Channel Aware Sparse Signaling for Ultra-low Latency TDD Access

Wonjun Kim, Hyoungju Ji, and Byonghyo Shim
Institute of New Media and Communications and Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea
Email: {wjkim, hyoungjuji, bshim}@islab.snu.ac.kr

Abstract—Future mobile communication systems need to support wide variety of new services and applications. In order to support these, ITU-R introduced new types of use cases. One such use case, called ultra reliable and low latency communication (URLLC), concerns the reduction of latency down to a millisecond level while ensuring the reliability of the transmission. In case of uplink transmission, supporting the stringent latency requirement of URLLC is quite challenging due to a time-consuming and complicated handshaking process. In the time division duplexing (TDD) systems, satisfying the latency requirement is far more difficult since the mobile device cannot transmit the data when the subframe is directed to the downlink. In this paper, we propose a new grant signaling scheme, referred to as channel-aware sparse signaling (CASS), to achieve a low latency access in the TDD-based URLLC systems. Key idea of CASS is to map the grant information into a small number of subcarriers and then decode it using a small number of early received samples. From the numerical evaluations, we demonstrate that the proposed CASS scheme achieves significant reduction in access latency over the conventional LTE-TDD systems.

I. INTRODUCTION

Future mobile communication systems are expected to change our life by supporting wide variety of services and applications such as tactile internet, remote control, smart factories, and driverless vehicles, to name just a few [1]. In order to support diverse services and applications, new requirements other than the classical throughput enhancement are needed [2]. One such requirement is the reduction of latency down to a millisecond level and ensuring the reliability of the transmission. To cope with this new requirement and related services, ITU introduced new use case called ultra-reliable and low latency communications (URLLC) [3]. Since it is not possible to satisfy the stringent latency requirement by a small makeshift of current 4G LTE systems, an entirely new transmission scheme to support URLLC is required.

Recently, there have been some studies to achieve the latency reduction in the downlink transmission [4]–[6]. One simple approach is to transmit an urgent data without any reservations [4]. Also, an approach reserving resources in prior to the data scheduling has been proposed [5]. In the uplink direction, however, these approaches might not be applicable since the uplink transmission is subject to the complicated handshaking procedure with heavy signaling overhead. Note that the signaling process requires a complicated interplay between the basestation and mobile device, and thus it takes quite a bit of time for a mobile device to initiate the data transmission. Indeed, it has been reported that the signaling for LTE scheduling takes more than 7ms even for the best scenario [7].

In the time division duplexing (TDD) systems, which is expected to be a popular duplexing scheme in the future cellular systems [8], satisfying the latency requirement is far more difficult since the mobile device cannot transmit the data in the uplink when the subframe is directed to the downlink (DL). Thus, even though there is an urgent information to transmit, a mobile device has no way but to wait until the transmit direction is switched to the uplink (UL). For example, current 4G LTE TDD systems switch from DL to UL with half-frame-level (5ms) or frame-level (10ms) period so that the URLLC requirements cannot be satisfied with an ordinary processing. One can naturally infer from this observation that a direct way to reduce the physical layer latency is to shorten the switching period up to the subframe-level (1ms) period or less. Even in this case, it is not easy to support the short switching period in the current 4G LTE systems due to the time-consuming handshaking process.

An aim of this paper is to propose a low latency uplink access scheme suitable for TDD-based URLLC systems. A key feature of the proposed scheme is to transmit the latency sensitive information without waiting for the transmit direction change. To be specific, the basestation switches the transmit direction to UL right after sending the URLLC grant information and hence a mobile device having the latency sensitive information can access the UL resources right away. To support the fast uplink access, we introduce a new grant signaling scheme, referred to as channel-aware sparse signaling (CASS). Key idea of CASS is to map the URLLC grant information into a sparse vector in the OFDM symbol and then decode it using a small amount of time-domain received samples. In doing so, we make the frequency-domain OFDM symbol vector sparse (see Fig. 1). This together with the fact that the submatrix of the inverse discrete Fourier transform (IDFT) matrix is used as a sensing matrix allows us to use the compressed sensing (CS) technique in the decoding of the grant signal. It is now well-known from the theory of
CS that an accurate recovery of a sparse vector is guaranteed with a relatively small number of measurements as long as the sensing (measurement) process preserves the energy of an input sparse vector [9], [10]. In our context, this means that a mobile device can accurately decode the grant information with a small number of early arrived received samples (see Fig. 1), which in turn means that the UL access latency (latency of buffering, transmission, and processing of the grant signal) can be reduced significantly.

