University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications from the Department of Electrical Engineering

Electrical Engineering, Department of

1-1-2011

Throughput Analysis of Buffer-Constrained Wireless Systems in the Finite Blocklength Regime

M. Cenk Gursoy
University of Nebraska - Lincoln, gursoy@engr.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/electricalengineeringfacpub

Part of the Electrical and Computer Engineering Commons

Cenk Gursoy, M., "Throughput Analysis of Buffer-Constrained Wireless Systems in the Finite Blocklength Regime" (2011). *Faculty Publications from the Department of Electrical Engineering.* Paper 168.
http://digitalcommons.unl.edu/electricalengineeringfacpub/168

This Article is brought to you for free and open access by the Electrical Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Department of Electrical Engineering by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

IEEE International Conference on Communications (ICC), 2011
doi: 10.1109/icc.2011.5963210

Throughput Analysis of Buffer-Constrained Wireless Systems in the Finite Blocklength Regime

Mustafa Cenk Gursoy
Department of Electrical Engineering
University of Nebraska-Lincoln, Lincoln, NE 68588
Email: gursoy@engr.unl.edu

***Abstract*—[1] In this paper, wireless systems operating under queueing constraints in the form of limitations on the buffer violation probabilities are considered. The throughput of wireless systems operating under such constraints is captured by the effective capacity formulation. It is assumed that finite blocklength codes are employed for transmission. Under this assumption, a recent result on the channel coding rate in the finite blocklength regime is incorporated into the analysis and the throughput achieved with such codes in the presence of queueing constraints is identified. Interactions between the effective rate, queueing constraints, error probabilities, and blocklength values are investigated. In particular, it is shown that for given signal-to-noise ratio, blocklength, and quality of service exponent, the effective rate is maximized at a unique decoding error probability value.**

I. Introduction

Providing quality of service (QoS) guarantees in the form of limitations on the queueing delays or buffer violation probabilities is of paramount importance in many delay-sensitive wireless systems, e.g., voice over IP (VoIP), and interactive and streaming video applications. In [1], effective capacity is proposed as a metric that can be employed to measure the performance in the presence of statistical QoS limitations. Effective capacity formulation uses the large deviations theory and incorporates the statistical QoS constraints by capturing the rate of decay of the buffer occupancy probability for large queue lengths. Hence, effective capacity can be regarded as the maximum throughput of a system operating under limitations on the buffer violation probability.

Recently, there has been much interest in the analysis of the effective capacity of fading channels (see e.g., [5] – [8] and references therein) in order to identify the performance of wireless systems operating subject to statistical queueing constraints. However, in almost all prior studies, the service rates of the queueing model (or equivalently the instantaneous transmission rates over the wireless channel) are assumed to be equal to the instantaneous capacity values although coding is assumed to be performed using a finite block of symbols. Moreover, transmissions are assumed to be error-free.

In this paper, we depart from these idealistic assumptions and consider channel coding rates achievable with finite blocklength codes, and incorporate the error probabilities and possible retransmission scenarios into the model. This analysis is facilitated mainly by the recent results of Polyanskiy, Poor, and Verdú in [9] where the authors identified an approximate maximal achievable rate expression for a given error probability in the finite blocklength regime. This expression can be regarded as a second-order asymptotic approximation of the channel coding rate at large but finite blocklength values. Throughout the rest of the paper, we use this coding rate expression in the evaluation of the effective throughput and analyze the interactions between the queueing constraints, error probabilities, effective rate, and blocklength.

II. Channel Model

We consider a frequency-flat channel model, and assume that the fading coefficients stay fixed for a block of m symbols and then change independently for the following block. Under this block-fading assumption, the channel input-output relation in one coherence block can be expressed as

$$\mathbf{y} = h\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the m-dimensional, complex, channel input and output vectors, respectively. h is the complex-valued fading coefficient with finite second moment, i.e., $\mathbb{E}\{|h|^2\} < \infty$. We assume that both the receiver and transmitter have perfect channel side information (CSI) and hence perfectly know the instantaneous realizations of the fading coefficients. Finally, $\mathbf{n}$ represents the Gaussian noise vector whose components are independent and identically distributed (i.i.d.), complex, circularly symmetric, Gaussian random variables with mean zero and variance N_0, i.e., $\mathbf{n} \sim \mathcal{CN}(0, N_0\mathbf{I}_m)$ where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix.

