Selección de Eventos para una Búsqueda de Materia Oscura de Baja Masa en Dispersión con Electrones usando DAMIC-M

Data Selection for a Low-Mass Dark Matter Search
in Dark Matter–Electron Scattering with DAMIC-M

Trabajo de Fin de Máster

para acceder al

MÁSTER EN FÍSICA DE PARTÍCULAS Y DEL COSMOS

Autor: Elena Muñoz Rivas
Directora: Nuria Castelló-Mor
Codirectora: Rocío Vilar Cortabitarte

Julio-2025

A mis padres, mi hermana y a Alejandro, por estar siempre ahí durante tantas horas de pruebas, errores y líneas de código, apoyándome y animándome cuando más lo necesitaba.
A Nuria, por su dedicación incansable y por acompañarme en cada paso, resolviéndome todas las dudas que me surgían, que no eran pocas.
A Rocío, por su apoyo constante y su cercanía a lo largo de este proyecto.
Y a toda la gente del IFCA, por acogerme con tanta amabilidad y hacerme sentir parte del equipo, en especial a los IFCarcelados, por cada conversación, risa y momento compartido.

Resumen

Comprender la naturaleza de la materia oscura (DM) sigue siendo un desafío fundamental en la física moderna. El experimento DAMIC-M (DArk MAtter In CCDs at Modane) emplea dispositivos skipper Charge-Coupled Devices (CCDs) para buscar interacciones de DM con una sensibilidad sin precedentes. Uno de los retos es la caracterización de diversos fondos, como la corriente oscura y la radiactividad de los materiales, que pueden limitar la sensibilidad de los detectores. Este estudio se centra en el proceso de selección de datos para la dispersión DM–electrón: se definen y aplican siete máscaras jerárquicas para suprimir artefactos instrumentales y depósitos de alta carga, y se implementan técnicas de identificación de patrones para clasificar eventos de 1–3 e^- y compararlos con los fondos esperados. En consecuencia, el estudio proporciona una base validada para futuras investigaciones con exposiciones mayores en DAMIC-M.

Palabras clave: materia oscura ligera; DAMIC-M; Skipper-CCD; selección de datos; enmascarado; identificación de patrones; fondos instrumentales.

Abstract

Understanding the nature of dark matter (DM) remains a fundamental challenge in modern physics. The DAMIC-M experiment (DArk MAtter In CCDs at Modane) employs skipper Charge-Coupled Devices (CCDs) to search for DM interactions with unprecedented sensitivity. One of the challenges is the characterization of various backgrounds, such as dark current and material radioactivity, which can limit the sensitivity of the detectors. This study focuses on the data-selection process for DM–electron scattering: seven hierarchical masks are defined and applied to suppress instrumental artifacts and high-charge deposits, and pattern-identification techniques are implemented to classify 1–3 e^- events and compare them with the expected backgrounds. Consequently, the study provides a validated basis for future investigations with larger exposures in DAMIC-M.

Keywords: light dark matter; DAMIC-M; Skipper-CCD; data selection; masking; pattern identification; instrumental backgrounds.

Content

Chapter 1

Introduction

This master's thesis presents a study of data selection for low-mass DM–electron scattering with DAMIC-M. Chapter 1 provides context on the current evidence for the existence of DM, the main strategies for its terrestrial detection, and the role played by the DAMIC-M experiment in probing sub-GeV dark matter candidates. Chapter 2 introduces the expected dark matter–electron signal and the role of Skipper-CCD technology in achieving single-electron resolution; it then reviews the main radiogenic and cosmogenic backgrounds and the shielding and material-selection strategies used to mitigate them. Finally, it discusses the dark current and instrumental-noise sources and their suppression. Chapter 3 describes the data-selection masking techniques developed for this analysis. Chapter 4 details the pattern-identification methods used to reconstruct and classify low-energy events, the validation of their performance, and the results obtained from unblinded data. Chapter 5 applies the full analysis pipeline to blinded data collected in four data sets from October to December 2024. Finally, Chapter 6 summarizes the main findings of this study and outlines prospects for future developments and applications of the methodology.

1.1 Dark Matter and Direct Detection

DM is an invisible component of the Universe that has been inferred only through its gravitational effects. In the ΛCDM cosmological model, cold DM contributes $\Omega_c h^2 = 0.1200 \pm 0.0012$,(using $h \simeq 0.674$, this gives $\Omega_c \sim 26\%$ of the present day energy budget, whereas ordinary baryons account for just $\sim 5\%$ [1]. A wide range of astrophysical and cosmological observations —including the dynamics of galaxy clusters, flat galactic rotation curves, gravitational lensing and the temperature anisotropies of the cosmic-microwave background —requires a non-baryonic matter density consistent with the above value [2]. Because baryons alone cannot supply this density, DM must be a neutral, long-lived particle that couples only feebly to the Standard Model (SM). On galactic scales it is modelled as an extended halo; in the Milky Way, the standard-halo model adopts a local density of $\rho_0 \sim 0.3$ GeVcm^{-3} [3]. As a result, the Solar System is immersed in a constant stream of DM particles that could, in principle, be detected on Earth.

Proposed DM candidates span a wide range of masses. A viable DM particle must (i) gravitate, (ii) be stable over cosmic time, (iii) be non-relativistic during structure formation, and (iv) interact weakly enough that self-interactions satisfy $\sigma/m \lesssim 1$ cm^{2}g^{-1}, a limit inferred from the survival of elliptical halos and from merging clusters such as the Bullet Cluster[4], which disfavour stronger elastic dark-matter scattering. Mass hypotheses span more than forty orders of magnitude. In the GeV-TeV regime, *Weakly Interacting Massive Particles (WIMPs)* dominated past searches [5]; however, the null results from LUX-ZEPLIN experiment (LZ), XENONnT and PandaX have shifted interest toward lighter hidden-sector candidates -sub-GeV DM, axion (like particles and dark photons) which demand sub-keV thresholds. A hidden sector is a set of new fields that carry no Standard-Model gauge charges. They communicate with ordinary matter only through extremely weak mixings—for example kinetic mixing with the photon, suppressed Higgs exchange, or very heavy mediator exchange. Such feeble couplings explain why the particles have eluded past searches while still allowing them to be probed by ultra-low-threshold detectors.

Searches for DM are conducted through three types of experiments. High-energy colliders search for missing-momentum signatures; indirect searches try to identify SM products of DM annihilation or decay in astrophysical sources; direct-detection experiments, the focus of this thesis, aim to observe the recoil of target nuclei or electrons after an elastic scattering with dark matter inside a well-shielded detector.

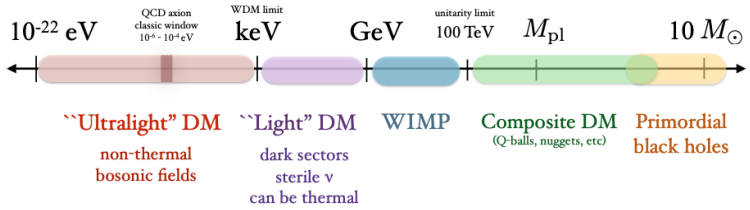


Figure 1.1: Landscape of dark matter candidates as a function of mass [6].

For DM masses $m_\chi \gtrsim 10$ GeV/c^2, experiments typically search for nuclear recoils with energies of a few keV. At lower masses, scattering off electrons becomes more favorable: MeV-scale DM can ionize one or two electron–hole pairs in a semiconductor target. Detecting such small signals requires detectors with readout noise well below the single-electron level, enabling sub-electron thresholds and resolution of individual charge carriers. The expected event rate falls steeply with recoil energy and can exhibit daily or annual modulation due to the Earth's motion through the halo [7, 8]. While shielding and material selection are designed to suppress backgrounds to the lowest possible levels, residual backgrounds may persist. In such cases, the distinctive time-dependent signature of a modulated DM signal, absent in most known backgrounds, can provide a valuable discriminant to improve sensitivity. Technologies under development include cryogenic bolometers, noble-liquid TPCs, superheated liquids, and semiconductors. Among these, silicon CCDs and high-voltage Ge arrays stand out for their sub-electron charge resolution.

SiliconCCDs offer a unique combination of large target mass, high spatial granularity and sub-electron noise. This makes them ideal to probe both WIMPs, induced nuclear recoils and single-electron events from MeV-scale DM. The DAMIC program at Sudbury Neutrino Observatory Laboratory (SNOLAB) demonstrated background rates below one event kg^{-1}day^{-1} in the 1-e^- bin; its successor, **DAMIC-M**, now under construction at LSM, will deploy up to 104 Skipper-CCDs in its first stage, with ultra-low thresholds and radio pure materials, pushing DM–electron searches into previously unexplored parameter space [9].

1.2 The DAMIC-M Experiment

DAMIC-M[1] is the next generation successor to DAMIC at SNOLAB. It relies on silicon Skipper-CCDs that achieve sub-electron charge resolution. The experiment is being installed at the Laboratoire Souterrain de Modane (LSM) under $\sim$ 1.7 km of rock, which attenuates the cosmic-muon flux by six orders of magnitude [10]. The baseline design foresees $\sim$50 modules, each holding four Skipper-CCDs, for a total of $\sim$200 sensors or $\sim$ 1 kg of active silicon, two orders of magnitude more target mass than DAMIC at SNOLAB [9].

The DAMIC program at SNOLAB demonstrated background rates below one event kg^{-1}day^{-1} in the 1-e^- bin; its successor, **DAMIC-M**, now under construction at LSM, will deploy up to 104 skipper-CCDs in its first stage, using ultra-low thresholds and radiopure materials to push DM–electron searches into previously unexplored parameter space [9].

DAMIC-M aims to probe dark matter candidates in the MeV-GeV mass range, with a focus on DM-electron scattering. With sub-electron thresholds and background levels below 1differential rate unit (dru)[2] the experiment can test scenarios that standard WIMPs detectors miss, including

- **Freeze-out:** DM particles were once in full thermal contact with ordinary matter. As the Universe expanded and cooled, the interaction rate became too slow to keep up with cosmic expansion,

[1] **D**ark **M**atter **I**n **C**CDs at **M**odane
[2] 1 event kg^{-1} day^{-1} in the 1-e^- bin.

"freezing" the DM abundance at its present value.

- **Freeze-in:** DM couples so weakly to the SM that it never reached equilibrium. Instead, a small population was produced from rare SM interactions early in the Universe; once temperatures dropped, production ceased, leaving today's very low interaction cross-section[11].

Both mechanisms lead to interaction strengths far below current nuclear-recoil limits, yet still accessible to single-electron detectors like DAMIC-M.

1.3 Charge-Coupled Devices (CCDs)

The DAMIC-M detector is a vertical copper tower of modules; each module hosts four Skipper-CCDs mounted on a high-purity silicon frame that provides excellent thermal contact and minimal radio-impurities.

Silicon CCDs are two dimensional pixelated ionization detectors whose low noise, high spatial granularity and gram-scale target mass make them ideal for rare-event searches. Invented at Bell Labs in 1969 as a solid-state analogue of magnetic-bubble memory [12], CCDs were soon adopted for imaging [13]. High-resistivity devices developed for astronomy now provide hundreds of microns of fully-depleted silicon, enabling experiments such as DAMIC and DAMIC-M.

1.3.1 Pixel Architecture and Image Layout

Each DAMIC-M sensor consists of a 2D array of 6144×1536 square pixels, each measuring $15 \times 15\,\mu\mathrm{m}^2$, resulting in an active area of approximately $9.2 \times 2.3\,\mathrm{cm}^2$. With a thickness of $675\,\mu$m, the fully depleted volume corresponds to ~ 3.3 g of silicon per CCD [14].

Pixels are formed as coupled Metal–Oxide–Semiconductor (MOS) capacitors on a high-resistivity n-type substrate. A shallow p channel lies beneath the gate oxide; three polysilicon gates per pixel provide the 3-phase clocking scheme (see Figure 1.2), with lateral confinement provided by channel stops between columns. A backside bias $V_{\mathrm{sub}} \gtrsim 40\,\mathrm{V}$ fully depletes the bulk, drifting photo-generated charge to the buried channel where holes are stored until readout.

Vertical transfer (parallel clocks) moves charge row-by-row into a horizontal serial register (effectively the last row of the matrix); faster serial clocks then shift the charge horizontally through the serial register to the output amplifier. In practice the 3-phase clock cycles the three gate potentials (P_1, P_2, P_3) through a six-step sequence, shifting a charge packet to the next pixel with each complete cycle; a schematic of the vertical and horizontal transfers is shown in Figure 1.2. Although each CCD includes four amplifiers (one per corner), DAMIC-M normally reads through a single corner to minimize electronic complexity (such as cross-talk as we will see later on).

Every frame —the two-dimensional image read out after one exposure— contains two reference regions that receive no exposure: a prescan before the first physical column and an overscan after the last. Their pixels track the electronic pedestal and noise on a row-by-row basis, allowing a pixel-level correction of baseline drifts. In the standard readout mode, correlated double sampling is employed to subtract the signal before and after charge transfer, suppressing reset noise and high-frequency fluctuations. In skipper mode, multiple non-destructive measurements of the same pixel further reduce statistical noise, enabling sub-electron resolution.

1.3.2 Charge Generation in Silicon

When a particle deposits energy in the silicon bulk of a CCD, it can ionize the medium, producing free charge carriers. If the deposited energy exceeds the band gap of silicon, $E_{\mathrm{gap}} = 1.12$ eV, an electron can

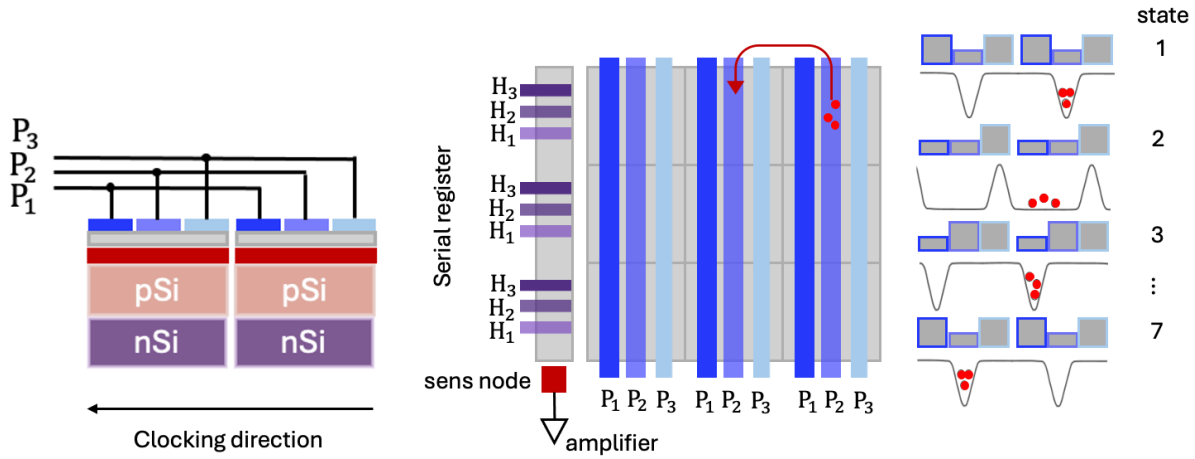


Figure 1.2: Schematic representation of charge transfer in a Skipper-CCD using a 3-phase clocking scheme. Left: Cross-section of the pixel structure showing the p-n junction and the three polysilicon gates (P_1, P_2, P_3) connected to the global clocks. Center: 2D layout of a small pixel matrix with vertical and horizontal registers. Charge is moved vertically by cycling the clocks, and finally read out through the serial register and sense node. Right: Evolution of the potential wells for different clocking states, and how the charge packet (red dots) is transferred from one pixel to the next.

be promoted to the conduction band, leaving behind a mobile hole. The resulting electron–hole pair then drifts under the influence of the internal electric field created by the back-side bias [15].

At energies above a few keV the charge yield in silicon is well described by the asymptotic relation

$$\langle N_{\mathrm{eh}} \rangle = \frac{E}{\varepsilon_{\mathrm{eh}}}, \qquad \sigma_N^2 = F \, \langle N_{\mathrm{eh}} \rangle, \tag{1.1}$$

with $\varepsilon_{\mathrm{eh}} = (3.752 \pm 0.002)\,\mathrm{eV}$ and $F = (0.119 \pm 0.003)$ at $T \simeq 123$ K, as measured with Skipper-CCDs [16]. Below ~ 50 eV the linear approximation fails because only a few carriers participate in each impact-ionisation cascade. In the interval $1.12\ \mathrm{eV} \leq E \leq 50\ \mathrm{eV}$ we therefore adopt the energy-dependent probability distributions $P(n, E)$ of Ref. [17], which give the likelihood of producing exactly n electron–hole pairs. The model reproduces the single-pair limit just above the band gap and converges smoothly to the asymptotic relation as additional channels open at higher energies.

1.3.3 Charge Transport and Diffusion

After a particle interaction generates electron-hole pairs in the fully depleted bulk of the CCD, the electric field created by the back-side bias drives the charges toward opposite directions. In the case of DAMIC-M, holes drift toward the front side of the detector, where they are stored in the potential wells defined by the gate electrodes. These holes constitute the signal charge that is eventually read out.

As holes drift along the vertical (z) direction, they undergo lateral thermal diffusion in the x-y plane. The magnitude of this diffusion depends on the transit time and on the electric field profile in the bulk. For point-like deposits the transverse charge distribution is approximately Gaussian with an root-mean-square (RMS) spread $\sigma(z)$ that depends on depth z [18]:

$$\sigma(z) = \sqrt{-A \ln|1 - Bz|}, \tag{1.2}$$

$$A = \frac{\varepsilon k_B T}{e\rho_n}, \qquad B = \frac{\varepsilon V_{\text{sub}}}{\rho_n z_D (z_D + 2)}, \tag{1.3}$$

where ε is the permittivity of silicon, ρ_n the donor density and z_D the detector thickness.

Measuring σ thus provides the interaction depth, supplying the third spatial coordinate and aiding surface-background rejection. Figure 1.3 left panel, illustrates the effect of diffusion on two point-like energy deposits at different depths. Events occurring deeper in the silicon exhibit wider charge distributions due to the longer drift time.

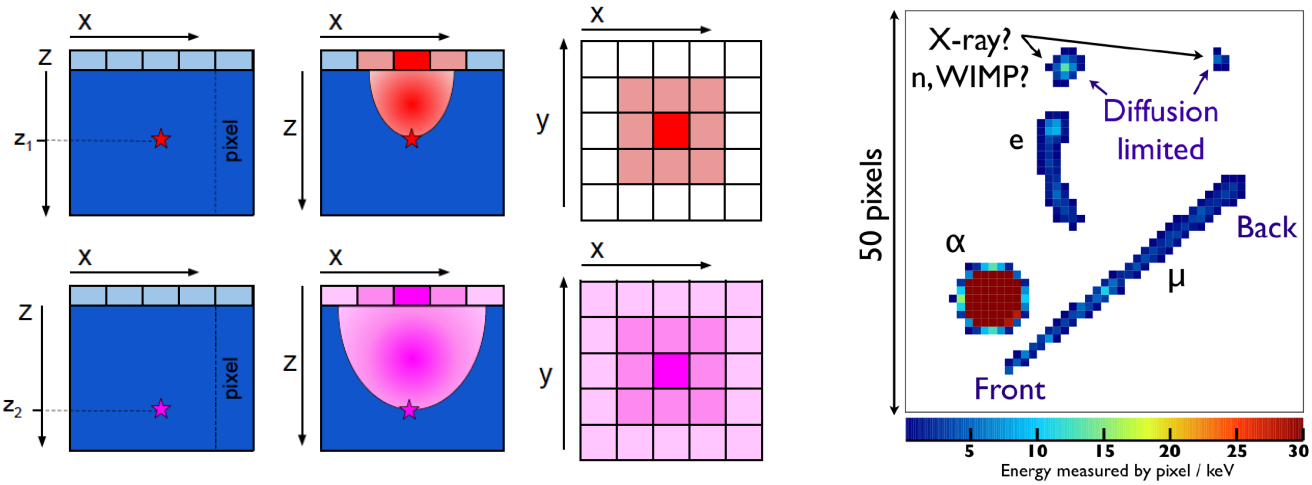


Figure 1.3: Left: schematic representation of charge diffusion in the CCD. Ionizing events at increasing depths ($z_1 < z_2$) result in broader lateral charge distributions [19]. Right: representative unbinned clusters showing the characteristic shapes of particle tracks from electrons, muons, alpha particles, and diffusion-limited events.

1.3.4 Skipper Readout: Multiple Non-Destructive Measurements

After each column has been clocked into the the serial register, the charge packets are clocked horizontally through three adjacent gates shown in Figure 1.4 (left panel). The Summing Well (SW) acts a temporary potential pocket that collects the packet; the the Output Gate (OG) is a clocked barrier that can open or close the passage, letting the packet move back and forth; and the Sense Node (SN)) is a tiny capacitor where the packet is finally dumped and converted into a voltage step $\Delta V = Q/C_S$. That step is digitized by a 16-bit Analog-to-Digital Converter (ADC), so the pixel values stored on disk are in Analog-to-Digital Unit (ADU).

In a skipper device the OG is moved back and forth repeatedly, letting the same packet be sampled N_{skip} times without destruction. Because the samples are uncorrelated, the readout noise fall as $\sigma_e(N_{\text{skip}}) = \sigma_1/\sqrt{N_{\text{skip}}}$. For the device of Figure 1.4 right panel, the single sample noise is $\sigma_1 \simeq 3.2\,e^-$, so the noise falls to $\sim 1.0\,e^-$ after $N_{\text{skip}} \approx 12$, to $\sim 0.4\,e^-$ after $N_{\text{skip}} \approx 64$, and reaches genuine single-electron resolution, $\sigma_e \lesssim 0.1\,e^-$, only after $\sim 10^3$ skipper samples. That precision multiplies the raw readout time by a comparable factor; DAMIC-M mitigates the penalty by operating in continuous mode, exposing the CCD while the previous image is still being read.

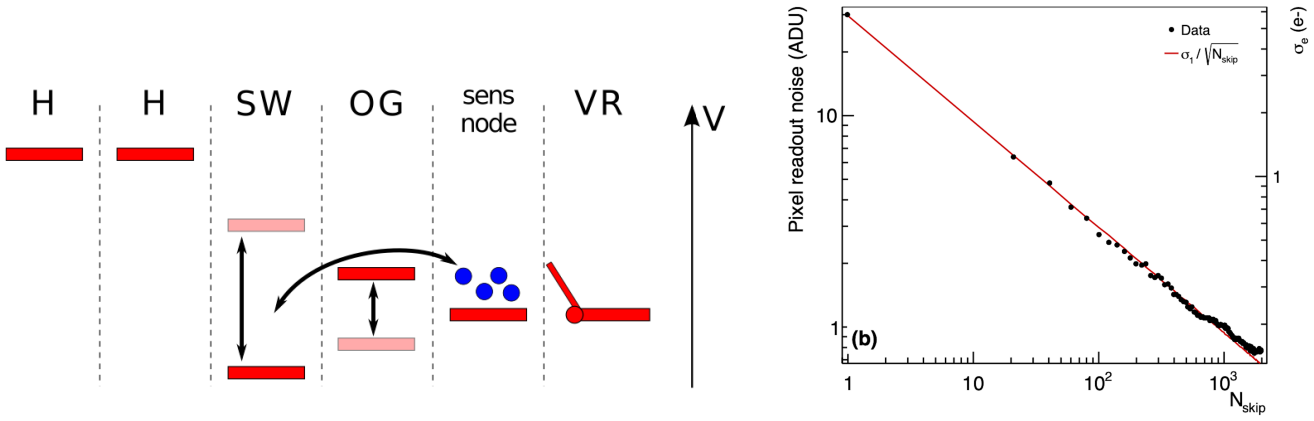


Figure 1.4: Left: schematic of the Skipper-CCD output stage (SW, OG, SN). Right: read-out noise versus number of non-destructive samples N_{skip} [20]. The left axis is in ADU, the right one in electrons; for the particular device shown one electron corresponds to $\simeq 5$ ADU, as determined from the 0- and 1-e^- peak separation.

1.4 Low-Background Chamber (LBC)

Before full-scale deployment, DAMIC-M technology was qualified in the LBC, a prototype cryostat equipped with two skipper-CCD modules. The LBC reproduces the ultra-high-vacuum environment ($P \simeq 10^{-7}$ mbar) and the same generation of front-end electronics foreseen for the final detector, though it is not surrounded by the full external shielding. Each module (Figure 1.5 left panel) holds four 675 μm thick skipper-CCDs wire-bonded to a monolithic silicon pitch adapter that serves both as low-radioactivity mechanical support and signal fan-out. The two modules are clamped to a copper cold finger inside the stainless-steel vacuum vessel shown in Figure 1.5 (right panel).

Installed at LSM in 2023, the LBC operates at $T \simeq 120$ K. With the new DAMIC-M electronics the read-out noise measured by the two CCDs is $\sigma_1 \lesssim 3\, e^-_{RMS}$ [21]. The data analyzed in this thesis were taken with $N_{\text{skip}} = 500$ non-destructive samples per pixel, yielding $\sigma_e \approx 0.10\, e^-_{RMS}$.

An exposure of 115 g·days has already produced the most stringent limits on DM-electron scattering above a few MeV/c^2, validating both the shielding concept and the Skipper read-out. Ongoing upgrades —replacing Oxygen-Free High-Conductivity copper (OFHC) with electro-formed copper near the CCDs and improving radon suppression— are expected to push the background below the design goal 1 dru, unlocking the full sensitivity of DAMIC-M.

1.5 Goals and Data-Selection Strategy

Dark matter in the sub-GeV mass range remains a compelling target in astroparticle physics. At these low masses, scattering off bound electrons in semiconductors offers the best sensitivity, as nuclear recoils fall below threshold [8]. Skipper-CCDs, such as those of DAMIC-M, combine sub-electron noise with gram-scale mass, enabling the identification of individual electron-hole pairs. Initial runs at LSM have achieved background levels below one event kg^{-1} day^{-1} in the 1e$^-$ bin [19].

To exploit this sensitivity a robust data-selection pipeline is essential. Its purpose is to retain genuine low-charge ionizations while removing instrumental artifacts -hot columns, crosstalk, charge-transfer trails -and residuals from energetic interactions that could mimic a signal.

This study develops and validates a full data-selection pipeline tailored to the DAMIC-M detector response. Starting from raw images, the procedure includes the pre-processing of them (pedestal subtraction, gain calibration); the identification and masking of high-energy deposits ($\gtrsim 10\, e^-$) using

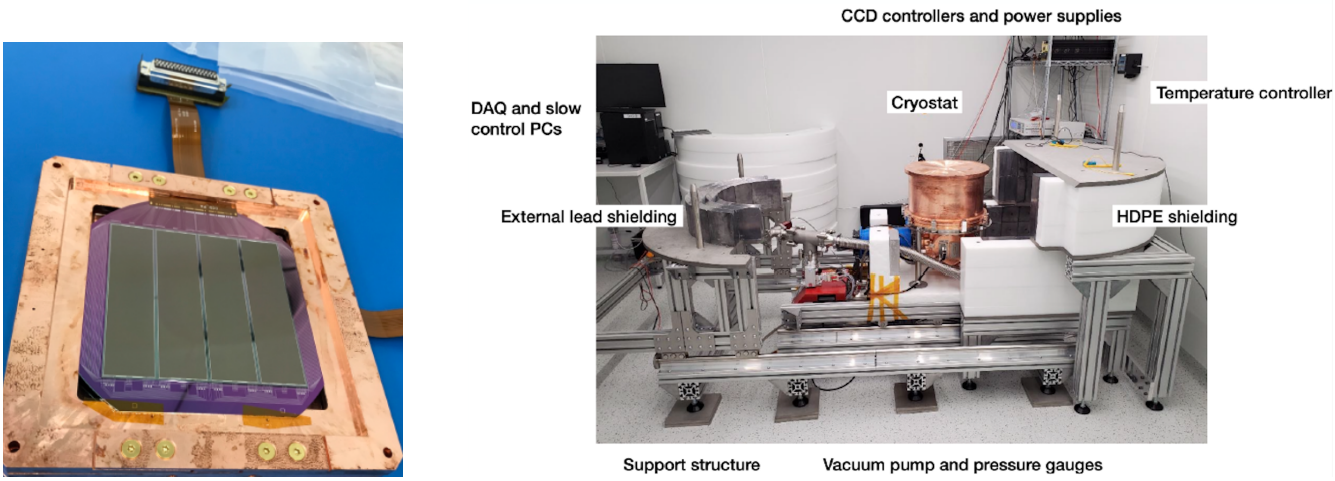


Figure 1.5: Low-Background Chamber and module hardware used for the DAMIC-M prototype studies. Left: DAMIC-M module hosting four Skipper-CCDs mounted on a high-purity silicon frame [22]. Right: The LBC at LSM. The vacuum vessel (centre) encloses the CCD stack; thin aluminum lids shield infrared light, while external lead and polyethylene (not shown) reduce γ and neutron backgrounds during commissioning runs [23].

WADERS algorithms and masking of CCD artifacts, hot-column and over-density algorithms, crosstalk veto, CTI/halo expansion, multiplicity and isolated-column filters. Advanced methods are introduced for the classification of 1-3 e^- pixels and statistical consistency checks with a Poisson Dark Current (DC) model.

The strategy is validated with dedicated Monte Carlo simulations that reproduce DAMIC-M operating conditions, as well as with real *unblinded* data from the LBC, achieving high efficiency for single-electron events with well-controlled false-rejection rates. All procedures follow the methodology established in the recent DAMIC-M analysis that delivered the first constraints on several hidden-sector benchmark models, based on an exposure of 141.5 g·d [24]. The data-selection and pattern identification presented in this thesis were developed within the framework of that analysis and are fully aligned with the strategy that led to unprecedented results in the search for DM–electron scattering in the MeV mass range. This work therefore provides a validated basis for future studies with larger exposures in DAMIC-M.

Chapter 2

Backgrounds for Dark Matter Searches

The region of interest (ROI) for DM–electron scattering corresponds to single-pixel events depositing between 1 and 10 electrons of total charge. In this ultra-low-charge regime, the only irreducible background is expected to be the dark current, modeled as a Poissonian distribution[1]. Other physical backgrounds (such as radiogenic γ-rays, cosmogenic neutrons, and muons) could in principle yield few-electron clusters, but the combination of layers of copper, lead, and polyethylene shielding, together with the deep-underground location of the LBC, attenuate these contributions by several orders of magnitude, preventing an appreciable number of such particles from reaching the active detector volume [24].

2.1 Expected DM-e$^-$ Signal

In the sub-GeV mass regime, DM particles are expected to interact predominantly with bound electrons in the detector material, producing ionization signals of only a few electron-hole pairs [25]. The DAMIC-M Skipper-CCDs, with their ability to resolve individual electrons, are uniquely suited to detect such small signals.

The expected DM-electron scattering rate depends on several physical inputs: the local dark matter density; the DM velocity distribution in the Galactic halo, usually modeled with a Maxwell-Boltzmann form truncated at the Galactic escape speed [26]; and the matrix element describing the interaction, which depends on the nature and mass of the mediator.

For a given DM mass m_χ, the differential scattering rate with respect to the deposited ionization energy can be written as

$$\frac{dR}{d\ln E_e} \propto \bar{\sigma}_e \frac{\rho_\chi}{m_\chi} \int dq \, |f_{\rm ion}(k', q)|^2 \, |F_{\rm DM}(q)|^2 \, \eta(v_{\rm min}), \tag{2.1}$$

where $\bar{\sigma}_e$ is the reference DM-electron scattering cross-section, $f_{\rm ion}(k', q)$ is the ionization form factor describing the initial and final electronic states in the silicon target, $F_{\rm DM}(q)$ is the DM form factor (unity for contact interactions, $1/q^2$ for light mediators), and $\eta(v_{\rm min})$ encodes the DM velocity distribution above the minimum required velocity $v_{\rm min}$ for a given energy transfer.

The form factor $f_{\rm ion}$ is computed using numerical wave functions for the valence and conduction-band states in silicon. Its structure leads to a sharply falling recoil spectrum, peaking at low energies ($\lesssim 1\,$eV). This makes single-electron resolution a critical requirement for probing light dark matter candidates.

In practice, once the theoretical rate is obtained, it is folded with a detector-response simulation that accounts for charge creation, diffusion in the silicon bulk, and read-out noise. The resulting spectra provide the expected event yields in each electron-multiplicity bin, which are later compared with the data.

Figure 2.1 (left panel) shows the predicted differential rate $R(Q_{\rm pix})$ for two representative DM–electron scattering models, with either a massive or a massless mediator. In both cases, the event rate drops steeply with the number of electrons per pixel, highlighting the importance of accurate background modeling in the 1–3 e^- region.

[1] For our data-taking conditions the measured mean dark current is $\lambda \simeq 2.2 \times 10^{-4}$ events pixel^{-1} image^{-1}.

2.2 Radiogenic and Cosmogenic Backgrounds

In addition to thermal dark current —which follows Poisson statistics— the DAMIC-M CCDs are exposed to ionizing backgrounds from natural radioactivity and from cosmic-ray induced interactions. These processes can generate low-energy clusters that imitate dark-matter interactions; many of them exhibit spatial or temporal correlations, or appear preferentially near detector surfaces. Careful characterization and mitigation are therefore essential to reach single-electron sensitivity.

2.2.1 Bulk and Surface Radioactivity

Radioactive impurities in detector materials can deposit energy in the CCD bulk or at its surfaces. Two long-lived isotopes are of particular concern:

- **^{32}Si.** Produced cosmogenically, ^{32}Si is incorporated during crystal growth and undergoes a two-step β-decay via ^{32}P. Both electrons fall in the 1-3 e^- region.
- **^{210}Pb.** Radon progeny that arrive onto CCD or copper surfaces decay through ^{210}Bi and ^{206}Pb. The ^{210}Bi β (Q = 63.5 keV) and the ^{206}Pb recoil can mimic low-energy events when emitted within a few μm of the active silicon.

Mitigation relies on nitrogen-purged handling, chemical polishing and packaging in radon-free cleanrooms (details in Sect. 2.2.5) [27]. With these controls, radio-induced single-electron candidates are reduced to $\lesssim 0.2$ dru, consistent with DAMIC-M's design goal[19].

2.2.2 Muon-Induced Backgrounds

Even at the depth of the LSM, cosmic-ray muons are not completely eliminated. The residual flux is $\Phi_\mu \simeq 5 \times 10^{-9}$ cm^{-2} s^{-1}, with a mean energy of ~ 250 GeV. Although rare, these muons generate two classes of backgrounds: (i) electromagnetic / hadronic showers that release fast neutrons and δ-rays in the shielding, and (ii) cosmogenic activation of copper and silicon that produces long-lived isotopes such as ^{60}Co and ^{3}H.

The DAMIC-M CCDs offer robust identification of muon tracks thanks to their excellent spatial resolution. Muons crossing the active area leave straight, high-energy tracks that are easily distinguished from low-energy clusters. Events correlated in time with such tracks, including delayed neutron or gamma interactions, can be rejected offline.

After all selection cuts, simulations that use the measured LSM spectrum predict a muon-related rate <0.1 dru in the 1-electron bin-sub-dominant with respect to radiogenic backgrounds and well within the DAMIC-M budget [19].

2.2.3 Neutrons from Radiogenic and Cosmogenic Sources

Neutrons are a dangerous background because nuclear recoils are, in principle, indistinguishable from DM scatters. They arise from two mechanisms: (i) *radiogenic* production within detector materials, and (ii) *cosmogenic* production by the residual muon flux. Radiogenic neutrons are primarily emitted through (α, n) reactions and spontaneous fission in the decay chains of ^{238}U and ^{232}Th present in construction materials. Their spectrum peaks below 10 MeV and the yield depends strongly on the material composition, particularly the abundance of light nuclei that enhance (α, n) cross-sections. Cosmogenic neutrons originate when high-energy muons interact with the surrounding rock or the lead shield; their flux at LSM is $\sim 6 \times 10^{-10}$ cm^{-2}s^{-1} above 1 MeV.

A 40 cm polyethylene moderator surrounding the cryostat thermalizes and captures most neutrons, while the inner 22 cm of lead attenuates accompanying γ rays. Residual events are further suppressed

with topology-based cuts (multi-pixel and time-correlated signatures; see Chapter 3). After all hardware and analysis measures, simulations predict a neutron contribution well below 0.05 dru, comfortably within the DAMIC-M background budget.

2.2.4 Electrons and Gamma Rays

Electrons and γ rays dominate the low-energy background because they are abundant and can deposit only a few electrons in the CCD. They originate from the decay chains of ^{238}U, ^{232}Th and ^{40}K in surrounding materials and from cosmogenic isotopes such as ^{60}Co.

Inside silicon, γ rays interact mainly through Compton scattering and photoelectric absorption, producing secondary electrons that form single- or multi-pixel clusters. Low-energy Compton electrons are particularly problematic: their 1–3 e^- charge is indistinguishable from a potential DM signal unless additional event-shape information is used.

External γ radiation is attenuated by the lead shield described in Sect. 2.2.5. Residual β/γ events are rejected in analysis: extended or irregular clusters (total charge $\gtrsim 10\,e^-$) are masked, and shallow interactions are removed with the depth-reconstruction cut detailed in Chapter 3. After these cuts the projected β/γ rate in the 1-electron bin is $\lesssim 0.3$ dru.

2.2.5 Low-Background Materials and Shielding Strategy

A graded shield and strict material screening reduce external and intrinsic backgrounds to the sub-dru level. The cryostat is enclosed by (from inside out) 2 cm of ancient lead, 20 cm of low-activity lead and 40 cm of polyethylene, as introduced in Sects. 2.2 and 2.2.3. Mechanical and thermal parts closest to the CCDs are machined from underground electro-formed Cu (U/Th < 0.1 ppt); outer structures use OFHC Cu. All components are assayed with high-purity germanium detector (HPGe) and inductively coupled plasma mass spectrometry (ICP-MS), and assembled in radon-free clean rooms. This combination of passive shielding and material radio purity supports the background estimates presented throughout this chapter.

Despite extensive shielding, during the data-taking period used in this work, the LBC shielding was left partially open (by ~ 5 cm), allowing a non-negligible contribution from ambient radiation. Simulations based on the detailed LBC geometry predict a nearly flat background rate in the ROI of approximately 10 dru, mostly due to ambient gamma rays and muon-induced neutrons.

2.3 Dark Current Background and Instrumental Noise

After the suppression of external backgrounds, the remaining events in the ROI are dominated by the dark current, a thermal background following Poisson statistics. In addition, a variety of instrumental artifacts can produce spurious low-energy signals.

Spurious low-energy pixels can arise from imperfections in the Skipper-CCD material or in the read-out chain and may mimic true DM interactions. Identifying and masking these *instrumental artifacts* is therefore crucial to preserve sensitivity to single-electron events.

One of the most prominent sources of artifacts are silicon defects in the CCD bulk, which can form charge traps that capture and later release charge carriers. These defects give rise to hot pixels—pixels with elevated dark current—or super-hot pixels, which saturate during the exposure and corrupt pixels during read-out; if contiguous in a column they form *hot columns*. Conversely, cold or bad pixels may arise from defects that reduce charge-collection efficiency, suppressing genuine signal.

Crosstalk between read-out channels can also create false signals by coupling the signal of one amplifier into neighboring regions of the image. These artifacts typically appear as ghost clusters at identical coordinates in neighboring CCDs. Additional sources of noise include transient read-out instabilities,

clocking imperfections and thermal drifts, all of which may contribute to baseline fluctuations or isolated spurious pixels.

During read-out, the accumulated charge is transferred across the CCD using a 3-phase clocking scheme (see Subsection 1.3.1). Small inefficiencies in this process, quantified by the CTI can lead to partial loss or distortion of signal packets.

Baseline drifts (slow shifts in the zero level of each pixel) are corrected with the prescan/overscan pedestal and, as detailed in Subsection 1.3.1, by correlated double sampling (CDS) or multiple Skipper reads. Through the combination of hardware design, online monitoring and offline masking, instrumental artifacts are kept at the sub-dru level in DAMIC-M. Nonetheless, their potential to mimic or obscure faint ionization signals requires continuous validation with control datasets and systematic masking strategies.

In the absence of irreducible physical backgrounds, the dark current becomes the dominant contributor in the 1–3 e^- range. Provided that instrumental artifacts can be reliably identified and suppressed, this leaves a clean window in which to search for DM–electron interactions. The expected shape of the DC spectrum is shown in Figure 2.1 (right panel), illustrating the contribution of each electron bin in the low-charge regime.

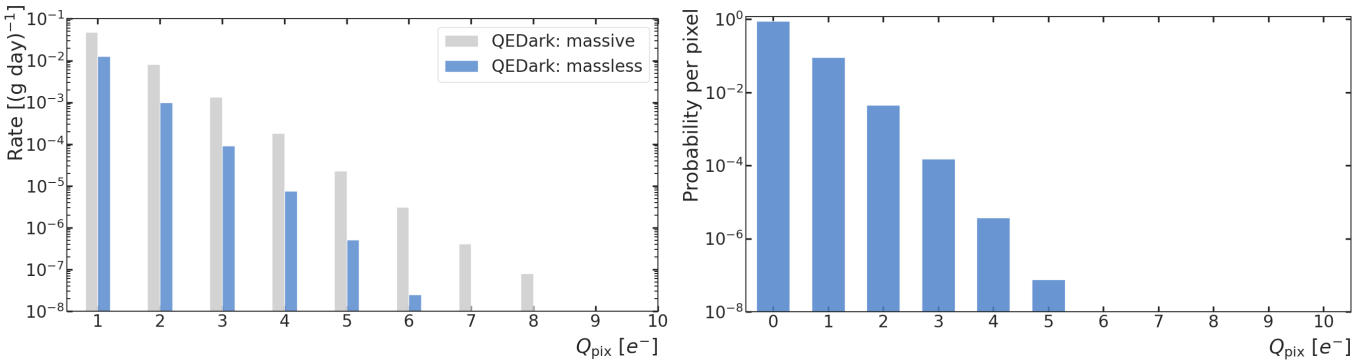


Figure 2.1: Left: predicted differential rate $R(Q_{\rm pix})$ for DM–electron scattering in two QEDark benchmarks (massive and massless mediator). Right: illustrative Poisson spectrum for a mean DC level $\lambda = 0.1\,e^-\,\text{pixel}^{-1}\,\text{image}^{-1}$; the real data average is $\lambda \simeq 2.2 \times 10^{-4}$ but the larger value is chosen for visual clarity.

The data-selection pipeline therefore aims to remove instrumental artifacts, external backgrounds that may survive shielding, and regions of the CCD affected by read-out issues or non-uniform response without suppressing potential DM signals or events consistent with dark current.

Chapter 3

Event Selection Framework

The suppression of instrumental and environmental backgrounds is essential for any search targeting single-electron signals from sub-GeV dark matter. This chapter details the methodology that converts the raw stream of Skipper-CCD images acquired in the LBC into a high-purity sample of pixels suitable for a low-mass DM search. The goal is to construct a robust and reproducible data-selection pipeline that maximizes the exposure while minimizing contamination from spurious instrumental or radiogenic sources. The selection strategy relies on a series of masking algorithms, some defined by the DAMIC-M collaboration and others developed specifically for this work.

3.1 Blinded Analysis Strategy

A blinded analysis strategy was implemented to avoid selection bias. This section presents the data-taking conditions, the characteristics of the dataset, and the structure of the blinded and unblinded samples used throughout the analysis.

3.1.1 Dataset and Run Conditions

The data analyzed in this work were acquired with the DAMIC-M LBC between October 2024 and January 2025. The dataset corresponds to two CCD modules, labeled PA07 (ID 104) and PA08 (ID 103), each equipped with four high-resistivity n-type silicon Skipper-CCDs. Every sensor contains 6144×1536 active pixels of $15 \times 15\,\mu\text{m}^2$ and is $670\,\mu$m thick, corresponding to a mass of 3.3 g per CCD. In this study, we focus exclusively on data from module PA08 to simplify the analysis workflow.

The CCDs were operated at a temperature of 130 K and a vacuum of 5×10^{-6} mbar, with a substrate bias of 45 V applied to ensure full depletion of the silicon bulk. The lateral diffusion of charge carriers during drift, which affects the spatial spread of low-energy events, depends on this bias voltage and is modeled accordingly in Section 1.3.3. A single amplifier per CCD was used for read-out.

Data acquisition was performed in continuous-read-out mode, meaning that images were read out continuously while simultaneously accumulating dark-matter exposure. Each image consists of 6300 columns and 16 binned rows, with each row integrating the charge from 100 physical rows summed vertically before entering the serial register (i.e., the last row of the CCD matrix). Note that more columns than the 6144 physical ones of the CCD are read out; these additional pixels, which receive significantly lower exposure, provide valuable information on data quality and on the dark current in the serial register. The corresponding exposure of each CCD pixel is approximately equal to the image read-out time of 1668 s, corresponding to an effective exposure of ~ 0.0193 days per pixel.

