突触束理论在脉冲驱动的感觉运动系统中的应用：八个以上的独立突触束会破坏奖励-STDP 学习过程

Takeshi Kobayashi, Shogo Yonekura, Yasuo Kuniyoshi
Laboratory for Intelligent Systems and Informatics, Department of Mechano-Informatics, Graduate School of Information Science and Technology, TheUniversity of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan
(10Dated: 2025 年 8 月 24 日)

神经元尖峰直接驱动肌肉，赋予动物敏捷的动作，但将基于尖峰的控制信号应用于人工感官运动系统中的执行器不可避免地导致学习崩溃。我们开发了一个可以改变传感器到运动连接中独立突触束的数量的系统。本文展示了以下四项发现：(i) 一旦运动神经元的数量或独立突触束的数量超过某个临界极限，学习就会崩溃。(ii) 运动神经元数量较少会增加学习失败的概率，而 (iii) 如果学习成功，则较少的运动神经元会导致更快的学习。(iv) 反向于最优权重移动的权重更新次数可以定量解释这些结果。尖峰的功能在很大程度上仍然未知。确定使用尖峰构建学习系统的参数范围将使研究以前由于学习难度而无法访问的尖峰功能成为可能。

动物通过神经元中的尖峰动作电位传递信息，从而实现高能量效率和适应性行为。脉冲神经网络（SNNs）明确将这些尖峰作为计算原语 [1]。SNNs 已被用于建模周围感觉系统 [2, 3]、被称为液态状态机的皮层微电路 [4] 以及贝叶斯推断的神经实现 [5–7]。从工程角度来看，它们提供高能量效率 [8]、逃离局部最优的能力 [9]、即时适应性出现 [10] 以及参数变化的元鲁棒性 [11]。SNNs 也被应用于机器人学中的感觉运动学习 [12–16]。动物同样使用尖峰来驱动它们的肌肉 [17]。然而，用尖峰来驱动机器人电机使得端到端的学习非常困难。在实践中，现有的框架使用发放率编码 [13, 15] 或由外部反馈控制器提供的平滑控制信号 [12, 14, 16] 驱动执行器，从而完全避免了不连续的脉冲基命令。感觉运动学习崩溃的机制尚不清楚。阐明这一问题将有利于神经启发的人工智能和自适应机器人技术（例如，整合即时适应性的出现与学习），深化我们对学习如何与动物的神经肌肉系统相互作用的理解，并从非平衡物理学的角度，帮助分析自我组织性能如何依赖于系统参数。

通常，突触权重在学习过程中是独立更新的（图 1(b)，顶部）。在此方法中，我们将权重共享约束纳入学习规则，使权重能够部分或完全耦合（图 1(b)，中部/底部）。我们定义耦合级别为 N_b，独立突触束的数量。然后，我们系统地考察了 N_b 和 N_m 的组合以及运动神经元的数量如何影响学习性能。这里，N_m 决定输出方差，而 N_b 设定可学习参数的数量。实验揭示了一个尖峰基础学习成功的参数区域：$N_b \leq 8$ 和 $6 \leq N_m \leq 20$。现有框架使用独立的突触权重（大的 N_b）。然而，我们的结果显示实际需要共享的突触权重（小的 N_b）进行尖峰学习。这一发现仅在我们的设置中能够改变 N_b 的情况下才得以揭示。结果还表明，自然界中不存在 $N_m \to \infty$ 和 $N_b \to \infty$ 生物的原因可能不仅在于能量限制，还在于固有的学习能力上限。

基准问题是双阱势中质点的稳定化，因为这个看似简单的系统在形式上等同于经典的不稳定平衡问题，如倒立摆的平衡和人类步态的控制。这使其成为学习和控制算法 [10] 的重要测试平台。质量 $x(t)$ 的动力学和势能 $U(x)$ 是

