

Capacity-Approaching Pseudorandom Code for Stochastic Edit Channels

Leshui Wang¹ and Kenji Yasunaga¹

Institute of Science Tokyo, 2-12-1 Ookayama, Meguro, Tokyo, Japan
wang.l.e512@m.isct.ac.jp
yasunaga@comp.isct.ac.jp

Abstract. Pseudorandom codes (PRCs) are error-correcting codes whose codewords are computationally indistinguishable from uniformly random strings to any efficient adversary. Most prior PRC work focuses on Hamming error channels; nevertheless, the few proposed constructions on edit channels are not optimized for a stochastic channel with both Hamming and synchronization errors. We study PRC over a stochastic memoryless Binary Substitution-Insertion-Deletion (BSID) channel [1] and its generalized q -ary version. At each current input symbol, the channel inserts a uniformly random symbol without advancing with probability P_i , deletes and advances with probability P_d , or transmits and advances otherwise; a transmitted symbol is then replaced uniformly by another alphabet symbol with probability P_s .

We first prove that the majority based PRC construction of Christ and Gunn [3] remains robust over the BSID channel whenever $1 - P_i - P_d > 0$ and $P_s < 1/2$, by proving its majority decoding after the BSID channel recovers more than half of the original coordinates by a constant margin with overwhelming probability. Then we propose a unique-decodable public-key multi-bit PRC on a constant alphabet constructed by letting the Christ–Gunn majority component protect control information and the Haeupler–Shahrasbi synchronization/list-recovery construction [8] encodes the high-rate payload. For every sufficiently large admissible constant alphabet q depending on the channel parameters, the PRC is robust against a q -ary SID channel bounded by $P_i + P_d < 1$, $P_s < 1$, and its rate approaches the genie-aided capacity upper bound of the channel when q approaches infinity.

1 Introduction

1.1 Motivation

We study a stochastic edit channel that introduces substitutions, insertions and deletions. In the real world, similar channels are widely encountered when communicating in public environments, for example, in wireless communication in low signal-to-noise-ratio or electromagnetically complex environments. Such a channel’s error model was first modeled in Briffa et al.’s work [1] as an expansion of the Davey–MacKay Insertion Deletion channel proposed in [2], and was given the name Binary Substitution Insertion Deletion (BSID) channel.

Assume Alice and Bob want to communicate through such an unreliable public channel, requiring not only confidentiality of the message, but also covertness of the communication itself: the very act of sending a meaningful message should remain hidden. To achieve this, we hope that the message transmitted has error-correcting ability and, simultaneously, does not contain any recognizable structure.

Introduced by Christ and Gunn [3], pseudorandom code (PRC) provide a natural cryptography primitive for this setting. PRC is a type of error-correcting code that is computationally indistinguishable from uniform randomness. Per [3], besides error-correcting ability (robustness) and pseudorandomness, PRC should also satisfy soundness, meaning a low probability of successfully decoding random fixed strings.

After Christ and Gunn [3]’s work, there has been a substantial amount of research on PRC; however, most of the research has focused on PRCs robust against Hamming-type errors, while synchronization error destroys their error-correcting structure. Four papers proposed constructions of zero-bit and multibit PRCs against adversarial or oblivious edit channels that can be used to correct stochastic Hamming and synchronization errors [3] [6] [7] [4], among them of these works provide multi-bit PRC constructions [3] [4]. However, these results do not characterize the achievable rate for the memoryless SID model considered here.

Additionally, a genie-aided capacity approaching multi-bit PRC against stochastic Hamming and synchronization errors could potentially serve as the base code of future high rate multi-bit PRC against oblivious and adversarial Hamming and synchronization errors. Historically, for Hamming error, Lipton proposed in [10] that worst case error, after evenly dispersed by a secret coordinate permutation, would look random and become decodable by stochastic channel oriented codes with high probability; based on this method, Guruswami and Smith in [11] mixed the permutation seed within the codeword to protect them, creating their capacity approaching codes against oblivious and adversarial Hamming error. If a similar-in-function method that smears both Hamming and synchronization error evenly throughout the codeword was discovered, applying it on our research could potentially make a high rate PRC against oblivious and adversarial Hamming and synchronization channel.

1.2 Our Contributions

This paper focuses on pseudorandom code for BSID and q -ary SID channels, a generalized version of BSID channel. The two channels are modeled in the same manner: for each current input symbol, the channel inserts a uniformly random bit with probability P_i without advancing the input, deletes the current bit with probability P_d and advances to the next input symbol, or transmits it with probability $1 - P_i - P_d$ and advances. Conditioned on transmission, the transmitted bit is independently flipped with probability P_s . Our two contributions are summarized as follows:

- In Section 3, we prove that the majority vote based construction of Christ and Gunn Construction 6 [3], previously analyzed for a binary deletion channel (BDC) followed by a binary substitution channel (BSC), remains robust over the BSID channel whenever $1 - P_i - P_d > 0$ and $P_s < 1/2$. The proof controls the insertion-induced displacement of all majority-block boundaries, computes the surviving majority correlation under substitutions, and converts the resulting local dependence into an overwhelming-probability global agreement bound.
- In section 4, we construct a polynomial-time unique-decodable public-key pseudorandom code for the q -ary SID channel bounded by $1 - P_i - P_d > 0, P_s < 1$. The construction uses the CG24 component to communicate the seeds of a pseudorandom mask and self-matching string, key and message's hash value of an almost-universal hash [12], while the payload is encoded by the high rate synchronization and list-recovery construction of Haeupler and Shahrasbi [8]. We proved the code's robustness, pseudorandomness, soundness, and rate. For fixed channel parameters and every sufficiently large admissible alphabet, the construction achieves a rate at least $(1 - \delta - \varepsilon)^{\frac{\log(q/2)}{\log q}}$ where δ is the fraction of deletion plus substitution, hence its unique-decoding rate approaches the genie-aided capacity upper bound as q grows and the parameter slack tends to zero.

Because the majority based construction, the self-matching string decoding on HRW code, and the almost-universal hash are all unconditional, beside those required by the underlying binary multi-bit PRC robust against key-independent p -bounded channel in the modular Christ and Gunn Construction 6 [3], the paper will only use assumption on standard pseudorandom generators (PRGs).

1.3 Literature Review

We are aware of two prior works giving multibit PRC constructions robust against edit channels: The first official PRC paper, by Christ and Gunn [3], provides Construction 6, a binary zero-bit/multibit public- or secret-key PRC over a stochastic memoryless channel $\text{BSC} \circ \text{BDC}$. Its majority-vote encoding limits the rate to $o(1)$, but the construction is robust against any constant fraction of deletions and any substitution fraction below $1/2$. Although it was originally proved over $\text{BSC} \circ \text{BDC}$, this construction is highly compatible with the BSID channel: we prove in this paper that it is robust with overwhelming probability for any BSID channel satisfying $1 - P_i - P_d > 0$ and $P_s < 1/2$.

A very recent work by Huang et al. [4] provides strong results on public-key PRC constructions against oblivious edit channels. Using a technique that mixes control information carried by zero-bit PRCs within the codeword, the paper proposes three multibit PRC constructions for edit channels, including a binary concatenated-code construction, a constant-sized-alphabet synchronization-string construction robust against sublinear oblivious edits, and another synchronization-string construction over an alphabet of size $\text{poly}(\lambda)$ that has rate $1 - \varepsilon$ and is

Table 1. Comparison of prior multibit PRC constructions against edit channels with our construction.

Work	Rate	Alphabet	channel model	error bound	PK/SK
CG24	$o(1)$	binary	$\text{BSC}_q \circ \text{BDC}_p$	$p < 1 - \varepsilon, q < 1/2$	PK/SK
HLMZ26	$1 - \varepsilon$	$\text{poly}(\lambda)$	oblivious	$O(n)$	PK/SK
HLMZ26	$1 - \varepsilon$	$\text{poly}(1/\varepsilon)$	oblivious	$O(n^{1-c})$	PK/SK
HLMZ26	$1/2 - \varepsilon$	binary	oblivious	$O(n^{1-c})$	PK/SK
Theorem2	$(1 - \delta - \varepsilon) \frac{\log(q/2)}{\log q}$	$q(P_i, P_d, P_s)$	q -ary SID	$1 - P_i - P_d > 0, P_s < 1$	PK

robust against an $(\varepsilon - \eta)n$ fraction of oblivious edits. If the $\text{poly}(\lambda)$ -sized alphabet is not considered a cost, the last construction is almost optimal in that it approaches the Singleton tradeoff in its stated parameter regime, correcting $(\varepsilon - \eta)n$ at rate $1 - \varepsilon$. However, in the context of the q -ary SID channel, the fraction of edits can exceed 1, which this construction cannot handle, and a better rate is expected because the Singleton bound is no longer relevant to such a channel.

Complementing the constructive results above, Dehmel and Li [5] recently established an LCS-based impossibility result for fixed-alphabet multibit public-key PRCs against public-key dependent adversarial deletion channels, that no single-output PRC can be robust against all such δ -deletion channels for $\delta > 1 - \gamma_q^{\text{LCS}}$, and extend the obstruction to constant-size list decoding. Their adversarial channel depend on both the public key and the transmitted codeword, although their bound does not apply to stochastic or other oblivious channels.

