

Mente e Consciência: Análise de um Modelo GNW Baseado em Redes de Hopfield Estocásticas

J N Tavares

Casa das Ciências/FCUP

16 de dezembro de 2025

Sumário

1 Introdução

2 GNW e GWT

3 GNW/Hopfield

4 Ignição

5 Fases e Consciência

A Teoria da Informação Integrada (IIT) e o modelo GNW - nova abordagem. Próxima palestra.

Notícia Ciência—Global. 26 de agosto de 2025

CIÊNCIA | GLOBAL

Onde nasce a consciência? A questão que opõe neurocientistas

26/08/2025

Pesquisa conduzida por sete anos analisou as duas principais teses sobre o tema. Resultados ampliam disputa sobre se máquinas e animais, por exemplo, teriam experiências conscientes.

Ciência Global. Notícia

O Desafio da Consciência

- A compreensão da consciência continua a ser um dos maiores desafios da neurociência.
- Distinção crucial (David Chalmers):
 - **Easy Problems:** Como o cérebro processa informação, integra dados sensoriais, gera comportamento.
 - **Hard Problem:** Como e por que surge a experiência subjetiva (os *qualia*), o "sentir" algo, a partir de processos físicos?
- O presente trabalho foca-se nos mecanismos e processos que podem levar à emergência de estados de consciência.

Teorias Computacionais da Consciência (I)

- **Global Workspace Theory (GWT) - Bernard Baars:**
 - **Metáfora do Teatro:** A consciência é um "palco" onde informações inconscientes competem por acesso. A informação selecionada torna-se globalmente acessível e difundida para múltiplos processadores inconscientes.
 - **Foco:** Disponibilidade global da informação.
- **Global Neuronal Workspace (GNW) - Stanislas Dehaene & Jean-Pierre Changeux:**
 - **Neuronalização da GWT:** Transforma a metáfora do teatro numa arquitetura neuronal específica (e.g., córtex pré-frontal e parietal).
 - **Conceito Chave:** "Ignição Neuronal" - explosão de atividade sincronizada que amplifica e distribui a informação, tornando-a consciente.

Teorias Computacionais da Consciência (II)

■ Teoria da Informação Integrada (IIT) - Giulio Tononi:

- **Postulado Central:** Consciência é informação integrada.
- **Princípios:** A experiência consciente é **estruturada** (diferenciada e integrada) e **causalmente eficaz**.
- **Métrica:** **Phi (Φ)** quantifica a quantidade de informação integrada (causalidade intrínseca) que um sistema gera.
- **Conexão:** Um sistema com $\Phi > 0$ possui consciência, e a qualidade da experiência é determinada pela estrutura causal que gera esse Φ .

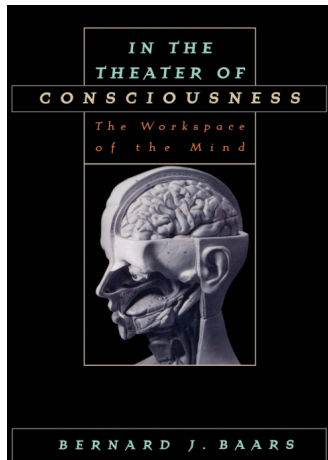
■ Teorias de Ordem Superior (HOTs):

- Um estado mental de primeira ordem torna-se consciente por ser objeto de um **estado mental de ordem superior** (e.g., um pensamento ou percepção sobre o estado original).

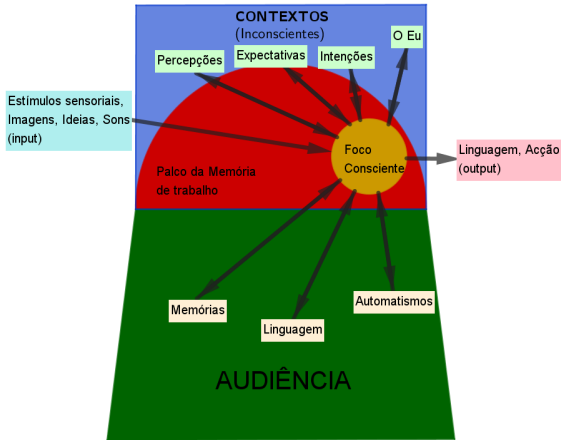
Objetivos do Trabalho: Uma Abordagem Integrada

- Formalizar o modelo GNW com redes de Hopfield estocásticas acopladas.
- Investigar o papel da estocasticidade, capacidade de armazenamento e acoplamento na *dinâmica da ignição*.
 - **Vantagens:** Robustez ao ruído, exploração eficiente do espaço de estados, realismo biológico.
- Formalizar a **dinâmica de Ignição** no GNW como um processo **guiado por padrões** e orquestrado por **campos externos**.
- Investigar a relação entre os **mecanismos do GNW** (estocasticidade, acoplamento, plasticidade) e a **emergência da informação integrada (Φ)**.
- Investigar **Transições de Fase** e a emergência de Consciência.

