Energy-Efficient Power Allocation for Delay-Sensitive Multimedia Traffic Over Wireless Systems

Lingjia Liu, *Member, IEEE*, Yang Yi, *Member, IEEE*, Jean-François Chamberland, *Senior Member, IEEE*, and Jianzhong Zhang, *Senior Member, IEEE*

***Abstract*—With the extremely high penetration rate of wireless broadband devices, the telecommunication industry predicts that the mobile data traffic will double almost every year for the next five years with more than two-thirds of the future traffic being mobile video. This trend clearly indicates that future mobile broadband networks will face the dual challenge of supporting large data volumes while providing reliable service for delay-sensitive mobile multimedia streams. This explosive demand for mobile multimedia communications has significantly increased not only spectrum requirements but also energy consumption. Energy-efficient communication is becoming a prime concern for mobile multimedia communications partially due to the fact that improvements in battery technology occur on a much slower scale than gains in processing power and energy consumption of electronic devices. Furthermore, the ever-increasing demand for wireless services and ubiquitous network access comes at the expense of a growing carbon footprint for the mobile wireless communication industry. This paper addresses resource allocation to maximize the energy efficiency of a wireless link under statistical quality-of-service (QoS) constraints for mobile multimedia traffic. In the energy-efficiency analysis, both circuit and transmit power values of mobile devices are taken into account. The joint impact of statistical QoS constraints, underlying circuit power consumption, transmission power, and spectral bandwidth is considered. The unique globally optimal power allocation scheme for delay-sensitive multimedia traffic is characterized. Intuition about optimal strategies is obtained by looking at both the low-signal-to-noise-ratio (SNR) regime and the high-SNR regime.**

***Index Terms*—Delay-sensitive traffic, energy efficiency, power allocation, quality of service (QoS), wireless systems.**

I. Introduction

RARELY have technical innovations changed everyday life as rapidly and profoundly as mobile wireless communication. Over the past two decades, mobile wireless services have grown from niche market applications to globally available components of daily life. According to the International Telecommunication Union [2], the number of mobile wireless subscriptions globally has surpassed 6.83 billion in year 2013, a figure that exceeds 96% of the world population.

The driving force behind this rapid development is the growing importance of wireless connectivity for social and economic interactions. In February 2013, Cisco [3] predicted a 66% compound annual growth rate for global mobile data traffic from 2012 to 2017, indicating an unsettling growth of 13 times by year 2017. Among the traffic types responsible for this rising demand, *mobile video* features the highest growth rate. By year 2017, two-thirds of the world's mobile data traffic will be mobile video. Since a significant portion of video applications are *delay sensitive*, these statistics and predictions clearly suggest that future mobile wireless data networks will face the dual challenge of supporting large traffic volumes and providing reliable service for delay-sensitive multimedia applications such as mobile video traffic.

With the evolution toward new multimedia systems and services, the expectations of typical users are no longer confined to plain connectivity. Rather, consumers now expect services to be delivered promptly and in the high-quality format they are accustomed to. Meanwhile, audiovisual systems are becoming more and more complex, and new possibilities of presenting content are available, including augmented reality and immersive environments. However, for wireless systems, there are intrinsic characteristics that can impact user experience. For instance, the time-varying nature of wireless channels can result in perceivable impairments, originated in the different steps of the value chain from content production to display techniques. Altogether, these phenomena influence the quality of service (QoS) perceived by end users.

Most delay-sensitive multimedia applications have stringent delay requirements. Due to the time-varying nature of the underlying wireless channel, it is impossible to impose fixed delay requirements for various real-time multimedia traffic volumes. Accordingly, the statistical delay-violation probability becomes an important QoS measure of the underlying delay-sensitive multimedia traffic. The statistical delay-violation probability is defined as the probability that the delay experienced by the traffic exceeds a certain predefined threshold. For example, in the 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) Advanced standard, the upper bound on the

Manuscript received August 24, 2013; revised January 19, 2014; accepted February 27, 2014. Date of publication March 26, 2014; date of current version June 12, 2014. This work was presented in part at the IEEE International Conference on Communications (ICC'12), Ottawa, ON, Canada [1]. The work of L. Liu was supported by the National Science Foundation under Grant ECCS-1228071. The review of this paper was coordinated by Prof. Y. Fang.

L. Liu is with the Department of Electrical Engineering and Computer Science, The University of Kansas, Lawrence, KS 66045-7608 USA (e-mail: lingjialiu@ittc.ku.edu).

Y. Yi is with the Electrical Engineering and Computer Science Department, University of Kansas, Lawrence, KS, USA.

J.-F. Chamberland is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA.

J. Zhang is with the Wireless Communications Laboratory, Samsung Research America, TX 75082-4124 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2313864

delay-violation probability for online gaming is 2% with a delay threshold set to 50 ms for radio access networks [4]. That is, for online gaming traffic, more than 98% of the data should experience delay less than 50 ms. Wireless channels are prone to attenuation, fading, and interference, which negatively impact packet loss probabilities, queue-length distributions, and statistical delay-violation probabilities. Accordingly, the overall performance of wireless systems under statistical delay constraints becomes very difficult to tackle. As such, designing mobile wireless systems for delay-sensitive multimedia applications that are subject to statistical delay constraints is a challenging and pertinent task.

To address the statistical delay requirements of certain multimedia traffic, we adopt a link-layer channel model termed the *effective capacity*, which was popularized by Wu and Negi [5], [6]. An accurate estimate of the effective capacity can ensure that QoS provisioning of multimedia traffic over a wireless link can be optimal with respect to the characteristics of the underlying physical environment. The notion of effective capacity is based on the maximum constant arrival data rate that a wireless system can support when subject to a service requirement θ defined in terms of asymptotic decay rate in the occupancy of the transmit buffer. Parameter θ reflects the QoS constraint for particular delay-sensitive multimedia traffic: A larger θ results in a more stringent QoS constraint. In particular, when θ approaches zero, i.e., the QoS requirement on the link vanishes, the effective capacity then converges to the maximum throughput of the wireless channel. Furthermore, QoS metric θ is known to be closely related to the statistical delay-violation probability of a wireless link [7].

