

TriFinger: An Open-Source Robot for Learning Dexterity

Manuel Wüthrich¹ Felix Widmaier¹ Felix Grimminger¹ Joel Akpo¹ Shruti Joshi¹ Vaibhav Agrawal¹
 Bilal Hammoud^{2,1} Majid Khadiv¹ Miroslav Bogdanovic¹ Vincent Berenz¹ Julian Viereck^{2,1}
 Maximilien Naveau¹ Ludovic Righetti^{2,1} Bernhard Schölkopf¹ Stefan Bauer¹

Abstract

Dexterous object manipulation remains an open problem in robotics, despite the rapid progress in machine learning during the past decade. We argue that a hindrance is the high cost of experimentation on real systems, in terms of both time and money. We address this problem by proposing an open-source robotic platform which can safely operate without human supervision. The hardware is inexpensive (about \$5000) yet highly dynamic, robust, and capable of complex interaction with external objects. The software operates at 1-kilohertz and performs safety checks to prevent the hardware from breaking. The easy-to-use front-end (in C++ and Python) is suitable for real-time control as well as deep reinforcement learning. In addition, the software framework is largely robot-agnostic and can hence be used independently of the hardware proposed herein. Finally, we illustrate the potential of the proposed platform through a number of experiments, including real-time optimal control, deep reinforcement learning from scratch, throwing, and writing.

1. Introduction

In the past decades, researchers have devoted great efforts to making computers more autonomous, such that they could take care of more and more tasks in our society. Today, algorithms make a great number of decisions which were made by humans previously: They decide on the best way to get from one place to another, what movie we may want to watch next and what articles we might be interested in. They take strategic decisions for companies, place advertisements and invest large sums of money. They play and win against us in video games (Mnih et al., 2013), chess (Campbell et al., 2002), and Go (Silver et al., 2017). Yet, the way we

¹Max Planck Institute for Intelligent Systems, Tübingen, Germany ²Tandon School of Engineering, New York University, Brooklyn, USA. Correspondence to: Manuel Wüthrich <manuel.wuthrich@gmail.com>.

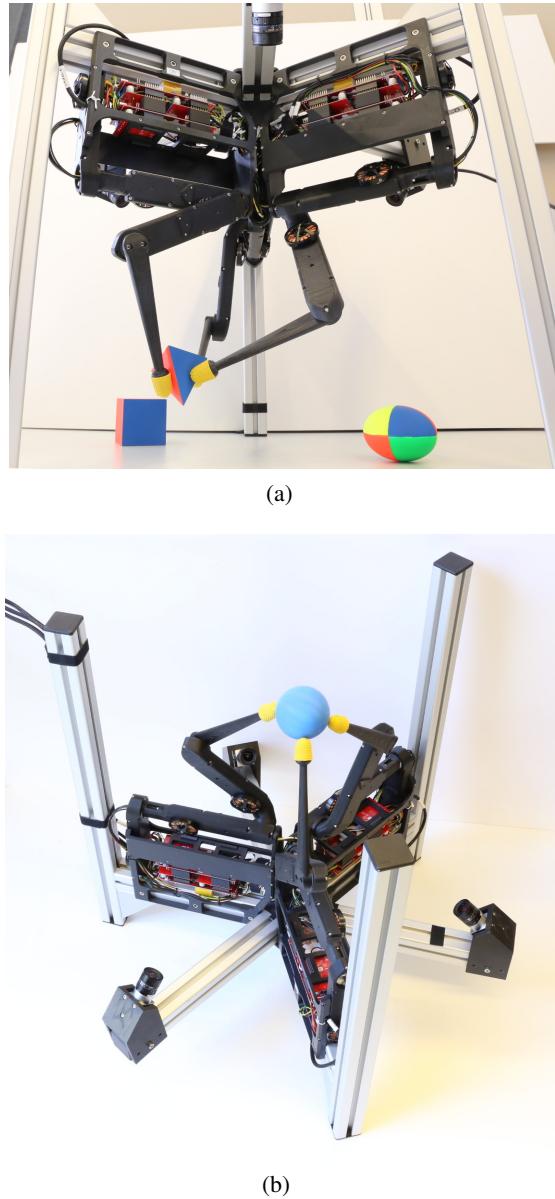


Figure 1. (a) The proposed platform can be used for object manipulation on the table. (b) The same platform can be flipped and used for e.g. catching and throwing.

construct buildings, clean, cook, dispose of our trash, plant and harvest food, search and rescue and assist people with physical disabilities has remained largely unaffected. Where does this striking contrast come from?

A difference which catches the eye is that the first class of tasks is primarily concerned with manipulation of information, whereas the second class is concerned with manipulation of matter, achieved through physical contact between the agent and the external world. We believe that a key issue which has hindered progress in the second class of problems is the high cost of experimentation. Robotic manipulators are typically very expensive and can break easily when entering into contact with the external world. Furthermore, they are usually operated only in the presence of human supervisors, ready to hit the emergency stop in case something goes wrong. These factors have largely prohibited systematic large scale experimentation on physical manipulation systems thus far. A large part of the robotic reinforcement learning (RL) community has hence focused on simulation experiments (e.g. Haarnoja et al., 2018; Fujimoto et al., 2018; Popov et al., 2017; Mnih et al., 2016; Heess et al., 2017; Duan et al., 2016; Henderson et al., 2018). However, results obtained in simulation experiments do often not translate to real systems (e.g. Tobin et al., 2017; James et al., 2019). The physics of contact interaction are nonsmooth and the outcome is highly sensitive to parameters and initial conditions (e.g. a slight difference in the shape of objects in contact can lead to very different motions).

Therefore, we believe that an experimental platform, capable of generating large amounts of data from a wide range of possible contact interactions with external objects at a low cost, could greatly support progress in autonomous robotic manipulation. The goal of this paper is to take a step in this direction. We present an open-source robotic platform called TriFinger. Its hardware and software design provide it with the following key strengths:

Dexterity: The robot design consists of three fingers and has the mechanical and sensorial capabilities necessary for complex object manipulation beyond grasping.

Safe Unsupervised Operation: The combination of robust hardware and safety checks in the software allows users to run even unpredictable algorithms without supervision. This enables, for instance, training of deep neural networks directly on the real robot.

Ease of Use: The C++ and Python interfaces are very simple and well-suited for reinforcement learning as well as optimal control at rates up to 1 kHz. For convenience, we also provide a simulation (PyBullet) environment of the robot.

