利用选择性状态空间模型增强可制造多层惠更斯超表面的半解析设计：第二部分——生成式逆向设计（MetaMamba）

NatanelNissan, *IEEE* 学生会员，Sherman W. Marcus, *IEEE* 终身会员，Dan Raviv, *IEEE* 会员，Raja Giryes, *IEEE* 高级会员，以及 Ariel Epstein, *IEEE* 高级会员

arXiv:2603.03877v1 [physics.app-ph] 4 Mar 2026

摘要——**本文提出一种用于五层透射式惠更斯超表面（HMSs）逆向设计的生成框架，旨在解决长期以来的难题：如何以尽可能少的全波仿真，实现全相位、高效率的单元胞设计。实现这一目标的关键在于采用本文第一部分提出的基于场的半解析（SA）方案，该方案能够快速且高效地合成此类多层复合结构，但精度有限。为克服传统流程对数据的巨大需求，我们采用适于长程序列建模的选择性状态空间模型 Mamba 作为学习框架的骨干网络。首先，使用半解析生成数据训练双向 Mamba（Bi-Mamba）正向代理模型，随后利用全波 CST 样本对其进行微调。对包含 1080 个样本的 CST 样本池进行消融实验后发现，仅需 270 个全波校准样本，即可在仿真成本仅为其一小部分的情况下达到近乎 CST 级的一致性。随后，使用代理模型增强数据训练自回归 Mamba 逆向生成器，将单元胞合成视为序列生成任务。由此得到的一对多生成模型可根据目标散射响应生成多样的单元胞几何结构。在 20 GHz 下，该模型能够生成经 CST 验证的设计，其场透射幅度达到 0.9，覆盖完整的 0–2π 相位范围。此外，经过训练、能够准确预测频率响应（18–22 GHz）的经 CST 校准的代理模型，可对逆向生成的设计进行功能后选择。总之，这项两部分汇编所提出的半解析—生成式混合方法，为多层惠更斯超表面合成提供了一种可扩展且数据高效的解决方案，并可自然扩展至宽带、斜入射及更高维电磁逆向设计问题。**

索引词—**惠更斯超表面、超表面综合、代理模型、逆向设计、深度学习、生成模型、状态空间模型、Mamba、序列建模、机器学习。**

I. 引言

电磁超表面——由亚波长可极化粒子（超原子）构成的低剖面阵列——提供了紧凑、平面化的手段来操控波前，从而实现对振幅、相位和偏振的灵活控制 [1], [2]。在这一类器件中，惠更斯超表面（全息超表面）尤其具有吸引力，因为其共置的电磁极化率能够以极小的反射实现范围广泛的功率守恒场变换 [3]–[7]。在宏观层面，惠更斯超表面的变换可通过广义片层跃迁条件（GSTC）进行严格描述，该条件将电磁场的不连续性与表面极化率张量相关联 [8]–[11]。如第一部分 [12], 所述，对于常见的波束成形应用（例如异常折射板或聚焦透镜），基于 GSTC 的设计通常会指定局部响应，使反射消除型超原子在法向入射平面波照射下产生预设的相移。然而，此类透射型惠更斯超表面单元胞的逆向设计（微观设计 [6]）仍然极具挑战性：如何在现实的亚波长平台上实现接近 1 的透射率，同时覆盖完整的 2π相位范围。尽管 GSTC 规定了所需的表面响应，但将其转化为可制造的几何结构仍然困难，因为目前缺少一种系统、可靠且省时的方法来生成满足严格散射约束的现实结构。

在微波频率下，实现这种所需共位电响应与磁响应的最自然方式，是采用定义于印刷电路板（PCB）上的导电环与导线组合 [3], [13]。然而，后者必须人工切割并重新排布，从而形成复杂的多板结构，可能不适用于许多应用。另一种方案表明，采用薄电极化片层（阻抗片层）的对称级联来实现，可提供足够的自由度，以支持对称模和反对称模的激励，从而有效模拟惠更斯超原子所需的电、磁表面极化率 [4], [14]。该结构与简单且标准的无过孔多层印刷电路板结构完全兼容，构成极具吸引力的实现平台 [6], [15]。尽管如此，高效率工作通常需要多个堆叠层 [16], [17],，而随着层数增加，暴力优化将变得难以处理。考虑均匀化阻抗片层之间多次反射的基于传输线的模型 [4], [7], [14], [16], [17] 可即时给出 S参数预

本研究由以色列创新局通过其超材料联盟提供支持。

N. Nissan、D. Raviv 和 R. Giryes 任职于特拉维夫大学电气与计算机工程学院，地址为以色列特拉维夫 69978（电子邮箱：natanel.nissan@gmail.com、raviv.dan@gmail.com、raja@tauex.tau.ac.il）。

S. W. Marcus 和 A. Epstein 任职于 Technion – 以色列理工学院 Andrew and Erna Viterbi 电气与计算机工程学院，地址为以色列海法 3200003（电子邮箱：shermanm@technion.ac.il、epsteina@ee.technion.ac.il）。

这两部分汇编第一部分和第二部分中开发并使用的代码已公开发布于 https://github.com/nati223/metamamba。

测，并可能揭示重要的物理相互作用，但由于忽略了层间近场耦合效应而存在显著误差；在常见超表面（MS）中层间距较小，这种效应更为明显。因此，在几乎所有情况下，都需要借助全波优化器进行最终微调，才能获得可用设计 [16],[18], [19]。试图通过在扩展电路模型中引入专用元件来纳入互耦，在这一意义上取得了一定成功，但需要迭代调整其值（或事先了解耦合特性），从而降低了此类方法的通用性与吸引力 [20], [21]。

在这篇论文的第一部分中，我们提出并系统地构建了一种基于场的半解析（SA）方法，该方法能够严格考虑多层加载导线型超原子构型中的这些复杂近场效应（体现为倏逝波谱） [9], [22], ，从而支持设计多种高保真、可制造的超表面 [23]–[27]。然而，尽管我们成功设计出实用且功耗高效的双极化、频率多样化惠更斯超原子印制电路板布局，可按需生成，模型中的某些假设仍会造成轻微误差。当自由度较多（即印制电路板堆叠结构中的层数较多）时，这些误差会逐渐累积，在某些情况下可能导致超原子的整体响应劣化。具体而言，虽然该模型恰当地考虑了各加载导线上感应电流的主导偶极矩所产生的近场耦合，但未考虑高阶多极矩，这可能导致散射参数预测出现偏差。因此，正如第一部分 [12],所述，基于SA方法得到一组合适的惠更斯超原子构型（印制电路板堆叠结构中每层的铜箔走线规格），并使其覆盖整个 2π 相位空间且具有高效率之后，还需借助高保真求解器（例如CST Microwave Studio）进行最终验证和筛选；仅保留仿真结果表现良好的解（视为“真实值”），并据此构建查找表（LUT），随后即可将其用于全息超表面的宏观设计 [6]。

尽管 SA 模型的相对高保真度使这一设计流程效率很高，只需对 SA 方案输出的适用候选几何结构进行数值仿真（与蛮力全波优化相比，数量仅为其一小部分），但验证 SA 预测结果的需求最终仍带来一些不足。首先，虽然查找表有效覆盖了整个2π 相位区间，但在 SA 模型往往不够准确的这些区域，相位分辨率可能有限（因为经过上述“筛选”流程后，仅保留了部分提出的解）。其次，尽管该方法能够快速生成适用于实际可制造设计的查找表，但对于所考虑的每种相移，难以评估所获得的超原子效率是否已达到所考虑的多层物理结构所能实现的最佳水平（因为 SA 预测的最优解可能与全波 “ 真实值 ” 并不完全一致）。

与基于物理的加速方法并行，机器学习（ML）已成为一种颇具前景的替代方案，可用于超表面设计中的暴力优化和计算成本高昂的全波仿真 [28]。早期工作聚焦于正向问题，采用代理模型将几何参数映射为电磁响应 [29], ，从而实现快速评估。为应对逆向设计这一挑战，后来引入了生成模型。早期研究采用变分自编码器（VAEs）和生成对抗网络（GANs），用于频率选择表面（FSSs）和多层多频带反射式极化超表面的单元胞设计 [30], [31]。随后，研究进一步转向更复杂的任务，包括透射式超表面单元胞的逆向合成 [32], 、全空间正交通道超表面图案 [33],，以及采用环形拓扑结构的宽角度双极化频率选择表面 [34]。与此同时，研究人员也探索了利用扩散模型生成高效率透射式超表面的单元胞，并取得了显著成果 [35], [36]。

尽管这些生成式方法在自动化超表面逆向设计方面展现出巨大潜力，但也暴露出一个关键瓶颈：庞大的数据需求。大多数研究需要数量级为 10^4–10^5 的高保真全波仿真，或相当数量的核心小时，才能完成 [30], [31], [33], [35]–[39]的训练。近期研究尝试减轻这一计算负担，例如采用先验知识引导的生成式深度学习方法进行多目标综合 [34]。然而，尽管利用物理与经验知识约束设计空间，并指导网络训练和数据采集，有效训练仍需要数千次高保真全波仿真 [34]。在光学领域，这一挑战通常可借助严格耦合波分析（RCWA）等快速近似求解器缓解，从而生成海量数据集 [40], [41]。然而，在微波多层惠更斯超表面设计中，通常依赖全波求解器（如 CST Microwave Studio 或 ANSYS HFSS），因此扩展到更深层堆叠的成本尤其高昂 [42]。

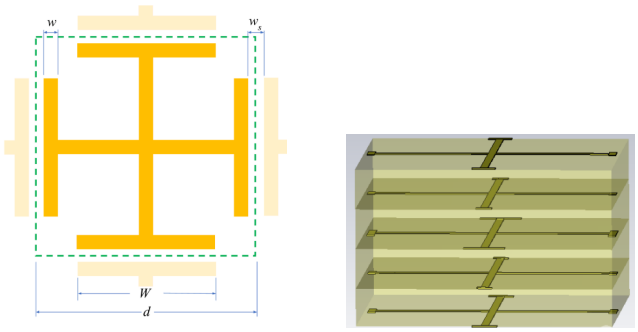
除了计算能力的限制，现有的透射式惠更斯超表面实现还面临固有的性能约束。在实用的多层惠更斯超原子中，同时实现接近 1 的场透射幅度（$|T|$）和完整的 $0-2\pi$ 相位（ϕ）覆盖，本身就是一项极具挑战性的目标 [36], [43], [44]。据我们所知，尚无无源印刷电路板兼容配置在整个相位范围内实现$|T|>0.9$ 。

在本第二部分中，我们介绍*MetaMamba*，一种用于多层透射式惠更斯超表面逆向设计的混合式SA与生成框架，旨在克服这些局限。其核心动机是结合两种方法的互补优势：SA模型可提供接近物理的精度，且运行时间比全波求解器快数个数量级，因而非常适合生成大规模合成数据集；生成模型则能够从数据中学习丰富的映射并实现近最优逆向设计——但训练往往需要大量数据集。通过在半解析生成数据上进行预训练，并利用少量高保真全波仿真进行微调，MetaMamba以最小的仿真成本实现了高精度，有效弥合了这一差距。为捕捉多层惠更斯超表面各层之间强烈的电磁耦合，我们将设计过程表述为序列建模任务 [45]。正向问题所涉及的

散射响应的预测本质上是双向的，因为所有层都会相互作用。相反，逆问题涉及自回归（AR）生成 [46]：每一层均根据目标响应以及先前已生成的层依次合成。这一表述自然引出状态空间模型（SSM）的使用——一类植根于线性动力系统与控制理论的高效序列模型 [47], [48], ，其中信息经由潜在隐藏状态传播，而这些状态按照学习得到的系统矩阵随时间演化。该表述类似于反馈系统或递归滤波器对物理动力学的建模方式，因而特别适用于涉及相互依赖的层或空间序列的问题。Mamba 架构作为最新的状态空间模型变体，扩展了这一概念，使模型能够在状态更新期间动态调节序列中哪些部分受到强调，从而在多个领域同时实现全局上下文传播与线性时间效率 [49]。第二代变体Mamba-2 [50],进一步稳定了训练、提升了并行性，并增强了长序列处理能力。它能够以较低开销在堆叠层之间传播信息，因此非常适合多层 HMS 设计；其中强烈的层间耦合要求对整个序列中的依赖关系进行建模。

