Correlated radiation erasures limit cross-chip erasure codes in modular superconducting processors: a simulation study

Dylan Henderson[1, *]

[1] *Cogitan, Coppell, Texas, USA*

(Dated: November 21, 2026)

Cross-chip erasure codes [Xu *et al.*, PRL **129**, 240502 (2022)] encode each chip's logical qubit in an inner surface code and combine chips into an outer erasure code. A radiation burst that erases one chip is then a located, correctable error, and the scheme's failure-rate formula assumes that chips fail independently. We test that assumption by simulation. Geant4 transport through a model refrigerator feeds a deposit-to-loss chain calibrated on Google's gap-engineered Willow processor, using a burst-to-error mapping that Google measured with an echo pulse in the stabilizer circuit. The chain does not transfer to a second gap-engineered array we tested, so it applies to Willow-class chips only. In a measured two-chip device, a burst on one chip coincided with one on the other 2.4% of the time (8 events). With one detector threshold fitted, and one Geant4 setting corrected after a first comparison, our transport predicts 1.5%. We consider a $[[4, 1, 2]]$ outer code of four data chips and an ancilla chip, with distance-7 inner codes and a 1 ms outer correction cycle, on chips mounted flat 1 to 3 mm apart in one sample box. Counting line gammas and muons at a measured laboratory gamma flux, these layouts fall 140 to 370 times short of the formula's protection, the upper end from nine-chip arrays. They fall 26 to 60 times short at the higher flux that Willow's burst rate implies. In our model, single events that erase several chips, mostly through material the chips share, cause the gap. Placing each of the five chips in its own sample box at least 5 cm apart cuts the block's correlated failure rate 40 to 60 times at inner-code distance 7 (95% intervals down to 24 times). The comparison is with the same chips on one carrier, 1 mm apart. With lost chips located perfectly, the separated block then recovers 15 to 25% of the formula's protection on our interpolated loss curve. Located from burst detections by the simplest attribution rules, under our nominal loss model, it recovers 0.3 to 0.5%. On chips using the echo, writing off every chip with a burst detected at the qubits' own threshold does worse than no outer code when all five chips share a carrier. These detection results assume that bursts too small to detect do not destroy the inner logical qubit, which our simulation cannot resolve. All rates isolate radiation, the floor that remains once intrinsic logical errors are suppressed.

I. INTRODUCTION

High-energy particles from cosmic rays and natural radioactivity deposit energy in the substrate of a superconducting quantum processor and raise its quasiparticle density [1, 2]. The energy spreads as phonons, breaks Cooper pairs across the die, and raises error rates on many qubits at once for tens of microseconds to tens of milliseconds, depending on the device [3–6]. Such bursts can overwhelm quantum error-correcting codes designed for errors that are rare and local. Channeling the deposited energy away from the qubits has been proposed [7] and demonstrated [8] as a way to suppress the mechanism, and gap engineering at the Josephson junctions suppresses the quasiparticle-tunneling errors [9]. On a gap-engineered array the bursts that remain are rarer, shorter (about 1 ms) and dominated by quasiparticle-induced frequency shifts [10]. They are not removed. On Google's Willow processors, correlated events occur about once an hour, and repetition codes up to distance 29 show an apparent logical error floor of about 10^{-10} per cycle [11]. Later work offers such bursts as a plausible explanation of that floor [10], and another study proposes hadronic cosmic rays as a major contributor to it [12].

Xu *et al.* proposed an architectural answer [13]. If each chip holds one surface-code logical qubit (the inner code), and the logical qubits of several chips are combined into an outer $[[n, k, D]]$ erasure code of distance D, then a burst that erases one chip is a located, correctable error. In their protocol an erased chip is replaced by a reserve chip. Let λ be the rate at which one chip is erased (in Xu *et al.*'s model every burst erases its chip, so λ is their per-chip burst rate), and let τ be the duration of the outer code's correction cycle. Xu *et al.* bound the rate at which D chips are erased within one cycle, the worst case for block failure. For $n\lambda\tau \ll 1$ it is

$$\Gamma_{\mathrm{ind}} = n\lambda\,\frac{[(n+1)\lambda\tau]^{D-1}}{(D-1)!}, \tag{1}$$

a first erasure on any of the n data chips, followed within τ by $D-1$ more on any of the $n+1$ chips, including the ancilla chip that measures the outer code. For $\lambda \sim 1\ \mathrm{h}^{-1}$ and $\tau = 1$ ms this rate is five orders of magnitude below λ at $D = 2$. The scheme has become a reference point for modular architectures. An experiment on a two-module processor, with one chip per module, found that only $\sim$2% of decay events in one module coincide with events in the other, and concluded that modular designs with bespoke codes are promising [14]. A recent concatenated-

* dylan@cogitan.ai

code proposal cites that measurement as evidence that catastrophic errors are predominantly localized to individual chips, and models different code blocks as failing independently [15]. Another proposal expects parts of a large memory that rely on independent refrigerators and power supplies to keep such noise small on the scale of the whole memory [18]. Device failures in distributed codes are modeled as separate events of fixed probability [16], and qubits have been split across substrates to avoid correlated errors carried through a common one [17]. Splitting a chip into two dies 0.2 mm apart did block phonons between them, but 2.8% of bursts were still recorded on both dies [19]. In a three-layer stack of chips, each in its own sample box, about 25% of events reached two detector layers, which the authors attribute to muons crossing both [20].

Equation (1) rests on independence, which Xu *et al.* state explicitly: "there is no correlation between CREs [cosmic-ray events] at different chips, when the interconnects are switched off" [13]. A single event that erases D chips at once defeats the block directly, at a rate $\Gamma_{\rm corr}$ that does not involve τ. For $D = 2$, writing $\Gamma_{\rm corr} = n\lambda p_2$,

$$\frac{\Gamma_{\rm corr}}{\Gamma_{\rm ind}} = \frac{p_2}{(n+1)\lambda\tau} \qquad (D = 2), \tag{2}$$

where p_2 is the probability that an event erasing one chip of the block also erases at least one other. The denominator is about 10^{-6}. A correlation that looks negligible in a laboratory measurement, p_2 of a fraction of a percent, therefore exceeds the independent term by orders of magnitude. What matters is how large the correlation is, what carries it, and how packaging changes it.

This paper studies those questions by simulation, anchored to three measurements. All results are for our model: one model refrigerator, chips calibrated on Willow, and a $[[4, 1, 2]]$ outer code. They are the following.

(i) At $\tau = 1$ ms, a $[[4, 1, 2]]$ block with distance-7 inner codes, its chips mounted flat 1–3 mm apart in one sample box, falls 140–370× short of Eq. (1) at a measured laboratory gamma flux, and 26–60× short at the highest gamma flux consistent with Willow's burst rate. Both ranges cover every erasure reading we consider (Sec. III A).

(ii) Chips are coupled mainly through material they share (the carrier, the sample box and the refrigerator), not by one particle crossing two chips. That is the case in our model of modules on a shared motherboard. For two dies with no shared carrier, a post hoc analysis in a box of assumed dimensions traces 14–24% of coincidences to particles from the neighboring die and most of the others to the box and the refrigerator (Sec. III B). By species, muons are the largest single source of correlated erasures at inner-code distance 7. Gammas and, in multi-chip modules, sea-level neutrons and protons supply the remainder (Sec. III D).

(iii) Chips in separate sample boxes at least 5 cm apart cut the block's correlated failure rate 40–60× against a module with 1 mm gaps, independently of the correction-cycle time. With lost chips located perfectly, they recover 15–25% of the formula's protection at $\tau = 1$ ms. These values are for inner-code distance 7, the recovered share on the analysis curve (Sec. IV), and the lowest bootstrap 95% bound is 24×. Most of the reduction comes from the separation between chips. In our model, separate boxes add a factor of about 2 over one shared box at the same distance (Sec. III C).

(iv) The reduction survives a heavier refrigerator (weakened by about 25%), larger separations, air-shower bundles, a realistic muon spectrum and, on few events, sea-level protons and neutrons. We reran only the shared-carrier array and the multi-chip module with a finer Geant4 production cut, not the separated layouts (Sec. III D).

(v) A larger inner-code distance lowers the absolute stakes steeply. Without echo, where the loss curve is resolved, it also raises the shortfall, from 180× at distance 7 to 404–429× at distance 11, because small single-chip losses vanish before large multi-chip ones (Sec. III E).

(vi) Writing off every chip with a burst detected at the qubits' own threshold is self-defeating in multi-chip modules of echoed, gap-engineered chips. This hard-erasure policy is our reading of the scheme at that threshold. Soft use of burst detections recovers about 100×, under an assumption about bursts too small to detect. A decoder that knew which chip each deposit struck would not need that assumption, but we have not quantified the per-chip hit sensor it would require (Sec. III F).

All rates isolate the radiation term. At the base error rate we simulate, the inner code's own unheralded logical errors are far more frequent (Sec. II C). For most studies we wrote an analysis plan before the runs, kept as a local file and not deposited in a registry. Section IV says which studies had no plan and lists the expectations that failed.

II. MODEL

A. Particle transport

We use Geant4 11.4.1 [21–23] with the FTFP_BERT physics list (QGSP_BIC_HP, with high-precision neutron transport, for the proton and neutron runs). Geant4 creates a secondary particle only if it would travel farther than a set range, the production cut. Below it, the energy is deposited along the parent particle's track [21]. The default 0.7 mm cut therefore suppresses the low-energy knock-on electrons that cross millimeter gaps between chips. Lowering it to 1 µm raises muon coincidences between the neighboring modules of the device in Sec. II D by a relative 17–28%, so the reproduction of that device uses the 1 µm cut. The array, pair and block runs use the default, since the 1 µm cut costs about 50 times more per muon. Of these, only the shared-carrier array A1 and the multi-chip module X1 were rerun at the 1 µm cut. Their central rates changed by up to 18% in either direction, with 95% intervals that allow larger changes (Sec. III D). The separated layouts were not rerun. The

two-die device of Sec. II D uses the 1 µm cut only inside a region around the dies and the default elsewhere. The region holds the dies, the third chip in the same box, the pedestals, the circuit board and the inner 1 mm of the box walls. An electron that the default cut would suppress cannot cross 1 mm of copper, so a 1 mm layer is enough. Electrons above the default cut are produced in every volume.

Sources. Sea-level muons of 4 GeV (the mean energy at the ground) with a $\cos^2\theta$ zenith distribution and an intensity of 1 $\mathrm{cm}^{-2}\,\mathrm{min}^{-1}$ through a horizontal plane [24, Sec. 30.3.1], launched from a disk covering the device. The five-chip block is repeated with energies and angles drawn jointly from a sea-level spectrum [25], and with sea-level protons and neutrons added (Sec. III D). Environmental gammas are emitted isotropically from a sphere around the device. They are the primordial lines of the uranium and thorium chains and ^{40}K at a total flux of 1.7 $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. That is the value measured with a NaI spectrometer in an above-ground qubit laboratory (1.7 ± 0.9 $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [26]). A scattered low-energy continuum serves as a systematic bracket.

Refrigerator. Muon runs include a box-shell model of a dilution refrigerator: nested copper and aluminum box shells and plates with the wall thicknesses and materials (not the radii) of the refrigerator of Loer *et al.* [27] [Fig. 1(c)]. The same refrigerator is used for every layout, so it does not move when the chips do.

Layouts. Chips are $10 \times 10 \times 0.5$ mm^3 silicon, flip-chip mounted 10 µm above 0.5 mm silicon carriers, inside a 5 mm aluminum sample box on an 8 mm copper plate. We study three kinds of layout. The first is a 3×3 array on one shared carrier with 1 mm gaps (A1) or 3 mm gaps (A3), on separate carriers (B3), or mounted vertically (V1), and two chips stacked face to face (S1). The second is a pair of chips on their own carriers, in one shared box at 1–50 mm (P1–P50) or in separate boxes at 20–200 mm (Q20–Q200). The third is a five-chip block in a plus pattern, the size of Xu *et al.*'s $[[4, 1, 2]]$ code with its ancilla chip. It is built either as a multi-chip module on one carrier in one box with 1 mm gaps (X1, the class of multi-die module demonstrated in Ref. [28]) or with each chip on its own carrier in its own box (XQ50, XQ100) [Fig. 1(a,b)]. We call every layout except V1 and S1 flat. The output is, per event, the energy deposited in every chip.

The two-die device of Ref. [19] is modeled from that paper's text and photographs. Two silicon dies, 13.5×10 mm^2 each, lie 0.2 mm apart, each on four copper pedestals at its corners. They sit in a copper box with a 36×36 mm^2 cavity, on the box floor inside a cutout of a circuit board. A third chip occupies the box's other slot. The two scale bars of the photographs disagree on the die size, so 10.5×9.2 mm^2 is run as a bracket. Other dimensions are not published, and we assume them. Die thickness (0.5 or 0.675 mm), pedestal height (0.5 or 1 mm) and cavity height (8 or 12 mm) are bracketed. The pedestal footprint (1×1 mm^2), the walls and floor (5 mm), the lid (3 mm) and the circuit board (0.5 mm) are fixed.