II. UPLINK ACCESS LATENCY IN TDD SYSTEMS

In this section, we briefly review the latency of TDD-based uplink transmission [7]. As mentioned, scheduling procedure is needed for the 4G LTE systems to initiate the UL data transmission\(^1\). As illustrated in Fig. 2, a mobile device sends a scheduling request (SR) signal to the basestation when there is an information to transmit. After receiving SR, the basestation allocates resources and then sends the resource grant (RG) signal to the mobile device. After receiving and decoding the RG signal, the mobile device begins to transmit the information to the basestation in the assigned timing (resources).

In the scheduling process, the uplink access latency \(T_{\text{up}}\), defined as the time duration from the transmission of the grant signal to the initiation of the data transmission, can be expressed as the sum of three distinct latency components (see Fig. 2):

\[
T_{\text{up}} = T_{\text{prop}} + T_{\text{proc}} + T_{\text{wait}}. \tag{1}
\]

- \(T_{\text{prop}}\), called the propagation latency, is the time for a grant signal to travel from the basestation and to the mobile device
- \(T_{\text{proc}}\) is the processing latency for the grant signal
- \(T_{\text{wait}}\) is the waiting latency for the transmit direction change

Among these latency components, we consider on the reduction of the major components \(T_{\text{proc}}\) and \(T_{\text{wait}}\). First, \(T_{\text{proc}}\) can be divided into two components: 1) the buffering latency \(T_{\text{buff}}\) (the time to receive the OFDM symbol containing the grant signal) and 2) the decoding latency \(T_{\text{dec}}\) (the time to decode the grant information). For example, it takes around 1ms to buffer and decode the grant signal in the current 4G LTE systems [7]. Clearly, this time would be too large to satisfy the URLLC latency requirement. \(T_{\text{wait}}\) is caused by the periodic direction change in the TDD systems (see Fig. 2). Since the current LTE TDD systems switch the transmit direction every 5ms or 10ms, a mobile device should wait until the direction is switched to UL to transmit the urgent data. Since this long switching period cannot satisfy the URLLC latency requirement, a dynamic TDD scheme supporting ultra short DL-to-UL switching period has been introduced [11].

When the switching period is short, one can notice that \(T_{\text{proc}}\) would be a bottleneck to support the fast UL access. In the conventional TDD systems, a mobile device has enough time to decode the grant signal since the switching period (e.g., 5ms in LTE TDD systems) is much larger than \(T_{\text{proc}}\). However, when the switching period is very short (e.g., 1ms subframe-level switching), conventional grant signaling mechanism requiring all the received samples (e.g., 1024 samples in one OFDM symbol) to decode the grant information would not be an acceptable option due to the large \(T_{\text{proc}}\) (e.g., 1ms in LTE systems). In the following section, we describe the CASS scheme that greatly reduces \(T_{\text{proc}}\) of the grant signal.

III. CHANNEL-AWARE SPARSE SIGNALING

A. System Description of CASS

Fig. 3 depicts the block diagram of the proposed CASS scheme. When designing the grant signal \(s\), the basestation picks a small number, say \(k\) out of \(N\), of subcarriers. For example, if the second and fifth subcarriers are chosen in the grant signal \(s\), then \(s = [0 \ s_1 \ 0 \ s_2 \ 0 \ \cdots \ 0]\) (\(s_1\) and \(s_2\) are the symbols) and thus the support of \(s\) is \(\Omega = \{2, 5\}\). In the CASS scheme, the granted user ID is encoded to the positions of the selected subcarriers\(^3\) and the remaining grant information (e.g., uplink timing and transmission band) is encoded into the symbols.

\(^3\)When the basestation picks \(k\) subcarriers out of \(N\), then there are \(\binom{N}{k}\) user IDs in total. In the above example, \(\Omega = \{2, 5\}\) is a user ID.
Fourier transform (IFFT), the time-domain sample vector $s$ we make the grant signal vector the basestation and mobile device in OFDM system.

