

Morpho-evolution with learning using a controller archive as an inheritance mechanism

Léni K. Le Goff, Edgar Buchanan, Emma Hart, Agoston E. Eiben, Wei Li, Matteo De Carlo, Alan F. Winfield, Matthew F. Hale, Robert Woolley, Mike Angus, Jon Timmis, Andy M. Tyrrell

Abstract—Most work in evolutionary robotics centres on evolving a controller for a fixed body-plan. However, previous studies suggest that simultaneously evolving both controller and body-plan could open up many interesting possibilities. However, the joint optimisation of body-plan and control via evolutionary processes can be challenging in rich morphological spaces. This is because offspring can have body-plans that are very different from either of their parents, leading to a potential mismatch between the structure of an inherited neural controller and the new body. To address this, we propose a framework that combines an evolutionary algorithm to generate body-plans and a learning algorithm to optimise the parameters of a neural controller. The topology of this controller is created once the body-plan of each offspring has been generated. The key novelty of the approach is to add an external archive for storing learned controllers that map to explicit ‘types’ of robots (where this is defined with respect to the features of the body-plan). By initiating learning from a controller with an appropriate structure inherited from the archive, rather than from a randomly initialised one, we show that both the speed and magnitude of learning increases over time when compared to an approach that starts from scratch, using two tasks and three environments. The framework also provides new insights into the complex interactions between evolution and learning.

Index Terms—Evolutionary robotics, Embodied Intelligence

I. INTRODUCTION

The idea of embodied intelligence — describing the design and behaviours of physical objects situated in the real-world — was first introduced by Brooks in 1991 [1]. Pfeifer and Bongard’s seminal text “How the body shapes the way we think” [2] expanded on the idea that intelligent control is not only dependent on the brain but at the same time both constrained and enabled by the body. Increasingly, artificial evolution approaches have been used in robotics to jointly optimise both the body-plan and controller of a robot to accomplish a desired task. This has the potential advantage of allowing evolution to discover the appropriate balance between morphological and brain complexity and functionality.

However, much of this work has taken place in restricted morphological spaces, for example using regular shaped modules to construct body-plans, in which each module can be

individually actuated [3], [4]. In richer spaces which can give rise to complex and irregular robot skeletons with multiple forms of sensing and actuation (e.g. joints and/or wheels), then more complex controllers that link multiple sensors and actuators are required. In addition, the evolutionary process becomes more challenging: reproduction between two morphologically distinct parents might result in a viable body-plan, but a directly inherited controller is at best unlikely to provide adequate control and, at worst, will not work at all because inputs and outputs do not correspond to the new body-plan.

One approach to address this is to evolve a morphology-independent control mechanism, for example using a compositional pattern producing network (CPPN) [5] to generate a controller, thereby enabling direct inheritance of the generator [6]. However, evolutionary algorithms (EA) using generative methods tend to need more generations to converge than EAs using fixed-size genomes. An alternative is to add a learning cycle into the evolutionary loop [7], [8]. This can either improve an inherited controller over an individual’s lifetime – when the inherited controller has an appropriate structure – or learn a new controller from scratch otherwise. Here, we follow the latter approach and propose a novel framework for combining evolution and learning that is capable of joint optimisation of body and control of robots in a complex morphological space when using controller encodings that do not permit direct inheritance, i.e. when the topology of a child controller does not match the inherited body.

The framework contains a morpho-evolutionary algorithm (MEA) to optimise the body-plan and a learning algorithm to optimise the parameters of the controller. Two optimisation processes are nested; for each body-plan produced with the MEA, the learning process is invoked to optimise its controller. The key novelty of the approach is the addition of an external *controller archive*. This multi-dimensional archive stores the best-found controller for a given ‘type’ of robots, where *type* is defined by a vector describing the robot’s morphological features (e.g. number of wheels, number of sensors of form A, number of sensors of form B, etc.). If a body-plan is produced that is of the same type as a controller already stored in the archive, the learning process is initiated with this controller, otherwise, it starts from scratch. The archive is updated over generations as better controllers are found. Essentially the archive can be viewed as a form of inheritance, storing successful controllers per robot type that can be used to bootstrap *learning* in future generations. Hence the framework is named MELAI: *morpho-evolution with learning using archive inheritance*. Specifically, for the MEA, we use the

L. K. Le Goff and E. Hart are with the School of Computing, Edinburgh Napier University, Scotland, UK, contact l.legoff2@napier.ac.uk

E. Buchanan, W. Li, R. Woolley, M. Angus, A. M. Tyrrell are with the Department of Electronic Engineering, University of York, England, UK

M. F. Hale and A. F. Winfield are with Bristol Robotics Laboratory, University of the West of England, UK

J. Timmis is with School of Computer Science, University of Sunderland, England, UK

M. De Carlo and A. E. Eiben are with Department of Computer Science, Vrije Universiteit Amsterdam, NL

matrix-based CPPN morpho-evolution (MCME) introduced in previous work [9] to evolve body-plans. The learning algorithm used is a novelty-driven evolution-strategy, that uses an increasing population size (NIP-ES), and was also introduced in previous work [10]. It learns the weights of a controller specified by an Elman network that has a topology matching the generated child body-plan.

The contributions of the method are two-fold: (1) it offers a novel approach for the joint optimisation of both body-plans and controllers of robots, that integrates evolution and learning – uniquely, it uses a morpho-evolutionary algorithm for the former and an evolution-strategy for the latter; (2) it proposes the use of an external archive as an efficient mechanism for transferring control knowledge from parents to offspring in situations where offspring are morphologically distinct from their parents.

We show the benefits of using an archive (which interacts only with the learning process) in terms of increasing the efficiency of the approach compared to methods that learn from scratch and provide new insights into the interplay of evolutionary and learning processes. Moreover, our results show the emergence of different types of robots for different tasks.

The rest of the paper is organized as follows: section II analyses previous studies related to the joint optimisation of robot body-plans/controllers. In section III MELAI is explained in detail. The experimental protocol is described in section IV and the results are presented in section V. Finally, sections VI and VII discuss the results and conclude the paper.

II. RELATED WORK

Among the numerous studies in the field of evolutionary robotics, the majority address either the evolution of the body-plan alone *or* the evolution of the controller alone. This section focuses on literature that describes methods for the *joint* optimisation of body-plans and controllers. We focus attention on approaches that permit the evolution of offspring that require a controller topology that is different to either parent. That is, methods that evolve changes to morphology that do not impact the topology of controllers (i.e. the number of inputs and outputs of a neural controller) are out of scope. For example, this excludes work in which morphological change is restricted to repositioning sensors [11] or altering the length, weight and size of joints, e.g. [12], [13], [14], [15]. We first discuss methods that create controllers that are directly correlated to a specific type of body-plan, followed by morphology-independent methods, i.e. those that are capable of *generating* a controller for any given body-plan.

A naïve approach to avoiding a potential mismatch between a controller and body-plan is to evolve only the body-plan and then learn a new controller with the correct topology from scratch for each child body-plan. The work of Gupta *et al.* [16] follows this approach by using evolution for the body-plans and reinforcement learning to optimize the controllers for simulated robots composed from articulated 3D rigid parts connected via motor actuated hinge joints. Learning starts from a randomly initialised controller for each body-plan,

and uses a distributed implementation across multiple CPU to minimise computational cost. Liao *et al.* [17] also proposes a similar nested optimisation process, with the aim of finding the best morphology for a walker micro-robot. Here, Bayesian optimisation is used to learn the controller. However, despite Bayesian optimisation being known to be sample efficient, it only works well for small parameter spaces [10].

