

# Neural reservoir control of a soft bio-hybrid arm

Noel Naughton<sup>1,5†</sup>, Arman Tekinalp<sup>2,5‡</sup>, Keshav Shivam<sup>3,5§</sup>, Seung Hung Kim<sup>2,5</sup>, Volodymyr Kindratenko<sup>3,5</sup>, and Mattia Gazzola<sup>2,3,4,5\*</sup>

<sup>1</sup>Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, <sup>2</sup>Mechanical Science and Engineering, University of Illinois Urbana-Champaign, <sup>3</sup>National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, <sup>4</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, <sup>5</sup>The Grainger College of Engineering, University of Illinois Urbana-Champaign

A long-standing engineering problem, the control of soft robots is difficult because of their highly non-linear, heterogeneous, anisotropic, and distributed nature. Here, bridging engineering and biology, a neural reservoir is employed for the dynamic control of a bio-hybrid model arm made of multiple muscle-tendon groups enveloping an elastic spine. We show how the use of reservoirs facilitates simultaneous control and self-modeling across a set of challenging tasks, outperforming classic neural network approaches. Further, by implementing a spiking reservoir on neuromorphic hardware, energy efficiency is achieved, with nearly two-orders of magnitude improvement relative to standard CPUs, with implications for the on-board control of untethered, small-scale soft robots.

Hyper-redundancy, underactuation, distributedness, and continuum mechanics are defining features of soft robots (artificial or biological (1–8)), intrinsic to their compliant, elastic constitutive materials. These traits are attractive in the pursuit of extreme reconfigurability, morphological adaptivity, delicacy and dexterity, for applications in medicine, defense, or agriculture (9–12). However, the same advantageous traits pose fundamental challenges in control, due to the associated vast space of degrees of freedom and the highly non-linear dynamics involved.

Recognizing the importance of the problem, significant effort has been devoted to the development of control approaches suitable for continuum and elastic structures (13–15). Model-based controllers have proven effective in quasi-static settings, but lack accuracy when inertial effects become significant and typically rely on simplifying assumptions that may overlook environmental interactions, anisotropy, or material nonlinearities (14). Data-driven approaches bypass such modeling difficulties by directly learning associated dynamics via artificial neural networks (16, 17). These techniques are generally resource-intensive and may employ neural architectures poorly suited to capture the long-range spatial and temporal dependencies that define soft systems (9, 18), hindering robustness.

An alternative, bioinspired route to soft robotic control hinges on the concept of ‘mechanical intelligence’ (19–21), whereby non-linear structural and environmental effects are purposefully exploited to passively achieve adaptivity, thus simplifying and robustifying control. Complementarily, rhythm-generation techniques inspired by the nervous system (e.g., central pattern generators) have also been shown effective, particularly for locomotion (22, 23), leading to bio-hybrid attempts where neural-tissue is directly integrated with soft scaffolds, muscles, and electronics, to drive simple actuation dynamics (7, 8, 24, 25).

Here, with the goal of intimately meshing compliant mechanics and neural dynamics for enhanced adaptivity and control, we numerically explore the integration of a soft bio-hybrid arm with Reservoir Computing (RC). Partially inspired by the mammalian neocortex (26), RC was developed (27) for processing real-time, analog, spatiotemporally correlated inputs, by means of a dynamically rich substrate (28–30), the reservoir. The reservoir, which

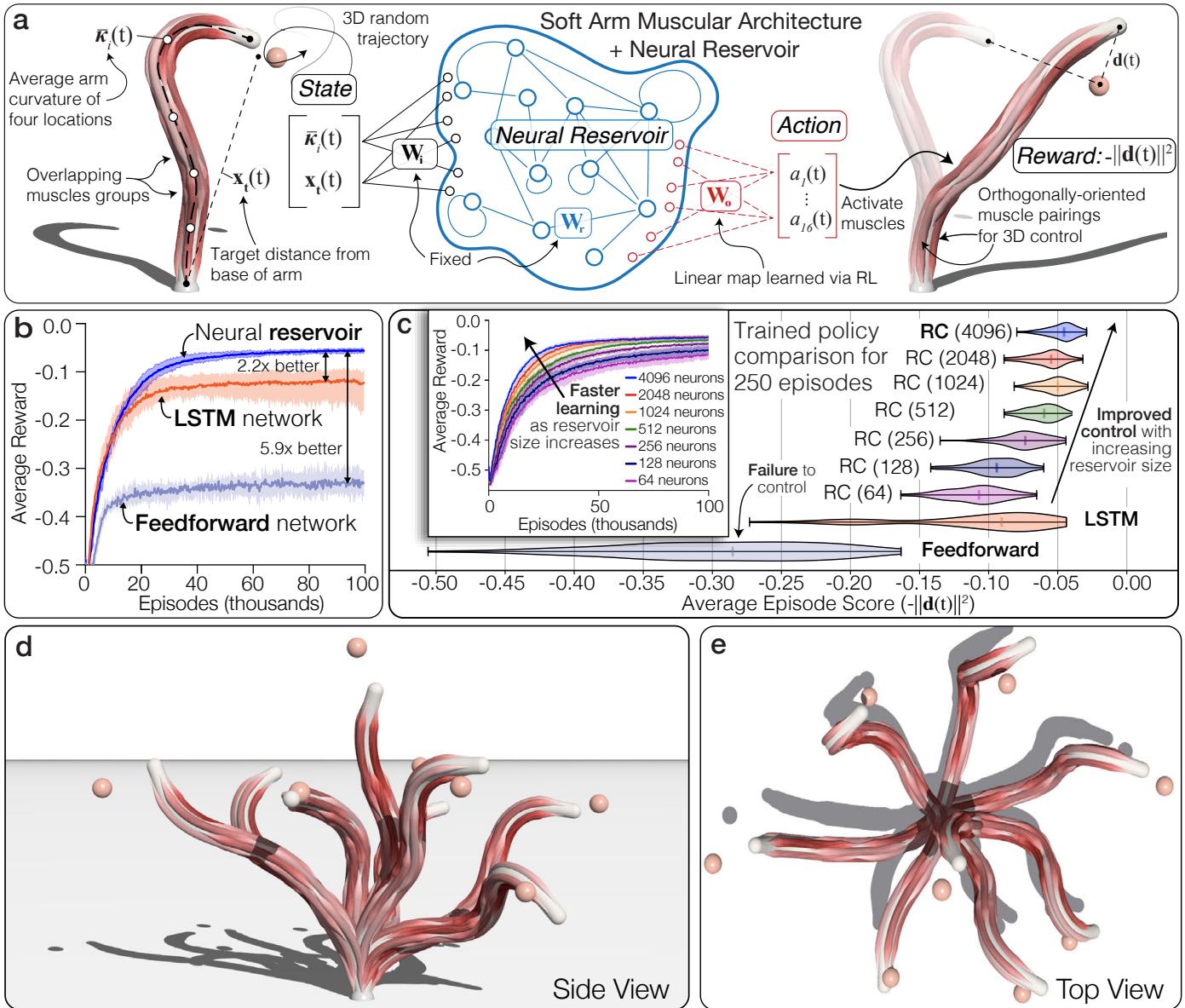
in principle can be any dynamical system (31), integrates and projects input data streams into a separable, high-dimensional latent space that decomposes non-linear correlations. Reservoir dynamics are then sampled and recombined via linear maps into desired computations. While a variety of algorithmic and physical implementations have been proposed (29, 31), in general the RC paradigm offers a number of appealing features (27, 32): it does not require expensive backpropagation (only output linear maps are learned), it intrinsically accommodates nonlinear, spatiotemporally correlated dynamics, it is natively parallel as multiple maps can be learned separately to achieve different tasks while running on the same reservoir, and can be matched with specialized hardware (e.g., neuromorphic systems for energy efficiency (33, 34)) or ‘wetware’ (neural tissue used as bio-hybrid reservoir (35)).

In this work, we consider a neural reservoir (modeled as a recurrently connected or spiking network) able to sense the shape (proprioception) of a simulated bio-hybrid arm made of muscles and tendons enveloping an elastic spine. Neural dynamics are coupled to muscle actuations through Reinforcement Learning (RL), to learn to control the system’s musculature in an unsupervised manner. We demonstrate that our approach produces control policies that outperform traditional deep-learning methodologies, with advantages seen to widen as the arm’s compliance increases. RC’s high-dimensional latent space is also leveraged, through parallel output maps, for concurrent self-modeling to improve control robustness during conditions of disturbance and sensing failure. Further, motivated by the potential relevance in bio-hybrid contexts, spiking neural reservoirs are considered, and subsequently mapped onto neuromorphic hardware, attaining a seventy-five fold reduction in energy usage. Neuromorphic RC is finally employed to drive the arm through a set of unstructured obstacles, learning to exploit solid objects to reshape and facilitate the reaching of a target — a hallmark of mechanical intelligence that is automatically recovered here.

Overall, this work not only advances soft robotic control, but also furthers neuro-mechano integration, a characteristic trait of organic systems. This, in turn, may lead to novel bio-hybrid designs bridging engineering and biology (7, 8, 25, 35–37) as well as new insights into biological processing.

