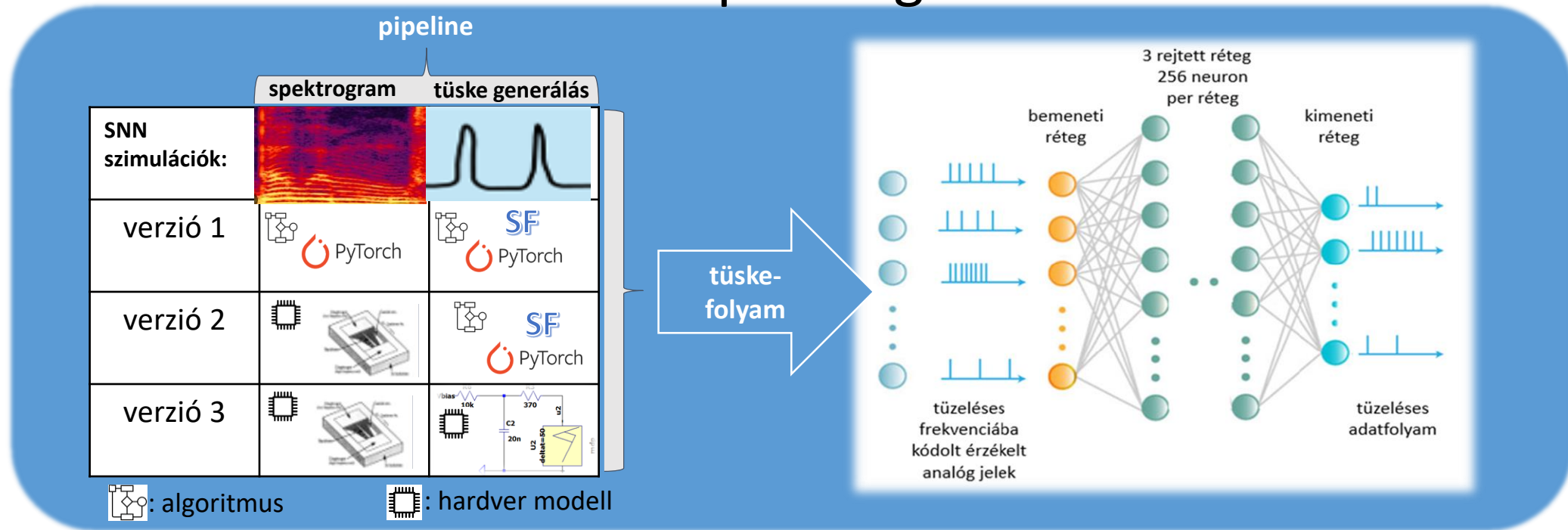
A harmadik generációs SNN hálózat és az érzékelő-hardver modellezése és együttes vizsgálata



- PyTorch és SNN Torch keretrendszerek segítségével
- Neurális modell tartalmaz 3 x 256 neuront, 3 rejtett rétegben (Speech2Spikes ⁽¹⁾ (S2S) cikk)
- Az előfeldolgozás két nagyobb részre bontható:
 - spektrogram és
 - tüske generálásra
- Az S2S pipeline ⁽¹⁾ algoritmusait a cikk mondatai alapján programozom. (nincs kód és hiányoznak matematikai leírások)
- A hálózatot a SZTAKI Titán nevű nvidia GPU hardverén futtatom

⁽¹⁾ „NICE 2023: Speech2Spikes: Efficient Audio Encoding Pipeline for Real-time Neuromorphic Systems”

Az érzékelő hardver és a 3. Generációs SNN hálózat együttes, modell alapú vizsgálata



A bemenő hangjelek előfeldolgozását három változatban végzem el (táblázat)

