

打破衍射极限：从阿贝成像理论到超分辨荧光显微

第 43 届决赛短训讲义组

2026 年 10 月

目录

1 导读	3
2 名词与符号	5
3 前置模块：波、空间频率与衍射	7
3.1 复振幅、平面波与空间频率	7
3.2 惠更斯-菲涅耳原理	8
3.3 夫琅禾费近似：远场衍射就是傅里叶变换	8
3.4 傅里叶分析：最少够用的部分	9
3.5 三个经典衍射图样	11
3.6 透镜把方向变成位置；数值孔径	13
4 阿贝成像理论与衍射极限	13
4.1 物体是许多光栅的叠加	13
4.2 光栅成像：两个衍射级才能成像	13
4.3 傅里叶光学的表述：成像是低通滤波	15
4.4 荧光是非相干成像：截止频率加倍	16
4.5 点扩散函数与瑞利判据	18
4.6 为什么“更亮的光”“更好的算法”都越不过截止频率	19
5 荧光、受激辐射与饱和	20
5.1 荧光分子的能级	20
5.2 自发辐射与受激辐射	20
5.3 速率方程	21
6 STED：用环形光把光斑“削”小	22

目录	2
6.1 原理	22
6.2 环形光：涡旋相位与中心零点	24
6.3 有效点扩散函数的宽度	25
6.4 需要多强的光？实际限制	27
7 单分子定位：测“中心”比测“宽度”准得多	27
7.1 前置：随机变量、方差与泊松分布	27
7.2 思想：让分子一个一个地亮	28
7.3 定位精度：光子统计	29
7.4 像素化与背景	29
7.5 图像分辨率：定位精度与标记密度	31
7.6 MINIFLUX：用“零点”定位	32
8 结构光照明：把高频“搬”进通带	33
8.1 莫尔条纹	33
8.2 结构光照明显微 (SIM)	33
8.3 非线性 SIM：利用饱和产生高次谐波	35
8.4 三类方法的比较	36
9 决赛风格例题	36
10 实验与数据处理：标定单分子定位精度	43
11 小结：推导链	46
A 圆孔衍射与 Airy 斑	50
B 非相干光学传递函数	51
C 涡旋光瞳的焦斑	51
D 概率与统计工具	52
E 采样定理与混叠	53
F SIM 的三相位分离与饱和谐波	53
G 最小二乘拟合公式	54

1 导读

用光学显微镜看一个细胞，能看清多小的东西？直觉的回答是“放大倍数越大越清楚”。19 世纪 70 年代，德国耶拿的物理学家阿贝（Ernst Abbe）指出这个直觉是错的：显微镜的分辨本领由光的波长和物镜能收集的光锥角度决定，放大倍数再大也只是把模糊的像放大。对可见光，这个极限约为 200 nm。一百多年里，这被看作光学显微镜不可逾越的“衍射极限”。

20 世纪 90 年代以来，物理学家找到了绕过它的办法。这些办法并没有推翻阿贝的结论，而是利用了阿贝理论成立的前提之一：成像过程是线性的，即像中每一点的亮度与物上各点的亮度成正比。只要让荧光分子的发光对照明光强度作非线性响应（例如饱和），或者让分子随机地、一个一个地亮起来，就能在远场光学显微镜中看到几十纳米甚至几纳米的结构。本讲沿着这条思路展开：先弄清衍射极限从哪里来、为什么在线性成像中不可逾越，再看三类超分辨方法分别怎样绕过它。

背景：阿贝、瑞利与“衍射极限”

1873 年，阿贝在《显微解剖学文献》上发表长文《显微镜理论与显微知觉》[1]。他指出：如果结构细到连第一束衍射光都不能与未衍射的光同时进入物镜，任何显微镜都无法把它分辨出来。文中用文字给出了结论：正入射照明时，可分辨的最小间距等于波长除以物镜半孔径角的正弦；最倾斜的照明时恰好减半；因此分辨极限“永远不会明显超过蓝光波长的一半”（原文第 19 节，第 455–456 页）。1873 年的论文里并没有公式。1881 年，阿贝在英国皇家显微学会杂志上明确定义了“数值孔径” $n \sin \alpha$ [2]，用它刻画浸没物镜收集光锥的大小。后人把阿贝的结论写成 $d = \lambda / (2n \sin \alpha)$ ，耶拿的阿贝纪念碑上刻的就是这个公式（图 1）。

另一条线索来自天文望远镜。1835 年，艾里（George Airy）算出圆形透镜所成的点像不是一个点，而是一个中央亮斑外围绕一圈圈暗环和亮环的图样（今称 Airy 斑）[3]。艾里的数值表给出：中央亮斑在 $v = 1.616$ 处降到一半，第一暗环在 $v = 3.832$ 处（ v 的定义见式 (3.11)）。1879 年，瑞利（Lord Rayleigh）在研究光谱仪的分辨本领时，讨论了两条谱线的像：当一条的中央极大与另一条的第一个零点重合时，两像之间中点的亮度是峰值的 81%，他认为这大约是“刚能看出分开”的界限[4]。把同一约定用到圆孔的 Airy 斑上，就是今天常说的瑞利判据，对显微镜给出最小可分辨距离 $0.61\lambda/\text{NA}$ （第 4.5 节）。

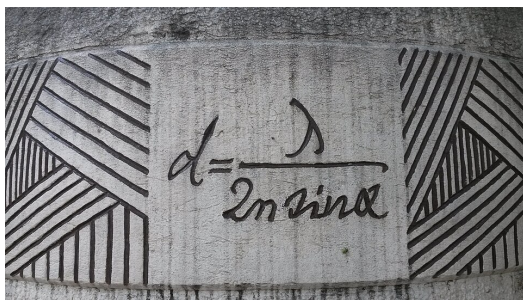


图 1 耶拿富尔斯滕格拉本街阿贝纪念碑上的公式 $d = \lambda / (2n \sin \alpha)$ 。（图源：Daniel Mietchen 摄，Wikimedia Commons, CC0）

背景：超分辨荧光显微：2014 年诺贝尔化学奖

2014 年诺贝尔化学奖授予 Eric Betzig、Stefan W. Hell 与 William E. Moerner，以表彰他们“发展了超分辨荧光显微技术” [5]。

- 受激发射损耗 (STED) 显微：Hell 与 Wichmann 于 1994 年提出 [6]：用一束耗尽光把焦斑外围的激发分子“强迫”退激发，只留下中心极小区域发荧光。1999 年 Klar 与 Hell 用一束偏置的耗尽光，在一个横向方向上把光斑从 150 nm 压到 106 nm [7]；2000 年用相位板产生中心为零的耗尽光，把轴向光斑从 490 nm 压到 97 nm，横向光斑同时缩小约一半 [8]；2009 年在金刚石色心上达到 5.8 nm [9]。
- 单分子定位显微：1989 年 Moerner 与 Kador 首次在固体中用光学方法探测到单个分子的吸收 [10]，1990 年 Orrit 与 Bernard 探测到单分子荧光 [11]；1997 年 Moerner 等人发现绿色荧光蛋白的突变体可以被光“开关” [12]。2006 年，Betzig 等人的 PALM (光激活定位显微) [13]、Hess 等人的 FPALM [14]、庄小威 (Xiaowei Zhuang) 小组的 STORM (随机光学重建显微) [15] 几乎同时发表：每次只让稀疏的少数分子发光，逐个精确定位，再把成千上万帧叠加成图像。
- 结构光照明显微 (SIM)：用条纹状照明把样品中超出截止频率的细节“搬”进物镜的通带。Heintzmann 与 Cremer [16]、Gustafsson [17] 分别在 1999、2000 年发表，线性 SIM 把分辨率提高到约 2 倍；Gustafsson 2005 年利用荧光饱和实现了非线性 SIM [18]。
- 2017 年，Hell 小组把“激发光的零点”与“单分子开关”结合，提出 MINFLUX：所需光子数比相机质心定位少约 22 倍，定位精度达到约 1 nm，分辨出相距 6 nm 的分子 [19]；2020 年在细胞中实现了 1–3 nm 的分辨率 [20]。

学习目标 学完本讲，你应当能够：

1. 用“空间频率”的语言描述光的衍射：把夫琅禾费衍射理解为孔径函数的傅里叶变换，算出单缝、光栅、圆孔的衍射图样，说明 Airy 斑第一暗环 $1.22\lambda/D$ 的来历；
2. 由阿贝成像理论推出最小可分辨周期 λ/NA (正入射) 与 $\lambda/(2NA)$ (斜照明)，理解成像是空间频率的低通滤波器，区分相干与非相干 (荧光) 成像的截止频率；
3. 说明为什么在线性成像中增大光强、改进算法都不能越过截止频率，并指出三条出路：非线性响应、关于样品的先验知识、照明对频率的搬移；
4. 由速率方程得到荧光饱和与受激辐射耗尽的规律，推出 STED 有效点扩散函数宽度 $w \approx w_0/\sqrt{1 + \beta I_{\max}/I_s}$ 并估算所需功率；
5. 由光子统计推出单分子定位精度 $\sigma_{\text{loc}} \approx \sqrt{(s^2 + a^2/12)/N}$ ，理解标记密度对分辨率的限制；
6. 由三角函数积化和差理解莫尔条纹与结构光照明的频率搬移，推出线性 SIM 的 2 倍上限；
7. 处理单分子反复定位的数据：线性化、拟合、不确定度与系统误差分析。

前置知识：本讲需要的普通物理与数学

- 光学：光的波动性、惠更斯原理、杨氏双缝干涉、单缝与光栅衍射的初步结论、折射定律、薄透镜成像；
- 近代物理：光子能量 $h\nu$ 、原子能级与跃迁；
- 数学：复数与欧拉公式 $e^{i\theta} = \cos \theta + i \sin \theta$ 、三角函数积化和差、泰勒展开、定积分、极坐标下的面积分、平均值与方差的初步概念。

本讲新引入的工具（复振幅、空间频率、傅里叶变换与卷积、贝塞尔函数、传递函数、速率方程、随机变量的方差、采样定理、最小二乘）都在首次使用前给出定义与推导。

建议学时 共 2 学时（约 90 分钟）：前置模块（波与傅里叶分析）20 分钟；阿贝理论与衍射极限 20 分钟；荧光饱和与 STED 20 分钟；单分子定位 15 分钟；结构光照明 15 分钟。例题与实验题供课后练习，附录供深入阅读。

推导路线 惠更斯-菲涅耳原理 $\Rightarrow$ （远场近似）衍射 = 傅里叶变换 $\Rightarrow$ （阿贝）物镜孔径截掉高空间频率，成像是低通滤波 $\Rightarrow$ 线性成像下截止频率外的信息丢失 $\Rightarrow$ 三条出路：（1）荧光饱和的非线性 $\Rightarrow$ STED；（2）“样品是一个点”的先验 + 光子统计 $\Rightarrow$ 单分子定位；（3）照明与样品频率相乘 $\Rightarrow$ 频率搬移 $\Rightarrow$ SIM。

2 名词与符号

几个容易混淆的说法：

- 分辨率与分辨本领：本讲统一指“能区分的最小间距”，数值越小越好；
- 点扩散函数（point spread function, PSF）：一个理想点光源经过成像系统后得到的像；
- 定位精度与分辨率：前者是“一个孤立分子的位置能测准到多少”，后者是“两个靠近的结构能否被区分”，两者不是一回事（第 7.5 节）；
- 本讲的“荧光”都指分子吸收一个光子后，经过纳秒量级的时间自发地发出一个波长较长的光子；分子之间发光互不相干。

表 1 本讲符号表

符号	含义	单位
λ	真空中的波长 (介质中波长为 λ/n)	nm
n, θ	折射率; 光线与光轴的夹角	1, rad
α, NA	物镜收光锥半角; 数值孔径 $\text{NA} = n \sin \alpha$	rad, 1
U	光场的复振幅 (第 3.1 节)	任意
I	光强 (单位面积功率), $I \propto U ^2$	W/m^2
$f_x, f_y; \mathbf{f}$	空间频率 (单位长度内的周期数); 空间频率矢量	μm^{-1}
$\tilde{g}(f)$	函数 $g(x)$ 的傅里叶变换 (式 (3.7))	
d	光栅 (周期性结构) 的周期	nm
$\ell, D; N_s$	单缝宽度; 圆孔直径; 光栅缝数	m; 1
F_1, F_2	物镜、筒镜的焦距	m
$T(x), c_m$	物的振幅透过率; 其第 m 个傅里叶系数	1
$h(\mathbf{r}), H(\mathbf{r})$	振幅点扩散函数; 强度点扩散函数 $H = h ^2$	
$P(f)$	光瞳函数 (相干传递函数)	1
$\text{OTF}(\mathbf{f})$	光学传递函数 (非相干)	1
f_c	非相干成像截止频率 $2\text{NA}/\lambda$	μm^{-1}
v	无量纲径向坐标 $v = 2\pi\text{NA} r/\lambda$	1
J_0, J_1	第一类贝塞尔函数 (附录 A)	1
w_0, w	衍射极限光斑与 STED 有效光斑的半高全宽	nm
τ, τ_p	荧光寿命; 耗尽光脉冲宽度	ns
σ_a, σ_{se}	吸收截面; 受激辐射截面	cm^2
I_s, ζ	饱和光强 (式 (5.6)、(5.7)); $\zeta = I_{\max}/I_s$	$\text{W}/\text{cm}^2, 1$
η	受耗尽光照射后分子仍能发荧光的概率	1
r_0, β	环形光中心抛物线的特征半径; STED 宽度公式中的系数	nm, 1
s	用高斯近似的单分子像 (PSF) 的标准差	nm
$N, \bar{N}$	一次定位所用的光子数; 其平均值	1
a, b^2	相机像素边长 (折算到样品面); 每像素背景计数的方差	nm, 1
Q	实验题中 $\sigma^2 \bar{N}$ (式 (10.1))	nm^2
L, γ	MINFLUX 两个零点的间距; 光子数比例	nm, 1
σ_{loc}	单分子定位精度 (定位误差的标准差)	nm
$\mathbf{f}_0, \mathbf{f}_s$	照明条纹、样品条纹的空间频率矢量	μm^{-1}
Λ, μ, ϕ	照明条纹的周期、调制度、相位	nm, 1, rad
$\mathcal{E}; \tilde{B}_0, \tilde{B}_{\pm}$	SIM 中的发光强度; 三个频带	
κ, e_n	脉冲饱和参数 $\sigma_a J_p / h\nu$; 第 n 次谐波系数	1
f_{rep}	脉冲重复频率	Hz
$\langle X \rangle, \text{Var } X$	随机变量 X 的期望与方差 (第 7.1 节)	

3 前置模块：波、空间频率与衍射

3.1 复振幅、平面波与空间频率

定义：单色光的复振幅

单色光的电场分量随时间作简谐变化，可写成 $E(\mathbf{r}, t) = A(\mathbf{r}) \cos[\omega t - \phi(\mathbf{r})]$ 。把振幅与相位合成一个复数

$$U(\mathbf{r}) = A(\mathbf{r}) e^{i\phi(\mathbf{r})}, \quad E(\mathbf{r}, t) = \operatorname{Re}[U(\mathbf{r}) e^{-i\omega t}], \quad (3.1)$$

称为复振幅。两列同频光叠加时，电场相加等价于复振幅相加（取实部与相加可以交换次序）。探测器（眼睛、相机）只响应时间平均能流，即光强

$$I = \overline{E^2} (\text{乘以常数}) \propto \frac{1}{2} A^2 = \frac{1}{2} |U|^2.$$

本讲只关心相对光强，写成 $I = |U|^2$ 。

注：相干与非相干叠加

同一光源分出的两束光，相位差固定，叠加后 $I = |U_1 + U_2|^2 = |U_1|^2 + |U_2|^2 + 2 \operatorname{Re}(U_1 U_2^*)$ ，最后一项是干涉项。两个独立发光的荧光分子，相位差在纳秒内随机变化，干涉项的时间平均为零，于是 $I = I_1 + I_2$ ：非相干叠加。这个区别在第 4.4 节决定了荧光显微镜的截止频率。

定义：平面波与空间频率

在折射率为 n 的均匀介质中，沿单位矢量 $\hat{\mathbf{k}}$ 传播的平面波的复振幅为 $U = U_0 e^{i\mathbf{k} \cdot \mathbf{r}}$ ，波矢 $\mathbf{k} = (2\pi n/\lambda) \hat{\mathbf{k}}$ ， λ 为真空波长。设光轴为 z 轴，传播方向在 xz 平面内且与 z 轴成 θ 角，则在某个 $z = \text{常数}$ 的平面上

$$U(x) = U_0 e^{ik_z z} e^{i(2\pi n \sin \theta / \lambda) x} \equiv U'_0 e^{i2\pi f_x x}, \quad f_x = \frac{n \sin \theta}{\lambda}.$$

$U(x)$ 的实部沿 x 方向是余弦起伏，单位长度内有 f_x 个周期。我们把 f_x 叫作这个平面波在该平面上的空间频率（单位 μm^{-1} ），它是“时间频率 = 单位时间内的周期数”在空间中的类比。一般方向上 $\mathbf{f} = (f_x, f_y)$ ， $|\mathbf{f}| = n \sin \theta / \lambda$ 。

由此得到全讲最重要的对应关系：传播方向 $\iff$ 横向空间频率。斜得越厉害，横向空间频率越高；但 $\sin \theta \leq 1$ ，所以在介质中能传播的平面波满足

$$|\mathbf{f}| \leq \frac{n}{\lambda}. \quad (3.2)$$

横向空间频率更高的成分对应 $k_z^2 = (2\pi n/\lambda)^2 - (2\pi |\mathbf{f}|)^2 < 0$ ， k_z 为虚数，场沿 z 方向指数衰减，只能存在于界面附近约一个波长之内，称为倏逝波。远处的透镜收不到它们。

要点： $n \sin \theta$ 在平行界面上守恒

光从折射率 n_1 的介质经平面界面进入 n_2 的介质，折射定律 $n_1 \sin \theta_1 = n_2 \sin \theta_2$ 正好说明：横向空间频率 $n \sin \theta / \lambda$ 在界面两侧相同（界面上两侧的场必须以同样的方式沿界面起伏）。因此光穿过盖玻片、浸油等平行分层介质时， $n \sin \theta$ 不变。显微镜物镜收光锥的大小用 $\text{NA} = n \sin \alpha$ （ n 为样品与物镜之间介质的折射率， α 为收光锥半角）刻画，正是因为它在各层中不变。空气

中 $\text{NA} < 1$ ；用折射率 1.52 左右的浸油，NA 可达 1.4–1.5。

3.2 惠更斯–菲涅耳原理

假设：标量衍射理论的适用条件

1. 只考虑电场的一个分量，把光当作标量波（标量近似）。当所涉及的角度不太大、开孔远大于波长时，这一近似很好；高 NA 物镜的焦点处偏振效应会使光斑形状改变约 10%，本讲在用到处会说明 [21]。
2. 介质均匀、线性，光为单色光。

定义：惠更斯–菲涅耳原理

惠更斯认为波前上每一点都是新的子波源；菲涅耳补充了“子波相互干涉”。定量形式是：已知平面 Σ （例如一个开孔所在的平面）上的复振幅 $U(x', y')$ ，则 Σ 后面一点 Q 处的复振幅为所有子波的叠加

$$U(Q) = \frac{1}{i\lambda_n} \iint_{\Sigma} U(x', y') \frac{e^{ikr}}{r} K dx' dy', \quad (3.3)$$

其中 r 是面元到 Q 的距离， $k = 2\pi/\lambda_n$ ， $\lambda_n = \lambda/n$ 为介质中波长， e^{ikr}/r 是从面元出发的球面子波， K 是一个与方向有关的倾斜因子（近轴时 $K \approx 1$ ）。前因子 $1/(i\lambda_n)$ 的大小与相位由“平面波通过这个公式应当复原自身”这一要求确定，不影响本讲关心的强度分布形状。

3.3 夫琅禾费近似：远场衍射就是傅里叶变换

下面从式 (3.3) 出发，看观察点很远时积分能化简成什么形式。

推导：从式 (3.3) 到远场

取开孔中心为原点，观察点 Q 在远处，坐标 (X, Y, Z) ，到原点距离 $R = \sqrt{X^2 + Y^2 + Z^2}$ 。开孔内一点 $(x', y', 0)$ 到 Q 的距离为

$$\begin{aligned} r &= \sqrt{(X - x')^2 + (Y - y')^2 + Z^2} \\ &= R \sqrt{1 - \frac{2(Xx' + Yy')}{R^2} + \frac{x'^2 + y'^2}{R^2}}. \end{aligned}$$

1. 对 $(1 + \epsilon)^{1/2} \approx 1 + \epsilon/2 - \epsilon^2/8$ 展开，小参量 $\epsilon \sim \rho'/R$ (ρ' 为开孔尺寸)。保留到 ρ'^2/R ：

$$r \approx R - \frac{Xx' + Yy'}{R} + \frac{x'^2 + y'^2}{2R} - \frac{(Xx' + Yy')^2}{2R^3}.$$

2. 分母中的 r 只影响振幅，换成 R 的相对误差为 ρ'/R ，可以忽略。指数中的 kr 影响相位，必须仔细：最后两项的相位不超过 $k\rho'^2/(2R)$ 。夫琅禾费条件是

$$R \gg \frac{\pi\rho'^2}{\lambda_n}, \quad (3.4)$$

此时这两项的相位远小于 1 rad，可以略去。

3. 剩下 $kr \approx kR - k(x' \sin \theta_x + y' \sin \theta_y)$, 其中 $\sin \theta_x = X/R$ 、 $\sin \theta_y = Y/R$ 是观察方向的方向余弦。代入式 (3.3),

$$U(Q) = \frac{e^{ikR}}{i\lambda_n R} \iint U(x', y') e^{-i2\pi(f_x x' + f_y y')} dx' dy', \quad (3.5)$$

$$f_x = \frac{n \sin \theta_x}{\lambda}, \quad f_y = \frac{n \sin \theta_y}{\lambda}.$$

式 (3.5) 的积分是一种标准运算, 叫作 $U(x', y')$ 的傅里叶变换。它的物理意义很直接: 沿方向 (θ_x, θ_y) 远去的光的振幅, 等于开孔处场中空间频率为 $(f_x, f_y) = (n \sin \theta_x, n \sin \theta_y)/\lambda$ 的那个“余弦成分”的大小。下一节把这个说法讲清楚。

检验：夫琅禾费条件的数值

直径 1 mm 的圆孔、 $\lambda = 500 \text{ nm}$ (空气): $\pi \rho'^2/\lambda = \pi \times (0.5 \text{ mm})^2/500 \text{ nm} \approx 1.6 \text{ m}$ 。要在几米之外才是远场。用透镜可以把“无穷远”搬到焦平面上 (第 3.6 节), 显微镜中正是这样做的。

3.4 傅里叶分析：最少够用的部分

从傅里叶级数说起 一个周期为 d 的函数 $g(x)$ (例如光栅的透过率) 可以写成不同空间频率的余弦与正弦之和。用复指数写最简洁:

$$g(x) = \sum_{m=-\infty}^{\infty} c_m e^{i2\pi m x/d}, \quad (3.6)$$

$$c_m = \frac{1}{d} \int_{-d/2}^{d/2} g(x) e^{-i2\pi m x/d} dx.$$

第二式的来历: 把第一式两边乘以 $e^{-i2\pi m' x/d}$ 并在一个周期上积分, 由于 $\int_{-d/2}^{d/2} e^{i2\pi(m-m')x/d} dx = d$ ($m = m'$) 或 0 ($m \neq m'$, 积分了整数个周期的正弦与余弦), 右边只剩 $m = m'$ 一项。若 g 是实函数, 则 $c_{-m} = c_m^*$, 第 m 项与第 $-m$ 项合起来是一个实的余弦 $2|c_m| \cos(2\pi m x/d + \arg c_m)$ 。所以, 周期函数是空间频率 $0, 1/d, 2/d, \dots$ 的余弦之和。

例题：方波光栅的分解

透光宽度为 $d/2$ 、中心在 $x = 0$ 的方波光栅: 一个周期内 $|x| < d/4$ 时 $g = 1$, 否则 $g = 0$ 。由式 (3.6),

$$c_m = \frac{1}{d} \int_{-d/4}^{d/4} e^{-i2\pi m x/d} dx = \frac{\sin(\pi m/2)}{\pi m} = \frac{1}{2} \text{sinc} \frac{m}{2},$$

$$\text{sinc } u \equiv \frac{\sin \pi u}{\pi u}.$$

$c_0 = 1/2$ (平均透过率), $c_{\pm 1} = 1/\pi$, 偶数级 $c_{\pm 2} = c_{\pm 4} = \dots = 0$, $c_{\pm 3} = -1/(3\pi)$ 。图 2(a) 显示: 只取 0 级与 ± 1 级, 得到周期正确的余弦; 加入越来越高的频率, 边缘越来越陡。精细的特征 (陡峭的边缘、窄的缝) 由高空间频率携带。

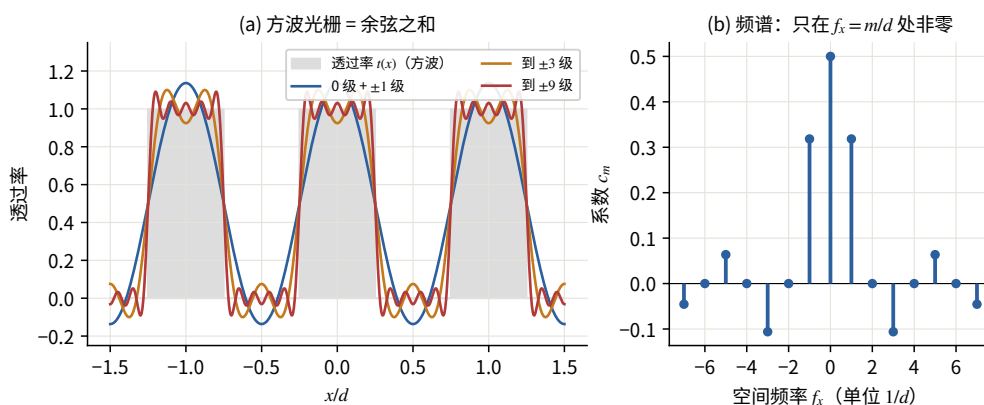


图 2 (a) 方波光栅 (灰色) 与它的傅里叶级数部分和; (b) 傅里叶系数 c_m , 只在空间频率 m/d 处非零。系数由式 (3.6) 计算, 并由数值积分核对 (scripts/fig_fourier.py)。

