

Giliojo mokymo modeliai kalbos signalui gryninti

Parengė: Daniel Zakševski

Darbo vadovas: Doc. Dr. Gintautas Tamulevičius

Doktorantūros studijų metai: 2023-2027

Studijų metai: 3-ieji

Doktorantūros studijų ir mokslinių tyrimų plano bei jo vykdymo suvestinė

Studijų metai	Egzaminai	
	Planas	Ivykdyta
I (2023/2024)	2	2
II (2024/2025)	1	1
III (2025/2026)	1	1
IV (2026/2027)	0	0
Iš viso:	4	4

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citavimo rodikliu			Be citavimo rodiklio		
	Planas	Ivykdyta	Planas	Ivykdyta	Planas	Ivykdyta	Būklė	Planas	Ivykdyta	Būklė
I (2023/2024)	0		0		0			0		
II (2024/2025)	0	1	0	2	0			1	1	Priimta
III (2025/2026)	1		1		1			0		
IV (2026/2027)	2		1		1			0		
Iš viso:	3	1	2	2	2			1		

Ataskaitinio pusmečio darbo planas ir jo vykdymas, 2024/2025 – II pusmetis

Egzaminai, 2025/2026 – I pusmetis		
Planas	Įvykdyta	Būklė
Kalbos signalų apdorojimas	Kalbos signalų apdorojimas, 2026-01-16	Išlaikytas

Dalyvavimas konferencijose, 2025/2026 – I pusmetis		
Planas	Įvykdyta	Konferencijos tipas
„6th International Conference on Communications, Information, Electronic and Energy Systems“, 2025 m. lapkričio 26 – 28 d., Rusės m., Bulgarija	D. Zakševski, G. Korvel, G. Tamulevičius, „A Decoupled Biophysically-Inspired Architecture for Speech Enhancement“, „6th International Conference on Communications, Information, Electronic and Energy Systems“, 2025 m. lapkričio 26 – 28 d., Rusės m. Bulgarija	Tarptautinė

Publikacijos, 2025/2026 – I pusmetis			
Planas	Įvykdyta	Būklė	Publikacijos tipas
2025 6th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)	D. Zakševski, G. Korvel and G. Tamulevičius, "A Decoupled Biophysically-Inspired Architecture for Speech Enhancement," <i>2025 6th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)</i> , Ruse, Bulgaria, 2025, pp. 1-5, doi: 10.1109/CIEES66347.2025.11300204.	Publikuota	Be cit. rodiklio

Doktorantūros studijų pasiekimai, tarptautiniai renginiai ir publikacijos

Dalyvavimas tarptautinėse konferencijose	
	Aprašas
1.	D. Zakševski, G. Korvel, G. Tamulevičius, „A Decoupled Biophysically-Inspired Architecture for Speech Enhancement“, „6th International Conference on Communications, Information, Electronic and Energy Systems“, 2025 m., lapkričio 26 – 28 d., Rusės m., Bulgarija

Publikacijos su citavimo rodikliu	
	Bibliografinis aprašas

Mokslinių tyrimų ir disertacijos rengimo etapai (1)

Darbo pavadinimas		Atlikimo terminai		Pastabos
		Nuo	Iki	
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):			1.1., 1.2., 1.3. įvykdyti, 1.4. vykdomas
1.1.	Kalbos signalo gryninimo problematikos apžvalga	2023 m. IV ketvirtis	2024 m. I ketvirtis	
1.2.	Kalbos signalo gryninimo problemos klasikinių sprendimų metodikos analizė	2024 m. I ketvirtis	2024 m. I ketvirtis	
1.3.	Kalbos signalo gryninimo problemos mašininio mokymu grįstų sprendimų metodikos analizė	2024 m. I ketvirtis	2024 m. II ketvirtis	
1.4.	Išankstinio duomenų apdorojimo metodų analizė	2024 m. I ketvirtis	2024 m. III ketvirtis	
2.	Mokslinio tyrimo vykdymas:			2.1.1 įvykdytas
2.1.	Tyrimo metodikos sudarymas:			
2.1.1.	Tyrimų tikslų ir užduočių suformulavimas, tyrimo metodikos sudarymas	2024 m. II ketvirtis	2024 m. III ketvirtis	
2.2.	Teorinis tyrimas:			2.2.1., 2.2.2. įvykdyti
2.2.1.	Klasikinių bei mašininio mokymu grįstų metodų palyginimas kalbos signalo gryninimo uždavinio sprendimui	2024 m. III ketvirtis	2025 m. I ketvirtis	
2.2.2.	Giliojo mokymo metodais grįstų kalbos signalo gryninimo algoritmų apibendrinimas	2025 m. I ketvirtis	2025 m. III ketvirtis	

Mokslinių tyrimų ir disertacijos rengimo etapai (2)

Darbo pavadinimas		Atlikimo terminai		Pastabos
		Nuo	Iki	
2.3.	Empirinis tyrimas:			2.3.1. įvykdytas, 2.3.2 pradėtas vykdyti
2.3.1.	Skirtingomis architektūromis paremtų giliojo mokymo algoritmų tyrimas	2025 m. III ketvirtis	2026 m. I ketvirtis	
2.3.2.	Giliojo mokymo metodais grįsto kalbos gryninimo algoritmų praktinio taikymo problematikos tyrimas	2026 m. I ketvirtis	2026 m. IV ketvirtis	
2.4.	Gautų duomenų analizė, apibendrinimas, išvadų parengimas:			
2.4.1.	Gautų duomenų analizė, apibendrinimas, išvadų parengimas	2026 m. IV ketvirtis	2027 m. III ketvirtis	
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:			
3.1.	Literatūros apžvalga	2026 m. IV ketvirtis	2027 m. I ketvirtis	
3.2.	Tyrimų metodikos aprašymas	2027 m. I ketvirtis	2027 m. II ketvirtis	
3.3.	Rezultatų aprašymas	2027 m. II ketvirtis	2027 m. III ketvirtis	
3.4.	Ginamų teiginių ir išvadų aprašymas	2027 m. III ketvirtis	2027 m. IV ketvirtis	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2027 m. IV ketvirtis	2027 m. IV ketvirtis	
5.	Daktaro disertacijos gynimas	2027 m. IV ketvirtis	2027 m. IV ketvirtis	

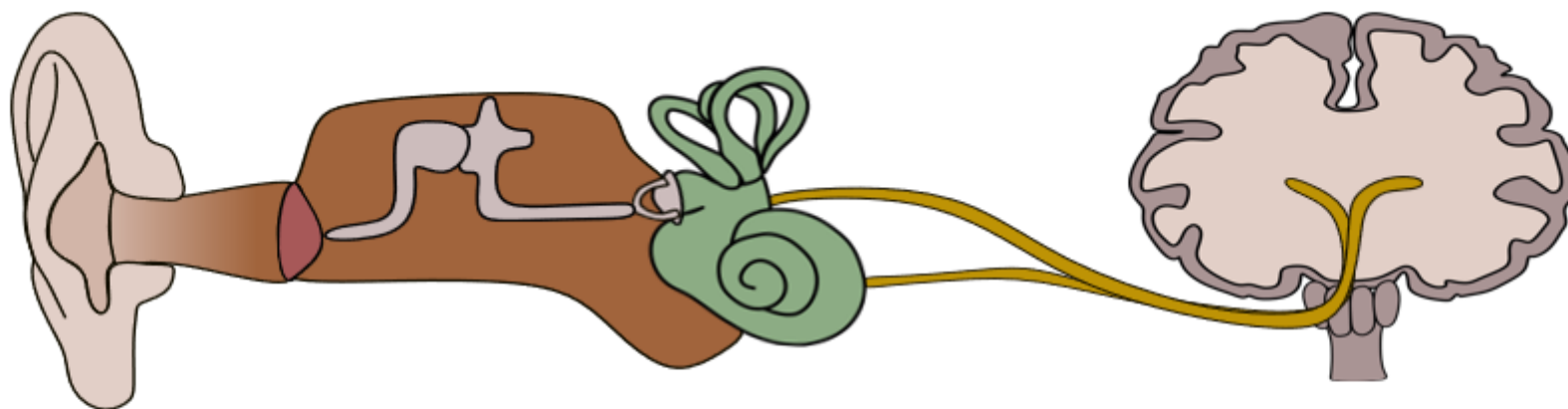
Tyrimo detalės

- **Tyrimo objektas:** šnekos signalas, balso signalo kokybė bei triukšmas.
- **Tyrimo tikslas:** pasiūlyti ir pritaikyti klausos modeliais grįstus giliojo mokymo tinklus kalbos signalui gryninti.
- **Tyrimo uždaviniai:**
 1. Įvertinti giliojo mokymo modelius naudojamus kalbos signalui gryninti.
 2. Apžvelgti ir įvertinti žmogiškosios klausos modelius.
 3. Integruoti klausos modelius, kitas žinias į kalbos signalo gryninimo modelį.
 4. Įvertinti pasiūlytųjų modelių modifikacijų efektyvumą, signalo gryninimo kokybę.

