LLR-Based Attention-Weighted Projection Decoding for Sparse Vector-Coded Short Packet Transmissions

Mody Sy
EURECOM, 06410 BIOT, France
mody.sy@eurecom.fr

***Abstract*—This paper presents an enhanced sparse vector coding scheme tailored for short-packet communication over noisy wireless channels. Conventional SVC often uses random (Gaussian, Bernoulli) or structured (partial Hadamard) spreading/projection matrices with baseline decoding based on naive projection scores to recover the sparse signal support, which may suffer from reliability degradation under fading conditions. To address this, we introduce an enhanced decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to their reliability, giving more importance to more reliable measurements. Simulation results over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels show that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Moreover, we extend our analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results highlight that the proposed SVC scheme, employing a partial Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal maximum likelihood (ML) decoding at very low BLER targets, while offering significant computational complexity advantages, making it a promising short-packet coding candidate for ultra-reliable, low-latency 6G short-block-length uplink/downlink channels.**

***Index Terms*—Sparse Vector Coding, Short Packet Communications, LLR-based Attention-Weighted Projection decoding.**

I. Introduction

Short-packet transmission has become a critical element in emerging communication systems, including ultra-reliable low-latency communications (URLLC) and Internet of Things (IoT) networks. In recent years, particularly with the advent of beyond-5G technologies, significant efforts have been devoted to improving short-packet transmission, with a strong focus on signal coding and receiver design aspects [1]–[4]. In this context, sparse vector coding (SVC) techniques, which encode information into sparse vectors, are gaining popularity due to their ability to convey information efficiently with reduced bandwidth, low latency, low complexity, and high reliability. Since its introduction in [5], SVC has been extensively investigated as an efficient transmission paradigm for short-packet URLLC [6]–[10]. Several extensions have focused on improving reliability by enriching the information-bearing structure of sparse vectors. In particular, enhanced SVC (ESVC) and sparse superposition–based schemes encode information jointly in the support and in the non-zero values, often relying on complex-valued constellations at the support elements to increase spectral efficiency [6], [7]. While these approaches achieve notable BLER gains, their reliance on complex signaling and sophisticated decoding may increase processing latency and hardware complexity, which can be undesirable in ultra-low-latency regimes. In parallel, conventional SVC methods, as presented in the seminal work of Ji *et al.* [5], often rely on random measurement matrices, typically Gaussian or Bernoulli ensembles, owing to their theoretical guarantees for sparse recovery. Despite these advantages, random matrices present practical challenges: they require substantial storage and computational resources, and their inherent randomness hinders structured hardware implementation. Complementary to symbol-domain and code-design enhancements, recent works have highlighted the role of structured measurement matrices in SVC and ESVC to reduce complexity and enable efficient implementations [8]. Structured measurement matrices for SVC, such as partial Hadamard matrices, therefore spark growing interest. Their deterministic construction enables fast transforms and low-complexity implementations. Nonetheless, classical decoding methods based on fixed correlations treat all measurements equally, which can lead to suboptimal performance when noise disproportionately corrupts certain measurements.

Motivated by these observations, the present work focuses on a real-valued, bipolar ESVC framework that is particularly well suited for short-packet transmission under stringent latency constraints. Unlike prior works that primarily enhance SVC through constellation design or code and spreading matrix construction, our contribution targets the receiver stage by introducing an adaptive decoding framework leveraging LLR-based attention-weighted Hadamard projections that dynamically reweight received signal measurements according to their reliability. In this respect, the attention weights are derived from per-element log-likelihood ratios (LLRs) that quantify the reliability of each received signal observations while accounting for the statistics of the transmitted sparsity-aware signal.

Simulation results demonstrate that the proposed LLR-based attention-weighted projection decoding consistently outperforms baseline projection decoding for both structured (partial Hadamard) and random (Gaussian) spreading matrices, particularly in challenging transmission scenarios involving frequency-selective fading channels. Moreover, the analysis is extended to a short-packet transmission framework with payloads of 3–11 bits over the 3GPP physical uplink control channel (PUCCH), comparing the proposed approach with the current standard based on 3GPP Reed–Muller coded transmission. We find that SVC using a partial Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding can competitively rival 3GPP RM codes under optimal maximum likelihood (ML) decoding at very low block error rate (BLER) targets. From a computational complexity perspective, the proposed SVC decoder exhibits significant advantages over conventional ML decoding methods, making it a promising ultra-reliable, low-latency short-packet coding solution for next-generation communication systems.
The remainder of this paper is organized as follows. Section II presents the general framework of sparse vector coding,

Section III describes the system model, Section IV focuses on the SVC receiver design, Section V provides numerical results and performance analysis, and Section VI concludes the paper.

II. General Framework of Sparse Vector Coding

A. Basics of Sparse Vector Coding

A binary message of length B is mapped to a sparse vector $\mathbf{x} \in \mathbb{R}^n$ with sparsity level $k = \|\mathbf{x}\|_0$, which counts its non-zero entries. The support of $\mathbf{x}$, denoted $\mathrm{supp}(\mathbf{x}) = \mathcal{S} \subset \{1, \ldots, n\}$, satisfies $|\mathcal{S}| = k$. Each entry in the support, $i \in \mathcal{S}$, is assigned a bipolar amplitude $\mathrm{x}_i \in \{-1, +1\}$. The total number of bits per message, including sign bits, is

$$B = \underbrace{\left\lfloor \log_2 \binom{n}{k} \right\rfloor}_{\texttt{Position bits}} + \underbrace{k}_{\texttt{Sign bits}} \quad \text{[bits]}. \tag{1}$$

Formally, the binary message is encoded into a sparse vector through the following two-step procedure:

- `Support (active position) selection:` The support set $\mathcal{S}$ is chosen from the $|\mathcal{C}| = \binom{n}{k}$ possible combinations, thereby encoding a first message block $\boldsymbol{m}_1$ of $\left\lfloor \log_2 \binom{n}{k} \right\rfloor$ bits.
- `Bipolar modulation (sign assignment):` Each active position is assigned a symbol ± 1 from a binary antipodal alphabet, thereby encoding a second message block $\boldsymbol{m}_2$ of k bits.