B. From deposited energy to a lost logical qubit

Whether a deposit erases a chip depends on the chip. We model gap-engineered chips of the Willow class, by which we mean chips like those of Ref. [10], on which our calibration was made. Most links of the chain are constrained by measurements on that device. The exceptions, listed in Appendix B, are the gamma channel, which is not used, a transport thickness taken from another device, and a muon and gamma split measured on a third. Other designs are out of scope. (1) A reduced-order ballistic phonon transport model gives each qubit's peak frequency shift δf. It is calibrated to Willow [10]. The muon burst rate sets the effective deposition thickness, and the median peak shift fixes one shift-per-energy constant. The burst-size distribution selects the film absorption within a passing band. The burst duration is then a consistency check rather than an independent prediction, since the recombination constant it uses was fitted to the measured burst recoveries. The 0.5 mm transport thickness follows Google's earlier flip-chip platform [4]. Only the shift-per-energy constant is fitted to a single measured value. The two thicknesses and the film absorption are set within ranges allowed by data or a published device, and the qubit-response constants are measured or bracketed (Appendix A, Table VI). The model, its calibration and its limits are given in Appendices A–B. (2) The shift relaxes by quasiparticle recombination with the measured rate constant $r = 1/(88 \pm 12$ ns) [10]. (3) The shift is converted to excess detection probability per error-correction round using the mapping measured with an echo in the stabilizer circuit: 2% at 1 MHz and 5% at 2 MHz, fitted as $0.02\,(\delta f/\mathrm{MHz})^{1.32}$ [10]. We call chips read out this way echoed. That mapping was measured in a repetition code. We apply it to the surface code, and the authors caution that their decoupling strategy may not carry over directly. The conversion from detection probability to per-qubit Pauli error is bracketed between 0.25 and 1.0. A no-echo variant (17% at 1 MHz, measured in the original circuit) is kept as a bracket. It reuses the echo exponent and so overstates the measured value at 2 MHz (42% against ≈35%), which makes it a pessimistic bound. (4) A distance-7 rotated surface-code memory (distances 5 and 11 in Sec. III E) at base error rate 10^{-3} is simulated in Stim [29] for the duration of the burst, with time-dependent noise and a decoder that is not told about the burst (PyMatching 2 [30]). The simulation gives P_{flip}, the excess probability, over the burst window, that the memory's logical Z value flips [Appendix E, Eq. (E1)]. A Z-basis memory registers only half of the logical errors of a fully scrambled state, so we take the probability that a chip's logical state is lost to be $P_{\mathrm{loss}} = \min(1, 2P_{\mathrm{flip}})$. This factor of 2 was added after the no-echo run, whose flip probability saturated near 0.5. The analysis plan had used the flip probability

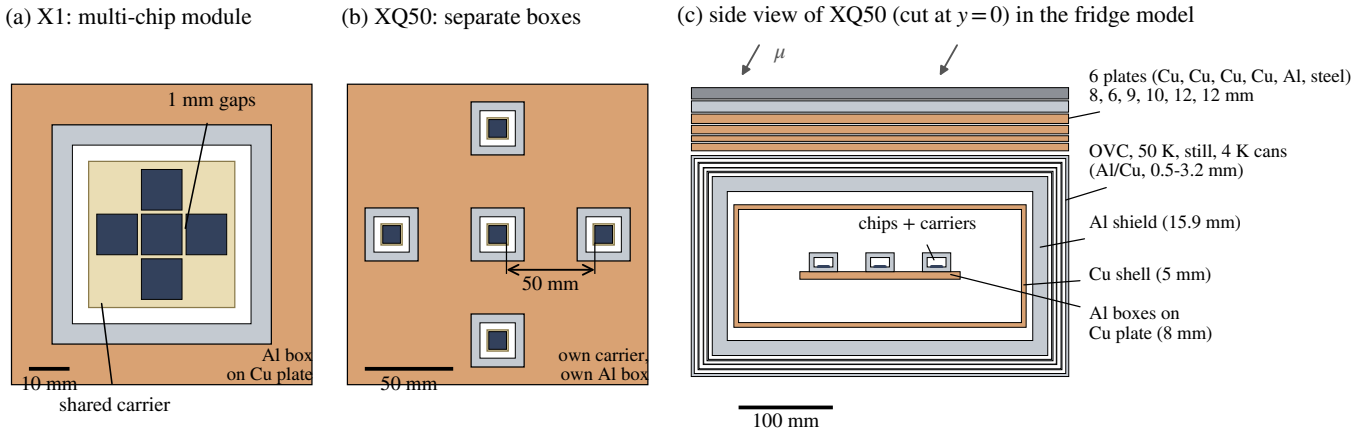


FIG. 1. Geometry. (a) Multi-chip module X1: five chips on one carrier in one sample box, 1 mm gaps. (b) Separate boxes XQ50: each chip on its own carrier in its own box, 50 mm edge gaps. (c) Side view of XQ50 in the refrigerator model, cut at the plane $y = 0$ of the simulation's horizontal coordinates: copper and aluminum shells and plates above the device, with wall thicknesses from Ref. [27]. The radii are our choice, not those of that refrigerator. OVC: outer vacuum can. Still: the still stage.

itself.

Figure 2 shows the result. With echo, a 2 MeV deposit loses the logical qubit only about 17% of the time. At conversions 0.25 and 0.5, deposits below 0.4 MeV almost never do (their bins are not resolved from zero). At conversion 1.0 they are resolved at 2–4% from 0.06 to 0.4 MeV. Without echo, the code is scrambled by ∼0.5 MeV. As a rough consistency check, the predicted rate of lost logical states on our $10 \times 10 \times 0.5$ mm^3 chip is about 1.2 h^{-1} (0.8–1.7 h^{-1} across the gamma-flux bracket), the same order as Willow's rate of about one correlated event per hour [11]. The comparison is not like for like. The chain is calibrated on Willow, and our chips are thicker than the effective deposition thickness of that calibration, so more muons exceed a given deposit.

Erasure readings. We carry five erasure readings through every result. In the flagged reading, a chip counts as erased whenever its deposit exceeds 180 keV. That is the deposit above which Willow's bursts are detected, as inferred from its burst rate [10, 31]. In the four lost readings, a chip counts as erased with probability $P_{\rm loss}$: at each of the three conversions with echo, and without echo at conversion 0.5. A result is reported as robust only if it holds under all five. Some results are also given as counts of events above a fixed deposit threshold (0.3 or 0.5 MeV), which use no loss curve.

C. Code-level rates

For each Geant4 event, with E_i the energy deposited in chip i, the chips of a block are lost independently with probabilities $P_i = P(E_i)$. The block fails if at least D are lost (Poisson-binomial). Weighting events by the source fluxes gives the per-chip erasure rate λ and the correlated failure rate $\Gamma_{\rm corr}$. For the five-chip block $\Gamma_{\rm corr}$ is computed directly, with no pair approximation, and counts any D of the five chips, the ancilla included. For the 3×3 array of Table I we use $\Gamma_{\rm corr} = n\lambda p_2$, with p_2 the probability that an event erasing one chip erases at least one other chip of the array. This is an upper-side estimate for a five-chip block, by an amount we have not quantified. We take $\tau = 1$ ms. Xu *et al.* estimate about 270 µs for the $[[4, 1, 2]]$ code [13], and Sec. III A gives the scaling between the two. We report the protection achieved against one unprotected chip, $\lambda/\Gamma_{\rm fail}$ with block failure rate $\Gamma_{\rm fail} = \Gamma_{\rm ind} + \Gamma_{\rm corr}$, beside the protection claimed by Eq. (1), $\lambda/\Gamma_{\rm ind}$. We call their ratio, claimed over achieved, $1 + \Gamma_{\rm corr}/\Gamma_{\rm ind}$, the shortfall. Counting a failure only when D chips are lost assumes, as Eq. (1) does, that the outer decoder knows which chips were lost. The achieved protection is therefore a packaging-limited ceiling, and Sec. III F asks what burst detections deliver.

What the rates isolate. All rates count radiation only. The inner code's own logical errors are unheralded. At the base error rate we simulate (10^{-3} per operation, 1 µs rounds) they are also far more frequent. The idle distance-7 code fails at 2.9×10^{-6} per round, about 10^4 h^{-1}, against a radiation loss rate of 1.2 h^{-1}, and the ratio is about 8500 at distance 5 and 1800 at distance 11. At distance 11 the radiation rate is itself poorly resolved (Sec. III E). Radiation at distance 7 corresponds to 3×10^{-10} per round, where a round is one error-correction cycle. That is the scale of the apparent floor of about 10^{-10} per cycle seen in repetition codes of distance 15 and above on Willow [11]. Our results therefore describe that floor: the regime, assumed by Xu *et al.* [13], in which intrinsic logical errors have been suppressed below the radiation term by a lower base error rate or a larger distance.

Counts carry exact Poisson 95% intervals. Two-sided intervals on a ratio of two counts k_1 and k_2 use a log-

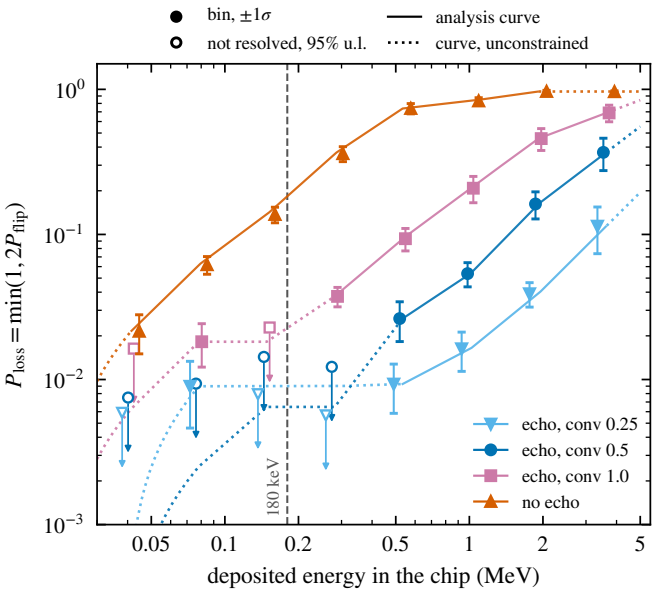


FIG. 2. Probability that a deposit destroys a distance-7 logical qubit on a gap-engineered chip, $P_{\rm loss} = \min(1, 2P_{\rm flip})$, for three detection-to-error conversions with echo and for no echo. Points: Stim results binned in deposit (11–20 simulated strikes per bin, 600 shots each), with 1σ errors (the larger of the per-strike spread and the propagated binomial error). Open symbols with arrows: bins consistent with zero, shown as 95% upper limits. Lines: the interpolated curve used in the analysis, dotted where it rests on unresolved bins or is extrapolated. Dashed vertical line: the 180 keV burst-detection threshold. The curve never decreases with deposit, so one noisy low-energy bin of the conversion-0.25 run (57–108 keV) holds it above the conversion-0.5 curve up to 0.4 MeV. Both runs use the same strikes, and paired strike by strike, the larger conversion gives the larger loss. Capping the conversion-0.25 curve at the conversion-0.5 one lowers its λ by 21%, leaves its shortfall unchanged at the quoted precision and keeps its separation reduction within the quoted range.

normal approximation with $\sigma_{\ln} = (1/k_1 + 1/k_2)^{1/2}$. One-sided lower bounds on reductions ("≥") divide the reference rate by the 95% Poisson upper limit of the separated rate, which ignores the reference's own uncertainty. A one-sided 95% conditional binomial bound, which includes it, is higher (19× rather than 17× for the 50 mm pair of Table II), so the quoted bounds are conservative. Intervals on reductions under the erasure readings are bootstrap 95% intervals over events.

D. Validation

Cross-module coincidences. Wu *et al.* [14] measured cross-module coincidences between two 3×7 mm sapphire modules flip-chip mounted on a shared motherboard, the device of Ref. [32]. Each module is one chip. We measure the module edges to be 2.1 mm apart from the square marks in Fig. 1(c) of that reference, which we take to be the SU-8 spacer studs. We simulate 2.2 mm, inside our ±0.15 mm reading uncertainty. Pooling their two runs (their Fig. 4) gives 8 coincidences among 329 events in one module, 2.4% (exact Poisson 95% interval 1.05–4.8%, our calculation). The pooling assumes that their run-1 coincidences with the two qubits of the other module are the same events. We model their device with the module detection threshold fixed by their single-module rate and nothing else fitted. For gamma fluxes up to our nominal 1.7 $\rm cm^{-2}\,s^{-1}$ we predict 1.1 to 1.65%, centrally 1.5% [Fig. 3(a)], and 0.8–1.65% over all brackets. The central prediction lies about 1.1 standard deviations below the measurement. Across the 1.1–1.65% range the distance runs from about 1.9 to about 0.9 standard deviations. Here we express the one-sided Poisson probability of 8 or more coincidences, given the prediction, as a normal quantile.

This is not the number our analysis plan produced. The plan fixed the method and a pass band of 1.0–4.8%, not a numerical prediction. Applied as planned, the method gave 0.7–1.3% (hatched in the figure), with Geant4's default 0.7 mm production cut, and the planned verdict was a marginal reproduction. That cut suppresses the knock-on electrons that carry energy between closely spaced modules. After comparing that result with the measurement, we reran the device the same day with a 1 µm cut. The finer cut raises the coincidence probability by a factor 1.17–1.28, measured without the refrigerator and assumed to carry over to it. At the 5 cm separations of our rule the cut was not tested (Sec. III D). This measurement is our anchor for the shared-material secondaries on which our results rest. Of the two checks that follow, the two-die one probes them in a second geometry and the muon-coincidence one does not.

Muon coincidences. A stack of three silicon detectors read out by kinetic-inductance detectors measured a muon-induced (background-subtracted) top-bottom coincidence rate of 0.192 ± 0.009 $\rm s^{-1}$ [33]. With the geometry shown in that paper's photograph, we obtain 0.189–0.207 $\rm s^{-1}$ at their muon flux. At our own flux normalization the rate is 16–29% higher. Our singles rate in the top detector is also above their Geant4 value, so our absolute muon rates may be about 20% high [Fig. 3(b)]. A muon-flux error scales the muon terms. It moves the separation reductions only through the share of non-muon events in the multi-chip module's correlated rate (Sec. III D), which we have not varied. The shortfall, which is 1 plus the ratio of Eq. (2), moves by at most 11%, because a muon excess raises $\Gamma_{\rm corr}$ and λ^2 together.

Two dies in one box. Moshel *et al.* [19] read two silicon dies about 0.2 mm apart in one sealed copper box with kinetic-inductance detectors. The dies share no carrier: each rests on its own corner pedestals. Of 352 bursts, 10 were recorded on both dies, 2.8% (exact Poisson 95% interval 1.4–5.2%, our calculation). We model the device as described in Sec. II A, with the sources, refrigerator and flux brackets of the other runs. We did not simu-

late hadrons for it, the muons were mono-energetic, and the site's gamma flux and overburden are unknown. Because the measurement time is not given, the detection threshold cannot be fixed by a rate as it was for the cross-module device. We therefore scan it. Over the die-size, die-thickness, gamma-spectrum and flux brackets we obtain 1.6–2.0% at thresholds of 10–20 keV and 1.1–1.5% at 100 keV. At the central fluxes and nominal geometry, the values at 10–20 keV lie 1.3 to 1.5 standard deviations below the measurement. We have not established whether thresholds this low are plausible for their 11 nm aluminum detectors. With the threshold free, this check is weaker than the cross-module one. What it adds is a geometry with no shared carrier (Sec. III B).

Two expectations in our plan for this study failed. We expected the coincidence probability either to reach the measured 2.8%, with secondaries supplying most coincidences, or to fall below the measured interval. Instead it lies inside the interval but below 2.8%. Moshel *et al.* report the "typically very high amplitude" of their coincident events. We therefore expected the median of the larger deposit in a coincident event to exceed, more than threefold, the median deposit of events recorded on one die only. In our model the ratio is 1.3–1.8. Their observation concerns the nonlinear phase response of their detectors, so the comparison is only indicative. It agrees with a stacked-detector measurement in which events seen by two detectors appear to deposit less energy per detector, on average, than events seen by one [20].

Environment. Our transport reproduces, chain by chain, the gamma emission of Fowler *et al.* [31] (to 1% for ^{40}K, U and one Th subchain, 11% for the other) and the fraction of substrate events above 1 MeV in their computed backgrounds [34]. Our absolute substrate gamma rate is about twice theirs, uniformly across decay chains, a difference we have not resolved. The gamma flux does not cancel from the separation reductions. Gammas carry 11–24% of the multi-chip module's correlated rate across the distance-7 readings (Sec. III D). No simulated gamma event contributes to the separate-box correlated rate, where muons carry 86–95% and hadrons the rest (at a proton-to-neutron ratio of 1). A higher gamma flux would therefore raise the module's correlated rate and leave the separated one unchanged, which would increase the reductions. We have not computed by how much.

