

# On the Effective Energy Efficiency of Ultra-reliable Networks in the Finite Blocklength Regime

Mohammad Shehab, Endrit Dosti, Hirley Alves, and Matti Latva-aho  
Centre for Wireless Communications (CWC), University of Oulu, Finland  
Email: firstname.lastname@oulu.fi

**Abstract**—Effective Capacity (EC) indicates the maximum communication rate subject to a certain delay constraint while effective energy efficiency (EEE) denotes the ratio between EC and power consumption. In this paper, we analyze the EEE of ultra-reliable networks operating in the finite blocklength regime. We obtain a closed form approximation for the EEE in Rayleigh block fading channels as a function of power, error probability, and delay. We show the optimum power allocation strategy for maximizing the EEE in finite blocklength transmission which reveals that Shannon’s model underestimates the optimum power when compared to the exact finite blocklength model. Furthermore, we characterize the buffer constrained EEE maximization problem for different power consumption models. The results show that accounting for empty buffer probability (EBP) and extending the maximum delay tolerance jointly enhance the EC and EEE.

## I. INTRODUCTION

The new generation of mobile communication is expected to support a multitude of smart devices interconnected via machine-to-machine (M2M) type networks, enabling the Internet of Things (IoT). Energy efficient transmission while guaranteeing quality-of-service (QoS) is an ultimate goal in the design of 5G. QoS constraints ranging from low latency in the order of few milliseconds and packet loss rate (PLR) ( $< 10^{-3}$ ) are key requirements for *Ultra-Reliable Low Latency Communication* (URLLC) [1]–[3]. In order to boost throughput and reliability while guaranteeing low latency, it becomes crucial to investigate and optimize the resources that are allocated for transmission. In most cases, URLL application devices have limited power resources which dictates careful planning of throughput maximization with wise energy consumption models. Furthermore, the information and communication technology industry is estimated to contribute to 6% of global CO<sub>2</sub> emission by 2020 [4]. This urges the invention of low power consumption, green communication schemes.

To satisfy extremely low latency in real time applications and emerging technologies such as e-health, wireless sensor networks, and smart grids, an attractive solution is communication with short blocklength messages [1]. When the packets are short and delay requirements are stringent, performance metrics, such as Shannon capacity or outage capacity, provide a poor benchmark [5], [6]. Therefore, fundamentally new approaches are needed [1], [7]. The maximum achievable rate of finite blocklength packets is defined in [6] in terms of blocklength and error probability.

The effective capacity (EC) metric was first introduced in [8] to guarantee statistical QoS requirements by capturing the

physical and link layers aspects. EC maps the maximum arrival rate that can be supported by a network with a maximum delay bound of  $\delta$  and a delay outage probability of  $\Lambda$ . In [7], Gursoy characterized the EC in bits per channel use (bpcu) for short packets in quasi-static fading channels where the channel coefficients remain constant for the whole time spanning one packet transmission. A cognitive transmission scheme was analyzed in [9], where the target was to maximize the secondary user EC with constraints on the interference affecting the primary user. The per-node EC in massive M2M networks was studied in [10] proposing three methods to alleviate interference namely power control, graceful degradation of delay constraint and the hybrid method.

Effective energy efficiency (EEE) is defined as the ratio between EC and the total consumed power [11]. The maximization of EEE is of great importance for the upcoming massive M2M communication and thus the IoT, where the goal is to maximize the throughput for each consumed unit of power. In [12], the empty buffer probability (EBP) model was considered as an EEE booster for long packets transmission. The trade off between EEE and EC was studied in [13] where the authors suggested an algorithm to maximize the EC subject to EEE constraint. However, the probability of transmission error that appears in finite blocklength communication due to imperfect coding was not considered. She et al. showed in [14] that the relation between EEE and delay in wireless systems is not always a tradeoff. They concluded that a linear relation between service rate and power consumption leads to an EEE-delay non-tradeoff region.

In this paper, we derive a mathematical expression for the EEE in quasi-static Rayleigh fading for delay limited networks when applying linear power consumption model. We characterize the optimum power allocation strategy for the maximization of EEE. Afterwards, we resort to the power consumption model accounting for the probability of empty buffer in finite blocklength transmission, and prove that this model is valid for blocklength limited packets. Then, we emphasize that considering the probability of emptying the buffer during transmission of short packets allows for a more precise estimation of the EC and EEE which are higher when compared to the full buffer scenario. We depict how the optimum power allocation is affected by limiting the packet length and the performance gap that appears accordingly. Finally, we evaluate the impact of delay limitations on EC and EEE. We deduce that extending the allowable delay provides

gains in EC and EEE in all cases.

The rest of the paper is organized as follows: in Section II, we introduce the system model and elucidate the relation between EC and EEE. Next, Section III presents the EEE analysis in Rayleigh block fading scenario and characterizes optimum error and power allocation strategies for EEE maximization when applying the linear power consumption model. We illustrate the EBP model in finite blocklength transmission and characterize the EEE maximization with delay, error and power constraints in Section IV. The results are discussed in Section V. Finally, Section VI concludes the paper.