Thus, the total message structure $\boldsymbol{m} \in \mathbb{F}_2^{\left\lfloor \log_2 \binom{n}{k} \right\rfloor + k}$ can be expressed as $\boldsymbol{m} = [\boldsymbol{m}_1 \ \ \boldsymbol{m}_2]$ of B bits.

The sparse vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\mathrm{x}_i = \begin{cases} +1 & \text{if } i \in \mathcal{S} \text{ and the associated sign bit is } 0, \\ -1 & \text{if } i \in \mathcal{S} \text{ and the associated sign bit is } 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This mapping establishes a bijection between the binary message $\boldsymbol{m} \in \mathcal{M}$ and the sparse vector $\mathbf{x} \in \mathcal{X}$, since each message $\boldsymbol{m}$ corresponds to exactly one sparse vector $\mathbf{x}$, and vice versa.

$$\mathcal{L} : \mathcal{M} \to \mathcal{X} \quad | \quad \mathbf{x} = \mathcal{L}(\boldsymbol{m}), \tag{3}$$

where $\mathcal{M}$ denotes the set of transmitted binary messages, and $\mathcal{X} = \{\mathbf{x} \in \{-1, 0, +1\}^n : \|\mathbf{x}\|_0 = k\}$ represents the set of encoded sparse vectors. The function $\mathcal{L}$ is a discrete mapping implemented as a lookup table (LUT), defining a bijective correspondence between the binary message $\boldsymbol{m}$ and the sparse vector $\mathbf{x}$. SVC, as defined here and conceptually illustrated in Figure 1 for parameters $n = 8$ and $k = 2$, can be viewed as a *combinatorial encoding combined with bipolar modulation*.

The mapping between binary message combinations and sparse-vector pair-wise patterns, implementable as an LUT, is given in Table I. Positions marked with ☒ in the sparse vector indicate the locations of the bipolar symbols, thereby defining the support of the sparse vector.

TABLE I
SVC MAPPING TABLE : $n = 8, k = 2$.

$\boldsymbol{m}_1 : \lfloor \log_2 \binom{n}{k} \rfloor$ bits	DEC Idx:$(\boldsymbol{m}_1)_{10}$	Active Pos : $\mathcal{C}(\mathrm{Idx}+1, :)$	SVC Mapping Patterns
0000	0	(1 , 2)	☒ ☒ 0 0 0 0 0 0
0001	1	(1 , 3)	☒ 0 ☒ 0 0 0 0 0
⋮	⋮	⋮	⋮
0101	5	(1 , 7)	☒ 0 0 0 0 0 ☒ 0
⋮	⋮	⋮	⋮
1111	15	(3 , 6)	0 0 ☒ 0 0 ☒ 0 0
-	16	(3 , 7)	0 0 ☒ 0 0 0 ☒ 0
⋮	⋮	⋮	⋮
-	27	(7 , 8)	0 0 0 0 0 0 ☒ ☒

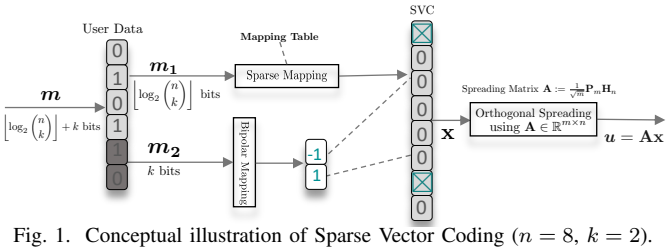


Fig. 1. Conceptual illustration of Sparse Vector Coding ($n = 8$, $k = 2$).

B. Structured SVC with Partial Hadamard Matrices

Partial Hadamard matrices are effectively well suited to overcome the limitations inherent in random measurement matrices in SVC. Let $\mathbf{H}_n \in \{\pm 1\}^{n \times n}$ denote the Hadamard matrix of order n, which is a structured orthogonal matrix. The projection matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is constructed from $\mathbf{H}_n$ by randomly selecting $m \leqslant n$ rows of $\mathbf{H}_n$:

$$\mathbf{A} := \frac{1}{\sqrt{m}} \mathbf{P}_m \mathbf{H}_n, \tag{4}$$

where $\mathbf{P}_m \in \{0, 1\}^{m \times n}$ is a selection matrix composed of m rows of the identity matrix corresponding to the chosen rows, i.e., $\mathbf{P}_m$ is a partial isometry. The normalization factor $\frac{1}{\sqrt{m}}$ ensures that each column of $\mathbf{A}$ has approximately unit norm (i.e., $\forall i \ \|\mathbf{a}_i\|_2^2 = 1$) and $\forall (i, j)$ the entry $a_{i,j} = \pm \frac{1}{\sqrt{m}}$. The *coherence* of $\mathbf{A}$: $\mu(\mathbf{A}) = \max_{i \neq j} |\mathbf{a}_i^\top \mathbf{a}_j|$ bounds the cross terms by $|\mathbf{a}_i^\top \mathbf{a}_\ell| \leq \mu$, s.t. $\mu \leq \sqrt{\frac{\log n}{m}}$.

The sparse vector is multiplied by the spreading matrix also referred to as the projection, sensing, or measurement matrix to generate a compressed signal, as in compressive sensing. At the receiver, the original sparse vector is reconstructed through sparse recovery. Conventional decoders are greedy or convex method-based like the well-known orthogonal matching pursuit (OMP) [11], and typically operate under a generic additive white Gaussian noise (AWGN) assumption and do not account for specific channel impairments like fading or interference.

In what follows, we show that SVC, employing orthogonal spreading via a partial Hadamard matrix, can be efficiently decoded using a sophisticated strategy based on LLR-driven attention-weighted projections.