In addition to supporting delay-sensitive multimedia applications, energy efficiency is also becoming increasingly important for mobile wireless communications. As higher capacity wireless links are being created to meet the increasing demand, the power consumed by wireless infrastructures is rising quickly. Although processing power and storage capacities of wireless devices have doubled approximately every 18 months according to Moore's law [8], processor power consumption is also expanding by 150% every two years [9]. In contrast, the improvement in battery technology is much slower, improving a modest 10% every two years [9], leading to an exponentially widening gap between the demand for energy and the battery capacity offered. Furthermore, the prevailing trend of shrinking device sizes imposes an ergonomic limit on the battery capacity available. This points to the fact that energy efficiency is becoming increasingly important for mobile wireless communications. With sufficient battery power, wireless communications can be geared toward peak performance. However, with limited energy reserves on mobile devices, a well-engineered system should be able to facilitate energy conservation to minimize battery drain. Additionally, the ever-increasing demand for wireless services and ubiquitous network access comes at the expense of a growing carbon footprint for the mobile communication industry. Estimates suggest that the entire information and communication technology sector was responsible for roughly 2% of global CO_2 emissions and about 1.3% of global CO_2 equivalent (CO_2e) emissions in 2007 [10]. This study estimates the corresponding figure for mobile wireless networks to be 0.4% of the global CO_2e emissions in 2020 based on the prediction that the footprint of mobile communications could almost triple from 2007 to 2020. Thus, reducing energy consumption of mobile communication networks has an ecological impact as well.

Recently, there have been many contributions and advances in energy-efficient communication schemes [11]–[14]. It is shown in [11] that when the transmission bandwidth approaches infinity, the minimum received signal energy per bit for reliable communication over additive white Gaussian noise channels approaches -1.59 dB. However, this study does not account for additional circuit power consumed during transmission. Energy dissipation of both transmitter circuits and radio-frequency (RF) output is investigated in [12], where the modulation level is adapted to minimize the energy consumption according to the simulation observations. In [13], energy-efficient link adaptation is proposed for orthogonal frequency-division multiplexing frequency-selective channels where the transmission power and the circuit power of a communication system are both incorporated in the optimization objective. In [14], energy-efficient proportional fair scheduling is proposed for downlink multiuser multiple-input–multiple-output systems to trade off the cell-edge and cell-average energy efficiency.

In this paper, we address an energy-efficient design for delay-sensitive multimedia applications over wireless systems based on the concept of delay-sensitive energy efficiency introduced in [1]. Unlike in [15] where only transmission power is considered, we take into account circuit and transmission power values when designing the energy-efficient power allocation schemes. The proposed scheme balances the energy consumption of circuit operations and RF to achieve the maximum energy efficiency of a wireless link, which is defined as the number of bits transmitted per joule of energy under statistical QoS constraints for multimedia traffic. We demonstrate the existence of a unique globally optimal power allocation that achieves the energy-efficiency capacity and shows the intuition of the optimal power allocation in both low-SNR regime and high-SNR regime.

The outline of the paper is as follows. In Section II, we present the system model and introduce the concept of effective capacity. The effective capacity is further related to the perceived QoS constraint of delay-sensitive multimedia traffic. In Section III, we define the concept of delay-sensitive energy efficiency. This will form the optimization objective of the underlying multimedia wireless system. In Section IV, an optimal power allocation scheme to maximize the delay-sensitive energy efficiency is characterized. Intuitions about the optimal power allocation in low- and high-SNR regimes are also discussed. Numerical evaluation of the proposed strategies is also included. Section V concludes this paper.

II. System Model and Effective Capacity

Real-time applications such as video conferencing and online gaming demand low end-to-end delays. Once a delay requirement is violated, the corresponding data packet becomes useless. For wireless systems, using deterministic delay bounds is

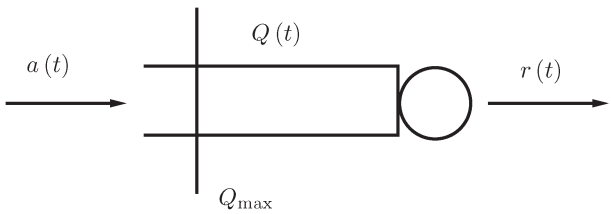

Fig. 1. Wireless queuing system model.

often impractical due to the fading nature of wireless channels. As such, we adopt a statistical QoS guarantee where the delay bound is violated with a very small probability.

Consider the single-user wireless communication system shown in Fig. 1. At time t, multimedia content arrives in the transmit buffer at rate $a(t)$ and it is served by the system at rate $r(t)$. Variable $Q(t)$ denotes the amount of information awaiting transmission at the source, and $Q_{\max}$ is a certain threshold. Fig. 1 can be regarded as an abstraction of the radio access link in a mobile broadband system. In general, the delay of the data in the radio access network is due to two major components, i.e., the queuing delay and the coding delay. The coding delay is related to the code word length, the interleaving process, and the probability of decoding failure [16]–[18]. In this paper, we restrict our attention to the queuing delay of a multimedia communication system, which amounts to the elapsed time that a packet spends in the communication system before it gets transmitted to its destination. Suppose that data arrive in the buffer at rate $a(t) = a$. At time t, the delay experienced by the data, which is about to get serviced, i.e., $D(t)$, is related to the queue length of the buffer $Q(t)$ through $Q(t) = aD(t)$. For a specified delay bound $D_{\max}$ (e.g., 50 ms for real-time video traffic in LTE Advanced systems), a service constraint may require the delay-violation probability to be no greater than a certain threshold ε, i.e., $P_{\text{delay}} \triangleq \Pr\{D(\infty) > D_{\max}\} \leq \varepsilon$, where $D(\infty)$ represents the probability distribution of the delay experienced by packets at steady state.

Theorem 1: Assume that the joint process $(Q(t), h(t))$, where $h(t)$ is the channel gain at time t is stationary and ergodic. When the system is stable

$$\Pr\{D(\infty) > D_{\max}\} \leq c\sqrt{\Pr\{Q(\infty) > Q_{\max}\}} \qquad (1)$$

where c is some positive constant, $Q(\infty)$ is the steady-state queue distribution of the buffer, and $Q_{\max} = aD_{\max}$.