Fig. 3. Block diagram of channel-aware sparse signaling (CASS) between the basestation and mobile device in OFDM system.

As mentioned, by using only small number of subcarriers, we make the grant signal vector $s$ sparse. After the inverse fast Fourier transform (IFFT), the time-domain sample vector $s_t = [s_1(1) \cdots s_1(N)]^T$ is transmitted through the fading channel. The relationship between the sparse grant signal $s$ and the received time-domain samples $y$ can be expressed as

$$\begin{align*}
y &= Hs_t + v \\
    &= HF^*s + v
\end{align*}$$

where $H \in \mathbb{C}^{N \times N}$ is the channel matrix, $F^* \in \mathbb{C}^{N \times N}$ is the IDFT matrix, and $v \sim \mathcal{CN}(0, \sigma_v^2)$ is the additive Gaussian noise vector. Since the channel matrix $H$ is the circulant matrix after removing the cyclic prefix, it can be eigen-decomposed by DFT matrix, i.e., $H = F^*\Lambda F$ where $\Lambda$ is the diagonal matrix whose diagonal entry $\lambda_{ii}$ is the frequency-domain channel response for the $i$-th subcarrier. Thus, we have

$$\begin{align*}
y &= (F^*\Lambda F)F^*s + v \\
    &= F^*\Lambda s + v \\
    &= F^*x + v
\end{align*}$$

where $x = \Lambda s$. It is worth mentioning that the supports of $s$ and $x$ are the same (i.e., nonzero positions of $s$ and $x$ are the same).

In the context of CS, $x$ is the sparse input vector and $F^*$ is the sensing matrix. Since $F^*$ preserves the signal energy of $x$, the sparse vector $x$ can be readily recovered with only a part of $y$ using the sparse recovery algorithm. This means that we only need a small number of early arrived samples in $y$ to decode the grant informations. The corresponding partial measurement vector $\hat{y} \in \mathbb{C}^{m \times 1}$ ($m \ll N$) constructed from early arrived samples can be expressed as

$$\begin{align*}
\hat{y} &= \Pi \hat{v} \\
      &= \Pi F^*x + \hat{v} \\
      &= \Lambda x + \hat{v}
\end{align*}$$

where $\Pi = [I_m \ 0_m \times (N-m)]$ is the matrix to select the first $m$ samples among $N$ time-domain samples, $\hat{v} = \Pi v$ is the modified noise vector, and $\Lambda = \Pi F^*$ is the partial IDFT matrix consisting of the first $m$ consecutive rows of $F^*$.

Algorithm 1 The proposed CASS encoding algorithm

\begin{itemize}
  \item [1.] $\omega^* = \arg \max_{\omega \in \mathbb{C}^m} \|H_{\omega}\|_2$
  \item [2.] $\Gamma = \{\gamma \in \mathbb{C}^m \mid f(\gamma, \omega^*) \approx 0\} \cup \{\omega^*\}$
  \item [3.] $\Omega^* = \arg \max_{|\Omega| = k, \Omega \subseteq \Gamma} \|H_{\Omega}\|_2$
\end{itemize}

Output: $\Omega^*$

As mentioned, the user ID and grant information are encoded into the position of subcarriers and symbols, respectively, so that the decoding process consists of two steps: 1) support identification to find out the nonzero positions of $s$ vector and 2) detection of symbols for nonzero positions of $s$. First, for the decoding of the granted user ID, a mobile device needs to identify the support of $x$, which is done by the sparse recovery algorithm [12]. After identifying the support $\Omega$, a mobile device decodes the remaining grant information by detecting the symbol vector $\hat{s}_\Omega$ corresponding to the nonzero elements. Note that, after removing the components associated with the non-support elements in (9), the system model is converted into the overdetermined system model ($m > k$). For example, if $\Omega = \{2, 5\}$, then the system model in (9) is simplified to $\hat{y} = [a_2 \ a_5] \begin{bmatrix} x_2 \\ x_5 \end{bmatrix} + \hat{v}$. Thus, the conventional technique such as the linear minimum mean square error (LMMSE) estimator followed by the symbol slicer can be used for this task.

