A Generative Prior-Guided Faster Inverse Design Framework for High-Quality Integrated Photonics

Junho Park[1,2,*], Taehan Kim[3,*], Mohammad Ali[1], SooHyuk Cho[4], 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
4. Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, USA
These authors contributed equally; † liangdi@umich.edu

***Abstract*—We experimentally show that generative priors distilled from inverse-design archives break the performance–manufacturability–modeling-speed tradeoff in C-band silicon arbitrary power splitters. In fabricated ultra-compact devices ($2\,\mu\mathrm{m}\times 2\,\mu\mathrm{m}$), they reduce mean IL from 3.40 to 1.20 dB, improve manufacturability 4×, and halve optimization steps.**

***Index Terms*—integrated photonics, inverse design, generative prior, arbitrary power splitter**

I. Introduction

Adjoint-based inverse design enables ultra-compact integrated photonic components, but its outcome remains highly sensitive to the starting layout [1]–[3]. Uninformed seeds force the optimizer to simultaneously discover a useful topology and tune the optical response, often producing restart-heavy workflows, fragmented geometries, and unnecessary full-wave simulation cost.

Years of inverse-design runs already encode past optimization effort that is usually discarded once a design fails to meet a particular performance or manufacturability target. However, these archived design datasets contain reusable interference scaffolds and condition-dependent structure that can be harvested. Here, as shown in Fig. 1, we distill this information into a learned generative prior: a conditional variational autoencoder (cVAE) [4] generates binary device geometries conditioned on wavelength and a device figure-of-merit, i.e., target split ratio for a power splitter here. A fast 2.5D FDTD screen ranks the generated candidates, then the best seed is refined under a fixed adjoint optimization budget to achieve optimal designs with multi-dimensional benefits.

We experimentally validate this generative prior-guided workflow by fabricating C-band silicon arbitrary power splitters. Random, midpoint (0.5 density), and cVAE-selected initializations are compared under the same 60-step refinement budget and the same objective formulation. **Across the fabricated C-band device set, cVAE seeding simultaneously improves measured IL, splitting ratios, manufacturability, and convergence speed, demonstrating that archived inverse-design knowledge can be reused to break the conventional performance–manufacturability–modeling-speed tradeoff rather than merely navigate it.**

II. Design and Measurement Method

Each splitter is represented by a binary silicon/cladding mask $\rho \in \{0,1\}^{H\times W}$ and condition $c = (\lambda, r_{\mathrm{top}})$, where r_{top} is the target top-port power fraction. The cVAE uses encoder $q_\phi(z|\rho, c) = \mathcal{N}(\mu_\phi, \mathrm{diag}(\sigma_\phi^2))$, latent sample $z \sim \mathcal{N}(0, I)$, and decoder $g_\theta(z, c)$. The training archive contains 6362 SPINS-B-generated arbitrary-ratio splitter layouts [5]; the encoder and decoder use two 512-unit fully connected layers and a 64-dimensional latent space. For each requested 1550-nm split ratio, 50 cVAE samples are generated, ranked using a reduced-cost 2.5D FDTD screen, and the top seed initializes FDTD adjoint refinement. All three initialization conditions use the same refinement budget. For each target ratio, we fabricated and measured one device from each initialization strategy under the same C-band characterization setup.

The reference power, P_{GC}, is taken as the average of ten C-band grating-coupler reference traces with the same

