Sparse Vector Coding for 5G Ultra-Reliable and Low Latency Communications

Hyoungju Ji, Member, IEEE, Sunho Park, Member, IEEE, and Byonghyo Shim, Senior Member, IEEE,
INMC, The Department of Electrical and Computer Engineering, Seoul National University, Korea.
Email:{hyoungjuji, shpark, bsim}@islab.snu.ac.kr

Abstract—Ultra reliable and low latency communication (URLLC) is a newly introduced service category in 5G to support delay-sensitive applications. In order to support this new service category, 3rd Generation Partnership Project (3GPP) sets an aggressive requirement that a packet should be delivered with $10^{-5}$ block error rate within 1 ms transmission period. Since the current wireless standard designed to maximize the coding gain by transmitting capacity achieving long code-block is not relevant for this purpose, entirely new transmission strategy is required. In this paper, we propose a new approach to transmit short packet information, called sparse vector coding (SVC). Key idea behind the proposed method is to transmit the control channel information after the sparse vector transformation. By mapping the transmit information into the position of nonzero elements and then transmitting it after the random spreading, we obtain underdetermined sparse system for which the principle of compressed sensing can be applied. From the numerical evaluations on realistic channel setting and decoder performance analysis, we demonstrate that the proposed SVC technique is very effective in URLLC transmission and outperforms the 4G LTE and LTE-Advanced physical downlink control channel (PDCCH) scheme.

I. INTRODUCTION

Ultra reliable and low latency communication (URLLC) is a newly introduced service category in 5G to support delay-sensitive applications such as the tactile internet, autonomous driving, factory automation, and remote controls. In order to support this new service category, 3rd Generation Partnership Project (3GPP) sets an aggressive requirement that a packet should be delivered with $10^{-5}$ block error rate (BLER) within 1 ms transmission period [1]. Since the current wireless transmission principle designed to maximize the coding gain by transmitting capacity achieving long code-block is not relevant to achieve this relentless goal, entirely new transmission strategy to meet the demanding requirement of 5G is required. While there have been some efforts to improve the latency and reliability of data channel [1–3], not much work has been done for the control channel except for the consideration of new channel coding scheme (e.g., Polar code) and multiple access scheme [4, 5]. In many future wireless applications including machine-type communications the amount of information to be transmitted is very small so that the efficient transmission of control-type information is of great importance.

In the current 4G systems, encoding of the control channel is based on the channel coding [6]. Encoding at the base station is done by the rate $\frac{1}{3}$ convolution coding, and the decoding at the mobile terminal is done by the maximum likelihood decoding (MLD) such as Viterbi decoding. While this approach has shown to be effective in 4G systems, use of this scheme in URLLC scenario would be problematic, since there is a stringent limitation on the packet length (and thus the parity size) yet the required reliability (target BLER = $10^{-5}$) is much higher than the current LTE-A systems (target BLER = $10^{-2}$–$10^{-3}$) [4].

The purpose of this paper is to propose a new type of short packet transmission scheme for URLLC that does not rely on the conventional channel coding principle. Key idea behind the proposed technique, henceforth referred to as sparse vector coding (SVC), is to transmit the information after the sparse vector transformation. By mapping the information into a sparse vector and then transmitting it after the random spreading, we obtain an underdetermined sparse system for which the principle of compressed sensing can be applied [7]. It is now well-known that the theory of compressed sensing guarantees an accurate recovery of a sparse vector with a relatively small number of measurements when the system matrix (a.k.a. sensing matrix) is generated at random [8], which can be achieved in our case via the random spreading. In the proposed scheme, encoding is done by the simple one-to-one (injective) mapping and the decoding is performed by the sparse signal recovery. Therefore, the proposed scheme is very simple to implement and can be applied to wide variety of future wireless environments in which latency and reliability should be guaranteed. From the numerical evaluations on realistic channel setting and decoder performance analysis, we demonstrate that the proposed SVC technique is very effective in URLLC transmission and outperforms the 4G LTE and LTE-Advanced physical downlink control channel (PDCCH) scheme by a large margin.

