Locally Recoverable Data Availability Sampling with Derived Group Commitments

Seunghyun Cho, Eun-Young Seo, Saehoon Jung, and Young-Sik Kim

Department of Electrical Engineering and Computer Science, DGIST
seunghyuncho@dgist.ac.kr, eyseo_dgist@dgist.ac.kr, shoon723@dgist.ac.kr, ysk@dgist.ac.kr

Abstract. Data availability sampling (DAS) enables a verifier to assess whether an encoded object is available for reconstruction without downloading it in full. Nodes that accumulate partial data additionally require disjoint recovery groups that they can certify and serve independently. We present LR-DAS, which commits to the source-group restrictions of a Tamo–Barg codeword, enabling every verifier to derive each group commitment as a public linear combination of the published commitments. While leveraging a linear derivation analogous to two-dimensional KZG designs, LR-DAS evaluates each derived group on its dedicated coset of a multiplicative subgroup, requiring only one diagonal scaling per group and achieving the optimal Singleton-type minimum distance. In the reference configuration, the minimum distance is 3.91 times that of the source-compressed product code with the same commitment size, reference string, and local serving interface, and the number of uniform samples required for a light client to detect any fixed unrecoverable withholding pattern with probability $1 - 2^{-128}$ decreases from 991 to 218. The reference string has the degree of the local dimension rather than that of the global polynomial, no Fiat–Shamir transform is used, and a node holding authenticated values at any r distinct positions of a group can produce valid openings for that group without auxiliary data. Against adaptive adversaries, we prove position binding, code binding, and canonicality in the algebraic group model under a discrete-logarithm assumption with auxiliary inputs, and we show by an explicit attack that degree enforcement is necessary.

Keywords: Data Availability Sampling · Locally Recoverable Codes · Tamo–Barg Codes · Erasure-Code Commitments

1 Introduction

Data availability sampling (DAS) enables a verifier to assess whether an encoded object is available for reconstruction without downloading it in full. The cost of replicating large objects across distributed systems [32,24] motivates checking availability without full downloads, which DAS enables through erasure coding and commitment-based verification [22]. Erasure coding makes sampling effective because preventing reconstruction requires withholding a substantial fraction of

the encoding [1]. Hall-Andersen, Simkin, and Wagner [22] formalize DAS through *erasure-code commitments*, which ensure that accepted openings are consistent with a codeword; we adopt their model, in which the adversary produces the commitment and answers queries adaptively.

A node that accumulates parts of an encoding, which we call a *collector*, requires two essential properties of the code: disjoint units that it can certify and then serve, and a global minimum distance large enough to minimize the sample complexity of a light client. Such a node stores no copy of the object and answers requests for a group once it holds r accepted values of it; the set of certified units is also useful to protocols that act on partial availability [34]. A single global commitment does not provide such units: in PeerDAS, each cell is the evaluation of the polynomial of one data block on a coset, and its proof is verified against the commitment to the whole polynomial [35], so a node that holds the data of a cell cannot produce new proofs for it before it has recovered that polynomial. A product of two Reed–Solomon codes provides disjoint units whose commitments are commitments to polynomials of degree less than r, but its distance is the product of two small distances. A Tamo–Barg code [33] provides the same units with the largest distance permitted by the Singleton-type bound of [29], but its codewords are evaluations of one polynomial of degree D; committing to that polynomial and linking each group to it by auxiliary data requires a reference string of degree D and per-group quotient and degree data that a node must obtain before it can serve (Appendix D).

Our construction overcomes both limitations. Specifically, write $k = ar + b$, fix a group labels γ_s, let L_s be the Lagrange basis on them, and set $B(T) := \prod_{s<a}(T - \gamma_s)$. Here, a is the number of source groups, f is the global encoding polynomial, u_s is its restriction polynomial on the sth source group, and e is a residual polynomial. Proposition 1 shows that

$$f(X) \;=\; \sum_{s<a} L_s\big(X^m\big)\, u_s(X) \;+\; B\big(X^m\big)\, e(X), \qquad \deg u_s < r,\ \deg e < b, \quad (1)$$

is a bijection onto the Tamo–Barg space V_D, with $f \bmod h_j = \sum_s L_s(\gamma_j)u_s + B(\gamma_j)e$ on group j. Committing to the u_s and to e therefore suffices: by linearity, every verifier *derives* $C_j = \sum_{s<a} L_s(\gamma_j)U_s + B(\gamma_j)E$ from the published commitments, and a position of group j is proven via a standard opening under C_j. The published commitment, which we call the *header*, has at most $a + 2$ group elements, the reference string has degree r, and there is neither per-group auxiliary data nor a Fiat–Shamir transform; when $b = 0$, the last term and its degree check vanish. Equation (1) restates the systematic encoding of Tamo and Barg [33, §III-D] in a form compatible with linear commitments; the new element is its cryptographic use: committing to low-degree source restrictions yields code binding for a code of much higher degree together with local serving.

The same derivation applies to the product-code design: two-dimensional KZG designs obtain the commitments of the extended rows as the Lagrange extension of the commitments of the original rows [7], so the product code of Section 7 has the same header, reference string, and local serving as our construction. The two

Table 1. Three designs on the same $n = 125{,}952$ positions with $k = 62{,}976$ and the same partition into $\ell = 123$ groups of $m = 1024$ positions. The one-dimensional design commits to a Reed–Solomon codeword of dimension k with one KZG commitment; the grouped designs use $r = 768$ and a dedicated reference string (Profile A, Section 5). "Values to serve a group" is the number of codeword values that a node requires before it can produce valid openings for one group without auxiliary data such as stored proofs. "Detection samples" is the number of uniform samples that one light client needs to detect any fixed unrecoverable withholding pattern with probability $1 - 2^{-128}$; the communication is that of such a client with single-point openings, header included.

	One-dimensional Reed–Solomon	Source-compressed product code	LR-DAS
Header	48 B	3,936 B	3,936 B
Published $\mathbb{G}_1$ powers	62,976	768	768
Values to serve a group	62,976	768	768
Minimum distance	62,977	10,794	42,241
Detection samples	128	991	218
Light-client communication	10,288 B	83,216 B	21,376 B

designs differ solely in the evaluation domain of each derived polynomial: on a single common set for the product code, and on the coset $A_j = \beta_j H$ for group j in our construction. By Proposition 6, this amounts to one diagonal scaling of the coefficients per group, yet it determines the achievable minimum distance: for free source blocks mixed by any matrix of rank a, a common evaluation set upper-bounds the distance by the product value $(\ell - a + 1)(m - r + 1)$ (Proposition 5), whereas the cosets attain $(\ell - a + 1)m - r + 1$, which is 3.91 times larger at our reference profile (Corollary 3). Table 1 summarizes the comparison, with a one-dimensional design as reference.

Deriving group commitments also has a security consequence. A supplied group commitment may commit to a different polynomial than the correct group data, causing proofs regenerated from the data to fail verification (Remark 5); this is the group-local instance of an obstruction that [22, Ex. 1] records for repair and excludes by message-bound openings and computational uniqueness [22, Def. 22, Def. 23]. With derived commitments, the corresponding property, which we call *canonicality*, holds by construction.

Related Work. One-dimensional sampling uses Reed–Solomon codes. PeerDAS extends each data block independently and authenticates its cells with KZG multiproofs; it was first analyzed in the algebraic group model [35] and later in the standard model under a falsifiable assumption [36], so a shorter reference string or a lower auxiliary degree does not inherently imply that our assumption is weaker than the assumptions established in that line of work. Two-dimensional schemes encode along rows and columns [1,7,10]; the tensor-code commitment of [22, §8] provides a rigorous formalization of this approach. Block circulant codes [30,31] exceed the relative distance of two-dimensional Reed–Solomon codes at high rate with KZG-compatible local codes; their local codes overlap and are defined by

local parity checks alone, so erasures are decoded from pairs of local codes. Our groups are disjoint and attain the Singleton-type bound for that structure, at the cost of global constraints (Proposition 7); we also fix the header, the reference string, and the recovery cost r, and a recovered group regenerates proofs from its r values and the header (Proposition 4). FRIDA compiles an interactive oracle proof of proximity into an erasure-code commitment with FRI [3,21], and two recent works extend that analysis beyond the unique decoding radius [27,37]. In ZODA, sampled rows and columns of a tensor encoding prove their own consistency without a trusted setup [17]. Boneh, Neu, Nikolaenko, and Partap [6] use locally correctable multiplicity codes and repair data and proofs at one coordinate from a few other proofs, whereas in our design a node that has reassembled r values of a group requires no auxiliary proof from other nodes to resume serving it (Appendix F.1). Locally recoverable codes were previously considered for DAS in a conceptual study with simulations [13]. A subsequent preprint [14] certifies disjoint groups of a Tamo–Barg codeword under a commitment to the global polynomial, with per-group local and quotient commitments linked by an out-of-domain check whose challenge is derived by the Fiat–Shamir transform; it analyzes non-adaptive withholding and instantiates this approach with KZG for $k = \ell r$, where the code is the direct sum of its local codes, and with a transparent proximity test. We remove the global commitment and derive all group commitments from commitments to the source-group restrictions (Construction 2), which yields code binding for every k at the optimal distance, canonicality by construction, a reference string of degree r, and security against adaptive adversaries without a random oracle (Theorem 4).

2 Preliminaries

Let $\mathbb{F}_q$ be a finite field and $Z_S(X) := \prod_{a \in S}(X - a)$. We fix an evaluation set $A \subseteq \mathbb{F}_q$ of size n partitioned into ℓ disjoint groups $A = \bigsqcup_{j<\ell} A_j$ with $|A_j| = m := r + \alpha$ and $n = \ell m$, where $r \geq 2$ and $\alpha \geq 1$ throughout. A code $C \subseteq \mathbb{F}_q^n$ has *disjoint* (m, r) *recovery groups* for this partition if each punctured code $C|_{A_j}$ is an $[m, r]$ maximum distance separable (MDS) code, so that any r symbols of A_j determine the remaining α. The local distance is then $\alpha + 1$, and the minimum distance obeys [29,25]

$$d \leq n - k + 1 - (\lceil k/r \rceil - 1)\alpha, \tag{2}$$

attained with equality by the codes of Definition 1. A product of two Reed–Solomon codes also has disjoint (m, r) groups, namely its columns, but does not attain (2); Section 7 quantifies the gap.

Assume $m \mid q - 1$, let $H \leq \mathbb{F}_q^\times$ have order m, let $\beta_0, \ldots, \beta_{\ell-1}$ represent $\ell \leq (q-1)/m$ distinct cosets, and set $A_j := \beta_j H$. Then $g(X) := X^m$ is constant on each A_j with distinct values $\gamma_j := \beta_j^m$ [33, §III-B], and $h_j := Z_{A_j} = X^m - \gamma_j$. Two algebraic properties are used below: $g(\eta) \notin \{\gamma_j\}_j$ for $\eta \notin A$, since $g - \gamma_j$ has exactly the roots A_j; and $[h_j(\tau)]G_2 = [\tau^m]G_2 - \gamma_j G_2$ is computable from a short reference string.

Definition 1 (Tamo–Barg Code with Degree Bound [33]). *For $r \le k \le \ell r$ set $D := k - 1 + (\lceil k/r \rceil - 1)\alpha$ and $V_D := \{f : \deg f \le D \wedge f \in \mathbb{F}_q[g] \cdot \langle 1, X, \ldots, X^{r-1} \rangle\}$, $C_{\mathrm{TB}} := \mathrm{ev}_A(V_D)$. Equivalently, V_D is the set of $f = \sum_{i<r} p_i(g(X))X^i$ of degree at most D, with no bound on the individual p_i.*

Theorem 1 (Dimension, Coefficient Degrees, Distance). *Write $k = ar + b$ with $0 \le b < r$. (i) The monomials $X^i g^t$ with $i < r$ have pairwise distinct degrees $i + tm$, so $V_D = \mathrm{span}\{X^i g^t : i + tm \le D\}$, $\dim V_D = k$, and every $f \in V_D$ is uniquely $f = \sum_{i<r} p_i(g)X^i$ with $\deg p_i < \kappa_i$, where $\kappa_i = a + 1$ for $i < b$ and $\kappa_i = a$ otherwise. (ii) $C_{\mathrm{TB}}|_{A_j} = \mathrm{RS}_r(A_j)$, the $[m, r, \alpha + 1]$ MDS code of evaluations of polynomials of degree less than r; equivalently, $f \bmod h_j$ has degree less than r and its evaluations on A_j are $f|_{A_j}$. (iii) $d(C_{\mathrm{TB}}) = n - D$, so (2) holds with equality.*

Corollary 1 (Characterization by Local Restrictions). *For every $r \le k \le \ell r$, $V_D = \{f : \deg f \le D \wedge \forall j,\ f|_{A_j} \in \mathrm{RS}_r(A_j)\}$.*

Definition 1 is the construction of [33, §V-C] with local codes of length $m = r + \alpha$, stated for general k through the degree bound D. Proofs are in Appendix A, where Remark 2 also shows that the degree bound of Definition 1 cannot be replaced by a bound on the individual p_i without losing optimality. All concrete figures use the *reference profile*: the BLS12-381 scalar field, admissible since $2^{32} \mid q - 1$, with $n = 125{,}952$, $\ell = 123$, $m = 1024$, $r = 768$, $\alpha = 256$, $k = 62{,}976$, $a = H_\star = 82$, $b = 0$, $D = 83{,}711$, $d = 42{,}241$, and rate $1/2$.

We use the framework of [22, Def. 1], whose numbered items we cite from its full version: a sampling scheme is *sound* if a designated set of T accepting clients forces the extractor Ext not to output $\perp$, and *consistent* if two runs of Ext under one commitment never disagree; we fix the client set in advance, and the variant in which an adversary drops transcripts follows with a factor $\binom{L}{T}$ [22, Def. 2, Lem. 1]. The compiler of [22, §6.3] turns an erasure-code commitment and an index sampler into such a scheme. We use KZG commitments [26] with extractability in the algebraic group model [19] and vanishing-polynomial batching for multipoint openings [5]; enforcing a degree bound below that of the reference string by committing to a shifted polynomial is standard [12].

3 Codewords as Tuples of Source-Group Restrictions

Proposition 1 (Source Representation). *Write $k = ar + b$ with $0 \le b < r$, so that $D = am + b - 1$ when $b > 0$ and $D = (a-1)m + r - 1$ when $b = 0$. Fix a distinct source-group indices $j_0, \ldots, j_{a-1}$, write γ_s for γ_{j_s}, let $L_0, \ldots, L_{a-1}$ be the Lagrange basis for $\gamma_0, \ldots, \gamma_{a-1}$, and set $B(T) := \prod_{s<a}(T - \gamma_s)$. Then $\Phi(\boldsymbol{u}, e) := \sum_{s<a} L_s(g)u_s + B(g)e$ is a linear bijection from $(\mathbb{F}_q[X]_{<r})^a \times \mathbb{F}_q[X]_{<b}$ onto V_D, and for every group index j,*

$$\Phi(\boldsymbol{u}, e) \bmod h_j \;=\; \textstyle\sum_{s<a} L_s(\gamma_j)\, u_s + B(\gamma_j)\, e. \tag{3}$$

In particular $\Phi(\boldsymbol{u}, e) \bmod h_{j_s} = u_s$*, and when* $b = 0$ *the term* $B(g)e$ *is absent and* $a = k/r$*.*

Remark 1 (Source Groups Do Not Determine the Residual Block). When $b > 0$, the tuple $(\boldsymbol{u}, e)$ does not admit straightforward local reconstruction: since $B(\gamma_{j_s}) = 0$ gives $f \bmod h_{j_s} = u_s$, the restrictions of the codeword to all source groups leave the b coefficients of e undetermined (as illustrated in Appendix A), so a certified source groups provide only $k - b$ message dimensions.

To resolve this issue, we adjust the message coordinates rather than altering the commitment: we designate one non-source group $j_\star$ and replace e by the low-order part of the codeword's restriction to it.

Proposition 2 (Message Coordinates). *Assume* $b > 0$ *and fix* $j_\star \notin J_{\mathrm{src}} := \{j_0, \ldots, j_{a-1}\}$*, so* $B(\gamma_{j_\star}) \neq 0$*. For* $\boldsymbol{u}$ *and* v *as above put*

$$F_\star := \textstyle\sum_{s<a} L_s(\gamma_{j_\star})\, u_s, \qquad e := \frac{v - \mathrm{Low}_{<b}(F_\star)}{B(\gamma_{j_\star})}, \qquad \Psi(\boldsymbol{u}, v) := \Phi(\boldsymbol{u}, e), \quad (4)$$

with $\mathrm{Low}_{<b}$ *the truncation to the coefficients of* $X^0, \ldots, X^{b-1}$*. Then* Ψ *is a linear bijection onto* V_D*, and for* $f = \Psi(\boldsymbol{u}, v)$*,*

$$u_s = f \bmod h_{j_s} \quad (s < a), \qquad v = \mathrm{Low}_{<b}\big(f \bmod h_{j_\star}\big). \quad (5)$$

With $J_{\mathrm{msg}} := J_{\mathrm{src}} \cup \{j_\star\}$*, of size* $H_\star := \lceil k/r \rceil = a + 1$*, every message coordinate is a coefficient of the degree-*$(< r)$ *restriction of the codeword to one group of* J_{msg}*. When* $b = 0$ *we set* $J_{\mathrm{msg}} := J_{\mathrm{src}}$ *and* $\Psi := \Phi$*.*

The proofs of Propositions 1 and 2 are in Appendix A. This representation is locally accessible in the sense of [22, §3.2] with query complexity r (Appendix E.3).

