

Surprisingly Robust In-Hand Manipulation: An Empirical Study

Aditya Bhatt* Adrian Sieler* Steffen Puhmann Oliver Brock

Abstract—We present in-hand manipulation skills on a dexterous, compliant, anthropomorphic hand. Even though these skills were derived in a simplistic manner, they exhibit surprising robustness to variations in shape, size, weight, and placement of the manipulated object. They are also very insensitive to variation of execution speeds, ranging from highly dynamic to quasi-static. The robustness of the skills leads to compositional properties that enable extended and robust manipulation programs. To explain the surprising robustness of the in-hand manipulation skills, we performed a detailed, empirical analysis of the skills’ performance. From this analysis, we identify three principles for skill design: 1) Exploiting the hardware’s innate ability to drive hard-to-model contact dynamics. 2) Taking actions to constrain these interactions, funneling the system into a narrow set of possibilities. 3) Composing such action sequences into complex manipulation programs. We believe that these principles constitute an important foundation for robust robotic in-hand manipulation, and possibly for manipulation in general.

I. INTRODUCTION

We present dexterous in-hand manipulation skills on a robotic hand (Figure 1) that substantially go beyond the state of the art in their robustness, generality, and fluidity of motion. As it is difficult to substantiate claims about real-world behavior in writing, we suggest that the interested reader watch the following video to form their own opinion: <https://youtu.be/Z6ECG3KHibI>.

The state of the art, at this time, is the groundbreaking work presented by OpenAI [1], who used Deep Reinforcement Learning to produce remarkably dexterous behavior on a five-fingered robotic hand, first manipulating a cube and later even an articulated Rubik’s cube [2]. Their learned skills feature contact-rich movements like finger-gaiting, pivoting, and the exploitation of gravity.

In contrast to OpenAI’s work, our approach uses a highly compliant hand and achieves such capabilities entirely without sensing, hand or object models, or machine learning. Yet, we demonstrate in-hand manipulation skills that

- transfer unmodified to objects of diverse shapes, weights, and sizes, e.g. cubes, cuboids, and cylinders, varying in size and weight by a factor of two,
- execute successfully and robustly when their execution speed is varied by a factor of 80,
- tolerate, without loss of performance, variations in translational object placement of over 90% of the object’s size and variations in rotational placement of 23° ,

*Both authors contributed equally to this work.

All authors are with the Robotics and Biology Laboratory, Technische Universität Berlin, Germany. We gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - EXC 2002/1 "Science of Intelligence" - project number 390523135, German Priority Program DFG-SPP 2100 "Soft Material Robotic Systems" - project number 405033880 and DFG project "Robotics-Specific Machine Learning (R-ML)" - project number 329426068.

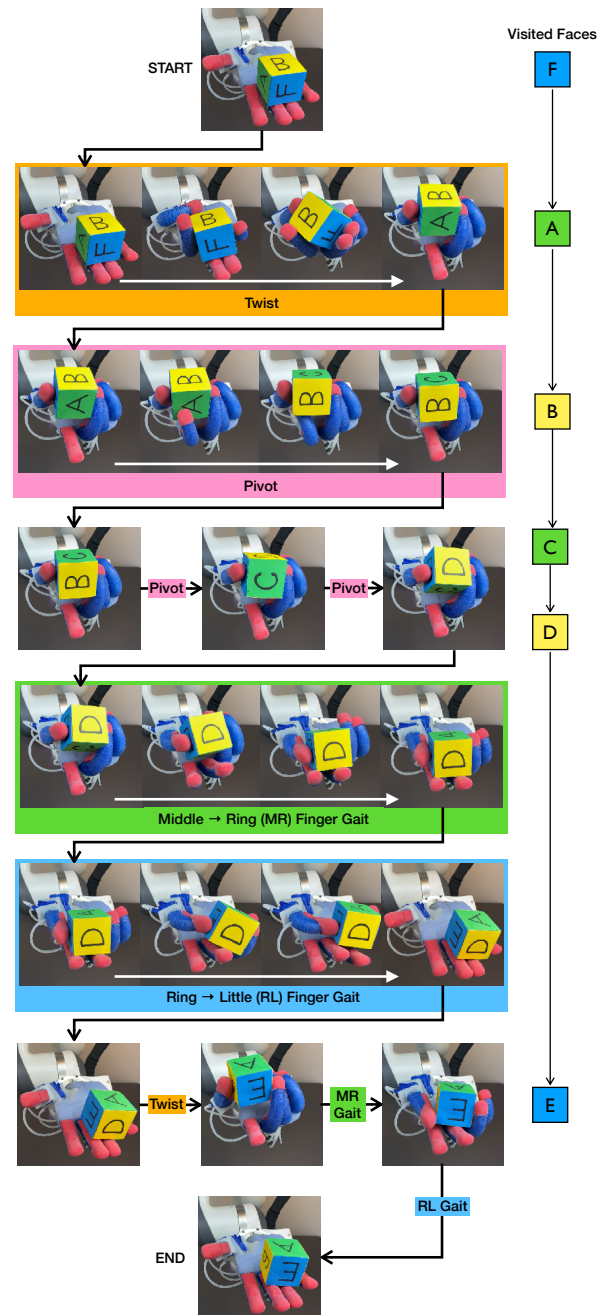


Fig. 1. **An open-loop in-hand manipulation program** that cycles through all cube faces in the order: $FABCDE$. This plan is composed entirely of just four skills executed at appropriate times. With these skills, we can achieve all possible face configurations ($24 = 6 \text{ faces} \times 4 \text{ orientations}$). The skills feature diverse contact types (sliding, pushing, and rolling) between the object and finger parts (dorsal, ventral, lateral, tip) and feature various capabilities like twisting, pivoting, finger-gaiting, and controlled dropping. Full video: <https://youtu.be/jRHibnE8Kfo>

- can reliably be composed into larger manipulation programs, letting us execute them consecutively up to 140 times,
- and do all of this with sensorless open-loop actuation.

Figure 1 shows the execution of such a program, composed of four repeatedly executed in-hand manipulation skills.

We demonstrate the existence of these remarkable generalization phenomena in real-world experiments. We then analyze the resulting behavior to identify, empirically, the reasons for this observed performance. Our key contribution is to illustrate *how* to achieve such robust manipulation behavior. Through our analysis, we hypothesize principles that lead to robust manipulation.

We believe these insights offer new opportunities for achieving human-level dexterity and facilitate the development of capable robotic manipulation systems. Much like convolution is a useful inductive bias for deep learning in computer vision, the exploitation of these principles may serve as a powerful inductive bias for Deep RL in robotic manipulation.

