arXiv:2611.30421v1 [physics.atom-ph] 23 Nov 2026

中性原子阵列的置信度感知判别式读出

Jaeick Bae, [1,2] Kangheun Kim, [1] Yunheung Song, [1] Junseo Ha, [1] Jeong Ho Han, [1] Jaewook Ahn, [2,3] Hyunsoo Ha, [4] Soonwon Choi, [4] 以及 Jongchul Mun[1,2,*]

[1] *Korea Research Institute of Standards and Science, Daejeon 34113, South Korea*
[2] *Graduate School of Quantum Science and Technology, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea* [3] *Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea* [4] *Center for Theoretical Physics—a Leinweber Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*（日期：2026年11月24日）

快速且可靠的检测是中性原子量子计算机的核心要求，但荧光成像仍然是快速实验循环、大规模阵列运行和量子纠错协议的限制步骤。在本研究中，我们利用判别式分类器，实现了对光镊中^{171}Yb原子的置信度感知读出。所提出的读出方法并不将每个原子位点图像直接强制转换为硬二元结果，而是生成测量可靠性评分。该评分由逻辑回归分类器生成，该分类器基于每个位点的总光子计数和荧光空间分布所提取的特征进行训练。所得分类器在成像时间约缩短一半的情况下，实现了相同的读出保真度。在10毫秒曝光时间下，其读出保真度达到99.83%，与传统的20毫秒光子计数阈值判定所达到的99.82%相当。更重要的是，分类器评分能够将低置信度结果转换为测量级擦除，从而在固定擦除预算下减少残余未标记错误。此外，该方法还显著抑制了假阳性；这对于原子重排等下游操作十分重要，因为假阳性可能产生目标缺陷。蒙特卡洛重排模拟表明，较低的假阳性率和基于置信度的擦除转换能够提升预测的无缺陷组装性能。本研究确立了利用判别式分类器进行置信度感知读出这一快速且可扩展的中性原子量子计算实用范式。

I. 引言

在当前的量子计算平台中，中性原子架构凭借其可扩展性和灵活连接性而崭露头角，展现出广阔前景 [1, 2]。然而，这些架构仍面临若干重要的系统级限制，包括相对缓慢的原子输运 [3, 4]，以及与物理量子比特操作相比仍然较长的基于荧光的测量时间 [5]。这种测量开销直接限制了中性原子处理器的周期时间。对于容错电路中的重复电路中途测量 [6, 7]，、连续装载架构中的快速子阵列重排和态制备 [8, 9]，，以及量子纠错（QEC）周期中空闲错误的抑制 [10]，这一点尤为重要。大规模算法可能需要持续数天进行容错运行，而面向中性原子的量子纠错和模拟架构越来越依赖于亚毫秒级的综合征提取、重排和少轮次综合征测量 [11–14]。这些应用要求可靠的荧光读出，其时间尺度必须与处理器门操作相匹配。

快速荧光成像的核心挑战不仅在于以更少的探测光子维持较高的二元读出保真度，还在于保留有关哪些读出决策可靠的信息。强烈的光子散射会加热原子并导致原子损失，而缩短测量时间会减少光子预算，并扩大存在原子与不存在原子这两类信号分布之间的重叠。然而，传统荧光读出会通过对标量光子计数值进行阈值判定，将每个感兴趣区域（ROI）压缩为一个二元决策。当读出光子预算有限时，部分感兴趣区域的测量结果会落在决策边界附近，因此其所分配标签具有较大的不确定性。硬阈值判定只返回二元结果，丢弃了这种不确定性，也损失了有关该结果可靠性的信息。

这促使我们采用置信度感知读出方法：读出结果不仅限于给出硬性的占据/空置标签，还包括一个衡量判定可靠性的标量。此类软 测量信息的价值已在量子纠错理论中得到充分证实，并且在超导量子比特处理器中得到了最直接的实验探索，其中模拟读出记录

可以纳入解码器，而不是通过硬阈值判定 [15, 16]将其丢弃。相关的解码研究思路也已应用于中性原子量子纠错模型，表明荧光读出的不确定性会影响逻辑性能 [17]。然而，在中性原子荧光实验中，图像信息通常会先被压缩为二元占据状态，然后再传递给重排、原子替换或重新初始化等下游操作。

* jcmun@kriss.re.kr

一个关键问题是，什么样的读出分数能够为荧光读出图像提供有用的可靠性信息。为解决这一问题，我们考察两大类基于软件的分类方法：生成式方法和判别式方法 [图1(a)]。生成式方法规定每个类别如何产生观测数据，例如通过类条件图像似然 $p(R|B)$和 $p(R|D)$，其中 R 是测得的荧光图像， B 和 D 分别表示有原子类和无原子类。随后使用贝叶斯定理推断后验类别概率。当图像信息模型和噪声模型准确时，这种方法具有物理可解释性，并且能够高效利用数据。然而，解析模型与实验图像分布之间的不匹配，可能限制其性能。相比之下，判别式方法使用标注数据直接学习有原子图像与无原子图像之间的决策边界。这可以使分类器对与分类任务无关的图像分布细节的模型设定错误不那么敏感 [18]，，并且随着获得更多具有代表性的标注训练数据，其性能通常会提高 [19, 20]。当类条件生成模型和类别先验均正确时，贝叶斯推断能够得到最优决策规则 [21]，；但在实验环境中，精确的生成式建模可能较为困难。这促使我们采用一种判别式读出策略，从标注数据中学习经验占据/空置边界。

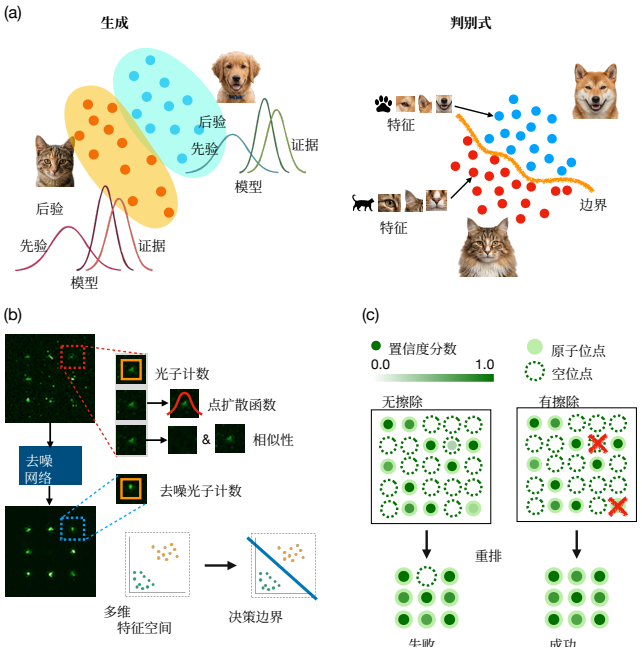


图 1. **判别式置信度感知读出的概述及其在原子重排中的应用。** (a) 生成式分类与判别式分类的概念比较。生成式方法对各类别如何生成测量数据进行建模，并利用先验信息和证据推断类别后验概率。相比之下，判别式方法从数据中提取特征，并直接学习一条用于区分两个类别的边界。(b) 本文采用的判别式读出流程。从每张荧光图像中提取标量感兴趣区域（ROI）特征，包括原始光子计数、去噪光子计数、点扩散函数（PSF）模板匹配分数，以及背景拟合相机噪声对数似然。这些特征构成多维特征空间，逻辑回归在该空间中学习占据/空置决策边界。(c) 利用置信度分数在原子重排中进行测量擦除转换。若不进行擦除转换，低置信度的假阳性事件可能被视为占据的源位点，从而导致重排失败。采用擦除感知读出时，同一低置信度事件会被标记并从源池中排除，从而实现成功重排。

在本工作中，我们证明了判别式分类器的输出值能够同时改善中性原子荧光读出图像的快速读出和基于置信度的错误标记。该分类器结合了基于物理的特征与人工智能学习的特征，包括原始光子计数、图像降噪后的光子计数、经验空间信息以及相机噪声似然特征 [图1(b)]。相较于传统阈值判定和贝叶斯后验规则基线，所得分数提高了经验占据读出保真度，并使达到相当读出保真度所需的测量时间减少约 $2\times$ 。除这一改进外，判别式分数还提供了更有用的可靠性信息：当将低置信度结果转换为测量级擦除时，

该分类器产生的残余错误少于传统阈值判定或贝叶斯基线。

具体而言，我们引入一种由人工智能驱动的去噪ROI特征，以减少类别重叠并降低假阳性率（FPR）；此处将其定义为无原子感兴趣区域被错误分类为有原子的概率。将去噪ROI特征纳入多特征判别式分类器后，所得的人工智能辅助读出可显著抑制假阳性，使FPR降低一个数量级以上。与基于生成对抗网络（GAN）的去噪等生成式图像方法不同， [22],我们的方法无需使用无噪目标图像训练去噪器，并且能够抑制幻觉。

这些改进与无缺陷原子量子比特阵列的组装直接相关。在重排过程中，假阳性占据错误尤其有害。如果一个空的储备位点被错误识别为已占据，它可能会被选作原子源，从而尝试将原子从空阱中移出，并在目标阵列中留下缺陷。基于置信度的擦除通过在重排前将含糊不清的位点排除出源池，缓解了这种失败情况 [图1(c)]。尽管增加读出曝光时间可以提高分类保真度，但我们的模拟表明，与不进行擦除、使用更长读出时间相比，快速读出结合基于置信度的擦除能够获得更高的无缺陷成功率，而无需进行更长时间的读出——

确定过滤。

II. 实验装置概述

本工作采用的成像方法遵循前期工作中所述的双频窄线宽成像方案 [23]。我们简要总结该方案，并重点介绍为同时采集读出图像和参考图像而进行的实验改动。我们使用双频驱动，在556纳米处对 ^{171}Yb原子在浅阱759.4 nm钟魔法波长光镊中俘获的 $|^1S_0,\ F=1/2,\ m_F=\pm1/2\rangle$ 和 $|^3P_1,\ F'=3/2,\ m'_F=\pm1/2\rangle$ 互组合跃迁进行成像。所使用的光镊阵列束腰为 $0.9\ \mu$m，阱深为 $200\ \mu$K。

原子荧光由数值孔径为 0.5 的物镜（Mitutoyo G Plan Apo $50\times$）收集，并由一个 50:50 分束器 [Fig.2] 等分。一条输出光路导向电子倍增电荷耦合器件（EMCCD）相机（Andor iXon Ultra 897），相机增益为 300，用于读出图像采集；另一条光路导向 qCMOS 相机（Hamamatsu C15550- 22UP），用于参考图像采集。读出图像和参考图像同时采集。参考图像以固定的 60 毫秒曝光时间采集，用于标签生成；而连续的读出图像对则以 $t_{img}=4$、5、8、10、12、14、16 和 *20* 毫秒的曝光时间记录，用于去噪器训练。相应的每个感兴趣区域的平均探测到的光子计数分别为 *4.*45、6.44、8.39、10.4、12.33、14.36、16.67 和 20.77。60 毫秒的参考曝光时间相较于约 5.26 秒的特征成像损失时间而言较短。

图像的尺寸为 64×64 像素，每个原子位点均使用包含 5 ×5个像素的感兴趣区域进行分析。测得的最近邻原子间距约为15像素，对应的实际原子间距为 $4.8\ \mu$米。对于每种读出曝光时间，数据集包含约 1.4×10^4至 3.0×10^4帧，对应约 6.9×10^3至 1.5×10^4 个配对采集组，因为每组记录了两张读出图像（见附录 A）。

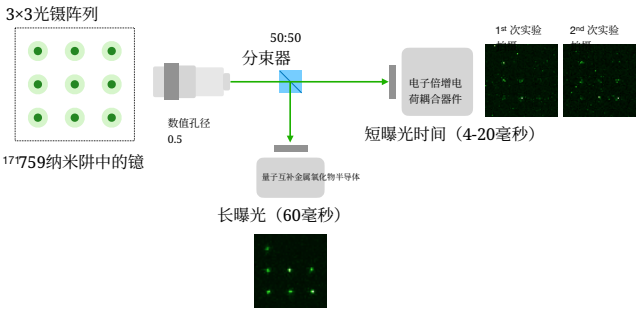


图2. **实验装置** 同时采集读出图像和参考荧光图像，拍摄对象为 3×3 ^{171}Yb光镊阵列。759纳米光镊阵列中受阱囚禁的原子发出的荧光，经556纳米成像光束和数值孔径为 $=0.5$ 的物镜收集。随后，荧光经50:50分束器分成两路。一路传至电子倍增电荷耦合器件（电子倍增电荷耦合器件）相机，以4–20毫秒的曝光时间采集读出图像；另一路传至量子互补金属氧化物半导体相机，以60毫秒的曝光时间采集参考图像，用于生成参考标签。每次实验拍摄中，电子倍增电荷耦合器件这一路连续采集了两张读出图像。

III. 数据驱动判别式读出

数据驱动判别式读出直接从带标签的实验数据中学习占据与空置位点之间的边界，而不是通过完全指定的像素级生成似然进行分类。在我们的实现中，每个原子位点感兴趣区域均由一组具有明确物理意义的标量特征表示，而逻辑回归——
回归分类器结合这些特征，得出原子存在概率。具体而言，我们从测量图像中提取 (i) 基于像素值、(ii) 基于去噪所得像素值、(iii) 基于背景拟合相机噪声模型的特征，以及 (iv) 基于点扩散函数（PSF）的特征，并将它们作为逻辑回归分类器的输入。