4 Locality-Aware Erasure-Code Commitments and Certification

Both Construction 2 and the variant of Appendix D provide one group commitment per group, obtained either by a public computation on com or with the help of auxiliary data dispersed with the group. The following syntax covers both, and all openings are openings under a group commitment.

Definition 2 (Locality-Aware Erasure-Code Commitment). *An LAECC adds two algorithms to an erasure-code commitment for* C_{TB} *in the sense of [22, Def. 8]. The algorithm* $\mathsf{DeriveGrp}(\mathsf{ck}, \mathsf{com}, j, \mathsf{aux}_j)$ *outputs a group commitment* C_j *or* $\perp$*. The algorithm* $\mathsf{Certify}(j, \{(a, v_a)\}_{a \in U})$*, where* $U \subseteq A_j$ *and* $|U| = r$*, interpolates* $\widehat{\delta}_j := \mathrm{Interp}_{<r}(U, \{v_a\})$ *and outputs* $\widehat{c}_j := (\widehat{\delta}_j(a))_{a \in A_j}$*. Openings are relative to a group:* Open *and* Verify *act on* $U \subseteq A_j$ *under* C_j*, and a pair* (a, v) *is an* accepting opening *if it lies in a set accepted by* Verify *under* $C_j \neq \perp$*. The auxiliary data* aux_j *is produced once by* Com *and dispersed with group* j*; it is empty in Section 5, where* $\mathsf{DeriveGrp}$ *depends only on* com *and* j*.*

Following [22, Def. 9, 10], we require two properties and add a third, specific to the local structure. Throughout, $\mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda)$, and adversaries are probabilistic polynomial time.

Definition 3 (Position and Code Binding). *An LAECC is $\varepsilon_{\mathrm{pos}}$-position binding if no adversary outputs, with probability more than $\varepsilon_{\mathrm{pos}}$, a commitment with $\mathsf{VerifyCom} = 1$ and two accepting openings (a, v), (a, v') with $v \neq v'$; it is $\varepsilon_{\mathrm{code}}$-code binding if the same holds, with $\varepsilon_{\mathrm{code}}$, for accepting openings $\{(a_i, v_i)\}_i$ at pairwise distinct positions such that no $c \in C_{\mathrm{TB}}$ satisfies $c(a_i) = v_i$ for all i.*

Definition 4 (Canonical Group Commitments). *Fix a deterministic map LCom from polynomials of degree less than r to group commitments. An LAECC is $\varepsilon_{\mathrm{can}}$-canonical if no adversary outputs, with probability more than $\varepsilon_{\mathrm{can}}$, a tuple $(\mathsf{com}, j, \mathsf{aux}_j, U, \{(v_a, \pi_a)\}_{a\in U})$ with $U \subseteq A_j$ and $|U| = r$, such that $\mathsf{VerifyCom} = 1$, $C_j := \mathsf{DeriveGrp}(\mathsf{ck}, \mathsf{com}, j, \mathsf{aux}_j) \neq \bot$, every (a, v_a, π_a) is an accepting opening, and $C_j \neq \mathsf{LCom}(\mathrm{Interp}_{<r}(U, \{v_a\}))$.*

Definition 4 is the group-local counterpart of computational uniqueness [22, Def. 23]. It does not follow from Definition 3: adding to C_j a verified commitment to a multiple of h_j leaves every opening on A_j and both binding properties unchanged, while C_j no longer equals LCom of the local polynomial (Remark 6). Code binding also implies local binding: accepting openings in one group are consistent with a common codeword of $C_{\mathrm{TB}}|_{A_j}$.

In one execution the committer publishes com; the verifier runs $\mathsf{VerifyCom}$, halts on failure, and otherwise samples positions with its own randomness; the adversary answers adaptively; and whenever r distinct accepting openings of one group have accumulated, the verifier certifies it and retains the evidence.

Definition 5 (Certified Group and Failure Events). *A group A_j is* certified *if the verifier holds r distinct accepting openings at $U \subseteq A_j$, with output $\widehat{c}_j := \mathsf{Certify}(j, \{(a, v_a)\}_{a\in U})$; J_{cert} is the certified index set, and $c^\star \in C_{\mathrm{TB}}$* completes *the execution if it agrees with every accepting opening of it. Let $\mathsf{FalseCert}$ be the event that a completion exists but some certified $\widehat{c}_j$ differs from $c^\star|_{A_j}$ for some completion, and $\mathsf{Inconsistent}$ that two certifications of one group differ; both occur by convention if no completion exists.*

Theorem 2 (Integrity of Certified Groups). *Let the LAECC be $\varepsilon_{\mathrm{pos}}$-position binding and $\varepsilon_{\mathrm{code}}$-code binding. For every adversary that outputs com and answers all queries adaptively, $\Pr[\mathsf{FalseCert} \vee \mathsf{Inconsistent}] \leq \varepsilon_{\mathrm{code}} + \varepsilon_{\mathrm{pos}}$; outside that event every completion induces the same restriction on every certified group, so each certified reconstruction is determined by the evidence retained for it.*

The proof, in Appendix A, uses only code binding and the unique recoverability of the units, so it holds for any erasure-code commitment for a code with disjoint MDS restrictions. Certification guarantees that the data a verifier holds is consistent and can be regenerated, not that it is stored persistently or that its holder remains online.

For $S \subseteq A$ let $\mathrm{cl}(S) := S \cup \bigcup\{A_j : |S \cap A_j| \geq r\}$ be the *closure* of a position set, the set of positions that an interpolation-only decoder can reconstruct directly; it adds no information, since $\mathrm{rank}(G_{\mathrm{cl}(S)}) = \mathrm{rank}(G_S)$ for a generator matrix G.

Construction 1 (LR-DAS). Encode encodes the data as $f \in V_D$, commits with an LAECC for C_{TB}, and sets $\pi_a := (f(a), \tau_a)$ for $a \in A$, with τ_a an opening under $C_{j(a)}$. A client queries the positions given by an index sampler and accepts if all openings verify. Given accepting transcripts, Ext selects for each position the first accepting opening, obtains a position set S, forms $\mathrm{cl}(S)$ by Theorem 1(ii), and, if $|\mathrm{cl}(S)| \geq D+1$, interpolates the unique f of degree at most D through $D+1$ of them, returning $\Psi^{-1}(f)$ if $f \in V_D$ and $\mathrm{ev}_A(f)$ agrees with the selected openings, and $\perp$ otherwise. Algorithm 1 in Appendix E implements Ext, and all of its recovery paths terminate at this common verification check.

Theorem 3 (Soundness and Consistency). *Let* LAECC *be* $\varepsilon_{\mathrm{pos}}$*-position binding and* $\varepsilon_{\mathrm{code}}$*-code binding, and put* $\nu^\star(T) := \Pr[|\mathrm{cl}(S_T)| \leq D]$ *for* S_T *the positions queried by* T *designated clients. Then Construction 1 satisfies* $\mathrm{Adv}^{\mathrm{sound}} \leq \nu^\star(T) + \varepsilon_{\mathrm{code}} + \varepsilon_{\mathrm{pos}}$ *and* $\mathrm{Adv}^{\mathrm{cons}} \leq \varepsilon_{\mathrm{code}} + \varepsilon_{\mathrm{pos}}$*, with* $\nu^\star(T) \leq \binom{n}{D}(D/n)^{QT}$ *for a uniform sampler of* Q *positions per client and* $\nu^\star(T) \leq \binom{\ell}{H_\star - 1}((H_\star - 1)/\ell)^T$ *for a sampler that takes one uniform group and* r *distinct positions inside it. The proof is in Appendix A.*

At the reference profile a light client uses 218 uniform queries and certifies no group, whereas a collector uses the group sampler; Lemma 1 gives tighter bounds.

5 Derived Group Commitments

Construction 2 (LAECC with Derived Group Commitments). Write $k = ar + b$. The public parameters specify $j_0, \ldots, j_{a-1}$, the basis L_s and the polynomial B of Proposition 1, and, when $b > 0$, the residual index $j_\star$ of Proposition 2.

- $\mathsf{Setup}(1^\lambda)$ runs a dedicated trusted setup giving

$$\mathsf{ck} := \big((\tau^i G_1)_{0 \leq i < r},\ (\tau^i G_2)_{0 \leq i \leq r}\big), \tag{6}$$

  with no $\mathbb{G}_1$ element above degree $r-1$ published anywhere.
- $\mathsf{Com}(\mathsf{ck}, \mathsf{m})$ parses m as $(\boldsymbol{u}, v)$ with $\deg u_s < r$ and $\deg v < b$, converts by

$$F_\star := \textstyle\sum_{s<a} L_s(\gamma_{j_\star})\, u_s, \qquad e := \big(v - \mathrm{Low}_{<b}(F_\star)\big)\, B(\gamma_{j_\star})^{-1}, \tag{7}$$

  which is (4), sets $f := \Psi(\boldsymbol{u}, v) \in V_D$, and outputs $\mathsf{com} := (U_0, \ldots, U_{a-1}, E, \bar{E})$ with $U_s := \mathsf{KZG.Com}(u_s)$, $E := \mathsf{KZG.Com}(e)$, $\bar{E} := \mathsf{KZG.Com}(X^{r-b} e)$, and $\mathsf{aux}_j := \varepsilon$; when $b = 0$ the last two are omitted.
- $\mathsf{VerifyCom}$ accepts if com has the correct shape and, when $b > 0$, $e(E, [\tau^{\,r-b}]G_2) = e(\bar{E}, G_2)$; no challenge and no opening are involved.

– $\mathsf{DeriveGrp}(\mathsf{ck}, \mathsf{com}, j, \varepsilon)$ outputs $C_j := \sum_{s<a} L_s(\gamma_j)U_s + B(\gamma_j)E$.
– $\mathsf{Open}(\mathsf{ck}, \mathsf{St}, j, U)$ for $|U| \le r$ returns $\{(a, \delta_j(a))\}_{a\in U}$ with $\delta_j := f \bmod h_j$ and $\pi_U := \mathsf{KZG.Com}((\delta_j - I_U)/Z_U)$; Verify checks $e(C_j - [I_U(\tau)]G_1, G_2) = e(\pi_U, [Z_U(\tau)]G_2)$, forming $[Z_U(\tau)]G_2$ from $(\tau^i G_2)_{i\le r}$.

We call com the header. The conversion (7) is needed for completeness, not for binding (Appendix E.4), and membership in V_D is structural. The commitment of a source group is a header element, because $L_{s'}(\gamma_{j_s}) = [s = s']$ and $B(\gamma_{j_s}) = 0$, so only the $\ell - a$ parity groups require a linear combination. When $|U| = r$, the proof is the identity element, because $\deg \delta_j < r$ implies $I_U = \delta_j$; certifying a whole group reduces to checking $\mathsf{KZG.Com}(I_U) = C_j$.

Theorem 4 (Binding and Canonicality of Construction 2). *Work in the algebraic group model, assume $r \ge 2$, and let $\varepsilon_{\mathrm{dl}}(t')$ be the best advantage of an algorithm of expected running time t' against the $(r-1, r)$-discrete-logarithm problem: given $(\tau^i G_1)_{0\le i<r}$ and $(\tau^i G_2)_{0\le i\le r}$, output τ. For an adversary running in time t, there is an event E with $\Pr[E] \le \varepsilon_{\mathrm{kzg}} := \varepsilon_{\mathrm{dl}}(t + t_{\mathrm{red}})$, where the expected time t_{red} of the reduction is polynomial in r, $\log q$, and the number of outputs of the adversary. Outside E, for every* com *with* $\mathsf{VerifyCom} = 1$ *that the adversary outputs: (a) the extracted $u_s^\star$, $e^\star$ satisfy $\deg u_s^\star < r$ and $\deg e^\star < b$, so $f^\star := \Phi(\boldsymbol{u}^\star, e^\star) \in V_D$ and $c^\star := \mathrm{ev}_A(f^\star) \in C_{\mathrm{TB}}$; (b) every accepting opening (a, v) has $v = c^\star(a)$; and (c) for every j, $\mathsf{DeriveGrp}(\mathsf{ck}, \mathsf{com}, j, \varepsilon) = \mathsf{KZG.Com}(f^\star \bmod h_j)$. Hence the scheme is position binding, code binding, and canonical with the single error $\varepsilon_{\mathrm{kzg}}$, and Theorem 2 applies with that quantity.*

Proof idea. Since the adversary is algebraic and ck publishes no $\mathbb{G}_1$ element of degree above $r-1$, its $\mathbb{G}_1$ outputs yield $u_s^\star$, $e^\star$, and $\bar{e}^\star$ of degree less than r; when $b > 0$, the degree equation gives $X^{r-b}e^\star = \bar{e}^\star$, which forces $\deg e^\star < b$. Proposition 1 then places $f^\star$ in V_D, which is (a). An accepting opening on $U \subseteq A_j$ asserts $P(\tau) - I_U(\tau) = \Pi_U(\tau)Z_U(\tau)$; if $R := P - I_U - Z_U\Pi_U$, of degree at most $2r-1$, is nonzero, then the reduction finds τ among its roots by factoring R and comparing θG_1 with τG_1, and otherwise the opened values are those of P. Since $\mathsf{DeriveGrp}$ and extraction are linear, the polynomial extracted from C_j is $f^\star \bmod h_j$ by (3), which is (c), and (b) follows from Theorem 1(ii). Appendix B gives the complete proof.

The argument relies on the fact that no $\mathbb{G}_1$ element $\tau^i G_1$ with $i \ge r$ is available to the adversary, and this degree enforcement is necessary:

Proposition 3 (Degree Enforcement Is Necessary). *Suppose the adversary holds $\tau^r G_1$ in addition to the* ck *of* (6)*, for instance because* ck *was obtained by truncating a longer string generated with the same τ. Then Construction 2 is not code binding.*

Appendix B proves Propositions 3 and 4 and details three deployment profiles. Under a setup dedicated to degree r (Profile A), Construction 2 applies as stated. Under a longer string that also publishes the $\mathbb{G}_2$ elements needed for shifted degree checks (Profile B), a further header elements and one batched check with a

verifier-sampled challenge restore the conclusions (Proposition 9), and the header doubles to 7,872 bytes. Under a longer string without those elements (Profile C), the construction cannot be securely instantiated. Profile A does not always require a new setup: the 2^{12} transcript of the Ethereum KZG ceremony has its own secret, $\mathbb{G}_1$ powers up to degree 4095, and $\mathbb{G}_2$ powers up to degree 64 [16], and for $r = 4096$ and $b = 0$ Corollary 4 (Appendix B) preserves Theorem 4 and local serving under this setup, with multipoint openings of at most 64 positions and whole groups certified by $\mathsf{KZG.Com}(I_U) = C_j$. This does not extend to $r = 768$: truncating that string leaves higher powers of the same τ public (Proposition 3).

Proposition 4 (Local Serving). *Let a node hold* com*, the reference-string powers* $(\tau^i G_1)_{0 \le i < r}$ *from* ck*, and the values at* r *distinct positions of group* j*, each carried by an accepting opening. Outside the event* E *of Theorem 4, and after the committer has become unavailable, the node can derive* C_j*, interpolate* $\widehat{\delta}_j$*, and produce for every* $U \subseteq A_j$ *with* $|U| \le r$ *a multipoint opening that* Verify *accepts. It requires no per-group auxiliary data, no reference-string element of index at least* r*, and no interaction.*

Under direct proof generation, without stored proofs or other proving state, the node maintains 24 KiB of values per group and forms each proof by a multiscalar multiplication of at most $r - 1 = 767$ terms over 36 KiB of compressed $\mathbb{G}_1$ powers; in the one-dimensional design of Table 1, direct generation of a single-point proof requires all k values (1.92 MiB) and up to $k - 1$ terms over 2.88 MiB of such powers. A node of that design may instead store the proofs of the positions it serves (48 KiB per group) or compute many proofs at once from the full polynomial [18]; in contrast, LR-DAS enables regeneration relying solely on the authenticated values of a group. For supplied group commitments, Proposition 4 fails (Remark 5).

6 Recovery from Certified Groups

Theorem 5 (Recovery from the Closure). *Let* $c = \mathrm{ev}_A(f)$ *with* $f \in V_D$ *and let* $S \subseteq A$ *be known. If* $|\mathrm{cl}(S)| \ge D + 1$ *then* f *is determined, by interpolation through any* $D + 1$ *positions of* $\mathrm{cl}(S)$*. Suppose further that* S *has exactly* r *positions in each of* h *groups, indexed by* H*, plus a set* S_{out} *disjoint from* $\bigcup_{j \in H} A_j$ *with* $|S_{\mathrm{out}} \cap A_j| < r$ *for* $j \notin H$*. Then* $|\mathrm{cl}(S)| = hm + |S_{\mathrm{out}}|$*, so the criterion reads* $|S_{\mathrm{out}}| \ge R(h) := \max(0, D + 1 - hm)$*; such an* S *exists precisely when* $R(h) \le (\ell - h)(r - 1)$*, and the smallest number of received values that meets the criterion is* $C(h) := hr + R(h)$*, equal to* $D + 1 - h\alpha$ *for* $h \le H_\star - 1$ *and to* hr *for* $h \ge H_\star$*. In particular* $C(H_\star - 1) = k$*, the minimum of* C*, and* $\lceil (D+1)/m \rceil = H_\star$*.*

Corollary 2 (Any $H_\star$ Groups Suffice). *If* $c|_{A_j}$ *is known for* $H_\star$ *indices, then* f *is determined, using* $rH_\star < k + r$ *symbols and only interpolation. If those indices are exactly* J_{msg}*, no global interpolation of* f *is required: the message is directly recovered blockwise by* (5) *from the degree-*$(< r)$ *local polynomials that certification has already interpolated. If all a source groups are known and* $b > 0$*,*

the residual e is determined by any b further known positions outside $\bigcup_{s<a} A_{j_s}$, which need not form a group, through

$$e(x) = \big(f(x) - \textstyle\sum_{s<a} L_s(g(x))\, u_s(x)\big) \big/ B(g(x)). \tag{8}$$

Proofs are in Appendix A, with Remark 3 (the criterion is sufficient, not necessary) and Proposition 8 (the closure almost never helps a uniform sampler).

