# Sparse VLSF Codes Optimization for Short-Packet Transmission via Saddlepoint Methods

Guodong Sun, Samir M. Perlaza, Philippe Mary, Jean-Marie Gorce

# Sparse VLSF Codes Optimization for Short-Packet Transmission via Saddlepoint Methods

Guodong Sun*, Samir M. Perlaza*, Philippe Mary†, Jean-Marie Gorce‡
* INRIA Université Côte d'Azur, 2004 Route des Lucioles, 06560 Valbonne, France
† INSA Rennes, 20 Avenue des Buttes de Coesmes, 35700 Rennes, France
‡ INRIA Lyon, 56 Boulevard Niels Bohr, 69100 Villeurbanne, France

***Abstract*—In this work, we present an optimization framework for sparse variable-length stop-feedback (VLSF) codes based on a saddlepoint approximation, which jointly optimizes the decoding configuration parameters. Thanks to the analytical tractability of a saddlepoint approximation, the framework enables efficient gradient-based optimization of such parameters for common memoryless channels, including the additive white Gaussian noise, binary symmetric, and binary erasure channels. We further propose a refined decoding rule that extends the conventional fixed-threshold rule and leads to a tighter achievability bound. Numerical results demonstrate that our framework provides near-optimal decoding configurations at low computational cost. Moreover, the results from our refined rule demonstrate that the fixed-threshold decoding rule is restrictive and that achievability bounds can be further tightened.**

***Index Terms*—Variable-length stop-feedback codes, sparse feedback, saddlepoint approximation, short packet transmission.**

## I. Introduction

Variable-length stop-feedback (VLSF) codes, where transmission continues until an acknowledgment is received, are central to many variable-length feedback schemes and play an important role in channel coding [1], [2]. For short-packet communication, VLSF codes can improve the capacity-achieving (first-order) rate and eliminate the finite-blocklength (second-order) penalty when the stop feedback is available after every channel use and no delay constraints are imposed [2]. In practice, however, these gains may be restricted by feedback limitation, such as delay constraints [2]. For instance, [3] shows that probabilistic delay constraints allow for second-order improvements but do not eliminate the penalty. Similarly, extensions to noisy feedback channels are studied in [4], which shows that in such settings, fixed-length codes without feedback may be preferable.

The practical use of VLSF codes is constrained by the cost of reserving feedback channel resources for possible transmission after each channel use, even when the resources are idle. This motivates the use of sparse VLSF codes, which perform decoding and send stop feedback only at a small number of instants [5]. These schemes are widely implemented through hybrid automatic repeat request with incremental redundancy [6]. While [5], [7] examine periodic sparse VLSF codes, more recent studies [8]–[11] focus on optimizing the decoding and feedback instants. These works show that the achievability bound of sparse VLSF can approach that of dense VLSF with only a few well-chosen decoding instants. In [8], random stopping times are approximated to be Gaussian-distributed, and a sequential differential optimization (SDO) algorithm is proposed to determine the optimal decoding instants. Building on this approach, the authors of [9] derive achievability bounds that characterize the non-asymptotic rate at moderate blocklengths, while leaving the development of tight bounds for short packets as an open challenge.

To address the short-packet regime, the authors of [10], [11] formulate the computation of the achievability bound as a two-step optimization problem: they first optimize the decoding instants using a refined SDO algorithm; then, they optimize the decoding threshold via search, a step that may incur high computational cost. To apply the SDO algorithm, the information-density distribution for the binary-input additive white Gaussian noise (BI-AWGN) channel is approximated using Edgeworth [12] and Petrov [13] expansions for different ranges of blocklength. For the binary symmetric (BSC) and binary erasure (BEC) channels, the information density distributions are computed exactly. These channel-specific techniques limit both generality and analytical tractability; for example, the AWGN channel remains unexplored. The main contributions of the present work are summarized as follows:

- We propose a refined decoding rule that anticipates the final decoding attempt by employing maximum-likelihood decoding, for which the achievability bound is determined by fixed-blocklength bounds [14].
- A saddlepoint approximation [15] is proposed to characterize the probability distribution of the information-density for a variety of channels, i.e. AWGN, BSC, and BEC. An accurate analytical expressions that support efficient gradient-based optimization is then obtained and can be solved using a gradient-based mixed-integer nonlinear programming (MINLP) solver, e.g., Juniper [16].
- The proposed framework yields optimal or near-optimal decoding instants. This significantly enhances both the efficiency of the optimization and its generality across different channels. In addition, the refined decoding rule provides a tighter achievability bound.

The remainder of this paper is organized as follows. Section II presents the problem and saddlepoint approximation. Section III details the optimization method. Section IV discusses numerical results, and Section V concludes our contribution.

