2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

# Acoustic Peg-in-hole assembly with Phased Transducer Array and Microscope

Saida Yang, Chengxi Zhong, Yujie Jiang, Qun Gao, Hu Su*, and Song Liu*, *Member, IEEE*

***Abstract*— Microassembly is becoming increasingly critical in modern smart manufacturing, placing higher demands on system performance—particularly for *in situ* and *in vivo* applications in biomedicine, photonics, sensors, and microrobotics. Non-contact mechanical microassembly has emerged as a promising solution, addressing challenges such as part contamination, limited environmental compatibility, and undesired microscopic forces. This paper presents an ultrasonic-driven non-contact microassembly system capable of performing a representative peg-in-hole assembly. The system primarily consists of an ultrasonic phased transducer array, which serves as a holographic acoustic end-effector, and a microscope that provides visual feedback. Elliptical and semi-elliptical holographic acoustic end-effectors are designed and generated by generative adversarial networks. A homogeneous transformation strategy is employed to generate pre-planned phase-only hologram (POH) sequence, while a closed-loop control strategy dynamically adjusts the end-effector's pose by integrating real-time visual feedback. Experimental results demonstrate that, with disturbances compensated by the closed-loop strategy, the system can stably adjust the peg's position and orientation to achieve acceptable alignment accuracy. It successfully manipulates high-aspect-ratio objects to complete the peg-in-hole assembly in fluidic and strong magnetic environments. Moreover, the system requires no preset object position or orientation and does not alter the object's form or structure during operation, demonstrating strong potential for broader *in situ* applications.**

## I. Introduction

Microassembly [1] refers to the process of precisely manipulating, positioning and connecting micro-parts at a microscopic scale to assemble functional devices or systems. It has been widely used in many fields such as biomedical equipment [2], micro-optical systems [3], high-precision sensors [4] and micro-robots [5]. Among various microassembly tasks, peg-in-hole microassembly [6] is one of the most typical, fundamental and critical task [7]. As pointed by [8], nearly half of the assembly work can be considered as peg-in-hole assembly tasks, at the same time, its assembly quality and reliability directly influence the performance, lifetime, and functionality of the resulting micro-systems.

Conventional peg-in-hole microassembly methods are predominantly contact-based, utilizing micro-grippers [9], probes [10], or other mechanical end-effectors [11] to manipulate objects. While widely used, these techniques face persistent challenges of assembly part contamination risk [12], limited adaptability to different environments (especially liquid [13] or confined spaces [14]), and susceptibility to adverse microscopic forces such as electrostatic, capillary, and van der Waals interactions [15]. These forces become dominant at microscopic scales and often lead to unstable grasping, sticking, or positioning errors, thus compromising assembly quality and reliability. In addition, the peg-in-hole assembly is becoming even more challenging in unstructured, complex, and confined environments, such as strong magnetic fields, *in vivo* conditions [16], and microfluidic systems. To overcome these limitations, non-contact manipulation methods are employed to perform the peg-in-hole microassembly task, enabling reliable assembly with in situ [17] capability and adaptability to diverse environments.

Current non-contact manipulation strategies using external fields such as optical [18], magnetic [19], acoustic [20], or electric fields [21], have been explored. Among these, acoustic manipulation, particularly through ultrasonic standing waves [22] or phased arrays [23], has attracted growing interest due to its compatibility with liquid, strong magnetic, and biological environments, minimal impact on objects, and ability to exert controllable forces without physical contact. Existing acoustic tweezers and levitation systems [24] have demonstrated promising capabilities in trapping, translating, and rotating micro-particles. However, through experiment we found that many current acoustic end-effector (such as single-focal acoustic trap, dual-focal acoustic trap, and vortex acoustic trap) lack flexibility to manipulate slender peg, which usually has high-aspect-ratio, especially when the dimension is larger than the wavelength of the acoustic wave, which hinders the development of high-precision and stable acoustic peg-in-hole assembly system.

This paper was supported by National Natural Science Foundation of China under Grant 62303321 and T2522022. (Corresponding author: *Hu Su* and *Song Liu*).

S. Yang and H. Su are with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. S. Yang is also with School of Artificial Intelligence, University of Chinese Academy of Science, Beijing 100049, China. (e-mail: yangsaida2025@ia.ac.cn, hu.su@ia.ac.cn).

C. Zhong, Y. Jiang and S. Liu are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China. S. Liu is also with the Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai 201210, China. (e-mail: zhongchx@shanghaitech.edu.cn, jiangyj12023@shanghaitech.edu.cn liusong@shanghaitech.edu.cn).

Q. Gao is with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China. (e-mail: qungao@ciomp.ac.cn)

This paper proposes an ultrasonic-driven non-contact microassembly system designed for *in situ*, multi-environment compatible, and contamination-free peg-in-hole microassembly. The system mainly consists of an ultrasonic phased transducer array capable of generating reprogrammable holographic acoustic fields that function as non-contact end-effectors and a microscope used to provide real-time visual feedback. The acoustic end-effectors are designed as elliptical and semi-elliptical trapping profiles respectively for peg pose-position adjustment and insertion, which are optimized using generative adversarial networks. A homogeneous transformation strategy and a vision-based closed-loop control strategy [25] are used for motion planning and real-time adjustment of the peg's position and orientation. The vision-based closed-loop control strategy includes dual-mode regulation: integral tracking for precise alignment and compensatory adjustment for disturbance recovery, enhancing operational robustness. Experimental validations confirm the effective of proposed system in peg-in-hole microassembly across diverse settings, including strong magnetic environments.