The benefits of CASS can be summarized as follows. First and foremost, support identification for the decoding of the grant signal $s$ is done with the small number of time-domain samples. When compared to the conventional signaling mechanism in which all received samples are needed to decode the grant information, buffering latency $T_{buf}$ can be reduced by the factor of $m/N$. For example, if $m = 128$ and $N = 1024$, then $T_{buf}$ would be reduced by the factor of 1/8. Second, since the channel information is unnecessary in the support identification process, an error caused by the channel estimation error can be prevented. This is because the sensing matrix $A$ in (9) is constructed only by the submatrix of IDFT matrix and what we need to do is to find out the nonzero positions of $x = \Lambda s$, not the actual values. Third, the implementation cost and the computational complexity of CASS is very low. In particular, since the sparsity $k$ is small and also known to the mobile device, one can decode the grant information using a simple sparse recovery algorithm such as orthogonal matching pursuit (OMP) [13]. We discuss the next subsections that by choosing nonzero positions deliberately, support identification can be finished in just two iterations.

B. Encoding Operation in CASS

Since the decoding of the grant signal is done by the support identification, accurate identification of the support is of great importance for the success of CASS. In general, when the system matrix is generated at random, the support identification performance would not be affected by the choice
of support. In the CASS scheme, however, the system matrix is constructed from IDFT matrix and the sparse vector $x = \Lambda s$ is the product of the frequency-domain channel $\Lambda$ and the sparse grant signal $s$ so that both system matrix and channel state affect the decoding performance.

First, the support identification performance depends heavily on the channel state. For example, if a selected subcarrier $s_i$ undergoes a deep fading in the frequency-selective channel (i.e., $\lambda_{ii} \approx 0$), then an accurate identification of the nonzero position $x_i = \lambda_{ii} s_i$ would not be possible. Since the DL channel information can be derived from the UL channel estimation via the channel reciprocity in TDD systems [14], it would be desirable to choose indices of subcarriers having the highest subchannel gains as support elements (i.e., $\Omega = \arg \max_{|\Omega| = k} ||\hat{h}_\Omega||_2$). In doing so, one can reduce the chance of the decoding failure significantly.

Second, the support identification performance depends also on the correlation between columns in the system matrix $A$. In most of greedy sparse recovery algorithms, such as OMP, an index of a column in $A$ maximally correlated to the partial measurement $\tilde{y}$ is chosen as an estimate of the support element. Therefore, if two columns of $A$ are strongly correlated and only one of these is associated with the nonzero values in $x$, then it might not be easy to distinguish the right column (column associated with the nonzero value) from wrong one in the presence of noise. Fortunately, since all entries of $A = \Pi F^*$ are known in advance, we can alleviate this event by considering the column correlation of $A$ in the support selection. Specifically, let $f(\omega_p, \omega_q)$ be the correlation between $\omega_p$ and $\omega_q$-th columns in $A$, then we have

$$f(\omega_p, \omega_q) = \frac{1}{m} \sum_{l=1}^{m} e^{-j2\pi(\omega_p-\omega_q)(l-1)/N} e^{j2\pi(\omega_q-\omega_p)(l-1)/N}$$

$$f(\omega_p, \omega_q) = \frac{1}{m} \sum_{l=1}^{m} \frac{\sin(\omega_p-\omega_q)l}{N} \sin(\pi(\omega_p-\omega_q)l/N)$$

Since $f(\omega_p, \omega_q)$ depends only on the absolute difference between $\omega_p$ and $\omega_q$, we will henceforth denote it as $f(|\omega_p - \omega_q|)$. One can easily see that columns $a_{\omega_p}$ and $a_{\omega_q}$ are (near) orthogonal (i.e., $f(|\omega_p - \omega_q|) \approx 0$) if $|\omega_p - \omega_q| \approx cN$ for a properly chosen integer $c$. Thus, by choosing the subcarrier indices from the set of the orthogonal columns in $A$, accuracy of the support identification can be improved significantly.