Instead of learning from scratch, an alternative approach is to use a morphology-independent control generation mechanism that can generate control parameters for any given body-plan. For example, Cheney *et al.* [18], [19] evolves soft-robots built from voxels, in which each voxel has a parameterised local controller. Both body-plans and controller parameters are outputs of two separate compositional pattern producing networks (CPPN) [5], both of which are evolved via the well-known neuro-evolution with augmenting topology (NEAT) algorithm [20]. However, due to the distributed nature of the controller, the variety of possible behaviours for the same body-plan is limited. Sims [21] also used a decentralised form of control, proposing a genotype that contains a nested graph; the graph specifies morphological nodes describing the robot shape, each of which contains another graph specifying the neural circuitry for that node. More recent work has achieved a similar effect with the use of Lindenmayer-Systems (L-systems) instead of a graph-representation, e.g. [22], [23], [24].

In the later work [24], an additional learning mechanism was applied to improve the inherited brains of newborn robots: the authors showed that learning not only influences the morphology of the resulting robots but also that the capacity to learn increases over generations. Jelisivcic *et al.* [25] also employed a learning mechanism. Their genome carried a pool of CPPNs used to specify the weights of a controller generated to match the child body-plan. In contrast to the work of Cheney *et al.* which encoded a single CPPN that undergoes evolution, here a child inherits a subset of CPPNs from each parent. A learning algorithm (HyperNEAT) is then applied to the inherited pool to evolve a new pool. The process is therefore Lamarckian.

To summarise, in the context of joint optimisation of body and controller, on the one hand, the literature has shown that using generative encodings (with and without additional learning) can mitigate the issues arising regarding inheritance of controllers that might not be applicable to a new child body-plan. However, these methods often require many evaluations to converge [6] and add additional hyper-parameters which may be difficult to optimise. On the other hand, neural controller encodings which are explicitly tied to a body-plan can be rapidly optimised as they only require weight rather than topological optimisation. Although they often cannot be inherited, this can be addressed by learning a controller from scratch, e.g. as in [16], although at the expense of ignoring any previously learned knowledge.

In this paper, we choose to use a fixed structure neural network for reasons of efficiency, motivated by the goal of eventually evolving directly in hardware. As in previous work, we use a learning algorithm to optimise a controller that has a fixed structure that matches the new body-plan [10]. However, in order to avoid starting from scratch for each body-plan, we

introduce a novel method for storing past solutions that can be accessed by the learning algorithm to bootstrap learning. This takes the form of an archive that stores the weights of a controller for each ‘type’ of robot that has previously been encountered as described in the previous section. This archive or ‘brain pool’ is dynamically composed and adapted during the evolutionary process.

Note that the term *archive* should not be confused with other uses of the word in the wider evolutionary literature. For example, archives are commonly used in multi-objective optimisation to either drive the population toward the Pareto front or to maintain population diversity, that is, they directly interact with an evolving population. Our approach has more in common with methods that try to enhance a search process by re-using past experience gained when solving related problems. For example, Louis and McDonnell [26] maintained a store of past solutions from similar instances which are periodically injected into an evolving population. However their approach is only applicable if instances share structural properties, hence cannot be applied to controllers with different topologies. Feng *et al.* [27] attempted to reuse structured common knowledge captured in the optimized solutions of past search experiences in a form independent of solution representation, however, their specific implementation was tailored to combinatorial optimisation. Here, we draw inspiration from [26] in maintaining an archive of past solutions. Extending this concept, we store solutions with different *types*, corresponding to different controller topologies, organised in the form of a grid. While our use of a grid-based container to organise solutions according to a set of features resembles the quality-diversity algorithm MAP-Elites [28], the role of the archive in our work is different: (1) in MAP-Elites, the grid is first filled using randomly generated solutions, which form a *population*; (2) the population is acted on by an evolutionary process that performs mutation and/or crossover to generate new solutions which are projected back to the grid. In our work, the archive does not act as a population but as a library. Cells are only filled with a controller if and when a particular ‘type’ is generated by the MEA. A single controller from the library is selected to initialise a separate learning algorithm. In summary, the processing of filling cells is governed by the MEA, while the stored controllers are used to inform the learning algorithm.

III. METHODS

A. Algorithm Description

Morpho-evolution with learning using an archive inheritance (MELAI) is a nested optimisation algorithm. As illustrated in figure 1, body-plans are optimised with an evolutionary algorithm, then for each body-plan, a learning process is used to optimise their controller.

The first optimisation algorithm (MEA) uses a generative encoding to produce the robot’s body-plan, based on our previous work described in [9]. This is a matrix-based CPPN morpho-evolution denoted MCME. The second optimization algorithm (learning) optimises the parameters of a controller which is a fixed size neural network structure. Therefore, the

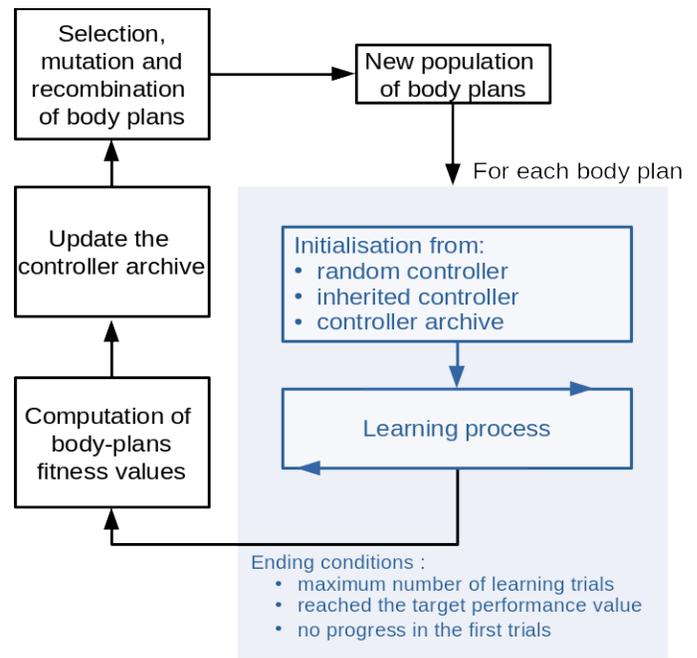


Fig. 1. Diagram illustrating the MELAI algorithm. MELAI has two nested optimisation processes. As the main process, a morpho-evolution algorithm, shown in black, divides into four main steps: *computation* of the fitness values, *update* the controller archive with the best ones from the current population, *Selection, mutation, and recombination*, and finally send the new population of new body-plans to the learning process. For each body-plan, a controller is learned, shown in blue. The learning process can either start from a random controller, an inherited controller, or a controller from the archive. Then, the learning process runs until reaching an end condition.

number of parameters is fixed. The novelty-driven increasing population evolutionary strategy (NIP-ES) algorithm [10] is used for learning. A detailed description of both MCME and NIP-ES is given in the supplementary materials. Although the instantiation of MELAI described in this paper uses MCME and NIP-ES, the framework itself is general in that any kind of MEA or learning algorithm could be used.