最终分类器所用的特征集，是通过系统评估一系列候选描述符并仅保留信息量最高的描述符而选定的。本文提出的框架并不限于此处考察的四种特征；原则上，任何基于物理或统计依据的感兴趣区域描述符都可以纳入判别式分类器。因此，我们还考察了其他候选特征，包括集中在感兴趣区域中心附近的原子信号比例、去噪器学习的噪声模型与观测噪声之间的相似性，以及从感兴趣区域中减去拟合的PSF模型后残留的空间结构。特征增补分析表明，除选定的四种特征外，这些额外描述符带来的变化很小（其定义见附录D 2，结果见表 D3）。尽管如此，该框架仍具有可扩展性，未来研究或可找到信息量更高的描述符，进一步提升分类性能或增强物理可解释性。

在本节中，我们将介绍选定的特征和逻辑回归模型。随后，我们分析由此带来的保真度提升，并量化各个特征的贡献。我们将该方法与两种具有代表性的读出策略进行基准比较：一种常规方法，根据检测到的光子计数进行阈值判定，以确定位点占据率 [24], ；另一种是 Martinez-Dorantes 等人 [25] 提出的基于EMCCD噪声的贝叶斯读出模型（见附录B）。

A. 特征描述与逻辑回归

我们选取了四个特征，以捕捉测量图像的互补信息。对于每个原子位点 s，我们定义一个以校准后位点位置为中心的 5×5 感兴趣区域（ROI） $c_s = (x_s, y_s)$，$\mathcal{R}_s = \{p = (x, y) : |x - x_s| \leq 2,\ |y - y_s| \leq 2\}$。其中，$p$ 用于索引相机像素。实验拍摄 t 中的原始图像记为 $I_t^{\mathrm{raw}}(p)$，单位为模数转换单位（ADU）；训练好的去噪器的输出记为 $I_t^{\mathrm{den}}(p)$。背景图像取自装载原子进入陷阱前采集的独立帧。相应的原始背景帧和去噪背景帧分别记为$I^{\mathrm{raw}}{}_{,b}{}^{\mathrm{bg}}(p)$和 $I^{\mathrm{den}}{}_{,b}{}^{\mathrm{bg}}(p)$，其中 b用于索引背景帧。原始背景和去噪背景的均值分别为$B^{\mathrm{raw}}(p) = \left\langle I^{\mathrm{raw}}{}_{,b}{}^{\mathrm{bg}}(p) \right\rangle_b$，$B^{\mathrm{den}}(p) = \left\langle I^{\mathrm{den}}{}_{,b}{}^{\mathrm{bg}}(p) \right\rangle_b$.我们从每个感兴趣区域中构造以下四个标量特征。

a. 原始图像中的光子计数。 第一个特征是5×5 感兴趣区域内背景扣除后的原始信号总量，

$$x_1(s,t) = G \sum_{p \in \mathcal{R}_s} \left[I_t^{\mathrm{raw}}(p) - B^{\mathrm{raw}}(p) \right], \qquad (1)$$

其中，$G = \kappa / g$ 用于在经过 EM增益校正后将 ADU计数转换为探测到的光电子。该特征估计从原子图像中收集到的光子信号总量。我们使用相机灵敏度 $\kappa = 4.15\ \mathrm{e}^-/\mathrm{ADU}$ 和 EM增益 $g = 300$，得到 $G \simeq 0.01383$。

b. 去噪图像中的光子计数。 第二个特征对去噪图像 [Fig.3(a)],进行相同的求和

$$x_2(s,t) = G \sum_{p \in \mathcal{R}_s} \left[I_t^{\mathrm{den}}(p) - B^{\mathrm{den}}(p) \right]. \qquad (2)$$

该特征捕捉经过训练好的去噪器抑制噪声后的类原子信号。去噪架构改编自 Noise2NoiseFlow 框架 [26, 27], ，该框架支持在没有干净目标图像的情况下进行训练。我们使用连续采集的两幅图像作为去噪器训练的成对含噪观测。具体而言，将两幅图像记为 $I_1 = S + n_1$ 和 $I_2 = S + n_2$，其中 S 表示这对图像中共有的底层荧光信号，n_1 和 n_2表示两幅图像中相互独立的噪声成分。首先，使用 DnCNN模块 [28] 对图像进行去噪，该模块通过生成干净图像估计来学习去除噪声。随后，使用归一化流噪声模型 [29, 30] 对残余噪声建模，并接入 DnCNN；该模型显式学习噪声分布的可逆概率密度函数。这样便能从数据中学习相机噪声统计特性，而不必通过人工设计的参数模型来加以规定。

对于我们的荧光数据，我们提高并优化了基于 DnCNN 的去噪模块的深度，并简化了流模型噪声模型。有关架构和优化后损失函数的更多细节见附录 E。如图 3(b) 所示，去噪降低了有原子和无原子感兴趣区域的信号分布方差。这种方差降低抑制了假阳性事件，如图 4(b) 和表 C2 所示。

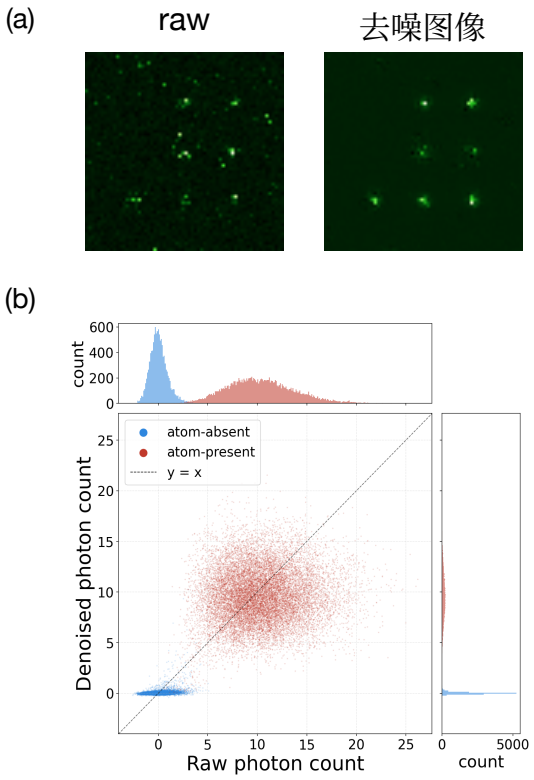


图 3. **去噪可降低快速测量中的光子计数方差。**(a) $t_{\mathrm{img}} = 10$ 毫秒时具有代表性的原始图像和去噪读出图像。去噪在保留原子荧光斑点的同时，抑制相机/背景波动。(b) $t_{\mathrm{img}} = 10$ 毫秒时原始光子计数与去噪后光子计数的分布，以及无原子和有原子感兴趣区域的直方图。去噪后的光子计数在降低无原子和有原子感兴趣区域中信号分布方差的同时，保留了有原子感兴趣区域的光子尺度，从而减少了两类之间的重叠。

c. ***PSF*** 模板匹配得分。 第三个特征用于衡量背景扣除后的感兴趣区域内计数的空间分布是否与通过经验测量获得的单原子图像一致。这里，点扩散函数（PSF）并不是指拟合得到的高斯模型或艾里模型。相反，我们直接利用参考成像臂识别出的有原子图像，构建经验原子图像模板。随后，我们定义一个归一化的模板匹配得分，该得分在数学上等价于背景扣除后的感兴趣区域与该经验模板之间的皮尔逊相关系数

以及该经验模板 [31]。对于每个训练折，经验模板 $T(p)$ 通过对所有含原子的样本中的背景扣除后的 5×5 感兴趣区域进行平均得到，

$$T(p) = \frac{1}{N_B} \sum_{(s,t):y_{s,t}=1} R_{s,t}(p) \tag{3}$$

其中， N_B 是含原子的训练样本数量，$y_{s,\,t}$ 表示参考标签，其中 $y_{s,\,t} = 1$表示含原子的样本，$y_{s,\,t} = 0$ 表示不含原子的样本。求平均后，将负值截为零，并将模板归一化至总和为 1。随后，将点扩散函数特征定义为背景扣除后的感兴趣区域与经验模板之间的皮尔逊相关，

$$x_3(s,t) = \frac{\sum_{p\in\mathcal{R}_s} \left(R_{s,t}(p) - \overline{R}_{s,t}\right)\left(T(p) - \overline{T}\right)}{\left\|R_{s,t} - \overline{R}_{s,t}\right\|_2 \left\|T - \overline{T}\right\|_2}. \tag{4}$$

上划线表示均值， $\|\cdot\|_2$ 表示感兴趣区域内像素的欧几里得范数。该特征旨在提供空间形状信息，用于衡量检测到的信号分布是否与单原子荧光图像相似。

d. 背景拟合的相机噪声对数似然。该特征描述图像像素分布与电子倍增电荷耦合器件噪声分布之间的相似性。噪声分布通过结合电子倍增统计、读出噪声和时钟诱导电荷的贡献来建模 [32, 33]。对于似然特征，相机模型参数由背景图像估计，得到 θ_{bg}。随后，我们将背景拟合的 EMCCD 对数似然特征定义为

$$x_4(s,t) = \frac{1}{|\mathcal{R}_s|} \sum_{p\in\mathcal{R}_s} \log P_{\mathrm{EMCCD}}\left(I_t^{\mathrm{raw}}(p); \hat{\theta}_{\mathrm{bg}}\right). \tag{5}$$

x_4 的值越小，表示观测到的感兴趣区域像素与根据背景图像拟合的电子倍增电荷耦合器件噪声分布越不相符，因此其中越可能包含原子荧光。这里使用解析的电子倍增电荷耦合器件输出噪声模型来实现探测器输出似然特征，因为探测器响应得到了充分表征，而且该特征在我们的数据中取得了最佳验证性能。

e. 逻辑回归分类器 我们采用逻辑回归作为一种简单的判别模型，为四个特征分配学习得到的权重，并将它们的加权和映射为原子存在概率。在拟合逻辑回归分类器之前，使用相应交叉验证折中训练样本计算得到的均值和标准差，对每个特征进行标准化，

$$\widetilde{x}_k(s,t) = \frac{x_k(s,t) - \mu_k}{\sigma_k}, \qquad k = 1, \ldots, 4. \tag{6}$$

标准化后的向量 $\widetilde{\mathbf{x}}(s,\ t) = (\widetilde{x}_1,\ \ldots,\ \widetilde{x}_4)$ 在四维特征空间中定义了一个点，其中每个维度对应一个标量图像描述符。该向量用作逻辑回归分类器的输入。原子存在概率建模为

$$P_{s,t} = P(y_{s,t} = 1 \mid s,t) = \mathrm{sigmoid}\left(w_0 + \sum_{k=1}^{4} w_k \widetilde{x}_k(s,t)\right), \tag{7}$$

其中 $\mathrm{sigmoid}(u) = 1/(1+e^{-u})$。令 w_0 表示偏置项，$\mathbf{w} = (w_1, \ldots, w_4)$ 表示特征权重向量。通过最小化 L2 正则化交叉熵损失来求得权重，

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}). \tag{8}$$

$$\begin{aligned} \mathcal{L}(\mathbf{w}) &= -\sum_{(s,t)\in\mathcal{T}} \ell_{s,t} + \frac{1}{2}\|\mathbf{w}_{1:4}\|_2^2, \\ \ell_{s,t} &= y_{s,t}\log P_{s,t} + (1 - y_{s,t})\log(1 - P_{s,t}). \end{aligned} \tag{9}$$

其中 $\mathcal{T}$ 表示训练集。使用固定阈值获得最终的二元预测，$\hat{}y_{s,t} = \mathbf{1}[P_{s,t} > 0.5]$。所有预处理步骤，包括 PSF 模板构建、特征标准化和分类器拟合，均在每个训练折内独立进行。报告的性能指标根据折外预测计算。

B. 保真度提升和误差分析

接下来，我们量化判别性特征表示相较于一维阈值判定 [24] 和贝叶斯基线 [25]对读出性能的提升。由于总保真度本身可能掩盖潜在的误差机制，我们还分别分析假阳性率（FPR）和假阴性率（FNR） [图4]。所有测量时间对应的完整保真度、假阳性率和假阴性率数值见附录 C。

完整的四特征分类器（4D-LR）融合光子计数、去噪强度、相机噪声似然和经验空间结构等互补信息，以提升读出保真度；在快速成像区间（$t_{\mathrm{img}} \le 10$ 毫秒）[, 其绝对增益最大]图 4(a)]。在 10 毫秒时，其保真度达到 99.83%，与 RAW-1D 在 20 毫秒时达到的 99.82% 相当，表明达到这一保真度水平所需的测量时间约减少了一半。从 12 毫秒起，它在所有测量时间下的保真度均超过 99.9%。

这一改进也改变了误差分布，大幅减少了假阳性判定 [图4(b)]。在 10 毫秒时，4D-LR 将假阳性率降至 0.12%，而 RAW-1D 和贝叶斯基线分别为 1.52% 和 1.22%。