For fixed n, ℓ, m, and k, increasing $r \in [k/\ell, m)$ improves the global distance at the cost of higher per-group recovery overhead, while reducing α (Table 3 in Appendix C); at $r = k/\ell$ the code is the direct sum of its local codes. We adopt $r = 768$, so $b = 0$ and $a = 82$.

7 Comparison with the Source-Compressed Product Code

The existing design with disjoint, independently committed recovery units is the tensor-code erasure-code commitment of [22, §8]. With an outer RS[123, 82] and KZG column commitments for the inner RS[1024, 768], it has 123 columns, each recoverable from 768 values, at our n and k and with the same need for degree enforcement. As published, it commits to every column; since the column polynomials form an outer Reed–Solomon codeword, publishing only the $a = 82$ source columns and deriving the others, as in [7], gives the baseline a header of 3,936 bytes, exactly ours, and the same serving procedure. We compare against this source-compressed form, since the uncompressed form would credit our construction with a header advantage that is due only to the encoding of the baseline.

Both constructions belong to a unified class of designs: fix free source polynomials u_s of degree less than r and a rank-a matrix $M \in \mathbb{F}_q^{\ell \times a}$, and let group j carry the evaluations of $\delta_j := \sum_{s<a} M_{j,s} u_s$. Every such scheme publishes a commitments, derives each group commitment as a public linear combination of them, opens a coordinate with one KZG proof, and lets a node holding r symbols of a group interpolate and resume serving. The remaining degree of freedom is the choice of evaluation points for δ_j, which solely determines the achievable minimum distance.

Proposition 5 (Distance Ceiling for a Common Evaluation Set). *If every δ_j is evaluated on one common set B of m points, the resulting code is $\mathrm{Im}(M) \otimes \mathrm{RS}_r(B)$, of dimension $ar = k$ and minimum distance*

$$d\big(\mathrm{Im}(M)\big)\,(m - r + 1) \;\le\; (\ell - a + 1)(m - r + 1) \;=\; d_{\mathrm{prod}}.$$

No choice of M exceeds the plain product-code distance within this common-evaluation subclass. Keeping M but replacing the common set by the group-dependent sets $B_j = \beta_j H$ departs from that subclass while remaining within the source-linear class, and with $M_{j,s} = L_s(\gamma_j)$ it attains $(\ell - a + 1)m - r + 1$, larger by $(\ell - a)(r - 1)$. If in addition M has no zero row, every group carries an $[m, r]$ MDS restriction, which is the locality required in Section 2; the distance bound itself holds for any M of rank a.

The proof is in Appendix A. On a common set, zeros are counted column by column, so the component distances multiply to $42 \cdot 257 = 10{,}794$ here; on the cosets, the codeword is the evaluation of one polynomial of degree at most D, giving $n - D = 42{,}241$. This proposition establishes an upper bound specific to this subclass, rather than an impossibility result for other commitments or codes.

Corollary 3 (Distance Gap, and Its Dependence on the Local Rate). *Let $r \mid k$ with $k = ar$ and $r > 1$, and compare C_{TB} with the product of* $\mathrm{RS}[\ell, a]$ *and* $\mathrm{RS}[m, r]$ *on the same $n = \ell m$, k, group size m, and per-group recovery cost r. Then $d_{\mathrm{TB}} = (\ell - a + 1)m - r + 1$ and $d_{\mathrm{prod}} = (\ell - a + 1)(m - r + 1)$, so that*

$$d_{\mathrm{TB}} - d_{\mathrm{prod}} = (\ell - a)(r - 1), \qquad \frac{d_{\mathrm{TB}}}{d_{\mathrm{prod}}} = \frac{m}{\alpha + 1}\Big(1 - \frac{r - 1}{m(\ell - a + 1)}\Big) < \frac{m}{\alpha + 1}. \quad (9)$$

The gap grows with the outer distance $\ell - a + 1$ and with the per-group recovery cost r, and it vanishes exactly at $a = \ell$, where both codes are the direct sum of their local codes. The ratio in (9) *is bounded by $m/(\alpha + 1) = (1 - r/m + 1/m)^{-1}$, which at fixed m depends only on the local rate r/m, and deviates from that bound by exactly $(r - 1)/((\alpha + 1)(\ell - a + 1))$.*

Table 4 in Appendix C evaluates (9) at other shapes.

Proposition 6 (Cost of the Evaluation Step). *Fix $r \mid k$, let $\mathbf{d}_j$ be the coefficient vector of δ_j, which is the same for the two constructions once the same source-to-group combination is used, and let V_H evaluate $1, X, \ldots, X^{r-1}$ on H. Then the product code encodes group j as $V_H \mathbf{d}_j$ and Construction 2 as $V_H \operatorname{diag}(1, \beta_j, \ldots, \beta_j^{r-1})\, \mathbf{d}_j$, so the two differ by one diagonal scaling per group, costing at most $\ell(2r - 3)$ field multiplications, or $\ell(r - 1)$ once the public powers β_j^i are tabulated.*

At the reference profile the scaling adds at most 94,341 field multiplications to identical inner transforms. Grouping the queries of a light client into one multipoint opening per touched group benefits both designs but reduces the relative advantage shown in Table 1, since the product code amortizes each proof across more queries: about 15,820 bytes against 41,550. Table 6 in Appendix C gives the ratios for larger objects.

Proposition 7 (An Erasure Pattern Precluding Local Completion while Permitting Row Repair). *Assume $m \geq \ell$ and $r \mid k$ with $k = ar$. Index the positions of group j by $t \in \mathbb{Z}_m$ through a fixed bijection and erase $W_j := \{\, t : (t - \lfloor mj/\ell \rfloor) \bmod m \leq \alpha \,\}$ in every group. Then (i) every group loses exactly $\alpha + 1$ symbols, so no group of C_{TB} is locally recoverable, $\mathrm{cl}(S) = S$ for the surviving set S, and the decoder of Construction 1 must interpolate f from $D + 1$ symbols; (ii) every row of the product code loses at most $\lceil \ell(\alpha + 1)/m \rceil$ symbols, so if that is at most $\ell - a$ then every row stays recoverable from a symbols along the other axis; and (iii) $|W| = \ell(\alpha + 1) < d_{\mathrm{TB}}$ if and only if $(\ell - 1)(r - 1) > m(a - 1)$, so under that condition the data remains recoverable under C_{TB}, but not locally.*

The proof is in Appendix A. At the reference profile the conditions hold: every product row loses at most $31 \le 41$ values and $|W| = 31{,}611 < d_{\mathrm{TB}}$, so the product code permits row-wise recovery, while the interpolation decoder of Construction 1 falls back to requiring $D+1$ values.

This requirement of $D+1$ values is an artifact of the decoder rather than an intrinsic limitation of C_{TB}. For a group with $s_j < r$ surviving positions, write $\delta_j = I_j + Z_j w_j$, where I_j interpolates the surviving values, Z_j vanishes on them, and $\deg w_j < r - s_j$. For $b = 0$, Proposition 1 turns $\delta_j = \sum_s L_s(\gamma_j)\delta_{j_s}$, $j \notin J_{\mathrm{src}}$, into $(\ell - a)r$ linear equations in the $\sum_j \max(0, r - s_j)$ unknown coefficients, and the codeword is determined exactly when this system has a unique solution (Proposition 10 in Appendix E.2). Under the pattern of Proposition 7, every group has $s_j = r - 1$, so the unknowns are $\ell = 123$ scalars; at the reference profile, the equations for the three leading coefficients of the local polynomials already have rank 123 over the BLS12-381 scalar field. In the tested random instances (Table 8), the decoder recovers both codes, also from sets with fewer than $D+1$ positions; we therefore claim the advantage of C_{TB} only for adversarial withholding, where the minimum distance is decisive.

We do not claim global optimality: the comparison fixes free low-degree source blocks, scalar derivation, and the per-group recovery cost. Subcodes of product codes with larger distance have recently been constructed by adding global parity constraints [4]; at matched dimension they leave the comparison intact, since $\dim(\mathrm{RS}[\ell, a] \otimes \mathrm{RS}[m, r]) = k$ forces a subcode of dimension k to be the product code itself.

8 Parameters, Costs, and Evaluation

Combining Theorems 4, 2, and 3, Construction 2 fails with probability at most $\varepsilon_{\mathrm{kzg}} + \nu^\star(T)$. The first term is computational: it depends on the curve, on the powers of τ that are public, on the running time of the adversary, and on the cost of the reduction, and we evaluate it separately from the sampling target (Appendix B discusses concrete estimates, under which BLS12-381 provides fewer than 128 bits). The second term is statistical, and we require $\nu^\star(T) \le 2^{-130}$. Neither term involves a random oracle. The following lemma, proved in Appendix A, tightens the bound of Theorem 3.

Lemma 1 (Client Counts Without the Union Bound). *For the uniform sampler with Q queries per client and T clients, put $p := (1 - 1/n)^{QT}$. If $(n - D)/n > p$, then $\nu^\star(T) \le \exp(-n\, D_{\mathrm{KL}}(\frac{n-D}{n} \| p))$.*

Requiring $\nu^\star(T) \le 2^{-130}$ gives, by Lemma 1, $T = 663$ clients at $Q = 218$ for the uniform sampler and, by evaluating the occupancy distribution in exact integer arithmetic, $T = 398$ at $Q = r$ for the group sampler; the union-bound forms of Theorem 3 require 904 and 399 (Appendix A). The group sampler needs fewer clients but $398 \cdot 768 = 305{,}664$ queries against $663 \cdot 218 = 144{,}534$. These counts rest on the cardinality criterion of Construction 1 and describe the extraction guarantee of LR-DAS alone; they are not a comparison with the

product code, whose recovery from random position sets depends on its decoder (Section 7).

Table 5 in Appendix C collects the costs. A light client drawing 218 uniform samples queries approximately 34.08 parity groups in expectation, and deriving their commitments requires roughly 2,795 scalar multiplications; a batched check eliminates this derivation overhead. Following the batched verification of KZG openings in PeerDAS [35, §4.2], a client that holds single-point openings (x_i, v_i, π_i) in groups j_i, $i < N$, samples $\rho_0, \ldots, \rho_{N-1}$ uniformly from $\mathbb{F}_q$ after receiving them and, for $b = 0$, checks

$$\begin{aligned} e\Big(\textstyle\sum_{s<a} c_s U_s - \big(\sum_i \rho_i v_i\big) G_1 + \sum_i \rho_i x_i \pi_i,\ G_2\Big) &= e\Big(\textstyle\sum_i \rho_i \pi_i,\ [\tau] G_2\Big), \\ c_s &:= \textstyle\sum_i \rho_i L_s(\gamma_{j_i}); \end{aligned} \tag{10}$$

when $b > 0$, the term $\big(\sum_i \rho_i B(\gamma_{j_i})\big)E$ is added on the left. Each individual check reads $e(C_{j_i} - v_i G_1 + x_i \pi_i, G_2) = e(\pi_i, [\tau]G_2)$, and every C_j is a fixed linear combination of the header, so $\sum_i \rho_i C_{j_i}$ folds into $\sum_s c_s U_s$. If some individual check fails, (10) holds with probability at most $1/q$, since the ratio of its two sides is a product of powers $\Delta_i^{\rho_i}$ in the prime-order target group with $\Delta_i \neq 1$ for every failing check. The check computes two elements of $\mathbb{G}_1$ by multiscalar multiplications of $a + N + 1$ and N terms, $a + 2N + 1$ in total, and one product of two pairings: 519 terms for $N = 218$, against 2,065 for the product code at $N = 991$, which is batched analogously. Over M batched checks, Construction 2 under Profile A thus fails with probability at most $\varepsilon_{\mathrm{kzg}} + \nu^\star(T) + M/q$, whose statistical part stays below 2^{-128} for $M \leq 2^{120}$; under Profile B, Proposition 9 replaces $\varepsilon_{\mathrm{kzg}}$ by $\varepsilon_{\mathrm{dl}}^{(L)}(t + t_{\mathrm{red}})$ and adds $(a-1)/q$, or a/q if $b > 0$, per execution of VerifyCom.

Appendix E.4 tests the specification for general k over small fields on 5,791 parameter sets, and Appendix F reports a collector policy that matches its closed-form query counts; the scripts accompany the submission as supplementary material.

8.1 Concrete Runtime Evaluation

Implementation and measurement conditions. We implemented LR-DAS and the source-compressed product code using arkworks 0.4 in Rust over BLS12-381. Timings were obtained from three independent single-threaded runs on an Intel Core i9-14900K, excluding setup, input preparation, and network communication; Appendix C reports Profile B and other group sizes.

Commitment and verification. Commitment generation costs the same for both designs. The measured client verifier does not use (10); starting without verified commitment state, it parses and validates the commitment, derives group commitments, and checks every sampled opening. With 218 queries against 991, it takes 74.9% less time for LR-DAS under Profile A and 74.3% less under Profile B, where shifted commitments double the commitment size; these figures apply to

Table 2. Measured costs at $(\ell, m, r, a, b) = (123, 1024, 768, 82, 0)$ under Profile A, and the analytic size of the batched check (10). Client verification covers 218 queries for LR-DAS and 991 for the product code, the counts that reach the same detection target.

	LR-DAS	Source-compressed product code
Commitment generation	864.6 ms	863.8 ms
Unbatched client verification	414.3 ms	1,650.0 ms
Full reconstruction, pattern of Proposition 7	6,247.1 ms	552.9 ms
Batched check (10): MSM terms, pairings	519, 2	2,065, 2

this unbatched verifier only. Under (10) both designs use two pairings, and the ratio depends on the multiscalar multiplications, deserialization, and subgroup checks, which we have not measured.

Full reconstruction under the pattern of Proposition 7. With $(\ell, a, b) = (123, 82, 0)$ and $r = 3m/4$, the implemented LR-DAS decoder interpolates globally from $D + 1$ values, whereas the product code recovers row by row; starting from authenticated values, LR-DAS requires 10.1 to 12.7 times more execution time for $m \in \{512, 1024, 2048\}$. The ratio concerns the implemented decoder: the structured decoder of Section 7 reduces the pattern to 123 local interpolations and a system in 123 unknowns, which a reference-scale run on random data solves exactly (Appendix E.2), and if $\beta_j = \omega^j$ for a primitive 2^{17}th root of unity ω, then A consists of 123 of the 128 cosets of H in the group of 2^{17}th roots of unity, so the FFT-based erasure recovery of PeerDAS [35, §4.3] applies to the global branch. Neither approach has been empirically benchmarked; hence, we do not claim any practical reduction of this gap.

9 Conclusion

LR-DAS combines disjoint, locally certifiable recovery groups with the optimal Singleton-type minimum distance by committing to the source-group restrictions of a Tamo–Barg codeword. It keeps the header, reference string, and local serving of the source-compressed product code, and evaluating each group on its own coset multiplies the minimum distance by 3.91 at the reference profile.

Three limitations remain. The construction requires degree enforcement (Proposition 3), provided by a dedicated string of degree r, by the 2^{12} transcript of an existing ceremony for $r = 4096$ (Corollary 4), or by the checks of Profile B. Each symbol has a single recovery set, which increases the reconstruction overhead of the interpolation decoder under specific erasure patterns (Proposition 7). The batched verification check, the structured decoder, and the time to resume serving a group remain to be empirically evaluated. Further directions include a proof without the algebraic group model along the lines of [36], a header of constant size, and a transparent instantiation.

