Throughput Maximization of Low-Latency Communication with Imperfect CSI in Finite Blocklength Regime

Yao Zhu[1], Yulin Hu[1*], Zheng Chang[2] and Anke Schmeink[1]
[1] ISEK Research Group, RWTH Aachen University, Email: $zhu|hu|schmeink$@ti.rwth-aachen.de
[2] Faculty of Information Technology, University of Jyvaskyla, Email: $zheng.chang$@jyu.fi

***Abstract*—We consider a low-latency communication network operating with finite blocklength (FBL) codes. During the transmission, the minimum mean squared error (MMSE) channel estimation is assumed to be applied to obtain the instantaneous but imperfect Channel State Information (CSI) for the rate selection. We aim at optimizing the FBL throughput of the system under given reliability constraints. First, we provide an optimal frame structure design by optimally allocating the total frame length for MMSE training of channel estimation and data transmission. In addition, we further improve the FBL throughput considering channel dynamics which optimally selects the coding rate per frame. Combining the frame structure and the coding rate selection, a joint optimization problem is studied and solved by a sub-optimal algorithm. In the simulation study, we validate the proposed analytical model and evaluate the FBL throughput of the proposed solution in comparison to benchmark schemes.**

Keywords*—*Finite blocklength, low-latency, ultra reliable communication, channel estimation.

I. Introduction

In the recent years, the interest in the new wave of fourth industrial revolution, in which communication must guarantee both low latency and high reliability has grown rapidly. In fact, such demands are key concerns in the design of many applications in the current and future wireless communication networks, e.g., autonomous vehicles, virtual/augmented reality, industrial automation and mission critical communications.

Due to the critical latency and reliability requirement, data transmissions are realized by codes with short blocklengths [2], i.e., the networks operate in the so-called finite blocklength (FBL) regime in which the transmissions are no longer arbitrarily reliable. To tackle this problem, the FBL information theoretic bound is developed in [3] for an additive Gaussian noise (AWGN) in a single hop transmission. In addition, the model has been extended to quasi-static fading channels [4], multiple antenna scenario [5], cooperative networks [6] multiple access systems [7]–[9], and wireless power transfer networks [10]. However, the results of above existing works are concluded under the assumption of either the perfect Channel State Information (CSI) or no instantaneous CSI availability at the transmitter. In a practical system, it is more likely that the system has the mechanism of channel estimation and CSI feedback, while the CSI is imperfect, e.g., due to the estimation errors. Nevertheless, more recently the work in [11] studies the FBL performance of a network operating with an imperfect CSI and with a fixed cost of channel estimation. However, for a practical channel estimator, spending a longer overhead (in terms of blocklength) provides a more accurate CSI. Note that it is shown in [3] that the shorter the blocklength is the lower the performance is. There clearly exists a tradeoff of blocklength allocation between CSI accuracy and data transmission blocklength. Beyond this tradeoff with respect to the blocklength, what is the optimal choice of the instantaneous coding rate selection based on the imperfect CSI? To the best of our knowledge, this problem has not been addressed before.

In order to acquire the CSI, there exists variant channel estimation mechanisms in practical systems [12]. Among those mechanisms, the training-based linear minimum mean-squared error (MMSE) channel estimator has been widely-used and is well-studied [13], [14]. Nevertheless, most of the works are based on the Shannon capacity by assuming infinite blocklength. In the FBL regime, the cost of the estimation needs to be considered comparing to the total blocklength resource in the packet frame.

In this work, we consider a network with a widely-used linear minimum mean-squared error (MMSE) channel estimator, and aim at maximizing the FBL throughput. We firstly develop a frame structure design by optimally allocating the total frame length to the channel estimation and the data transmission. Moreover, we exploit the estimated imperfect CSI optimally choosing the instantaneous coding rate. Via numerical evaluation, we validate our analytical model and evaluate the proposed designs.

The rest of the paper is organized as follows. We first describe the system model in Section II. The FBL throughput maximization are studied in Section III, where a frame structure design, an instantaneous coding rate selection and a joint design are provided. Finally, we provide numerical results a in Section IV and conclude the whole work in Section V.

II. System Model

We consider a single link wireless transmission system, where the transmitter sends data with a transmission rate r based on an estimated CSI. Due to a latency constraint, the length of each transmission frame is required to be less than M_{d} channel uses. Moreover, each frame is further divided into three slots. First, the transmitter takes the first slot of l channel uses to transmit training signals to the receiver. Secondly, the receiver estimates the channel based on the training signals and sends CSI feedback to the transmitter via a slot with a fixed length of F symbols. Finally, the payload is transmitted to the receiver with adjusted coding rate over the slot of n channel uses. Hence, we have $n + l \leq M_{\mathrm{d}} - F \coloneqq M$.