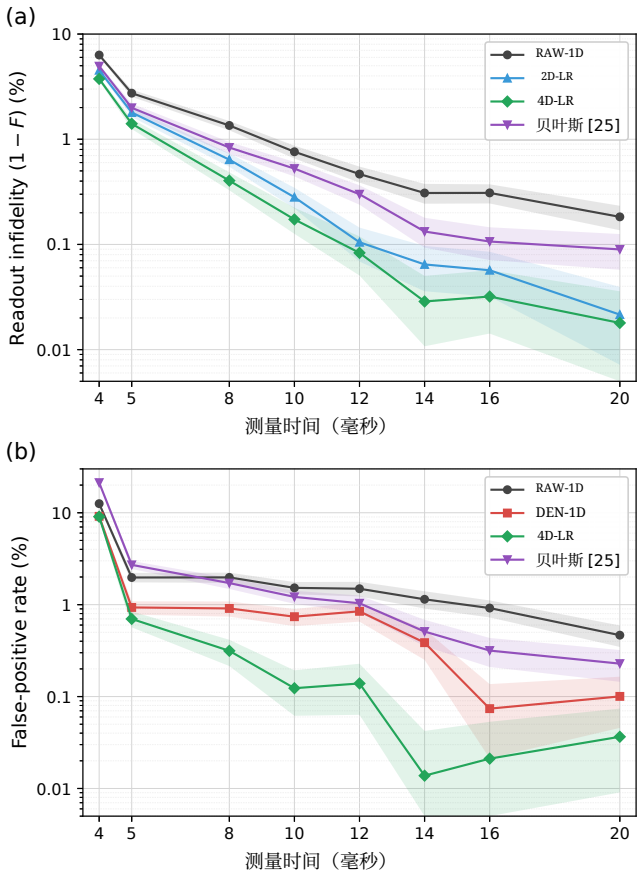


图4. **读出保真度与假阳性抑制。**(a) 传统阈值判定应用于原始图像（RAW-1D）、使用两个特征（原始ROI光子计数和背景拟合相机噪声模型似然）的双特征逻辑回归（2D-LR）、四特征逻辑回归（4D-LR）以及贝叶斯基线的读出不保真度 $1 - F$ 随测量时间的变化。判别式分类器2D-LR和4D-LR在低光子区间带来了最大的性能提升。(b) 各读出方法的假阳性率（FPR），其中包括对去噪图像进行传统阈值判定（DEN-1D）。与RAW-1D和贝叶斯基线相比，去噪ROI特征显著减少了假阳性错误。阴影表示基于固定折外预测的5000个帧自助法重采样样本计算的逐点95%置信区间。贝叶斯：根据 Martinez-Dorantes *et al.* [25]改编的贝叶斯基线。

与贝叶斯基线相比，这种抑制是以较高的假阴性率为代价的，这说明仅凭总保真度无法充分表征读出性能。

与更简单的读出方法进行比较有助于解释这一增益。使用原始光子计数和相机噪声对数似然的双特征分类器（2D-LR）已优于RAW-1D和贝叶斯基线，而对去噪图像进行阈值判定（DEN-1D）主要抑制假阳性。详细的误差率比较见附录 C。

这些结果凸显了直接从实验数据中学习决策边界的实际优势。尽管当类别条件分布和先验得到正确指定时，贝叶斯决策规则能够最小化分类误差，但在实验中准确地对这些分布建模可能颇具挑战。我们的判别式方法无需对观测图像进行完整的生成式描述，便取得了比所实现的贝叶斯基线更高的经验保真度。

为量化各特征对4D-LR的贡献，我们从完整分类器中每次移除一个特征（表 I）。在8毫秒时，移除相机噪声对数似然导致最大的保真度下降，其次是原始光子计数。移除去噪光子计数导致假阳性率增幅最大，凸显其在抑制假阳性判定中的作用。PSF模板特征带来较小但可测的保真度增益。这些结果支持组合互补特征，以同时改善读出保真度和误差分布。各测量时间的结果及特征构建分析见附录 D。

表 I. **4D-LR 分类器的留一特征消融分析。** 消融结果在 $t_{\mathrm{img}} = 8$ 毫秒时进行评估。每次从完整模型中移除一个特征。正值 ΔF 表示保真度下降，正值 Δ假阳性率表示假阳性率增加。变化以百分点 (pp) 报告。每列中的最大值以粗体显示。

移除的特征 re	ΔF (pp)	Δ假阳性率 (pp)
原始图像光子计数	0.070	0.026
去噪图像光子计数	0.062	**0.060**
PSF模板相关性	0.059	0.009
相机噪声对数似然	**0.140**	0.051

IV. 置信度感知的测量擦除转换

传统的原子读出通常会将每个感兴趣区域转换为二元决策，即使测量信号接近决策边界也是如此。然而，对于擦除感知纠错或原子阵列组装等下游应用，与其强行将歧义测量归为错误的二元赋值，不如明确将其标记为不确定，可能更为有用。因此，我们评估标量读出输出值是否也可用于识别可能出错的决策。

令 s 表示标量读出分数，其中，对于概率分类器，$s = p_{\mathrm{atom}}$；对于阈值方法，$s = \mathrm{sigmoid}[(x - x_{\mathrm{th}})/\sigma_{\mathrm{train}}]$，其中 x 为感兴趣区域求和，阈值 x_{th} 和尺度 σ_{train} 由训练折确定。我们将置信度定义为其与相应决策阈值$c = |s - s_{\mathrm{th}}|$之间的距离，其中 $s_{\mathrm{th}} = 0.5$。因此，接近阈值的判定被视为具有更大不确定性，并可标记为擦除。这类似于在量子纠错中使用软 测量

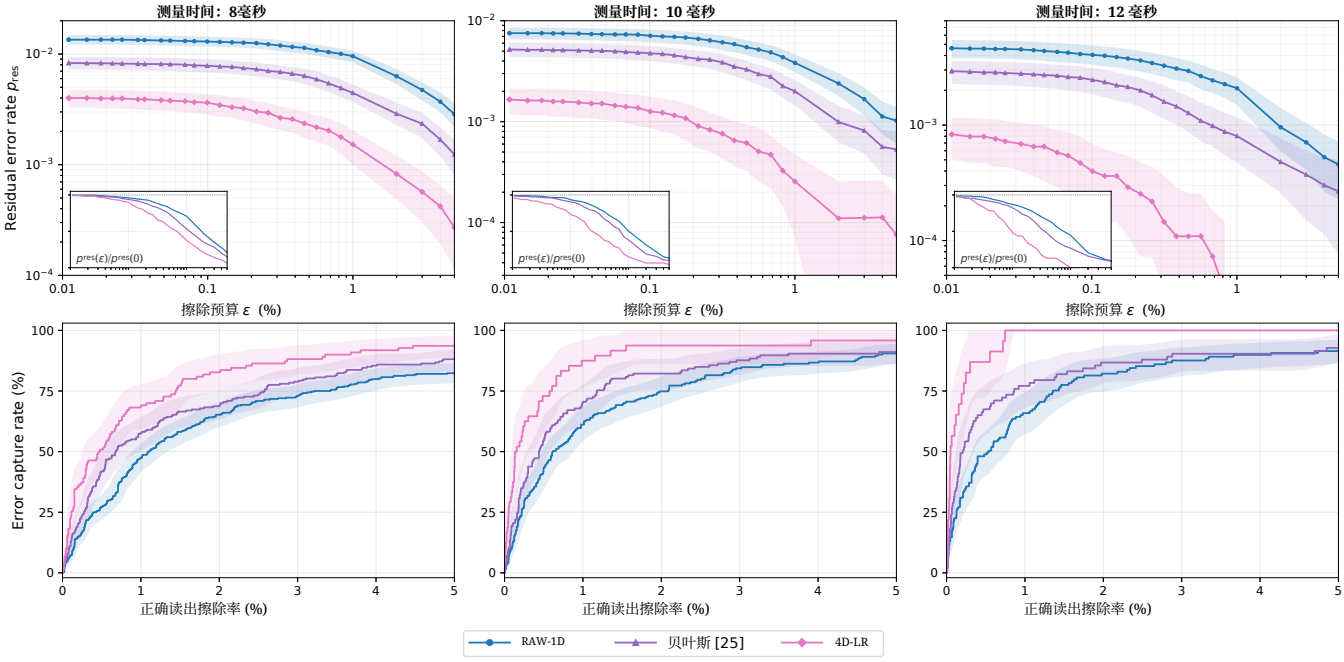


图 5. **置信度感知的测量擦除转换。** 上方各图：残余未标记错误率 $p_{\mathrm{res}}(\epsilon)$ 随擦除预算 ϵ 的变化，展示了具有代表性的测量时间。感兴趣区域按决策置信度从低到高排序，并将置信度最低的比例 ϵ 标记为擦除。插图展示归一化残余错误率 $p_{\mathrm{res}}(\epsilon)/p_{\mathrm{res}}(0)$。下方各图：通过改变置信度阈值获得的错误检测ROC曲线。横轴为被标记为擦除的正确读出结果（TP+TN）所占百分比，纵轴为被标记为擦除的读出错误（FP+FN）所占百分比。各分母分别为擦除筛选前对应项目的总数。越靠近左上角的曲线，表示能够捕获更多错误，同时将更少的正确结果标记为擦除。阴影带表示逐点95%置信区间，该区间由对采集帧进行3,000次配对自助重采样样本计算得出；每帧均保留全部九个感兴趣区域的结果，并固定折外预测。

信息：在量子纠错中，似然或后验分数的大小量化了测量比特 [15]的可靠性。这也与擦除转换策略有关，在这些策略中，泄漏检测事件会被转换为有定位的擦除，而不是任其成为未标记错误 [34, 35]。

令 $\mathcal{D}$ 表示给定曝光时间下、进行擦除筛选之前的所有感兴趣区域决策集合。我们按置信度从低到高排列这些决策，并将置信度最低的决策标记为擦除。擦除预算 $\epsilon = |\mathcal{E}_\epsilon|/|\mathcal{D}|$ 是被标记决策占所有决策的比例，其中 $\mathcal{E}_\epsilon$ 为被标记决策的集合。令 $\mathcal{E}_\epsilon^c \equiv \mathcal{D} \setminus \mathcal{E}_\epsilon$ 表示相应的未擦除感兴趣区域决策集合，$\mathcal{M} \subset \mathcal{D}$ 表示读出错误集合。对于每个擦除预算 ϵ，残余错误率是未擦除决策中错误决策所占的比例，

$$p_{\mathrm{res}}(\epsilon) = \frac{|\mathcal{M} \cap \mathcal{E}_\epsilon^c|}{|\mathcal{E}_\epsilon^c|}. \tag{10}$$

图 5 上方各图显示残余错误率 p_{res} 随擦除预算 ϵ 的变化情况。

图 5 的下方各图比较了在标记正确读出结果的成本相匹配时的错误检测效果。其中，$\mathcal{C} = \mathcal{D} \setminus \mathcal{M}$ 表示正确读出结果的集合，横轴和纵轴分别给出了被标记为擦除的正确和错误读出结果所占的比例，

$$r_{\mathrm{corr}}(\epsilon) = \frac{|\mathcal{E}_\epsilon \cap \mathcal{C}|}{|\mathcal{C}|}, \tag{11}$$

$$r_{\mathrm{err}}(\epsilon) = \frac{|\mathcal{E}_\epsilon \cap \mathcal{M}|}{|\mathcal{M}|}. \tag{12}$$

在给定正确读出擦除率下，错误捕获率越高，表明错误决策与正确决策的区分效果越好。

读出分数的价值不仅在于其硬判决的准确性，还在于能够识别哪些判决可能是错误的。图 5 显示，4D-LR 在这两方面都具有优势。在所测试的擦除预算范围内，它实现了最低的残余未标记错误率；同时，其下降更快的归一化曲线表明，低置信度结果涵盖了其初始错误中更大的比例。因此，这种优势不仅体现在更高的读出保真度上，还体现在能够更有效地识别不可靠判决。这样一来，就可以更有效地利用有限的擦除预算，减少仍未被标记的错误。

V. 通过测量擦除进行原子重排的改进

原子重排是置信度感知读出的一个直接应用，因为成功组装取决于被选作原子源的位点是否可靠。假阳性判定可能导致选中空位点，从而在目标阵列中留下缺陷；而假阴性判定主要会减少可用的原子池。这种不对称性促使我们在重排之前利用测量置信度，排除不确定的占据位点预测结果。这种具备擦除感知能力的源选择可以减少错误的源分配，前提是剩余的原子足够填满目标。

我们使用蒙特卡洛模拟量化了这一效应。在装载概率为 $p_{\mathrm{load}} = 0.5$ 的 30×30 储备阵列上初始化原子，并应用重排算法填满一个方形目标阵列。将测得的假阳性率和假阴性率作为模拟输入，并使用 $P_{\mathrm{surv}}(t_{\mathrm{img}}) = \exp(-\gamma_{\mathrm{loss}} t_{\mathrm{img}})$（$\gamma_{\mathrm{loss}} = 1.9 \times 10^{-4}$毫秒$^{-1}$）纳入成像诱导的原子损失。

我们评估了单轮和多轮重排，并将成功率定义为在每种方案下实现无缺陷组装的蒙特卡洛试验所占的比例。单轮方案要求在一个读出并移动周期后填满固定的 $L \times L$ 目标，且该轮内不进行额外读出或重新装载。多轮方案最多允许 20 个周期，并根据检测到的候选原子数量调整每轮的目标大小。每轮失败后，仅保留尝试目标区域内剩余的原子供下一轮使用；未使用的储备原子将被丢弃，不再重复使用。成功要求在允许的轮数内获得至少 $L \times L$ 的无缺陷阵列，因此 L 指定了该方案中的最小目标尺寸。对于考虑擦除的方法，每次移动前都会将被判定为已占据的低置信度位点排除在源池之外。重排算法和模拟流程见附录 F。

