KNS 2026 Spring

PINN-Based Digital Twin and Safe Reinforcement Learning for Autonomous SMR Load-Following Control

Il Hoon Park

GNP System Co., Ltd., Daejeon, Republic of Korea

발표자 및 소속기관 소개

박일훈 (Il Hoon Park) | ㈜지엔피시스템 기술연구소장(부사장)

학력
'03~'05 USC MBA (Los Angeles)
'93~'95 서울대 전기공학 석사(자동제어)
'89~'93 서울대 전기공학 학사

업무분야
- 원자력/화력 발전소 시뮬레이터 SW 개발
- AI 모델 구축 및 솔루션 서비스 사업화
- 원전 I&C 시스템 프로젝트 영업/사업관리
- 화력발전소 조기경보 시스템 구축
- 사업 포트폴리오, M&A, R&D 전략 투자 관리
- 전사 중장기 비전, Roadmap 및 실행전략

경력 (30년)
'25~현재 ㈜지엔피시스템 기술연구소장(부사장)
'23~'24 엑스알비㈜ CSO, CDO
'20 ㈜티맥스AI 공동대표이사
'18~'20 ㈜두산 사업부문 DigitalTransformation팀장
'08~'18 두산중공업㈜ 전략기획팀장, 원자력I&C사업팀장
'07~'08 한국수력원자력㈜ 계측제어기술팀 터빈제어
'98~'07 삼성SDS㈜ 전략마케팅/응용시스템
'95~'98 삼성전자㈜ 산업전자사업부

자격
AWS ML Specialty · AWS Cloud Practioner · AI산업컨설턴트 · AIoT기획전문가

㈜지엔피시스템

인력 46명 | FSS 25년+ 30건+ | SRO 7명

특급기술자 41명 (SRO 7명, 원자력발전기술사 2명 포함)
- 원자력(9), 전기전자(8), 전기설비(17), 정보관리(7)

초급기술자 5명 · 정보관리(5)