3.1.2 Blindness Protocol

To avoid analysis bias and ensure the integrity of the results, a blinded analysis strategy was adopted. Initial data selection criteria and image processing procedures were defined using an unblinded dataset corresponding to data acquisition on 2024-10-06, referred to as **UB0**. This dataset accounts for approximately 9 % of the total exposure (7.33 days of live time). Once the selection pipeline was finalized, it was applied without modification to the rest of the dataset. For this work, the blinded exposure was divided into four acquisition periods-2024-10-14, 2024-10-28, 2024-11-19, and 2024-12-17-corresponding to blinded datasets **B1**-**B4**. This segmentation, motivated by the large data volume, facilitates manageable

data handling, and the final results are expected to be consistent with those obtained from the full image dataset. Table 3.1 summarizes all datasets used in this study, including the number of images, the live time per CCD, and the relative contribution of each dataset to the total exposure.

Dataset	Start date	Images	Live time [d]	Exposure [%]
UB0	06-Oct-2024	380	7.33	-
B1	14-Oct-2024	470	9.07	12 %
B2	28-Oct-2024	935	18.00	25 %
B3	19-Nov-2024	1386	26.70	37 %
B4	17-Dec-2024	1006	19.40	26 %
Total	—	**4177**	**80.50**	**100 %**

Table 3.1: Summary of PA08-103 datasets used in this study. "Live time" refers to the effective exposure time per CCD, accounting for continuous-read-out operation. Since the four CCDs record data simultaneously, the total exposure is four times the live time before masking. Columns, from left to right, give the dataset label, the start date, the number of images acquired, the corresponding live time in days, and the percentage contribution to the total exposure. UB0 was used only to define the analysis cuts and is therefore excluded from the exposure calculation; the percentages listed for B1–B4 are normalized relative to the combined exposure of those datasets alone.

3.2 Selection Strategy Overview

The analysis pipeline consists of two main stages: a standard preprocessing sequence developed by the DAMIC-M collaboration, and a set of advanced masking algorithms, some newly introduced and others optimized in this work, to target subtle or spatially correlated backgrounds.

To guide this selection, we define a ROI that targets single-pixel events depositing between 1 and 10 electrons of total charge[1]. While the expected signal is concentrated in the lowest-charge bins, the 10 e^- upper bound ensures sensitivity to potential signal tails and non-standard scenarios. Within this ROI, the data selection aims to remove instrumental artifacts, external backgrounds that may survive shielding, and CCD regions with unstable response, while preserving events consistent with DC or a genuine DM signal.

The initial image processing (including Skipper compression, pedestal subtraction, charge calibration, cluster[2] identification) was performed using the softWAre for Dark matter ExpeRiments with Skippers (WADERS) framework. This software incorporates a modular library of processes (e.g., CompressSkipperProcess, PedestalSubtractionProcess, CalibrationProcess, etc.), each corresponding to a predefined algorithm for CCD data handling. The full processing sequence is described in detail in Appendix A.

Subsequent masking steps, developed specifically in the context of this work, go beyond the standard preprocessing and clustering procedures. These include the identification of hot columns (Sec. 3.3.3), overdensity regions (overdensity (OD)) (Sec. 3.3.4), high-charge pixel trails (CTI) (Sec. 3.3.5), charge multiplicity anomalies (Sec. 3.3.6), and isolated column artifacts (Sec. 3.3.7). The design and implementation of these masking algorithms constitute a central contribution of this analysis. Each masking criterion is physically motivated and its performance validated on unblinded data and/or Monte Carlo (MC) simulations. Table 3.2 summarizes the masks applied, their purpose, whether they were inherited or developed here, and are presented in the order in which they should be applied. These steps, their

[1] Above $Q_{\rm pix} \sim 5\,e^-$ the predicted dark-matter rates for our benchmark models become negligible; see Figure 2.1 left panel.

[2] A cluster is defined as a contiguous group of pixels with charge higher than a given threshold, typically produced by the passage of an ionizing particle through the CCD. These pixels share spatial proximity and exceed a defined charge threshold set according to the requirements of the dark matter search, as well as, the background sources.

motivation, and performance on unblinded data are detailed in the following sections, with supporting figures provided in Appendix B.

Mask Name	**Physical Motivation**	**Contribution Level**
Cluster Mask	As mentioned, the DM-electron scattering search targets the 1-10 e^- range, where ambient backgrounds are negligible. Therefore, any cluster with energy $> 10\,\text{eV}$ is excluded to remove ionizing radiation events.	Applied
Cluster Crosstalk	Cross-talk, whereby signals from one read-out channel unintentionally affect others, can induce spurious pixels during the read-out of high-charge clusters. These are identified and masked across CCDs.	Applied
Hot Columns	Columns with persistently elevated charge occupancy, often called hot columns or regions, typically caused from defects or damage in the sensor, leading to anomalously high-pixel charge even in the absence of ionizing particles.	Optimize
Overdensity Regions	Extended vertical regions with excess low-energy events, not captured by individual hot columns.	Designed
CTI Mask	Due to charge transfer inefficiencies (CTI), high-charge pixels can leak charge along transfer paths (vertically and horizontally). In addition, halos of low-charge pixels (1-5 e^-) can appear around these clusters, possibly due to Cherenkov radiation.	Optimize
Charge Multiplicity Mask	Suppress rows or columns with anomalously high numbers of charged pixels.	Optimize
Isolated Columns	Remove unmasked columns located between two fully masked ones, ensuring spatial continuity.	Implemented

Table 3.2: Summary of masking steps in the data-cleaning pipeline, indicating their purpose and origin.

3.2.1 Motivation for Pixel Masking

The spatial correlations presented in this subsection anticipate the two masking families applied later in the analysis:

(i) **Multiplicity / hot-column masks** (Section 3.3.6), designed to suppress persistent low-multiplicity excesses that appear repeatedly in the same pixel columns or in clusters of neighbouring pixels; and

(ii) **CTI-halo mask** (Section 3.3.5), which removes low-charge pixels that trail bright, high-energy clusters due to charge-transfer inefficiency (CTI) during read-out.

Before presenting the performance of the masking algorithms, it is essential to investigate the spatial behavior of low-charge events in the unmasked dataset. Specifically, this analysis explores whether pixels with 1, 2, or 3 e^- events show (a) correlations with the nearest high-energy cluster and/or (b) correlations

among themselves. The presence of such correlations would signal the influence of instrumental effects and justify the need for masking strategies.

Correlations with high-energy clusters

Figure 3.1 shows the distance from isolated low-charge pixels to the closest high-energy deposit either below (left panel) in the same column-pixels read out later during the vertical transfer-or to the right (right panel) in the same row-pixels shifted horizontally towards the amplifier. The solid black line shows the average result for 1 e^- pixels over 500 such pseudo-images, and the shaded grey band represents the corresponding 1σ statistical spread. Monte-Carlo predictions for 2 (red) and 3 e^- (blue) events are also shown.

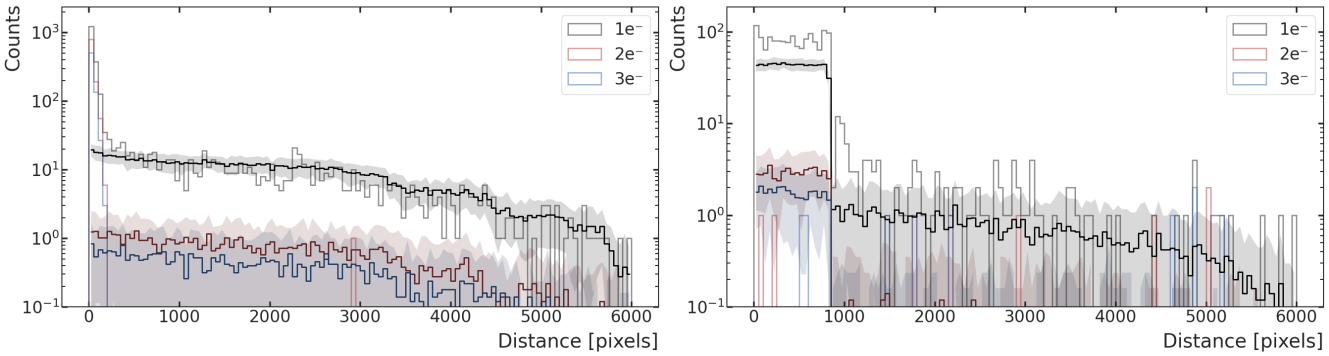


Figure 3.1: Distances from 1, 2, and 3 e^- pixels (step histograms in grey, red, and blue) to the closest high-energy cluster (left) bellow and (right) on the right in the unmasked dataset. Darker lines show the average of 500 MC images in which the same numbers of 1, 2, and 3 e^- pixels are placed randomly over the cluster-free area; shaded bands indicate the 1σ spread.

In the left panel (vertical transfer) the three colored step histograms (grey, red and blue for 1, 2 and 3 e^-, respectively) displays a pronounced excess in the first bins (up to $\lesssim 200$ rows). This is a classical signature of vertical charge–transfer inefficiency: a fraction of the charge stored in each cluster is trapped during the upward clocking of the column and then released a few transfers later, creating a short trail of single-electron pixels immediately below the cluster. Beyond ~ 200 rows the data settle back onto the MC expectation, indicating that isolated low-charge pixels at larger separations are consistent with uncorrelated DC statistics. Note that the MC toys use the same total of 1-, 2- and 3-e^- pixels as the data; if we had relied on the DC rates instead, 2- and 3-e^- pixels would be absent. In the right panel (horizontal transfer), one might expect similar behavior; however, this CCD exhibits a peculiar feature. A significant number of clusters reside in a damaged column at $x \simeq 5315$—likely a hot column, which will be discussed further in Section 3.3.3. When CTI effects release charge during the left-to-right serial transfer, the first affected pixel lies immediately to the right of the cluster. Since these pixels are typically adjacent to high-charge pixels, they are frequently masked by the cluster-selection algorithm (see Figure 3.3, left panel). As the active region extends only up to column 6144, this fixed one-column offset translates into apparent separations ranging from 1 up to ~ 800 pixels for clusters in that column. As a result, a nearly flat plateau emerges in the data for all three charge classes. Because the Monte Carlo simulations preserve the exact cluster positions observed in the data, they naturally reproduce this feature. Only at the largest separations, where clusters from other columns dominate, do the data and the MC predictions begin to converge.

Correlations with low-charge pixels

Figure 3.2 shows the pairwise distances between identical low-charge events (1, 2 and 3e^-) along columns (left panel) and rows (right panel). The same gray, red, and blue lines are used here as in the

previous figure to denote $1e^-$, $2e^-$, and $3e^-$, and the Monte Carlo is carried out in exactly the same way (same random re-distribution and clustering algorithm) as before. Because this MC preserves only overall counts but removes any spatial correlation, its $2e^-$ (dark red) and $3e^-$ (dark blue) curves remain nearly flat. In contrast, the MC $1e^-$ curve falls off smoothly with distance. In the real data, $1e^-$ exhibits a spike at low distance-far above the MC band. Beyond a few tens of pixels, however, the 1 e^- histogram re-enters the MC envelope, showing that single-electron events behave like Poisson DC at larger separations. The $2e^-$ and $3e^-$ data show modest peaks at low distances.

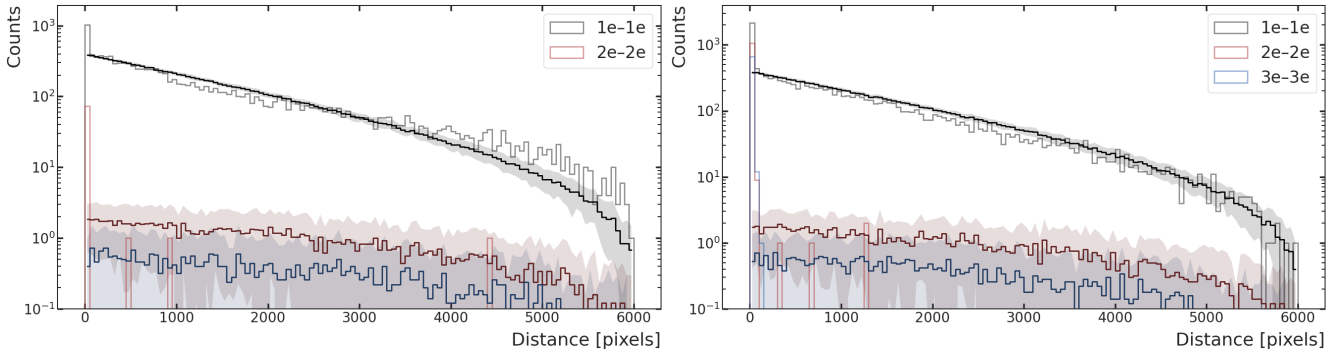


Figure 3.2: Pairwise distance distributions for equal-charge pixels before masking. Step histograms show the data for 1–1 e^- (grey), 2–2 e^- (red) and 3–3 e^- (blue) pairs. Solid colored curves with shaded bands represent the mean and 1σ spread of 500 toy Monte-Carlo images in which the observed numbers of 1, 2 and 3 e^- pixels are randomly placed on the cluster-free area, thus preserving global occupancies while removing spatial correlations.

This analysis highlights the need to apply and optimize the series of masking techniques previously introduced in Table 3.2 in order to effectively suppress multiplicity events within the ROI defined for the dark matter search. Such events are correlated with high-energy ionization, structural defects, or inhomogeneities in the silicon crystal, and-although they may mimic genuine dark matter interactions-can be discriminated based on their characteristic spatial correlations.

The selection strategy described in this section is applied to fully processed images. Standard image preparation steps—including pedestal subtraction, charge calibration, and preliminary clustering—are performed using the WADERS software developed by the collaboration. A detailed description of this processing framework is provided in Appendix A.

3.3 Custom Masking Techniques for low-background selection

3.3.1 Cluster Mask (WADERS)

Clusters are groups of contiguous charged pixels, produced when ionizing particles-such as muons, neutrons, or radiogenic backgrounds-deposit energy in the active silicon of the CCD. These interactions generate extended, often irregular charge distributions whose total signal typically exceeds the few-electron scale relevant for dark-matter searches. Because our analysis targets low-multiplicity events (0–10 e^-), any cluster whose summed charge exceeds 10 e^- must be removed to suppress background contamination. In this work we employ the standard WADERS clustering framework, which first identifies high-energy "seed" pixels and their neighbors via the `ClusterFinder` step, then builds a binary mask of all pixels belonging to each cluster using `BuildClusterMask`.

The ClusterFinder algorithm identifies individual pixels as signal if they fulfill the condition:

$$p_{th} > n_{cut} \cdot \sigma, \tag{3.1}$$

where n_{cut} is a configurable threshold obtained in Subsection A.4 and σ represents the read-out noise of the CCD, estimated independently for each CCD amplifier channel. In this analysis, a pixel is considered as signal if $p_{th} > 0.7\,\mathrm{e}^-$. To initiate the formation of a cluster, the algorithm requires the presence of a seed pixel, defined as one with signal significantly above the expected dark current fluctuations. Specifically, a pixel is considered a seed if its charge satisfies:

$$q_{\mathrm{pix}} > q_{\mathrm{seed}} = n_{\mathrm{seed}} \cdot \sigma, \tag{3.2}$$

with $n_{\mathrm{seed}} = 27$ in this analysis. This high threshold ensures that the probability of a random noise fluctuation producing a false seed is negligible (less than 10^{-12} for Gaussian noise with $\sigma \sim 0.15\,e^-$). Given a typical dark current rate of $\lambda \sim 3.6{\cdot}10^{-6}\,e^-$/pix/img, the occurrence of such high-charge pixels from DC alone is extremely rare. Therefore, pixels passing this cut are robust indicators of ionizing radiation. Once a seed is identified, the algorithm recursively includes neighboring pixels that pass the lower threshold defined in Equation 3.1, constructing the full cluster. An example of a reconstructed cluster, showing only the pixels that exceed the threshold and are therefore masked, is displayed in Figure 3.3 (left panel). Such clusters, dominated by ionizing radiation, are removed from the low-multiplicity analysis using the `BuildClusterMask` procedure.

It is important to note that the elongated vertical appearance of the reconstructed clusters is not an intrinsic feature of the ionizing radiation, but rather a consequence of the vertical binning applied during read-out, which groups the charge from 100 consecutive rows into a single bin. As a result, all clusters in this analysis appear vertically stretched, and their characteristic morphologies —such as the compact shape of electron recoils, the track-like signatures of alpha particles, or the more diffuse profiles of gammas— are no longer distinguishable. Representative examples of unbinned clusters with characteristic shapes can be found in Figure 1.3 (right panel).

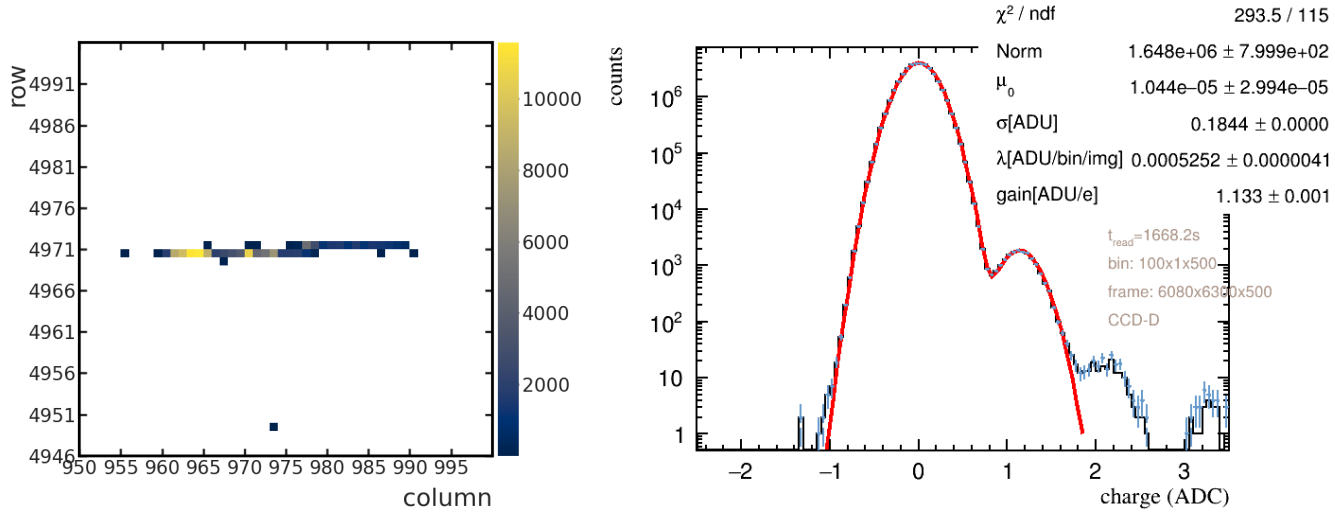


Figure 3.3: Illustration of the cluster masking procedure. Left: Example cluster detected in UB0 CCD D. Only pixels above the clustering threshold are shown; the seed (brightest pixel at the center) triggers the inclusion of neighboring signal pixels. This entire structure is excluded from the dark-matter ROI by the `BuildClusterMask`. Right: Pixel charge distribution in the same CCD after masking. he black histograms represent the charge distributions after applying cluster mask. The red fit corresponds to a Poissonian fit of two peaks.

A dedicated toy-MC study was performed to estimate the maximum total charge that could result from the accidental aggregation of dark current electrons. In this simulation, 10^6 images containing only Poisson-distributed DC ($\lambda = (2{-}5) \times 10^{-4}\,e^-\,\mathrm{pix}^{-1}\,\mathrm{img}^{-1}$) were generated, and clusters were identified

using the same 4-connected clustering algorithm applied to real data (pixels connected via their edges). The simulation results show that the largest contiguous cluster was ever observed carries at most $4\,e^-$, and the probability of creating a cluster with $Q_{\text{cluster}} \geq 4\,e^-$ is below $\sim 5 \times 10^{-6}$. Therefore, in real data any group of adjacent pixels whose summed charge exceeds this limit cannot plausibly arise from random pile-up of dark current electrons. In this analysis, only clusters extending at least one pixel in both horizontal and vertical directions and with a total charge greater than 10 e^- are masked. As illustrated in Figure 2.1 (left panel), the predicted differential rate for DM-electron scattering already drops by more than two orders of magnitude at $Q = 8\,e^-$ and is negligible by $10\,e^-$. Consequently, the $10\,e^-$ cut defines a region of interest where the expected signal-to-noise ratio is most favorable for dark matter searches, as it preserves low-charge events -where a potential dark matter signal would lie- while effectively suppressing backgrounds from ionizing radiation.

The impact of this cluster mask procedure is illustrated in Figure 3.3 (right panel) which shows pixel charge distributions from the unblinded dataset before (in blue) and after (in black) masking. The red curve represents a two-peak Poissonian fit used to extract the dark current and noise parameters from the cleaned distribution; the blue line shows the raw charge distribution before any masking and the black histogram corresponds to the distribution after applying the cluster mask.

3.3.2 Cluster Crosstalk

Cluster crosstalk refers to the phenomenon in which a charge signal in one CCD induces a spurious signal in neighboring CCDs due to electronic interference, resulting in artificial charge contamination. To mitigate this effect, a dedicated crosstalk mask is applied. This mask excludes all pixel positions where a cluster has been identified by the clustering algorithm (see Subsection 3.3.1), regardless of which CCD the cluster originated from. In other words, any pixel identified as part of a cluster in one CCD is masked at the same position across all CCDs. An example of this effect is illustrated in Figure 3.4. A high-energy cluster is observed in CCD C (third panel of Figure 3.4) with pixel values exceeding 10^4 electrons (approximately 37 keV). At the same coordinates-and read out simultaneously-signals of up to 1.5 electrons are visible in CCDs A and B, while CCD D exhibits a negative signal reaching -0.8 electrons. This sign inversion likely results from capacitive coupling between the read-out channels, where a strong signal in one amplifier induces a response of opposite polarity in others. The matching cluster topology across all CCDs further supports the interpretation of these signals as crosstalk, rather than coincidental charge deposits.

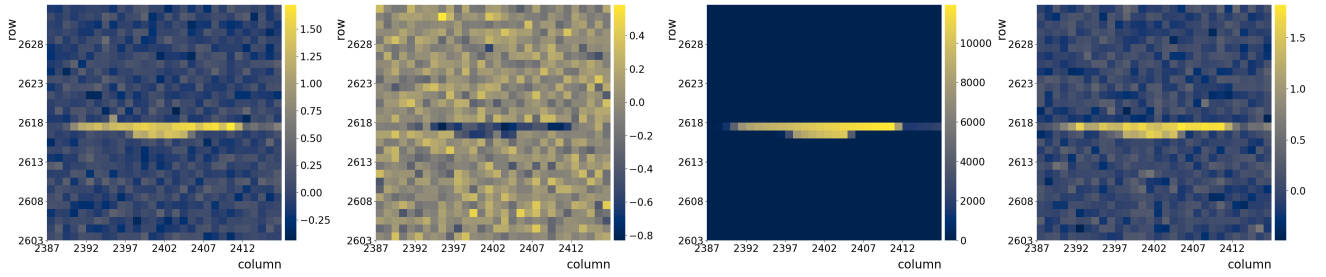


Figure 3.4: Example of cluster crosstalk in CCD module 103. A high-charge cluster is observed in CCD C, while fainter, spurious signals appear at the same pixel coordinates. Sorted top to bottom: CCD A, B, C and D, respectively.

3.3.3 Hot Columns

Hot columns are individual CCD columns whose occupancy of low-charge pixels is significantly higher than that of their neighbours. These features typically arise from defects in the silicon bulk, radiation-

induced damage, or instabilities in the read-out electronics. If not properly identified and masked, hot columns may mimic low-energy events or distort the charge distribution, compromising the accuracy of the analysis.

Detection and masking follow a two-stage procedure: (i) identification of candidate columns through a statistical threshold, and (ii) block expansion to capture neighbouring columns that belong to the same defect.

Candidate hot columns are first identified via their *column occupancy*, $M_{\mathrm{col},i}$, which is computed using only unmasked pixels as:

$$M_{\mathrm{col},i} = \frac{N_i^{\mathrm{low}}}{N_i^{\mathrm{unm}}} \cdot N_{\mathrm{rows}}, \tag{3.3}$$

where N_i^{low} is the number of unmasked pixels in column i with charge in the range $[0.7\,\mathrm{e}^-, 3.7\,\mathrm{e}^-]$, N_i^{unm} is the total number of unmasked pixels in the same column, and N_{rows} is the number of rows in the image. The $M_{col,i}$ values are then sorted in ascending order to form a one-dimensional profile. The algorithm searches for the largest discontinuity in this profile, defined as the maximum increase between two consecutive values in the sorted list. This jump indicates an abrupt transition between background and anomalous behavior. The upper value of the largest jump defines the threshold $M_{\mathrm{col}}^{\mathrm{th}}$; all columns with $M_{\mathrm{col}} > M_{\mathrm{col}}^{\mathrm{th}}$ are marked as hot. The procedure is repeated iteratively on the remaining columns until no additional discontinuities are found, requiring in each step that the new threshold satisfies $M_{\mathrm{col}}^{\mathrm{th}} \geq \mu_{\mathrm{bck}} + 3\sigma_{\mathrm{bck}}$ (with $\mu_{\mathrm{bck}}, \sigma_{\mathrm{bck}}$ computed from the $\sim 60\%$ lowest-occupancy columns). While this iterative detection efficiently identifies isolated columns with abnormally high occupancy, it may miss extended structures where the anomaly affects not only a single column but also its neighbors. In particular, hot columns with strongly elevated occupancy can induce a spatially correlated increase in nearby columns, forming a profile that decreases gradually (commonly referred to as "exponential" hot columns). An illustrative example from unblinded data is shown in Figure 3.5.

The left panel shows a "single" hot column: the occupancy rises sharply in one column and drops back to background in the neighbors. The right panel on the other hand, presents an "exponential" hot column whose tail (red curve: exponential fit) decays over several adjacent columns. A second step therefore analyses the spatial continuity of the signal and expands each candidate block.

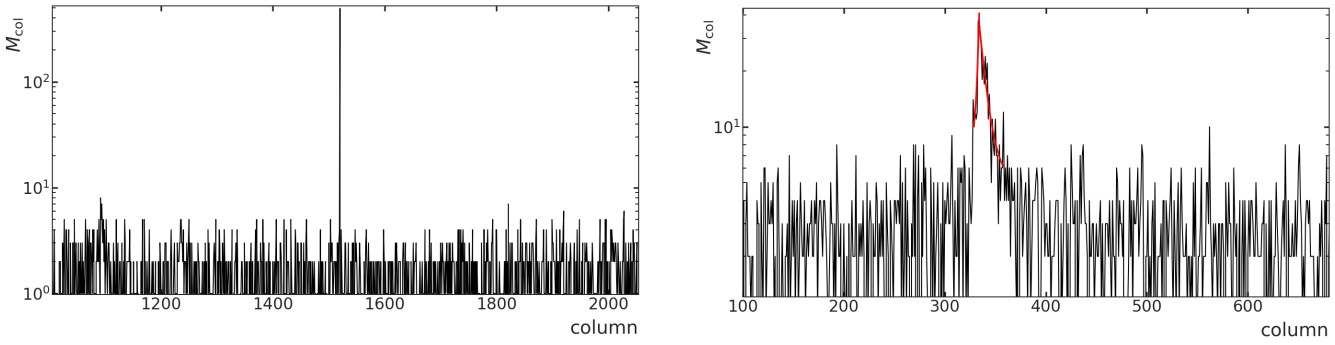


Figure 3.5: M_{col} profiles of two hot-column types observed in the UB0 dataset that motivate the block-expansion algorithm. Left: Single hot column; Right: Exponential hot column (red: fit).

The detected hot columns are grouped into contiguous blocks; two columns belong to the same block if they are separated by at most one clean column (gap ≤ 1). Each identified block is then expanded by examining a sliding window of 20 columns on either side. In this region, additional columns are included if at least two exceed a relaxed threshold defined by:

$$M_{\mathrm{col}}^{\mathrm{exp}} = M_{\mathrm{col}}^{\mathrm{th}} - \Delta M_{\mathrm{col}}, \quad \Delta M_{\mathrm{col}} = 6\mathrm{e}^-. \tag{3.4}$$

Expansion stops when no further columns meet the criterion, and adjacent blocks separated by one clean column are finally merged.

The choice of $\Delta M_{\text{col}} = 6\text{e}^-$ is based on Monte Carlo validation studies, ensuring that the mask expansion efficiently captures spatially correlated activity while minimizing false positives.

The robustness of the algorithm was tested with a MC study that reproduces the statistical conditions of the PA08 dataset. Each simulated image has a size of 6800×6300 pixels and includes a realistic background of dark current and electronic noise. The DC is drawn from a Poisson distribution with mean $\mu_{\text{DC}} = 2.7 \times 10^{-4}\,\text{e}^-\,\text{pix}^{-1}\,\text{day}^{-1}$, and a Gaussian read-out noise with standard deviation $\sigma_{\text{read}} = 0.16\,\text{e}^-$. Seven synthetic hot columns are added to every image: five isolated columns, whose occupancies are randomly chosen between 20 and 100 low-energy pixels, and two "exponential" columns whose central core contains 30-100 low-energy pixels and whose exponentially decreasing tails contaminate neighboring columns. The length of the tails and their decay constants are drawn randomly: the left and right tails extend up to 25 columns, with decay rates μ sampled independently from a uniform distribution between 0.01 and 0.25. A total of 10^3 images were generated to provide good statistical precision.

The refinement stage of the mask has two free parameters: the expansion radius (`neighbor_range`), which is the number of neighboring columns inspected on each side of a candidate hot region (10, 20, 30, 40, 50, 60); the relaxed threshold ΔM_{col} $(0, 2, 4, 5, 6, 8, 10\,\text{e}^-)$, which allows a neighboring column to be included when its multiplicity is within ΔM_{col} of the main threshold. To find the best configuration, different combinations of these two parameters were tested. For each case, three metrics were calculated:

- The **efficiency** (or true positive rate), defined as the fraction of real hot columns correctly identified;
- The **false negative rate (FNR)**, corresponding to missed hot columns ;
- The **false positive rate (FPR)**, or the fraction of clean columns wrongly masked.

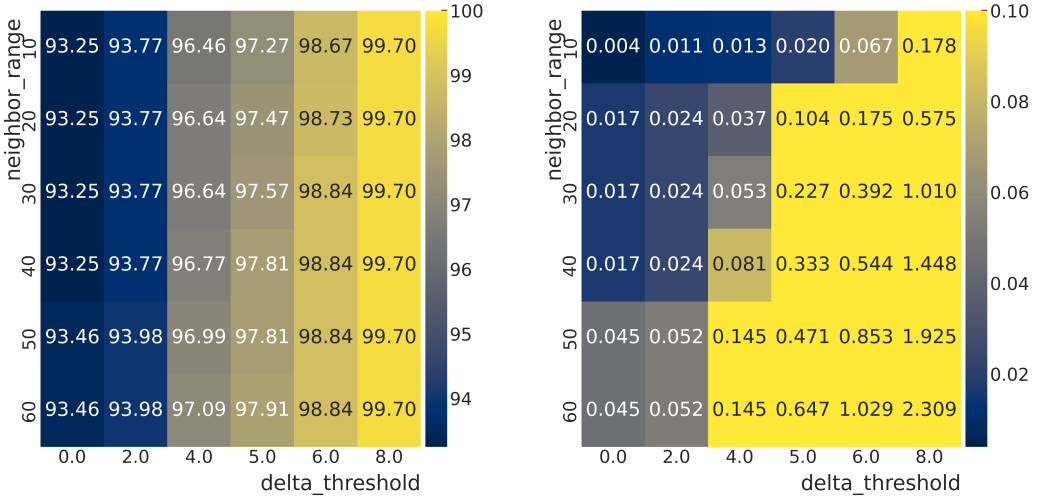


Figure 3.6: Performance of the hot-column-detection algorithm for the full parameter scan (ΔM_{col} and `neighbor_range`). Left: detection efficiency (eff). Right: FPR

Left panel of Figure 3.6 shows that the algorithm reaches efficiencies above 93% over most of the parameter space. The absolute maximum, $\varepsilon = 99.70\%$, appears at `neighbor_range` $= 10$ and $\Delta M_{\text{col}} = 8$ e$^-$; however, that working point increase the masked area to FPR $\simeq 0.18\%$. Conversely, a tighter

setting of `neighbor_range` = 20 combined with a relaxed threshold of $\Delta M_{\text{col}} = 6$ e$^-$ provides an excellent compromise for the simulated data: $\varepsilon = 97.27\%$, FPR = 0.020%, and FNR = 2.73%.[3]

$$\boxed{\texttt{neighbor\_range} = 10,\ \Delta M_{\text{col}} = 5\,\text{e}^-}$$

This configuration retains the majority of genuine hot columns while keeping the masked fraction very low.

The performance of the hot column detection algorithm was also evaluated on the unblinded dataset using the parameter values obtained during validation. Figure 3.7 presents the normalized column occupancy (M_{col}) for two representative sensors: CCD A and CCD C. The corresponding distributions for CCDs B and D are relegated to Appendix B for brevity. Red dashed lines indicate the positions of columns flagged as hot by the algorithm.

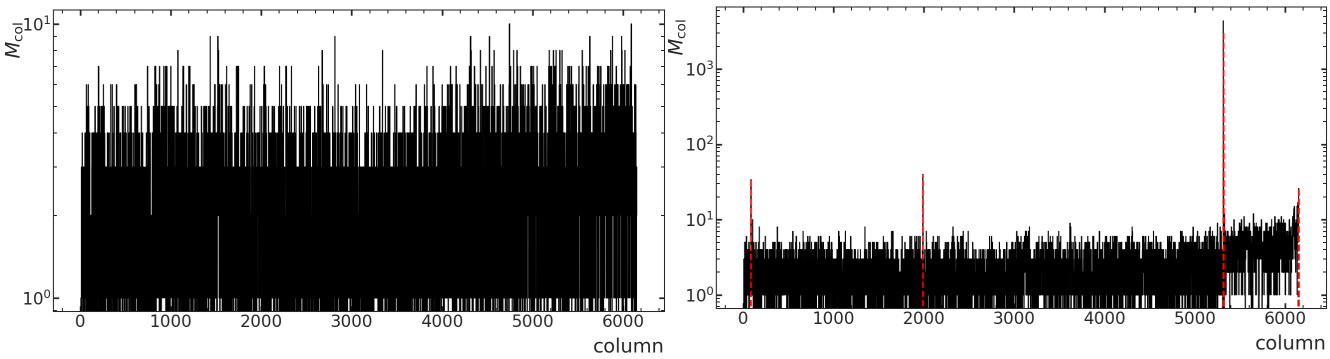


Figure 3.7: Normalized column occupancy M_{col} for CCD A (left) and CCD C (right) in UB0, displayed on a logarithmic scale. Red dashed lines mark columns flagged as hot by the algorithm. The corresponding plots for CCD B and CCD D are shown in Appendix B, Figure B.1.

In CCD A (left panel of Figure 3.7), no column exhibits significant deviations from the background level, confirming the absence of hot columns and confirming the absence of localized defects in the low-charge regime. In CCD B (left panel of Figure B.1), a narrow and intense peak appear at columns 1520. This feature correspond to a hot column with column occupancy reaching values hundreds of times above the average. Such extreme activity is likely caused by severe localized defects in the silicon substrate or electronic leakage paths, which generate persistent charge independent of true ionizing events. The behavior of CCD C (right panel of Figure 3.7) reveals multiple types of hot column activity. Several isolated features are visible across the sensor, with varying intensity. In particular, a group of very hot columns centered around column 5316 stands out, with thousands of pixels showing elevated occupancy. Notably, column 5316 also marks the beginning of a broader region characterized by persistently elevated occupancy values, which extends beyond the localized peak and will be analyzed in more detail in the following section. Additionally, a wide block of consecutive hot columns is observed starting at column 6132. Finally, CCD D (right panel of Figure B.1) exhibits a broad exponential-shaped region spanning columns 328 to 358 (see right panel of Figure 3.5). Such an exponential tail is more consistent with trapping effects, possibly originating in the serial register, rather than simple lateral diffusion of charge.

A summary of the hot regions detected in each CCD is provided in Table 3.3. The second column lists the columns initially flagged by the thresholding method, while the third column shows the final

[3]For comparison, increasing the relaxed threshold to $\Delta M_{\text{col}} = 6$ e$^-$ boosts the efficiency to 98.66% but raises the false-positive rate by more than a factor three (FPR = 0.067%).

expanded blocks. The fourth column reports the total fraction of masked columns with respect to the full CCD width. The masked area remains below 0.5% in all cases, confirming that the algorithm preserves exposure while reliably identifying problematic regions. The results demonstrate that the procedure is effective in real data, capturing both isolated and extended hot columns with high sensitivity and minimal over-masking.

CCDs	HC Detected	Expanded HC Blocks	%
A	None	None	0.000
B	1520	1520	0.016
C	86, 1989, [5315, 5318], 6141, 6146, [6150, 6151]	86, 1989, [5315, 5318], [6141, 6151]	0.277
D	328, [333, 342], 344	[328, 350]	0.374

Table 3.3: Summary of hot column detection in the UB0 dataset. The second column lists the columns initially flagged by the thresholding algorithm. The third column shows the final blocks after expansion. The last column reports the fraction of masked pixels with respect to the total number of columns in the CCD.

3.3.4 Overdensity Regions

Overdensity regions are areas of the CCD exhibiting a statistically excess in the local multiplicity of low-energy events ($1 - 3e^-$), with respect to the surrounding columns as we see in Figures 3.8 and B.2. Such structures are attributed to extended sensor defects, thermal gradients that break the flat-field uniformity, surface traps, or radiation-induced damage. Since these anomalies can span tens to hundreds of columns, they are not captured by the hot-column mask and must be detected and masked explicitly.

The detection algorithm proceeds in three sequential stages: (i) sliding-window scan to compute the local multiplicity median, (ii) sub-window refinement to profile continuous excesses, and (iii) merging and length-filtering to define the final overdense bands. Starting from the column-multiplicity profile $M_{\mathrm{col},i}$ defined in Section 3.3.3, a sliding window of width $W = 100$ columns is used to compute the median column multiplicity μ_{global} of low-energy pixels $M_{\mathrm{col},i}$:

$$\mu_{w,i} \;=\; \mathrm{median}\big\{M_{\mathrm{col},j}\big\}_{j=i}^{i+W-1}, \quad i = 0, \ldots, N_{\mathrm{cols}} - W.$$

A window is considered as "candidate" whenever $\mu_{w,i}$ exceeds $\mu_{w,i} > \mathrm{w_{thr}} \cdot \mu_{global}$, where $\mathrm{w_{thr}} = 1.7$. This threshold was optimized via MC studies to detect genuine extended excesses while keeping the overall masked fraction low. For each candidate window $[i,\, i + W - 1]$, we examine sub-windows of width $W_{\mathrm{sub}} = 10$ columns. Inside these sub-windows, the threshold is relaxed to $\mu_{subw,i} > \mathrm{subw_{thr}} \cdot \mu_{global}$ with $\mathrm{subw_{thr}} = 1.2$ so that the slowly varying edges of real OD bands are kept continuous. Finally, neighboring slices are merged and the resulting band is retained only if its length is at least 150 columns. Shorter bands are considered statistical fluctuations and discarded. All columns within the surviving bands are masked in subsequent low-energy analysis.

To quantify the performance of the overdensity algorithm, a MC study-built on the same framework used for the hot-column validation-reproduced the statistical conditions of the PA08 dataset. Each synthetic image (6800×6300 pixels) contains a baseline dark current $\lambda = 2.7{\times}10^{-4} e^-/\mathrm{pix/img}$ and a Gaussian read-out noise $\sigma = 0.16\, e^-$. A shallow parabolic gradient in λ is added to mimic the flat-field curvature observed in real data. Zero, one, or two OD bands are then injected at random positions; every band spans 150-800 columns and its mean DC rate is boosted by a factor 1.2-1.4. The boost is implemented as an additional Poisson draw that contributes on average 1-3 single-electron pixels per column, following

either a linear "ramp" or a flat "plateau" profile. More than two thousand parameter combinations of the sliding-window algorithm were then exercised: principal widths $W = 50, 100, 150, 200$; sub-windows $W_{\mathrm{sub}} = 5, 10, 20$; global thresholds $\mathrm{w_{thr}} = 1.25$-1.95; local thresholds $\mathrm{subw_{thr}} = 1.20$-$1.50$; band lengths $L_{\mathrm{min}} = 50, 100, 150, 200$ and gap tolerances gap_max $= 0, 20, 30, 50$. For each tuple 500 statistically independent images were generated and the true-positive (efficiency), false-positive and false-negative rates were averaged. Two limiting regimes emerged. Loose thresholds ($\mathrm{w_{thr}} \leq 1.35$) recovered virtually every band, but masked a quarter of the sensor (FPR $\gtrsim 0.25$). Conversely, very tight thresholds ($\mathrm{w_{thr}} \geq 1.75$) reduced the masked area below 1×10^{-2} yet missed one band in three (FNR $\gtrsim 0.15$). The configuration finally adopted is $W = 100$, $W_{\mathrm{sub}} = 10$, $\mathrm{w_{thr}} = 1.70$, $\mathrm{subw_{thr}} = 1.20$, $L_{\mathrm{min}} = 150$ for which the MC gives eff $= 0.923$, FPR $= 4.6 \times 10^{-3}$ and FNR $= 0.073$. To verify that no simpler or more sophisticated tool could outperform the tuned sliding-window, the same MC sample was processed with six change-point algorithms from the `ruptures` library-*PELT*, *Window*, *Binseg*, *BottomUp*, *Dynp* and *KernelCPD*-plus an unsupervised anomaly-detection scheme based on *Isolation Forest*. Each detector underwent an independent grid search on a calibration subset. The best scores obtained were eff $= 0.55$, FPR $= 0.65$ for PELT (`pen`=100), eff $= 0.50$, FPR $= 0.53$ for sliding Window (width 80, $n_{\mathrm{bkps}} = 3$), and eff $= 0.41$, FPR $= 0.32$ for Isolation Forest (trained on pixel-multiplicity profiles with default hyper-parameters). All other ruptures modes-Binseg, BottomUp, Dynp and KernelCPD-yielded similar or worse trade-offs: they retrieved less than 50 % of the true bands and misclassified 30-100 % of clean columns as defective. Although the selected sliding-window parameters do not provide perfect recall, they combine the highest efficiency with the lowest masking overhead among thousands of tested variants and outperform every generic change-point or isolation-based method. They are therefore adopted as the baseline for the blinded analysis.

Figure 3.8 shows the normalized column occupancy M_{col} after hot-column masking for CCD A (left panel) and CCD C (right panel). Black bars show the raw per-column counts, the blue line is a Savitzky-Golay smoothing of the same data and the red curve is the quadratic baseline obtained from a robust fit. Columns flagged by the OD algorithm are highlighted in orange.

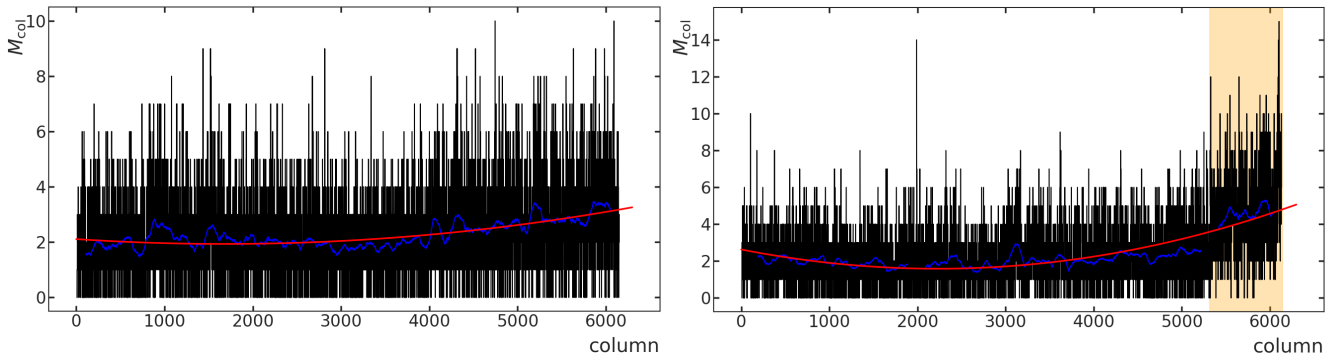


Figure 3.8: Normalized charge distribution per column after hot-column masking for CCD A (left) and CCD C (right) in UB0. Orange bands mark the regions masked as overdense; the blue curve is a 201-point, second-order Savitzky-Golay smoothing, and the red line the quadratic baseline fit. The corresponding plots for CCD B and CCD D are shown in Appendix B, Figure B.2.