**Modeling of a soft bio-hybrid arm.** While no soft, synthetic actuator has yet reached biologically comparable levels of performance, improved artificial (38, 39) or tissue-bioengineered

<sup>†</sup>Current address: Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA. <sup>‡</sup>Current address: Department of Mechanical Engineering, University of Maryland, College Park, MD. <sup>§</sup>Current address: Google, Mountain View, CA, \*email: mgazzola@illinois.edu



**Figure 1.** Neural reservoir control of a soft bio-hybrid arm. (a) Schematic depicting the coupling of a neural reservoir with a soft bio-hybrid arm made of sixteen muscle-tendon units enveloping an elastic spine to achieve autonomous control. (b) Learning performance for the control of a soft bio-hybrid arm (backbone stiffness: 125 kPa) tracking a 3D moving target: neural reservoir and traditional feedforward and LSTM network architectures. The solid lines depict the average learning performance (average episode return), each obtained by considering five neural architectures instances initialized using different random number generator seeds. The shaded regions depict the spread in learning performance of the five architectures with different initial seeds. (c) Violin plots of average performance of trained policies evaluated over 250 episodes, for different reservoir sizes and reference architectures (FF and LSTM). Inset shows learning performance as reservoir size increases, illustrating how both more rapid and better overall learning performance is achieved as the reservoir size increases. For FF and LSTM, different numbers of neurons and layers / stacks were considered (FF: [64×64], [128×128], [64×64×64], [64×64×64×64]; LSTM: [64×1 stack], [128×1 stack], [256×1 stack], [64×2 stacks], [128×2 stacks], [256×2 stacks]), with no significant differences observed. Here we report data for the best performing networks (SI for full details). Overlaid snapshots of (d) side view and (e) top view of the bio-hybrid arm successfully tracking a 3D target (see also SI Video 1) using a trained neural reservoir control policy. Intensity of muscle color (pale to dark red) denotes a muscle's activation level.

(8) prototypes are continually being developed. Cognizant of this rapid development, we consider the automated, dynamic control of a computational neuromuscular arm, characterized by multiple, antagonistic and overlapping muscle-tendon pairs arranged around and along an elastic supporting spine (Fig. 1a).

Such a design captures key organizational motifs commonly encountered in biology and, increasingly, in robotics (21, 23). Antagonistic muscles are, for example, ubiquitous in both limbed (40) and limbless animals (41, 42), and popular in bio-hybrid robots (2, 4, 8, 43). Their asymmetric contractions enable rota-

tions (i.e., joints (40)) or continuum bending and twisting (e.g., octopus arms (41, 44)), while graded co-contractions can improve structural stability (45, 46) by actively increasing body stiffness. Similarly, overlapping muscle groups (of which the many staggered spinal muscles of snakes provide an extreme example (42)) enable precise, distributed, and efficient actuation (47). Our prototypical model arm then allows us to explore compliant actuation and control in a biologically and engineering relevant setting, while offering a generalizable blueprint for experimental implementations. For example, the elastic spine

may be realized at various scales via molding or 3D printing, and decorated with arrays of compliant actuators, from advanced supercoiled artificial muscles (48) to cultured muscles (8, 49).

To computationally instantiate our arm, we employ a modeling approach based on assemblies of Cosserat rods (Methods, (50, 51)). These are slender, one-dimensional elements able to undergo all modes of deformation —bend, stretch, twist, and shear—to dynamically reconfigure in three-dimensional space. Cosserat rods are a convenient representation since they naturally map to elastic beams (spine), tendons, and muscles, can actively contract and stretch along their length according to prescribed force-length relations, and can be connected together (via appropriate boundary conditions) into arbitrary architectures. They are thus well-suited to capture the heterogeneous, anisotropic, and distributed nature of our bio-hybrid arm. Assemblies of rods and their governing equations are numerically discretized and solved via the open-source software Elastica (52), whose quantitative utility has been demonstrated in a range of biophysical settings, including animal locomotion (47, 53, 54), invertebrate manipulation (44), plant dynamics (55), fibrous metamaterials (56, 57), and soft robotic design and control (8, 47, 49, 58–63).

We begin with the soft elastic spine/backbone, represented as a single passive rod, onto which 16 muscle-tendon units are patterned (Fig. 1a, Fig. S1). Three-dimensional arm deformations are achieved by organizing muscle-tendon units into four layers of orthogonal agonist-antagonistic pairs, with adjoining layers overlapping by 50% of their length and offset 45° to avoid intersection. This architecture enables bending in all directions, via the coordinated activation of orthogonal muscles combined with the omni-directional bending/twisting capability of the elastic backbone. Further, muscle overlap allows for continuously graded contractions along the arm despite the limited number of individual muscles (47). Each muscle-tendon unit is modeled as an individual rod with an actively-contractible muscle belly and tapered tendinous ends that insert into the backbone. The entire unit is glued lengthwise to the spine, to conform to it while bending, mimicking the presence of surrounding fascia tissue or encapsulating materials.

The active force-length and force-velocity relationships of the muscle belly and its passive hyperelastic material properties as well as the tendon's material properties are based on reported biomechanical data (Methods). Noteworthy is the non-linear passive stress-strain behavior of muscle and tendon tissues (64, 65), which are characterized by a ‘J-curve’ response that allows for compliant deformations at small strains, before stiffening at larger strains for structural stability (66). This is a key mechanical feature of biological musculoskeletal tissues, and significant engineering effort has been recently devoted to recapitulate similar properties (67–69). Finally, each muscle is independently controllable via a continuous tetanic activation ( $a_i \in [0, 1]$ ,  $i \in \{1, \dots, 16\}$ ). Full details are provided in Methods.

We challenge the arm to learn to coordinate its muscle activations (action space  $\mathbf{a}_t(t)$ ) to dynamically track a target moving along a smoothly-varying, random 3D trajectory (Fig. 1a). First, we outfit the arm with a compact set of sensory capabilities. Analogous to strain-sensing muscle spindles in vertebrates (70),

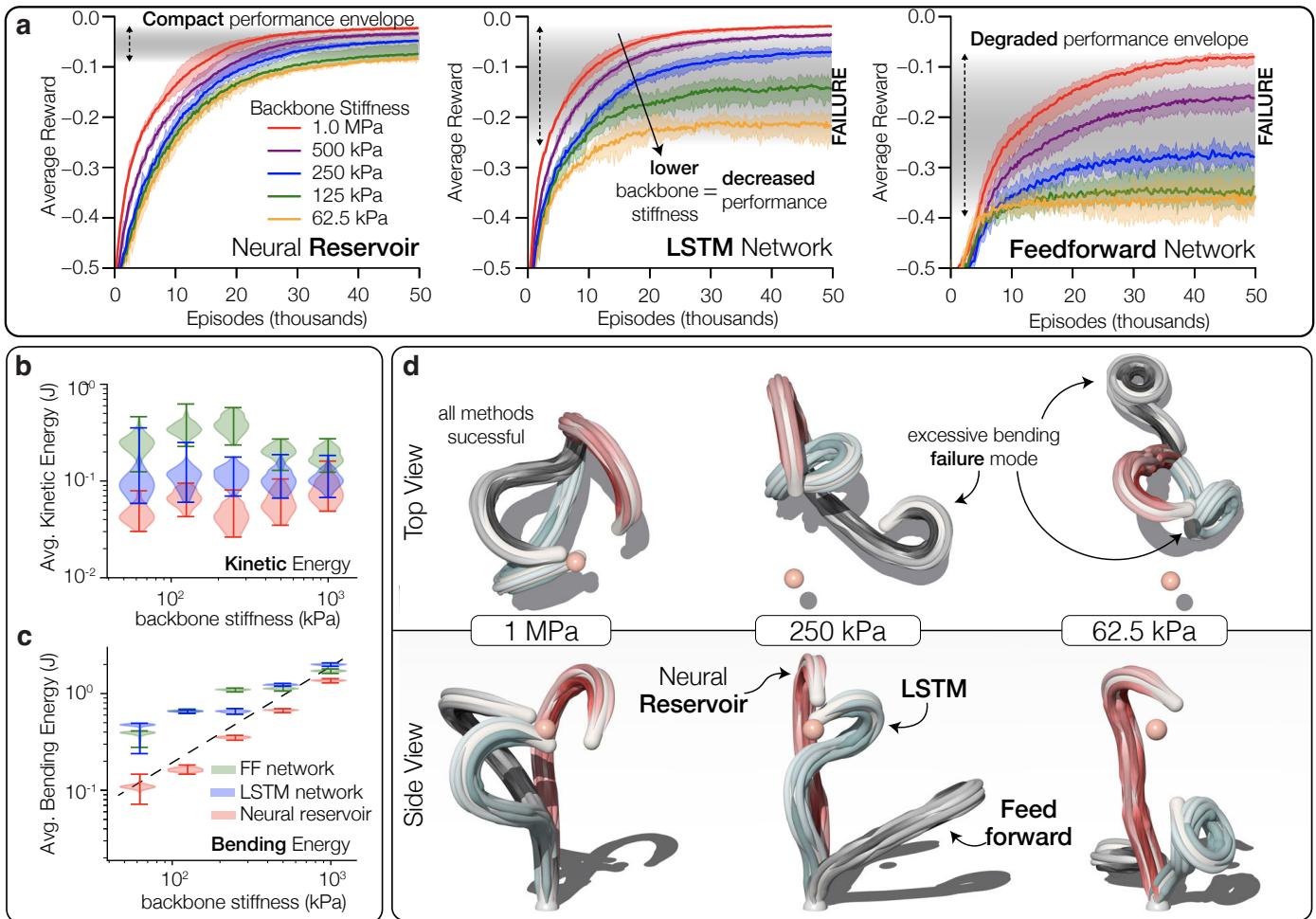
proprioception is provided here through the arm's curvature  $\bar{\kappa}_i(t)$ , estimated at four equally spaced locations (Methods). Environmental information is limited to the distance vector  $\mathbf{x}_t(t)$  between the (fixed) base of the arm and the moving target, representing visual or acoustic tracking. Notably, this sensory arrangement is laboratory-frame invariant. Indeed, local curvature is sufficient to recover the arm shape, while the target is sensed relative to the arm base. Thus, overall, the arm state is constituted by  $\bar{\kappa}_i(t)$  and  $\mathbf{x}_t(t)$ , with all higher-order information (velocities or accelerations) implicitly captured through the memory of the reservoir (29), described next.

