COLLECTIONELSS MACHINE LEARNING
THE HAMILTONIAN FRAMEWORK

Marco Gori
Università di Siena

MESS-2024

OUTLINE

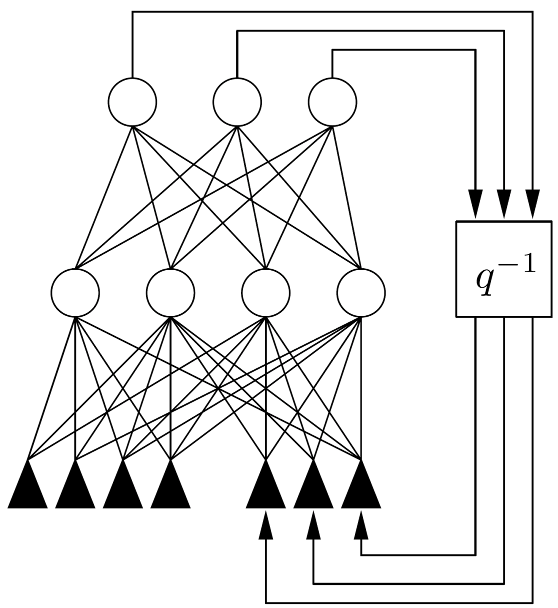
Part I

1. The path of Artificial Intelligence
2. A paradigm-shift: The connectionist wave
3. Collectionless AI
4. Intelligence and laws of Nature

Part II

5. The Hamiltonian framework of learning
6. Cognidynamics: A Theory of Neural propagation
7. Neural vs wave propagation

LEARNING IN THE TEMPORAL DIMENSION



Time-continuous model

$$\dot{x}(t) = f(x(t), w(t), u(t))$$

$$J_{[0,T]} = J_T + \int_0^T dt\, L(x(t), w(t), u(t))$$

$$w^\star = \arg\min_w J_{[0,T]}(w)$$

birth: the single reset   neither training nor test!   death t

0 **environmental interactions** T

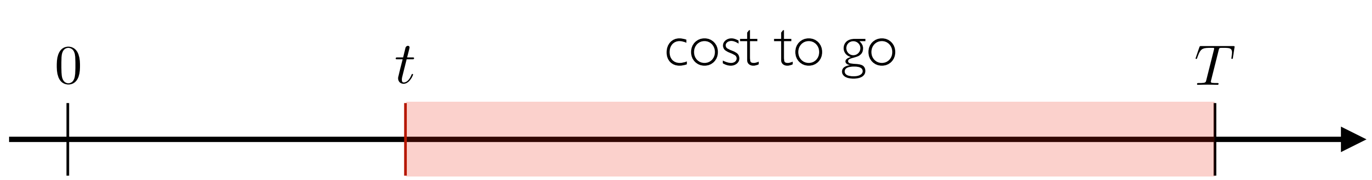
Genetic inheritance?

How big the lifespan is supposed to be? Virtually infinite?

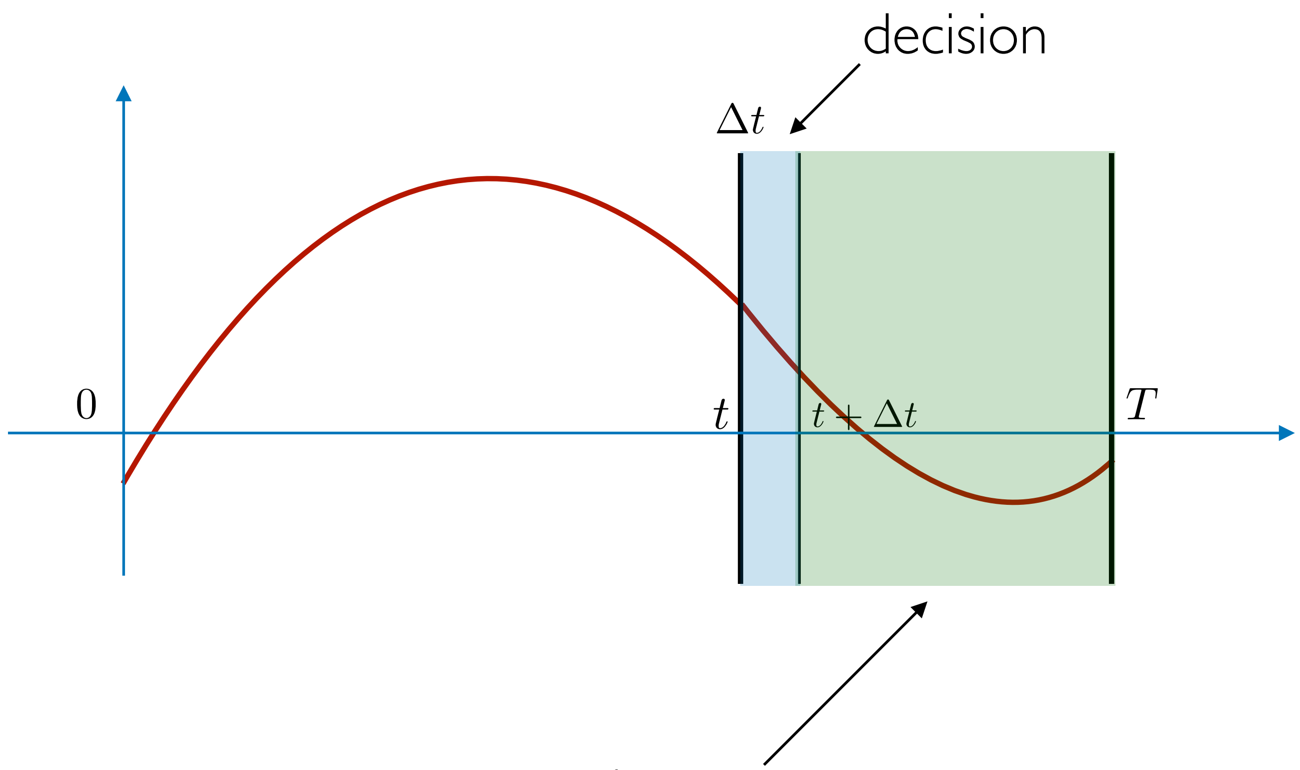
VALUE FUNCTION

$$\textit{Value Function } V : [0,T] \times \mathcal{X} \to \mathbb{R} : \ (t,\xi) \mapsto V(t,x)$$

$$V(t,x) := J_T + \min_w \int_t^T ds\ L(\xi(s), w(s), s)$$



USING BELLMAN' PRINCIPLE



$$V(t, x^\star) = \min_{w([t,T])} \left(V(t + \Delta t, x + \Delta x) + \int_t^{t+\Delta t} ds\ L(x(s), w(s), s) \right)$$

$$\begin{aligned} V(t, x^\star) &= V(t + \Delta t, x^\star + \Delta x^\star) + \min_{w([t,t+\Delta t])} L(x(t), w(t), t)\Delta t + o(\Delta t) \\ &= V(t, x^\star) + V_s(t, x^\star)\Delta t + V_x(t, x^\star)\Delta x^\star + o(\Delta x^\star) + o(\Delta t) \\ &+ \min_{w([t,t+\Delta t])} L(x(t), w(t), t)\Delta t, \end{aligned}$$

HJB EQUATIONS

$$V(t, x^\star) = V(t + \Delta t, x^\star + \Delta x^\star) + \min_{w([t,t+\Delta t])} L(x(t), w(t), t)\Delta t + o(\Delta t)$$

$$= V(t, x^\star) + V_s(t, x^\star)\Delta t + V_x(t, x^\star)\Delta x^\star + o(\Delta x^\star) + o(\Delta t)$$

$$+ \min_{w([t,t+\Delta t])} L(x(t), w(t), t)\Delta t,$$

$$\dot{x}(t)\Delta t = f(x^\star, w^\star, t)\Delta t$$

$$o(\Delta t) = V_x(t, x^\star) \cdot f(x^\star, w^\star, t)\Delta t + V_s(t, x^\star(t))\Delta t + \min_{w([t,t+\Delta t])} L(x(t), w(t), t)\Delta t$$

$$V_s(t, x^\star) = -\min_{\omega} \Big(L(x^\star, \omega, t) + V_x(t, x^\star) \cdot f(x^\star, \omega, t) \Big)$$

HAMILTONIAN AND HJB EQUATIONS

$$H(x, p, s) := \min_{\omega} \big(L(x, \omega, s) + p \cdot f(x, \omega, s) \big)$$ Hamiltonian

$$V_s(t, x^\star) = -\min_{\omega} \left(L(x^\star, \omega, t) + V_x(t, x^\star) \cdot f(x^\star, \omega, t) \right)$$

Partial Differential Equations

unknown $V(t, x)$

$$V_s(t, x^\star) + H(x^\star, V_x(t, x^\star), t) = 0$$

$V(T, x) = g(x)$ terminal condition

HJ(B) EQUATIONS AND METHOD OF CHARACTERISTICS

HAMILTONIAN DYNAMICS IS SUFFICIENT

Let us consider the following (HJ) initial-point problem

$$(\mathrm{HJ}) \qquad \begin{cases} V_s(t,x) + H(x, V_x(t,x,t) = 0. \\ V(0,x) = g(x). \end{cases}$$

CAUCHY PROBLEM: GOING FORWARD

We want to convert this PDE problem into an ODE that can open a dramatically different computational perspective. We use the method of characteristic. Now, let us introduce the *co-state* p as $p := V_x$ and consider the total derivative[18] of its κ coordinate

HJ(B) EQUATIONS AND METHOD OF CHARACTERISTICS

co-state p as $p := V_x$

How does it evolve?

