Integrating Contact-Aware CPG System for Learning-Based Soft Snake Robot Locomotion Controllers

Xuan Liu, *Member, IEEE*, Cagdas D. Onal, *Member, IEEE*, and Jie Fu, *Member, IEEE*

***Abstract*—Contact-awareness poses a significant challenge in the locomotion control of soft snake robots. This article is to develop bioinspired contact-aware locomotion controllers, grounded in a novel theory pertaining to the feedback mechanism of the Matsuoka oscillator. This mechanism enables the Matsuoka central pattern generator (CPG) system to function analogously to a "spinal cord" in the entire contact-aware control framework. Specifically, it concurrently integrates stimuli, such as tonic input signals originating from the "brain" (a goal-tracking locomotion controller) and sensory feedback signals from the "reflex arc" (the contact reactive controller), for generating different types of rhythmic signals to orchestrate the movement of the soft snake robot traversing through densely populated obstacles and even narrow aisles. Within the "reflex arc" design, we have designed two distinct types of contact reactive controllers: 1) a reinforcement learning-based sensor regulator that learns to modulate the sensory feedback inputs of the CPG system, and 2) a local reflexive controller that establishes a direct connection between sensor readings and the CPG's feedback inputs, adhering to a specific topological configuration. These two reactive controllers, when combined with the goal-tracking locomotion controller and the Matsuoka CPG system, facilitate the implementation of two contact-aware locomotion control schemes. Both control schemes have been rigorous tested and evaluated in both simulated and real-world soft snake robots, demonstrating commendable performance in contact-aware locomotion tasks. These experimental outcomes further validate the benefits of the modified Matsuoka CPG system, augmented by a novel sensory feedback mechanism, for the design of bioinspired robot controllers.**

***Index Terms*—Adaptive systems, bioinspired control, haptic interfaces, neurodynamics, reinforcement learning (RL), soft robotics.**

Received 1 July 2024; revised 30 November 2024; accepted 18 January 2025. Date of publication 5 February 2025; date of current version 25 February 2025. This work was supported by the National Science Foundation under Grant 1728412. This article was recommended for publication by Associate Editor D. Navarro-Alarcon and Editor A. Menciassi upon evaluation of the reviewers' comments. *(Corresponding author: Jie Fu.)*

Xuan Liu is with the School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China (e-mail: xliu9@zju.edu.cn).

Cagdas D. Onal is with the Robotics Engineering Department, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: cdonal@wpi.edu).

Jie Fu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: fujie@ufl.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TRO.2025.3539173, provided by the authors.

Digital Object Identifier 10.1109/TRO.2025.3539173

I. Introduction

SOFT continuum robots have unique advantages in traversing through cluttered and confined environments, due to their flexible body structure and deformable materials. Applications of soft continuum robots in contact-aware environments include search-and-rescue [1], pipe inspection [2], and medical surgery [3]. In particular, soft robotic snakes have the unique potential that any part of their body, if properly controlled, could adapt to and reduce the impact from collisions, or even benefit from the propulsion force generated by the contacts with obstacles. In this article, we investigate the following two questions.

- How to design the contact-aware controller for the soft snake robot that can intelligently adapt to and employ the contact force from crowded obstacles during locomotion?
- How to design the sensory and body structure on a snake-like soft robot to make the tactile perception more sensitive while avoiding jamming, and making the contact-aware locomotion more energy efficient?

In literature, several research groups have studied this unique topic in snake robot locomotion. The solutions to the contact-aware locomotion [4], [5], [6] are mainly studied and implemented on rigid snake robots and most of the control methods are model-based. Transeth et al. [7], [8] first defined this property as the *obstacle-aided locomotion*, wherein the snake robot actively employs external objects to generate propulsion forces during the locomotion. Their pioneer work proposed a two-module framework of obstacle-aided locomotion that includes: 1) a path planner that searches for a trajectory with more active contact chance for the rigid snake robot, and 2) a motion controller that controls the snake robot's real-time body movements to optimally utilize the contacts between the robot and the environment and generate desired propulsion force for the locomotion. In [9], [10], and [11], a hybrid controller is developed, where a contact event is treated individually by a reactive controller that maximizes the total propulsion force at the contacting moment. This controller has been applied to a rigid snake robot and showed its reliability in maintaining beneficial propulsion force. Kano et al. [12], [13] proposed local reflexive mechanisms that interrogate the contact status between the snake robot and the obstacles to determine whether the contact is beneficial to the locomotion. In this approach, only a segment of the robot links neighboring to the link in

contact react to the sensory feedback. On the basis of the local reflexive control method, a Tegotae heuristic scoring function is established by authors in [5], [14], and [15], for selecting which kind of reaction should be applied to the contacting link of the robot given certain situations including the snake robot's shape and contacting part of the robot. From the bioinspired perspective, inspired by the entrainment properties of neural oscillators that allow the systems' output to be synchronized with sensory feedback, several studies [6], [16] introduce CPG systems to the control loop of snake robots to process the sensory feedback signals during locomotion. However, in most existing work the locomotion control inputs of the feedback CPG systems are usually constant or basic sinusoidal signals due to the difficulty of coordinating multiple complex signals through a CPG system. Performing learning-based goal-tracking and contact reactive control on a CPG-driven snake robot is a promising, yet underexplored topic.

So far, the results on contact-aware control for soft snake robots are scarce. Although a few end-to-end soft robot controllers [17] perform well in simulation by assuming fully proprioceptive observations, it would be appealing to enable such a capability for soft snake robots in the real world. Moving from rigid snake robots to soft snake robots in contact-aware locomotion control faces many challenges as follows.

1) Due to the continuum of the pneumatic actuators, it is infeasible to construct an accurate dynamic model for a soft snake robot, rendering model-based control ineffective or inapplicable.
2) The pneumatic actuators in soft snake robots have nonlinear, delayed, and stochastic dynamical response given inputs, making it difficult to achieve fast responses through model-based control compared to rigid snake robots.
3) It is hard to embed tactile sensors in the soft material since the contact-free deformation of the soft body may interfere with the sensory data. As a result, the tactile sensors cannot be densely placed on the soft robot.
4) Equipping tactile sensors could introduce more contact friction due to the material of the sensors, or cause more contact jamming due to the bumped shape of the sensors.
5) When a soft snake robot is traversing unknown obstacles, its tactile sensory inputs usually present discrete and unpredictable impulsive features, which can result in overshoot and signal interference to a feedback control system.

For the above-mentioned problems, our solutions and contributions are summarized as follows.

1) *Development of a Novel Feedback Mechanism for the Matsuoka CPG:* Our modification on the Matsuoka oscillators allows the Matsuoka CPG system to effectively process both the locomotion control signals and tactile sensory feedback signals during the contact-aware locomotion of the soft snake robot. Through theoretical analysis, we leverage the unique advantages of the Matsuoka oscillator's feedback mechanism for reducing the overshoot and latency despite the interference of unexpected sensory feedback signals with the feedback control signals throughout the contact events.
2) *Design of Contact Reactive Controllers:* Based on our modification of the Matsuoka oscillator, we designed two different contact reactive control schemes [see Fig. 1(a) and (b)] for the contact-aware locomotion control of the soft snake robot. These two control schemes have disparate sensory feedback patterns: the *hybrid learning* controller uses a neural network-based policy to process the contact information; while the *local reflexive* method builds local signal flows between the contact sensor data and the CPGs' feedback inputs. We use these two reactive controllers to verify the adaptability of our modified Matsuoka CPG system to various sensory feedback input patterns.
3) *Design of a Gait Switching Mechanism in the CPG System:* Inspired by the way of natural snakes traversing the combination of plain ground and narrow aisles [18], [19], we include two classical snake locomotion gaits—the slithering gait and the rectilinear gait in the soft snake robot's contact-aware locomotion. These gaits are generated by the identical Matsuoka CPG system. We use different connections between the tactile sensors and the CPGs' feedback inputs to realize these two gaits and their switching mechanism. The triggering conditions for the gait switching are independent to the type of reactive mechanisms (*hybrid learning* or *local reflexive* modules for contact processing), and thus allows us to control the influence of gaits when comparing the performance of different contact reactive controllers.

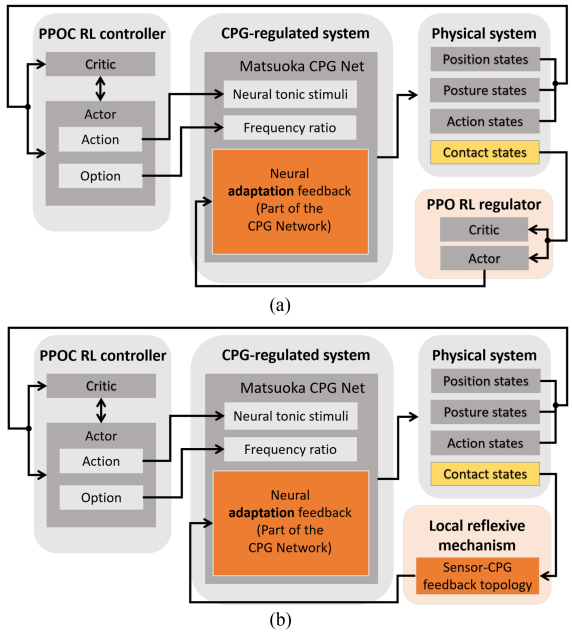


Fig. 1. Schematic view of (a) hybrid learning proximal policy optimization option-critics with central pattern generator (PPOC-CPG), and (b) local reflexive PPOC-CPG controllers.

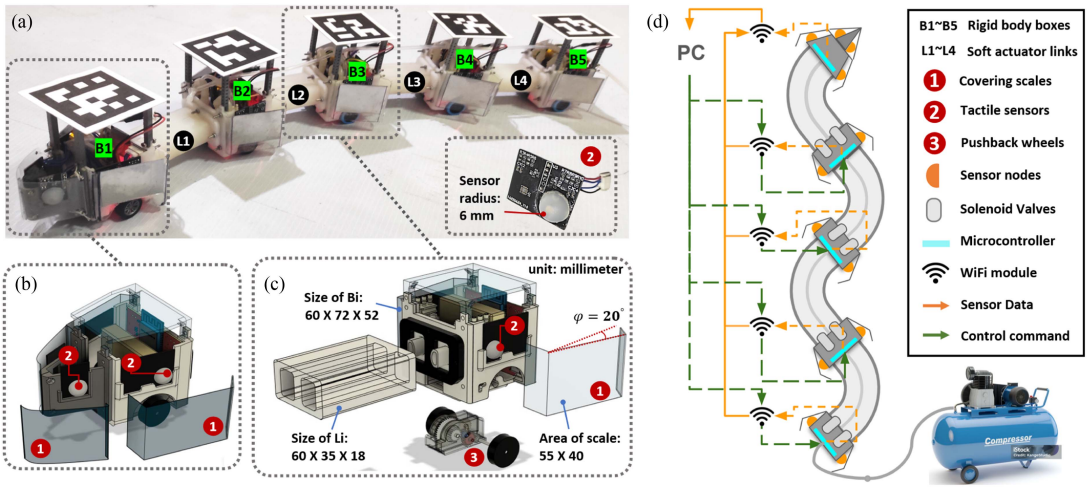


Fig. 2. (a) Soft snake robot in reality. (b) In rigid head and (c) rigid body components (Number 1, 2, and 3 in the above-mentioned figures are used for marking the covering scale, silicone sensor node, and torsion spring passive wheels, respectively). (d) Signal communication flow of soft snake robot circuit. (e) Example of sensor-CPG connection model for one link of a soft snake robot.

4) *Design and Allocation of the Tactile Sensors:* Soft snake robots' whole-body contact-aware locomotion faces challenges due to conventional tactile sensors' fragility, communication limits, high costs, and integration issues. To overcome these hurdles, we developed magnetic-based sensors inspired by Wang et al. [20], mimicking *Scale Sensilla* [21] for larger sensing area, fewer sensors, and improved sensitivity to dynamic contacts in soft snake robots.
5) *Experimental Implementations and Validations:* The efficacy of the proposed reactive control schemes and the gait switching module are evaluated in the real soft snake robot's contact-aware locomotion tasks. Both control schemes achieve promising locomotion performance for traversing crowded obstacles, while maintaining accurate target tracking and the ability to handle sharp turns. The gait switching mechanism is also successfully triggered when the soft snake robot detects narrow aisles. These results comprehensively verify the advantages of our modification on the Matsuoka CPGs' feedback mechanism.

This article focuses on the theoretical advancements and practical application of the Matsuoka oscillator's feedback mechanism for contact-aware locomotion control in soft snake robots. First, we provide an overview of our schematic design of a soft snake robot equipped with contact sensors in Section II. In Section III, we present the novel feedback mechanism of the Matsuoka CPGs with theoretical analyses to show its advantages in leveraging sensory feedback signals for contact-aware locomotion control. Based on this, we design two contact reactive controllers and a slithering-rectilinear gait switching module in Section IV. In Section V, we use several experiments to show the improvement of the snake robot's locomotion performance in the cluttered environments under closed-loop control. Finally, Section VI concludes this article.

II. Hardware Design for Contact-Aware Soft Snake Robot Locomotion

Inspired by the mechanical design of the soft pneumatic actuators in [22], our soft snake robot consists four pneumatically actuated soft links [see L1∼L4 in Fig. 2(a)] [23], [24], [25]. The links are connected by five rigid bodies [see B1∼B5 in Fig. 2(a)]. Each of them encloses a set of electronic components that are necessary to control the snake robot, including a ESP32 microcontroller unit with onboard WiFi module for link-to-link and link-to-computer communication [see Fig. 2(d)], a pair of SMC-S070C-SCG 3-2 (three-way two-position) solenoid valves [25], an 800 mAh LiPo battery, a pair of one-direction pull back wheels [see Fig. 2(c)] mimicking the anisotropic friction property,[1] and a couple of magnetic-based tactile sensors [see Fig. 2(b) and (c)], which will be introduced in detail later in this section.

A. Pulsewidth Modulation (PWM) for the Solenoid Valves

In the earlier episode of our soft snake robot [25], each soft link [such as L1 depicted in Fig. 2(a)] comprises two chambers, both of which are independently controlled by solenoid valves; yet, only one chamber per link is pressurized through a single valve at any given moment. In this work, we allow both chambers' valves to be activated simultaneously. Each of the two chambers controlling the joint were connected to a 20 psi [138 kPa] pressurized air line and each controlled by a separate solenoid valve (operated with 1 kHz PWM signal). The solenoid valve being used is tested to have fully open status when the PWM signal is at or above 59% duty cycle. This change allows the the soft snake robot to perform both slithering gait and rectilinear gait (for traversing narrow aisles) during the locomotion. It is

[1] Check supplementary document Section S1 for an explanation of the pull-back torque mechanism, which enhances the energy efficiency during the snake robot's contact-aware locomotion.

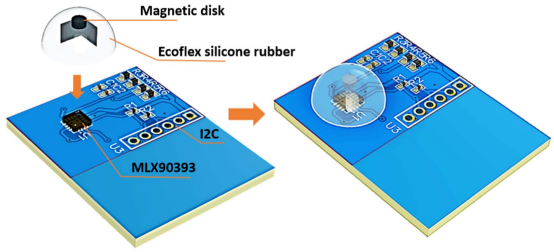


Fig. 3. Electronic design of touch sensor.