Proof: See Appendix A. ■

For delay-sensitive traffic and a wireless channel, Theorem 1 connects the delay-violation probability to the buffer overflow probability. Defining the performance metric in terms of queue length rather than delay leads to a much simpler characterization. Under certain conditions on the arrival process and service process [19], the queue-length process $Q(t)$ converges in distribution to a random variable Q that satisfies

$$-\lim_{Q_{\max}\to\infty} \frac{\log \Pr\{Q \geq Q_{\max}\}}{Q_{\max}} = \theta$$

where θ is called the QoS exponent or, alternatively, the asymptotic decay rate in buffer occupancy. The result indicates that the probability of the steady-state queue length exceeding a threshold $Q_{\max}$ decays exponentially fast as $Q_{\max}$ grows larger. Informally, this relation can be written as

$$P_{\text{buffer}} \triangleq \Pr(Q(\infty) > Q_{\max}) \approx e^{-\theta Q_{\max}} \qquad (2)$$

and it holds in an asymptotic sense. For a wireless system where the value of $Q_{\max}$ is typically large, this approximation of the buffer overflow probability becomes very accurate. Taking both (1) and (2) into account, the service constraint will be met provided that

$$\theta \geq \theta_0 = -\frac{2\log(\varepsilon/c)}{aD_{\max}}. \qquad (3)$$

In (3), it is shown that a constraint on statistical delay-violation probability can be transformed into a requirement θ_0 on the decay rate of buffer occupancy. Accordingly, parameter θ_0 plays a critical role in meeting the service requirement of delay-sensitive traffic. A larger θ_0 leads to a faster decay rate of the delay-violation probability of the multimedia traffic. For instance, as θ_0 approaches zero, it indicates that the traffic can tolerate arbitrarily long delays. This is usually the case for passive data traffic. On the other hand, when θ_0 tends to infinity, it suggests that the underlying traffic cannot tolerate any delay, a situation more aligned to live streams. Altogether, the QoS constraint on the asymptotic decay rate of buffer occupancy allows us to analyze data traffic and delay-sensitive traffic in a unified manner.

QoS exponent θ is closely related to the concept of effective capacity, which is defined as the maximum constant arrival rate that a wireless system can support subject to a requirement that $\theta \geq \theta_0$ [5], [6], [20]. Mathematically, the effective capacity can be formally defined as follows. Let $r(t)$ be the instantaneous service rate of a wireless link at time t. Let $\tilde{S}(t) = \int_0^t r(\tau)d\tau$ be the service offered by the wireless link in the interval from 0 to t. Suppose that the service process is stationary and the Gärtner–Ellis limit of $\tilde{S}(t)$

$$\Lambda(-\theta) = \lim_{t\to\infty} \frac{1}{t} \log E\left[e^{-\theta \int_0^t r(\tau)d\tau}\right] = \lim_{t\to\infty} \frac{1}{t} \log E\left[e^{-\theta \tilde{S}(t)}\right]$$

exists and is differentiable for all $\theta > 0$. Then, the effective capacity of the service process is defined by

$$\alpha(\theta) = \frac{-\Lambda(-\theta)}{\theta} = -\lim_{t\to\infty} \frac{1}{\theta t} \log E\left[e^{-\theta \tilde{S}(t)}\right]. \qquad (4)$$

It can be shown [5], [21] that QoS exponent θ satisfies $\theta \geq \theta_0$ whenever the constant arrival rate a fulfills $a \leq \alpha(\theta_0)$. In other words, QoS constraint θ_0 will be fulfilled if and only if $a \leq \alpha(\theta_0)$.

Consider a block fading model for the underlying wireless link. That is, the channel coefficients stay invariant within a block of duration T and vary independently from block to block. Under this assumption, an equivalent discrete-time channel model can be constructed. Assuming that the service process is stationary and ergodic, let r be a random variable representing the system throughput during one block, the

Gärtner–Ellis limit of $\tilde{S}(t)$, and the effective capacity of (4) reduce to

$$\Lambda(-\theta) = \frac{1}{T}\log E\left[e^{-\theta r}\right] \tag{5}$$

$$\alpha(\theta) = \frac{-\Lambda(-\theta)}{\theta} = -\frac{1}{\theta T}\log E\left[e^{-\theta r}\right]. \tag{6}$$

Let W denote the spectral bandwidth; we can express the achievable system throughput during one block as [22]

$$r = WT\log\left(1 + \frac{|h|^2 P_T}{\sigma^2\Phi}\right) \tag{7}$$

where h is the channel gain of the wireless link, P_T is the transmission power, $\sigma^2 = N_0 W$ is the noise variance, and Φ is the SNR gap that defines the gap between the channel capacity and a practical modulation and coding scheme (MCS). The SNR gap depends on the MCS used and the targeted error probability. For a coded quadrature amplitude modulation system, the gap is given by [22]

$$\Phi = 9.8 + \eta_m - \eta_c \quad (\text{dB})$$

where η_m is the system design margin, and η_c is the coding gain. For Shannon capacity, it can be seen that $\Phi = 0$ dB.

III. Delay-Sensitive Energy Efficiency

For energy-efficient communication, it is desirable to maximize the amount of data sent within a given energy budget [11], rather than a time constraint. That is, given an amount of energy Δe to be consumed over a time interval of Δt, i.e., $\Delta e = P_O \Delta t$ where P_O is the overall power consumption, the transmitter wants to send a maximum amount of data. Equivalently, the transmitter will have to maximize the energy efficiency as

$$U = \frac{r \cdot \Delta t}{\Delta e} \tag{8}$$

where r is the achievable system throughput shown in (7). As discussed in [13] and [23], in addition to radiated power, a mobile device also dissipates circuit power during transmission. While the transmit power accounts for the power radiated during reliable data transmission, the circuit power represents the average energy consumption of device electronics such as mixers, filters, and digital-to-analog converters. This portion of the energy consumption excludes the needs of the power amplifier, and it is usually independent of the transmission state [12]. If we denote the circuit power as P_C, the overall power consumption of a mobile device can be expressed as

$$P_O = P_C + P(r) \tag{9}$$

where $P(r)$ is the total transmission power necessary to achieve rate r. We note that $P(r)$ is related to the transmission power P_T found in (7) through

$$P(r) = \frac{P_T}{\xi}$$

where $\xi \in [0, 1]$ is the power amplifier efficiency. This latter quantity depends on the design and implementation of the transmitter.

Under statistical QoS constraint θ_0 and given energy budget Δe consumed in small duration Δt, we can define the delay-sensitive energy efficiency for multimedia traffic as

$$U(\theta_0) = \frac{\alpha(\theta_0)\Delta t}{\Delta e}. \tag{10}$$

We call function $U(\theta_0)$ the delay-sensitive energy efficiency, which represents how much data (bits) can be transmitted through a unit of energy (in joules) under QoS constraint θ_0 for the underlying multimedia traffic. It can be verified using (10) that the delay-sensitive energy efficiency depends heavily on the underlying QoS constraint θ_0. This observation is pertinent because multimedia traffic types can have vastly different QoS constraints. For example, θ_0 can be very large for live teleconferencing to reflect the stringent QoS requirements of real-time traffic. On the other hand, θ_0 can be relatively smaller for streaming video traffic where buffering is acceptable. Taking (9) into account, we get

$$\Delta e = P_O \Delta t = (P_C + P(r))\,\Delta t = \left(P_C + \frac{P_T}{\xi}\right)\Delta t. \tag{11}$$

This equation assumes that the transmission power of P_T does not change during Δt. For a fixed QoS constraint of θ_0 for multimedia traffic, the energy efficiency of the system can be then expressed as

$$U(\theta_0) = \frac{\alpha(\theta_0)}{P_C + \frac{P_T}{\xi}}. \tag{12}$$

Assuming that the transmitter only has the knowledge related to channel statistics, the optimal power allocation strategy achieves the maximum steady-state energy efficiency under the corresponding QoS constraint, i.e.,

$$P_T^* = \arg\max_{P_T} U(\theta_0) = \arg\max_{P_T} \frac{\xi\alpha(\theta_0)}{\xi P_C + P_T}. \tag{13}$$

We note that, if the overall transmission power is fixed, the objective of maximizing the energy efficiency in (13) is equivalent to maximizing the effective capacity under the QoS requirement of θ_0. Furthermore, from (13), it is important to note that although the circuit power is not a design parameter, it indeed affects the optimal transmission power to maximize the overall system energy efficiency.