C. System Model

The system under consideration encodes a sparse vector $\mathbf{x} \in \{-1, 0, +1\}^n$ with sparsity $k = \|\mathbf{x}\|_0 \ll n$ into a measurement vector $\mathbf{u} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ via multiplication by a partial Hadamard matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. The resulting signal is impaired by a memoryless complex fading channel, where the real and imaginary components of h_i have variance σ_h^2 each, with means given by the real and imaginary parts of μ_h, respectively. The channel is modeled by a diagonal fading matrix $\mathbf{H} = \mathrm{diag}(\mathrm{h}_1, \mathrm{h}_2, \ldots, \mathrm{h}_m)$, where $\mathrm{h}_i \sim \mathbb{C}\mathcal{N}(\mu_h, 2\sigma_h^2)$, and is subsequently transmitted over an additive white Gaussian noise channel $\mathbf{z} \sim \mathbb{C}\mathcal{N}(0, 2\sigma^2 \mathbf{I}_m)$. The received signal can thus be expressed as

$$\mathrm{y}_i = \mathrm{h}_i \, (\mathbf{a}_i^\top \mathbf{x}) + \mathrm{z}_i = \mathrm{h}_i \, u_i + \mathrm{z}_i \quad i = 1, 2, \ldots, m, \tag{5}$$

where $\mathbf{a}_i^\top \in \mathbb{R}^{1 \times n}$ denotes the i-th row of $\mathbf{A}$ and $u_i = \langle \mathbf{a}_i, \mathbf{x} \rangle = \sum_{j \in \mathcal{S}} a_{ij} x_j$ is a sum of at most k terms, each being $\pm 1/\sqrt{m}$, hence $|u_i| \leq k/\sqrt{m}$.

Equivalently, (5) can be expressed in a more compact vectorized form as

$$\mathbf{y} = \mathbf{h} \odot (\mathbf{A}\mathbf{x}) + \mathbf{z} = \mathbf{H}\mathbf{A}\mathbf{x} + \mathbf{z}, \tag{6}$$

where $\odot$ denotes the Hadamard (element-wise) product.

III. Receiver Design

At the receiver end, the objective is to reliably recover $\mathbf{x}$ from the noisy observations $\mathbf{y}$. A classical approach in SVC consists of estimating the support of the sparse vector using *correlation* or *projection* scores.

A. Baseline Projection Decoding (Simplified OMP)

First, in the considered scenario, namely transmission over fading channels, an equalization procedure is applied to mitigate channel-induced impairments:

$$\tilde{\mathrm{y}}_i = \Re\left(\frac{\mathrm{h}_i^* \cdot \mathrm{y}_i}{|\mathrm{h}_i|^2 + \varepsilon}\right), \quad \forall\, i, \quad (\mathrm{y}_i, \mathrm{h}_i) \in \mathbb{C}^2, \quad \forall\, \varepsilon > 0. \tag{7}$$

The baseline decoding technique, referred to as projection-based support recovery (simplified OMP), involves using the projection matrix $\mathbf{A}$ to correlate with the equalized received signal.

$$\mathbf{c} = \mathbf{A}^\top \tilde{\mathbf{y}} \quad \text{where} \quad \mathrm{c}_j = \mathbf{a}_j^\top \tilde{\mathbf{y}} = \langle \mathbf{a}_j, \tilde{\mathbf{y}} \rangle,\ j = 1, 2, \ldots, n. \tag{8}$$

$\mathbf{a}_j \in \mathbb{R}^m$ denotes the $j-th$ column of $\mathbf{A}$, and $\mathbf{c} \in \mathbb{R}^n$ contains the correlation scores.

The k indices with the largest correlation magnitudes are subsequently selected to form the estimated support $\widehat{\mathcal{S}}$.

$$\widehat{\mathcal{S}} = \underset{\mathcal{S} \subset \{1,\ldots,n\}, |\mathcal{S}|=k}{\operatorname{argmax}} \sum_{j \in \mathcal{S} = \operatorname{supp}(\mathbf{x})} |\mathrm{c}_j| = \operatorname{top}_k\left(|\mathbf{c}|\right). \tag{9}$$

where the operator $\operatorname{top}_k(|\cdot|)$ extracts indices of the k largest-magnitude entries.

Finally, a *naïve* estimate of the sparse vector $\mathbf{x}$ is obtained by assigning signs to the entries corresponding to the estimated support.

$$\hat{\mathrm{x}}_j = \begin{cases} \operatorname{sign}(\mathrm{c}_j) & \text{if } j \in \widehat{\mathcal{S}}, \\ 0 & \text{Otherwise.} \end{cases} \tag{10}$$

At the message bit-level recovery stage, LUT-based demapping via the inverse function $\mathcal{L}^{-1}$ can be employed to reconstruct the original binary message. This is made possible by $\mathcal{L}$'s bijective nature, which guarantees a unique inverse mapping:

$$\mathcal{L}^{-1} : \mathcal{X} \to \mathcal{M} \quad | \quad \hat{\boldsymbol{m}} = \mathcal{L}^{-1}(\hat{\mathbf{x}}). \tag{11}$$

B. LLR-Based Attention Weighted Projection Decoding

The conventional support recovery based on naive estimation via baseline projection scores is sensitive to column correlations in $\mathbf{A}$, particularly in the presence of noise, challenging channel conditions, or partial misdetection of the true support (i.e., $\widehat{\mathcal{S}} \neq \mathcal{S}$). To address this limitation, we propose a refined decoding strategy using LLR-driven attention-weighted projections, where adaptive weights derived from per-element LLRs reflect the reliability of each received signal observation. Formally, the approach assigns different weights to the elements of the received vector based on a reliability assessment of each element. In other respects, the LLR computation must account for the actual constellation statistics (i.e., the symbol distribution) of the measurement vector $\mathbf{u}$. Let $u_i = \mathbf{a}_i^\top \mathbf{x}$ be a discrete distribution with the set of possible values $u_i \in \left\{ \frac{2l-k}{\sqrt{m}} \;\middle|\; l = 0, 1, \ldots, k \right\}$ with probability $\mathbb{P}\left(u_i = \frac{2l-k}{\sqrt{m}}\right) = \binom{k}{l}\frac{1}{2^k}$, for all $k \ll m$. Hence, the actual range is $u_i \in [-\boldsymbol{s}, 0, +\boldsymbol{s}]$ where $\boldsymbol{s}$ represents the typical signal levels. Indeed, $\max_{s>0}(\boldsymbol{s}) = \frac{+k}{\sqrt{m}}$ and $\min_{s<0}(\boldsymbol{s}) = \frac{-k}{\sqrt{m}}$.

