Spiking Neural Network チュートリアル for Physical AI

京都大学大学院情報学研究科　D2 羽原丈博

自己紹介

羽原 丈博（はばら たけひろ）

所属

京都大学大学院情報学研究科 通信情報システム
集積システム工学講座 D2

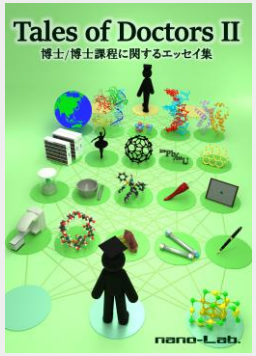
趣味

ものづくり（料理や同人誌など）
ゲーム全般（アナログもディジタルも）

C107 2025/12/31

研究テーマ

- ✓スパイキングニューラルネットワークの高速・低消費エネルギ推論
- ✓人流解析、スタイル変換などCV系



東テ33a

東テ25a

目次

1. スパイキングニューラルネットワークの現状
   A) ANNとSNN
   B) 低消費電力/エネルギ
   C) イベント駆動
   D) その他事例
2. 課題と解決への取り組み
3. まとめ

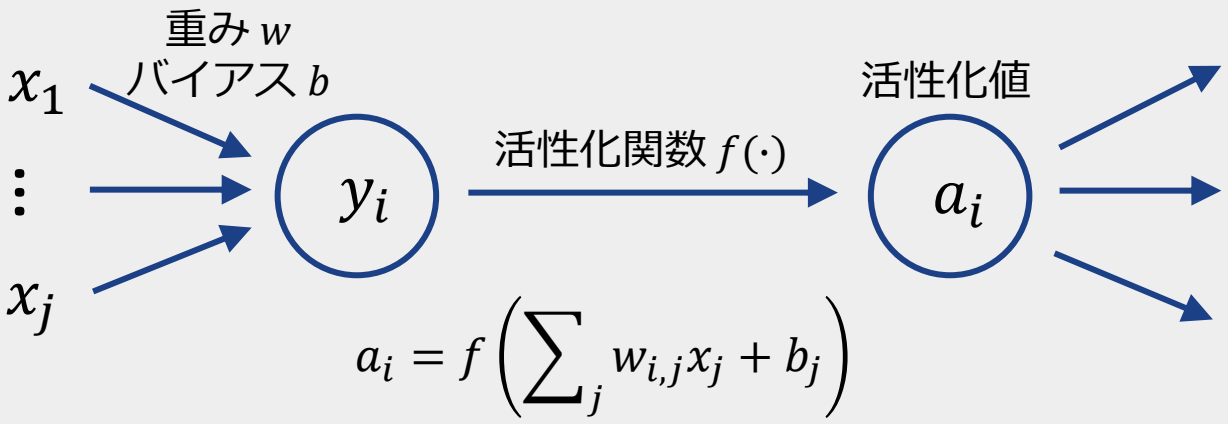
A) Artificial Neural Network (ANN)

深層学習で一般的に用いられるネットワークモデル

✓ 学習が容易で、様々なタスクに応用可能

✓ 実数×実数の計算を繰り返しており、高消費電力

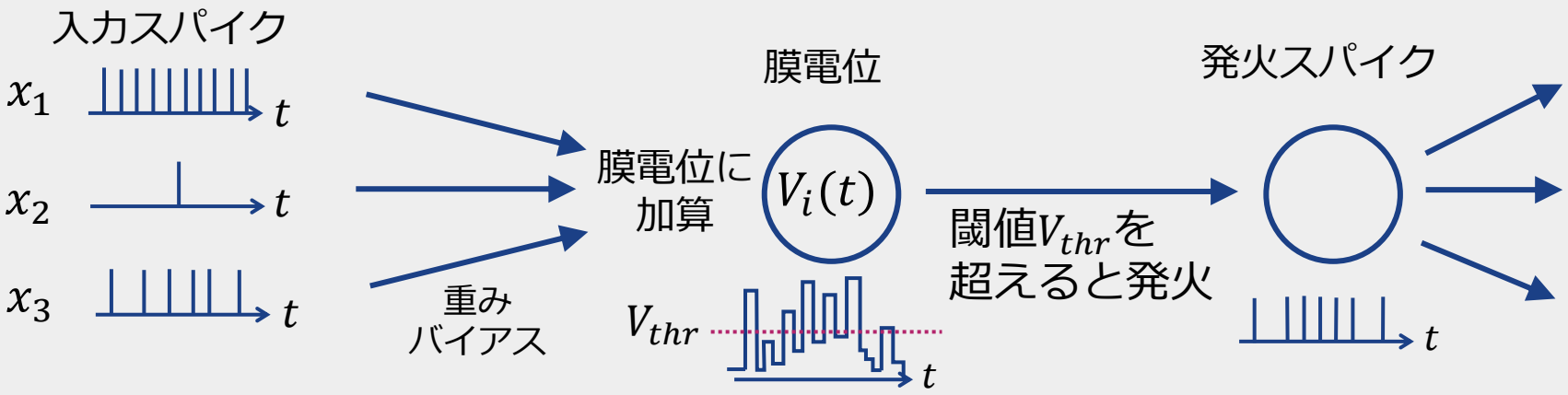
ChatGPTは年間アメリカの3万5000世帯分の電力を消費している噂



A) Spiking Neural Network (SNN)

生体の神経活動を模したネットワークモデル

✓スパイク×重み（実数）で計算

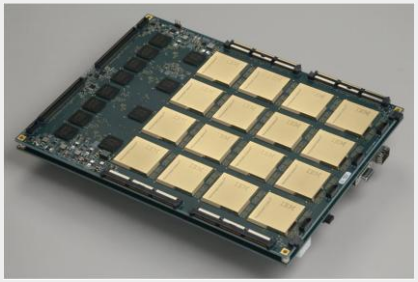


✓ スパース且つイベント駆動で低消費電力 / 低消費エネルギ

SNN専用のニューロモーフィックチップが開発

IBM TrueNorth, Intel Loihi, SynSense DYNAP-SE2

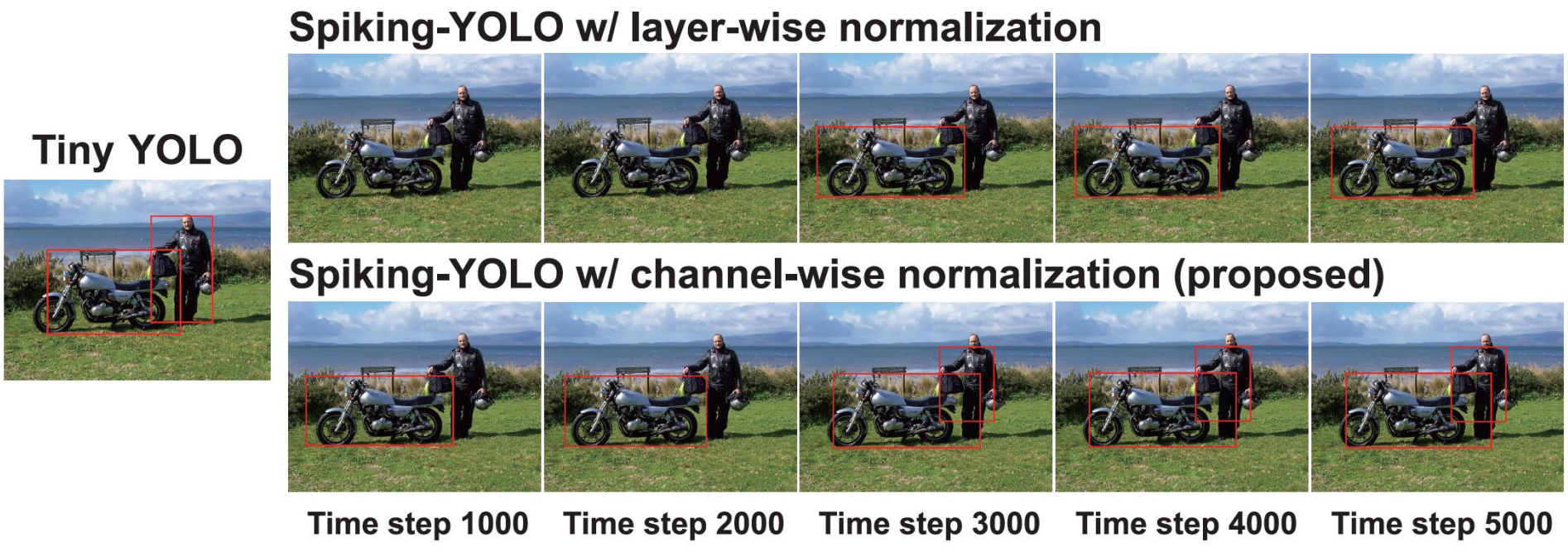
65mWで動作可能



B) 事例：物体検出

SNNで実装したYOLOによる物体検出

四角のバウンディングボックスとして画像中の物体を見つけ、クラス分類を行う



Kim, Seijoon & Park, Seongsik & Na, Byunggook & Yoon, Sungroh. (2020). Spiking-YOLO: Spiking Neural Network for Energy-Efficient Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence. 34. 11270-11277. 10.1609/aaai.v34i07.6787.

B) 事例：物体検出

ANNに比べて消費エネルギを99.6%削減可能

Tiny YOLO					
	Power (W)	GFLOPS	FLOPs		Energy (J)
	250	14,000	6.97E+09		0.12
Spiking-YOLO					
Norm. methods	GFLOPS / W	FLOPs	Power (W)	Time steps	Energy (J)
Layer	400	5.28E+07	1.320E-04	8,000	1.06E-03
Channel	400	4.90E+07	1.225E-04	**3,500**	**4.29E-04**

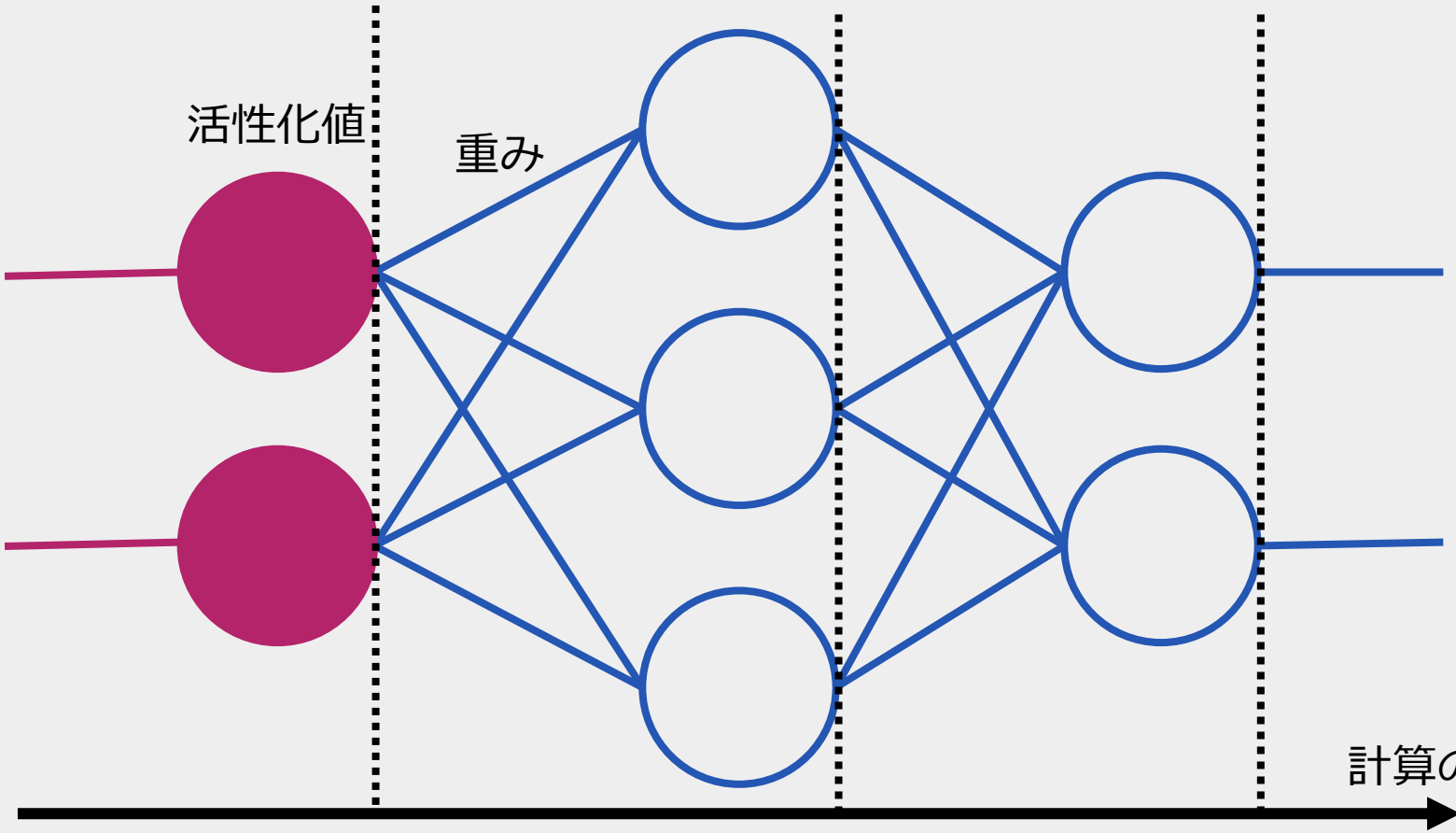
ANN: Titan V100で実行したときのパフォーマンス
SNN: TrueNorthで実行したときのパフォーマンス