CCD A (left panel of Figure 3.8) exhibits a uniform charge distribution: the blue Savitzky-Golay trace oscillates symmetrically about the quadratic baseline, with excursions fully compatible with Poisson noise from DC. Because no section meets the OD criteria, the mask leaves the entire sensor untouched and A is regarded as free of any large-scale structure. CCD B (left panel of Figure B.2) exhibits a single overdense slice at the far serial edge, spanning approximately columns 5823-6056. In this region, the smoothed profile rises almost linearly above the baseline-a shape characteristic of charge-injection or bias-edge phenomena.

The algorithm marks this entire block, ensuring that the modest but persistent excess of one- to three-electron hits does not bias the overall DC estimate. CCD C (right panel of Figure 3.8) carries a broad excess beginning near column 5311 and extending to the end of the image. Both the smoothed curve and the quadratic fit show a gentle upward curvature, revealing a low-frequency spatial gradient superimposed on a genuine surplus of single-electron counts. This flagged region overlaps the hot columns previously removed, indicating that beyond the sharp spikes there is a diffuse tail of low-level activity spanning many tens of columns. CCD D (right panel of Figure B.2) contains two overdensity blocks: one from roughly 5657 to 5836 and a second from about 5908 to 6160. Each island produces a local bump of two to four extra one-electron pixels above the baseline, and between them the profile returns to the global mean.

Overall, the OD algorithm successfully identifies and masks extended regions of excess charge. Table 3.4 summarizes the coordinates of the bands for each CCD.

CCDs	**Overdensity band(s) [columns]**	**%**
A	none	0.0
B	[5823, 6056]	3.8
C	[5311, 6140]	13.5
D	[5657, 5836], [5908, 6160]	7.0

Table 3.4: Coordinates of overdensity regions identified in the UB0 dataset. The last column reports the fraction of masked pixels with respect to the total number of columns in the CCD.

3.3.5 CTI Mask

Extended blobs or high-energy clusters can degrade the charge-transfer efficiency during read-out. When a very energetic pixel is shifted through the CCD, a small fraction of its charge may be left behind, creating vertical or horizontal trails. Because these CTI trails consist of $1 - 10e^-$ pixels, they can mimic low-energy events and leak into the region of interest defined for the dark matter search. The presence of such contamination is supported by the observed excess of low-charge pixels found near clusters, as shown in Figure 3.1. Moreover, by introducing spurious pixels that are not Poisson-distributed, they distort the expected statistical behavior of the dark current background. CTI effects cannot be accurately simulated due to the complexity and variability of their physical origin. Therefore, the algorithm for their identification is defined using real unblinded data (UB0), where their morphological characteristics can be directly observed and studied. In addition to the vertical and horizontal trails, bright clusters often display a faint halo of 1-5 e^- pixels surrounding the core. This halo is produced by a combination of lateral charge diffusion and secondary emission inside the silicon (e.g. internal Cerenkov light or mild bremsstrahlung), generating a sparse ring of low-energy pixels a few columns wide. Although much weaker than the main cluster, these pixels fall exactly in the energy region of interest and must therefore be masked with dedicated expansion rules (HaloCTI in Table 3.5).

This mask is carried out in two steps: identification of seed pixels based on charge thresholds, and application of expansion rules tailored to the expected trail morphology. Two thresholds are used to categorize cluster-energy seeds: $50 \leq Q < 100$ e^- and $Q \geq 100$ e^-. The split reflects the observation that brighter seeds generate longer CTI tails. For every seed, pixels are masked according to Table 3.5. Three trail morphologies are considered: vertical CTI (VCTI), horizontal CTI (HCTI) and a faint symmetric halo (HaloCTI) around very bright seeds. The VCTI mask targets pixels read after the seed along the vertical direction (i.e., subsequent rows in the read-out sequence), while the HCTI mask corresponds to pixels read after the seed horizontally, during the serial transfer to the amplifier. In contrast, the HaloCTI mask covers pixels immediately surrounding the seed, assuming an isotropic origin unrelated to

the read-out direction. In Table 3.5, each value indicates the number of pixels masked on either temporal side of the seed. For example, '10 pixels' in VCTI corresponds to ten rows read after the seed in both physical directions (upward and downward), since both are shifted later in time. Because 100 physical rows are summed into one binned row, a genuine vertical trail spanning hundreds of rows appears only a few rows long in the binned image; hence the comparatively small VCTI radii. Horizontal trails, on the other hand, extend across many columns in the serial register, justifying more aggressive masking.

CTI Type	$50 \leq Q < 100$ **e**$^-$	$Q \geq 100$ **e**$^-$
VCTI	10 pixels	100 pixels
HCTI	100 pixels	6300 pixels (entire CCD row)
HaloCTI	No expansion	10 pixels

Table 3.5: Mask expansion applied to High-energy seeds. Values indicate the number of pixels masked in each direction.

The masking strategy was defined using unblinded data by testing different expansion configurations for each CTI type. The aim is to maximize the containment of charge trails while minimizing the loss of unaffected area where the DM-electron scattering search will be performed. Figure 3.9 shows how the masking fraction evolves for three CTI components and for four low-energy charge ranges (1 e$^-$, 2 e$^-$, 3 e$^-$, and ≥ 4 e$^-$) in UB0 CCD D. The grey histogram corresponds to the smallest expansion, while the blue gradient marks highest radii. All panels were produced with the seed ($Q_{\text{seed}} \geq 100$ e$^-$); seeds in the range $50 \leq Q < 100$ e$^-$ give the same trends with proportionally lower percentages. As the left panel shows, the masked fraction for 2e$^-$ and 3e$^-$ pixels remains essentially constant once the expansion exceeds about 5 px binned, indicating that CTI trails in those bins are already fully captured at very small radii. By contrast, the 1e$^-$ population-the dominant CTI contaminant-continues to grow up. In CCD D, increasing the V-CTI radius from 5 px to 50 px raises the masked 1e$^-$ fraction from 0.67% to 1.07% while only increasing the total masked area from 0.034% to 0.310%. Extending further to 100 px binned removes 1.37% of all 1e$^-$ pixels at the cost of masking 0.615% of the sensor. Doubling the radius again to 200 px only boosts the 1e$^-$ rejection by 0.63% but doubles the masked area to 1.21%. Therefore, a 100 px binned expansion (10 000 physical rows) is chosen for seeds with $Q_{\text{seed}} \geq 100$ e$^-$. This radius captures the vertical CTI trail seen in unbinned data while keeping the total V-CTI-masked area below 1%, leaving margin for the remaining CTI masks (Halo and H-CTI) and other filters. For medium seeds (50-100 e$^-$), whose CTI tails are correspondingly shorter, a 10 px binned radius is sufficient.

Horizontal trails can cross thousands of columns. In right panel the masked fraction for 2-3 e$^-$ pixels levels off as soon as the expansion reaches ~ 500 pixels; extending the mask further does not cause changes in the percentage, indicating that the trail has been fully covered. To be conservative the entire row (6300 pixels) is masked when $Q_{\text{seed}} \geq 100$ e$^-$, ensuring that no residual trail survives. For $50 \leq Q < 100$ e$^-$ the horizontal smear is shorter, and a 100-pixel margin is enough.

The halo consists of a faint corona only a few columns wide. As shown in the center panel of Figure 3.9, a radius of 10 pixels already masks $\sim 60 - 70\%$ of the ≥ 4 e$^-$ pixels (mostly corresponding to the high-charge seeds themselves) and $\lesssim 30$ % of 2–3 e$^-$ pixels. Larger expansions lead to significantly more masking at low energies without a proportional gain, increasing the overall masked area of the CCD. To balance containment and area preservation, a 10-pixel radius is fixed for $Q_{\text{seed}} \geq 100$ e$^-$ and apply no expansion for lower energies, where the CTI effect is milder.

The final CTI mask is obtained by combining masks for both charge regimes. Figure 3.10 illustrates typical examples in unblinded CCD A. In Left panel of Figure 3.10, a bright pixel is observed with a long horizontal charge trail. The horizontal trail spans multiple pixels, which are correctly masked by the

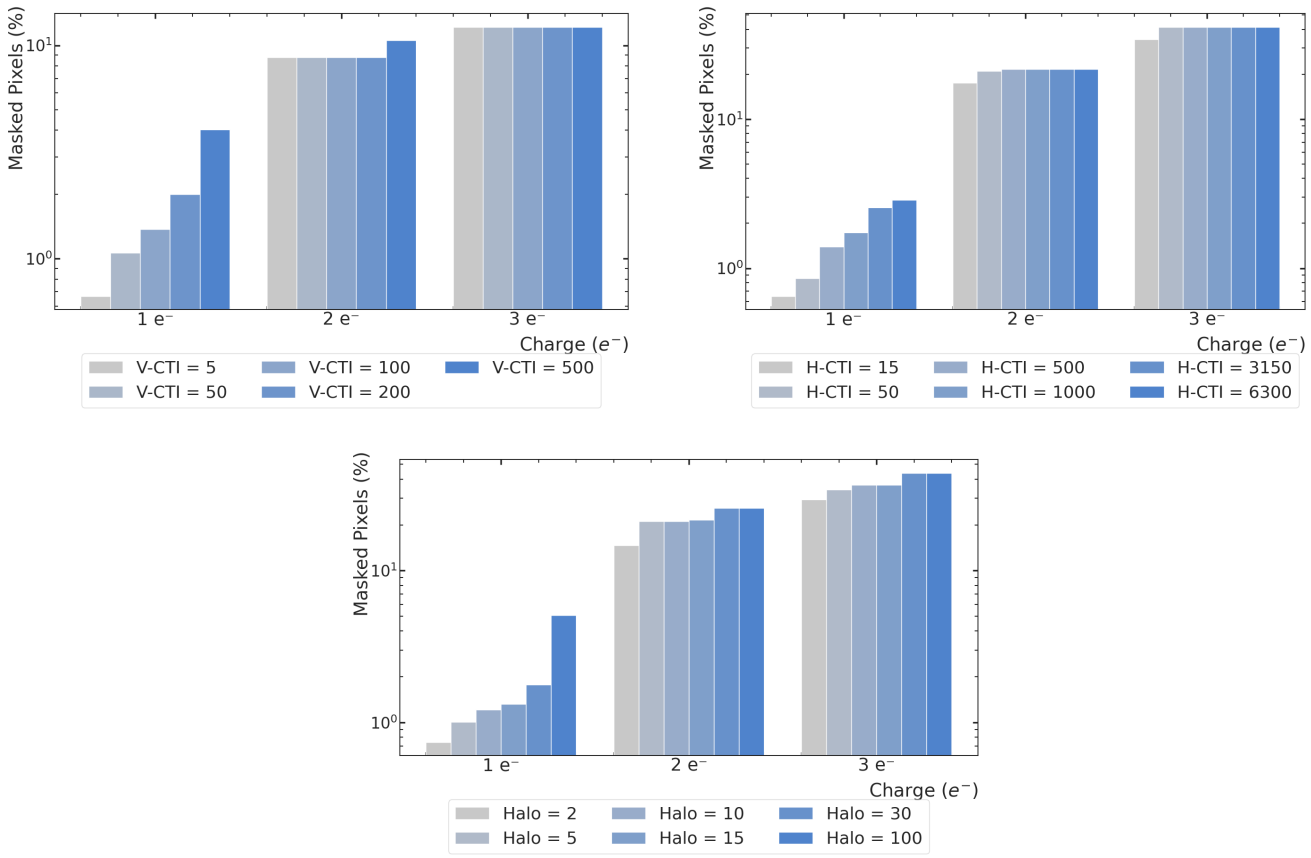


Figure 3.9: Masked-pixel fraction in each low-energy charge range as a function of the expansion radius around high-Q seeds ($Q_{\text{seed}} \geq 100$ e$^-$). The grey bar is the minimum expansion; deeper blues indicate larger radii Left: VCTI; Center: Halo CTI; Right: HCTI.

HCTI expansion. The Right panel of Figure 3.10 shows a zoomed-in region from a CCD after applying the full CTI masking. The corresponding VCTI, HCTI, and HaloCTI masks applied to each cluster are clearly visible.

A quantitative check confirms that the mask mainly removes truly energetic pixels (Table 3.6). More than 98 % of pixels with $Q > 4e^-$ are excluded-indicating that these $4e^-$ pixels are highly correlated with genuine high-energy clusters in their vicinity. A similarly high fraction of $3e^-$ pixels is also removed, reflecting that many of them arise as part of energetic cluster substructures. In contrast, fewer than 2 % of single-electron ($1e^-$) and $2e^-$ pixels are lost overall, as expected, since dark-current-induced hits are not correlated with high-energy clusters and dominate the $1e^-$ population.

Total Masked [%]	$1e^-$ [%]	$2e^-$ [%]	$3e^-$ [%]	$> 4e^-$ [%]
0.65	1.67	11.49	47.06	98.09

Table 3.6: Masking impact per charge bin for the selected CTI configuration. The last column reports the fraction of masked pixels with respect to the total number of columns in the CCD.

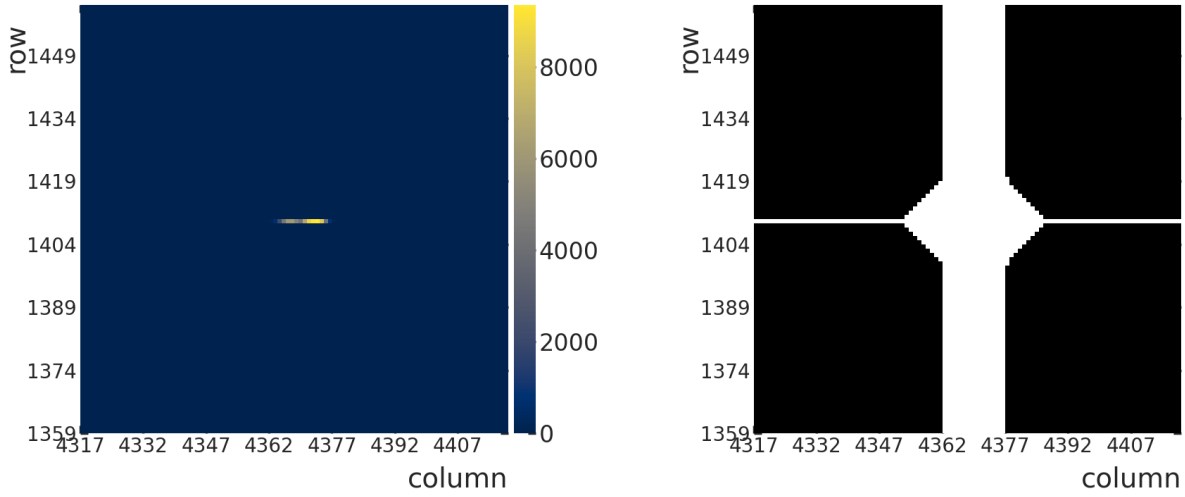


Figure 3.10: Application of the CTI mask in UB0 CCD A. Left panel show a high-energy cluster and the right panel a zoom-in of the CTI mask.

3.3.6 Charge Multiplicity

Rows and columns with an unusually high *multiplicity*-defined as the number of pixels in that row or column with charge $Q > 0.7$ e$^-$-are potential indicators of defects or correlated noise. After all previous masks have been applied, two complementary criteria are used to decide whether a whole row/column must be discarded: The multiplicity distributions for rows and columns are computed, which are expected to be purely statistical: for a uniform DC rate λ, the distribution follows a binomial law that approaches a Gaussian in the limit of large statistics. Because only the first few populated bins (m = 0-6) have sufficient statistics, a full-range Gaussian fit is unstable. Instead, the fit is performed only on the upper tail (bins above the 60$^{\text{th}}$ percentile), and a row/column is masked once the data exceed that fit by more than 3σ. If no bin satisfies the 3σ excess, the cut is placed at the 99$^{\text{th}}$ percentile, providing a conservative upper bound:

$$T_{\text{mask}} = \begin{cases} \min_{m \geq m_{60}} \left\{ m \mid N_{\text{data}}(m) - N_{\text{fit}}(m) > 3\sigma_{\text{data}} \right\}, & \text{if deviation found} \\ 99\text{th percentile of distribution}, & \text{otherwise} \end{cases} \tag{3.5}$$

where m_{60} is the multiplicity at the 60% percentile, N_{data} is the observed bin content, and N_{fit} is the value from the Gaussian fit. This procedure suppresses only those rows/columns whose multiplicities are statistically incompatible with a pure dark current process while leaving the bulk of the detector untouched.

Rows or columns are directly masked if they contain more than one non-consecutive pixel exceeding $Q \geq 2\,\text{e}^-$, since under normal operational conditions the probability that two non-adjacent pixels in the same row or column both fluctuate above $2\,\text{e}^-$ from DC alone is $\ll 10^{-6}$ for the measured λ. If two adjacent pixels exceed $2\,\text{e}^-$, they may originate from a single energetic cluster.

The validity of the two masking criteria was assessed with UB0 data and a dedicated MC simulation of dark current. The MC assumes that the charges per pixel follow a Poisson law with mean λ, convolved with Gaussian read-out noise. To reproduce detector non-idealities it includes (i) the measured per-column $\lambda(x)$ profile, (ii) the live-pixel mask, and (iii) the read-noise distribution. Figure 3.11 shows the row (left) and column (right) multiplicity distributions for UB0 CCD A. The grey bars are the data; black markers are the same points with Poisson uncertainties; blue squares are the MC prediction; the red curve is the

Gaussian tail fit; and the orange line marks the 99$^{\text{th}}$ percentile.

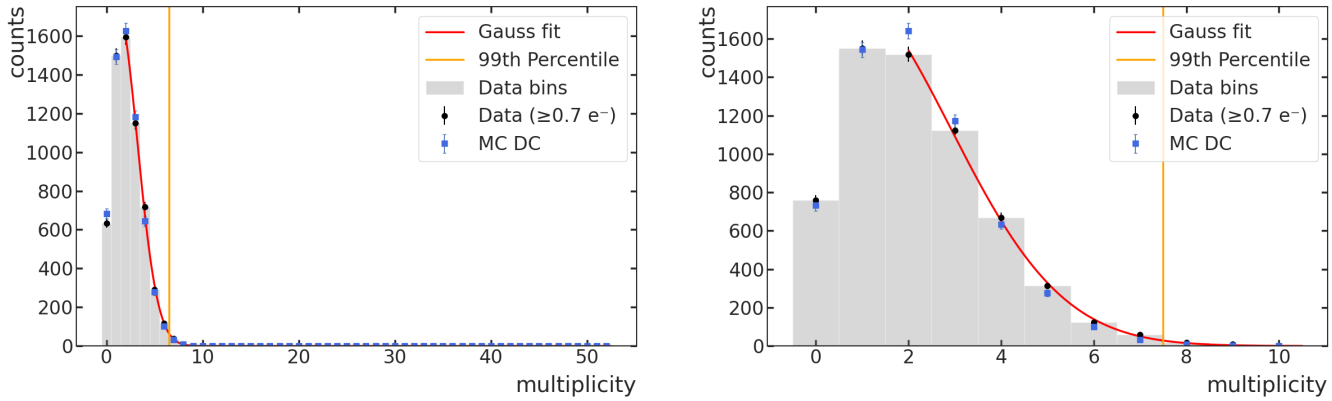


Figure 3.11: Row (left panel) and column (right panel) multiplicity histograms in CCD A. Gaussian fits (red) are applied to the upper tail. Orange lines indicate the 99th percentile thresholds.

In an ideal detector with uniform λ and independent pixels, the multiplicity distributions would follow a nearly Gaussian shape, and the data and MC would be indistinguishable. In CCD A, we observe good agreement between the data and the MC prediction up to a $m \sim 6$ (columns) and $m \sim 7$ (rows), confirming that the simulation correctly models the typical DC fluctuations. Beyond this range (above the orange line), the statistics become sparse and isolated outliers begin to dominate. These outliers are precisely the types of structures that the masking procedure is designed to suppress. Even in the high-multiplicity tail we still expect the points to fall along a bell-shaped (Gaussian) curve. The reason is the central-limit theorem: when you add up many independent pixel 'trials'-each pixel can either contribute a hit or not-the total number of hits tends to follow a Gaussian distribution once the number of pixels is large. Small departures from an ideal Gaussian remain visible and are expected for three reasons: (i) the per-column variation of $\lambda(x)$ broadens the distribution; (ii) rows and columns have different lengths, so their variances scale differently; and (iii) the statistics in the tail are limited, giving a coarse multiplicity spectrum.

Table 3.7 lists the number of rows and columns that were masked by criterion (2) in the UB0 sample (pixels with $Q \geq 2\,\mathrm{e}^-$). Across all CCDs, only a small number of rows were masked, and no columns were flagged. This is likely because other masking procedures-particularly those targeting column-related anomalies such as hot columns and crosstalk-have already removed most pathological columns earlier in the processing pipeline.

CCDs	**Masked Rows**	**Masked Columns**	**%**
A	2	0	0.033
B	1	0	0.016
C	0	0	0.000
D	1	0	0.016

Table 3.7: Masked rows and columns in unblinded data using Method 2 $Q \geq 2\,\mathrm{e}^-$ in the UB0 dataset.

Figure 3.12 complements the previous plot by displaying the full charged-pixel count per row/column in CCD A of UB0, together with the thresholds (red vertical lines).

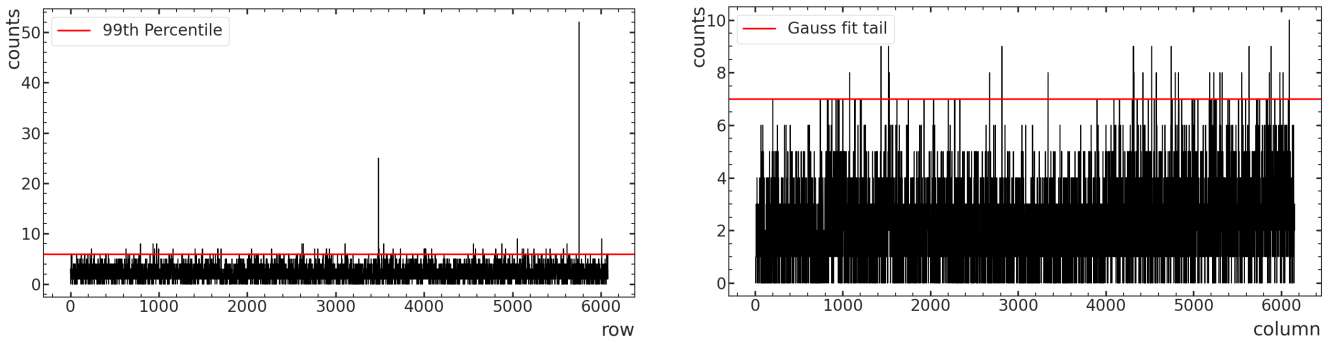


Figure 3.12: Charged pixel counts per (left panel) row and (right panel) column in CCD A. Red lines indicate thresholds.

3.3.7 Isolated Columns

Due to the masking procedures described previously, some columns may remain unmasked between two fully masked neighbors. To preserve spatial consistency and avoid fragmented masking patterns, an additional step is applied to identify and mask such isolated columns. A column is classified as isolated if both its immediate neighbors (left and right) are entirely masked, while the column itself is not. After applying all previous masking steps, a small number of isolated columns were identified and masked. This correction affected fewer than 0.1% of the pixels, ensuring minimal impact on data efficiency while improving the coherence of the final mask (see Table 3.8).

CCDs	Masked Isolated Columns	%
A	1	0.016
B	2	0.032
C	3	0.049
D	3	0.049

Table 3.8: Number of isolated columns masked in the UB0 dataset. The last column reports the fraction of masked pixels with respect to the total number of columns in the CCD.

3.3.8 Efficiency Summary of the Data Selection

After applying all masking procedures described in Sections 3.3.1 to 3.3.7, this section summarizes the cumulative impact of the selection process on UB0 dataset. Table 3.9 presents the resulting efficiency and number of masked pixels at each step, focusing on the active region (AR) of each CCD. Efficiency is calculated as:

$$\epsilon_{\text{acc}} = \frac{n_{\text{umskAR}}}{n_{\text{AR}}}, \tag{3.6}$$

where n_{umskAR} is the number of unmasked pixels in the active region after a given masking step, and n_{AR} is the total number of pixels in the active region of the CCD, 6080×6144, (see Section 3.1.1 for details).

As observed, the Clustering mask (Section 3.3.1) preserves nearly all pixels in CCDs A and B, while slightly affecting CCDs C and D. More pronounced effects are seen with Hot Columns (Section 3.3.3) and Overdensity (Section 3.3.4), particularly in CCDs B, C and D, reflecting vertical defects and high-charge accumulations identified in Figures 3.7,B.1 and 3.8, B.2.

Mask	CCD A		CCD B		CCD C		CCD D	
	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}
Clustering+AR	1.0000	383	1.0000	931	0.9998	3004	1.0000	1646
Cluster Crosstalk	0.9998	5964	0.9998	5964	0.9998	5964	0.9998	5964
Hot Columns	0.9998	0	0.9997	6080	0.9957	158080	0.9961	139840
Overdensity	0.9998	0	0.9616	1422720	0.8627	5046400	0.9271	2577920
CTI	0.9933	245962	0.9511	403697	0.8536	395221	0.9116	646881
Charge Multiplicity	0.9804	482156	0.9337	670296	0.8397	573059	0.8966	591294
Isolated Columns	0.9802	6080	0.9334	12160	0.8392	18240	0.8961	18240
Final	0.9802	737994	0.9334	2487142	0.8392	6007578	0.8961	3879813

Table 3.9: Cumulative efficiency and number of masked pixels at each step of the data selection process applied to the UB0 dataset. Values are reported for each CCD in the active region (6080 rows · 6144 columns). The final row shows the overall result after applying all masking criteria.

CCD C experiences the largest impact, consistent with the multiple hot-column groups listed in Table 3.3 and the substantial overdensity near the overscan region. Meanwhile, CCD A shows minimal disruption, suggesting stable operational conditions and uniformity. The CTI mask (Section 3.3.5) affects CCD D by addressing prominent charge-transfer inefficiencies (CTI), whereas CCD A maintains a higher efficiency level, implying fewer bright pixels from CTI. Further refinements by the Charge Multiplicity mask (Section 3.3.6) affect all CCDs similarly, identifying regions with elevated pixel multiplicity, while the Isolated Columns step (Section 3.3.7) finalizes the process by merging fragmented or single-column artifacts. CCD A retains the highest efficiency after all cuts, with 98.02% of its active pixels still usable-consistent with its lack of large-scale defects. CCD B follows at 93.34%; most of the loss stems from the wide OD band on the serial edge that had to be fully masked. CCD D finishes at 89.61%, although trailing from very hot columns required sizable masks, the sensor remains largely exploitable once charge-transfer and multiplicity artifacts are removed. CCD C is the most affected, ending at 83.92% because multiple hot-column clusters and the broad OD near the overscan region combine to mask several million pixels. Overall, the results confirm high data quality in all CCDs, with most masking attributed to real defects or instrumental effects.

Figure 3.13 provides a final overview of the masking procedure applied to UB0. For CCD A the left panel shows the two-dimensional occupancy map, where masked pixels appear in white. The right panel displays the corresponding Pixel Charge Distribution (PCD) after all masking steps. The corresponding plots for CCDs B, C, and D are presented in Appendix B.

3.4 Deviations from Poisson Statistics

After the full masking described in Section 3.3, the cleaned charge distributions from the unblinded dataset were analyzed to determine their compatibility with the expected dark current model. Each distribution was fitted with a convolution of a Poisson distribution (to model the nature of DC) and a Gaussian (to account for read-out noise):

$$f(x) = A \cdot \sum_{k=0}^{k_{\max}} \frac{\lambda^k e^{-\lambda}}{k!} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-k)^2}{2\sigma^2}}, \tag{3.7}$$

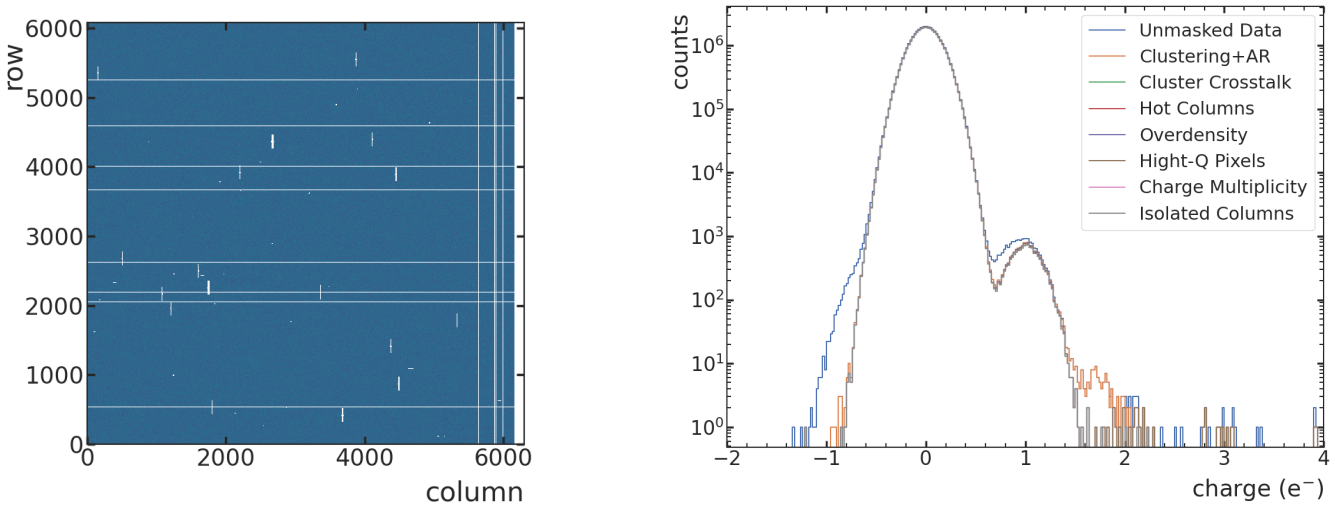


Figure 3.13: Left CCD map with masked pixels in white. Right: PCD for UB0 CCD A. The corresponding plots for CCDs B, C and D are provided in Appendix B, Figure B.5.

where λ is the average number of electrons per pixel, σ is the charge resolution obtained from the Gaussian fit, and A is a normalization factor. Figure 3.14 shows the results of the fits for CCDs A and C, and the equivalent plots for CCDs B and D are included in Appendix B, Figure B.6. Table 3.10 summarizes the extracted parameters. The best-fit values for the dark current after masking (λ) and their respective uncertainties obtained from these fits are expressed in units of electrons per binning per image (e^-/bin/img). To facilitate interpretation and comparison with other results, these values were converted into daily rates and normalized by detector mass.

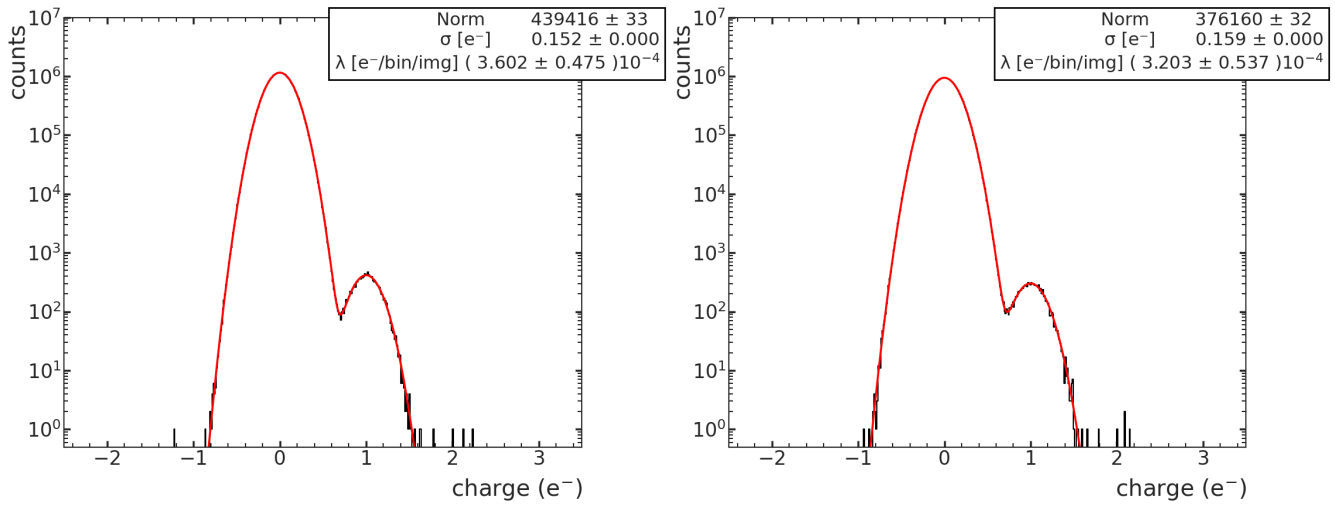


Figure 3.14: Dark current distributions from CCD A (left) and CCD C (right) in UB0. The black histograms represent the charge distributions after all masking procedures. The red fit corresponds to the convolution of Poisson and Gaussian distributions. The corresponding plots for CCD B and CCD D are shown in Appendix B, Figure B.6.

The CCD has a vertical binning of 100 pixels, so the dark current in terms of electrons per pixel

per image will be $\lambda_{e^-/\text{bin/img}} = \frac{\lambda_{e^-/\text{pix/img}}}{100}$. The read-out of each CCD image is performed in continuous acquisition mode, where the exposure time per pixel corresponds to the total read-out duration. This duration was computed from the metadata time using the difference between the start and end of the acquisition:

$$t_{\text{read-out}} = \texttt{DATE\_END} - \texttt{DATE\_INI} = 2200.627970 - 532.347327 = 1668.280643 \text{ s},$$

which corresponds to approximately 0.0193 days. Thus, the average DC per day is obtained as $\lambda_{\text{day}} = \frac{\lambda_{pix}}{0.0193}$. To express the rate in units of electrons per gram per day (e^-/g/day), the mass of a single CCD pixel was estimated. Using the measurements for an LBC CCD described in Section 1.3, with a pixel size of 15×15 μm^2 and a thickness of 670 μm, the volume of one pixel is:

$$V_{\text{pix}} = (15 \cdot 10^{-4} \text{ cm})^2 \cdot 670 \cdot 10^{-4} \text{ cm} = 1.5075 \cdot 10^{-7} \text{ cm}^3.$$

Assuming the density of silicon to be $\rho = 2.33$ g/cm^3, the corresponding pixel mass is:

$$m_{\text{pix}} = V_{\text{pix}} \cdot \rho = 3.51 \cdot 10^{-7} \text{ g}.$$

The dark current rate per gram per day is then given by $\lambda_{\text{g/day}} = \frac{\lambda_{\text{day}}}{m_{\text{pix}}}$. These derived quantities are summarized for each CCD in Table 3.10.

CCDs	**e^-/bin/img**	**e^-/pix/day**	**e^-/g/day**	σ
A	$(3.6 \pm 0.5) \cdot 10^{-4}$	$(1.9 \pm 0.2) \cdot 10^{-4}$	531 ± 70	0.152 ± 0.000
B	$(3.0 \pm 0.6) \cdot 10^{-4}$	$(1.6 \pm 0.3) \cdot 10^{-4}$	443 ± 90	0.156 ± 0.000
C	$(3.2 \pm 0.5) \cdot 10^{-4}$	$(1.7 \pm 0.3) \cdot 10^{-4}$	472 ± 80	0.159 ± 0.000
D	$(4.0 \pm 0.6) \cdot 10^{-4}$	$(2.1 \pm 0.3) \cdot 10^{-4}$	591 ± 90	0.163 ± 0.000

Table 3.10: Extracted parameters from Poisson-Gaussian fits for each CCD, including statistical uncertainties. The values correspond to the average charge per pixel per image (e^-/bin/img), per day (e^-/pix/day), and normalized per gram per day (e^-/g/day), along with the charge resolution parameter σ.

3.4.1 Excess over Poisson Statistics

Although the Poisson-Gaussian model in Figure 3.14 and B.6 is tuned to reproduce the $0\,e^-$ pedestal and the $1\,e^-$ peak, it consistently under-predicts the 2 and 3 e^- bins in every CCD. This deficit is hard to spot on the logarithmic scale of the figure, but becomes evident when the number of isolated single-pixel events is integrated and compared with the pure-Poisson expectation $P(k;\lambda) = \lambda^k e^{-\lambda}/k!$ obtained from the fitted mean λ. Table 3.11 summarizes the result.

Three effects can raise the probability of observing two or more electrons in the same pixel. First, burst-like dark current traps occasionally release several electrons at once, violating the assumption of independent arrivals. Second, a few-electron halo left by bright clusters, CTI tails or blooming can escape the masking cuts and appear as isolated multi-electron pixels. Third, genuine multi-electron deposits from low-energy X-rays, Compton electrons, bremsstrahlung photons-or even sub-GeV dark-matter scatters-can populate the same charge range. In fact, recent Skipper-CCD studies (Barak et al. 2022) show that even after removing bulk dark current, two additional processes-amplifier-emitted photons during read-out and spurious charge from clock transitions-produce residual single-electron events whose stochastic combination can boost the 2-3 e^- bins above a pure Poisson expectation.

q / CCDs	1 e$^-$		2 e$^-$		3 e$^-$	
	Exp $[10^4]$	Obs $[10^4]$	Exp	Obs	Exp $[10^{-4}]$	Obs
A	1.28 ± 0.17	1.231 ± 0.011	2.3 ± 0.6	4 ± 2	2.8 ± 1.1	0 ± 0
B	1.04 ± 0.20	0.949 ± 0.010	1.6 ± 0.6	4 ± 2	1.6 ± 1.0	0 ± 0
C	0.98 ± 0.16	0.942 ± 0.010	1.6 ± 0.5	5 ± 2	1.7 ± 0.9	0 ± 0
D	1.34 ± 0.19	1.260 ± 0.011	2.8 ± 0.8	4 ± 2	3.6 ± 1.6	0 ± 0

Table 3.11: Isolated single-pixel counts after masking for UB0. "Exp" values stem from the Poisson mean fitted to the 1 e^- peak; "Obs" are the data. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. All CCDs exhibit a significant excess at 2 e^-.

$$\mu = \lambda_{\mathrm{EXP}}\, t_{\mathrm{EXP}} + \lambda_{\mathrm{RO}}\, t_{\mathrm{RO}} + \mu_{\mathrm{SC}},$$

quantifies these three contributions (thermal, read-out, and clocking), and naturally explains the observed excess in 2-3e$^-$ even after masking the bulk Poisson DC [28]. EXP refers to the exposure duration, RO to the read-out phase, and SC to the spurious charge contribution.

The observed excess therefore motivates an alternative approach to improve the sensitivity of our dark matter analysis. We introduce a method-referred to as pattern analysis-where the background model exhibits better agreement with the data by identifying random coincidences of single-electron events. This strategy will be described extensively in the next chapter.

Chapter 4

Pattern-base Observable for Dark Matter Detection

Building on the deviations highlighted in Section 3.4.1, this chapter develops a systematic method to identify and characterize the multi-pixel charge patterns that survive the masking. A pattern is defined as a group of two or three adjacent pixels —only patterns found in rows are considered for DM searches because the images are binned over 100 physical rows— in which every pixel carries non-zero charge and the total charge lies between 2 and $4\,e^-$. While isolated pixels with charges $\geq 2\,e^-$ already show discrepancies from Poissonian expectations, it is essential to assess whether such multi-pixel patterns also exhibit inconsistencies with statistical predictions.

The dark matter search presented in this work is conducted through the identification of these low-charge patterns. Specifically, we consider the ordered sets $\{mn\}$ and $\{mnl\}$, where the number corresponds to the pixel charge in e^- of each pixel. When permutations are taken into account, this yields $\{11\}, \{12\}, \{13\}, \{21\}, \{31\}$ for two-pixel patterns and $\{111\}, \{211\}, \{121\}, \{112\}$ for three-pixel patterns.

Patterns with total charge exceeding $4, e^-$ can still appear, as the clustering algorithm only discards patterns with total charge above $10\,e^-$. However, for the analysis presented here, the energy threshold is conservatively reduced to $6\,e^-$ to account for the open shielding conditions. Patterns in the intermediate range $4 < Q \leq 6\,e^-$ are not assigned to specific combinations; they are grouped into a single high-charge category that is treated identically in data and in the dark-matter signal simulations, ensuring an unbiased analysis.

The following sections detail the methodology for selecting, isolating, and evaluating these patterns, providing the foundation for the subsequent dark-matter analysis.

4.1 Score metrics for Pattern Classification

After masking, the analysis searches for charge patterns formed by one, two, or three neighboring pixels. Charge patterns in the CCD are identified by analyzing small groups of neighboring pixels and evaluating their compatibility probabilistically. The method models the likelihood of a measured charge being produced by a given electron deposition using a Gaussian cumulative distribution function (CDF):

$$\mathrm{CDF}(q \,|\, m, \sigma) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{q-m}{\sigma\sqrt{2}}\right)\right), \tag{4.1}$$

where q is the measured pixel charge, m is the expected charge $m = 0, 1, \ldots, 4$, and σ[1] is the charge resolution. A small CDF value means q is unlikely to originate from m electrons. For a pattern that spans k pixels, the probability is given by:

$$\Lambda_{m,n,\ldots} \;=\; -\sum_{i=1}^{k} \log\big[\mathrm{CDF}(q_i \,|\, m_i, \sigma)\big], \tag{4.2}$$

so that Λ_m, $\Lambda_{m,n}$ and $\Lambda_{m,n,l}$ are the one-, two- and three-pixel scores that enter the cascade. For convenience we work with the negative logarithm of the CDF, so that good candidates correspond to

[1] σ is taken from the dark-current fits of Figures 3.14 and B.6 and depends on CCD and amplifier.

small values of Λ. To determine the most probable charge pattern m, n, l, the algorithm evaluates multiple charge combinations across one-, two-, and three-pixel configurations, as summarized in Table 4.1. Within each configuration, patterns are tested sequentially, starting from the smallest total charge. Candidate patterns are evaluated using a cascade approach that sequentially applies thresholds tailored to each pattern group size.

The pattern selection proceeds by testing all one-pixel patterns first, followed by two-pixel patterns, and finally three-pixel patterns. Each configuration is fully evaluated before moving to the next. If multiple patterns satisfy the threshold criteria, only the last valid pattern (i.e., the most extended structure) is retained. This strategy favors selecting the largest consistent structure while minimizing false identifications. Although the dark-matter analysis ultimately uses only horizontal patterns because the images are binned over 100 physical rows, the search itself is performed along rows, columns and both main diagonals.

The three charge thresholds are set to $\mathrm{thr}_1 = 3.5$, $\mathrm{thr}_2 = 4.0$, and $\mathrm{thr}_3 = 5.5$. These classification thresholds were chosen based on simulation results to strike a balance between efficiency and misclassification. Their optimization procedure—and the resulting efficiencies and false-rejection rates—is discussed in detail in Section 4.2.

Single pixel (m)	**Pattern** (m, n)	**Pattern** (m, n, l)
$\Lambda_1 < \mathrm{thr}_1$	$\Lambda_{1,1} < \mathrm{thr}_2$	$\Lambda_{1,1,1} < \mathrm{thr}_3$
$\Lambda_2 < \mathrm{thr}_1$	$\min\{\Lambda_{1,2}, \Lambda_{2,1}\} < \mathrm{thr}_2$	$\min\{\Lambda_{1,1,2}, \Lambda_{1,2,1}, \Lambda_{2,1,1}\} < \mathrm{thr}_3$
$\Lambda_3 < \mathrm{thr}_1$	$\min\{\Lambda_{1,3}, \Lambda_{2,2}, \Lambda_{3,1}\} < \mathrm{thr}_2$	$\min\{\Lambda_{1,2,2}, \Lambda_{2,1,2}, \Lambda_{2,2,1}\} < \mathrm{thr}_3$
	$\min\{\Lambda_{2,3}, \Lambda_{3,2}\} < \mathrm{thr}_2$	

Table 4.1: Cascade logic for one-, two- and three-pixel patterns ($m, n, l = 1, 2, 3$ electrons). Only the lowest Λ in each group is compared to the corresponding threshold.

