A Probabilistic Model for Photon Pileup in Imaging Detectors

17th IACHEC Meeting
Osaka, Japan

Robert Zimmerman[1]
with David A. van Dyk[1] and Vinay Kashyap[2]

[1]Imperial College London
[2]Harvard & Smithsonian

May 13, 2025

Pileup

- **Pileup** occurs when multiple photons strike the same detector region during a single frame
- The detector cannot resolve them as separate events
- Their combined charge is interpreted as a *single* event:
  - Total energy is roughly the sum of the individual photon energies
  - Resulting charge cloud may be irregular → *bad grade*
- These effects distort the observed spectrum and reduce usable counts

Figure: Pileup distorts the observed image, producing a “hole” in bright regions where events are lost due to bad grades

Why Model Pileup Probabilistically?

- In high-energy regimes, photon coincidence distorts observed energy channels and grades
- Traditional corrections (Ballet, 1999, 2003; Davis, 2001) are deterministic and can fail when...
  - Pileup is frequent
  - The source spectrum is complex
  -
- We propose a **fully generative statistical model** for pileup, which enables principled inference

What Actually Happens During Pileup?

- Multiple photons in the same frame $\rightarrow$ Overlapping charge clouds $\rightarrow$ Sum of energies is recorded as a single event with pulse height approximately equal to the sum of energies
- Detector logic applies:
  - Charge spreading $\rightarrow$ Multiple pixel signal
  - The center of the pixel island is picked
  - Grading $\rightarrow$ Determines the measure of spatial distribution of charge
- Observed data: energy channel + event grade (and pixel island location and detector time)

Ingredients

- Within each time frame and detector region (i.e., pixel island) r, we observe...

  A channel $C_r \in \{1, \ldots, 1024\}$: an observed measure of the photon's (or piled photons') energy with measurement error quantified by a redistribution matrix function

  A grade $G_r \in \{\text{good}, \text{bad}\}$ based on the electron charge cloud pattern generated by the photon(s)

  - Events with pileup are usually assigned bad grades due to irregular charge cloud shapes
  - Events with bad grades are typically discarded in practice — but in our model, they are critical for inference!

- We do **not** observe...

  P_r, the number of photons *actually* emitted by the source (and then piled up)

  $E_{r,1}, \ldots, E_{r,P_r}$ the energies of those individual photons

  $H_{r,1}, \ldots, H_{r,P_r}$ the would-be grades of those individual photons

A Graphical Model

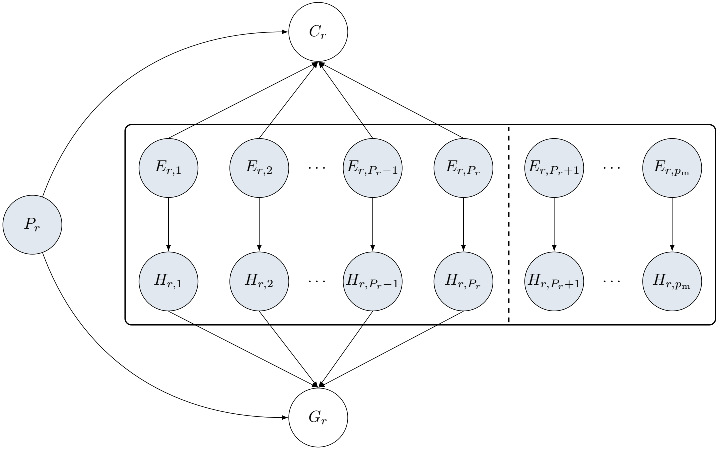


Figure: Graphical model representing the conditional independence structure of the observed and latent variables within a single region r; shaded nodes represent latent variables

Statistical Inference

- We are mainly interested in the vector of parameters $\boldsymbol{\theta}$ corresponding to the source spectrum
- We aim to perform frequentist statistical inference by maximizing the likelihood function

$$L(\boldsymbol{\theta}) = \prod_{r=1}^{R} \prod_{i=1}^{n_r} \mathbb{P}_{\boldsymbol{\theta}}(G_r = g_{r,i}, C_r = c_{r,i})$$

  where $(g_{r,1}, c_{r,1}), \ldots, (g_{r,n_r}, c_{r,n_r})$ are independent grade-channel observations in region r