## II. System Model

### A. Sparse VLSF codes

Consider a memoryless channel $P_{Y|X}$ with input $X \sim P_X$ and output $Y$ taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. For a blocklength $n \in \mathbb{N}_{>0}$, the input-output sequence is denoted by

$$(x^n, y^n) = \big((x_1, y_1), \ldots, (x_n, y_n)\big) \in \mathcal{X}^n \times \mathcal{Y}^n. \tag{1}$$

The information density function is defined as

$$\imath_n : \begin{cases} \mathcal{X}^n \times \mathcal{Y}^n \longrightarrow \mathbb{R} \\ (x^n, y^n) \longmapsto \log \frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}P_{Y^n}}(y^n), \end{cases} \tag{2}$$

where $P_{Y^n|X^n} = (P_{Y|X})^n$ and $P_{Y^n} = P_Y^n$ for the memoryless channel, with $P_Y$ the marginal on $\mathcal{Y}$ induced by $P_{Y|X}P_X$. We assume the Radon–Nikodym derivative $\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}$ exists for $P_X$-almost every $x$ and the information density function can be written as

$$\imath_n(x^n, y^n) = \sum_{i=1}^{n} \log \frac{\mathrm{d}P_{Y|X=x_i}}{\mathrm{d}P_Y}(y_i). \tag{3}$$

When the blocklength $n$ is clear from context, we omit the subscript and write $\imath(x^n, y^n)$.

A sparse VLSF code specified by $(\mathcal{T}, M, \epsilon)$, where $\mathcal{T} \triangleq \{n_1, \ldots, n_t\} \in \mathbb{N}_{>0}$, where $t = |\mathcal{T}|$, is the set of admissible decoding instants, $M$ is the number of messages, and $\epsilon$ is the average error probability constraint. A message, denoted by $W$, is uniformly distributed over the message set $\mathcal{M} \triangleq \{1, \ldots, M\}$. A random codebook with independent and identically distributed (i.i.d.) $n_t$ length codewords is a tuple

$$\mathcal{C} \triangleq \big(X_1^{n_t}, \ldots, X_M^{n_t}\big), \quad X_m^{n_t} \sim P_X^{n_t}, \quad m \in \mathcal{M}. \tag{4}$$

Let $\gamma > 0$ denote an information density threshold. At each decoding instant $n \in \mathcal{T}$, the decoder evaluates the information density for all codewords and proceeds as follows: (1) Continue decoding if no codeword's information density exceeds $\gamma$; (2) Make a decision if there is a unique codeword whose information density exceeds $\gamma$, outputting the index of this codeword; (3) Declare an error if more than one codeword's information density exceeds $\gamma$. For each $m \in \mathcal{M}$, define the stopping time of the $m$-th codeword:

$$\tau_m \triangleq \min\{n \in \mathcal{T} : \imath(X_m^n, Y^n) \geq \gamma\}. \tag{5}$$

The stopping time of the decoder is

$$\tau^* \triangleq \min_{m \in \mathcal{M}} \tau_m, \tag{6}$$

with the convention $\tau^* = n_t$ if no codeword reaches the threshold by the final decoding instant $n_t$. The achievable rate of the code is defined as the ratio of the message size (in bits) to the expected minimum stopping time [2]

$$r = \log_2(M)/\mathbb{E}[\tau^*]. \tag{7}$$

Yavas et al. [9, Theorem 1] established an achievability bound for sparse VLSF codes. Building on this result, Yang et al. [11] obtained the maximal achievability bound by solving the following optimization problem, referred as P. 8

$$\min_{\gamma, \mathcal{T}} \quad n_1 + \sum_{j=1}^{|\mathcal{T}|-1} (n_{j+1} - n_j)\mathbb{P}[\imath(X_1^{n_j}, Y^{n_j}) < \gamma], \tag{8a}$$

$$\text{s.t.} \quad \mathbb{P}[\imath(X_1^{n_t}, Y^{n_t}) < \gamma] + (M-1)e^{-\gamma} \leq \epsilon, \tag{8b}$$

where (8a) is an upper bound on the expected decoding time $\mathbb{E}[\tau^*]$ under the threshold decoding rule. The error probability constraint (8b) is determined by the final decoding attempt. Specifically, $(M-1)e^{-\gamma}$ upper-bounds the probability that any non-transmitted codeword exceeds the threshold before the transmitted one (false-alarming), while $\mathbb{P}[\imath(X_1^{n_t}, Y^{n_t}) < \gamma]$ represents the probability that the information density of the transmitted codeword, assuming it is $X_1^n$, fails to reach the threshold at the final decoding attempt (missed-detection). This formulation also accounts for intermediate decoding attempts, since the false-alarming term $(M-1)e^{-\gamma}$ is already included in the constraint for each prior decoding attempt. Note that this error probability constraint aligns with the one used to determine an achievability bound for the fixed-blocklength codes, i.e., Shannon's bound [14, Theorem 2].