Viability: The hardware design, based on (Grimminger et al., 2020), is very simple and inexpensive (about

\$5000 for the complete system), such that as many researchers as possible will be able to build their own platforms. All the necessary information to reproduce and control the platform is available here¹.

An important point to note is that most of the software framework is robot-agnostic and new systems can be integrated easily. Hence, it is a contribution in its own right and can be used independently of the robotic system proposed here.

In the remainder of the paper, we discuss related work and then describe the design of the hardware and software in detail. Finally, we present experiments in learning and optimal control to illustrate the aforementioned capabilities.

2. Related Work

In the past years, a large part of the RL community has focused on simulation benchmarks, such as the deepmind control suite (Tassa et al., 2018) or OpenAI gym (Brockman et al., 2016) and extensions thereof (Zamora et al., 2016). These benchmarks internally use physics simulators, typically Mujoco (Todorov et al., 2012) or pyBullet (Coumans & Bai, 2016).

These commonly accepted benchmarks allowed researchers from different labs to compare their methods, reproduce results, and hence build on each other’s work. Very impressive results have been obtained through this coordinated effort (see e.g. Haarnoja et al., 2018; Fujimoto et al., 2018; Popov et al., 2017; Mnih et al., 2016; Heess et al., 2017; Duan et al., 2016; Henderson et al., 2018).

In contrast, no such coordinated effort was possible on real robotic systems, since there is no shared benchmark. There have been isolated successes on real systems (Levine et al., 2018; Zhu et al., 2020; Pinto & Gupta, 2016; Andrychowicz et al., 2020), but these results cannot be compared or reproduced, since each lab has their own robotic setup.

2.1. Robot Hardware

This lack of standardized real-world benchmarks has been recognized by the robotics and reinforcement learning community (Behnke, 2006; Bonsignorio & del Pobil, 2015; Calli et al., 2015a;b; Amigoni et al., 2015; Murali et al., 2019a). Recently, there have been renewed efforts in this direction: For instance, Pickem et al. (2017) propose the ROBOTARIUM, a real-world benchmark for mobile robots, and Grimminger et al. (2020) propose an open-source quadruped.

For manipulation, there are a number of affordable robots, such as Franka Emika, Baxter, and Sawyer. Yang et al. (2019) propose Replab, a simple manipulation platform which has an even lower cost. Similarly, CMU proposed

¹<https://sites.google.com/view/trifinger>

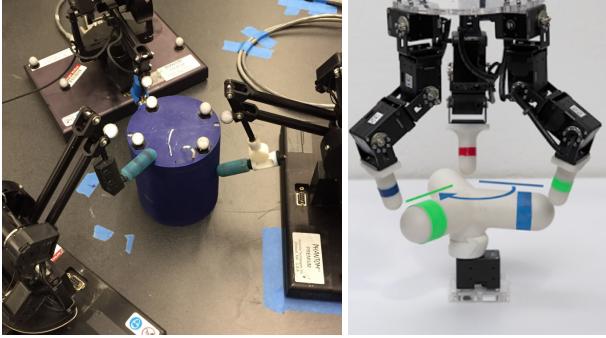


Figure 2. Phantom Manipulator (left, image from (Lowrey et al., 2018)) and D’Claw (right, image from (Ahn et al., 2019)). As the TriFinger, both platforms consist of three manipulators with 3 DoF each.

LoCoBot, a low-cost, open-source platform for mobile manipulation. However, all of these platforms have very simple end-effectors, typically 1-D grippers, which limit the possibilities of interaction with the environment.

For dexterous manipulation, there are a number of robotic hands on the market (e.g. BarrettHand, Shadow Hand, Schunk Hand) which typically cost over \$50 000 per hand. In addition, Dollar & Howe (2010); She et al. (2015); Xu & Todorov (2016) proposed some innovative, exploratory hand designs. However, none of these hands are suited for long-term unsupervised operation. In addition, to have a sufficient workspace for manipulation, they have to be mounted on a robot arm, which increases the complexity and risk of damage even further.

The two setups which are most similar to the TriFinger are the D’Claw, a three-fingered robotic hand (Ahn et al., 2019) and the Phantom Manipulation Platform, consisting of three Phantom Haptic Devices (Lowrey et al., 2018), see figure 2. As the TriFinger, both of these setups consist of three manipulators with 3 DoF each. However, the workspace where these manipulators can interact with objects is much larger for the TriFinger (see figure 6). In addition, the Phantom Manipulation Platform is at \$30 000 about six times more expensive than the proposed setup. The D’Claw robot, at \$3500, is in a similar price range as the proposed setup. However, its actuators allow for far less dynamic motion and are less robust to impacts, as they are only backdrivable with substantial force, i.e. they do not give in as easily.

2.2. Robot Software

Robot software can roughly be divided into two classes:

- Low-level software allows to communicate with the robot in real-time. Essentially, it sends motor commands to the robot and retrieves sensory measurements.

- High-level software which does not need to run in real-time and often operates on a more abstract level. The most widely-used high-level software is by far ROS. More recently Murali et al. (2019b) proposed PyRobot, a python interface for motion generation and learning.

Here we are mainly concerned with the low-level robot control software. Unfortunately, there is no framework which is widely used across labs, instead there is a large number of diverse solutions. These solutions typically rely on programming languages close to machine language, in order to ensure that control loops run in real-time (i.e. at a constant rate, e.g. 1 kHz). For example SL (Schaal, 2009) uses C while ROS-control (Chitta et al., 2017), LAAS-CNRS Stack-of-Tasks (Mansard et al., 2009), and ETH control-toolbox (Gifthaler et al., 2018) are implemented in C++. A particular case is the IHMC Robotics software (*ihm*) which uses a real-time Java with a modified garbage collector. These frameworks differ in how end-users integrate their controllers. SL and control-toolbox provide static methods or classes to fill-in. ROS control uses the concept of ros-services to load and change controllers online. Stack-of-Tasks relies on Python bindings to dynamically interact with a control graph.

Unfortunately, these frameworks are only accessible to experienced users who spent substantial amounts of time getting used to the particular approach at hand. Implementation of controllers must usually follow a strict structure which does not easily accommodate e.g. neural networks. In contrast, we designed the user interface to be simple enough that even inexperienced users are able to write controllers easily. It essentially consists of only two functions: appending actions to a queue and accessing a history of observations. Both functions are exposed in C++ as well as Python. The user may employ these functions in any desired manner, for real-time or non-real-time control (using e.g. a neural network). An additional benefit of the proposed design is that it makes integration into high-level frameworks very simple.