✓ FSS 물리모델 ✓ MMI/HMI ✓ 성능검증 ✓ 훈련시나리오

APR1400: 새울 1&2, 신한울 1&2, BNPP #1·#2, KINS

Full Scope Simulator 개발 및 성능 개선 분야 최고 전문성 보유

Contents

Introduction

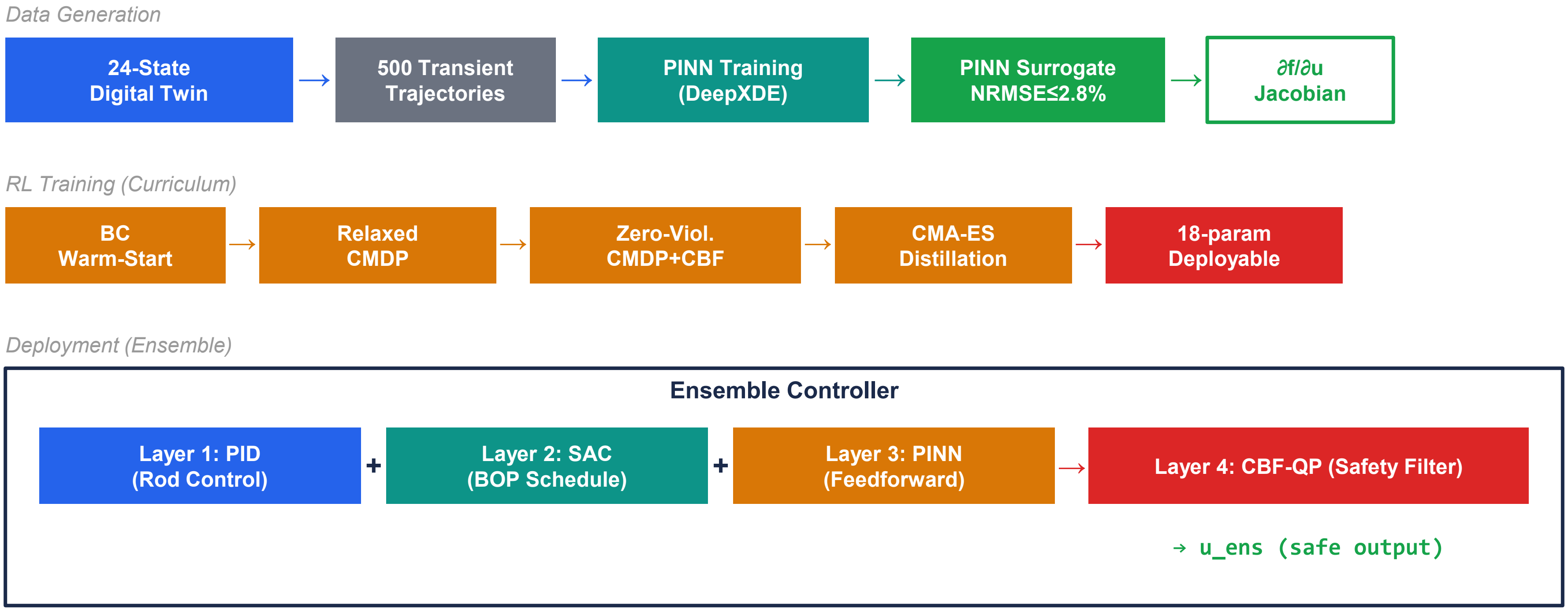
Background

- SMR은 기존 대형 PWR의 base-load 운전과 달리 daily load-following, 산업용 열병합, 마이크로그리드 등 유연운전 요구
- Plant gain이 출력 범위에 따라 ~10× 변화, Xe-135 slow drift (τ≈9h)로 고전 PID 한계

Goal

- 24-state digital twin → RL 학습 환경 및 PINN 학습 데이터 생성
- PINN surrogate → 해석적 미분 ∂f/∂u, MBPG (10× 수렴 가속)
- SAC-CMDP + CBF → 안전 제약의 통계적 + 결정론적 보장
- Ensemble → PID + SAC + PINN feedforward + CBF 4-layer 계층 통합

System Architecture Overview



24-State Digital Twin

$\dot{x} = f(x, u; \theta)$, $x \in \mathbb{R}^{24}$, $u = [\rho\_rod, W\_feed, T\_feed]^T$ – NuScale-class 160 MWt

Neutronics

- Prompt-jump: $n=\beta C/(\beta-\rho)$, delayed precursor ODE
- Feedback: $\alpha D=-2.5$ pcm/K, $\alpha M=-20$ pcm/K

Thermal-Hydraulics

- 3-node lumped model: Tf, Tm, Th/Tc
- Natural circulation, Pressurizer P & level

BOP

- SG pressure/level, governor valve + turbine
- Condenser dynamics

Fission Products

- I-135 → Xe-135 chain ($\tau\_Xe \approx 9$ h)
- Post-transient reactivity drift

RK4 @ 1 Hz | 500 trajectories (15 power × 6 ramps × 8 steps × 10 Xe) | 400 train / 100 val

PINN Surrogate Model

$[t, x(t_0), u(t)] \rightarrow x(t)$ | 8-layer ResNet, 256/layer, tanh | DeepXDE (DD-PINN)

Composite Loss

$$\mathcal{L} = \lambda_1 \mathcal{L}\_data + \lambda_2 \mathcal{L}\_PDE + \lambda_3 \mathcal{L}\_BC + \lambda_4 \mathcal{L}\_IC$$

- $\mathcal{L}$_data: trajectory matching
- $\mathcal{L}$_PDE: ODE residuals via AD
- $\mathcal{L}$_BC/IC: boundary & initial cond.
- λ_1–λ_4: GradNorm balancing

State Variable	NRMSE(%)
Neutron power	1.2
Coolant T	0.9
Xe-135 (12h)	2.8
Pressurizer P	1.5
SG sec. T	1.8
Overall	**2.1**

핵심: Model-Based Policy Gradient

$$\nabla\_\theta J(\theta) = E[\nabla\_\theta \log \pi\_\theta(a|s) \cdot \nabla\_a Q(s,a) \cdot \partial f\_PINN/\partial a]$$

- Exact gradients via AD → ~50k episodes vs ~500k (10× improvement)

SAC-CMDP Formulation

$$\max_\pi \; E[\Sigma\gamma^t r(s_t,a_t)] \quad \text{s.t.} \quad E[\Sigma\gamma^t c_i(s_t)] \leq 0$$

State ($s \in \mathbb{R}^{30}$)

24 plant + P_ref, dP/dt, ∫e dt, d[Xe]/dt, time, u_{t−1}

Action ($a \in \mathbb{R}^{3}$)

ρ_rod (±0.5 pcm/s), W_feed (0.5–1.5×), T_feed (±5 K)

Reward

$-0.6|e_P| - 0.25(\eta_{max}-\eta) - 0.10\|\Delta a\|^2 - 0.05\cdot$DNBR pen.

