Efficient Bilevel Source Mask Optimization

Guojin Chen[1], Hongquan He[2], Peng Xu[1], Hao Geng[2], Bei Yu[1]

[1]The Chinese University of Hong Kong, [2]ShanghaiTech University

Introduction

- RETs critical for advanced technology nodes, with SMO pivotal for optimizing source and mask together to expand the process window.
- Traditional SMO methods are limited by sequential optimizations, leading to long runtimes and no performance guarantees.
- The paper introduces a unified SMO framework using accelerated Abbe forward imaging, enhancing precision and efficiency.
- The innovative **BiSMO** framework, using bilevel optimization and three gradient-based methods, achieves 40% error reduction and 8x runtime efficiency increase.

Contribution

- First unified Abbe-based SMO framework with process window considerations, parallel computation accelerates Abbe imaging to Hopkins' method speeds.
- Modeled SMO as a unified bilevel framework, developed three efficient gradient-based methods for better solution space exploration.
- Experimental results: 40% error reduction and 8x throughput increase compared to SOTA SMO methods. Error metrics are half compared to SOTA MO methods.

Bilevel SMO

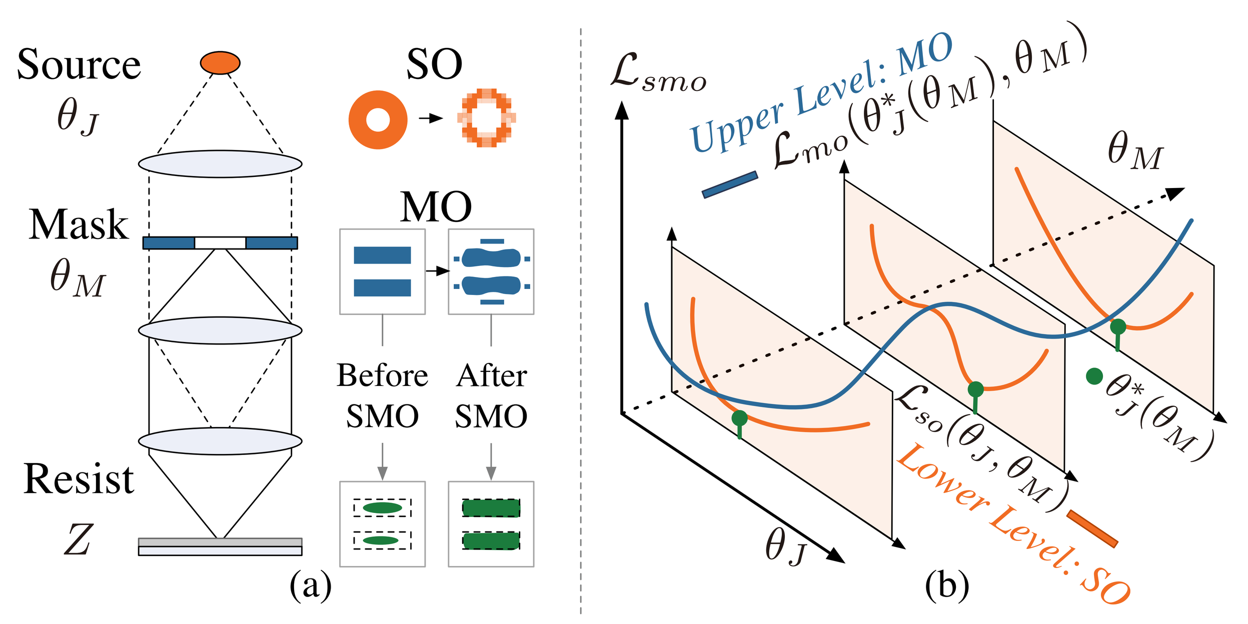

Figure 1. (a) The forward lithography and SMO process. (b) Bilevel SMO with upper-level MO and lower-level SO.

