

Stability of a TDMA Network Subject to Finite Blocklength Constraints

Christos K. Kourtellis, Constantinos Psomas and Ioannis Krikidis
Department of Electrical and Computer Engineering, University of Cyprus, Cyprus
Email: {kourtellis.christos, psomas, krikidis}@ucy.ac.cy

Abstract—Recent advances in information theory have provided a novel framework regarding finite blocklength analysis, which can be employed to address the current and future demands of communication networks. In this paper, we investigate the performance of the Time-Division Multiple-Access (TDMA) scheme with bursty traffic, subject to finite blocklength constraints. In contrast to previously reported work, where the analysis of such communication networks was performed under the infinite blocklength assumption, we develop a comprehensive framework that takes finite blocklength constraints into account. In particular, we employ the recent results in finite blocklength analysis, to prove necessary stability conditions for the overall system at the finite blocklength regime, and to identify the optimal trade-off between data length and channel blocklength. The later one is evaluated both numerically and via the proposed linear and quadratic approximations that result to closed form expressions.

Index Terms—Finite-blocklength communication, Multiple Access Channel, Queueing network stability, Q-function approximations.

I. INTRODUCTION

Future wireless services will support a wide range of applications with a high discrepancy of their performance criteria, such as, very low latency ($<1\text{ms}$), very high data rates ($>1\text{Gbps}$), and ultra-high reliability (block error probability $<10^{-8}$). The simultaneous consideration of these criteria is a key factor for many vertical markets, including, autonomous vehicles, remote healthcare, industrial automation and mission critical communications. It is crucial therefore, to have a comprehensive framework for communication networks that addresses the above demands and identifies the optimal trade-off between the performance criteria.

Although Shannon's information theory evolved over the years to include applications in a wide range of fields related to communications, such as, compression, coding and statistics, it has failed to leave its distinct mark in the field of communication networks [1]. One of the main reasons that led to this result, is the asymptotic nature of information theory which cannot sufficiently address the finite blocklength constraints in communication applications [1], [2]. Thus, up until recently

This work was supported in parts by the EU's Horizon 2020 programme under the Marie Skłodowska-Curie grant agreement No 690750, in parts by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (Project: INFRASTRUCTURES/1216/0017) and in parts by the European Regional Development Fund and the Republic of Cyprus through the Research Promotion Foundation (Project: EXCELLENCE/1216/0365).

where finite blocklength rates were characterized for various channels (see [2] and references within), the analysis of communication networks was performed under the infinite blocklength assumption [3].

The purpose of this work is two-fold. First, it aims to identify necessary conditions for the stability of the Time-Division Multiple-Access (TDMA) network, subject to finite blocklength constraints. Since classical approaches consider infinite large blocklength, both for the data and the channel codes, it is necessary to re-examine the stability of the queueing systems at the finite blocklength regime. This leads to necessary stability conditions that connect the overall rate of the system to the throughput. Further, we address the problem of throughput maximization for fixed channel blocklength, and provide the optimal trade-off between data length, k , and channel blocklength, n . Since, the Additive White Gaussian Noise (AWGN) Q-function is encountered in the throughput optimization problem, no closed form solution regarding the optimal throughput can emerge. Therefore, throughput optimization is performed both numerically and via the proposed linear and quadratic approximations, that result to closed form expressions. Overall, this work delivers insight into the fundamental aspects of the finite blocklength communication networks, and poses interesting questions for future work.

The remainder of this paper is organized as follows. Section II briefly reviews the recent results in finite blocklength analysis and describes the system model and the underlying assumption. Section III proves the stability conditions for the overall queueing system in the finite blocklength regime. The evaluation of the throughput as well as the optimal trade-off between data length and channel's blocklength are given in Section IV.

II. PROBLEM FORMULATION

In this section we recall some preliminaries about the finite blocklength analysis and introduce the system model.