MetaMamba 流程包含两个主要组成部分。首先，训练一个双向 Mamba（Bi-Mamba）正向代理模型，分两个阶段进行：先使用 524,000 个 SA 样本进行预训练，以捕捉广泛的散射趋势；随后利用 270 次高保真 CST 仿真进行微调，以尽可能缩短训练和仿真时间，同时达到 CST 级精度。其次，AR-Mamba 逆向生成器以期望的散射响应为条件，逐层采样几何结构，其方式仿照自回归语言模型的结构。这使得一对多生成成为可能，可产生多样且有效的多层 HMS 设计。此外，MetaMamba 可自然扩展至宽带单元胞设计；借助校准代理模型，可高效预测不同频率范围内的性能，并在后续筛选出功能性设计。这种混合策略生成了经 CST 验证的五层 HMS 单元胞，其透射率接近 1，且具有完整的相位覆盖范围；与纯全波仿真流程相比，所需数据量减少了数个数量级。除本文两部分中针对透射式 HMS 透镜所展示的具体应用外，所提出的框架还可自然推广至一大类能够表述为序列生成任务的电磁逆问题。

综上所述，第一部分和第二部分确立了一种可扩展且灵活的方法，可用于合成可制造的双极化、多频全息超表面设计。第一部分提出的计算高效 SA 方案与本研究引入的 MetaMamba 算法的生成精度相结合，形成了一类新型的物理与机器学习混合设计工具。这些工具兼具两方面的优势：物理可解释性、快速评估能力和接近 CST 级精度，同时仅需少量全波仿真进行校准。这种混合方法为多层 PCB 兼容超表面的按需逆向设计提供了实用途径，并具备深层 MS 堆叠结构和大型设计空间所需的可扩展性。



(a) (b)

图1. 单元几何结构。(a) 单层参数化。一个 JC 贴片位于一个周期 P内，板宽为 w，支腿长度 W可变。(b) 由垂直堆叠的 JC 铜图案构成的五层全息超表面单元，其中每层的 JC 支腿长度 W_n 以及堆叠结构的集体电磁相互作用共同决定最终的散射参数。

因此，它有望显著加快基于全息超表面的高性能波束成形与波前工程组件在多种时效性应用中的部署，包括5G/6G通信系统、卫星链路和下一代微波成像平台 [15], [52]–[55]。

II. 方法

A. 总体框架与设计依据

我们考虑一种兼容印制电路板（PCB）制造的 N层透射式惠更斯超表面，其设计工作频率为标称频率 f_0（图1）。与第一部分介绍的配置类似，每层均由周期排列的耶路撒冷十字（JC）铜图案构成，并以可调节的 JC 支腿长度 W （图1a）作为参数；这些图案制作在分隔相邻金属化层的介质层压板上（图1b）。如第一部分所述，SA LAYERS模型能够快速基于场预测此类堆叠结构的散射响应，同时严格考虑近场（倏逝波）耦合 [12]。

为了与本方法逆向设计阶段所需的离散生成建模相衔接（见第 III 节），每个层参数都被离散化为由有限词表 V 表示的一组可能值。具体而言，

$$W_n \in V = \{W_{\min}, W_{\min}+\Delta W, \ldots, W_{\max}\},\ n=1,\ldots,N \quad (1)$$

其中，$W_{\min}$、$W_{\max}$ 和 ΔW 定义了可用几何状态的范围和分辨率。这种量化将每个连续层尺寸转换为从词表 V 中选取的离散符号，从而形成由 AR 生成器处理的词元空间。$W_{\min}$、$W_{\max}$ 和 ΔW 的选择应与物理可实现性和建模分辨率相匹配，即 PCB 基板上可打印或可蚀刻的 JC 腿尺寸的最小值/最大值，以及在不引入冗余相邻状态的前提下，捕捉透射和相位有意义变化所需的粒度。请注意，这种离散化是 AR 逆向生成器所必需的，而正向代理模型则被训练为针对标量 W 值的连续回归模型。

一个关键观察是，本文提出并研究的特定基于 JC 的配置（图 1）中的多层全息超表面

可以自然地表示为序列：每个 W_n 表示有序堆栈中第 n 层的设计参数，而输出 $\mathbf{S}$则概括了它们的联合散射响应。从这一视角出发，令 $\mathbf{W}_{1:N} = (W_1, \ldots, W_N)$ 表示设计序列，$\mathbf{S}$ 表示散射S参数序列。电磁正向问题为 $\mathbf{S} = f_{\mathbf{S}}(\mathbf{W}_{1:N})$，即预测给定几何结构对应的散射响应。[1]

更一般地，上述表述并不将散射响应$\mathbf{S}$限制于单一工作条件。相反，$\mathbf{S}$ 可被视为沿频率、入射角或偏振状态等辅助维度索引的结构化序列。在这种情况下，正向代理模型的任务是学习条件映射 $f_{\mathbf{S}}(\mathbf{W}_{1:N}, \mathbf{c})$，其中$\mathbf{c}$ 表示一组用于指定所需工作状态的条件变量。这种抽象使同一序列模型能够扩展到更丰富的电磁目标，而无需改变其核心架构；这一属性在第III-D节中被用于对宽带频率响应进行建模，其中$\mathbf{c}$对应于覆盖工作频带的离散频率索引。

与正向问题不同，逆任务旨在寻找一个或多个能够产生期望响应$\mathbf{S}^*$的层序列$\mathbf{W}_{1:N}$ 。然而，映射$\mathbf{S} \mapsto \mathbf{W}_{1:N}$ 通常并非双射，因为许多不同的多层结构配置可以产生相似甚至相同的散射响应。这种固有的一对多结构，再加上各层之间复杂的电磁相互作用，使得直接求逆从根本上是病态的。为克服这一困难，我们采用受语言模型启发的自回归形式：正如这类模型根据其上下文 [58],预测下一个词，我们寻求在先前各层与目标$\mathbf{S}^*$的条件下预测下一个层令牌。

上述定义的正向问题和逆向问题都涉及处理有序层序列，其全局电磁响应源于层间相互作用的累积。如第 I 节所述，这种结构自然促使我们采用状态空间模型：随着序列的遍历，潜在状态逐步汇聚全局上下文。在一般的离散状态空间形式中，输入 $\mathbf{x}_n$ 会更新内部隐藏状态 $\mathbf{h}_n$ ，并生成输出 $\mathbf{y}_n$，其关系如下：

$$\mathbf{h}_n = \mathbf{A}_n \mathbf{h}_{n-1} + \mathbf{B}_n \mathbf{x}_n, \qquad \mathbf{y}_n = \mathbf{C}_n \mathbf{h}_n, \tag{2}$$

其中学习算子设置下 $\mathbf{A}_n, \mathbf{B}_n, \mathbf{C}_n$ 随……变化。至关重要的是，正向任务和逆向任务对序列信息的处理方式有着不同的结构要求。正向问题是一项确定性回归任务，其中散射响应

[1]需要注意的是，鉴于本部分所提出算法的有效性与灵活性，我们允许自己扩展可用的自由度，并同时考虑非对称阻抗片层堆叠（例如，$W_1 \neq W_N$）。对于理想（无损耗、单一工作频率）的全息超表面，这种扩展不应形成额外路径来提高指定透射相位下的单元透射率 [4], [7], ；而在更实际的情形中（不可忽略的介质损耗与导体损耗、在一段频率范围内工作），它可能能够生成新的有用解。虽然这种对称性破缺必然会引入双各向异性 [7], [16], [56], [57], ，但这预计不会对通常可以忽略波阻抗失配的设计（例如金属透镜）的性能产生负面影响。

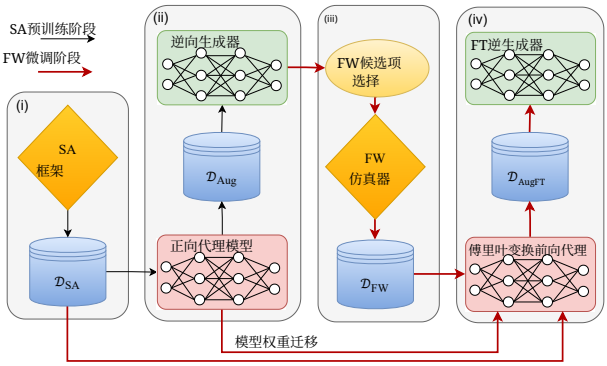


图 2。MetaMamba 处理流程。(i) 使用 SA模型生成大型合成数据集 $\mathcal{D}_{\mathrm{SA}}$，为数据驱动学习奠定基础。(ii) 在 $\mathcal{D}_{\mathrm{SA}}$上预训练 Bi-Mamba 前向代理模型，并用其生成增强数据集 $\mathcal{D}_{\mathrm{Aug}}$；随后使用该代理模型生成的数据训练初始 AR-Mamba 逆向生成器。(iii) 预训练逆向模型生成候选单元胞几何结构；筛选出高透射率代表项，并通过全波（FW）仿真进行评估，以构建校准集 $\mathcal{D}_{\mathrm{FW}}$。(iv) 使用 $\mathcal{D}_{\mathrm{FW}}$ 对正向代理模型进行微调（FT），并用其合成增强型高保真语料库 $\mathcal{D}_{\mathrm{AugFT}}$，从而训练最终经 FW 校准的 AR-Mamba 逆向模型。

$\mathbf{S}$ 通过相互电磁耦合依赖于所有层。因此，正向预测受益于双向处理，使信息能够沿层索引对称流动。相比之下，逆问题本质上是病态的且一对多：有效设计必须按顺序构建，同时与先前生成的层以及目标响应保持一致。这自然倾向于采用因果的自回归形式，其中每个输出$\mathbf{y}_n$ 对应于下一层参数的条件分布。基于这些考虑，我们使用Mamba实例化状态空间先验，采用双向变体进行正向代理建模，并采用因果自回归变体进行逆向生成。

B. 详细处理流程

基于这些观察，我们按照第 I 节所述继续推进；为建立可扩展且数据高效的解决方案，我们构建了一个混合框架，将 SA 建模、有限的高保真仿真以及序列感知的生成式学习相结合（见图 2）。为清晰起见，图 2 所示的工作流可从两个互补层面加以理解。在流程层面，它由四个步骤 (i)–(iv) 组成；而在概念层面，这些步骤自然归入两个训练阶段：一个*SA*预训练阶段（步骤 (i)–(ii)，细黑箭头），其中学习模型仅接触 SA 数据；以及随后的*FW*微调阶段（步骤 (iii)–(iv)，粗红箭头），其中代理模型被校准至全波仿真，同时保留从SA模型学到的广泛泛化能力，从而为最终的逆向模型训练形成高保真基础。

如图 2 的步骤 (i) 所示，混合框架首先使用 rapidSA 模型生成大型合成数据集 $\mathcal{D}_{\mathrm{SA}}$ 。该数据集为正向代理模型的预训练奠定基础，使其能够捕捉整个设计空间中的全局散射趋势。

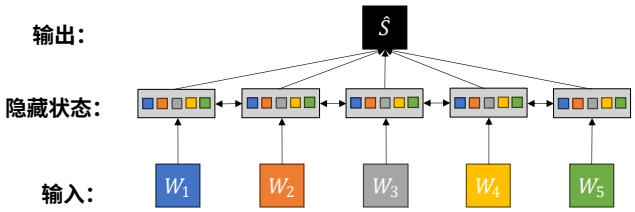


图 3. Bi-Mamba正向代理模型（图 2，步骤 (ii)、(iv)，粉色方块）。融合从左到右和从右到左的扫描，以捕捉全局层间耦合并预测散射响应 $\hat{}\mathbf{S}$。该架构通过在所有层元素之间传播上下文，高效地建模多层相互作用。

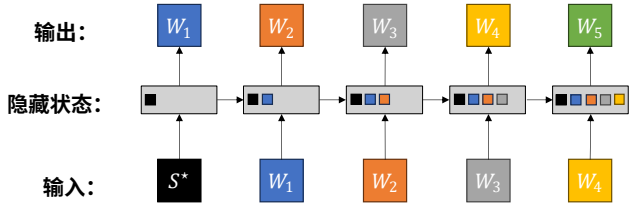


图 4. AR-Mamba 逆向生成器（图 2，步骤 (ii)、(iv)，绿色方块）。以目标响应 $\mathbf{S}^\star$为条件，模型从左到右依次预测层词元。在每个步骤 n，下一个词元 W_n 由前序词元 $\mathbf{W}_{<n}$ 和条件序列 $\mathbf{S}^\star$通过因果方式生成。这种自回归形式类似于语言模型，可为全息超表面设计生成多样化序列。

