

Atti

DELLA ACCADEMIA ROVERETANA DEGLI AGIATI

CCLXXIII ANNO ACCADEMICO

2023 ser. X, vol. V, B

Classe di Scienze matematiche, fisiche e naturali



SCRIPTA EDIZIONI

Diego Bisero

Neurocomputer e dispositivi elettronici basati sullo spin

ABSTRACT: Current computers spend huge computational resources to process cognitive and perception-related functions, which humans routinely perform every day. This has recently resulted in a great change in the field of computation where research efforts have been addressed to the development of a neurocomputer that attempts to mimic the human brain starting from the underlying physical processes characterizing the nanoelectronic components and thereby using its efficiency in recognition problems. This paper attempts to provide a review of the recent developments in the field of spintronic device based neuromorphic computing.

KEY WORDS: Neuro-Inspired Computing, Spintronic Devices, Domain Walls, Spin Torque, Tunnel Magnetoresistance, Artificial Neurons and Synapses.

RIASSUNTO: I computer attuali spendono enormi risorse computazionali per elaborare funzioni cognitive e relative alla percezione, che gli esseri umani svolgono abitualmente ogni giorno. Per questo gli sforzi di ricerca sono stati indirizzati negli ultimi anni verso lo sviluppo di quello che viene definito un "neurocomputer", ovvero un computer che tenta di imitare il cervello umano a partire dalla realizzazione di componenti nanoelettronici di base, che somigliano nel loro funzionamento alle strutture elementari (neuroni e sinapsi) del cervello, sfruttando in questo modo l'efficienza del cervello stesso nei problemi di riconoscimento. Questo articolo vuole fornire una rassegna dei recenti sviluppi nel campo del Calcolo Neuromorfico, basato sull'utilizzo specifico di dispositivi spintronici, che tentano di colmare il divario tra neuroscienze e nanoelettronica, superando i dispositivi CMOS attuali, che risultano particolarmente energivori.

PAROLE CHIAVE: Calcolo Neuromorfico, Dispositivi Spintronici, Pareti di Dominio, Effetto Spin Torque, Magnetoresistenza Tunnel, Neuroni e Sinapsi Artificiali.

Anche se il funzionamento del cervello non è ancora compreso pienamente, il calcolo neuromorfico, che tenta di emulare alcuni aspetti delle funzionalità del cervello e dell'inter-connettività, sta diventando sempre più popolare in

relazione a compiti di apprendimento automatico e sta superando le capacità degli esseri umani in specifici compiti cognitivi di elevata complessità. Possiamo citare come esempio il fatto che “Google DeepMind” ha sconfitto il campione del mondo di Go 19x19 (Silver, 2016), che è unanimemente considerato il gioco esistente più complesso strategicamente, anche rispetto agli scacchi; per chi non fosse pratico delle regole di Go può risultare illuminante un proverbio che sostiene che nessuna partita di questo gioco sia mai stata giocata due volte, circostanza verosimile se si pensa che ci sono 2×10^{170} differenti posizioni possibili.

L’ispirazione chiave dietro lo sviluppo di algoritmi e paradigmi di calcolo con alto grado di bio-fedeltà è guidata dall’aspettativa che emulando alcune caratteristiche del cervello umano, potremmo avvicinarci alle sue capacità cognitive molto efficienti e a basso consumo energetico. Per esempio, l’attuazione di paradigmi bio-realistici “spiking” di calcolo neurale, che verranno illustrati più avanti, ha recentemente permesso di realizzare funzionalità di apprendimento sinaptico a basso consumo.

Mentre questi modelli di calcolo neuro-ispirati sono ancora implementati in architetture tradizionali di von-Neumann, che includono logica booleana e circuiti di memoria, il cervello “artificiale” che si vorrebbe realizzare deve essere altamente parallelo, interconnesso e dotato di memorie sinaptiche in situ. Va detto anche che i transistor CMOS, che costituiscono le basi degli attuali sistemi informatici, sono sostanzialmente interruttori on-off adatti per l’informatica booleana, ma non sono intrinsecamente adeguati allo sviluppo di algoritmi neuro-mimetici.