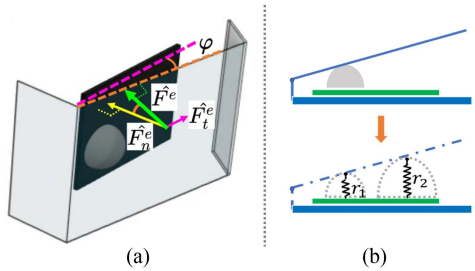


Fig. 4. (a) Illustration of contact force decomposition for a scale covered tactile sensor on the left of a rigid body. (b) Tactile sensor + covering scale structure (right top) versus its approximation in simulation (right bottom).

noted that the 3–2 valves do not offer the proportional pressure control option, the bending and elongation movements are not coupled on a soft link of the snake robot. Therefore, any bending movements of the soft links controlled by two valves in this article are equivalent to the "single valve plan" in our previous work [23], [25].

B. Design of the Tactile Sensor

1) Core Sensing Unit: For any contact-aware robot locomotion task, the tactile sensors should always be time efficient to ensure safe and prompt reaction of the robot controller. Due to its whole body contact feature when locomoting through the obstacles, the soft snake robot brings different challenges to the design of tactile sensors on the energy and communication cost, durability, accuracy, size, deformability, and cover area.

Considering the tradeoff among these requirements, we choose a magnetic field soft tactile sensor based on [20]. As shown in Fig. 3, the major component of the soft tactile sensor is comprised of a small magnet cylinder disk (with 2 mm diameter and a height of 1 mm) and a Melexis MLX90393 Hall effect module (3 mm $\times$ 3 mm $\times$ 0.8 mm, QFN-16 package) separated by a hemisphere-shaped elastomer (made of Ecoflex$^{\text{TM}}$ 00-30 silicone rubber). The magnet piece is sealed in the elastomer through molding of the silicone first and then the elastomer is glued to the top of the hall sensor on the printed circuit board. The detailed fabrication steps are similar to [20]. The working principle of this tactile sensor is based on the detection of the presence and magnitude of a magnetic field using the Hall effect. The magnetic field varies when the elastomer deforms and causes positional changes in the small magnet disk inside the elastomer. These changes can be detected and calculated by the hall sensor. The data collected by the hall sensor is sent to the motherboard via interintegrated circuit (I2C) bus.

2) Covering Scales: As shown in Fig. 2(a), (b), and (c), there are in total 12 tactile sensors installed on the robot. The components marked by number 2 are the installation positions of the tactile sensors. On top of each tactile sensor, the components marked by number 1 are the scales. Each scale is made of two layers of materials—an acrylic layer attached with a steel plate layer (glued by polyurethane and reinforced with adhesive tape). The scales are designed for three major purposes as follows.

1) To significantly increase the contact sensitivity and sensing area of the snake robot [in Fig. 2(c), the contact area is about 20 times to the uncovered version].
2) To reduce friction resistance on the contact surface (friction coefficient reduced from 1.7 of dry silicone to around 0.3 of polished acrylic board).
3) To protect the silicone tactile node from frequent collisions.

C. Preprocessing and Data Structure of the Sensory Information

1) Real Robot Sensors: Based on [20, (12)∼(14)], the tactile force elements $\mathbf{F} = \{F_x, F_y, F_z\}$ are calculated by

$$F_z = \Sigma_{k=0}^{n}\Sigma_{i=0}^{k} C_{zj} B_z^i B_r^{(k-i)}, \quad j = 1, \ldots \frac{(n+1)n}{2}$$

$$F_r = \Sigma_{k=0}^{n}\Sigma_{i=0}^{k} C_{rj} B_z^i B_r^{(k-i)}, \quad j = 1, \ldots \frac{(n+1)n}{2}$$

where F_z represents the normal contact force and F_r represents the shear contact force, which can be further decomposed to

$$F_x = \frac{B_x}{\sqrt{B_x^2 + B_y^2}} F_r, \quad F_y = \frac{B_y}{\sqrt{B_x^2 + B_y^2}} F_r.$$

Parameters B_z and B_r are the normal and shear magnetic intensity, where B_r can be decomposed to B_x and B_y. C_{zj} and C_{rj} are the jth coefficients of best fitting polynomials of F_z and F_r calculated by moving least squares (MLS) method which generates a data-based model that maps the measured magnetic field to the estimated force on the silicone elastomer. The number n is the polynomial order in the MLS method.

Due to the constraints of the locomotion's dimensions and the relatively low friction between the covering scale and the contact object (with a friction coefficient of 0.3), the direction of the contact force can be approximately regarded as perpendicular to the surface of the covering scale. As Fig. 4(a) shows, the contact force $\hat{F}^e$ acting on the left side scale covered sensor can be decomposed to two components—the vertical component $\hat{F}^e_n$ is transmitted to the normal pressure on the silicone node, which mainly contributes to the sensor reading F_z; and the tangent component $\hat{F}^e_t$ is transmitted to the shear squeezing force, which can be treated as the source of F_r. As a result, the contact force

on the scale sensor can be approximated by the equivalent effect of deformation on the silicone elastomer, and translated to the three axis measurement of the magnetic force $\mathbf{F}$. We reduce the sensor data dimension by calculating the norm of $\mathbf{F}$, such that

$$F = ||\mathbf{F}|| = \sqrt{F_x^2 + F_y^2 + F_z^2}. \tag{1}$$

Furthermore, we introduce a sigmoid function to process the value of the soft tactile sensor, such that

$$\sigma(F) = \frac{2}{1 + \exp^{-aF}} - 1 \tag{2}$$

where $a \in \mathbf{R}^+$ is a positive constant. This step helps to limit the force value within the range $[0, 1]$ and amplify the weak signals to increase the sensitivity to the contact events, such that the small contact signals' wave shapes become closer to square-like and easier to be identified.

Inspired by a previous study on obstacle-aided locomotion of rigid snake robots [10], we approximate the total contact forces acted on the rigid bodies $B1 \sim B5$ based on the inputs sampled from the force sensors.

According to (2), let F_i^e, F_i^f represent the norm of contact force vectors detected by the left and right sensor respectively on the ith rigid body of the soft snake robot, then $N_i^e = \sigma(F_i^e)$, and $N_i^f = \sigma(F_i^f)$ (the superscripts e, f denote left and right of the sensors' placement in reference to the heading direction of the soft snake robot). We use the collection of contact forces in the soft snake robot from head to tail to represent the contact states of the soft snake robot as follows:

$$\mathbf{N} = \left[N_1^e, N_1^f, N_2^e, N_2^f, N_3^e, N_3^f, N_4^e, N_4^f, N_5^e, N_5^f\right]^T$$

where $N_1^e = N_{11}^e + N_{12}^e$, and $N_1^f = N_{11}^f + N_{12}^f$ specially, since the robot has two pairs of tactile sensors on the head rigid link (B1 in Fig. 2).

2) Simulated Sensors: In order to simulate the robot for reinforcement learning (RL) and sim-to-real transfer of the learned controller, we developed a physics-based high-fidelity simulator that models the inflation and deflation of the air chamber and the resulting deformation of the soft bodies with tetrahedral finite elements [23]. To simplify the tactile sensing function of the scale structure in simulation, we use two hemisphere elastic force fields with different sensing radii to model the tactile sensor node + scale structure in reality [as shown in Fig. 4(a)]. The elastic force fields have equilibrium positions (where elastic force equals zero) everywhere on the surface of the hemispheres and have no friction on the hemispheres. The tactile readings are modeled by the elastic forces when an object's distance is smaller than the radius of any simulated tactile node. In the simulation, the readings of the two hemisphere force fields are averaged to simulate the contact force signal of one tactile sensor in the real robot.

III. Modified Matsuoka Oscillator With Sensory Feedback

In order to effectively integrate the tactile information into the contact-aware locomotion framework of the soft snake robot, we study the effect of an additional variable on the Matsuoka oscillator for handling the feedback force signals from the tactile sensors. In this section, we analyze the properties of our method and the conventional approach [6], [16] from a theoretical perspective.

In our previous work [24], [25], we presented a control scheme that employs sensor-free Matsuoka oscillators to generate undulating control signals as actuation inputs for the soft snake robot to perform Serpentine locomotion. The original Matsuoka oscillator is a piecewise linear dynamical system mimicking a neuron's behavior, which has the form

$$\begin{aligned}
k_f \tau_r \dot{x}_i^e &= -x_i^e - a z_i^f - b y_i^e - \Sigma_{\text{coupling}}^e + u_i^e + c \\
k_f \tau_a \dot{y}_i^e &= z_i^e - y_i^e \\
k_f \tau_r \dot{x}_i^f &= -x_i^f - a z_i^e - b y_i^f - \Sigma_{\text{coupling}}^f + u_i^f + c \\
k_f \tau_a \dot{y}_i^f &= z_i^f - y_i^f
\end{aligned} \tag{3}$$

where the subscripts e and f represent variables related to the extensor neuron and flexor neuron, respectively. x_i^e, x_i^f represent the activation states (or membrane potentials) and y_i^e, y_i^f denote the self-inhibitory states (or adaptation states [26], [27]) of the ith neuron respectively, z_i^e, z_i^f are the outputs of the ith neuron. Tonic inputs u_i^e, u_i^f are the major coefficients that can be controlled to affect the output bias and amplitude of the Matsuoka oscillator. The frequency ratio $k_f \in \mathbf{R}$ can be manipulated to affect the natural oscillation frequency of the system. The interneuron couplings have common expressions by assuming fully connected topology, which are $\Sigma_{\text{coupling}}^e = \sum_{j=1}^N (w_{ji}^{ee} z_j^e + w_{ji}^{fe} z_j^f)$ and $\Sigma_{\text{coupling}}^f = \sum_{j=1}^N (w_{ji}^{ef} z_j^e + w_{ji}^{ff} z_j^f)$, where the subscript and superscript of a coupling weight is defined by the connection between the CPG nodes, e.g., w_{ji}^{ef} indicates the coupling weight from jth CPG node's extensor to the ith CPG node's flexor. The free-response input is denoted as parameter c in the equation, which is used for amplifying the free-response oscillation of the CPG system. The remaining parameters are all constant weights. In system (3), all coupled states including $x_i^e, x_i^f, y_i^e, y_i^f$ and z_i^e, z_i^f are inhibited (negatively weighted), except that the tonic inputs are activating signals (positively weighted). In the context of this article, the extensor and flexor subsystems in the CPG system play distinct roles in actuating left and right chambers of the snake robot respectively (taking the head direction of the robot as reference).

Based on the form of the original Matsuoka oscillator, the question is—how to integrate sensory feedback into the Matsuoka CPG system to enable effective response to contact events?

A conventional approach is to directly add positive force feedback (as activation signals) to the membrane potential state equations ($\dot{x}_i^e, \dot{x}_i^f$) of the original Matsuoka oscillator [16, (5)]. Such form of feedback Matsuoka oscillator has been used in some snake robot locomotion studies [6], [16], where the tonic inputs (for locomotion control) of the CPG systems in these applications are mostly constant or regular sinusoidal waves. We summarize the dynamic equations of the conventional feedback Matsuoka oscillator as follows.

Membrane potential feedback form (MPF) Matsuoka oscillator:

$$\begin{aligned} k_f\tau_r\dot{x}_i^e &= -x_i^e - az_i^f - by_i^e - \Sigma_{\text{coupling}}^e + u_i^e + bp_i^e + c \\ k_f\tau_a\dot{y}_i^e &= z_i^e - y_i^e \\ k_f\tau_r\dot{x}_i^f &= -x_i^f - az_i^e - by_i^f - \Sigma_{\text{coupling}}^f + u_i^f + bp_i^f + c \\ k_f\tau_a\dot{y}_i^f &= z_i^f - y_i^f \end{aligned} \tag{4}$$

where the sensory force feedback signals are represented by p_i^e and p_i^f. The reason for naming MPF to this type of Matsuoka oscillator is because the sensory feedback signals are directly added to the membrane potential states as activation (positive) signals.

However, we have a concern about the above-mentioned conventional form. In the case when the tonic inputs are complicated wave signals and the sensory feedback are irregular signals, these two types of inputs may intervene each other, and therefore fail to effectively present the impact of sensory feedback to the system output.

Sato et al. [28] mentioned the addition of the CPG state coupling terms not only to the fast dynamic states' equations (potential membrane $\dot{x}_i^e, \dot{x}_i^f$) but also to the slow dynamic states' equations (adaptation states $\dot{y}_i^e, \dot{y}_i^f$) of the Matsuoka oscillator with opposite signs to improve the dynamic impact of the coupling signals. Given the inspiration, we are interested in knowing whether it is possible to add sensory feedback signals which are external impulse signals to the adaptation states of the Matsuoka oscillator. After in-depth theoretical analysis and experimental comparison, we construct a novel branch of feedback mechanism in the Matsuoka oscillator as follows.

Adaptation feedback (AF) form Matsuoka oscillator:

$$\begin{aligned} k_f\tau_r\dot{x}_i^e &= -x_i^e - az_i^f - by_i^e - \Sigma_{\text{coupling}}^e + u_i^e + c \\ k_f\tau_a\dot{y}_i^e &= z_i^e - y_i^e - p_i^e \\ k_f\tau_r\dot{x}_i^f &= -x_i^f - az_i^e - by_i^f - \Sigma_{\text{coupling}}^f + u_i^f + c \\ k_f\tau_a\dot{y}_i^f &= z_i^f - y_i^f - p_i^f. \end{aligned} \tag{5}$$

In this design, the tonic inputs u_i^e, u_i^f as well as the free oscillation tonic input c are still added to the potential membrane states (x_i^e, x_i^f) as fast dynamic inputs, while the sensory feedback p_i^e and p_i^f are inhibitive signals[2] added to the equations of adaptation states (y_i^e, y_i^f) of Matsuoka oscillator as slow dynamic feedback inputs. In this article, we name this version of the Matsuoka oscillator as the AF form Matsuoka oscillator.

To explore the feasibility of the AF form Matsuoka oscillator, and find out the advantage of the AF form design, we discuss the difference between the AF and MPF form of Matsuoka oscillator when both the tonic inputs and the sensory feedback signals are variables. The discussion is organized by the following derivations.

[2] For the reason of making inhibitive sensory feedbacks, please refer to the supplementary document S-Section 3.

Considering AF form Matsuoka oscillator described in system (5) and MPF form Matsuoka oscillator method described in system (4). For the AF-form Matsuoka CPG system, consider the coupling term based on the CPG network connection topology of the soft snake robot (as shown in Fig. 7), we have the mathematics formulation as follows:

$$\begin{aligned} k_f\tau_r\dot{x}_i^e &= -x_i^e - az_i^f - by_i^e - w_{i-1,i}z_{i-1}^f - w_{i+1,i}z_{i+1}^e \\ &\quad + u_i^e + c \\ k_f\tau_a\dot{y}_i^e &= z_i^e - y_i^e - p_i^e \\ k_f\tau_r\dot{x}_i^f &= -x_i^f - az_i^e - by_i^f - w_{i-1,i}z_{i-1}^e - w_{i+1,i}z_{i+1}^f \\ &\quad + u_i^f + c \\ k_f\tau_a\dot{y}_i^f &= z_i^f - y_i^f - p_i^f \end{aligned} \tag{6}$$

with $i = 1, 2, 3, 4$, $w_{0,1} = w_{5,4} = 0$.