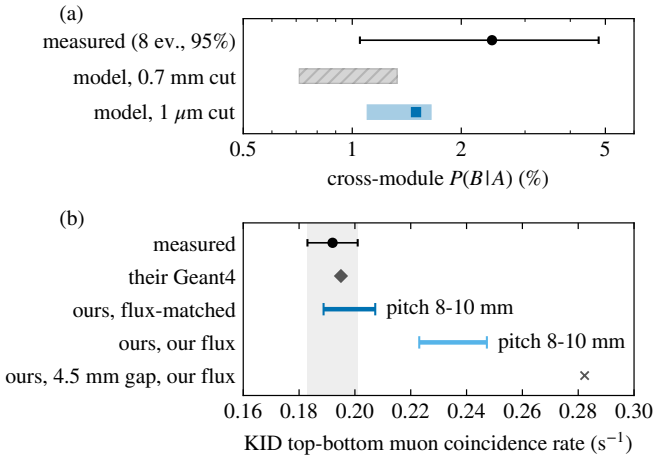


FIG. 3. Validation. (a) Cross-module coincidence probability $P(B \mid A)$, the probability that an event in module A coincides with one in module B, for the device of Ref. [14]. Shown are the measured value (8 events, 95% interval) and our prediction band with a 1 µm production cut. Hatched: the result of the planned method, with the default 0.7 mm cut. (b) Top-bottom muon coincidence rate for the stack of kinetic-inductance detectors (KIDs) of Ref. [33]. Shown are the measured value (shaded: its quoted uncertainty), their Geant4 value, and ours at their flux and at our flux, for the wafer pitch of 8–10 mm read from that paper's photograph. Cross: our first, planned geometry, which took the 4.5 mm interlayer spacing stated in the text as the gap between wafers. We read the photograph as showing that spacing between their copper holders (Sec. IV).

III. RESULTS

A. The block falls far short of its formula

Table I and Fig. 4(b) give the protection of a $[[4, 1, 2]]$ block in the 3×3 array layouts. On a shared carrier, the achieved protection is 370× below the claim when only chips whose logical qubit is actually lost are counted, and 250–340× below it under the other readings. Separate carriers (B3) help by a factor of about 2. Mounting the package vertically makes it more than 10 times worse. Across readings, brackets and the presence or absence of echo, the shortfall in the flat 3×3 arrays stays within 140–370× for the measured gamma-line flux. With the scattered-continuum bracket the shared-carrier value rises to about 580×. The absolute protection varies far more. It runs from about 9× to about 1000×, set mainly by the erasure reading, and how burst detections are used changes it further (Sec. III F).

The gamma flux is the least certain input (Sec. IV). At the higher flux we infer from the burst rates of a second laboratory's device (3.2–4.5 cm^{-2} s^{-1} by spectrum, Appendix B), the shortfall of the flat 3×3 arrays is 88–290×. At a flux that reproduces Willow's gamma burst rate at the calibrated thickness (11–19 cm^{-2} s^{-1}), λ rises 2.9- to 3.6-fold and the shortfall falls to 26–60×. Gammas raise λ much more than they raise Γ_{corr}, so more of them shrink the shortfall. They do not cancel from the separation reductions, since they carry 11–24% of the multi-chip module's correlated rate, and no simulated gamma event enters the separate-box rate (Sec. II D). The effect of a higher gamma flux on the reductions is to increase them, by an amount we have not computed.

The shortfall also depends on the correction-cycle time. Since $\Gamma_{\mathrm{ind}} \propto \tau$ and Γ_{corr} does not depend on τ, the claimed protection, the shortfall and the share of the claim recovered all scale with τ, while reductions in Γ_{corr}

TABLE I. Protection of a $[[4,1,2]]$ block of the 3×3 array against one chip ($\tau = 1$ ms). Lost: a chip counts as erased with probability P_{loss} (echo, conversion 0.5 unless stated). Flagged: every chip above 180 keV is written off. $\Gamma_{\text{corr}} = n\lambda p_2$ with the array's p_2 (Sec. II C). Claimed: $\lambda/\Gamma_{\text{ind}}$. Achieved: $\lambda/\Gamma_{\text{fail}}$. Shortfall: claimed over achieved.

Reading	Layout	λ (h^{-1})	Claimed	Achieved	Shortfall
Lost	A1 shared	1.24	1.45×10^5	390	370×
Lost	B3 separate	1.24	1.45×10^5	738	200×
Lost	V1 vertical	1.47	1.23×10^5	30	4100×
Lost, no echo	A1	38	4.8×10^3	17	280×
Flagged*	A1	79	2.3×10^3	9	250×

*Protection against the flag rate. Measured against the true loss rate of a chip, writing off flagged chips is worse than no outer code in a module (Sec. III F).

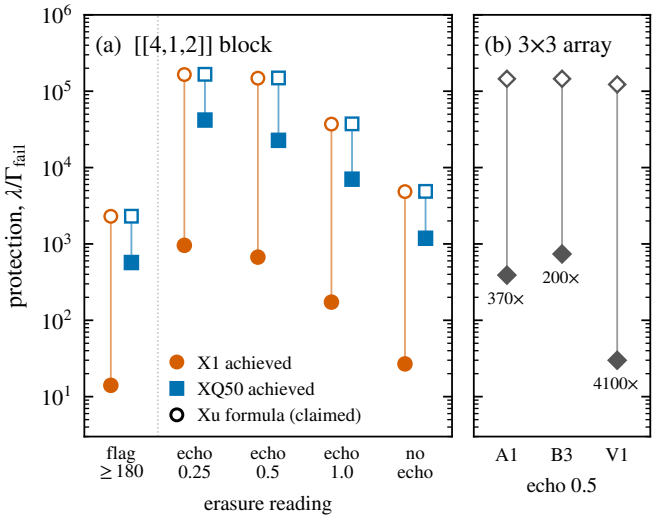


FIG. 4. Protection claimed by Eq. (1) (open symbols) and achieved with lost chips located perfectly (filled), $\tau = 1$ ms, with $\Gamma_{\text{fail}} = \Gamma_{\text{ind}} + \Gamma_{\text{corr}}$. (a) The five-chip $[[4,1,2]]$ block as a multi-chip module (X1) and in separate boxes (XQ50), under each erasure reading. Under the flagged reading ("flag ≥180": deposits of at least 180 keV) both are normalized to the flag rate, as in Table I. Measured against the loss rate, writing off flagged chips is worse than no outer code in the module (Sec. III F). (b) The 3×3 array layouts of Table I (echo, conversion 0.5), labeled with the shortfall.

do not. At Xu *et al.*'s estimate of 270 µs every shortfall is about 3.7 times larger and every recovered share about 3.7 times smaller than the values we quote at 1 ms.

B. What couples two chips

In the device of Wu *et al.*, the usual way the two modules are hit together is that a muon passes through the shared motherboard and knocks out an electron energetic enough to cross the gap into the other module. Tagging primary tracks in that geometry (muons only, default production cut) shows that one particle crossing both modules is rare. In 301 muon coincidences with the enclosure alone, the muon crossed both modules in 5. In 263 it crossed one module and the shared motherboard, and a knock-on electron carried a median 355 keV into the other module. Adding the refrigerator raised the muon coincidence probability from 1.13% to 1.75%. These runs are muons only at the default cut, so they are not comparable with the all-species 1.5% at the 1 µm cut of Sec. II D. The refrigerator also raised the share of coincidences in which the muon crossed neither module from 11% to 25%: secondaries showering down from the copper of the refrigerator above [Fig. 5(a)].

The two-die device of Sec. II D has no shared carrier. Moshel *et al.* consider it likely that its coincidences come from one particle crossing both dies at a shallow angle [19]. In our model, most do not. The criterion in our plan was the share of coincidences in which the primary particle deposited energy in both dies, with crossing taken as the main route if that share reached one half. Between 20 and 300 keV it is 2–8% at the central fluxes in the nominal geometry and 2–16% over all brackets. A photon deposits energy only through the electrons it sets in motion, so this measure undercounts gamma crossings. After the main runs we therefore added two analyses that the plan did not include. The first counts coincidences whose primary entered both dies. For no particle species, bracket or threshold between 20 and 300 keV does that share exceed 22%. The second is a separate run of the nominal geometry. For each coincidence it records the birth volume that supplied most of the energy entering the die the primary missed. At the central fluxes and the same thresholds, the primary entered both dies in 2–12% of coincidences. The dominant birth volume was the copper box in 52–66%, the refrigerator in 12–21% and the neighboring die, across the 0.2 mm gap, in 14–24%.

A count of straight lines through the photographed geometry, for die thicknesses of 0.5–0.675 mm, agrees. Of isotropic lines that meet one die, 1.2–2.2% also meet the other. For the $\cos^2\theta$ zenith intensity of sea-level muons the share is 0.4–0.9%. The ray model of Moshel *et al.*, also isotropic, gives 4.0%, and we have not identified the source of the difference. They name a $\cos\theta$ zenith distribution as a refinement of that model. Our analysis and theirs both fit the measured rate and differ on its cause. They rule out pair-breaking radiation reflected inside the box, and phonons through the box and the circuit board, as significant paths. Our explanation does not rely on either.

Most of the coupling in both devices therefore runs through *shared material*. In the cross-module device that material is mainly the motherboard under the modules, and in the two-die device the box around the dies. In both, the refrigerator above adds to it. In the two-die device, particles from the neighboring die and direct crossings account for the remaining coincidences, an attribution that rests on a post hoc run in a box of assumed dimensions. In separate boxes there is no shared carrier

or box, and the refrigerator carries the coupling.

This has two consequences. First, among muons and gammas, muons carry 88–100% of multi-chip events in the flat 3×3 arrays, although they carry only 40–53% of single-chip erasures [Fig. 5(b)]. Sea-level neutrons and protons add to the correlated rate of the multi-chip module X1 (Sec. III D), but at inner-code distance 7 muons remain its largest single source. We did not simulate shielding or overburden. From these shares we infer that lead shielding would lower λ much more than Γ_{corr}, since gammas carry 11–24% of the module's correlated rate (Sec. III D). Operating underground would lower the correlated term directly. Muons cannot be screened by local shielding, and shallow underground sites attenuate the muon flux by up to a factor of 35 at 100 m [35]. In line with this, coincidences between two detectors are used to tag muons [5]. Second, Γ_{corr} does not depend on the outer code's correction-cycle time τ, so a faster outer-code cycle shrinks only Γ_{ind}, which is already the smaller term. Shorter quasiparticle bursts (faster trapping, stronger gap engineering) do help, by lowering each chip's loss probability. But they lower correlated and independent losses together. If every (small) P_{loss} is scaled by a factor s, both Γ_{corr} and Γ_{ind} scale as s^2 and the shortfall is unchanged. Section III E tests this by changing the inner code's distance.

A shared bump-bonded carrier might also couple chips through phonons. A ballistic transport model with bump transmission bracketed from 0.003 to 0.3 shows that a carrier larger than the die acts as a drain rather than a bridge. For realistic indium bumps a neighbor receives at most 0.16% of a struck chip's film energy, which changes p_2 by a relative 1–11%. Only full-area bonding reaches a factor of about 2. This agrees with Liu *et al.*, who expect phonons reaching the indium bumps of a bump-bonded module to relax to energies too low to enter the qubit chip's niobium ground plane, and acoustic mismatch to inhibit transfer from chip to chip [36].

C. Separate boxes

If the coupling is shared material, the remedy is to share less of it. Table II and Fig. 6 give the two-chip erasure probability for chip pairs, at a 0.5 MeV threshold in the table. Distance inside one shared box helps slowly at first, 3.5× from 1 to 20 mm. It then helps strongly: two chips 50 mm apart in one shared box (P50) are 30× below the 1 mm value at this threshold, and 25× at 0.3 MeV, where the counts are larger. A separate box at equal distance helps a further factor of about 2: 2.6× at 20 mm, and at 50 mm 1.8× (95% interval 1.1–3.0 at 0.3 MeV). The 50 mm comparison followed a written plan, whose verdict was that the box walls matter. Distance therefore does most of the work, and separate boxes add a smaller factor whose 95% interval at 0.3 MeV excludes 1. In separate boxes at 50 mm the two-chip probability is about 50× below the 1 mm shared-box value (95% bound

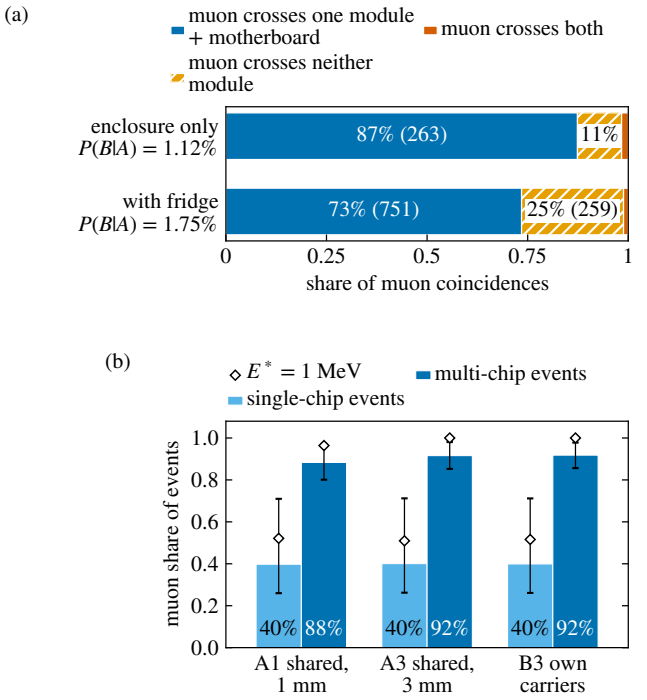


FIG. 5. What couples two chips. (a) Muon coincidences between the modules of Ref. [14] by primary-track topology, with and without the refrigerator (diagnostic runs at the default production cut, 15 keV module threshold). $P(B \mid A)$ is the cross-module coincidence probability of Fig. 3(a). (b) Rate-weighted muon share, among the simulated species (muons and gammas), of events erasing exactly one chip and of events erasing two or more in the 3×3 array. Bars: a chip counts as erased above a deposit threshold $E^* = 0.5$ MeV, at the nominal gamma flux. Error bars: the range over the gamma-flux bracket (two spectra, three fluxes). Diamonds: the same at $E^* = 1$ MeV.

$\geq 17\times$). Replacing the threshold with the physics erasure curves gives 44–51× at 50 mm under every reading. At 100 mm the reduction is larger still, but it rests on one event at this threshold. Under the lost readings, which weight every event by its loss probability, the equivalent number of equally weighted events is 9–18. We therefore report it only as continued improvement and do not use it as a headline.

The rule carries over to a full code block. For the five-chip block, separate boxes at 50 mm cut the block's correlated failure rate by 40–59× against the multi-chip module under every reading. Bootstrap 95% intervals over events run from 50–58× for the flagged reading to 24–60× for echo at conversion 0.5, the reading most dominated by a few large-deposit events. By counts at the 0.5 MeV threshold the reduction is 38× (36 events in 2.4×10^8 muons, 95% interval 26–54×). At 100 mm it is larger again (2 events, not quoted). In absolute terms (echo, conversion 0.5), the multi-chip module achieves 672× against a claimed 1.48×10^5, 0.45% of the claim. Separate boxes achieve 22 700×, about 15% of it (Fig. 4).

TABLE II. Two-chip erasure probability p_2 of Eq. (2) (muon events erasing both chips over those erasing at least one) at a 0.5 MeV threshold, and reduction against P1. P: one shared box. Q: separate boxes. Events: the number of muon events erasing both chips. Last column: reduction under the five erasure readings (blank: not computed).