Le architetture neuromorfiche basate su CMOS, essendo limitate da questa discrepanza tra le unità di calcolo e il sottostante hardware, consumano risorse e potenza che sono ordini di grandezza superiori a quelle coinvolte nel cervello biologico. Per colmare questa lacuna si rende necessaria l’esplorazione di dispositivi, circuiti e architetture che forniscono una migliore corrispondenza ai processi biologici e che richiedono un significativo ripensamento dell’informatica tradizionale basata su architettura di von-Neumann.

Mentre l’utilizzo di **dispositivi spintronici** in applicazioni di memoria ha raggiunto la maturità ed è ormai prossimo all’utilizzo sul mercato, esperimenti recenti in dispositivi magnetici basati sul moto di **pareti di dominio** stanno rivelando enormi possibilità di attuare funzionalità neurali e sinaptiche, attraverso l’utilizzo di singole strutture spintroniche, che sono in grado di operare a bassissime tensioni (Sengupta, 2016). Il semplice controllo delle dimensioni o della regione di polarizzazione dei dispositivi può favorire l’emulazione di funzionalità che variano dal comportamento di spiking dei

neuroni, alle abilità di apprendimento sinaptico nella stessa struttura magnetica. Le prospettive di miglioramento nella densità di integrazione e nel consumo di energia rendono i dispositivi spintronici un percorso promettente verso la realizzazione di una nanoelettronica bio-ispirata (Daniels, 2020).

I due fenomeni fisici principali che possono essere sfruttati per realizzare dispositivi neuromimetici basati sullo spin sono l'effetto spin-torque (STE, meccanismo di scrittura) e la magnetoresistenza tunnel (TMR, meccanismo di lettura). Ci riferiremo a questi due effetti nel seguito utilizzando i loro acronimi: STE e TMR, rispettivamente. La possibilità di manipolare gli stati di magnetizzazione tramite STE, senza l'utilizzo di un campo magnetico esterno, è stata prevista teoricamente nel 1996 (Slonczewski, 1996). Da allora numerosi esperimenti hanno dimostrato l'inversione della magnetizzazione di un sistema magnetico attraverso STE. D'altra parte, la possibilità di leggere lo stato di magnetizzazione tramite TMR è nota già dal 1975.

L'STE è un effetto in cui l'orientamento di uno strato magnetico può essere modificato utilizzando una corrente polarizzata in spin. Facciamo un passo indietro: i portatori di carica (come gli elettroni) hanno una proprietà nota come spin, che è una quantità piccola di momento angolare intrinseco. Una corrente elettrica, costituita da elettroni, è generalmente non polarizzata (contiene una metà di elettroni con spin su e l'altra metà con spin giù); una corrente polarizzata in spin, invece, possiede più elettroni di un determinato spin rispetto all'altro. Facendo passare una corrente attraverso uno strato magnetico, si può produrre una corrente spin-polarizzata. In un certo senso potremmo dire che lo strato magnetico svolge un ruolo di "filtro" di spin. Se una corrente spin-polarizzata attraversa un secondo strato magnetico più sottile, il momento angolare di spin può essere trasferito a questo strato, cambiando il suo orientamento. Questo effetto può essere usato per capovolgere l'orientamento dello strato magnetico.

La TMR è un effetto magnetoresistivo che si verifica in una giunzione magnetica a effetto tunnel, che è costituita da due ferromagneti sottili separati da un isolante ancora più sottile. Se lo strato isolante è di pochi nanometri, gli elettroni possono passare da un ferromagnete all'altro per effetto tunnel, fenomeno peculiare della fisica quantistica, non previsto dalla fisica classica.

I dispositivi magnetici basati sul moto di **pareti di dominio** hanno il vantaggio di poter associare più stati allo stesso sistema fisico. I magneti mono-dominio, infatti, sono caratterizzati da una magnetizzazione che può rappresentare solo due stati. I magneti multi-dominio possono essere associati a più di due stati, grazie alla presenza di regioni di transizione denominate pareti di dominio al loro interno. Lo stato del dispositivo può essere rappre-

sentato dalla posizione della parete di dominio o dalla proporzione relativa fra i due domini magnetici polarizzati in modo opposto. Il movimento di pareti di dominio indotto dal passaggio di una corrente è stato osservato in numerosi esperimenti e attribuito a STE associato a effetti locali di magnetizzazione indotta dal passaggio di elettroni nel materiale ferromagnetico (Beach, Tsoi, Erskine, 2008).

