

# Apprendimento dinamico della memoria di lavoro: una realizzazione in elettronica analogica

Daniel J. Amit, Paolo Del Giudice e Stefano Fusi

June 9, 1998

## **La memoria di lavoro e la classificazione**

La classificazione degli stimoli ambientali viene considerata uno dei misteri fondamentali del sistema cognitivo realizzato dal cervello ([1]), in grado di elaborare stimoli ambientali che si presentano ad esso in gran numero e possono cambiare rapidamente. Questi stimoli inoltre si presentano nel tempo in sequenze imprevedibili, ed anche quando lo stesso stimolo appare più volte, non si presenta alla retina nello stesso modo. Malgrado ciò il cervello riesce a creare delle classificazioni, che da un lato raggruppano stimoli simili e dall'altro consentono un'efficace discriminazione di stimoli diversi. Le classificazioni naturalmente vengono apprese e utilizzate a diversi livelli: al livello più semplice verranno raggruppati stimoli che si presentano con caratteristiche simili agli apparati sensoriali, mentre una gerarchia di classificazioni via via più complesse corrisponderà per esempio alla dipendenza dal contesto o alle relazioni con il linguaggio.

In questa sede affronteremo il problema della classificazione al livello più semplice, e restringeremo ulteriormente il contesto focalizzandoci sui tipi di classificazioni non pre-programmate geneticamente ma apprese dall'esperienza.

## **La memoria attiva e la classificazione in neurofisiologia**

Un modo per legare il fenomeno (psicologico o psicofisico) della classificazione al substrato cerebrale, consiste nello studiare la neurofisiologia di animali su-

teriori durante lo svolgimento di compiti che comportano classificazione. Nel corso di una lunga tradizione sperimentale, iniziata con il lavoro pionieristico di Fuster e Niki[2, 3] dei primi anni settanta, si è giunti ad un paradigma sperimentale in cui delle scimmie vengono addestrate a seguire una lunga sequenza di ‘trial’, ognuno secondo il seguente protocollo (figura 1): la scimmia fissa uno schermo su cui si presenta per breve tempo una figura astratta ‘campione’ (*sample*) e, dopo un ritardo relativamente lungo, la deve paragonare con un’altra figura astratta di ‘confronto’ (*test*) che viene scelta a caso (con il 50% di probabilità è uguale alla figura campione). In una versione recente di questo tipo di esperimenti (detti DMS, da *delayed match to sample*, corrispondenza ritardata con il campione), il gruppo di Miyashita [11] è riuscito ad addestrare delle scimmie ad ottenere buone prestazioni su un centinaio di stimoli campione del tipo rappresentato in figura 1. Queste figure ricadono in due categorie: “frattali” (fila in alto) e descrittori di Fourier (in basso). Entrambi i tipi vengono generati da algoritmi che contengono degli elementi casuali; la natura astratta e casuale di queste immagini assicura che esse non appartengano al patrimonio di esperienza passata, o a quello ereditario, della scimmia. L’elaborazione di queste immagini da parte del sistema biologico è molto probabilmente il risultato di un apprendimento a partire dall’esperienza.

Dopo un periodo di addestramento relativamente lungo, in cui ogni stimolo viene presentato molte volte sia come *campione* che come stimolo di confronto, si osserva un insieme di distribuzioni di attività neuronale durante il periodo di ritardo tra la presentazione della figura campione e quella di confronto, nella parte anteriore della corteccia inferotemporale della scimmia (si veda la figura 1): dopo ogni stimolo *campione*, in assenza di stimolazione, nell’intervallo temporale che precede lo stimolo di confronto, si osserva che alcune delle cellule mantengono frequenze di emissione di impulsi significativamente più alte rispetto a quelle precedenti lo stimolo campione. Ogni stimolo campione seleziona un insieme diverso di cellule candidate ad avere una frequenza più elevata. In questo senso possiamo dire che la distribuzione di frequenze delle cellule nel periodo di *ritardo* (DAD, da *Delay Activity Distributions*) costituisce una rappresentazione interna dello stimolo campione, e preserva informazione su di esso fino all’arrivo dello stimolo di confronto, alcuni secondi dopo la scomparsa dello stimolo campione. Inoltre, per tutte le 100 immagini le diverse distribuzioni di attività si manifestano in un’area ristretta (1 mm<sup>2</sup>) della corteccia inferotemporale, parte C in figura.

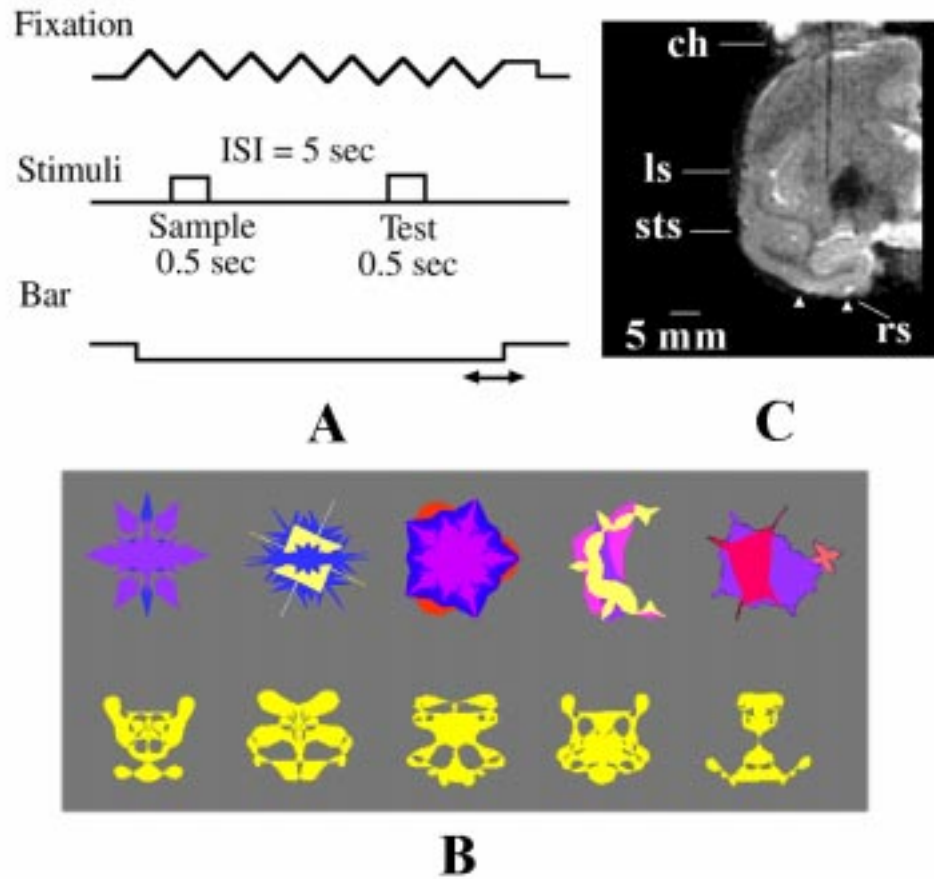


Figure 1: A. Schema del protocollo temporale di un 'trial': all'inizio appare sullo schermo un punto di fissazione intermittente (riga in alto) che cattura l'attenzione della scimmia durante tutto il trial. La scimmia all'inizio del trial deve premere una barra e tenerla abbassata finchè, dopo la presentazione dello stimolo di confronto, deve muovere la barra a destra o a sinistra a seconda che esso sia o meno uguale allo stimolo campione. B. esempio di stimoli usati nell'esperimento: frattali (in alto) e descrittori di Fourier (in basso). C. immagine di una sezione del cervello ottenuta con tecniche di risonanza magnetica nucleare: la striscia scura indica la zona dove passa l'elettrodo mentre i triangoli bianchi delimitano la parte della corteccia inferotemporale dove avvengono le registrazioni ([16]).

In che senso si possono considerare le distribuzioni di attività osservate nel periodo di ritardo come delle classificazioni? In primo luogo il numero di tali distribuzioni, che si formano durante l'addestramento, sono una piccola frazione del numero totale di stimoli visivi cui la scimmia è stata sottoposta durante quel periodo. Sia durante le sessioni di lavoro, che costituiscono una parte relativamente piccola del periodo di veglia della scimmia, sia nel resto del tempo, la scimmia riceve un numero enorme di stimoli. Inoltre ogni stimolo, quando viene presentato sullo schermo, provoca delle risposte visive alquanto diverse nella corteccia della scimmia, a causa di piccoli movimenti oculari, o del rumore presente lungo il percorso visivo dalla retina alla corteccia inferotemporale. Queste variazioni si possono osservare nelle diverse risposte delle cellule alla stessa immagine. Ciò nonostante per ogni 'classe' (insieme di stimoli simili) corrispondente ad una data immagine viene propagato un solo schema di attività nel periodo di ritardo che segue la scomparsa dello stimolo. D'altra parte, sovrapponendo del rumore alle immagini stesse, come nel lavoro di Amit, Fusi e Yakovlev [14], si osserva che tutto un insieme di stimoli visivi simili ad una delle immagini generate provoca la stessa distribuzione di attività nel periodo di ritardo.

Infine, immagini che non siano state usate nel corso dell'addestramento non sembrano provocare alcuna attività selettiva nel periodo di ritardo, almeno in questa regione della corteccia.

Il quadro che emerge da questi esperimenti è di un piccolo modulo (più o meno di  $1 \text{ mm}^2$ , con circa  $10^5$  cellule) nella corteccia infero-temporale anteriore che, dopo aver ricevuto una sequenza molto lunga di stimoli (ogni immagine viene presentata centinaia di volte) riesce ad estrarre la presenza di insiemi di stimoli tra di loro simili, nel contesto di un compito dato, e a generare una sola rappresentazione per ognuno di questi insiemi, cioè la corrispondente distribuzione di frequenze di emissione durante il periodo di ritardo. Questa distribuzione di frequenze costituisce una rappresentazione interna della classe.

## **Ruolo psicologico delle DAD**

Gli esperimenti DMS che abbiamo descritto suggeriscono un ruolo utile e necessario delle distribuzioni di attività nel periodo di ritardo, come meccanismo per trasmettere informazione su uno stimolo ai fini di una elaborazione che deve avvenire parecchio dopo la scomparsa dello stimolo stesso. Al momento

in cui appare sullo schermo la seconda immagine, la prima è scomparsa da tempo, ma l'informazione su di essa è essenziale per il confronto con la seconda immagine e svolgere così il compito. Compiti di questo tipo sono molto comuni nella esperienza cognitiva quotidiana degli esseri umani: ad esempio il caso in cui si debba andare ad incontrare all'aeroporto qualcuno che si conosce, e ci viene detto il nome della persona da incontrare. L'incontro potrà avvenire dopo un tempo relativamente lungo dal momento in cui si è ascoltato il nome; ciò nonostante l'informazione relativa all'aspetto della persona sarà mantenuta attiva, e risulterà disponibile al momento giusto.

Questa situazione è simile a quella osservata negli esperimenti compiuti dal gruppo della Goldman-Rakic[4], in cui si addestrano delle scimmie a spostare gli occhi verso la direzione indicata da uno stimolo visivo che funge da indizio, alcuni secondi dopo la scomparsa dell'indizio. Anche in questi esperimenti si trovano delle cellule in una porzione molto localizzata della corteccia che mostrano, nel periodo di ritardo, una attività selettiva rispetto alla direzione.

Come si collocano queste DAD nella discussione sui tipi di memoria a breve, lungo o medio termine? In effetti, il tipo di memoria realizzata dalla attività nel periodo di ritardo dovrebbe essere comune a tutte e tre le categorie. L'informazione trasmessa oltre il periodo di ritardo non viene appresa durante il compito, ma nella lunga esperienza che lo precede. Le immagini sono dunque familiari; una di queste immagini familiari deve essere posta in uno stato attivo dal primo stimolo, così da consentire delle operazioni su di esso nel periodo successivo alla sua scomparsa. Il suggerimento è di riservare i termini breve, lungo termine ecc. al periodo della vita dell'animale in cui una particolare immagine rimane familiare, il periodo cioè in cui quella immagine può ancora generare una specifica distribuzione di attività nel periodo di ritardo.

## DAD e paradigma Hebbiano

Per usare la terminologia di Hebb, potremmo dire che la memoria, il processo di familiarizzazione, è il processo nel quale, durante l'addestramento, si imprime un *engramma* nelle sinapsi [5]. I cambiamenti sinaptici indotti dall'apprendimento producono il sostrato per la propagazione della attività nel periodo di ritardo da parte dell'insieme locale di neuroni (la *riverberazione* nella terminologia di Hebb). La sede della memoria è dunque la sinapsi, e

l'intervallo di tempo in cui vengono preservati i cambiamenti sinaptici stabilisce la persistenza della memoria. L'attività riverberante nel periodo di ritardo corrisponde, in questa interpretazione, al fatto che uno stimolo 'familiare' (appreso) pone la memoria corrispondente, codificata nella struttura sinaptica, in uno stato attivo, in grado di mantenere disponibile nel tempo, in modo autonomo, l'informazione sullo stimolo appreso.

L'apprendimento ha luogo durante la presentazione di una immagine (stimolo) che aumenta, in un modo selettivo, la frequenza di emissione di un sottoinsieme dei neuroni. Le sinapsi che connettono due neuroni con risposte forti allo stimolo (frequenze elevate di emissione) si potenziano (diventano più efficaci), mentre quelle che connettono coppie di cellule con attività anticorrelate si indeboliscono (depressione sinaptica). Questi due effetti sono denominati rispettivamente LTP (da *Long Term Potentiation*, potenziamento a lungo termine) e LTD (da *Long Term Depression*, depressione a lungo termine). Ci si attende dunque che una simile dinamica della matrice sinaptica, in risposta agli stimoli in ingresso, possa generare una struttura sinaptica in grado di sostenere una distribuzione di attività simile a quella indotta dallo stimolo, anche dopo la scomparsa dello stimolo. Inoltre la struttura sinaptica generata deve essere in grado di sostenere in modo autonomo un'ampia varietà di riverberazioni, come nell'esperimento descritto sopra.

La Figura 2 illustra, con una simulazione, lo sviluppo di questo schema, in cui diverse immagini che provocano una risposta visiva in un gruppo di cellule, creano delle DAD attraverso un processo Hebbiano.

La riga in alto mostra il protocollo temporale di due trial consecutivi: il primo stimolo campione (simbolo "S") è seguito da un periodo in assenza di stimoli, e quindi dalla presentazione dello stimolo di confronto ("T"), che conclude il primo trial; dopo un ulteriore periodo in assenza di stimoli, viene presentato lo stimolo campione del secondo trial (secondo simbolo "S"). La sequenza dei trial viene ripetuta più volte per accumulare statistica sulla attività dei neuroni.

Nelle porzioni  $R_1$ ,  $I_1$ ,  $R_2$ ,  $I_2$  dei due riquadri  $A$  e  $B$  si mostra una tipica registrazione (simulata) di due neuroni durante il protocollo sopra descritto. Partendo dall'alto in ogni riquadro, per ogni neurone sono riportati i 'raster'  $R_1$  e  $R_2$  (ogni sequenza orizzontale di tracce bianche rappresenta la successione temporale di impulsi emessi dal neurone in una ripetizione del trial), e in  $I_1$ ,  $I_2$  gli istogrammi, derivati dai raster, della attività media del neurone nelle ripetizioni dei trial con lo stesso stimolo campione. Tali istogrammi,

costruiti a partire dalla attività relativa ad ogni stimolo, vengono per questo chiamati ‘peristimulus histograms’ (‘PSTH’). La parte inferiore di ogni riquadro,  $N$ , mostra una porzione di 8 neuroni della rete cui i due neuroni registrati appartengono, nelle varie fasi dei trial. Ogni neurone è rappresentato da una sferetta, tanto più grande e luminosa, quanto più il neurone è attivo. Lo spessore delle barre che uniscono i due neuroni è proporzionale alla efficacia della sinapsi che li connette.

