

ISP仿射重量化对光子转移特性及感知质量的影响与抑制策略

by 知乎/Bilibili@姜尧耕

个人站: <https://y-g-jiang.github.io>

ISP仿射重量化对光子转移特性及感知质量的影响与抑制策略

P 代序

1 引言

2 标准解码框架与黑电平机制

2-1 标准解码框架

2-1-1 零点良好空间中的色彩空间变换

2-1-2 白平衡解码

2-1-3 白平衡解码向对角的简化

2-2 黑电平的意义及确定机制

3 极低噪下的量化噪声模型修正

3-1 平场中的噪声分析

3-1-1 理论噪声模型

3-1-2 综合噪声模型与实验验证

3-1-2-1 综合噪声模型

3-1-2-2 实验验证

3-2 高斯模型下的量化噪声

3-3 极低噪声下基座偏移的模型修正

3-4 高斯对均匀假设补偿干扰的PTC模拟

4 量化网格漂移致偏色及改进的解码办法

4-1 漂移量的感知结果

4-2 改进的解码办法

5 Prescaling仿射重量化模型分析

5-1 Prescaling的发现与建模

5-2 Prescaling对PTC光子转移曲线的影响

5-2-1 黑电平精确的情况下PTC的刻画

5-2-1-1 实验模拟

5-2-1-2 数学模型

5-2-1-3 对DR测量的影响

5-2-2 黑电平漂移对非整数倍重量化的PTC的影响

5-2-2-1 数学模型模拟

5-2-2-2 对DR测量的影响

5b z8的特殊情况

6 附录

7 参考文献

P 代序

本文是我在25.7~25.11的工作中一小部分的总结。另一部分工作大概在JiangtherapeeOnline那篇文档中 (<https://y-g-jiang.github.io/JOindex.html>) 。并不都深刻，但足以启发。有个课程作业需要我提交一个与概率统计有关的论文，于是整理如下。

这篇中我也会给出一些我这两周的新东西。快速过一遍我下面会提及的东西：

1, 黑电平以及prescaling的机制探索（上月内容）；2, 黑电平矫正来改善暗部偏色办法的提出（上月内容）；3, 仿射量化导致PTC弯曲以及相位移动的机理分析（本周的新内容）；4, 极低噪声下量化网格精确对提升动态范围的意义（本周的新内容）；5, 量化网格漂移量的感知结果（本周的新内容）。

1 引言

在现代ISP链路中，非线性量化调制例如对数曲线编码与 Gamma 校正因其对信号统计特性的显著改变而广受关注。工业界通常假设，只要排除了这些非线性模块，获取的数据即为“Linear Raw Data”。然而，这一假设往往与消费电子的事实相悖。大量被认为“底层”或“Raw”的传感器输出数据，实际上在 ISP 前端经历了粗糙的仿射重量化处理。这种操作虽然维持了宏观的线性度，但其引入的量化网格相位失配以及PTC弯折、DR测量失准等问题却往往被忽略。

这是因为在早期的消费级数字成像系统中，受限于较高的模拟读出噪声和暗电流，微观的量化调制伪影往往被掩盖。但多增益合成等新技术的引入带来的极低底噪和宽动态范围使得这些被掩盖的结构缺陷在测试以及实际使用中愈发显著。

本文旨在填补消费级相机在极高信噪比时代的这一研究空白。通过建立数学模型并精确实验，我量化了消费级传感器的诸多过程对传感器光子转移曲线、色彩解码保真度以及最终感知质量的级联影响，并提出了意义重大的相应抑制策略。

2 标准解码框架与黑电平机制

2-1 标准解码框架

2-1-1 零点良好空间中的色彩空间变换

人对色彩解码的感知可以被拆解为如下三个步骤：1，物体光谱反射率与照明光源相乘，得到进入人眼睛的有效光谱；2，人眼用三个视锥的响应函数对光谱做内积。

$$\mathbf{H} = \begin{bmatrix} L \\ M \\ S \end{bmatrix} = \int E(\lambda) \begin{bmatrix} l(\lambda) \\ m(\lambda) \\ s(\lambda) \end{bmatrix} d\lambda$$

数码相机的色彩响应函数显然是与人类视觉系统响应函数往往不同的。

$$\mathbf{C} = \begin{bmatrix} R_{cam} \\ G_{cam} \\ B_{cam} \end{bmatrix} = \int E(\lambda) \begin{bmatrix} r(\lambda) \\ g(\lambda) \\ b(\lambda) \end{bmatrix} d\lambda$$

为了使得相机空间的色彩被转换到人的色彩空间中，我们希望使用 3×3 的一个线性变换矩阵把相机的三个基向人的三个基进行粗略变换。设矩阵 \mathbf{T} 为从人眼 lms 到相机rgb响应曲线的变换阵，对于一个CIS的CMF近似人眼CMF的线性组合的情况，我们可以推出这种情况下的解码办法：

$$\begin{aligned} \begin{bmatrix} r(\lambda) \\ g(\lambda) \\ b(\lambda) \end{bmatrix} &= \mathbf{T} \cdot \begin{bmatrix} l(\lambda) \\ m(\lambda) \\ s(\lambda) \end{bmatrix} \\ \mathbf{C} &= \mathbf{T} \cdot \left(\int E(\lambda) \begin{bmatrix} l(\lambda) \\ m(\lambda) \\ s(\lambda) \end{bmatrix} d\lambda \right) = \mathbf{T} \cdot \mathbf{H} \\ \mathbf{H} &= \mathbf{T}^{-1} \cdot \mathbf{C} = \mathbf{T}^{-1} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{scene} \end{aligned}$$

其中 \mathbf{T} 被Adobe命名为ColorMatrix。为保持符号简洁，下文中的“ \mathbf{T} ”均指ColorMatrix的逆。

受限于尺寸对膜层数的限制以及成本、量子效率的取舍，一般CIS的CMF并不是人眼CMF的精确线性组合。因此CCM 仅能提供针对特定训练样本集的最小均方误差解，而非全光谱域的解析解。一般商业流程中需要更为细致的LUT操作。

2-1-2 白平衡解码

人类视觉系统会自然地使用一种色度适应机制来调整其对场景光源白点的感知，从而在变化的照明条件下实现色彩恒常性。

以sRGB举例，他的白点被定义为CIE标准光源D65的色度坐标。也即当一个sRGB图像中的某个像素的R、G、B值相等且达到最大值时(例如[255, 255, 255])，它在理想的sRGB显示设备上应该呈现出与标准D65日光(色温约为6500K)下的一块完美白色反光板相同的颜色外观。因此为了匹配人眼的色彩恒常性，RAW在经过T变换后，我希望把现场白迁到D65。这一步变换我们称其为CAT。

在数学上，最基础的色度适应模型是Von-Kries假说。该假说认为，色彩适应可以通过对人眼三种视锥细胞响应进行独立的对角增益调节来实现。然而，直接在CIE XYZ空间或标准的LMS空间中执行这种对角缩放，往往无法产生视觉上最准确的适应结果。

为了解决这一问题，我们引入了Bradford变换。他可以通过寻找一个锐化的响应空间，使得在该空间内执行Von-Kries对角缩放能产生最低的感知误差。

首先，利用Bradford矩阵 \mathbf{M}_{BFD} 将CIE XYZ值映射到锐化的RGB/LMS中间域：

$$\begin{bmatrix} \rho \\ \gamma \\ \beta \end{bmatrix} = \mathbf{M}_{BFD} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

在该中间域内，利用源白点 (X_S, Y_S, Z_S) 和目标白点 (X_D, Y_D, Z_D) 计算出的比例系数，构建对角矩阵 \mathbf{D} 进行白点对齐。

$$\mathbf{D} = \begin{bmatrix} \rho_D/\rho_S & 0 & 0 \\ 0 & \gamma_D/\gamma_S & 0 \\ 0 & 0 & \beta_D/\beta_S \end{bmatrix}$$

最后，通过逆矩阵 \mathbf{M}_{BFD}^{-1} 将数据转换回标准的CIE XYZ空间。

综合起来，这形成了一个完整的线性变换矩阵 \mathbf{M}_{CAT} ：

$$\mathbf{M}_{CAT} = \mathbf{M}_{BFD}^{-1} \cdot \mathbf{D} \cdot \mathbf{M}_{BFD}$$

2-1-3 白平衡解码向对角的简化

近年来Adobe选择把CAT与色彩空间转换矩阵封装在一起成为ForwardMatrix。解码的时候默认以ForwardMatrix进行运算。

$$\begin{bmatrix} R_L \\ G_L \\ B_L \end{bmatrix}_{D65} = \underbrace{\left(M_{sRGB}^{-1} \cdot \text{CAT}_{AW \rightarrow D65} \cdot \underline{T}_i \cdot \underline{D}_i^{-1} \right)}_{\text{ForwardMatrix}} \cdot \underbrace{\underline{D}}_{\text{WB}} \cdot \begin{bmatrix} \mathcal{R} \\ \mathcal{G} \\ \mathcal{B} \end{bmatrix}_{\text{scene}}$$

再对不同白点下的ForwardMatrix插值出最终使用的阵。

而类似dcraw的很多解码办法会选择在相机基下先对灰，再进行色彩空间变换。

$$\begin{bmatrix} R_L \\ G_L \\ B_L \end{bmatrix}_{D65} \approx M_{sRGB}^{-1} \cdot T_{D65} \cdot \begin{bmatrix} \text{Gain}_R & 0 & 0 \\ 0 & \text{Gain}_G & 0 \\ 0 & 0 & \text{Gain}_B \end{bmatrix} \cdot \begin{bmatrix} \mathcal{R} \\ \mathcal{G} \\ \mathcal{B} \end{bmatrix}_{\text{scene}}$$

其中各颜色增益为D65下码值与AdoptedWhite码值之比。

2-2 黑电平的意义及确定机制

为了使得色彩空间转换可以在一个零点良好的线性空间中进行，正确的处理流程必须首先在AFE或ISP预处理阶段构建一个能够容纳完整噪声统计分布的基座。基座的创建办法为连接一带偏置电阻的反相输入端，使放大器撬起输出的基准线。

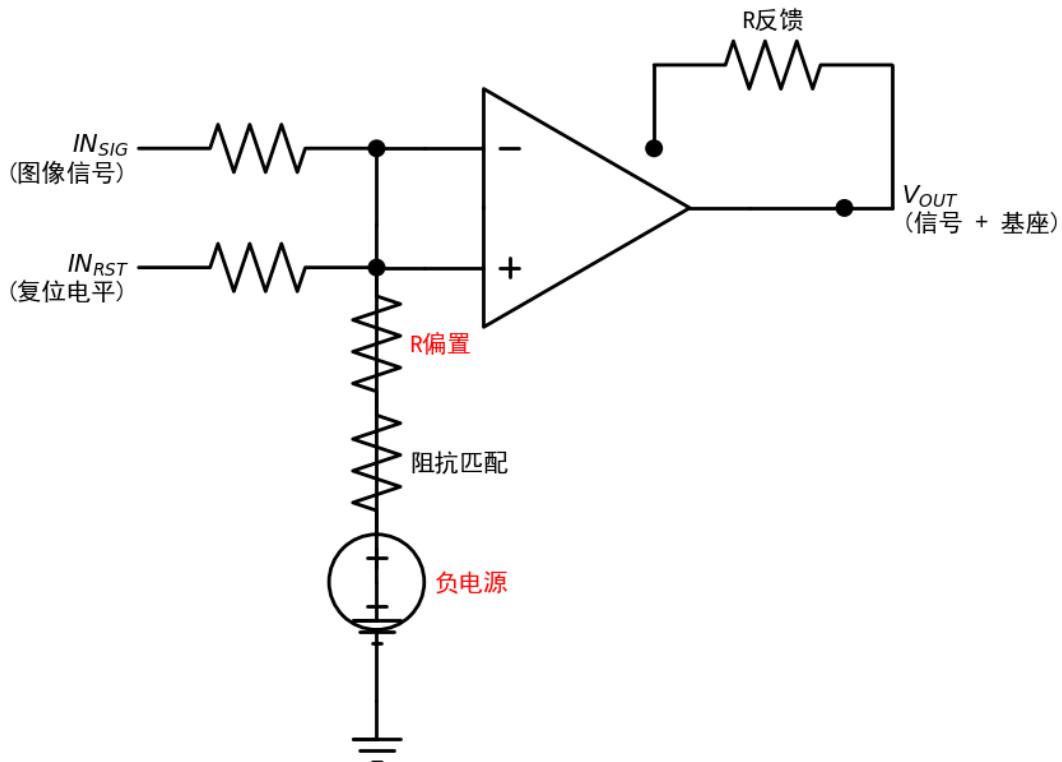


图2-2-1

针对光学黑电平的扣减，工业界常采用混合信号负反馈回路来实现：系统在模拟域实时采样OB像素的电压，将其与寄存器中预设的Clamp Target进行比较，计算出误差后控制数模转换器DAC产生一个模拟修正电压。这个电压被直接叠加在原始信号上，从而将信号的物理零点强行锚定在设定的基座之上。