The channel is assumed to experience Rayleigh fading, i.e., $h \sim \mathcal{CN}(0, 1)$. Note that the channel estimation is not perfect, i.e., the exact channel coefficient h is likely to be not the same as the estimated one $\hat{h}$. In particular, the following relationship holds

$$h = \hat{h} + \Delta h \tag{1}$$

where Δh is the estimation error. Due to the MMSE estimation, $\hat{h}$ and Δh are statistically independent and satisfying $\hat{h} \sim \mathcal{CN}(0, \frac{lP_S}{1+lP_S})$ and $\Delta h \sim \mathcal{CN}(0, \frac{1}{1+lP_S})$, where P_S is the transmit power.

At the receiver, an equalizing filter $\omega = \frac{\hat{h}^*}{|\hat{h}|^2}$ is designed according to the MMSE estimation. Analog to h, the received signal after filtering can also be represented by the sum of the estimated received signal together with the "interference" signal caused by the estimation error

$$\begin{aligned} y = \hat{y} + \Delta y &= \sqrt{P_S}\omega(\hat{h} + \Delta h)x + \omega a \\ &= \sqrt{P_S}\omega\hat{h}x + \sqrt{P_S}\omega\Delta h x + \omega a, \end{aligned} \tag{2}$$

where x is the signal at the transmitter and $a \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN). The instants SNR γ can be obtained as

$$\gamma = \frac{|\hat{h}|^2 P_S}{\frac{|\omega\Delta h|^2}{|\omega|^2} P_S + \sigma^2} = \frac{P}{I+1}, \tag{3}$$

where $P = \frac{P_S|\hat{h}|^2}{\sigma^2}$ and $I = \frac{P_S|\Delta h|^2}{\sigma^2}$. Note that the channel experiencing Rayleigh fading. The closed-form cumulative distribution function (CDF) of SNR γ is

$$\begin{aligned} F_\gamma(z) &= \int_0^{+\infty} F_P(z(p+1)) f_P(p) dp \\ &= 1 - \exp\left(-\frac{z}{\kappa l P_S}\right)(1 + z/(lP_S))^{-1}, \end{aligned} \tag{4}$$

Finally, we have the probability distribution function (PDF) of γ

$$f_\gamma(z) = \frac{lP_S \exp\left(-\frac{z}{\kappa l P_S}\right)}{(lP_S + z)^2} + \frac{\exp\left(-\frac{z}{\kappa l P_S}\right)}{\kappa(lP_S + z)}, \tag{5}$$

where $\kappa = l/(lP_S + 1)$.

III. The FBL Throughput Maximization

In the FBL regime, the transmission/decoding error is no longer arbitrarily small even if the transmission rate is smaller than the Shannon capacity. According to [3], for a given coding rate r and a given channel gain z, the instantaneous (block) error probability at the receiver is

$$\varepsilon = Q\left(\sqrt{\frac{n}{V(\gamma)}}(\mathcal{C}(\gamma) - r)\log_e 2\right), \tag{6}$$

where $\mathcal{C}(\gamma) = \log 2(1+\gamma)$ is the Shannon capacity and $V(\gamma) = 1 - \frac{1}{(1+\gamma)^2}$ is the channel dispersion.

The throughput μ of the considered network is defined as the average (over channel fading) successfully transmitted information within a given total channel use M. Considering the channel distribution provided in (5), μ is given by

$$\mu = \int_0^{+\infty} \left(1 - Q\left(\sqrt{\frac{n}{V(\gamma)}}(\mathcal{C}(\gamma) - r)\log_e 2\right)\right) f_\gamma(z) r \cdot \frac{n}{M} dz. \tag{7}$$

Following this throughput model, in the following subsections we provide a constant and an instantaneous design to maximize the FBL throughput. To further exploiting the acquired CSI, we also propose a joint optimal design combing both frame structure design and instantaneous coding rate selection.

