Prove the Camera, Not the Cloud: Anonymous, Hardware-Anchored Photo Provenance

Jamie Newton
Apertrue
jamie@apertrue.photo

***Abstract*—Industry provenance signs a manifest at capture with a device or service key: a signing story, not a sensor story, with linkable per-photo certificates. We take the sensor road and remove its two blockers. PRNU, the sensor's physical fingerprint, leaks from every photo and cannot be reissued. We register a keyed BioHash template of the stability-selected PRNU bits, never the bits themselves, and match it inside a 170k-gate threshold circuit, 4.4 s on-phone (an iPhone 13 Pro Max, the system platform): an estimate from public photos cannot form the template (impostor template BER 0.499 against $\tau = 0.475$; genuine 0.069), and a leaked template is retired by rotating the device-resident key. Forgery resistance and revocability, for a biometric that cannot change. A 415k-gate delegation circuit reduces platform contact to one Apple attestation per device, ever; the capture path contacts no server, and verification runs entirely in the browser: an anonymous "genuine capture from a registered camera" verdict with nothing to log. We anchor on PRNU having measured and rejected the leading alternative: dark-current challenge-response is a unique, stable fingerprint on clamp-free sensors yet *not* possession-hard—a per-pixel Meyer–Neldel proportionality collapses the thermal challenge to a global scale knob that a handful of the victim's dark frames reproduce.**

I. Introduction

A. The problem: provenance without surveillance

As synthetic imagery becomes indistinguishable from photographs, "was this taken by a real camera?" is turning into infrastructure. The industry answer, C2PA [1], signs a manifest at capture with a device or service key. That establishes origin, but at two costs. First, it is a *signing* story, not a *sensor* story: a compromised signing key, or a service authorised to sign on a user's behalf, can produce valid provenance for an image no physical camera captured. Second, and less discussed, per-photo signatures are *linkable*: an X.509 certificate is, in Paquin's phrase, "inescapably linkable" [2]. The strongest deployed design shows how far the signing story can go and where it stops. Pixel 10 signs with Titan M2-resident keys under a unique certificate per photo, its camera app certified at C2PA Assurance Level 2, so verifiers cannot link a user's photos to each other [3]. But the certificates are provisioned from Google's CA while the device is online; the issuer's view of who obtained which certificate — and, when connected, of per-capture timestamp requests [4] — is bounded by a published no-logging *policy* rather than by construction [3]; and, as in any trust-list ecosystem, which signers are accepted is a governance question separate from the cryptography [5]. A provenance layer that tells the world which photos a journalist, a dissident, or an ordinary person took (or whose only privacy guarantee is its operator's promise) addresses one risk while introducing another.

We see the two as complementary rather than competing: the signing ecosystem supplies transport, tooling, and adoption, and a sensor-anchored proof can travel inside a C2PA manifest's assertion store [1] as an additional assertion, giving that ecosystem a physics-rooted, CA-free signal it does not otherwise carry. Our contribution is that anchor.

We take the position that a provenance system must satisfy two properties at once, and that existing systems satisfy at most one:

- **Hardware-anchored.** The claim should rest on something physical about the capturing device, not solely on possession of a key. Photo-Response Non-Uniformity (PRNU) [6] (the sensor's manufacturing fingerprint, present in every image and rooted in the physical silicon rather than issued by any authority) is the natural anchor, complemented by hardware attestation for the app-integrity half of the claim. It is also the *measured* choice: we characterise the leading alternative anchor, dark-current thermal challenge-response, and show (§V-E) that even on the sensors where its signal survives it is not possession-hard, so PRNU is the candidate that survives.
- **Privacy-preserving by construction.** Verification should reveal nothing beyond the single bit "genuine capture from a registered device", should not link a person's photos to each other, and should not require sending anything to a server that could log it. Not "we promise not to log" — *nothing to log*.

Neither PRNU nor attestation gives this for free. PRNU *leaks by design*: the fingerprint is estimable from any target's public photos (the fingerprint-copy attack), so the obvious construction — register a hash of the raw fingerprint and later prove a photo matches it, as in the zero-knowledge PRNU-membership line (§III-Fb) — is *forgeable*: the estimated bits are a valid preimage. And it is a biometric that *cannot be changed*: a compromised template cannot be re-issued the way a key can, yet our own hot path exposes the raw bits in app memory on every proof. Hardware attestation, meanwhile, is the thing that calls home: Apple App Attest mints a fresh Apple-certified key per use, so using it per capture reintroduces exactly the linkability we set out to avoid.

B. This paper: prove the camera, not the cloud

We show that a commodity phone can produce a self-contained, publicly verifiable proof of "genuine capture from a registered camera" that is anonymous, forgery-resistant, revocable, and whose capture path sends nothing to any server (the platform is contacted once, at enrolment; §IV-F prices the remaining metadata surface), and that the whole thing runs at interactive latencies (seconds per proof) on hardware people already own (an iPhone 13 Pro Max as the system platform). The two hard sub-problems (making PRNU registration forgery-resistant *and* revocable, and getting hardware attestation without per-capture surveillance) are each solved by moving the trust into a zero-knowledge proof the device generates itself (Fig. 1). One trust statement belongs up front rather than five pages in: the RAW→bits extraction runs in the attested app, outside the circuits (§IV-D); everything downstream of extraction is proven, while extraction itself is app-asserted pending IVC-style extraction proofs. Concretely, our contributions are:

1) **Cancellable PRNU verification in zero-knowledge (§III).** We register a keyed BioHash [7] *template* of the stability-selected sensor bits (bit selection and the two-tier design are the companion paper's contributions [8], consumed here) rather than the bits themselves, so that an adversary who estimates the raw fingerprint from public photos cannot form the template, because the key never leaves the device (forgery-resistance), and a leaked template is retired by rotating the key (revocability). The template is matched inside the threshold membership circuit (170k gates, 4.4 s on-phone), both tiers hardware-verified against *unchanged* verification keys. To our knowledge this is the first cancellable-template treatment of PRNU and the first scheme to compute a cancellable-template match *inside* a SNARK; the cancellable-biometrics literature's usual weakness, transform-domain leakage (ARM, inversion), is removed here because the template is never released and the key is never transmitted (§III-D, §III-F).
2) **The delegation pattern: one Apple contact per device, ever (§II).** A network attestation at enrolment authorises per-capture *local* assertions (`generateAssertion`, no network), proven in a 415k-gate circuit stating "this capture's binding message was authorised by a key a genuine copy of this app minted on genuine Apple hardware" — with the certificate, the credential key and the enrolment challenge all private, and nothing linkable across captures. This retrofits Direct Anonymous Attestation's goal [9] onto Apple's non-DAA attestation.
3) **On-phone SNARK-circuit App Attest in seconds (§II).** A 322k-gate Noir [10] / UltraHonk [11] circuit proves a private Apple-issued credential certificate (a P-384 issuance signature over $\mathrm{SHA256(TBS)}$, the nonce extension, keyID and appID) in 8.5 s on-device, verifying in 2.0 s. The base idea (App Attest in zero-knowledge) is not ours; we cite a00012025/zk-app-attest (2024, RISC Zero zkVM, server-proved) [12]. But the on-phone SNARK, the on-device latency, and the provenance application are.
4) **A complete verifiable-capture pipeline with tiered trust labels (§IV, §V).** RAW→HEIC one-time-key binding (unlinkable), drand beacon freshness [13] (BLS-verified [14] against a pinned chain key), PDQ soft binding (§IV-B), the per-capture zero-knowledge hardware anchor, and off-device recursive aggregation of the sensor and anchor proofs into one. The verifier surfaces tiered trust labels rather than a single opaque verdict.
5) **No-logging as a structural property (§V).** Verification can run entirely in the browser (WASM barretenberg), so the proof bundle never leaves the page: the server is asked only for public data (verification keys, pinned circuit hashes, registry roots). Absence of *content* logging cannot be proven in zero-knowledge; we make it structural instead, and price the remaining metadata surface explicitly (§IV-F).
6) **The measured physics boundary that justifies the PRNU anchor (§V-D, §V-E).** A characterisation of the leading alternative anchor, dark-current thermal challenge-response: on the phone it is suppressed 3–4 orders of magnitude by on-die calibration, and on the clamp-free sensor class where it survives (two same-model consumer DSLRs) it is a unique, stable fingerprint that still fails possession-hardness: a per-pixel Meyer–Neldel proportionality collapses the challenge space to a global scale knob reproducible from a handful of the victim's dark frames. This is the measurement behind the anchor choice in §I-A, and it rescopes thermal-PUF proposals to sensors without on-die suppression.

C. Measurement status and scope

Every *system* number in this paper is measured on an iPhone 13 Pro Max against a Barretenberg 5.0 nightly toolchain; the anchor-boundary study (§V-D, §V-E) and the DSLR two-body import-audit demonstration (§V-C) are measured on that device and on two same-model Canon EOS 1100D bodies, and the impostor demonstration below uses a second iPhone. The circuits, the app, and the verifier are implemented, and the two cancellable tiers and the delegation anchor were cross-verified on desktop `bb` against unmodified verification keys. Generality is claimed at that scope: this paper demonstrates the design end-to-end on one platform and one enrolled sensor identity; population-level false-accept statistics come from the companion's 15-camera corpus [8, §6.3], which §V-H carries through the keyed transform: zero template-domain accepts under the companion's pairing (30 000 trials) and under exhaustive all-pairs testing everywhere except one same-model telephoto pair, which crosses the bound and is reported as the measured limit of same-model margins (§III-C, §V-H). Single-device results here should be read accordingly. Device-uniqueness / Sybil resistance is *out of scope* (§III-E, the decentralised-identity trilemma): the claim is forgery-resistance and revocability, not one-device-one-person. The

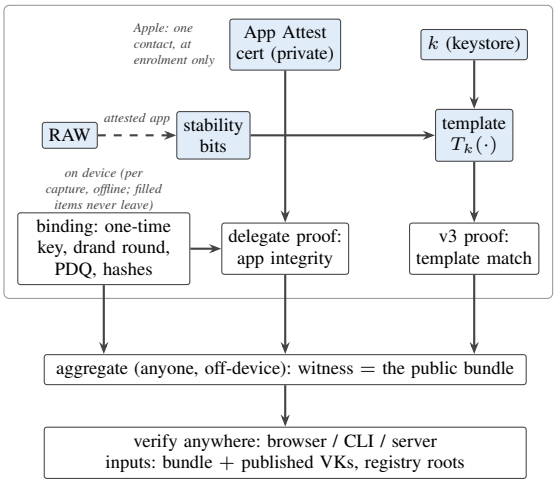

Fig. 1. The capture pipeline and its trust boundaries. Filled boxes are private and never leave the device; open boxes are public. Solid arrows are proven (in-circuit or by signature); the dashed arrow is the one app-asserted step (§IV-D). Aggregation is off-device because its witness is exactly the published bundle (§II-E).

second-device impostor test is now *demonstrated* on hardware (2026-07-10): a second iPhone (an iPhone 11 Pro, iOS 17) photographing ordinary scenes against the enrolled device's registry is rejected on all eight captures. Bind-tier Hamming clustered at chance (mean $\approx$16 344 of 32 704, the nearest 694 bits above the $\tau \cdot N = 15\,534$ bound) $\rightarrow$ "NOT THIS SENSOR". The membership circuit *refuses* to prove (the constraint $\mathrm{popcount} < \tau \cdot N$ is unsatisfiable) and the same device, re-enrolled as its own identity, verifies genuine — so the refusal is the sensor, not a broken prover. This is the false-accept side end-to-end, complementing the 15-camera corpus ([8, §6.3]: zero false accepts over 300 trials, and its template-domain extension, §V-H); the full eight-capture breakdown, the circuit refusal, and the night-proof control are in §V-C.

II. zkAppAttest: proving Apple's attestation without Apple watching

A. The problem: hardware anchors vs. unlinkability

A sensor-physics proof establishes that *some enrolled camera* produced a RAW capture, but says nothing about the published derivative (the HEIC the photographer actually shares). Binding the two requires software the verifier can trust; otherwise an enrolled adversary proves their own RAW and attaches an arbitrary image (the *laundering attack*, §IV-B). Platform attestation (Apple App Attest) supplies exactly this "genuine app on genuine hardware" fact, but at a privacy price that undermines the system's anonymity goal: every attestation is a fresh Apple-issued certificate minted through Apple's servers, giving the platform vendor a per-capture event stream. A provenance system that promises unlinkable photos while calling home on every shutter press is incoherent.