3. Hardware Design

The hardware design is loosely inspired by thumb, index and middle finger of a human hand, see figure 3. We will therefore refer to the individual manipulators as fingers and to the whole setup as TriFinger. In figure 1(b) we can see the main components of the robotic platform, including the fingers, the frame and the three cameras. In the following we will describe each of these components. All the details necessary for building an instance of the proposed platform are open source (footnote 1).



Figure 3. Analogy between the human hand and the proposed platform.

3.1. Finger Mechanics and Electronics

The mechanics and electronics of the proposed robot are based on a recently published open-source quadruped ([Grimminger et al., 2020](#)) consisting of inexpensive high-performance motors, off-the-shelf parts, and 3D printed shells (see [figure 4](#)). Grimminger et al. (2020) originally



Figure 4. Figure from ([Grimminger et al., 2020](#)) showing the actuator module which is the main building block of their quadruped legs and our fingers.

proposed a 2 DoF leg, which they extended to 3 DoF in the meanwhile (see ² for an overview of the designs based on ([Grimminger et al., 2020](#))). The fingers of the proposed platform are identical to the 3 DoF version of the quadruped leg, apart from some slight modifications to the mounting, the 3D-printed shells, and the end-effector. We hence inherit all the favorable properties of the design from [Grimminger et al. \(2020\)](#):

²https://github.com/open-dynamic-robot-initiative/open_robot_actuator_hw/tree/master/mechanics

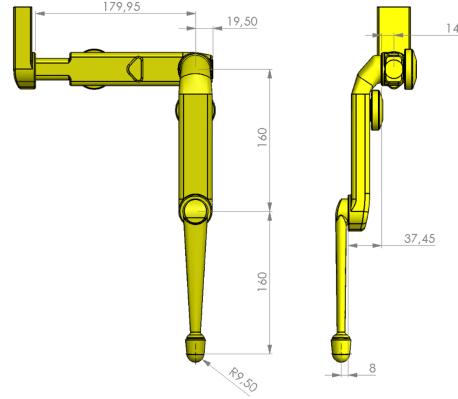


Figure 5. Technical drawing of a single finger, measurements are in mm.

- The high-performance brushless DC motors provide **high-torque** actuation while having a **low weight**, which allows for dynamic manipulation of objects up to a few hundred grams in weight.
- Transparency of the transmission enables **force control and sensing** and **robustness to impacts**, both of which are crucial for robotic manipulation. Transparency means that forces applied at the end-effector directly translate to torques at the motors, rather than being absorbed by a high-gear-ratio transmission. This implies that end-effector forces can be obtained by measuring the motor currents and that impacts will not break the transmission.
- The motors can be **controlled at high frequency (1 kHz)** from a consumer computer with a realtime-patched Ubuntu. This allows the robot to sense external forces and react to them extremely quickly.
- The **design is very simple** and consists of **inexpensive** off-the-shelf parts and 3D printed shells. This will allow other researchers to build their own platforms.

A small but important addition in our design is a soft tip which mimics the human finger tip. This increases the stability of interaction with external objects greatly, as impacts are damped and contact extends to a surface rather than a single point.

3.2. Kinematics

The kinematics were designed to maximize the workspace where all three fingers can interact with an object simultaneously, see [figure 6](#).

Each of the manipulators has 3 DoF, which implies that the finger tip can move in any direction (see [figure 5](#) for a



Figure 6. An overlay of different robot configurations to illustrate the workspace.

technical drawing). This is important for dexterity and it makes the finger robust to impacts, as it can give in to forces from any direction.

3.3. Frame and Boundary

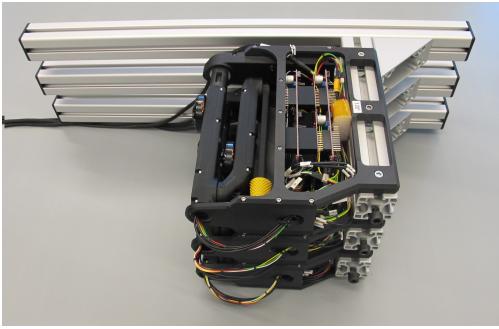


Figure 7. The platform can be disassembled into three modules for transportation and storage.

We use an aluminium frame to attach the fingers and cameras. The height of the fingers can be adjusted easily according to the requirements of a specific task. In addition, the entire platform can simply be flipped (see figure 1(b)) for tasks such as throwing and catching.

For manipulation on the table, we designed an optional boundary to confine objects to the workspace of the platform, see figure 8. This is essential for learning during extended periods of time without human supervision.

Finally, the platform can be disassembled easily into three compact modules, see figure 7.

3.4. Cameras

As there are three fingers interacting closely with the target objects, there is a lot of potential for occlusions. Therefore we place three cameras around the platform (see figure 1(b)),

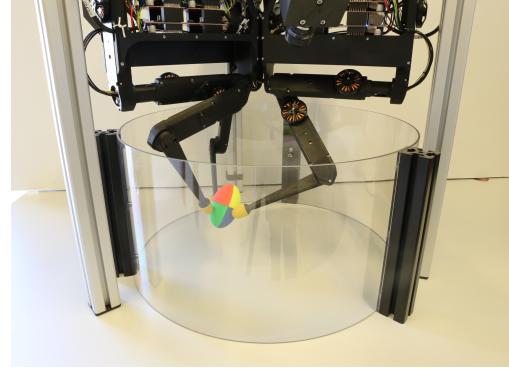


Figure 8. A boundary to prevent objects from leaving the workspace.

ensuring that the object will at all times be visible from at least one camera. We use Basler acA720-520uc cameras as they have a high rate of up to 525 fps using global shutter and a low latency. We mount a Basler C125-0418-5M-2000034830 lens on each camera, which leads to an appropriate field of view, see figure 9.



Figure 9. Images taken from each of the three cameras. Having three cameras ensures that the object will always be visible from at least one camera.

3.5. Measurements and Control Signals

As in (Grimminger et al., 2020), the signals are transferred between the control computer (standard consumer PC) and the motors through CAN at 1 kHz. In accordance with RL terminology, we will call the control signal action and the measurements observation.