References

1. Al-Bassam, M., Sonnino, A., Buterin, V., Khoffi, I.: Fraud and data availability proofs: Detecting invalid blocks in light clients. In: Borisov, N., Díaz, C. (eds.) Financial Cryptography and Data Security (FC 2021). Lecture Notes in Computer Science, vol. 12675, pp. 279–298. Springer (2021). https://doi.org/10.1007/978-3-662-64331-0_15
2. Barbulescu, R., Duquesne, S.: Updating key size estimations for pairings. Journal of Cryptology **32**(4), 1298–1336 (2019). https://doi.org/10.1007/s00145-018-9280-5, https://eprint.iacr.org/2017/334, full version: Cryptology ePrint Archive, Report 2017/334
3. Ben-Sasson, E., Bentov, I., Horesh, Y., Riabzev, M.: Fast reed–solomon interactive oracle proofs of proximity. In: International Colloquium on Automata, Languages, and Programming (ICALP 2018). LIPIcs, vol. 107, pp. 14:1–14:17 (2018). https://doi.org/10.4230/LIPIcs.ICALP.2018.14, https://doi.org/10.4230/LIPIcs.ICALP.2018.14
4. Berman, A., Shany, Y., Tamo, I.: Codes with large minimum distance in product codes: Explicit constructions and bounds (2026), https://arxiv.org/abs/2604.15080, arXiv:2604.15080
5. Boneh, D., Drake, J., Fisch, B., Gabizon, A.: Efficient polynomial commitment schemes for multiple points and polynomials (2020), https://eprint.iacr.org/2020/081, cryptology ePrint Archive, Report 2020/081
6. Boneh, D., Neu, J., Nikolaenko, V., Partap, A.: Data availability sampling with repair (2025), https://eprint.iacr.org/2025/1414, cryptology ePrint Archive, Report 2025/1414
7. Buterin, V.: 2d data availability with kate commitments. Ethereum Research Blog (2020), https://ethresear.ch/t/2d-data-availability-with-kate-commitments/8081
8. Buterin, V., Feist, D., Loerakker, D., Kadianakis, G., Garnett, M., Taiwo, M., Dietrichs, A.: EIP-4844: Shard blob transactions. Ethereum Improvement Proposal (2022), https://eips.ethereum.org/EIPS/eip-4844
9. Cai, H., Miao, Y., Schwartz, M., Tang, X.: Repairing schemes for Tamo–Barg codes (2023), https://arxiv.org/abs/2312.12803, arXiv:2312.12803
10. Chaudhuri, A., Basak, S., Kiraly, C., Ryajov, D., Bautista-Gomez, L.: On the design of Ethereum's data availability sampling: A comprehensive simulation study. arXiv preprint arXiv:2407.18085 (2024), https://arxiv.org/abs/2407.18085
11. Cheon, J.H.: Security analysis of the strong Diffie–Hellman problem. In: Vaudenay, S. (ed.) Advances in Cryptology – EUROCRYPT 2006. Lecture Notes in Computer Science, vol. 4004, pp. 1–11. Springer (2006). https://doi.org/10.1007/11761679_1
12. Chiesa, A., Hu, Y., Maller, M., Mishra, P., Vesely, N., Ward, N.: Marlin: Preprocessing zkSNARKs with universal and updatable SRS. In: Advances in Cryptology – EUROCRYPT 2020. Lecture Notes in Computer Science, vol. 12105, pp. 738–768. Springer (2020)
13. Cho, S.H., Kim, Y.S.: Exploring the applicability of locally recoverable codes in data availability sampling schemes. In: Proc. Summer Conference of the Korean Institute of Communications and Information Sciences (KICS) (2025), in Korean
14. Cho, S., Seo, E., Kim, Y.S.: Locally recoverable data availability sampling (2025), https://eprint.iacr.org/2025/1851, cryptology ePrint Archive, Paper 2025/1851

15. Dubhashi, D., Ranjan, D.: Balls and bins: A study in negative dependence. Random Structures and Algorithms **13**(2), 99–124 (1998)
16. Ethereum: Powers of tau specification for the KZG ceremony. GitHub repository (2023), https://github.com/ethereum/kzg-ceremony-specs
17. Evans, A., Mohnblatt, N., Angeris, G.: ZODA: Zero-overhead data availability (2025), https://eprint.iacr.org/2025/034, cryptology ePrint Archive, Report 2025/034
18. Feist, D., Khovratovich, D.: Fast amortized KZG proofs (2023), https://eprint.iacr.org/2023/033, cryptology ePrint Archive, Report 2023/033
19. Fuchsbauer, G., Kiltz, E., Loss, J.: The algebraic group model and its applications. In: Advances in Cryptology – CRYPTO 2018. Lecture Notes in Computer Science, vol. 10992, pp. 33–62. Springer, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-96881-0_2
20. von zur Gathen, J., Gerhard, J.: Modern Computer Algebra. Cambridge University Press, 3 edn. (2013)
21. Hall-Andersen, M., Simkin, M., Wagner, B.: FRIDA: Data availability sampling from FRI. In: Reyzin, L., Stebila, D. (eds.) Advances in Cryptology – CRYPTO 2024. Lecture Notes in Computer Science, vol. 14925, pp. 289–324. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-68391-6_9
22. Hall-Andersen, M., Simkin, M., Wagner, B.: Foundations of data availability sampling. IACR Communications in Cryptology **1**(4), 34 (2025). https://doi.org/10.62056/a09qudhdj, https://eprint.iacr.org/2023/1079, full version: Cryptology ePrint Archive, Report 2023/1079
23. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association **58**(301), 13–30 (1963)
24. Kadhe, S., Chung, J., Ramchandran, K.: SeF: A secure fountain architecture for slashing storage costs in blockchains (2019), https://arxiv.org/abs/1906.12140, arXiv:1906.12140
25. Kamath, G.M., Prakash, N., Lalitha, V., Kumar, P.V.: Codes with local regeneration. In: IEEE International Symposium on Information Theory (ISIT 2013). pp. 1606–1610 (2013)
26. Kate, A., Zaverucha, G.M., Goldberg, I.: Constant-size commitments to polynomials and their applications. In: Advances in Cryptology – ASIACRYPT 2010. Lecture Notes in Computer Science, vol. 6477, pp. 177–194. Springer (2010). https://doi.org/10.1007/978-3-642-17373-8_11
27. Mohnblatt, N., Wagner, B.: Extending FRIDA beyond unique decoding for free (2026), https://eprint.iacr.org/2026/1055, cryptology ePrint Archive, Report 2026/1055
28. Papamanthou, C., Shi, E., Tamassia, R.: Signatures of correct computation. In: Theory of Cryptography (TCC 2013). Lecture Notes in Computer Science, vol. 7785, pp. 222–242. Springer (2013)
29. Prakash, N., Kamath, G.M., Lalitha, V., Kumar, P.V.: Optimal linear codes with a local-error-correction property. In: IEEE International Symposium on Information Theory (ISIT 2012). pp. 2776–2780 (2012). https://doi.org/10.1109/ISIT.2012.6284028, full version: arXiv:1202.2414
30. Sasidharan, B., Viterbo, E., Dau, S.H.: Block circulant codes with application to decentralized systems. IEEE Transactions on Communications **73**(10), 8542–8557 (2025), full version: arXiv:2406.12160
31. Sasidharan, B., Viterbo, E., Feist, D.: Block circulant codes for Ethereum PeerDAS (2026), https://eprint.iacr.org/2026/708, cryptology ePrint Archive, Paper 2026/708

32. Sforzin, A., Maso, M., Soriente, C., Karame, G.: On the storage overhead of proof-of-work blockchains (2022), https://arxiv.org/abs/2205.04108, arXiv:2205.04108
33. Tamo, I., Barg, A.: A family of optimal locally recoverable codes. IEEE Transactions on Information Theory **60**(8), 4661–4676 (2014). https://doi.org/10.1109/TIT.2014.2321280, https://arxiv.org/abs/1311.3284, full version: arXiv:1311.3284
34. Tonkikh, A., Arun, B., Xiang, Z., Li, Z., Spiegelman, A.: Raptr: Prefix consensus for robust high-performance BFT (2025), https://arxiv.org/abs/2504.18649v2, arXiv:2504.18649, version 2
35. Wagner, B., Zapico, A.: A documentation of ethereum's PeerDAS (2024), https://eprint.iacr.org/2024/1362, cryptology ePrint Archive, Paper 2024/1362
36. Wagner, B., Zapico, A.: Proving the security of PeerDAS without the AGM (2025), https://eprint.iacr.org/2025/1683, cryptology ePrint Archive, Report 2025/1683
37. Yang, M., Zhang, X., Deng, Y.: Deep-FRIDA: Transparent data availability sampling with smaller size. Cybersecurity **9**(1), 145 (2026). https://doi.org/10.1186/s42400-026-00564-9

A Proofs of the Coding-Theoretic Results

This appendix proves the coding-theoretic and probabilistic statements of Sections 2 to 4 and 6 to 8.

Proof (of Theorem 1). (i) If $i + tm = i' + t'm$ with $0 \le i, i' < r \le m$, then $i \equiv i' \pmod m$ forces $i = i'$ and $t = t'$. Polynomials of distinct degrees are independent and $\deg \sum c_{i,t} X^i g^t = \max\{i + tm : c_{i,t} \neq 0\}$, so $\deg f \le D$ holds if and only if $c_{i,t} = 0$ for $i + tm > D$. The number of admissible t at index i is $\kappa_i = \lfloor (D-i)/m \rfloor + 1$. If $b > 0$, then $\lceil k/r \rceil = a+1$ and $D = am + b - 1$; for $i < b$, the inequalities $0 \le b - 1 - i \le m - 1$ give $\kappa_i = a + 1$, and for $b \le i \le r - 1$, the inequalities $am - (r - b) \le D - i \le am - 1$ give $\kappa_i = a$. If $b = 0$, then $\lceil k/r \rceil = a$ and $D = (a-1)m + r - 1$, and $0 \le r - 1 - i \le m - 1$ gives $\kappa_i = a$ throughout. In both cases $\sum_i \kappa_i = ar + b = k$, and evaluation is injective on V_D since $D < n$.

(ii) For $a \in A_j$, $f(a) = \sum_{i<r} p_i(\gamma_j) a^i$, an evaluation of a polynomial of degree less than r; since h_j vanishes exactly on A_j and $\deg h_j = m > r$, that polynomial is $f \bmod h_j$. Conversely $k \ge r$ gives $\kappa_i \ge 1$, so each p_i may be an arbitrary constant and $(p_i(\gamma_j))_{i<r}$ ranges over $\mathbb{F}_q^r$. As $|A_j| = m$ with distinct points, $\mathrm{RS}_r(A_j)$ is $[m, r, \alpha + 1]$ and MDS.

(iii) A nonzero $f \in V_D$ has at most D roots, so $d \ge n - D$; part (ii) supplies the locality required by (2), so $d \le n - k + 1 - (\lceil k/r \rceil - 1)\alpha$. If $b > 0$, then $n - D = n - k + 1 - a\alpha$ with $\lceil k/r \rceil - 1 = a$; if $b = 0$, then $n - D = n - k + 1 - (a-1)\alpha$ with $\lceil k/r \rceil - 1 = a - 1$. □

Proof (of Corollary 1). Left to right is Theorem 1(ii). Conversely, let $\deg f \le D < \ell m$ and write $f = \sum_{i<m} p_i(g) X^i$, the base-$m$ decomposition of exponents, so that $\deg p_i < \ell$. On A_j the functions $a \mapsto a^i$, $i < m$, are independent because the associated $m \times m$ Vandermonde matrix has distinct nodes; hence $f|_{A_j} \in \mathrm{RS}_r(A_j)$ if and only if $p_i(\gamma_j) = 0$ for $r \le i < m$. Imposing this for all j gives each such p_i at least ℓ roots while $\deg p_i < \ell$, so $p_i \equiv 0$. □

Remark 2 (The Degree Bound Is Part of the Definition). Requiring only $\deg p_i < \kappa_i$ for some allocation with $\kappa_i \in \{\lfloor k/r \rfloor, \lceil k/r \rceil\}$ and $\sum_i \kappa_i = k$ does not determine the code, and such an allocation need not attain (2). Over $\mathbb{F}_{17}$ with $m = 4$, $\ell = 4$, $r = 3$, and $k = 4$, the allocation $(1, 2, 1)$ yields the span $\{1, X, X^2, X^5\}$, whose minimum distance is 11 against the bound's 12: the polynomial $(X-1)(X-2)(X-7)(X-8)(X-16) = X^5 + 7X^2 + 16X + 10$ has vanishing X^3 and X^4 coefficients, hence lies in that span, and has exactly five roots in $\mathbb{F}_{17}^{\times}$. The allocation $(2, 1, 1)$ of Theorem 1(i) attains the bound.

Proof (of Proposition 1). Every summand lies in $\mathbb{F}_q[g] \cdot \langle 1, \ldots, X^{r-1} \rangle$. For the degrees, $\deg(L_s(g)u_s) \le (a-1)m + r - 1$ and $\deg(B(g)e) \le am + b - 1$; when $b > 0$ the second equals D and the first is at most D because $r \le m + b$, and when $b = 0$ the second term is absent and the first equals D. Hence Φ maps into V_D. Fix j and put $\delta := \sum_s L_s(\gamma_j)u_s + B(\gamma_j)e$, of degree less than r since $b \le r$. For $a' \in A_j$ we have $g(a') = \gamma_j$, hence $\Phi(\boldsymbol{u}, e)(a') = \delta(a')$; as $\deg \delta < r < m = |A_j|$, Theorem 1(ii) identifies δ with $\Phi(\boldsymbol{u}, e) \bmod h_j$, which is (3). At $j = j_s$ we obtain $L_{s'}(\gamma_s) = [s = s']$ and $B(\gamma_s) = 0$, so $\Phi(\boldsymbol{u}, e) \bmod h_{j_s} = u_s$. For injectivity, suppose $\Phi(\boldsymbol{u}, e) = 0$; reducing modulo h_{j_s} gives $u_s = 0$ for every s, so $B(g)e = 0$, and as $B(g)$ is nonzero this forces $e = 0$. The domain has dimension $ar + b = k = \dim V_D$, so Φ is bijective. □

Proof (of Proposition 2). By Proposition 1, $f \bmod h_{j_\star} = F_\star + B(\gamma_{j_\star})e$. Since $\deg e < b$, the polynomial $B(\gamma_{j_\star})e$ equals its own low-order part, so by linearity of $\mathrm{Low}_{<b}$ we obtain $\mathrm{Low}_{<b}(f \bmod h_{j_\star}) = \mathrm{Low}_{<b}(F_\star) + B(\gamma_{j_\star})e = v$ by (4); the first half of (5) is Proposition 1. The map $(\boldsymbol{u}, v) \mapsto (\boldsymbol{u}, e)$ is linear, and for fixed $\boldsymbol{u}$ it is a bijection in the second argument because $B(\gamma_{j_\star}) \neq 0$; composing with the bijection Φ gives the claim. □

An example for Remark 1. Over $\mathbb{F}_{17}$ with $m = 4$, $r = 3$, and $k = 4$, so that $a = b = 1$, take $\gamma_0 = 1$. The codewords $f = u_0$ and $f = u_0 + (X^4 - 1)$ agree on all of $A_0 = \{1, 4, 13, 16\}$, so the restriction of the codeword to the source group does not determine e.

Lemma 2 (Restriction Equals Divisibility). *For $f, \delta \in \mathbb{F}_q[X]$ we have $f|_{A_j} = \delta|_{A_j}$ if and only if $f - \delta = h_j q$ for some q. If moreover $\deg \delta < r < m$, then $\delta = f \bmod h_j$.*

Proof. The difference $f - \delta$ vanishes on A_j exactly when Z_{A_j} divides it, and $\deg h_j = m$; a representative of degree less than m is unique. □

Proof (of Theorem 2). Let $\mathcal{O}$ be the set of accepting openings produced in the execution, let E_{pos} be the event that $\mathcal{O}$ contains two openings at one position with different values, and let E_{code} be the event that $\neg E_{\mathrm{pos}}$ holds and yet no $c \in C_{\mathrm{TB}}$ agrees with every element of $\mathcal{O}$. An adversary causing E_{pos} yields, by forwarding com and the two openings, an adversary against Definition 3 with the same success probability, so $\Pr[E_{\mathrm{pos}}] \le \varepsilon_{\mathrm{pos}}$. Under $\neg E_{\mathrm{pos}}$ each position occurring in $\mathcal{O}$ carries a single value, so discarding duplicates leaves a collection

at pairwise distinct positions with the same consistency status; forwarding it gives an adversary against the code binding clause of Definition 3, so $\Pr[E_{\mathrm{code}}] \le \varepsilon_{\mathrm{code}}$.

Assume neither event occurs and let $c^\star$ be any completion. Suppose a verifier certifies A_j from r distinct accepting openings at $U \subseteq A_j$; their values are $c^\star|_U$. By Theorem 1(ii) there is a $\delta_j^\star$ of degree less than r with $c^\star|_{A_j} = (\delta_j^\star(a))_{a \in A_j}$. The algorithm $\mathsf{Certify}$ interpolates the unique polynomial of degree less than r through the r distinct points of U with values $c^\star|_U$; since $\delta_j^\star$ is such a polynomial and the interpolant is unique in that space, the two coincide and $\widehat{c}_j = c^\star|_{A_j}$. As $\widehat{c}_j$ is computed from the retained evidence alone, it is independent of the choice of completion, so every completion induces the same restriction on A_j and $\mathsf{FalseCert}$ fails. Applying the argument to two certifications of one group shows that both output $c^\star|_{A_j}$, so $\mathsf{Inconsistent}$ fails. The argument never refers to how C_j was obtained nor to the order of queries, so it is unaffected by adaptivity. □

Proof (of Theorem 3). The scheme is the compiler of [22, §6.3] applied to C_{TB}, whose reception efficiency is $D+1$ by Theorem 1(iii), with the extractor allowed to use cl. Suppose all T designated clients accept and Ext outputs $\perp$. If two accepting openings at one position carry different values, forwarding them breaks Definition 3; this contributes $\varepsilon_{\mathrm{pos}}$, and outside it the selection made by Ext is the only value at each position. Otherwise either $|\mathrm{cl}(S_T)| \le D$, an event of probability at most $\nu^\star(T)$ over the independent randomness of the clients, or $|\mathrm{cl}(S_T)| \ge D+1$ and the final check failed. In the latter case the values on $\mathrm{cl}(S_T)$ are inconsistent with every codeword: those at positions of S_T come from accepting openings at distinct positions, and those filled by cl are, by Theorem 1(ii), the unique completion of r accepting openings inside an MDS restriction. Passing com and the selected openings to the code binding game therefore succeeds, contributing $\varepsilon_{\mathrm{code}}$. Consistency is [22, Lem. 10] with [22, Lem. 2].