B) どうして低消費電力/エネルギなのか

ANN

層ごとに密な実数積和演算を実行する必要あり

→ 高い並列性かつ高い粒度で計算するGPU

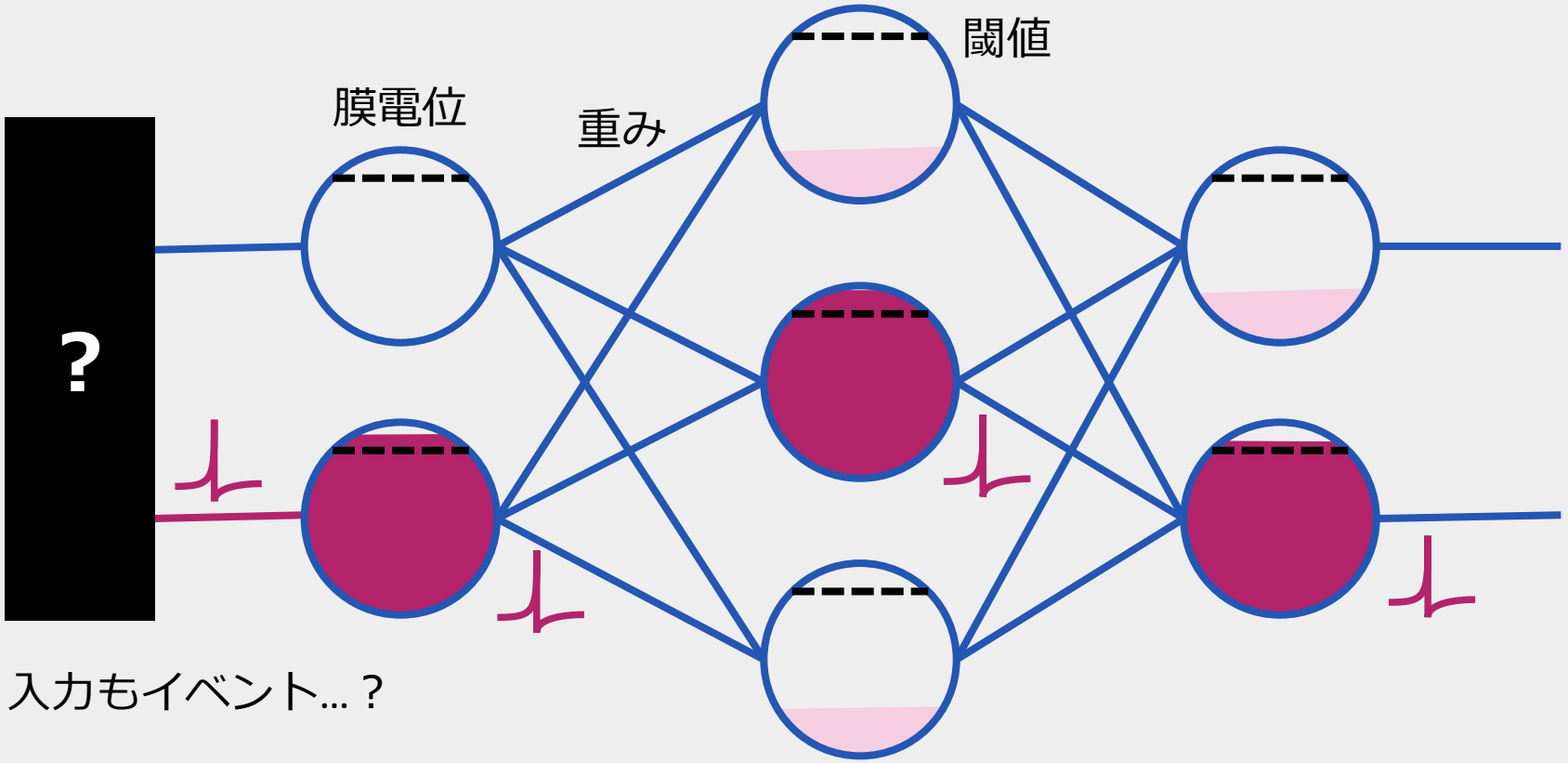


B) どうして低消費電力/エネルギなのか

SNN

スパイクごとに疎な実数加算演算を実行

→ イベント駆動可能なニューロモルフィックチップ



入力もイベント…？

C) イベントカメラ

ピクセルの輝度が大きく変化した“イベント”のみを情報として出力するカメラ

C) イベントカメラ

ピクセルの輝度が大きく変化した“イベント”のみを情報として出力するカメラ

- ☺ 高フレームレート
  高速なロボットやスポーツへの応用が期待
- ☺ スパースかつ非同期で低消費電力
- ☺ ハイダイナミックレンジ
  自動運転や宇宙・深海など光環境が大きく変化する場所での運用

- ☹ 静止しているものは苦手
  形状、テクスチャ等の把握は難しい
- ☹ ノイズが多く、データ形式や前処理が複雑
- ☹ 市場やツールチェーンが未成熟

C) 事例：物体追跡システム

SNNとイベントカメラを用いた卓球ロボット

Related Work

Spiking Neural Network (SNN)

UNIVERSITÄT TÜBINGEN

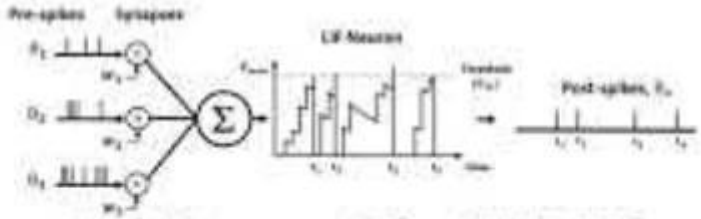


Figure: LIF neuron model [Lee, FNINS, 2020]

Background:

- SNNs model the spiking behavior of biological neurons more closely

Advantages:

- No multiplications needed (only additions) -> efficient implementations on hardware

Detection of Fast-Moving Objects with Neuromorphic Hardware — Andreas Ziegler et al. 3/5

Ziegler, A., Vetter, K., Gossard, T., Tebbe, J., Otte, S., & Zell, A. (2024). Detection of Fast-Moving Objects with Neuromorphic Hardware. arXiv preprint arXiv:2403.10677.

C) 事例：物体追跡システム

性能

予測座標の正確さは十分

DynapCNNはUSB接続、Loihi2はクラウドサーバなので通信オーバーヘッドがあり、One forward passは大きい

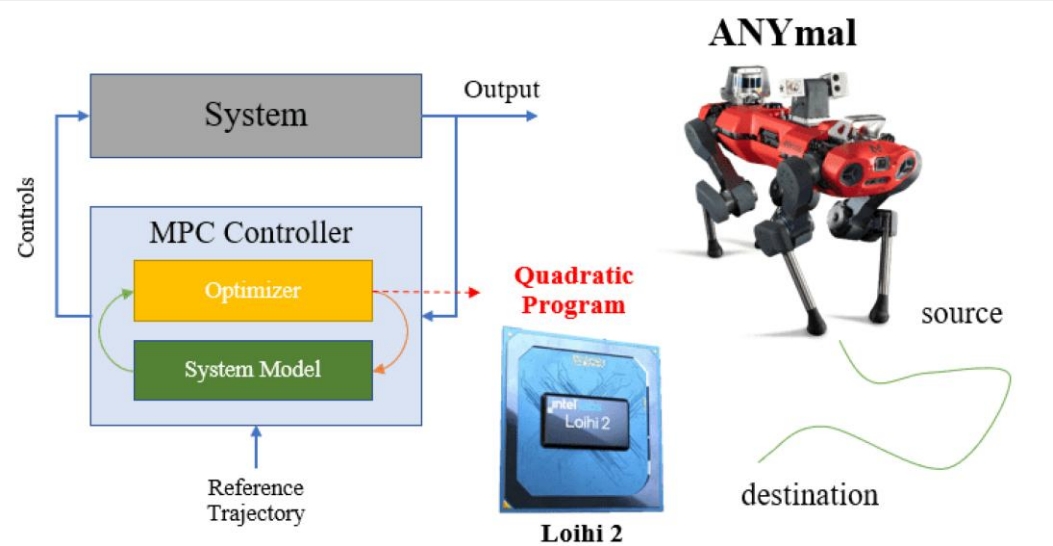
ハードウェア	Error [pixel]	One forward pass [ms]	Inference time [ms]
BrainChip Akida	1.44 ± 1.41	2.20 ± 0.35	0.89 ± 0.28
DynapCNN	1.59 ± 1.83	46.04 ± 21.38	0.82 ± 0.24
Loihi2	1.57 ± 1.55	1458.57 ± 230.50	0.62 ± 0.01

Error：追跡したボール座標のずれ
One forward pass：データ入力から座標出力までの時間
Inference time：ハードウェアで推論した時間

D) その他事例

ロボット

IntelのSNNアクセラレータLoihiを活用
四足歩行ロボットの制御や工業用ロボットハンドの制御



Mangalore, Ashish Rao, et al. "Neuromorphic quadratic programming for efficient and scalable model predictive control: Towards advancing speed and energy efficiency in robotic control." IEEE Robotics & Automation Magazine (2024).

Amaya, Camilo, et al. "Neuromorphic force-control in an industrial task: validating energy and latency benefits." arXiv preprint arXiv:2403.08928 (2024).

