

From the Lab to the Desert: Fast Prototyping and Learning of Robot Locomotion

Kevin Sebastian Luck^{*§}, Joseph Campbell^{*§}, Michael Andrew Jansen^{†§}, Daniel M. Aukes[‡] and Heni Ben Amor^{*}

^{*}School of Computing, Informatics, and Decision Systems Engineering,

[†]School of Life Sciences,

[‡]The Polytechnic School,

Arizona State University, Tempe, Arizona 85281

Email: {ksluck, jacampb1, majanse1, danaukes, hbenamor}@asu.edu

[§]Authors contributed equally

Abstract—We present a methodology for fast prototyping of morphologies and controllers for robot locomotion. Going beyond simulation-based approaches, we argue that the form and function of a robot, as well as their interplay with real-world environmental conditions are critical. Hence, fast design and learning cycles are necessary to adapt robot shape and behavior to their environment. To this end, we present a combination of laminate robot manufacturing and sample-efficient reinforcement learning. We leverage this methodology to conduct an extensive robot learning experiment. Inspired by locomotion in sea turtles, we design a low-cost crawling robot with variable, interchangeable fins. Learning is performed using both bio-inspired and original fin designs in an artificial indoor environment as well as a natural environment in the Arizona desert. The findings of this study show that static policies developed in the laboratory do not translate to effective locomotion strategies in natural environments. In contrast to that, sample-efficient reinforcement learning can help to rapidly accommodate changes in the environment or the robot.

I. INTRODUCTION

Robots are often tasked with operating in challenging environments that are difficult to model accurately. Search-and-rescue or space exploration tasks, for example, require robots to navigate through loose, granular media of varying density and unknown composition, such as sandy desert environments. A common approach is to use simulations in order to develop ideal locomotion strategies before deployment. Such an approach, however, requires prior knowledge about ground composition which may not be available or may fluctuate significantly. In addition, the sheer complexity of such terrain necessitates the use of approximations when simulating interactions between the robot and its environment. However, inaccuracies inherent to approximations can lead to substantial discrepancies between simulated and real-world performance. These limitations are especially troublesome as robot design is also guided by simulations in order to overcome time constraints and material deterioration associated with traditional physical testing.

In this work we argue that the design of effective locomotion strategies is dependent on the interplay between (a) the shape of the robot, (b) the behavioral and adaptive capabilities of the robot, and (c) the characteristics of the environment. In particular, adverse and dynamic terrains require a design pro-

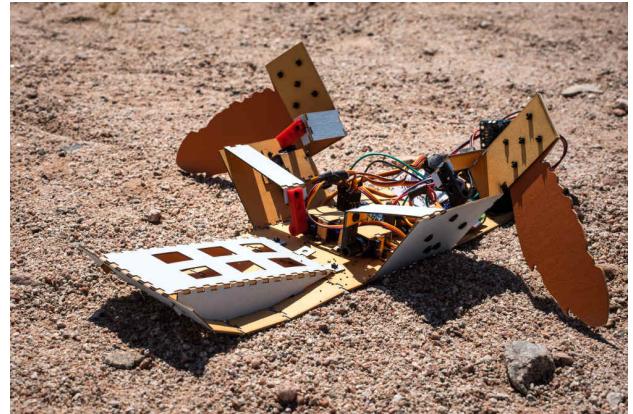


Fig. 1: A robot made from a multi-layer composite learns how to move across sand in the Arizona desert.

cess in which both *form* and *function* of a robot can be rapidly adapted to numerous environmental constraints. To this end, we introduce a novel methodology employing a combination of fast prototyping and manufacturing with sample-efficient reinforcement, thereby enabling practical, physical testing-based design.

First, we describe a manufacturing process in which *foldable robotic devices* (Fig. 1) are constructed out of a *single planar shape* consisting of multiple laminated layers of material. The overall production time of a robot using this manufacturing method is in the range of a few hours, i.e., from the first laser-cut to the deployment. As a result, changes to the robot shape can be performed by quickly iterating over several low-cost design cycles.

In addition to rapid design refinement and iteration, the synthesis of effective robot control policies is also of vital importance. Variations in terrain, the assembly process, motor properties, and other factors can heavily influence the optimal locomotion policy. Manual coding and adaptation of control policies is, therefore, a laborious and time-intensive process which may have to be repeated whenever the robot or terrain properties change, especially drift in actuation or changes in media granularity. Reinforcement learning (RL) methods [21] are a potential solution to this problem. Using a trial-and-error process, RL methods explore the policy space in search of

solutions that maximize the expected reward, e.g., the distance traveled while executing the policy. However, RL algorithms typically require thousands or hundreds of thousands of trials before they converge on a suitable policy [17]. Performing large numbers of experiments on a physical robot causes wear-and-tear on hardware, leads to drift in sensing and actuation, and may require extensive human involvement. This severely limits the number of learning experiments that can be performed within a reasonable amount of time.

A key element of our approach is a sample-efficient RL [11] method which is used for swift learning and adaptation whenever the changes occur to the robot or the environment. By leveraging the low-dimensional nature and periodicity of locomotion gaits, we can rapidly synthesize effective control policies that are best adapted to the current terrain. We show that using this method, the learning process quickly converges towards appropriate policy parameters. This translates to learning times of about 2-3 hours on the physical robot.