Precursor: A Metáfora do Teatro de Bernard Baars (GWT)



Precursor: A Metáfora do Teatro de Bernard Baars (GWT)



Precursor: A Metáfora do Teatro de Bernard Baars (GWT)

- Proposta por Bernard Baars nos anos 80.
- Consciência como um "teatro mental" onde:
 - Processamento inconsciente compete para "aceder ao palco central".
 - Informações seleccionadas tornam-se globalmente acessíveis ao sistema cognitivo.
- **Conceito-Chave:** Uma informação amplificada e difundida por toda a rede cerebral (via "ignição" neuronal) provoca a sincronização de diversos processos.
- O aspeto crítico da consciência é a **disponibilidade global** da informação.

Metáfora do Teatro: Componentes

- **O Palco (Memória de Trabalho):** Espaço limitado para pensamentos, imagens, sensações que estão conscientes num dado momento.
- **Os Atores (Conteúdos Mentais):** Pensamentos, sensações, memórias que competem pelo acesso ao palco. A qualidade da atuação determina a seleção.
- **O Holofote da Atenção (Foco Seletivo):** Ilumina certos atores, tornando-os o foco da nossa consciência.
- **A Audiência (Processadores Inconscientes):** Vasta gama de processos mentais inconscientes que recebem a informação difundida do palco.
- **A Equipa nos Bastidores (Contexto Inconsciente):** Expectativas, crenças que influenciam a experiência consciente.
- **O Realizador (Funções Executivas):** Controla e coordena as atividades no palco (supervisão).

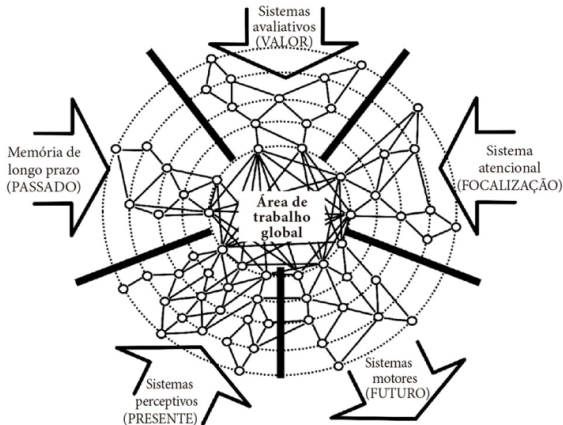
A man with a bald head and glasses, wearing a dark suit jacket over a blue and white striped shirt, is smiling and holding a large, colorful, segmented brain model. The brain model is composed of many small, rounded segments in shades of blue, green, and yellow, with a prominent orange segment on the right side. The background is a plain, light gray.



Jean-Pierre Changeux (1936 -) é um neurobiologista francês. Em 2010 foi eleito membro da Académie royale des sciences, des lettres et des beaux-arts de Belgique.

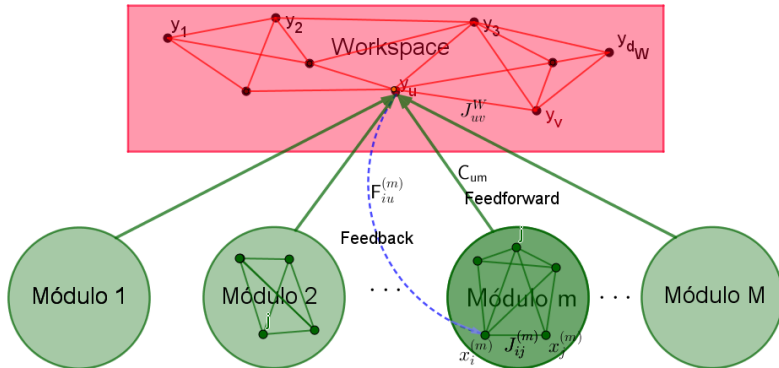
Mente e Consciência: Análise de um Modelo GNW Baseado em Redes de Hopfield Estocásticas

Do GWT ao GNW: A Neuronalização



Mente e Consciência: Análise de um Modelo GNW Baseado em Redes de Hopfield Estocásticas

A Estrutura do GNW



Mente e Consciência: Análise de um Modelo GNW Baseado em Redes de Hopfield Estocásticas

Principais inovações deste trabalho

- Modelar o GNW com **Redes de Hopfield** estocásticas.
- Modificação da dinâmica estocástica da rede

$$\text{GNW} = \{\text{Módulos}\} \cup \{\text{Workspace}\}$$

introduzindo energias modificadas pela introdução de campos externos, que permitem uma dinâmica "guiada" pelos padrões armazenados.