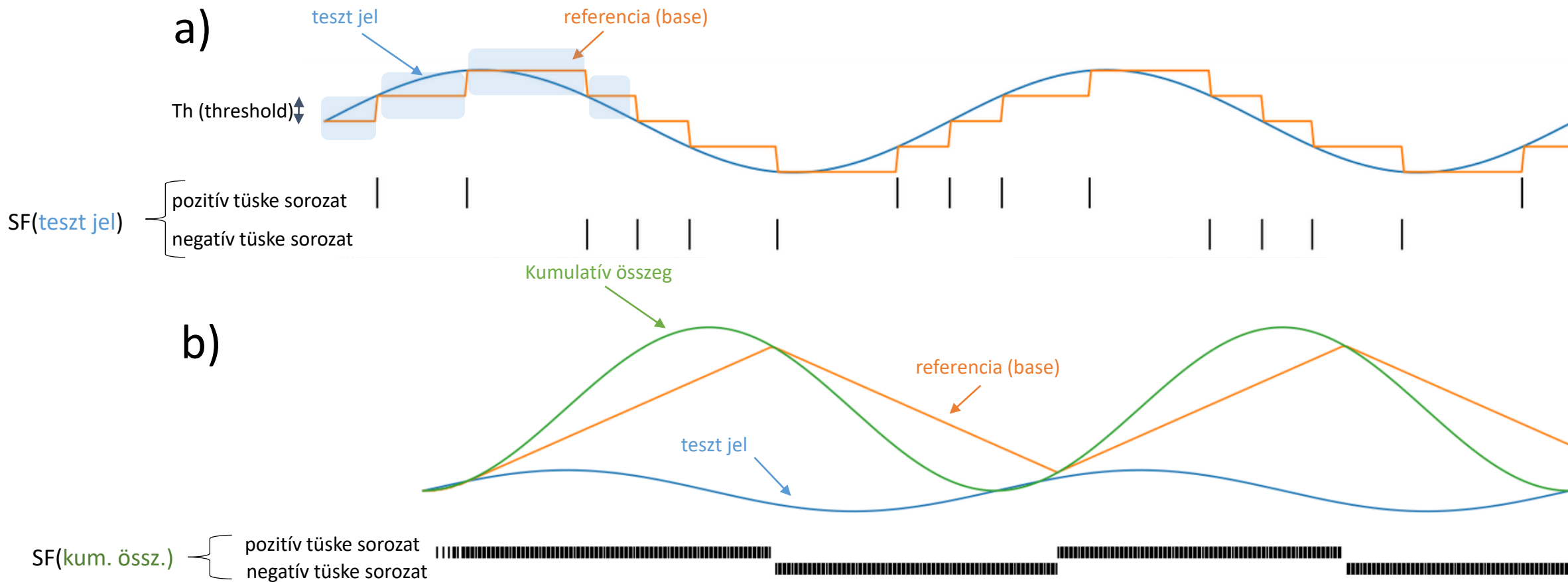
Vizsgálom a modell teljesítményét mindhárom esetre

- *verzió 1*: cikk alapján a Speech2Spikes⁽¹⁾ (S2S) pipeline alkalmazása: algoritmusok: spektrum + step-forward (SF)
- *verzió 2*: laborban *legyártott rezgőnyelvsor mért adataival* + az S2S algoritmust
- *verzió 3*: algoritmusok kiváltása a rezgőnyelvsor spektrumjával + a tuskéket generáló oszcilláló áramkör modelljével

⁽¹⁾ NICE 2023: Speech2Spikes: Efficient Audio Encoding Pipeline for Real-time Neuromorphic Systems

SF (Speech2Spikes) tüske-kódolás szinuszos teszttel függvényekkel

- SF(Szinusz értékek) -> a) ábrán: pozitív és negatív tüzelések => két neuronnak bemenet
- SF(CSUM(Szinusz értékek)) -> b) ábra: pozitív és negatív tüzelések => két neuronnak bemenet