Layout	p_2	Events	Reduction	Five readings
P1 (1 mm)	0.265%	108	1	1
P5	0.126%	26	2.1	
P10	0.102%	20	2.6	
P20	0.076%	13	3.5	6
P50	0.009%	5	30	
Q20	0.029%	23	9	10–11
Q50	0.005%	3	∼50 (≥17)	44–51

≥: one-sided 95% lower bound (Sec. II C). Not used, on one event: Q100 (100 mm), p_2 of 0.003% (1 event), reduction ∼100 (≥18), 143–175 under the five readings.

Across readings, with lost chips located perfectly, separate boxes recover 15–25% of the claim, against 0.45–0.61% for the module. At 270 µs the values are 5–8% against 0.12–0.17%. Section III F gives what burst detections deliver.

D. Robustness

Heavier refrigerator. Doubling every copper wall and plate raises the shower contribution between separated boxes. The 50 mm reduction falls from about 45–50× to 35–40× under the five readings (29× by counts at 0.3 MeV, ≥ 18.6× at 95%). The plan's floor test required the separated pair to stay at least 10× below P1 at 95% confidence, without naming a threshold. It passes at 0.3 MeV. At 0.5 MeV, our headline threshold, the counts give 22× with a 95% bound of ≥ 8.7×, below the planned 10×, on 4 separated events. The rule therefore depends on the refrigerator: doubling its copper once weakened it by about 25%.

Larger separation. At 200 mm, rerun with corrected muon coverage, the two-chip probability keeps falling (1 event in 6.6×10^4 at 0.3 MeV), consistent with continued decline. On one event, a floor beyond 10 cm is not constrained.

Air-shower bundles. One cosmic-ray air shower can send separate particles into two chips however far apart they are. On our estimate this adds a few percent to the simulated two-chip rate at 5 cm. Single-primary transport cannot produce two particles from one air shower hitting two chips, a term that does not fall with separation. We bound it with the measured sea-level density spectrum of extensive air showers. That spectrum gives the hourly rate of showers exceeding a local density ρ of charged particles, $N(>\rho) = B\rho^{-\gamma}$. We weight it by the probability $(1 - e^{-\rho A_c})^2$ that a shower of local density ρ puts at least one particle in each of two

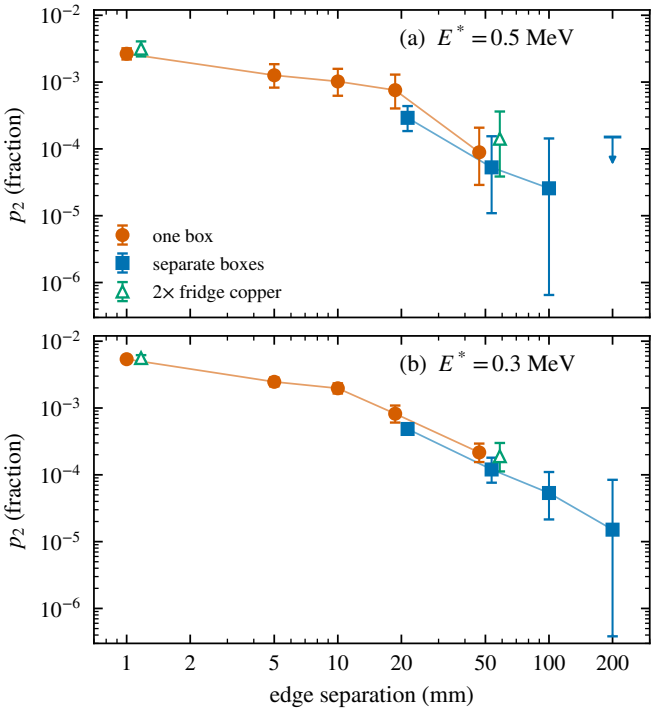


FIG. 6. Two-chip erasure probability p_2 of muon events against edge separation, for chips in one shared box (1–50 mm) and in separate boxes (20–200 mm), with the heavier-refrigerator points. The heavier-refrigerator points are shifted slightly right for visibility, and the one-box and separate-box points at 20 and 50 mm slightly apart. A chip counts as erased above a deposit threshold E^* of (a) 0.5 MeV and (b) 0.3 MeV. As in Table II, p_2 is conditional on at least one chip being erased, and here it is plotted as a fraction (the table gives percent). Error bars are exact Poisson 95% intervals. Arrow: 95% upper limit at 200 mm, where no muon event exceeds 0.5 MeV in both chips.

chips of area A_c. For 1 cm^2 chips the bound is set by densities near 10^4 m^{-2}, above every measured range, so the high-density behavior matters. Single power laws (B = 620 h^{-1}, γ = 1.425 over 5–500 m^{-2} [37], and B = 540 h^{-1}, γ = 1.39 below 500 m^{-2} [38]) give a loose envelope. The measured spectrum steepens above 500 m^{-2}, its exponent reaching 2.2 at 1000 m^{-2} [38]. Holding it at 2.2 above that, an extrapolation, gives our estimate. We count every shower particle as reaching the chip with no absorption, and use our muon data for the probability that a hit erases a chip. Bundles then add 2–7% of the simulated 50 mm two-chip rate across the erasure readings (4–13% for the single power laws). Since the correlated rate keeps falling with separation while the bundle term does not, bundles may matter beyond ∼10 cm, but they do not limit the 5 cm result.

Muon spectrum. We repeated the five-chip block with muon energies and angles drawn from a sea-level muon spectrum parametrized from measurements [25], in which 23% of muons are below 1 GeV. The counts at 0.5 MeV are unchanged. Of the muon events that erase

at least one chip, 0.850% erase two or more in the module (0.816% at 4 GeV) and 0.0216% in separate boxes (0.0217% at 4 GeV), a reduction of 39× (95% interval 27–58×). Across the five readings it is 54–66×. The one reading that moved most (echo, conversion 0.5, from 40 to 66×) has overlapping bootstrap intervals, 24–60× with 4 GeV muons and 50–85× with the spectrum, so the change is consistent with statistical fluctuation (Fig. 7).

Protons and neutrons. Neutrons hit the module harder than the separated boxes, so separation helps against them too, although the separated-box count is small. A neutron makes large multi-chip deposits through nuclear recoils and fragments from interactions in the shared carrier and copper. We added sea-level neutrons, using the JESD89A reference spectrum [39], which integrates to 19.6 $\mathrm{cm^{-2}\,h^{-1}}$ above 1 MeV, and high-precision neutron physics. We ran 4×10^7 neutrons on the separated boxes XQ50 and 6×10^6 on the module X1. We also added protons, 3×10^7 on XQ50 and 6×10^6 on X1, normalized to the neutrons with a proton-to-neutron ratio p/n bracketed at 0.5–2. That ratio is about 2 for nucleons above 1 GeV/c [24, Sec. 30.3.3]. For the protons we reuse the neutron spectral shape, an untested approximation, and the ratio bracket covers only the normalization. The block reduction stays at 49–70× across the readings (Fig. 7). Hadrons add 4–28% to the separated-box correlated rate, relative to its muon and gamma terms, and more to the module's. All three two-chip events above 2 MeV in the module are neutrons, against none from 1.6×10^7 muons. The module's neutron sample corresponds to about 1450 h of sea-level exposure and its muon sample to about 1260 h. In the module, hadrons carry 2–42% of the correlated rate across the distance-7 readings and gammas 11–24%, so the module's correlated term is cosmic-ray driven rather than muon driven. At the nominal reading (echo, conversion 0.5) muons still carry the largest share. In separate boxes muons carry 86–95% at $p/n = 1$. Because hadrons interact strongly, overburden should remove them faster than muons, which would strengthen the case for depth. We did not simulate overburden. The hadronic counts are small (three two-chip events in separate boxes at 0.5 MeV).

Production cut. Two layouts were rerun at the 1 µm cut. On the shared-carrier array A1 (muons, enclosure only) the finer cut raises p_2 by a relative 18% at 0.3 MeV and by 11% at 0.5 MeV, where the change is consistent with zero (43 against 40 events). Repeating the multi-chip module with 4×10^6 muons at the 1 µm cut changes its two-chip event rate by +8% at 180 keV (95% interval 1.7–14%), −3% at 300 keV and −16% at 0.5 MeV (both consistent with zero). Its correlated rate under the five readings changes by −10% to +12%. These two layouts therefore change by at most 18% in either direction. The separated layouts (the Q pairs and XQ50) were not rerun. Their coupling runs through box walls and refrigerator showers rather than millimeter gaps, so we expect the cut to matter less there, but that is an argument, not a result.

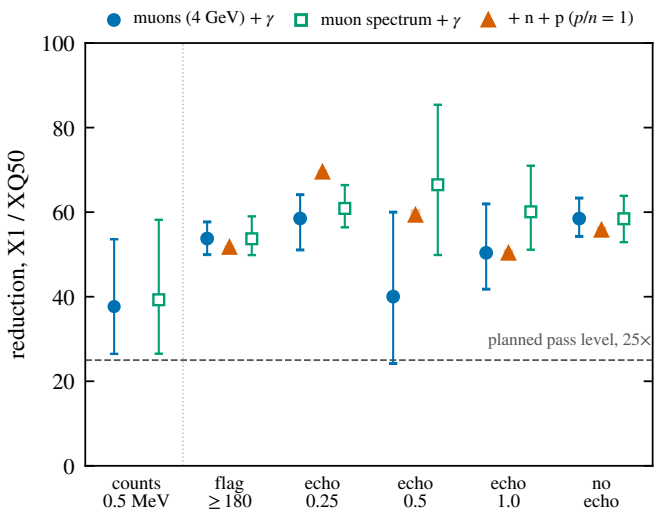


FIG. 7. Reduction of the five-chip block's correlated failure rate by separate boxes at 50 mm (XQ50) against the module (X1). Category "counts 0.5 MeV": ratio of the fraction of muon events depositing ≥0.5 MeV in two or more chips, among those depositing it in at least one. It is 37.7× (26–54) for 4 GeV muons and 39.3× (27–58) for a sea-level spectrum [25] (log-normal 95% intervals). Other categories: ratio of Γ_{corr} under each erasure reading (inner-code distance $d = 7$), with "flag ≥180" the flagged reading at 180 keV. Circles: muons and gammas. Squares: with the muon spectrum. Both carry bootstrap 95% intervals over events. Triangles: with protons and neutrons added at $p/n = 1$. The faint bar behind each triangle spans p/n from 0.5 to 2, at most 2.5 in the reduction factor. The triangles carry no interval, since the separated-box hadronic term rests on three events. Dashed line: the 25× pass level fixed in the analysis plan.

Layouts to avoid. Vertical mounting turns sea-level muons into grazing tracks along the array and raises p_2 3–9×. This is consistent with the larger above-MeV deposits measured in vertically oriented wafers [35]. Stacking chips face to face raises it about 8×. For the two-module device of Ref. [14], our model predicts that rotating the package so that one module sits above the other raises its cross-module coincidence rate (events per hour) 1.7–2×. Standing it on the other edge leaves it unchanged in our model. Distinguishing the two at 2 standard deviations needs 43–63 h of data per orientation.

E. Inner-code distance

The loss curve of Fig. 2 is for a distance-7 inner code. We repeated the chain at distances 5 and 11 (echo, conversion 0.5, 120 strikes each), and at distance 11 also at conversion 1.0 and without echo. The loss probability falls steeply with distance wherever it is resolved (Fig. 8, Table III). Echoed at conversion 0.5, distance 11 is resolved only above about 1.4 MeV, whereas at conversion 1.0 resolved bins carry 72% of λ, and without echo 95%.

Both expectations that the plan's addendum wrote for these two added runs held. Between 2.6 and 5 MeV the distance-11 loss probability is 0.073 at conversion 1.0 (expected below 0.1) and 0.76 without echo (expected above 0.3).

The distance has two effects. First, the shortfall grows with distance where the curves are resolved. Lowering every loss probability by one factor would lower correlated and independent failures alike (Sec. III B). Instead the small, single-chip losses vanish first and the large, multi-chip deposits last. Without echo the shortfall rises from 180× at distance 7 to 404–429× at distance 11, and at conversion 1.0 from 214× to about 2500× on the analysis curve (234–6100×). These ranges are set by treating the unresolved bins as zero or at their 95% upper limits. At conversion 1.0 the distance-11 range lies above the distance-7 value and overlaps the distance-7 range. At conversion 0.5 the shortfall goes from 184× at distance 5 to 200–380× at 7. At 11 that bracket is not resolved: only 1% of λ and 6% of the module's correlated rate come from resolved bins, and the data bound the shortfall only from below, at about 230×. Second, the absolute stakes fall steeply. From distance 7 to 11 the per-chip loss rate falls about 300-fold at conversion 1.0 and 21-fold without echo, and the losses that remain come from deposits above ∼1 MeV.

Those deposits are rare in our samples. Above 1 MeV the separate-box sample holds one two-chip muon event, against 13 in the module. Without echo, where deposits from 0.3 MeV up still matter at distance 11, separate boxes cut the correlated rate 32× (bootstrap 95% interval 19–65×, against 58× at distance 7). With echo the separated-box term is not resolved (7.5×, interval 2–134×, at conversion 1.0). The counts and the flagged reading, which use no loss curve, remain the evidence for the rule. In our model, cross-chip codes matter most where bursts still destroy inner logical qubits: without echo, with smaller inner codes, and for the largest deposits. Chips without gap engineering, where this may also hold, are outside our calibration. Where the loss curves are resolved, the larger the inner code, the further the outer code falls short of its formula.

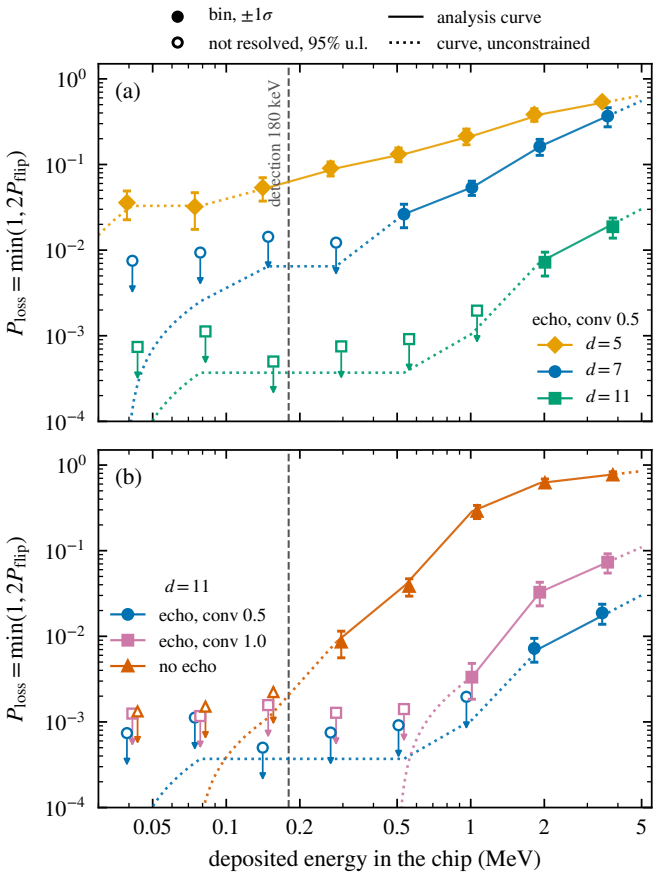


FIG. 8. Loss probability $P_{\rm loss}$ against deposited energy, with symbols as in Fig. 2. (a) Inner surface codes of distance $d = 5$, 7 and 11 (echo, conversion 0.5). At distance 11 the loss is resolved only above about 1.4 MeV (about 7×10^{-3} at 2 MeV and 2×10^{-2} at 3.6 MeV). Lower bins are upper limits of about 5×10^{-4} to 2×10^{-3}. (b) Distance 11 at conversion 0.5 and 1.0 with echo, and without echo. Below 0.7 MeV the conversion-0.5 curve lies above the conversion-1.0 one only because it is drawn through unresolved bins. Both are consistent with zero there.

