

Robust Peg-in-Hole Assembly under Uncertainties via Compliant and Interactive Contact-Rich Manipulation

Yiting Chen¹, Kenneth Kimble², Howard H. Qian¹, Podshara Chanrungaeneekul¹, Robert Seney², Kaiyu Hang¹

¹Department of Computer Science, Rice University

²U.S. National Institute of Standards and Technology

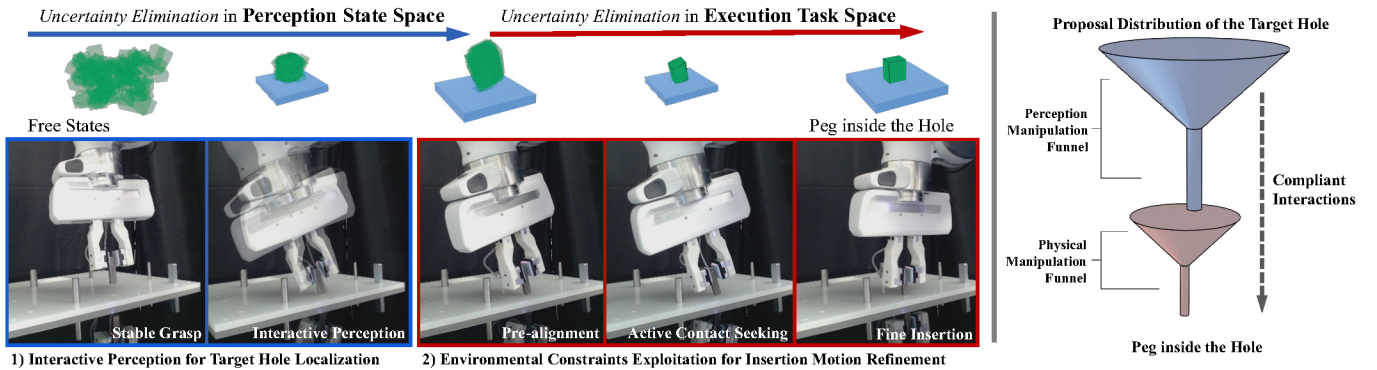


Fig. 1: **Motivation.** Acknowledging that real-world uncertainties are inevitable, we exploit environmental constraints to shape the manipulation process toward the desired outcome rather than expecting the robot to precisely execute any trajectories. The example demonstrates the system’s two major components: 1) progressively identify environmental constraints to localize the target hole; 2) form a constant contact between the inclined peg and its corresponding corner from the hole to serve as motion constraints for insertion refinement. Such an uncertainty-absorbing paradigm is formulated based on the composition of **manipulation funnels** —a concept that guarantees a strict space shrinkage through task-specific interactions, in the perception state space and execution task space.

Abstract—Robust and adaptive robotic peg-in-hole assembly under tight tolerances is critical to various industrial applications. However, it remains an open challenge due to perceptual and physical uncertainties from contact-rich interactions that easily exceed the allowed clearance. In this paper, we study how to leverage contact between the peg and its matching hole to eliminate uncertainties in the assembly process under unstructured settings. By examining the role of compliance under contact constraints, we present a manipulation system that plans collision-inclusive interactions for the peg to 1) iteratively identify its task environment to localize the target hole and 2) exploit environmental contact constraints to refine insertion motions into the target hole without relying on precise perception, enabling a robust solution to peg-in-hole assembly. By conceptualizing the above process as the composition of funneling in different state spaces, we present a formal approach to constructing manipulation funnels as an uncertainty-absorbing paradigm for peg-in-hole assembly. The proposed system effectively generalizes across diverse peg-in-hole scenarios across varying scales, shapes, and materials in a learning-free manner. Extensive experiments on a NIST Assembly Task Board (ATB) and additional challenging scenarios validate its robustness in real-world applications.

I. INTRODUCTION

Robotic peg-in-hole assembly is a foundational skill in various industrial applications and beyond, involving inserting

a peg into its matching hole on another object under tight clearances [28]. Common strategies seek to optimize the insertion trajectory under certain conditions, either through explicitly modeling the task environment or formulating the problem as a POMDP [29] (partially observable Markov decision process) to enable reactive policies based on sensory input. Regardless of the debate on how accurately a model describes reality, both types of approach neglect that the onus of fine manipulation lies in the execution [52]. It is the *execution* that brings the optimized result from planning into the real world through motor actuation by the controller. However, real-world perception challenges and physical uncertainties, such as those arising from hardware limitations, can introduce small deviations between planned and executed trajectories. Under the tight tolerance of peg-in-hole assembly, even minor uncertainties can lead to undesired manipulation outcomes. Therefore, robotic peg-in-hole assembly requires non-trivial effort to tackle real-world uncertainties beyond solely planned trajectories.

Acknowledging that planned trajectories are rarely executed perfectly in the real world due to uncertainties, focusing on the outcomes of interactions between the system and the environment [15, 14], rather than on explicit trajectories, offers

a promising approach to achieving robust manipulation. This formulation often requires the role of compliance, either from mechanical design [12, 42] or computational simulation [22], to passively adapt to external contacts. Rather than causing undesired outcomes, environmental constraints can shape the compliant manipulation process toward the desired outcome under deliberate exploitation [34], even without elaborate contact models [44, 8]. With this insight, we propose leveraging the contacts between the peg and its matching hole as an advantage to eliminate uncertainties for fine insertions.

In this paper, we introduce a manipulation system (shown in Fig. 1) that plans compliant interactions between the manipulated peg and its task environment to identify and exploit contact constraints for robust insertion. Safe contact is ensured by impedance control [22], which allows external contact to alter the pose of the peg by mimicking a spring-damper behavior. Since the steady pose of the peg after each interaction is observable and is a combined effect of the environmental constraints and potential from impedance control, the system can 1) identify its encountered constraints to shrink the target hole’s proposal distribution progressively and 2) exploit the environmental constraints to shape the motion of the peg to enter its matching hole. Specifically, the exploitation process involves maintaining constant contact between the peg and the hole to serve as motion constraints. Since the environmental constraints can shape a large margin of uncertainties and the interaction mechanism of force-pose regulation is independent of the specific trajectory taken, the proposed system can tolerate uncertainties smaller than its execution precision, even under tight tolerances.

As the system dynamics progressively eliminates uncertainty in perception and execution, we formulate the proposed system as a composition of *manipulation funnels* in different state spaces. The original concept of the *manipulation funnel* was introduced by Mason [39] to eliminate uncertainties in object location, emphasizing that the robustness of funnel-based manipulation requires the co-design of robot interaction and task mechanics. A general formulation of the funneling process can be defined in any task-relevant space as a progressive space shrinkage for uncertainty elimination. Inspired by the planar theory of automatic motion synthesis [34], we establish a formal approach to constructing manipulation funnels for robust fine insertion of general prismatic pegs in the real world. The primary contributions of this paper are:

- We present an uncertainty-absorbing paradigm for general peg-in-hole insertion based on the concept of manipulation funnel, which does not plan on the explicit trajectories but rather the outcome of each compliant interaction.
- We present a formal approach to constructing the funneling process for both perception and execution, which can be further composed for general peg-in-hole insertion.
- We provide a system with detailed algorithms to implement the above funnel-based manipulation paradigm on real-world assembly tasks.
- We conduct ablation studies through extensive experiments on a standard assembly benchmark NIST ATB and

additional challenging tasks to validate the robustness and generalizability of the proposed paradigm and system.

The rest of this paper is organized as follows: Section II reviews related works. Section III introduces the preliminaries and problem statement for funnel-based manipulation planning in peg-in-hole assembly. Section IV introduces the task mechanics and presents a formal approach to constructing manipulation funnels in perception and execution. Section V demonstrates the robustness of the proposed system in simulation and real-world experiments. We discuss the limitations under current settings in Section VI and summarize the contributions in Section VII.

II. RELATED WORK

A. Peg-in-Hole Assembly

Over the past decade, pipeline-based and end-to-end approaches have driven advancements in robotic peg-in-hole systems. Pipeline-based approaches aim to explicitly estimate intermediate system states for motion planning. Previous works [25, 51, 27] leverage tactile or force/torque sensors for contact state estimation. Vision-based methods [54, 21] estimate alignment deviations for correction or search motion planning. This modular design provides interpretability and flexibility, allowing for the integration of both analytical and data-driven methods. In contrast, end-to-end approaches [30] aim at deriving an implicit function to map sensory inputs to motor skills from trial-and-error. The entire system connects feature extraction to planning in a unified manner [13, 37, 50, 35, 24, 36].

However, previous methods are mainly based on the principle that assembly is a relative positioning task [49], assuming the robot will precisely execute the issued trajectory. Our research targets a different perspective [34], as the fine insertion motion is a passive refinement process that exploits environmental constraints.

B. Compliance-enabled Manipulation

Compliance mitigates unknown external disturbances through passive adaptability. The role of compliance in robot manipulation often involves compensation for uncertainty and safe interaction with the environment, which has been adopted in a wide range of tasks, including dexterous grasping [6, 33, 12], in-hand manipulation [32, 41, 2], grasp adaptation [31, 19, 26], assembly [40, 38, 55] and human-robot co-manipulation [47, 43].

Additionally, compliance-enabled contact can safely exploit environmental constraints to eliminate uncertainties, ensuring robust and dexterous manipulation [14, 20, 11, 46, 56, 23, 7]. In this work, we propose a formal approach for compliance-enabled contact between the peg and the hole to eliminate uncertainty in insertion motions.

C. Funnel-based Manipulation

Mason [39] proposed the original concept of *manipulation funnels* to eliminate uncertainties based on task mechanics and task-specific interaction. The importance of exploiting

task mechanics to eliminate uncertainties is also revealed in concepts such as *pre-image* [34] and *backprojections* [16]. Early works have applied funnel-based designs in manipulation tasks, including ball batting [3], part orienting [18, 17] and feeding [1], grasping [9], and peg-in-hole assembly [53]. Bhatt et al. [2] investigated funnel-based action primitives for in-hand manipulation through an empirical study. Canberk et al. [5] proposes “canonicalized-alignment” as task mechanics to funnel the large space of possible cloth configurations into a smaller, structured one.

The major challenges in developing funnel-based manipulation lie twofold: 1) formulating the general task mechanics for a given manipulation task and 2) defining task-specific interaction beyond hardware-associated control. To tackle the above challenges in real-world peg-in-hole tasks, we provide a general formulation of the task mechanics of peg-in-hole insertion and an object-centric interaction mechanism, which is not dependent on any specific hardware.

III. PRELIMINARIES AND PROBLEM STATEMENT

In this work, we consider the peg-in-hole assembly problem as inserting a peg into its matching hole on a planar board as illustrated in Fig. 2-(a). We aim to plan compliant motions of a manipulated peg that frequently makes and breaks contact with its task environment, first to perceive the location of the matching hole on the task board and then insert the peg into it under tight clearance.

A. Preliminaries

1) *Geometry of the Peg and Hole*: The 3-D geometry of the hole is described in the *hole reference frame* $\{H\}$, as an intrusion of a 2-D polygonal base defined by an ordered sequence of vertices $\mathcal{V}_h = (v_h^0, v_h^1, \dots, v_h^n)$, $v_h \in \mathbb{R}^3$, $z_{v_h} = 0$ on the XY -plane of $\{H\}$. $S(\mathcal{V}_h) = \cup_{i=0}^n \{v_h^i\} \subset \mathbb{R}^3$ is the corresponding unordered set of the \mathcal{V}_h . The intrusion is along the negative direction of the Z -axis of $\{H\}$. \mathcal{V}_h defines a planar region $\mathcal{A}_h \subset \mathbb{R}^3$ on the XY -plane of $\{H\}$ as:

$$\mathcal{A}_h = \{(x, y, z) \in \mathbb{R}^3 \mid f_{\mathcal{A}_h}(x, y) \leq 0 \wedge z = 0\} \quad (1)$$

where $f_{\mathcal{A}_h}(x, y)$ is an implicit distance function to describe the boundary of \mathcal{A}_h as $f_{\mathcal{A}_h}(x, y) = 0$ and its interior area as $f_{\mathcal{A}_h}(x, y) < 0$, exterior area as $f_{\mathcal{A}_h}(x, y) > 0$.

The 3-D geometry of the peg is described in the *peg reference frame* $\{E\}$, as an extrusion of a 2-D polygon base $\mathcal{V}_e = (v_e^0, v_e^1, \dots, v_e^n)$, $v_e \in \mathbb{R}^3$, $z_{v_e} = 0$ on the XY -plane of $\{E\}$. Its corresponding unordered set is denoted as $S(\mathcal{V}_e) = \cup_{i=0}^n \{v_e^i\} \subset \mathbb{R}^3$. The extrusion is along the positive direction of Z -axis of $\{E\}$, and \mathcal{V}_e is considered approximated equal to \mathcal{V}_h but defined in different frames.

