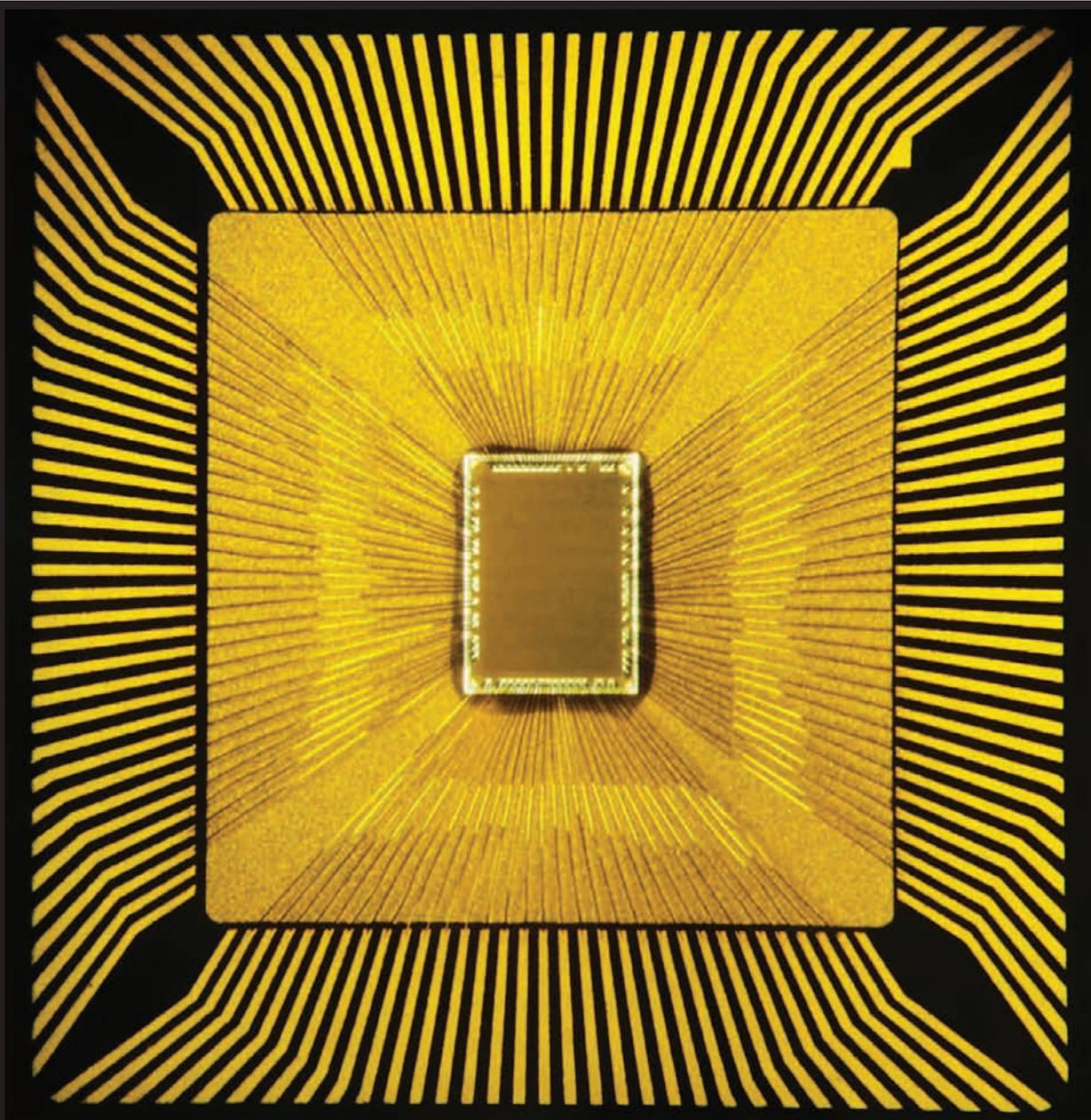


PENSIERI AL SILICIO

I microchip modellati sul cervello umano eccellono in compiti che mettono in serie difficoltà i computer attuali.

Tom Simonite



Questo chip, prodotto da IBM nel 2011, riproduce le caratteristiche di 256 neuroni e 1.024 sinapsi.

Fotografia: per gentile concessione di IBM Research

Immaginiamo una persona che sta leggendo queste parole su un laptop in una caffetteria. L'apparecchio fatto di metallo, plastica e silicio consuma circa 50 watt di energia nel tentativo di convertire i bit informativi – una lunga stringa di 1 e 0 – in una schiera di puntini sullo schermo. Contemporaneamente, nella mente di quella persona, un appiccicoso ammasso di proteine, sale e acqua utilizza una frazione di quell'energia per riconoscere gli schemi di lettere, parole, frasi, ascoltando anche la musica diffusa dalla radio.