## II. PRELIMINARIES

Consider a point to point transmission scenario in which two nodes communicate through a Rayleigh block fading channel with blocklength  $n$ . The received vector  $\mathbf{y} \in \mathbb{C}^n$  is given by

$$\mathbf{y} = h\mathbf{x} + \mathbf{w}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{C}^n$  is the transmitted packet, and the fading coefficient is denoted by  $h$  which is assumed to be quasi-static and Rayleigh distributed. This implies that  $h$  remains constant over  $n$  symbols which span the whole packet duration.  $\mathbf{w}$  is the additive complex Gaussian noise vector whose entries are of unit variance. Furthermore, we assume that CSI is available at each node.

### A. Communication at Finite Blocklength

In finite blocklength transmission, short packets are conveyed at rate that depends not only on the SNR, but also on the blocklength and the probability of error  $\epsilon$  [5]. In this case,  $\epsilon$  has a small value but not vanishing. For error probability  $\epsilon \in [0, 1]$ , the normalized achievable rate in bpcu is given by

$$r \approx \log_2(1 + \rho|h|^2) - \frac{Q^{-1}(\epsilon) \log_2(e)}{\sqrt{(n)}} \sqrt{1 - \frac{1}{(1 + \rho|h|^2)^2}}, \quad (2)$$

where  $Q(\cdot) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$  is the Gaussian Q-function, and  $Q^{-1}(\cdot)$  represents its inverse,  $\rho$  is the SNR and  $|h|^2$  is the Rayleigh fading envelope coefficients, which is distributed as  $f_Z(z) = e^{-z}$ .

### B. The relation between Effective Capacity and Effective Energy Efficiency

The concept of EC ( $C_e$ ) indicates the capability of communication nodes to exchange data with maximum rate under statistical delay constraint. A statistical delay violation model implies that an outage occurs when a packet delay exceeds a maximum delay bound  $\delta$  and the outage probability is defined as [8]

$$\Lambda = \Pr(\text{delay} \geq \delta) \approx e^{-\theta \cdot C_e \cdot \delta}, \quad (3)$$

where  $\Pr(\cdot)$  denotes the probability of a certain event. Conventionally, the tolerance of a system to long delays is measured by the delay exponent  $\theta$ . The system tolerates large delays for small values of  $\theta$  (i.e.,  $\theta \rightarrow 0$ ), and it becomes stricter

delay-wise for large values of  $\theta$ . For instance, a network with unity EC and an outage probability  $\Lambda = 10^{-3}$  can tolerate a maximum delay of  $\delta = 691$  symbol periods for  $\theta = 0.01$  and  $\delta = 23$  symbol periods when  $\theta = 0.3$ . In quasi-static fading, the channel remains constant within each transmission period  $n$  [15], and the EC in bits per channel use (bpcu) is [7]

$$C_e(\rho, \theta, \epsilon) = -\frac{\log \psi(\rho, \theta, \epsilon)}{n\theta}, \quad (4)$$

where

$$\psi(\rho, \theta, \epsilon) = \mathbb{E}_Z [\epsilon + (1 - \epsilon)e^{-n\theta r}], \quad (5)$$

and  $\log$  is the natural logarithm. In [7], the effective capacity is statistically studied for single node scenario, but never to a closed form expression. It has been proven that the EC is concave in  $\epsilon$  and hence, has a unique maximizer.

Indeed the raise of transmission power increases the transmission rate  $r$  boosting the EC. However, this comes at the cost of energy consumption, which is not feasible for energy-limited systems such as smart grids, and massive M2M type systems, which are the main concern of our analysis. Typically, these systems are isolated from stationary power sources. Therefore, it would be of high interest to study the energy consumption of these networks.

Defined as the ratio between EC and consumed power, the EEE metric indicates the network's capability of achieving a certain latency restricted rate with minimum energy consumption [16]. In this paper, we characterize the EEE and the optimum power allocation for different power models in the finite blocklength regime.

## III. LINEAR POWER CONSUMPTION MODEL

Consider the linear model in which power consumption is defined by [17]

$$P_t(\rho) = \zeta\rho + P_c, \quad (6)$$

with  $\zeta$  being the inverse drain efficiency of the transmit amplifier and  $P_c$  the hardware power dissipated in circuit. For this model, the EEE is given by

$$\eta_{ee} = \frac{-\frac{1}{n\theta} \log (\mathbb{E}_Z [\epsilon + (1 - \epsilon)e^{-n\theta r}])}{\zeta\rho + P_c}. \quad (7)$$

Notice that here the noise is normalized so that the SNR  $\rho$  frankly represents the transmit power. This scenario assumes an always full buffer and does not account for EBP. In [7], a stochastic model for EC was studied, but never to a closed form expression. Here, we present a tight approximation for the EC and hence, the EEE.