Come abbiamo già detto nel calcolo neuromorfico lo sforzo di ricerca è diretto verso l'imitazione dei meccanismi sinaptici e neurali a partire dai principi fisici di funzionamento dei dispositivi. Un pioniere di questo settore è stato Carver Mead che negli anni '80 ha proposto l'utilizzo di transistori CMOS che operano in un particolare regime definito sottosoglia, nel quale il meccanismo principale del trasporto dei portatori è la diffusione, che emula il meccanismo di flusso di ioni nei canali biologici dei neuroni. Sebbene i neuroni e le sinapsi CMOS siano ancora oggetto di studio da parte di numerosi gruppi di ricerca, essi richiedono transistor multipli e meccanismi di feedback per mimare le funzionalità di neuroni e sinapsi, che i dispositivi spintronici sono in grado di emulare in modo più diretto.

Il primo lavoro su calcolo neuromorfico spintronico può essere fatto risalire a Krysteczko *et al.*, che hanno esplorato la possibilità di implementazione delle funzionalità memristive nelle strutture magnetiche a giunzione tunnel (MTJ), attraverso fenomeni di commutazione indotti dalla tensione (Vincent, 2015). In questo caso la MTJ imita in modo diretto il funzionamento di una sinapsi, che si comporta come un memristore, che riceve in input una corrente e la moltiplica per la propria resistenza, giocando in questo modo il ruolo del peso sinaptico. La stabilità della magnetizzazione nelle MTJ permette di conservare il valore del peso sinaptico. Visto che la magnetizzazione è naturalmente bistabile, le MTJ sono ottimi candidati per realizzare reti neurali con pesi binari. E' anche possibile modificare i materiali o le geometrie in modo che le variazioni di magnetizzazione passino attraverso stati non-uniformi. Quest'idea ha permesso di realizzare sperimentalmente sinapsi analogiche, ma anche treni di piccole reti neurali con sinapsi con multi-stati magnetici.

Nella maggior parte degli algoritmi di reti neurali i neuroni semplicemente applicano una funzione non-lineare agli input sinaptici che ricevono. Le caratteristiche della dinamica non-lineare tipiche della spintronica possono essere sfruttate per imitare più da vicino i processi biologici, con un conseguente aumento delle funzionalità di calcolo. I neuroni biologici trasformano la differenza di potenziale sulla loro membrana in treni di spike elettrici, con una frequenza media che è dipendente non-linearmente dalla

differenza di potenziale. Le MTJ trasformano input di corrente continua in potenziali oscillanti con una frequenza che dipende non-linearmente dalla corrente iniettata. Questa proprietà è intrinsecamente ideale per imitare il funzionamento dei neuroni biologici. In questo senso può essere citata come esempio la realizzazione di un sistema di quattro nano-oscillatori spintronici addestrato a riconoscere le vocali attraverso dei pattern di sincronizzazione ad input RF.

Volendo tentare di approfondire ulteriormente il funzionamento fisico dei **dispositivi neuromorfici spintronici** bisogna sottolineare che esiste una similarità tra la funzione di trasferimento a “step” neuronale e un evento di inversione di una MTJ mono-dominio. La MTJ commuta fra due stati stabili (parallelo e antiparallelo) a patto che la corrente superi un determinato valore di soglia.

È possibile immaginare anche l’implementazione di funzionalità neuronali di tipo “non-step” in dispositivi spintronici. Una MTJ con strato magnetico mono-dominio è costituita da due stati stabili e di conseguenza solo due stati di output del neurone possono essere rappresentati da questo dispositivo. Se consideriamo MTJ con strati magnetici multi-dominio, costituiti da due regioni di magnetizzazione opposta separate da una parete di dominio, possiamo rappresentare stati multi-resistivi. E’ stato dimostrato sperimentalmente che un dispositivo resistivo multi-livello basato sul movimento di una parete di dominio può mostrare stati resistivi intermedi, realizzando un resistore programmabile a tre livelli.