BiSMO vs. Traditional SMO Flow

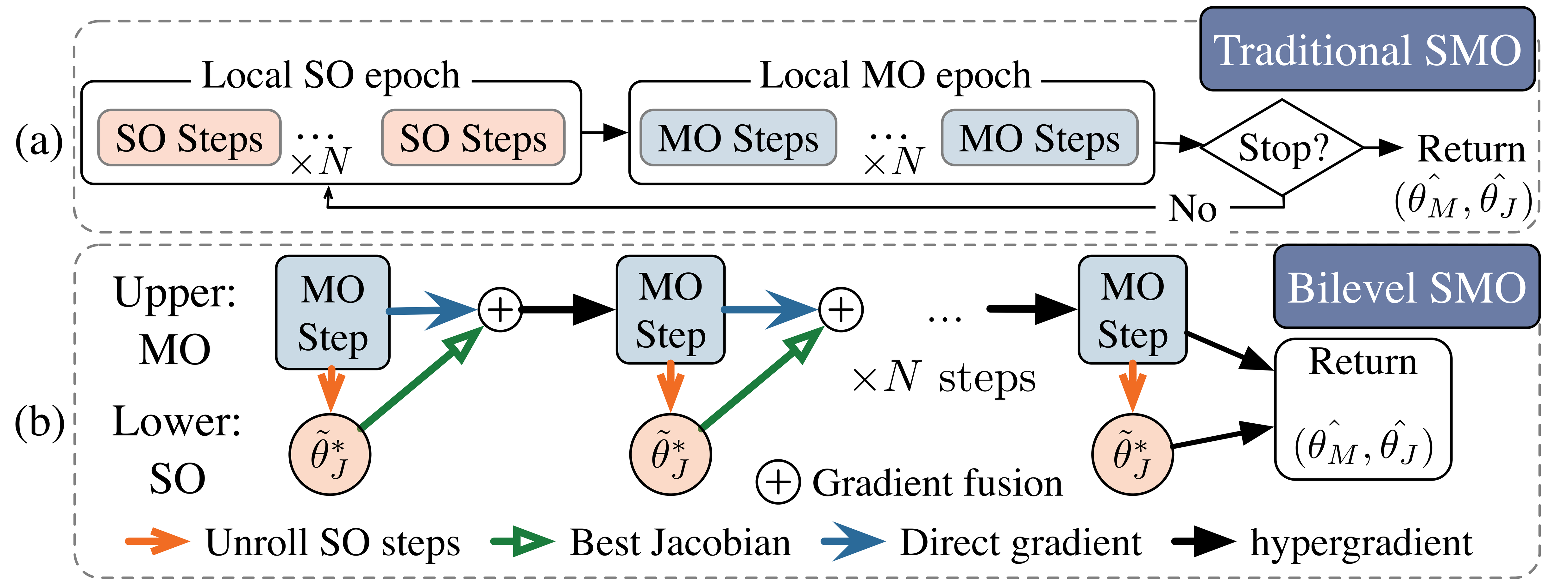

Figure 2. (a)Previous AM-SMO flow. (b)Our BiSMO flow.

Reformulate into bilevel format.

$$(\hat{\boldsymbol{\theta}}_J, \hat{\boldsymbol{\theta}_M}) = \underset{(\boldsymbol{\theta}_J, \boldsymbol{\theta}_M)}{\operatorname{argmin}}\ \mathcal{L}_{smo}(\boldsymbol{\theta}_J, \boldsymbol{\theta}_M), \tag{1}$$

$$\Downarrow$$

$$\min_{\boldsymbol{\theta}_M} \mathcal{L}_{mo}(\boldsymbol{\theta}_J^*(\boldsymbol{\theta}_M), \boldsymbol{\theta}_M), \qquad \triangleright \textit{Upper-Level: MO}$$

$$\text{s.t. } \boldsymbol{\theta}_J^*(\boldsymbol{\theta}_M) = \operatorname{argmin}_{\boldsymbol{\theta}_J} \mathcal{L}_{so}(\boldsymbol{\theta}_J, \boldsymbol{\theta}_M). \qquad \triangleright \textit{Lower-Level: SO}$$

Solve the bilevel SMO : Hypergradient

$$\nabla_{\boldsymbol{\theta}_M} \mathcal{L}_{mo} = \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_M} + \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_J} \frac{\partial \boldsymbol{\theta}_J^*(\boldsymbol{\theta}_M)}{\partial \boldsymbol{\theta}_M}. \tag{2}$$

- `BiSMO-FD` : bilevel SMO with finite difference.
- `BiSMO-NMN`: bilevel SMO with Neumann series.
- `BiSMO-CG`: bilevel SMO with conjugate gradient.