**Neural reservoir control.** Controlling this structure is challenging because of the arm's continuum elastic nature, characterized by highly non-linear and long-range stress propagation effects. Indeed, distributed and localized loads (musculation/environment) are communicated throughout the entire system, potentially leading to mechanical instabilities and global morphological reconfigurations. Compounding this complexity are the non-linear contractile properties of the muscle-tendon units, the anisotropic and heterogeneous quality of the arm, as well as its inertial dynamics.

To establish control (Fig. 1a), we seek to learn, in an unsupervised, model-free fashion, muscle activations (action:  $a = \{a_j(t)\}$ ,  $j \in \{1, \dots, 16\}$ ) that minimize at all times the distance  $\mathbf{d}(t)$  between the arm tip and the target (reward:  $r = -||\mathbf{d}(t)||^2$ ), based on available sensory information (state:  $s = \{\mathbf{x}_t(t), \bar{\kappa}_i(t)\}$ ,  $i \in \{1, \dots, 4\}$ ). The mapping (control policy) between state and action space may be represented by any neural network topology (71, 72), and it is here that we insert our neural reservoir (Fig. 1a). This is initially realized as a recurrent artificial neural network (Methods) paired with a single, external linear recombination layer (map) whose function is to transform the reservoir's internal dynamics (sampled at each neuron) into muscle contractions  $a = \{a_j(t)\}$ . To learn this linear map, we employ the reinforcement learning (on-policy) proximal policy optimization (PPO) strategy (73). We emphasize here that since RC requires learning the output map weights only, all connections from the input state to the reservoir and within the reservoir itself, remain fixed, removing the need for backpropagation.

To contextualize the reservoir performance, we additionally consider feedforward (FF) and long short-term memory (LSTM) networks, reference architectures commonly employed in learning-based control (74, 75). Training is again performed using PPO, where this time the network weights are all learned (as opposed to RC) through backpropagation. To facilitate a meaningful comparison, all network topologies are evaluated against the same control problem, and setup/trained according to the same protocol (Methods).

As illustrated in Fig. 1b-e and SI Video 1, upon training, the neural reservoir successfully learns to track the target, substantially outperforming the FF and LSTM networks. Indeed, the FF network is found to quickly reach a (poor) performance limit beyond which it is unable to improve (Fig. 1b). The LSTM network instead converges to a better performance, although the ceiling remains significantly below the reservoir's (which is 2.2x better). The neural reservoir's performance is also found to correlate



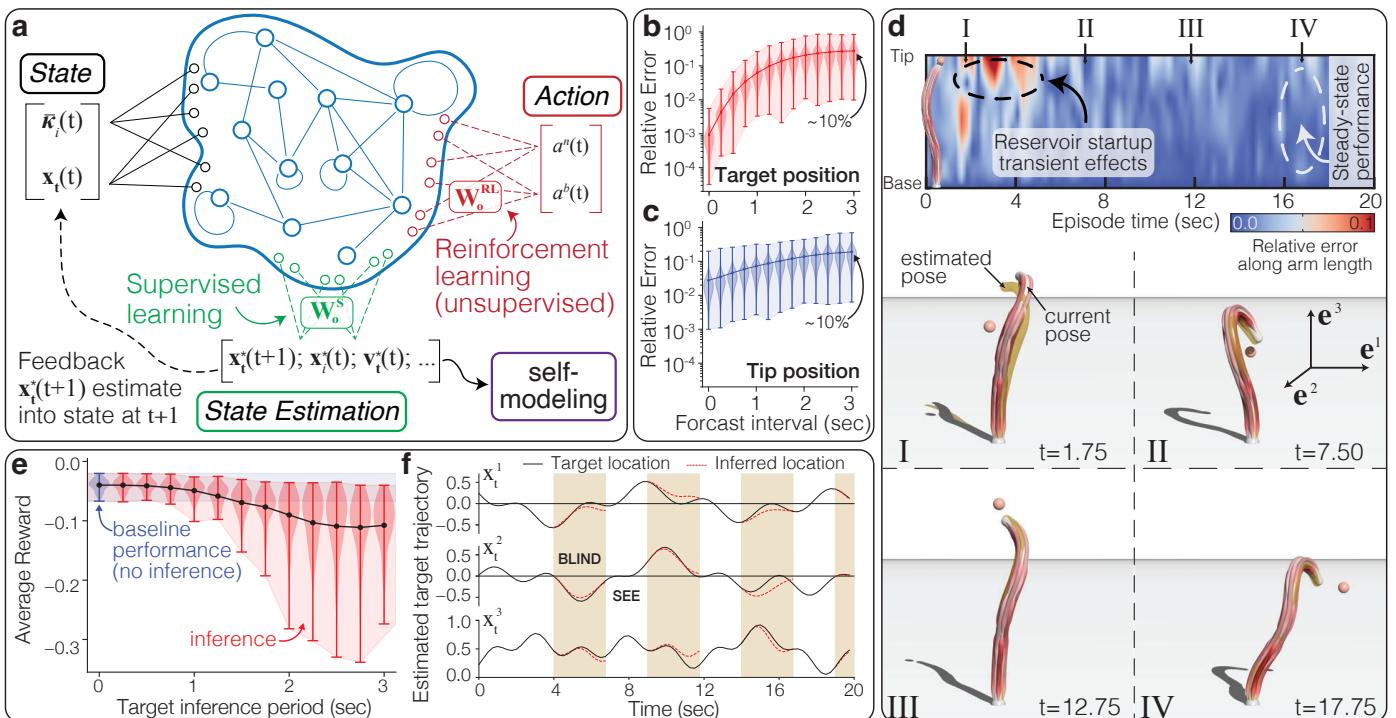
**Figure 2. Control of increasingly compliant bio-hybrid arms.** (a) Learning performance of a neural reservoir (4096 neurons) compared to LSTM and FF networks for a soft bio-hybrid arm with decreasing backbone elastic moduli. The neural reservoir exhibits a compact performance envelope while the LSTM and FF networks exhibit much wider performance envelopes with control performance drastically decreasing as the backbone softens. Solid and shaded regions denote the performance average and spread relative to five randomly initialized network instances (as in Fig. 1b). Violin plots of the average (b) kinetic and (c) bending energy of the arm's spine, when RC, LSTM, and FF network architectures are used for controlling systems of decreasing backbone stiffness. Values reported are total kinetic/bending energy integrated over the episode and normalized by episode length. The RC network exhibits lower energies in all cases. Bending energy is observed to decrease linearly in the case of RC, in keep with the linear softening of the spine and in contrast with FF and LSTM, which exhibits sublinear scalings. (d) Snapshots of trained policy performance for different backbone stiffness levels illustrating control failure modes. While all three network architectures successfully control a soft arm with a backbone stiffness of 1 MPa, for a stiffness of 250 kPa the feedforward network fails to coordinate muscle contractions, demonstrating an excessive bending failure mode (with an associated increased bending energy compared to RC, see panel (b)). While The LSTM network exhibits better performance, it also produces excessive bending and fails as the backbone stiffness continues to decrease (62.5 kPa). Video comparison of tracking performances for all backbone stiffness cases is available in SI Video 2. As in Fig. 1, for FF and LSTM, different numbers of neurons and layers / stacks were considered, with no significant differences observed across arm's stiffnesses. Here we report data for the best performing networks (SI for full details).

with reservoir size, whereby learning speed and tracking accuracy increase as the number of neurons grows (Fig. 1c). This is in keeping with the known ability of larger reservoirs to capture complex dynamics (29, 76), and supports the hypothesis that the control of elastic systems may benefit from a specific focus on highly non-linear, spatiotemporally correlated dynamics.

To further test this hypothesis, we investigate the effect of the arm's elastic spine stiffness by progressively decreasing its elastic modulus (while holding mass and geometry constant). This softening enables larger modes of deformations, further exacerbating the system's non-linear response, rendering control harder and allowing us to tease out differences between RC, FF and LSTM approaches. We begin with a relatively stiff backbone (1 MPa, equivalent to rubber (77) and  $\sim 10$ x stiffer than Fig. 1b) for which all three network topologies successfully learn to track

the moving target (Fig. 2a, SI Video 2). However, as soon as the spine softens, differences start to manifest. The learned FF controllers are indeed observed to rapidly degrade, practically failing to coordinate the arm already at 500 kPa (SI Video 2). The LSTM network instead initially maintains control, although below 250 kPa its performance too rapidly degrades (SI Video 2). In contrast, the neural reservoir exhibits a compact performance envelope and maintains sustained control down to a 62.5 kPa stiffness (comparable to mammalian (78) or octopus (79) muscle tissue (78), for reference).