For the uniform sampler, $|\mathrm{cl}(S_T)| \ge |S_T|$, so the event implies $|S_T| \le D$ and the bound is [22, Lem. 3] with $\Delta = D$. For the group sampler, a client that queries r distinct positions of one group completes it, so $|\mathrm{cl}(S_T)| \ge m \cdot |\{\text{groups touched}\}|$; as $\lceil (D+1)/m \rceil = H_\star$ by Theorem 5, the event implies that the clients touched at most $H_\star - 1$ groups, and the bound is the same lemma with $N = \ell$, $Q = 1$, $\Delta = H_\star - 1$. □

Proof (of Theorem 5). Every position of $\mathrm{cl}(S) \setminus S$ lies in a group of which r positions are in S; by Theorem 1(ii) the restriction of c to such a group is a codeword of an (m, r) MDS code, so those r values determine the whole restriction and all values on $\mathrm{cl}(S)$ are known. As $\deg f \le D$, any $D+1$ of them determine f. Under the stated hypothesis the groups completed by S are exactly those of H, so $\mathrm{cl}(S) = \bigsqcup_{j \in H} A_j \sqcup S_{\mathrm{out}}$; without it the union may be larger and the equality becomes an inequality. Finally $C(h) = hr + (D+1-hm) = D+1-h\alpha$ whenever $R(h) > 0$.

For the case distinction write $k = ar + b$. If $b > 0$, then $H_\star = a+1$, $D = am+b-1$, and $D+1-hm > 0$ exactly for $h \le a = H_\star - 1$, with $D+1-am = b$, so $C(H_\star - 1) = ar + b = k$. If $b = 0$, then $H_\star = a$, $D = (a-1)m + r - 1$, and $D + 1 - hm > 0$ exactly for $h \le a - 1$, with $D + 1 - (a-1)m = r$, so

$C(H_\star - 1) = (a-1)r + r = k$. In both cases C decreases by α per certified group on $[0, H_\star - 1]$ and equals $hr \geq k$ afterwards; the computation also shows $(H_\star - 1)m < D + 1 \leq H_\star m$, that is, $\lceil (D+1)/m \rceil = H_\star$. □

Proof (of Corollary 2). The first claim is immediate from $\lceil (D+1)/m \rceil = H_\star$. For an arbitrary set of $H_\star$ groups one interpolates each δ_j, reads off $p_i(\gamma_j)$ by Theorem 1(ii), and interpolates each p_i from its $H_\star \geq \kappa_i$ known values. For J_{msg} the blocks are read directly by Proposition 2. For (8), if $x \notin \bigcup_{s<a} A_{j_s}$, then $g(x) \neq \gamma_{j_s}$ for every s, so $B(g(x)) \neq 0$ and the identity follows by evaluating $f = \sum_s L_s(g)u_s + B(g)e$ at x; as $\deg e < b$, any b distinct such positions determine e. □

Remark 3 (The Criterion Is Sufficient, Not Necessary). Recovery of a linear code from a position set S is possible exactly when $\mathrm{rank}(G_S) = k$, and since local completion adds only linear combinations of observed values, $\mathrm{rank}(G_{\mathrm{cl}(S)}) = \mathrm{rank}(G_S)$. The condition $|\mathrm{cl}(S)| \geq D+1$ is therefore sufficient for the interpolation-only decoder of Construction 1, not a characterization. It can fail while recovery is possible: over $\mathbb{F}_{17}$ with $m = 4$, $\ell = 4$, $A = \mathbb{F}_{17}^{\times}$, $r = 3$, $\alpha = 1$, and $k = 4$ we have $D = 4$ and $V_D = \mathrm{span}\{1, X, X^2, X^4\}$; the set $S = \{1, 2, 3, 4\}$ meets three distinct groups, so $\mathrm{cl}(S) = S$ and $|\mathrm{cl}(S)| = 4 < 5 = D + 1$, yet the evaluation matrix of that basis at those four points has determinant 1 and the message is determined.

Proof (of Proposition 7). (i) holds by construction, since $|W_j| = \alpha + 1$ leaves $m - \alpha - 1 = r - 1 < r$ positions in each group.

(ii) Row t is erased in group j exactly when $c_j := \lfloor mj/\ell \rfloor$ lies in the cyclic window of the $w := \alpha + 1$ residues $t - \alpha, \ldots, t$ modulo m. Since $m \geq \ell$, the map $j \mapsto c_j$ is strictly increasing with values in $[0, m)$, and $c_j \in [x, x + w)$ holds exactly for the integers j in an interval of length $w\ell/m$, of which there are at most $\lceil w\ell/m \rceil$. If the window wraps, write it as $[x, m) \cup [0, y)$ with $u := m - x$ and $u + y = w$; the count is $\lfloor u\ell/m \rfloor + \lceil y\ell/m \rceil$, which is smaller than $w\ell/m + 1$ and hence again at most $\lceil w\ell/m \rceil$. Each row therefore keeps at least $\ell - \lceil w\ell/m \rceil \geq a$ symbols, and the outer code $\mathrm{RS}[\ell, a]$ recovers it.

(iii) By Corollary 3, $d_{\mathrm{TB}} = (\ell - a + 1)m - r + 1$, and $\alpha + 1 = m - r + 1$. Hence $\ell(m - r + 1) < (\ell - a + 1)m - r + 1$ is equivalent to $-\ell r + \ell < -am + m - r + 1$, that is, to $m(a-1) < \ell r - \ell - r + 1 = (\ell - 1)(r - 1)$. □

Proof (of Corollary 3). By Theorem 1(iii), $d_{\mathrm{TB}} = n - D$ with $D = (a-1)m + r - 1$, which gives the first expression; the second is the product of the component distances $\ell - a + 1$ and $m - r + 1$. Subtracting, $(\ell - a + 1)m - r + 1 - (\ell - a + 1)(m - r + 1) = (\ell - a + 1)(r - 1) - (r - 1) = (\ell - a)(r - 1)$, and at $a = \ell$ both expressions equal $m - r + 1 = \alpha + 1$. For the ratio, write $A := \ell - a + 1$ and divide the two expressions: $d_{\mathrm{TB}}/d_{\mathrm{prod}} = (Am - r + 1)/(A(m - r + 1)) = (m - (r-1)/A)/(m - r + 1) = \frac{m}{\alpha+1}(1 - \frac{r-1}{mA})$, using $m - r + 1 = \alpha + 1$. Since $0 < (r-1)/A \leq r - 1 < m$, the bracket lies strictly between 0 and 1, which gives the upper bound $m/(\alpha + 1)$; expanding the product shows that the deficit from it is $(r-1)/((\alpha+1)A)$, which is the stated correction and tends to 0 as A grows. □

Proof (of Proposition 5). Write U for the $a \times m$ array whose rows are $\mathrm{ev}_B(u_s)$, so that the codeword is the $\ell \times m$ array MU. Every column of MU is M applied to a column of U, hence lies in $\mathrm{Im}(M)$, and every row of MU is a linear combination of rows of U, hence lies in $\mathrm{RS}_r(B)$. Conversely, let an array Z have all columns in $\mathrm{Im}(M)$ and all rows in $\mathrm{RS}_r(B)$; since M has rank a it has a left inverse M^+, each column of Z is M applied to the corresponding column of $U := M^+Z$, so $Z = MU$, and each row of U is a linear combination of rows of Z and therefore lies in $\mathrm{RS}_r(B)$. The code is thus exactly the tensor product $\mathrm{Im}(M) \otimes \mathrm{RS}_r(B)$, and the map $(u_s)_s \mapsto MU$ is injective because M^+ recovers U, so the dimension is ar. The minimum distance of a tensor product is the product of the component distances, $\mathrm{Im}(M)$ is an $[\ell, a]$ code and so has distance at most $\ell - a + 1$ by the Singleton bound, and $d(\mathrm{RS}_r(B)) = m - r + 1$. The final claim is Corollary 3. □

Proof (of Proposition 6). A point of A_j is $\beta_j h$ with $h \in H$, and $\delta_j(\beta_j h) = \sum_{i<r} d_{j,i}\beta_j^i h^i$, which is the evaluation on H of the polynomial with coefficients $d_{j,i}\beta_j^i$. Building the powers $\beta_j^2, \ldots, \beta_j^{r-1}$ costs $r-2$ multiplications per group and applying them to the coefficients of $X, \ldots, X^{r-1}$ costs $r-1$ more; with the powers tabulated, only the latter remains. □

Proof (of Lemma 1). The event $|\mathrm{cl}(S_T)| \le D$ implies $|S_T| \le D$, that is, at least $n - D$ positions are untouched by all QT draws. For position i, the indicator that i is untouched is a non-increasing function of the occupancy of i, and the occupancy numbers of balls thrown independently into bins are negatively associated [15]; non-increasing functions of disjoint sets of negatively associated variables are again negatively associated, so these n indicators are negatively associated, each with mean p. The Chernoff–Hoeffding upper-tail bound [23] applies to sums of negatively associated indicators exactly as to independent ones, and the stated inequality is that bound at the deviation $(n - D)/n > p$. □

The union-bound forms of Theorem 3 require $T = 904$ for the uniform sampler and $T = 399$ for the group sampler, against 663 and 398 in Section 8. Under the uniform sampler the closure almost never adds a position (Proposition 8), so under the cardinality criterion the clients must collectively receive $D + 1$ distinct positions, which requires $T \ge (D+1)/Q = 384$.

Proposition 8 (Relative Benefit of the Closure). *Let N uniform samples with replacement be drawn from A, let S_N be the set of distinct positions obtained, and put $\nu_0(N) := \Pr[|S_N| \le D]$ and $\nu_{\mathrm{cl}}(N) := \Pr[|\mathrm{cl}(S_N)| \le D]$. Assume $D/n < r/m$ and set $\delta := \min\{1,\ \ell \exp(-m\, D_{\mathrm{KL}}(r/m \| D/n))\}$. Then for every N with $\nu_0(N) > 0$ we have $0 \le (\nu_0(N) - \nu_{\mathrm{cl}}(N))/\nu_0(N) \le \delta$. At the reference profile the per-group bound is $2.27 \cdot 10^{-8}$ and $\delta = 2.79 \cdot 10^{-6}$.*

Proof. Since $|\mathrm{cl}(S_N)| \ge |S_N|$ we have $\{|\mathrm{cl}(S_N)| \le D\} \subseteq \{|S_N| \le D\}$, giving $\nu_{\mathrm{cl}} \le \nu_0$ and the left inequality. For the right one, condition on $|S_N| = s$ with $s \le D$; given its size, S_N is a uniformly random s-subset of A, so for each group $N_j := |S_N \cap A_j|$ is hypergeometric with parameters (n, m, s) and mean ms/n. The comparison of the hypergeometric with the binomial distribution [23]

gives the same upper-tail bound as for m independent Bernoulli variables of mean s/n, namely $\Pr[N_j \geq r] \leq \exp(-m\, D_{\mathrm{KL}}(r/m \| s/n))$, valid because $r/m > D/n \geq s/n$; the same bound follows from the negative association of balls-and-bins occupancy [15]. As $D_{\mathrm{KL}}(r/m\|\cdot)$ is decreasing on $[0, r/m)$, this is at most $\exp(-m\, D_{\mathrm{KL}}(r/m\|D/n))$, and a union bound over the ℓ groups gives $\Pr[\mathrm{cl}(S_N) \neq S_N \mid |S_N| = s] \leq \delta$ for every $s \leq D$. The closure can help only when the raw set fails and the closure does not, which requires some group to be complete, so $\nu_0(N) - \nu_{\mathrm{cl}}(N) = \Pr[|S_N| \leq D < |\mathrm{cl}(S_N)|] \leq \delta\, \nu_0(N)$. The numbers follow from $D/n = 0.664626$, $r/m = 0.75$, $D_{\mathrm{KL}} = 0.017190$, $m = 1024$, and $\ell = 123$. □

At the reference profile, local completion therefore reduces the failure probability of a uniform sampler under the cardinality criterion by a relative amount of at most $2.8 \cdot 10^{-6}$. The statement concerns the uniform sampler only and does not apply to samplers that bias their queries toward groups.

B Security Proofs and Deployment Profiles

This appendix proves the statements of Section 5, details the deployment profiles, and shows that canonicality does not follow from the binding properties.

Root finding and the cost of the reduction. The reduction below must find τ among the roots of a nonzero polynomial $R \in \mathbb{F}_q[X]$ of degree d with $R(\tau) = 0$. It computes $\gcd(R, X^q - X)$ with $O(\mathsf{M}(d) \log q)$ field operations, where $\mathsf{M}(d)$ is the cost of multiplying two polynomials of degree at most d, and splits the result into linear factors by equal-degree factorization [20, §14], which uses random splitting elements and runs in expected time polynomial in d and $\log q$. Each of the at most d roots θ is compared with τ by checking $\theta G_1 = \tau G_1$. Since the correct root is always recognized, the procedure never outputs a wrong value, and only its running time is random. Reading the algebraic representations of N outputs and forming the relations below costs $O(Nr + Gar + N\,\mathsf{M}(r) \log r)$ field operations when $G \leq \ell$ distinct groups occur. These costs, with $d \leq 2r - 1$, form the expected time t_{red} of Theorem 4.

Proof (of Theorem 4). Because the adversary is algebraic, every $\mathbb{G}_1$ element it outputs comes with a representation as a linear combination of the elements of ck and of elements it has received; since ck contains no $\mathbb{G}_1$ element of degree above $r - 1$ and none is published elsewhere, reading off coefficients yields polynomials of degree at most $r - 1$. Applied to the header, this gives $u_s^\star$, $e^\star$, and $\bar{e}^\star$, all of degree less than r. Extraction is deterministic and uses no rewinding, incurring no security loss with respect to a.

A single reduction. Define E as the event that the adversary produces any of the three failures below. Each exhibits an explicit polynomial R that is nonzero, has degree at most $2r-1$, and satisfies $R(\tau) = 0$. The reduction runs the adversary once and then, in order, reads off the algebraic representations of its $\mathbb{G}_1$ outputs and forms the corresponding polynomials, which for a derived commitment is the

same a-term combination that DeriveGrp performs; scans the openings and the degree equation for a relation that the verifier accepted while the corresponding R is not the zero polynomial, testing each candidate by comparing coefficients; and finds τ among the roots of that R as described above. No rewinding and no guessing of a target relation occur, so the reduction succeeds whenever E occurs, and no second cryptographic assumption is used.

Single-point openings. Let an accepting opening of a commitment C at x carry v, with extracted P and proof polynomial Π, both of degree at most $r-1$. Put $R := P - v - (X-x)\Pi$, of degree at most r. Acceptance states $P(\tau) - v = \Pi(\tau)(\tau - x)$, that is, $R(\tau) = 0$. If the opening is incorrect, then $R(x) = P(x) - v \neq 0$, so $R \neq 0$. Hence outside E every accepting single-point opening satisfies $v = P(x)$.

Multipoint openings. For U with $|U| \leq r$ the equation reads $P(\tau) - I_U(\tau) = \Pi_U(\tau) Z_U(\tau)$ with $\deg \Pi_U \leq r-1$ and $\deg Z_U = |U| \leq r$. Put $R := P - I_U - Z_U \Pi_U$, of degree at most $2r-1$, so that acceptance is $R(\tau) = 0$. If $R \neq 0$, event E occurs; otherwise Z_U divides $P - I_U$ and the opened values are $P|_U$.

The degree equation, when $b > 0$. Acceptance states $\tau^{r-b} e^\star(\tau) = \bar{e}^\star(\tau)$. Put $R := X^{r-b} e^\star - \bar{e}^\star$, of degree at most $2r-b-1 \leq 2r-1$, with $R(\tau) = 0$. If $R \neq 0$, event E occurs; otherwise $X^{r-b} e^\star = \bar{e}^\star$ as polynomials, and $\deg \bar{e}^\star \leq r-1$ forces $\deg e^\star \leq b-1$.