$$\texttt{BiSMO-FD}: \nabla_{\boldsymbol{\theta}_M} \mathcal{L}_{mo}^{\text{FD}} = \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_M} - \xi \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_J} \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_M \partial \boldsymbol{\theta}_J}, \tag{3}$$

Implicit Function Theorem (IFT)

$$\nabla_{\boldsymbol{\theta}_M} \mathcal{L}_{mo} = \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_M} - \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_J} \left[\frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_J \partial \boldsymbol{\theta}_J}\right]^{-1} \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_M \partial \boldsymbol{\theta}_J}. \tag{4}$$

Neumann series:

With a matrix $\boldsymbol{A}$ that $\|\boldsymbol{\mathcal{I}} - \boldsymbol{A}\| < 1$, we have:

$$\boldsymbol{A}^{-1} = \sum_{k=0}^{\infty} (\boldsymbol{\mathcal{I}} - \boldsymbol{A})^k. \tag{5}$$

Truncate the Neumann series to K terms, the `BiSMO-NMN` is given by,

$$\nabla_{\boldsymbol{\theta}_M} \tilde{\mathcal{L}}_{mo}^{\text{NMN}} = \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_M} - \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_J} \sum_{k=0}^{K} \left[\boldsymbol{\mathcal{I}} - \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_J \partial \boldsymbol{\theta}_J}\right]^k \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_M \partial \boldsymbol{\theta}_J}. \tag{6}$$

Conjugate Gradient Bilevel SMO:

The vector $\vec{w}$ can be obtained by solving the optimization problem:

$$\min_{\vec{w}} \vec{w}^\top \left[\frac{\partial^2 \mathcal{L}_{so}}{\partial \vec{\theta}_J \partial \vec{\theta}_J}\right] \vec{w} - \vec{w}^\top \frac{\partial \mathcal{L}_{mo}}{\partial \vec{\theta}_J}. \tag{7}$$

The conjugate gradient (CG) algorithm is well-suited for this task. `BiSMO-CG` is computed as:

$$\nabla_{\boldsymbol{\theta}_M} \tilde{\mathcal{L}}_{mo}^{\text{CG}} = \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_M} - \left[\underset{\boldsymbol{w}}{\operatorname{argmin}} \left(\boldsymbol{w}^\top \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_J \partial \boldsymbol{\theta}_J} \boldsymbol{w} - \boldsymbol{w}^\top \frac{\partial \mathcal{L}_{mo}}{\partial \boldsymbol{\theta}_J}\right)\right] \frac{\partial^2 \mathcal{L}_{so}}{\partial \boldsymbol{\theta}_M \partial \boldsymbol{\theta}_J}. \tag{8}$$