III. Throughput under Statistical Queueing Constraints

In [1], Wu and Negi defined the effective capacity as the maximum constant arrival rate that a given service process can support in order to guarantee a statistical QoS requirement specified by the QoS exponent θ [2]. If we define Q as the stationary queue length, then θ is the decay rate of the tail of the distribution of the queue length Q:

$$\lim_{q\to\infty} \frac{\log P(Q \geq q)}{q} = -\theta. \tag{2}$$

Therefore, for large $q_{\max}$, we have the following approximation for the buffer violation probability:

$$P(Q \geq q_{\max}) \approx e^{-\theta q_{\max}}.$$

[1]This work was supported by the National Science Foundation under Grants CCF – 0546384 (CAREER) and CCF–0917265.

[2]For time-varying arrival rates, effective capacity specifies the effective bandwidth of the arrival process that can be supported by the channel.

978-1-61284-233-2/11/$26.00 ©2011 IEEE

Hence, while larger θ corresponds to more strict QoS constraints, smaller θ implies looser QoS guarantees. Similarly, if D denotes the steady-state delay experienced in the buffer, then $P(D \geq d_{\max}) \approx e^{-\theta\delta d_{\max}}$ for large $d_{\max}$, where δ is determined by the arrival and service processes [5]. Therefore, effective capacity formulation provides the maximum constant arrival rates that can be supported by the time-varying wireless channel under the queue length constraint $P(Q \geq q_{\max}) \leq e^{-\theta q_{max}}$ for large q_{max} or the delay constraint $P(D \geq d_{\max}) \leq e^{-\theta\delta d_{\max}}$ for large $d_{\max}$. Since the average arrival rate is equal to the average departure rate when the queue is in steady-state [4], effective capacity can also be seen as the maximum throughput in the presence of such constraints.

The effective capacity is given by ([1], [2], [3])

$$R_E = -\lim_{t\to\infty} \frac{1}{\theta t} \log_e \mathbb{E}\{e^{-\theta S[t]}\} \tag{3}$$

where $S[t] = \sum_{i=1}^{t} R_i$ is the time-accumulated service process and $\{R_i, i = 1, 2, \ldots\}$ denotes the discrete-time stationary and ergodic stochastic service process. We would like to note that in the remainder of the paper, we will refer to R_E as the effective rate rather than the effective capacity since R_E is in our setup will be the throughput when the service rates are equal to the approximate channel coding rates in the finite blocklength regime.

IV. Effective Throughput with Finite Blocklength Codes

In [9], the authors have studied the channel coding rate in the finite blocklength regime. For general classes of channels, they have obtained new achievability and converse bounds on the coding rate for a given finite blocklength and error probability. In particular, for the real, additive white Gaussian noise (AWGN) channel, the transmission rate (in bits per m channel uses) with error probability $0 < \epsilon < 1$, signal-to-noise ratio (SNR), and coding blocklength m is shown to have the following asymptotic expression [9, Theorem 54]:

$$\begin{aligned} r =& \frac{m}{2}\log_2(1+\text{SNR}) \\ &- \sqrt{\frac{m}{2}\left(1-\frac{1}{(\text{SNR}+1)^2}\right)} Q^{-1}(\epsilon)\log_2 e + O(\log m) \end{aligned} \tag{4}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ is the Gaussian Q-function. Denoting the rate in bits per channel use by $\bar{r}$, we can write

$$\begin{aligned} \bar{r} = \frac{r}{m} =& \frac{1}{2}\log_2(1+\text{SNR}) \\ &- \sqrt{\frac{1}{2m}\left(1-\frac{1}{(\text{SNR}+1)^2}\right)} Q^{-1}(\epsilon)\log_2 e + \frac{O(\log m)}{m} \\ \approx& \frac{1}{2}\log_2(1+\text{SNR}) \\ &- \sqrt{\frac{1}{2m}\left(1-\frac{1}{(\text{SNR}+1)^2}\right)} Q^{-1}(\epsilon)\log_2 e \end{aligned} \tag{5}$$

where the approximation is accurate for sufficiently large m. Note that the above results are for the AWGN channel with real input and real output.