傅里叶变换 非周期函数 (例如单个狭缝) 可以看作周期 $d \rightarrow \infty$ 的极限: 允许的频率 m/d 越来越密, 变成连续变量 f 。定义

$$\begin{aligned}\tilde{g}(f) &= \int_{-\infty}^{\infty} g(x) e^{-i2\pi f x} dx, \\ g(x) &= \int_{-\infty}^{\infty} \tilde{g}(f) e^{i2\pi f x} df.\end{aligned}\quad (3.7)$$

第一式叫傅里叶变换, 第二式叫逆变换, 它说 $g(x)$ 是频率连续分布的复指数 $e^{i2\pi f x}$ 的叠加, $\tilde{g}(f) df$ 是频率在 f 到 $f + df$ 之间那部分的“权重”。(逆变换成立的证明见数学分析教材; 对本讲用到的函数, 可以把它看作式 (3.6) 在 $d \rightarrow \infty$ 时的极限: 令 $\tilde{g}(m/d) = d c_m$, 求和 $\sum_m(\cdot)$ 乘以频率间隔 $1/d$ 变成积分。) 二维函数的变换就是对 x, y 各做一次: $\tilde{g}(f_x, f_y) = \iint g(x, y) e^{-i2\pi(f_x x + f_y y)} dx dy$ 。

对照式 (3.5): 远场衍射的振幅分布就是开孔处复振幅的傅里叶变换, 频率与方向的关系为 $f = n \sin \theta / \lambda$ 。

下面几条性质, 本讲都要用。每条都可以把定义式 (3.7) 直接代入验证。

1. 线性: $ag_1 + bg_2$ 的变换是 $a\tilde{g}_1 + b\tilde{g}_2$ 。
2. 平移: $g(x - x_0)$ 的变换是 $e^{-i2\pi f x_0} \tilde{g}(f)$ 。(令 $x' = x - x_0$ 换元即得。) 平移只改变各频率成分的相位, 不改变其大小。
3. 伸缩: $g(x/L)$ 的变换是 $L \tilde{g}(Lf)$ 。函数在 x 方向变宽 L 倍, 频谱就变窄 L 倍: 越细的结构, 频谱越宽。
4. 调制: $g(x)e^{i2\pi f_0 x}$ 的变换是 $\tilde{g}(f - f_0)$ 。乘上一个复指数, 频谱整体搬移 f_0 。这是第 8 节结构光照明的数学核心。
5. 卷积定理 (证明见下)。

定义：卷积

两个函数的卷积定义为

$$(g \otimes h)(x) = \int_{-\infty}^{\infty} g(x') h(x - x') dx'. \quad (3.8)$$

物理图像：把 g 看作许多点源的分布 (x' 处强度 $g(x')$)，每个点源都产生同样形状、以 x' 为中心的响应 $h(x - x')$ ，总响应是它们的叠加。成像中， g 是物， h 是点扩散函数，卷积就是“物上每一点都被模糊成一个 h ”。

推导：卷积定理：卷积的变换 = 变换的乘积

对式 (3.8) 做傅里叶变换，交换积分次序，并在内层积分中令 $u = x - x'$ ：

$$\begin{aligned} & \iint g(x')h(x - x')e^{-i2\pi fx} dx' dx \\ &= \int g(x')e^{-i2\pi fx'} dx' \int h(u)e^{-i2\pi fu} du. \end{aligned}$$

这里用了 $e^{-i2\pi fx} = e^{-i2\pi fx'}e^{-i2\pi fu}$ 。于是

$$\widetilde{g \otimes h}(f) = \tilde{g}(f)\tilde{h}(f). \quad (3.9)$$

反过来(把 x 与 f 的角色对调, 同样的推导), 两函数乘积的变换是它们变换的卷积: $\widetilde{gh} = \tilde{g} \otimes \tilde{h}$ 。

定义： δ 函数

$\delta(x)$ 是一个“无限窄、面积为 1”的尖峰的理想化：对任何连续函数 g , $\int g(x)\delta(x-x_0)dx = g(x_0)$ 。由定义式 (3.7), $\delta(x - x_0)$ 的变换是 $e^{-i2\pi fx_0}$ 。反过来, 把 $\tilde{g}(f) = \delta(f - f_0)$ 代入逆变换, 得 $g(x) = \int \delta(f - f_0)e^{i2\pi fx}df = e^{i2\pi f_0 x}$; 由于变换与逆变换一一对应, $e^{i2\pi f_0 x}$ 的变换就是 $\delta(f - f_0)$ 。于是

$$\begin{aligned} \cos(2\pi f_0 x) &= \frac{1}{2} \left(e^{i2\pi f_0 x} + e^{-i2\pi f_0 x} \right) \\ &\longrightarrow \frac{1}{2} [\delta(f - f_0) + \delta(f + f_0)]. \end{aligned}$$

一个余弦只含 $\pm f_0$ 两个频率。周期函数式 (3.6) 的变换是位于 $f = m/d$ 的一串 δ 峰, 权重 c_m (图 2(b))。

3.5 三个经典衍射图样

推导：单缝

宽 ℓ 的狭缝被正入射平面波照明, 缝内 $U = 1$ 。由式 (3.5) (只保留 x 方向),

$$\begin{aligned} \tilde{U}(f) &= \int_{-\ell/2}^{\ell/2} e^{-i2\pi fx} dx = \frac{\sin \pi f \ell}{\pi f} = \ell \operatorname{sinc}(f\ell), \\ I(\theta) &\propto \operatorname{sinc}^2\left(\frac{\ell \sin \theta}{\lambda}\right) \end{aligned}$$

(空气中 $n = 1$)。暗纹在 $\ell \sin \theta = m\lambda$ ($m = \pm 1, \pm 2, \dots$), 中央亮纹的角半宽 $\sin \theta = \lambda/\ell$: 缝越窄, 衍射越开 (伸缩性质)。见图 3(a)。

推导：光栅

N_s 条宽 ℓ 、周期 d 的狭缝。透过率可以写成“单缝”与“ N_s 个 δ 峰”的卷积，由卷积定理，远场振幅 = 单缝因子 $\times$ 多缝因子：

$$\begin{aligned}\tilde{U}(f) &= \ell \operatorname{sinc}(f\ell) \sum_{j=0}^{N_s-1} e^{-i2\pi fjd} \\ &= \ell \operatorname{sinc}(f\ell) e^{-i\pi f(N_s-1)d} \frac{\sin(N_s\pi fd)}{\sin(\pi fd)}.\end{aligned}$$

第二步是等比数列求和。当 fd 为整数 m 时多缝因子的模达到最大值 N_s ：光栅方程

$$\begin{aligned}d(n \sin \theta_m - n \sin \theta_i) &= m\lambda, \\ m &= 0, \pm 1, \pm 2, \dots\end{aligned}\tag{3.10}$$

这里顺便写出了斜入射（入射方向 θ_i ）的一般形式：斜入射的平面波在光栅面上带有相位 $e^{i2\pi(n \sin \theta_i/\lambda)x}$ ，由调制性质，整个频谱搬移 $n \sin \theta_i/\lambda$ 。 N_s 很大时，衍射光集中在一系列分立的方向上，即衍射级，第 m 级的振幅正比于傅里叶系数 c_m 。见图 3(b)。

要点：圆孔：Airy 斑

直径 D 的圆孔被正入射平面波照明，远场强度为（推导见附录 A）

$$I(\theta) = I_0 \left[\frac{2J_1(v)}{v} \right]^2, \quad v = \frac{\pi D \sin \theta}{\lambda},\tag{3.11}$$

其中 J_1 是一阶贝塞尔函数，可以用积分 $J_1(v) = \frac{1}{\pi} \int_0^\pi \cos(\varphi - v \sin \varphi) d\varphi$ 定义（附录 A），它像一个振幅缓慢衰减的正弦。 $J_1(v)$ 的第一个正零点为 $v_1 = 3.8317$ ，所以第一暗环在

$$\sin \theta_1 = \frac{3.8317}{\pi} \frac{\lambda}{D} = 1.22 \frac{\lambda}{D}.\tag{3.12}$$

中央亮斑（Airy 盘）集中了总光能的 84%（附录 A）。

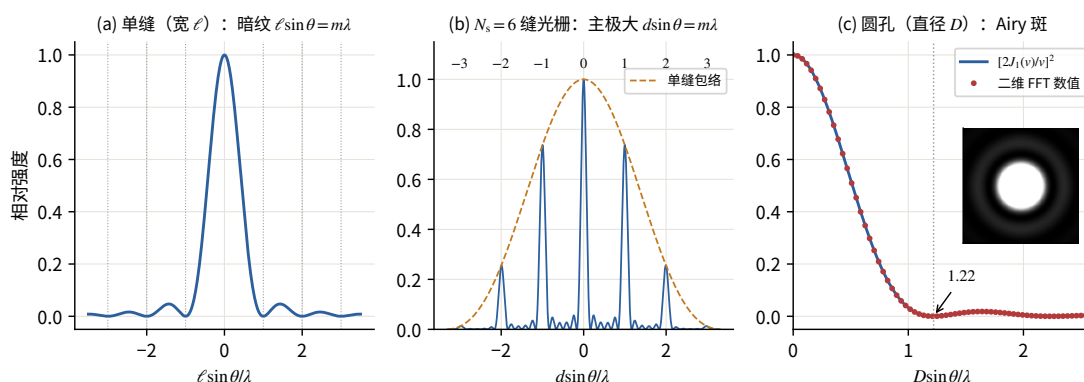


图 3 三种夫琅禾费衍射图样，均由解析式计算。(a) 单缝；(b) 6 缝光栅（缝宽 $0.3d$ ），橙色虚线为单缝包络；(c) 圆孔的 Airy 斑：蓝线为式 (3.11)，红点为对圆孔透过率直接做二维快速傅里叶变换的数值结果，最大偏差 10^{-3} ；插图为二维图样（压缩了亮度以显出暗环）。第一暗环位置 1.2197 由脚本核对。

3.6 透镜把方向变成位置；数值孔径

推导：焦平面上的场是物面场的傅里叶变换

先看空气中 ($n = 1$) 的薄透镜。与光轴成 θ 角的一束平行光，经焦距 F 的透镜后会聚到后焦面上 $x = F \tan \theta \approx F \sin \theta$ 处 (近轴, $\theta \ll 1$)。由第 3.1 节，物面上的场可以分解为许多平面波，空间频率为 f_x 的那个平面波沿 $\sin \theta = \lambda f_x$ 的方向传播，因而会聚到

$$x_{\text{BFP}} = \lambda F f_x. \quad (3.13)$$

所以后焦面上的每一点对应物面场的一个空间频率。若物恰好放在透镜的前焦面上，可以证明后焦面上的复振幅分布严格正比于物面场的傅里叶变换；物不在前焦面时，后焦面上的场还要乘一个随位置二次变化的相位因子 (相位 $\propto x_{\text{BFP}}^2$)，强度分布不变 [22]。显微镜正是把样品放在物镜前焦面附近。这相当于把式 (3.5) 中的“无穷远”搬到了焦平面上。

注：显微物镜与正弦条件

高 NA 物镜的收光角很大，近轴近似不再成立。设计良好的显微物镜满足阿贝提出的正弦条件：从焦点出发、与光轴成 θ 角的光线，离开物镜后与光轴的距离正比于 $n \sin \theta$ 。于是后焦面 (光瞳) 上的径向坐标严格正比于 $n \sin \theta = \lambda |\mathbf{f}|$ ，“光瞳上的位置 = 空间频率”的对应不依赖近轴近似 [22, 23]。物镜能收集的最大空间频率是

$$|\mathbf{f}|_{\text{max}} = \frac{n \sin \alpha}{\lambda} = \frac{\text{NA}}{\lambda}, \quad (3.14)$$

它对应后焦面上光阑 (光瞳) 的边缘。

4 阿贝成像理论与衍射极限

4.1 物体是许多光栅的叠加

设物是一张薄的透明片，振幅透过率为 $T(x, y)$ ：照明光的复振幅经过它后乘以 T 。由式 (3.7) 的逆变换，

$$T(x, y) = \iint \tilde{T}(f_x, f_y) e^{i2\pi(f_x x + f_y y)} df_x df_y,$$

即任何物都可以看作许多余弦光栅的叠加，空间频率为 (f_x, f_y) 的光栅 (周期 $1/|\mathbf{f}|$ ，条纹垂直于 $\mathbf{f}$) 的“强度”是 $\tilde{T}(\mathbf{f})$ 。物的细节越精细，所需的空间频率越高。由线性性质，只要弄清一个余弦光栅怎样成像，就弄清了一切物怎样成像。这是阿贝的出发点 [1, 22]。

4.2 光栅成像：两个衍射级才能成像

假设：阿贝理论的设定

1. 照明是相干的：一束 (或一系列互不相干的单个) 单色平面波照到物上。
2. 物是薄的透射光栅，周期 d ，样品与物镜之间充满折射率 n 的介质；物镜的收光锥半角为 α ， $\text{NA} = n \sin \alpha$ 。

3. 物镜本身无像差，只是有限大小：方向在收光锥之外的光进不了物镜。

推导：阿贝的两步成像

第一步：物产生衍射。由光栅方程 (3.10)，入射方向 θ_i 的平面波经过光栅后分成一系列平面波（衍射级），第 m 级的方向满足

$$n \sin \theta_m = n \sin \theta_i + m \frac{\lambda}{d},$$

振幅正比于傅里叶系数 c_m 。在“ $n \sin \theta$ 轴”上，各衍射级等间距排列，间距为 λ/d (图 4(c))。第二步：进入物镜的衍射级在像面上相互干涉。物镜只收 $|n \sin \theta_m| \leq \text{NA}$ 的衍射级。这些平面波经物镜和筒镜（显微镜中物镜后面的第二块透镜，把物镜出射的平行光会聚到相机上成像）后又变成平面波，在像面上叠加（像面的坐标已按放大率折算回物面尺度）：

$$U_{\text{img}}(x) \propto e^{i2\pi n \sin \theta_i x / \lambda} \sum_{m: |n \sin \theta_m| \leq \text{NA}} c_m e^{i2\pi m x / d}. \quad (4.1)$$

公共因子 $e^{i2\pi n \sin \theta_i x / \lambda}$ 的模为 1，不影响光强。

- 若只收到一个衍射级（比如只有 0 级）， $I_{\text{img}} = |c_0|^2$ 处处相同：像是一片均匀的亮，看不到光栅。
- 若收到 0 级与 +1 级， $I_{\text{img}} = |c_0 + c_1 e^{i2\pi x/d}|^2 = |c_0|^2 + |c_1|^2 + 2|c_0 c_1| \cos(2\pi x/d + \phi)$ ，出现周期为 d 的条纹：光栅被“分辨”了。

所以，要看到周期为 d 的结构，至少要有两个相邻的衍射级同时进入物镜。

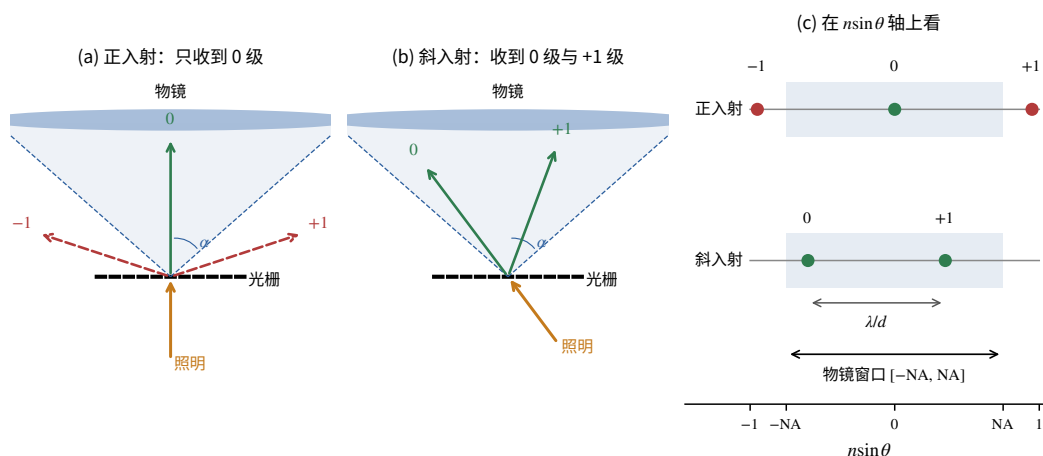


图 4 阿贝成像。(a) 正入射： ± 1 级的方向超出物镜收光锥，只有 0 级进入，像中没有条纹。(b) 同一光栅，照明斜入射：0 级与 +1 级都进入物镜，形成周期为 d 的条纹。(c) 在 $n \sin \theta$ 轴上看：衍射级等间距（间距 λ/d ），物镜是宽 2NA 的“窗口”。图中参数 $\text{NA} = 0.75$ （空气）， $\lambda/d = 0.95$ ，斜入射 $\sin \theta_i = -0.60$ ；哪些级次被接收由光栅方程计算（scripts/fig_abbe.py）。

要点：阿贝分辨极限

1. 正入射 ($\theta_i = 0$) : 0 级在窗口正中, ± 1 级在 $\pm \lambda/d$ 。要让 ± 1 级中至少一个进入, 需 $\lambda/d \leq \text{NA}$:

$$d_{\min} = \frac{\lambda}{\text{NA}} \quad (\text{正入射相干照明}). \quad (4.2)$$

2. 斜照明: 把 0 级移到窗口的一边 ($n \sin \theta_i = -\text{NA}$), $+1$ 级在 $-\text{NA} + \lambda/d$, 只要它不超出另一边: $\lambda/d \leq 2\text{NA}$,

$$d_{\min} = \frac{\lambda}{2\text{NA}} = \frac{\lambda}{2n \sin \alpha} \quad (\text{最倾斜的照明}). \quad (4.3)$$

照明光由聚光镜提供, 其数值孔径 NA_c 限制了 $|n \sin \theta_i|$, 一般情形为 $d_{\min} = \lambda/(\text{NA} + \text{NA}_c)$ 。

式 (4.3) 就是阿贝极限。取 $\lambda = 500 \text{ nm}$ 、 $\text{NA} = 1.4$, $d_{\min} \approx 180 \text{ nm}$ 。

注：放大倍数与分辨率无关

上面的论证中没有出现放大倍数。被物镜挡掉的衍射级所携带的信息根本没有进入仪器, 后面的目镜放大多少倍也找不回来。阿贝把超出分辨能力的放大称为“空放大”。另外注意 λ/d 也不能超过 $2n$: 由式 (3.2), 周期小于 $\lambda/(2n)$ 的光栅, 其 ± 1 级在任何照明下都是倏逝波, 根本传不出样品附近。

交互演示：阿贝成像。调节光栅周期 d 、物镜 NA 、照明方向 $n \sin \theta_i$ 和样品与物镜之间的介质 (NA 与 $|n \sin \theta_i|$ 都不能超过介质折射率 n ; $|n \sin \theta_m| > n$ 的级次是倏逝波, 不画出): 上图是各衍射级在 $n \sin \theta$ 轴上的位置 (绿色 = 进入物镜), 下图是由式 (4.1) 算出的像。试试: (1) 正入射时把 d 减小到 λ/NA 以下, 条纹消失; (2) 加上斜照明, 条纹重新出现; (3) 选“挡住 0 级”, 观察像的变化, 并与例题 1(3) 的计算对照。

网页版中此处为可调参数的交互动画 (演示编号 superres-abbe)。

4.3 傅里叶光学的表述：成像是低通滤波

把阿贝的图像推广到任意物。考虑图 5(a) 的“4f 系统” (两块焦距相同的透镜相距 $2F$, 物在第一块的前焦面, 像在第二块的后焦面, 物到像总长 $4F$, 故名): 物在物镜的前焦面, 物镜后焦面 (光瞳) 上放光阑, 筒镜把后焦面上的场再变换一次, 在像面上成像。由第 3.6 节, 后焦面上的场是物面场的傅里叶变换; 筒镜再做一次变换, 得到 (倒立、放大的) 物面场。

推导：相干成像的传递函数与点扩散函数

设正入射照明, 物面场 $U_{\text{obj}} = T(x, y)$ 。

1. 后焦面上, 空间频率 $\mathbf{f}$ 处的场正比于 $\tilde{T}(\mathbf{f})$ (式 (3.5)、(3.13))。
2. 光阑让 $|\mathbf{f}| \leq \text{NA}/\lambda$ 的成分通过, 挡住其余。用光瞳函数 $P(\mathbf{f})$ 表示: 光阑内 $P = 1$, 光阑外 $P = 0$ 。通过后为 $P(\mathbf{f})\tilde{T}(\mathbf{f})$ 。

3. 筒镜做逆变换（把像面坐标按放大率折算回物面，并把倒像翻正），像的复振幅为

$$U_{\text{img}}(\mathbf{r}) = \iint P(\mathbf{f}) \tilde{T}(\mathbf{f}) e^{i2\pi \mathbf{f} \cdot \mathbf{r}} d^2 f.$$

4. 由卷积定理 (3.9)，频域相乘等于空间域卷积：

$$U_{\text{img}} = T \otimes h, \quad h(\mathbf{r}) = \iint P(\mathbf{f}) e^{i2\pi \mathbf{f} \cdot \mathbf{r}} d^2 f. \quad (4.4)$$

h 叫作振幅点扩散函数：若物是一个点 ($T = \delta$)，像就是 h 。 $P(\mathbf{f})$ 叫作相干传递函数。对半径 NA/λ 的圆形光瞳， h 就是圆孔的变换（附录 A）：

$$h(r) \propto \frac{2J_1(v)}{v}, \quad v = \frac{2\pi \text{NA}}{\lambda} r. \quad (4.5)$$

这就是阿贝理论的现代表述：成像系统是一个空间频率的低通滤波器，截止频率（相干照明）为 NA/λ 。周期 d 的光栅只有在 $1/d \leq \text{NA}/\lambda$ 时才能通过，与式 (4.2) 一致。

一个可以预见的疑问：相干成像的强度像 $|U_{\text{img}}|^2$ 是振幅的平方，两个振幅分量 $e^{i2\pi \mathbf{f}_1 \cdot \mathbf{r}}$ 、 $e^{i2\pi \mathbf{f}_2 \cdot \mathbf{r}}$ 相乘会产生差频 $\mathbf{f}_1 - \mathbf{f}_2$ ，最高可达 $2\text{NA}/\lambda$ ，为什么还说截止频率是 NA/λ ？因为截止是对振幅而言的：相干成像中被忠实传递的是物的振幅透过率 T 的频谱；强度像中 NA/λ 到 $2\text{NA}/\lambda$ 之间的成分是振幅分量之间的差频，并不等于物的强度透过率 $|T|^2$ 在该频率的成分，甚至可能是物上根本没有的结构（例题 1(3) 的“假细节”就是一例）。更一般地，相干成像的强度像对物是非线性的 ($|T \otimes h|^2$)，即使在 NA/λ 以内，强度频谱也不等于把 $|T|^2$ 的频谱简单滤波的结果；只有非相干成像中“像的强度频谱 = 物的强度频谱 $\times$ OTF”才严格成立。

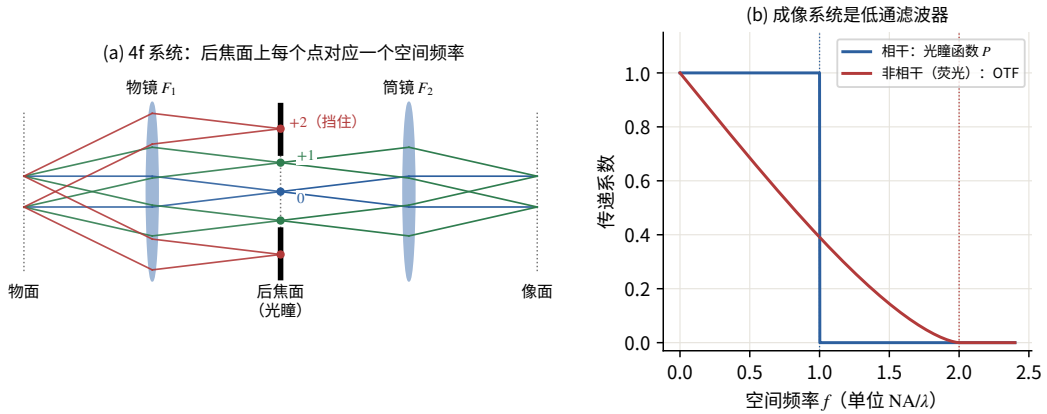


图 5 (a) 4f 成像系统（示意）：物面上同一方向的平行光（同一空间频率）会聚到后焦面上的同一点，光阑挡住高频成分（ ± 2 级）。光线按薄透镜规则计算， $F_1 = F_2$ 。(b) 相干传递函数（光瞳函数，截止于 NA/λ ）与非相干光学传递函数 OTF（第 4.4 节；光瞳的自相关，截止于 $2\text{NA}/\lambda$ ）。OTF 由二维数值自相关与式 (4.9) 核对（scripts/fig_transfer.py）。

