



## 반도체 검사 장비를 위한 강화학습 기반 이중 서보 제어

서승우<sup>a</sup>, 전용호<sup>a</sup>, 이문구<sup>a,\*</sup>

### Design of Reinforcement Learning-Based Dual-Servo Controller for Semiconductor Inspection Equipment

Seung Woo Seo<sup>a</sup>, Yongho Jeon<sup>a</sup>, Moon Gu Lee<sup>a,\*</sup><sup>a</sup> Department of Mechanical Engineering, Ajou University

#### ARTICLE INFO

##### Article history:

Received	1	January	2026
Revised	20	January	2026
Accepted	21	January	2026

##### Keywords:

Reinforcement learning  
PID controller  
Dual-servo stage  
Semiconductor inspection equipment

#### ABSTRACT

Ultra-precision stages for semiconductor inspection must simultaneously achieve long-stroke motion, nanometer-level accuracy, and fast settling. Dual-servo stages comprising coarse and fine stages are widely used, but coordination performance degrades in practice due to dynamic mismatch, alignment errors, and nonlinear coupling. Conventional approaches such as feedforward compensation and model-based control rely on accurate models and show limited robustness to disturbances. This study proposes a reinforcement learning-based PID (RL-PID) controller for the fine stage of a one-degree-of-freedom dual-servo system, while a conventional PID is applied to the coarse stage. The RL-PID optimizes PID scaling factors using DDPG, with stability ensured through a Lyapunov-based reward and FRF-derived safe exploration bounds. Simulation and dSPACE-based experiments demonstrate significant improvements in settling time, overshoot, and tracking accuracy, achieving up to a 33% reduction in RMSE under high-speed scanning. The results validate the effectiveness of safe RL-PID in mitigating nonlinear coordination issues and enhancing industrial applicability.

## 1. 서론

반도체 산업은 인공지능, 고성능 컴퓨팅, 자율주행, 5G 통신 등 첨단 산업 발전을 뒷받침하는 핵심 기반이다. 반도체 칩의 집적도 증가와 회로 선폭의 미세화에 따라 제조 장비에는 나노미터 수준의 해상도와 높은 생산성이 동시에 요구된다. 이러한 요구를 충족하기 위해 반도체 검사 장비에는 웨이퍼 및 레티클의 위치를 정밀하게 제어하는 초정밀 스테이지가 필수적으로 사용된다. 해당 스테이지는 수백 밀리미터 이상의 장거리 이송 범위, 나노미터 수준의 위치 정밀도, 그리고 수십 밀리 초 이하의 빠른 정착

시간을 동시에 만족해야 한다<sup>[1]</sup>. 이러한 성능 요구를 충족하기 위한 구조로 조동 스테이지와 미동 스테이지로 구성된 이중 서보 스테이지가 널리 사용되고 있다<sup>[2, 3]</sup>. 조동 스테이지는 리니어 모터 기반으로 장거리 이송과 고속 응답을 담당하며, 미동 스테이지는 보이소 코일 모터와 플렉서 구조를 통해 정밀 위치 보정을 수행한다<sup>[4]</sup>.

그러나 두 스테이지 간의 동특성 불일치, 정렬 오차, 기구적 결합 효과로 인해 상호 간섭이 발생하고, 이로 인해 협조성이 저하되어 전체 시스템의 정밀도와 응답 성능이 제한된다. 따라서 이중 서보 스테이지의 성능을 극대화하기 위해서는 스테이지 간 상호작용을

\* Corresponding author. Tel.: +82-31-219-2338

E-mail address: moongulee@ajou.ac.kr (Moon Gu Lee).

고려한 협조 제어 전략이 필수적이다<sup>[5]</sup>.

기존에는 이러한 문제를 해결하기 위해 Feedforward 보상, Gain Scheduling, 모델 기반 보상 제어 등 다양한 기법이 제안되었으나, 이들 방법은 시스템의 선형 모델링 정확도에 크게 의존하며, 마찰, 온도 변화, 구조 공진과 같은 비선형 요인과 외란이 존재하는 실제 환경에서는 성능 유지에 한계가 있다<sup>[6, 7]</sup>. 이에 따라 최근에는 퍼지 제어, 신경망, 강화학습과 같은 지능형 제어 기법이 도입되어 비선형성과 불확실성이 큰 시스템에서도 강인한 제어 성능을 확보하려는 연구가 활발히 진행되고 있다<sup>[8, 9]</sup>.

특히 강화학습(reinforcement learning, RL)은 환경과의 상호 작용을 통해 제어 정책을 데이터 기반으로 학습할 수 있어, 이중 서보스테이지와 같이 복잡한 비선형 협조 특성을 갖는 시스템에 적합한 접근법으로 주목받고 있다. 연속 제어가 가능한 Deep Deterministic Policy Gradient (DDPG), Twin Delayed DDPG (TD3), Proximal Policy Optimization (PPO) 등의 알고리즘을 활용한 연구에서는 Proportional-Integral-Derivative (PID) 계인의 자동 조정 및 제어 성능 개선이 보고되었다<sup>[10-12]</sup>. 그러나 대부분의 선행 연구는 단일 서보 시스템이나 단순화된 시뮬레이션 환경을 대상으로 수행되었으며, 조동-미동 스테이지 간 상호작용을 실제로 반영한 실험적 연구는 제한적이다. 또한, 보상 함수 설계가 경험적 기준에 의존하는 경우가 많아, 학습 과정 중 불안정한 제어 입력이 발생할 가능성이 존재하며, 이는 고신뢰성이 요구되는 산업 장비 적용에 있어 중요한 한계로 작용한다.

최근에는 이러한 문제를 해결하기 위해 Lyapunov 안정성 이론을 강화학습에 결합한 Safe Reinforcement Learning (Safe RL) 기법이 제안되고 있다. Lyapunov 기반 보상 설계나 제약 조건을 통해 학습 과정 전반에서 안정성을 보장하려는 연구가 보고되었으나<sup>[13]</sup>, 반도체 검사 장비와 같은 고속·고정밀 이중 서보 스테이지를 대상으로 한 실증 연구는 여전히 부족한 실정이다.