An energetic analysis reveals that the neural reservoir consistently minimizes (relative to FF and LSTM, Fig. 2b,c) both the kinetic and bending energy of the spine. For each network architecture, kinetic energy (Fig. 2b) is found to remain approximately consistent across Youngs' moduli, albeit settling at dif-



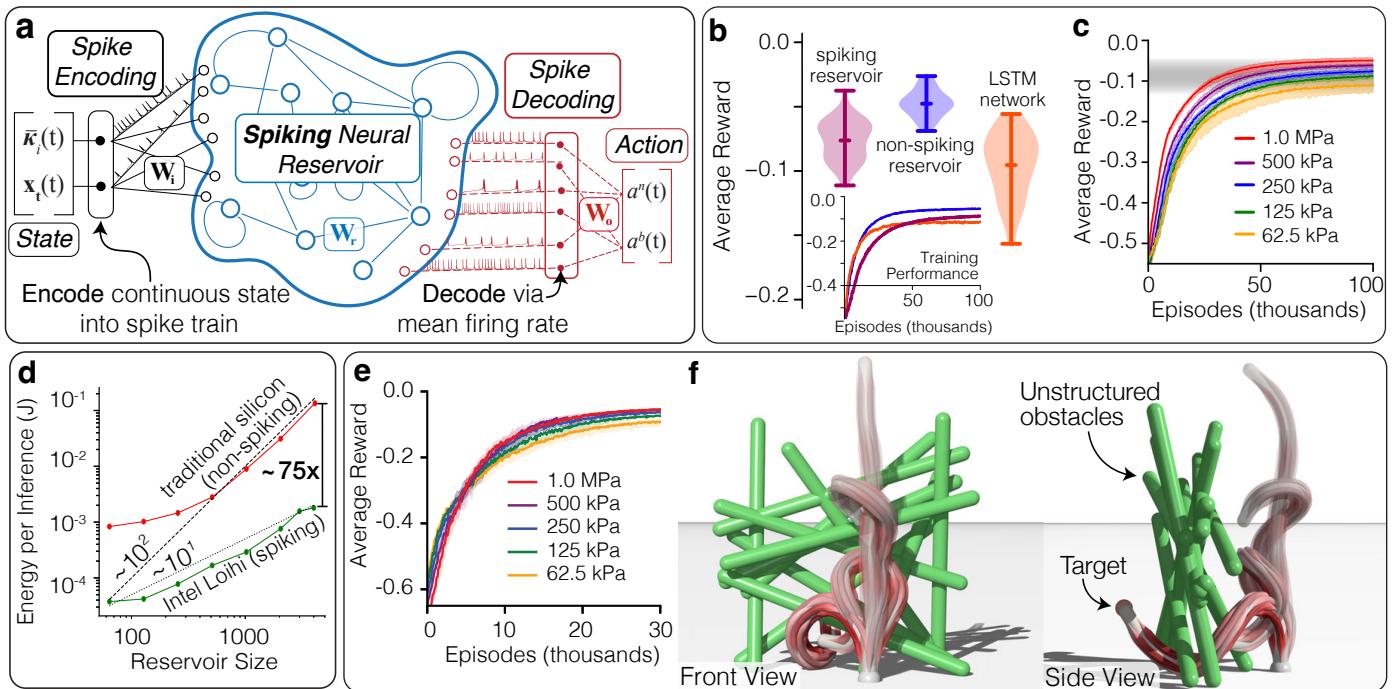
**Figure 3. Parallel maps for self-modeling and robust control.** (a) Schematic of neural reservoir control equipped with additional, parallel maps to infer/predict state information. All reported results are for an arm with a backbone stiffness of 250 kPa controlled by an already trained (via RL) neural reservoir with 4096 neurons, as described in Fig. 1. (b) Accuracy of parallel map estimates of future target positions, for increasing time-windows into the future. (c) Accuracy of parallel map estimates of future arm's tip position for increasing time-windows into the future. For both (b) and (c), the relative error is defined as  $\|\hat{y} - y\|/\ell$  where  $\|\cdot\|$  is the L2-vector norm,  $\hat{y}$  is the predicted position,  $y$  is the true position, and  $\ell$  is the length of the soft arm. (d) Performance of reservoir self-modeling for estimation of current arm pose (top row). Heat map of accuracy of pose estimation along the length of the arm. Color denotes relative error of estimation of the position of a point  $s \in [0, \ell]$  along the soft arm of length  $\ell$ . Relative error is defined as  $\|\hat{y}(s) - y(s)\|/s$  where  $\hat{y}(s)$  is the predicted location of point  $s$  and  $y(s)$  is the true location of that point. Accuracy is initially lower due to transient startup effects in the reservoir before reaching a consistent level of high accuracy as confirmed (mid/bottom row) by visualization of the estimated (gold) and true (full color) arm pose at selected time instances. (e) Tracking performance of the neural reservoir when the target position becomes unavailable for increasing lengths of time. The target oscillates between being measured ('seen') and inferred ('blind') for equal time intervals that range in length from 0.25 seconds to 3 seconds. Violin plots show performance over 50 trials for increasing periods of time during which the arm is blinded. The blue shaded region shows the baseline performance of the reservoir when the target location is always known. (f) Comparison of inferred target's 3D position for a three-second inference period. After approximately 1.5 seconds, the estimate of the target trajectory sometimes drifts from the true trajectory, leading to the drop in tracking performance seen in panel (e).

ferent levels, with LSTM and FF entailing  $\sim 3x$  and  $\sim 10x$  larger energies than RC, respectively. Phenomenologically, higher kinetic energies manifest as faster swings or pronounced structural vibrations (SI Video 2), and thus the RC-based controller (which produces reduced kinetic energies) exhibits more precise and smoother actuation, without extensively relying on corrective muscle contractions. The neural reservoir also demonstrates clear superiority in minimizing the spine's bending energy (Fig. 2c). It is worth noting that bending stiffness scales linearly with the elastic modulus. We would thus expect that a controller able to effectively coordinate the arm would produce bending deformations (and therefore energies) congruent with this linear scaling as the spine softens. This trend is indeed met by RC, while LSTM and FF networks are found to exhibit a sub-linear decrease. This indicates that LSTM and FF networks resort to deformations that are excessive in relation to the spine's stiffness, eventually leading to failure (SI Video 2). Recorded muscle activation patterns (SI Fig. S2) suggest that RC's greater control may be enabled by the use of antagonistic muscle co-contractions, to locally stiffen and stabilize the arm as it softens. LSTM and FF controllers instead do not discover this strategy, and forgo co-contractions in favor of one-sided activations.

Therefore, RC-based controllers are shown to achieve higher tracking accuracies and, concurrently, avoid unnecessary bending, twisting, or corrective accelerations, reflecting their ability to capture (and predict) the arm's dynamics, as quantified next.

**Self-modeling for robust control.** Reservoir dynamics can also be used for self-modeling (80, 81), that is, for the explicit estimation or prediction of the system's configuration, which in turn can be used to robustify the controller. Self-modeling is achieved here through three dedicated maps (Fig. 3a), to extract the arm's full pose, future target location, and future tip location. Maps are obtained in a supervised fashion via Ridge regression based on data from the 250kPa case of Fig. 2a (Methods). The maps are then directly appended to the reservoir, exploiting RC's intrinsic parallelism whereby several computations can be carried out concurrently using the same underlying dynamics.

We show how the reservoir, using target location and curvature proprioception history, accurately synthesizes future target locations (Fig. 3b), current arm pose (Fig. 3d, SI Video 3), and future tip positions (by anticipating forthcoming muscle activations selected by the control policy based on the predicted target trajectory, Fig. 3c). We note that in all cases target trajectories are randomly generated so that the reservoir must constantly



**Figure 4. Spiking neural reservoir control on neuromorphic hardware** (a) Schematic showing how the state is encoded into a spike train (SI) that is sent to a spiking neural reservoir running on an Intel Loihi chip. Spike train outputs are then decoded (SI) into continuous actions (contractions of the arm muscles). (b) Comparison between the performance over 250 episodes of a trained spiking neural reservoir on Loihi (backbone stiffness: 125 kPa; reservoir size: 2048 neurons, which is the largest implementable on Loihi), a neural reservoir running on traditional silicon using non-spiking artificial neurons, and an LSTM network (same as Fig. 1 and Fig. 2). Inset shows the training performance for the spiking reservoir, the non-spiking reservoir, and the LSTM. (c) The spiking reservoir exhibits a compact performance envelope as the backbone stiffness decreases. All cases are trained using five random initialization seeds as in Fig. 1b. (d) Energy use of a non-spiking reservoir running on traditional silicon (Intel Xeon W-2665) and the spiking reservoir running on the Intel Loihi chip. The non-spiking reservoir on traditional silicon exhibits quadratic energy scaling as the reservoir size increases compared to the linear energy scaling of the spiking reservoir running on the Intel Loihi chip. (e) Initial 30k episodes of learning performance of neural reservoirs ( $n=1024$ ) controlling arms of decreasing backbone stiffness tasked with reaching through a cluttered nest of obstacles to a fixed target (SI for expanded results). Video of the performance of final trained policies for all stiffness levels is available in SI Video 6. (f) Timelapse front/side views of a spiking reservoir guiding a soft arm through unstructured obstacles to reach a target (backbone stiffness: 125 kPa).

produce new predictions. These capabilities are then directly used to improve control in a scenario where the arm intermittently loses sight of the target, mimicking a faulty sensor or environmental disturbances. For the arm to keep ‘seeing’, the target’s next location is predicted and fed back into the reservoir in lieu of the missing data. This in turn enables the forecasting of muscle actuations and future tip positions, allowing the blinded arm to continue tracking (Fig. 3e,f). Using this strategy, inference periods of less than one second (interval during which the arm is ‘blind’, approximately equivalent to the target traveling a distance of 50% the arm’s length) result in no performance degradation relative to perfect knowledge of the target location. If the disturbance lasts beyond one second, then tracking quality begins to decrease, although modestly. Indeed, average performance remain comparable to an LSTM network with complete knowledge of the target location (SI). However, the spread of the arm performance (violin plot bars in Fig. 3e) increases for large disturbance periods, implying that the arm does occasionally lose track of the target while blinded.

Thus, a mixed strategy where unsupervised learning is used for control and supervised learning is employed for self-modeling, with the second supporting the first, is found to be effective in robustifying performance (SI Video 4). It is worth noting that the inclusion of self-modeling entails minimal computational overhead since all linear maps use the same reservoir.

**Spiking reservoirs on neuromorphic hardware.** Here, we illustrate how our approach is also well-positioned to enhance computational energy efficiency, a concern in many robotic applications (82, 83). To this end, we consider a spiking neural reservoir on neuromorphic hardware, demonstrating almost two orders of magnitude improvements relative to standard CPUs, while retaining control performance.

The use of spiking networks, inspired by cortical dynamics (29, 84), well-aligns with both RC (27) and specialized neuromorphic chips (33, 34) that physically implement energy efficient artificial spiking neurons (85, 86). It is thus only natural to explore the use of spiking networks within our framework, also envisioning potential bio-hybrid applications where spiking neural tissue may replace synthetic reservoirs.

