DexDrummer: In-Hand, Contact-Rich, and Long-Horizon Dexterous Robot Drumming

Hung-Chieh Fang[1,2], Amber Xie[1], Jennifer Grannen[1], Dorsa Sadigh[1]

***Abstract*— Performing in-hand, contact-rich, and long-horizon dexterous manipulation remains an unsolved challenge in robotics. Prior hand dexterity works have considered each of these three challenges in isolation, yet do not combine these skills into a single, complex task. To further test the capabilities of dexterity, we propose drumming as a testbed for dexterous manipulation. Drumming naturally integrates all three challenges: it involves in-hand control for stabilizing and adjusting the drumstick with the fingers, contact-rich interaction through repeated striking of the drum surface, and long-horizon coordination when switching between drums and sustaining rhythmic play. Our key insight is leveraging contact-targeted rewards to address *in-hand contacts* (finger—stick) and *external contacts* (stick—drum). We instantiate this idea with DexDrummer, a dexterous drumming policy learned via reinforcement learning in simulation, with sim-to-real transfer for real-world drumming. DexDrummer leverages minimal hand priors to encourage stable in-hand contact; a trajectory reward and contact curriculum to mitigate the challenges with external contact; and a reactive grasp to support long-horizon playing. In simulation, we show our policy can play two styles of music: multi-drum, bimanual songs and challenging, technical exercises that require increased dexterity. Across simulated bimanual tasks, our dexterous, reactive policy outperforms a fixed grasp policy by 1.87x across easy songs and 1.22x across hard songs F1 scores. In real-world tasks, we show song performance across a multi-drum setup. DexDrummer is able to play our training song and its extended version with an F1 score of 1.0. Project website and videos: `https://sites.google.com/view/dexdrummer/`**

I. INTRODUCTION

Dexterous hand manipulation is an attractive problem in robotics because it unlocks a broad set of real-world tasks. Existing works have tackled challenges such as in-hand object reorientation [1], [2], grasping [3], [4], [5], [6], [7], [8], [9], and tool-based manipulation [10], all of which require managing complex finger–object interactions. While these works provide useful insights on dexterity, these studies typically emphasize short-horizon tasks or narrow aspects of dexterity in isolation.

In contrast, many real-world tasks such as assembly or cooking require dexterous skills that combine in-hand control, robustness to external perturbations, and long-horizon robustness. For example, assembling parts often involves reorienting a fastener in the hand while applying force to connect components, and cooking requires both holding utensils stably and stirring against resistance.

Motivated by the need for a compelling testbed, we propose drumming, a long-horizon, contact-rich dexterous manipulation task. Drumming inherently requires balancing in-hand control – maintaining and adjusting the grasp of the stick with fine finger control – and external contact – forcefully and repeatedly striking drums. To play long songs, this control becomes even more crucial: drumming requires a policy robust to these contacts for extended periods of time.

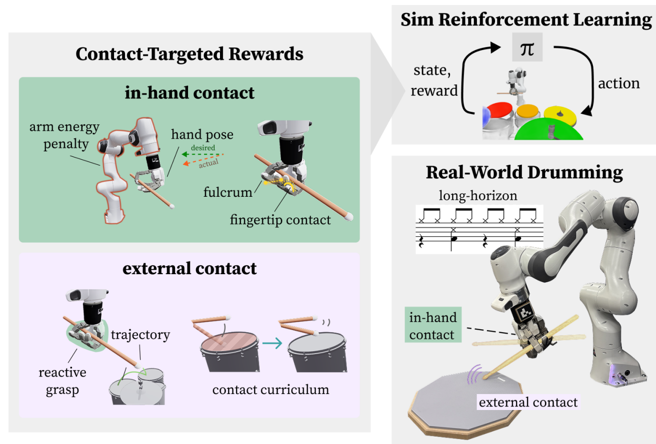

Fig. 1. **DexDrummer** is a dexterous, drum-playing RL policy that addresses the unified challenge of in-hand manipulation, external contacts, and long-horizon control by leveraging a contact-targeted reward. ***Left***: We decouple the contact challenges into two categories: (1) in-hand contacts which include fingertip contacts with the stick, a fulcrum incentive between the thumb and index finger, and an arm energy penalty to incentivize finger-biased control; and (2) external contacts such as a trajectory reward for controlled drum-hitting motions, a reactive grasp reward to prevent stick dropping, and a contact curriculum for gradually emphasizing drum playing. ***Right***: We demonstrate the efficacy of our drum playing policy both in simulation and real environments.

To address the unified challenges of in-hand control, external forces, and long-horizon robustness, our key insight is to leverage reinforcement learning with contact-targeted rewards. We denote two main categories of these rewards: *in-hand contact* and *external contact*. Inspired by in-hand manipulation and reorientation tasks which often require fine-grained finger and object control, our in-hand contact refers to finger and stick interactions, which requires fine-grained dexterity. Present in hammering or other tool-based manipulation tasks, external contact refers to the tool and environment interaction (in this case stick and drum), which includes initiating and reacting to the external forces. By encouraging robustness across both types of contact, we improve long-horizon control, which is essential to playing songs that require repeated drum hits.

We propose DexDrummer, which uses contact-targeted rewards to train a reinforcement learning policy in simula-

[1]Stanford University
[2]National Taiwan University

tion, and demonstrates real-world drumming capabilities. To address internal contacts over long horizons, we incorporate finger-based rewards to encourage fine-grained manipulation of the drum stick. To train a policy capable of both initiating hits and responding to external contacts, we introduce a guiding trajectory rewards, a contact curriculum and a reactive grasp. In simulation, we show DexDrummer performance on two styles of music: bimanual, multi-drum songs, and a uni-drum technical exercise. We show sim-to-real transfer for real-world drumming, showcasing DexDrummer's capability to perform dexterous, contact-rich control of the drum stick across a long horizon.

