

Multi-Stage Etch Recipe Inverse Design: Bridging Expert Behavioral Cloning and Reinforcement Learning

Jianming Guo¹, Mingqiang Geng¹, Zhengming Liu², Dianyu Qi^{1,*}, Dong Ni^{1,*}, and Dawei Gao^{1,2}

¹College of Integrated Circuits, Zhejiang University, Hangzhou 311200, China

²Zhejiang ICsprout Semiconductor Co., Ltd., Hangzhou 311200, China

Email: Jianming Guo(21826022@zju.edu.cn)

* Corresponding author: dni@zju.edu.cn, qidianyu@zju.edu.cn

Abstract—Multi-stage plasma etching recipe development still relies heavily on expert experience and costly trial-and-error, especially when new target profiles must be realized across coupled film-transfer stages. In this work, we propose a target-driven inverse design framework that bridges expert behavioral cloning (BC) and offline reinforcement learning (RL) for multi-stage etch recipe generation. BC is first used to learn stage-wise recipe generation from historical expert trajectories, providing a stable baseline for recipe prediction. Building on this BC-initialized baseline, offline RL is then introduced as a conservative refinement mechanism on a fixed dataset generated from a calibrated digital twin, enabling the model to move beyond memorization and learn how recipe adjustments should change with different target profiles. The framework is evaluated on a four-stage etch flow built on a fab-calibrated simulation environment. Results show that the proposed method preserves the strong in-distribution accuracy of BC while substantially improving generalization to unseen targets, achieving a CD error of 9.4 nm and SSIM of 0.968. These results support AI-assisted etch recipe inverse design as a practical complement to expert-guided process development.

Keywords—plasma etching, semiconductor manufacturing, process simulation, recipe inverse design, offline reinforcement learning

I. INTRODUCTION

Multi-stage plasma etching is a key process in semiconductor manufacturing, yet recipe development remains challenging because each etched profile becomes the geometric boundary condition for the next stage. As a result, the problem is sequential, strongly coupled, and still largely explored by costly empirical design of experiments (DOE). To reduce this dependence on wafer-consuming trial-and-error, simulation-driven virtual manufacturing and data-driven process modeling have attracted growing interest. Physics-based TCAD and calibrated digital twins provide a practical route to generate process knowledge beyond limited fab experiments, while AI offers an efficient way to translate this process knowledge into target-driven recipe design and optimization for plasma etching[1].

However, most existing methods still treat etching as a static mapping problem. Behavioral Cloning (BC) is attractive in this context because it directly learns from expert trajectories and naturally captures process-of-record (POR)-centered engineering practice. Yet BC mainly reproduces known

mappings and does not readily infer how recipes should shift for unseen targets.

To address this limitation, we propose a multi-stage etch recipe inverse design framework that bridges expert Behavioral Cloning and Reinforcement Learning (RL). BC first provides a stable expert prior by learning stage-wise recipe decisions from historical POR-centered trajectories. Offline RL is then used as a conservative refinement mechanism in a fab-calibrated digital twin, enabling target-dependent recipe adjustment beyond direct imitation. In this way, the framework moves from experience imitation toward target-oriented inverse design while maintaining consistency with physically meaningful process conditions.

II. METHODOLOGY

The target application is a shallow trench isolation (STI) pattern-transfer process developed on a 12-inch production line. The etch is performed on a Lam Research Kiyo45 inductively coupled plasma (ICP) chamber, which is widely deployed in high-volume manufacturing. The film stack under study consists of photoresist (PR), bottom anti-reflective coating (BARC), SiO₂, Si₃N₄, pad oxide, and silicon. As illustrated in Fig. 1, the complete pattern transfer is decomposed into four sequential stages: Stage 1 (BARC open), Stage 2 (oxide hard-mask etch), Stage 3 (nitride hard-mask etch), and Stage 4 (silicon main etch). Each stage is governed by 8-10 controllable parameters, including source and bias RF power, chamber pressure, process time, and gas flow rates.