To match algorithmic and hardware infrastructures, we consider a reservoir of leaky integrate and fire (LIF) spiking neurons on Intel’s neuromorphic Loihi chip (33) (Fig. 4a, Methods). This spiking reservoir is challenged to control our soft arm and is compared with the LSTM and the non-spiking reservoir approaches of Fig. 1, which were run on traditional CPUs (Intel Xeon).

As can be seen in Fig. 4b, both reservoir-based approaches outperform the LSTM network in the target-tracking test. While the non-spiking reservoir exhibits slightly better overall performance, inspection of the trained policies (SI Videos 1 & 5) shows that both approaches are in fact successful in controlling the soft

arm, which is instead found to frequently collapse (as captured by the larger score spread) in the LSTM case. Further, consistent with the non-spiking reservoir, the spiking reservoir exhibits a compact performance envelope as the arm softens (Fig. 4c), and larger reservoirs improve performance (SI Fig. S4).

However, the spiking reservoir on the neuromorphic Loihi chip provides a clear advantage when energy use is considered (Fig. 4d). Indeed, power consumption is observed to reduce by up to 75x. Furthermore, and critically, energy use on Loihi exhibits linear scaling with reservoir size compared to quadratic scaling for reservoirs deployed on traditional silicon hardware. This implies that larger (thus more capable) spiking reservoirs may be used in energy-constrained applications.

**Environmental interactions.** A key advantage of soft robots is their ability to comply with the environment without damaging themselves or their surroundings. We explore these boundary interactions, and the possibility of taking advantage of them, by considering the reaching of a target through a set of unstructured obstacles (Fig. 4e,f). We underscore the steep control challenge associated with this test, exacerbated by a complex environment, geometric constraints, and contact dynamics.

In Fig. 4e,f, we show results obtained using a spiking neural reservoir for arms characterized by spines of decreasing stiffness. As can be seen, all arms successfully learn to reach through the obstacles with softer ones exhibiting faster initial learning, ostensibly due to their ability to better conform to obstacles and push through gaps without precise actuation. However, the more precise control associated with stiff arms is found to eventually achieve better performance (Fig. 4e). In all cases though, throughout training, arms are observed to extensively exploit (instead of avoiding) solid boundaries, leaning against them to passively reconfigure and favorably redirect the tip (SI Video 6). This facilitates learning and control, as supported by the fact that the target can be reached even by the softest arm (see Video), despite being the most difficult to coordinate (Fig. 2). Thus, our RC approach is found to naturally recover a hallmark of mechanical intelligence, whereby passive mechanics and the environment are leveraged to simplify the task at hand.

**Discussion.** In this work, we consider the challenging problem of learning to control a heterogeneous, soft bio-hybrid arm. The arm’s musculoskeletal architecture is designed to encompass a range of biomechanically relevant elements and is modeled via assemblies of Cosserat rods. The physical infrastructure is matched with a neural infrastructure (the reservoir), to encode the arm’s sensing and spatiotemporal dynamics. This representation is coupled with reinforcement learning to achieve muscle coordination and allow the arm to track and reach targets.

Our approach is demonstrated to be effective, outperforming classic neural network methods. In particular, the synergy between neural and physical dynamics is observed to produce control policies that minimize (relative to FF and LSTM) the arm’s bending and kinetic energies, across a range of material stiffnesses. In contrast, FF and LSTM approaches are found unable to cope with material property changes. Parallel self-modeling on the same reservoir is also demonstrated for robust control, in the case of unavailable or corrupted sensory informa-

tion. Further, spiking neural reservoirs on neuromorphic hardware are shown to deliver over 75-fold energy savings relative to non-spiking reservoirs, while conserving control performance. Spiking reservoirs are finally employed to drive the arm through a set of unstructured obstacles, taking advantage of solid boundaries and compliance, a hallmark of mechanical intelligence.

The capacity of neural reservoirs to synthesize compliant dynamics presents much potential for soft robotic control, combining learning, versatility, adaptivity, and robustness to mechanical changes or environmental disturbances. Higher-order situational information (velocities or accelerations) can be extracted with minimal additional computational costs (parallel maps running on the same reservoir), reducing onboard sensing and energy requirements. This aspect is compounded by the use of spiking neural reservoirs on neuromorphic hardware, to deliver significant energy savings, a necessity in many robotic applications.

These results not only advance soft robotic control, but may also further the use of biological constructs in bio-hybrid robotics and engineering, with both spike-activated cultured muscle tissue (6–8, 49) and living reservoirs of cultured neural tissue (35) amenable to integration within this computing paradigm.

## References

1. L. Ricotti *et al.*, *Science robotics* **2**, eaqq0495 (2017).
2. Y. Morimoto, H. Onoe, S. Takeuchi, *Science robotics* **3**, eaat4440 (2018).
3. T. Li, S. Takeuchi, *Biophysics Reviews* **6** (2025).
4. X. Ren, Y. Morimoto, S. Takeuchi, *Science Robotics* **10**, eadr5512 (2025).
5. S. Park *et al.*, *Science* **353**, 158–162 (2016).
6. G. J. Pagan-Diaz *et al.*, *Advanced Functional Materials* **28**, 1801145 (2018).
7. O. Aydin *et al.*, *Proceedings of the National Academy of Sciences* **116**, 19841–19847 (2019).
8. Y. Kim *et al.*, *Science Robotics* **8**, eadd1053 (2023).
9. O. Yasa *et al.*, *Annual Review of Control, Robotics, and Autonomous Systems* **6**, 1–29 (2023).
10. P. Polygerinos *et al.*, *Advanced engineering materials* **19**, 1700016 (2017).
11. M. Cianchetti, C. Laschi, A. Menciassi, P. Dario, *Nature Reviews Materials* **3**, 143–153 (2018).
12. G. Chowdhary, M. Gazzola, G. Krishnan, C. Soman, S. Lovell, *Sustainability* **11**, 6751 (2019).
13. M. Russo *et al.*, *Advanced Intelligent Systems* **5**, 2200367 (2023).
14. C. Della Santina, C. Duriez, D. Rus, *IEEE Control Systems Magazine* **43**, 30–65 (2023).
15. T. George Thuruthel, Y. Ansari, E. Falotico, C. Laschi, *Soft robotics* **5**, 149–163 (2018).
16. Z. Chen *et al.*, *IEEE Transactions on Automation Science and Engineering* (2024).
17. E. Falotico *et al.*, *Advanced Intelligent Systems*, 2400344 (2024).
18. G. Mengaldo *et al.*, *Nature Reviews Physics* **4**, 595–610 (2022).
19. R. Pfeifer, *How the body shapes the way we think: A New View of intelligence*, 2006.
20. N. T. Ulrich, “Grasping with mechanical intelligence”, tech. rep.
21. T. Wang *et al.*, *Science Robotics* **8**, eadi2243 (2023).
22. A. J. Ijspeert, *Neural networks* **21**, 642–653 (2008).
23. P. Ramdy, A. J. Ijspeert, *Science Robotics* **8**, eadg0279 (2023).