Let $\mathbf{X^e} = [x_1^e, x_2^e, x_3^e, x_4^e]^T$, $\mathbf{X^f} = [x_1^f, x_2^f, x_3^f, x_4^f]^T$, and apply this notation analogously to $\mathbf{Y^e}$, $\mathbf{Y^f}$, $\mathbf{Z^e}$, $\mathbf{Z^f}$, $\mathbf{U^e}$, $\mathbf{U^f}$, $\mathbf{P^e}$, $\mathbf{P^f}$. Let $\mathbf{C} = [c, c, c, c]^T \in \mathbf{R}^4$. Then, the vector form of (6) can be expressed as follows:

$$\begin{aligned} k_f\tau_r\dot{\mathbf{X}}^\mathbf{e} &= -\mathbf{X^e} - a\mathbf{Z^f} - b\mathbf{Y^e} - \mathbf{W_u}\mathbf{Z^e} - \mathbf{W_d}\mathbf{Z^f} + \mathbf{U^e} + \mathbf{C} \\ k_f\tau_a\dot{\mathbf{Y}}^\mathbf{e} &= \mathbf{Z^e} - \mathbf{Y^e} - \mathbf{P^e} \\ k_f\tau_r\dot{\mathbf{X}}^\mathbf{f} &= -\mathbf{X^f} - a\mathbf{Z^e} - b\mathbf{Y^f} - \mathbf{W_u}\mathbf{Z^f} - \mathbf{W_d}\mathbf{Z^e} + \mathbf{U^f} + \mathbf{C} \\ k_f\tau_a\dot{\mathbf{Y}}^\mathbf{f} &= \mathbf{Z^f} - \mathbf{Y^f} - \mathbf{P^f} \end{aligned} \tag{7}$$

where

$$\mathbf{W_u} = \begin{bmatrix} 0 & w_{21} & 0 & 0 \\ 0 & 0 & w_{32} & 0 \\ 0 & 0 & 0 & w_{43} \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{W_d} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ w_{12} & 0 & 0 & 0 \\ 0 & w_{23} & 0 & 0 \\ 0 & 0 & w_{34} & 0 \end{bmatrix}.$$

Let $\mathbf{W} = \mathbf{W_u} - \mathbf{W_d}$, such that

$$\mathbf{W} = \begin{bmatrix} 0 & w_{21} & 0 & 0 \\ -w_{12} & 0 & w_{32} & 0 \\ 0 & -w_{23} & 0 & w_{43} \\ 0 & 0 & -w_{34} & 0 \end{bmatrix}.$$

Let $\mathbf{X} = \mathbf{X^e} - \mathbf{X^f}$, and apply this notation analogously to $\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{P}$. Subtracting $\dot{\mathbf{X}}^\mathbf{e}$ and $\dot{\mathbf{X}}^\mathbf{f}$, $\dot{\mathbf{Y}}^\mathbf{e}$ and $\dot{\mathbf{Y}}^\mathbf{f}$ in (7), we can obtain

$$\begin{aligned} k_f\tau_r\dot{\mathbf{X}} &= -\mathbf{X} + a\mathbf{Z} - \mathbf{W}\mathbf{Z} - b\mathbf{Y} + \mathbf{U} \\ k_f\tau_a\dot{\mathbf{Y}} &= \mathbf{Z} - \mathbf{Y} - \mathbf{P}. \end{aligned} \tag{8}$$

When x_i^e and x_i^f are perfectly entrained for $i = 1, 2, 3, 4$, $r_{x_i^e} \approx r_{x_i^f}$ (r is the ratio of bias to the amplitude of the signal x), we

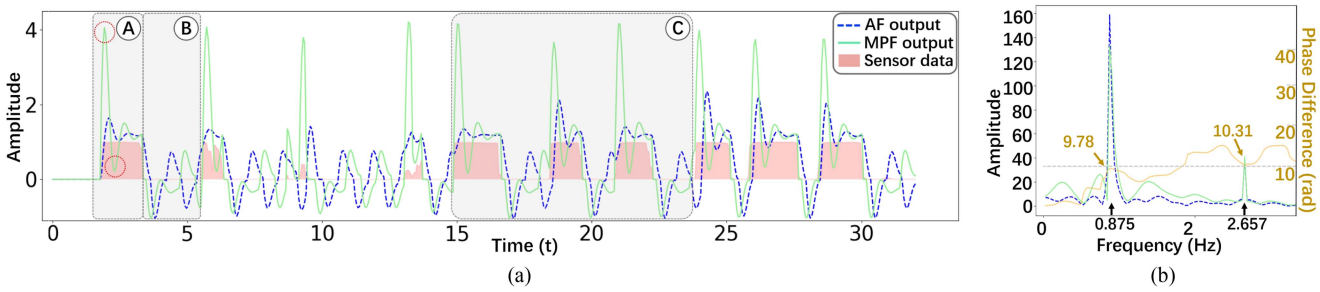


Fig. 5. (a) Output of AF form and MPF form Matsuoka oscillator given sensory feedback data. Highlighted region A presents overshoots problem, region B presents phase distortion issue, and region C shows the influence of continuous feedback waves. (b) Spectrum and phase difference between the outputs of AF form and MPF form Matsuoka oscillator in responding to a square like sensory input.

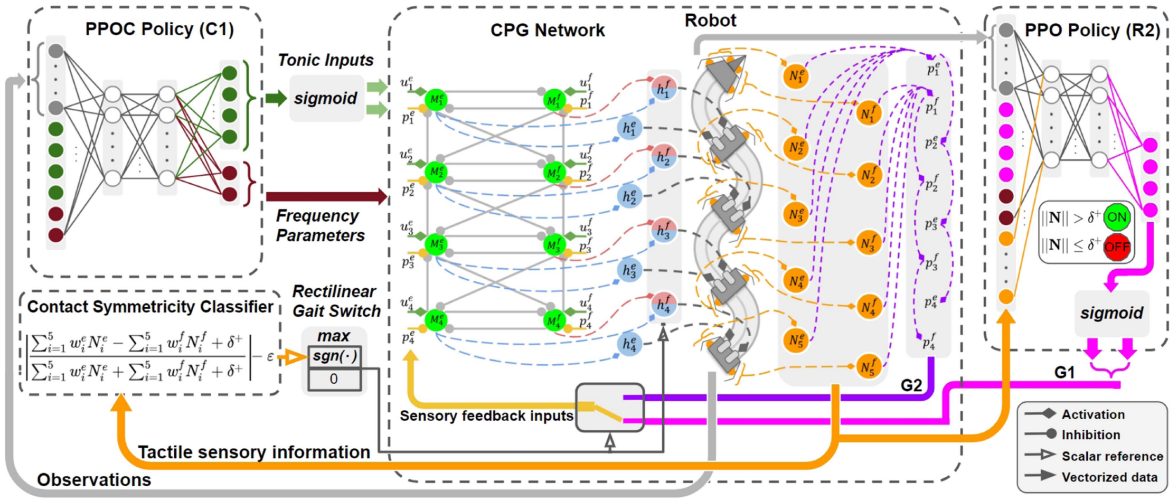


Fig. 6. AF-learning control scheme with rectilinear gait module and the switching method. G1 represents the sensory feedback data flow for the slithering gait, while G2 indicates the data flow for the rectilinear gait. The CPG nodes M_i^e and M_i^f colored in light green represent the extensor half and flexor half of the primitive Matsuoka oscillator corresponding to the ith soft link's pattern. The ON/OFF switch signifies that the R2 module is activated only when $||\mathbf{N}|| > \delta^+$ (where $\delta^+ \approx 0$), indicating that contact has been detected. This scheme also applies to the MPF-learning method except that the primitive CPG nodes are replaced by the MPF form Matsuoka oscillator.

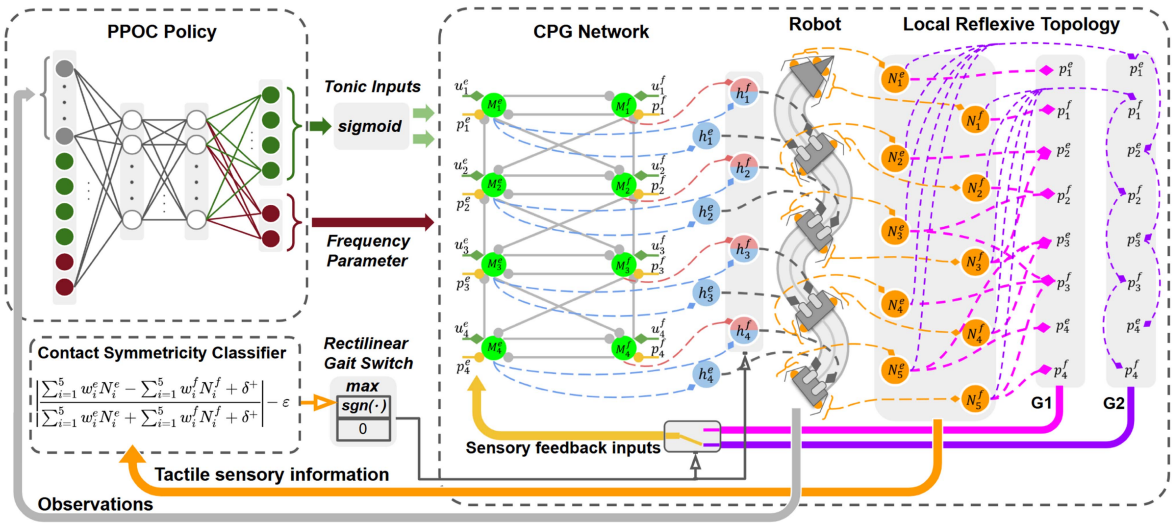


Fig. 7. Control scheme of the AF-local method with rectilinear gait module and the switching method. G1 represents the sensory feedback inputs flow for the slithering gait, while G2 indicates the feedback inputs flow for the rectilinear gait. This scheme also applies to the MPF-local method except that the CPG nodes are replaced by the MPF form Matsuoka oscillator.

have $z_i \approx K_i x_i$ [according to (B.2)[3]]. Denote

$$\mathbf{\Lambda_K} = \begin{bmatrix} K_1 & 0 & 0 & 0 \\ 0 & K_2 & 0 & 0 \\ 0 & 0 & K_3 & 0 \\ 0 & 0 & 0amp; K_4 & \end{bmatrix}$$

where $K_i = K(r_{x_i})$ (see Appendix B), and $\mathbf{E}$ as a 4-D identity matrix, then (8) can be simplified to

$$k_f \tau_r \dot{\mathbf{X}} = (-\mathbf{E} + a\mathbf{\Lambda_K} - \mathbf{W}\mathbf{\Lambda_K})\mathbf{X} - b\mathbf{Y} + \mathbf{U} \quad (9)$$

$$k_f \tau_a \dot{\mathbf{Y}} = \mathbf{\Lambda_K}\mathbf{X} - \mathbf{Y} - \mathbf{P}. \quad (10)$$

Differentiate (9) on time t to obtain

$$k_f \tau_r \ddot{\mathbf{X}} = (-\mathbf{E} + a\mathbf{\Lambda_K} - \mathbf{W}\mathbf{\Lambda_K})\dot{\mathbf{X}} - b\dot{\mathbf{Y}} + \dot{\mathbf{U}}. \quad (11)$$

Calculate the equations through $(9) - b\,(10) + k_f \tau_a\,(11)$, we can obtain the second order ODE form of the CPG system as

$$k_f^2 \tau_a \tau_r \ddot{\mathbf{X}} + k_f((\tau_a + \tau_r)\mathbf{E} - a\tau_a \mathbf{\Lambda_K} + \tau_a \mathbf{W}\mathbf{\Lambda_K})\dot{\mathbf{X}} + (\mathbf{E} + (b-a)\mathbf{\Lambda_K} + \mathbf{W}\mathbf{\Lambda_K})\mathbf{X} = k_f \tau_a \dot{\mathbf{U}} + \mathbf{U} + b\mathbf{P}. \quad (12)$$

For MPF-form Matsuoka oscillator, according to (4) and the CPG network structure in Fig. 7, we can obtain the vector form as follows:

$$k_f \tau_r \dot{\mathbf{X}} = -\mathbf{X} + a\mathbf{Z} - \mathbf{W}\mathbf{Z} - b\mathbf{Y} + \mathbf{U} + b\mathbf{P}$$

$$k_f \tau_a \dot{\mathbf{Y}} = \mathbf{Z} - \mathbf{Y}. \quad (13)$$

Similar to the derivations in the AF-form Matsuoka CPG system, the second order ODE of the MPF-form Matsuoka CPG can be derived as

$$k_f^2 \tau_a \tau_r \ddot{\mathbf{X}} + k_f((\tau_a + \tau_r)\mathbf{E} - a\tau_a \mathbf{\Lambda_K} + \tau_a \mathbf{W}\mathbf{\Lambda_K})\dot{\mathbf{X}} + (\mathbf{E} + (b-a)\mathbf{\Lambda_K} + \mathbf{W}\mathbf{\Lambda_K})\mathbf{X} = k_f \tau_a \dot{\mathbf{U}} + \mathbf{U} + k_f \tau_a b\dot{\mathbf{P}} + b\mathbf{P}. \quad (14)$$

From the right-hand side of (12) and (14), the derivation (12) of MPF form Matsuoka oscillator has an additional free term $k_f \tau_a b\dot{\mathbf{P}}$ compared to the derivation (14) of the AF form Matsuoka oscillator. According to the superposition property of solutions of the second order ODE, when $\mathbf{P}$ is a variable with complex waveform (e.g., collision force signals), the interference of $k_f \tau_a b\dot{\mathbf{P}}$ will be relatively large. Concluding the above-mentioned discussion yields the following remark.

Remark 1: For the AF form and MPF form feedback Matsuoka systems satisfying perfect entrainment condition [27], when the feedback inputs $\mathbf{P^e}, \mathbf{P^f}$ are variables, an additional first order derivative input disturbance $k_f \tau_a b\dot{\mathbf{P}}$ (where $\mathbf{P} = \mathbf{P^e} - \mathbf{P^f}$) is introduced to the MPF form Matsuoka oscillator, which could cause *overshoot* and *phase distortion*[4] problems to the system. Thus, the feedback inputs of AF form Matsuoka oscillator are more effective than the feedback inputs of MPF form Matsuoka oscillator.

Next, in the AF form and the MPF form Matsuoka oscillator, in order to compare the impact of tonic inputs u_i^e, u_i^f and sensory feedback inputs p_i^e, p_i^f to the output amplitude bias, we introduce the following proposition.

Proposition 1: For the AF form or MPF form Matsuoka oscillator, if the tonic inputs u_i^e and u_i^f are complementary to each other (Definition 1), then the oscillation bias of z_i, u_i, p_i satisfy the following relationship:

$$\mathbf{B_Z} = ((2 + b - a)\mathbf{E} + \mathbf{W})^{-1}(b\mathbf{B_P} + \mathbf{B_U}) \quad (15)$$

where

$$\mathbf{B_Z} = [\text{bias}(z_1), \text{bias}(z_2), \text{bias}(z_3), \text{bias}(z_4)]^T$$

$$\mathbf{B_P} = [\text{bias}(p_1), \text{bias}(p_2), \text{bias}(p_3), \text{bias}(p_4)]^T$$

$$\mathbf{B_U} = [\text{bias}(u_1), \text{bias}(u_2), \text{bias}(u_3), \text{bias}(u_4)]^T$$

and $z_i = z_i^e - z_i^f$, $u_i = u_i^e - u_i^f$, $p_i = p_i^e - p_i^f$.