$$m\ddot{x} = -\gamma\dot{x} - \frac{dU(x)}{dx} + A(t) \tag{1a}$$

$$U(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2 \tag{1b}$$

其中，m 是质量，γ 是摩擦系数，而 $A(t)$ 是由 SNN 控制器产生的控制力，其定义如下。控制的目标是保持质量在区间 $[-0.1, 0.1]$ 内。

SNN 控制器架构如图 1(A) 所示。对于一个运动神经元 i，其瞬时放电概率用膜电位 $u_i(t)$ 表示：

$$\rho_i(t) = \exp\big(u_i(t)\big)dt \tag{2}$$

基于脉冲的运动指令 $A(t)$ 是从正向和负向运动池中的突触后电位的带符号总和 $y(t)$ 中获得的。

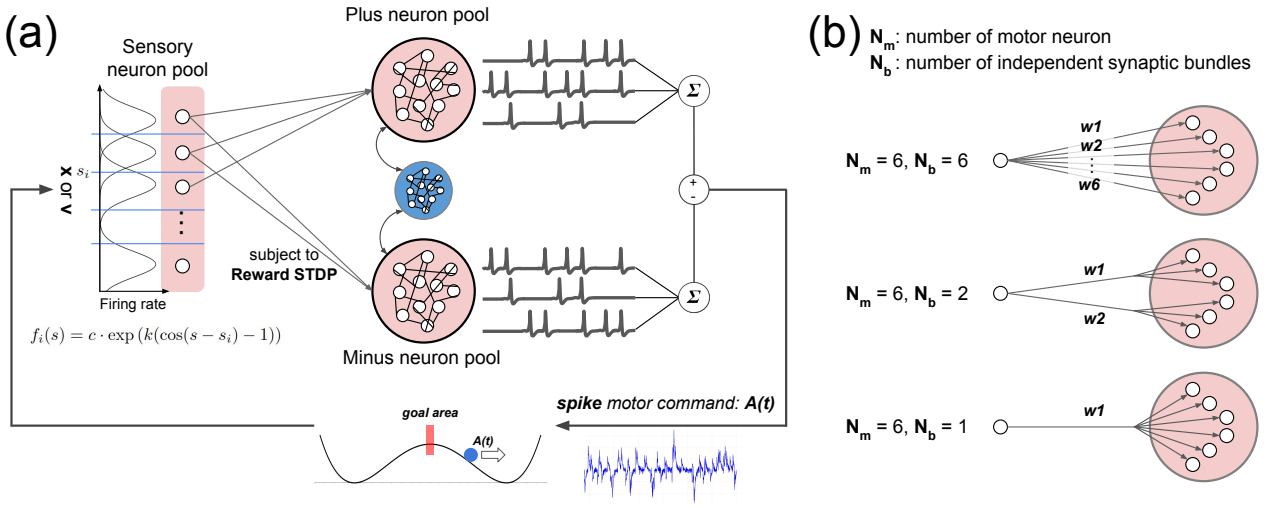

图 1. (a) 双井势任务和网络架构。目标是通过调整左右推力 $A(t)$ 来稳定点质量在红色目标区域内。SNN 控制器由单独的感觉神经元池和运动神经元池组成。位置 x 和速度 v 由两个不同的感觉池编码（仅显示其中一个）。每个感觉池包含 30 个神经元，其调谐曲线为 $f_i(s) = c\exp\big(k(\cos(s-s_i)-1)\big)$，其中 $c=40$、$k=12.5$ 和 s_i 在 $[-1.5, 1.5]$ 范围内以 30 个等步长分布。运动神经元被分为产生正推力和负推力的池。这些池通过侧抑制形成胜者全得电路。来自正池和负池的均值 PSPs 的带符号总和产生了尖峰运动命令 $A(t)$（方程 3b）。感觉-运动突触通过奖励调制的 STDP 进行训练。(b) 学习性能通过改变 N_b、独立突触束的数量进行系统研究，其中 N_m 突触从单个感觉神经元到运动池可分配的权重值的数量即为此参数。当 $N_m=6$ 和 $N_b=1$ 时，所有六个突触共享一个权重。当 $N_b=2$ 时，突触形成两个各含三个成员的组且权重相同。当 $N_b=6$ 时，每个突触具有独立的权重。

$$\tau_a \frac{dy(t)}{dt} = -y(t) \pm \frac{1}{N} \sum_{i \in \pm\text{motor pool}}^{N_m} \delta(t - t_f^{(i)}) \tag{3a}$$

$$A(t) = g_a y(t) \tag{3b}$$

电机指令的时间常数是 $\tau_a = 10\text{ms}$，其增益是 $g_a = 200.0$，而 N_m 表示每个（正或负）池中的运动神经元数量。

突触权重在感觉神经元和运动神经元之间的更新是由一种奖励-STDP 类型的在线强化学习规则 [18] 完成的。为了用单一变量表示连接的存在及其强度，引入了突触参数 θ_i，并定义相应的突触权重 w_i 如下：

$$w_i(t) = \begin{cases} \exp(\theta_i(t) - \theta_0) & \text{if } \theta_i(t) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