A. Preliminaries on finite blocklength analysis

Let $R^*(n, \epsilon)$, denote the optimal rate for fixed blocklength, n , and block error probability, ϵ . While, in general, $R^*(n, \epsilon)$ is an NP-hard problem [2], [4], a recent work [5], among others, refines Strassen's normal approximation of $R^*(n, \epsilon)$ [6], and provides an attractive tight approximation for it. In particular,

they proved that for a class of channel models with positive capacity, C , $R^*(n, \epsilon)$ is given by

$$R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad (1)$$

where C is the ergodic capacity, V is the channel's dispersion, which is by definition the minimum variance of information density over all capacity achieving input distributions [5], $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q-function and $\mathcal{O}(\log n/n)$ comprises the higher order terms. Providing closed form expressions for the channel's dispersion, is perhaps the most crucial challenge regarding the evaluation of the finite blocklength rate. Over the past few years, this challenge was successfully addressed for various channels (see [2] and references within). In particular, for the AWGN channel, the channel's capacity and dispersion are given by

$$C = \frac{1}{2} \log_2(1 + \text{SNR}), \quad (2)$$

$$V = \frac{\text{SNR}}{2} \frac{\text{SNR} + 2}{(\text{SNR} + 1)^2} (\log_2 e)^2, \quad (3)$$

respectively, where SNR denotes the signal to noise ratio. The finite blocklength rate subject to equal-power constraint is then approximated by

$$R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon). \quad (4)$$

Substituting $R^*(n, \epsilon) = \frac{k}{n}$, where k denotes the length of the data packet, and solving with respect to the block error probability ϵ , we obtain $\epsilon(k, n) \approx Q\left(\frac{nC - k}{\sqrt{nV}}\right)$. The probability of successful transmission for a code of blocklength n , is the cumulative distribution function (cdf) of the normal distribution, and it is expressed as

$$P_c(k, n) = 1 - \epsilon(k, n) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{nC - k}{\sqrt{nV}}} e^{-\frac{z^2}{2}} dz. \quad (5)$$

A recent work [7], refined the approximation given in (4), by providing the third order term in the normal approximation for the AWGN channel, that resulted in the following expressions.

$$R^*(n, \epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{\log_2(n)}{2n}, \quad (6)$$

$$P_c(k, n) \approx 1 - Q\left(\frac{nC - k + 0.5 \log_2 n}{\sqrt{nV}}\right). \quad (7)$$

B. Model description

We consider a model with two source terminals, A and B , with infinite buffer memories, and a single destination node D , as depicted in Fig. 1. At each time slot, data packets of length k_i , $i \in \{A, B\}$, arrive at the source terminal $i \in \{A, B\}$, according to a Bernoulli distribution with probability p_i . The expected value of arrivals at each time slot is $\lambda_i = p_i$, $\forall i \in \{A, B\}$. The terminals then encode the data packet into a codeword of length n , and access the channel through a TDMA scheduling with probability ω_i , where $0 \leq \omega_i \leq 1$,

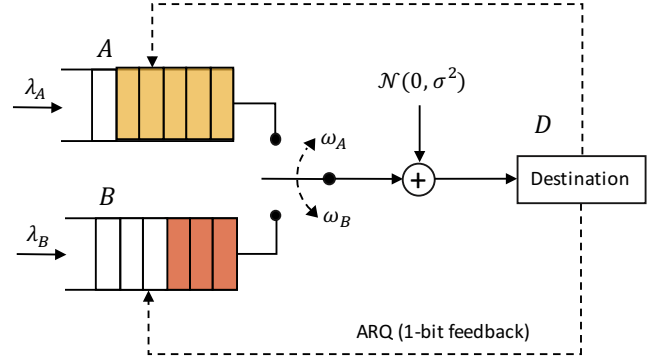


Fig. 1: Model of a TDMA network with ACK/NACK feedback.

and $\omega_A + \omega_B = 1$ [3]. We assume that at each time slot, n channel uses are employed and solely allocated to source terminal i , with probability ω_i . The channel is an AWGN channel with zero mean and variance σ^2 . The destination, after receiving and decoding the codeword, sends Acknowledgement/ Negative-Acknowledgement (ACK/NACK) back to the respective source terminal, to inform it about the status of the transmission. In the case of a correct transmission, the respective source terminal discards the data packet from its buffer memory. In the opposite case, the data packet remains in the buffer memory and waits for the next available time slot for retransmission.