The rest of this paper is organized as follows: Section II introduces the proposed acoustic manipulation system for automated microassembly of peg-in-hole, along with its calibration and task specification. Section III discusses the acoustic end-effector design and simulation. Section IV decomposes the assembly process and details the vision alignment algorithm and insertion control strategy. Section V validates the proposed system through experiments.

## II. System Configuration and Calibration

### A. Task Specification

To demonstrate the capability of the acoustic end effector proposed in this paper to non-contact manipulate slender peg, we design an ultrasonic-driven automatic manipulation platform to complete a two-dimensional peg-in-hole microassembly task, as shown in Fig. 1. In a wide variety of operating environments, the water surface is a representative operational scenario [26, 27], so we chose assemble peg-in-hole on the water surface. In addition, we carried out the assembly within the constrained space of the petri dish and used an acoustic end effector for non-contact manipulation, which is contamination free and did not require the initial pose of the target to be preset. Thus, this assembly task is also suitable for *in situ* assembly scenarios. The assembly task includes alignment stage and insertion stage. And the system guides peg-in-hole alignment and manipulates peg insertion with vision-based closed-loop control, achieving peg-in-hole assembly finally.

### B. System Configuration and Calibration

The ultrasonic-driven non-contact peg-in-hole microassembly system is composed of a phased transducer array (PTA), a water tank, a high resolution vertical microscope, a customized field programmable gate array (FPGA)-based driver board and a host computer, as shown in Fig. 1. During the assembly process, the hole is fixed and the peg is placed in a petri dish filled with water. The microscopic camera captures the image information in real time and feedback to the host computer, which is responsible for processing the image to get the peg position and calculates phase information, then sends instructions to the FPGA driver board. The driver board generates square actuation signals to control the PTA system to generate dynamic acoustic trap. The trap exerts controllable acoustic radiation forces (ARF) on the peg in both force orientation and magnitude, enabling non-contact manipulation.

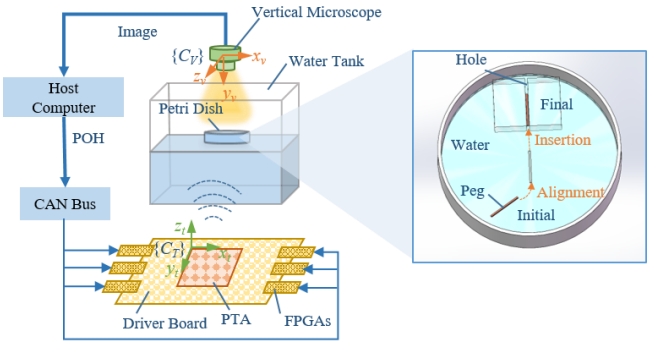


**Fig. 1.** Schematic diagram of the ultrasonic-driven system for in-plane assembly of peg-in-hole.

To obtain the transformation relationship between the image space coordinates and the operation space coordinates, the system needs to calibrate the image Jacobian matrix. The Cartesian coordinate system $\{\boldsymbol{C_T}\}$ is defined on the PTA plane, with its origin at the upper-left corner, while the image coordinate system $\{\boldsymbol{C_V}\}$ is aligned with the microscope plane. In fact, as the end-effector of the micro assembly system, the acoustic tweezers are invisible. Therefore, getting the position of the acoustic trap focal points needs a needle hydrophone during the calibration process. To calibrate the image Jacobian matrix, multiple calibration points are chosen randomly with their coordinates in $\{\boldsymbol{C_T}\}$ and $\{\boldsymbol{C_V}\}$ recorded. Due to the planar motion of peg during manipulation in Petri dish, the motion increment along the $z$-axis can be neglected. Therefore, the Jacobian matrix $\boldsymbol{J_V}$ can be simplified as a two-dimensional matrix following:

$$\begin{bmatrix} x_{ptc} \\ y_{ptc} \\ z_{ptc} \end{bmatrix} = \begin{bmatrix} x_{ref} \\ y_{ref} \\ z_{ref} \end{bmatrix} + \begin{bmatrix} J_V^* \\ 0 \quad 0 \end{bmatrix} \begin{bmatrix} u_{ptc} - u_{ref} \\ v_{ptc} - v_{ref} \end{bmatrix} \tag{1}$$

where $[x\_, y\_, z\_]^{\mathrm{T}}$ refer to the coordinates in $\{C_T\}$ and $[u\_, v\_]^{\mathrm{T}}$ refer to $x$-/$y$-axis in $\{C_V\}$. The subscript *ptc* and *ref* denote acoustic focal points and the selected reference points.