$$\dot{p}^{\kappa}(t) := \dot{p}_{x_\kappa}(t) = V_{x_\kappa t}(t, x(t)) + V_{x_\kappa x_i} \cdot \dot{x}_i.$$

Now, if V solves (HJ) then

$$V_{x_\kappa t}(x,t) = -H_{x_\kappa}(x, V_x(x,t), t) - H_{p_i}(x, V_x(x,t), t) \cdot V_{x_i x_\kappa}(x,t)$$

$$\begin{aligned}\dot{p}^{\kappa}(t) = &-H_{x_\kappa}(x(t), \underbrace{V_x(x(t),t)}_{p(t)}, t) \\ &+ \left(\dot{x}_i(t) - H_{p_i}(x(t), \underbrace{V_x(x(t),t)}_{p(t)}, t)\right) \cdot V_{x_\kappa x_i}(t, x(t))\end{aligned}$$

HJB EQUATIONS AND METHOD OF CHARACTERISTICS (CON'T)

$$\dot{p}^{\kappa}(t) = -H_{x_\kappa}(x(t), \underbrace{V_x(x(t),t)}_{p(t)}, t)$$
$$+ \big(\dot{x}_i(t) - H_{p_i}(x(t), \underbrace{V_x(x(t),t)}_{p(t)}, t)\big) \cdot V_{x_\kappa x_i}(t, x(t))$$

Now we can promptly see that the following choice

(H) $$\begin{cases} \dot{x}(t) = H_p(x(t), p(t), t) \\ \dot{p}(t) = -H_x(x(t), p(t), t) \end{cases}$$

↓

(HJ) $$\begin{cases} V_s(t,x) + H(x, V_x(t,x,t) = 0. \\ V(0,x) = g(x). \end{cases}$$

NON-HOLONOMIC CONSTRAINTS

$$J_{[0,T]}(w) = J_T + \int_0^T dt\ L(x(t), w(t), t)$$

$$\dot{x}(t) = f(x(t), w(t), t)$$

LAGRANGIAN APPROACH

$$J_L = J_T + \int_0^T dt \Big(L\big(x(t), w(t), t\big) + \lambda(t) \cdot (f(x(t), w(t), t)) - \dot{x}(t) \Big)$$

$$\mathcal{H}\big(x(t), \lambda(t), w(t), t\big) := L\big(x(t), w(t), t\big) + \lambda(t) \cdot f(x(t), w(t), t).$$

$$J_L = J_T + \int_0^T dt \Big(\underbrace{\mathcal{H}(x(t), \lambda(t), w(t), t) - \lambda(t) \cdot \dot{x}(t)}_{\mathcal{L}^x} \Big)$$

A CLASSIC "TRICK"

$$\int_0^T dt\ \lambda(t)\cdot\dot{x}(t) = \Big[\lambda(t)\cdot x(t)\Big]_0^T - \int_0^T dt\ \dot{\lambda}(t)\cdot x(t)$$

$$J_L = J_T + \int_0^T dt\Big(\underbrace{\mathcal{H}(x(t),\lambda(t),w(t),t) - \lambda(t)\cdot\dot{x}(t)}_{\mathcal{L}^x}\Big) \longleftarrow$$

$$J_L(x,\lambda) = J_T - \Big[x(t)\cdot\lambda(t)\Big]_0^T + \int_0^T dt\big(\underbrace{\mathcal{H}\big(x(t),\lambda(t),w(t),t\big) + x(t)\cdot\dot{\lambda}(t)}_{\mathcal{L}^\lambda}\big) \longleftarrow$$

two different ways of writing the functional

EULER LAGRANGE EQUATIONS

$$0 = \frac{d}{dt}\mathcal{L}^x_{\dot{x}} - \mathcal{L}^x_x \rightarrow \quad \dot{\lambda}(t) + \mathcal{H}_x(x(t), \lambda(t), w(t), t) = 0$$

$$0 = \frac{d}{dt}\mathcal{L}^\lambda_{\dot{\lambda}} - \mathcal{L}^\lambda_\lambda \rightarrow \quad \dot{x}(t) - \mathcal{H}_\lambda(x(t), \lambda(t), w(t), t) = 0$$

$$0 = \frac{d}{dt}\mathcal{L}^\lambda_{\dot{w}} - \mathcal{L}^\lambda_w \rightarrow \quad \mathcal{H}_x(x(t), \lambda(t), w(t), t) = 0.$$

$$H(x, \lambda, t) = \min_w \mathcal{H}(x, \lambda, w, t).$$

Finally, this leads to the Hamiltonian equations

$$\begin{cases} \dot{\lambda}(t) = -H_x(x(t), \lambda(t), t) \\ \dot{x}(t) = -\mathcal{H}_\lambda(x(t), \lambda(t), t). \end{cases}$$

CHARACTERISTIC EQUATIONS OF HJB HAMILTONIAN "LAWS"

$$\begin{cases} \dot{x}(t) & = H_p(x(t), p(t), u(t), t) = \mathrm{f}(x(t), w(t), u(t)) \\ \dot{p}(t) & = -H_x(x(t), p(t), u(t), t), \end{cases}$$

$$x(0) = x_0 \ \ and \ p(T) = p_T = V_x(T, x(T))$$

CLASSIC CASE OF LINEAR QUADRATIC (LQ) CONTROL

LINEAR QUADRATIC (LQ) CONTROL

$$\dot{x} = Ax + Bw$$

$$L(x, w, t) = \frac{1}{2}x'Qx + \frac{1}{2}w'Rw$$

$$\begin{aligned} w^\star &= \min_w \left(\frac{1}{2}x'Qx + \frac{1}{2}w'Rw + p'(Ax + Bw) \right) \\ &= -R^{-1}B'p = -R^{-1}B'Px := Fx. \end{aligned}$$

feedback control

$$\dot{x} = (A + BF)x$$

THE HAMILTONIAN

$$w^\star = \min_w \left(\frac{1}{2} x' Q x + \frac{1}{2} w' R w + p'(Ax + Bw) \right)$$
$$= -R^{-1} B' p = -R^{-1} B' P x := F x.$$

$$
\begin{aligned}
H(x,p,w)\big|_{w^\star} &= \frac{1}{2} x' Q x + \left[\frac{1}{2} w' R w + p' \cdot \big(Ax + Bw \big) \right]_{w^\star} \\
&= \frac{1}{2} x' Q x + \frac{1}{2} (R^{-1} B' p)' R (R^{-1} B' p) - p' \cdot \big(Ax + B(R^{-1} B' p) \big) \\
&= \frac{1}{2} x' Q x + \frac{1}{2} p' \underbrace{B R^{-1} B'}_{S} p - p' \cdot \big(Ax + B R^{-1} B' p \big) \\
&= \frac{1}{2} x' Q x - \frac{1}{2} p' S p + p' A x
\end{aligned}
$$

SOLVING HJB EQUATIONS

$$V(t,x) = \frac{1}{2}x'P(t)x$$

Let's assume a quadratic function

$$V_t + H(x, V_x) = 0$$

terminal condition $V(T,x) = g(x)$

$$\frac{1}{2}x'\dot{P}x + \frac{1}{2}x'Qx - \frac{1}{2}p'Sp + p'Ax = \frac{1}{2}\cancel{x}'\dot{P}\cancel{x} + \frac{1}{2}\cancel{x}'Q\cancel{x} - \frac{1}{2}\cancel{x}'P'SP\cancel{x} + \cancel{x}'P'A\cancel{x} = 0$$

$$A'P \rightsquigarrow a_{\kappa i}p_{\kappa j} = p_{j\kappa}a_{\kappa i} \rightsquigarrow PA$$

Riccati equation

$$\dot{P} + Q + PA + A'P - PSP = 0$$

SOLVING HJB EQUATIONS

$$V_t + H(x, V_x) = 0$$ terminal condition $$V(T, x) = g(x)$$

We solve

$$\dot{P} + Q + PA + A'P - PSP = 0$$

$$V(t, x) = \frac{1}{2} x' P(t) x$$

solution of the PDE

We find the control law

$$\begin{aligned} w^\star &= \min_w \left(\frac{1}{2} x' Q x + \frac{1}{2} w' R w + p'(Ax + Bw) \right) \\ &= -R^{-1} B' p = -R^{-1} B' P x := F x. \end{aligned}$$

ASYMPTOTIC STABILITY

$$\dot{x} = Ax + Bw$$

$$\dot{x} = (A + BF)x$$ feedback control

$$W(t) = \frac{1}{2}x'(t)\bar{P}x(t)$$ Lyapunov function

$$\begin{aligned} \dot{W}(t) &= \dot{x}'\bar{P}x(t) = x'\bar{P}(A - BR^{-1}B'\bar{P})x(t) \\ &= \frac{1}{2}x'(t)\Big(\bar{P}(A - BR^{-1}B'\bar{P}) + (A' - \bar{P}BR^{-1}B')\bar{P}\Big)x(t) \end{aligned}$$

$$\begin{aligned} &\bar{P}(A - BR^{-1}B'\bar{P}) + (A' - \bar{P}BR^{-1}B')\bar{P} \\ &= \bar{P}A + A'\bar{P} - 2\bar{P}S\bar{P} = \underbrace{Q + \bar{P}A + A'\bar{P} - \bar{P}S\bar{P}}_{0} - Q - \bar{P}S\bar{P} = -Q - \bar{P}S\bar{P} \end{aligned}$$

