

Spiking Neural Network プロセッサ tinyODIN の並列化の検討

1250302 門屋 陽丈 【 コンピュータ構成学研究室 】

1 はじめに

近年, IoT 技術の発展に伴い, AI を搭載したエッジデバイス数が増加傾向にある. これらのエッジデバイスには高性能, かつ低消費電力な動作が必要とされ, 解決案として非同期な動作により, 低消費電力が期待される SNN(Spiking Neural Network) 専用ハードウェアの将来性が注目されている.

本研究では, 既存の SNN プロセッサ tinyODIN[1][2] に対し, エッジデバイスに必要とされる要件の 1 つである高性能化の観点から並列化手法を検討し, 回路規模と実行時間を評価した.

2 SNN プロセッサ ODIN/tinyODIN

SNN プロセッサ ODIN[1] は, N 個のニューロンが $N \times N$ のシナプスにより完全結合されたモデルをエミュレートできるようにクロスバー網を備えている. 各ニューロンには, 膜電位, 閾値, リーク電流の時定数が設定でき, 積分発火モデル LIF(Leaky Integrate-and-Fire) や Izhikevich モデルを選択できる機能を有している. これに対して, エッジデバイスの実装に適した tinyODIN が提案されている [2]. しかし, 回路規模を削減するために, 各ニューロンの計算を逐次的に処理する構成となっている. そこで, 本研究では tinyODIN の並列化による高速化を検討する.

3 tinyODIN の並列化

tinyODIN では, 各ニューロンおよびそれらのシナプス結合を SPI バス経由で初期設定する機能, および, 設定された SNN をシミュレートする機能を有しており, いずれの機能も逐次的に実行されている. エッジ AI デバイスでは一般に初期設定は一度であるが, SNN による推論は定期的に行われることを考慮すると, 推論機能を並列化するほうがより高性能化に寄与する.

図 1(a) に tinyODIN の回路構成を示す. 提案並列化構成では, 並列化度 p に対して, シナプス回路内の SRAM ワード長を p 倍にし, 深さを $1/p$ とする. これにより, シナプス回路から重み $weight$ を並列にニューロン回路に供給できる. また, ニューロン回路では, シナプスの重みを受け取り, ニューロン状態 NS (膜電位, 閾値, リーク電流時定数) を SRAM から読み出し, LIF モデルに基づく膜電位の計算と発火検出を行う. LIF の計算を p 並列で実行するために, p 個の LIF 回路を用い, それぞれに深さ $1/p$ の SRAM を接続する. この時, 各 SRAM のアドレスは, (ニューロン ID $\gg \log_2 p$) となる. 図 1(b) には, 最も簡単な $p=2$ の場合の提案並列化構成を示している.

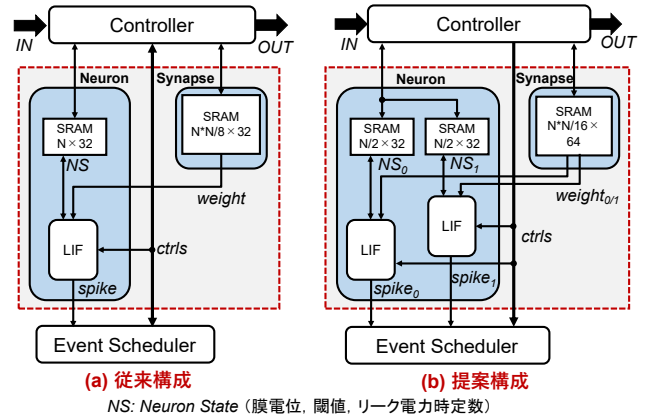


図 1 tinyODIN の構成とその並列化構成

4 回路設計・評価

本研究では, 提案並列化構成のうち並列化度 $p=2$ の回路の設計・合成・配置配線を試みた. 対象 FPGA チップは AMD 社 Zynq7010, EDA ツールは AMD 社 vivado を用いた. 比較のための tinyODIN 回路には公開ソースコード [2] を用いた. ニューロン数 $N=256$, シナプス数 $N \times N$ とした. 表 1 に, RTL シミュレーションで 256 ニューロンの LIF 計算に要する時間 $T[\text{ns}]$ (クロック周波数 250MHz 時) を測定した結果を示す. また, 回路規模として, LUT 数, レジスタ数, スライス数を比較した結果を示す.

並列化により回路規模 (Slices 数) 1.05 倍増に対して, ニューロン回路の計算性能は 2 倍に向上できる見通しが得られた. 今回は, シナプスの初期設定機能に若干バグが残っている状況での評価であるため, この点は今後の課題として残されている.

参考文献

- [1] C. Frenkel, et al., “A 0.086-mm² 12.7-pj/sop 64k-synapse 256-neuron online-learning digital spiking neuromorphic processor in 28-nm cmos,” IEEE Trans. Biomed. Circuits Syst., 2019.
- [2] C. Frenkel, “tinyODIN,” <https://github.com/ChFrenkel/tinyODIN> (最終閲覧日 2025/2/4).

表 1 ニューロン回路の動作時間 T と SNN プロセッサ全体の回路規模の比較

	T [ns]	LUTs	Registers	Slices
tinyODIN	1600	1116	1744	746
本提案 ($p=2$)	800	1263	1871	790