Formally, we adopt a generalized ternary hypothesis: (1) *Null hypothesis* ($\boldsymbol{\mathcal{H}}_0$): $u_i = 0$, (2) *Positive sign hypothesis* ($\boldsymbol{\mathcal{H}}_+$): $u_i = +s$, and (3) *Negative sign hypothesis* ($\boldsymbol{\mathcal{H}}_-$): $u_i = -s$, with probabilities $\mathbb{P}(u_i = \pm s) = \frac{1}{2}(1 - \mathbb{P}(u_i = 0))$, $\mathbb{P}(u_i = 0) = \binom{k}{k/2}\frac{1}{2^k}$, for all even k.

Under the coherent Gaussian model described in equation (5), the complex Gaussian probability density function can be expressed as $p(\mathrm{y}_i \mid u_i, \mathrm{h}_i) := \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\mathrm{y}_i - \mathrm{h}_i u_i|^2}{2\sigma^2}\right)$.

However, note that the LLR $\{\Lambda_i\}_{i=1}^m$ we are seeking is for the sign, even though in reality u_i can be equal to zero. We will therefore consider an LLR that compares the hypotheses $\boldsymbol{\mathcal{H}}_+ : u_i > 0$ and $\boldsymbol{\mathcal{H}}_- : u_i < 0$, ignoring $\boldsymbol{\mathcal{H}}_0 : u_i = 0 \Rightarrow \mathrm{y}_i = \mathrm{z}_i$ since in this case the LLR reduces to pure noise projected onto the direction of h_i. We rather want a metric that reflects the reliability of the sign estimation. In the context of weighting, we want a function that is positive if the estimated sign is positive, negative if negative, and whose absolute value reflects the reliability. The LLR is then obtained by applying Bayes' rule a priori, conditioned on h_i:

$$p(u_i = \pm s \mid \mathrm{y}_i, \mathrm{h}_i) = \frac{p(\mathrm{y}_i \mid u_i = \pm s, \mathrm{h}_i)\, p(u_i = \pm s)}{p(\mathrm{y}_i \mid \mathrm{h}_i)}. \tag{12}$$

Taking the ratio and logarithm, we obtain:

$$\Lambda_i = \log \frac{p(\mathrm{y}_i \mid u_i = +s, \mathrm{h}_i)\, p(u_i = +s)}{p(\mathrm{y}_i \mid u_i = -s, \mathrm{h}_i)\, p(u_i = -s)}. \tag{13}$$

Assuming equal priors, $p(u_i = +s) = p(u_i = -s)$, they cancel out, yielding:

$$\Lambda_i = \log \frac{p(\mathrm{y}_i \mid u_i = +s, \mathrm{h}_i)}{p(\mathrm{y}_i \mid u_i = -s, \mathrm{h}_i)}. \tag{14}$$

Substituting the probability density function into (14) and expanding the quadratic terms gives the simplified LLR expression for the i-th measurement:

$$\begin{aligned} \Lambda_i &= \log\left[\exp\left(-\frac{|\mathrm{y}_i - \mathrm{h}_i(+s)|^2}{2\sigma^2}\right)\right] - \log\left[\exp\left(-\frac{|\mathrm{y}_i - \mathrm{h}_i(-s)|^2}{2\sigma^2}\right)\right] \\ &= \frac{-|\mathrm{y}_i - \mathrm{h}_i(+s)|^2 + |\mathrm{y}_i - \mathrm{h}_i(-s)|^2}{2\sigma^2} = \frac{2 \cdot s \cdot \Re(\mathrm{h}_i^* \mathrm{y}_i)}{\sigma^2}. \end{aligned} \tag{15}$$

Notwithstanding, the LLR expression (15) assumes that the amplitude of u_i is exactly equal to s (i.e., the maximum amplitude). In practice, however, the amplitude may vary, particularly when $k > 2$. A more accurate approach is therefore to use the *conditional mean amplitude*, i.e., the *conditional expectation*, which accounts for the statistical distribution of the transmitted signal and better captures its average magnitude.

Proposition 1. *(LLR Metrics for Sparse-Aware Signaling.)* In sparse-aware signaling under the typical sparsity regime $k \ll m$, the LLR reflecting the reliability of the sign estimates for the support of the sparse vector $\mathbf{x}$ is computed by considering the conditional expectation of the magnitude of the transmitted signal observations, namely the measurement entries u_i, as follows:

$$\Lambda_i \approx \tfrac{2\,\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2} \cdot \mathbb{E}\big[|\mathbf{a}_i^\top \mathbf{x}| \;\big|\; \mathbf{x} \neq 0\big] = \tfrac{2\,\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2} \cdot \mathbb{E}\big[|u_i| \;\big|\; \mathbf{x} \neq 0\big], \tag{16}$$

where explicitly $\mathbb{E}[|u_i| \mid \mathbf{x} \neq 0] = \frac{k}{2^{k-1}\sqrt{m}}\binom{k-1}{\lfloor (k-1)/2 \rfloor}$.

Proof. See Appendix Section A[1] ∎

For a more compact notation of (16), we define the *sufficient statistic* $T_i := \frac{2\,\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2}$ and the *sparsity-dependent constant* $C_{k,m} := \frac{k}{2^{k-1}\sqrt{m}}\binom{k-1}{\lfloor (k-1)/2 \rfloor}$.