I due riquadri  $A$  e  $B$  descrivono la tipica evoluzione temporale dell’attività dei due neuroni, prima ( $A$ ) e dopo l’apprendimento ( $B$ ). Prima dell’apprendimento la rete, in assenza di stimoli, è sempre in attività spontanea (i neuroni emettono impulsi a frequenze basse); durante la presentazione dello stimolo, i neuroni della rete che hanno una ‘risposta visiva’ aumentano di molto la loro frequenza di emissione e, in base ad un meccanismo Hebbiano di LTP, le connessioni sinaptiche tra di loro tendono a potenziarsi. Dopo molte ripetizioni del trial la rete raggiunge la situazione descritta nella porzione  $N$  del riquadro  $B$ : le sottopopolazioni attivate da stimoli diversi hanno sinapsi potenziate al loro interno, e questo potenziamento selettivo è in grado di sostenere collettivamente lo schema di attività (l’attrattore) provocato da ogni stimolo, anche dopo la scomparsa dello stimolo stesso: i neuroni eccitati dallo stimolo rimangono attivi dopo la sua rimozione, ed eccitandosi l’un l’altro emettono a frequenze nettamente maggiori di quelle spontanee, a causa della intensa retroazione. Alla presentazione del secondo stimolo campione, nella rete strutturata ( $N$ ), si mostra un esempio in cui l’insieme dei neuroni attivati dallo stimolo non coincide completamente con quello dei neuroni attivi durante la fase di apprendimento; questo corrisponde ad una situazione in cui si presenta alla rete una versione leggermente modificata dello stimolo: il neurone indicato dalla freccia gialla non è attivato dallo stimolo e ha attività di ritardo. Dalla attività che segue la rimozione dello stimolo si vede che la struttura sinaptica permette la ricostruzione dinamica dello schema di attività appreso.

Quanto abbiamo esposto non è certamente un quadro completo dell’apprendimento e della memoria. E’ però un quadro molto ricco, è una spiegazione promettente dei risultati di Miyashita e Goldman-Rakic e, se l’interpretazione di questi esperimenti è corretta, lo schema Hebbiano potrebbe esserne il meccanismo di base. Inoltre è attraente l’ipotesi che ciò sia sufficiente per produrre un classificatore con apprendimento. Il problema diventa quindi la verifica che questo schema funzioni.

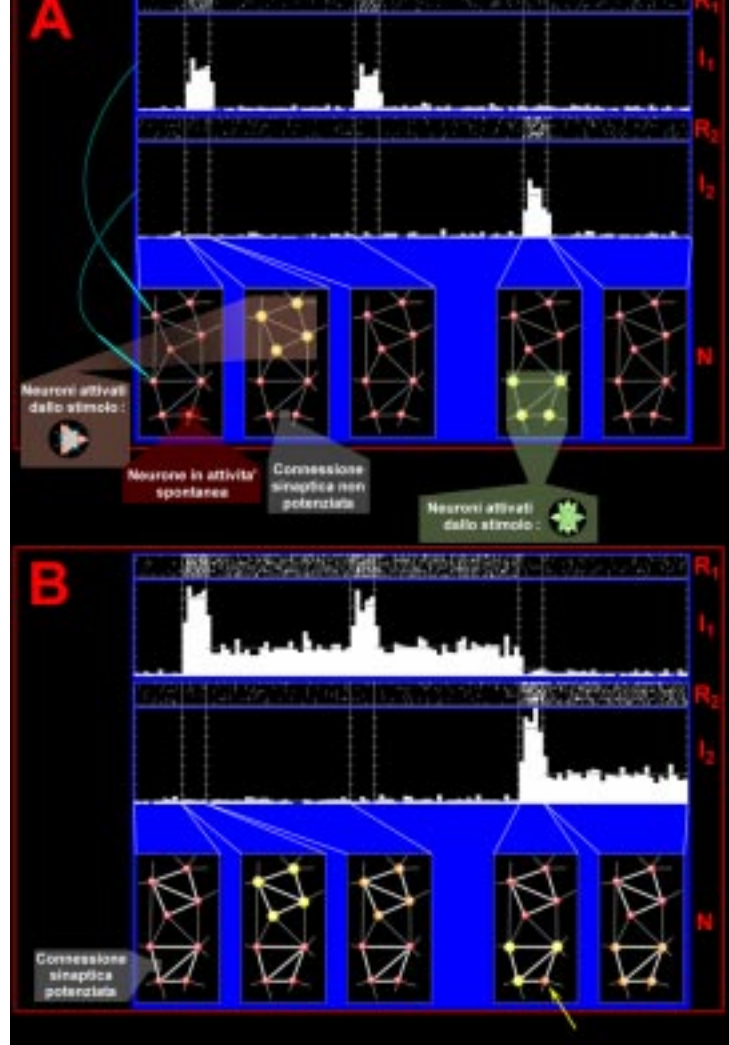


Figure 2: L'evoluzione di una DAD. Riquadro A: prima dell'apprendimento. Riquadro B: dopo l'apprendimento. La riga in alto descrive il protocollo di due trial successivi. Nei due riquadri vengono riportate le registrazioni di due neuroni simulati ( $R_1, R_2, I_1, I_2$ ) e la struttura della rete a diversi tempi durante il 'trial' ( $N$ ). Le registrazioni per ogni neurone sono divise in due parti: i 'raster' ( $R_1, R_2$ ), in cui ogni riga rappresenta la sequenza di impulsi per un 'trial', e i 'PSTH' (*peristimulus histogram*,  $I_1, I_2$ ) della frequenza media ogni 50ms (ordinata dell'istogramma). La struttura della rete viene illustrata nelle righe  $N$  dei due riquadri: ogni rettangolo contiene una rappresentazione di una porzione della rete (8 neuroni); ogni sfera rappresenta un neurone, e il colore e le dimensioni codificano il livello di attivita', mentre le connessioni sono rappresentate dalle linee bianche. Le registrazioni simulate si riferiscono ai neuroni indicati dagli "elettrodi" in alto a sinistra. Inizialmente (A) le connessioni sono casuali e della stessa intensita'. Gli stimoli presentati rafforzano le connessioni tra i neuroni attivati, e nel secondo stadio, (B) la struttura delle connessioni e' tale da sostenere un'attivita' piu' elevata di quella spontanea durante gli intervalli di ritardo tra lo stimolo campione e il confronto.

## Gli aspetti teorici

La prima domanda da porsi è se esiste una matrice sinaptica in grado di sostenere una moltitudine di distribuzioni persistenti di attività selettiva, ognuna con un suo bacino di attrazione che definisce una classe di stimoli. Poi è necessario chiedersi se una semplice dinamica hebbiana possa portare ad una tale struttura sinaptica, e infine occorre analizzare gli aspetti teorici relativi ai vincoli imposti da una realizzazione materiale, sia essa biologica o elettronica.

Alla prima domanda viene data una risposta dal modello di Hopfield [6] (per una panoramica su questa classe di modelli vedi anche [12]). Per comprendere meglio la soluzione proposta, è utile rifarsi alla metafora del “paesaggio”: la dinamica della rete è analoga alla caduta di un grave in un paesaggio fatto di valli e alture. Ogni punto del paesaggio corrisponde ad una configurazione di stati di attività della rete, e la tendenza ad andare verso il basso è una conseguenza della dinamica neuronale. Ogni fondovalle corrisponde ad uno stato di equilibrio (attrattore) dove vanno a finire tutti gli stimoli (condizioni iniziali della rete) che sono nel suo bacino di attrazione, ovvero che appartengono alla stessa classe. Lo stato a fondovalle è rappresentativo dell'intera classe di stimoli e costituisce una rappresentazione interna della classe. La struttura sinaptica determina l'altitudine (energia del sistema) del paesaggio in ogni punto, e dunque contiene l'informazione su dove sono i fondovalle e quanto sono grandi i bacini di attrazione. A sua volta la struttura sinaptica viene costruita come risultato dell'apprendimento e dunque dipende dalla struttura degli stimoli presentati.

Il modello di Hopfield fornisce un esempio di struttura sinaptica che genera un paesaggio con le proprietà appena descritte (vedi anche il capitolo **modelli neuronali?** di questo volume). I neuroni hanno due stati (+1,-1), corrispondenti ad alta e bassa frequenza di emissione. Una rete di  $N$  neuroni di questo tipo può avere  $2^N$  stati differenti, e lo spazio sul quale il paesaggio viene disegnato si estende a tutti questi stati. Ogni stato della rete è una ‘parola’ a  $N$  bit. Denotiamo una tale parola con  $\xi_i^\mu$ , dove  $i$  è l'indice del neurone della rete e  $\mu$  indica lo stimolo preso in considerazione. Per  $\mu$  fissato, questa parola a  $N$  bit corrisponde uno stato di tutti i neuroni nella rete quando viene presentato lo stimolo  $\mu$ . Ad ogni neurone viene assegnato un valore +1 o -1.

Una matrice sinaptica che garantisce che un insieme particolare di stati

neuronali  $\{\xi_i^\mu\}$  sia uno stato di equilibrio (un attrattore) della dinamica neuronale, è data da:

$$J_{ij} = \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \quad (1)$$

dove  $J_{ij}$  è l'efficacia della sinapsi che modula il segnale emesso dal neurone pre-sinaptico  $j$  per depolarizzare il neurone post-sinaptico  $i$ .  $P$  è il numero di stimoli che sono stati impressi nella rete come stati persistenti, e sono i minimi (fondivalle) del paesaggio. L'energia del sistema (l'altitudine nel paesaggio) è data dalla espressione:

$$E = - \sum_{i \neq j} J_{ij} S_i S_j \quad (2)$$

La dinamica della rete viene concepita nel modo seguente: dato uno stato  $\{S_i(t)\}$  di tutti i neuroni al tempo  $t$ , ogni neurone calcola il proprio input sinaptico (la sua depolarizzazione)

$$h_i(t) = \sum_{j=1}^N J_{ij} S_j(t) \quad (3)$$

e assumendo una soglia uguale a 0, il nuovo stato del neurone  $i$  vale:

$$S_i(t + \delta t) = \begin{cases} 1 & \text{se } h_i > 0 \\ -1 & \text{se } h_i < 0 \end{cases} \quad (4)$$

Ad ogni passo di questa dinamica l'energia (2) diminuisce (il punto rappresentativo dello stato della rete nel paesaggio scende verso un fondovalle).

Questo sistema dinamico ha la maggior parte delle caratteristiche richieste. Quando  $P$  non è troppo elevato, le  $P$  distribuzioni di attività  $\{\xi_i^\mu\}$  sono proprio in corrispondenza dei fondivalle del paesaggio di energia.

Il modello di Hopfield non è molto realistico dal punto di vista neurobiologico e sono stati costruiti dei modelli più elaborati per analizzare i fenomeni di memoria attiva e passiva. Tuttavia presenta molte delle caratteristiche dei modelli più realistici. E' pertanto utile per discutere le promesse e le limitazioni dei sistemi di classificazione neuronale.

La prima domanda che ci si pone è se questo modello sia in grado di apprendere a classificare. Supponiamo che i  $P$  stimoli  $\xi_i^\mu$  siano presentati alla

rete uno dopo l'altro e che ogni stimolo imponga ai neuroni lo stato prescritto dalla sua struttura. Supponiamo anche che le  $J_{ij}$  cambino semplicemente aggiungendo all'efficacia sinaptica il prodotto  $\xi_i^\mu \xi_j^\mu$  degli stati imposti dallo stimolo ai due neuroni che sono connessi dalla sinapsi. Questo non è molto diverso da un meccanismo hebbiano. Se non interviene nessuna ulteriore modifica, il sistema risultante è un classificatore che ha imparato a classificare dalla propria esperienza. Ognuno degli stati appresi attrae un grosso numero di stimoli futuri (il suo bacino di attrazione) verso la stessa distribuzione di attività persistente.

In termini di paesaggio possiamo pensare ad una condizione iniziale della rete con sinapsi casuali, che corrisponde in generale ad una successione molto irregolare di alture e di valli (v. figura 3); la posizione del fondo delle valli sarà essenzialmente scorrelata dagli stimoli. Per effetto delle modificazioni sinaptiche il paesaggio si modifica gradualmente, e valli sempre più ampie e profonde si scavano in corrispondenza delle configurazioni che codificano gli stimoli. Alla fine, se la rete è sotto il limite di capacità di memoria, queste valli dominano il paesaggio (pur essendovi la possibilità di valli minori o di piccole valli entro quelle maggiori).

Un tale sistema funziona sorprendentemente bene, anche quando viene esposto ad un flusso 'rumoroso' di stimoli, in cui alcuni appartengono a classi di stimoli simili, mentre gli altri sono scorrelati tra di loro e con le classi. Anche se il numero di stimoli transitori presentati alla rete è molto più grande del numero di quelli appartenenti alle classi, queste manterranno un grosso bacino di attrazione. Il flusso transitorio, anche se ognuno dei suoi membri cambia la matrice sinaptica come gli altri stimoli, produce cambiamenti che tendono a cancellarsi statisticamente e non disturbano il richiamo delle classi. Viene formato un attrattore per ognuno degli stimoli transitori, tuttavia appaiono molti membri di ogni classe, e la rete è in grado di estrarre un 'prototipo' rappresentativo di ognuna di esse, il cui bacino di attrazione sarà più profondo e più esteso.