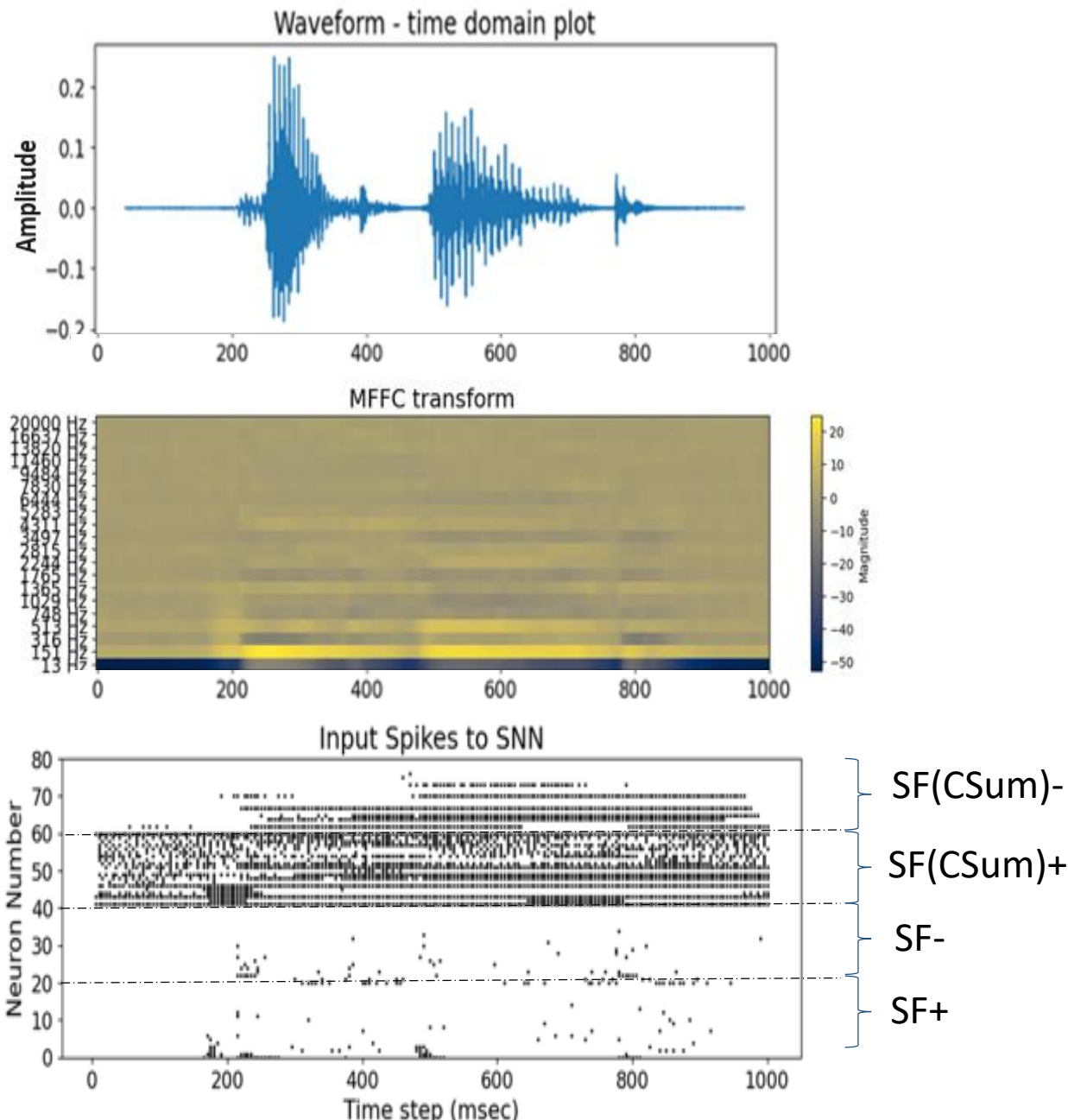
Verzió 1: Speech2Spikes algoritmus

Mel Spektrogram képzése:

1. Csuszóablakos Fourier-transzformáció alkalmazása a hangjelen
2. Amplitudót decibelbe, a frekvenciát mel skálába konvertálva (az emberi hallás érzékenységére optimalizált)