24. O. Aydin *et al.*, *APL bioengineering* **4**, 016107 (2020).
25. H. Tetsuka, S. Gobbi, T. Hatanaka, L. Pirrami, S. Shin, *Science Robotics* **9**, eado0051 (2024).
26. L. F. Seoane, *Philosophical Transactions of the Royal Society B* **374**, 20180377 (2019).
27. W. Maass, T. Natschläger, H. Markram, *Neural computation* **14**, 2531–2560 (2002).
28. L. Gonon, J.-P. Ortega, *IEEE transactions on neural networks and learning systems* **31**, 100–112 (2019).
29. M. Lukoševičius, H. Jaeger, *Computer science review* **3**, 127–149 (2009).
30. W. Maass, P. Joshi, E. Sontag, *Advances in neural information processing systems* **18** (2005).
31. G. Tanaka *et al.*, *Neural Networks* **115**, 100–123 (2019).
32. H. Jaeger, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report **148**, 13 (2001).
33. M. Davies *et al.*, *IEEE Micro* **38**, 82–99 (2018).
34. P. A. Merolla *et al.*, *Science* **345**, 668–673 (2014).
35. X. Zhang *et al.*, *Advanced Science* **11**, 2306826 (2024).
36. H. Cai *et al.*, *Nature Electronics* **6**, 1032–1039 (2023).
37. B. J. Kagan *et al.*, *Neuron* **110**, 3952–3969 (2022).
38. M. Duduta, E. Hajiesmaili, H. Zhao, R. J. Wood, D. R. Clarke, *Proceedings of the National Academy of Sciences* **116**, 2476–2481 (2019).
39. J. Wang, D. Gao, P. S. Lee, *Advanced Materials* **33**, 2003088 (2021).
40. J. Gray, *Journal of experimental biology* **20**, 88–116 (1944).
41. W. M. Kier, K. K. Smith, *Zoological journal of the Linnean Society* **83**, 307–324 (1985).
42. B. C. Jayne, *Journal of Morphology* **197**, 159–181 (1988).
43. K. Y. Lee *et al.*, *Science* **375**, 639–647 (2022).
44. A. Tekinalp *et al.*, *Proceedings of the National Academy of Sciences* **121**, e2318769121 (2024).
45. P. L. Gribble, L. I. Mullin, N. Cothros, A. Mattar, *Journal of neurophysiology* **89**, 2396–2405 (2003).
46. A. D. Koelewijn, A. J. Van Den Bogert, *PeerJ* **10**, e13085 (2022).
47. X. Zhang, F. K. Chan, T. Parthasarathy, M. Gazzola, *Nature communications* **10**, 1–12 (2019).
48. G. M. Spinks, N. D. Martino, S. Naficy, D. J. Shepherd, J. Foroughi, *Science Robotics* **6**, eabf4788 (2021).
49. J. Wang *et al.*, *Advanced Intelligent Systems* **3**, 2000237 (2021).
50. S. S. Antman, *Nonlinear Problems of Elasticity*, 513–584 (2005).
51. M. Gazzola, L. Dudte, A. McCormick, L. Mahadevan, *Royal Society open science* **5**, 171628 (2018).
52. A. Tekinalp *et al.*, *PyElastica*, Feb. 2023, (<https://doi.org/10.5281/zenodo.7658872>).
53. X. Zhang, N. Naughton, T. Parthasarathy, M. Gazzola, *Nature communications* **12**, 6076 (2021).
54. A. Tekinalp, Y. Bhosale, S. Cui, F. K. Chan, M. Gazzola, *arXiv preprint arXiv:2401.09506* (2024).
55. A. Porat, A. Tekinalp, Y. Bhosale, M. Gazzola, Y. Meroz, *Proceedings of the National Academy of Sciences* **121**, e2312761121 (2024).
56. Y. Bhosale *et al.*, *Physical review letters* **128**, 198003 (2022).
57. N. Weiner, Y. Bhosale, M. Gazzola, H. King, *Journal of Applied Physics* **127** (2020).
58. N. Naughton *et al.*, *IEEE Robotics and Automation Letters* **6**, 3389–3396 (2021).
59. H.-S. Chang *et al.*, presented at the 2020 59th IEEE Conference on Decision and Control (CDC), pp. 3913–3920.
60. H.-S. Chang *et al.*, *Proceedings of the Royal Society A* **479**, 20220593 (2023).
61. T. Wang *et al.*, presented at the 2022 IEEE 61st Conference on Decision and Control (CDC), pp. 1059–1066.
62. C.-H. Shih *et al.*, *Advanced Intelligent Systems* **n/a**, 2300088.
63. N. Charles, R. Chelakkot, M. Gazzola, B. Young, L. Mahadevan, *arXiv preprint arXiv:2303.15482* (2023).
64. B. Calvo *et al.*, *Journal of biomechanics* **43**, 318–325 (2010).
65. C. N. Maganaris, J. P. Paul, *In vivo human tendon mechanical properties*, 1999.
66. R. F. Ker, *International Journal of Fatigue* **29**, 1001–1009 (2007).
67. M. Vatankhah-Varnosfaderani *et al.*, *Nature* **549**, 497–501 (2017).
68. Z. Tu *et al.*, *Nature communications* **12**, 2916 (2021).
69. W. L. Lim, L. L. Liau, M. H. Ng, S. R. Chowdhury, J. X. Law, *Tissue engineering and regenerative medicine* **16**, 549–571 (2019).
70. P. Matthews, *Physiological Reviews* **44**, 219–288 (1964).
71. M. T. Hagan, H. B. Demuth, presented at the Proceedings of the 1999 American control conference (cat. No. 99CH36251), vol. 3, pp. 1642–1656.
72. A. Perrusquia, W. Yu, *Neurocomputing* **438**, 145–154 (2021).
73. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, *arXiv:1707.06347* (2017).
74. Y. Yu, X. Si, C. Hu, J. Zhang, *Neural computation* **31**, 1235–1270 (2019).
75. J. Zou, Y. Han, S.-S. So, *Artificial neural networks: methods and applications*, 14–22 (2009).
76. J. Pathak, B. Hunt, M. Girvan, Z. Lu, E. Ott, *Physical review letters* **120**, 024102 (2018).
77. K. Nagdi, *Rubber as an engineering material: guideline for users* (Hanser Verlag, 1993).
78. D. Fish, J. Orenstein, S. Bloom, *Circulation research* **54**, 267–276 (1984).
79. A. Di Clemente, F. Maiole, I. Bornia, L. Zullo, *Journal of Experimental Biology* **224**, jeb242644 (2021).
80. B. Chen, R. Kwiatkowski, C. Vondrick, H. Lipson, *Science Robotics* **7**, eabn1944 (2022).
81. Y. Hu, J. Lin, H. Lipson, *Nature Machine Intelligence*, 1–11 (2025).
82. M. Zhu, O. G. Schmidt, *MRS Bulletin* **49**, 115–124 (2024).
83. Y. Sandamirskaya, M. Kaboli, J. Conradt, T. Celikel, *Science Robotics* **7**, cabl8419 (2022).
84. P. F. Dominey, *Biological cybernetics* **73**, 265–274 (1995).
85. B. Rajendran, A. Sebastian, M. Schmuker, N. Srinivasa, E. Eleftheriou, *IEEE Signal Processing Magazine* **36**, 97–110 (2019).
86. A. Javanshir, T. T. Nguyen, M. P. Mahmud, A. Z. Kouzani, *Neural Computation* **34**, 1289–1328 (2022).
87. H. Herr, R. G. Dennis, *Journal of neuroengineering and rehabilitation* **1**, 1–9 (2004).
88. J. Van Leeuwen, *Journal of theoretical biology* **149**, 229–256 (1991).
89. J. Van Leeuwen, W. M. Kier, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **352**, 551–571 (1997).
90. R. L. Lieber, C. G. Brown, C. L. Trestik, *Journal of biomechanics* **25**, 421–428 (1992).
91. L. F. Abbott, *Brain research bulletin* **50**, 303–304 (1999).
92. N. Hansen, S. D. Müller, P. Koumoutsakos, *Evolutionary computation* **11**, 1–18 (2003).
93. T. Bekolay *et al.*, *Frontiers in Neuroinformatics* **7**, 1–13, ISSN: 1662-5196 (2014).
94. A. Raffin *et al.*, *Journal of Machine Learning Research* **22**, 1–8, (<http://jmlr.org/papers/v22/20-1364.html>) (2021).

## Acknowledgements

**Funding:** This study was jointly funded by ONR MURI N00014-19-1-2373 (M.G.), ONR N00014-22-1-2569 (M.G.), NSF EFRI C3 SoRo #1830881 (M.G.), NSF CAREER #1846752 (M.G.), NSF Expeditions in Computing ‘Mind in Vitro’ #2123781 (M.G.), and NSF EFRI BEGIN OI #2422340 (N.N.). Computational support was provided by the Bridges2 supercomputer at the Pittsburgh Supercomputing Center through allocation TG-MCB190004 from the Extreme Science and Engineering Discovery Environment (XSEDE; NSF grant ACI-1548562). **Author contributions statement:** N.N., A.T., K.S., and M.G. conceptualized and designed the research. N.N., A.T., K.S., and S.H.K. performed the research. All analyzed the data. All wrote the paper. **Competing interests:** The authors have no competing interests to declare.

## Methods

### Modeling a neuromuscular arm

We model the muscular arm as an assembly of Cosserat rods, which are slender, one-dimensional elastic structures that can undergo all modes of deformation: bending, twisting, stretching, and shearing. Representing the backbone and muscle groups as Cosserat rods entails a number of advantages. Assembled together they naturally capture the heterogeneity of muscular architectures while the Cosserat formulation can be extended to account for non-linear material properties, connectivity, active stresses, internal pressure, and environmental loads. The numerical implementation of Cosserat rods is computationally efficient as they accurately capture large 3D deformations through a one-dimensional representation, alleviating time-consuming remeshing difficulties and compute costs of 3D elasticity.

#### Dynamics of a single Cosserat rod.

An individual Cosserat rod is described by its center line position  $\bar{\mathbf{x}}(s, t) \in \mathbb{R}^3$  and an oriented reference frame of (row-wise) orthonormal directors  $\mathbf{Q}(s, t) \in \mathbb{R}^{3 \times 3} = [\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2, \bar{\mathbf{d}}_3]^{-1}$  along its length  $s \in [0, L(t)]$ , where  $L(t)$  is the current length, for all time  $t \in \mathbb{R} \geq 0$ . Any vector defined in the global *lab frame* ( $\bar{\mathbf{v}}$ ) can be transformed into the *local frame* ( $\mathbf{v}$ ) via  $\mathbf{v} = \mathbf{Q}\bar{\mathbf{v}}$  and from the local to the lab frame via  $\bar{\mathbf{v}} = \mathbf{Q}^T\mathbf{v}$ . For an unshearable and inextensible rod,  $\bar{\mathbf{d}}_3$  is parallel to the rod’s local tangent ( $\partial_s \bar{\mathbf{x}} = \bar{\mathbf{x}}_s$ ), and  $\bar{\mathbf{d}}_1$  (normal) and  $\bar{\mathbf{d}}_2$  (binormal) span the rod’s cross-section. However, under shear or extension, the rod’s tangent direction  $\bar{\mathbf{x}}_s$  and  $\bar{\mathbf{d}}_3$  are no longer the same, with the difference represented by the shear strain vector  $\boldsymbol{\epsilon} = \mathbf{Q}(\bar{\mathbf{x}}_s - \bar{\mathbf{d}}_3)$ . The gradient of the directors ( $\bar{\mathbf{d}}_j$ ) with respect to the rod’s length is defined by the curvature vector  $\bar{\boldsymbol{\kappa}}(s, t) \in \mathbb{R}^3$  through the relation  $\partial_s \bar{\mathbf{d}}_j = \bar{\boldsymbol{\kappa}} \times \bar{\mathbf{d}}_j$ . In the local frame ( $\boldsymbol{\kappa} = \mathbf{Q}\bar{\boldsymbol{\kappa}}$ ), the components of the curvature vector  $\boldsymbol{\kappa} = [\kappa_1, \kappa_2, \kappa_3]$  relate to bending ( $\kappa_1$  and  $\kappa_2$ ) and twisting ( $\kappa_3$ ) of the rod. Similarly, the gradient of the directors with respect to time is defined by the angular velocity vector  $\bar{\boldsymbol{\omega}}(s, t) \in \mathbb{R}^3$  through the relation  $\partial_t \bar{\mathbf{d}}_j = \bar{\boldsymbol{\omega}} \times \bar{\mathbf{d}}_j$ . The linear velocity of the center-line is  $\bar{\mathbf{v}}(s, t) \in \mathbb{R}^3 = \partial_t \bar{\mathbf{x}}$  while the second area moment of inertia  $\mathbf{I}(s, t) \in \mathbb{R}^{3 \times 3}$ , cross-sectional area  $A(s, t) \in \mathbb{R}$ , and density  $\rho(s) \in \mathbb{R}$  are defined based on the rod’s material properties. The dynamics of a Cosserat rod are then described as (51)