θ_i 的更新规则表示为

$$d\theta_i(t) = \eta g_i(t) dt + \sqrt{2\eta T} dW \tag{5}$$

学习率是 $\eta = 1.5 \times 10^{-4}$，温度参数是 $T = 0.1$，$g_i(t)$ 表示奖励梯度的局部估计，而 dW 代表一个维纳过程。

$$\frac{de_i(t)}{dt} = -\frac{e_i(t)}{\tau_e} + w_i(t) I_i(t) (S_{\text{post}}(t) - \rho_{\text{post}}(t)) \tag{6}$$

$$\frac{dg_i(t)}{dt} = -\frac{g_i(t)}{\tau_g} + r(t) e_i(t) \tag{7}$$

资格迹是 $e_i(t)$，突触后放电率是 $\rho_{\text{post}}(t)$，资格迹时间常数是 $\tau_e = 1.9\text{s}$，奖励梯度时间常数是 $\tau_g = 50\text{s}$，奖励信号是 $r(t)$，由附录定义。

变化 N_b（图 1(B)）的方法是通过用虚拟的一对多突触替换常规的一对一突触连接，这些突触的权重被共享。为了启用这种一对多连接的 STDP 风格学习，将受相同权重约束的突触的相关资格迹线近似为其平均值（方程 6）（方程 8）。

$$\frac{de_i(t)}{dt} = -\frac{e_i(t)}{\tau_e} + w_i(t) I_i(t) \frac{1}{N_b} \sum_{j=1}^{N_b} (S_{\text{j}}(t) - \rho_{\text{j}}(t)) \tag{8}$$

Although motor neurons connected to the same synaptic bundle have the same membrane potential, spike variability is maintained since spiking is determined by equation (2). The parameter N_b can be chosen freely

within $N_b \leq N$, but for implementation convenience, it was restricted to values that divide N_m exactly.

学习是根据方程 (5) 在线进行的。为候选人准备了五个初始位置 x_0 和五个初始速度 v_0，每种组合每个周期训练一次。这导致每个周期有 25 个时段。位置候选是 $x_0 \in \{-1.0, -0.5, 0.0, 0.5, 1.0\}$，速度候选是 $v_0 \in \{-0.35, -0.20, 0.0, 0.20, 0.35\}$。每个时段持续 45 秒，随后有一个 5 秒的间隔来重置网络。

每个周期后的性能评估使用附录中定义的分数。该量表经过归一化处理，使得在整个时段内保持 $x = 0$ 和 $v = 0$ 的值为 1.0。θ 的初始值从 [3.0, 5.0] 的均匀分布中抽取。

学习成功率如图 2(a) 所示，当突触权重共享 ($N_b = 1$) 和 $N_m = \{1, 2, \ldots, 9, 10\}$。成功率达到 $N_m > 7$ 以上时超过 90%。对于较小的 N_m，成功率下降。这种行为源于方程 (3b) 中读出依赖于 N_m(图 2(b))，表明当与固定增益 $g_a = 200.0$ 结合时，控制质点存在一个最优的 N 范围。

$N_b = 1$ 和 $N_m = \{6, 10, 15, 20, 25, 30\}$ 的学习曲线如图 2 所示。N_m 的值越大，学习过程就越慢。对于 $N_m > 25$，没有观察到学习进展，因为方差太小，无法生成足够的输出来移动质点（图 2(b)）。

最终，图 2(d) 显示了独立突触权重 ($N_b = N$) 与 $N_b = \{6, 7, \ldots, 10, 11\}$ 的学习曲线。学习在 $N_b \leq 8$ 处进行，在 $N_b = 9$ 处变得不稳定，并且对于 $N_b \geq 10$ 无法取得进展。

为了更精确地分析学习性能，在每个时间步长对每个突触的错误转换的数量进行了计数。这是方程 7 中 $g_i(t)$ 符号指向与正确的突触权重更新方向相反次数的数量。目标权重是从共享权重 ($N_b = 1$) 和 $6 \leq N_m \leq 20$ 的学习成功运行中获得的。每个生成的权重向量都被视为真实值，计算了它们之间的成对余弦相似性。由于每一对得分至少为 0.95，选择了 $N_b = 1, N_m = 10$ 学习权重作为最优权重。