We leverage this methodology to conduct an extensive robot learning experiment. Inspired by locomotion in sea turtles, we design a low-cost crawler robot with variable, interchangeable fins. Learning is performed with different bio-inspired and original fin designs in both an indoor, artificial environment, as well as a natural environment in the Arizona desert. The findings of this experiment indicate that artificial environments consisting of poppy seeds, plastic granulates or other popular loose media substitutes may be a poor replacement for true environmental conditions. Hence, even policies that are not learned in simulation, but rather on granulate substitutes in the lab may not translate to reasonable locomotion skills in the real-world. In addition, our findings show that reinforcement learning is a crucial component in adapting and coping with variability in the environment, the robot, and the manufacturing process.

We thus demonstrate that the combination of a rapid prototyping process for robot design (form) and the fast learning of robot policies (function) enables environment-adaptive robot locomotion.

II. RELATED WORK

Prior studies have indicated that locomotion in granulate media is dependent upon successful compaction of the substrate, without causing fluidization [9, 15, 14]. Unfortunately, the dynamic response of granulate media during locomotion is difficult to predictively simulate or replicate [9, 15, 1]. In desert environments, this difficulty is compounded by the *heterogeneous* composition of the loose, sandy topsoil, making it nearly impossible to predict the effectiveness of any locomotor strategy *a priori* [9, 1]. In practice, the performance of robotic systems in heterogeneous granulate media, particularly in xeric habitats, must be evaluated *post-hoc* and iteratively improved (see for example [15]) through successive design refinements and adaptive learning of locomotion.

Finned animals, and sea turtles in particular, have achieved highly stable and efficient locomotion through heterogeneous granulate substrata [14, 9, 15]. Of the many animals capable

of effective locomotion in sand, we drew most heavily upon the sea turtle due to the simplicity and stability of its motion [25]. Unlike sand-swimming animals, like sand lizards [13], finned animals (such as sea turtles) require fewer degrees of freedom and actuated joints to achieve forward motion.

A robotic analogue to sea turtles, FlipperBot (FBot), was designed to provide a two-limbed approximation of sea turtle locomotion in an ongoing effort to characterize the motion of finned animals through sand [15]. Unlike other robotic devices inspired by turtles, FBot was designed for locomotion in granulate media and not for swimming [10, 26]. FBot features two degrees of freedom for each limb; however, the fins were configured such that they could be either fixed (relative to the arm) or free to rotate. The combined quasi-static motion of the limbs was similar to a “breast stroke”, dragging the body through the sand [15].

In general, the “bio-inspired robotics” approach [2] has proven fruitful for designing laboratory robots with new capabilities (new gaits, morphologies, control schemes), including rapid running [3, 18], slithering [22], flying [12], and “swimming” in sand [13]. In addition, using the biologically inspired robots as “physical models” of the organisms has revealed scientific insights into the principles that govern movement in biological systems, as well as new insights into low-dimensional dynamical systems (see for example [7] and references therein). Our work differs fundamentally from these works, not only in execution, but also in principle: we aim to generate optimal motion through bio-mimicry and learning, rather than learning how optima are generated in a biological system.

III. METHODOLOGY

In this section, we describe our methodology for fast robot prototyping and learning. We discuss a sample-efficient reinforcement learning method that enables fast learning of new locomotion skills. In combination with a laminate robot manufacturing process, our method allows for rapid iterations over both form and function of a robot. The main rationale behind this approach is that environmental conditions are often hard to reproduce outside of the natural application domain. Hence, the development cycle should be informed by experiences in the real application domain, e.g., on challenging terrain such as desert environments. Our approach facilitates this process and significantly reduces the underlying development time. Consequently, we will describe the methodology for prototyping form and function in more detail.

A. FORM: Laminate Robot Manufacturing

Laminate manufacturing can be used in order to construct affordable, light-weight robots. Laminate fabrication processes (known as SCM [24], PC-MEMS [20], popup-book MEMS [23], lamina-emergent mechanisms [6], etc.) permit rapid construction from planar sheets of material which are iteratively cut, aligned, stacked, and laminated to form a composite material.

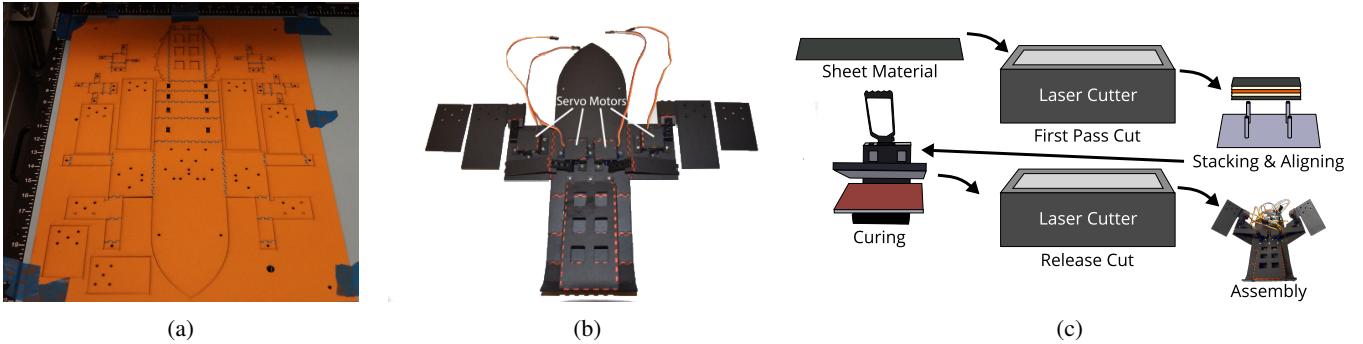


Fig. 2: Manufacturing of laminate robotic mechanisms: (a) The robot components are engraved using a laser cutter on planar sheets and later laminated. (b) The components are folded into a robotic structure. The motors, the control board, and the battery are added manually. (c) The full fabrication process.