基于置信度的源选择可提升多轮和单轮重排中的无缺陷组装效果。对于最小目标尺寸为 20×20 的情况，采用擦除感知选择的 4D-LR 在 $t_{\mathrm{img}} \geq 8$ 毫秒 [Fig.6(a)] 下实现了超过 90% 的多轮成功率。在 10 毫秒时，擦除感知选择将预测的多轮成功率从 88.09% 提高到 98.44%。仅允许一个重排周期时，其优势尤为明显。基于 4D-LR 的方法是唯一能在 8–12 毫秒时取得显著成功率的方法[Fig.6(b)], , 因为此时错误的源分配无法通过后续读出和重排来纠正。在 10 毫秒时，擦除感知选择将预测的单轮成功率从 59.87% 提高到 96.37%；相比之下，DEN-1D 为 4.90%，贝叶斯基线为 0.86%，RAW-1D 为 0.22%。

擦除预算的影响揭示了源可靠性与原子可用性之间的平衡。拒绝不确定的候选项可以减少错误的源分配，但过度拒绝会导致可用于填满目标的原子过少 [Fig.6(c)]。在适当的擦除预算下，采用基于置信度的选择并缩短读出曝光时间，效果可能优于不进行源过滤而延长读出曝光时间。例如，采用擦除筛选的 10 毫秒单轮 4D-LR 成功率为 96.37%，而 20 毫秒且未进行擦除筛选时为 85.53%。

VI. 讨论

我们的贝叶斯基线采用基于物理的似然模型，明确考虑了光子计数、相机噪声和原子损失。其性能取决于该模型对占据与空置位点的荧光图像描述得有多准确。

若类别条件似然和先验指定正确，将每幅图像分配给后验概率较高的类别，可使预期分类错误率最小。4D-LR 更高的读出保真度以及更有效的读出错误识别能力，表明我们的贝叶斯基线所用的似然模型仍有改进空间。4D-LR 使用的特征为检验贝叶斯似然模型提供了具体方法。例如，比较空位点的模拟与测量图像，可以揭示该模型能否再现背景涨落的出现频率，尤其是那些看起来像原子荧光斑点的背景涨落。

因此，一种可能的扩展方案是保留基于物理的似然模型作为读出主干，同时利用校准图像调整最终的占据/空置判定规则。这种混合策略 [20] 既能保留贝叶斯模型的可解释性和数据效率，又能纠正由解析似然模型未能充分涵盖的实验效应所导致的系统性判定偏差。

判别式分类器基于特征的结构使该框架具有天然的可扩展性。更系统的特征发现或可揭示其他有用的可观测量，而这些可观测量又可能为改进贝叶斯似然的生成式建模提供指导。一个特别有前景的方向是利用时间信息，例如光子到达统计或逐次强度动态；这些信息可能提供超出本文所考虑的空间感兴趣区域特征之外的信息。

其中有几个方面与未来的实验部署相兼容。首先，该去噪架构不依赖于测量设备的特定模型。它从成对含噪图像中学习相机噪声分布，因此一旦获得适当的训练数据，便可适配其他荧光成像系统。这样的变体将使逻辑回归分类器无需依赖明确定义的相机输出模型即可运行，从而为构建相机模型无关的读出流程开辟途径。其

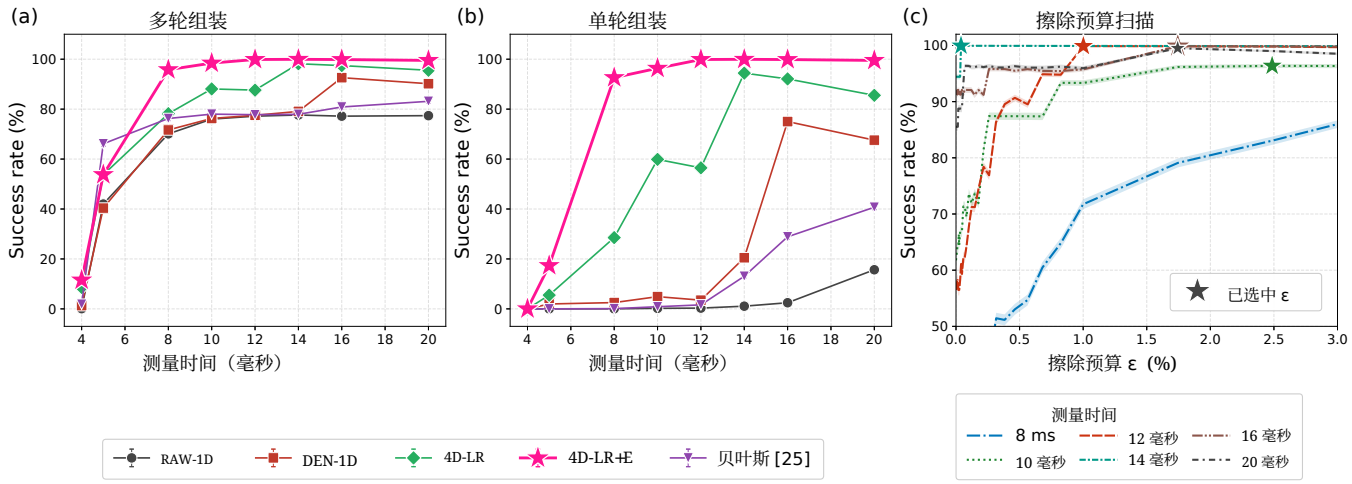


图 6. **置信度感知读出和测量擦除可改善原子重排**。(a) 多轮成功率随测量时间 t_{img} 的变化，最小目标大小为 20×20。模拟从装载概率为 $p_{\mathrm{load}} = 0.5$ 的 30×30 储备阵列开始。采用感知擦除的源选择的 4D-LR，在 $t_{\mathrm{img}} \geq 8$ 毫秒时实现了超过 90% 的成功率。(b) 固定 20×20 目标时，单轮成功率随测量时间的变化。基于 4D-LR 的方法是唯一能在 8–12 毫秒时获得显著单轮成功率的方法。在 10 毫秒时，成功率从未进行擦除筛选时的 59.87% 提升至进行擦除筛选后的 96.37%。(c) 固定 20×20 目标时，4D-LR 的单轮成功率随擦除预算的变化。对于每种曝光时间，星号标示在对最高 25% 的预算进行扫描时，成功率与所找到的最大值相差不超过一个蒙特卡洛标准误的最小擦除预算。图中显示的范围为 8–20 毫秒、0–3% 的擦除预算，以及至少 50% 的成功率。仅当选定点落在此范围内时才显示星号。采用基于置信度的源选择时，较短的读出曝光可优于较长且未进行源过滤的读出曝光：单轮 4D-LR 在 10 毫秒时进行擦除筛选，成功率达到 96.37%；而在 20 毫秒时不进行擦除筛选，成功率为 85.53%。

次，推理延迟可以通过多种方式降低。由于主要的计算组件是卷积去噪器，该流程非常适合进行硬件加速。近期研究已经展示了用于基于荧光的量子比特读出的现场可编程门阵列加速卷积神经网络（CNN）推理 [36]。采用面向张量架构的神经处理单元（NPUs） [37]也能够高效执行神经网络推理工作负载。利用整数量化和剪枝等标准神经网络压缩技术 [38–40]，还可以进一步减少内存流量和延迟。

我们的方法能够减少达到给定读出保真度所需的探测光子数。若以更高的速率收集这些光子，还能进一步加快成像。例如，在我们目前的装置中，50:50分束器会将收集到的一半荧光分流至qCMOS相机，用于同步参考成像。分类器经过训练和验证后，常规读出时便可移除分束器，使到达电子倍增电荷耦合器件的荧光光子通量大约翻倍。这样一来，只需约一半的曝光时间便可获得相近的光子计数，并有望达到相当的读出保真度。更高的数值孔径、更高的散射率或更有效的冷却等改进，也有助于在保持较高原子存活率的同时实现更快速的成像。

在量子纠错中，更短的读出时间可以减少测量期间数据量子比特上的空闲错误—时间 [10, 41], ，而基于置信度的擦除转换则可以将不可靠的测量结果标记为擦除，从而帮助解码 [34]。这些效应共同作用，可以降低逻辑错误率。对成像时间、测量保真度、擦除预算与测量轮次数量之间权衡关系的系统性研究，仍是未来工作的重要方向。

七、结论

我们展示了如何利用基于特征的判别分类器，对光镊中的单个 ^{171}Yb原子进行置信度感知的荧光读出。逻辑回归模型结合原始图像和去噪图像的光子计数、PSF模板相关性得分，以及基于背景拟合的相机噪声对数似然，直接从带标签的实验数据中学习占据/空置决策边界。这种方法保留了具有物理动机的特征，同时无需为荧光图像建立完整的类别条件似然模型，即可学习决策规则。

在所评估的数据集中，四特征分类器取得了高于传统光子计数阈值判定和所实现贝叶斯基线的占据读出保真度。在 10 毫秒的成像曝光时间下，其保真度达到 99.83%，与传统阈值判定在 20 毫秒时获得的 99.82% 相当。在这一工作点，该分类器因此实现了相当的

保真度，而所使用的曝光时间仅为阈值基线的一半。这一改进还伴随着假阳性判定的大幅减少。

除了提升二元分类性能外，分类器得分还提供了有关各个读出结果可靠性的有用信息。将低置信度结果标记为测量级擦除后，在相同擦除预算下，残留的未标记错误比阈值法和贝叶斯基线更少。对于原子重排，这些信息使得不确定的占据位点预测可以被排除在源池之外。基于测得的错误率进行的蒙特卡洛模拟预测，通过这种基于置信度的源选择，可以提高组装无缺陷阵列的成功率。这些结果表明，读出置信度以及假阳性与假阴性判定后果的不对等性，可用于原子重排等下游操作。

基于特征的表述具有良好的可扩展性。针对特定成像系统或下游任务，可以纳入并评估其他基于物理原理或统计学依据的描述符。这种灵活性使得随着新的、具有实验信息价值的特征不断被发现，可以进一步优化读出模型。因此，我们的框架为结合准确的占据状态判别与置信度引导的原子选择提供了实用基础，也为研究其在更大阵列和未来电路中途测量协议中的应用提供了基础。

致谢

我们感谢 Kangjin Kim、JuYoung Park、Sunhwa Hwang 和 Seungtaek Oh 提供的宝贵讨论。本研究得到了韩国国家研究基金会（NRF）资助，资助编号为 RS-2025-25464182、RS-2023-NR119928 和 RS-2023-00283259。

附录 A：数据准备与交叉验证协议

在监督评估中，我们使用 qCMOS 参考图像将每个位置标记为占据或空置。电子倍增电荷耦合器件的读出图像与 qCMOS 参考图像同时采集，曝光时间分别为 4–20 毫秒和 60 毫秒。若原子在曝光初期发出荧光，随后丢失或停止散射，其信号可能足以在短曝光读出图像中被检测到，但积分信号却不足以超过参考图像中判定位置为占据的阈值。我们标记出这样的帧：其中至少有一个位置在 qCMOS 参考图像中被标记为空置，但其电子倍增电荷耦合器件图像中的背景扣除后的感兴趣区域信号超过阈值。我们对标记的帧进行了目视检查，保留了被判定为电子倍增电荷耦合器件假阳性的帧，并将其余标记帧整体排除。此外，我们还排除了在 3×3 阵列的任何位置均未检测到原子的采集组，因为经验发现，完全空置的图像会降低去噪模型训练的效果。

表 A1 总结了采集帧数、筛选后剩余的测试帧数，以及筛选后的测试集中的原子加载比例。加载比例的计算方法是：将有原子的位点数除以筛选后的测试集中分析的位点总数。报告正文中定义的先验加权读出保真度时，也使用了这些加载比例。

我们使用五折交叉验证，在相同的滤波帧上评估所有方法。每次划分中，我们使用其中四折来拟合读出方法及其预处理参数，然后将其应用于第五折。所有折中均固定使用 Noise2NoiseFlow 去噪器。我们根据留出集预测结果计算所有分类指标。

表 A1. **数据划分、筛选与加载比例**。 总帧数表示每个测量时间采集的电子倍增电荷耦合器件读出图像数量。数据按 6:2:2 的比例划分为训练集、验证集和测试集。筛选后的测试帧表示排除完全为空的采集组，以及读出图像与参考图像中原子出现情况标签不一致的样本后剩余的测试帧数。加载比率根据筛选后的测试集计算，即有原子的位点占所有分析位点的比例。由于每个测量时间条件都是在独立的实验运行中采集的，因此加载比率会有所不同。

Exp. (ms)	总计帧	Test 帧	已筛选测试帧	正在加载比率（%）
4	30,000	5,357	5,284	80.28
5	13,770	2,458	2,444	37.61
8	15,952	3,190	3,024	56.78
10	15,986	3,197	3,084	59.09
12	16,000	3,200	3,073	71.37
14	16,000	3,200	3,099	73.99
16	15,994	3,198	3,128	66.35
20	15,996	3,199	3,098	60.72

附录B：读出基线与评估指标

1. 阈值基线

作为传统基线，我们采用一种基于感兴趣区域积分信号 $x = \sum_{(i,\ j)\ \in 感兴趣区域} c_{ij}$ 的阈值判定方法，其中 c_{ij} 表示感兴趣区域坐标（i, j）处的增益校准像素值。我们根据真实标签样本对无原子和有原子感兴趣区域的信号分布进行建模：无原子分布用一个高斯峰表示，而全曝光时间下的有原子信号分布 $P(x \mid B)$ 则用一个均值和方差分别估计的高斯峰表示。通过将初始占据位点的信号分布表示为存活-损失混合分布，将成像期间的原子损失纳入考量，

$$P(x \mid B_{\mathrm{init}}) = s\,P(x \mid B) + (1-s)\,P(x \mid L), \tag{B1}$$

其中， $P(x \mid L)$ 是指数衰减尾部，用于描述在原子丢失前仅发出完整曝光信号一部分的原子。存活概率的一阶近似为

$$s = 1 - r_{\mathrm{loss}} t_{\mathrm{exp}}, \qquad r_{\mathrm{loss}} = 0.019\%\ \mathrm{ms}^{-1}. \tag{B2}$$