由于图像传感器的RN本质上服从以黑电平为均值的高斯分布，在极低照度下，部分像素的物理信号电平会因噪声波动而低于理论黑电平均值。若简单地将ADC的零输入参考电平与物理黑电平完全对齐，或者在数字域执行黑电平扣除时强制将负值截断，将导致噪声概率密度函数的左侧尾部被非物理地移除。这种零点截断会引入整流误差，迫使暗部信号的统计均值人为抬升，从而导致暗部线性度的丧失。

近十年越来越多的相机可以通过AFE中的偏置DAC，将光学黑像素的钳位目标值设定在稍高于ADC物理零点的位置。BL等于零的相机大部分未采取此办法。这一人为抬升的“模拟基座”为负向的噪声波动预留了足够的下溢空间。在随后的ISP链路中，系统允许产生并保留有符号的结果，或在更高位深的无符号空间中维持一个

计算基座。在这样一个保留了负信号统计完整性的线性空间中，传感器才可以做到在高噪声下的解码插值及响应的完全线性，保证后续的正确处理。

3 极低噪下的量化噪声模型修正

这个偏置往往不是精确的——偏置不总可以精确地落在比零输入低给定的黑电平处（例如1008）。可能是1008.2，也可能是1008.7。由于他落在了1008和1009之间，解码时只能当作1008或者1009这样的整数来处理。这引入了很多问题。我在3-2中给出了一个改进的解码办法，3-1中我讨论这种偏离带来的噪声以及动态范围微弱差异。

3-1 平场中的噪声分析

3-1-1 理论噪声模型

设平场总噪声 N_{total} 由 M 个独立随机过程叠加构成：

$$N_{total} = \sum_{i=1}^M X_i$$

其中 $M \sim 10^9$ （对应百万像素传感器的载流子数），每个 X_i 满足 $E[X_i^2] < \infty$ 。
根据 Lindeberg-Lévy 中心极限定理：

$$\frac{\sum_{i=1}^M X_i - M\mu}{\sqrt{M}\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

该收敛的充分条件在半导体器件中完全满足：① 各电子运动相互独立；② 载流子数 $M \gg 1$ ；③ 涨落能量有限。以下分别证明各噪声源的具体高斯性。

光子到达过程服从泊松分布。对于量子效率为 η 的像素，在曝光时间 T_{exp} 内收集的光电子数：

$$N_{pe} \sim \text{Poisson}(\lambda), \quad \lambda = \frac{P \cdot T_{exp} \cdot \eta}{h\nu}$$

当 $\lambda \rightarrow \infty$ 时，标准化变量：

$$Z = \frac{N_{pe} - \lambda}{\sqrt{\lambda}}$$

依分布收敛于标准正态分布。下证。泊松分布的特征函数为：

$$\varphi_P(t) = \exp [\lambda(e^{it} - 1)]$$

对于标准化变量 Z ，其特征函数：

$$\varphi_Z(t) = E[e^{itZ}] = \exp(-it\sqrt{\lambda}) \cdot \exp[\lambda(e^{it/\sqrt{\lambda}} - 1)]$$

对指数项进行二阶泰勒展开：

$$e^{it/\sqrt{\lambda}} = 1 + \frac{it}{\sqrt{\lambda}} - \frac{t^2}{2\lambda} + O(\lambda^{-3/2})$$

当 $\lambda \rightarrow \infty$, 高阶项消失, $\varphi_Z(t) \rightarrow \exp(-t^2/2)$, 即标准正态分布的特征函数。

对于典型传感器, $\lambda \approx 10^3\text{-}10^5 \text{ e}^-$, 远超收敛阈值 ($\lambda > 20$), 故高斯近似误差 $< 1\%$ 。

MOS电容复位噪声服从涨落-耗散定理, 其均方根电压为

$$V_n = \sqrt{\frac{kT}{C}}$$

该噪声严格高斯分布, 与载流子统计无关。

源跟随器噪声包含热噪声与闪烁噪声:

$$i_{n,total}^2 = i_{n,thermal}^2 + i_{n,flicker}^2 = \frac{8kT}{3} g_m \Delta f + K_f \frac{I_D^{A_f}}{f^{E_f}}$$

沟道热噪声与复位噪声同源, 均源于载流子热运动, 其电流涨落为高斯分布。

闪烁噪声源于Si-SiO₂界面陷阱的随机俘获/发射, 每个陷阱产生双态马尔可夫过程, 总电流扰动为所有陷阱贡献的叠加。根据中心极限定理, 大量独立双态过程的叠加收敛于高斯分布。

独立高斯噪声的线性组合仍服从高斯分布。设热噪声分量 $i_{th} \sim \mathcal{N}(0, \sigma_{th}^2)$, 闪烁噪声分量 $i_f \sim \mathcal{N}(0, \sigma_f^2)$, 且相互独立, 则总噪声:

$$i_{total} = i_{th} + i_f \sim \mathcal{N}(0, \sigma_{th}^2 + \sigma_f^2)$$

$$\sigma_{SF}^2 = \frac{8kT}{3g_m} \Delta f + K_f I_D^2 \ln \left(\frac{f_{max}}{f_{min}} \right)$$

3-1-2 综合噪声模型与实验验证

3-1-2-1 综合噪声模型

根据线性叠加原理, 像素输出噪声为各独立分量的和:

$$N_{total} = N_{shot} + N_{dark} + N_{reset} + N_{SF}$$

基于各分量的高斯性, 总噪声服从:

$$N_{total} \sim \mathcal{N}(\mu, \sigma_{total}^2)$$

其中:

$$\sigma_{total}^2 = \sigma_{shot}^2 + \sigma_{dark}^2 + \sigma_{reset}^2 + \sigma_{SF}^2$$

3-1-2-2 实验验证

我使用远距离单灯 (~10m) , 并给相机卡口覆盖一个均匀的帘幕, 这种情况下我让相机连续拍摄两张, 再使用JiangtherapeeOnline点对点进行差分, 输出差分帧。在这种情况下, 噪声功率是原先两倍。不应当在现在补偿倍数。在输出差分帧的时候补偿容易出现scaling导致的ptc弯曲, 下文中我会详细建模。

取二平场图像为例: 图3-1-2-2-1为ILCE-7M3在6.46码值的的红通道以及一个Mean=518.46、Sigma=1.65的高斯拟合。



图3-1-2-2-1

图3-1-2-2-2为ILCE-7M3在17.12码值的的红通道以及一个Mean=531.47、Sigma=2.03的高斯拟合。

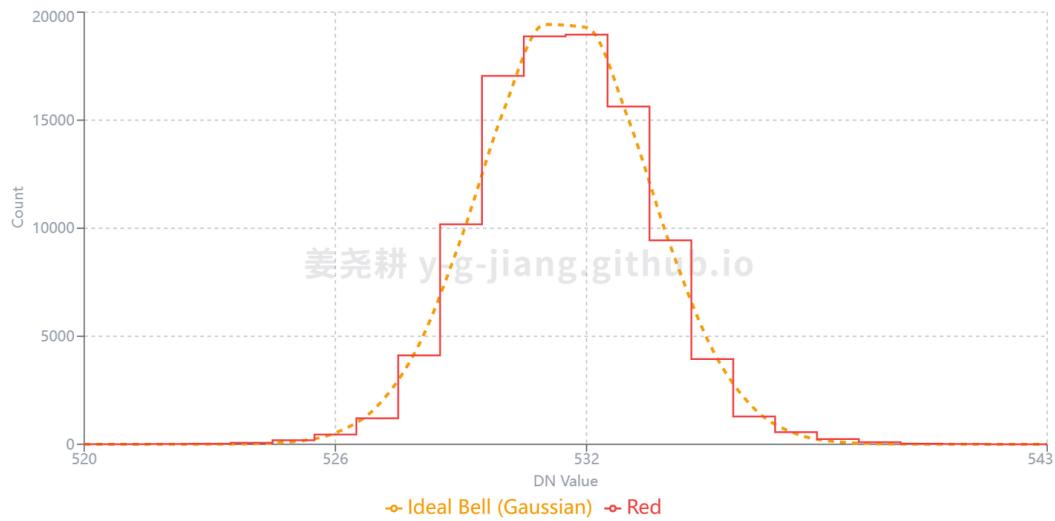


图3-1-2-2-2

图3-1-2-2-3为该码值下使用我的办法构建的平场能量谱 (堆叠平均后) 。

低码值处例机的非线性降噪会让结果略微偏离高斯。但低码值平场的码值跨度较小, 降噪在此码值范围内可近似视作线性。

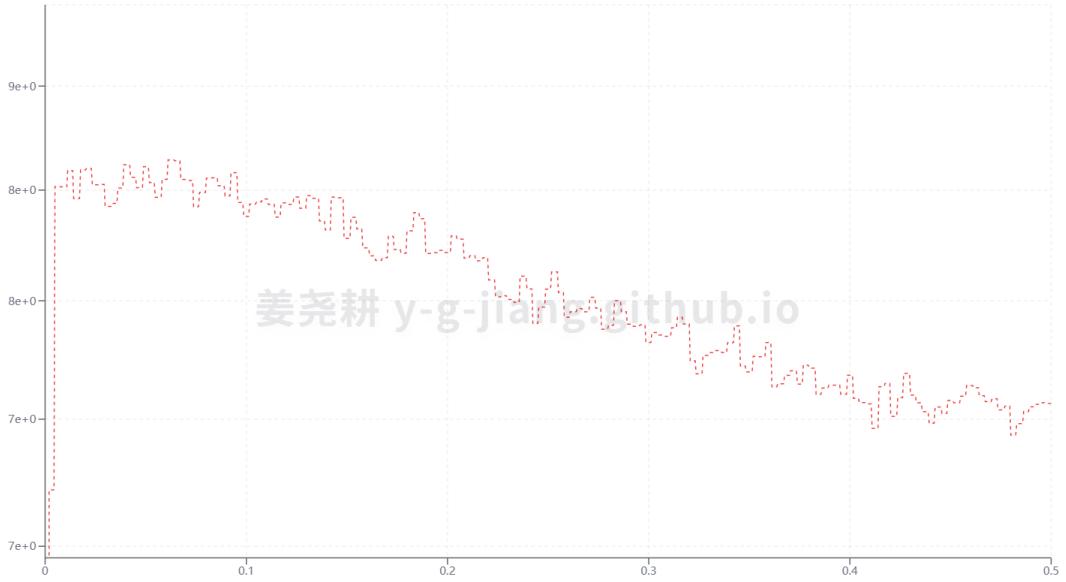


图3-1-2-2-3

对于线性降噪，设原始噪声 $n(i, j) \sim \mathcal{N}(0, \sigma^2)$ ，滤波器核为 $h(k, l)$ 。

输出噪声：

$$\tilde{n}(i, j) = (h * n)(i, j) = \sum_k \sum_l h(k, l) \cdot n(i - k, j - l)$$