The probability of an erroneous transmission for a packet, generated by terminal i at a given time slot, is denoted by $P_e(k_i, n)$. The service (departure) process is Bernoulli distributed with probability $q_i = \omega_i(1 - P_e(k_i, n))$. Since both the arrivals and departures are Bernoulli distributed, the time of an arrival and the time for a departure to occur, measured in slots, is characterized by a geometric distribution. The system at each terminal $i \in \{A, B\}$ can be described by a discrete time Markov process with states $\{S_j, j \geq 0\}$, which denote the number of packets in the system.

III. STABILITY FOR THE NON COOPERATIVE SCHEME ON THE FINITE BLOCKLENGTH REGIME

In this section, we characterize the stability region and evaluate the performance of the TDMA scheme, in the finite blocklength regime. Moreover, we evaluate the optimal throughput and the trade off between data size and blocklength, both numerically and via the proposed approximations.

Our first objective is to study the maximum rate that can be supported by the network. Towards this direction, we prove that network stability is possible, if and only if, the overall rate of the system is less than the throughput.

Theorem 1. *The TDMA network is stable, if and only if, the following conditions hold*

$$\lambda_i < \omega_i P_c(k_i, n), \quad \forall i \in \{A, B\}, \quad (8)$$

$$\lambda_A + \lambda_B < \omega_A P_c(k_A, n) + \omega_B P_c(k_B, n). \quad (9)$$

Proof. See Appendix A. \square

The following corollary is a straightforward consequence of Theorem 1.

Corollary 1. *Let $X(k, n) \triangleq \frac{k}{n}(\lambda_A + \lambda_B)$ denote the rate of the scheme, $u(k, n) \triangleq \frac{k}{n}P_c(k, n)$ denote the overall throughput of the scheme [5], and $k_A = k_B = k$. Then,*

$$X(k, n) \triangleq (\lambda_A + \lambda_B) \frac{k}{n} < \frac{k}{n} P_c(k, n) \triangleq u(k, n). \quad (10)$$

Proof. For the special case where $k_A = k_B = k$, then $P_{c,A}(k_A, n) = P_{c,B}(k_B, n) = P_c(k, n)$, thus from (8), we have

$$\lambda_i < \omega_i P_c(k, n), \quad \forall i \in \{A, B\}. \quad (11)$$

Since $\omega_A + \omega_B = 1$, the stability for the overall scheme consisted of the two terminals A and B, is calculated, as follow

$$\lambda_A + \lambda_B < (\omega_A + \omega_B) P_c(k, n) = P_c(k, n). \quad (12)$$

Multiplying both sides of (12) with $\frac{k}{n}$, we obtain (10). \square

The assumption $k_A = k_B = k$ is imposed to keep the notation clean. However, the general case of different data packet size can emerge directly by employing the proposed analysis.

Next, we employ Corollary 1, to recast the classical problem of maximizing the overall rate of the network by imposing a blocklength (latency) constraint. That is, given a channel and a fixed blocklength n , we ask what is the optimal size of the data packets that maximizes the rate. For the rest of this work, we will consider the case where the size of the data arriving at the two terminals is identical, that is, $k_A = k_B = k$, investigate the impact of the blocklength, n , on the throughput, and provide numerical evaluation and closed-form approximations for the throughput. As proved in Corollary 1, the overall rate that guarantees stability can be arbitrary close to the throughput of the system. Thus, to maximize rate we need to identify the optimal value of k that maximizes $u(k, n)$. The resulted optimization problem is given by

$$u^*(k, n) = \max_k \frac{k}{n} P_c(k, n). \quad (13)$$

Before we proceed to the solution of the above optimization problem, we investigate the convexity properties of the objective function $u(k, n)$. Towards this direction, we state the necessary definition of log-concavity and a lemma which highlights an important property of log-concave functions.