II. DESIGNING IN-HAND MANIPULATION SKILLS

In order to study the problem of in-hand manipulation from first principles, we tried the simplest possible approach. We observed ourselves rotating a cube in our human hands, and attempted to replicate this behavior with a robotic hand by manually actuating its parts. This effort resulted in a skill that was easy to design, yet surprisingly advanced in its robustness and manipulation complexity. Before we explain how we designed the skill, we must first introduce the hardware we used, as it plays a salient role.

A. The RBO Hand 3

To explore in-hand manipulation, we used the compliant anthropomorphic RBO Hand 3 [3] (Figure 2). Its silicone fingers are pneumatically actuated, soft continuum actuators with two air chambers each. The hand additionally possesses a highly dexterous opposable thumb, consisting of a stack of differently oriented soft bellow actuators made of coated nylon fabric. Additional bellows inside the hand can spread the fingers wide and flex the palm inwards. In total, the hand has 16 actuators and virtually unlimited degrees of freedom due to its flexible morphology.

We actuate the hand by controlling the air-mass enclosed in each actuator [4]. These inflation levels can be commanded by a computer or through a re-tasked electronic mixing board (commonly used in music synthesis).

B. Coding In-Hand Manipulation Skills

The first behavior we chose to replicate, after observing our own hands, was to repeatedly rotate a cube counter-clockwise. To faithfully mimic this behavior on the robot hand, we did not restrict ourselves to the common practice of fingertip-only contacts. As we shall see later, this is crucial to the success of our approach.

We placed the cube at a specific location on the area spanned by the ring and middle fingers. Using a mixing board, we

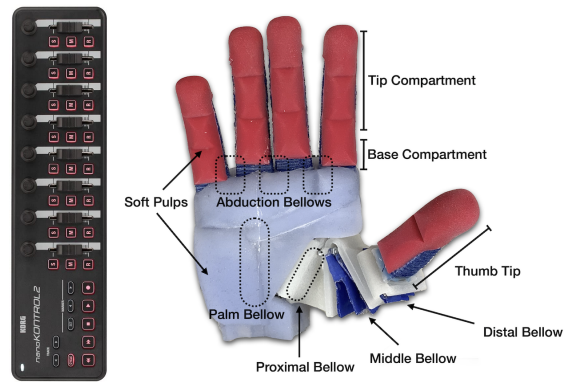


Fig. 2. **The RBO Hand 3**, an anthropomorphic, pneumatically actuated, soft robotic hand with 16 degrees of actuation. It is highly compliant to support the diverse mechanical interactions encountered during manipulation. Soft silicone paddings, called pulps (red and white), cover the fingers and palm. An easy way to manually command hand postures is with a mixing board (pictured left); each slider maps to one actuator.

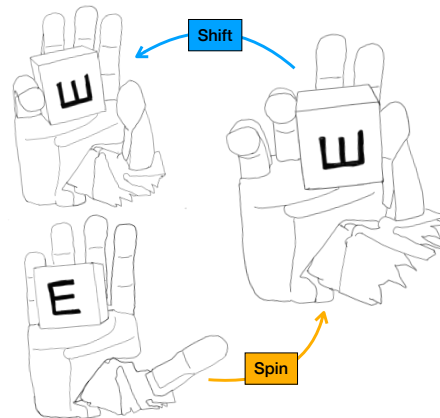


Fig. 3. **Start and end states of the spin and shift skills.** We replicated this behavior on the robotic hand by observing ourselves perform the same behavior.

manually adjusted the actuators' inflations and observed the resulting hand-object interactions. With this process, we were able to produce a similar counter-clockwise rotation.

Next, we captured these commanded inflation levels into a sequence of *keyframes*, corresponding to the enclosed air-masses in all actuators $\mathbf{a} = \{a^1, a^2, \dots, a^{16}\}$. We then replayed this sequence onto the hand by linearly interpolating through the intermediate keyframes.

Without the visual feedback we used when actuating the hand interactively, we expected the rotation to fail during automated replay. Instead, it worked! With additional fine-tuning of the keyframes, we were able to reliably spin the cube by 90° each time and (with another sequence of keyframes) shift it back to its starting location as illustrated in Figure 3. Designed in this manner, the *spin* and *shift* skills are open-loop interpolations through a total of 13 hand-crafted keyframes (Figure 5), depicted visually in Figure 4. In the same way, we designed three additional skills (depicted in Figure 1)

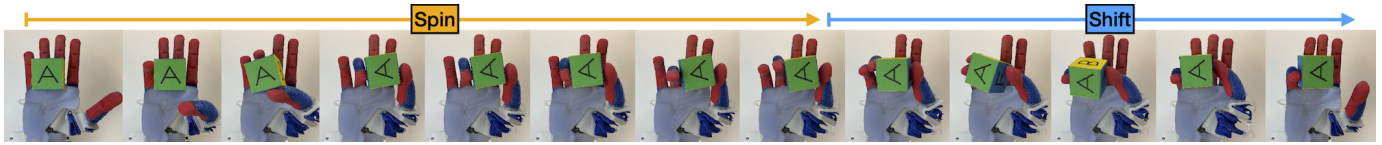


Fig. 4. All keyframes of the designed *spin* and *shift* skills (Viewed from above): *Spin* uses the thumb and ring finger to rotate an object counter-clockwise. *Shift* gaits the cube from the ring finger to the little finger, moving it to the left.

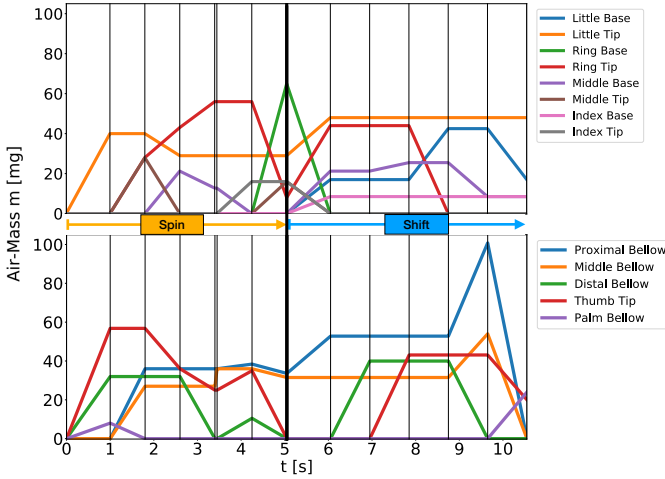


Fig. 5. Air-mass actuation signal for the *spin* + *shift* skill. Each vertical line corresponds to an intermediate keyframe. Appendix A contains a detailed description of each of these keyframes.

that work together to reorient a cube to all possible face configurations. We summarize all of these:

- **Spin** uses the thumb to rotate the cube outer-clockwise by 90° around the ring-finger, and places it to rest on the index and middle finger. It does so with 7 keyframes.
- **Shift**, also called **Ring→Little (RL) Finger Gait**, gaits the cube from the ring-finger to the little finger, and places it to rest on the ring and middle finger (5 keyframes).
- **Twist** uses the thumb, middle, and ring fingers to lift the cube into a precision grip, in the process rotating the cube counter-clockwise by 90° (9 keyframes).
- **Pivot** maintains this posture, contacting the cube with the index finger to rotate it by 90° around the grip axis (6 keyframes).
- **Middle→Ring (MR) Finger Gait**, a variant of *shift*, gaits the cube from this posture to a grasp between the thumb and middle finger, and places it to rest on the index and ring fingers (6 keyframes).