**Lemma 1.** *For a Rayleigh block fading channel with blocklength  $n$ , the EEE of the linear power consumption model is approximated as*

$$\eta_{ee}(\rho, \theta, \epsilon) \approx -\frac{\log [\epsilon + (1 - \epsilon) \mathcal{J}]}{n\theta (\zeta\rho + P_c)}, \quad (8)$$

where

$$\mathcal{J} = e^{\frac{1}{\rho}} \rho^\alpha \left[ \left( \frac{\beta^2}{2} + \beta + 1 \right) \Gamma \left( \alpha + 1, \frac{1}{\rho} \right) - \left( \frac{\beta^2}{2} + \beta \right) \frac{\Gamma \left( \alpha - 1, \frac{1}{\rho} \right)}{\rho^2} \right], \quad (9)$$

$\Gamma(\cdot, \cdot)$  is the upper incomplete gamma function [18],  $\alpha = \frac{-\theta n}{\log 2}$ ,  $\beta = \theta \sqrt{n} Q^{-1}(\epsilon) \log_2 e$ , and  $\gamma = \sqrt{(1 - \frac{1}{(1+\rho z)^2})}$ .

*Proof.* Please refer to Appendix A.  $\square$

**Lemma 2.** There is a unique local and global maximizer in  $\epsilon$  for the EEE in Rayleigh block fading channels which is given by

$$\epsilon^*(\rho, c, d) \approx \arg \min_{0 \leq \epsilon \leq 1} \epsilon + (1 - \epsilon) \mathcal{J}. \quad (10)$$

*Proof.* The expectation in (5) is shown to be convex in  $\epsilon$  in [7] independent of the distribution of channel coefficients  $Z$ . Thus, it has a unique minimizer  $\epsilon^*$  which is consequently the EC maximizer given by (10). Presuming constant transmit power and full buffer, there is a unique maximizer in  $\epsilon$  for both EC and EEE. This is because the denominator of (7) does not depend on  $\epsilon$ , thus maximizing EC, also maximizes the EEE.  $\square$

Lemma 1 provides a numerical solution for  $\epsilon^*$  which can be obtained via linear search. Note that  $\beta$  is not a function of  $z$  or  $\rho$  which simplifies the problem. The maximum effective energy efficiency  $\eta_{ee_{max}}$  can be obtained by substituting  $\epsilon^*$  into (8).

Since the logarithmic term is dominant in the rate equation given in (2), it is quite straightforward to verify that the rate function has a negative second derivative for practical SNR regions and therefore is concave in power. This firmly holds for non-extremely low SNR (i.e.  $\geq -10$  dB) regions and the mathematical proof will be shown later in a journal version. Following a similar procedure as in [13] based on [19], we can conclude that the EEE in the finite blocklength regime is also a quasi-concave function of power and strictly concave in its upper contour. Hence, the optimum power allocation for maximizing the EEE is obtained by differentiating (8) with respect to  $\rho$  as follows

$$\begin{aligned} \frac{\partial \eta_{ee}}{\partial \rho} &= - \left[ \frac{\frac{(1-\epsilon)\mathcal{J}'(\zeta\rho+P_c)}{\epsilon+(1-\epsilon)\mathcal{J}} - \zeta \log(\epsilon + (1-\epsilon)\mathcal{J})}{n\theta(\zeta\rho + P_c)^2} \right] \\ &\approx - \left[ \frac{\frac{\mathcal{J}'(\zeta\rho+P_c)}{n\theta\mathcal{J}(\zeta\rho+P_c)} - \frac{\zeta \log(\epsilon+(1-\epsilon)\mathcal{J})}{n\theta(\zeta\rho+P_c)}}{(\zeta\rho + P_c)} \right] = 0. \end{aligned} \quad (11)$$

Manipulating, we obtain the optimum power allocation  $\rho^*$  as the solution to

$$\eta_{ee}(\rho^*) = \frac{-1}{n\theta} \left( \frac{\mathcal{J}'(\rho^*)}{\mathcal{J}(\rho^*)} \right). \quad (12)$$



Fig. 1. Transmission with empty buffer probability in quasi-static channel with blocklength  $n$ .

