A1 Optical field simulation

Analyzing how light field propagate through those components are critical to device optimization and photonic integrated circuit design. Given a linear isotropic optical component, we will shine time-harmonic continuous-wave light on its input ports and analyze the steady-state electromagnetic field distributions $\mathbf{E} = \hat{\mathbf{x}}\mathbf{E}_x + \hat{\mathbf{y}}\mathbf{E}_y + \hat{\mathbf{z}}\mathbf{E}_z$ and $\mathbf{H} = \hat{\mathbf{x}}\mathbf{H}_x + \hat{\mathbf{y}}\mathbf{H}_y + \hat{\mathbf{z}}\mathbf{H}_z$ in it, each of which includes horizontal (x), vertical (y), and longitudinal (z) components. The light field follows the Maxwell PDE under certain absorptive boundary conditions [15],

$$\nabla \times \mathbf{E}(\boldsymbol{r}, t) = \mu_0 \frac{\partial \mathbf{H}(\boldsymbol{r}, t)}{\partial t} + \mathbf{J}_e(\boldsymbol{r}, t),\ \nabla \times \mathbf{H}(\boldsymbol{r}, t) = -\epsilon_0 \boldsymbol{\epsilon}_r(\boldsymbol{r}) \frac{\partial \mathbf{E}(\boldsymbol{r}, t)}{\partial t} + \mathbf{J}_e(\boldsymbol{r}, t), \tag{7}$$

where $\nabla\times$ is the curl operator of a vector function, μ_0 is the vacuum magnetic permeability, ϵ_0 and $\boldsymbol{\epsilon}_r$ are the vacuum and relative electric permittivity, $\mathbf{J}_m$ and $\mathbf{J}_e$ are the magnetic and electric current sources. Since the input light is time-harmonic at a vacuum angular frequency ω, the time-domain PDE can be transformed to the frequency domain for the steady state as follows,

$$\nabla \times \mathbf{E}(\boldsymbol{r}) = j\omega\mu_0 \mathbf{H}(\boldsymbol{r}) + \mathbf{J}_m(\boldsymbol{r}),\ \nabla \times \mathbf{H}(\boldsymbol{r}) = -j\omega\epsilon_0\boldsymbol{\epsilon}_r(\boldsymbol{r})\mathbf{E}(\boldsymbol{r}) + \mathbf{J}_e(\boldsymbol{r}). \tag{8}$$

A simple variable substitution gives us the *curl-of-curl* Maxwell PDE,

$$\big((\mu_0^{-1}\nabla\times\nabla\times) - \omega^2\epsilon_0\boldsymbol{\epsilon}_r(\boldsymbol{r})\big)\mathbf{E}(\boldsymbol{r}) = j\omega\mathbf{J}_e(\boldsymbol{r}),\ \big(\nabla\times(\boldsymbol{\epsilon}_r^{-1}(\boldsymbol{r})\nabla\times) - \omega^2\mu_0\epsilon_0\big)\mathbf{H}(\boldsymbol{r}) = j\omega\mathbf{J}_m(\boldsymbol{r}). \tag{9}$$

To restrict a unique solution without boundary reflection, complicated boundary conditions will be inserted [15]. An artificial material, i.e., coordinate-stretched perfectly matched layer (SC-PML), will be padded around the solving domain. Such PML materials have large imaginary parts in the permittivities to introduce strong energy absorption and changes the derivative operator to $\nabla = (\frac{1}{s_x(x)}\frac{\partial}{\partial x}, \frac{1}{s_y(y)}\frac{\partial}{\partial y}, \frac{1}{s_z(z)}\frac{\partial}{\partial z})$, where s is a location-determined complex value. Solving the above PDEs will give the steady-state frequency-domain complex magnitude of the optical fields.

A2 Dataset generation

We generate our customized MMI device simulation dataset using an open-source FDFD simulator angler [15]. The tunable MMI dataset has 5.5 K *single-source* training data, 614 validation data, and 1.5 K multi-source test data. The etched MMI dataset has 12.4 K *single-source* training data, 1.4 K validation data, and 1.5 K *multi-source* test data. We summarize how we generate random devices in Table A4. We randomly sample the physical dimension of the MMI, input/output waveguide width, the width of the perfectly matched layer (PML), device border width away from PML, controlling pad sizes, input light source frequencies, etched cavity sizes and ratio (determines the number of cavities in the MMIs), and permittivities in the controlling region.