In this paper, we consider a fading Gaussian channel model with complex-valued input and output, and assume that channel coding is performed in each coherence interval of m symbols, during which the fading stays fixed. Under these assumptions, coding over a fading Gaussian channel can be seen as coding over a real Gaussian channel (with a certain channel gain) using a coding blocklength of $2m$. The following arguments provide a detailed description of this approach. Knowing the channel fading coefficient h, the receiver can multiply the received signal with $e^{-j\theta_h}$, where θ_h is the phase of h, and obtain

$$\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_r + j\tilde{\mathbf{y}}_i = \mathbf{y}e^{-j\theta_h} = |h|\mathbf{x} + \tilde{\mathbf{n}} = |h|\mathbf{x}_r + \tilde{\mathbf{n}}_r + j(|h|\mathbf{x}_i + \tilde{\mathbf{n}}_i) \tag{6}$$

where $\tilde{\mathbf{y}}_r$, $\mathbf{x}_r$, $\tilde{\mathbf{n}}_r$ and $\tilde{\mathbf{y}}_i$, $\mathbf{x}_i$, $\tilde{\mathbf{n}}_i$ denote the real and imaginary components, respectively, of the output vector $\tilde{\mathbf{y}}$, input vector $\mathbf{x}$, and noise vector $\tilde{\mathbf{n}}$. It can be easily verified that $\tilde{\mathbf{n}} = \mathbf{n}e^{-j\theta_h}$ has the same statistics as $\mathbf{n}$ and hence $\tilde{\mathbf{n}} \sim \mathcal{CN}(0, N_0\mathbf{I}_m)$. Now, the above channel input-output relation can also be written as

$$[\tilde{\mathbf{y}}_r\ \tilde{\mathbf{y}}_i] = |h|\,[\mathbf{x}_r\ \mathbf{x}_i] + [\tilde{\mathbf{n}}_r\ \tilde{\mathbf{n}}_i] \tag{7}$$

where $[\tilde{\mathbf{y}}_r\ \tilde{\mathbf{y}}_i]$ denotes the vector formed by concatenating $\tilde{\mathbf{y}}_r$ and $\tilde{\mathbf{y}}_i$. Since the real and imaginary components are m-dimensional vectors, the above channel model is a real Gaussian channel with $2m$ dimensional input and output and with channel gain $|h|$. Note that the real and imaginary noise components $\tilde{\mathbf{n}}_r$ and $\tilde{\mathbf{n}}_i$ are independent due to the assumption of the circular symmetry of the additive complex Gaussian noise. For this channel, the coding rate (in bits per m channel uses) in the ith block achieved with block error probability ϵ is

$$\begin{aligned} r_i =& m\log_2(1+\text{SNR}|h_i|^2) \\ &- \sqrt{m\left(1-\frac{1}{(\text{SNR}|h_i|^2+1)^2}\right)} Q^{-1}(\epsilon)\log_2 e + O(\log 2m) \end{aligned} \tag{8}$$

where h_i denotes the fading coefficient in the ith block. Note that the expression in (8) is obtained from that in (4) by replacing m with $2m$, and SNR with $\text{SNR}|h_i|^2 = \frac{P}{N_0}|h_i|^2$, which is the received signal-to-noise ratio in the ith block. Now, the normalized rate in bits per channel use is approximately

$$\begin{aligned} \bar{r}_i = \frac{r_i}{m} =& \log_2(1+\text{SNR}|h_i|^2) \\ &- \sqrt{\frac{1}{m}\left(1-\frac{1}{(\text{SNR}|h_i|^2+1)^2}\right)} Q^{-1}(\epsilon)\log_2 e \end{aligned} \tag{9}$$