此时每个输出像素是有限个独立高斯变量的线性组合：

$$\tilde{n} = \sum_{m=1}^M a_m \cdot X_m, \quad X_m \sim \mathcal{N}(0, \sigma^2)$$

根据高斯分布的线性不变性：

$$\tilde{n} \sim \mathcal{N}\left(0, \sigma^2 \sum_m a_m^2\right)$$

在高码值处，PRNU等乘性噪声使得分布偏离高斯。但该时的噪声已经足够大至均匀假设也可完美适配模型，故不作分析。

3-2 高斯模型下的量化噪声

如上文所述。对信噪比的分析一般在平场中进行，而平场噪声是无数独立随机过程的总和，其分布是趋近于高斯分布的。

常见情况下人们通常假设：信号跨越了大量的量化台阶且信号的变化相对于量化步长是随机且剧烈的。

此时，量化误差 $e = Y - X$ 被假设为在区间 $[-0.5, 0.5]$ 上服从均匀分布。对于均匀分布，其方差为：

$$\sigma_q^2 = \int_{-0.5}^{0.5} x^2 dx = \left[\frac{x^3}{3} \right]_{-0.5}^{0.5} = \frac{1}{12}$$

总方差通常被认为是输入噪声与量化噪声的平方和:

$$\sigma_{total}^2 = \sigma^2 + \frac{1}{12} \implies \sigma_{total} = \sqrt{\sigma^2 + 0.0833}$$

但如果设模拟输入信号 X 服从高斯分布 $X \sim \mathcal{N}(\mu, \sigma^2)$, 其概率密度函数 (PDF) 为:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

量化器 $Q(\cdot)$ 仍定义为具有单位步长 $\Delta = 1$ 的四舍五入操作:

$$Y = Q(X) = \text{round}(X)$$

输出 Y 取整数 k 的概率 P_k 为输入 PDF 在量化区间 上的 $(k - 0.5, k + 0.5]$ 积分:

$$P_k = P(Y = k) = \int_{k-0.5}^{k+0.5} f_X(x) dx = \Phi\left(\frac{k+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{k-0.5-\mu}{\sigma}\right)$$

其中 $\Phi(\cdot)$ 为 CDF。

$$E[Y] = \sum_{k=-\infty}^{\infty} k \cdot P_k$$

$$E[Y^2] = \sum_{k=-\infty}^{\infty} k^2 \cdot P_k$$

$$\sigma_{out,exact}^2(\mu, \sigma) = E[Y^2] - (E[Y])^2$$

下图 (图3-2-1) 为我对输入标准差、完全对齐量化网格后的高斯分布的标准差的解析解、均匀分布假设的解析解以及蒙特卡洛采样模拟点的绘制。

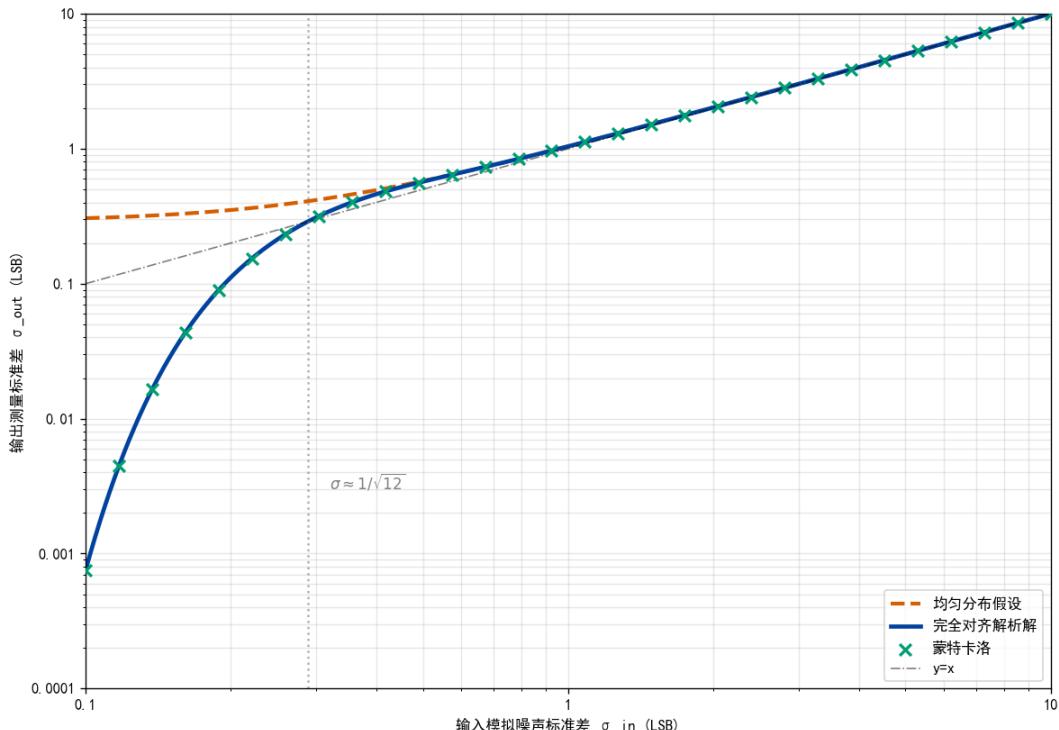


图3-2-1

下表为我按照对数均匀撒点给出的。三列分别为输入噪声、贝奈特模型也就是均匀输入模型的噪声源以及我对高斯模型下的推导。对于输入噪声大于0.5LSB的时候，贝奈特模型就可以非常完美地拟合高斯模型了。

Input_Sigma	Bennett_Theory	Aligned_Exact
0.1000	0.3055	0.0008
0.1274	0.3155	0.0093
0.1624	0.3312	0.0456
0.2069	0.3552	0.1252
0.2637	0.3910	0.2407
0.3360	0.4430	0.3698
0.4281	0.5164	0.4942
0.5456	0.6172	0.6143
0.6952	0.7527	0.7526
0.8859	0.9317	0.9317
1.1288	1.1652	1.1652
1.4384	1.4671	1.4671
1.8330	1.8556	1.8556
2.3357	2.3535	2.3535
2.9764	2.9903	2.9903
3.7927	3.8037	3.8037
4.8329	4.8415	4.8415
6.1585	6.1652	6.1652
7.8476	7.8529	7.8529
10.0000	10.0042	10.0042

Input_Sigma	Bennett_Err	Aligned_Err
0.1000	0.2055	-0.0992
0.1274	0.1881	-0.1181
0.1624	0.1688	-0.1168
0.2069	0.1483	-0.0817
0.2637	0.1273	-0.0230
0.3360	0.1070	0.0338
0.4281	0.0882	0.0661
0.5456	0.0717	0.0687
0.6952	0.0576	0.0575
0.8859	0.0458	0.0458
1.1288	0.0363	0.0363
1.4384	0.0287	0.0287
1.8330	0.0226	0.0226
2.3357	0.0178	0.0178
2.9764	0.0140	0.0140
3.7927	0.0110	0.0110
4.8329	0.0086	0.0086
6.1585	0.0068	0.0068
7.8476	0.0053	0.0053
10.0000	0.0042	0.0042

3-3 极低噪声下基座偏移的模型修正

此时我们加入对量化网格偏移的影响的讨论。

在3-2章中我们有：

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

定义量化函数 $Q(x) = \text{round}(x)$ ，量化后的均值：

$$\mu_N = E[Q(X)] = \int_{-\infty}^{\infty} \text{round}(x) \cdot f(x; \mu, \sigma) dx$$

最终带入方差定义式开方：

$$\sigma_N(\mu, \sigma) = \sqrt{E[N^2] - \mu_N^2} = \sqrt{\left(\int_{-\infty}^{\infty} \text{round}(x)^2 f(x; \mu, \sigma) dx \right) - \left(\int_{-\infty}^{\infty} \text{round}(x) f(x; \mu, \sigma) dx \right)^2}$$