The corresponding air-mass actuation sequences for these particular three skills can be found in the Appendix A.

The main point of this section is to illustrate the simplicity of the programmed in-hand manipulation skills: they are open-loop sequences of keyframes, programmed using human intuition. One might call this approach simplistic. Still, the resulting skills are highly robust. In the next section, we present and analyze experimental evidence for this robustness.

III. EVIDENCE OF SURPRISING ROBUSTNESS

We now experimentally characterize the robustness of the hand-programmed skills. We report results for the *spin* + *shift* skill, as well as for the combined *twist* + *pivot* + *MR-gait* + *shift*¹ skill. The sequencing of these particular skills allows us to restore the manipulated object — if executed successfully — to a configuration that can be handled by the *twist* as well as the *spin* skill.

A. Robustness to Object Placement

Due to its sensorless design, it is reasonable to expect the skill to only work for a specifically chosen initial cube pose. However, this is not the case. <https://youtu.be/-gzeFdHbvIM> shows successful and unsuccessful trials of *spin* + *shift* with different initial placements. Surprisingly, the skill works robustly for a wide set of reasonable poses. We can identify regions of the cube’s pose space which would lead to failure. The *twist* + *pivot* + *MR-gait* + *shift* skill is similarly robust (video: <https://youtu.be/gzvRXcdolpc>). In the next section, we will investigate why.

B. Generalizing to New Objects

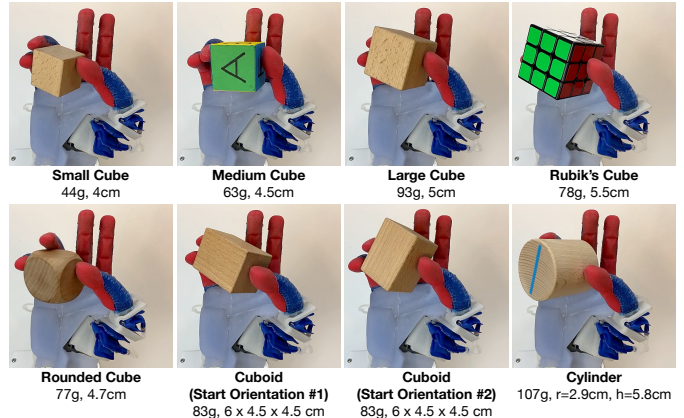


Fig. 6. *Spin* + *shift* successfully manipulating seven objects of varying shape, size, weight, and surface properties, using the same open-loop actuation signal. The sizes of the cubes and the cuboid refer to the edge length(s). Cylinder - d represents the diameter and h the height.

Figure 6 and its corresponding video (<https://youtu.be/TejM-kfllVE>) show the same *spin* + *shift* skill executing successfully with seven different objects: variously sized cubes,

¹The *RL-gait* (*shift*) skill is a slightly updated version of the one used in *spin* + *shift*. Nevertheless, all manipulations referenced in the subsequent videos for the combined *twist* + *pivot* + *MR-gait* + *shift* skill are performed with the exact same open-loop actuation signal.

a cuboid, and even a cylinder. Having designed the skill using only two different cubes, we were surprised by this immediate generalization across shapes, sizes, and weights. This observation is not restricted to the *spin + shift* skill. The *twist + pivot + MR-gait + shift* skill exhibits similar transfer across objects (video: <https://youtu.be/NNf5UBgQQZk>).

The actuation signal is the same in all cases, so clearly our skills cannot manipulate every possible object (e.g. *spin + shift* fails to rotate pyramids or spheres), but the straightforward generalization we observe makes detailed knowledge of objects’ properties unnecessary.

Why, then, do we see this object-adaptive behavior? We will discuss this in a later section.

C. Robustness to Execution Speed

As designed, *spin + shift* executes for approximately 11 seconds. However, this duration is arbitrary. We recorded executions with a range of different speeds; a video (<https://youtu.be/JU6BlpaqjVA>) shows the results. The slowest executions we tested ran for over 1 minute, while the fastest reliable runs took only 0.8 seconds, suggesting a temporal scaling tolerance of at least 80. This degree of robustness to timing changes was surprising: the skill had only ever been tested at the same speed of about 11 seconds, and the dynamics of the manipulandum and the relative influence of gravity vary significantly over these different time scales. The more complex *twist + pivot + MR-gait + shift* skill was designed to execute over a duration of 21 seconds. We found it robust to execution time-scales as low as 7.5 seconds, and as high as 5.5 minutes. (video: <https://youtu.be/-T6DMi6veE>).

D. Composing Longer Manipulations

The *spin + shift* skill leaves the object in approximately the same position at which it starts. Having observed the skill’s robustness to variability in initial positions, we looped its execution. In our demonstration (video: <https://youtu.be/6OAWxvhpfDQ>), we were able to spin a Rubik’s cube 140 times before it falls out due to an unexpected slippage. The fact that this is possible without any kind of explicit feedback control makes it all the more surprising. Again, this observation is not limited to the *spin + shift* skill. <https://youtu.be/jr3UxtbFCs> shows the *twist + pivot + MR-gait + shift* skill repeatedly manipulating a 4.5cm cube 54 consecutive times.

IV. EMPIRICAL ANALYSIS OF ROBUSTNESS

Having established that these skills are highly robust, we now attempt to identify the reasons for this robustness by empirical analysis. Again, due to space constraints, and the impracticality of motion-capture for the other skills, we only analyze the *spin + shift* skill. The analyses of the other skills should lead to identical conclusions.

