

水谷天祐, 清山浩司

(長崎総合科学大学大学院 工学研究科)

1. はじめに

近年、IoT(Internet of Things)の発展と共にエッジ側でのAI(Artificial Intelligence: 人工知能)処理需要が増加しており、低消費電力かつリアルタイムな推論を可能にするAIアクセラレータ(以降AIチップと略す)の開発が進められている。とりわけアナログ演算とデジタル処理を組み合わせたAIチップは、従来のデジタル中心のアーキテクチャに比べてエネルギー効率に優れ、センシングから推論までを一体化した処理が可能になる点で注目されている。本発表では、アナログ演算を積和回路に用いる検討を述べる。

2. アナログ演算の方法及び構成概略

図1に人工ニューロンのモデルと回路ブロックの対応を示す。図1(a)は、2個の入力ニューロン x_1, x_2 と掛け算の係数に対応する重みが積和演算後 y_1, y_2 として出力されるモデル例である。図中の丸(○)はニューロンで信号の入出力、矢印(→)はシナプスで重み g の掛け算の係数を示している。アナログ演算を用いるAIチップは、並列で積和演算を実現するため一般的にクロスバー構成が用いられる。本研究でも同様の構成を用いており、図1(b)に構成の概略を示す。図1(b)は図1(a)に対応させた2個の入出力部を示したクロスバー構成のブロック図である。クロスバー構成とは、入力側となるDAC(デジタル・アナログ変換器)に接続されたX方向の2本の信号線と、出力側になるADC(アナログ・デジタル変換器)に繋がるY方向の2本の信号線のクロスポイントに接合点を設け、 2^2 個の接続点を設ける。このクロスポイントにコンダクタンス素子等掛け算の係数に相当する重みを設けることにより行列積和演算が並列で実行可能となる。

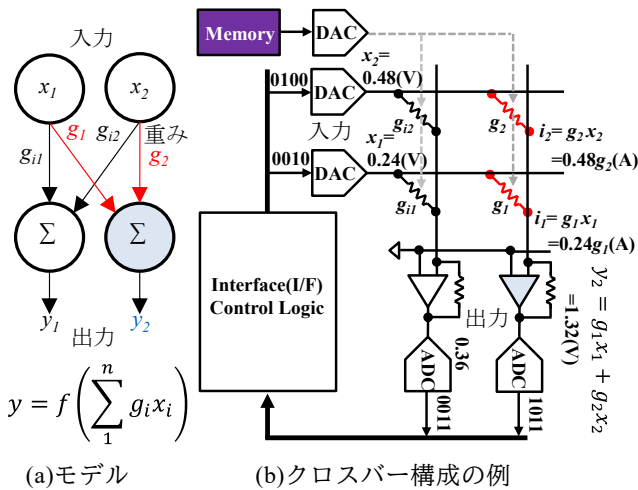


図1.ニューロンモデルと回路ブロックの概要

3. 積和演算回路の検討及びまとめ

クロスバー構成においては、コンダクタンス素子としてReRAMやPCM、MRAM、フラッシュメモリなど多様な不揮発性デバイスを利用する方式が提案・開発されている。本発表では図2に示すMOSFETの電流源機能を利用する回路の検討を述べる。図2右側に示す回路は1個所の接合点(シナプス)

である。nMOSのM1は、容量 C に書き込み保持されたゲート電圧 V_G によりドレイン電流 I_{ds} を制御する電流源であり重みの役割を果たす。nMOSのM2は、スイッチとして機能して入力信号の値に応じたパルス波形幅(Δt)で制御して信号と重みの積演算を実現する。和演算はY軸方向に設けた容量 C_{INT} から引き抜かれる電流で表現する。この回路は、最初に C_{INT} を電源電圧 V_{DD} にリセットおよび V_G の書き込みを行う。次にX軸方向の全ての行に入力信号を与えて積和演算を実行する。1個の回路の出力電圧 V_{out} は以下になる。

$$V_{out} = V_{DD} - I_{ds} \cdot \Delta t$$

積和演算終了後、演算結果は C_{INT} に保持してADCを介して外部に取り出される。

図3は、CMOS 65nmのSPICEモデルを用いて積和回路(1個の回路)の動作確認を行ったシミュレーション結果である。図の上段は C_{INT} のリセット制御および入力信号パルス x_i 、下段は出力電圧である。M1のサイズはチャネル長65nm、チャネル幅120nmとして電源電圧は1.0Vである。この結果の一例は、入力信号 Δt を固定して異なる4種類の V_G 電圧を設定した時の積和演算結果である(V_G の異なる出力電圧 V_{out} を同一プロットに重ねて表示)。この出力 V_{out} は、はじめに出力信号線に接続された容量 C_{INT} がリセットされ1Vになる。次に入力信号にあたる Δt を一定時間印可して容量 C_{INT} の電圧は、線形的に低下している。その後、2nsまで結果を保持させており、想定通りの結果が確認出来る。今回の検討では、エッジ推論で必要とされる最小分解能4bitを演算精度に設定した(入力信号と重みは共に $2^4=16$ 値)。回路動作や構成およびシミュレーション結果の詳細は、発表時に述べる。

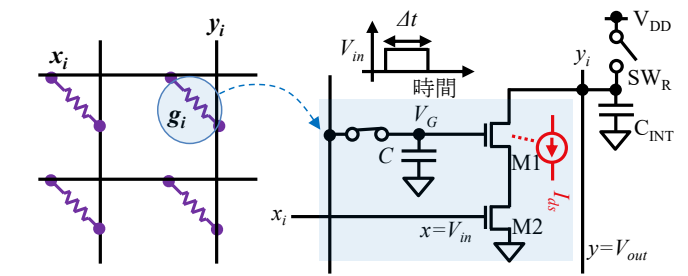


図2.接合点(ニューロン結合・積和演算)の構成

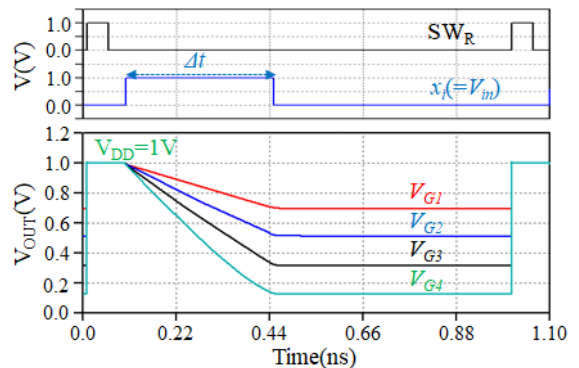


図3.積和回路のシミュレーション結果