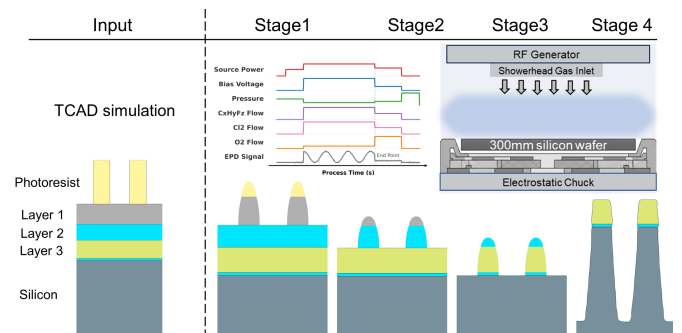


Fig. 1. Overview of the multi-stage etch flow and inverse-design problem. From left to right: initial film stack and simulation input, stage-wise process parameters, profile evolution across the four etch stages, and schematic of the ICP chamber.

The digital twin is built on Synopsys Sentaurus Topography. A reactor-scale plasma model provides ion and neutral flux information under given recipe conditions, and a feature-scale particle Monte Carlo model predicts trench evolution through ion-assisted etching, chemical etching, redeposition, and passivation. Surface kinetic parameters were calibrated against TEM and CD-SEM data from the Kiyo45 platform using our earlier simulation-assisted inverse optimization workflow. After calibration, the simulator reproduces trench profiles with depth error within 7% and good agreement in CD and sidewall shape across all four stages[2,3]. Representative stage-wise comparisons between TEM observations and simulated profiles are shown in Fig. 2. This level of fidelity is critical: the downstream learning framework relies on physically meaningful intermediate-state transitions, not only final profile shapes.

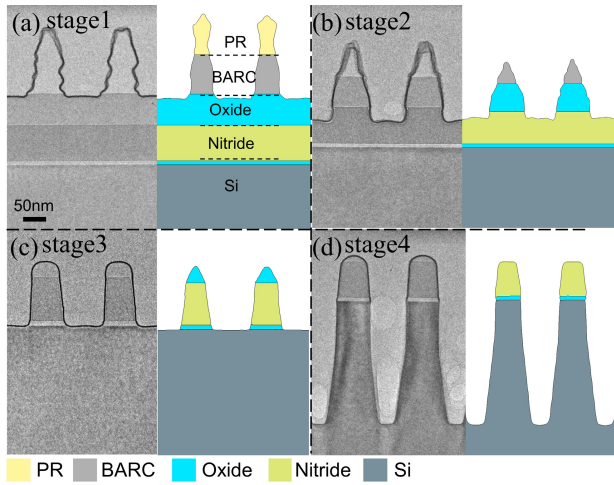


Fig. 2. Stage-wise validation of the calibrated digital twin against experimental TEM cross-sections. (a)–(d) correspond to Stage 1 BARC open, Stage 2 oxide hard-mask etch, Stage 3 nitride hard-mask etch, and Stage 4 silicon main etch.

The multi-stage recipe generation problem is formulated as a finite-horizon target-conditioned decision problem. The horizon is fixed to four, corresponding to the four etch stages. At each stage, the state consists of the current intermediate profile, the desired final target profile, and the stage index. The action is the recipe vector of the current stage, including RF power, pressure, etch time, and gas flow rates. The transition is provided by the calibrated Sentaurus Topography digital twin, which maps the current profile and selected recipe to the next etched profile. The terminal objective is defined according to the mismatch between the achieved final profile and the target profile, including CD error, trench depth error, and pixel-level profile loss. This formulation allows the model to optimize a sequence of physically executable stage-wise recipes rather than a single static recipe vector.

Based on the above target-conditioned formulation, the proposed method follows a two-phase training strategy, as illustrated in Fig. 3. In Phase 1, a multi-stage actor network π_θ is pre-trained by Behavioral Cloning (BC) on the original POR-centered expert trajectories. The actor uses a shared backbone with stage-specific output heads to handle the different recipe dimensions of the four etch stages. The BC training objective is

to minimize the mean squared error between the actor-predicted recipe and the expert recipe:

$$L_{BC} = E_{(s_t, a_t) \sim D_{exp}} \left[\|\pi_\theta(s_t) - a_t\|_2^2 \right] \quad (1)$$

where s_t denotes the stage-wise input state, a_t denotes the expert recipe action, $\pi_\theta(s_t)$ is the recipe predicted by the actor network, and D_{exp} is the historical expert trajectory dataset. This step provides a stable expert prior and anchors the policy within the physically meaningful POR-centered process window.