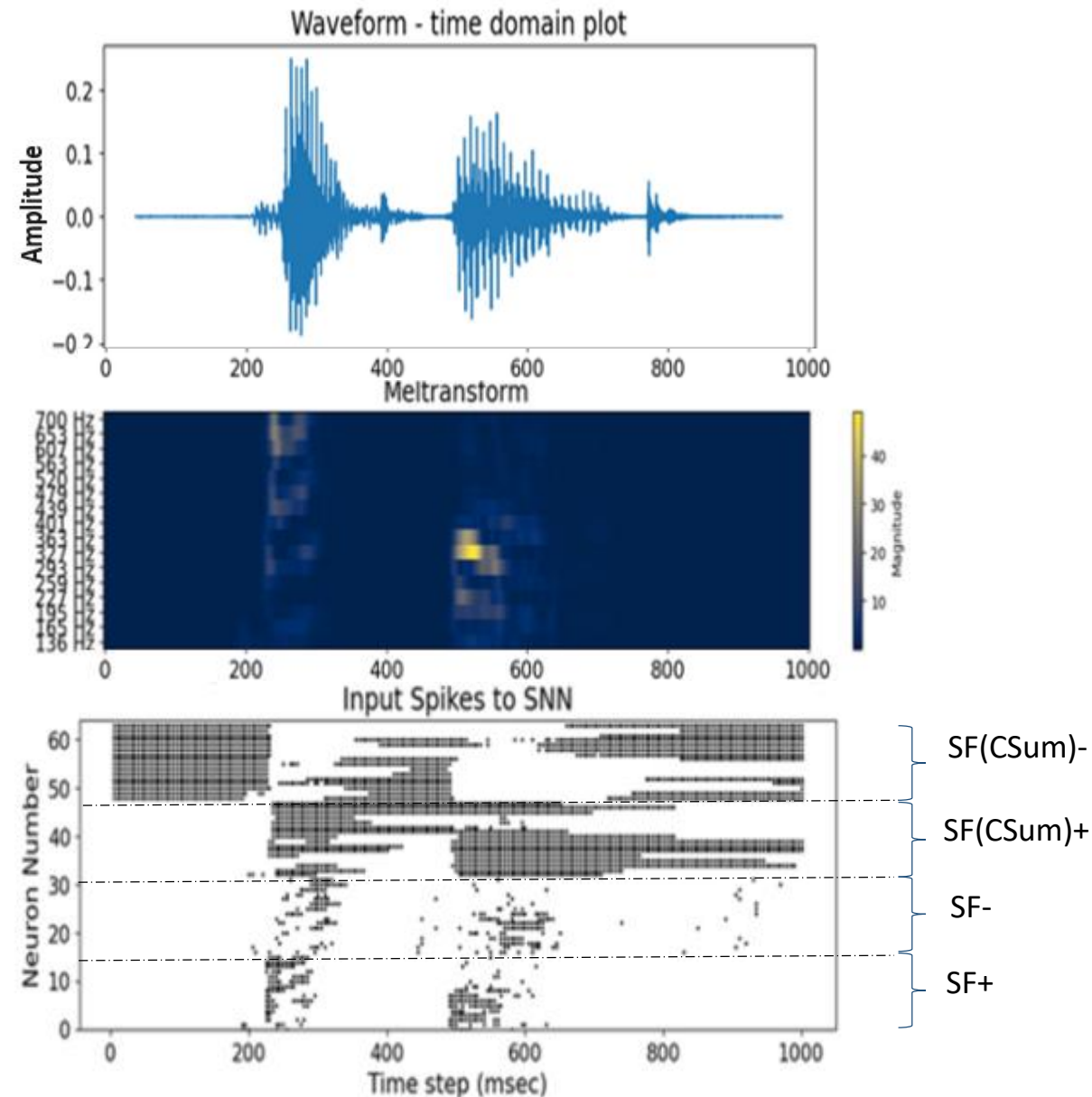
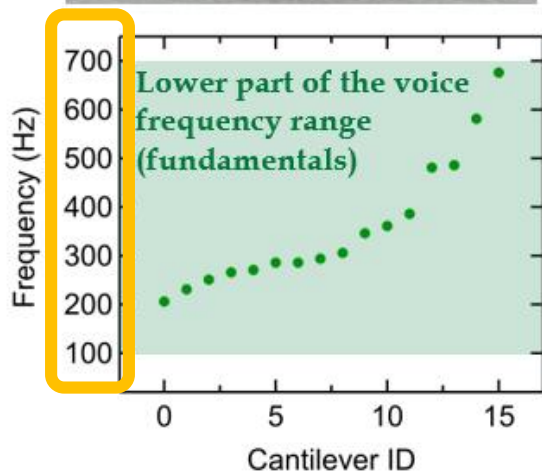
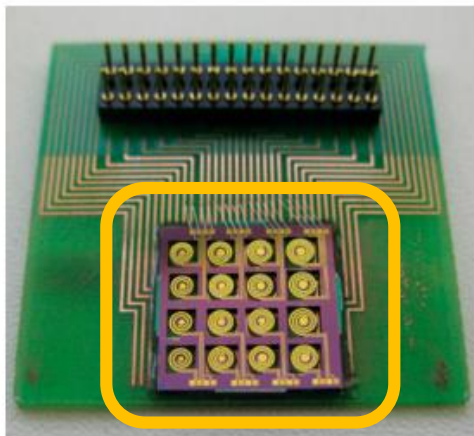
Tüske raszter képzése:

1. 4 sávos tüskeraszter a Step-Forward segítségével: SF-, SF+, SF(CSum)+, SF(CSum)-
2. sávonként 20 frekvencia csatorna
3. 4x20 tüskesor az SNN 80 bemeneti neuron számára



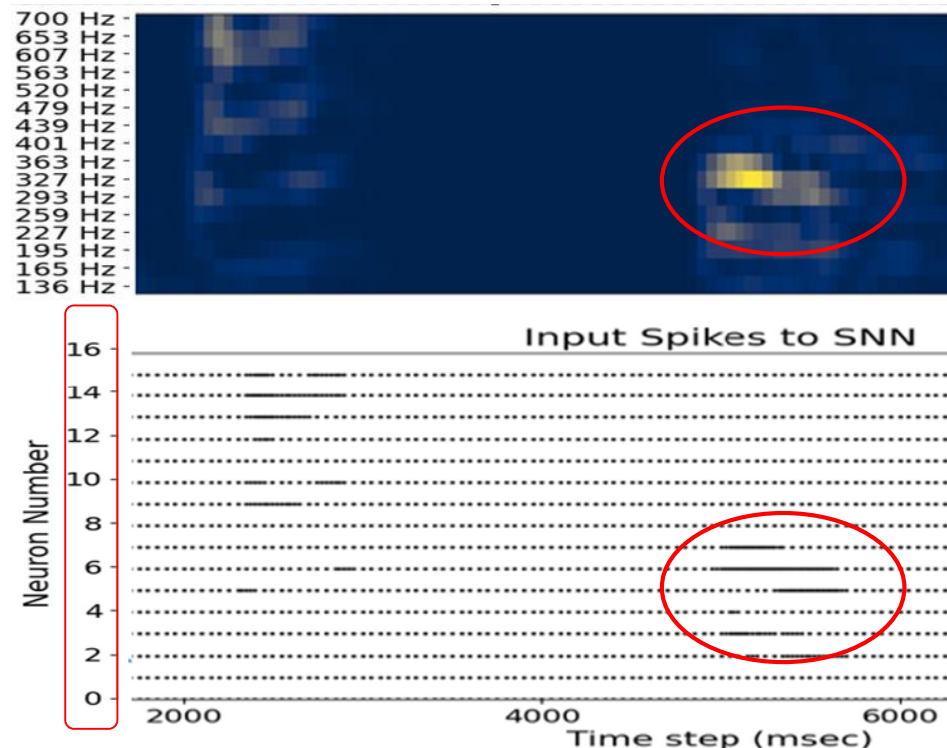
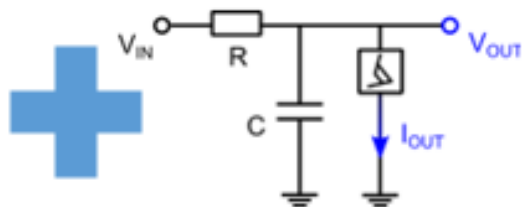
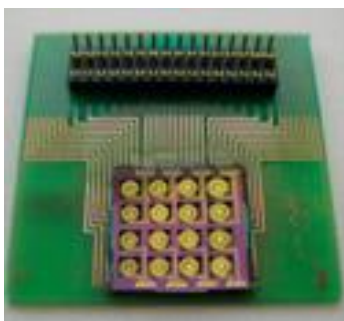
Verzió 2: S2S algoritmus szűkebb frekvenciatartománnyal

- Laborban mért MEMS rezgőnyelv-sort modellezem
- 200 - 700Hz frekvenciatartományt és 16 csatornát állítok be a spektrogram előállításához
- Továbbra is az s2s algoritmust alkalmazom -> 4x16 neuron bemenet