A. Frame Structure Design

We first consider a constant design on the frame structure. Note that in each frame, the two slots for channel estimation and data transmission share at most M symbols, i.e., $l + n \leq M$. To design the frame structure actually requires to determine the optimal lengths for these two slots. On the one hand, the estimation slot l influences the quality of channel estimation. As stated in (1), the variance of SNR is a monotonic increasing function with respect of l. In other words, a longer l makes the estimated SNR be statistically closer to the real one, i.e., $\Delta h \to 0$ and $\hat{h} \to 1$. On the other hand, as the slot for the packet transmission, n directly influences the system performance. In particular, for a fixed transmission rate r_0, the error probability ε is monotonic decreasing in n. If n is infinite or sufficiently long, the error probability is arbitrarily small for an AWGN channel, which leads us to the infinite blocklength regime. Since the overall blocklength $M = n + l$ indicates the resource of the system, there exists clearly a trade-off between n and l in terms of resource allocation. Therefore, we aim at maximizing the throughput at the receiver by scheduling the parameter n and l. Mathematically, the optimization problem is formulated as,

$$\underset{n,l}{\text{maximize}} \quad \mu \tag{8a}$$

$$\text{subject to} \quad l + n \leq M, \ (l, n) \in \mathbb{Z}_+, \tag{8b}$$

$$\varepsilon \leq \varepsilon_t. \tag{8c}$$

Firstly, we provide the following lemma to replace the inequality constraint in (8b) by a equality constraint.

Lemma 1. *In order to maximize the throughput μ, the latency constraint in* (8b) *is required to hold with equality.*

Proof: To prove Lemma 1 by contradiction, we assume that there exists an optimal solution (l', n') which satisfies the constraint with strict inequality, i.e., $M - (l' + n') = \alpha > 0$. Hence, the optimal result μ' with this optimal solution (l', n') is always the global maximum, namely $\mu'(l', n') \geq \mu(l, n)$. On the other hand, we can find another feasible solution $(l'' = l', n'' = n' + \alpha) \in \{l, n | l + n \leq M\}$. In our previous works [6], we showed the throughput is a strictly increasing function with respect to the blocklength of transmission n. Hence, we conclude that the solution (l'', n'') results in a throughput $\mu'' > \mu'$, i.e., the assumption that (l', n') is the optimal solution of the problem is violated. ■

We can therefore eliminate the parameter n by substituting n by $M - l$. Furthermore, we replace the original problem by the relaxation of l. We denote $l^o \in (0, M)$ the optimal solution of the relaxed problem. Then, the optimal solution l^* of the original problem can be determined by comparing the neighbor integers of l^o, i.e., we can obtain $l^* = \underset{l \in \{l_{\text{ceil}}, l_{\text{floor}}\}}{\arg\max}\, \mu$, where $l_{\text{ceil}} = \lceil l^o \rceil$ and $l_{\text{floor}} = \lfloor l^o \rfloor$ are the results of ceil and floor functions, respectively. Noting that the error probability is a monotonic decreasing function in n, the constraint (8c) can be replaced equivalently by $n \geq n_t$, where $n_t = \varepsilon_t^{-1}(n)$. Hence, Problem (8) is reformulated as

$$\begin{aligned}
\underset{l}{\text{maximize}} \quad & \mu && \text{(9a)}\\
\text{subject to} \quad & l \leq M, \qquad l \in \mathbb{R}_+, && \text{(9b)}\\
& n = M - l, && \text{(9c)}\\
& n \geq n_t, \qquad n \in \mathbb{R}_+. && \text{(9d)}
\end{aligned}$$

To solve this problem, we establish following lemma:

Lemma 2. *There exists only one optimal solution l^*, so that $\mu(l^*) = \max \mu(l)$, where $l \in (0, M)$.*

Proof: As stated in [16], the error probability in the finite blocklength regime can be approximated as follows by applying the Laplace Method:

$$\begin{aligned}
\mathbb{E}(\varepsilon) &= \int_0^{+\infty} Q\left(\sqrt{\frac{n}{V(z)}}(\mathcal{C}(z) - r)\log_e 2\right) f_\gamma(z) dz \\
&\approx \sqrt{\frac{n}{\pi}} \sqrt{\frac{2\pi}{Tg''(\gamma_0)}} v(\gamma_0) \exp(-Tg(\gamma_0|l)) \\
&\approx F_\gamma(\gamma_0),
\end{aligned} \tag{10}$$

where $\gamma_0 = 2^r - 1$, $v(\gamma_0) = \sqrt{1/(\gamma_0(\gamma_0 + 2))} F_\gamma(\gamma_0)$, $T = \frac{n}{2}$ and $g(\gamma_0|l) = \frac{(C(\gamma_0|l) - r)^2}{V(\gamma_0|l)}$. Note that $M \geq n + l$. Based on (10), we reformulate μ as

$$\mu = (1 - F(\gamma_0|l)) r \frac{M - l}{M}. \tag{11}$$

We have the first order derivative of μ of l

$$\frac{\partial \mu}{\partial l} = A \cdot K(l), \tag{12}$$

where $A = -\exp\left(\frac{-\gamma_0 l P_S - 1}{l P_S^2}\right) (l P_S(\gamma_0 + l P_S)^2)^{-1}$ is a non-zero real constant. $K(l)$ is a cubic function with respect of l and it can be expressed as

$$\begin{aligned}
K(l) &= P_S^3 l^3 + (\gamma_0 P_S + 2\gamma_0 P_S^2) l^2 \\
&\quad + (\gamma_0^2 - \gamma_0 P_S^2 - \gamma_0 P_S) l - \gamma_0^2 \\
&= a l^3 + b l^2 + c l + d.
\end{aligned} \tag{13}$$