Let  $T_1 = \frac{\beta^2}{2} + \beta + 1$  and  $T_2 = \frac{\beta^2}{2} + \beta$  in (9), then we have

$$\mathcal{J} = e^{\frac{1}{\rho}} \rho^\alpha \left( T_1 \Gamma \left( \alpha + 1, \frac{1}{\rho} \right) - \frac{T_2}{\rho^2} \Gamma \left( \alpha - 1, \frac{1}{\rho} \right) \right). \quad (13)$$

To differentiate  $\mathcal{J}$ , we apply the derivative of the upper incomplete gamma function [20] as follows

$$\begin{aligned} \mathcal{J}' &= \frac{\partial \mathcal{J}}{\partial \rho} \\ &= -\frac{1}{\rho^2} \left[ \mathcal{J} + \frac{\alpha}{\rho} \mathcal{J} - \frac{T_1}{\rho^\alpha} e^{-\frac{1}{\rho}} - \frac{2T_2}{\rho} \Gamma \left( \alpha - 1, \frac{1}{\rho} \right) + \frac{e^{-\frac{1}{\rho}}}{\rho^\alpha} \right] \\ &= -\frac{1}{\rho^2} \left[ \left( 1 + \frac{\alpha}{\rho} \right) \mathcal{J} + (1 - T_1) \frac{e^{-\frac{1}{\rho}}}{\rho^\alpha} - \frac{2T_2}{\rho} \Gamma \left( \alpha - 1, \frac{1}{\rho} \right) \right]. \end{aligned} \quad (14)$$

Although we could differentiate  $\mathcal{J}$ , a closed form solution for (12) does not exist. For this purpose, we can utilize Matlab root-finding functions, e.g., `fzero` or plotting in a similar way to [13].

#### IV. EMPTY BUFFER PROBABILITY MODEL

In this section, we discuss the EEE in the finite blocklength regime for the power model which considers the EBP. Previously, we assumed that the buffer is always full which practically is not always the case. In real scenarios, there would be instants in which a certain node becomes idle and therefore has no data to transmit. Thus, we need to account for the case when the buffer is empty. Accordingly, we apply the model considered in [12] to networks operating in the finite blocklength regime with non-vanishing probability of error  $\epsilon$ . We investigate the effect of EEE maximization with EBP and compare it to the case of always full buffer. After accounting for EBP, the transmission probability  $P_{nb}$  is equal to  $(1 - \text{the probability of empty buffer})$  and the transmission process appears in Fig. 1. For an arrival rate of  $\lambda$  and a stable queue, the power consumption becomes

$$P_t(\rho) = P_{nb} \zeta \rho + P_c = \frac{\lambda}{\mathbb{E}[r]} \zeta \rho + P_c, \quad (15)$$

with  $P_{nb} = \frac{\lambda}{\mathbb{E}[r]}$  denoting non-empty buffer probability (NBP). Note that  $P_{nb} = 1$  indicates that the buffer is always full. The EEE with EBP is given by

$$\eta_{ee} = \frac{-\frac{1}{n\theta} \log[\epsilon + (1-\epsilon)\mathcal{J}]}{\frac{\lambda}{\mathbb{E}[r]} \zeta \rho + P_c}, \quad (16)$$

where the numerator represents the effective capacity in the finite blocklength regime.

### A. Verifying the effective energy efficiency model with empty buffer probability in finite blocklength

The power consumption model considering the probability of empty buffer fulfills the characteristic properties of an energy efficiency function for the Shannon model [12]. According to [11], an energy efficiency function must be non-negative, must be zero when the transmit power is zero, and must tend to zero as the transmit power tends to infinity. We start by verifying that this power consumption model is valid as well for short packets transmission.

**Lemma 3.** *The EEE in (16) is zero for  $\rho = 0$  and tends to 0 when  $\rho \rightarrow \infty$ .*

*Proof.* Please refer to Appendix B.  $\square$

### B. Effective energy efficiency maximization with buffer constraints

We investigate the EEE maximization with EC, delay, and power constraints. EC should be higher than the arrival rate  $\lambda$  to guarantee a stable queue, while the transmission SNR  $\rho$  is bounded by  $\rho_{max}$ . Thus, the optimization problem is formulated as

$$\begin{aligned} \max_{\rho \geq 0, \theta \geq 0} \quad & \eta_{ee} = \frac{-\frac{1}{n\theta} \log[\epsilon + (1-\epsilon)\mathcal{J}]}{P_{nb}\zeta\rho + P_c}, \\ \text{s.t.} \quad & C_e(\rho, \theta, \epsilon) \geq \lambda \\ & P_{nb}e^{-\theta\lambda\delta} \leq \Lambda \\ & \rho \leq \rho_{max} \\ & \epsilon \leq \epsilon_t \end{aligned} \quad (17)$$

For the full buffer model, we set  $P_{nb}$  to 1. We perform a line search for  $\rho$  in the interval  $[0, \rho_{max}]$ . The optimum error probability is  $\min[\epsilon^*, \epsilon_t]$  where  $\epsilon^*$  is obtained from Lemma 1. When analyzing the empty buffer scenario, we set  $P_{nb} = \frac{\lambda}{\mathbb{E}[r]}$ . Here,  $\Lambda$  is the maximum allowed delay outage probability. In all cases, the optimal value of  $\theta$  can be obtained from the second constraint at equality as

$$\theta^*(\rho) = \frac{1}{\lambda\delta} \log \frac{P_{nb}}{\Lambda}. \quad (18)$$

## V. RESULTS AND DISCUSSION

In Fig. 2, we plot the EEE in Rayleigh block fading channel for different delay exponents using the expectation in (7) and Lemma 1. The network parameters are  $n = 500$  symbol periods and  $\epsilon = 10^{-3}$ . The figure proves the accuracy of Lemma 1. Note that the EEE declines when the delay constraint becomes more strict. Furthermore, the figure shows the concavity of the upper contour of the EEE in the transmit power and the approximation in Lemma 1 captures this concavity precisely.