In summary, our main contributions are threefold:

1) We introduce drumming as a challenging testbed for dexterous manipulation, unifying the challenges of in-hand manipulation, external contacts, and long-horizon robustness.
2) We propose DexDrummer, a RL policy for drumming that leverages contact-targeted rewards to study the challenges of in-hand and external contacts combined in a single complex task.
3) We showcase simulated drumming across six different musical genres at two difficulty levels, with performances ranging from 20 to 40 seconds, as well as a high-speed exercise piece designed to test finger dexterity. In addition, we evaluate our method on a real-world drumming task using a cymbal and drum pad setup.

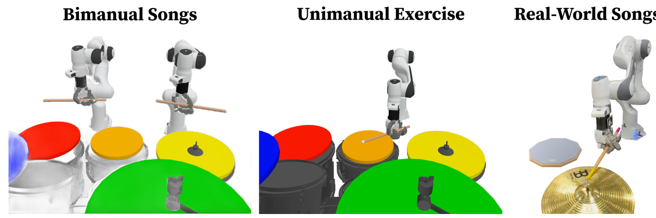

Fig. 2. **Drumming Environments.** Our first simulation environment includes bimanual, multi-drum song-playing. Our second environment involves unimanual, uni-drum control for a high-speed technical exercise. Finally, in the real-world, we play songs with a drum pad and cymbal.

II. Related Work

Dexterous Manipulation Dexterous manipulation has been a long-standing challenge in robotics. Many prior works have focused on tasks falling under these broad categories: in-hand manipulation, grasping, and post-grasp tool use. First, in-hand manipulation typically consists of rotating or translating objects with multi-finger control [11], [12], [13], [14], [15]. Next, dexterous grasping is a useful skill for automation or household tasks [4], [3], [16]. Then, grasping an object, the robot may place the object (pick-and-place) or use it as a tool, such as for drilling [17], pouring [10], wiping [18], scissor cutting [19], and more. While these tasks are challenging and important for researching dexterity, we propose drumming, a task that requires both in-hand control, post-grasp tool-based contact, and long-horizon robustness for playing long songs.

To learn dexterous policies, many classical methods use planning with precise models [20], [21], [22] for in-hand manipulation, and propose grasps based on collision detection or optimization-based methods [5], [23], [7], [24], [6], [25]. Other works use sim-to-real training with reinforcement learning [26], [2], [1], [27] for in-hand tasks, or use synthetically generated grasp datasets in simulation with learning-based methods can leverage [28], [29], [8], [4], [3], [16]. Imitation learning from human demonstrations has been used for many dexterous applications [30], [31], [10], [32], [9], [17]. However, to reduce the data burden on designing simulated environments or teleoperating dexterous data, many works retarget human data [33], [34], [19], which may also utilize reinforcement learning in simulation [35], [36], [37], [38], [39]. Following many prior works, we train an RL policy in sim, which reduces the need to hire expert drummers to generate demonstration data. We use simple sim-to-real techniques, such as domain randomization, to showcase real-world drumming.

Robot Drumming Robot drumming is an exciting subset of research in robotic music-playing. Many prior works design custom hardware for drum-playing [40], [41], [42], [43]. Our project assumes a more general embodiment – robot arm and hand – which requires learned dexterity to hit the drums, instead of special hardware designs. Recent works explore learning-based approaches to drumming, but likewise use either custom embodiments [44], or directly fix the stick to the embodiment in simulation [45], without examples of real-world drumming. Like [45], we learn our policy in simulation via reinforcement learning, but instead of focusing on humanoid control, we assume a realistic stick-holding embodiment, which allows sim-to-real transfer.

III. METHOD

We present DexDrummer, a bimanual, dexterous policy that can play drums in simulated and real environments. Following prior work [37], [2], we use reinforcement learning in simulation and use domain randomization to facilitate transfer to real-world scenes. However, to achieve dexterous movements under rich contact dynamics, we introduce contact-targeted rewards to address both in-hand and external contacts.

A. Problem Statement

We would like to learn a robot drumming policy via reinforcement learning. Under this framework, our objective is to learn a policy $\pi(a|s)$ that maximizes the expected discounted cumulative reward across trajectories. We assume an environment with observation o_t, r_t per timestep t, and discount factor γ.

In our drumming environment, we assume either a unimanual with $n_{\text{hand}} = 1$, or bimanual environment with $n_{\text{hand}} = 2$, where n_{hand} corresponds to the number of hands. We detail the policy inputs in table II and reward functions in table I.

TABLE I

Reward functions and curriculum. n_{contacts} is the number of fingertip contacts, $\mathbf{p}_{\text{thumb}}, \mathbf{p}_{\text{index}}, \mathbf{p}_{\text{stick}}$ are positions, $g(\cdot)$ is a shaping function from [46] mapping distances to $[0, 1]$, $\mathbf{x}_{\text{palm}}, \mathbf{z}_{\text{palm}}$ are palm orientation vectors, $\mathbf{v}_{\text{right}}, \mathbf{v}_{\text{down}}$ are world reference vectors, $\tau_{\text{arm\_joints}}, v_{\text{arm\_joints}}$ are arm torques and velocities, and $\mathbb{1}$ denotes indicator functions for grasp. *Contact Curr.* refers to contact curriculum. For the *exercise* setting, which mainly challenges finger dexterity, we disable the hand pose and grasping rewards, as the hand remains largely stable in this task.

Reward	Formula	Explanation	Weight (Song)	Weight (Exercise)
		In-Hand Contact		
Fingertip	$\exp\left(-1/(n_{\text{contacts}} + \varepsilon)\right)$	Make fingertips contact the stick	1.0	1.0
Fulcrum	$g\big((\lVert\mathbf{p}_{\text{thumb}} - \mathbf{p}_{\text{stick}}\rVert_2 + \lVert\mathbf{p}_{\text{index}} - \mathbf{p}_{\text{stick}}\rVert_2)/2\big)$	Position thumb and index finger to hold the fulcrum	0.0	1.0
Hand Pose	$(\mathbf{x}_{\text{palm}} \cdot \mathbf{v}_{\text{right}} + \mathbf{z}_{\text{palm}} \cdot \mathbf{v}_{\text{down}})/2$	Maintain a stable palm orientation in the world frame	0.5	0.0
Arm	$\lVert\tau_{\text{arm\_joints}}\rVert + \lVert v_{\text{arm\_joints}}\rVert$	Penalize arm movement to incentivize finger movement	0.0	-2.0
		External Contact		
Trajectory	$\mathbb{1}_{\text{is\_grasped}} \cdot g(\lVert\mathbf{p}_{\text{stick}} - \hat{\mathbf{p}}_{\text{stick}}\rVert_2)$	Guide the stick with a reference trajectory	1.5	2.0
Grasp	$\mathbb{1}_{\text{is\_grasped}}$	Check grasp with the palm-stick distance.	1.0	0.0
Contact Curr.	N/A	Disable stick-drum contact for first N steps	N/A	✓
		Task		
Drum Hit	$\mathbb{1}_{\text{is\_stick\_hit\_drum}} \cdot \mathbb{1}_{\text{is\_hit\_window}}$	Hit drum according to music	1.0	1.0