分别对不同均值的钟形曲线绘图：

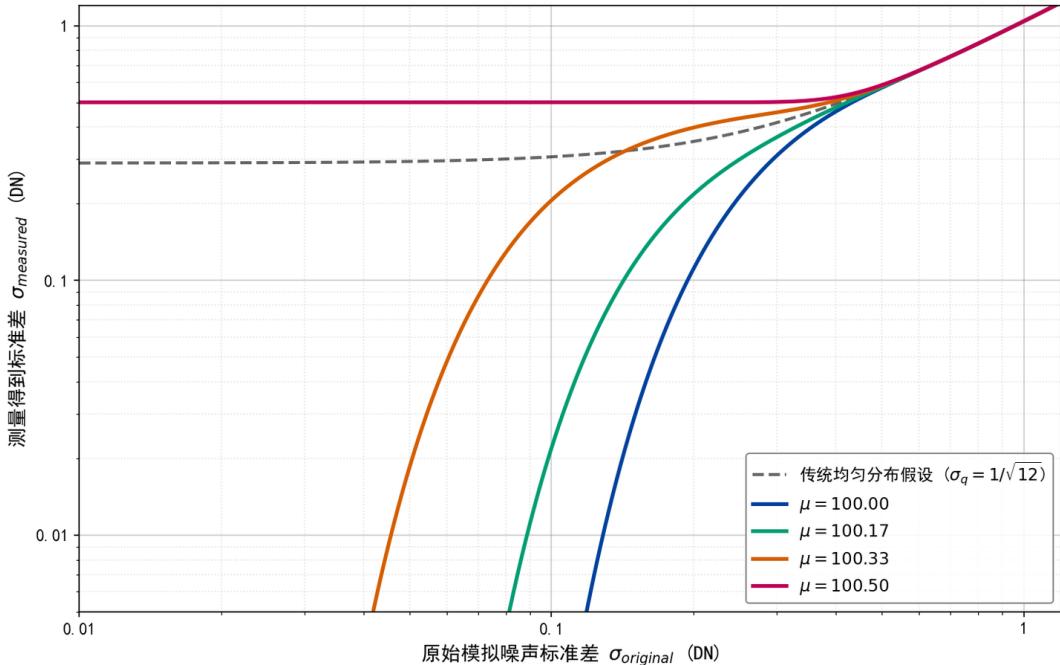


图3-3-1

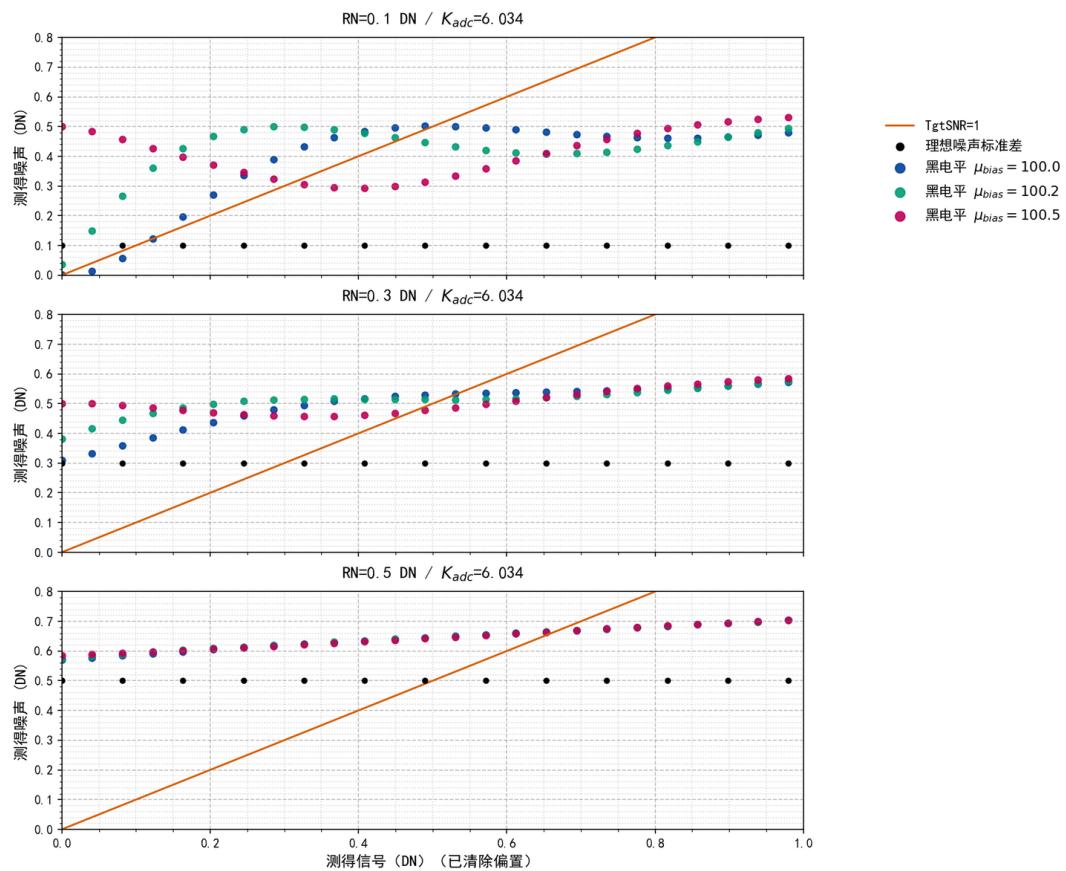


图3-3-2

在上图可以得到印证：14bit全幅，在RN>=0.5时，普通ff的阴影点附近完全可以以均匀分布假设来计算量化噪声来分析动态范围测试结果；在RN<0.3时，普通ff阴影点附近的量化噪声严重偏离均匀分布假设的计算结果。

目前14bit全幅RNDN最低的一批机器刚好在这个阈值附近。如果FFCmos进一步发展，CIS厂商就必须关注量化网格的差异导致的动态范围测量结果差异了。

3-4 高斯对均匀假设补偿干扰的PTC模拟

以ILCE-7M3为蓝本，我使用JiangtherapeeOnline模拟出了在类似CIS中进行低噪声合成（例如DGO）后的一些结果。

对于给定的入射光强，我对上文的讨论简化得，对于FPN/PRNU不强的地方，传感器输出的数字信号的统计特性由光子散粒噪声、读出噪声以及量化噪声共同决定。总时间噪声在电子域的方差 $\sigma_{total_e}^2$ 可表示为：

$$\sigma_{total_e}^2 = \sigma_{shot_e}^2 + \sigma_{read_e}^2$$

其中， $\sigma_{shot_e}^2$ 为光子散粒噪声方差，服从泊松分布特性，其值等于光电子产生的平均信号量 μ_e ； σ_{read_e} 为等效输入噪声。根据系统增益 K ，理想的无噪声数字信号与光电子数的关系为 $\mu_e = S_{ideal} \cdot K$ 。

我针对上述每个采样点，采用蒙特卡洛方法生成 $N = 2000$ 个独立的像素样本以计算统计矩。模拟数据生成时先将目标净数字信号 S_{net} 转换为平均光电子数 $\mu_e = S_{net} \cdot K$ ，再基于高斯近似，生成服从分布 $\mathcal{N}(\mu_e, \mu_e + \sigma_{read_e}^2)$ 的随机样本序列，模拟光子到达与读出过程的随机涨落。然后将含噪电子信号转换回数字域并叠加黑电平偏置： $D_{raw} = (S_{electron}/K) + BL$ 。最终模拟ADC的离散化特性，对 D_{raw} 执行四舍五入取整操作，并限制在 $[0, 2^{14} - 1]$ 范围内。

共三组模拟图（图3-4-1），每组左右两图。自上至下分别为RN=0.7、0.45、0.3的模拟。（图3-4-1 2,4,6）的PTC拟合线（黑线）为三参数直接的模拟，（图3-4-1 1,3,5）的PTC拟合线为三参数仅更改RN至均匀分布的量化噪声假设。M.Tgt紫色水平线为无FPN情况下8mp中SNR=1的TgtSNR线。



图3-4-1

近期索尼发布了一系列读出噪声接近0.5LSB的系统（例如ILCE-7M5）。在其情形下，对应上组图中第二组的模式。量化噪声的拟合还是相当精确的——对于PTP DR以及8mpSNR=1的LandscapeDR在PTC正确以均匀分布量化假设补偿下的影响偏差在0.01EV以内。但注意第三组的上下两图对比：在读出噪声（Measured）在0.4DN的情况下，对于类似ILCE-7M3的模型，计算得的DR结果就会受量化网格相位影响较大，均匀量化模型也不再适用了。这也是与前面我的推论基本相同的。

4 量化网格漂移致偏色及改进的解码办法

消费级图像传感器在极低照度下因上面描述的黑电平漂移和低位深量化会导致量化网格的错位误差。尽管该误差的绝对幅值极小 ($|\delta| < 0.5 \text{ DN}$)，但在执行白平衡等色彩解码时，由于 R、G、B 通道的增益系数 g_c 往往差异显著（通常 $g_b > g_r > 1$ ），各通道残留的小数漂移被非一致性放大。在极低照度区域，这种被放大的差异性偏置破坏了 RGB 信号的辐射度比例，导致视觉上无法消除的底色污染。

为此本文提出了一种“位深扩展-重对齐-逆归一化”的预处理策略。我通过注入精确的物理黑电平并模拟错误的量化基座，将原本发生在低位的非线性截断误差搬运至高位深空间中的可忽略区域，从而在最终输出中实现暗部线性度的物理保真。

4-1 漂移量的感知结果

为评估量化网格漂移在 ISP 链路中的感知放大效应，我构建了一个前向逆向耦合的小信号色偏模型。令相机标定得到的 $\text{XYZ} \rightarrow \text{Camera RGB}$ 矩阵即为 ColorMatrix 为

$$M_{\text{fwd}} \in \mathbb{R}^{3 \times 3},$$

则 Camera RGB $\rightarrow \text{XYZ}$ 的校色矩阵 (CCM) 为其逆

$$M_{\text{ccm}} = M_{\text{fwd}}^{-1}.$$

对于任意目标明度 $L^* \in [0.5, 50]$ ，首先构造中性色

$$\text{Lab}_{\text{ideal}} = (L^*, 0, 0),$$

并通过 CIE D65 白点将其转换为 XYZ 三刺激值

$$\text{XYZ}_{\text{ideal}} = \text{Lab}_{\text{D65}}^{-1}(L^*, 0, 0).$$

继而利用 M_{fwd} 完成「逆向编码」：

$$\text{Cam}_{\text{lin}} = M_{\text{fwd}} \text{XYZ}_{\text{ideal}}, \quad \text{Cam}_{\text{DN}} = \text{Cam}_{\text{lin}} \cdot (2^{14} - 1),$$

得到 14-bit 线性码值。此时按照黑电平真实漂移情况注入亚 LSB 级漂移

$$\Delta \text{Cam}_{\text{DN}} \in \mathbb{R}^3,$$

其分量可正可负，且按 2 的幂次提亮档位 $G = 2^n$ 进行比例缩放：

$$\text{Cam}'_{\text{DN}} = \text{Cam}_{\text{DN}} + G \cdot \Delta \text{Cam}_{\text{DN}}.$$

完成扰动后，执行向前解码：

$$\text{Cam}_{\text{wb}} = \text{Cam}'_{\text{DN}} \odot \mathbf{g}_{\text{wb}}$$

$$XYZ_{\text{drift}} = M_{\text{ccm}} \left(\frac{\text{Cam}_{\text{wb}}}{2^{14} - 1} \right)$$

$$\text{Lab}_{\text{drift}} = \text{XYZ}_{\text{D65}}^{-1}(\text{XYZ}_{\text{drift}})$$

最终感知误差由 CIEDE2000 度量：

$$\Delta E_{00} = \Delta E_{00}(\text{Lab}_{\text{ideal}}, \text{Lab}_{\text{drift}}).$$

通过在线性均匀采样的 L^* 序列上重复上述流程，即可获得下图（图4-1-1、4-1-2）所示的色偏衰减曲线。图中箭头高度直接对应 ΔE_{00} ，颜色则由提亮后偏色方向的 sRGB 坐标给出，从而在同一图表内同时承载误差大小与误差方向两种信息。