## III. Acoustic End-effector Design

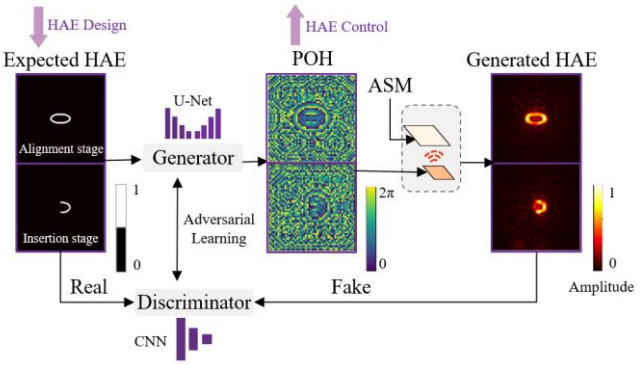


**Fig. 2.** Pipeline of the physics-based generative learning method for POH calculation and HAE generation.

### *A. Acoustic End-Effector Design and Generation*

To generate target acoustic end-effector, we use acoustic holography [28], namely holographic acoustic end-effector (HAE). The POH of the target HAE is generated by deep learning, sending an independent phase to each transducer in the ultrasonic PTA to generate the target acoustic trap. Traditional algorithms for POH calculation have limited resolution as constrained by the space-bandwidth product. Therefore, a generative physics-based deep learning-based algorithm is applied to combine the super-resolution concept with HAE generation. This method uses generative adversarial networks [29], which consists of a generator and a discriminator, as shown in Fig. 2. Specifically, the generator generates POH with reference to U-Net and adjusts the architecture to adapt to the super-resolution case (the input HAE has a higher resolution than the output POH). POH is applied to reconstruct the HAE through Angular Spectrum Method (ASM), which is a commonly used acoustic waves propagation simulation model). The discriminator is used to distinguish between the reconstructed HAE and the expected HAE. Through the adversary between the discriminator and the generator, mutual enhancement is achieved. Traditional error metrics (e.g., mean squared error) often produce overly smooth holograms with reduced trapping efficiency. The GAN-based approach learns to preserve sharp features and generates phase patterns that better match the desired acoustic trap shape, especially for complex geometries like elliptical acoustic trap.

In this paper, an elliptical acoustic trap as HAE was used to manipulate the slender peg. To verify the effectiveness of the proposed acoustic holographic acoustic trap construction method, we performed acoustic trap simulations using Python 3.10 and the ASM. The simulation result is shown in Fig. 2. The HAE is centered at [25, 25] mm on $\{\boldsymbol{Cr}\}$, with 5 mm semi-major axis, 1.5 mm semi-minor axis, and 1 mm linewidth. The PTA in simulation is in size of 50×50 with pitch of 1 mm, working at 2.32 MHz. And sound velocity is 1480 m/s. At this frequency, the wavelength in water is approximately 0.638 mm, and the pitch is slightly larger than one wavelength. While this may introduce some spatial aliasing effects, we selected these parameters to balance between spatial resolution and hardware constraints. The resulting holographic acoustic trap remains effective for our manipulation task, as demonstrated experimentally.

As can be seen, the sound pressure exhibits an elliptical distribution on the target plane, which conforms to the target acoustic trap shape. In fact, the shape of our end-effector is flexible. During alignment and the initial insertion phase, we use an elliptical acoustic trap to better control position and angle. After the peg enters the hole, we use a semi-elliptical acoustic end-effector to propel the peg further into the hole. The POH results in some pressure non-uniformity within the elliptical trap region. While iterative algorithms like MRAF could improve uniformity, they require significantly more computation time. For our peg manipulation task, the achieved uniformity was sufficient, as the peg's motion is governed by the gradient of the acoustic field rather than absolute pressure values.

### *B. Homogeneous Transformation of Acoustic End-Effector*

To achieve efficient dynamic control of the acoustic end-effector, this paper employs a homogeneous transformation method for manipulating the POH. The core physical principle is that applying a linear phase tilt to the wavefront equivalently modulates the acoustic propagation direction, enabling precise displacement of the HAE. This approach avoids the computational burden of re-running the deep learning model for each new pose, making it suitable for real-time control. The steady-state behavior of sound waves propagating in a homogeneous, lossless medium can be described by the Helmholtz equation:

$$(\nabla^2 + k^2)p(r) = 0 \tag{2}$$

where $p(r)$ is the complex sound pressure. The ASM is an efficient numerical solution to this equation that relates the acoustic end-effector in the plane of the transducer array $p(x, y, 0)$ to its angular spectrum $P(k_x, k_y; 0)$ via a Fourier transform:

$$P(k_x, k_y; 0) = \iint p(x, y, 0) e^{-j(k_x x + k_y y)} dxdy \tag{3}$$

The process of the acoustic end-effector propagating to the target plane is reflected as a phase delay in the angular spectrum domain:

$$P(k_x, k_y; z) = P(k_x, k_y; 0) e^{jk_z z} \tag{4}$$

According to the Fourier shift theorem in wave diffraction theory, the translation of a function in the spatial domain is equivalent to multiplying its spectrum in the frequency domain by a linear phase factor.

$$\mathcal{F}\{p(x - \Delta x, y - \Delta y)\} = P(k_x, k_y) e^{-j(k_x \Delta x + k_y \Delta y)} \tag{5}$$

Therefore, the rigid body motion of the acoustic

end-effector in a two-dimensional plane can be achieved through the homogeneous transformation of the POH. Note that Eqs. 2-5 are formulated in the continuous domain, whereas the actual phased array is discrete and spatially undersampled (pitch > wavelength). Theoretical analysis suggests that this discretization introduces spatial harmonics, and the undersampling could lead to spatial aliasing, which potentially limits the steering range of the HAE. However, in our system, this effect is mitigated by two factors. First, the finite size of the transducer elements provides strong intrinsic spatial low-pass filtering. This filtering naturally suppresses energy leakage into large-angle grating lobes. Second, the microassembly task involves small-range manipulation near the center of the array, where the steering angles are small. Consequently, the generated HAE and steering performance remain sufficiently robust for stable peg manipulation, as validated by the experiments.