D) その他事例

画像生成

Fashion MNIST

MNIST

CIFAR-10

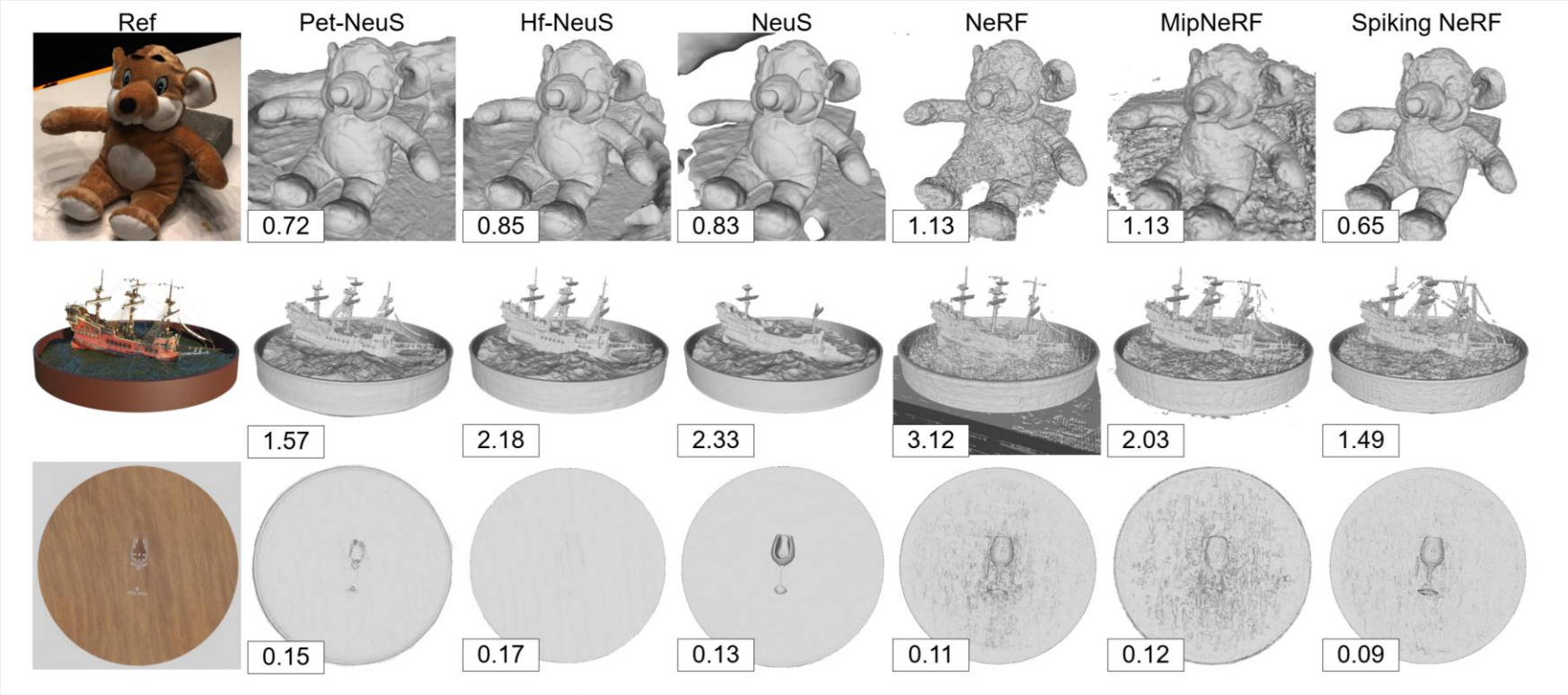
CelebA

LSUN-bed

D) その他事例

NeRF

従来のANNでは幾何学表現が連続であるため、空気と物体表面の界面が分離されない
そこで、SNNの非連続性を活用



Liao, Zhanfeng, et al. "Spiking NeRF: Representing the Real-World Geometry by a Discontinuous Representation." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. No. 12. 2024.

SNNの現状

- ✓ANNよりも低消費電力 / エネルギなネットワークモデルとして様々な応用先が模索されている
- ✓ANNで達成できている認識タスクや生成タスクはあらたかSNNは推論可能
- ✓イベント情報と相性が良く、センサからアクセラレータまでSNNに特化したシステムを構築する取り組みも進められている

目次

1. スパイキングニューラルネットワークの現状
2. 課題と解決への取り組み
   A) 学習則に関する課題
   B) ハードウェアに関する課題
3. まとめ

SNNの課題

学習則が未成熟

スパイクは微分不可のため、誤差逆伝搬法を直接利用不可

✓**代理勾配法**

✓ANN-SNN変換

ハードウェア周りが未成熟

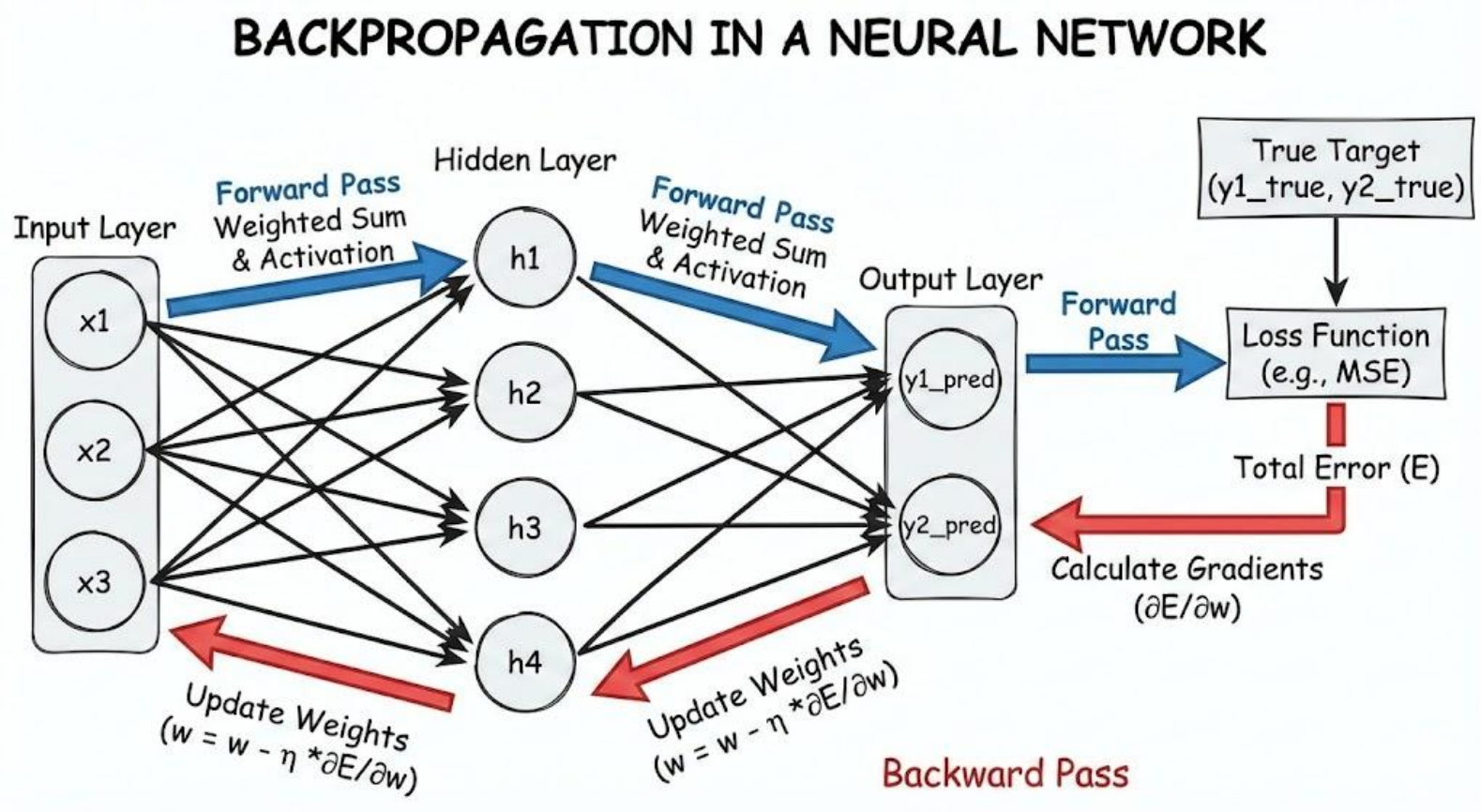
ツールが整備されておらず、学習コストが高い

ハードウェアに改善の余地あり

誤差逆伝搬法

関数で定義された「誤り」を小さくするように、重みを更新するANNの学習方法

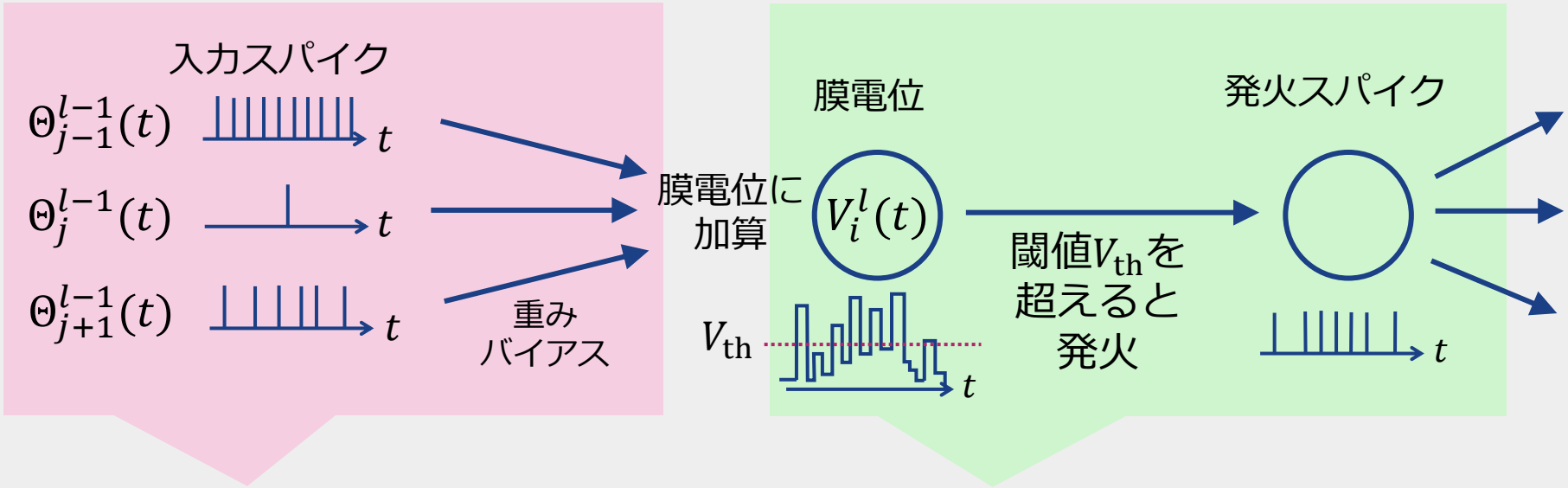
微分の連鎖率を活用



Made by Gemini (Prompt: Please create a figure to explain "Backpropagation")

SNNの微分不可能性

スパイク発火に用いるHeaviside関数が微分不可能



$$m_i^l(t) = V_i^l(t) + \sum_j w_{i,j}^l \Theta_{\mathrm{j}}^{l-1}(t) + b_i^l$$

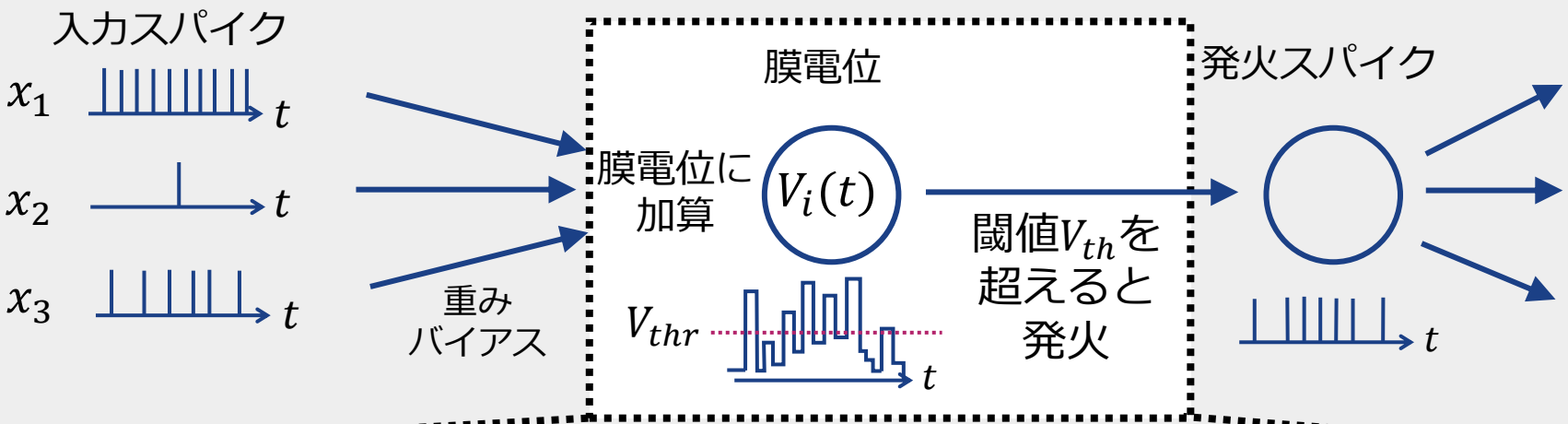
$$V_i^l(t+1) = m_i^l(t) - V_{\mathrm{th}}\Theta_i^l(t)$$