对于每种测量时间条件 ，选择分类阈值 x_{th}，使先验加权保真度最大化 ，

$$\begin{aligned}\mathcal{F}(x_{\mathrm{th}}) = 1 - \Big[&(1-f)\,P(x \ge x_{\mathrm{th}} \mid D) \\ &+ f\,P(x < x_{\mathrm{th}} \mid B_{\mathrm{init}})\Big].\end{aligned} \tag{B3}$$

其中， f 是根据训练数据得出的经验原子装载率。

2. 贝叶斯基线

作为一种基于物理的基线方法，我们遵循 Martinez-Dorantes 等人 [25]实现了一个贝叶斯读出模型。给定观测信号值 c 和占据类别 $S \in \{B, D\}$，根据贝叶斯定理可得到后验概率，

$$P^{\mathrm{post}}(S \mid c) = \frac{P(c \mid S)\,P^{\mathrm{pri}}(S)}{\sum_{S'} P(c \mid S')\,P^{\mathrm{pri}}(S')}. \tag{B4}$$

在我们的实现中，类别条件似然 $P(c \mid S)$明确纳入了依赖类别的光子发射统计特性，包括泄漏、泊松时钟诱导电荷（CIC）、随机电子倍增增益放大以及高斯读出噪声。具体而言，总电子计数分布被建模为类别相关光子分布与泊松时钟诱导电荷过程的卷积，而得到的相机响应则建模为埃尔朗放大分布与高斯读出噪声的卷积。成像期间的原子损失通过有原子似然中的存活-损失混合分布加以纳入，并使用测得的损失率 γ_{loss} 对所有 ROI 像素共享的荧光持续时间进行建模。在对共同损失时间进行积分之前，先在给定该持续时间的条件下将像素似然相乘。相机噪声参数保持为其校准值。先验概率$P^{\mathrm{pri}}(S)$根据每个交叉验证划分的训练分区进行估计。具体而言，将 $P_{\mathrm{pri}}(B)$设为训练标签中的有原子比例，并将 $P_{\mathrm{pri}}(D) = 1 - P^{\mathrm{pri}}(B)$。随后，在相应的留出测试折上评估贝叶斯后验时，这些先验保持不变。

与参考文献 [25], 中原有的逐列公式不同，我们的几何结构允许直接在每个位点周围固定的二维感兴趣区域内进行推断，因为最近邻间距远大于成像点扩散函数宽度。因此，我们在每个感兴趣区域内逐像素进行贝叶斯推断，并使用每个训练折中测得的有原子图像均值作为类点扩散函数二维信号模板。随后对所得后验对数概率应用阈值，以获得贝叶斯判别器。

3. 评估指标

在本研究中，读出保真度直接根据混淆矩阵与参考标签进行评估，而不是通过对观测到的计数分布进行参数拟合来估算。令 $f = N_{\mathrm{atom}}/N_{\mathrm{total}}$ 表示

原子加载比例，$F_1 = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$ 表示有原子读出保真度，$F_0 = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP})$ 表示无原子读出保真度。随后，我们将总保真度定义为

$$F = (1 - f)F_0 + fF_1, \tag{B5}$$

相应的不保真度为

$$\epsilon = 1 - F. \tag{B6}$$

其中 f 是加载比例。所有报告的分类指标均通过交叉验证评估。特别是，除非另有说明，阈值和监督分类器均在训练折上确定，然后在 5 折交叉验证中应用于测试折。采集帧中的全部九个站点均属于同一折。统计不确定性为逐点 95%百分位区间，通过对完整帧进行自助重采样并固定折外预测得到：保真度和错误率采用 5,000 次重采样，擦除曲线采用 3,000 次配对重采样。这里，$\mathrm{FPR} = \mathrm{FP}/N_{\mathrm{empty}}$，$\mathrm{FNR} = \mathrm{FN}/N_{\mathrm{atom}}$，且 $F = 1 - (\mathrm{FP} + \mathrm{FN})/N_{\mathrm{total}}$。

附录 C：详细读出指标

为求完整，表 C1 和表 C2 分别报告了所有方法在整个测量时间范围内的读出保真度以及相应的假阳性率和假阴性率。这些数值结果补充了图 4 所示的趋势，并支持正文中的比较。

在本研究采用的固定决策阈值下，2D-LR 在所有测量时间内的读出保真度均高于 RAW-1D 和贝叶斯基线。在快速读出区间（$t_{\mathrm{img}} \leq 10$ 毫秒）内，提升最为显著；此时有原子和不含原子的感兴趣区域的信号分布高度重叠，生成模型中的残余失配预计会产生最强的影响。

对于 DEN-1D，在图像去噪后采用与 RAW-1D 相同的一维阈值判定流程。DEN-1D 达到了与贝叶斯基线相当的保真度，这表明图像域噪声抑制能够恢复通过显式相机噪声建模所获得的大部分增益。不过，其误差分布有所不同，因为去噪主要抑制了假阳性。例如，在 4 毫秒时，假阳性率从 RAW-1D 的 12.56% 降至 DEN-1D 的 9.14%。

完整的四维特征集进一步改善了决策边界。在快速成像区间，保真度的绝对增益最大；此时分类错误主要由有原子与无原子信号分布之间的重叠所致。对于 $t_{\mathrm{img}} > 10$ 毫秒，基线保真度已接近 1，提升空间有限。尽管如此，从 12 毫秒起，4D-LR 的保真度均超过 99.9%。其在 10 毫秒时达到的 99.83% 保真度，与 RAW-1D 在 20 毫秒时达到的 99.82% 相当，这支持了正文中讨论的测量时间缩短。

贝叶斯基线倾向于将模糊的位点分类为占据状态，从而抑制假阴性，但代价是假阳性率相对较高。这一特性在 4 ms 时尤为明显：其假阴性率为 0.94%，假阳性率则为 21.26%。相比之下，4D-LR 将假阳性率降至 9.06%，同时将假阴性率保持在 2.44%。假阳性率相应降低了 12.20 个百分点，而假阴性率仅增加了 1.50 个百分点。在整个测量时间范围内，4D-LR 的假阳性率始终低于贝叶斯基线，而假阴性率始终高于贝叶斯基线（表 C2）。

表 C1. **读出保真度**。 表中数值为保真度 [95% CI], , 单位为百分比。区间基于固定折外预测的 5,000 个整帧自助重采样样本计算。粗体数值表示各测量时间下的最高保真度。RAW-1D：对原始图像进行传统阈值判定；DEN-1D：对去噪图像进行传统阈值判定；贝叶斯：改编自 Martinez-Dorantes 的贝叶斯基线 *et al.* [25]。2D-LR：使用原始图像特征的二维逻辑回归；4D-LR：四维逻辑回归；

测量 (ms)	Mean 光子	RAW-1D 保真度（%）	DEN-1D 保真度（%）	贝叶斯 [25] 保真度（%）	2D-LR 保真度（%）	4D-LR 保真度（%）
4	4.45	93.68 [93.46, 93.90]	94.97 [94.76, 95.16]	95.05 [94.85, 95.25]	95.47 [95.29, 95.66]	**96.25** [96.08, 96.42]
5	6.44	97.26 [97.04, 97.47]	97.99 [97.81, 98.17]	98.01 [97.82, 98.19]	98.20 [98.02, 98.37]	**98.60** [98.44, 98.75]
8	8.39	98.65 [98.51, 98.78]	99.16 [99.04, 99.27]	99.17 [99.06, 99.28]	99.36 [99.27, 99.45]	**99.60** [99.51, 99.67]
10	10.40	99.24 [99.13, 99.34]	99.58 [99.51, 99.66]	99.47 [99.39, 99.56]	99.72 [99.65, 99.78]	**99.83** [99.78, 99.87]
12	12.33	99.53 [99.45, 99.61]	99.73 [99.67, 99.79]	99.70 [99.63, 99.76]	99.90 [99.86, 99.93]	**99.92** [99.88, 99.95]
14	14.36	99.69 [99.62, 99.76]	99.89 [99.85, 99.93]	99.87 [99.82, 99.91]	99.94 [99.90, 99.96]	**99.97** [99.95, 99.99]
16	16.67	99.69 [99.63, 99.75]	99.96 [99.94, 99.99]	99.89 [99.85, 99.93]	99.94 [99.91, 99.97]	**99.97** [99.94, 99.99]
20	20.77	99.82 [99.77, 99.86]	99.96 [99.94, 99.98]	99.91 [99.87, 99.94]	99.98 [99.96, 99.99]	**99.98** [99.96, 100.00]

表 C2. **假阳性率和假阴性率**。 FPR 和 FNR 以百分比表示。RAW-1D：对原始图像进行传统阈值判定；DEN-1D：对去噪图像进行传统阈值判定；贝叶斯：改编自 Martinez-Dorantes *et al.* [25] 的贝叶斯基线。2D-LR：使用原始图像特征的二维逻辑回归；4D-LR：四维逻辑回归；

测量 (ms)	Mean 光子	RAW-1D 假阳性率 (%)	RAW-1D 假阴性率 (%)	DEN-1D 假阳性率 (%)	DEN-1D 假阴性率 (%)	贝叶斯 [25] 假阳性率（%）	贝叶斯 [25] 假阴性率（%）	2D-LR 假阳性率（%）	2D-LR 假阴性率（%）	4D-LR 假阳性率（%）	4D-LR 假阴性率（%）
4	4.45	12.56	4.78	9.14	4.02	21.26	0.94	11.62	2.78	9.06	2.44
5	6.44	1.97	4.01	0.93	3.80	2.71	0.80	1.01	3.11	0.70	2.57
8	8.39	1.98	0.87	0.91	0.79	1.72	0.16	0.54	0.72	0.31	0.47
10	10.40	1.52	0.23	0.74	0.20	1.22	0.05	0.25	0.30	0.12	0.21
12	12.33	1.49	0.06	0.85	0.04	1.04	0.01	0.13	0.10	0.14	0.06
14	14.36	1.14	0.01	0.39	0.01	0.51	0.00	0.07	0.06	0.01	0.03
16	16.67	0.92	0.00	0.07	0.02	0.32	0.00	0.05	0.06	0.02	0.04
20	20.77	0.47	0.00	0.10	0.00	0.23	0.00	0.04	0.01	0.04	0.01

附录 D：4D-LR 消融研究

1. 选定的四个特征的消融

为确定 4D-LR 分类器中各项特征的作用，我们进行了两项互补的消融分析。首先，我们进行了逐一剔除特征研究，即从完整的 4D-LR 分类器中逐一移除各项特征。这项互补分析量化了在考虑所有特征相关性后，最终分类器对每项特征的依赖程度。其次，我们进行了特征逐步累加研究：从原始光子计数特征开始，依次加入背景拟合的 EMCCD 对数似然、去噪光子计数特征和 PSF 模板匹配得分。该分析展示了每增加一项特征，与已有特征结合时如何提升分类器的性能。

正文中的表 I 总结了报告的在 $t_{\mathrm{img}} = 8$ 毫秒时保真度和假阳性率的变化，表 D1 则提供了各个测量时间下的结果，包括假阴性率的变化。留一特征分析量化了每个物理或人工智能衍生特征对完整分类器的贡献。对于 $t_{\mathrm{img}} \leq 10$ 毫秒的测量时间，相机噪声对数似然通常对保真度的影响最大，而去噪强度在假阳性抑制中发挥着明确作用。移除 PSF 模板也会导致可测量的保真度损失，尽管其影响通常小于相机噪声似然的影响。由于特征逐步构建结果取决于添加特征的顺序，我们将该分析视为一种可解释性诊断。

表 D1 确认，没有任何单一特征能够单独解释全部性能提升：移除 EMCCD 对数似然通常会在 $t_{\mathrm{img}} \leq 10$ 毫秒时造成最大程度的性能下降。这些结果支持这样一种解释：4D-LR 通过结合光子数、相机输出似然、去噪强度和经验空间结构中的互补信息来改善读出。表 D2 显示，相较于仅使用原始计数的逻辑回归，EMCCD 对数似然带来了最大的初始改进，尤其是在低光子区间。去噪光子计数特征进一步降低了假阳性率，而 PSF 模板匹配得分则带来了较小的最终改进。

表 D1。**4D-LR 分类器的逐一特征剔除消融实验**。每个条目报告从完整 4D-LR 模型中移除一个特征后的性能下降情况。 $\Delta F = F_{\mathrm{full}} - F_{\mathrm{drop}}$、 Δ假阳性率 $=$ 假阳性率$_{\mathrm{drop}}$ $-$ 假阳性率$_{\mathrm{full}}$，以及 Δ假阴性率 $=$ 假阴性率$_{\mathrm{drop}}$ $-$假阴性率$_{\mathrm{full}}$。所有数值均以百分点表示。

Exp.	移除原始振幅			移除去噪振幅			丢弃电子倍增电荷耦合器件似然			移除 PSF 模板		
(ms)	ΔF	ΔFPR	ΔFNR	ΔF	ΔFPR	ΔFNR	ΔF	ΔFPR	ΔFNR	ΔF	ΔFPR	ΔFNR
4	0.18	0.30	0.15	0.11	0.54	0.01	0.26	0.62	0.18	0.18	0.54	0.09
5	0.11	0.03	0.24	0.05	0.17	-0.16	0.25	0.06	0.56	0.17	0.05	0.37
8	0.07	0.03	0.10	0.06	0.06	0.06	0.14	0.05	0.21	0.06	0.01	0.10
10	0.01	-0.02	0.02	0.05	0.04	0.06	0.04	0.01	0.07	0.02	0.00	0.04
12	0.00	0.01	0.00	0.01	0.01	0.01	0.03	0.06	0.02	0.02	0.01	0.02
14	0.00	0.00	0.00	0.02	0.06	0.00	0.01	0.00	0.01	0.00	0.00	0.00
16	0.00	0.00	-0.01	0.00	0.03	-0.02	0.01	0.00	0.02	0.01	0.00	0.01
20	0.00	0.00	0.01	0.01	0.02	0.00	0.01	0.00	0.02	0.00	-0.01	0.01