In summary, the support selection rule considering the channel state and system matrix is given by

$$\Omega = \arg \max_{|\Omega| = k} ||\hat{h}_\Omega||_2$$

(11)

where $\Gamma$ is the index set of the orthogonal columns (see Algorithm 1). Overall grant procedure is as follows. First, each and every mobile device finds its own support $\Omega$ (user ID) using (11). Exploiting the channel reciprocity, basestation can also figure out the user IDs of all mobile devices using (11). Second, after receiving SR, the basestation transmits the CASS-based grant signal to the desired mobile device. Using the small number of early arrived received samples, the mobile device decodes the grant signal. If the decoded support $\hat{\Omega}$ is equivalent to its own support $\Omega$ (i.e., $\hat{\Omega} = \Omega$), the grant signal is decoded successfully and thus the mobile device sends the (latency sensitive) information immediately (see Fig. 4).

C. Decoding Process in CASS

1) Basic Decoding: Key operation of the CASS decoding is to find out the support $\Omega$. In other words, main task of decoding is to find $k$ non-zero positions of $x$ vector from the received vector $\tilde{y} = Ax + \nu$. By exploiting the orthogonality of the columns associated with non-zero positions of $x$, we can greatly simplify the support identification process.

To be specific, in the first iteration, a column maximally correlated with $\tilde{y}$ is chosen as an estimate of the support element $\hat{\omega}_1$. Since the columns associated with the support $\Omega$ are chosen from the set of orthogonal columns, remaining columns should be orthogonal to the column chosen in the first iteration. In the second iteration, therefore, we choose $k - 1$ best columns among those orthogonal to the firstly chosen column. Thus, when compared to the conventional greedy sparse algorithm in which $k$ iterations are required, the proposed CASS decoding is finished in only two iterations.

Algorithm 2 The proposed CASS decoding algorithm

Input: $\tilde{y} \in \mathbb{C}^m$, $A \in \mathbb{C}^{m \times N}$, $k \in \mathbb{N}$, $\tau \in \mathbb{N}$, $h \in \mathbb{C}^N$

1: $\hat{\omega}_1 = \arg \max ||a_{\omega} \tilde{y}||_2$

2: $\Gamma = \{\gamma \in \Omega | f(\gamma, \hat{\omega}_1) \approx 0\}$

3: (Identification) Select indices $\{\hat{\omega}_i\}_{i=2,...,k}$ corresponding to $k - 1$ largest entries in $A_{\hat{\omega}_1}^T \tilde{y}$

4: $\hat{\Omega} = \{\hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_k\}$

5: ($\tau$-close support identification) Check $|\hat{\omega}_i - \hat{\omega}_i| < \tau$ for $i \in \{1, \cdots, k\}$

6: if $\tau$-close support identification is successful then

7: (Estimation of $s_{\hat{\Omega}}$) $\hat{s}_{\Omega} = \arg \max_{|\Omega| = k} ||\tilde{y} - A_{\hat{\Omega}} \hat{\Lambda}_{\hat{\Omega}} u||_2$

8: (Symbol slicing) $\hat{s}_{\Omega} = Q(\hat{s}_{\Omega})$

9: end if

Output: $\hat{\Omega}, \hat{s}_{\Omega}$
After this, a mobile device checks if it is granted by comparing the decoded support $\hat{\Omega}$ and $\Omega$. If $\hat{\Omega} = \Omega$, remaining grant information is obtained by decoding the symbols associated with the support position.

2) $\tau$-close Support Identification: Since the correlation between the adjacent columns in $A = \mathbf{F}^\dagger$ is large (see (10)), a column adjacent to the correct one might be chosen by mistake. To avoid this type of mistake, we propose an improved scheme relaxing the success condition in the support identification. Basic idea of the proposed strategy, called $\tau$-close support identification, is to regard the selected index as correct one if the selected position is close to the true one. That is, a chosen index $\hat{\omega}_i$ is considered as the correct one if it is not too far away from the true index $\omega_i \in \Omega$, i.e., $\hat{\omega}_i \in \{\omega_i - \tau + 1, \ldots, \omega_i, \ldots, \omega_i + \tau - 1\}$. Since $\mathbf{x}$ is the sparse vector and hence the number of nonzero elements is small, as long as the difference between $\hat{\omega}_i$ and $\omega_i$ is small, there would not be any confusion caused by the $\tau$-close support identification. In Algorithm 2, we summarize a refined CASS decoding algorithm incorporating the $\tau$-close support identification (see [15] for details.).