II. CONTROL CHANNEL IN LTE AND LTE-ADVANCED

To make our exposition simple, we focus on the control channel transmission in this work. However, extension of SVC to the data channel is straightforward. In 4G LTE systems, PDCCH carries essential information for the mobile terminal when it tries to transmit or receive the data [6]. To be specific, PDCCH carries the scheduling information to decode the data
A. SVC Encoding and Transmission

The key idea of the proposed SVC technique, when compared to the conventional transmission mechanism, is to map the information into the position of the sparse vector \( \mathbf{s} \). Figuratively speaking, SVC encoding can be thought as marking a few dots to the empty table. As illustrated in Fig. 1, if we try to mark dots to two cells out of 10, then there would be \( \binom{10}{2} = 45 \) choices. In general, when we choose \( K \) out of \( N \) symbol positions, we can encode \( \log_2 \left( \binom{N}{K} \right) \) bits of information.

\[
\begin{array}{cccccccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & & & & & & & \\
\vdots & & & & & & & & & & & \\
1 & 1 & 1 & 1 & 1 & & & & & & & \\
\hline
\end{array}
\]

\[\downarrow b\text{-bit control information}(b=5)\]
\[\downarrow K\text{-sparse vector } \mathbf{s}(K=2)\]

After the sparse mapping, each nonzero element in \( \mathbf{s} \) is spread into \( m \) resources using the codeword (spreading sequence) in the spreading codebook \( \mathbf{C} \). While it is possible to allocate resources either in time, frequency axis, or hybrid of these, in this work we assume that they are allocated in the frequency axis (see Fig. 2(a)). This choice will not affect the system model but will minimize the transmission latency. As a result of this spreading process called the \textit{multi-code spreading}, the resource mapping matrix \( \mathbf{R} \) in (1) is replaced with the codebook matrix \( \mathbf{C} = [c_1 \ c_2 \ \cdots \ c_N] \) where \( c_i = [c_{i1} \ c_{i2} \ \cdots \ c_{im}]^T \) is the spreading sequence. For example, if the first and the third elements of \( \mathbf{s} \) are nonzero, then the transmit vector after multi-code spreading would be

\[
\mathbf{x} = \mathbf{C}\mathbf{s} = s_1\mathbf{c}_1 + s_3\mathbf{c}_3.
\]

Since the positions of nonzero elements are chosen at random, the codebook matrix \( \mathbf{C} \) should be designed such that the transmit vector \( \mathbf{x} \) contains enough information to recover the sparse vector \( \mathbf{s} \) irrespective of the selection of the nonzero positions.

It has been shown that if entries of the codebook matrix \( \mathbf{C} \) are generated at random, e.g., sampled from Gaussian or Bernoulli distribution, then an accurate recovery of the sparse vector is possible as long as \( m = \mathcal{O}(K \log N) \) [8]. Example of \( \mathbf{C} \) for \( m=5 \) and \( N=10 \), when elements of \( \mathbf{c}_i \) are chosen from the Bernoulli distribution, is given by

\[
\mathbf{C} = \frac{1}{\alpha} \begin{bmatrix}
1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & 1 \\
1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 \\
-1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 \\
\end{bmatrix},
\]

where \( \alpha = \sqrt{2(M-1)/3} \) is the normalization factor depending on the modulated symbols (M=2 for QPSK).