4.4 荧光是非相干成像：截止频率加倍

荧光显微镜中，“物”是样品里荧光分子的分布 $C(\mathbf{r})$ （单位面积的发光功率）。每个分子独立发光，彼此不相干。

推导：非相干成像：强度卷积

1. 位于 $\mathbf{r}'$ 的一个分子，在像面上产生的复振幅是 $h(\mathbf{r} - \mathbf{r}')$ 乘以一个随机相位因子，光强为 $C(\mathbf{r}') |h(\mathbf{r} - \mathbf{r}')|^2$ 。
2. 不同分子的随机相位彼此独立，干涉项的时间平均为零（第 3.1 节注），光强直接相加：

$$I_{\text{img}}(\mathbf{r}) = \iint C(\mathbf{r}') H(\mathbf{r} - \mathbf{r}') d^2 r' = C \otimes H, \quad (4.6)$$

$$H \equiv |h|^2.$$

H 就是一个点状荧光分子的像，称为（强度）点扩散函数。

3. 由卷积定理，像的频谱为 $\tilde{I}_{\text{img}} = \tilde{C} \cdot \tilde{H}$ 。通常把 $\tilde{H}$ 归一化，定义光学传递函数

$$\text{OTF}(\mathbf{f}) = \frac{\tilde{H}(\mathbf{f})}{\tilde{H}(0)}. \quad (4.7)$$

推导：OTF 的截止频率是 $2\text{NA}/\lambda$

由式 (4.4)， h 是 P 的逆变换； $H = h \cdot h^*$ 是一个乘积。由卷积定理的反向形式（乘积的变换 = 变换的卷积），并注意 $h^*(\mathbf{r})$ 的变换是 $P^*(-\mathbf{f})$ （取复共轭并把 $\mathbf{f} \rightarrow -\mathbf{f}$ ），

$$\tilde{H}(\mathbf{f}) = \iint P(\mathbf{f}') P^*(\mathbf{f}' - \mathbf{f}) d^2 f'.$$

这是光瞳函数与它自己平移 $\mathbf{f}$ 后的重叠积分，称为自相关。对圆形光瞳， $P(\mathbf{f}')$ 与 $P(\mathbf{f}' - \mathbf{f})$ 是两个半径 NA/λ 、圆心相距 $|\mathbf{f}|$ 的圆盘，积分等于它们的重叠面积。当 $|\mathbf{f}| > 2\text{NA}/\lambda$ 时两圆不重叠， $\tilde{H} = 0$ 。于是非相干成像的截止频率为

$$f_c = \frac{2\text{NA}}{\lambda}. \quad (4.8)$$

重叠面积除以一个圆的面积，即得（附录 B）

$$\text{OTF}(f) = \frac{2}{\pi} \left[\arccos \frac{f}{f_c} - \frac{f}{f_c} \sqrt{1 - \left(\frac{f}{f_c}\right)^2} \right], \quad (4.9)$$

$$0 \leq f \leq f_c.$$

要点：非相干成像与阿贝极限的关系

荧光（非相干）成像能传递的最短周期为 $1/f_c = \lambda/(2\text{NA})$ ，恰与阿贝斜照明极限 (4.3) 相同。物理上，强度是振幅的平方，两个振幅分量 $e^{i2\pi f_1 \cdot \mathbf{r}}$ 、 $e^{i2\pi f_2 \cdot \mathbf{r}}$ 相乘时产生差频 $\mathbf{f}_1 - \mathbf{f}_2$ ； $\mathbf{f}_1, \mathbf{f}_2$ 各自可取到光瞳的两端，差频最大为 $2\text{NA}/\lambda$ 。但 OTF 在接近 f_c 时趋于零（图 5(b)）：周期接近 $\lambda/(2\text{NA})$ 的结构虽然传过去了，对比度却很低，实际能看到的取决于噪声。

4.5 点扩散函数与瑞利判据

由式 (4.5)，一个荧光分子的像是

$$H(r) = \left[\frac{2J_1(v)}{v} \right]^2, \quad v = \frac{2\pi NA}{\lambda} r. \quad (4.10)$$

这与式 (3.11) 是同一个函数，两者的变量一一对应：圆孔衍射中的（远场空间频率 $\sin \theta / \lambda$ ，孔半径 $D/2$ ）对应显微 PSF 中的（焦面位置 r ，光瞳的频率半径 NA/λ ）。由 J_1 的第一个零点 $v_1 = 3.8317$ ，第一暗环半径和半高全宽分别为

$$r_R = \frac{3.8317}{2\pi} \frac{\lambda}{NA} = 0.61 \frac{\lambda}{NA}, \quad (4.11)$$

$$w_0 = 0.51 \frac{\lambda}{NA} \approx \frac{\lambda}{2NA}.$$

（半高全宽由 $[2J_1(v)/v]^2 = 1/2$ 的数值解 $v = 1.6163$ 得到。）

定义：瑞利判据与斯帕罗判据

两个强度相同的非相干点光源，相距 Δx ，像为两个 Airy 斑之和。

- 瑞利判据：一个斑的中心落在另一个斑的第一暗环上， $\Delta x = r_R = 0.61\lambda/NA$ ，算作“刚好可分辨”。此时两峰之间中点的强度是峰值的 73.5%（图 6(b)），人眼能看出凹陷。
- 斯帕罗判据 [24]：再靠近，凹陷消失、中点处曲率变为零， $\Delta x = 0.47\lambda/NA$ （数值求解 $\partial^2 I / \partial x^2 = 0$ ）。

两个判据都是约定，不是物理定律；它们和截止周期 $\lambda/(2NA)$ 是同一量级，描述的是同一个限制的不同侧面。

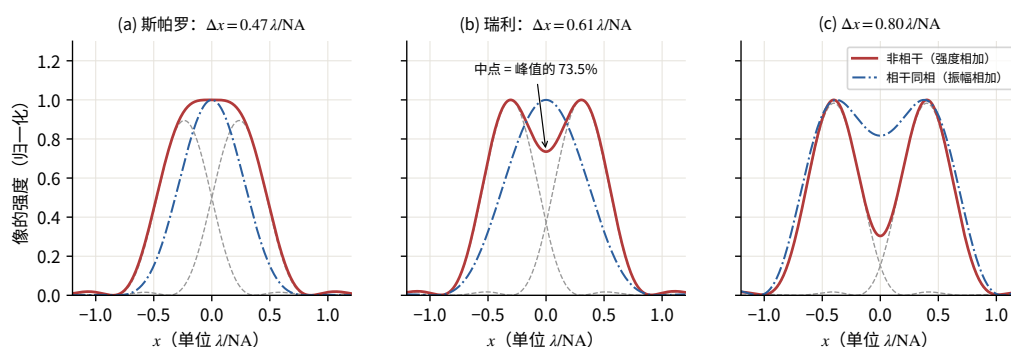


图 6 两个点光源的像沿连线的强度剖面（各自归一化到最大值）。红：非相干（强度相加）；灰虚线：单个 Airy 斑；蓝点划线：相干同相（振幅相加）。(a) 斯帕罗间距：非相干的像中点恰好不再凹陷，顶部变平；(b) 瑞利间距下中点为峰值的 73.5%；相干同相的两点在此间距下中点反而最亮，完全看不出是两个点；(c) $0.80\lambda/NA$ ：非相干时凹陷很深，相干同相时中点仍有峰值的 82%，分辨要难得多。数值由 `scripts/fig_rayleigh.py` 计算并核对。

交互演示：两个点光源的像。调节间距、波长与 NA ，比较非相干（荧光）与相干同相、反相时的像；读数给出连线中点与峰值的强度比。按“设为瑞利间距”可检验 73.5%。

网页版中此处为可调参数的交互动画（演示编号 `superres-psf`）。

4.6 为什么“更亮的光”“更好的算法”都越不过截止频率

要点：线性成像的信息极限

在式 (4.6) 描述的线性成像中，像的频谱 $\tilde{I}_{\text{img}}(\mathbf{f}) = \text{OTF}(\mathbf{f}) \tilde{C}(\mathbf{f}) \tilde{H}(0)$ 。对 $|\mathbf{f}| > f_c$ ，OTF 等于零。

1. 信息丢失，而不是被噪声淹没：取任意一个只含 $|\mathbf{f}| > f_c$ 成分的函数 $\Delta C(\mathbf{r})$ （例如周期 100 nm 的余弦起伏；只要 C 本身含足够大的常数， $C + \Delta C \geq 0$ 就仍是合理的浓度分布），物 C 与 $C + \Delta C$ 的像完全相同。即使没有一点噪声，也不可能从像判断物是哪一个。
2. 增大光强只是把 C 整体乘一个常数，OTF 的零区不变。
3. 去卷积等算法可以把 $0 < \text{OTF} < 1$ 的区域“放大”回来（相当于除以 OTF），但除以零是不可能的。

注：严格的数学例外及其代价

若事先知道物只占据有限区域，数学上它的频谱是“解析函数”（处处可以展开成收敛的泰勒级数的函数；知道它在任意一小段区间上的值，就确定了它在所有地方的值），原则上可由截止频率以内的频谱唯一延拓出去。但这种延拓对噪声极其敏感，所需信噪比随要扩展的频率范围指数增长，实际上不可行 [22]。这正说明：要越过截止频率，必须引入新的物理或强的先验知识。

要点：三条出路

上面的论证依赖三个前提：(i) 发光强度与照明光强成正比（线性）；(ii) 对物一无所知；(iii) 照明是均匀的。分别放弃其中一个，就得到三类超分辨方法：

1. 放弃 (i)：利用荧光的饱和，使有效点扩散函数变窄——STED（第 6 节）、非线性 SIM；
2. 放弃 (ii)：已知“这一帧里亮的是一个个孤立的点”，位置就能测得比光斑小得多——单分子定位（第 7 节）；
3. 放弃 (iii)：用条纹照明与样品频率相乘，把截止频率外的成分搬进通带——SIM（第 8 节）。

另一类方法（近场扫描光学显微）把探针放到样品表面几十纳米内，直接收集倏逝波，绕开了“远场”这个前提，本讲不讨论。

5 荧光、受激辐射与饱和

5.1 荧光分子的能级

定义：荧光分子的能级与荧光寿命

有机染料分子或荧光蛋白的电子能级之间间隔约 $2\text{ eV} \sim 3\text{ eV}$ （对应可见光）。每个电子能级上又叠加着一组间隔约 $0.1\text{ eV} \sim 0.2\text{ eV}$ 的振动物能级（分子中原子的振动）。记电子基态为 S_0 ，第一电子激发态为 S_1 （图 7(a)）。

- 吸收：分子吸收一个光子，从 S_0 的最低振动能级跃迁到 S_1 的某个较高振动能级。
- 振动弛豫：多余的振动能在约 1 ps 内以热的形式交给周围溶剂，分子落到 S_1 的最低振动能级。
- 荧光：分子从 S_1 自发地跃迁到 S_0 的某个振动能级，发出一个光子；随后再经振动弛豫回到 S_0 最低振动能级。因为吸收与发射之间损失了振动能，荧光波长比激发光长（斯托克斯位移）；因为发射可以终止于 S_0 的不同振动能级，荧光是一个宽的光谱带。
- 荧光寿命 τ ：处在 S_1 的分子单位时间内离开 S_1 的概率为 $1/\tau$ （包括发光与不发光两种途径）。常见染料与荧光蛋白 $\tau \approx 1\text{ ns} \sim 5\text{ ns}$ ，例如罗丹明 6G 在水中为 4.08 ns 、荧光素为 4.0 ns [25]。发光途径所占比例叫量子产率，好的染料接近 1。

定义：吸收截面与光子通量

光强为 I 、频率为 ν 的光，单位时间穿过单位面积的光子数（光子通量密度）为 $\Phi = I/(h\nu)$ 。一个分子单位时间吸收的光子数正比于 Φ ，写成

$$k_a = \sigma_a \Phi = \frac{\sigma_a I}{h\nu}, \quad (5.1)$$

比例系数 σ_a 具有面积量纲，叫吸收截面：可以想象分子是一个面积为 σ_a 的“靶”，落在靶上的光子都被吸收。化学中常用摩尔消光系数 ε （比尔定律 $I = I_0 10^{-\varepsilon c l}$ ， c 为摩尔浓度， l 为光程）。比较 $dI/dl = -\sigma_a n_m I$ （ n_m 为分子数密度）与 $dI/dl = -(\ln 10)\varepsilon c I$ ，得 $\sigma_a = (\ln 10)\varepsilon/(1000 N_A)$ ，即 $\sigma_a/\text{cm}^2 = 3.82 \times 10^{-21} \varepsilon/(\text{L mol}^{-1} \text{cm}^{-1})$ 。例如罗丹明 6G 在 530 nm 处 $\varepsilon = 1.16 \times 10^5 \text{ L mol}^{-1} \text{cm}^{-1}$ [26]， $\sigma_a = 4.4 \times 10^{-16} \text{ cm}^2$ ，约为分子几何截面（ $\sim 1\text{ nm}^2 = 10^{-14} \text{ cm}^2$ ）的百分之几。

5.2 自发辐射与受激辐射

定义：受激辐射

处在激发态的分子，除了自发发光，还可以在一个频率合适的入射光子“诱导”下跃迁到低能级，同时发出一个与入射光子频率、传播方向、偏振、相位都相同的光子：一个光子进去，两个相同的光子出来（图 7(a) 红色箭头）。这就是受激辐射，也是激光的工作原理。爱因斯坦 1917 年从热平衡的要求推出：对同一对能级、同一频率，受激辐射与吸收的单位光子通量下的跃迁

概率相等（两能级简并度相同时）。仿照式 (5.1)，受激辐射速率写成

$$k_{se} = \frac{\sigma_{se} I_{STED}}{h\nu_{STED}}, \quad (5.2)$$

σ_{se} 为受激辐射截面，它大致随该波长处荧光光谱的强度而变（荧光强的波长，受激辐射也强）；有机染料在荧光谱带的长波一侧的典型值为 $\sigma_{se} \sim 10^{-17} \text{ cm}^2$ [27]。它比第 5.1 节的吸收截面（ $\sim 4 \times 10^{-16} \text{ cm}^2$ ）小一个量级，与爱因斯坦关系并不矛盾：“两者相等”只对同一对能级、同一频率成立；分子的谱带很宽，吸收截面是在吸收峰处取值，而耗尽光工作在荧光谱的长波尾部，那里的跃迁强度小得多。

注：为什么耗尽光只“关”不“开”

STED 中耗尽光的波长选在荧光谱带的长波端（例如激发 640 nm、耗尽 775 nm）。(1) 此波长的光子能量低于 $S_0 \rightarrow S_1$ 吸收所需，基本不能把基态分子激发上去；(2) 受激辐射把分子送到 S_0 的较高振动能级，后者在约 1 ps 内弛豫到最低振动能级，来不及被同一束光“激发回去”。于是从 S_1 到 S_0 的耗尽是单向的： S_1 、 S_0 两个电子能级加上各自快速弛豫的振动能级，等效于一个“四能级系统”。另外，受激辐射出的光子与耗尽光同波长、同方向，用滤光片就能挡掉，探测器只收自发荧光。

5.3 速率方程

假设：单分子速率方程模型

1. 只考虑 S_0 、 S_1 两个电子态；振动弛豫（约 1 ps）比其他所有过程都快，可以认为瞬间完成。
2. 忽略三重态（分子从 S_1 偶尔转入的一种寿命长达微秒以上、不发荧光的“暗态”）、光漂白（第 6.4 节讨论其后果）。
3. 用 $p_1(t)$ 表示分子处在 S_1 的概率， $p_0 = 1 - p_1$ 。各跃迁都是随机事件，单位时间的跃迁概率（速率）为：激发 k_a （式 (5.1)），离开 S_1 的自发过程 $1/\tau$ ，受激辐射 k_{se} （式 (5.2)）。

在时间 dt 内， S_0 上的分子以概率 $k_a dt$ 被激发， S_1 上的分子以概率 $(1/\tau + k_{se})dt$ 回到 S_0 。于是

$$\frac{dp_1}{dt} = k_a(1 - p_1) - \left(\frac{1}{\tau} + k_{se}\right)p_1. \quad (5.3)$$

分子发出自发荧光光子的速率正比于 p_1 。

推导：激发饱和

只有激发光（ $k_{se} = 0$ ）、连续照明，稳态 $dp_1/dt = 0$ ：

$$p_1 = \frac{k_a}{k_a + 1/\tau} = \frac{I/I_{sat}}{1 + I/I_{sat}}, \quad I_{sat} = \frac{h\nu}{\sigma_a \tau}. \quad (5.4)$$

弱光（ $I \ll I_{sat}$ ）时 $p_1 \approx I/I_{sat}$ ，荧光与光强成正比——这就是“线性成像”的前提；强光时 $p_1 \rightarrow 1$ ，荧光不再增加，称为饱和。 I_{sat} 的含义：在这个光强下，激发速率恰好等于自发衰减速率 $1/\tau$ 。数值：罗丹明 6G， $\sigma_a = 4.4 \times 10^{-16} \text{ cm}^2$ 、 $\tau = 4.08 \text{ ns}$ 、 $\lambda = 530 \text{ nm}$ （ $h\nu = 3.75 \times 10^{-19} \text{ J}$ ），

$I_{\text{sat}} \approx 2.1 \times 10^5 \text{ W/cm}^2$ 。把 1 mW 聚焦到 $0.1 \mu\text{m}^2$ 就有 $1 \times 10^6 \text{ W/cm}^2$ ，已经饱和。

推导：受激辐射耗尽：脉冲情形

STED 常用脉冲光。设 $t = 0$ 时一个极短的激发脉冲使 $p_1 = p_1(0)$ ；紧接着一个宽度为 τ_p 、光强为 I (脉冲期间近似恒定) 的耗尽脉冲。脉冲期间没有激发光，式 (5.3) 变为 $dp_1/dt = -(1/\tau + k_{\text{se}})p_1$ ，解为

$$p_1(\tau_p) = p_1(0) e^{-\tau_p/\tau} e^{-k_{\text{se}}\tau_p}.$$

脉冲之后分子照常自发发光，发出的荧光光子数正比于 $p_1(\tau_p)$ 。与没有耗尽光时相比，荧光被乘上因子（剩余荧光概率）

$$\eta = e^{-k_{\text{se}}\tau_p} = \exp\left(-\frac{I}{I_s}\right), \quad (5.5)$$

$$I_s = \frac{h\nu_{\text{STED}}}{\sigma_{\text{se}}\tau_p} \quad (\text{脉冲耗尽的饱和光强}). \quad (5.6)$$

推导用了两个条件： $\tau_p \ll \tau$ (脉冲期间自发衰减可略， $e^{-\tau_p/\tau} \approx 1$ ，且这段时间里发出的少量荧光可略，或用时间门剔除：探测器只计耗尽脉冲结束之后到达的光子)； τ_p 远长于振动弛豫时间（约 1 ps，否则被送到 S_0 高振动能级的分子会被同一脉冲受激吸收而重新激发）。实际 τ_p 取 0.1–1 ns。 $I\tau_p$ 是脉冲的能流密度，所以式 (5.5) 也常写成 $\eta = \exp(-\sigma_{\text{se}}J/h\nu)$ ， J 为单个脉冲的能流密度。

推导：受激辐射耗尽：连续情形

激发光与耗尽光都连续照射，且激发很弱 ($k_a \ll 1/\tau$)。稳态时由式 (5.3)， $p_1 \approx k_a/(1/\tau + k_{\text{se}})$ ；与 $k_{\text{se}} = 0$ 时相比，

$$\eta = \frac{1}{1 + k_{\text{se}}\tau} = \frac{1}{1 + I/I_s},$$

$$I_s = \frac{h\nu_{\text{STED}}}{\sigma_{\text{se}}\tau} \quad (\text{连续耗尽}). \quad (5.7)$$

要点：饱和光强的统一含义

两种情形下 I_s 都是“受激辐射速率 $\times$ 分子可供耗尽的时间 = 1”的光强：脉冲情形可供耗尽的时间是 τ_p ，连续情形是寿命 τ 。 $I \gg I_s$ 时荧光几乎被完全“关掉”；脉冲时随 I 指数下降，连续时只按 $1/I$ 下降（图 7(b)）。这种强烈的非线性——荧光对耗尽光强度的依赖不是线性的——正是 STED 越过衍射极限的关键。上面三个解析式都已用式 (5.3) 的数值积分核对 (scripts/fig_levels.py)。

6 STED: 用环形光把光斑“削”小

6.1 原理

STED 显微镜是在共聚焦显微镜的基础上改造的。共聚焦显微镜是一台扫描显微镜：一个聚焦光斑在样品上逐点移动，每个位置记录一次荧光总量，组成图像；探测器前放一个小针孔，只让焦点附近发出的荧光通过（主要作用是挡掉焦面上下的杂散荧光，横向分辨率仍受衍射限制）。它

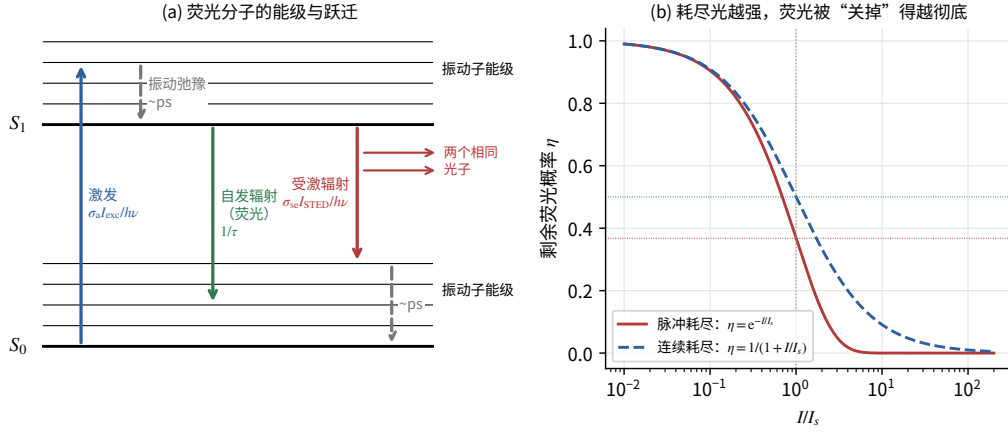


图 7 (a) 荧光分子的能级与四种过程 (示意): 激发、振动弛豫、自发辐射 (荧光)、受激辐射 (STED)。(b) 剩余荧光概率 η 随耗尽光强的变化: 脉冲 (式 (5.5)) 与连续 (式 (5.7))。两式均由速率方程 (5.3) 的数值积分核对 (scripts/fig_levels.py)。

的光斑由两束脉冲光叠加而成 [6, 8]:

1. 激发光: 普通聚焦光斑, 强度分布为 $H_{\text{exc}}(r)$ (式 (4.10)), 半高全宽 $w_0 \approx \lambda_{\text{exc}}/(2NA)$, 把光斑内的分子激发到 S_1 ;
2. 耗尽光: 紧接着到达的一束“环形”光, 中心强度为零, 四周很强。环上的分子被受激辐射强迫回到基态, 发不出荧光; 只有中心零点附近的分子保留激发态。

所以, “能发荧光的区域” 只剩环形光零点周围一个小区域。耗尽光越强, 被关掉的区域越向中心逼近, 这个小区域就越小。图 8 是 2000 年的实验结果: 荧光随耗尽光强度迅速下降, 光斑被显著压缩。

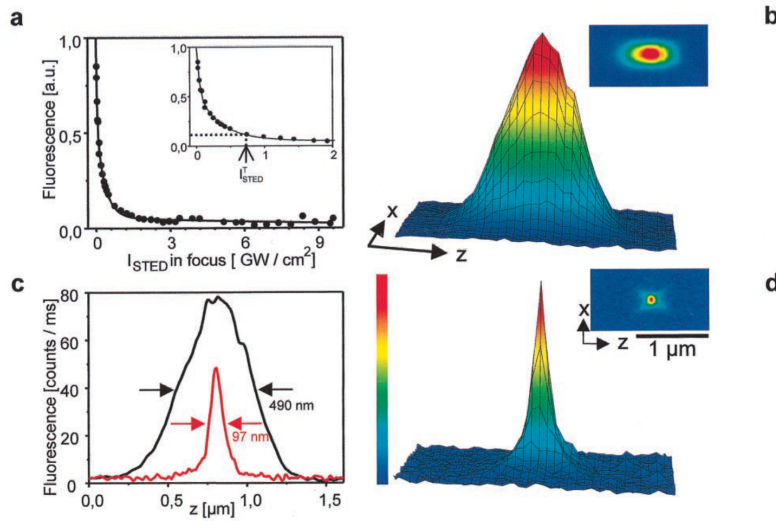


图 8 2000 年 STED 的早期实验。(a) 荧光随耗尽光强度 I_{STED} 非线性地下降 (插图中 I_{STED}^T 处剩余 10%), 与式 (5.5) 定性一致 (强光下有残余平台, 原文归因于振动弛豫的限制); (b)(d) 分别为关闭、打开耗尽光时荧光光斑的 xz 截面 (立体图, 插图为平面图); (c) 沿光轴的剖面, 半高全宽从 490 nm (黑) 降到 97 nm (红)。该实验的耗尽光图样设计成主要压缩轴向。(图源: Klar 等, PNAS 97, 8206 (2000), Fig. 2. Copyright (2000) National Academy of Sciences, U.S.A.)