The LLR can then be equivalently expressed as

$$\Lambda_i = C_{k,m}\, T_i, \quad i = 1, 2, \ldots, m, \quad \text{with } C_{k,m} > 0. \tag{17}$$

[1] https://github.com/modiisii/IEEE-ICC2026/blob/main/LLR_Metric_Derivation.pdf

Although the weights could be defined directly from the LLR magnitudes via $\beta_i = |\Lambda_i| = C_{k,m} \cdot |T_i|$, such a straightforward formulation provides no guarantees on the range of β_i. It inherits the potentially unbounded scale of the LLRs and may suffer from numerical overflow ; specifically, $|T_i|$ can grow arbitrarily large under high signal-to-noise ratio (SNR) conditions. Hence, to obtain well-behaved weights, we employ a weighting function $f(|\mathbf{\Lambda}|)$ that maps the LLR magnitudes to weights bounded between 0 and 1, where $f(\cdot)$ denotes a suitable bounded and monotonic transformation. More importantly, the magnitude of an LLR directly reflects its reliability: larger magnitudes indicate higher decision confidence. This intrinsic reliability measure naturally justifies using LLR magnitudes to form attention weights.

Thereafter, we define the attention weights $\boldsymbol{\beta}$ from the magnitudes of the LLRs by applying a max-normalization function $f : \mathbb{R}_+ \longrightarrow [0, 1], \quad |\Lambda_i| \longmapsto f(|\Lambda_i|)$.

Formally, the weights β_i are positive and max-normalized based on the LLR magnitudes, expressed as

$$\beta_i \coloneqq f\big(|\Lambda_i|\big) = \frac{|\Lambda_i|}{\max\limits_{1\leqslant \ell \leqslant m} |\Lambda_\ell| + \varepsilon} = \frac{C_{k,m}|T_i|}{\max\limits_{1\leqslant \ell \leqslant m} C_{k,m}|T_\ell| + \varepsilon}, \tag{18}$$

where $\varepsilon > 0$ is a small regularization constant. If strictly positive weights are required, we assume $|\Lambda_i| > 0$ for all i ; hence $\beta_i \in (0, 1]$. Subsequently, the diagonal weighting matrix is constructed as

$$\mathbf{W} = \mathrm{diag}(\beta_1, \ldots, \beta_m) = \mathrm{diag}(\boldsymbol{\beta}), \tag{19}$$

which reweights the equalized received vector $\tilde{\mathbf{y}}$ prior to correlation with the transpose of the projection matrix.

Proposition 2. *(LLR-Based Attention-Weighted Projection.)*

The decoding stage follows the procedure described in the previous subsection. In particular, the decoder evaluates the weighted correlation vector defined as

$$\mathbf{c}^{(\mathsf{w})} = \mathbf{A}^\top (\mathbf{W}, \tilde{\mathbf{y}}) \;\text{ where }\; \mathrm{c}_j^{(\mathsf{w})} = \mathbf{a}_j^\top (\boldsymbol{\beta} \odot \tilde{\mathbf{y}}) = \langle \mathbf{a}_j, \boldsymbol{\beta} \odot \tilde{\mathbf{y}} \rangle. \tag{20}$$

The support of $\mathbf{x}$ is then estimated by selecting the indices corresponding to the k largest-magnitude entries of $\mathbf{c}^{(\mathsf{w})}$. Figure 2 provides a conceptual schematic of the SVC receiver employing LLR-based attention-weighted projection decoding, while Algorithm 1 summarizes the associated decoding procedure.

Algorithm 1: LLR-Based Attention-Weighted Decoding.

Require: Projection matrix $\mathbf{A}$, received vector $\mathbf{y}$, channel diagonal matrix $\mathrm{diag}(\mathbf{h})$, sparsity level k, Noise variance σ^2 .

1: Equalized received vector: $\tilde{\mathrm{y}}_i = \Re\left(\frac{\mathrm{h}_i^* \mathrm{y}_i}{|\mathrm{h}_i|^2 + \varepsilon}\right)$.
2: Compute LLRs: $\Lambda_i = \frac{2\Re\{\mathrm{h}_i^* \mathrm{y}_i\}}{\sigma^2} \cdot \mathbb{E}[|\mathbf{a}_i^\top \mathbf{x}| \mid \mathbf{x} \neq 0]$.
3: Compute attention weights: $\beta_i = \frac{|\Lambda_i|}{\max_\ell |\Lambda_\ell| + \varepsilon}$.
4: Form diagonal weighting matrix $\mathbf{W} = \mathrm{diag}(\boldsymbol{\beta})$.
5: Compute weighted correlation: $\mathbf{c}^{(\mathsf{w})} = \mathbf{A}^\top (\mathbf{W}\tilde{\mathbf{y}})$.
6: Estimate support set: $\widehat{\mathcal{S}} = \mathrm{top}_k\left(|\mathbf{c}^{(\mathsf{w})}|\right)$.
7: Reconstruct sparse vector $\hat{\mathbf{x}}$ with: $\hat{\mathrm{x}}_j = \mathrm{sign}(\mathrm{c}_j^{(\mathsf{w})})$ for $j \in \widehat{\mathcal{S}}$, and $\hat{\mathrm{x}}_j = 0$ elsewhere.
8: **return** $\hat{\mathbf{x}}$.

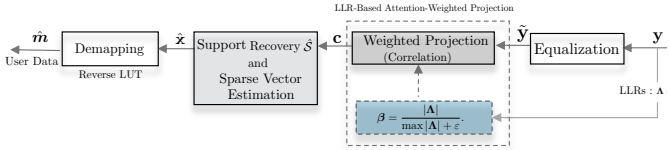


Fig. 2. Conceptual illustration of the SVC receiver employing LLR-based attention-weighted projection decoding.

C. SVC for 3GPP PUCCH Formats 2, 3, 4

In the 3GPP standard, uplink control information (UCI) packets ranging from 3 to 11 bits are conveyed on the PUCCH using Reed Muller encoding, followed by rate matching, scrambling, complex-valued QPSK modulation, in general, even if complex-valued $\pi/2$-BPSK modulation is allowed for PUCCH format 3 / 4. The resulting complex-modulated symbols are then mapped to subcarriers across multiple resource blocks using one or multiple OFDM symbols. For channel estimation, a pseudo random QPSK sequence is used as demodulation reference signals, enabling the base station to resolve channel ambiguities in time, frequency, or space prior to performing coherent detection. Furthermore, the principle of Reed Muller coded transmission for the 3GPP PUCCH is detailed in [12].