处理流程的下一部分（步骤 (ii)）从训练 Bi-Mamba 前向代理模型 $f_{\mathbf{S}}$ 开始：$\mathbf{W}_{1:N} \mapsto \mathbf{S}$，该模型经过训练，用于近似图 3 所示的正向映射。在这种状态空间表述中，层参数序列经过两个循环方向（正向和反向）的处理，其隐状态根据学习到的状态转移和输入依赖调制而演化，使模型能够突出最具信息量的层，同时保持线性时间评估。这种双向流模拟了层间的相互多重散射相互作用，使代理模型能够以线性时间效率在整个堆栈中传播信息。首先使用标准均方误差（MSE）损失函数对模型进行优化，训练数据为 $\mathcal{D}_{\mathrm{SA}}$：

$$\mathcal{L}_{\mathrm{fwd}} = \frac{1}{d_{\mathbf{S}}} \sum_{i=1}^{d_{\mathbf{S}}} \left(S_i^\star - \hat{S}_i \right)^2 . \tag{3}$$

其中 $d_{\mathbf{S}}$ 是$\mathbf{S}$ 序列的维度。此过程（图 2，步骤 (ii)，粉色方块）使代理模型能够捕捉从 LAYERS 学得、贯穿整个设计空间的全局散射趋势。

正向代理模型训练完成后，我们用它生成大型数据集 $\mathcal{D}_{\mathrm{Aug}}$，并通过Sobol采样 [59], 确保可行散射响应的密集覆盖，以用于逆向模型训练（图 2，步骤 (ii)，绿色块）。逆向模型学习概率函数：

$$p_W(\mathbf{W_{1:N}}|\mathbf{S}^\star) = \prod_{n=1}^{N} p_W(W_n|\mathbf{W_{<n}}, \mathbf{S}^\star), \tag{4}$$

使用非因果 AR-Mamba。序列首先将 $\mathbf{S}^\star$ 投影到高维嵌入空间，随后进行自回归词元预测，如图 4 所示。在训练过程中，采用教师强制 [60]，以最小化交叉熵：

$$\mathcal{L}_{\mathrm{inv}} = -\sum_{n=1}^{N} \log p(W_n \,|\, \mathbf{W}^\star_{<n}, \mathbf{S}^\star) . \tag{5}$$

为评估逆向模型的性能，重构平均绝对误差（平均绝对误差）指标利用代理模型量化逆向预测$\hat{}\mathbf{W_{1:N}}$ 的散射响应 $\mathbf{S}$ 与期望散射$\mathbf{S}^\star$之间的匹配程度：

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{d_{\mathbf{S}}} \sum_{i=1}^{d_{\mathbf{S}}} \left| S_i^\star - \hat{S}_i \right|, \tag{6}$$

由于并非任意期望的$\mathbf{S}^\star$ 都可物理实现，逆向模型旨在逼近可行分布，并返回接近$\mathbf{S}^\star$的设计。若$\mathbf{S}^\star$ 远离该分布，所得序列可能无法实现，而这可通过代理模型评估轻易识别。

设计由逆向模型通过top-k 或top-p 采样 [61], [62], 生成，再由校准代理模型$f_{\mathbf{S},\mathrm{cal}}$ 重新排序，从而构成单元胞几何$\mathbf{W}_{1:N}$的查找表，以实现期望响应$\mathbf{S}^\star$。在*top-k* 采样中，下一个词元从 k 个最可能的候选项中随机抽取；而在*top-p* 采样中，则从累积概率超过 p的最小候选项集合中抽取。至此完成图 2 的步骤 (ii)，得到仅在代理模型增强数据集上训练的初始逆向生成器。

为衔接高保真物理并实现步骤 (iii)，我们选取 M候选单元胞（M 针对图 1 所示配置的$\mathbf{W}_{1:N}$ 特定组合），并使用通用电磁求解器（CST）进行仿真。图5所示的候选样本选择流程遵循通用的两阶段策略。首先，以感兴趣的特性为条件（在本研究中为高功率传输效率与宽相位覆盖范围相结合），预训练逆向模型（图 2，步骤 (ii)，绿色模块）生成大量单元几何结构。随后，通过代理模型评估每个候选项，以获得其预测的散射特性（例如（$|T|^2$、ϕ））；我们以极坐标形式对其进行可视化，从而揭示生成的设计如何分布于幅度–相位空间中（图5a）。为确保全波校准集覆盖整个目标范围，同时避免几何结构冗余，我们将候选项划分为大小为 θ 的固定相位扇区，并在每个扇区内的几何参数空间（W_1、...、W_N）中应用 K均值算法[63] 聚类（图5b）。尽管相位扇区的选择是针对当前任务的，但在几何空间中进行聚类之前，也可以根据任何与任务相关的响应维度对候选项进行分箱，因此同样的方法能够更广泛地适用。随后，我们从每个聚类中选取具有最高透射率的代表项，形成一组紧凑且几何结构多样的全波仿真，其记为 $\mathcal{D}_{\mathrm{FW}}$ （图5c）。

全波校准是释放混合方案在精度–效率权衡方面优势的关键步骤。我们不从稀缺的全波数据开始训练，而是在图 2 的步骤 (iv) 中，使用在 $\mathcal{D}_{\mathrm{SA}}$ 上预训练的正向代理模型 $f_{\mathbf{S}}$ [64] 继续进行优化，并在微调过程中交替使用 SA 批次和全波批次（这种重演策略可缓解灾难性遗忘 [65]，对应于

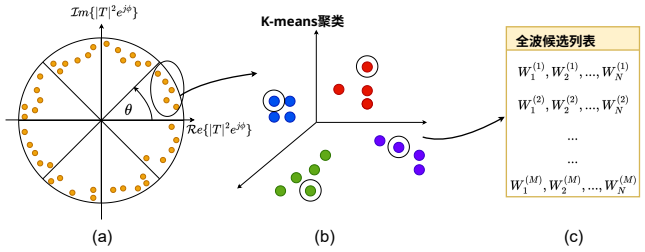


图5. 候选选择策略。(a) 预训练逆向模型生成大量高透射候选项，其预测响应（$|T|^2$, ϕ）以极坐标形式呈现，并划分到固定的相位扇区中。(b) 在每个相位扇区内，使用 K-means 聚类对几何参数空间 $\mathbf{W}_{1:N}$ 中的候选项进行聚类，以促进多样性。(c) 选取每个簇中透射率最高的代表项（在 (b) 中用圆圈标出），组成紧凑且多样的校准集；该校准集与全波仿真结果共同构成 $\mathcal{D}_{\mathrm{FW}}$。

图 2 中步骤 (iv) 的粉色模块所接收的红色粗箭头表示的过程）。校准目标函数混合了两种保真度水平：

$$\mathcal{L}_{\mathrm{cal}} \;=\; \lambda_{\mathrm{FW}}\,\mathbb{E}_{\mathcal{D}_{\mathrm{FW}}}\big[\mathcal{L}_{\mathrm{fwd}}\big] \;+\; \lambda_{\mathrm{SA}}\,\mathbb{E}_{\mathcal{D}_{\mathrm{SA}}}\big[\mathcal{L}_{\mathrm{fwd}}\big], \qquad (7)$$

其中，λ_{FW} 和 λ_{SA} 控制高精度锚定与广泛泛化之间的平衡。这既保留了从 $\mathcal{D}_{\mathrm{SA}}$ 中学到的全局覆盖范围，也使预测结果与高保真“真实值”保持一致。

校准后，代理模型 $f_{\mathbf{S},\ \mathbf{cal}}$ 会合成增强语料库 $\mathcal{D}_{\mathrm{AugFT}} = \{(\mathbf{W}_{1:N},\ f_{\mathbf{S},\ \mathbf{cal}}(\mathbf{W}_{1:N}))\}$，其精度接近全波精度，从而能够高效训练经过校准且准确的 AR-Mamba 逆向模型（图 2，步骤 (iv)，绿色方框）。至此，MetaMamba 流程已完成，校准后的代理模型可支持最终的高保真逆向生成器。

III. 结果与讨论

*A. HMS*配置与*SA*模型精度

我们将第II节中开发的MetaMamba框架应用于工作在20 GHz的一个代表性多层透射式HMS配置。最终目标是构建一个高保真、可用于制造的五层双极化单元查找表，以尽可能高的效率覆盖完整的 2π 透射相位范围。

如第II节所述，图1示意了本文研究的物理配置，该配置由五层堆叠的金属层组成，每层均带有JC图案。每个图案由一条支腿长度可变的水平铜线构成 W_n （第 n个JC支腿长度），并由介电基板和粘接层支撑。表I [12]总结了SA模型所用超原子的几何结构和材料参数。相应地，JC支腿长度的允许最小值和最大值分别为 $W_{\min} = 0$ 密耳和 $W_{\max} = 80$ 密耳，采样分辨率为 $\Delta W = 1$ 密耳，由此构成包含81个离散值的词表 V，以及大小为 $|V|^5 = 81^5 \approx 3.5 \times 10^9$ 种可能层组合的设计空间。选择 $\Delta W = 1$ 密耳是为了符合标准PCB制造能力，并确保相邻词元对应的散射响应平滑变化。

表I 用于SA模型的几何和材料常数 [12]

参数	值	材料 / 属性
周期， d	3.06 毫米	$\approx \lambda_0/4.9$ 在 20 GHz
波长， λ_0	14.99 毫米	在 20 GHz
十字宽度， w	4 密耳	铜
距单元边缘 $w_s/2$	2 密耳	—
铜厚度	18 μ米	$\sigma = 5.8 \times 10^7$ 西门子/米
基板厚度	30 密耳	Isola Astra MT77 ， $\varepsilon_r = 3,\ \tan\delta = 0.001$
粘结层厚度	2 密耳	Isola Astra MT77 ， $\varepsilon_r = 3,\ \tan\delta = 0.001$

在介绍MetaMamba学习流程（图 2，步骤(ii)–(iv)）之前，首先评估底层SA LAYERS模型的固有精度颇有裨益；前向代理模型最初便是在该模型上进行预训练的（步骤(i)）。为此，图6a给出了极坐标误差可视化结果，将留出单元胞配置集上的LAYERS预测结果与CST 真值进行了比较。重要的是，图6中所示的点并非查找表（LUT），也不是穷举设计扫描的结果。相反，这些点对应于第III-B节所述前向模型微调过程中使用的留出测试集，在此作为诊断性预览展示，以凸显校准的必要性。单位圆上的每个蓝色标记代表由LAYERS模型（a）或校准后的前向代理模型（b）预测的复数透射系数（$|T|^2e^{j\phi}$）。对于每种配置，红线在复平面中连接预测值与对应的CST结果，因此线段长度直接反映预测误差。如图6a所示，LAYERS预测结果相较于CST呈现系统性的、依赖相位的偏差，显然具有结构性而非随机性。在极坐标平面的多个区域，这些差异表现为较大的角度偏移，有时还伴随着 $|T|^2$的明显劣化。若逆向设计模型仅使用LAYERS生成的数据进行训练，这些差异会直接误导模型，因此需要在接下来的第III-B节引入校准阶段，使前向模型的预测与 CST 真值相一致。

B. 前向代理模型校准

为实现正向代理模型并随后评估其性能，我们遵循第II节（图 2）中所述的两阶段训练流程（图 2）。首先，在包含 $\approx$ 524,000 个 SA生成样本（$\mathcal{D}_{\mathrm{SA}}$）的数据集上预训练 Bi-Mamba [2]模型，以学习广泛的散射趋势，从而高效地对层到响应映射进行建模（图 2 中的步骤 (i) 和 (ii)）。接下来，为了减轻图6a所示 SA模型的残余误差，我们通过在数量有限的全量-

[2]Bi-Mamba 前向模型的完整架构细节，包括层维度和参数数量，见附录A（表 X）。本次校准阶段所采用的训练计划和超参数汇总于附录 B（表XIII）。

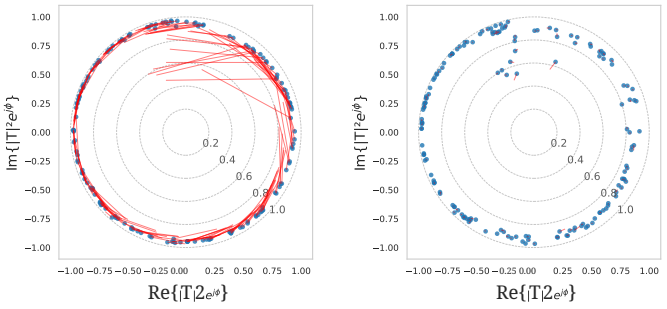


图6. 极坐标误差图比较了 CST 真值与 (a) SA LAYERS 模型以及 (b) 校准后的前向代理模型。每个蓝色标记表示用于校准研究（第 III-B 节）中某个留出测试配置的预测复透射率 $|T|^2e^{j\phi}$。红色线段将每个预测值与其对应的 CST 结果连接在复平面中；因此，其长度直接表示预测误差的幅度。面板 (a) 显示 LAYERS 模型存在明显的相位相关结构差异，而面板 (b) 表明校准后这些偏差大多已消除。