4.2 Selection Performance: Efficiency and Misclassification

The performance of the pattern-selection cascade was quantified with dedicated Monte Carlo samples in which every allowed charge pattern was injected 10^6 times and Gaussian readout noise was added ($\sigma = 0.16\, e^-$). Each replica was classified through the algorithm described in Section 4.1 and compared to its original label. To reproduce the expected physical population every event was weighted by its Poisson probability, and the resulting ensemble was used to build a confusion matrix $P_{(m,n,l)\to(p,q,r)}$ that counts how often a true pattern (m, n, l) is finally classified as (p, q, r). The efficiency, $\epsilon_{(m,n,l)}$, is the fraction of genuine (m, n, l) patterns that are correctly identified:

$$\epsilon_{(m,n,l)} = \frac{P_{(m,n,l)\to(m,n,l)}}{\sum_{(p,q,r)} P_{(m,n,l)\to(p,q,r)}}, \tag{4.3}$$

where $P_{(m,n,l)\to(m,n,l)}$ is the number of correctly identified patterns, and the denominator corresponds to the total number of events whose true pattern is (m, n, l). The misidentification rate (MisID) rate quantifies how often a detected pattern (p, q, r) actually originates from a different true pattern (m, n, l). It is defined as:

$$\mathrm{MisID}_{(m,n,l)} = 1 - \frac{P_{(m,n,l)\to(m,n,l)}}{\sum_{(p,q,r)} P_{(p,q,r)\to(m,n,l)}}, \tag{4.4}$$

where $P_{(m,n,l)\to(m,n,l)}$ represents the number of correctly identified (m, n, l) patterns, and the denominator accounts for all patterns detected as (m, n, l), regardless of their true origin. Large efficiency and a small

MisID therefore indicate an optimal operating point.

Figure 4.1 compares the efficiency and the mis-identification probability for one-, two-, and three-pixel patterns, showing how both metrics vary as the score threshold Λ_m, $\Lambda_{m,n}$, or $\Lambda_{m,n,l}$ —defined in Equation 4.2 — is adjusted. The efficiency rises rapidly as the score threshold moves above the noise tail, stabilizing at around 98% for single-pixel and 91% for multi-pixel patterns. Mis-identification begins at a few percent due to overlaps between charge patterns and increases steadily as the thresholds is lowered, reaching 100% when almost all candidates are accepted. Thresholds values are determined by selecting points on the efficiency plateau where MisID remains at the acceptable level. Optimizing the figure of merit $S = \epsilon - \text{MisID}$ under the constraint $\epsilon \gtrsim 90\,\%$ yields the thresholds $\text{thr}_1 = 3.5$, $\text{thr}_2 = 4.0$, and $\text{thr}_3 = 5.5$. At these thresholds, the efficiency reaches about 98% for single-pixel patterns and 91% for both two- and three-pixel patterns. These thresholds ensure the preservation of the vast majority of genuine patterns while limiting contamination to approximately 11% for two-pixel and 18% for three-pixel patterns. Finally, the confusion matrix $P_{(m,n,l)\to(p,q,r)}$ is shown in Figure 4.2 at a representative noise level of $\sigma = 0.16\text{e}^-$. Within the noise range of 0.15-0.17 e^- observed during data acquisition, the matrix remains essentially stable, with only minor degradation in the two- and three-pixel channels.

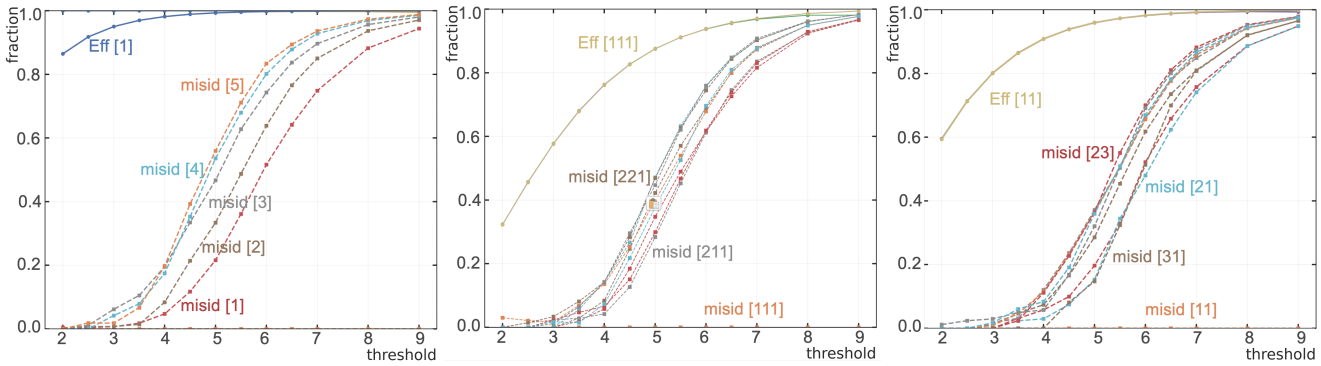


Figure 4.1: Efficiency (solid) and mis-identification (dashed) as a function of the score threshold Λ. Left: single-pixel candidates Λ_m; Center: two-pixel $\Lambda_{m,n}$; Right: three-pixel $\Lambda_{m,n,l}$.

4.2.1 Pattern Isolation

A detected pattern is considered valid only if it is isolated from other charge deposits. For a single–pixel pattern this means that the eight neighboring pixels (horizontal, vertical and diagonal) must exhibit charge values consistent with zero electron. This condition is evaluated using the one electron CDF demanding that Λ_1 and requires $\Lambda_1 > \text{thr}_1$. For larger patterns the isolation region is defined as every pixel that shares an edge or a corner with any pixel belonging to the pattern, excluding the pattern pixels themselves.

$$\Lambda_1 = -\log \text{CDF}(q \,|\, m = 1, \sigma)) > \text{thr}_1, \tag{4.5}$$

where q is the measured charge, $m = 1$ is the one-electron hypothesis, σ is the charge resolution from Table 3.10, and $\text{thr}_1 = 3.5$ is the same single-pixel threshold used in the selection cascade (Section 4.1). The pattern is kept only if *all* surrounding pixels pass the test; if a neighbor fails the inequality in Equation (4.5), its charge is not compatible with zero and the entire pattern is discarded.

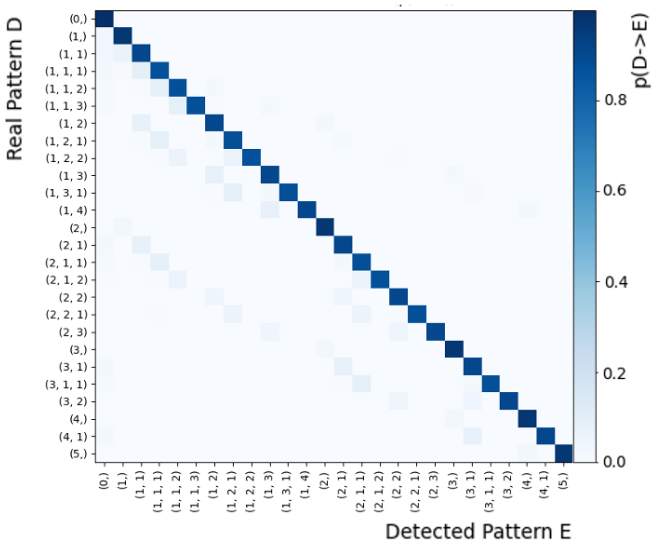


Figure 4.2: Confusion matrix for a readout noise of $0.16e^{-}$. Color intensity indicates the transition probability $P_{(m,n,l)\to(p,q,r)}$.

4.3 Validation of the Classification Methodology

The thresholds introduced in Section 4.1 were validated by analyzing the score distributions of representative patterns obtained from dedicated simulations, ensuring properly isolation of the intended configurations.

Figure 4.3 compares the first–step score Λ for the three candidate types examined at the start of the cascade: single pixels tested against the one-electron hypothesis (1) (black), doublets against $(1,1)$ (blue), and triplets against $(1,1,1)$ (magenta). The red dashed lines indicate the threshold cuts $\{\mathrm{thr}_1, \mathrm{thr}_2, \mathrm{thr}_3\}$. A low Λ value signals good compatibility with the hypothesis, while high values point to poor matches. As expected, most entries lie well below the corresponding thresholds, reflecting that all simulated patterns were generated with at least one full electron of charge, to which only readout noise was added. The few events exceeding the cuts result from noise shifting one or more pixel charges so low that the pattern no longer fits. In multi-pixel patterns, this may happen when several pixels are simultaneously affected by downward fluctuations. While moderate shifts (e.g., 0.9, 0.9) can still pass, stronger losses (e.g., 0.6, 0.8) may lead to rejection. These misidentifications are rare and do not significantly impact the overall performance, which remains highly efficient for all pattern types, as detailed in Section 4.2.

To confirm the clustering of charge combinations in multi-pixel patterns, 2D projections were studied. Figure 4.4 displays representative charge combinations and their correlations for $(1,1)$ and $(1,1,1)$ patterns, highlighting the region isolated by the selection. Green dashed line represents $0.6\,e^{-}$ charge. The left panel plots the $(q[i], q[i+1])$ pairs selected as $(1,1)$ doublets. Most events cluster tightly around $(1\,e^{-}, 1\,e^{-})$. The red dashed lines at $0.7\,e^{-}$ mark the single-pixel acceptance threshold to consider a pixel with charge. Charges up to $\simeq 2.3\,e^{-}$ on one axis are therefore still compatible with the $(1,1)$ hypothesis provided that the partner pixel sits very close to the $0.7\,e^{-}$ cut. The center panel shows the $(1,1,1)$ triplets, coloring each point by the charge of the third pixel. A color island centered at $(1,1)\,e^{-}$ shows that the combination of the other charge is around $\approx 1.8\,e^{-}$. Only a few peripheral points extend beyond the red selection lines and to almost none to the green lines, illustrating the efficiency of the threshold impose. The right panel presents a smoothed density map of the same $(1,1)$ sample. The high-density core is clearly visible, while the population drops steeply as soon as one of the charges approaches the

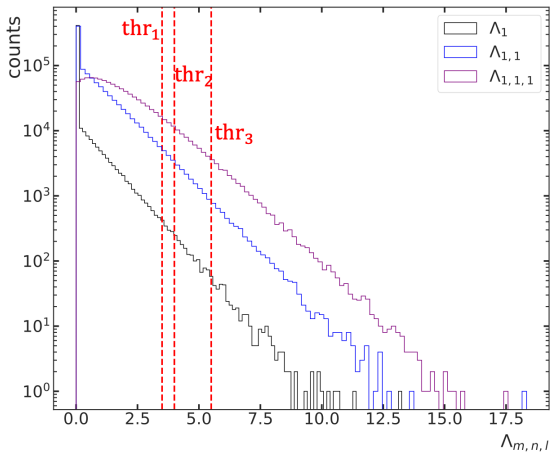


Figure 4.3: First–stage score ($\Lambda_{m,n,l}$) distributions for one-pixel (black), two-pixel (blue) and three-pixel (magenta) candidates, tested against the (1), (1, 1) and (1, 1, 1) hypotheses, respectively. The red dashed lines mark the working thresholds $\{\mathrm{thr}_1, \mathrm{thr}_2, \mathrm{thr}_3\}$; everything to its right is rejected.

$0.7\,e^-$ boundary. Additional visualizations for all studied patterns are provided in Appendix C.

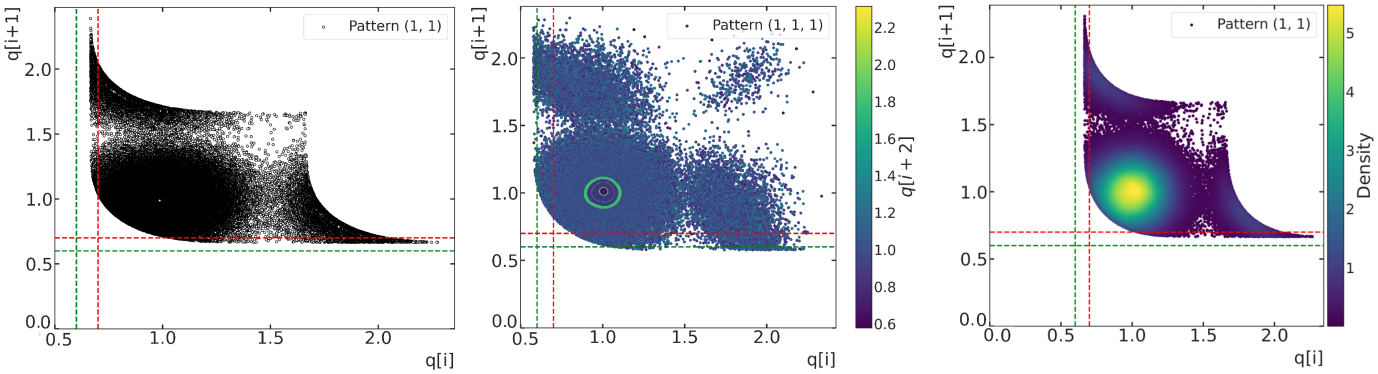


Figure 4.4: Charge-correlation plots. Left: raw $(q[i], q[i{+}1])$ pairs for the $(1, 1)$ hypothesis; Center: triplet $(1, 1, 1)$ colored by the third pixel; Right: density map of the $(1, 1)$ sample. Green dashed lines delimit the $1\,e^-$ core and the red lines indicate the selection borders.

Additional pixel-charge histograms illustrating the same behavior are shown in Appendix C.2.3 (Figure C.4).

4.4 Diffusion-Induced Pattern Formation Probabilities

Up to this point the study has evaluated the classification efficiency of pre-defined patterns, assuming they already exist in the image. That description is sufficient for dark-current events, whose signal is created at the CCD surface and recorded pixel-by-pixel. For ionizing interactions, however, charge is produced at arbitrary positions inside the $675\,\mu$m-thick bulk and drifts to the pixel array while diffusing laterally. In this section, the focus shifts to evaluating the pattern formation efficiency resulting from diffusion, which plays a crucial role in shaping the observable patterns of ionizing signals, such as those expected from dark matter interactions.

The lateral charge spread follows the depth-dependent model of Equation(1.3) in Section1.3.3. MC simulations were used to model this process through the diffusion probability $p_{\mathrm{diff}}(q \to p)$, which quantifies

the likelihood that an initial charge deposit of q electrons evolves into a final pattern p due to lateral diffusion. Using the work [17], the measured creation probabilities $P(n_{eh} \mid E)$ are sampled to obtain an integer number ($q = 1$–5) of electron-hole pair created given an energy lost E. Then, the charge cloud is placed at a random (x, y, z) inside the CCD and diffused according to:

$$\sigma_{xy}^2(z, E) = -A \log|1 - bz| \cdot (\alpha + \beta E)^2 \,, \tag{4.6}$$

using the following values:

$$A = 803.5\ \mu\text{m}^2, \qquad b = 6.5 \cdot 10^{-4}\ \mu\text{m}^{-1},$$
$$\alpha = 0.83, \qquad \beta = 0.0112\ \text{eV}^{-1}.$$

calibrated with cosmic-ray tracks at $V_{\text{sub}} = 45$ V. Following diffusion, the resulting charge distributions were vertically binned (100 pixels per bin), Gaussian readout noise and the same pattern detection algorithm used for data analysis (see Sections 4.1) was applied to the simulated charge matrices.

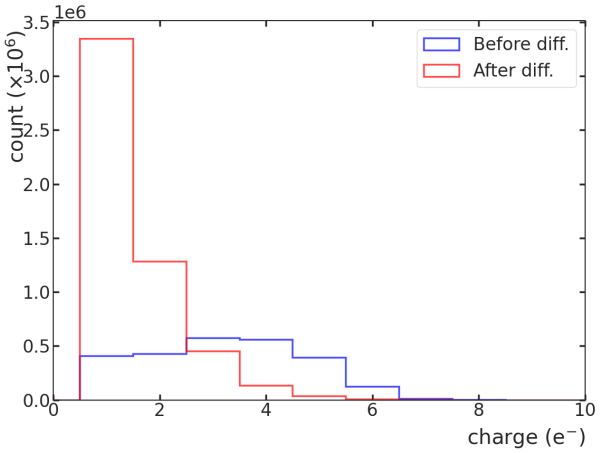


Figure 4.5: Simulated distribution of pixel charges before (blue) and after (red) applying charge diffusion. A Gaussian noise with $\sigma = 0.16$ was included after diffusion. Binning: 100 pixels per bin.

Figure 4.5 shows the simulated distribution of pixel charges before and after applying diffusion. Before diffusion, the charge is localized, resulting in a broader distribution of higher charge values per pixel. After diffusion, charge is spread over multiple pixels, leading to a concentration of counts at lower pixel charges. Using this simulation setup, diffusion probabilities $p_{\text{diff}}(q \to p)$ were extracted by counting how often each true deposited charge q resulted in a specific detected pattern p. Table 4.2 shows the most relevant results, corresponding to pattern types also analyzed in real data. These simulations confirm that multi-pixel structures naturally arise from the diffusion of low-charge deposits. For instance, a charge of $q = 2e^-$ leads to the appearance of a two-pixel pattern ($\{11\}$) in approximately 38% of cases. Additionally, patterns where both electrons remain confined to the same pixel ($\{2\}$) or are split as isolated single electrons ($\{1\}$) occur with relative frequencies of 35% and 46%, respectively. Because several patterns can arise from one event, probabilities need not sum to unity. Similarly, for $q = 3e^-$, the combined frequency of patterns such as $\{11\}$, $\{21\}$, and $\{111\}$ exceeds 45%, highlighting the relevance of multi-pixel detection in low-energy event reconstruction.

4.5 Expected Pattern Rates under a Poissonian Background

In the context of this DM search, the dedicated masking procedures described in the previous chapter (Chapter 3) have been designed to suppress any population of low-multiplicity (< 3) events that cannot

q \ p	{1}	{2}	{3}	{11}	{12}	{111}	{13}	{211}
2	0.462	0.351	0.000	0.380	0.000	0.000	0.000	0.000
3	0.311	0.143	0.163	0.133	0.187	0.138	0.000	0.000
4	0.250	0.090	0.144	0.061	0.063	0.043	0.097	0.062

Table 4.2: Diffusion probabilities $p_{\text{diff}}(q \to p)$ obtained from $5 \cdot 10^4$ simulated events ($\sigma_{\text{read}} = 0.16\, e^-$). q is the number of initial electrons and p is the detected pattern.

be attributed to dark current. As a result, after these cuts the residual background in the low-charge region should originate exclusively from uncorrelated, purely random charge depositions, which are well modeled by a Poisson process. Under this assumption, this section focuses on estimating the expected occurrence rates of each pattern providing a reference distribution against which to compare the observed data.

The expected number of observed patterns (p, q, r) is calculated by summing over all true patterns (m, n, l), weighted by their Poisson occurrence probability and the corresponding confusion matrix elements:

$$N_{\text{exp}}(p, q, r) = N_{\text{selec}} \cdot \sum_{(m,n,l)} \left[p_{\text{dc}}(m, n, l; \lambda) \cdot P_{(m,n,l)\to(p,q,r)} \right], \tag{4.7}$$

where N_{exp} is the expected number of patterns that will be classified as (p, q, r), N_{selec} the total number of pixels that survive the masking, $p_{\text{dc}}(m, n, l; \lambda)$ is the Probability that the dark-current model produces exactly the set of integer charges (m, n, l) in the same exposure, and $P_{(m,n,l)\to(p,q,r)}$ is the corresponding transition probability. Because DC electrons are produced independently in each pixel, the probability of a given pattern is modeled as the product of independent pixel contributions:

$$p_{\text{dc}}(m, n, l; \lambda) = \prod_{k_i \in (m,n,l)} \text{Poisson}(k_i; \lambda), \tag{4.8}$$

where k_i are the integer charge values within the pattern and λ is the average number of electrons per pixel obtained from the fit in Table 3.10. For example, the probability of a genuine {11} doublet is $\text{Poisson}(1; \lambda)^2$

The uncertainties on the expected pattern counts combine two contributions. The dominant term arises from the uncertainty in the Poisson mean λ, extracted from the Poisson+Gaussian fit (Figures 3.14 and B.6) to the single-pixel charge distribution. This contribution is estimated using a symmetric variation around the best-fit value:

$$\sigma_\lambda(E) = \tfrac{1}{2} \left| N_{\text{exp}}(\lambda + \delta\lambda) - N_{\text{exp}}(\lambda - \delta\lambda) \right|. \tag{4.9}$$

A second, smaller contribution comes from the statistical uncertainty in the transition probabilities $P_{(m,n,l)\to(p,q,r)}$, derived from MC simulations. This term is propagated as:

$$\sigma^2_{\text{stat}}(p, q, r) = \sum_{(m,n,l)} \left[N_{\text{selec}}\, p_{\text{dc}}(m, n, l; \lambda)\, \delta P_{(m,n,l)\to(p,q,r)} \right]^2, \tag{4.10}$$

where $\delta P_{(m,n,l)\to(p,q,r)}$ is the MC error on each matrix element. The total uncertainty on the expected number of patterns is then given by:

$$\sigma_N(E) = \sqrt{\sigma^2_\lambda(E) + \sigma^2_{\text{stat}}(E)}. \tag{4.11}$$

No additional uncertainty is assigned to the number of selected pixels N_{selec}, to the electronic noise width σ_e, which is well constrained by the fit, or to the pattern selection thresholds, which are kept fixed across all analyses.

4.6 Do the Observed Patterns in UB0 Follow Poissonian Expectations?

In this section, the pattern classification methodology developed throughout this chapter is applied to the UB0 dataset in order to quantify the observed pattern rates and assess their compatibility with the expectations from a purely Poissonian background.

For each CCD we compare the observed pattern counts (Obs) with the Poisson expectations (Exp) derived from Equation 4.7; the results are summarized in Tables 3.11 and 4.3. Numbers without square brackets refer to patterns oriented along rows, which are the only candidates evaluated in this analysis. Values in square brackets give the total over all directions. All four CCDs (A, B, C, and D) show similar results supporting the reliability of the data selection and pattern identification methods.

CCDs \ **p**		**{11}** [10^0]	**{12}** [10^{-3}]	**{13}** [10^{-7}]	**{111}** [10^{-3}]	**{112}** [10^{-7}]
A	Exp	4.3 ± 1.1	0.8 ± 0.3	0.9 ± 0.5	1.6 ± 0.6	2.9 ± 1.5
	Obs	$3.0 \pm 1.7[15]$	0	0	0	0
B	Exp	2.9 ± 1.2	0.5 ± 0.2	0.4 ± 0.3	0.9 ± 0.5	1.5 ± 1.2
	Obs	$4 \pm 2[14]$	0	0	0	0
C	Exp	3.0 ± 1.0	0.5 ± 0.2	0.5 ± 0.3	1.0 ± 0.5	1.7 ± 1.1
	Obs	$1 \pm 1[6]$	0	0	0	0
D	Exp	5.1 ± 1.5	1.0 ± 0.5	1.4 ± 0.8	2.2 ± 1.0	4.6 ± 2.7
	Obs	$9 \pm 3[22]$	0	0	0	0
Total	Exp	15.4 ± 2.4	2.7 ± 0.7	3.3 ± 1.1	5.7 ± 1.4	11 ± 4
	Obs	$17 \pm 4[57]$	0	0	0	0

Table 4.3: Expected (Exp) and observed (Obs) counts for multi-pixel patterns in each CCD of the UB0 dataset. Results are shown in the format $x \pm \delta x\,[N]$, where $x \pm \delta x$ refers to the expected or measured counts in rows, and N is the total number of patterns detected over all directions. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. Each pattern symbol implicitly includes all charge permutations of the same topology.

After masking, the vast majority of pixels appear singly, so the only multi-pixel topology with non-zero statistics is the nearest-neighbor pair {11}. For every CCD the measured row-oriented {11} count lies close to the Poisson prediction. Combining the four sensors, the row count is 17 ± 4 against a prediction of 15.4 ± 2.4, a difference of only $0.4\,\sigma_{exp}$. All higher-order shapes ({12}, {13}, {111}, {112} and permutations) are completely absent, exactly as one would expect in a sample of this size and looking at the Poisson expectation.

Because the masking and pattern filters behave as expected on UB0, it demonstrates that the masking and pattern filters effectively suppress instrumental artifacts. The same pipeline is therefore applied unchanged to the four blinded periods analyzed in Chapter 5; any departure from the baseline established here will hence be readily identifiable.

Chapter 5

Event Selection on Blinded Data

The complete data-selection and pattern-identification pipeline defined in Chapters 3 and 4 is now applied without any modification to the four blinded data taking periods B1–B4 listed in Table 3.1.

5.1 Final Pixel Charge distribution

This section (i) compares the masked structures observed in each run, (ii) reviews the resulting masking efficiencies, and (iii) presents the residual pixel-charge spectra after all selections.

5.1.1 Consistency of Masked Features Across Datasets

The unblinded reference set (UB0) already revealed two persistent defects: column 1807 in CCD B and the block [5315, 5318] in CCD C, together with a broader band in C and an exponential tail in D (Table 3.3).

Hot-column masks in the blinded datasets show the same picture with minor run-to-run variations. In CCD A no hot column persists through all periods: columns 1532 and 2350 appear only in B1, whereas the block [1531, 1533] and column 6092 recur in B2–B4 (with [3764, 3765] present in B2–B4), while isolated columns 796 and 1018 appear only in B3 and 979 and 1017 only in B4. CCD B is more stable: column 1807 is masked from B1 to B3, the block [4206, 4207] reappears in B2–B3 and 4206 in B4; column 1520 shows up in B1 and B4 but not in B2–B3. CCD C exhibits the most robust defect: the cluster [5315, 5318] is present in every run, accompanied by columns 86 and 1989 throughout often flanked by activity near ~ 6100. CCD D retains an exponential-like regions at [323, 360] in all periods (with slight endpoint shifts) and [3800, 3810] in B2-B4. Thus, the dominant structures first seen in UB0—[5315, 5318] in C and 1807 in B—remain evident in all blinded runs, whereas lower-intensity defects fluctuate with time, as summarized in Table 5.1.

	CCD A	**CCD B**	**CCD C**	**CCD D**
B1	1532, 2350	1520, 1807, 5313	86, 1989, [3355, 3356], [5315, 5318], [6141, 6151]	[329, 346], 4328
B2	[1531, 1533], [3764, 3765], 6092	1807, 3929, 3938, [4206, 4207], 5967	86, 1989, [5315, 5318], [6093, 6098], [6140, 6151]	[324, 360], [3798, 3810]
B3	796, 1018, [1531, 1533], [3764, 3765], 6092	[8, 9], 1807, [3930, 3941], [4205, 4207]	86, 1989, [5315, 5318], 5571, 5918, 6085, [6090, 6103], 6135, [6142, 6151]	[323, 357], [3800, 3809]
B4	979, 1017, [3764, 3765], 6092	1520, 4206	8, 86, 1989, [5315, 5318], 6086, 6094, [6097, 6109], 6126, [6131, 6151]	[325, 357], 1169, 1171, 2144, [3798, 3810], [4328, 4333], 6146, 6150

Table 5.1: Hot columns detected in each blinded dataset. Columns 1807 (CCD B) and 5315–5326 (CCD C) persist through all runs, matching the UB0 baseline.

UB0 already revealed a broad overdensity in CCD C [5311–6140], a narrower edge band in CCD B [5823–6056] and two tails in CCD D [5657–5836], [5908–6160], while CCD A showed none (Table 3.4).

The four blinded datasets confirm and extend this picture. In CCD A every run displays a shallow parabolic excess around column 950, with widths of [820–1031] (B1), [789–1101] (B2), [792–1074] (B3) and [791–1077] (B4). In CCD B the serial-edge overdensity appears as follows: no mask is applied in B1; in B2 two bands at [5591, 5898] and [5905, 6057] are flagged; in B3 the band spans [5703, 6058]; and in B4 it covers [5727, 6059]. In CCD C a diffuse excess starts near 5310 and reaches the image edge in all periods, exactly reproducing the UB0 overdensity in both position and shape. CCD D remains essentially clean: only B1 masks at [297, 451] and [5143, 5307], while no overdensity cut is required in B2–B4.

Table 5.2 summarizes the overdensity bands identified in each blinded run; therefore, the algorithm removes the same low-level charge gradients seen in UB0, with only minor adjustments.

	CCD A	**CCD B**	**CCD C**	**CCD D**
B1	[820, 1031]	none	[5311, 6149]	[297, 451], [5143, 5307]
B2	[789, 1101]	[5591, 5898], [5905, 6057]	[5313, 6096]	none
B3	[792, 1074]	[5703, 6058]	[5310, 6150]	none
B4	[791, 1077]	[5727, 6059]	[5298, 6139]	none

Table 5.2: Overdensity regions detected in the blinded datasets B1–B4. The broad band in CCD C and the serial-edge excess in CCD B reproduce the UB0 structures listed in Table 3.4.

Layout-independent masks —clustering in the active region and crosstalk suppression— remove practically the same fraction of pixels in every CCD because they target transient, uniformly distributed events rather than fixed silicon defects. Clustering+AR excludes only 0.01−0.02 % of the active area in all runs, whereas the crosstalk mask rejects 0.04 % in B1 and increases to at most 0.10 % in B2–B4. These fractions vary by less than a factor of two from B1 to B4 and are identical across the four CCDs within each dataset. Compared with the percent-level losses produced by the hot-column and overdensity masks, the impact of the layout-independent cuts is negligible (see Tables D.1–D.7).

Complete lists of hot columns, overdensity bands, and masked-pixel maps are provided in Appendix D.1 (see Figs. D.3–D.20).

5.1.2 Final Masking Efficiency

Table 5.3 lists the overall acceptance $\epsilon_{\rm acc}$ after all masks. Each CCD changes by less than two percentage points across the four blinded periods, showing the stability of the selection pipeline.

	CCD A	**CCD B**	**CCD C**	**CCD D**
B1	0.942	0.977	0.840	0.923
B2	0.924	0.907	0.842	0.968
B3	0.929	0.917	0.837	0.964
B4	0.928	0.919	0.835	0.960

Table 5.3: Global acceptance $\epsilon_{\rm acc}$ after all masks for every blinded dataset.

For CCD A the acceptance oscillates between 0.924 and 0.942, limited chiefly by the shallow overdensity at columns [820–1031] and the hot-column pair at [3764–3771]. CCD B shows the largest drop, from 0.977 in B1 to 0.907 in B2, as the edge overdensity expands toward the serial register; afterwards it stabilizes near 0.91–0.92. CCD C, dominated by its broad overdense band starting at 5310, remains close to 0.84 in every dataset. For CCD D , the acceptance remains very high in B2–B4 (0.960–0.968) since no overdensity cuts are applied there, but dips to 0.923 in B1 due to two narrow masked bands. Overall, D

shows minimal loss outside that single run, with no trend of progressive degradation.

Step-by-step efficiency tables together with the row- and column-multiplicity histograms for every period are collected in Appendix D.1 (Tables D.1–D.7).

5.1.3 Dark-Current Stability and Charge Resolution

After masking, the single-pixel charge spectra of B1–B4 are described by the same Poisson–Gaussian model validated on UB0. The fitted Poisson means vary by less than 10 % from run to run: CCDs A and B lie in the range $(3.3{-}4.3) \cdot 10^2\,\mathrm{e^-\,g^{-1}\,day^{-1}}$, CCD C clusters around $4.0 \cdot 10^2\,\mathrm{e^-\,g^{-1}\,day^{-1}}$, and CCD D is consistently higher at $(4.5{-}4.8) \cdot 10^2\,\mathrm{e^-\,g^{-1}\,day^{-1}}$.

The Gaussian width, which quantifies single-electron resolution, is equally stable: every sensor returns $\sigma = 0.153{-}0.167\,\mathrm{e^-}$ with no drift over the three-month campaign. Table 5.4 lists the mean values per CCD; run-to-run fluctuations stay below 10 %.

	CCD A	**CCD B**	**CCD C**	**CCD D**
$\langle\boldsymbol{\lambda}\rangle$ $[\mathbf{e^-\,g^{-1}\,day^{-1}}]$	$4.2 \cdot 10^2$	$3.4 \cdot 10^2$	$4.0 \cdot 10^2$	$4.7 \cdot 10^2$
$\langle\boldsymbol{\sigma}\rangle$ $[\mathbf{e^-}]$	0.154	0.157	0.160	0.165

Table 5.4: Average dark-current rate λ and Gaussian width σ for each CCD, obtained by combining the four blinded datasets. Quoted values are the mean of B1–B4; run-to-run fluctuations stay below 10 %.

The constancy of both parameters confirms that the PA08 module operated under stable environmental conditions and that the masking strategy does not distort the low-energy spectrum. The Poisson–Gaussian template derived from UB0 can therefore be applied to all blinded data without further adjustment.

5.2 Pattern Identification in Blinded Data

Having validated the pattern identification method and its compatibility with Poisson expectations on unblinded data, we now apply the same selection pipeline to the four blinded periods.

The analysis aims to test the presence of statistically significant deviations in multi-pixel structures beyond those expected from dark current and electronic noise. Particular attention is given to patterns with total charges of two or three electrons, whose occurrence rates may hint at low-energy ionization signals.

Each of the following subsections presents the results for an individual data-taking period, comparing the number of observed patterns with the predicted counts from Poisson convolutions. The analysis includes per-CCD breakdowns, directional pattern identification, and an evaluation of the potential excesses relative to statistical expectations.

5.2.1 Dataset B1 (2024-10-14)

Table D.9 compares the measured population of isolated one-, two- and three-electron pixels with the expectations of the Poisson model for the first blinded dataset B1. The one-electron population agrees with the Poisson model in every CCD by definition. For isolated two-electron hits we find $22{\pm}5$ candidates compared with 6.6 ± 1.2 expected, an overall excess of about 3σ. It is distributed over the four CCDs rather than being driven by a single CCD. Exactly one three-electron pixel survives the cascade. Although the expected rate is very low ($((6.4 \pm 1.8) \cdot 10^{-4})$), the observed count of 1 ± 1 remains consistent within 1σ, due to the large statistical uncertainty.

Multi-pixel results for B1 are given in Table D.10. A global summary appears in the first row of Table 5.6 (and Table 5.7 for three-pixel patterns). The only topology that appears is the {11} doublet.

Restricting the comparison to row-aligned pairs-the numbers in parentheses in the tables-we found 14 ± 4 events versus 12 ± 2 expected, a marginal 1σ. No patterns such as {12}, {13}, {111} or permutations, are observed; their expected rates are below one event and the non-observation is fully compatible with Poisson noise. A compact overview of all pattern topologies is already given by Tables D.9–D.10 (see also Appendix D.2). Dataset B1 therefore shows a modest excess of two-electron single pixels with no significant anomaly in pair topologies.

5.2.2 Dataset B2 (2024-10-28)

The one-electron sample matches the background model in every CCD by construction. In the second row of Table 5.5, an excess is seen for isolated $2e^-$ pixels: observed 44 ± 7, whereas the Poisson model predicts 15 ± 2. The excess, about 4σ is shared by all four CCDs with the largest share in CCD C (see Tables D.11 and D.12). Five isolated $3e^-$ pixels are also found, well above the expectation $(1.4 \pm 0.3) \cdot 10^{-3}$); given the very low background, these few events are simply noted for future statistics.

In agreement with the Poisson expectations of 29 ± 4, a total of 29 ± 5 row-aligned {11} patterns are found. No excess appears in any other two- or three-pixel pattern topology; all are consistent with accidental coincidences at the sub-percent level (Tables D.11–D.12 show all patterns per-CCD).

5.2.3 Dataset B3 (2024-11-19)

As in the previous datasets, we observe an excess in the two- and three-electron event rates. The population of isolated one-electron patterns follows the expectation by definition. In contrast, single-pixel two-electron events show a pronounced excess 53 ± 7 against 21 ± 3 predicted, corresponding to a $\sim 5\sigma$ upward fluctuation. Similarly, 11 ± 3 triple-electron singlets survive the selection cascade, while the Poisson model predicts only $(2.0 \pm 0.4) \cdot 10^{-3}$, confirming the same trend of two and three electron patterns seen in earlier datasets. Extending the search to neighboring pixels reveals a comparable excess. Row-aligned {11} pairs are measured at 40 ± 6, in line with the predicted 39 ± 5. Only one candidate each is seen for {12} and {13}; given their sub-milli-level probabilities, the appearance of a single event is statistically improbable. No three-pixel patterns are identified. The full per-CCD tables for B3 are provided in Appendix D.2 (Tables D.13 and D.14).

5.2.4 Dataset B4 (2024-12-17)

The one-electron population again matches the background model by construction. Isolated $2e^-$ pixels are measured at 33 ± 6 while the background model predicts 14.6 ± 1.8, a $\sim 3\sigma$ excess distributed over all sensors. For $3e^-$ isolated pixels, 3 ± 2 are found; the Poisson prediction is practically zero $(1.5 \pm 0.3) \cdot 10^{-3}$. Row-aligned pattern doublets of type {11} remain consistence 29 ± 5 candidates versus 27 ± 3 expected. No other two- or three-pattern survives the isolation cuts, exactly as anticipated from their millipercent-level probabilities (see Tables D.15–D.16 in Appendix D.2).

5.2.5 Global pattern Blinded Data

Table 5.5 collects the single-pixel yields for the four blinded runs, while Tables 5.6 and 5.7 summarize the two- and three-pixel topologies, respectively[1]. Poisson–Gaussian expectations are computed run-by-run from the 1-e^- peaks and then summed; quoted errors on the data are statistical ($\sqrt{N}$), and the treatment of expected errors follows Section 4.5.

The $1\,e^-$ population matches the dark-current model by construction, confirming the stability of the background fit over the entire three-month dataset. Conversely, isolated multi–electron pixels remain

[1]Columns with bracketed powers of ten must be multiplied accordingly; Obs in red exceed the Poisson expectation. Each multiplicity-pattern symbol implicitly includes all charge permutations of the same topology.

an open issue: the method records 152 ± 13 two–electron and 20 ± 4 three–electron single pixels, to be compared with expectations of 57 ± 4 and $(5.5 \pm 0.6) \times 10^{-3}$, respectively. The corresponding fluctuations reach $\sim 7\sigma$ and $\sim 4.7\sigma$, and echo the behavior was already seen in the unblinded reference data (for $2e^-$). Because these extra events are spread over all four runs and all CCDs, they are unlikely to originate from a single localized defect.

When examining patterns, no three-pixel configuration survives the cascade, exactly as expected from the negligible Poisson probabilities. Row-aligned {11} doublets total 112 ± 11, in agreement with the prediction 107 ± 7. Only two exotic topologies appear: {21} and {13}. A cross–check at the same coordinates in the companion sensors shows that the {21} candidate's $2\,e^-$ pixel registers just $0.67\,e^-$ elsewhere, identifying it as crosstalk; this event is therefore discarded. The single {13} hit is kept for completeness but, given its vanishing Poisson probability and lack of spatial or temporal correlation, is likewise attributed to residual instrumental background.

	{1}		**{2}**		**{3}**	
	Exp $[10^4]$	Obs $[10^4]$	Exp	Obs	Exp $[10^{-3}]$	Obs
B1	4.6 ± 0.4	4.37 ± 0.04	6.6 ± 1.2	22 ± 5	0.64 ± 0.18	1 ± 1
B2	11.0 ± 0.8	10.62 ± 0.03	15 ± 2	44 ± 7	1.4 ± 0.3	5 ± 2
B3	14.0 ± 0.9	13.39 ± 0.12	21 ± 3	53 ± 7	2.0 ± 0.4	11 ± 3
B4	10.0 ± 0.6	9.50 ± 0.01	15 ± 2	33 ± 6	1.5 ± 0.3	3 ± 2
Total	39.6 ± 1.4	37.88 ± 0.13	57 ± 4	152 ± 13	5.5 ± 0.6	20 ± 4

Table 5.5: Expected and observed counts for patterns {1}, {2} and {3} across all datasets.

	{11}		**{12}**		**{13}**	
	Exp	Obs	Exp $[10^{-3}]$	Obs	Exp $[10^{-7}]$	Obs
B1	12 ± 2	14 ± 4	1.8 ± 0.5	0	1.7 ± 0.7	0
B2	29 ± 4	29 ± 6	4.0 ± 0.9	0	4.0 ± 1.2	0
B3	39 ± 5	40 ± 6	5.7 ± 1.1	1	5.7 ± 1.5	1
B4	27 ± 3	29 ± 5	4.0 ± 0.7	0	4.0 ± 1.0	0
Total	107 ± 7	112 ± 11	15.5 ± 1.7	1	15 ± 2	1

Table 5.6: Expected and observed counts for patterns {11}, {12}, and {13} for all datasets.

	{111}		**{112}**	
	Exp $[10^{-3}]$	Obs	Exp $[10^{-6}]$	Obs
B1	3.6 ± 1.0	0	0.59 ± 0.23	0
B2	8.4 ± 1.8	0	1.3 ± 0.4	0
B3	12 ± 2	0	1.9 ± 0.5	0
B4	8.2 ± 1.5	0	1.3 ± 0.3	0
Total	32 ± 3	0	5.1 ± 0.7	0

Table 5.7: Expected and observed counts for patterns {111}, {112} for all datasets.

Chapter 6

Conclusions

The main goal of this work has been to implement and validate a complete masking strategy for the analysis of low-charge events in the DAMIC-M Skipper-CCDs. A total of seven masks were applied to remove instrumental backgrounds and select a clean and reliable data sample. These masks include the removal of clusters, crosstalk, hot columns, overdensities, CTI trails and halos, high-multiplicity pixels and isolated columns. After this procedure, the final area efficiencies obtained for the blind datasets B1 to B4 range between 84 % and 98 %, depending on the CCD, with the lowest efficiencies found in CCD C (about 0.84, due to a large overdensity near the prescan) and the highest in CCD D (about 0.96).

Some of the masking steps are already stable and do not require further changes, like the cluster and crosstalk masks. Others, such as the hot column mask, could benefit from improving the method used to detect the end of exponential tails, which currently relies on a fixed discontinuity. A more precise method, based on fitting the decay and identifying deviations from the expected behavior, could allow for a more accurate definition of the masked region. The overdensity mask, while effective, still depends on manually tuned parameters such as the window size and detection thresholds. Although alternative approaches based on change-point or isolation-forest methods were tested, their performance in efficiency was too low. Future work should therefore focus on alternative, less parameter-dependent statistics to refine this mask.

Once the cleaned data is obtained, the charge spectrum deviates from the expectation of a purely Poissonian dark current. In particular, the number of isolated $2\,e^-$ and $3\,e^-$ pixels is significantly higher than predicted by the model. Regarding isolated $1\,e^-$ pixels, since their rate is directly measured from data we cannot know whether there is an excess. The true origin of these events remains unknown, and may include contributions from DC, infrared photons, or even a dark matter signal. A promising discriminator is a search for daily modulation in the $1\,e^-$ rate, which would be a clear signature of dark-matter interactions. With the current background model, which includes only random coincidences of $1\,e^-$ pixels, the expected rate for a $\{31\}$ pattern is $B_{31} = 1.5 \times 10^{-6}$ events, so the observation of a single event represents a clear upward fluctuation. This estimate, however, does not include radiogenic backgrounds that can mimic multi-pixel patterns. Using the same dataset, Ref. [24] finds a background expectation of $B_{31} \simeq 7 \times 10^{-2}$ events for this topology, which renders the observation statistically consistent with background fluctuations. In either case, the single candidate is insufficient to claim evidence of a dark-matter signal.

For the excesses observed in $2\,e^-$ and $3\,e^-$ isolated pixels, a dark matter interpretation is unlikely. Dark matter is expected to interact uniformly in the bulk of the detector, and due to charge diffusion, many of the resulting energy deposits would appear as multi-pixel patterns. Diffusion simulations (Section 4.4) show that 38 % of true $2\,e^-$ deposits and 65 % of true $3\,e^-$ deposits should form multi-pixel patterns. In the combined blind dataset (B1+B2+B3+B4), we observed 152 isolated $2\,e^-$ pixels and 21 isolated $3\,e^-$ pixels. Using the simulated diffusion probabilities $p_{\text{diff}}(2 \to 2) = 0.351$ and $p_{\text{diff}}(3 \to 3) = 0.163$, we estimate a true number of deposits of $N_{\text{real}}(2e^-) \approx 433$ and $N_{\text{real}}(3e^-) \approx 129$. Such populations would yield clear excesses of patterns like $\{11\}$, $\{2\}$, $\{21\}$, $\{12\}$ or $\{111\}$, which we do not observe. For instance, we would expect an excess of about 166 $\{11\}$ patterns, but the observed numbers are compatible with background expectations. We therefore conclude that the excesses arise from a mixture of radiogenic events and residual instrumental background, not from dark-matter interactions uniformly distributed in the CCD bulk.

The identification of patterns based on charge distribution offers a more sensitive observable than the simple pixel-charge histogram. By reconstructing and classifying 1-, 2-, and 3-pixel patterns, and

comparing their observed rates with a background model based on Poissonian dark current plus radiogenic components, we are able to increase the sensitivity of the analysis. A dedicated study of the systematic uncertainties introduced by the pattern selection, similar in spirit to the analysis carried out by Paula [TFM Paula Pérez], is ongoing and will be crucial for future limit calculations.