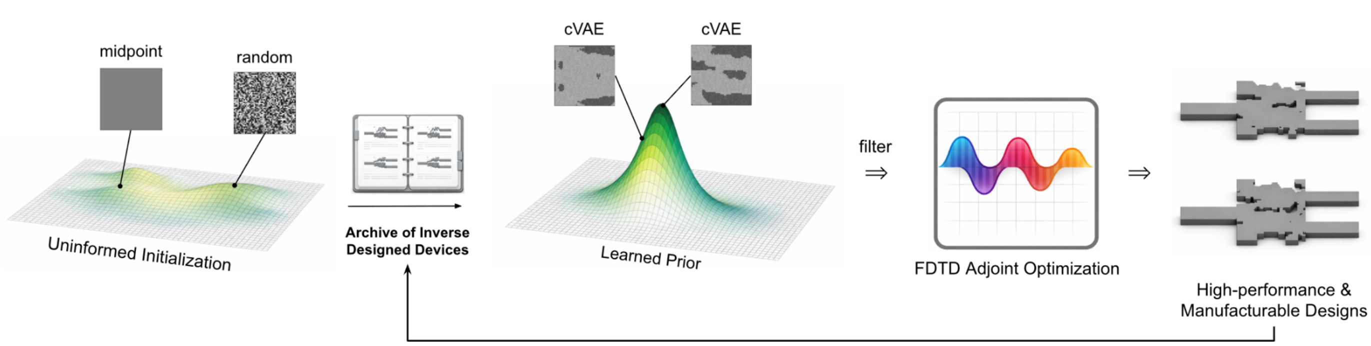

Fig. 1. Generative prior-guided inverse-design workflow. Archived splitter layouts train a condition-dependent cVAE prior; sampled seeds are FDTD-screened and refined with the same fixed adjoint budget as random and midpoint baselines.

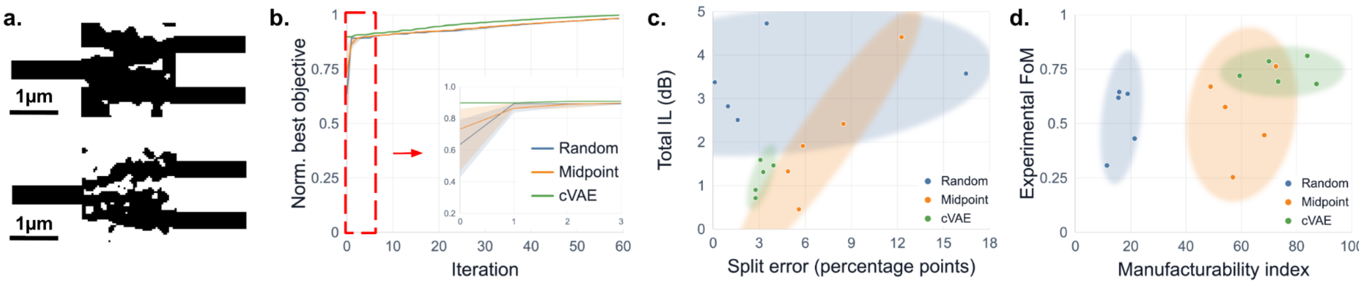

Fig. 2. (a) Random and cVAE-optimized 20:80 layouts show the improved regularity from prior seeding. (b) cVAE reaches the high-objective region within the first few iterations under the same 60-step budget. (c,d) Experimental endpoints show cVAE shifting devices toward lower loss, lower split error, higher manufacturability, and higher FOM.

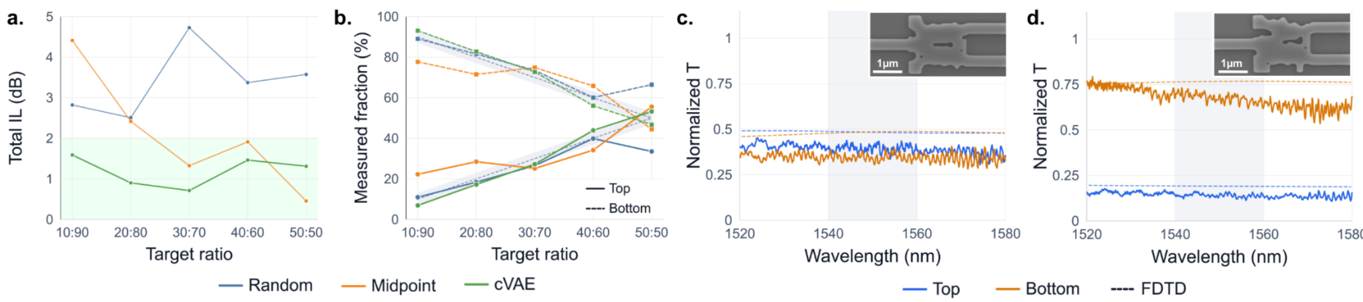

Fig. 3. Fabricated C-band splitter performance. (a) Total IL for five target top:bottom ratios, averaged over 1540–1560 nm and normalized by the ten-trace GC reference. (b) Measured top and bottom fractions from window-averaged powers. (c,d) Representative cVAE 20:80 and 50:50 spectra compare measured normalized top/bottom transmission with FDTD predictions across 1520–1580 nm; shaded regions mark the metric window.