Since the ensuring reliability is the top priority in URLLC, QPSK modulation scheme might be used exclusively. In order to use the QPSK modulation in SVC, we set one of the nonzero entries in \( \mathbf{s} \) to 1 and the other to \( j \). Thus, the transmit vector \( \mathbf{x} \) can be expressed as

\[
\mathbf{x} = 1\mathbf{c}_1 + j\mathbf{c}_3.
\]

From (3), we can easily see that elements of the transmit vector \( \mathbf{x} \) is mapped to the QPSK symbol (i.e., \( x_i \in \{1+j, 1-j, -1+j, -1-j\} \)). The corresponding received signal \( \mathbf{y} \) is

\[
\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v} = \begin{bmatrix}
\mathbf{H}\mathbf{c}_1 \\
\mathbf{H}\mathbf{c}_3
\end{bmatrix} \begin{bmatrix}
s_1 \\
s_3
\end{bmatrix} + \mathbf{v}.
\]

In general, the received vector \( \mathbf{y} \) is given by

\[
\mathbf{y} = \mathbf{HC}\mathbf{s} + \mathbf{v}.
\]

It is worth mentioning that an accurate recovery of the sparse vector \( \mathbf{s} \) is unnecessary in SVC since the decoding of control information is achieved by the \textit{identification of nonzero positions}, not the actual values of this vector. The fact that
the decoding is done by the support identification\(^1\) greatly simplifies the decoding process and also reduces the chance of decoding failure. The overall structure of the proposed SVC is depicted in Fig. 2(b).

The benefits of SVC can be summarized as follows: First, the transmission power of the control channel is concentrated on the nonzero elements of an information vector. Thus, when compared to the conventional system in which the transmission power is uniformly distributed across all symbols, effective transmit power per symbol is higher. Second, the SVC decoding process achieved by the sparse recovery algorithm is performed by the support detection and demapping. In particular, since the sparsity \(K\) is small and also known to the receiver, one can decode the SVC packet using a simple sparse recovery algorithm such as orthogonal matching pursuit (OMP) \([7]\).

**B. SVC Decoding**

1) **Support Identification:** SVC decoding is done by the identification of the support and any sparse recovery algorithm can be used to this end. In this work, we employ the greedy sparse recovery algorithm in the decoding of the SVC packet \([7]\). Since \(s\) has only \(K\) nonzero elements, the received vector \(y = \mathbf{HC}s + v\) can be expressed as a linear combination of \(K\) columns in \(\mathbf{H} = \mathbf{HC}\) perturbed by the noise. In view of this, main task of the SVC decoding is to identify the columns in \(\mathbf{H}\) participating in the received vector. In each iteration, greedy sparse recovery algorithm identifies the column of \(\mathbf{H}\) one at a time using a greedy strategy. Specifically, let \(\mathbf{H}_n^{-1}\) be the submatrix of \(\mathbf{H}\) that only contains columns indexed by \(\Omega_n^{-1}\), then the index \(\omega_j\) chosen at the \(j\)-th iteration of the greedy algorithm is given by\(^3\)

\[ \omega_j = \arg \max_i |\mathbf{H}_n^T y_i|^2, \]

where \(y_i = y - \mathbf{H}_n^{-1} \hat{s}_n^{j-1}\) is the residual vector and \(\hat{s}_n^{j-1} = \mathbf{H}_n^+ y\) is the estimate of \(s\) at \((j-1)\)-th iteration.\(^4\)

A better way to improve the decoding performance might be to use the maximum likelihood (ML) detection. Recalling that the sparsity \(K\) is known to both transmitter and receiver, the ML detection problem for the system model in (6) is expressed as

\[ s^* = \arg \max_{\|s\|_0 = K} P_r(s | y, \mathbf{H}, \mathbf{C}), \]

where \(\|s\|_0\) is the \(\ell_0\)-norm of \(s\) counting the number of nonzero elements in \(s\). Since our goal is to find out the support of \(s\), we alternatively have

\[ \Omega_n^* = \arg \max_{|\Omega_n| = K} P_r(s | y, \mathbf{H}, \mathbf{C}). \]

To find out the ML solution, we need to enumerate all possible combinations of candidate supports with cardinality \(K\). Unfortunately, this exhaustive search would not be feasible for most practical scenarios. In this work, we alternatively use the multipath match pursuit (MMP) algorithm, a recently

\(^1\)Support is the set of nonzero elements. For example, if \(s = [0 0 1 0 0 1]\), the \(\Omega_n = \{3, 6\}\).