Moksliniai rezultatai, 2025/2026 – I pusmetis (1)

- Pagrindinis kalbos gryninimo (angl. speech enhancement) tikslas – padidinti akustinių kalbos signalų kokybę ir suprantamumą, pašalinant įvairius triukšmus bei jų sukeltus iškraipymus, atkuriant ar išryškinant spektrines kalbos signalo savybes.
- Geriausia šnekos gryninimo sistema yra žmogaus klausa. Savybių, leidžiančių žmogaus klausai efektyviai veikti, integravimas į skaitmeninius šnekos gryninimo algoritmus, potencialiai leistų toliau juos tobulinti.

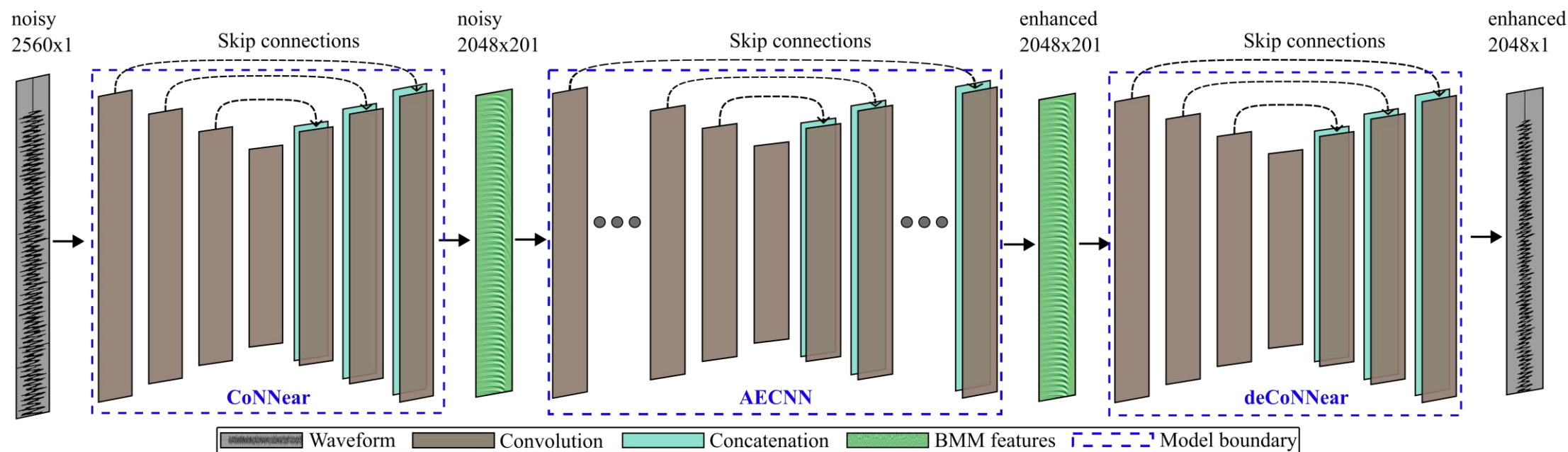
Moksliniai rezultatai, 2025/2026 – I pusmetis (2)



	Išorinė ausis	Vidurinė ausis	Vidinė ausis	Smegenų kamienas ir smegenys
Funkcija:	Garso surinkimas	Vibracijų sustiprinimas	Dažnių analizė	Garso suvokimas ir interpretavimas

1 pav. Žmogaus klausos sistema, jos dalių funkcijos

Moksliniai rezultatai, 2025/2026 – I pusmetis (3)



2 pav. Pasiūlytas 3-jų dalių šnekos gryninimo giliojo mokymosi modelis – d-CoSE.

Sudarytas iš 3 dalių:

CoNNear – požymių (*BMM*) ištraukimas iš signalo (pagrindinės (baziliarinės) membranos nervinių šaknelių poslinkių vertės)

AECNN – triukšmo šalinimas, šnekos kokybės gerinimas požymių erdvėje

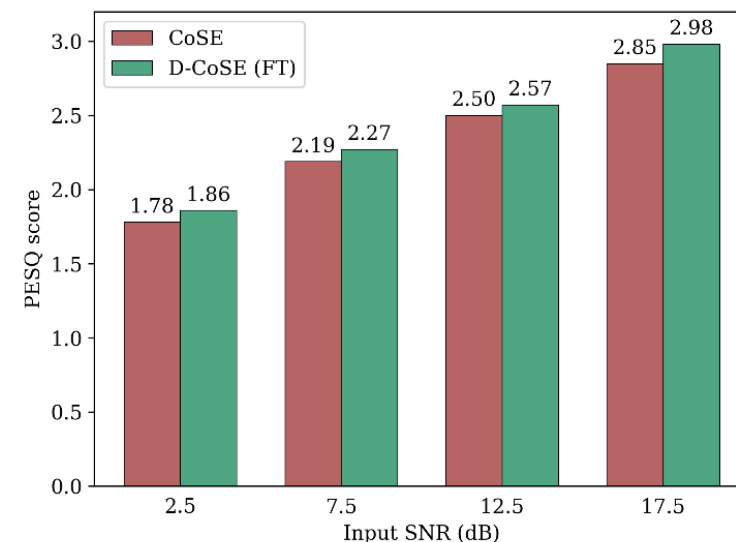
deCoNNear – signalo atstatymas iš BMM požymių

Moksliniai rezultatai, 2025/2026 – I pusmetis (4)

- Pasiūlyta d-CoSE sistema – moduliari, biofizikos principais grindžiama architektūra, kuri šnekos gryninimo procesą suskirsto į analizės, gryninimo ir sintezės etapus.

Model	Loss function	SNR, dB	SSNR, dB	PESQ	STOI
Noisy speech	–	8.39	-0.50	1.94	0.76
CoSE	$L1_{\text{waveform}}$	18.35	11.63	2.33	0.93
d-CoSE	$L1_{\text{waveform}}$	16.68	10.59	2.26	0.92
d-CoSE	$L1_{\text{BMM}}$	0.92	-5.99	1.69	0.91
d-CoSE	$L1_{\text{hybrid}}$	15.78	9.62	2.30	0.92
d-CoSE ^{FT}	$L1_{\text{hybrid}}$	17.83	11.31	2.42	0.93

1 lentelė. Monolitinės (CoSE) ir atskirtos (d-CoSE) architektūrų vertinimas taikant įvairias mokymo konfigūracijas.