IV. SIMULATION RESULTS

In this section, we present the numerical results to evaluate the decoding performance and access latency of the proposed CASS. In our simulations, we consider the OFDM-based TDD systems with $N = 1024$ subcarriers. As a channel model, we use the i.i.d Rayleigh fading channels. For comparison, we use two different approaches in the support selection. In the first approach, we choose the subcarriers uniformly at random among $N$ subcarriers. In the second approach, we choose the support by the proposed selection rule (Algorithm 1). In the decoding process, we use the proposed decoding algorithm (Algorithm 2) with $\tau$-close support identification ($\tau = 2$). As performance metrics, we use the success probability of support identification, block error rate (BLER) and average access latency. The access latency is defined as the sum of the waiting latency $T_{\text{wait}}$ and processing time $T_{\text{proc}}$ in (1).

In Fig. 5, we evaluate the success probability of the support identification as a function of $m$ for various SNRs (SNR = −3dB, 0dB, and 5dB). Since the user decoding of CASS is done by the support identification, this probability is equivalent to the user decoding success probability. We observe that the proposed CASS scheme achieves a significant reduction in the number of received samples. When compared to the conventional signaling mechanism in which all received samples are needed to decode the grant information, CASS requires much smaller number of samples. For example, CASS requires only 7.8% ($m = 80$ at 5 dB) of the received samples, which directly implies that the buffering latency $T_{\text{buff}}$ can be reduced by the factor of 92.2% (see Section III.A).

In Fig. 6, we evaluate the success probability of the support identification for various sparsity levels ($k = 4, 8,$ and $12$). We observe that only 10% ($k = 4$) and 15% ($k = 12$) of the received samples are needed to decode the grant information. These results clearly demonstrate that the proposed support selection rule (in Sec III.B) is very effective in reducing the latency. This behavior, however, cannot be achieved in the random support selection approach. For instance, if $k$ increases from 4 to 12, required number of samples to achieve 40% success probability increases from 38 samples to 75 samples in the proposed support selection rule but that for the random support selection rule increases from 57 to 256.

In Fig. 7, we compare the BLER performance of the proposed CASS scheme and the control channel of 4G LTE (physical downlink control channel (PDCCH)) using the convolution code (1/2) rate and Viterbi decoding. In this simulation, we encode 72 bits of grant information using 16-QAM for the proposed CASS scheme (i.e., $k = 18$) and QPSK for the PDCCH. We observe that the proposed scheme outperforms the PDCCH by a large margin, achieving more than 6 dB gain at 10$^{-2}$ BLER point. This is because 1) the transmission power of the data symbol is higher than that of the PDCCH and further 2) the basestation selectively use the subcarriers with good channel condition selectively for the data transmission.

Finally, we evaluate the access latency of CASS-based TDD system in Table. I. In our simulation, we consider the LTE-TDD system (Rel. 13) as a reference. The access latency is defined as $T_{\text{up}} = T_{\text{wait}} + T_{\text{proc}} = T_{\text{wait}} + \left(\frac{m}{f_s} + T_{\text{dec}}\right)$ where $m$ is the number of received samples and $f_s$ is the sampling frequency. When carrying out the CASS-based access, the basestation changes the transmit direction into UL right after sending the grant signal and thus the mobile device can transmit the latency sensitive data without waiting for the periodic transmit direction change (i.e., $T_{\text{wait}} \approx 0$). We use two TDD configurations with the different DL-UL ratio (9:1 and 8:2) and generate one URLLC packet in every two subframes. In case of DL:UL=9:1 configuration, the access latency of the CASS-based TDD system (0.71 ms) is reduced.
by the factor of 87% over the LTE TDD system (5.56 ms). In a similar way, the access latency is also reduced by the factor of 82% for the DL:UL=8:2 configuration. These results demonstrate that the CASS-based access is very effective in the URLLC packet transmission. This is because 1) $T_{\text{wait}}$ is quite small in CASS-based access and 2) $T_{\text{proc}}$ is reduced substantially by using only 15% of the received samples.

V. CONCLUSION

In this paper, we have proposed the ultra low latency access scheme based on the CASS for URLLC. Our work is motivated by the observation that the waiting time to switch the transmit direction and processing time to decode the grant signal are quite large in TDD systems. The key idea behind the proposed CASS scheme is to transform the

**References**