Action: The input of this platform is a nine-dimensional vector: A desired torque (which is proportional to the current) for each joint. The robot expects this signal to be sent at a rate of 1 kHz.

Observation: The output consists of proprioceptive measurements (joint angles, joint velocities, joint torques), acquired at 1 kHz, and images from the three cameras, obtained typically at 100 Hz.

4. Software Design

The key strengths of our software framework are:

- The user interface in C++ and Python is very **simple**, yet well-suited for real-time (1 kHz) optimal control and reinforcement learning.
- It performs **safety checks** to prevent the robot from breaking. This frees the user from this burden and allows them to execute even complex and unpredictable algorithms without surveilling the platform. This opens, for instance, the possibility of training a deep neural network policy during several days directly on the robot.
- A synchronized **history of all the inputs and outputs** of the robot is available to the user and can be logged.
- The software is designed such that new robots and simulators can easily be integrated. This may facilitate reuse of algorithms across robots.

4.1. Control Modes

There are two modes of control supported by our software:

Definition 4.1 (Real-time control). By real-time control we mean that actions have to be sent to the system at a fixed rate of Δ seconds. This is necessary for real-world dynamical systems, as they evolve in time and can hence not wait for the next control to arrive. For instance, a falling humanoid robot cannot stand still in the air to wait for computation of the next action to be completed.

Definition 4.2 (Non-real-time control). By non-real-time control we mean that a change in the time at which actions are applied does not change the outcome, and that actions may take varying amounts of time. An example of such a system is a simulator: it will wait until the next action is provided and simulation of an action may take varying amounts of time for computational reasons. Another example is a mobile robot which is controlled through highlevel actions, such as moving to a specified goal location. In between actions the robot will stand still, and execution time of an action will vary according to the distance to the goal etc.

An important feature of our software design is that both modes are supported through the same interface, which makes it easy to run the same code in simulation and on the real robot.

4.2. Overview

The software framework has three main components (see figure 10):

- The **back-end** communicates with the robot through the driver,

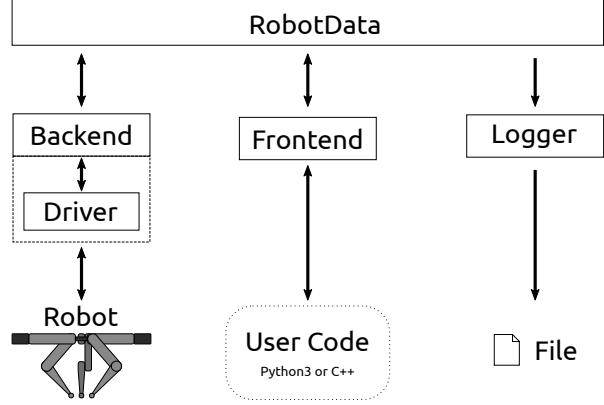


Figure 10. Software Architecture. The modules communicate through the RobotData and do not have any direct connections.

- the **front-end** allows the user to control the robot through C++ or Python3,
- the **logger** logs all the inputs and outputs of the robot.

Each of these components can run in a separate process, which is advantageous for computational reasons and it separates the back-end from the user code. Since all the commands sent to the robot flow through the back-end, we can implement checks which ensure safety no matter what happens in the user code.

In the following we will describe the front-end and the back-end. The back-end section is relevant for readers interested in setting up their own robot, users only need to know about the front-end.

4.3. Front-end

In our design, the only objects that exist from the user-perspective are

- a time-series of desired actions a computed by the user,
- a time-series of actions actually applied to the robot a' , which is identical to a except for potential modifications to satisfy safety constraints,
- and a time-series of observations y .

Figure 11 shows the temporal relations of these variables. The user may perform two operations: They may append actions to the desired-action time-series a and they may read from any of the three time-series a, a', y , where a', y are filled-in by the back-end.

The user interface implementation is equivalent to the fol-

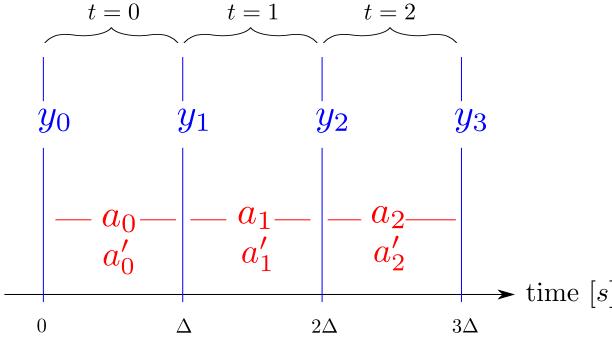


Figure 11. This figure shows the temporal sequence of variables. In real-time mode (4.1) there is a fixed control and observation rate Δ . Each observation y_t corresponds to a single instant in time $t\Delta$, while each action a_t corresponds to a time interval $[t\Delta, (t+1)\Delta)$.

lowing pseudo code:

```
def append_desired_action(x) :
    a ← (a, x)
    return len(a)-1
def get_observation(t) :
    wait until len(y)>t
    return yt
def get_desired_action(t) : as above
def get_applied_action(t) : as above
```

This interface provides access to a synchronized history of all the inputs and outputs of the robot. The user has complete freedom how to use this data and when to append actions, as long as they make sure to do so on time, before the action is needed by the robot. For instance, they may choose to run a typical real-time control loop where a new action a_t is computed periodically, or they may choose to compute entire action sequences at a lower rate and then append them in a burst by repeatedly calling `append_desired_action`.

If the user attempts to access a future observation through `get_observation(t)`, this function will wait and return as soon as this observations is acquired. For instance, in the case of real-time control (definition 4.1), if the call `get_observation(2)` is made at time $< 2\Delta$ seconds, the function will wait until the observation y_2 is received at time 2Δ seconds and then return. This feature allows for synchronization with the real system.

The function `append_desired_action(x)` will append action x to the action time-series. For convenience it returns the timeindex of the appended action, but if the user keeps track of timeindices externally, the return value can be ignored. The instant of the first call to `append_desired_action` marks time 0, after which the back-end will start filling in the a' , y timeseries and

expect the action timeseries a to be filled in by the user.

A basic control loop can be written as follows:

Example 4.1 (Basic control loop).

```
robot.append_desired_action(a0)
for t in (0,...,T):
    yt = robot.get_observation(t)
    at+1 = some_policy(yt)
    robot.append_desired_action(at+1)
```

No explicit wait is necessary, synchronization with the back-end is ensured through the call to `robot.get_observation(t)`, which will wait until the back-end has appended y_t .