6.2 环形光：涡旋相位与中心零点

怎样得到中心严格为零的光斑？最常用的办法是在物镜光瞳处加一块涡旋相位板：光瞳上方位角为 φ 的点，相位延迟 φ ，绕光轴一周相位增加 2π （图 9(a)）。

推导：焦点中心为什么是零、附近为什么 $\propto r^2$

由式 (4.4)，焦面上的振幅是光瞳函数的逆变换。加了相位板后光瞳函数为 $P(\mathbf{f})e^{i\varphi_f}$ (φ_f 是 $\mathbf{f}$ 的方位角)：

$$h_v(\mathbf{r}) = \iint P(\mathbf{f}) e^{i\varphi_f} e^{i2\pi\mathbf{f}\cdot\mathbf{r}} d^2f.$$

1. 中心： $\mathbf{r} = 0$ 时 $h_v(0) = \int_0^{NA/\lambda} f df \int_0^{2\pi} e^{i\varphi_f} d\varphi_f = 0$ 。光瞳上相对的两点 (φ_f 与 $\varphi_f + \pi$) 到达焦点时相位相反，两两抵消。这个零点由对称性保证，不依赖光强。
2. 中心附近：设 $\mathbf{r} = r(\cos\psi, \sin\psi)$, $r \ll \lambda/NA$ 。展开 $e^{i2\pi\mathbf{f}\cdot\mathbf{r}} \approx 1 + i2\pi fr \cos(\varphi_f - \psi)$ (小参量 $2\pi fr \leq 2\pi NA r/\lambda$)。第一项给出零 (上一步)。第二项中 $\cos(\varphi_f - \psi) = \frac{1}{2}[e^{i(\varphi_f - \psi)} + e^{-i(\varphi_f - \psi)}]$ ，与 $e^{i\varphi_f}$ 相乘后，只有 $e^{-i(\varphi_f - \psi)}$ 那一半对 φ_f 积分不为零 (得 $2\pi e^{i\psi}$)：

$$h_v(\mathbf{r}) \approx i2\pi r \cdot \frac{1}{2} \cdot 2\pi e^{i\psi} \int_0^{NA/\lambda} f^2 df = i \frac{2\pi^2}{3} \left(\frac{NA}{\lambda}\right)^3 r e^{i\psi}.$$

振幅正比于 r ，所以强度正比于 r^2 ：中心附近是一个抛物线形的“碗底”。

完整的径向分布需要数值积分或贝塞尔函数 (附录 C)。结果是：峰值在 $r = 0.39\lambda/NA$ 处，中心附近

$$I_{\text{STED}}(r) \approx I_{\text{max}} \left(\frac{r}{r_0}\right)^2, \quad r_0 = 0.243 \frac{\lambda_{\text{STED}}}{NA}, \quad (6.1)$$

I_{max} 是环上的峰值强度， r_0 是“抛物线外推到 I_{max} ”处的半径 (比峰值半径小)。环的峰值强度只有同功率普通聚焦光斑中心强度的 26%。

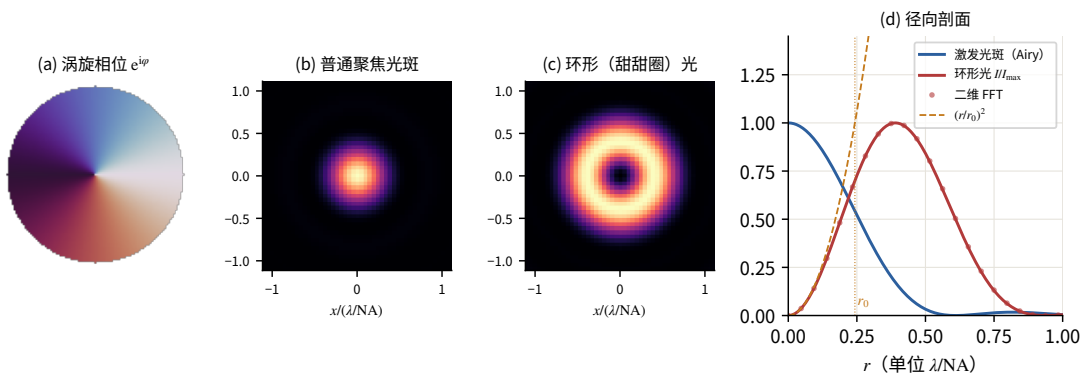


图 9 涡旋相位板产生的环形光 (标量近似, 均匀照明的圆形光瞳)。(a) 光瞳上的相位 φ (颜色表示 $0-2\pi$)；(b) 不加相位板的聚焦光斑；(c) 加相位板后的环形光 (二维 FFT 直接计算)；(d) 径向剖面：环形光的闭式结果 (附录 C, 红线) 与二维 FFT (红点) 一致，中心附近与抛物线 $(r/r_0)^2$ 重合 ($r_0 = 0.243\lambda/NA$)。scripts/fig_sted.py 核对了中心严格为零、峰值位置与抛物线系数。

注：偏振的影响

高 NA 物镜的焦点处，光的偏振方向会随入射角改变（标量近似不再严格）。若用线偏振的耗尽光，焦点中心会残留可观的纵向电场，零点“填不满”。实验中用与涡旋方向匹配的圆偏振光，中心的各电场分量仍相互抵消 [21, 28]。

6.3 有效点扩散函数的宽度**假设：STED 宽度公式的假设**

1. 激发光很弱（不饱和），激发概率正比于 $H_{\text{exc}}(r)$ ；
2. 耗尽脉冲满足式 (5.5) 的条件，剩余荧光概率 $\eta(r) = \exp[-I_{\text{STED}}(r)/I_s]$ ；
3. 探测器收集分子发出的全部荧光（不管分子在哪里）。实际的 STED 显微镜前有共聚焦针孔，有效 PSF 应再乘以探测 PSF H_{det} ；只要针孔足够大（约一个 Airy 斑直径以上）， H_{det} 在 H_{eff} 的宽度（几十纳米）内近似为常数，可以略去。这也是为什么在 STED 中针孔对分辨率的作用很小，主要用来压低远离中心处的微弱荧光与杂散光。

在这些条件下，位于 $\mathbf{r}$ 的分子在扫描到原点时贡献的荧光正比于 $H_{\text{exc}}(r)\eta(r)$ 。这个乘积就是 STED 的有效点扩散函数

$$H_{\text{eff}}(r) = H_{\text{exc}}(r) \exp\left[-\frac{I_{\text{STED}}(r)}{I_s}\right]. \quad (6.2)$$

图像仍是分子分布与 H_{eff} 的卷积。

推导：宽度公式

为了得到解析式，做两个近似：

1. 用半高全宽相同的高斯函数近似激发光斑： $H_{\text{exc}}(r) \approx \exp(-4 \ln 2 r^2/w_0^2)$ ， $w_0 = 0.51\lambda_{\text{exc}}/\text{NA}$ 。可以验证 $r = w_0/2$ 时它等于 1/2。
2. 用式 (6.1) 的抛物线近似环形光。这一近似在 $r \ll r_0$ 时成立；耗尽光足够强时 H_{eff} 只在 $r \ll r_0$ 内不为零，近似自治。

记 $\zeta = I_{\text{max}}/I_s$ 。代入式 (6.2)，两个指数合并：

$$H_{\text{eff}}(r) \approx \exp\left[-r^2\left(\frac{4 \ln 2}{w_0^2} + \frac{\zeta}{r_0^2}\right)\right].$$

它仍是高斯函数。令它等于 1/2，半高全宽 w 满足 $(w/2)^2(4 \ln 2/w_0^2 + \zeta/r_0^2) = \ln 2$ ，即

$$\boxed{w = \frac{w_0}{\sqrt{1 + \beta\zeta}}}, \quad (6.3)$$

$$\beta = \frac{w_0^2}{4 \ln 2 r_0^2} = 1.62 \left(\frac{\lambda_{\text{exc}}}{\lambda_{\text{STED}}}\right)^2.$$

最后一步代入了 $w_0 = 0.514\lambda_{\text{exc}}/\text{NA}$ 与 $r_0 = 0.243\lambda_{\text{STED}}/\text{NA}$ 。

检验：极限、量纲与数值核对

- $\zeta = 0$ (无耗尽光): $w = w_0$, 回到衍射极限。 $\zeta \gg 1$: $w \approx w_0/\sqrt{\beta\zeta} \propto I_{\max}^{-1/2}$, 原则上可以任意小——分辨率不再受衍射限制, 而受可用光强与分子耐受力的限制。
- 量纲: ζ 、 β 都无量纲, w 与 w_0 同量纲。
- 与精确数值比较: 用精确的 Airy 激发斑与精确的环形光 (附录 C), 数值求 H_{eff} 的半高全宽 (图 10(b) 红点)。 $\lambda_{\text{STED}} = \lambda_{\text{exc}}$ 时, 式 (6.3) 在 $\zeta \geq 1$ 时偏差 $\leq 6\%$, $\zeta \geq 3$ 时 $\leq 3.3\%$, $\zeta \geq 10$ 时 $\leq 1.3\%$ 。偏差来自小 ζ 时抛物线近似超出适用范围。
- 常见写法: 文献中常写成 $w \approx \lambda/(2\text{NA}\sqrt{1+I_{\max}/I_s})$ [27, 29]。它与式 (6.3) 的差别只是把 β 吸收进 I_s 的定义 (典型的 $\lambda_{\text{STED}}/\lambda_{\text{exc}} \approx 1.2$ 时 $\beta \approx 1.1$, 本来就接近 1)。Westphal 与 Hell 用驻波型耗尽光的模型给出 $w \approx 0.45\lambda/(\text{NA}\sqrt{1+\zeta})$, 并把 I_s 定义为荧光降到 $1/e$ 时的光强 [30], 与式 (5.5) 一致。比较不同文献的 I_s 数值时, 要注意它们的定义可能相差这样的因子。
- 饱和光强的数值核对: Rittweger 等人报告金刚石氮-空位色心 $\sigma_{\text{se}} \approx 1.2 \times 10^{-17} \text{ cm}^2$ 、耗尽脉冲宽 3.2 ns 、 $\lambda_{\text{STED}} = 775 \text{ nm}$ 、 $I_s = 6.6 \text{ MW/cm}^2$ [9]; 由式 (5.6), $I_s = h\nu/(\sigma_{\text{se}}\tau_p) = 6.7 \text{ MW/cm}^2$, 一致。注意 NV 色心的寿命为 11.6 ns , $\tau_p/\tau = 0.28$, “ $\tau_p \ll \tau$ ”只是勉强满足; 但脉冲期间的自发衰减因子 $e^{-\tau_p/\tau}$ 在有、无耗尽光时相同, 在 η 的比值中相消, 所以式 (5.5) 仍可近似使用 (若不作时间门控, 还要考虑脉冲期间发出的荧光, 见例题 2)。

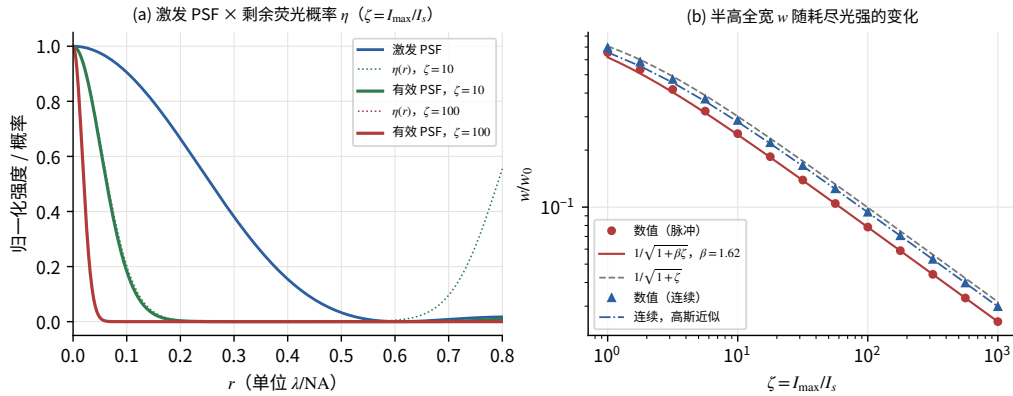


图 10 STED 有效点扩散函数 ($\lambda_{\text{STED}} = \lambda_{\text{exc}}$)。 (a) 激发 PSF (蓝) 乘以剩余荧光概率 $\eta(r)$ (点线) 得到有效 PSF (粗线), $\zeta = 10$ 与 100 。 $r > 0.6\lambda/\text{NA}$ 处 η 回升, 是因为环形光在那里有一个暗环, 但激发光在那里本来就弱。 (b) 半高全宽随 ζ 的变化: 数值 (点) 与式 (6.3) (实线, $\beta = 1.62$); 灰虚线为 $\beta = 1$ 的常见写法; 蓝色三角为连续耗尽 ($\eta = 1/(1+I/I_s)$) 的数值结果, 点划线为其高斯近似 (例题 2)。scripts/fig_sted.py。

要点：STED 越过衍射极限的本质

耗尽光的强度分布 $I_{\text{STED}}(r)$ 本身受衍射限制 (只含 $|f| \leq 2\text{NA}/\lambda$ 的空间频率), 但 $\exp(-I_{\text{STED}}/I_s)$ 是非线性函数。由泰勒展开 $e^{-x} = 1 - x + x^2/2 - \dots$, I_{STED}^2 、 I_{STED}^3 等项的频谱分别延伸到 $2f_c$ 、 $3f_c \dots$ (两函数乘积的频谱是两频谱的卷积, 宽度相加)。所以 H_{eff} 含有任意高的空间频率, 第 4.6 节的前提 (i) 被打破了。

交互演示：STED 有效点扩散函数。用滑块改变 $\zeta = I_{\max}/I_s$ 与耗尽光波长，看有效 PSF（红）怎样收缩；读数比较数值半高全宽与式 (6.3)。右（或下）方是相距可调的两个分子在共聚焦与 STED 下的像。也可切换为连续耗尽对比。

网页版中此处为可调参数的交互动画（演示编号 superres-sted）。

6.4 需要多强的光？实际限制

由式 (6.3)，要把分辨率从 $w_0 \approx 230 \text{ nm}$ 提高到 30 nm ，需要 $\beta\zeta \approx (230/30)^2 - 1 \approx 60$ ，即环上峰值强度约为饱和光强的 50 倍以上。代入典型参数（例题 2），脉冲 STED 的平均功率约为几十毫瓦，峰值强度达每平方厘米几百兆瓦。随之而来的限制：

1. 光漂白与光损伤：分子在强光下被反复激发、受激，进入三重态或发生光化学反应而永久失去荧光；活细胞也受不了太强的光。
2. 零点不完美：光学像差、偏振误差使中心残留一点强度 $\epsilon I_{\max}$ ，它会关掉中心本该发光的分子，降低信号（例题 2 定量讨论）。
3. 降低 I_s ：式 (5.6)、(5.7) 说明 $I_s \propto 1/(\sigma \times \text{可供耗尽的时间})$ 。若“关闭”不是靠受激辐射，而是靠把分子切换到一个寿命长达毫秒到秒的暗态（可光开关的荧光蛋白），所需光强可降低到每平方厘米几瓦到几千瓦，这类方法统称 RESOLFT（reversible saturable optical fluorescence transitions，“可逆可饱和的光学荧光跃迁”：用一束光把分子可逆地切换到不发光态并使之饱和）[27]。

作为参照：Westphal 与 Hell 用约 800 MW/cm^2 的环形耗尽光（ $\zeta \approx 36$ ）得到 $30\text{--}35 \text{ nm}$ 的光斑[30]；在耐受强光的金刚石氮-空位色心上，用 8.6 GW/cm^2 的一维耗尽光达到了 5.8 nm [9]。生物样品中受光漂白限制，STED 的分辨率一般为几十纳米。

7 单分子定位：测“中心”比测“宽度”准得多

7.1 前置：随机变量、方差与泊松分布

定义：期望、方差与标准差

随机变量 X 以概率密度 $p(x)$ 取值（ $p(x)dx$ 是 X 落在 x 到 $x+dx$ 之间的概率）。它的期望（平均值）与方差为

$$\langle X \rangle = \int xp(x) dx,$$

$$\text{Var } X = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2,$$

标准差 $\sqrt{\text{Var } X}$ 刻画 X 偏离平均值的典型幅度。对离散变量，把积分换成求和。

要点：方差的三条性质

1. $\text{Var}(cX) = c^2 \text{Var} X$ (c 为常数)，由定义直接可得。
2. 若 X 、 Y 相互独立（一个的取值不影响另一个的概率分布），则 $\langle XY \rangle = \langle X \rangle \langle Y \rangle$ ，于是

$$\begin{aligned}\text{Var}(X + Y) &= \text{Var} X + \text{Var} Y + 2 \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \\ &= \text{Var} X + \text{Var} Y,\end{aligned}$$

交叉项因独立而等于 $\langle X - \langle X \rangle \rangle \langle Y - \langle Y \rangle \rangle = 0$ 。

3. 由前两条： N 个独立、同分布（方差 σ^2 ）的随机变量的平均值 $\bar{X} = \frac{1}{N} \sum X_j$ ，方差为 $\frac{1}{N^2} \cdot N\sigma^2$ ，即

$$\text{Var} \bar{X} = \frac{\sigma^2}{N}, \quad \text{标准差} = \frac{\sigma}{\sqrt{N}}. \quad (7.1)$$

定义：两个常用分布

高斯（正态）分布： $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-x_0)^2/(2\sigma^2)}$ ，期望 x_0 ，方差 σ^2 ；它的半高全宽为 $2\sqrt{2\ln 2}\sigma = 2.355\sigma$ 。

均匀分布： X 等可能地落在 $[-a/2, a/2]$ 内， $p = 1/a$ ，期望为零，方差

$$\text{Var} X = \frac{1}{a} \int_{-a/2}^{a/2} x^2 dx = \frac{a^2}{12}. \quad (7.2)$$

推导：泊松分布：光子计数的涨落

在一段时间 t 内数到达探测器的光子。把 t 分成 J 个极短的小段，每段内到达一个光子的概率 q 很小 ($q \ll 1$)，各段互相独立，平均光子数 $\bar{N} = Jq$ 。每一小段的计数 n_j 只取 0 或 1，这种变量叫伯努利变量： $\langle n_j \rangle = q$ ， $\langle n_j^2 \rangle = q$ ，所以 $\text{Var} n_j = q(1 - q)$ 。 J 个独立伯努利变量之和服从二项分布，由独立性其方差为 $Jq(1 - q)$ 。这里

$$\text{Var} N = J(q - q^2) \rightarrow \bar{N} \quad (q \rightarrow 0).$$

即光子数的方差等于其平均值，相对涨落 $1/\sqrt{\bar{N}}$ 。取 $J \rightarrow \infty$ 的极限还可得到完整的分布 $P(N) = \bar{N}^N e^{-\bar{N}} / N!$ （泊松分布），本讲只用到它的方差。

7.2 思想：让分子一个一个地亮

一个孤立荧光分子的像是一个宽约 250 nm 的光斑。如果视野里只有它一个，我们就知道“这个光斑来自一个点”；光斑的中心可以测得比光斑本身的宽度准得多，正如一堆散落的弹孔虽然很散，平均位置却可以算得很准。困难在于生物样品中的分子很密，许多光斑重叠在一起。2006 年 PALM、FPALM 与 STORM 的共同办法是：用可开关的荧光分子标记样品，让它们大部分时间处于不发光的“关”态；每一帧只随机地“打开”极少数分子，使它们的光斑彼此不重叠；把这些分子逐个定位，然后关闭（或漂白）；重复成千上万帧，把所有定位点画在一张图上 [13–15]。开关机制有多种：PALM 用光激活荧光蛋白（紫外光把它从暗态激活）[13]；STORM 用 Cy3–Cy5 染料对，红光把 Cy5 关闭、绿光借助邻近的 Cy3 把它重新打开 [15]。

要点：单分子定位绕过衍射极限的方式

它没有改变点扩散函数，而是利用了先验知识——每个光斑来自一个点——并把“同时分辨两个靠近的分子”变成“在不同时刻分别定位它们”。第 4.6 节的前提 (ii) 被打破了。

7.3 定位精度：光子统计**假设：理想定位模型**

1. 分子位于 x_0 （只看 x 方向； y 方向完全类似）。它发出、被探测到的每个光子在像面上的落点（折算回样品面）是一个随机变量，概率密度就是归一化的点扩散函数。
2. 各光子的落点相互独立（荧光分子每次发光都独立）。
3. 先忽略相机像素和背景光。
4. 点扩散函数用高斯函数近似，标准差为 s （理由见下面的注）。

推导： $\sigma_{\text{loc}} = s/\sqrt{N}$

探测到 N 个光子，落点为 $x_1, \dots, x_N$ ，每个的期望为 x_0 、方差为 s^2 。最自然的位置估计是它们的平均（质心） $\hat{x} = \frac{1}{N} \sum x_j$ 。 $\langle \hat{x} \rangle = x_0$ ，即估计的期望等于真值，这样的估计叫无偏估计；由式 (7.1)，

$$\sigma_{\text{loc}} = \sqrt{\text{Var } \hat{x}} = \frac{s}{\sqrt{N}}. \quad (7.3)$$

即定位误差 = 平均值的标准误。 $s \approx 100 \text{ nm}$ 、 $N = 1000$ 时 $\sigma_{\text{loc}} \approx 3 \text{ nm}$ 。

注：高斯近似与 $s = 0.21\lambda/\text{NA}$

真实的点扩散函数是 Airy 斑 (4.10)。用高斯函数对它做最小二乘拟合，得 $s \approx 0.21\lambda/\text{NA}$ （按半高全宽匹配则为 $0.22\lambda/\text{NA}$ ；scripts/verify_numeric.py）。需要注意：Airy 斑的外环衰减很慢（ $\propto r^{-3}$ ），它的二阶矩 $\int x^2 H \text{d}^2r$ 是发散的，所以对真实 Airy 斑直接取“所有光子的平均”并不好用；实际中在一个有限区域内用模型函数拟合光斑。可以证明，对理想 Airy 斑，任何无偏估计的精度都不可能好于 $\lambda/(2\pi\text{NA}\sqrt{N}) \approx 0.16\lambda/(\text{NA}\sqrt{N})$ （这是统计学中“克拉默-拉奥下限”的结果 [31]，附录 D）。综合起来，定位精度写成 $c\lambda/(\text{NA}\sqrt{N})$ 时，系数 c 取决于估计方法：最优估计（下面注中的极大似然估计）接近下限 0.16；用高斯 PSF 模型、质心法估计（本节的推导）约为 0.21；常用的高斯最小二乘拟合还要乘约 $\sqrt{16/9}$ （第 7.4 节注），约为 0.28。对不加截断的理想 Airy 斑，质心法不收敛。

7.4 像素化与背景**推导：像素化：多出 $a^2/12$**

相机像素的边长（折算到样品面）为 a 。一个光子落在某个像素里，我们只知道它在这个像素内，于是把它记为像素中心。记录位置 = 真实落点 + 量化误差 u 。当点扩散函数比像素宽（ $s \gtrsim a/2$ ）、且分子相对像素格点的位置是随机的时， u 近似在 $[-a/2, a/2]$ 内均匀分布，并且

与落点的高斯涨落近似独立。由方差相加与式 (7.2)，每个记录位置的方差为 $s^2 + a^2/12$ ，于是

$$\sigma_{\text{loc}}^2 = \frac{s^2 + a^2/12}{N}. \quad (7.4)$$

图 11(b) 的蒙特卡罗模拟（直接产生光子、按像素记录、取平均）在 $a = 0, s, 2s$ 时都与式 (7.4) 相符到 3% 以内。（对固定的亚像素位置， $a = 2s$ 时单个光子的方差会随位置有 $\pm 3\%$ 的起伏，对位置平均后才严格等于 $s^2 + a^2/12$ ；scripts/verify_numeric.py 有核对。）

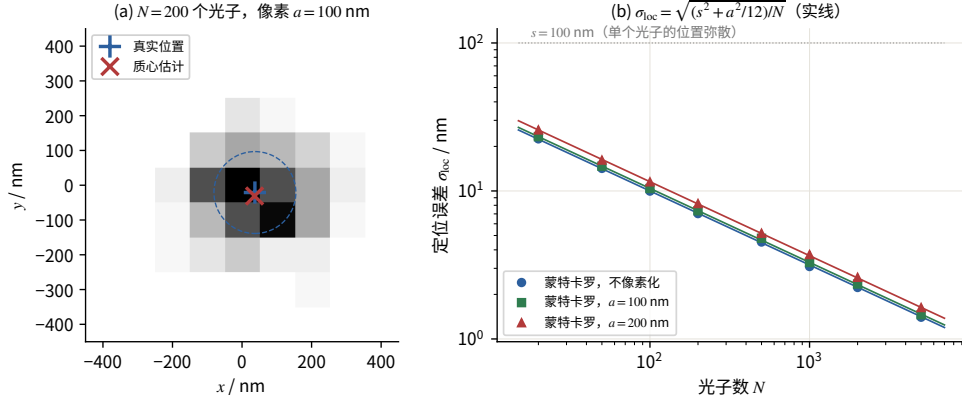


图 11 单分子定位的光子统计。(a) 一个分子的 200 个光子落在 100 nm 像素上形成的像（模拟），蓝十字为真实位置，红叉为质心估计，虚线圆直径为光斑的半高全宽 $2.355s$ 。(b) 质心定位误差随光子数 N 的变化：点为蒙特卡罗（每点 1500–4000 次），线为式 (7.4)， $s = 100$ nm。scripts/fig_localization.py。

要点：加上背景：Thompson 公式

实际图像上还有均匀的背景光（样品自发荧光、杂散光、相机读出噪声），设每个像素的背景计数涨落的方差为 b^2 （纯泊松背景时 b^2 等于每像素平均背景光子数）。Thompson、Larson 与 Webb 对“用高斯函数做最小二乘拟合”的定位方法给出 [32]

$$\sigma_{\text{loc}}^2 = \frac{s^2 + a^2/12}{N} + \frac{8\pi s^4 b^2}{a^2 N^2}. \quad (7.5)$$

第一项是上面推出的光子统计项，第二项来自背景。

推导：背景项的标度

定位相当于在光斑周围半径 $R_c \sim 2s$ 的区域内求“计数加权的平均位置” $\hat{x} = \sum_p x_p n_p / N$ (n_p 为计数，像素坐标 x_p 从分子位置即区域中心量起)。每个像素的背景计数独立地涨落，方差 b^2 ，它给分子 $\sum_p x_p n_p$ 带来的方差为 $b^2 \sum_p x_p^2$ 。区域内有 $\pi R_c^2 / a^2$ 个像素，把求和换成积分： $\sum_p x_p^2 \approx \frac{1}{a^2} \iint x^2 dA = \frac{\pi R_c^4}{4a^2}$ 。除以 N^2 ，

$$\sigma_{\text{bg}}^2 \sim \frac{b^2}{N^2} \cdot \frac{\pi R_c^4}{4a^2} \Big|_{R_c=2s} = \frac{4\pi s^4 b^2}{a^2 N^2}.$$