Proof: (See Appendix C). □

Proposition 1 shows that in both the AF form and the MPF form Matsuoka oscillator, there exists a binary linear relationship between the bias of u_i, p_i, and the bias of z_i. When both u_i^e, u_i^f are limited within $[0, 1]$ (according to the tonic input control design in [25]), since p_i^e, p_i^f are both nonnegative, the impact of p_i is larger than u_i as long as the coefficient of bias(p_i) is larger than the coefficient of bias(u_i). Hence, this attribute guarantees the efficacy and flexibility of the feedback mechanism within the Matsuoka CPGs, enabling sensory feedback inputs to significantly influence the CPG system's output pattern whenever contact occurs, regardless of the pattern of tonic inputs.

The disturbance caused by $k_f \tau_a b\dot{\mathbf{P}}$ in the MPF form Matsuoka oscillator (in Remark 1) and the dominance of $\mathbf{B_P}$ (in Proposition 1) during the contacts are further testified in a controlled variable test [as shown in Fig. 5(a)]. In this test, both the primitive AF and MPF forms of the Matsuoka oscillator (configured by Table I) were subjected to an identical time series of sensory signals (depicted in red) that were directly fed into each p_i. The tonic inputs were held constant for both CPGs' configurations. It is evident from the figure that the outputs of both CPGs are notably influenced by contact events whenever they occur, thereby confirming Proposition 1. By comparing the output patterns of the AF and MPF Matsuoka oscillators during and following a contact feedback event, we can highlight the significant drawbacks of the MPF form in processing these feedback signals as follows.

1) In segment A [in Fig. 5(a)], the rising edge of the contact signal produces significantly large $\dot{\mathbf{P}}$, prompting the MPF form of the Matsuoka oscillator to output considerably larger *overshoots* compared to the AF form. Due to the inherent elasticity of the Matsuoka CPG's oscillation, these enlarged *overshoots* can introduce substantial fluctuations into the system, resulting in rapid high-low-high control command of the corresponding actuator.

[3] The detailed derivation of this approximation can be found in the supplementary document S-Section 2.

[4] *Phase distortion* [29] is an important concept in signal processing, which refers to the alteration of the phase relationship between the various frequency components of a signal, resulting in changes to the original waveform shape.

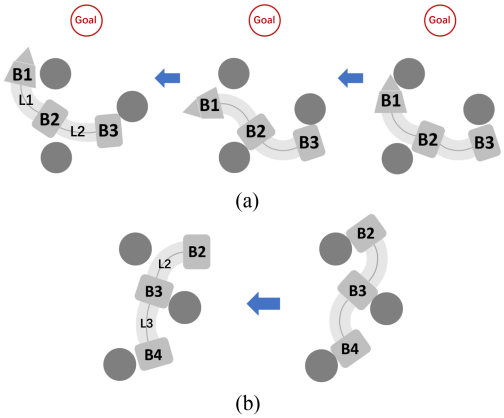


Fig. 8. (a) Example of reflexive mechanism on link L1. The red circle represents the target position. (b) Example of reflexive mechanism on L2 and L3 links.

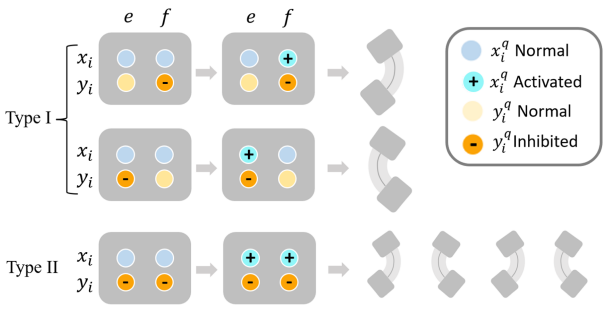


Fig. 9. Two types of soft body link actuation determined by the feedback inputs of the Matsuoka oscillator (assuming the outputs of the extensor and flexor controls left chamber and right chamber, respectively). In Type I actuation, $p_i^e > 0, p_i^f = 0$ or $p_i^e = 0, p_i^f > 0$, leads to the inhibition of y_i^e or y_i^f, and activation of x_i^e or x_i^f on the same side, which causes the opposite bending direction to the nonzero feedback input. In Type II actuation, $p_i^e > 0, p_i^f > 0$ inhibit y_i^e, y_i^f, leading to activation of x_i^e, x_i^f, which means free-response oscillatory actuation of the chambers.

2) In segment B, it is noticed that upon the disappearance of contact, the AF model's output signal promptly returns to its regular oscillation phase controlled by the constant tonic inputs. Conversely, the MPF model undergoes significant disruption due to the descending spike in the feedback input, leading to discernible *phase distortion*. Based on this observation, we conducted a detailed spectrum analysis of both output signals. Our findings reveal a phase difference of 9.78 and 10.31 rad for the primary frequency components at 0.875 and 2.657 Hz, respectively. This indicates a phase delay of approximately 1.5 periods, further emphasizing the distinct impact of the feedback inputs on the MPF model's performance.
3) In segment C, when unexpected feedback waves persist, the issues of *overshoots* and *phase distortion* occur frequently, resulting in a dilemma for the MPF form Matsuoka oscillator. It faces a challenge in accurately responding to contact and recovering from an incorrect control state once contact has been established.

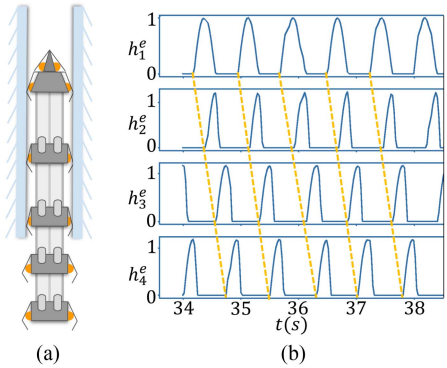


Fig. 10. (a) Soft snake robot traversing a narrow aisle using rectilinear gait. (b) Phase pattern illustration of CPG control command that can be used for both the slithering and rectilinear gait.

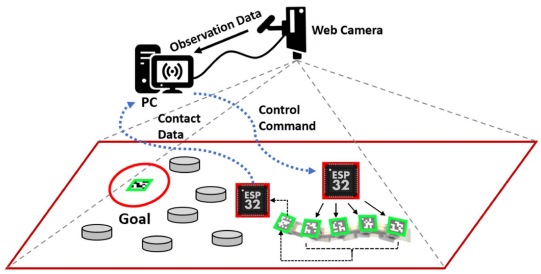


Fig. 11. Experiment setup of the contact-aware goal tracking locomotion task.

Overall, the properties of AF form Matsuoka oscillator show its flexibility and accuracy in reacting to the contact events. Based on this, we can further develop contact-aware controllers for the soft snake robot locomotion.

IV. Design of Controllers

In this section, we devise two distinct reactive control methods for the contact-aware locomotion of the soft snake robot, tailored to manipulate the sensory feedback inputs of the AF-form Matsuoka CPG system based on contact events. Specifically, 1) the AF-learning method integrates the notion of *hybrid control* [9] with biological insights drawn from reptiles' brains, specifically their utilization of distinct dorsal cortex regions for processing navigation commands and somatosensory signals [30]. It employs a learning-based, contact event-triggered controller to monitor the sensory data and adjust the sensory feedback coefficients within the Matsuoka CPGs of the soft snake robot [as depicted in Fig. 1(a)]. 2) The AF-local method, on the other hand, utilizes the *reflex control* paradigm [12]. It directly interfaces the contact sensory signals with the sensory feedback coefficients of the Matsuoka CPGs, following localized reflexive rules [as illustrated in Fig. 1(b)].

We consider contact events pertaining to the *passable environment* for the soft snake robot. This implies that the minimum gap between the surfaces of any two distinct obstacles (with convex geometry) exceeds the width of our soft snake robot,

considering fully compressed scale sensors. Specifically, this width measures approximately 73 mm.

Common contact situations for the soft snake robot can be categorized into two distinct groups as follows.

- Group I: None of the robot's body links detects simultaneous contact signals from both sides of its sensors.
- Group II: At least one body link of the snake robot registers contact signals from both sensor sides concurrently.

Taking into account the varying contact scenarios, Group I can be effectively managed using the slithering gait as the primary locomotion mode. However, for Group II, a specific rectilinear gait is necessary for successful locomotion. Our design of the reactive controllers incorporates this, where the AF-learning and AF-local methods initially operate on the slithering gait and transition to the rectilinear gait whenever a Group II situation arises. The unified switching module can be seamlessly integrated into the AF-learning and AF-local frameworks, as depicted in Figs. 6 and 7. In the latter part of this section, we will elaborate on the design of this gait switching mechanism.

A. Event-Triggered Learning-Based Sensory Reactive Controller

In the AF-learning method, we introduce the concept of hybrid control to a model-free learning-based control framework, which is composed of two controllers in the contact-aware goal-tracking task of soft snake robot—including a regular RL controller (C1) for goal-tracking locomotion control, and an event-triggered RL controller (R2) for contact reactive control, which only acts when the contact event-triggering condition is satisfied. The scheme of the controller is shown in Fig. 6. The reason for constructing the AF-learning method with two RL controllers is mainly due to the difference between goal-reaching and contact-reacting objectives.

- In the goal-reaching tasks, the actuators are controlled consistently to maintain propulsive slithering for the soft snake robot to approach the target positions, which requires continuous and periodic operations of the corresponding CPG's tonic inputs [25]. While the contact-reactive control aims to output the adjustments to the snake robot's behavior only when the robot is in contact with the obstacles. In this case, the contact events are neither continuous nor cyclic. Based on our analysis of feedback signals of the the Matsuoka oscillator in Section III and the sampled curve in Fig. 5(a), the features and timing of the contact-reactive actions are significantly different from the goal-reaching actions. As a result, we need different policies for the above-mentioned two objectives during the contact-aware locomotion tasks.
- In literature, people have proposed similar idea that adopts hybrid control concept to RL. For instance, a recent work [31] uses the parameterized action RL to solve control problem that requires both continuous actions and discrete actions to operate the agent, which has separated policy networks and a shared value network in the control scheme. In another state-of-the-art approach that study the bipedal locomotion control, the authors use totally separated learning-based sub-controllers for walking gait generator and reflexive balancing module [32].

A goal-tracking controller named free-response oscillation constrained proximal policy optimization option-critics with central pattern generator (FOC-PPOC-CPG) [25], is introduced as the C1 controller. The C1 controller takes the soft snake robot's dynamic states and a one-step action history [25, Figs. 3 and 4] as input observations and generates actions with a policy network to manipulate the tonic inputs and frequency parameters of the Matsuoka CPG network. The 4-D primitive action of C1 is $\boldsymbol{a} = [a_1, a_2, a_3, a_4]^T \in \mathbf{R}^4$ and are mapped to the tonic input vector $\mathbf{u}$ as follows:

$$u_i^e = \frac{1}{1+e^{-a_i}}, \text{ and } u_i^f = 1 - u_i^e, \text{ for } i = 1, \ldots, 4.$$

This function bounds the tonic input within $[0, 1]$.

For the R2 controller, we define the *contact event-triggering condition* as follows: At each time step, given the contact force vector $\bar{\mathbf{N}}$ and contact detection threshold ϵ_c. The event-triggering condition for the contact-aware scenario is $\|\bar{\mathbf{N}}\| > \epsilon_c$. When the event-triggering condition is satisfied, R2 is triggered to join the manipulation of the CPG system. The R2 controller shares the same reward function with C1 and operates on the same AF form CPG system as C1 does, but has different observations and actions.

Although it is not necessary for R2 to use the same learning algorithm as C1, for simplicity we also train R2 with PPO [33] in this work. In the obstacle-based locomotion scenario, there are in total 19 observation states for R2, denoted as $\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \ldots, \zeta_{19}\}$. $\zeta_1 \sim \zeta_4$ represents the dynamic states of the robot referenced on the goal position (ζ_1 represents the distance between snake's head COM to the goal position, ζ_2 is the velocity value on the direction pointing from snake's head to the goal, ζ_3 is the angle between goal and the snake's heading direction, and ζ_4 is the turning angular velocity of the snake's head [25]). $\zeta_5, \ldots, \zeta_8$ represent the real-time body curvatures of the 4 soft links. $\zeta_9, \ldots, \zeta_{12}$ record the R2 control actions from the last time step. ζ_{13} and ζ_{14} represents the frequency control option and its terminating probability by the C1 controller in the last time step, which contains the frequency information of the CPG system. $\zeta_{15}, \ldots, \zeta_{19}$ are the preprocessed contact forces. Similar to C1 that maps the 4-D actions to 8-D tonic inputs, the 4-D actions of R2 are mapped to fit the 8-D sensory feedback signals of the Matsuoka CPG network of the soft snake robot. Next, the 4-D action of R2 is $\tilde{\mathbf{a}} = [\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4]^T \in \mathbf{R}^4$ and map $\tilde{\mathbf{a}}$ to sensory feedback vector $\mathbf{p}$ as follows:

$$p_i^e = \frac{1}{1+e^{-\tilde{a}_i}}, \text{ and } p_i^f = 1 - p_i^e, \text{ for } i = 1, \ldots, 4. \quad (16)$$

This function also bounds p_i^e and p_i^f within $[0, 1]$. It is noted that, the definition of p_i^e, p_i^f varies depending on the specific contact reactive control method employed. Equation (16) only describes a way that maps R2's actions to the Matsuoka CPGs' sensory feedback inputs in the *hybrid learning* context.

The learning process of the whole control scheme [as shown in Fig. 1(a)] is: C1 is first trained in an obstacle-free environment in simulation. After C1 is converged, we fix C1 as a regular

controller for goal-tracking purposes. C1 policy is always effective regardless of the triggering of the contact events. Then we train R2 in the environment with randomly generated obstacle mazes in simulation until convergence. R2 is effective only when the contact event-triggering condition is satisfied. According to Remark 1 and Proposition 1, when the parameters of the AF form Matsuoka CPG system satisfy Table I, when R2 is effective, it will dominate the control of the CPG system (contact-awareness over goal-awareness).

B. Design of the Shared Reward Function

Now we present our design for the reward function shared by both learning based goal-tracking locomotion (C1) and contact-aware (R2) controllers. Our design will ensure that by maximizing the discounted sum of reward, the learned controller can achieve efficient locomotion and accurate set-point tracking.