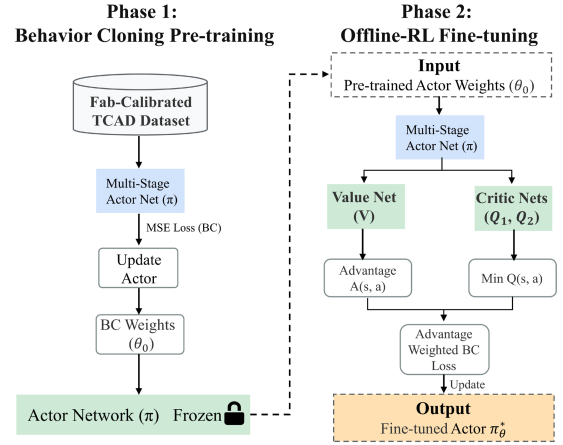


Fig. 3. Training framework of the proposed method. Phase 1: Behavioral Cloning pre-training learns a multi-stage actor from calibrated TCAD trajectories. Phase 2: offline RL fine-tuning initializes from the pretrained actor, trains value and critic networks with the actor frozen, then refines the actor through conservative advantage-weighted updates.

In Phase 2, the BC-initialized actor is refined using offline reinforcement learning on the cross-goal augmented dataset. The offline RL stage is implemented using Implicit Q-Learning (IQL) [4], which is suitable for conservative policy improvement from fixed datasets because it avoids explicit out-of-distribution action maximization. During refinement, the actor is first frozen while the value network and two critic networks are trained using the offline trajectory data. The advantage value is estimated as:

$$A(s_t, a_t) = \min[Q_1(s_t, a_t), Q_2(s_t, a_t)] - V(s_t) \quad (2)$$

where Q_1 and Q_2 are the two critic networks, and V is the value network. The estimated advantage measures whether a recipe action is expected to produce a better target-profile match than the average action observed at the same state.

After the value and critic networks become stable, the actor is unfrozen and fine-tuned through advantage-weighted behavioral cloning. The actor update is written as:

$$L_{actor} = E_{(s_t, a_t) \sim D} \left[w(s_t, a_t) \|\pi_\theta(s_t) - a_t\|_2^2 \right] \quad (3)$$

Actions with higher estimated advantage receive larger weights during actor fine-tuning, allowing the policy to selectively strengthen recipe adjustments that are predicted to reduce the target-profile mismatch. In this way, IQL does not replace expert imitation; instead, it conservatively refines the BC policy while keeping the updated actor close to the POR-centered process domain. In this work, the actor is frozen during

the 200-epoch Q/V warm-up stage and is then fine-tuned for up to 400 epochs with a conservative learning rate of 5×10^{-5} .

The hyperparameters used for the offline RL refinement are summarized in Table I; the value network is trained by expectile regression ($\tau = 0.7$), the critics by temporal-difference regression ($\gamma = 0.99$) with target networks updated by Polyak averaging (0.005), and the actor is refined by advantage-weighted updates with inverse temperature $\beta = 1.0$, using a Q/V warm-up before a conservative actor fine-tuning.

Table I. Hyperparameters of the offline RL (IQL) refinement.

Hyperparameter	Value
Backbone / head	2-layer MLP, 512 units, LayerNorm + ReLU / Tanh head
Critic & value nets	2-layer MLP, 512 units, LayerNorm + ReLU
Expectile (τ)	0.7
Inverse temperature (β)	1.0 (weight clip 100)
Discount factor (γ)	0.99
Target update rate (Polyak)	0.005
Reward scale	10
Optimizer / batch size	Adam / 256
Q/V warm-up	200 epochs, lr 3e-4 (actor frozen)
Actor fine-tune	up to 400 epochs, lr 5e-5, early stopping

III. RESULTS AND DISCUSSION

To evaluate the proposed framework, the predicted recipes were executed in the calibrated Sentaurus simulator, and the resulting profiles were compared with the ground-truth targets using CD error, depth error, MSE, and SSIM, where SSIM values closer to 1 indicate better structural agreement[5]. Two evaluation scenarios were considered: in-distribution targets drawn from the training domain and completely unseen targets held out during training. The results are summarized in Table II.