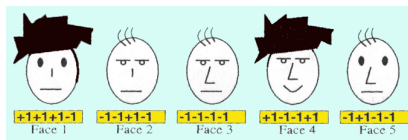
- Análise das Transições de Fase e da emergência da Consciência.
- Adaptação e aplicação da IIT para quantificar a integração de informação no GNW, relacionando Φ com os parâmetros de ordem do sistema (parte 2).

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ 🔍 ↺

Redes de Hopfield

Tipo de rede neuronal recorrente conhecida pela capacidade de armazenar e recuperar **padrões** de forma estável.

Atributos	+1	-1
$x_1 = \text{Cabelo}$		
$x_2 = \text{Olhos}$		
$x_3 = \text{Nariz}$		
$x_4 = \text{Boca}$		



Redes de Hopfield

face	x1	x2	x3	x4
1	1	1	1	-1
2	-1	-1	1	-1
3	-1	-1	-1	-1
4	1	-1	-1	1
5	-1	1	-1	-1

				
+1+1+1-1	-1-1+1-1	-1-1-1-1	+1-1-1+1	-1+1-1-1
Face 1	Face 2	Face 3	Face 4	Face 5

J_12	J_13	J_14	J_23	J_24	J_34
0,25	0,25	0,75	0,25	-0,25	-0,25

Pesos pela Regra de Hebb J_{ij}				
	1	2	3	4
1	0	0,25	0,25	0,75
2	0,25	0	0,25	-0,25
3	0,25	0,25	0	-0,25
4	0,75	-0,25	-0,25	0

Redes de Hopfield

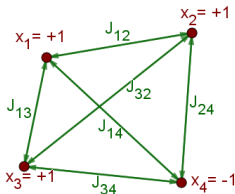
$$\xi^1 = (+1, +1, +1, -1)$$

$$\xi^2 = (-1, -1, +1, -1)$$

$$\xi^3 = (-1, -1, -1, -1)$$

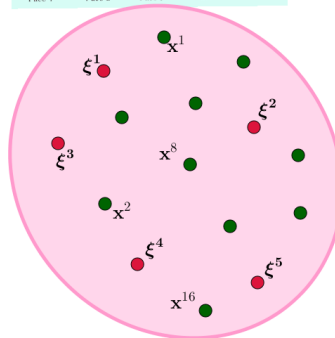
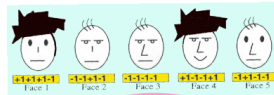
$$\xi^4 = (+1, -1, -1, +1)$$

$$\xi^5 = (-1, +1, -1, -1)$$



$$J_{12} = \frac{1}{4} \sum_{\mu=1}^5 \xi_1^{(\mu)} \xi_2^{(\mu)} = +0.25$$

$$J_{14} = \frac{1}{4} \sum_{\mu=1}^5 \xi_1^{(\mu)} \xi_4^{(\mu)} = 0.75, \text{ etc.}$$



S = Espaço de estados

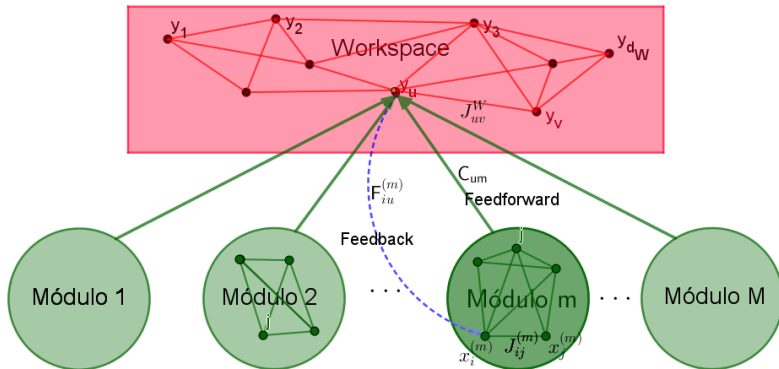
$2^4 = 16$ estados

O Objetivo do Modelo GNW Híbrido

Criar um sistema que capte:

- Existência de **Módulos Locais** especializados.
- **Workspace Global** que integra sinais distribuídos.
- Dinâmica da **ignição** (transições all-or-none e limiares críticos).
- Noção de **difusão** (broadcasting) e acessibilidade global.
- **Plasticidade sináptica**, fenómenos de atenção e de aprendizagem.