在 CST Microwave Studio 中采用周期性Floquet边界进行的波仿真（步骤 (iii) 和 (iv)）。

为在此校准步骤（步骤 (iii)）中选取信息丰富且具有代表性的样本，我们按照图5所示步骤，使用第II节所述的 AR逆向生成器（图 2，步骤 (ii)，绿色方框）生成 1080 个候选单元胞。所选采样策略为 top-k，其中 $k = 20$。这些单元胞几何结构通过在 SA 数据上训练的正向代理模型（图 2，步骤 (ii)，粉色方框）映射为相应的散射响应。随后，我们将结果划分为 $\theta = 5°$ 个扇区，并对设计几何序列（W_1，…，W_5）进行 K-means聚类（使用$K = 15$）。我们选取每个簇中透射率最高的几何结构，并使用 CST 对其全波响应进行仿真；共进行 1080 次仿真，构成数据集 $\mathcal{D}_{\mathrm{FW}}$（图 2，步骤 (iii)）。这种基于聚类的采样确保校准集涵盖多样且有意义的电磁行为，从而在最大限度提升微调效果的同时，尽量减少仿真预算。最后，使用式 (7) 的损失函数 $\mathcal{L}_{\mathrm{cal}}$ 优化校准模型$f_{\mathrm{S,\ cal}}$（步骤 (iv) 中的粉色方框）；完整的校准计划及相关超参数见附录 C。

在表 II 中，我们报告了式 (3) 的前向损失 $\mathcal{L}_{\mathrm{fwd}}$，其计算方式是在归一化目标[的正弦 ϕ、余弦 ϕ，$|T|^2$]上计算均方误差（MSE），并分别在 SA 和 CST 测试集上进行评估。相位通过其正弦和余弦分量进行编码，以避免环绕不连续性并确保表示平滑、连续，随后归一化至范围 [0,1]，从而促进稳定且均衡的训练。表中直接呈现出若干重要趋势。首先，尽管在所研究的 **W** 案例中，经过 SA 训练的前向模型在 SA 测试集上误差较低（2.3×10^{-4}），即与 LAYERS 得出的结果相比表现良好，但在 CST 真值上评估时，其误差增大了近两个数量级（1.9×10^{-2}），反映出 SA 求解器固有的建模误差。微调后，校准代理模型有意牺牲了部分一致性

表 II 低保真（LF）SA 和高保真（HF）CST 数据集上正向代理模型的测试均方误差。

	SA 测试	CST 测试
正向代理模型	2.3×10^{-4}	1.9×10^{-2}
傅里叶变换正向代理模型	4.5×10^{-4}	7.3×10^{-5}

正向代理模型对应图 2 的步骤 (ii)；微调后的傅里叶变换正向代理模型对应图 2 的步骤 (iv)。

与 SA 数据的一致性——其 SA 测试误差小幅上升至 4.5×10^{-4}——但同时在 CST 上取得了显著改善，将 CST 测试误差降低了 260 倍至7.3×10^{-5}。这一表现表明，微调（图 2，步骤 (iv)）成功地使代理模型向真实的全波物理特性靠拢，并证实没有发生灾难性遗忘。这些数值趋势印证了图6中观察到的定性表现。值得注意的是，图 6b 中红线的收缩清晰地显示出，图6a 原始 SA 预测中存在的系统性相位和幅度偏差得到了抑制。总体而言，结果证实，全波校准对于使正向代理模型达到 CST 级精度至关重要，从而实现可靠的高保真逆向设计。

作为合理性检查，我们还仅使用 CST 校准数据训练了一个正向代理模型，未进行 SA 预训练。该模型的 CST 测试均方误差为 9×10^{-4}，单独来看似乎颇具竞争力，但仍不如经过 SA 预训练和 CST 校准的代理模型。然而，在全局 LAYERS 测试集上评估时，仅使用 CST 训练的模型误差显著更高，达到 1.3×10^{-2}，这表明其表面上的准确性仅限于设计空间中的狭窄区域。鉴于 CST 校准集的多样性有限且存在很强的结构偏差，这种表现并不意外；该数据集主要包含高效率设计。仅用这些数据训练，得到的模型只能在局部进行插值，却缺乏全局的几何—响应背景。相比之下，尽管仅经 SA 预训练的代理模型在 CST 数据上的误差更高，尤其是在其置信度最低的区域，但它能够捕捉整个设计空间中的广泛趋势。校准代理模型兼具两者的优势：SA 预训练提供全局覆盖和物理背景，而 CST 校准则修正高效率区域中的系统性误差，因此在所有评估指标上都表现出一致的优越性。

表 III 中的数据进一步印证了这些观察结果。该表将 SA模型（LAYERS）和校准后的前向代理模型（图 2，步骤 (iv)，粉色方框）与来自 CST仿真数据集的留出测试集进行了比较。校准后的前向代理模型将效率误差降低了一个数量级，相位误差降低了超过 $30\times$，这证明了校准过程（步骤 (iii) 和 (iv)）的益处。图 7 的一致性图也清楚地体现了这种差异。(a,b) 面板对应 LAYERS预测结果与 CST 的比较，显示透射幅值和相位均存在显著散布和系统性偏差，这反映在中等水平的$\mathcal{R}^2$ 值上，分别为 0.591 和 0.677。相比之下，(c,d) 面板展示了校准后的前向代理模型与 CST 的比较，数据点紧密聚集在对角线附近，校准后的 $\mathcal{R}^2$ 得分在幅度方面接近 1

表 III 测试集上与 CST 的一致性。

	效率误差	相位[a] 误差
LAYERS[b] （均值）	0.065 (0.067)	21.13° (21.07°)
FT 前向代理（均值）	0.0067	0.656°
LAYERS[b] (标准差)	0.102 (0.092)	13.23° (12.97°)
FT 前向代理（标准差）	0.01	0.72°
LAYERS[b] (最小值)	≈ 0 (≈ 0)	4.5° (1.06°)
FT 前向代理（最小值）	≈ 0	≈ 0
LAYERS[b] （最大值）	0.59 (0.65)	80.39° (80.78°)
FT 前向代理（最大值）	0.09	4.68°

[a] 相位误差是以度为单位的包裹绝对差异。[b] 对于 LAYERS，括号外的值为测试集误差，括号内的值则是相对于完整 CST 数据集计算的。效率误差是以 $|T|^2$ 为单位的绝对差异。

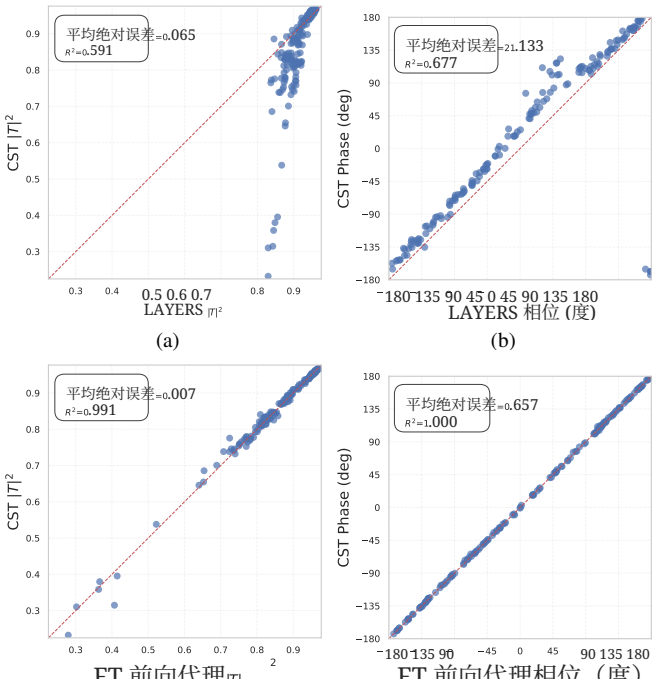


图 7. LAYERS 和微调前向代理模型与留出测试集上的 CST 对比。透射率 $|T|^2$ 和相位 ϕ 的一致性图对比了基于 SA 的预测和校准后的前向代理模型与 CST 的结果。（a、b）为 LAYERS 与 CST 的对比，（c、d）为校准后的代理模型与 CST 的对比。校准后的一致性显著提升，$\mathcal{R}^2$ 值从 0.59 提高到 0.99，$|T|^2$ 从 0.67 提高到 1.0。

和相位方面均接近 1。此处，$\mathcal{R}^2$ 表示决定系数，计算方法为 LAY-ERS 或代理模型预测结果与 CST 参考值之间的皮尔逊相关系数的平方；接近 1 的值表明，该模型能够准确捕捉设计空间内效率和相位的变化，证实校准模型在这方面优于 SA模型。

为量化 Bi-Mamba 前向代理模型的校准效率，我们比较了其在不同校准预算下与 CST 的一致性。所有变体均使用从包含 1080 次 CST 仿真的公共池中随机选取的子集进行训练，并在相同且严格保留的 162 个候选项测试集上进行评估（占总数的 15%）。对于每个预算（270、540 和 1080 个样本），均采用一致的 70%/15%

表 IV 不同 CST 预算下前向代理模型的校准消融实验（CAL）。误差为 $|T|^2$的绝对差异及以度为单位的包裹相位误差，均在相同的 162 个候选测试集上评估。

预算	指标	Mean	Std	Min	Max
270	相位（度）	1.1	2.02	≈ 0	19.81
	幅度	0.0125	0.0175	≈ 0	0.137
540	相位（度）	0.87	1.685	≈ 0	15.91
	幅度	0.0088	0.015	≈ 0	0.154
1080	相位（度）	0.656	0.72	≈ 0	4.68
	幅度	0.0067	0.01	≈ 0	0.09

训练/验证集划分，确保在不同校准规模之间进行直接且公平的比较。

如表 IV 所总结，即使只有 270 个 CST 校准样本，Bi-Mamba 前向代理模型的平均包裹相位误差也达到 1.1°，平均幅度误差为 0.0125；尽管校准预算有限，其结果仍与全波仿真高度吻合。将 CST 样本数量增加到 540 和 1080 后，相位误差和幅度误差均持续且单调下降，但精度提升幅度逐渐减小。随着校准预算增加，均值误差和标准差误差平稳下降，但最大误差的变化值得简要讨论。在较小预算（270 和 540 个样本）下，最大相位误差和幅度误差主要由少数离群配置造成，这些配置位于设计空间中样本稀疏的区域。这些区域在随机校准子集中的统计代表性不足，因此从有限的微调中获益较少。随着校准预算增加，这些罕见配置逐渐被纳入，最坏情况误差因而显著下降。值得注意的是，即使在最具挑战性的情况下，校准代理模型在所有预算下仍显著优于 SA LAYERS 模型（参见表 III），这表明剩余的大误差既局限于特定区域，也不具系统性。这些结果表明，只需数百个经过策略性选择的仿真，即可达到可靠且接近 CST 级的精度，凸显了所提出的 SA 预训练并选择性微调代理建模方法的数据效率。

C. 逆向设计性能

在建立高保真的正向代理模型（图 2，步骤 (iv)，粉色模块）后，我们接下来评估逆向生成器 [3] （步骤 (iv)，绿色模块）的性能。逆向模型的目标是合成多层 HMS 单元，使其散射响应与给定目标$\mathbf{S}^\star$相匹配，理想情况下实现高透射率并覆盖整个2π 相位范围。AR-Mamba生成器基于经校准后的正向代理模型生成的 $\approx 2 \times 10^6$ 个 Sobol序列样本进行训练，从而学习预测逐层几何结构

[3]AR-Mamba逆向模型的架构详见附录A（表 XI），其训练设置则见附录B。

以期望响应和堆叠中的先前层为条件。

推理过程中，生成器会以自回归方式预测每个层令牌，并以目标和先前生成的层为条件。最简单的策略 贪心解码是在每一步都选择概率最高的单个词元（top-$k = 1$）。这样会得到一个确定性设计序列，相当于模型针对给定目标给出的“最佳猜测”。采用贪心解码时，逆向模型在测试集上的重构损失为 $\mathcal{L}_{\text{rec}} = 3.9 \times 10^{-3}$ （参见 (6)），这表明模型能够以一对一的方式恢复与 CST 仿真一致的解。