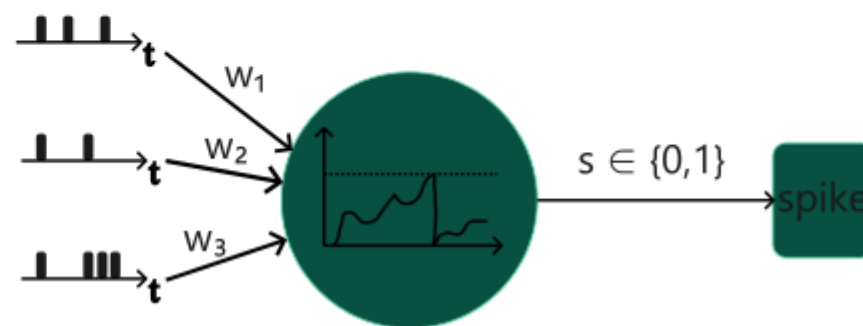
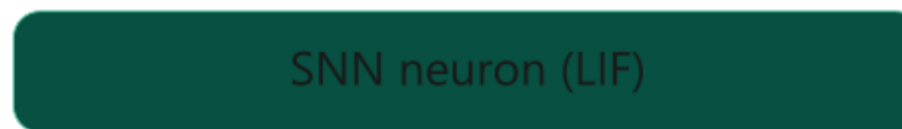
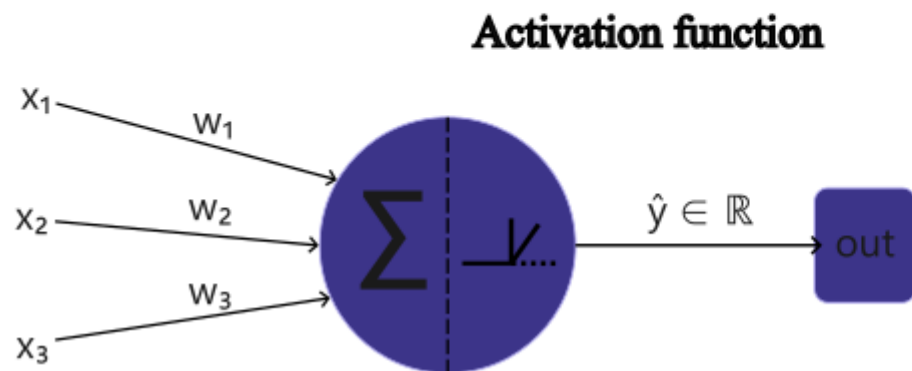
3 pav. CoSE ir d-CoSE^{FT} modelių PESQ įverčio palyginimas esant skirtingam pradinio signalo SNR.

Moksliniai rezultatai, 2025/2026 – I pusmetis (5)

- Šnekos gryninimo dalis pakeičiama impulsiniu neuroniniu tinklu (angl. Spiking Neural Network).
- **Privalumai:** energinis našumas, tinkluose veikia laiko, kaip parametro, dedamoji, signalų negausumas, įvykiais grįstas veikimas.
- **Trūkumai:** treniravimo sudėtingumas, SNN šnekos signalo gryninimo uždaviniuose atsilieka nuo ANN, pilnam paleidimui reikalinga speciali įranga, tolydžių įvesties duomenų kodavimo problema.

Moksliniai rezultatai, 2025/2026 – I pusmetis (6)

- Gauna įvestį, turi svorinius koeficientus, išvestį
- Gali būti jungiami į sluoksnius



4 pav. Dirbtinio ir impulsinio neuronų veikimo principo palyginimas

Moksliniai rezultatai, 2025/2026 – I pusmetis (7)

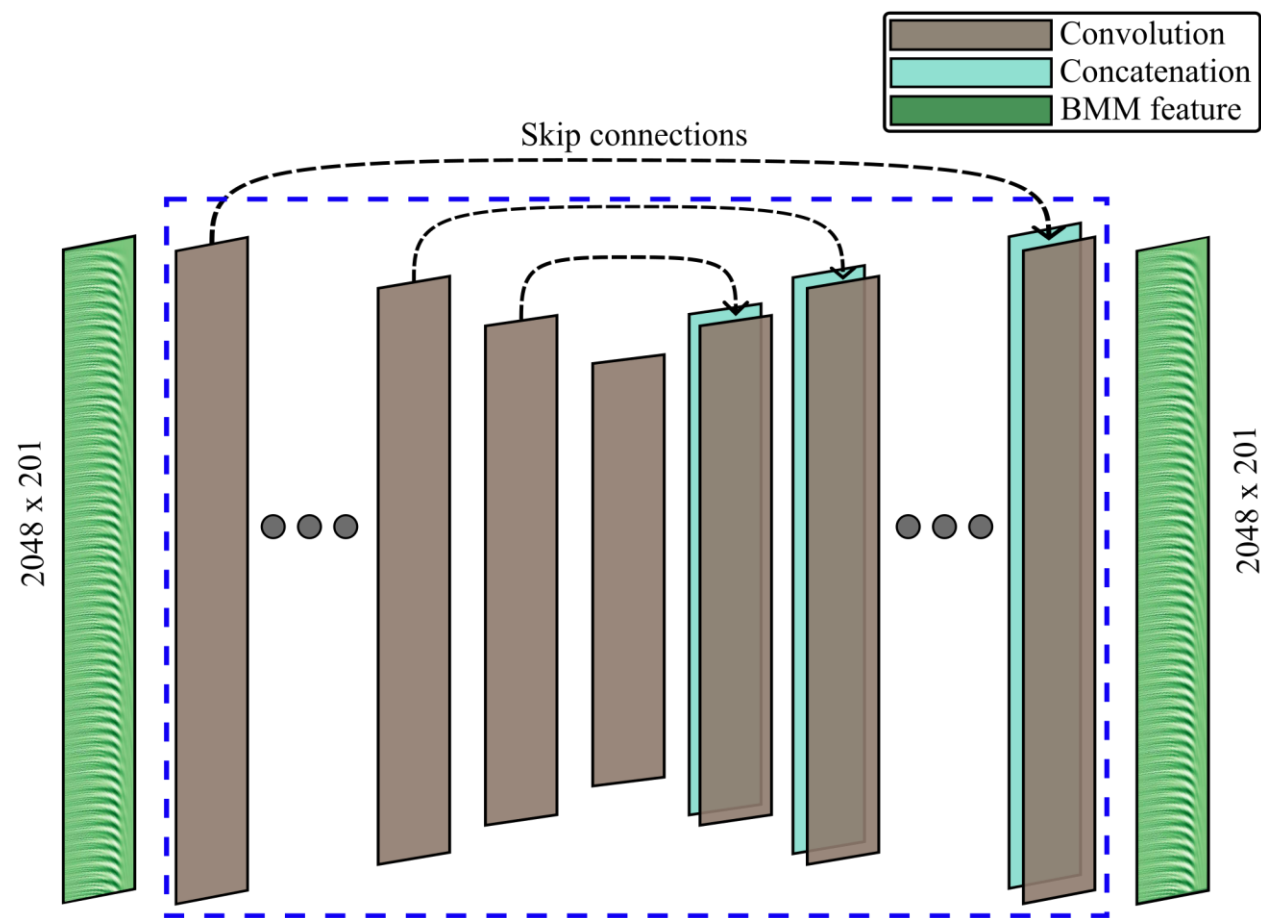
	Dirbtinis neuronas	Impulsinis neuronas
Aktyvacijos funkcija	tolydi	impulsas
Informacijos perdavimas	slankiojo kablelio skaičiai	impulsinis kodavimas
Laiko dedamoji	diskretūs laiko žingsniai	būdinga laikinė dinamika
Skaičiavimai	tankūs, sinchroniniai	reti, grįsti įvykiais
Tikslinė įranga	vaizdo plokštės	neuromorfiniai procesoriai

2 lentelė. Dirbtinio ir impulsinio neuronų techniniai, taikymo skirtumai

Moksliniai rezultatai, 2025/2026 – I pusmetis (8)