Since no further transmissions occur after the final decoding attempt, we propose a two-stage decoding strategy: during the intermediate decoding attempts, the decoder applies the threshold-based rule, while at the final decoding attempt, it selects the codeword with the maximal information density. For this last decoding attempt, the achievable codebook size $M_{fb}^*(n_t, \epsilon)$ has been investigated via random coding in [14]. We now formulate the optimization problem for the refined decoding rule, which we refer to as P. 9:

$$\min_{\gamma, \mathcal{T}} \quad n_1 + \sum_{j=1}^{|\mathcal{T}|-1} (n_{j+1} - n_j)\mathbb{P}[\imath(X_1^{n_j}, Y^{n_j}) < \gamma], \tag{9a}$$

$$\text{s.t.} \quad (M-1)e^{-\gamma} \leq \epsilon, \tag{9b}$$

$$M \leq M_{fb}^*(n_t, \epsilon). \tag{9c}$$

As shown in P. 8 and P. 9, the key component is the CDF $\mathbb{P}[\imath(X_1^{n_j}, Y^{n_j}) < \gamma]$ at any decoding instant $n_j \in \mathcal{T}$. For notational convenience, for $n \in \mathbb{N}_{>0}$ we define

$$S_n \triangleq \imath(X_1^n, Y^n) = \sum_{i=1}^{n} \imath(X_{1,i}, Y_i) = \sum_{i=1}^{n} Z_i, \tag{10}$$

where $Z_i \triangleq \imath(X_{1,i}, Y)$ denotes the single-letter information density of the transmitted codeword $X_1^{n_t}$ at time $i$. We also write $Z \triangleq \imath(X_1, Y)$ for a generic single-letter information density.

### B. Saddlepoint approximation

Next, we introduce the saddlepoint approximation of $\mathbb{P}[S_n < \gamma]$. Let $K_Z(s)$ denote the cumulant generating function (CGF) of $Z$. For $S_n = \sum_{i=1}^n Z_i$, the CGF is $K_{S_n}(s) = nK_Z(s)$ since $Z_i$ are i.i.d. The saddlepoint $\hat{s}$ is then computed by solving the $K'_{S_n}(\hat{s}) = \gamma$, where $K'_{S_n}(s)$ is the first derivative of the CGF.

If $Z$ has a density, the CDF of $\mathbb{P}[S_n < \gamma]$ for each $n \in \mathbb{N}_{>0}$ can be accurately approximated using the Lugannani-Rice saddlepoint approximation [15, 1.2.1] [17]

$$\mathbb{P}[S_n < \gamma] \approx \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}) & \text{if } \gamma \neq \mathbb{E}[S_n], \\ \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{3/2}} & \text{if } \gamma = \mathbb{E}[S_n], \end{cases} \tag{11}$$

where $\hat{w} = \text{sgn}(\hat{s})\sqrt{2(\hat{s}\gamma - K(\hat{s}))}$ and $\hat{u} = \hat{s}\sqrt{K''(\hat{s})}$. The functions $\phi$ and $\Phi$ denote the standard normal probability density and CDF, respectively, and sgn(.) is the sign function.

For discrete $Z$, a first-order continuity correction [15, 1.2.3] is applied to account for the lattice structure of $S_n$. Let $k$ denote the smallest attainable lattice point above $\gamma$ and the saddlepoint $\hat{s}$ is then computed by solving $K'_{S_n}(\hat{s}) = k$. $\mathbb{P}[S_n < \gamma] = \mathbb{P}[S_n < k]$ is approximated by the discrete saddlepoint formula:

$$\mathbb{P}[S_n < \gamma] \approx \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\bar{u}) & \text{if } k \neq \mathbb{E}[S_n], \\ \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(\frac{K'''(0)}{6K''(0)^{3/2}} - \frac{1}{2\sqrt{K''(0)}}\right) & \text{if } k = \mathbb{E}[S_n], \end{cases} \tag{12}$$

where $\hat{w}$ is defined as in the continuous case, and the curvature correction term is given by $\bar{u} = (1 - e^{-\hat{s}})\sqrt{K''(\hat{s})}$.

For both the continuous and discrete saddlepoint approximations, when $\gamma \neq \mathbb{E}[S_n]$ in (11) or $k \neq \mathbb{E}[S_n]$ in (12), the quantity $\Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u})$ is smooth with respect to the saddlepoint $\hat{s}$, since $\Phi$, $\phi$, $\hat{w}$, and $\hat{u}$ are smooth functions of $\hat{s}$. The special cases at $\gamma = \mathbb{E}[S_n]$ or $k = \mathbb{E}[S_n]$ are defined by continuation, and hence the saddlepoint approximation is a continuous function in all cases.