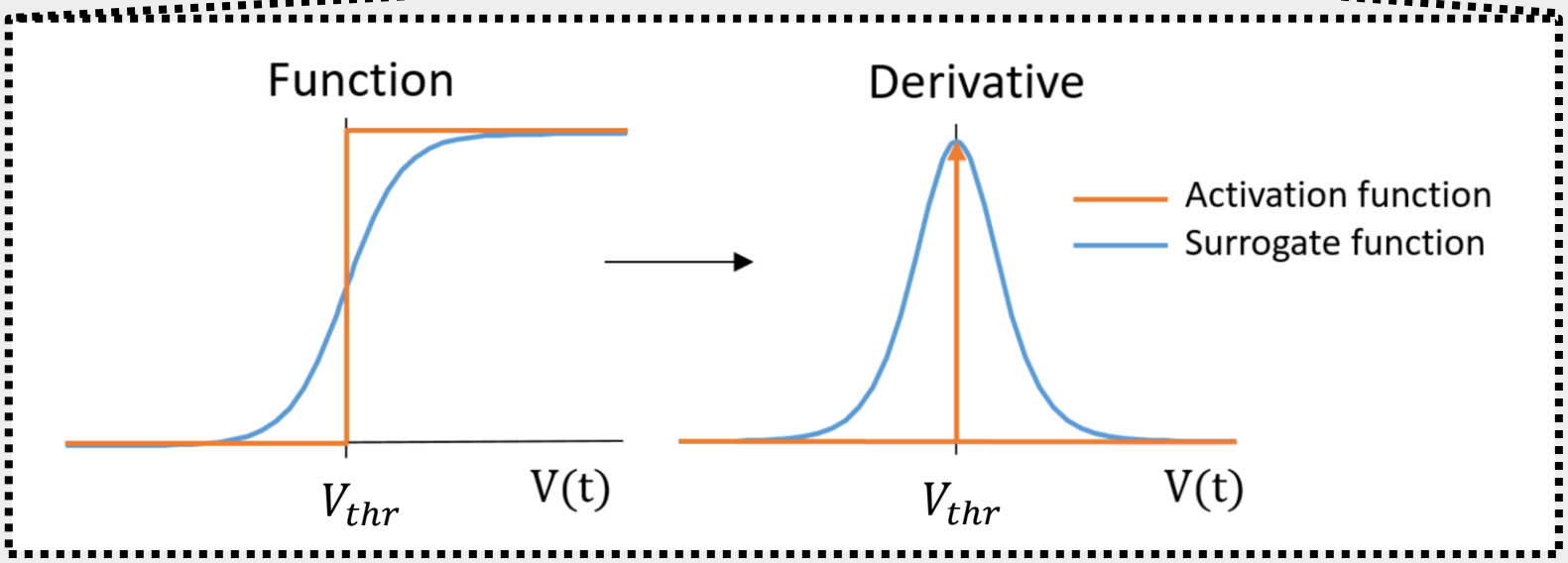
$$\Theta_i^l(t) = \begin{cases} 1 \left(m_i^l(t) \geq V_{\mathrm{th}}\right) \\ 0 \left(m_i^l(t) < V_{\mathrm{th}}\right) \end{cases}$$

↑こいつ

代理勾配法に関する課題

発火の式を**勾配計算時のみ**微分可能な式に近似し、
Backpropagation Trough Timeで学習する手法

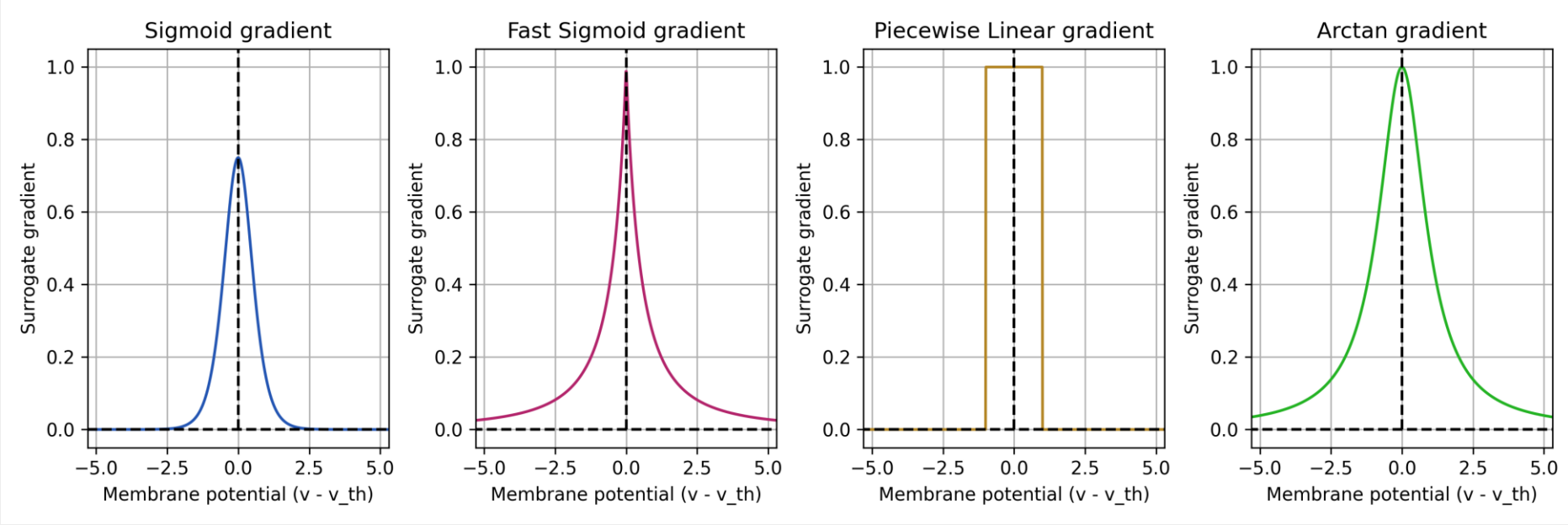




代理勾配法に関する課題

発火の式を**勾配計算時のみ**微分可能な式に近似し、Backpropagation Trough Timeで学習する手法

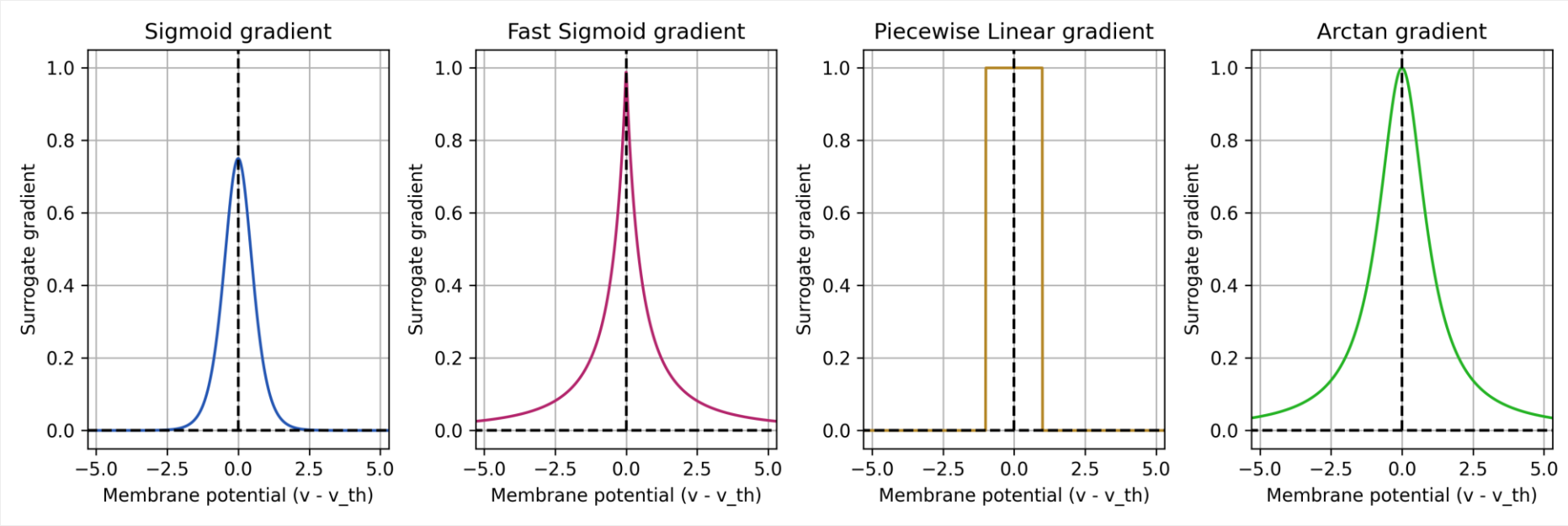
- Sigmoid : $\frac{1}{1+e^{-kx}}$
- Fast Sigmoid : $\frac{x}{1+|x|}$
- Arctangent : $\arctan(x)$
- Piecewise Linear : $\begin{cases} 0\ (x \leq -1) \\ \frac{x+1}{2}\ (-1 < x < 1) \\ 1\ (x \geq 1) \end{cases}$



代理勾配法に関する課題

発火の式を**勾配計算時のみ**微分可能な式に近似し、Backpropagation Trough Timeで学習する手法

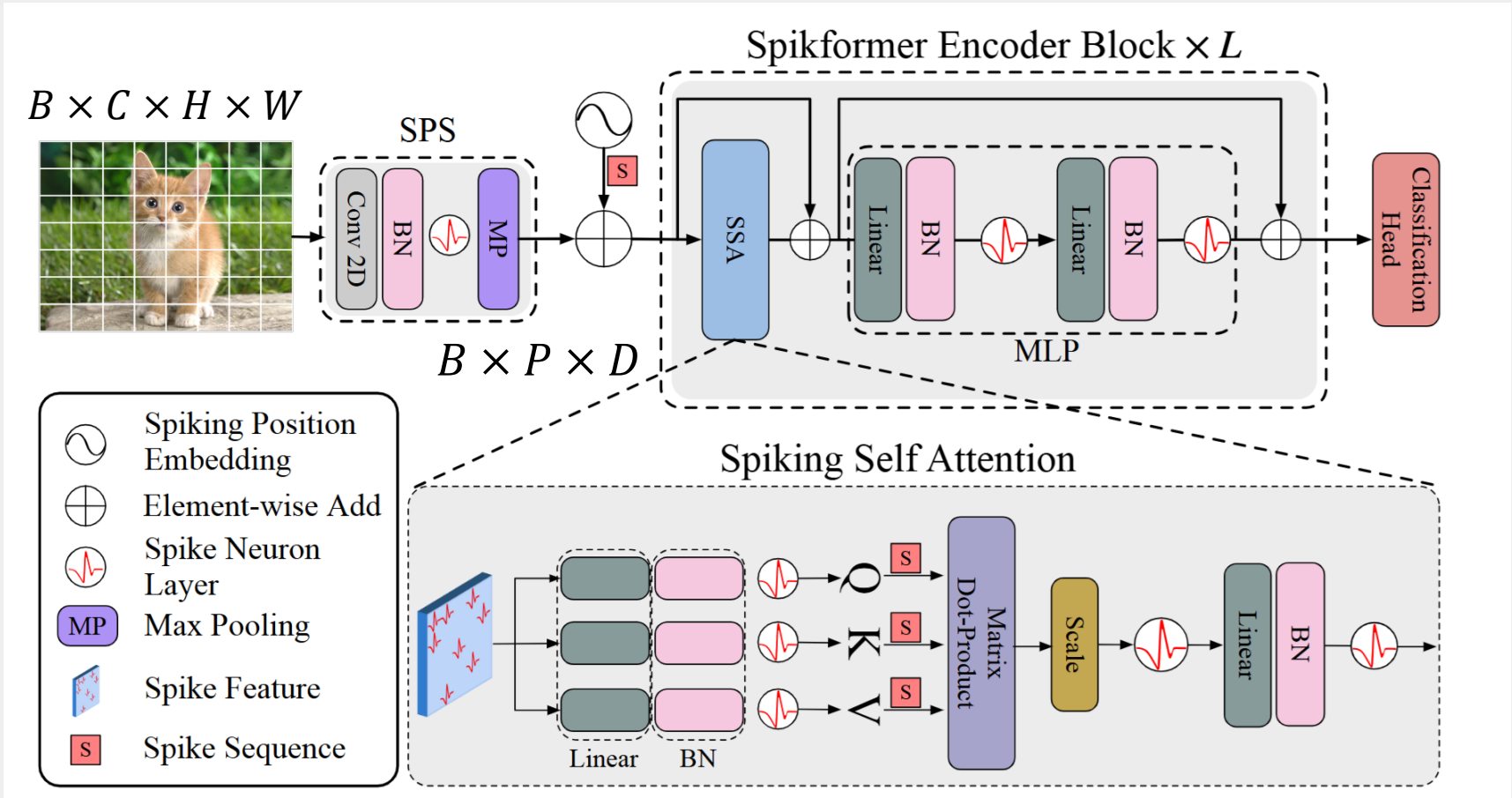
- ☺少ないタイムステップで推論可能
- ☹ 近似/勾配消失/爆発で高精度に達することが難しい
- ☹学習に十分必要なメモリを確保することが難しい



代理勾配法に関する課題：実施例

Spikfomer

SNNでTransformerを実装したモデルで、スパイクのみでself-attentionを実現できた初めての事例



Zhou, Z., Zhu, Y., He, C., Wang, Y., Yan, S., Tian, Y., & Yuan, L. (2022). Spikformer: When spiking neural network meets transformer. arXiv preprint arXiv:2209.15425.