As a relevant topic, related zero-bit PRC constructions against edit channels include: Golowich and Moitra’s [6] secret-key zero-bit PRC construction over an alphabet of size $\text{poly}(\lambda)$ achieves a relatively strong result over adaptive PPT edit channels, with the edit radius theoretically approaching 1. Their technique converts a Hamming-error-oriented PRC into an edit-error-robust PRC by encoding the index of the support of its original binary codeword into a large-alphabet codeword. While Golowich and Moitra’s [6] construction was presented mainly in the secret-key formulation, their paper stated that the arguments extend identically to public-key PRCs. Christ et al. [7] proposed a secret-key binary zero-bit PRC construction robust against an adaptive PPT adversarial edit channel with a constant error rate bounded by $1/400$. The construction is based on the permuted-codes assumption introduced in the same paper. Error correction is achieved through random indexing, exhaustive enumeration within a small edit ball around each potentially corrupted index and symbol, and list recovery. Huang et al.’s work [4] also contains a public-key binary zero-bit PRC construction against oblivious edit channels with a sublinear error rate. The code achieves edit-error correction using CGK embedding.

1.4 Technical Overview

Initially, we identify a high-performance PRC specially designed for a stochastic memoryless edit channel as a gap to fill in current PRC research. The BSID channel proposed by Briffa, Schaathun, and Wesemeyer [1], with its generalized q -ary SID version, is a rather natural stochastic edit channel to work on. From our literature review, we found that CG24 Construction 6 [3], a PRC for the deletion channel, can potentially survive a BSID channel: the channel has a consistent expected output length per unit input length, with an asymptotically smaller fluctuation in the actual output length; hence, with high probability, the equal-length partitioning of the received string should recover corrupted subblocks of the original codeword with only a vanishing fraction of positions misaligned. Moreover, the correlation between the majority-encoded symbol and the majority vote of the recovered block should remain a positive constant as long as a positive proportion of symbols from the original codeword survive and are correctly positioned. We formalize this intuition with a proof as follows: Bernstein’s inequality bounds the cumulative output length deviation at any position by $u = O(\sqrt{mn} \log n)$ from expected value with overwhelming probability, and $m = \Omega(n \log^6 n)$ makes $u = o(m)$, thus each reconstructed majority vote block from equal length partitioning the received string contains $am + o(m)$ surviving positions from its corresponding original majority vote block, where $am = (1 - P_i - P_d)m/(1 - P_i)$ is the expected number of survived positions from the original block. After accounting directly for substitutions over survived positions, the correlation coefficient between the original and reconstructed majority statistics converges to $\rho = (1 - P_i - P_d)(1 - 2P_s)/\sqrt{(1 - P_i)(1 - P_d)}$, which is positive when $1 - P_i - P_d > 0$ and $P_s < 1/2$. A bivariate central limit theorem then gives majority agreement probability $1/2 + \arcsin(\rho)/\pi + o(1)$. Finally, although neighboring agreement events can be correlated, separating them into three residue classes after fixing the channel geometry makes the events within each class independent, so Chernoff bounds could show that more than half of all coordinates are recovered by a constant margin with overwhelming probability.

The CG24 construction is proved to survive the BSID channel, but due to the internal parameter requirement of $m = \Omega(n \log^6 n)$, where the block length equals mn , the rate essentially approaches zero when the block length increases, so we need a high-rate wrapper around it to boost the overall rate. To obtain a non-vanishing rate, we combine this as control component with the large-alphabet list-decodable construction of HS21 Theorem 2.8 [8], which, when the deletion fraction is δ , has a rate arbitrarily close to $1 - \delta$. Combining the two constructions, with large enough constant alphabet $q(P_i, P_d, P_s)$, we built a public key multi-bit PRC robust against q -ary SID channel bounded by $1 - P_i - P_d > 0, P_s < 1$ and approaches the channel’s capacity upper bound. In our construction, the CG24 code is used to carry 4 equal length control information strings including two fixed length seeds b of pseudorandom mask R and pseudorandom self-matching string S , and almost-universal hash component including a hash key and the hash value of m , while the HS21 code encodes message m with the Hemenway–Ron–Zewi–Wootters (HRW) list-recoverable code [9] and

carries the code symbol x . The transmitted symbol is $c_i = (x_i + R_i, S_i, b_i)$, and the alphabet size of c_i is q , satisfying $q \geq \left(\frac{1+\gamma}{\varepsilon^2}\right)^{O(\frac{1+\gamma}{\varepsilon^3})}$, $q > \frac{1}{(1-P_s)}, 2 \mid q$ with $\gamma = \frac{P_i + (1-P_i-P_d)P_s}{1-P_i} + \eta$. Under a uniformly chosen q -ary substitution, b_i flips with probability $qP_s/(2(q-1)) < 1/2$ whenever $P_s < 1$ and $q > 1/(1-P_s)$, plus the channel is bounded by $1 - P_i - P_d > 0$, the first part of the proof recovers the seeds and hash components with overwhelming probability. Once the seeds are recovered, the decoder regenerates R and S . Similarly, substitution changes the $(x_i + R_i, S_i)$ projection with probability $r = (q-2)/(q-1)$, although, to simplify the parameters we set up the embedded HS21 code by assuming $r = 1$, or one global q -ary substitution on c_i alters $(x_i + R_i, S_i)$ with probability equals to 1. Treating each such substitution change as one deletion and one insertion places the projected channel within the chosen HS21 code deletion and insertion parameters. The self-matching string generated from S then lets HS21 algorithm reposition the received symbols into bounded positional candidate sets, after which the decoder unmask the candidates with recovered R and applies HRW list recovery. This results a list containing the transmitted message, of size at most $\exp(\exp(\exp(\log^* N)))$, from which, with overwhelming probability, after looking for any entry in the list who produces the recovered hash value when hashed by the recovered hash key, the correct m will be the eventual output. The random mask, pseudorandom self-matching string, and binary PRC together ensure pseudorandomness, while soundness follows from the binary PRC. The final rate is at least $(1 - \delta - \varepsilon) \log(q/2)/\log q$, where $\delta = \frac{P_d + (1-P_i-P_d)P_s}{1-P_i} + \eta$ resembling the expected fraction of deletion plus substitution, and for fixed channel parameters the resulting rate approaches the genie-aided capacity upper bound in the large-alphabet limit with arbitrarily small parameter slack.

2 Preliminaries

2.1 Notation

- *Indices.* For $n \in \mathbb{N}$, let $[n] := \{1, \dots, n\}$.
- *Sampling and concatenation.* We write $x \leftarrow \mathcal{D}$ for sampling from a distribution $\mathcal{D}$, $x \xleftarrow{\$} S$ for uniform sampling from a finite set S , and $x \parallel y$ for string concatenation.
- *Distances.* For equal-length strings, $\text{dist}(x, y)$ denotes their Hamming distance; for sets A, B , $A \Delta B$ is their symmetric difference.
- *Asymptotic notation.* All logarithms are base two unless a base is displayed. A function $\text{negl}(t)$ is negligible if it is smaller than t^{-c} for every constant $c > 0$ and all sufficiently large t ; an event holds with overwhelming probability if its complement has negligible probability. We write $\text{poly}(t)$ for an unspecified polynomial in t .
- *Computational indistinguishability.* We write $X \approx_c Y$ for computational indistinguishability.

2.2 Hash Function

ε_H -Almost-Universal Hashing [12] A ε_H -almost-universal hash family $\mathcal{H} = \{H_\tau : \mathcal{X} \rightarrow \mathcal{Y}\}_{\tau \in \mathcal{K}}$ is ε_H -almost-universal (ε_H -AU) if, for every distinct $x, x' \in \mathcal{X}$,

$$\Pr_{\tau \leftarrow \mathcal{K}} [H_\tau(x) = H_\tau(x')] \leq \varepsilon_H.$$

We use polynomial evaluation hashing over $\mathbb{F}_{2^\kappa}$ introduced in [13]. For $x = (x_0, \dots, x_{d-1}) \in (\mathbb{F}_{2^\kappa})^d$ and $\tau \leftarrow \mathbb{F}_{2^\kappa}$, define

$$H_\tau(x) = \sum_{i=0}^{d-1} x_i \tau^i.$$

This family is $\varepsilon_H = (d-1)/2^\kappa$ -almost-universal. When $d = \text{poly}(\kappa)$, $\varepsilon_H = \text{negl}(\kappa)$.

2.3 Pseudorandom Objects

Pseudorandom Generators A PRG is a polynomial-time map $G : \{0, 1\}^\kappa \rightarrow \Sigma^N$ whose output on a uniform seed is computationally indistinguishable from uniform over Σ^N .

Pseudorandom Codes We follow the definition of public-key pseudorandom codes of Christ and Gunn [3]. A public-key PRC over an alphabet Σ consists of polynomial-time randomized algorithms

$$(\text{Init}, \text{Enc}, \text{Dec}),$$

where $\text{Init}(1^\lambda)$ outputs (pk, sk) , $\text{Enc}_{pk}(m) \in \Sigma^N$, and $\text{Dec}_{sk}(c)$ outputs either $\perp$, or a message. We use the following notions.

- *Pseudorandomness*. For every PPT adversary given pk , adaptive oracle access to $\text{Enc}_{pk}(\cdot)$ is computationally indistinguishable from access to an ideal oracle that returns a fresh independent uniform element of Σ^N on every query, including repeated queries.
- *Soundness*. For every fixed $c \in \Sigma^*$,

$$\Pr_{(pk, sk) \leftarrow \text{Init}(1^\lambda)} [\text{Dec}_{sk}(c) \neq \perp] = 1 - \text{negl}(\lambda).$$

- *Robustness*. For a channel $\mathcal{E}$, for every message m ,

$$\Pr[m = \text{Dec}_{sk}(\mathcal{E}(\text{Enc}_{pk}(m)))] \geq 1 - \text{negl}(\lambda),$$

where the probability is over key generation, encoding, the channel, and decoding randomness.