Assume that the underlying wireless link has a Rayleigh fading profile. Then, the distribution of the channel magnitude gain $|h|$ in (7) follows the Rayleigh distribution. For a single-input–single-output wireless system, the effective capacity of the corresponding wireless system is equal to [7], [24]

$$\alpha(\theta) = W\log\left(\frac{P_T}{\sigma^2\Phi}\right) - \frac{1}{\theta T} \times \left(\log\left(\Gamma\left(1 - \theta WT, \frac{\sigma^2\Phi}{P_T}\right)\right) + \frac{\sigma^2\Phi}{P_T}\right)$$

where $\Gamma(z,x)$ is the *upper incomplete gamma function*. The optimal power allocation scheme to maximize the effective capacity $\alpha(\theta_0)$ for QoS constraint θ_0 under Rayleigh fading can be found in [24]. As θ_0 approaches zero, the traffic over the wireless system becomes delay insensitive. Accordingly, the effective capacity converges to the ergodic capacity of the underlying wireless channel. In this case, the optimal power allocation scheme becomes the water-filling approach. On the other hand, when θ_0 goes to infinity, the traffic cannot tolerate any delay. Therefore, the optimal power allocation strategy converges to a strategy that effectively inverts the channel gain. Herein, we investigate optimal power allocation and link adaptation strategies that maximize the delay-sensitive energy efficiency introduced in (12).

IV. Optimal Power Allocation for Delay-Sensitive Energy Efficiency

Here, we derive the optimal power allocation and link adaptation strategies for the delay-sensitive energy efficiency introduced in Section III. More specifically, we demonstrate that a unique globally optimal power allocation always exists, and we provide necessary and sufficient conditions for a power allocation scheme to maximize its delay-sensitive energy efficiency.

Assuming that the QoS constraint for the underlying multimedia traffic is θ_0 and the transmitter only has the knowledge related to channel statistics, as shown in (13), the optimal power allocation scheme for delay-sensitive energy efficiency is given by

$$P_T^* = \arg\max_{P_T} \frac{\xi\alpha(\theta_0)}{\xi P_C + P_T} = \arg\max_{P_T} -\frac{\xi \log E[e^{-\theta_0 r}]}{\theta_0 T(\xi P_C + P_T)}.$$

As shown, increasing the transmission power of P_T will increase both the denominator and numerator. Therefore, it is not immediately clear what is the optimal power allocation strategy. Let

$$V(\theta_0, P_T) = -\frac{\xi \log E[e^{-\theta_0 r}]}{\theta_0 T(\xi P_C + P_T)}$$

where r and P_T are related through (7). For any $\theta > 0$, $V(\theta, x)$ defines a mapping from $[0, P_{\max}]$ to $[0, +\infty)$, where $P_{\max}$ is the maximum transmission power at the transmitter. Furthermore, $V(\theta, x)$ is in fact the achievable energy efficiency of the multimedia traffic under QoS constraint θ and a transmission power value of x. Suppose that, for any $x \in X$

$$Q(\theta, x) \equiv \{z \in X : V(\theta, z) \geq V(\theta, x)\} \quad \forall \theta > 0$$

denotes the *better* set of x.

Theorem 2: Let $V(\theta, x) = \mu$, it is clear that $\mu \geq 0$ (energy efficiency is nonnegative). For any value of μ, the better set of x, i.e., $Q(\theta, x)$, is strictly convex. Furthermore, the energy-efficiency function $V(\theta, x)$ is strictly quasi-concave in x. Therefore, there exists a unique globally optimal transmission power value for delay-sensitive energy efficiency in (12).

Proof Sketch: The better set of x, i.e., $Q(\theta, x)$, can be expressed as

$$\begin{aligned} Q(\theta, x) &= \{z \in [0, P_{\max}] : V(\theta, z) \geq \mu\} \\ &= \left\{ z \in [0, P_{\max}] : -\frac{\xi \log E\left[e^{-\theta W T \log\left(1 + \frac{|h|^2 z}{\sigma^2 \Phi}\right)}\right]}{\theta T(\xi P_C + z)} \geq \mu \right\} \\ &= \left\{ z \in [0, P_{\max}] : \mu(\xi P_C + z)\theta T \right. \\ &\qquad \left. + \xi \log E\left[e^{-\theta W T \log\left(1 + \frac{|h|^2 z}{\sigma^2 \Phi}\right)}\right] \leq 0 \right\}. \end{aligned} \tag{14}$$

It is clear that $r(z) = WT\log(1 + (|h|^2 z/\sigma^2\Phi)$ is a concave function of z. Accordingly, $-\theta r(z)$ is a convex function, and $e^{-\theta r(z)}$ is a log-convex function of z. Since log-convexity is preserved under summation and positive scaling [25], $E[e^{-\theta r(z)}]$ is also a log-convex function of z. This implies that $\log E[e^{-\theta r(z)}]$ is a convex function of z, and thus, $Q(\theta, x)$ is strictly convex for any value of μ. From convex optimization, we know that $V(\theta, x)$ is strictly quasi-concave in x [26] if $Q(\theta, x)$ is strictly convex. For a strictly quasi-concave function, if a local maximum exists, it is also globally optimal [26]. Hence, a uniquely global optimal power allocation scheme always exists. ■