Inheritance of controllers from parents to children is challenging for MELAI as previously noted, since children might have different body-plan configurations than their parents. One way to address this issue is to learn the controller for each robot from scratch as in [16]. However, this has a number of disadvantages, including the fact that previously learned information from past learning cycles is wasted. In order to address this issue MELAI, has three alternative initialisation options. These are:

- 1) Select a controller from the archive with the same number of sensors and actuators assuming one exists, or
- 2) Start from a randomly initialised controller, or
- 3) Direct controller inheritance if the parent and child share the same number and type of actuators and sensors.

In this paper, the third option of direct inheritance is not considered because the encoding and the morphological space used in MCME make it unlikely that a parent and child will share the same number and type of actuators and sensors. Thus, the benefit of direct inheritance is likely to be negligible.

This allows experiments to focus directly on determining the benefit of the archive over random initialisation.

As noted, the learning algorithm used is NIP-ES, first described in [10]. The core of this method is a co-variance matrix adaptation evolutionary strategy (CMA-ES) algorithm in which a normal multivariate distribution (MVND) is used to sample a new population at each iteration. When using a controller from the archive, it is used to provide the starting mean of the MVND and thus the starting population is sampled in the surrounding parameters of this controller. When starting from scratch, CMA-ES starts from a random mean.

The learning process stops when one of the ending conditions is reached:

- *A satisfactory solution is reached:* the learning algorithm finds a controller with a task-performance value above a certain threshold.
- *The maximum number of evaluations is reached:* each optimisation process has a maximum number of updates. For the MEA, this parameter is the number of generations and for the learning process is the number of evaluations. The values of these parameters are chosen according to the difficulty of the task and environment, and according to the constraints of the system on which the algorithm is running. Here, a constraint of 100000 maximum evaluations is used. Given this overall budget, an additional choice that must be made is to decide how to divide it between the MEA and the learning process.
- *The performance of the robot stays very low during a trial period.* The trial period is defined by a fixed number of evaluations (50 in all experiments). If a robot has not moved by the end of this period then the learning process stops.

In the rest of the paper, the *fitness* indicates the value used by the MEA for selection. The term *task-performance* is used by the learning algorithm to assess the quality of behaviour. Several fitness functions can be used for the MEA. The most natural fitness function is the *best* task-performance value found during the learning. This is the one used in this paper. However, the learning process produces additional data which could also be exploited in future, such as statistical information regarding the progress of the learning algorithm or the novelty-scores used by NIP-ES.

B. Controller Archive

The controller archive stores the best controller found for different ‘types’ of robot. A ‘type’ is defined as a n -tuple, where each dimension represents a component of the body-plan, e.g. a specific form of sensor or actuator. An archive thus consists of an n -dimensional cube, where each dimension is discretised into cells corresponding to the number of each kind of component. A cell contains the best controller found for a body-plan described by a given tuple generated by the MEA, and is empty otherwise.

For instance, consider a morphological space with one kind of actuator (x) and sensor (y) as shown in figure 2. Assume a body-plan is generated with 2 actuators and 3 sensors.

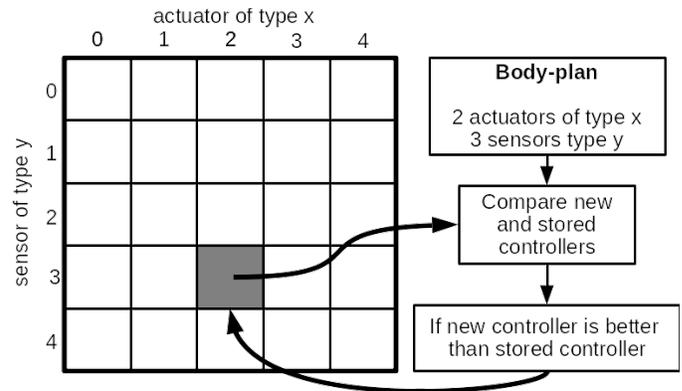


Fig. 2. Diagram illustrating the update of the controller archive. A body-plan with 2 actuators of type x and 3 sensors of type y has a new controller output of the learning process. If the cell corresponding to 2 actuators of type x and 3 sensors of type y is not empty, the new controller is compared with the stored one. The new controller replaces the stored one if its task-performance is greater.

When the learning process has ended, its controller’s task-performance value is compared with the stored one in the corresponding cell and replaces it if its task-performance value is greater. If the cell is empty the new controller is added to the cell.

We consider two kinds of actuators (wheel and joint), and one type of sensor. So, a ‘type’ of robot, in this case, is defined by a tuple $\langle \text{num_sensors}, \text{num_wheels}, \text{num_joints} \rangle$. The controller archive can be considered as a new form of inheritance. All the behavioural knowledge from past generations is stored in a common archive to be used by future generations. In this way, new child robots can leverage the learned behaviours of their ancestors.

IV. EXPERIMENTS

A. Experimental protocol

The experiments presented in this article aim to answer the following questions:

1. To what extent does using a controller-archive to bootstrap learning improve effectiveness and efficiency when compared to learning from scratch? And
2. To what extent does the distribution of effort between the morpho-evolution process and the learning process influence performance?

Experiments are conducted with and without the controller archive to answer question 1. In this way, the role of the controller archive in MELAI can be isolated. The variant of MELAI that does not use a controller archive is denoted morpho-evolution with learning (MEL).

Two tasks are used in experiments: exploration and phototaxis with multiple targets (explained in detail in section IV-D). For both tasks, the learning process has a budget of 200 evaluations. This learning budget is optimal for our experiments (see figure 8). For the exploration task, the evolution runs for 20 generations and for the photo-taxis task it runs for 15 generations. Experiments on the photo-taxis task run for fewer

generations because each robot needs to be evaluated for each target, thus, it is computationally expensive.

To answer question 2., more experiments are conducted only on the exploration task. For all of them, a fixed budget of 100000 evaluations is shared between the two optimisation processes. This has the objective of studying trade-offs in resource allocation between the two components of the framework. Parameter values tested are the following: [100,40], [150,30], [200,20], [400,10], [800,5], where the first value corresponds to the number of evaluations for each body-plan during the learning phase and the second to the number of generations of the MEA. Also, these variants of MELAI are compared with a baseline algorithm in which the learning process is replaced by a process that simply generates controllers using Latin Hypercube Sampling (LHS)¹ [29], i.e. there is no directed learning, only random sampling of controllers. This variant is called Morpho-Evolution with Latin Hypercube Sampling (MELHS). MELHS runs for 40 generations and for each body-plan 100 random controllers are sampled. These experiments are conducted only with the controller archive.

All experiments feature a population of 25 body-plans for the MEA. The hyper-parameters used for the experiments are given in the supplementary materials. All experiments are conducted in the three environments described in section IV-D and shown in figure 4. Twenty replicates are performed for each experiment. The robots and environments are simulated using the CoppeliaSim version 4.2.0 simulator². The source code to run these experiments and their data are available here: https://bitbucket.org/autonomousroboticsevolution/tcdis-2022/src/tcdis_rev/.

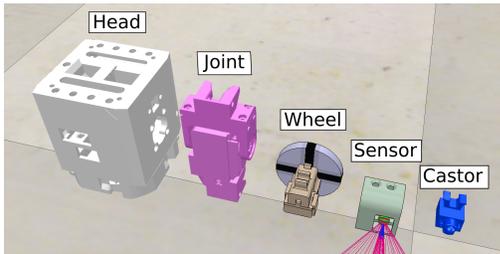


Fig. 3. Active and passive components are used for the experiments shown in this paper. The active components are the wheel, joint and sensor. The passive component is the castor wheel. The head is the central processing unit of the robot.