This control loop is valid for both real-time (4.1) and non-real-time (4.2) control:

If the back-end is running in **real-time mode**, it will add observations y_t periodically and the loop above will hence run at a fixed rate, unless the call `some_policy` is too slow. If that is the case, the back-end will detect that the user did not append the next action on time and it will shut down the robot and raise an error. Alternatively, the back-end can be configured to simply repeat the previous action if the next action has not been appended on time.

In contrast, if the back-end is running in **non-real-time mode**, it will wait for the next action, and the function `append_desired_action` may be called with arbitrary delay. Similarly, the execution of actions may take varying amounts of time, and hence the call to `get_observation` will not return at a predetermined time. This mode makes sense e.g. for simulation, where the simulator and the policy add actions and observations whenever they are done with their respective computations.

If we wish to control the robot at a lower rate, this can also be implemented very easily:

Example 4.2 (Basic control loop at a reduced rate).

```
robot.append_desired_action(a0)
for t in (0,...,T):
    yt = robot.get_observation(t·k)
    at+1 = some_policy(yt)
    for _ in range(k)
        robot.append_desired_action(at+1)
```

Here, the control frequency is reduced by a factor of k . Note that the index t refers to the control cycle here, not the robot cycle.

4.4. Relation to the Standard Reinforcement Learning Framework

As we shall see, there is a gap between the standard Markov decision process (MDP) formulation of reinforcement learning and the presented framework. This is due to the realtime

constraints we face when working with real robots. Here we show how to close this gap, which allows us to e.g. wrap our interface into the OpenAI Gym interface.

Some Simplifications: For simplicity of exposition, we will only consider the case where the observations y_t are Markovian, the extension to the non-Markovian case is straightforward. In addition, we will not treat the applied actions a'_t explicitly anymore, as they do not affect the independence structure between the desired actions a_t (henceforth referred to simply as actions) and observations y_t . They can either be simply ignored or they can be added to the observations if one wishes to give the RL algorithm access to this information.

4.4.1. DEPENDENCE STRUCTURE OF A REAL-TIME SYSTEM

We can deduce the dependences between variables from their temporal ordering represented in figure 11. Naturally, any variable can only causally depend on variables which precede it in time. However, in real-time systems there is the additional constraint that the controller needs some time to compute actions, hence they may not depend on observations they coincide with in time. More precisely, action a_t cannot depend on y_t , since a_t starts at precisely the time when observation y_t is acquired. In fact, we already took this into account in example 4.1, where a_{t+1} is a function of y_t rather than y_{t+1} . These considerations lead to the Bayes net in figure 12.

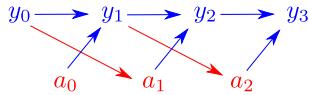


Figure 12. The Bayes net showing the dependence between variables in a real-time system.

net in figure 12. By comparing with the Bayes net of an MDP in figure 13 we see that we cannot simply identify the state s_t with y_t , even in the Markovian case. In the following, we propose two ways in which this structure can be mapped to a standard MDP.

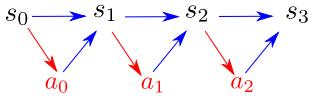


Figure 13. The dependence structure (Bayes net) of an MDP.

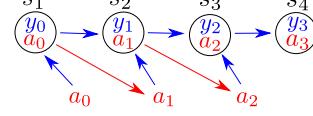


Figure 14. A Bayes net showing the dependence between variables obtained from figure 12 by defining the state as $s_{t+1} = (y_t, a_t)$. This yields the same dependency structure as in figure 13.

4.4.2. STATE AUGMENTATION

By defining the state in the slightly counter-intuitive way $s_{t+1} = (y_t, a_t)$, we retrieve the dependence structure of the MDP, see figure 14. This allows us to apply any of the RL algorithms which have been formulated for standard MDPs to our system. Ramstedt & Pal (2019) use a similar trick and propose a method to take the resulting structure into account when performing RL.

Given these considerations, we can now define an OpenAI Gym environment for our robot interface. For our discussion here, only the step function is relevant:

Example 4.3 (Gym Env - State Augmentation).

```
def step(a) :
    t = robot.append_desired_action(a)
    y = robot.get_observation(t)
    s' = (y, a)
    return s', some_other_stuff
```

Where the ' in s' denotes the subsequent time index.

This opens up the possibility of applying numerous implementations of RL algorithms out-of-the-box to our platform. For instance, we can directly apply the algorithms from (Hill et al., 2018), (garage contributors, 2019) and (Dhariwal et al., 2017), as we will demonstrate in the experimental section.

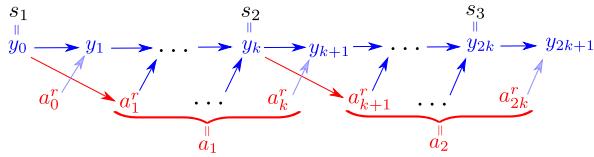
It is often the case that we want to operate at a control rate which is lower than the communication rate of the robot. For instance, in many tasks it is not necessary to control the proposed system at 1000 Hz, we may want to control it at e.g. 100 Hz. We can define an according OpenAI Gym environment.

Example 4.4 (Gym Env - State Augmentation at Reduced Rate). The logic here is exactly the same as in [example 4.3](#), except that each action is applied k times, where k is the rate reduction factor.

```
def step(a):
    t = robot.append_desired_action(a)
    for _ in range(k - 1):
        robot.append_desired_action(a)
    y = robot.get_observation(t)
    s' = (y, a)
    return s', some_other_stuff
```

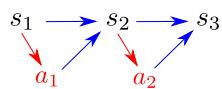
4.4.3. APPROXIMATE MAPPING FOR CONTROL AT LOWER RATES

In the case of low-rate control, we can also define an approximate mapping of the real-time system to a standard MDP which does not require state augmentation. This may in some cases be preferable, because it is more intuitive and it does not lead to an observation space of increased dimensionality. To see how this can be done, consider [figure 15](#). If we control at a rate reduced by a factor k , each control



[Figure 15](#). When we control at a reduced rate each control action a_t is applied k times and hence produces k robot actions $a_{(t-1)k+1:t k}^r$. To approximately map this to the dependence structure of a standard MDP ([figure 13](#)), we can ignore the light-blue arrows.

action a_t is applied k times and hence affects k observations. For k sufficiently large, a reasonable approximation is to ignore the dependence of the last of these k observations on a_t , see [figure 15](#). Doing so leads again to standard MDP dependencies, see [figure 16](#). We can now define again an according OpenAI Gym environment:



[Figure 16](#). This is the dependence structure of the states s and actions \bar{a} implied by [figure 15](#), ignoring the light-blue arrows.