Direct Anonymous Attestation [9] solves this in the abstract, but Apple's attestation is not DAA and cannot be changed from outside. Our contribution is retrofitting DAA's *effect* onto App Attest as deployed: the platform contact happens **once per device, at enrolment**, and every subsequent capture carries a zero-knowledge proof that a *locally minted* assertion chains back to that single attestation, with the certificate, the credential key, and the enrolment challenge all private.

B. What App Attest actually signs (reverse-engineered facts)

Apple's documentation [15] specifies verification recipes but not byte-level signing semantics. Building in-circuit verifiers forced exactness; two facts cost us real debugging time against live Secure Enclave artifacts and appear to be undocumented:

F1 (attestation issuance). The credential certificate (credCert) is signed `ecdsa-with-SHA256` by the App Attest CA 1 intermediate, whose key is P-384. One P-384 scalar-multiplication chain therefore suffices for in-circuit issuance verification when CA-1's public key is pinned as a circuit constant; the CA-1$\rightarrow$root link is a static, publicly checkable fact that belongs outside the circuit. The certificate's nonce extension (OID 1.2.840.113635.100.8.2) embeds $\mathrm{SHA256}(\mathrm{authData} \,\|\, \mathrm{SHA256}(\mathrm{challenge}))$ at a DER-predictable offset within the signed TBS bytes.

F2 (assertion signing). For assertions, Apple signs the nonce *as a message*, not as a prehash: the ECDSA digest is $\mathrm{SHA256}(\mathrm{SHA256}(\mathrm{authenticatorData} \,\|\, \mathrm{clientDataHash}))$, a double hash. Our circuit was one SHA short until a real device assertion refused to verify. Additionally, Apple's DER signatures are not low-S normalised, while barretenberg's secp256r1 gadget requires low-S; on-device normalisation is mandatory.

Both facts were validated against hardware-generated artifacts and are encoded in our tooling; we believe this byte-level table (§II-F) is independently useful to anyone building on App Attest.

C. The attestation circuit (zk_attest, 322k gates)

Public inputs: challenge (66 bytes as field elements), aaguid (16 bytes; dev/production is verifier policy). Private witness: the credCert TBS bytes (padded to 1024), the CA-1 signature (r, s), authData (164 bytes), and the DER offsets of the nonce extension and the credential public key.

The circuit enforces: (1) ECDSA-P384 verification of $\mathrm{SHA256}(\mathrm{TBS})$ under the pinned CA-1 key (Apple issued this certificate); (2) the nonce equation: `TBS[nonce_off..]` equals $\mathrm{SHA256}(\mathrm{authData} \,\|\, \mathrm{SHA256}(\mathrm{challenge}))$, behind an asserted 19-byte DER context (OID + wrapping), so the certificate attests THIS challenge; (3) key binding: SHA256 of the SPKI point at the witnessed offset equals authData's credentialId (offsets are witness values guarded by structural DER asserts, following but correcting the inverted-assert idiom of [16]); (4) authData layout: rpIdHash equals the pinned $\mathrm{SHA256}(\mathrm{appID})$, counter is zero, aaguid is exposed. Buffer parsing beyond these anchored offsets stays off-circuit: the nonce equation cryptographically ties every byte that matters.

Measured: 322k gates; 8.5 s proving / 2.0 s verification on an iPhone 13 Pro Max (UltraHonk, poseidon2 transcript); proofs cross-verify against desktop barretenberg with byte-identical verification keys.

D. Delegation: one Apple contact, ever (zk_attest_delegate, 415k gates)

App Attest's `generateAssertion` is fully local: the Secure Enclave signs without network. The delegation circuit extends §II-C's statement with: (5) an assertion by the *same credential key* (extracted in-circuit from the private certificate) over the capture's binding message, which is the circuit's public input. The composite statement:

> "This capture's binding message was authorised by a key that a genuine copy of this app, on genuine Apple hardware, minted — attested by an Apple-issued certificate I hold but do not show."

The enrolment challenge moves to the private witness (it was public at registration; per-capture proofs need not reveal it). Verifier-side obligations are minimal and static: CA-1 chains to Apple's root, and aaguid policy. Nothing in the proof is linkable across captures: the binding message is per-capture, the certificate never appears, and the one-time binding keys (§III) are fresh per shutter.

Cost of privacy: Apple's per-capture attestation is a network round-trip plus CBOR validation; ours is a 415k-gate proof, measured 9 965 ms to prove and 2 596 ms to verify on-device (iPhone 13 Pro Max, 2026-07-09). The deployed default is the ZK path (one Apple contact at enrolment, local assertions per capture); per-capture Apple attestation remains only as the fallback when the zk anchor is unavailable (no attested enrolment or no beacon), and the verifier labels which anchor a bundle carries.

E. Off-device aggregation is free of trust cost

A recursive circuit (1.48M gates) verifies the delegation proof and the sensor-membership proof in one UltraHonk proof, rebuilding both inner public-input sets from its own public inputs and exposing the inner verification-key hashes for pinning. *Where proving must happen is determined by witness secrecy, not system aesthetics.* The aggregator's witness (two finished proofs, their verification keys, their public values) is exactly the content of the published bundle. Aggregation on a server (16.3 s, 1.9 GB on an M-series laptop) therefore reveals nothing the bundle didn't, while the phone proves only the two witness-bearing circuits. Each recursive ZK verification costs ≈740k gates in our stack, an order of magnitude above folklore estimates from earlier proving systems.

F. The Apple facts table

Table I collects the byte-level facts established in §II-B.

III. Cancellable PRNU: revoking the unrevokable biometric

A. The problem: a biometric you cannot change and cannot help publishing

PRNU is a hardware biometric with two properties no other deployed biometric combines, and both are bad news for a naive enrolment system:

1) **It cannot be rotated.** The fingerprint is silicon non-uniformity; a compromised template cannot be re-issued the way a password or a certificate can. The only physical "rotation" is buying a new camera, unacceptable for exactly the users a provenance system serves (journalists, fleets, courts).
2) **It leaks by design.** Every published photo carries a trace of the fingerprint; given enough of a target's public photos an adversary can *estimate* the raw fingerprint (the fingerprint-copy attack, Goljan et al. [17]). This is not an implementation defect — it is what PRNU *is*.

Property 2 breaks raw-bit enrolment outright. Our membership proof takes fresh bits as a private input and proves them close (Hamming $< \tau \cdot N$) to an enrolled commitment; soundness against forgery therefore rests entirely on the *secrecy of a valid preimage* to that commitment. If the enrolled leaf commits to raw sensor bits, an estimated fingerprint **is** a valid preimage: the adversary proves "genuine, from the victim's camera" for an image the victim never took. The proof system is sound; the enrolment statement is wrong.

Property 1 breaks recovery. Even without estimation attacks, our own hot path guarantees the raw bits transit app memory on every prove — a standing exposure by construction. A scheme with no revocation story treats every future memory-disclosure bug as a permanent, unfixable identity compromise.

Both problems have a joint, standard-shaped answer: cancellable biometric templates [18]. But the standard constructions carry documented attacks that our architecture happens to close (§III-D).

B. Construction: keyed BioHash under a device-resident key

We register a keyed template of the fingerprint, never the fingerprint:

$$\begin{aligned}\text{template}[m] &= \operatorname{sign}\big(\textstyle\sum_s w[m][s] \cdot b[\text{taps}[m][s]]\big) \ \text{ over } \pm 1,\\ &= \big(\text{agreements}(w[m],\, b@\text{taps}[m]) > S/2\big), \ \ S \text{ odd}\end{aligned}$$

with $S = 15$ taps per template bit over the $Z = 32\,704$ stability-selected raw bits b. Tap indices and weight signs are derived by a PRF (HMAC-SHA256) from a 256-bit transform key k that is generated on-device at enrolment and held in the OS keychain, access-controlled to this device and only while unlocked (never synced or backed up), and is **never published, never registered, and never part of any proof's public data**. Stated exactly: k is not enclave-resident (the Secure Enclave holds only asymmetric keys), and the PRF evaluation brings its bytes and the derived taps into app memory at each transform — the exposure class of §III-A's raw bits, with the same mitigation, rotation. One key per registry; distinct registries get independent templates.

One transform covers both proof tiers, with only M varying:

- **identity tier** (exact-key circuits v1/v2; the companion paper's *exact-key tier* [8]): $M = 1\,024$ template bits, of which 896 feed the $r = 7$ repetition fuzzy extractor;
- **bind tier** (per-photo threshold circuit v3; the companion's *threshold tier*): $M = 32\,704 = Z$, so the template is *shape-identical to the raw input* and the deployed v3

TABLE I
THE APPLE FACTS TABLE.

Fact	Value	Discovered via
credCert signature	`ecdsa-with-SHA256` by P-384 CA-1 key	real attestation dissection
nonce extension	OID 1.2.840.113635.100.8.2, $\mathrm{SHA256}(\mathrm{authData} \,\|\|\, \mathrm{SHA256}(\mathrm{challenge}))$, fixed 19-byte DER prefix	ditto
authData (attestation)	rpIdHash(32) flags(1) counter(4)=0 aaguid(16) credIdLen(2)=32 credId(32)	Apple docs + verification
assertion authenticatorData	37 bytes: rpIdHash flags counter	real assertion
assertion ECDSA digest	$\mathrm{SHA256}(\mathrm{SHA256}(\mathrm{authData} \,\|\|\, \mathrm{clientDataHash}))$ — nonce signed as MESSAGE	refused single-hash circuit
assertion DER signatures	NOT low-S normalised (gadget requires low-S)	refused unnormalised witness
attestation CBOR	{fmt, attStmt:{x5c:[cred,ca1], receipt}, authData}	dissection

circuit (same N, same τ, same verification key) consumes template bits without any modification. Verification is unchanged; only what is *enrolled* changes.

Enrolment mints the template registries once (an on-device helper-circuit run at enrolment time); each capture then transforms its fresh bits through the same PRF-derived taps (milliseconds, off-circuit, native) and proves the unchanged tier statement in the template domain.

What the attacker equation becomes. A forger with a perfect estimate of the victim's raw bits must still produce template bits that match the enrolled template within τ. Without k, taps and weights are unknown; the attacker's best template is uncorrelated with the enrolled one: measured impostor template BER 0.499 against $\tau = 0.475$ (§III-C). Fingerprint estimation from public photos, the attack that kills raw-bit enrolment, buys nothing.

What rotation becomes. Discard k (a keychain delete), re-enrol: the new template is statistically independent of the old (measured same-camera-two-keys correlation $\approx$0, $|\mathrm{corr}| < 0.02$), so a leaked template is dead the moment the key rotates — burn-notice semantics for a biometric that physically cannot change. The same mechanism gives cross-registry compartmentalisation: distinct k per registry means a template stolen from one registry is worthless in every other, and the two registrations are unlinkable — measured at corpus scale as $D_{\mathrm{sys}} = 0.0064$ over 1 500 templates (§III-C).

C. Measured: the transform preserves both tiers' margins

All numbers from real iPhone 13 Pro Max enrolment bits (the same bits the deployed system uses), `tools/cancelable_e1.py` and `tools/cancelable_bind_margin.py`; see Table II.

The transform amplifies genuine noise roughly 3× (majority-of-15 taps over noisy ± 1 bits), but 0.069 remains far below the 0.475 acceptance bound while impostors stay at chance — separation is fully preserved. The identity-tier template BER ($M = 1\,024$, recomputed from the shipped transform, 2026-07) is 0.076 per capture, well under the 0.475 threshold, but higher than the raw key tier's 0.006. At BER 0.076 the $r = 7$ repetition decoder ($\leq$3 flips per 7-bit block, 128 blocks) fails with probability ≈ 0.117 per capture, versus $\approx 10^{-5}$ at 0.006. Single-ordinary-capture exact-key recovery is marginal by the decoder's own numbers; the identity/exact-key tier is therefore intended for controlled, enrolment-grade captures (locked ISO, bright flat), consistent with §V's tiering. (The E1 pilot reported 0.035 for a keyed transform over a more-stable input set; the production $M = 1\,024$ identity template over the bind bits is the 0.076 above.) [1]

TABLE II
RAW-DOMAIN VS. TEMPLATE-DOMAIN MARGINS.