2) *Task Environment*: We specify the *Cartesian world frame* $\{O\}$ attached to the task board, as its XY -plane is aligned with the board surface and its orthogonal Z -axis points outward the task board. The pose of the hole in the world frame is denoted as $T_{\{OH\}} \in \text{SE}(3)$. We describe the task

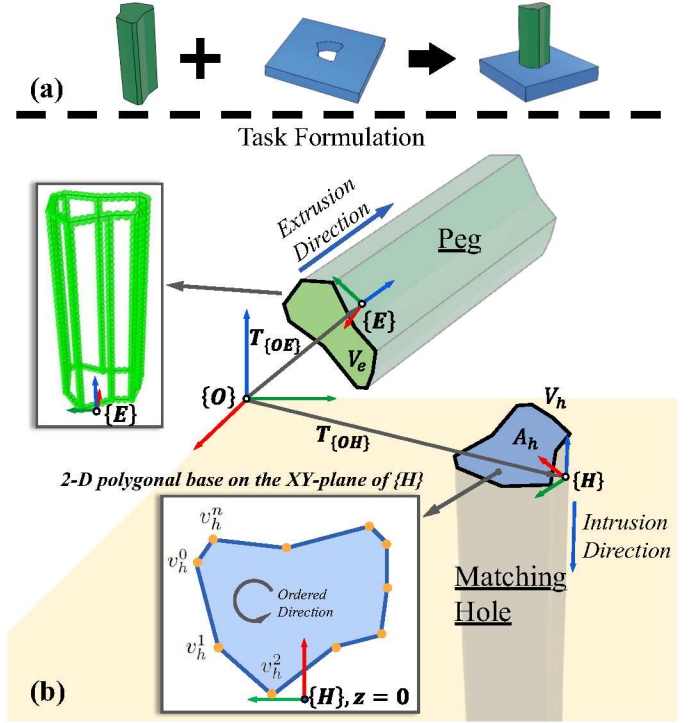


Fig. 2: (a) The peg-in-hole problem is considered as inserting a peg into its matching hole on a planar board (a randomly generated peg is adopted as the example). (b) An overview of the task formulation is presented, including the geometry representation of the peg, the hole, and the overall task environment.

environment using an occupancy function $\Gamma : \mathbb{R}^3 \mapsto [-1, 0, 1]$, which is defined as follows:

$$\Gamma(x, y, z) = \begin{cases} -1, & z < 0 \wedge f_{T_{\{OH\}}\mathcal{A}_h}(x, y) > 0 \\ 1, & z > 0 \vee (z \leq 0 \wedge f_{T_{\{OH\}}\mathcal{A}_h}(x, y) < 0) \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

in which -1 denotes occupied by the task board, 0 denotes on the surface of the task board, and 1 denotes the free space. An initial area of interest $\mathcal{A}_{\text{board}} \subset \mathbb{R}^3$ is specified on the XY -plane of $\{O\}$ that guarantees $T_{\{OH\}}\mathcal{A}_h \subset \mathcal{A}_{\text{board}}$.

3) *Compliant Interactions for the Peg*: As the peg is in a prism-shaped geometry as defined in Sec. III-A1, we use the densely sampled points $\mathcal{P}_e = \{p_e^i \in \mathbb{R}^3\}_{i=1}^M$ on the edges of the peg for collision checking (considered continuous along the lateral and base edges), in which $p_e = (x, y, z)$ is in the Cartesian coordinates of $\{E\}$. The sampled points are assumed to be rigidly attached to each other, as the entire point set undergoes the same rigid body transformation (rotation and translation). The pose of frame $\{E\}$ with reference to frame $\{O\}$ at time t is represented as $\mathbf{x}_t = [\mathbf{p}; \mathbf{r}] \in \mathbb{R}^6$, in which $\mathbf{p} \in \mathbb{R}^3$ is the position vector and $\mathbf{r} \in \mathbb{R}^3$ is the orientation represented with Euler angle. The target peg-in-hole state is denoted as \mathbf{x}^* . The corresponding transformation matrix of \mathbf{x}_t is denoted as $T_{\{OE\},t} \in \text{SE}(3)$. An overview of the task environment with the manipulated peg is illustrated in Fig. 2-(b).

To enable safe contact between the peg and its task environment, we formulate the impedance control based on the desired state $\mathbf{x}_t^d \in \mathbb{R}^6$ at time t . A wrench $\mathbf{F} \in \mathbb{R}^6$ (consisting of 3-D force and torque) is applied to the peg by a Cartesian impedance controller as formulated:

$$\mathbf{F} = \mathbf{K}_d(\mathbf{x}_t^d - \mathbf{x}_t) - \mathbf{D}_d\dot{\mathbf{x}}_t \quad (3)$$

in which $\mathbf{K}_d \in \mathbb{R}^{6 \times 6}$ is the diagonal stiffness matrix to decouple compliance in each degree of freedom, $\mathbf{D}_d \in \mathbb{R}^{6 \times 6}$ is the damping matrix for movement stabilizing and $\dot{\mathbf{x}}_t \in \mathbb{R}^6$ is the velocity of $\{E\}$ with reference to $\{O\}$ at time t .

An interaction command $\mathbf{c}_t = (\mathbf{x}_t, \mathbf{x}_t^d)$ at time t is defined by its starting state \mathbf{x}_t (considered steady as $\dot{\mathbf{x}}_t = 0$) and a desired state \mathbf{x}_t^d . A compliant interaction $\mathbf{x}_{t+1} = \Pi(\mathbf{x}_t, \mathbf{x}_t^d)$ between the peg and its task environment is an unknown but observable process $\Pi : \mathbb{R}^6 \times \mathbb{R}^6 \mapsto \mathbb{R}^6$ from the starting state \mathbf{x}_t to the next steady state \mathbf{x}_{t+1} , under the drive of \mathbf{x}_t^d and the environmental constraints in the frame of $\{O\}$. Except for the virtually defined desired state \mathbf{x}_t^d , any physical existing state \mathbf{x}_{t+1} during this process is constrained by its task environment as follows:

$$\forall p_o \in T_{\{OE\},t+1} \mathcal{P}_e, \Gamma(x_{p_o}, y_{p_o}, z_{p_o}) \neq -1 \quad (4)$$

We do not explicitly predict or trace the trajectory of the peg during this process but only observe its steady state at \mathbf{x}_t and \mathbf{x}_{t+1} . At the steady state \mathbf{x}_{t+1} , any external contact wrench $\mathbf{F}_{\text{ext}} \in \mathbb{R}^6$ in frame $\{O\}$ is balanced by the pose deviation as:

$$\mathbf{F}_{\text{ext}} = \mathbf{K}_d(\mathbf{x}_t^d - \mathbf{x}_{t+1}) \quad (5)$$

while the velocity $\dot{\mathbf{x}}_{t+1} = 0$ of the peg remains zero.

We use the intersection point set $\mathcal{P}_{o,\text{footprint},t} \subset \mathbb{R}^3$ between the peg's edges and the XY -plane of $\{O\}$ at the steady pose \mathbf{x}_t to describe the outcome of each interaction as:

$$\mathcal{P}_{o,\text{footprint}} = \{p_o \in T_{\{OE\},t} \mathcal{P}_e \mid \forall p_o = (x, y, z), z = 0\} \quad (6)$$

$\mathcal{P}_{o,\text{footprint}}$ reflects the spatial relation between the peg and the task board.

B. Problem Statement

We formulate the peg-in-hole assembly task as a progressive uncertainty elimination process in the state space of perception and execution. We aim to plan interactions $\mathbf{c}_t = (\mathbf{x}_t, \mathbf{x}_t^d)$ between the peg and its task environment to 1) iteratively reduce the uncertainty of the perception result of the target hole and 2) finish the peg insertion under execution and perception uncertainty. An overview of the peg-in-hole assembly process is outlined in Alg. 1.

Perception State Space: We use $\tilde{T}_{\{OH\}}$ to denote the estimated state of the ground truth $T_{\{OH\}}$. The estimation result at time t is represented with a probabilistic density function $P_t(\tilde{T}_{\{OH\}})$ bounded by a proposal region $\mathcal{X}_t \subset \text{SE}(3)$, which is defined as follows:

$$P_t(\tilde{T}_{\{OH\}}) = \begin{cases} \frac{1}{|\mathcal{X}_t|}, & \tilde{T}_{\{OH\}} \in \mathcal{X}_t \\ 0, & \tilde{T}_{\{OH\}} \notin \mathcal{X}_t \end{cases} \quad (7)$$

where $|\mathcal{X}_t|$ denotes the volume of region \mathcal{X}_t .

We aim to form a recursive sequence of proposal regions based on the outcomes $\mathcal{P}_{o,\text{footprint},t}$ of each interaction \mathbf{c}_t :

$$\mathcal{X}_0 \supseteq \mathcal{X}_1 \supseteq \mathcal{X}_2 \supseteq \cdots \supseteq \mathcal{X}_{T_1} \quad (8)$$

As \mathcal{X}_t shrinks over steps, the expected spread of $\tilde{T}_{\{OH\}}$ decreases and the uncertainty range of the perceived hole's state is reduced.

Execution Task Space: Let $\Delta \mathbf{x} \in \mathbb{R}$ be the deviation between the steady state \mathbf{x}_t and the peg-in-hole state \mathbf{x}^* . Based on the estimated state distribution of $P_{T_1}(\tilde{T}_{OH})$, we aim to shrink $\Delta \mathbf{x}$ at each step of interaction \mathbf{c}_t to progressively reach the peg-in-hole configuration as follows:

$$\Delta \mathbf{x}_0 > \Delta \mathbf{x}_1 > \Delta \mathbf{x}_2 > \cdots > \Delta \mathbf{x}_{T_2} \quad (9)$$

We consider a successful peg-in-hole assembly when $\Delta \mathbf{x}$ reaches zero.

Algorithm 1 Peg-in-Hole Assembly

Input: base polygon \mathcal{V}_h , initial area of interest $\mathcal{A}_{\text{board}}$, peg geometry as \mathcal{P}_e , world frame $\{O\}$

- 1: $t \leftarrow 0$ ▷ Initialize time step for perception
- 2: $\mathcal{X}_{T_1} \leftarrow \text{InteractivePerception}(\mathcal{V}_h, \mathcal{A}_{\text{board}}, \mathcal{P}_e, \{O\})$ ▷ Alg. 2
- 3: $t \leftarrow 0$ ▷ Reset time step for execution
- 4: $\mathbf{x}_t \leftarrow \text{CornerAlignment}(\mathcal{X}_{T_1})$ ▷ Alg. 3
- 5: **while** not reach \mathbf{x}^* **do**
- 6: $\mathbf{x}_t^d \leftarrow \text{InsertionPlanning}(\mathbf{x}_t)$ ▷ Alg. 4
- 7: $\mathbf{x}_{t+1} \leftarrow \mathbf{c}_t(\mathbf{x}_t, \mathbf{x}_t^d)$
- 8: $t \leftarrow t + 1$
- 9: **end while**

IV. FUNNEL-BASED MANIPULATION PLANNING

A manipulation funnel represents a convergence process by iterative contractions from a larger entrance to a smaller exit state space, ensured by task-specific interactions. A general funnel formulation is defined as *Definition 4.1*.

Definition 4.1. Let S be a task-relevant space that could represent any system parameters. Given a target state \mathbf{s}^* , we aim to find subspaces $S_{\text{in}} \subset S$ so that after applying some state transition functions, denoted as $\Pi : S \rightarrow S$, the resultant output space $S_{\text{out}} \subset S$ is guaranteed to be strictly smaller than the input space: $|S_{\text{in}}| > |S_{\text{out}}|$. If these conditions are satisfied, $\mathbf{s}^* \in S_{\text{in}}$ and $\mathbf{s}^* \in S_{\text{out}}$, a tuple $\mathcal{F} = (S_{\text{in}}, \Pi, S_{\text{out}})$ is called a manipulation funnel.

We first define the task-specific interactions based on the task mechanics in Section IV-A. Then, we introduce the formal approach to construct manipulation funnels in perception state space (Section IV-B) and execution task space (Section IV-C).

A. Task Mechanics and Interaction Primitives

Funnel-based manipulation planning requires deep exploitation of the intrinsic task mechanics [39]. As perception and physical uncertainties are inevitable [45], our key insight is

forming an aligned *corner* between the inclined peg and the target hole to create contact constraints for undesired motion freedoms and progressively enter the allowed clearance (as illustrated in Fig. 3-(a)). A *corner* from the hole is defined by arbitrary vertex $v_h^{\text{corner}} \in S(\mathcal{V}_h)$ as the local angle $\angle v_h^{\text{corner}}$ formed by v_h^{corner} and its nearby edges from the interior side. In this paper, we use the case of $\angle v_h^{\text{corner}} < \pi$ for motion funnel construction. However, theoretically, it also applies to the case of $\angle v_h^{\text{corner}} > \pi$ as long as the local convexity exists.