In effetti qualsiasi rumore presente nella rete tende a destabilizzare le piccole valli che corrispondono agli attrattori minori. Genericamente si può associare un 'rumore' a qualunque causa di stocasticità che si sovrapponga alla dinamica deterministica del neurone (equazione 4). Questo rumore provoca occasionalmente un movimento 'contro la gravità' del punto rappresentativo della rete nel paesaggio, in cui si risale brevemente il pendio invece di scendere verso il fondovalle. Se la sorgente di rumore non è troppo forte, il suo

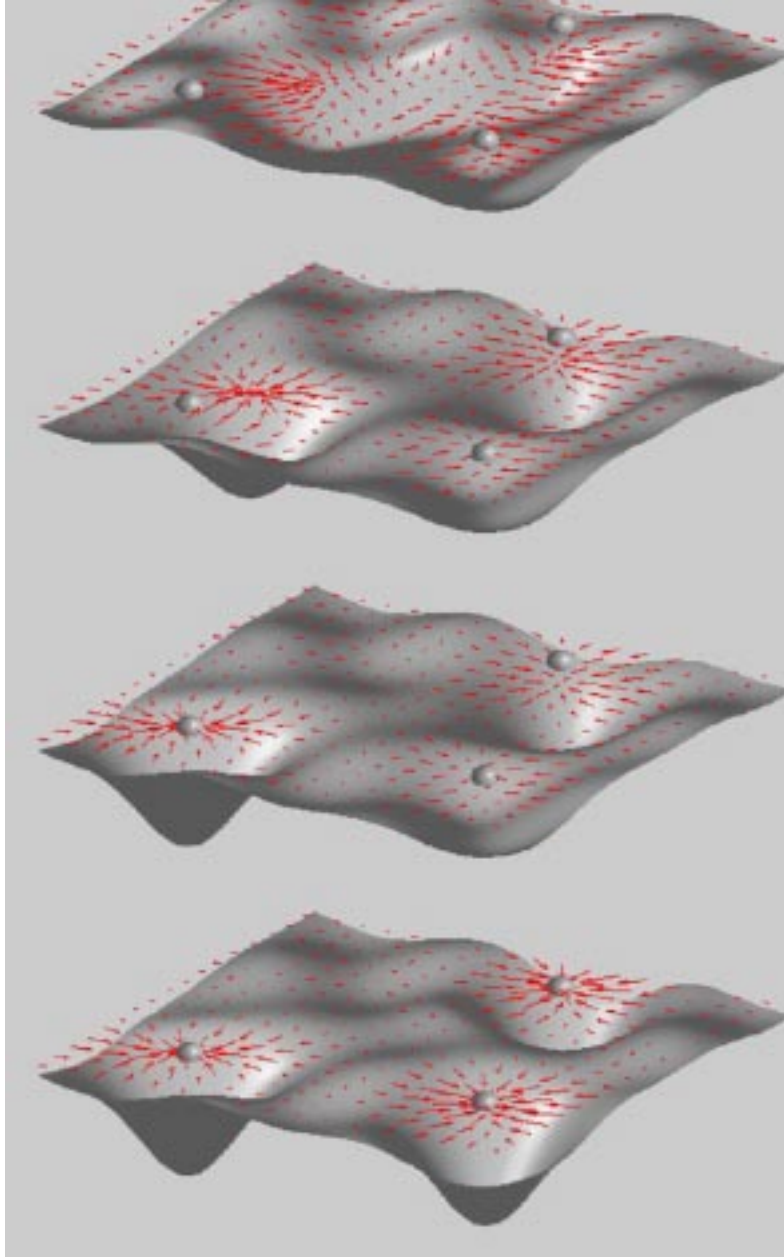


Figure 3: La metafora del paesaggio per il processo d'apprendimento di una rete ad attrattori. Ogni figura corrisponde ad uno stadio diverso di apprendimento: in alto il paesaggio iniziale, e andando verso il basso gli stadi più strutturati. Ogni punto del paesaggio rappresenta uno stato della rete (v. il testo) e le frecce rosse indicano la direzione di evoluzione della rete in ogni punto. La direzione e la lunghezza delle frecce indicano la via più breve verso il fondovalle. Le sfere indicano la posizione degli stati corrispondenti agli stimoli ripetutamente presentati. All'inizio (in alto) la rete non ha ancora alcuna esperienza degli stimoli e le valli sono poco profonde e in posizione casuale rispetto alle sfere. Negli stadi successivi gli stimoli scavano sotto le sfere delle valli che diventano sempre più ampie e profonde e che negli stadi finali dominano il paesaggio.

effetto non cambia la capacità da parte delle valli ampie e profonde di attrarre al proprio fondo gli stati della rete che si trovano sul pendio. Può essere però sufficiente ad offrire una via di fuga agli stati altrimenti intrappolati in piccole valli create da stimoli occasionali.

## **Imparare a classificare in un dispositivo reale**

La difficoltà maggiore che si incontra nel realizzare (per esempio in elettronica) un sistema come quello appena descritto sta nella *profondità analogica* richiesta per le sinapsi: la regola per cambiare l'efficacia sinaptica prevede che il dispositivo che realizza la sinapsi sia in grado di memorizzare un numero di valori differenti almeno pari al massimo numero di classi che devono essere imparate e questo numero può essere molto elevato. Il dispositivo sinaptico, che in biologia è così piccolo rispetto al neurone, deve essere in grado di preservare il valore dell'efficacia sinaptica in modo stabile per scale di tempo dell'ordine di giorni, se non di anni. Tuttavia, allo stesso tempo, deve essere in grado di adattarsi e di imparare in tempi molto brevi, dell'ordine di secondi.

Questi sono vincoli molto stringenti e un dispositivo che soddisfi tutte queste richieste non è disponibile allo stato attuale. E' piuttosto improbabile che la soluzione adottata dalla biologia preveda un dispositivo così difficile da realizzare.

Il nostro gruppo, a Roma, si è proposto di realizzare una serie di dispositivi sinaptici di capacità più limitate rispetto a quelle elencate nel paragrafo precedente. Tuttavia sono dispositivi facilmente realizzabili e, ciò che è più importante, non sacrificano nessuna delle capacità di elaborazione relative all'apprendimento e all'espressione di DAD come memoria attiva.

E' interessante notare come la costruzione della 'memoria' di un calcolatore consista nell'organizzare semplici elementi a due stati, 'bit', ognuno dei quali possa mantenere il suo stato per tempi lunghi (a spese dell'energia fornita dall'esterno). L'immagazzinamento di grandezze con buona profondità analogica viene recuperato organizzando la memoria in 'parole' di  $B$  bit, in grado di assumere (e mantenere)  $2^B$  diverse configurazioni. Occorre poi una 'intelligenza' (il processore) che assegni un valore ad ognuno dei bit in base alla posizione. A patto di definire e gestire una opportuna organizzazione di elementi di memoria binari stabili, si può costruire una memoria 'universale'. Una logica a più valori ridurrebbe il numero di elementi necessari a parità

di informazione, ma complicherebbe molto la struttura dei singoli elementi stabili.

Queste osservazioni generali suggeriscono anche per il nostro contesto una scelta di semplicità per l'elemento base della memoria (la sinapsi).

Dunque la riduzione drastica della profondità analogica della sinapsi potrebbe risolvere il problema del mantenimento a lungo termine, ma solleva altre due questioni. La prima riguarda la funzionalità: è possibile che una rete di sinapsi binarie sia in grado di sostenere la varietà di DAD che sono espressione del processo di classificazione? La seconda riguarda l'apprendimento: esiste una dinamica dell'efficacia sinaptica che sia realizzabile e che porti ad una matrice sinaptica con le caratteristiche desiderate?

La prima questione ha avuto una interessante risposta quando Sompolinsky [7, 8] chiarì che una rete di Hopfield perde poco dal punto di vista delle capacità di elaborazione, anche se si riducono a due i valori di ogni sinapsi (si mantiene solo il segno della sommatoria dell'equazione 1). Ma se si ha una risposta positiva alla questione della funzionalità, dall'altro sorgono nuove difficoltà legate alla questione dell'apprendimento. Nel contesto che abbiamo descritto la matrice alla Hopfield deve prima formarsi e, solo alla fine, ridursi ad uno dei due valori.

Da un lato, non è chiaro come l'apprendimento possa proseguire, dopo questa riduzione, al presentarsi di nuovi stimoli. Dall'altro se la riduzione esprime il fatto che il dispositivo sinaptico può mantenere una buona profondità analogica solo per piccoli intervalli di tempo  $\Delta t$ , e possiede un piccolo numero di stati stabili, si pone il problema di scegliere a quale stadio della formazione della matrice sinaptica effettuare la riduzione. Questa riduzione deve essere effettuata entro un tempo minore di  $\Delta t$ , altrimenti nuovi stimoli possono ribaltare il segno finale portando ad una matrice che non ha più le caratteristiche desiderate. Se la riduzione interviene su questa scala, bisogna assicurarsi che tutti gli stimoli da memorizzare vengano presentati alla rete in una finestra temporale relativamente breve.

Una alternativa semplice è di assumere che la sinapsi sia un dispositivo analogico solo su brevi scale di tempo, e che su lunghe scale di tempo intervenga un meccanismo di ripristino che preservi solo uno dei valori di un insieme discreto. Tutti gli altri valori sono instabili e, in assenza di stimoli, vengono attratti verso uno dei valori di questo insieme, che vengono mantenuti indefinitamente. Tra un valore stabile e l'altro vi è una soglia che discrimina i valori analogici che vengono attratti verso un valore e quelli che

vengono attratti verso l'altro.

Quando viene presentato uno stimolo viene attivata una sorgente che dipende dall'attività dei due neuroni connessi dalla sinapsi e che tende a portare l'efficacia sinaptica verso uno degli altri valori stabili. Se al momento della rimozione dello stimolo, l'efficacia ha attraversato una delle soglie, la sinapsi verrà attratta verso uno dei valori stabili e sarà avvenuta una transizione. Altrimenti il valore dell'efficacia verrà nuovamente attratto verso il valore stabile iniziale e lo stimolo non avrà indotto alcun cambiamento.

Un tale schema di dinamica di apprendimento risulta il più naturale per la realizzazione di un dispositivo sinaptico reale, sia esso elettronico o biologico. Prima di intraprendere la descrizione dell'hardware realizzato sulla base di questo schema, dobbiamo premettere alcune considerazioni sulle implicazioni a livello funzionale.

## **L'apprendimento stocastico**

### **I vincoli di un apprendimento realistico**

La dinamica di apprendimento appena descritta può essere schematizzata come una serie di transizioni da uno stato stabile all'altro: durante la presentazione di ogni stimolo la sinapsi viene spinta verso l'alto (potenziamento) o verso il basso (depressione) e alla fine può accadere che abbia lasciato lo stato precedente per ritrovarsi in un altro stato stabile. Questo nuovo stato stabile viene preservato fino alla presentazione di un nuovo stimolo.

In questo scenario la situazione è assai diversa da quella descritta nel paragrafo precedente: l'informazione sugli stimoli presentati nel passato si dissolve con estrema rapidità, in quanto ogni nuovo cambiamento della matrice sinaptica tende a cancellare la traccia del passaggio degli stimoli precedenti. Per capire come ciò possa accadere consideriamo il caso estremo di una rete con sinapsi binarie. Nel momento in cui si presenta un nuovo stimolo la sinapsi cerca di assecondare le "spinte" della sorgente Hebbiana in modo da acquisire l'informazione sulla struttura dello stimolo presentato. Se l'apprendimento è deterministico, per ognuna delle sinapsi lo stato successivo alla presentazione dello stimolo sarà scelto univocamente tra i due soli possibili, sulla base della sorgente Hebbiana da essa vista, indipendentemente dal suo stato di partenza e, quindi, indipendentemente dalla sua storia passata. In questo senso, la struttura sinaptica perde memoria degli stimoli passati ad

ogni presentazione di un nuovo stimolo (la sola ‘memoria’ residua dipende, in questo caso, dalle accidentali somiglianze tra gli stimoli che si succedono, che possono rendere implicitamente richiamabili gli stimoli tra di loro sufficientemente simili, indipendentemente dalla loro posizione nella sequenza temporale). La fenomenologia descritta, giustificata in termini quantitativi in [13, 15], non muta qualitativamente aumentando il numero (finito) di stati stabili delle sinapsi.

### La scappatoia dell’apprendimento stocastico

L’esempio presentato ci suggerisce anche una possibile scappatoia: non tutte le sinapsi devono essere modificate per acquisire abbastanza informazione da poter richiamare in futuro lo stimolo presentato. In effetti modificando solo una parte delle sinapsi, si può avere in memoria almeno la traccia di un numero più elevato di stimoli. Questo non vuol dire che questa traccia sia sufficiente a richiamarli.

Una dinamica plausibile dell’efficacia sinaptica, che non usa esplicitamente informazioni globali sul flusso dei stimoli, prevede i passi seguenti:

1) per ogni sinapsi viene deciso se essa debba essere potenziata o depressa, sulla base dell’attività dei due neuroni che essa connette.

2) la transizione permessa viene effettuata con probabilità  $q$ . A parità di condizioni sull’attività dei neuroni pre- e post-sinaptico, talvolta si hanno le transizioni previste e talvolta l’efficacia sinaptica rimane inalterata.

3) Se lo stato della sinapsi è già potenziato e fosse stato deciso di effettuare un ulteriore potenziamento, allora la sinapsi rimane inalterata perchè non si può andare oltre al livello massimo. Analogamente se si parte da uno stato depresso ed è stata decisa una ulteriore depressione.

Con questo tipo di apprendimento stocastico le sinapsi che cambiano sono estratte a caso ogni volta che un nuovo stimolo viene presentato. In seguito alla presentazione di uno stimolo generico si avranno in media  $qN^2$  sinapsi scelte per effettuare una transizione. Per queste sinapsi, tutto il passato viene dimenticato per adattarsi al meglio al nuovo stimolo. Quando viene presentato lo stimolo successivo, una frazione  $q$  di quelle che ricordano lo stimolo precedente saranno estratte per effettuare una transizione e distruggeranno parzialmente l’informazione del pattern precedente. Solo le restanti  $(1 - q)qN^2$  preserveranno la memoria del primo stimolo.

Se  $P$  stimoli vengono mostrati alla rete uno dopo l’altro una sola volta,

dopo la presentazione di tutti gli stimoli rimarranno solo  $q(1 - q)^{P-1}N^2$  sinapsi con memoria del primo. Una condizione necessaria perchè questo stimolo sia richiamabile, è che almeno una sinapsi lo “ricordi”:

$$q(1 - q)^{P-1}N^2 > 1 \Rightarrow P < \frac{\log(qN^2)}{\log[(1 - q)^{-1}]} + 1$$

Come si può vedere dall’ultima disuguaglianza, il numero di memorie possibili è tanto più grande quanto  $q$  è piccola. Invece, se  $q$  rimane fissa con l’aumentare di  $N$ , il numero di stimoli che la matrice sinaptica è in grado di tenere in memoria è minore di  $C \log(N)$  ( $C$  indipendente da  $N$ ). Però la  $q$  non può diventare troppo piccola, poichè deve cambiare in media almeno una sinapsi per stimolo ( $qN^2 > 1$ , ovvero  $q = k/N^2$  con  $k > 1$ ), da cui si trova che diminuendo  $q$  al massimo, si può arrivare, per  $N$  grande, a:

$$P < \frac{\log(k)}{\log[(1 - k/N^2)^{-1}]} + 1 \simeq N^2,$$

pur mantenendo una traccia del passaggio di tutti gli ultimi  $P$  stimoli, il che è una condizione necessaria perchè siano richiamabili.

Se gli stimoli presentati sono casuali ed è bassa la frazione media di neuroni che risultano attivi per ogni stimolo (il “livello di codifica”) allora solo un sottoinsieme di sinapsi tende ad essere modificato. In tal caso la struttura stessa degli stimoli fornisce un ulteriore elemento stocastico per ripartire meglio le risorse tra le diverse memorie. Infatti ogni nuovo stimolo presentato, essendo basso il livello di codifica, sarà molto diverso da quelli precedenti e dunque tenderà a modificare sinapsi che non venivano utilizzate dalle memorie preesistenti. Ci sono in effetti due estrazioni casuali dietro la modifica di ogni sinapsi: la prima è decisa dalla struttura dello stimolo casuale, che estrae a caso un sottoinsieme di sinapsi candidate a essere cambiate; la seconda è il meccanismo stocastico intrinseco ad ogni sinapsi, che effettua una ulteriore sottoselezione delle sinapsi che si adatteranno effettivamente allo stimolo.

Se le probabilità di transizione ‘efficaci’, che tengono conto anche della struttura degli stimoli, sono tali da avere un numero medio di potenziamenti uguale al numero medio di depressioni, allora le  $N^2$  memorie sono non solo presenti nella struttura sinaptica, ma di fatto anche richiamabili. Questo riproduce il risultato classico di Willshaw sulla capacità ottimale [9].