Quantity	Raw domain	Template domain
genuine key-tier BER ($M = 1\,024$)	0.006	0.076
genuine bind-tier BER ($M = 32\,704$)	0.026	0.069
impostor bind-tier BER	0.500	0.499
bound (τ/N)	0.475	0.475 (unchanged)
same camera, two keys: template correlation	—	$\approx$0 ($|\mathrm{corr}| < 0.02$)
impostor bind-tier BER, 15-camera corpus, per-photo all-pairs mean (§V-H)	0.4998	0.4999
unlinkability D_{sys}, 15 cameras × 100 keys (§V-H)	—	0.0064

Corpus-scale template statistics. The same quantities now exist at population scale: §V-H pushes the companion's 15-camera corpus through the shipped transform (production KDF, parity-checked bit-for-bit against the reference implementation) under 100 independent keys per enrolled camera. Per-photo impostor template BER has mean 0.4999 (sd 0.0037). Under the companion's pairing, zero of 30 000 trials accept; exhaustive all-pairs testing (42 000 trials) yields 7 accepts, every one from a single same-model telephoto pair, with the remaining 208 ordered pairs at zero accepts across 41 600 trials (the measured same-model boundary, §V-H). That crossing concerns the threshold leg alone — the physical distinctiveness the raw design inherits, probed with the enrolled camera's own key. The key leg, the forgery claim, stays at chance everywhere measured: an attacker holding a perfect estimate of the raw bits but not k (same camera, independent key) gets no closer than 376 bits outside the bound in 74 250

[1] An earlier design memo concluded the bind tier could not survive the transform (genuine BER ≈ 0.42, "thin margin"); the number came from a broken measurement script and reversed on real bits. It stays in the paper trail because the feasibility of cancellable-everything was decided by measurement, not argument.

TABLE III
ON-DEVICE COSTS OF THE CANCELLABLE TIERS.

Proof	Prove	Verify (on-device)
v3 cancellable (per-photo threshold, 170k gates)	4 305 ms	1 281 ms
v2 cancellable (identity tier, 28.7k gates)	613 ms	182 ms

comparisons. Unlinkability, measured in the Gomez-Barrero framework [19] over 1 500 templates (15 cameras $\times$ 100 keys): mated and non-mated cross-template score distributions coincide, $D_{\text{sys}} = 0.0064$, mated $|\text{corr}|$ mean 0.0044 and max 0.027 against a non-mated max of 0.027, upgrading the single-camera two-key row of Table II to a population statement.

Where τ comes from. The threshold predates every corpus number in this paper. The project record shows $\tau \cdot N = 15\,534$ ($= 0.475N$) frozen in the specification on 2026-07-06, before any corpus measurement: it is the point where the idealised independent-bit binomial impostor tail at $N = 32\,704$ reaches 7.7×10^{-20} (9.0 standard deviations below chance), chosen with false-reject headroom against the then-measured worst genuine scene BER of 0.446, and it replaced a demo threshold of $0.4965N$. The deployed circuit adopted the frozen constant on 2026-07-08, closing a demo/production threshold gap (§V-C, refusal item 5). τ was therefore derived analytically, not tuned on the corpus the previous paragraph evaluates.

Correlation-aware margins. The 7.7×10^{-20} figure prices per-capture noise only, and the corpus shows that this is the smaller part of the story. Per-capture noise is indeed independent-bit-like: the eight second-device captures of §V-C have a Hamming standard deviation of 89 bits against 90.4 predicted. Pair-level structure dominates the population spread: across the corpus the per-photo impostor BER has sd 0.0037 (effective $N \approx 18\,300$ rather than 32 704), and enrolment-grade fingerprint estimates widen the pair-mean sd to 0.0077 (effective $N \approx 4\,300$), so $\tau = 0.475$ sits 6.8 (per-photo) or 3.1 (enrolment-grade) standard deviations below the impostor mean, not 9.0. The tail is not Gaussian either: every observed crossing is one same-model pair, so false-accept claims below corpus scale are stated empirically in §V-H rather than extrapolated from a binomial model. The night operating point inherits the same discipline: the burst-rescue margins of §V-G ($+116$ bits at two frames, against a single-frame genuine BER of 0.483) are of the order of one capture-noise standard deviation, which is why the production burst engages adaptively instead of assuming idealised headroom.

On-device costs, measured July 9, iPhone 13 Pro Max, both proofs cross-verified on a desktop verifier against the **unchanged** production verification keys, are given in Table III.

Per-photo cost is indistinguishable from the raw-domain figures (§V table): the transform itself is off-circuit and native.

The in-circuit enrolment proof. Enrolment-time consistency ("the registered v2 leaf really is T_k applied to *a* fingerprint witness the enroller holds") is provable in one one-time circuit (`cancelable_enroll`): 270 825 gates at production parameters ($M = 1\,024$, $S = 15$, $Z = 32\,704$), 3.64 s / 735 MB on desktop, proves and verifies on production-size goldens. Since the 2026-07-09 rewire its single public output *is* the registered v2 leaf (the registry-helper leaf computation runs in-circuit; a cross-circuit golden checks both produce the identical field), so the proof speaks about the registered value, not a free-floating template commitment. Scope stated exactly: it proves the v2 (identity-tier) leaf is T_k of a held witness; it does not yet link the bind-tier template to the same witness (future work). A gate-cost note of independent interest to Noir authors: the transform's natural signed multiply-accumulate form costs $\sim$45 gates per tap (overflow range checks on signed arithmetic), while the equivalent agreement-count form (S odd) costs 8.65 gates per tap: a $\sim 5.2\times$ reduction on the tap primitive itself, from a boolean reformulation with bit-identical outputs. At the whole-circuit level the win dilutes to $3.1\times$ (210k $\rightarrow$ 68k gates at the $M = 256$ reduced parameters) because the fixed per-entry ROM-init cost ($\sim$4 gates/entry) is unchanged by the reformulation.

D. Security: ZK + a device-resident key close the classic cancellable-biometric attacks

Cancellable binary-template schemes (the cancellable-iris / BioHash lineage this construction belongs to) have a well-documented attack literature. Every documented attack requires at least one of two preconditions, and our architecture removes both *by construction* rather than by parameter tuning:

- **Attack via Record Multiplicity (ARM)** reconstructs raw bits by correlating multiple transformed templates — *given the templates and transform parameters*. Here templates are never exposed: they exist as hash commitments and private proof witnesses only.
- **Invertibility / known-key attacks** invert a *given* transformed template. No template is ever given.
- **Stolen-token attacks** assume the secret transform parameters leak — the scenario under which BioHash-family performance collapses [20]. k is device-resident and participates in no protocol message.

Pattern: the attacks need (a) a stored/transmitted template or (b) the transform secret. Zero-knowledge matching hides (a); the device-held key hides (b). This is a contribution, not a reuse: the standard weakness of cancellable biometrics, transform-domain leakage, is removed by *where* matching happens (in a SNARK) and *where* the key lives (in the device keystore), not by a stronger transform.

Residual risk is total device compromise (k **and** raw bits exfiltrated together), which is irreducible for any cancellable scheme; the response is the burn-notice rotation above. Cross-registry leakage is additionally bounded by the measured near-zero inter-template correlation ($|\text{corr}| < 0.02$ single-device; $D_{\text{sys}} = 0.0064$ at corpus scale, §III-C).

Complementarity: the zkAppAttest anchor (§II) and the cancellable template defend different claims. The anchor proves *app integrity*: that a genuine app performed the capture-and-bind protocol; the template defends the *sensor statement itself* against preimage forgery by an adversary who never touches the device. Neither subsumes the other.

E. Scope: what the template does not claim

Early design iterations aimed the keyed template at a third goal: Sybil-resistant registration via a deterministic per-sensor nullifier ("one sensor, one registry entry"). We de-scoped it, for reasons worth stating in a security paper:

1) Our own adversarial measurement killed the cheap construction: a malicious re-enroller corrupting 30–50 % of the device-ID core evades the nullifier linkage while still passing template verification (measured on real bits; `tools/op1_evasion_sweep.py`). Honest-enrolment duplicate detection survives; adversarial uniqueness does not.
2) Platform attestation does not rescue it: App Attest is unlinkable by design (many keys per device, reset on reinstall), so neither mechanism provides privacy-preserving one-device-one-identity.
3) This is the Decentralized Identity Trilemma (uniqueness, privacy, self-sovereignty) [21] rather than an engineering gap: Sybil resistance bottoms out in an external scarce resource, and even a perfect sensor-ID is "one per purchasable sensor", never "one per person".

Crucially, the core provenance claim loses nothing: registration requires proof-of-possession of the bits, so a Sybil can only register sensors they physically hold — which still yields only genuine photos from genuine sensors. The nullifier we ship retains what survives measurement: burn-notice rotation and verbatim-duplicate refusal. Even *honest* cross-enrolment duplicate detection via sketch recovery is measured dead on hardware (2026-07-09): two independent enrolments of the same sensor published sketches linking at residual 0.451 (chance) because the core is ordered by each enrolment's own stability ranking, so bit positions never align across enrolments. (A registry-defined canonical pixel lattice would fix alignment without publishing device-specific selection, itself an unkeyed fingerprint, but its viability rests on an unmeasured parameter, unselected-pixel stability; recorded as a fallback design, not built.) Duplicate and Sybil policy therefore layer on the enrolment attestation, the platform layer that natively provides device identity. The paper claims cancellable PRNU as **forgery-resistance + revocability**, not as Sybil resistance or duplicate detection.

F. Positioning against prior art

We looked adversarially for work that would pre-empt each claim below. Three neighbouring lines of work exist; none occupies our exact point.

(a) Privacy-preserving PRNU attribution hides the fingerprint from the verifier but does not make it *cancellable*. This is one group's line of work across two versions, cited once at the definitive version: the 2021 preprint SSS-PRNU (Jena, Singh & Mohanty, arXiv:2106.07029, "Privacy-Preserving PRNU Based Camera Attribution using Shamir Secret Sharing") was extended into the 2024 peer-reviewed journal paper PP-PRNU (Jena, Singh, Mohanty & Das, *Computing* 106:3309–3333) [22], which names the arXiv preprint as its preliminary version and adds a co-author over the same Shamir-Secret-Sharing core. Its architecture is worth stating precisely, because the contrast with ours is on every axis. PP-PRNU is a *forensic* scheme with three trusted entities (a fingerprint source that extracts the fingerprint inside a TrustZone enclave, a *match-maker* that the paper names as "the judge or law-enforcement authority" and which holds the query image **in plaintext**, and a match-maker server) plus n honest-but-curious cloud servers that hold $(2l-1, n)$ Shamir shares of the fingerprint and PRNU noise and compute the Pearson correlation on shares (SSS being homomorphic to addition and a single multiplication; the final square-root/division is done in the clear at the trusted match-maker server). It is "perfectly secure" (their Thm 1) but *conditionally*: the fingerprint is recoverable if l of the n servers collude (their Thm 2). Its privacy goal is to hide the *photographer's* identity from the servers in, e.g., whistle-blower cases, not to give the prover a portable proof. L-PRNU [23] is a separate low-complexity member of the same family. Against this: our scheme has **no trusted party** (a single prover emits a self-contained, publicly verifiable proof: no server holds shares, no honest-but-curious match-maker sees the query), makes **no non-collusion assumption**, never reveals the fingerprint or the query to anyone (ZK-hidden, not merely split), is **anonymous** rather than a plaintext-query forensic match, and — the point of this section — is **revocable**: a leaked template is rotated by discarding the device key, whereas a compromised PP-PRNU share configuration is not re-issuable. A separate neighbour on the PRNU axis is Valsesia et al.'s PRNU-based PUF [24], which compresses the fingerprint with adaptive random projections and extracts a key via a polar-code fuzzy extractor. The random projection is mechanically adjacent to a BioHash, but there it serves compression and key extraction: the scheme claims neither revocability nor unlinkability, and does not address template inversion or record multiplicity — the axes that define a cancellable template.