Through a discriminant test, where the discriminant of a cubic function is $\Delta = 18abcd - 4b^3 d + b^2 c^2 - 4ac^3 - 27a^2 d^2 > 0$, we know that $K(l)$ possesses three real roots, i.e., there exists three local extrema $\mu(l_i^*)$ so that $\mu'(l_i^*) = 0$, where $i = 1, 2, 3$. To further determine whether those extrema are maxima or minima, the second derivative of μ is required to be explored, which seems intractable. Therefore, we have

$$\frac{\partial K}{\partial l} = 3al^2 + 2bl + c. \tag{14}$$

The roots of $\frac{\partial K}{\partial l}$ is then $l_{1,2}^o = \frac{-2b - \sqrt{4b^2 - 16ac}}{8a}$. Noting $b = \gamma_0 P_S + 2\gamma_0 P_S^2 > 0$, one of the roots $l_{1,2}^o$ is always negative, which is infeasible. This implies that only two of $l_i^* \in (0, M)$ exists. In other word, there exists at most one local maximum and one local minimum with feasible $l \in (0, M)$. Moreover, we have $0 \leq F(\gamma_0) \leq 1$ and $l \leq M$, which results $\mu = (1 - F(\gamma_0)) r \frac{M - l}{M} \geq 0$. Readily, we can also obtain that $\mu(0) = 0$ and $\mu(M) = 0$ are the minimum of objective function. As a result, μ is a quasi-concave function with a global maximum $\mu(l^*)$, where $l^* \in (0, M)$. ■

According to the above lemma, the quasi-concave problem can be efficiently solved by many optimization techniques.

B. Instantaneous Coding Rate Selection

Next, we exploit the channel estimation to optimize the coding rate in the transmission frame-wise. Recall that the transmitter takes the first slot of l channel uses to transmit training signals to the receiver. The receiver estimates the channel based on the training signals and sends CSI feedback to the transmitter. Finally, the payload is transmitted to the receiver with adjusted coding rate over the slot of n channel uses. This operation is executed in each frame and the coding rate can be modified based on the currently estimated CSI. We assume the feedback transmission of the feedback takes a constant cost, i.e., F in symbols. Hence, the instantaneous throughput (namely, the statistically transmitted information bits in a single frame duration) can be improved by solving the following optimization problem:

$$\begin{aligned}
\underset{r}{\text{maximize}} \quad & \mu_i = (1 - \varepsilon(r|z_i, n))\, r_i \frac{n}{M} && \text{(15a)}\\
\text{subject to} \quad & \varepsilon \leq \varepsilon_t, && \text{(15b)}
\end{aligned}$$

where $i \in \{0, 1, ..., N\}$ indicates the index of the frame. To solve this optimization problem, we prove the concavity of the objective function with respect to r, as stated in following lemma.

Lemma 3. *μ_i is a concave function with respect to r_i, where $r_i \in (0, \mathcal{C})$.*

Proof: The concavity is proven by the second derivative test of μ_i as follows,

$$\frac{\partial^2 \mu_i}{\partial r_i^2} = \left(-\frac{n}{M} 2e^{-\frac{W^2}{2}}(2 + W)\right) \tag{16}$$

where $W = \sqrt{\frac{n}{V}}(\mathcal{C} - r) \geq 0$. Readily, we can obtain $\frac{\partial^2 \mu_i}{\partial r_i^2} \leq 0$, i.e., μ_i is concave in r_i. ■

C. Joint Design

Note that the proposed designs with respect to (l, n) and r in the above two subsections are independent. In the following, we combine them together. it can be considered as the following joint optimization problem with respects to

(l, n, r):

$$\underset{l,n,r}{\text{maximize}} \quad \int_0^{+\infty} \max_r (1 - \varepsilon(z, r, n)) r \frac{n}{M} f_\gamma(z|n) dz \tag{17a}$$

$$\text{subject to} \quad \varepsilon \leq \varepsilon_t, \tag{17b}$$

$$l + n \leq M, \, (l, n) \in \mathbb{Z}_+. \tag{17c}$$

However, the global solution of Problem (17) is analytically unpredictable as the system outage probability is a double integral form with the up-bounds given by the instantaneous throughput. Therefore, in the following we propose a sub-optimal solution.