A. Robustness During Manipulation

To understand the high reliability of our sequenceable skills, we must look closely at how these keyframe transitions affect

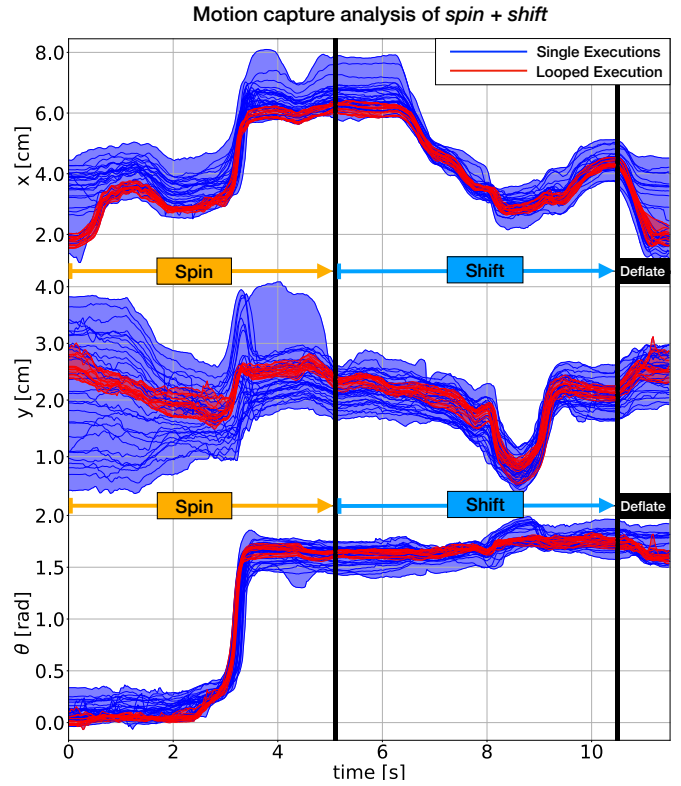


Fig. 7. **Quantitative evidence of robustifying disturbance rejection**, through tracking of the (x, y, θ) pose of a wooden cube, while executing the *spin + shift* skill. The blue curves are 33 independent trials. The 15 red curves are traced by a single episode of uninterrupted looped execution. After each execution episode, we deflate all compartments to rule out any drift in the low-level air-mass controller.

the object. We placed the hand in a 16-camera motion capture system and tracked a 4.5cm wooden cube’s pose in the (x, y) plane coincident with the palm. The outstretched fingers point along the y -axis (like in Figure 8).

As our manipulation skills are open-loop, the most straightforward way to quantify their robustness is to execute the skill for different initial (x, y) positions, thereby characterizing the region of success. Figure 7 shows all 33 successful trajectories of the *spin + shift* skill in blue. Outside of the observed initial range of 3.5cm, 4.5cm, and 23° along x , y , and θ , the manipulation fails. This indicates a tolerance margin of over 90% of the cube’s own width along the y axis.

More interesting than the skill’s “range of robust applicability” is *why* the manipulation is robust. Within the envelope of successful trials, we see these independent blue trajectories being funneled from a wide range of initial positions into a dense inner trough as the manipulation executes. This suggests that the manipulation achieves robustness by actively restricting the cube’s position to a narrow range of possibilities.

We can test this idea by repeatedly looping the same skill, without intervention, to see if the trajectories remain restricted. Figure 7 shows these 15 looped executions (plotted in red) over the blue envelope. Despite unintended slippages, imprecise contact locations, and sensorless actuation, the cube keeps

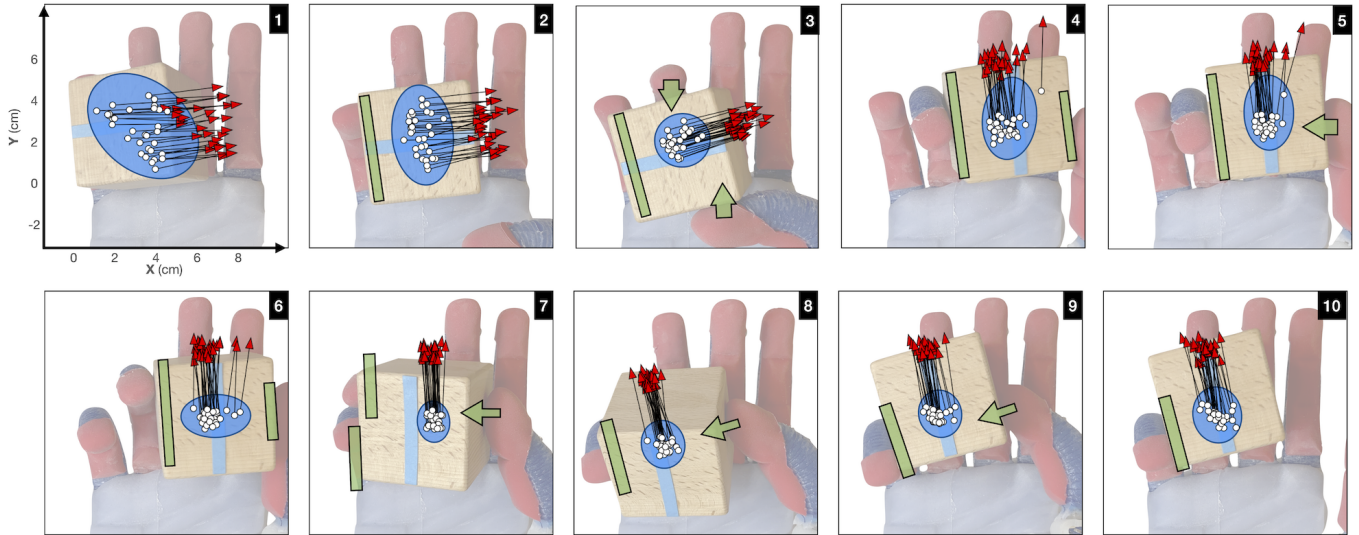


Fig. 8. **Robust manipulation as a series of uncertainty-restricting constraint exploitations:** The plots depict a keyframe-by-keyframe breakdown of cube poses (x, y, θ) gathered over 33 independent trials of *spin + shift*. The underlaid photographs are sourced from only the most salient keyframes of one illustrative execution. A white dot marks the cube position from a single execution, and the associated arrow indicates the planar rotation θ . The blue region coarsely indicates the set of observed cube positions over all trials. The green bars represent physical walls implemented by the fingers, and the green arrows represent active pushing interactions.

orbiting in a narrow basin of attraction without deviating.

We conclude that the robustness is somehow achieved via attractor dynamics that reject disturbances. Using the same trajectory data, we will now study *how* this robustifying uncertainty-restriction works, keyframe-by-keyframe.

B. The Building Blocks of a Robust Skill

Using the same trajectory data, we now study *how* this robustifying uncertainty-restriction works at the level of each keyframe. Each plot in Figure 8 depicts the (x, y, θ) pose of the cube after commanding a keyframe. We now describe what happens in the keyframes, referring to the figure.

The cubes start out with a large scattering in initial positions. In keyframe 2 (KF2), the hand raises the little finger to ensure the cube’s placement near the middle-ring area. This is a robust operation that restricts the cube’s horizontal scattering. In KF3, the ring finger and thumb tips act as a constraint to actively clamp the cube from either side, now squashing their vertical scattering. Clearly, this cannot succeed if the thumb cannot move behind the cube as shown. The skill would fail if the cube were too large or positioned entirely on the palm. It would also fail if the cube were placed too close to the fingertips.