B. Body-plans

The body-plans evolved using the MEA (as described in section III-A) have two main features: skeletons can have complex and widely differing shapes and different numbers of components³. The components can be active (e.g. a wheel driven by motors, sensors) or passive (e.g. a castor-wheel)

¹LHS samples evenly the parameter space. This gives a better sampling than using a simple uniform distribution.

²<https://www.coppeliarobotics.com/>

³All components have been designed to match the physical ones which are used in the ARE project [30] in order to be able to conduct experiments directly in hardware in the future.

where the active ones interact with the controller. The different component types are shown in Figure 3. The controller interacts differently with each active component:

- Each *wheel* takes one output from the controller which is translated to the speed of the rotational movement of the wheel.
- Each *joint* takes one output from the controller which is translated to the frequency of the oscillatory movement of the joint. Oscillatory control has previously been demonstrated to give good results for locomotion [25], [24].
- Each *sensor* provides two inputs to the controller where the first input is binary denoting detection of a beacon and the second input is the distance from the closest obstacle. The detection of a beacon uses a simulated IR sensor and the distance measure uses a simulated time-of-flight sensor.

The head is the central processing unit of the robot and therefore is a special component of a body-plan. It is always positioned in the centre of the skeleton (see figure 10). To obtain an accurate simulation, the mass of the body-plan has to be estimated. This is calculated as the sum of its components' and skeleton's weights. To estimate the mass of the skeleton, which can take various shapes, the density of a common plastic used for 3d printing is used to estimate the mass of a voxel. Skeleton mass is then obtained by multiplying the voxel mass by the number of voxels used. This results in robots that have a mass between 0.5kg and 2kg depending on the size of the skeleton.

C. Controllers

The controller used in this study is a modified version of an Elman network [31]. An Elman network is a recurrent neural network with two hidden layers (see the figure in section C in Appendix II of the supplementary materials). The first hidden layer is fully connected to the input and output layers. Then, each neuron is forward connected to one neuron of the second hidden layer, called the context layer. Neurons in the context layer (context units) are recursively connected to themselves, and the context units are also fully backward connected to the hidden layer. Each neuron uses a sigmoid function as its activation function. The context layer acts as a short term memory and allows the network to process real number sequences such as time-series [31]. Elman networks have been shown to be more efficient as controllers for a navigation task than a simple feed-forward network [6] due to their ability to capture time-dependent information from sensors.

For each body-plan, each Elman network has a number of inputs and outputs corresponding to its body-plans' number of sensors and actuators. The hidden and context layers have a fixed structure for all the body-plans. In all experiments, the hidden layers have 8 neurons each. Transferring a trained network from one body-plan to another is not possible unless they have the same number of sensors and actuators.

D. Task and environments

Two tasks are considered: exploration and photo-taxis.

a) *Exploration task*: In this task, the robot has to maximise the number of zones visited in a limited time (30 seconds). Zones are equal-sized squares forming a grid. Task-performance is computed by counting the number of zones visited and dividing the count by the total number of zones. The grid is 8 by 8 with cells of 25cm sides, so the total number of zones is 64. Each evaluation lasts 30 seconds and takes place in the *obstacles* environment (see figure 4). In this task, there is no target performance value therefore the first stopping criterion defined in section III does not apply. Although a target performance value could have been set, it is impossible for the robot to visit all zones in 30 seconds, therefore the choice of a number of tiles would have been arbitrary.

b) *Photo-taxis task*: In this task, the robot starts at one point and has to reach a target where a beacon is placed. The robot must first find the beacon in the arena and then go towards it. As the beacon is detected using a simulated IR sensor, the robot can not see it when it is occluded by an obstacle. The robot is evaluated three times with the target at different positions. The task performance is then the average of the task performance obtained in each evaluation. The task-performance function is the normalised distance between the final position (p_f) of the robot (at the end of the evaluation) and the position of the beacon (p_b) (see equation 1). This distance is subtracted from 1 in order to define a maximisation function. The distance is normalized by the length of the diagonal of the arena. As the arenas are squares of two by two metres the diagonal measures $D = \sqrt{2^3} \approx 2.83$.

$$F = 1 - \frac{\|p_f - p_b\|}{D} \quad (1)$$

The success threshold used to stop the learning process is equal to 0.95 for this task. This value corresponds to a circle with a radius of 14 cm around the target.

c) *Environments*: Three different environments (figure 4) are used in the experiments dubbed *obstacles*, *escape room* and *arena*. They are all square (2 metres sides) and have a tiled floor. Each is designed to be reproducible in reality. The tiles are spaced with a small gap of 1 millimetre which corresponds to the floor of our real arenas and the walls in the obstacles environment have feet to hold them standing. These constraints in design introduce small irregularities.

V. RESULTS

This section is split into four parts: experiments relating to the efficiency and effectiveness of the algorithm, the controller archive dynamics, the influence of learning, and the robots' diversity. The first part focuses on comparing the algorithm with and without an archive by measuring the quality of the solutions produced and the efficiency of both variants. The second part examines the controllers stored in the archive in terms of their number and quality. The third part studies the influence of learning by comparing MELAI with different learning budgets and a variant without learning. Finally, the last part investigates the influence of the task and environment over the type of robots generated. Where it is relevant a statistical test is conducted. The test used is the Mann-Whitney

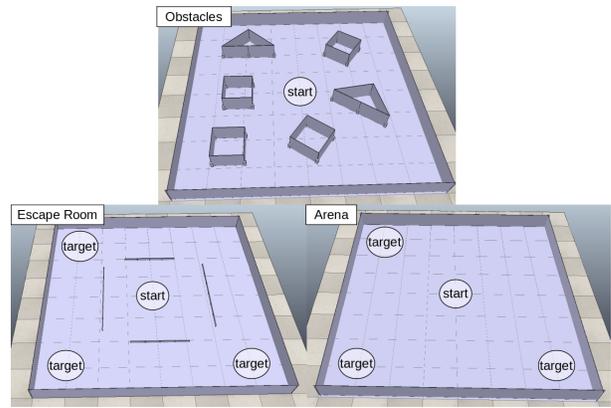


Fig. 4. The three environments used for the experiments in this paper: obstacles, escape room and arena.

U [32] under the null hypothesis using the distribution from the final generation. The test is two-tailed. Where the distributions are significantly different the p-value and the critical value (the U value of the Mann-Whitney U test) are indicated on the plots. All experiments have been replicated 20 times.

A. Efficiency and effectiveness

Three measures are used to assess the efficiency and effectiveness of MELAI: (a) the best fitness for each population of MEA, (b) the best and average initial task-performance of the learning process of the population and, (c) the total number of evaluations. The best fitness (a) is calculated for each population after the learning has finished. The initial task-performance (b) is the lowest task-performance from the first iteration of NIP-ES. The number of evaluations (c) used during one generation is the sum of the number of evaluations used by the learning process for each body-plan in the population. The best fitness and the initial task-performance are measures of the performance (effectiveness) of the complete MELAI algorithm while the number of evaluations measures the efficiency of the algorithm.