To improve learning efficiency, we employ a potential field-based reward function. Artificial potential field (APF) is widely applied in planning problems and potential game theory [34], [35], [36] to accelerate the process of searching for the optimal strategy. The potential field can be classified into two categories – the attracting field for target reaching and the repulsive field for obstacle avoidance. In this work, we only use the attracting field, and consider the obstacles' positions as unobservable environment information. The attracting field function is defined as follows:

$$U_{\text{att}}(\boldsymbol{\rho}) = \frac{1}{2} k_{\text{att}} ||\boldsymbol{\rho} - \boldsymbol{\rho}_g||^2$$

where $\boldsymbol{\rho}$ is the coordinate of the agent and $\boldsymbol{\rho}_g$ is the coordinate of the goal. Coefficient k_{att} is a positive constant indicating the strength of the attractive potential field. Since the attracting gravity is always pointing toward the goal coordinate from any position of the map, the value of gravity force should be negative. By taking the negative gradient of U_{att}, we have the attracting force function

$$\mathbf{F}_{\text{att}}(\boldsymbol{\rho}) = -\nabla U_{\text{att}} = -k_{\text{att}}(\boldsymbol{\rho} - \boldsymbol{\rho}_g).$$

The reward is designed to encourage the goal-reaching, guided by the APF. We designed the reward to be composed of two rewards

$$R = \omega_1 R_{\text{goal}} + \omega_2 R_{\text{att}} \tag{17}$$

where $\omega_i, i = 1, 2, 3$ are constant weights. R_{goal} is the termination reward for reaching a circular accepting area centered at the goal

$$R_{\text{goal}} = \cos\theta_g \sum_{k=0}^{i} \frac{1}{l_k} \max\{\text{sgn}(l_k - \rho_g^r), 0\}$$

where θ_g is the deviation angle between the locomotion direction of the snake robot and the direction of the goal, l_k denotes the radius of the accepting area in task-level k defined by the learning curriculum in [25], for $k = 0, \ldots, i$. $\rho_g^r = \|\boldsymbol{\rho} - \boldsymbol{\rho}_g\|$ is the reference linear distance between the head of the robot and the goal, and the term $\max\{\text{sgn}(l_k - \rho_g^r), 0\}$ is used for determining whether the robot's head is within the accepting area of the goal. R_{att} is the reward function of the attracting potential field

$$R_{\text{att}} = \mathbf{v} \cdot \mathbf{F}_{\text{att}}(\boldsymbol{\rho})$$

where $\mathbf{v}$ is the velocity vector. The dot product $\mathbf{v} \cdot \mathbf{F}_{\text{att}}(\boldsymbol{\rho})$ represents the extent of the agent's movement on following the potential flow in the task space.

C. Local Reflexive Control of Contact-Aware Locomotion

In this section, we introduce another novel contact reactive controller, referred to as AF-local, which incorporates the sensory feedback control module grounded in the local reflexive mechanism proposed by Kano [12] and the sensory feedback characteristics of the AF form Matsuoka oscillator.

The local reflexive mechanism, originally described in [5] and [12], functions by allowing only the snake robot's links proximal to a contact sensor to respond to contact events. However, given the distinct structural features of our pneumatically actuated soft snake robot, including antagonistic actuators and partially tunable chambers, we have tailored specific rules for establishing the reflexive loop between the sensors and the sensory feedback inputs of the CPG network. These rules are designed to work synergistically with a learning-based controller (C1), which focuses on manipulating the tonic inputs of the AF form Matsuoka CPG to achieve goal-reaching objectives (see Fig. 7).

Considering the typical contact cases encountered by our soft snake robot while locomoting through obstacles (with only Group I contact situations), our local reflexive rules can be succinctly described as follows.

1) Since the snake robot maneuvered by the goal-tracking controller tends to adjust its heading direction toward the goal, the robot could be blocked by the obstacles right in front the robot's heading direction if the head joint cannot properly react to the contact and turn away from the obstacle. Thus, the snake robot's head should always bend in the opposite direction to the major contact event, which means that the ipsilateral chamber of L1 link to the contact side of B1's sensor will be actuated. Fig. 8(a) provides an example showing the reflexive behavior of the L1 link when the head sensor on B1 touches an obstacle.
2) In biological snakes' scaffold-based locomotion [13], a lot of snake species bend a large portion of their body links, such as compressed springs to generate propulsion forces. In our four-link soft snake robot, the latter two links act as the major source of the scaffold-based propulsion. Therefore, we let the snake robot's latter half body bend against any obstacle detected by the tail sensors to create scaffold-like propulsion. To be more specific, the ipsilateral chamber of L3 and L4 links to the contact side of B5's sensor are actuated in this contact situation.
3) Except for the head and tail links, the corresponding CPG nodes of the rest of the soft body links should refer to their neighboring links' contact states to determine their reflexive behaviors accordingly. To design the connection

between these CPG nodes' feedback inputs and their neighboring sensors, we need to account for the jamming case. As shown in Fig. 8(b), the B2, B3, and B4 rigid parts detect the contact events in a "left-right-left" order (could be "right-left-right" otherwise). A way of escaping from jamming is to decrease the bending curvature of L3, and inflate L2's ipsilateral chamber to B2's contact side to create more space for the slithering gait controlled by the goal-reaching controller.

Based on the above-mentioned features and former experience in designing local reflexive control rules [12], we design the topology of the sensor connection to each CPG node in the soft snake robot's "vertebrate" system. We first define $\mathcal{D}_i, i = 1, 2, 3, 4$ as the sets of paired sensors installed on B1∼B5 that are connected to the ith Matsuoka CPG node. For example, for the third CPG node in Fig. 7, the sensory feedback inputs p_3^e, p_3^f are connected to the sensors with subscripts listed by $\mathcal{D}_3 = \{3, 4, 5\}$, which indicates that the sensor signals $\{N_3^e, N_3^f, N_4^e, N_4^f, N_5^e, N_5^f\}$ are involved. In addition, we define the connection marker array $\mathbf{J} = [J_1, J_2, J_3, J_4, J_5] = [-1, -1, 1, 1, -1]$. The value in $\mathbf{J}$ is assigned based on the way of connection between the sensors and the CPG network

$$\begin{aligned} p_i^e &= \Sigma_{k\in\mathcal{D}_i}\left(I^e(J_k)N_k^e + (1 - I^e(J_k))N_k^f\right) \\ p_i^f &= \Sigma_{k\in\mathcal{D}_i}\left(I^f(J_k)N_k^e + (1 - I^f(J_k))N_k^f\right) \end{aligned} \tag{18}$$

where

$$\begin{aligned} I^e(x) &= \max\{0, \text{sgn}(-x)\} \\ I^f(x) &= \max\{0, \text{sgn}(x)\} \end{aligned}$$

for any $x \in \mathbf{R}^+$.

More specifically, the mechanism of (18) acting on the actuators of the soft snake robot can be explained as follows (referred to Fig. 9).

1) In the L1 CPG node, the sensors are connected to the same side of sensory feedback inputs p_i^e, p_i^f of L1 CPG. When one side of the B1 sensors is in contact, the L1 link is actuated in Type I, which bends toward the opposite direction to the triggered sensors.
2) In the L2 CPG node, the sensors on B2 are connected to the same side of sensory feedback inputs of L2 CPG, while the sensors on B3 are connected to the opposite side of sensory feedback inputs of L2 CPG. When only the B2 or B3 sensor is triggered, or both B2 and B3 receive contact feedback from the opposite side, L2 will behave in Type I. When B2 and B3 have contacts on the same side, both y_2^e and y_2^f will be inhibited, leading to Type II behavior of L2.
3) In the L3 CPG node, the sensors on B3 and B4 are connected to the opposite side of sensory feedback inputs of L3 CPG, while the B5 sensors are connected to the same side of sensory feedback inputs of L3 CPG. Consider a single sensor-triggered case, when only the B3, or B4, or B5 sensor is triggered, L3 will also behave in Type I. For two sensors triggered case: when only (B3 and B4) are triggered on the same side, or (B3 and B5) or (B4 and B5) are triggered on the opposite side, L3 will behave in Type I; when only (B3 and B4) are triggered on the opposite side, or (B3 and B5) or (B4 and B5) are triggered on the identical side, L3 will oscillate like Type II. For three sensor-triggered cases, when B3 and B4 are triggered on the same side opposite to the contact side of B5, L3 will behave in Type I, otherwise, L3 follows Type II oscillation.
4) In the L4 CPG node, the B5 sensors are connected to the same side of sensory feedback inputs of L4 CPG. When one side of the B5 sensor is in contact, the actuation of the L4 link follows Type I, which bends toward the opposite direction to the triggered sensors.

In order to compare the AF form Matsuoka CPG system with the conventional MPF form Matsuoka CPG system, we also develop MPF-local and MPF-learning controllers by replacing the AF form Matsuoka oscillator with MPF form Matsuoka oscillator in the two control methods introduced in Sections IV-C and IV-A. In Section V, we will comprehensively compare the performance of the AF-local, AF-learning, MPF-local, and MPF-learning methods.

D. Gait Switching Module for the Two-Sided Contact Situation

Consider the contact situations belonging to Group II, this occurs when the soft snake robot transverses a just passable narrow aisle [see Fig. 10(a)]. This situation leaves no room for the operations on slithering gait (including the scaffold-based reactions) mentioned in [12] to traverse through.

To solve this problem, we introduce the rectilinear gait to help the snake robot squeeze through narrow aisles using worm-like locomotion pattern. Realizing the rectilinear gait for the soft snake robot requires both the left and right chambers to be actuated simultaneously with stable phase delay (less than $\pi/2$, or equivalently expressed as a wave speed above 1 body length per period of the four-link soft snake robot in this work). According to the Type II actuation, since both $p_i^e > 0, p_i^f > 0$ lead to free-response oscillation of the ith CPG node, then the uniform value assignment of the whole CPG's sensory feedback inputs guarantees stable phase delay. Therefore we change the connection between sensor readings and CPG feedback inputs to the following topology:

$$p_i^e = \sum_{i=1}^{5} N_i^e, \quad p_i^f = \sum_{i=1}^{5} N_i^f. \tag{19}$$

In our CPG system configured by Fig. 7 and Table I, the phase delay in the slithering gait pattern can be applied to the rectilinear gait [as shown in Fig. 10(b)] by assigning one-sided outputs of the CPG system to both left and right chambers of the soft link.

In order to effectively integrate this gait with the feedback reactive patterns generated by the Matsuoka CPG system within our control framework, we develop a mechanism for sensing and quantifying the narrowing degree of the contacting obstacles. First, we design a new pair of outputs as control command for

the soft pneumatic chambers on top of z_i^e and z_i^f, denoted as

$$h_i^e = z_i^e, h_i^f = (1-\mu(\mathbf{N}))z_i^e + \mu(\mathbf{N})z_i^f \tag{20}$$

where

$$\mu(\mathbf{N}) = \max\{\text{sgn}(|\xi(\mathbf{N})| - \epsilon), 0\}$$
$$\xi(\mathbf{N}) = \frac{\sum_{i=1}^5 \hat{w}_i N_i^e - \sum_{i=1}^5 \hat{w}_i N_i^f + \delta^+}{\sum_{i=1}^5 \hat{w}_i N_i^e + \sum_{i=1}^5 \hat{w}_i N_i^f + \delta^+}. \tag{21}$$

The threshold $\epsilon > 0$ is used for determining whether the number of contact events on both sides of the robot are close enough, and the total number of contacts are dense enough. A small positive constant δ^+ is added to avoid division by zero problem, and avoid $\mu(\mathbf{N}) < \epsilon$ when no contacts are detected ($\mathbf{N} = \mathbf{0}$). The weight coefficients $\hat{\mathbf{W}} = \{\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \hat{w}_5\}$ and ϵ are selected based on the following principles.

1) Simultaneous contact on both sides of the head (B1) sensors must trigger rectilinear gait, regardless of the contact situations of the other body links.
2) When there are only both sides contact situation on the snake robot's body links (either two-sided contact or no contact), the gait pattern must be switched to rectilinear.
3) When the head (B1) sensors are in one-sided contact situation and no two-sided contact is detected by the B2 sensors, the gait must be switched to slithering.

In this article, we select $\hat{\mathbf{W}} = \{16, 8, 6, 4, 2\}$ and $\epsilon = 0.39$. Our experiment results in the later section show good performance to support this design. It is important to highlight that the rectilinear gait and its associated switching mechanism are universally applicable to AF-local, AF-learning, MPF-local, and MPF-learning methods. Essentially, the rectilinear gait and switching mechanism remain consistent across the various reactive control schemes presented in Figs. 6 and 7. As a result, in the experimental section, we have employed the AF-local method shown in Fig. 7 as a typical instance to showcase the effectiveness of this adaptable switching mechanism.

V. Experiments

A. Signal Communication and Obstructed Environment Setting

The planar dynamic states of the soft snake robot [25, Fig. 4] are captured and calculated by a web camera (works under 120 Hz) hanging on the ceiling of the experiment room. We use Aruco [37] to detect and localize QR codes attached to every rigid body of the snake robot and the goal position. Fig. 11 shows the experiment setup for the real snake robot goal-reaching tasks. In this work, we update two major parts of the experiment settings compared to our previous work [25] as follows.

1) In the signal communication part, each ESP32 chip collects contact sensor information from local I2C and shares the data with the head chip through WiFi. In every time step, the head ESP32 chip packs all the sensor data and sends it back to the PC controller. The controller program running on a desktop computer receives the observation states from the web camera and the robot, generates the 8-D PWM duty ratio vector mapped from the CPG output and passes it to the ESP32 chips on B1∼B5 through WiFi communication. A three-step procedure is designed to translate the CPG output to the PWM control commands for the robot actuators.
   a) STEP 1, determine the gait and select the corresponding output from the CPG system according to (20).
   b) STEP 2, multiply the modified CPG outputs h_i^e, h_i^f with a normalization ratio a_φ (see Table I) obtained by calculating the boundary of z_i^e and z_i^f sampled from contact-free trials. When there are no contacts, $p_i^e + p_i^f = 0$, the CPG outputs are normalized to the values within $[0, 1]$ (according to the BIBO property of the Matsuoka CPG [26], [38]). When $p_i^e + p_i^f > 0$, the values of $a_\varphi h_i^e$ and $a_\varphi h_i^f$ could exceed 1 and need to be clipped with $\min(\cdot, 1)$.
   c) STEP 3, map $\min(a_\varphi h_i^e, 1), \min(a_\varphi h_i^f, 1)$ to the PWM duty cycle in $0\% \sim 100\%$ to control the left chamber's valve and right chamber's valve, respectively.

   The communication rate between the PC controller and the snake robot is 30 Hz, and the maximum communication delay is below 0.03 s.
2) In the environment setting, a number of tin cans filled with stones and sand are placed in the experiment field as obstacles. Each vertical peg in Fig. 11 represents a cylinder tin can with a diameter of 100 mm and height of 80 mm. The average weight of the obstacles is around 1.1 kg each, and the weight of the soft snake robot is 0.7 kg (including batteries). It has been tested to ensure that any collision caused by the soft snake robot will not move the obstacles.

B. Simulated Training and Evaluation

RL configuration: In the simulated training part, the goal-reaching controller C1 is a pretrained module as configured in [25, Sections V.A and VI.A.2)]. In this work, the contact-aware regulator R2 in AF-learning and MPF-learning controllers is trained in a goal-reaching task with a randomly generated 6×5 obstacle maze. During the training process of R2, the distance between the robot and the goal is fixed to 1.5 m. The deviation angle between the snake robot and the goal is initially sampled from $0 \sim 60°$ with a uniform distribution. In the simulator, the distance between every two obstacles is sampled between $120 \sim 180$ mm. The coordinate of each obstacle is added by an additional clipped standard Gaussian noise ($\omega \sim \mathcal{N}(0, 1)$, clipped by $-0.01 < \omega < 0.01$). The method of simulating contact sensors is introduced in Section II-B. To compensate for the mismatch between the simulation and the real environment, we employ a domain randomization (DR) technique [39], in which a subset of physical parameters are sampled from several uniform distributions. The range of distributions of DR

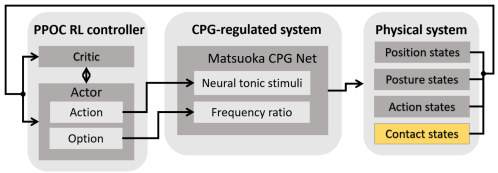


Fig. 12. Flowchart depicting the C1+ Method. The term "C1+" encapsulates two distinct aspects. (a) RL component of C1+ integrates contact states to the aforementioned C1 module as additional observations. (b) Its CPG component is modeled using the original Matsuoka oscillator, as described in (3), to generate rhythmic patterns.

parameters used for training can be found in the supplementary document.[5]

Matsuoka CPG configuration: In this study, the constant parameters of the Matsuoka oscillator, as configured in Table I, were determined using an evolutionary parameter optimization method tailored for CPG systems [25], [40] in a contact-free environment. Given this condition, where $b \gg 1$, Proposition 1 ensures that $\mathbf{B_P}$ exerts a dominant influence on the system's output compared to $\mathbf{B_U}$. Therefore, this configuration is applicable to both the AF and MPF forms, as well as the original form of the Matsuoka oscillator. The optimality of the configured parameters to both AF form and MPF form Matsuoka CPGs has also been confirmed in a contact-rich environment (refer to the Supplementary document).