Fig. 3: The laminates involved are constructed as a sandwich of five layers: poster-board, adhesive, polyester, adhesive, and poster-board.

The laminate mechanisms discussed in this paper were printed in five layers. As shown in Figure 3, two rigid layers of 1mm-thick poster-board are sandwiched around two adhesive layers of Drytac MHA heat-activated acrylic adhesive, which is itself sandwiched around a single layer of 50 μm -thick polyester film from McMaster-Carr. Flat sheets of each material are cut out on a laser cutter, then stacked and aligned using a jig with holes pre-cut in the first pass of the laser cut. Once aligned, these layers are fused together using a heated press in order to create a single laminate. The adhesive cures at around 85-104 degrees Celsius, and comes with a paper backing which allows it to be cut, aligned with the poster-board, and laminated. The paper backing is then removed, and the two adhesive-mounted poster-board layers are aligned with the center layer of polyester and laminated again. This laminate is returned to the laser, where a second release cut permits the device to be separated from surrounding scrap material and erected into a final three-dimensional configuration.

Laminate mechanisms resulting from this process are capable of a high degree of precision through bending of flexure-based hinges created through the selective removal of rigid material along desired bend axes. With fewer rolling contacts(bearings) than would typically be found in traditional robots, laminate mechanisms are ideal in sandy environments, where sand infiltration can shorten service life. Connections between layers can be established through adhesive layers, in addition to plastic rivets which permit quick attachment between laminates. Mounting holes permit attaching a variety of off-the-shelf components including motors, microcontrollers, and sensors. Rapid attachment/detachment is a highly desired feature for this platform, as different flipper designs can be tested using the same base platform. In all, this fabrication method permits *rapid iteration* during the

design phase, and *rapid re-configuration* for testing a variety of designs across a wide range of force and length scales, due to its compatibility with a wide range of materials. Fig. 2(a) depicts the basic planar sheets after cutting. Fig. 2(b) shows the individual components of the robot after they are detached from the sheets and folded into a structure. Fig. 2(c) is the full fabrication process. The whole manufacturing process of one robot takes up to 50 minutes while the 3D-printing process of four horns, which serve as connections between the motors and the paper, takes 58 minutes.

B. FUNCTION: Sample-Efficient Reinforcement Learning

In this section we discuss a sample-efficient RL method that converges on optimal locomotion policies within a small number of robot trials. Our approach leverages two key insights about human and animal locomotion. In particular, locomotion is (a) inherently low-dimensional and based on a small number of motor synergies [8], as well as (b) highly periodic in nature.

To implement these insights within a reinforcement learning framework, we build upon the Group Factor Policy Search (GrouPS) algorithm introduced by Luck et al. [11]. GrouPS jointly searches for a low-dimensional control policy as well as a projection matrix \mathbf{W} for embedding the results into a high-dimensional control space. It was previously shown [11] that the algorithm is able to uncover optimal policies after a few iterations with only hundreds of samples. Group Factor Policy Search models the joint actions as $\mathbf{a}_t^{(m)} = (\mathbf{W}^{(m)} \mathbf{Z} + \mathbf{M}^{(m)} + \mathbf{E}^{(m)}) \phi(\mathbf{s}, t)$ for each time step t of a trajectory and each m -th group of actions. The matrix \mathbf{W} represents the transformation matrix from the low dimensional to the high dimensional space (exploitation) and \mathbf{M} the parameters of the current mean policy. The entries of the matrices \mathbf{Z} and \mathbf{E} are Gaussian distributed with $z_{ij} \sim N(0, 1)$ for the latent values and $e_{ij} \sim N(0, \tau_m^{-1})$ for the isotropic exploration. The function $\phi(\mathbf{s}, t)$ consists of basis functions $\phi_i(\mathbf{s}, t)$ and depends in our experiments only of the time step t and not of the full state \mathbf{s} . In contrast to the work in [11], however, we incorporate periodicity constraints into the search process by focusing on periodic feature functions. We use periodic basis

functions over 20 time steps for the control signal, see Fig. 4. Given a point in time t we compute each control dimension a_i by

$$a_i = \sum_j (\tilde{w}_{ij} + m_{ij} + e_{ij}) \sin \left(\frac{t}{T} 720^\circ + \frac{j-1}{J} 360^\circ \right) \quad (1)$$

with $\tilde{w}_{ij} = \sum_k w_{ik} z_{kj}$ and J being the number of basis functions in $\phi(s, t)$.

GrouPS also takes prior information about potential groupings of joints into account when searching for an optimal transformation matrix \mathbf{W} . For our robotic device we used two groups: the first group consists of the two fin-joints and the second group of the two base-joints. Thus, we exploit the symmetry of the design. The number of dimensions of the manifold was set to three and the rank parameter, controlling the sparsity and structure of \mathbf{W} , to one. The outline of the algorithm can be found in Algorithm 1. Incorporating dimensionality reduction, periodicity, and information about the group structure yields a highly sample-efficient algorithm.

