Solver-Native Adjoint Attribution for Trimming Inverse-Designed Photonic Layouts

Junho Park[1,2,*], Taehan Kim[3,*], and Di Liang[1, 2 †]
1. Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
2. Quantum Research Institute, University of Michigan, Ann Arbor, MI, USA
3. College of Computing, Data Science, and Society, University of California, Berkeley, CA, USA
*These authors contributed equally; † liangdi@umich.edu

***Abstract*—Inverse-designed photonic components often retain high-index material (e.g., silicon) chunks that contribute little to performance. We introduce AIG and OPA, solver-native attribution tools for necessity-driven cleanup. Across sharp bends and multiplexers, simulations remove 8.5% design-region silicon with sub-0.15-dB penalties; OPA needs zero post-optimization adjoint solves.**

I. Introduction

Adjoint inverse design enables ultra-compact photonic devices with many geometric degrees of freedom [1]–[3]. The resulting binary masks, however, often retain peripheral islands and disconnected fragments w hose optical value is unclear, increasing patterned area, contour complexity, and foundry-handoff burden.

Because inverse-designed geometries are non-intuitive, manual ”cleanup” is risky: small features can be critical, while larger fragments may contribute little. Cleanup therefore needs physics-aware ranking and electromagnetic verification.

Here, we study a post-optimization trimming workflow for adjoint-designed photonic layouts. The method scores connected silicon chunks, tests low-attribution peripheral regions, and accepts each edit only after a forward electromagnetic check. We evaluate WDMs and bends, but **the workflow applies broadly because it requires only a final layout, adjoint-gradient information, and forward verification.**

II. Trimming Method

Fig. 1 summarizes the method. We compare two solver-native attribution scores. Adjoint integrated gradients (AIG) samples gradients along a baseline-to-design path [4]–[6]:

$$\mathrm{AIG}_i = (\rho_i^\star - \rho_{0,i})\frac{1}{m}\sum_{k=1}^{m}\frac{\partial F(\rho_0 + \alpha_k(\rho^\star - \rho_0))}{\partial \rho_i}, \qquad (1)$$

where F is the device figure of merit, ρ_0 is the baseline layout, $\rho^\star$ is the final device layout, and $m = 24$. AIG therefore requires additional post-optimization adjoint evaluations.

Optimization-path attribution (OPA) instead *reuses* the gradients already computed during inverse design:

$$\mathrm{OPA}_j = \sum_{t=0}^{T-1}\frac{g_{t,j} + g_{t+1,j}}{2}\left(p_{t+1,j} - p_{t,j}\right), \qquad (2)$$

where p_t is the optimizer parameter vector and $g_t = \partial F/\partial p_t$. Thus, unlike AIG, OPA adds zero post-optimization adjoint runs because it uses information already produced during the adjoint optimization process. For cleanup, each attribution map is reduced to chunk-level scores over connected peripheral

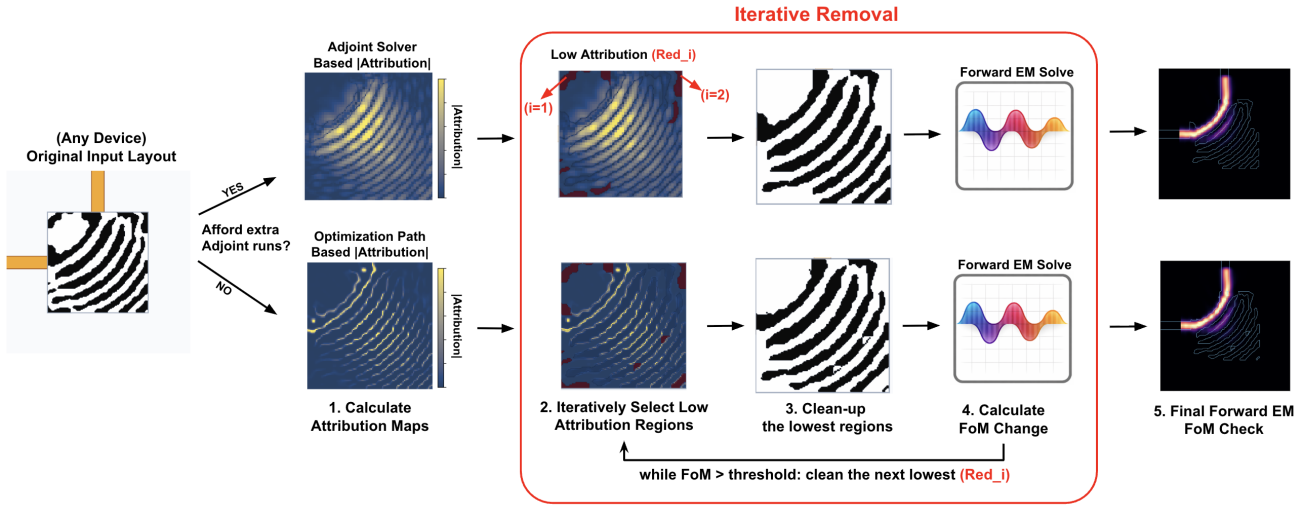

Fig. 1. Generic iterative clean-up flow. AIG uses additional adjoint samples, while OPA reuses gradients already produced during optimization. Low-attribution peripheral chunks are removed only when a forward EM solve confirms the FoM change.

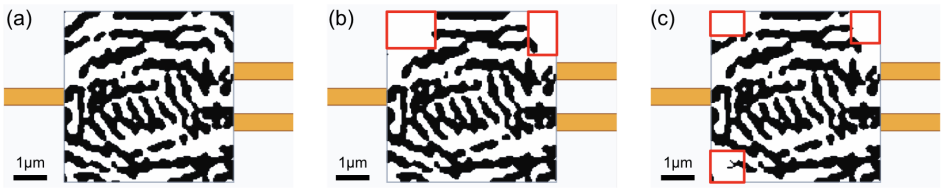

Fig. 2. Representative 1×2, 5-μm WDM trimming result. (a) Original layout. (b) AIG-cleaned layout. (c) OPA-cleaned layout. Both methods remove peripheral silicon chunks while preserving the main routing geometry; full access ports are shown and port corridors are preserved.