F. Erasure policy

Xu *et al.*'s protocol replaces a chip "upon detecting erasure errors" on it [13]. Later work describes the scheme as treating a ray detection as a heralded erasure [40] and expects bursts to reach an outer code as erasures of the inner-code tiles they hit [18]. In Xu *et al.*'s model a burst always destroys the chip, so detection and loss coincide. On echoed, gap-engineered chips they do not. Per chip, bursts above the detection threshold occur at about 78 h^{-1}, while the inner logical qubit is lost at about 1 h^{-1}. Writing off every flagged chip, which we call hard erasure on detection or the FLAG-ALL policy, therefore discards mostly healthy data. It is the rule of the flagged reading of Sec. II B, but measured here against the true loss rate of a chip rather than the flag rate. Two simultaneous detections lose the block although nothing was damaged. This is the chip-level analogue of false-positive erasure flags, which add to the effective erasure rate of erasure qubits [41]. For a burst-estimating decoder within one code, point estimates suggest that false alarms cost more than the decoder saves at measured impact rates, though the net benefit is uncertain [42]. A loss-detection scheme proposed for trapped-ion chains lists a working chain flagged as lost among its failure modes [43]. Those are detections of events that did not happen. Here the bursts are real: the detection is correct, and the inner code survives the burst. In a different chip model too, most strikes leave the code intact. Over 64 simulated muon strikes on a distance-7 surface code, a logical failure followed in 5.3% of strikes under a decoder

TABLE III. Inner-code distance d. Columns: $P_{\rm loss}$ at four deposits, the per-chip loss rate λ, the shortfall of the multi-chip module, and the share of the claim recovered by separate boxes, all at $\tau = 1$ ms.

d	Reading	$P_{\rm loss}$ (0.3/1/2/4 MeV)	λ (h^{-1})	Shortfall	Share
5	echo, conv. 0.5	0.094/0.21/0.39/0.57	9.7	184×	24%
7	echo, conv. 0.5	0.007*/0.053/0.17/0.42	1.2	220× (200–380)	15% (6–19)
11	echo, conv. 0.5	0.0004*/0.001*/0.008/0.022	0.046†	≥230×†	0–17%†
7	echo, conv. 1.0	0.041/0.21/0.47/0.73	4.9	214× (217–255)	19% (16–19)
11	echo, conv. 1.0	0*/0.003/0.035/0.083	0.016	2500× (234–6100)	0.3% (0.2–11)
7	no echo	0.39/0.83/0.97/0.97	37	180×	24%
11	no echo	0.010/0.28/0.63/0.79	1.8	414× (404–429)	7% (7–8)

*Interpolated through bins consistent with zero (Fig. 8). The 95% upper limits of those bins are about 0.01 at $d = 7$ and 5×10^{-4} to 2×10^{-3} at $d = 11$. Ranges: the unresolved bins set to zero and to their upper limits. †Resolved bins carry 1% of λ at $d = 11$. Here λ lies between 9×10^{-4} and 0.084 h^{-1}, and the shortfall between 230× and 6×10^4× (198× on the analysis curve). At $d = 11$, conversion 1.0, resolved bins carry 72% of λ, and without echo 95%.

with uniform priors and in 4.2% under one that knew the quasiparticle density [44]. On echoed chips, hard erasure on detection is worse than no outer code in a multi-chip module (0.13–0.74× across conversion brackets). Without echo, where most bursts do corrupt the inner code, it helps: 5.7× in the module, against 17× when detections are used as location hints by largest deposit, as below.

A soft policy instead treats detections as location hints, used only when the outer syndrome reports an error. The difference matters because a distance-2 outer code corrects an error whose location is known but can only detect one whose location is not. Blaming a random or the largest-deposit flagged chip, a soft policy is 70–110× better than hard erasure with no extra information. The remaining gap to the oracle, which knows which chips were lost, is misattribution in events that flag two or more chips.

We compute policies in three models. Table IV uses the loss model of Sec. II C. In it each chip is either lost, with probability $P_{\rm loss}$, or intact, and a policy fails when it cannot restore the lost chips (Appendix F). The flip model draws, for each flagged chip, whether its inner logical value flipped and the inner decoder's confidence from simulated bursts of similar deposit. It is used only where that confidence is needed (Appendix F). Table V works at the Pauli level, in which a lost chip carries a uniformly random Pauli error and the $[[4, 1, 2]]$ code is decoded exactly. Its protections are higher, and are compared only within that table. Figure 9 shows the resulting protection ladder in the loss and flip models.

Bursts too small to detect. Every policy result except the oracle rests on one assumption: bursts too small to detect do not destroy the inner logical qubit. Our distance-7 chain does not resolve the loss probability of deposits below about 0.4 MeV from zero (binned values up to $\sim 6 \times 10^{-3}$ per hit, 95% upper limits near 10^{-2}, Fig. 2). Such deposits are frequent ($\sim$39 h^{-1} per chip between 100 and 180 keV). If they carried even that much loss, the unflagged losses would dominate and every policy except the oracle would fall to about 1×. For the soft-policy values to stand, the unflagged loss rate, 39 h^{-1} times the loss probability per undetectable hit, must stay below λ divided by the soft protection (1.0/770 h^{-1} for separate boxes). That probability must therefore be below about 3×10^{-5}. Resolving that directly at distance 7 is out of reach. The idle code alone fails in about 2% of burst windows (Sec. II C), and separating so small an excess from that background at two standard deviations would take of order 10^9 shots per deposit bin. At distance 11 the upper limits fall to 5×10^{-4}–2×10^{-3} per hit (Fig. 8), still 15 to 60 times above that level. Whether a realistic inner code reaches it is open, and here it is an assumption. The separation reductions do not depend on it. The counts and the flagged reading use no loss curve, and the lost readings use the same curve for both layouts.

Without the assumption, the flag-only decoders of Table V (described below, no floor) fall to 2–4×. A de-

TABLE IV. Protection of the five-chip $[[4, 1, 2]]$ block against the true loss rate of one chip, loss model (echo, conversion 0.5). FLAG-ALL: every detected chip is erased. SOFT: detections are location hints used only when the outer syndrome fires. The blamed chip is chosen at random, by largest deposit, or by smallest inner-decoder complementary gap (GAP, computed in the flip model of Sec. III F and Appendix F). ORACLE: the decoder knows which chips were lost.

Policy	X1 (module)	XQ50 (separate boxes)
FLAG-ALL	0.18	7.2
SOFT-RANDOM	12	493
SOFT-RANKED	20	773
SOFT-GAP*	128	5610
ORACLE	717	22 000

*Flip model. In the same model SOFT-RANKED is 16 (X1) and 855 (XQ50), and ORACLE is 995 (X1) and 29 200 (XQ50). Loss probabilities below the 180 keV threshold are set to zero here, so ORACLE differs slightly from the achieved protection quoted in Sec. III C (717 vs 672 for X1). In the loss model the misattribution ceiling of ≈ 2900× (below) caps any soft policy in separate boxes, so SOFT-GAP's gain over SOFT-RANKED there is at most about 3.7×.

coder that knows which chip received any deposit, however small, keeps about 2000× for separate boxes, because a subthreshold hit still lands on one identifiable chip. That decoder uses every simulated deposit, with no sensor threshold and exact energies. What threshold and energy resolution a real sensor would need we have not quantified. Per-chip hit sensing with colocated sensors, as proposed for fault mitigation in Ref. [45], with a threshold below the qubits' own burst detection, could in principle make the policy results independent of the unresolved small-burst loss.

Detection threshold. With that assumption stated, we swept the threshold from 100 keV to 2 MeV, counting a loss on an unflagged chip as a failure, since the outer code detects but cannot locate it, and keeping losses below 180 keV at zero (the assumption above). A lower threshold adds false flags. A higher one leaves losses unflagged, and above the deposit at which losses begin every policy falls to about 0.2×, a five-chip block failing 5 times as often as one chip. At this floor, no threshold rescues hard erasure, and 180 keV is close to the best. The conclusion depends on the floor. Suppose every unflagged burst is harmless (the floor raised with the threshold). Then hard erasure in an echoed module overtakes no outer code at 0.5–0.7 MeV (0.6–6.8× at 0.5 MeV, 1.4–16× at 0.7 MeV across the echo brackets), while staying 4–70× below the soft policies at the same threshold. With the floor at 0.4 MeV, the top of the unresolved region, hard erasure stays below 1× at every threshold and conversion except one, conversion 1.0 with a 0.3 MeV threshold (1.5×).

An exact outer decoder. The soft rules blame one chip and ignore the code's weight-two stabilizer. We built an exact maximum-likelihood decoder of the $[[4, 1, 2]]$ code (stabilizers $XXXX$, $ZZZZ$, $ZZII$) that enumerates every Pauli configuration of the four data chips. It improves on the blame-one rule by only 1.0–1.7× (Table V, with floor). The gap to the oracle is missing information about which flagged chip was lost, not decoding. For separate boxes every soft policy approaches the same ceiling, $1/(16\lambda_{\text{flag}}\tau) \approx 2900\times$. Here λ_{flag} is the per-chip rate of burst detections, and $16 = n^2$ for $n = 4$ data chips (Appendix F). The ceiling is set by our deliberately pessimistic accounting of a loss coinciding with an unrelated burst within τ, so the true separate-box values are higher.

Inner-decoder confidence. Inner surface-code decoders can report their own confidence: the complementary gap, the weight difference between the best corrections in the two logical classes, is the soft output passed to the outer decoder in yoked surface codes [46]. That work credits related gap quantities to Refs. [47, 48], the latter under the name logical gap. Efficient soft-output decoders built on it have been proposed for hierarchical codes [49]. Soft inner-code information has been proposed for another concatenated scheme [18] and anticipated for algebraic outer codes [15]. Thresholding the gap to herald a logical erasure for an outer code was proposed in Ref. [50]. Under bursts the gap still separates failed from healthy shots (area under the curve 0.96 for 0.15–0.6 MeV deposits, falling to 0.79 at 2.5–5 MeV). In the flip model, blaming the flagged chip with the smallest gap gains a further 7–8× over SOFT-RANKED in both layouts, about half of the remaining distance to the oracle in log terms (Table IV). In the loss model the gain for separate boxes is capped at about 3.7× (Table IV footnote). The methods in this section are established. What we add is their size in our model, measured against hard erasure on detection for radiation bursts.

TABLE V. Exact maximum-likelihood (ML) outer decoding, Pauli model (echo, conversion 0.5), protection against one chip's loss rate. With floor: deposits below 180 keV assumed harmless. No floor: the loss curve as computed. Blame one: the largest-deposit flagged chip is blamed, as in SOFT-RANKED of Table IV. Pauli-level values exceed those of Table IV, because a scrambled qubit is error-free a quarter of the time, so compare within this table.

	With floor		No floor	
Policy	X1	XQ50	X1	XQ50
FLAG-ALL	0.27	9.5	0.30	2.7
Blame one flagged chip	83	1770	3.4	3.6
ML, flags known	89	1820	3.4	3.6
ML, deposits known	141	2130	109	2060
ORACLE	1090	33 100	1020	33 200

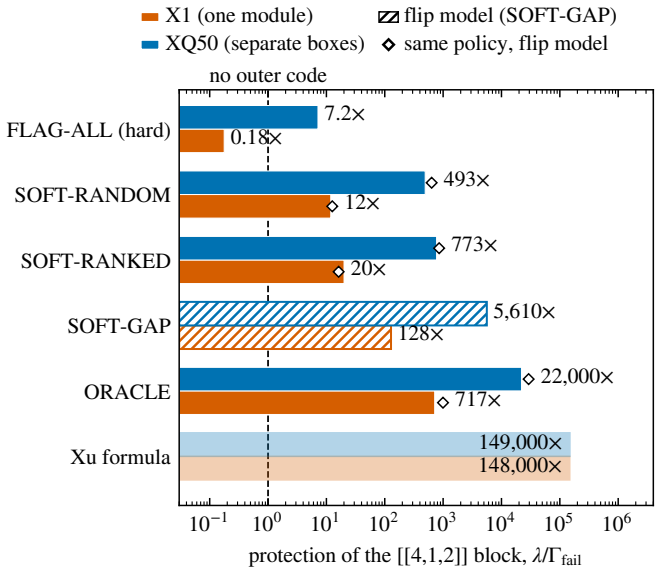


FIG. 9. Protection ladder for the five-chip block under four erasure policies and the oracle, for the multi-chip module and for separate boxes, beside the claim of Eq. (1). Here Γ_{fail} is the block's failure rate under each policy. The vertical line marks no outer code. Bars: loss model (Table IV), except the hatched SOFT-GAP bars, which are computed in the flip model. Diamonds: the same policy in the flip model. The misattribution ceiling that caps soft policies for separate boxes in the loss model (Sec. III F) is not drawn.

IV. DISCUSSION

Design rules. For a modular processor that protects against radiation with a cross-chip code, our results suggest seven rules. They hold in our model: one box-shell refrigerator, Willow-class chips and the [[4, 1, 2]] block. The first rule weakens by about 25% with doubled refrigerator copper, and none has been tested with copper sample boxes, real refrigerator radii or another electromagnetic physics model. (1) Place the chips of one code block at least 5 cm apart, in separate sample boxes, which in our model add a factor of about 2 over one shared box. (2) Do not mount packages vertically or stack chips face to face. (3) For the correlated term, prefer depth to lead shielding. This rule is an inference from the source shares of Sec. III B. We simulated neither. (4) Choose the inner code's distance with the correlated term in mind. Where our loss curves are resolved, a larger distance lowers independent failures faster than correlated ones, so the correlated term dominates more. (5) Treat burst detections as soft location hints, not erasures, on chips whose inner code usually survives a burst. (6) Rank flagged chips by the inner decoder's confidence. (7) Consider per-chip hit sensing, so that the decoder knows which chip any deposit struck. The sensor threshold and energy resolution this needs are not quantified here, and the benefit rests on an idealized decoder (Sec. III F).