Input: Reward function $R(\cdot)$ and initializations of parameters. Choose number of latent dimension n and rank r . Set hyper-parameter and define groupings of actions.

```

while reward not converged do
  for  $h=1:H$  do # Sample H rollouts
    for  $t=1:T$  do
       $\mathbf{a}_t = \mathbf{WZ}\phi + \mathbf{M}\phi + \mathbf{E}\phi$ 
      with  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  and  $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \tilde{\tau})$ ,
      where  $\tilde{\tau}^{(m)} = \tilde{\tau}_m^{-1} \mathbf{I}$ 
      Execute action  $\mathbf{a}_t$ 
    Observe and store reward  $R(\tau)$ 

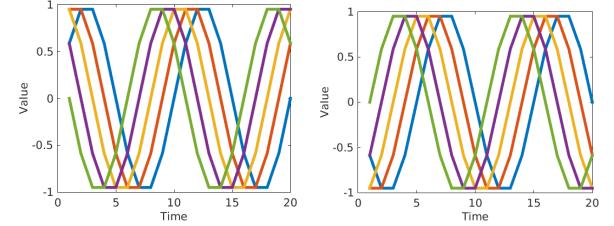
    Initialization of q-distribution
    while not converged do
      Update  $q(\mathbf{M})$ ,  $q(\mathbf{W})$ ,  $q(\tilde{\mathbf{Z}})$ ,  $q(\alpha)$  and  $q(\tilde{\tau})$ 
       $\mathbf{M} = \mathbb{E}_{q(\mathbf{M})}[\mathbf{M}]$ 
       $\mathbf{W} = \mathbb{E}_{q(\mathbf{W})}[\mathbf{W}]$ 
       $\alpha = \mathbb{E}_{q(\alpha)}[\alpha]$ 
       $\tilde{\tau} = \mathbb{E}_{q(\tilde{\tau})}[\tilde{\tau}]$ 
```

Result: Linear weights \mathbf{M} for the feature vector ϕ , representing the final policy. The columns of \mathbf{W} represent the factors of the latent space.

Algorithm 1: Outline of the Group Factor Policy Search (GrouPS) algorithm as presented in [11].

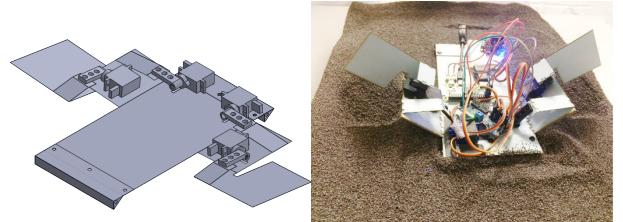
IV. A FOLDABLE ROBOTIC SEA TURTLE

With the general methodology established, this section introduces the design of the robotic device used in this research. As discussed in Sec. II, our design takes inspiration from sea turtles. By necessity, the design also conforms to



(a) Basis functions $\phi(t)_{1:5}$. (b) Basis functions $\phi(t)_{6:10}$.

Fig. 4: The sinusoidal basis functions $\phi(t)$ used for learning in this paper. Each basis function is based on a sine curve and shifted in time. The final policy is based on a linear combination of these functions.



(a) (b)

Fig. 5: The initial “flat body” design of the robot. The front of the robot buried into the sand during motion. The body was later curved

the constraints of the laminate fabrication techniques being employed – primarily that it is composed of a single planar layer. The salient aspects of Chelonioid morphology integrated into our design are described below.

A. Biological Inspiration

The design of our laminate device was primarily inspired by the anatomy and locomotion of sea turtles. We chose to focus on the terrestrial locomotion of *adult* sea turtles, rather than juveniles or hatchlings, emphasizing the greater load-bearing capacity and stability of their anatomy and behavior. There are seven recognized species in Cheloniodea in six genera [19]. In spite of considerable inter-specific differences in morphology, all sea turtles use the same set of anatomical features to generate motion. Specifically, adult sea turtles support themselves on the radial edge of the forelimbs to (1) elevate the body (thus reducing or eliminating drag) and (2) generate forward motion [25]. This unique behavior allows these large and exceedingly heavy animals (up to 915 kg in *Dermochelys coriacea* (Vandelli, 1761)¹) to move in a stable and effective manner through granular media [5].

B. Robot Design

Focusing on the turtle’s forelimb for generating locomotion, the robot form and structure was determined within an iterative

¹Pursuant to the International Code of Zoological Nomenclature, the first mention of any specific epithet will include the full genus and species names as a binomen (two part name) followed by the author and date of publication of the name. This is *not* an in-line reference; it is a part of the name itself and refers to a particular species-concept as indicated in the description of the species by that author.

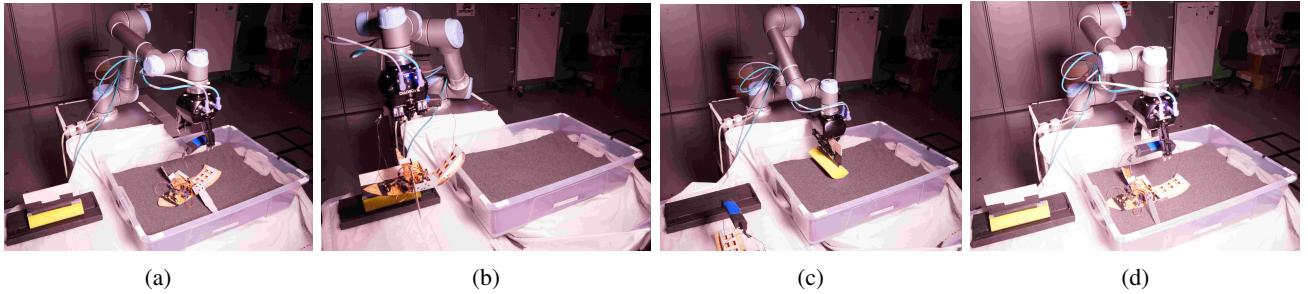


Fig. 6: Sequence of actions the robotic arm executes in each learning cycle: (a) First the robot under test is located in the testbed, grasped and then (b) subsequently moved into a resting position. The robotic arm proceeds to (c) smooth the testbed with a tool. Finally, the robot under test is (d) put into its initial position and the next trajectory is executed.

design cycle. In all designs, the body was suitably broad to prevent sinking during forward motion, and remained in contact with the ground at rest. This provides stability while removing the need for the limbs to bear the weight of the body at all times. A major benefit of this configuration is that only the two forelimbs are needed to generate forward thrust. Transmission of load occurs primarily under tension (as in muscles), to accommodate the laminate material and to provide dampening to reduce joint wear. The limbs have 2 rotational degrees of freedom, such that the fins move down and back into the substrate, while the body moves up and forward. This two degree of freedom arm was sufficient to replicate the circular motion of the fins (and particularly of the radial edge) observed in sea turtles (see [25]).

