Det ville være nyttigt med en algoritme, der modsat OBD kunne lægge forbindelser ind, der hvor de var mest nødvendige. En sådan algoritme kendes desværre ikke.

Måske er det mest hensigtmæssigt simpelthen at skabe et overskud af forbindelser for derefter at skære dem bort, der ikke er påkrævede. Når vi fødes, er der for eksempel flere forbindelser til musklerne end nødvendigt, men under brugen af musklerne reduceres dette antal hurtigt. Naturens egne metoder til neural vækst er blevet udviklet gennem en milliard år og berettiger nok til et indgående studium.

Benny Lautrup er lektor i teoretisk fysik ved Niels Bohr Institutet og leder af det danske forskningscenter for neurale netværk: CONNECT.

Biologisk beregning

John Hertz

Hvis vi vælger at kalde det, hjernen gør, for beregning, er det klart nok af en temmelig anden form, end den vi kender til i de sædvanlige computere eller i *feedforward* neurale netværk. Så simple modeller kan ikke beskrive systemer med så udstrakt et *feedback*, som hjernen har. I de senere år har neurovidenskaben taget ideer op fra fysikken, for eksempel fra teorien for dynamiske systemer og fra den statistiske mekanik, for at finde nye formuleringer på de neurovidenskabelige problemstillinger.

Neurovidenskabens mål er at karakterisere psykologiske fænomener, som for eksempel sansning, opmærksomhed, hukommelse og bevidsthed, gennem dynamikken i hjernens neurale netværk. I lang tid vægrede neurobiologerne sig mod at studere sådanne problemer på grund af vanskeligheder med at foretage eksperimenter, som kunne kaste et klart lys over dem. I de seneste år har begreber importeret fra fysikken imidlertid gjort det muligt at konstruere et teoretisk underlag for eksperimentel undersøgelse af disse fundamentale problemer. I denne korte artikel vil jeg kort beskrive de vigtigste ideer og nogle eksperimenter til at teste dem med.

Attraktorer

Hjernen er langt mere kompleks end noget kunstigt neuralt netværk, som vil blive konstrueret inden for en overskuelig fremtid. Ikke desto mindre er den et dynamisk system, som er underkastet den klassiske fysiks love. Desuden er den et såkaldt *dissipativt* system, hvilket vil sige, at den forbruger energi. Vi ved, at den asymptotiske opførsel af sådanne systemer kan beskrives ved *attraktorer* — karakteristiske bevægelsesmønstre som systemet til sidst ender op i.

Der er tre slags attraktorer. Den første type er et *fikspunkt*, som i dette tilfælde ville beskrive et neuronalt aktivitetsmønster, der ikke ændrer sig i tiden. Den anden type er en *grænsecyklus*, hvor aktivitetsmønstret er periodisk i tiden. Endelig findes der *sære attraktorer* hvor aktivitetsmønstret ændrer sig kaotisk i tiden.

For alle tre typer af attraktorer er systemets bevægelse stærkt begrænset i forhold til det fuldstændige tilstandsrum, som det i princippet kunne bevæge sig i. Dette er klart for fikspunkter: attraktoren er et enkelt punkt i et stort rum. Men selv sære attraktorer har en effektiv (fraktal) dimension som er mindre end det fulde tilstandsrum, så at attraktoren har målet nul. At systemets udvikling tiltrækkes mod et så specielt sæt tilstande er en konsekvens af dissipationen og er et meget kraftigt matematisk udsagn om dynamikken.

Walter Freeman fra Berkeley universitetet i Californien har brugt disse resultater fra teorien for dynamiske systemer og har fremsat den formodning, at elementære psykologiske begivenheder kan forbindes med, at hjernen (eller en del af den) falder ind i en attraktor. Ifølge denne tanke, kan enhver sansning, enhver erindring, o.s.v. identificeres med en eller anden attraktor. For eksempel, når et dyr sanser en bestemt duft, så vil de områder som i hjernen forbindes med lugtesansen (det olfaktoriske cortex) falde ind i en attraktor, og sanseoplevelsen af duften er simpelthen identisk med denne proces. Hvis to forskellige kemikalier påvirker duftmodtagecellerne i næsen, vil de

opfattes som forskellige dufte, hvis og kun hvis de driver systemet ind i forskellige attraktorer.

Freeman gennemførte omfattende målinger af de lokale elektriske potentialer på et sæt elektroder indsat i det olfaktoriske cortex. Resultaterne var konsistente med attraktorantagelsen. Der var en attraktor (en sær en) for systemets hviletilstand og andre (tilnærmelsesvis grænsecykler) forbundet med sansede dufte.