이에 본 연구에서는 1자유도의 이중 서보 스테이지를 대상으로, 조동 스테이지에는 기존 PID 제어를 적용하고 미동 스테이지에는 강화학습 기반 PID 제어기(RL-PID)를 적용하는 제어 구조를 제안한다. RL-PID는 Ziegler-Nichols 방법으로 초기 PID 계인을 설정한 후, 강화학습을 통해 스케일링 인자( $\alpha$ ,  $\beta$ ,  $\gamma$ )를 학습함으로써 제어기의 적응성과 최적화 성능을 동시에 확보한다. 또한, 주파수 응답 함수(FRF)를 활용하여 Gain Margin (GM)과 Phase Margin (PM) 기준을 만족하는 안정 영역 내에서만 학습을 수행하고, Lyapunov 안정성 조건을 보상 함수에 통합함으로써 학습 과정에서의 안정성과 수렴성을 동시에 보장한다. 제안된 방법은 MATLAB/Simulink 시뮬레이션과 dSPACE 기반 실험을 통해 기존 PID 제어기와 비교되며, 이를 통해 강화학습 기반 제어의

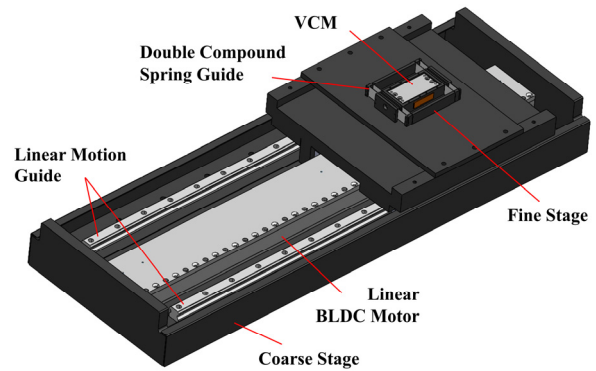


Fig. 1 Schematic of dual-servo stage

산업적 적용 가능성을 실증적으로 평가한다.

## 2. 시스템 및 제어기 설계

### 2.1 이중 서보 스테이지 시스템

본 연구에서 대상으로 하는 이중 서보 스테이지는 장거리 이송과 고정밀 위치 제어를 동시에 만족하기 위해 Fig. 1과 같이 조동 스테이지(coarse stage)와 미동 스테이지(fine stage)로 구성된다. 조동 스테이지는 리니어 BLDC 모터와 리니어 모션 가이드를 기반으로 수백 밀리미터 이상의 이송 범위와 고속 응답을 담당하며, 미동 스테이지는 보이시 코일 모터(VCM)와 플래셔 구조를 이용하여 나노미터 수준의 정밀 위치 보정을 수행한다. 이러한 역할 분담을 통해 전체 스테이지는 넓은 작업 영역과 높은 위치 정밀도를 동시에 확보할 수 있다.

조동 스테이지는 비교적 큰 관성 및 마찰 특성을 가지며, 주로 저주파 영역에서의 위치 추종 성능이 지배적이다. 반면, 미동 스테이지는 작은 질량과 높은 기계적 강성을 바탕으로 고주파 영역에서의 빠른 응답과 미세 보정에 유리하다. 두 스테이지는 직렬 구조로 결합되어 동일한 위치 출력에 기여하므로 조동-미동 스테이지 간의 동특성 불일치와 기구적 결합 효과로 인해 상호 간섭이 발생할 수 있다. 이러한 상호작용은 협조 제어 성능을 저하시켜 오버슈트 증가, 정착 시간 지연, 미세 추종 오차를 유발하는 주요 원인이 된다.

본 연구에서는 미동 스테이지의 동특성을 정밀하게 파악하기 위해 주파수 응답 함수(frequency response function, FRF)를 측정하였다. Fig. 2는 HP DSA 35670A FRF 측정 결과를 나타내며, 제어기 설계 및 안정성 확보를 위한 기준으로 활용된다. 특히, FRF 기반 분석은 강화학습 과정에서 불안정한 제어 입력이 발생하지 않도록 학습 가능 영역을 제한하는 데 중요한 근거를 제공한다. 또한 측정된 FRF 결과는 시스템 식별을 통해 시뮬레이션용 대리 모델을 구성하는 데 활용되었으며, 강화학습이 수행되는 시뮬레이션 환경에 실제 미동 스테이지의 동특성을 반영하기 위한

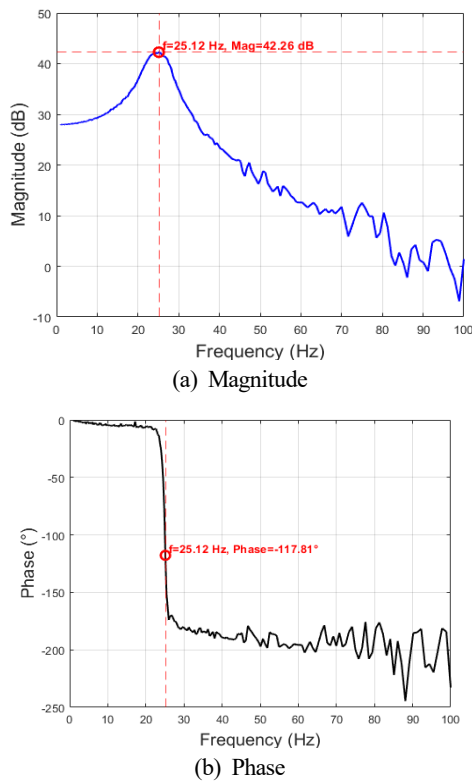


Fig. 2 Frequency response function

기준으로 적용되었다. 나아가 FRF 분석을 통해 확인된 공진 주파수 및 안정 여유 정보를 바탕으로 제어기 설계 시 허용 가능한 제어 입력 범위를 정의하였으며, 해당 범위는 강화학습 과정에서 PID 스케일링 인자의 탐색 범위를 제한하는 안정 영역으로 적용되어 학습 중 불안정한 제어 입력이 발생하지 않도록 하였다.