So far, we have inspected translational manipulations. During KF3→KF4, we strongly curl the ring finger inwards while uncurling the thumb. This greatly increases the staggered opposing forces, causing a rapid rotation that could increase uncertainty. However, in KF4 the cube’s poses remain well-restricted over the index-middle area, since the thumb and ring finger now act as constraining walls to the left and right of the cube. Additionally, the ring finger’s presence shifts the cube *exactly one finger-width* to the right. The rapid motion still

increases the uncertainty vertically, as there are no guards in that direction, other than friction.

In KF5, the thumb moves, pushing the cube against the curved ring finger. This aligns them, robustly squashing rotational uncertainty. This is an instance of a well-understood class of parts-orienting sensorless movements, previously studied in domains other than in-hand manipulation [5].

The remaining keyframes belong to the *shift* skill; they apply a sequence of the same push-to-align operations. These constraints, while enforced by different fingers at different times, remain similar. The constraints are also mobile, due to the fingers’ intrinsic lateral compliance. Using them, *shift* performs a guarded translation by one finger-width to the left.

We see that *spin+shift* is robust because it is itself composed of a series of highly robust primitive manipulations. These primitives work well, because they constrain the set of possible outcomes using physical constraints, produced and carefully positioned by the hand.

C. Limits of the Spin and Shift Movement

Having investigated the robustness of our skill, we now discuss the ways in which it can fail. A collection of failure cases can be found in this video: (<https://youtu.be/roLVk5pKQX8>)

To successfully trigger a rotation in the *spin* skill, the object must first be firmly clamped between the thumb and ring-finger, as in KF3 (Figure 8). The skill cannot be executed if this precondition is not satisfied, which could happen for the following reasons.

Positional and Orientational Limits

- The skill begins executing with the object sitting on the little finger, because raising this finger would push the object into the palm.

- The object starts out resting very close to the fingertips, making it difficult for the ring finger to curl and guard the cube from falling out.
- The object rests mainly on the palmar region so the thumb cannot move behind it.
- The object is placed close to the index finger, in which case the thumb is unable to block the object from falling off the right side.
- The object is not oriented in a way that allows the thumb and ring finger to make contact on opposing sides. Therefore, the spin transition cannot be executed.

Shape and Size limits

- The object is too big, obstructing finger motion, preventing the establishment of constraints
- The object is too small for the thumb to make contact
- The object is too heavy, bending the compliant fingers, changing the constraints provided by them

Similar limits apply to the *shift* skill. A collection of failures due to initial object placement for the *twist + pivot + MR-gait + shift* skill can be found in this video: (<https://youtu.be/w2-NNi3Nnro>).

V. PRINCIPLES FOR ROBUST MANIPULATION

Having studied these skills' inner workings, we now discuss our findings and identify three design principles for achieving dexterous in-hand manipulation. We will see that these principles are not new, in fact, they have been known for decades. We will thus also speculate on why the progress reported here became possible only now. The subsequent section will then discuss the related work that is also based on the three principles identified below.

A. Outsourcing Control of Contact Dynamics

We exploit the deeply researched insight that compliant mechanisms can free us from several challenging computational problems [6, 7]. We describe two of these ways below.

The most straightforward advantage of compliant hardware is its tendency to adapt to an object's shape, due to which simple open-loop control (e.g. closing all fingers together) is sufficient to grasp unknown objects; an ability that is nontrivial to achieve with fully actuated hands [8, 9, 10, 11]. In this paper, we see the effect of shape adaptation in Figure 6, where *spin + shift* manipulates 7 different objects of various shapes, sizes, and weights using the same actuation signal. We therefore conclude that compliant shape adaptation is useful not only in grasping, but also in fine manipulation.

Less obviously, soft compliant hardware frees us from the unfulfillable need to accurately model contact interactions. We explain this with an example.

Figure 9 shows a within-grasp manipulation with a rolling contact (best seen in a video: <https://youtu.be/Nor2QEtM4W8>) This manipulation consists of just two keyframes. In the first, we hold a wooden donut in a precision grip between the thumb and the index finger. With actuation, the donut twists stably without losing contact. Despite actuating only the thumb mechanism, we see that the index finger moves significantly,

remaining diametrically opposed to the thumb. Its motion is *solely* due to mechanical compliance and contact friction, which in this case is crucial for the motion's stability.

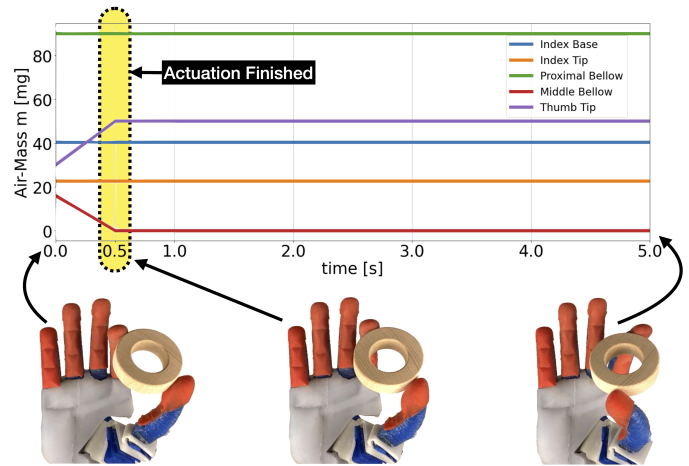


Fig. 9. **The hand twists a wooden donut in a stable rolling pinch grasp, without any feedback control.** Between 0 and 0.5 seconds, we actuate the thumb tip and the distal bellow. The actuation is finished long before the majority of the manipulation unfolds, showing the temporally-extended nature of the invoked dynamic response of the hand. This interaction is fully reversible by restoring the original inflation levels.

Note that the manipulation of the donut continues long after the actuation has completed at 0.5s. We argue that the hardware itself executes an object-adaptive *closed-loop* manipulation primitive. We never compute this explicitly; the process plays out entirely as a result of the hand's morphology, in the sense of *morphological computation* [12, 13]. The demonstrated contact dynamics are prohibitively complex to model analytically, but nevertheless perform very robustly in the real world. This manipulation behavior also transfers to different objects, and can be executed in reverse, as shown in this video: <https://youtu.be/6yiSLq59yFA>. Notably, the speed of motion varies for differently sized objects, and also at different intermediate postures.

This approach is closely related to another well-known control framework. Our air-mass actuation scheme resembles commanding a *virtual pose* in the sense of impedance control [14]. In impedance control, linear spring-like behavior encoded in a control law generates interaction forces whenever the hand/fingers deviate from their free-motion poses [15]. In our case, we do not design a control law but merely expect the actuator's elastic properties to evoke such a response. For example, the amount of air enclosed in a finger uniquely determines its virtual pose as well as a stiffness profile. On the RBO Hand 3 platform, the combination of 16 actuators lets us cycle through a vast number of virtual poses, each of which recruits different dynamic responses of the hand when in contact with an object. This computation-free alternative to managing mechanical interactions was suggested three decades ago by Hogan [14]:

“Exploiting the intrinsic properties of mechanical hardware can also provide a simple, effective and reliable way of dealing

with mechanical interaction.”