I computer sono incredibilmente poco efficienti nello svolgimento di numerosi compiti che vengono facilmente risolti anche dal più semplice dei cervelli, come per esempio il riconoscimento di immagini e l'orientamento in uno spazio non familiare. Gli apparecchi dei laboratori di ricerca o i grandi centri dati sono in grado di offrire queste prestazioni, ma le loro dimensioni sono imponenti, consumano quantità spropositate di energia e devono venire programmati da esperti. Di recente, Google ha conquistato le cronache con un software che può riconoscere gatti e volti umani nei videoclip, ma a prezzo di una dotazione di oltre 16mila potenti processori.

Un nuovo tipo di chip modellato sul cervello dei mammiferi potrebbe colmare la distanza tra il calcolo naturale e quello artificiale, tra circuiti che svolgono operazioni logiche a velocità impetuose e un meccanismo affinato dall'evoluzione al fine di elaborare gli input sensoriali del mondo reale e interagirvi. I progressi delle neuroscienze e della tecnologia dei chip hanno reso possibile la produzione di congegni che, almeno su piccola scala, elaborano i dati allo stesso modo del cervello dei mammiferi. Questi chip "neuromorfici" potrebbero rappresentare il pezzo mancante di molti promettenti, ma incompleti, progetti di intelligenza artificiale, come le automobili che si guidano da sole in qualsiasi condizione o gli smartphone che interagiscono nella conversazione.

«I computer moderni sono un'eredità dei calcolatori, specializzati nel macinare numeri», sostiene Dharmendra Modha, un ricercatore di IBM Research, ad Almaden, in California. «Il cervello si è invece evoluto nel mondo reale». Modha guida uno dei due gruppi che hanno prodotto chip dotati di un'architettura di base copiata dal cervello dei mammiferi all'interno di un progetto da 100 milioni di dollari chiamato Synapse, finanziato dalla Defense Advanced Research Projects Agency del Pentagono.

I prototipi hanno già mostrato di possedere i primi barlumi d'intelligenza, elaborando le immagini con grande efficacia e acquisendo nuove capacità in modi che replicano l'apprendimento biologico. IBM ha creato una serie di strumenti per permettere agli ingegneri del software di programmare questi chip modellati sul cervello. L'altro prototipo, presso gli HRL Laboratories a Malibu, in California, verrà presto installato all'interno di un minuscolo aereo robotico, per imparare a riconoscere le zone che sorvola.

L'evoluzione dei chip modellati sul cervello è cominciata nei primi anni Ottanta con Carver Mead, professore del California Institute of Technology e uno dei padri delle moderne tecnologie informatiche. Mead ha legato il suo nome alla sigla "progettazione VLSI" (Very Large Scale Integration), che ha permesso alle aziende di creare microprocessori molto più complessi, che

hanno alimentato l'esplosiva crescita della potenza di calcolo e la diffusione generalizzata dei chip. Ma l'industria informatica non ha mai manifestato l'intenzione di abbandonare la strada intrapresa all'inizio, che risale al 1945: l'architettura di von Neumann, dal nome del matematico e informatico ungherese John von Neumann, che prevede l'esecuzione di una serie lineare di sequenze. I computer attuali, dagli smartphone ai supercomputer, sono dotati di due sole componenti principali: una unità di elaborazione centrale, o CPU, per manipolare i dati, e una memoria ad accesso casuale, o RAM, per archiviare i dati e le istruzioni su come gestirli. La CPU riceve i comandi iniziali dalla memoria, seguiti dai dati necessari a eseguirli. Immediatamente dopo avere svolto il compito assegnato, il risultato viene inviato alla memoria e il ciclo si ripete. Anche i chip multicore che gestiscono i dati in parallelo, si limitano a pochi processi lineari in simultanea.

Questo modo di procedere è figlio della riflessione logico-matematica che tende a risolvere i problemi affidandosi a catene lineari di ragionamento. Tuttavia, il metodo si è dimostrato poco adatto a elaborare grandi quantità di dati, specialmente input sensoriali come le immagini e i suoni.