Visão Geral do GNW



Visão Geral do GNW

- O GNW é um sistema acoplado de redes de Hopfield estocásticas.
- Consiste em M **Módulos Locais** e um **Workspace** central.
- Cada subsistema (módulo ou Workspace) é modelado como uma rede de Hopfield estocástica.
- **Estado Global:** $x(t) = (\{x^{(m)}(t)\}_{m=1}^M, y_W(t))$.

Componentes: Módulos de Hopfield

- Cada módulo $m \in \{1, \dots, M\}$ possui N_m neurónios.
- **Estado:** $x^{(m)}(t) = (x^{(1)}(t), \dots, x^{N_m}(t))$, com $x_i^{(m)} \in \{-1, +1\}$.
- **Padrões Memorizados:** $\{\xi^{(m,\mu)}\}_{\mu=1}^{p_m}$ (vetores binários).
 - Representam "matéria-prima" cognitiva (e.g., características sensoriais, conceitos de baixo nível).
 - Atuam como guias para a dinâmica do módulo.
- **Matriz de Pesos:** $J^{(m)}$, calculada pela regra de Hebb:

$$J_{ij}^{(m)} = \frac{1}{N_m} \sum_{\mu=1}^{p_m} \xi_i^{(m,\mu)} \xi_j^{(m,\mu)} \quad (i \neq j)$$

- **Energia Clássica:**

$$E_m(x^{(m)}(t)) = -\frac{1}{2} \sum_{i \neq j} J_{ij}^{(m)} x_i^{(m)}(t) x_j^{(m)}(t)$$

Componentes: workspace ($\mathcal{W}$)

- Possui N_W neurónios.
- **Estado:** $y(t) = (y_1(t), \dots, y_{N_W}(t))$, com $y_u(t) \in \{-1, +1\}$.
- **Padrões Memorizados:** $\{\zeta^{(\kappa)}\}_{\kappa=1}^{P_W}$ (vetores binários).
 - Representam **conceitos de alto nível**, esquemas integrativos, metas ou intenções.
 - São os **alvos da ignição** e guiam a integração.
- **Matriz de Pesos:** J^W , calculada pela regra de Hebb de forma análoga.
- **Energia Clássica:**

$$E_W(y(t)) = -\frac{1}{2} \sum_{u \neq v} J_{uv}^W y_u(t) y_v(t)$$

Conectividades Inter-subsistemas

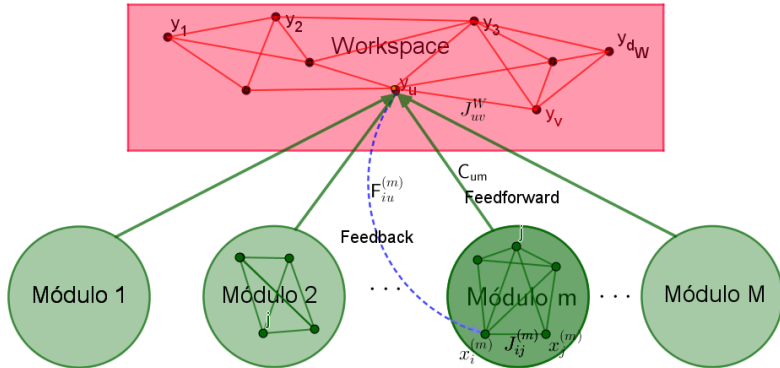
■ Feedforward (Módulos → Workspace):

- C_{um} : Conecta o **módulo m** , **como um todo**, ao neurónio u do workspace.
- Representa o fluxo de informação "bottom-up" (do específico ao geral).

■ Feedback (Workspace → Módulos):

- $F_{iu}^{(m)}$: Conecta o neurónio u do Workspace ao neurónio i do módulo m .
 - Representa o fluxo de informação "top-down" (do geral ao específico).
- A força de acoplamento entre módulos e Workspace é controlada por λ .