Example 4.5 (Gym Env - Approximation at Reduced Rate).

```
def step(a):
    for _ in range(k):
        t = robot.append_desired_action(a)
    s' = robot.get_observation(t)
    return s', some_other_stuff
```

Here, the time index of the last appended robot action is t , and we return $s' = y_t$. Hence, all robot actions up to time index $t - 1$ have to be executed before we can retrieve this observation, and only then the step function will return. Nevertheless, the step function will have to be called again before robot action $t + 1$ can start. This means that the controller will only have one robot cycle to compute the next action, despite the lower control rate.

Summarizing, we can say that this approximation is a good solution if the computation of the next action is substantially faster than the control rate. Otherwise, the solution in [example 4.4](#) is preferable, since there the controller may use the entire control cycle to compute the next action.

4.5. Back-end and Driver

To integrate a new robot into our software framework, the only necessary steps are to define an `Action`, an `Observation`, and a `RobotDriver` class in C++ with the following functions:

```
RobotDriver
{
    Observation get_latest_observation();
    Action apply_action(Action a_t);
}
```

The first function is self-explanatory. The second function takes a desired action as input, it may perform safety checks, and then returns the applied action.

The back-end takes care of calling the right functions at the right times. It will read the desired actions, appended by the user to the timeseries a , and pass them to the `RobotDriver`, and it will fill the observation y and applied action a' timeseries with the outputs from the `RobotDriver`. For details we refer the interested reader to the documentation in the code.

4.5.1. IMPLEMENTATION OF THE TRIFINGER ROBOT

The implementations of the classes above for the TriFinger robot are:

- **Observation:** Joint position, velocity and torque of each joint. The camera images are retrieved through a separate interface with the same structure, since they arrive at a different rate.

- Action: Torque to be applied at each joint and optionally a joint position along with the gains of a PD controller. The driver will then sum the torque with the feedback from the user-specified position controller.
- RobotDriver: Here, the implementation of `get_latest_observation` returns the latest measurements which were received from the motorboards through CAN, see (Grimminger et al., 2020) for details. The `apply_action` makes sure that the joint velocity and torques do not become too large, since otherwise the robot may break or the motors may overheat. It then sends the modified action to the motorboards and returns it.

5. Experiments

The purpose of this section is not to improve the state-of-the-art in robotic manipulation, but rather to illustrate the capabilities of the proposed hardware and software. Each experiment highlights different aspects:

- [Section 5.1](#) demonstrates the 1 kHz real-time torque-control abilities.
- [Section 5.2](#) shows that the back-end safety features allow for deep reinforcement learning from scratch, without any safety checks on the user side. This experiment also shows that the hardware is robust against collisions, which will necessarily occur during the learning of manipulation tasks. In addition, this use case shows that the software interface allows for application of out-of-the-box implementations of deep RL methods.
- In [section 5.3](#), we perform a throwing experiment to show that the actuators allow for highly-dynamic motions.
- Finally, in [section 5.4](#) we show through demonstration experiments that the platform is capable of fine-manipulation.

Videos of all experiments are available at [footnote 1](#).

5.1. Optimal Control

Controlling robot interactions with the environment is challenging due to the unilateral nature of the contact constraints and the stiff behavior of contact forces. We tackle this problem, for picking up and moving a cube, with the control loop depicted in [figure 18](#).



Figure 18. Control loop at 1kHz with force optimization.

Center of Mass Wrench: The first step is to compute the wrench (force and moment) which need to be applied to the object to maintain it on the desired trajectory. We do so using a simple PD law

$$F_{\text{com}} = P\delta x_{\text{com}} + D\delta \dot{x}_{\text{com}} - m\vec{g} \quad (1)$$

$$M_{\text{com}} = P\delta q + D\delta \omega \quad (2)$$

which will compute a force F_{com} and moment M_{com} at the object center-of-mass, to correct for errors in position (δx_{com}), linear velocity ($\delta \dot{x}_{\text{com}}$), orientation (δq) and angular velocity ($\delta \omega$). P, D are the controller gains, \vec{g} is the gravity vector and m is the object mass.

Force Optimization: The next question is what forces the finger tips must apply to the object in order to achieve the desired wrench calculated above. Therefore, we formulate a quadratic program to find optimal distribution of contact forces:

$$F_{\text{tip}} = \arg \min_y \frac{1}{2} y^\top y \quad (3)$$

$$\text{subject to } Gy \leq h \quad (4)$$

$$Ay = \begin{pmatrix} F_{\text{com}} \\ M_{\text{com}} \end{pmatrix} \quad (5)$$

where the optimization variable y is the stack of contact forces at the three finger tips expressed in the local frame of the object. The equality constraint ensures that the tip forces produce the desired wrench on the object (A transforms tip forces to center-of-mass wrench given known contact locations, see e.g. (Murray, 2017)).

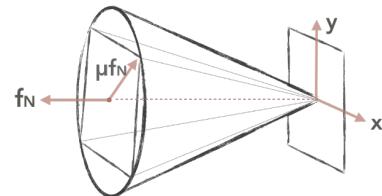


Figure 19. The friction cone ($|f_p| \leq \frac{\mu}{\sqrt{2}} f_n$) can be approximated using linear inequalities such that the approximation lies inside the actual cone. μ is the static friction coefficient and f_n, f_p are the force components normal and parallel to the contact surface, respectively.

The inequality constraint enforces that the tips can push but not pull, and it ensures that the object does not slip. It is a



Figure 17. TriFinger performing the pickup task.

linear approximation to the friction cone (see figure 19 and (Murray, 2017) for details).

Impedance Control: The final step is to compute the torques τ to be applied at the robot joints to produce the desired tip forces F_{tip} computed in the previous step. This can easily be achieved via Jacobian (J) transpose control. For stabilization, we also add a feedback on the fingertip position and velocity errors $\delta x_{\text{tip}}, \delta \dot{x}_{\text{tip}}$ (the desired tip trajectories are obtained given the desired object trajectory and known contact locations). This yields a simplified version of the impedance controller introduced by Hogan (1984)

$$\tau = J^T (F_{\text{tip}} + P' \delta x_{\text{tip}} + D' \delta \dot{x}_{\text{tip}}) \quad (6)$$

where P' and D' are hand-tuned controller gains.