What remains. Even with separate boxes and lost chips located perfectly, a block recovers only 15–25% of the protection its formula promises at $\tau = 1$ ms, and 5–8% at the 270 µs Xu *et al.* estimate. That is the packaging-limited ceiling. These values are for inner-code distance 7, on the analysis curve. At the nominal reading the share is 6–19% when the unresolved bins are set to zero or to their upper limits. At distance 11 the share is 7% without echo and is not resolved with echo (Table III). Locating lost chips from burst detections reaches 0.3–0.5% (soft policies, Table IV, at one reading). The misattribution ceiling sits at about 2% of the claim. Its accounting is deliberately pessimistic, so it does not bound what detection-based location can reach. No decoder in this paper closes the gap between detection-based location and the packaging-limited ceiling. In the Pauli model deposit-aware decoding reaches about 2130× for separate boxes, about 6% of the oracle's 33 100× (Table V), and per-chip hit sensing is untested here. Conversely, separate boxes keep beating a multi-chip module until τ reaches 160–220 ms across the erasure readings, so slower links between boxes are not an objection on these grounds. Lossier links are a separate cost we do not bound. They raise each chip's unheralded error rate, which a distance-2 outer code can detect but not correct. The coupling that remains between separate boxes is shared refrigerator mass and, possibly beyond about 10 cm, air showers, where our bundle estimate rests on an extrapolated spectrum. Chips in different refrigerators would remove the first. They would require links between refrigerators fast enough for the outer code's syndrome measurements. State transfer and entanglement between two refrigerators have been demonstrated [51], but not at the rate an outer code's syndrome cycle needs, and the noise of such links has a cost of its own [52].

Analysis plans and the expectations that failed. For most studies we wrote an analysis plan before the runs, with predictions, verdict criteria, or both. The plans are local files, not deposited in a registry. Their only timestamps are file dates and local version-control commits, the first made on 16 to 20 October 2026, after several of the runs, so neither record shows on its own that a plan preceded its runs. Three were amended with dated addenda. Two addenda were written before their runs. One of them, for the two-die device, restricts the 1 µm cut to a region around the dies. It followed a short test of run time and coincidence counts. With 20 000 muons each, the 1 µm cut everywhere gave 12 coincidences and the region cut 8. The two agree within the statistics, which cannot exclude a shift of tens of percent. The third was written after first counts of the study it extends had been seen. Some planned studies had predictions or an expected outcome but no pass and fail criteria: the erasure-policy comparison and the production-cut and inner-code distance runs. The 20 cm run was planned with neither, only to see whether the two-chip probability keeps falling. The detection-threshold sweep and the maximum-likelihood decoder were named in a plan without predictions. In the two-die device, the count of primaries entering both dies and the birth-volume run had no written plan. Other analyses without one include the 1 µm rerun and the orientation prediction for the device of Ref. [14], the comparisons with Fowler *et al.*, the resolved share of the distance results, the floor of the threshold sweep, the inner-code background, the particle shares of the correlated rates, the transfer to a second array, the G4CMP comparison and the bootstrap intervals.

Ten expectations written in the plans failed, and all are reported here. (1) The separation reduction was expected to change by less than a factor 1.5 between inner-code distances 5, 7 and 11. It falls 1.7-fold at conversion 0.5 and 1.8-fold without echo, though the distance-11 values carry wide intervals (Sec. III E). (2) With echo, the deposit at which half of logical qubits are lost lies above 5 MeV rather than in the predicted 0.1–3 MeV range. (3) The soft policy was expected to come within 2× of the oracle. It falls 10–90× short with echo (about 30× at the nominal conversion). (4) Hard erasure on detection was expected to be about 50 times worse, without saying worse than what. It is about 110 times worse than SOFT-RANKED and 3100–4100 times worse than the oracle (Table IV). (5) Under the realistic muon spectrum, one erasure reading moved by more than the planned factor of 1.5 (40 to 66×, consistent with statistical fluctuation). By the plan's rule this requires quoting the spectrum values alongside the mono-energetic ones, as we do. (6) The hadronic term in separate boxes was ex-

pected to stay below 20% of the muon term. Relative to the muon and gamma terms it reaches 26–28% for two readings at a proton-to-neutron ratio of 2. (7) The same runs were expected to move the 40–60× reduction by less than 20%. At echo, conversion 0.5, it rose from 40× to 58–61×. (8) The finer production cut was expected to raise the module's two-chip event rate at 0.5 MeV by 0–25%. It changed it by −16%, consistent with zero, and the only significant rise, +8%, is at 180 keV. (9) In the two-die device, the coincidence probability was expected either to reach the measured value or to fall below its interval. It lies inside the interval but below the measured value (Sec. II D). (10) In the same device, the median of the larger deposit in coincident events was expected to exceed the median deposit of events recorded on one die only more than threefold. The ratio is 1.3–1.8.

Three tests passed only marginally or in part. The muon-tagger reproduction passed cleanly only once the detector spacing was read from the published photograph rather than the text. The planned verdict on the first cross-module prediction, before the production-cut correction, was a marginal reproduction (Sec. II D). The heavier-refrigerator floor test passed at 0.3 MeV but not at 0.5 MeV (Sec. III D). One plan set no expectation but three verdicts. The complementary gap reached the middle one, closing about half the remaining distance to the oracle rather than coming within 3× of it.

Limitations. All results are simulations, anchored to the three measurements of Sec. II D. The erasure physics is calibrated to gap-engineered, echoed chips of the Willow class. Designs with heavy quasiparticle trapping or without gap engineering behave differently. On a second gap-engineered array [12], our shift-per-energy scale is 16–80 times off. The factor depends on the deposit energy and on whether the peak or the array-average shift is compared (Appendix B). Transport and qubit response cannot be separated as the cause. Absolute rates are for 0.5 mm dies, and the separation reductions are ratios at fixed die thickness. With the gamma flux inferred from a second device, our gamma burst rate on Willow is 3.4–4.2 times below the measured one at the calibrated thickness, so the gamma flux is not validated. It enters λ, and so the claimed protection, the shortfall and the recovered share (Sec. III A gives the size of the effect). It also enters the separation reductions, since gammas carry 11–24% of the module's correlated rate, and no simulated gamma event enters the separate-box rate. A higher gamma flux would increase the reductions, by an amount we have not computed. The refrigerator is a box-shell model. The separation rule has not been tested against an alternative electromagnetic physics constructor (for knock-on electrons and backscatter), copper rather than aluminum sample boxes or added magnetic shielding, refrigerator mass placed closer to the boxes, or real refrigerator radii. The lost readings apply a loss curve averaged over strike position. In the module, cross-chip secondaries may land preferentially near the facing chip edges, where the edge sinks remove phonons. The second chip's loss may then be overstated and the lost-reading separation ratios biased upward (the counts and the flagged reading are unaffected). The loss curve was computed on a 12 mm die and is applied to 10 mm chips, a mismatch we have not bounded. Radon, material contamination and cosmogenic activation are not modeled. Our muon flux normalization may be 20% high. The cross-module anchor rests on 8 events, and the muon-tagger anchor tests only the muon side of the chain. The two-die anchor rests on 10 events and leaves the detection threshold free. Its cavity height is bracketed, and its wall thickness and circuit-board thickness are assumed. The share of its coincidences traced to the box may depend on the distance to the lid. The cavity-height bracket was run for muons only and checked only for the share of direct crossings. The five-chip separate-box result rests on 36 events at the 0.5 MeV threshold. At distance 11 with echo the separation reduction is not resolved (without echo it is 19–65×). The inner decoder is not told about bursts. At distance 7 with 10^{-3} noise it cannot resolve whether bursts below the detection threshold destroy the logical qubit, and the erasure-policy results assume they do not (Sec. III F). Burst-aware decoding would change both the loss probabilities and the complementary gap. Such decoders reweight the decoder at the burst time [53], detect bursts from the syndrome and deform the code [54], or adapt the code around defects [55]. Correlated events not caused by radiation are not modeled. Examples are phonon bursts from stress relaxation or, in some packagings, mechanical perturbation from the cryocooler pulse tube [56]. A refrigerator-wide source such as the pulse tube could couple chips regardless of their boxes.

A direct test. The separation rule is testable on existing hardware. The same two-module measurement as Ref. [14] could be repeated with the modules in separate sample boxes at least 5 cm apart. Carrying over our results for 1 cm^2 silicon chip pairs, we expect a cross-module coincidence probability tens of times lower than on a shared motherboard. We have not simulated their sapphire modules in separate boxes, or their keV-scale thresholds, at which gamma coincidences may matter more. At 2 cm our pairs show only a 9–11× reduction.

V. CONCLUSION

The suppression promised by cross-chip erasure codes rests on chips failing independently. In our simulation of Willow-class chips in one model refrigerator, they do not. A single cosmic-ray particle passing through the material the chips share erases several chips at a rate that exceeds the independent term by one to three orders of magnitude. A faster outer-code cycle cannot reduce that rate. Where our loss curves are resolved, it also grows relative to the independent term as the inner code grows. Separate sample boxes at least 5 cm apart cut it 40 to 60 times at inner-code distance 7. With lost chips located

perfectly, the block then recovers 15–25% of its formula's protection on the analysis curve. Located from burst detections, it recovers 0.3–0.5%. On chips whose inner code usually survives a burst, the erasure policy decides whether the outer code helps at all. Modular architectures that count on independence between chips or code blocks should budget for the correlated term explicitly.

ACKNOWLEDGMENTS

The author used Claude (Anthropic) as an assistant in this work. This work received no external funding.

Competing interests. The author is the founder of Cogitan, which develops simulation software for superconducting quantum hardware, including the phonon-transport model used here.

Data availability. Geant4 geometry and sweep scripts, analysis code, and the per-event deposit files behind every figure and table are deposited at Zenodo, doi:10.5281/zenodo.24412856. The phonon-transport model and its calibration are specified in the Appendices. Remaking the figures and tables also needs the figure script and the script that rechecks every number in the text, together with the stored inputs they read: the surface-code loss curves, the complementary-gap map and the cached SOFT-GAP policy result. The loss curves and the gap map are computed with the phonon-transport code, which is part of a Cogitan product and is not released. The deposit includes these stored inputs together with the figure script and the checking script. Every lost reading, Figs. 2, 8 and 9, and Tables III–V can therefore be remade from released files. Recomputing the loss curves themselves requires an implementation of the model in the Appendices.

Appendix A: Phonon transport

Appendices A–G specify the chain of Sec. II B, and Table VI collects its constants.

Propagation. All phonons from a deposit start at the strike point with isotropic directions ($\cos\theta$ uniform on $[-1,1]$, azimuth ϕ uniform) and reflect specularly between the faces of a slab of thickness $t_{\rm tr}$. Successive face encounters are therefore separated by a fixed lateral step, and after k encounters a phonon is at

$$\mathbf{r}_k = (x_0, y_0) + k\, t_{\rm tr}\, |\tan\theta|\, (\cos\phi, \sin\phi), \quad k = 1, 2, \ldots. \tag{A1}$$

Encounters alternate between faces, starting with the top if $\cos\theta > 0$. The birth depth is not tracked (the first step is a full thickness). At each top-face encounter the phonon is absorbed by the film, breaking pairs at $\mathbf{r}_k$, with probability

$$p_{\rm f} = f_{\rm film}\, p_{\rm abs} = 0.3 \times 0.8 = 0.24, \tag{A2}$$

and otherwise reflects. Coverage $f_{\rm film}$ and per-encounter absorption $p_{\rm abs}$ enter only as this product, and we quote $f_{\rm film}$ at fixed $p_{\rm abs} = 0.8$. The bottom face reflects every phonon. A phonon that leaves the die laterally is removed (edges and mount are perfect sinks). There is no energy tracking and no downconversion length. More detailed phonon-transport models of qubit arrays exist [57]. We chose this one for speed, and because only one of its constants, K, is fitted to a single measured value (the others are set within bands, Appendix B). Phonons alive after 400 encounters are discarded, and $N_{\rm ph} = 2\times10^5$ phonons are traced. Since lateral paths are straight rays and the die is convex, a phonon that leaves never returns. One run on an unbounded slab therefore serves every strike: the hit cloud is translated by (x_0, y_0) and hits off the die are dropped, which is exact.

Energy and shift. For a deposit of energy E, qubit q absorbs

$$E_q = E\, n_q / N_{\rm ph}, \tag{A3}$$

with n_q the film absorptions inside its cell, a square of area $(1.15\ \text{mm})^2$ centered on the qubit. The peak frequency shift of qubit q is

$$\delta f_q = K\, E_q, \qquad K = 119.13\ \text{kHz/keV}, \tag{A4}$$

the chain's only constant fitted to a single measured value ($t_{\rm tr}$, $t_{\rm dep}$ and $p_{\rm f}$ are set within bands, Appendix B). The constant K absorbs the fraction of E reaching the film as pair-breaking phonons, the sensitive area within a cell and $\delta f / x_{\rm qp}$, where $x_{\rm qp}$ is the quasiparticle density (Appendix C). Transport is instantaneous on the 1 µs scale of a round, so δf_q is the shift at $t = 0$.

Two thicknesses. The *deposition* thickness $t_{\rm dep} = 0.3$ mm sets the Geant4 deposit spectrum used in calibration. The *transport* thickness $t_{\rm tr} = 0.5$ mm enters only Eq. (A1). The muon burst rate pins $t_{\rm dep}$: 16.7% of sea-level muons deposit ≥180 keV at 0.3 mm and 62.9% at 0.5 mm, giving 9.55 h^{-1} at 0.3 mm and 35.9 h^{-1} at 0.5 mm, against 9.3 h^{-1} measured (Table VII). The footprint needs a longer step. At $t_{\rm tr} = 0.3$ mm, with K refitted, the median burst is 9 qubits and 0.1–0.5% of bursts exceed half the array, against the measured 15 qubits and 5.4%. The value 0.5 mm is the substrate thickness reported for Google's earlier flip-chip platform [4]. We first attributed the split to phonons crossing the bump bonds into the carrier chip, but our own model does not support that. Joining a carrier to the die with a per-encounter bump transmission $T_{\rm b}$ always *shrinks* the in-die footprint, because the carrier drains the long-traveling phonons that make the far field. No $T_{\rm b} \le 0.76$ reproduces the 0.5 mm slab even with a die-sized carrier (half-absorption radius at most 1.36 mm against 1.70 mm). The transport thickness is therefore an effective parameter whose cause is open. Candidates are a die thicker than the muon rate implies, diffuse surface scattering, or the qubit-response model. We compared it with a full phonon Monte Carlo (Geant4 Condensed Matter Physics, G4CMP [58]) for 0.381 mm silicon with a patterned aluminum film, in an unpublished comparison. The radial pair-breaking distribution of Eq. (A1)

TABLE VI. Constants of the chain. Fitted: adjusted to data (Appendix B). Constrained: set within a band allowed by data or a published device (Appendix B). Measured: from Ref. [10]. Chosen: a modeling choice. Window end: the per-qubit error p_q below which the burst window ends (Appendix D).

Quantity	Value	Status
t_{tr}, t_{dep}	0.5, 0.3 mm	Constrained
Top-face absorption p_{f}	0.3×0.8	Constrained
Qubit cell	$(1.15\ \mathrm{mm})^2$	Chosen
K	119.13 kHz/keV	Fitted
Detection threshold	180 keV	[10, 31]
Burst-size threshold	200 kHz	[10]
a, f_q	0.77, 6 GHz	Measured
$1/r$	88 ns	Measured
Echo A, β	0.02, 1.32	Measured
No-echo A	0.17	Measured
Conversion c	0.25, 0.5, 1.0	Chosen
Base noise, round	10^{-3}, 1 µs	Chosen
Window end	2×10^{-4}	Chosen

misplaces 10.2% of the distribution against a 6.6% statistical floor, G4CMP being more peaked near the strike.