## III. Decoding Schedule Optimization

In this section, we show that for the AWGN, BSC, and BEC channels, the CDF $\mathbb{P}[S_n < \gamma]$ is continuous w.r.t. the channel parameters, enabling the application of gradient-based methods to solve the optimization problems.

### A. AWGN channel

For the AWGN channel $Y = X + N$, where $X \sim \mathcal{N}(0, p_0)$ and $N \sim \mathcal{N}(0, 1)$, where $p_0$ denotes the normalized SNR. Here, $\mathcal{N}$ denotes the normal distribution. The information density is

$$Z = \frac{1}{2}\log(1 + p_0) + \frac{1}{2}\left(\frac{Y^2}{p_0 + 1} - (Y - X)^2\right). \tag{13}$$

Let $\tilde{Z} = Z - \frac{1}{2}\log(1 + p_0)$ denote the shifted information density of $Z$ and define $\tilde{S}_n = \sum_{i=1}^n \tilde{Z}_i$ with the corresponding shifted threshold at time $n$:

$$\tilde{\gamma}_n = \gamma - \frac{n}{2}\log(1 + p_0). \tag{14}$$

The following lemma provides a closed-form expression for the saddlepoint:

**Lemma 1.** *The CGF of $\tilde{S}_n$ is given by*

$$K_{\tilde{S}_n}(s) = -\frac{n}{2}\log\left(1 - \frac{p_0 s^2}{p_0 + 1}\right), \tag{15}$$

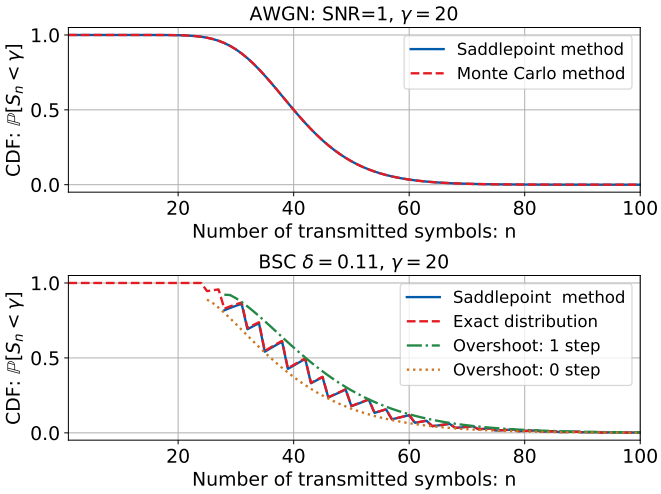


Fig. 1. Saddlepoint approximation versus the true distribution. For AWGN the true distribution is estimated by Monte Carlo simulation, and for the BSC it is computed exactly. For BSC with discrete $S_n$, overshoot at threshold crossings is between zero and one step; Saddlepoint approximations for both fixed overshoot extremes are shown.

*and the corresponding saddlepoint $\hat{s}$ is given by*

$$\hat{s} = \frac{-n + \sqrt{n^2 + 4\tilde{\gamma}^2(p_0 + 1)/(p_0)}}{2\tilde{\gamma}}. \tag{16}$$

*Proof.* See Appendix A □

Substituting this closed-form saddlepoint into (11) allows $\mathbb{P}[S_n < \gamma]$ to be evaluated as a closed-form function with respect to $n$ and $\gamma$. As illustrated in the upper subfigure of Fig. 1, the tail distribution obtained via the saddlepoint approximation closely matches the true tail distribution computed through extensive Monte Carlo simulations ($10^6$ trials). This highlights both the accuracy and the inherent smoothness of the saddlepoint approximation. This continuity allows the optimization problem to be solved using a standard gradient-based MINLP solver, which relaxes $n \in \mathbb{N}_{>0}$ to $n \in \mathbb{R}$ and then applies a local search. The corresponding results are presented in the next section.

### B. BSC or BEC

For a BSC with crossover probability $\delta \in (0, \frac{1}{2})$ and input $X \sim \text{Bernoulli}(\frac{1}{2})$, the information density of each transmitted symbol takes two values:

$$Z = \begin{cases} \log 2(1 - \delta) & \text{w.p. } 1 - \delta, \\ \log 2\delta & \text{w.p. } \delta, \end{cases} \tag{17}$$

where "w.p." means "with probability." Define the shifted information density $\tilde{Z} \triangleq Z - \log 2\delta$. At time $n$, $\tilde{S}_n = \sum_{i=1}^n \tilde{Z}_i = \sum_{i=1}^n Z_i - n\log 2\delta$. The CGF of $\tilde{S}_n$ is

$$K_{\tilde{S}_n}(s) = n\log\left(\delta + (1 - \delta)e^{\log\frac{1-\delta}{\delta}s}\right). \tag{18}$$

Recall $k$ is the smallest attainable lattice point exceeding $\gamma$. After shifting, the corresponding lattice point becomes $k -$

$n \log 2\delta$. Solving $K'_{S_n}(s) = k - n\log 2\delta$ gives a closed form expression for the saddlepoint:

$$\hat{s} = \log\left(\frac{(k/n - \log 2\delta)\delta}{(\log\big(2(1-\delta)\big) - k/n)(1-\delta)}\right) \Big/ \log\left(\frac{1-\delta}{\delta}\right). \tag{19}$$

Substituting $\hat{s}$ into (12) allows $\mathbb{P}[S_n < \gamma]$ to be in closed-form as a function of $n$ and the overshooting lattice point $k$.