Future work

There are several directions in which this work could be extended. First, the overdensity mask could be made more robust by exploring new statistical approaches that reduce its dependence on manually tuned parameters. Similarly, the method used to define the extent of hot columns could be improved by replacing the current thresholding strategy with a fit to the exponential tail followed by a dynamic cutoff. Another relevant improvement would be the implementation of a mask to remove residual correlated noise. This type of noise can be identified by detecting spatial patterns that systematically repeat across exposures or CCDs and deviate from statistical expectations. A suitable criterion would be to mask pixels with abnormally high variance or those exhibiting consistent correlations across modules, as done in recent DAMIC-M analyzes.

Although current masks have shown high efficiency in the available datasets, it would be important to validate their performance on larger exposures, where new sources of instrumental noise might appear or current thresholds might become suboptimal. In particular, it would be worth checking whether the combined dataset B1 to B4 yields consistent results with those obtained from the individual runs, both in terms of masking efficiency and pattern statistics. This comparison has not been performed in the present work because the computer used for the analysis was not able to process the full dataset at once. In addition, the model used to describe charge multiplicity, assumed to follow a Gaussian distribution, should be re-evaluated once higher-statistics datasets become available, since the limited sample size in the present analysis did not allow a reliable fit to the data. In terms of physics analysis, the next natural step is the study of daily modulation in the $1\,e^-$ event rate, which could help discriminate a possible dark matter component from known backgrounds.

In conclusion, the masking and pattern-based methodology developed in this thesis provides a clean and efficient framework for analyzing low-charge events in DAMIC-M. It allows us to identify and understand deviations from the expected background and improves our ability to search for dark matter signals with unprecedented sensitivity.

Bibliography

[1] Planck Collaboration, N. Aghanim, et al. Planck 2018 results. i. overview and the cosmological legacy of planck. *Astronomy & Astrophysics*, 641:A1, 2020. [Online]. Available: https://www.aanda.org/articles/aa/full_html/2020/09/aa33880-18/aa33880-18.html.

[2] G. Bertone, D. Hooper, and J. Silk. Particle dark matter: Evidence, candidates and constraints. *Physics Reports*, 405, 2005. [Online]. Available: https://arxiv.org/pdf/hep-ph/0404175.

[3] T. Marrodan and L. Rauch. Dark matter direct-detection experiments. *arXiv preprint*, 2017. [Online]. Available: https://arxiv.org/abs/1509.08767.

[4] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac. Constraints on the self-interaction cross section of dark matter from merging galaxy clusters. *Astrophys. J.*, 679:1173–1180, 2008.

[5] M. Schumann. Direct detection of wimp dark matter: Concepts and status. *Journal of Physics G: Nuclear and Particle Physics*, 46(10):103003, 2019. [Online]. Available: https://arxiv.org/abs/1903.03026.

[6] T. Lin. Dark matter models and direct detection. In *Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics: Theory in an Era of Data (TASI 2018)*. University of Colorado Boulder, SISSA Medialab, 2019. PoS(TASI2018)009. [Online]. Available: https://scispace.com/pdf/dark-matter-models-and-direct-detection-1tdh4geocn.pdf.

[7] B. J. Kavanagh and A. M. Green. Model independent determination of the dark matter mass from direct detection experiments. *Phys. Rev. Lett.*, 117(10):101301, 2016. [Online]. Available: https://arxiv.org/abs/1303.6868.

[8] R. Essig, J. Mardon, and T. Volansky. Direct detection of sub-gev dark matter with semiconductor targets. *Journal of High Energy Physics*, 2012(5):046, 2012. [Online]. Available: https://arxiv.org/abs/1108.5383.

[9] I. Arnquist *et al.* The damic-m experiment: status and first results. *Proceedings of Science*, 441:066, 2023. [Online]. Available: https://pos.sissa.it/441/066/pdf.

[10] DAMIC-M Collaboration. Dark matter in ccds at modane (damic-m) experiment. https://damicm.web.cern.ch/, 2022. [Online]. Accessed: 2025-05-08.

[11] M. Battaglieri and A. Belloni et al. Us cosmic visions: New ideas in dark matter 2017 – community report. arXiv:1707.04591 [hep-ph], 2017. [Online]. Available: https://arxiv.org/abs/1707.04591.

[12] W. S. Boyle and G. E. Smith. Charge coupled semiconductor devices. *Bell System Technical Journal*, 49:587–593, 1970.

[13] G. F. Amelio, M. F. Tompsett, and G. E. Smith. Experimental verification of the charge-coupled device concept. *Bell System Technical Journal*, 49:593–600, 1970.

[14] I. Arnquist, N. Avalos, and DAMIC-M Collaboration. First constraints from damic-m on sub-gev dark-matter particles interacting with electrons. *Physical Review Letters*, 130(17):171003,

2023. [Online]. Available: https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.130.171003.

[15] M. Hans-Gunther. Silicon detector systems in high energy physics. *Prog. Part. Nucl. Phys.*, 63:186–237, 2009.

[16] D. Rodrigues and K. Andersson et al. Absolute measurement of the fano factor using a skipper-ccd. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1010:165511, 2021. [Online]. Available: https://arxiv.org/abs/2004.11499.

[17] K. Ramanathan and N. Kurinsky. Ionization yield in silicon for ev-scale electron-recoil processes. *Phys. Rev. D*, 104(3):032006, 2021.

[18] A. Aguilar-Arevalo et al. Search for low-mass wimps in a 0.6 kg day exposure of the damic experiment at snolab. *Physical Review D*, 94(8):082006, 2016.

[19] L. De Domenicis, F. Bellini, G. Bruno, and *et al.* Radiopurity and radon mitigation strategies for the damic-m low-background chamber. *Journal of Instrumentation*, 17(07):T07001, 2022.

[20] D. Norcini and N. Castelló-Mor et al. Precision measurement of compton scattering in silicon with a skipper ccd for dark matter detection. *arXiv preprint*, 2022. [Online]. Available: https://arxiv.org/abs/2207.00809.

[21] I. Arnquist *et al.* The DAMIC-M Low Background Chamber. *JINST*, 2024. Prepared for submission.

[22] J. Tiffenberg. Counting electrons with the skipper-ccd. Presentation at the 52nd Annual Fermilab Users Meeting, August 2020. [Online]. Available: https://indico.fnal.gov/event/23109/contributions/193269/attachments/132888/163601/ura_eca.pdf.

[23] DAMIC-M Collaboration. DAMIC-M: Direct search for dark matter with Skipper-CCDs. Presentation at the Multidark Meeting, Santander, October 2024. Accessed: 2025-05-12.

[24] K. Aggarwal and et al. (DAMIC-M Collaboration). Probing benchmark models of hidden-sector dark matter with damic-m. *arXiv preprint arXiv:2503.14617*, 2025. [Online]. Available: https://arxiv.org/abs/2503.14617.

[25] R. Essig, T. Volansky, and T.-T. Yu. New constraints and prospects for sub-gev dark matter scattering off electrons in xenon. *Phys. Rev. D*, 96(4):043017, 2017.

[26] J. D. Lewin and P. F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astroparticle Physics*, 6(1):87–112, 1996.

[27] J. Tiffenberg, M. Sofo-Haro, A. Drlica-Wagner, R. Essig, J. Estrada, G. Fernandez, E. Paolini, and T. Volansky. Sub-electron readout noise in a skipper ccd. *Physical Review Letters*, 119(13):131802, 2017.

[28] L. Barak and et al. (SENSEI Collaboration). Characterization of single-electron events using a skipper-ccd. *arXiv preprint arXiv:2106.08347*, 2022. [Online]. Available: https://arxiv.org/abs/2106.08347.

[29] DAMIC-M Collaboration N. Castelló-Mor. Waders: A data processing framework for damic-m, 2024. Accessed: March 2025.

Appendix A

Data Pre-Processing

The initial processing of CCD images was performed using the WADERS framework [29], a modular Python-based tool developed within the DAMIC-M collaboration. WADERS is designed to apply a chain of processes over a real data CCD image. The sequence of processes applied to each raw FITS image is detailed below.

A.1 Compress of the Skipper images

The `CompressSkipperProcess` is a process that combines all individual skip measurements into a single image, reducing readout noise, by using a statistical function which can be set by the user. In this analysis, the statistical function used is the mean. This produces a two-dimensional array in which each pixel value represents the average charge measured during the exposure. This compressed image, serves as the input for all subsequent processing steps.

A.2 Pedestal Subtraction Process

The `PedestalSubtractionProcess` removes the electronic baseline (pedestal) from each pixel introduced during the CCD readout. In this analysis, the pedestal, $\mu_{\text{ped}}^{\text{row}}$, is obtained via a Gaussian fit to the pixel charge distribution in each row of the compressed image. To ensure stability, pixels with charges deviating by more than 3σ from the median were excluded, and the right tail of the distribution was suppressed.

This process corrects for the variation of the baseline with row index, due to charge accumulation during the readout. increase of baseline with row index due to charge accumulation during readout. The row-by-row standard deviation, σ_{row}, remains approximately constant across the sensor and is later used to define charge thresholds for cluster identification.

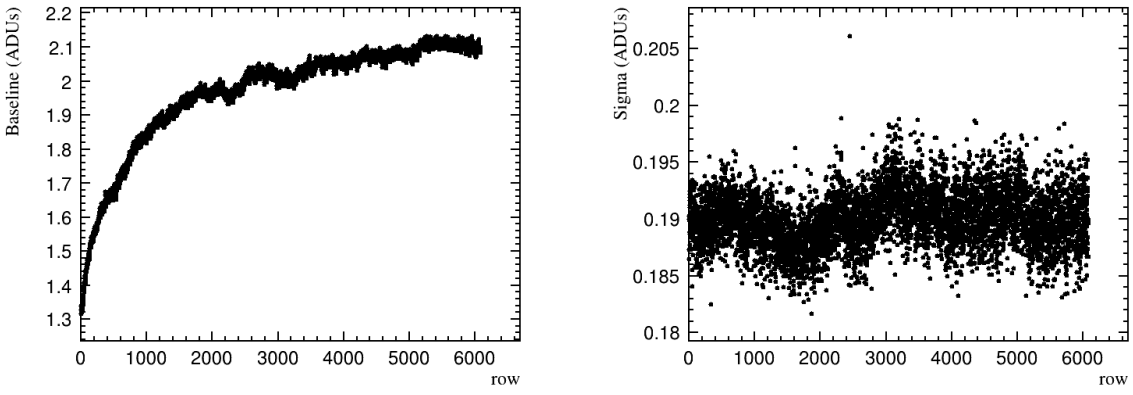


Figure A.1: Left panel: Baseline $\mu_{\text{ped}}^{\text{row}}$ as a function of row index, showing the gradual increase due to exposure time accumulation. Right panel: Standard deviation σ_{row} of the pedestal fit per row, used to define clustering thresholds. Note that both the baseline and electronic noise are expressed in analog-to-digital units (ADUs); see Section A.3 for their conversion to electrons.

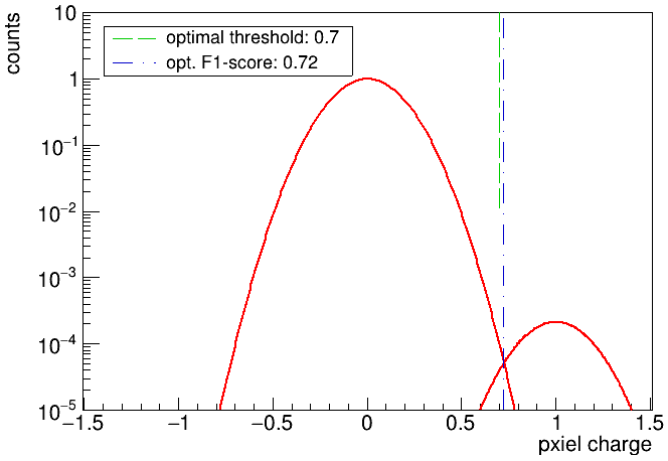


Figure A.2: PCD with optimal threshold indicated (green dashed line), determined using the get_energy_threshold tool for CCD A values.

A.3 Calibration Process

The `CalibrationProcess` converts the pixel charge from analog-to-digital units (ADU) to physical units of electrons (e^-) using a gain constant k determined for each amplifier. This constant is obtained by analyzing the pixel charge distribution with single-electron resolution, identifying the positions of peaks corresponding to 0, 1, 2, ... e^-, and fitting them individually with Gaussian functions. The gain k is then extracted as the mean distance between adjacent peaks, typically between the 0 and 1 e^- peaks. After this step, all pixel values are expressed in physical units of charge.

A.4 Threshold Optimization

The identification of signal pixels relies on the definition of a threshold above the noise level, beyond which a pixel is considered signal. To determine this optimal threshold, the script get_energy_threshold, developed by the DAMIC-M collaboration, was used. The tool takes as input the gain (in ADU/e^-), the electronic noise σ (in e^-), and the dark current rate λ (in e^-/pixel/image), all obtained from the Poisson-Gaussian fit described in Section A.3.

Rather than directly setting a fixed pixel charge threshold, the threshold optimization script implements two complementary statistical methods to determine the optimal value of n_{cut}, which defines the cut as a multiple of the calibrated readout noise σ, expressed in units of electrons.

The first method maximizes the *F1-score* for signal-pixel detection, balancing efficiency (true positives) and purity (precision) in the identification of $1\,e^-$ events. The second method finds the threshold $p_{\text{th}} = n_{\text{cut}} \cdot \sigma$ that equalizes the number of misclassified pixels on each side: the area of the $1\,e^-$ Gaussian left of the cut (missed signal) is set equal to the area of the $0\,e^-$ Gaussian right of the cut (false positives). This criterion inherently depends on the expected dark-current rate, which determines the relative normalization between the two peaks.

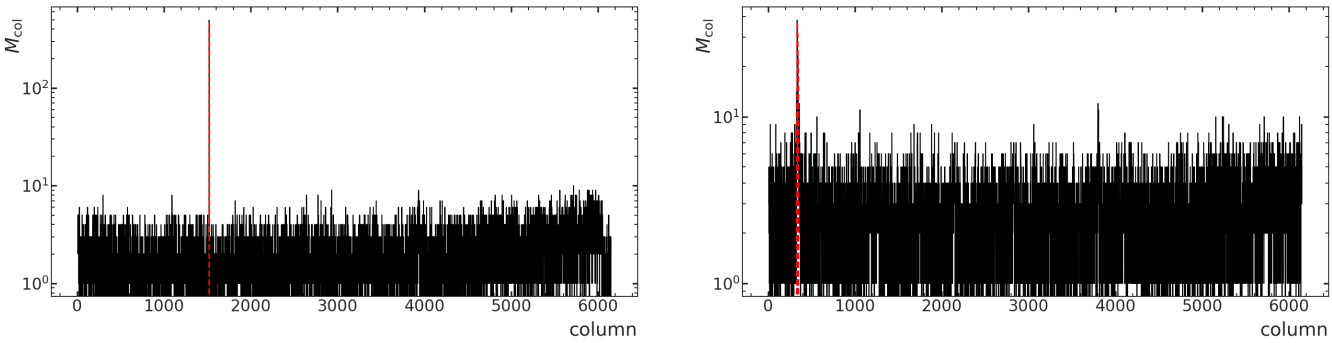
Since the readout noise σ differs between CCD amplifiers, the resulting n_{cut} is channel-specific and used as input to the `ClusterFinder` process. Figure A.2 illustrates the optimization result for CCD A, showing that both methods yield consistent results and identify a charge threshold of $0.7\,e^-$ as optimal for separating signal from noise.

Appendix B

Complementary Event Selection on Unblinded Data

This appendix collects supplementary figures from the unblinded UB0 dataset, complementing the data-selection procedure described in Chapter 3. While the main text focused on CCD A, here we present the corresponding figures for CCDs B, C and D.

The normalized column occupancy M_{col} is shown for CCDs B and D. Dashed red lines mark the columns flagged by the hot-column algorithm. Together with Fig. 3.7 (CCD A and C), this completes the survey of hot columns in UB0.



(a) CCD B (b) CCD D

Figure B.1: Normalized column occupancy M_{col} for UB0 sensors B and D.

The second plot (Fig. B.2) shows the charge per column in the range $[0.7, 3.7]\,e^-$ after masking hot columns for CCDs B and D. Orange bands indicate overdensity regions, the blue curve is a Savitzky–Golay smoothing and the red line the quadratic baseline fit. This supplements Fig. 3.8 for CCD A and C.

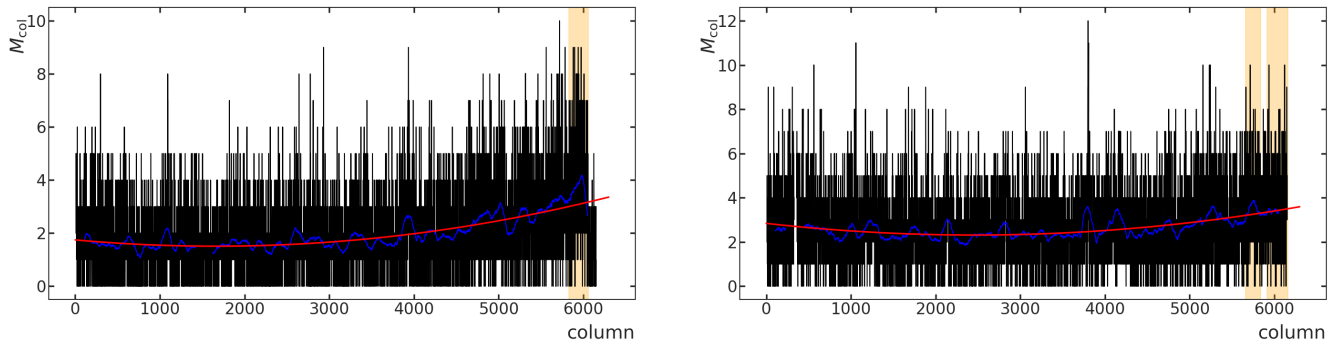


Figure B.2: Overdensity masks for UB0 sensors B and D.

Figures B.3–B.4 shows the multiplicity distributions of the charge multiplicity mask (Subsection 3.3.6). Left panels show charged-pixel multiplicity per row; right panels per column, for CCDs B, C, D. Figure B.3 uses threshold lines (red) to illustrate the mask cut; Fig. B.4 overlays Gaussian fits (red) and 99th-percentile thresholds (orange).

Mask maps and PCDs (Figure B.5). For each CCD we show (left) the 2D mask map and (right) the PCD after all masking steps. From top to bottom: B, C and D. See Fig. 3.13 for CCD A.

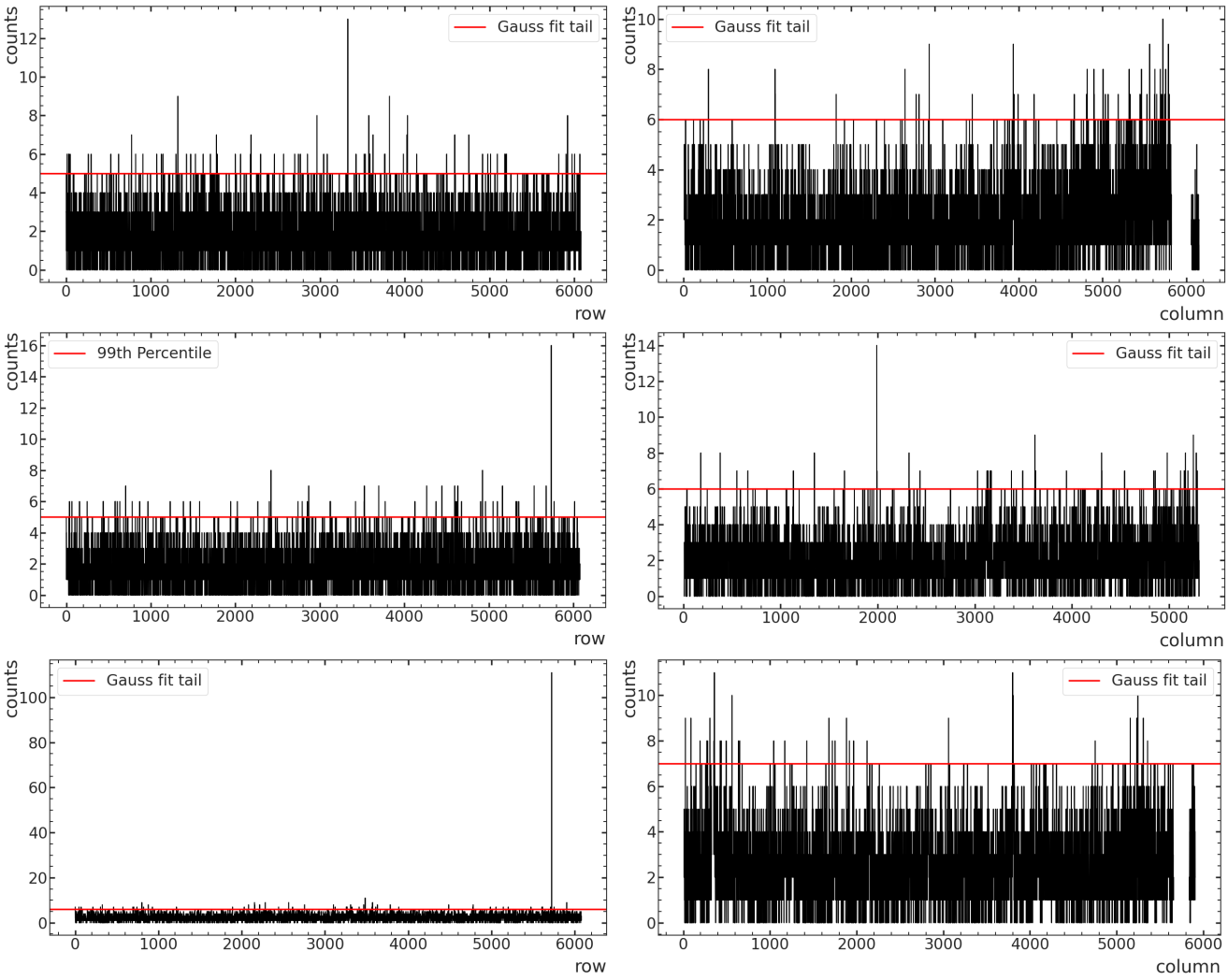


Figure B.3: Multiplicity per row/column with masking thresholds for UB0 CCDs B–D.

Dark-current distributions (Figure B.6). The histograms (black) show the post-mask charge distributions; the red curve is the Poisson–Gaussian fit of 2 picks. These complete Fig. 3.14 for CCDs A and C.

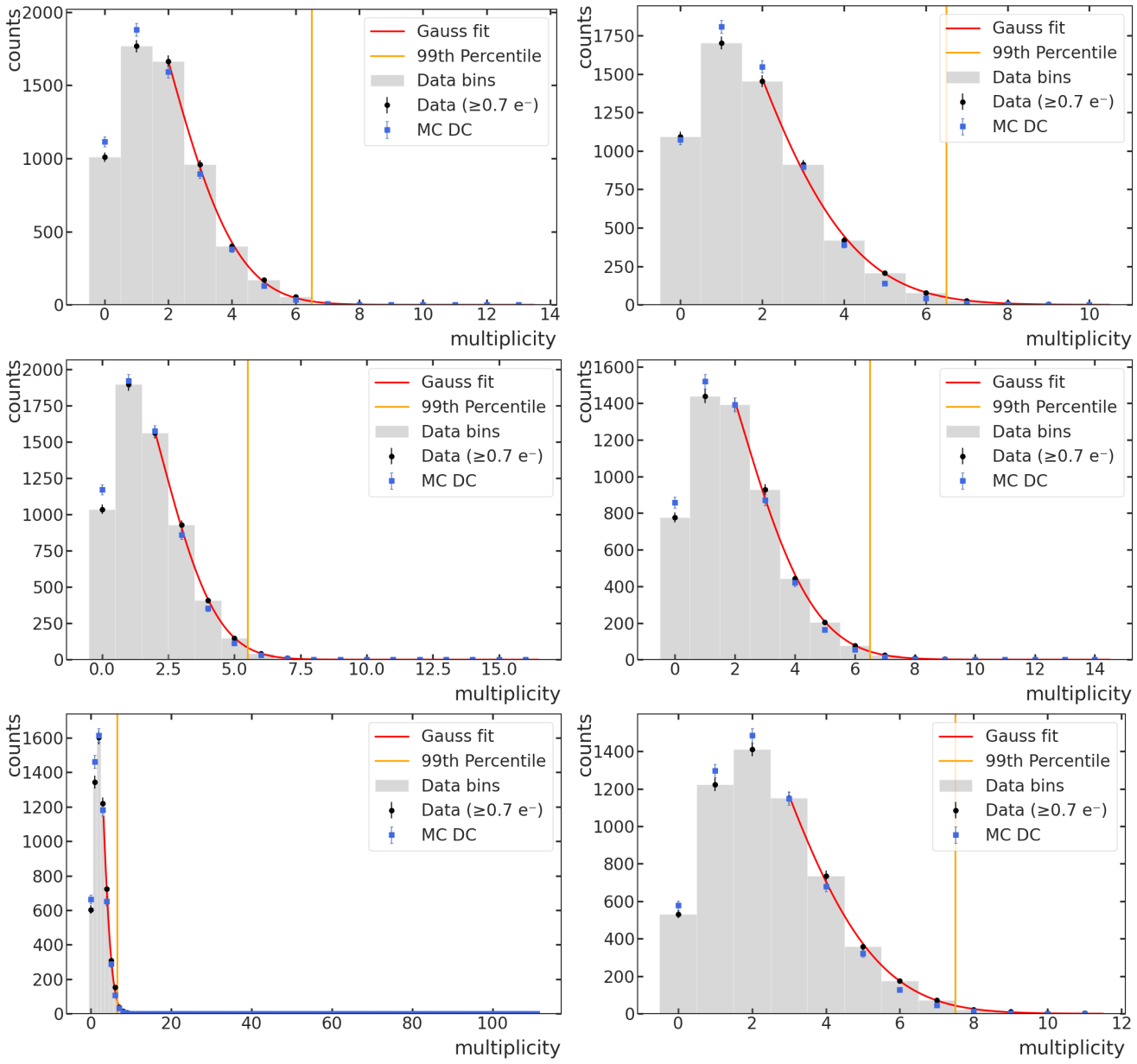


Figure B.4: Multiplicity per row/column with Gaussian fits (red) and 99% thresholds (orange) for UB0 CCDs B–D.

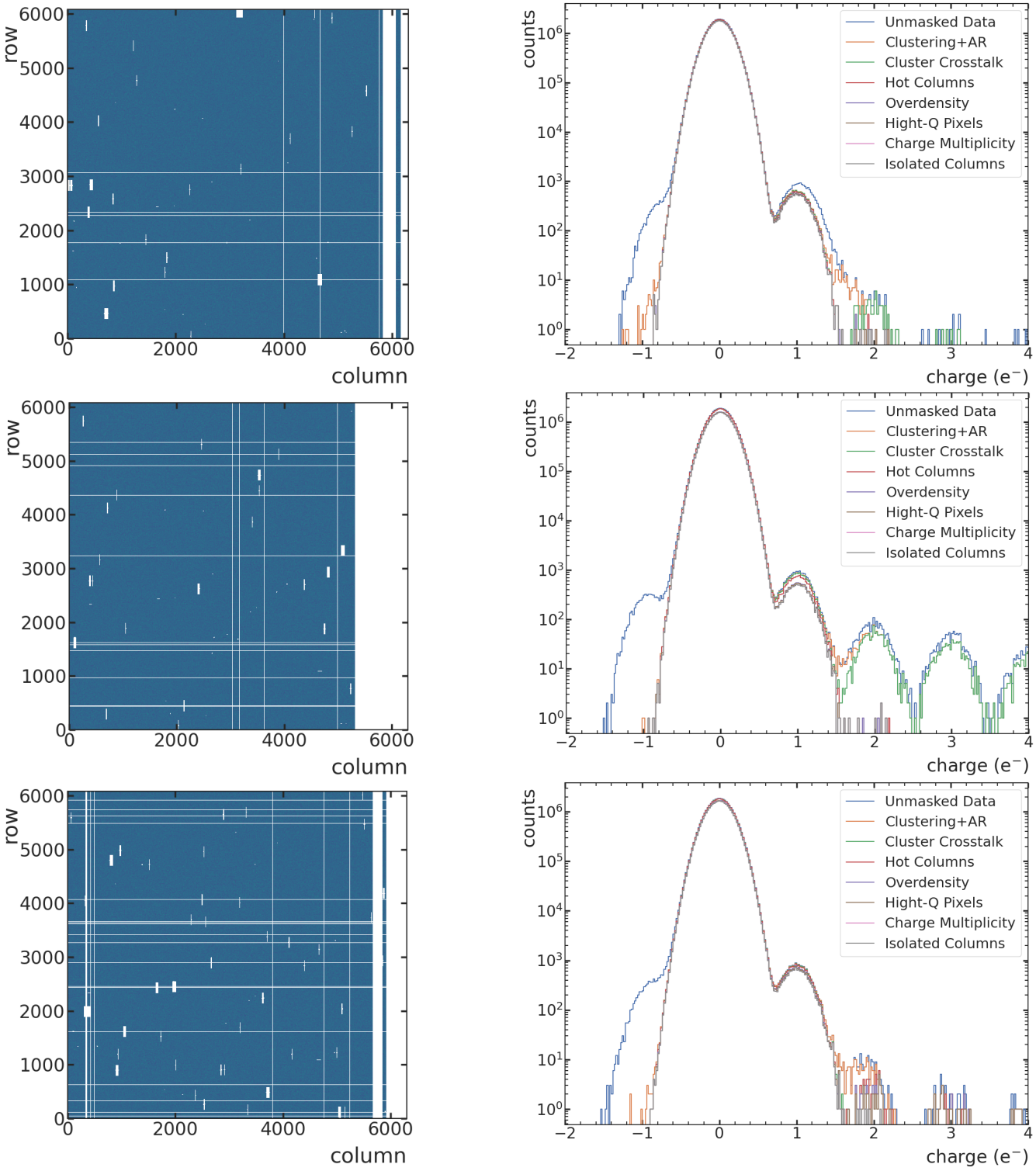


Figure B.5: Left: 2D mask map; Right: PCD after masking for UB0 dataset. From top to bottom: B, C and D. Continuation of Fig. 3.13 (CCD A).

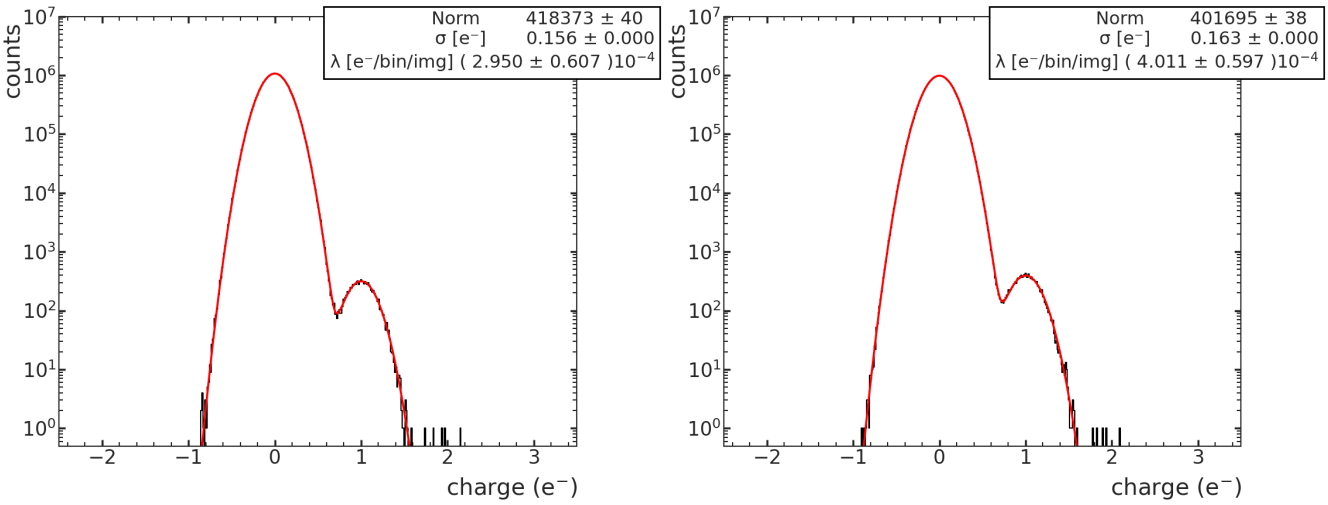


Figure B.6: Dark-current spectra and Poisson–Gaussian fits for UB0 CCDs B and D.

Appendix C

Pattern Distributions and Pixel Charge Correlations

This appendix compiles all figures related to the distributions and pixel charge properties of the patterns identified in the masked images. These plots serve to visually inspect the quality of the thresholds used in the pattern selection algorithm and to verify that they retain pixels consistent with low-charge ionization events. Only patterns with total charge up to $5\,e^-$ are shown.

C.1 One-Pixel Pattern Distributions

The total score distributions Λ_m for isolated patterns with $m = 1$ to $5\,e^-$ are shown below. These distributions reflect the global likelihood score assigned to each candidate, as computed from the CDF-based selection algorithm.

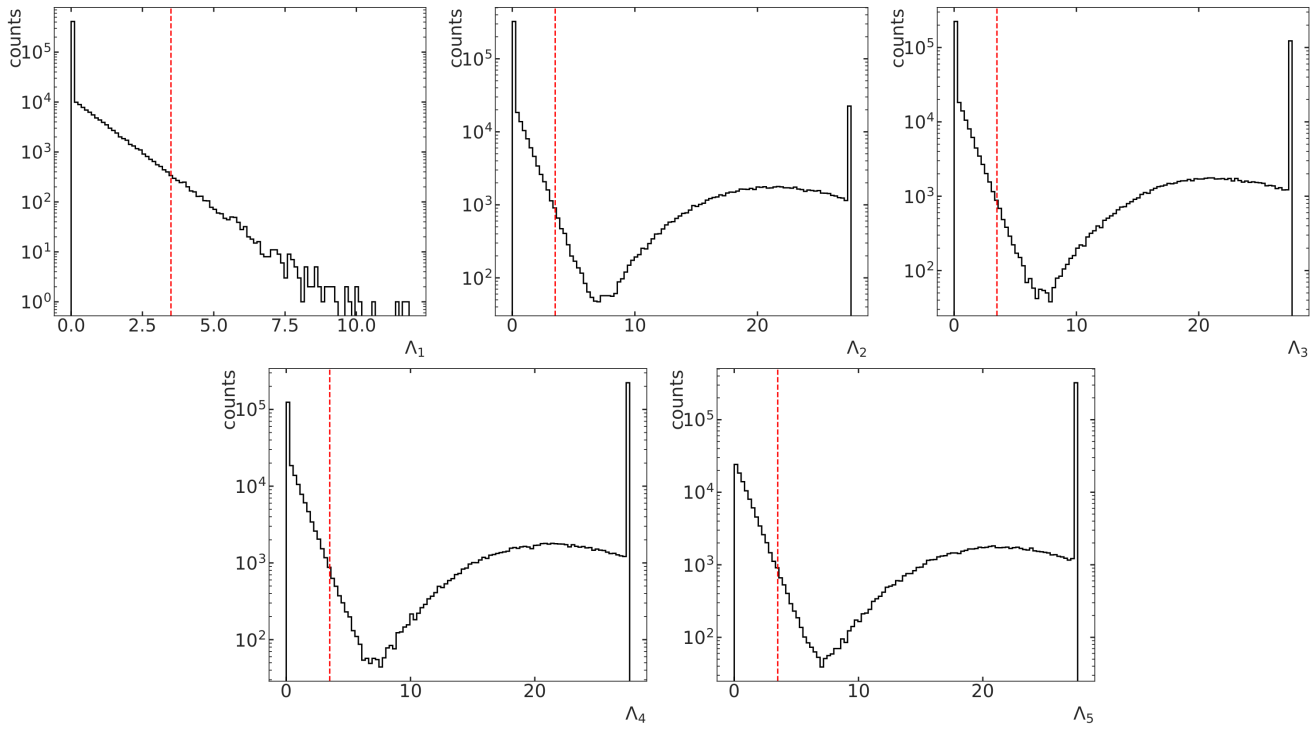


Figure C.1: Score distributions Λ_m for single-pixel patterns with total charge from 1 to $5\,e^-$. These scores are calculated using the likelihood-based pattern selection method.

C.2 Two-Pixel Pattern Distributions and Charge Correlation

This section presents detailed results for two-pixel patterns. Each pattern corresponds to a fixed charge configuration (m, n) such that $m + n \leq 5$. The figures include score distributions, pixel-pixel charge

correlations, and charge histograms. These visualizations assess the selection quality and help identify potential mis-identified combinations.

C.2.1 Score distributions for specific two-pixel patterns Λ_{mn}

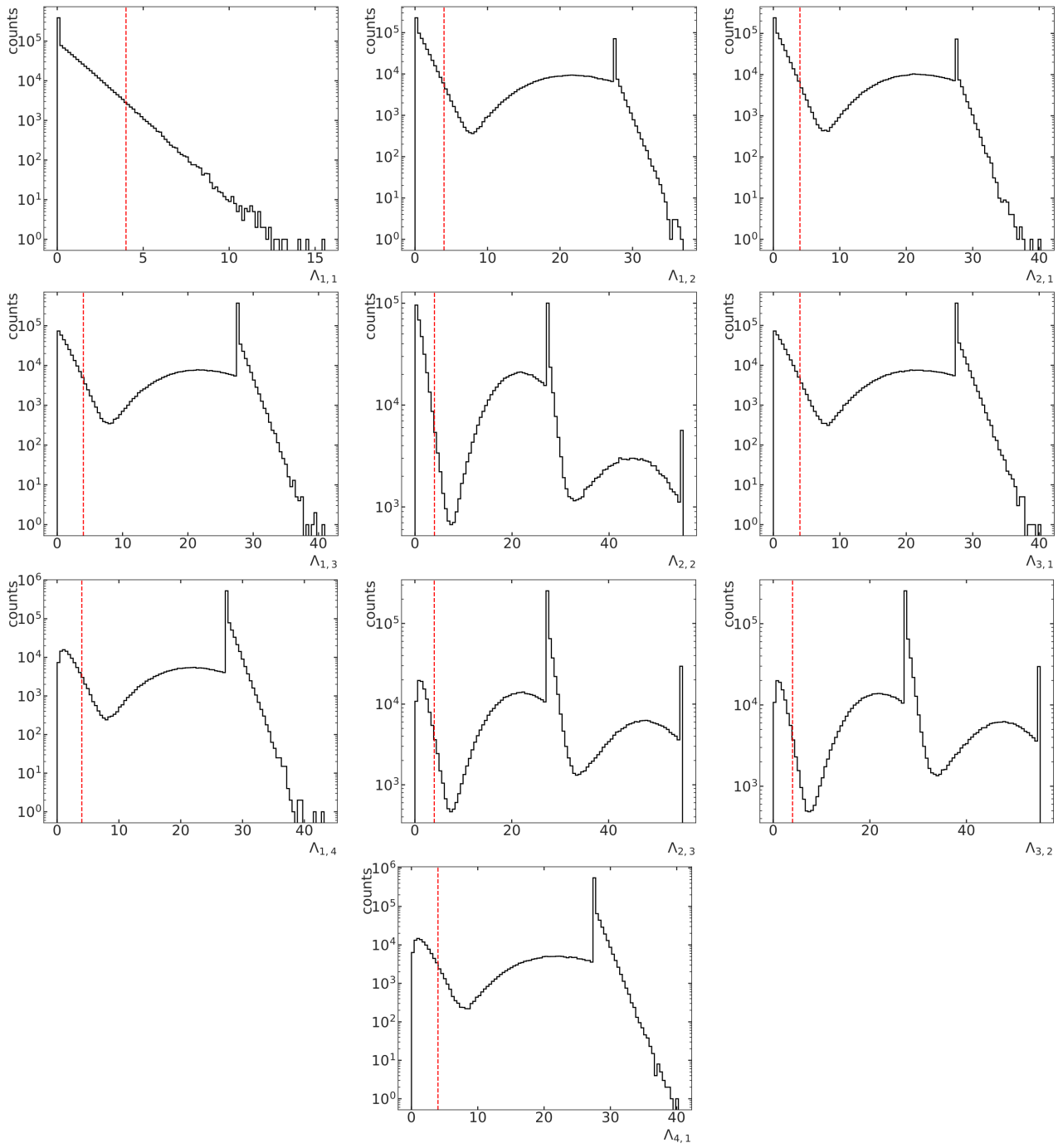


Figure C.2: Score distributions Λ_{mn} for selected two-pixel patterns, ordered as follows (left to right, top to bottom): (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (3,1), (3,2), (4,1). Each plot includes all two-pixel configurations with fixed charges (m, n), evaluated using the CDF-based score.

C.2.2 2D charge correlation maps

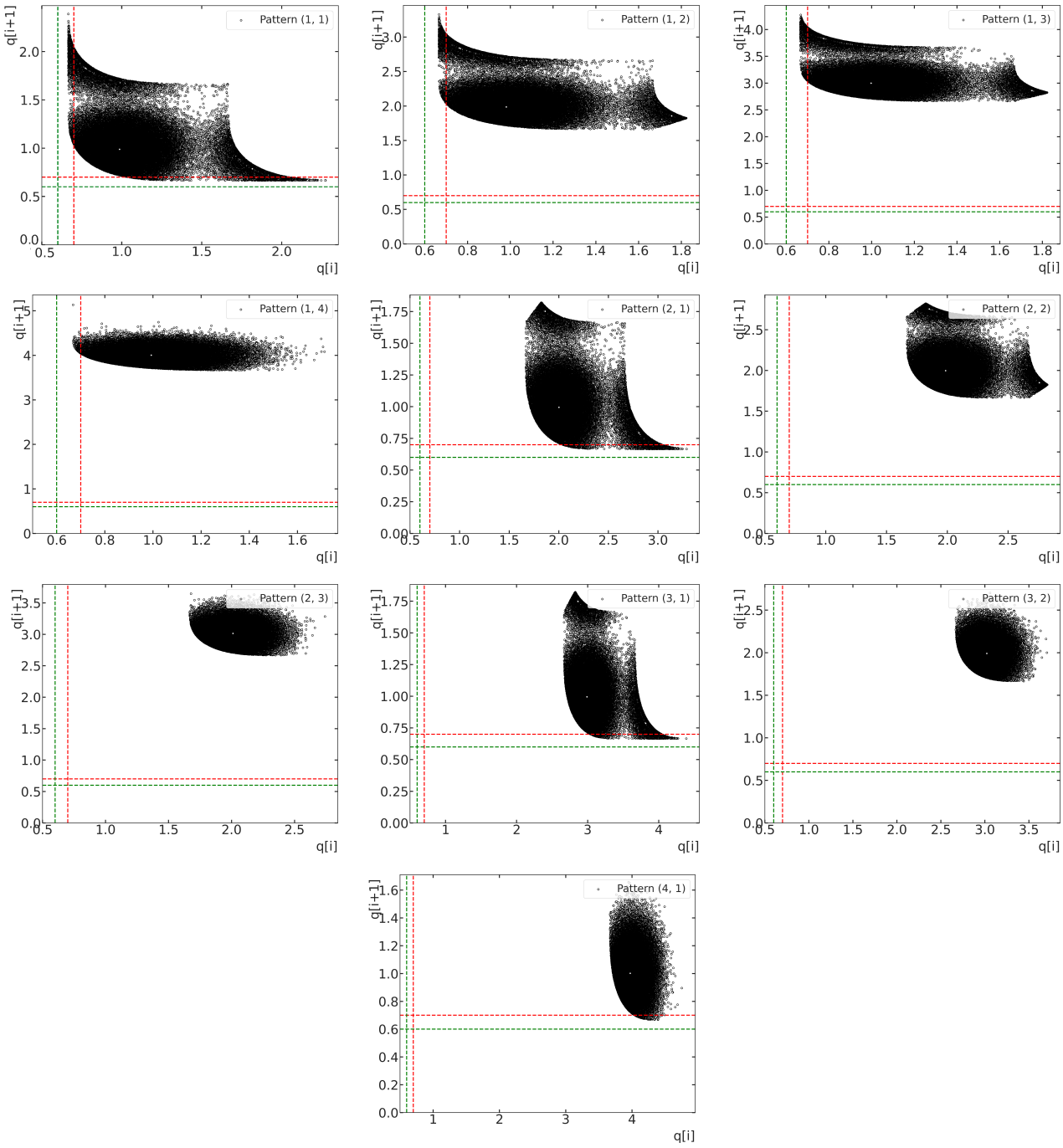


Figure C.3: Two-dimensional charge correlation maps for accepted two-pixel patterns Λ_{mn}, ordered as follows (left to right, top to bottom): (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (3,1), (3,2), (4,1). Each heatmap displays the charge distribution in both pixels for a given pattern.