除了贪心解码之外，MetaMamba 还可以利用随机解码策略生成多个候选设计。采用任一种采样策略时，模型都能在 GPU 上并行生成数千个候选项（每秒 ∼8 批次解码，每个批次的规模可达 10^3）。因此，我们转而采用基于采样的解码策略，并评估五层 JC 设计配置的物理边界。为此，我们通过步进扫描$|T|^2 \in [0.81, 1.0]$，以 0.01 为步长，对（$|T|^2$，ϕ）平面进行密集探测，并以 $2°$ 步扫描 $\phi \in [0°, 360°$。对于网格上的每个目标，AR-Mamba 生成器都会使用 top-$k = 20$ 采样生成一批包含 $B = 512$个候选设计的批次。每个相位区间内可达到的最高效率定义了可行性包络——在所考虑结构下，可实现的高效率响应的上边界。在单块 GPU 上，生成完整包络所需时间不到 10 分钟。图 8 所示的结果前沿揭示了非均匀的物理景观：许多相位的效率远高于 90%（浅绿色背景），而某些区域则因几何自由度受限以及实际金属中不可避免的导体损耗而存在固有局限（浅红色背景）。值得注意的是，在估算的可行性包络内，可实现的最小功率传输效率为 $|T|^2 = 0.8385$ （对应于$|T| = 0.916$），出现在 $\phi = 69°$。这一值代表包络中限制最严的点，并构成所考虑的五层 JC 配置下物理可实现的高效率响应的下界。我们通过额外的 CST 仿真（绿色实线）验证了包络上的若干选定点，结果证实该包络准确反映了底层的可实现响应集。

位于包络之外的目标，在当前设计空间中被视为分布外（OOD）目标。对于此类目标，逆向模型自然无法生成成功的设计，而是生成距离最近的可行包络点；如果没有点满足阈值，则不予生成。

确定可行响应集后，我们接下来分析生成模型在该包络内的行为。如果生成的设计同时满足相位精度约束 $|\phi_{\text{pred}} - \phi^\star| \le 5°$ 和效率约束 $|T_{\text{pred}}|^2 \ge 0.95\,|T^\star|^2$，则视为成功，其中（$|T^\star|^2$，$\phi^\star$）表示期望的目标响应，而（$|T_{\text{pred}}|^2$，$\phi_{\text{pred}}$）表示与逆向模型预测的设计序列对应的响应。

为量化随机解码运行中的性能，我们定义了若干互补指标。首先，通过成功率 （SRof (9)）评估可靠性，其定义为

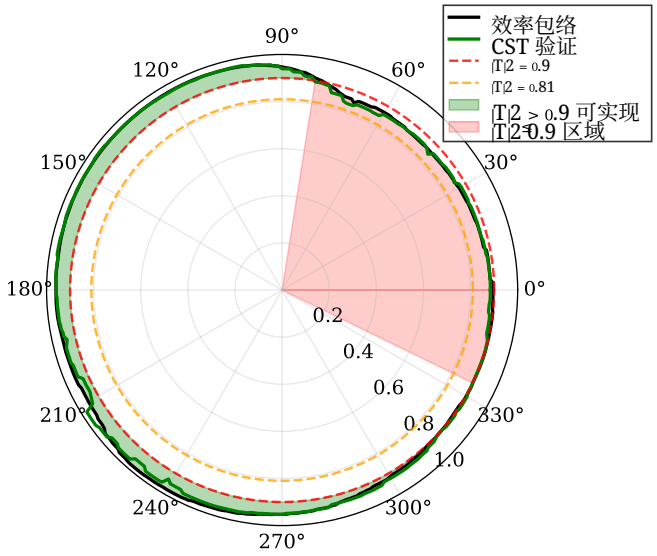


图 8. 使用校准后的逆向模型估计的可行性包络。红色虚线圆表示 $|T|^2 = 0.9$。该包络刻画了各相位下可达到的最大功率传输效率，揭示了（$|T|^2$，ϕ）域中物理上可实现和不可实现的区域。更多 CST 仿真证实了预测边界的有效性。

成功样本数（记为 n_s）与批次中生成样本总数 B之比。形式上，成功设计的数量为

$$n_s = \sum_{i=1}^{B} \mathbb{I}\Big(|\phi_{\text{pred}}^{(i)} - \phi^\star| \le 5° \ \wedge \ |T_{\text{pred}}^{(i)}|^2 \ge 0.95\,|T^\star|^2\Big), \tag{8}$$

其中 $\mathbb{I}(\cdot)$表示指示函数。相应的成功率为

$$\text{SR} = \frac{n_s}{B}. \tag{9}$$

其次，为了捕捉结构多重性，我们报告成功设计序列的唯一数量，其定义为满足成功判据的 n_s 个设计中不同几何序列$\{(W_1, \dots, W_5)\}$ 构成集合的基数。最后，为了量化逆向模型对离散设计空间的探索范围，我们引入多样性指标，定义为所有成功设计之间词元 ℓ_1 距离的两两均值：

$$\text{Diversity} = \frac{2}{n_s(n_s-1)} \sum_{i=1}^{n_s-1} \sum_{j=i+1}^{n_s} \frac{1}{N} \sum_{n=1}^{N} \left| W_n^{(i)} - W_n^{(j)} \right|, \tag{10}$$

其中 n_s 取自式（8）， N 为层序列长度。

定义成功率、唯一性和多样性指标后，我们现在考察这些量在可行响应空间中的变化情况。图 9 展示了在不同可行目标相位下，对 $B = 512$ 个样本进行 top-$k = 20$解码时，成功率与结构多样性之间的关系。每个标记对应一个不同的目标相位 $\phi^\star$，颜色表示唯一成功序列的数量。

显而易见，成功率和多样性强烈依赖于目标相位。 240°–270° 附近的区域兼具较高的成功率和最大的多样性，可生成

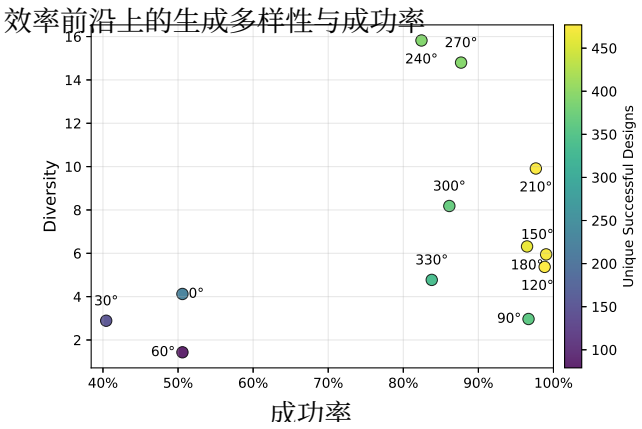


图 9. 采用排名前 $k = 20$ 的解码策略、预算为 $B = 512$个样本时，生成的设计的多样性 (10) 与成功率。每个标记对应可行性包络中的一个目标相位，颜色表示唯一成功序列的数量。

表 V 可行性包络内解码策略的性能。

策略	成功率[a]	多样性[b]	唯一性[c]
前-k (10)	0.822	4.900	338
前-k (20)	0.800	5.600	383
前-k (50)	0.788	5.660	373
前-p (0.90)	0.834	5.370	378
前-p (0.95)	0.818	5.470	366

[a] 满足 $|\Delta\phi| \le 5^\circ$ 和 $|T_{\mathrm{pred}}|^2 \ge 0.95 \cdot |T^\star|^2$ 的生成样本比例。[b] 多样性：成功设计中词元两两平均距离的中位数。[c] 唯一性：成功设计的不同设计序列数量的中位数。

数百种满足效率和相位约束的独特实现。相比之下， $0°$–$60°$ 区域的多样性和成功率始终较低——这正是可行性包络所表明的区域，其原因在于堆叠式 JC 几何结构的可调谐性有限且损耗特性存在内在物理约束。为补充这些发现，表 V 对比了几种解码策略。由此可见，top-k 和 top-p 采样均能保持较高的成功率（通常为 0.8 或更高）；按照式 (10) 的定义，top-$k = 20$在可靠性与结构多样性之间实现了较好的平衡。

这些结果表明，MetaMamba并不局限于单一的确定性解。相反，对于给定目标，它可以即时生成数百个不同的高保真单元实现。这种多样性具有重要的实际价值，既能为制造提供灵活性（例如特征尺寸约束），也有助于满足次级目标（例如带宽，另见第III-D节）。更为重要的是，所观察到的相位依赖性表明，该框架所获得的多样性与超表面响应潜在的物理可行性相关。

D. 宽带代理模型与功能后选择

为进一步展示所提出的混合学习方法的普适性和可扩展性，我们将校准后的前向代理模型扩展为能够预测18–22 GHz频段内的宽带散射响应，其中包含 $\approx 65,000$个频率响应（FR）。这一扩展得益于第一部分中报告的LAYERS所具备的光谱预测功能 [12]，该功能能够高效生成密集的频率响应，从而为学习宽工作频带内的色散行为提供了自然且可扩展的监督来源。

值得注意的是，扩展至宽带并不需要重新设计代理模型架构，也不需要改变训练流程。相反，前向模型通过引入一组学习得到的频率嵌入进行增强，这些嵌入依据离散频率索引对共享潜在表示进行条件化。在本研究中，将连续的 [18,22] GHz频段以 0.1 GHz 的均匀间隔离散化，得到 41 个频率索引。具体而言，将逐层几何表示聚合为紧凑的隐藏状态后，该状态会沿频率索引复制，并与可训练的频率嵌入结合，使网络能够利用共用的骨干网络解码频率相关效率和相位响应。通过简单调整条件分箱的数量（见附录A中的表 X），这一机制便可使同一模型预测单个工作频率或完整的频率响应，彰显了该方法的模块化与适应性。

宽带校准流程其余部分遵循第II节和图 2中所述的相同混合策略。首先，使用 LAYERS 在 [18,22]GHz频段内生成的约 65，000 个 SA 频率响应训练模型（如图 2步骤 (ii) 中的粉色方框所示），随后使用单频校准所采用的同一 组 1080 个 CST 候选样本进行微调（如步骤 (iv) 中的粉色方框所示）。这些 CST 仿真提供了整个频段内按频率分辨的真实值，无需重复候选样本选择流程即可完成宽带校正。

表 VI 比较了校准前后 $|T|^2$和相位响应的预测结果与 CST 仿真结果，涵盖 [18,22]GHz频段。SA LAYERS预测结果虽能捕捉总体频率变化趋势，但在整个频段内的效率和相位仍存在明显偏差。相比之下，经过校准的宽带正向代理模型将均值效率误差降低了 7.2 倍，并将相位误差降低了一个数量级以上，表明该校准流程成功地将 CST 级保真度迁移至各个频率。

图1 0 展示了两个测试集单元的代表性频谱。LAYERS 响应（橙色曲线）遵循总体色散趋势，但相较于全波结果存在系统性偏差；而校准代理模型（绿色曲线）在整个频段内与 CST 响应（蓝色曲线）高度重合，在图示尺度下几乎无法区分差异。这些结果证实，该宽带代理模型能够准确捕捉多层惠更斯超表面单元胞的效率和相位色散。

为补充这些代表性频谱，图11给出了多频率一致性图，比较了经校准的代理模型与CST在{v1}GHz频段内五个等间隔频率上的表现。

表 VI 与 CST 在留出测试集上的宽带一致性。

	效率误差	相位误差
LAYERS（均值）	0.0596	10.34°
FT 前向代理（均值）	0.0082	0.84°
LAYERS（标准差）	0.0827	15.64°
FT 前向代理（标准差）	0.0137	1.83°
LAYERS（最小值）	0.0088	2.71°
FT 前向代理（最小值）	0.0018	0.18°
LAYERS（最大值）	0.1326	57.96°
FT 前向代理（最大值）	0.0646	13.31°

[a] 相位误差为以度为单位的包裹绝对差异。[b] 所有指标均针对整个 [18,22] GHz 频段计算。

表VII 宽带正向代理模型在不同CST预算下的宽带校准消融实验。误差在 [18,22] GHZ 频段内评估，并报告为 $|T|^2$（效率）的绝对差异和以度为单位的包裹相位误差，评估采用相同的 162 个候选测试集。

预算	指标	Mean	Std	Min	Max
270	相位（度）	1.72	5.16	0.24	23.94
	效率	0.0165	0.0265	0.0035	0.0857
540	相位（度）	1.26	2.94	0.21	15.16
	效率	0.0130	0.0209	0.0029	0.1057
1080	相位（度）	0.84	1.83	0.18	13.31
	效率	0.0082	0.0137	0.0018	0.0646