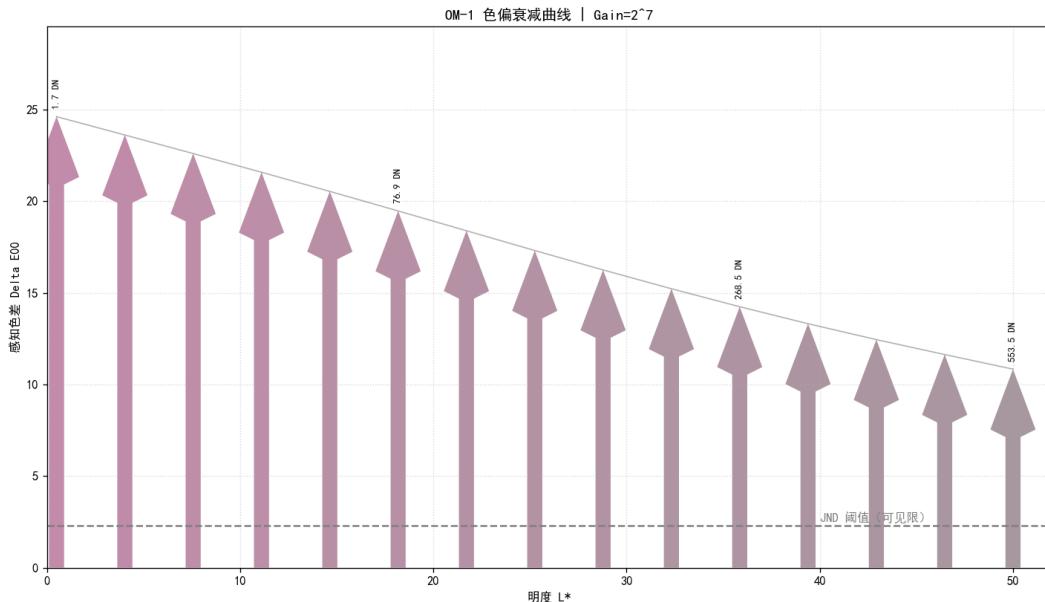


图4-1-1

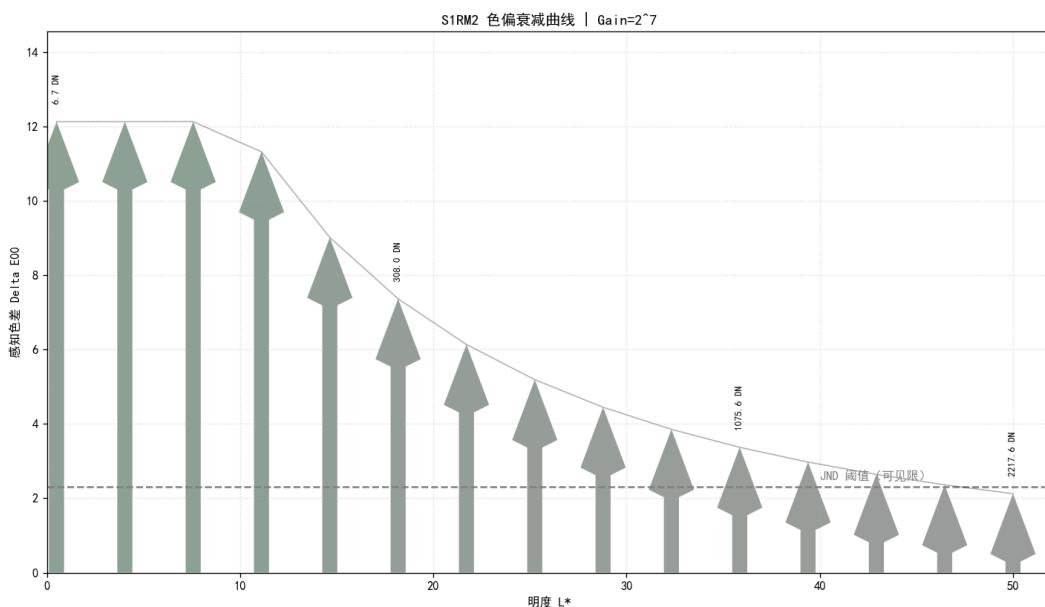


图4-1-2

4-2 改进的解码办法

如4-1-1图所模拟，OM-1在提亮后有剧烈偏紫倾向。4-2-1中左图为奥林巴斯OM-1的dpr ettl6标版使用AdobeCameraRaw解码并提亮六档的标版，右图为经过本文算法处理后的解码。4-2-2 中左图为LUMIX G9的dpr iso200 light off标版使用AdobeCameraRaw解码并提亮六档的标版，右图为经过本文算法处理后的解码。可见本文提出的解码办法可对现有解码管线下黑电平钳位不准导致的暗部失线性做到极佳的改善。

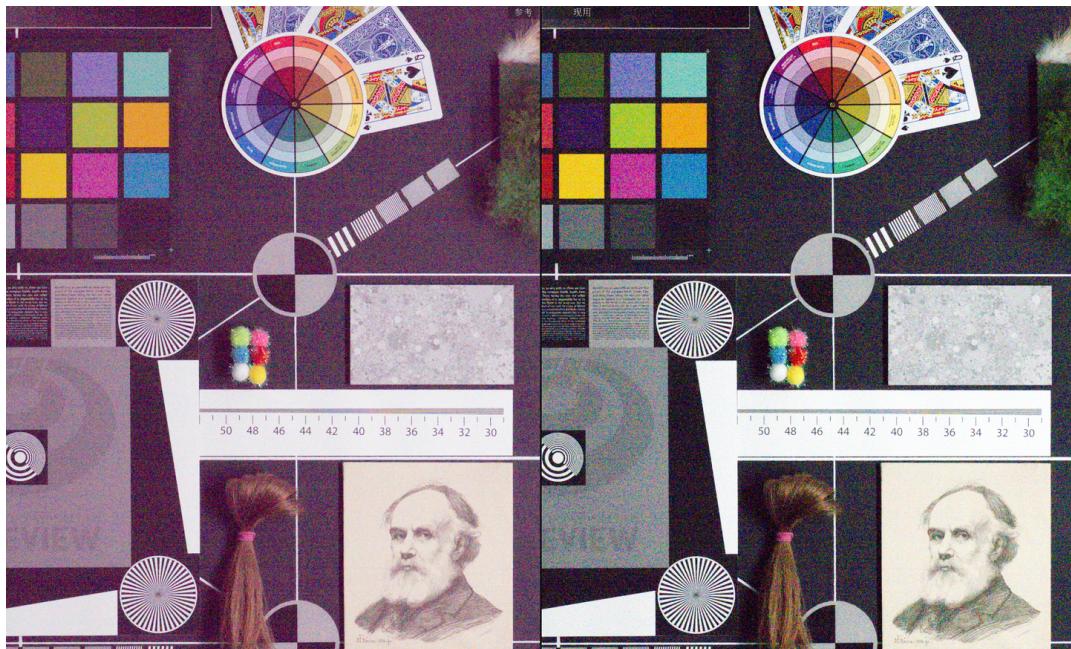


图4-2-1



图4-2-2

定义以下变量：

- x_{raw} : 传感器输出的原始 Raw 数据（整数）。
- B_{wrong} : ISP 管线默认使用的、可能存在漂移的黑电平。
- B_{true} : 通过暗场校准测得的物理真值黑电平。
- N_{in} : 输入数据的位深。

- N_{target} : 处理过程中的高位深。
- k_{gain} : 位深扩展系数, 定义为 $k_{gain} = 2^{(N_{target}-N_{in})}$ 。

首先, 我们在高精度域中, 从原始观测值中扣除物理真值黑电平, 得到真实的物理光子信号分量:

$$S_{true} = x_{raw} - B_{true}$$

为了适配原有的整数解码管线 (该管线通常假设输入为 N_{in} 位深且黑电平为 B_{wrong}) , 我们将真实的物理信号放大到高位深域, 并人为叠加原管线预期的黑电平基座。

构建修正后的高位深数值 Y_{high} :

$$Y_{high} = \text{round}(S_{true} \times k_{gain} + B_{wrong})$$

此时原本的暗部信号 S_{true} 被放大了 $2^{N_{target}-N_{in}}$ 倍, 原本的基座误差 ($B_{wrong} - B_{true}$) 相对于被放大后的信号而言, 其权重被显著稀释。

我将构建好的 Y_{high} 输入到标准的转换流程中。

在原有的错误管线中, 解码输出 Out_{orig} 为:

$$Out_{orig} = \text{Process}(x_{raw} - B_{wrong})$$

由于 $B_{wrong} \neq B_{true}$, 当 $x_{raw} \approx B_{true}$ (极暗部) 时, $(x_{raw} - B_{wrong})$ 会产生显著的相对误差, 导致色偏。在使用本方法后, 虽然管线依然执行减去 B_{wrong} 的操作, 但实际上它处理的是:

$$\text{Input}_{pipe} = Y_{high} \quad (\text{视为高位深数据})$$

管线执行的等效去黑操作为:

$$\text{Net}_{pipe} = Y_{high} - B_{wrong} = (x_{raw} - B_{true}) \cdot k_{gain}$$

最后进行解码, 在线性色彩空间中除回原值。

经过这一系列变换, 原本由 B_{wrong} 引入的固定偏置误差在最终输出图像中被等效为了一个极小量。其视觉影响等效于在原位深下、距离暗部 ($N_{target} - N_{in}$) 个档位处的偏色。

例如, 若从 8-bit 扩展到 16-bit ($k_{gain} = 2^8 = 256$), 原先 1 DN 的黑电平漂移带来的色偏, 现在被压缩到了 1/256 DN 的量级, 在视觉上完全不可见。这实现了在不修改 ISP 固件代码的前提下, 通过数据预处理彻底消除暗部基座色偏。此算法现在已完全落地(<https://y-g-jiang.github.io/DCBL.html>)。

5 Prescaling仿射重量化模型分析

5-1 Prescaling的发现与建模

在25年7月，我在对商用图像传感器的 RAW 数据输出进行特性分析时，观测到了广泛存在于诸多图像传感器输出的码值断裂。基于下文中对其产生机理的溯源，我把它命名为“对角预调制”。因为他等效于一个对线性良好的数据一个乘减的操作，也即作用于与色彩空间转换同零点的线性空间中的对角阵。英文中我命名其为“wb prescaling”。因为在我于2-1-3章中给出的第二种以旋转矩阵规范色彩空间的流程中，白平衡矫正采取的也是一个对角阵。wb prescaling对数据的影响是类似白平衡预调制的。下文中我会对这种操作进行建模并讨论他对动态范围的影响。



图5-1-1

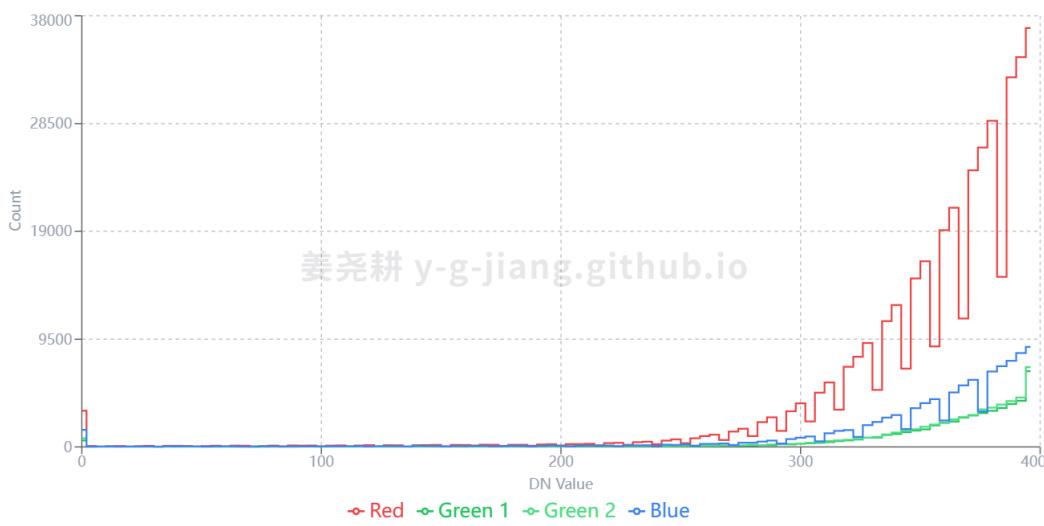


图5-1-2

看上二图：D850的黑位码值是400附近。使用JiangtherapeeOnline，我把低于黑电平的直方图给出。低于黑位的部分也是有可见的断码。这表明其ISP采取了一种扣减黑位后乘倍率再加回的处理方式，其中扣减是不截零的。

图5-1-2可见码值粘连，也即并不是完全的错开。这种情况往往是由于四种原因：
1，相位点插值；2，降噪；3，坏点补偿；4，降采样/压缩算法。D850在12800的直方图粘连来自第二种原因。