I limiti erano anche di tipo strutturale: per rendere i computer più potenti, le aziende hanno prodotto chip sempre più complessi, in grado di velocizzare le operazioni sequenziali, ponendo gli ingegneri di fronte ai problemi connessi alle maggiori efficienze e al raffreddamento, perché più i chip erano rapidi più disperdevano calore. Mead, ora 79enne e professore emerito, aveva già intuito che ci poteva essere una strada diversa. «Più ci pensavo, più mi sembrava imperfetto», dice, seduto nel suo ufficio alla Caltech. All'inizio, Mead sognava chip che elaborassero una serie di istruzioni – addirittura milioni – in parallelo. Questi chip avrebbero dovuto svolgere nuovi compiti, gestendo efficacemente grandi quantità di informazioni non strutturate come i video o i suoni. Il sistema sarebbe stato più compatto e avrebbe consumato meno energia, anche se avesse dovuto svolgere operazioni molto particolari. La natura offriva esempi evidenti di tali sistemi: «I cervelli degli animali sono dotati di strutture che agiscono in parallelo».

I cervelli calcolano in parallelo e le cellule attive elettricamente al loro interno, chiamate neuroni, operano simultaneamente e ininterrottamente. Collegati in reti intricate da prolungamenti filiformi, i neuroni si influenzano inviando i rispettivi impulsi elettrici attraverso connessioni chiamate sinapsi. Quando l'informazione passa attraverso il cervello, i dati vengono elaborati come se fossero una serie serrata di picchi che si diffondono tra neuroni e sinapsi. Per esempio, il lettore riconosce le parole di un paragrafo grazie a un particolare schema di attività elettrica nel cervello alimentato dagli input che provengono dagli occhi. Inoltre, l'hardware neurale è anche flessibile: nuovi input alterano le sinapsi, modificando la gerarchia dei neuroni, così favorendo l'apprendimento. In termini informatici, si tratta di un sistema massivamente parallelo in grado di riprogrammare se stesso.

Paradossalmente, anche se ha ispirato il design informatico ancora oggi dominante, von Neumann ha allo stesso tempo aperto la strada ai nuovi computer modellati sul cervello. Nel suo libro incompiuto *The Computer and the Brain*, pubblicato

un anno dopo la morte nel 1957, si stupiva delle dimensioni, dell'efficienza e della potenza del cervello in confronto al computer: «Gli studi matematici più approfonditi del sistema nervoso possono modificare il nostro modo di osservare la matematica e la logica». Due decenni dopo, quando Mead arrivò alle stesse conclusioni, si rese conto che nessuno aveva provato a realizzare un computer sul modello del cervello: «Nessuno in quegli anni pensava: "Come costruirne uno?". Non avevamo la minima idea di come potesse funzionare».

A metà degli anni Ottanta, Mead produsse i suoi primi chip neuromorfici, come decise di battezzare i suoi congegni ispirati al cervello, dopo avere collaborato con alcuni neuroscienziati per comprendere come i neuroni elaborano i dati. Mead dispose dei normali transistor con voltaggi insolitamente bassi in reti di feedback che sembravano molto differenti dai raggruppamenti neuronali, ma funzionavano in modo simile. Cercò quindi di emulare i circuiti di elaborazione dati della retina e della coclea, con chip che rivelavano i contorni degli oggetti e le caratteristiche dei segnali audio. Ma i problemi da affrontare erano complessi e la tecnologia per la produzione di chip limitata. Considerando inoltre che allora i computer neuromorfici erano poco più di una "stranezza", Mead decise di passare ad altri progetti: «Era più complicato di quanto pensassi. Il cervello di un moscerino non sembra complicato, ma fa cose che noi ancora oggi non riusciamo a fare. Vorrà pure dire qualcosa!».

Dentro i neuroni

L'Almaden Lab di IBM si trova nelle vicinanze della Silicon Valley: il luogo ideale per ripensare le fondamenta dell'industria informatica. Per arrivarci, si prende una strada tra due file di piante di magnolie ai confini della città e ci si inerpica per 3 chilometri di curve. Il laboratorio si trova all'interno di 2.317 acri protetti di dolci colline. Al suo interno, i ricercatori si muovono lungo corridoi ampi e senza fine, avvolti nel silenzio, riflettendo sulle soluzioni dei problemi da affrontare. Modha dirige il più grande dei due

gruppi ai quali il DARPA ha affidato l'incarico di sdoganare il mondo dei computer dalla dipendenza dal pensiero di von Neumann. La filosofia di fondo è la stessa di Mead: la produzione di chip al silicio con componenti che funzionino come i neuroni. Ma il vantaggio è costituito dai progressi che nel frattempo si sono fatti nei campi delle neuroscienze e della produzione di chip. «Il momento è fondamentale; non era quello giusto per Carver», afferma Modha, che ha l'abitudine di chiudere gli occhi per pensare, respirare profondamente e infine parlare.