这与式 (7.5) 第二项的形式 $s^4 b^2 / (a^2 N^2)$ 一致；系数 8π 要由最小二乘拟合的具体权重算出。两点物理含义：(1) 背景项 $\propto 1/N^2$ ，光子少时它占主导，光子多时第一项占主导；(2) 区域取得越大，背景项越大。另外，背景光子平均分布在整个区域、平均位置在区域中心，它们计入分母后会把质心拉向区域中心，造成系统偏差，并按同一比例 $N/(N + n_b)$ (n_b 为区域内背景光

子数) 压缩信号光子的涨落。这两点都说明不能对整个视野求质心，而要在光斑附近拟合（或扣除背景）。

注：更精细的修正

式 (7.5) 是近似结果。Mortensen 等人 [33] 指出：最小二乘拟合本身并不是最优的，其方差比式 (7.5) 的第一项还要大一个因子（约 16/9），而极大似然估计（选取使观测到的这组光子数据出现概率最大的参数值）可以接近理论下限。电子倍增 CCD (EMCCD) 在读出前把每个光电子随机地放大成成百上千个电子，放大倍数本身有涨落，使信号的方差增大一倍（过量噪声因子 $F^2 \approx 2$ ），等效于有效方差加倍。比较不同实验的定位精度时，要弄清用的是哪种相机和哪种估计方法。

7.5 图像分辨率：定位精度与标记密度

单分子定位得到的是一堆点。由它们重建的图像能分辨多细的结构，取决于两件事：

1. 定位精度：每个点都有 σ_{loc} 的随机偏差，一个点状物在重建图上变成标准差 σ_{loc} 的“云”，其半高全宽 $2.355\sigma_{\text{loc}}$ 。这相当于新的点扩散函数。
2. 标记密度：结构只在有标记分子的地方被“采样”。即使每个点都定位得极准，如果标记稀疏，也看不出连续的形状（图 12(c)）。

要点：奈奎斯特采样判据

一个只含空间频率 $\leq f_{\text{max}}$ 的函数，只要采样间隔 $\Delta \leq 1/(2f_{\text{max}})$ ，就能由样点唯一确定；采样间隔更大时，频率为 f 的成分与频率为 $1/\Delta - f$ 的成分在样点上取值相同，无法区分（混叠，推导见附录 E）。要分辨周期为 d 的结构 ($f_{\text{max}} = 1/d$)，相邻标记的间距应不超过 $d/2$ [34]。二维中，这要求标记的面密度至少约 $(2/d)^2$ ： $d = 20 \text{ nm}$ 时为每平方微米 10^4 个分子。

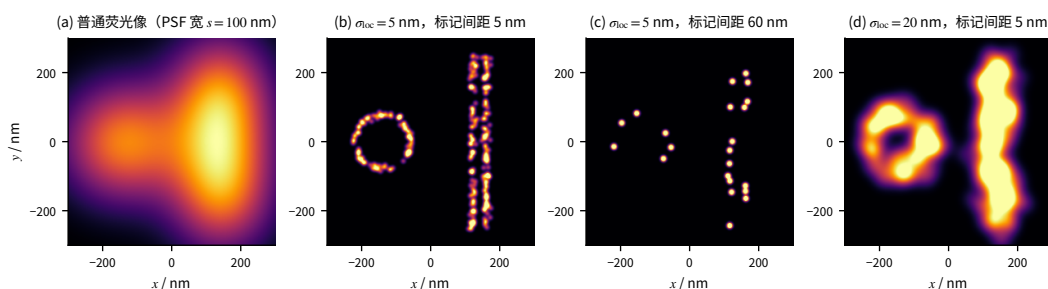


图 12 单分子定位重建（模拟）：半径 80 nm 的环与两条相距 40 nm 的平行线。(a) 普通荧光像；(b) 定位精度 5 nm、标记平均间距 5 nm：清楚；(c) 精度相同但标记间距 60 nm：看不出环与线的形状；(d) 标记密集但精度 20 nm：两条线合在一起（中点强度为峰值的 99%）。每个定位点画成标准差为 σ_{loc} 的高斯斑。(d) 的 99% 是这一组有限样本的结果；按解析式，两条相距 40 nm 的线各自被展宽为标准差约 $\sqrt{2} \times 20 \text{ nm}$ （定位误差与绘制宽度叠加）后已合为单峰。scripts/fig_nyquist.py。

注：代价：时间

每帧只能让很稀疏的分子发光（例题 3 给出估算：每平方微米约 0.26 个），而样品需要每平方微米上万个标记，所以一幅图通常要上万帧、几分钟。单分子定位用时间换取了空间分辨率，因而不适合观察快速运动的结构。

交互演示：随机开关的单分子定位。左：相机的当前一帧（像素 100 nm），蓝圈是本帧被激活分子的真实位置；右：累积的定位图（可切换为普通荧光像对比）。每个光子都按“高斯落点 + 像素中心记录”模拟，定位用质心。调大平均光子数 $\bar{N}$ ，图像变锐；调大激活概率，采集变快，但重叠光斑造成的错误定位（红点）增多。判断光斑是否重叠用的是分子的真实位置（“上帝视角”），阈值为 2.5s。读数比较实测定位误差与式 (7.4)。

网页版中此处为可调参数的交互动画（演示编号 superres-palm）。

7.6 MINIFLUX：用“零点”定位

式 (7.4) 说明，用相机定位时，精度受光斑宽度 $s \sim 100$ nm 的限制，要达到 1 nm 需要上万个光子。2017 年提出的 MINIFLUX（名字取自 minimal photon fluxes，“最少的光子流”）反其道而行：不用分子发出的光斑找中心，而是用 STED 中那种中心为零的环形激发光去“探”分子 [19]。分子越靠近零点，被激发得越少、发出的光子越少；把零点依次放在分子附近几个已知位置，比较各处的光子数，就能推出分子位置。因为零点附近强度 $\propto r^2$ 变化很陡，位置的微小变化会引起光子数比例的显著变化。图 13 是一维的情形：零点先后放在相距 L 的两处，在两个零点之间 ($|x_m| < L/2$)，两次光子数的比例随分子位置单调变化， L 越小变化越陡；在两零点之外比例会折返，同一比例对应两个位置。因此 MINIFLUX 要先用相机粗定位，把分子限制在 L 的范围内，实际实现中还会加一次零点在中心的曝光来排除歧义 [19]。 L 可以取得远小于光斑尺寸（几十纳米），所以同样的光子数下精度高得多（定量结果见例题 3）。Balzarotti 等人取 $L = 50$ nm、每个分子约一千个光子，定位精度达到约 1 nm，他们在 DNA 折纸（用 DNA 链自组装成的纳米结构，可以把荧光分子安放在设计好的位置上）上分辨了相距 6 nm 的分子 [19]，后来在细胞中实现了 1–3 nm 的分辨率 [20]。

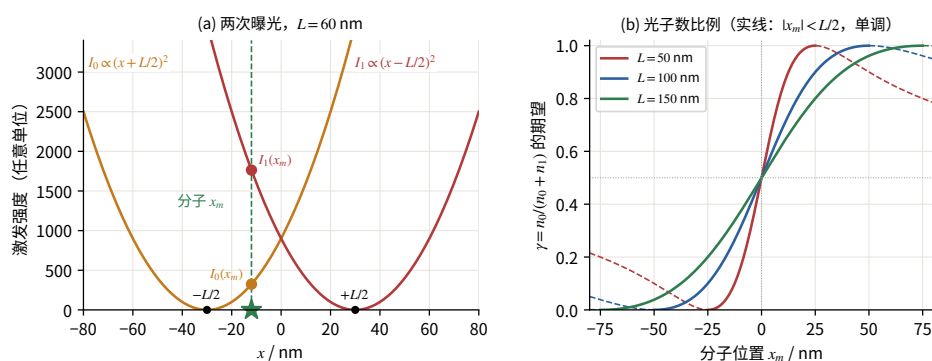


图 13 MINIFLUX 的一维原理（自绘，按 $I \propto (x - x_p)^2$ 计算）。(a) 激发强度的零点先后放在 $x = -L/2$ 与 $+L/2$ ，位于 x_m 的分子两次受到的激发强度 $I_0(x_m)$ 、 $I_1(x_m)$ 不同，两次的平均光子数与之成正比；(b) 光子数比例 $\gamma = n_0/(n_0 + n_1)$ 的期望：在两零点之间 ($|x_m| < L/2$ ，实线) 随分子位置单调变化，由测得的比例可反推位置， L 越小曲线越陡；零点之外（虚线）曲线折返，需先粗定位（在 $x_m = 0$ 处斜率由脚本核对为 $2/L$ ）。scripts/fig_miniflux.py。

8 结构光照明：把高频“搬”进通带

8.1 莫尔条纹

把两块细密的网纱叠在一起，会看到粗大的、缓缓移动的波纹，这就是莫尔条纹。它的数学本质是积化和差：

$$\begin{aligned} & \cos(2\pi \mathbf{f}_1 \cdot \mathbf{r}) \cdot \cos(2\pi \mathbf{f}_2 \cdot \mathbf{r}) \\ &= \frac{1}{2} \cos[2\pi(\mathbf{f}_1 + \mathbf{f}_2) \cdot \mathbf{r}] + \frac{1}{2} \cos[2\pi(\mathbf{f}_1 - \mathbf{f}_2) \cdot \mathbf{r}]. \end{aligned} \quad (8.1)$$

两个条纹相乘（例如光依次透过两块光栅，透过率相乘），产生和频与差频。若 $\mathbf{f}_1$ 与 $\mathbf{f}_2$ 相近，差频 $|\mathbf{f}_1 - \mathbf{f}_2|$ 很小，对应粗大的条纹。即使两组原始条纹都细得看不清，它们的差频条纹也可能看得清。例如两组周期都为 Λ 、夹角为 Θ 的条纹， $|\mathbf{f}_1 - \mathbf{f}_2| = 2 \sin(\Theta/2)/\Lambda$ ，莫尔条纹周期为 $\Lambda/[2 \sin(\Theta/2)]$ ，夹角越小越粗：周期 200 nm、夹角 5° 的两组条纹产生周期 2.3 μm 的莫尔条纹。

8.2 结构光照明显微 (SIM)

假设：线性 SIM 的设定

1. 宽场荧光显微镜：整个视野同时照明，相机成像；荧光对照明光强是线性的 ($I \ll I_{\text{sat}}$ ，式 (5.4))。
2. 照明光不是均匀的，而是正弦条纹： $I_{\text{ill}}(\mathbf{r}) = I_0[1 + \mu \cos(2\pi \mathbf{f}_0 \cdot \mathbf{r} + \phi)]$ ， $\mu \leq 1$ 为调制度， ϕ 为相位，二者都可控制。

推导：照明条纹怎样产生

把一束激光分成两束，经物镜以 $\pm\theta$ 角（在样品介质中）射到样品上。两束平面波 $U_{\pm} = e^{ik_z z} e^{\pm i2\pi(n \sin \theta / \lambda_{\text{exc}})x}$ 相干叠加（偏振都垂直于入射面，使电场方向相同、干涉对比度最大），

$$|U_+ + U_-|^2 = 2 + 2 \cos\left(2\pi \cdot \frac{2n \sin \theta}{\lambda_{\text{exc}}} x\right),$$

得到 $\mu = 1$ 的条纹，空间频率 $f_0 = 2n \sin \theta / \lambda_{\text{exc}}$ 。两束光都必须通过物镜， $n \sin \theta \leq \text{NA}$ ，所以

$$f_0 \leq \frac{2\text{NA}}{\lambda_{\text{exc}}}. \quad (8.2)$$

照明条纹本身也受衍射限制，最细只能到截止频率附近。

推导：频率搬移

发光强度 $\mathcal{E}(\mathbf{r})$ 是荧光分子密度 $C(\mathbf{r})$ 与照明强度之积。把余弦写成两个复指数，

$$\begin{aligned} \mathcal{E}(\mathbf{r}) &= C(\mathbf{r}) I_{\text{ill}}(\mathbf{r}) \\ &= I_0 \left[C + \frac{\mu}{2} e^{i\phi} C e^{i2\pi \mathbf{f}_0 \cdot \mathbf{r}} + \frac{\mu}{2} e^{-i\phi} C e^{-i2\pi \mathbf{f}_0 \cdot \mathbf{r}} \right]. \end{aligned}$$

由调制性质（第 3.4 节第 4 条），各项的频谱为

$$\tilde{\mathcal{E}}(\mathbf{f}) = I_0 \left[\tilde{C}(\mathbf{f}) + \frac{\mu}{2} e^{i\phi} \tilde{C}(\mathbf{f} - \mathbf{f}_0) + \frac{\mu}{2} e^{-i\phi} \tilde{C}(\mathbf{f} + \mathbf{f}_0) \right]. \quad (8.3)$$

相机得到的像的频谱是 $\text{OTF}(\mathbf{f})\tilde{\mathcal{E}}(\mathbf{f})$ （式 (4.6)），只在 $|\mathbf{f}| \leq f_c$ 内不为零。看第二项：像的频谱在 $\mathbf{f}$ 处含有样品在 $\mathbf{f}' = \mathbf{f} - \mathbf{f}_0$ 处的成分。所以，凡是满足 $|\mathbf{f}' + \mathbf{f}_0| \leq f_c$ 的样品频率 $\mathbf{f}'$ ，都在这一项中被“搬”进了通带。第三项同理：样品频率 $\mathbf{f}'$ 出现在像频谱的 $\mathbf{f}' - \mathbf{f}_0$ 处，凡 $|\mathbf{f}' - \mathbf{f}_0| \leq f_c$ 的都能被看到。图 15(a)：样品的频率 $\mathbf{f}_s$ 在通带圆外，但 $\mathbf{f}_s - \mathbf{f}_0$ 在圆内——这正是式 (8.1) 的差频，即莫尔条纹（图 14）。

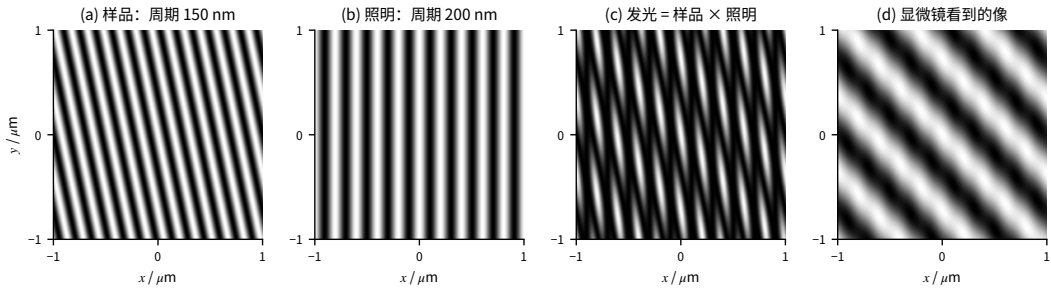


图 14 莫尔条纹把看不见的细节带进像中（模拟， $\text{NA} = 1.4$ 、发射波长 520 nm，截止周期 186 nm）。(a) 样品：周期 150 nm 的荧光条纹（方向 12° ），超出截止频率，普通显微镜下像是均匀的；(b) 照明条纹：周期 200 nm；(c) 发光强度 = 样品 $\times$ 照明；(d) 显微镜的像：每个频率成分乘以 OTF 后叠加，只剩下差频的莫尔条纹（周期 486 nm， $\text{OTF} = 0.53$ ）和很弱的照明条纹（ $\text{OTF} = 0.02$ ）。scripts/fig_sim.py。

推导：三个相位分离三个分量

同一像素（同一 $\mathbf{f}$ ）上，式 (8.3) 的三个未知量 $\tilde{C}(\mathbf{f})$ 、 $\tilde{C}(\mathbf{f} - \mathbf{f}_0)$ 、 $\tilde{C}(\mathbf{f} + \mathbf{f}_0)$ 混在一起。改变照明相位 ϕ 拍三张图， $\phi_k = 0, 2\pi/3, 4\pi/3$ ，得到三个方程

$$\tilde{I}_k = \tilde{B}_0 + \frac{\mu}{2} e^{i\phi_k} \tilde{B}_+ + \frac{\mu}{2} e^{-i\phi_k} \tilde{B}_-, \quad k = 1, 2, 3,$$

其中 $\tilde{I}_k$ 是第 k 张原始像的频谱， $\tilde{B}_0 = I_0 \text{OTF}(\mathbf{f}) \tilde{C}(\mathbf{f})$ 、 $\tilde{B}_\pm = I_0 \text{OTF}(\mathbf{f}) \tilde{C}(\mathbf{f} \mp \mathbf{f}_0)$ 是三个“频带”。系数矩阵的行列式不为零（三个相位均匀分布时，利用 $\sum_k e^{i\phi_k} = 0$ 可直接解出 $\tilde{B}_0 = \frac{1}{3} \sum_k \tilde{I}_k$ 、 $\tilde{B}_+ = \frac{2}{3\mu} \sum_k e^{-i\phi_k} \tilde{I}_k$ ，附录 F）。把分离出的 $\tilde{B}_\pm$ 在频率空间移回原位（ $\tilde{B}_+$ 中位于 $\mathbf{f}$ 处的是样品在 $\mathbf{f} - \mathbf{f}_0$ 处的成分），与 $\tilde{B}_0$ 拼在一起，就得到一个更大的频谱。为了在各个方向上都扩展，条纹要转三个方向（例如 0° 、 60° 、 120° ），共拍 9 张原始图（图 15(b)）。

要点：线性 SIM 的分辨率上限：2 倍

能恢复的样品频率满足 $|\mathbf{f}'| \leq f_c + f_0$ 。由式 (8.2)， f_0 最大为 $2\text{NA}/\lambda_{\text{exc}}$ ，于是

$$f_{\text{max}}^{\text{SIM}} = \frac{2\text{NA}}{\lambda_{\text{em}}} + \frac{2\text{NA}}{\lambda_{\text{exc}}} \approx 2f_c, \quad d_{\text{min}}^{\text{SIM}} \approx \frac{\lambda}{4\text{NA}}. \quad (8.4)$$

上限是 2 倍，原因有两条：照明条纹本身要通过同一个受衍射限制的物镜（ $f_0 \lesssim f_c$ ）；发光对照明强度是线性的，照明里只有 $0, \pm \mathbf{f}_0$ 三个频率，只能搬移 $\pm \mathbf{f}_0$ 一次。Gustafsson 2000 年用 NA 1.4 的物镜，把肌动蛋白纤维像的宽度从约 290 nm 减小到约 115 nm[17]。它的优点是

宽场成像、所需光强低（与普通荧光显微镜相当），适合活细胞。

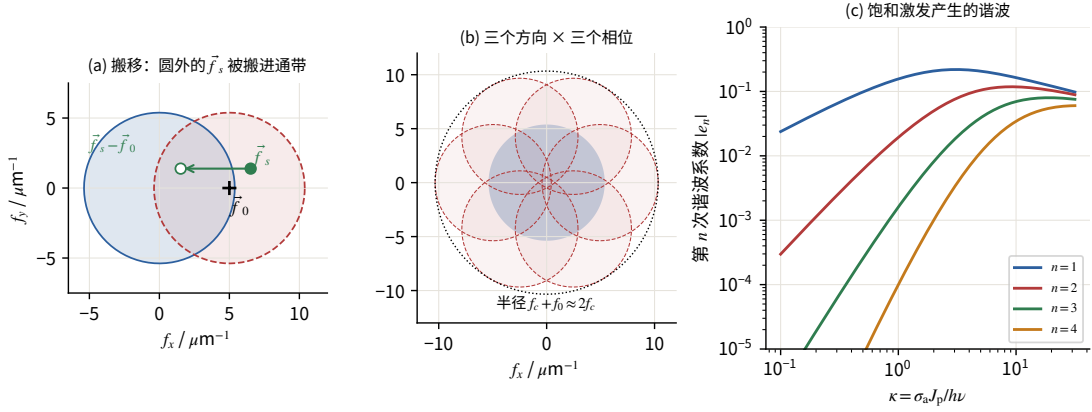


图 15 频率空间中的 SIM。(a) 蓝圆为通带 $|f| \leq f_c$ ；红虚线圆是以 f_0 为中心、半径 f_c 的圆，圆内的样品频率被搬移 $-f_0$ 后落入通带。图中 f_s （周期 150 nm）在通带外、在红圆内。(b) 三个方向的照明（ $f_0 = 0.92 f_c$ ）合成的扩展通带，半径约 $f_c + f_0$ 。(c) 饱和激发时（式 (8.5)）发光中第 n 次谐波的系数 $|e_n|$ 随 $\kappa = \sigma_a J_p / h\nu$ 的变化：解析式（修正贝塞尔函数，附录 F）与数值 FFT 一致。scripts/fig_sim.py。

交互演示：结构光照明（NA = 1.4）。调节样品条纹与照明条纹的周期、方向与相位，以及激发波长（照明条纹周期不能小于 $\lambda_{\text{exc}}/(2\text{NA})$ ，式 (8.2)），观察：普通显微像（均匀照明）在样品周期小于截止周期时是一片均匀；结构光照明下出现莫尔条纹；“SIM 重建”用三相位分离后的简单拼接恢复样品条纹（对比度按附录 F 的公式计算）。右下角为频率空间示意。

网页版中此处为可调参数的交互动画（演示编号 superres-sim）。

8.3 非线性 SIM：利用饱和产生高次谐波

要越过 2 倍，就要让照明的“有效频率”含有 $2f_0, 3f_0 \dots$ 。线性响应做不到，非线性响应可以：若发光正比于 $C(\mathbf{r})F[I_{\text{ill}}(\mathbf{r})]$ ， F 是非线性函数，则 $F[I_{\text{ill}}]$ 仍以 $1/f_0$ 为周期，但不再是单纯的余弦，它的傅里叶级数含有各次谐波 nf_0 。每一个谐波都把样品频谱搬移 nf_0 ，可恢复的频率扩大到 $f_c + n_{\text{max}}f_0$ ，原则上没有上限。Gustafsson 2005 年用饱和激发实现了这一点 [18]。

推导：脉冲激发的饱和

用宽度远小于荧光寿命的短脉冲激发，且激发前分子处于基态。脉冲期间自发衰减可以忽略，式 (5.3) 变为 $dp_1/dt = k_a(1 - p_1)$ ，从 $p_1 = 0$ 积分到脉冲结束：

$$p_1 = 1 - e^{-k_a t_{\text{pulse}}} = 1 - \exp\left(-\frac{\sigma_a J}{h\nu}\right), \quad (8.5)$$

J 为脉冲能流密度。 J 小时 $p_1 \approx \sigma_a J / h\nu$ （线性）， J 大时 $p_1 \rightarrow 1$ （饱和）。照明 $J(\mathbf{r}) = J_p[1 + \cos u]/2$ （ $u = 2\pi f_0 \cdot \mathbf{r}$ ， J_p 为峰值）。记 $\kappa = \sigma_a J_p / h\nu$ ，对小 κ 做泰勒展开到二阶：

$$p_1 \approx x - \frac{x^2}{2}, \quad x = \frac{\kappa}{2}(1 + \cos u).$$

用 $\cos^2 u = \frac{1}{2}(1 + \cos 2u)$ ， $x^2 = \frac{\kappa^2}{4}\left(\frac{3}{2} + 2\cos u + \frac{1}{2}\cos 2u\right)$ ，所以

$$p_1 \approx \left(\frac{\kappa}{2} - \frac{3\kappa^2}{16}\right) + \left(\frac{\kappa}{2} - \frac{\kappa^2}{4}\right)\cos u - \frac{\kappa^2}{16}\cos 2u.$$

出现了二次谐波 $\cos 2u$ ，振幅 $\kappa^2/16$ ，与一次谐波之比约 $\kappa/8$ 。精确的各次谐波振幅可用修正贝塞尔函数写出（附录 F），见图 15(c)。

注：非线性 SIM 的代价

n 次谐波的振幅在弱饱和时按 κ^n 增长，饱和不深时高次谐波很弱，信号很容易被噪声淹没；用到 $n_{\max}$ 次谐波时每个方向需要 $2n_{\max} + 1$ 个相位。强饱和需要强光，荧光分子容易漂白。Gustafsson 用 532 nm、640 ps 的脉冲激发，在荧光小球样品上利用到第 3 次谐波，得到优于 50 nm 的分辨率 [18]（他的饱和公式还包含了脉冲期间的自发衰减，式 (8.5) 是脉冲远短于寿命时的极限）。后来改用可光开关的荧光蛋白（类似第 6.4 节 RESOLFT 的思路），降低了所需光强。连续光饱和激发产生谐波的情形见例题 4。

8.4 三类方法的比较

表 2 从打破的前提、关键物理、分辨率、光强和速度几方面比较三类方法。

表 2 三类超分辨方法的比较（数值为典型量级）

	STED	单分子定位 (PALM/STORM)	线性 SIM
打破的前提	线性响应	对样品一无所知	均匀照明
关键物理	受激辐射饱和	光子统计 + 分子开关	频率搬移 (莫尔)
分辨率	几十 nm ($\propto I_{\max}^{-1/2}$)	约 20 nm ($\propto N^{-1/2}$)	约 100 nm (2 倍)
所需光强	很强 ($\gg I_s$)	中等 (需要开关与大量光子)	弱
成像方式与速度	点扫描，视野小时快	宽场，需上万帧	宽场，9 帧一幅

9 决赛风格例题

例题 1：硅藻、斜照明与“假细节”