使用JiangtherapeeOnline进行量化阶梯分析：

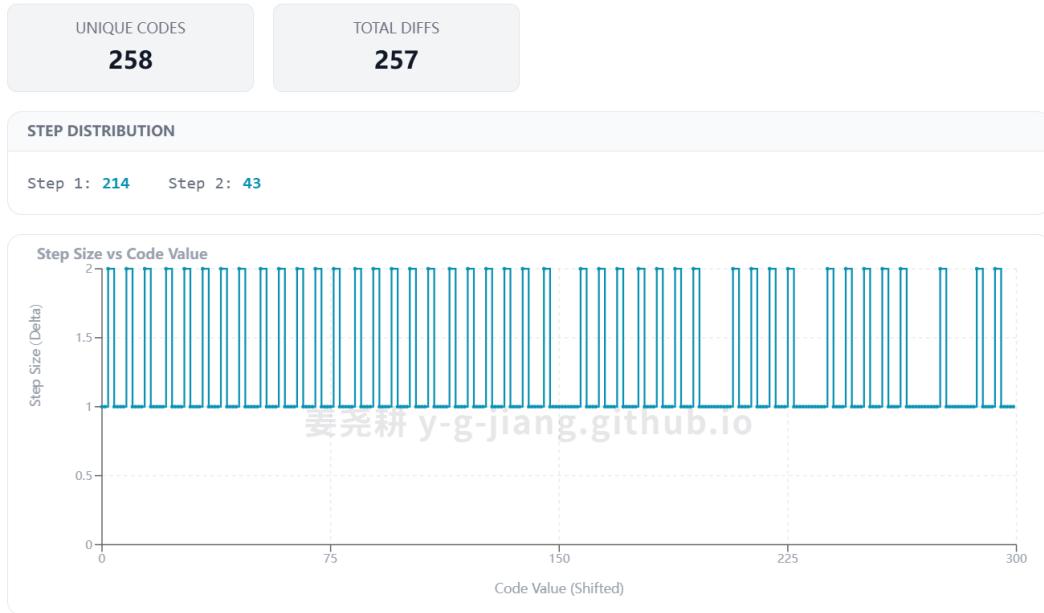


图5-1-3

我们发现D850有着明确的六个码值一空的规律。

为了保证线性性以及零点的正确性，这个ISP操作一定是乘减的。我可以下式构建：

$$y = k \cdot (\alpha - B) + B$$

式中 α 即为prescaling的倍率。经上面测量不难发现D850的prescaling倍率在1.165左右。

5-2 Prescaling对PTC光子转移曲线的影响

5-2-1 黑电平精确的情况下PTC的刻画

5-2-1-1 实验模拟

如3-4（高斯对均匀假设补偿干扰的PTC模拟）一章所讨论，在RNDN>0.5、Kadc正常时（也即绝大部分情况——我有理由相信当RNDN普遍小于0.5时厂商会选择更高位深的容器），我们可以放心来用对0.001以上码值的结果进行PTC拟合。

同样设输入信号 x 为服从泊松-高斯混合分布的连续随机变量，变换公式为：

$$y = Q(k \cdot x) = \text{round}(k \cdot x)$$

其中 k 为预缩放系数（例如上文中的典型值 1.165/1.167）。

当 $k > 1$ 且为非整数时，输出信号 y 的取值空间虽然仍为整数集 \mathbb{Z} ，但其对应的输入信号区间不再恒定。

对于输出整数 m ，其对应的输入区间宽度 Δ_m 为 $\Delta_m = \frac{1}{k}$ 。我可以把他等效为一个周期性变化的量化噪声的引入。不同于贝内特假设中的恒定方差 $\frac{1}{12}$ ，此时的量化误差方差 σ_q^2 成为信号均值 μ_x 的函数。

图5-2-1-1 a中即为我以蒙特卡洛模拟后的结果；图5-2-1-1 b为我直接对S5M2的PTC测试拍摄的RAW进行DN后乘减处理再输入JiangtherapeeOnline后的输出结果。

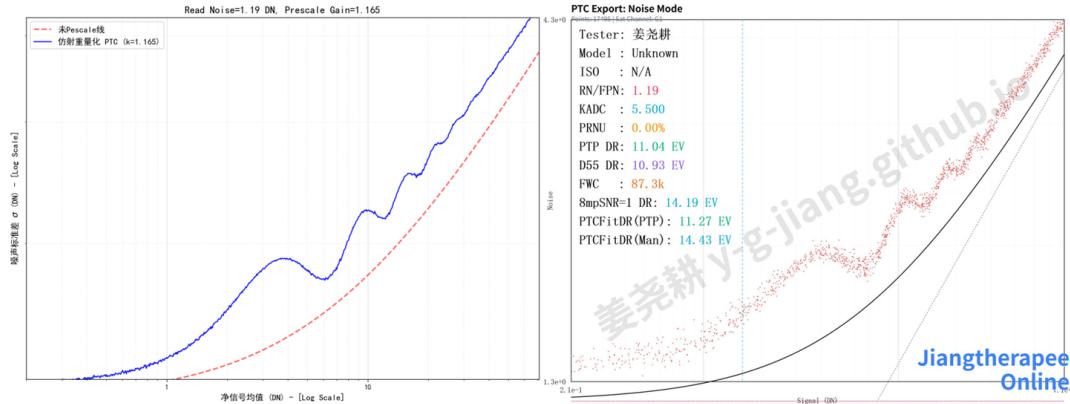


图5-2-1-1 a/b

5-2-1-2 数学模型

3-2章中，我已经给出了输出整数 n 的概率 $P(n)$ 严格等于模拟信号落在量化区间 $(n - 0.5, n + 0.5]$ 内的概率积分：

$$P_{in}(n) = \int_{n-0.5}^{n+0.5} f_X(x) dx = \Phi\left(\frac{n + 0.5 - \mu}{\sigma_{in}}\right) - \Phi\left(\frac{n - 0.5 - \mu}{\sigma_{in}}\right)$$

定义 $g(n)$ 为数字增益映射。在黑电平对齐且增益为 k 的条件下可视作：

$$m = g(n) = \text{round}(k \cdot n)$$

该映射将输入整数空间 \mathbb{Z}_{in} 映射到输出整数空间 \mathbb{Z}_{out} 。由于 k 为非整数（如 1.165），该映射发散且非均匀的。

$$E[Y] = \sum_{n=-\infty}^{\infty} \text{round}(k \cdot n) \cdot P_{in}(n)$$

$$E[Y^2] = \sum_{n=-\infty}^{\infty} [\text{round}(k \cdot n)]^2 \cdot P_{in}(n)$$

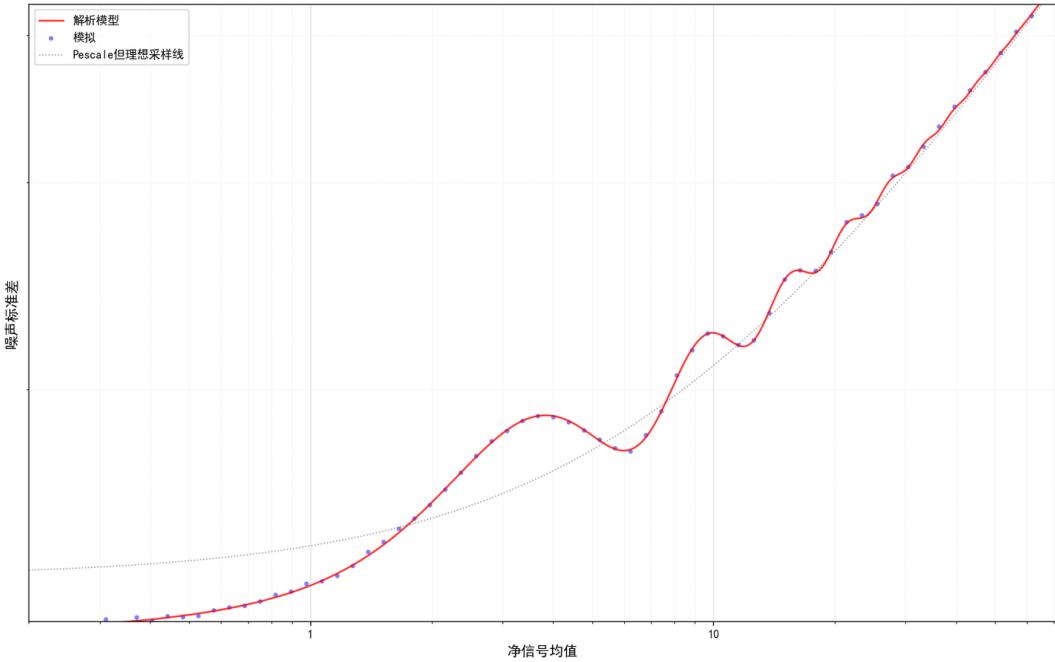
最终测量方差为

$$\sigma_{\text{measured}}^2 = \sum_{n=-\infty}^{\infty} [\text{round}(kn)]^2 P_{in}(n) - \left[\sum_{n=-\infty}^{\infty} \text{round}(kn) P_{in}(n) \right]^2$$

在PTC图表中，我们将 σ_y 对 μ_y 作图。由于上述震荡项的存在，测量得到的噪声标准差将在理论值 $k\sigma_x$ 上下波动。

对于 $k = 1.165$ ，震荡的周期由 k 的小数部分 $0.165 \approx 1/6$ 决定。这意味着大约每隔6个输入码值，量化误差的分布形态就会经历一个完整的“压缩-扩张”循环。

PTC曲线不再平滑，而是呈现出规律的波纹。即使黑电平完美对齐，这种由 k 本身数学属性决定的震荡也是不可消除的固有特征。



5-2-1-3 对DR测量的影响

prescaling最显著的影响是降低满阱容、增大噪声。对PTC模型而言他可以等效为一个在前方的乘子。

$$\begin{aligned}\sigma_{out}^2 &= (k_{pre} \cdot \sigma_{raw})^2 \\ &= k_{pre}^2 \left(K_{sys} \cdot \frac{\mu_{out}}{k_{pre}} + \sigma_{read}^2 \right) \\ &= \underbrace{(k_{pre} \cdot K_{sys})}_{\text{新}Kadc} \cdot \mu_{out} + \underbrace{(k_{pre} \cdot \sigma_{read})^2}_{\text{新}NoiseFloor}\end{aligned}$$

但因为这个震荡，在TgtSNR=1的测试中，若单纯只考虑对满阱容和读出噪声的干扰，假设完全理想的采样，prescaling只会影响约0.22档（虽然已经不少）；若考虑到这个震荡，实测的DR会比无prescaling情况下低0.32档。如果只对R和B通道采取这个操作，最后的平均DR会下降0.16档之多。图5-2-1-3-1中黑色拟合线为理想采样下的prescaling，图5-2-1-3-2中黑色拟合线为无prescaling情况下的结果。