Nevertheless, it appears increasingly clear that 3GPP Reed–Muller codes, as currently designed to be decoded using well-known maximum likelihood decoding, are not optimal in certain respects, particularly from a computational complexity standpoint. Decoding via exhaustive ML search is effectively prohibitive, especially in the context of short-packet transmissions targeting URLLC use cases. Therefore, in what follows, we seek to address the following question: how can sparse-aware vector-coded transmission emerge as a viable alternative, potentially rivaling or even surpassing standard 3GPP RM-coded transmission in terms of both performance and decoding complexity?

Figure 3 presents a conceptual illustration of sparse aware vector coded transmission adapted to the 3GPP PUCCH structure, using PUCCH Format 2 as a representative instance, specifically in the resource mapping fashion.

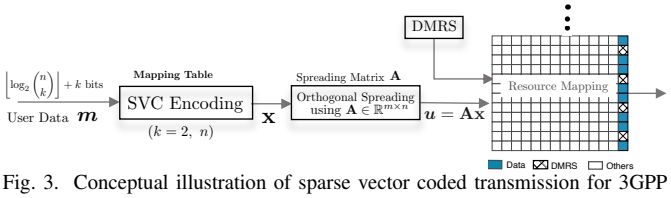


Fig. 3. Conceptual illustration of sparse vector coded transmission for 3GPP PUCCH (e.g., Format 2).

IV. Numerical Results

A. Proposed SVC vs Conventional Approaches

For illustrative purposes, we consider an SVC transmission with a sparse vector length $n = 64$, sparsity level $k = 2$, and a spreading matrix $\mathbf{A} \in \mathbb{R}^{32\times n}$. This configuration operates in the short blocklength regime and corresponds to an information message $\boldsymbol{m}$ carrying a payload of $B = 12$ bits, which lies at the lower end of the short-packet communication range, typically on the order of a few tens of bits. The performance comparison includes: (– –⊗– –) the proposed *enhanced projection decoding* based on an LLR-based attention-weighted approach, (—⊗—, · · ⊗ · ·) the *baseline projection decoding* employing either a Hadamard projection matrix or a random Gaussian matrix, and the optimal ML decoder (—), performing exhaustive search either at the sparse vector level $\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathcal{X}}{\mathrm{argmin}} \|\tilde{\mathbf{y}} - \mathbf{A}\mathbf{x}\|^2$, or at the message-bit level $\hat{\boldsymbol{m}} = \underset{\boldsymbol{m}\in\mathcal{M}}{\mathrm{argmax}} 2\tilde{\mathbf{y}}^\top \mathbf{A}\mathbf{x}(\boldsymbol{m}) - \|\mathbf{A}\mathbf{x}(\boldsymbol{m})\|^2$, thereby exhaustively evaluating all $|\mathcal{M}|=2^{\lfloor \log_2 \binom{n}{k} \rfloor + k}$ candidate messages. In addition, the performance evaluation in this subsection assumes perfect channel state information ; however, the analysis can be straightforwardly extended to scenarios wherein the channel state information is unknown. In such cases, a pilot-assisted transmission scheme can be employed, which will also be illustrated in the subsequent subsection

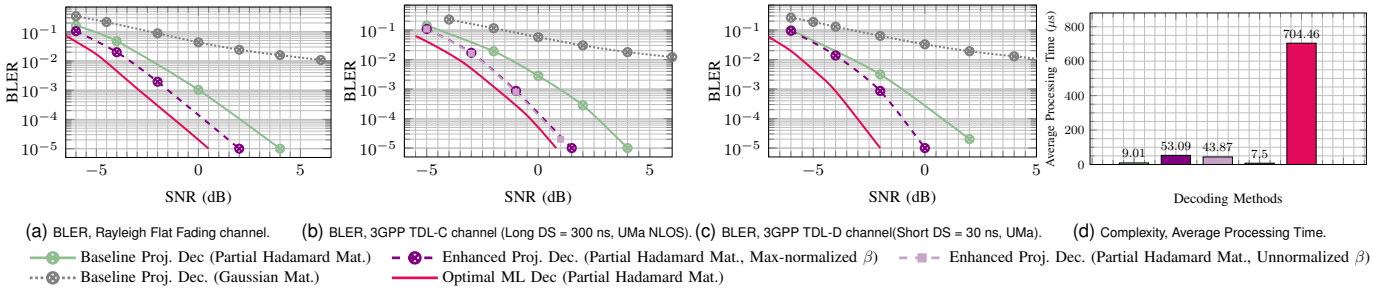


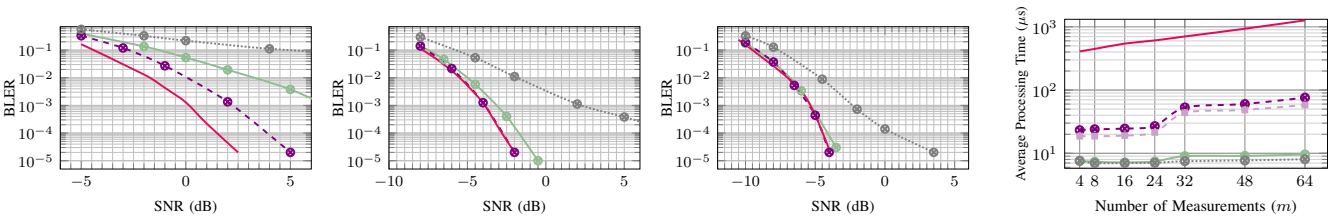
Fig. 4. Block error rate (BLER) and average processing complexity analysis. **SVC parameters:** sparse vector length $n = 64$, sparsity $k = 2$, spreading matrix $\mathbf{A} \in \mathbb{R}^{32\times 64}$, message $\boldsymbol{m}$ of 12 bits ($\boldsymbol{m}_1 = 10$ bits + $\boldsymbol{m}_2 = 2$ bits), coding rate $12/64$. **Decoders:** baseline partial Hadamard projection, baseline random Gaussian projection, and LLR-based attention-weighted Hadamard projection. **Channel configurations:** perfect CSI, 1×4 SIMO with (a) Rayleigh flat fading, (b) 3GPP TDL-C NLOS (long delay spread = 300 ns), and (c) 3GPP TDL-D LOS+NLOS (short delay spread = 30 ns); sampling rate $f_s = 30.72$ MHz.