C.2.3 Pixel-Charge Histograms

Complementary information is provided by Figure C.4. Here each histogram displays the individual pixel charges that survive the selection. For the $(1, 1)$ pattern (left) the two curves sit almost perfectly on top of

each other, both peaking at $\approx 1.0\,e^-$. The triplet case (right) mirrors this behavior. The histograms show a secondary shoulder around 1.7–1.8 e^- and a sparse tail extending to $\sim 2.3\,e^-$. These entries correspond to pairs in which one pixel sits at $\approx 0.7\,e^-$ while the partner pixel carries nearly the full two-electron charge. Such borderline combinations are formally accepted by the score-based selection and constitute the dominant source of the mis-identification rate. Their low statistics, however, ensure that the cascade still preserves the integrity of the main, quantized $1\,e^-$ population and keeps the impact on the overall charge spectrum negligible.

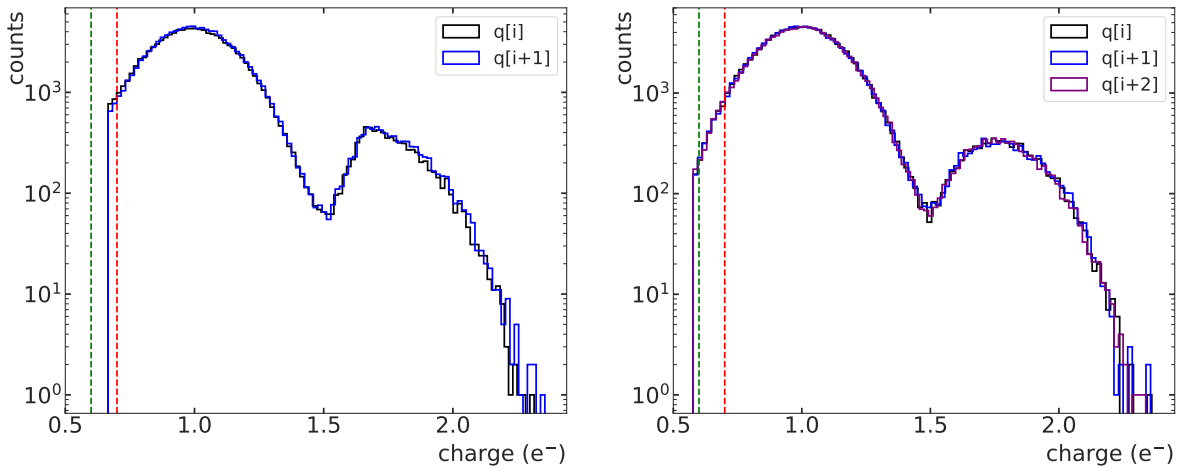


Figure C.4: Charge histograms for the pixels inside each accepted pattern. Left: the two members of a $(1, 1)$ doublet; Right: the three members of a $(1, 1, 1)$ triplet. The agreement of the peaks at $1\,e^-$ confirms correct quantification.

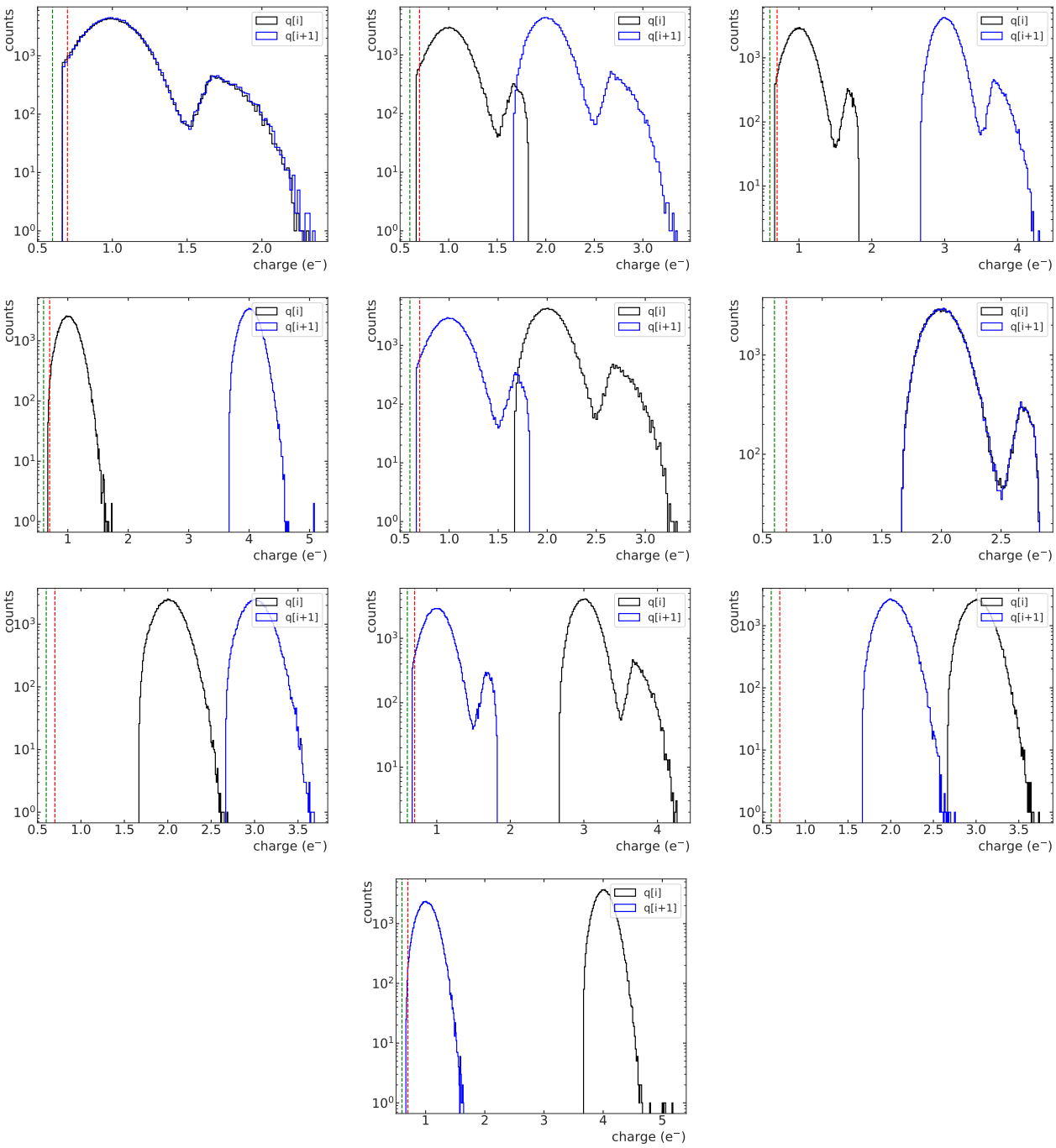


Figure C.5: Histograms of pixel charges for selected two-pixel patterns Λ_{mn}, ordered as follows (left to right, top to bottom): (1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (3,1), (3,2), (4,1). Each histogram shows the individual pixel charges retained by the selection algorithm.

C.3 Three-Pixel Pattern Distributions and Charge Correlation

Figures show the total scores Λ_{mnl} for patterns with specific three-pixel charge combinations, such as $(1, 2, 3)$. These include all possible permutations.

C.3.1 Score distributions for specific two-pixel patterns Λ_{mnl}

Figure C.6 displays the score distributions Λ_{mnl} for selected three-pixel patterns. Each histogram includes all permutations of the indicated configuration, scored using the CDF-based selection algorithm.

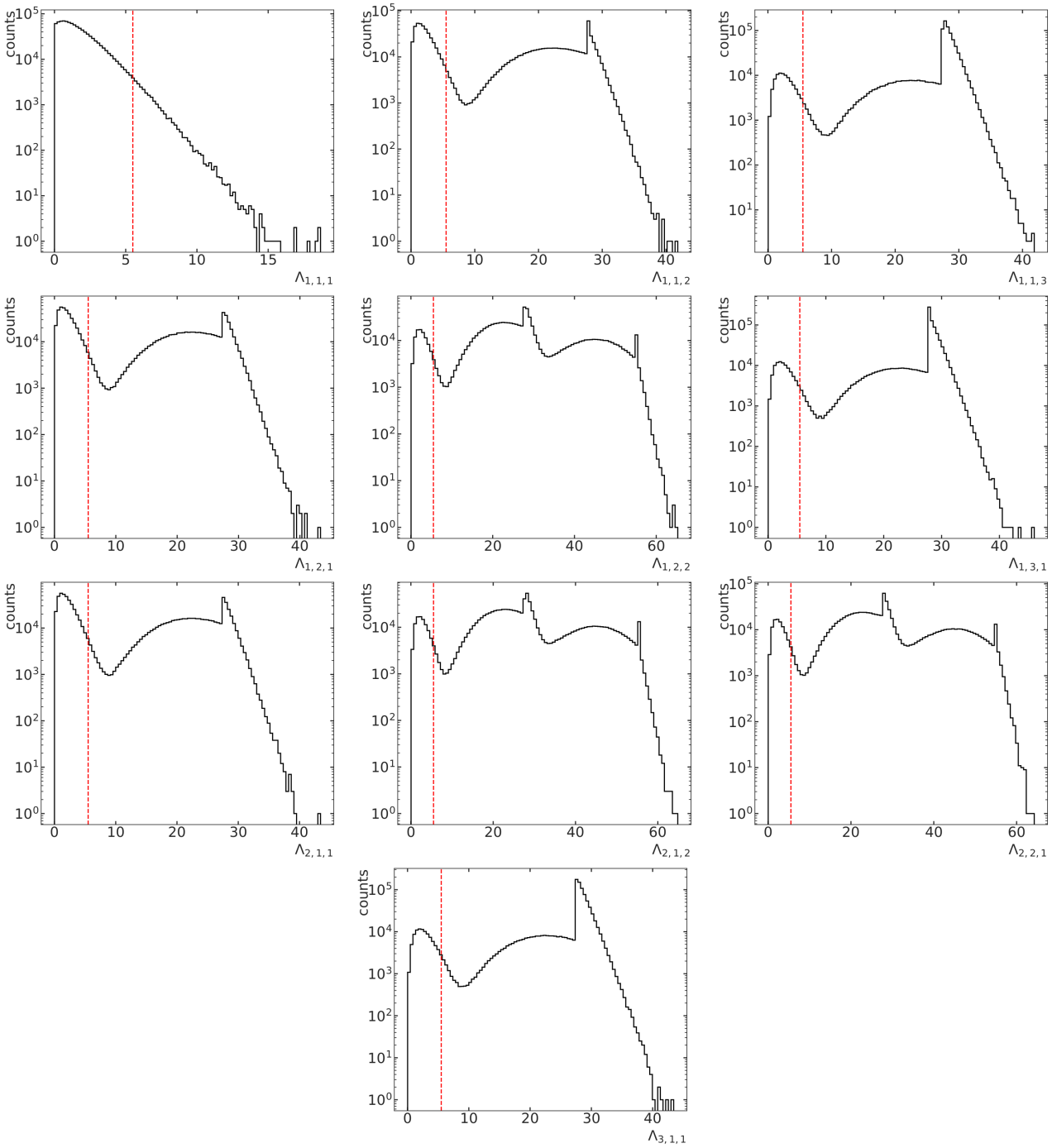


Figure C.6: Score distributions Λ_{mnl} for selected three-pixel patterns, ordered as follows (left to right, top to bottom): (1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), (1,3,1), (2,1,1), (2,1,2), (2,2,1), (3,1,1). Each distribution includes all permutations of the given charge configuration, scored using the CDF-based selection method.

C.3.2 2D charge correlation maps

Figure C.7 shows the 2D correlation maps of individual pixel charges within each three-pixel pattern. These maps visualize how the charge is typically distributed among the three components, and can reveal systematic asymmetries or tails caused by borderline or mixed-charge configurations.

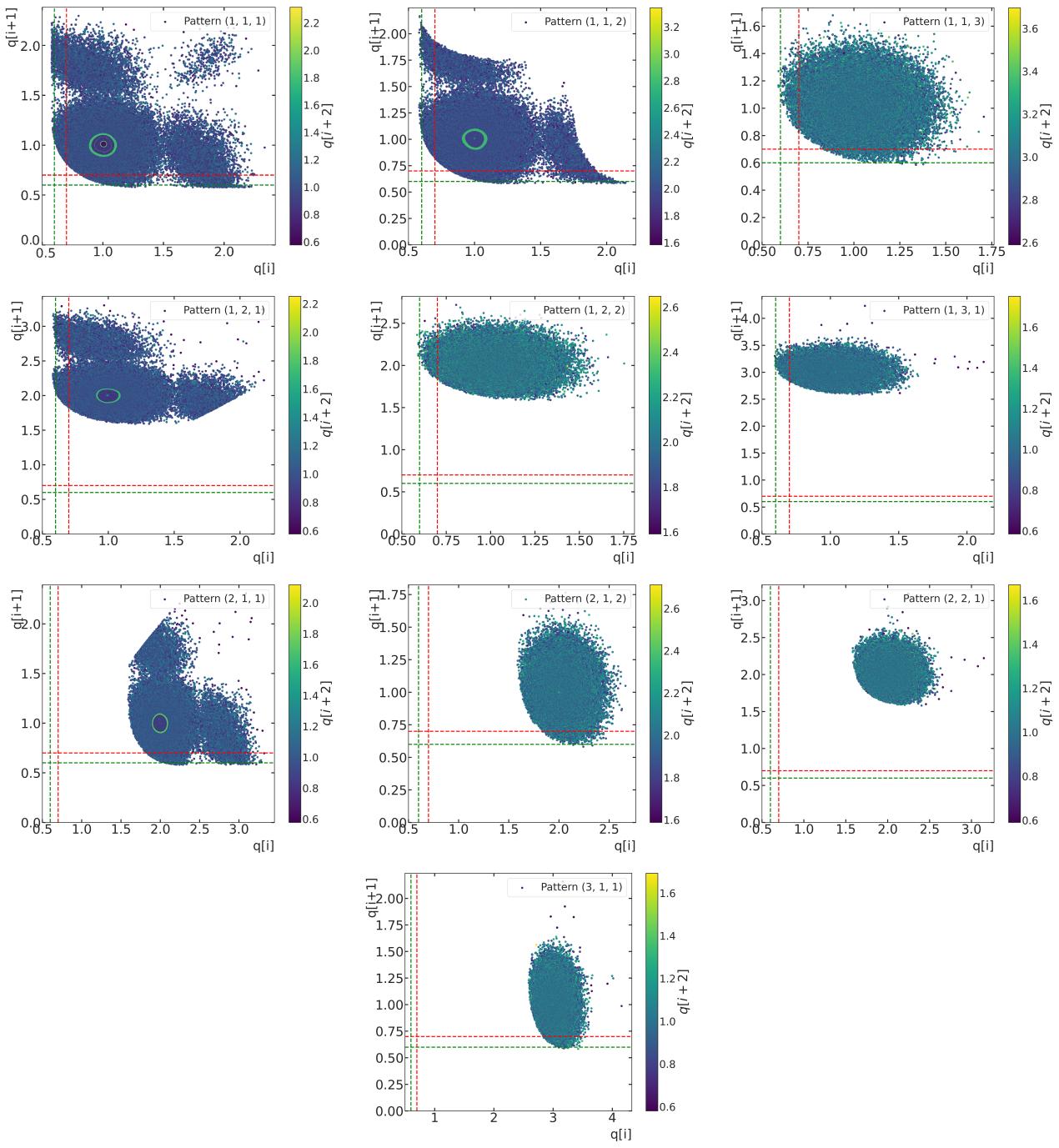


Figure C.7: 2D charge correlation maps for selected three-pixel patterns, in the same order as in Figure C.6: (1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), (1,3,1), (2,1,1), (2,1,2), (2,2,1), (3,1,1).

C.3.3 Pixel-Charge Histograms

Figure C.8 presents histograms of the individual pixel charges for each accepted three-pixel pattern. As with two-pixel patterns, the presence of secondary peaks or shoulders at sub-electron values indicates the inclusion of borderline combinations, particularly in cases where one or more pixels carry only partial charge.

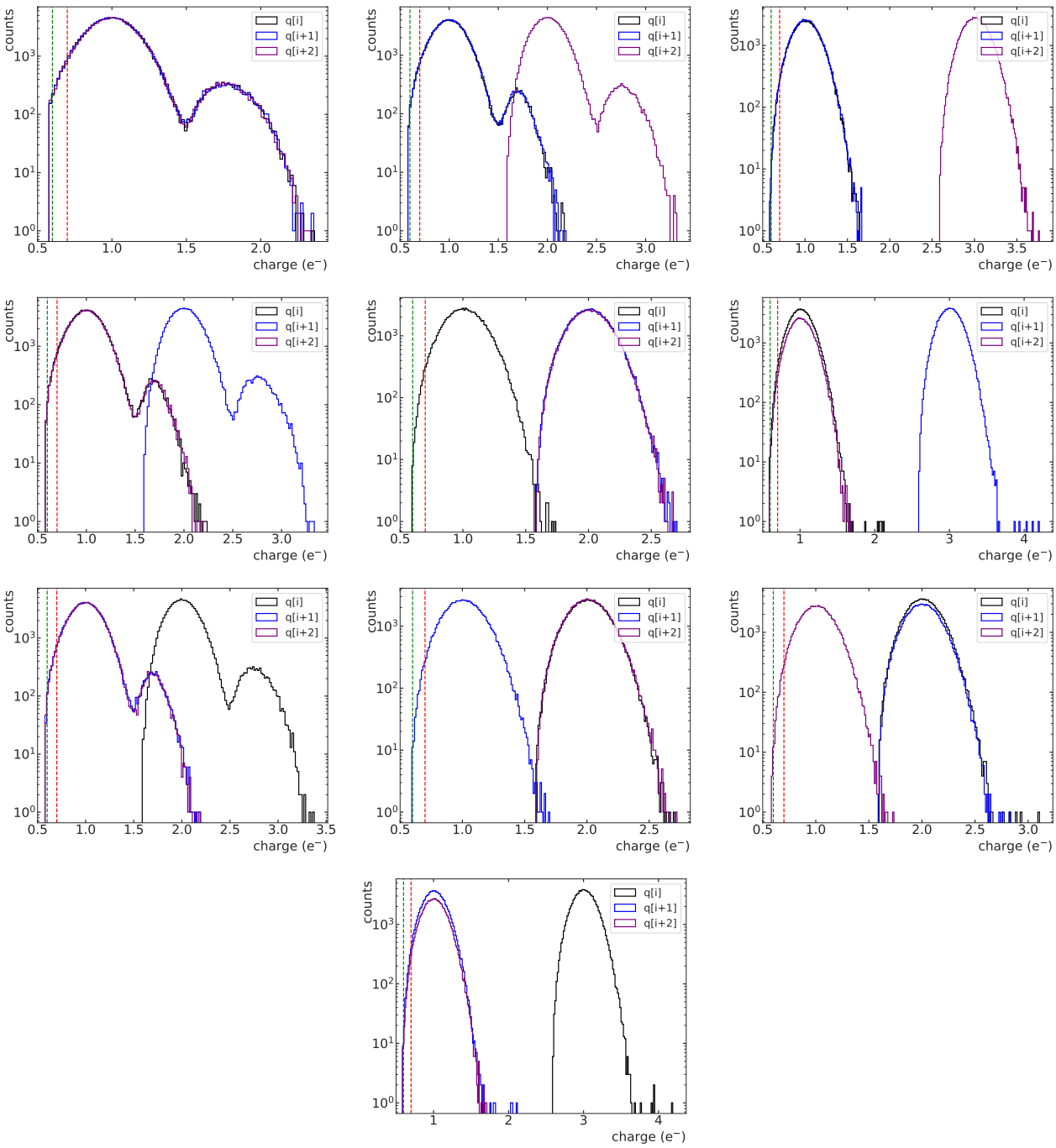


Figure C.8: Histograms of individual pixel charges in selected three-pixel patterns, following the same order as above: (1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), (1,3,1), (2,1,1), (2,1,2), (2,2,1), (3,1,1). These show the internal charge structure of each accepted pattern.

Appendix D

Extended Event Selection from Blinded Data

This appendix provides the full breakdown of masking efficiencies, pixel–charge distributions, and pattern counts for each blinded dataset (B1–B4), complementing the summary tables in Chapter 5. For every run you'll find: (i) Cumulative masking efficiencies ϵ_{acc} and counts of masked pixels at each cut stage; (ii) final mask maps and PCDs per CCD; (iii) extended diagnostics: hot-column scans, overdensity fits, and multiplicity histogram; (iv) single-pixel and multi-pixel pattern tables, with "Exp" vs. "Obs" for each CCD and for the total.

D.1 Extended Masking Efficiencies and Figures of Blinded Data

Each subsection includes a detailed table reporting the cumulative efficiency ϵ_{acc} and number of masked pixels n_{msk} at every step of the selection. These steps include clustering, hot column detection, overdensity removal, and filters based on charge quality, multiplicity and column isolation as described in Chapter 3.

For each dataset, the final masked-pixel maps and the associated PCDs is also shown. In addition, extended figures that illustrate the identification of structural anomalies are provided: Column-wise normalized occupancy profiles used to identify hot columns; Sliding-window scans employed in the detection of overdensity regions; and Multiplicity histograms in rows and columns used to reject structured residuals.

D.1.1 Dataset B1 (2024-10-14)

Table D.1 summarizes the cumulative masking efficiency and pixel counts at each selection step for dataset B1.

Mask	**CCD A**		**CCD B**		**CCD C**		**CCD D**	
	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}
Clustering+AR	0.9999	3108	0.9999	4108	0.9999	5929	0.9999	4787
Cluster Crosstalk	0.9996	17932	0.9996	17932	0.9996	17932	0.9996	17932
Hot Columns	0.9993	15040	0.9991	22560	0.9965	142880	0.9965	160809
Overdensity	0.9648	1594240	0.9991	0	0.8621	6309280	0.9474	2430990
High-Q Pixels	0.9560	418875	0.9884	500907	0.8506	650670	0.9341	3044504
Charge Multiplicity	0.9425	640015	0.9775	506844	0.8401	517206	0.9242	3503133
Isolated Columns	0.9424	7520	0.9773	7520	0.8401	0	0.9234	37600
Final	0.9424	2662121	0.9773	1046963	0.8401	7386197	0.9234	3540204

Table D.1: Cumulative efficiency ϵ_{acc} and number of masked pixels n_{msk} at each step of the data selection process for B1. Pixel counts refer to the active region (7520 rows · 6144 columns). Values of n_{msk} are per-mask (not cumulative).

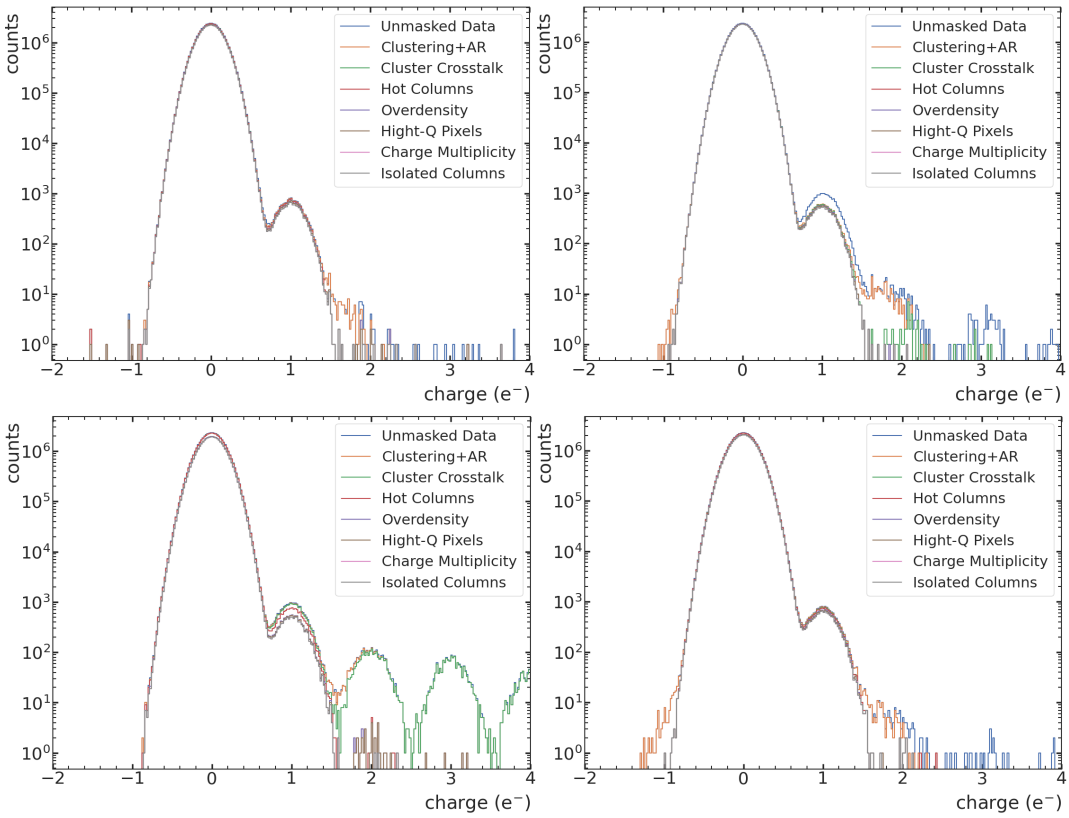


Figure D.1: Pixel ChargeDistribution (PCD) for B1. From top to bottom: A, B, C and D.

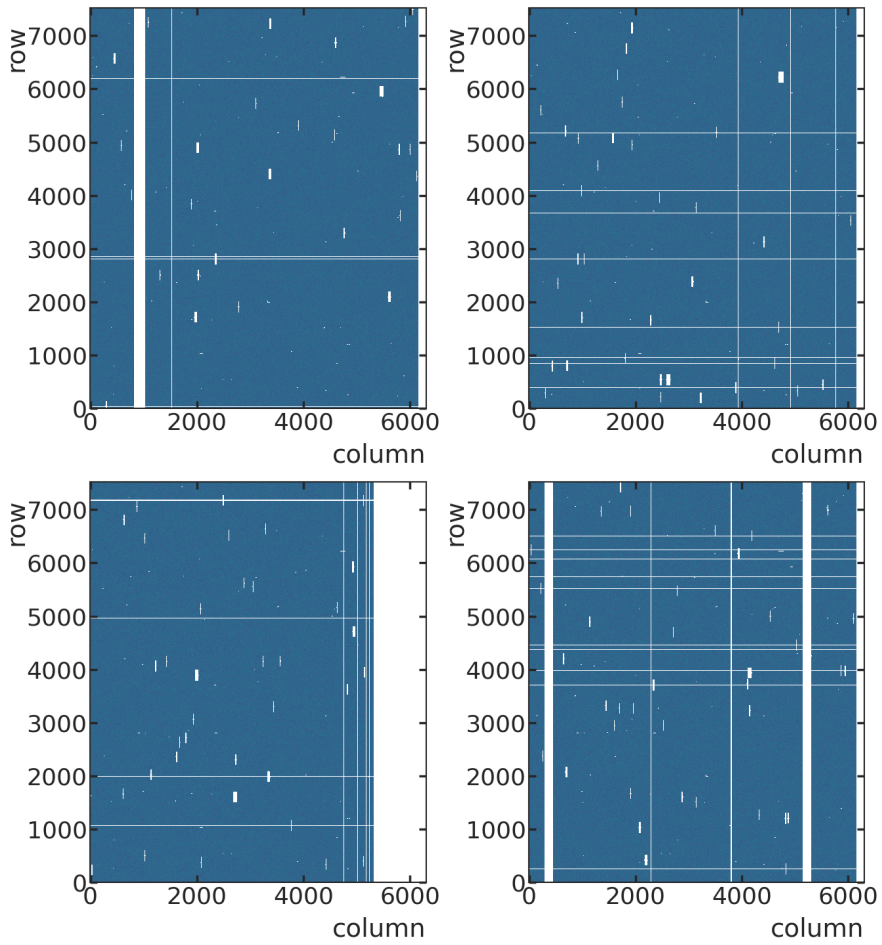


Figure D.2: Masked-pixel maps for B1. White regions are excluded pixels. From top to bottom: A, B, C and D.

Additional figures used in the identification of hot columns, overdensity regions, and row/column multiplicity anomalies for this dataset are shown in Figs. D.3, D.4 and D.5.

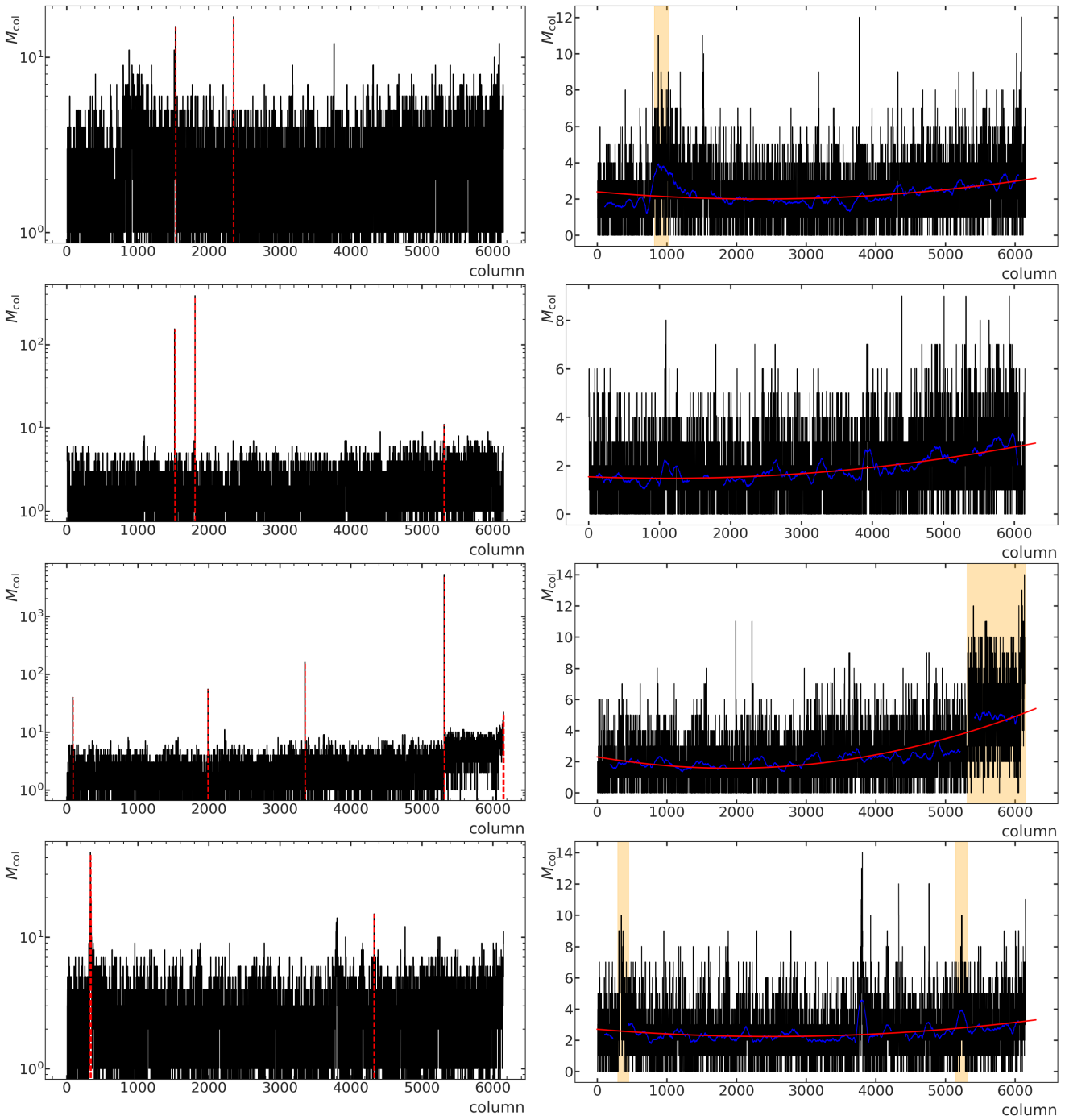


Figure D.3: Structural-defect identification for dataset B1. Left panels: hot-column scan with vertical red dashed lines indicating the columns removed. Right panels: column–multiplicity profiles; the red curve is the parabolic baseline and the orange band marks the detected overdensity. From top to bottom: CCDs A, B, C and D.

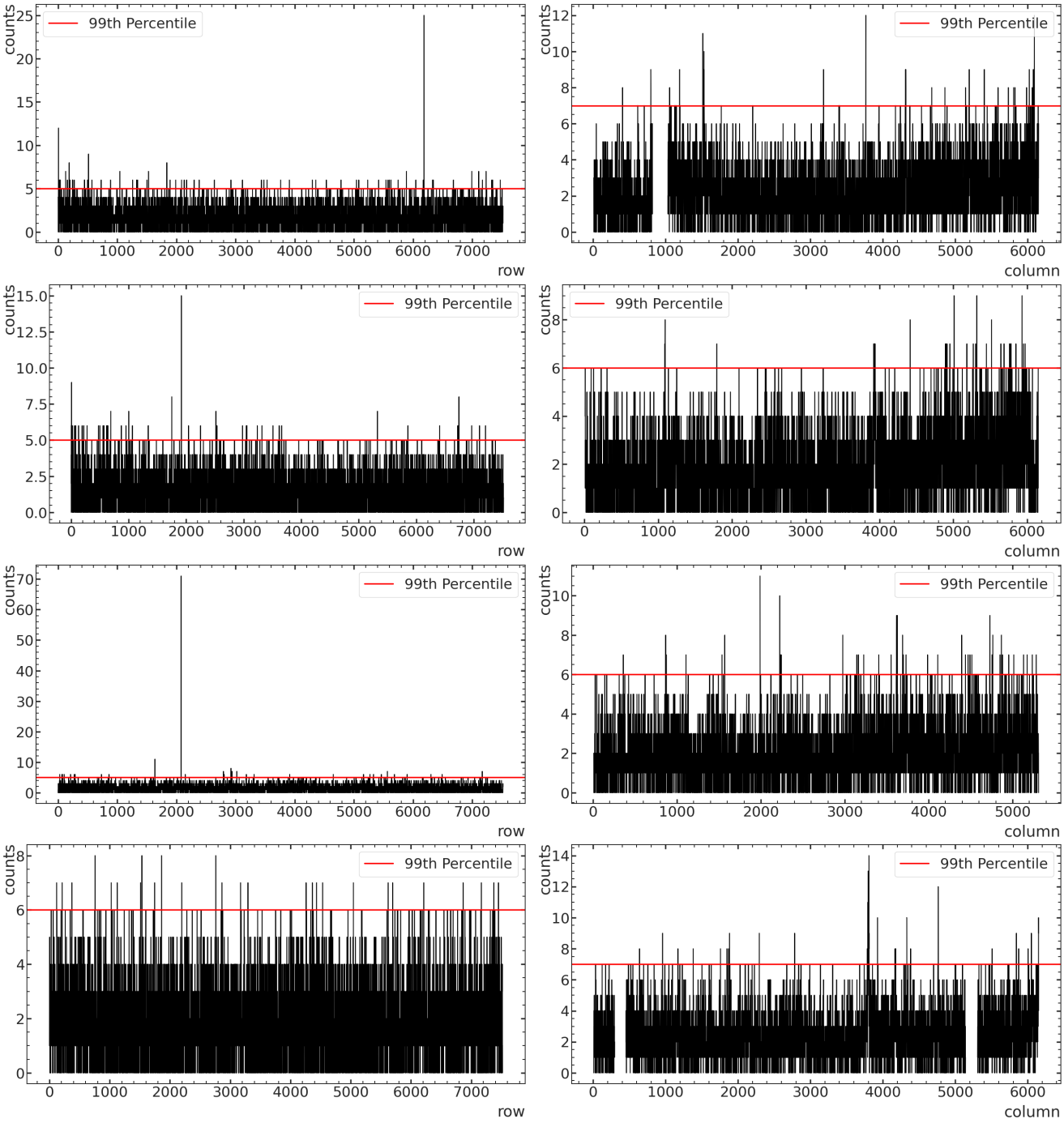


Figure D.4: Charged–pixel multiplicity distributions per row (left) and per column (right) for dataset B1. Red horizontal lines mark the masking thresholds. From top to bottom: CCDs A, B, C and D.

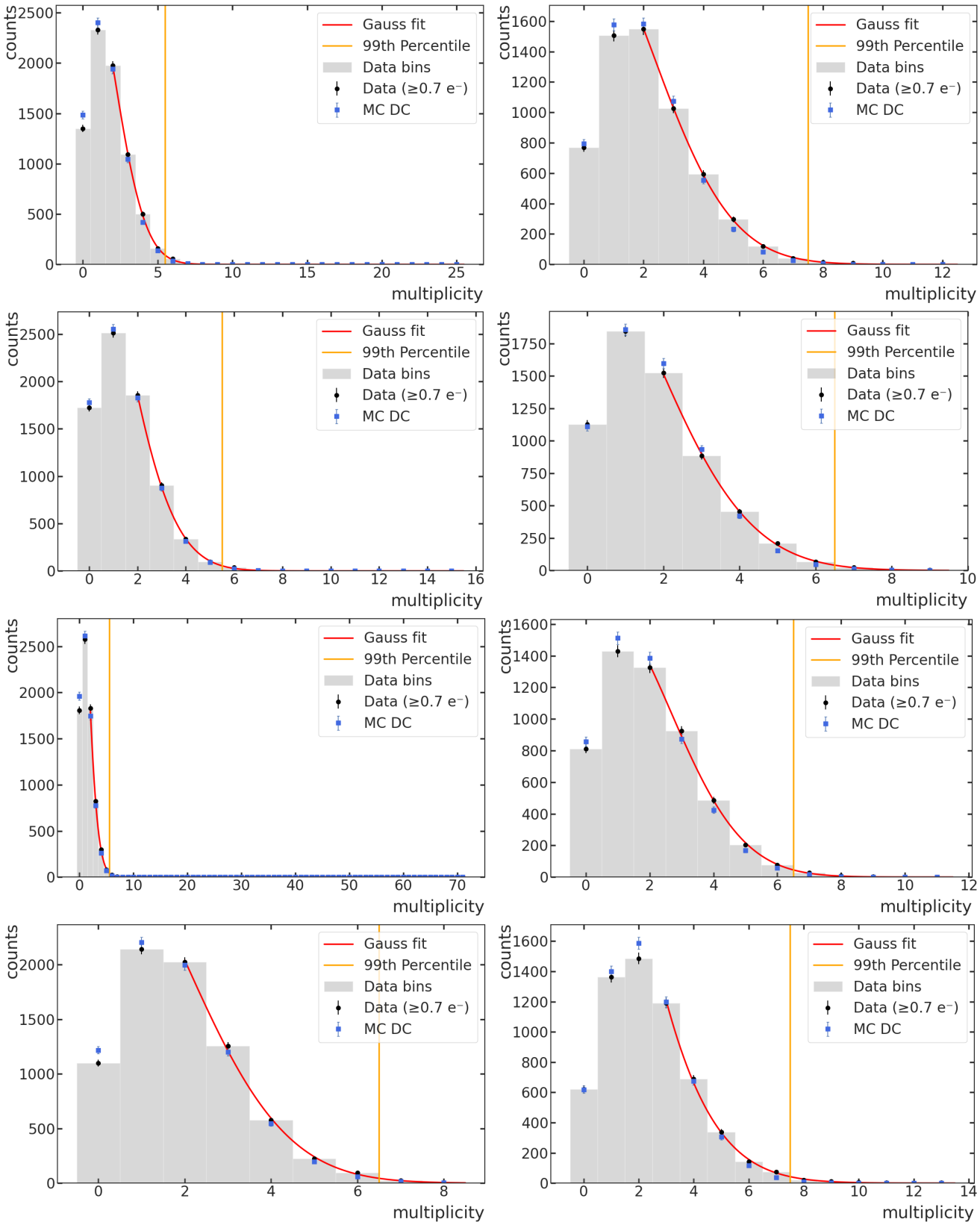


Figure D.5: Charged–pixel multiplicity histograms per row (left) and per column (right) for dataset B1. Gaussian fits (red) are applied to the upper tail. Orange line indicates the 99th percentile threshold. From top to bottom: CCDs A, B, C and D.

After masking, the four PCDs are well described by a Poisson–Gaussian convolution (Fig. D.6). The fit parameters are collected in Table D.2; charge-resolution values cluster around $\sigma \simeq 0.16\ e^-$, matching UB0. The mean dark-current rates span 3.46×10^2–$4.55 \times 10^2\ e^-\,\mathrm{g}^{-1}\,\mathrm{day}^{-1}$, fully compatible (within 10 %) with the UB0 reference.

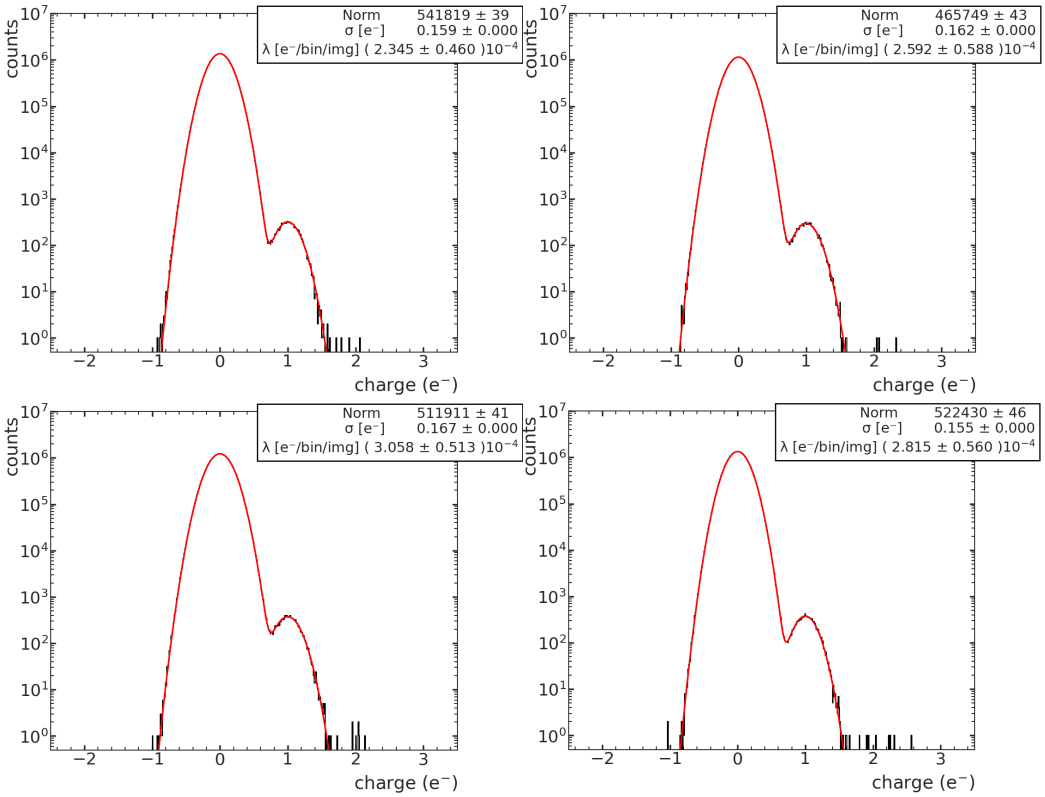


Figure D.6: PCDs for dataset B1 after all masking. The black histograms represent the charge distributions after all masking procedures. The red fit corresponds to the convolution of Poisson and Gaussian distributions (Table D.2). Sorted top to bottom: CCD A, B, C and D, respectively.

CCD	**e⁻/bin/img**	**e⁻/pix/day**	**e⁻/g/day**	σ
CCD A	$(2.8 \pm 0.6) \cdot 10^{-4}$	$(1.5 \pm 0.3) \cdot 10^{-4}$	415 ± 80	0.155 ± 0.000
CCD B	$(2.3 \pm 0.5) \cdot 10^{-4}$	$(1.2 \pm 0.2) \cdot 10^{-4}$	346 ± 70	0.159 ± 0.000
CCD C	$(2.6 \pm 0.6) \cdot 10^{-4}$	$(1.3 \pm 0.3) \cdot 10^{-4}$	382 ± 90	0.162 ± 0.000
CCD D	$(3.1 \pm 0.5) \cdot 10^{-4}$	$(1.6 \pm 0.3) \cdot 10^{-4}$	451 ± 80	0.167 ± 0.000

Table D.2: Poisson–Gaussian fit parameters for dataset B1, including statistical uncertainties. The values correspond to the average charge per pixel per image (e⁻/bin/img), per day (e⁻/pix/day), and normalized per gram per day (e⁻/g/day), along with the charge resolution parameter σ.

D.1.2 Dataset B2 (2024-10-28)

Table D.3 summarizes the cumulative masking efficiency and number of masked pixels at each step of the selection for dataset B2.