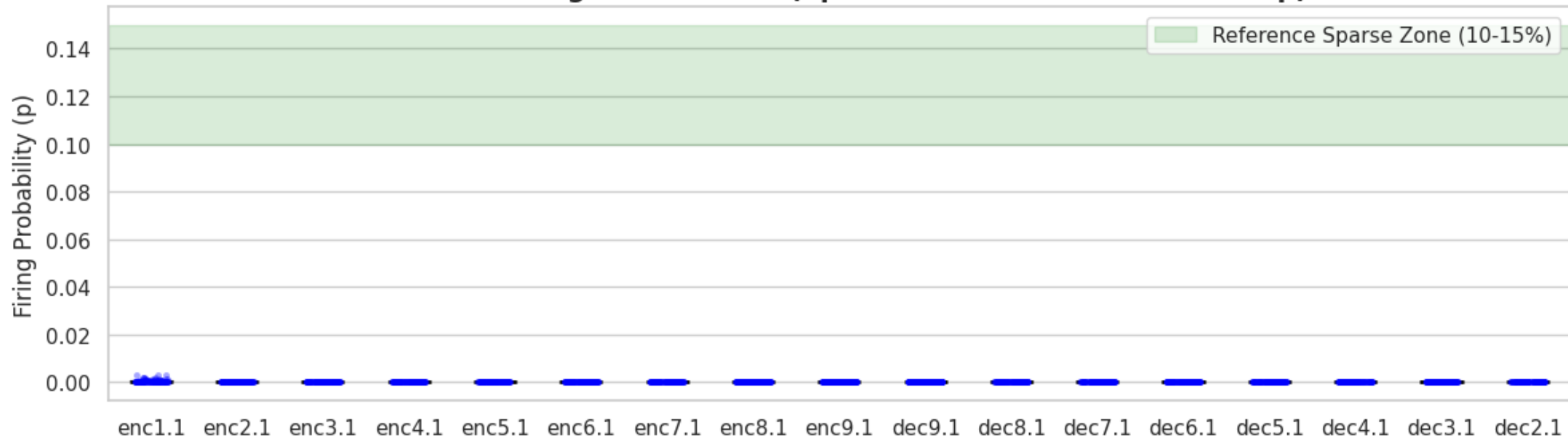
- Galima išskirti tam tikrus svarbius parametrus:
- Slenkstinė vertė V_{th} – membranos potencialas, kurį pasiekus įvyksta impulsas
- Nuotėkio parametras τ – kaip greitai neuronas „pamiršta“ ankstesnių laiko žingsnių poveikį
- Simuliacijos trukmė T – diskrečių simuliacijos laiko vienetų kiekis

Moksliniai rezultatai, 2025/2026 – I pusmetis (9)



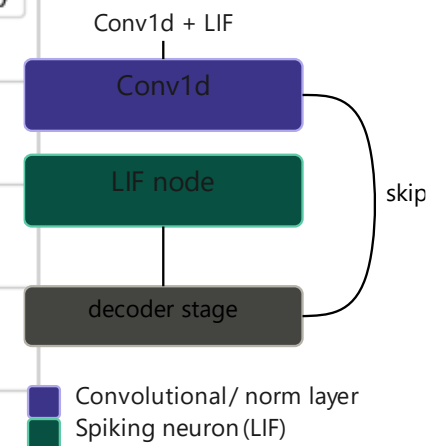
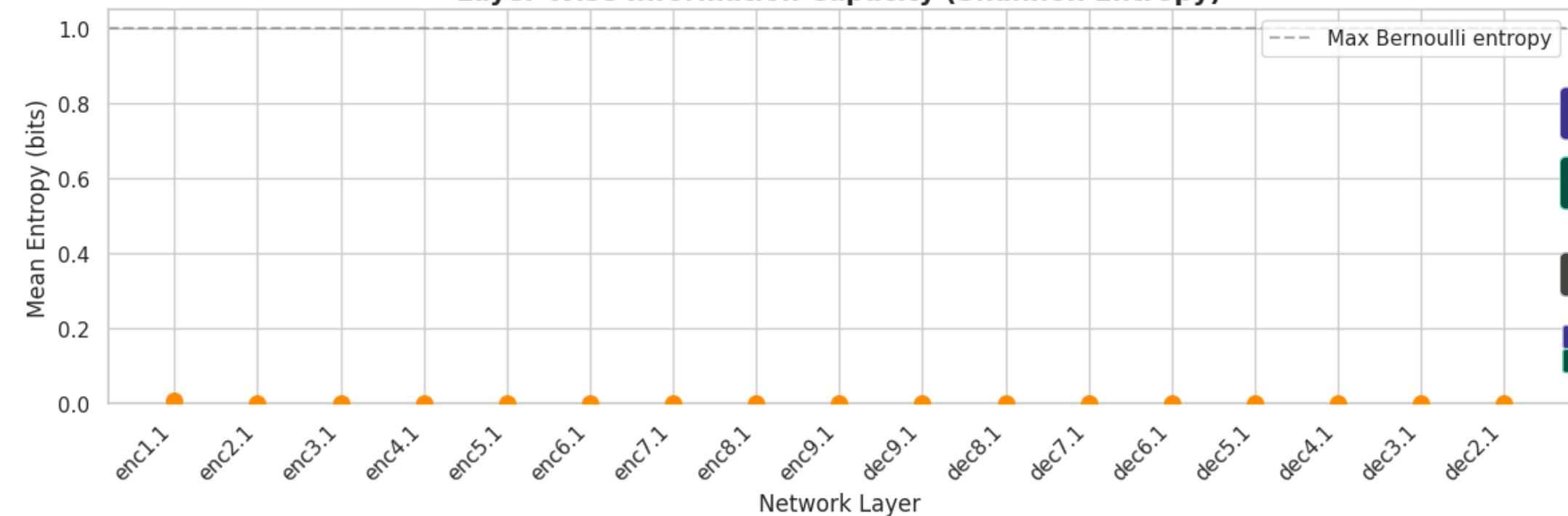
5 pav. AECNN – šnekos gryninimo modelis, veikiantis požymių (BMM) erdvėje

Channel Firing Probabilities (Spatial Distribution of Bernoulli p)

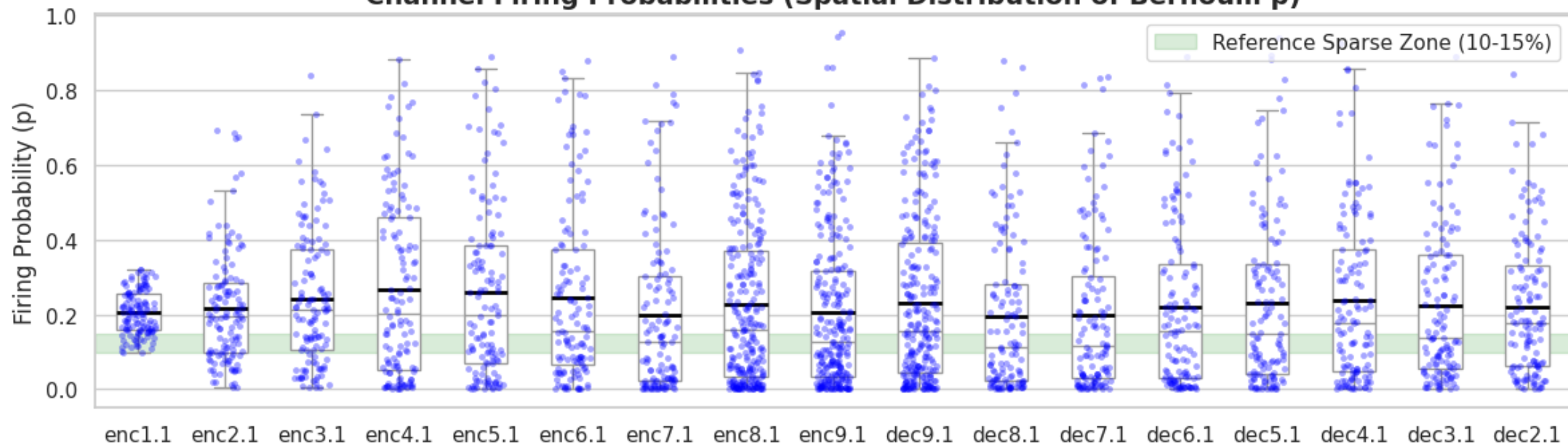


$$V_{th} = 1.0$$
$$T = 4$$

Layer-Wise Information Capacity (Shannon Entropy)