Conectividades Inter-subsistemas



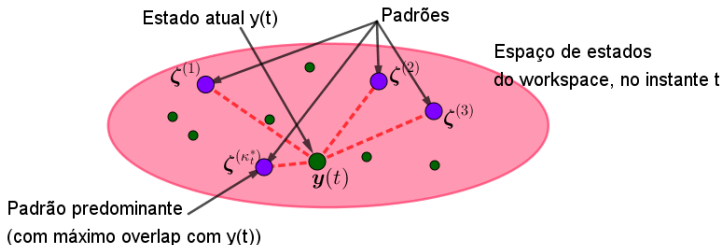
Mecanismos de Competição Modular

A hipótese é que os padrões memorizados, dentro de cada subsistema (módulos e workspace), influenciam ativamente o fluxo de informação, a tomada de decisão (competição modular) e a formação de estados integrados (ignição do workspace).

Padrões e Overlaps. Notações.

- Padrões do módulo m: $\{\xi^{(\mu)}\}_{\mu=1}^{p_m}$.
- Padrões do workspace: $\{\zeta^{(\kappa)}\}_{\kappa=1}^{p_W}$.
- Overlaps:
 - $Ov_m^{(\mu)}(t) = \text{Overlap} \left(x^{(m)}(t), \xi^{(\mu)} \right), \quad \mu = 1, 2, \dots, p_m$
 - $Ov_W^{(\kappa)}(t) = \text{Overlap} \left(y(t), \zeta^{(\kappa)} \right), \quad \kappa = 1, 2, \dots, p_W$
 - $o_m(t) = \max_{\mu} \{Ov_m^{(\mu)}(t)\}_{\mu=1}^{p_m}$
 - $o_W(t) = \max_{\kappa} \{Ov_W^{(\kappa)}(t)\}_{\kappa=1}^{p_W}$
 - $\zeta_W^{(\kappa_t^*)}$ padrão predominante no workspace.
 - $\kappa_t^* = \arg \max_{\kappa} \text{Overlap} \left(y(t), \zeta^{(\kappa)} \right)$

Padrões Memorizados no Workspace



$$\zeta^{(\kappa_t^*)} = (-1, +1, +1, +1, -1, -1, +1, +1)$$

$$y(t) = (+1, +1, +1, +1, -1, -1, -1, -1)$$

$$\text{Ov}_W^{(\kappa_t^*)} = -1 + 1 + 1 + 1 + 1 + 1 - 1 - 1 = 2/8$$

$$o_W(t) = \max_{\kappa=1, \dots, 4} \{\text{Ov}_W^{(\kappa)}\} = \text{Ov}_W^{(\kappa_t^*)} = 2/8$$

Competição Lateral Inter-Módulos

- O Overlap máximo do estado atual do módulo m com os seus padrões:

$$o_m(t) = \max_{\mu} \text{Overlap} \left(x^{(m)}(t), \xi^{(m,\mu)} \right)$$

- Mede a relevância do módulo.
- **Competição Lateral Inter-Módulos** ($\text{Comp}_m(t)$): Mede a "força relativa" do módulo m . É a base para o MECANISMO DE ATENÇÃO SELETIVA entre os módulos do GNW

$$\text{Comp}_m(t) = o_m(t) - \frac{1}{M-1} \sum_{n \neq m} o_n(t)$$

- A competição premeia os módulos que estão mais "claros" (com alto overlap com os seus padrões).

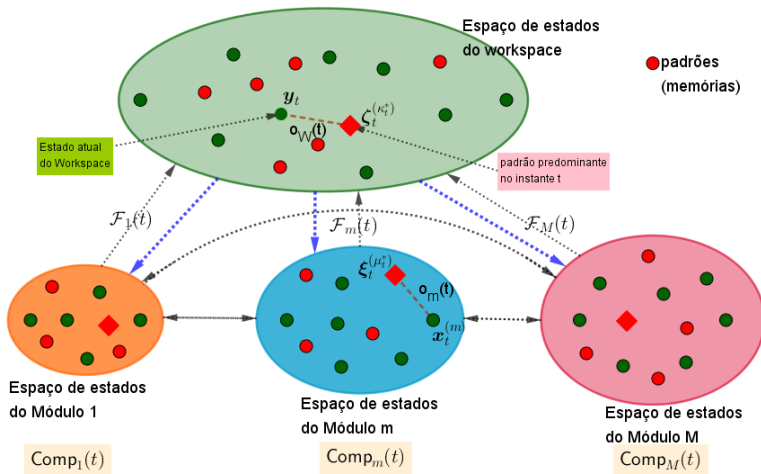
Mecanismos de Competição Modular (Suas Definições)

- **Função de Competição** ($\mathcal{F}_m(t)$): Normaliza e aplica "ganho seletivo".

$$\mathcal{F}_m(t) = \frac{e^{\beta_s \text{Comp}_m(t)}}{\sum_{n=1}^M e^{\beta_s \text{Comp}_n(t)}}$$

- são todos positivos e somam 1.
- β_s controla a seletividade.
- Um $\mathcal{F}_m(t)$ alto, indica prioridade para o Workspace.