$$\partial_t^2 (\rho A \bar{\mathbf{x}}) = \partial_s (\mathbf{Q}^T \mathbf{n}) + \bar{\mathbf{f}} \quad (1)$$

$$\partial_t (\rho \mathbf{I} \boldsymbol{\omega}) = \partial_s \boldsymbol{\tau} + \boldsymbol{\kappa} \times \boldsymbol{\tau} + (\mathbf{Q} \bar{\mathbf{x}}_s \times \mathbf{n}) + (\rho \mathbf{I} \boldsymbol{\omega}) \times \boldsymbol{\omega} + \mathbf{Q} \bar{\mathbf{c}} \quad (2)$$

where Eq. 1 (lab frame) and Eq. 2 (local frame) represent the change of linear and angular momentum at every cross-section, respectively,  $\mathbf{n}(s, t) \in \mathbb{R}^3$  and  $\boldsymbol{\tau}(s, t) \in \mathbb{R}^3$  are internal forces and couples, respectively, developed due to elastic deformations and muscle contractions while  $\bar{\mathbf{f}}(s, t) \in \mathbb{R}^3$  and  $\bar{\mathbf{c}}(s, t) \in \mathbb{R}^3$  capture external forces and couples applied to the arm, respectively.

The above continuous representation is discretized into  $(n_{\text{elem}} + 1)$  nodes of position  $\bar{\mathbf{x}}_i$  that are connected by  $n_{\text{elem}}$  cylindrical elements. Linear displacements are determined by the internal and external forces acting at the nodes, while rotations are accounted for via couples applied to the cylindrical elements. Energy losses due to internal friction and viscoelastic effects are captured through use of Rayleigh potentials. The dynamic behavior of a rod is then computed by integrating the discretized set of equations, along with appropriate boundary conditions, in time via a second-order position Verlet scheme (51). In this study, we used PyElastica (52), which is an open-source, Python-based implementation of this numerical scheme.

#### Assembling a muscular arm.

We assemble our muscular arm by arranging active and passive rods into a representative muscular architecture. The arm consists of an elastic backbone (modeled as a single rod) onto which sixteen muscle-tendon units (each modeled as a single rod) are attached around the outside for a total of seventeen rods. This mechanical connectivity enables the arm to translate the one-dimensional internal contraction forces generated by individual muscles into the three-dimensional dynamic motions of the arm as a whole. A description of the arm’s geometry and material parameters is provided in the SI.

*Modeling the elastic backbone.* The backbone is modeled as a constant cross-section, elastic rod. For a linear elastic material such as the backbone, the internal forces  $\mathbf{n} = [n_1, n_2, n_3]$  are proportional to the shear strain of the rod  $\mathbf{n} = \mathbf{S}\boldsymbol{\epsilon}$  where  $\mathbf{S} = \text{diag}(\alpha_c GA, \alpha_c GA, EA)$  is the rod’s shear/stretch stiffness matrix,  $E$  is the rod’s Young’s modulus,  $G$  is the rod’s shear modulus, and  $\alpha_c$  is the Timoshenko shear correction factor (51). We additionally model internal elastic torques  $\boldsymbol{\tau}$  using the linear relationship  $\boldsymbol{\tau} = \mathbf{B}\boldsymbol{\kappa}$  where  $\mathbf{B} \in \mathbb{R}^{3 \times 3} = \text{diag}(EI_1, EI_2, GI_3)$  is the rod’s bend/twist stiffness matrix, and  $I_1$ ,  $I_2$ , and  $I_3$  are the rod’s second moments of inertia about  $\bar{\mathbf{d}}_1$ ,  $\bar{\mathbf{d}}_2$ , and  $\bar{\mathbf{d}}_3$ , respectively. Finally, zero-displacement and zero-rotation boundary conditions are defined on the node and element of the backbone nearest to the base, respectively, to anchor the arm.

*Modeling the muscle-tendon unit.* We model muscle-tendon units based on frog semitendinous muscle, in line with previous bio-hybrid demonstrations (87). Biological muscles actively generate internal forces that cause them to axially contract while both muscle and tendon tissues exhibit hyperelastic passive behavior when stretched. Both effects render a linear stress-strain treatment of the Cosserat rod’s axial stretch inaccurate. Instead, we model the axial component ( $n_3$ ) of the internal force vector  $\mathbf{n}$  by considering the active  $\sigma_a(s, t) \in \mathbb{R}$  and passive  $\sigma_p(s, t) \in \mathbb{R}$  axial stresses developed along the muscle  $n_3 = A(\sigma_a^m + \sigma_p^m)$  and the passive response of the tendon  $n_3 = A(\sigma_p^t)$  while modeling its shear components ( $n_1, n_2$ ) using the above presented linear elastic formulation. Internal elastic torques are represented us-

ing a similar approach as the elastic backbone, though we note that the non-linear effects of the muscle's axial stretch extends to encompass angular momentum via the third term in Eq. 2. Material properties (provided in the SI) of the muscle-tendon unit are based on the sliding-filament model of muscle (88, 89) and experimental measurements (90).

*Patterning muscle-tendon units onto backbone.* We pattern the sixteen muscle-tendon units onto the elastic backbone. Units are arranged in four layers. Each layer consists of four units, organized in two agonist-antagonist pairs oriented orthogonal to each other. Neighboring pairs overlap lengthwise by 50% while being rotated 45° relative to each other to avoid intersecting. Muscle-tendon units are arranged on the elastic backbone's surface such that the muscle-tendon unit's center-line lies one (backbone) radius away from the backbone's center-line. The arm in its rest configuration presents no muscle activation or residual stresses.

To join the muscle-tendon units to the backbone, we define zero-displacement boundary conditions at the ends of the muscle-tendon unit relative to point on the outer surface of the backbone to model to first order the enthesis point where tendon and bone join together. The location of these connections a radial distance from the backbone's center-line then generates a bending torque on the backbone then the muscle-tendon unit linear shortens when activated. Beyond these primary connections, we also implement distributed displacement-force boundary conditions to lightly adhere the muscle-tendon unit to the backbone during bending motions (44, 47), analogous to the effect of connective fascia tissue that helps maintain muscle organization.

### 3D tracking environment

The muscular arm defined above is challenged to track a target moving in 3D space along to a smoothly varying, random trajectory. The target trajectory  $\bar{\mathbf{x}}_t(t) = [\bar{x}_1, \bar{x}_2, \bar{x}_3]$  is defined at the beginning of each episode with each spatial component of the trajectory generated according to

$$\bar{x}_j(t) = b_j \sin(f_{j,1}2\pi t) \sin(f_{j,2}2\pi t) \sin(f_{j,3}2\pi t) \quad \forall j \in [1..3] \quad (3)$$

where  $f_{j,i} \in \mathbb{R}$  is a randomly selected constant from the interval  $[0.5, 1.0]$  drawn at the beginning of each episode and  $b_j$  is a velocity scaling constant, whose sign is also randomly selected each episode. This structure of Eq. 3 results in a smoothly varying trajectory of the target while the random selection of  $f_{j,i}$  each episode renders it not possible for the controller to simply memorize a target trajectory. The state and muscle activations of the arm are updated at a 4 Hz control frequency ( $\Delta t = 250$  ms), leading to a mapping between the physical time  $t$  of the simulation environment and the control time step number  $n$  of  $n = [4t]$ . At each timestep, muscle activations are directly mapped from the action space  $\mathbf{a}^n = [a_i] \in \mathbb{R}^{16}, a_i \in [-\infty, \infty]$  to an individual muscle activation level  $\alpha_i(t) = (\tanh(a_i(t)) + 1)/2 \in [0, 1]$  which is applied for 0.25 seconds (until the next control time step). The arm's state consists of the current 3D target location  $\bar{\mathbf{x}}_t(t)$  and the current averaged normalized curvature of the arm  $\mathbf{K}_i(t)$  over  $i = 4$  equidistant intervals in the rod's local

normal and binormal directions

$$\mathbf{K}_i(t) = \frac{L}{2\pi} \frac{\int_{d_i}^{d_{i+1}} \mathbf{k}(s, t) ds}{d_{i+1} - d_i} \quad \forall i \in [0..3] \quad (4)$$

where  $\mathbf{k}(s, t) \in \mathbb{R}^2$  is the local normal and binormal curvature of the arm,  $d = \{kL/4 | k \in [0..4]\}$ , and  $L$  is the length of the arm. This results in a state  $\mathbf{s}^n = [\bar{\mathbf{x}}_t(t), \mathbf{K}(t)] \in \mathbb{R}^{3+8}$ . The reward is the square of the Euclidean distance between the tip of the arm  $\bar{\mathbf{x}}(L, t)$  and the 3D moving target  $\bar{\mathbf{x}}_t(t)$  integrated over the control time step

$$r_n = \int_t^{t+\Delta t} \|\bar{\mathbf{x}}(L, t') - \bar{\mathbf{x}}_t(t')\|^2 dt' \quad (5)$$

The integration of the reward over the time step ensures the tip of the arm continuously follows the target trajectory over the entire time step. Each episode is terminated after 100 seconds.

### Unstructured nest environment

In Fig. 4, the arm is challenged to reach through an unstructured nest of rigid cylindrical obstacles towards a stationary target. Here, the target and nest are stationary and fixed in the same location for all episodes. Between the target and the arm's initial configuration, an unstructured nest of cylindrical obstacles is placed, which the arm must navigate through to reach the target on the other side. The location of the obstacles is the same for all episodes and each episode is terminated after 5 seconds. External contact forces between the arm and the nest are implemented (described in SI) as well as contact between the arm and the floor, allowing the arm to interact with its external environment. The state, action, and reward of this environment are the same as with the 3D tracking environment, but here successful control of the neuromuscular arm requires addressing a more complex physical environment by learning to interact with and account for the presence of solid obstacles that restrict reaching.