- *Rate*. The rate is

$$R = \frac{\log |\mathcal{M}|}{N \log |\Sigma|}.$$

2.4 The Majority Wrapper

Majority Encoding For odd m , $\text{MajEnc}_m(x_1, \dots, x_n)$ independently samples, for every j , a uniform block in $\{0, 1\}^m$ conditioned on having majority x_j , and concatenates the blocks.

Majority Decoding Given an arbitrary received string, MajDec_n partitions it into n consecutive blocks whose lengths differ by at most one and outputs their majorities, using a fixed rule for ties.

p -Bounded Channels

Definition 1 (p -bounded channel). A length-preserving binary channel $\mathcal{E} : \{0, 1\}^* \rightarrow \{0, 1\}^*$ is p -bounded if, for every $n \in \mathbb{N}$,

$$\Pr_{x \leftarrow \{0, 1\}^n} [\text{dist}(\mathcal{E}(x), x) > pn] \leq \text{negl}(n),$$

where the probability is taken over the uniform choice of x and the internal randomness of $\mathcal{E}$.

2.5 Synchronization and List Recovery

List Recovery Let $\mathcal{C} \subseteq \Sigma_C^N$ be a code. It is (a, l, L) -list-recoverable if, given sets $A_1, \dots, A_N \subseteq \Sigma_C$ of size at most l , its list-recovery algorithm outputs at most L codewords and includes every $c \in \mathcal{C}$ satisfying $|\{i : c_i \in A_i\}| \geq aN$.

Self-Matching Strings A monotone matching of a string $S \in \Sigma_S^N$ with itself consists of pairs (i_r, j_r) , $r \in [t]$, such that both index sequences are strictly increasing and $S_{i_r} = S_{j_r}$. The string is ε_s -self-matching if every such matching contains fewer than $\varepsilon_s N$ pairs with $i_r \neq j_r$.

LCS Repositioning We write ListRep_K for the K -round longest-common-subsequence (LCS) repositioning procedure of Haeupler and Shahrabi [8], which uses an index string to produce at most K candidate symbols for each transmitted position. Their Theorem 2.8 combines this procedure with a high-rate list-recoverable code: for constants $0 \leq \delta < 1$, $\gamma \geq 0$, and $\varepsilon > 0$, it gives a polynomial-time code tolerating δN deletions and γN insertions, of rate at least $1 - \delta - \varepsilon$ over a constant-size alphabet, with list size $L(N) = \exp(\exp(\exp(\log^* N)))$.

HS21 repositioning guarantee. The analysis in the proof of Theorem 2.8 of Haeupler and Shahrabi [8] gives the following interface. Let $((u_i, S_i))_{i \in [N]}$ be an indexed sequence, where $S \in \Sigma_S^N$ is ε_s -self-matching. If the sequence undergoes at most δN deletions and γN insertions, then ListRep_K outputs candidate sets $A_1, \dots, A_N$ satisfying $|A_i| \leq K$ and

$$|\{i \in [N] : u_i \in A_i\}| \geq \left(1 - \delta - \frac{1 + \gamma}{K} - K\varepsilon_s\right) N. \quad (1)$$

2.6 Channel Model

We consider a q -ary generalization of the Binary Substitution, Insertion, and Deletion (BSID) channel of Briffa, Schaathun, and Wesemeyer [1], derived from the random insertion/deletion channel of Davey and MacKay [2]. Let Σ be an alphabet of size $q \geq 2$ and let $x = (x_1, \dots, x_N) \in \Sigma^N$. The channel has parameters $P_i, P_d, P_s \in [0, 1]$ with $P_i + P_d < 1$. At the current input position i , independently of previous events, it performs one of three actions:

- With probability P_i , it outputs $u \xleftarrow{\$} \Sigma$ and remains at position i .
- With probability P_d , it deletes x_i and advances to $i + 1$.
- With probability $1 - P_i - P_d$, it transmits x_i and advances to $i + 1$.

Since an insertion does not advance the input position, arbitrarily many inserted symbols may precede any deletion or transmission.

Conditioned on transmission, the output equals x_i with probability $1 - P_s$; for every $a \in \Sigma \setminus \{x_i\}$, it equals a with probability $P_s/(q - 1)$. Thus a substitution replaces x_i uniformly by one of the other $q - 1$ symbols. The channel maps $x \in \Sigma^N$ to a random output $y \in \Sigma^{N'}$ of random length N' . For $q = 2$, every substitution flips the transmitted bit and the model reduces to the binary BSID channel; we call its non-binary extension the q -ary *SID channel*.

3 The CG24 Code over the BSID Channel

Christ and Gunn construct a pseudorandom code for the deletion channel by composing a PRC robust to bounded Hamming errors with the majority encoder and decoder (MajEnc, MajDec) [3]. Their original analysis proves this construction robust over the composed channel $\text{BSC}_q \circ \text{BDC}_p$, which independently deletes bits and then flips surviving bits, but does not account for insertions. We instead consider the BSID channel, a natural memoryless channel in which insertions, deletions, and substitutions occur within a single process. Since inserted bits change the received length and shift the block boundaries used by MajDec, the CG24 robustness result does not apply directly. We show that, when $1 - P_i - P_d > 0$, $P_s < \frac{1}{2}$, and $m = \Omega(n \log^6 n)$ is odd, the majority code still recovers more than half of the original coordinates by a constant margin with overwhelming probability. This is the key step that extends the CG24 construction to the BSID channel and is stated formally in the following theorem.

Theorem 1 (Majority-Voted Code Survives BSID). *Let $\mathcal{E}$ be a BSID channel with parameters $\{P_i, P_d, P_s\}$, and let $x \leftarrow \{0, 1\}^n$. If $1 - P_i - P_d > 0$, $P_s < \frac{1}{2}$, and $m = \Omega(n \log^6 n)$ is odd, then there exists a constant $\varepsilon > 0$ such that*

$$\Pr[\text{dist}(x, \text{MajDec}_n(\mathcal{E}(\text{MajEnc}_m(x)))) \leq (\tfrac{1}{2} - \varepsilon)n] \geq 1 - \text{negl}(n).$$

3.1 Channel Geometry and Block Alignment

We begin the proof by letting $Y \in \{0, 1\}^{mn}$ be $\text{MajEnc}_m(X)$ and considering what happens when Y goes through the BSID channel $\mathcal{E}$. By Lemma 1 below, Y is also a uniform string. The i -th bit of Y is labeled as Y_i ; when the memoryless channel $\mathcal{E}$ takes Y_i in, the output would be some inserted uniform bits in the front, concatenated with the original Y_i with probability $\frac{(1-P_i-P_d)(1-P_s)}{1-P_i}$, or a flipped Y_i with probability $\frac{(1-P_i-P_d)P_s}{1-P_i}$, or nothing if Y_i is deleted with probability $\frac{P_d}{1-P_i}$. For every Y_i we name the number of inserted uniform bits as K_i , where $\Pr[K_i = k] = (1 - P_i)P_i^k$, and name the concatenated string channel $\mathcal{E}$ outputs on input Y_i as G_i , so $\mathcal{E}(Y) = G_1 \parallel G_2 \parallel \dots \parallel G_{mn} = G$. For convenience, we name $G_{(j-1)m+1} \parallel G_{(j-1)m+2} \parallel \dots \parallel G_{jm}$ as H_j , to represent the section of G originated from X_j .

Lemma 1 (Uniformity of Majority Encoding). *For uniformly random $X \in \{0, 1\}^n$ and odd m , $\text{MajEnc}_m(X)$ is uniformly distributed over $\{0, 1\}^{mn}$. Hence its mn coordinates may be regarded as i.i.d. uniform bits.*

Proof. For odd m , exactly 2^{m-1} strings in $\{0, 1\}^m$ have majority 0, and exactly 2^{m-1} have majority 1. The encoder samples each block uniformly conditioned on its majority being the corresponding bit of X . Averaging over a uniform input bit therefore makes each block uniform over $\{0, 1\}^m$; independence across the input bits and encoder samples makes the concatenation uniform over $\{0, 1\}^{mn}$. This is also the pseudorandomness observation in the proof of Theorem 4 of Christ and Gunn [3]. $\square$

$\text{MajDec}_n(G)$ will first partition G into n equal length blocks, which we call $S_1, S_2, \dots, S_n$, with length deviation of 1 at most, then it will solve the majority symbol within each S_j and output them in sequence. To let $\text{dist}(\text{MajDec}_n(G), X) \leq (\frac{1}{2} - \epsilon)n$ with overwhelming probability, the sufficient condition is that $\Pr[\text{MajDec}_1(S_j) = X_j] \geq \frac{1}{2} + \gamma$ where $\gamma > \epsilon$, plus the correlation between $\text{MajDec}_1(S_j) = X_j$ agreements must not be too far from independence.

We explore two properties to determine whether the sufficient condition of Theorem 1 above is met. First, how much do S_j and H_j overlap and misalign? Second, under such a misalignment, what is the probability that $\text{MajDec}_1(S_j) = X_j$? Besides directly influencing this probability, the misalignment also determines the correlation between agreement events, because correlation arises when a misaligned S_j overlaps with H_k for $k \neq j$. In other words, if $S_j = H_j$ for every j , then the events $\text{MajDec}_1(S_j) = X_j$ are independent because the X_j are mutually independent and each H_j depends only on its corresponding input block.