For the following simulations, we fix the network parameters as follows  $\Lambda = 10^{-2}, 10^{-3}$ ,  $P_c = 0.2$  W,  $\zeta = 0.2$ ,  $\lambda = 1$ ,  $\delta = 500$  symbol periods, and  $n = 500$  symbol periods, unless stated otherwise. In Fig. 3, we evaluate the EEE as a function of error  $\epsilon$  in case of EBP and compare it to the case where the buffer is always full while fixing the transmit power at  $\rho = 10$  dB. We observe that the EEE is concave

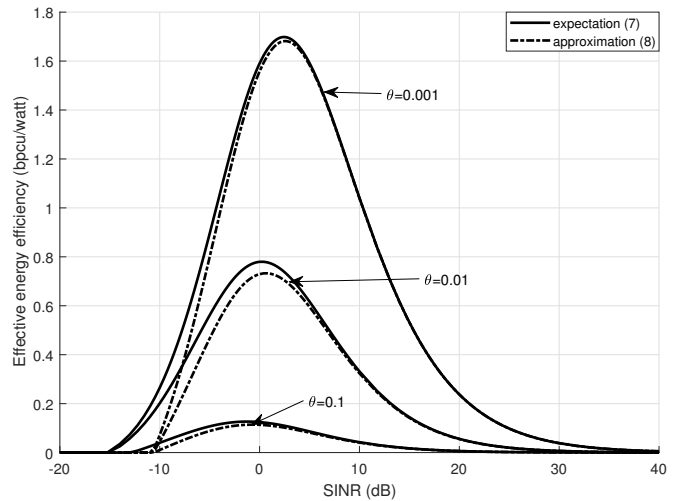


Fig. 2. Effective energy efficiency vs SNR in Rayleigh block fading for  $n = 500$ ,  $\epsilon = 10^{-3}$ ,  $P_c = 0.2$ ,  $\zeta = 0.2$ ,  $\lambda = 1$  and different delay exponents  $\theta$ .

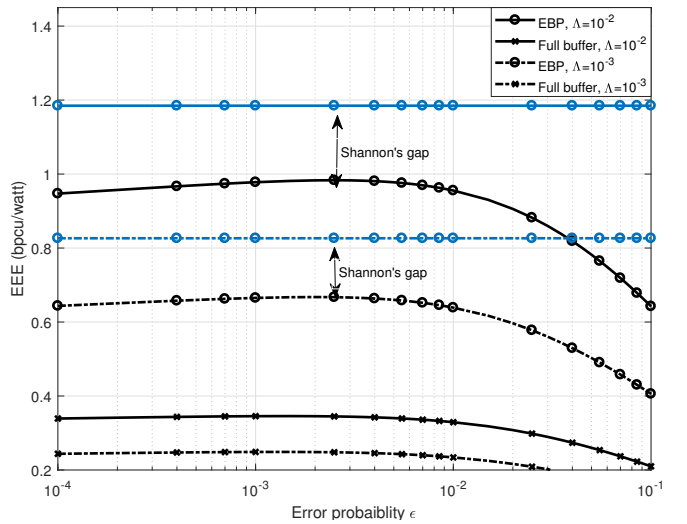


Fig. 3. EEE vs  $\epsilon$  with and without empty-buffer probability for  $\Lambda = 10^{-2}, 10^{-3}$ ,  $P_c = 0.2$ ,  $\zeta = 0.2$ ,  $\lambda = 1$ ,  $\delta = 500$ , and  $n = 500$ .

in  $\epsilon$  as stated in Lemma 2. It is obvious that considering the probability of empty buffer reflects a gain in the EEE over the full buffer model. Moreover, the figure depicts the EEE gap between the finite blocklength model and Shannon's bound when considering EBP where the Shannon's model considered in [12] overestimates the EEE by more than 20% when compared to the exact finite blocklength model.