### Neural reservoir formulation

We consider two formulations of a neural reservoir: artificial neuron reservoirs and spiking neuron reservoirs. Artificial neuron reservoirs consist of a discrete-time, recurrent network of artificial neurons typically used in machine learning applications while spiking neuron reservoirs utilize a continuous-time, spike-based approach.

### Neural reservoir with artificial neurons

Neural reservoirs using artificial neurons are implemented using an Echo State Network (ESN) approach (32). The reservoir consists of a large pool of  $n$  sparsely and recurrently connected artificial neurons. Recurrent connections within the reservoir are represented using the reservoir connection matrix  $W_r \in \mathbb{R}^{n \times n}$  while connections into the reservoir from the input state (of size  $s$ ) are represented using the input connection matrix  $W_i \in \mathbb{R}^{n \times s}$ . The state of the reservoir  $\mathbf{u}_r^n$  at time step  $n$  is then

$$\mathbf{u}_r^n = f(W_i \mathbf{s}^n + W_r \mathbf{u}_r^{n-1}) \quad (6)$$

where  $f(\cdot)$  is a non-linear activation function. Here, a hyperbolic tangent function (Tanh) is used. Notably, the connection

weight matrices are fixed at initialization. No adjustment to their weights is made during the learning process. Thus, their proper initialization is critical to the information processing capability of the reservoir. Following the ESN approach, both the reservoir connection matrix  $W_r$  and the input connection matrix  $W_i$  are randomly initialized from a normal distribution with zero mean, a standard deviation of 0.50, and a sparse connection density  $\rho$  of 0.10. The reservoir connection matrix is then normalized based on its spectral radius  $\lambda$  (the magnitude of its largest eigenvector) to ensure the echo state property, which is that information from initial conditions will asymptotically disappear over time. To achieve the echo state property it is generally sufficient (though not necessary) to enforce  $\lambda < 1$  (32).

The mapping from the reservoir state  $\mathbf{u}_r^n$  to the output (in this case action, with size  $j$ ) is determined according to the linear output mapping  $W_o \in \mathbb{R}^{j \times n}$  according to

$$a^n = W_o \mathbf{u}_r^n \quad (7)$$

The linear map  $W_o$  can be learned using a variety of different algorithms. For the control policy, the unsupervised reinforcement learning Proximal Policy Optimization (PPO) algorithm was used while for the parallel output maps of Fig. 3, supervised learning algorithms such as ridge regression were used.

### **Neural reservoir with spiking neurons**

To implement a neural reservoir using spiking neurons, we use leaky-integrate and fire (LIF) neurons (91) whose dynamics are governed by the differential equation

$$\tau_m \frac{dv}{dt} = (v_r - v(t)) + RI(t) \quad (8)$$

where  $\tau_m$  is the membrane constant of the RC circuit simulating voltage leakage,  $R$  is the resistance,  $I(t)$  is the input current,  $v(t)$  is the output voltage, and  $v_r$  is the reset membrane potential. When  $v(t)$  exceeds some threshold voltage level ( $v(t) > v_t$ ), a spike in voltage occurs at time  $t$  followed by the neuron returning to  $v_r$  which it is held at for some refractory time  $\tau_r$ .

We begin by converting the discrete time, continuous state input of the soft muscular arm to a continuous time, spike-based input compatible with a spiking neural reservoir. Each state variable is presented as current into two LIF input neurons for a time of  $\Delta$  (which we decouple from the control time  $\Delta t$ ). These two neurons have opposite tuning curves such that one will positively rate encode increasing values while the other will rate encode decreasing values. The resulting spike trains are multiplied by input weights  $W_i$  and fed into the spiking reservoir. The input weights  $W_i$  are randomly drawn from a uniform distribution between -1 and 1. The spiking neural reservoir consists of a set of  $N$  LIF neurons recurrently connected by a sparse set of weights  $W_r$ . The recurrent connection weights have a density of 0.1 and are drawn from the normal distribution  $\mathcal{N}(0, 0.5^2)$ . LIF neuron properties such as refractory period and membrane voltage decay as well as reservoir connection properties were optimized using a CMA evolutionary search algorithm (92) as detailed in the SI. For each control time step, the number of spikes over the time interval  $\Delta$  is counted for each reservoir neuron. This spike

count is then treated as the output of the reservoir for that control time step. This approach then enables the same RL algorithms as used to train the artificial neural reservoir to be employed to learn the output weights  $W_o$  to control the arm.

This neural reservoir setup was deployed on Intel's Kapoho Bay device (33), which contains 2 Loihi chips. The reservoir was defined using the Nengo Python package (93) and its NengoLoihi extension. The input spike conversion layer and output spike counting layer are implemented off-chip, with the communication between the CPU host and the Loihi chip occurring only via spike communication.

### **Energy efficiency**

The energy usage of neural reservoirs consisting of either artificial neurons or spiking neurons was analyzed by measuring the average energy consumption for each inference, or control time step. For a neural reservoir with artificial neurons, this refers to a single evaluation of the system defined in Eq. 6 while for spiking neurons it refers to the presentation of the input state for a time  $\Delta$ . The energy consumption of the artificial neural reservoir was measured using the pyRAPL Python package on an Intel Xeon W-2265 CPU to identify the running power with idle power subtracted. Energy usage on Loihi was measured using utilities provided by Intel's NxSDK API for Loihi to similarly measure running power.

### **Reinforcement learning integration**

To learn the mapping from the neuromuscular arm's state to the action required to control it, reinforcement learning (RL) is used. We employ the state-of-the art, model-free, on-policy Proximal Policy Optimization (PPO) algorithm to optimize  $W_o$  (73) based on the open-source Stable Baselines 3 implementation (94). When used with feed-forward and LSTM network architectures, the environment state  $s^n$  was used as the input state into the artificial neural network in the usual manner. When integrated with the neural reservoir, the current reservoir state  $\mathbf{u}_r^n$  is used as the input state to the PPO algorithm, with the PPO algorithm then tasked with learning the single layer, linear mapping  $W_o$  from the reservoir state to the optimal muscle action. When a neural reservoir consisting of artificial neurons is used, the reservoir state is directly used as the input into the PPO algorithm while for the spiking neural reservoir, output spikes are accumulated and counted over each control time step to determine an analogous reservoir state that can be then directly used by the PPO algorithm.

### **Training and evaluation**

All network architectures were trained using the same approach. For the 3D tracking case (Figs. 1, 2, 4b,c), the policy was trained using 120 parallel environments. The policy was updated every 4800 time steps (1200 seconds at 4 Hz), resulting in each environment collecting ten seconds of data between updates. Each environment was reset every 100 seconds with a new trajectory generated and the arm initiated in a straight configuration. Policies were trained for either two or four million time steps and five separate policies with random initial seeds were trained for each case. Trained policies were evaluated on new,

unseen trajectories for up to 100 seconds (400 time steps). Hyperparameter tuning for all algorithms was performed as detailed in the SI. All training and hyperparameter tuning was performed on the Pittsburgh Supercomputing Center’s Bridges 2 cluster.

For the spiking neural reservoir, only one Kapaho Bay device was available, so NengoLoihi’s emulation environment, which emulates Loihi hardware on CPUs, was used to train using the same 120 parallel environment approach for the 3D tracking case (Fig. 4b,c). For the unstructured obstacle case (Fig. 4e,f), 120 parallel environments were also used (using the NengoLoihi emulator), resulting in each environment having a five-second episode, however, environments were then reset after each episode. All training using the Loihi emulation environment was performed on the Pittsburgh Supercomputing Center’s Bridges 2 cluster. Trained policies were then evaluated on Loihi hardware to determine performance and energy usage. All evaluation and visualization results reported in Fig. 4 and in the SI videos were run on Loihi hardware.

### Kinetic and bending energy measurements

To report the kinetic and bending energy measurements in Figure 2b,c, the average kinetic and bending energy of the arm’s backbone over an episode of length  $T$  was measured as

$$\bar{e}_i = \frac{\int_0^T e_i(t) dt}{\int_0^T dt} \quad (9)$$

where  $e_i \in \{e_k, e_b\}$ . Here  $e_k$  is the kinetic energy of the backbone and  $e_b$  is the elastic bending energy of the rod. The kinetic energy of the rod at each time is

$$e_k(t) = \frac{1}{2} \int_0^L \rho A(\mathbf{v} \cdot \mathbf{v}) ds \quad (10)$$

where  $\rho$  is the local density,  $A$  is the local cross-sectional area, and  $\mathbf{v}$  is the local velocity. The bending energy of the rod at each time is

$$e_b(t) = \frac{1}{2} \int_0^L \mathbf{k} \cdot \boldsymbol{\tau} ds \quad (11)$$

where  $\mathbf{k}$  is the local curvature of the rod,  $\boldsymbol{\tau} = \mathbf{B}\mathbf{k}$  is the local internal elastic torque vector, and  $\mathbf{B}$  is the rod’s bend/twist stiffness matrix. Arm dynamics from trained policies were recorded for 100 unique episodes (100 seconds each) and the average energy of each episode was calculated to determine the distribution of values reported in Fig. 2b.

### Training of parallel maps

Figure 3 shows the use of parallel maps to increase the information extracted from the neural reservoir. To generate these parallel maps, an RL policy trained to track a moving target was recorded for 100 seconds over 400 different target trajectories. The state, action, reservoir state, and environment were recorded. These data were then used as the training data (using an 80/20 training/testing data split) for parallel maps using a supervised learning method via ridge regression. For target correction (Fig. 3d), the parallel map was evaluated every time step to predict the 3D position the target would be one time step in the future and used in place of the target’s true position in the state vector as required.