Similarly, for the BEC with erasure probability $\delta$ under the input $X \sim \text{Bernoulli}(\frac{1}{2})$, the information density $Z$ of each transmitted symbol is a discrete random variable:

$$Z = \begin{cases} 1 & \text{w.p. } 1-\delta, \\ 0 & \text{w.p. } \delta. \end{cases} \tag{20}$$

which can be viewed as a special case of BSC in (17), and we omit the derivation of the saddlepoint for BEC.

For BSC and BEC, the cumulative information density $S_n$ is defined on a lattice. For lattice-valued $S_n$, the smallest lattice point $k$ exceeding $\gamma$ lies above the latter by a non-negative margin $D \triangleq k - \gamma$. This margin is smaller than the step size $|\log(2(1-\delta)) - \log(2(\delta))| = \log\big((1-\delta)/\delta\big)$ for $\delta \in (0, \frac{1}{2})$. For BSC with $\gamma \in \mathbb{R}$, we have $D \in [0, \log\left((1-\delta)/\delta\right))$. To enable gradient-based MINLP optimization, we first apply a continuous correction for approximating the overshoot.

By fixing the overshoot $D$ within this range, the CDF with overshoot lies between the CDF approximations without overshoot $\mathbb{P}[S_n < \gamma]$ and $\mathbb{P}[S_n < \gamma + \log(\frac{1-\delta}{\delta})]$, as shown in the lower subfigure of Fig. 1. This continuous correction ensures that the saddlepoint approximation is smooth in $n$ and $\gamma + D \in \mathbb{R}$, and hence suitable for gradient-based optimization. We first solve this continuously corrected version for both $D = 0$ and $D = 1$, following the approach used for AWGN channels, which provides an optimistic set of decoding instants ($D = 0$) and a pessimistic set ($D = 1$). Then we refine the solution via a local search within these decoding ranges to obtain the optimal parameters.

### C. Instability of saddlepoint approximation

A technical difficulty for applying the saddlepoint approximation, specifically for discrete channels, is the numerical instability near the mean, i.e., when $\gamma \approx \mathbb{E}[S_n]$ (12). In this region, the saddlepoint $\hat{s}$, obtained by solving, $K'(\hat{s}) = \gamma$ is approximately zero. Then, $\hat{w} = \text{sgn}(\hat{s})\sqrt{2(\hat{s}\gamma - K(\hat{s}))}$ and $\bar{u} = 1 - e^{-\hat{s}}\sqrt{K''(\hat{s})}$ are also approximately zero. The difference $\frac{1}{\hat{w}} - \frac{1}{\bar{u}}$ in (12) is then numerically ill-conditioned, and the impact of approximating overshoot is huge. To address this, we apply a second-order Taylor expansion of the CGF $K(\hat{s})$ for $\hat{s}$ around 0:

$$K(\hat{s}) = \mu\hat{s} + \frac{1}{2}\sigma^2\hat{s}^2 + O(\hat{s}^3), \tag{21}$$

where $\mu$ and $\sigma^2$ are the mean and variance of $S_n$. Then $K'(s) \approx \mu + \sigma^2\hat{s}$ and solving $K'(\hat{s}) = \gamma$ gives $\hat{s} \approx \frac{\gamma - \mu}{\sigma^2}$. Substituting into $\hat{w}$, we have

$$\hat{w} = \sqrt{2(\hat{s}\gamma - K(\hat{s})} \approx \frac{\gamma - \mu}{\sigma}. \tag{22}$$

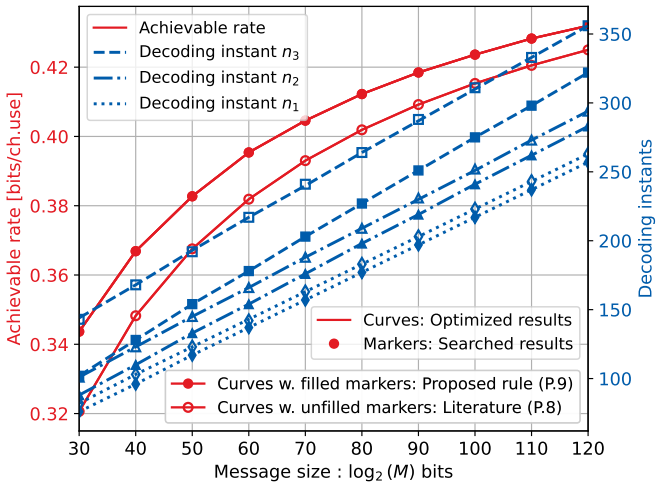


Fig. 2. Achievable rate (left) and optimal decoding instants (right) versus message size (in bits) in an AWGN channel (SNR $= 1$, $t = 3$ attempts).