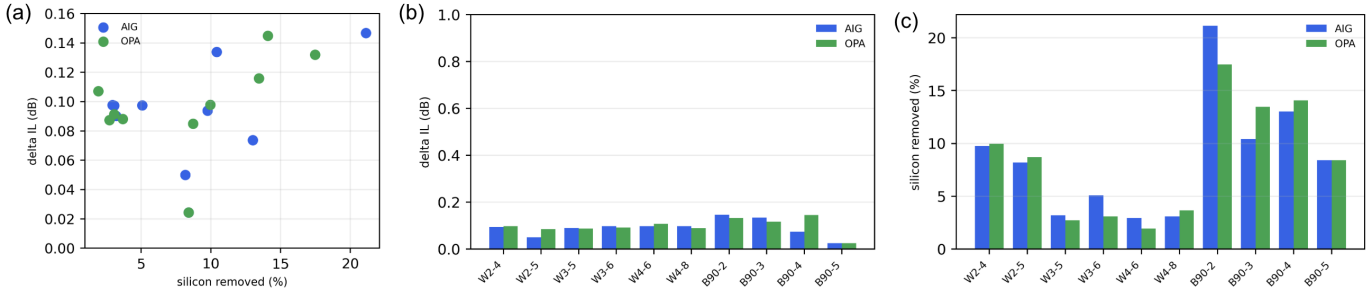

Fig. 3. Ten-device cleanup summary. (a) Removed-silicon versus insertion-loss tradeoff. (b) Per-device insertion-loss penalty. (c) Per-device design-region silicon removed. AIG and OPA achieve similar cleanup behavior across WDMs and bends.

silicon regions, with port corridors protected. For a chunk C, we score its boundary-filtered pixels $B(C)$ by

$$S(C) = \text{median}_{i \in B(C)} |A_i|, \tag{3}$$

where A_i is the AIG or OPA attribution. Low-score chunks are tested first, and accepted only if a forward solve satisfies the gate: insertion-loss change below 0.10 dB for wavelength-division multiplexers (WDMs) and 0.15 dB for bends, with WDM leakage constrained.

III. Simulation Study

We evaluated ten frequency-domain cases at 80-nm grid spacing: six WDMs and four 90-degree bends. Labels encode function and size: Wn-L is a 1×n WDM with an L-μm design region, and B90-L is a 90-degree bend with an L-μm design region. This tests whether one trimming rule can serve as a device-agnostic cleanup step.

Figs. 2 and 4 show a representative W2-5 WDM. AIG and OPA remove visible peripheral silicon chunks while preserving the main routing geometry. The cleaned S21/S31 spectra remain close to the original response across the evaluated wavelength range, showing that the removed regions have limited optical impact. Across all ten cases, AIG removed a median 8.29% of design-region silicon with a median insertion-loss penalty of 0.0956 dB, while OPA removed a median 8.56% with a median penalty of 0.0945 dB using zero extra post-optimization adjoint evaluations. Fig. 3 highlights the broader impact: removing roughly 8.5% of the original design-region silicon reduces patterned silicon area in the cleaned layout, while the forward EM propagation modeling prevents visually plausible but optically harmful edits.

IV. Conclusion

AIG and OPA provide a solver-native workflow for trimming inverse-designed photonic layouts. By ranking connected

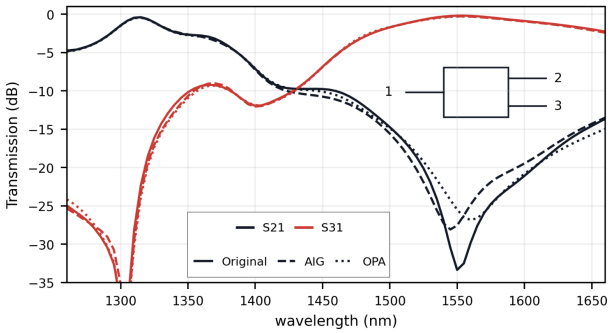

Fig. 4. Representative 1×2 WDM spectrum. AIG and OPA mostly preserve the O/C-band routing response after trimming.

chunks and forward-verifying each edit, the method reduces pattern area and layout complexity while preserving optical performance. This supports foundry-facing mask review, targeted inspection, and manufacturability-aware handoff. OPA is especially practical because it requires zero additional post-optimization adjoint solves. Future work includes 3D modeling, process-bias analysis, and fabrication experiments; we thank the Stanford NQP group for releasing SPINS-B.

References

[1] A. Y. Piggott *et al.*, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nat. Photonics*, vol. 9, pp. 374–377, 2015.

[2] S. Molesky *et al.*, "Inverse design in nanophotonics," *Nat. Photonics*, vol. 12, pp. 659–670, 2018.

[3] L. Su *et al.*, "Nanophotonic inverse design with SPINS: Software architecture and practical considerations," *Appl. Phys. Rev.*, vol. 7, Art. no. 011407, 2020.

[4] C. M. Lalau-Keraly *et al.*, "Adjoint shape optimization applied to electromagnetic design," *Opt. Express*, vol. 21, pp. 21693–21701, 2013.

[5] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. ICML*, pp. 3319–3328, 2017.

[6] J. Park, T. Kim, M. Ali, and D. Liang, "Interpretable geometry sensitivity for inverse design of integrated photonics," *arXiv:2510.22176*, 2026.