In una sequenza di singole presentazioni, la memoria corrispondente all'ultimo stimolo presentato è il più facilmente richiamabile. Andando indietro nel passato la traccia di memoria degli stimoli via via più vecchi si indebolisce finché, oltre il  $P$ -esimo stimolo nella sequenza, i pattern non sono più richiamabili. Se invece la  $q$  è così piccola da richiedere molte presentazioni dello stesso stimolo prima che sia richiamabile, allora l'ultimo stimolo non è più privilegiato e le  $P$  memorie sono tutte sullo stesso piano [15]. Ogni stimolo presentato modifica così poco la struttura sinaptica che difficilmente perturba le sinapsi che si sono adattate agli altri stimoli. La frazione di sinapsi che viene modificata per acquisire informazione sul nuovo stimolo è molto bassa, e per questo sono richieste molte presentazioni. Tuttavia è molto bassa anche la frazione di sinapsi che perde la memoria del passato; alla fine la struttura sinaptica non dipende più dall'ordine temporale in cui queste modifiche vengono effettuate, e la statistica delle correlazioni tra stimolo e struttura sinaptica rimane invariata tra due ripetizioni successive dello stesso stimolo. Si recupera in questo modo la situazione ideale, in cui le risorse di memoria (le efficacie sinaptiche) vengono ripartite uniformemente tra tutti gli stimoli che devono essere richiamati.

## **Modelli, simulazioni e reti neuronali elettroniche**

Molta della conoscenza e dell'esperienza acquisita sul comportamento dettagliato di questo tipo di reti deriva dalla loro simulazione al computer, in cui le equazioni che descrivono la dinamica del sistema (le equazioni di evoluzione per i neuroni e per le sinapsi) vengono approssimate introducendo dei passi temporali discreti, ad ognuno dei quali lo stato del sistema viene calcolato e aggiornato. Queste simulazioni costituiscono dei veri e propri "esperimenti numerici", in cui si è in grado di riprodurre, fatte salve le molte semplificazioni implicite nel modello, molte delle metodologie degli esperimenti reali: "registrazioni" dei singoli neuroni, "registrazioni" multiple, uso di diversi "protocolli" controllati di presentazione degli stimoli ecc.

Questo approccio basato sulle simulazioni, che conserverà comunque nel futuro un ruolo essenziale, soffre di due limitazioni. La prima, quantitativa, è legata all'onere computazionale richiesto per simulazioni di reti complesse, in particolare quando si vuole far evolvere nel contempo la dinamica neuronale e quella sinaptica, caratterizzate da scale di tempo molto diverse. Sebbene siano in corso di sviluppo dei metodi di simulazione che alleviano il prob-

lema rispetto a quelli convenzionali, è prevedibile che sistemi neuronali estesi e multi-modulari in grado di realizzare funzioni complesse rimarranno non trattabili numericamente su calcolatori convenzionali.

La seconda limitazione è di carattere metodologico. Una simulazione non può fornire più informazione di quella contenuta nelle equazioni che definiscono il modello. Il tipo di astrazione necessaria per la definizione matematica del modello può d'altra parte rivelarsi incompatibile con la sua effettiva realizzabilità, sia essa biologica, elettronica o altro. Con riferimento a quanto esposto nei capitoli precedenti, una richiesta di profondità analogica infinita per una variabile dinamica (limitata nelle simulazioni solo dalla rappresentazione binaria scelta), o specifiche caratteristiche imposte ad un processo stocastico, sono esempi di possibili sorgenti di incompatibilità tra la caratterizzazione matematica di un modello e la sua realizzabilità in un dispositivo reale. E' chiaro che in generale sarà possibile *a posteriori* modificare le simulazioni, e tener conto dei vincoli emersi; manca spesso, però, una guida per l'identificazione *a priori* di tali vincoli.

### **Motivazioni per una rete neuronale elettronica**

Le due limitazioni di cui si è detto rendono desiderabile un approccio complementare alla simulazione numerica di modelli neuronali, basato sulla loro realizzazione elettronica. Da un lato, questo consente in prospettiva la realizzazione di sistemi neuronali complessi che interagiscano con l'ambiente in tempo reale. Dall'altro, svolge una preziosa funzione euristica nella individuazione dei vincoli cui i modelli devono soddisfare.

La considerazione contemporanea di questi due aspetti conduce ad un genere di "hardware neuronale" relativamente nuovo, in rapido sviluppo attualmente, in cui gli elementi costitutivi della rete sono direttamente ispirati alla loro controparte biologica [10]. Questo indirizzo si differenzia in modo marcato da quello, affermato in vari settori applicativi, in cui si realizzano dei "coprocessori neuronali" da affiancarsi ai computer convenzionali, in grado di accelerare significativamente l'esecuzione di una simulazione. La novità dell'approccio ha motivato un neologismo che lo identifica come "hardware neuromorfo".

Per quanto riguarda la dinamica neuronale, almeno al livello di descrizione intermedio tra il macroscopico e il microscopico al quale i modelli in questione si collocano, sono in gran parte le caratteristiche elettriche del neurone

che giocano il ruolo fondamentale (le conducibilità della membrana neuronale rispetto alle diverse correnti ioniche, la capacità della membrana ecc. [17]); va ricordato che, a partire dalle equazioni fenomenologiche di Hodgkin e Huxley (che descrivono in modo dettagliato l'emissione del potenziale di azione), fino al modello semplificato di neurone “integrate and fire” cui si accennerà in seguito, molte caratteristiche funzionali del neurone (e in particolare quelle che si ritengono essenziali ai fini del comportamento collettivo di molti neuroni interagenti) sono state descritte con successo in termini di semplici circuiti elettrici equivalenti. In ultima analisi, comunque, sarà la capacità del modello neuronale di produrre comportamenti collettivi in accordo con i dati sperimentali (al livello di descrizione scelto) a fornirne la legittimazione.

La corrispondenza tra i modelli di dinamica sinaptica proposti e i dati sperimentali noti è più difficile da stabilire, e molto lavoro rimane da fare al riguardo (sia dal punto di vista sperimentale che teorico). In questo caso, l'interazione tra le fasi di modellizzazione teorica, di simulazione numerica e di realizzazione elettronica risulta particolarmente feconda.

Il tipo di apprendimento descritto nel capitolo precedente emerge come versione minimale di regola di modificazione sinaptica in grado di 1) soddisfare genericamente un vincolo di realizzabilità legato ad un numero limitato e prefissato di stati stabili discreti per la sinapsi 2) garantire la località spaziale del meccanismo di apprendimento (la modificazione di una sinapsi dipende solo dall'attività dei due neuroni da essa connessi) 3) garantirne la località temporale (la modificazione sinaptica viene indotta dallo stimolo in modo indipendente dagli stimoli precedenti) 4) consentire una alta capacità di memoria, in funzione del numero di neuroni della rete, adottando un meccanismo stocastico di modificazione sinaptica.

## La LANN27

Nel seguito descriveremo le prime due fasi di un progetto che si propone di realizzare un sistema neuromorfo integrato su chip, ispirato ai principi di modellizzazione sopra esposti. La prima realizzazione, la LANN27 [19, 20], è una rete asincrona ad attrattori, in elementi analogici discreti, composta da 27 neuroni e 351 sinapsi simmetriche (è quindi completamente connessa), con apprendimento stocastico (v. figura 4).

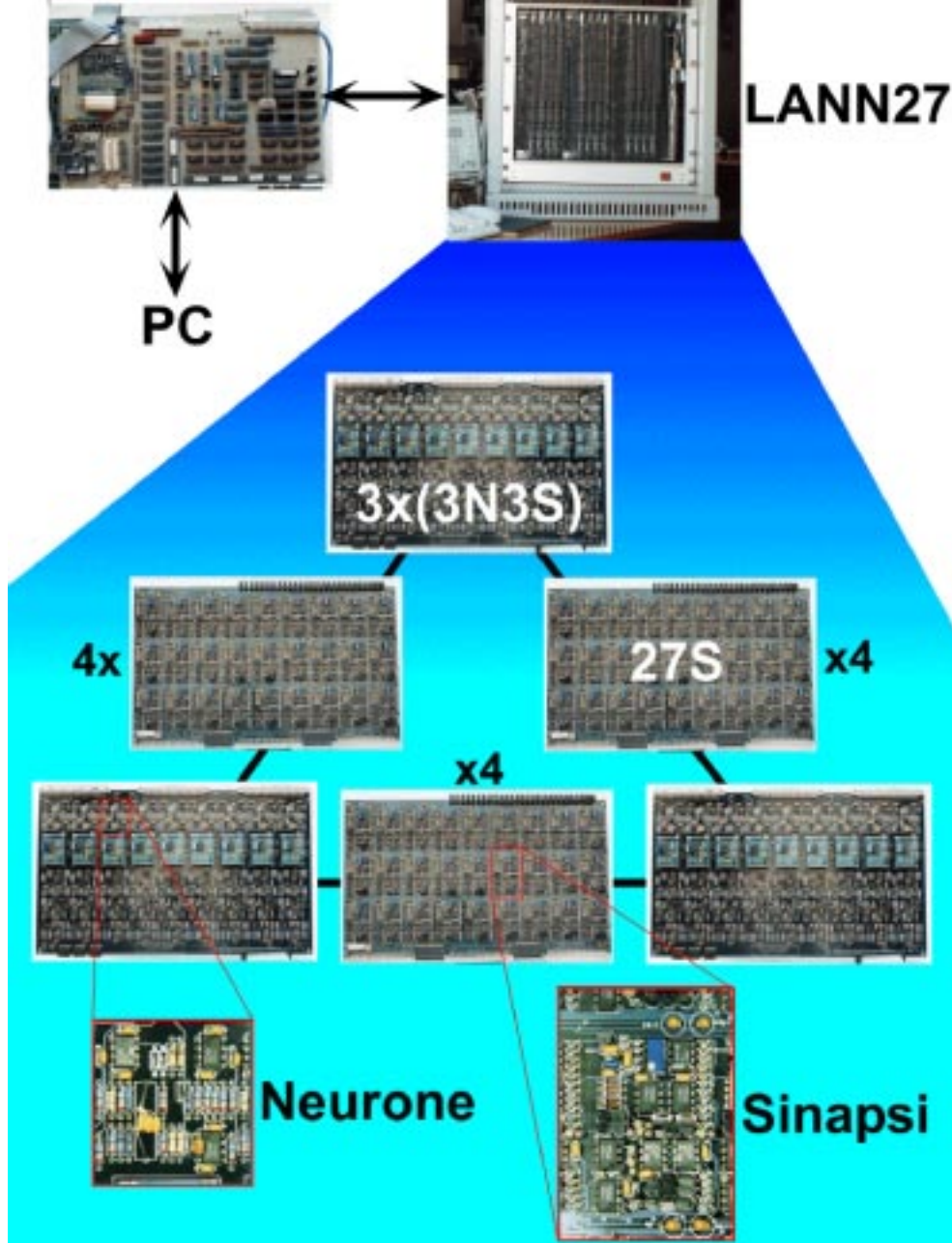


Figure 4: Schema della LANN27: la macchina è stata realizzata in elettronica analogica a componenti discreti. In alto, al centro, viene rappresentata la macchina completa. L'architettura è illustrata nella parte inferiore dove si possono vedere le schede che contengono 3 sottoreti di 3 neuroni completamente connessi tra loro (ai vertici del triangolo) e quelle che ospitano le sinapsi per le connessioni tra le diverse sottoreti. La macchina comunica con l'esterno tramite la scheda in alto a sinistra che consente la connessione con un PC. In basso vengono ingranditi gli elementi principali: il neurone e la sinapsi.

## Dinamica dei neuroni

Data una configurazione ad un certo istante delle sinapsi  $J_{ij}$  ( $j$  indica il neurone pre-sinaptico,  $i$  il post-sinaptico;  $J_{ij} = J_{ji}$ , sinapsi simmetriche), l'ingresso al neurone generico  $k$  è la media, su un breve intervallo di tempo, della somma delle attività  $S_j$  degli altri neuroni, pesata dalle efficacie sinaptiche  $J_{kj}$ . Se nell'intervallo di tempo considerato la rete è sottoposta ad uno stimolo esterno, questo si traduce in un termine aggiuntivo (e predominante) nell'input al neurone. Il neurone effettua una trasformazione non lineare del suo input totale, determinando il suo stato di attivazione  $S_k(t)$ . Questa trasformazione è realizzata attraverso amplificatori operazionali, e se il loro guadagno è molto alto (come nelle situazioni che descriveremo) lo stato di attività del neurone può considerarsi una variabile binaria (+1, -1).

Le sinapsi hanno una loro dinamica temporale, molto più lenta della dinamica neuronale, il che permette di considerare costanti le efficacie sinaptiche durante il tempo in cui viene mediato l'input al neurone e viene determinato il suo stato di attività.

La dinamica collettiva dei neuroni, per una data configurazione sinaptica, si può descrivere qualitativamente così: se è presente uno stimolo esterno, questo domina l'attività della rete per tutto il tempo della sua persistenza ("tempo di presentazione") che, come vedremo, è anche il tempo 'utile' per l'apprendimento. Alla scomparsa dello stimolo, la attività ricorrente dei neuroni è determinata dalla struttura sinaptica, che ne fissa gli stati di equilibrio (attrattori), la cui esistenza è garantita dalla simmetria delle sinapsi.

## Dinamica delle sinapsi

La dinamica (stocastica) delle sinapsi realizza il meccanismo di apprendimento, che cambia gradualmente nel tempo la struttura dei punti di equilibrio della dinamica neuronale (le 'memorie' della rete). Il meccanismo, di tipo Hebbiano, è illustrato nella figura 5; esso rappresenta una particolare realizzazione dell'idea già illustrata del meccanismo di transizione tra gli stati stabili di una sinapsi. Le curve in figura che fluttuano in modo irregolare intorno ai valori  $(-J_0, J_0)$  sono due soglie fluttuanti che realizzano il meccanismo stocastico di transizione; i valori  $(-J_c, 0, J_c)$  sono i tre stati stabili che la sinapsi può assumere su lunghe scale di tempo. In questo caso, lo stimolo (presentato alla rete al tempo 0) induce lo stesso stato di attività nei neuroni

pre- e post-sinaptico, e questa covarianza tende a potenziare la sinapsi (la curva continua è l'efficacia sinaptica istantanea), che parte dallo stato stabile 0.

L'efficacia sinaptica, sotto l'effetto dello stimolo, tende verso un valore asintotico che dipende dall'intensità dello stimolo. Nel caso in figura, durante la presentazione del primo stimolo, la soglia fluttuante rimane sempre al di sopra del valore della efficacia sinaptica. In queste condizioni, non si hanno transizioni tra gli stati stabili della sinapsi. Quando lo stimolo viene rimosso, essa ritorna allo stato stabile di partenza.

Nella stessa figura si osserva la dinamica seguente alla presentazione di un secondo stimolo tale, di nuovo, da potenziare la sinapsi. In questo caso, a causa delle fluttuazioni, la soglia per caso scende al di sotto del valore della efficacia sinaptica; questo fatto provoca un istantaneo contributo di potenziamento ulteriore della sinapsi, tale da portarne l'efficacia al di sopra del valore stabile  $J_c$ , al quale l'efficacia rilassa appena lo stimolo viene rimosso. In questo modo, il secondo stimolo ha indotto la transizione ( $0 \rightarrow J_c$ ) della efficacia sinaptica. Il valore  $J_c$  così raggiunto si manterrà stabile in assenza di stimoli successivi.