Definition 1. *A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is log-concave if $f(x) > 0 \forall x$, and $\log f$ is concave.*

Lemma 1. *Log-concavity is closed under multiplication, that is, if f and g are log-concave, the pointwise product is also log-concave [8, Section 3.5].*

We now state the theorem regarding the log-concavity of the objective function $u(k, n)$.

Theorem 2. *For any fixed $n > 1$, $u(k, n)$ is log-concave function of k .*

Proof. Let $f(k) = \frac{k}{n}$ and $h(k) = P_c(k, n)$. The objective function can be rewritten as $u(k) = f(k)h(k)$. By Definition 1, $f(k)$ is log-concave since $f(k) > 0$ and $\log f(k)$ is concave. The function $h(k)$ is by definition the cdf of a normal distribution, which is shown to be log-concave [8, Section 3.5]. Since both the functions $f(k)$ and $h(k)$ are log-concave, then by Lemma 1, the function $u(k)$ is also log-concave. \square

By virtue of Theorem 2, $u(k, n)$ is unimodal, that is, there are no local maxima that are non-global ones. This property eliminates the risk for the optimization algorithm getting trapped into a local maxima that is no global. Moreover, log-concavity allows transforming the original optimization problem into a convex optimization problem, that inherits all useful properties and tools of convex optimization.

Unfortunately, no closed form solutions can emerge from the optimization problem (13), since no explicit expression is known for $P_c(k, n)$. To overcome this problem, we capitalize the properties of the objective function, $u(k, n)$, and provide numerical evaluation of the optimal value of k via exhaustive search. Additionally, we propose first order and second order approximations of $P_c(k, n)$, which are applied in order to evaluate closed form approximations of the optimal data packet size, k^* , and the optimal throughput, $u^*(k, n)$, with a view to identify the optimal trade off between the optimal size of the data packet, k , and the blocklength n .

Remark 1. *In our analysis, we do not address the issue of control signals (metadata), which are necessary, inter alia, for the error detecting schemes required for the ACK/NACK protocol. Thus, the results of this work should be interpreted in the light of this consideration. This is translated as a genie aided destination [9], [10], that can identify possible errors, and requests, or does not request, data retransmission.*

IV. THROUGHPUT EVALUATION AND APPROXIMATIONS

The optimal solution of optimization problem (13) can be found via exhaustive search over all possible values of $k \geq 1$. This approach is computationally efficient due to the log-concavity of $u(k, n)$, which results to a unique global maxima.

The exhaustive search algorithm simply compares the objective function, $u(k, n)$, for successive values of k , and terminates the search when $u(k = i + 1, n) < u(k = i, n)$, $i \in [1, \infty)$. Then, the optimal solution is given by, $k^* = i$. By substituting the value of k^* in (13), we obtain the value of the throughput.

The analytical evaluation of the throughput involves the AWGN Q-function, and since it cannot be integrated in closed form, tight approximations should be employed in order to evaluate closed form expressions for the throughput and for the trade off between the size of the data, k , and the channel's blocklength, n . Despite the significant work on approximations of the Gaussian Q-function (see [11] and references within), these cannot be employed to provide closed form expressions of the throughput, due to their complex structure. Towards this direction, we propose linear and quadratic approximations on the probability of successful transmission, that result in closed form expressions.

Remark 2. *It has been observed, via numerical evaluation of the throughput, that the approximation given in (7), though tighter than (5) for relatively large blocklength, $n > 10^3$, may produce inconsistent results for very small blocklengths, $n < 10^2$ (the approximated rates are greater than channel's capacity). This observation holds especially for small values of SNR ($\text{SNR} < 1$). Thus, we employ the pessimistic expression (5) rather than (7). Nevertheless, the proposed methodology and results can be straightforwardly extended to any possible expression of $P_c(k, n)$.*

A. Linear approximation

Linear approximation, though not the tightest, is attractive since it provides simple expressions that can be physically interpreted. Recent works on topics related to finite blocklength analysis employ such approximations, for the finite blocklength analysis of the incremental redundancy Hybrid ARQ (HARQ) [12] and for full-duplex and half-duplex relaying for short packet communications [13]. Let, the linear approximation of the probability of successful transmission be denoted by $\hat{P}_c(k, n)$, and the resulting approximations of the throughput and of the data packet size be denoted by $\hat{u}(k, n)$ and \hat{k} , respectively.