Figure 5 shows the plots of the best fitness (first row) and the best and average initial task-performance (second row) over the generations. For the exploration task and the arena with three targets, MELAI achieves better performance than MEL, corresponding to visiting more zones during the exploration task and minimising the distance from the target for the photo-taxis task. With the exploration task, MELAI generates robots that are able to visit between 26 and 28 zones out of a potential 64, whereas with MEL, the best robots visit between 22 and 24 zones. In the arena, MELAI finds a solution that reaches the success threshold (approximately 0.14 metre from the target) in fewer generations (~ 4 generations) than MEL (~ 6 generations). All replicates find a solution that reaches the target threshold with MELAI, while with MEL, some replicates do not reach the success threshold. However, both MELAI and MEL produce similar results on the escape room. More interestingly, on the three environments, the best and average initial task-performance (see the second row of figure 5) of MELAI is above the one obtained by MEL. This

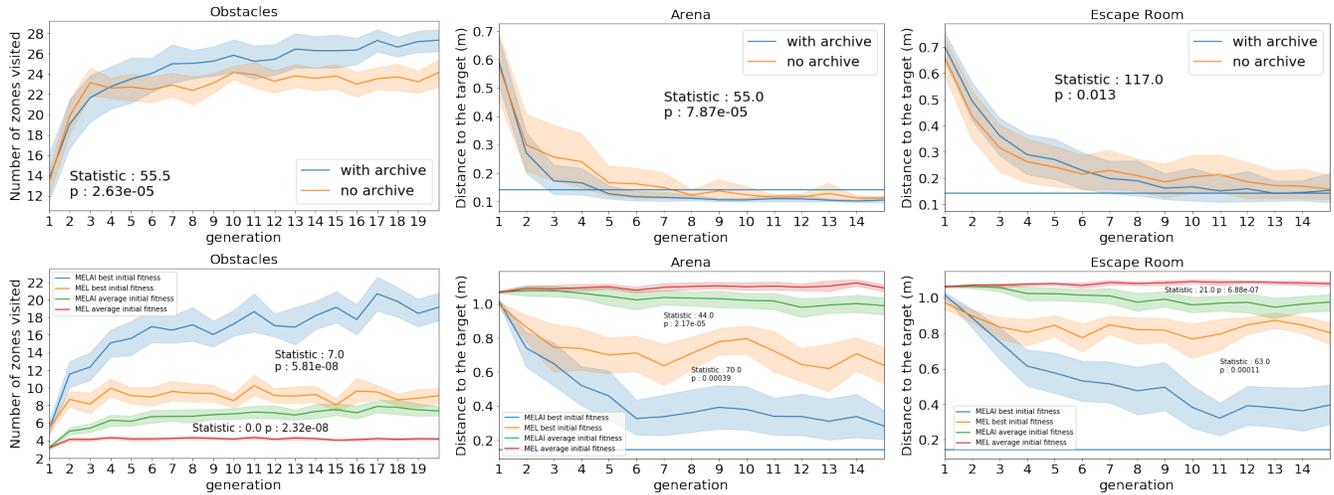


Fig. 5. Measures of effectiveness of the algorithms. In the first row, the best fitness values over the generations and in the second row the best and average of the *initial* task-performance values over the generations. These experiments have been conducted with a budget of 200 evaluations per body-plan and 20 generations for the exploration task and 15 for the photo-taxis task. The coloured areas correspond to the confidence interval around the mean. The difference between the distributions of the last generation is significant when a p-value and its critical value are indicated. The significant test is the Mann-Whitney U test.

shows that starting from a controller from the archive provides a better start for the learning algorithm.

Another benefit of the controller archive is with respect to efficiency (figure 6). On the photo-taxis task, the total number of evaluations used per generation decreases over time for both algorithms (MEL and MELAI). The number of evaluations used by MELAI is fewer than MEL, and also decreases faster (see figure 6). This dynamic does not appear in the exploration task because the learning algorithm does not have a target performance value for this task. So, when possible, transferring the controllers through the generations speeds up learning. In other words, the learning process increases its efficiency over generations when the archive is used.

Moreover, the difference between the initial and best performance values (learning delta) stays constant over the generations for both MELAI and MEL. As the archive allows learning to start from a better solution, MELAI can reach a better solution after learning and in a shorter time for the photo-taxis task. Additional figures can be found in the supplementary material which plots the learning delta over the initial task-performance and the learning delta over the generations and support the above interpretation.

B. Controller archive dynamics

To analyse the dynamics associated with using the controller archive, two metrics related to the controllers stored in the archive are monitored: the average and best task-performance value (first row of figure 7) and the number (second row of figure 7). The number of controllers in the archive corresponds to the number of ‘types’ of body-plan generated by the MEA (according to the 3-dimensional descriptor used).

The best task-performance in the archive corresponds to the best task-performance found so far (see figure 5). The average and best task-performance of the controllers stored in the archive follow the same dynamics as the best fitness

of the population of body-plans (see figure 5). The *average* task-performance value remains low as the controller archive retains controllers in cells that correspond to robots that have unsuitable body-plans for the tasks (e.g. no wheels).

The accumulation of controllers is shown in the second row of figure 7. The controller archive accumulates controllers through the generations: some controllers are replaced over time by higher-performing versions, while others may never be updated if the type of body-plan they belong to is never re-used. The rate of increase slows over time. As the number of controllers corresponds to the number of ‘types’ discovered by the algorithm, figure 7 shows that in the first generations tens of ‘types’ are discovered. Then, it slows down to 2 or 3 ‘types’ per generation.

C. Influence of learning

In addition to the controller archive, the combined evolution of morphology and learning of behaviours in MELAI leads to added complexity in analysing results. To gain insight into the interaction between these two optimisation processes, MELAI was run with five different budgets ([800,5], [400,10], [200,20], [150,30], and [100,40]) which correspond respectively to [learning budget, number of generations]. All the variants have a population of 25 body-plans. Thus each variant tests a different total number of body-plans, for instance, the variant [400,10] tests 250 body-plans. However, all the variants have the same total number of evaluations of 100000. These variants are compared with MELHS as the baseline, conducted on the exploration task.

Figure 8 shows the best task-performances (number of zones visited) over the number of evaluations. As expected, the advantage of using learning is clear. The best individuals produced by MELHS visit on average between 8 and 9 zones while all the variants of MELAI reach between 23 and 28 on average. With MELHS, the quality of the solutions increases

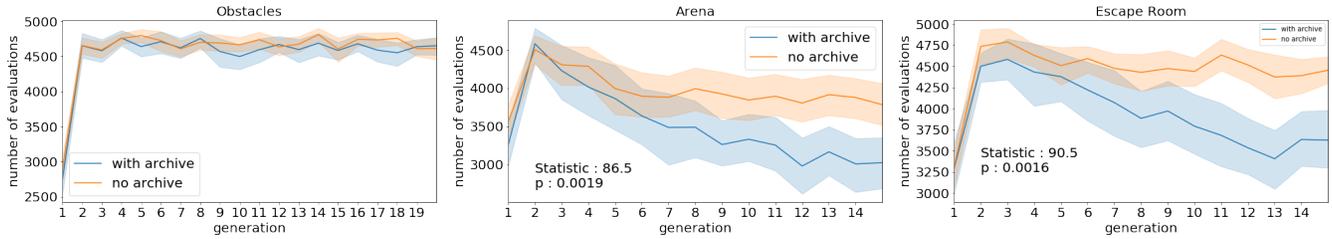


Fig. 6. The number of evaluations per generation. The values are plotted for MEL and MELAI. These experiments have been conducted with a budget of 200 evaluations per body-plan and 20 generations for the exploration task and 15 for the photo-taxis task. The coloured areas correspond to the confidence interval around the mean. The difference between the distributions of the last generation is significant when a p-value and its critical value are indicated. The significant test is the Mann-Whitney U test.