用油浸物镜 (NA = 1.30，浸油 $n = 1.515$) 在 $\lambda = 550$ nm 的绿光下观察硅藻壳上的周期性纹路（看作振幅透过率为方波、透光宽度为半个周期的光栅，周期为 d ）。照明是相干的平面波。

1. 正入射照明时，能看到条纹的最小周期是多少？
2. 改用斜照明。若聚光镜与载玻片之间也加浸油，照明方向可达 $n \sin \theta_i = 1.30$ ，最小周期是多少？若聚光镜是“干”的（与载玻片之间是空气），最小周期变为多少？说明理由。
3. 取 $d = 600$ nm、正入射。列出哪些衍射级能进入物镜。若在物镜后焦面中心放一个小挡片，只挡住 0 级，像是什么样子？这对“用显微镜判断样品结构”有何警示？
4. 硅藻壳几乎透明，更接近一个弱相位光栅： $T(x) = \exp[i\phi_0 \cos(2\pi x/d)]$ ， $\phi_0 = 0.1$ ， $d = 600$ nm。证明在普通（明场）照明下像几乎没有衬度；若在后焦面上让 0 级多走 $1/4$ 波长（相位延迟 $\pi/2$ ），像的衬度变为多少？

解答

(1) 由式 (4.2), $d_{\min} = \lambda/\text{NA} = 550/1.30 = 423 \text{ nm}$ 。

(2) 0 级与 +1 级在 $n \sin \theta$ 轴上相距 λ/d 。0 级在 $n \sin \theta_i$ 处, +1 级在 $n \sin \theta_i + \lambda/d$ 处, 两者都要落在 $[-\text{NA}, \text{NA}]$ 内。取 $n \sin \theta_i = -\text{NA}_c$ ($\text{NA}_c \leq \text{NA}$), 条件为 $-\text{NA}_c + \lambda/d \leq \text{NA}$, 得 $d_{\min} = \lambda/(\text{NA} + \text{NA}_c)$ 。

浸油聚光镜: $\text{NA}_c = 1.30$, $d_{\min} = 550/2.60 = 212 \text{ nm}$ 。

干聚光镜: 照明光在到达样品前要穿过一层空气, 而光穿过平行界面时 $n \sin \theta$ 守恒 (第 3.1 节), 空气中 $n \sin \theta < 1$, 所以到达样品时仍有 $|n \sin \theta_i| < 1$, 即 $\text{NA}_c < 1$ 。 $d_{\min} > 550/(1.30 + 1) = 239 \text{ nm}$ (实际聚光镜 $\text{NA}_c \approx 0.95$ 时为 244 nm)。这就是高分辨显微观察时聚光镜也要上油的原因。

(3) $\lambda/d = 0.917$ 。各级 $n \sin \theta_m = 0.917m$: $m = 0, \pm 1$ 满足 $|n \sin \theta| \leq 1.30$, 进入物镜; $|m| \geq 2$ 时 $|n \sin \theta| \geq 1.83 > n = 1.515$, 是倏逝波, 根本不传播 (而且方波光栅的 $c_{\pm 2} = 0$)。

挡住 0 级后, 像的振幅只剩 ± 1 级: $U \propto c_1 e^{i2\pi x/d} + c_{-1} e^{-i2\pi x/d} = \frac{2}{\pi} \cos \frac{2\pi x}{d}$ ($c_{\pm 1} = 1/\pi$),

$$I \propto \cos^2 \frac{2\pi x}{d} = \frac{1}{2} \left(1 + \cos \frac{4\pi x}{d} \right),$$

周期为 $d/2 = 300 \text{ nm}$: 像中出现了样品上并不存在的、周期减半的条纹。阿贝的光栅实验正是用这种“在后焦面上挡住部分衍射级”的方法说明: 像的样子由哪些衍射级参与干涉决定, 像不一定忠实于物 [1, 35]。判断精细结构时, 必须知道哪些空间频率被系统传递了。

(4) 把 T 按 ϕ_0 展开到一阶: $T \approx 1 + i\phi_0 \cos(2\pi x/d) = 1 + \frac{i\phi_0}{2} (e^{i2\pi x/d} + e^{-i2\pi x/d})$, 即 $c_0 = 1$, $c_{\pm 1} = i\phi_0/2$ 。由 (3), 这三级都进入物镜。

明场: $I = |1 + i\phi_0 \cos(2\pi x/d)|^2 = 1 + \phi_0^2 \cos^2(2\pi x/d)$ 。一阶项消失, 因为 ± 1 级与 0 级相位差 $\pi/2$, 干涉项 $2\text{Re}(c_0^* c_1 \dots)$ 为零。衬度只有二阶小量 $\phi_0^2/2 = 0.005$ (而且周期为 $d/2$)。

0 级相位延迟 $\pi/2$ (乘以 i): $U = i + i\phi_0 \cos(2\pi x/d) = i[1 + \phi_0 \cos(2\pi x/d)]$,

$$I = 1 + 2\phi_0 \cos \frac{2\pi x}{d} + \mathcal{O}(\phi_0^2),$$

$$\text{衬度} \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \approx 2\phi_0 = 0.20.$$

相位起伏变成了强度起伏。这就是泽尼克 (F. Zernike) 的相衬显微镜原理 (1953 年诺贝尔物理学奖)。用精确系数 $c_m = i^m J_m(\phi_0)$ 的数值计算给出衬度 0.198, 明场衬度 5.0×10^{-3} (scripts/examples.py)。

例题 2: STED 需要多强的光

某染料荧光寿命 $\tau = 3.0 \text{ ns}$, 在 $\lambda_{\text{STED}} = 775 \text{ nm}$ 处受激辐射截面 $\sigma_{\text{se}} = 3.0 \times 10^{-17} \text{ cm}^2$ 。激发波长 $\lambda_{\text{exc}} = 640 \text{ nm}$, 物镜 $\text{NA} = 1.40$ 。耗尽光为矩形脉冲, 宽 $\tau_p = 0.30 \text{ ns}$, 重复频率 $f_{\text{rep}} = 40 \text{ MHz}$, 用涡旋相位板形成环形光斑。

1. 求脉冲耗尽的饱和光强 I_s 。

2. 第 5.3 节推导式 (5.5) 时略去了耗尽脉冲期间分子自发发出的荧光。若探测器不作时间门控、这部分荧光也被计入, 从速率方程 (5.3) 出发求剩余荧光比例 $\eta(I)$ (以无耗尽光时的总荧光为 1)。证明 $I \gg I_s$ 时 $\eta \approx (\tau_p/\tau)(I_s/I)$, 并说明这对环形光“关掉”荧光的效果

意味着什么。

3. 要求有效光斑半高全宽 $w = 30 \text{ nm}$ (按门控情形, 式 (6.3)), 求 $\zeta = I_{\max}/I_s$ 、环上峰值强度 $I_{\max}$ 和耗尽光的平均功率。(已知同功率下环形光斑满足 $\int I dA = 1.23 I_{\max}(\lambda_{\text{STED}}/\text{NA})^2$, 见附录 C。)
4. 若改用连续耗尽光 (激发光也连续且很弱), 达到同样的 w 需要多大的平均功率? 解释两者差别的物理原因。
5. 实际的环形光中心并非严格为零, 残留强度为 $\epsilon I_{\max}$ 。求光斑中心的荧光信号与无耗尽光时之比, 并计算 $\epsilon = 1\%$ 时 $w = 30 \text{ nm}$ 与 20 nm 两种情况下的值。若要求中心信号不低于无耗尽时的一半, 这台仪器能达到的最小 w 是多少?

解答

(1) $h\nu = hc/\lambda_{\text{STED}} = 2.56 \times 10^{-19} \text{ J}$ 。由式 (5.6),

$$I_s = \frac{h\nu}{\sigma_{\text{se}}\tau_p} = \frac{2.56 \times 10^{-19} \text{ J}}{(3.0 \times 10^{-21} \text{ m}^2)(3.0 \times 10^{-10} \text{ s})} = 2.85 \times 10^{11} \text{ W/m}^2 = 28 \text{ MW/cm}^2.$$

(2) 设激发后 $p_1(0) = 1$ 。脉冲期间 $dp_1/dt = -kp_1$, $k = k_{\text{se}} + 1/\tau$, $p_1(t) = e^{-kt}$ 。自发荧光的发射速率为 p_1/τ (量子产率是公共因子, 略去), 脉冲期间发出

$$\int_0^{\tau_p} \frac{p_1}{\tau} dt = \frac{1 - e^{-k\tau_p}}{k\tau},$$

脉冲之后剩余的 $p_1(\tau_p) = e^{-k\tau_p}$ 全部自发衰减。没有耗尽光时总荧光为 1 ($k = 1/\tau$ 时上两项之和为 1)。所以

$$\eta = e^{-k\tau_p} + \frac{1 - e^{-k\tau_p}}{k\tau},$$

$$k\tau_p = \frac{I}{I_s} + \frac{\tau_p}{\tau}, \quad k\tau = \frac{\tau}{\tau_p} \frac{I}{I_s} + 1.$$

$I \gg I_s$ 时第一项指数地趋于零, 第二项 $\approx 1/(k\tau) \approx (\tau_p/\tau)(I_s/I)$ 。本题 $\tau_p/\tau = 0.1$, 在环上 ($I = I_{\max} = 55I_s$) $\eta \approx 0.1/55 = 1.8 \times 10^{-3}$, 而门控时为 $e^{-55} \approx 10^{-24}$ 。物理上: 在耗尽脉冲的前一小段时间里, 分子还来不及被受激辐射“关掉”, 就有一定概率先自发发出一个光子, 这部分泄漏随光强只按 $1/I$ 下降, 不再是指数下降。对光斑宽度的影响不大 (数值计算: w 从门控时的 30.0 nm 变为 31.0 nm , scripts/examples.py), 但它给远离中心的区域留下了不随光强指数消失的“底”。实际仪器用时间门 (第 5.3 节) 去掉这部分荧光。

(3) 衍射极限光斑 $w_0 = 0.514\lambda_{\text{exc}}/\text{NA} = 235 \text{ nm}$ 。系数 $\beta = 1.62(640/775)^2 = 1.10$ 。由式 (6.3),

$$\zeta = \frac{(w_0/w)^2 - 1}{\beta} = \frac{(235/30)^2 - 1}{1.10} = 55,$$

$$I_{\max} = \zeta I_s = 1.6 \text{ GW/cm}^2.$$

(用精确 Airy 斑与精确环形光数值求解得 $\zeta = 55.1$, 差 0.5% 。) 环形光的总功率 (脉冲期间) $P_{\text{pk}} = 1.23 I_{\max}(\lambda_{\text{STED}}/\text{NA})^2 = 1.23 \times 1.56 \times 10^{13} \text{ W/m}^2 \times (0.554 \mu\text{m})^2 = 5.9 \text{ W}$ 。每个周期只

在 τ_p 内有光, 占空比 $\tau_p f_{\text{rep}} = 0.012$, 平均功率

$$\bar{P} = P_{\text{pk}} \tau_p f_{\text{rep}} \approx 71 \text{ mW}.$$

这与 Westphal 与 Hell 用约 800 MW/cm^2 得到 $30\text{--}35 \text{ nm}$ 光斑的量级一致 [30]。

(4) 连续耗尽时 $\eta = 1/(1 + I/I'_s)$, $I'_s = h\nu/(\sigma_{\text{se}}\tau) = 2.85 \text{ MW/cm}^2$ (式 (5.7))。宽度公式要重新推: 仍用高斯激发斑与抛物线环形光, 记 $u = 4 \ln 2 r^2/w_0^2$, 则 $\zeta I_{\text{STED}}/I_{\text{max}} = \beta\zeta u$,

$$H_{\text{eff}} = \frac{e^{-u}}{1 + \beta\zeta u}.$$

半高处 $u_{1/2} = \ln 2 (w/w_0)^2 = 0.0113$, 由 $e^{-u_{1/2}}/(1 + \beta\zeta u_{1/2}) = 1/2$ 解得 $\beta\zeta = (2e^{-u_{1/2}} - 1)/u_{1/2}$, $\zeta_{\text{cw}} = 78.5$ (精确数值 79.0)。所需功率 $P = 1.23 \zeta_{\text{cw}} I'_s (\lambda_{\text{STED}}/\text{NA})^2 \approx 0.84 \text{ W}$, 是脉冲平均功率的 12 倍。

为什么差这么多? 把比值拆开:

$$\frac{P_{\text{cw}}}{\bar{P}} = \frac{\zeta_{\text{cw}}}{\zeta} \cdot \frac{I'_s}{I_s \tau_p f_{\text{rep}}} = \frac{\zeta_{\text{cw}}}{\zeta} \cdot \frac{1}{f_{\text{rep}} \tau} = 1.43 \times 8.3.$$

第二个因子的含义: 脉冲方案中, 每个周期只需在激发之后的 τ_p 内送去能流密度约 $\zeta h\nu/\sigma_{\text{se}}$, 就能在分子自发光之前把它关掉; 连续方案中, 受激辐射必须在每一时刻与自发衰减速率 $1/\tau$ 竞争, 相当于每经过一个寿命 τ 就要送去同量级的能流。一个周期 $1/f_{\text{rep}}$ 内有 $1/(f_{\text{rep}}\tau) \approx 8$ 个寿命, 所以多耗约 8 倍。换一种说法: 连续光的能量平均分摊在所有时刻, 而一个分子在被激发之后的大约 τ 时间内, 耗尽光才对它起作用。第一个因子 1.43 来自 $1/(1+x)$ 比 e^{-x} “关”得慢 (图 7(b))。

(5) 中心处 $I_{\text{STED}} = \epsilon I_{\text{max}} = \epsilon \zeta I_s$, 由式 (5.5), 中心的荧光乘以 $e^{-\epsilon\zeta}$ 。 $w = 30 \text{ nm}$: $\zeta = 55$, 剩 $e^{-0.55} = 0.58$; $w = 20 \text{ nm}$: $\zeta = [(235/20)^2 - 1]/1.10 = 124$, 剩 $e^{-1.24} = 0.29$ 。要求 $e^{-\epsilon\zeta} \geq 1/2$, 即 $\zeta \leq \ln 2/\epsilon = 69$, 代入式 (6.3): $w_{\text{min}} = w_0/\sqrt{1 + \beta \cdot 69} = 27 \text{ nm}$ 。分辨率每提高一倍, ζ 约增大 4 倍, 残余强度造成的信号损失按 $e^{-\epsilon\zeta}$ 指数增长, 所以零点质量直接决定了可用的最高分辨率。

例题 3: 定位需要多少光子; MINFLUX

单分子像用高斯函数近似, $s = 100 \text{ nm}$; 相机像素折算到样品面 $a = 100 \text{ nm}$ 。

1. 用高斯最小二乘拟合定位 (式 (7.5) 适用), 要使定位精度 $\sigma_{\text{loc}} = 5 \text{ nm}$, 无背景时需要多少光子? 若每个像素平均有 10 个背景光子 (泊松涨落, $b^2 = 10$), 需要多少?
2. 要分辨周期 $d = 20 \text{ nm}$ 的结构, 按奈奎斯特判据, 标记的面密度至少多大? 对 $10 \mu\text{m} \times 10 \mu\text{m}$ 的视野, 若要求每个被激活分子半径 250 nm (约 $2.5s$) 内有另一个被激活分子的概率不超过 5%, 每帧最多激活多少个分子? 以每秒 100 帧采集, 至少要多长时间? (设每个分子只被定位一次。被激活分子在视野内随机均匀分布时, 给定区域内被激活分子数服从泊松分布。)
3. (MINFLUX 的一维模型) 激发光强度在零点附近为 $I(x) \propto (x - x_p)^2$ 。把零点先后放在 $x_p = -L/2$ 与 $+L/2$ (图 13), 两次曝光时间相同, 分别探测到 n_0, n_1 个光子, $n_0 + n_1 = N$ 。

分子位于 x_m 附近, 由比值 $\hat{\gamma} = n_0/N$ 反推 x_m 。求 $x_m = 0$ 时的定位精度 σ_x 。 $L = 50 \text{ nm}$ 时, 达到 1 nm 需要多少光子? 与 (1) 中相机定位 (无背景、 $a \rightarrow 0$) 相比如何?

4. 若每次曝光还有平均 n_b 个背景光子 (已知, 可扣除), 分子在 $x_m = 0$ 时每次曝光的信号光子平均为 n_s , 求定位精度, 用信号光子总数 $N = 2n_s$ 表示。为什么 L 不能无限缩小?

解答

(1) 无背景由式 (7.4): $N = (s^2 + a^2/12)/\sigma_{\text{loc}}^2 = (10^4 + 833)/25 = 433$ 。

有背景时由式 (7.5) ($b^2 = 10$): $25 = \frac{10833}{N} + \frac{8\pi \times 10^8 \times 10}{10^4 N^2} = \frac{10833}{N} + \frac{2.51 \times 10^6}{N^2}$ 。乘以 N^2 得二次方程 $25N^2 - 10833N - 2.51 \times 10^6 = 0$, 取正根 $N = 601$ 。背景使所需光子数增加约 40%。(若再按 Mortensen 等人对最小二乘的修正 [33], 第一项乘以约 16/9, 所需光子数还要再乘约 1.8。)

(2) 奈奎斯特: 相邻标记间距 $\leq d/2 = 10 \text{ nm}$, 面密度 $\rho \geq (2/20 \text{ nm})^2 = 10^4 \mu\text{m}^{-2}$; 视野中共 10^6 个分子。

设每帧被激活分子的面密度为 ρ_a 。一个被激活分子周围半径 $r = 0.25 \mu\text{m}$ 的圆内, 其他被激活分子数服从平均值 $\rho_a \pi r^2$ 的泊松分布, “至少有一个” 的概率为 $1 - e^{-\rho_a \pi r^2} \leq 0.05$, 得 $\rho_a \leq -\ln 0.95/(\pi r^2) = 0.26 \mu\text{m}^{-2}$, 每帧最多约 26 个。帧数 $\geq 10^6/26 = 3.8 \times 10^4$, 以 100 帧/秒计需约 380 s, 即 6 分多钟。

(3) 两次曝光的平均光子数正比于分子处的强度: $\langle n_0 \rangle \propto (x_m + L/2)^2$, $\langle n_1 \rangle \propto (x_m - L/2)^2$ 。给定总数 N , 每个光子独立地以概率

$$\gamma(x_m) = \frac{(x_m + L/2)^2}{(x_m + L/2)^2 + (x_m - L/2)^2}$$

来自第一次曝光, 所以 n_0 服从二项分布 (第 7.1 节), $\hat{\gamma} = n_0/N$ 的方差为 $\gamma(1-\gamma)/N$ 。 $x_m = 0$ 时 $\gamma = 1/2$, 方差 $1/(4N)$ 。

把 $\hat{\gamma}$ 的涨落换成位置的涨落要除以斜率 (泰勒展开到一阶: $\delta\hat{\gamma} \approx \gamma'(x_m) \delta x$)。记 $u = x_m + L/2$, $v = x_m - L/2$, $\gamma = u^2/(u^2 + v^2)$,

$$\frac{d\gamma}{dx_m} = \frac{2u(u^2 + v^2) - u^2(2u + 2v)}{(u^2 + v^2)^2} = \frac{2uv(v - u)}{(u^2 + v^2)^2}.$$

$x_m = 0$ 时 $u = L/2$, $v = -L/2$, $v - u = -L$, 斜率 $= \frac{2(L/2)(-L/2)(-L)}{(L^2/2)^2} = \frac{2}{L}$ 。于是

$$\sigma_x = \frac{\sqrt{1/(4N)}}{2/L} = \frac{L}{4\sqrt{N}}.$$

蒙特卡罗核对 ($N = 400$, 2 万次): $\sigma_x = 0.624 \text{ nm}$, 理论 0.625 nm 。这与 Balzarotti 等人给出的一维结果一致 [19]。 $L = 50 \text{ nm}$ 、 $\sigma_x = 1 \text{ nm}$ 需 $N = (L/4)^2 = 156$; 相机定位 $\sigma = s/\sqrt{N}$ 需 $N = (s/1 \text{ nm})^2 = 10^4$, 多 64 倍 $= (4s/L)^2$ 。MINFLUX 的 “有效 PSF 宽度” 是 $L/4$ 而不是 s 。

(4) 每次曝光的平均计数为 $n_s + n_b$, 扣除已知背景后用 $\hat{\gamma} = (n_0 - n_b)/(n_0 + n_1 - 2n_b)$ 。在 $x_m = 0$ 处, $\hat{\gamma}$ 对两次计数的偏导数分别为 $\pm 1/(4n_s)$, 两次计数的泊松方差都是 $n_s + n_b$, 所以

$$\text{Var } \hat{\gamma} \approx \frac{2(n_s + n_b)}{(4n_s)^2} = \frac{1}{4N} \left(1 + \frac{n_b}{n_s}\right), \quad N = 2n_s.$$

斜率不变 (背景已扣除), 故 $\sigma_x = \frac{L}{4\sqrt{N}} \sqrt{1 + n_b/n_s}$ 。蒙特卡罗 ($n_s = 50$ 、 $n_b = 10$): 1.39 nm, 理论 1.37 nm。

L 越小, 零点两侧的探测位置离分子越近, 分子处的激发强度 $\propto (L/2)^2$ 越小, 相同曝光时间内的信号 $n_s \propto L^2$ 下降, 而背景不变, $n_b/n_s \propto L^{-2}$ 。 L 小到 $n_b/n_s \gtrsim 1$ 时, $\sigma_x \propto L \cdot L^{-1}$, 精度不再随 L 改善 (若靠延长曝光补足信号, 背景也随之增大, 比值不变)。实际 MINFLUX 中零点的残余强度起着与背景相同的作用。

例题 4: 饱和结构光、相位误差与倏逝波照明

本题讨论结构光照明在第 8 节之外的三个问题。

- 连续光饱和激发。用连续照明代替短脉冲, 分子处于激发态的稳态概率为 $p_1 = x/(1+x)$ (式 (5.4), $x = I/I_{\text{sat}}$)。照明 $x = x_0(1 + \cos u)$, $u = 2\pi \mathbf{f}_0 \cdot \mathbf{r}$ 。
 - 利用等比级数证明: 当 $|q| < 1$ 时 $1 + 2 \sum_{n \geq 1} q^n \cos nu = \frac{1 - q^2}{1 - 2q \cos u + q^2}$ 。
 - 由此求 p_1 的各次谐波系数 ($p_1 = c_0 + \sum_{n \geq 1} c_n \cos nu$), 给出 q 与 x_0 的关系。检验 $x_0 \rightarrow 0$ 的线性极限。
 - 计算 $x_0 = 1$ 与 $x_0 = 10$ 时 $|c_2/c_1|$ 、 $|c_3/c_1|$ 。
 - 强饱和时, 发光图样在照明零点 $u = \pi$ 处出现一条窄暗线。求暗线的宽度 (取 p_1 降到 $1/2$ 处的全宽, 用照明周期 $\Lambda = 1/f_0$ 表示), 说明它与 STED 有什么相同之处。
- 相位步进误差。线性 SIM 中第二张原始像的实际相位为 $2\pi/3 + \delta$, 但重建时仍按 $0, 2\pi/3, 4\pi/3$ 计算 (附录 F 的分离公式)。求分离出的各分量与真值的关系, 保留到 δ 的一阶; $\delta = 10^\circ$ 时, “错误分量”混入的比例是多少? 说明它在重建图像中造成什么样的伪影。
- 倏逝波照明能否突破 $f_0 \leq 2NA/\lambda_{\text{exc}}$? 样品是水溶液 ($n_w = 1.33$), 物镜 $NA = 1.49$ (油浸, 玻璃 $n_g = 1.518$), 激发 488 nm, 荧光 520 nm。若两束照明光在玻璃中的入射角超过全反射临界角, 样品中只剩沿界面传播、垂直界面指数衰减的倏逝波, 两束倏逝波干涉成条纹。求: 两束光在水中传播时条纹的最高空间频率; 在全反射条件下的最高空间频率; 相应的 SIM 最短可恢复周期 (检测截止频率按 $2NA/\lambda_{\text{em}}$ 计); 倏逝波光强的穿透深度。全反射照明能否突破式 (8.4) 的 2 倍上限?