This forms our first design principle: we can identify and exploit useful and robust hand-object interactions, and compose them into dexterous manipulations. We lose the ability to command arbitrary interaction dynamics. What we gain in exchange is a greatly simplified control problem, with several substantial benefits, as explained above.

B. Constraining Object Motion

The dynamic response of the hand can produce complex behaviors when invoked by actuation, but at the expense of precise object motion. It is difficult to guarantee, for example, a specific object pose as a result of the interaction. In the absence of such certainty, we can instead restrict how the dynamics unfold. This idea is central to the study of *constraint exploitation* [16, 17], and has been used to ease manipulation with the aid of pre-existing environmental constraints such as walls, corners, and fixtures [18, 19]. In our skills, we also use constraints as tools to make manipulation work, but these are not external environmental features.

We explain this with the *spin + shift* example from Figure 8. During the KF3→KF4 transition, we trigger a very dynamic interaction: a rapid spin of the cube. The reason this transition does not spin the cube more than 90°, or send it flying out of the hand, is that KF4 blocks the cube with the ring finger and thumb on either side. Constraining the dynamics of the manipulandum in this way makes the resulting behavior highly regular, assuring a specific outcome with great certainty.

Each keyframe transition in our skills is robust manipulation in its own right. All of these primitives offload the fine details of interactions to the hand’s compliant morphology, and recruit physical constraints from that same morphology to prevent undesirable outcomes.

This forms our second design principle: the morphology itself serves as a platform for establishing *constraints* on the dynamics of the hand/object interaction. Under this framework, the skills we can design are limited by the available constraints, and we lose the ability to command arbitrary object poses. In practice, this need not be a problem; by actuating a sufficiently complex morphology, we can generate a combinatorially large number of possible constraints, letting us robustly achieve diverse object configurations.

C. Composing Manipulation Funnels

As Figure 8 shows, our open-loop primitives restrict uncertainty in object location by geometrically squashing, stretching, and moving the set of possible placements. Restricting uncertainty through mechanical action is the very idea of *manipulation funnels*, pioneered by Mason [20] over three decades ago. We quote from the paper:

“Using a funnel, the goal to position an object can be accomplished despite variation in the initial locations and shapes of the objects.”

As our primitives also restrict uncertainty in this way, we regard our primitives as funnels. The set of states from which a funnel can be executed is called its *entrance*. Upon

execution, it is guaranteed to map these states to another set called the *exit*. The power of funnels lies in their potential for composition [17, 21, 22]. If the exit of one funnel lies in the entrance of another, then by executing both funnels, we are guaranteed to end up in the second funnel’s exit. Importantly, such a composition is also a funnel in its own right.

This constitutes our third design principle: by sequencing several of these primitive funnels, we can design long complex manipulation funnels like *spin* and *shift*. Some funnels have exits that lie in their own entrances, e.g. *spin+shift* as well as *twist + pivot + MR-gait + shift*; we can loop their execution indefinitely.

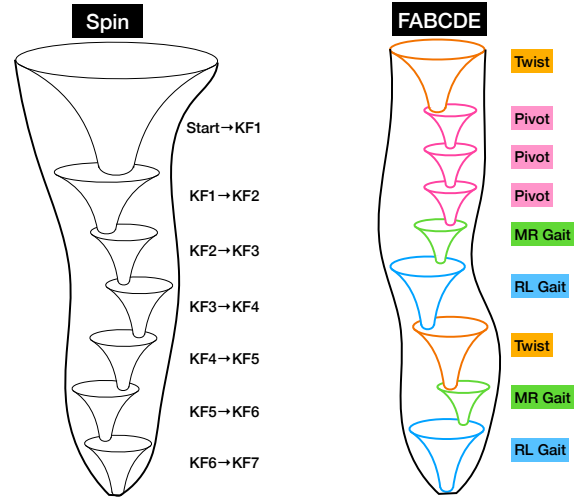


Fig. 10. **Robust manipulation through funnel composition:** Like *spin*, each skill is composed of funnel-like robust primitive manipulations, so it is a robust funnel in itself. Each funnel reliably transforms a set of hand-object configurations (its entrance) into another set (its exit). We designed our skills to funnel into each other, letting us compose longer manipulation plans like FABCDE.

We have designed all of our dexterous skills as funnel compositions, letting us incorporate them into long manipulation plans like the cube reorientation program in Figure 1. Figure 10 informally depicts the funnel composition used in the program.

In our final demonstration, we rely on the cross-object generalization of our chainable skills to manipulate multiple lettered objects in one open-loop execution, to spell a word (video: <https://youtu.be/tnq0xXMUbhc>). This demo combines sequences of all five discussed skills, as well as open-loop grasping and placement skills involving a 7-DoF Franka Emika Panda robot arm. The Appendix A contains a detailed description of the funnels used in this demonstration.

D. Why Now?

The three principles identified above are known for decades, begging the question why the progress presented here was not accomplished decades ago. We believe there are several reasons for this.

While the principles were known in their respective disciplines, they were not identified as *the* key ingredients. As a re-

sult, few researchers made it their focus to combine them into a single approach. Instead, our community debated and explored the advantages and limitations of analytical approaches versus data-driven approaches. In retrospect, analytical approaches mostly ignored the fact that we may not need accurate models for robust manipulation. (Note, however, that our insights do not call into question the value of those analytical models.) And data-driven approaches usually tackled problems that were much more complicated than necessary, because they largely ignored the inductive biases provided by the three principles.

The dependence on capable hardware posed another obstacle for progress in the past. In this study, we demonstrate that the most suitable realization of the three principles requires the careful co-design of hand morphology and associated control strategies [23]. Only when the implementation of the three design principles is orchestrated across both hardware and software, can it achieve its full potential. The advent and success of soft robotics as a research discipline has produced many insights necessary for the design of hardware sufficiently competent for the full realization of the three principles [3]. Centrally among them is the ability of soft hands to establish safe contact over large surface areas, as opposed to the point-based contact models underlying most analytic approaches to manipulation.

The explanations offered above can be illustrated by returning to the work of OpenAI, mentioned in the beginning of this paper [2, 1]. The work pursued a strongly data-centric approach, based on domain randomization. This approach forwent the advantages of outsourcing the handling of contact dynamics to clever hardware, did not leverage the inductive bias represented by the exploitation of physical constraints provided by the hand, and did not explicitly incorporate a compositionality prior *à la* funnels. In some sense, it attempted something much harder than what the problem required. We are excited to see the power of data-driven approaches and, in particular, deep reinforcement learning joined with the inductive biases presented here.

