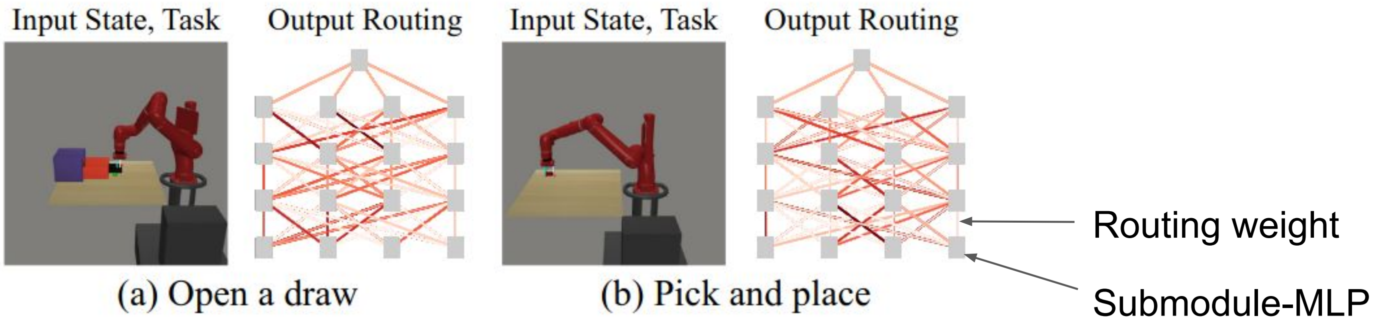
# Astrocytes for Neural Information Routing in Reinforcement Learning

Tianqin Li

May 23

# Overview

- Multi-Task Reinforcement Learning with Soft Modularization
- Property of astrocyte
- Potential idea for incorporate astrocytes property in routing submodulerized neural network



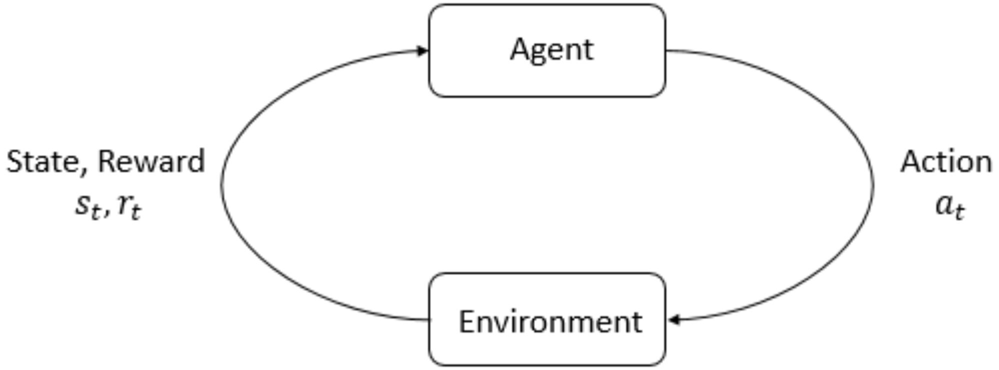

(a) Open a draw (b) Pick and place

# Multi-Task Reinforcement Learning with Soft Modularization

Ruihan Yang, Huazhe Xu, Yi Wu, Xiaolong Wang
2020 March

# Reinforcement Learning

- State value $S_t$
- Action value $a_t \sim \pi(\cdot|S_t)$
- Transition distribution $P(S_{t+1}|S_t, a_t)$
- Reward $R(S_t, a_t)$
- Policy $\pi_\phi(a_t|S_t)$
- Learn policy that maximize the cumulative rewards