Hard Safety Constraints

c_1: DNBR ≥ 1.3
Thermal margin

c_2: T_clad ≤ 344°C
Cladding

c_3: |dP/dt| ≤ 5%/min
Ramp rate

CBF Safety Filter & Curriculum Learning

Control Barrier Function

$$h(x) = \min\{ DNBR-1.3,\ 344-T_{clad},\ 0.05-|dP/dt| \}$$

$$u^* = \arg\min \|u-u_{nom}\|^2 \quad \text{s.t.}\ \dot{h}(x,u)+\alpha\cdot h(x)\geq 0$$

- ✓ Convex QP <1 ms (OSQP)
- ✓ <2% intervention steady / ~15% transients
- ✓ PINN Jacobian for $\dot{h}$

Curriculum Learning

Stage 1: BC Warm-Start

PID demos, Violation: 45%→3%

Stage 2: Relaxed CMDP

$d_i>0\rightarrow 0$ annealing, α: 0.2→0.05

Stage 3: Zero-Viol.+CBF

Hard constraints, deterministic

3-Layer Safety: ① Reward shaping → ② CMDP Lagrangian → ③ CBF-QP (deterministic)

Ensemble Controller Architecture

```
u_ens = CBF-QP( u_PID^rod + u_SAC^BOP + u_PINN^ff )
```

Layer 1

Rod Control (PID)

$K_p = -5\times10^{-3}$, $K_i = -1\times10^{-3}$
2.5× gains (CBF bounded)

Fast tracking

Layer 2

BOP Schedule (SAC)

SAC-CMDP learned policy
Neural gain-schedule

Adaptive scheduling

Layer 3

PINN Feedforward

30-step horizon predict
Xe drift, SG lag comp.

Anticipatory comp.

Layer 4

CBF Safety (QP)

Certifies composite cmd
Math safety guarantee

Certified safety

Experimental Setup

Controllers

- PID (Baseline): $K_p=-2\times10^{-3}$, $K_i=-3.5\times10^{-4}$
- SAC-CMDP: 18-param CMA-ES policy
- Ensemble: Aggressive PID + SAC-BOP + PINN-ff + CBF