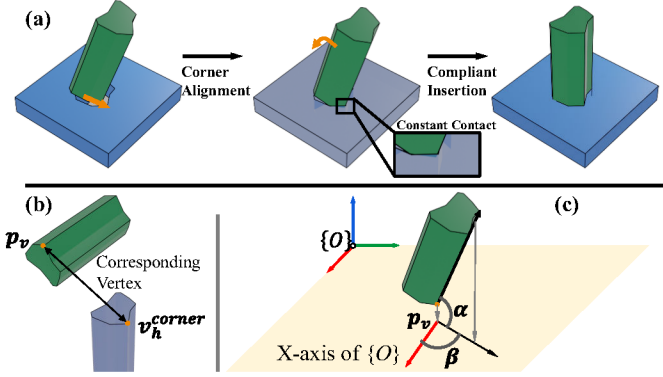


Fig. 3: (a) The task mechanics of peg-in-hole insertion: first, constant contact between the peg and the hole is formed; second, the formed constraints are leveraged to refine the insertion motions. (b) A paired corner on the peg and hole: this local geometry enables the downstream iterative insertion process. (c) The inclined state is constrained by its supporting vertex p_v , inclined angle α and rotation angle β .

Despite the trajectory being a dominant action representation in manipulation planning, it is unsuitable for funnel-based manipulations as interactions with the task environment are allowed to alter the motion of the manipulator [39]. Given the target corner v_h^{corner} of the hole, its corresponding vertex on the base of the peg is denoted as $p_v = v_e^{\text{corner}} \in S(\mathcal{V}_e) \subset \mathcal{P}_e$, as shown in Fig. 3-(b). To integrate the alignment mechanism as an inductive bias for interactions, the task-specific interaction $\mathbf{c}_t = (\mathbf{x}_t^d, \mathbf{x}_t)$ is extended as $\mathbf{c}_t^v = (\mathbf{x}_t^d, \mathbf{x}_t, p_v)$ in Definition 4.2.

Definition 4.2. For a task-specific interaction \mathbf{c}_t^v , the starting state \mathbf{x}_t and desired state \mathbf{x}_t^d is constrained by a common supporting vertex p_v defined as follows:

$$\forall \mathbf{x} \in [\mathbf{x}_t^d, \mathbf{x}_t] : \forall p_o, p_o \in T_{\{O\}} \mathcal{P}_e \wedge p_o \neq T_{\{O\}} p_v, z_{p_o} > z_{T_{\{O\}} p_v} \quad (10)$$

Such a formulation ensures that the vertex p_v on the peg base possesses the lowest value along the Z-axis of $\{O\}$, while the interaction process is unchanged as $\mathbf{x}_{t+1} = \Pi(\mathbf{x}_t, \mathbf{x}_t^d)$.

An inclined state is visualized in Fig. 3-(c). Let \mathbf{n} represents the positive direction of Z-axis of $\{E\}$ with reference to $\{O\}$ at the inclined state, we use α as the inclined angle and β as

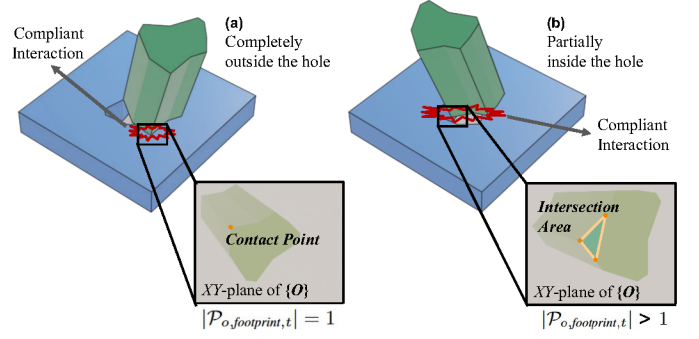


Fig. 4: The observation $\mathcal{P}_{o,\text{footprint},t}$ from \mathbf{c}_t^v can be divided into binary categories, (a) contact point when $|\mathcal{P}_{o,\text{footprint},t}| = 1$ or (b) intersection area when $|\mathcal{P}_{o,\text{footprint},t}| > 1$, based on its cardinality.

the rotation angle to define \mathbf{n} as follows:

$$\mathbf{n} = \begin{bmatrix} \cos(\alpha) \cos(\beta) \\ \cos(\alpha) \sin(\beta) \\ \sin(\alpha) \end{bmatrix} \quad (11)$$

Interaction with inclined states is designed to identify and exploit its environmental contact constraints. Such a formulation offers several advantages: 1) when the peg engages with the task board from above (as illustrated in Fig. 4), the observation $\mathcal{P}_{o,\text{footprint},t}$ is divided into binary categories based on the cardinality of the set $|\mathcal{P}_{o,\text{footprint},t}|$: whether it is a contact point or an intersection area between the peg and hole; 2) an inclined state relaxes the requirement for a peg to be partially entered into the hole [34]; 3) before an insertion task is finished (considered vertical w.r.t. the board surface plane), all peg states can be described as inclined states during the insertion process.

B. Perception Manipulation Funnel

The perception manipulation funnel aims to iteratively reduce the proposal region \mathcal{X}_t of the target hole's distribution $P_t(\tilde{T}_{\{OH\}})$ through observation from each interaction. By strategically placing the supporting vertex p_v of the desired state \mathbf{x}_t^d beneath the XY-plane of $\{O\}$, we can intentionally capture the encountered contact constraints reflected on the resulted steady state \mathbf{x}_{t+1} . As illustrated in Fig. 4, after each interaction, the observation $\mathcal{P}_{o,\text{footprint},t}$ can be interpreted as an inequality constraint as defined in Definition 4.3 to limit the distribution of $\tilde{T}_{\{OH\}}$.

Definition 4.3. An inequality constraint on the possible space of $\tilde{T}_{\{OH\}}$ is defined as $g_t(\tilde{T}_{\{OH\}}) \geq 0$ based on the observation $\mathcal{P}_{o,\text{footprint},t}$, where $g_t(\tilde{T}_{\{OH\}})$ is defined as:

$$\forall p_o \in \mathcal{P}_{o,\text{footprint},t}, \quad g_t(\tilde{T}_{\{OH\}}) = \begin{cases} f_{\tilde{T}_{\{OH\}} \mathcal{A}_h}(p_o), & |\mathcal{P}_{o,\text{footprint},t}| = 1 \\ -\max f_{\tilde{T}_{\{OH\}} \mathcal{A}_h}(p_o), & \text{Otherwise} \end{cases} \quad (12)$$

in which $|\cdot|$ is the cardinality of the set. For a potential hole's state $\tilde{T}_{\{OH\}}$, any contact point (as $|\mathcal{P}_{o,\text{footprint},t}| = 1$) should be excluded from the hole's area, while any intersection area

(as $|\mathcal{P}_{o,\text{footprint},t}| > 1$) should be fully contained by the hole's area.

The proposal region at $t = 0$ is derived from the initial area of interest $\mathcal{A}_{\text{board}}$ as follows:

$$\mathcal{X}_0 = \{\tilde{T}_{\{OH\}} \mid \tilde{T}_{\{OH\}}\mathcal{A}_h \subset \mathcal{A}_{\text{board}}\} \quad (13)$$

The proposal region at step t is defined as:

$$\mathcal{X}_t = \{\tilde{T}_{\{OH\}} \in \mathcal{X}_0 \subset SE(3) \mid g_i(\tilde{T}_{\{OH\}}) \geq 0, i = 1, 2, \dots, t\} \quad (14)$$

which is updated at step $t + 1$ with the new added constraint $g_{t+1}(\tilde{T}_{\{OH\}}) \geq 0$ as:

$$\mathcal{X}_{t+1} = \mathcal{X}_t \cap \{\tilde{T}_{\{OH\}} \in SE(3) \mid g_{t+1}(\tilde{T}_{\{OH\}}) \geq 0\} \quad (15)$$

Lemma 4.4. As defined in Eq. (15), the volume $|\mathcal{X}_t|$ of the feasible region \mathcal{X}_t is monotonically decreasing ($\mathcal{X}_{t+1} \subseteq \mathcal{X}_t$) over nonidentical interactions \mathbf{c}_t^v .

Proof: Since \mathcal{X}_{t+1} is the intersection of \mathcal{X}_t with another constraint set $g_{t+1}(\tilde{T}_{\{OH\}}) \geq 0$ and $\{g_t(\tilde{T}_{\{OH\}}) \geq 0\} \neq \{g_{t+1}(\tilde{T}_{\{OH\}}) \geq 0\}$ under nonidentical interactions, $|\mathcal{X}_t|$ is decreasing or stays the same when new constraints are added. ■

As \mathcal{X}_t shrinks over steps in Lemma 4.4, the expected spread of $\tilde{T}_{\{OH\}}$ decreases, and the allowable state space of the target hole narrows.

Lemma 4.5. The ground truth state of the hole is always included in the allowable state space as $T_{\{OH\}} \in \mathcal{X}_t$ over interactions \mathbf{c}_t^v .

Proof: As $T_{\{OH\}} \in \mathcal{X}_0$ and $T_{\{OH\}} \in \{\tilde{T}_{\{OH\}} \mid g_t(\tilde{T}_{\{OH\}}) \geq 0\}$, thus proven $T_{\{OH\}} \in \mathcal{X}_t$ as defined in Eq. (14). ■

Since the proposal region $\mathcal{X}_{t+1} \subseteq \mathcal{X}_t$ of the target hole is monotonically shrinking over interactions \mathbf{c}_t^v , meanwhile the target state $T_{\{OH\}}$ is guaranteed to be contained by \mathcal{X}_t and \mathcal{X}_{t+1} , we introduce the perception manipulation funnel as defined in Theorem 4.6.

Theorem 4.6. The process of proposal region shrinkage $\mathcal{X}_{t+1} \subseteq \mathcal{X}_t$ through compliant interactions \mathbf{c}_t^v is a manipulation funnel in the perception state space.

Proof: As proven in Lemma 4.4 and 4.5, the proposal region \mathcal{X}_t is shrinking and approaching $T_{\{OH\}}$ over random interactions \mathbf{c}_t^v , thus proving that the perception manipulation funnel is established based on the general Definition 4.1. ■

Additionally, a maximum entropy-based method is introduced to improve convergence efficiency. An overview is outlined in Alg. 2. The probability distribution $P_t(\tilde{T}_{\{OH\}})$ of the target hole's state is defined by Eq. 7 and the initial area for exploration $\mathcal{A}_{\text{board}}$ is divided into grid points $\mathcal{G} = \{\mathbf{g}_i \mid i = 1, 2, \dots, N\}$ where $\mathbf{g}_i \in \mathbb{R}^3$ on the XY -plane of $\{O\}$. During each interaction, number K of potential hole's states is sampled as $\{\tilde{T}_{\{OH\},i}\}_{i=1}^K \sim P_t(\tilde{T}_{\{OH\}})$. The probability P_{in} that \mathbf{g} is included by the target hole is defined by:

$$P_{\text{in}}(\mathbf{g}) = \frac{\sum_{i=1}^K h(\tilde{T}_{\{OH\},i}, \mathbf{g})}{K}, \quad (16)$$

$$h(\tilde{T}_{\{OH\}}, \mathbf{g}) = \begin{cases} 1, & \mathbf{g} \in \tilde{T}_{\{OH\}}\mathcal{A}_h \\ 0, & \text{Otherwise} \end{cases} \quad (17)$$

Let $\mathcal{H}(\cdot)$ represent the entropy of the given probabilistic density function. The expected entropy reduction is formulated as:

$$\Delta\mathcal{H}(\mathbf{g}) = \mathcal{H}(P_t(\tilde{T}_{\{OH\}})) - E[\mathcal{H}(P_t(\tilde{T}_{\{OH\}}) \mid \mathbf{g})] \quad (18)$$

where $E[\mathcal{H}(P_t(\tilde{T}_{\{OH\}}) \mid \mathbf{g})]$ is the expected entropy after evaluating \mathbf{g} . To maximize information gain, the desired state should spatially overlap with the selected grid point $\mathbf{g}^* \in \mathcal{G}$ that maximizes $\Delta\mathcal{H}(\mathbf{g})$ for contact constraints identification:

$$\mathbf{g}^* = \arg \max_{\mathbf{g} \in \mathcal{G}} \Delta\mathcal{H}(\mathbf{g}) \quad (19)$$

The convergence of the proposed entropy-based exploration is proven in Theorem 4.6.