代理勾配法に関する課題：実施例

Spikfomer

SNNでTransformerを実装したモデルで、スパイクのみでself-attentionを実現できた初めての事例

Methods	Architecture	Param (M)	OPs (G)	Power (mJ)	Time Step	Acc
Hybrid training(Rathi et al., 2020)	ResNet-34	21.79	-	-	250	61.48
TET(Deng et al., 2021)	Spiking-ResNet-34	21.79	-	-	6	64.79
	SEW-ResNet-34	21.79	-	-	4	68.00
Spiking ResNet(Hu et al., 2021a)	ResNet-34	21.79	65.28	59.295	350	71.61
	ResNet-50	25.56	78.29	70.934	350	72.75
STBP-tdBN(Zheng et al., 2021)	Spiking-ResNet-34	21.79	6.50	6.393	6	63.72
SEW ResNet(Fang et al., 2021a)	SEW-ResNet-34	21.79	3.88	4.035	4	67.04
	SEW-ResNet-50	25.56	4.83	4.890	4	67.78
	SEW-ResNet-101	44.55	9.30	8.913	4	68.76
	SEW-ResNet-152	60.19	13.72	12.891	4	69.26
Transformer	Transformer-8-512	29.68	8.33	38.340	1	**80.80**
Spikformer	Spikformer-8-384	16.81	6.82	7.734	4	70.24
	Spikformer-6-512	23.37	8.69	9.417	4	72.46
	Spikformer-8-512	29.68	11.09	11.577	4	**73.38**
	Spikformer-10-512	36.01	13.67	13.899	4	**73.68**
	Spikformer-8-768	66.34	22.09	21.477	4	**74.81**

Zhou, Z., Zhu, Y., He, C., Wang, Y., Yan, S., Tian, Y., & Yuan, L. (2022). Spikformer: When spiking neural network meets transformer. arXiv preprint arXiv:2209.15425.

代理勾配法の課題解決方針

- ☹ ANNと同等の精度に達することが難しい
  - ✓Optimizerなどのハイパーパラメータの調整する方針

    Deng, Lei, et al. "Rethinking the performance comparison between SNNS and ANNS." Neural networks 121 (2020): 294-307.

  - ✓**時間変化も学習する方針**

    Deng, S., Li, Y., Zhang, S., & Gu, S. (2022). Temporal efficient training of spiking neural network via gradient re-weighting. arXiv preprint arXiv:2202.11946.

- ☹ 学習時の消費メモリが大きい
  - ✓学習済みANNで初期化する方針

    Rathi, Nitin, and Kaushik Roy. "Diet-SNN: Direct input encoding with leakage and threshold optimization in deep spiking neural networks." arXiv preprint arXiv:2008.03658 (2020).

  - ✓スパイクの発火時刻で情報を表現する方針

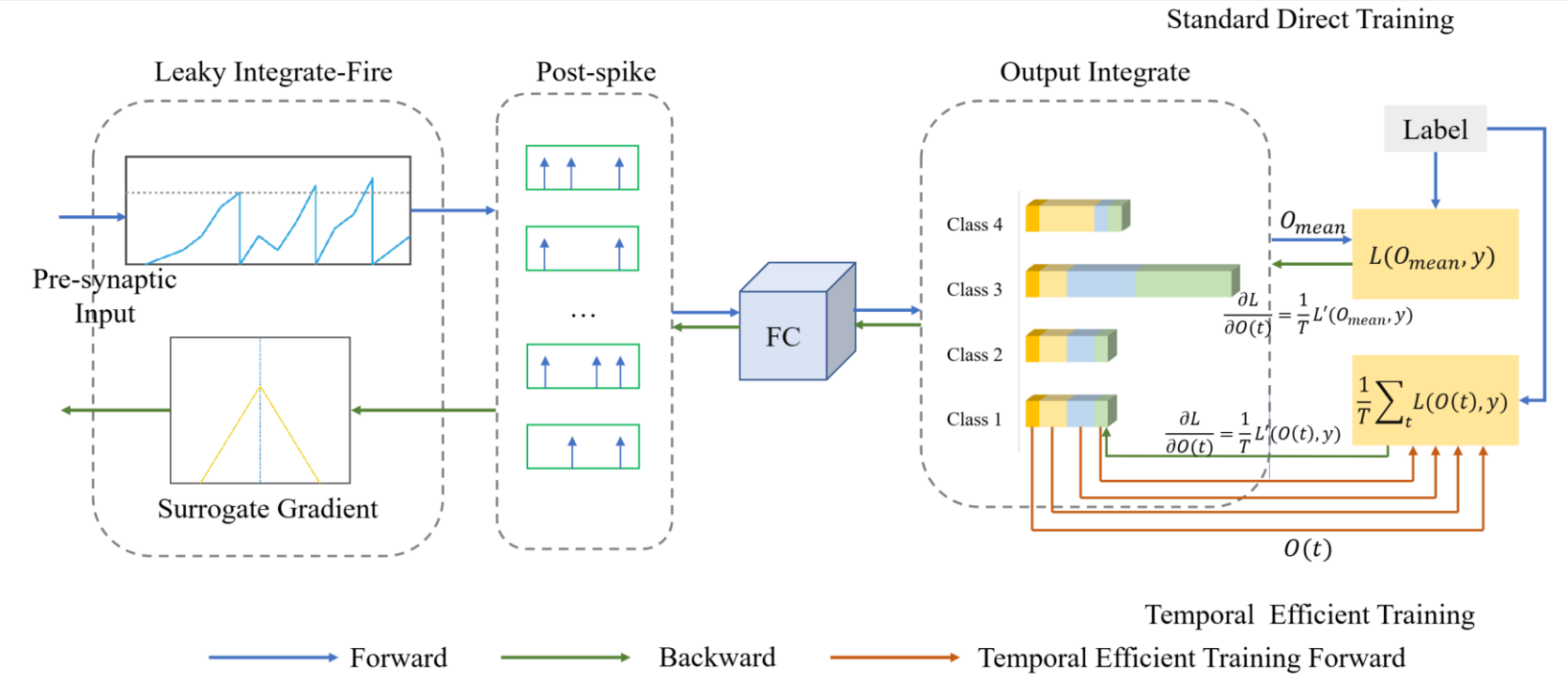
    Comşa, Iulia-Maria, et al. "Temporal coding in spiking neural networks with alpha synaptic function: learning with backpropagation." IEEE transactions on neural networks and learning systems 33.10 (2021): 5939-5952.

代理勾配法の課題解決方針

時間変化も学習する方針

例）Temporal efficient training method

推論時間ごとに損失を計算することで、より強い勾配降下を与え、局所最適に陥ることを防ぐ



Deng, S., Li, Y., Zhang, S., & Gu, S. (2022). Temporal efficient training of spiking neural network via gradient re-weighting. arXiv preprint arXiv:2202.11946.

代理勾配法の課題解決方針

小規模なデータセットではANNと同等の精度を達成

Dataset	Model	Methods	Architecture	Simulation Length	Accuracy
CIFAR10	Rathi et al. (2019)	Hybrid training	ResNet-20	250	92.22
	Rathi & Roy (2020)	Diet-SNN	ResNet-20	10	92.54
	Wu et al. (2018)	STBP	CIFARNet	12	89.83
	Wu et al. (2019)	STBP NeuNorm	CIFARNet	12	90.53
	Zhang & Li (2020)	TSSL-BP	CIFARNet	5	91.41
	Zheng et al. (2021)	STBP-tdBN	ResNet-19	6	93.16
				4	92.92
				2	92.34
	our model	TET	ResNet-19	6	**94.50**±0.07
				4	**94.44**±0.08
				2	**94.16**±0.03
	ANN*	ANN	ResNet-19	1	94.97
CIFAR100	Rathi et al. (2019)	Hybrid training	VGG-11	125	67.87
	Rathi & Roy (2020)	Diet-SNN	ResNet-20	5	64.07
	Zheng et al. (2021)*	STBP-tdBN	ResNet-19	6	71.12±0.57
				4	70.86±0.22
				2	69.41±0.08
	our model	TET	ResNet-19	6	**74.72**±0.28
				4	**74.47**±0.15
				2	**72.87**±0.10
	ANN*	ANN	ResNet-19	1	75.35

代理勾配法の課題解決方針

大規模なデータセットでは改善の余地あり

ImageNet	Rathi et al. (2019)	Hybrid training	ResNet-34	250	61.48
	Sengupta et al. (2018)	SPIKE-NORM	ResNet-34	2500	69.96
	Zheng et al. (2021)	STBP-tdBN	Spiking-ResNet-34	6	63.72
	Fang et al. (2021)	SEW ResNet	SEW-ResNet-34	4	67.04
	our model	TET	Spiking-ResNet-34	6	**64.79**
		TET	SEW-ResNet-34	4	**68.00**
	Pytorch公式実装	ANN	ResNet34	1	73.31

https://pytorch.org/vision/main/models/generated/torchvision.models.resnet34.html

イベント時系列データセットでは最も良い結果

DVS-CIFAR10	Zheng et al. (2021)	STBP-tdBN	ResNet-19	10	67.8
	Kugele et al. (2020)	Streaming Rollout	DenseNet	10	66.8
	Wu et al. (2021)	Conv3D	LIAF-Net	10	71.70
	Wu et al. (2021)	LIAF	LIAF-Net	10	70.40
	our model	TET	VGGSNN	10	**77.33**±0.21
		TET†	VGGSNN	10	**83.17**±0.15