for large enough m for which $\frac{O(\log 2m)}{m}$ is negligible. Henceforth, we assume that the instantaneous transmission rate in each coherence block of the fading channel is given by the expression in (9). Since the block error rate is ϵ, this rate is attained with probability $1-\epsilon$. We assume that the receiver reliably detects the errors, employs a simple ARQ mechanism and sends a negative acknowledgement requesting the retransmission of the message in case of an erroneous reception. Therefore, the data rate is effectively zero when error occurs. Under this assumption, the

service rate (in bits per m channel uses) in each block is

$$R_i = \begin{cases} 0 & \text{with prob. } \epsilon \\ m\bar{r}_i & \text{with prob. } (1-\epsilon) \end{cases}. \tag{10}$$

With the above service rate characterization, we immediately obtain the following expression for the effective rate.

Proposition 1: The effective rate (in bits per channel use) at a given SNR, error probability ϵ, blocklength m, and QoS exponent θ is

$$R_E(\theta) = -\frac{1}{m\theta}\log_e \mathbb{E}_{|h|^2}\left\{\epsilon + (1-\epsilon)e^{-\theta m\bar{r}}\right\} \tag{11}$$

where $\bar{r}$ is given in (9) and the expectation is with respect to $|h|^2$.

Proof: We first note that the service rate $\{R_i\}$ is an i.i.d. process due to the facts that the fading process is i.i.d. in different blocks and the noise is an i.i.d. process leading to the independence of error events in different blocks. Now, we have

$$\begin{aligned}
R_E(\theta) &= -\lim_{t\to\infty}\frac{1}{\theta t}\log_e \mathbb{E}\{e^{-\theta S[t]}\} && (12)\\
&= -\lim_{t\to\infty}\frac{1}{\theta t}\log_e \mathbb{E}\{e^{-\theta\sum_{i=1}^{t}R_i}\} && (13)\\
&= -\lim_{t\to\infty}\frac{1}{\theta t}\log_e \mathbb{E}\left\{\prod_{i=1}^{t}e^{-\theta R_i}\right\} && (14)\\
&= -\lim_{t\to\infty}\frac{1}{\theta t}\log_e \prod_{i=1}^{t}\mathbb{E}\left\{e^{-\theta R_i}\right\} && (15)\\
&= -\lim_{t\to\infty}\frac{1}{\theta t}\log_e \left(\mathbb{E}\left\{e^{-\theta R_i}\right\}\right)^t && (16)\\
&= -\lim_{t\to\infty}\frac{1}{\theta t}t\log_e \mathbb{E}\left\{e^{-\theta R_i}\right\} && (17)\\
&= -\frac{1}{\theta}\log_e \mathbb{E}\left\{e^{-\theta R_i}\right\} && (18)\\
&= -\frac{1}{\theta}\log_e \mathbb{E}_{|h|^2}\left\{\epsilon + (1-\epsilon)e^{-\theta m\bar{r}_i}\right\} && (19)
\end{aligned}$$

Above, (15) follows from the independence of the service process and (16) is due to its being identically distributed. The expression inside the expectation in (19) is obtained by evaluating the expected value of $e^{-\theta R_i}$ for fixed $|h|^2$. Finally, (11) is obtained by normalizing (19) by m to have the effective rate in the units of bits per channel use, and by dropping the time index i. □

Note that the effective rate is a function of the QoS exponent θ, blocklength m, signal-to-noise ratio SNR and error probability ϵ. Since we assume that coding is performed in each coherence interval, the blocklength m is determined by the statistics of the fading process. The value of θ can be dictated by the application requirements and SNR depends on the power budget. Given the values of these parameters, the remaining parameter ϵ can be optimized to maximize the throughput. Note that large ϵ implies that the transmitter attempts to transmit the data at a high rate but at the risk of more frequent errors and hence retransmissions. On the other hand, if ϵ is small, the instantaneous transmission rate is low but the reliability of the transmissions is high. The following result shows that the optimal ϵ is unique.