Fig. 4 depicts the achieved maximum EEE obtained from (17) for different delay limits  $\delta$  where  $\rho_{max} = 10$  dB (variable transmission power), and  $\epsilon_t = 10^{-3}$ . We observe that the EEE increases when extending the delay  $\delta$  and relaxing the delay outage probability  $\Lambda$ . Again, the EEE is significantly higher when considering EBP. It is clear that the sporadic transmission scenario allows for a better modeling of the power consumption, thus is a more realistic model. This reflects that full buffer is the worst case, where we assume that all power will be consumed, while NBP models the fraction of

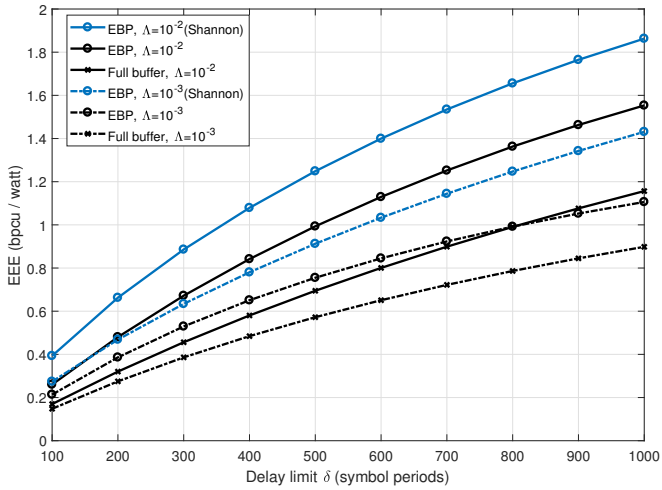


Fig. 4. EEE vs  $\delta$  with and without empty buffer probability for  $\Lambda = 10^{-2}, 10^{-3}$ ,  $P_c = 0.2$  W,  $\zeta = 0.2$ ,  $\lambda = 1$ ,  $n = 500$ , and  $\epsilon_t = 10^{-3}$ .

time that is actually used for transmission of packets according to the queue congestion which highlights the gain of this model compared to always full buffer. Furthermore, the figure verifies the inaccuracy of Shannon's model when computing the EEE for relatively small packets where the inaccuracy gap reaches 0.3 bpcu/W in higher delay region.

In Fig. 5, we plot the optimum power allocation for maximizing the EEE as a function of the maximum delay  $\delta$  in case of EBP and always full buffer where  $\rho_{max} = 10$  dB. The error outage probability is fixed at  $\epsilon = 10^{-3}$ , and the rest of the parameters are the same as in Fig. 3. The figure shows that the optimum power allocation which maximizes the EEE is significantly higher when EBP is considered. The figure also depicts that Shannon's model does not render an accurate power allocation to maximize the EEE as it underestimates the optimum power allocation when compared to the exact finite blocklength model. The power gap is ranges from 1 to 2 dB for the  $\Lambda = 10^{-2}$  as shown in the figure. Thus, we can exploit the extra power allocation that results from considering empty buffer and applying the finite blocklength model in order to boost the EC without losing energy efficiency. Moreover, the value of optimum power decays when the arrival rate  $\lambda$  declines as shown for  $\lambda = 0.3$  in the full buffer model.

Finally, we show the gain obtained in EC when considering the EBP model by plotting the EC obtained at the maximum EEE vs the maximum allowable delay  $\delta$  in Fig. 6 for the same parameters as in Fig. 5. This figure shows that considering EBP not only maximizes the EEE, but also provides a significant boost in the EC when applying the optimum power allocation strategy.

## VI. CONCLUSION

In this work, we presented a detailed analysis of the EEE for delay constrained networks in the finite blocklength regime. For Rayleigh block fading channels, we proposed an approximation for the EEE and characterized the EEE maximizers in terms of optimum error probability and power

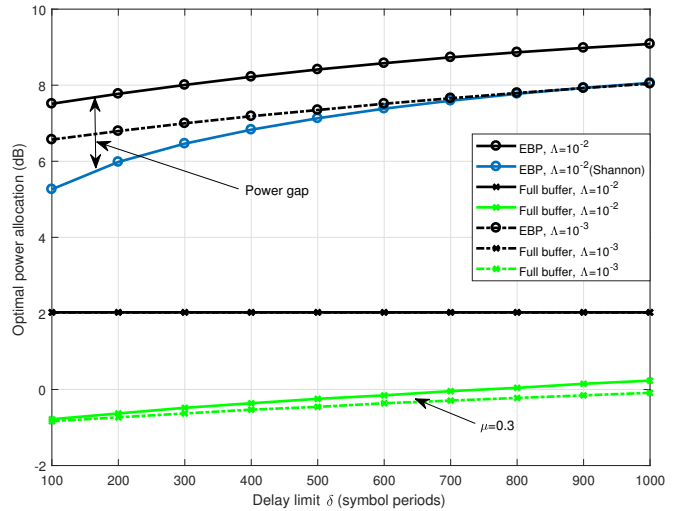


Fig. 5. Optimal power allocation vs  $\delta$  with and without empty-buffer probability for  $\Lambda = 10^{-2}, 10^{-3}$ ,  $P_c = 0.2$ ,  $\zeta = 0.2$ ,  $\lambda = 1$ ,  $\rho_{max} = 10$  dB and  $\epsilon = 10^{-3}$ .