Mecanismos de Competição Modular



Dinâmica Guiada por Padrões e Campos Externos

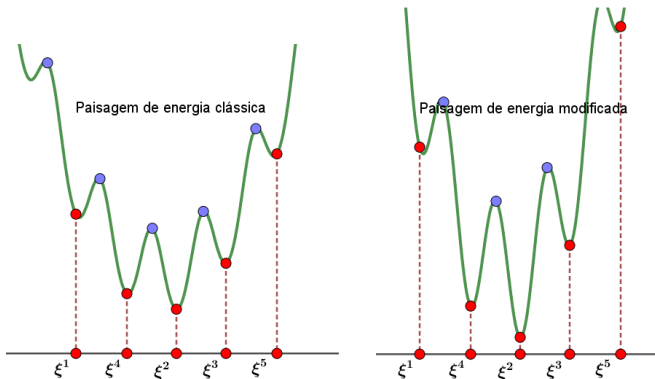
- Os padrões memorizados atuam como **guias ativos** para a dinâmica do GNW.
- As interações entre subsistemas são mediadas por **Campos Externos** (h) adicionados às Energias.
- Estes campos representam:
 - **Atenção seletiva** (módulos $\rightarrow$ Workspace).
 - **Direcionamento interno** (Workspace para si mesmo).
 - **Controle top-down** (Workspace $\rightarrow$ módulos).

Campos externos

O objetivo é adicionar "campos externos" à energia de cada subsistema, de tal forma que a dinâmica seja influenciada pelos padrões memorizados.

- Os campos externos são os veículos pelos quais a interação dinâmica e a modulação ativa ocorrem no GNW. Eles traduzem a "intenção" (padrões ativos), a "atenção" (competição modular) e a "coordenação" (feedback) das forças que direcionam a dinâmica de Metropolis de cada subsistema.

Paisagem de energia



Campo Externo no Workspace ($h_{W,u}^{\text{mod}}(t)$)

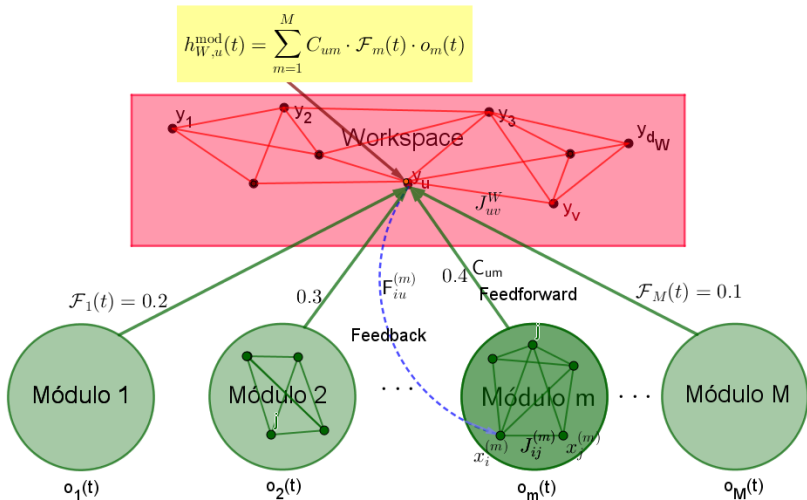
Recebe influências de duas fontes principais:

1. Input Modulado dos Módulos (Feedforward):

$$h_{W,u}^{\text{mod}}(t) = \sum_{m=1}^M C_{um} \cdot \mathcal{F}_m(t) \cdot o_m(t)$$

- Informação agregada "*bottom-up*" de todos os módulos para o neurónio u do Workspace, modulada pela função de competição $\mathcal{F}_m(t)$:
- Implementa a **atenção seletiva "bottom-up"**

Campo Externo no Workspace ($h_{W;u}^{\text{mod}}(t)$)