waveguide routing but no splitter device. Total transmission is computed from the window-averaged output powers as $T_{\text{exp}} = (\langle P_{\text{top}}\rangle + \langle P_{\text{bot}}\rangle)/\langle P_{\text{GC}}\rangle$, and IL is reported in dB. The measured top-port split is $S_{\text{top}} = \langle P_{\text{top}}\rangle/(\langle P_{\text{top}}\rangle + \langle P_{\text{bot}}\rangle)$, with split error defined as $\epsilon_{\text{split}} = 100|S_{\text{top}} - r_{\text{top}}|$ in percentage points. We also report a figure of merit, $F_{\text{exp}} = 0.7T_{\text{exp}} + 0.3\max(0, 1 - \epsilon_{\text{split}}/10)$, and the manufacturability index is defined as $\sqrt{A}/P$, where P and A are the total perimeter and silicon area, quantifying geometric regularity.

III. Experimental Results

Fig. 2 shows how the learned prior changes both the optimization path and the fabricated-device tradeoff. For the representative 20:80 target in Fig. 2(a), random initialization produces a fragmented scattering region, whereas cVAE seeding forms a more connected interference scaffold, which is reflected in manufacturability score: 16.6 for random, 60.2 for midpoint, and 74.8 for cVAE. Fig. 2(b) further shows that cVAE starts from a higher normalized objective and reaches the high-performance region within the first few iterations, amortizing topology discovery under the same 60-step budget.

Fig. 2(c,d) show that this acceleration does not come at the cost of experimental performance. In the IL–split-error plane, cVAE devices cluster closest to the desired lower-left region, while in the manufacturability–FOM plane they occupy the high-regularity, high-FOM region most consistently. Thus, the learned prior shifts the fabricated population toward a better combined optical and geometric tradeoff.

Fig. 3(a,b) gives the ratio-by-ratio C-band measurements. Across the ratios, cVAE gives the lowest mean total IL, 1.20 dB, compared with 2.10 dB for midpoint and 3.40 dB for random. It also gives the best mean split tracking, with mean absolute top-port split errors of 3.16, 7.41, and 4.52 percentage points for cVAE, midpoint, and random, respectively. Combining transmission and split fidelity, cVAE achieves the highest experimental figure of merit, $F_{\text{exp}} = 0.738$.

The representative spectra in Fig. 3(c,d) show smooth broadband splitter responses rather than isolated single-wavelength performance. Over the 1540–1560 nm metric window, the measured top and bottom transmissions preserve the intended power ordering and follow the FDTD-predicted partition. The cVAE 20:80 device measures a 17.24:82.76 split with 0.90 dB IL, while the cVAE 50:50 device measures a 53.26:46.74 top:bottom split with 1.31 dB IL, confirming robust fabricated C-band splitting.

IV. Conclusion

We propose and experimentally validate generative prior-guided inverse design for breaking the coupled performance–modeling-speed–manufacturability tradeoff in integrated photonics. This generic AI-enabled methodological innovation can have a profound impact on a wide range of PIC designs.

References

[1] J. S. Jensen and O. Sigmund, *Laser Photon. Rev.*, vol. 5, p. 308, 2011.
[2] C. M. Lalau-Keraly *et al.*, *Opt. Express*, vol. 21, p. 21693, 2013.
[3] S. Molesky *et al.*, *Nat. Photon.*, vol. 12, p. 659, 2018.
[4] K. Sohn, H. Lee, and X. Yan, *NeurIPS*, vol. 28, p. 3483, 2015.
[5] L. Su *et al.*, arXiv:1910.04829, 2019.