To reveal the misalignment between S_j and H_j , we check H_j 's property first. We define the estimated length of every G_i as $\mu := \mathbb{E}[|G_i|] = \mathbb{E}[K_i] + \mathbb{E}[\text{Ber}(\frac{1-P_i-P_d}{1-P_i})] = \frac{1-P_d}{1-P_i}$, and the cumulative length from G_1 to G_t as $C_t := \sum_{i=1}^t |G_i|$. The only factor causing the misalignment between S_j and H_j is the accumulated deviation between the actual $|G_i|$ and μ : if every $|G_i|$ equaled μ , then every H_j would have the fixed length μm and would be parsed precisely as S_j by MajDec . Here we measure the deviation of $|G_i|$ from μ first.

Lemma 2 (Channel Output Length per Input Bit). *There exists constant C and c , such that for every integer $x \geq 1$ $\Pr[|G_i| - \mu \geq x] \leq Ce^{-cx}$.*

Proof. The expression of $|G_i| - \mu$ is as follows: $|G_i| - \mu = |G_i| - \mathbb{E}[|G_i|] = K_i + \text{Ber}\left(\frac{1-P_i-P_d}{1-P_i}\right) - \mathbb{E}[|G_i|]$. Within G_i , K_i follows a geometric distribution, where $\Pr[K_i \geq k] = P_i^k$, thus for $0 < P_i < 1$, $\Pr[|G_i| - \mu \geq x] < \Pr[K_i \geq x-1] = P_i^{x-1} = \frac{1}{P_i} e^{-\log(1/P_i)x}$; for $P_i = 0$, the probability is outright zero. $\square$

With the zero mean and subexponential nature of $|G_i| - \mu$, we can now research the behavior of accumulated deviation $C_t - t\mu$ by applying Bernstein's Inequality.

Lemma 3 (Channel Output Length per Input String). *There exists a constant $c > 0$ such that, for $u := c\sqrt{mn} \log n$ and all j ,*

$$\Pr[|C_{jm} - jm\mu| \geq u] = \text{negl}(n).$$

Proof. Because $|G_i| - \mu$ is subexponential, its sub-exponential norm is constant, so we let $k := \| |G_i| - \mu \|_{\psi_1}$. We have $C_t - t\mu = \sum_{i=1}^t (|G_i| - \mu)$. Per Bernstein's Inequality, there exist a constant c_1 such that

$$\Pr\left[\left|\sum_{i=1}^{jm} (|G_i| - \mu)\right| \geq c_2 \sqrt{mn} \log n\right] \leq 2 \exp\left(-c_1 \min\left\{\frac{c_2^2 n \log^2 n}{jk^2}, \frac{c_2 \sqrt{mn} \log n}{k}\right\}\right) = \text{negl}(n).$$

$\square$

Next, we locate each S_j and compare it with H_j , thereby quantifying their misalignment.

Lemma 4 (S_j, H_j Misalignment). *For all j , the length $|S_j \Delta H_j|$ of the misaligned portion of S_j and H_j satisfies*

$$\Pr[|S_j \Delta H_j| \leq O(u)] \geq 1 - \text{negl}(n).$$

Proof. Let E_j be the j -th boundary produced by $\text{MajDec}_n(G)$ partitioning G . Since it is an equal partition but $|G|$ is not always divisible by n , we have $\left|E_j - \frac{j|G|}{n}\right| \leq 1$. Per Lemma 3, since $||G| - \mu mn| \leq u$ with overwhelming probability, we can estimate the distance between the right boundaries of S_j and H_j , which are E_j and C_{jm} :

$$|E_j - C_{jm}| \leq \left|E_j - \frac{j|G|}{n}\right| + \frac{j}{n} ||G| - \mu mn| + |\mu jm - C_{jm}| \leq 2u + 1.$$

Therefore,

$$|S_j \Delta H_j| \leq |E_{j-1} - C_{(j-1)m}| + |E_j - C_{jm}| \leq 4u + 2 = O(u).$$

With the length of misalignment between S_j and H_j measured, to evaluate the probability of $\text{MajDec}_1(S_j) = X_j$, we need to determine the number of shared bits between S_j and $\text{MajEnc}_m(X_j) = Y_{(j-1)m+1} \parallel Y_{(j-1)m+2} \parallel \dots \parallel Y_{jm}$. For convenience, we define $B_j := Y_{(j-1)m+1} \parallel Y_{(j-1)m+2} \parallel \dots \parallel Y_{jm}$. We distinguish bit values from position labels as follows. Every original bit Y_i has label i , and the r -th bit inserted before Y_i has label (i, r) , so all labels lie in the common universe $\mathcal{U} := [mn] \cup ([mn] \times \mathbb{N})$. Let $B'_j := \{(j-1)m+1, \dots, jm\}$, and let H'_j and S'_j be the sets of labels of the output symbols appearing in H_j and S_j , respectively. A deleted original label belongs to B'_j but to neither H'_j nor S'_j , while a substituted bit retains its original label. Since all emitted symbols have unique labels, $|S'_j \Delta H'_j| = |S_j \Delta H_j|$.

Since $\mathcal{E}(B_j) = H_j$, we first determine the number of surviving B'_j bit positions in H'_j .

Lemma 5 (Codeword Bit Positions Surviving the Channel). *Let $a := \frac{1-P_i-P_d}{1-P_i}$. For all j ,*

$$\Pr[|B'_j \cap H'_j| = am + o(m)] \geq 1 - \text{negl}(n).$$

Proof. Since every original label in B'_j independently has probability a of surviving the channel and becoming part of H'_j , $|B'_j \cap H'_j| \sim \text{Bin}(m, a)$. Chernoff's inequality plus a union bound imply $|B'_j \cap H'_j| = am + o(m)$ simultaneously for all j with overwhelming probability. $\square$

With the lemma above we can determine the number of shared bit positions between B'_j and S'_j . Here we inherit m 's value from Christ and Gunn [3], so that the uncertain tail can be asymptotically smaller than the certain part.

Lemma 6 (Codeword Bit Positions in Reconstructed Blocks). *Let $m = \Omega(n \log^6 n)$. For all j ,*

$$\Pr[|B'_j \cap S'_j| = am + o(m)] \geq 1 - \text{negl}(n).$$

Proof. $B'_j \cap S'_j \subseteq B'_j \cap H'_j$, and $(B'_j \cap H'_j) \setminus (B'_j \cap S'_j) \subseteq S'_j \Delta H'_j$, so

$$|B'_j \cap H'_j| - |S'_j \Delta H'_j| \leq |B'_j \cap S'_j| \leq |B'_j \cap H'_j|.$$

Therefore, with overwhelming probability,

$$am + o(m) - O(u) \leq |B'_j \cap S'_j| \leq am + o(m).$$

Taking $m = \Omega(n \log^6 n)$ makes u asymptotically smaller than m , since $\frac{u}{m} = c\sqrt{\frac{n}{m}} \log n = O\left(\frac{1}{\log^2 n}\right) = o(1)$. Thus $u = o(m)$, and both sides of the bound become $am + o(m)$, proving $|B'_j \cap S'_j| = am + o(m)$ with overwhelming probability. $\square$

3.2 Majority-Vote Agreement

We now use the preceding overlap bound to compute the correlation between the majority of S_j and the majority of B_j (representing X_j). Let $\mathcal{F}$ denote the channel geometry generated by all insertion counts and all deletion/survival decisions. Conditioning on $\mathcal{F}$ fixes every labeled set B'_j, H'_j, S'_j and every decoder boundary, but leaves the original bit values, inserted bit values, and substitution indicators mutually independent. Fix a good geometry $\mathcal{F} = \omega$ for which the preceding simultaneous bounds hold, and define

$$T_j := B'_j \cap S'_j, \quad N_j := |T_j| = am + o(m), \quad |S'_j| = \mu m + o(m).$$

For every $r \in B'_j$, let $W_r \in \{-1, 1\}$ be the sign representation of the original bit at position r . For every $r \in T_j$, independently let

$$F_r = \begin{cases} +1, & \text{with probability } 1 - P_s, \\ -1, & \text{with probability } P_s, \end{cases}$$

so the corresponding received bit has sign $F_r W_r$. For every output label $q \in S'_j \setminus T_j$, let $Q_q \in \{-1, 1\}$ denote its received sign. Conditional on $\mathcal{F} = \omega$, the variables W_r, F_r , and Q_q used below are independent, the W_r and Q_q are uniform Rademacher variables, and $\mathbb{E}[F_r] = 1 - 2P_s$. Define

$$\begin{aligned} W^{(j)} &:= \sum_{r \in T_j} W_r, & \widetilde{W}^{(j)} &:= \sum_{r \in T_j} F_r W_r, \\ V^{(j)} &:= \sum_{r \in B'_j \setminus T_j} W_r, & Z^{(j)} &:= \sum_{q \in S'_j \setminus T_j} Q_q. \end{aligned}$$

Then

$$B_j^{\text{Maj}} = W^{(j)} + V^{(j)}, \quad S_j^{\text{Maj}} = \widetilde{W}^{(j)} + Z^{(j)}.$$

Therefore, conditional on $\mathcal{F} = \omega$,

$$\text{Var}(B_j^{\text{Maj}}) = m, \quad \text{Var}(S_j^{\text{Maj}}) = |S'_j| = \mu m + o(m),$$

and only the paired contributions $(W_r, F_r W_r)$ for $r \in T_j$ contribute covariance. Hence

$$\begin{aligned} \text{Cov}(B_j^{\text{Maj}}, S_j^{\text{Maj}} \mid \mathcal{F} = \omega) &= \sum_{r \in T_j} \mathbb{E}[F_r W_r^2] \\ &= (1 - 2P_s)N_j \\ &= (1 - 2P_s)am + o(m). \end{aligned}$$

If we define $a' := (1 - 2P_s)a$, the conditional correlation coefficient is

$$\begin{aligned} \rho_{j,\omega} &= \frac{a'm + o(m)}{\sqrt{m(\mu m + o(m))}} \\ &= \frac{a'}{\sqrt{\mu}} + o(1) \\ &= \frac{(1 - P_i - P_d)(1 - 2P_s)}{\sqrt{(1 - P_i)(1 - P_d)}} + o(1). \end{aligned}$$

Finally, with conditional correlation coefficient $\rho_{j,\omega}$ derived as above uniformly in all j and good geometry ω , we can compute the probability that $\text{MajDec}_1(S_j)$ equals X_j .