# Reinforcement Learning

- Trajectories: sequence of states and actions in the world

$$\tau = (S_0, a_0, S_1, a_1, \dots)$$

- At first, $S_0$ is random sampled from a start state distribution

$$S_0 \sim \rho_0(\cdot)$$

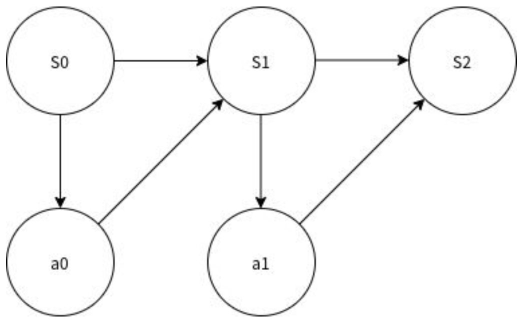
- Given $S_t$ and $a_t$ , the state at t+1 $S_{t+1}$ is produced stochastically:

$$S_{t+1} \sim P(\cdot | S_t, a_t)$$

# Reinforcement Learning

- Reward at time t: $r_t = R(S_t, a_t)$
- Cumulative reward in infinite time:

$$R(\tau) = \sum_{t=0}^{\infty} \gamma^t r_t$$




- RL fundamental problem:
  - Probability over trajectories

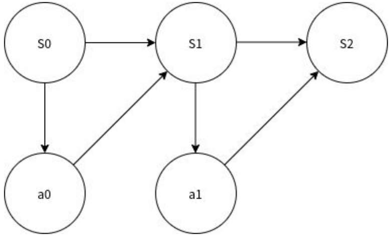
$$P(\tau|\pi) = \rho_0(S_0) \prod_{t=0}^{\infty} P(S_{t+1}|S_t, a_t)\pi(a_t|S_t)$$

  - Optimize policy to obtain the max expected return for all observed $\tau$

$$\pi^* = \operatorname{argmax}_{\pi} \int_{\tau} P(\tau|\pi)R(\tau) = \operatorname{argmax}_{\pi} \mathbb{E}_{\tau\sim\pi}[R(\tau)]$$

$$\pi^* = \text{argmax}_\pi \int_\tau P(\tau|\pi)R(\tau) = \text{argmax}_\pi \mathbb{E}_{\tau\sim\pi}[R(\tau)]$$

# Reinforcement Learning




- Value functions / state-action pair

$$V^\pi(s) = \mathbb{E}_{\tau\sim\pi}[R(\tau)|S_0 = s]$$

$$Q^\pi(s,a) = \mathbb{E}_{\tau\sim\pi}[R(\tau)|S_0 = s, a_0 = a]$$

$$V^\pi(s) = \mathbb{E}_{a\sim\pi}[Q^\pi(s,a)]$$

- Recursively express the state value function / state-action pair

$$V^\pi(s) = \mathbb{E}_{a\sim\pi(\cdot|s),s'\sim P(\cdot|s,a)}[R(s,a) + \gamma V^\pi(s')]$$

$$Q^\pi(s,a) = \mathbb{E}_{s'\sim P(\cdot|s,a)}[R(s,a) + \gamma \mathbb{E}_{a'\sim\pi(\cdot|s')}[Q^\pi(s',a')]]$$

$$\pi^* = \text{argmax}_\pi \int_\tau P(\tau|\pi)R(\tau) = \text{argmax}_\pi \mathbb{E}_{\tau\sim\pi}[R(\tau)]$$

# Reinforcement Learning

$$R(\tau) = \sum_{t=0}^{\infty} \gamma^t r_t$$

$$Q^\pi(s,a) = \mathbb{E}_{s'\sim P(\cdot|s,a)}[R(s,a) + \gamma\mathbb{E}_{a'\sim\pi(\cdot|s')}[Q^\pi(s',a')]]$$

- Policy / Q-function update:
  - Denote ***D*** as data
  - Maximize w.r.t. $\pi$ function

$$J(\pi) = \mathbb{E}_{s_t\sim D}[\mathbb{E}_{a_t\sim\pi(\cdot|s_t)}[Q(s_t,a_t)]]$$