IBM produce chip neuromorfici con 6mila transistor per emulare il comportamento a picchi elettrici dei neuroni e favorire le connessioni tra questi neuroni al silicio. Per imitare il funzionamento del cervello, Modha si è ispirato agli studi sulla corteccia cerebrale, il grinzoso strato esterno. Anche se le diverse parti della corteccia svolgono funzioni differenti, come il controllo del linguaggio e dei movimenti, sono tutte formate da cosiddette microcolonne, con blocchi reiterati che variano dai 100 ai 250 neuroni. Modha ha reso pubblica la sua versione di una microcolonna nel 2011: una particella di silicio più piccola di una punta di spillo, con 256 neuroni al silicio e una memoria in grado di definire le proprietà di 262mila connessioni sinaptiche tra loro. La programmazione corretta di queste sinapsi può creare una rete che elabora le informazioni e reagisce come i veri neuroni del cervello.

Per vedere come il chip risolve un problema, è necessario programmare una simulazione del chip su un computer tradizionale e poi trasferire la configurazione al chip reale. In un esperimento, il chip ha riconosciuto cifre da 0 a 9 scritte a mano, riuscendo anche a prevedere il numero che si stava tracciando con una penna digitale. In un altro esperimento, la rete di chip è stata programmata per giocare una partita al videogame Pong. In un terzo, il sistema di chip ha permesso a un piccolo velivolo senza equipaggio di sorvolare la doppia linea gialla sulla strada che porta al laboratorio di IBM. Tutte queste azioni sono realizzabili anche con il software tradizionale, ma in questi casi sono state portate a termine con una frazione di codice, potenza e hardware normalmente richiesti.

Storia dei computer neurali

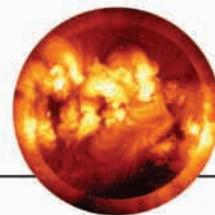
In *The computer and the Brain* (Il computer e il cervello, Bompiani), John von Neumann scrive: «Una comprensione matematica più profonda del sistema nervoso può cambiare il modo in cui guardiamo alla matematica e alla logica».

Carver Mead pubblica un articolo su un chip al silicio che replica il processo di elaborazione visiva della retina.

1958



1987



1996

Due retine artificiali al silicio vengono caricate a bordo di un pallone aerostatico sull'Antartide, per aiutare l'osservatorio Flare Genesis a mantenere la sua telecamera puntata verso il Sole.

Fotografie: Per gentile concessione di Wikimedia Commons (Von Neumann), per gentile concessione di Jaxa (Il sole)

Tradizionalmente si è cercato di accrescere la capacità di calcolo e la potenza degli algoritmi, ma questo approccio mostra ormai la corda.

Modha sta sperimentando le prime versioni di un chip più complesso, costituito da una griglia di nuclei neurosinaptici inseriti in una corteccia rudimentale, che arriva nel complesso a oltre un milione di neuroni. La scorsa estate, IBM ha annunciato un progetto di architettura neuromorfica basata su blocchi modulari di codice chiamati *corelets*. L'intenzione è di organizzare un menu di blocchi costruttivi componibili e riutilizzabili, senza trovarsi a lottare con i problemi legati alle sinapsi e i neuroni di silicio. Oltre 150 *corelets* sono già stati ideati, per compiti che vanno dal riconoscimento delle persone nei video alla capacità di distinguere tra la musica di Beethoven e di Bach.