Proposition 2: Assume that the values of m, $\theta > 0$, and SNR > 0 are fixed. Then, the function

$$\Psi(\epsilon) = \mathbb{E}_{|h|^2}\left\{\epsilon + (1-\epsilon)e^{-\theta m\bar{r}}\right\} \tag{20}$$

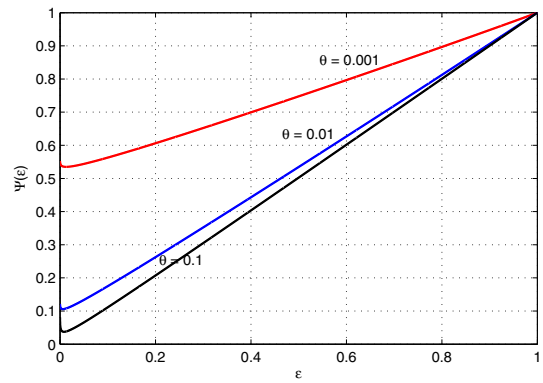

Fig. 1. The function $\Psi(\epsilon)$ vs. the error probability ϵ in the Rayleigh fading channel. SNR $= 0$ dB and the blocklength is $m = 1000$.

is strictly convex in ϵ and therefore the optimal value of ϵ that minimizes this function or equivalently maximizes the effective rate in (11) is unique.

Proof: Please see the extended journal version [12].

Note that the convexity result indicates that the optimal error probability ϵ^* can be easily found using standard convex optimization methods. The analysis and the resulting ϵ^* provide guidelines on the design of the channel codes and their strength. Note further that the above result is shown for the case in which $\theta > 0$. If there are no QoS constraints and hence $\theta = 0$, then we have the following corollary to Proposition 1.

Corollary 1: When $\theta = 0$, the effective capacity becomes

$$R_E(0) = \lim_{\theta\to 0} R_E(\theta) = (1-\epsilon)\mathbb{E}_{|h|^2}\{\bar{r}\} \tag{21}$$

where $\bar{r}$ is given in (9).

Note that the $R_E(0)$ is the average transmission rate averaged over the fading states. Below, we show that $R_E(0)$ is a strictly concave function of ϵ.

Proposition 3: Assume that the values of m, and SNR > 0 are fixed. Then, the function

$$R_E(0) = (1-\epsilon)\mathbb{E}_{|h|^2}\{\bar{r}\} \tag{22}$$

is strictly concave in ϵ and therefore the optimal value of ϵ that maximizes this effective rate is unique.

Proof: Please see [12].

Next, we provide numerical examples to illustrate the results. Although the preceding analysis is applicable to any fading distribution with finite power, we consider a Rayleigh fading channel in the numerical analysis, and assume that the fading power $z = |h|^2$ is exponentially distributed with unit mean (i.e., has the probability density function $f_z(z) = e^{-z}$).

In Figure 1, we plot $\Psi(\epsilon) = \mathbb{E}_{|h|^2}\left\{\epsilon + (1-\epsilon)e^{-\theta m\bar{r}}\right\}$ as a function of the error probability ϵ in the Rayleigh fading channel. In the figure, SNR $= 0$ dB and the blocklength $m = 1000$. We provide curves for different values of the QoS exponent $\theta > 0$. In all cases, we immediately observe the strict convexity of the curves, confirming the result in Proposition 2. Indeed, the optimal error probabilities that minimize $\Psi(\epsilon)$ are unique and are equal to $\epsilon^* = 0.0127, 0.0061, 0.0084$ for $\theta = 0.001, 0.01, 0.1,$

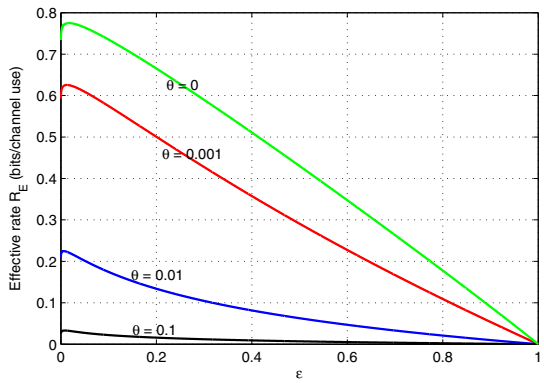

Fig. 2. Effective rate R_E vs. the error probability ϵ in the Rayleigh fading channel. SNR $= 0$ dB and the blocklength is $m = 1000$.