B. *Drum Environment*

We create a simulated drum environment in the ManiSkill framework [47] that consists of a bimanual robot setup and a full drum set (snare, tom, ride, hi-hat, and crash). In particular, this requires us to control and coordinate two arms and hands under a single policy, that can simultaneously play different drums.

We identify three main dexterous challenges in our drum environment. First, the robot must maintain *in-hand contact* to orient and grasp the stick firmly while performing. Second, the act of hitting each drum requires responding to *external contacts*. Third, we require *long-horizon control* to move between drums and respond robustly to in-hand and external contacts.

C. *Reward Design*

To address the different challenges in the drumming task, we propose reward design to guide policy learning. We categorize these rewards into 3 categories: (1) *In-Hand Contact Rewards* for finger—stick interaction (2) *External Contact Rewards* for stick—drum interaction and (3) *Task Rewards* for drum-playing.

In-Hand Contact Rewards In-hand control for the drum stick is paramount to drum playing. To enable this, we incorporate rewards targeting in-hand contact. First, we encourage finger-stick contact through a fingertip contact reward, a general reward function used in prior work [18] for finger-object interactions. Next, following human drum priors, we introduce the fulcrum reward, which specifically encourages the thumb and index finger to grasp the “fulcrum,” the center of the drum stick, following human priors in drumming. This further enhances finger-stick contact based on drumming priors. The fingertip contact reward encourages the fingers to successfully touch and grasp the object, which in this case, allows us to hold and manipulate the stick.

Moreover, unlike works with a fixed hand, drumming requires both arm and hand control, and we propose two rewards to further encourage in-hand contact in this arm and hand system. This is important, as arm control can either complicate or facilitate hand control. For example, if the arm moves excessively or to unnatural positions, in-hand contact may be difficult to maintain, but synergistic hand and arm movements may make in-hand stick control much easier. To address this issue, we design two complementary rewards. A hand-pose reward encourages the palm to face the drum set (left palm facing right and vice versa), guiding the arm to move to positions where the hand can better manipulate the stick. Next, we propose an arm penalty constraint, which reduces excessive arm movements, making in-hand contact more natural. This prevents the agent from manipulating the stick with arm movements, incentivizing the agent to develop fine-grained finger control. Energy minimization has been widely used in locomotion [48], [49], [50] to induce diverse gaits, and here, we adapt it for in-hand contact with the stick.

External Contact Rewards To address external contact, we divide the challenges into two stages: (1) initiating contact and (2) maintaining long-horizon contact. We address the first with a trajectory reward and contact curriculum, and we address the second with a reactive grasp reward.

First, we would like our policy to successfully initiate external contact between the stick and drum. To achieve this, we apply a trajectory reward, which explicitly guides the speed and motion, hence guiding precise control of the contact force. In our case, we pre-compute the desired trajectories of the drumstick tip and end by modeling drum hits with a sinusoidal wave and interpolating between drum positions across hits. This approach is inspired by prior work on high-level trajectory planning and low-level control [18], [51], [39], [37]. While this reward guides the hitting motion of the stick, the reward by itself is often insufficient to learn

TABLE II

OBSERVATION SPACE. L denotes lookahead horizon, and n_{hand} is 1 for unimanual and 2 for bimanual.

Observation	Dimension
Arm Proprioception	$7 \times n_{\text{hand}}$
Hand Proprioception	$20 \times n_{\text{hand}}$
Stick Head Proprioception	$3 \times n_{\text{hand}}$
Stick Tail Proprioception	$3 \times n_{\text{hand}}$
Trajectory Plan: Stick Head	$3 \times n_{\text{hand}} \times L$
Trajectory Plan: Stick Tail	$3 \times n_{\text{hand}} \times L$
Stick is Grasped	$1 \times n_{\text{hand}}$
Previously Played Drum	7 (discrete)
Next Drum to Play	7 (discrete)
Time Before Next Drum Hit	1

how to properly hit the drum. Specifically, during learning, the stick often rests on the drumhead, which often blocks the exploratory finger motions necessary for controlling the stick. To address this issue, we introduce a contact curriculum. Initially, contact between the stick and the drum is disabled, allowing the agent to practice trajectory following in free space while following the trajectory reward. Contact is later reintroduced, enabling the policy to more effectively initiate drum hits. This curriculum helps the policy learn to initiate external contact, as we decompose the problem into first following a motion, and then learning reactive behaviors to continue following the trajectory with the external forces. This curriculum shares similarities with DexMachina [52], which uses virtual object controllers to prevent early failures caused by gravity. In contrast, our curriculum targets contact-related randomness – a more challenging source of instability – and is simpler, requiring no modifications to object assets.

Lastly, we would like to repeatedly initiate external contacts for a long horizon. Notably, these external contacts can make in-hand control of the stick unstable. Thus, to enforce in-hand contact with the stick during external contact between the stick and drum, we propose a reactive grasp reward, as it adapts to dynamic interactions and provides stability across extended sequences of drum hits. Reward details are included in table I.

Task Rewards Similar to prior works [45], we add a sparse hit reward to check whether the drum is hit at a specified time.

IV. EXPERIMENTS

We seek to answer the following questions:

1) Is dexterity *essential* for achieving robust long-horizon drumming? Can we learn such a dexterous policy?
2) How can we enable finger-driven control for precise drumming?
3) How do our design decisions affect dexterous control?
4) Can the learned behaviors transfer to real world?