ASYMPTOTIC STABILITY (CON'T)

$$\dot{W}(t) = \dot{x}'\bar{P}x(t) = x'\bar{P}(A - BR^{-1}B'\bar{P})x(t)$$
$$= \frac{1}{2}x'(t)\Big(\bar{P}(A - BR^{-1}B'\bar{P}) + (A' - \bar{P}BR^{-1}B')\bar{P}\Big)x(t)$$

$$\bar{P}(A - BR^{-1}B'\bar{P}) + (A' - \bar{P}BR^{-1}B')\bar{P}$$
$$= \bar{P}A + A'\bar{P} - 2\bar{P}S\bar{P} = \underbrace{Q + \bar{P}A + A'\bar{P} - \bar{P}S\bar{P}}_{0} - Q - \bar{P}S\bar{P} = -Q - \bar{P}S\bar{P}$$

Riccati's equation

$$\dot{W}(t) = -\frac{1}{2}x'\big(Q + \bar{P}S\bar{P}\big)x \leq 0$$

$Q + \bar{P}S\bar{P} \geq 0$ $\quad Q > 0, R > 0.$ asymptotic stability $\bar{P}$

The "magic" of asymptotic stability: we need to solve Riccati's equation

LQ: HAMILTONIAN EQUATIONS

$$w^\star = \min_w \left(\frac{1}{2} x' Q x + \frac{1}{2} w' R w + p'(Ax + Bw) \right)$$
$$= -R^{-1} B' p = -R^{-1} B' P x := F x.$$

$$\dot{x} = (A + BF) x$$

$$\begin{aligned} H(x,p,w)\big|_{w^\star} &= \frac{1}{2} x' Q x + \left[\frac{1}{2} w' R w + p' \cdot \big(Ax + Bw\big) \right]_{w^\star} \\ &= \frac{1}{2} x' Q x + \frac{1}{2} (R^{-1} B' p)' R (R^{-1} B' p) - p' \cdot \big(Ax + B(R^{-1} B' p)\big) \\ &= \frac{1}{2} x' Q x + \frac{1}{2} p' \underbrace{B R^{-1} B'}_{S} p - p' \cdot \big(Ax + B R^{-1} B' p\big) \\ &= \frac{1}{2} x' Q x - \frac{1}{2} p' S p + p' A x \end{aligned}$$

LQ HAMILTONIAN EQUATIONS

$$\dot{x} = Ax + Bw = Ax - \underbrace{BR^{-1}B}_{S}p$$

$$\dot{p} = -Qx - A'p.$$

$$\dot{\begin{pmatrix} x \\ p \end{pmatrix}} = \begin{pmatrix} A & -S \\ -Q & -A' \end{pmatrix} \cdot \begin{pmatrix} x \\ p \end{pmatrix}$$

1-Dim

$$\begin{pmatrix} a & -s \\ -q & -a \end{pmatrix} \qquad \det \begin{pmatrix} \rho - a & -s \\ -q & \rho + a \end{pmatrix} = 0$$

positive eigenvalues … we need of the crystal ball!

$$(\rho^2 - a^2) - qs = 0 \rightarrow \rho = \pm\sqrt{a^2 + qs}$$

HAMILTON and RICCATI EQUATIONS

When considering the circuital assumption $p = Px$ we get $\dot{p} = \dot{P}x + P\dot{x}$. From the state equation $P\dot{x} = PAx - PSPx$ and, therefore,

$$\dot{p} = -Qx - A'Px = \dot{P}x + PAx - PSPx$$

That is, for any x:

$$Qx + A'Px + \dot{P}x + PAx - PSPx = 0 \rightarrow \dot{P} + Q + A'P + PA - PSP = 0.$$

$$\dot{P} + Q + A'P + PA - PSP = 0$$

it cannot be solved "forward in time"!

H AND HJB EQUATIONS:

CAN WE FIND THE VALUE FUNCTION? WHY IS QUADRATIC?

$$V_t + H(x, V_x) = V_t + \frac{1}{2}x'Qx - \frac{1}{2}p'Sp + p'Ax = 0$$ Hamiltonian

$$\dot{x} = Ax + Bw = Ax - \underbrace{BR^{-1}B}_{S}p$$

$$\dot{p} = -Qx - A'p.$$

$$p'\dot{x} - x'\dot{p} = p'Ax - p'Sp + x'Qx + x'A'p \leftarrow \begin{cases} p'\dot{x} = p'Ax - p'Sp \\ x'\dot{p} = -x'Qx - x'A'p \end{cases}$$

$$V_t = \tfrac{1}{2}\left(p'\dot{x} - x'\dot{p}\right)$$

Since $p(t) = P(x)x(t)$ we get

$$V_t + \frac{1}{2}\left(x'P\dot{x} - x'\left(\dot{P}x + P\dot{x}\right)\right) = V_t - \frac{1}{2}x'\dot{P}(t)x' = 0$$

and, finally

$$V(t, x) = \frac{1}{2}x'(t)\int_0^t ds\dot{P}(s)x(t) = \frac{1}{2}x'(t)P(t)x(t)$$

TO SUM UP

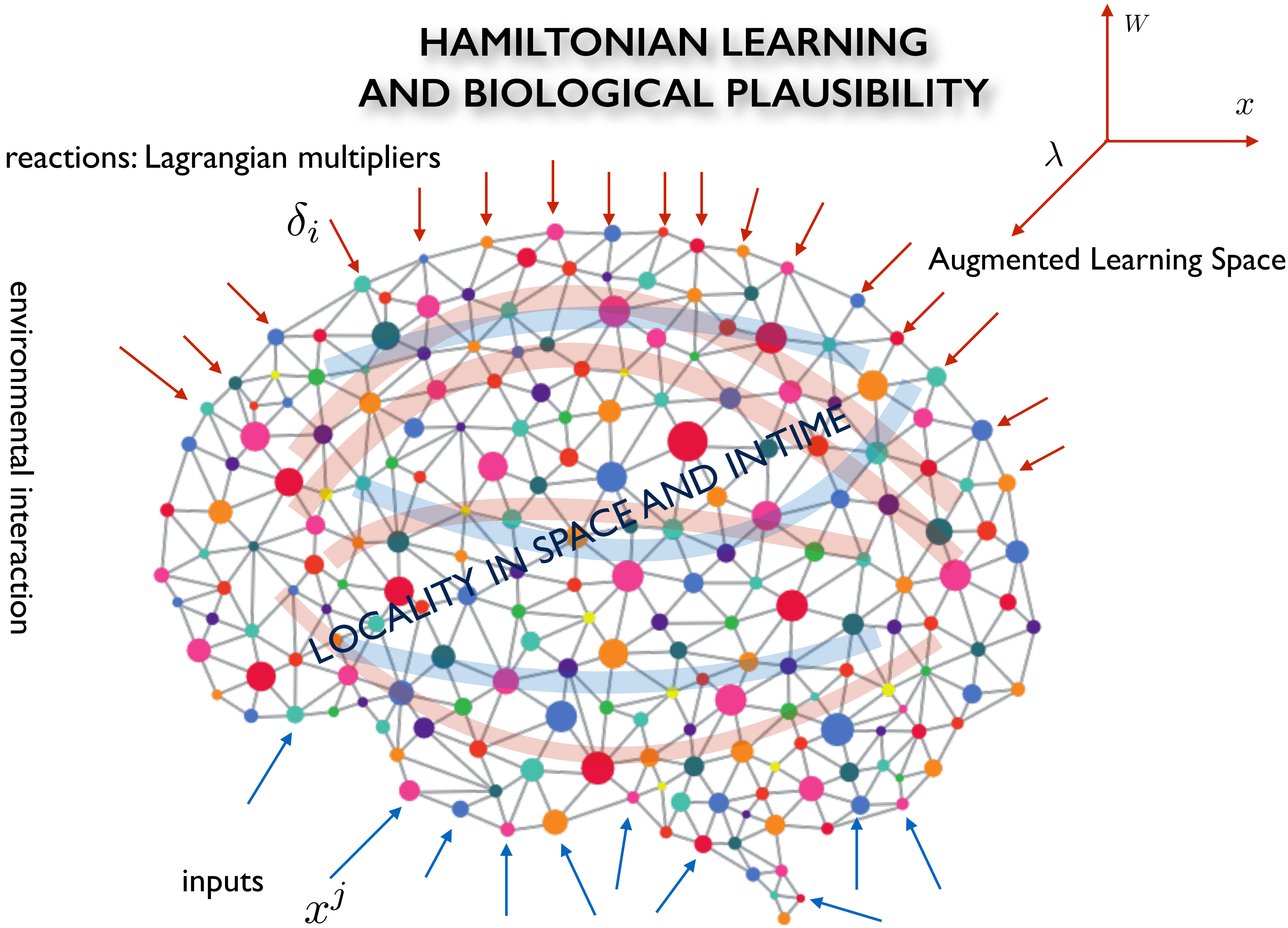
- HJB: necessary and SUFFICIENT conditions!
- H equations are characteristic for the HJ PDE
- Links with Lagrangian approach - Pontryagin's Maximum Principle - PMP)
- The perspective of H Learning

COGNIDYNAMICS:
A THEORY OF NEURAL PROPAGATION

"Life can only be understood backwards; but it must be lived forwards."