\(^2\)The correlation between two distinct columns of the random matrix decreases exponentially with the dimension of a column (see, e.g., \([9\), Theorem 1\]).

\(^3\)If \(\Omega = \{1, 3\}\), then \(\mathbf{H}_n = [\mathbf{H}_1, \mathbf{H}_3]\).

\(^4\)\(\mathbf{H}_n^+ = (\mathbf{H}_n^T \mathbf{H}_n)^{-1} \mathbf{H}_n^T\) is the pseudo-inverse of \(\mathbf{H}\).
The proposed near-ML detection algorithm based on the greedy tree search (see [10] for details).

One clear advantage of MMP, in the perspective of SVC decoding, is that it deteriorates the quality of incorrect candidate yet does not impose any estimation error to the correct one. This is because the quality of incorrect candidates gets worse due to the error propagation while there is no such effect for the correct one. In particular, since the nonzero values of an original sparse vector s are known to the receiver, no estimation error will be introduced in the correct candidate.

2) Identification of False Alarm: Overall, there are two kinds of false alarm events causing the decoding failure: 1) support detection when the basestation transmits information to the different user and 2) support detection when there is no transmission at the basestation. In order to prevent these events, we need to examine the residual magnitude in each iteration. Firstly, when the control channel for a different user is received, the codebook between two distinct users would be different from each other so that the magnitude of the correlation μij between two codewords chosen from two district codebooks would be small. In this case, clearly, one cannot expect a substantial reduction in the residual magnitude. Secondly, when there is no transmission, the received vector will measure the noise only (i.e., y = v) and thus some column in Φ, say Φl, will be added to the residual in each iteration r = r - Φl. Based on these observations, if ∥r∥2 > F−1(1 - Pth) (e.g., F−1(0.01) and F−1(0.99) is the inverse cumulative distribution function of Chi-square random variable, then we declare the hypothesis is not true (i.e., decoding is not successful) and discard the decoded packet.

The proposed MMP-based SVC decoding algorithm is summarized in Table I.

### IV. SVC PERFORMANCE ANALYSIS

In this section, we analyze the success probability of the proposed SVC technique. As mentioned, decoding of the SVC-encoded packet is successful when all support elements are chosen by the sparse recovery algorithm. Thus, we analyze the probability that the support is identified accurately. In our analysis, we assume that the greedy algorithm is used in the decoding process and analyze the lower bound of the success probability.

Let S1 be the success probability that the support element is chosen in the j-th iteration. Since K = 2 and thus the required number of iterations to decode the information vector is two, the probability that the SVC packet is successfully decoded can be expressed as

\[
P_{\text{suc}} = P(S_1) = P(S_1, S_2)
\]

Our main result in this section is as follows (see [11] for details).

**Theorem 1.** The probability that the SVC-encoded packet is decoded successfully satisfies

\[
P_{\text{suc}} \geq 1 - \exp \left(-m(1 - \mu^*)^2 \right) - \exp \left(-\frac{m}{\sigma^2} \right)N^2
\]

where m is the number of measurements (resources), N is the size of sparse vector, \( \sigma^2 \) is the noise variance, and \( \mu^* = \max_{i \neq j} |\mu_{ij}| \) is the maximum correlation between spreading sequences.

Noting that the block error rate is BLER = 1 - P_{suc}, the upper bound of BLER when m is sufficiently large is

\[
\text{BLER}_{\text{svc}} \leq 1 - \left(1 - \exp \left(-\frac{m(1 - \mu^*)^2}{\sigma^2} \right)N^2\right)
\]

In Fig. 3, we plot the BLER performance of SVC as a function of SNR. To judge the effectiveness of Theorem 1, we perform the empirical simulation for m = 42, N = 96, and \( \mu^* \approx 0.7 \) in this setup. When we apply this value to the upper bound in (11), we could observe that the obtained bound is tight across
the board. In order to prove Theorem 1, we first analyze the success probability \( P(S^1) \) for the first iteration.