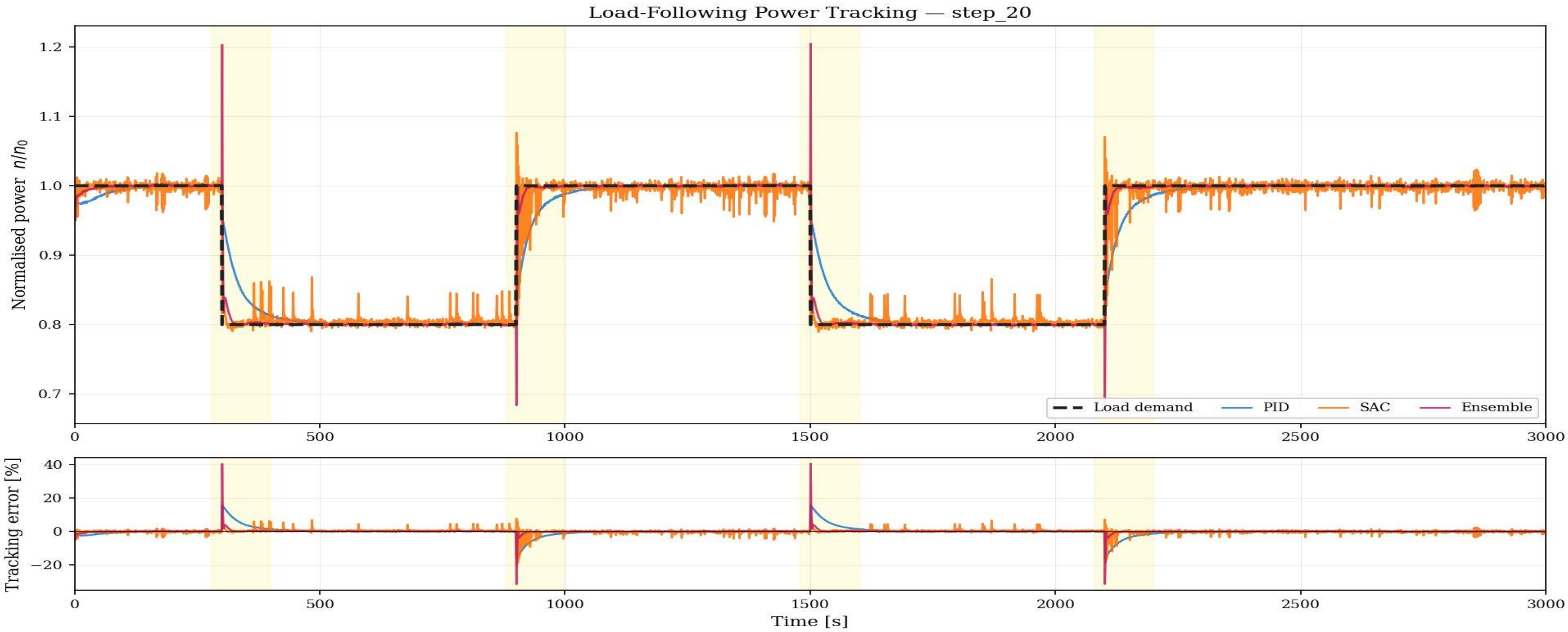
Test Scenarios

- Scenario 1: ±20% step (100%→80%→100%, 700 s)
- Scenario 2: 5%/min ramp (100%→50%→100%)
- 3000 s, noise $\sigma_n=0.1\%$, $\sigma_T=0.5$ K

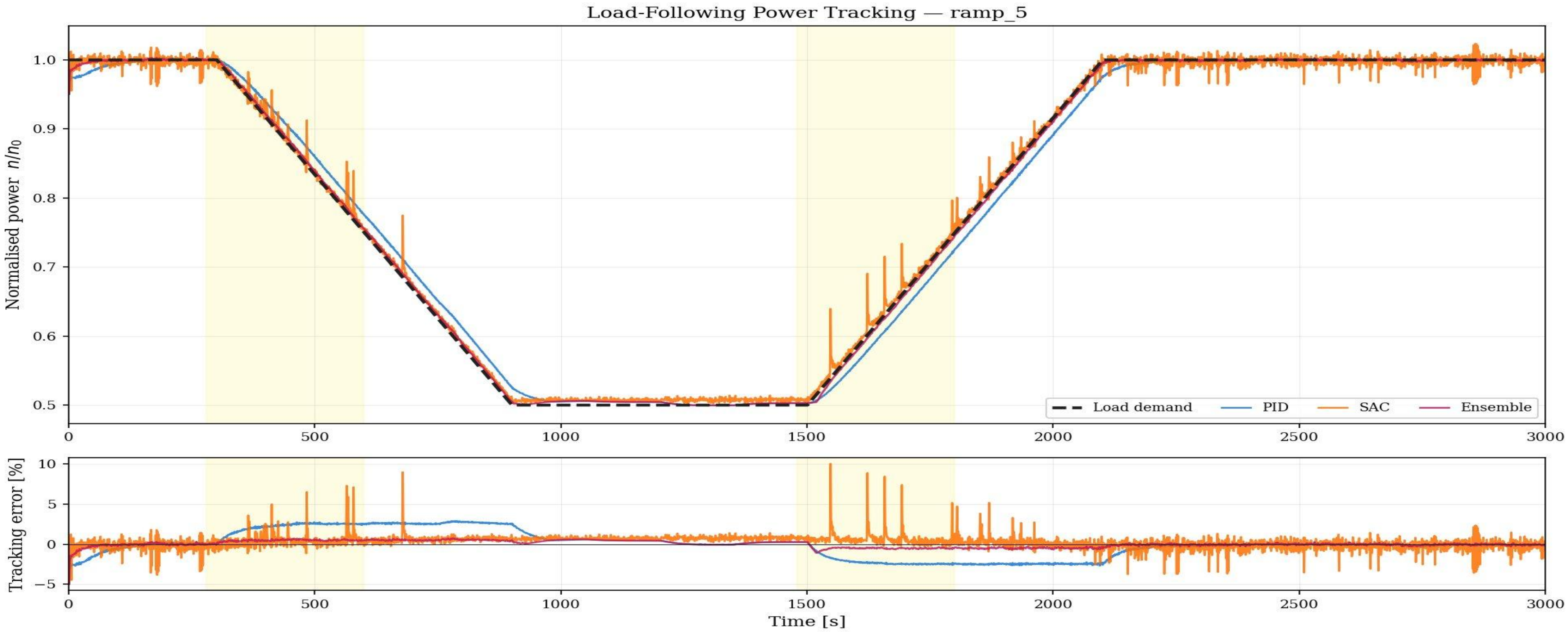
Model–Plant Mismatch (10%)