Table II. Performance comparison on in-distribution and unseen targets.

Method		CD Error (nm)	Depth Error (nm)	MSE↓	SSIM↑
BC	In-dist.	10.66	22.86	0.0059	0.958
Ours	In-dist.	10.26	30.32	0.0056	0.961
BC	Unseen	(failed)	(failed)	—	—
Ours	Unseen	9.44	28.44	0.0045	0.968

For in-distribution targets, both methods achieve comparable accuracy, showing that offline RL refinement preserves the expert prior learned by BC. The proposed method slightly improves CD error, MSE, and SSIM, reducing the CD error from 10.66 to 10.26 nm and increasing SSIM from 0.958 to 0.961. Using a 200 nm pitch as reference, these CD errors correspond to relative deviations of 5.33% and 5.13%,

respectively. Fig. 4 shows a representative in-distribution example. In this case, the recipe predicted by the proposed method reproduces the overall trench shape, sidewall profile, and dual-feature layout of the ground-truth target. The predicted critical dimension is 245.9 nm versus 236.9 nm for the ground truth (a 9.0 nm difference), and the predicted trench depth is 703.8 nm versus 682.8 nm (a 21.0 nm difference), consistent with the averaged in-distribution errors reported in Table II. Both the masking stack and the etched silicon profile are well preserved, indicating that the offline-RL refinement maintains the structural fidelity of the BC prior on in-distribution targets.

The main difference appears on unseen targets. BC fails to produce valid profiles (predicted recipes fall outside the calibrated process window and yield non-physical or non-converging simulations), whereas the proposed method remains stable, achieving 9.44 nm CD error, 28.44 nm depth error, 0.0045 MSE, and 0.968 SSIM. Relative to a 200 nm pitch, the CD deviation is 4.72%. Relative to a 500 nm depth scale, the depth deviation is 5.69%. This result indicates that BC mainly interpolates among known expert trajectories, while the proposed framework can generalize recipe adjustments to previously unseen target geometries.

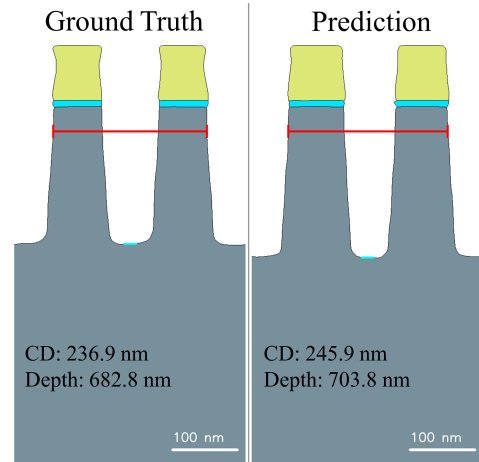


Fig. 4. Representative in-distribution result. Ground-truth profile (left) vs. profile predicted by the proposed offline-RL method (right).

From a process-development perspective, this is the more relevant capability, since engineers often need recipes for targets beyond the existing POR database. The proposed framework generates viable candidate recipes within the calibrated simulator, reducing dependence on additional wafer-consuming experiments.