In the proposed sub-optimal solution, we start with considering the offline optimization problem:

$$\underset{l,n,r}{\text{maximize}} \quad \int_0^{+\infty} (1 - \varepsilon(z, r, n)) r \frac{n}{M} f_\gamma(z|n) dz \tag{18a}$$

$$\text{subject to} \quad \varepsilon \leq \varepsilon_t, \tag{18b}$$

$$l + n \leq M, \, (l, n) \in \mathbb{Z}_+. \tag{18c}$$

In order to solve this problem, we provide the following Lemma.

Lemma 4. *If the coding rate is constant during the transmission, the average throughput μ is a concave function with respect to r, where $r \in (0, \varepsilon_t^{-1}(r))$.*

Proof: The average throughput μ is the mean value of instantaneous throughput μ_i with the frame length N going to infinite. It reads as

$$\mu = \lim_{N \to \infty} \frac{1}{N} \sum \mu_i(r), \tag{19}$$

According to Lemma 3, μ_i is a concave function with respect to r. Hence, μ as a summation of the concave functions is also concave. ∎

We determine the optimal solution l^*_{off}, n^*_{off} and r^*_{off} of Problem (18) depending on the average channel gain before the transmission. In each frame, we solve Problem (15) according to Lemma 3 by setting $l = l^*_{\text{off}}$ and $n = n^*_{\text{off}}$ and obtain the optimal solution r^*_i for the i-th frame.

Note that the optimal solution l^* and n^* is based on the assumption that the coding rate r_{off} cannot be adjusted frame-wise and constant over the whole transmission. Therefore, the result μ^* that we obtained by l^*, n^* and r^*_i, is a sub-optimal result. However, the impact of the instantaneous sub-optimal coding rate to the overall throughput is limited, since the influence of estimated Rayleigh channel over whole transmission depends on the accuracy of estimation instead of the channel variance. In the next section, we show numerically the performance of the proposed algorithm is close to performance of the global optimal solution.

IV. Simulation Results

In this section, we provide the numerical results. In particular, we validate our analytical models by Monte Carlo simulations. Subsequently, we show the performance advantage of the proposed design by comparing the numerical results with other benchmark schemes. In all the simulations, we consider

Algorithm 1 for solving the joint optimization problem

1: Initialize $l^k_{\text{off}} = n^k_{\text{off}} = M/2, k = 0$.
2: Solving the local problem $\max_r \mu(r, l^k_{\text{off}}, n^k_{\text{off}})$ according to Lemma 4, and denote the solution by r^{k+1}_{off}.
3: Solving the local problem $\max_{l,n} \mu(r^{k+1}_{\text{off}}, l, n)$ according to Lemma 2, and denote the solution by l'_{off}.
4: Let $l_{\text{off,ceil}} = \lceil l'_{\text{off}} \rceil$ and $l_{\text{off,floor}} = \lfloor l'_{\text{off}} \rfloor$. The optimal solution is then $l^{k+1}_{\text{off}} = \underset{l \in \{l_{\text{off,ceil}}, l_{\text{off,floor}}\}}{\arg\max} \mu(r^{(k+1)}, l, n)$.
5: Updating the iteration indicator k=k+1.
6: Repeating steps 1)-6) till the solution converges, and denote the solution $l^*_{\text{off}} = l^k_{\text{off}}, n^*_{\text{off}} = n^k_{\text{off}}$.
7: In each i-th frame, solving the problem $\max_{r_i} \mu_i(r_i, l^*_{\text{off}}, n^*_{\text{off}}, \hat{h})$ after acquiring estimated CSI $\hat{h}$ according to Lemma 3, and denote the solution by r^*_i.
8: Outputting $l^{\text{opt}} = l^*_{\text{off}}, n^{\text{opt}} = n^*_{\text{off}}$ and $r^{\text{opt}}_i = r^*_i$.

the following setups. First, we set transmit power $P_S = 10$ dB and receiver noise power $\sigma^2 = 0.01$. In addition, we assume the distance between transmitter and receiver $d = 50$ m. Moreover, the reliability constraint is set to 99.9999%, i.e., the error probability threshold $\varepsilon_t = 0.0001$. Finally, l_0 is set to 80, following the same portion of pilot symbols in LTE [17] with $M = 3000$ per frame.