Algorithm 2 Entropy-based Interactive Perception

Input: $\mathcal{V}_h, \mathcal{A}_{\text{board}}, \mathcal{P}_e, \{O\}$

1: $t \leftarrow 0$

2: Sample grid points $\mathcal{G} \leftarrow \mathcal{A}_{\text{board}}$

3: Initialize proposal region $\mathcal{X}_t \leftarrow \mathcal{A}_{\text{board}}$ ▷ Eq.(13)

4: **while** not converged **do**

5: $\mathbf{g}^* \leftarrow$ Select grid point for interaction ▷ Eq.(19)

6: $\mathcal{P}_{o,\text{footprint},t} \leftarrow$ Active contact with the selected point

7: $\mathcal{X}_{t+1} \leftarrow$ ConstraintsUpdate($\mathcal{P}_{o,\text{footprint},t}$) ▷ Eq.(15)

8: $t \leftarrow t + 1$

9: **end while**

Output: bounded region \mathcal{X}_{T1}

C. Physical Manipulation Funnel

The physical manipulation funnel aims to leverage environmental contacts as physical constraints in the execution task space to iteratively reduce $\Delta\mathbf{x}$ for peg-in-hole insertion. Based on the task mechanics as Fig. 3-(a), we consider a sequential process as 1) aligning an inclined state to its corresponding corner and 2) adjusting the aligned inclined state to insertion. All the states \mathbf{x}_t described in this section are considered inclined states, which are defined by p_v, α , and β as introduced in Eq. (10) and (11). We describe the spatial relationship between the \mathbf{x}_t and \mathbf{x}^* (the target peg-in-hole state) based on the $\mathcal{P}_{o,\text{footprint},t}$ and the target corner v_h^{corner} . Specifically, $\mathcal{P}_{o,\text{footprint},t}$ forms a 2-D polygon as $\mathcal{V}_{o,\text{footprint},t} = (p_o^0, p_o^1, \dots, p_o^k), p_o \in \mathcal{P}_{o,\text{footprint},t}$ with a corresponding area $\mathcal{A}_{o,\text{footprint},t} \subset \mathbb{R}^3$ on the XY -plane of $\{O\}$. The state deviation $\Delta\mathbf{x}$ is defined based on $\mathcal{V}_{o,\text{footprint},t}$ and v_h^{corner} as Definition 4.7.

Definition 4.7. Given an inclined peg state \mathbf{x}_t with the lateral edge of the supporting vertex p_v intersects with the XY -plane of $\{O\}$, we project its state as $\Psi : \mathbb{R}^6 \mapsto \mathbb{R}^4$. We use the

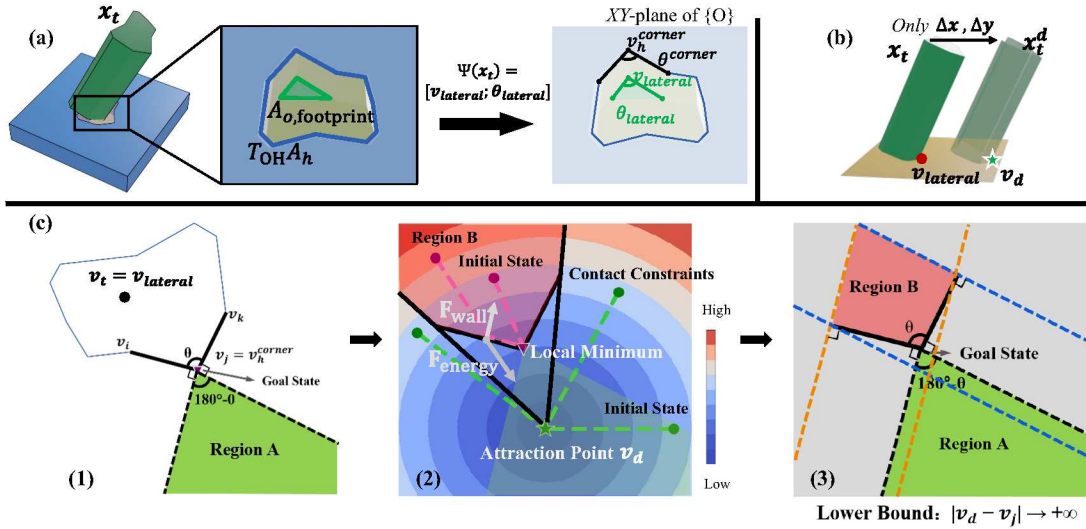


Fig. 5: (a) The peg's state \mathbf{x}_t is projected as $[v_{\text{lateral}}; \theta_{\text{lateral}}]$. (b) By setting the translation deviation between the desired state \mathbf{x}_t^d and the starting state \mathbf{x}_t only on the XY -plane of $\{O\}$, a 2-D potential energy field is formulated. (c) The corner alignment process is formulated as aligning the point $v_t = v_{\text{lateral}}$ to the corner $v_j = v_h^{\text{corner}}$ under the drive of the created potential energy field, while the edges are considered non-penetration walls.

$v_{\text{lateral}} \in \mathbb{R}^3$ and the angle value $\theta_{\text{lateral}} \in \mathbb{R}$ of $\angle v_{\text{lateral}}$ to describe the pose as $\Psi(\mathbf{x}_t) = [v_{\text{lateral}}; \theta_{\text{lateral}}]$, as illustrated in Fig. 5-(a). v_{lateral} is the intersection point between the lateral edge of p_v and the XY -plane of $\{O\}$, $\angle v_{\text{lateral}}$ is the angle formed by v_{lateral} and its nearby vertices in $\mathcal{V}_{o, \text{footprint}, t}$. Given the target corner for alignment v_h^{corner} and its corresponding angle $\angle v_h^{\text{corner}}$, the deviation $\Delta \mathbf{x}$ is transformed into $\Delta \mathbf{s}$ as follows:

$$\Delta \mathbf{s} = \left\| v_{\text{lateral}} - T_{\{OH\}} v_h^{\text{corner}} \right\| + \left| \theta^{\text{corner}} - \theta_{\text{lateral}} \right| \quad (20)$$

where $\theta^{\text{corner}}, \theta_{\text{lateral}}$ is the angle value of $\angle v_h^{\text{corner}}$ and $\angle v_{\text{lateral}}$ respectively. $\Delta \mathbf{x} \rightarrow 0$ is considered equal to $\Delta \mathbf{s} \rightarrow 0$.

Since the footprint fully contains the projection of the intersection part beneath the board surface plane, the environmental constraints as defined in Eq. 4 are projected into the XY -plane of $\{O\}$ as $\mathcal{A}_{o, \text{footprint}, t} \subset T_{\{OH\}} \mathcal{A}_h$.

In the following, we describe the process of inclined corner alignment as Δv reaches zero and the process from inclined to insertion as $\Delta \theta$ reaches zero. Since each degree of freedom in the Cartesian space of $\{O\}$ is decoupled as \mathbf{K}_d is a diagonal stiffness matrix, we aim to leverage the potential caused by translation deviation for corner alignment and rotational deviation for insertion based on Lemma 4.8.

Lemma 4.8. Under the quasi-static assumption, the motion of the peg \mathbf{x} is considered driven by a potential energy field:

$$U(\mathbf{x}) = \frac{1}{2} (\mathbf{x}_d - \mathbf{x})^\top \mathbf{K}_d (\mathbf{x}_d - \mathbf{x}) \quad (21)$$

Given an interaction $\mathbf{c}_t^v = (\mathbf{x}_t, \mathbf{x}_t^d)$, the resulted steady pose \mathbf{x}_{t+1} automatically falls into its nearby local minimum energy under the non-penetration environmental constraints as $U(\mathbf{x}_t) \geq U(\mathbf{x}_{t+1}) \geq U(\mathbf{x}_t^d)$.

Proof: The peg tends to rest at a lower energy state in the potential energy field. Also, the peg cannot move to a higher energy state without external force. ■

Inclined Corner Alignment: By constraining the deviation between the starting state \mathbf{x}_t and desired state \mathbf{x}_t^d only as planar translation in the XY -plane of $\{O\}$ (as illustrated in Fig. 5-(b)), the applied external force to the peg defined by Eq. (5) can be projected into the XY -plane as follows:

$$\mathbf{F}_{\text{ext}}^{\text{planar}} = \mathbf{K}_d [\Delta x, \Delta y, 0, 0, 0, 0]^\top \quad (22)$$

The formulation of $\mathbf{F}_{\text{ext}}^{\text{planar}}$ can be rewritten into a potential energy field based on the desired point $v_d \in \mathbb{R}^3$ and any point $v \in \mathbb{R}^3$ on the XY -plane of $\{O\}$ in a quadratic form:

$$U_{XOY}(v) = \frac{1}{2} (v_d - v)^\top \mathbf{K}_d|_{1:3, 1:3} (v_d - v) \quad (23)$$

As illustrated in Fig. 5-(c12), for an angle formed by the target corner vertex $v_j = T_{\{OH\}} v_h^{\text{corner}}$ and its nearby vertices, its edges are considered non-penetration walls in $U_{XOY}(v)$. Let v_t be the position of v_{lateral} at time t ; we aim to leverage such constraints to shape the motion of v_t to rest at v_j under the drive of $U_{XOY}(v)$. Specifically, our objective is to formulate a *potential well* to let v_j be the local minimum in a potential energy field so that v_t tends to rest at v_j without escaping. This process involves two fundamental requirements: 1) finding an area $\mathcal{A}_{\text{well}}$ to place the attraction point v_d (defined by $\Psi(\mathbf{x}_t^d)$) and 2) finding the *basin of attraction* $\mathcal{A}_{\text{basin}}$ where all initial states v_t inside it will eventually move to v_j . Both $\mathcal{A}_{\text{well}}$ and $\mathcal{A}_{\text{basin}}$ should only condition on the non-penetration constraints from angle $\angle v_i v_j v_k$.

We start with a straightforward example of aligning a point v_t to a determinate vertex v_j in the small interior area of an angle $\angle v_i v_j v_k$ on the XY -plane of $\{O\}$. First, the formulation of $\mathcal{A}_{\text{well}}$ is defined in Lemma 4.9. Then, we introduce the formulation of $\mathcal{A}_{\text{basin}}$ in Lemma 4.10.

Lemma 4.9. Area $\mathcal{A}_{\text{well}} \subset \mathbb{R}^3$ exist on the XY -plane of $\{O\}$ and only condition on the angle $\angle v_i v_j v_k$ (as visualized in Fig. 5-(c1): Region A), for any desired point $v_d \in \mathcal{A}_{\text{well}}$, v_j is always the local minimum state in the created basin of attraction $\mathcal{A}_{\text{basin}}^{v_d} \subset \mathbb{R}^3$ (as visualized in Fig. 5-(c2): Region B). $\mathcal{A}_{\text{well}}$ is defined as:

$$\mathcal{A}_{\text{well}} = \{v \in \mathbb{R}^3 \mid \frac{\pi}{2} \leq \angle v_i v_j v \leq (\frac{3\pi}{2} - \angle v_i v_j v_k)\} \quad (24)$$

while $\mathcal{A}_{\text{basin}}^{v_d}$ is defined as:

$$\mathcal{A}_{\text{basin}}^{v_d} = \{v \in \mathbb{R}^3 \mid \angle v_i v_d v \leq \angle v_i v_d v_k \cap \angle v_i v_j v \leq \angle v_i v_j v_k\} \quad (25)$$

Proof: The absolute distance between $v_d \in \mathcal{A}_{\text{well}}$ and v_t represents the energy level of a quadratic-formed energy field. As $\forall v_d \in \mathcal{A}_{\text{well}}, \forall v_t \in \mathcal{A}_{\text{basin}}^{v_d}, v_t \neq v_j : |v_d - v_t| > |v_d - v_j|$, thus proven v_j is the local minimum state in $\mathcal{A}_{\text{basin}}^{v_d}$, $\mathcal{A}_{\text{well}}$ and $\mathcal{A}_{\text{basin}}^{v_d}$ establish. ■

Lemma 4.10. Area $\mathcal{A}_{\text{basin}} \subset \mathbb{R}^3$ exist on the XY -plane of $\{O\}$ and only condition on the angle $\angle v_i v_j v_k$ (as visualized in Fig. 5-(c3): Region B). For any desired point $v_d \in \mathcal{A}_{\text{well}}$ defined in Eq. (24), v_j is always the local minimum state in $\mathcal{A}_{\text{basin}}$. $\mathcal{A}_{\text{basin}} = \cap \{\mathcal{A}_{\text{basin}}^{v_d} \mid \forall v_d \in \mathcal{A}_{\text{well}}\}$ is defined as

$$\mathcal{A}_{\text{basin}} = \{v \in \mathbb{R}^3 \mid \angle v v_i v_j \leq \frac{\pi}{2} \cap \angle v v_k v_j \leq \frac{\pi}{2}\} \quad (26)$$

when $|v_d - v_j| \rightarrow \infty \wedge v_d \in \mathcal{A}_{\text{well}}$

Proof: The original energy gradient of $U_{\text{XOY}}(v)$ is derived as $\mathbf{F}_{\text{energy}} = -\nabla U_{\text{XOY}}(v)$. While $|v_d - v_j| \neq 0$, any $v_t \in \mathcal{A}_{\text{basin}}$ moves along with the force gradient and will get in contact with the non-penetration wall first. As long as v_t is in contact with the wall, the component of the energy gradient $\mathbf{F}_{\text{energy}}$ that is normal to the wall is canceled out by the non-penetration constraint as illustrated in Fig. 5-(b2). Given the wall's normal vector \mathbf{n}_{wall} at the contact point, the modified force gradient is derived as $\mathbf{F}_{\text{total}} = \mathbf{F}_{\text{energy}} + \mathbf{F}_{\text{wall}} = -\nabla U_{\text{XOY}}(v) + (\nabla U_{\text{XOY}}(v) \cdot \mathbf{n}_{\text{wall}}) \mathbf{n}_{\text{wall}}$. For any v_t in contact with the wall, it performs physical gradient descent under force $\mathbf{F}_{\text{total}}$ and reaches the local minimum as the force gradient $\mathbf{F}_{\text{total}} = 0$ when \mathbf{n}_{wall} is on the same line of $\mathbf{F}_{\text{energy}}$. As proven in Lemma 4.9, the local minimum of energy field is at v_j , thus proven $\mathcal{A}_{\text{basin}}$ establish. ■

Based on $\mathcal{A}_{\text{well}}$ and $\mathcal{A}_{\text{basin}}$, the corner alignment process is introduced in Lemma 4.11 through potential well construction.