Macchine che apprendono

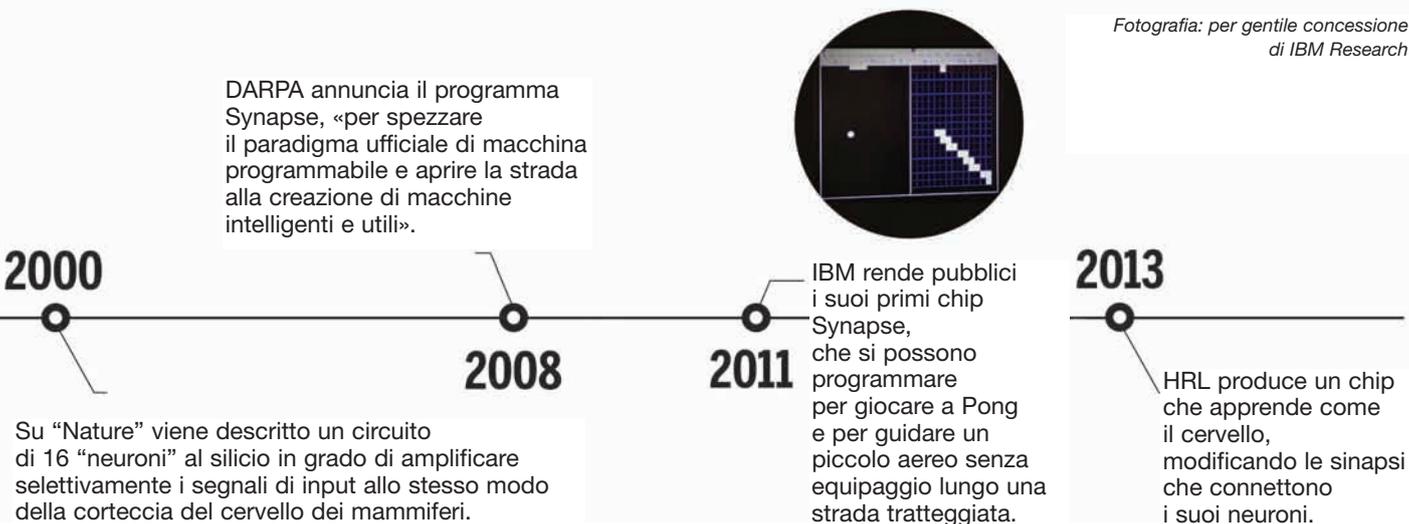
Su un'altra collina californiana, circa 500 km più a sud, la seconda parte del progetto DARPA mira a produrre chip che imitino ancora più da vicino il funzionamento del cervello. HRL, che spinge il suo sguardo oltre Malibu dalle colline della catena montuosa delle Santa Monica Mountains, è stata fondata da Hughes Aircraft e opera oggi sul mercato come joint venture tra General Motors e Boeing. Con un *koi pond* (un piccolo stagno), le palme e le piante di banane, l'ingresso ricorda gli hotel dell'epoca d'oro di Hollywood. Spicca, tra le altre cose, una placca commemorativa dedicata al primo laser funzionante, prodotto nel 1960 da quelli che allora erano chiamati gli Hughes Research Labs.

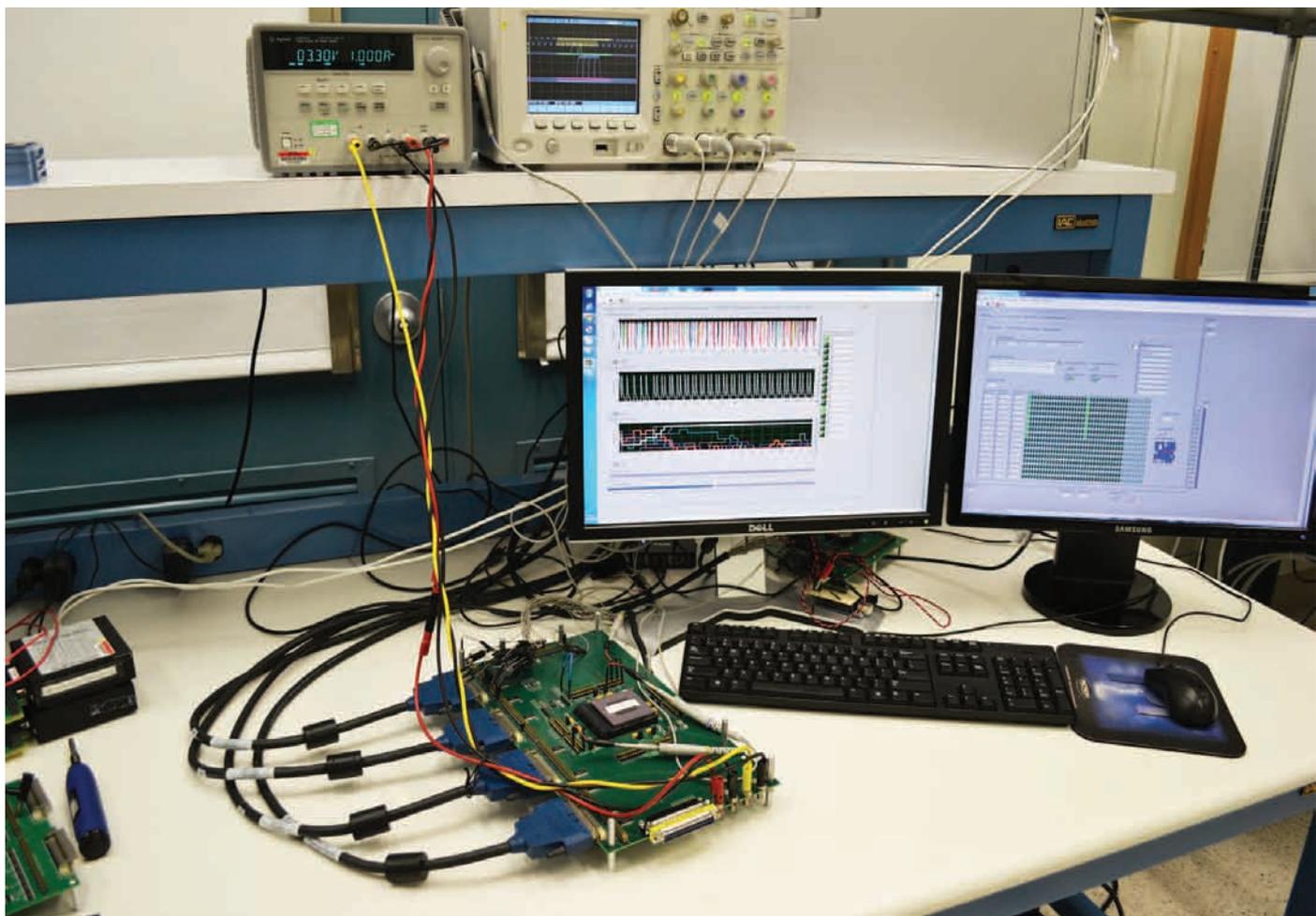
Su un banco in un laboratorio senza finestre, il chip di Naran Srinivasa s'intravede all'interno di un nugolo di cavi. L'attività dei suoi 576 neuroni artificiali appare su uno schermo di computer sotto forma di una sfilata di picchi, una specie di elettroencefalogramma di un cervello al silicio. Il chip di HRL ha neuroni e sinapsi molto simili a quelli di IBM. Ma a somiglianza dei neuroni nei nostri cervelli, quelli del chip di HRL modificano le loro connessioni se esposti a nuove informazioni. In altre parole, il chip apprende con l'esperienza.