Mask	CCD A		CCD B		CCD C		CCD D	
	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}
Clustering+AR	0.9999	5947	0.9999	7061	0.9999	11223	0.9999	11890
Cluster Crosstalk	0.9997	36121	0.9997	36121	0.9997	36121	0.9997	36121
Hot Columns	0.9987	592416	0.9987	107712	0.9958	430848	0.9915	897600
Overdensity	0.9478	5618976	0.9238	8275872	0.8626	15043776	0.9915	0
High-Q Pixels	0.9401	892419	0.9157	971507	0.8511	1449164	0.9781	1502261
Charge Multiplicity	0.9242	1697806	0.9081	865755	0.8423	1043732	0.9682	1102632
Isolated Columns	0.9236	71808	0.9073	89760	0.8418	53856	0.9676	71808
Final	0.9236	8430401	0.9073	10225977	0.8418	17449064	0.9676	3750483

Table D.3: Cumulative efficiency ϵ_{acc} and number of masked pixels n_{msk} at each step of the data selection process for the B2 dataset. Pixel counts refer to the active region (14960 rows · 6144 columns). Values of n_{msk} are per-mask (not cumulative).

Figures D.7 and D.8 show the final masked-pixel maps and the corresponding PCDs for each CCD:

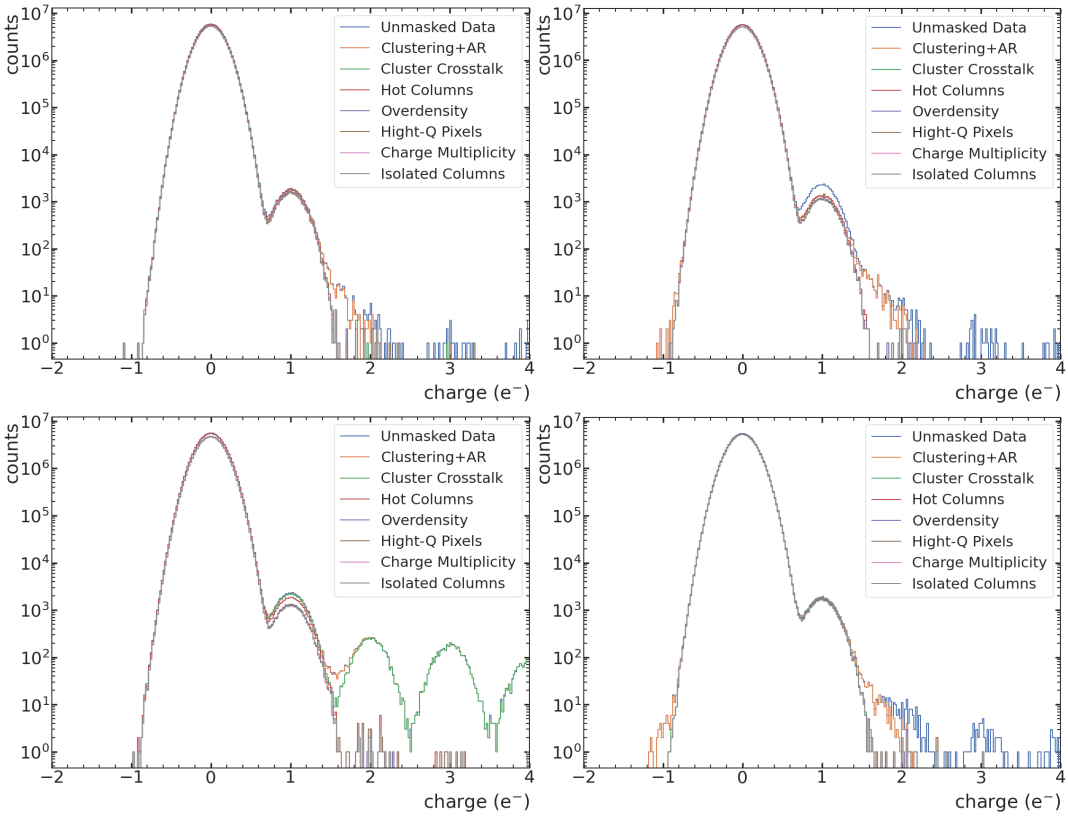


Figure D.7: Pixel Charge Distribution (PCD) for B2. From top to bottom: A, B, C and D.

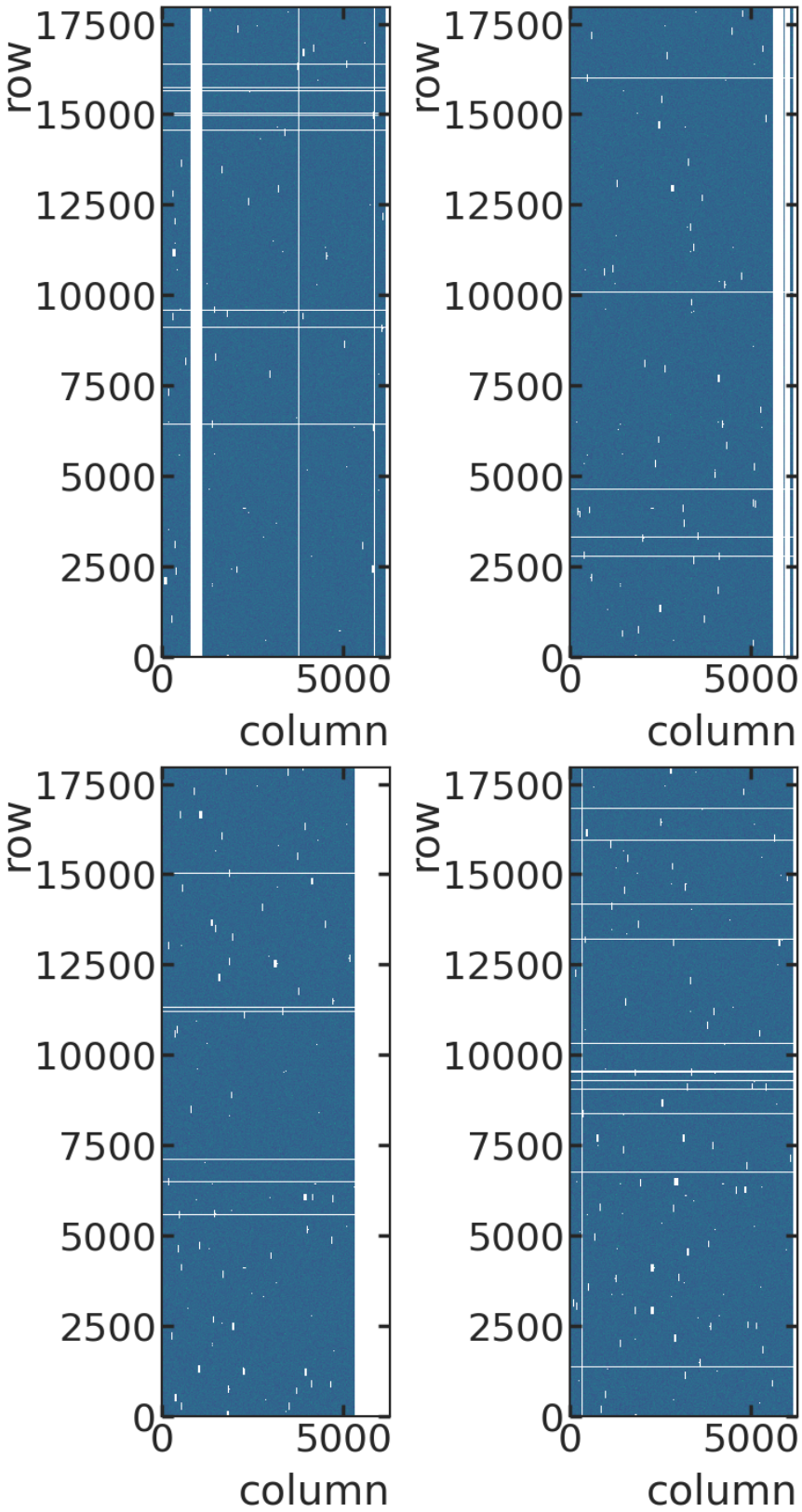


Figure D.8: Masked-pixel occupancy maps for B2. White regions denote masked pixels. From top to bottom: A, B, C and D.

Additional figures used in the identification of hot columns, overdensity regions, and row/column multiplicity anomalies for this dataset are shown in Figs. D.9, D.10 and D.11.

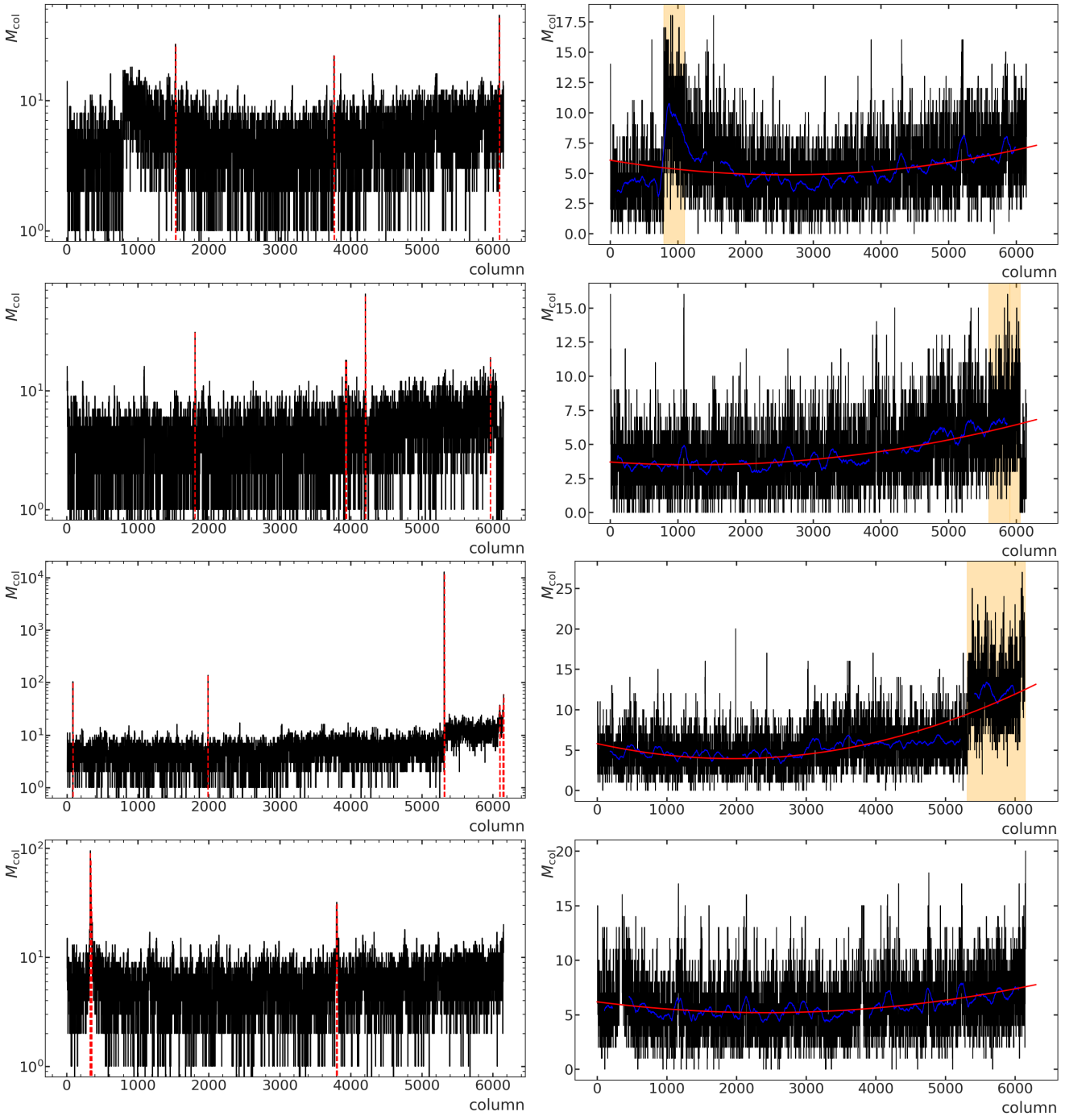


Figure D.9: Structural-defect identification for dataset B2. Left: hot-column scan with vertical red dashed lines indicating the columns removed. Right: column–multiplicity profiles; the red curve is the parabolic baseline and the orange band marks the detected overdensity. From top to bottom: CCDs A, B, C and D.

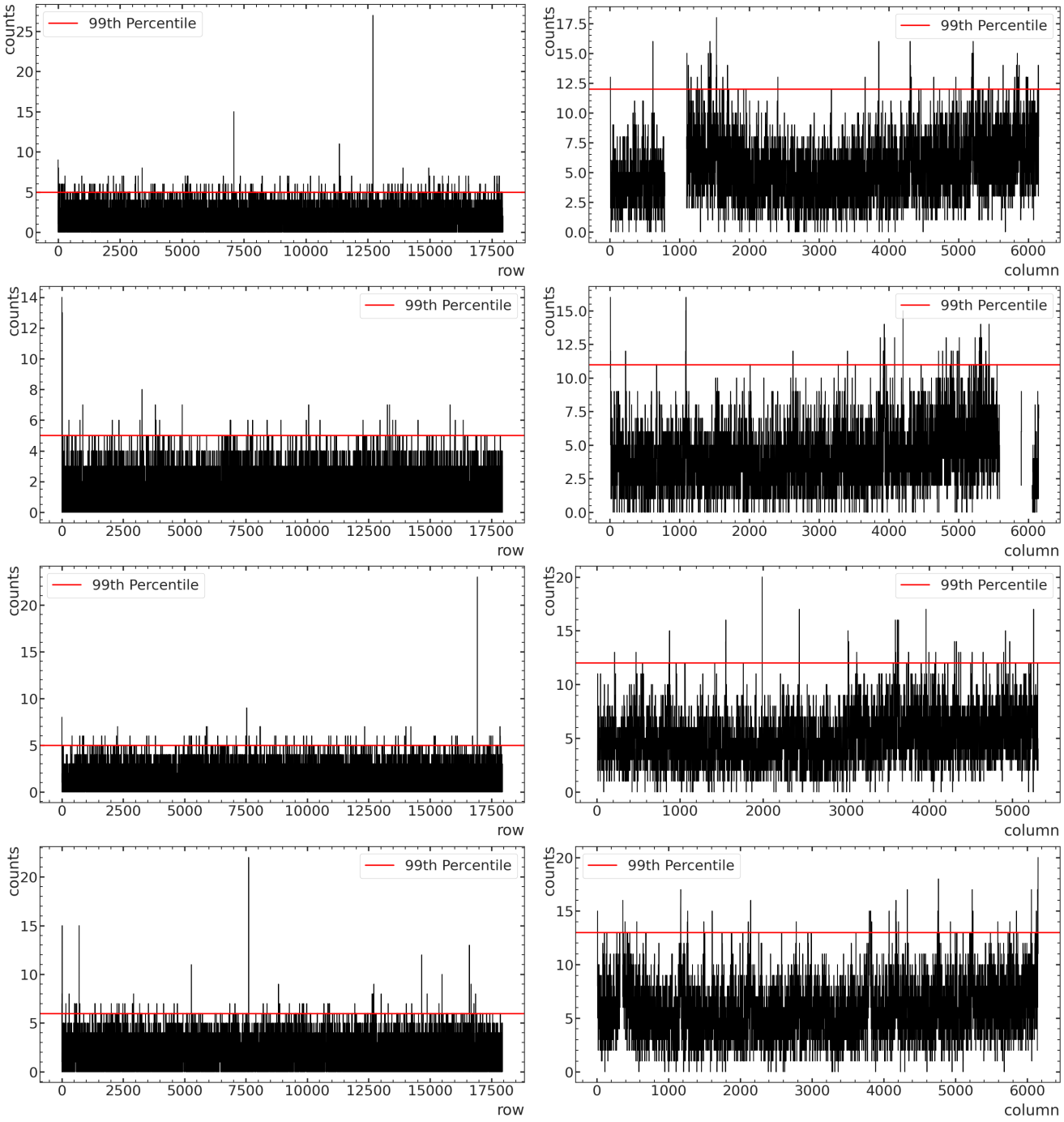


Figure D.10: Charged–pixel multiplicity distributions per row (left) and per column (right) for dataset B2. Red horizontal lines mark the masking thresholds. From top to bottom: CCDs A, B, C and D.

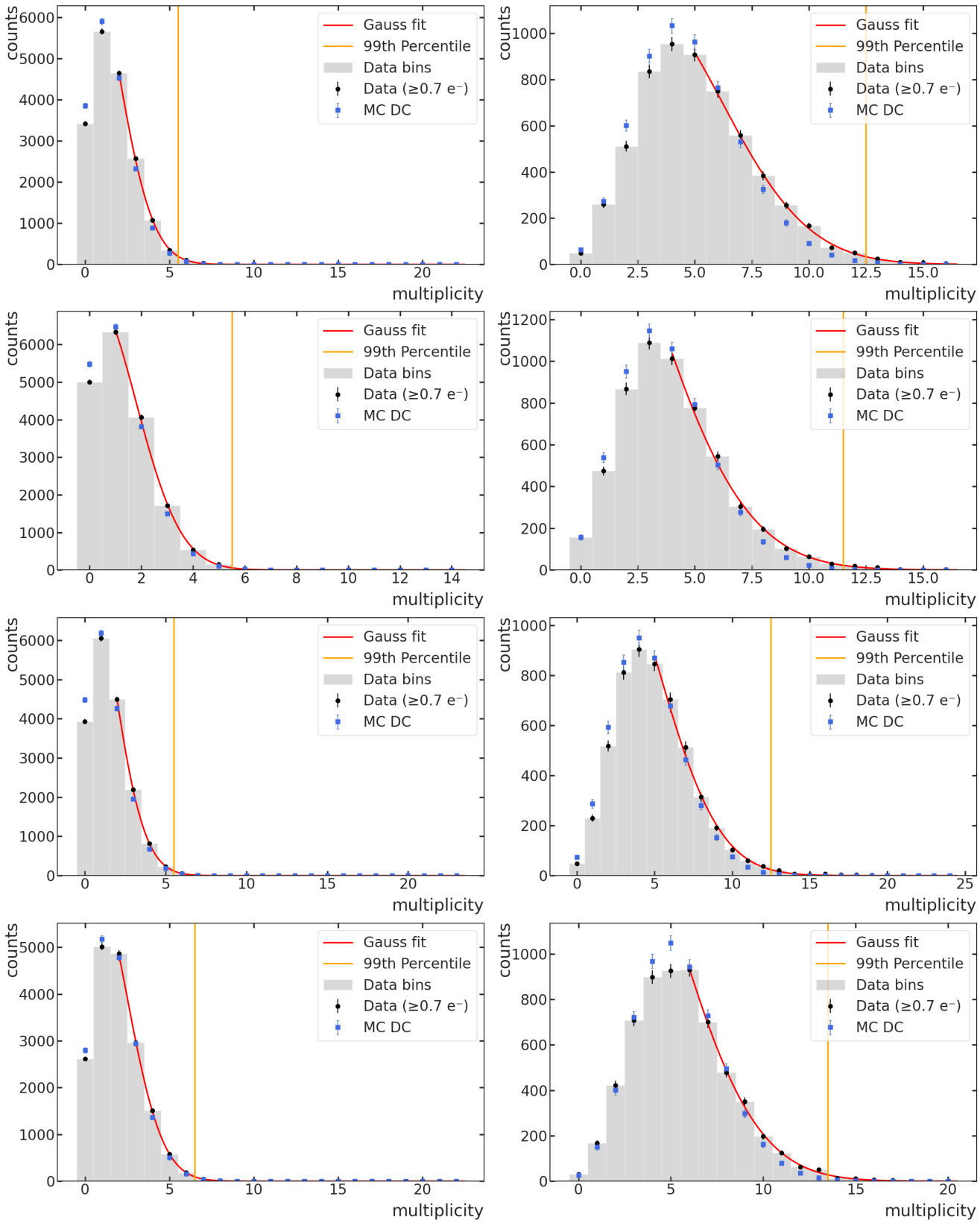


Figure D.11: Charged–pixel multiplicity histograms per row (left) and per column (right) for dataset B2. Gaussian fits (red) are applied to the upper tail. Orange line indicates the 99th percentile threshold. From top to bottom: CCDs A, B, C and D.

After masking, the four Pixel Charge Distributions (PCDs) are well described by a Poisson–Gaussian convolution (Fig. D.12). Fit parameters-summarized in Table D.4- show charge resolutions clustered around $\sigma \simeq 0.16\, e^-$ and mean dark-current rates that agree with those of B2 to better than 5 %.

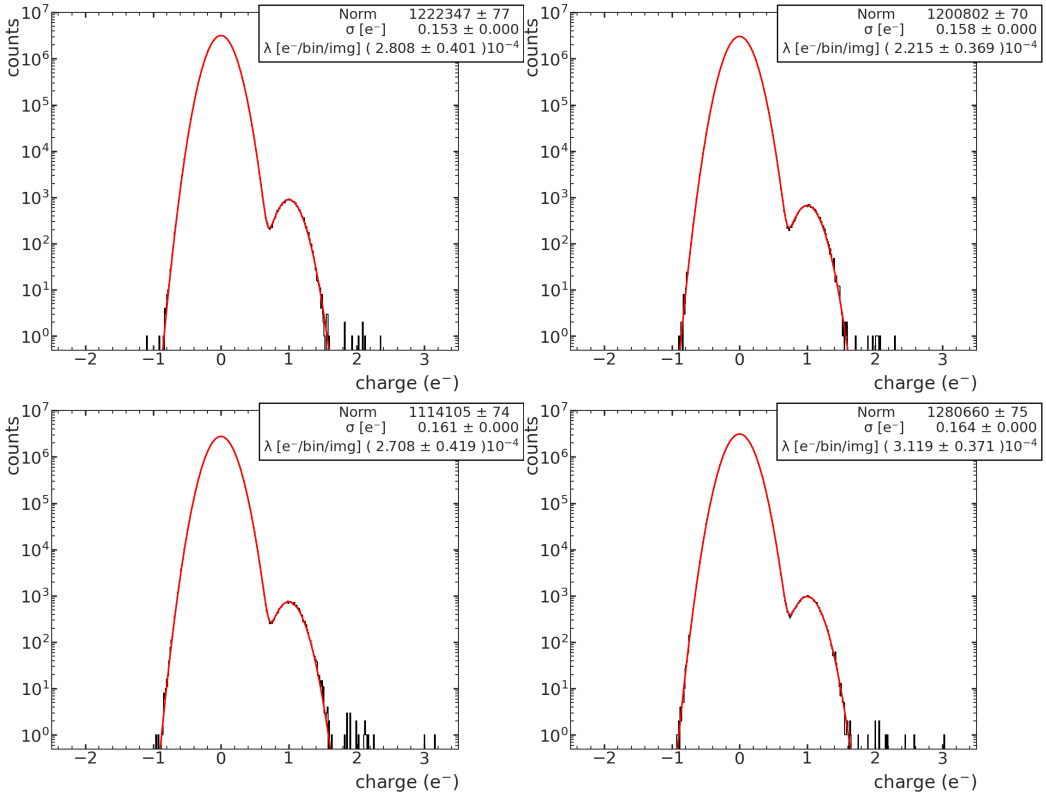


Figure D.12: PCDs for dataset B2 after all masking. The black histograms represent the charge distributions after all masking procedures. The red fit corresponds to the convolution of Poisson and Gaussian distributions (Table D.4). Sorted top to bottom: CCD A, B, C and D, respectively.

CCD	**e$^-$/bin/img**	**e$^-$/pix/day**	**e$^-$/g/day**	σ
CCD A	$(2.8 \pm 0.4) \cdot 10^{-4}$	$(1.5 \pm 0.2) \cdot 10^{-4}$	414 ± 60	0.153 ± 0.000
CCD B	$(2.2 \pm 0.4) \cdot 10^{-4}$	$(1.15 \pm 0.19) \cdot 10^{-4}$	327 ± 50	0.158 ± 0.000
CCD C	$(2.7 \pm 0.4) \cdot 10^{-4}$	$(1.40 \pm 0.2) \cdot 10^{-4}$	400 ± 60	0.161 ± 0.000
CCD D	$(3.1 \pm 0.4) \cdot 10^{-4}$	$(1.61 \pm 0.19) \cdot 10^{-4}$	459 ± 50	0.164 ± 0.000

Table D.4: Poisson–Gaussian fit parameters for dataset B2, including statistical uncertainties. The values correspond to the average charge per pixel per image (e$^-$/bin/img), per day (e$^-$/pix/day), and normalized per gram per day (e$^-$/g/day), along with the charge resolution parameter σ.

D.1.3 Dataset B3 (2024-11-19)

Table D.5 summarizes the cumulative masking efficiency and number of masked pixels at each step of the selection for dataset B3.

Mask	CCD A		CCD B		CCD C		CCD D	
	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}
Clustering+AR	0.9999	7916	0.9999	12341	0.9999	18889	0.9999	16279
Cluster Crosstalk	0.9996	55425	0.9996	55425	0.9996	55425	0.9996	55425
Hot Columns	0.9983	177408	0.9967	399168	0.9941	753984	0.9923	997920
Overdensity	0.9526	6275808	0.9387	7894656	0.8623	18650016	0.9923	0
High-Q Pixels	0.9446	1151739	0.9277	1595704	0.8473	2333495	0.9774	2055706
Charge Multiplicity	0.9294	2132256	0.9178	1386739	0.8376	1440684	0.9653	1677244
Isolated Columns	0.9286	110880	0.9173	66528	0.8371	66528	0.9641	155232
Final	0.9286	9723762	0.9173	11261760	0.8371	22191666	0.9641	4886641

Table D.5: Cumulative efficiency ϵ_{acc} and number of masked pixels n_{msk} at each step of the data selection process for the B3 dataset. Pixel counts refer to the active region (22176 rows · 6144 columns). Values of n_{msk} are per-mask (not cumulative).

Figures D.13 and D.14 show the final masked-pixel maps and the corresponding PCDs for each CCD:

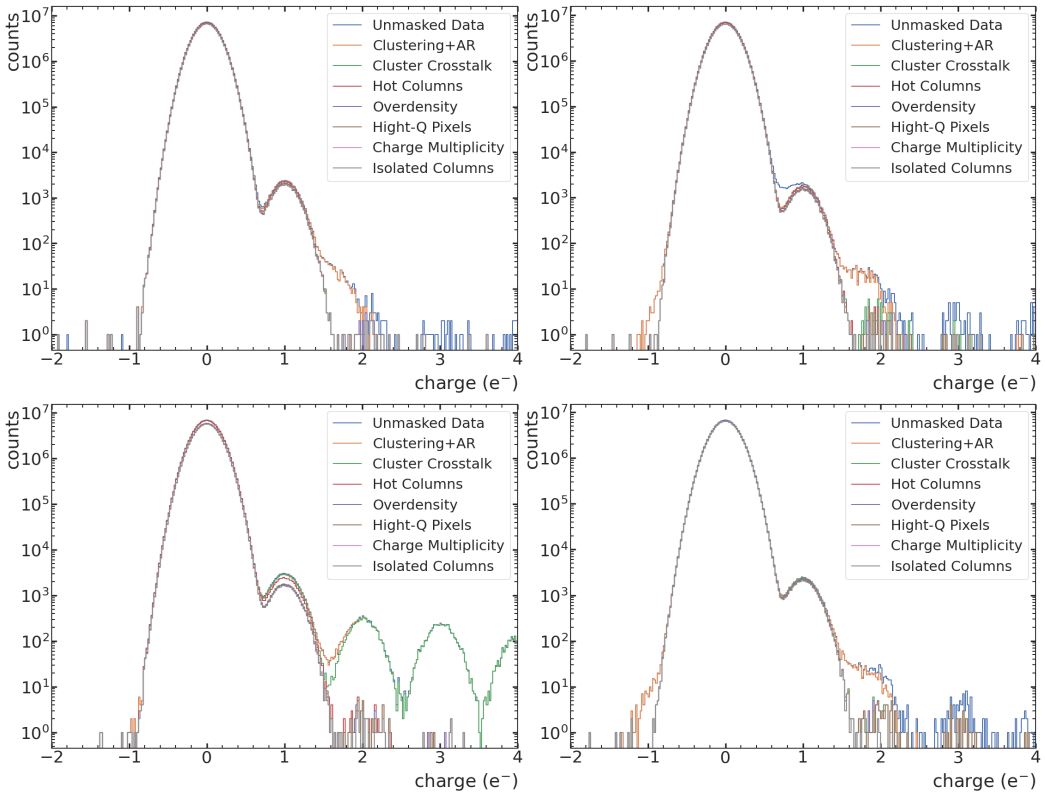


Figure D.13: Pixel ChargeDistribution (PCD) for B3. From top to bottom: A, B, C and D.

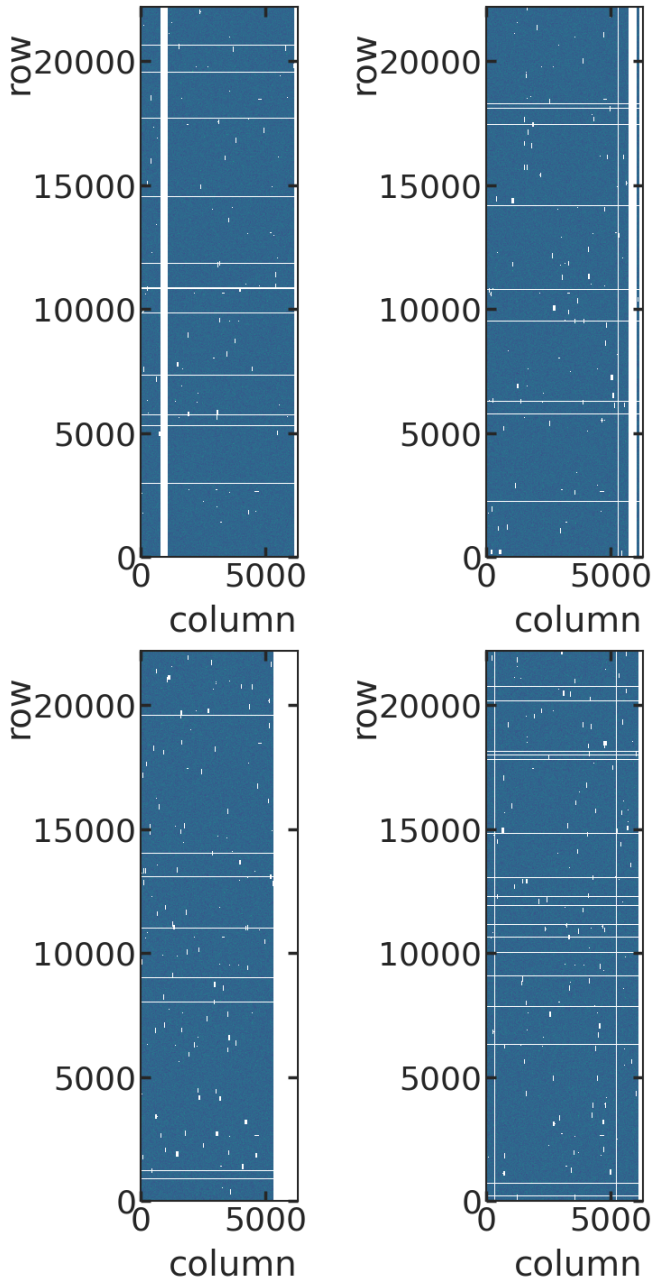


Figure D.14: Masked-pixel occupancy maps for B3. White areas denote masked pixels. From top to bottom: A, B, C and D.

Additional diagnostics for hot-column detection (Fig. D.15), overdensity removal and multiplicity anomalies (Figs. D.16–D.17) follow.

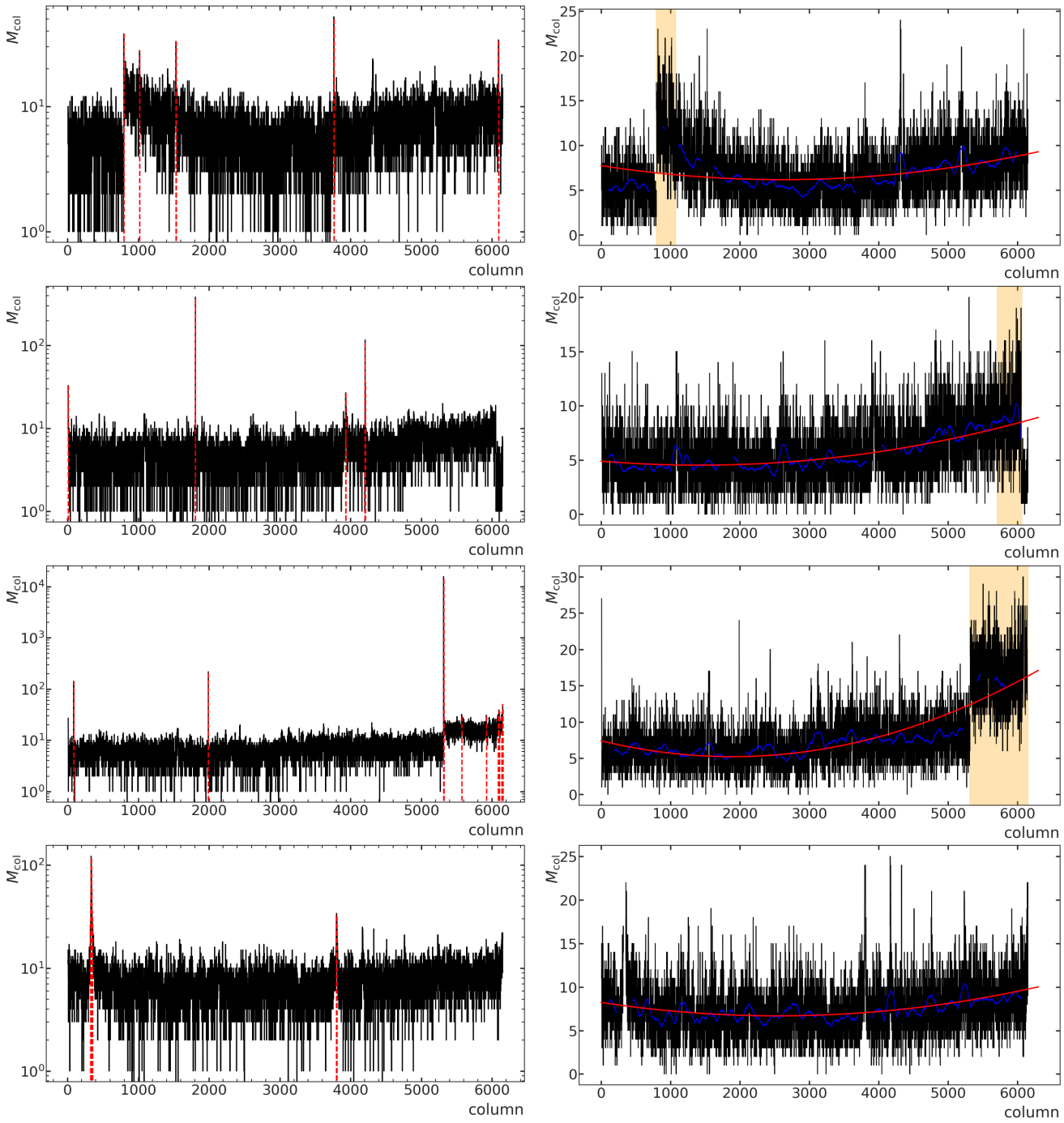


Figure D.15: Structural-defect identification for dataset B3. Left: hot-column scan with vertical red dashed lines indicating the columns removed. Right: Column occupancy profiles; the red curve is the parabolic baseline and the orange band marks the detected overdensity. From top to bottom: CCDs A, B, C and D.

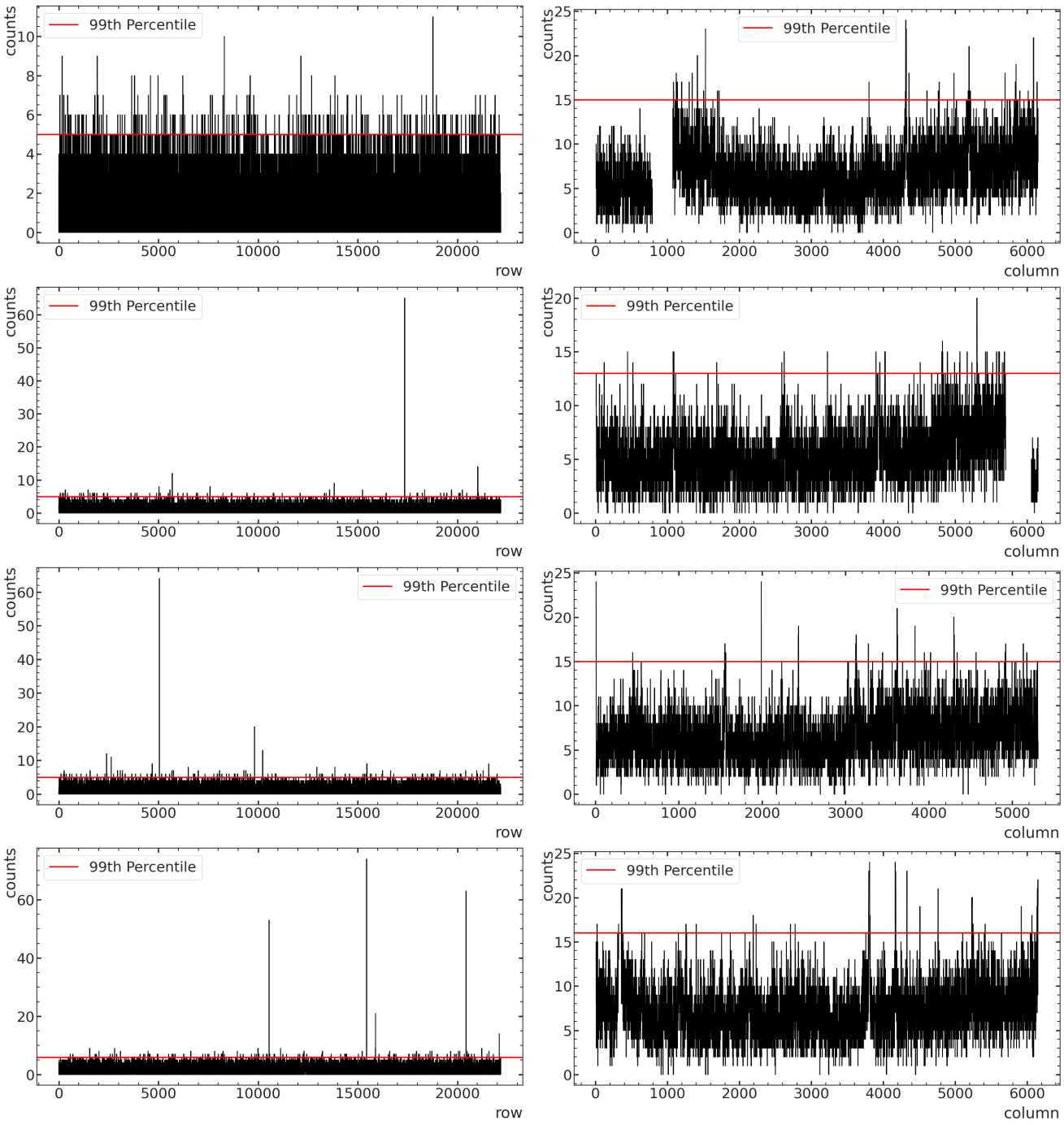


Figure D.16: Charged–pixel multiplicity distributions per row (left) and per column (right) for dataset B3. Red horizontal lines mark the masking thresholds. From top to bottom: CCDs A, B, C and D.

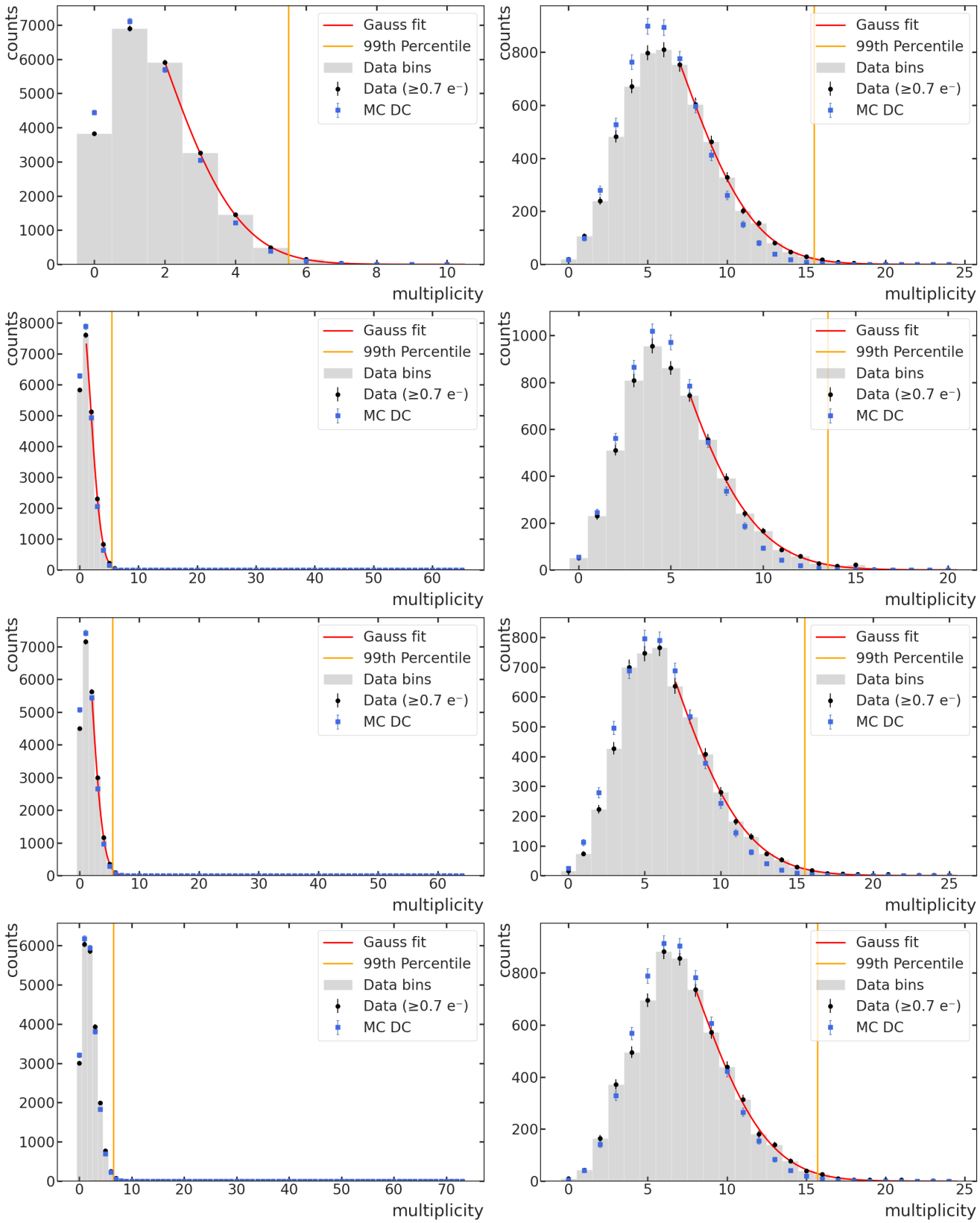


Figure D.17: Charged–pixel multiplicity histograms per row (left) and per column (right) for dataset B3. Gaussian fits (red) are applied to the upper tail. Orange line indicates the 99th percentile threshold. From top to bottom: CCDs A, B, C and D.

Finally, the Poisson–Gaussian fits to the PCDs are shown in Fig. D.18, and the fit parameters are collected in Table D.6; charge resolutions cluster around $\sigma \simeq 0.16\ e^-$ and the mean dark-current rates remain consistent with previous runs, the highest being $4.74 \times 10^2\ e^-\ \mathrm{g}^{-1}\ \mathrm{day}^{-1}$ in CCD D.