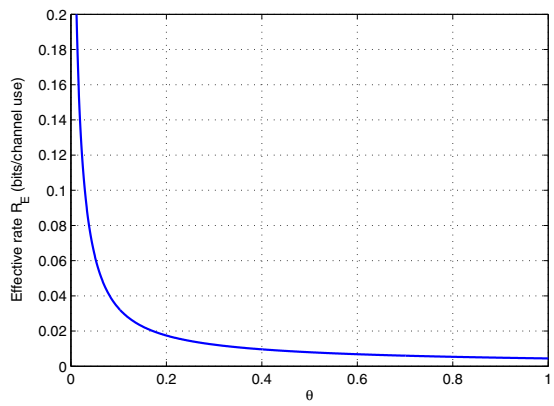

Fig. 3. The optimal effective rate R_E vs. QoS exponent θ in the Rayleigh fading channel. SNR $= 0$ dB and the blocklength is $m = 1000$.

respectively.

In Fig. 2, we plot the effective rate in (11) as a function of the error probability ϵ. The other parameters are the same as in Fig. 1. Notice that we have also included in this figure the throughput curve for the case in which $\theta = 0$. Note that if $\theta = 0$, the system does not have any queueing constraints. In Proposition 3, we have shown that $R_E(0)$ is a strictly concave function of ϵ and the optimal ϵ^* that maximizes $R_E(0)$ is unique. The strict concavity is observed in Fig. 2. The optimal value of the error probability in the case of $\theta = 0$ is $\epsilon^* = 0.0171$. For $\theta > 0$, the effective rate curves are not necessarily concave. In Fig. 2, we observe that these curves are quasiconcave and, as predicted by Proposition 1, they are maximized at a unique ϵ^*. The optimal error probabilities for the cases in which $\theta > 0$ are equal to the same ones obtained in Fig. 1. At the optimal error probabilities, the maximum effective rate values are $R_E = 0.7750, 0.6256, 0.2246, 0.0329$ bits/channel use for $\theta = 0, 0.001, 0.01, 0.1$, respectively. Note that increasing θ leads to more stringent QoS constraints, and we observe that the effective rate and hence the effective throughput diminishes as θ increases. This trend is also clearly seen in Fig. 3 where we plot the maximum effective rate values (i.e., effective rate at the optimal error probability ϵ^*) as a function of θ.

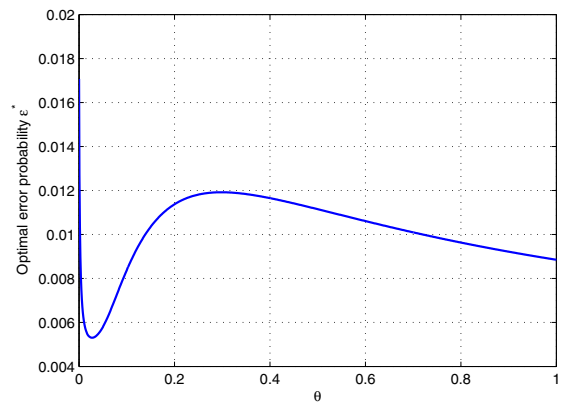

Fig. 4. The optimal error probability ϵ^* vs. QoS exponent θ in the Rayleigh fading channel. SNR $= 0$ dB and the blocklength is $m = 1000$.