A. Validation

We start the validation with Fig. 1 to present the average throughput performance while varying the blocklength of the first slot l. Readily, it confirms our analytical model that μ is concave with respect to l for all the cases with different coding rate selections. In particular, the throughput increases dramatically with the initially increasing l due to the increase of the SNR after filtering, namely the increase of the reliability. The more portion of frame that l possesses, thehigher accuracy of estimation is achieved by the transmission. Nevertheless, the error probability is a non-linear function and becomes quasi

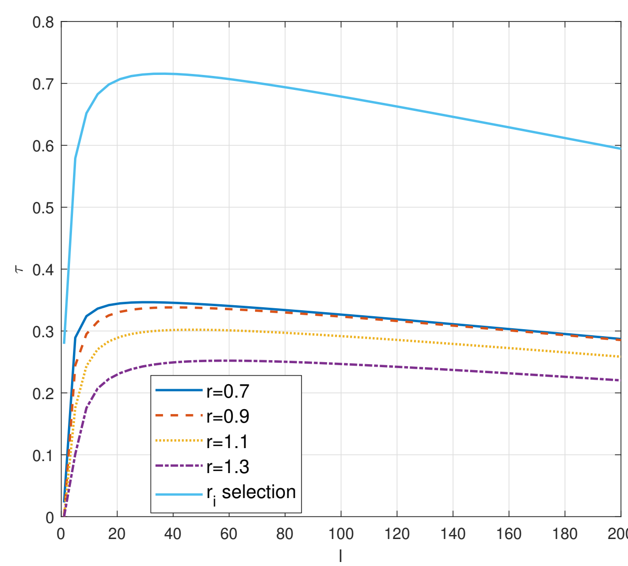

Fig. 1. The throughput μ versus blocklength l under $M = 900$ over Rayleigh fading channel.

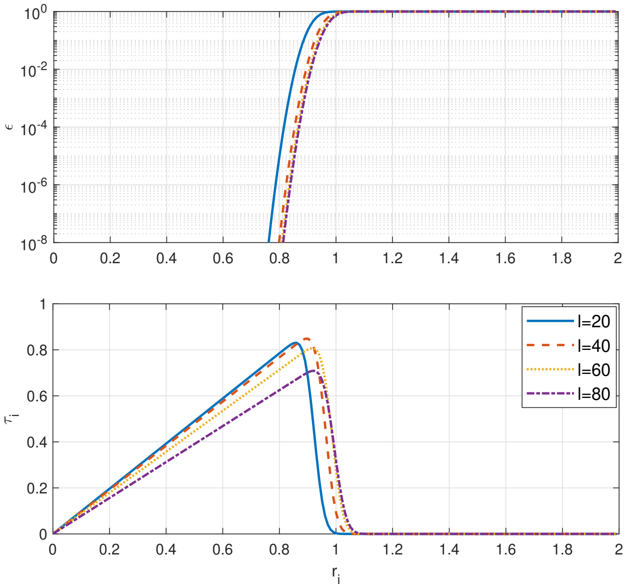

Fig. 2. The instantaneous throughput μ and error probability ε versus coding rate r_i under $M = 900$ with constant channel gain $h = 1$ under variant frame structures.

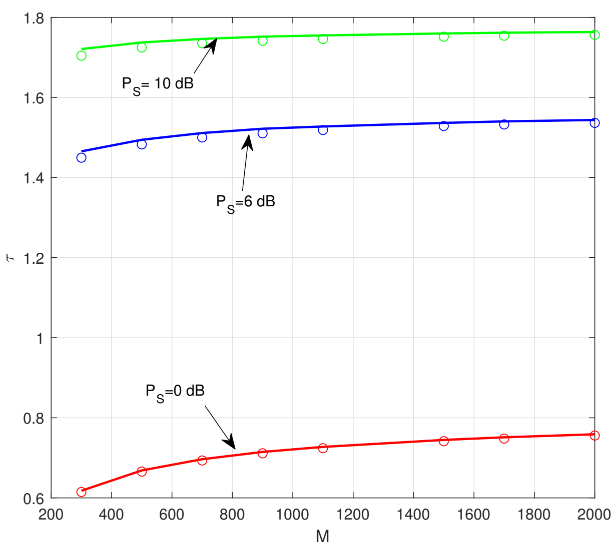

Fig. 3. The optimal result μ^* versus the total blocklength M under varying transmit power P_S over a Rayleigh fading channel $h \sim \mathcal{CN}(0, 1)$. The results of the proposed algorithm and exhaustive search are illustrated by marks and curves, respectively.

flat-topped for a high SNR. In this regime, the contribution to the throughput is dominated by the length of the second slot $n = M - l$ instead of l, which represents the accuracy of the estimation. Therefore, there exists a trade-off between l and n. It can be observed intuitively from the figure. Moreover, The shape of the trade-off is also influenced by the coding rate. Clearly, the system with instantaneous coding rate selection significantly outperforms the system with constant coding rate in the offline optimization scheme, since the instantaneous coding rate influences not only the transmitted information in the frame, but also the error probability according to (9). It motivates us to further investigate the relationship between the error probability and the throughput in the perspective of r.