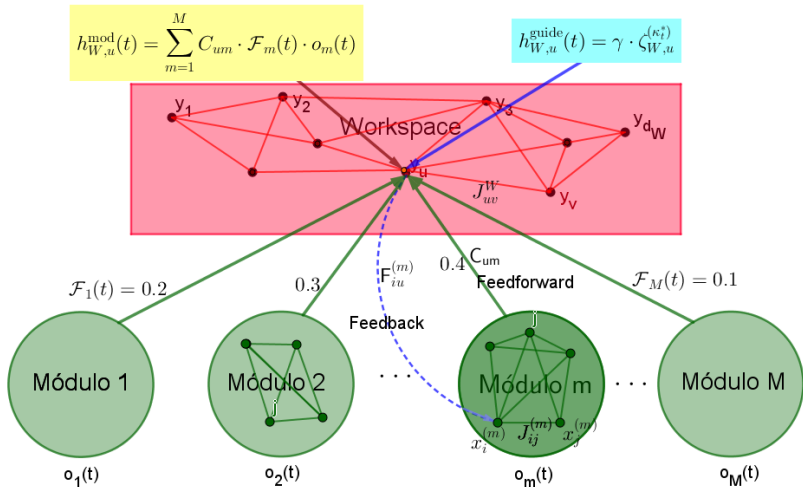
Campo Externo no Workspace ($h_{W,u}^{\text{guide}}(t)$)

2. Atração Ativa para o Padrão predominante do Workspace:

$$h_{W,u}^{\text{guide}}(t) = \gamma \cdot \zeta_{W,u}^{(\kappa_t^*)}, \quad \gamma : \text{força de atração.}$$

- $\zeta_{W,u}^{(\kappa_t^*)}$: neurónio u do padrão dominante no Workspace (κ_t^* é o índice do padrão com maior overlap $Ov_W(t)$).
- Guia ativamente o workspace para o padrão predominante.

Campo Externo no Workspace ($h_{W,u}^{\text{guide}}(t)$)



Energia Modificada do Workspace

A Energia total do Workspace é:

$$E_W^{\text{mod}}(y(t)) = -\frac{1}{2} \sum_{u \neq v} J_{uv}^W y_u(t) y_v(t) - \sum_{u=1}^{N_W} \left(h_{W,u}^{\text{mod}}(t) + h_{W,u}^{\text{guide}}(t) \right) y_u(t)$$

- A Energia clássica é modificada pela adição dos campos externos.
- Esta função de energia é utilizada no Método de Metropolis para a atualização do estado do Workspace.

Campo Externo nos Módulos ($h_{m,i}(t)$)

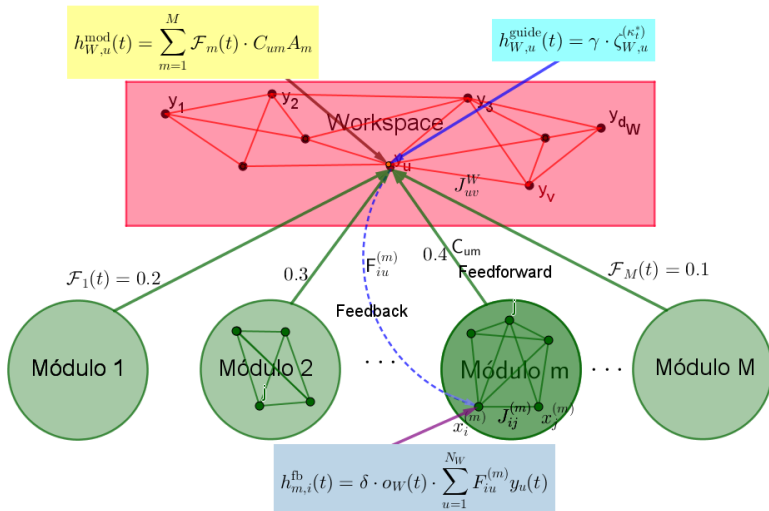
O neurónio i do módulo m recebe feedback do Workspace:

- **Feedback Dirigido por Padrões do Workspace**
(Top-Down) - $h_{m,i}^{\text{fb}}(t)$:

$$h_{m,i}^{\text{fb}}(t) = \delta \cdot o_w(t) \cdot \sum_{u=1}^{N_W} F_{iu}^{(m)} y_u(t)$$

- δ : força do feedback.
- $o_w(t)$ é o overlap máximo do workspace. Atua como um modulador da força do feedback. Se o workspace está num estado claro e coerente com um dos seus padrões ($o_w(t)$ alto), o feedback é mais forte.
- $\sum_{u=1}^{N_W} F_{iu}^{(m)} y_u(t)$: sinal bruto do Workspace para o neurónio i do módulo m .
- Direciona a atividade modular para ser consistente com o conceito integrado no Workspace.