Outside that event $(\boldsymbol{u}^\star, e^\star)$ is an admissible tuple, so Proposition 1 places $f^\star := \Phi(\boldsymbol{u}^\star, e^\star)$ in V_D, which is (a). Because DeriveGrp is a fixed linear combination and algebraic extraction is linear, the polynomial extracted from C_j is $\delta_j^\star := \sum_s L_s(\gamma_j) u_s^\star + B(\gamma_j) e^\star$, which by (3) equals $f^\star \bmod h_j$; that is (c), with $\mathsf{LCom} := \mathsf{KZG.Com}$. By the single-point and multipoint opening arguments above, every accepting opening at $a \in A_j$ satisfies $v = \delta_j^\star(a) = f^\star(a) = c^\star(a)$ by Theorem 1(ii), which is (b). Two accepting openings at one position therefore agree, which is position binding; all accepting openings agree with the single $c^\star \in C_{\mathrm{TB}}$, which is code binding; and for canonicality, the r accepting openings that the adversary must produce carry the values $c^\star|_U$, whose degree-$(< r)$ interpolant is $\delta_j^\star$ by uniqueness, while DeriveGrp returns $\mathsf{KZG.Com}(\delta_j^\star)$ by (c). □

Proof (of Proposition 3). Take $b = 0$ and set $U_s := \tau^r G_1$ for every $s < a$. The Lagrange basis is a partition of unity, $\sum_{s<a} L_s(T) \equiv 1$, since the right-hand side is the unique polynomial of degree less than a interpolating the value 1 at the a source labels. Hence DeriveGrp returns $C_j = \sum_s L_s(\gamma_j) U_s = \tau^r G_1 = \mathsf{KZG.Com}(X^r)$ for every j. The adversary answers a query at $x \in A_j$ with $v := x^r$ and $\pi := \mathsf{KZG.Com}((X^r - x^r)/(X-x))$; the quotient has degree $r-1$, so π is computable from $(\tau^i G_1)_{i<r}$ alone, and every such opening verifies. But $r < m$ gives $X^r \bmod h_j = X^r$, of degree r, so by Theorem 1(ii) no codeword of C_{TB} agrees with these values on A_j: already $r+1$ accepting openings inside one group are inconsistent with every codeword, since no polynomial of degree less than r interpolates $r+1$ points of the graph of X^r. When $b > 0$ the same adversary sets $E := \bar{E} := \mathcal{O}$, which satisfies the degree equation. The identical choice breaks the source-compressed product-code baseline of Section 7, whose derived column

commitments are the same linear combinations and whose column code is RS_r on $m > r$ points. □

We now state the three deployment profiles in full. In *Profile A*, a dedicated setup of degree r, Construction 2 applies as stated. In *Profile B* the reference string carries $\mathbb{G}_1$ powers up to $L \geq r$ and also $[\tau^{L-r+1}]_{\mathbb{G}_2}$; the committer then publishes a further elements $\hat{U}_s := \mathsf{KZG.Com}(X^{L-r+1}u_s)$, and $\mathsf{VerifyCom}$ samples $\rho \leftarrow \mathbb{F}_q$ with its own randomness and checks

$$e\Big(\textstyle\sum_{s<a} \rho^s U_s,\ [\tau^{L-r+1}]_{\mathbb{G}_2}\Big) \ = \ e\Big(\textstyle\sum_{s<a} \rho^s \hat{U}_s,\ G_2\Big). \tag{11}$$

This introduces no Fiat–Shamir transform, since ρ is sampled after com is fixed, as in the row check of the tensor-code commitment of [22, §8]. Proposition 9(i) covers $r \mid k$ and hence the reference profile; for $b > 0$ the residual block needs a shift of its own (Proposition 9(ii) and Remark 4). In *Profile C* the string carries a long sequence of $\mathbb{G}_1$ powers but not $[\tau^{L-r+1}]_{\mathbb{G}_2}$; then neither Theorem 4 nor (11) applies, and the construction must not be run without a new setup.

Proposition 9 (Security and Cost of Profile B). *Let the reference string publish $(\tau^i G_1)_{0\leq i\leq L}$ with $L \geq r$ and $(\tau^i G_2)_{0\leq i\leq r}$, and in both cases below let $\rho \leftarrow \mathbb{F}_q$ be sampled by the verifier after* com *is fixed. Write $\varepsilon_{\mathrm{dl}}^{(L)}$ for the advantage against discrete logarithm with that string as auxiliary input.*

(i) Let $r \mid k$, so that $b = 0$. Let the string also publish $[\tau^{L-r+1}]_{\mathbb{G}_2}$, let the header carry the a further elements $\hat{U}_s := \mathsf{KZG.Com}(X^{L-r+1}u_s)$, and let $\mathsf{VerifyCom}$ *perform* (11)*. Then the conclusions of Theorem 4 hold with $\Pr[E] \leq \varepsilon_{\mathrm{dl}}^{(L)}(t + t_{\mathrm{red}}) + (a-1)/q$, and the check costs two pairings and two a-term multiscalar multiplications.*

(ii) Let $b > 0$. Let the string also publish $[\tau^{L-r+1}]_{\mathbb{G}_2}$ and $[\tau^{L-b+1}]_{\mathbb{G}_2}$, let the header carry the $\hat{U}_s$ of (i) together with $\hat{E} := \mathsf{KZG.Com}(X^{L-b+1}e)$, and let $\mathsf{VerifyCom}$ *check*

$$\begin{aligned} &e\Big(\textstyle\sum_{s<a} \rho^s U_s,\, [\tau^{L-r+1}]_{\mathbb{G}_2}\Big) \cdot e\big(\rho^a E,\, [\tau^{L-b+1}]_{\mathbb{G}_2}\big) \\ &\qquad = e\Big(\textstyle\sum_{s<a} \rho^s \hat{U}_s + \rho^a \hat{E},\, G_2\Big). \end{aligned} \tag{12}$$

Then the same conclusions hold with $(a-1)/q$ replaced by a/q, at three pairings; the header element $\bar{E}$ and the degree equation of Construction 2 become redundant and may be dropped.

Neither case introduces a Fiat–Shamir transform.

Proof. The extracted polynomials now have degree at most L. In case (i) write $\Delta_s := X^{L-r+1}u_s^\star - \hat{u}_s^\star$ and $R_\rho := \sum_{s<a} \rho^s \Delta_s$; in case (ii) write in addition $\Delta_e := X^{L-b+1}e^\star - \hat{e}^\star$ and $R_\rho := \sum_{s<a} \rho^s \Delta_s + \rho^a \Delta_e$. In both cases $\deg_X R_\rho \leq 2L-b+1$. The differences are determined by com and hence fixed before ρ is sampled, so if one of them is nonzero then R_ρ, viewed as a polynomial in ρ with coefficients in $\mathbb{F}_q[X]$, is nonzero of degree at most $a-1$ in case (i) and at most a in case (ii), and

therefore vanishes identically for at most that many values of ρ; for every other value, acceptance states $R_\rho(\tau) = 0$ with $R_\rho \neq 0$, and the reduction of Theorem 4 recovers τ from its roots, using the root finding above at degree at most $2L + 1$. If all differences vanish, then $X^{L-r+1}u_s^\star = \hat{u}_s^\star$ as polynomials and $\deg \hat{u}_s^\star \leq L$ forces $\deg u_s^\star < r$; in case (ii) likewise $X^{L-b+1}e^\star = \hat{e}^\star$ with $\deg \hat{e}^\star \leq L$ forces $\deg e^\star \leq b - 1$, which is exactly the bound that Theorem 4(a) needs and that Remark 4 shows the shift X^{L-r+1} does not deliver. The proof of Theorem 4 then applies verbatim, with the single-point and multipoint opening relations now having degree at most $L + r$. Since $\mathsf{VerifyCom}$ is randomized, Definitions 3 to 4 are read with ρ sampled by the challenger after com is output; no random oracle is used, so no Fiat–Shamir term appears. A deployment in which M verifiers each sample their own challenge is covered by a union bound, at $M(a-1)/q$ or Ma/q. The costs are immediate from the shapes of (11) and (12). □

Remark 4 (The Residual Block Needs a Shift of Its Own). When $b > 0$, applying the shift X^{L-r+1} of the source blocks to the residual block as well does not enforce $\deg e < b$. With $u_s = 0$ for every s, $e = X^b$, $\bar{e} = X^r$, and $\hat{e} = X^{L-r+1+b}$, all polynomials have degree at most L, the identities $X^{r-b}e = \bar{e}$ and $X^{L-r+1}e = \hat{e}$ hold, and every check passes for every ρ; yet the committed $f = B(X^m)X^b$ has degree $D + 1$, so $f \notin V_D$, while $f \bmod h_j = B(\gamma_j)X^b$ has degree $b < r$ for every j, so that no local check detects the violation. The shift X^{L-b+1} repairs this, because $\deg(X^{L-b+1}e) \leq L$ forces $\deg e \leq b - 1$. The failure concerns degree enforcement only and affects neither Profile A nor the reference profile, where $b = 0$.

At the reference profile the header grows from 3,936 to 7,872 bytes, and the same growth applies to the source-compressed baseline, so the ratios of light-client communication, 3.89 for single-point and 2.63 for group-batched openings under Profile A, become 3.44 and 2.30.

A deployment should bind a designated identifier to the reference string that each parameter profile uses and treat Profile A and Profile B as separate setups, since Proposition 3 concerns the public availability of reference elements rather than what a specific party possesses: if any protocol sharing the same τ publishes $\mathbb{G}_1$ powers above degree $r - 1$, Profile A no longer applies.

Corollary 4 (Reference Strings with Few $\mathbb{G}_2$ Powers). *Let $b = 0$ and $1 \leq s_2 < r$, and replace (6) by $\mathsf{ck}_{r,s_2} := ((\tau^i G_1)_{0 \leq i < r}, (\tau^i G_2)_{0 \leq i \leq s_2})$, with no $\mathbb{G}_1$ power of τ above degree $r - 1$ published anywhere. Let Verify accept multipoint openings only for $|U| \leq s_2$, and let a set of r values in group j be accepted exactly when $\mathsf{KZG.Com}(I_U) = C_j$. Then Theorem 4 holds with $\varepsilon_{\mathrm{dl}}$ replaced by the advantage $\varepsilon_{\mathrm{dl}}^{(r-1,s_2)}$ against the problem of computing τ from ck_{r,s_2}, which is at most $\varepsilon_{\mathrm{dl}}$, and Proposition 4 holds for $|U| \leq s_2$ and for $|U| = r$.*

Proof. The reduction receives exactly ck_{r,s_2}, and an algorithm for this problem solves the $(r-1, r)$ problem by ignoring the additional inputs. With $b = 0$, $\mathsf{VerifyCom}$ checks only the shape of com, and every remaining check uses $\mathbb{G}_2$ powers of degree at most s_2: a multipoint opening with $|U| \leq s_2$ needs $[Z_U(\tau)]G_2$,

and the whole-group check uses no pairing. The polynomials R in the proof of Theorem 4 are unchanged for $|U| \le s_2$, and for the whole-group check acceptance means $P(\tau) = I_U(\tau)$ for the polynomial P extracted from C_j, so $R := P - I_U$, of degree less than r, is either zero or yields τ. The proofs produced in Proposition 4 lie in $\mathbb{G}_1$ and use only $(\tau^i G_1)_{i<r}$. □

The 2^{12} transcript of the Ethereum ceremony is $\mathsf{ck}_{4096,64}$ for a secret used only in that transcript [16]. A matching profile is $r = 4096$, $m = 5632 = 2^9 \cdot 11$, which divides $q-1$ for the BLS12-381 scalar field, $\ell = 32$, and $a = 22$: then $n = 180{,}224$, $k = 90{,}112$ (rate 1/2), $D = 122{,}367$, $d_{\mathrm{TB}} = 57{,}857$ against $d_{\mathrm{prod}} = 16{,}907$, a ratio of 3.42, and one light client needs 230 uniform samples against 901 for the same detection target, with a header of 1,056 bytes. Each source polynomial has the size of one EIP-4844 blob [8]. We present this profile as a valid setting compatible with the existing ceremony and report no timings for it, since its evaluation domain differs from that of Section 8.1.

Concrete security of the assumption. Cheon's algorithm [11] recovers τ from τG and $\tau^d G$ for a divisor d of $q-1$ with about $\sqrt{q/d} + \sqrt{d}$ group operations, a loss of about $\frac{1}{2}\log_2 d$ bits relative to the generic bound. For the BLS12-381 scalar field, 768 divides $q-1$, so the reference string of Construction 2 admits a loss of about 4.8 bits, suggesting roughly 123 bits of generic security; a design that commits directly to f publishes powers up to degree D, where the largest divisor of $q-1$ not exceeding D is 81,416, suggesting roughly 119 bits. The pairing group constrains both designs further: updated key-size estimates for pairings [2] recommend BLS12 curves over larger base fields than that of BLS12-381 for 128-bit security. These figures are estimates of attack work, not bounds on $\varepsilon_{\mathrm{dl}}$; the 128-bit targets elsewhere in this paper are statistical.

Proof (of Proposition 4). Deriving C_j is a public linear combination of com. By Theorem 4(b) the r held values are $c^\star|_U$, so by Theorem 1(ii) their degree-$(<r)$ interpolant is $f^\star \bmod h_j$, which by (c) is exactly what C_j commits to. The proof for U is $\mathsf{KZG.Com}((\widehat{\delta}_j - I_U)/Z_U)$, a polynomial of degree less than $r - |U| \le r$, so only $(\tau^i G_1)_{i<r}$ is used, and the verification equation holds by correctness. □

Remark 5 (Rationale for Derived over Supplied Commitments). If the prover supplies each group's commitment, canonicality must be enforced and does not follow from the shape of the header. Consider the variant in which the header commits to f itself under a reference string of degree D and each group carries a commitment $C^{(j)}_{\mathrm{loc}}$ to a polynomial δ_j together with a commitment $C^{(j)}_q$ to a quotient q_j, accepted when the pairing equation corresponding to $f - \delta_j = h_j q_j$ holds. Take $f = 0$ and $\delta_j := h_j = X^m - \gamma_j$ with $q_j := -1$. Both are admissible at degree $D \ge m$, the quotient identity holds as polynomials, and $\delta_j(a) = a^m - \gamma_j = 0 = f(a)$ for every $a \in A_j$, so the group data and all openings agree with the global codeword; yet a node that collects the r symbols, all zero, interpolates $\widehat{\delta}_j = 0$ and computes proofs against $\mathsf{KZG.Com}(0) = \mathcal{O}$, which do not verify against the published $C^{(j)}_{\mathrm{loc}} = [\tau^m - \gamma_j]G_1$ unless $\tau \in A_j$, an event of

probability at most $m/(q-1)$ for uniform $\tau \in \mathbb{F}_q^\times$. Although the reconstructed data is correct, the regenerated proofs fail to verify, representing the group-local manifestation of the obstruction recorded in [22, Ex. 1]: δ_j is congruent to f modulo h_j without being the representative of degree less than m. Appendix D restores canonicality with a second pairing equation and states the resulting costs. Under Construction 2 no such commitment exists, because C_j is not supplied.

Remark 6 (Canonicality Is Not Implied by Binding). Consider the following modification of Construction 2. The reference string also publishes $\mathbb{G}_1$ powers up to a degree $L \geq m + r - 1$ and $[\tau^m]G_2$; the degrees of the source and residual blocks are enforced as in Profile B; and aux_j carries two elements (Y_j, W_j), accepted if $e(Y_j, G_2) = e(W_j, [\tau^m]G_2 - \gamma_j G_2)$, in which case $\mathsf{DeriveGrp}$ returns $C'_j := \sum_{s<a} L_s(\gamma_j)U_s + B(\gamma_j)E + Y_j$. In the algebraic group model, outside an event bounded as in Proposition 9, the polynomial extracted from Y_j is $h_j w_j^\star$, so the polynomial extracted from C'_j is $\delta_j^\star + h_j w_j^\star$ with $\delta_j^\star = f^\star \bmod h_j$ and $f^\star \in V_D$. Since h_j vanishes on A_j, every accepting opening on A_j carries the value $\delta_j^\star(a) = c^\star(a)$, and the arguments of Theorem 4 give position binding and code binding. Yet an adversary that sets $W_j := G_1$ and $Y_j := [\tau^m]G_1 - \gamma_j G_1$ obtains $C'_j = \mathsf{KZG.Com}(\delta_j + h_j) = \mathsf{KZG.Com}(\delta_j) + h_j(\tau)G_1$, while r honestly computed openings of group j under C'_j are accepted; canonicality therefore fails unless $h_j(\tau) = 0$, that is, unless $\tau \in A_j$, an event of probability at most $m/(q-1)$ for uniform $\tau \in \mathbb{F}_q^\times$. Remark 5 exhibits the same phenomenon for supplied group commitments.

C Parameters and Costs

This appendix collects the parameter tables and cost derivations at the reference profile, with group elements of 48 bytes and field elements of 32 bytes.

Table 3. Allocation of redundancy at fixed $n = 125{,}952$, $\ell = 123$, $m = 1024$, $k = 62{,}976$. Here $X_{\mathrm{glob}} := 1-(D+1)/n$; X_{loc} is the largest independent erasure rate for which a union bound over the groups, with the binomial tail evaluated exactly, keeps the probability that some group retains fewer than r symbols below 10^{-3}; and $s := \lceil 128/\log_2(n/D) \rceil$ is the number of uniform samples one client needs for a 2^{-128} detection error against a fixed missing pattern, a single-client figure and not the $\nu^\star(T)$ of Theorem 3. A rate-1/2 Reed–Solomon code gives $s = 128$.

r	α	D	$d = n - D$	X_{glob}	X_{loc}	$H_\star$	s	header
512	512	125,439	513	0.004	0.433	123	21,739	5,904
640	384	100,607	25,345	0.201	0.312	99	395	4,800
768	256	83,711	42,241	0.335	0.195	82	218	3,936
896	128	71,935	54,017	0.429	0.086	71	159	3,456

Table 4. The ratio in (9) at five shapes with $r \mid k$, with the single-client sample counts of Table 3 applied to both codes. The third row is the square shape $a/\ell = r/m = 1/2$ that a two-dimensional design with equal expansion in both directions produces.