이중 서보 스테이지의 제어 구조는 조동 스테이지와 미동 스테이지가 각각 독립적인 제어 루프를 가지는 형태로 구성된다. 조동 스테이지는 기존 PID 제어기를 통해 기준 위치를 추종하며, 미동 스테이지는 조동 스테이지의 잔여 오차를 보상하는 역할을 수행한다. 본 연구에서는 이러한 구조적 특성을 고려하여 조동 스테이지에는 기존 PID 제어기를 적용하고, 미동 스테이지에는 강화학습 기반 PID 제어기를 적용함으로써 이중 서보 스테이지의 협조 성능을 향상시키고자 한다.

## 2.2 강화학습 기반 PID 제어기

이중 서보 스테이지의 협조 제어에서 미동 스테이지는 조동 스테이지의 잔여 오차를 보상하며 전체 위치 정밀도를 결정하는 핵심 요소이다. 일반적인 PID 제어기는 구조가 단순하고 안정성이 우수하지만, 마찰, 공진, 정렬 오차 등 비선형성과 불확실성이 존재하는 실제 이중 서보 시스템 환경에서는 성능이 제한적이다. 특히 미동 스테이지는 고주파 영역에서의 빠른 응답과 정밀 제어가 요

구되므로, 고정된 PID 게인만으로는 다양한 동작 조건에 대해 일관된 성능을 유지하기 어렵다.

본 연구에서는 이러한 한계를 극복하기 위해 미동 스테이지에 강화학습 기반 PID제어기(RL-PID)를 적용한다. 제안된 RL-PID는 기존 PID 구조를 유지하되, 비례(proportional), 적분(integral), 미분(derivative) 항에 대한 스케일링 인자( $\alpha, \beta, \gamma$ )를 강화학습을 통해 조정하는 방식으로 설계된다. 이를 통해 제어기의 기본적인 안정성과 해석 가능성은 유지하면서도, 환경 변화와 비선형 특성에 대한 적응 능력을 확보할 수 있다. 초기 PID 게인은 Ziegler-Nichols 방법을 통해 설정하여 학습 초기 단계에서의 불안정한 응답을 완화하였다.

강화학습 에이전트의 상태(state)는 미동 스테이지의 위치, 속도, 위치 오차, 적분 오차로 구성되며, 이는 시스템의 현재 동작 상태와 누적 제어 성능을 동시에 반영한다. 행동(action)은 PID 제어기의 세 스케일링 인자( $\alpha, \beta, \gamma$ )로 정의되며, 연속적인 값으로 출력된다. 이러한 연속 제어 문제를 효과적으로 처리하기 위해 본 연구에서는 Deep Deterministic Policy Gradient (DDPG) 알고리즘을 사용하였다. DDPG는 Actor-Critic 구조를 기반으로 연속 행동 공간에서의 정책 학습이 가능하며, PID 스케일링 인자의 실시간 조정에 적합하다. 또한 비교적 안정적인 학습 특성을 가지므로, 산업 제어 환경에 적용 가능한 강화학습 알고리즘으로 판단하였다.

강화학습 기반 제어기의 안정성 확보를 위해, 본 연구에서는 성능 지표와 안정성 조건을 동시에 고려한 통합형 보상 함수 식 (2.1)을 설계하였다. 보상 함수는 위치 오차 및 누적 오차를 감소시키는 성능 항 식 (2.2)와 함께, Lyapunov 안정성 조건을 반영한 식 (2.3)을 포함한다. 식 (2.4)는 Lyapunov Function을 나타내며,  $K_1, K_2, K_3, \omega_1$ , 그리고  $\omega_2$ 는 제어 안정성과 응답 성능 간의 균형을 고려하여 실험적으로 결정하였다. 이때,  $\omega_1$ 과  $\omega_2$ 는  $\omega_1 + \omega_2 = 1$ 의 관계 식을 만족한다. 이를 통해 학습 과정에서 제어 성능 향상뿐만 아니라 안정적인 수렴을 유도한다. 또한, 주파수 응답 함수(FRF) 분석을 통해 도출된 Gain Margin 및 Phase Margin 기준을 바탕으로, 강화학습의 탐색 범위를 Fig. 3과 같이 MATLAB Reinforcement Learning Toolbox에서 요구하는 연속형 박스 기반 탐색 구조로 설계하였다. 이러한 설계는 강화학습 에이전트가 안전한 범위 내에서만 게인을 조정하도록 하여 학습 과정의 안정성을 확보하는 동시에, 성능 향상을 위한 충분한 탐색 자유도를 제공한다.

$$Reward = \omega_1 Reward_P + \omega_2 Reward_S \quad (2.1)$$

$$Reward_P = -K_1 e^2 - K_2 \int e^2 dt - K_3 \left( \int e dt \right)^2 \quad (2.2)$$

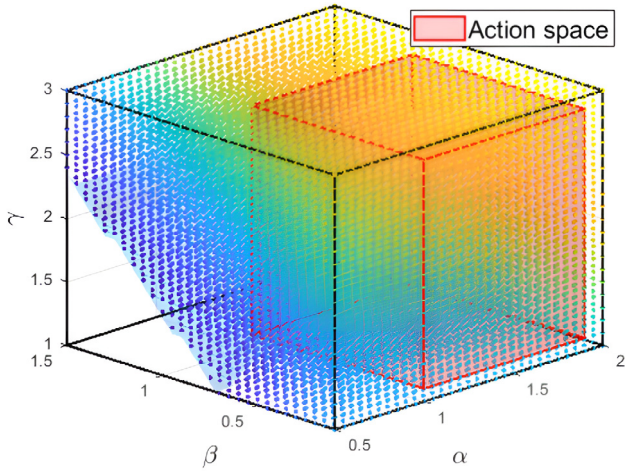


Fig. 3 Action space

$$Reward_s = \begin{cases} -1, & \dot{V} > 0 \\ 0, & else \end{cases} \quad (2.3)$$

$$V = \frac{1}{2}m\dot{e}^2 + \frac{1}{2}ke^2 > 0, \quad \dot{V} = m\dot{e}\ddot{e} + ke\dot{e} \quad (2.4)$$

결과적으로 제안된 RL-PID 제어기는 기존 PID 제어기의 구조적 장점을 유지하면서, 강화학습을 통해 미동 스테이지의 비선형성과 외란에 대한 적응성을 향상시키도록 설계되었다. 이러한 제어 구조는 이중 서보 스테이지의 협조 성능 저하 문제를 완화하고, 고속, 고정밀 반도체 검사 장비에 적용 가능한 안정적인 지능형 제어 프레임워크를 제공한다.