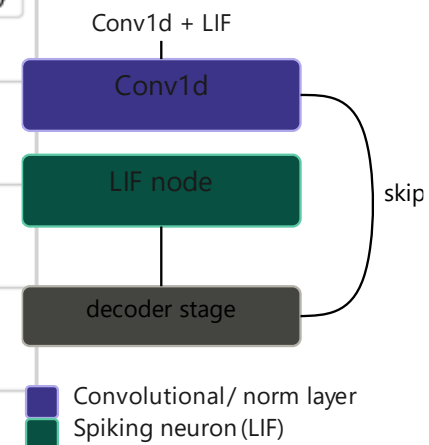
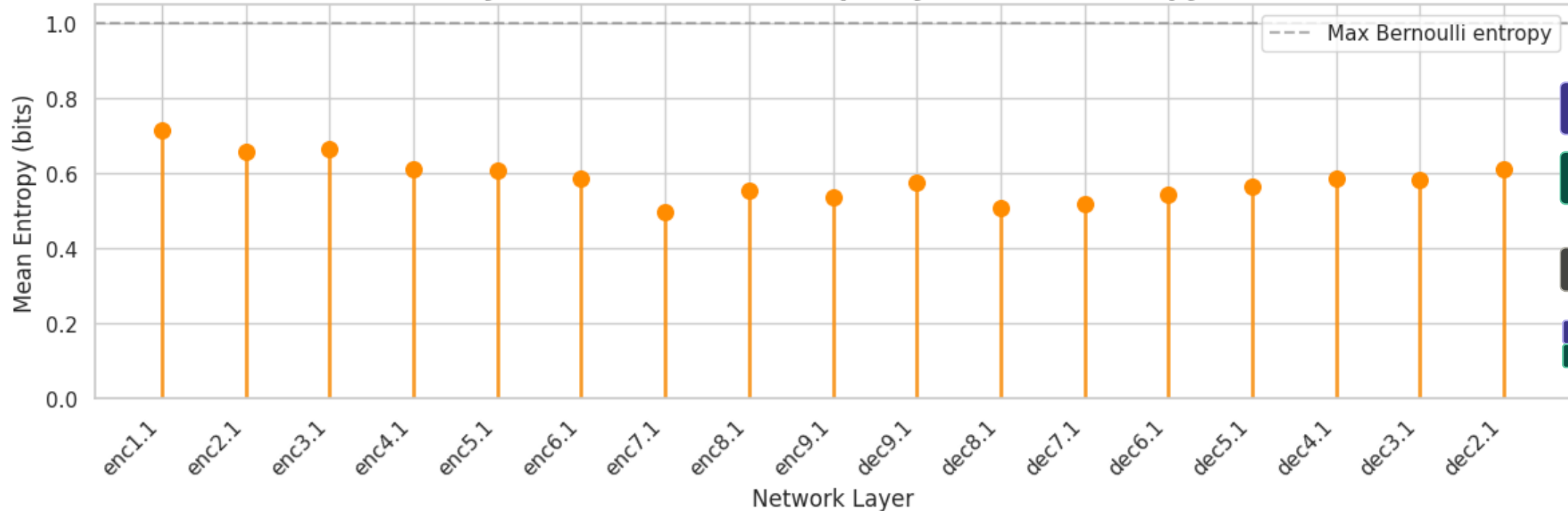


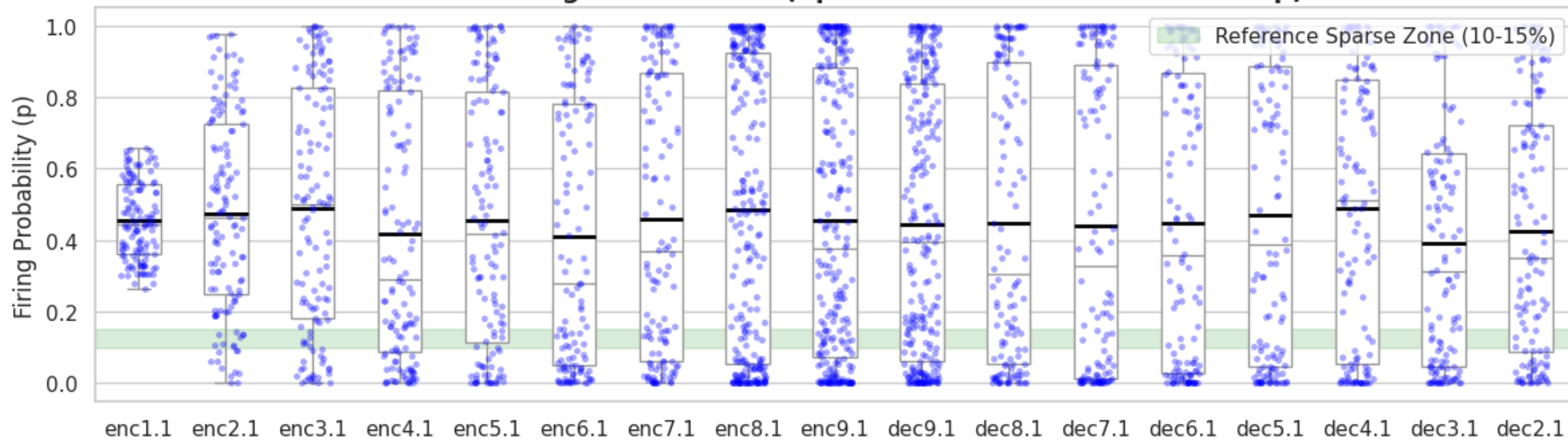
Channel Firing Probabilities (Spatial Distribution of Bernoulli p)



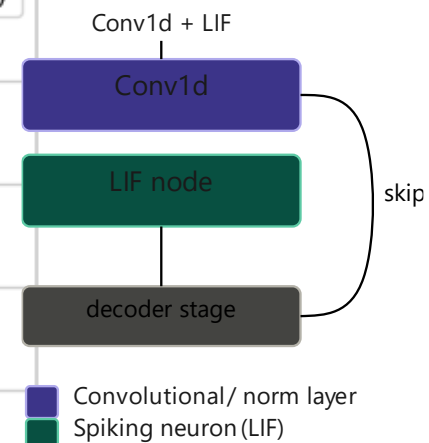
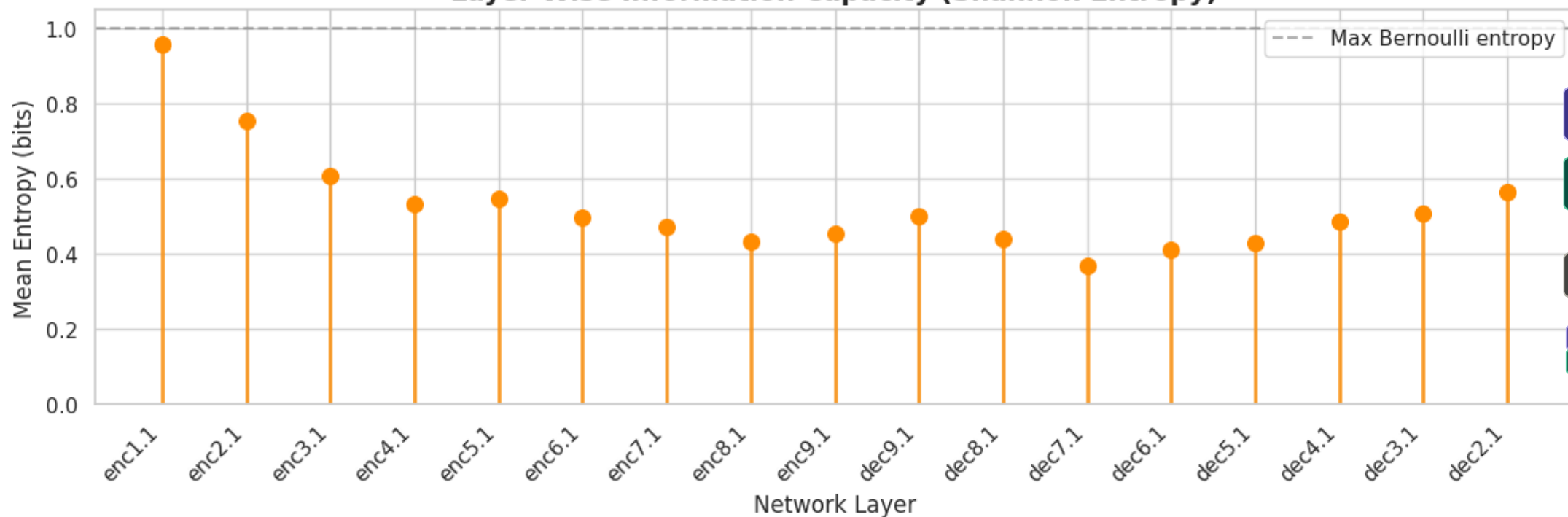
$$V_{th} = 0.1$$
$$T = 4$$