In questa dinamica l'efficacia sinaptica assume valori continui determinati dallo stimolo, su scale di tempo brevi, mentre assume valori discreti nell'insieme  $(-J_c, 0, J_c)$  su scale lunghe di tempo.

Nella realizzazione elettronica la presenza di uno stimolo (e la conseguente attività incrementata dei neuroni) viene espressa attraverso un segnale di input ai neuroni, aggiuntivo a quello determinato dalla attività ricorrente, che entra anche nella determinazione del contributo hebbiano alla dinamica sinaptica. L'intensità di questo contributo influisce sulla probabilità di transizione sinaptica (determina infatti il valore asintotico cui tende il valore analogico della sinapsi – si veda la figura). Nella dinamica sinaptica, inoltre, un meccanismo di “adattamento” fa sì che l'effetto di uno stimolo che persista per tempi lunghi si attenui gradualmente fino ad annullarsi, in modo che tempi di presentazione molto lunghi non determinino una cancellazione della memoria legata agli stimoli precedenti.

Quando lo stimolo scompare, persiste l'attività riverberante dei neuroni nell'attrattore in cui la rete ha rilassato. Per ogni sinapsi, quindi, permane una attività pre- e post-sinaptica potenzialmente in grado di indurre transizioni, sia pure con probabilità molto bassa, se la rete permane a lungo nell'attrattore. Anche in questo caso, un meccanismo di adattamento limita

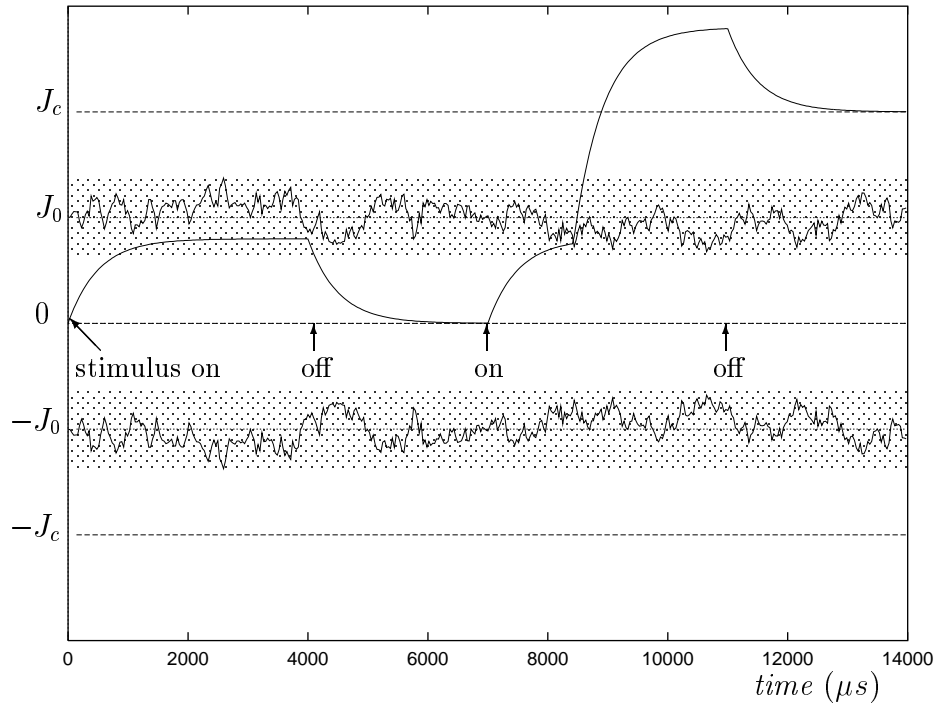


Figure 5: Il meccanismo stocastico di modificazione sinaptica: l'efficacia sinaptica  $J$  in funzione del tempo durante due presentazioni dello stesso stimolo. Nel primo caso la soglia fluttuante non scende mai al di sotto della  $J$  e dopo la rimozione dello stimolo l'efficacia sinaptica ritorna allo stato iniziale ( $J = 0$ ). Durante la seconda presentazione la soglia finisce per caso sotto il valore della  $J$  e la corrente che porta l'efficacia allo stato alto  $+J_c$  viene attivata. Alla rimozione dello stimolo il nuovo stato stabile è  $J_c$ : è avvenuta una transizione.

l'effetto ad una finestra temporale.

Il carattere stocastico delle transizioni indotte dagli stimoli tra gli stati stabili è determinato dalla natura casuale delle fluttuazioni delle soglie intorno ai valori  $-J_0$  e  $J_0$ . Come si è visto, un apprendimento stocastico efficiente richiede in generale probabilità di transizione piccole per le sinapsi. Si pone quindi il problema del controllo di queste probabilità nel caso in esame. In principio si può agire su diversi parametri che entrano in gioco nella dinamica, come per esempio l'intensità degli stimoli. Risulta però che per la maggior parte di essi le regolazioni richieste, nella regione di probabilità molto piccole, sono molto fini, e questo rende il comportamento della rete instabile e poco controllabile (è uno degli esempi di come delle 'specifiche di progetto' facilmente formulabili in linea di principio incontrino dei vincoli di implementazione non banali). Una grandezza che si è rivelata idonea al fine di controllare le probabilità di transizione è il contenuto in frequenza delle fluttuazioni delle soglie; in termini qualitativi, è intuitivo che la lunghezza di correlazione di queste fluttuazioni (la 'memoria' del processo stocastico associato) è legata alla possibilità di ottenere eventi (transizioni) rari. Si è quindi scelto, a livello elettronico, di operare un filtraggio in frequenza di queste fluttuazioni. La frequenza di taglio (insieme alla pendenza) del filtro fornisce un buon controllo delle probabilità di transizione sinaptiche.

### **Dinamica collettiva della LANN27**

Questo sistema, pur così limitato nel numero di unità, esibisce un comportamento complesso a causa della retroazione intensa tra queste unità. Non si può disporre nè di una previsione deterministica della sua dinamica, per la natura intrinsecamente stocastica, nè di una previsione teorica accurata, poichè il numero di elementi non è abbastanza elevato). D'altra parte, proprio la natura cooperativa del suo comportamento lo rende molto 'robusto', per esempio rispetto alle inevitabili disomogeneità elettroniche tra i componenti.

Per illustrare la strategia seguita per estrarre e comprendere il comportamento della rete, riassumiamo innanzitutto le aspettative: la rete deve riflettere, nella sua struttura sinaptica, la statistica spaziale e temporale del flusso di stimoli cui è sottoposta. Dovrà in particolare:

- sviluppare *in modo non supervisionato* delle rappresentazioni interne robuste per gli stimoli apparsi più di frequente (le basse probabilità

di transizione per le sinapsi implicano che molte presentazioni di uno stimolo siano necessarie affinché sia appreso);

- poter richiamare gli stimoli appresi in modo associativo (definire cioè dinamicamente delle classi di stimoli rappresentati dallo stesso attrattore);
- poter ‘dimenticare’ le rappresentazioni di stimoli non più presenti per tempi lunghi, a favore di stimoli nuovi che entrino nel flusso di input.

### **Protocolli per la generazione di flussi di stimoli**

Sebbene la rete sia in grado di adattarsi ad un flusso libero di stimoli esterni è utile, per mettere in luce aspetti specifici del suo comportamento dinamico, adottare degli opportuni “protocolli” di presentazione degli stimoli.

Descriveremo sommariamente la dinamica della rete per quattro protocolli di presentazione degli stimoli:

1. *protocollo incrementale*: si generano a caso 3 prototipi (stringhe di 27 elementi  $\pm 1$ ); in una prima fase si presenta ripetutamente il primo prototipo alla rete, con tempi abbastanza lunghi da indurre modificazioni sinaptiche. Si presentano quindi ripetutamente alla rete i primi due prototipi, in ordine casuale, e infine i tre prototipi, sempre in ordine casuale.
2. *protocollo di palinsesto*: si generano 3 prototipi come nel caso precedente. In una prima fase si presentano, in ordine casuale, i primi due prototipi alla rete. Si smette quindi di presentare il primo prototipo, e si presentano ripetutamente in ordine casuale il secondo e il terzo prototipo.
3. *protocollo di generalizzazione*: per ogni prototipo, si genera un insieme di configurazioni (stimoli) ottenute invertendo a caso lo stato di un numero assegnato di neuroni. Si presentano quindi alla rete in ordine casuale queste versioni “degradata” di tutti i prototipi.
4. *protocollo rumoroso*: vengono generati uno o due prototipi, e degli stimoli casuali vengono interposti tra i prototipi nella sequenza di presentazione.

Il protocollo *incrementale*, in cui “l’ambiente” esterno alla rete si arricchisce gradualmente, mette in luce il processo graduale di formazione degli attrattori (rappresentazioni interne dei prototipi) in relazione alla diversità di stimoli esterni, e gli effetti che si producono quando la memoria della rete giunge al suo limite di capacità.

Il protocollo *palinsesto* misura la capacità della rete di “dimenticare”: ci si attende che le rappresentazioni interne di stimoli non più presenti nell’ambiente scompaiano gradualmente a causa e a favore di quelle corrispondenti a stimoli nuovi.

Con il protocollo *di generalizzazione* si intende studiare la fondamentale proprietà della rete di estrarre, da una classe di stimoli simili, una rappresentazione prototipale della classe.

Questi protocolli corrispondono quindi a condizioni di apprendimento in ambienti ‘controllati’, in cui si presentano alla rete solo stimoli appartenenti alle classi da memorizzare (i prototipi stessi, o versioni poco perturbate di essi). Ricordando la nostra discussione generale sull’apprendimento, però, le capacità di classificazione della rete dovrebbero mantenersi anche nel caso in cui nel flusso di stimoli appaiano stimoli casuali, scorrelati dalle classi da memorizzare. Ci si attende che la struttura asintotica del paesaggio nello spazio degli stati mentenga valli ampie e profonde in corrispondenza alle classi, anche se piccole valli indotte dagli stimoli casuali ne corrugano il profilo. Allo scopo di controllare queste aspettative, ai tre protocolli sopra descritti si aggiunge quindi un *protocollo rumoroso*, in cui stimoli casuali si inseriscono nel flusso di presentazioni ripetute delle classi. In questo caso, i prototipi vengono imparati perchè presentati molte volte, mentre ogni stimolo casuale produce del rumore che tende a far dimenticare i prototipi.

Per sondare la struttura delle rappresentazioni interne degli stimoli generate in corrispondenza ai diversi protocolli di presentazione, si deve verificare, durante le sequenze di apprendimento, la risposta della rete ad una varietà di stimoli. Questi stimoli, che chiameremo ‘stimoli di richiamo’ devono essere presentati per un tempo abbastanza breve da non indurre a loro volta modificazioni sinaptiche, e devono rappresentare un buon campionamento dello spazio di tutti i possibili stimoli, in modo da fornire una buona descrizione dell’insieme delle risposte della rete.

Riprendendo ancora una volta la metafora del paesaggio, immaginiamo ora di essere in volo al di sopra di questa complicata orografia, e di volerne studiare la struttura; che una nebbia ci impedisca di distinguere le valli,

ma che al fondo di ciascuna di esse qualcuno accenda una luce appena una palla che noi lanciamo dall'alto raggiunge il fondo della valle, con un colore diverso per ogni valle. Un modo ragionevole di ricostruire la distribuzione e l'ampiezza delle valli consisterebbe nel sorvolare l'intero spazio sovrastante il paesaggio, lanciare con regolarità delle palle verso terra (annotando ogni volta la posizione di partenza), e registrare il colore e la posizione della luce che viene accesa in seguito al lancio. Alla fine avremmo una mappa di tutte le posizioni iniziali di lancio, scomposta in regioni "colorate" che corrispondono al "bacino di attrazione" di ogni valle.

Il problema che si deve affrontare per ricostruire la struttura dello spazio degli stati della LANN27 è in linea di principio abbastanza simile. Però il numero di possibili stati (di "posizioni di lancio") da usare come sonde è enorme, e inoltre la dimensionalità alta dello spazio degli stati rende difficile una semplice rappresentazione del "paesaggio".

### **Una rappresentazione grafica dello spazio degli stati**

Si è dovuto quindi: 1) cercare una rappresentazione dello spazio degli stati (a 27 dimensioni) della LANN27 che rendesse intellegibile la sua struttura, e 2) definire delle grandezze in grado di descrivere quantitativamente (e sinteticamente) il processo di classificazione.

L'apprendimento nel nostro caso consiste nella creazione di attrattori intorno ai prototipi delle classi. Il monitoraggio del comportamento della rete consiste nel presentare alla rete un numero elevato di stimoli che costituiscano un buon campione dello spazio degli stimoli possibili, e nel classificare lo stato di attività persistente che segue ognuno degli stimoli. Si parla di un campione, perchè il numero totale di stimoli possibili, anche per una rete così piccola, è enorme ( $2^{27}$ ).

Il risultato del "test" è una rappresentazione di tutti gli stimoli insieme al loro punto d'arrivo. Visto che lo spazio degli stimoli è a 27 dimensioni, la rappresentazione grafica della dinamica della rete richiede una proiezione su uno o più spazi di 2 o 3 dimensioni. Una possibilità è illustrata nella figura 6.

Siccome l'aspettativa è che la struttura dinamica si formi intorno ai prototipi, gli spazi di proiezione vengono scelti uno per ogni coppia di prototipi apparsi nell'apprendimento. Ad ogni prototipo viene assegnato un colore fondamentale (rosso, verde, blu). Per ogni coppia di prototipi ( $P_1$  e  $P_2$  nella figura) si fissano sul piano due punti rappresentativi a distanza uguale alla

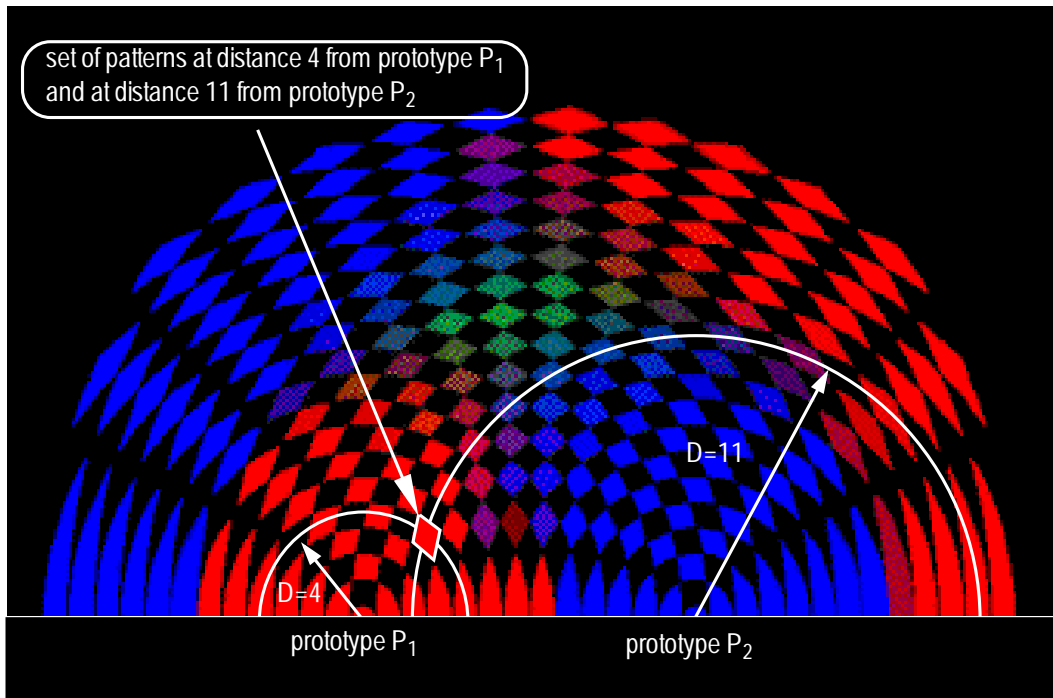


Figure 6: Rappresentazione grafica della dinamica della rete per una coppia di prototipi ('rosso' e 'blu'). Ogni losanga rappresenta l'insieme di tutti gli stati della rete che sono a distanza fissata dal prototipo rosso e da quello blu (ad esempio, quella selezionata è a distanza  $P_1 = 4$  dal rosso e  $P_2 = 11$  dal blu). Quando la rete parte da questi stati, la dinamica può evolvere verso uno dei prototipi o verso un altro stato e i colori delle losanghe rispecchiano questo comportamento dinamico come descritto nel testo.

distanza di Hamming tra i due prototipi (il numero di neuroni in stati diversi nei due prototipi). Una losanga nel piano della figura rappresenta l'insieme di stimoli di richiamo che sono ad una data distanza da ognuno dei due prototipi. Il colore della losanga rappresenta la statistica dei punti d'arrivo degli stimoli che partono da questa distanza: ogni stimolo di richiamo che porta la rete ad un attrattore vicino ad un dato prototipo, dà un contributo alla miscela di colori della losanga del colore corrispondente a questo prototipo.