Lemma 4.11. Under the condition that energy caused by translation deviation is significantly larger than that caused by rotation deviation, by setting the desired state \mathbf{x}_t^d with its projected $v_d \in \mathcal{A}_{\text{well}}$, any starting state \mathbf{x}_t with $v_t \in \mathcal{A}_{\text{basin}}$ will automatically align with the target corner (as the lateral edge of p_v intersects with v_j at state $\mathbf{x}_{t+1} = \Pi(\mathbf{x}_t, \mathbf{x}_t^d)$).

Proof: As the command from impedance control and the target corner creates a potential well with v_j as its local minimum energy state, as proved in Lemma 4.9 and 4.10, \mathbf{x}_{t+1} automatically rests at the local minimum state as $v_{t+1} \rightarrow v_j$. ■

We relax the assumption on a determinate $v_j = T_{\{OH\}} v_h^{\text{corner}}$ and expand the potential well construction process on the probabilistic distribution of the target hole's state $P(\tilde{T}_{\{OH\}})$, as outlined in Alg. 3. K states are sampled as $\{\tilde{T}_{\{OH\},i}\}_{i=1}^K \sim P(\tilde{T}_{\{OH\}})$ and the ground truth of $T_{\{OH\}}$ is assumed to be covered the samples. By calculating the intersection area of $\tilde{\mathcal{A}}_{\text{basin}} = \cap \{\mathcal{A}_{\text{basin},i}\}_{i=1}^K$ and $\tilde{\mathcal{A}}_{\text{well}} = \cap \{\mathcal{A}_{\text{well},i}\}_{i=1}^K$ over pose samples $\{\tilde{T}_{\{OH\},i}\}_{i=1}^K$, it is guaranteed to find a sub-area as $\tilde{\mathcal{A}}_{\text{well}} \subset T_{\{OH\}} \mathcal{A}_{\text{well}}$ and $\tilde{\mathcal{A}}_{\text{basin}} \subset T_{\{OH\}} \mathcal{A}_{\text{basin}}$ for potential well construction. Theoretically, the robustness of the corner alignment is conditioned on $\tilde{\mathcal{A}}_{\text{well}}$ and $\tilde{\mathcal{A}}_{\text{basin}}$ instead of the geometric size of \mathcal{V}_h .

Algorithm 3 Nondeterministic Corner Alignment

Input: bounded region \mathcal{X}_{T1} , $P(\tilde{T}_{\{OH\}}) \leftarrow \mathcal{X}_{T1}$

- 1: Initialize samples $\{\tilde{T}_{\{OH\},i}\}_{i=1}^K \sim P(\tilde{T}_{\{OH\}})$, $T_{\{OH\}} \in \tilde{T}_{\{OH\},i}\}_{i=1}^K$, $\tilde{\mathcal{A}}_{\text{well}} = \tilde{\mathcal{A}}_{\text{basin}} = XY\text{-plane of } \{O\}$
- 2: **for** $i = 0, 1, \dots, K-1$ **do**
- 3: $\tilde{\mathcal{A}}_{\text{well}} = \tilde{\mathcal{A}}_{\text{well}} \cap \tilde{T}_{\{OH\},i} \mathcal{A}_{\text{well}}$
- 4: $\tilde{\mathcal{A}}_{\text{basin}} = \tilde{\mathcal{A}}_{\text{basin}} \cap \tilde{T}_{\{OH\},i} \mathcal{A}_{\text{basin}}$
- 5: **end for**
- 6: $v_{\text{lateral}} \leftarrow \Psi(\mathbf{x})$, $v_{\text{lateral}} \in \tilde{\mathcal{A}}_{\text{basin}}$
- 7: $v_{\text{lateral}}^d \leftarrow \Psi(\mathbf{x}^d)$, $v_{\text{lateral}}^d \in \tilde{\mathcal{A}}_{\text{well}}$
- 8: Execute interaction $c^v = (\mathbf{x}, \mathbf{x}^d)$

Insertion under Motion Constraints: As constant contact between the peg and its matching hole is established, we aim to leverage the motion constraints from the environment to synthesize insertion motions. Successful insertion motions are formulated as a sequence of interactions $S = [c_0^v, c_1^v, \dots, c_t^v]$ that connect the initial inclined state to the target peg-in-hole configuration. We keep the alignment tight by keeping the projection v_d of the desired state inside $v_d \in \mathcal{A}_{\text{well}}$. The alignment is formed by the energy field caused by the translation deviation; additionally, we aim to leverage the energy field caused by the rotation deviation to drive the insertion process.

Definition 4.12. Let $\alpha^* = \frac{\pi}{2}$, the insertion process is formulated as follows: under the condition that the corner contact is formed as a pivot point, a complete insertion is regarded as progressively turning the inclined angle of the peg approaching vertical as $\alpha \rightarrow \alpha^*$.

We describe the relative motion between the peg and its matching hole using a *contact reference frame* and a *manipulation frame*. As illustrated in Fig. 6, a *contact reference frame* $\{C\}$ is attached to the on the corner $T_{\{OH\}} v_h^{\text{corner}}$, where its origin is at $T_{\{OH\}} v_h^{\text{corner}}$ and its Z -axis shares the same direction with $\{O\}$. The positive direction of its X -axis is aligned with the rotation angle β from the inclined state of the peg on the XY -plane of $\{O\}$. The Y -axis is placed orthogonal with reference to the XZ -plane of $\{C\}$. By definition, $\{C\}$ is not fixed with reference to $\{O\}$ as its X -axis is defined by the rotation angle β of the peg's inclined state.

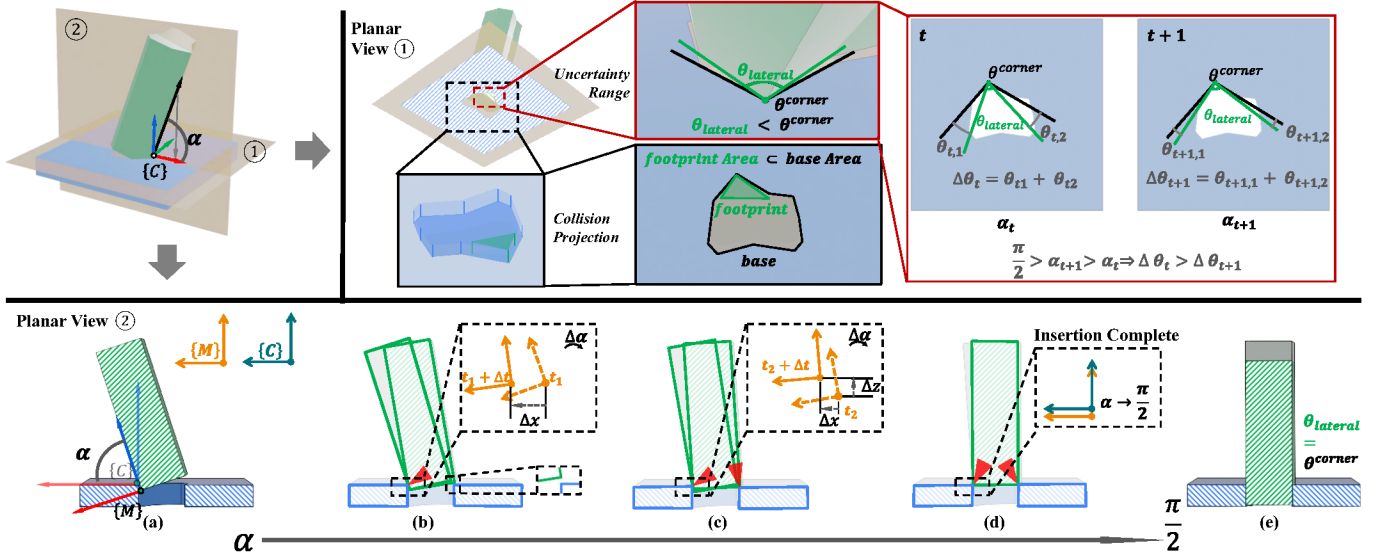


Fig. 6: An overview of the motion constraints based on the XY -plane (planar view 1) and XZ -plane (planar view 2) of $\{C\}$. **Planar View 1:** An uncertainty w.r.t to rotation is constrained by the range of $\Delta\theta$ through interactions. Meanwhile, $\Delta\theta$ is also decreasing significantly over insertion motions. **Planar View 2:** The translation of $\{M\}$ w.r.t $\{C\}$ induced by the potential energy field caused by rotation deviation and environmental constraints (red friction cones) is a passive result to balance \mathbf{F}_{ext} , thus the position of the origin of $\{M\}$ will not reach the XY -plane of $\{C\}$ as the process finishes at $\alpha \rightarrow \frac{\pi}{2}$.

Lemma 4.13. *During the insertion process (Definition 4.12), $\{C\}$ is a constrained frame with a fixed position and a limited orientation range with reference to $\{O\}$.*

Proof: As the lateral edge of p_v intersects with the corner vertex as illustrated in the *Planar View 1* from Fig. 6, the possible pose of the peg is restricted by its projection $\theta_{\text{lateral}} < \theta_{\text{corner}}$ and the uncertainty range is subject to $\Delta\theta$. Since the orientation of the inclined peg is subject to $\Delta\theta$, proving that $\{C\}$ is a constrained frame with reference to $\{O\}$. ■

Selecting a *manipulation frame* $\{M\}$ attached to the peg is critical to describe its relative motion in $\{C\}$. Since the interactions \mathbf{c}_t^v are constrained as *Definition 4.2* by a supporting vertex p_v , we set the origin of $\{M\}$ to p_v with its Z -axis's positive direction towards \mathbf{n} . The X -axis of $\{M\}$ is placed orthogonally in the plane formed by the lateral edge of p_v and the X -axis of $\{C\}$ as illustrated in the *Planar View 2* of Fig. 6. We denote the instantaneous relative motion of $\{M\}$ with reference to $\{C\}$ at time t as a twist vector $\xi_t = [\boldsymbol{\omega} \ \mathbf{d}]^T$, in which $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z] \in \mathbb{R}^3$ is the rotation rate along each axis of $\{C\}$ and $\mathbf{d} = [d_x, d_y, d_z] \in \mathbb{R}^3$ is the rate of change of the position.

Lemma 4.14. *During the insertion process (Definition 4.12), ξ_t is restricted to translation in the XZ -plane of $\{C\}$ and rotation about the Y -axis of $\{C\}$, forming a subset of $SE(3)$.*

Proof: Since the XZ -plane of $\{C\}$ and the XZ -plane of $\{M\}$ lies in the same geometric plane in $\{O\}$, thus proving that the relative motion is restricted as $\xi = [0, \omega_y, 0, d_x, 0, d_z]$. ■

Since frame $\{C\}$ is within a limited range, it can be regarded as a static frame at any instant movement ξ_t described

in *Lemma 4.15*.

Lemma 4.15. *Under the condition that the energy caused by rotation deviation is significantly larger than that caused by translation deviation, by setting the desired state's orientation as a small rotation $\omega_y \Delta t$ on the current state, the deviation $\Delta\alpha = \alpha^* - \alpha$ is decreasing, and the uncertainty range $\Delta\theta$ of $\{M\}$ is shrinking.*

Proof: As illustrated in the *Planar View 2* of Fig. 6, the translation motion \mathbf{v} is a result refined by the passive force from the environmental contacts. The peg cannot break the alignment according to *Lemma 4.8*, as the result $\{M\}$ is always lower than $\{C\}$ in the world frame. Such formulation also guarantees that an enlarged inclined angle represents a smaller potential energy state in the caused energy field; the transition between the peg's state automatically leads to a higher inclined angle under its environmental motion constraints according to *Lemma 4.8*. As illustrated in the *Planar View 1* of Fig. 6, the rotation motion ω_x and ω_z is constrained by the limited range of $\Delta\theta$, as the α increases, $\Delta\theta$ decreases. ■

By incrementally increasing the inclined angle α and updating the measurement of the inclined state's rotation angle β , the peg's state will progressively move towards the target state while its uncertainty range $\Delta\theta$ is shrinking towards zero. Theoretically, the robustness of the insertion process is conditioned on the peg's state \mathbf{x}_t instead of its geometric size.