A. Experimental Setup

1) Hardware Setup: We use a 7-DOF Franka Panda arm and a 20-DOF Tesollo DG-5F hand in both simulated and real-world tasks (with bimanual setups in simulation). We run policy inference at 20 Hz, and we use a PID joint position controller that runs at 100 Hz.

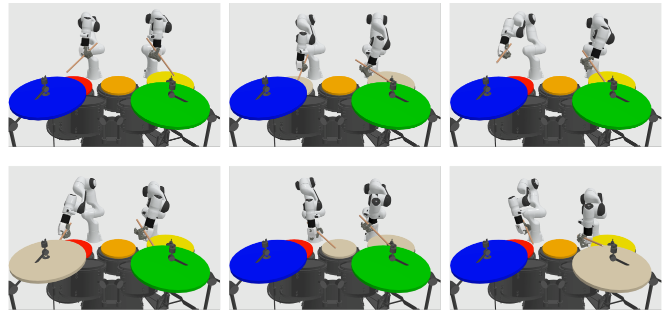

Fig. 3. **Bimanual Song-Playing Rollout.** We visualize 6 frames across a single song trajectory, with lighter colored drums and cymbals corresponding to a hit. Every song requires multiple combinations of drums to be hit.

We showcase real-world drumming with a unimanual, drum pad and cymbal setup. We create a digital twin of the real world by matching the drum, cymbal, and stick positions and sizes. We apply randomization to improve robustness for sim-to-real. At each step, uncorrelated Gaussian noise sampled from $\mathcal{N}(0, 0.05^2)$ is added independently to the proprioception and stick positions in the observation space. The stick's friction coefficient is perturbed with noise drawn from a uniform distribution $\mathcal{U}(-0.2, 0.2)$. In addition, control gains are scaled by a random factor sampled from $\mathcal{U}(0.9, 1.1)$ at environment initialization, and this factor is kept fixed for each environment throughout training.

To track the stick proprioception, we paint the end of the drumstick. We use color segmentation, depth readings from a RealSense camera, camera intrinsics, and camera extrinsics to project it into robot frame.

2) Policy Training: We train our policies with Proximal Policy Optimization [53], running for 60M steps on bimanual tasks and 40M steps on unimanual tasks. The training setup uses a discount factor of $\gamma = 0.8$, a clipping parameter of 0.2, and 1024 parallel environments. We adopt generalized advantage estimation (GAE) with $\lambda = 0.9$ and employ a 3-layer MLP policy network with hidden dimensions of size 512.

3) Evaluation Metrics: For bimanual song playing, we evaluate performance using two metrics: (1) the F1 score for song performance, and (2) the stick-hold ratio, defined as the fraction of time the stick remains held in hand over the total duration, which reflects the effectiveness of in-hand control. For dexterous fine control, we use (1) trajectory error, measuring how accurately the policy follows fine-grained trajectories, and (2) energy consumption, capturing the overall efficiency of the system.

4) Drum Songs: Like prior works [46], [45], we specify songs by importing MIDI files into the environment. MIDI is a widely used representation that encodes instruments and timing in a concise, discrete format. We obtain MIDI songs from an open-access website[1] . From this collection, we

[1] https://mididrumfiles.com/tag/midi-files/

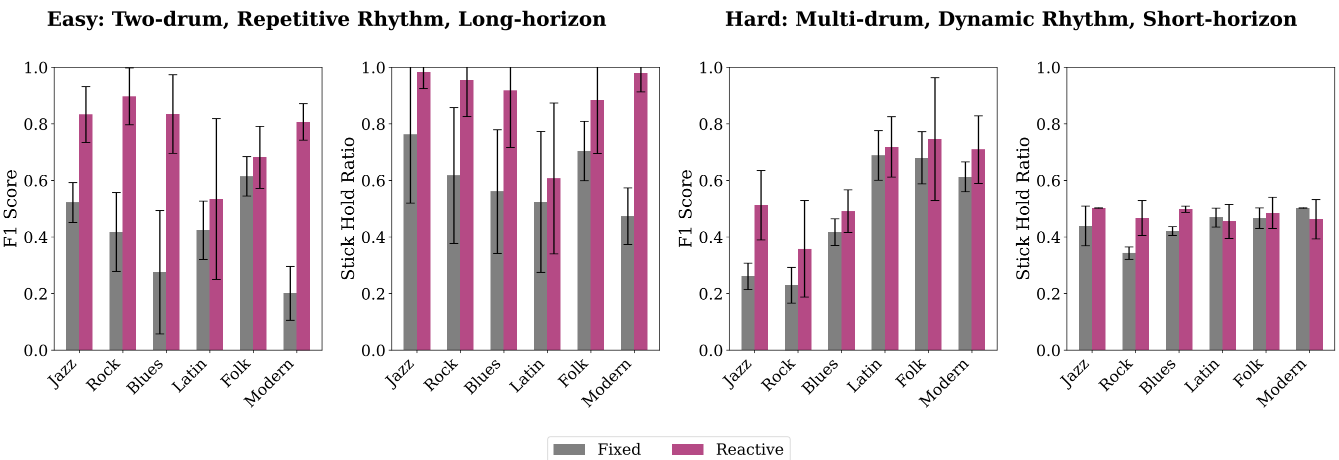

Fig. 4. **Results for Dexterous Song-Playing.** ***Left***: Reactive grasp outperforms fixed grasp by a large margin in long-horizon contacts. ***Right***: For more challenging songs requiring frequent drum-to-drum transitions, reactive grasp still improves performance, but with a smaller margin, primarily due to the reduced action space of fixed grasp.