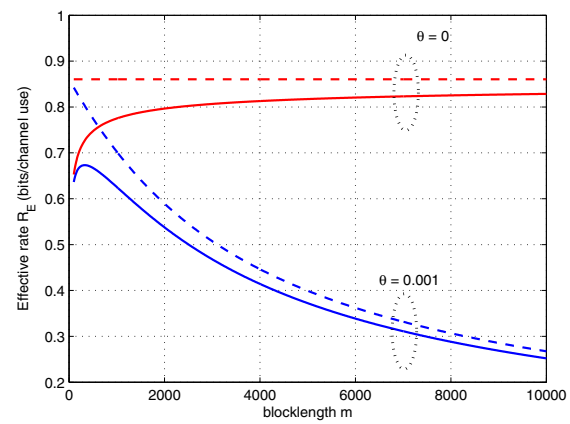

Fig. 5. The optimal effective rate R_E vs. the blocklength m in the Rayleigh fading channel. SNR $= 0$ dB and the QoS exponent is $\theta = 0.001$. Dashed curves correspond to the effective rate of the ideal model in which the service rate is equal to the instantaneous channel capacity and error probability is zero.

Another interesting analysis is the behavior of ϵ^* as a function of θ. This is depicted in Fig. 4. Here, we observe that as θ increases and therefore the QoS limitations become more stringent, the value of ϵ^* initially decreases sharply. Hence, the transmitter opts for more reliable but low-rate transmissions. On the other hand, as θ increases beyond approximately 0.028, the trend reverses and ϵ^* starts to increase. The transmitter increases the transmission rate at the cost of increased ϵ^* and hence more retransmissions. When θ exceeds 0.298, ϵ^* starts decreasing again. Note that for high values of θ, the effective rate is small. This small effective rate can be supported by low-rate transmissions. Hence, when θ is high beyond a threshold, the transmitter chooses to transmit at low rates and keep the error probability and the number of retransmissions low as well.

In Fig. 5, we plot the effective rate as a function of the blocklength m for $\theta = 0$ and $\theta = 0.001$. The solid-lined curves correspond to the effective rate in (11) optimized over ϵ. The dashed curves correspond to the effective rate of the ideal model in which the service rate is equal to the instantaneous capacity,

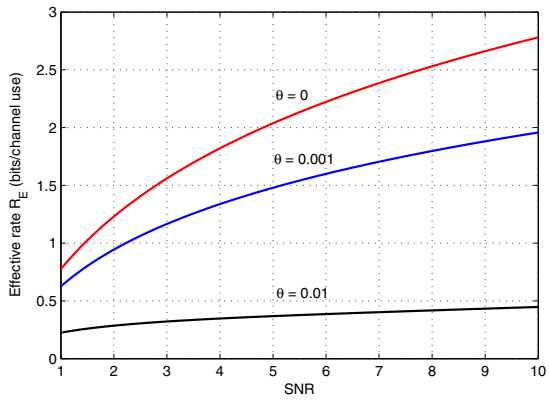

Fig. 6. The optimal effective rate R_E vs. signal-to-noise ratio (SNR) in the Rayleigh fading channel. The blocklength is $m = 1000$.

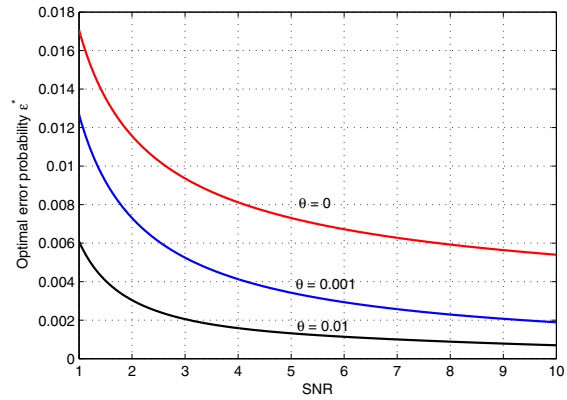

Fig. 7. The optimal error probability ϵ^* vs. signal-to-noise ratio (SNR). The blocklength is $m = 1000$.