Initial experiments attempted on early prototypes revealed a critical design flaw: the anterior end was prone to “plowing” into the substrate (see Fig. 5). This limitation was solved by mimicking two features of turtle anatomy. First, the apical portion of the design is shaped to elevate the body above sand, with an upturned apex, similar to upturned intergular and gular scales of the anterior sea turtle plastron (see [19]). Second, the back end of the body was tapered to reduce drag (as compared to a rectangular end of equal length).

In the final design cycle, we also sought to mimic and explore the morphology of the fins. Extant sea turtle species exhibit a variety of fin shapes and include irregularities seen on the outer edges, such as scales and claws. These features are known to be used for terrestrial locomotion by articulating with the surface directly (rather than being buried in the substrate) [14, 4]. In order to understand how fin shape affects locomotion performance, we designed four pairs of fins: two generated from outlines of sea turtle fins which include all irregularities (*Caretta caretta* (Linnaeus, 1758) and *Natator depressus* (Garman, 1880), from [19]), and two based on artificial shapes with no irregularities, as shown in Fig. 7. All of these were attached to the main body at a position equivalent to the anatomical location of the humeroradial joint (part of the elbow in the fin), and scaled to the width of the body. The arms of the robot were designed such that the fins can be interchanged at will, allowing for easy comparison of fin performance.

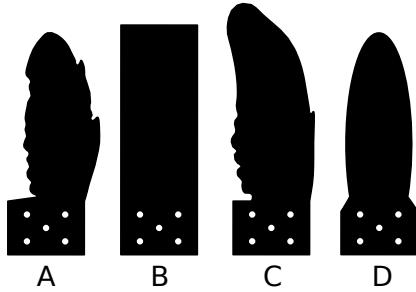


Fig. 7: The four different design of fins used for the presented robotic device. Designs A and C are accurate reproductions of the actual shape of sea turtle fins, namely *Caretta caretta* (A) and *Natator depressus* (C). Designs B and D are simple rectangular and ellipsoid shapes.

V. EXPERIMENTS

In this section, we focus on evaluating the locomotion performance of the prototypes generated with our laminate fabrication process. In particular, the robustness to variations stemming from the terrain and manufacturing process, and the sensitivity to changes in the physical fin shape.

More formally, there are three hypotheses that we experimentally evaluate:

H1 *Group Factor Policy Search is able to find an improved locomotion policy – with respect to distance traveled forward – in a limited number of trials, despite the presence of variations in the rapidly prototyped robotic device and the environment.*

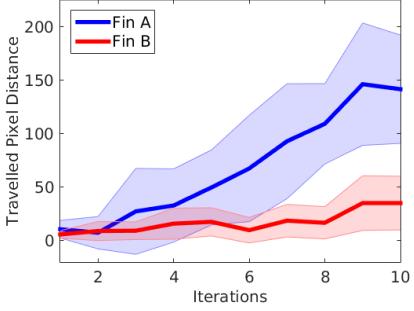
H2 *The shape of the fin influences the performance of the locomotion policy.*

H3 *The locomotion policies learned in the natural environment out-perform those learned in the artificial environment, when executed in the natural environment.*

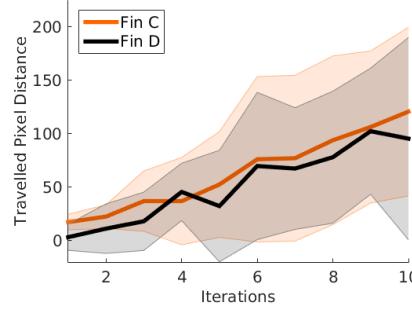
These hypotheses are tested through the following experiments.

A. Evaluation of Fin Designs

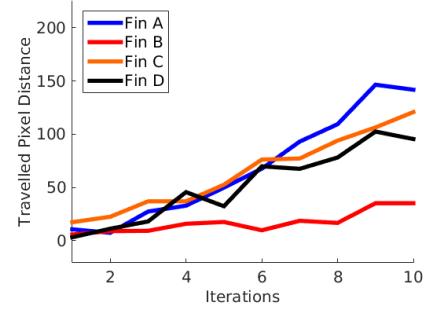
This experiment is designed to evaluate the effectiveness of locomotion policies generated for the four types of fins described in Sec. IV. Five independent learning sessions were conducted for each fin, consisting of 10 policy search iterations



(a) Comparison between fin A and fin B.



(b) Comparison between fin C and fin D.



(c) Comparison between all fins.

Fig. 8: Comparison between the learning for different fin designs. Each experiment was performed five times and mean/standard deviations were computed. The learning process was performed on poppy seeds.

each for a total of 1050 policy executions per fin. The experiment was performed in an indoor, artificial environment utilizing poppy seeds (similar to [16]) as a granulate material substitute for sand – they are less abrasive and increase the longevity of prototypes. Human involvement, and thus randomness, was minimized during the learning process by employing an articulated robotic arm (UR5). This arm was responsible for placing the robot under test in the artificial environment prior to each policy execution, then subsequently removing it and resetting the environment with a leveling tool. This sequence of actions is depicted in Fig. 6.