(b) Zero-knowledge PRNU membership is the closest prior art and is the baseline we improve. The PhotoNecf line (Web Photo Source Identification, WWW 2023, arXiv:2302.09228) [25] registers a hash digest $h = \mathrm{SHA256}(K)$ of the raw fingerprint K and later proves a photo derives from the registered camera. This is precisely the raw-bit registration our §III-A shows is *forgeable*: because PRNU is estimable from a target's public photos (fingerprint-copy, Goljan et al. [17]; see also the leakage analysis of Fernández-Menduiña and Pérez-González [26]), the estimated bits are a valid preimage to h. Our contribution is the cancellable keyed template $\mathrm{commit}(T_k(\mathrm{bits}))$ with k device-resident, which removes exactly this attack and, as a bonus, adds revocability. We cite PhotoNecf as the ZK-PRNU-membership baseline, not as a competitor we outperform on accuracy. On the image-claim side, PhotoProof [27] pioneered proof-carrying image transformations (edits provably permissible relative to an authentic original); our RAW→HEIC binding takes the cheaper attested-software route (§IV-B), trading transformation generality for on-phone cost.

(c) ZK biometric matching exists for *human* biometrics, but no scheme matches a cancellable template *inside* the proof. Guo et al. (zk-SNARK biometric identification, Secur. Commun. Netw. 2022) [28] and BioZero (arXiv:2409.17509, Pedersen commitments + Groth16) [29] prove possession/match of a committed biometric template inside a proof system, but the registered object is a commitment to the biometric itself, not a *revocable, keyed* transform of it: a leaked commitment is not rotatable. Where cancellable templates and ZK proofs do co-occur, the matching stays outside the circuit: Kothari et al. [30] pair a keyed random-matrix cancelable fingerprint template with a Groth16 threshold circuit, but the template-to-probe similarity matrix enters that circuit as a *public* input (the matching itself runs in plaintext off-circuit), and Mao et al. [31] commit cancelable multimodal templates yet compute the distance check outside the proof. None of these keeps the transformed template as a private witness matched in-circuit, and none addresses the transform-domain attacks (ARM, invertibility) that §III-D closes.

The cancellable-iris lineage is our direct ancestry, and the fit is close enough to be the paper's framing device. Iris recognition, like ours, matches a *binary code* under a Hamming/Jaccard threshold, and the cancellable-iris literature builds keyed transforms over that binary code that preserve the distance while satisfying the ISO/IEC 24745 requirements (irreversibility, unlinkability, renewability) [32]. The two schemes we build on: Rathgeb, Breitinger & Busch's alignment-free cancellable iris templates via adaptive Bloom filters [33] (2013, the foundational Bloom-filter iris work; see also *Computers & Security* 42:1–12, 2014, on mixing iris-codes with adaptive Bloom filters [34]), and Indexing-First-One (IFO) hashing (Lai et al., *Pattern Recognition* 64:105–117, 2017) [35], which maps the IrisCode to a distance-preserving hashed index that is non-invertible, unlinkable and renewable (0.54 % EER on CASIA-v3). Our keyed BioHash [7] over the PRNU stability code is the same move ported from a *human* biometric to a *sensor* one.

What we add on top of that lineage is exactly what it cannot do for itself. These iris schemes carry documented transform-domain weaknesses: ARM (Attack via Record Multiplicity) [36] and inversion attacks that operate on *stored/transmitted* templates [37]. Their whole security argument is spent bounding how much a released template leaks (hence the arms race of double-Bloom-filter, reverse-merge, and security-analysis papers). We sidestep that arms race: the template is never released (ZK-hidden, §III-D) and the key never leaves the device, so ARM/inversion have no input to work on. We also face a threat iris does not — the biometric is estimable from the subject's public photos (fingerprint-copy) — which is why forgery-resistance, not just privacy, is a first-class goal here.

Claims, qualified. Against this landscape we state, to our knowledge: (1) the first *cancellable-template* treatment of PRNU (revocability plus forgery-resistance for a hardware biometric), where the privacy-preserving PRNU family gives neither and the PRNU-PUF line's random projections serve key extraction, not template protection (a); (2) the first scheme to compute the cancellable-template match itself *inside* a SNARK (the transformed template is a private witness; nothing beyond the threshold verdict is revealed), where prior combinations either prove over commitments to the raw biometric or leave the matching outside the circuit (b/c); (3) the device-held-key property (no protocol message ever carries the template or the key), which is what closes the standard cancellable-biometric attack surface (§III-D). We frame (1)–(3) as "to our knowledge" pending a final librarian pass, per the project's rule to distrust a fast negative. The in-circuit `cancelable_enroll` consistency proof (§III-C) additionally makes the enrolment-time transform *verifiable* rather than assumed — a property none of (a)–(c) offers.

Open items (status as of submission draft): the enrolment consistency proof `cancelable_enroll` (270 825 gates, 3.64 s / 735 MB desktop) is now wired into the app's registration flow (2026-07-09): the app proves it once at enrolment and the verifier labels the registry entry "enrolment-consistency PROVEN" after checking the public output equals the registered cancellable v2 leaf (an invalid or mismatched proof refuses the registration; an absent one leaves the entry unlabelled). A one-time enrolment cost, not per-photo; the per-photo tiers are unaffected. Its on-phone prove time is now measured: 10.7–14.6 s across four enrolment runs (iPhone 11 Pro, iOS 17, 2026-07-09), just above the $\sim$10 s estimate; each value is recorded in its registry entry.

IV. Threat model & security argument

A. Claims, separated

The system makes two claims a verifier can confuse but the design must not: the **sensor claim** ("the RAW behind this proof came from an enrolled sensor") rests on physics (PRNU) and cryptography (the threshold circuit) alone; and the **image claim** ("the published HEIC is that RAW's derivative"), which no physics can carry, because the derivation happens in software. Every trust decision in the system flows from keeping this separation strict.

A third claim is deliberately absent. "Genuine capture from a registered camera" is literally true for a photograph of a screen or print displaying synthetic content: rephotography (the analogue hole) defeats any capture-authentication system at the semantic layer, because the capture event *is* genuine. The system authenticates that a registered sensor captured this light, never that the depicted scene is real; verifier labels are worded as capture claims, not veracity claims, and Table IV carries the scope row.

B. The laundering attack (why app integrity is load-bearing)

Adversary: a legitimately enrolled device owner. They photograph a wall, prove their own RAW (the sensor claim holds: it IS their sensor), and attach someone else's image as the "derivative". Nothing in the sensor claim resists this; the RAW$\rightarrow$HEIC binding is only as good as the software that computed the hashes at capture time. This is why the capture binding must be anchored in *app integrity* (Apple attestation or the zkAttest delegation proof), and why an unanchored

binding is labelled as exactly that ("sensor proof stands; the RAW→HEIC link is self-asserted") rather than silently passing. We adopted this position mid-design when the zero-third-party default briefly shipped without it; the verifier's tiered labels are the durable form of the lesson.

PDQ soft binding (the re-encode case). The exact RAW→HEIC hash binds only the byte-identical derivative, so the signed binding also carries a 256-bit perceptual hash of the derivative: a PDQ-shaped [38] implementation (same 16×16 DCT-and-median shape, own box downsample and cosine tables, no quality score) computed at capture and recomputed by the verifier from the received image with the identical C. Measured distances: byte-identical files 0/256, JPEG re-encodes 2–6/256, different images ∼130/256, against an acceptance threshold of 31 — so an honest re-encode of the captured derivative stays soft-bound to the signature while a substituted image does not. Two caveats: the hash is *not* reference-PDQ-compatible (cross-ecosystem PDQ database lookups will not match; anyone needing that interop can recompute reference PDQ from the image, which travels with the bundle), and swapping in reference PDQ later is binding-format-compatible (same 32-byte slot) but is a protocol version change, the same class as an extractor change.

C. The refusal matrix

What each layer independently refuses, all demonstrated on hardware, is given in Table IV.

D. Residual trust

1) **RAW→bits extraction is app-asserted.** The stability-bit extraction runs outside the circuit (in-circuit demosaic/denoise of a 12MP frame is infeasible today; we measured the adjacent problem, in-circuit resize, at 45–82 h native). The zkAttest anchor reduces this to "trust the attested app binary", the same root every attestation scheme bottoms out in; IVC-style extraction proofs are future work.
2) **Apple, once.** Enrolment trusts Apple's attestation root and ceremony. We remove Apple from the per-capture path, not from the TCB. A compromised Secure Enclave or forged attestation defeats the image claim (not the sensor claim).
3) **aaguid policy.** Development vs production attestation environments are distinguishable and exposed as a public input; verifiers choose.
4) **Registration is near-open in the MVP.** The nullifier gate (shipped 2026-07-09) refuses verbatim duplicates and gives rotation semantics, but measured on hardware it cannot catch an independent re-enrolment of the same sensor (same-sensor sketch linkage $0.451 \approx$ chance: the fingerprint is an authenticity primitive, not an identity primitive). Duplicate/Sybil policy layers on the enrolment attestation instead (one Apple contact, already in the flow; attest key recorded with the registration). Registration $\neq$ verification: an enrolled Sybil still cannot pass for anyone else's sensor.
5) **Verifier honesty.** Server-side verification could lie or log; both are addressed structurally, not by policy: anyone can verify the bundle (bb CLI, or the in-page WASM verifier whose only server requests are for public constants), so a lying verifier is detectable and a logging one learns nothing it wasn't sent.
6) **The sensor's owner is outside the sensor claim's scope.** The sensor claim authenticates a capture against any party who does not hold the sensor (the non-owner refusals demonstrated in §IV-C: the 15-camera corpus and the second-device rejection). Against the sensor's own owner, who holds the fingerprint and key and controls the noise model, we find no passive statistic that reliably separates a matched graft from a genuine capture, so owner-side integrity reduces to the capture-binding and accountability anchors §IV-B already makes load-bearing. A detector study is left to future work.

E. Stolen-device revocation

Burn-notice rotation (§III) requires holding the device; the remaining threat is an adversary holding the *unlocked* device (enclave, keys and enrolment included) producing genuine-looking proofs indefinitely. The countermeasure is registry-side and account-less: enrolment mints a 128-bit *recovery credential* (shown once; synced to the user's iCloud Keychain; the registry stores only its SHA-256 on every root of that identity). Presenting the credential at a public endpoint revokes those roots **at the current drand round**. Because every capture binding commits to a drand round, verification *time-cleaves*: proofs whose round precedes the revocation keep verifying (they were true when made, with an explicit "later revoked, provably predates" label); later captures — the thief's — are refused; and a proof with no provable capture time is refused once its root is revoked (fail-closed: the round is the only thing separating the owner's history from the thief's future). The freshness beacon thus does double duty (anti-backdating at capture and the revocation cleaving line), with its BLS signature verified against the pinned drand chain key, so neither side can forge its position in time.

F. Metadata and the anonymity set

The zero-knowledge layer hides everything inside the proofs; what remains is the metadata around them and the pixels themselves, and we price both explicitly rather than let "anonymous" overclaim.

1) **The beacon round discloses capture time.** Every capture binding commits to a drand round (§IV-E), so a published bundle reveals its capture time to the beacon's round granularity. That is the anti-backdating and revocation-cleaving mechanism doing its job, but it is also a correlation handle: an anonymous photo whose round matches a known event narrows who could have taken it. Mitigations are protocol work, not policy — committing to a coarsened round bucket, or proving round membership in a range in zero knowledge — both trading anti-backdating precision for temporal privacy.

TABLE IV
THE REFUSAL MATRIX.