The proposed linear approximation is given by

$$\hat{P}_c(k, n) = \begin{cases} 1 & \text{if } \chi \geq \delta_1, \\ \frac{1}{2\delta_1}\chi + \delta_0 & \text{if } -\delta_1 \leq \chi < \delta_1, \\ 0 & \text{if } \chi < -\delta_1, \end{cases} \quad (14)$$

where

$$\chi = \frac{nC - k}{\sqrt{nV}}. \quad (15)$$

The parameters $\{\delta_0, \delta_1\} \in \mathbb{R}$, are evaluated by minimizing the integral of the absolute error, that is

$$\{\delta_0^*, \delta_1^*\} = \arg \min_{\delta_0, \delta_1} \int_{-\infty}^{\infty} |\hat{P}_c(k, n) - P_c(k, n)| d\chi, \quad (16)$$

which results to $\delta_0 = 0.5$ and $\delta_1 = 1.545$. Then, by employing the above approximation, the optimization problem is given by

$$\begin{aligned} \hat{u}(k, n) &= \max_k \frac{k}{n} \hat{P}_c(k, n) \\ &= \max_k \begin{cases} \frac{k}{n} & \text{if } \chi \geq \delta_1, \\ \frac{k}{n} \left(\frac{1}{2\delta_1}\chi + \delta_0 \right) & \text{if } -\delta_1 \leq \chi < \delta_1. \end{cases} \end{aligned} \quad (17)$$

We first perform the optimization in the region $-\delta_1 \leq \chi < \delta_1$. By substituting χ and the value of δ_1 , we rewrite the predefined region as a function of k , that is

$$nC - 1.545\sqrt{nV} < k \leq nC + 1.545\sqrt{nV}. \quad (18)$$

The optimization problem is solved by differentiating the objective function, $\hat{u}(k, n)$, with respect to k , and verifying that the second derivative is negative. The optimal value of k is then given by

$$\hat{k}^* = 0.5 (nC + 1.545\sqrt{nV}). \quad (19)$$

The analytical calculations are omitted due to space limitations. Since the value of k must lay in the region defined by (18), the optimal value of k is valid only if

$$nC - 1.545\sqrt{nV} < \hat{k}^* \leq nC + 1.545\sqrt{nV}. \quad (20)$$

By substituting (19) in (20) and solving with respect to the blocklength n , we obtain the region of n for which the optimal solution given by (19) holds, which yields $0 \leq n < 13.905V/C^2$. For the region $\chi \geq \delta_1$, or equivalently for $k \leq nC - 1.545\sqrt{nV}$, the maximization of k/n with respect to k , occurs on the boundary, that is, $k = nC - 1.545\sqrt{nV}$, and this solution holds for $n \geq 13.905V/C^2$. Summarizing the above results, the optimal size of the data packet that resulted from the linear approximation of $P_c(k, n)$, is given by

$$\hat{k}^* = \begin{cases} nC - 1.545\sqrt{nV} & \text{if } n \geq \frac{9\delta_1^2 V}{C^2}, \\ 0.5 (nC + 1.545\sqrt{nV}) & \text{if } 0 < n < \frac{9\delta_1^2 V}{C^2}. \end{cases} \quad (21)$$

The above equation provides the optimal trade-off between data size and channel's blocklength. Note, that since the data size is integer, the optimal solution given in (21) should be rounded to the nearest integer. Since we are interested in an approximation of the throughput and not its exact calculation, the effect of the selected rounding function (i.e., round, ceil or floor) is negligible. The approximation of the throughput is then obtained by substituting the rounded value of (21) in (17).