Verzió 3: frekvencia túskekódolás alkalmazása

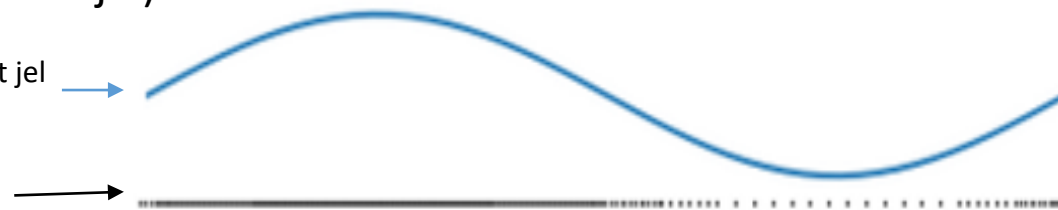
- Továbbra is 16 csatorna 200-700 Hz tartományban
- A túskeket az előző szemeszterben bemutatott memrisztoros oszcillációs hardver állítja elő
- mindezt a rate-encoding féle kódolásnak megfelelően
- a túskek időbeni gyakorisága arányos a csatornák pillanatnyi amplitúdójával (rate encoding)
- Az idegi hálózat bemenetének száma a hardverben 16, azaz a negyedére csökken










RATE ENCODING: Jelintenzitás frekvencia kódolása (pozitív értékű, szinuszos tesztjel):

szinuszos teszt jel

túske sorozat



Eredmény: a hálózat százalékos pontossággal predikciót végez

elő-feldolgozás verzió	Mel bin sávok száma	frekvencia-tartomány [Hz]	tüske kódolása	modell teljesítmény
Ver. 1	 20	 20...16.000	 S2S	70%
Ver. 2	 16	 200-700	 S2S	40%
Ver. 3			 frekvencia	Nincs

 : algoritmus

 : hardver modell

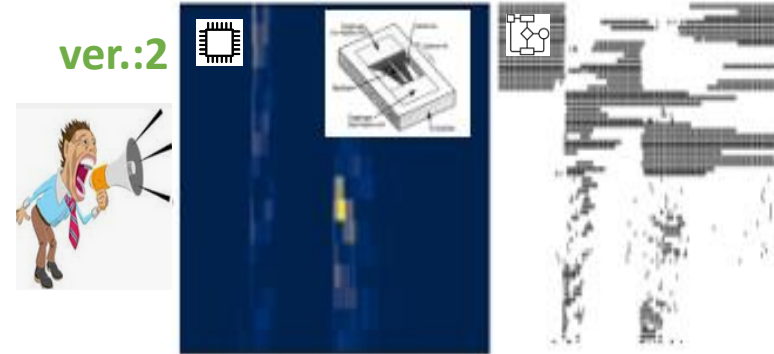
A validációs adatpontok százalékos pontossága:

- Verzió 1: 70% kontra cikk adat (88.5%)
- Verzió 2: 40%
- Verzió 3: nem tanul, lehetséges okok:
 1. szegényebb az információ (csak 16 bemeneti neuron)
 2. hardver kódolás elégtelensége

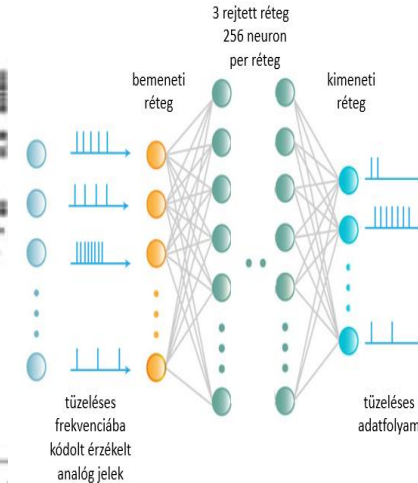
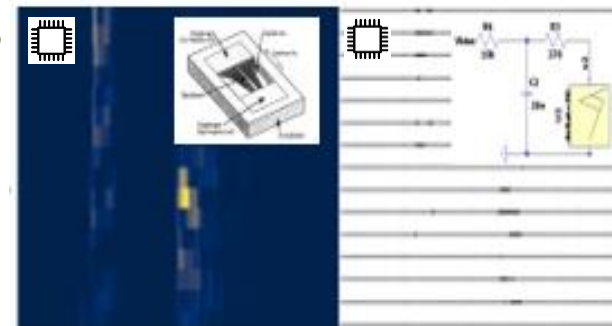
ver.:1