Table A4: Summary of device design variable's sampling range, distribution, and unit.

Variables	Value/Distribution		Unit
	$\lvert\mathbf{J}\rvert \times \lvert\mathbf{J}\rvert$ Tunable MMI	$\lvert\mathbf{J}\rvert \times \lvert\mathbf{J}\rvert$ Etched MMI	
Length	$\mathcal{U}$(20, 30)	$\mathcal{U}$(20, 30)	μm
Width	$\mathcal{U}$(5.5, 7)	$\mathcal{U}$(5.5, 7)	μm
Port Length	3	3	μm
Port Width	$\mathcal{U}$(0.8, 1.1)	$\mathcal{U}$(0.8, 1.1)	μm
Border Width	0.25	0.25	μm
PML Width	1.5	1.5	μm
Pad Length	$\mathcal{U}$(0.7, 0.9)×Length	$\mathcal{U}$(0.7, 0.9)×Length	μm
Pad Width	$\mathcal{U}$(0.4, 0.65)×Width/$\lvert\mathbf{J}\rvert$	$\mathcal{U}$(0.4, 0.65)×Width/$\lvert\mathbf{J}\rvert$	μm
Wavelengths λ	$\mathcal{U}$(1.53, 1.565)	$\mathcal{U}$(1.53, 1.565)	μm
Cavity Ratio	-	$\mathcal{U}$(0.05, 0.1)	-
Cavity Size	-	0.027 Length × 0.114 Width	μm^2
Relative Permittivity $\boldsymbol{\epsilon}_r$	$\mathcal{U}$(11.9, 12.3)	{2.07, 12.11}	-

A3 Training settings

We implement all models and training logic in PyTorch 1.10.2. All experiments are conducted on a machine with Intel Core i7-9700 CPUs and an NVIDIA Quadro RTX 6000 GPU. For training from scratch, we set the number of epochs to 200 with an initial learning rate of 0.002, cosine learning rate decay, and a mini-batch size of 12. For the tunable MMI dataset, we split all 7,680 examples into 72% training data, 8% validation data, and 20% test data. For the etched MMI dataset, we split all 15,360 examples into 81% training data, 9% validation data, and 10% test data. For device adaptation, we first perform linear probing for 20 epochs with an initial learning rate of 0.002 and cosine learning rate decay; then we perform finetuning for 30 epochs with an initial learning rate of 0.0002 and a cosine learning rate decay. We apply stochastic network depth with a linear scaling strategy and a maximum drop rate of 0.1.

A4 Model architectures

UNet. We construct a 4-level convolutional UNet with a base channel number of 34. The total parameter count is 3.47 M.

FNO-2d. For Fourier neural operator (FNO), we use 5 2-D FNO layers with a channel number of 32. The Fourier modes are set to (#Mode$_z$=32, #Mode$_x$=10). The final projection head is CONV1$\times$1(256)-GELU-CONV1$\times$1(2). The total parameter count is 3.29 M.

F-FNO. For factorized Fourier neural operator (F-FNO), we use 12 F-FNO layers with a channel number of 48. The Fourier modes are set to (#Mode$_z$=70, #Mode$_x$=40). The final projection head is CONV1$\times$1(256)-GELU-CONV1$\times$1(2). The total parameter count is 3.16 M.

`NeurOLight`. For our proposed `NeurOLight`, we use 12 F-FNO layers for tunable MMIs and 16 layers for etched MMIs with a base channel number C=64. The convolution stem is BSConv3$\times$3(32)-BN-ReLU-BSConv3$\times$3(64)-BN-ReLU, where BSConv is blueprint convolution [12]. The Fourier modes are set to (#Mode$_z$=70, #Mode$_x$=40). The channel expansion ratio in the FFN is set to s=2. The final projection head is CONV1$\times$1(256)-GELU-CONV1$\times$1(2). The total parameter count is 1.58 M.

A5 Animation of `NeurOLight`

We animate the prediction process of `NeurOLight` on a 3$\times$3 tunable MMI. The prediction throughput reaches 120 frames per second (FPS) which allows designers to freely tune input frequencies, device sizes, permittivities, and input light sources with *real-time* simulation result feedback.