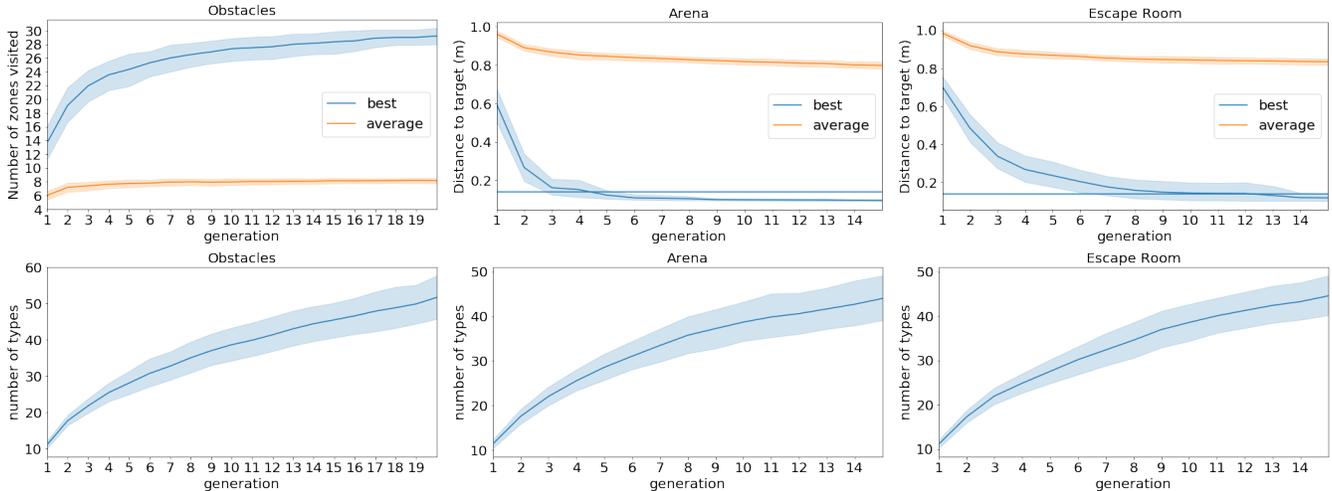


Fig. 7. Plots of three metrics over the generations related to the controllers in the archive. First row: the average and best task-performance. Second row: the number of controllers in the archive. The coloured areas correspond to the confidence interval around the mean.

very slowly, starting from 8 and reaching barely 9 at the end. However, MELAI demonstrates an obvious learning curve for all variants tested.

On the other hand, the difference between the different budgets using MELAI is small. The variant [800,5] is sub-optimal compared with the others. The variant [100,40] is the fastest to reach a satisfactory solution with an average of 25 zones visited after around 15000 evaluations. Generally, reducing the learning budget speeds up convergence. This is due to the generational aspect of MELAI: for a given number of *evaluations*, smaller learning budgets enable more *generations* to be conducted. Ultimately, apart from [800,5] all the variants converge to similar task-performances.

D. Robot diversity

Finally, figure 9 shows the distribution of the number of wheels, joints and sensors of the robots with the highest task-performance over the generations. These show that the majority of successful robots have between three and five wheels for all the environments and both tasks. In contrast, the majority of successful robots have no joints. This is unsurprising as the three environments have a flat floor rendering joints unnecessary. MELAI discovers that for photo-taxis at least one sensor is required, while for the exploration task the majority of the best solutions does not feature sensors. Indeed,

to reach the target at three different locations in the photo-taxis, a sensor is required. On the contrary, ‘blind’ robots can easily visit multiple zones.

These results are not surprising given that the main optimisation process in MELAI is an MEA. Evolution is most likely to proceed along the ‘easiest’ path that enables it to maximise the fitness function. In this case, this corresponds to robots with only wheels. This type of robot is easier to control and therefore it is easier to learn a controller than for robots with joints and sensors. Sensors emerge only if they are necessary as in the photo-taxis task. Thus, MELAI is able to produce different ‘types’ of robots depending on the task. This is evidenced in additional plots provided in the Supplementary Material. Illustrations of the most common and efficient designs found by MELAI for each task are shown in figure 10 as examples. Only one robot is shown for the photo-taxis task because for both environments *arena* and *escape room* the algorithm converges to the same ‘type’. However, additional pictures are provided in the Supplementary Material. Worthy of note is that similarity and ‘minimalism’ of both body-plans. The similarity confirms the lack of diversity in the best robots generated. The simplistic design shows that MELAI reduced the cost of production of such robots despite the lack of an explicit objective to do so.

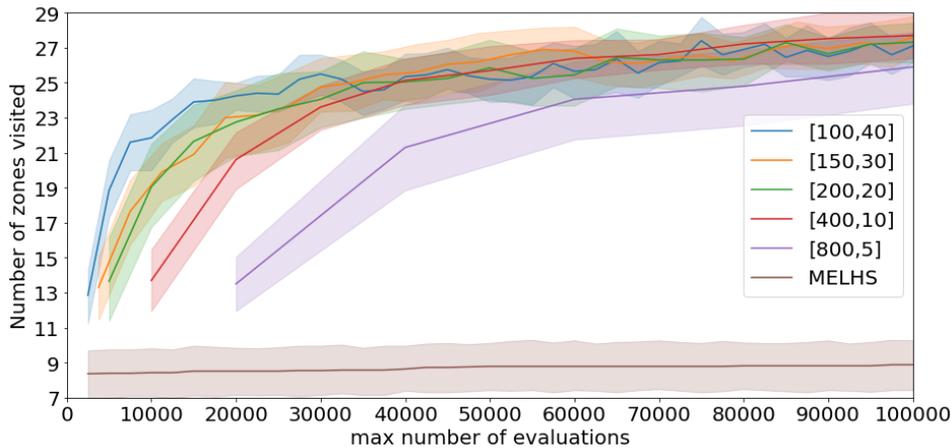


Fig. 8. Best fitness over the number of evaluations on the exploration task for six variants: $[evaluations, generations]$ of $[800,5]$, $[400,10]$, $[200,20]$, $[150,30]$, and $[100,40]$ and MELHS. The coloured areas correspond to the confidence interval around the mean.

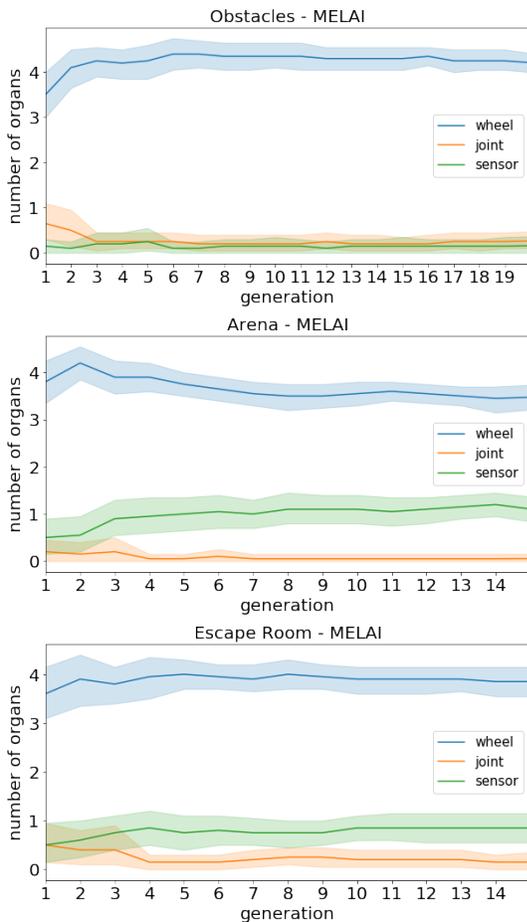


Fig. 9. For each of the three environments, distribution of the number of wheels, joints and sensors over the best robots of each generation. The coloured areas correspond to the confidence interval and the solid curves to the central tendency.