这些结果显示，效率和相位始终紧密聚集在对角线附近，证实校准流程能够泛化至整个工作频带，而不仅仅是在标称设计频率处。这些结果还证实，使用同一紧凑训练语料库即可捕捉宽带散射行为，凸显了混合校准方案的数据效率和可扩展性。从物理角度看，该模型成功学习了多层HMS响应的连续色散特性。

为进一步评估宽带校准的数据效率，表VII报告了在不同 CST 校准预算下进行的消融研究。随着全波仿真次数从 270 次增加到 1080 次，整个频带内的相位误差和效率误差均持续降低，这表明额外的校准数据主要提高了色散精度，而非修正个别频率点。值得注意的是，即使只有 270 个 CST 样本，代理模型也已能相当精确地捕捉频率响应；而完整的 1080 个样本预算则使整个工作频带内的平均相位误差低于 1 度，效率偏差低于 0。01。

除了展示可扩展性外，宽带代理模型还能够评估逆向设计的功能多样性
生成的设计。当模型用于评估 AR-Mamba 逆向生成器为同一标称 20 GHz 目标合成的多个单元胞时，它表明，在 20 GHz 下传输幅度和相位几乎相同的设计，在 [18,22]GHz 范围内呈现出不同的频率相关特性。图 12 展示了这类代表项的示例，其中每条频率响应曲线都对应一种独特的逆向生成几何结构。

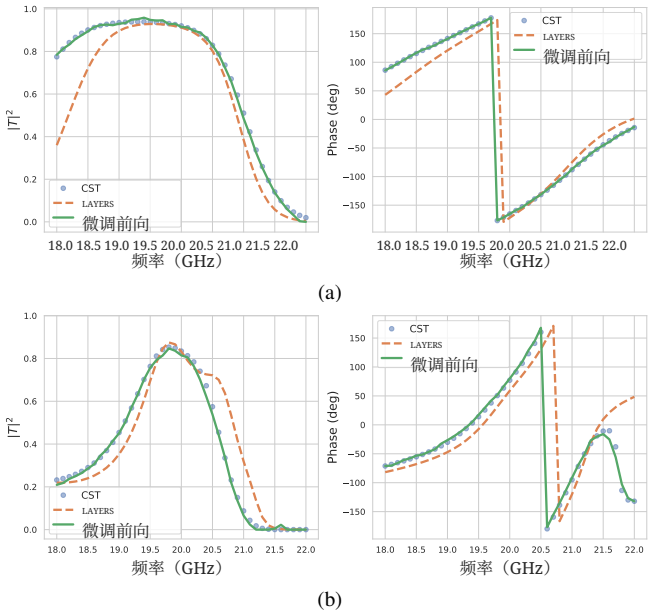


图10. 具有以下 JC支腿长度的代表性超原子的效率（左）和相位（右）响应：(a) **W** $=(14,58,11,55,15)$ 密耳；(b)**W** $=(22,17,15,20,24)$ 密耳，覆盖 [18,22] GHz频段。图中将 SA LAYERS 预测结果（橙色虚线）与 CST 真值（蓝色圆形标记）以及校准后的宽带正向代理模型（绿色实线）进行了比较。

这些示例见图 12，其中每条频率响应曲线都对应一种独特的逆向生成几何结构；这些结构在 20 GHz 下实现相同的目标相位 $\phi \approx 300^\circ$ 和传输功率效率 $|T|^2 \approx 0.9$，但在频带外呈现出不同的斜率和带宽。出现这种功能差异是意料之中的，因为即使多种几何结构实现了相同的标称散射，其内部模态结构和层间耦合仍各不相同，因而会产生不同的非谐振行为。

该能力补充了第一部分 [12, 第III-D节],中介绍的基于优化的宽带设计工作流。在该工作流中，SA 求解器通过在标称频率目标与频带平均性能之间进行权衡，一次定义一个单元。相比之下，MetaMamba利用一对多解集，能够即时生成许多满足标称要求的几何结构，并使用校准后的宽带代理模型筛选其频率响应。

在实践中，这种多样性带来了一个强大的全新设计自由度：获得一组满足标称频率规格的单元后，宽带代理模型便可用于后选出具有理想次要特性的变体，例如更宽的带宽、更平坦的群时延，或在谐振附近具有更高的稳定性，而无需重新训练生成器或运行额外的全波仿真。因此，生成多样性支持按频率进行功能专门化，并进一步支持针对不同设计任务的专门化。

E. 计算效率与运行时间分析

所提出的混合式 SA–生成式框架相较于传统的全波驱动工作流程，具有一项重要优势：计算开销极低。在此背景下，表 VIII 汇总了处理流程（图 2）中所有主要阶段在单频和宽带工作模式下的样本

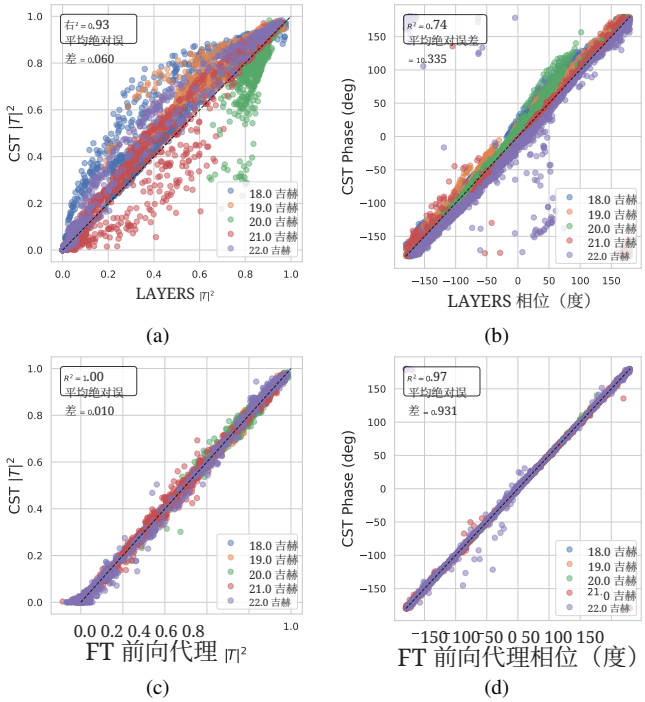


图11. 在 [18,22] GHz频段内，比较 (a,b) LAYERS 和 (c,d) 校准后的宽带正向代理模型与 CST 的多频率一致性图。(a,c) 透射率 $|T|^2$ 和 b,d) 相位的一致性图展示了五个代表项频率（18、19、20、21 和 22 吉赫）。

数量和运行时间；所有测量均使用附录 D 中详述的计算资源完成。结果揭示了一种明确且十分理想的趋势：扩展到宽带运行时，只有数据生成组件的成本会显著增加——具体而言，即构建 SA 数据集 $\mathcal{D}_{\mathrm{SA}}$ 和精简的 CST 校准集 $\mathcal{D}_{\mathrm{FW}}$。所有学习阶段，包括代理模型预训练、校准微调、逆向模型训练和 AR 解码，开销都很小，可在单块 GPU 上于几分钟到数小时内完成。这种分工直接源于框架的架构。LAYERS 模型以较低成本提供大规模监督信号，捕捉宽广设计空间中的底层物理机制；而 CST 校准集仅提供少量高保真修正。获得校准代理模型后，正向预测和逆向生成式设计几乎都能即时完成，从而可在数秒内评估或生成数千个单元。值得注意的是，即使在传统上计算负担极重的宽带情形下，总运行时间仍主要耗费在一次性的 CST 批次上，而学习和推理成本几乎保持不变。这些发现凸显了 MetaMamba 的实用可扩展性，以及它以远低于传统全波方法所需仿真成本实现 CST 级精度的能力。

另外两项观察进一步凸显了该工作流的实用性。首先，尽管我们将 1080 个 CST 样本作为保守的上限用于校准，

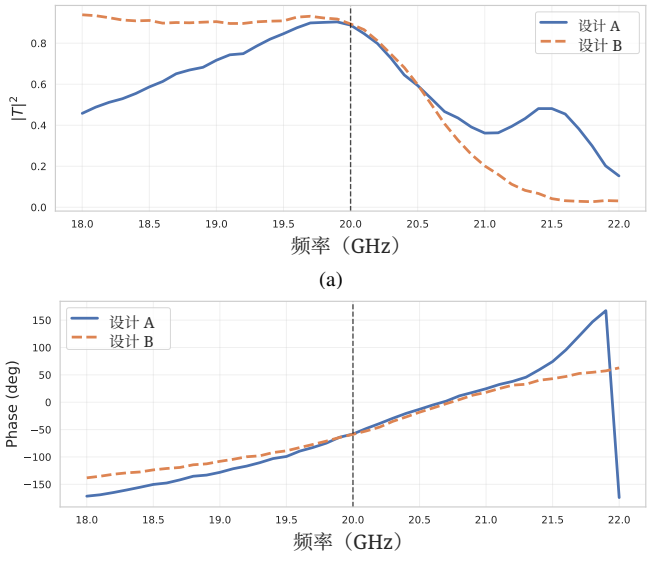


图12. 两个逆向生成单元的代表性频率响应：(a) 效率和 (b) 相位。两种设计在 20 GHz 处均与目标散射响应匹配，但在 [18.22]GHz 频段内呈现出不同的行为。设计 A（蓝色实线）和设计 B（橙色虚线）的生成序列分别为 $\mathbf{W}_A = (15,75,11,15,26)$ 密耳和 $\mathbf{W}_B = (6,5,17,8,5)$ 密耳。

我们的消融分析（见表 IV）表明，只需少得多的仿真，便足以实现与 CST 级别近乎一致的结果。这意味着，在保持较高预测保真度的同时，整个设计周期可以大幅缩短。相比之下，传统的机器学习驱动超表面逆向设计方法通常需要数千至数万个高保真全波仿真来进行有效训练 [30], [31], [33], [35]–[39]。其次，如果预先知道目标应用需要准确的宽带特性，则可仅 基于频率响应仿真进行 CST 校准。由于每次宽带 CST 仿真本身就包含标称的单频切片，因此无需再进行单频 CST 仿真，从而进一步减轻校准负担。因此，MetaMamba 工作流仅需传统方法一小部分的仿真成本，即可实现 CST 级精度，为多层超表面设计提供了一种可扩展且实用可部署的工具。

IV. 结论

本文提出了 MetaMamba，这是一种面向序列、用于透射型单元逆向设计的生成式处理流程。通过结合第一部分提出的 SA方案、经 CST 校准的 Bi-Mamba 前向代理模型以及 AR-Mamba 逆向生成器，该框架能够生成多样化的、经 CST 验证的五层单元胞，实现跨越完整 0–2π 相位范围的场透射幅度 $|T| > 0.9$，同时在约 72% 的相位范围内保持 90% 以上的功率透射效率；仅需 270 个带 CST标签的校准样本，即可在留出的全波测试集上达到近似 CST 级别的一致性。

表 VIII MetaMamba 工作流主要阶段的样本数量和运行时间，对应图 2 所示的处理流程。硬件规格见附录 D。

阶段	**样本**	**运行时间**
单频		
SA 数据集 ($\mathcal{D}_{\mathrm{SA}}$)	5.24×10^5	2 h
CST 数据集（$\mathcal{D}_{\mathrm{FW}}$）	1080	≈ 2 天
代理模型预训练	5×10^5	1 h
代理微调	$270 - 1080$	$< 1/2$ 小时
正向求解	$O(10^4)$	1 秒
逆向模型训练	2×10^6	1.5 小时
生成式解码	$O(10^4)$	1秒
宽带		
SA 数据集（$\mathcal{D}_{\mathrm{SA}}$）	65×10^3	12 h
CST 数据集（$\mathcal{D}_{\mathrm{FW}}$）	1080	≈ 10 天
代理模型预训练	64×10^3	$1/2$ 小时
代理微调	$270 - 1080$	$< 1/2$ 小时
正向求解	$O(10^4)$	1秒

本研究得出了若干重要结论。首先，在 SA方案奠定的基础上，前向模型校准仅需适量的 CST预算。表 IV 和表 VII 中的消融结果表明，即使校准集远小于我们的初始假设，仍能获得近乎相当的精度，凸显了该方法的数据效率。其次，逆向生成器能够针对每个目标即时生成多个各不相同的高保真设计，从而充分利用逆映射的一对多特性。第三，将元胞合成表述为序列生成任务，为扩展到更深层堆叠、更丰富的几何结构以及其他复杂电磁目标提供了自然路径。最后，与第一部分提出的 LAYERS SA框架相结合，体现了这项两部分研究的更广泛愿景：将快速的基于物理的建模与现代生成式序列学习相结合，从而为多层超表面构建可扩展、数据高效的逆向设计流程。