Attack	Refused by	Demonstrated
Black/underexposed frame	v3 constraint unsatisfiability (no fingerprint to match)	on-device, production τ
Photo from a different sensor	τ threshold (impostor Hamming $\approx 50\,\%$ vs $47.5\,\%$ bound)	15-camera corpus: 0 false accepts / 300 trials [8, §6.3]; template-domain and all-pairs extension in Table IX (zero accepts outside one same-model telephoto pair). **2nd-device demo (2026-07-10):** a second iPhone rejected on all 8/8 captures (mean $16\,344$ vs $15\,534$ bound), v3 proof refused, own-identity control verifies (§V-C).
Foreign HEIC on own RAW (laundering)	binding signature + attestation/zkAttest anchor over the pair	live verifier labels
Replayed/backdated capture	drand round pinned in the signed binding; verifier checks round randomness + time window	round mismatch $\rightarrow$ refuse
Modified app / emulated device	App Attest at enrolment; per-capture local assertions chain to it in ZK	real cert chain in-circuit
Substituted inner circuit in aggregation	pinned inner VK hashes as public outputs	tampered pin $\rightarrow$ refuse
Proof/bundle tampering	UltraHonk verification (server AND in-browser)	4 tamper vectors $\times$ 2 circuits
Cross-capture linkage by verifier	one-time binding keys; ZK anchors; per-capture assertion certs never revealed	bundle inspection
Per-capture platform surveillance	delegation: Apple contact = enrolment only	network-free capture path
Rephotography of a displayed image (analogue hole)	not refused: the capture event is genuine; labels claim capture, never scene veracity (§IV-A)	out of scope by design

Sources for whom capture time is itself sensitive should treat the current design accordingly.

2) **Fetch patterns partition the anonymity set.** In-browser verification requests public constants, including registry roots; the serving host (or a network observer) learns which *registry* a bundle verifies against, plus requester IP and timing. The anonymity set of a proof is therefore its registry's enrolled population, not "all cameras": a single-user registry identifies its user. Nothing in the proof layer prevents large shared registries; deployments should prefer them.
3) **Bundle fields are distinguishing metadata.** Anchor type (zkAttest vs. the per-capture attestation fallback), aaguid environment, tier, and night-proof frame count appear in the clear and partition bundles into equivalence classes. None identifies a device; together they shrink the crowd a bundle hides in.
4) **Enrolment-time surfaces.** Apple observes one attestation per device at enrolment (never per capture), and the recovery credential syncs through the user's iCloud Keychain (§IV-E), putting existence-of-enrolment (not captures) within reach of that account's compromise or compulsion.

The pixel layer is a separate channel. The four points above are metadata; the shared *image* is its own leak. A published photo retains its PRNU — the physical sensor fingerprint every photo carries — which a forensic observer can correlate to cluster a source's photos, or match against a named reference, with no access to the proof, the registry, or any of our machinery. Zero knowledge makes the proof unlinkable; it cannot reach the pixels. The anonymous-share path therefore suppresses the fingerprint in the shared derivative: an informed suppression in the linear sensor domain, in the lineage of flatfield fingerprint removal [39] and adaptive PRNU denoising [40], but driven by the device's own enrolled fingerprint and applied single-frame at develop time, so the bound derivative is the scrubbed one and the sensor claim over the original RAW is untouched. On its first hardware validation (one phone's scenes developed through the same single-frame path the shared image takes, then scored by an adversary who estimates the fingerprint from the source's own *developed* photos and matches it against the developed shared image, the attack an observer actually mounts on a published photo), unscrubbed captures were attributable on 6 of 12 scenes (peak-to-correlation energy, PCE, up to 700); all 12 scrubbed captures fell below the conventional detection threshold of 60 (max 43). This is a single-device data point, not an evaluation. What this does *not* defeat is an adversary who pools many of a source's scrubbed photos and averages their residuals to re-estimate the fingerprint [41]; aggregation-robust removal is an open problem we do not claim to solve, and a source depending on pixel-level unlinkability should treat single-image removal as the boundary. This is a distinct contribution track and we develop it separately; here it closes the pixel half of the anonymity account.

The registry trust model, stated once. A registry stores, per identity: the enrolled roots and cancellable leaf, the enrolment-consistency label, the recovery credential's SHA-256, and the enrolment attestation key (§IV-D). It learns registration events and revocations; it does not see captures, proofs, or verification queries (verification touches only published roots). What compulsion yields is therefore the enrolment record: proof that a device enrolled, never which photos it took — captures remain unlinkable even to the registry. Who

operates a registry is a deployment parameter (self-hosted, organisational, or public), and point 2 above is the governing criterion: bigger, shared registries give their members a bigger crowd.

G. What the negative results contribute

The 15-camera corpus of the companion paper ([8, §6.3]) provides the statistical false-accept side (zero accepts / 300 trials, extended through the transform and to all camera pairs in §V-H); the on-device second-device demonstration is now *done* (2026-07-10): a second iPhone rejected on all eight captures (bind Hamming mean $\approx$16 344 vs the 15 534 bound, nearest 694 bits clear), the v3 membership proof refused (unsatisfiable constraint), and the same device, re-enrolled as its own identity, verified genuine (§V-C). The Dark Probe result (§V-D) closes the tempting "second fingerprint" avenue: no exposure-scaling or temperature-responsive dark signal survives the sensor's on-die calibration on this hardware — claims built on iPhone DSNU would have been built on read-noise FPN. And where the thermal signal *does* survive, consumer DSLRs without on-die suppression (§V-E), it is still not possession-proving: a per-pixel Meyer–Neldel proportionality collapses the challenge to a global scale knob a handful of the victim's dark frames reproduce. The alternative anchor fails not just on the phone but on the physics. The system's security budget is spent where the physics actually is: PRNU.

V. Evaluation

System platform: iPhone 13 Pro Max (A15), iOS 26; desktop = Apple M-series. The anchor-boundary study (§V-D, §V-E) additionally uses two Canon EOS 1100D bodies. Stack: Noir (nargo 1.0.0-beta.20), barretenberg 5.0.0-nightly.20260324 (identical nightly across phone FFI [Swoirenberg/noir_rs], desktop CLI, and browser WASM [bb.js]): one proving stack, three verification surfaces, byte-identical verification keys throughout.

A. Circuits

Table V lists the deployed circuits and their measured costs.

Proof size: 14 656 B (458 field elements, UltraHonk ZK) for every top-level circuit; public inputs vary (1 field for v3; 160 for the delegate; 164 for the aggregate). VK: 115 fields + hash.

Notes. (a) The P-384 issuance verification dominates zk_attest ($\sim$207k of 322k gates measured on an isolated spike circuit). (b) Recursive ZK verification costs $\approx$740k gates per inner proof in this stack; aggregate accordingly. (c) Aggregation runs off-device because its witness holds no secrets (§II-E); phone peak memory stays $\approx$1.0 GB, within iOS limits. (d) The v2/v3 timings here and in Table III come from separate measurement sessions (this census vs the July-9 cross-verified run); both are reported unmodified.

B. Verification surfaces

The three verification surfaces are compared in Table VI. The browser surface makes no-logging structural: the page's only server requests are for public constants (VK, pinned circuit hashes, registry roots, CRS ranges), all served same-origin; the bundle never leaves the client. (The in-page verification path is validated 1:1 against the server verdict in Node.)

C. End-to-end demonstrations (all on real hardware)

1) Capture $\rightarrow$ extract (RAW discarded) $\rightarrow$ on-device v3 proof $\rightarrow$ AirDrop $\rightarrow$ desktop bb cross-verification against original VKs.
2) Self-serve loop by the device owner, unassisted: enrol (24 flats, quality-gated) $\rightarrow$ register $\rightarrow$ capture $\rightarrow$ prove $\rightarrow$ web verdict.
3) zkAttest full chain: enrolment attestation (the one Apple contact) $\rightarrow$ capture with zero network $\rightarrow$ delegation proof from a per-device persisted certificate $\rightarrow$ three-check verdict on the public verifier.
4) Aggregation: real delegate + v3 proofs $\rightarrow$ one recursive proof $\rightarrow$ 4-check verdict (anchor, membership, pinned inner VKs, binding) live.
5) Refusals: black frame (constraint-unsatisfiable), a demo-`TAU` soundness gap caught and fixed by moving to the production `TAU`, tampered TBS/challenge/authData/signature/binding/pins each refused.
6) Impostor 2×2 matrix, demonstrated on hardware (2026-07-10). Device A (the enrolled iPhone 13 Pro Max) passes against its own enrolment (GENUINE; v3/v2 proofs verify). Device B (an iPhone 11 Pro, iOS 17) against A's enrolment is rejected on all eight captures: bind-tier Hamming 16 228 / 16 233 / 16 312 / 16 318 / 16 351 / 16 405 / 16 437 / 16 469 of 32 704 (mean $\approx$16 344, chance) against $\tau \cdot N = 15\,534$, the nearest impostor 694 bits outside the bound: "NOT THIS SENSOR" (R1); the v3 membership proof refuses to generate (unsatisfiable constraint, R2); and re-enrolled as *its own* identity, device B verifies genuine end-to-end (its sensor and anchor proofs check true on the live verifier), so R1/R2 are the sensor, not the install (R3). One impostor capture that auto-triggered a 12-frame night proof still scored 16 419: burst averaging amplifies a signal the impostor does not have, so it cannot manufacture a false accept. The classic each-device-passes-its-own / fails-the-other's biometric result, on real hardware.
7) DSLR sensor class, same pipeline (2026-07-13). The identical normative extractor, fed each sensor's own measured geometry and levels (no phone constants), runs on the two Canon EOS 1100D bodies of §V-E: 32 flats captured per body, quality-gated to 30–32 (the gate auto-dropped 2 of 64 as moved or underlit), half enrolled as the fingerprint and half held out; decode is the native path the desktop host will ship (LibRaw), verified byte-identical to the reference decoder on 28/28 corpus frames. Between two physically identical same-model bodies, the standard PRNU separation [6] holds with wide margin: 31/31 held-out flats accept against their own fingerprint (PCE 1.1–1.3×10^5) and 0/31

TABLE V
CIRCUITS: STATEMENTS, GATE COUNTS, AND MEASURED PROVING AND VERIFICATION TIMES.

Circuit	Statement	Gates	Phone prove	Phone verify	Desktop prove
prnu_key (v1)	key-tier FE decode + commitment	14.3k	430 ms	—	—
prnu_key_v2	anonymous identity (key tier, Merkle)	28.7k	595 ms (self-test)	—	—
prnu_bind_v3	per-photo threshold + anonymous membership	170k	4431 ms	—	∼0.9 s
zk_attest	Apple attestation in ZK	322k	8469 ms	2031 ms	3.7 s
zk_attest_delegate	+ local assertion over capture binding	415k	9965 ms	2596 ms	3.9 s
capture_aggregator	recursive: delegate + v3, one proof	1 479 937	(off-device by design)	—	16.3 s / 1.9 GB

TABLE VI
VERIFICATION SURFACES.

Surface	Verifier	Delegate proof	v3 / aggregate
on-device (Swift/FFI)	same bb nightly	2.6 s	sub-second
server (bb CLI, fly.io 512 MB VM)	subprocess	∼0.1 s	∼0.1 s
in-browser (bb.js WASM, single-threaded)	in-page	783 ms incl. warmup	$<$ 10 ms

TABLE VII
DARK-SIGNAL THERMAL PROBE.

State	Dark signal $\Delta(1\,\mathrm{s} - 0.1\,\mathrm{s})$	Self-NCC (1 s)	Long-phase mean DN
cold (genuinely pre-cooled)	0.53 DN	0.093	2112.20
hot (post-proving)	0.96 DN	0.127	2112.63
(warm, plugged-in baseline)	1.20 DN	0.130	2112.87

against the other body's (max 32, threshold 60); 12 real scene photographs from body A, imported from the card as a photographer would, all match A's flat-enrolled fingerprint (PCE 1.6×10^4 to 1.1×10^5) and none match B's (max 30). That is the import audit's PRNU check on genuine content, not calibration frames. This is a two-body validation that the credential and import audit run unmodified on a second sensor class, not a uniqueness evaluation (that is the fifteen-camera corpus, §V-H).

D. Negative result: a measurable but unusable thermal channel

Protocol: rear lenses covered, ISO 800 locked, exposure verified at 1.000 s / 0.100 s (frame duration pinned; quality prioritisation `.speed`; without both, the pipeline silently normalises exposure), offset-free metrics (the 14-bit Bayer buffers carry 12-bit data left-shifted 2: black ≈ 2112 DN, and naive black-level subtraction manufactures a phantom constant signal). Results in Table VII.