The local optimal of $V(\theta, P_T)$ when $\theta = \theta_0$ can be obtained by setting the derivative of $V(\theta_0, P_T)$ with respect to P_T to zero. That is

$$\left.\frac{d\,V(\theta_0, P_T)}{dP_T}\right|_{P_T = P_T^*} = 0.$$

Since $V(\theta_0, P_T)$ is quasi-concave in P_T, the local optimum is also the global optimum. The derivative of $V(\theta_0, P_T)$ with respect to P_T can be expressed as

$$\begin{aligned} \frac{d\,V(\theta_0, P_T)}{d\,P_T} &= \frac{\xi}{\theta_0 T} \frac{\theta_0(\xi P_C + P_T)\frac{E\left[r' e^{-\theta_0 r}\right]}{E[e^{-\theta_0 r}]} + \log E[e^{-\theta_0 r}]}{(\xi P_C + P_T)^2} \\ &= \frac{\xi}{(\xi P_C + P_T)^2}\left(w(\theta_0, P_T) - \alpha(\theta_0, P_T)\right) \end{aligned} \tag{15}$$

where

$$w(\theta_0, P_T) = \frac{(\xi P_C + P_T) E\left[r' e^{-\theta_0 r}\right]}{T E[e^{-\theta_0 r}]}$$

$$r' = \frac{dr}{d\,P_T} = \frac{|h|^2 W T}{\sigma^2 \Phi + |h|^2 P_T}.$$

Some interesting observations can be obtained by looking at (15). It is well known that, for band-limited transmission, the lowest order modulation should be used to maximize the energy efficiency of the system [27]. This is not true for the

case where circuit power is considered. When $P_C \neq 0$, it is clear that $w(\theta_0, 0) > 0$ and $\alpha(\theta_0, 0) = 0$. This suggests that $dV(\theta_0, P_T)/dP_T|_{P_T \to 0} > 0$. Thus, the lowest order modulation/lowest transmission power is not necessarily the optimal operation strategy in this context. In general, when circuit power is nonnegligible, the optimal operation regime for energy efficiency is not the low-transmission-power regime as predicted by information theory in [11].

Let us further analyze the characteristic of the optimal power allocation strategy for the delay-sensitive energy efficiency of multimedia traffic. When $P_T^* \in [0, P_{\max}]$, the optimal power allocation for the delay-sensitive energy efficiency is the solution of

$$w(\theta_0, P_T^*) = \alpha(\theta_0, P_T^*). \tag{16}$$

Accordingly, the maximum delay-sensitive energy efficiency for the wireless system is

$$V^* = \frac{\xi\alpha(\theta_0, P_T^*)}{\xi P_C + P_T^*} = \frac{\xi E\left[r' e^{-\theta_0 r}\right]}{TE\left[e^{-\theta_0 r}\right]}. \tag{17}$$

It is important to note that solving (16) for the optimal power allocation is not straightforward. In general, various search algorithms such as binary search and gradient search [28] can be implemented to solve the implicit power allocation function. In this paper, we will not focus on introducing algorithms to numerically identify the optimal power allocation strategy; rather, we identify the characteristics of the optimal power allocation strategy in both low- and high-SNR regimes to gain intuitions about optimal delay-sensitive energy efficiency.

A. Low-SNR Regime

At low SNR, the approximation

$$\log\left(1 + \frac{|h|^2 P_T}{\sigma^2\Phi}\right) \approx \frac{|h|^2 P_T}{\sigma^2\Phi}$$

can be applied. Accordingly, the delay-sensitive energy efficiency of the wireless communication system can be expressed as

$$V(\theta_0, P_T) = \frac{\xi \log\left(1 + \frac{\theta_0 P_T W T}{\sigma^2\Phi}\right)}{\theta_0 T(\xi P_C + P_T)}.$$

Taking the first derivative of the delay-sensitive energy efficiency, we get

$$\begin{aligned}\frac{dV(\theta_0, P_T)}{dP_T} &= \frac{\xi}{\theta_0 T(\xi P_C + P_T)} \\ &\quad \times \left(\frac{\theta_0 W T}{\sigma^2\Phi + \theta_0 P_T W T} - \frac{\log\left(1 + \frac{\theta_0 P_T W T}{\sigma^2\Phi}\right)}{\xi P_C + P_T}\right) \\ &\triangleq A(\theta_0, P_T)\left(w_1(\theta_0, P_T) - w_2(\theta_0, P_T)\right) \\ &\triangleq A(\theta_0, P_T) B(\theta_0, P_T).\end{aligned} \tag{18}$$

It is shown that when the circuit power is nonnegligible, i.e., $P_C \neq 0$, $B(\theta_0, 0) > 0$. Not surprisingly, this suggests that the lowest order modulation/lowest transmission power is not the optimal power allocation scheme. Furthermore, when the circuit power is negligible, i.e., $P_C = 0$, we have

$$\begin{aligned}\frac{dV(\theta_0, P_T)}{dP_T} &= \frac{\xi}{\theta_0 T P_T} \\ &\quad \times \left(\frac{\theta_0 W T}{\sigma^2\Phi + \theta_0 P_T W T} - \frac{\log\left(1 + \frac{\theta_0 P_T W T}{\sigma^2\Phi}\right)}{P_T}\right).\end{aligned}$$

We can further characterize the derivative of the delay-sensitive energy efficiency when the transmission power is small, i.e.,

$$\begin{aligned}&\left.\frac{dV(\theta_0, P_T)}{dP_T}\right|_{P_T \to 0} \\ &= \lim_{P_T \to 0} \frac{\xi}{\theta_0 T P_T}\left(\frac{\theta_0 W T}{\sigma^2\Phi + \theta_0 P_T W T} - \frac{\theta_0 W T}{\sigma^2\Phi} \right. \\ &\qquad \left. + \frac{\theta_0^2 P_T W^2 T^2}{2\sigma^4\Phi^2}\right) \\ &= \lim_{P_T \to 0} \xi\left(\frac{\theta_0 W^2 T}{2\sigma^4\Phi^2} - \frac{\theta_0 W^2 T}{\sigma^2\Phi\left(\sigma^2\Phi + \theta_0 P_T W T\right)}\right) \\ &= \lim_{P_T \to 0} \xi\theta_0 W^2 T\left(\frac{\theta_0 P_T W T - \sigma^2\Phi}{2\sigma^4\Phi^2\left(\sigma^2\Phi + \theta_0 P_T W T\right)}\right) \\ &= -\frac{\xi\theta_0 W^2 T}{2\sigma^4\Phi^2} < 0.\end{aligned} \tag{19}$$

This result suggests that, in the low-SNR regime, the lowest order modulation/lowest transmission power is the optimal power allocation strategy for all QoS constraints when the circuit power is negligible.