  - Minimize w.r.t. Q-function

$$J(Q) = \mathbb{E}_{(s_t,a_t)\sim D}[\tfrac{1}{2}(Q(s_t,a_t) - (R(s_t,a_t) + \gamma\mathbb{E}_{s_{t+1}\sim P(\cdot|s_t,a_t)}V(s_{t+1})))^2]$$

$$V^\pi(s) = \mathbb{E}_{a\sim\pi}[Q^\pi(s,a)]$$

- Policy: Actor
- Q-function: Critic

# Multi-task Soft Actor Critic

- Using entropy regularization to encourage exploration

$$J(\pi) = \mathbb{E}_{s_t \sim D}[\mathbb{E}_{a_t \sim \pi(\cdot|s_t)}[Q(s_t, a_t) + \boxed{\alpha H(\pi(\cdot|s_t))}]]$$

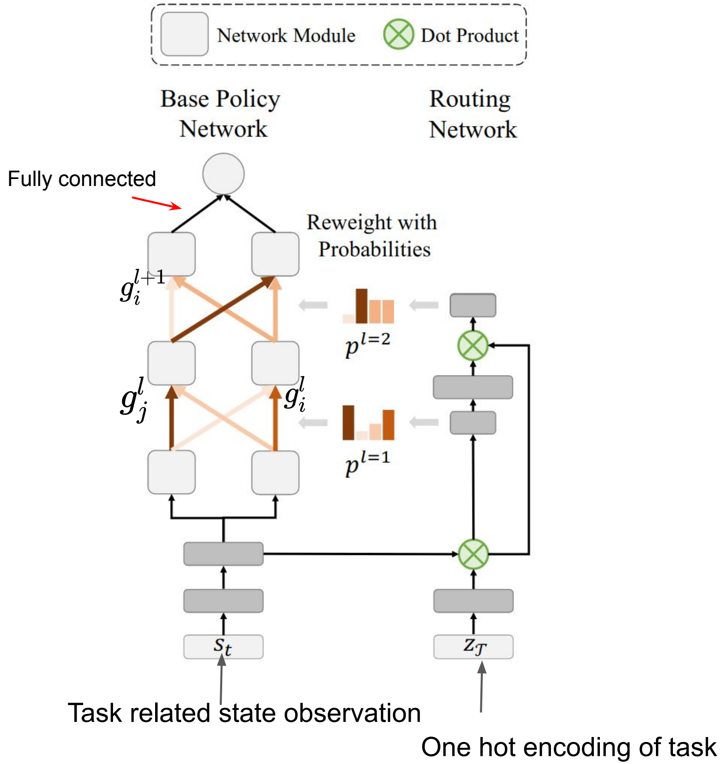
$$J(Q) = \mathbb{E}_{(s_t,a_t)\sim D}[\tfrac{1}{2}(Q(s_t, a_t) - (R(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1}\sim P(\cdot|s_t,a_t)} V(s_{t+1})))^2]$$

- Multi-task learning
  - Task follows certain distribution p(T)
  - Marginalize different T out

$$J(\pi) = \mathbb{E}_{T \sim p(T)}[J(\pi, T)]$$

$$J(Q) = \mathbb{E}_{T \sim p(T)}[J(Q, T)]$$

# Routing network for modularization




Calculating routing weights:

$$p^{l+1} = \mathcal{W}_d^l(\mathrm{ReLU}(\mathcal{W}_u^l p^l \cdot (f(s_t) \cdot h(z_\mathcal{T}))))$$