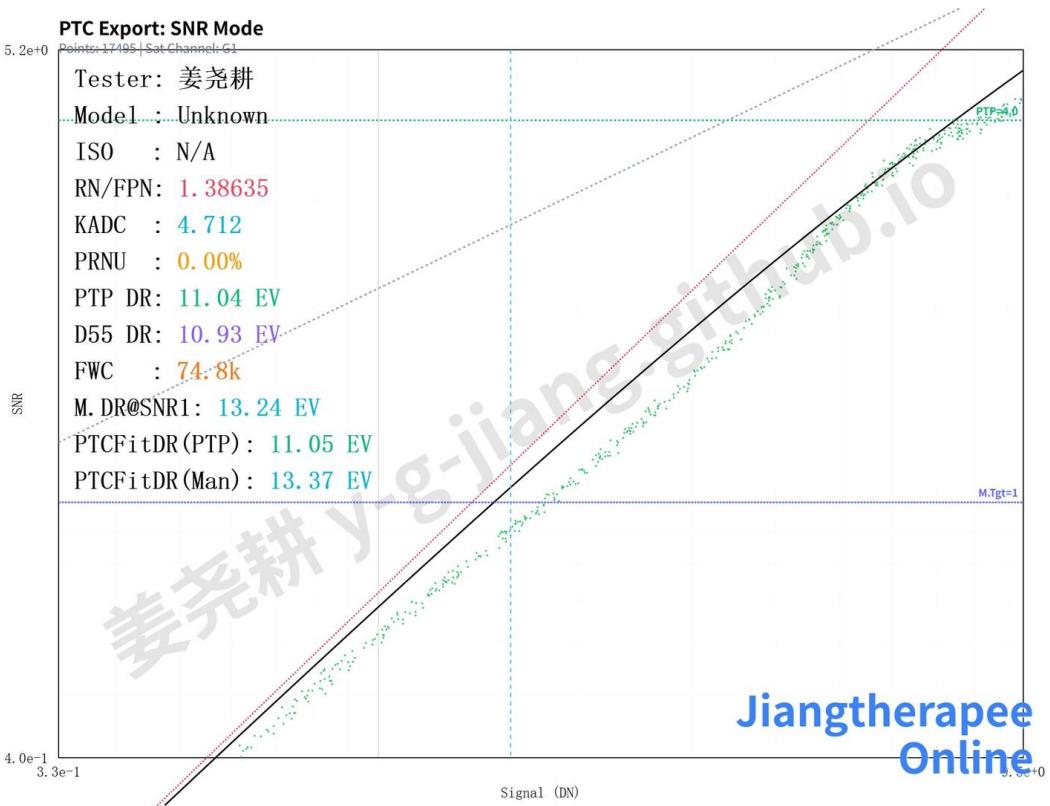


图5-2-1-3-1

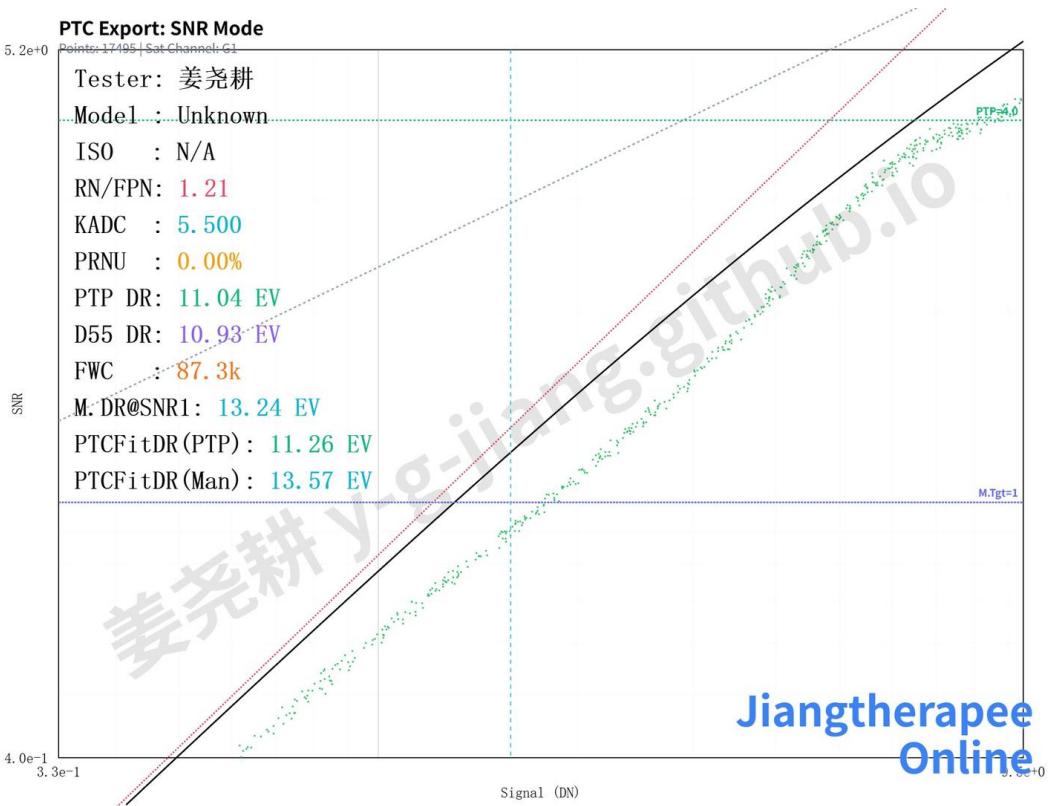


图5-2-1-3-2

5-2-2 黑电平漂移对非整数倍重量化的PTC的影响

5-2-2-1 数学模型模拟

考虑到即便我使用了手动定义的数字域的BL，解码结果仍然会有两族个截然不同的曲线，亦或者你的PTC结果与官方不符（具体为波动差一点相位）。那么可能出现的问题共被我归结为以下两点：

- 1, 解码时未恰当处理黑电平；
- 2, 乘减时取黑电平不同。

对于第一点，他可以被简单归类为PTC曲线的水平移动。在对ZF的测量中我经常看到此情况。这导致PTC像被数族曲线拼贴起来的一般，很好辨识。例如下图（图5-2-2-1-1）中暗部不仅有湖蓝色与绿色重叠，还有一簇湖蓝在绿色之左。这是因为拍摄该簇时的图片的黑电平均未被正确处理。

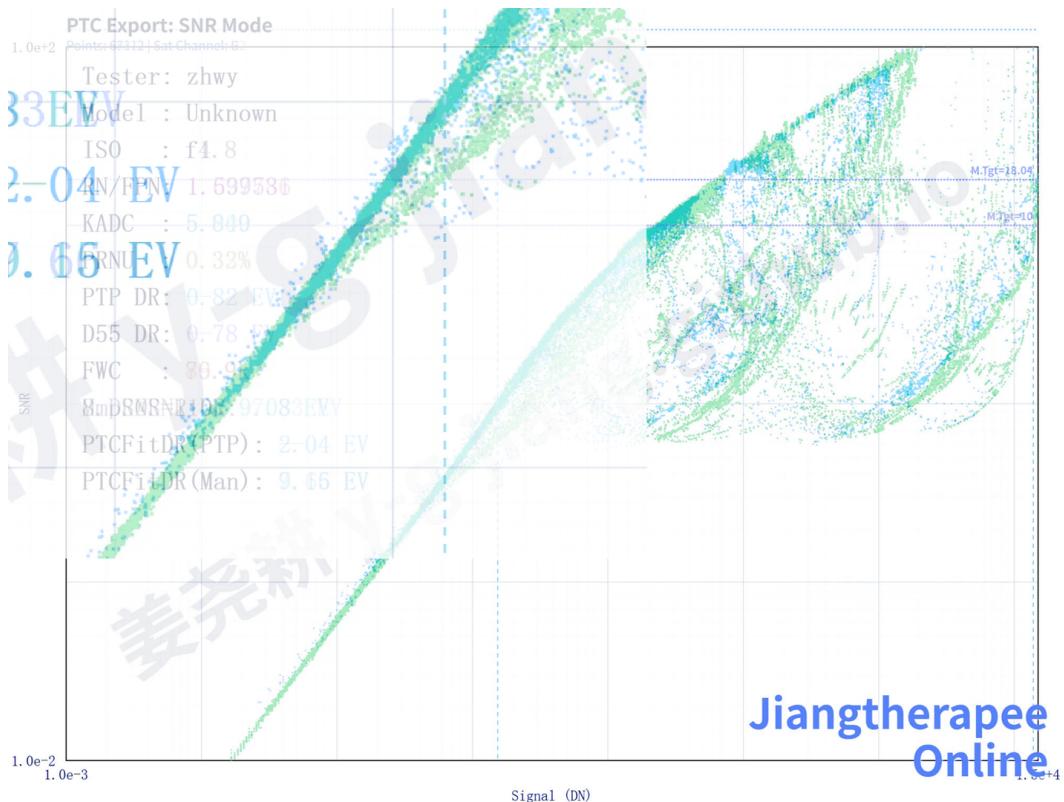


图5-2-2-1-1

第二点则会导致震荡的相位移动。下图（图5-2-2-1-2）中红线与绿线是我模拟的按（等效）511进行乘减与按512进行乘减的结果。尽管最后解码时我都按照统一的黑电平进行解码了，一个不统一的乘减仍然会导致波动的相位移动。移动的相位会带给DR测量极大的不确定性。