VI. RELATED WORK

The principles we described here are prevalent in past manipulation research. It is impossible to do justice to the many outstanding works that implemented them. Still, we structure our discussion of related work based on these principles. We also discuss research on analytical and data-driven approaches, as we feel these are two important categories of ongoing manipulation research.

A. Outsourcing Control of Contact Dynamics

Appropriately designed hardware can simplify the control problem for interaction tasks that require handling of difficult contact dynamics. Two prominent examples are the passive dynamic walker [6] and the universal robotic gripper [7]. Outsourcing aspects of control and perception to hardware is at the very core of the growing field of soft robotics [24, 25]. In the domain of robotic manipulation compliant hands like

the Pisa/IIT SoftHand 2 [11], the RBO Hand 2 [10], the i-Hy Hand [26] as well as the dexterous gripper presented in [27] have shown to be well suited for grasping and in-hand manipulation. Besides hands, the dynamic properties of a system composed of a soft arm with a soft gripper can be used to robustly solve various interaction tasks [28].

B. Constraining Object Motion

Deliberate contacts with the environment can be utilized to constrain the motion of the object to be manipulated. By following this principle, even simple gripper-like systems can attain comprehensive manipulation skills [29, 30]. Humans employ similar strategies when grasping. The observed environmental constraints can be used to execute robust grasping strategies with a soft robotic hand [18]. Environmental constraints can also simplify the learning of peg-insertion tasks [19]. The concept of constraint exploitation has been researched for in-hand manipulation with an underactuated gripper. This time the end-effector itself provides the constraints to the objects motion [31].

C. Using Funnels and Their Compositionality

A funnel, as a primitive to reduce uncertainty, is a well-suited abstraction to solve various tasks. The composition of funnels follows a hybrid nature; executing a funnel represents the continuous aspect while switching between funnels is by definition a discrete operation. The roots of funnels for manipulation planning can be traced back to Mason's *Manipulation Funnel* [20] and pre-image back chaining [16]. Robust motion strategies for orienting planar objects [17] and pushing [32] that reduce uncertainty purely through mechanical interaction are an effective way of outsourcing computation if the task mechanics can be described. This approach has been complemented to obtain the respective motion strategies from observational data [21]. Mason's funnel metaphor and the pre-image back chaining by Lozano-Pérez, Mason, and Taylor were extended to also cover sensory feedback to switch between funnels and/or feedback-controllers that can act robustly in the presence of uncertainties. Examples include paddle juggling [22], move-to-grasp [33], real-time motion planning [34], and robot grasping [35].

D. Analytic Approaches

The main platforms for in-hand manipulation research in the last decades have been fully-actuated [36] or hyper-actuated robot hands [37] with tightly integrated sensory devices, where every degree of freedom is controllable. Different analytic approaches to manipulation using these systems are, for example, impedance control [38], trajectory optimization [39], and hybrid position/force-control [40, 41]. Most of the approaches depend on explicit models of contact, mostly point contacts, which is not always a realistic assumption in real-world applications.

We discussed the data-driven approach by OpenAI in subsection V-D. Data-driven approaches to grasping [42] and in-hand manipulation [43, 44] have gained substantial popularity in recent years. They often do not make constraining assumptions about contact types or models, thus holding the promise of delivering more general manipulation behavior. These approaches often require massive computational resources and produce strategies encoded in black-box representations. Learning-based methods can complement the capabilities of analytic approaches, e.g. to obtain impedance gains for grasping and within-grasp manipulation [45], to hierarchically structure a cube manipulation problem [46], or to leverage dynamic movement primitives for within-grasp manipulation [47].

VII. CONCLUSION

We attempted to create the simplest possible in-hand manipulation skill on the RBO Hand 3, as a starting point for further investigations. What we thought to be the simplest approach, also turned out to be surprisingly competent and robust. So robust, in fact, that we think the skills presented in this paper substantially advance the state of the art in in-hand manipulation. In the paper, we substantiate this claim with extensive experimental evidence. This evidence shows how the skills transfer unmodified to objects of diverse shapes, weights, and sizes. It shows that the execution speed of the identical skill can be changed 80-fold without loss in performance, that the skills are robust to substantial disruptions in initial conditions, and that they can be composed into complex in-hand manipulation programs. Our surprise about these results is amplified by the fact that none of the skills we describe here depends on explicit sensing or feedback control.

The near-accidental discovery of very robust in-hand manipulation skills triggered an extensive empirical investigation into the causes of the surprising robustness. Through this investigation, we identified three well-known principles as the underlying pillars of robust in-hand manipulation (and probably of robust manipulation in general). First, it is important to outsource the control of complex and difficult-to-model contact dynamics to the hand's morphology. Second, the exploitation of constraints to limit the manipulandum's motion leads to significant reduction of uncertainty and thus to simplification of perception and control. Such constraints can arise from the hand's morphology itself or can be offered by the environment. Third, by viewing the exploitation of constraints as manipulation funnels, we can exploit their compositionality to produce complex manipulation programs. We believe that the insights presented in this paper offer new opportunities for robotic manipulation systems, in the quest of achieving human-level dexterity.

APPENDIX

The supplementary material can be found at this URL:
<http://dx.doi.org/10.14279/depositonce-12071>

- [1] M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.
- [2] 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>
- [3] Anonymous, "Xyz hand - a platform for soft dexterous manipulation," in preparation.
- [4] R. Deimel, M. Radke, and O. Brock, "Mass control of pneumatic soft continuum actuators with commodity components," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 774–779.
- [5] K. Y. Goldberg, "Orienting polygonal parts without sensors," *Algorithmica*, vol. 10, no. 2, pp. 201–225, 1993.
- [6] R. Tedrake, T. W. Zhang, Ming-fai Fong, and H. S. Seung, "Actuating a simple 3d passive dynamic walker," in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 5, 2004, pp. 4656–4661.
- [7] E. Brown, N. Rodenberg, J. Amend, A. Mozeika, E. Steltz, M. R. Zakin, H. Lipson, and H. M. Jaeger, "Universal robotic gripper based on the jamming of granular material," *Proceedings of the National Academy of Sciences*, vol. 107, no. 44, pp. 18 809–18 814, 2010.
- [8] A. Bicchi, M. Gabbicini, and M. Santello, "Modelling natural and artificial hands with synergies," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1581, pp. 3153–3161, 2011.
- [9] M. G. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi, "Adaptive synergies for the design and control of the pisa/IIT SoftHand," *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 768–782, 2014.
- [10] R. Deimel and O. Brock, "A novel type of compliant and underactuated robotic hand for dexterous grasping," *The International Journal of Robotics Research*, vol. 35, no. 1, pp. 161–185, 2016.
- [11] C. D. Santina, C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi, "Toward dexterous manipulation with augmented adaptive synergies: The pisa/IIT SoftHand 2," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1141–1156, 2018.
- [12] C. Paul, "Morphological computation: A basis for the analysis of morphology and control requirements," *Robotics and Autonomous Systems*, vol. 54, no. 8, pp. 619–630, 2006.
- [13] K. Ghazi-Zahedi, R. Deimel, G. Montúfar, V. Wall, and O. Brock, "Morphological computation: The good, the