ℓ	m	r	a	rate	r/m	$d_{\mathrm{TB}}/d_{\mathrm{prod}}$	$m/(\alpha+1)$	samples
123	1024	768	82	0.500	0.750	3.91	3.98	$991 \to 218$
123	1024	768	41	0.250	0.750	3.95	3.98	$479 \to 81$
128	1024	512	64	0.250	0.500	1.98	2.00	$303 \to 127$
123	1024	640	98	0.498	0.625	2.60	2.66	$1072 \to 384$
123	1024	896	70	0.498	0.875	7.81	7.94	$1560 \to 157$

Table 5. Costs of Construction 2 at the reference profile under Profile A, with $s = 218$, group elements of 48 bytes, and field elements of 32 bytes. The two collector rows correspond to two alternative verification modes (Appendix F) and are mutually exclusive.

Quantity	Value
Payload	$k \cdot 32 = 2{,}015{,}232$ B $= 1.92$ MiB
Header; per-group auxiliary data	$a \cdot 48 = 3{,}936$ B; none
Reference string, $\mathbb{G}_1$ / $\mathbb{G}_2$	768 / 769 elements = 36.0 / 72.1 KiB
Assumption	$(r-1, r)$-discrete logarithm, auxiliary degree 768
Per-query response	80 B
Light client, s samples	$3{,}936 + s \cdot 80 = 21{,}376$ B $= 20.9$ KiB
Group certification, $\lvert U\rvert = r$	$r \cdot 32 = 24$ KiB, proof omitted
Collector to $H_\star$ groups, whole-group verification	$H_\star\, r \cdot 32 = 1.92$ MiB, payload values only
Collector to $H_\star$ groups, per-response verification	$H_\star\, r \cdot 80 = 4.80$ MiB, plus failed responses
Serving state	r field elements per group; header and 36.0 KiB of compressed $\mathbb{G}_1$ powers shared
Derivation, parity group; light client	$a = 82$ terms, cached; $\approx 2{,}795$ terms

Derivations. Under Profile A, ck holds 768 elements of $\mathbb{G}_1$ (36,864 bytes) and 769 elements of $\mathbb{G}_2$ at 96 bytes. A query response is a field element and a proof, 80 bytes. Certifying a group opens exactly r positions with the identity element as proof, at $r \cdot 32 = 24{,}576$ bytes and r scalar multiplications for the check $\mathsf{KZG.Com}(I_U) = C_j$. A collector that requests each of $H_\star = 82$ groups as one set of r values transfers 2,015,232 bytes, exactly the payload; if every response is verified on arrival and therefore carries its own proof, the same k values cost $k \cdot 80 = 5{,}038{,}080$ bytes, plus the responses from abandoned groups. A client with $s = 218$ uniform samples touches $41(1 - (1 - 1/\ell)^{218}) \approx 34.08$ parity groups in expectation, hence about 2,795 scalar multiplications for derivation, against 3,362 for all 41 parity commitments. The detection counts are $\lceil 128/\log_2(n/D) \rceil = 218$ and $\lceil 128/\log_2(n/(n - d_{\mathrm{prod}})) \rceil = 991$; under Profile B the single-point light-client costs become 25,312 and 87,152 bytes.

Table 6. Communication ratios as the object grows, at fixed $m = 1024$, $r = 768$ and fixed global and local rates, so that ℓ and $a = 2\ell/3$ scale together. "First" charges each client for the header, as on the first retrieval of an object; "repeated" does not. Sample counts are the single-client detection figures; the distance ratio is (9).

Payload	Header	s_{TB}	s_{prod}	single-point		group-batched	
				first	repeated	first	repeated
1.92 MiB	3,936 B	218	991	3.89	4.55	2.63	3.17
19.2 MiB	39,360 B	219	1,013	2.12	4.63	1.87	3.94
192 MiB	393,600 B	219	1,016	1.16	4.64	1.15	4.55

General k. At $k = 60{,}000$ and $r = 768$ one has $a = 78$, $b = 96$, $D = 79{,}967$, and $d = 45{,}985$, with a header of $(a+2)\cdot 48 = 3{,}840$ bytes. Encoding the same message by padding to $k' = 60{,}672$ and using the case $b = 0$ would give $D' = 80{,}639$ and $d' = 45{,}313$ with a header of 3,792 bytes, so the general construction incurs an overhead of only 48 additional header bytes while recovering 672 message dimensions and increasing the minimum distance by 672; it also avoids having to state separately that the padded coordinates are zero, which the security target of a padded scheme would have to include.

Single-value repair under the pattern of Proposition 7. At the reference profile each group loses 257 values, each product row loses at most $31 \le 41$, and the total is $31{,}611 < d_{\mathrm{TB}} = 42{,}241$. In the product code an erased value at a common evaluation point x of group j is recovered from $a = 82$ surviving values in the same row: with $\delta_j = \sum_{s\in J} \mu_s \delta_s$ for the outer interpolation coefficients μ_s of a helper set J, linearity gives $(\delta_j(X) - \delta_j(x))/(X - x) = \sum_{s\in J} \mu_s (\delta_s(X) - \delta_s(x))/(X - x)$, so the opening proof is $\pi_j(x) = \sum_{s\in J} \mu_s \pi_s(x)$, as observed for two-dimensional KZG designs [7]. Thus 82 authenticated value–proof pairs, $82(32 + 48) = 6{,}560$ bytes, repair one value together with its proof.

Runtime measurements. Table 7 reports the complete measurements summarized in Table 2, under the conditions of Section 8.1.

D A Variant with Supplied Group Commitments

Construction 2 publishes no commitment to f itself. The natural alternative, which shares with the design of [14] the use of a commitment to the global polynomial, publishes one and lets each group carry its own commitment, linked to the global one by a pairing equation. We record this variant to make the cost of supplying group commitments explicit and to substantiate Remark 5; nothing in the body depends on it.

Construction 3 (Supplied Group Commitments). The reference string is $\mathsf{ck}' := ((\tau^i G_1)_{0\le i\le D}, (\tau^i G_2)_{0\le i\le m}, [\tau^{D-r+1}]G_2)$, and we assume $m \le D$. The

Table 7. Runtime and serialized sizes at $m = 1024$ under Profiles A and B, and full reconstruction under the erasure pattern of Proposition 7 at three group sizes with $r = 3m/4$, starting from authenticated values. "Product" denotes the source-compressed product code. Commitment costs include the shifted commitments in Profile B, which uses $L = 83{,}711$. Client verification covers 218 queries for LR-DAS and 991 for the product code and does not use (10). Each entry reflects three independent runs.

	Profile A		Profile B	
	LR-DAS	Product	LR-DAS	Product
Commitment generation	864.6 ms	863.8 ms	1,732.0 ms	1,728.6 ms
Unbatched client verification	414.3 ms	1,650.0 ms	427.3 ms	1,663.1 ms
Commitment size	3,936 B	3,936 B	7,872 B	7,872 B
$\mathbb{G}_1$ reference-string size	36,864 B	36,864 B	4,018,176 B	4,018,176 B
Reconstruction, $m = 512$	2,746.8 ms	271.3 ms		
Reconstruction, $m = 1{,}024$	6,247.1 ms	552.9 ms		
Reconstruction, $m = 2{,}048$	14,037.2 ms	1,102.8 ms		

committer parses the message exactly as in Construction 2 and computes $f := \Psi(\boldsymbol{u}, v) \in V_D$. It publishes $\mathsf{com} := C_{\mathrm{glob}}$, a commitment to f, and disperses with group j the auxiliary triple $\mathsf{aux}_j := (C_{\mathrm{loc}}^{(j)}, C_q^{(j)}, \bar{C}_{\mathrm{loc}}^{(j)})$, committing to $\delta_j := f \bmod h_j$, to $q_j := (f - \delta_j)/h_j$, and to $X^{D-r+1}\delta_j$. Then $\mathsf{DeriveGrp}$ returns $C_{\mathrm{loc}}^{(j)}$ if

$$e\big(C_{\mathrm{glob}} - C_{\mathrm{loc}}^{(j)},\, G_2\big) = e\big(C_q^{(j)},\, [\tau^m]G_2 - \gamma_j G_2\big), \tag{13}$$

$$e\big(C_{\mathrm{loc}}^{(j)},\, [\tau^{D-r+1}]G_2\big) = e\big(\bar{C}_{\mathrm{loc}}^{(j)},\, G_2\big), \tag{14}$$

and $\perp$ otherwise. Openings are as in Construction 2, against $C_{\mathrm{loc}}^{(j)}$. The extractor is that of Construction 1 and returns $\Psi^{-1}(f)$, so both instantiations use one message representation.

In Construction 3, equation (14) is what Remark 5 shows to be necessary: without it, the commitment with $\delta_j = h_j$ and $q_j = -1$ passes (13) while being non-canonical. Restoring a global membership guarantee at the level of the header alone, without relying on the number of verified triples or on sampling, would require binding the coefficient structure of f rather than the values it takes on each group: a canonical route is to commit to the bivariate polynomial $F(X, T) := \sum_{i<r} p_i(T)X^i$ via a bivariate commitment scheme [28] and verify $f(X) = F(X, X^m)$, which would strictly constrain the exponent residues modulo m exploited in Remark 7; we leave this extension outside the scope of the present work.

Theorem 6 (Binding and Canonicality of the Supplied Variant). *Work in the algebraic group model and let $\varepsilon'_{\mathrm{dl}}(t')$ denote the best advantage of an algorithm of expected running time t' against the discrete-logarithm problem with ck' as auxiliary input. For an adversary running in time t, there is an event*

E' with $\Pr[E'] \le \varepsilon'_{\mathrm{kzg}} := \varepsilon'_{\mathrm{dl}}(t + t'_{\mathrm{red}})$, where t'_{red} is polynomial in D, $\log q$, and the number of outputs of the adversary, outside which the following holds. Let $f^\star$ be the polynomial extracted from C_{glob}, write $f^\star = \sum_{i<m} p_i(g)X^i$ in the base-m decomposition of exponents, and put $f_{<r} := \sum_{i<r} p_i(g)X^i$. Then (a) $f_{<r} \in V_D$, so $c^\star := \mathrm{ev}_A(f_{<r}) \in C_{\mathrm{TB}}$; (b) every accepting opening (a, v) satisfies $v = c^\star(a)$; and (c) for every j with $\mathsf{DeriveGrp} \neq \bot$, $\mathsf{DeriveGrp}(\mathsf{ck}', \mathsf{com}, j, \mathsf{aux}_j) = \mathsf{KZG.Com}(f_{<r} \bmod h_j)$. Hence the variant is position binding, code binding, and canonical with error $\varepsilon'_{\mathrm{kzg}}$. The conclusion concerns the projection $f_{<r}$, not $f^\star$: it asserts neither $f^\star \in V_D$ nor that $f^\star$ and $f_{<r}$ agree outside the groups for which aux_j verifies, and Remark 7 exhibits an adversary for which both fail.

Proof (sketch). Extraction yields $f^\star, \delta_j^\star, q_j^\star, \bar{\delta}_j^\star$ of degree at most D. As in the proof of Theorem 4, every accepted single-point or multipoint opening, linking equation (13), and degree equation (14) either holds as a polynomial identity or yields a nonzero polynomial of degree at most $2D$ that vanishes at τ, from which the reduction recovers τ; let E' be the union of these events. Outside E', (14) gives $X^{D-r+1}\delta_j^\star = \bar{\delta}_j^\star$, so $\deg \delta_j^\star < r$, and (13) gives $f^\star - \delta_j^\star = h_j q_j^\star$. In the base-$m$ decomposition $\deg p_i < \ell$, so $\deg f_{<r} \le D$ and $f_{<r} \in V_D$, which is (a). For a group j with a verifying triple, $f^\star|_{A_j} = \delta_j^\star|_{A_j}$; since the functions $a \mapsto a^i$, $i < m$, are independent on A_j, comparing the two representations forces $p_i(\gamma_j) = 0$ for $r \le i < m$, hence $\delta_j^\star = f_{<r} \bmod h_j$ by Lemma 2, which is (c). By Theorem 1(ii) the accepted openings in group j carry the values $c^\star|_{A_j}$, which is (b), and canonicality follows from the uniqueness of the interpolant of r accepted values. □

Remark 7 (The Global Commitment Is Not Bound to the Code). Theorem 6 cannot be strengthened to a statement about $f^\star$. Take $f_{\mathrm{glob}} := (X^m - \gamma_0)X^r$ and supply for A_0 the triple $\delta_0 := 0$, $q_0 := X^r$, $\bar{\delta}_0 := 0$. At the reference profile $\deg f_{\mathrm{glob}} = m + r \le D$, both (13) and (14) hold as polynomial identities, and every opening on A_0 correctly returns 0. Both exponents of f_{glob} are congruent to r modulo m, so $f_{\mathrm{glob}} \notin V_D$, whereas $f_{<r} = 0$; for $\eta \notin A$ with $\eta \neq 0$, $f_{\mathrm{glob}}(\eta) \neq 0 = f_{<r}(\eta)$. An opening of C_{glob} outside A is therefore not an evaluation of the codeword to which the group interface is bound. This gap concerns degree and residue enforcement, not the hardness assumption, and does not arise in Construction 2, where membership is structural.

Lemma 3 (Conditional Global Membership). *Outside E', if the auxiliary triples of $H_\star - 1$ distinct groups verify, then $f^\star = f_{<r}$ and hence $f^\star \in V_D$. The threshold is tight: for every set J of $H_\star - 2$ group indices there is an $f \notin V_D$ of degree at most D whose triple verifies at every $j \in J$.*

Proof. For $r \le i < m$ the term $p_i(g)X^i$ of $f^\star$ has degree $i + m \deg p_i \le D$, so $\deg p_i \le \lfloor (D - r)/m \rfloor$, and that floor equals $H_\star - 2$ in both cases: for $b = 0$, $(D - r)/m = (a - 1) - 1/m$; for $b > 0$, $(D - r)/m = a + (b - 1 - r)/m$ with $-m < b - 1 - r < 0$. The proof of Theorem 6 gives $p_i(\gamma_j) = 0$ for every $i \ge r$ at each group whose triple verifies, so $H_\star - 1$ distinct labels are more roots than the degree allows and every such p_i vanishes, leaving $f^\star = f_{<r} \in V_D$. For tightness

put $f_J := X^r \prod_{j\in J}(X^m - \gamma_j)$, of degree $r + (H_\star - 2)m$, which is at most D by the same two computations, and whose exponents are all congruent to r modulo m, so $f_J \notin V_D$; at each $j \in J$ the triple $(\delta_j, q_j, \bar{\delta}_j) := (0,\ f_J/h_j,\ 0)$ satisfies (13) and (14) as polynomial identities. □

Corollary 5 (Sampling Against Global Nonmembership). *Let the adversary receive* ck' *and output* com *together with the algebraic representations of its group elements, and let a client then query* s *positions of* A *drawn uniformly and independently with its own randomness. Over the whole experiment,*

$$\Pr\big[f^\star \notin V_D \ \wedge\ \textit{all } s \textit{ queries are answered by accepting openings}\big] \le \varepsilon'_{\mathrm{kzg}} + \left(\tfrac{H_\star - 2}{\ell}\right)^s.$$

Proof. Outside E', Lemma 3 bounds by $H_\star - 2$ the number of groups whose triple can verify when $f^\star \notin V_D$, and that set is determined by $f^\star$, hence by com, before any query is made; adaptive answering cannot enlarge it. A query outside those groups has $\mathsf{DeriveGrp} = \perp$ and admits no accepting opening, and the groups partition A into equal parts, so each query lands in the admissible set with probability at most $(H_\star - 2)/\ell$, independently. □

At the reference profile $(80/123)^{218} = 2^{-135.29}$, so a client of the supplied variant that performs the sampling of Section 8 also rules out global nonmembership at no extra queries; this is a joint bound, not a bound conditioned on acceptance. By Lemma 3, 81 verified triples force $f^\star \in V_D$ at the reference profile, whereas 80 do not.

Uniqueness, serving, and cost. Outside E', two triples that $\mathsf{DeriveGrp}$ accepts at the same j are equal: the identities $f^\star - \delta = h_j q$ and $f^\star - \delta' = h_j q'$ give $\delta' - \delta = h_j(q - q')$, and $\deg\delta, \deg\delta' < r < m$ forces $\delta = \delta'$ and $q = q'$. This uniqueness is computational: if $C_{\mathrm{glob}} = \mathcal{O}$ and $\tau^m \neq \gamma_j$, then for every $z \in \mathbb{F}_q$ the triple $(zG_1,\ -z(\tau^m - \gamma_j)^{-1}G_1,\ z\tau^{D-r+1}G_1)$ satisfies (13) and (14), and only the hardness assumption prevents an efficient adversary from producing one with $z \neq 0$. A node that holds r accepted values of group j and a triple that verifies can serve as in Proposition 4, using Theorem 6(c) in place of Theorem 4(c); a node that holds correct data but no valid triple cannot, and obtaining the triple is a question of repair liveness in the sense of [22, §3.2], which Construction 2 avoids. At the reference profile the header of this variant is the single element C_{glob} of 48 bytes, the reference string holds $D+1 = 83{,}712$ elements of $\mathbb{G}_1$ and $m+2$ elements of $\mathbb{G}_2$, and a light client must also fetch the triple of every group it touches, at 144 bytes and two pairing equations each: about $48 + 218 \cdot 80 + 102.25 \cdot 144 \approx 32{,}212$ bytes on the first retrieval of an object, against 21,376 bytes for Construction 2, a ratio of 1.51. Construction 2 incurs an additional 3,888 bytes in the header but eliminates 144 bytes per queried group, thereby achieving lower overall communication whenever more than 27 groups are queried.