Un aspetto interessante dei dispositivi basati sul movimento di pareti di dominio che riteniamo di dover sottolineare è la somiglianza fra la capacità di integrare gli effetti della corrente di una parete di dominio e la funzionalità specifica di un neurone “spiking” (Maass, 1997). Considerando impulsi di corrente (“spike” di input) che fluiscono attraverso il dispositivo a tempi diversi, la parete di dominio si sposta di una quantità proporzionale alla grandezza dello spike ogni volta che un impulso di corrente attraversa il dispositivo. Si noti che una rete neurale spiking è una rete neurale artificiale a impulso, che tenta di mimare le reti neurali naturali. Oltre allo stato sinaptico e neuronale una rete di questo tipo incorpora anche il concetto di tempo nel suo modello operativo. L’idea è che i neuroni artificiali attivino un ciclo di propagazione non in maniera automatica, ma piuttosto quando un potenziale di membrana – una intrinseca qualità del neurone correlata alla carica della sua membrana elettrica – raggiunge uno specifico valore. Quando un neurone si attiva genera un segnale che viaggia verso altri neuroni, che a turno incrementano o decrementano i loro potenziali in accordo a questo segnale.

La sfida futura sarà quella di creare circuiti spintronici neuromorfici su larga scala in grado di risolvere compiti utili a basso consumo energetico, specialmente per l'intelligenza artificiale (Grollier, 2020). Ciò richiede l'integrazione di strati di sinapsi e neuroni spintronici interfacciati in reti.

Nel breve termine, la spintronica neuromorfica dovrebbe vedere la commercializzazione di chip che immagazzinano pesi sinaptici nelle attuali e future generazioni di ST-MRAM. A questo primo passo potrebbe seguire lo sviluppo di circuiti neuronali hardware sfruttando le proprietà dinamiche di MTJ per implementare sinapsi e neuroni per inferenza e apprendimento. A lungo termine, materiali più esotici potrebbero offrire l'affascinante prospettiva del calcolo basato su effetti fisici complessi.

Bibliografia

- Beach G., Tsoi M., Erskine J., 2008, *Current-induced domain wall motion*, «J. Magn. Magn. Mater.» 320, 1272, <https://www.sciencedirect.com/science/article/abs/pii/S0304885307010141>.
- Daniels M.W. et al., 2020, *Energy-efficient stochastic computing with superparamagnetic tunnel junctions*, «Phys. Rev. Appl.» 13, 034016, <https://journals.aps.org/prapplied/abstract/10.1103/PhysRevApplied.13.034016>.
- Grollier J. et al., 2020, *Neuromorphic spintronics*, «Nat. Electron.» 3, 360, <https://www.nature.com/articles/s41928-019-0360-9>.
- Maass W., 1997, *Networks of spiking neurons: The third generation of neural network models*, «Neural Networks» 10, 1659, <https://www.sciencedirect.com/science/article/abs/pii/S0893608097000117>.
- Sengupta A., Banerjee A., Roy K., 2016, *Hybrid spintronic-CMOS spiking neural network with on-chip learning: Devices, circuits and systems*, «Phys. Rev. Appl.» 6, 064003, <https://journals.aps.org/prapplied/abstract/10.1103/PhysRevApplied.6.064003>.
- Silver D. et al., 2016, *Mastering the game of Go with deep neural networks and tree search*, «Nature» 529, 484, <https://www.nature.com/articles/nature16961>.
- Slonczewski J. C., 1996, *Current-driven excitation of magnetic multilayers*, «J. Magn. Magn. Mater.» 159, L1, <https://www.sciencedirect.com/science/article/abs/pii/0304885396000625>.
- Vincent A.F. et al., 2015, *Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems*, «IEEE Trans. Biomed. Circuits Syst.» 9, 166, <https://ieeexplore.ieee.org/document/7086084>.