Theorem 4.16. *The insertion process introduced by inclined angle adjustment after corner alignment is a manipulation funnel in the execution task space based on the general Definition 4.1.*

Proof: The deviation in the peg's state $\Delta s = \Delta v + \Delta \theta$ is decreasing decoupled. As Δv decreases to zero as proven in Lemma 4.11 and $\Delta \theta$ decreases to zero as proven in Lemma 4.15 over environmentally constrained interactions c_t^v , thus proving the physical manipulation funnel established. ■

Algorithm 4 MPC-based Insertion Planning

Input: system dynamics model $\tilde{\Pi}$, initial state \mathbf{x}_t , target inclined angle α^*

- 1: **while** not $|\alpha_t - \alpha^*| \rightarrow 0$ **do**
- 2: $(\mathbf{x}_t^d, \dots, \mathbf{x}_{t+T-1}^d) \leftarrow \text{ConstrainedOptimization}$ ▷
 Eq.(27)
- 3: Execute the interaction $c_t^v = (\mathbf{x}_t, \mathbf{x}_t^d)$
- 4: Re-measure \mathbf{x}_{t+1} , $\alpha_{t+1} \leftarrow \mathbf{x}_{t+1}$
- 5: Update system dynamics model $\tilde{\Pi}$ ▷ Appendix
- 6: $t \leftarrow t + 1$
- 7: **end while**

Additionally, we introduce a Model Predictive Control (MPC) framework to find the formulated insertions process S . We initialize the system transition function $\tilde{\Pi}$ with a linear prior and update it online with a Recursive Least Squares (RLS) adaptive filter (see Appendix). At time step t , the system sets a desired state \mathbf{x}_t^d , and the state of the peg transitions from \mathbf{x}_t to \mathbf{x}_{t+1} under environmental constraints. Interactions during insertion are successive, which means the resulting steady state from the previous interaction serves as the starting state for the next interaction. The cost function is defined by the inclined angle α_t at state \mathbf{x}_t and the target inclined angle α^* at \mathbf{x}^* as $\mathcal{J}(\mathbf{x}_t, \mathbf{x}^*) = |\alpha_t - \alpha^*|$. A sequence of actions $(\mathbf{x}_t^d, \dots, \mathbf{x}_{t+T-1}^d)$ is selected to minimize the defined loss over a finite horizon T as:

$$\begin{aligned}
 (\mathbf{x}_t^d, \dots, \mathbf{x}_{t+T-1}^d) &= \arg \min \sum_{i=0}^{T-1} (\mathcal{J}(\Pi(\mathbf{x}_{t+i}, \mathbf{x}_{t+i}^d), \mathbf{x}^*) + \mathbf{u}_{t+i}) \\
 \text{subject to: } \Delta \mathbf{x}_{\min} &\leq |\mathbf{x}_{t+i} - \mathbf{x}_{t+i}^d| \leq \Delta \mathbf{x}_{\max} \\
 \Psi(\mathbf{x}_{t+i}^d) &= [v_{\text{lateral}}, \theta_{\text{lateral}}], v_{\text{lateral}} \in \mathcal{A}_{\text{well}} \\
 i &= 0, \dots, N-1
 \end{aligned} \tag{27}$$

where $\mathbf{u}_{t+i} = \|\mathbf{x}_{t+i} - \mathbf{x}_{t+i}^d\|$ is added to ensure smooth control inputs by discouraging unnecessarily large or aggressive control actions. At each step t , the system applies the first selected desired state to execute and updates the transition model $\tilde{\Pi}$. An overview of the MPC framework is outlined in Alg. 4, and its convergence is proved in Lemma 4.15.

V. EXPERIMENTS

In this section, we evaluate the proposed framework across a diverse range of peg-in-hole tasks to demonstrate its robustness and generalizability. Ablation studies are conducted to gain deeper insights into how the funnel-based manipulation approach mitigates uncertainties arising from perception and execution. Specifically, these experiments aim to validate: (1) the perception manipulation funnel eliminates uncertainties in the proposal distribution of the target hole's state

while tolerating uncertainties from the imprecise interactions; (2) the physical manipulation funnel absorbs perception and execution uncertainties to achieve precise insertion; (3) by leveraging funnel-based manipulation planning, the system achieves robust and precise insertion tasks with a tolerance level surpassing the robot's inherent execution precision.

A. Experiment Setup

Our experimental setup is shown in Fig. 7-(a). We test the proposed funnel-based manipulation framework on a robotic system consisting of a Franka Emika Panda manipulator with GelSight Mini attached to the fingertips of its parallel jaw gripper. No camera is required as our system eliminates uncertainties purely through physical interactions. This design removes any system dependence on variable lighting conditions and textures, making it easily generalizable to different scenarios. It is worth noting that, though visual modality is not mandatory, it can be easily integrated as providing the initial estimation of the target hole, which can be further refined by interactive perception or directly bridged with insertion. The in-hand pose of the grasped peg is observed from its corresponding tactile imprints, as discussed in the Appendix. Our system design does not require slipping-free conditions since the proposed funnel theory is object-centric. Two types of controllers are adopted in our real-world experiments: 1) a Cartesian impedance controller with a translation error up to 2mm and 2) a Cartesian position controller with a translation error up to 1mm due to imprecise internal models. A detailed formulation of Cartesian impedance control for a manipulator is discussed in the Appendix. The Cartesian position controller is applied in position-based commands, and the Cartesian impedance controller is applied in commands requiring physical interactions. Under external contacts, the object undergoes only minor elastic movements on the tactile sensor's planar surface, while major pose deviations are regulated by the robot's Cartesian impedance controller through the end-effector.

The system is tested on a NIST ATB benchmark and additional tight-clearance tasks, with detailed parameters of the peg-in-hole tasks shown in Fig. 8. The experimental pegs are selected with different base symmetries (ranging from central symmetry and axial symmetry to asymmetry), materials, scales, and clearance levels for a comprehensive evaluation. It is also worth noting that, although *chamfer*¹ on the peg and the hole would make the insertion process easier, as it relatively enlarges the entry space, it is excluded in our experiments for a more challenging task formulation.

B. Perception Manipulation Funnel

We conducted detailed experiments in PyBullet [10] to evaluate the robustness of the perception manipulation funnel for uncertainty elimination in the perception state space (as

¹In the context of peg-in-hole tasks, the chamfer refers to the beveled or tapered edge at the entrance of the hole (and sometimes at the tip of the peg). A chamfer added to the edges can help reduce contact forces, reduce jamming and tolerate misalignment.

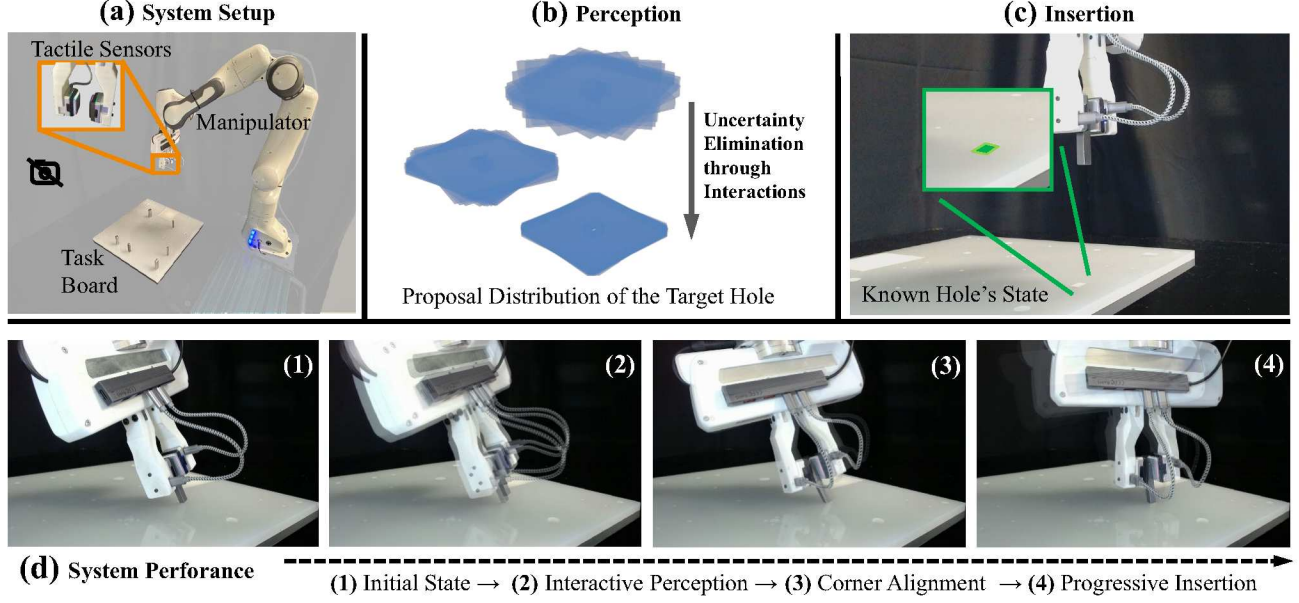


Fig. 7: (a) Overview of the System Setup; (b) Ablation study on the perception manipulation funnel; (c) Ablation study on the physical manipulation funnel; (d) Performance evaluation on the overall funnel-based manipulation system.



Name	Scale*	Clearance*	Material	Symmetry
Round 8	$\Phi 8$	~ 0.8	metal	central
Round 12	$\Phi 12$	~ 0.8	metal	central
Round 16	$\Phi 16$	~ 0.8	metal	central
Rectangle 8	8×7	~ 0.6	metal	axial
Rectangle 12	12×8	~ 0.7	metal	axial
Rectangle 16	16×10	~ 0.8	metal	axial
Random #1**	20×16	~ 0.4	polylactide	asymmetric
Random #2**	22×25	~ 0.4	polylactide	asymmetric
Random #3**	23×17	~ 0.4	polylactide	asymmetric

* Scale and clearance are denoted in millimeters.

** Pegs with random generated polygonal base; scales are described with the bounding box of the base.

Fig. 8: Overview of the Peg-in-Hole Tasks in Real-world Experiments.

illustrated in Fig. 7-(b)). Specifically, we aim to demonstrate that 1) the perception manipulation funnel is robust against a large range of action uncertainties and 2) the entropy-based exploration significantly accelerates the uncertainty elimination process.

To quantify the uncertainty level of actions, given a desired interaction point $\mathbf{g}_d = (x_d, y_d, 0)$ on the board surface plane (XY -plane of $\{O\}$), the actual interacted point $\mathbf{g}_a = (x_a, y_a, 0)$ incorporates additional positional noise as

action uncertainty from a 2-D Gaussian distribution as follows:

$$f(x_a, y_a) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \Sigma^{-1} \mathbf{z}\right) \quad (28)$$

$$\mathbf{z} = \begin{bmatrix} x_a - x_d \\ y_a - y_d \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sigma \quad (29)$$

where σ represents the noise level of distance. We use the Jaccard index to calculate the distance between the ground truth hole area and the union of sampled areas $\{\tilde{\mathcal{A}}\} = \{\tilde{T}_{\{OH\},i} \cap \mathcal{A}_h\}_{i=1}^{K=200} \sim P(\tilde{T}_{\{OH\}})$ based on the refined proposal distribution as the convergence level:

$$J(\{\tilde{\mathcal{A}}\}) = \frac{|T_{\{OH\}} \cap \mathcal{A}_h \cap \cup\{\tilde{\mathcal{A}}\}|}{|T_{\{OH\}} \cap \mathcal{A}_h \cup \cup\{\tilde{\mathcal{A}}\}|} \approx \frac{|\mathcal{A}_h|}{|\cup\{\tilde{\mathcal{A}}\}|} \quad (30)$$

Random exploration serves as a baseline policy to demonstrate that 1) the intrinsic convergence mechanics of the perception manipulation does not rely on any specific action strategy and 2) the proposed entropy-based exploration benefits the convergence efficiency.

The results, presented in Fig. 9, show the system's performance after eight steps of interaction under different levels of action noises and a random policy. Pegs with smaller scales are more sensitive to the system's uncertainty, as the relative error is significantly larger than the pegs with larger scales. Despite high action noise relative to the peg's scale, all the test objects successfully converge to a high convergence level. Even random actions would significantly shrink the proposal distribution of the target hole due to the mechanics of the formulated perception funnel.

C. Physical Manipulation Funnel

We evaluate the robustness of the physical manipulation funnel through peg insertion into the target hole as shown

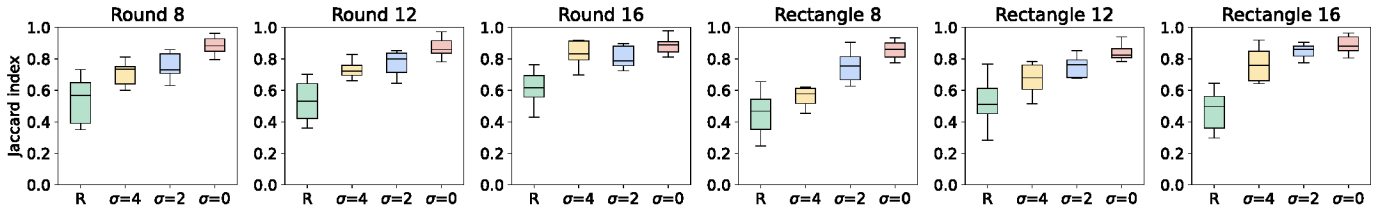


Fig. 9: Experimental result of the uncertainty elimination process in simulation with different levels of action uncertainty and object scales. “R” denotes the random policy; σ denotes our entropy-based policy with different uncertainty levels in millimeters.