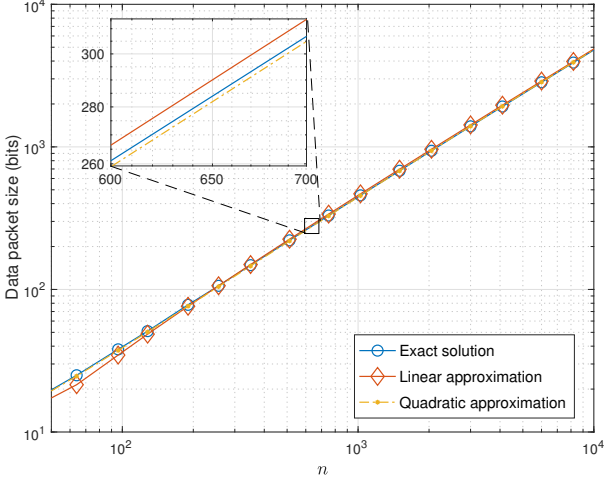


Fig. 2: Optimal size of data packets as a function of channel's blocklength, n , and comparison with the expressions resulted from the linear and quadratic approximation of $P_c(k, n)$, for $\text{SNR} = 1$.

B. Quadratic approximation

Next, we propose an approximation of $P_c(k, n)$, that, in general, gives tighter results compared to the linear approximation. This approximation is quadratic in a defined region of χ and linear in the rest of the region. Let, the quadratic approximation of the probability of successful transmission be denoted by $\tilde{P}_c(k, n)$, and the resulting approximations of the throughput and of the data packet size be denoted by $\tilde{u}(k, n)$ and \tilde{k} , respectively. Then, the proposed approximation is given by

$$\tilde{P}_c(k, n) = \begin{cases} 1 & \text{if } \chi \geq \theta_1, \\ \theta_2 \chi (2\theta_1 - \chi) + \theta_0 & \text{if } 0 \leq \chi < \theta_1, \\ \theta_2 \chi (2\theta_1 + \chi) + \theta_0 & \text{if } -\theta_1 < \chi < 0, \\ 0 & \text{if } \chi \leq -\theta_1, \end{cases} \quad (22)$$

where χ is given by (15) and $\{\theta_0, \theta_1, \theta_2\} \in \mathbb{R}$. Since, (i) the approximation given by (22) is odd-symmetric with respect to $\chi = 0$, and (ii) $P_c(k, n)|_{\chi=0} = 0.5$, then the optimal value of θ_0 that minimizes the absolute value of the error between $\tilde{P}_c(k, n)$ and $P_c(k, n)$ is, $\theta_0 = 0.5$.

Next, we evaluate the parameters $\{\theta_1, \theta_2\}$, by imposing an additional constraint regarding the continuity of the first derivative with respect to k , which significantly simplifies the optimization problem. The proposed quadratic form guarantees continuity in the region $\chi \in (-\theta_1, \theta_1)$. The conditions that ensure continuity of the first derivative in the regions $\chi \in (-\infty, -\theta_1]$ and $\chi \in [\theta_1, \infty)$, and thus for the whole region

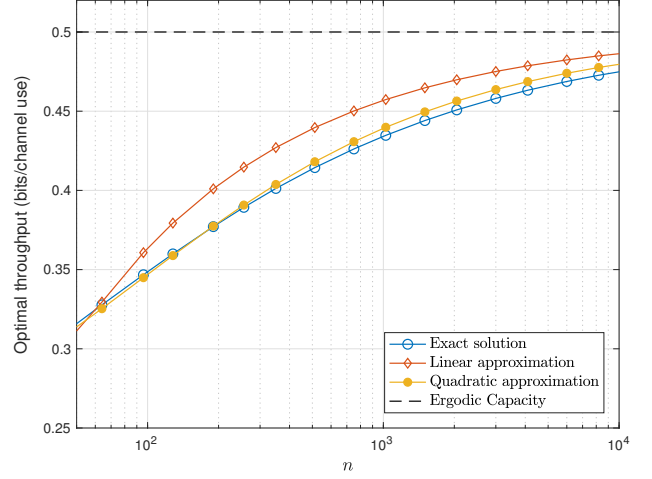


Fig. 3: Optimal throughput, $u^*(k, n)$, and comparison with the expressions resulted from the linear and quadratic approximation of $P_c(k, n)$, for $\text{SNR} = 1$.