Layer-Wise Information Capacity (Shannon Entropy)

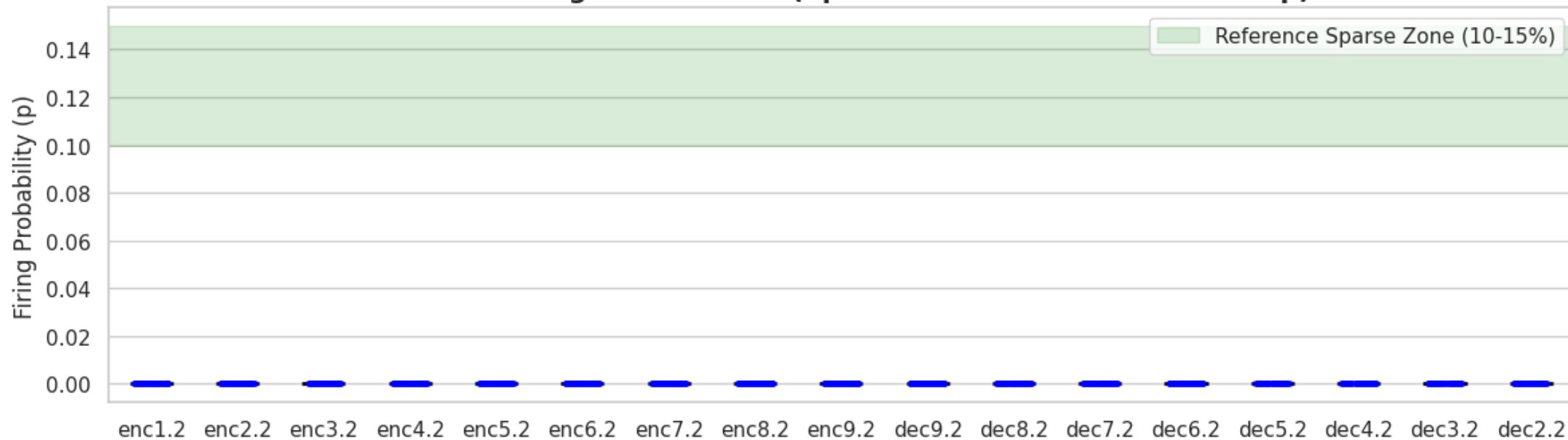


Channel Firing Probabilities (Spatial Distribution of Bernoulli p)

$$V_{th} = 0.01$$
$$T = 4$$

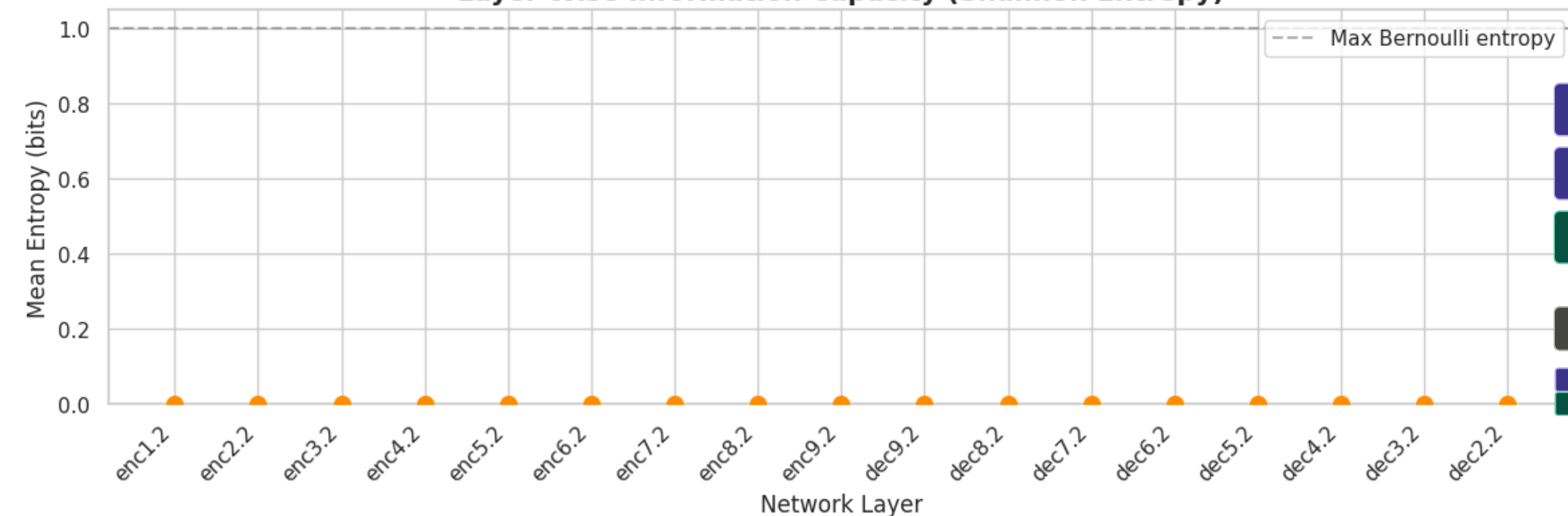
Layer-Wise Information Capacity (Shannon Entropy)

Channel Firing Probabilities (Spatial Distribution of Bernoulli p)



$V_{th} = 0.1$
 $\alpha = 1.0$
 $T = 4$

Layer-Wise Information Capacity (Shannon Entropy)