Dunque, se per esempio intorno al punto rappresentativo del prototipo 'rosso' si estende un'area rosso vivo, questo vuol dire che la maggior parte degli stimoli corrispondenti alle losanghe interne all'area provoca nella rete DAD vicine a quel prototipo. Questa rappresentazione fornisce una misura del bacino di attrazione del prototipo (la dimensione della 'valle' associata). Analogamente, una zona di colore blu, ma poco intensa, indica un insieme di stimoli che il più delle volte ha provocato nella rete DAD lontane dai tre prototipi, e DAD vicine al prototipo 'blu' negli altri casi. Una zona nera indica un insieme di stimoli che ha prodotto in tutti i casi risposte lontane da tutti i prototipi. Da notare che le losanghe nere nella figura 6 non sono dovute a buchi nel bacino d'attrazione blu o rosso. Risultano invece dalla struttura discreta dello spazio degli stati, e non corrispondono a nessuno stimolo. Le zone blu e rosse lontane dai prototipi e vicine al bordo, corrispondono ad anti-attrattori e sono dovute ad una simmetria particolare del modello tra gli stati attivi e quelli inattivi dei neuroni.

Durante l'apprendimento si presenta alla rete una lunga sequenza di stimoli secondo uno dei protocolli. Ogni stimolo viene presentato per un tempo sufficientemente lungo da provocare dei cambiamenti sinaptici (stimoli di apprendimento). Periodicamente si effettuano dei "cicli di richiamo" per controllare il funzionamento acquisito della rete. La figura 7, a sinistra, mostra l'evoluzione degli attrattori con un protocollo di presentazione *incrementale*.

Inizialmente, prima dell'apprendimento, il nero dominante indica una struttura sinaptica scorrelata dai prototipi, tale che le risposte della rete alle prime presentazioni del prototipo ('rosso') risultano in media lontane da tutti e tre i prototipi. Si vede traccia di questo nella prima 'istantanea' dello spazio degli stati (primo riquadro in alto nella colonna a sinistra in figura 7), presa dopo cinque presentazioni del prototipo 'rosso'. Le zone nere, che sono piuttosto estese, non sono dovute solo alla discretizzazione ma indicano una effettiva mancanza di bacini di attrazione per i tre prototipi.

In seguito, solo lo stimolo 'rosso' viene presentato ripetutamente alla rete.

L'attrattore rosso si forma, ed il suo bacino copre presto la maggior parte dello spazio. Quando il flusso di stimoli si arricchisce (prima il prototipo blu viene presentato insieme al rosso, ed infine anche il prototipo verde entra nel flusso di input), le nuove informazioni competono con le vecchie per creare nella matrice sinaptica una traccia sufficiente a determinare un attrattore. Alla fine, quando il numero di stimoli ricorrenti nel flusso di input supera il *limite di capacità* della rete, le sinapsi disponibili non saranno più sufficienti a mantenere una rappresentazione interna stabile per ognuno di essi. Come vedremo, la LANN27 raggiunge il suo limite di capacità già per  $p = 3$  stimoli. Per tre prototipi la rete è ancora in grado di strutturare il suo spazio degli stati in un numero corrispondente di ampie valli, con il fondovalle vicino ai prototipi. Rispetto ai casi  $p = 1$  e  $p = 2$  però, oltre a queste valli principali (e per lo più all'interno di esse) compaiono valli minori, corrispondenti ad altrettanti attrattori della dinamica.

La parte destra della figura 7 illustra la proprietà esibita dalla rete di *dimenticare* informazioni già apprese, che però non si presentano più per lungo tempo nel flusso di stimoli, a favore di informazioni nuove. In questo caso si è adottato il protocollo di *palinsesto* per l'apprendimento. Come si vede dalla figura, dopo aver formato attrattori stabili in seguito alla presentazione ripetuta dei prototipi rosso e blu, la rete reagisce al cambiamento nel flusso di stimoli: il prototipo blu non compare più, e l'attrattore corrispondente tende gradualmente a restringere il suo bacino, mentre vengono presentati ripetutamente i prototipi rosso e verde, e si forma l'attrattore corrispondente al prototipo verde, che convive ora con il rosso.

### Capacità di memoria della LANN27

Abbiamo accennato al fatto che per  $p = 3$  la rete ha raggiunto il suo limite di capacità. Infatti il comportamento della rete in vari scenari di apprendimento è stato descritto solo fino a tre prototipi nel flusso degli stimoli. Una rete piccola esibisce un comportamento complesso intorno al limite di capacità; sia appena sotto, che appena sopra di esso, si presenta un numero elevato di attrattori spuri accanto a quelli che rappresentano gli stimoli appresi, con bacini ridotti. Questo è quello che si osserva, per esempio, nel modello di Hopfield. Per controllare le caratteristiche dell'apprendimento della rete è essenziale avere un criterio per stabilire il limite di capacità, perchè a questo punto il comportamento della rete diventa sensibile alle piccole fluttuazioni

nella sequenza degli stimoli.

Se la rete è al di sotto del suo limite di capacità, ogni stimolo appreso scava un attrattore, e quando vengono presentati, per richiamo, il prototipo appreso o i suoi vicini, la rete rilassa nella stessa distribuzione di attività neuronale, lo stesso attrattore. Quando anche l'intorno del prototipo non è più compatto, la rete è arrivata al limite di capacità.

Nella figura 8, l'istogramma a sinistra mostra la distribuzione delle distanze massime (su molti insiemi diversi di  $p$  prototipi,  $p = 1, 2, 3, 4$ ) tra gli attrattori dei prototipi (stimoli appresi) e quelli dei loro primi vicini. Per  $p \leq 2$  i prototipi e i loro primi vicini provocano nella rete la stessa DAD (distanza massima nulla) in quasi tutti i casi (99%), e quindi la rete è al di sotto del limite di capacità. Il caso  $p = 3$  è al confine: la coincidenza richiesta vale ancora in un numero elevato di casi (47%), ma molti stimoli che differiscono da un prototipo per un solo stato neuronale portano ad un attrattore lontano dall'attrattore a cui porta il prototipo stesso: la struttura del bacino d'attrazione si è frantumata. Il caso  $p = 4$  risulta nettamente al di sopra del limite di capacità, in quanto per quasi tutti gli insiemi di prototipi appresi non c'è un attrattore con bacino compatto, come attesta la popolazione di distanze elevate nell'istogramma corrispondente a  $p = 4$ . La parte destra della figura mostra la stessa analisi per i vicini dei prototipi a distanza di Hamming pari a due, in cui si vede che già per  $p = 3$  le distanze di Hamming tra gli attrattori dei secondi vicini e quelli dei prototipi hanno un'ampia dispersione.

L'analisi riassunta in questo paragrafo e nel precedente ci porta dunque a concludere che per  $p = 1, 2$  si hanno essenzialmente  $p$  attrattori, quasi sempre coincidenti con i prototipi. Quando si raggiunge il limite di capacità ( $p = 3$ ) lo spazio degli stati si affolla di attrattori, che tendono ancora a raggrupparsi intorno ai prototipi; un prototipo ha ancora lo stesso attrattore dei suoi primi vicini in molti casi. Per  $p > 3$  gli attrattori si distribuiscono su tutto lo spazio degli stati.

### **Generalizzazione e robustezza al rumore**

Tra le aspettative più importanti per la rete, c'è la sua capacità di *generalizzazione*: la capacità cioè di riconoscere un insieme di stimoli simili come una classe, associando ad essi una unica rappresentazione interna (attrattore). La rete dovrà dunque essere capace di estrarre un *prototipo* della classe, che

potrà non coincidere con alcuno degli stimoli visti durante l'apprendimento, ma dovrà catturare le loro caratteristiche comuni. Nel protocollo di generalizzazione si generano a caso molti insiemi di  $p$  prototipi ( $p = 1, 2, 3, 4$ ), e per ogni prototipo si effettua l'apprendimento utilizzando stimoli scelti a caso tra i primi vicini del prototipo. Nella fase di richiamo si sonda la risposta della rete a tutti i primi vicini dei  $p$  prototipi, e ai prototipi stessi. I prototipi stessi non vengono mai usati durante l'apprendimento.

Per ottenere delle osservabili rappresentative della capacità di generalizzazione si calcola per ogni  $p$ : 1) la distribuzione, su tutti gli insiemi di  $p$  prototipi, della massima distanza di Hamming tra i prototipi e gli attrattori dei loro primi vicini; 2) la analoga distribuzione delle distanze tra gli attrattori dei prototipi e gli attrattori dei loro primi vicini.

Dagli istogrammi a sinistra in figura 9 si vede che per  $p=1$  e 2 l'attrattore dei primi vicini del prototipo coincide con l'attrattore del prototipo stesso: si è formata dalla classe di stimoli una DAD che rappresenta anche il prototipo da cui è stata generata la classe, nonostante il fatto che questo prototipo non fosse comparso durante l'apprendimento. Un quadro simile risulta se si suppone che la DAD sia simile al prototipo stesso (istogrammi a destra). Al crescere di  $p$  la distribuzione di distanze si allarga, finchè per  $p = 4$  gli attrattori dei primi vicini sono essenzialmente scorrelati dal prototipo. Questi risultati mostrano, nei limiti imposti dal limite di capacità, come la rete generi dinamicamente una rappresentazione interna per una classe di stimoli, sulla base delle correlazioni tra di essi.

Come abbiamo osservato sopra, ci si attende che le proprietà di classificazione della rete si mantengano quando si inserisce nel flusso di stimoli un certo numero di stimoli casuali, scorrelati dalle classi. Il basso limite di capacità della rete limita i test quantitativi di questa proprietà.

Nella Figura 10 riportiamo l'analisi della robustezza rispetto al rumore per il caso  $p = 1$ , con 1 o 2 stimoli casuali per ogni presentazione del prototipo. I grafici mostrano la frazione di stimoli di richiamo, situati a varie distanze dal prototipo, che vengono attratti da uno stato a distanza massima 3 ('tolleranza') dal prototipo, in seguito alle sequenze di apprendimento menzionate. In ogni figura vengono costruiti questi grafici per cinque diverse scelte del prototipo. Si può dedurre dalla figura che nei casi presentati si hanno attrattori vicini ai prototipi (o coincidenti con essi), con ampi bacini di attrazione nella maggioranza dei casi. L'effetto degli stimoli casuali è di restringere i bacini d'attrazione degli attrattori.

Nel caso  $p = 2$ , l'inserimento di uno stimolo casuale dopo ogni presentazione dei prototipi porta la rete al limite di capacità, e solo nel 20% circa dei casi si sviluppano bacini di attrazione ampi per entrambi i prototipi.

## **Dalla LANN27 a reti con neuroni impulsati in VLSI**

La LANN27 costituisce, sia nella sua ricchezza che nelle sue limitazioni, un importante dispositivo pilota per diversi motivi:

- ha fornito il primo sistema completamente analogico e asincrono in grado di strutturarsi in modo non supervisionato, adattandosi a flussi arbitrari di stimoli, creando una rappresentazione interna delle loro caratteristiche statistiche in modo associativo;

- ha messo in luce le difficoltà, e le possibili strategie risolutive, connesse all'analisi della dinamica di apprendimento in un dispositivo reale;

- con le sue limitazioni, ha reso più chiara la strada da percorrere per la realizzazione di reti neuronali con apprendimento continuo di maggiori dimensioni e complessità, e più realistiche dal punto di vista biologico;

Tra le limitazioni della LANN27, appare ovvia quella di una bassa capacità di memoria, che impone la realizzazione di reti con un numero molto maggiore di neuroni. Questa prospettiva si scontra però con difficoltà in primo luogo tecniche, ma rappresentative anche di questioni di principio.

Innanzitutto c'è il problema delle dimensioni fisiche della rete. Nella realizzazione circuitale della LANN27 (che, ricordiamo, è a elementi discreti), lo spazio occupato dalla singola sinapsi eccede largamente quello occupato dal neurone. Questo fatto da un lato rappresenta un ostacolo fondamentale rispetto alla costruzione di grandi reti, perchè con il crescere del numero di neuroni il numero di sinapsi cresce molto rapidamente (quadraticamente); dall'altro è un indizio di un problema di principio.

Le grandi dimensioni della sinapsi sono dovute primariamente alla sorgente di rumore associata ad ogni sinapsi, che sostiene il meccanismo stocastico di apprendimento, e in particolare al circuito che genera e filtra in frequenza il rumore, che a sua volta è necessario per avere basse probabilità di transizione tra gli stati stabili sinaptici. L'esigenza di una sorgente di rumore 'lento' associata ad ogni singola sinapsi appare molto problematica da un punto di vista realizzativo, ed è innaturale da un punto di vista neurobiologico, visto che la sinapsi biologica è molto più piccola del neurone.

Inoltre, con reti grandi si presenta il problema dei consumi, che già sono enormi nella rete descritta (circa 200 W). Questa considerazione spinge verso la realizzazione di reti con molti neuroni in elettronica integrata (VLSI). Questa scelta, a sua volta, impone dei nuovi vincoli realizzativi che implicano delle diverse scelte sugli elementi costitutivi del modello.

Da un punto di vista generale, vi sono poi altre caratteristiche della LANN27 che richiedono di essere riconsiderate in vista di una realizzazione ‘biologicamente plausibile’:

- l’attività della rete dovrebbe riflettere in modo autonomo le diverse condizioni corrispondenti alla presenza dello stimolo (attività molto alta), all’attività selettiva nell’attrattore (di livello medio) e all’attività spontanea (di livello basso). Nella LANN27, come abbiamo visto, la presenza dello stimolo è codificata da un termine aggiuntivo ad hoc nell’input al neurone.