### 3. 시뮬레이션 및 실험 결과

#### 3.1 시뮬레이션 결과

제안한 강화학습 기반 PID (RL-PID) 제어기의 성능을 검증하기 위해 MATLAB/Simulink 환경에서 시뮬레이션을 수행하였다. 미동 스테이지 모델은 앞서 제시한 미동 스테이지의 동특성과 FRF 분석 결과를 반영하여 구성되었다. 강화학습 에이전트는 DDPG 알고리즘을 기반으로 500 에피소드 학습되었으며, 학습 과정 동안 FRF 기반 안정 영역과 Lyapunov 안정성 조건을 만족하도록 탐색 범위를 제한하였다.

Fig. 4는 미동 스테이지의 RL-PID가 100 μm 계단 입력에 대해 어떻게 적용되는지를 나타낸 것으로, 스케일링 인자, 제어 입력, 보상 값, 그리고 위치 응답의 변화를 도사하였다. 계단 입력 직후에는 Ziegler-Nichols 기반 초기 PID 계인으로 인해 큰 과도 오차가 발생하지만, 강화학습 에이전트는 오차와 적분 오차의 변동에 반응하여 스케일링 인자를 빠르게 조정하며 안정 영역으로 수렴하는

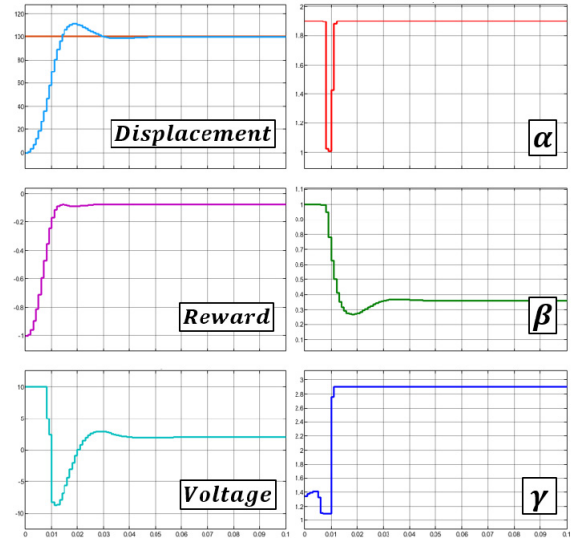


Fig. 4 Step response learning behavior of the RL-PID

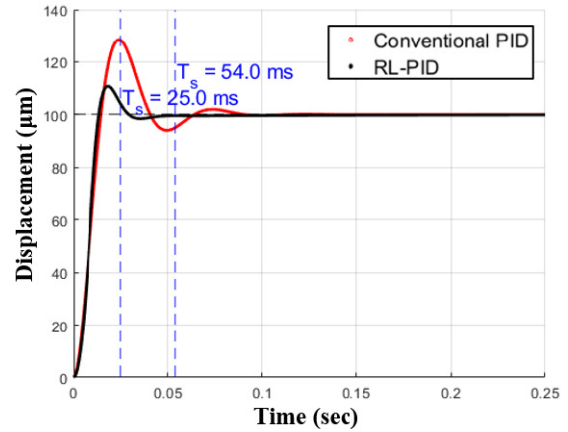


Fig. 5 Step response simulation of the fine stage

정책을 학습한다. 보상 값 역시 시간이 지날수록 꾸준히 증가하여, 에이전트가 안정성과 성능을 동시에 만족하는 제어 정책으로 수렴하고 있음을 보여준다.

기존 입력에 대한 RL-PID 제어기의 성능을 기존 PID 제어기와 비교하였다. 비교군으로 사용된 기존 PID는 MATLAB 내장 Auto-Tuner를 이용해 자동 튜닝하였으며, 이는 산업 현장에서 널리 사용되는 초기 설정 방식이다. 먼저 Fig. 5는 미동 스테이지 단독에 대한 100 μm 계단 입력 응답 비교를 나타낸다. 시뮬레이션 결과, 기존 PID는 28.5%의 오버슈트와 54.0 ms의 정착 시간(5%)을 보였으나, RL-PID는 오버슈트를 10.7%로 억제하고 정착 시간(5%)을 25.0 ms로 단축하여 과도응답에서 뚜렷한 성능 개선을 확인할 수 있다.