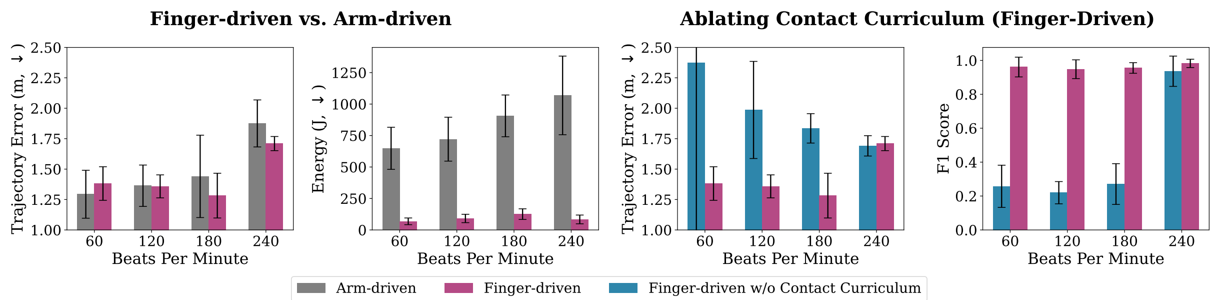

Fig. 5. **Results for Finger-Driven Control.** ***Left:*** As tempo (beats per minute) increases, trajectory error for finger-driven control decreases and the gap to arm-driven control widens, showing the superior dexterity of finger-driven control. Energy consumption is also substantially lower compared to arm-driven control. ***Right:*** Without contact-targeted rewards, finger-driven control struggles to manage contact interactions effectively.

select six different genres of music, and for each track, we extract both an Easy and Hard segment. To ensure playability for the robot policy, all songs are slowed down by a factor of three.

The Easy tracks feature repetitive loops, which we evaluate over 400 timesteps (20 sec) to test long-horizon control. The Hard tracks require hitting multiple drums with dynamic rhythms, and we evaluate these with 200 timesteps (10 sec). All tracks are performed using a bimanual setup.

For our Finger-Driven Control experiment (Section IV-C), we use an exercise that emphasizes finger dexterity by playing a single drum at a very high speed of up to 240 beats per minute (i.e., 4 hits per second). We train with 100 timesteps in a uni-manual setup.

B. Dexterity for Bimanual Song Playing

In our first set of experiments, we evaluate the importance of dexterity for long-horizon tool manipulation. For drumming, grasping the drum stick and repeatedly hitting a drum for extended periods of time likely necessitates reorientation and adjustment of the drum stick. To test the effectiveness of our dexterous policy, we compare *Fixed Grasp*, where the finger joints are frozen after reaching an initial grasp of the stick, to our method, *Reactive Grasp*, where the agent exerts dexterous control of the stick.

We evaluate the F1 Score and Hold Duration of the stick. The F1 Score represents how well the policy can play the song, showing that DexDrummer can learn a successful dexterous drumming policy. The Hold Duration evaluates how *Fixed Grasp* and *Reactive Grasp* are affected by slippage as the robot continues to move and hit the drum stick.

In fig. 4, we present two setups: an easier scenario with a repetitive loop but long horizon (*left*), and a more challenging scenario with multiple drums, dynamic rhythm, but shorter horizon (*right*). In the left case, the reactive grasp clearly outperforms the fixed grasp in both song performance and stick hold duration, highlighting the necessity of reactive, closed-loop dexterous control for long-horizon contact. In the right case, the reactive grasp still outperforms the fixed grasp, but with a much smaller margin. This is mainly because the fixed grasp only needs to learn arm motion within a lower-dimensional action space, which makes it easier to handle complex drum-to-drum transitions. The result shows a trade-off between dexterity and learning complexity, making it an interesting direction for balancing the two.

C. Finger-Driven Control

Next, we explore how to enable fine-grained, dexterous motions instead of relying on unnatural, whole-arm movements. Drumming can be attempted through arm movements or dexterous finger movements, but the precise finger control can often be overshadowed by initial arm exploration and movement. In this experiment, we evaluate how DexDrummer guides dexterity, and whether this leads to better performance of the song.

To evaluate dexterity, we choose an exercise that requires fine-grained finger movements. This exercise requires playing a single drum very rapidly, which is difficult for arm movements to follow and motivates dexterous finger control. To enable finger-driven control, we incorporate the arm energy penalty and contact curriculum, with reward terms listed in table I. The arm energy penalty limits arm movement in favor of finger movements, and the contact curriculum helps the finger-driven policy learn to handle external contact. The arm-driven policy optimizes the same reward terms, with the exception of the arm penalty and contact curriculum.

For this task, we evaluate the F1 Score, Trajectory Error, and Energy Consumption across a range of speeds for the exercise, denoted by Beats Per Minute. The F1 Score represents how well the policy plays the song. The Trajectory Error shows, more precisely, how well the drum stick can follow the desired trajectory. In particular, because this song requires greater rotation of the stick up and down, this captures how well arm or finger driven movements are able to reproduce this motion. Finally, high Energy Consumption implies extra, unnecessary movements, which is less desirable due to safety and sustainability concerns. We evaluate these metrics across a range of Beats Per Minute (BPM) for our exercise. A higher BPM implies quicker hits and less time in between hits, and is thus more challenging.

As shown in fig. 5, finger-driven control outperforms arm-driven control, particularly as the BPM increases. This highlights the necessity of fingers for fine-grained, high-speed motions, as arm-driven motion is unable to accurately replicate the quick stick movements. Moreover, finger-driven motion results in significantly lower energy consumption, which is advantageous for practical deployment. Qualitatively, in fig. 6, we find finger-driven motions to look more natural and human-like, whereas arm-driven motions are clunky and dangerous.

The fig. 5 (right) investigates what enables this dexterous finger movement. We ablate our finger-driven motion policy by removing the contact curriculum. Across different BPMs, removing the contact curriculum leads to much higher trajectory errors, implying that the curriculum is necessary to learn finger dexterity for following the stick trajectory. An exception occurs at a very high tempo (240 BPM), where the robot only needs to lift the stick slightly above the drum before hitting it again. Here, the effect of the contact curriculum is limited because the dexterous motion is small and easier to learn. This is further exemplified through the F1 scores, which shows a drastic gap when the contact curriculum is not enabled, besides for the fastest tempo (240 BPM). Qualitatively, in fig. 5, without the contact curriculum, the stick often rests on the drum head, preventing effective finger exploration. We hypothesize that because finger movements are relatively small, they can easily be canceled out by external contact.