2. 其他候选特征

我们还测试了第三节中提到的另外三个候选描述符组。这里我们定义它们的具体实现形式，并通过分别将每个组加入选定的四特征分类器来评估各组。这一测试考察的是其边际贡献，而不是其单独使用时的预测能力。令 $R(p) = I^{\mathrm{raw}}(p) - B^{\mathrm{raw}}(p)$ 表示以模拟数字单元（ADU）为单位的背景扣除后的感兴趣区域，同时省略位置和拍摄索引。令 $\mathcal{C}$ 表示其中心 3×3 像素，$\mathcal{O} = \mathcal{R} \setminus \mathcal{C}$ 表示该 5×5 感兴趣区域外围的16个像素。

a. 中心集中度。 中心像素中正的背景扣除信号所占的比率为

$$c_{\mathrm{center}} = \frac{\sum_{p \in \mathcal{C}} R_+(p)}{\sum_{p \in \mathcal{R}} R_+(p)}, \qquad R_+(p) = \max[R(p), 0]. \tag{D1}$$

截去负值可以避免在低信号感兴趣区域中发生抵消。该标量描述空间集中度。

b. 学习噪声评分。 第二个候选值采用训练好的 Noise2NoiseFlow$\widetilde{\ }$ $\widetilde{\ }$噪声模型的潜在表示。原始图像和平均背景使用以模拟数字单元为单位的偏移量和尺度，分别归一化为 I 和 B 。

表 D2。**4D-LR 分类器的特征逐步构建消融**。该分类器从原始光子计数特征开始构建，并依次添加背景拟合相机噪声对数似然、去噪光子计数和 PSF 模板匹配得分。数值以百分比表示。括号中的数字表示相对于上一步构建的变化，单位为百分点。对于保真度，正变化表示有所提升；对于假阳性率和假阴性率，负变化表示错误率降低。最后一步对应完整的 4D-LR 分类器。

Exp. (ms)	特征集	保真度（%）	假阳性率（%）	假阴性率（%）
4	原始振幅	93.99	15.71	3.63
4	+EMCCD LL	95.47（+1.48）	11.62（-4.08）	2.78（-0.85）
4	+去噪振幅	96.07（+0.60）	9.61 (-2.02)	2.53 (-0.25)
4	+ PSF 模板	96.25 (+0.18)	9.06 (-0.54)	2.44 (-0.09)
5	原始振幅	97.14	1.67	4.84
5	+EMCCD LL	98.20 (+1.06)	1.01 (-0.66)	3.11 (-1.73)
5	+去噪振幅	98.42 (+0.22)	0.75 (-0.26)	2.95 (-0.16)
5	+ PSF 模板	98.60 (+0.17)	0.70 (-0.05)	2.57 (-0.37)
8	原始振幅	98.76	1.09	1.35
8	+EMCCD LL	99.36 (+0.60)	0.54 (-0.55)	0.72 (-0.63)
8	+去噪振幅	99.54 (+0.18)	0.32 (-0.21)	0.57 (-0.15)
8	+ PSF 模板	99.60 (+0.06)	0.31 (-0.01)	0.47 (-0.10)
10	原始振幅	99.44	0.52	0.59
10	+EMCCD LL	99.72 (+0.28)	0.25 (-0.27)	0.30 (-0.29)
10	+去噪振幅	99.81 (+0.09)	0.12 (-0.12)	0.24 (-0.06)
10	+ PSF 模板	99.83 (+0.02)	0.12 (0.00)	0.21 (-0.04)
12	原始振幅	99.74	0.43	0.20
12	+EMCCD LL	99.90 (+0.16)	0.13 (-0.30)	0.10 (-0.10)
12	+去噪振幅	99.90 (0.00)	0.15 (+0.03)	0.08 (-0.02)
12	+ PSF 模板	99.92 (+0.02)	0.14 (-0.01)	0.06 (-0.02)
14	原始振幅	99.88	0.21	0.09
14	+EMCCD LL	99.94（+0.06）	0.07（-0.14）	0.06（-0.03）
14	+去噪振幅	99.97（+0.03）	0.01（-0.06）	0.04 (-0.02)
14	+ PSF 模板	99.97 (0.00)	0.01 (0.00)	0.03 (0.00)
16	原始振幅	99.91	0.04	0.11
16	+EMCCD LL	99.94 (+0.03)	0.05 (+0.01)	0.06 (-0.05)
16	+去噪振幅	99.96 (+0.02)	0.02 (-0.03)	0.05 (-0.01)
16	+ PSF 模板	99.97 (+0.01)	0.02 (0.00)	0.04 (-0.01)
20	原始振幅	99.95	0.05	0.05
20	+EMCCD LL	99.98（+0.03）	0.04（-0.01）	0.01（-0.04）
20	+去噪振幅	99.98 (0.00)	0.03 (-0.01)	0.01 (0.00)
20	+ PSF 模板	99.98 (0.00)	0.04 (+0.01)	0.01（-0.01）

$\tilde{}$ $\tilde{}$归一化图像被复制到已保存噪声模型使用的两个输入通道 c 中。将 $I - B$输入其前向流，可得到潜在值 z_c（p）。ROI 分数为

$$a_{\mathrm{NF}} = -\frac{1}{2|\mathcal{R}|} \sum_{p\in\mathcal{R}} \sum_{c=1}^{2} z_c(p)^2. \tag{D2}$$

分数越高，表示与学习到的背景噪声的兼容性越高。

c. ***PSF****残差结构。* 使用与 PSF 相关特征相同的非负、和为一的经验模板 $T(p)$，我们通过 $R(p) \simeq AT(p)$ 近似背景扣除后的感兴趣区域，在保持模板形状不变的情况下，仅拟合其振幅。我们通过非加权最小二乘法选择无约束振幅，

$$\widehat{A} = \arg\min_{A\in\mathbb{R}} \sum_{p\in\mathcal{R}} [R(p) - AT(p)]^2.$$

$\widehat{}$令关于 A 的导数为零，得到 $\sum_{p\in\mathcal{R}} T(p)[R(p) - AT(p)] = 0$，从而得出拟合振幅和残差，

$$\widehat{A} = \frac{\sum_{p\in\mathcal{R}} R(p)T(p)}{\sum_{p\in\mathcal{R}} T(p)^2}, \qquad r(p) = R(p) - \widehat{A}T(p). \tag{D3}$$

残差组由三个标量组成，

$$E_{\text{center}} = \sum_{p \in \mathcal{C}} r(p)^2, \tag{D4}$$

$$E_{\text{outer}} = \sum_{p \in \mathcal{O}} r(p)^2, \tag{D5}$$

$$q_{\text{res}} = \frac{\sum_{p \in \mathcal{C}} r(p) - \sum_{p \in \mathcal{O}} r(p)}{\sqrt{\sum_{p \in \mathcal{R}} R(p)^2}}, \tag{D6}$$

这些量描述了中心像素和外围像素中的残差能量，以及带符号的中心与外围不平衡。

表 D3 报告了相对于 4D-LR 的匹配变化。观测到的最大保真度增幅为 0.025 个百分点，这是在 4 毫秒时加入 PSF 残差结构所带来的。所有报告的保真度变化 95% 区间均包含零，而且没有任何候选特征能在不同测量时间下带来一致的提升。这些结果支持当前线性分类器继续采用四特征表示，但并不排除其他分类器或数据集可能从中受益。

表 D3. **添加到所选 4D-LR 分类器中的其他候选特征**。 每行仅在相同的四项特征基础上，添加指定的标量或描述符组。变化定义为增强后的性能减去基线性能：$F_{\text{aug}} - F_{\text{4D}}$ Δ = aug − 4D Δ = aug − 4D ΔF、假阳性率 FPR FPR FPR 和假阴性率 FNR FNR FNR。所有数值均以百分点表示；正值表示性能提升，而比率变化为正则表示错误增加。方括号给出 ΔF 的配对 95% 百分位置信区间，采用 5,000 次帧自助抽样，两种分类器使用相同的抽样帧。在全部八个测量时间点，基线折外判定均与主要分析中使用的判定完全一致。

实验（毫秒）	新增候选特征	ΔF [95% CI]	ΔFPR	ΔFNR
4	中心集中度	0.000 [-0.050, +0.050]	+0.117	-0.029
4	学习噪声评分	+0.006 [-0.013, +0.025]	-0.021	-0.003
4	PSF残差结构	+0.025 [-0.011, +0.061]	-0.032	-0.024
5	中心集中度	+0.014 [-0.032, +0.059]	+0.022	-0.073
5	学习噪声评分	0.000 [-0.045, +0.050]	+0.022	-0.036
5	PSF残差结构	-0.005 [-0.036, +0.023]	-0.007	+0.024
8	中心集中度	-0.004 [-0.029, +0.018]	-0.009	+0.013
8	学习噪声评分	+0.004 [-0.007, +0.018]	+0.009	-0.013
8	PSF残差结构	+0.015 [-0.004, +0.037]	-0.017	-0.013
10	中心集中度	-0.004 [-0.018, +0.007]	0.000	+0.006
10	学习噪声评分	+0.004 [0.000, +0.011]	0.000	-0.006
10	PSF残差结构	+0.004 [-0.007, +0.014]	-0.009	0.000
12	中心集中度	0.000 [0.000, 0.000]	0.000	0.000
12	学习噪声评分	-0.007 [-0.018, 0.000]	+0.013	+0.005
12	PSF残差结构	0.000 [-0.022, +0.018]	+0.013	-0.005
14	中心集中度	0.000 [0.000, 0.000]	0.000	0.000
14	学习噪声评分	0.000 [0.000, 0.000]	0.000	0.000
14	PSF残差结构	0.000 [0.000, 0.000]	0.000	0.000
16	中心集中度	+0.004 [0.000, +0.011]	0.000	-0.005
16	学习噪声评分	+0.004 [0.000, +0.011]	0.000	-0.005
16	PSF残差结构	-0.004 [-0.011, 0.000]	0.000	+0.005
20	中心集中度	-0.004 [-0.011, 0.000]	0.000	+0.006
20	学习噪声评分	0.000 [0.000, 0.000]	0.000	0.000
20	PSF残差结构	-0.004 [-0.011, 0.000]	+0.009	0.000

附录 E：去噪架构

我们的去噪模型基于 Maleky等人 [26], 提出的 Noise2NoiseFlow 框架。该框架使用成对的含噪图像联合训练图像去噪器和噪声模型，无需任何干净图像进行训练。原始框架旨在适应不同相机传感器和 ISO 设置所带来的噪声特性变化。相比之下，我们的数据集仅包含使用同一台 Andor iXon Ultra 897 相机、在固定 EM 增益 300 下采集的单通道电子倍增电荷耦合器件荧光图像。因此，我们简化了原始架构，移除了相机/ISO条件调节和信号相关相机噪声层，同时保留了利用成对含噪观测进行学习的核心思想。

对于每种曝光时间，我们将连续采集的两张读出图像用作一对噪声图像。这些图像包含相同的名义原子构型，但相机噪声分布彼此独立。去噪器将一张噪声图像映射为去噪估计，而去噪估计与配对噪声观测之间的残差则由归一化流噪声模型评估。

去噪器本身是一个单通道 DnCNN [28]，包含 25 个卷积层，输入和输出均为单通道。原始 Noise2NoiseFlow 实现使用的是 9 层 DnCNN；我们测试了去噪器的深度，发现 25 层 DnCNN 在我们的电子倍增电荷耦合器件荧光数据上取得了最佳验证性能。

残差噪声模型采用一种简化的无条件流。该流架构如下：

$$\begin{aligned}&\mathrm{squeeze}\,|\,\mathrm{unc}\,|\,\mathrm{unc}\,|\,\mathrm{gain}\,|\,\mathrm{unc}\,|\,\mathrm{unc}\\&|\,\mathrm{gain}\,|\,\mathrm{unc}\,|\,\mathrm{unc}\,|\,\mathrm{unsqueeze}.\end{aligned}$$

挤压 操作将单通道图像重排为四个通道，从而使后续层能够进行通道拆分。每个无条件 块由一个可逆 1×1 卷积和一个无条件仿射耦合层组成。这些模块的作用是通过一系列可逆变量变换，将简单的高斯分布转换为更复杂的残差噪声分布。堆叠多个这样的模块，使流能够表示单个高斯模型无法捕捉的非高斯且不对称的 EMCCD残差统计特性。增益 层则调整所建模残差分布的整体尺度。

训练目标遵循 Noise2NoiseFlow 形式，

$$\mathcal{L}=\mathcal{L}_{\mathrm{nm}}(\tilde{I}_1,\tilde{I}_2)+\lambda\mathcal{L}_{\mathrm{dn}}(\tilde{I}_1,\tilde{I}_2),$$