如图 3（左）所示，错误转换的数量随 N_m 增加。与之前观察结果一致，学习速度在 N_m 较小的情况下更快，具有较少运动神经元的运行表现出更少的错误转换。这证实了较低数量的错误转换与较快的学习相吻合。

如图 3（右）所示，错误转移计数的数量随 $N_b < 10$ 增加。在大约 8 附近的 N_b 处，错误转移计数的数量急剧增加。这一准确结果表明较小的 N_b 更有利于学习。

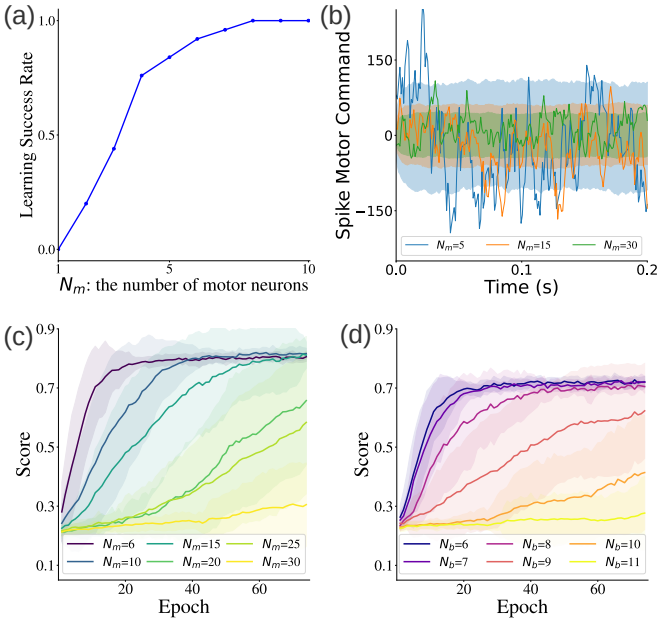

图 2. (a)$N_m \leq 10$ 的学习成功率。对于每个 N_m 值执行了 25 次训练运行。如果最终周期分数超过 0.65，则认为该运行为成功。(b)N_m 对尖峰运动指令幅度（方差）的影响。对于每个 N_m，进行了 1000 次 0.2 秒的模拟，每个神经元的平均放电概率为 0.15。输出是通过等式 (3b) 生成的。阴影带表示输出方差的大小，实线代表单个随机选择的输出轨迹。(c) 带有 $N_m = \{6, 10, 15, 20, 25, 30\}$ 的 $N_b = 1$ 学习曲线。每个 N_m 都进行了 20 次训练运行。实线表示均值，阴影带表示标准差。(d) 带有 $N_m = N_b$ 的 $N_b = \{6, 7, 8, 9, 10, 11\}$ 的学习曲线。每个 N_b 都进行了 20 次训练运行。实线表示均值，阴影带表示标准差。

图 4 展示了学习分数与错误转移计数的散点图，其中 $5 \leq N_m \leq 25$ 的 N_b 值变化。当错误转移计数较低时，学习分数较高。这一情况也对应于 N_b 较小的值。

本研究完全使用脉冲表示感觉输入和运动输出，并将奖励调制的 STDP 强化学习应用于最简单的感知-运动任务：双势阱中质点的稳定。将独立突触来的数量，N_b 视为可控制参数，允许独立改变从传感器到运动神经元突触的数量和运动神经元的数量。这使得能够全面探索参数空间。

分析揭示了三个关键发现：(i) 学习仅在神经元数量为 $6 \leq N_m \leq 20$ 的范围内保持稳定，(ii) 较小的 N_m 导致更快的收敛，以及 (iii) 仅当 $N_b \leq 8$ 时观察到学习进展。当 $N_m < 6$，突触后电位读出的方差变得如此之大，以至于根据公式 3b 得到的结果尖峰运动命令不适合控制质点（图 2(b)）。相反，当 $N_m > 20$，方差变