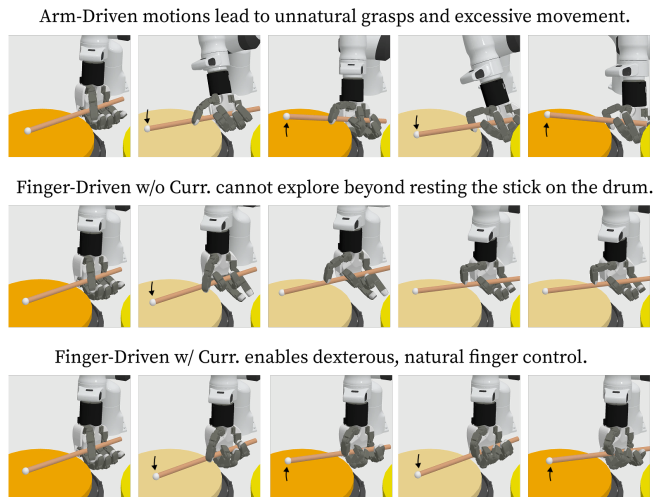

Fig. 6. **Sample Rollouts for the Dexterous Exercise.** For fast-playing exercises, the arm-driven policy moves to unnatural positions, leading to energy-intensive and potentially dangerous positions. Finger-driven control without the contact curriculum is unable to effectively learn dexterous control, whereas adding the contact curriculum leads to the most natural, effective drum hits. The lighter-colored drum head denotes that the drum is being hit by the stick.

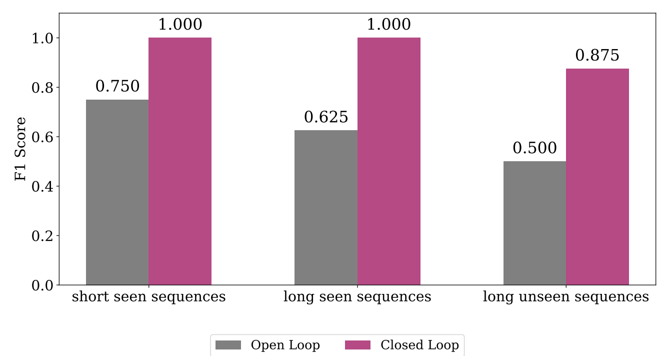

Fig. 7. **Results for Real-world Unimanual Drum Playing.** We denote the cymbal as c and the drum pad as d. The short seen sequence used for training is *ccdd*. The long seen sequence is *ccccdddd*, while the long unseen sequence is *ccddccdd*. Open-loop replay achieves reasonable performance but cannot adapt to real-world dynamics. Our policy is able to play unseen sequences (e.g., $d \rightarrow c$) by conditioning on the desired trajectory.

D. Real-World Drumming

Finally, we show results on real-world drumming with a drum pad and cymbal. We evaluate the effectiveness of our sim-to-real policy across multiple songs in an open and closed-loop setting, similar to [54]. First, we evaluate F1 performance on the seen song our reinforcement learning agent is trained on. Then, we evaluate on an extended version of the song, which should have hits and trajectories that are

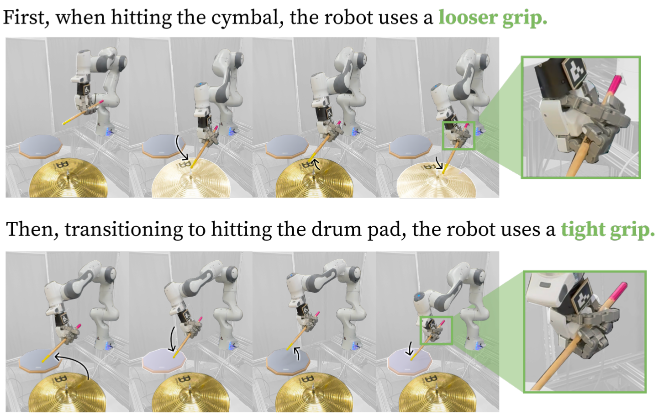

Fig. 8. **Real World Rollouts.** We visualize part of one closed-loop trajectory. The lighter colors for the cymbal and drum show when the robot hits them with the stick, and the arrows visualize the direction the stick was moving in. In this continuous rollout, we two cymbal hits and two drum hits. Notably, for the cymbal, which is not fixed and can rotate around the world vector, the robot strikes with a relatively loose grip (top). However, after hitting the drum pad, which is stable, the hand adjusts and forms a firmer grip (bottom), as shown by the movement of the index finger and thumb.

seen in the train song. Lastly, our most difficult song not only requires more hits than the train song, but it also includes unseen drum transitions. This is to evaluate the generalization of our policy to out-of-distribution songs. For our open-loop setting, we directly run the policy in simulation and replay actions, whereas for closed-loop song-playing, we run policy inference based on real-world states.

In fig. 7, we find that our closed-loop policy consistently outperforms the open-loop policy. This shows that closed-loop control is essential for real-world drumming, as the robot is able to react to stick movements to better play the song. Our closed-loop policy is able to play both the train song and an extended version of the train song with an F1 score of 1.0, showing that sim-to-real can enable effective drum playing. Additionally, the closed-loop policy can play songs with unseen drum transitions (e.g., from drum pad to cymbal), which may be due to high-level trajectory guidance that our policy is conditioned on table II.

In Fig. 8, we show a sample rollout from real-world drum playing. Notably, we highlight that the dynamics of hitting a cymbal and a drum pad differ: the cymbal is not fixed and involves less contact, whereas the drum pad is fixed and introduces more substantial contact. When striking the cymbal, the grip remains relatively loose since the interaction does not cause significant instability. In contrast, after striking the drum pad, the grip becomes firmer to handle the stronger contact. This adaptation shows the effectiveness of the reactive grasp.