解答

(1a) 令 $\xi = qe^{iu}$, $|\xi| < 1$ 。等比级数 $\sum_{n \geq 0} \xi^n = 1/(1 - \xi)$, 所以

$$\begin{aligned} 1 + 2 \sum_{n \geq 1} q^n \cos nu &= 2 \operatorname{Re} \frac{1}{1 - \xi} - 1 = \frac{2(1 - q \cos u)}{1 - 2q \cos u + q^2} - 1 \\ &= \frac{1 - q^2}{1 - 2q \cos u + q^2}. \end{aligned}$$

(第二步: $\frac{1}{1 - \xi} = \frac{1 - \xi^*}{|1 - \xi|^2}$, $|1 - \xi|^2 = 1 - 2q \cos u + q^2$ 。)

(1b) $p_1 = 1 - \frac{1}{1+x} = 1 - \frac{1}{\mathcal{A} + \mathcal{B} \cos u}$, $\mathcal{A} = 1 + x_0$, $\mathcal{B} = x_0$ 。要把 $\frac{1}{\mathcal{A} + \mathcal{B} \cos u}$ 写成 (1a) 的形式, 需 $1 - 2q \cos u + q^2 \propto \mathcal{A} + \mathcal{B} \cos u$, 即 $(1 + q^2)/(-2q) = \mathcal{A}/\mathcal{B}$ 。取 $|q| < 1$ 的根:

$$q = \frac{\sqrt{\mathcal{A}^2 - \mathcal{B}^2} - \mathcal{A}}{\mathcal{B}} = \frac{\sqrt{1 + 2x_0} - 1 - x_0}{x_0} < 0.$$

比较常数 (在 u 的任一值处) 得 $\frac{1}{\mathcal{A} + \mathcal{B} \cos u} = \frac{1}{\sqrt{\mathcal{A}^2 - \mathcal{B}^2}} \cdot \frac{1 - q^2}{1 - 2q \cos u + q^2}$, 所以

$$c_0 = 1 - \frac{1}{\sqrt{1 + 2x_0}}, \quad c_n = -\frac{2q^n}{\sqrt{1 + 2x_0}} \quad (n \geq 1).$$

各次谐波按 $|q|^n$ 递减。线性极限: $x_0 \rightarrow 0$ 时 $\sqrt{1 + 2x_0} \approx 1 + x_0 - x_0^2/2$, $q \approx -x_0/2$, $c_1 \approx x_0$, $c_2 \approx -x_0^2/2$, 即 $p_1 \approx x_0(1 + \cos u)$, 与线性响应 $p_1 \approx x$ 一致。(数值 FFT 核对, 偏差 $< 10^{-10}$ 。)

(1c) $|c_{n+1}/c_n| = |q|$ 。 $x_0 = 1$: $q = \sqrt{3} - 2 = -0.268$, $|c_2/c_1| = 0.268$, $|c_3/c_1| = 0.072$ 。 $x_0 = 10$: $q = -0.642$, $|c_2/c_1| = 0.642$, $|c_3/c_1| = 0.412$ 。饱和越深, 高次谐波越强, 可利用的谐波越多。

(1d) 在 $u = \pi$ 附近 $1 + \cos u \approx (u - \pi)^2/2$, $x \approx x_0(u - \pi)^2/2$ 。 p_1 在远离暗线处接近 1, 在 $x = 1$ 处降到一半, 即 $|u - \pi| = \sqrt{2/x_0}$ 。暗线半高全宽为 $\Delta u = 2\sqrt{2/x_0}$, 换成长度是 $\Lambda \Delta u / (2\pi) = \sqrt{2}\Lambda / (\pi\sqrt{x_0})$ ($x_0 = 10$ 时为 0.142Λ , 精确数值 0.144Λ ; 若改按“半深度”即 p_1 降到其最大值一半处计, 精确值为 0.137Λ)。与 STED 相同之处: 都是利用照明强度的零点附近的抛物线形状加上饱和响应, “暗区”的宽度随饱和程度按 (强度/饱和强度) $^{-1/2}$ 收缩, 不受衍射限制。区别只是 STED 中零点附近的分子发光, SIM 中零点附近的分子不发光。

(2) 第 k 张原始像 $\tilde{I}_k = \tilde{B}_0 + \frac{\mu}{2}e^{i\phi_k}\tilde{B}_+ + \frac{\mu}{2}e^{-i\phi_k}\tilde{B}_-$, 实际相位 ϕ_k 只在 $k = 2$ 时比名义相位 ϕ_k^0 多 δ , 其余相等。按名义相位分离:

$\tilde{B}_0$ 的估计 $\frac{1}{3} \sum \tilde{I}_k$ 中, $\tilde{B}_0$ 的系数为 1; $\tilde{B}_\pm$ 的系数为 $\frac{\mu}{6} \sum_k e^{\pm i\phi_k} = \frac{\mu}{6} e^{\pm i\phi_2^0} (e^{\pm i\delta} - 1) \approx \pm \frac{i\mu\delta}{6} e^{\pm i\phi_2^0}$, 即混入模约为 $\mu\delta/6$ 的调制分量 ($\delta = 10^\circ$ 、 $\mu = 1$ 时 2.9%)。

$\tilde{B}_+$ 的估计 $\frac{2}{3\mu} \sum_k e^{-i\phi_k^0} \tilde{I}_k$ 中: $\tilde{B}_0$ 的系数为 $\frac{2}{3\mu} \sum e^{-i\phi_k^0} = 0$; $\tilde{B}_+$ 的系数为 $\frac{1}{3}(2 + e^{i\delta}) \approx 1 + i\delta/3$; $\tilde{B}_-$ 的系数为 $\frac{1}{3} \sum_k e^{-i\phi_k^0} e^{-i\phi_k} = \frac{1}{3} e^{-2i\phi_2^0} (e^{-i\delta} - 1)$, 模为 $\frac{2}{3} \sin(\delta/2) \approx \delta/3$ 。

$\delta = 10^\circ = 0.175 \text{ rad}$ 时, “错误分量” $\tilde{B}_-$ 以约 $\delta/3 = 5.8\%$ 混入 $\tilde{B}_+$ (数值 5.81%)。 $\tilde{B}_-$ 携带的是 $\tilde{C}(\mathbf{f} + \mathbf{f}_0)$, 却被当作 $\tilde{C}(\mathbf{f} - \mathbf{f}_0)$ 移回原位, 样品频谱于是被错移了 $2\mathbf{f}_0$ 。样品中最强的成分为零频 (平均亮度), 它被移到 $\pm 2\mathbf{f}_0$ 处, 在重建图像中表现为周期为 $\Lambda/2$ 、沿照明条纹方向的“残余条纹”伪影。另一方面, $\tilde{B}_0$ 中混入的 $\tilde{B}_\pm$ 在 $\pm \mathbf{f}_0$ 处带着样品的零频成分 $I_0 \text{OTF}(\mathbf{f}_0) \tilde{C}(0)$, 不经搬移就留在像中, 表现为周期 Λ 的残余照明条纹, 系数 $\mu\delta/6$, 与前一种同量级 (约为其一半)。两种伪影的幅度都正比于 $\text{OTF}(\mathbf{f}_0)$: 在不除以 OTF 的简单拼接中很弱 ($\mathbf{f}_0 = 0.92f_c$ 时 $\text{OTF} = 0.027$), 但实际重建要把各频带除以 OTF (去卷积或加权) 以压平频谱, 伪影随之被放大。所以实际 SIM 要从原始数据中精确估计每张图的实际相位。

(3) 两束平面波以 $\pm\theta$ 入射时条纹频率 $f_0 = 2n \sin \theta / \lambda_{\text{exc}}$, 而 $n \sin \theta$ 在各平行层中守恒 (第 3.1 节)。

- 在水中传播: $n \sin \theta = n_w \sin \theta_w < 1.33$, $f_0 < 2 \times 1.33 / 0.488 \mu\text{m} = 5.45 \mu\text{m}^{-1}$ (周期 183 nm)。
- 全反射: 光在玻璃中以 $n_g \sin \theta_g$ 入射, 它可以大于 n_w 但不超过物镜的 $\text{NA} = 1.49$ 。界面上横向空间频率不变, 水中的场为 $e^{ik_x x} e^{-\Gamma z}$, $k_x = 2\pi n_g \sin \theta_g / \lambda_{\text{exc}}$, 两束干涉的条纹频率最高为 $2\text{NA} / \lambda_{\text{exc}} = 6.11 \mu\text{m}^{-1}$ (周期 164 nm)。

- $f_c = 2 \times 1.49 / 0.520 \mu\text{m} = 5.73 \mu\text{m}^{-1}$ 。SIM 最短周期：传播光 $1/(5.73 + 5.45) \mu\text{m} = 89 \text{ nm}$ ，全反射 $1/(5.73 + 6.11) \mu\text{m} = 84 \text{ nm}$ 。注意：水中分子发出的、 $n \sin \theta > n_w$ 的荧光（“超临界角荧光”）只来自离界面约一个波长以内的分子，所以按 $2\text{NA}/\lambda_{\text{em}}$ 估算探测截止只对全反射照明的薄层合理；对传播光照明的厚样品，探测截止更接近 $2n_w/\lambda_{\text{em}}$ ，上面的 89 nm 偏乐观。
- 穿透深度：由 $k_x^2 - \Gamma^2 = (2\pi n_w/\lambda)^2$ ， $\Gamma = (2\pi/\lambda)\sqrt{(n_g \sin \theta_g)^2 - n_w^2}$ ；光强按 $e^{-2\Gamma z}$ 衰减， $1/e$ 深度为 $\lambda/[4\pi\sqrt{1.49^2 - 1.33^2}] = 58 \text{ nm}$ 。

结论：全反射照明能使条纹频率达到 $2\text{NA}/\lambda_{\text{exc}}$ ，在样品折射率低于物镜 NA 时比传播光更细，但不能超过这一值：两束光都必须经过物镜，横向空间频率受 NA 限制，所以式 (8.4) 的 2 倍上限仍然成立。代价是只能照亮界面附近约 60 nm 的薄层。

10 实验与数据处理：标定单分子定位精度

例题：实验题：定位精度与光子数、像素的关系

背景。单分子定位显微镜在使用前要标定：定位精度是否像理论预言的那样随光子数变化？点扩散函数的实际宽度 s 是多少？本题数据为模拟数据：由 `scripts/exp_data.py` 按下述机制逐个光子生成，随机数种子 20261017（事先固定）。

装置与数据。把直径约 20 nm 的荧光小球稀疏地固定在盖玻片上，用油浸物镜（ $\text{NA} = 1.45$ ）在荧光波长 $\lambda = 600 \text{ nm}$ 下成像。背景已扣除，可视为零。对 8 个小球各连续拍 $M = 200$ 帧，改变激发光强度使每个小球的每帧平均光子数 $\bar{N}$ 不同。每一帧的光子落点同时按两种像素记录：相机原始像素（折算到样品面 $a_1 = 100 \text{ nm}$ ），以及在软件中把 2×2 个像素合并后的大像素（ $a_2 = 200 \text{ nm}$ ，合并后的格点与原格点对齐）。各小球相对像素格点的位置是随机的。两种记录各自用“光子记录位置的平均”（质心）定位一次。对每个小球给出 $\bar{N}$ 和两种像素下 200 次定位结果在 x 、 y 方向的样本标准差，见表 3。

问题。

1. 从第 7.3、7.4 节的模型出发，写出 $\sigma^2 \equiv (\sigma_x^2 + \sigma_y^2)/2$ 与 $\bar{N}$ 的关系，说明为什么可以合并 x 、 y 两个方向。已知 M 次测量的样本方差的相对标准差为 $\sqrt{2/(M-1)}$ ，估计每个 σ^2 的相对不确定度。
2. 对两种像素分别作 $\ln \sigma - \ln \bar{N}$ 直线拟合，求斜率及其不确定度，检验 $\sigma \propto \bar{N}^{-1/2}$ 。
3. 固定斜率 $-1/2$ ，对每种像素求 $Q \equiv \sigma^2 \bar{N}$ 的最佳值与不确定度，并检验各点是否与常数相符。
4. 由 Q 求 s 。两种像素来自同一批光子，利用这一点设计一个不依赖 s 、尽量灵敏的检验，验证像素项 $a^2/12$ 。结果说明什么？
5. 先检验两种像素给出的 s 是否一致，再给出 s 的最佳值，与理想 Airy 斑的高斯近似 $0.21\lambda/\text{NA}$ 比较，讨论差异的可能原因。

6. 讨论系统误差：(a) 若每个像素有平均 0.05 个背景光子，在以小球为中心的 7×7 个原始像素内求质心，背景对 σ^2 有多大贡献？还有什么别的影响？(b) 光子数逐帧涨落、样品漂移、改用电子倍增 CCD 相机分别会怎样影响结果？如何从数据中检验漂移？

表 3 荧光小球反复定位的模拟数据 ($M = 200$ 帧; σ_x, σ_y 单位 nm)

$\bar{N}$	$\sigma_x(a_1)$	$\sigma_y(a_1)$	$\sigma_x(a_2)$	$\sigma_y(a_2)$
121	9.39	9.02	10.66	9.33
201	7.27	6.96	8.17	7.57
348	6.06	5.41	6.67	5.88
598	4.17	4.14	4.73	4.25
998	3.77	3.24	4.11	3.69
1703	2.35	2.58	2.56	2.89
3001	2.09	1.85	2.26	1.96
4999	1.36	1.53	1.50	1.67

解答

(1) 模型。由式 (7.4)，每帧定位误差在每个方向上的方差为 $(s^2 + a^2/12)/N$ 。光子数 N 逐帧涨落（泊松，均值 $\bar{N}$ ）， $\langle 1/N \rangle = (1/\bar{N})[1 + \mathcal{O}(1/\bar{N})]$ ， $\bar{N} \geq 120$ 时修正 $< 1\%$ ，可略。所以

$$\sigma^2 = \frac{Q}{\bar{N}}, \quad Q = s^2 + \frac{a^2}{12}. \quad (10.1)$$

x, y 两个方向的光子落点相互独立、模型相同 (PSF 圆对称、像素是正方形)， σ_x^2, σ_y^2 是同一个量的两次独立估计。两者平均后自由度为 $2(M-1)$ ， σ^2 的相对不确定度为 $\sqrt{2/[2(M-1)]} = 1/\sqrt{199} = 7.1\%$ 。由 $\ln \sigma = \frac{1}{2} \ln \sigma^2$ ， $\delta(\ln \sigma) = \frac{1}{2} \delta(\ln \sigma^2) = \frac{1}{2} \times 7.1\% = 0.035$ 。

(2) 对数拟合。令 $X_i = \ln \bar{N}_i$ 、 $Y_i = \ln \sigma_i$ 。各点 Y_i 的不确定度相同，用不加权最小二乘（附录 G）。 a_1 ：斜率 -0.492 ± 0.011 ； a_2 ： -0.493 ± 0.012 （不确定度由残差估计；由 (1) 的 0.035 传递得 0.010，两者相符）。两者都与 $-1/2$ 相差不到 1σ ：数据支持 $\sigma \propto \bar{N}^{-1/2}$ （图 16(a)）。

(3) 求 Q 。每点算 $Q_i = \sigma_i^2 \bar{N}_i$ （单位 nm^2 ）：

a_1 ：10257, 10180, 11483, 10324, 12331, 10370, 11690, 10474；

a_2 ：12141, 12467, 13757, 12090, 15224, 12692, 13428, 12595。

各点的相对不确定度都是 7.1%，等权平均即为加权平均：

$$Q_1 = 10889 \pm 291, \quad Q_2 = 13049 \pm 373 \quad (\text{nm}^2),$$

不确定度取 Q_i 的样本标准差除以 $\sqrt{8}$ （按 7.1% 传递分别为 273、327，相符）。检验常数假设： $\chi^2 = \sum_i [(Q_i - \bar{Q})/(0.071\bar{Q})]^2$ ，两组分别为 7.95、9.10，自由度 7。 χ^2 大于这两个值的概率（附录 G）分别为 34% 与 25%，可以接受。 Q_i 与 $\bar{N}$ 也没有系统趋势（图 16(b)）。

(4) 求 s ；像素项检验。 $\delta s = \delta Q/(2s)$ ：

$$s_1 = \sqrt{Q_1 - a_1^2/12} = \sqrt{10889 - 833} = 100.3 \pm 1.4 \text{ nm},$$

$$s_2 = \sqrt{13049 - 3333} = 98.6 \pm 1.9 \text{ nm}.$$

不依赖 s 的检验： $Q_2 - Q_1 = (a_2^2 - a_1^2)/12 = 2500 \text{ nm}^2$ 。两种像素记录的是同一批光子，光子落点的高斯涨落对两者相同，在逐个小球的差 $\Delta Q_i = Q_{2,i} - Q_{1,i}$ 中大部分相消，所以应当配对相减，而不是把 Q_1 、 Q_2 当作独立量相减：

$$\begin{aligned}\Delta Q_i &= 1885, 2287, 2274, 1766, \\ &2893, 2322, 1738, 2121; \\ \overline{\Delta Q} &= 2161 \pm 134 \text{ nm}^2.\end{aligned}$$

(若当作独立量相减，不确定度为 $\sqrt{291^2 + 373^2} = 473$ 。配对后误差的期望约减小一半：同一机制重复模拟 400 次，配对与不配对误差估计的中位数分别为 189 与 399，比值约 2.1；本组数据中比值达 3.5，部分是因为配对误差 134 偶然偏小，只有约 13% 的模拟会这么小。) 结论有两层：

- 像素项确实存在： $\overline{\Delta Q}$ 与零相差 16 倍标准差。
- 它的大小比预言的 2500 低 2.5σ 。只有 8 个小球时，误差本身也是估计出来的，应按自由度 7 的 t 分布 (附录 G) 估计偶然出现的概率，双侧约 4%。这是一个轻微的张力：单凭这组数据不足以否定模型，也不能说“完全相符”。若按重复模拟得到的期望误差 (约 200) 计算，偏离约 1.7σ 。这正说明为什么要用 t 分布：8 个点估出的误差本身就有约 25% 的不确定度。只有当它在重复实验中持续出现时，才需要怀疑像素模型。(对同一生成机制重复模拟 400 次， $\overline{\Delta Q}$ 的平均值为 2494 ± 10 ，说明模型本身无偏，见 VERIFY.md。)

各 ΔQ_i 之间的散布 (约 ± 380) 部分来自亚像素位置效应： $a_2 = 200 \text{ nm} \approx 2s$ 时，单个光子的方差随小球相对像素格点的位置有约 $\pm 3\%$ 的起伏 (第 7.4 节)。这一效应对随机位置平均为零，只增加散布、不造成系统偏差，而这份散布已包含在由样本标准差估计的误差 134 中。

(5) 一致性与比较。两种像素给出的 s 相差 1.7 nm，按独立误差计算为 0.7σ ；由于来自同一批光子，两者高度相关，逐个小球配对比较给出 $s_2 - s_1 = -1.8 \pm 0.7 \text{ nm}$ ，与 (4) 中 $\overline{\Delta Q}$ 略低于预言是同一件事 (2.6σ)。这个差 (1.8 nm) 远小于 s 与理想值之差 (约 12 nm)，合并不影响下面的结论：对每个小球取两种像素的 s 的平均，再对 8 个小球求平均，并把半差 0.9 nm 作为系统误差，得

$$s = 99.3 \pm 1.6 (\text{统计}) \pm 0.9 (\text{系统}) \text{ nm} \approx 99.3 \pm 1.8 \text{ nm}.$$

理想值 $0.21\lambda/\text{NA} = 0.21 \times 600/1.45 = 87 \text{ nm}$ 。测得 s 大 14%，偏离约 7σ ，差异显著。可能的原因：离焦 (小球不在焦面上，PSF 变宽)；油的折射率与样品介质不匹配造成的球差；高 NA 下的矢量衍射效应；荧光光谱有一定宽度 (长波成分的 PSF 更宽)。小球本身的大小 (半径 10 nm，投影方差约 $R^2/5 = 20 \text{ nm}^2$) 只使 s^2 增加 0.2%，可以忽略。注意：本题是模拟数据，生成时把 s 设为 98 nm，用来代表这些因素的总效果；单凭这组数据无法区分是哪一种原因，要区分需要改变离焦量、换浸油等对照实验。

(6) 系统误差。

- (a) 背景：区域内背景光子平均数 $n_b = 0.05 \times 49 = 2.45$ ，质心为 $\hat{x} = (\sum_{\text{信号}} x + \sum_{\text{背景}} x_p)/(N + n_b)$ ， x_p 从小球位置 (区域中心) 量起。背景有两个作用：分子上多出方

差 $b^2 \sum_p x_p^2$ ；分母变大，使信号涨落按 $N/(N + n_b)$ 被压缩（第 7.4 节）。合起来

$$\sigma^2 \approx \frac{NQ + b^2 \sum_p x_p^2}{(N + n_b)^2}.$$

7×7 个 100 nm 像素， $\sum_p x_p^2 = 7 \times 2 \times (100^2 + 200^2 + 300^2) \text{ nm}^2 = 1.96 \times 10^6 \text{ nm}^2$ ， $b^2 = 0.05$ 。 $\bar{N} = 120$ 时 $\sigma^2 = 90.1 \text{ nm}^2$ ，比无背景时的 87.0 nm^2 大 3.6%（若只计分子上的背景项会误得 7.8%，因为忽略了分母对信号涨落的压缩）； $\bar{N} = 1000$ 时只大 0.4%。光子级蒙特卡罗给出 +3.7% 与 +0.5%（统计误差约 1%，scripts/exp_data.py）。所以背景会使小 $\bar{N}$ 端的 Q_i 上翘约 3–4%，约为单点统计误差（7%）的一半。此外背景把质心拉向区域中心， $\bar{N} = 120$ 时约 2%，是系统偏差。实际做法是缩小求质心的区域，或用包含背景项的模型拟合。本题已假设背景扣除为零。

- (b) 光子数涨落：使 σ^2 偏大约 $1/\bar{N}$ ，在 $\bar{N} = 120$ 时为 0.8%，远小于 7% 的统计误差。
- 漂移：样品在 200 帧内的机械漂移给每个小球的定位加上一个与 $\bar{N}$ 无关的方差 σ_d^2 ，模型变为 $\sigma^2 = Q/\bar{N} + \sigma_d^2$ ，在大 $\bar{N}$ 端表现为 Q_i 上翘。检验方法：以 $1/\bar{N}$ 为自变量，对 σ^2 作带截距的加权直线拟合（附录 G）。两种像素的截距分别为 0.03 ± 0.17 、 $-0.01 \pm 0.20 \text{ nm}^2$ ，都与零相符，漂移可以忽略。
- 电子倍增 CCD：倍增过程的随机性使信号的涨落增大，有效方差约加倍 [33]（过量噪声因子 $F^2 \approx 2$ ），即 $Q \rightarrow 2Q$ 。若不知道这一点而直接用式 (10.1)，会得到 $s = \sqrt{2Q_1 - 833} \approx 145 \text{ nm}$ ，严重偏大。

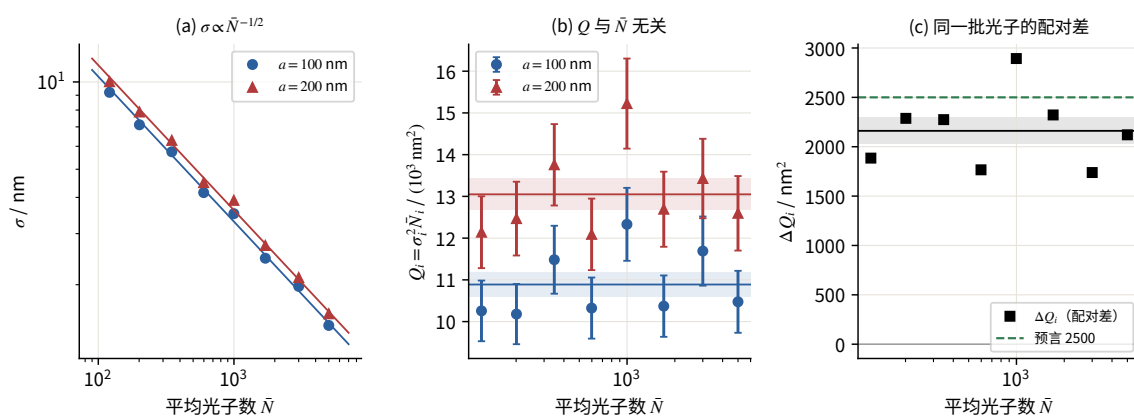


图 16 实验题数据处理。(a) 双对数坐标下 σ 与 $\bar{N}$ ，直线为 $\sigma = \sqrt{Q/\bar{N}}$ ；(b) $Q_i = \sigma_i^2 \bar{N}_i$ ，误差棒为 7.1% 的统计不确定度，横线与色带为平均值及其标准误；(c) 逐个小球的配对差 ΔQ_i ，黑线与灰带为平均值 2161 ± 134 ，绿虚线为预言值 2500。scripts/exp_data.py。

11 小结：推导链

1. 波与空间频率：平面波 $\Leftrightarrow$ 横向空间频率 $n \sin \theta / \lambda$ ；惠更斯-菲涅耳原理 + 远场近似 $\Rightarrow$ 衍射振幅 = 孔径函数的傅里叶变换（式 (3.5)） $\Rightarrow$ 单缝、光栅方程 (3.10)、Airy 斑 (3.11)。
2. 阿贝：物 = 光栅的叠加；成像需要至少两个衍射级进入物镜 $\Rightarrow d_{\min} = \lambda / \text{NA}$ （正入射）、

- $\lambda/(2NA)$ (斜照明)。傅里叶光学：相干成像 = 乘以光瞳函数 (截止 NA/λ)；荧光是非相干成像，强度卷积 $\Rightarrow$ OTF = 光瞳自相关，截止 $f_c = 2NA/\lambda$ (式 (4.8))；PSF 宽度 $0.51\lambda/NA$ ，瑞利距离 $0.61\lambda/NA$ 。
3. 衍射极限：线性成像中 $|f| > f_c$ 的信息为零，与光强和算法无关；出路是打破“线性”“无先验”“均匀照明”之一。
 4. 荧光动力学：速率方程 (5.3) $\Rightarrow$ 激发饱和 $I_{\text{sat}} = h\nu/(\sigma_a\tau)$ ；受激辐射耗尽 $\eta = e^{-I/I_s}$ (脉冲, $I_s = h\nu/\sigma_{\text{se}}\tau_p$) 或 $1/(1 + I/I_s)$ (连续, $I_s = h\nu/\sigma_{\text{se}}\tau$)。
 5. STED：涡旋相位 $\Rightarrow$ 中心零点、 $I \approx I_{\text{max}}(r/r_0)^2$ ；有效 PSF = 激发 PSF $\times \eta \Rightarrow w = w_0/\sqrt{1 + \beta\zeta}$ (式 (6.3))；功率、零点质量与漂白决定实际极限。
 6. 单分子定位：独立光子的平均值 $\Rightarrow \sigma_{\text{loc}}^2 = (s^2 + a^2/12)/N$ (+ 背景项)；重建图像还受奈奎斯特标记密度限制；MINIFLUX 用零点把有效宽度从 s 换成 $L/4$ 。
 7. SIM：积化和差 / 调制定理 $\Rightarrow$ 频谱搬移 $\pm f_0$ ；三相位 $\times$ 三方向分离与拼接 $\Rightarrow$ 线性上限 $f_c + f_0 \approx 2f_c$ ；饱和产生谐波 $\Rightarrow$ 非线性 SIM 原则上无上限。

参考文献

- [1] Ernst Abbe. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–468, 1873. doi: 10.1007/BF02956173.
- [2] Ernst Abbe. On the estimation of aperture in the microscope. *Journal of the Royal Microscopical Society*, 1(3):388–423, 1881. doi: 10.1111/j.1365-2818.1881.tb05909.x.
- [3] George Biddell Airy. On the diffraction of an object-glass with circular aperture. *Transactions of the Cambridge Philosophical Society*, 5:283–291, 1835. URL <https://archive.org/details/transactionsofca05camb>.
- [4] Lord Rayleigh. Investigations in optics, with special reference to the spectroscope. *Philosophical Magazine (Series 5)*, 8(49):261–274, 1879. doi: 10.1080/14786447908639684.
- [5] The Royal Swedish Academy of Sciences. The Nobel prize in chemistry 2014 – press release (8 october 2014). NobelPrize.org, 2014. URL <https://www.nobelprize.org/prizes/chemistry/2014/press-release/>.
- [6] Stefan W. Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782, 1994. doi: 10.1364/OL.19.000780.
- [7] Thomas A. Klar and Stefan W. Hell. Subdiffraction resolution in far-field fluorescence microscopy. *Optics Letters*, 24(14):954–956, 1999. doi: 10.1364/OL.24.000954.