## IV. Visual Alignment and Insertion Control

### A. Feature Extraction and Visual Alignment

We process images captured by the microscope camera to extract features of the hole and peg. Before placing the peg, we capture a background image and perform feature extraction to determine the hole edge location. During the peg-in-hole assembly process, feature extraction is performed on the real-time camera image. We first implement background subtraction and grayscale transformation, then extract black points from the image. Linear fitting is analyzed on these points to determine the orientation of the part's principal axis and center coordinates.

We set the starting insertion point as A, 7 mm from the hole along the hole. This means the peg must be aligned before reaching this point and then inserted. Peg-in-hole alignment involves both angular and positional alignment. Image processing allows us to determine the peg's current pose, then calculate its relative position delta($u$, $v$) and angle delta($\theta$) relative to point A and the hole direction. After obtaining the image pose, the pseudo image Jacobian matrix $\boldsymbol{J}_V^*$ is used to convert the image coordinates into Cartesian coordinates to determine the pose of the peg in the actual space and how it needs to move. The peg is captured and manipulated using the elliptical acoustic end-effector proposed in Section III. However, due to factors such as environmental interference (water flow or acoustic flow and so on), system calibration errors, etc., it is not enough to simply move the acoustic end-effector according to the preset path. The actual pose of the peg may deviate from the acoustic end-effector or even completely detach from the acoustic end-effector, resulting in manipulation failure. Therefore, this paper introduces closed-loop control under the condition of planning the motion trajectory in advance. The deviation between the peg's pose $p_n$ of real-time visual feedback and acoustic end-effector pose $x_n$ is used to dynamically adjust the pose $x_{n+1}$ of the acoustic end-effector to ensure that the peg is always captured and guided to move along the target path. The specific control strategy is as follows:

$$x_{\mathrm{n+1}} = T_{n+1} + K_i \sum_{i+1}^{n} (T_i - p_i), \quad |p_n - x_n| \le L \tag{6}$$

$$x_{n+1} = p_n + K_i (T_n - p_n), \quad |p_n - x_n| > L \tag{7}$$

where $T_i$ and $T_n$ represent the pose along the planning trajectory, $p_i$ and $p_n$ represent the pose of the peg, which is the pose transferred from the microscope using pseudo image Jacobian matrix $\boldsymbol{J}_V^*$, $K_i$ is the integral coefficient. Before the experiment, the threshold $L$ is determined first, and the $L$ is determined by the prior knowledge in the experiment as the scope of the acoustic end-effector. When the deviation is less than the threshold $L$, the peg is considered to be within the effective range of the acoustic end-effector, and integral closed-loop control is employed. When the deviation exceeds the threshold $L$, the peg is considered to have escaped or is about to escape the acoustic end-effector, and a compensation point is introduced. Sudden strong interference (such as a violent fluid impact) may cause the peg to momentarily move too far from the acoustic end-effector, making it impossible to directly re-enter the trajectory using integral control.

After executing the planned motion path, the deviation between the current and target poses of the peg is calculated based on image features. The error is then determined to be less than a threshold. If not, the peg pose is further adjusted using Eq. 9 until alignment is achieved.

$$\vec{e}_n = T_n - p_n \tag{8}$$

$$x_{n+1} = x_n + k_i \vec{e}_n \tag{9}$$

where $x_n$ represents the trap position, the vector $e_n$ is the pose misalignment error of the target feature obtained in the image with respect to the desired target, $n$ represents the control loop index, and $k_i$ represents the integral coefficient.

### B. Insertion Control Strategy

In traditional microassembly, a mechanical gripper physically contacts and rigidly holds the target part. This ensures that the target's pose remains unchanged after the vision system completes precise alignment, making the visual alignment and insertion processes independent. Ultrasonic microassembly, however, manipulates the target non-contact, lacking rigid pose constraints. Furthermore, the acoustic end-effector exhibits nonlinear effects. After precise alignment, the peg's pose may shift during insertion, leading to assembly failure. Therefore, the proposed vision system continuously and precisely monitors the peg's pose throughout the insertion process. This system provides real-time feedback to the motion control system, actively compensating for angle deviations through Eq. 9, forming a closed-loop control system that ensures the motion trajectory strictly conforms to the expected trajectory. The specific

process is shown in Fig. 3: Each time the peg moves, the relative angle error is calculated and is adjusted to ensure successful insertion.