Gælder et lignende billede for resten af hjernen? I eksperimenter udført af Charles Gray og Wolf Singer på Max Planck institutet i Frankfurt og af Reinhold Eckhorn og medarbejdere fra universitetet i Marburg, blev elektrodepar placeret i det primære visuelle cortex på en kat. Begge elektroder reagerede på bevægelse af en rektangulær figur inden for et lille *receptivt felt* centreret omkring et bestemt punkt i synsfeltet (de to receptive felter overlappede ikke). Groft taget udviste begge celler en cyklisk opførsel, nogenlunde som de duftattraktorer, der blev observeret af Freeman i det olfaktoriske cortex. Der er altså belæg for attraktorer i forbindelse med sansning af bevægelse i bestemte områder af synsfeltet.

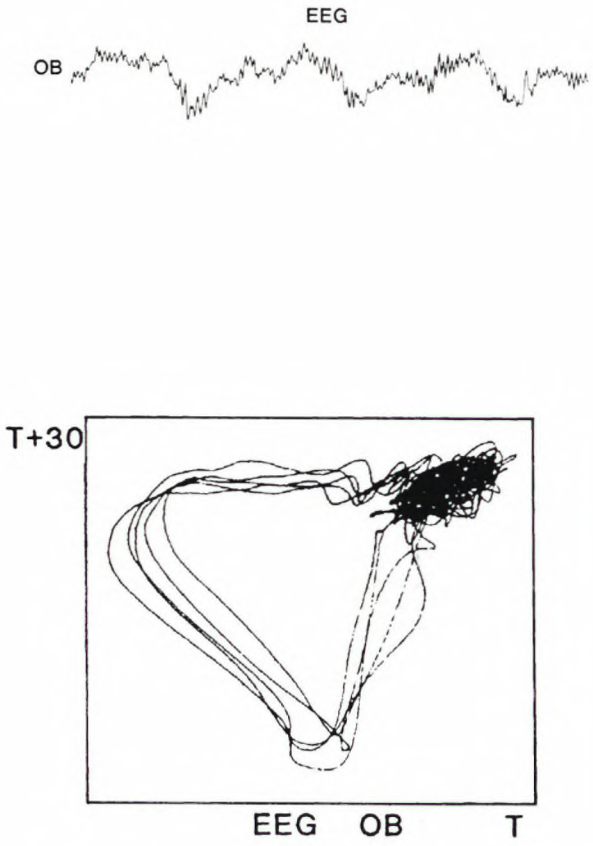
Men hermed ender historien ikke. Det blev overraskende fundet, at når en enkelt stor rektangulær figur blev ført gennem de to receptive felter, så var der en tendens til synkronisation i svingingerne i de to cellers aktivitet. I vort billede svarer synkroniserede og usynkroniserede svinginger til forskellige attraktorer. Det så altså ud til at sansning af et enkelt sammenhængende objekt kunne identificeres med en fælles attraktor.

Ideen om at der skulle være en sådan generel forbindelse mellem neurofysiologi og psykologi er klart nok meget tiltrækkende. Det er dog stadig uklart hvor omfattende den er. Hos aber ses sådanne svingninger kun sjældent, så der synes ikke at være noget, der kan synkroniseres. Det er naturligvis stadig muligt at tidsligt forbundne og tidsligt adskilte stimuli forbindes med forskellige (muligvis sære) attraktorer, men det er endnu ikke undersøgt.

Attraktorbilledet kan faktisk kun være en tilnærmelse, ikke en absolut lov for neural funktionalitet. Det er klart, at de dele af hjernen som modtager signaler fra de sanseområder, vi har diskuteret her, vil begynde at reagere så snart, de modtager meningsfyldte signaler fra disse områder. De vil ikke vente på at områderne evt. falder ind i attraktorer. Den funktionelle relevans af attraktorerne kan blot være, at de karakteriserer bestemte vedblivende aktivitetsmønstre, som de efterfølgende hjerneområder er kommet til at opfatte som meningsfyldte.

Indlæring og hukommelse

Hjernen er meget plastisk og ændrer sin opførsel ud fra sine erfaringer. For eksempel, ændrer de duftspecifikke attraktorer sig, når dyret udsættes for nye lugte. Hvorledes foregår disse forandringer, og hvordan påvirker de hjernens attraktorer? Psykologen Donald Hebb foreslog i 40'erne at styrken af de synaptiske forbindelser mellem neuronerne blev forstærket i samklang med deres aktiviteter, og at dette var den basale mekanisme bag indlæring og hukommelse.