In the general case where the circuit power is nonzero, the optimal power allocation scheme is the solution to

$$\frac{\theta_0 W T}{\sigma^2\Phi + \theta_0 P_T^* W T} = \frac{\log\left(1 + \frac{\theta_0 P_T^* W T}{\sigma^2\Phi}\right)}{\xi P_C + P_T^*}.$$

The maximum delay-sensitive energy efficiency can be written as

$$V^* = \frac{\xi \log\left(1 + \frac{\theta_0 P_T^* W T}{\sigma^2\Phi}\right)}{\theta_0 T\left(\xi P_C + P_T^*\right)} = \frac{\xi W}{\sigma^2\Phi + \theta_0 P_T^* W T}.$$

Thus, the optimal power allocation can be expressed as

$$P_T^* = \frac{\xi}{\theta_0 T V^*} - \frac{\sigma^2\Phi}{\theta_0 W T}. \tag{20}$$

It is shown in (20) that the optimal power allocation scheme follows a water-filling structure where the water level is determined by the optimal delay-sensitive energy efficiency and the QoS constraint θ_0.

In the case where θ_0 is small (delay insensitive traffic), the delay-sensitive energy efficiency at low SNR becomes

$$\lim_{\theta_0 \to 0} V(\theta_0, P_T) = \frac{\xi P_T W}{\sigma^2 \Phi(\xi P_C + P_T)}.$$

When the circuit power is negligible, the energy efficiency is independent of the transmission power. This is because, in the low-SNR regime, the achievable system throughput in (7) is linear in the transmission power. In the case where the circuit power is nonnegligible, i.e., $P_C \neq 0$, the optimal power allocation in this case is $P_T^* = P_{\max}$.

On the other hand, in the case where θ_0 is nonnegligible, we can evaluate the change of the delay-sensitive energy efficiency as a function of QoS constraint θ. To be specific, assume that P_T^* is the optimal transmission power that maximizes $V(\theta, P_T)$. That is, $B(\theta_0, P_T^*) = 0$. Taking the partial derivative of $A(\theta, P_T)B(\theta, P_T)$ in (18) with respect to θ, we obtain

$$\begin{aligned}\left.\frac{\partial^2 V(\theta, P_T)}{\partial\theta\partial P_T}\right|_{\theta=\theta_0, P_T=P_T^*} &= \left.\frac{\partial A(\theta, P_T)B(\theta, P_T)}{\partial\theta}\right|_{\theta=\theta_0, P_T=P_T^*} \\ &= \left.\frac{\partial A(\theta, P_T)}{\partial\theta} B(\theta, P_T)\right|_{\theta=\theta_0, P_T=P_T^*} \\ &\quad + A(\theta, P_T) \left.\frac{\partial B(\theta, P_T)}{\partial\theta}\right|_{\theta=\theta_0, P_T=P_T^*} \\ &= A(\theta, P_T) \left.\frac{\partial B(\theta, P_T)}{\partial\theta}\right|_{\theta=\theta_0, P_T=P_T^*}.\end{aligned}$$

It can be verified that $\partial B(\theta, P_T)/\partial\theta|_{\theta=\theta_0, P_T=P_T^*} < 0$, and therefore, we have

$$\left.\frac{\partial^2 V(\theta, P_T)}{\partial\theta\partial P_T}\right|_{\theta=\theta_0, P_T=P_T^*} < 0.$$

This implies that, in the low-SNR regime, the optimal transmission power for delay-sensitive energy efficiency decreases as the QoS constraint increases. In fact, this observation can be also obtained by looking at (20). When the QoS constraint becomes more stringent, transmission power has to be reduced. For the case where $P_C = 0$, it is known that the lowest order modulation/lowest transmission power is the optimal strategy for $\theta_0 = 0$. Hence, as θ_0 increases, the optimal transmission power should be reduced. This suggests that the lowest order modulation/lowest transmission power is the optimal strategy for all values of θ_0 when $P_C = 0$, which coincides with the observation we have from (19).

B. High-SNR Regime

At high SNR, the approximation

$$1 + \frac{|h|^2 P_T}{\sigma^2 \Phi} \approx \frac{|h|^2 P_T}{\sigma^2 \Phi}$$

can be applied. Accordingly, the delay-sensitive energy efficiency can be expressed as

$$V(\theta_0, P_T) = -\frac{\xi \log \int_0^\infty \left(\frac{z P_T}{\sigma^2 \Phi}\right)^{-\theta_0 W T} e^{-z} dz}{\theta_0 T (\xi P_C + P_T)}. \tag{21}$$

This implies that the energy efficiency decays to 0 as $P_T \to \infty$. The derivative of the achievable system throughput during a block with respect to P_T, r', can be approximated as

$$r' = \frac{dr}{d P_T} = \frac{|h|^2 W T}{\sigma^2 \Phi + |h|^2 P_T} \approx \frac{W T}{P_T}. \tag{22}$$

Considering both (15) and (22), we have

$$\begin{aligned}\frac{d V(\theta_0, P_T)}{d P_T} &= \frac{\xi}{(\xi P_C + P_T)^2} \left(w(\theta_0, P_T) - \alpha(\theta_0, P_T)\right) \\ &= \frac{\xi}{(\xi P_C + P_T)^2} \left(\frac{W(\xi P_C + P_T)}{P_T} - \alpha(\theta_0, P_T)\right).\end{aligned} \tag{23}$$

Note that (23) can be also obtained by taking the derivative of (21) with respect to P_T. As such, the optimal power allocation scheme is the solution of

$$\frac{W(\xi P_C + P_T^*)}{P_T^*} = \alpha(\theta_0, P_T^*).$$

Assume that P_T^* is the optimal transmission power to maximize the energy efficiency when $P_C = 0$ (i.e., when circuit power is negligible) and P_T^{C*} is the optimal transmission power when $P_C \neq 0$. It is clear that $P_T^{C*} \geq P_T^*$ since $\alpha(\theta_0, P_T^*)$ is a monotonically increasing function of P_T^*. Furthermore, the difference between P_T^{C*} and P_T^* also depends on the QoS constraint of the underlying multimedia traffic. Therefore, the circuit power and the QoS constraint of the underlying multimedia traffic have a nonnegligible impact on the optimal transmission power. In (12), we have $V^* = \xi\alpha(\theta_0, P_T^*)/\xi P_C + P_T^*$. Accordingly, the maximum delay-sensitive energy efficiency can be expressed as

$$V^* = \frac{\xi W}{P_T^*}.$$

This expression indicates that, in the high-SNR regime, the optimal energy efficiency is reversely proportional to the optimal transmission power.