Task specification: In the contact-aware locomotion task, the robot is required to traverse an array of obstacles and reach the randomly generated goals. Similar to the real-world setting in Fig. 11, there is also an accepting radius in the simulation for each goal-reaching task, which means that the robot needs to be close enough to the goal in order to succeed and receive a terminal reward. At each time step, the robot also receives a reward from the potential field defined in Section IV-B. If the agent reaches the accepting region of the current goal, a new goal is randomly sampled. In the failing situation, when the robot is jammed by obstacles for a certain amount of time, the desired goal will be resampled and updated. We set the time threshold for this failing condition to 900 ms. In addition, if the linear velocity of the snake robot stays negative in the goal direction for over 360 time steps (each time step is about 20 ms), the goal-reaching task is also judged as a failure and trigger the resampling of the new task. After the end of each learning episode, both the physical state of the soft snake robot and the dynamic states of all Matsuoka CPG nodes within the system are initialized to zero.

Training/average evaluation score comparison: According to the above-mentioned task specification, we train the AF-learning, MPF-learning, and C1+ methods, and compare their training scores with the average evaluation scores of AF-local, MPF-local in the same environment. As illustrated in Fig. 12, C1+ functions as a benchmark for comparison due to its direct end-to-end mapping from contact and environmental observations to CPG tonic inputs. In this simulated experiment, the shared C1 controller has been pre-trained in the obstacle-free environment till convergence [25]. This means all AF-local, AF-learning, MPF-local, and MPF-learning methods use the identical frozen C1 policy throughout learning/evaluation in contact-aware locomotion experiments. It is worth noting that both AF-local and MPF-local methods utilize fixed contact reaction strategies, serving as benchmarks for the learning-based contact reactive methods (AF-learning, MPF-learning, C1+). Consequently, their final scores are calculated by averaging the episode scores throughout the evaluation process. On the other hand, the R2 controllers of AF-learning and MPF-learning methods are trained respectively in the obstacle-based goal-tracking tasks for 3000 episodes till convergence. Different from the other four methods, the C1+ controller is first trained in the contact-free environment, then transferred to the obstacle-based environment, and is also trained for 3000 episodes. As the rectilinear gait and its switching mechanism constitute a fixed, independent module compatible with AF-learning, AF-local, MPF-learning, and MPF-local methods and remains unaffected by any learning algorithms, there is no need to incorporate the gait into the training score comparison experiment. Furthermore, as C1+'s CPG component utilizes the original Matsuoka oscillator without incorporating sensory feedback, it is incompatible with the rectilinear gait switching module illustrated in Fig. 6. Consequently, all methods outlined in Fig. 13 are trained or evaluated excluding the rectilinear switching module.

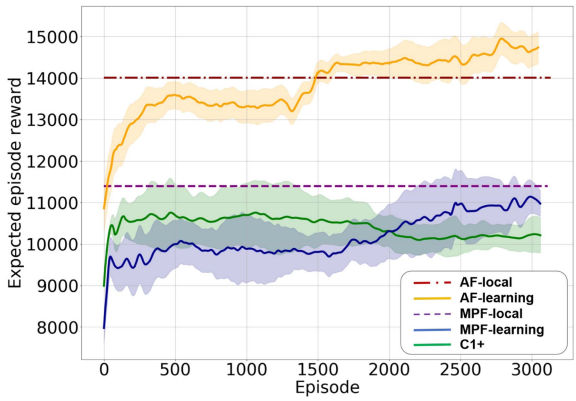


Fig. 13. Comparison of learning process rewards (AF-learning, MPF-learning, C1+) and average evaluation scores (AF-local, MPF-local) obtained in an obstacle-based training environment, averaged across three seeded trials. The shaded area depicts the standard deviation of the value curves for each method.

From Fig. 13, it can be observed that the AF-learning method attains the highest reward and stands out as the sole learning approach that continues to progress throughout the learning process. Among the other methods, AF-local is notably the closest to the AF-learning method in terms of average reward, highlighting the superiority of AF-related approaches. In addition, although the MPF-learning method initially lags slightly behind the C1+ method, it continuously improves and eventually converges to a higher score, surpassing the reward of C1+. As discussed in Remark 1, the lower scores observed in the MPF series methods compared to the AF series methods may be attributed to the influence of $k_f \tau_a b \dot{\mathbf{P}}$ (introduced by the varying contact signals) in the MPF form Matsuoka oscillator, which complicates the R2

[5] The whole training process of each method runs on four simulated soft snake robots (Rendered by Nvidia Flex) on a workstation equipped with an Intel Core i7-9700K, 32GB of RAM, and one NVIDIA RTX2080 Super GPU.

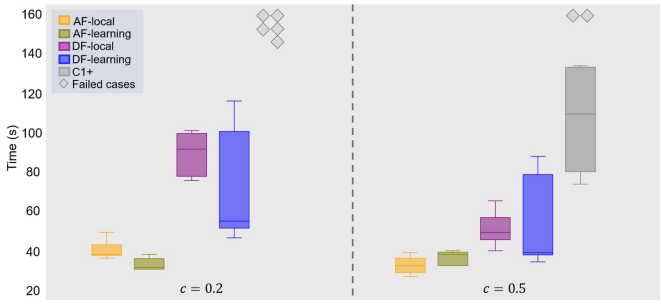


Fig. 14. Statistics of escaping time of the proposed methods and the baseline.

RL controller's ability to process sensory feedback signals from the CPG system.

C. Performance Analysis in Real Robot Experiments

In this section, we compare the performance of all five methods (mentioned in Section V-B) in contact-aware soft snake robot locomotion tasks in the real world. Furthermore, we test the performance of the top two methods in more challenging obstacle-based environments.

1) Escaping Experiment: In the real-world contact-aware locomotion scenario, we design an escaping task to distinguish the strength and weakness of the contact-aware controllers (listed in Section V-B).

Environment settings: The escaping task is designed for the following principles.

1) The allocation of the obstacles should create a narrow passage for the snake robot, with more contact opportunities and a sharper tuning angle to test the overall capability of the controllers in escaping the jamming situations. In addition, the narrow space also limits the amplitude for regular body oscillation of the snake robot.
2) The obstacles should be allocated to obstruct the goal-reaching behavior. This is to test the coordination of the goal-reaching module (C1 controller) and contact reactive module (local reflexive or R2 method) in the compared controllers.
3) The allocation of the obstacles should include the situation where only the latter half links of the robot are stuck in the obstacles. This is for telling whether the controller relies mostly on its head steering to escape from the obstacles.
4) The obstacles should be placed more densely in reality to test the generality of the compared controllers.

Based on the above-mentioned principles, the obstacles in the escaping task are allocated, as shown in Fig. 15. In the escaping task, the distance between every two obstacles ranged from 85 to 150 mm. The robot is initially bending to its left, and placed at a position where four rigid bodies are in contact with the obstacles from different sides. The exit direction (left) of the obstacle region is intentionally set opposite to the goal direction (right). The distance between the exit of the obstacle region and the goal is 540 mm, which is close to the length of the snake robot.

Performance statistics: According to the free oscillation tonic input property of coefficient c in [25, Appendix B-D], as the value of c increases, it can increase the oscillation amplitude of the outputs of FOC-PPOC-CPG controller and therefore improve its sim-to-real adaptability in the locomotion tasks. However, the value of c should not be larger since a higher free oscillation tonic input could decrease the goal-tracking accuracy. As a result, we separate the experiment into two groups with $c = 0.2$ and $c = 0.5$, respectively. For each value of c, we run five trials for each control method.[6]

We record and compare the finishing time of the escaping task of each controller. As shown in Fig. 14, AF-local and AF-learning methods outperform the other methods in the escaping task in both speed and stability. The increase of c from 0.2 to 0.5 does not significantly improve the performance of both AF methods. The main reason is that the sim-to-real adaptability of the AF methods is already good. MPF-learning method's average finishing time is shorter than MPF-local when $c = 0.2$, but is less stable than MPF-local, with the task finishing time varying from 42 to 120 s. When $c = 0.5$, MPF-local method outperforms MPF-learning in both speed and stability. C1+ method cannot reach the goal in every trial when $c = 0.2$. However, with the increase of c to 0.5, the adaptability of the C1+ controller is also improved so that it succeeds in a few of the trials. It is noted that, although MPF-learning converges to a lower reward level than C1+ method during the learning process (see Fig. 13), its adaptability to the harder unseen task (in sim-to-real) is better than C1+ method. Generally, the results in Fig. 14 further verify the advantages of AF feedback Matsuoka oscillator predicted by Remark 1 and Proposition 1.

Discussion: We can further compare the sample output trajectories of contact feedback signals and control commands for different control methods (refer to Fig. 16, and the Supplementary document) to analyze the special features of AF-local and AF-learning methods. It is noted that in these figures, the positive and negative values are related to the extensor and flexor of the CPG system, as well as the left and right of the snake body, respectively.

First, we investigate the trajectory sample of the AF-local method in the escaping task on the basis of AF-local mechanisms illustrated by Fig. 7. As shown in Fig. 16, we highlight four time intervals of the trajectory that present typical local reflexive control in the AF-local controller [the robot's body postures before and after contacts at intervals (a)∼(d) are captured by Fig. 15(a)∼(d)]. Here, we select time intervals (a) and (c) for discussion. At time interval (a) of Fig. 16, both CPG nodes at L3 and L4 are first influenced by the contact from the N_5^f ($N_5 < 0$), so the flexors of CPG nodes in L3 and L4 are activated to open the right valves of L3 and L4, which results in both links bend to the left in Fig. 15(a). Then L3's CPG output is influenced by N_3^f, which will activate L3's left chamber. At time interval (c) of Fig. 16, the CPG node at L2 is influenced by the superposition of N_3^e and N_2^f, and is supposed to activate its flexor to open the right valve of L2, which results in L2 bend

[6] The related video is available at: https://youtu.be/FyiTn4AyBjY.

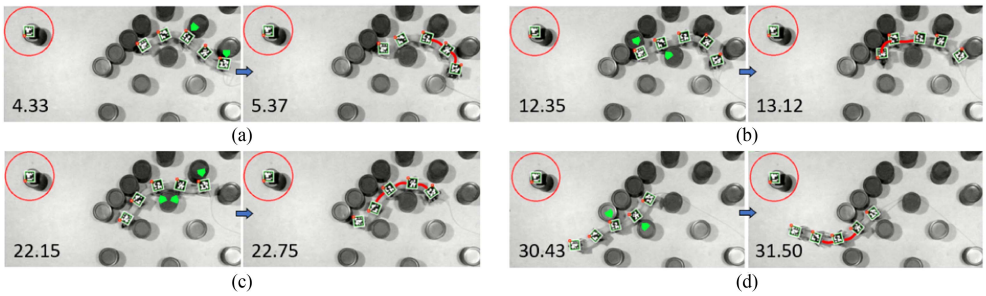


Fig. 15. Sample screenshots of performance of the AF-local method in a goal oriented escaping task from the obstacles. Each pair of pictures shows the local reactive behavior of the soft snake robot before and after contacts.

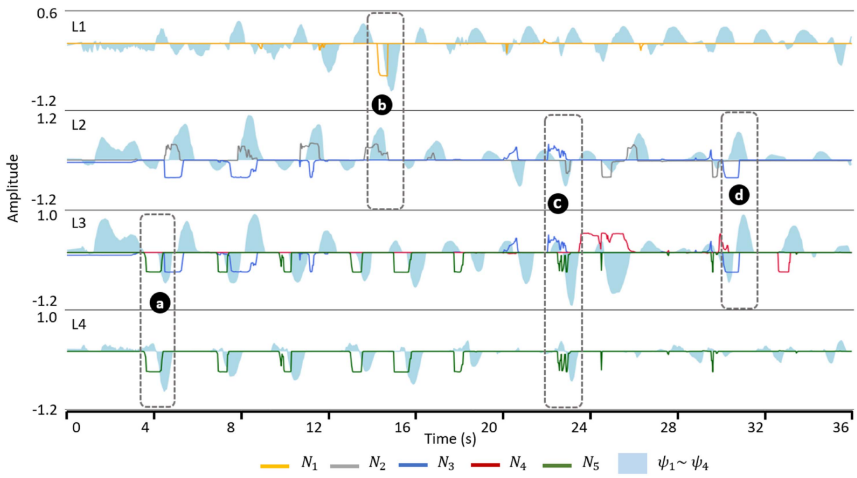


Fig. 16. Recorded sensory input and CPG output of each body link of the soft snake robot controlled by the AF-local method in the goal-oriented escaping task. The variable ψ_i refers to the Matsuoka CPG's output signal defined in [25, (3)], which is equivalent to $a_\varphi(h_i^e - h_i^f)$.

leftward in Fig. 15(c). Due to the whole snake robot's tendency to turn right toward the target position, the amplitude of L2's CPG output signal is smaller than expected. The CPG node at L3 is influenced by the superposition of N_3^e and N_5^f, which also causes L3's right chamber activated to bend to the left side. The CPG node at L4 is influenced by N_5^f, which activates L4's right chamber and bends link L4 to the left. From the above-mentioned behavior of the CPG outputs, we can verify that the experiment results match the local reflexive mechanism illustrated in Fig. 7.

Similarly, from the sampled trajectories of the MPF-local method (refer to the Supplementary document), we can conclude that the sensory inputs and the CPG outputs for all body links satisfy the local reflexive mechanism determined by Fig. 7. However, when comparing MPF-local behavior to the AF-local behavior in Fig. 16, the MPF-local controller produces significantly larger *overshoots* even when the contact signals are small. The MPF-local controller also suffers from the *phase distortion* issue in most contact events. However, the combination of numerous contact events and the turning locomotion command complicates the clear identification of phase distortion. To address this, we devise a targeted miniexperiment to demonstrate the presence of phase distortion in the MPF form of the Matsuoka CPG system during real robot contact-aware locomotion tasks. As depicted in Fig. 17, a brief contact force of approximately 2 N, lasting for a duration of $0.5 \sim 0.7$ seconds, is exerted once on the tactile sensor located on the right side of the robot's head joint. This operation occurs when the robot is moving toward its target on the left side, guided by the AF-local and MPF-local policies, respectively. Following the local reflexive connection rule outlined in Section IV-C, the L1 link promptly turns left after experiencing contact on the right side. We then monitor the recovery process postcontact, plot, and analyze the joint trajectory of the CPG's output command to the L1 link actuator. The head reaction behavior is selected owing to the clean and decoupled sensory feedback connection to the L1's CPG node, which makes the reaction phase particularly observable and easy to assess.