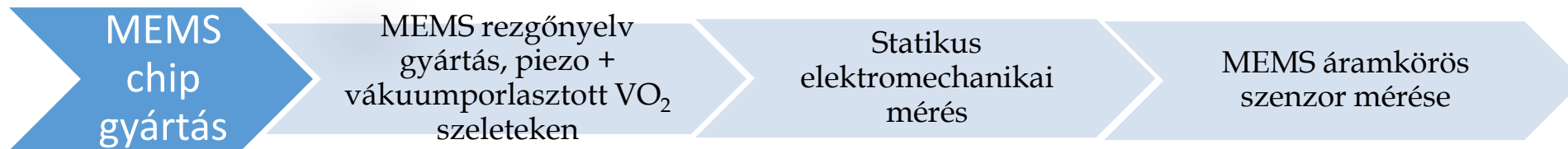
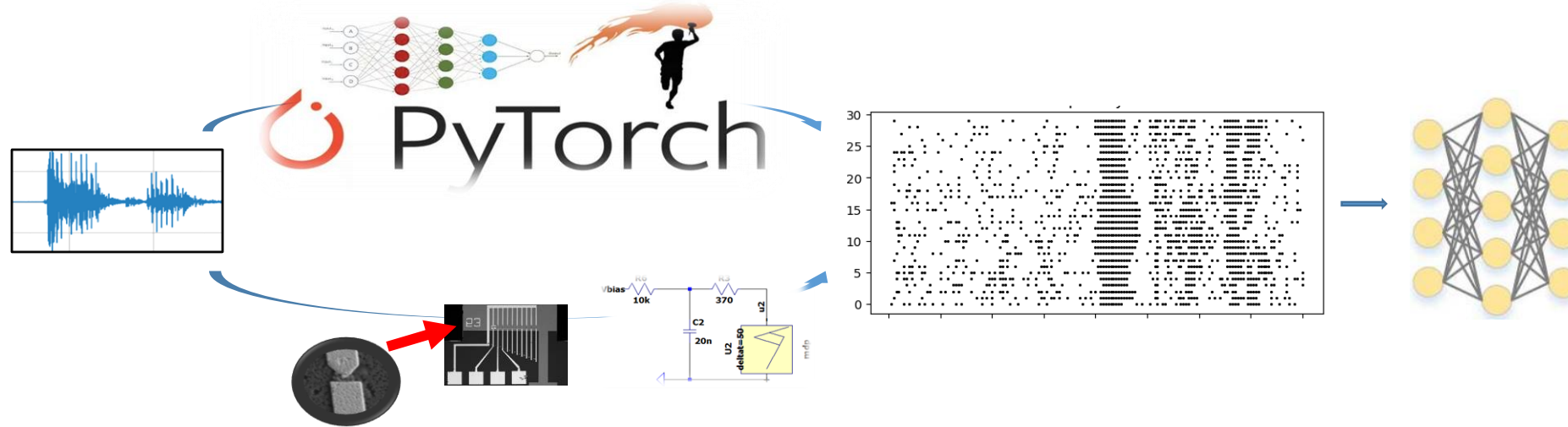
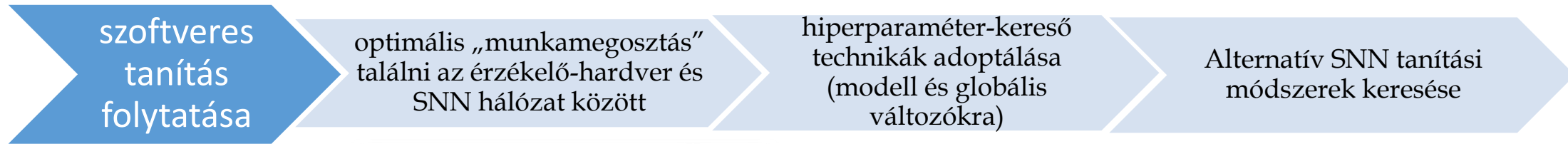
ver.:2



ver.:3



Kitekintés



Elfogadott konferencia absztrakt 2024: "Towards fully hardware-based neuromorphic encoding for efficient vibration signal recognition", EUROSENSORS

Tervezett cikkek

1. Statikus mérésből egy **anyagtudományi** folyóirat cikk
2. Dinamikus mérésből egy **szenzoros** folyóirat cikk (például: Sensors and Actuators)
3. IEEE cikk a MEMS chip rendszerszintű alkalmazásáról

Vége

Köszönöm a figyelmet.