Fig. 6은 본 연구의 직렬형 이중 서보 제어 블록 다이어그램을 나타내며, 레이저 간섭계를 주 센서로 사용하여 전체 스테이지의 고해상도 위치 정보를 기반으로 제어 루프가 구성된다. 미동 스테이

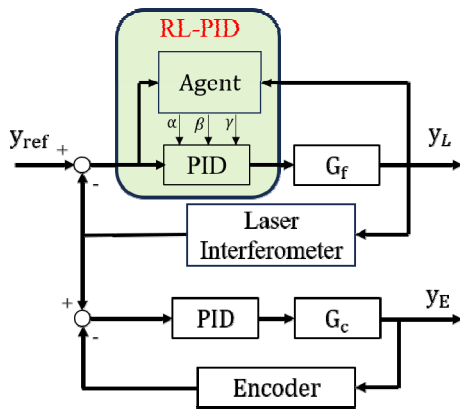


Fig. 6 Serial-type dual-servo control

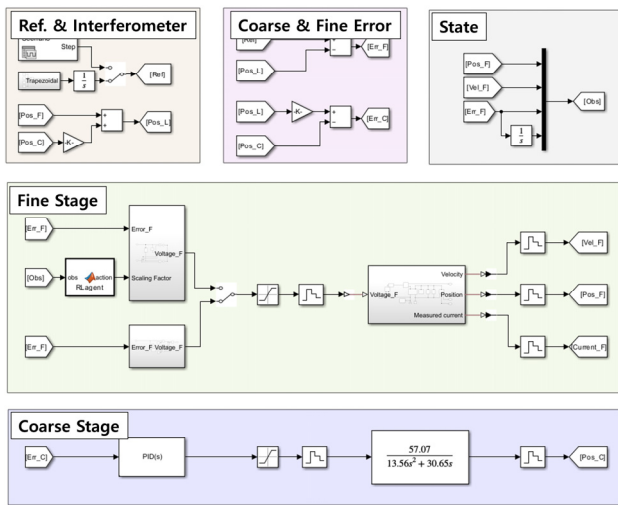


Fig. 7 Simulink model of serial-type dual-servo control

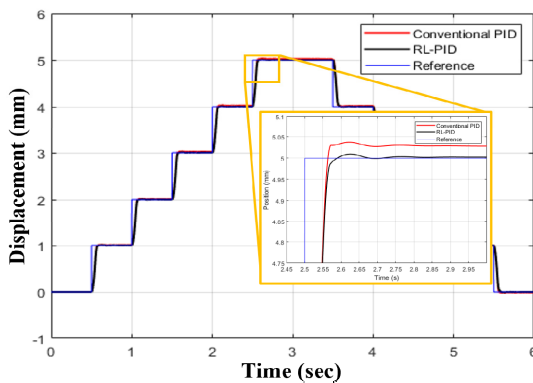


Fig. 8 Step response simulation of the dual-servo stage

지는 참조 입력과 간섭계로 측정된 변위의 차이를 직접 오차로 받아 정밀 보정을 수행한다. 반면, 조동 스테이지는 간섭계 측정값과 리니어 엔코더 측정값의 차이를 오차로 사용하여, 보조 센서인 엔코더의 절대 위치 정보를 활용한 안정적 장거리 이송 제어를 담당한다. 이와 같은 구성은 간섭계의 고해상도 및 고선형성 장점을 유지하면서도, 엔코더의 장거리 안정성 및 강인성을 활용해 두 센서가 상호

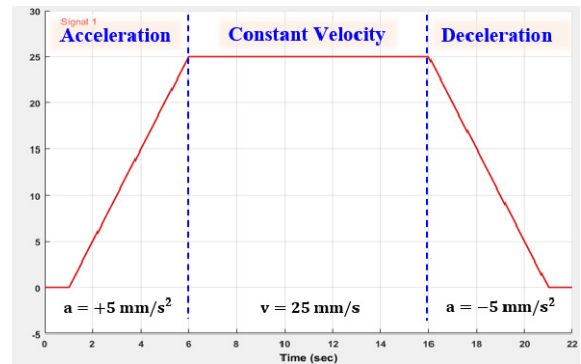


Fig. 9 Trapezoidal velocity profile

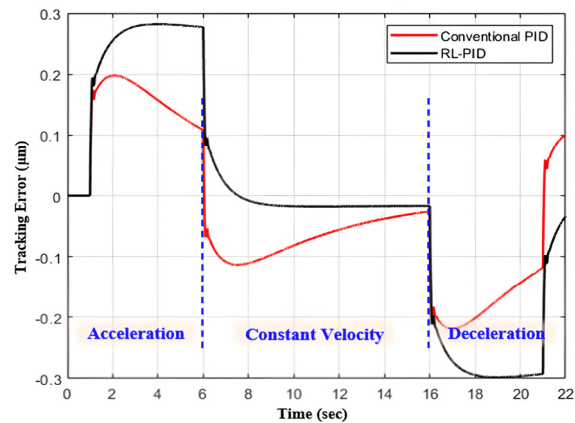


Fig. 10 Trapezoidal velocity response simulation of the dual-servo stage

보완적으로 동작하도록 설계된 것이다. Fig. 7은 MATLAB/Simulink 기반으로 구현한 이중 서보 제어기의 전체 블록도이다.

다음으로 이중 서보 스테이지 전체에 대한 시뮬레이션에서는 조동-미동 스테이지 간 협조 제어 성능을 평가하였다. 먼저 1 mm 계단 입력 시뮬레이션에서 조동 스테이지가 기준 위치를 추종하는 동안 미동 스테이지는 잔여 오차를 보상하며, RL-PID 제어를 적용한 경우 전체 응답에서 오버슈트 감소와 빠른 수렴 특성이 확인되었다. 반면, 기존 PID 제어기만을 적용한 경우에는 동특성 불일치와 결합 효과로 인해 잔류 진동과 정착 시간 지연이 상대적으로 크게 나타났다. Fig. 8은 이중 서보 스테이지에 계단 입력을 인가했을 때의 시뮬레이션 결과를 나타낸다.