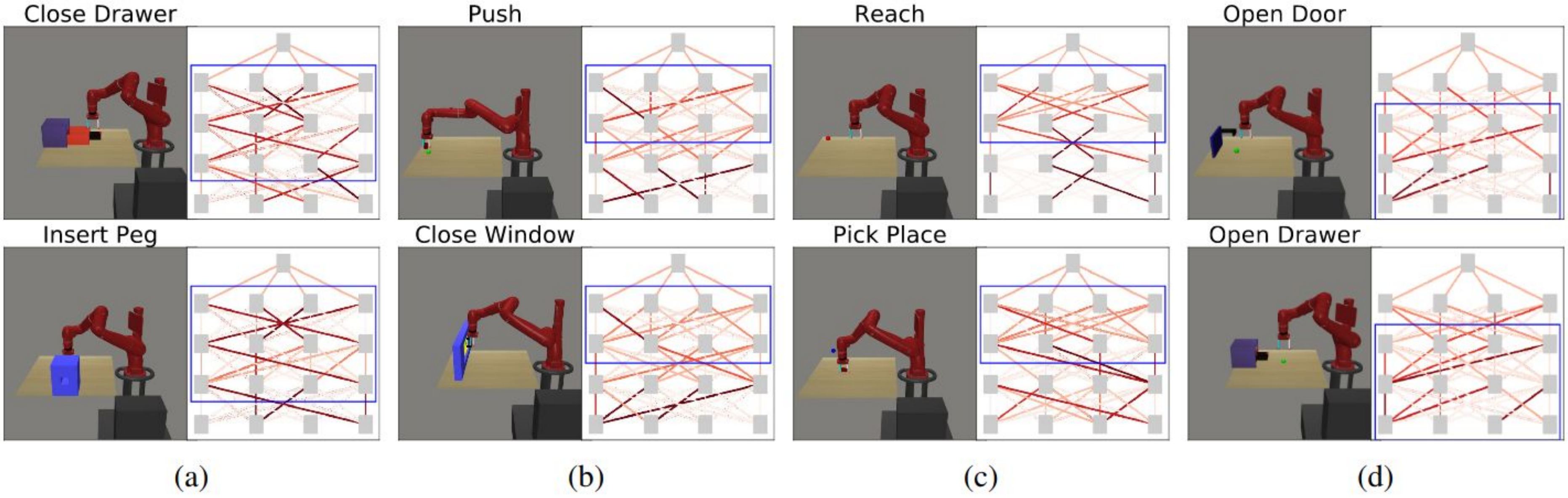
$$p^{l=1} = \mathcal{W}_d^{l=1}(\mathrm{ReLU}(f(s_t) \cdot h(z_\mathcal{T})))$$

$$\hat{p}_{i,j}^l = \frac{\exp(p_{i,j}^l)}{\sum_{j=1}^n \exp(p_{i,j}^l)}$$

Rerouting the modular network subcomponents:

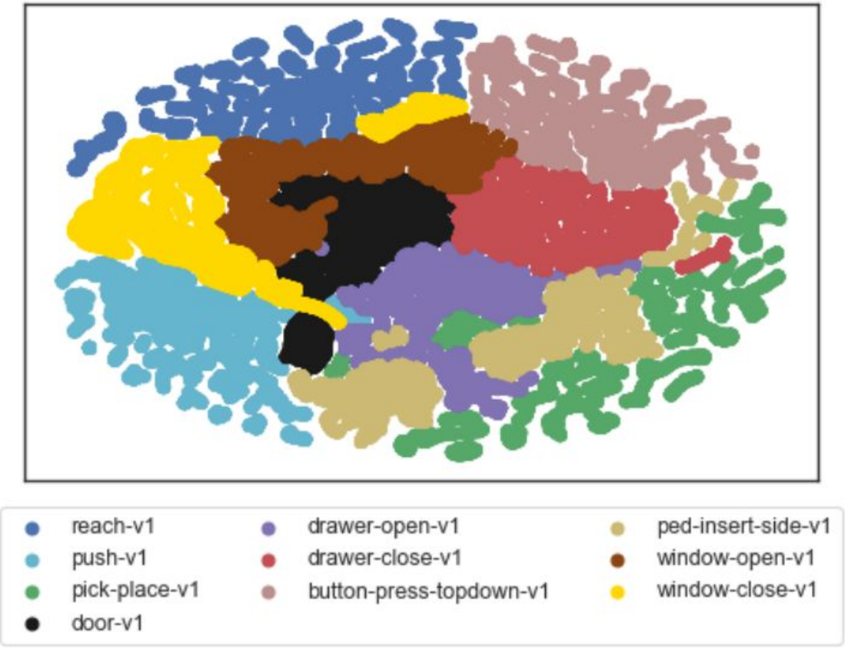
$$g_i^{l+1} = \sum_{j=1}^n \hat{p}_{i,j}^l(\mathrm{ReLU}(W_j^l g_j^l)) \qquad g_j^l \in R^d$$