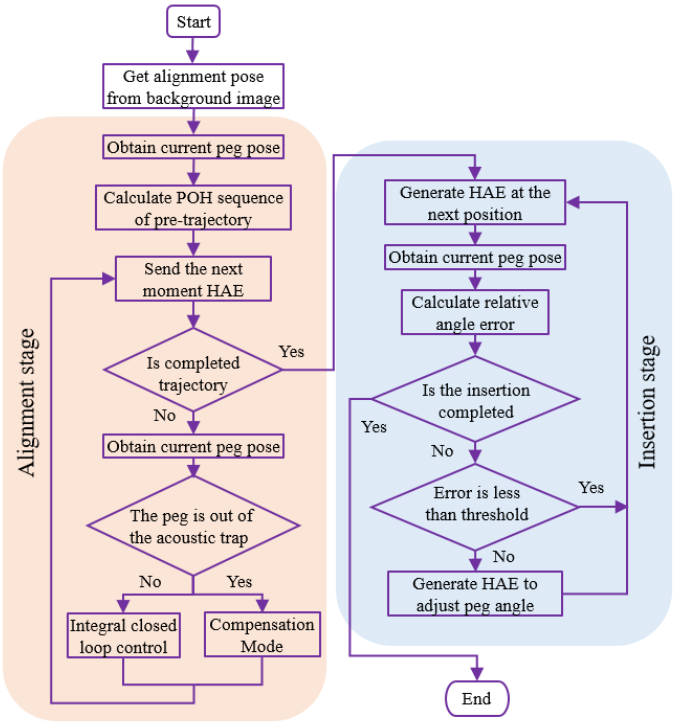


**Fig. 3.** Peg-in-hole trajectory control strategy in alignment process and insertion control strategy for real-time posture adjustment.

## V. Experiments and Results

### A. Experiment Details

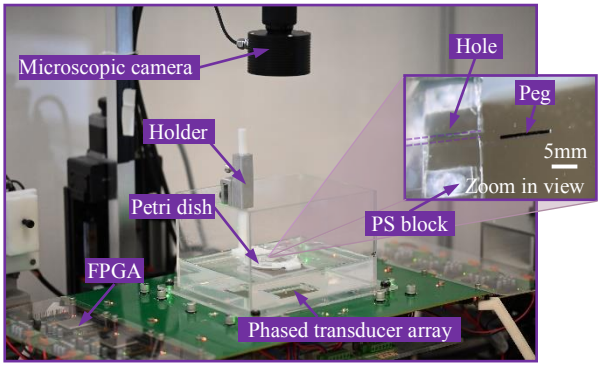


**Fig. 4.** Acoustic microassembly platform, which primarily consists of PTA, a microscopic camera.

**System Setup.** The physical prototype of the ultrasonic manipulation experimental platform for peg-in-hole microassembly described in Section II-A is shown in Fig. 4. The core of the platform is a phased array system with 50×50 channels. The PTA is fabricated on 1 mm thick lead piezoelectric ceramics (PZT) by lithographic printing and wet etching of nickel electrodes on both sides. The transducer dimensions for each channel are 800×800 μm, with a 1 mm pitch. The driver board consists of six FPGAs operating at 200 MHz and a 10 V drive voltage. The system communicates via a CAN-Bus controller with a baud rate of 500 kbps, capable of updating the POH at 11 frames per second (FPS). The imaging system uses a Prosilica GC2450 microscope camera equipped with a Navitar zoom lens, capable of capturing images at 2448×2050 resolution and 15 FPS. The microscope has a field of view (FOV) of approximately 30×35 mm. A square hole (800 μm wide, 15 mm long) made of PDMS is fixedly mounted in a commercial petri dish (60 mm diameter). A plastic peg (600 μm diameter, 10 mm long, black) is placed on the water surface in a petri dish containing water. The size of the square hole is 200 microns larger than the diameter of the peg, so there is a clearance fit between them. Although the acoustic impedance mismatch at the water-Petri dish interface causes some reflection, the energy loss is limited. Given the material properties and thickness of the Petri dish, experimental results confirm that sufficient acoustic energy is transmitted through the dish bottom to generate the required trapping forces on the peg.

**Calibration.** The calibration hand-eye relationship of system was calculated by feature points in $\{\boldsymbol{C_V}\}$ and $\{\boldsymbol{C_T}\}$. The centroid of the calibration data pairs were $[u_{avg}, v_{avg}]^T$=[1329, 718]$^T$ pixel and $[x_{avg}, y_{avg}, z_{avg}]^T$=[25.00, 25.00, 50.00]$^T$ mm. The image Jacobian matrix was calibrated as

$$J_V^* = \begin{bmatrix} -15.8062 & 1.1411 \\ -0.9098 & -15.5441 \end{bmatrix} \mu m / pixel \qquad (10)$$

The calibrated image Jacobian matrix error was assessed by comparing the calculated position with actual position. The result showed the calibration error to be 54.41 μm, which was within acceptable accuracy range for the peg-in-hole microassembly experiment.

### B. Acoustic End-Effector Optimization

To identify the optimal acoustic end-effector for manipulating a 10mm long peg, we used elliptical acoustic end-effector of different sizes (with the major axis decreasing from 12mm to 10mm and the semi-minor axis decreasing from 4mm to 2mm, all with a line width of 1mm) to do the peg trap and positioning experiments. The results are shown in Fig. 5. The values in parentheses in the figure, like (2,10), represent the size of the elliptical acoustic end-effector. The first item refers to the size of the minor axis of the ellipse, and the second item refers to the size of the major axis of the ellipse. The two curves in the figure represent the trap success rate, and positioning error for different acoustic end-effectors. The positioning error includes positional error and angular error, and all error values are normalized. The results confirm that an elliptical acoustic end-effector with a major axis of 10mm and a minor axis of 3mm achieves a high trap success rate and better positioning accuracy. Therefore, an HAE of this size was used to proceed other peg manipulation and peg-in-hole assembly experiments.