SNNの課題

学習則が未成熟

スパイクは微分不可のため、誤差逆伝搬法を直接利用不可

✓代理勾配法

✓**ANN-SNN変換**

ハードウェア周りが未成熟

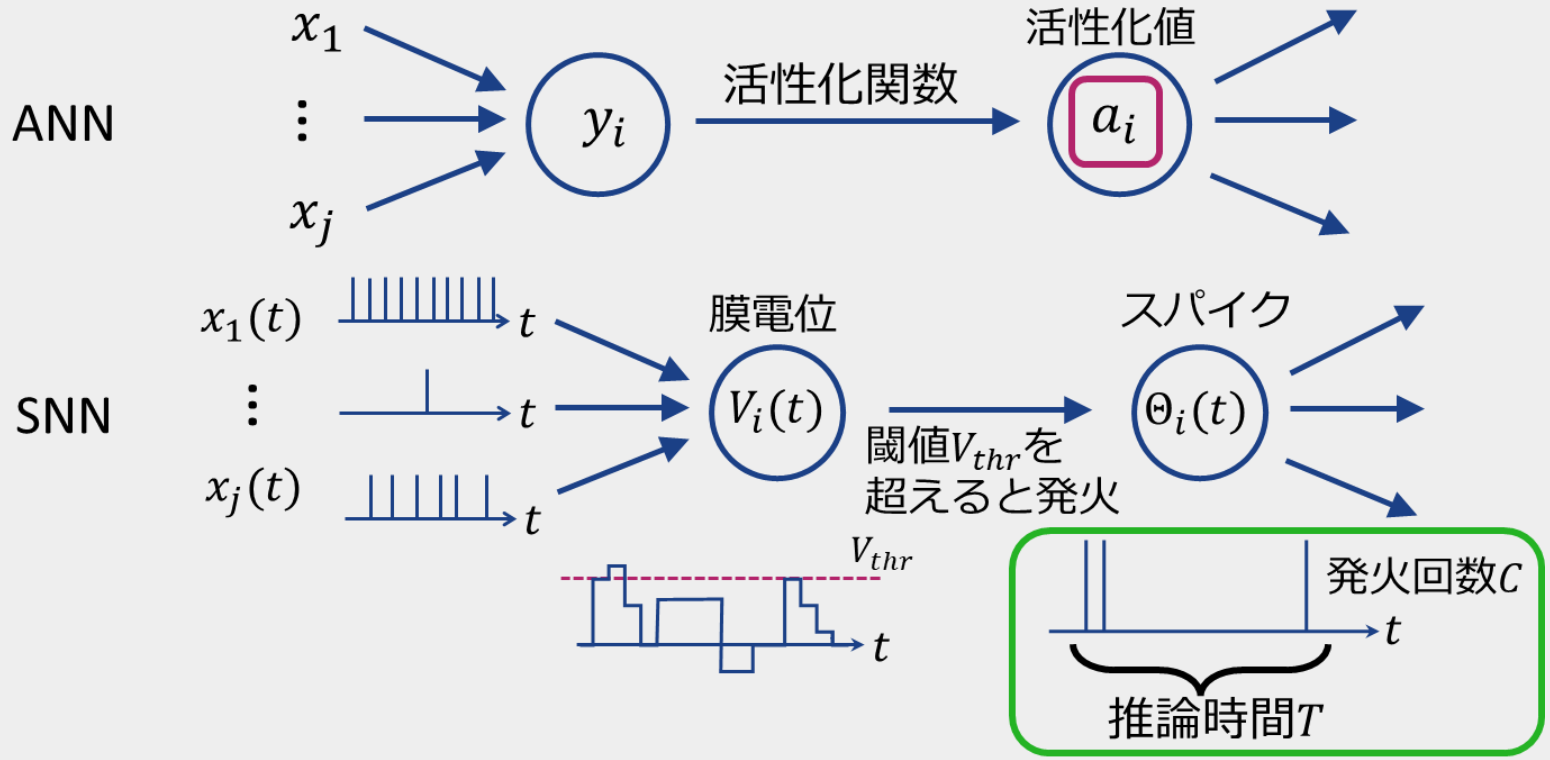
ツールが整備されておらず、学習コストが高い

ハードウェアに改善の余地あり

ANN-SNN変換に関する課題

活性化値と発火率を一致させるように学習済みANNをSNNにマッピングする学習手法

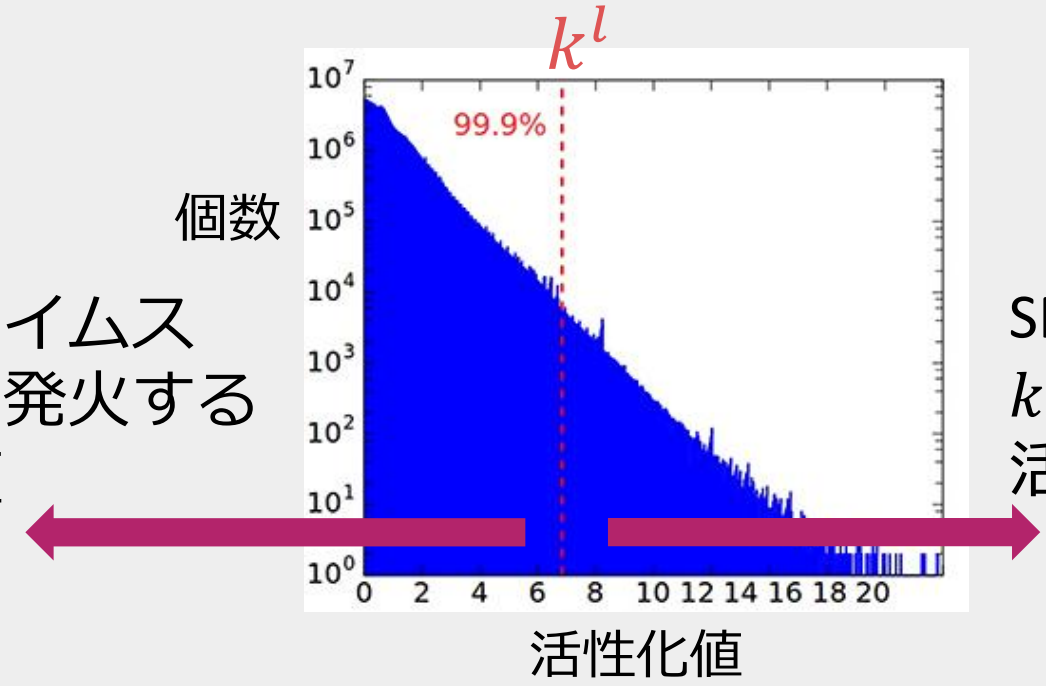
- ☺ 従来のANNモデルと同等の精度を達成可能
- ☹ 莫大な推論時間が必要
- ☹ 変換できないモデル構造がある



ANN-SNN変換に関する課題

重み・バイアスを**ANNの活性化値のうち99.9パーセンタイル値**で正規化

$$\bar{w}_{i,j}^{l} = w_{i,j}^{l} \frac{k^{l-1}}{k^{l}} \qquad \bar{b}_{i}^{l} = b_{i}^{l} \frac{1}{k^{l}}$$



複数のタイムステップで発火する活性化値

SNN側で発火しても、k^lの値に丸められる活性化値

Rueckauer, B., Lungu, I. A., Hu, Y., Pfeiffer, M., & Liu, S. C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. Frontiers in neuroscience, 11, 682.

SNNの学習：ANN-SNN変換

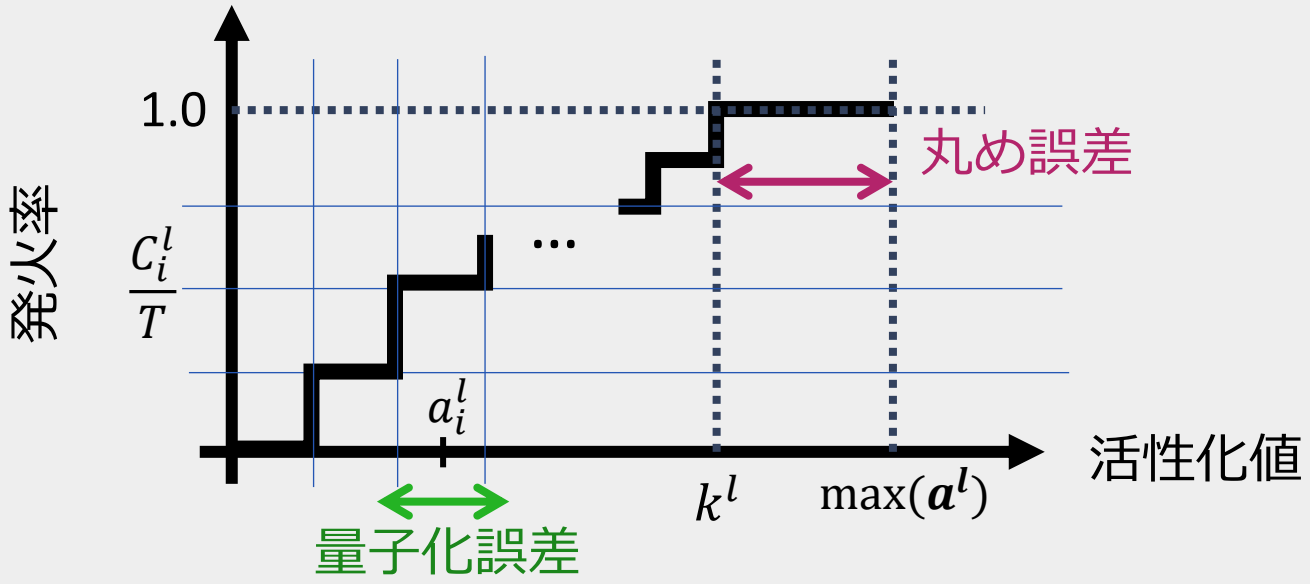
ANN-SNN変換の誤差

- ✓ 丸め誤差

  SNNの発火率で表現可能な値の範囲と元のANNの活性化値で表現可能な値の範囲の差

- ✓ 量子化誤差

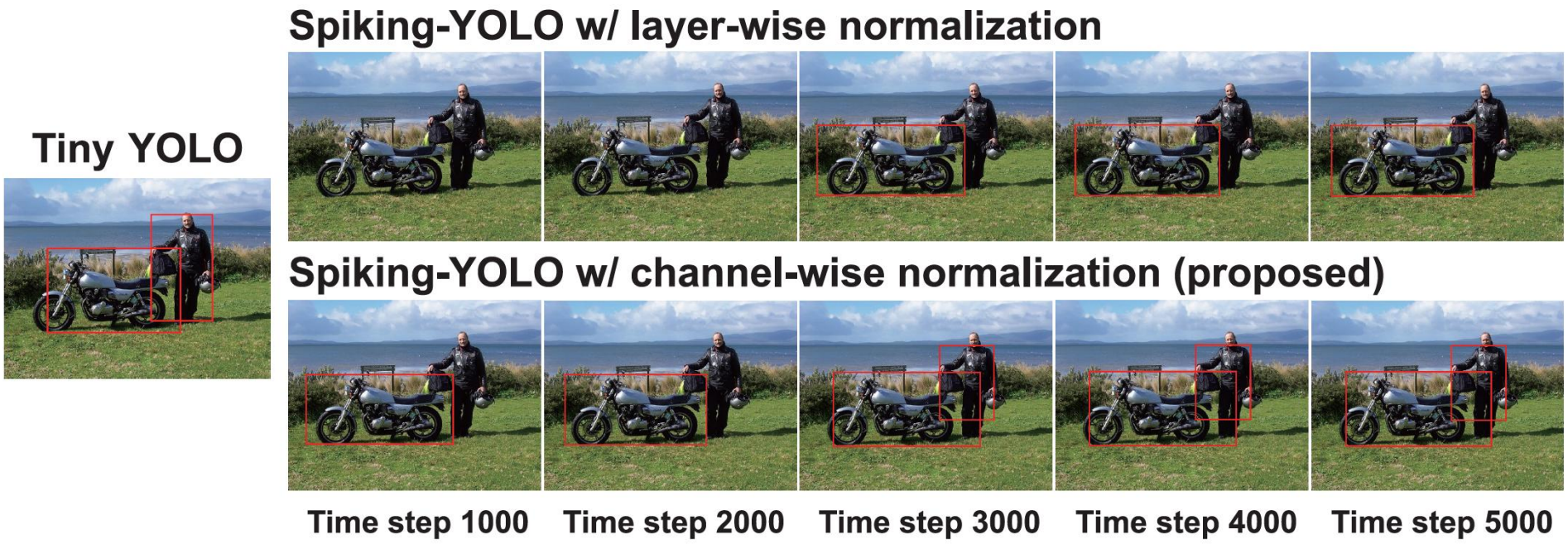
  発火率は離散値であるが活性化値は実数
  推論時間Tを大きくすることで削減可能