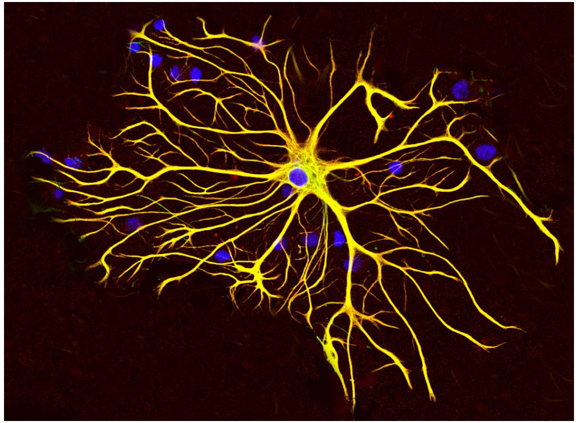
# Results - probability visualization



# Results - probability visualization

- Concatenate all the routing weight together and perform tSNE.
- Clear distinction from different tasks



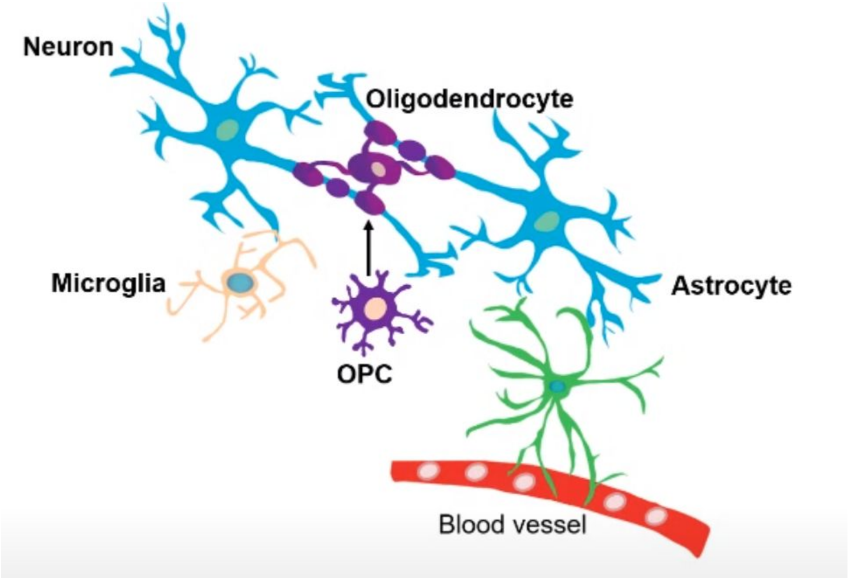


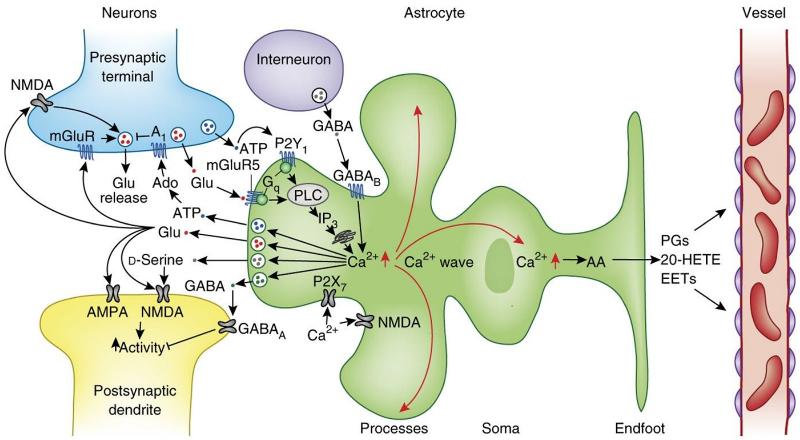
# Astrocyte inside the brain

Their potential biological evidence in routing neural network for different task