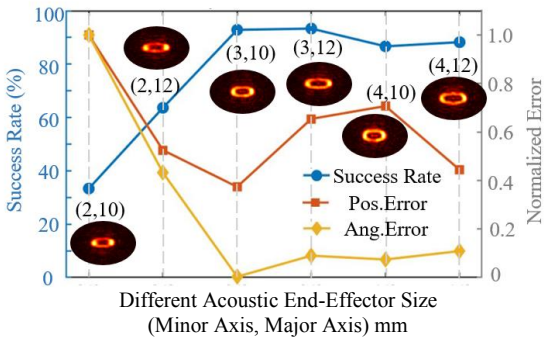


**Fig. 5.** Optimization experiment, which shows the trap and positioning capabilities of elliptical acoustic traps of different sizes for

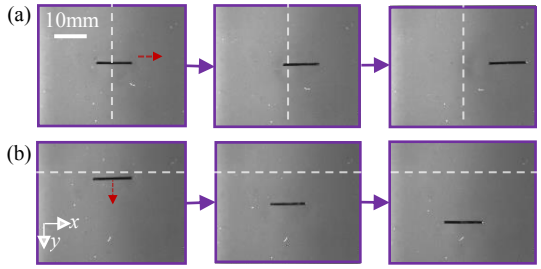


**Fig. 6** The translation experiment result. (a) shows the movement of the peg in the horizontal direction, while (b) shows the movement of the peg in the vertical direction.

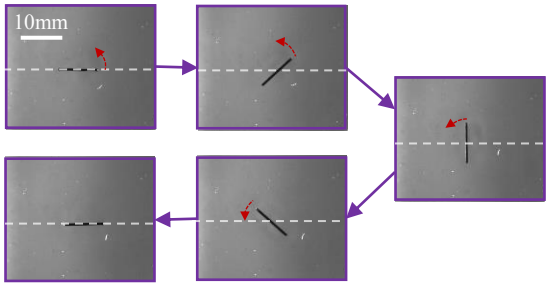


**Fig. 7** The rotation experiment result, which shows the peg rotating 180 degrees around its center

### *C. Peg Manipulation Experiments*

The peg manipulation experiment can be divided into two parts: translation and rotation. In the translation experiment, we let the peg translate 5mm along the $x$-axis and $y$-axis respectively, as shown in Fig. 6. In the rotation experiment, we rotated the peg 180° around its center point, as shown in Fig. 7. The feasibility of peg manipulation was verified by this experiment.

### *D. Peg-in-Hole Assembly Experiments*

**Alignment.** Since microassembly can be divided into two steps, alignment and insertion, we first conducted an alignment experiment on the peg. Before placing the peg, the camera captured a background image to determine the center point of the hole edge. The insertion starting point was set 7 mm from this point along the hole, with the actual coordinates [$x_a$, $y_a$]. (The hole was oriented along the x-axis, meaning its angle was 0°.) Therefore, during the alignment experiment, the peg need to be moved to the [$x_a$, $y_a$, 0°] pose and accurately positioned. Initially, the peg can be in any pose. The camera captures the image and identifies it, generating an elliptical acoustic end-effector at that position to capture the peg. After obtaining the initial and target poses, a series of acoustic end-effector trajectories are obtained through linear interpolation. The corresponding POH is calculated by rigid transformation of the POH with the known pose [25mm, 25mm, 0°].

Alignment experiments were conducted using both open-loop and closed-loop control. Open-loop control directly executes the pre-generated trajectory without visual feedback; closed-loop control uses the control law described in Section IV-A. The integral coefficient ki is experimentally determined to be 0.3, and the termination criteria for closed-loop control are a position error of less than 100μm and an angular error of less than 2°. The results of repeated experiments are shown in Table 1. The average position alignment error for open-loop control was 1626 μm, and the average angular alignment error was 2.32°. Closed-loop control achieved alignment errors as small as 60 μm, with an average angular alignment error less than 0.8°, which validated the effectiveness of closed-loop control algorithm.

**Insertion.** After alignment is completed, an insertion experiment is conducted. We specify that the insertion is successful when the end of the peg is recognized to enter the hole. During this process, the camera provides real-time feedback and compensation for the peg angle, achieving precise control. Therefore, when the alignment is completed, we record the distance from the end of the peg to the hole in the experiment and draw a curve to characterize the insertion process, as shown in Fig. 8.

**Assembly.** The entire assembly process is shown in Fig. 9. The two curves in the figure represent the position error and angular error changes during the assembly process. The period before 25 seconds represents the alignment process, and the period after 25 seconds represents the insertion process. It can be seen that during the alignment process, the position error and angular error gradually decrease. During the insertion phase, the error curves fluctuate within a certain range. Even with some disturbance, the acoustic end-effector can adjust the peg's position and pose in a timely manner to ensure successful insertion.

TABLE I. ALIGNMENT ERROR OF OPEN-LOOP AND CLOSED-LOOP CONTROL

| Exp. No | *Open-loop error* (*μm*, °) | *Closed-loop error* (*μm*, °) |
|---|---|---|
| 1 | (861.9, 2.053) | (53.8, 0.364) |
| 2 | (1868.8, 2.620) | (33.8, 1.030) |
| 3 | (1786.8, 1.221) | (42.9, 0.371) |
| 4 | (1143.2, 4.243) | (80.0, 0.611) |
| 5 | (284.6, 1.455) | (63.4, 1.188) |
| Average Value | (1189.06, 2.3184) | (54.78, 0.7128) |