5.1.1. RESULTS

We apply this methodology to two tasks: Lifting a cube 20 cm along a vertical line (see figure 17) and sliding that same cube in a circle along the table. As can be seen in the videos at footnote 1, the forces applied to the object are appropriate for moving it along the desired trajectory without slippage.

5.2. Reinforcement Learning

We illustrate the suitability of the platform for real-time reinforcement learning by training a DDPG (Lillicrap et al., 2015) agent from scratch on a reaching task, using the DDPG implementation from stable-baselines (Hill et al., 2018).

In this task, the goal is for each finger-tip to reach a randomly-sampled target location as accurately as possible. The episode length is set to 2 seconds. The observation space of the policy consists of joint positions, joint velocities, and target positions for each finger specified in task space. The action is the desired joint configuration of the fingers (the software back-end provides a PD controller, see section 4.5.1). The reward at each timestep is the negative Euclidean distance between the end-effector and the target. We train the system for 700 episodes, corresponding to 23 minutes of execution on the robot, plus a few minutes of computation time used by DDPG.

At the beginning of training, the fingers’ motions are jerky, and they often collide, see the video at footnote 1. As the

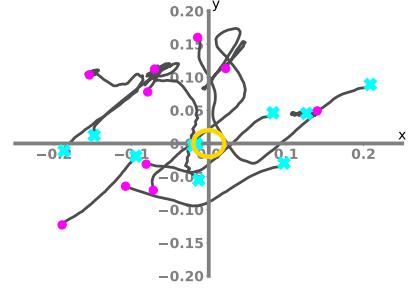


Figure 20. Beginning of training

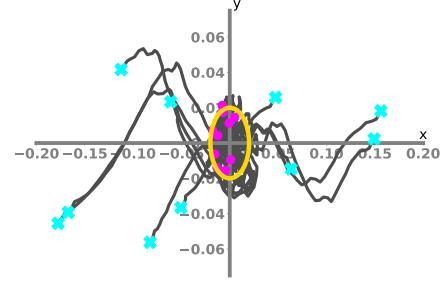


Figure 21. After 200 episodes

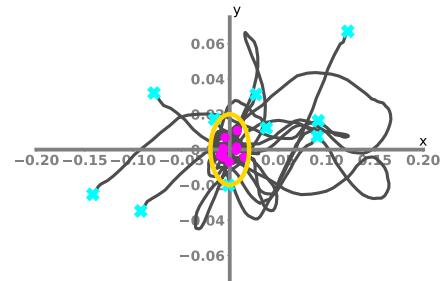


Figure 22. End of training

Figure 23. Trajectories of the finger tip relative to the goal position in the xy-plane. Each trajectory corresponds to one episode (start: cyan cross, end: pink dot). The yellow ellipse marks 2 cm position error on each axis.

training progresses, the motion becomes smoother and more accurate, see figure 23. At the end of training, the fingers

are able to reach the target positions consistently within an error of about 2cm.

5.3. Throwing

To showcase the ability of performing highly dynamic tasks, we execute throwing motions recorded through kinesthetic teaching (i.e. the motion was demonstrated by guiding the robot fingers). The videos at [footnote 1](#) show that the TriFinger is able to throw light objects several meters. We expect that using appropriate controllers, instead of kinesthetic teaching, one could improve considerably on these results.

5.4. Fine Manipulation

Finally, to illustrate the dexterity of the platform, we perform several fine manipulation motions, which, as above, were demonstrated through kinesthetic teaching. As can be seen in the videos at [footnote 1](#), the experiments include flipping a cube, turning it with one finger while the others hold it, balancing a flat cuboid on its side, and picking up a pen and drawing.

6. Conclusion

We presented an open-source robotic platform with novel hardware and software. We have shown its i) capabilities for dexterous manipulation, ii) suitability for deep RL from scratch thanks to the robustness of the hardware and safety checks of the software and iii) ease of use for both real-time optimal control as well as deep RL. We hope that these factors, in combination with the simple and inexpensive hardware design, will motivate many researcher to adopt this platform as a shared benchmark for real-world dexterous-manipulation. This could lead to a more coordinated effort among different labs and the generation of orders of magnitude more real-robot data than was possible thus far.

References

- IHMC Open Robotic Software. <https://github.com/ihmcrobotics>.
- Ahn, M., Zhu, H., Hartikainen, K., Ponte, H., Gupta, A., Levine, S., and Kumar, V. Robel: Robotics benchmarks for learning with low-cost robots. *arXiv preprint arXiv:1909.11639*, 2019.
- Amigoni, F., Bastianelli, E., Berghofer, J., Bonarini, A., Fontana, G., Hochgeschwender, N., Iocchi, L., Kraetzschmar, G., Lima, P., Matteucci, M., Miraldo, P., Nardi, D., and Schiaffonati, V. Competitions for Benchmarking: Task and Functionality Scoring Complete Performance Assessment. *IEEE robotics & automation magazine / IEEE Robotics & Automation Society*, 22(3):53–61, September 2015.
- Andrychowicz, O. M., Baker, B., Chociej, M., Jozefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.
- Behnke, S. Robot competitions-ideal benchmarks for robotics research. In *Proc. of IROS-2006 Workshop on Benchmarks in Robotics Research*. Institute of Electrical and Electronics Engineers (IEEE), 2006.
- Bonsignorio, F. and del Pobil, A. P. Toward Replicable and Measurable Robotics Research [From the Guest Editors]. *IEEE robotics & automation magazine / IEEE Robotics & Automation Society*, 22(3):32–35, September 2015.
- Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.
- Calli, B., Walsman, A., Singh, A., Srinivasa, S., Abbeel, P., and Dollar, A. M. Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set. *IEEE robotics & automation magazine / IEEE Robotics & Automation Society*, 22(3):36–52, September 2015a.
- Calli, B., Walsman, A., Singh, A., Srinivasa, S., Abbeel, P., and Dollar, A. M. Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols. February 2015b.
- Campbell, M., Hoane Jr, A. J., and Hsu, F.-h. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.
- Chitta, S., Marder-Eppstein, E., Meeussen, W., Pradeep, V., Rodríguez Tsouroukdissian, A., Bohren, J., Coleman, D., Magyar, B., Raiola, G., Lüdtke, M., and Fernández Perdomo, E. ros_control: A generic and simple control framework for ros. *The Journal of Open Source Software*, 2017. doi: 10.21105/joss.00456. URL <http://www.theoj.org/joss-papers/joss.00456/10.21105.joss.00456.pdf>.
- Coumans, E. and Bai, Y. Pybullet, a python module for physics simulation for games, robotics and machine learning. *GitHub repository*, 2016.
- Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S., Wu, Y., and Zhokhov, P. Openai baselines. <https://github.com/openai/baselines>, 2017.
- Dollar, A. M. and Howe, R. D. The highly adaptive sdm hand: Design and performance evaluation. *The international journal of robotics research*, 29(5):585–597, 2010.