其中，$\tilde{I}_1$ 和 $\tilde{I}_2$ 是连续获取的两张含噪原子图像。噪声模型项 $\mathcal{L}_{\mathrm{nm}}$ 是图像残差的负对数似然，其定义为含噪图像与对应去噪图像之差。去噪项 $\mathcal{L}_{\mathrm{dn}}$ 将根据一张含噪图像得到的去噪预测，与从相同底层场景中独立获取的另一张含噪图像进行比较。这种跨样本形式使去噪器无需无噪目标图像即可完成训练。系数 λ 控制去噪项的相对权重。我们采用了与参考文献 [26] 相同的损失权重 $\lambda=2^{18}$。模型使用 Adam 进行优化，采用 $\beta_1=0.9$、$\beta_2=0.999$、$\varepsilon=10^{-8}$，以及恒定学习率 10^{-3}。批次大小为 8。

每个曝光时间均独立进行训练。对于每个曝光时间，电子倍增电荷耦合器件的原始图像强度使用单独选择的 $[v_{\min},v_{\max}]$ 范围进行线性归一化至 $[0,1]$。训练使用 64×64 张灰度图像，每张图像包含一个图像块。训练至少进行 200 个 epoch，最多进行 500 个 epoch。每 10 个 epoch 执行一次完整验证；在第 200 个 epoch 之后，如果验证负对数似然在连续五次验证检查中均未改善超过 5×10^{-4}，则触发早停。

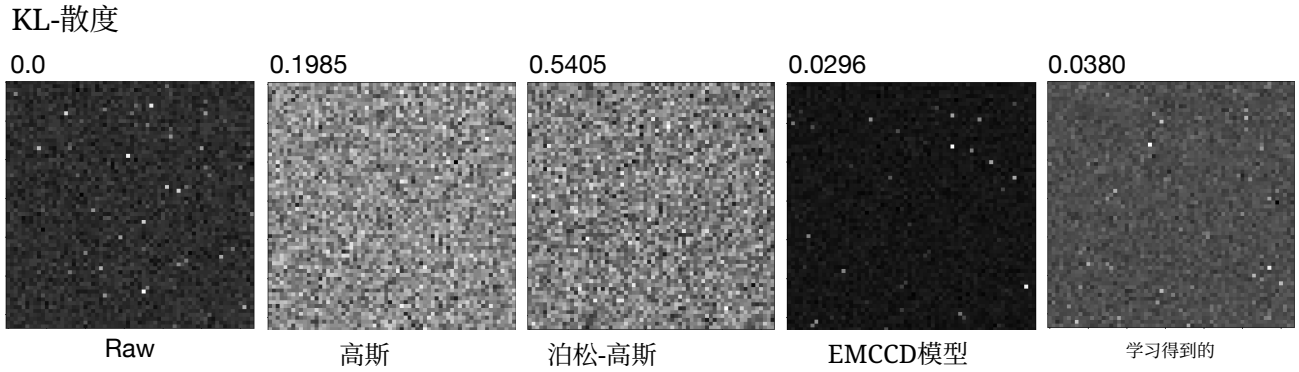
我们评估了所学习的噪声模型能否复现实测的 EMCCD背景统计。对于每个噪声模型，我们生成了合成背景图像，并将其像素值直方图与实测背景图像的直方图进行比较 [图 E1]。我们使用 KL散度对差异进行了量化

$$D_{\mathrm{KL}}(P_{\mathrm{bg}}\|P_{\mathrm{model}})=\sum_b P_{\mathrm{bg}}(b)\log\frac{P_{\mathrm{bg}}(b)}{P_{\mathrm{model}}(b)},\tag{E1}$$

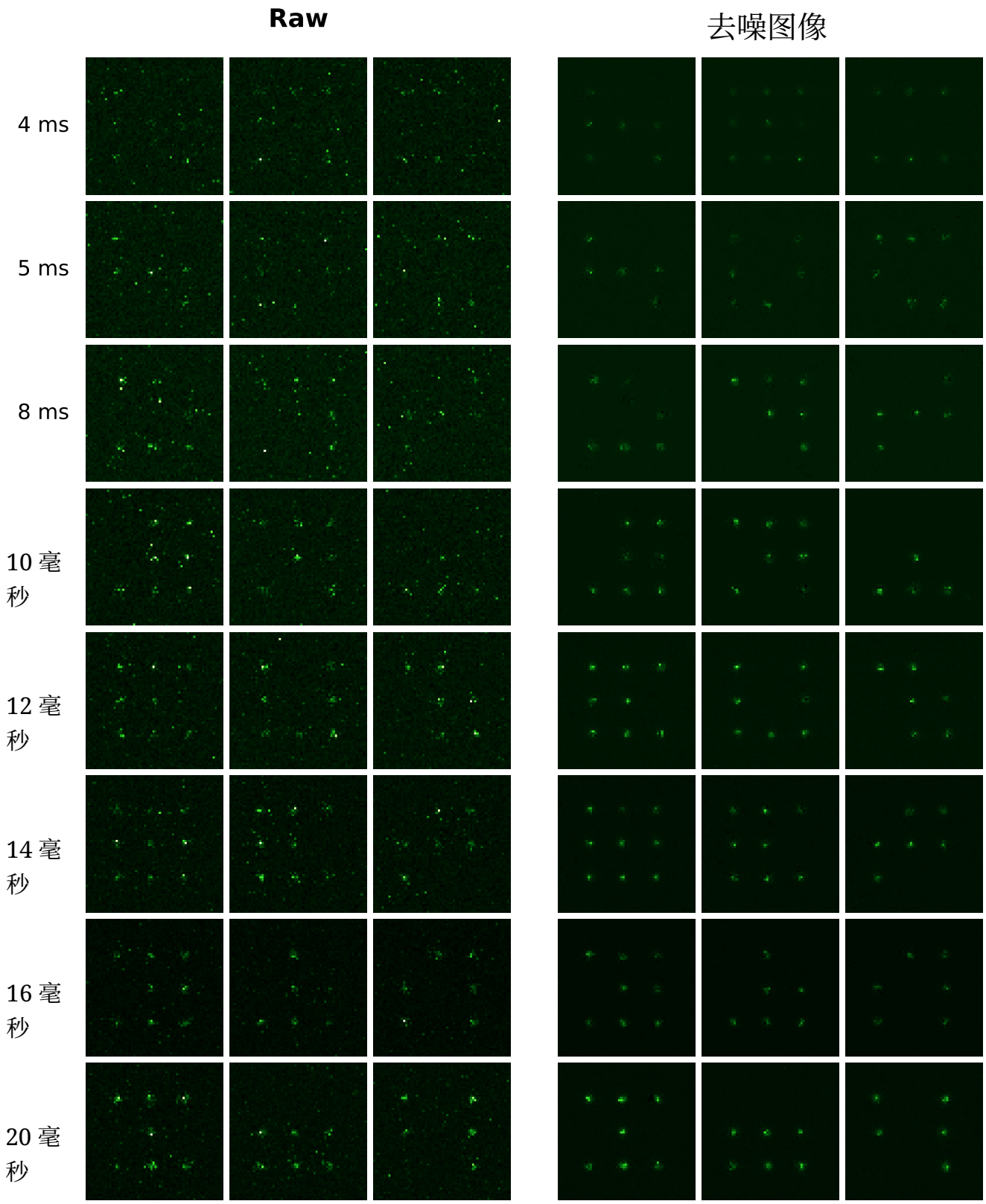
其中，P_{bg} 是经验背景像素值分布，P_{model} 是候选噪声模型生成的分布，b 为直方图分箱索引。D_{KL} 越低，表示与测得的 EMCCD 背景统计越吻合。图E1比较了高斯、泊松-高斯模型、解析 EMCCD [25], 以及学习得到的 Noise2NoiseFlow 背景模型。在所测试的模型中，解析 EMCCD 模型与测得的背景分布最为吻合（$D_{\mathrm{KL}}=0.0296$），而学习得到的 Noise2NoiseFlow 残差模型也达到了相近的数值（$D_{\mathrm{KL}}=0.0380$）。这表明学习得到的模型捕捉到了测得的 EMCCD 背景统计中的大部分特征。

如果去噪在空位点产生了类似原子荧光的光斑，该位点可能会被误判为占据状态。这种虚假的特征就是图像幻觉的一个例子。基于生成对抗网络（GAN）的图像复原方法可以通过匹配输出图像分布来提升感知真实感，但这一目标可能会以牺牲失真保真度为代价，还可能生成输入数据并不支持的结构 [42, 43]。当目标侧重于视觉上的合理性，而非与测量数据的统计一致性时，GAN 类或基于图像转换的去噪方法也存在类似问题。

所采用的去噪架构在这方面更加保守。DnCNN模块能够预测残差图像，而归一化流模块则对该残差的统计特性进行建模。因此，去噪后的残差被约束为与测得的相机噪声分布相符。如果去噪器生成或去除了类似原子的特征，相应的残差就会包含类似信号的成分，而非类似背景噪声的成分。因此，基于流的残差似然和背景统计验证可作为安全检查，防止出现此类伪影。与对抗性图像生成目标相比，DnCNN+Flow的形式更符合保守的科学去噪，因为它对噪声残差进行显式建模，并且可以通过测得的背景统计进行检验。图E2展示了各曝光时间下原始荧光图像及其对应的去噪图像示例。



图E1。**学习型与解析EMCCD背景噪声模型的比较**。将真实EMCCD背景帧与由四种噪声模型生成的代表性合成背景帧进行比较：高斯模型、泊松–高斯模型、解析EMCCD模型 [25], 以及学习型残差Noise2NoiseFlow [26] 模型。比较针对 $t_{\mathrm{exp}} = 8$ 毫秒背景数据进行。对于每种噪声模型，我们生成合成背景图像，并将其像素值直方图与实测背景图像的直方图进行比较。所有模型均使用相同的直方图分箱，因此KL散度仅衡量生成像素值统计量之间的差异。每幅图像上方的数值为 $D_{\mathrm{KL}}(P_{\mathrm{bg}} \| P_{\mathrm{model}})$，其中 P_{bg} 为实测背景直方图， P_{model} 为生成背景图像的直方图。真实背景帧作为视觉参考，其散度按定义为零。KL散度越低，表示与实测EMCCD背景统计的一致性越高。



图E2. **原始荧光图像和去噪荧光图像集**。 展示了所有测量时间下的原始图像和去噪图像样例： $t_{\mathrm{img}} = 4$、5、8、10、12、14、16和20毫秒。每行中，前三张图像为原始电子倍增电荷耦合器件图像，后三张为相应的去噪输出。

附录 F：重排模拟

本附录介绍用于评估读出错误对无缺陷原子阵列制备影响的蒙特卡洛重排模拟。该模拟重点关注源分配过程中假阳性与假阴性之间的不对称性。假阳性可能会将一个空的储备位点选为源原子，并直接造成目标缺陷；而假阴性主要会减少可用的原子池。因此，我们采用一种简化模型，仅分析由读出引起的源分配错误，而不对特定重排算法的详细传输轨迹、碰撞约束或运动控制误差进行建模 [44, 45]。

每次试验均从一个具有独立装载概率 p_{load} 的 $M \times M$ 储备阵列开始。对于给定的读出方法和曝光时间，首先使用 $P_{\mathrm{surv}}(t_{\mathrm{img}})$ 对成像存活率进行抽样。测得的假阴性率p_{FN} 应用于存活原子，假阳性率 p_{FP} 应用于成像前为空的位点。随后，重排算法根据测得的占据位点选择源原子，并尝试填充一个方形目标阵列。

对于单轮重排，目标固定为 20×20，且仅允许一个读出与移动周期。对于多轮重排，第 k 轮的目标边长为 $L_k = \lfloor \sqrt{B} \rfloor$，其中 B 是经过擦除筛选后保留的检测到的候选原子数量。只有当组装后的阵列无缺陷且边长至少为 $L_{\mathrm{floor}} = 20$，并且在不超过 20 轮内完成时，才将该次试验计为成功。某一轮失败后，下一轮仅使用实际位于尝试构建的目标区域内的原子进行初始化。储备区中剩余的所有未使用原子均会丢弃，不再重复使用。

对于感知擦除的源选择，在重排前，会从候选源池中移除低置信度占据位点预测。这可以抑制由假阳性导致的错误源分配，但也会丢弃一些实际被占据的原子。因此，对擦除预算进行扫描，可以衡量移除不可靠的源分配与保留足够大的原子池之间的权衡。模拟器的有效输入为

$$p_{\mathrm{FP}}^{\mathrm{eff}} = \frac{\mathrm{FP}_{\mathrm{kept}}}{N_{\mathrm{empty,total}}} = p_{\mathrm{FP}}(1 - p_{\mathrm{era}|F}), \tag{F1}$$

$$p_{\mathrm{FN}}^{\mathrm{eff}} = \frac{\mathrm{FN}_{\mathrm{kept}} + N_{\mathrm{erased,atom}}}{N_{\mathrm{atom,total}}} = 1 - (1 - p_{\mathrm{FN}})(1 - p_{\mathrm{era}|T}). \tag{F2}$$

此处，分母包括擦除筛选前所有标记为无原子或有原子的感兴趣区域结果。算法 1 展示了这种等效的两步抽稀过程，它使用未筛选的错误率和条件候选擦除概率。从保留的候选项中无放回地均匀选择源。每种条件均进行 10,000 次试验，缺陷容忍度为零。对于考虑擦除的方法，在同一轮扫描中，选择成功率最高值的一个蒙特卡洛标准误差范围内预算最低的方案，预算上限为 25%。每次模拟试验的成功与否，均根据重排后的已知模拟真实占据状态进行评估。这是模拟预言机判据，而非实验中实际实施的停止测量。

在模拟中，真阳性源候选取自成像后存活的原子，假阳性候选项则取自成像前为空的位点。我们假定，损失相关假阳性分配已包含在经验假阳性率中。因此，成像过程中损失的原子不会作为单独的假阳性源候选项加入。每轮内不进行额外的读出或重新装载。