In the case where θ_0 is nonnegligible, as in the low-SNR regime, we can evaluate the change of the delay-sensitive energy efficiency as a function of QoS constraint θ. Taking the partial derivative of $V(\theta, P_T)$ in (23) with respect to θ, we get

$$\frac{\partial^2 V(\theta, P_T)}{\partial\theta\partial P_T} = \frac{-\xi}{(\xi P_C + P_T)^2} \frac{\partial\alpha(\theta, P_T)}{\partial\theta}.$$

Since $\alpha(\theta, P_T)$ is a monotonically decreasing function for $\theta > 0$ [7], we have $\partial^2 V(\theta, P_T)/\partial\theta\partial P_T \geq 0$. This result suggests that, in the high-SNR regime, the optimal transmission power for delay-sensitive energy efficiency increases as the QoS constraint becomes more stringent. Furthermore, since the optimal energy efficiency is reversely proportional to the optimal transmission power, the energy efficiency of the system reduces as the QoS requirement becomes more stringent.

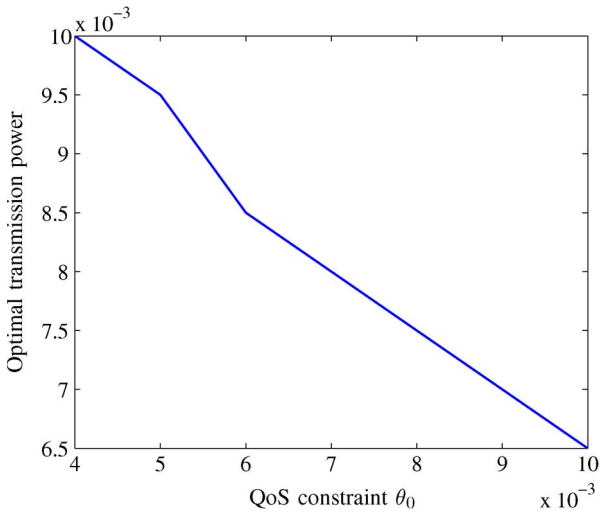

Fig. 2. Optimal transmission power as a function of QoS constraint θ_0.

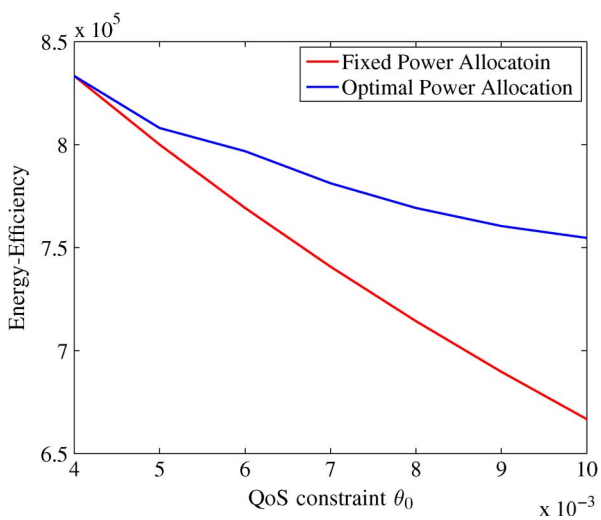

Fig. 3. Energy efficiency for different power allocation strategies.

C. Performance Evaluation

Here, we evaluate the energy-efficiency performance of delay-sensitive multimedia traffic under different power allocation strategies. For example, when $P_{\max} = 10$ mW, $P_C = 1$ mW, $W = 1$ MHz, $T = 5$ ms, $N_0 = 10^{-6}$ W/Hz, $\Phi = 0$, and $\xi = 1$, the optimal transmission power to maximize the delay-sensitive energy efficiency can be found in Fig. 2. As suggested by the analytical results in Section IV-A, the optimal transmission power decreases as the QoS constraint increases. In Fig. 3, we evaluate the energy efficiency of various power allocation strategies for the same set of system parameters. To be specific, we compare two power allocation strategies: 1) a fixed power allocation strategy where we set the power allocation to a certain level irrespective of the underlying QoS constraint for the multimedia traffic; 2) an optimal power allocation strategy where we choose different power allocation rules for different QoS constraints. The corresponding energy efficiency values of the two systems are illustrated in Fig. 3. Under fixed power allocation, we select the power allocation strategy, which is optimized for $\theta_0 = 0.004$. This is the reason why the two energy-efficiency curves start at the same point at $\theta_0 = 0.004$. Still, as the QoS requirement of the multimedia traffic becomes more stringent, the transmission power has to be reduced to optimize the energy efficiency. This is the reason why the QoS-aware power allocation significantly outperforms the fixed power allocation strategy for different QoS requirements.

V. Conclusion

In this paper, we connect the delay-violation probability to buffer overflow probability for delay-sensitive multimedia traffic. Furthermore, we develop the notion of delay-sensitive energy efficiency for multimedia traffic. Based on this concept, optimal power allocation is characterized to maximize the energy efficiency of a wireless system under statistical QoS constraints. The statistical QoS requirement and the circuit power play a significant role in the optimal transmission power allocation strategies for energy-efficient multimedia communications. In general, in the low-SNR regime, the optimal transmission power should decrease as the statistical QoS constraint increases. However, in the high-SNR regime, the optimal transmission power should increase when the statistical QoS requirement becomes more stringent. In both cases, the optimal transmission power should increase as circuit power becomes nonnegligible.

Appendix A
Delay-Violation Probability

Here, we provide a proof for Theorem 1. Suppose that the joint process of the queue length and channel, i.e., $(Q(t), h(t))$, is stationary and ergodic. Then, this process converges in distribution to a probability law μ. Let $D(\infty)$ be a random variable whose distribution coincides with the delay experienced by packets at steady state and $D_{\max}$ be the delay constraint imposed on the traffic. Similarly, let $Q(\infty)$ be a random variable whose distribution is equal to the queue-length distribution of the buffer at steady state and $Q_{\max} = aD_{\max}$ be the delay-violation threshold for the queue. Note that this relationship holds because of the constant arrival rate in the buffer. The probability that the queue length at steady state exceeds $Q_{\max}$ is

$$\Pr\{Q(\infty) > Q_{\max}\} = \int_{\mathcal{R}^+ \times \mathcal{H}} \mathbf{1}_{\{q > Q_{\max}\}} d\,\mu(q, h)$$

where $\mathbf{1}_A$ is the indicator function of set A.