展望未来，MetaMamba框架无需进行根本性的架构调整，即可轻松扩展以实现多项重要功能。具体而言，可以丰富条件序列，以编码更广泛的电磁规格，包括反射特性、宽带频率响应、入射角依赖性和偏振多样性。总体而言，我们的结果表明，面向电磁超表面的数据高效生成式逆向设计不仅切实可行，而且只需极低的计算成本，便可达到全波仿真的精度。

附录A 架构

为确保内容完整且便于复现，本附录详细说明了正向代理模型（Bi-Mamba）和逆向生成器（AR-Mamba）所采用的确切网络架构，以及两个模型共用的 Mamba-2 构建模块。第 III 节报告的所有定量结果均采用本文所述架构获得。第 II 节从概念层面阐述了这些选择的依据，

表IX 正向与逆向Mamba架构（摘要）。

方面	前向模型	逆向模型
主架构	Bi-Mamba + 多层感知机头	因果 Mamba (语言模型头)
状态空间模型变体	Mamba-2	Mamba-2
d_{model}	256	256
中间维度	512	512
LAYERS	6	6
状态空间模型状态 d_{state}	64	64
参数	5.22M	5.00M
输入长度	5	可变（自回归）
输出长度	3	可变（自回归）
输入类型	连续长度 W	连续 S 和 离散值 W 词元
输出类型	连续 S	离散值 $W_{1:5}$
归一化	RMSNorm	RMSNorm
融合加法归一化	yes	yes
目标函数	回归（均方误差）	教师强制交叉熵

而本附录则提供实现层面的细节，包括维度、参数数量和数据流，以便忠实复现并有据地修改架构。

正向网络和逆向网络共享同一个状态空间骨干网络，但在序列信息的处理和输出方式上有所不同。正向代理模型处理的是确定性回归任务，将固定长度的层序列映射为全局散射响应，因此采用双向处理来捕获层间的双向耦合。相比之下，逆向生成器处理的是本质上病态的一对多合成任务，并被构建为非因果自回归模型，按顺序逐个输出层参数。表IX总结了这些高层次的架构差异，而后续的表 X和表 XI则分别给出了各模型的详细规格。

从信号流视角来看，正向模型首先将逐层几何参数嵌入共享潜在空间，通过 Bi-Mamba 骨干网络处理所得序列，并在映射至散射输出之前对序列维度进行池化，如图 3 所示。在宽带工作模式下，该池化后的几何状态会在离散值条件索引之间进行复制，并与学习式条件嵌入相结合（在本研究中，这些嵌入按频率索引），从而利用共享骨干网络实现频率相关解码。逆向模型则遵循相反的映射过程：目标响应的紧凑嵌入用于初始化序列，随后通过因果 Mamba 骨干网络和语言模型头（图 4 ）逐词元生成离散值层参数。尽管存在这些差异，两种模型都依赖相同的底层状态空间机制，以便在序列中高效传播上下文信息，无论采用双向方式还是因果方式。

前向网络与逆向网络均由相同的Mamba-2模块组成，其内部结构总结于表 XII。该模块将局部卷积混合与状态空间模块相结合，从而能够高效建模其中的短程依赖与长程依赖

表 X 前向模型的更多细节。

组件	规格
输入投影	线性层： $1 \to 256$
Mamba 主干网络	$6\times$ Mamba-2层（双向）
双向处理	正向和反向序列遍历 采用加性隐状态融合
池化（序列维度）	线性层： $5 \to 1$
频率索引 (FI)	1（单频 20 GHz）； 41（宽带 [18,22] GHz）
学习式条件嵌入	$\lvert FI\rvert \times 256$，添加到池化状态
幅度头	多层感知机： $256 \to 512 \to 1 \cdot \lvert FI\rvert$
相位头	多层感知机： $256 \to 512 \to 2 \cdot \lvert FI\rvert$
输出	拼接相位与幅度
激活函数	SiLU
参数量	5.22M

表 XI 逆向（AR-MAMBA）模型附加详情。

组件	规格
混合嵌入	浮点数： $1 \to 256$ 离散值：词汇表 (81) $\to 256$
Mamba 主干网络	$6\times$ Mamba-2层（因果）
因果性	$W_{1:5}$的自回归生成
语言模型头	线性层： $256 \to 81$
浮点序列长度	3（$\sin\,\phi$，$\cos\,\phi$，$\lvert T\rvert^2$）
词表	81 个词元（每 W_n）
参数量	5.00M
解码	top-k/ top-p

层序列。

附录B 训练超参数

本附录总结了用于Bi-Mamba代理模型和AR-Mamba逆向生成器的关键超参数与训练调度。这些设置是对附录A中架构规格的补充。

训练配置体现了两种模型各自不同的作用。正向代理模型经过训练以最小化平滑的回归目标，因此仅依靠验证损失进行学习率调度和早停。相比之下，逆向生成器仅使用交叉熵损失进行优化，而其下游物理保真度则通过重建误差指标进行评估。后者仅用于模型评估，并作为辅助早停标准，不纳入优化目标函数。以下汇总的超参数设置体现了这些选择。

两种模型均采用 AdamW优化器进行训练（权重衰减 10^{-2}），在回归和 AR 目标函数上都实现了稳定收敛。训练过程中采用平台期降低学习率调度策略：如果验证损失在固定训练轮数（耐心参数）内未能改善，学习率便会降低至原来的 0.1 倍，具体设置见表 XIII。采用早停准则，在进一步改善微乎其微时终止训练，从而

表 XII Mamba-2 块结构（两个模型均使用）。

组件	Op.	Dim.	备注
输入	x	$(B,\ L,\ 256)$	输入序列
预归一化	RMSNorm(x)	$(B,\ L,\ 256)$	输入归一化
输入投影	线性层	$256 \to 512$	扩展 d_{model}。
激活函数	SiLU	$(B,\ L,\ 512)$	非线性函数
一维卷积	一维卷积（k=4）	$(B,\ L,\ 512)$	局部上下文混合
SSM	Mamba-2	$(B,\ L,\ 512)$	状态空间（d_{state}=64）
Gate	逐元素 $\times$	$(B,\ L,\ 512)$	门控激活
输出投影	线性层	$512 \to 256$	返回模型维度。
残差	$x+$ 输出投影	$(B,\ L,\ 256)$	跳跃连接

表XIII 训练超参数。

参数	**Bi-Mamba**	**AR-Mamba**
批次大小	1024	1024
优化器	AdamW	AdamW
权重衰减	0.01	0.01
随机失活	0.0	0.0
初始学习率	0.001	0.001
调度器	ROP	ROP
学习率耐心值	2	2
学习率衰减因子	0.1	0.1
L_{rec} 停滞耐心参数	—	5
终止学习率	1×10^{-5} 学习率 < 终止学习率	1×10^{-5} 学习率 <终止学习率**或** L_{rec} 停滞
提前停止条件		

防止过拟合，避免不必要的计算成本。

附录 C 微调超参数

本附录报告了在 CST 校准数据上微调代理模型时所用的超参数，同时保留预训练 SA 代理模型中的知识（表 XIV）。校准策略采用了若干关键设计选择。首先，回放计划 以 2:1 的比例交替使用 CST 批次和 SA 批次，其中 CST 批次包含 128 个样本，SA 批次包含 1024 个样本，确保模型既能看到高保真修正，又能保持广泛覆盖。其次，采用差分学习率： 预训练骨干网络使用 1×10^4，而预测头使用 5×10^4，以便更快地适应任务专属层，同时保留学习到的表征。第三，使用两个超参数来控制高保真和低保真数据之间的平衡：

- 回放损失权重 （公式 7）： $\lambda_{FW} = 1.0$ 和 λ_{SA} =0.2 将完整 CST 损失与缩放后的 20% SA 损失混合，用于梯度计算。

- 验证加权： $\alpha = 0.7$ 在选择模型检查点时优先考虑 CST 验证性能。

训练采用带预热的余弦调度器，进行 100个训练周期。此配置可实现稳定校准，在贴近 CST 真值的同时保留 SA 泛化能力，并有效缓解灾难性遗忘。

附录 D 计算资源

本附录详细介绍用于数据生成、代理模型校准和

表 XIV 用于代理校准的微调超参数。

参数	值
训练轮数	100
批次大小（CST，高保真）	128
批次大小（SA，重演）	1024
重演率（CST:SA）	2:1
骨干网络学习率	1.0×10^{-4}
头部学习率	5.0×10^{-4}
骨干网络权重衰减	1.0×10^{-2}
头部权重衰减	1.0×10^{-4}
随机失活	0.0
调度器	带预热的余弦调度
重演损失权重 λ_{FW}； λ_{SA}	1.0；0.2
验证加权 α_{FW}； α_{SA}	70%；30%

逆向设计实验的硬件和软件环境。

A. 模型训练与推理

所有代理模型和逆向模型的训练、微调及推理均在单张 NVIDIA L40S GPU 上完成。软件环境如下：

- 操作系统：Linux 6.8.0-62-generic（x86 64，glibc 2.39）
- Python：CPython 3.10.16
- CUDA：12.8
- PyTorch：2.2

B. SA 数据生成

SA 数据集是在一台标准个人电脑上生成的：

- 操作系统：Windows 11 专业版
- 处理器：Intel Core i7-1355U（10 核，12 线程）
- 内存：16 GB

C. CST 仿真

所有全波仿真均使用*CST* 微波工作室*2023*年 在 Comm-Lab 设施的专用工作站上完成，规格如下：

- 工作站：HP Z8 G4（型号 Z3Z16AV）
- 操作系统：Microsoft Windows 11 专业版（工作站版）
- CPU： 2×Intel Xeon Gold 6244 @ 3.60 吉赫（每个 8 核，每个 CPU 16 个逻辑处理器）
- 已安装内存：192 GB RAM
- GPU：NVIDIA Quadro GV100（32 GB）

使用了频域求解器，该求解器不支持 GPU 加速。因此，表 VIII 中报告的所有 CST 运行时间均对应仅 CPU 执行。

致谢

作者感谢 FVMat 的 Doron Klepach 提供富有成效的讨论。

参考文献

[1] C. L. Holloway、M.A. Mohamed、E. F. Kuester 和 A. Dienstfrey，“超表面薄膜的反射和透射特性：及其在由谐振粒子构成的可控表面中的应用”，*IEEE* 电磁兼容汇刊，第 47 卷，第 4 期，第 853–865 页，2005 年。