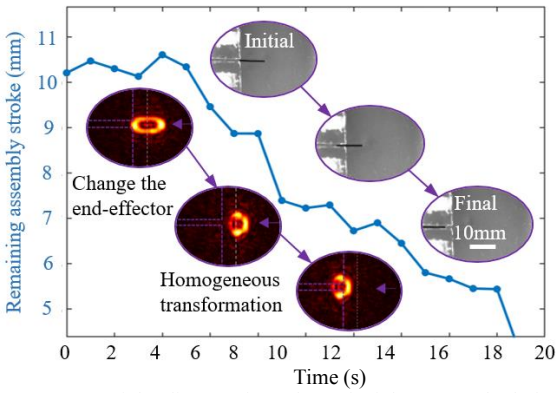


**Fig. 8.** Curve of the distance from the end of the peg to the hole during insertion process.

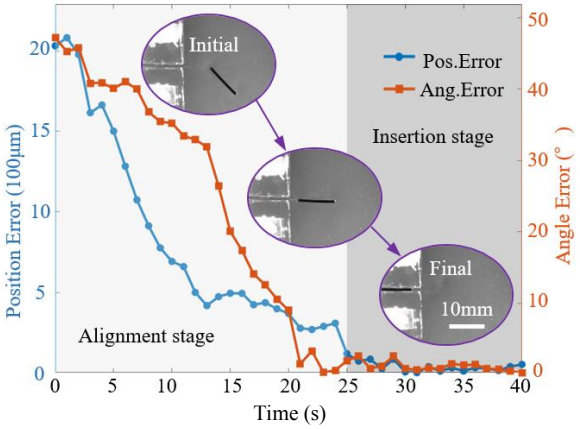


**Fig. 9.** Error convergence curve during peg-in-hole assembly, including position error and angle error.

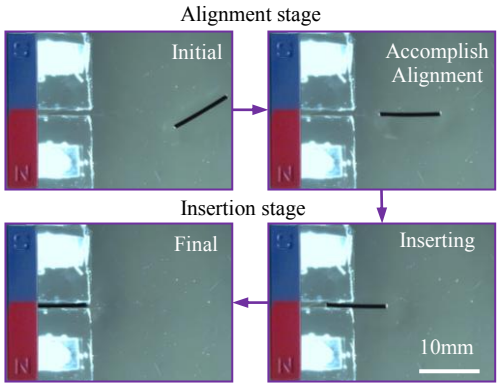


**Fig. 10.** Assembly process in strong magnetic environment.

### *E. Peg-in-Hole Assembly Experiments in Strong Magnetic Environment*

To demonstrate the environmental compatibility of this method, we conducted assembly experiments under strong magnetic field conditions, with a magnet suspended above the petri dish, as shown in Fig. 10. The plastic peg is non-magnetic and thus unaffected by the field directly; the purpose is to show that our acoustic manipulation system operates reliably in environments where magnetic field is present. It can be observed that the acoustic manipulation of the peg was not influenced by the magnetic field environment. It still exhibited excellent alignment accuracy and insertion control, and finally accomplished the assembly task successfully.

### *F. Discussion and Future Works*

Experimental results show that the closed-loop average alignment error for peg-in-hole assembly is 54.78 μm and 0.71°. This is because we only considered the acoustic properties related to the sound pressure distribution in the end-effector design, without considering the nonlinear effects caused by acoustic-fluid-structure coupling. As a result, the alignment accuracy during the assembly process is less than ideal, decreasing compared to traditional methods. To further improve the assembly precision, more research efforts should be dedicated towards optimized end-effector design, better understanding of the acoustic-fluid-structure coupling, by developing high-fidelity multiphysics simulations to accurately predict these interactions in the future. Furthermore, we will explore advanced control strategies to actively compensate for these disturbances in real-time. We anticipate that these improvements will significantly enhance the alignment accuracy, potentially surpassing the performance of traditional assembly methods, and enable the application of our ultrasonic-driven non-contact microassembly system in more demanding high-precision manufacturing scenarios.

## VI. Conclusion

In this work, we testified the feasibility precision acoustic assembly over peg-in-hole in micron-scale by applying robotic approach, in terms of auto trapping, auto positioning and motion planning. We propose a non-contact microassembly system for peg-in-hole tasks using an ultrasonic phased array and visual feedback. Experimental results demonstrate high stability in peg positioning, alignment, and insertion, and show robustness against disturbances during assembly under closed-loop control. Due to its ease of deployment and non-contact nature, ultrasonic assembly is suitable for specialized applications, such as *in situ* and *in vivo* assembly, and holds great potential for microassembly applications such as biomedical devices and micro-robotics.

## References

[1] J. Agnus, N. Chaillet, C. Clévy, *et al.*, "Robotic microassembly and micromanipulation at FEMTO-ST," *Journal of Micro-Bio Robotics*, vol. 8, pp. 91–106, 2013.

[2] P. Li, *et al.*, "3D magnetic assembly of cellular structures with 'printing' manipulation by microrobot-controlled microfluidic system," *IEEE International Conference on Robotics and Biomimetics*, pp. 1967–1972, 2015.

[3] A. N. André, *et al.*, "Automating robotic micro-assembly of fluidic chips and single fiber compression tests based on Θ visual measurement with high-precision fiducial markers," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 1, pp. 353–366, 2024.

[4] Y. Huang, X. Liu, J. Ke, X. Zhang, L. Yang, and J. Ota, “Variable parameter admittance control of robot based on a virtual force/torque sensor,” *IEEE/ASME Transactions on Mechatronics*, in press, 2025.