Søren Kierkegaard

HAMILTONIAN LEARNING AND BIOLOGICAL PLAUSIBILITY



LEARNING IN RECURRENT NETS

$$\mathcal{N}: \begin{cases} \xi_i(t) = u_i(t) & i \in \mathcal{I} \\ \dot{\xi}_i(t) = \alpha_i(t)\Big[-\xi_i(t) + \sigma\Big(\sum_{j\in\mathcal{V}} \mathrm{w}_{ij}(t)\xi_j(t)\Big)\Big] & i \in \bar{\mathcal{V}} \\ \dot{w}_{ij}(t) = \psi_{ij}(t)\nu_{ij}(t) & (i,j) \in \mathcal{A} \\ \mathrm{w}_{ij}(t) = \omega_{ij}(t) w_{ij}(t) & (i,j) \in \mathcal{A} \end{cases}$$

$$x \sim [\xi_i, w_{ip}]$$

environmental interaction

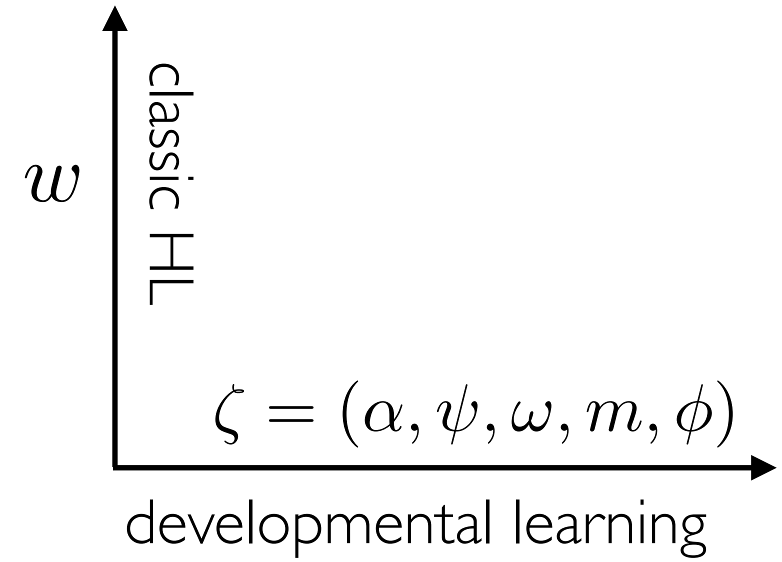
dissipative parameters

$$R(\nu, T) = \int_0^T \left(\frac{1}{2}\sum_{i\in\bar{\mathcal{V}}}\sum_{j\in\mathcal{V}} m_{ij}(s)\nu_{ij}^2(s) + \gamma \sum_{i\in\mathcal{O}} \phi_i(s) V(\xi_i(s), s)\right) ds$$

kinetic energy

potential energy

THE HAMILTONIAN

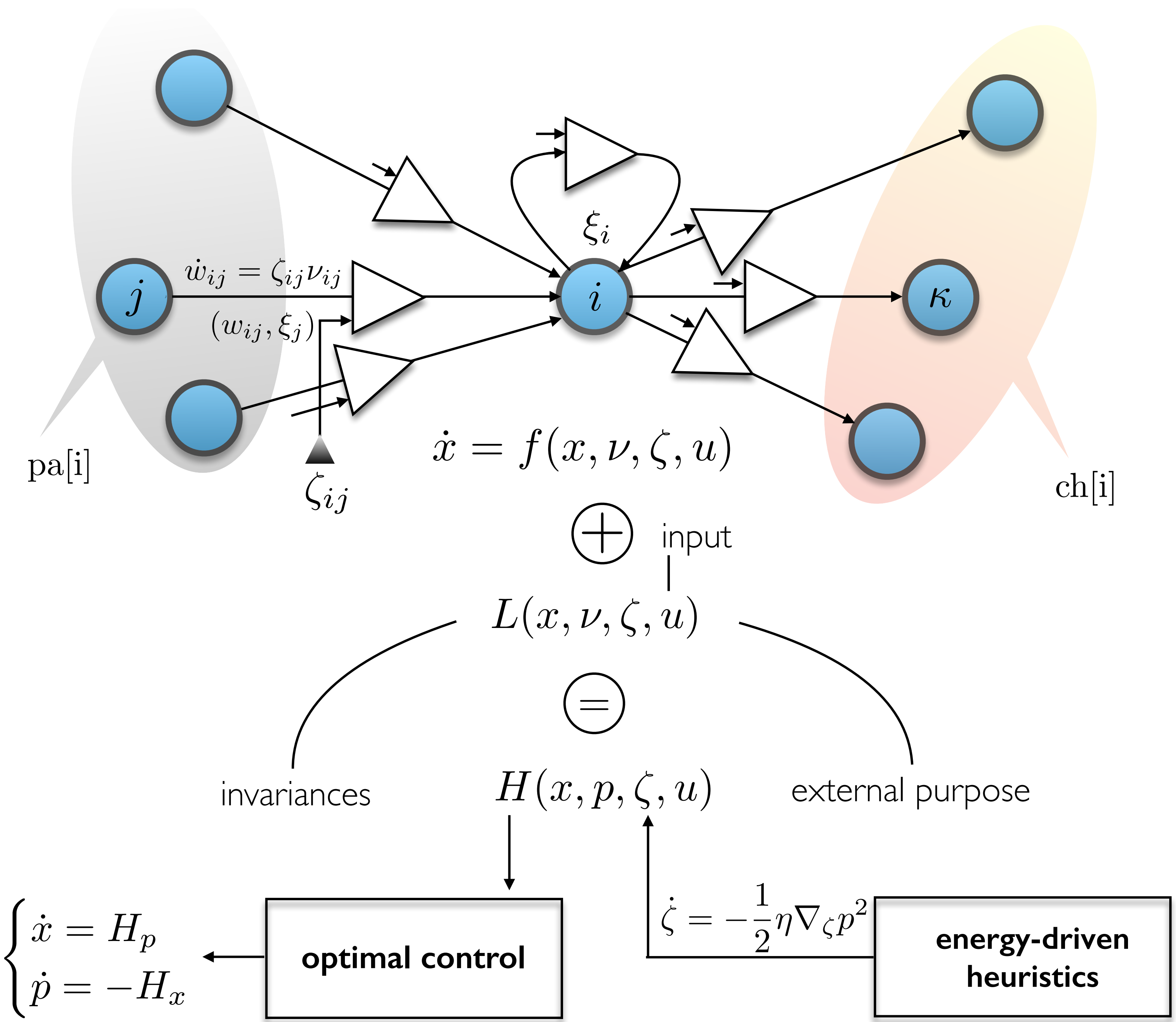


$$
\begin{aligned}
&H(\xi_i, w_{ij}, p_i, p_{ij}, t) \\
&= \min_v \Bigg(\frac{1}{2} \sum\nolimits_{ij} m_{ij}(t)\nu_{ij}^2(t) + \gamma\phi_i(t) V\big(\xi_i, t\big) \\
&\quad + \sum_i \alpha_i(t) p_i(t) \Big[-\xi_i(t) + \sigma\Big(\sum\nolimits_j \omega_{ij}(t) w_{ij}(t) \xi_j(t) \Big) \Big] \\
&\quad + \sum_{ij} p_{ij}(t) \psi_{ij}(t) \nu_{ij}(t) \Bigg).
\end{aligned}
\qquad \beta_{ij}(t) := \frac{\psi_{ij}^2(t)}{m_{ij}(t)}
$$

$$
\begin{aligned}
H = &-\frac{1}{2} \sum\nolimits_{ij \in \mathcal{A}} \beta_{ij}(t) p_{ij}^2(t) + \gamma \sum\nolimits_{i \in \mathcal{O}} \phi_i(t) V\big(\xi(t), t\big) \\
&+ \sum\nolimits_{i \in \bar{\mathcal{V}}} \alpha_i(t) p_i(t) \Big[-\xi_i(t) + \sigma\Big(\sum\nolimits_{j \in \mathcal{V}} \omega_{ij}(t) w_{ij}(t) \xi_j(t) \Big) \Big]
\end{aligned}
$$

THE HAMILTONIAN (CON'T)

full characterization of learning



$$H = -\frac{1}{2}\sum_{ij\in\mathcal{A}} \beta_{ij}(t)p_{ij}^2(t) + \gamma\sum_{i\in\mathcal{O}} \phi_i(t)V(\xi(t), t)$$
$$+\sum_{i\in\bar{\mathcal{V}}} \alpha_i(t)p_i(t)\Big[-\xi_i(t) + \sigma\Big(\sum_{j\in\mathcal{V}} \omega_{ij}(t)w_{ij}(t)\xi_j(t)\Big)\Big]$$