Lemma 7 (Majority-Vote Matching Probability). *For $1 - P_i - P_d > 0$, $P_s < \frac{1}{2}$, sufficiently large odd m , every good geometry $\mathcal{F} = \omega$, and all j , there exists a constant $\gamma > 0$ independent of j and ω such that*

$$\Pr[\text{MajDec}_1(S_j) = X_j \mid \mathcal{F} = \omega] \geq \frac{1}{2} + \gamma.$$

Proof. Since m is odd, X_j is the sign of the majority statistic of B_j . Thus, conditional on $\mathcal{F} = \omega$,

$$\Pr[\text{MajDec}_1(S_j) = X_j \mid \mathcal{F} = \omega] = \Pr[\text{sgn}(S_j^{\text{Maj}}) = \text{sgn}(B_j^{\text{Maj}}) \mid \mathcal{F} = \omega].$$

To find the probability that S_j^{Maj} and B_j^{Maj} share the same sign, we apply the central limit theorem to R_j , defined as follows:

$$\begin{aligned} R_{j,\omega} &:= \left(\frac{S_j^{\text{Maj}}}{\sqrt{|S'_j|}}, \frac{B_j^{\text{Maj}}}{\sqrt{m}} \right) \\ &= \sum_{r \in T_j} \left(\frac{F_r W_r}{\sqrt{|S'_j|}}, \frac{W_r}{\sqrt{m}} \right) + \sum_{q \in S'_j \setminus T_j} \left(\frac{Q_q}{\sqrt{|S'_j|}}, 0 \right) + \sum_{r \in B'_j \setminus T_j} \left(0, \frac{W_r}{\sqrt{m}} \right). \end{aligned}$$

These are independent, centered, bounded vector summands conditional on $\mathcal{F} = \omega$. Since $|S'_j| = \Theta(m)$, their largest Euclidean norm is $O(m^{-1/2})$, meaning all summands uniformly approaches 0, so the Lindeberg condition holds. Define

$\rho := \frac{a'}{\sqrt{\mu}}$. The covariance matrices $\begin{pmatrix} 1 & \rho_{j,\omega} \\ \rho_{j,\omega} & 1 \end{pmatrix}$ converge uniformly over good

geometries to $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ because $\rho_{j,\omega}$ uniformly converges to ρ . Applying the multivariate Lindeberg–Feller central limit theorem, for good ω ,

$$R_{j,\omega} \xrightarrow{d} R \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Since the convergence holds for every sequence of pairs of j and good ω , the convergence is uniform in j and ω . For such an $R = (R_1, R_2)$, represent its coordinates using two independent standard normal variables X, Y as $R_1 = X$ and $R_2 = \rho X + \sqrt{1 - \rho^2} Y$. The event $R_1 > 0, R_2 > 0$ is equivalent to $X > 0$ and $Y > -\frac{\rho}{\sqrt{1 - \rho^2}} X$. These inequalities define a sector in the (X, Y) -plane with angular size $\frac{\pi}{2} + \arcsin(\rho)$. By rotational symmetry, its probability is $\frac{1}{4} + \frac{\arcsin \rho}{2\pi}$. Symmetry also gives $\Pr[R_1 < 0, R_2 < 0] = \Pr[R_1 > 0, R_2 > 0]$, so

$$\Pr[\text{sgn}(R_1) = \text{sgn}(R_2)] = \frac{1}{2} + \frac{\arcsin \rho}{\pi}.$$

Since this convergence is uniform over good geometries and the limiting normal distribution assigns zero probability to either coordinate being zero, majority ties contribute only $o(1)$. Therefore, for every good ω ,

$$\begin{aligned} \Pr \left[\operatorname{sgn} \left(\frac{S_j^{\text{Maj}}}{\sqrt{|S'_j|}} \right) = \operatorname{sgn} \left(\frac{B_j^{\text{Maj}}}{\sqrt{m}} \right) \mid \mathcal{F} = \omega \right] \\ = \Pr[\operatorname{sgn}(S_j^{\text{Maj}}) = \operatorname{sgn}(B_j^{\text{Maj}}) \mid \mathcal{F} = \omega] \\ = \frac{1}{2} + \frac{\arcsin \rho}{\pi} + o(1). \end{aligned}$$

Let $0 < \gamma < \frac{\arcsin \rho}{\pi}$. Uniform convergence then implies that for sufficiently large odd m , every good ω , and every j ,

$$\Pr[\text{MajDec}_1(S_j) = X_j \mid \mathcal{F} = \omega] \geq \frac{1}{2} + \gamma.$$

□

3.3 Global Concentration and the CG24 Corollary

We now prove Theorem 1 and derive its consequence for the CG24 construction.

Proof (Proof of Theorem 1). Let **Good** be the event, measurable with respect to $\mathcal{F}$, on which all the preceding simultaneous output-length, boundary-misalignment, and surviving-position bounds hold. These lemmas give $\Pr[\text{Good}] \geq 1 - \text{negl}(n)$. Fix any geometry $\mathcal{F} = \omega \in \text{Good}$. On this geometry,

$$|E_j - C_{jm}| = o(m), \quad |H'_j| = C_{jm} - C_{(j-1)m} = \mu m + o(m) = \Theta(m)$$

simultaneously for all j . Hence, for sufficiently large m , neither endpoint of S_j can cross an entire adjacent H -block, and therefore $S'_j \subseteq H'_{j-1} \cup H'_j \cup H'_{j+1}$, with the nonexistent boundary blocks omitted. Thus the event $I_j := \mathbf{1}[\text{MajDec}_1(S_j) = X_j]$ depends only on the original bits, inserted bits, and substitution indicators associated with input blocks $j-1, j, j+1$. Conditional on the fixed geometry ω , these underlying random variables are independent across disjoint input-block neighborhoods. In particular, the events I_j and $I_{j'}$ are independent whenever $|j - j'| \geq 3$.

We therefore separate the events into three groups $A^{(1)}, A^{(2)}, A^{(3)}$ according to $j \equiv 0, 1, 2 \pmod{3}$. Within each group the events are mutually independent conditional on $\mathcal{F} = \omega$, and the preceding lemma gives $\Pr[I_j = 1 \mid \mathcal{F} = \omega] \geq \frac{1}{2} + \gamma$. Thus, for $l \in \{1, 2, 3\}$, a Chernoff bound gives

$$\Pr \left[\sum_{j \in A^{(l)}} \mathbf{1}[\text{MajDec}_1(S_j) = X_j] < |A^{(l)}| \left(\frac{1}{2} + \frac{\gamma}{2} \right) \mid \mathcal{F} = \omega \right] \leq \exp(-\Omega(\gamma^2 n)).$$

Taking a union bound over the three groups, for every fixed good geometry ω we have

$$\Pr \left[\sum_{j=1}^n \mathbf{1}[\text{MajDec}_1(S_j) = X_j] < n \left(\frac{1}{2} + \frac{\gamma}{2} \right) \mid \mathcal{F} = \omega \right] \leq 3 \exp(-\Omega(\gamma^2 n)) = \text{negl}(n).$$

Since this conditional bound is uniform over $\omega \in \text{Good}$,

$$\begin{aligned} \Pr[\text{failure}] &\leq \Pr[\neg \text{Good}] + \sup_{\omega \in \text{Good}} \Pr[\text{failure} \mid \mathcal{F} = \omega] \\ &= \text{negl}(n). \end{aligned}$$

Let $\varepsilon = \gamma/2$. We obtain the main theorem when $1 - P_i - P_d > 0$, $P_s < \frac{1}{2}$, and odd $m = \Omega(n \log^6 n)$:

$$\Pr[\text{dist}(x, \text{MajDec}_n(\mathcal{E}(\text{MajEnc}_m(x)))) \leq (\tfrac{1}{2} - \varepsilon)n] \geq 1 - \text{negl}(n).$$

□

The Majority-Voted Code Survives BSID Theorem, which parallels Lemma 15 of Christ and Gunn [3], immediately implies that their deletion-channel PRC works for the BSID channel under the parameter constraints above.

Corollary 1 (CG24 Code for BSID). *Let $\mathcal{E}$ be a BSID channel with parameters $\{P_i, P_d, P_s\}$, and security parameter be λ . Let a binary multi-bit pseudorandom code, such as the code of Theorem 3 of Christ and Gunn [3], be robust to every key-independent p -bounded Hamming channel with block length $n \geq \lambda$, and accept all assumptions this PRC is established on. If $\frac{1}{2} - \varepsilon < p < \frac{1}{2}$, $1 - P_i - P_d > 0$, $P_s < \frac{1}{2}$, and $m = \Omega(n \log^6 n)$ is sufficiently large and odd, then Construction 6 of Christ and Gunn [3] yields a PRC that is robust to $\mathcal{E}$, pseudorandom to PPT adversary with $1 - \text{negl}(\lambda)$ probability, and is sound with $1 - \text{negl}(\lambda)$ probability.*

Proof. By Theorem 1, $\text{dist}(x, \text{MajDec}_n(\mathcal{E}(\text{MajEnc}_m(x))))$ is $\frac{1}{2} - \varepsilon$ bounded for uniform x . The same bound holds when x is a pseudorandom codeword: otherwise, sampling the channel, applying MajEnc_m and MajDec_n , and testing the resulting distance would distinguish a PRC codeword from a uniform string. Choosing the underlying PRC to tolerate any $p > \frac{1}{2} - \varepsilon$ bounded channel, robustness follows exactly as in the proof of Theorem 4 of Christ and Gunn [3]. The soundness and pseudorandomness arguments of their Construction 6 are unchanged. □

4 A Unique-Decodable Pseudorandom Code for the q -ary SID Channel

We construct a unique-decodable pseudorandom code under the q -ary SID channel $\mathcal{E}'$ by combining the Corollary 1 PRC and the error-correcting code against synchronization errors from Theorem 2.8 of Haeupler and Shahrasbi [8]. Its rate is at least $(1 - \delta - \varepsilon) \frac{\log(q/2)}{\log q}$.