V. CONCLUSION

We introduced DexDrummer, a testbed for dexterous manipulation that unifies the challenges of in-hand control, contact-rich interaction, and long-horizon tasks. To address this entangled problem, we proposed two categories of contact-targeted rewards: **in-hand contact rewards**, which incorporate hand priors to stabilize movements and sustain long-duration contact with the object, and **external contact rewards**, which guide contact initiation, and reacting to external perturbations across long horizons. We demonstrated the effectiveness of our framework through bimanual song playing and fine-grained control in simulation, as well as unimanual multi-drum performance on the real robot. In future work, we aim to extend these insights to broader problems that involve interaction with in-hand objects and the physical world.

Limitations and Future Work We are excited about future directions for dexterous drumming. For one, DexDrummer cannot play bimanual, multi-drum songs at human speed, and we slow down our songs in order to make it feasible. Future directions may explore how to play to drum tracks in real-time. Similarly, current experiments test performance for up to 400 timesteps, whereas real-world songs are often 3-5 minutes (up to 6000 steps). Improving speed, robustness, and control is paramount to improved song performances.

For our real-world experiments, we currently only show a uni-manual performance with two drums. Future work may showcase bimanual dexterous drumming with a full drum set. Furthermore, we do not incorporate sim-to-real techniques besides domain randomization, and further research into reducing the sim-to-real gap may lead to stronger song-playing.

REFERENCES

[1] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, "Solving rubik's cube with a robot hand," 2019. [Online]. Available: https://arxiv.org/abs/1910.07113

[2] J. Wang, Y. Yuan, H. Che, H. Qi, Y. Ma, J. Malik, and X. Wang, "Lessons from learning to spin "pens"," in *CoRL*, 2024.

[3] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," *arXiv preprint arXiv:2210.00722*, 2022.

[4] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "Dexgrasp anything: Towards universal robotic dexterous grasping with physics awareness," *arXiv preprint arXiv:2503.08257*, 2025.

[5] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[6] R. Wang, J. Zhang, J. Chen, Y. Xu, P. Li, T. Liu, and H. Wang, "Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation," *arXiv preprint arXiv:2210.02697*, 2022.

[7] S. Brahmbhatt, A. Handa, J. Hays, and D. Fox, "Contactgrasp: Functional multi-finger grasp synthesis from contact," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2386–2393.

[8] Y. Xu, W. Wan, J. Zhang, H. Liu, Z. Shan, H. Shen, R. Wang, H. Geng, Y. Weng, J. Chen *et al.*, "Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4737–4746.

[9] Y. Zhong, X. Huang, R. Li, C. Zhang, Z. Chen, T. Guan, F. Zeng, K. N. Lui, Y. Ye, Y. Liang, Y. Yang, and Y. Chen, "Dexgraspvla: A vision-language-action framework towards general dexterous grasping," 2025. [Online]. Available: https://arxiv.org/abs/2502.20900

[10] K. Shaw, Y. Li, J. Yang, M. K. Srirama, R. Liu, H. Xiong, R. Mendonca, and D. Pathak, "Bimanual dexterity for complex tasks," in *8th Annual Conference on Robot Learning*, 2024.

[11] A. Bhatt*, A. Sieler*, S. Puhlmann, and O. Brock, "Surprisingly robust in-hand manipulation: An empirical study," in *Robotics: Science and Systems XVII*, ser. RSS2021. Robotics: Science and Systems Foundation, Jul. 2021. [Online]. Available: http://dx.doi.org/10.15607/RSS.2021.XVII.089

[12] Y. Bai and C. K. Liu, "Dexterous manipulation using both palm and fingers," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1560–1565.

[13] J. Yin, H. Qi, J. Malik, J. Pikul, M. Yim, and T. Hellebrekers, "Learning in-hand translation using tactile skin with shear and normal force sensing," 2025. [Online]. Available: https://arxiv.org/abs/2407.07885

[14] Z.-H. Yin, B. Huang, Y. Qin, Q. Chen, and X. Wang, "Rotating without seeing: Towards in-hand dexterity through touch," 2023. [Online]. Available: https://arxiv.org/abs/2303.10880

[15] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik, "General in-hand object rotation with vision and touch," 2023. [Online]. Available: https://arxiv.org/abs/2309.09979

[16] T. G. W. Lum, A. H. Li, P. Culbertson, K. Srinivasan, A. D. Ames, M. Schwager, and J. Bohg, "Get a grip: Multi-finger grasp evaluation at scale enables robust sim-to-real transfer," 2024. [Online]. Available: https://arxiv.org/abs/2410.23701

[17] H.-S. Fang, B. Romero, Y. Xie, A. Hu, B.-R. Huang, J. Alvarez, M. Kim, G. Margolis, K. Anbarasu, M. Tomizuka, E. Adelson, and P. Agrawal, "Dexop: A device for robotic transfer of dexterous human manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2509.04441

[18] V. de Bakker, J. Hejna, T. G. W. Lum, O. Celik, A. Taranovic, D. Blessing, G. Neumann, J. Bohg, and D. Sadigh, "Scaffolding dexterous manipulation with vision-language models," 2025. [Online]. Available: https://arxiv.org/abs/2506.19212

[19] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "Dexcap: Scalable and portable mocap data collection system for dexterous manipulation," 2024. [Online]. Available: https://arxiv.org/abs/2403.07788

[20] L. Han and J. C. Trinkle, "Dextrous manipulation by rolling and finger gaiting," in *Proceedings. 1998 IEEE International Conference on Robotics and Automation (Cat. No. 98CH36146)*, vol. 1. IEEE, 1998, pp. 730–735.

[21] Y. Bai and C. K. Liu, "Dexterous manipulation using both palm and fingers," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1560–1565.

[22] I. Mordatch, Z. Popović, and E. Todorov, "Contact-invariant optimization for hand manipulation," in *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation*, 2012, pp. 137–144.

[23] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Deep differentiable grasp planner for high-dof grippers," *arXiv preprint arXiv:2002.01530*, 2020.

[24] T. Liu, Z. Liu, Z. Jiao, Y. Zhu, and S.-C. Zhu, "Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 470–477, 2021.

[25] N. C. Dafle, A. Rodriguez, R. Paolini, B. Tang, S. S. Srinivasa, M. Erdmann, M. T. Mason, I. Lundberg, H. Staab, and T. Fuhlbrigge, "Extrinsic dexterity: In-hand manipulation with external forces," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1578–1585.