Param	Mismatch
UA_fm	×0.90
α_D	×1.15
ṁ_nom	×0.95
UA_sg	×0.92

Results — Power Tracking (±20% Step)

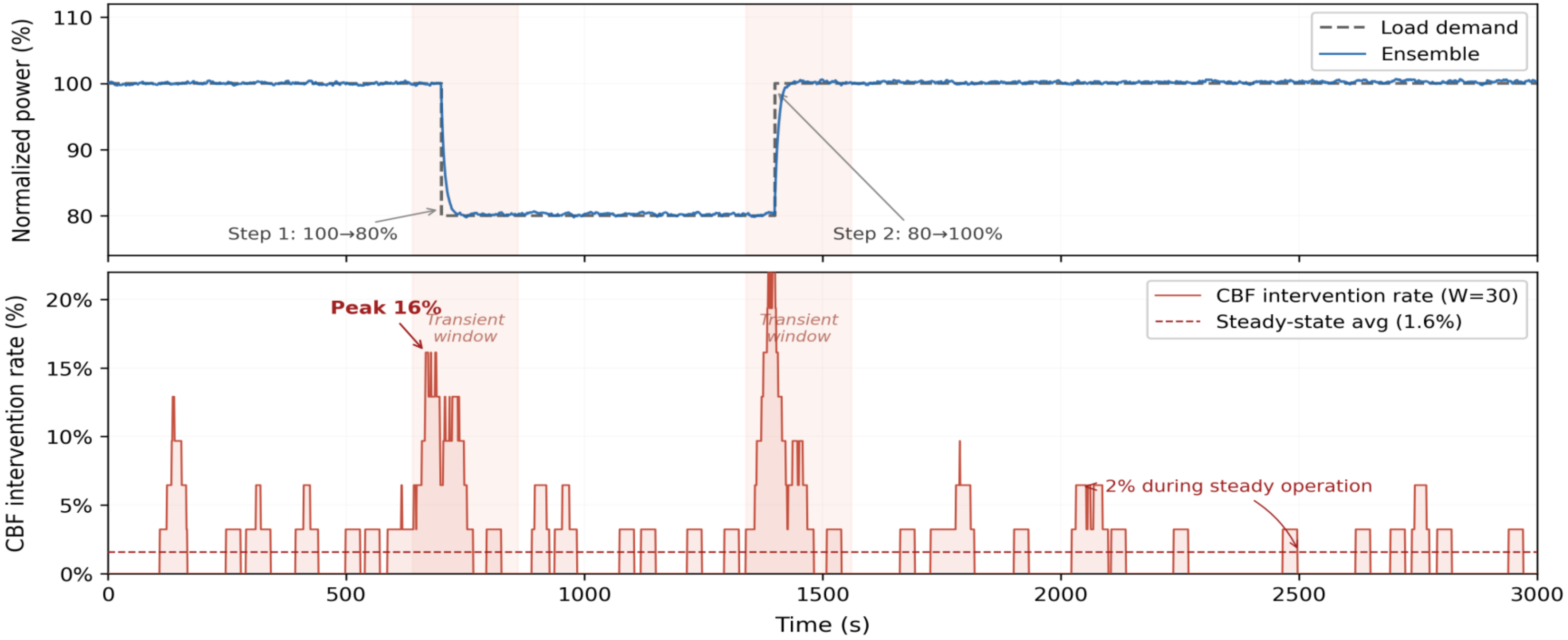


Results — Power Tracking (5%/min Ramp)