- bad, and the ugly,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 464–469.
- [14] N. Hogan, “Impedance control: An approach to manipulation: Part 1 — theory,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 1–7, 1985.
- [15] —, “Impedance control: An approach to manipulation: Part 2 — implementation,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 8–16, 1985.
- [16] T. Lozano-Pérez, M. T. Mason, and R. H. Taylor, “Automatic synthesis of fine-motion strategies for robots,” *The International Journal of Robotics Research*, vol. 3, no. 1, pp. 3–24, 1984.
- [17] M. A. Erdmann and M. T. Mason, “An exploration of sensorless manipulation,” *IEEE Journal on Robotics and Automation*, vol. 4, no. 4, pp. 369–379, 1988.
- [18] C. Eppner, R. Deimel, J. Álvarez-Ruiz, M. Maertens, and O. Brock, “Exploitation of environmental constraints in human and robotic grasping,” *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 1021–1038, 2015.
- [19] L. Shao, T. Migimatsu, and J. Bohg, “Learning to scaffold the development of robotic manipulation skills,” *arXiv*, 2019. [Online]. Available: <http://arxiv.org/abs/1911.00969>
- [20] M. T. Mason, “The mechanics of manipulation,” in *International Conference on Robotics and Automation (ICRA)*, vol. 2, 1985, pp. 544–548.
- [21] A. D. Christiansen, “Manipulation planning for empirical backprojections,” in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Computer Society, 1991, pp. 762–768.
- [22] R. R. Burridge, A. A. Rizzi, and D. E. Koditschek, “Sequential composition of dynamically dexterous robot behaviors,” *The International Journal of Robotics Research*, vol. 18, no. 6, pp. 534–555, 1999.
- [23] R. Deimel, P. Irmisch, V. Wall, and O. Brock, “Automated co-design of soft hand morphology and control strategy for grasping,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1213–1218.
- [24] M. McEvoy and N. Correll, “Materials science. materials that couple sensing, actuation, computation, and communication,” *Science*, vol. 347, p. 1261689, 2015.
- [25] D. Rus and M. Tolley, “Design, fabrication and control of soft robots,” *Nature*, vol. 521, pp. 467–75, 2015.
- [26] L. U. Odhner, L. P. Jentoft, M. R. Claffee, N. Corson, Y. Tenzer, R. R. Ma, M. Buehler, R. Kohout, R. D. Howe, and A. M. Dollar, “A compliant, underactuated hand for robust manipulation,” *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 736–752, 2014.
- [27] S. Abondance, C. B. Teeple, and R. J. Wood, “A dexterous soft robotic hand for delicate in-hand manipulation,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5502–5509, 2020.
- [28] H. Jiang, Z. Wang, Y. Jin, X. Chen, P. Li, Y. Gan, S. Lin, and X. Chen, “Hierarchical control of soft manipulators towards unstructured interactions,” *The International Journal of Robotics Research*, 2021.
- [29] N. C. Daffe, 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 *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 1578–1585.
- [30] N. Chavan-Daffe and A. Rodriguez, “Prehensile pushing: In-hand manipulation with push-primitives,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 6215–6222.
- [31] R. R. Ma, W. G. Bircher, and A. M. Dollar, “Toward robust, whole-hand caging manipulation with underactuated hands,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1336–1342.
- [32] M. T. Mason, “Mechanics and planning of manipulator pushing operations,” *The International Journal of Robotics Research*, vol. 5, no. 3, pp. 53–71, 1986.
- [33] R. Platt, R. Burridge, M. Diftler, J. Graf, M. Goza, E. Huber, and O. Brock, “Humanoid mobile manipulation using controller refinement,” in *IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 94–101.
- [34] A. Majumdar and R. Tedrake, “Funnel libraries for real-time robust feedback motion planning,” *The International Journal of Robotics Research*, vol. 36, no. 8, pp. 947–982, 2017.
- [35] C. Eppner and O. Brock, “Planning grasp strategies that exploit environmental constraints,” *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2015, pp. 4947–4952, 2015.
- [36] M. Grebenstein, M. Chalon, W. Friedl, S. Haddadin, T. Wimböck, G. Hirzinger, and R. Siegwart, “The hand of the DLR hand arm system: Designed for interaction,” *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1531–1555, 2012.
- [37] J. Butterfass, M. Grebenstein, H. Liu, and G. Hirzinger, “DLR-hand II: next generation of a dextrous robot hand,” in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, 2001, pp. 109–114.
- [38] M. Pfanne, M. Chalon, F. Stulp, H. Ritter, and A. Albu-Schäffer, “Object-level impedance control for dextrous in-hand manipulation,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2987–2994, 2020.
- [39] B. Sundaralingam and T. Hermans, “Relaxed-rigidity constraints: kinematic trajectory optimization and collision avoidance for in-grasp manipulation,” *Autonomous Robots*, vol. 43, no. 2, pp. 469–483, 2019.
- [40] P. Michelman and P. Allen, “Forming complex dextrous manipulations from task primitives,” in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 4, 1994, pp. 3383–3388.
- [41] K. Or, M. Tomura, A. Schmitz, S. Funabashi, and S. Sug-

- ano, “Position-force combination control with passive flexibility for versatile in-hand manipulation based on posture interpolation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 2542–2547.
- [42] J. Bohg, A. Morales, T. Asfour, and D. Kragic, “Data-driven grasp synthesis—a survey,” *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.
- [43] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, “Learning complex dexterous manipulation with deep reinforcement learning and demonstrations,” in *Proceedings of Robotics: Science and Systems*, 2018.
- [44] A. Nagabandi, K. Konolige, S. Levine, and V. Kumar, “Deep dynamics models for learning dexterous manipulation,” in *Conference on Robot Learning (CoRL)*, 2019.
- [45] M. Li, H. Yin, K. Tahara, and A. Billard, “Learning object-level impedance control for robust grasping and dexterous manipulation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 6784–6791.
- [46] F. Veiga, R. Akrou, and J. Peters, “Hierarchical tactile-based control decomposition of dexterous in-hand manipulation tasks,” *Frontiers in Robotics and AI*, vol. 7, no. article no. 521448, 2020.
- [47] G. Solak and L. Jamone, “Learning by demonstration and robust control of dexterous in-hand robotic manipulation skills,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 8246–8251.