Campo Externo nos Módulos ($h_{m,i}(t)$)



Energia Modificada dos Módulos

A Energia de cada módulo m é:

$$E_m^{\text{mod}}(x^{(m)}(t)) = -\frac{1}{2} \sum_{i \neq j} J_{ij}^{(m)} x_i^{(m)}(t) x_j^{(m)}(t) \\ - \sum_{i=1}^{N_m} h_{m,i}^{\text{fb}}(t) x_i^{(m)}(t)$$

- O campo de feedback $h_{m,i}^{\text{fb}}(t)$ é utilizado no Método de Metropolis para a atualização do estado dos módulos.

Padrões Memorizados no Workspace e Ignição

- **Ignição do Workspace**, $lg_W(t)$: Transição do Workspace para um estado coerente e significativo.

$$lg_W(t) = 1 \iff (o_W(t) > \theta \text{ e } \forall \tau \in [t, t + \Delta t_s], o_W(\tau) > \theta')$$

- θ, θ' : limiares de overlap. Δt_s : duração mínima.
- A ignição é ativamente dirigida pelos padrões (via $h_{W,u}^{\text{guide}}$) e pelos módulos (via $h_{W,u}^{\text{mod}}$).

Relação Formal entre Ignição e Φ

- A **Ignição do Workspace** é o correlato físico-dinâmico de um estado de alta integração de informação.
- **Hipótese Central:**

$$\lg(t) = 1 \implies \Phi(t) \text{ alto}$$

- A ignição implica:
 - **Diferenciação Aumentada:** Convergência para padrão específico.
 - **Informação Altamente Integrada:** Interdependência causal forte via feedforward e feedback.
 - **Causalidade Efetiva Forte:** Atração robusta a atrator (padrão).
- Todos estes fatores contribuem para um **Φ elevado**.

Parâmetros de Ordem para o GNW

- **Parâmetros de Ordem:** Quantidades macroscópicas que descrevem o grau de "ordem" de um sistema.
- **Overlap com Padrões nos módulos m:**

- Para o módulo m:

$$o_m(t) = \max_{\mu} \text{Overlap} \left(x^{(m)}(t), \xi^{(m,\mu)} \right)$$

- Para o workspace:

$$o_W(t) = \max_{\kappa} \text{Overlap} \left(y(t), \zeta^{(\kappa)} \right)$$

- Um $\langle o \rangle$ alto ($\langle o_m \rangle$ ou $\langle o_W \rangle$) (próximo de 1) indica **fase ordenada** (recuperação de padrão).

Parâmetros de Ordem para o GNW

■ Overlapps quadráticos (q):

- $q_m(t) = \frac{1}{N_m} \sum_{i=1}^{N_m} \langle x_i^{(m)} \rangle_{\text{temp}}^2$. Mede quão "estáveis" ou "polarizados" estão os neurónios individuais do módulo m . ao longo do tempo
- $q_W(t) = \frac{1}{N_W} \sum_{u=1}^{N_W} \langle y_u \rangle_{\text{temp}}^2$. Mesma interpretação.
- Um q alto (q_m ou q_W) (com o baixo) indica **fase de confusão** (atratores espúrios).
- Um q baixo (com m baixo) indica **fase desordenada** (caos).

Consciência e as Fases do GNW.

- A hipótese central é que a consciência (ou os estados de cognição integrada) corresponde a uma fase ordenada do GNW, ativamente modulada pelas suas interações.

A emergência da consciência no GNW é caracterizada pela operação do sistema na FASE ORDENADA, onde os padrões memorizados são robustamente recuperados e integrados.

Esta fase é distintamente caracterizada por $\langle o_m \rangle, \langle o_W \rangle \gg 0$ e $q_m, q_W \gg 0$ para o Workspace e módulos relevantes, e é ativamente modulada pela dinâmica do GNW.

Capacidade de Armazenamento ($\mathcal{C} = p/N$)

- $\mathcal{C}$ é um parâmetro de controle fundamental que induz **transições de fase**.
- **Fase Ordenada (Clareza/Consciência):**
 - Condição: $\mathcal{C} < \mathcal{C}_c$.
 - Comportamento: Recupera padrões robustamente.
 - Parâmetros de Ordem: $\langle o \rangle \gg 0$.
- **Fase de Confusão (Sobrecarga/Indecisão):**
 - Condição: $\mathcal{C}_c < \mathcal{C} < \mathcal{C}_L$.
 - Comportamento: Atratores espúrios, não padrões puros.
 - Parâmetros de Ordem: $\langle o \rangle \approx 0$ e $q \gg 0$.

Obrigado pela vossa atenção!

OBRIGADO

Pré-print CMUP:

<https://www.cmup.pt/bibcite/reference/8037>