Il chip di HRL imita due processi d'apprendimento cerebrali. Uno avviene quando i neuroni diventano più o meno sensibili ai segnali di un altro neurone a seconda della frequenza con cui riceve questi segnali. L'altro appare più complesso: un meccanismo che si ritiene sostenga l'apprendimento e la memoria, a cui è stato dato il nome di plasticità sinaptica, dipendente dalla distribuzione temporale degli impulsi nervosi. Questo fenomeno fa diventare i neuroni più "vicini" ad altri neuroni con cui hanno avuto più contatti in passato. Se gruppi di neuroni collaborano in modo costruttivo, le connessioni tra loro si rinforzano, mentre quelle meno utili cadono in uno stato dormiente.

I risultati degli esperimenti con versioni simulate di chip sono impressionanti. Il chip ha giocato a Pong, esattamente come quello di IBM. A differenza di quest'ultimo, però, il chip di HRL non era programmato per fare il gioco, ma solo per muovere la racchetta, rilevare la pallina e ricevere il feedback positivo o negativo sul colpo effettuato. Dopo solo 5 incontri, i 120 neuroni del chip erano diventati un temibile giocatore. «Nessuna programmazione. Le uniche indicazioni sono: "Ben fatto" e "Pessimo colpo" e il chip prefigura quello che dovrebbe fare», spiega Srinivasa. Se si aggiungono più palline, altre racchette o nuovi avversari, la rete di neuroni si adatta rapidamente ai cambiamenti.

Questo approccio potrebbe consentire agli ingegneri di creare un robot che passi attraverso una sorta di stadio "infantile", in cui sperimenta come muoversi e cosa fare. «Non si può tenere conto della varietà delle cose che accadono nel mondo reale. L'opzione





Un microchip prodotto da HRL apprende come il cervello biologico, rafforzando o indebolendo le connessioni sul modello delle sinapsi.
 Fotografia: per gentile concessione di HRL

migliore sarebbe che il sistema interagisse direttamente con l'ambiente circostante», sostiene Srinivasa. Le altre macchine potrebbero incorporare quanto è stato appreso da quella originale. Ma sarebbe un vantaggio considerevole se i robot potessero apprendere oltre i limiti prefissati, in quanto potrebbero adattarsi se danneggiati o modulare il loro passo a seconda del terreno.