LEARNING IN THE TEMPORAL DIMENSION

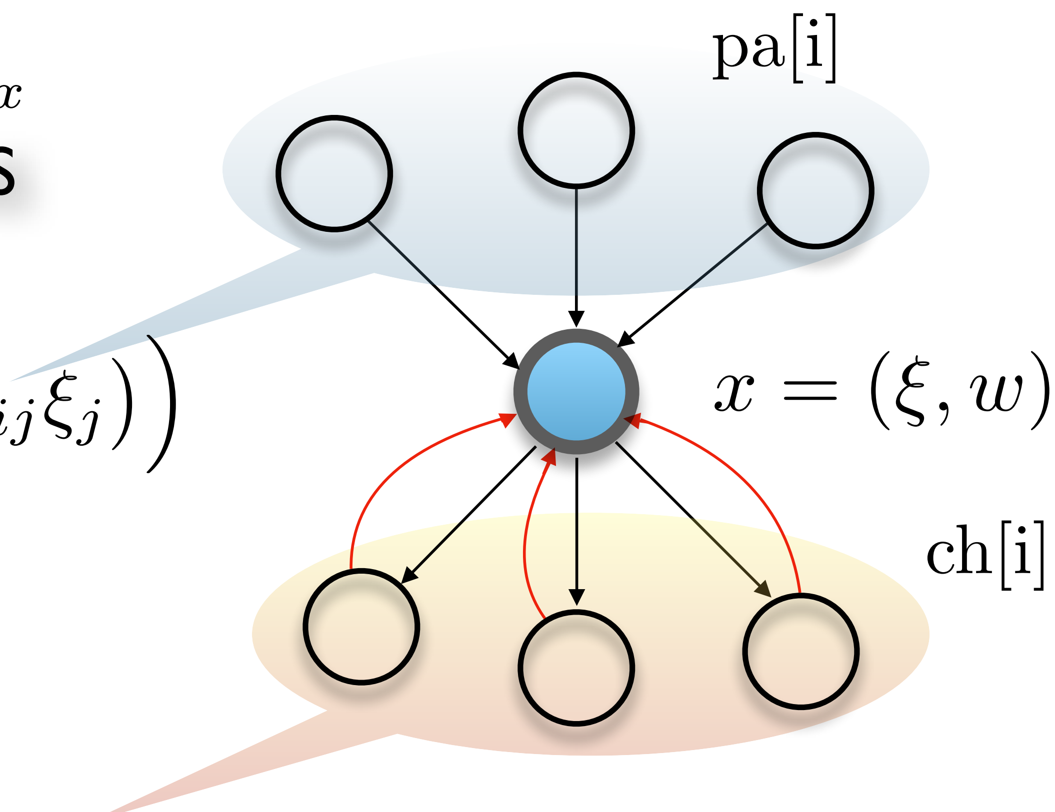
$$\dot{x} = H_p \qquad \dot{p} = -H_x$$

HAMILTONIAN EQUATIONS

$$\dot{\xi}_i = \alpha_i \Big(\xi_i - \sigma \big(\sum_{j \in \mathrm{pa[i]}} \mathrm{w}_{ij} \xi_j \big) \Big)$$

$$\dot{w}_{ij} = \psi_{ij} \nu_{ij}$$

$$\mathrm{w}_{ij} = w_{ij} \omega_{ij}$$



Energy-Driven Heuristics

$$\dot{p}_i = -s_i \alpha_i \phi_i V_{\xi_i} + s_i \alpha_i p_i - s_i \sum_{\kappa \in \mathrm{ch[i]}} \alpha_\kappa \sigma'(a_\kappa) \mathrm{w}_{\kappa i} p_\kappa$$

$$\dot{p}_{ij} = -s_i \alpha_i \omega_{ij} \sigma'(a_i) p_i \xi_j$$

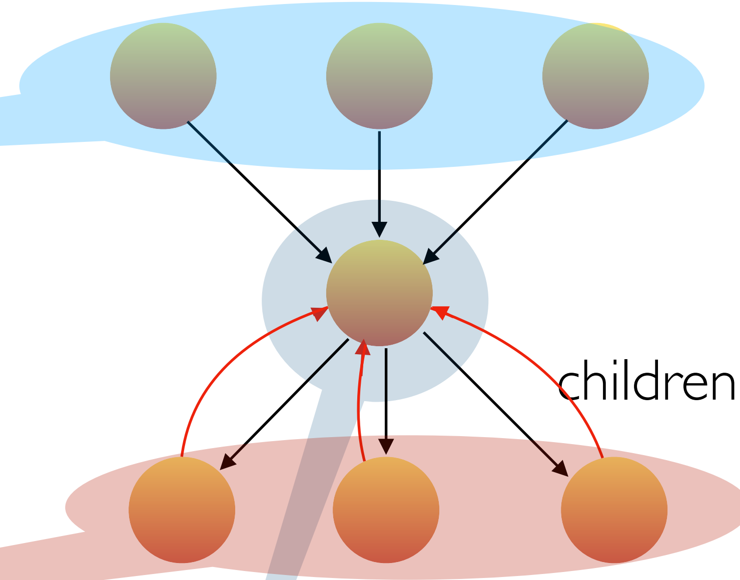
LOCAL SPATIOTEMPORAL PROPAGATION

Forward

$$\dot{\xi}_i(t) = \alpha_i(t)\Big[-\xi_i(t) + \sigma\Big(\textstyle\sum_{j\in\mathcal{V}} \mathrm{w}_{ij}(t)\xi_j(t)\Big)\Big]$$

$$\dot{w}_{ij}(t) = \psi_{ij}(t)\nu_{ij}(t)$$

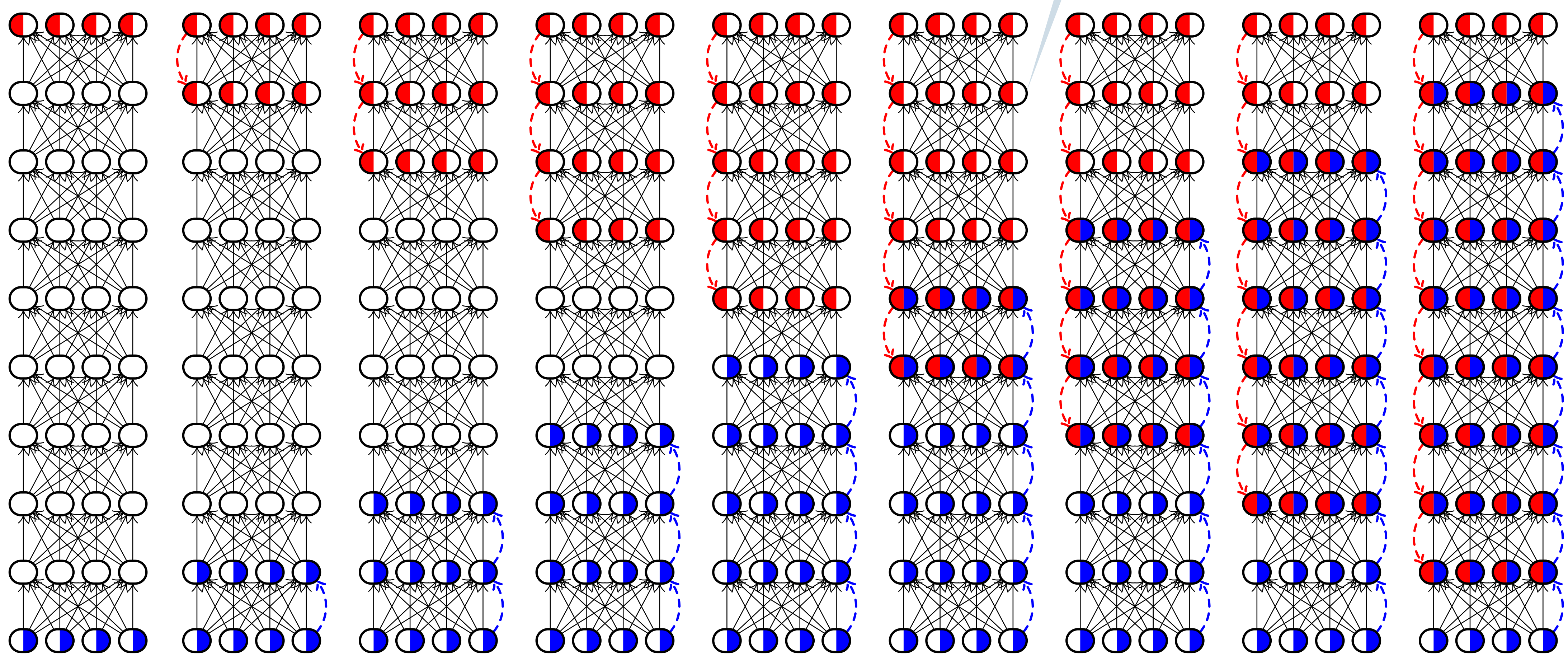
$$\mathrm{w}_{ij}(t) = \omega_{ij}(t)w_{ij}(t)$$



Backward

$$\dot{p}_i(t) = -\,s\gamma\phi_i(t)V_{\xi_i}(\xi(t),t) + s\alpha_i(t)p_i(t)$$
$$-\,s\sum_{\kappa\in\mathrm{ch}[i]}\alpha_\kappa(t)\omega_{\kappa i}(t)\sigma'\Big(\sum_{j\in\mathrm{pa}[\kappa]}\omega_{\kappa j}(t)w_{\kappa j}(t)\xi_j(t)\Big)w_{\kappa i}(t)p_\kappa(t);$$

$$\dot{p}_{ij}(t) = -s\alpha_i(t)\omega_{ij}(t)\sigma'\Big(\sum_m w_{im}(t)\xi_m(t)\Big)p_i(t)\xi_j(t).$$



t $t+1$ $t+2$ $t+3$ $t+4$ $t+5$ $t+6$ $t+7$ $t+8$

GRADIENT-BASED INTERPRETATION OF HAMILTONIAN LEARNING

$$\theta_{ij} := \frac{\dot{m}_{ij}}{m_{ij}} - 2\frac{\dot{\psi}_{ij}}{\psi_{ij}} \qquad \beta_{ij}(t) := \frac{\psi_{ij}^2(t)}{m_{ij}(t)}$$

$$\dot{\beta}_{ij} = \frac{d}{dt}\frac{\psi_{ij}^2}{m_{ij}} = \frac{\psi_{ij}^2}{m_{ij}}\left(2\frac{\dot{\psi}_{ij}}{\psi_{ij}} - \frac{\dot{m}_{ij}}{m_{ij}}\right) = -\frac{\psi_{ij}^2}{m_{ij}}\theta_{ij} = -\beta_{ij}\theta_{ij}$$

$$\beta_{ij}(t) = \beta_{ij}(0) \cdot \exp - \int_0^t \theta_{ij}(s)ds$$

I - gradient-based interpretation

$$\dot{w}_{ij}(t) = -\beta_{ij}(t)p_{ij}(t) \qquad g_{ij} = -\int_0^t \alpha_i \omega_{ij}\sigma'(a_i)p_i\xi_j$$