ANN-SNN変換に関する課題：実施例

SNNで実装したYOLOによる物体検出

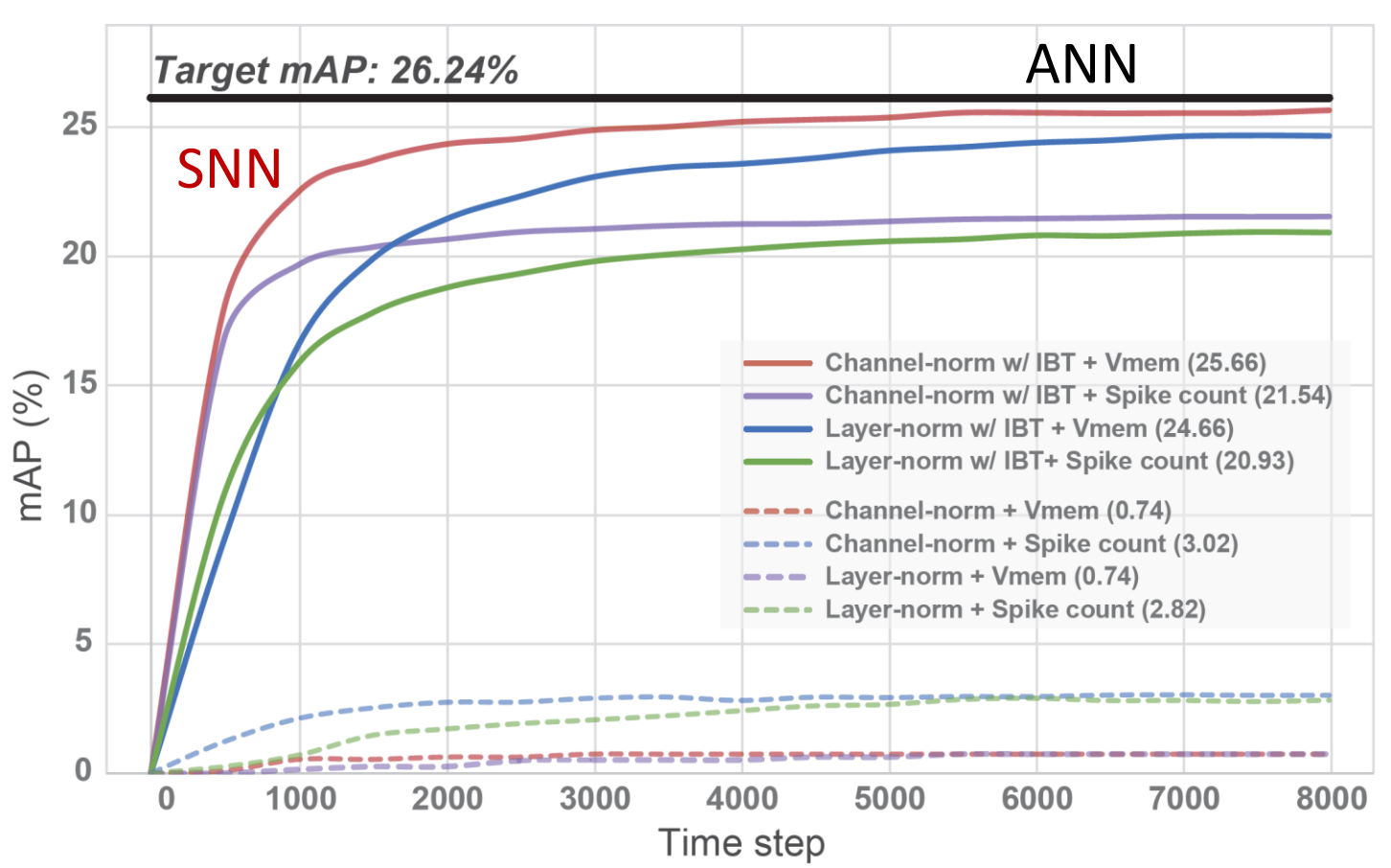
四角のバウンディングボックスとして画像中の物体を見つけ、クラス分類を行う



Kim, Seijoon & Park, Seongsik & Na, Byunggook & Yoon, Sungroh. (2020). Spiking-YOLO: Spiking Neural Network for Energy-Efficient Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence. 34. 11270-11277. 10.1609/aaai.v34i07.6787.

ANN-SNN変換に関する課題：実施例

4000タイムステップぐらい推論することでANNと同等の精度に達することが可能



MS COCOデータセットにおける結果

ANN-SNN変換に関する課題：実施例

ANNの推論時間：0.12 / 250 = 4.8×10^{-4}秒

SNNの推論時間：4.29×10^{-4} / 1.225×10^{-4} = 3.5秒

Tiny YOLO					
	Power (W)	GFLOPS	FLOPs		Energy (J)
	250	14,000	6.97E+09		0.12
Spiking-YOLO					
Norm. methods	GFLOPS / W	FLOPs	Power (W)	Time steps	Energy (J)
Layer	400	5.28E+07	1.320E-04	8,000	1.06E-03
Channel	400	4.90E+07	1.225E-04	**3,500**	**4.29E-04**

ANN: Titan V100で実行したときのパフォーマンス

SNN: TrueNorthで実行したときのパフォーマンス

ANN-SNN変換の課題解決方針

☹ 莫大な推論時間（レイテンシ）が必要

✓変換前の重みや活性化値を量子化する方針

Bu, T., Fang, W., Ding, J., Dai, P., Yu, Z., & Huang, T. (2023). Optimal ANN-SNN conversion for high-accuracy and ultra-low-latency spiking neural networks. arXiv preprint arXiv:2303.04347.

✓膜電位などのハイパーパラメータを調整する方針

Hwang, Sungmin, et al. "Low-latency spiking neural networks using pre-charged membrane potential and delayed evaluation." Frontiers in Neuroscience 15 (2021): 629000.

✓事前知識や出力以外の情報を利用する方針

Habara, Takehiro, Takashi Sato, and Hiromitsu Awano. "BayesianSpikeFusion: accelerating spiking neural network inference via Bayesian fusion of early prediction." Frontiers in Neuroscience 18 (2024): 1420119.

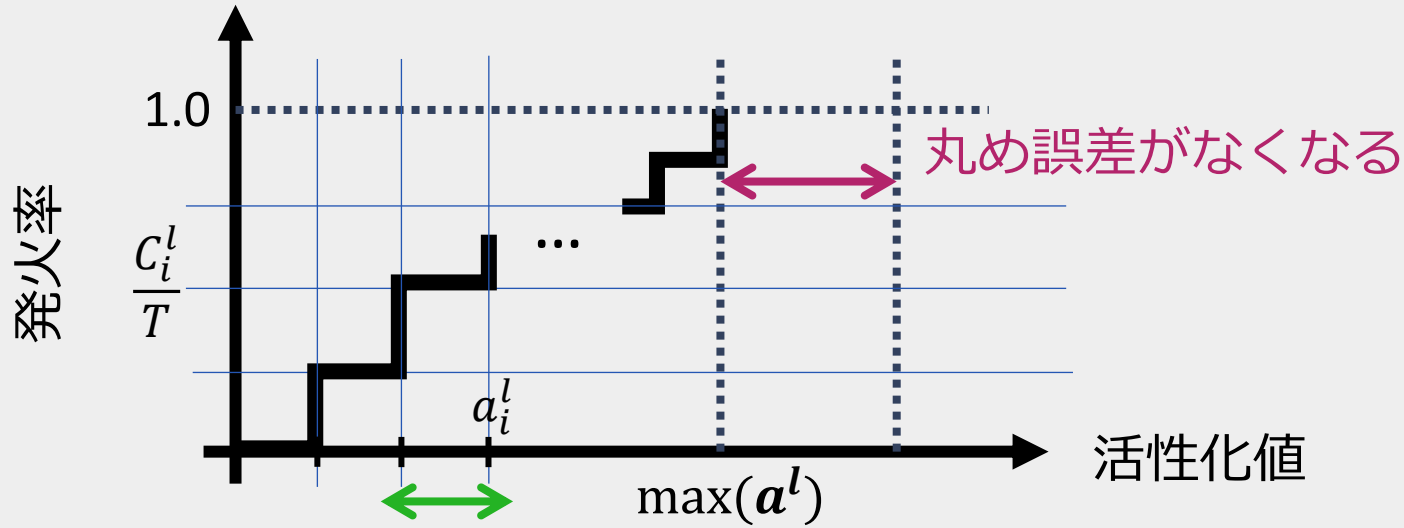
ANN-SNN変換の課題解決

ANNの学習時に活性化値を量子化する手法

例）Quantization clip-floor-shift activation function

ReLU関数の代わりとなる下のような式

$$a^l = \lambda^l \mathrm{clip}\left(\frac{1}{L}\left\lfloor \frac{z^l L}{\lambda^l} + \varphi \right\rfloor, 0, 1\right)$$



床関数により学習時に離散値であることを明示することで量子化誤差を削減

Bu, T., Fang, W., Ding, J., Dai, P., Yu, Z., & Huang, T. (2023). Optimal ANN-SNN conversion for high-accuracy and ultra-low-latency spiking neural networks. arXiv preprint arXiv:2303.04347.

ANN-SNN変換の課題解決

QCFSを用いることで推論時間を大幅に削減

Architecture	Method	ANN	T=16	T=32	T=64	T=128	T=256	T≥1024
ResNet-34	RMP	70.64%	-	-	-	-	-	65.47%
	TSC	70.64%	-	-	-	-	61.48%	65.10%
	RTS	75.66%	-	0.09%	0.12%	3.19%	47.11%	75.08%
	SNNC-AP	75.66%	-	64.54%	71.12%	73.45%	74.61%	75.45%
	Ours	74.32%	59.35%	69.37%	72.35%	73.15%	73.37%	73.39%
VGG-16	RMP	73.49%	-	-	-	-	48.32%	73.09%
	TSC	73.49%	-	-	-	-	69.71%	73.46%
	RTS	75.36%	-	0.114%	0.118%	0.122%	1.81%	73.88%
	SNNC-AP	75.36%	-	63.64%	70.69%	73.32%	74.23%	75.32%
	Ours	74.29%	50.97%	68.47%	72.85%	73.97%	74.22%	74.32%

ImageNetデータセットでの実施例：OursがQCFS

学習則の課題まとめ

	代理勾配法	ANN-SNN変換
メリット	✓ 低レイテンシ	✓ 従来のANNと同等の精度
デメリット	✓ ANNよりも精度劣化 ✓ 消費メモリが大きい	✓ 高レイテンシ ✓ 複雑なモデル構造の変換は難しい
解決の方針	✓ 時間変化も同時に学習 ✓ スパイクの発火時刻で情報を表現	✓ ANN学習時に活性化関数を工夫 ✓ ハイパーパラメータの調整 ✓ 事前知識などの活用

SNNの課題

学習則が未成熟

スパイクという微分不可能な性質上、誤差逆伝搬法以外の手法で学習する必要あり

✓代理勾配法

✓ANN-SNN変換

ハードウェア周りが未成熟

ツールが整備されておらず、学習コストが高い

ハードウェアに改善の余地あり

ハードウェアアクセラレータ

ANN

単位時間あたりに多くの行列演算が可能なGPU、TPU

SNN

イベント駆動（非同期）なニューロモーフィックチップ

- ✓ IBM TrueNorth
  初めて作られたニューロモーフィックチップで現在アクセス不可
- ✓ Intel Loihi
  基礎から応用まで利用されている
  研究者のみアクセス可能
- ✓ SynSense SPECK™
  イベントカメラと相性が良く、エッジデバイスとして実用に向けた開発