- Duan, Y., Chen, X., Houthooft, R., Schulman, J., and Abbeel, P. Benchmarking deep reinforcement learning for continuous control. In *International Conference on Machine Learning*, pp. 1329–1338, 2016.
- Fujimoto, S., van Hoof, H., and Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. February 2018.
- garage contributors, T. Garage: A toolkit for reproducible reinforcement learning research. <https://github.com/rlworkgroup/garage>, 2019.
- Giftthaler, M., Neunert, M., Stäuble, M., and Buchli, J. The Control Toolbox - an open-source C++ library for robotics, optimal and model predictive control, May 2018.
- Grimminger, F., Meduri, A., Khadiv, M., Viereck, J., Wüthrich, M., Naveau, M., Berenz, V., Heim, S., Widmaier, F., Fiene, J., Badri-Spröwitz, A., and Righetti, L. An Open Torque-Controlled Modular Robot Architecture for Legged Locomotion Research. In *International Conference on Robotics and Automation (ICRA)*, 2020.
- Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. January 2018.
- Heess, N., Dhruva, T. B., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Ali Eslami, S. M., Riedmiller, M., and Silver, D. Emergence of Locomotion Behaviours in Rich Environments. July 2017.
- Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., and Meger, D. Deep reinforcement learning that matters. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- Hill, A., Raffin, A., Ernestus, M., Gleave, A., Kanervisto, A., Traore, R., Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S., and Wu, Y. Stable baselines. <https://github.com/hill-a/stable-baselines>, 2018.
- Hogan, N. Impedance control: An approach to manipulation. In *1984 American control conference*, pp. 304–313. IEEE, 1984.
- James, S., Wohlhart, P., Kalakrishnan, M., Kalashnikov, D., Irpan, A., Ibarz, J., Levine, S., Hadsell, R., and Boumalis, K. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12627–12637, 2019.
- Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., and Quillen, D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 37(4-5):421–436, 2018.
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.
- Lowrey, K., Kolev, S., Dao, J., Rajeswaran, A., and Todorov, E. Reinforcement learning for non-prehensile manipulation: Transfer from simulation to physical system. In *2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, pp. 35–42. IEEE, 2018.
- Mansard, N., Stasse, O., Evrard, P., and Kheddar, A. A versatile generalized inverted kinematics implementation for collaborative working humanoid robots: The stack of tasks. In *International Conference on Advanced Robotics (ICAR)*, pp. 119, June 2009. URL http://hal-lirmm.ccsd.cnrs.fr/file/index/docid/796736/filename/2009_icar_mansard-Stack_of_Tasks.pdf.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1928–1937, New York, New York, USA, 2016. PMLR.
- Murali, A., Chen, T., Alwala, K. V., Gandhi, D., Pinto, L., Gupta, S., and Gupta, A. PyRobot: An Open-source Robotics Framework for Research and Benchmarking. June 2019a.
- Murali, A., Chen, T., Alwala, K. V., Gandhi, D., Pinto, L., Gupta, S., and Gupta, A. Pyrobot: An open-source robotics framework for research and benchmarking. *arXiv preprint arXiv:1906.08236*, 2019b.
- Murray, R. M. *A mathematical introduction to robotic manipulation*. CRC press, 2017.
- Pickem, D., Glotfelter, P., Wang, L., Mote, M., Ames, A., Feron, E., and Egerstedt, M. The robotarium: A remotely accessible swarm robotics research testbed. In *2017 IEEE*

- International Conference on Robotics and Automation (ICRA)*, pp. 1699–1706. IEEE, 2017.
- Pinto, L. and Gupta, A. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 3406–3413. IEEE, 2016.
- Popov, I., Heess, N., Lillicrap, T., Hafner, R., Barth-Maron, G., Vecerik, M., Lampe, T., Tassa, Y., Erez, T., and Riedmiller, M. Data-efficient Deep Reinforcement Learning for Dexterous Manipulation. April 2017.
- Ramstedt, S. and Pal, C. Real-time reinforcement learning. In *Advances in Neural Information Processing Systems*, pp. 3067–3076, 2019.
- Schaal, S. The sl simulation and real-time control software package. Technical report, Los Angeles, CA, 2009. URL <http://www-clmc.usc.edu/publications/S/schaal-TRSL.pdf>. clmc.
- She, Y., Li, C., Cleary, J., and Su, H.-J. Design and fabrication of a soft robotic hand with embedded actuators and sensors. *Journal of Mechanisms and Robotics*, 7(2), 2015.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354–359, 2017.
- Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D. d. L., Budden, D., Abdolmaleki, A., Merel, J., Lefrancq, A., et al. Deepmind control suite. *arXiv preprint arXiv:1801.00690*, 2018.
- Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., and Abbeel, P. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 23–30. IEEE, 2017.
- Todorov, E., Erez, T., and Tassa, Y. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012.
- Xu, Z. and Todorov, E. Design of a highly biomimetic anthropomorphic robotic hand towards artificial limb regeneration. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3485–3492. IEEE, 2016.
- Yang, B., Zhang, J., Pong, V., Levine, S., and Jayaraman, D. Replab: A reproducible low-cost arm benchmark platform for robotic learning. *arXiv preprint arXiv:1905.07447*, 2019.
- Zamora, I., Lopez, N. G., Vilches, V. M., and Cordero, A. H. Extending the openai gym for robotics: a toolkit for reinforcement learning using ros and gazebo. *arXiv preprint arXiv:1608.05742*, 2016.
- Zhu, H., Yu, J., Gupta, A., Shah, D., Hartikainen, K., Singh, A., Kumar, V., and Levine, S. The ingredients of real world robotic reinforcement learning. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=rJe2syrtvS>.