스캐닝 환경을 모사하기 위해 Fig. 9와 같이 사다리꼴 속도 프로파일(감가속도  $\pm 5 \text{ mm/s}^2$ , 정속  $25 \text{ mm/s}$ )을 적용하여 추종 성능을 평가하였다. Fig. 10은 해당 사다리꼴 속도 입력에 대한 이중 서보 스테이지의 시뮬레이션 응답을 나타낸다. RL-PID는 정속 구간에서 RMSE 27.5 nm로 매우 우수한 추종 성능을 보였으며, 기존 PID의 73.8 nm 대비 약 62.7% 향상된 결과를 나타냈다. 반면, 가감속 구간에서는 RL-PID가 기존 PID보다 다소 큰 과도 오차를 보이는 구간이 관찰되었다. 이는 강화학습의 보상 구조가 정속 구

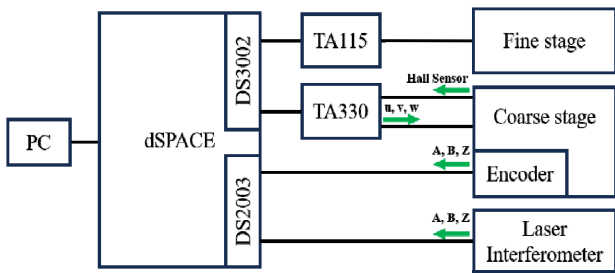


Fig. 11 Hardware configuration block diagram

Table 1 Hardware configuration block diagram

Component	Value	Units
<b>Linear BLDC motor</b>	<b>SWL-DM50</b>	-
Continuous/Max current	3.65 / 11	$A_{rms}$
Force constant	57.07	$N/A_{rms}$
Stroke	470	mm
<b>Voice coil motor</b>	<b>AVA2-20</b>	-
Continuous/Max current	1.4 / 4.2	$A_{rms}$
Force constant	8.35	$N/A_{rms}$
Stroke	20	mm
<b>Coarse motor driver</b>	<b>TA330-E01</b>	-
Supply voltage	48	V
Torque gain	1.0	A/V
Bandwidth	5.0	kHz
<b>Fine motor driver</b>	<b>TA115-E01</b>	-
Supply voltage	24	V
Torque gain	0.2	A/V
Bandwidth	5.0	kHz
<b>Linear encoder</b>	<b>AK MS 15 TTLx50</b>	-
Supply voltage	5	V
Resolution	0.2	$\mu m$
Output frequency	250	kHz
<b>Laser interferometer</b>	<b>RLU10 laser</b>	-
Supply voltage	24	V
Resolution	158	nm
Output update rates	5	MHz

간 오차 최소화를 우선적으로 학습하도록 설계되어, 속도 변화가 급격한 전이 구간에 대한 최적화가 상대적으로 덜 이루어진 결과로 해석된다. 그럼에도 불구하고, 실제 검사 품질은 정속 구간에서의 추종 정밀도가 가장 큰 영향을 미치므로, RL-PID는 전체 스케닝 과정에서 기존 PID 대비 명확한 성능 우위를 제공한다.

이러한 시뮬레이션 결과를 통해 제안한 RL-PID 제어기는 기존 PID 제어기 대비 정착 시간, 오버슈트 및 추종 오차 측면에서 향상된 성능을 보였으며, 이중 서보 스테이지의 협조 제어 성능 저하 문제를 효과적으로 완화할 수 있음을 확인하였다. 다음 절에서는

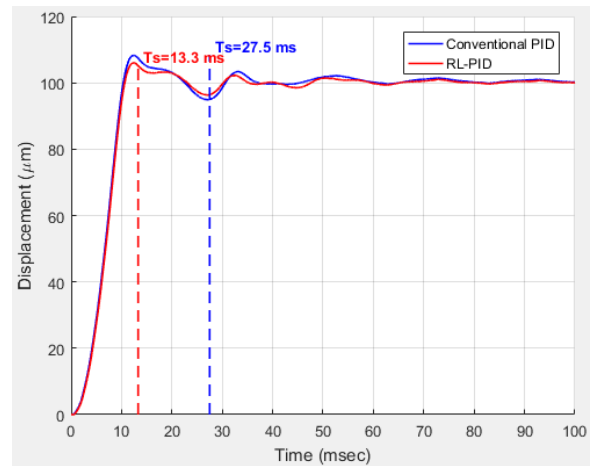


Fig. 12 Step response experiment of the fine stage

dSPACE 기반 실험 시스템을 통해 시뮬레이션 결과의 타당성을 실제 하드웨어 환경에서 검증한다.

### 3.2 실험 결과

제안한 강화학습 기반 PID(RL-PID) 제어기의 실효성을 검증하기 위해 dSPACE 기반의 실시간 제어 시스템을 이용한 실험을 수행하였다. 실험 제어 시스템은 Fig. 11의 제어 선도와 동일한 구조로 구현되었으며, 강화학습을 통해 학습된 RL-PID 정책은 실시간 모델로 변환된 뒤 dSPACE 보드에 삽입되었다. 이때, 고주파(50 Hz, 1  $\mu m$ ) 및 저주파(5 Hz, 5  $\mu m$ ) 외란 특성을 반영하여 추가 학습을 진행하였다. 제어 연산은 dSPACE DSP 프로세서에서 수행되며, 미동 스테이지의 빠른 동 특성을 반영하기 위해 제어 주기는 0.1 ms로 설정하였다.

Table 1은 실험에 사용된 전체 구성 요소의 제원을 정리한 것으로, 조동 스테이지의 BLDC 모터 사양, 미동 스테이지의 VCM 특성, 리니어 엔코더 및 레이저 간섭계의 해상도, 모터 드라이버의 출력 등 핵심 파라미터를 포함한다.

먼저 미동 스테이지에 100  $\mu m$ 의 계단 입력을 인가하여 단일 스테이지에서의 과도응답 특성을

비교하였다. Fig. 12에 나타난 바와 같이, 기존 PID 제어기는 약 27.5 ms의 정착 시간(5%)과 8.3%의 오버슈트를 보였으며, 약 50 Hz 대역의 구조적 외란으로 인해 응답 후반부에서 잔류 진동이 비교적 크게 발생하였다. 반면, RL-PID 제어기는 정착 시간(5%)이 13.3 ms로 기존 PID 대비 약 51% 단축되었고, 오버슈트 또한 6.1%로 감소하여 더욱 빠르고 안정적인 과도응답을 나타냈다. 두 제어기 모두에서 50 Hz 성분의 공통 외란이 관측되었으나, RL-PID는 해당 고주파 성분의 진폭을 억제하여 미동 스테이지의 실제 환경에서 발생하는 주기적 외란에 대해 더 높은 강인성을 보였다.