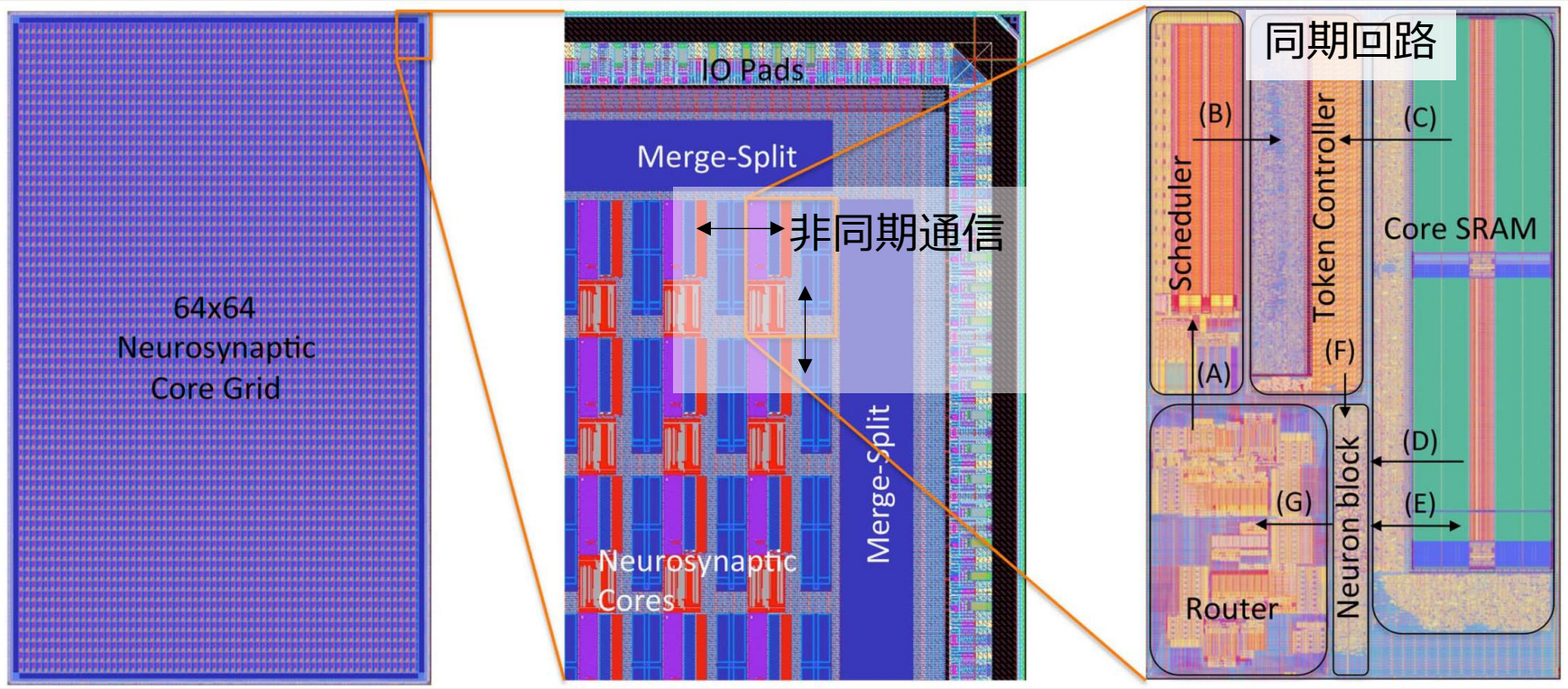
Loihiを大量に積んだサーバ
6ラックユニットで2600W
イベント駆動で15TOPS/W

https://www.intel.co.jp/content/www/jp/ja/newsroom/news/intel-builds-world-largest-neuromorphic-system.html

ハードウェアアクセラレータ

従来のニューロモーフィックチップ

- ✓大域非同期・局所同期方式で実装
  - 重み乗算など要素ごとの回路は消費電力の大きい同期回路
  - 要素間の通信は消費電力の小さい非同期回路



F. Akopyan et al., "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, no. 10, pp. 1537-1557, Oct. 2015, doi: 10.1109/TCAD.2015.2474396.

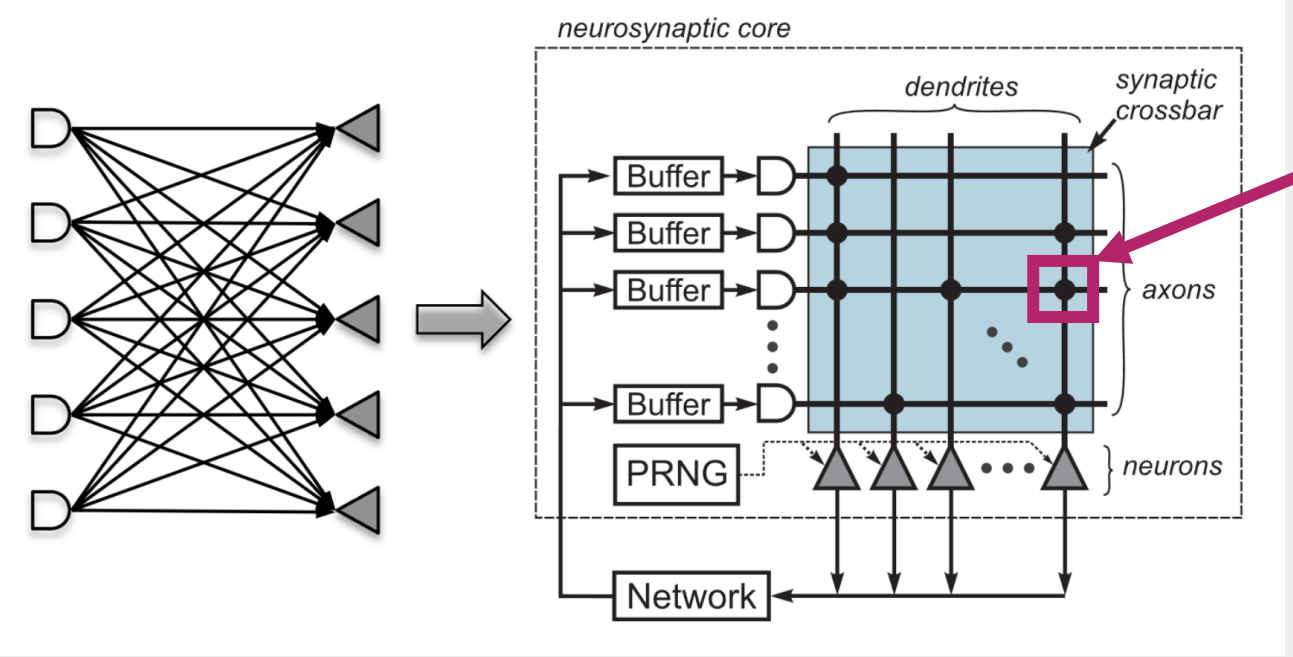
ハードウェアアクセラレータ

従来のニューロモーフィックチップ

✓コンピューティングインメモリ(CIM)の採用

計算回路とメモリを融合させることでメモリボトルネックを解消

メモリにはSRAMを採用



重みが保存されているSRAM

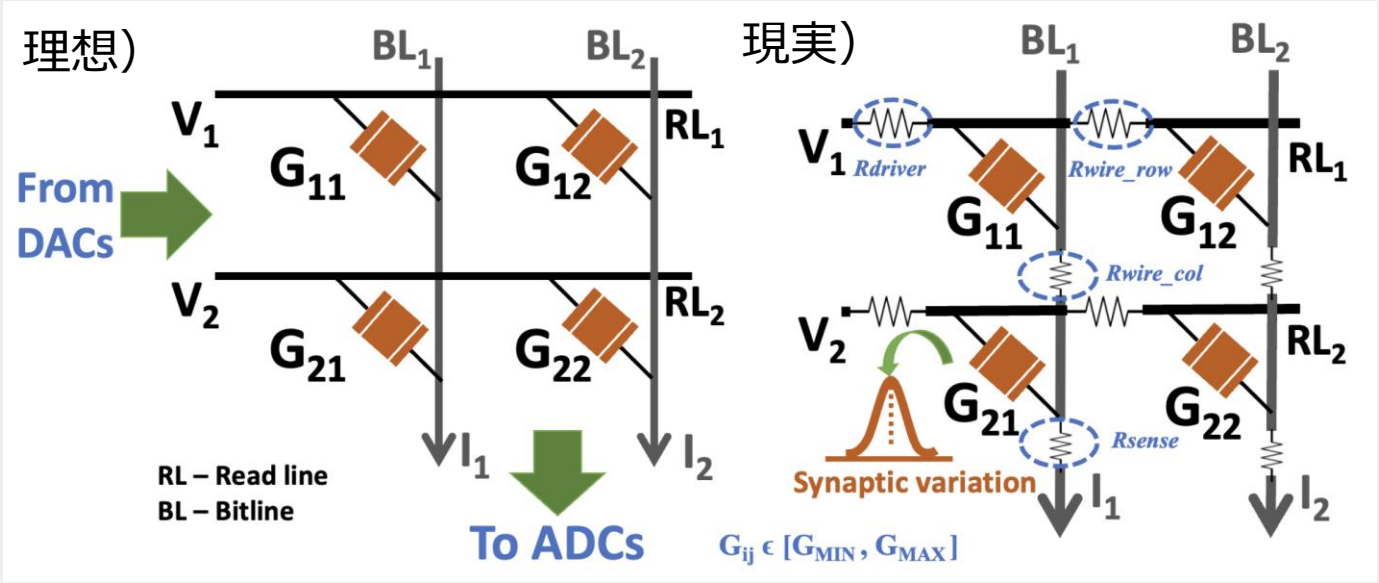
☹SRAMは揮発性で、データを再度書き込む必要あり

☹回路面積/消費電力/レイテンシの削減余地あり

ハードウェアアクセラレータ

現在目指されているニューロモーフィックチップ

- ✓CIMで採用するメモリとして不揮発性メモリの採用
  - 抵抗値として値を保存可能な抵抗変化型メモリ（RRAM）や相変化メモリ（PCM）



☹ このような不揮発性メモリはノイズや配線抵抗の影響を受けやすく、推論時に理想的な値から大きく変動してしまう

Abhiroop Bhattacharjee, Youngeun Kim, Abhishek Moitra, and Priyadarshini Panda. 2022. Examining the Robustness of Spiking Neural Networks on Non-ideal Memristive Crossbars. In Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED '22). Association for Computing Machinery, New York, NY, USA, Article 1, 1–6. https://doi.org/10.1145/3531437.3539729

ハードウェアの課題解決方針

- ✓Write-Verify
  - 重みの値が理想的な値となるように一つ一つ確認する手法
  - しかし、確認するための人的コストが大きい
- ✓補正回路を導入する方針
  - 正解データを保持し、定期的に重みと正解データを比較して補正する
  - しかし、回路のオーバーヘッドが大きい
- ✓Noise-aware学習
  - 学習時に使用するデバイスの特性を明らかにして学習
  - ノイズモデルが対象デバイスごとに異なり、汎用性がない

ハードウェアの課題まとめ

一般的な課題として…

- ✓ツールなどが未成熟なため学習コストが高い
- ✓非同期に通信を実施するため既存のシステムとの統合が難しい

	従来チップ	次世代チップ
回路面積	大きい	小さい
消費電力	大きい	小さい
レイテンシ	大きい	小さい
精度	高い	低い

目次

1. スパイキングニューラルネットワークの現状
2. 課題と解決への取り組み
3. まとめ
   A) 現状のまとめ
   B) 課題とその解決方針のまとめ

まとめ：現状

- ✓ANNよりも低消費電力 / エネルギなネットワークモデルとして様々な応用先が模索されている
- ✓ANNで達成できている認識タスクや生成タスクはあらたかSNNは推論可能
- ✓イベント情報と相性が良く、センサからアクセラレータまでSNNに特化したシステムを構築する取り組みも進められている

まとめ：課題と解決方針

✓学習：代理勾配法

推論時間が短いが、低精度であり学習コストも高い

損失関数やスパイク表現の方式を工夫する必要あり

✓学習：ANN-SNN変換

従来のANNと同等の精度を達成できるが、高レイテンシで複雑なモデル構造の変換は難しい

新たな活性化関数や事前知識を導入する必要あり

✓ハードウェア

ツールなどが未整備で学習コストが高い

抵抗値をメモリとしたハードウェアは回路面積・レイテンシ・消費電力が改善されるが、ノイズによって値が変動しやすく精度が低い

学習から推論のうちどこかで補正をする必要あり

これらの問題が解決すれば、Physical AIを支える
エッジAIモデルとして採用されるかもしれない

その他

今後、以下のレポジトリに機械学習向けのSNNの資料を作成していきます。

Github: https://github.com/hanebarla/SNN-tutorial