算法 1蒙特卡洛重排模拟

要求：储存库大小 M、装载概率 p_{load}、读出错误率 p_{FP} 和 p_{FN}、条件擦除概率 $p_{\text{era}|T}$ 和 $p_{\text{era}|F}$、成像存活概率 P_{surv}、协议模式、最大轮数 K、固定目标边长 L、最小目标边长 L_{floor}和缺陷容忍度 τ。

其中，$p_{\text{era}|T}$ 是真阳性候选体被擦除的概率，$p_{\text{era}|F}$ 是假阳性候选体被擦除的概率。若不采用感知擦除的源选择，则这两个概率均设为零。

$A \sim$ 二项分布(M^2, p_{load})，$S \leftarrow M^2$ ▷ A：真实原子，S：可用位置
对于 $k = 1 \dots, K$ **do**
  $A_{\text{pre}} \leftarrow A$
  $A \leftarrow$ 二项分布$(A_{\text{pre}}, P_{\text{surv}})$ ▷成像后的存活原子
  $B_{\text{TP}} \leftarrow$ 二项分布$(A,\ 1 - p_{\text{FN}})$ ▷ 检测为占据的存活原子
  $B_{\text{FP}} \leftarrow$ 二项分布$(S - A_{\text{pre}},\ p_{\text{FP}})$ ▷ 最初为空、但被误检为占据的位点
  $B_{\text{TP}} \leftarrow$ 二项分布$(B_{\text{TP}},\ 1 - p_{\text{era}|T})$ ▷ 经擦除筛选后保留的真阳性候选项
  $B_{\text{FP}} \leftarrow$ 二项分布$(B_{\text{FP}},\ 1 - p_{\text{era}|F})$ ▷ 擦除筛选后保留的假阳性候选项
  $B \leftarrow B_{\text{TP}} + B_{\text{FP}}$
  如果协议模式为单轮**则**
    $L_k \leftarrow L$
    如果 $B < L^2$ **则**
      返回失败
    结束如果
  else
    $L_k \leftarrow \lfloor \sqrt{B} \rfloor$
    如果 $L_k < L_{\text{floor}}$ **，则**
      返回 失败
    结束如果
  结束如果
  $T \leftarrow$ 超几何分布$(B,\ B_{\text{TP}},\ L_k^2)$ ▷总体 B，真实候选者 B_{TP}，抽取次数 L_k^2
  如果 $(L_k^2 - T)/L_k^2 \le \tau$**，则**
    返回 成功
  else
    $A \leftarrow T$, $S \leftarrow L_k^2$ ▷ 保守清理：仅重复使用已尝试的目标区域
  结束如果
结束循环
返回失败

[1] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, 基于可重构原子阵列的逻辑量子处理器, 《自然》626, 58（2024）。[2] H. J. Manetsch, G. Nomura, E. Bataille, X. Lv, K. H. Leung 和 M. Endres, 具有 6,100 个高度相干原子量子比特的光镊阵列, 《自然》647, 60（2025）。[3] C. Cicali, M. Calzavara, E. Cuestas, T. Calarco, R. Zeier 和 F. Motzoi, 光镊之间的快速中性原子输运与转移, 《Phys. Rev. Applied》24, 024070（2025）。[4] S. Hwang, H. Hwang, K. Kim, A. Byun, K. Kim, S. Jeong, M. P. Soegianto 和 J. Ahn, 通过光镊实现快速且可靠的原子输运, 《Optica Quantum》3, 64（2025）。[5] M. Saffman, 使用原子量子比特阵列进行量子计算：直面连通性的代价（2025）, arXiv:2505.11218 [quant-ph]。[6] D. Bluvstein, A. A. Geim, S. H. Li, S. J. Evered, J. P. Bonilla Ataides, G. Baranes, A. Gu, T. Manovitz, M. Xu, M. Kalinowski, *et al.*, 用于通用量子计算的容错中性原子架构, 《自然》649, 39（2026）。[7] E. Deist, Y.-H. Lu, J. Ho, M. K. Pasha, J. Zeiher, Z. Yan 和 D. M. Stamper-Kurn, 中性原子阵列中的中途腔测量, 《Phys. Rev. Lett.》129, 203602（2022）。[8] Y. Li, Y. Bao, M. Peper, C. Li 和 J. D. Thompson, 中性原子量子比特阵列中的快速、连续且相干的原子替换（2025）, arXiv:2506.15633 [quant-ph]。[9] N.-C. Chiu, E. C. Trapp, J. Guo, M. H. Abobeih, L. M. Stewart, S. Hollerith, P. L. Stroganov, M. Kalinowski, A. A. Geim, S. J. Evered, *et al.*, 相干 3,000 量子比特系统的连续运行, 《自然》646, 1075（2025）。[10] S. Veroni, M. M¨uller 和 G. Giudice, 面向中性原子的优化无测量容错量子纠错, 《Phys. Rev. Research》6, 043253（2024）。[11] M. Cain, Q. Xu, R. King, L. R. B. Picard, H. Levine, M. Endres, J. Preskill, H.-Y. Huang 和 D. Bluvstein, Shor算法仅需 10,000 个可重构原子量子比特即可实现（2026）, arXiv:2603.28627 [quant-ph]。[12] C. Zhao, C. Duckering, A. Gu, N. Maskara 和 H. Zhou, 迈向使用可重构原子阵列的超高比率量子纠错（ 2026 2604 16209 ）, [ ] arXiv: . quant-ph .[13] R. Ismail, I.-C. Chen, C. Zhao, R. Weiss, F. Liu, H. Zhou, S.-T. Wang, A. Sornborger 和 M. Kornjaˇca, 使用中性原子实现百万量子比特规模量子模拟的横向架构, 《PRX Quantum》7, 020343（2026）。[14] C. Gidney 和 M. Ekerˆa, 如何利用 2,000 万个含噪量子比特在 8 小时内分解 2048 位 RSA 整数, 《Quantum》5, 433（2021）, arXiv:1905.09749 [quant-ph]。[15] H. Ali, J. Marques, O. Crawford, J. Majaniemi, M. Serra-Peralta, D. Byfield, B. Varbanov, B. M. Terhal, L. Di-Carlo 和 E. T. Campbell, 利用模拟读出信息降低超导逻辑量子比特的错误率, 《Phys. Rev. Applied》22, 044031（2024）。[16] M. D. Hanisch, B. Het´enyi 和 J. R. Wootton, 使用超导量子比特进行软信息解码, 美国物理学会 Open Science 1, 000019 (2026), arXiv:2411.16228 [quant- ph].[17] J. Majaniemi 和 E. S. Matekole, 利用软信息降低量子纠错开销, Quan- tum Sci. Technol. **11**, 025024 2026 2504 03504 ( ), arXiv: . [ ] quant-ph .[18] I. Ulusoy 和 C. M. Bishop, 目标识别中的生成式方法与判别式方法, 载于 *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2 2005 , 第 ( ) 卷, 第 258–265 页。[19] A. Y. Ng 和 M. I. Jordan, 关于判别式与生成式分类器：逻辑回归和朴素贝叶斯的比较, 载于 *Advances in Neural Information Processing Systems* 14 2001 841 848 , 第 ( ) 卷, 第 – 页。[20] R. Raina、Y. Shen、A. Y. Ng 和 A. McCallum, 使用混合生成式/判别式模型进行分类, 载于 *Advances in Neural Information Processing Systems*, 第 16 卷（2003）, 第 545–552 页。[21] R. O. Duda、P. E. Hart 和 D. G. Stork, *Pattern Classification* 2 2001 , 第版（Wiley, 纽约, ）。[22] C. N. Mude、L. Phuttitarn、S. Maurya、K. Sinha、M. Saffman 和 S. Tannu, 通过图像去噪实现快速、准确的中性原子读出（2025）, arXiv:2510.25982 [quant-ph]。[23] Y. Song、K. Kim、J. H. Han、S. Oh 和 J. Mun, 通过交替双音窄线冷却, 在浅钟魔术光镊中对 ^{171}Yb原子进行高性能成像（2026）, arXiv:2603.27498 [physics.atom-ph]。[24] A. Cooper、J. P. Covey、I. S. Madjarov、S. G. Porsev、M. S. Safronova 和 M. Endres, 光镊中的碱土原子, Phys. Rev. X 8, 041055 (2018)。[25] M. Martinez-Dorantes、W. Alt、J. Gallego、S. Ghosh、L. Ratschbacher、Y. V¨olzke 和 D. Meschede, 光学势中中性原子寄存器的快速无损并行读出, Phys. Rev. Lett. 119, 180503 (2017)。[26] A. Maleky、S. Kousha、M. S. Brown 和 M. A. Brubaker, Noise2NoiseFlow：无需干净图像的真实相机噪声建模, 载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), 第 17632–17641 页。[27] A. Abdelhamed 、 M. A. Brubaker 和 M. S. Brown, Noise flow：使用条件归一化流进行噪声建模, 载于 *Proceedings of the IEEE/CVF International Conference on Computer Vision* 2019 3165 3173 ( ), 第 – 页。[28] K. Zhang、W. Zuo、Y. Chen、D. Meng 和 L. Zhang, 超越高斯去噪器：用于图像去噪的深度卷积神经网络残差学习, IEEE Trans. Image Process. 26, 3142 (2017)。[29] D. J. Rezende 和 S. Mohamed, 使用归一化流的变分推断, 载于 *Proceedings of the 32nd International Conference on Machine Learning* , 《Machine Learning Research》论文集, 第 37 卷（PMLR, 2015）, 第 1530–1538 页。[30] D. P. Kingma 和 P. Dhariwal, Glow：带可逆 1×1 卷积的生成流, 载于 *Advances in Neural Information Processing Systems* 31 2018 , 第 ( ) 卷。[31] F. Zhao、Q. Huang 和 W. Gao, 基于归一化互相关的图像匹配, 载于 *Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing* 2 2006 729 732 , 第 II– –II– 页。

[32] A. G. Basden，电子倍增电荷耦合器件和科学级互补金属氧化物半导体读出噪声模型对夏克–哈特曼波前传感器精度的影响分析，J. As- tron. Telesc. Instrum. Syst. **1**, 039002 (2015)。[33] K. B. W. Harpsøe、M. I. Andersen 和 P. Kjægaard，使用电子倍增电荷耦合器件（EMCCD）的贝叶斯光子计数，Astron. Astrophys. **537**, A50 (2012)。[34] Y. Wu、S. Kolkowitz、S. Puri 和 J. D. Thompson，碱土金属里德伯原子阵列中用于容错量子计算的擦除转换，Nat. Commun. **13**, 4657 2022( )。[35] M. N. H. Chow、V. Buchemmavari、S. Omanakuttan、B. J. Little、S. Pandey、I. H. Deutsch 和 Y.-Y. Jau，中性原子量子处理器中基于电路的泄漏到擦除转换，《PRX Quantum》**5**, 040343 (2024)。[36] X. Su、M. Liang、T. Zhang、Z. Zhao、P. Yin、Y. Yang、W. Dai、X. Luo、X. Li 和 L. You，面向基于荧光的量子比特读出的现场可编程门阵列加速曝光感知卷积神经网络，Phys. Rev. Applied **25**, 054004 (2026)。[37] H. Kim、Y. Choi、J. Park、B. Bae、H. Jeong、S. M. Lee、J. Yeon、M. Kim、C. Park、B. Gu、 *et al.*，TCP：面向人工智能工作负载工业产品的张量收缩处理器，载于 *Proceedings of the 2024 ACM/IEEE 51st Annual International Symposium on Computer Architecture (ISCA)* 2024 890 902 (IEEE,)

第 – 页。[38] B. Jacob、S. Kligys、B. Chen、M. Zhu、M. Tang、A. Howard、H. Adam 和 D. Kalenichenko，用于高效纯整数算术推理的神经网络量化与训练，载于 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2018 2704 2713 ( )

第 – 页。[39] S. Han、H. Mao 和 W. J. Dally，《深度压缩：通过剪枝、训练量化和霍夫曼编码压缩深度神经网络》，载于 *International Conference on Learning Representations (ICLR)* (2016), arXiv:1510.00149。[40] S. K. Esser、J. L. McKinstry、D. Bablani、R. Appuswamy 和 D. S. Modha，《学习步长量化》，载于 *International Conference on Learning Representations (ICLR)* 2020 ( )。

[41] T. M. Graham、L. Phuttitarn、R. Chinnarasu、Y. Song、C. Poole、K. Jooya、J. Scott、A. Scott、P. Eichler 和 M. Saffman，《单一物种中性碱金属原子量子处理器上的电路中测量》，Phys. Rev. X **13**, 041051 2023( )。[42] Y. Blau 和 T. Michaeli，《感知–失真权衡》，载于 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), 第 6228–6237 页。[43] J. P. Cohen、M. Luck 和 S. Honari，《分布匹配损失可能在医学图像转换中产生虚构特征》，载于 *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* 2018 (Springer,)，第 529–536 页。[44] H. Kim、M. Kim、W. Lee 和 J. Ahn，《用于在光镊陷阱中快速高效地进行原子重排的Gerchberg-Saxton算法》，Opt. Express **27**, 2184 (2019)。[45] R. Lin、H.-S. Zhong、Y. Li、Z.-R. Zhao、L.-T. Zheng、T.-R. Hu、H.-M. Wu、Z. Wu、W.-J. Ma、Y. Gao、 *et al.*，《人工智能驱动的数千个无缺陷中性原子阵列并行组装》，Phys. Rev. Lett. **135**, 060602 (2025)。