Theorem 2. *Let PRC be the binary public-key pseudorandom code from Corollary 1, accept assumptions needed by this PRC and standard PRG. For security parameter λ , any constants P_i, P_d, P_s and $\varepsilon, \eta > 0$, and any admissible sufficiently large constant alphabet size q satisfying $q \geq \left(\frac{1+\gamma}{\varepsilon^2}\right)^{O(\frac{1+\gamma}{\varepsilon^3})}$, $q > \frac{1}{(1-P_s)}$, $2 \mid q$, $\gamma = \frac{P_i + (1-P_i-P_d)P_s}{1-P_i} + \eta$, with $8(1+\gamma)/\varepsilon \in \mathbb{Z}$, if $P_i + P_d < 1$ and $P_s < 1$ then there exists a polynomial-time public-key pseudorandom code of rate $R \geq (1 - \delta - \varepsilon) \frac{\log(q/2)}{\log q}$, where $\delta = \frac{P_d + (1-P_i-P_d)P_s}{1-P_i} + \eta$, over an alphabet of size q that is unique-decodable over the q -ary SID channel with parameters (P_i, P_d, P_s) except with negligible probability in λ , and its codewords are computationally indistinguishable in λ from uniform strings over the code alphabet, and sound with overwhelming probability.*

In this construction, the code of Christ and Gunn [3] carries the control information including the seeds for a random mask and a self-matching string, a hash key and hash value of the message, through $\mathcal{E}'$ and is decoded first. If the control information is recovered correctly and the generated string is a valid self-matching string, the remaining decoding procedure is the original list decoder from Theorem 2.8 of Haeupler and Shahrabi [8] operating on a q -ary SID channel with a hash value verification to pinpoint the transmitted message.

4.1 Parameters

Per input symbol through $\mathcal{E}'$, the expected insertion, deletion, and substitution ratios are respectively $\frac{P_i}{1-P_i}$, $\frac{P_d}{1-P_i}$, and $\frac{(1-P_i-P_d)P_s}{1-P_i}$. Over a length- N codeword, these ratios have $o(1)$ fluctuations with overwhelming probability as proved in the following lemma:

Lemma 8 (Concentration of SID error ratios). *Let I_N, D_N, S_N be the numbers of insertions, deletions, and substitutions produced by feeding N symbols through $\mathcal{E}'$, then with $1 - \text{negl}(N)$ probability,*

$$\frac{I_N}{N} = \frac{P_i}{1-P_i} + o(1), \quad \frac{D_N}{N} = \frac{P_d}{1-P_i} + o(1), \quad \frac{S_N}{N} = \frac{(1-P_i-P_d)P_s}{1-P_i} + o(1).$$

Proof. For each input coordinate, let K_i denote the number of preceding insertions. As observed in the proof of Lemma 2, the K_i 's are independent geometric random variables with $\mathbb{E}K_i = \frac{P_i}{1-P_i}$ and constant subexponential norm. The deletion and substitution indicators are independent across input coordinates, bounded, and have respective means $\frac{P_d}{1-P_i}$ and $\frac{(1-P_i-P_d)P_s}{1-P_i}$. Applying Bernstein's inequality as in Lemma 3 to each of the three length- N sums gives deviations $O(\sqrt{N} \log N) = o(N)$ from their expectations, except with probability $e^{-\Omega(\log^2 N)} = \text{negl}(N)$. A union bound over the three counts and division by N demonstrate fluctuation of $o(1)$ and prove then claim. $\square$

Since Theorem 2.8 of Haeupler and Shahrabi [8] is formulated in edit distance, we interpret each substitution as one deletion followed by one insertion.

Fix a small constant concentration slack $\tau > 0$; with overwhelming probability, the full codeword therefore has insertion and deletion ratios at most $\gamma_0 = \frac{P_i + (1 - P_i - P_d)P_s}{1 - P_i} + \tau$ and $\delta_0 = \frac{P_d + (1 - P_i - P_d)P_s}{1 - P_i} + \tau$, respectively.

Every symbol of the construction's codeword is a bundling of ((HRW Code + mask) $\times$ self-matching string $\times$ binary PRC), in which (HRW Code $\times$ self-matching string) belongs to the self-matching-string list-recovery construction in Theorem 2.8 of Haeupler and Shahrabi [8]. For each substitution of a codeword symbol, assuming that the mask seed is recovered correctly from the binary PRC, the probability that the (HRW Code $\times$ self-matching string) pair is altered is $r = (q - 2)/(q - 1)$. Thus, with overwhelming probability, the insertion and deletion ratios seen by the HS construction are at most

$$\gamma_{HS} = \frac{P_i + (1 - P_i - P_d)rP_s}{1 - P_i} + \tau, \quad \delta_{HS} = \frac{P_d + (1 - P_i - P_d)rP_s}{1 - P_i} + \tau.$$

For parameter setup, choose another small constant $\eta \geq \tau$ and define the conservative bounds

$$\gamma = \frac{P_i + (1 - P_i - P_d)P_s}{1 - P_i} + \eta \geq \gamma_{HS}, \quad \delta = \frac{P_d + (1 - P_i - P_d)P_s}{1 - P_i} + \eta \geq \delta_{HS}.$$

Finally, choose $\varepsilon > 0$ such that $\eta + \varepsilon < \frac{(1 - P_i - P_d)(1 - P_s)}{1 - P_i}$. With γ , δ , and ε fixed, bounded by $8(1 + \gamma)/\varepsilon \in \mathbb{Z}$, All parameters involved in the self-matching-string list-recovery construction of Theorem 2.8 in Haeupler and Shahrabi [8] can be determined from the original text, except we re-instantiate the HRW list-recoverable code with a larger maximum input size of $l = \frac{8(1 + \gamma)}{\varepsilon}$, rather than using the original $l = \frac{2(1 + \gamma)}{\varepsilon}$ appearing in the proof of HS21 Theorem 2.8 [8]. This is a direct re-instantiation of HRW Theorem A.7 [9] (see also HS21 Theorem 2.9 [8]), whose rate $\rho = 1 - \delta - \varepsilon/2$ and agreement threshold $a = 1 - \delta - \varepsilon/4$ are independent of ℓ . Since $\ell = O_{\gamma, \varepsilon}(1)$, this constant-factor increase preserves the asymptotic alphabet-size and output-list bounds used in HS21; in particular, $|\Sigma_C| \geq \left(\frac{1 + \gamma}{\varepsilon^2}\right)^{O((1 + \gamma)/\varepsilon^3)}$ and $|\mathcal{L}| = L(N) = \exp(\exp(\exp(\log^* N)))$. In the repositioning algorithm ListRep, the LCS repositioning executes $K = \frac{8(1 + \gamma)}{\varepsilon}$ times; and the reference string S is proved later by Lemma 10 to be $\varepsilon_s = \frac{\varepsilon^2}{64(1 + \gamma)}$ self-matching with overwhelming probability when $|\Sigma_S| = \Omega\left(\frac{(1 + \gamma)^2}{\varepsilon^4}\right)$. Applying equation (1) with $u_i = x_i + R_i$ and the chosen parameters shows that the candidate sets contain the correctly positioned masked symbol in at least

$$\left(1 - \delta - \frac{1 + \gamma}{K} - K\varepsilon_s\right) N = \left(1 - \delta - \frac{\varepsilon}{4}\right) N = aN$$

positions, as required by the HS21 list-recovery construction.

At last, block length N requires special attention because both binary PRC of Corollary 1 and the self-matching-string code have constraints over it, the former requires $N = mn$ where m is odd and $m = \Omega(n \log^6(n))$, the latter needs N to be subjugated to the internal HRW code block length, where $N =$

$|\Sigma_C|^{e/2}(\sqrt{|\Sigma_C|} - 1)$ for $e \in \mathbb{Z}$. The scheme must choose admissible N to meet these conditions. We hereby provide a way to find some of the valid N s and prove infinitely many admissible N s exist in a short lemma.