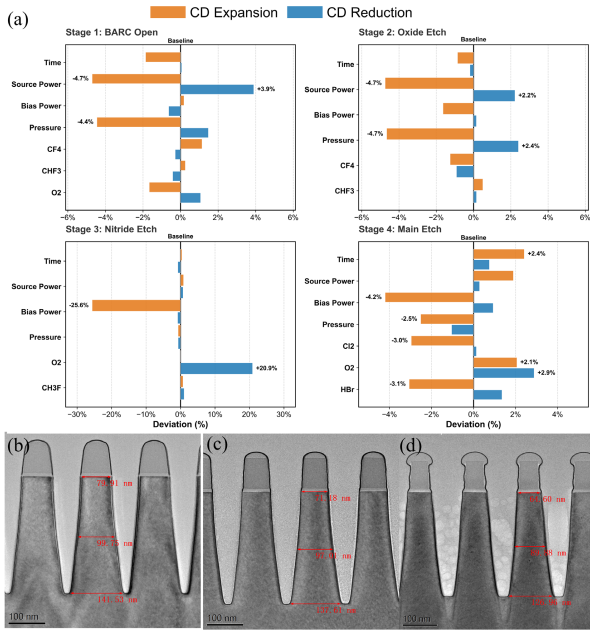


Fig. 5. Process-interpretability analysis of the recipe adjustments generated by the proposed BC-initialized IQL framework. (a) Stage-wise recipe deviations relative to the baseline recipe for CD expansion and CD reduction targets. Orange bars denote the CD expansion strategy, while blue bars denote the CD reduction strategy. (b–d) Cross-sectional TEM profiles corresponding to the CD expansion, baseline, and CD reduction cases, respectively.

To further examine whether the generated recipes are physically meaningful, Fig. 5 analyzes the stage-wise recipe adjustments predicted by the proposed BC-initialized IQL framework for two opposite targets: CD expansion and CD reduction. The diverging bar chart in Fig. 5(a) shows that the model does not simply apply a uniform shift to all process parameters. Instead, it selects different control knobs at different etch stages, indicating that the learned policy captures the coupled nature of multi-stage profile transfer.

For the CD reduction target, the most pronounced adjustment occurs in Stage 3, where the O₂ flow is increased by 20.9%. This suggests that the model uses enhanced chemical trimming during the nitride etch stage to reduce the lateral feature dimension before the final silicon transfer. The resulting TEM profile in Fig. 5(d) confirms this trend, with the top CD reduced from 71.18 nm in the baseline case to 64.60 nm. The middle and bottom CDs are also reduced from 97.01 nm and 137.61 nm to 89.88 nm and 126.96 nm, respectively, indicating that the CD reduction is transferred through the full profile rather than being limited to the top opening.

For the CD expansion target, the model chooses a different strategy. Instead of simply reversing the O₂ adjustment, it strongly reduces the Stage 3 bias power by 25.6%, which weakens directional ion bombardment and relaxes the anisotropic confinement of the profile. As shown in Fig. 5(b), this recipe produces a wider profile, increasing the top CD from 71.18 nm to 79.91 nm. The middle and bottom CDs also increase to 99.75 nm and 141.53 nm, respectively. These opposite but physically interpretable adjustment patterns

demonstrate that the proposed framework learns target-dependent recipe control rather than memorizing a fixed correction rule.

IV. CONCLUSION

In this work, we proposed a target-driven inverse design framework for multi-stage plasma etch recipe generation by combining expert Behavioral Cloning with offline reinforcement learning. BC captures POR-centered engineering practice and provides a stable expert prior, while IQL refines this policy in a fab-calibrated Sentaurus digital twin to improve target-dependent recipe adjustment. Validation on a four-stage STI etch flow shows that the proposed method maintains strong in-distribution accuracy and enables valid recipe generation for unseen target profiles. In addition, the stage-wise recipe adjustment analysis indicates that the generated recipes are physically interpretable: the model selects different process levers for CD expansion and CD reduction, such as bias-power relaxation and O₂-assisted chemical trimming. These results demonstrate the potential of expert-guided offline RL as a practical path toward AI-assisted plasma etch process development with reduced dependence on wafer-consuming trial-and-error.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (Grant No. 2022YFF0605800), the Zhejiang Key R&D Program (Grant No. 2025C01206), and the Scientific Research Fund of Zhejiang University (Grant No. XY2025020).

References

- [1] K. J. Kanarik et al., *Nature*, vol. 616, pp. 707–711, 2023. doi: 10.1038/s41586-023-05773-7.
- [2] J. Guo, M. Geng, K. Ren, D. Ni, and D. Gao, *IEEE Access*, vol. 12, pp. 127065–127073, 2024. doi: 10.1109/ACCESS.2024.3444454.
- [3] J. Guo, K. Ren, D. Ni, and D. Gao, *J. Vac. Sci. Technol. A*, vol. 43, 023003, 2025. doi: 10.1116/6.0004166.
- [4] I. Kostrikov, A. Nair, and S. Levine, “Offline reinforcement learning with implicit Q-learning,” *arXiv:2110.06169*, 2021. doi: 10.48550/arXiv.2110.06169.
- [5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. doi: 10.1109/TIP.2003.819861.