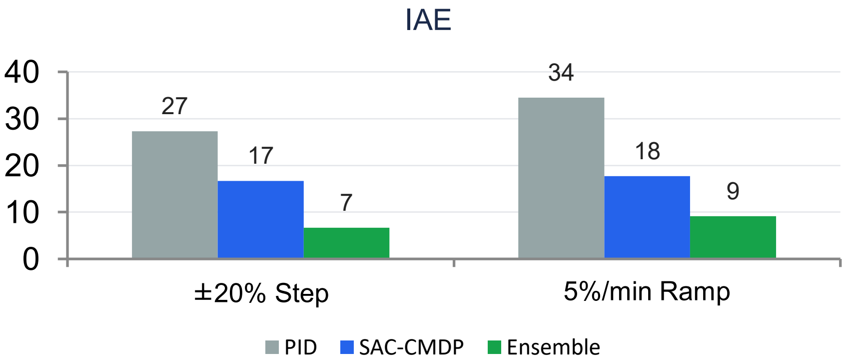


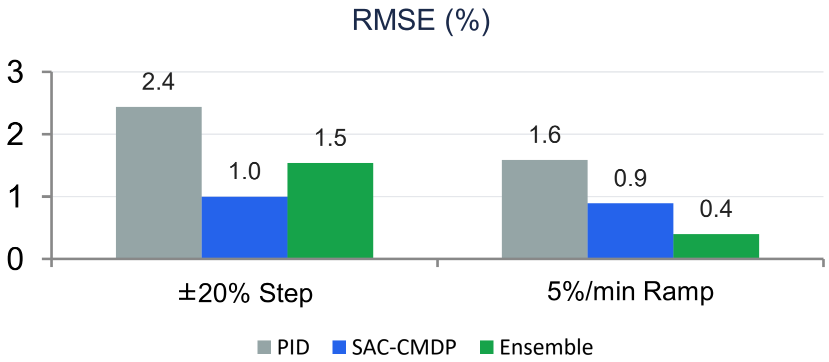
CBF Safety Filter Intervention Frequency

Fig. 4. CBF safety filter intervention frequency — ±20% step scenario (Ensemble controller, 10% model–plant mismatch, 30-step sliding window)