Appendix B: Calibration

All targets are Willow measurements from Ref. [10]. They record 265 bursts in 5.2 h on 60 qubits, a total rate of 1/(71 s). The median burst has 15 qubits shifted by more than ≈200 kHz, and the median peak shift is 2 MHz. Bursts over half the array occur at 1/(22 min), and bursts last 1–2 ms. The 200 kHz threshold is their $3\sigma_f$, with σ_f the width of a qubit's frequency noise. A deposit makes a burst when it reaches 180 keV, the threshold at which the measured rate is consistent with the computed rate of deposits [10, 31].

Geant4 (Sec. II A) scores deposits in a bare $10\times10\ \mathrm{mm}^2$ silicon die of thickness t_{dep}. The sources are 4 GeV muons with a $\cos^2\theta$ zenith intensity, uniform over the die, and gammas from an enclosing sphere of radius R with cosine-law inward directions. A rate is then $\Phi\,\pi R^2$ times the fraction of generated primaries above threshold, with Φ the source flux. The Willow die area is taken as 0.95 cm^2 (72 cells of 1.15 mm). The 60 measured qubits are an 8×8 grid minus corners at 1.15 mm pitch on a 10.2 mm die, with axis-aligned cells. Strikes are uniform over the die with energies resampled from the deposits ≥180 keV. Since every shift is proportional to K, the fit to the median peak shift is exact,

$$K = 2\ \mathrm{MHz} \,/\, \mathrm{median}_{\mathrm{events}} \max_q E_q, \tag{B1}$$

and the pinned value is the mean over 8 seeds of 3000 strikes on the gamma-line spectrum (seed-to-seed s.d. 1.5 kHz/keV). Burst size is the number of qubits with $\delta f_q > 200$ kHz, and a large burst has more than 30.

TABLE VII. Calibration against Willow [10]. Muon and gamma targets are our split of the total rate by the 18.4% muon fraction measured on another device [5], close to the 17% cosmic-ray share of correlated events measured independently in Ref. [6].

Observable	Measured	Model	Role
Muon burst rate	9.3 h^{-1}	9.55 h^{-1}	Sets t_{dep}
Gamma burst rate	41.4 h^{-1}	9.8–12.0 h^{-1}	Not used
Median peak shift	2 MHz	2 MHz	Fits K
Median burst size	15 of 60	14	Constrains p_{f}
Bursts over half	5.4%	3.4–5.5%	Constrains p_{f}
Burst duration	1–2 ms	2.03 ms	Check

Fitted and checked (Table VII). One constant, K, is fitted. The transport thickness is the published prior [4], and the deposition thickness is where the muon rate agrees. The split is bounded (between die and die plus carrier) but inferred, since Willow's thicknesses are unpublished. The film coverage is set to $f_{\mathrm{film}} = 0.3$, the value closest to both footprint observables within a band (0.25–0.4, found with 0.5 mm deposits) that passes both preset bands. These are median size 11–19 (a chosen tolerance) and large-burst share 2.9–8.9% (Poisson 95% on ∼14 bursts). The size distribution therefore constrains one transport parameter and is not an independent prediction. The test is that a band passes rather than a point. The duration has no parameter of ours fitted to it (Appendix C) and does not involve the transport. It is still a consistency check rather than an independent prediction, because the recombination constant r was fitted by Ref. [10] to the measured burst recoveries.

The gamma rate is not used. We infer the gamma flux from the burst rates of a second device in another laboratory [5], which gives 3.2–4.5 $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ by spectrum. At that flux the predicted Willow gamma rate at $t_{\mathrm{dep}} = 0.3$ mm is 3.4–4.2 times below its share of the total, and no single thickness satisfies both channels within a factor of 1.5. Since the muon channel has no free normalization, we read this as a difference in site gamma flux, which varies more than this between buildings. Gammas carry 11–24% of the multi-chip module's correlated rate, and no simulated gamma event enters the separate-box rate (Sec. III D), so this flux also enters the separation reductions (Sec. II D). As a check, K implies $\approx 3 \times 10^4\ \mu\mathrm{m}^3$ of low-gap aluminum (57% pair-breaking efficiency, $2\Delta = 0.36$ meV, with Δ the superconducting gap), the scale of transmon pads.

Transfer fails. We tested the chain on a gap-engineered aluminum array with energy-calibrated shifts [12]. With K scaled by that device's frequency, gap, coverage and cell area, the chain overpredicts the shift 16× at 699 keV (1.6 MHz against an upper bound of 100 kHz). For the array-average shifts it overpredicts 43× as one scale fit-

ted through their Ramsey measurement at all five energies (57–699 keV). Fitted energy by energy, the scale falls from 70–80× at 57–216 keV to 22× at 699 keV. The measured shift grows faster than linearly with energy, which the model lacks. A ground-plane drain, film volume, the junction side on which quasiparticles settle, and the two devices' packaging all scale the same factor and cannot be separated. The chain is calibrated on Willow and fails on the one other gap-engineered array we tested, so we apply it only to Willow-class chips. Burst sizes on other designs are not validated and are probably too large.

Appendix C: Quasiparticle recovery

Following Ref. [10], $\delta f_q = a f_q x_{\mathrm{qp},q}$, where $x_{\mathrm{qp},q}$ is the quasiparticle density (per Cooper pair) at qubit q, $f_q = 6$ GHz is the qubit frequency, and $a = 0.77$ is a dimensionless coefficient (their value for $\Delta/h = 52$ GHz, gap difference 12 GHz). Without traps the density decays by recombination, $\dot{x}_{\mathrm{qp}} = -r x_{\mathrm{qp}}^2$, $r = 1/(88\ \mathrm{ns})$ (central value of $1/(88 \pm 12\ \mathrm{ns})$), so

$$x_{\mathrm{qp},q}(t) = \frac{x_{q0}}{1 + r x_{q0} t}, \qquad x_{q0} = \frac{\delta f_q}{a f_q}, \tag{C1}$$

i.e. $\delta f_q(t) = \delta f_q/(1 + t/t_q)$ with $t_q = a f_q/(r\,\delta f_q)$. A harder burst recovers faster (x_{q0} is floored at $\delta f_q = 1$ Hz). The time to fall to 200 kHz, $(a f_q/r)(1/200\ \mathrm{kHz} - 1/\delta f_q)$, saturates at 2.03 ms for any hard burst (1.83 ms at a 2 MHz peak). This is the modeled duration. At late times $\delta f_q(t) \to a f_q/(rt) = 406.6\ \mathrm{Hz\,s}/t$ whatever the initial shift, so all strongly hit qubits share one tail and the burst window (Appendix D) has nearly the same length for every event. The 35 µs T_1 burst of Ref. [10], which needs quasiparticles above the gap difference, is not modeled.

Appendix D: From shift to error

The shift becomes excess detection probability per round through the mapping measured with an echo in the stabilizer circuit [10], 2% at 1 MHz and 5% at 2 MHz:

$$\varepsilon(\delta f) = A\,(\delta f/1\ \mathrm{MHz})^{\beta}, \qquad A = 0.02,\ \beta = 1.32. \tag{D1}$$

The no-echo variant has $A = 0.17$ (17% at 1 MHz, measured) and the same, assumed, β. It gives 42% at 2 MHz against ≈35% measured (the two no-echo points imply $\beta \approx 1.04$), so it is a pessimistic bound. The mapping was measured in a repetition code. The per-qubit Pauli error per round, with c the conversion from detection probability to Pauli error, is

$$p_q(t) = \min\big[0.5,\ c\,\varepsilon(\delta f_q(t))\big], \tag{D2}$$

with conversion $c = 0.5$, bracketed by 0.25 and 1.0, applied alike to data and measure qubits.

Segments. The burst is cut into segments of n_j rounds, $n_0 = 20$, $n_{j+1} = \lceil 1.25 n_j \rceil$. Every round of segment j uses p_q at the segment midpoint, $t_j + n_j \tau_{\mathrm{r}}/2$, where t_j is the segment's start time and $\tau_{\mathrm{r}} = 1$ µs the round time. Segments are added until $\max_q p_q < 2 \times 10^{-4}$, or until one would start at or after round 9000 (the straddling segment is kept whole). By the common tail of Appendix C the window is set by the bracket, not the deposit: 4549 rounds for $c = 0.25$, 7164 for $c = 0.5$, and 11 253 for $c = 1.0$ and no echo. These two brackets hit the cap and so bias P_{flip} slightly low. Qubits are grouped by $\log_{10} p_q$ rounded to 0.01, with one noise instruction per group, and $p_q < 10^{-5}$ is dropped. These segments, this stopping rule and the deposit range of the loss curve (30–5000 keV, Appendix E) differ from the analysis plan. The plan fixed 50-round segments, run until every shift was under 50 kHz, and deposits from 50 keV to 5 MeV.

Appendix E: Surface-code simulation

The inner code is Stim's `surface_code:rotated_memory_z` at $d = 7$ [29] (49 data and 48 measure qubits), with as many rounds as the window. Base noise is 10^{-3} on all four generator channels: depolarizing after every Clifford gate (two-qubit after CNOTs), a flip after every reset, depolarizing on data qubits at the start of each round, and a flip before every measurement. The last two are set to 0.999×10^{-3} and 0.998×10^{-3} to mark their positions. Directly after these markers the burst adds `DEPOLARIZE1`(p_q) on each data qubit at the start of each round and `X_ERROR`(p_q) on each qubit before its measurement.

Stim's coordinates are scaled so nearest neighbors (data to measure) are 1.15 mm apart, centered on a 12×12 mm^2 die (the patch spans 11.4 mm). The qubits form a square lattice rotated by 45°, and each qubit's cell in Eq. (A3) is its Voronoi cell, a rotated square of side 1.15 mm. Absorptions are assigned by rounding rotated lattice coordinates. Transport uses the calibrated t_{tr} and p_{f}.

The decoder is PyMatching 2 [30] built from the detector error model of the *base* circuit with the same rounds, so it is not told about the burst. From the logical failure fractions of the burst circuit, f_{b}, and of the base circuit, f_0 (same shots, independent seed),

$$P_{\mathrm{flip}} = \max[0,\ (f_{\mathrm{b}} - f_0)/(1 - f_0)]\,. \tag{E1}$$

A Z memory sees X-type logical errors only, while a burst causes Z-type ones at a similar rate. A fully scrambled logical flips its Z value half the time. In both limits the loss probability is close to twice the flip probability, so $P_{\mathrm{loss}} = \min(1, 2P_{\mathrm{flip}})$ (this factor was added after the no-echo run, Sec. II B).

Loss curve. Per bracket, 120 simulated strikes have E log-uniform on $[30, 5000]$ keV and (x_0, y_0) uniform over

the die, 600 shots each. The flip probability $P_{\rm flip}$ is averaged, unclipped, in 8 log-spaced bins on $[30, 5000]$ keV ($\approx \pm 0.002$ per bin), which averages over strike position. Negative bin means are set to zero, and a running maximum makes the curve monotone. The curve is interpolated in $\log E$–$\log P$ between bin centers (geometric means), ramps linearly to zero at half the lowest nonzero center, is extrapolated above the highest center on the last two bins' slope, and is clipped to $[0, 1]$. With echo the bins below 0.4 MeV are not resolved from zero. The running maximum carries their noise upward, moving λ by up to $\sim 2\times$ but the achieved protection by about 10% (tested by zeroing below 300 and 400 keV). The curve depends on E only and is applied to the deposits in the $10 \times 10 \times 0.5$ mm^3 chips of Sec. II A, which are thicker than $t_{\rm dep}$ and smaller than the 12 mm die used here.

Appendix F: Code-level rates and erasure policies

Each primary of source s stands for $w_s = \Phi_s \pi R_s^2 / N_s$ events per second, with N_s primaries generated on a horizontal disk of radius R_s (muons, $\Phi = 1$ cm^{-2} min^{-1} through a horizontal plane [24]) or a sphere (gamma lines, $\Phi = 1.7$ cm^{-2} s^{-1} [26]). In event e, with weight $w_e = w_s$ for an event from source s, chip i is erased with probability $P_i^{(e)} = P(E_i)$. The chips are erased independently given the deposits, since the chips' noise is independent. Here P is $\min(1, 2P_{\rm flip})$ for the lost readings and $\Theta(E_i - 180\ \text{keV})$, with Θ the unit step function, for the flagged reading. With n_c chips,

$$\lambda = \frac{1}{n_c} \sum_e w_e \sum_i P_i^{(e)}, \tag{F1}$$

$$\Gamma_{\rm corr} = \sum_e w_e \Pr\Big[\sum_i L_i^{(e)} \geq D\Big], \tag{F2}$$

where $L_i^{(e)} \sim \text{Bernoulli}(P_i^{(e)})$ indicates that chip i is lost. The Poisson-binomial tail is computed by recursion over chips. For the five-chip block (four data chips and the ancilla) $n_c = 5$ and $D = 2$, and Eq. (F2) is exact, with no pair approximation. The independent rate $\Gamma_{\rm ind}$ is Eq. (1) with $n = 4$, $D = 2$ and $\tau = 1$ ms. Achieved protection is $\lambda/(\Gamma_{\rm ind} + \Gamma_{\rm corr})$, claimed $\lambda/\Gamma_{\rm ind}$. For the 3×3 arrays (Table I) no block is placed. There λ is averaged over nine chips and $\Gamma_{\rm corr} = n\lambda p_2$, with $p_2 = \Gamma_{\geq 2}/\Gamma_{\geq 1}$ the ratio of the rates of events erasing ≥ 2 and ≥ 1 of the nine. This is an upper-side estimate for a block of fewer chips.

Policies (five-chip block, Table IV). Chip i is flagged, $F_i = 1$, if $E_i \geq 180$ keV, and lost with probability $P_i = F_i P_{\rm loss}(E_i)$, so losses below threshold are zeroed. Let $u_j = P_j \prod_{k \neq j}(1 - P_k)$ be the probability that only j is lost, $P_{\geq 2}$ the probability that two or more are lost, and $n_F = \sum_i F_i$. The per-event failure probabilities are then $\Theta(n_F - 2)$ for FLAG-ALL, $P_{\geq 2}$ for ORACLE, $P_{\geq 2} + (1 - 1/n_F)\sum_j u_j$ for SOFT-RANDOM, and $P_{\geq 2} + \sum_{j \neq j^*} u_j$ for SOFT-RANKED. Here $j^* = \arg\max_k P_k$ is the flagged chip of largest deposit, since $P_{\rm loss}$ is nondecreasing. A soft policy corrects a single loss only if the chip it blames when the outer syndrome fires is the lost one. Let $\lambda_{\rm loss}$ denote the per-chip loss rate, the λ of the main text, and $\lambda_{\rm flag}$ the per-chip rate of burst detections. To the event-weighted sums we add the independent terms $n\lambda_{\rm flag}(n+1)\lambda_{\rm flag}\tau$ (FLAG-ALL) or $n\lambda_{\rm loss}(n+1)\lambda_{\rm loss}\tau$ (others). For soft policies we also add a misattribution term $n\lambda_{\rm loss}\, n\lambda_{\rm flag}\tau$, for a loss and an unrelated flag within τ, always counted as misattributed (an upper bound). Protection is $\lambda_{\rm loss}$ over the failure rate, so every policy is measured against the true loss rate of one chip.