The policy search reward was automatically computed by measuring the distance (in pixel values) that a target affixed to the robot traveled with a standard 2D high-definition webcam. This was computed from still frames captured before and after policy execution. After learning, the mean iteration policies were manually executed and measured in order to produce metric distance rewards for comparison.

B. Policy Learning in a Desert Environment

The second experiment was designed to test how well policies transfer between environments, and whether policies learned in-situ are more effective than policies learned in other environments. Over the course of two days, the policies generated for each fin in the artificial environment from the first experiment were executed in a desert environment in the Tonto National Forest of Arizona in order to measure their distance rewards. We opted to create a flattened test bed as shown in Fig. 9, rather than using untouched ground, in order to reduce locomotion bias due to inclines, rocks, and plants.

Furthermore, two additional learning sessions were conducted for fins A and C in the same test bed in order to provide a point of comparison. To maintain consistency with the first experiment, learning was performed with 10 Policy Search iterations and reward values were measured via camera. Manually measured distance values for each mean iteration policy were obtained after learning. A video of the learning process and supplementary material can be found on <http://www.c-turtle.org>.



Fig. 9: The testbed in the Arizona desert used for evaluating and learning policies in a real environment. The surface of the testbed is flattened in order to increase comparability between the values measured for each policy.

VI. RESULTS

The rewards achieved by policies learned on poppy seeds are presented in Figure 8 with their mean and standard deviation over the conducted experiments. Figure 8 (a) compares the biologically inspired fin A (*C. caretta*) and the simple rectangular shape. The second biologically inspired fin C (*N. depressus*) and the artificial oval fin can be found in Figure 8 (b), both with a similar performance. The mean values of the learned policies are given in Figure 8 (c). The reward in these plots is given as pixel distances, as recorded by the camera, covered by the robot with its movement, which means that fin A (*C. caretta*) outperforms all other fin designs. On the opposite, the rectangular shaped fin shows the worst performance. This can also be seen in Figure 10 which compares the mean and standard deviation of achieved rewards in the last iteration of the learning process between the four different fin designs.

Two different fin designs, A (*C. caretta*) and C (*N. depressus*), were selected for the comparison between policies learned on poppy seeds and policies learned in a natural environment. Figure 11 (a) and (b) show the covered distances in centimeters for policies learned and executed on poppy seeds as well as executed in the desert for each iteration. The third policy for each fin was learned and evaluated in the desert. It can be seen that the policy learned in the natural environment outperforms the policies learned on the substitute in the laboratory environment.

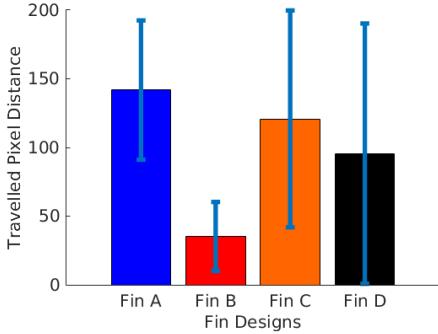


Fig. 10: The mean and standard deviation of policies for each fin design in the last iteration of the learning process. The rewards represent the distance the robot moved forward.

A series of images from the executions of the policies are shown in Figure 12. The pictures show the final position after execution of policies learned in iteration one, four, six, eight and ten. The images in Figure 12 (b) and (c) show the difference in covered length between policies learned on poppy seeds and the policies learned in the natural environment.

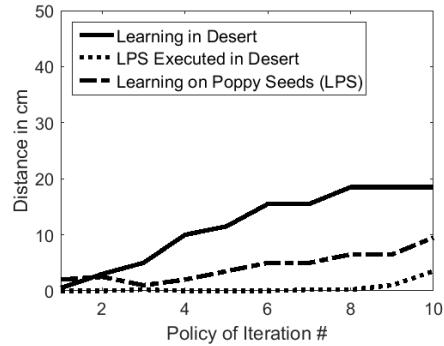
VII. DISCUSSION

The results shown in Fig. 8 and Fig. 11 indicate that for every fin that underwent learning, in both artificial and natural environments, the final locomotion policy shows some degree of improvement with regard to distance traveled by the robot after only 10 iterations. This supports hypothesis H1 which postulated that Group Factor Policy Search would find an improved locomotion policy in a limited number of trials, despite variations in the environment and fin shape.

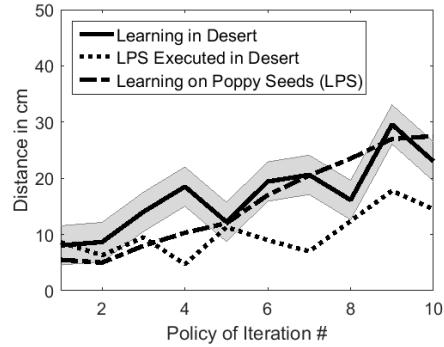
However, the results also indicate that some fins clearly performed better than others. For example, fin B only achieved a mean pixel reward of 35.2 in the artificial environment, while fin A saw a mean pixel reward of 141.8, as shown in Fig. 8a. This supports H2, which hypothesized that the physical shape of the fin affects locomotion performance.

It is interesting to note that the biologically inspired fins (A and C) out-performed the artificial fins (B and D) on average. At least part of this may be due to the intersection of the fin and the ground when they make contact at an angle, as is the case in our robotic design. The biological fins have a curved design which increases the surface area that is in contact with granulate media when compared to the artificial fins while the overall surface areas of artificial fins and biologically inspired fins are comparable to each other. Furthermore, fin B exhibited significant deformation when in contact with the ground which likely reduced its effectiveness in producing forward motion.