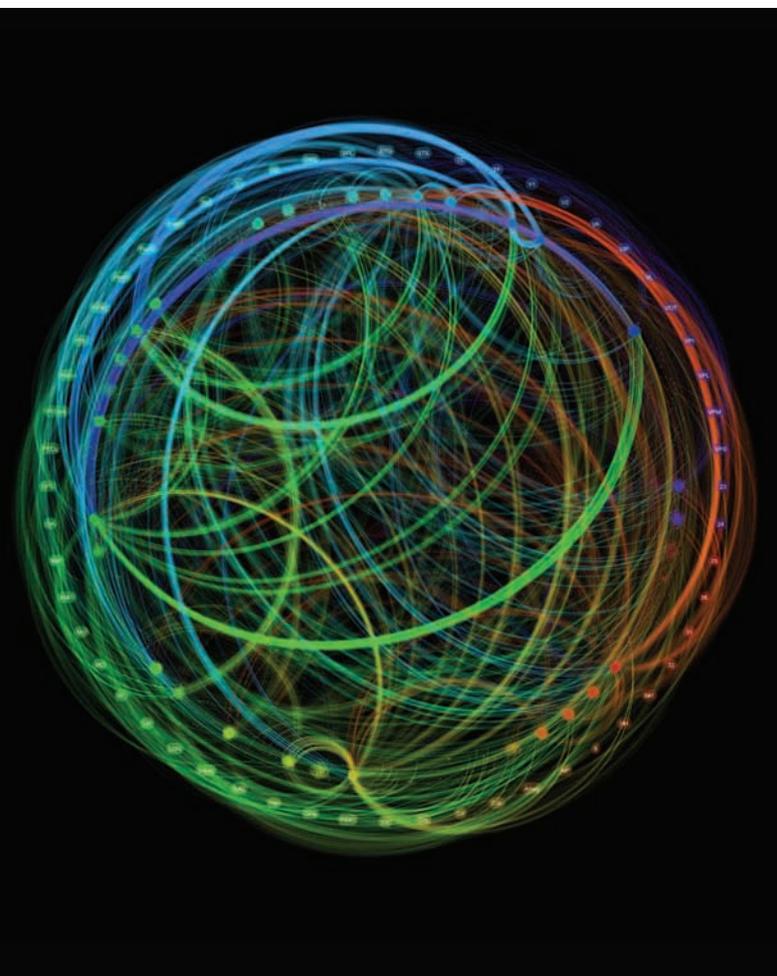
Il primo vero test per i computer neuromorfici arriverà la prossima estate, quando il chip di HRL dovrà uscire dal laboratorio e prendere posto nello Snipe, un velivolo dotato di ali, grande come il palmo di una mano. Mentre un umano utilizza i comandi remoti per guidare il piccolo aereo attraverso una serie di stanze, il chip raccoglierà i dati grazie ai sensori e alla telecamera. A un certo punto al chip sarà dato un segnale che significa "Fai attenzione ora". Al passaggio successivo in quella stanza, il chip dovrà accendere una luce per segnalare che si ricorda dell'istruzione ricevuta. Oggi, questo tipo di prestazione con un piccolo aeromobile richiederebbe normalmente una potenza di calcolo e un consumo di elettricità di gran lunga superiori.

Intelligenza artificiale e apprendimento profondo

Al di là dei modesti, ma significativi, successi dei chip Synapse, non è ancora chiaro se la produzione su scala industriale di questi chip porterà alla creazione di macchine dotate di facoltà sofisticate come i cervelli umani. Alcuni critici dubitano che sarà mai possibile per gli esperti di tecnologie copiare la biologia al punto di riprodurre queste abilità.

Il neuroscienziato Henry Markram, che ha scoperto la plasticità sinaptica dipendente dalla distribuzione temporale degli impulsi nervosi, ha attaccato il lavoro di Modha sulle reti di neuroni simulati, sostenendo che il loro comportamento è semplicistico. Markram ritiene che per emulare le facoltà cerebrali con successo sia necessario replicare le sinapsi su scala molecolare; il comportamento dei neuroni è influenzato dalle interazioni di decine di canali ionici e migliaia di proteine, egli spiega, e ci sono numerosi tipi di sinapsi che agiscono in modo caotico e non lineare. Secondo Markram, per replicare le capacità di un cervello reale, gli scienziati dovrebbero incorporare tutte queste caratteristiche.