GRADIENT-BASED INTERPRETATION OF HAMILTONIAN LEARNING

$$\ddot{w}_{ij} + \theta_{ij}\dot{w}_{ij} + \beta_{ij}\dot{p}_{ij} = 0$$

$$\dot{p}_{ij} = -\alpha_i \omega_{ij} \sigma'(a_i) p_i \xi_j$$

$$g_{ij}$$

II - gradient-based interpretation

LOCAL SPATIO TEMPORAL PROPAGATION

Algorithmic Analysis

LSTP

```
for t in range(n-1):
    #
    # Hamilton's statiotemporal propagation
    #
    p_x_dot[t,0] =  q * (z[t] - x[t,0]) * gamma[t]
    for i in range(m):
        x_dot[t,i] = SignFlip[t]*(-x[t,i]+Sigma(f[i]))
        x[t+1,i] = x[t,i] + tau*x_dot[t,i]
        p_b_dot[t,i] = - SignFlip[t]*(DSigma(f[i])*p_x[t,i]*u[t] + r_0w*b[t,i])
        p_b[t+1,i] = p_b[t,i] + tau*p_b_dot[t,i]
        p_x_dot[t,i] = p_x_dot[t,i] + p_x[t,i] - r_0x * x[t,i]
        b_dot[t,i] = -SignFlip[t]*p_b[t,i]/r_w
        b[t+1,i] = b[t,i] + tau*b_dot[t,i]
        for j in range(m):
            f[i] +=  w[t,i,j]*x[t,j]
            p_x_dot[t,i] = p_x_dot[t,i] - DSigma(f[i])*p_x[t,j]*w[t,j,i]
            p_w_dot[t,i,j] = - SignFlip[t]*(DSigma(f[i]) *p_x[t,i]*x[t,j] + r_0w*w[t,i,j])
            p_w[t+1,i,j] = p_w[t,i,j] + tau*p_w_dot[t,i,j]
            w_dot[t,i,j] = -SignFlip[t]*p_w[t,i,j]/r_w
            w[t+1,i,j] = w[t,i,j] + tau*w_dot[t,i,j]
        p_x_dot[t,i] = SignFlip[t]*p_x_dot[t,i]
        p_x[t+1,i] = p_x[t,i] + tau*p_x_dot[t,i]
```

target

response

time

neuron

“forward” H_p

“backward” $-H_x$

connection

ENERGY BALANCE

$$E := \int_0^t \phi_i(\tau) V_s(\xi_i(\tau), \tau)\, d\tau$$ input/output information

$$D := D_\phi + D_\beta + D_\alpha + D_\omega$$

$$D_\phi := -\int_0^t \dot{\phi}_i(\tau) V(\xi_i(\tau), \tau)\, d\tau$$ variation of loss gating

$$D_\beta := \frac{1}{2}\int_0^t \sum_{ij} \dot{\beta}_{ij}(\tau) p_{ij}^2(\tau)\, d\tau$$ variation of weight velocity gating

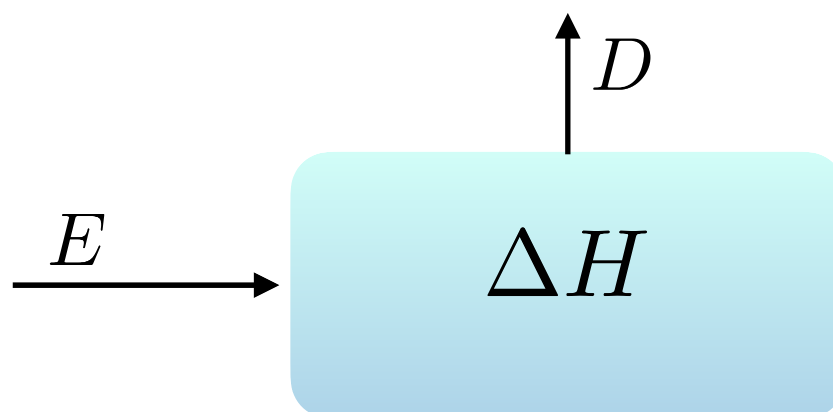
$$D_\alpha := -\int_0^t \sum_{i \in \bar{\mathcal{V}}} \dot{\alpha}_i(\tau) p_i(\tau) \Big[-\xi_i(\tau) + \sigma\big(a_i(\tau)\big) \Big] d\tau$$ variation of activation gating

$$D_\omega := -\int_0^t \sum_{i \in \bar{\mathcal{V}}} \alpha_i(\tau) p_i(\tau) \sigma'\big(a_i(\tau)\big) \sum_j \dot{\omega}_{ij}(\tau) w_{ij}(\tau) \xi_j(\tau)\, d\tau$$ variation of weight gating

ENERGY BALANCE (con't)

I Principle of Cognidynamics

$$E = \Delta H + D$$



All energy term can either be positive or negative!

ENERGY BALANCE (con't)

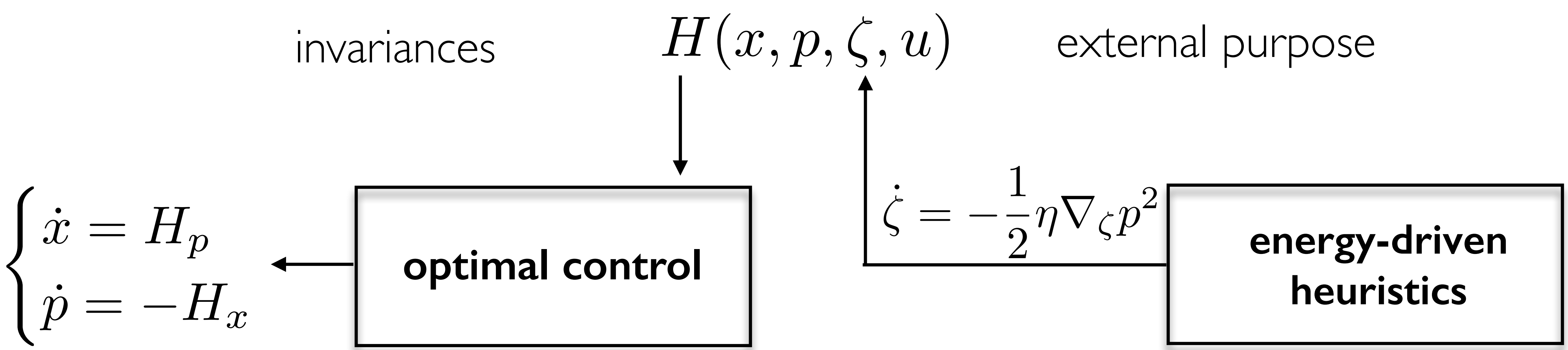
Proof.

$$
\begin{aligned}
H_s\big|_{s=\tau} &= \frac{\partial}{\partial s}\Big(-\frac{1}{2}\sum\nolimits_{ij} \beta_{ij}(s)p_{ij}^2(s) + \sum_{i\in\mathcal{O}} \phi_i(s)V\big(\xi(s), s\big) \\
&\quad + \sum\nolimits_{i\in\bar{\mathcal{V}}} \alpha_i(s)p_i(s)\Big[-\xi_i(s) + \sigma\Big(\sum\nolimits_j \omega_{ij}(s)w_{ij}(s)\xi_j(s)\Big)\Big]\Big)\Big|_{s=\tau} \\
&= -\sum\nolimits_{ij} \dot{\beta}_{ij}(\tau)p_{ij}^2(\tau) + \sum_{i\in\mathcal{O}} \phi_i(\tau)V_s(\xi_i(\tau), \tau) + \dot{\phi}_i(\tau)V(\xi_i(\tau), \tau) \\
&\quad + \sum\nolimits_{i\in\bar{\mathcal{V}}} \dot{\alpha}_i(\tau)p_i(\tau)\Big[-\xi_i(\tau) + \sigma\Big(\sum\nolimits_j \omega_{ij}(\tau)w_{ij}(\tau)\xi_j(\tau)\Big)\Big] \\
&\quad + \sum\nolimits_{i\in\bar{\mathcal{V}}} \alpha_i(\tau)p_i(\tau)\sigma'\big(a_i(\tau)\big)\sum\nolimits_j \dot{\omega}_{ij}(\tau)w_{ij}(\tau)\xi_j(\tau)
\end{aligned}
$$