The results shown in Fig. 11 support hypothesis H3, in that policies learned in the natural environment outperform the policies that were learned in the artificial environment. We reason that part of this discrepancy is due to the composition of the granulate material. The poppy seeds used in the artificial environment have an average density of 0.54 g/ml with a – qualitatively speaking – homogeneous seed size, while the sand grains in the desert have an average density of 1.46



(a) Comparison between policies learned for fin design C. The algorithm was initialized with the same random number generator for learning.



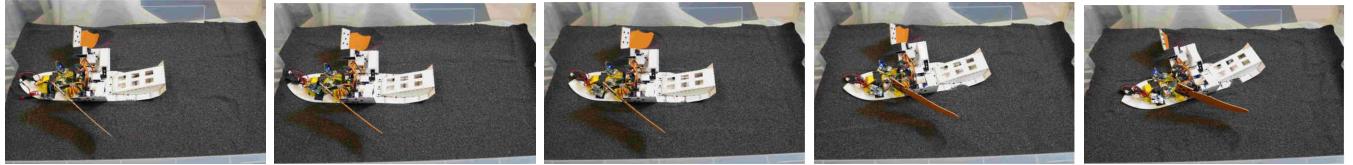
(b) Comparison between policies learned for fin design A. The algorithm was initialized with the same random number generator for learning. Due to a technical issue only pixel distances were recorded for learning in the desert. For comparability those pixel distances were transformed into centimeters but are attached with a variance of about 3.5cm.

Fig. 11: Comparison between policies learned on poppy seeds and executed on poppy seeds (LPS), learned on poppy seeds and executed in a desert environment, and policies learned and executed in a desert environment.

g/ml and a heterogeneous grain size. These results indicate that artificial environments consisting of popular granulate substitutes, such as poppy seeds, may not yield performance comparable to the real-world environments that they are mimicking. Thus, it is not only simulations that can yield performance discrepancies, but also physical environments.

Additionally, we observed that the composition of the natural environment itself fluctuated over time. For instance, we measured a difference in the moisture content of the sand of nearly 82% between the two days in which we performed experiments: 1.59% and 0.87% by weight. These factors may serve to make the target environment difficult to emulate, and suggest that not only are discrepancies possible between simulated environments, artificial environments, and actual environments, but also between the same actual environment over time. We suspect that lifelong learning might be a possible solution to this problem.

Yet another interesting observation can be made from the gaits shown in Fig. 13. The cycle produced by the fin during a more effective policy extends deeper and further than that



(a) Executions of policies learned on poppy seeds. The start position of the robot was on the wall of the testbed on the left side.



(b) Executions of policies learned on poppy seeds and executed in a real desert environment. The white line shows the start position of the robot.



(c) Executions of policies learned in a desert environment. The white line shows the start position of the robot.

Fig. 12: Executions of learned policies on poppy seeds and in a real desert environment. Row (a) shows the execution of the policies learned on poppy seeds which are also executed in a real desert environment in (b). Finally, (c) shows the policies learned and executed in the desert. For both learning experiments the same initial values and random number generators were used. The images show the executions of trajectories after 1, 4, 6, 8 and 10 iterations.

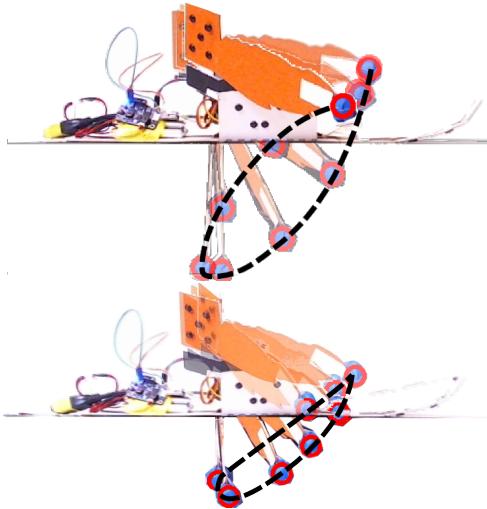


Fig. 13: Top: the gait produced by the right fin after iteration 10 with fin A. Bottom: the gait produced by the right fin of the robot after iteration 3.

produced during a less effective policy. Intuitively, we can reason that this more effective policy pushes against a larger volume of sand, generating more force for forward motion.

VIII. CONCLUSION

In this paper, we presented a methodology for rapid prototyping of robotic structures for terrestrial locomotion. A combination of laminate robot manufacturing and sample-efficient reinforcement learning enables re-configuration and adaptation of both form and function to best fit environmental constraints.

In turn, this approach decreases the amount of time for the development-production-learning-deployment cycle. With the presented techniques, it is possible to construct a robot out of raw material and learn a controller for locomotion in under a day. We designed a bio-inspired robotic device using the new methodology and, consequently, conducted an extensive robot learning study which involved several thousand executions. The experiment was performed with different sets of fins, both inside the lab, as well as in the desert of Arizona. Our results indicate the approach is well-suited for fast adaptation to new ground.

The results also show that granulates which are commonly used as a replacement for sand in robotics laboratories may not be an effective replacement. More specifically, the efficiency of robot control policies learned on such granulates in the laboratory were not as effective when deployed outside. A variety of factors such as variability in actuation, energy supply, the manufacturing process, or the terrain may contribute to this phenomenon. Consequently, learning and adaptation is of crucial importance. The discussed sample-efficient reinforcement learning algorithm enabled robots to quickly adapt an existing policy or learn a new one. Learning time was typically in the range of 2 – 3 hours. The results also show that biological inspiration in the fin design can lead to significant advantages in the resulting policies, even when learning was employed.

For future work we aim to investigate life-long learning approaches that do not separate between a training and a deployment phase. Using an accelerometer, the robot could continuously calculate rewards and update the control policy.