$\chi \in (-\infty, \infty)$, are

$$\frac{d}{dk} [\tilde{P}_c(k, n)] \Big|_{\chi=\theta_1} = 0, \quad \frac{d}{dk} [\tilde{P}_c(k, n)] \Big|_{\chi=-\theta_1} = 0 \quad (23)$$

$$\tilde{P}_c(k, n) \Big|_{\chi=\theta_1} = 1, \quad \tilde{P}_c(k, n) \Big|_{\chi=-\theta_1} = 0 \quad (24)$$

Equations in (23) are satisfied directly by the proposed quadratic form, whereas equations in (24) are satisfied, if and only if, $\theta_2 = 0.5/\theta_1^2$. The remaining parameter, θ_1 , is evaluated by minimizing the integral of the absolute error

$$\{\theta_1^*\} = \arg \min_{\theta_1} \int_{-\infty}^{\infty} |\tilde{P}_c(k, n) - P_c(k, n)| d\chi, \quad (25)$$

which yields $\theta_1 = 2.35$. The optimization problem for the case of the quadratic approximation is

$$\tilde{u}(k, n) = \max_k \frac{k}{n} \tilde{P}_c(k, n), \quad (26)$$

where $\tilde{P}_c(k, n)$ is given by (22). By employing the methodology discussed in Section IV-A, we obtain the following results regarding the optimal length of the data size

$$\tilde{k}^* = \begin{cases} \frac{2}{3} (nC - \theta_1 \sqrt{nV}) + \theta_3 & \text{if } n \geq \frac{\theta_1^2 V}{4C^2}, \\ \frac{1}{3} (nC - \theta_1 \sqrt{nV}) & \text{if } 0 < n < \frac{\theta_1^2 V}{4C^2}, \end{cases} \quad (27)$$

where

$$\theta_3 = \frac{\sqrt{n}}{3} (nC^2 - 7\theta_1^2 V - 2\theta_1 C \sqrt{nV})^{\frac{1}{2}}.$$

Then, $\tilde{u}^*(k, n)$ emerges by substituting the values of (22) and (27), in $\frac{k^*}{n} \tilde{P}_c(k, n)$.

The optimal trade-off between the data packet size and the channel's blocklength, n , as well as the comparison with the

provided approximations \hat{k}^* and \tilde{k}^* , are depicted in Fig. 2. While both approximations perform well, the optimal data packet size emerged from the quadratic approximation, \tilde{k}^* , is almost identical to k^* . The optimal throughput and the throughput approximations are illustrated in Fig. 3. Again, the solution emerged from the quadratic approximation approaches very well the numerical evaluation of the optimal throughput.

Remark 3. *The results of this work can be employed in order to identify the optimal throughput of various schemes, such as, cognitive communication schemes. For example, assume that data of length k_A arrive at the terminal of the primary user (e.g. Terminal A) with rate λ_A . By (8) we can determine the minimum ω_A , denoted by ω_A^* , such that (8) holds (assuming it exists). Then, $\omega_B^* = 1 - \omega_A^*$, while the throughput of the secondary user (e.g. Terminal B) is $\omega_B^* \frac{k_B}{n} P_c(k_B, n)$. Thus, optimizing over k_B , either numerically or via the approximations, we can identify the optimal data length k_B and the optimal throughput both for the secondary user and the overall network.*

V. CONCLUSION

In this work we provide a novel framework regarding the analysis of a TDMA network with bursty traffic, in the finite blocklength regime. In particular, we examine the stability of the network, evaluate the optimal throughput for fixed latency constraints, and identify the optimal trade-off between data length and latency, both numerically and via the proposed closed form approximations. The impact of metadata is worth investigating as part of future work.