[26] H. Qi, A. Kumar, R. Calandra, Y. Ma, and J. Malik, "In-hand object rotation via rapid motor adaptation," 2022. [Online]. Available: https://arxiv.org/abs/2210.04887

[27] M. Yang, C. Lu, A. Church, Y. Lin, C. Ford, H. Li, E. Psomopoulou, D. A. W. Barton, and N. F. Lepora, "Anyrotate: Gravity-invariant in-hand object rotation with sim-to-real touch," 2024. [Online]. Available: https://arxiv.org/abs/2405.07391

[28] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez, "Ganhand: Predicting human grasp affordances in multi-object scenes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5031–5041.

[29] C. Rosales, L. Ros, J. M. Porta, and R. Suárez, "Synthesizing grasp configurations with specified contact regions," *The International Journal of Robotics Research*, vol. 30, no. 4, pp. 431–443, 2011.

[30] S. P. Arunachalam, I. Güzey, S. Chintala, and L. Pinto, "Holo-dex: Teaching dexterity with immersive mixed reality," 2022. [Online]. Available: https://arxiv.org/abs/2210.06463

[31] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," in *Robotics: Science and Systems*, 2023.

[32] T. Lin, Y. Zhang, Q. Li, H. Qi, B. Yi, S. Levine, and J. Malik, "Learning visuotactile skills with two multifingered hands," 2024. [Online]. Available: https://arxiv.org/abs/2404.16823

[33] T. Tao, M. K. Srirama, J. J. Liu, K. Shaw, and D. Pathak, "Dexwild: Dexterous human interactions for in-the-wild robot policies," *Robotics: Science and Systems (RSS)*, 2025.

[34] M. Xu, H. Zhang, Y. Hou, Z. Xu, L. Fan, M. Veloso, and S. Song, "Dexumi: Using human hand as the universal manipulation interface for dexterous manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2505.21864

[35] A. Agarwal, S. Uppal, K. Shaw, and D. Pathak, "Dexterous functional grasping," in *7th Annual Conference on Robot Learning*, 2023.

[36] K. Shaw, S. Bahl, and D. Pathak, "Videodex: Learning dexterity from internet videos," *CoRL*, 2022.

[37] T. G. W. Lum, O. Y. Lee, C. K. Liu, and J. Bohg, "Crossing the human-robot embodiment gap with sim-to-real rl using one human demonstration," 2025. [Online]. Available: https://arxiv.org/abs/2504.12609

[38] T. Lin, K. Sachdev, L. Fan, J. Malik, and Y. Zhu, "Sim-to-real reinforcement learning for vision-based dexterous manipulation on humanoids," *arXiv:2502.20396*, 2025.

[39] Y. Chen, C. Wang, Y. Yang, and K. Liu, "Object-centric dexterous manipulation from human motion data," in *8th Annual Conference on Robot Learning*.

[40] A. Kapur, T. Singer, A. Suleman, and G. Tzanetakis, "A comparison of solenoid-based strategies for robotic drumming," 01 2007.

[41] S. Jordà, "Afasia: The ultimate homeric one-man-multimedia-band." 01 2002, pp. 132–137.

[42] Y. G. Kim, M. Garabini, J. Park, and A. Bicchi, "Drum stroke variation using variable stiffness actuators," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 3892–3897.

[43] E. Berdahl, B. Verplank, J. O. Smith III, and G. Niemeyer, "A physically intuitive haptic drumstick," in *ICMC*, 2007.

[44] S. M. Karbasi, A. R. Jensenius, R. I. Godoy, and J. Torresen, "Embodied intelligence for drumming: A reinforcement learning approach to drumming robots," in *Front Robot AI*, 2024.

[45] A. A. Shahid, F. Braghin, and R. Loris, "Robot drummer: Learning rhythmic skills for humanoid drumming," *arXiv preprint arXiv:2507.11498*, 2025.

[46] K. Zakka, P. Wu, L. Smith, N. Gileadi, T. Howell, X. B. Peng, S. Singh, Y. Tassa, P. Florence, A. Zeng, and P. Abbeel, "Robopianist: Dexterous piano playing with deep reinforcement learning," in *Conference on Robot Learning (CoRL)*, 2023.

[47] S. Tao, F. Xiang, A. Shukla, Y. Qin, X. Hinrichsen, X. Yuan, C. Bao, X. Lin, Y. Liu, T. kai Chan, Y. Gao, X. Li, T. Mu, N. Xiao, A. Gurha, V. N. Rajesh, Y. W. Choi, Y.-R. Chen, Z. Huang, R. Calandra, R. Chen, S. Luo, and H. Su, "Maniskill3: Gpu parallelized robotics simulation and rendering for generalizable embodied ai," *Robotics: Science and Systems*, 2025.

[48] J. Tan, T. Zhang, E. Coumans, A. Iscen, Y. Bai, D. Hafner, S. Bohez, and V. Vanhoucke, "Sim-to-real: Learning agile locomotion for quadruped robots," 2018. [Online]. Available: https://arxiv.org/abs/1804.10332

[49] X. Song, Y. Yang, K. Choromanski, K. Caluwaerts, W. Gao, C. Finn, and J. Tan, "Rapidly adaptable legged robots via evolutionary meta-learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 3769–3776.

[50] Z. Fu, A. Kumar, J. Malik, and D. Pathak, "Minimizing energy consumption leads to the emergence of gaits in legged robots," in *Conference on Robot Learning (CoRL)*, 2021.

[51] C. Qian, J. Urain, K. Zakka, and J. Peters, "Pianomime: Learning a generalist, dexterous piano player from internet demonstrations," in *8th Annual Conference on Robot Learning*, 2024. [Online]. Available: https://openreview.net/forum?id=t0LkF9JnVb

[52] Z. Mandi, Y. Hou, D. Fox, Y. Narang, A. Mandlekar, and S. Song, "Dexmachina: Functional retargeting for bimanual dexterous manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2505.24853

[53] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017. [Online]. Available: https://arxiv.org/abs/1707.06347

[54] "Learning to play piano in the real world," 2025.