So that near the mean, the CDF of $S_n$ can be well approximated by $\Phi(\frac{\gamma-\mu}{\sigma})$. The CDF is then given by

$$\mathbb{P}[S_n < \gamma] = \begin{cases} \Phi(\frac{\gamma-\mu}{\sigma}) & |\frac{\gamma-\mu}{\sigma}| \le \epsilon_s \\ (12) & |\frac{\gamma-\mu}{\sigma}| > \epsilon_s, \end{cases} \tag{23}$$

where $\epsilon_s$ is a small empirically chosen threshold (typically $\epsilon_s = 0.1$) that separates the unstable central region from the tails, where the full saddlepoint approximation in (12) is used. Note that the piecewise smooth approximation may introduce a discontinuity at the transition. This has little practical impact, as the optimizer still converges reliably given the narrowness of the corrected region, as confirmed in Section IV.

## IV. Numerical Results

We evaluate our framework by comparing the optimized decoding attempts obtained with our method against those from a brute-force search. The gradient-based MINLP problem is solved using the Juniper package in Julia [16].

### A. Evaluation of the optimization framework

In the following, $t = 3$ decoding attempts, message sizes between 30 to 120 bits, SNR $= 1$, and $\epsilon = 10^{-3}$ are considered, unless otherwise mentioned.

*a) Computational efficiency:* For both AWGN and BSC channels, the gradient-based optimization is far more efficient than brute-force search: it produces results under one second, whereas the brute-force approach can take several hours. Moreover, brute-force search is computationally feasible only for $t \le 3$ due to the exponential growth of the search space, which highlights the advantage of the gradient-based method.

*b) AWGN:* Fig. 2 presents the achievable rate (vertical left axis) and the corresponding decoding instants (vertical right axis) with respect to the message size. The gradient-based optimization (curves) closely matches the brute-force search (markers), confirming the effectiveness of our approach.

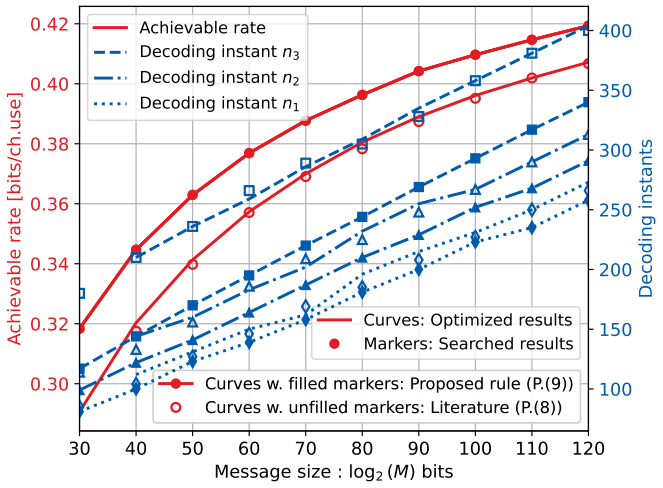


Fig. 3. Achievability rate (left) and optimal decoding instants (right) versus message size (in bits) in a BSC ($\delta = 0.11$, $t = 3$ attempts).

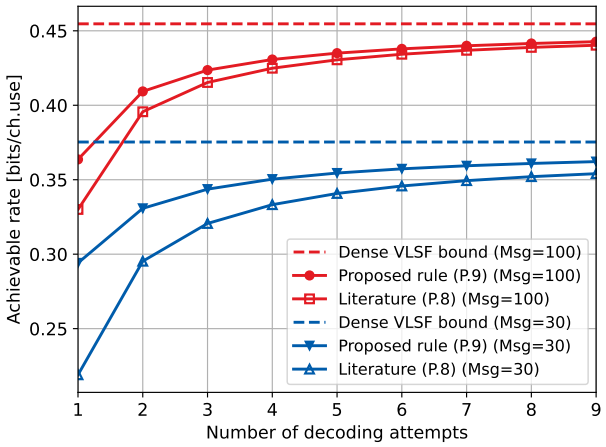


Fig. 4. Achievable rate of sparse VLSF codes versus number of decoding times.