for 3GPP PUCCH using DMRS-assisted transmission. Here, in the scenario of interest, SIMO diversity is exploited using multiple receive antennas in conjunction with the maximal ratio combining (MRC) algorithm, which performs optimal weighted combining based on the channel gains to enhance transmission reliability by mitigating fading and improving the SNR. Both Rayleigh flat fading channels (Figure 4a) and frequency-selective channels modeled using the 3GPP tapped delay line (TDL) models, specifically TDL-C (long delay spread of 300 ns) and TDL-D (short delay spread of 30 ns) for urban macro (UMa) scenarios [13] are considered, as illustrated in Figures 4b and 4c, respectively. TDL-C and TDL-D capture multipath propagation with distinct delay profiles and are tailored for non-MIMO channel assessments, thus enabling a realistic evaluation of the proposed decoding scheme under practical propagation conditions. In this multiple-receive-antenna scenario, the attention weights are computed by averaging the β values across the $\mathrm{N_R}$ receive antenna branches. It is noteworthy that the LLR-based attention-weighted decoder exhibits superior performance under challenging channel conditions characterized by severe fading, noise, or interference. For instance, in the TDL-C NLOS scenario, signal dispersion over multiple weak multipath components complicates detection at low SNR, as the receiver must reconstruct information from fragmented and noisy signals. At a very low BLER threshold of 0.001%, the proposed enhanced decoder—based on LLR-based attention-weighted Hadamard projection demonstrates an approximate 2 dB gain over the baseline Hadamard projection decoder. These gains and performance differences are consistently observable across all three channel types considered, as shown in Figure 4. Importantly, performance critically depends on the properties of the projection matrix, particularly its orthogonality and robustness to noise, fading, and interference. Notably, configurations employing random Gaussian projection decoding consistently exhibit the poorest performance, with an error floor around 1% BLER. In contrast, configurations utilizing a partial Hadamard matrix achieve significantly lower BLER, with the LLR-based attention-weighted projection decoding providing an additional performance improvement, nearing the optimal performance bound achieved by ML decoding.

Moreover, computational complexity, assessed via average processing time on a semi-logarithmic scale, is reported in Figure 4d. The proposed decoder is evaluated both with weights formed from max-normalized and unnormalized LLR magnitudes, and is compared against the baseline and optimal ML decoders. The enhanced decoder cycle comprises LLR computation, weight formulation, and projection calculation, simulated over 10^4 Monte Carlo trials at a given SNR. Relative to the baseline, the proposed decoder incurs approximately a $6\times$ increase in processing time ; however, it remains nearly $13\times$ less complex than full ML decoding. Notably, both the unnormalized and max-normalized LLR weighting configurations yield identical BLER performance. Nonetheless, the unnormalized variant is computationally more efficient, as the max-normalization step introduces an additional overhead of approximately 10 μs per decoding instance. Furthermore, Figure 5 illustrates the BLER performances and average processing time complexity as a function of the measurement dimension m. It is worth noting that the proposed receiver achieves higher gains over the baseline counterpart when the measurement dimension satisfies $m \lesssim n/2$. For larger dimensions, particularly as m approaches n (i.e., when a full Hadamard matrix is employed) and under perfect CSI conditions, the performance improvement relative to baseline projection becomes negligible. Specifically, when $m = n$ under perfect CSI, the performance of the proposed projection decoder and its baseline counterpart is identical, as the full orthogonality of the Hadamard projection matrix is already fully exploited in both cases. From a complexity perspective, the computational cost increases with the number of measurements for all benchmark decoders. Hence, for low-latency applications, a smaller measurement dimension m should be adopted, while the attention-weighting mechanism can compensate for the performance gap relative to the optimal ML performance bound.

B. Proposed SVC vs 3GPP RM Coding over PUCCH 2

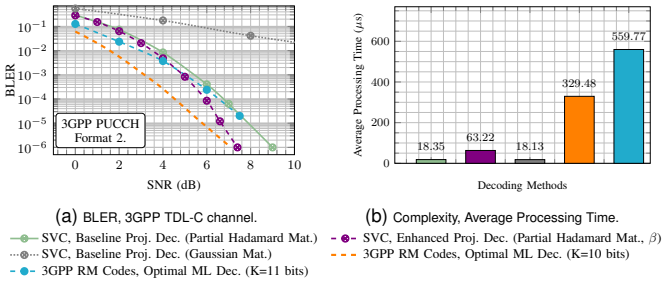
As described above, for illustrative purposes, we consider a sparse vector-coded transmission on a 3GPP PUCCH Format 2 with $n = 32$, $k = 2$, and measurement dimension $m = 32$. The information message $\boldsymbol{m}$ has a payload of $K = 10$ bits, which falls within the 3–11 bit short-packet range specified by the 3GPP standard for RM-coded transmissions. A payload of $K = 11$ bits is also considered for additional performance analysis. Accordingly, we consider a scenario with unknown channel conditions, and thus a DMRS-assisted transmission configuration is deployed. A single OFDM symbol spans four physical resource blocks (PRBs), in which 32 resource elements (i.e., m elements of the measurement vector after orthogonal spreading) are allocated to data REs, while the remaining 16 REs are reserved for demodulation reference signals (DMRSs). Channel estimation is performed using the least squares method, followed by linear interpolation to track the channel at the data positions. Consistent with previous



(a) BLER, 3GPP TDL-C channel , m=24. (b) BLER, 3GPP TDL-C channel, m=48. (c) BLER, 3GPP TDL-C channel, m=64. (d) Complexity, Average Processing Time.