A thermal response *is* present but sub-DN. With a genuinely cold start, the dark-current delta rises ∼1.8× under load (0.53 → 0.96 DN) and the long-exposure dark mean tracks temperature monotonically across runs (2112.20 → 2112.87) while the 0.1 s baseline is pinned at 2111.67 ±0.01. This is physically consistent with dark current doubling per $\sim 7\,^\circ\mathrm{C}$, the thermal dependence forensic dark-current work relies on to recover capture temperature [42]. (An earlier warm baseline masked this: with the phone pre-heated, the "cold" delta was already elevated to 1.20, so cold ≈ hot.) Crucially the effect is ∼0.5–1 DN against a ∼16 000 DN range: on-die calibration suppresses it 3–4 orders of magnitude below any fingerprinting floor. The surviving read-noise FPN (NCC ≈ 0.1) has an extreme-pixel tail that is only ∼33 % stable ACROSS power cycles (52 % within a session; most of the tail is session/thermal, not device-persistent). Net: a *measurable but unusable* thermal channel, consistent with what dark-image-ID work exploits post-clamp [43]. iPhone-specific DSNU literature is otherwise absent; we offer this as a citable negative that rescopes thermal-PUF proposals to sensors without on-die suppression — a class we then measure directly in §V-E.

E. The anchor boundary: dark-current challenge-response is a fingerprint, not a PUF

PRNU is our anchor; this subsection is the measured case against the leading alternative. A challenge-response variant of dark current (vary the exposure/temperature "challenge" and check the sensor's dark "response") would be a stronger anchor than static PRNU if it were possession-hard: a copied response could not answer a fresh challenge. The need is documented: ABC [44] authenticated smartphones by their PRNU fingerprint and was broken two years later by fingerprint substitution from public photos [45] (see also the at-scale re-evaluation [46]); a possession-hard challenge is exactly what a static fingerprint cannot offer. We show dark current is not it, on the one sensor class where the signal even survives. Consumer DSLRs are exactly that class (§V-D rescoped to sensors with no on-die dark-current clamp). We measured two same-model Canon EOS 1100D bodies (the uniqueness claim is distinguishing physically identical hardware) over a $4\times$ exposure ladder in two thermal states: 192 temperature-labelled RAW dark frames. The full capture protocol, the capture-time thermal-drift confound and its per-pixel remedy, and the four physics gates are in Appendix A. The headline: **all four gates pass**. The temperature-corrected dark-rate map α is a first-class device fingerprint — cross-body NCC $= 0.0011$ with *zero* shared hot-pixel positions, split-half NCC $0.978/0.986$, thermal response $4.1\times$ / $5.0\times$ from 32 to $52\,^\circ\mathrm{C}$, held-out exposure NCC 0.85 (Table X, Fig. 4).

The security gate: fail. A fingerprint is necessary but not sufficient for a *challenge-response* PUF [47], [48], whose premise is that the response to a fresh challenge cannot be produced without the physical sensor. We ran the modeling-cost simulation: attackers predict the held-out hot-state long-exposure response, scored by NCC against a 0.5 match threshold. The public-photo attacker (no dark frames) scores ≈ 0: dark frames never survive into shared derivatives, so the fingerprint is *data-hard*. But the partial-RAW attacker, holding as few as *four* of the victim's ambient dark frames and *no* thermal sweep, forges the hot-state challenge at NCC 0.80 (0.89 with sixteen), matching the possession attacker at 0.85 (Fig. 3). Modeling = a handful of dark frames, not possession: the pre-registered failure condition of the modeling-cost gate.

Mechanism, measured. The reason is not conjectured (Fig. 2): per-pixel thermal sensitivity γ is ≈proportional to base dark rate α (γ–α correlation $0.96/0.99$, the Meyer–Neldel relation [49] at per-pixel granularity, as Widenhorn et al. measured for CCD dark current [50]), so temperature is a single *global* scale knob; and correlation-based verification is scale-invariant, so that knob carries no per-pixel challenge entropy. Residual thermal identity beyond proportionality is noise-dominated (split-half $0.34/0.35$). Consumer-sensor dark current is a superb *static* fingerprint and *not* a PUF; possession-hardness would need a challenge dimension with per-pixel diversity that survives scale-invariant verification, which this physics does not provide. The static fingerprint remains directly useful: because dark-frame features never leak into shared derivatives, it is a data-hard import-audit credential even though it is not a PUF.

What the failure closes. Even a runtime challenge issued at the shutter is dead on this class: a holder of a few of the victim's dark frames synthesises the response for any exposure or temperature delta, so challenge nudges are detectable but not unforgeable. A corollary: dark current gives DSLRs no "taken-now" mechanism; the ceiling for capture-time claims on this class is appeared-within windows. Scope: one model, two bodies; the Meyer–Neldel mechanism is generic silicon physics, while the measured constants are model-specific.

Positioning, measured. To our knowledge (qualified as in §III-F, pending a final librarian pass), no image-sensor PUF in the academic literature uses thermal state or exposure time as its *challenge* dimension. The CIS-PUF line (Cao et al.'s pixel-pair design [51], CamPUF's dark-signal FPN keys [52], the Okura et al. transistor-variation series [53]) selects pixels spatially as the challenge and engineers temperature *out*: differential readout, post-processing circuits, reliability quoted across wide thermal ranges. Rosenfeld et al.'s founding sensor-PUF concept [54] made the sensed stimulus part of the challenge, but for light through a random coating on a single photodetector, not the thermal state of an imager. This study inverts the field's framing (temperature as the challenge rather than the nuisance to stabilise away); the inversion does not survive measurement. It is also, to our knowledge, the first dark-current PUF evaluation on a consumer DSLR (prior work is custom silicon or phones) and the first use of per-pixel thermal sensitivity, previously dark-current metrology [55], [50], as a candidate authentication credential; the go/no-go gated protocol with thresholds fixed before capture (Table X) is reusable beyond this instance. These are the reasons the measurement was worth running, and why its negative verdict carries information for the anchor choice.

F. Deployment findings: what multi-sensor hardware actually looks like

Extending enrolment to every rear lens (each physical sensor is its own biometric) surfaced hardware behaviour absent from the PRNU literature, all measured on-device.

The padding trap. The 13 Pro Max ultrawide's classic-Bayer buffer is 4224×3024 against a ≈4032-wide active array; the margin is ISP edge-fill: *copied* pixel data. Copied pixels have residual exactly zero, hence *perfect* sign stability, and a stability-ranked selector packs the fingerprint with them: 99 % of selected positions landed in the outer band (uniform baseline ≈21 %). The resulting template is built of device-independent bits — predictable from the model's fill convention, the inverse of a biometric. The diagnostic signature is reusable: Hamming distance constant to 1 bit in $32\,704$ across opposite scenes ($19\,356$ dark / $19\,355$ bright: frozen bits, not weak signal). The remedy is a selection rule, not a spec change: a position may join the fingerprint only with *photosite evidence*, per-position residual energy above 1 % of the sensor's own sampled median (self-calibrating, no device constants) and a light response above the darkness floor. Validated: edge share $99\,\% \rightarrow 7\,\%$, raw distance $19\,356 \rightarrow 12\,324$ (verifies, $+3\,210$ margin); security-relevant because predictable bits are exactly the direction false accepts come from.

Capability and CFA heterogeneity. Classic-Bayer RAW availability varies *by lens within one phone*: the 13 Pro Max exposes it on all three rear sensors (UW 4224×3024 RGGB, wide 4032×3024 BGGR, tele 4032×3024 RGGB); the 11 Pro on wide and tele but not its ultrawide — and the same lens role differs in CFA order across devices (wide: BGGR vs. RGGB). Enrollability is therefore *measured per lens on device* (a one-time capability census), never assumed from a model table; sensor black/white levels are read from each lens's own RAW metadata (both devices report a 12-bit black of 528 in a 14-bit container, i.e. 2112) rather than hardcoded.

G. Night proofs: burst-averaged threshold verification

Dim scenes are read-noise-limited (PRNU is multiplicative). Averaging N RAW frames cuts uncorrelated noise $\sim \sqrt{N}$ while the photosite-fixed fingerprint adds coherently; and, unlike every burst technique in the imaging literature, *no frame alignment is required*: the pattern cannot move between frames, and drifting scene content blurs into the average, which helps the residual. Table VIII gives the rescue curve on a failing indoor scene (13 Pro Max).

Signal depth below chance, measured against the burst's mean single-frame depth (the table's single-frame row is the burst's first frame, which frame-to-frame variance places

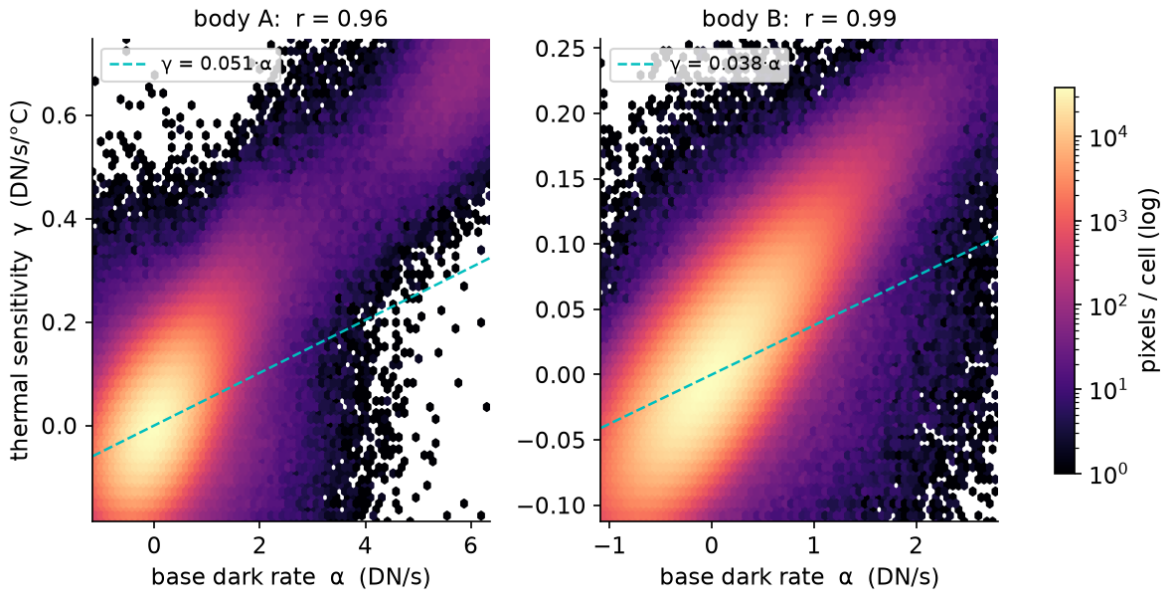

Fig. 2. Meyer–Neldel: per-pixel thermal sensitivity γ is proportional to base dark rate α ($r = 0.96/0.99$; hexbin density over the full map). Temperature is a global scale knob, and scale-invariant (NCC) verification sees no challenge entropy in it: the mechanism behind the failure.

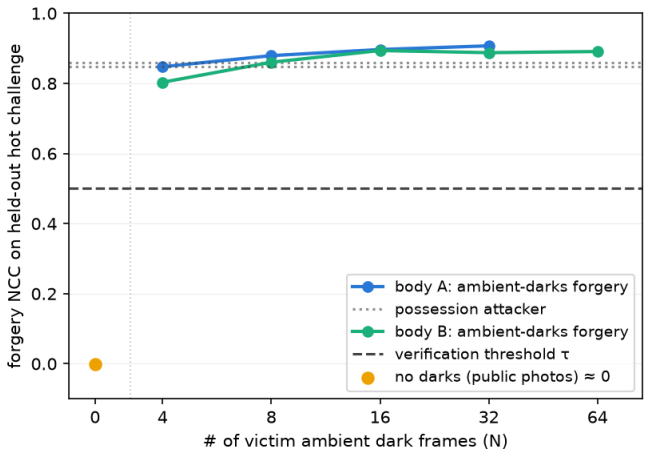

Fig. 3. Modeling-cost gate. An attacker with N of the victim's ambient dark frames forges the held-out hot-state challenge at possession-level NCC from $N = 4$ upward; the no-darks (public-photo) attacker scores ≈ 0. Modeling equals a few dark frames, not possession.

TABLE VIII
NIGHT-PROOF RESCUE: BURST AVERAGING VS. THRESHOLD $\tau \cdot N$.