In this experiment, we consistently observed the absence of one left turning phase in the period following the contact reaction phase in cases controlled by MPF-local, whereas AF-local controlled cases did not exhibit this issue.[7] As illustrated in

[7] The related video is available at https://youtu.be/iH38aop9ZDg

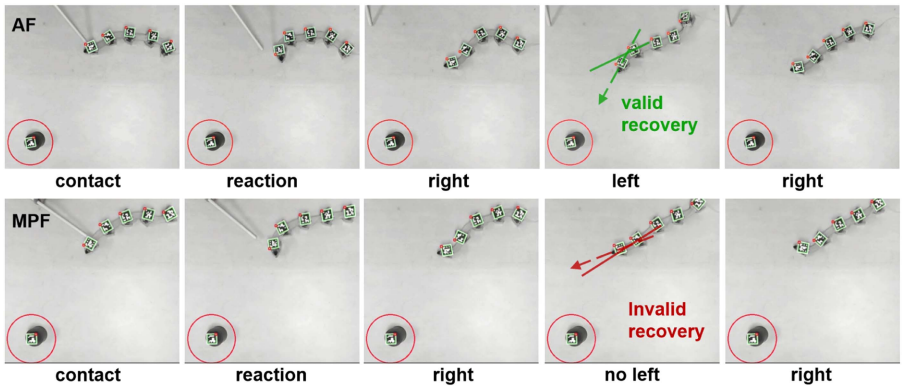


Fig. 17. Isolated phase distortion test on AF-local and MPF-local methods.

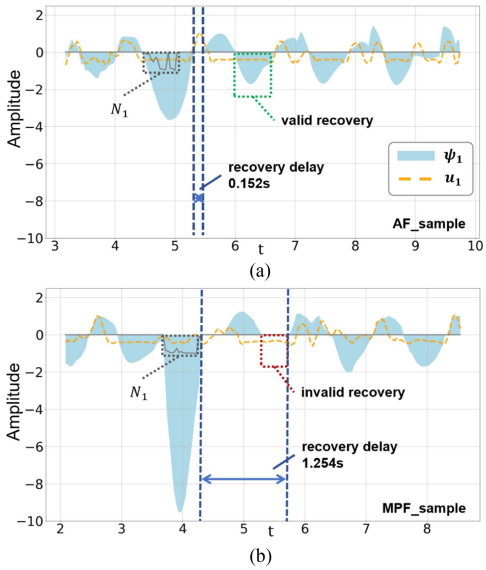


Fig. 18. (a) AF-local and (b) MPF-local phase patterns of the L1 link's CPG node in response to an isolated contact force. The blue shaded regions are bounded by the curve of ψ_1, which refers to the Matsuoka CPG's output signal defined in [25, (3)], equivalent to $a_\varphi \left(h_1^e - h_1^f\right)$.

Fig. 18(b), the missing left turning phase results in an extra cycle of delay in the recovery from the contact reactive pattern back to its original oscillation pattern, thus is named as an invalid recovery. The AF-local controller demonstrates the ability to perform a valid recovery, featuring a minimal latency interval between the contact reactive phase and the subsequent period, and distinguished by a clear left turning phase after the disturbance. We conduct ten trials for both the AF-local and MPF-local methods, with the results indicating that the average recovery delay for the AF-local method is around 0.148 ± 0.031 s, while for the MPF-local method, counting the additional invalid recovery time, is 1.285 ± 0.103 s. This invalid recovery signifies the failure to resume the precontact locomotion pattern within the CPG system, as evidenced in Fig. 17, where a proper recovery from the right-side contact event during left turning should manifest as a consistently left-biased oscillation. The results of this experiment corroborate the phase distortion issue observed in the MPF form of the Matsuoka oscillator, as anticipated by the output trajectory of the primitive CPG node depicted in Fig. 5.

These observations further verify Remark 1, that the first order derivative term $k_f \tau_a b \dot{\mathbf{P}}$ will seriously interfere with the control of MPF form CPG system when the contact feedback signals are densely emerging, and therefore hinder the performance of contact-aware locomotion.[8]

The issue of the output wave response can also be observed in MPF-learning (refer to the Supplementary document). With more chaotic sensory feedback signals from the RL event-based controller R2, the CPG outputs also show disturbed behaviors, which significantly slow down the locomotion in the escaping task. It is worth noting that, due to the black-box property of the learning-based method, both AF-learning and MPF-learning methods send more complex sensory feedback signals to their CPG systems. However, we can still observe clear and coordinated oscillation in the sample performance of the AF-learning method (refer to the Supplementary document). This is also because AF series methods are free from the disturbances of the $k_f \tau_a b \dot{\mathbf{P}}$ term.

In conclusion, the results and analyses in the escaping tasks show AF-local and AF-learning controllers excel in contact-aware locomotion of soft snake robots.

2) General Performance of AF Series Methods in Difficult Contact-Aware Locomotion Tasks: In this section, we test the two methods with the best performance in a more complicated scenario, which requires the soft snake robot to traverse a densely distributed obstacle maze to reach multiple targets by sequence. Although the positions of the initial targets are predetermined by specific geometries devoid of randomness, the positional information pertaining to the subsequent goals are only disclosed to the robot controllers upon their successful achievement of the preceding objectives.

The primary challenges in this task stem from the intricate interplay between contact awareness and goal-tracking objectives, which are densely intertwined. Given the control agents' lack of knowledge regarding the positions of obstacles, the soft snake

[8] The locomotion performance of the MPF-local method can be observed in videos https://youtu.be/_pFegKgscn8 and https://youtu.be/TZakvxSAFSY.

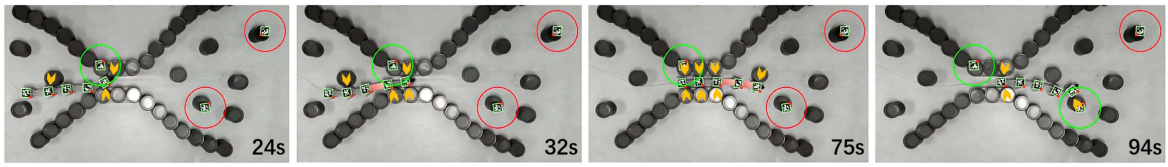


Fig. 19. Key frame screenshot of gait switching before and after the snake robot traversing through a narrow aisle (the green circles indicate reached goals and red circles indicate desired goals, the green arrows point to the sensors in contact at certain time step captured by the screenshot).

robot's navigation toward each goal is frequently disrupted and diverted by these unseen barriers. Consequently, this results in the robot encountering significantly sharper turning angles and unforeseen jamming scenarios. Furthermore, the unpredictable nature of contacts within the obstacle maze introduces an element of randomness that renders the timing and location of detours utterly unforeseeable for any control method during the robot's locomotion toward its goals. We show the detoured paths and shortcuts for each task (goals allocated on the square or triangle vertices) in video "square.mp4" and "triangle.mp4" to justify the adaptability of the two controllers without the help of path preplanning or any positional information of the obstacles. Both methods demonstrate their proficiency in visiting the goals in the prescribed sequence, even amidst challenging tracks characterized by a heightened density of blocking obstacles and steeper turning angles (as shown in the Supplementary document). Notably, the experimental outcomes underscore the superiority of the AF series controllers in locomoting through environments abundant with contacts, as well as the remarkable adaptability of the AF-form Matsuoka oscillator in responding to feedback signals characterized by their high complexity and variability.

3) Performance of AF-Local Method Equipped With Rectilinear Gait Mechanism: To investigate the efficacy of the gait switching mechanism between the slithering and rectilinear gait during the intelligent goal-tracking locomotion in a contact-rich environment with the passable narrow aisles, we design an experiment with goals and obstacles allocated as Fig. 19 shows. We integrate the switching module to the AF-local controller according to the scheme in Fig. 7. To guide the soft snake robot traverse the aisle to complete this task, we allocate goals on both sides of a narrow aisle with about half length of the snake robot (260 mm) and set the width of the path as 74 mm, which is 2 mm wider than the minimum width [see Fig. 2(c)] of the rigid body of the snake robot after the scale sensors are fully squeezed.

The sample trial (video "gait switching performance.mp4") shows a successfully task completion process of the soft snake robot with clear gait switching behaviors to the rectilinear gait when traversing the aisle. From Fig. 20, it is observed that $\mu(\mathbf{N})$ only drops below zero when the head or the major part of the snake robot links are inside the aisle. Fig. 19 shows the screenshots of the switching process in the experiment. First, it takes the robot about 8 s to recognize the entrance of the aisle. Using the contact buffer introduced in Section IV-D, the switching condition could be satisfied without rigorously simultaneous contacts on both sides of the snake's head. It takes about 60 s for the soft snake robot to squeeze through the aisle (head in, tail out). It is noticed that when $t = 75$ s, the one-sided contact to the

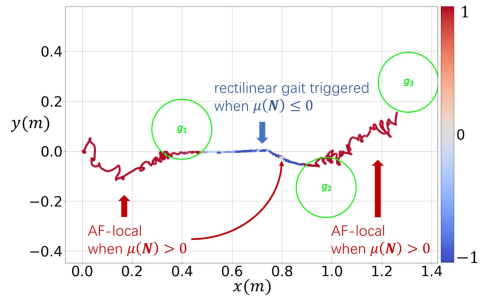


Fig. 20. Example head trajectory of pattern switching between local reflexive slithering gait and rectilinear gait governed by the switching function $\mu(\mathbf{N})$.

head sensor significantly breaks the sum value balance between all the left and right contacts, and causes $\mu(\mathbf{N}) > 0$. As the slim red arrow emphasizes in Fig. 20, the controller shortly switches to the slithering gait and turns the robot's head toward the second goal and then switch back to the rectilinear gait after the head contact vanishes. At $t = 94$ s, the head sensor receives one side contact signal after reaching the second goal, and switches to the slithering gait again, until reaching the target.

In summary, the experimental findings conclusively demonstrate that our switching mechanism, rooted in the feedback characteristics of the AF-form Matsuoka CPGs, has effectively facilitated the detection and traversal of navigable narrow corridors during the soft snake robot's contact-aware locomotion tasks. This experiment not only validates the AF-form Matsuoka CPG system's resilience in managing simultaneous bilateral contact events but also underscores the versatility of the soft snake robot, capable of executing both slithering and rectilinear gaits utilizing the same set of actuators.

VI. Conclusion

This article establishes a novel framework for the contact-aware intelligent locomotion control of a soft snake robot. This framework is an organic integration of hardware design, feedback mechanism study through a bioinspired CPG system, and implementation of sensory feedback control schemes. The proposed approaches are able to achieve promising performance in both simulation and real robots in several contact-aware locomotion tasks with densely allocated obstacles. Our novel method tackles jointly contact sensing, contact reacting controls and a gait switching mechanism in the contact-aware locomotion control of the soft snake robot. Our work brings inspiration for both the distributed reflexive method and learning-based control method and forms the basis to design and

control of soft snake robots that can pass through environments with unpredictable and dense obstacles.

For future study, the contact module and design can be enhanced with the consideration of more advanced materials and structures to improve contact sensitivity and locomotion efficiency for more challenging environments (e.g., underwater contact or uneven and compliant terrains). The tactile information and the locomotion gait can also be enriched by increasing the number of body links of the snake robot. More investigation is also needed to understand the influence of couplings among primitive AF form feedback Matsuoka oscillators in the CPG network so that the variation of couplings can be utilized to improve the control performance.

The proposed learning-based controller is mainly reactive to contacts but does not leverage obstacles to aid the locomotion. To achieve so, the control system may need to employ depth visual information and physics properties of the obstacles and a trajectory planning module. Another limitation is that we do not allow the soft snake robot to perform any mixed gait (e.g., half-slithering half-rectilinear). However, such mixed gaits for the snake robot are important to enable the artificial snakes to achieve better environmental adaptability. From bioinspired perspective, the natural snakes have been witnessed using mixed gaits when traversing the cluttered terrains [18], [19]. Developing locomotion controllers with mixed gaits could be an interesting future direction.

APPENDIX A
DATA

TABLE I
PARAMETER CONFIGURATION OF THE MATSUOKA CPG NET CONTROLLER FOR THE SOFT SNAKE ROBOT ($i = 1, 2, 3, 4$, $w_{0,1} = w_{5,4} = 0.0$)

Parameters	Symbols	Values
Amplitude ratio	a_ψ	2.0935
Mutual inhibition weight	a	4.6062
Self-inhibition weight	b	10.0355
Discharge rate	τ_r	0.7696
Adaptation rate	τ_a	1.7728
Period ratio	k_f	1.0
Coupling weights	$w_{i-1,i}$	8.8669
	$w_{i+1,i}$	0.7844

This section includes the parameter configuration of the Matsuoka CPG network and the hyperparameter setting of DR for the experiment.

APPENDIX B
PRELIMINARY

This section is mainly cited from [25] as the key preliminary information for the theoretical derivations in this article.

A. Important Concepts by Definitions

Definition 1: (Complementation) For two real signals $u(t)$ and $v(t)$, and a known bounded range $\mathcal{D} : [a, b]$ where $\mathcal{D} \subseteq \mathbf{R}$, we say $u(t)$ and $v(t)$ are complementary to each other in range $\mathcal{D}$ when $u(t), v(t) \in \mathcal{D}$ for all $t \in \mathbf{R}^+$ and $u(t) + v(t) \equiv b - a$.

Definition 2: (Entrainment) Given a neural oscillator system with its natural frequency $\omega_n > 0$. If the neural oscillator's output is synchronized to the coupled input with frequency ω, then this system is entrained with the coupled input signal. The relation between the neural oscillator's output and the coupled input signal is called *entrainment*. If the two signals are *perfectly entrained*, they are supposed to have the same oscillation amplitude and bias in addition to the synchronized oscillation frequency.

B. Describing Function Analysis of the Matsuoka Oscillator

According to Fourier theory, we denote the main sinusoidal and constant component in Fourier expansion of the vanilla state $x(t)$ as

$$x_{\mathcal{F}}(t) = A\cos(\omega t) + d = A(\cos(\omega t) + r) \tag{B.1}$$

where $r = d/A, r \in R$ is the ratio of bias to the amplitude of the signal. We assume $x_{\mathcal{F}}(t)$ only contains cosine term for simplicity. And because this chapter only discusses amplitude and bias properties of the signals, such simplification will not affect the following derivations. We use $z_{\mathcal{F}}(t) = g(x_{\mathcal{F}}(t)) - \epsilon(t) = \max(x_{\mathcal{F}}(t), 0) - \epsilon(t)$ to represent the main sinusoidal property of $z(t) = g(x(t)) = \max(x(t), 0)$. In a single period $\left[-\frac{\pi}{\omega}, \frac{\pi}{\omega}\right]$

$$g(x_{\mathcal{F}}(t)) = \begin{cases} 0 & \text{elsewhere} \\ A(\cos(\omega t) + r) & t \in \left[-\frac{\arccos(-r)}{\omega}, \frac{\arccos(-r)}{\omega}\right]. \end{cases}$$

Using Fourier expansion, the output state $z_{\mathcal{F}}(t)$ can also be expressed as

$$\begin{aligned} g(x_{\mathcal{F}}(t)) &= g(A(\cos(\omega t) + r)) \\ &= z_{\mathcal{F}}(t) + \epsilon(t) \\ &= A(K(r)\cos(\omega t) + L(r)) + \epsilon(t) \quad (n \geq 1) \end{aligned} \tag{B.2}$$

where $\epsilon(t)$ is the summation of all remaining high frequency terms in the Fourier expansion of $z_{\mathcal{F}}(t)$

$$K(r) = \begin{cases} 0 & (r < -1) \\ \frac{1}{\pi}\left(r\sqrt{1-r^2} - \cos^{-1}(r)\right) + 1 & (-1 \leq r \leq 1) \\ 1 & (r > 1) \end{cases} \tag{B.3}$$

$$L(r) = \begin{cases} 0 & (r < -1) \\ \frac{1}{\pi}\left(\sqrt{1-r^2} - r\cos^{-1}(r)\right) + r & (-1 \leq r \leq 1) \\ r & (r > 1). \end{cases} \tag{B.4}$$

The derivation of $K(r)$ and $L(r)$ are based on Fourier series analysis. Both $K(r)$ and $L(r)$ are constrained by $-1 \leq r \leq 1$ for $x_{\mathcal{F}}(t)$ to be nonnegative in the period $\left[-\frac{\pi}{\omega}, \frac{\pi}{\omega}\right]$.