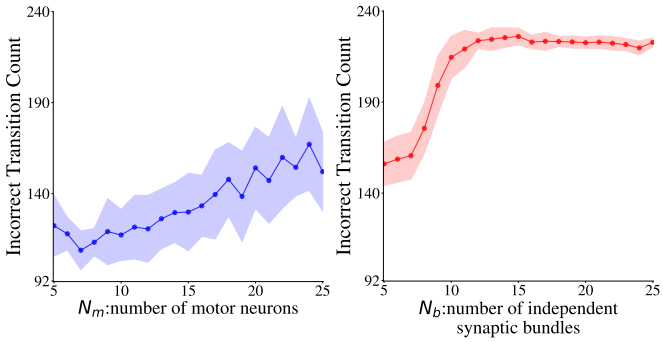

图 3. (左) 不正确转换计数 (纵轴) 与错误突触权重方向相反的权值更新次数之间的关系，以及与 N_m 和 $N_b = 1$ 的关系。(右) 对于具有 $N_m = N_b$ 的网络，相同的指标绘制在 N_b 上。对于每个 N_m (或 N_b) 的值进行了 15 次训练运行。实线表示平均值，阴影带表示标准差。不正确的转换计数代表每秒每个突触的不正确转换次数的平均值。

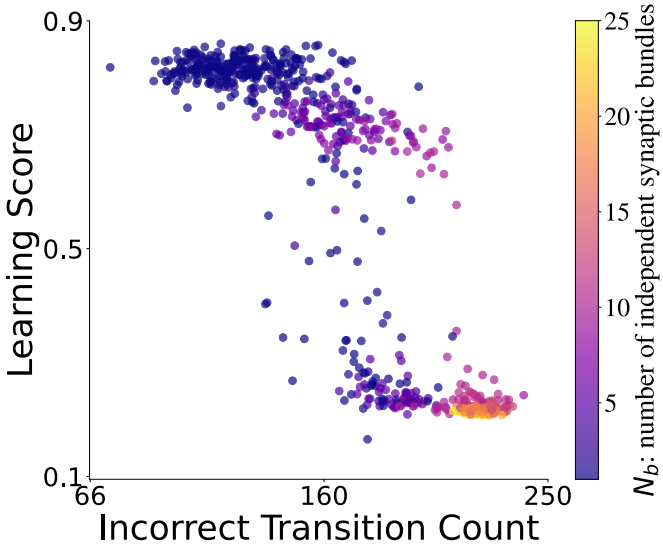

图 4. 学习得分与错误转换次数的散点图，其中错误转换次数是指权重更新朝最优突触权重相反方向移动的次数。对于参数组合进行的学习运行中，变化了 N_b 而 $5 <= N_m <= 25$ 保持不变。每个点代表一次单独的训练运行。每个点的颜色对应于 N_b 的值。对于每组参数组合，进行了 15 次训练运行。错误转换次数表示每个突触每秒平均发生的错误转换数量。

得太小而无法生成足够大的控制输入。发现 (ii) 是因为在小的 N_m 处较大的方差增强了探索驱动并加速了学习。发现 (iii) 是反直觉的。虽然原则上更多的突触应该增加表示能力，但实际上较少的不同权重的学习效果更好。这种效应归因于最终运动指令是正负运动池活动之和。过多独立的突触束会导致相反方向频繁更新权重，这干扰了收敛。例如，当 10 个正向运动神经元和 5 个负向运动神经元同时放电时，净力加速质点朝正方向移动。当这种运动受到奖励时，投射到正池中的突触被正确强化。然而，投射到负池的突触被错误地强化了。因此，在单一符号运动池内分配超过八个不同的突触权重 ($N_b > 8$) 揭示了一个空间信用分配问题。

为了更精确地分析学习性能，将性能绘制为错误转换计数的函数。如图 3 和图 4 所示，观察值 (iiii) 都一致地由这个单一指标解释。

在生物系统中，运动神经元池通常由一个共同输入同步驱动——这种共同驱动策略已在神经生理学上得到了记录 (例如，[19])。在我们的模型中，将 N_b 视为可调参数创建了对该策略的计算表示。共享输入缓解空间信用分配问题并加速学习这一发现为共同驱动的功能益处提供了新的计算证据——这种效果仅当整个感觉运动环路由尖峰表示时才明显。当 N_m 过小时学习崩溃，这与先前研究观察 [17] 一致，即单个运动神经元无法在嘈杂的神经系统中可靠地向肌肉传递共同输入。最后，模拟结果表明，只有少数神经元和突触时学习最有效。这一发现支持生物系统的特点资源效率，在这种情况下，显著的学习能力从最少的神经资源中产生。

从工程的角度来看，在保持运动输出的尖峰表示的同时展示传感器-运动学习允许将由尖峰诱导的顺序 (*SIO*) [10] 与强化学习相结合。海上输入输出是一种适应性效应，预计会在具有高度自由度的系统中出现。虽然本研究仅限于单个自由度，但未来的研究可以包括扩展到具有更高自由度的系统，探索与肌肉协同作用 [20] 的联系，并更广泛地应用 *SIO*。

本工作得到了日本学术振兴会 (KAKENHI) 研究补助金 JP24KJ0674 的支持。

[1] W. Maass, Networks of spiking neurons: The third generation of neural network models, Neural Networks **10**, 1659 (1997) .