Figur 1. Biologisk attraktor. Øverste del af figuren viser det elektriske potentiale målt i en rottes lugteområde under en kunstigt fremkaldt lokal krampe. Nederste del viser et Ruelle plot af en lignende optagelse (potentialstyrkemålinger med 30 millisekunders afstand afbildet mod hinanden). Man ser hvorledes plotpunktet vandrer rundt en attraktor bestående af store løkker svarende til en sinuskurve med høj amplitude, afbrudt af en tilfældig, kaotisk bevægelse med lille amplitude.

De synapsestyrker som dannes på grundlag af denne såkaldte Hebb'ske indlæring fører til dannelsen af mange attraktorer, der hver er knyttet til en bestemt erindring. Den præcise form for denne proces er i årenes løb blevet studeret i mange forskellige modeller. En særlig simpel model blev indført af fysikeren John Hopfield fra CalTech for lidt over 10 år siden. Modellen viste sig at kunne løses eksakt ved hjælp af moderne metoder fra den statistiske mekanik. Selv om den var alt for simpel til at kunne anvendes i neurobiologien, så viste den dog eksplicit hvorledes udefra påtrykte aktivitetsmønstre kunne skabe en attraktor for hvert mønster.

Nylige eksperimenter udført af Y. Miyashita og medarbejdere har kastet nyt lys på indlæringsmekanismen, og hvorledes den påvirker strukturen af attraktorerne for hjernens dynamik. De studerede en bestemt klasse neuroner

i en del af hjernen som menes at have at gøre med visuel hukommelse. Det viste sig faktisk, at neuronerne her var stærkt aktive mellem præsentationer af mønstrene, som om de bevarede en erindring om det foregående mønster. Aberne blev trænet til at genkende et sæt mønstre, som altid blev præsenteret i samme orden. De resulterende attraktorer blev derefter studeret med mønstrene præsenteret i vilkårlig orden.

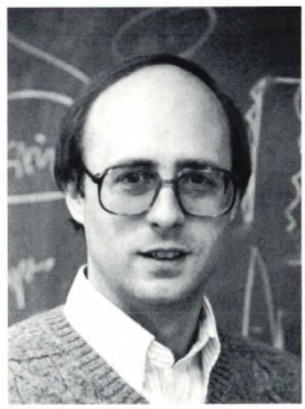
I overensstemmelse med attraktorbilledet fandt man forskellige attraktorer for hvert mønster. Men selv om forskerne havde gjort alt for at eliminere korrelationer mellem mønstrene, fandt de korrelationer mellem attraktorerne. Korrelationsgraden mellem et vilkårligt par attraktorer afhang af deres indbyrdes afstand i træningssekvensen.

Hvordan kan det nu forstås? Det Hebb'ske model for indlæring beskrevet ovenfor har plads til en naturlig udvidelse. Når der under indlæringen optræder en ny stimulus, vil der være et kort tidsrum i hvilket en neuron har skiftet over til den attraktor som svarer til det nye mønster medens en efterfølgende neuron som modtager signaler fra den første, endnu ikke har gjort det. Det vil føre til ændringer i de synaptiske styrker proportionalt med produktet af aktiviteterne i den præsynaptiske celle for attraktoren til det nye mønster og den postsynaptiske celle i det gamle. Alle attraktorerne vil få ændret deres struktur af disse ændringer i synapserne.

Det er ikke klart, hvorledes ændringerne præcist finder sted, og det er derfor nødvendigt at bruge modelberegninger for at få gennemskuelige resultater. Fysikerne Meir Griniasty, Mischa Tsodyks og Daniel Amit undersøgte nogle modeller som meget lignede Hopfield's, men hvor disse nye synaptiske ændringer var inkluderet. På trods af modellernes relative enkelhed er forbindelsen mellem de ekstra bidrag til synapsestyrkerne og de resulterende attraktorer ikke-triviel. Ikke desto mindre kunne de for et stort parameterområde vise, at de attraktorer der kom frem havde lige præcis den slags korrelationer som Yamashita havde observeret. Dette resultat giver tiltro til attraktorbilledet, den Hebb'ske indlæringshypotese, og den tanke, at idealiserede teoretiske modeller kan give indsigt i den virkelige neurovidenskab.

Perspektiv

Det er klart, at disse lidt forsigtige forsøg højst er de første skridt i den lange rejse mod en fuldstændig forståelse af neural beregning. Måske vil det vise sig, at de går i den forkerte retning. Men der er for tiden intet andet forslag på markedet, som giver et tilsvarende frugtbart grundlag for det samspil mellem teori og eksperiment, som er nødvendigt for at opnå en forståelse af nervesystemet. I det mindste har det bragt ideen om *teori* — i den form som fysikerne bruger den — til et felt som har kritisk behov for det. I det mindste på denne måde kan det ikke undgå at have en virkning.



John Hertz Professor ved NORDITA.
For tiden tilknyttet National Institute of Health, USA