Evidence	Hamming	Margin vs. $\tau \cdot N$
single frame	15 806	−272 (refused)
2-frame average	15 418	+116 (verifies)
4-frame average	15 029	+505
6-frame average	14 517	+1 017

on either side of that mean), grew 796 → 1 835 over six frames (2.31×) against $\sqrt{6} \approx 2.45$ predicted; a 2026-07-12 replication on a dimmer scene reproduced the scaling (269 → 620, 2.30×, with the 2- and 4-frame points at 1.48× and 1.97× against $\sqrt{2} \approx 1.41$ and $\sqrt{4} = 2.00$). Two darker scenes characterise the limits: deep-dim gains ∼400 bits per 6 frames without crossing (an adaptive 12-frame burst closes it: production rescues sealed at +43 and +9 margin), and near-black (single ≈ chance) is unrescuable: the floor. In production the capture engages the burst only when frame 1 would not verify *and* raw signal is present (below chance −200); the sealed witness and enclave binding cover the averaged mosaic (the exact evidence the proof consumes) within the unchanged 144-byte signed-message format, and the frame count travels in the record and both proof bundles. Enrolment needed no change (it has averaged 24 flats from the start); verification borrows the same physics when light is scarce. The security side is symmetric: 12 frames of averaging moved an impostor from the ≈16 344 cluster to 16 419 — nowhere.

H. Corpus transfer statistics: methodology and sample sizes

The population-level false-accept evidence in this paper comes from the companion's 15-camera corpus [8, §6.3], carried here through the keyed transform so the statistics exist in the domain the deployed v3 circuit actually consumes. Protocol: fifteen iPhone camera modules, each contributing 30 photos for a maximum-likelihood fingerprint (512-pixel central crop, four Bayer planes, 1 048 576 positions) and 20 held-out probe photos; the corpus operating point is each camera's own top-N = 32 704 magnitude mask, with the enrolled code the fingerprint signs at that mask and probe codes the wavelet-residual signs of single photos. Template-domain trials use the production KDF (the reference implementation the shipped transform is unit-tested against) at the bind tier, with independent 256-bit keys per enrolled camera and the vectorised pipeline parity-checked bit-for-bit against

TABLE IX
CORPUS-THROUGH-TRANSFORM STATISTICS AT $\tau = 0.475$. "ONE PAIR" IS THE CORPUS'S TWO SAME-MODEL TELEPHOTO MODULES (SEE TEXT).

Statistic	Trials	Outcome
raw per-photo impostor, companion pairing	300	0 accepts; nearest $+501$ bits
raw per-photo impostor, all 210 ordered pairs	4 200	6 accepts, all one pair; worst -527 bits
template per-photo impostor, companion pairing, 100 keys	30 000	0 accepts; nearest $+389$ bits
template per-photo impostor, all pairs, 10 keys	42 000	7 accepts, one photo of the same pair; worst -220 bits
template impostor from enrolment-grade fingerprint estimates, 100 keys	21 000	200 accepts: the same pair, every key; worst -835 bits
perfect raw-bit estimate under a wrong key (mated cross-key)	74 250	0 accepts; nearest $+376$ bits
unlinkability: non-mated vs mated cross-key scores	1 050 000	distributions coincide; $D_{\mathrm{sys}} = 0.0064$

the reference in every run; key-generation seeds are recorded with the results. Table IX gives every corpus statistic with its sample size.

The same-model boundary. Every threshold crossing in Table IX traces to one unordered pair: the corpus's two iPhone 13 Pro back-telephoto modules (identified from capture metadata). They are distinct devices — cross-pair per-photo BER (mean 0.4875) sits well apart from own-genuine (mean 0.470) — but they share roughly 40 % of the genuine deviation from chance, and the corpus's third telephoto module (an iPhone 8 Plus) is elevated against both while staying outside the bound (enrolment-grade pair BER 0.482–0.485). The reading is lens-class/model-common structure that survives the extractor's row/column zero-meaning and concentrates exactly where the magnitude mask selects; a stability-selected mask reproduces the effect, so mask choice does not close it. The standard forensic remedy, suppression of non-unique artefacts shared across a camera model, is not part of this pipeline; it is recorded as a required measurement, not assumed to close the gap.

Two scope statements delimit what the table claims. First, the crossing concerns the threshold leg only: it is measured with the enrolled camera's own key applied to foreign bits. The forgery claim (the key leg) is unaffected and separately measured at chance (§III-C). The deployed enrolment also differs from the corpus transfer protocol (flat-field enrolment, stability selection, own-device capture), and the on-device second-device demonstration (§V-C) is cross-model; a same-model, same-lens-class pair has not been tested on the deployed protocol, so same-model margins are bounded by the corpus result above rather than by the idealised tail (§III-C). Second, the corpus does not measure the deployed genuine operating point: natural-photo enrolment and single-photo probes leave corpus genuine transfer at raw BER mean 0.477 (300 trials; template mean 0.485 over 30 000 trials), far from the deployed flat-field enrolment's 0.069, so corpus evidence in this paper is impostor-side only.

I. Availability and ethics

The biometric data in this study are the authors' own devices' sensor fingerprints; no third-party subjects or images are involved. No fingerprint-bearing data (raw captures, enrolment bits, templates, or per-pixel maps) is released: published figures are aggregate-only by construction. The verifier is publicly reachable, in-browser verification included; circuits, verification keys, and measurement tooling are available from the authors. Per-statistic sample sizes for the corpus package are in Table IX; single-device statistics carry their sample sizes at the point of claim (eight second-device captures, §V-C; four enrolment-proof runs, §III-C).

VI. CONCLUSION

A provenance system does not have to choose between proving the camera and protecting the photographer. On a commodity phone, Apple's attestation becomes a zero-knowledge fact (one platform contact per device, ever), the sensor's fingerprint becomes a cancellable template matched inside a SNARK (forgery-resistant because the key never leaves the device; revocable because the key can rotate), and verification runs where no log can exist — in the verifier's own browser. Each layer's refusal was demonstrated on hardware, including the one that matters most: a second device photographing ordinary scenes is refused by the physics, not by policy. The anchor itself was chosen by measurement: the leading alternative, dark-current challenge-response, fails possession-hardness even on the sensor class where its signal survives (§V-E). The remaining limits: bit extraction is app-asserted pending IVC-style proofs (§IV-D), the enrolment consistency proof does not yet bind the bind-tier template to the same witness (§III-C), and the cross-device impostor measurement extends next to the ultrawide and telephoto lenses. The security budget is spent where the physics is; the privacy budget, where the logs cannot be.

APPENDIX A
DSLR CAPTURE PROTOCOL, THERMAL CONFOUND, AND PHYSICS GATES

This appendix carries the measurement detail behind the anchor-boundary summary in §V-E.

Capture protocol. Two same-model Canon EOS 1100D bodies, a $4\times$ exposure ladder ($1/10$–15 s) at ISO 800, lens-capped, in two thermal states (rested $\approx 30\,^{\circ}\mathrm{C}$; self-heated to $\approx 60\,^{\circ}\mathrm{C}$ by ten minutes of tethered live-view), 8–16 frames per cell (the quieter body re-captured at 16): 192 RAW dark frames, each carrying its own sensor temperature in MakerNotes. On-body long-exposure noise reduction was disabled — it dark-subtracts the signal.

Temperature is a confound, not a nuisance. Fitting dark signal against exposure per nominal state fails backwards: the hot/cold slope ratio comes out below unity, impossible for real dark current. The cause is capture-time thermal drift: the hot ladders *cool* across increasing exposure (the long frames, which dominate the slope, land coolest) while cold ladders warm, an anti-correlation that inverts the contrast. The remedy

TABLE X
THE FOUR PHYSICS GATES, BODIES A / B. THE DARK-RATE FINGERPRINT PASSES EVERY PHYSICAL TEST; THE FAILURE THAT FOLLOWS IN §V-E IS SECURITY, NOT PHYSICS (FIG. 3).

Gate	Threshold	Measured (A / B)	Verdict
M1 exposure-fit R^2	≥ 0.80	0.81 / 0.85	pass
M2 repeatability (split-half NCC)	≥ 0.50	0.978 / 0.986	pass
M3 cross-body NCC	< 0.10	0.0011	pass
M4 thermal ratio (32→52 °C)	> 1.3	4.1× / 5.0×	pass

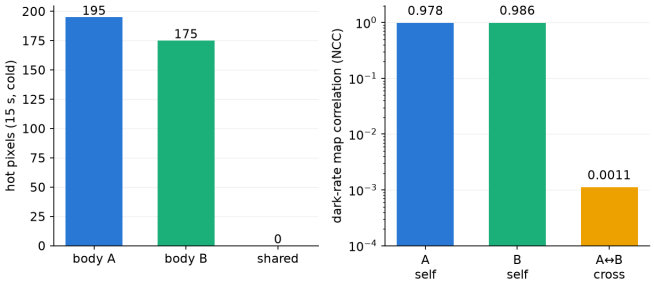

Fig. 4. Two same-model 1100D bodies. Left: 15 s cold hot-pixel counts; zero shared positions. Right: the dark-rate map correlates with itself (split-half) but ≈0 across bodies — uniqueness is positional, not distributional.

is to fit, per pixel, the physical model over all cells at once using each frame's *measured* temperature, $s(t, T) = t\,[\alpha + \gamma\,(T - T_{\text{ref}})]$, where α is the temperature-corrected dark rate and γ its thermal sensitivity. This is the sequel to the offset-free lesson of §V-D: label temperatures, do not trust state labels.

The physics gates. Corrected, the dark-rate map α passes all four pre-registered gates (Table X). Two physically identical bodies are orthogonal: cross-body NCC $= 0.0011$, and their 15 s hot-pixel maps (195 vs 175 defects) share *zero* positions (Fig. 4). The map is stable (split-half NCC $0.978/0.986$), strongly thermally responsive (α rises $4.1\times/5.0\times$ from 32 to 52 °C), and exposure-predictable (held-out-exposure NCC 0.85).

GENERATIVE AI USAGE

The author used Claude for circuit and tooling implementation and for drafting the manuscript. All results were independently verified by the author against the controls reported herein. The author takes full responsibility for the content.

REFERENCES

[1] Coalition for Content Provenance and Authenticity, "C2PA specification 2.x," https://c2pa.org.

[2] C. Paquin, "Privacy for C2PA signers," Christian's Corner, https://christianpaquin.github.io/2026-03-13-privacy-for-c2pa-signers.html, 2026, 2026-03-13; prototype: github.com/christianpaquin/c2pa-signer-privacy.

[3] E. Lynch and S. Hanna, "How Pixel and Android are bringing a new level of trust to your images with C2PA content credentials," Google Security Blog, https://security.googleblog.com/2025/09/pixel-android-trusted-images-c2pa-content-credentials.html, 2025, 2025-09-10; StrongBox/Titan M2 keys, one certificate per image, Assurance Level 2 (Pixel Camera app), no-logging CA policy, on-device offline TSA. Accessed July 2026.

[4] N. Krawetz, "Google Pixel 10 and massive C2PA failures," The Hacker Factor Blog, https://hackerfactor.com/blog/index.php?/archives/1077-Google-Pixel-10-and-Massive-C2PA-Failures.html, 2025, 2025-09-05; the per-capture query to Google's trusted timestamp service lets Google associate signatures with IP addresses and capture times, with no user opt-out. Accessed July 2026.

[5] K. Kaye and P. Dixon, "Privacy, identity and trust in C2PA: A technical review and analysis of the C2PA digital media provenance framework," World Privacy Forum, https://worldprivacyforum.org/posts/privacy-identity-and-trust-in-c2pa/, 2025, 2025-09-03; documents unpublished trust-list and conformance criteria under the C2PA steering committee, and identity-assertion privacy risks. Accessed July 2026.

[6] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006, dOI 10.1109/TIFS.2006.873602.

[7] A. B. J. Teoh, D. C. L. Ngo, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition*, vol. 37, no. 11, pp. 2245–2255, 2004, dOI 10.1016/j.patcog.2004.04.011.

[8] J. Newton, "Stable bits, not strong ones: Selecting PRNU bits for on-device cryptographic camera provenance," Companion manuscript, 2026.

[9] E. Brickell, J. Camenisch, and L. Chen, "Direct anonymous attestation," in *ACM Conference on Computer and Communications Security (CCS)*, 2004.