[2] S. B. Glybovski、S. A. Tretyakov、P. A. Belov、Y. S. Kivshar 和 C. R. Simovski，《超表面：从微波到可见光》，《物理报告》，第 634 卷，第 1–72 页，2016年5月。[3] C. Pfeiffer 和 A. Grbic，《超材料惠更斯表面：利用无反射薄片定制波前》，《物理评论快报》，第 110 卷，第 19 期，第 197401 页，2013年5月。[4] F. Monticone、N. M. Estakhri 和 A. Al`u，《利用复合超表屏全面控制纳米尺度光传输》，《物理评论快报》，第 110 卷，第 20 期，第 203903 页，2013年5月。[5] M. Selvanayagam 和 G. V. Eleftheriades，《利用正交电流和磁流产生不连续电磁场以调控波前》，《光学快报》，第 21 卷，第 12 期，第 14409 页，2013年6月。[6] A. Epstein 和 G. V. Eleftheriades，《基于等效原理的惠更斯超表面：设计与应用》，《美国光学学会杂志B辑》，第 33 卷，第 2 期，第 A31–A50 页，2016年2月。[7] A. Epstein 和 G. V. Eleftheriades，《利用无源无损欧米伽型双各向异性超表面实现任意功率守恒场变换》，IEEE 天线与传播汇刊，第 64 卷，第 9 期，第 3880–3895 页，2016年9月。[8] E. F. Kuester、M. A. Mohamed、M. Piket-May 和 C. L. Holloway，《超表面薄膜上电磁场的平均过渡条件》，IEEE 天线与传播汇刊，第 51 卷，第 10 期，第 2641–2651 页，2003年10月。[9] S. Tretyakov，《应用电磁学中的解析建模》。Artech，2003年。[10] S. A. Tretyakov，《用于电磁场一般变换的超表面》，《英国皇家学会哲学汇刊A辑：数学、物理与工程科学》，第 373 卷，第 2049 期，第 20140362 页，2015年。[11] K. Achouri、M. A. Salem 和 C. Caloz，《基于极化率张量的一般超表面综合》，IEEE 天线与传播汇刊，第 63 卷，第 7 期，第 2977–2991 页，2015年7月。[12] S. W. Marcus、N. Nissan、V. K. Killamsetty、R. Yadav、D. Raviv、R. Giryes 和 A. Epstein，《利用选择性状态空间模型增强可制造多层惠更斯超表面的半解析设计：第一部分——基于场的半解析综合》，审稿中。[13] M. Chen 和 G. V. Eleftheriades，《用于无反射宽角度折射的欧米伽双各向异性线环惠更斯超表面》，IEEE 天线与传播汇刊，第 68 卷，第 3 期，第 1477–1490 页，2020年3月。[14] C. Pfeiffer 和 A. Grbic，《用于波前与极化控制的毫米波透射阵列》，IEEE 微波理论与技术汇刊，第 61 卷，第 12 期，第 4407–4417 页，2013年12月。[15] Z. N. Chen、T. Li、X. Qing、J. Shi、S. Li、Y. Su、W. E. I. Liu、C. Xue、Q. Lou、Z. H. Jiang、R. Xu、P. Liu 和 H. Sheng，《微波金属透镜天线》，《IEEE会刊》，第 111 卷，第 8 期，第 978–1010 页，2023年。[16] C. Pfeiffer 和 A. Grbic，《用于最优极化控制的双各向异性超表面：分析与综合》，《物理评论应用》，第 2 卷，第 4 期，第 044011 页，2014年10月。[17] E. Abdo-Sanchez、M. Chen、A. Epstein 和 G. V. Eleftheriades，《利用双各向异性惠更斯超表面实现受控辐射的漏波天线》，IEEE 天线与传播汇刊，第 67 卷，第 1 期，第 108–120 页，2019年1月。[18] M. Chen、E. Abdo-S´anchez、A. Epstein 和 G. V. Eleftheriades，《用于宽角度折射的无反射双各向异性惠更斯超表面的理论、设计与实验验证》，《物理评论B辑》，第 97 卷，第 12 期，第 125433 页，2018年3月。[19] G. Lavigne、K. Achouri、V. Asadchy、S. Tretyakov 和 C. Caloz，《无杂散衍射折射超表面的极化率推导与实验演示》，IEEE 天线与传播汇刊，第 66 卷，第 3 期，第 1321–1330 页，2018年。[20] G. Xu、S. V. Hum 和 G. V. Eleftheriades，《多层多阻带频率选择表面的设计技术》，IEEE 天线与传播汇刊，第 66 卷，第 2 期，第 780–789 页，2018年。[21] A. E. Olk 和 D. A. Powell，《纳入近场耦合效应的精确超表面综合》，《物理评论应用》，第 11 卷，第 6 期，第 064007 页，2019年。[22] O. Rabinovich 和 A. Epstein，《利用多层多元件超光栅实现任意衍射设计》，IEEE 天线与传播汇刊，第 68 卷，第 3 期，第 1553–1568 页，2020年3月。[23] S. Levy、Y. Kerzhner 和 A. Epstein，《纳入层间耦合的超表面微观设计严格解析模型》，载于 IEEE国际天线与传播研讨会美国国家委员会—国际无线电科学联盟无线电科学会议（APS/URSI）论文集，2019年，第 195–196 页。[24] S. Levy、Y. Kerzhner 和 A. Epstein，《基于纳入近场层内与层间耦合的严格解析模型的多层超表面综合与增强》，载于国际电磁学进展与应用会议（ICEAA）论文集，2019年，第 0528–0528 页。

[25] V. K. Killamsetty 和 A. Epstein，《按需多功能超表面的半解析综合方案》，载于第*15*届人工材料与新型波现象国际大会（超材料）论文集，2021年9月，第 123–125 页。[26] S. Kuznetsov、M. Tumashov、V. K. Killamsetty、P. Lazorskiy、A. Epstein 和 S. Glybovski，《利用多层金属-聚合物惠更斯超表面实现太赫兹波束的高效异常折射》，*IEEE* 太赫兹科学与技术汇刊，第 14 卷，第 1 期，第 109–121 页，2024年1月。[27] P. Mateos-Ruiz、V. K. Killamsetty、A. Epstein 和 E. Abdo-S´anchez，《惠更斯超表面漏波天线中任意且严格的孔径照明综合》，*IEEE* 天线与传播汇刊，第 73 卷，第 7 期，第 4208–4223 页，2025年7月。[28] K. V. Mishra、A. M. Elbir 和 A. I. Zaghloul，超表面设计的机器学习及其应用。约翰·威立父子出版公司有限公司，2023年，第 9 章，第 281–317 页。[29] T. Qiu、X. Shi、J. Wang、Y. Li、S. Qu 和 Q. Cheng，《深度学习：通往超表面自动设计的快速高效途径》，《先进科学》，第 6 卷，第 12 期，第 1900128 页，2019年。[30] P. Naseri 和 S. V. Hum，《一种基于生成式机器学习的多层超表面逆向设计方法》，*IEEE* 天线与传播汇刊，第 69 卷，第 9 期，第 5725–5739 页，2021 年 9 月。[31] P. Naseri 、 G. Goussetis 、 N. J. G. Fonseca 和 S. V. Hum，《利用生成对抗网络综合多频带反射式极化超表面》，《科学报告》，第 12 卷，第 1 期，第 17071 页，2022年。[32] H. P. Wang、D. M. Cao、X. Y. Pang、X. H. Zhang、S. Y. Wang、W. Y. Hou、C. C. Nie 和 Y. B. Li，《基于生成对抗网络、针对频带定制传输特性的超表面逆向设计》，《光学快报》，第 31 卷，第 23 期，第 37 763–37 777 页，2023年11月。[33] X. Liu、X. Cao、T. Hong 和 W. Jiang，《深度卷积生成对抗网络辅助的四通道全空间超表面逆向设计》，《光学快报》，第 32 卷，第 26 期，第 46 542–46 555 页，2024年。[34] Y. Lyu、Y. Hong 和 Z. N. Chen，《利用生成式深度学习实现双极化宽角度环形族频率选择表面的多目标综合》，*IEEE* 天线与传播汇刊，第 74 卷，第 1 期，第 456–466 页，2026年。[35] C. Niu、M. Phaneuf 和 P. Mojabi，《用于多层超表面单元综合的扩散模型》，*IEEE*《开放获取》天线与传播开放期刊，第 4 卷，第 654–666 页，2023 年。[36] R. Chen 、 C. W. L. Lee 、 P. K. Tan 和 R. Solanki，《通过ControlNet扩散生成模型实现高效率超透镜天线设计》，*IEEE* 天线与传播汇刊，第 24 卷，第 4 期，第 938–942 页，2025年。[37] G. Oliveri、M. Salucci 和 A. Massa，《迈向高效反射阵数字孪生——电磁驱动的机器学习视角》，*IEEE* 天线与传播汇刊，第 70 卷，第 7 期，第 5078–5093 页，2022年。[38] H. Yang、C. Yang 和 H. Li，《利用串联生成网络提升高自由度超原子设计的精度与速度》，《*ACS*光子学》，2025年。[39] J. Jung、S. Peng 和 Y. Rahmat-Samii，《可重构电磁表面的最新进展：工程设计、全波分析与大规模优化》，《电磁科学》，第 4 卷，第 35 期，2024年，第 1–15 页。[40] C. Yang、Y. Qin、Z. Zhang、Z. Zheng、J. Feng 和 H. Li，《利用多目标优化和扩散概率模型解决超表面逆向设计中的高性能数据稀疏问题》，《光学快报》，第 32 卷，第 23 期，第 40 869–40 884 页，2024年。[41] L. Hen、E. Yosef、D. Raviv、R. Giryes 和 J. Scheuer，《利用扩散模型进行衍射超表面逆向设计》，《*ACS*光子学》，第 13 卷，第 1 期，第 38–46 页，2026年。[42] M. S. I. Sarker、E. Podder、T. Mondal 和 S. Shafin，《机器学习与深度学习在天线设计、优化和选择中的应用：综述》，*IEEE*《开放获取》，第 11 卷，第 47 673–47 697 页，2023年。[43] C. Xue、Q. Lou 和 Z. Chen，《适用于5G毫米波系统的宽带双层惠更斯超表面透镜天线》，*IEEE* 天线与传播汇刊，第 68 卷，第 3 期，第 1468–1476 页，2020年。

[44] Y. Wang、F. Tian、L. Guo、H. Yang 和 X. Huang，《具有 360° 相位覆盖范围的高透射率惠更斯超表面及其超透镜应用》，《应用物理快报》，第 122 卷，第 23 期，第 231703 页，2023年。[45] I. Goodfellow、Y. Bengio 和 A. Courville，《深度学习》。MIT 出版社，2016年，http://www.deeplearningbook.org。[46] A. Graves，《使用循环神经网络生成序列》，2014年。[47] A. Gu、K. Goel 和 C. R´e，《利用结构化状态空间高效建模长序列》，2021年，arXiv:2111.00396。[48] C. A. Alonso、J. Sieber 和 M. N. Zeilinger，《作为基础模型的状态空间模型：控制理论综述》，2024年，arXiv:2403.16899。[49] A. Gu 和 T. Dao，《Mamba：利用选择性状态空间进行线性时间序列建模》，2024年，arXiv:2312.00752。[50] T. Dao 和 A. Gu，《Transformer 即状态空间模型：通过结构化状态空间对偶性实现广义模型和高效算法》，2024年，arXiv:2405.21060。[51] A. Liang、X. Jiang、Y. Sun、X. Shi 和 K. Li，《Bi-Mamba：用于时间序列预测的双向 Mamba》，2024年，arXiv:2404.15772。[52] Z. Wu、L. Li、Y. Li 和 X. Chen，《用于卫星通信的宽带圆极化超表面覆层天线》，IEEE《天线与无线传播快报》，第 15 卷，第 374–377 页，2016年。[53] M. F. Imani、J. N. Gollub、O. Yurduseven、A. V. Diebold、M. Boyarsky、T. Fromenteze、L. Pulido-Mancera、T. Sleasman 和 D. R. Smith，《计算微波成像超表面天线综述》，IEEE 天线与传播汇刊，第 68 卷，第 3 期，第 1860–1875 页，2020年。[54] S. Tariq、S. I. Naqvi、N. Hussain 和 Y. Amin，《一种基于超表面的 5G 毫米波应用 MIMO天线》，IEEE《开放获取》，第 9 卷，第 51 805–51 817 页，2021年。[55] N. Shlezinger、G. C. Alexandropoulos、M. F. Imani、Y. C. Eldar 和 D. R. Smith，《面向 6G 极大规模 MIMO通信的动态超表面天线》，IEEE《无线通信》，第 28 卷，第 2 期，第 106–113 页，2021年。[56] J. P. S. Wong、A. Epstein 和 G. V. Eleftheriades，《无反射宽角度折射超表面》，IEEE《天线与无线传播快报》，第 15 卷，第 1293–1296 页，2015年。[57] V. S. Asadchy、M. Albooyeh、S. N. Tcvetkova、A. D´ıaz-Rubio、Y. Ra’di 和 S. A. Tretyakov，《利用空间色散超表面完美控制反射和折射》，《物理评论B辑》，第 94 卷，第 7 期，第 075142 页，2016年。[58] T. Brown、B. Mann、N. Ryder、M. Subbiah 等，《语言模型是少样本学习器》，《神经信息处理系统进展》，第 33 卷，第 1877–1901 页，2020年。[59] I. M. Sobol’，《立方体中点的分布与积分的近似计算》，《苏联计算数学与数学物理》，第 7 卷，第 4 期，第 86–112 页，1967年。[60] R. J. Williams 和 D. Zipser，《一种持续运行的全循环神经网络学习算法》，《神经计算》，第 1 卷，第 2 期，第 270–280 页，1989年。[61] A. Fan、M. Lewis 和 Y. Dauphin，《层次化神经故事生成》，载于《计算语言学协会年会论文集》（ACL），2018年，第 889–898 页。[62] A. Holtzman、J. Buys、L. Du、M. Forbes 和 Y. Choi，《神经文本退化的奇特案例》，2020年，arXiv:1904.09751。[63] J. MacQueen，《多变量观测的分类与分析方法》，载于《第五届伯克利数学统计与概率研讨会论文集》，第 1 卷。美国加利福尼亚州伯克利：加利福尼亚大学出版社，1967年，第 281–297 页。[64] J. Yosinski、J. Clune、Y. Bengio 和 H. Lipson，《深度神经网络中的特征可迁移性如何？》，载于《神经信息处理系统进展》（NeurIPS），第 27 卷，2014年，第 3320–3328 页。[65] R. M. French，《联结主义网络中的灾难性遗忘》，《认知科学趋势》，第 3 卷，第 4 期，第 128–135 页，1999年。