- con i neuroni ‘pseudobinari’ della LANN27, che assumono valori  $(-1, 1)$ , per ogni sinapsi tutte le quattro coppie possibili di valori pre- e post-sinaptici inducono un tentativo di modificare l’efficacia sinaptica. Inoltre il livello di codifica (che ricordiamo essere la frazione media di neuroni attivi - nello stato “1” - in ogni stimolo) è 50%. Entrambe queste caratteristiche sono innaturali dal punto di vista biologico (si ritiene che in assenza di segnale pre-sinaptico non vi sia modificazione dell’efficacia sinaptica, e sperimentalmente la frazione di neuroni eccitati da uno stimolo in una popolazioni corticali di aree associative e di altre aree ‘profonde’ è molto bassa). Dal punto di vista computazionale, entrambe queste caratteristiche tendono a mantenere bassa la capacità della rete. Tendono infatti a ‘velocizzare’ la dinamica sinaptica, aumentando il numero delle sinapsi che tentano di cambiare stato ad ogni stimolo facendo scomparire più in fretta la traccia degli stimoli precedenti dalla memoria.

Questi problemi trovano una soluzione naturale nel passaggio a neuroni che comunichino attraverso impulsi, come in effetti accade nel sistema nervoso reale.

Come vedremo in un caso specifico nel prossimo paragrafo, la frequenza di emissione di un neurone impulsante è determinata dall’intensità della ‘corrente’ ad esso afferente, a sua volta determinata sia dagli impulsi emessi dai neuroni della stessa popolazione, pesati dalle corrispondenti efficacie sinaptiche, sia da quelli provenienti dall’esterno. La successione di impulsi prodotti dalla rete appare molto simile ad un processo di estrazione casuale di eventi.

In assenza di stimoli, le connessioni tra i neuroni, e la corrente esterna,

sono tali da mantenere una bassa attività (basse frequenze di emissione) nella rete – ‘attività spontanea’. La presenza di uno stimolo si esprime in un aumento marcato della corrente esterna ad un piccolo sottoinsieme di neuroni, che innalzano molto la loro frequenza di emissione. Se lo stimolo è ‘familiare’ (è già impresso nella struttura sinaptica), alla sua scomparsa la rete rilassa in una DAD, in cui un sottoinsieme di neuroni (selettivo per quello stimolo) continua ad emettere a frequenze nettamente superiori a quelle di attività spontanea.

In questo contesto appare un nuovo candidato per la sorgente della stocasticità intrinseca delle transizioni sinaptiche: l’aleatorietà degli intervalli tra gli impulsi emessi dai neuroni pre- e post-sinaptico. Una sinapsi in grado di sfruttare questa sorgente di stocasticità non avrebbe bisogno di un generatore di rumore, come nel caso della LANN27, e il generatore di rumore risulterebbe distribuito su tutta la rete.

## **Dinamica di neuroni impulsati e sinapsi plastiche**

Gli elementi di base di un dispositivo integrato ispirato ai principi esposti sono il neurone IF (“integrate-and-fire”) e la sinapsi plastica. Il neurone IF quando riceve degli impulsi da altri neuroni, agisce come un circuito integratore in cui il potenziale ai capi del condensatore rappresenta la somma dei contributi degli impulsi arrivati. Quando il potenziale supera una certa soglia viene emesso un impulso e il potenziale del neurone torna ad un valore iniziale da cui ricomincia il processo di integrazione. Amit e Brunel hanno dimostrato [18] che per la stabilità dell’attività spontanea di una rete di elementi di questo tipo occorre che vi siano due tipi di neuroni: eccitatori e inibitori, come accade nella corteccia. I neuroni inibitori devono dare un contributo alla dinamica della rete sufficiente ad evitare che gli altri neuroni si eccitino l’un l’altro fino a giungere ad un’esplosione incontrollata di attività.

Le sinapsi che connettono i neuroni eccitatori tra loro sono quelle plastiche e la loro dinamica costituisce un’altra realizzazione del meccanismo di transizioni stocastiche schematizzato in precedenza. La sinapsi ha due stati stabili ed è stata progettata in modo da avere una rete a neuroni impulsati che si comporta come un classificatore degli stimoli in ingresso, analogamente a quanto veniva fatto dalla LANN27. Per ottenere questo comportamento, durante la presentazione di uno stimolo che deve essere imparato l’efficacia deve

tendere ad essere potenziata quando entrambi i neuroni sono in uno stato di attività elevata (meccanismo “hebbiano”). Una volta arrivati nello stato alto, deve esservi la possibilità di tornare nello stato basso (depressione). Nel nostro caso questo avviene quando il neurone pre-sinaptico emette impulsi a frequenza elevata e quello post-sinaptico ha solo attività spontanea. Inoltre, quando non vengono presentati stimoli e la rete è in condizioni di attività spontanea la sinapsi deve preservare la propria efficacia.

Le transizioni descritte devono essere stocastiche e il meccanismo che le realizza deve essere in grado di sfruttare l’aleatorietà del processo di emissione di impulsi da parte dei due neuroni connessi dalla sinapsi.

Lo schema di funzionamento della sinapsi che è stata realizzata in VLSI analogico e che soddisfa a queste condizioni è illustrato nella figura 11 dove vengono mostrati due casi in cui, a parità di frequenza pre-sinaptica, le diverse frequenze post-sinaptiche determinano transizioni verso l’alto (parte superiore della figura) o verso il basso (parte inferiore) dell’efficacia sinaptica.

Nel primo riquadro dall’alto di ognuna delle due parti viene rappresentata l’efficacia sinaptica in funzione del tempo durante la presentazione di uno stimolo. La dinamica si svolge nella regione compresa tra 0 e 1, che rappresentano gli stati stabili dell’efficacia sinaptica. In assenza di impulsi, tutti i valori che si trovano sotto la soglia  $J_0$  (la linea mediana) vengono attratti verso lo stato basso, mentre quelli sopra soglia vanno allo stato alto. Lo stato alto costituisce anche il limite superiore per l’efficacia sinaptica, e non può essere superato (analogamente, lo stato basso è il limite inferiore). Quindi questo meccanismo garantisce che in assenza di impulsi la sinapsi preservi uno dei due valori di efficacia.

La sorgente “Hebbiana” che spinge l’efficacia verso l’alto o verso il basso è funzione delle attività dei due neuroni connessi dalla sinapsi (nei due riquadri inferiori nella figura 11). L’efficacia riceve una spinta verso l’alto ogni volta che il neurone pre-sinaptico emette un impulso e il potenziale del neurone post-sinaptico si trova al di sopra di una certa soglia  $\Theta_{Hebb}$ . La spinta è invece verso il basso quando il neurone pre-sinaptico sta emettendo un impulso e il potenziale del post-sinaptico si trova al di sotto di  $\Theta_{Hebb}$ .

Quando il neurone post-sinaptico emette impulsi ad una frequenza elevata, la probabilità che il suo potenziale sia al di sopra della soglia  $\Theta_{Hebb}$  è molto elevata e dunque la maggior parte degli impulsi del pre-sinaptico tenderà a spingere l’efficacia verso l’alto. Se per caso si presentano abbastanza

eventi di questo tipo in un tempo breve, allora la sinapsi supererà la soglia e transirà allo stato alto.

Analogamente, quando il neurone post-sinaptico emette a basse frequenze, le spinte attivate dagli impulsi del pre-sinaptico saranno prevalentemente verso il basso (parte inferiore della figura 11) permettendo all'efficacia sinaptica di transire allo stato depotenziato. Le transizioni avvengono solo se durante la presentazione dello stimolo arrivano abbastanza impulsi dal pre-sinaptico, il che succede con una certa probabilità minore di 1.

Nella figura 12 viene riportato il layout di un chip aVLSI (*analog VLSI*, VLSI analogico) che realizza una rete di 21 neuroni IF impulsati, di cui 14 eccitatori e 7 inibitori. Le sinapsi tra neuroni eccitatori sono plastiche, mentre tutte le altre sono fisse. Questa piccola rete pilota è stata realizzata per testare i vari elementi e la loro interazione in un regime di correnti estremamente basse. Il passo successivo sarà la costruzione di un chip con alcune centinaia di neuroni e un sistema multi-chip, per arrivare ad una rete di migliaia di neuroni, che già rappresenterebbe un dispositivo d'interesse funzionale e computazionale. L'integrazione di un sistema multi-chip richiederà la soluzione del problema della comunicazione che viene descritto in questo stesso volume (Douglas).

## References

- [1] Gerald M. Edelman 1987 Neural Darwinism: The Theory of Neuronal Group Selection *Basic Books*
- [2] Fuster JM 1973 Unit activity prefrontal cortex during delayed-response performance: neuronal correlate of transit memory, *J. Neuropsychol.*, **36** 61
- [3] Niki H 1974 Prefrontal unit activity during delay alternation in the Monkey, *Brain Res.* **68** 185.
- [4] Goldman-Rakic P 1987 Circuitry of primate prefrontal cortex and regulation of behaviour by representational memory, *Handbook of Physiology - The nervous system V* chapter 9.
- [5] Hebb D 1949, The organization of behaviour, Wiley, NY
- [6] Hopfield JJ 1982 Neural networks and physical systems with emergent selective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554

- [7] Sompolinsky H., Neural networks with nonlinear synapses and a static noise, *Phys. Rev. A*, **34**, 2571 (1986)
- [8] Sompolinsky H., The theory of neural networks: The Hebb rule and beyond, in L. van Hemmen and I. Morgenstern eds. *Heidelberg Colloquium on Glassy Dynamics*, 485-527 (Springer-Verlag, Heidelberg, 1987)
- [9] D Willshaw 1969, Non-holographic associative memory, *Nature*, London, **222**, 960
- [10] Mead C 1989, Analog VLSI and neural systems, Addison Wesley.
- [11] Miyashita Y 1988 Neuronal correlate of visual associative long-term memory in the primate temporal cortex, *Nature* **335** 817
- [12] D.J. Amit 1989 Modeling brain function, Cambridge UK: Cambridge University Press. (trad. it.: Modellizzare le funzioni del cervello, CEDAM, 1995)
- [13] D.J. Amit and S. Fusi 1994 Learning in neural networks with material synapses, *Neural Computation* **6** 957
- [14] D.J. Amit, S. Fusi, V. Yakovlev 1997 A paradigmatic working memory (attractor) cell in IT cortex , *Neural Computation*, **9** 1101
- [15] N. Brunel, F. Carusi, S. Fusi 1998 Slow stochastic Hebbian learning of classes of stimuli in a recurrent neural network, *Network*, **9**, 123
- [16] V. Yakovlev, S. Fusi, E. Berman, E. Zohary 1998 Inter-trial neuronal activity in infero-temporal cortex: a putative vehicle to generate long term associations, inviato a *Nature Neuroscience*
- [17] J G Nicholls, A R Martin e B G Wallace 1992, From neuron to brain, Sinauer Associates Inc.  
I B Katz 1966,  
Nerve, muscle, synapse, Mc Graw Hill NY
- [18] D.J. Amit and N. Brunel 1997 Model of global spontaneous activity and local structured (learned) delay activity during delay periods in cerebral cortex, *Cerebral Cortex*, **7**, 237-252
- [19] D Badoni, S Bertazzoni, S Buglioni, G Salina, D J Amit e S Fusi 1995, *Network* **6**, 125
- [20] P. Del Giudice, S. Fusi, D. Badoni, V. Dante, D.J. Amit 1998 Learning attractors in an asynchronous, stochastic electronic neural network, *Network*, **9**, 183

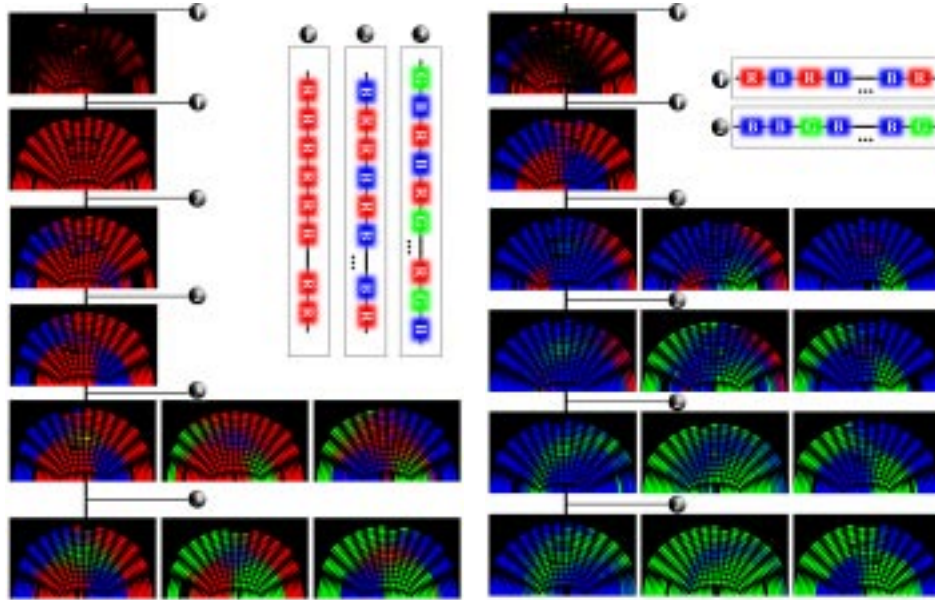


Figure 7: Evoluzione temporale (dall'alto, stato iniziale, verso il basso, stato finale) dello schema che rappresenta la dinamica della rete per due protocolli d'addestramento: l'incrementale a sinistra e quello di palinsesto a destra. Per ogni protocollo, in orizzontale, sono riportate le proiezioni per ogni coppia di stimoli che sono stati presentati alla rete. Le sfere numerate indicano le sequenze di presentazione nelle diverse fasi dell'apprendimento, di cui si fornisce un esempio nei riquadri associati. Nel protocollo incrementale, all'inizio vi è una sola proiezione (prototipo rosso e poi, dalla terza riga, prototipi rosso e blu), mentre alla comparsa del prototipo verde (quinta riga) sono riportate anche le altre due possibili proiezioni (coppia rosso-verde e verde-blu). Analogamente per il protocollo di palinsesto. Vedi il testo per la descrizione della fenomenologia.