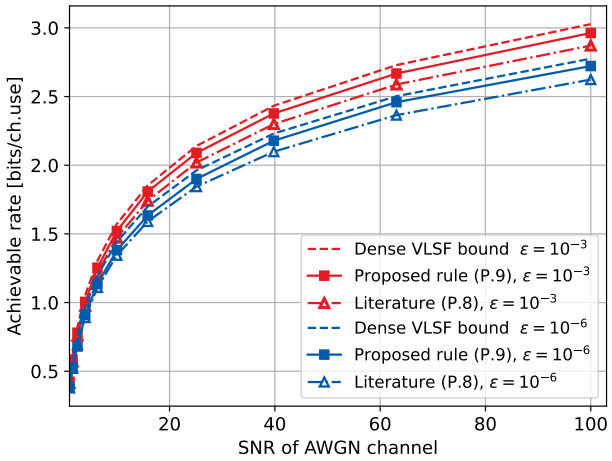


Fig. 5. Achievable rate with respect to SNR. The target decoding error is set as $\epsilon_0 = \{10^{-3}, 10^{-6}\}$, respectively.

Compared with the rule in the literature (P. 8), the proposed decoding rule (P. 9) achieves higher rates, indicating a closer approach to the fundamental limit of sparse VLSF codes. This improvement is more significant for small message sizes, with nearly an 8% gain at 30 bits, decreasing to approximately 2% at 120 bits. Hence, the proposed rule accurately captures the performance limits for short-packet communication. On the right axis, the optimized decoding instants under the proposed rule (filled markers) are consistently earlier than those from the literature rule (unfilled markers). The gap is especially evident at later decoding stages (e.g., $n_2, n_3$), suggesting that threshold-based decoding is restrictive in the later stage. Further work is needed to refine the decoding rule for stages that allow early stopping.

*c) BSC:* Fig. 3 shows the results for a BSC with crossover probability $\delta = 0.11$, corresponding to a Shannon capacity of approximately $0.5$ bit/s/Hz.

Fig. 3 shows a similar advantage for the BSC as observed in the AWGN channel: the proposed decoding rule achieves higher rates as the decoding times are anticipated. This demonstrates that the saddlepoint approximation, combined with gradient-based optimization, provides a general approach applicable across different channels, enabling a systematic investigation of decoding rules. However, as seen in comparison with brute-force search, gradient-based optimization does not guarantee globally optimal decoding schedules for the BSC, since gradients cannot fully capture the exact CDF of threshold crossing with overshoot (shown in the lower panel of Fig. 1). Nevertheless, the results are near-optimal, as the decoding instants and achievable rates obtained by optimization closely align with those obtained by brute force search.

### *B. Evaluation of the proposed achievable bound*

Next, we focus on the AWGN channel, which is of primary interest in communication problems, having noted that similar performance trends are observed for the BSC.

Fig. 4 shows achievable rates versus the number of sparse decoding attempts for SNR$=1$ and message sizes of $\{30, 100\}$ bits. Note that the reference curve representing dense decoding after every received symbol is also reported [2]. The gap to the reference closes quickly, indicating that near-optimal performance can be achieved with only a few attempts. The proposed rule lies closer to the reference, achieving substantially better performance even with one or two attempts. As the number of decoding attempts increases, the gap narrows, showing that the refined rule is most significant for fewer attempts.

In general, the performance of finite blocklength codes depends on parameters such as message size, blocklength, target error probability, and channel conditions. The current framework allows us to analyze the interactions among these parameters. Fig. 5 shows achievable rates for both decoding rules w.r.t. SNR for message size of 100 bits and $\epsilon = \{10^{-3}, 10^{-6}\}$. The gap between sparse and dense decoding widens at higher SNR, and the difference between the lit-

erature rule and the proposed rule also grows. Nevertheless, overall trends remain almost identical over $\epsilon$, indicating that SNR has a greater impact than the target error on sparse VLSF performance.

## V. Conclusion

In this work, we developed a gradient-based optimization framework to compute the achievability bound of sparse VLSF codes, leveraging the analyticity of the saddlepoint approximation of the information density's CDF. This framework enables joint optimization of decoding schedules, providing an efficient framework for characterizing the performance of sparse VLSF codes. Beyond its practical application, this method facilitates the investigation of alternative decoding rules and offers insights into finite blocklength coding through refinement of the final decoding attempt. Future work may investigate alternatives to the fixed-threshold decoding scheme in the intermediate decoding steps to achieve further rate improvements, which may otherwise be restrictive.

## VI. Acknowledgment

This work is supported by the French National Agency for Research project titled France 2030 PEPR réseaux du Futur under grant ANR-22-PEFT-0010 and France 2030 under grant ANR-23-CMAS-0023.