When $t \in \left[-\frac{\arccos(-r)}{\omega}, \frac{\arccos(-r)}{\omega}\right]$, $z_{\mathcal{F}}(t) = x_{\mathcal{F}}(t)$, we have

$$\begin{aligned} \epsilon(t) &= x_{\mathcal{F}}(t) - A\{K(r)\cos(\omega t) + L(r)\} \\ &= -\frac{A}{\pi}\left\{\left(r\sqrt{1-r^2} - \arccos r\right)\cos(\omega t) + \sqrt{1-r^2} \right. \\ &\quad \left. - r\arccos r\right\}. \end{aligned}$$

When $t \in \left[-\frac{\pi}{\omega}, -\frac{\arccos(-r)}{\omega}\right] \cup \left[\frac{\arccos(-r)}{\omega}, \frac{\pi}{\omega}\right]$, $z_{\mathcal{F}}(t) = 0$, we have

$$\begin{aligned}\epsilon(t) &= 0 - A\{K(r)\cos(\omega t) + L(r)\} \\ &= -A\left\{\left[\frac{1}{\pi}\left(r\sqrt{1-r^2} - \arccos r\right) + 1\right]\cos(\omega t) \right. \\ &\quad \left. - \frac{1}{\pi}\left(\sqrt{1-r^2} - r\arccos r\right) - r\right\}.\end{aligned}$$

Appendix C
Proof of Proposition 1

Proof: For an AF form Matsuoka oscillator, using Taylor expansion at around $r = 0$, we have

$$L(r) \approx \frac{1}{2}r + \frac{1}{\pi}.$$

Let $\mathbf{B_X}$ denote the amplitude bias vector of $\mathbf{X}$, and apply this notation analogously to $\mathbf{B_Z}, \mathbf{B_U}, \mathbf{B_P}$. Let

$$\mathbf{A^e} = \begin{bmatrix} A_{x_1^e} & 0 & 0 & 0 \\ 0 & A_{x_2^e} & 0 & 0 \\ 0 & 0 & A_{x_3^e} & 0 \\ 0 & 0 & 0 & A_{x_4^e} \end{bmatrix}$$

$$\mathbf{A^f} = \begin{bmatrix} A_{x_1^f} & 0 & 0 & 0 \\ 0 & A_{x_2^f} & 0 & 0 \\ 0 & 0 & A_{x_3^f} & 0 \\ 0 & 0 & 0 & A_{x_4^f} \end{bmatrix}$$

where $A_{x_i^e}, A_{x_i^f}$ are the amplitude of x_i^e, x_i^f, respectively.

Since $\mathbf{U^e} = \mathbf{E} - \mathbf{U^f}$ with all elements being positive satisfies the *complementary* definition, we can assume x^e, x^f have similar amplitudes when the CPG system for the soft snake robot is mostly under forced-response oscillation and slithering gait, thus $A^e - A^f \approx 0$. Let $\mathbf{I} = [1,1,1,1]^T$, then we have

$$\begin{aligned}\mathbf{B_Z} &= \mathbf{A^e L(r^e)} - \mathbf{A^f L(r^f)} \\ &= \frac{1}{2}\left(\mathbf{A^e r^e} - \mathbf{A^f r^f}\right) + \frac{1}{\pi}\left(\mathbf{A^e I} - \mathbf{A^f I}\right) \\ &\approx \frac{1}{2}\left(\mathbf{A^e r^e} - \mathbf{A^f r^f}\right) \\ &= \frac{1}{2}\mathbf{B_X}. \end{aligned} \tag{C.1}$$

Taking the bias of (8) yields

$$\mathbf{B_X} = ((-b+a)\mathbf{E} - \mathbf{W})\mathbf{B_Z} + b\mathbf{B_P} + \mathbf{B_U}. \tag{C.2}$$

Plugging (C.1) back to (C.2) yields

$$((2+b-a)\mathbf{E} + \mathbf{W})\mathbf{B_Z} = b\mathbf{B_P} + \mathbf{B_U}.$$

Therefore, we have the relation among inputs' and outputs' biases in vector form as follows:

$$\mathbf{B_Z} = ((2+b-a)\mathbf{E} + \mathbf{W})^{-1}(b\mathbf{B_P} + \mathbf{B_U}).$$

Similarly, this conclusion for the MPF form Matsuoka oscillator can be proven by the same method. □

References

[1] E. W. Hawkes, L. H. Blumenschein, J. D. Greer, and A. M. Okamura, "A soft robot that navigates its environment through growth," *Sci. Robot.*, vol. 2, no. 8, 2017, Art. no. eaan3028.

[2] C. Majidi, "Soft robotics: A perspective–Current trends and prospects for the future," *Soft Robot.*, vol. 1, no. 1, pp. 5–11, 2014.

[3] H. Wang, R. Zhang, W. Chen, X. Wang, and R. Pfeifer, "A cable-driven soft robot surgical system for cardiothoracic endoscopic surgery: Preclinical tests in animals," *Surg. Endoscopy*, vol. 31, no. 8, pp. 3152–3158, 2017.

[4] F. Sanfilippo, J. Azpiazu, G. Marafioti, A. A. Transeth, Ø. Stavdahl, and P. Liljebäck, "A review on perception-driven obstacle-aided locomotion for snake robots," in *Proc. IEEE 14th Int. Conf. Control, Automat., Robot. Vis.*, 2016, pp. 1–7.

[5] T. Kano, R. Yoshizawa, and A. Ishiguro, "Tegotae-based decentralised control scheme for autonomous gait transition of snake-like robots," *Bioinspiration Biomimetics*, vol. 12, no. 4, Aug. 2017, Art. no. 046009.

[6] Y. Fukuoka, K. Otaka, R. Takeuchi, K. Shigemori, and K. Inoue, "Mechanical designs for field undulatory locomotion by a wheeled snake-like robot with decoupled neural oscillators," *IEEE Trans. Robot.*, vol. 39, no. 2, pp. 959–977, Apr. 2023.

[7] A. A. Transeth, P. Liljeback, and K. Y. Pettersen, "Snake robot obstacle aided locomotion: An experimental validation of a non-smooth modeling approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 2582–2589.

[8] A. A. Transeth, R. I. Leine, C. Glocker, K. Y. Pettersen, and P. Liljebäck, "Snake robot obstacle-aided locomotion: Modeling, simulations, and experiments," *IEEE Trans. Robot.*, vol. 24, no. 1, pp. 88–104, Feb. 2008.

[9] P. Liljeback, K. Y. Pettersen, Ø. Stavdahl, and J. T. Gravdahl, "Hybrid modelling and control of obstacle-aided snake robot locomotion," *IEEE Trans. Robot.*, vol. 26, no. 5, pp. 781–799, Oct. 2010.

[10] P. Liljeback, K. Y. Pettersen, Ø. Stavdahl, and J. T. Gravdahl, "Experimental investigation of obstacle-aided locomotion with a snake robot," *IEEE Trans. Robot.*, vol. 27, no. 4, pp. 792–800, Aug. 2011.

[11] I. Gravdahl, Øyvind Stavdahl, A. Koushan, J. Løwer, and K. Y. Pettersen, "Modeling for hybrid obstacle-aided locomotion (HOAL) of snake robots," in *IFAC-PapersOnLine 10th Vienna Int. Conf. Math. Modelling*, 2022, vol. 55, no. 20, pp. 247–252.

[12] T. Kano, T. Sato, R. Kobayashi, and A. Ishiguro, "Local reflexive mechanisms essential for snakes' scaffold-based locomotion," *Bioinspiration Biomimetics*, vol. 7, no. 4, 2012, Art. no. 046008.

[13] T. Kano and A. Ishiguro, "Obstacles are beneficial to me! scaffold-based locomotion of a snake-like robot using decentralized control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 3273–3278.

[14] T. Kano and A. Ishiguro, "Decoding decentralized control mechanism underlying adaptive and versatile locomotion of snakes," *Integrative Comp. Biol.*, vol. 60, no. 1, pp. 232–247, 2020.

[15] T. Kano, N. Matsui, and A. Ishiguro, "3D movement of snake robot driven by tegotae-based control," in *Biomimetic and Biohybrid Systems*, U. Martinez-Hernandez et al., Eds. Cham, Switzerland: Springer, 2019, pp. 346–350.

[16] R. Thandiackal et al., "Emergence of robust self-organized undulatory swimming based on local hydrodynamic force sensing," *Sci. Robot.*, vol. 6, no. 57, 2021, Art. no. eabf6354.

[17] S. Min, J. Won, S. Lee, J. Park, and J. Lee, "Softcon: Simulation and control of soft-bodied animals with biomimetic actuators," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 1–12, 2019.

[18] M. Hamidreza, B. Jacob, and H. D. L. Hu, "Snakes mimic earthworms: Propulsion using rectilinear travelling waves," *J. Roy. Soc. Interface*, vol. 10, 2013, Art. no. 20130188.

[19] B. C. Jayne, "What defines different modes of snake locomotion?," *Integrative Comp. Biol.*, vol. 60, no. 1, pp. 156–170, 2020.

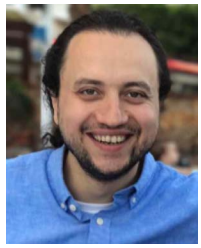
[20] H. Wang et al., "Design methodology for magnetic field-based soft tri-axis tactile sensors," *Sensors*, vol. 16, no. 9, 2016, Art. no. 1356.

[21] J. M. Crowe-Riddell, E. P. Snelling, A. P. Watson, A. K. Suh, J. C. Partridge, and K. L. Sanders, "The evolution of scale sensilla in the transition from land to sea in elapid snakes," *Open Biol.*, vol. 6, no. 6, 2016, Art. no. 160054.

[22] M. Luo et al., "Toward modular soft robotics: Proprioceptive curvature sensing and sliding-mode control of soft bidirectional bending modules," *Soft Robot.*, vol. 4, no. 2, pp. 117–125, 2017.
[23] R. Gasoto et al., "A validated physical model for real-time simulation of soft robotic snakes," in *Proc. 2019 Int. Conf. Robot. Automat.*, 2019, pp. 6272–6279.
[24] X. Liu, R. Gasoto, Z. Jiang, C. Onal, and J. Fu, "Learning to locomote with artificial neural-network and CPG-based control in a soft snake robot," in *Proc. 2020 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 7758–7765.
[25] X. Liu, C. D. Onal, and J. Fu, "Reinforcement learning of CPG-regulated locomotion controller for a soft snake robot," *IEEE Trans. Robot.*, vol. 39, no. 5, pp. 3382–3401, Oct. 2023.
[26] K. Matsuoka, "Sustained oscillations generated by mutually inhibiting neurons with adaptation," *Biol. Cybern.*, vol. 56, no. 5-6, pp. 367–376, 1985.
[27] K. Matsuoka, "Analysis of a neural oscillator," *Biol. Cybern.*, vol. 104, no. 4-5, pp. 297–304, 2011.
[28] Y. D. Sato, K. Nakada, and K. Matsuoka, "Singular perturbation approach with Matsuoka oscillator and synchronization phenomena," in *Artificial Neural Networks and Machine Learning–ICANN*, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Berlin, Heidelberg: Springer, 2011, pp. 269–276.
[29] H. Nyquist and S. Brand, "Measurement of phase distortion," *Bell Syst. Tech. J.*, vol. 9, no. 3, pp. 522–549, 1930.
[30] R. K. Naumann et al., "The reptilian brain," *Curr. Biol.*, vol. 25, no. 8, pp. R317–R321, 2015.
[31] Z. Fan, R. Su, W. Zhang, and Y. Yu, "Hybrid actor-critic reinforcement learning in parameterized action space," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 2279–2285.
[32] G. Li, A. Ijspeert, and M. Hayashibe, "AI-CPG: Adaptive imitated central pattern generators for bipedal locomotion learned through reinforced reflex neural networks," *IEEE Robot. Automat. Lett.*, vol. 9, no. 6, pp. 5190–5197, Jun. 2024.
[33] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.
[34] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *Int. J. Robot. Res.*, vol. 5, no. 1, pp. 90–98, 1986.
[35] M. G. Park, J. H. Jeon, and M. C. Lee, "Obstacle avoidance for mobile robots using artificial potential field approach with simulated annealing," in *IEEE Int. Symp. Ind. Electron. Proc.*, pp. 1530–1535, 2001, vol. 3.
[36] J. Dong, X. Zhang, and X. Jia, "Strategies of pursuit-evasion game based on improved potential field and differential game theory for mobile robots," in *Proc. Int. Conf. Instrum., Meas., Comput., Commun. Control*, 2012, pp. 1452–1456.
[37] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.
[38] K. Matsuoka, "Mechanisms of frequency and pattern control in the neural rhythm generators," *Biol. Cybern.*, vol. 56, no. 5–6, pp. 345–353, 1987.
[39] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 23–30.
[40] A. J. Ijspeert, J. Hallam, and D. Willshaw, "Evolving swimming controllers for a simulated lamprey with inspiration from neurobiology," *Adaptive Behav.*, vol. 7, no. 2, pp. 151–172, 1999.

Xuan Liu (Member, IEEE) received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015, and the M.S. degree from the University of Southern California, Los Angeles, CA, USA, in 2017, both in computer science, and the Ph.D. degree in robotics engineering from Worcester Polytechnic Institute, Worcester, MA, USA, in 2023.

He is currently a Postdoctoral Researcher with the Department of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China. His research interests include formal methods, stochastic control, reinforcement learning, and embodiment control in bioinspired soft robotics.



Cagdas D. Onal (Member, IEEE) received the B.Sc. and M.Sc. degrees from Mechatronics Engineering Program, Sabanci University, Istanbul, Turkey, in 2003 and 2005, respectively, and the Ph.D. degree in mechanical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2009.

He is currently the Dean's Associate Professor and Arseneault Faculty Fellow in robotics engineering with Worcester Polytechnic Institute (WPI), Worcester, MA, USA. He is the Director of WPI Soft Robotics Lab and leads the NSF Future of Robots in the Workplace Research and Development (FORW-RD) NRT Program with WPI. Before joining WPI, he was a Postdoctoral Associate with Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include soft robotics, origami-inspired printable robotics, alternative actuation/sensing mechanisms, bioinspiration, dynamic modeling, and control theory.

Dr. Onal was the recipient of an NSF CAREER award for origami-inspired soft robotic systems.

Jie Fu (Member, IEEE) received the B.S. and M.S. degrees in Electrical Engineering and Automation, from the School of Automation, from the Beijing Institute of Technology, Beijing, China, in 2007 and 2009, respectively, and the Ph.D. degree in mechanical engineering from the University of Delaware, Newark, Delaware, in 2013.

She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. Her research interests include stochastic control, reinforcement learning, algorithmic game theory, and formal methods.