Conv1d + tdBn + LIF

Conv1d

tdBatchNorm

LIF node

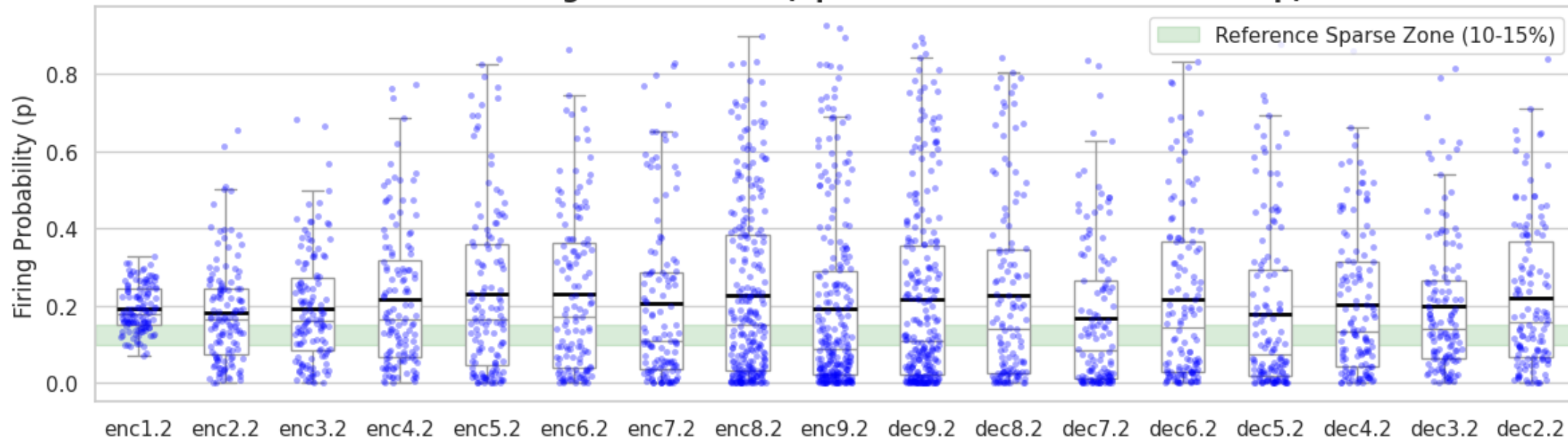
decoder stage

Convolutional / norm layer

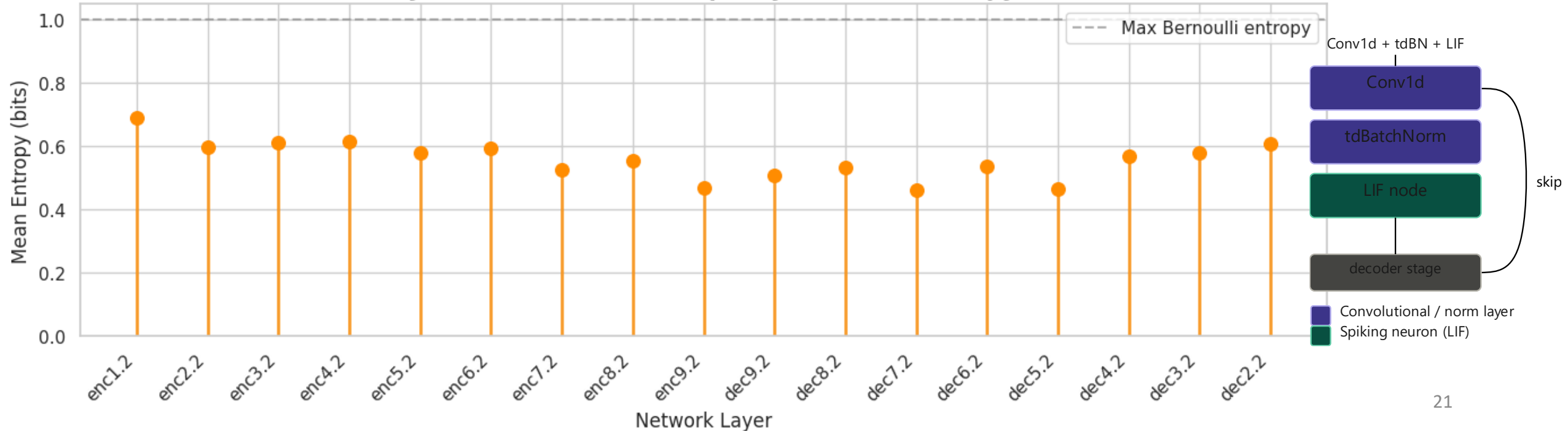
Spiking neuron (LIF)

skip

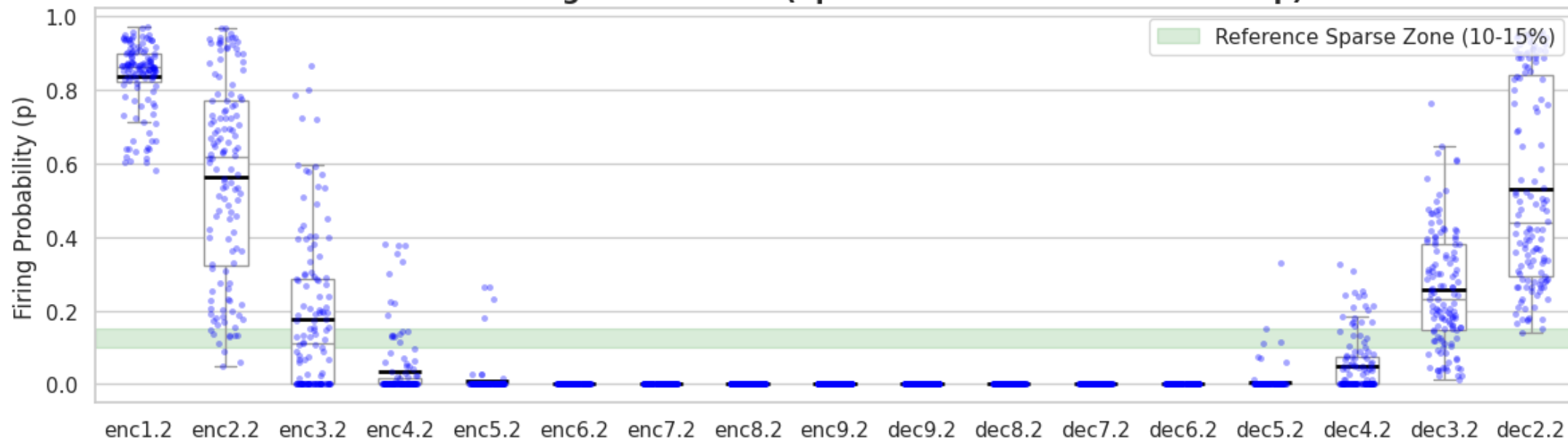
Channel Firing Probabilities (Spatial Distribution of Bernoulli p)



Layer-Wise Information Capacity (Shannon Entropy)

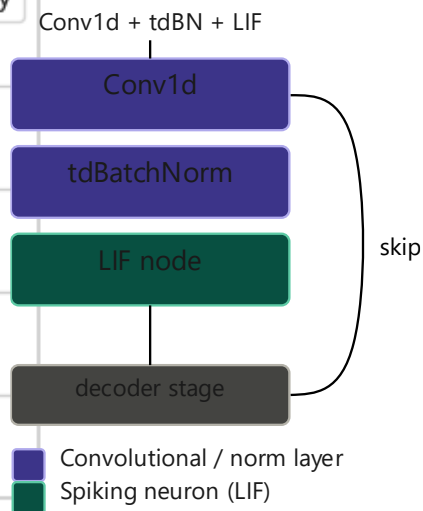
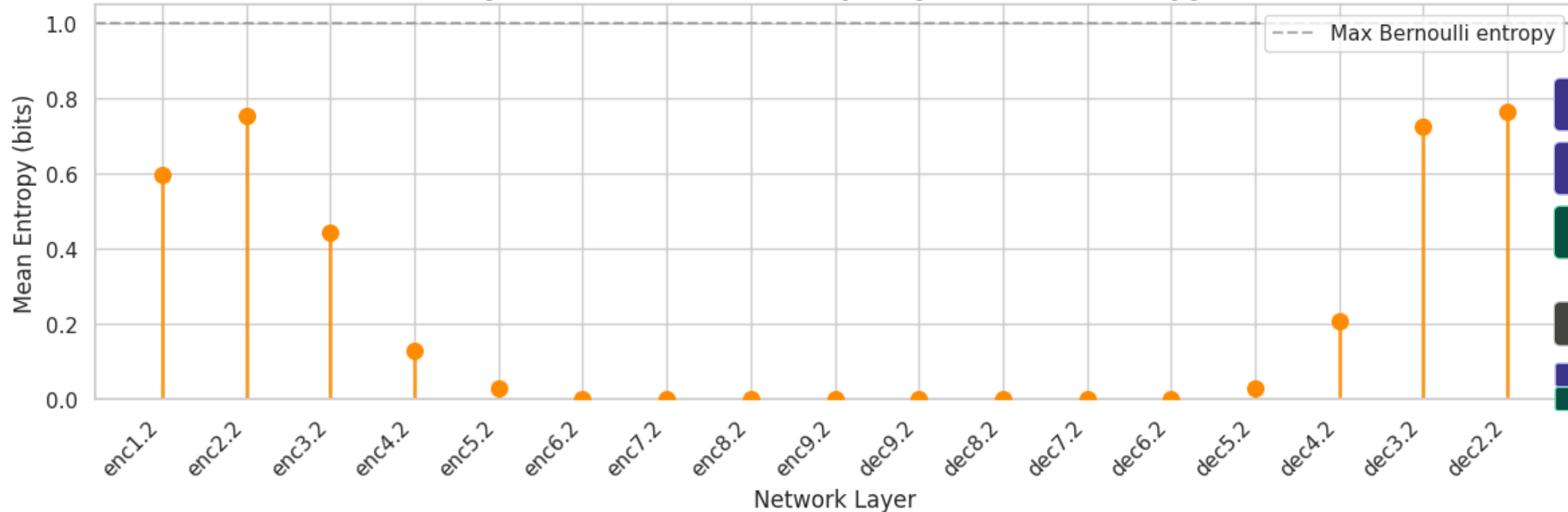