## Appendix

### A. Computing saddlepoint for the AWGN channel

The shifted $\tilde{Z}$ is

$$\tilde{Z} = \frac{1}{2}\left(\frac{Y^2}{p_0+1} - (Y-X)^2\right) = \frac{1}{2}\left(\frac{(X+N)^2}{p_0+1} - N^2\right), \tag{24}$$

Define

$$V = \frac{X+N}{\sqrt{p_0+1}} \sim \mathcal{N}(0,1), \quad W = N \sim \mathcal{N}(0,1). \tag{25}$$

Since both depend on the same noise variable $N$, $V$ and $W$ are correlated. Their covariance is

$$\rho \triangleq \operatorname{cov}(V,W) = \mathbb{E}[VW] = \frac{\mathbb{E}[XN]+\mathbb{E}[N^2]}{\sqrt{p_0+1}} = \sqrt{\frac{1}{p_0+1}}. \tag{26}$$

From (24), $\tilde{Z}$ is a quadratic form

$$\tilde{Z} = \frac{1}{2}(V^2 - W^2) = \begin{pmatrix} V & W \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} V \\ W \end{pmatrix}. \tag{27}$$

The moment generating function (MGF) of such a quadratic form is given by [18, Theorem 5.2b]

$$\begin{aligned} M_{\tilde{Z}}(s) &= \left(\det\left(I - 2s\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}\right)\right)^{-1/2} \\ &= \frac{1}{\sqrt{1-s^2(1-\rho^2)}} \end{aligned} \tag{28}$$

MGF exists for values of $s$ within the region of convergence

$$s < \sqrt{(p_0+1)/p_0}. \tag{29}$$

Since $\tilde{S}_n = \sum_{i=1}^n \tilde{Z}$, the CGF of $\tilde{S}_n$ is

$$K_{\tilde{S}_n}(s) = \log\left(M_{\tilde{Z}}(s)\right)^n = -\frac{n}{2}\log\left(1 - \frac{p_0 s^2}{p_0+1}\right). \tag{30}$$

To find the saddlepoint, we have $K'(s) = \frac{np_0 s}{(p_0+1)-p_0 s^2}$. The saddlepoint $\hat{s}(\tilde{\gamma})$ is the solution to $K'(s) = \tilde{\gamma}$, which gives

$$\hat{s}(\tilde{\gamma}) = \frac{-n + \sqrt{n^2 + 4\tilde{\gamma}^2(p_0+1)/(p_0)}}{2\tilde{\gamma}}, \tag{31}$$

where the other root does not satisfy the convergence condition in (29).

## References

[1] G. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, 2011.

[3] Y. Altuğ, H. V. Poor, and S. Verdú, "Variable-length channel codes with probabilistic delay guarantees," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 642–649, IEEE, 2015.

[4] J. Östman, R. Devassy, G. Durisi, and E. G. Ström, "Short-packet transmission via variable-length codes in the presence of noisy stop feedback," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 214–227, 2020.

[5] A. R. Williamson, T.-Y. Chen, and R. D. Wesel, "Variable-length convolutional coding for short blocklengths with decision feedback," *IEEE Transactions on Communications*, vol. 63, no. 7, pp. 2389–2403, 2015.

[6] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[7] S. H. Kim, D. K. Sung, and T. Le-Ngoc, "Variable-length feedback codes under a strict delay constraint," *IEEE Communications Letters*, vol. 19, no. 4, pp. 513–516, 2015.

[8] K. Vakilinia, S. V. Ranganathan, D. Divsalar, and R. D. Wesel, "Optimizing transmission lengths for limited feedback with nonbinary ldpc examples," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2245–2257, 2016.

[9] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length sparse feedback codes for point-to-point, multiple access, and random access channels," *IEEE Transactions on Information Theory*, vol. 70, no. 4, pp. 2367–2394, 2023.

[10] H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Variable-length stop-feedback codes with finite optimal decoding times for bi-awgn channels," in *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 1527–1532, IEEE, 2022.

[11] H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Incremental redundancy with ack/nack feedback at a few optimal decoding times," *arXiv preprint arXiv:2205.15399*, 2022.

[12] P. Hall, *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, 2013.

[13] V. V. Petrov, *Sums of independent random variables*, vol. 82. Springer Science & Business Media, 2012.

[14] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[15] R. W. Butler, *Saddlepoint approximations with applications*, vol. 22. Cambridge University Press, 2007.

[16] O. Kröger, C. Coffrin, H. Hijazi, and H. Nagarajan, "Juniper: An open-source nonlinear branch-and-bound solver in julia," in *International conference on the integration of constraint programming, artificial intelligence, and operations research*, pp. 377–386, Springer, 2018.

[17] R. Lugannani and S. Rice, "Saddle point approximation for the distribution of the sum of independent random variables," *Advances in applied probability*, vol. 12, no. 2, pp. 475–490, 1980.

[18] A. C. Rencher and G. B. Schaalje, *Linear models in statistics*. John Wiley & Sons, 2008.