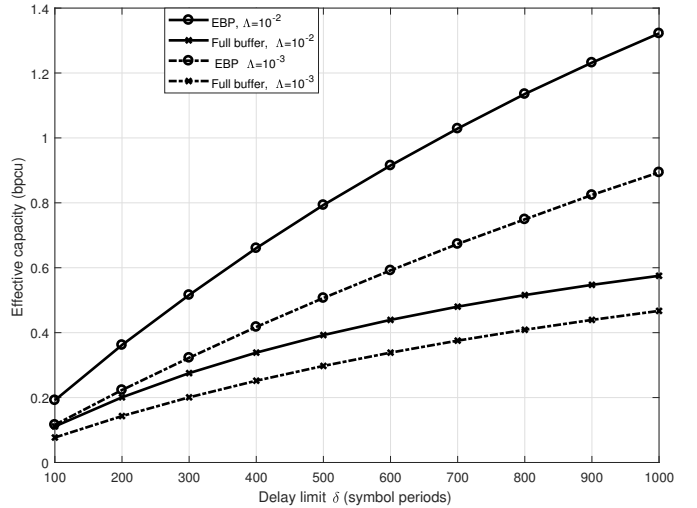


Fig. 6. Effective capacity vs  $\delta$  with and without empty-buffer probability for  $\Lambda = 10^{-2}, 10^{-3}$ ,  $P_c = 0.2$  W,  $\zeta = 0.2$ ,  $\lambda = 1$ ,  $\rho_{max} = 10$  dB and  $\epsilon = 10^{-3}$ .

allocation. We showed that the advantage of considering non-empty buffer probability and flexible transmission power is twofold since it significantly improves both the EC and EEE of networks operating in the finite blocklength regime. However, Shannon's model overestimates the EEE and underestimates the optimum power allocation when compared to the exact finite blocklength model. We investigated the trade off between EEE and maximum delay limit  $\delta$ . The results show that in order to achieve some target latency, there is a sacrifice in EC and EEE while allowing for larger delays significantly boosts the EC and EEE.

## ACKNOWLEDGMENTS

This work is partially supported by Aka Project SAFE (Grant no. 303532), and by Finnish Funding Agency for Technology and Innovation (Tekes), Bittium Wireless, Keysight

APPENDIX A  
PROOF OF LEMMA 1

Applying (5), we attain

$$\psi(\rho, \theta, \epsilon) = \int_0^\infty (\epsilon + (1 - \epsilon)e^{-\theta nr}) e^{-z} dz. \quad (19)$$

From (2), we have

$$e^{-\theta nr} = e^{-\theta n \log_2(1+\rho z)} e^{\theta \sqrt{n(1-\frac{1}{(1+\rho z)^2})} Q^{-1}(\epsilon) \log_2 e}, \quad (20)$$

$$e^{-\theta n \log_2(1+\rho z)} = (1 + \rho z)^\alpha \quad (21)$$

which leads to

$$e^{\theta \sqrt{n(1-\frac{1}{(1+\rho z)^2})} Q^{-1}(\epsilon) \log_2 e} = e^{\beta \gamma}. \quad (22)$$

We resort to the second order Taylor expansion to obtain  $e^{\beta \gamma} = 1 + (\beta \gamma) + \frac{(\beta \gamma)^2}{2}$ . It follows from (20), (21), and (22) that the expression in (19) can be written as

$$\begin{aligned} \psi(\rho, \theta, \epsilon) = & \epsilon + (1 - \epsilon) \left[ \int_0^\infty (1 + \rho z)^\alpha e^{-z} dz + \right. \\ & \left. \beta \int_0^\infty (1 + \rho z)^\alpha \gamma e^{-z} dz + \frac{\beta^2}{2} \int_0^\infty (1 + \rho z)^\alpha \gamma^2 e^{-z} dz \right]. \end{aligned} \quad (23)$$

The first integral can be written as  $e^{\frac{1}{\rho}} \rho^\alpha \Gamma\left(\alpha + 1, \frac{1}{\rho}\right)$ . By applying Laurent's expansion for  $\gamma$  [21], we obtain  $\gamma \approx 1 - \frac{1}{2(1+\rho z)^2}$ . Hence, the second and third integrals can be written as  $e^{\frac{1}{\rho}} \beta \rho^\alpha \left( \Gamma\left(\alpha + 1, \frac{1}{\rho}\right) - \frac{\Gamma(\alpha-1, \frac{1}{\rho})}{\rho^2} \right)$ , and  $e^{\frac{1}{\rho}} \frac{\beta^2}{2} \rho^\alpha \left( \Gamma\left(\alpha + 1, \frac{1}{\rho}\right) - \frac{\Gamma(\alpha-1, \frac{1}{\rho})}{\rho^2} \right)$ , respectively leading to (9).