**Lemma 1.** Consider the received signal \( y = \gamma \Phi s + v \) where \( \gamma = \frac{\text{SNR}}{\mathcal{N}^2} \) and \( \phi_i = [h_{11}c_{i1} h_{22}c_{i2} \cdots h_{mm}c_{im}]^T \). The probability that the support element is chosen in the first iteration satisfies

\[
P(S^1) \geq \left( 1 - \exp \left( -\frac{m(1 - \mu)^2}{\sigma_v^2} \right) - \exp \left( -\frac{m}{\sigma_v^2} \right) \right)^{N-1}.
\]

We now move to the success condition for the second iteration.

**Lemma 2.** The probability that the support element is chosen at the second iteration under the condition that the first iteration is successful satisfies

\[
P(S^2|S^1) \geq \left( 1 - \exp \left( -\frac{m(1 - \mu)^2}{\sigma_v^2} \right) - \exp \left( -\frac{m}{\sigma_v^2} \right) \right)^{N-2}.
\]

It is worth mentioning that the lower bounds of \( P(S^1) \) and \( P(S^2) \) are the same except for the exponent. We are now ready to prove the main theorem.

**Proof of Theorem 1:** By combining Lemma 1 and 2, we can obtain the lower bound of the success probability \( P_{\text{suc}} \) as

\[
P_{\text{suc}} = P(S^2|S^1) P(S^1)
\geq \left( 1 - \exp \left( -\frac{m(1 - \mu)^2}{\sigma_v^2} \right) - \exp \left( -\frac{m}{\sigma_v^2} \right) \right)^{N-2}(N-1)
\geq \left( 1 - \exp \left( -\frac{m(1 - \mu)^2}{\sigma_v^2} \right) - \exp \left( -\frac{m}{\sigma_v^2} \right) \right)^{N^2},
\]

which completes the proof.

We skip the detailed steps due to the lack of space.

**V. SIMULATIONS AND DISCUSSIONS**

In this section, we examine the performance of the proposed SVC technique. Our simulation setup is based on the OFDM system in the 3GPP LTE-Adv Rel.13. As a channel model, we use AWGN and realistic ITU channel models including extended typical urban (ETU) and extended pedestrian-A (EPA) channel model [6]. For comparison, we also investigate the performance of the conventional PDCCH of LTE-Advanced system, polar code-based PDCCH [12], and AWGN lower bound. In the conventional PDCCH method, the convolution code with rate \( \frac{1}{2} \), along with the 16-bit CRC is employed. Since the block size of the polar code is not flexible, we set the rate \( \frac{3}{4} \) to test similar conditions \( (b_i = 12 \text{ and } b_c = 16) \).

In the proposed SVC algorithm, we set the random binary spreading codebook with \( N = 96 \) and \( K = 2 \). To ensure the fair comparison, we use the same number of resources \( (b_i = 12 \text{ and } m = 42 \text{ with } L = 8 \text{ repetitions}) \) in the control packet transmission. As a performance measure, we use block error rate (BLER) of the code blocks.

In Fig. 4(a), we investigate the BLER performance of the proposed SVC method and competing schemes under AWGN channel condition. We observe that the proposed SVC technique outperforms the conventional PDCCH and polar
SVC, achieving more than 4 dB gain over the conventional PDCCH and about 1.1 dB gain over the polar code-based scheme at $10^{-5}$ BLER point. In realistic scenarios such as EPA and EVA channels in LTE-Advanced, we observe that the performance gain of proposed SVC scheme over competing schemes is maintained (see Fig. 4(b)).