Name	Clearance	Position-based ¹	Funnel-based ²
Round 8	~ 0.8	4/10	10/10
Round 12	~ 0.8	5/10	10/10
Round 16	~ 0.8	5/10	10/10
Rectangle 8	~ 0.6	3/10	10/10
Rectangle 12	~ 0.7	5/10	10/10
Rectangle 16	~ 0.8	4/10	10/10
Random #1	~ 0.4	1/10	9/10
Random #2	~ 0.4	0/10	8/10
Random #3	~ 0.4	0/10	10/10
Average		3.0/10	9.7/10

¹ Baseline: Actuated with Cartesian position controller.

² Ours: Actuated with Cartesian impedance controller.

TABLE I: Experimental Result of Real-world Insertion Tasks

in Fig. 7-(c), which is considered a known position. We compare the robustness of funnel-based manipulation with position-based manipulation as a baseline, which formulates the insertion process by positioning the peg into the hole without collisions. The collision-free trajectory is defined as a top-down insertion with a vertical peg state after a stable grasp. Since execution uncertainties are inevitable in real-world scenarios, we aim to demonstrate that 1) when system uncertainty exceeds the allowed clearance, funnel-based manipulation for peg insertion is more robust against position-based manipulation in real-world settings, and 2) precise manipulation tasks can be achieved through imprecise funnel-based manipulation.

The experimental result is shown in Table I. Despite the tight clearance being smaller than the robot execution precision, our funnel-based insertion effectively absorbs unmodeled uncertainties and inserts the peg at a *near-perfect* success rate. In contrast, position-based manipulation expects a precisely executed trajectory while lacking the ability to tolerate such uncertainty. Without the help of chamfers on the peg and the hole, it can easily fail due to the unexpected misalignments or impact caused by the wedging phenomenon [53]. Especially for asymmetric pegs with a tighter tolerance in both translation and orientation, we observe that position-based control is ineffective at small scales.

D. Real-world System Performance

Ultimately, we combined the perception and physical manipulation funnel in a unified system to perform robust insertion without visual feedback in an open loop, as illustrated in Fig. 7-(d). To comprehensively evaluate the proposed system,

we consider two levels of prior on the target hole’s state: 1) the initial pose of the peg is placed partially inside the target hole by kinesthetic teaching and 2) a larger search area (bounding circle with $\sim 30\%$ dilation) is specified, containing the target hole on the board surface. The robot needs to identify the possible state of the target hole on the board surface to finish the insertion. The results (as presented in Table II) depict the steps of interactions, uncertainty level ($1 - J(\{\tilde{\mathcal{A}}\})$) and the successful rate for each tested peg. Since the proposed system effectively absorbs uncertainties from an imprecise target hole state and interactions, we observe very few failure cases from the overall performance.

We acknowledge the gap between the established object-centric theory and real-world implementation; the failure mode can be divided into the following categories: 1) high contact force which breaks the condition of in-hand stable grasp, i.e., the gripper is opened and the peg beyond the observation range of the tactile sensor; 2) high friction force caused by interactions between the peg and the hole that causes the peg largely slipped and is out of the tactile perception range and 3) the inaccurate dynamics model of the manipulator for impedance control that leads to unsatisfied compliance performance during physical contacts. We believe that using a more forceful end-effector or a more advanced compliance mechanism would effectively resolve this issue.

VI. LIMITATIONS

Our work has certain limitations, which naturally reveal the potential future research directions. First, the current system relies on a stable in-hand grasp of the peg to perform the insertion process. A re-grasping policy [4] can be incorporated to initialize a new insertion trial for drastic slippage that exceeds the tactile perception range. Second, current exploration relies on a priori of the planar exploratory area. By combining visual modalities and learning-based estimation methods, the system can perform environmental identification and insertion without any prior. Third, as our method doesn’t model friction explicitly, a variable impedance controller is considered a requirement when encountering interaction with high resistance.

We envision a future iteration of the proposed robotic system that leverages multi-modal perception to partially perceive the task environment and further plan funnel-based manipulation strategies to robustly shape the interaction process toward

TABLE II: System Performance under Different Prior Knowledge on the Target Hole

Object	Partially Inside the Hole			Bounded Area			Overall
	Success	Interactions (\pm std)	Uncertainty(\pm std)	Success	Interactions (\pm std)	Uncertainty(\pm std)	Success
Round 8	5 / 5	4.0 \pm 0.7	0.152 \pm 0.029	5/5	4.6 \pm 1.5	0.148 \pm 0.040	10/10
Round 12	5 / 5	4.4 \pm 1.1	0.155 \pm 0.048	5/5	5.0 \pm 1.7	0.150 \pm 0.036	10/10
Round 16	5 / 5	4.6 \pm 0.9	0.163 \pm 0.060	5/5	5.6 \pm 1.5	0.162 \pm 0.063	10/10
Rectangle 8	5 / 5	6.8 \pm 1.6	0.178 \pm 0.031	4/5	7.8 \pm 1.9	0.174 \pm 0.043	9/10
Rectangle 12	5 / 5	6.8 \pm 1.3	0.174 \pm 0.046	5/5	7.6 \pm 1.5	0.170 \pm 0.033	10/10
Rectangle 16	5 / 5	7.2 \pm 0.8	0.178 \pm 0.031	5/5	7.8 \pm 2.1	0.193 \pm 0.055	10/10
Random #1	5 / 5	9.2 \pm 2.2	0.172 \pm 0.029	4/5	9.0 \pm 2.9	0.158 \pm 0.036	9/10
Random #2	4 / 5	9.4 \pm 2.7	0.164 \pm 0.062	4/5	10.2 \pm 1.9	0.181 \pm 0.065	8/10
Random #3	4 / 5	9.0 \pm 2.9	0.170 \pm 0.045	5/5	9.6 \pm 1.8	0.178 \pm 0.061	9/10
Average	4.78 / 5	6.8 \pm 2.6	0.167 \pm 0.041	4.67/5	7.5 \pm 2.6	0.168 \pm 0.046	9.44/10

the desired outcome beyond the peg-in-hole insertion task.

VII. CONCLUSION AND DISCUSSION

In this paper, we have presented a funnel-based manipulation paradigm for robust peg-in-hole insertion, along with a real-world robotic implementation. The advantages of the proposed system have been demonstrated through careful experimentation in simulation and in the real world. A comprehensive evaluation of the proposed system is provided on a standard NIST ATB benchmark and additional challenging tasks with a satisfactory overall success rate. Most critically, we highlighted the power of compliance and demonstrated how a robot can leverage imprecise compliant interactions to perform precise tasks, similar to how humans achieve dexterous manipulation. For future work, we aim to expand the current theory to non-planar scenarios and develop automated learning methods capable of abstracting funnel-based manipulation mechanics from natural language descriptions or video demonstrations and their interaction mechanisms.

ACKNOWLEDGMENTS

This work was supported by the award 60NANB24D296 from U.S. Department of Commerce, National Institute of Standards and Technology.

REFERENCES

- [1] Srinivas Akella, Wesley H Huang, Kevin M Lynch, and Matthew T Mason. Parts feeding on a conveyor with a one joint robot. *Algorithmica*, 26:313–344, 2000.
- [2] Aditya Bhatt, Adrian Sieler, Steffen Puhlmann, and Oliver Brock. Surprisingly Robust In-Hand Manipulation: An Empirical Study. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi: 10.15607/RSS.2021.XVII.089.
- [3] Robert R Burridge, Alfred A Rizzi, and Daniel E Koditschek. Sequential composition of dynamically dexterous robot behaviors. *The International Journal of Robotics Research*, 18(6):534–555, 1999.
- [4] Roberto Calandra, Andrew Owens, Dinesh Jayaraman, Justin Lin, Wenzhen Yuan, Jitendra Malik, Edward H Adelson, and Sergey Levine. More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robotics and Automation Letters*, 3(4):3300–3307, 2018.
- [5] Alper Canberk, Cheng Chi, Huy Ha, Benjamin Burchfiel, Eric Cousineau, Siyuan Feng, and Shuran Song. Cloth funnels: Canonicalized-alignment for multi-purpose garment manipulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5872–5879. IEEE, 2023.
- [6] Sirui Chen, Jeannette Bohg, and Karen Liu. SpringGrasp: Synthesizing Compliant, Dexterous Grasps under Shape Uncertainty. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024. doi: 10.15607/RSS.2024.XX.042.
- [7] Yiting Chen, Ahmet Ercan Tekden, Marc Peter Deisenroth, and Yasemin Bekiroglu. Sliding touch-based exploration for modeling unknown object shape with multi-fingered hands. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8943–8950. IEEE, 2023.
- [8] Xianyi Cheng, Eric Huang, Yifan Hou, and Matthew T Mason. Contact mode guided motion planning for quasidynamic dexterous manipulation in 3d. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2730–2736. IEEE, 2022.
- [9] Alan D Christiansen. Manipulation planning for empirical backprojections. In *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pages 762–763. IEEE Computer Society, 1991.
- [10] Erwin Coumans and Yunfei Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning, 2016.
- [11] Nikhil Chavan Daffe, Alberto Rodriguez, Robert Paolini, Bowei Tang, Siddhartha S Srinivasa, Michael Erdmann, Matthew T Mason, Ivan Lundberg, Harald Staab, and Thomas Fuhlbrigge. Extrinsic dexterity: In-hand manipulation with external forces. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1578–1585. IEEE, 2014.
- [12] Raphael Deimel and Oliver Brock. A novel type of compliant and underactuated robotic hand for dexterous grasping. *The International Journal of Robotics Research*, 35(1-3):161–185, 2016.
- [13] Siyuan Dong, Devesh K Jha, Diego Romeres, Sangwoon Kim, Daniel Nikovski, and Alberto Rodriguez. Tactile-rl for insertion: Generalization to objects of unknown

- geometry. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6437–6443. IEEE, 2021.
- [14] Clemens Eppner and Oliver Brock. Planning grasp strategies that exploit environmental constraints. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 4947–4952. IEEE, 2015.
- [15] Clemens Eppner, Raphael Deimel, José Alvarez-Ruiz, Marianne Maertens, and Oliver Brock. Exploitation of environmental constraints in human and robotic grasping. *The International Journal of Robotics Research*, 34(7):1021–1038, 2015.
- [16] Michael Erdmann. Using backprojections for fine motion planning with uncertainty. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 549–554. IEEE, 1985.
- [17] Michael A Erdmann and Matthew T Mason. An exploration of sensorless manipulation. *IEEE Journal on Robotics and Automation*, 4(4):369–379, 1988.
- [18] Kenneth Y Goldberg. Orienting polygonal parts without sensors. *Algorithmica*, 10(2):201–225, 1993.
- [19] Kaiyu Hang, Miao Li, Johannes A Stork, Yasemin Bekiroglu, Florian T Pokorny, Aude Billard, and Danica Kragic. Hierarchical fingertip space: A unified framework for grasp planning and in-hand grasp adaptation. *IEEE Transactions on robotics*, 32(4):960–972, 2016.
- [20] Kaiyu Hang, Andrew S Morgan, and Aaron M Dollar. Pre-grasp sliding manipulation of thin objects using soft, compliant, or underactuated hands. *IEEE Robotics and Automation Letters*, 4(2):662–669, 2019.
- [21] Rasmus Haugaard, Jeppe Langaa, Christoffer Sloth, and Anders Buch. Fast robust peg-in-hole insertion with continuous visual servoing. In *Conference on Robot Learning*, pages 1696–1705. PMLR, 2021.
- [22] Neville Hogan. Impedance control: An approach to manipulation. In *1984 American control conference*, pages 304–313. IEEE, 1984.
- [23] Yifan Hou, Zhenzhong Jia, and Matthew T Mason. Manipulation with shared grasping. *arXiv preprint arXiv:2006.02996*, 2020.
- [24] Tadanobu Inoue, Giovanni De Magistris, Asim Munawar, Tsuyoshi Yokoya, and Ryuki Tachibana. Deep reinforcement learning for high precision assembly tasks. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 819–825. IEEE, 2017.
- [25] Shiyu Jin, Xinghao Zhu, Changhao Wang, and Masayoshi Tomizuka. Contact pose identification for peg-in-hole assembly under uncertainties. In *2021 American Control Conference (ACC)*, pages 48–53. IEEE, 2021.
- [26] Farshad Khadivar and Aude Billard. Adaptive fingers coordination for robust grasp and in-hand manipulation under disturbances and unknown dynamics. *IEEE Transactions on Robotics*, 39(5):3350–3367, 2023.
- [27] Sangwoon Kim and Alberto Rodriguez. Active extrinsic contact sensing: Application to general peg-in-hole insertion. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 10241–10247. IEEE, 2022.
- [28] Kenneth Kimble, Karl Van Wyk, Joe Falco, Elena Messina, Yu Sun, Mizuho Shibata, Wataru Uemura, and Yasuyoshi Yokokohji. Benchmarking protocols for evaluating small parts robotic assembly systems. *IEEE robotics and automation letters*, 5(2):883–889, 2020.
- [29] Mikko Lauri, David Hsu, and Joni Pajarinen. Partially observable markov decision processes in robotics: A survey. *IEEE Transactions on Robotics*, 39(1):21–40, 2022.
- [30] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.
- [31] Miao Li, Yasemin Bekiroglu, Danica Kragic, and Aude Billard. Learning of grasp adaptation through experience and tactile sensing. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3339–3346. Ieee, 2014.
- [32] Miao Li, Hang Yin, Kenji Tahara, and Aude Billard. Learning object-level impedance control for robust grasping and dexterous manipulation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6784–6791. IEEE, 2014.
- [33] Miao Li, Kaiyu Hang, Danica Kragic, and Aude Billard. Dexterous grasping under shape uncertainty. *Robotics and Autonomous Systems*, 75:352–364, 2016.
- [34] Tomas Lozano-Perez, Matthew T Mason, and Russell H Taylor. Automatic synthesis of fine-motion strategies for robots. *The International Journal of Robotics Research*, 3(1):3–24, 1984.
- [35] Jianlan Luo, Eugen Solowjow, Chengtao Wen, Juan Aparicio Ojea, and Alice M Agogino. Deep reinforcement learning for robotic assembly of mixed deformable and rigid objects. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2062–2069. IEEE, 2018.
- [36] Jianlan Luo, Eugen Solowjow, Chengtao Wen, Juan Aparicio Ojea, Alice M Agogino, Aviv Tamar, and Pieter Abbeel. Reinforcement learning on variable impedance controller for high-precision robotic assembly. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3080–3087. IEEE, 2019.
- [37] Jianlan Luo, Oleg Sushkov, Rugile Pevceviute, Wenzhao Lian, Chang Su, Mel Vecerik, Ning Ye, Stefan Schaal, and Jon Scholz. Robust multi-modal policies for industrial assembly via reinforcement learning and demonstrations: A large-scale study. *arXiv preprint arXiv:2103.11512*, 2021.
- [38] Jianlan Luo, Zheyuan Hu, Charles Xu, You Liang Tan, Jacob Berg, Archit Sharma, Stefan Schaal, Chelsea Finn, Abhishek Gupta, and Sergey Levine. Serl: A software suite for sample-efficient robotic reinforcement learning. *arXiv preprint arXiv:2401.16013*, 2024.
- [39] Matthew Mason. The mechanics of manipulation. In

- Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 544–548. IEEE, 1985.
- [40] Andrew S Morgan, Bowen Wen, Junchi Liang, Abdeslam Boularias, Aaron M Dollar, and Kostas Bekris. Vision-driven compliant manipulation for reliable, high-precision assembly tasks. *arXiv preprint arXiv:2106.14070*, 2021.
- [41] Andrew S Morgan, Kaiyu Hang, Bowen Wen, Kostas Bekris, and Aaron M Dollar. Complex in-hand manipulation via compliance-enabled finger gaiting and multi-modal planning. *IEEE Robotics and Automation Letters*, 7(2):4821–4828, 2022.
- [42] Lael U Odhner, Leif P Jentoft, Mark R Claffee, Nicholas Corson, Yaroslav Tenzer, Raymond R Ma, Martin Buehler, Robert Kohout, Robert D Howe, and Aaron M Dollar. A compliant, underactuated hand for robust manipulation. *The International Journal of Robotics Research*, 33(5):736–752, 2014.
- [43] Luka Peternel, Nikos Tsagarakis, Darwin Caldwell, and Arash Ajoudani. Robot adaptation to human physical fatigue in human–robot co-manipulation. *Autonomous Robots*, 42:1011–1021, 2018.
- [44] Michael Posa, Cecilia Cantu, and Russ Tedrake. A direct method for trajectory optimization of rigid bodies through contact. *The International Journal of Robotics Research*, 33(1):69–81, 2014.
- [45] Alberto Rodriguez. The unstable queen: Uncertainty, mechanics, and tactile feedback. *Science Robotics*, 6(54): eabi4667, 2021.
- [46] Lin Shao, Toki Migimatsu, and Jeannette Bohg. Learning to scaffold the development of robotic manipulation skills. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5671–5677. IEEE, 2020.
- [47] Yifei Simon Shao, Tianyu Li, Shafagh Keyvanian, Pratik Chaudhari, Vijay Kumar, and Nadia Figueroa. Constraint-Aware Intent Estimation for Dynamic Human-Robot Object Co-Manipulation. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024. doi: 10.15607/RSS.2024.XX.028.
- [48] Jianbo Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.
- [49] Sergio Natalio SimunoviÄ SimunoviÄ. *An information approach to parts mating*. PhD thesis, Massachusetts Institute of Technology, 1979.
- [50] Bingjie Tang, Iretiayo Akinola, Jie Xu, Bowen Wen, Ankur Handa, Karl Van Wyk, Dieter Fox, Gaurav S. Sukhatme, Fabio Ramos, and Yashraj Narang. AutoMate: Specialist and Generalist Assembly Policies over Diverse Geometries. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024. doi: 10.15607/RSS.2024.XX.064.
- [51] Te Tang, Hsien-Chung Lin, Yu Zhao, Wenjie Chen, and Masayoshi Tomizuka. Autonomous alignment of peg and hole by force/torque measurement for robotic assembly. In *2016 IEEE international conference on automation science and engineering (CASE)*, pages 162–167. IEEE, 2016.
- [52] Marc Toussaint, Nathan Ratliff, Jeannette Bohg, Ludovic Righetti, Peter Englert, and Stefan Schaal. Dual execution of optimized contact interaction trajectories. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 47–54. IEEE, 2014.
- [53] Daniel E Whitney et al. Quasi-static assembly of compliantly supported rigid parts. *Journal of Dynamic Systems, Measurement, and Control*, 104(1):65–77, 1982.
- [54] Liang Xie, Hongxiang Yu, Yinghao Zhao, Haodong Zhang, Zhongxiang Zhou, Minhang Wang, Yue Wang, and Rong Xiong. Learning to fill the seam by vision: Sub-millimeter peg-in-hole on unseen shapes in real world. In *2022 International conference on robotics and automation (ICRA)*, pages 2982–2988. IEEE, 2022.
- [55] Xiang Zhang, Liting Sun, Zhian Kuang, and Masayoshi Tomizuka. Learning variable impedance control via inverse reinforcement learning for force-related tasks. *IEEE Robotics and Automation Letters*, 6(2):2225–2232, 2021.
- [56] Wenxuan Zhou and David Held. Learning to grasp the ungraspable with emergent extrinsic dexterity. In *Conference on Robot Learning*, pages 150–160. PMLR, 2023.

APPENDIX

A. Cartesian Impedance Controller

The state of Franka Emika Panda in the configuration space at time t is denoted as $\mathbf{q}_t \in \mathbb{R}^7$, with its velocity as $\dot{\mathbf{q}}_t \in \mathbb{R}^7$ and acceleration as $\ddot{\mathbf{q}}_t \in \mathbb{R}^7$. Given the desired state \mathbf{x}_t^d and current state \mathbf{x}_t of the manipulated frame, the robot motion is regulated by the Cartesian impedance controller based on the simulated force \mathbf{F} as:

$$\mathbf{M}(\mathbf{q}_t)\ddot{\mathbf{q}}_t + \mathbf{C}(\mathbf{q}_t, \dot{\mathbf{q}}_t)\dot{\mathbf{q}}_t + \mathbf{g}(\mathbf{q}_t) = \mathbf{J}(\mathbf{q}_t)^\top \mathbf{F} \quad (31)$$

where $\mathbf{M}(\mathbf{q}_t) \in \mathbb{R}^{7 \times 7}$ is the inertial matrix, $\mathbf{C}(\mathbf{q}_t, \dot{\mathbf{q}}_t)$ is the Coriolis and centrifugal matrix, $\mathbf{g}(\mathbf{q}_t)$ is the gravity vector and $\mathbf{J}(\mathbf{q}_t)^\top \in \mathbb{R}^{7 \times 6}$ is the transpose of the Jacobian matrix.

B. Tactile Pose Estimation

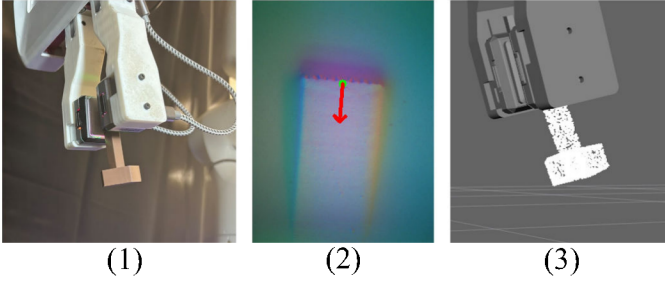


Fig. 10: (1) A stable grasp of the peg for insertion; (2) The corresponding tactile image with the object's imprint; (3) The estimated in-hand pose of the peg.

We estimate the peg's in-hand pose based on its tail end's tactile imprint (as shown in Fig. 10). After a stable grasping, the motion of the object is limited in the SE(2) space as $[x, y, \omega]$ on the planar gel surface. We identify the keypoints from the corner of the peg in the tactile image with a gradient-based feature extraction method [48]. As the connecting segment between the extracted corners represents the base of the peg, the peg's position is defined by its center, and the peg's orientation is perpendicular to the formed segment. To alleviate the perception noise for a stable estimation, we apply an Extended Kalman Filter (EKF) to smooth the estimated result and only take it into consideration when it reaches stability. As long as the peg after slippage does not exceed the observation range of the fingertip tactile sensors, the estimated in-hand pose is further leveraged for object-centric manipulation planning. For our experiments, the perception range of the Gelsight Mini tactile sensor is around $25\text{mm} \times 20\text{mm}$, and the end of the peg is required to show up completely.

C. System Transition Function for MPC

We model the system transition function with a linear prior, as if there is no external contact, the steady state will reach the desired state during each interaction. The discrete linear state-space model of the system is formulated as:

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{u}_t \text{ where } \mathbf{u}_t = \mathbf{x}_t^d - \mathbf{x}_t \quad (32)$$

where $\mathbf{A} \in \mathbb{R}^{6 \times 6}$ is the state transition matrix and $\mathbf{B} \in \mathbb{R}^{6 \times 6}$ is the input matrix. Both \mathbf{A} and \mathbf{B} are initialized as an Identity matrix.

However, under the un-modeled contact after corner alignment, the steady state of the peg is the result of the interaction between the potential from the impedance control and its environmental constraints. We aim to approximate the intricate transition function online, as \mathbf{A} and \mathbf{B} from Eq. (32) are updated by an RLS adaptive filter. The system dynamics is rewritten as:

$$\mathbf{x}_{t+1} = \Phi_t \gamma_t + \epsilon_t, \text{ where } \epsilon_t \in \mathbb{R}^6$$

$$\Phi_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix}^\top \in \mathbb{R}^{6 \times 12}, \quad \gamma_t = [\mathbf{A}_t \quad \mathbf{B}_t] \in \mathbb{R}^{12 \times 6} \quad (33)$$

Φ_t is the regression vector, γ_t is the system to update and ϵ_t is the prediction error. Given a covariance matrix \mathbf{P}_t and a forgetting factor λ , the gain \mathbf{K}_t is calculated based on the prediction error after each interaction as:

$$\epsilon_t = \mathbf{x}_{t+1} - \Phi_t \gamma_t \quad (34)$$

$$\mathbf{K}_t = \mathbf{P}_t \Phi_t^\top (\lambda + \Phi_t \mathbf{P}_t \Phi_t^\top)^{-1} \quad (35)$$

and the system parameters are updated as:

$$\gamma_{t+1} = \gamma_t + \mathbf{K}_t \epsilon_t \quad (36)$$

$$\gamma_{t+1} = [\mathbf{A}_{t+1} \quad \mathbf{B}_{t+1}] \quad (37)$$

The covariance matrix is then updated as:

$$\mathbf{P}_{t+1} = \frac{1}{\lambda} (\mathbf{P}_t - \mathbf{K}_t \Phi_t \mathbf{P}_t) \quad (38)$$

Thus, the approximated system transition function $\tilde{\Pi}_t(\mathbf{x}_t, \mathbf{x}_t^d)$ is defined as:

$$\tilde{\Pi}_t(\mathbf{x}_t, \mathbf{x}_t^d) = \mathbf{A}_t \mathbf{x}_t + \mathbf{B}_t (\mathbf{x}_t^d - \mathbf{x}_t) \quad (39)$$

and updated online to identify the underlying dynamics.

D. Computational Complexity Analysis

We perform an additional analysis of the computational efficiency of the proposed system. For Alg. 2, we adopted a rejection-based sampling algorithm with the theoretical time complexity per accepted sample as $\mathcal{O}(M)$ (M is the reciprocal of the acceptance rate conditioned on the gathered inequality constraints defined in Definition 4.3). Alg. 3 calculates the intersection of the union over a constant number of samples and thus possesses a time complexity of $\mathcal{O}(1)$. For Alg. 4, since we are using a linear state-space model as defined in Eq. (39), which leads to a convex QP problem that can be efficiently solved, the MPC planner runs at $\sim 30\text{Hz}$.