APPENDIX B  
PROOF OF LEMMA 3

For  $\rho = 0$ , the achievable rate  $r = 0$  and the numerator of (7) becomes 0. Applying L'Hopital's rule for the denominator, we have

$$\lim_{\rho \rightarrow 0} \frac{\rho}{\mathbb{E}[r]} = \lim_{\rho \rightarrow 0} \frac{1}{\mathbb{E}\left[ z \left( \frac{1}{(1+\rho z) \log 2} - \frac{Q^{-1}(\epsilon) \log_2(e)}{\sqrt{n(1+\rho z)^3} \gamma} \right) \right]} = 0. \quad (24)$$

Thus the denominator of (7) equals to  $P_c$  yielding 0 for the EEE.

For the second condition, the numerator of (7) is upper bounded by  $-\frac{\log \epsilon}{n\theta}$ , while L'Hopital's rule for the denominator, we obtain

$$\lim_{\rho \rightarrow \infty} \frac{1}{\mathbb{E}\left[ z \left( \frac{1}{(1+\rho z) \log 2} - \frac{Q^{-1}(\epsilon) \log_2(e)}{\sqrt{n(1+\rho z)^3} \gamma} \right) \right]} = \infty. \quad (25)$$

Thus, the denominator of (7) tends to infinity which nulls the EEE. Hence, (16) holds as well under finite blocklength regime, which concludes the proof.

- [1] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 146–151.
- [2] E. Dosti, U. L. Wijewardhana, H. Alves, and M. Latva-aho, "Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length," in *IEEE ICC 2017*, Paris, France, May 2017, pp. 5168–5173.
- [3] Nokia, "5G for Mission Critical Communication: Achieve ultra-reliability and virtual zero latency," *Nokia White Paper*, 2016.
- [4] C. H. Liu and K. L. Fong, "Fundamentals of the downlink green coverage and energy efficiency in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3271–3287, Dec 2016.
- [5] G. Durisi, T. Koch, and P. Popovski, "Toward Massive, Ultrareliable, and Low-Latency Wireless Communication with Short Packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.
- [6] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-Fading Channels at Finite Blocklength," *ISWCS 2013*, vol. 9, pp. 410–413, 2013.
- [7] M. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," in *EURASIP Journal on Wireless Communications and Networking* 2013, 2013.
- [8] 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, 2003.
- [9] L. Musavian and T. Le-Ngoc, "QoS-based power allocation for cognitive radios with AMC and ARQ in Nakagami-m fading Channels," in *Trans. Emerging Tel. Tech.*, vol. 27, 2014, pp. 266–277.
- [10] M. Shehab, E. Dosti, H. Alves, and M. Latva-aho, "On the effective capacity of MTC networks in the finite blocklength regime," in *EUCNC 2017*, Oulu, Finland, Jun. 2017.
- [11] A. Zappone and E. Jorswieck, *Digital Communication over Fading Channels: A Unified Approach to Performance Analysis*. Energy efficiency in wireless networks via fractional programming theory. Found. Trends Commun. Inf. Theory., 2015, vol. 11.
- [12] M. Sinaie, A. Zappone, E. A. Jorswieck, and P. Azmi, "A Novel Power Consumption Model for Effective Energy Efficiency in Wireless Networks," *IEEE Wireless Communications Letters*, vol. 5, no. 2, pp. 152–155, 2016.
- [13] L. Musavian and Q. Ni, "Effective capacity maximization with statistical delay and effective energy efficiency requirements," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3824–3835, July 2015.
- [14] C. She and C. Yang, "Energy Efficiency and Delay in Wireless Systems: Is Their Relation Always a Tradeoff?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7215–7228, 2016.
- [15] Y. Hu, A. Schmeink, and J. Gross, "Blocklength-Limited Performance of Relaying under Quasi-Static Rayleigh Channels," *IEEE Transactions on Wireless Communications*, vol. 15, 2016.
- [16] N. Petreska, H. Al-Zubaidy, R. Knorr, and J. Gross, "Power-Minimization under Statistical Delay Constraints for Multi-Hop Wireless Industrial Networks," *arXiv*, 2016. [Online]. Available: <http://arxiv.org/abs/1608.02191>
- [17] A. Helmy, L. Musavian, and T. Le-Ngoc, "Energy-efficient power adaptation over a frequency-selective fading channel with delay and power constraints," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4529–4541, September 2013.
- [18] S. I. Abramowitz M, *Handbook of mathematical functions*. New York: Dover, 1965.
- [19] L. Liu, Y. Yi, J. F. Chamberland, and J. Zhang, "Energy-efficient power allocation for delay-sensitive multimedia traffic over wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2038–2047, Jun 2014.
- [20] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. London: Academic Press, 1994.
- [21] K. I. R. Rodriguez and J. Gilman, *Complex Analysis: In the Spirit of Lipman Bers*. Springer, 2012.