The impact of the coding rate (within transmission frame i) on the instantaneous throughput and error probability are shown in Fig. 2, where the instantaneous channel gain is set to $h = 1$. As expected, the throughput shows quasi-concavity with respect to r, which is confirmed by Lemma 3. In particular, it is observed that μ_i is concave in the coding rate r_i when $\varepsilon < 0.5$, i.e., the error probability has a value of practical interests in a reliable transmission scenario. It can be also observed that the throughput increases almost linearly in r_i in the low rate region, since in this case the bottleneck of the system performance is the coding rate. In addition, as r_i approaches closer to the capacity, the throughput drops dramatically, as the transmission error becomes considerable. According to (7), when r_i is set to the Shannon capacity, the ε becomes 0.5. Hence, in the FBL regime guaranteeing a reliable transmission requires to set the coding rate much lower than the Shannon bound. Note that in this work, our designs are under reliability constraints. The figure actually indicates that when the system has a stringent error probability threshold, e.g., $\varepsilon_t = 10^{-4}$, the top of the throughput curves can not be achieved. In other words, the cost (in terms of throughput) of guaranteeing a reliable transmission is significant. In particular, the feasible throughput can be approximated by a linear function $\mu \approx r\frac{M-l}{M}$ under stringent reliability constraints.

To confirm the performance of the proposed joint design, we compare it with the exhaustive search scheme. The results are provided in Fig. 3, where different transmit power setups are considered. First of all, it is shown that the throughput is monotonously increasing in the total blocklength M. In addition, a high transmit power P_S also leads to a high throughput. Secondly, the performance of our proposed sub-optimal algorithm provides a performance close to the exhaustive search scheme representing the global optimal solution. This confirms the effectiveness of the proposed joint design. Finally, we also observe that throughput improvement by the joint design is more significant when M is short.

B. Performance Comparison

In this subsection, to show the advantage of the proposed algorithm, we compare the performance of the proposed design with the following three benchmark schemes:

1) Perfect CSI scheme. The perfect CSI is prior-acquired by both transmitter and receiver without any estimation, i.e., $h = \hat{h}$ and $l = 0$. Moreover, the coding rate r is still adjustable in each frame depending on the CSI. This scheme provides an upper bound of the system performance, although this bound is not tight due to the fact that perfect CSI is unrealistic in practical systems.
2) Average CSI scheme. in this case, no instantaneous CSI is obtained, the frame structure is designed purely based on the average CSI. Therefore, the coding rate is also selected according to the average CSI, i.e., $r_i^* = r_{\text{off}}^*$. In other words, as long as the coding rate is determined, it is fixed over all frames. This scheme is expected to represent a lower bound performance of the considered network.
3) MMSE estimation scheme with a constant length. In this scheme, the length of the slot for channel estimation is constant, i.e., l has a constant value

l_0. In this case, the Problem (17) degenerates into Problem (15) with the constraint of $l = l_0$.

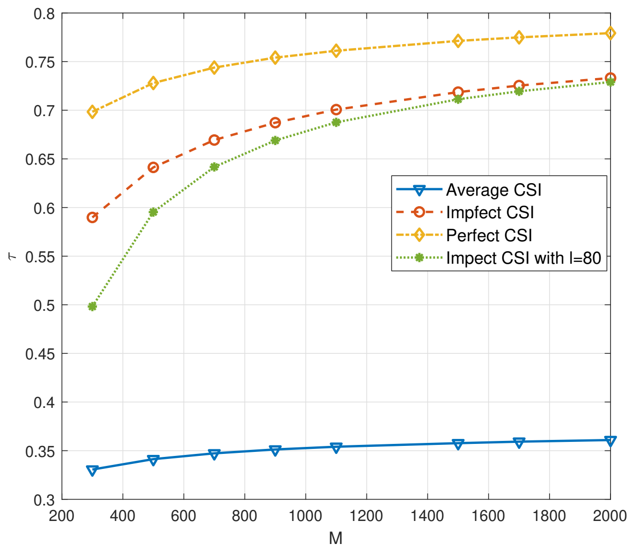

Fig. 4. Performance comparison with different schemes.

The comparison results are provided in Fig. 4, where the average throughput is evaluated while varying the frame length M. Note that the instantaneous coding rate selection cannot be applied to the average CSI scheme. Instead, in the simulation, we consider a constant coding rate optimization to maximize the average throughput.