In Fig. 5, we evaluate the BLER performance PDCCH and SVC as a function of SINR for various information bit size ($b = 12, 24, 48$, and $96$). The results demonstrate that the proposed SVC technique supports more control bits than the conventional PDCCH. For example, SVC can transmit twice more information than PDCCH in the low SNR region (for example, see $b = 12$ of PDCCH and $b = 24$ of SVC in Fig. 5). It is worth mentioning that the coding gain of the conventional PDCCH improves with the codeblock size so that the gap between the SVC and PDCCH diminished gradually as the number of information bits increases. Finally, we investigate the BLER performance of cell edge or small cell scenario where the received signal contains a considerable amount of interference from adjacent basestations. In this simulations, we set the power level of interference to half of the amount of interference from adjacent basestations. In this paper, we have proposed a novel shot packet transmission strategy for the URLLC scenario. The key idea of the proposed SVC transmission scheme is to transform the transmit information into a sparse vector in the transmitter and to exploit a sparse recovery algorithm as a decoder in the receiver. Metaphorically, SVC can be thought as a marking dots to the empty table. As long as the number of dots is small enough and the measurements contains enough information to find out the marked cell positions, accurate decoding of SVC packet can be guaranteed. We showed from the numerical evaluations that the proposed SVC scheme is very effective in 5G URLLC scenarios.

VI. CONCLUSION

In this paper, we have proposed a novel shot packet transmission strategy for the URLLC scenario. The key idea of the proposed SVC transmission scheme is to transform the transmit information into a sparse vector in the transmitter and to exploit a sparse recovery algorithm as a decoder in the receiver. Metaphorically, SVC can be thought as a marking dots to the empty table. As long as the number of dots is small enough and the measurements contains enough information to find out the marked cell positions, accurate decoding of SVC packet can be guaranteed. We showed from the numerical evaluations that the proposed SVC scheme is very effective in 5G URLLC scenarios.

REFERENCES

[1] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, E. Riedel, A. Puschmann, A. Mitschelethiel, M. Mller, T. Elste, and M. Windisch, “Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture,” *IEEE Commun. Mag.*, vol.55, no.2, pp.70-78, 2017.

[2] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, “Ultra Reliable and Low Latency Communications in 5G Downlink: Physical Layer Aspects,” *submitted to IEEE Wireless Commun.*

[3] O. Yilmaz, Y. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and Joachim Sachs, “Analysis Ultra-Reliable and Low-Latency 5G Communication for a Factory Automation Use Case,” In *Proc. IEEE Int. Conf. on Comm. (ICC) Workshop*, pp.1190 - 1195, 2015.

[4] 3GPP Technical Report 38.913, “Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14),” v14.2.0, 2017.

[5] N. A. Johansson, Y.Wang, E. Eriksson, and M. Hessler, “Radio Access for Ultra-Reliable and Low-Latency 5G Communications,” *In Proc. IEEE Int. Conf. on Comm. (ICC) Workshop*, pp.1185 - 1189, 2015.

[6] S. Sesia, M. Baker, and I. Toufik, *LTE-the UMTS Long Term Evolution: from Theory to Practice*, John Wiley & Sons., 2012.

[7] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, “Compressed Sensing for Wireless Communications: Useful Tips and Tricks,” *IEEE Commun. Survey and Tutorials.*, vol.19, pp.1527-1550, 2017.

[8] D. L. Donoho, “Compressed Sensing,” *IEEE Trans. on Inform. Theory*, vol.52, no.4, pp.1289-1306, 2006.

[9] T.T. Cai and J. Tiefeng, “Phase Transition in Limiting Distributions of Coherence of High-dimensional Random Matrices,” *Journal of Multivariate Analysis*, vol.107, pp.24-39, 2012.

[10] S. Kwon, J. Wang and B. Shim, “Multipath Matching Pursuit,” *IEEE Trans. on Inform. Theory*, vol.60, no.5, pp.2986-3001, May 2014.

[11] H. Ji, S. Park, and B. Shim, “Sparse Vector Coding of Control Channel for Ultra-Reliable and Low Latency Communications,” *submitted to IEEE Trans. Wireless Commun.*, 2017.

[12] H. Vangala, Y. Hong, and E. Viterbo, “Efficient Algorithms for Systematic Polar Encoding,” *IEEE Commun. Lett.*, vol.20, no.1, pp.17-20, Jan. 2016.