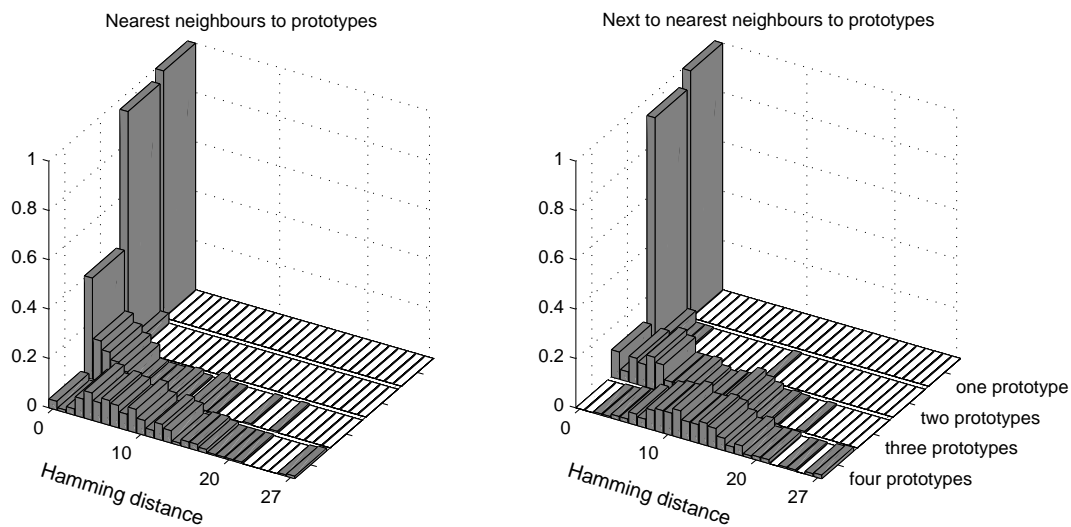


Figure 8: Capacità di memoria della rete: distribuzioni per diverso numero di prototipi delle distanze massime tra gli attrattori dei prototipi e quelli dei loro primi vicini (a sinistra) e dei vicini a distanza 2 (a destra). L'allargamento della distribuzione (per  $p = 3$ ) indica che la rete ha raggiunto il limite di capacità e non è in grado di dare una risposta (DAD) coerente a stimoli simili ai prototipi imparati.

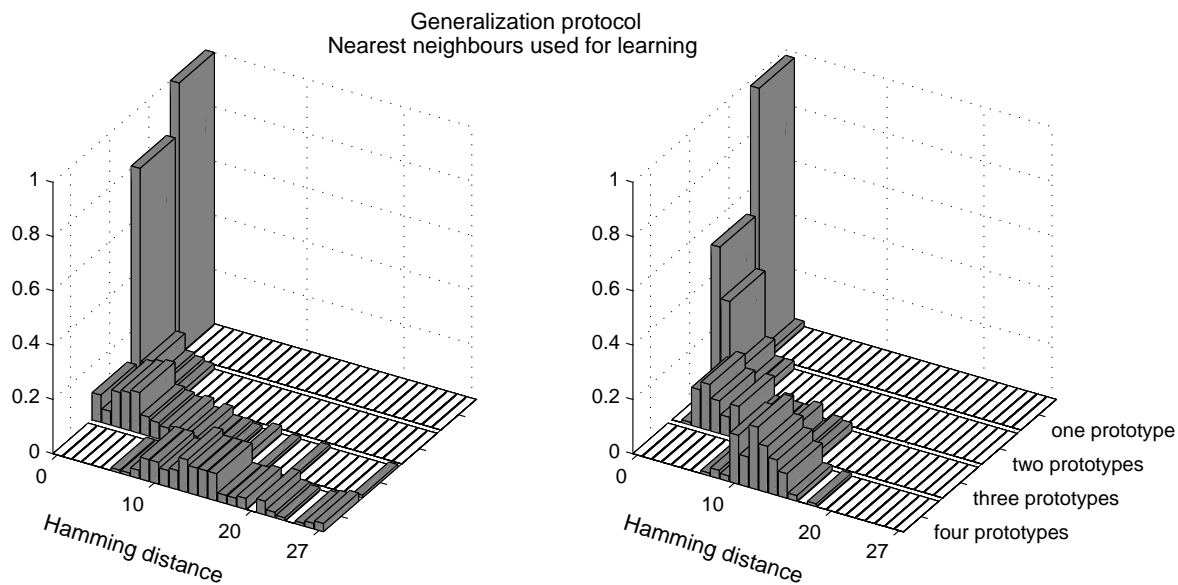


Figure 9: Estrazione del prototipo (per numero diverso di classi): A sinistra: la distribuzione della massima distanza tra gli attrattori dei prototipi e gli attrattori dei primi vicini. Si vede che per  $p=1$  e 2 il prototipo, che non é stato presentato durante l'apprendimento, condivide l'attrattore con la la classe da lui generata. Al superamento della capacità di memoria si perde anche la capacità di generalizzazione. A destra: la distribuzione della massima distanza tra i prototipi usati per generare le classi e gli attrattori delle classi. Quando questa distanza è quasi sempre nulla, l'attrattore della classe coincide con il prototipo.

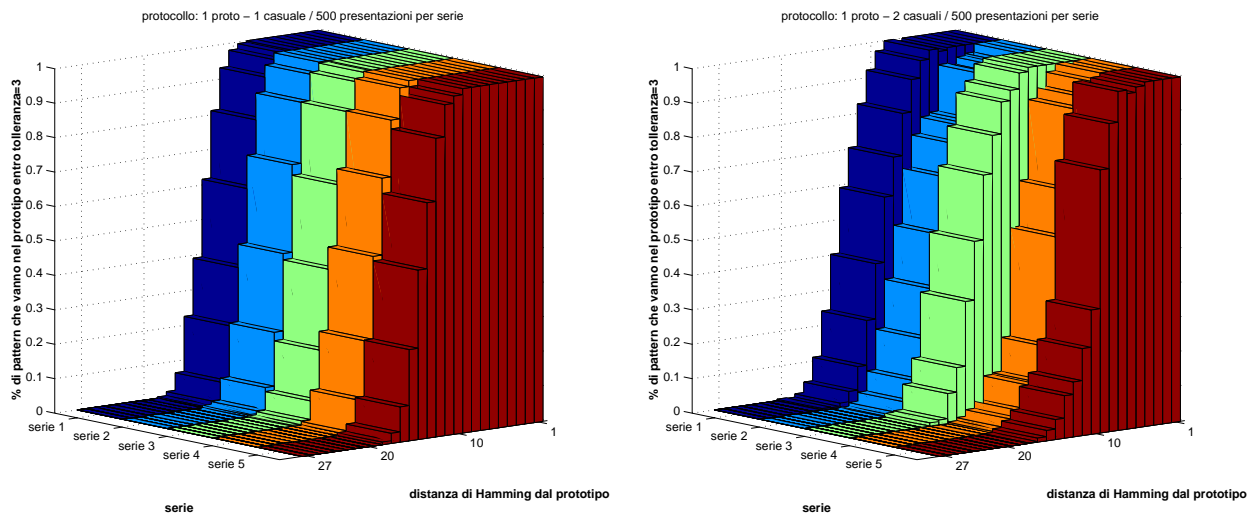
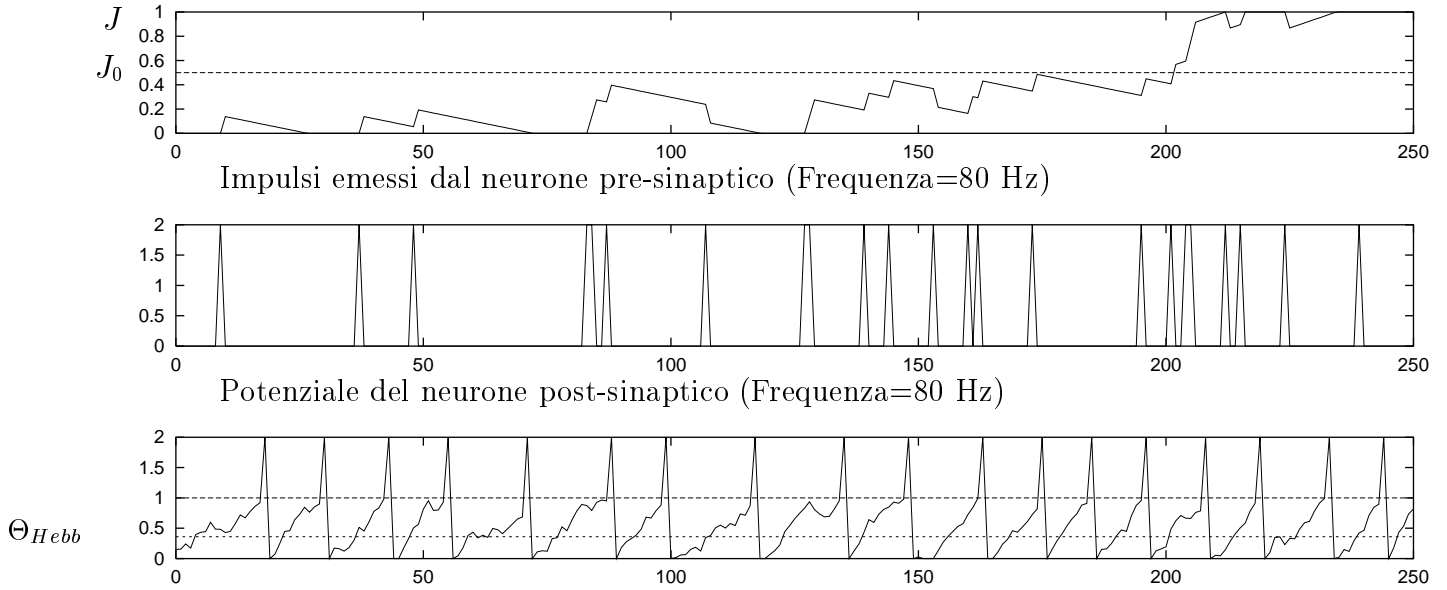


Figure 10: Robustezza al rumore nella sequenza degli stimoli: istogrammi della frazione di stimoli che partendo da diverse distanze vengono attratti dai prototipi imparati. Sono riportati i risultati per due diversi protocolli. A sinistra vengono presentati, durante l'apprendimento, alternativamente un prototipo e uno stimolo ogni volta scelto a caso, per cinque scelte diverse del prototipo (5 serie). A destra le presentazioni del prototipo durante l'apprendimento sono disturbate da due stimoli casuali. Come si può vedere, nonostante il disturbo degli stimoli casuali, l'attrattore del prototipo viene imparato e ha un grosso bacino di attrazione.

### POTENZIAMENTO A LUNGO TERMINE



### DEPRESSIONE A LUNGO TERMINE

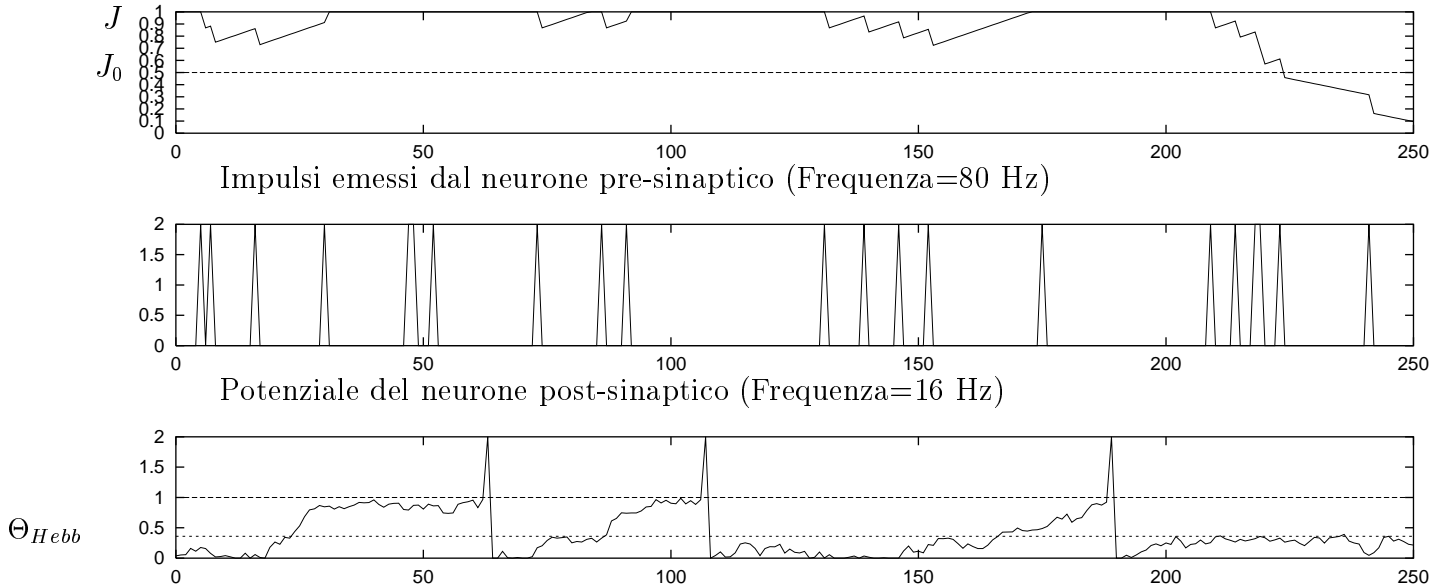


Figure 11: Dinamica dell'efficacia sinaptica: due esempi di transizioni in condizioni diverse. In ogni riquadro sono riportati in funzione del tempo (partendo dall'alto): l'efficacia sinaptica  $J$ , gli impulsi del neurone pre-sinaptico, e il potenziale del neurone post-sinaptico. In alto i neuroni pre- e post-sinaptico sono attivi e la sinapsi effettua una transizione allo stato alto. In basso il neurone pre-sinaptico è attivo mentre quello post-sinaptico è inattivo e la sinapsi, partendo dallo stato alto, si ritrova alla fine nello stato basso. Vedi il testo per la spiegazione della dinamica.

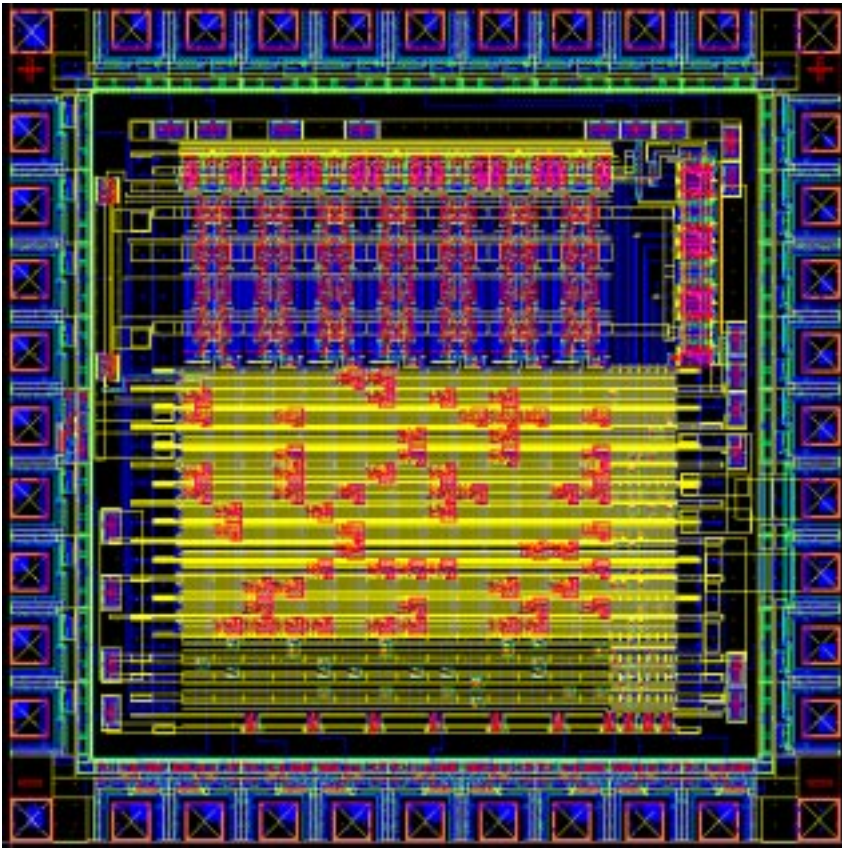


Figure 12: Layout del chip LANN21