- For this, we need a model that allows us to compute (or estimate) the joint distribution $\mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c)$

Key Modelling Idea

- Model the number of piled-up photons P_r in region r in one time frame as having a Poisson distribution
- Within a given region r and time frame, let (g, c) be the observed grade-channel pair
- We marginalize over all possible photon configurations and energies:

$$
\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c) \\
= &\sum_{p=1}^{\infty} \mathbb{P}_{\boldsymbol{\theta}}(G = g, C = c \mid P_r = p) \cdot \mathbb{P}_{\boldsymbol{\theta}}(P_r = p) \\
= &\sum_{p=1}^{\infty} \int \mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c \mid P_r = p, \boldsymbol{E}_{1:p} = \boldsymbol{e}_{1:p}) \cdot q_{p,\boldsymbol{\theta}}(\boldsymbol{e}_{1:p}) \, \mathrm{d}\boldsymbol{e}_{1:p} \cdot \mathbb{P}_{\boldsymbol{\theta}}(P_r = p)
\end{aligned}
$$

  - ...where $q_{p,\boldsymbol{\theta}}(\boldsymbol{e}_{1:p}) = \prod_{k=1}^{p} q_{\boldsymbol{\theta}}(e_k)$ is the joint density of the vector of p energies $\boldsymbol{E}_{1:p} = (E_1, \ldots, E_p)$

Complications....

- The distribution of C_r depends on $\boldsymbol{E}_{1:p}$ through the total energy $\sum_{k=1}^{p} E_k$, which follows a p-fold convolution of $q_{\boldsymbol{\theta}}$

- The distribution of the observed grade-channel pair depends on photon-level grades, which are unobserved (so we must marginalize over them too):

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c \mid P_r = p, \boldsymbol{E}_{1:p} = \boldsymbol{e}_{1:p}) \\
&= \mathbb{P}_{\boldsymbol{\theta}}(C_r = c \mid \boldsymbol{E}_{1:p} = \boldsymbol{e}_{1:p}) \cdot \sum_{\boldsymbol{h}_{1:p} \in \{\text{good},\text{bad}\}^p} \mathbb{P}_{\boldsymbol{\theta}}(G_r = g \mid \boldsymbol{H}_{1:p} = \boldsymbol{h}_{1:p}) \\
&\qquad \cdot \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{H}_{1:p} = \boldsymbol{h}_{1:p})
\end{aligned}$$

- The detector has a maximum threshold $\nu_{\max}$: if $\sum_{k=1}^{p} e_k > \nu_{\max}$, then the event is unrecorded (we must account for this too)

- We handle all of these!

Statistical Computation

- The likelihood function can be written down, but it is analytically intractable and impossible to compute exactly (mainly due to integrals over convolutions)
- Instead, we approximate it by Monte Carlo
- If $E_p^{(1)}, \ldots, E_p^{(m)} \overset{iid}{\sim} q_{\boldsymbol{\theta}}$ for each $k \geq 1$ and m is large, then

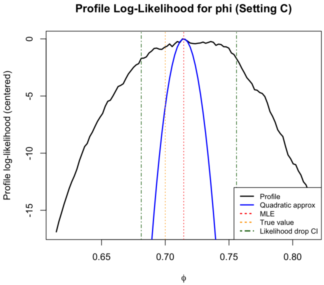
$$\int \mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c \mid P_r = p, \boldsymbol{E}_{1:p} = \boldsymbol{e}_{1:p}) \cdot q_{p,\boldsymbol{\theta}}(\boldsymbol{e}_{1:p}) \, \mathrm{d}\boldsymbol{e}_{1:p}$$
$$\approx \frac{1}{m} \sum_{j=1}^{m} \mathbb{P}_{\boldsymbol{\theta}}(G_r = g, C_r = c \mid P_r = p, \boldsymbol{E}^{(j)})$$