Clearly, throughputs are increasing in M for all schemes. In particular, the system under the perfect CSI scheme always outperforms other schemes and the system under the average CSI scheme performs poorly even with a significantly long M. These two schemes actually provide an upper bound and a lower bound of the considered system with an inaccurate CSI, respectively. We can observe that the proposed design is closer to the upper-bound. Finally, we can observe that the proposed design also shows a performance advantage in comparison to the MMSE scheme with constant length of estimator. In particular, this performance advantage is more significant when M is short. It implies that the frame design with imperfect CSI is desirable in low-latency short blocklength scenarios.

V. Conclusion

In this work, we considered a single link wireless transmission system with MMSE estimated imperfect CSI. Aiming at the FBL throughput under given reliability constraints, we proposed a frame structure design by allocating the slot of channel estimation and data transmission. Moreover, we provided an instantaneous adapted coding rate selection based on an estimated CSI. We futher also studied the joint optimization problem by combing the frame structure design and coding rate selection. We solved the problem with a sub-optimal solution. Via numerical simulations, we validated the analytic model and compared the performance of our proposed algorithm with other benchmark schemes.

References

[1] B. Makki, T. Svensson and M. Zorzi, "Finite Block-Length Analysis of the Incremental Redundancy HARQ," in IEEE Wireless Communications Letters, vol. 3, no. 5, pp. 529-532, Oct. 2014.

[2] Y. Hu, M. C. Gursoy and A. Schmeink, "Relaying-Enabled Ultra-Reliable Low-Latency Communications in 5G," *IEEE Network*, vol. 32, no. 2, pp. 62-68, March-April 2018.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010

[4] S. Xu, T. Chang, S. Lin, C. Shen and G. Zhu, "Energy-Efficient Packet Scheduling With Finite Blocklength Codes: Convexity Analysis and Efficient Algorithms," *IEEE Trans. Wireless. Commn.*, vol. 15, no. 8, pp. 5527-5540, Aug. 2016.

[5] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy and W. Yang, "Short-Packet Communications Over Multiple-Antenna Rayleigh-Fading Channels," *IEEE Trans. Commn.*, vol. 64, no. 2, pp. 618-629, Feb. 2016.

[6] Y. Hu, J. Gross and A. Schmeink, "On the capacity of relaying with finite blocklength," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 17901794, Mar. 2016.

[7] Xiaofang Sun, Shihao Yan, Nan Yang, Zhiguo Ding, Chao Shen, Zhangdui Zhong, "Short-Packet Downlink Transmission with Non-Orthogonal Multiple Access," *IEEE Trans. Wireless. Commn.*, Early Access, Apr. 2018.

[8] Y. Hu, M. C. Gursoy and A. Schmeink, "Efficient Transmission Schemes for Low-Latency Networks: NOMA vs. Relaying", in *Proc. IEEE PIMRC*, Montreal, QC, Oct. 2017.

[9] Y. Hu, M. Ozmen, M. C. Gursoy and A. Schmeink, "Optimal Power Allocation for QoS-Constrained Downlink Multi-User Networks in the Finite Blocklength Regime," *IEEE Trans. Wireless. Commn.* vol. 17, no. 9, pp. 5827-5840, Sept. 2018.

[10] Y. Hu, Y. Zhu, M. C. Gursoy and A. Schmeink, "SWIPT-Enabled Relaying in IoT Networks Operating With Finite Blocklength Codes," *IEEE JSAC*, vol. 37, no. 1, pp. 74-88, Jan. 2019.

[11] Y. Hu, J. Gross and A. Schmeink, "Optimal Scheduling of Reliability-Constrained Relaying System under Outdated CSI in the Finite Blocklength Regime", *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6146-6155, July. 2018

[12] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Borjesson, "On channel estimation in OFDM systems," in Proc. IEEE 45th Vehicular Technology Conference. *Countdown to the Wireless Twenty-First Century*, Chicago, IL, USA, 1995, pp. 815-819, vol. 2.

[13] Baoguo Yang, Zhigang Cao and K. B. Letaief, "Analysis of low-complexity windowed DFT-based MMSE channel estimator for OFDM systems," *IEEE Transactions on Communications*, vol. 49, no. 11, pp. 1977-1987, Nov. 2001.

[14] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884-893, March 2006.

[15] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, April 2003.

[16] P. Mary, J. Gorce, A. Unsal and H. V. Poor, "Finite Blocklength Information Theory: What Is the Practical Impact on Wireless Communications?," in Proc. *IEEE Globecom Workshops*, Washington, DC, 2016, pp. 1-6.

[17] 3GPP TS 36.141. "Base Station (BS) conformance testing." *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA)*. URL: http://www.3gpp.org.