[2] K. Wiesenfeld and F. Moss, Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs, Nature **373**, 33 (1995) .

[3] J. J. Collins, C. C. Chow, and T. T. Imhoff, Stochastic resonance without tuning, Nature **376**, 236 (1995) .

[4] W. Maass, T. Natschlger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural

Computation **14**, 2531 (2002) .

[5] L. Buesing, J. Bill, B. Nessler, and W. Maass, Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons, PLoS Computational Biology **7**, e1002211 (2011) .

[6] S. Habenschuss, Z. Jonke, and W. Maass, Stochastic computations in cortical microcircuit models, PLoS Computational Biology **9**, e1003311 (2013) .

[7] S. Habenschuss, H. Puhr, and W. Maass, Emergence of optimal decoding of population codes through STDP, Neural Computation **25**, 1371 (2013) .

[8] H. A. Gonzalez, J. Huang, F. Kelber, K. K. Nazeer, T. Langer, C. Liu, M. Lohrmann, A. Rostami, M. Schöne, B. Vogginger, *et al.*, Spinnaker2: A large-scale neuromorphic system for event-based and asynchronous machine learning, arXiv preprint arXiv:2401.04491 (2024) .

[9] Z. Jonke, S. Habenschuss, and W. Maass, Solving constraint satisfaction problems with networks of spiking neurons, Frontiers in Neuroscience **10**, 118 (2016) .

[10] S. Yonekura and Y. Kuniyoshi, Spike-induced ordering: Stochastic neural spikes provide immediate adaptability to the sensorimotor system, Proceedings of the National Academy of Sciences **117**, 12486 (2020) .

[11] Y. Garipova, S. Yonekura, and Y. Kuniyoshi, Noise and dynamical synapses as optimization tools for spiking neural networks, Entropy **27**, 219 (2025) .

[12] J. C. V. Tieck, P. Becker, J. Kaiser, I. Peric, M. Akl, D. Reichard, A. Roennau, and R. Dillmann, Learning target reaching motions with a robotic arm using brain-inspired dopamine modulated STDP, 2019 IEEE 18th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC) **00**, 54 (2019) .

[13] E. Rueckert, D. Kappel, D. Tanneberg, D. Pecevski, and J. Peters, Recurrent spiking networks solve planning tasks, Scientific Reports **6**, 21142 (2016) .

[14] A. Juarez-Lora, V. H. Ponce-Ponce, H. Sossa, and E. Rubio-Espino, R-STDP spiking neural network architecture for motion control on a changing friction joint robotic arm, Frontiers in Neurorobotics **16**, 904017 (2022) .

[15] L. Zanatta, F. Barchi, S. Manoni, S. Tolu, A. Bartolini, and A. Acquaviva, Exploring spiking neural networks for deep reinforcement learning in robotic tasks, Scientific Reports **14**, 30648 (2024) .

[16] G. L. Chadderdon, S. A. Neymotin, C. C. Kerr, and W. W. Lytton, Reinforcement learning of targeted movement in a spiking neuronal model of motor cortex, PLoS ONE **7**, e47251 (2012) .

[17] D. Farina, F. Negro, and J. L. Dideriksen, The effective neural drive to muscles is the common synaptic input to motor neurons, The Journal of Physiology **592**, 3427 (2014) .

[18] D. Kappel, R. Legenstein, S. Habenschuss, M. Hsieh, and W. Maass, A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning, eNeuro **5**, ENEURO.0301 (2018) .

[19] C. J. D. Luca and Z. Erim, Common drive of motor units in regulation of muscle force, Trends in Neurosciences **17**, 299 (1994) .

[20] F. Hug, S. Avrillon, J. Ibáñez, and D. Farina, Common synaptic input, synergies and size principle: Control of spinal motor neurons for movement generation, The Journal of Physiology **601**, 11 (2023) .