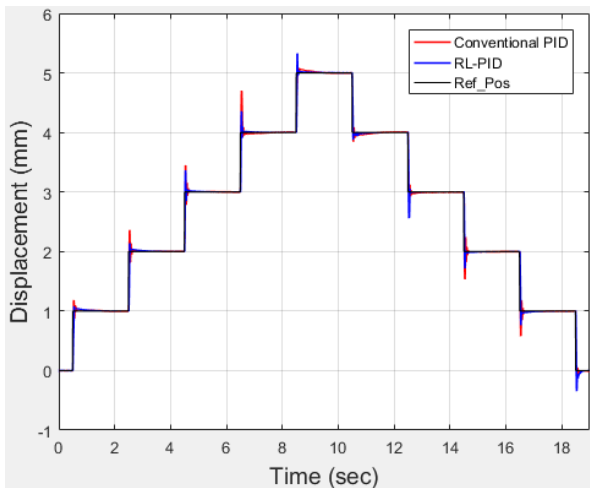


Fig. 13 Step response experiment of the dual-servo stage

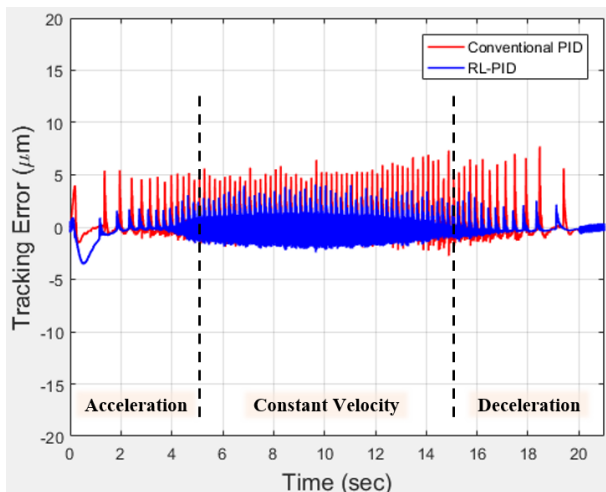


Fig. 14 Trapezoidal velocity response experiment of the dual-servo stage

이후 이중 서보 스테이지에 대해 1 mm 계단 입력을 인가하여 협조 제어 성능을 평가하였다. Fig. 13의 실험 결과에서 기존 PID는 조동 스테이지로부터 전달되는 비선형성, 관성 영향, 그리고 잔류 진동에 의해 단계 전환 구간에서 비교적 큰 오버슈트가 발생하였다. 반면, RL-PID 제어기는 조동 스테이지의 동 특성 변화를 실시간으로 보상하며 상승 및 하강 구간 모두에서 더 작은 진동과 더 빠른 감쇠를 보였다. 정상 상태에서도 RL-PID의 변위 추종 값이 기준 궤적에 더 가깝게 유지되었으며, 조동-미동 스테이지 간의 Coupling Error를 보다 효과적으로 억제한 것으로 나타났다. 이는 RL-PID가 이중 서보 스테이지에서도 조동 스테이지로부터 전달되는 복잡한 동적 영향을 적응적으로 대응하며 협조 제어 성능을 실질적으로 향상시킨 결과이다.

Fig. 14는 사다리꼴 속도 프로파일 입력 실험 결과를 나타낸다. 정속 구간에서의 RMSE는 기존 PID가 1.453  $\mu\text{m}$ 였던 반면 RL-

PID는 0.966  $\mu\text{m}$ 로 약 33% 감소하여, RL-PID가 스캐닝 구간에서 한층 정밀한 추종 성능을 유지함을 확인할 수 있었다. 또한, 실험 결과에서 약 5 mm 간격으로 반복적으로 나타나는 미세한 피크는 조동 스테이지의 코깅 토크의 영향으로, 조동 스테이지의 영구자석 간격(약 30 mm)과 6극 구조가 결합될 때 발생하는 특성적 힘 변동에 의해 설명될 수 있다. 이러한 비선형 구간에서도 RL-PID는 기존 PID보다 오차 진폭을 더 효과적으로 억제하여 강인한 추종 특성을 보였다.

#### 4. 결론

본 논문에서는 반도체 검사 장비의 이중 서보 스테이지 협조 제어 성능 향상을 위해 강화학습 기반 PID(RL-PID) 제어 기법을 제안하였다. 조동 스테이지에는 기존 PID 제어기를 적용하고, 미동 스테이지에는 강화학습을 통해 PID 스케일링 인자를 조정하는 구조를 채택함으로써, 기존 시스템의 안정성을 유지하면서 비선형성과 외란에 대한 적응성을 향상시켰다.

제안된 RL-PID 제어기는 Ziegler-Nichols 방법으로 초기 PID 게인을 설정한 후, DDPG 기반 강화학습을 통해 스케일링 인자를 학습하도록 설계되었다. 또한 FRF 분석을 통해 안정성이 보장되는 학습 영역을 정의하고, Lyapunov 안정성 조건을 보상 함수에 포함시켜 강화학습 과정에서 발생할 수 있는 불안정한 탐색과 과도한 제어 입력을 효과적으로 억제하였다.

MATLAB/Simulink 시뮬레이션 및 dSPACE 기반 실험 결과, 제안한 RL-PID 제어기는 기존 PID 제어기 대비 정착 시간, 오버슈트 및 위치 추종 오차 측면에서 향상된 성능을 보였으며, 고속 동작 조건에서도 안정적인 협조 제어 성능을 유지함을 확인하였다. 이를 통해 제안한 Safe RL-PID 구조가 실제 이중 서보 스테이지 시스템에 적용 가능함을 실험적으로 검증하였다.