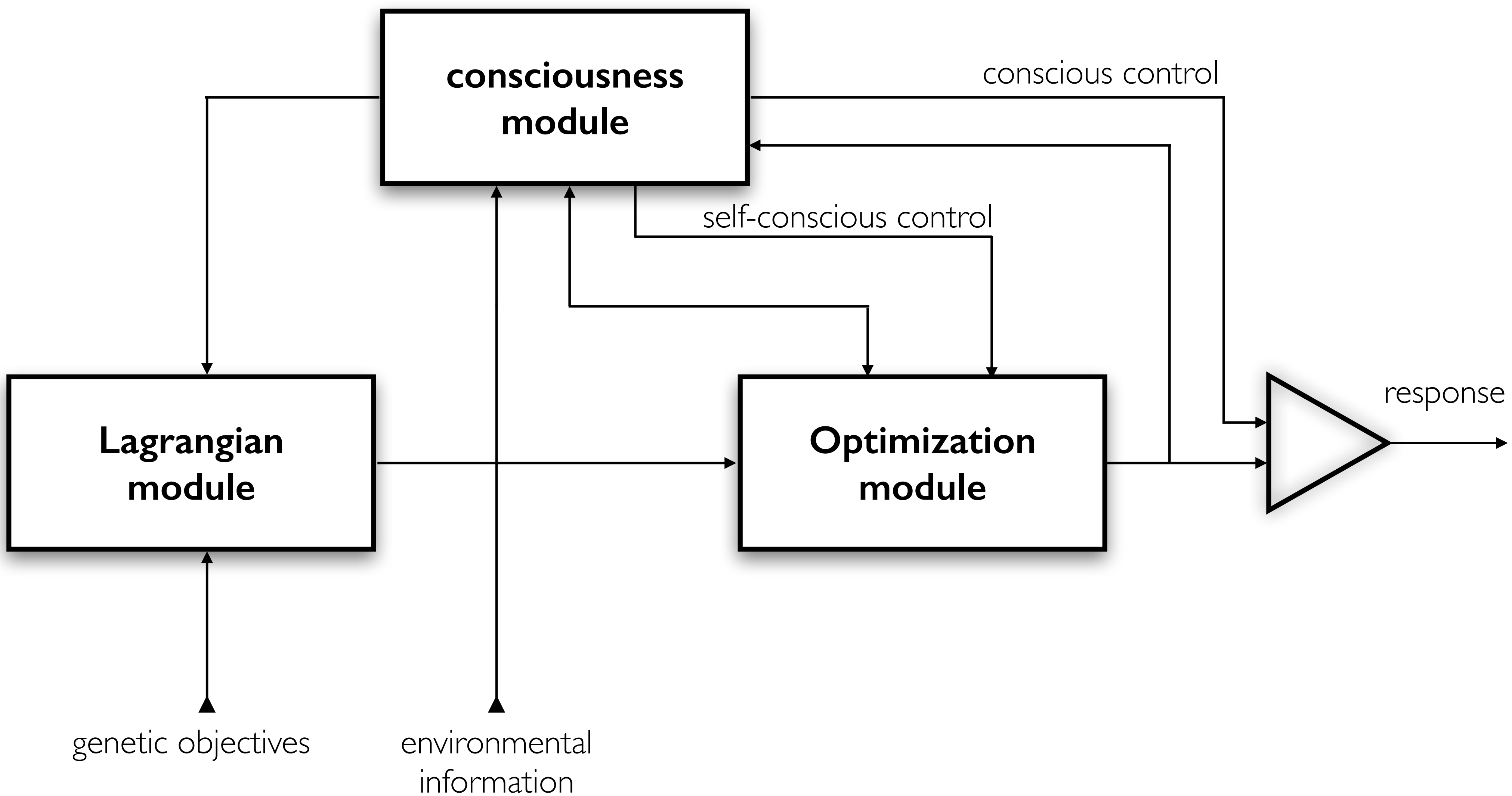
Now, let $\Delta H := H(\xi(t), w(t), p_\xi(t), p_w(t), t) - H(\xi(0), w(0), p_\xi(0), p_w(0), 0)$ be. If we integrate over $[0, t]$ we get

$$
\begin{aligned}
\Delta H = &\int_0^t \sum_{i\in\mathcal{O}} \big(\phi_i(\tau)V_s(\xi_i(\tau), \tau)\big)d\tau && \leftarrow E \\
&+ \int_0^t \dot{\phi}_i(\tau)V(\xi_i(\tau), \tau)d\tau && \leftarrow -D_\phi \\
&- \frac{1}{2}\int_0^t \sum\nolimits_{ij} \dot{\beta}_{ij}(\tau)p_{ij}^2(\tau)d\tau && \leftarrow -D_\beta \\
&+ \int_0^t \sum\nolimits_{i\in\bar{\mathcal{V}}} \dot{\alpha}_i(\tau)p_i(\tau)\Big[-\xi_i(\tau) + \sigma\big(a_i(\tau)\big)\Big]d\tau && \leftarrow -D_\alpha \\
&+ \int_0^t \sum\nolimits_{i\in\bar{\mathcal{V}}} \alpha_i(\tau)p_i(\tau)\sigma'\big(a_i(\tau)\big)\sum\nolimits_j \dot{\omega}_{ij}(\tau)w_{ij}(\tau)\xi_j(\tau)d\tau && \leftarrow -D_\omega
\end{aligned}
$$

DEVELOPMENTAL LEARNING AND ENERGY-DRIVEN HEURISTICS



CONSCIOUSNESS ISSUES



Neural vs Electromagnetic Wave Propagation

WAVE PROPAGATION

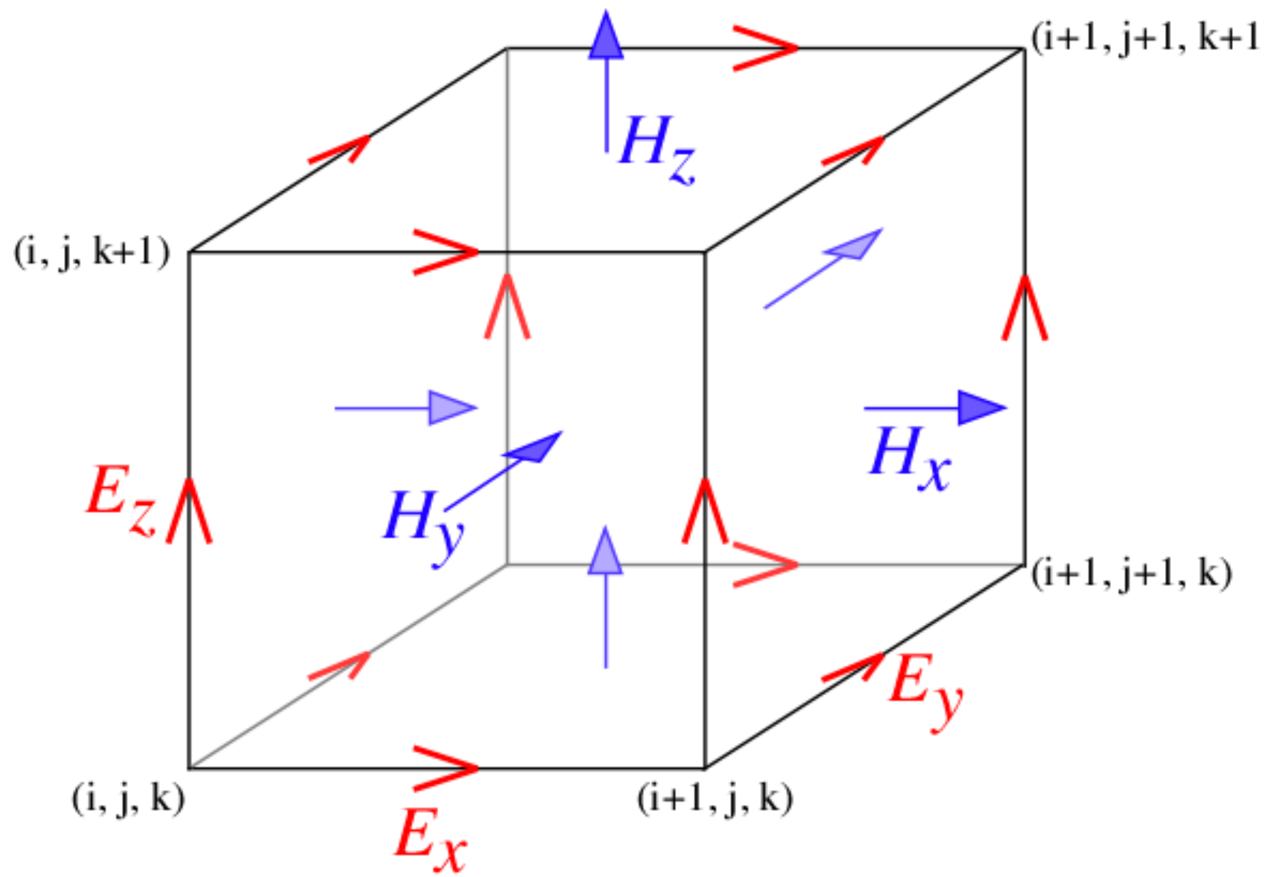
$$\nabla \cdot E = \frac{\rho}{\epsilon}$$

$$\nabla \times B = \mu J + \epsilon\mu\frac{\partial E}{\partial t}$$

$$\nabla \cdot B = 0$$

$$\nabla \times E = -\frac{\partial B}{\partial t}$$

Finite Difference Time Domain (FDTD)



MAXWELL'S EQUATIONS

$$\frac{E_x|^{n+1}_{i,j,k} - E_x|^{n-1}_{i,j,k}}{2\Delta t} = \frac{1}{\varepsilon}\left(\frac{H_z|^{n}_{i,j+1,k} - H_z|^{n}_{i,j-1,k}}{2\Delta y} - \frac{H_y|^{n}_{i,j,k+1} - H_y|^{n}_{i,j,k-1}}{2\Delta z}\right)$$

$$\frac{E_y|^{n+1}_{i,j,k} - E_y|^{n-1}_{i,j,k}}{2\Delta t} = \frac{1}{\varepsilon}\left(\frac{H_x|^{n}_{i,j,k+1} - H_x|^{n}_{i,j,k-1}}{2\Delta z} - \frac{H_z|^{n}_{i+1,j,k} - H_z|^{n}_{i-1,j,k}}{2\Delta x}\right)$$

$$\frac{E_z|^{n+1}_{i,j,k} - E_z|^{n-1}_{i,j,k}}{2\Delta t} = \frac{1}{\varepsilon}\left(\frac{H_y|^{n}_{i+1,j,k} - H_y|^{n}_{i-1,j,k}}{2\Delta x} - \frac{H_x|^{n}_{i,j+1,k} - H_x|^{n}_{i,j-1,k}}{2\Delta y}\right)$$

$$\frac{H_x|^{n+1}_{i,j,k} - H_x|^{n-1}_{i,j,k}}{2\Delta t} = \frac{-1}{\mu}\left(\frac{E_z|^{n}_{i,j+1,k} - E_z|^{n}_{i,j-1,k}}{2\Delta y} - \frac{E_y|^{n}_{i,j,k+1} - E_y|^{n}_{i,j,k-1}}{2\Delta z}\right)$$