- [8] Thomas A. Klar, Stefan Jakobs, Marcus Dyba, Alexander Egner, and Stefan W. Hell. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences USA*, 97(15):8206–8210, 2000. doi: 10.1073/pnas.97.15.8206.
- [9] Eva Rittweger, Kyu Young Han, Scott E. Irvine, Christian Eggeling, and Stefan W. Hell. STED microscopy reveals crystal colour centres with nanometric resolution. *Nature Photonics*, 3(3):144–147, 2009. doi: 10.1038/nphoton.2009.2.
- [10] W. E. Moerner and L. Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters*, 62(21):2535–2538, 1989. doi: 10.1103/PhysRevLett.62.2535.
- [11] M. Orrit and J. Bernard. Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Physical Review Letters*, 65(21):2716–2719, 1990. doi: 10.1103/PhysRevLett.65.2716.
- [12] Robert M. Dickson, Andrew B. Cubitt, Roger Y. Tsien, and W. E. Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355–358, 1997. doi: 10.1038/41048.
- [13] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006. doi: 10.1126/science.1127344.
- [14] Samuel T. Hess, Thanu P. K. Girirajan, and Michael D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11):4258–4272, 2006. doi: 10.1529/biophysj.106.091116.
- [15] Michael J. Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–795, 2006. doi: 10.1038/nmeth929.
- [16] Rainer Heintzmann and Christoph G. Cremer. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. In *Optical Biopsies and Microscopic Techniques III*, volume 3568 of *Proc. SPIE*, pages 185–196, 1999. doi: 10.1117/12.336833.
- [17] M. G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy*, 198(2):82–87, 2000. doi: 10.1046/j.1365-2818.2000.00710.x.
- [18] Mats G. L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences USA*, 102(37):13081–13086, 2005. doi: 10.1073/pnas.0406877102.

- [19] Francisco Balzarotti, Yvan Eilers, Klaus C. Gwosch, Arvid H. Gynnå, Volker Westphal, Fernando D. Stefani, Johan Elf, and Stefan W. Hell. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science*, 355(6325):606–612, 2017. doi: 10.1126/science.aak9913.
- [20] Klaus C. Gwosch, Jasmin K. Pape, Francisco Balzarotti, Philipp Hoess, Jan Ellenberg, Jonas Ries, and Stefan W. Hell. MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nature Methods*, 17(2):217–224, 2020. doi: 10.1038/s41592-019-0688-0.
- [21] B. Richards and E. Wolf. Electromagnetic diffraction in optical systems, II. structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London A*, 253:358–379, 1959. doi: 10.1098/rspa.1959.0200.
- [22] Joseph W. Goodman. *Introduction to Fourier Optics*. Roberts & Company, Englewood, Colorado, 3rd edition, 2005.
- [23] Max Born and Emil Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, 7th edition, 1999. doi: 10.1017/CBO9781139644181.
- [24] C. M. Sparrow. On spectroscopic resolving power. *The Astrophysical Journal*, 44:76–86, 1916. doi: 10.1086/142271.
- [25] ISS Inc. Lifetime data of selected fluorophores. URL https://www.iss.com/resources/reference/data_tables/FL_LifetimeStandards.html. accessed 2026-10.
- [26] Scott Prahl. Rhodamine 6G (photochemcad data; extinction coefficient from birge 1987, quantum yield from kubin & fletcher 1982). Oregon Medical Laser Center, 2017. URL <https://omlc.org/spectra/PhotochemCAD/html/083.html>.
- [27] Michael Hofmann, Christian Eggeling, Stefan Jakobs, and Stefan W. Hell. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proceedings of the National Academy of Sciences USA*, 102(49):17565–17569, 2005. doi: 10.1073/pnas.0506010102.
- [28] Peter Török and Peter R. T. Munro. The use of Gauss-Laguerre vector beams in STED microscopy. *Optics Express*, 12(15):3605, 2004. doi: 10.1364/OPEX.12.003605.
- [29] Måns Ehrenberg. Super-resolved fluorescence microscopy. scientific background on the Nobel prize in chemistry 2014, October 2014. URL <https://www.nobelprize.org/uploads/2018/06/advanced-chemistryprize2014.pdf>.
- [30] Volker Westphal and Stefan W. Hell. Nanoscale resolution in the focal plane of an optical microscope. *Physical Review Letters*, 94(14):143903, 2005. doi: 10.1103/PhysRevLett.94.143903.

- [31] Raimund J. Ober, Sripad Ram, and E. Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical Journal*, 86(2):1185–1200, 2004. doi: 10.1016/S0006-3495(04)74193-4.
- [32] Russell E. Thompson, Daniel R. Larson, and Watt W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, 82(5):2775–2783, 2002. doi: 10.1016/S0006-3495(02)75618-X.
- [33] Kim I. Mortensen, L. Stirling Churchman, James A. Spudich, and Henrik Flyvbjerg. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nature Methods*, 7(5):377–381, 2010. doi: 10.1038/nmeth.1447.
- [34] Hari Shroff, Catherine G. Galbraith, James A. Galbraith, and Eric Betzig. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature Methods*, 5(5):417–423, 2008. doi: 10.1038/nmeth.1202.
- [35] Albert B. Porter. On the diffraction theory of microscopic vision. *Philosophical Magazine (Series 6)*, 11(61):154–166, 1906. doi: 10.1080/14786440609463433.

A 圆孔衍射与 Airy 斑

贝塞尔函数 本讲用到两个函数，用积分定义：

$$\begin{aligned} J_0(x) &= \frac{1}{2\pi} \int_0^{2\pi} e^{ix \cos \varphi} d\varphi, \\ J_1(x) &= \frac{1}{2\pi i} \int_0^{2\pi} e^{i\varphi} e^{ix \cos \varphi} d\varphi. \end{aligned} \quad (\text{A.1})$$

一般地，

$$J_n(x) = \frac{1}{2\pi i^n} \int_0^{2\pi} e^{in\varphi} e^{ix \cos \varphi} d\varphi.$$

把 $e^{ix \cos \varphi}$ 展开成幂级数并逐项积分，利用 $\frac{1}{2\pi} \int_0^{2\pi} \cos^{2k} \varphi d\varphi = \frac{(2k)!}{2^{2k}(k!)^2}$ （奇次幂积分为零），得

$$\begin{aligned} J_0(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(k!)^2} \left(\frac{x}{2}\right)^{2k}, \\ J_1(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+1)!} \left(\frac{x}{2}\right)^{2k+1} \end{aligned}$$

（ J_1 的级数同理，用 $\frac{1}{2\pi} \int \cos \varphi \cos^{2k+1} \varphi d\varphi$ ）。逐项求导可验证

$$\frac{d}{dx} [xJ_1(x)] = xJ_0(x). \quad (\text{A.2})$$

小 x 时 $J_0 \approx 1 - x^2/4$, $J_1 \approx x/2$ ；大 x 时它们像振幅按 $x^{-1/2}$ 衰减的余弦。 J_1 的前三个正零点为 3.8317、7.0156、10.1735（scripts/verify_numeric.py 用 scipy 核对，并与式 (A.1) 的数值积分比较）。

圆孔的远场 直径 D 的圆孔, 用极坐标 (ρ', φ') 表示孔内一点, 观察方向的空间频率 $\mathbf{f}$ 的大小为 $f = \sin \theta / \lambda$ (空气中), 方位角为 ψ 。式 (3.5) 的积分为

$$\begin{aligned}\tilde{U}(f) &= \int_0^{D/2} \rho' d\rho' \int_0^{2\pi} e^{-i2\pi f \rho' \cos(\varphi' - \psi)} d\varphi' \\ &= 2\pi \int_0^{D/2} J_0(2\pi f \rho') \rho' d\rho' .\end{aligned}$$

(内层积分把 $\varphi' - \psi$ 换成积分变量, 并注意 J_0 是偶函数。) 令 $x = 2\pi f \rho'$, 用式 (A.2):

$$\begin{aligned}\tilde{U}(f) &= \frac{2\pi}{(2\pi f)^2} \int_0^{\pi f D} x J_0(x) dx = \frac{2\pi}{(2\pi f)^2} \pi f D J_1(\pi f D) \\ &= \frac{\pi D^2}{4} \cdot \frac{2J_1(v)}{v}, \quad v = \pi f D = \frac{\pi D \sin \theta}{\lambda} .\end{aligned}$$

$\pi D^2/4$ 是孔的面积, 即 $\theta = 0$ 时的振幅。取模平方即得式 (3.11)。

显微镜的点扩散函数 式 (4.4) 中的 h 是半径 NA/λ 的圆盘 (光瞳) 的逆变换, 计算完全相同, 只是把 “孔半径 $D/2$ ” 换成 “频率半径 NA/λ ”、把 “频率 f ” 换成 “位置 r ”: $h(r) \propto 2J_1(v)/v$, $v = 2\pi \text{NA} r / \lambda$, 即式 (4.5)。

几个数值 强度 $[2J_1(v)/v]^2$ 降到一半处 $v = 1.6163$, 第一零点 $v = 3.8317$, 与艾里 1835 年的表 (1.616 与 3.832) 一致 [3]; 第一暗环以内的能量占 83.8% (scripts/verify_numeric.py 数值积分核对)。

B 非相干光学传递函数

由第 4.4 节, $\tilde{H}(\mathbf{f})$ 等于两个半径 $f_p = \text{NA}/\lambda$ (光瞳的频率半径)、圆心相距 $f = |\mathbf{f}|$ 的圆盘的重叠面积。重叠区被两圆的公共弦分成两个相同的弓形。设 $q = f/(2f_p)$, 弦到每个圆心的距离为 $f/2 = qf_p$, 弦所对圆心角为 2χ , $\cos \chi = q$ 。一个弓形的面积 = 扇形面积 - 三角形面积 = $f_p^2 \chi - f_p^2 \sin \chi \cos \chi$ 。重叠面积

$$S_o(f) = 2f_p^2 (\chi - \sin \chi \cos \chi) = 2f_p^2 (\arccos q - q\sqrt{1 - q^2}) .$$

除以 $f = 0$ 时的面积 πf_p^2 , 并注意 $q = f/(2f_p) = f/f_c$, 即得式 (4.9)。检验: $f = 0$ 时为 1; $f = f_c$ 时为 0; $f = f_c/2$ 时为 0.391。scripts/fig_transfer.py 用二维离散自相关逐点核对 (偏差 < 0.01)。

C 涡旋光瞳的焦斑

光瞳用归一化半径 $\rho = |\mathbf{f}|/(\text{NA}/\lambda) \in [0, 1]$ 与方位角 φ 表示, 焦面上一点用 $v = 2\pi \text{NA} r / \lambda$ 与方位角 ψ 表示, 则 $2\pi \mathbf{f} \cdot \mathbf{r} = v\rho \cos(\varphi - \psi)$ 。加涡旋相位后

$$\begin{aligned}h_v &\propto \int_0^1 \rho d\rho \int_0^{2\pi} e^{i\varphi} e^{iv\rho \cos(\varphi - \psi)} d\varphi \\ &= 2\pi i e^{i\psi} \int_0^1 J_1(v\rho) \rho d\rho \equiv 2\pi i e^{i\psi} g_v(v) .\end{aligned}$$

第二步令 $\varphi' = \varphi - \psi$ 并用式 (A.1) 中 J_1 的定义。强度 $\propto g_v(v)^2$, 与方位角无关 (圆环)。

- 中心附近: $J_1(x) \approx x/2$, $g_v(v) \approx \frac{v}{2} \int_0^1 \rho^2 d\rho = v/6$, 强度 $\propto v^2/36$, 与第 6.2 节的展开一致。
- 峰值: 数值求 g_v^2 的极大, 在 $v = 2.4516$ ($r = 0.390\lambda/\text{NA}$), $g_{\max}^2 = 0.06469$ 。于是 $I/I_{\max} \approx (v/6)^2/0.06469 = 0.4294v^2$, 即式 (6.1): $r_0 = \lambda/(2\pi\sqrt{0.4294}\text{NA}) = 0.243\lambda/\text{NA}$ 。
- 与普通聚焦光斑比较: 无涡旋时中心振幅为 $2\pi \int_0^1 J_0(0)\rho d\rho = \pi$, 此处峰值振幅为 $2\pi g_{\max}$, 强度比 $4g_{\max}^2 = 0.259$ 。
- 总功率: 相位板不改变光瞳上的光强, 焦面上的总功率不变。对普通光斑积分式 (4.10) (利用 $\int_0^\infty J_1^2(v)dv/v = 1/2$),

$$\int H dA = 4\pi \left(\frac{\lambda}{2\pi\text{NA}} \right)^2 H(0).$$

同一功率的环形光峰值 $I_{\max} = 0.259H(0)$, 所以

$$\int I dA = \frac{4\pi}{0.259} \left(\frac{\lambda}{2\pi\text{NA}} \right)^2 I_{\max} = 1.23 I_{\max} \left(\frac{\lambda}{\text{NA}} \right)^2,$$

即例题 2 所用的关系。

- $g_v(v)$ 也可写成闭式 $g_v(v) = \frac{\pi}{2v} [J_1(v)\mathbf{H}_0(v) - J_0(v)\mathbf{H}_1(v)]$ ($\mathbf{H}_n$ 为 Struve 函数), scripts/optics.py 用它作图, 并与二维 FFT 及直接数值积分核对 (最大偏差 < 0.01)。

D 概率与统计工具

样本方差的涨落 对 M 个独立的高斯变量 (期望 x_0 、方差 σ^2), 样本方差 $\hat{\sigma}^2 = \frac{1}{M-1} \sum (x_j - \bar{x})^2$ 的期望为 σ^2 。它的涨落可以这样估计: M 很大时 $\bar{x} \approx x_0$, $\hat{\sigma}^2 \approx \frac{1}{M} \sum y_j$, $y_j = (x_j - x_0)^2$ 。 y_j 相互独立, $\langle y_j \rangle = \sigma^2$, $\langle y_j^2 \rangle = \langle (x - x_0)^4 \rangle = 3\sigma^4$ (高斯分布的四阶矩, 由 $\int t^4 e^{-t^2/2} dt = 3\sqrt{2\pi}$ 得到), 所以 $\text{Var } y_j = 2\sigma^4$, $\text{Var } \hat{\sigma}^2 \approx 2\sigma^4/M$ 。把用 $\bar{x}$ 代替 x_0 所损失的一个自由度计入, $M \rightarrow M-1$, 相对标准差为 $\sqrt{2/(M-1)}$ (scripts/verify_numeric.py 用蒙特卡罗核对)。由 $\delta\hat{\sigma}/\hat{\sigma} = \frac{1}{2}\delta\hat{\sigma}^2/\hat{\sigma}^2$, 样本标准差的相对标准差约为 $1/\sqrt{2(M-1)}$ 。

定位精度的理论下限 (选读) 统计学中的“克拉默–拉奥下限”说: 若每个光子的落点概率密度为 $\varrho(x-x_0)$, 则由 N 个独立光子对 x_0 作任何无偏估计, 方差不小于 $1/(N\mathcal{F})$, 其中 $\mathcal{F} = \int (\partial_x \varrho)^2 / \varrho dA$ 称为 (单光子的) 费希尔信息。对高斯 PSF, $\mathcal{F} = 1/s^2$, 下限就是 s^2/N , 质心法已经达到。对 Airy 斑, $\varrho \propto h^2$ (h 为实的振幅 PSF), $(\partial_x \varrho)^2 / \varrho = 4(\partial_x h)^2$, 由帕塞瓦尔定理 (函数平方的积分等于其傅里叶变换模平方的积分) 与 “求导对应乘以 $i2\pi f_x$ ”,

$$\begin{aligned} \mathcal{F} &= \frac{4 \int (\partial_x h)^2 dA}{\int h^2 dA} = \frac{4 \int (2\pi f_x)^2 |P|^2 d^2 f}{\int |P|^2 d^2 f} \\ &= 16\pi^2 \langle f_x^2 \rangle_{\text{光瞳}} = 4\pi^2 \left(\frac{\text{NA}}{\lambda} \right)^2, \end{aligned}$$

最后一步用了均匀圆盘上 $\langle f_x^2 \rangle = f_p^2/4$, $f_p = \text{NA}/\lambda$ 。于是 $\sigma_{\min} = \lambda/(2\pi\text{NA}\sqrt{N})$, 与 Ober 等人的结果一致 [31], 比高斯近似的 $0.21\lambda/(\text{NA}\sqrt{N})$ 小约 25%。scripts/verify_numeric.py 用数值积分核对了 $\mathcal{F} = 4\pi^2(\text{NA}/\lambda)^2$ 。

E 采样定理与混叠

在间隔为 Δ 的点 $x_j = j\Delta$ 上对余弦 $\cos(2\pi f x)$ 采样。由于 $2\pi(1/\Delta)x_j = 2\pi j$,

$$\cos(2\pi f x_j) = \cos(2\pi j - 2\pi f x_j) = \cos\left[2\pi\left(\frac{1}{\Delta} - f\right)x_j\right],$$

频率 f 与 $1/\Delta - f$ 在所有样点上取值完全相同, 无法区分 (混叠)。若已知函数只含 $0 \leq f \leq f_{\max}$ 的成分, 为使任何两个允许的频率都不混叠, 需 $1/\Delta - f_{\max} \geq f_{\max}$, 即

$$\Delta \leq \frac{1}{2f_{\max}} \quad (\text{奈奎斯特判据}).$$

可以进一步证明, 满足这一条件时用适当的插值公式能由样点唯一复原原函数 (香农采样定理), 此处不展开。单分子定位中“样点”是随机分布的标记分子, 判据只作数量级估计用 [34]。

F SIM 的三相位分离与饱和谐波

分离 记 $\omega = e^{i2\pi/3}$, $\phi_k = 2\pi(k-1)/3$ 。三个方程 $\tilde{I}_k = \tilde{B}_0 + \frac{\mu}{2}\omega^{k-1}\tilde{B}_+ + \frac{\mu}{2}\omega^{-(k-1)}\tilde{B}_-$ 。利用 $1 + \omega + \omega^2 = 0$:

$$\begin{aligned} \frac{1}{3} \sum_k \tilde{I}_k &= \tilde{B}_0, \\ \frac{2}{3\mu} \sum_k \omega^{-(k-1)} \tilde{I}_k &= \tilde{B}_+, \quad \frac{2}{3\mu} \sum_k \omega^{k-1} \tilde{I}_k = \tilde{B}_-. \end{aligned}$$

(例如第二式中 $\tilde{B}_0$ 的系数为 $\frac{2}{3\mu} \sum \omega^{-(k-1)} = 0$, $\tilde{B}_+$ 的系数为 $\frac{2}{3\mu} \cdot \frac{\mu}{2} \cdot 3 = 1$, $\tilde{B}_-$ 的系数为 $\frac{2}{3\mu} \cdot \frac{\mu}{2} \sum \omega^{-2(k-1)} = 0$ 。) 系数矩阵行列式的模为 $3\sqrt{3}\mu^2/4$ 。

演示中的简单重建 在空间域, 相位为 ϕ 的原始像为 $I_\phi = H \otimes \{C[1 + \cos(2\pi \mathbf{f}_0 \cdot \mathbf{r} + \phi)]\}$ (取 $\mu = 1$, 略去常数 I_0)。按上式分离出三幅“频带像” $B_0 = H \otimes C$ 、 $B_\pm = H \otimes (C e^{\pm i2\pi \mathbf{f}_0 \cdot \mathbf{r}})$, 再乘以 $e^{\mp i2\pi \mathbf{f}_0 \cdot \mathbf{r}}$ 把频谱移回原位, 相加得重建像 $I_{\text{rec}} = B_0 + B_+ e^{-i2\pi \mathbf{f}_0 \cdot \mathbf{r}} + B_- e^{i2\pi \mathbf{f}_0 \cdot \mathbf{r}}$ 。对样品 $C = 1 + \cos(2\pi \mathbf{f}_s \cdot \mathbf{r})$ 逐项计算 (每个频率分量乘以该频率处的 OTF),

$$\begin{aligned} I_{\text{rec}} &= 1 + 2 \text{OTF}(f_0) + [\text{OTF}(|\mathbf{f}_s|) \\ &\quad + \text{OTF}(|\mathbf{f}_s - \mathbf{f}_0|) + \text{OTF}(|\mathbf{f}_s + \mathbf{f}_0|)] \\ &\quad \times \cos(2\pi \mathbf{f}_s \cdot \mathbf{r}). \end{aligned}$$

只要三个 OTF 中有一个不为零, 样品条纹就被恢复。演示 superres-sim 显示的是这个结果 (未作任何去卷积与加权, 实用的算法会对各频带除以 OTF 并加权以压平频谱)。

饱和谐波的精确系数 修正贝塞尔函数可以由生成函数定义 (与式 (A.1) 类比):

$$e^{z \cos u} = \sum_{n=-\infty}^{\infty} I_n(z) e^{inu}, \quad I_n(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{z \cos u} \cos nu \, du.$$

由式 (8.5), 把 $z = -\kappa/2$ 代入并用 $I_n(-z) = (-1)^n I_n(z)$:

$$p_1 = 1 - e^{-\kappa(1+\cos u)/2} = 1 - e^{-\kappa/2} \sum_n (-1)^n I_n(\kappa/2) e^{inu}.$$

所以 e^{inu} ($n \neq 0$) 的系数为 $e_n = -(-1)^n e^{-\kappa/2} I_n(\kappa/2)$, $\cos nu$ 的系数为 $2e_n$ 。小 κ 时 $I_n(z) \approx (z/2)^n/n!$, 与第 8.3 节的二阶展开一致。图 15(c) 画出 $|e_n|$, 并用数值 FFT 核对 (偏差 $< 10^{-10}$)。

G 最小二乘拟合公式

直线 $y = y_0 + Kx$, **各点等权** 记 $\bar{x} = \frac{1}{n} \sum x_i$, $\bar{y}$ 同理, $S_{xx} = \sum (x_i - \bar{x})^2$ 。使 $\sum (y_i - y_0 - Kx_i)^2$ 最小 (对 y_0 、 K 求偏导令其为零), 得

$$K = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{S_{xx}}, \quad y_0 = \bar{y} - K\bar{x}.$$

若各点 y_i 的标准差同为 σ_y , K 是 y_i 的线性组合, 由方差相加 $\sigma_K = \sigma_y/\sqrt{S_{xx}}$ 。 σ_y 未知时用残差估计 $\sigma_y^2 \approx \sum (y_i - y_0 - Kx_i)^2/(n-2)$ (减 2 是因为两个参数已由数据定出)。

加权 各点标准差 σ_i 不同时, 最小化 $\chi^2 = \sum [(y_i - y_0 - Kx_i)/\sigma_i]^2$ 。记 $w_i = 1/\sigma_i^2$, $W = \sum w_i$, $W_x = \sum w_i x_i$, $W_y = \sum w_i y_i$, $W_{xx} = \sum w_i x_i^2$, $W_{xy} = \sum w_i x_i y_i$, $\Delta = WW_{xx} - W_x^2$:

$$K = \frac{WW_{xy} - W_x W_y}{\Delta}, \quad y_0 = \frac{W_{xx} W_y - W_x W_{xy}}{\Delta},$$

$$\sigma_K = \sqrt{\frac{W}{\Delta}}, \quad \sigma_{y_0} = \sqrt{\frac{W_{xx}}{\Delta}}.$$

只拟合一个常数 $y = \bar{Q}$ 时, $\bar{Q} = \sum w_i y_i / \sum w_i$, $\sigma_{\bar{Q}} = (\sum w_i)^{-1/2}$; 各点相对误差相同、数值又相近时, 退化为普通平均。

χ^2 检验与 p 值 若模型正确、各 σ_i 估计正确, 每一项 $(y_i - \text{模型})/\sigma_i$ 是一个期望为 0、方差为 1 的高斯变量, 它们的平方和 χ^2 的期望等于自由度 ν (数据点数减去拟合参数个数), 标准差约为 $\sqrt{2\nu}$ 。 χ^2 明显大于 ν (例如超过 $\nu + 2\sqrt{2\nu}$) 说明模型或误差估计有问题。更精确地, “模型正确时 χ^2 偶然大于测得值的概率” 称为 p 值, 可由 χ^2 分布表查得 (或用 `scipy.stats.chi2.sf` 计算), p 只有百分之几或更小时才认为不相符。比较一个平均值与预言值时, 若误差本身是由少数几个点的散布估计的 (n 个点), 偏离倍数 $t = (\text{测量} - \text{预言})/\text{标准误}$ 服从自由度 $n-1$ 的 t 分布, 它比高斯分布 “尾巴” 更长; $n=8$ 时 $|t| > 2.5$ 的概率约 4%, 而高斯分布为 1.2%。