한편, 본 연구에서는 실시간 학습 과정에서 발생할 수 있는 불안정성과 계산 부담을 고려하여 Sim-to-Real 방식으로 정책을 적용하였다. DDPG 기반 연속 제어 강화학습은 계산량이 크고 하이퍼파라미터 설정에 민감한 특성을 가지므로, 본 연구에서는 오프라인 학습된 정책을 제어기에 탑재하는 구조를 채택하였다. 향후 연구에서는 제어 연산과 학습 연산을 분리한 온라인 강화학습 구조, 하이퍼파라미터 최적화, 그리고 다양한 초기 조건 및 외란에 대한 일반화 성능 검증을 통해 본 연구의 적용 범위를 확장할 계획이다. 또한 FRF 기반 안정 영역을 박스형 행동 공간으로 제한한 접근은 탐색 효율 측면에서 개선의 여지가 있으므로, 조건부 행동 제한이나 동적 안전 필터를 적용한 보다 유연한 안정성 보장 기법으로의 확장이 가능할 것으로 기대된다.

## 후 기

이 논문은 2025년도 중소벤처기업부의 기술개발사업 지원을 받아 연구되었음(No. S3273805).

이 연구는 대한민국 정부(산업통상자원부)에서 지원하고 한국산업기술진흥원(KIAT)에서 주관하는 RS-2025-02304112 산업혁신 인력양성(HRD) 프로그램의 연구비 지원을 받아 수행되었음.

## References

- [1] Zhu, H., Pang, C. K., Teo, T. J., 2017, A Flexure-Based Parallel Actuation Dual-Stage System for Large-Stroke Nanopositioning, *IEEE Transactions on Industrial Electronics*, 64:7 5553-5563, <https://doi.org/10.1109/TIE.2017.2677306>.
- [2] Choi, Y.-M., Gweon, D.-G., 2011, A High-Precision Dual-Servo Stage Using Halbach Linear Active Magnetic Bearings, *IEEE/ASME Transactions on Mechatronics*, 16:5 925-931, <https://doi.org/10.1109/TMECH.2010.2056694>.
- [3] Pahk, H. J., Lee, D. S., Park., J. H., 2001, Ultra Precision Positioning System for Servo Motor-Piezo Actuator Using the Dual Servo Loop and Digital Filter Implementation, *Int. J. Mach. Tools Manuf.*, 41:1 51-63, [https://doi.org/10.1016/S0890-6955\(00\)00061-4](https://doi.org/10.1016/S0890-6955(00)00061-4).
- [4] Kim, K.-H., Choi, Y.-M., Gweon, D.-G., Hong, D.-P., Kim, K.-S., Lee, S.-W., Lee, M.-G., 2006, Design of Decoupled Dual Servo Stage with Voice Coil Motor and Linear Motor for XY Long Stroke Ultra-Precision Scanning System, *Proc. SPIE 6040, ICMIT 2005: Mechatronics, MEMS, and Smart Materials*, 60401C, <https://doi.org/10.1117/12.664218>.
- [5] Ahn, D., Jin, J.-W., Yun, H., Jeong, J., 2022, Development of a Novel Dual Servo Magnetic Levitation Stage, *Actuators*, 11:6 147, <https://doi.org/10.3390/act11060147>.
- [6] Zhou, Z., Gao, J., Zhang, L., 2024, A Positioning and Tracking Performance-Enhanced Composite Control Algorithm for the Macro-Micro Precision Stage, *Actuators*, 13:11 433, <https://doi.org/10.3390/act13110433>.
- [7] Wang, J., Li, Z., Sun, W., Al-Haddad, K., Gao, H., 2025, Adaptive Robust Tracking Control for Precision Motion Stages via Fully Actuated System Approach, *IEEE 34th International Symposium on Industrial Electronics (ISIE)*, 1-6, <https://doi.org/10.1109/ISIE62713.2025.11124629>.
- [8] Qin, Y., Zhang, W., Shi, J., Liu, J., 2018, Improve PID controller through reinforcement learning, *IEEE CSAA Guidance, Navigation and Control Conference (CGNCC)*, 1-6, <https://doi.org/10.1109/GNCC42960.2018.9019095>.
- [9] Huang, Y., Tang, J., Jiang, J., Li, Y., Liu, M., Chen H., Cao, D., 2025, Physics-Informed Neural Networks Identification and Reinforcement Learning Control for a Nonlinear Vibration System, *J. Vib. Eng. Technol.*, 13 301, <https://doi.org/10.1007/s42417-025-01861-2>.
- [10] Zhu, Y., Bing, Z., Zhao, D., Li, Y., Lai, Z., 2024, Self-tuning of Servo System PID Parameters based on TD3 Algorithm, *International Core Journal of Engineering*, 10:1 262-273, [https://doi.org/10.6919/ICJE.202401\\_10\(1\).0038](https://doi.org/10.6919/ICJE.202401_10(1).0038).
- [11] Ding, Y., Ren, X., Xiaochuan Zhang, X., Liu, X., Wang, X., 2023, Multi-Phase Focused PID Adaptive Tuning with Reinforcement Learning, *Electronics*, 12:18 3925, <https://doi.org/10.3390/electronics12183925>.
- [12] Chowdhury, M. A., Lu, Q., 2023, A Novel Entropy-Maximizing TD3-based Reinforcement Learning for Automatic PID Tuning, *American Control Conference (ACC)*, 2763-2768, <https://doi.org/10.23919/ACC55779.2023.10156246>.
- [13] Sinaei, K., Yazdi., M. R. S., 2021, PID Controller Tuning with Deep Reinforcement Learning Policy Gradient Methods, *Proceedings of the 29th International Conference of Iranian Society of Mechanical Engineers & 8th Conference on Thermal Power Plants*.



### Seung Woo Seo

M.Sc. candidate in the Department of Mechanical Engineering, Ajou University. His research interests include design and control of high precision positioning systems. E-mail: seosw99@ajou.ac.kr



### Yongho Jeon

Professor in the Department of Mechanical Engineering, Ajou University. His research interests include additive manufacturing, welding. E-mail: princaps@ajou.ac.kr



### Moon Gu Lee

Professor in the Department of Mechanical Engineering, Ajou University. His research interests include design and control of high precision positioning systems. E-mail: moongulee@ajou.ac.kr