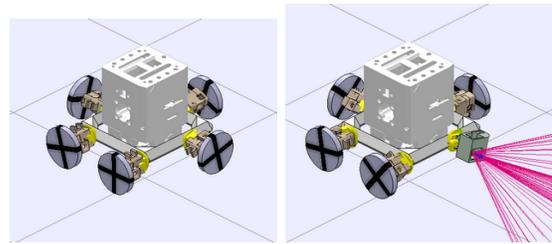


Fig. 10. The most common and efficient design found by MELAI for each task. On the left for the exploration task and on the right for the photo-taxis task.

VI. DISCUSSION

Advancing previous work in the domain of body-brain evolution, we have proposed a method to evolve robots in a rich morphological space that includes a variety of sensors and actuators, and that can realise skeletons with diverse forms and sizes. Hence, a considerably more diverse range of body-plans can be produced in this space than in previous work that either uses modular systems [24] or spaces in which the components have common control mechanisms [16]. The richer space increases the likelihood that a controller produced via evolutionary operators will not match a new body-plan. Although using a generative (morphology-independent) encoding can address this inheritance issue, the time-complexity associated with these methods can be prohibitive when working with physical robots.

To address this we have proposed the use of an external archive that stores a learned controller associated with a ‘type’ of robots. As described in section V, we have demonstrated that the archive significantly improves the quality of the solution and efficiency in evolving a body-plan capable of solving multiple tasks. The role of the archive and the various components of the framework that lead to this result is discussed below.

Each cell of the controller archive stores the best controller learned during an individual lifetime. The archive thus represents a history of knowledge discovered in previous

generations that can be passed to future generations. Therefore, it acts as a novel form of *inheritance*. In the sense that it stores information learned during an individual lifetime, it shares characteristics with Lamarckian artificial evolutionary systems [25]. Note that the tuple defining a ‘type’ is deliberately simple. However, it should be clear that many different body-plans can be mapped to a single cell, given that for any given combination of sensors and actuators, the robot’s skeleton they are attached to can vary enormously in shape and size, and the configuration of components can also vary. Given this variation, it might be expected that inheriting a controller of the correct ‘type’ would not necessarily bring many benefit. However, it is clear from figure 5 that shows the best and initial task-performance, that starting from a controller from the archive brings a significant advantage. Interestingly, this suggests that there is some generalisation of controllers across a range of body-plans. The results, shown in the second row of figure 5, show that inheriting from the archive bootstraps the learning process, and the size of this effect increases in magnitude as the generations progress.

It would of course be possible to define each cell using a higher degree of granularity, although it is reasonable to assume there is a balance to be struck in not making the archive too granular (which at the extreme would map every robot to an individual cell). Another way to approach this would be to store multiple controllers per cell, and either try them all, select one at random, or use clustering or species system to select the most suitable one. Also, increasing the granularity of the archive requires defining an appropriate morphological descriptor. This is not an easy task. In our previous work [9], different morphological descriptors were studied in the context of novelty search, designed to reflect different types of control (e.g. due to symmetry). The results showed that in fact, the most simple descriptor that simply counted components was better in producing a greater diversity of body-plans.

Recall that the framework consists of two components: an MEA that learns body-plans and a learning algorithm based on an evolutionary strategy that learns controllers. The former is selected for its ability to explore a diverse space of plans and based on previous work [9]. The selection of NIP-ES as the learner is deliberate in that this algorithm demonstrates high *exploration* capabilities. This is essential as the learner might have to start from scratch if no controller is available in the archive, or a selected controller might not be well adapted to a new body. In contrast, previous work which has used learning as a mechanism to enhance a controller selected by evolution (e.g. [24]) can afford to be much more exploitative.

Despite the diversity produced by MCME, the successful robots are all of the same ‘type’: mostly wheels, a few sensors, and almost no joints. In fact, MELAI falls into the local optima of the robots for which it is easy to learn a controller. This issue is common in evolutionary robotics and it is accentuated by the joint optimisation of body-plan and controllers. In the work of Cheney *et al.* [18], this issue is explained by the fact that promising body-plans for which it takes longer to learn a controller are quickly dropped by the EA. One of their later works [19] proposed the *morphological innovation protection*

mechanism. Each body-plan has an attribute corresponding to its age, which increases with each generation. The selection mechanism favours younger body-plans and thus protects new body-plans that need more time to learn a controller. In MELAI, this solution could be implemented by attributing bigger budgets to the learning process for younger body-plans and decreasing their budgets as they age. Another possible reason for the premature convergence to a local optimum is the generational aspect of the EA. At each generation, the selection mechanism is applied to the whole population with the result that a high performing solution will often invade the population. An asynchronous parallel evolution (APE) approach proposed by Gupta *et al.* [16] provides a possible solution. As a side-effect, they observe that this results in more diversity in the highest performing body-plans. APE applies tournament selection to small groups of four individuals asynchronously. A low performing robot can be preserved for longer as it is not always compared to the highest performing individuals.

Given the importance of the learning loop just discussed when jointly optimising body-plans and controllers, it is natural then to discuss how a computational budget should be balanced between the outer evolutionary loop and the inner learning loop. The results shown in figure 8 shed some light on this by varying the budget assigned to the learning from 100 to 800 evaluations. The smaller learning budget delivers a faster bootstrap in both environments as generations are shorter. However, 200 evaluations is the necessary minimum budget to allow NIP-ES to reach its full potential [10] (see supplementary materials). So, the choice of the budget is dependent on the learning algorithm used in MELAI. Also, it is worth noting that the decision of how to split this budget is influenced by whether one is working in simulation or on physical robots: in simulation, generating a body-plan has negligible cost whereas in reality, producing a physical robot can take weeks [17]. In contrast, evaluations are cheap in both environments hence this may influence the choice.

Finally, many design choices made in this work are made with the intention of applying the scheme to a physical system [30]. In particular, NIP-ES and the controller archive reduce the number of evaluations needed to reach a satisfactory solution. However, further progress is still required in this direction: on the photo-taxis task in the arena, MELAI needs around 15000 evaluations shared among 100 body-plans tested to find a robot that completes the task. In future work, hybrid methods using both simulated and real robots will therefore be investigated.

VII. CONCLUSION

This paper has proposed a new framework MELAI for the joint optimisation of body-plans and controllers in a diverse and complex morphological space. The framework intertwines an evolutionary algorithm MCME for evolving body-plans with an evolution strategy NIP-ES for learning individual controllers. Its key novelty is in the use of an external archive for storing learned controllers for different ‘types’ of robots. This acts as a means of transferring learned information

between multiple generations and is used to bootstrap the learning mechanism. Hence it can be seen as a form of non-genetic inheritance. It is shown to bring benefits with respect to efficiency, leading to increased rates and magnitude of learning over generations. Finally, it provides new insights into the complex interactions between evolution and learning, adding to the growing amount of recent literature on this subject (e.g. [16], [17], [24]).