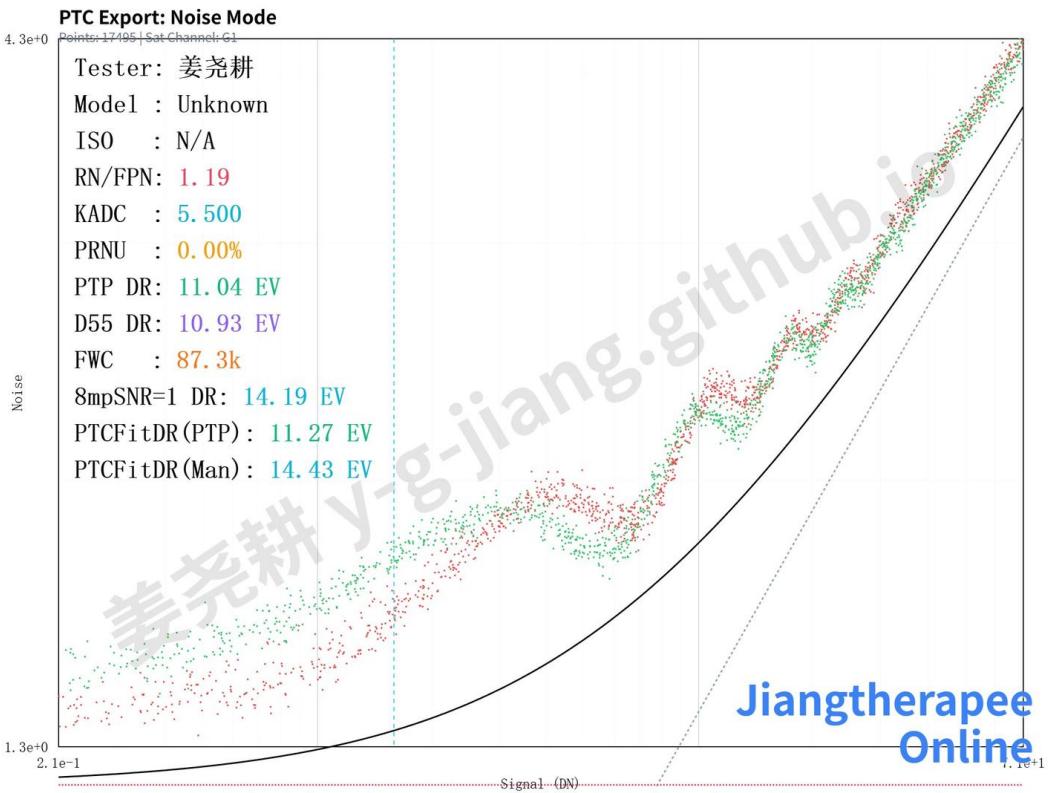


图5-2-2-1-2

注意上图黑线为无prescale下的理想最佳PTC，并非“prescale后的理想采样”。

5-2-2-2 对DR测量的影响

解码时取恰当黑电平是必须做的。你如果想测得准确的结果，起码应当统一拍摄PTC时尽可能少改变其他变量。使用JiangtherapeeOnline你们不难做到。

但是第二点切实会带给DR测量管线一些影响了。

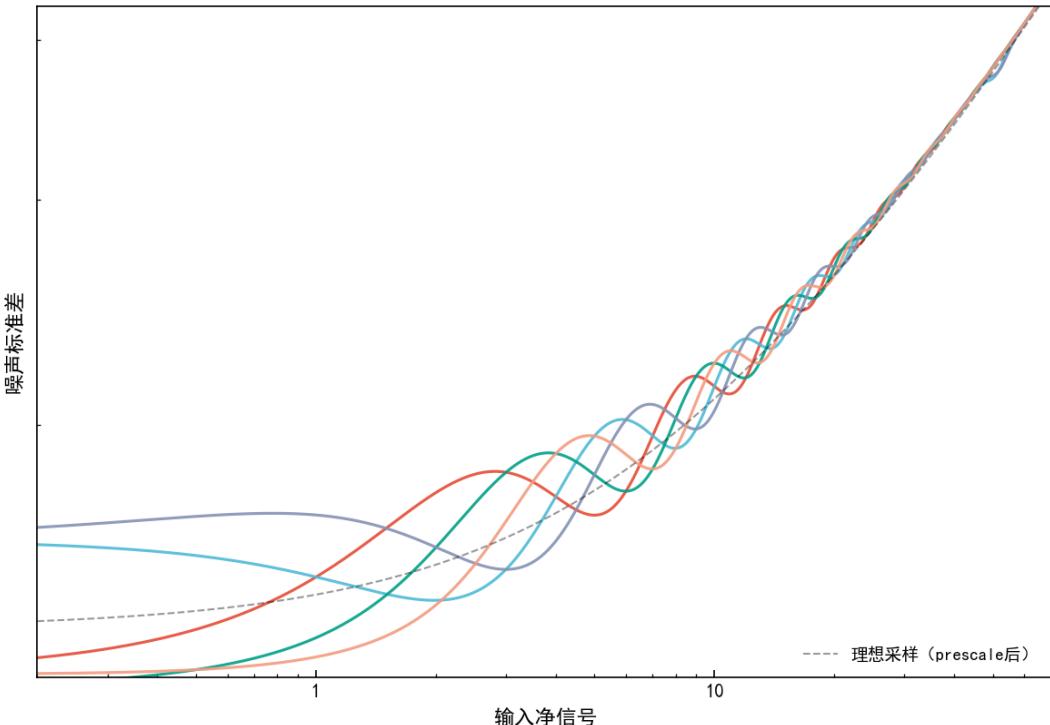


图5-2-2-2-1

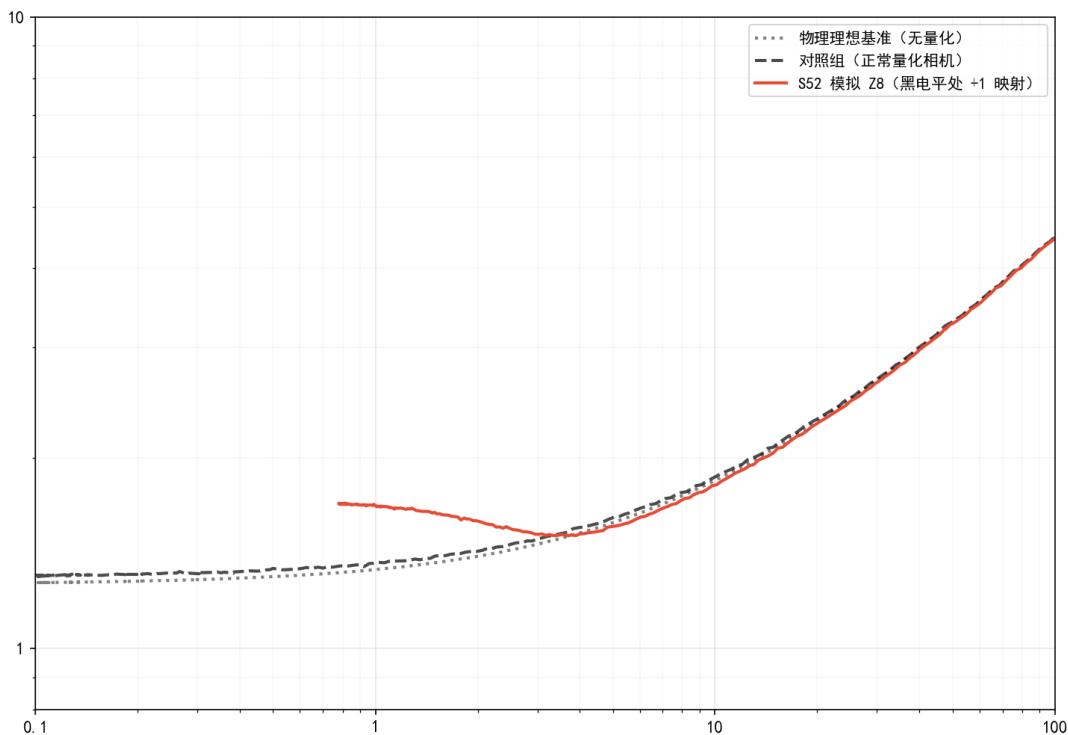
不难发现乘减时采取的黑电平可以遍历一切相位。在我以S52这个模拟例子中，乘减时的黑电平的更改，会导致如图5-2-1-3-1所示、大于0.1档DR(8mp, SNR=1)的差异。这是远低于JiangtherapeeOnline的测量流程精度的干扰。若对其他机型模拟分析，差异或会更加显著。

5b z8的特殊情况

在近期我又发现了一个奇怪的现象：z8的PTC存在反曲情况。也即PTC线先降低再升高。

经研究，我猜测其每个像素都会自动跳过1008附近的某个值。姑且不讨论他的原因，我来给出概括性建模。不妨假设是在数字化后进行的移位，那么有代表性的选项有两个：1，等效于将大于等于1008的码值+1；2，等效于将小于等于1008的码值全部-1。

作为更复杂的情况，先对+1讨论。不过起码对z8而言是减法，后面有证明。

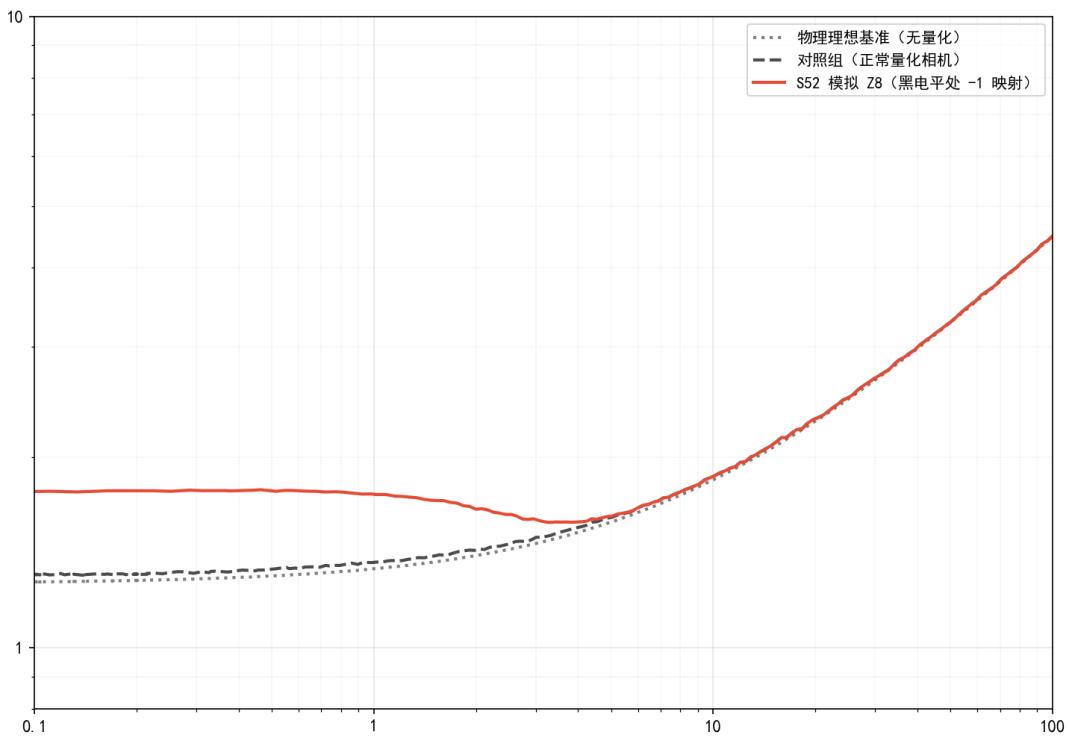


看上图不难看出现在有两个问题：1，edr虚低；2，在pdr阴影点附近区段，反而出现了snr虚高的情况。

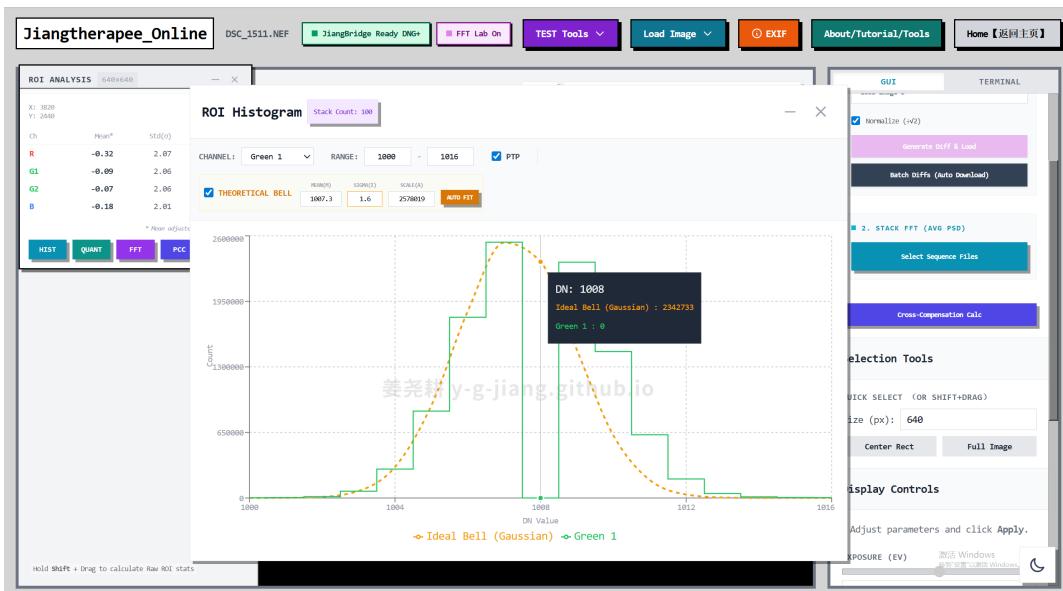
这是因为在极暗部受此映射影响，直方图从中间直接裂开，导致方差虚高；而对于那些在直方图主要区域不在奇点附近的暗部，不受这个“分裂”影响的地方，PTC相当于直接右移了1ADU，这使得同DN处的噪声要低于原先的值。PTP的TgtSNR线对应的阴影点处的高斯，对现有机型来说，均不会受此分裂影响。因此这个映射降低了edr，却提升了pdr。

最好的补偿办法是先拍摄暗场，将高于1008的码值左移来重新标定黑电平；再把这个黑电平+1，输入jiangtherapee。这样就可以获得极准确的PDR了。这个操作也等效于校正此映射后的暗部偏色。

但z8并没有出现前文中我提及过的黑电平不够高导致的暗部偏品——甚至是偏绿的。经过线性分析可知z8的行为是更加近似等效于将小于等于1008的码值减一的。那么不难推得此时情形为PTC后段重合，前段虚高。



这足以描述z8的行为了——在prescaling的基础上，z8有此种操作，导致了PDR不受干扰的情况下，EDR测量结果严重偏低。反推z8的运算用EDR，可以反演raw到单峰形式。为了绝对严谨，请确保拼合后的直方图构成钟形曲线。经我的验证，贴合极度良好。



图：极低码值平场，非黑场。

注意：当反推减法移位的动态时，黑电平应当保持不变来进行动态范围的计算。

6 附录

代码仓库：[y-g-jiang/IIARPTC](https://github.com/y-g-jiang/IIARPTC)

后续本文更新信息见该仓库README。

7 参考文献

无