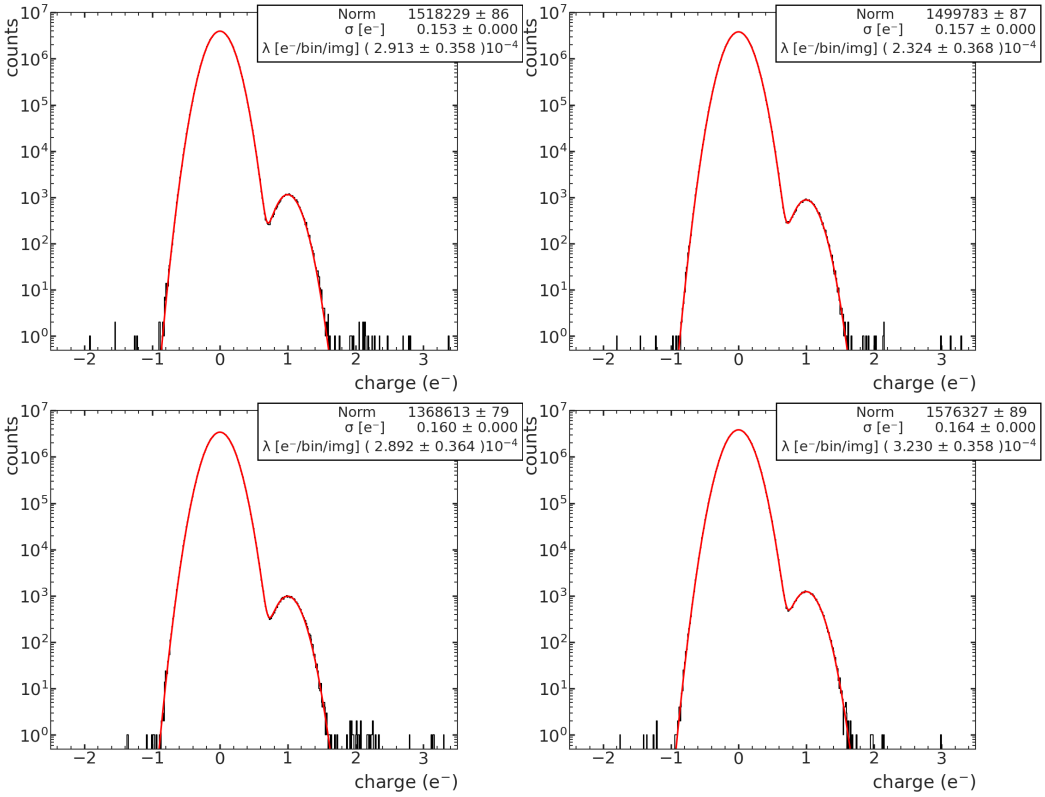


Figure D.18: PCDs for dataset B3 after all masking, with Poisson–Gaussian fit to two picks (red) from Table D.6. Sorted top to bottom: CCD A, B, C and D, respectively.

CCD	**e⁻/bin/img**	**e⁻/pix/day**	**e⁻/g/day**	σ
CCD A	$(2.9 \pm 0.4) \cdot 10^{-4}$	$(1.51 \pm 0.19) \cdot 10^{-4}$	429 ± 50	0.153 ± 0.000
CCD B	$(2.3 \pm 0.4) \cdot 10^{-4}$	$(1.20 \pm 0.19) \cdot 10^{-4}$	342 ± 50	0.157 ± 0.000
CCD C	$(2.9 \pm 0.4) \cdot 10^{-4}$	$(1.50 \pm 0.19) \cdot 10^{-4}$	427 ± 50	0.160 ± 0.000
CCD D	$(3.2 \pm 0.4) \cdot 10^{-4}$	$(1.66 \pm 0.19) \cdot 10^{-4}$	474 ± 50	0.164 ± 0.000

Table D.6: Poisson–Gaussian fit parameters for dataset B3, including statistical uncertainties. The values correspond to the average charge per pixel per image (e⁻/bin/img), per day (e⁻/pix/day), and normalized per gram per day (e⁻/g/day), along with the charge resolution parameter σ.

D.1.4 Dataset B4 (2024-12-17)

Table D.7 summarizes the cumulative masking efficiency and number of masked pixels at each step of the selection for dataset B4.

Mask	CCD A		CCD B		CCD C		CCD D	
	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}	ϵ_{acc}	n_{msk}
Clustering+AR	0.9999	5109	0.9999	10411	0.9999	13454	0.9999	12013
Cluster Crosstalk	0.9996	40936	0.9996	40936	0.9996	40936	0.9996	40936
Hot Columns	0.9988	80480	0.9993	32192	0.9924	708224	0.9903	917472
Overdensity	0.9524	4619552	0.9451	5359968	0.8602	13552832	0.9903	0
High-Q Pixels	0.9452	748545	0.9324	1339118	0.8453	1746571	0.9754	1508488
Charge Multiplicity	0.9283	1722660	0.9202	1236216	0.8354	1045174	0.9606	1489650
Isolated Columns	0.9278	48288	0.9194	80480	0.8351	32192	0.9601	48288
Final	0.9278	7139111	0.9194	7971351	0.8351	16306637	0.9601	3946779

Table D.7: Cumulative efficiency ϵ_{acc} and number of masked pixels n_{msk} at each step of the data selection process for the B4 dataset. Pixel counts refer to the active region (16096 rows · 6144 columns). Values of n_{msk} are per-mask (not cumulative).

Figures D.19 and D.20 show the final masked-pixel maps and the corresponding PCDs for each CCD:

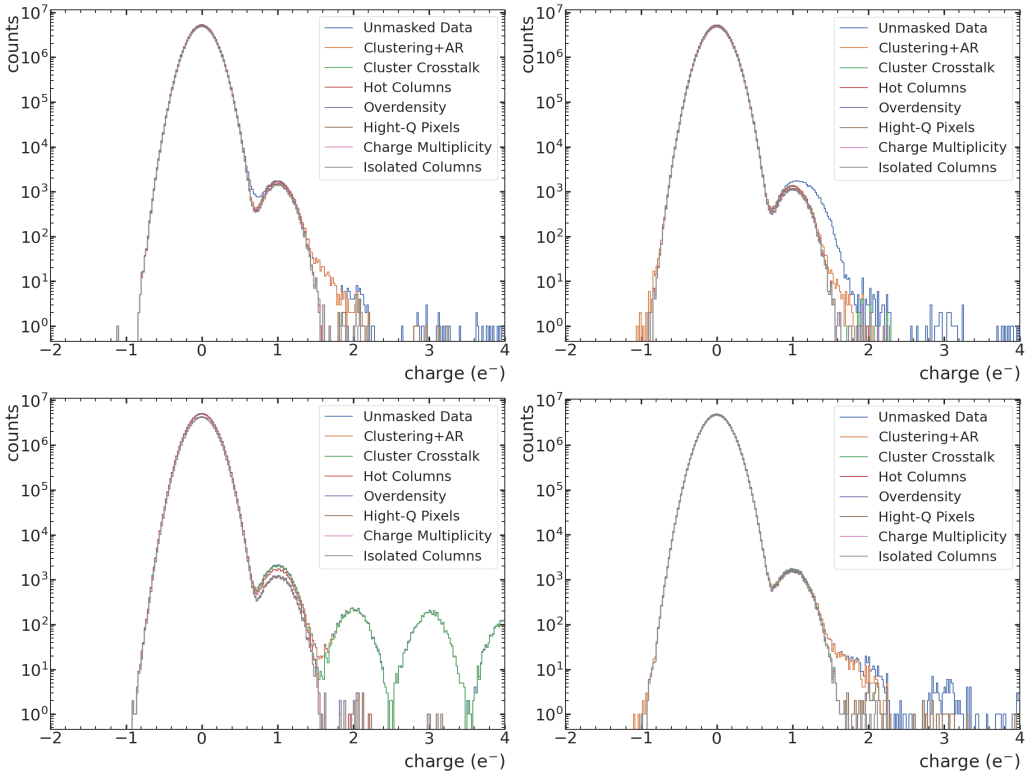


Figure D.19: PCDs for B4. From top to bottom: A, B, C and D.

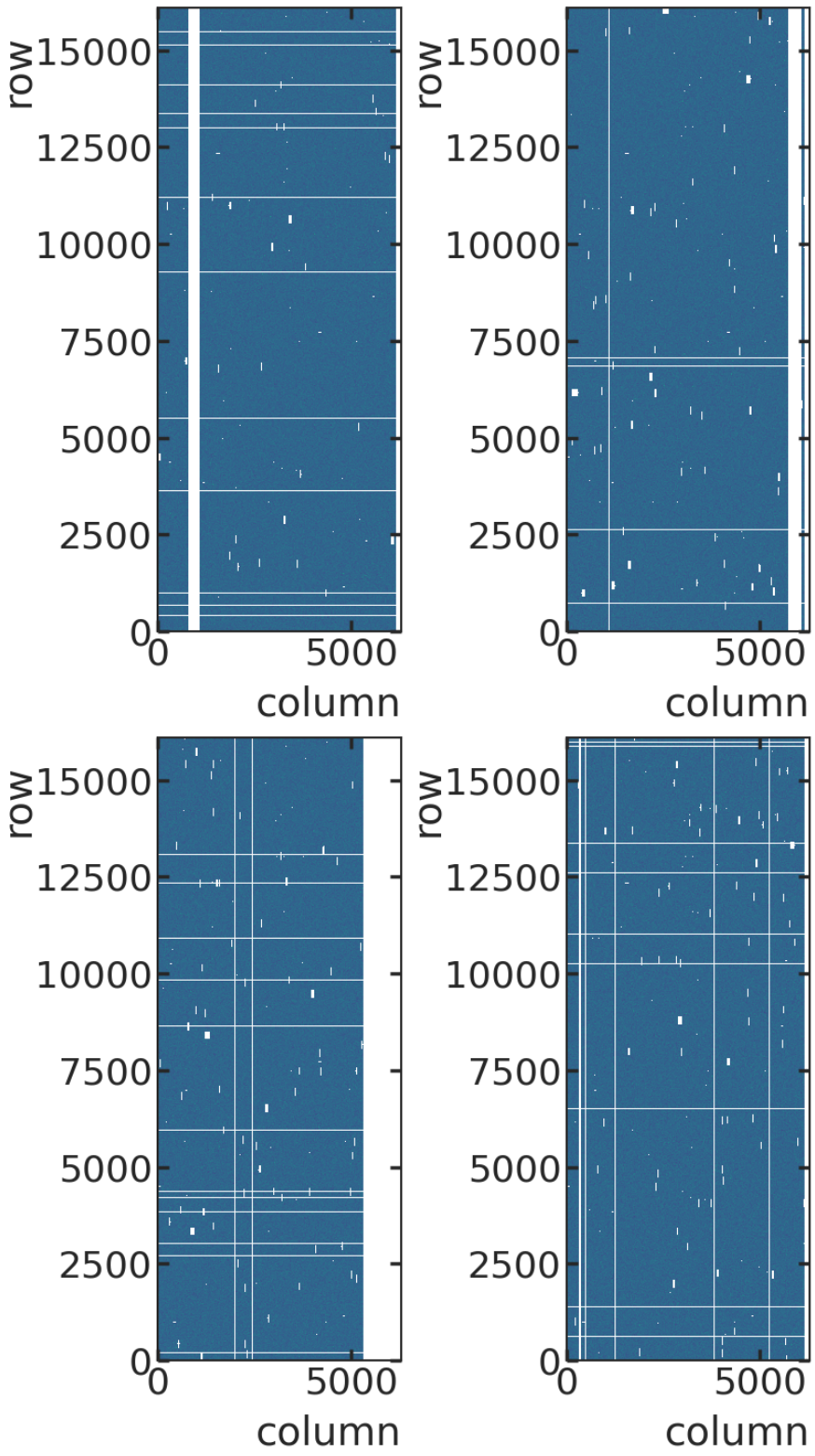


Figure D.20: Masked-pixel maps for B4. White regions are excluded pixels. From top to bottom: A, B, C and D.

Additional figures for hot-column detection (Fig. D.21), overdensity regions and row/columnmultiplicity anomalies (Figs. D.22–D.23) follow.

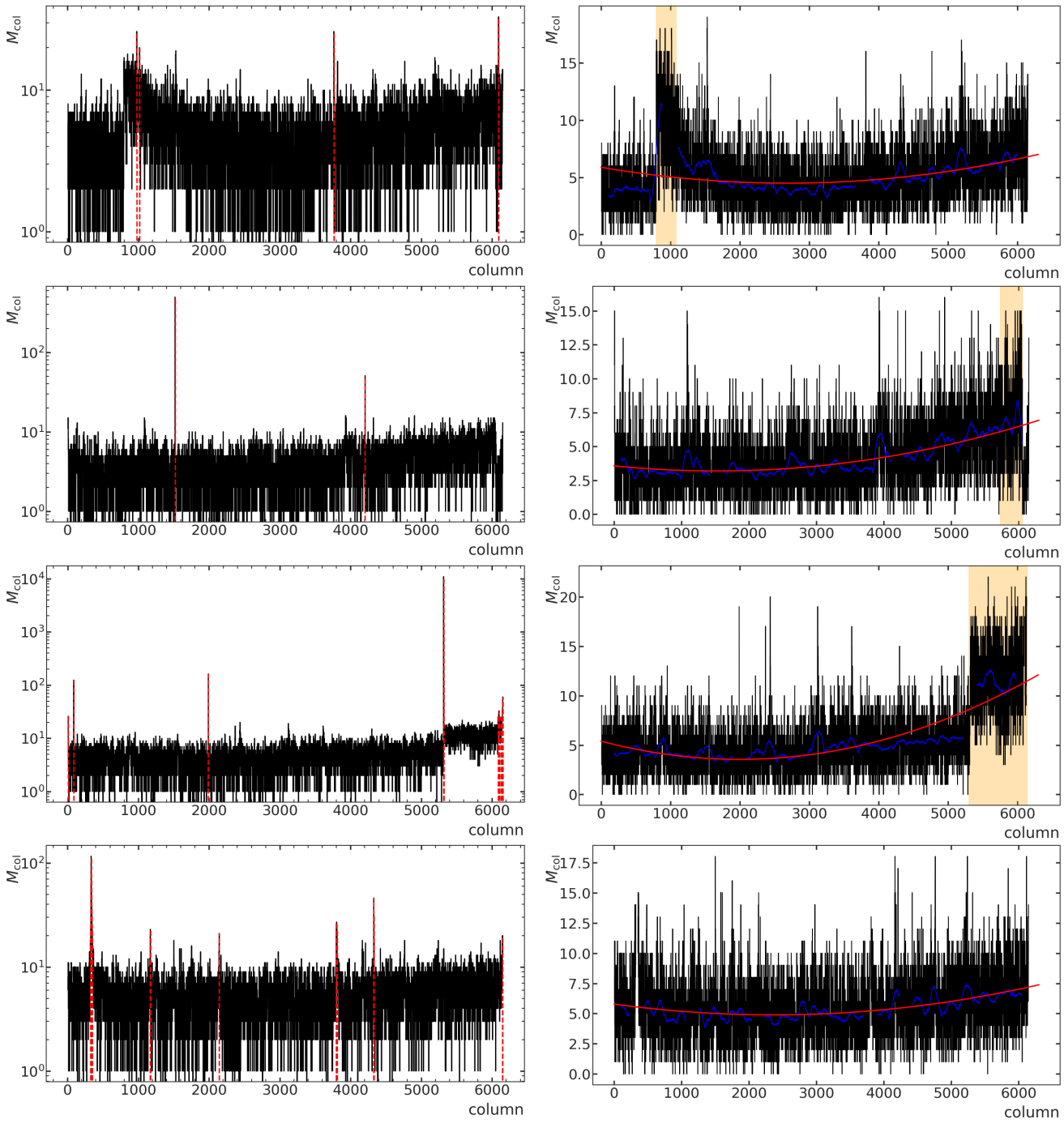


Figure D.21: Structural-defect identification for dataset B4. Left: hot-column scan with vertical red dashed lines indicating the columns removed. Right: column–multiplicity profiles; the red curve is the parabolic baseline and the orange band marks the detected overdensity. From top to bottom: CCDs A, B, C and D.

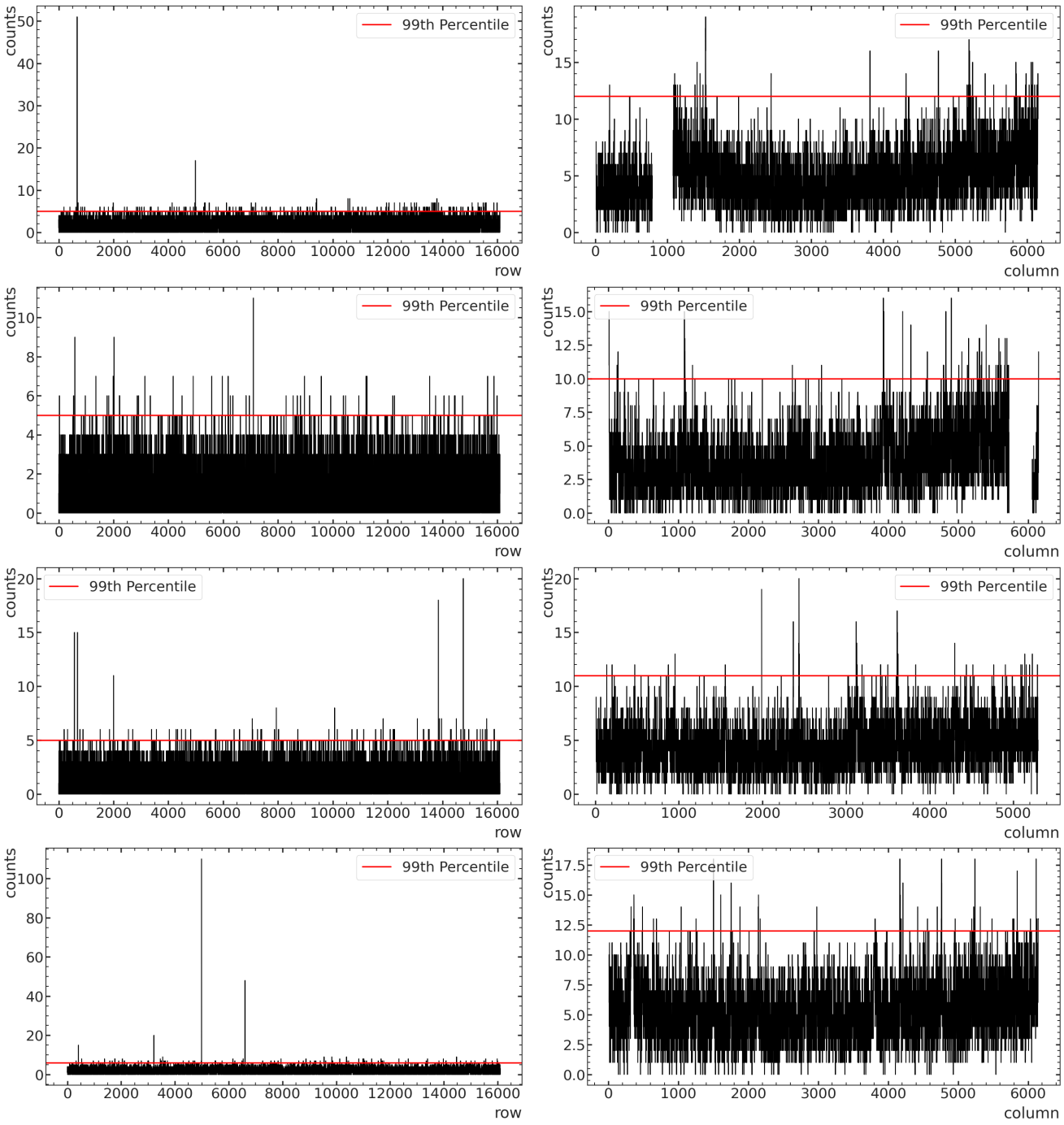


Figure D.22: Charged–pixel multiplicity distributions per row (left) and per column (right) for dataset B4. Red horizontal lines mark the masking thresholds. From top to bottom: CCDs A, B, C and D.

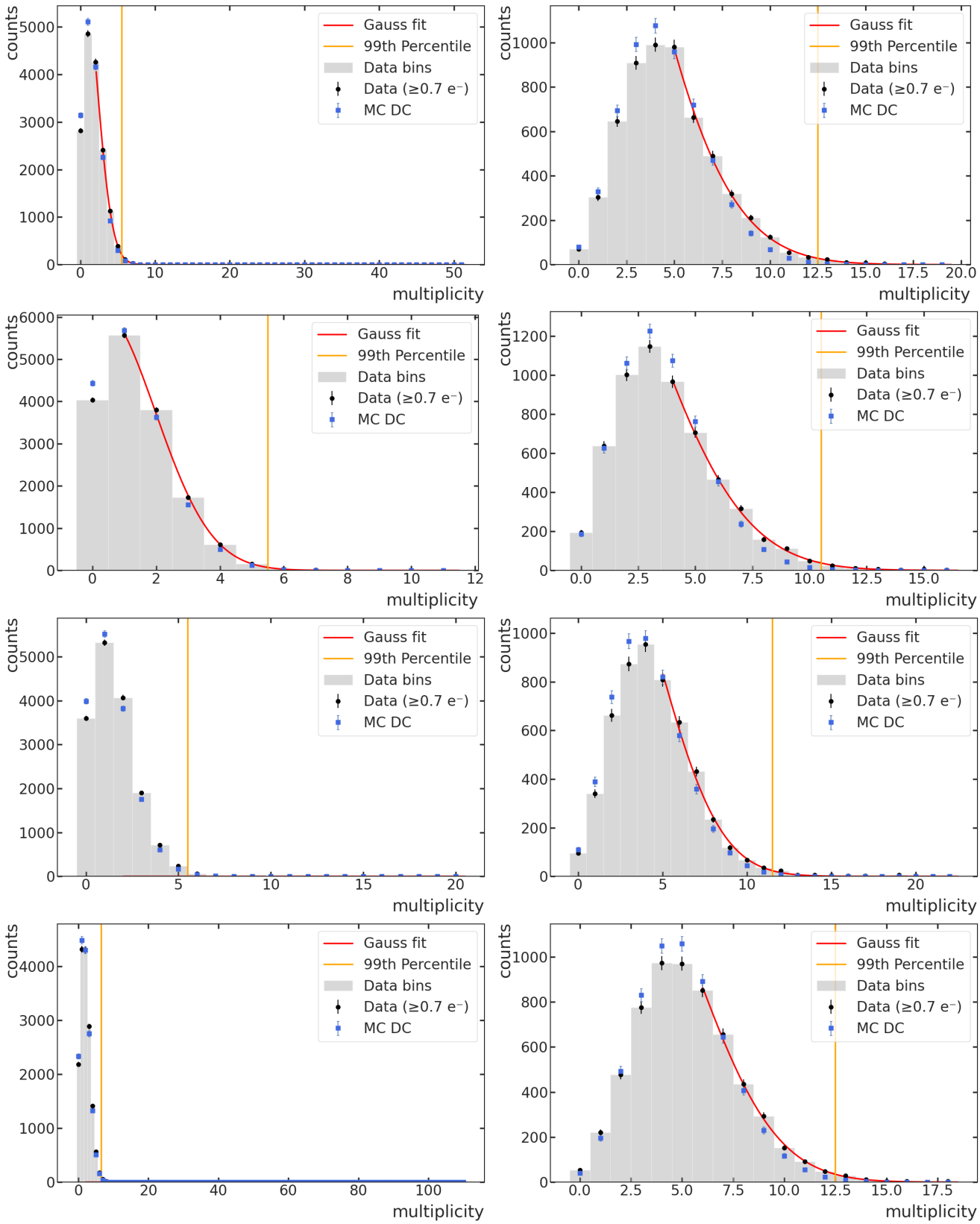


Figure D.23: Charged–pixel multiplicity histograms per row (left) and per column (right) for dataset B4. Gaussian fits (red) are applied to the upper tail. Orange line indicates the 99th percentile threshold. From top to bottom: CCDs A, B, C and D.

After masking, the probability–charge distributions are well described by a Poisson–Gaussian convolution (Fig. D.24). Fit results are summarized in Table D.8: charge resolutions cluster around $\sigma \simeq 0.16\ e^-$ and the mean dark-current rates remain fully consistent with earlier periods, the highest being $4.74 \times 10^2\ e^-\,\mathrm{g}^{-1}\,\mathrm{day}^{-1}$ in CCD D.

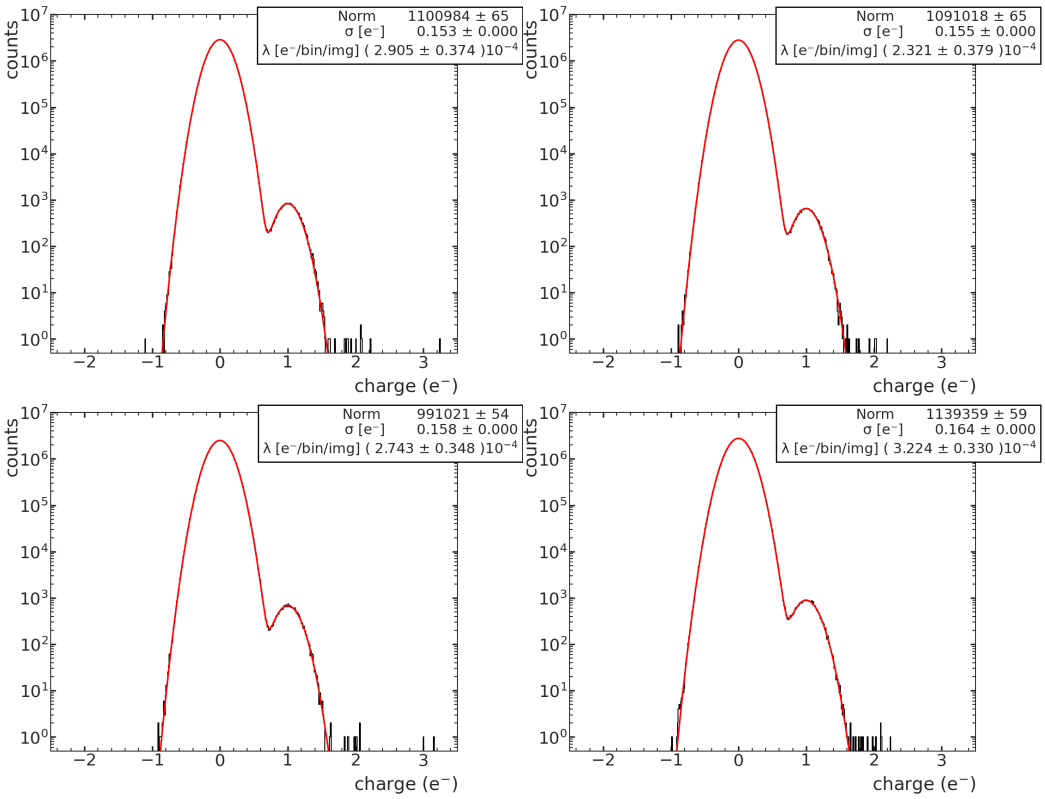


Figure D.24: PCDs for dataset B4 after all masking. The black histograms represent the charge distributions after all masking procedures. The red fit corresponds to the convolution of Poisson and Gaussian distributions (Table D.8). Sorted top to bottom: CCD A, B, C and D, respectively.

CCD	**e⁻/bin/img**	**e⁻/pix/day**	**e⁻/g/day**	σ
CCD A	$(2.9 \pm 0.4) \cdot 10^{-4}$	$(1.50 \pm 0.19) \cdot 10^{-4}$	428 ± 50	0.153 ± 0.000
CCD B	$(2.3 \pm 0.4) \cdot 10^{-4}$	$(1.2 \pm 0.2) \cdot 10^{-4}$	342 ± 60	0.155 ± 0.000
CCD C	$(2.7 \pm 0.4) \cdot 10^{-4}$	$(1.42 \pm 0.18) \cdot 10^{-4}$	404 ± 50	0.158 ± 0.000
CCD D	$(3.2 \pm 0.3) \cdot 10^{-4}$	$(1.67 \pm 0.17) \cdot 10^{-4}$	475 ± 50	0.164 ± 0.000

Table D.8: Poisson–Gaussian fit parameters for dataset B4, including statistical uncertainties. The values correspond to the average charge per pixel per image (e⁻/bin/img), per day (e⁻/pix/day), and normalized per gram per day (e⁻/g/day), along with the charge resolution parameter σ.

D.2 Complete Pattern Statistics

This section presents the full pattern statistics from the four blinded datasets analyzed in Chapter 5.

Each subsection provides two tables per dataset. The first includes the counts of single-pixel patterns with charges of one, two, and three electrons. The second lists all the detected multi-pixel structures, grouped by topology and compared to the expected number of occurrences under the Poisson background model.

Measured values ("Obs") are shown with statistical uncertainties estimated as $\sqrt{N}$, and the "Exp" values correspond to the expected counts based on the fitted dark current and detector noise. These extended results complement the summaries in Chapter 5.2, enabling a more detailed cross-check of any potential deviations or excesses across CCDs and data periods.

D.2.1 Dataset B1 (2024-10-14)

Pattern		**{1}**	**{2}**	**{3}**
CCD A	Exp	$(1.2 \pm 0.2) \cdot 10^4$	(1.7 ± 0.7)	$(1.6 \pm 0.9) \cdot 10^{-4}$
	Obs	$(1.142 \pm 0.011) \cdot 10^4$	8 ± 3	1 ± 1
CCD B	Exp	$(1.1 \pm 0.2) \cdot 10^4$	(1.3 ± 0.5)	$(1.1 \pm 0.6) \cdot 10^{-4}$
	Obs	$(1.003 \pm 0.010) \cdot 10^4$	3 ± 2	0 ± 0
CCD C	Exp	$(1.0 \pm 0.2) \cdot 10^4$	(1.4 ± 0.6)	$(1.2 \pm 0.8) \cdot 10^{-4}$
	Obs	$(0.949 \pm 0.010) \cdot 10^4$	3 ± 2	0
CCD D	Exp	$(1.3 \pm 0.2) \cdot 10^4$	(2.3 ± 0.7)	$(2.5 \pm 1.2) \cdot 10^{-4}$
	Obs	$(1.281 \pm 0.011) \cdot 10^4$	8 ± 3	0
All	Exp	$(4.6 \pm 0.4) \cdot 10^4$	(6.6 ± 1.2)	$(6.4 \pm 1.8) \cdot 10^{-4}$
	Obs	$(4.37 \pm 0.04) \cdot 10^4$	22 ± 5	1 ± 1

Table D.9: Isolated single-pixel counts after masking for B1. "Exp" values come from the Poisson mean fitted to the 1 e^- peak; "Obs" are the data. All CCDs exhibit a significant excess at 2 e^- and 3 e^-.

CCDs \ **p**		**{11}** $[10^0]$	**{12}** $[10^{-3}]$	**{13}** $[10^{-7}]$	**{111}** $[10^{-3}]$	**{112}** $[10^{-7}]$
A	Exp	3.3 ± 1.3	0.5 ± 0.3	0.4 ± 0.3	0.9 ± 0.5	1.3 ± 1.1
	Obs	$1 \pm 1[8]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
B	Exp	2.4 ± 0.9	0.27 ± 0.16	0.3 ± 0.2	0.6 ± 0.3	0.8 ± 0.6
	Obs	$4 \pm 2[15]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
C	Exp	2.5 ± 1.1	0.3 ± 0.2	0.3 ± 0.3	0.7 ± 0.5	1.0 ± 0.9
	Obs	$4 \pm 2[7]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
D	Exp	3.9 ± 1.3	0.7 ± 0.3	0.8 ± 0.5	1.4 ± 0.7	2.8 ± 1.7
	Obs	$5 \pm 2[11]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
All	Exp	12 ± 2	1.8 ± 0.5	1.7 ± 0.7	3.6 ± 1.0	5.9 ± 2.3
	Obs	$14 \pm 4[41]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table D.10: Expected (Exp) and observed (Obs) counts for multi-pixel patterns in each CCD of the B1 dataset. Results are shown in the format $x \pm \delta x\,[N]$, where $x \pm \delta x$ refers to the expected or measured counts in rows, and N is the total number of patterns detected over all directions. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. Each pattern symbol implicitly includes all charge permutations of the same topology.

D.2.2 Dataset B2 (2024-10-28)

Pattern		{1}	{2}	{3}
CCD A	Exp	$(2.8 \pm 0.4) \cdot 10^4$	3.9 ± 1.1	$(3.7 \pm 1.6) \cdot 10^{-4}$
	Obs	$(2.651 \pm 0.016) \cdot 10^4$	8 ± 3	1 ± 1
CCD B	Exp	$(2.2 \pm 0.4) \cdot 10^4$	2.4 ± 0.8	$(1.9 \pm 0.9) \cdot 10^{-4}$
	Obs	$(2.125 \pm 0.015) \cdot 10^4$	9 ± 3	0 ± 0
CCD C	Exp	$(2.5 \pm 0.4) \cdot 10^4$	3.5 ± 1.1	$(3.3 \pm 1.5) \cdot 10^{-4}$
	Obs	$(2.414 \pm 0.016) \cdot 10^4$	17 ± 4	3 ± 2
CCD D	Exp	$(3.3 \pm 0.4) \cdot 10^4$	5.5 ± 1.3	$(5 \pm 2) \cdot 10^{-4}$
	Obs	$(3.210 \pm 0.018) \cdot 10^4$	10 ± 3	1 ± 1
All	Exp	$(1.1 \pm 0.8) \cdot 10^5$	15 ± 2	$(1.4 \pm 0.3) \cdot 10^{-3}$
	Obs	$(1.062 \pm 0.003) \cdot 10^5$	44 ± 7	5 ± 2

Table D.11: Isolated single-pixel counts after masking for B2. "Exp" values come from the Poisson mean fitted to the 1 e^- peak; "Obs" are the data. All CCDs exhibit a significant excess at 2 e^- and 3 e^-.

CCDs \ p		{11} $[10^0]$	{12} $[10^{-3}]$	{13} $[10^{-7}]$	{111} $[10^{-3}]$	{112} $[10^{-7}]$
A	Exp	7 ± 2	1.1 ± 0.5	1.0 ± 0.6	2.1 ± 0.9	3.0 ± 1.7
	Obs	$5 \pm 2[25]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
B	Exp	5 ± 2	0.5 ± 0.3	0.4 ± 0.3	1.0 ± 0.5	1.3 ± 0.8
	Obs	$6 \pm 2[18]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
C	Exp	7 ± 2	0.9 ± 0.4	0.9 ± 0.5	1.9 ± 0.9	2.8 ± 1.7
	Obs	$5 \pm 2[26]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
D	Exp	10 ± 2	1.6 ± 0.6	1.8 ± 0.8	3.3 ± 1.2	6 ± 3
	Obs	$13 \pm 4[36]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
All	Exp	29 ± 4	4.0 ± 0.9	4.0 ± 1.2	8.4 ± 1.8	13 ± 4
	Obs	$29 \pm 6[105]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table D.12: Expected (Exp) and observed (Obs) counts for multi-pixel patterns in each CCD of the B2 dataset. Results are shown in the format $x \pm \delta x\,[N]$, where $x \pm \delta x$ refers to the expected or measured counts in rows, and N is the total number of patterns detected over all directions. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. Each pattern symbol implicitly includes all charge permutations of the same topology.

D.2.3 Dataset B3 (2024-11-19)

Pattern		{1}	{2}	{3}
CCD A	Exp	$(3.5 \pm 0.4) \cdot 10^4$	5.2 ± 1.3	$(5.0 \pm 1.9) \cdot 10^{-4}$
	Obs	$(3.424 \pm 0.019) \cdot 10^4$	16 ± 4	4 ± 2
CCD B	Exp	$(2.8 \pm 0.4) \cdot 10^4$	3.3 ± 1.0	$(2.6 \pm 1.2) \cdot 10^{-4}$
	Obs	$(2.754 \pm 0.017) \cdot 10^4$	8 ± 3	2 ± 1
CCD C	Exp	$(3.3 \pm 0.4) \cdot 10^4$	4.9 ± 1.2	$(4.7 \pm 1.8) \cdot 10^{-4}$
	Obs	$(3.129 \pm 0.018) \cdot 10^4$	23 ± 5	5 ± 2
CCD D	Exp	$(4.3 \pm 0.5) \cdot 10^4$	7.1 ± 1.5	$(8 \pm 3) \cdot 10^{-4}$
	Obs	$(4.09 \pm 0.02) \cdot 10^4$	6 ± 2	0 ± 0
All	Exp	$(1.4 \pm 0.9) \cdot 10^5$	21 ± 3	$(2.0 \pm 0.4) \cdot 10^{-3}$
	Obs	$(1.339 \pm 0.012) \cdot 10^5$	53 ± 7	11 ± 3

Table D.13: Isolated single-pixel counts after masking for B3. "Exp" values come from the Poisson mean fitted to the 1 e^- peak; "Obs" are the data. All CCDs exhibit a significant excess at 2 e^- and 3 e^-.

CCDs \ p		{11} $[10^0]$	{12} $[10^{-3}]$	{13} $[10^{-7}]$	{111} $[10^{-3}]$	{112} $[10^{-7}]$
A	Exp	10 ± 2	1.5 ± 0.5	1.4 ± 0.6	2.9 ± 1.1	4 ± 2
	Obs	9 ± 3[34]	1 ± 1	0 ± 0	0 ± 0	0 ± 0
B	Exp	6.3 ± 1.9	0.7 ± 0.3	0.6 ± 0.4	1.5 ± 0.7	1.8 ± 1.1
	Obs	9 ± 3[29]	0 ± 0	1(col) $\pm$ 1	0 ± 0	0 ± 0
C	Exp	9 ± 2	1.4 ± 0.5	1.4 ± 0.7	2.8 ± 1.0	4 ± 2
	Obs	10 ± 3[33]	0 ± 0	0 ± 0	0 ± 0	0 ± 0
D	Exp	14 ± 3	2.2 ± 0.7	2.5 ± 1.1	4.5 ± 1.5	9 ± 4
	Obs	12 ± 3[47]	0 ± 0	1 ± 1	0 ± 0	0 ± 0
All	Exp	39 ± 5	5.7 ± 1.1	5.7 ± 1.5	12 ± 2	19 ± 5
	Obs	40 ± 6[143]	1 ± 1	1 ± 1	0 ± 0	0 ± 0

Table D.14: Expected (Exp) and observed (Obs) counts for multi-pixel patterns in each CCD of the B3 dataset. Results are shown in the format $x \pm \delta x\,[N]$, where $x \pm \delta x$ refers to the expected or measured counts in rows, and N is the total number of patterns detected over all directions. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. Each pattern symbol implicitly includes all charge permutations of the same topology.

D.2.4 Dataset B4 (2024-12-17)

Pattern		{1}	{2}	{3}
CCD A	Exp	$(2.6 \pm 0.3) \cdot 10^4$	3.8 ± 0.9	$(3.7 \pm 1.4) \cdot 10^{-4}$
	Obs	$(2.575 \pm 0.016) \cdot 10^4$	9 ± 3	1 ± 1
CCD B	Exp	$(2.1 \pm 0.3) \cdot 10^4$	2.4 ± 0.7	$(1.9 \pm 0.9) \cdot 10^{-4}$
	Obs	$(1.985 \pm 0.014) \cdot 10^4$	6 ± 2	0 ± 0
CCD C	Exp	$(2.2 \pm 0.3) \cdot 10^4$	3.0 ± 0.8	$(2.9 \pm 1.1) \cdot 10^{-4}$
	Obs	$(2.132 \pm 0.015) \cdot 10^4$	8 ± 3	2 ± 1
CCD D	Exp	$(3.1 \pm 0.3) \cdot 10^4$	5.4 ± 1.1	$(6.2 \pm 1.8) \cdot 10^{-4}$
	Obs	$(2.912 \pm 0.017) \cdot 10^4$	10 ± 3	0 ± 0
All	Exp	$(1.00 \pm 0.06) \cdot 10^5$	14.6 ± 1.8	$(1.5 \pm 0.3) \cdot 10^{-3}$
	Obs	$(9.500 \pm 0.001) \cdot 10^4$	33 ± 6	3 ± 2

Table D.15: Isolated single-pixel counts after masking for B4. "Exp" values come from the Poisson mean fitted to the 1 e^- peak; "Obs" are the data. All CCDs exhibit a significant excess at 2 e^- and 3 e^-.

CCDs \ p		{11} $[10^0]$	{12} $[10^{-3}]$	{13} $[10^{-7}]$	{111} $[10^{-3}]$	{112} $[10^{-7}]$
A	Exp	7.1 ± 1.8	1.0 ± 0.4	1.0 ± 0.5	2.1 ± 0.8	3.1 ± 1.6
	Obs	$5 \pm 2[28]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
B	Exp	4.6 ± 1.5	0.5 ± 0.3	0.4 ± 0.3	1.1 ± 0.5	1.3 ± 0.9
	Obs	$4 \pm 2[10]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
C	Exp	5.7 ± 1.4	0.8 ± 0.3	0.7 ± 0.4	1.6 ± 0.6	2.5 ± 1.3
	Obs	$7 \pm 3[17]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
D	Exp	9.5 ± 1.9	1.6 ± 0.5	1.9 ± 0.7	3.3 ± 0.98	6 ± 2
	Obs	$13 \pm 4[45]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0
All	Exp	27 ± 3	4.0 ± 0.7	4.0 ± 1.0	8.2 ± 1.5	13 ± 3
	Obs	$29 \pm 5[100]$	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Table D.16: Expected (Exp) and observed (Obs) counts for multi-pixel patterns in each CCD of the B4 dataset. Results are shown in the format $x \pm \delta x\,[N]$, where $x \pm \delta x$ refers to the expected or measured counts in rows, and N is the total number of patterns detected over all directions. Columns whose headings contain a power of ten indicate a scale factor: multiply every entry in that column by the bracketed factor to obtain the absolute count. Each pattern symbol implicitly includes all charge permutations of the same topology.

Appendix E

Data Selection and Pattern Identification Scripts

Tis appendix summarizes the Python scripts developed for the selection of low-energy events, masking, and pattern identification in CCD images from the DAMIC-M experiment. These tools allow the identification of instrumental artifacts, estimation of residual backgrounds, and classification of charge patterns.

All scripts are publicly available at: `https://github.com/elenamruc/dataselectionCCDs.git`
The repository is organized as follows:

- **data_selection_masks/**: Implements pixel-level masks to remove instrumental backgrounds such as hot columns, charge transfer inefficiency (CTI), overdense regions, and high multiplicity areas. A summary script computes the residual dark current and overall masking efficiency.
- **pattern_identification_and_efficiencies/**: Detects isolated single- and multi-pixel patterns in masked images using CDF-based scoring. Also includes scripts to compute recall, misidentification, and pattern expectations from Poisson statistics.
  - **diffusion_probabilities/**: Contains a Monte Carlo simulator to compute the diffusion probabilities $p_{\text{diff}}(q \to p)$, based on a physics-motivated charge diffusion model. Used to correct the expected pattern distributions.

Most scripts can be executed with a command like:

```
python3 pattern_identification.py -f image1.fits image2.fits ...
```

Optional parameters are documented within each script.

Acronyms

ΛCDM Lambda Cold Dark Matter. 96

ADC Analog-to-Digital Converter. 5, 96

ADU Analog-to-Digital Unit. 5, 6, 96, 97

AR active region. 29, 30, 43, 96

CCD Charge-Coupled Device. 1–18, 21–34, 38, 39, 41–48, 51–57, 66, 69–73, 75–79, 81–85, 87–94, 96–104

CDF cumulative distribution function. 34, 36, 96

CDS correlated double sampling. 11, 96

CTI Charge-Transfer Inefficiency. 7, 8, 11, 13–15, 24–27, 30, 32, 47, 96, 98, 102

DAMIC-M Dark Matter In CCDs at Modane. 1–13, 47, 48, 96, 97

DC Dark Current. 7, 11, 13, 15–17, 20, 22–24, 27, 28, 30, 32, 33, 40, 47, 96, 97

DM Dark Matter. 1–3, 6–14, 18, 25, 34, 39, 96, 97

dru differential rate unit. 2, 6, 9–11, 96

HPGe high-purity germanium detector. 10, 96

ICP-MS inductively coupled plasma mass spectrometry. 10, 96

LBC Low-Background Chamber. 6–8, 10, 12, 32, 96, 97

LSM Laboratoire Souterrain de Modane. 2, 6, 7, 9, 96, 97

LZ LUX-ZEPLIN experiment. 1, 96

MC Monte Carlo. 13, 15–17, 20, 22, 23, 27, 28, 38, 40, 96, 97

MisID misidentification rate. 35, 36, 96

MOS Metal–Oxide–Semiconductor. 3, 96

OD overdensity. 13, 22–24, 30, 96

OFHC Oxygen-Free High-Conductivity copper. 6, 10, 96

OG Output Gate. 5, 6, 96, 97

PandaX Particle and Astrophysical Xenon experiment. 96

PCD Pixel Charge Distribution. 30, 31, 52, 53, 56, 66, 67, 72, 73, 78, 79, 84, 85, 90, 96, 98–101

RMS root-mean-square. 4, 6, 96

ROI region of interest. 8, 10, 13, 16, 17, 96, 98

SM Standard Model. 1, 3, 96

SN Sense Node. 5, 6, 96, 97

SNOLAB Sudbury Neutrino Observatory Laboratory. 2, 96

SW Summing Well. 5, 6, 96, 97

TPC Time Projection Chamber. 96

WADERS softWAre for Dark matter ExpeRiments with Skippers. 13, 16, 96

WIMP Weakly Interacting Massive Particle. 1, 2, 96

XENONnT XENON n-Tonne experiment. 96

List of Figures

List of Tables