Results and Analysis

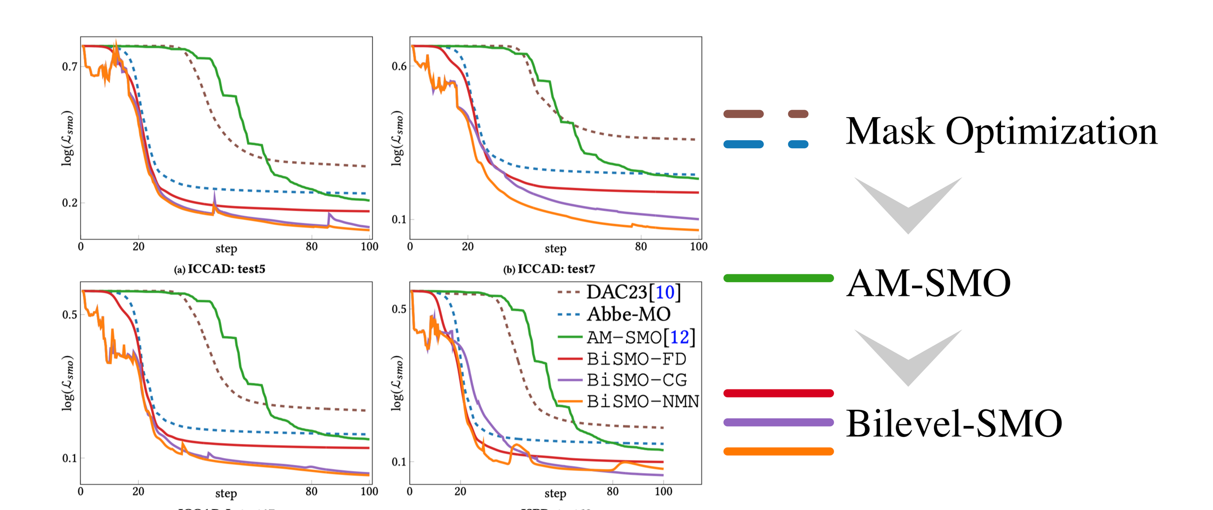

Figure 3: Loss comparison between different MO methods (dashed lines) and SMO methods (solid lines).

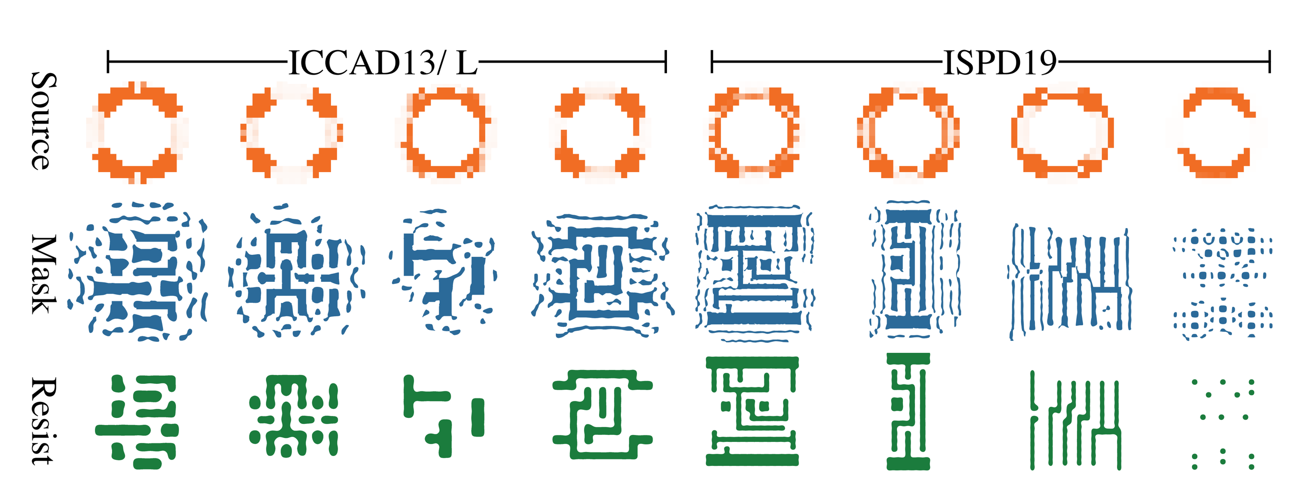

Figure 3. Result samples from ICCAD13 and ISPD19 datasets.

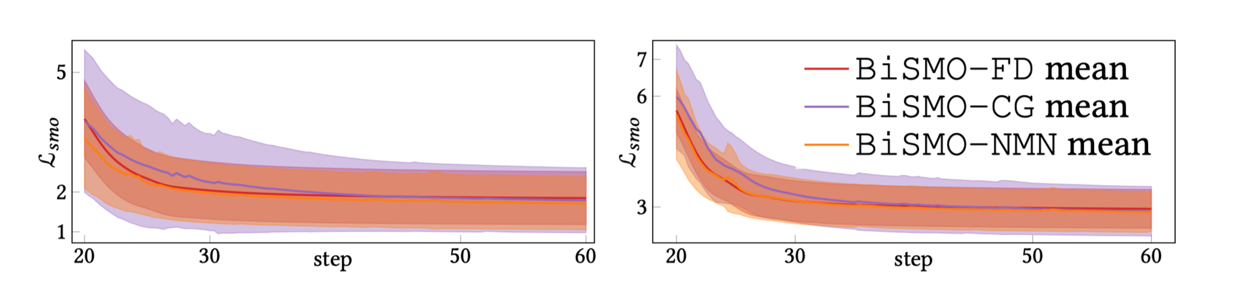

Figure 5: Mean and STD of (a) ICCAD (b) ICCAD-L datasets.