Looking ahead, the work provides a foundation for moving towards applying the framework to the evolution of robots in a hybrid system that mixes evolution in hardware and simulation. In such a space, increasing the efficiency of the evolutionary cycle is key for reasons that include time, cost of materials, and wear and tear on robotic parts.

ACKNOWLEDGMENT

This work is funded by EPSRC ARE project, EP/R03561X, EP/R035733, EP/R035679, and the Vrije 610 Universiteit Amsterdam.

REFERENCES

- [1] R. A. Brooks, "Intelligence without reason," in *Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1*, ser. IJCAI'91. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1991, p. 569–595.
- [2] R. Pfeifer and J. Bongard, *How the body shapes the way we think: a new view of intelligence*. MIT press, 2006.
- [3] S. Kriegman, D. Blackiston, M. Levin, and J. Bongard, "A scalable pipeline for designing reconfigurable organisms," *Proceedings of the National Academy of Sciences*, vol. 117, no. 4, pp. 1853–1859, 2020. [Online]. Available: <https://www.pnas.org/content/117/4/1853>
- [4] V. Vujovic, A. Rosendo, L. Brodbeck, and F. Iida, "Evolutionary developmental robotics: Improving morphology and control of physical robots," *Artificial Life*, vol. 23, no. 2, pp. 169–185, 2017, pMID: 28513207.
- [5] K. O. Stanley, "Compositional pattern producing networks: A novel abstraction of development," *Genetic programming and evolvable machines*, vol. 8, no. 2, pp. 131–162, 2007.
- [6] L. K. Le Goff, E. Hart, A. Coninx, and S. Doncieux, "On pros and cons of evolving topologies with novelty search," in *Artificial Life Conference Proceedings*. MIT Press, 2020, pp. 423–431.
- [7] A. Eiben, N. Bredeche, M. Hoogendoorn, J. Stradner, J. Timmis, A. Tyrrell, and A. Winfield, "The triangle of life: Evolving robots in real-time and real-space," in *Proc. of the 12th European Conference on the Synthesis and Simulation of Living Systems (ECAL 2013)*, P. Lio, O. Miglino, G. Nicosia, S. Nolfi, and M. Pavone, Eds. MIT Press, 2013, pp. 1056–1063.
- [8] A. E. Eiben and E. Hart, "If it evolves it needs to learn," in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1383–1384.
- [9] E. Buchanan, L. K. Le Goff, W. Li, E. Hart, A. E. Eiben, M. De Carlo, A. F. Winfield, M. F. Hale, R. Woolley, M. Angus, J. Timmis, and A. M. Tyrrell, "Bootstrapping artificial evolution to design robots for autonomous fabrication," *Robotics*, vol. 9, no. 4, p. 106, 2020.
- [10] L. K. Le Goff, E. Buchanan, E. Hart, A. E. Eiben, W. Li, M. de Carlo, M. F. Hale, M. Angus, R. Woolley, J. Timmis, A. Winfield, and A. M. Tyrrell, "Sample and time efficient policy learning with cma-es and bayesian optimisation," in *Artificial Life Conference Proceedings*. MIT Press, 2020, pp. 432–440.
- [11] G. Buason, N. Bergfeldt, and T. Ziemke, "Brains, bodies, and beyond: Competitive co-evolution of robot controllers, morphologies and environments," *Genetic Programming and Evolvable Machines*, vol. 6, no. 1, pp. 25–51, 2005.
- [12] T. F. Nygaard, E. Samuelsen, and K. Glette, "Overcoming initial convergence in multi-objective evolution of robot control and morphology using a two-phase approach," in *European Conference on the Applications of Evolutionary Computation*. Springer, 2017, pp. 825–836.
- [13] T. F. Nygaard, C. P. Martin, E. Samuelsen, J. Torresen, and K. Glette, "Real-world evolution adapts robot morphology and control to hardware limitations," in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2018, pp. 125–132.
- [14] T. F. Nygaard, C. P. Martin, D. Howard, J. Torresen, and K. Glette, "Environmental adaptation of robot morphology and control through real-world evolution," *arXiv preprint arXiv:2003.13254*, 2020.
- [15] J. Juárez-Guerrero, S. Muñoz-Gutiérrez, and W. M. Cuevas, "Design of a walking machine structure using evolutionary strategies," in *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, vol. 2. IEEE, 1998, pp. 1427–1432.
- [16] A. Gupta, S. Savarese, S. Ganguli, and L. Fei-Fei, "Embodied intelligence via learning and evolution," *arXiv preprint arXiv:2102.02202*, 2021.
- [17] T. Liao, G. Wang, B. Yang, R. Lee, K. Pister, S. Levine, and R. Calandra, "Data-efficient learning of morphology and controller for a microrobot," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2488–2494.
- [18] N. Cheney, J. Bongard, V. SunSpiral, and H. Lipson, "On the difficulty of co-optimizing morphology and control in evolved virtual creatures," in *Artificial Life Conference 2016*, 2016, p. 226.
- [19] —, "Scalable co-optimization of morphology and control in embodied machines," *Journal of The Royal Society Interface*, vol. 15, no. 143, p. 20170937, 2018.
- [20] K. O. Stanley and R. Miikkulainen, "Evolving neural networks through augmenting topologies," *Evolutionary computation*, vol. 10, no. 2, pp. 99–127, 2002.
- [21] K. Sims, "Evolving 3d morphology and behavior by competition," *Artificial life*, vol. 1, no. 4, pp. 353–372, 1994.
- [22] G. S. Hornby, H. Lipson, and J. B. Pollack, "Generative representations for the automated design of modular physical robots," *IEEE transactions on Robotics and Automation*, vol. 19, no. 4, pp. 703–719, 2003.
- [23] K. Miras, E. Haasdijk, K. Glette, and A. Eiben, "Search space analysis of evolvable robot morphologies," in *International Conference on the Applications of Evolutionary Computation*. Springer, 2018, pp. 703–718.
- [24] K. Miras, M. De Carlo, S. Akhatou, and A. Eiben, "Evolving-controllers versus learning-controllers for morphologically evolvable robots," in *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*. Springer, 2020, pp. 86–99.
- [25] M. Jelisavcic, K. Glette, E. Haasdijk, and A. Eiben, "Lamarckian evolution of simulated modular robots," *Frontiers in Robotics and AI*, vol. 6, p. 9, 2019.
- [26] S. J. Louis and J. McDonnell, "Learning with case-injected genetic algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 4, pp. 316–328, 2004.
- [27] L. Feng, Y.-S. Ong, A.-H. Tan, and I. W. Tsang, "Memes as building blocks: a case study on evolutionary optimization + transfer learning for routing problems," *Memetic Computing*, vol. 7, no. 3, pp. 159–180, 2015.
- [28] J.-B. Mouret and J. Clune, "Illuminating search spaces by mapping elites," *arXiv preprint arXiv:1504.04909*, 2015.
- [29] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 42, no. 1, pp. 55–61, 2000.
- [30] M. F. Hale, M. Angus, E. Buchanan, W. Li, R. Woolley, L. K. Le Goff, M. De Carlo, J. Timmis, A. F. Winfield, E. Hart *et al.*, "Hardware design for autonomous robot evolution," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 2140–2147.
- [31] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
- [32] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.