# Astrocyte inside the brain

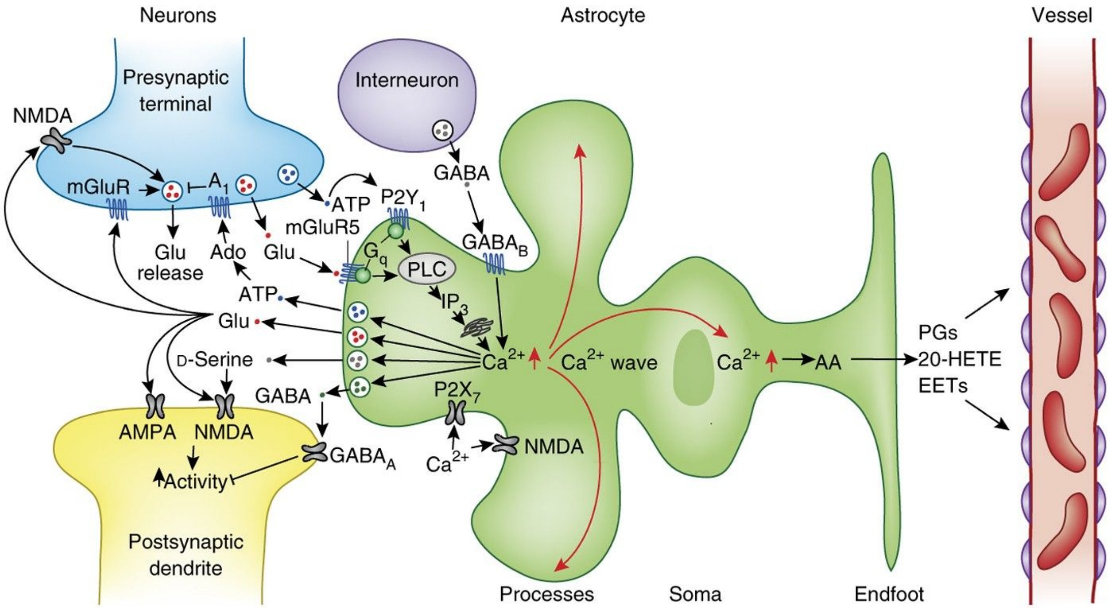
- Known for forming triple synapse with neurons
- Each astrocytes covers 140,000 synapses (Bushorg et. al, 2002)
- Integration of neural signals in time and spatial manner



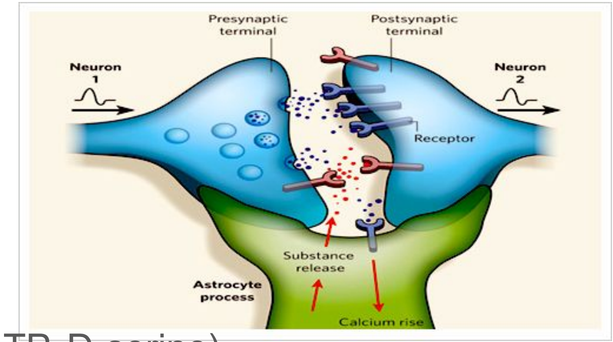



# Astrocyte-neuron interaction

- Astrocytes uses Ca2+ elevation to change behaviour
- Ca2+ level in astrocytes can be activated by neurons