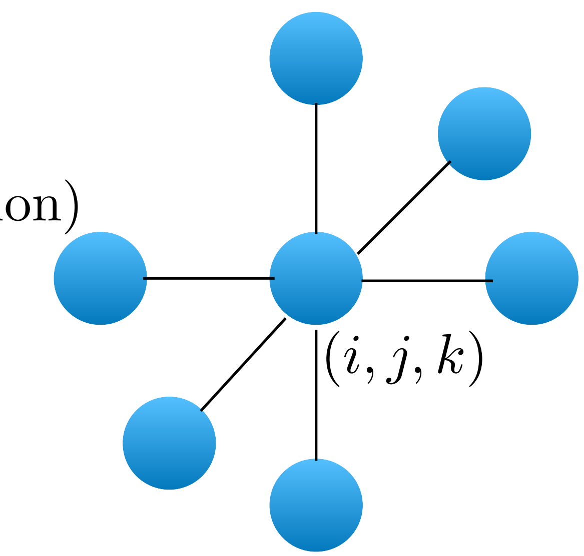
$$\frac{H_y|^{n+1}_{i,j,k} - H_y|^{n-1}_{i,j,k}}{2\Delta t} = \frac{-1}{\mu}\left(\frac{E_x|^{n}_{i,j,k+1} - E_x|^{n}_{i,j,k-1}}{2\Delta z} - \frac{E_z|^{n}_{i+1,j,k} - E_z|^{n}_{i-1,j,k}}{2\Delta x}\right)$$

$$\frac{H_z|^{n+1}_{i,j,k} - H_z|^{n-1}_{i,j,k}}{2\Delta t} = \frac{-1}{\mu}\left(\frac{E_y|^{n}_{i+1,j,k} - E_y|^{n}_{i-1,j,k}}{2\Delta x} - \frac{E_x|^{n}_{i,j+1,k} - E_x|^{n}_{i,j-1,k}}{2\Delta y}\right)$$

… plus divergence equations

MAXWELL EQS: INVERSE PROBLEM

$$\begin{cases} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \quad \text{(Gauss's Law)} \\ \nabla \cdot \mathbf{B} &= 0 \quad \text{(Gauss's Law for Magnetism)} \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \quad \text{(Faraday's Law)} \\ \nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad \text{(Ampère's Law with Maxwell's correction)} \end{cases}$$



$$D_{\text{div}} E(t) = \frac{\rho(t)}{\epsilon_0}, \quad D_{\text{div}} B(t) = 0$$

$$\frac{d}{dt} \begin{pmatrix} E \\ B \end{pmatrix} (t) = \begin{pmatrix} 0 & -D_{\text{curl}} \\ D_{\text{curl}} & 0 \end{pmatrix} \begin{pmatrix} E \\ B \end{pmatrix} (t) + \frac{1}{\varepsilon_0} \begin{pmatrix} S \\ 0 \end{pmatrix} J(t)$$

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) := C \begin{pmatrix} E \\ B \end{pmatrix} (t) = Cx(t)$$

INVERSE PROBLEM AS OPTIMAL CONTROL

divergence equations on B, E

$$u^\star = \arg\min_u \int_0^\infty \Big[(Cx(t) - z(t))' Q (Cx(t) - z(t)) + u(t)' R u(t)\Big] dt$$

↓

$$A'P + PA - PBR^{-1}B'P + C'QC = 0$$

It's likely very hard to solve!

↓

$$u(t) = -R^{-1}B'Px(t)$$

HAMILTONIAN SOLUTION

$S = BR^{-1}B'$.

$$H(x,p) = \frac{1}{2}(Cx(t) - z(t))'Q(Cx(t) - z(t)) - \frac{1}{2}p'Sp$$

$$\dot{x}(t) = Ax(t) - Sp(t) \qquad \dot{x}(t) = Ax(t) + Bu(t)$$

$$\dot{p}(t) = -C'Q\big(Cx(t) - z(t)\big) - A'p(t)$$

$$x(0) = x_0$$

$$p(0) = p_0$$

generally hard to solve

$$u(t) = -R^{-1}B'p(t)$$

TIME SYMMETRY

time symmetry

A eigenvalues on the imaginary axis

$$\dot{x}(t) = Ax(t) - Sp(t) \qquad \dot{x}(t) = Ax(t) + Bu(t)$$

$$\dot{p}(t) = -C'Q\big(Cx(t) - z(t)\big) - A'p(t)$$

$$x(0) = x_0$$

$$\cancel{p(0) = p_0}$$

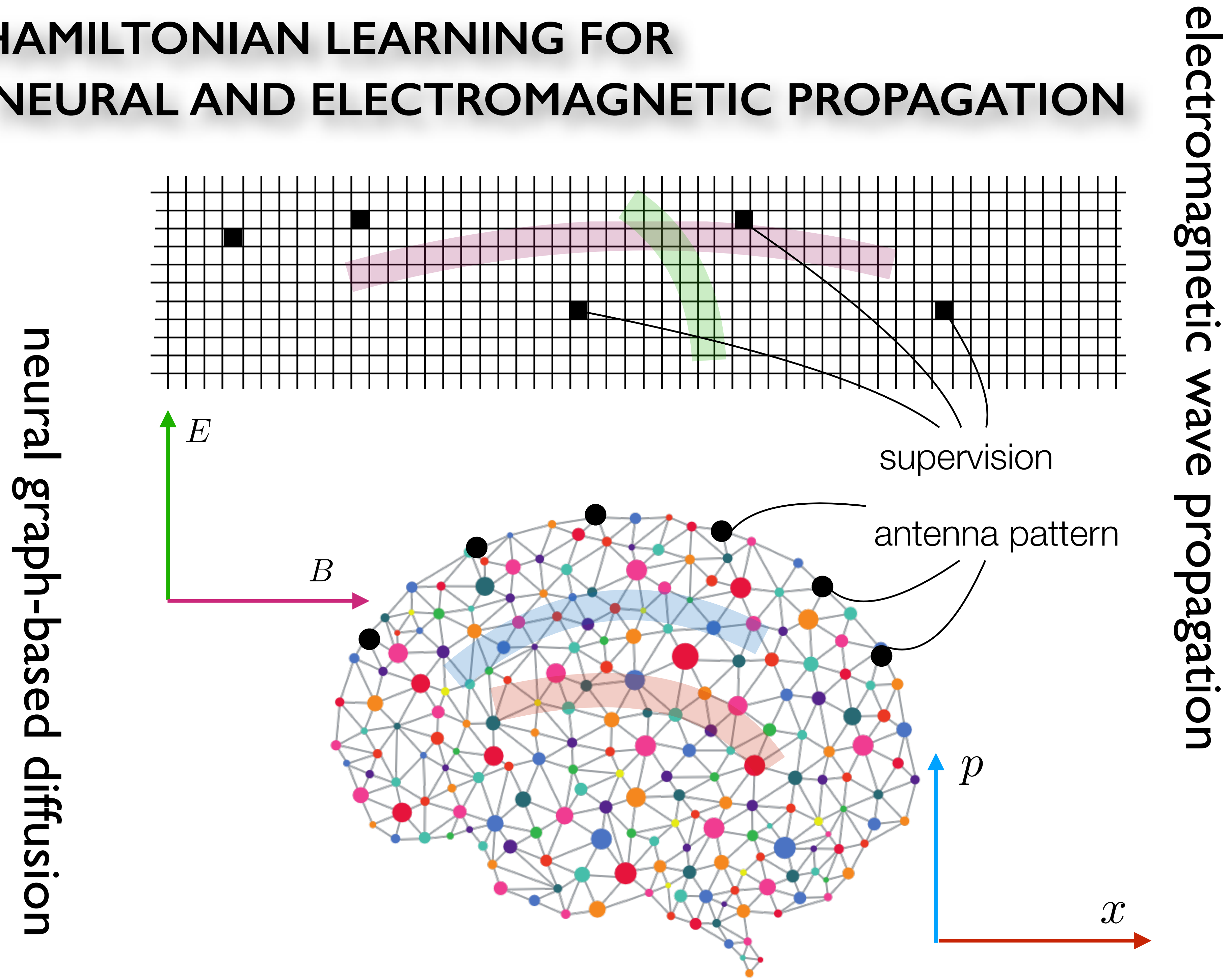
boundary conditions

$$u(t) = -R^{-1}B'p(t)$$

"forward propagation also for the co-state!"

$$\dot{u}(t) = A'u(t) + R^{-1}B'C'Q\big(Cx(t) - z(t)\big)$$

HAMILTONIAN LEARNING FOR NEURAL AND ELECTROMAGNETIC PROPAGATION



S3P-2024 - Neural vs wave propagation

CONCLUSIONS

- Regulated access to data collections and the challenge of CollectionLess AI - emphasis on environmental interactions
- Learning theory inspired from Theoretical Physics; a pre-algorithmic step: Cognitive Action, natural laws vs algorithms)
- Hamiltonian Learning and dissipation
- Local SpatioTemporal Propagation (LSTP) as a proposal to replace Backpropagation in “temporal learning environments”
- Electromagnetic wave propagation

HIRING AT SAILAB on Collectionless AI

Two postdoc positions

2 years (50 KEuro/year)