SOFT-GAP blames the flagged chip with the smallest complementary gap (Appendix G) and needs per-shot outcomes, so it is computed by Monte Carlo in a *flip model*. Each flagged chip takes 400 draws of (flip, gap). Each draw is a random shot of one of the 4 simulated strikes nearest in deposit. A draw fails if ≥ 2 chips flipped, or if one flipped and another was blamed. The other soft policies and ORACLE are recomputed in the same model. It differs from the loss model in four ways: no factor of 2, base-noise flips included, burst window ended at 10^{-3} added error, and independent terms omitted. Policies are compared only within one model.

Pauli model (Table V). The outer code has stabilizers $XXXX$, $ZZZZ$ and $ZZII$ and logical operators $XXII$ and $ZIZI$. A lost chip carries I, X, Y or Z with probability 1/4 each, and under FLAG-ALL every flagged chip does. A lost ancilla (the center chip of the plus) leaves the round intact only if no data chip carries an error. Per event, all 256 Pauli configurations of the four data chips are enumerated. For each syndrome, the decoder picks the logical class of largest prior probability given what it knows: the flags, the per-chip deposits, or, for ORACLE, the lost chips.

Appendix G: Complementary gap

The soft output is the complementary gap of yoked surface codes [46]. The logical observable becomes an extra detector, the syndrome is decoded with it forced to 0 and to 1, and the gap $|m_0 - m_1|$ is the weight difference between the best corrections in the two classes. The error model stays graphlike, with every error touching at most two detectors so that matching applies, only if every observable-flipping error touches one detector. Stim places the observable on an interior data row, so we move it to a boundary row of the final data measurements. That is an equivalent logical, since rows differ by measured Z stabilizers. We take the first of $y_{\min}$, $y_{\max}$, $x_{\min}$, $x_{\max}$ for which every observable-flipping mechanism of the decomposed error model touches at most one detector (chosen once, on three rounds). In the base model's detector error model the observable is replaced by a new detector D_N. Two PyMatching decoders are built from the base model: the ordinary one, giving the prediction

and weight m, and the one with D_N, run with $D_N = 0$ and 1 to give m_0, m_1.

Validation. On a $d = 5$, 10-round memory with 3×10^{-3} on all four channels (2000 shots), every observable edge is single-detector, $\min(m_0, m_1) = m$ in every shot, and the smaller-weight class equals the prediction in 100% of shots.

Under bursts. The gap map uses Appendices A–E at $d = 7$, echo, $c = 0.5$, with the window truncated when the added error falls to 10^{-3} (2283 rounds instead of 7164) to afford the forced decodes: 80 events, E log-uniform on $[150, 5000]$ keV, uniform strikes, 300 shots each. The decoders are built from the base model and are not told about the burst. Per deposit bin, the area under the curve for "small gap means failed" is 0.96 at 0.15–0.6 MeV, falling to 0.79 at 2.5–5 MeV as the gaps of successful shots shrink.

[1] A. P. Vepsäläinen *et al.*, Impact of ionizing radiation on superconducting qubit coherence, Nature **584**, 551 (2020), author correction: Nature **586**, E8 (2020), doi:10.1038/s41586-020-2754-2.

[2] L. Cardani *et al.*, Reducing the impact of radioactivity on quantum circuits in a deep-underground facility, Nat. Commun. **12**, 2733 (2021).

[3] C. D. Wilen *et al.*, Correlated charge noise and relaxation errors in superconducting qubits, Nature **594**, 369 (2021).

[4] M. McEwen *et al.*, Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits, Nat. Phys. **18**, 107 (2022).

[5] X. Li, J. Wang, Y.-Y. Jiang, G.-M. Xue, *et al.*, Cosmic-ray-induced correlated errors in superconducting qubit array, Nat. Commun. **16**, 4677 (2025).

[6] P. M. Harrington, M. Li, M. Hays, W. Van De Pontseele, *et al.*, Synchronous detection of cosmic rays and correlated errors in superconducting qubit arrays, Nat. Commun. **16**, 6428 (2025).

[7] J. M. Martinis, Saving superconducting quantum processors from decay and correlated errors generated by gamma and cosmic rays, npj Quantum Inf. **7**, 90 (2021).

[8] V. Iaia, J. Ku, A. Ballard, C. P. Larson, E. Yelton, C. H. Liu, S. Patel, R. McDermott, and B. L. T. Plourde, Phonon downconversion to suppress correlated errors in superconducting qubits, Nat. Commun. **13**, 6425 (2022).

[9] M. McEwen, K. C. Miao, J. Atalaya, A. Bilmes, *et al.*, Resisting high-energy impact events through gap engineering in superconducting qubit arrays, Phys. Rev. Lett. **133**, 240601 (2024).

[10] V. D. Kurilovich, G. Roberts, L. S. Martin, M. McEwen, *et al.*, Correlated phase error bursts in a gap-engineered superconducting qubit array, Phys. Rev. X **16**, 021025 (2026).

[11] R. Acharya *et al.* (Google Quantum AI and Collaborators), Quantum error correction below the surface code threshold, Nature **638**, 920 (2025), author correction: Nature **653**, E5 (2026), doi:10.1038/s41586-026-10559-8.

[12] H. D. Pinckney, T. McJunkin, A. W. Hunt, P. M. Harrington, *et al.*, Characterization of radiation-induced errors in superconducting qubits protected with various gap-engineering strategies, PRX Quantum 10.1103/plgq-byn7 (2026), published online 22 October 2026.

[13] Q. Xu, A. Seif, H. Yan, N. Mannucci, B. O. Sane, R. Van Meter, A. N. Cleland, and L. Jiang, Distributed quantum error correction for chip-level catastrophic errors, Phys. Rev. Lett. **129**, 240502 (2022).

[14] X. Wu, Y. J. Joshi, H. Yan, G. Andersson, A. Anferov, C. R. Conner, *et al.*, Mitigating cosmic-ray-like correlated events with a modular quantum processor, Phys. Rev. Appl. **24**, 044022 (2025).

[15] A. Wills, M. E. Beverland, L. S. Bishop, J. M. Gambetta, P. Rall, V. Siddhu, and A. W. Cross, Concatenating algebraic codes over high-rate quantum LDPC codes (2026), arXiv:2605.21898.

[16] E. Sutcliffe and C. M. Westoby, Tolerating device failure in distributed quantum computing (2026), arXiv:2605.11088.

[17] G. J. Norris, K. Dalton, D. Colao Zanuz, A. Rommens, A. Flasby, M. Bahrami Panah, F. Swiadek, C. Scarato, C. Hellings, J.-C. Besse, and A. Wallraff, Performance characterization of a multi-module quantum processor with static inter-chip couplers, EPJ Quantum Technol. **13**, 29 (2026).

[18] C. A. Pattison, A. Krishna, and J. Preskill, Hierarchical memories: Simulating quantum LDPC codes with local gates, Quantum **9**, 1728 (2025).

[19] G. Moshel, O. Rabinowitz, E. Blumenthal, and S. Hacohen-Gourgy, Die separation for mitigation of phonon bursts in superconducting circuits, Appl. Phys. Lett. **127**, 052601 (2025).

[20] A. R. Castelli, K. M. Beck, L. D. H. Alegria, L. A. Martinez, K. R. Chaves, S. R. O'Kelley, N. Materise, J. L. DuBois, and Y. J. Rosen, Superconducting-qubit decoherence correlated with detected radiation events, Phys. Rev. Appl. **26**, 044052 (2026).

[21] S. Agostinelli *et al.*, Geant4: a simulation toolkit, Nucl. Instrum. Methods Phys. Res. A **506**, 250 (2003).

[22] J. Allison *et al.*, Geant4 developments and applications, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[23] J. Allison *et al.*, Recent developments in Geant4, Nucl. Instrum. Methods Phys. Res. A **835**, 186 (2016).

[24] R. L. Workman *et al.* (Particle Data Group), Review of particle physics, Prog. Theor. Exp. Phys. **2022**, 083C01 (2022), cosmic-ray review, Sec. 30.

[25] M. Guan, M.-C. Chu, J. Cao, K.-B. Luk, and C. Yang, A parametrization of the cosmic-ray muon flux at sea-level (2015), arXiv:1509.06176.

[26] F. De Dominicis, T. Roy, A. Mariani, M. Bal, *et al.*, Evaluating radiation impact on transmon qubits in above and underground facilities, EPJ Quantum Technol. **13**, 47 (2026).

[27] B. Loer, P. M. Harrington, B. Archambault, E. Fuller, *et al.*, Abatement of ionizing radiation for superconducting quantum devices, J. Instrum. **19**, P09001.

[28] A. Gold, J. P. Paquette, A. Stockklauser, M. J. Reagor, *et al.*, Entanglement across separate silicon dies in a mod-

ular superconducting qubit device, npj Quantum Inf. **7**, 142 (2021).
[29] C. Gidney, Stim: a fast stabilizer circuit simulator, Quantum **5**, 497 (2021).
[30] O. Higgott and C. Gidney, Sparse Blossom: correcting a million errors per core second with minimum-weight matching, Quantum **9**, 1600 (2025).
[31] J. W. Fowler, P. Szypryt, R. Bunker, E. R. Edwards, *et al.*, Spectroscopic measurements and models of energy deposition in the substrate of quantum circuits by natural ionizing radiation, PRX Quantum **5**, 040323 (2024).
[32] X. Wu, H. Yan, G. Andersson, A. Anferov, M.-H. Chou, C. R. Conner, J. Grebel, Y. J. Joshi, *et al.*, Modular quantum processor with an all-to-all reconfigurable router, Phys. Rev. X **14**, 041030 (2024).
[33] A. Mariani, L. Cardani, M. Bal, N. Casali, *et al.*, A cryogenic muon tagging system based on kinetic inductance detectors for superconducting quantum processors, Quantum Sci. Technol. **11**, 025053 (2026).
[34] J. Fowler, I. Fogarty Florang, N. Nakamura, D. Swetz, P. Szypryt, and J. Ullom, Computed models of natural radiation backgrounds in qubits and superconducting detectors, IEEE Trans. Appl. Supercond. **35**, 1700305 (2025).
[35] E. Bertoldo, V. Pérez Sánchez, M. Martínez, M. Martínez, H. Khalife, and P. Forn-Díaz, Cosmic muon flux attenuation methods for superconducting qubit experiments, New J. Phys. **27**, 023014 (2025).
[36] C. H. Liu, A. Ballard, D. Olaya, D. R. Schmidt, J. Biesecker, T. Lucas, J. Ullom, S. Patel, O. Rafferty, A. Opremcak, K. Dodge, V. Iaia, *et al.*, Single flux quantum-based digital control of superconducting qubits in a multichip module, PRX Quantum **4**, 030310 (2023).
[37] D. Broadbent, E. W. Kellermann, and M. A. Hakeem, The density spectrum and structure of extensive cosmic-ray air showers at sea level, Proc. Phys. Soc. A **63**, 864 (1950).
[38] R. J. Norman, The energy spectrum of extensive air showers, Proc. Phys. Soc. A **69**, 804 (1956).
[39] JEDEC Solid State Technology Association, *Measurement and reporting of alpha particle and terrestrial cosmic ray-induced soft errors in semiconductor devices*, Standard JESD89A (JEDEC, 2006).
[40] J. D. Chadwick, C. Kang, J. Viszlai, S. F. Lin, and F. T. Chong, Averting multi-qubit burst errors in surface code magic state factories, in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)* (2024) pp. 1089–1101.
[41] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the T_1 limit in superconducting circuits, Phys. Rev. X **13**, 041022 (2023).
[42] J. Kim, Circuit-aware protection against burst errors in lattice surgery (2026), arXiv:2611.14583.
[43] N. J. Coble, M. Ye, and N. Delfosse, Correction of chain losses in trapped ion quantum computers (2025), arXiv:2511.16632.
[44] A. K. Nayak, P. G. Baity, P. J. Love, N. Jeon, B.-J. Yoon, A. Hoisie, and L. R. Varshney, Iterative decoding of stabilizer codes under radiation-induced correlated noise (2026), arXiv:2603.18231.
[45] J. L. Orrell and B. Loer, Sensor-assisted fault mitigation in quantum computation, Phys. Rev. Appl. **16**, 024025 (2021).
[46] C. Gidney, M. Newman, P. Brooks, and C. Jones, Yoked surface codes, Nat. Commun. **16**, 4498 (2025).
[47] A. Hutter, J. R. Wootton, and D. Loss, Efficient Markov chain Monte Carlo algorithm for the surface code, Phys. Rev. A **89**, 022326 (2014).
[48] H. Bombín, M. Pant, S. Roberts, and K. I. Seetharam, Fault-tolerant postselection for low-overhead magic state preparation, PRX Quantum **5**, 010302 (2024).
[49] N. Meister, C. A. Pattison, and J. Preskill, Efficient soft-output decoders for the surface code (2024), arXiv:2405.07433.
[50] S. C. Smith, B. J. Brown, and S. D. Bartlett, Mitigating errors in logical qubits, Commun. Phys. **7**, 386 (2024).
[51] P. Magnard, S. Storz, P. Kurpiers, J. Schär, *et al.*, Microwave quantum link between superconducting circuits housed in spatially separated cryogenic systems, Phys. Rev. Lett. **125**, 260502 (2020).
[52] J. Ramette, J. Sinclair, N. P. Breuckmann, and V. Vuletić, Fault-tolerant connection of error-corrected qubits with noisy links, npj Quantum Inf. **10**, 58 (2024).
[53] S. J. S. Tan, C. A. Pattison, M. McEwen, and J. Preskill, Resilience of the surface code to error bursts, Phys. Rev. A **113**, 022450 (2026).
[54] Y. Suzuki, T. Sugiyama, T. Arai, W. Liao, K. Inoue, and T. Tanimoto, Q3DE: A fault-tolerant quantum computer architecture for multi-bit burst errors by cosmic rays, in *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)* (2022) pp. 1110–1125.
[55] A. Siegel, A. Strikis, T. Flatters, and S. Benjamin, Adaptive surface code for quantum error correction in the presence of temporary or permanent defects, Quantum **7**, 1065 (2023).
[56] E. Yelton, C. P. Larson, K. Dodge, K. Okubo, and B. L. T. Plourde, Correlated quasiparticle poisoning from phonon-only events in superconducting qubits, Phys. Rev. Lett. **135**, 123601 (2025).
[57] E. Yelton, C. P. Larson, V. Iaia, K. Dodge, *et al.*, Modeling phonon-mediated quasiparticle poisoning in superconducting qubit arrays, Phys. Rev. B **110**, 024519 (2024).
[58] M. H. Kelsey, R. Agnese, Y. F. Alam, I. Ataee Langroudy, *et al.*, G4CMP: Condensed matter physics simulation using the Geant4 toolkit, Nucl. Instrum. Methods Phys. Res. A **1055**, 168473 (2023).