Lemma 9 (Compatible block lengths). *There are infinitely many admissible HRW block lengths N for which $N = nm$, where m is odd and $m = \Omega(n \log^6(n))$.*

Proof. Let $|\Sigma_C| \geq \left(\frac{1+\gamma}{\varepsilon^2}\right)^{O((1+\gamma)/\varepsilon^3)}$ be r^2 where r is some odd integer. Write an admissible HRW block length as $N = r^e(r-1)$, and let 2^s be the largest power of 2 dividing $r-1$, let c be any constant such that $0 < c < 1/2$. Set $n = 2^s r^{\lfloor ce \rfloor}$, $m = \frac{r-1}{2^s} r^{e-\lfloor ce \rfloor}$, so that $N = nm$, and m is odd. Moreover,

$$\frac{m}{n} = \frac{r-1}{2^{2s}} r^{e-2\lfloor ce \rfloor} = r^{\Omega(e)},$$

whereas $\log n = \Theta(e)$. Hence, for all sufficiently large e , $m = \Omega(n \log^6 n)$. $\square$

4.2 Construction

Building Blocks

- PRC of Christ and Gunn [3] from Corollary 1: $\{\text{PRC.Init}, \text{PRC.Enc}, \text{PRC.Dec}\}$, instantiated with message capacity $n_{\text{PRC}} \geq 4\kappa \geq 4\lambda$ to encode a 4κ long control information series into a length- N binary codeword. Here n_{PRC} is the block length of the internal PRC directly encoding the 4κ length message, $N = n_{\text{PRC}} m_{\text{PRC}}$, $m_{\text{PRC}} = \Omega(n_{\text{PRC}} \log^6 n_{\text{PRC}})$ is sufficiently large and odd. For an efficient encoding and decoding, $N = \text{poly}(\lambda)$.
- The HRW list recoverable code used in Theorem 2.8 of Haeupler and Shahrasbi [8]: $\mathcal{C}_{\text{HRW}} \subseteq \Sigma_C^N : \{\text{Enc}_{\text{HRW}}, \text{ListRec}_{\text{HRW}}\}$, in rate ρ , block length N , minimum list recovery agreement ratio a , maximum input list size l and maximum output list size $L(N)$.
- PRG $G_0() : \{0, 1\}^\kappa \rightarrow \Sigma_C^N$.
- PRG $G_1() : \{0, 1\}^\kappa \rightarrow \Sigma_S^N$.
- ε_H -almost-universal hash function $H() : \Sigma_C^{\rho N} \rightarrow \{0, 1\}^\kappa$. Since we would like to hash the message $m \in \Sigma_C^{\rho N}$, and m if represented in binary will take $\log_2(\Sigma_C^{\rho N}) = \lceil \rho N \log_2 |\Sigma_C| \rceil$, according to Preliminaries 2.3, we initiate $H()$ with parameter key length κ and $d = \left\lceil \frac{\lceil \rho N \log_2 |\Sigma_C| \rceil}{\kappa} \right\rceil = \text{poly}(\kappa)$, hence $\varepsilon_H = \text{negl}(\kappa) = \text{negl}(\lambda)$.
- The payload alphabet Σ_C is a finite additive group, so masking and unmasking are well-defined.
- The HS21 repositioning procedure ListRep_K returns candidate sets $A_i \subseteq \Sigma_C$ with $|A_i| \leq K$.
- The alphabet size of the codeword is $q = 2|\Sigma_C||\Sigma_S| = \left(\frac{1+\gamma}{\varepsilon^2}\right)^{O(\frac{1+\gamma}{\varepsilon^3})}$.

Initialization $\text{Init}(\lambda)$

- $(pk, sk) \leftarrow \text{PRC.Init}(\lambda)$.

Encoding $\text{Enc}_{pk}(m)$

- Sample two independent uniform seeds for random-mask and self-matching-string generation: $\sigma_0, \sigma_1, \tau \xleftarrow{\$} \{0, 1\}^\kappa$, where $\kappa \geq \lambda$.
- Compute hash value of the message m : $h = H_\tau(m) \in \{0, 1\}^\kappa$.
- Encode the message with the HRW code: $x = \text{Enc}_{\text{HRW}}(m) = (x_1, \dots, x_N) \in \Sigma_C^N$.
- Generate the random mask and self-matching string:
 $R = G_0(\sigma_0) = (R_1, \dots, R_N) \in \Sigma_C^N, S = G_1(\sigma_1) = (S_1, \dots, S_N) \in \Sigma_S^N$.
- Encode the seeds with the PRC: $b = \text{PRC.Enc}_{pk}(\sigma_0 \parallel \sigma_1 \parallel \tau \parallel h) = (b_1, \dots, b_N) \in \{0, 1\}^N$.
- Output the codeword $c = (c_1, \dots, c_N)$ defined coordinate-wise by $c_i = (x_i + R_i, S_i, b_i) \in \Sigma_C \times \Sigma_S \times \{0, 1\}$.

Decoding $\text{Dec}_{sk}(c')$

- Let the corrupted codeword be $c' = ((y'_1, z'_1, b'_1), \dots, (y'_{N'}, z'_{N'}, b'_{N'}))$.
- Decode the seeds from the PRC codeword: $(\hat{\sigma}_0, \hat{\sigma}_1, \hat{\tau}, \hat{h}) \leftarrow \text{PRC.Dec}_{sk}(b'_1 \dots b'_{N'})$.
If the result is $\perp$, output $\perp$ and abort.
- Regenerate the random mask and self-matching string: $\hat{R} = G_0(\hat{\sigma}_0), \hat{S} = G_1(\hat{\sigma}_1)$.
- Run the modified repositioning algorithm from Theorem 2.8 of Haeupler and Shahrabi [8]:

$$(A_1, \dots, A_N) = \text{ListRep}_K \left(\hat{S}, ((y'_j, z'_j))_{j \in [N']} \right).$$

- Unmask each positional candidate: $B_i := A_i - \hat{R}_i = \{a - \hat{R}_i : a \in A_i\}$.
- Compute the HRW list-recovery result: $\mathcal{L} \leftarrow \text{ListRec}_{\text{HRW}}(B_1, \dots, B_N)$.
- Finally, for every entry in $\mathcal{L}$, output the first $\hat{m}_k$ such that $H_{\hat{\tau}}(\hat{m}_k) = \hat{h}$; if no matching $\hat{m}_k$ is detected, output $\perp$.

4.3 Analysis

In this section we discuss the robustness on q -ary SID channel $\mathcal{E}'$, pseudorandomness to PPT adversary, soundness, and rate of the construction above.

Robustness The q -ary SID channel $\mathcal{E}'$ satisfies $1 - P_i - P_d > 0$ and $P_s < 1$. Projecting each received q -ary symbol onto its binary coordinate induces a BSID channel with parameters $\{P_i, P_d, \frac{(q/2)P_s}{q-1}\}$, the substitution ration being $\frac{(q/2)P_s}{q-1}$ is because exactly $q/2$ of the $q - 1$ possible substitution targets have

binary coordinate opposite to that of the transmitted symbol. Since $P_s < 1$ and $q > \frac{1}{1-P_s}$ implies $\frac{(q/2)P_s}{q-1} < \frac{1}{2}$, Corollary 1 proves that the embedded PRC is robust, and hence $\Pr[(\hat{\sigma}_0, \hat{\sigma}_1, \hat{\tau}, \hat{h}) = (\sigma_0, \sigma_1, \tau, h)] \geq 1 - \text{negl}(\lambda)$. With the correct seeds recovered, Theorem 2.8 of Haeupler and Shahrasi [8] applies because the actual projected edit ratios satisfy $\gamma_{HS} \leq \gamma$ and $\delta_{HS} \leq \delta$, the parameters obey $1 - \delta - \varepsilon > 0$, and $S = G_1(\sigma_1)$ is ε_s self-matching with overwhelming probability. Indeed, by the choice of η and ε above, $1 - \delta - \varepsilon = \frac{(1-P_i-P_d)(1-P_s)}{1-P_i} - \eta - \varepsilon > 0$.

Next, we examine whether the pseudorandom string S is ε_s self-matching. By Chapter 2.1.1 of HS21, a uniform random string of length N has probability at most $\left(\frac{e^2}{|\Sigma|\varepsilon^2}\right)^{N\varepsilon}$ of failing to be ε self-matching. Thus, for $|\Sigma| = \Omega(\varepsilon^{-2})$ with a sufficiently large hidden constant, the failure probability is $\text{negl}(N)$. With this fact we can prove the following lemma.

Lemma 10 (S Is ε_s Self-Matching). *When $|\Sigma_S| = \Omega\left(\frac{(1+\gamma)^2}{\varepsilon^4}\right)$, S is ε_s self-matching with probability at least $1 - \text{negl}(\lambda)$.*

Proof. For a length- N uniform random string U^N over Σ_S , the conclusion above implies that it is ε_s self-matching with probability at least $1 - \text{negl}(N)$, because $\varepsilon_s^{-2} = \frac{4096(1+\gamma)^2}{\varepsilon^4}$ and $|\Sigma_S| = \Omega(\varepsilon_s^{-2})$.

For $S = G_1(\sigma_1)$, suppose instead that the probability of failing the ε_s self-matching property exceeded the uniform-string failure probability by a non-negligible function of λ . Since $N = \text{poly}(\lambda)$ and the self-matching property can be checked in time $\text{poly}(N)$ [8], this checker would distinguish $G_1(U^\kappa)$ from U^N with non-negligible advantage, contradicting the security of G_1 . Therefore S is ε_s self-matching with probability at least $1 - \text{negl}(\lambda)$. $\square$

For $\mathcal{L}$, where $m \in \mathcal{L}$, and correctly recovered τ, h : because for all $\hat{m}_k \in \mathcal{L} \wedge \hat{m}_k \neq m$, $\Pr_\tau[H_\tau(\hat{m}_k) = h] \leq \varepsilon_H = \text{negl}(\lambda)$, with a union bound over all other $|\mathcal{L}| - 1$ list entries, $\Pr_\tau[\exists \hat{m}_k \neq m, \tau(\hat{m}_k) = h] = (|\mathcal{L}| - 1)\text{negl}(\lambda) = \text{negl}(\lambda)$, hence the probability for the correct m to be selected from $\mathcal{L}$ as output is $1 - \text{negl}()$.