REFERENCES

- [1] Hesam Askari and Ken Kamrin. Intrusion rheology in grains and other flowable materials. *Nature Materials*, 15(12):1274–1279, 2016.
- [2] Bharat Bhushan. Biomimetics: lessons from nature—an overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1893):1445–1486, 2009.
- [3] Jonathan E Clark, Jorge G Cham, Sean A Bailey, Edward M Froehlich, Pratik K Nahata, Robert J Full, and Mark R Cutkosky. Biomimetic design and fabrication of a hexapedal running robot. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 4, pages 3643–3649, 2001.
- [4] C Kenneth Dodd Jr. Synopsis of the biological data on the loggerhead sea turtle caretta caretta (linnaeus 1758). Technical report, DTIC Document, 1988.
- [5] Karen L Eckert and Chris Luginbuhl. Death of a giant. *Marine Turtle Newsletter*, 43:2–3, 1988.
- [6] Paul S. Gollnick, Spencer P. Magleby, and Larry L. Howell. An Introduction to Multilayer Lamina Emergent Mechanisms. *Journal of Mechanical Design*, 133(8):081006, 2011. ISSN 10500472.
- [7] Philip Holmes, Robert J Full, Dan Koditschek, and John Guckenheimer. The dynamics of legged locomotion: Models, analyses, and challenges. *Siam Review*, 48(2):207–304, 2006.
- [8] Nedialko Krouchev, John F. Kalaska, and Trevor Drew. Sequential activation of muscle synergies during locomotion in the intact cat as revealed by cluster analysis and direct decomposition. *Journal of Neurophysiology*, 96(4):1991–2010, 2006. ISSN 0022-3077.
- [9] Chen Li, Tingnan Zhang, and Daniel I. Goldman. A terradynamics of legged locomotion on granular media. *Science*, 339:1408–1412, 2013.
- [10] Kin-Huat Low, Chunlin Zhou, TW Ong, and Junzhi Yu. Modular design and initial gait study of an amphibian robotic turtle. In *Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on*, pages 535–540. IEEE, 2007.
- [11] Kevin Sebastian Luck, Joni Pajarinen, Erik Berger, Ville Kytki, and Heni Ben Amor. Sparse latent space policy search. In *AAAI*, pages 1911–1918, 2016.
- [12] Kevin Y Ma, Pakpong Chirarattananon, Sawyer B Fuller, and Robert J Wood. Controlled flight of a biologically inspired, insect-scale robot. *Science*, 340(6132):603–607, 2013.
- [13] Ryan D Maladen, Yang Ding, Chen Li, and Daniel I Goldman. Undulatory swimming in sand: subsurface locomotion of the sandfish lizard. *science*, 325(5938):314–318, 2009.
- [14] Nicole Mazouchova, Nick Gravish, Andrei Savu, and Daniel I Goldman. Utilization of granular solidification during terrestrial locomotion of hatchling sea turtles. *Biology Letters*, 6:398–401, 2010.
- [15] Nicole Mazouchova, Paul B Umbanhowar, and Daniel I Goldman. Flipper-driven terrestrial locomotion of a sea turtle-inspired robot. *Bioinspiration and Biomimetics*, 8(2):026007, 2013.
- [16] Nicole Mazouchova, Paul B Umbanhowar, and Daniel I Goldman. Flipper-driven terrestrial locomotion of a sea turtle-inspired robot. *Bioinspiration & biomimetics*, 8(2):026007, 2013.
- [17] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 02 2015.
- [18] Robert Playter, Martin Buehler, and Marc Raibert. Big-dog. In Douglas W. Gage Grant R. Gerhart, Charles M. Shoemaker, editor, *Unmanned Ground Vehicle Technology VIII*, volume 6230 of *Proceedings of SPIE*, pages 62302O1–62302O6, 2006.
- [19] Peter Pritchard and Jeanne Mortimer. Taxonomy, external morphology, and species identification. *Research and management techniques for the conservation of sea turtles*, 4:21, 1999.
- [20] Pratheevo S Sreetharan, John P Whitney, Mark D Strauss, and Robert J Wood. Monolithic fabrication of millimeter-scale machines. *Journal of Micromechanics and Microengineering*, 22(5):055027, may 2012. ISSN 0960-1317.
- [21] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998. ISBN 0262193981.
- [22] Matthew Tesch, Kevin Lipkin, Isaac Brown, Ross Hatton, Aaron Peck, Justine Rembisz, and Howie Choset. Parameterized and scripted gaits for modular snake robots. *Advanced Robotics*, 23(9):1131–1158, 2009.
- [23] John P Whitney, Pratheevo S Sreetharan, Kevin Y Ma, and Robert J Wood. Pop-up book MEMS. *Journal of Micromechanics and Microengineering*, 21(11):115021, nov 2011. ISSN 0960-1317.
- [24] Robert J Wood, Srinath Avadhanula, Ranjana Sahai, Erik Steltz, and Ronald S Fearing. Microrobot Design Using Fiber Reinforced Composites. *Journal of Mechanical Design*, 130(5):052304, 2008. ISSN 10500472.
- [25] Jeanette Wyneken. Sea turtle locomotion: Mechanics, behavior, and energetics. In Peter L Lutz, editor, *The Biology of Sea Turtles*, pages 168–198. CRC Press, 1997.
- [26] Guocai Yao, Jianhong Liang, Tianmiao Wang, Xingbang Yang, Qi Shen, Yucheng Zhang, Hailiang Wu, and Weicheng Tian. Development of a turtle-like underwater vehicle using central pattern generator. In *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*, pages 44–49. IEEE, 2013.