Channel Firing Probabilities (Spatial Distribution of Bernoulli p)

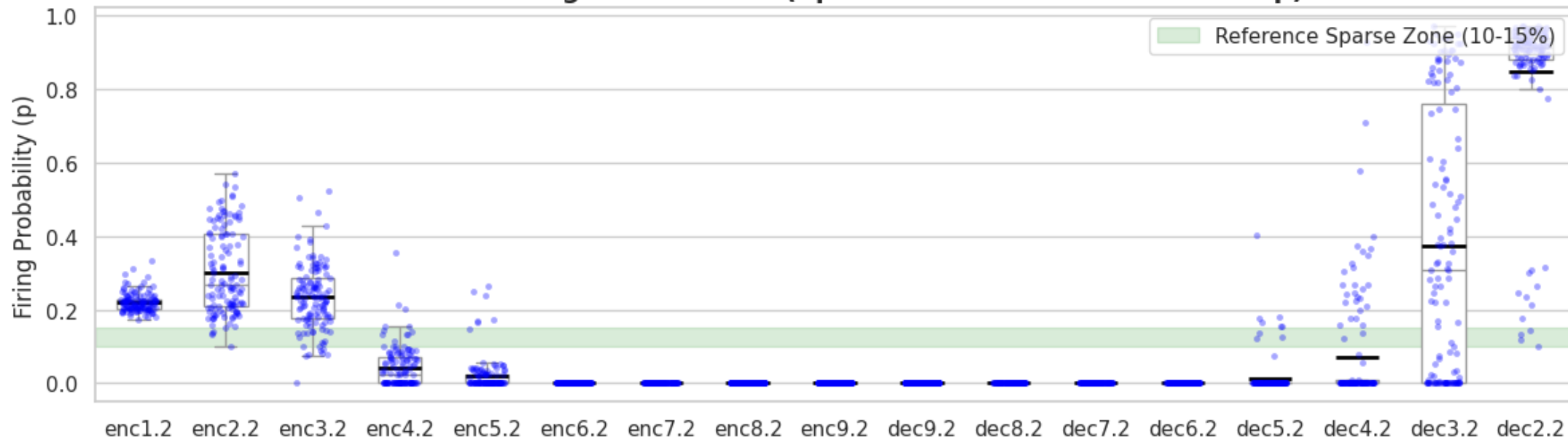


$V_{th} = 0.1$
 $\alpha = 10.0$
 $T = 4$
 treniruotas

Layer-Wise Information Capacity (Shannon Entropy)

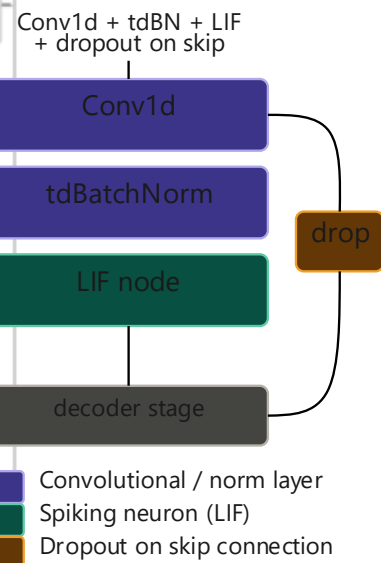
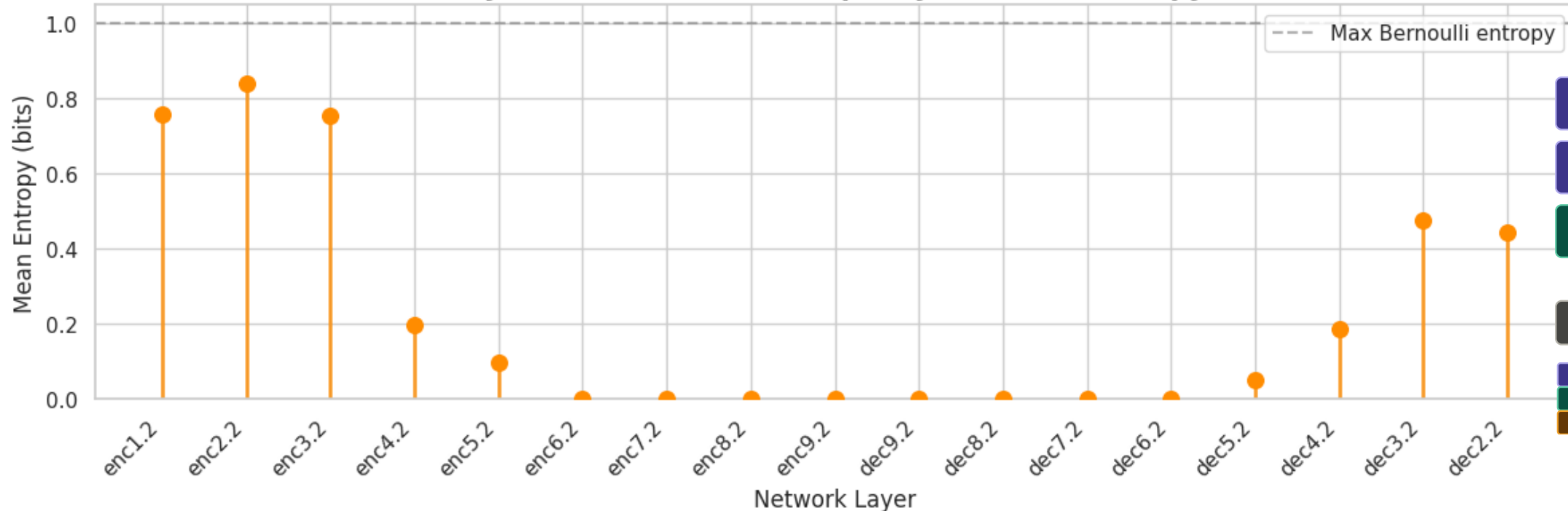


Channel Firing Probabilities (Spatial Distribution of Bernoulli p)



$V_{th} = 0.1$
 $\alpha = 10.0$
 $T = 16$
 $p_{drop} = 0.35$
 treniruotas

Layer-Wise Information Capacity (Shannon Entropy)



Sekančio pusmečio darbo planas, 2025/2026

– II pusmetis

- Tęsti impulsinių neuroninių tinklų taikymo problematikos tyrimą, pradėti šių tinklų empirinį tyrimą.
- Pradėti pasiruošimo darbus publikacijai su cit. rodikliu.

Klausimai