[10] Aztec Labs, "Noir: A domain-specific language for zero-knowledge proofs," github.com/noir-lang/noir; docs https://noir-lang.org, 2026, accessed July 2026.

[11] ——, "Barretenberg: The UltraHonk proving system," github.com/AztecProtocol/aztec-packages (subdirectory `barretenberg/`), 2026, pLONK-lineage prover (Gabizon, Williamson & Ciobotaru, ePrint 2019/953) with sumcheck over the hypercube. Accessed July 2026.

[12] a00012025, "zk-app-attest," github.com/a00012025/zk-app-attest, 2024, rISC Zero zkVM, server-proved App Attest.

[13] League of Entropy, "drand: Distributed randomness beacon," https://drand.love; source github.com/drand/drand, 2020, implements Syta et al., "Scalable Bias-Resistant Distributed Randomness," IEEE S&P 2017.

[14] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the Weil pairing," *Journal of Cryptology*, vol. 17, no. 4, pp. 297–319, 2004, dOI 10.1007/s00145-004-0314-9; preliminary version in ASIACRYPT 2001.

[15] Apple Inc., "Establishing your app's integrity," Apple Developer Documentation, DeviceCheck framework, https://developer.apple.com/documentation/devicecheck/establishing-your-app-s-integrity, 2026, accessed July 2026.

[16] themighty1, "zk_cert_verification.nr," https://github.com/themighty1/zk_cert_verification.nr, in-browser zero-knowledge verification of X.509 certificate chains in Noir; the inverted-assert idiom for witnessed DER offsets.

[17] M. Goljan, J. Fridrich, and M. Chen, "Defending against fingerprint-copy attack in sensor-based camera identification," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 227–236, 2011, dOI 10.1109/TIFS.2010.2099220.

[18] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001, dOI 10.1147/sj.403.0614.

[19] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch, "General framework to evaluate unlinkability in biometric template protection systems," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1406–1420, 2018, dOI 10.1109/TIFS.2017.2788000.

[20] A. W.-K. Kong, K.-H. Cheung, D. Zhang, M. S. Kamel, and J. You, "An analysis of BioHashing and its variants," *Pattern Recognition*, vol. 39, no. 7, pp. 1359–1368, 2006, dOI 10.1016/j.patcog.2005.10.025; the stolen-token collapse analysis.

[21] M. Laskus, "Decentralized identity trilemma," Essay, https://maciek.blog/p/dit, 2018, self-sovereignty, privacy, Sybil-resistance: pick two.

[22] R. Jena, P. Singh, M. Mohanty, and M. L. Das, "PP-PRNU: PRNU-based source camera attribution with privacy-preserving applications," *Computing*, vol. 106, no. 10, pp. 3309–3333, 2024, preliminary version: R. Jena, P. Singh, M. Mohanty, "SSS-PRNU: Privacy-Preserving PRNU Based Camera Attribution using Shamir Secret Sharing," arXiv:2106.07029, 2021.

[23] A. Huang and J. S.-T. Juan, "L-PRNU: Low-complexity privacy-preserving PRNU-based camera attribution scheme," *Computers*, vol. 12, no. 10, p. 212, 2023, dOI 10.3390/computers12100212.

[24] D. Valsesia, G. Coluccia, T. Bianchi, and E. Magli, "User authentication via PRNU-based physical unclonable functions," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1941–1956, 2017, dOI 10.1109/TIFS.2017.2697402.

[25] F. Qian, S. He, H. Huang, H. Ma, X. Zhang, and L. Yang, "Web photo source identification based on neural enhanced camera fingerprint," in *Proceedings of the ACM Web Conference (WWW)*, 2023, dOI 10.1145/3543507.3583225; arXiv:2302.09228.

[26] S. F.-M. na and F. Pérez-González, "On the information leakage of camera fingerprint estimates," arXiv:2002.11162, 2020, mutual-information and membership-inference analysis of PRNU fingerprint estimates.

[27] A. Naveh and E. Tromer, "PhotoProof: Cryptographic image authentication for any set of permissible transformations," in *IEEE Symposium on Security and Privacy (SP)*, 2016, pp. 255–271, dOI 10.1109/SP.2016.23.

[28] C. Guo, L. You, and G. Hu, "A novel biometric identification scheme based on zero-knowledge succinct noninteractive argument of knowledge," *Security and Communication Networks*, vol. 2022, p. 2791058, 2022, dOI 10.1155/2022/2791058.

[29] Z. Lin, T. Wang, J. Lai, S. Zhang, Q. Yang, and S. C. Liew, "BioZero: Privacy-preserving and publicly verifiable on-chain biometric authentication via homomorphic commitments and zero-knowledge proofs," arXiv:2409.17509, 2024, revised April 2026; Pedersen commitments + Groth16. The 2024 v1 (Lai, Wang, Zhang, Yang, Liew) was titled "An Efficient and Privacy-Preserving Decentralized Biometric Authentication Protocol on Open Blockchain".

[30] P. Kothari, D. Chopra, M. Singh, S. Bhardwaj, and R. Dwivedi, "Incorporating zero-knowledge succinct non-interactive argument of knowledge for blockchain-based identity management with off-chain computations," arXiv:2310.19452, 2023, cancelable fingerprint template (keyed random-matrix transform) + Groth16 threshold circuit; the similarity matrix is a public input, matching runs off-circuit.

[31] X. Mao, X. Zhou, X. Zhao, and Y. Chen, "A ZKP-based anonymous biometric authentication scheme for the E-health systems," *PLOS ONE*, vol. 20, no. 6, p. e0324289, 2025, dOI 10.1371/journal.pone.0324289; cancelable multimodal templates with Pedersen commitments, distance check outside the proof.

[32] ISO/IEC, "ISO/IEC 24745:2022, information security, cybersecurity and privacy protection — biometric information protection," International Organization for Standardization, 2nd ed., 2022.

[33] C. Rathgeb, F. Breitinger, and C. Busch, "Alignment-free cancelable iris biometric templates based on adaptive Bloom filters," in *Int. Conf. on Biometrics (ICB)*, 2013.

[34] C. Rathgeb and C. Busch, "Cancelable multi-biometrics: Mixing iriscodes based on adaptive Bloom filters," *Computers & Security*, vol. 42, pp. 1–12, 2014.

[35] Y. L. Lai, Z. Jin, A. B. J. Teoh, B. M. Goi, W. S. Yap, T. Y. Chai, and C. Rathgeb, "Cancellable iris template generation based on indexing-first-one hashing," *Pattern Recognition*, vol. 64, pp. 105–117, 2017, dOI 10.1016/j.patcog.2016.10.035. The 0.54 % EER / CASIA-v3 figure is stated in the authors' companion conference version, DOI 10.1007/978-3-319-46298-1_29.

[36] C. Li and J. Hu, "Attacks via record multiplicity on cancelable biometrics templates," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 8, pp. 1593–1605, 2014, dOI 10.1002/cpe.3042.

[37] X. Dong, J. Park, Z. Jin, A. B. J. Teoh, M. Tistarelli, and K. Wong, "On the risk of cancelable biometrics," arXiv:1910.07770, 2019, pre-image attacks and leakage analysis across six cancelable-biometric schemes.

[38] Facebook, Inc., "The TMK+PDQF video-hashing algorithm and the PDQ image-hashing algorithm," Technical report, distributed via the ThreatExchange repository (github.com/facebook/ThreatExchange, `hashing/hashing.pdf`), 2019.

[39] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *Proc. 15th ACM Int. Conf. on Multimedia (MM)*, 2007, pp. 78–86, dOI 10.1145/1291233.1291252; informed flatfield fingerprint removal.

[40] A. Karaküçük and A. E. Dirik, "Adaptive photo-response non-uniformity noise removal against image source attribution," *Digital Investigation*, vol. 12, pp. 66–76, 2015, dOI 10.1016/j.diin.2015.01.017.

[41] S. Taspinar, M. Mohanty, and N. Memon, "PRNU-based camera attribution from multiple seam-carved images," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 3065–3080, 2017, dOI 10.1109/TIFS.2017.2737961; pooling individually anonymized images restores attribution.

[42] R. Matthews, M. Sorell, and N. Falkner, "Determining image sensor temperature using dark current," arXiv:1901.02113, 2019, dark current carries a recoverable thermal signal in image sensors — the channel we measure on clamp-free DSLRs.

[43] A. Berdich and B. Groza, "Smartphone camera identification from low-mid frequency DCT coefficients of dark images," Entropy, vol. 24, no. 8, art. 1158, 2022, dOI 10.3390/e24081158.

[44] Z. Ba, S. Piao, X. Fu, D. Koutsonikolas, A. Mohaisen, and K. Ren, "ABC: Enabling smartphone authentication with built-in camera," in *Network and Distributed System Security Symposium (NDSS)*, 2018, dOI 10.14722/ndss.2018.23099.

[45] C. Benegui and R. T. Ionescu, "A breach into the authentication with built-in camera (ABC) protocol," in *Applied Cryptography and Network Security (ACNS), LNCS 12146*, 2020, pp. 3–20, dOI 10.1007/978-3-030-57878-7_1; fingerprint substitution from external photos passes at 54.1 % FAR.

[46] D. Maier, H. Erb, P. Mullan, and V. Haupert, "Camera fingerprinting authentication revisited," in *Int. Symposium on Research in Attacks, Intrusions and Defenses (RAID), USENIX Association*, 2020, pp. 31–46.

[47] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002, dOI 10.1126/science.1074376.

[48] C. Herder, M.-D. Yu, F. Koushanfar, and S. Devadas, "Physical unclonable functions and applications: A tutorial," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1126–1141, 2014, dOI 10.1109/JPROC.2014.2320516; the strong/weak and challenge-response PUF definitions.

[49] W. Meyer and H. Neldel, "Über die beziehungen zwischen der energiekonstanten und der mengenkonstanten a in der leitwerts-temperaturformel bei oxydischen halbleitern," Z. Tech. Phys. (Leipzig), vol. 18, pp. 588–593, 1937.

[50] R. Widenhorn, L. Mündermann, A. Rest, and E. Bodegom, "Meyer–neldel rule for dark current in charge-coupled devices," *Journal of Applied Physics*, vol. 89, no. 12, pp. 8179–8182, 2001, dOI 10.1063/1.1372365.

[51] Y. Cao, L. Zhang, S. S. Zalivaka, C.-H. Chang, and S. Chen, "CMOS image sensor based physical unclonable function for coherent sensor-level authentication," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 11, pp. 2629–2640, 2015, dOI 10.1109/TCSI.2015.2476318.

[52] Y. Kim and Y. Lee, "CamPUF: Physically unclonable function based on CMOS image sensor fixed pattern noise," in *Proc. 55th Annual Design Automation Conference (DAC)*, 2018, dOI 10.1145/3195970.3196005.

[53] S. Okura, Y. Nakura, M. Shirahata, M. Shiozaki, T. Kubota, K. Ishikawa, I. Takayanagi, and T. Fujino, "A proposal of PUF utilizing pixel variations in the CMOS image sensor," in *Int. Image Sensor Workshop (IISW)*, 2017, dOI 10.60928/dgut-pu73; the line continues through Sensors 21(18):6079, 2021 (DOI 10.3390/s21186079).

[54] K. Rosenfeld, E. Gavas, and R. Karri, "Sensor physical unclonable functions," in *IEEE Int. Symposium on Hardware-Oriented Security and Trust (HOST)*, 2010, pp. 112–117, dOI 10.1109/HST.2010.5513103.

[55] E. A. G. Webster, R. L. Nicol, L. Grant, and D. Renshaw, "Per-pixel dark current spectroscopy measurement and analysis in CMOS image sensors," *IEEE Transactions on Electron Devices*, vol. 57, no. 9, pp. 2176–2182, 2010, dOI 10.1109/TED.2010.2052399.

[56] D. Kang, T. Hashimoto, I. Stoica, and Y. Sun, "ZK-IMG: Attested images via zero-knowledge proofs to fight disinformation," arXiv:2211.04775, 2022.

[57] M. Frigo and A. Shelat, "Anonymous credentials from ECDSA," Cryptology ePrint Archive 2024/2010; library github.com/google/longfellow-zk, 2024, deployed in Google Wallet.