E Decoders, Local Access, and Specification Tests

E.1 Receiver-Side Decoder

Algorithm 1 is the receiver-side procedure implied by Corollary 2 and Theorem 5, and it is the extractor of Construction 1. The first two branches are the source-group path of Corollary 2, and the third is Theorem 5; that the output equals the committed message follows from Theorem 2 with Proposition 2. When $b > 0$, the source groups determine $u_0, \dots, u_{a-1}$ but not the residual block (Remark 1), so the second branch requires b further known positions outside the source groups, at which (8) evaluates e pointwise, rather than a further certified group. All branches end at one verification step, so a given commitment and set of evidence yield the same output regardless of the path. Under Construction 2 the test $f \in V_D$ always succeeds by Proposition 1, and by Remark 3 the stopping rule is conservative.

Algorithm 1 LR-DAS receiver-side decoder. Every branch produces a candidate f; the single exit validates it and returns $\Psi^{-1}(f)$.

Require: Accepting openings $S \subseteq A \times \mathbb{F}_q$, one value selected per position, and the commitment com.

$H \leftarrow \emptyset$
for each group A_j with at least r accepting openings in S **do**
  $\delta_j \leftarrow$ the degree-$(< r)$ interpolant of any r of them; extend the known values to all of A_j; $H \leftarrow H \cup \{j\}$
end for
$h \leftarrow |H|$; $S_{\text{out}} \leftarrow$ positions of S outside $\bigcup_{j \in H} A_j$
if $J_{\text{src}} \subseteq H$ and $b = 0$ **then**
  $f \leftarrow \sum_{s<a} L_s(g)\, \delta_{j_s}$
else if $J_{\text{src}} \subseteq H$ and $b > 0$ and $|P| \geq b$ for $P :=$ known positions outside $\bigcup_{s<a} A_{j_s}$ **then**
  evaluate e at b distinct points of P by (8); interpolate e of degree $< b$
  $f \leftarrow \sum_{s<a} L_s(g)\, \delta_{j_s} + B(g)\, e$
else if $hm + |S_{\text{out}}| \geq D + 1$ **then**
  $f \leftarrow$ the degree-$(\leq D)$ interpolant through any $D + 1$ positions of $\mathrm{cl}(S)$
else
  return $\perp$
end if
if $f \notin V_D$ or $\mathrm{ev}_A(f)$ disagrees with a selected opening of S **then**
  return $\perp$
end if
return $\Psi^{-1}(f)$

E.2 A Structured Erasure Decoder

Proposition 10 (Structured Erasure Decoding). *Let $b = 0$, let $S \subseteq A$ be a set of positions at which the values of a codeword $c = \mathrm{ev}_A(f)$, $f \in V_D$, are*

known, and let $s_j := |S \cap A_j|$*. For each* j *with* $s_j \geq r$*, let* δ_j *be the interpolant of degree less than* r *of any* r *of these values; for each* j *with* $s_j < r$*, let* I_j *be the interpolant of degree less than* s_j *of the values on* $S \cap A_j$*, let* $Z_j := Z_{S \cap A_j}$*, and write* $\delta_j = I_j + Z_j w_j$ *with an unknown polynomial* w_j *of degree less than* $r - s_j$*. Then the tuples of local polynomials of the codewords of* C_{TB} *that agree with* c *on* S *are exactly the solutions of the linear system*

$$\delta_j \;=\; \textstyle\sum_{s<a} L_s(\gamma_j)\, \delta_{j_s} \qquad (j \notin J_{\mathrm{src}}), \tag{15}$$

read coefficientwise, in the $U := \sum_j \max(0, r - s_j)$ *unknown coefficients of the* w_j*. In particular,* c *is determined by its values on* S *if and only if* (15) *has a unique solution, and a decoder that solves* (15) *recovers* c *whenever any decoder can.*

Proof. By Theorem 1(ii), a codeword c' of C_{TB} is determined by its local polynomials $\delta'_j = f' \bmod h_j$, each of degree less than r, and c' agrees with c on S exactly when every δ'_j interpolates the values on $S \cap A_j$. For $s_j \geq r$ this fixes $\delta'_j = \delta_j$ by the uniqueness of the interpolant. For $s_j < r$, the polynomials of degree less than r through the s_j given points are exactly the $I_j + Z_j w_j$ with $\deg w_j < r - s_j$, since two of them differ by a polynomial of degree less than r that vanishes on $S \cap A_j$. By Proposition 1 with $b = 0$, a tuple $(\delta'_j)_j$ of polynomials of degree less than r is the tuple of local polynomials of an element of V_D if and only if $\delta'_j = \sum_s L_s(\gamma_j)\delta'_{j_s}$ for every j: the forward direction is (3), and conversely $\Phi((\delta'_{j_s})_s)$ lies in V_D and has the local polynomials $\sum_s L_s(\gamma_j)\delta'_{j_s} = \delta'_j$, which determine it because ev_A is injective on V_D. Hence the solutions of (15) correspond bijectively to the codewords that agree with c on S. □

The decoder costs one interpolation per group, $O(\ell\, \mathsf{M}(r) \log r)$ field operations, and the solution of a linear system with $(\ell - a)r$ equations in U unknowns; for the pattern below, the system reduces to 123 equations in 123 unknowns, whose elimination alone takes about $123^3/3 \approx 6.2 \cdot 10^5$ field multiplications. The interpolations, the construction of the system, the inversions and back substitution, and the reconstruction of the local polynomials come in addition, so this count is not the cost of full recovery; we report the decoder's correctness, not its running time or memory use. Under the pattern of Proposition 7, $s_j = r - 1$ for every j, so each w_j is a scalar, $U = \ell$, and w_j equals the leading coefficient of δ_j, because Z_j is monic of degree $r - 1$ and $\deg I_j < r - 1$. At the reference profile, with the positions of group j indexed as $\beta_j \omega^t$ for a generator ω of H, the equations of (15) for the three leading coefficients, $3(\ell - a) = 123$ equations in 123 unknowns, have full rank over the BLS12-381 scalar field, both for $\beta_j = 7^j$ and for $\beta_j = \omega_{17}^j$ with ω_{17} a primitive 2^{17}th root of unity. The coefficients of the unknowns in these equations depend only on the positions, so this check does not involve the data. A reference-scale run on random data, included in the artifact, solves this 123×123 system for $\beta_j = 7^j$ and recovers the leading coefficient w_j of every δ_j, and hence every local polynomial $\delta_j = I_j + Z_j w_j$, exactly.

Table 8 reports an implementation of the decoder over $\mathbb{F}_{65537}$ on random data. For the product code, the same decoder applies with the common set H

in place of the cosets. The first row is the pattern of Proposition 7; each other row keeps a uniformly random set of positions of the stated size, drawn once from a fixed seed. In every row the system has full rank for both codes and the decoder returns the encoded data, including the rows that keep fewer than $D+1$ positions; the configuration $(m, r, \ell, a) = (32, 24, 24, 16)$ behaves in the same way on the pattern of Proposition 7, with 24 unknowns. The tested random instances therefore do not separate the two codes; being few and drawn from single seeds, they illustrate the decoder and do not characterize the random-erasure behavior of either code, whose difference we claim only for the worst-case patterns that the minimum distance captures. For such patterns U grows with the number of erasures, up to 884 here, so the count of ℓ unknowns is specific to the pattern of Proposition 7.

Table 8. The structured decoder over $\mathbb{F}_{65537}$ with $(m, r, \ell, a) = (64, 48, 60, 40)$, so that $n = 3{,}840$, $k = 1{,}920$, and $D + 1 = 2{,}544$. "Unknowns" is U of Proposition 10; the system has $(\ell - a)r = 960$ equations. The rank equals U in every row, and both codes are recovered exactly.

Erasure pattern	Kept	Fraction	Unknowns	Rank, C_{TB}	Rank, product
Proposition 7	2,820	0.734	60	60	60
Random	2,688	0.700	223	223	223
Random	2,534	0.660	347	347	347
Random	2,304	0.600	576	576	576
Random	2,112	0.550	768	768	768
Random	1,996	0.520	884	884	884

E.3 Local Access

Proposition 11 (Local-Access Consistency of the Message Representation). *Let* $\mathsf{Access}(\mathsf{ck}, \mathsf{com}, i)$ *be the following procedure, in which* $j \in J_{\mathrm{msg}}$ *is the group whose block contains message coordinate* i *under* (5) *and* $U_j \subseteq A_j$ *is a fixed set of* r *positions: if* $\mathsf{VerifyCom}(\mathsf{ck}, \mathsf{com}) = 0$*, return* $\perp$*; set* $C_j \leftarrow \mathsf{DeriveGrp}(\mathsf{ck}, \mathsf{com}, j, \mathsf{aux}_j)$ *and return* $\perp$ *if it is* $\perp$*; query the positions of* U_j *and return* $\perp$ *if fewer than* r *answers arrive; set* $\widehat{\delta}_j \leftarrow \mathrm{Interp}_{<r}(U_j, \{v_x\})$ *and check* $\mathsf{KZG.Com}(\widehat{\delta}_j) = C_j$*, which is the* $|U| = r$ *case of* Verify*, returning* $\perp$ *if the check fails; otherwise return the coefficient of* $\widehat{\delta}_j$ *that* (5) *assigns to index* i*. Then* Access *makes* r *queries for every index* $i < k$*; on an honest encoding it returns the correct symbol; and against any adversary, except with probability* $\varepsilon_{\mathrm{code}} + \varepsilon_{\mathrm{pos}}$*, every non-*$\perp$ *output is a coordinate of the single message* $\Psi^{-1}(f^\star)$*, where* $c^\star = \mathrm{ev}_A(f^\star)$ *is the codeword of Theorem 2. The representation of Proposition 2 is therefore locally accessible in the sense of [22, §3.2] with query complexity* r*, against the QT of the trivial construction [22, Ex. 2].*

Proof. Completeness holds because the r queried values of an honest encoding are $\delta_j|_{U_j}$, whose interpolant of degree less than r is δ_j. For consistency, a non-$\perp$ output means that the r pairs $(x, v_x)_{x\in U_j}$ form an accepting opening set under C_j; outside the event of Theorem 2 they agree with $c^\star = \mathrm{ev}_A(f^\star)$, so $\widehat{\delta}_j = f^\star \bmod h_j$ by Theorem 1(ii), and the returned value is the coordinate of the single message $\Psi^{-1}(f^\star)$ that (5) assigns to i. Withheld or failing answers force $\perp$, which is not an inconsistent output. □

The claim concerns this message representation, not the code. Local accessibility itself is obtained in [22, §3.2] from generalized systematic encodings; Proposition 11 shows that the coordinates of Proposition 2 retain it for every k, including the residual block. The number of certified groups is also the natural measure of progress for a collector: the certified restrictions have rank $\min(k, hr)$ at h certified groups, so the object is determined once $h \geq H_\star$.

E.4 Specification Tests and the Mutation Condition

The enumeration referred to in Section 8 is deterministic: for each $(q, m, \ell_{\max})$ in $\{(17, 4, 4), (97, 8, 6), (97, 12, 8), (193, 16, 8)\}$ we take every ℓ with $2 \leq \ell \leq \ell_{\max}$, every r with $2 \leq r < m$, and every k with $r \leq k \leq \ell r$, which gives 5,791 parameter sets, 978 with $b = 0$ and 4,813 with $b > 0$; each set is tested on one message drawn from a fixed seed. No further filtering is applied, since $k \leq \ell r$ and $b > 0$ force $a < \ell$, so a residual group $j_\star \notin J_{\mathrm{src}}$ always exists.

Suppose now that Com omits the conversion (7) and uses the last message block v directly as e, so that it commits to $f = \Phi(\boldsymbol{u}, v)$. By Proposition 1 the restriction of that codeword to $A_{j_\star}$ is $F_\star + B(\gamma_{j_\star})v$, and the extractor of Construction 1, which returns Ψ^{-1}, reports $\widetilde{v} = \mathrm{Low}_{<b}(F_\star) + B(\gamma_{j_\star})v$. The round-trip decoding condition is therefore satisfied exactly by messages fulfilling

$$\mathrm{Low}_{<b}(F_\star) + \big(B(\gamma_{j_\star}) - 1\big)v = 0, \tag{16}$$

which is b linear conditions on $(\boldsymbol{u}, v)$. Because $j_\star \notin J_{\mathrm{src}}$, every $L_s(\gamma_{j_\star})$ is nonzero, so for uniform $\boldsymbol{u}$ the term $\mathrm{Low}_{<b}(F_\star)$ is uniform on $\mathbb{F}_q^b$ and independent of v; the left side of (16) is therefore uniform regardless of $B(\gamma_{j_\star})$, and a message satisfies this condition with probability q^{-b}. Summing q^{-b} over the 4,813 sets with $b > 0$ predicts 6.62 survivors. At $(q, m, \ell, r, k) = (17, 4, 4, 3, 4)$ with $\gamma_0 = 1$ and $\gamma_{j_\star} = 16$, (16) reads $u_{0,0} + 14v_0 = 0$ and is satisfied by 17^3 of the 17^4 messages.

F Collector Case Study and Repair-Oriented Comparison

Each certified group lowers the number of values still to be received by m positions and the total by α (Theorem 5), enables local serving for that group (Proposition 4), and the stopping point $H_\star$ is known in advance. The *sequential group policy* visits groups in uniformly random order and, inside a group, queries positions without replacement, leaving the group as soon as r openings are accepted or $\alpha + 1$ have failed, so that a group is decided within m queries;

in *recovery mode* the collector stops at $|J_{\text{cert}}| = H_\star$, and in *coverage mode* it continues until its query budget is exhausted. Every response is verified on arrival and therefore carries its own proof, which is why the traffic of this policy is the 4.80 MiB row of Table 5 rather than the 1.92 MiB of whole-group verification. We simulate three patterns of missing positions over 200 runs: *S1* withholds 41 groups entirely, leaving exactly $H_\star = 82$ recoverable; *S2* erases 200 positions in each of 82 groups and $300 > \alpha$ positions in each of the other 41; and *S3* erases each position independently with probability 0.2.

Table 9. Sequential group policy at the reference profile, 200 runs. Recovery mode reports the number of queries at which $|J_{\text{cert}}|$ first reaches $H_\star = 82$; coverage mode reports the mean $|J_{\text{cert}}|$ at budget exhaustion. The two modes are different executions.

	Queries to $H_\star$			Mean $\lvert J_{\text{cert}}\rvert$ at budget B		
Pattern	mean	min	max	50k	100k	126k
S1: 41 groups withheld	73,381	72,742	73,513	55.4	82.0	82.0
S2: partial losses	113,760	110,493	114,381	35.7	71.5	82.0
S3: independent, $X = 0.2$	78,722	78,284	79,821	51.8	103.8	123.0

The measurements of Table 9 match the closed-form predictions. In S1 every group must be visited, so a run costs at most $82 \cdot 768 + 41 \cdot 257 = 73{,}513$ queries, the observed maximum, and $82 \cdot 768 + \frac{41 \cdot 82}{83} \cdot 257 \approx 73{,}386$ in expectation. In S3, certifying $H_\star$ groups requires 62,976 accepted openings at a failure rate of 0.2, hence about 78,720 queries; in S2, a recoverable group costs about 186 failed openings in addition to its 768 accepted ones. The collector can serve each certified group from its own evidence and the header (Theorem 2, Proposition 4); the study illustrates the policy and measures no running times.

F.1 Position Relative to Repair-Oriented Sampling

Local repair of data and of opening proofs is not by itself a distinguishing feature: Boneh, Neu, Nikolaenko, and Partap [6] formalize local repair and construct a scheme from locally correctable multiplicity codes, with a multivariate commitment that opens partial derivatives, batches the generation of opening proofs, and recomputes the opening proof at one point from a few other proofs; they compare their scheme with PeerDAS in storage overhead, local repair bandwidth, and coordination. Short reference strings, low assumption degree, and dedicated generation represent distinct design dimensions, and we do not claim that either underlying cryptographic assumption is strictly weaker. Table 10 lists the axes on which the two designs differ; its entries for [6] describe the interface, not measured quantities.

The distinguishing property here is the absence of per-group auxiliary state after recovery; which interface is preferable depends on whether isolated erasures or correlated losses of whole blocks dominate.

Table 10. Interface comparison with repair-oriented sampling. Entries for the present work refer to Construction 2 under Profile A.

Axis	[6]	This work
Code	locally correctable multiplicity code	Tamo–Barg code, optimal distance for the prescribed groups
Commitment	multivariate, with partial-derivative openings and batched proof generation	univariate KZG; every group commitment derived from the header
Repair granularity	one coordinate, together with its opening proof	one group of m coordinates, from any r accepted symbols of it
Input to regenerating a proof	a few opening proofs at other points	the r field elements, the header, and $(\tau^i G_1)_{i<r}$
Per-unit auxiliary state	the assisting proofs must be obtainable	none
Prescribed recovery groups per symbol	several	one group of size m, any r of whose symbols suffice
Global membership	not needed: repair is local correction	structural, from the extracted source tuple
Setup	trusted setup for the multivariate scheme	dedicated univariate string of degree r (Proposition 3)
Reported evaluation	measured, against PeerDAS	analytic and measured, against the source-compressed product code

Other recovery structures. Codes with several disjoint recovering sets [33, §IV] and rack-aware repair of Tamo–Barg codes [9] concern a fixed honest codeword; whether they extend to our adversarial commitment model is open.

G Experimental Code

The code used for the experiments in this paper is publicly available in an anonymous repository at `https://anonymous.4open.science/status/fc27-440`.