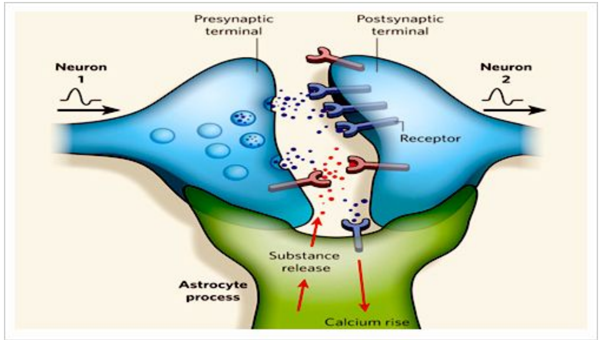


# Astrocyte-neuron interaction

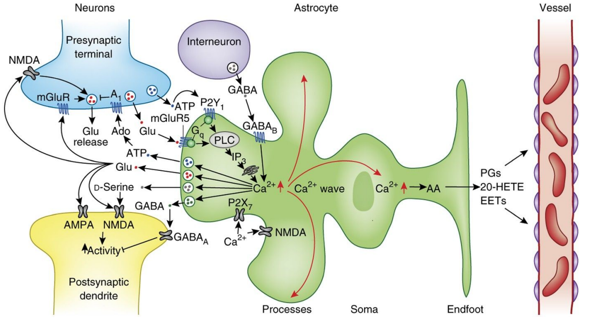



- How can astrocytes affect synaptic plasticity?
  - Synaptic signals introduce Calcium elevation in astrocytes
  - Calcium elevation -> release of glia-transmitter (glutamate, ATP, D-serine)
  - Many ways to affect the synapse:
    - Glutamate induces postsynaptic slow inward current (SIC) which leads to postsynaptic action potential
    - Glutamate also alters frequency of miniature postsynaptic current (mPSCs), which leads to increase of presynaptic transmitter release
- Compartmentalization of astrocytes behavior
  - Microdomain of astrocytes behave differently on Ca2+ elevation
  - Local regulation and soma level Ca2+ propagation is seperated

# Astrocyte-neuron interaction



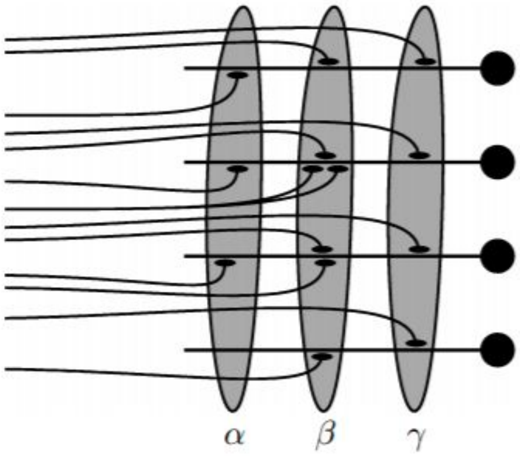

- Two level of Ca2+ elevation in astrocytes
  - Micro domain level:
    - Happen in local process far away from soma
    - Take 0.2-5 seconds to receive neuron signals
    - Last for 0.3 - 10 seconds locally
    - Sufficient modulate short term synaptic efficiency
  - Somatic level:
    - Robust and happen in somatic level
    - Take longer to activate but last tens of seconds




- Well suited for routing signals in neuron circuits by somatic level Ca2+ wave

# Astrocyte-neuron interaction

- Hypothesis:
  - Different microdomains is activated by initial task
  - Downstream synapses are grouped by astrocytes and enhanced together
  - The time scale of astrocytes enhancement and activation may help in on-policy learning




Caroline et al,. Glial Cells for Information Routing? Cognitive Systems Research, doi:10.1016/j.cogsys.2006.07.001

Fig. 3. Network of four target neurons with three microdomains $\alpha$, $\beta$ and $\gamma$ and afferent fibers. The dendritic trees of the neurons are symbolized by horizontal straight lines.