[5] Y. Zhao, et al., “Automated assembly of magnetic soft microrobots with chopstick-like two-fingered microhand,” *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 19824–19835, 2025.

[6] Y. Zang, P. Wang, F. Zha, W. Guo, S. Ruan, and L. Sun, “Geometric-feature representation based pre-training method for reinforcement learning of peg-in-hole tasks,” *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3478–3485, 2023.

[7] W. Rong, L. Sun, and L. Chen, "Research on integrated micromanipulator and its application in microassembly," *Proceedings of the IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*, Changsha, China, pp. 831-835, vol. 2, 2003.

[8] Z. Wang, B. Zhang, J. Li, Z. Luo, and C. Zheng, "Data-driven robotic shaft-hole assembly: A review," *Journal of Nanjing University of Aeronautics & Astronautics (Natural Science Edition)*, vol. 57, no. 3, Jun. 2025.

[9] X. Huang, J. Cai, M. Wang, and X. Lv, “A piezoelectric bimorph micro-gripper with micro-force sensing,” *IEEE International Conference on Information Acquisition*, Hong Kong, China, 2005.

[10] H. Wang, *et al.*, “Reconfigurable formation control for coordinated micromanipulation of cross-scale targets in railed multiprobe robot,” *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 2, pp. 1319–1330, 2024.

[11] Q. Fan, Y. Wu, K. Bi, and Y. Liu, “Autonomous vision-guided two-arm collaborative microassembly using learned manipulation model,” *IEEE Robotics and Automation Letters*, vol. 9, no. 3, pp. 2375–2382, 2024.

[12] F. Mo and W. Cai, “Discussion of failure cases for contamination in electronics assembly process,” *International Conference on Reliability, Maintainability and Safety*, Hangzhou, China, pp. 1–4, 2016.

[13] X. Hui, J. Luo, H. You, and H. Sun, “Noncontact manipulator for sedimented/floating objects via laser-induced thermocapillary convection,” *IEEE Transactions on Robotics*, vol. 41, pp. 1476–1490, 2025.

[14] F. W. Heger, “Generating robust assembly plans in constrained environments,” *IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA, pp. 4068–4073, 2008.

[15] B. Chang, B. K. Wang, J. L. Jin, and Q. Zhou, “Capillary pick-and-place of glass microfibers,” *IEEE Access*, vol. 9, pp. 15074–15083, 2021.

[16] D. Y. Hou, D. B. Cheng, N. Y. Zhang, *et al.*, “In vivo assembly enhanced binding effect augments tumor specific ferroptosis therapy,” *Nature Communications*, vol. 15, p. 454, 2024.

[17] Z. Li, P. Z. Liang, L. Xu, *et al.*, “In situ orderly self-assembly strategy affording NIR-II-J-aggregates for in vivo imaging and surgical navigation,” Nature Communications, vol. 14, p. 1843, 2023.

[18] N. Carlisle, M. A. K. Williams, C. P. Whitby, V. Nock, J. L. Y. Chen, and E. Avci, “Manipulation of optical force-induced micro-assemblies at the air-liquid interface,” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Detroit, MI, USA, pp. 2709–2714, 2023.

[19] Y. Zhang, *et al.*, “Enhanced automated cell micromanipulation via programmable magnetic microgripper design,” IEEE Transactions on Automation Science and Engineering, vol. 22, pp. 10107–10117, 2025.

[20] M. Wang, S. An, Z. Sun, J. Li, Y. Wang, and S. Liu, “Selective, robust, and precision manipulation of particles in complex environments with ultrasonic phased transducer array and microscope,” *IEEE Transactions on Robotics*, vol. 41, pp. 887–904, 2025.

[21] I. Matei, M. Zhenirovskyy, J. de Kleer, and A. Plochowietz, “Microscale 2D particle position control: The individual and group cases,” *IEEE Control Systems Magazine*, vol. 43, no. 4, pp. 42–62, 2023.

[22] S. H. Cho, A. K. M. Ariful Haque Siddique, C. G. Kim, and B. Ahn, “Translocation of microparticles in a fluid-flow by adjusting the operating frequency of ultrasonic standing wave (USW),” *IEEE International Ultrasonics Symposium*, Orlando, FL, USA, pp. 1510–1513, 2011.

[23] Z. Zhang, M. Koch, and D. Ahmed, “Soft acoustic end-effector,” *IEEE International Conference on Robotics and Automation*, Yokohama, Japan, pp. 12772–12778, 2024.

[24] I. I. I. Al-Nuaimi, M. N. Mahyuddin, and N. K. Bachache, “A non-contact manipulation for robotic applications: A review on acoustic levitation,” *IEEE Access*, vol. 10, pp. 120823–120837, 2022.

[25] M. Wang, *et al.*, “Vision-based closed-loop control with spatiotemporal multiplexing strategy for noncontact trapping of multiple micro-particles,” *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 4063–4075, 2025.

[26] Y. S. Guan, *et al.*, “Air/water interfacial assembled rubbery semiconducting nanofilm for fully rubbery integrated electronics,” *Science Advances*, vol. 6, no. 26, p. eabb3656, 2020.

[27] B. Wang, Z. Che, C. Cheng, *et al.*, "Topological water-wave structures manipulating particles," *Nature*, vol. 638, pp. 394–400, 2025.

[28] C. Zhong, *et al.*, “Real-time acoustic holography with physics-based deep learning for robotic manipulation,” *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 4155–4164, 2024.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014.