i.e., $\bar{r} = \log_2(1 + \text{SNR}|h|^2)$, and the error probability is assumed to be zero, i.e., $\epsilon = 0$. Here, we have interesting observations. When $\theta = 0$ and the ideal model is considered, then the effective rate is $R_E(0) = \mathbb{E}_{|h|^2}\{\log_2(1 + \text{SNR}|h|^2)\}$, which is the ergodic capacity of the fading channel and is clearly independent of the blocklength. On the other hand, if the service rate is given by $\bar{r}$ in (9), the effective rate $R_E(0) = (1 - \epsilon)\mathbb{E}_{|h|^2}\{\bar{r}\}$ increases with blocklength m as seen in Fig. 5. In the presence of QoS constraints, i.e., when $\theta > 0$, we have stark differences. Under the idealistic assumption of transmitting at the instantaneous capacity with no errors, we see from the behavior of the dashed curve for $\theta = 0.001$ that effective rate decreases with increasing m. The reason is that since m is the coherence duration over which the fading state remains fixed, larger m corresponds to slower fading and slow fading is detrimental for buffer-constrained systems. In a slow-fading scenario, deep-fading can be persistent causing long durations of low rate transmissions leading to buffer overflows. In the finite blocklength regime, as seen in the behavior of the solid-lined curve of the case of $\theta = 0.001$, there is a certain tradeoff. Initially, increasing m improves the performance as this allows the system to perform transmissions with longer codewords and to have higher transmission rates. However, if m increases beyond a threshold, slowness of the fading starts to degrade the performance.

In all cases in Fig. 5, the gap between the dashed and solid-lined curves diminishes as m increases since the idealistic model becomes more accurate. On the other hand, for moderate values of m (e.g., when $m < 2000$), the idealistic assumptions lead to significant overestimations of the performance.

Finally, we provide numerical results for the optimal effective rate and optimal error probability as a function of SNR in Figs. 6 and 7, respectively, for $\theta = 0, 0.001$, and 0.01. We see that, for fixed θ, increasing the SNR improves the throughput and also the reliability of the transmissions by lowering the error probabilities.

V. Conclusion

We have analyzed the performance of buffer-constrained wireless systems in the practical scenario in which transmissions are performed using finite blocklength codes with possible decoding errors at the receiver. Employing a recent result on coding rate in the finite blocklength regime, we have determined the effective rate expression as a function of the QoS exponent, coding blocklength, decoding error probability, and signal-to-noise ratio, and characterized the throughput under statistical QoS constraints. For an extended treatment of this subject, we refer to [12] where we have additionally studied the impact of different transmission strategies such as power adaptation, fixed-rate transmissions, and sending multiple codewords in a single coherence interval.

References

[1] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol.2,no. 4, pp.630-643. July 2003

[2] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queuing networks," *IEEE Trans. Auto. Control*, vol. 39, no. 5, pp. 913-931, May 1994

[3] C.-S. Chang, *Performance Guarantees in Communication Networks*, New York: Springer, 1995

[4] C.-S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," *Proceedings of IEEE Infocom,* pp. 1001-1009, 1995

[5] J. Tang and X. Zhang, "Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp.2318-2328, June 2008.

[6] L. Liu, P. Parag, J. Tang, W.-Y. Chen and J.-F. Chamberland, "Resource allocation and quality of service evaluation for wireless communication systems using fluid models," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1767-1777, May 2007

[7] M.C. Gursoy, D. Qiao, and S. Velipasalar, "Analysis of energy efficiency in fading channel under QoS constrains," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4252-4263, Aug. 2009.

[8] D. Qiao, M.C. Gursoy, and S. Velipasalar, "The impact of QoS constraints on the energy efficiency of fixed-rate wireless transmissions," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5957-5969, Dec. 2009.

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

[10] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: How reliable should the PHY be?," Proc. of the IEEE Global Communications Conference (GLOBECOM), Dec. 2009.

[11] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[12] M.C. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," submitted. Available online at *http://arxiv.org/abs/1009.2955*.