- So

$$L(\boldsymbol{\theta}) = \prod_{r=1}^{R} \prod_{i=1}^{n_r} \mathbb{P}_{\boldsymbol{\theta}}(G_r = g_{r,i}, C_r = c_{r,i})$$
$$\approx \prod_{i=1}^{n_r} \sum_{p=1}^{\infty} \left(\frac{1}{m} \sum_{j=1}^{m} \mathbb{P}_{\boldsymbol{\theta}}(G_r = g_{r,i}, C_r = c_{r,i} \mid P_r = p, \boldsymbol{E}^{(j)}) \right) \cdot \mathbb{P}_{\boldsymbol{\theta}}(P_r = p)$$

Uncertainty Quantification

- In many situations, one can obtain confidence intervals for $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$ by inspecting the Hessian of $\log(L(\boldsymbol{\theta}))$ at $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$
- However, our log-likelihood surfaces are locally "wobbly"
  - We suspect this is due to the discrete nature of observed grades/channels and (maybe) Monte Carlo noise



- Instead, we approximate the log-likelihood surface by dropping points around the MLE and fitting a quadratic surface using linear regression
  - This (usually) works well!

Simulation Study: Setup

- Write $\lambda(\boldsymbol{\theta})$ for the total expected photon count per unit frame time
- We simulate data under four scenarios to assess parameter recovery:

Setting A: Low pileup, no emission line
$\lambda(\boldsymbol{\theta}) \approx 2.1$, power law spectrum

Setting B: High pileup, no emission line
$\lambda(\boldsymbol{\theta}) \approx 4.5$, power law spectrum

Setting C: Moderate pileup, mixed spectrum
$\lambda(\boldsymbol{\theta}) \approx 3.3$, power law + emission line spectrum

Setting D: High pileup, mixed spectrum
$\lambda(\boldsymbol{\theta}) \approx 5.6$, power law + emission line spectrum

- These simulations use a fixed RMF matrix and PSF; we fit the model via maximum likelihood using our Monte Carlo likelihood approximation

Simulation Study: Parameter Recovery ($n = 500$)

- Each simulation involves estimating $\boldsymbol{\theta} = (\alpha, \theta)$ or $\boldsymbol{\theta} = (\alpha, \theta, \phi)$, with

  α: Grade migration parameter which controls probability that piled photons yield a good grade

  θ: Power law slope

  ϕ: Relative weight of the power law component vs. the emission line

Setting	α	$\widehat{\alpha} \pm$ SE	θ	$\widehat{\theta} \pm$ SE	ϕ	$\widehat{\phi} \pm$ SE
A	0.70	0.71 $\pm$ 0.048	0.70	0.67 $\pm$ 0.005	1.00	(fixed)
B	0.70	0.70 $\pm$ 0.025	1.50	1.63 $\pm$ 0.013	1.00	(fixed)
C	0.70	0.77 $\pm$ 0.030	0.70	0.68 $\pm$ 0.006	0.70	0.69 $\pm$ 0.004
D	0.70	0.68 $\pm$ 0.022	1.50	1.63 $\pm$ 0.017	0.30	0.28 $\pm$ 0.008

- Estimated parameters recover true values well across all conditions
  - Slight upward bias in $\widehat{\theta}$ under high pileup (Settings B and D), as expected

Future Directions

- Although we deal with frequentist inference, our work extends easily to the Bayesian regime
  - Use the same model and use our Monte Carlo based evaluation of the likelihood
- The modular nature of our model allows us to easily add a refined sub-model for the probability of grade migration
- Incorporation of additional instrumental effects
- Accounting for background contamination
- Etc.

Thank you!

References

Ballet, J. (1999). Pile-up on X-ray CCD instruments. *Astronomy and Astrophysics Supplement Series*, 135(2):371–381.

Ballet, J. (2003). Pile-up on X-ray CCD instruments. *Advances in Space Research*, 32(10):2077–2082.

Davis, J. E. (2001). Event Pileup in Charge-coupled Devices. *The Astrophysical Journal*, 562(1):575–582.