I gruppi del progetto DARPA controbattono che non è necessario riproporre la complessa struttura del cervello per ottenere dei risultati e che le future generazioni dei loro chip si avvicineranno sempre più al corrispettivo biologico. HRL spera d'implementare i suoi chip facendo in modo che i neuroni ai sili-



IBM ha utilizzato questa simulazione di lunghi tracciati neurali in un macaco per impostare il design dei chip neuromorfici.
Fotografia: per gentile concessione di DARPA

cio si attivino come quelli cerebrali e IBM sta cablando le connessioni tra i nuclei del suo ultimo chip neuromorfico in modo nuovo, sfruttando le indicazioni provenienti dalle simulazioni delle connessioni tra le diverse aree della corteccia di un macaco. Modha crede che queste connessioni potrebbero risultare importanti per attivare le funzioni cerebrali superiori. Comunque, anche considerando questi passi in avanti, i chip sono ancora lontani dal cogliere la complessa realtà mentale. Sembra improbabile che i microchip possano replicare in un singolo centimetro quadrato le 10 miliardi di connessioni sinaptiche, anche se HRL sta conducendo esperimenti con una forma più densa di memoria basata sul memristor, vale a dire un componente elettronico passivo.

Il design neuromorfico è quasi del tutto assente dai computer attuali. Appare più logico considerare questi chip come un mondo completamente a parte, una nuova, aliena forma d'intelligenza.

Sarà anche aliena, ma Zachary Lemnios, il responsabile della ricerca di IBM, prevede che diventeranno ben presto un oggetto

familiare. Molte grandi aziende sono già alla ricerca di nuovi tipi di intelligenza computazionale. Spiega Lemnios che «l'approccio tradizionale consiste nell'aggiunta di capacità di calcolo e di algoritmi più potenti, ma si arriva a un limite dello sviluppo e ci si sta già scontrando con questo problema».

Come esempi, cita Siri, l'assistente personale di Apple, e le macchine che si guidano da sole, di Google. Queste tecnologie non sono molto avanzate nella comprensione del mondo intorno, aggiunge Lemnios.

Per muoversi, le macchine di Google si affidano in buona parte ai dati delle mappe precaricate, mentre Siri si collega ai remoti server del cloud per il riconoscimento vocale e l'elaborazione del linguaggio, provocando cospicui ritardi.

Oggi, la punta di diamante del software d'intelligenza artificiale è una disciplina conosciuta con il nome di "apprendimento profondo", a cui fanno riferimento, tra gli altri, Google e Facebook. Questo sistema si basa sulle reti neurali, circuiti costruiti in modo analogo alle sinapsi del cervello, che possono "apprendere" modificando le proprie connessioni fino al momento in cui a ogni ingresso corrisponde l'uscita desiderata.

Ma il *deep learning*, che ha generato il software per il riconoscimento dei gatti di Google, si affida a estesi cluster di computer per simulare le reti neurali e alimentarle con i dati.

Le macchine neuromorfiche dovrebbero permettere di produrre apparecchi compatti ed efficienti dotati di queste capacità per affrontare le situazioni in cui è complicato connettersi a un centro dati remoto.

IBM ha già preso contatti con clienti interessati ai sistemi neuromorfici. L'analisi dei video per la sicurezza e la prevenzione delle frodi finanziarie sono la linea avanzata di questa ricerca, in quanto entrambi i settori richiedono un apprendimento complesso e un riconoscimento di schemi in tempo reale.

Qualsiasi sarà l'utilizzo finale dei chip neuromorfici, non potrà prescindere dalla collaborazione con le macchine di von Neumann. I numeri dovranno sempre venire manipolati e anche nei sistemi che si scontreranno con i problemi relativi alle analisi delle immagini, sarà più semplice ed efficace disporre di un computer tradizionale. I chip neuromorfici potranno venire impiegati per compiti particolari, così come il cervello si affida a diverse aree specializzate per compiere lavori differenti.

Come è successo spesso nella storia dei computer, il primo di tali sistemi sarà probabilmente utilizzato a scopi militari. «Niente di mistico o magico», dice Gill Pratt, che dirige il progetto Synapse alla DARPA, a proposito dei computer neuromorfici. «Si tratta di una differenza strutturale che porta a un diverso equilibrio tra energia e performance». Pratt sostiene che, in particolare, gli aerei a pilotaggio remoto (UAV) godranno dei vantaggi di questa tecnologia. I chip neuromorfici potrebbero riconoscere contrassegni od obiettivi senza il trasferimento di dati ingombranti e la potenza dei computer tradizionali ora necessari per elaborare le immagini.

In questa visione di un nuovo chip si sarebbero certamente rispecchiati Mead e von Neumann. ■

Tom Simonite è responsabile dell'area software e hardware della edizione americana di *MIT Technology Review*.