In summary, over $\mathcal{E}'$ the PRC recovers $(\sigma_0, \sigma_1, \tau, h)$ with overwhelming probability; given the correct σ_1 , S is ε_s self-matching with overwhelming probability; and given the correct σ_0 and S , ListRep and ListRec return a list $\mathcal{L}$ containing m with overwhelming probability, and hash detects correct m from $\mathcal{L}$ with overwhelming probability. By a union bound,

$$\Pr[m = \text{Dec}_{sk}(\mathcal{E}'(\text{Enc}_{pk}(m)))] \geq 1 - \text{negl}(\lambda).$$

Pseudorandomness Fix a PPT adversary A given pk and making at most $Q = \text{poly}(\lambda)$ adaptive message queries. Every game samples $(pk, sk) \leftarrow \text{PRC.Init}(\lambda)$ once. The real oracle returns $\text{Enc}_{pk}(m)$, whereas the ideal oracle returns a fresh independent uniform string in Σ^N , where $\Sigma = \Sigma_C \times \Sigma_S \times \{0, 1\}$ and $|\Sigma| = q$.

Game H_0 . Answer every query m with $\text{Enc}_{pk}(m)$, with fresh independent seeds σ_0, σ_1 , fresh hash key τ , and fresh encoding randomness. Write $x = \text{Enc}_{\text{HRW}}(m)$.

Game H_1 . Replace every binary PRC component by $U_B \xleftarrow{\$} \{0,1\}^N$. A distinguisher for the underlying PRC oracle simulates A by passing it pk , sampling the two seeds and hash key on each query, submitting $\sigma_0 \parallel \sigma_1 \parallel \tau \parallel H_\tau(m)$ to its oracle, and computing the other two components locally. Thus the difference in A 's acceptance probabilities between H_0 and H_1 is at most $\varepsilon_{\text{PRC}}(\lambda) = \text{negl}(\lambda)$ by the Corollary 1 binary PRC's pseudorandomness.

Game H_2 . Starting from H_1 , replace every mask $G_0(\sigma_0)$ by $U_C \xleftarrow{\$} \Sigma_C^N$. Use intermediate games replacing the first t mask outputs, for $t = 0, \dots, Q$. A G_0 distinguisher embeds its challenge at a uniformly chosen query and simulates all other replies, continuing to answer adaptive queries. This is possible because σ_0 is no longer used in any other component. If $\varepsilon_0(\lambda)$ is this distinguisher's negligible advantage, the transition costs at most $Q\varepsilon_0(\lambda)$.

Game H_3 . Starting from H_2 , replace every index string $G_1(\sigma_1)$ by $U_S \xleftarrow{\$} \Sigma_S^N$. The same query-by-query argument applies to G_1 , since σ_1 is no longer used in any other component. If $\varepsilon_1(\lambda)$ is the corresponding distinguisher's negligible advantage, this transition costs at most $Q\varepsilon_1(\lambda)$.

In H_3 , conditional on pk , the prior transcript, and the current query, the reply is $(x + U_C, U_S, U_B)$. Translation by x is a bijection of Σ_C^N , so this reply is uniform over Σ^N , independently of all previous replies. Hence H_3 is exactly the ideal oracle, whose Q replies are jointly independent uniform q -ary strings. The total distinguishing advantage is at most $\varepsilon_{\text{PRC}}(\lambda) + Q(\varepsilon_0(\lambda) + \varepsilon_1(\lambda)) = \text{negl}(\lambda)$, proving the required oracle pseudorandomness.

Soundness The soundness of this construction immediately follows from the soundness of the Corollary-1 PRC. Let a q -ary input string c' satisfy the conditions of PRC soundness definition, its binary projection will satisfy the same conditions, so $\text{PRC.Dec}_{sk}(b'_1 \dots b'_{N'})$ returns $\perp$ with overwhelming probability. By the decoder definition, the construction would also outputs $\perp$ and aborts. Hence, $\text{Dec}_{sk}(c')$ returns a non $\perp$ value on a (pk, sk) independent input already with at most $\text{negl}(\lambda)$ probability.

Rate According to Theorem 2.8 of Haeupler and Shahrasbi [8], without the binary PRC coordinate, the complete HS code over $\Sigma_C \times \Sigma_S$ has rate

$$R_{HS} \geq 1 - \delta - \varepsilon = \frac{(1 - P_i - P_d)(1 - P_s)}{1 - P_i} - \eta - \varepsilon.$$

The binary PRC coordinate contains only control information while doubles the alphabet size. Since $q = 2|\Sigma_C||\Sigma_S|$, the rate of the complete construction is therefore $R = R_{HS} \frac{\log(q/2)}{\log q}$.

Proof (Proof of Theorem 2). Polynomial running time follows from the corresponding guarantees of the underlying components. Robustness, pseudorandomness, and soundness are established in preceding discussions, and the rate

calculation above gives

$$R \geq (1 - \delta - \varepsilon) \frac{\log(q/2)}{\log q}.$$

Together these prove the theorem. $\square$

Finally, we compare the achieved rate R with the genie aided channel capacity.

Proposition 1 (Genie-aided capacity upper bound). *Let $C_{\text{SID}}^{(q)}$ denote the genie aided capacity of the q -ary SID channel. Then*

$$C_{\text{SID}}^{(q)} \leq \frac{1 - P_i - P_d}{1 - P_i} (1 - h_q(P_s)),$$

where

$$h_q(p) = p \log_q(q - 1) - p \log_q p - (1 - p) \log_q(1 - p).$$

Moreover, the rate R of Theorem 2 approaches this upper bound, up to the arbitrarily small parameter slack $\eta + \varepsilon$, as $q \rightarrow \infty$ over admissible alphabet sizes.

Proof. A genie revealing all insertion and deletion locations leaves a fraction

$$\frac{1 - P_i - P_d}{1 - P_i}$$

of the input symbols passing through a q -ary symmetric channel with substitution probability P_s , giving the stated upper bound. On the other hand,

$$R \geq (1 - \delta - \varepsilon) \frac{\log(q/2)}{\log q} = \left(\frac{(1 - P_i - P_d)(1 - P_s)}{1 - P_i} - \eta - \varepsilon \right) \frac{\log(q/2)}{\log q}.$$

Since $h_q(P_s) \rightarrow P_s$ and $\log(q/2)/\log q \rightarrow 1$ as $q \rightarrow \infty$, the claimed asymptotic comparison follows.

References

1. Briffa, J.A., Schaathun, H.G., Wesemeyer, S.: An improved decoding algorithm for the Davey–MacKay construction. In: IEEE International Conference on Communications, pp. 1–5. IEEE (2010). <https://doi.org/10.1109/ICC.2010.5502293>
2. Davey, M.C., MacKay, D.J.C.: Reliable communication over channels with insertions, deletions, and substitutions. IEEE Transactions on Information Theory **47**(2), 687–698 (2001). <https://doi.org/10.1109/18.910582>
3. Christ, M., Gunn, S.: Pseudorandom error-correcting codes. In: Advances in Cryptology—CRYPTO 2024, Part VI, LNCS, vol. 14925, pp. 325–347. Springer (2024). https://doi.org/10.1007/978-3-031-68391-6_10
4. Huang, S., Li, X., Mao, S., Zhou, Z.: High-rate public-key pseudorandom codes for edit errors. CoRR abs/2605.19402 (2026). <https://doi.org/10.48550/arXiv.2605.19402>

5. Dehmel, F., Li, S.: A public-key-dependent adversarial-deletion ceiling for fixed-alphabet multi-bit pseudorandom codes. CoRR abs/2609.02943 (2026). <https://doi.org/10.48550/arXiv.2609.02943>
6. Golowich, N., Moitra, A.: Edit distance robust watermarks via indexing pseudorandom codes. In: Advances in Neural Information Processing Systems 37 (NeurIPS 2024), pp. 20645–20693 (2024). <https://doi.org/10.52202/079017-0652>
7. Christ, M., Golowich, N., Gunn, S., Moitra, A., Wichs, D.: Improved pseudorandom codes from permuted puzzles. In: Proceedings of the 58th Annual ACM Symposium on Theory of Computing (STOC), pp. 2130–2139. ACM (2026). <https://doi.org/10.1145/3798129.3800916>
8. Haeupler, B., Shahrasbi, A.: Synchronization strings and codes for insertions and deletions—a survey. IEEE Transactions on Information Theory **67**(6), 3190–3206 (2021). <https://doi.org/10.1109/TIT.2021.3056317>
9. Hemenway, B., Ron-Zewi, N., Wootters, M.: Local list recovery of high-rate tensor codes and applications. In: 58th IEEE Symposium on Foundations of Computer Science (FOCS), pp. 204–215. IEEE (2017). <https://doi.org/10.1109/FOCS.2017.27>
10. Lipton, R.J.: A new approach to information theory. In: STACS 94, LNCS, vol. 775, pp. 699–708. Springer (1994). https://doi.org/10.1007/3-540-57785-8_183
11. Guruswami, V., Smith, A.: Optimal rate code constructions for computationally simple channels. Journal of the ACM **63**(4), 35:1–35:37 (2016). <https://doi.org/10.1145/2936015>
12. Stinson, D.R.: Universal hashing and authentication codes. Designs, Codes and Cryptography **4**(4), 369–380 (1994). <https://doi.org/10.1007/BF01388651>
13. Sarkar, P.: A trade-off between collision probability and key size in universal hashing using polynomials. Designs, Codes and Cryptography **58**, 271–278 (2011). <https://doi.org/10.1007/s10623-010-9408-6>
14. Wang, P., Li, Y., Zhang, L., Zheng, K.: Related-key almost universal hash functions: Definitions, constructions and applications. In: Fast Software Encryption (FSE 2016), LNCS, vol. 9783, pp. 514–532. Springer (2016). https://doi.org/10.1007/978-3-662-52993-5_26