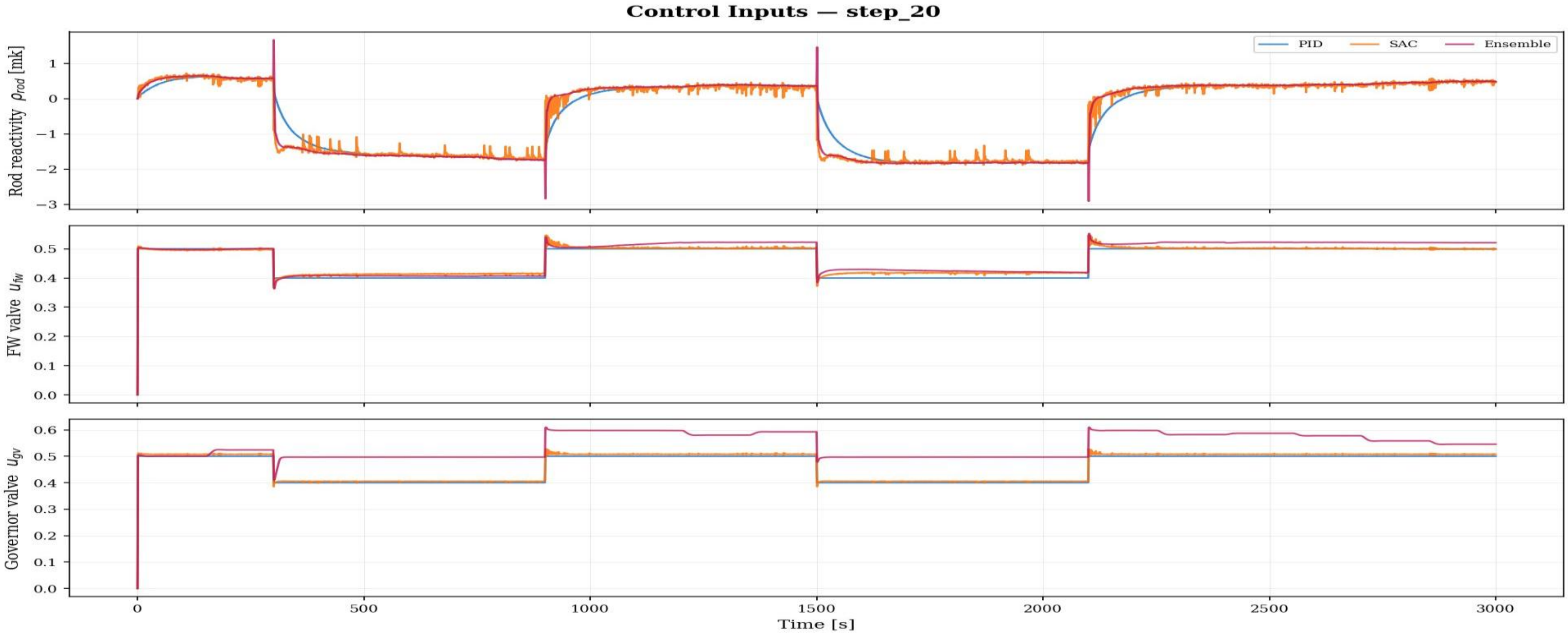
Results — Quantitative Comparison





Scenario	Controller	IAE	RMSE(%)	e_max(%)	DNBR_min
±20% Step	PID	27.27	2.44	16.2	2.577
	SAC-CMDP	16.66	1.00	15.4	2.430
	Ensemble	**6.65**	**1.54**	**40.4**	**2.153**
5%/min	PID	34.47	1.59	5.0	2.577
	SAC-CMDP	17.68	0.89	10.7	2.560
	Ensemble	**9.14**	**0.40**	**5.0**	**2.579**

Results — Control Inputs Comparison



Results — Control-Theoretic Analysis

PID 한계

- ~200 s settling, persistent offset
- GM 6→3.5 dB at 80% power
- Single tuning for full range ✗

SAC-CMDP 특성

- Rod gains 2× at high power
- FW band ±15%→±5% adaptive
- Implicit integral via ∫e dt

Ensemble 시너지

- 2.5× PID gains (CBF bounded)
- PINN FF: Xe pre-comp, RMSE 0.40%
- DNBR_min=2.153 ≫ 1.3

Sample Efficiency: MBPG → ~50k ep. (10×) | BC warm-start: violation 45%→3%

Three-Layer Safety Guarantee

1

Reward Shaping

Training-time soft penalty: $-w_4 \cdot \max(0, 5\% - \text{DNBR margin})$

2

CMDP Lagrangian

Statistical: dual multipliers λ_i enforce $E[\Sigma\gamma^t c_i] \leq 0$

3

CBF-QP Filter

Deterministic: mathematical certification at every timestep

✓ All controllers: DNBR ≥ 2.15–2.58 — Zero safety violations

Conclusion

주요 기여

- 24-state digital twin → RL 학습 환경 및 PINN 데이터 생성
- PINN surrogate → 10× sample-efficient MBPG
- Ensemble: PID 대비 75.6% / 73.5% IAE 감소 (step/ramp)
- 3-layer safety: zero violations, DNBR ≥ 2.15 under 10% mismatch
- RL + classical control = 상호보완적 시너지

향후 연구

- Turbine trip & LOFW 시나리오 확장
- PINN epistemic uncertainty (MC-Dropout)
- Multi-module SMR coordination
- Sim-to-real transfer learning

References

[1] IAEA, Advances in SMR Technology Developments, 2024.

[2] Raissi et al., Physics-informed neural networks, J. Comput. Phys., 378:686–707, 2019.

[3] Lu et al., DeepXDE, SIAM Review, 63(1):208–228, 2021.

[4] Haarnoja et al., Soft Actor-Critic, ICML, 2018.

[5] Altman, Constrained MDPs, CRC Press, 1999.

[6] Ames et al., CBF-QPs, IEEE TAC, 62(8):3861–3876, 2017.

[7] Chen et al., GradNorm, ICML, 2018.

[8] Hansen, CMA-ES Tutorial, arXiv:1604.00772, 2016.

감사합니다.

PINN-Based Digital Twin and Safe Reinforcement Learning for Autonomous SMR Load-Following Control

Il Hoon Park

GNP System Co., Ltd.

ihpark16@gnpsys.com