Fig. 5. Block Error Rate (BLER) and average processing complexity for SVC with $n = 64$, $k = 2$, varying measurement dimension m, and 12-bit messages. Perfect CSI, 1×4 SIMO TDL-C NLOS channel (long delay spread = 300 ns, $f_s = 30.72$ MHz).

experiments, we adopt the 3GPP TDL-C NLOS channel model with a long delay spread of 300 ns. The performance evaluation considers three schemes: the proposed *SVC with enhanced projection decoding*, the *SVC with baseline projection decoding* employing either a Hadamard projection matrix or a random Gaussian matrix, and *3GPP RM codes with optimal ML decoding* (– – –) serving as a benchmark. The optimal ML decoder is implemented via exhaustive search. For a fair comparison, the same number of message bits is maintained across both coding schemes, i.e., K bits (3GPP RM) $\equiv B = \left\lfloor \log_2 \binom{n}{k} \right\rfloor + k$ bits (SVC). As reported in the previous



(a) BLER, 3GPP TDL-C channel. (b) Complexity, Average Processing Time.

Fig. 6. Block Error Rate comparison for 3GPP PUCCH Format 2 transmission. **PUCCH Configuration:** 1 OFDM symbol, 4 PRBs, 32 data REs, 16 DMRS REs. **SVC Parameters:** sparse vector length $n = 32$, sparsity $k = 2$, message $\boldsymbol{m}$ of $B = 10$ bits, coding rate $10/32$. **3GPP Reed–Muller:** $\mathscr{C}(32, 10)$ code, message of $K = 10$ bits, coding rate $10/32$, QPSK modulation. **Decoders:** baseline Hadamard projection vs. LLR-based attention-weighted Hadamard projection vs. ML decoding for RM codes. **Channel configurations:** unknown CSI, LS estimation with linear interpolation, 1×4 SIMO TDL-C NLOS channel (long delay spread = 300 ns, $f_s = 30.72$ MHz).

subsection, the conventional approach using *SVC with a Gaussian random spreading matrix and baseline projection decoding* remains less performant than all other configurations considered in Figure 6. In contrast, SVC with a Hadamard spreading matrix combined with LLR-based attention-weighted projection decoding achieves performance comparable to 3GPP RM-coded transmission under optimal ML decoding at very low BLER targets. For example, at a BLER of 0.0001%, both schemes demonstrate equivalent performance on PUCCH Format 2. As the BLER target becomes more stringent, the proposed SVC receiver is expected to outperform the conventional ML receiver for 3GPP RM codes. Furthermore, from a computational complexity perspective, as depicted in Figure 6b, the proposed decoder offers approximately a $5\times$ reduction in processing time compared to 3GPP RM under ML decoding, thereby contributing to the fulfillment of URLLC requirements.

V. Conclusions

In this work, we have presented an enhanced SVC scheme tailored for short-packet communication over noisy wireless channels. We have shown that conventional SVC approaches relying on random (Gaussian, Bernoulli) or structured (partial Hadamard) spreading/projection matrices with baseline projection decoding may suffer from reliability degradation under fading conditions. To overcome this limitation, we have introduced an enhanced decoding approach leveraging LLR-based attention-weighted projection, which dynamically reweights received signal measurements according to their reliability, giving more importance to the most reliable measurements. Through simulations over both Rayleigh flat fading and 3GPP TDL-C/TDL-D frequency-selective channels, we have demonstrated that the proposed LLR-based attention-weighted projection decoding outperforms baseline projection decoding with both structured and random projection matrices. Furthermore, we have extended the analysis to a short-packet (3–11 bits) transmission framework over the 3GPP PUCCH. The results indicate that the proposed SVC scheme, when combined with a partial Hadamard spreading matrix and LLR-based attention-weighted projection decoding, can competitively rival 3GPP RM codes under optimal ML decoding at very low BLER targets, while exhibiting significant computational complexity advantages.

References

[1] B. Lee, S. Park, D. J. Love, H. Ji, and B. Shim, "Packet structure and receiver design for low latency wireless communications with ultra-short packets," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 796–807, Feb. 2018.

[2] M. Sy and R. Knopp, "Enhanced low-complexity receiver design for short block transmission systems," in *Proc. IEEE 34th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Toronto, ON, Canada, Sep. 2023.

[3] N. Doan, "Low-complexity decoding of short linear block codes with machine learning," Ph.D. dissertation, McGill Univ., Montréal, QC, Canada, May 2022.

[4] C. Yue, V. Miloslavskaya, M. Shirvanimoghaddam, B. Vucetic, and Y. Li, "Efficient decoders for short block length codes in 6G URLLC," *IEEE Commun. Mag.*, vol. 61, no. 4, pp. 84–90, Apr. 2023.

[5] H. Ji, S. Park, and B. Shim, "SVC for ultra reliable and low latency communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6693–6706, Oct. 2018.

[6] W. Kim, S. K. Bandari, and B. Shim, "Enhanced sparse vector coding for ultra-reliable and low latency communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5698–5702, May 2020.

[7] X. Zhang, D. Zhang, B. Shim, G. Han, D. Zhang, and T. Sato, "Sparse superimposed coding for short-packet URLLC," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5275–5289, Apr. 2022.

[8] L. Yang and P. Fan, "Improved sparse vector code based on optimized spreading matrix for short-packet in URLLC," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 728–732, Apr. 2023.

[9] Y. Zhang *et al.*, "Sparse superimposed vector transmission for short-packet high-mobility communication," *IEEE Wireless Commun. Lett.*, vol. 12, no. 11, pp. 1961–1965, Nov. 2023.

[10] Y. Zhang, X. Zhu, Y. Liu, H. Liang, Y. L. Guan, and V. K. N. Lau, "Block sparse vector coding based ultra-reliable and low-latency short-packet transmission," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Cape Town, South Africa, 2024.

[11] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 1993.

[12] 3GPP TS 38.212 v16.2.0, "Multiplexing and channel coding (Release 16)," Jul. 2020.

[13] 3GPP TR 38.901 v16.1.0, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)," Nov. 2020.