APPENDIX

A. Proof of Theorem 1.

The stability conditions of the underlying Markov chains at the two terminals depend on the existence, or non-existence, of a stationary distribution, defined by

$$\pi_{i,j} = \lim_{m \rightarrow \infty} P(S_m = j), \quad j \geq 0, i \in \{A, B\}. \quad (28)$$

The characterization of the stationary distribution for the emerged *Geo/Geo/1* queue is obtained by employing the global balance equations [14], which yield

$$\pi_{i,0} = \frac{1 - q_i}{-q_1 + \sum_{m=0}^{\infty} \left(\frac{p_i(1 - q_i)}{q_i(1 - p_i)} \right)^m} = \frac{q_i - p_i}{q_i}, \quad i \in \{A, B\}, \quad (29)$$

$$\pi_{i,j} = \left(\frac{p_i(1 - q_i)}{q_i(1 - p_i)} \right)^j \frac{1}{1 - q_i} \pi_{i,0}, \quad j \geq 1, i \in \{A, B\}. \quad (30)$$

Therefore, the stationary distribution is non-zero, only if

$$\frac{p_i(1 - q_i)}{q_i(1 - p_i)} < 1, \quad \forall i \in \{A, B\}, \quad (31)$$

or equivalently

$$q_i > p_i, \quad \forall i \in \{A, B\}. \quad (32)$$

Otherwise, $\sum_{i=0}^{\infty} (p_i(1 - q_i)/q_i(1 - p_i))$ would be infinite and all $\{\pi_{i,j}, j \geq 0\}$ would be zero. By substituting the average arrival rate, $\lambda_i = p_i$, and average departure rate $\mu_i = q_i = \omega_i(1 - P_{e,i}(k_i, n))$, in (32), we obtain the following stability condition

$$\lambda_i < \omega_i(1 - P_{e,i}(k_i, n)) \triangleq \omega_i P_{c,i}(k_i, n), \quad \forall i \in \{A, B\}. \quad (33)$$

Summing over all $i \in \{A, B\}$, we obtain (9), which completes the proof.

REFERENCES

- [1] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.
- [2] 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, Sept. 2016.
- [3] I. Krikidis, B. Rong, and A. Ephremides, "Network-level cooperation for a multiple-access channel via dynamic decode-and-forward," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7759–7770, Dec. 2011.
- [4] R. A. Costa, M. Langberg, and J. Barros, "One-shot capacity of discrete channels," in *2010 IEEE International Symposium on Information Theory*, Jun. 2010, pp. 211–215.
- [5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [6] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," 1962, pp. 689–723. [Online]. Available: <http://www.math.cornell.edu/~pmlut/strassen.pdf>
- [7] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the awgn channel," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2430–2438, May 2015.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [9] C. Steger and A. Sabharwal, "Single-input two-way simo channel: diversity-multiplexing tradeoff with two-way training," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4877–4885, Dec. 2008.
- [10] K. R. Kumar and G. Caire, "Coding and decoding for the dynamic decode and forward relay protocol," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3186–3205, Jul. 2009.
- [11] G. K. Karagiannis and A. S. Lioumpas, "An improved approximation for the gaussian q-function," *IEEE Communications Letters*, vol. 11, no. 8, pp. 644–646, Aug. 2007.
- [12] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of the incremental redundancy harq," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 529–532, Oct. 2014.
- [13] Y. Gu, H. Chen, Y. Li, and B. Vucetic, "Ultra-reliable short-packet communications: Half-duplex or full-duplex relaying?" *IEEE Wireless Communications Letters*, vol. pp, no. 99, pp. 1–1, 2017.
- [14] T. Robertazzi, *Networks and Grids: Technology and Theory*, ser. Information Technology: Transmission, Processing and Storage. Springer New York, 2007.