At time t, the delay $D(t)$ experienced by a packet that is about to leave the buffer is related to the queue length of the buffer $Q(t)$ through $Q(t) = aD(t)$. For specific realization of the system, the empirical probability that a packet transmitted during time interval $[0, T]$ exceeds $D_{\max}$ is given by

$$\frac{\int_0^T \mathbf{1}_{\{Q(t) > Q_{\max}\}} v(t) dt}{\int_0^T v(t) dt}$$

where $v(t)$ is the instantaneous departure rate of the system at time t. Note that when the buffer is nonempty, $v(t)$ is equal

to $r(h(t))$, i.e., the instantaneous service rate of the wireless channel. Thus, the limiting delay-violation probability is equal to

$$\lim_{T\to\infty} \frac{\frac{1}{T}\int_0^T \mathbf{1}_{\{Q(t)>Q_{\max}\}} r\left(h(t)\right) dt}{\frac{1}{T}\int_0^T v(t)dt} = \lim_{T\to\infty} \frac{\frac{1}{T}\int_0^T \mathbf{1}_{\{Q(t)>Q_{\max}\}} r\left(h(t)\right) dt}{\frac{1}{T}\left(aT + Q(0) - Q(T)\right)}.$$

The stability of the system implies that

$$\lim_{T\to\infty} \frac{1}{T}\left(aT + Q(0) - Q(T)\right) = a.$$

Since the joint process $(Q(t), h(t))$ is stationary and ergodic, we can compute the delay-violation probability using the limiting distribution μ

$$\Pr\left\{D(\infty) > D_{\max}\right\} = \frac{1}{a} \int\limits_{\mathcal{R}^+\times\mathcal{H}} \mathbf{1}_{\{q>Q_{\max}\}} r(h) d\,\mu(q,h).$$

Accordingly, the delay-violation probability can be bounded as follows,

$$\begin{aligned}\Pr\left\{D(\infty) > D_{\max}\right\} &= \frac{1}{a} \int\limits_{\mathcal{R}^+\times\mathcal{H}} \mathbf{1}_{\{q>Q_{\max}\}} r(h) d\,\mu(q,h) \\ &\le \frac{1}{a}\sqrt{\int\limits_{\mathcal{R}^+\times\mathcal{H}} \mathbf{1}^2_{\{q>Q_{\max}\}} d\,\mu(q,h)}\sqrt{\int\limits_{\mathcal{R}^+\times\mathcal{H}} r^2(h) d\,\mu(q,h)} \\ &= \frac{1}{a}\sqrt{\int\limits_{\mathcal{R}^+\times\mathcal{H}} \mathbf{1}_{\{q>Q_{\max}\}} d\,\mu(q,h)}\sqrt{\int\limits_{\mathcal{R}^+\times\mathcal{H}} r^2(h) d\,\mu(q,h)} \\ &= \frac{1}{a}\sqrt{\Pr\left\{Q(\infty) > Q_{\max}\right\}}\sqrt{\int\limits_{\mathcal{R}^+\times\mathcal{H}} r^2(h) d\,\mu(q,h)}\end{aligned}$$

where the first inequality comes from the Cauchy–Schwarz inequality, and the second equality comes from the fact that $\mathbf{1}_A^2 = \mathbf{1}_A$. Let ν be the probability law for the marginal distribution of the channel in steady state, we have

$$\int\limits_{\mathcal{R}^+\times\mathcal{H}} r^2(h) d\,\mu(q,h) = \int\limits_{\mathcal{H}} r^2(h) d\,\nu(h).$$

Therefore

$$\Pr\left\{D\left((\infty\right) > D_{\max}\right\} \le c\sqrt{\Pr\left\{Q(\infty) > Q_{\max}\right\}}$$

where $c = (1/a)\sqrt{\int_{\mathcal{H}} r^2(h) d\,\nu(h)}$ is a constant independent of the queue distribution.

References

[1] L. Liu, "Energy-efficient power allocation for delay-sensitive traffic over wireless systems," in *Proc. IEEE ICC*, Jun. 2012, pp. 5901–5905.

[2] ITU-D, World Telecommunication/ICT Indicators Database 2013, Jun. 2013.

[3] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017, Cisco, Feb. 2013.

[4] "Further Advancements for E-UTRA," 3rd Generation Partnership Project (3GPP), Sophia-Antipolis Cedex, France, Tech. Rep. 36 814, Aug. 2009.

[5] 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, Jul. 2003.

[6] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1547–1557, Sep. 2004.

[7] L. Liu and J.-F. Chamberland, "On the effective capacities of multi-antenna Gaussian channels," in *Proc. IEEE ISIT*, Jul. 2008, pp. 2583–2587.

[8] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, Apr. 1965.

[9] K. Lahiri, A. Raghunathan, S. Dey, and D. Panigrahi, "Battery-driven system design: A new frontier in low power design," in *Proc. Design Autom. Conf., 7th Asia South Pacific 15th IEEE Int. Conf. VLSI Design*, Jan. 2002, pp. 261–267.

[10] J. Malmodin, A. Moberg, D. Lunden, G. Finnveden, and N. Lovehagen, "Greenhouse gas emissions and operational electricity use in the ICT and entertainment & media sectors," *J. Ind. Ecolog.*, vol. 14, no. 5, pp. 770–790, Oct. 2010.

[11] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[12] A. Y. Wang, S. Cho, C. G. Sodini, and A. P. Chandrakasan, "Energy efficient modulation and MAC for asymmetric RF microsensor system," in *Proc. IEEE Int. Symp. Low Power Electron. Design*, Aug. 2001, pp. 106–111.

[13] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[14] L. Liu, G. Miao, and J. Zhang, "Energy-efficient scheduling for downlink multi-user MIMO," in *Proc. IEEE ICC*, Jun. 2012, p. 4394.

[15] D. Qian, 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.

[16] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

[17] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

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

[19] C.-S. Chang, *Performance Guarantees in Communication Networks*. New York, NY, USA: Springer-Verlag, 2000.

[20] D. Wu and R. Negi, "Utilizing multiuser diversity for efficient support of quality of service over a fading channel," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 1198–1206, May 2005.

[21] 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. Inf. Theory*, vol. 53, no. 5, pp. 1767–1777, May 2007.

[22] J. M. Cioffi, A Multicarrier Primer, ANSI T1E1 1999.

[23] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[24] 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, no. 6, pp. 2318–2328, Jun. 2008.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[26] E. Wolfstetter, *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[27] F. Meshkati, H. V. Poor, S. C. Schwartz, and N. B. Mandayam, "An energy-efficient approach to power control and receiver design in wireless networks," *IEEE Trans. Commun.*, vol. 5, no. 1, pp. 3306–3315, Nov. 2006.

[28] J. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms (Applied Optimization)*. New York, NY, USA: Springer-Verlag, Dec. 2005.

Lingjia Liu (S'03–M'07) received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2003 and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2008.

Currently, he is an Assistant Professor with the Department of Electrical Engineering and Computer Science (EECS), The University of Kansas (KU), Lawrence, KS, USA. Prior to joining EECS at KU, he was a Senior Research Engineer with Samsung Research America where he led Samsung's research activities on multiuser multiple-input–multiple-output, coordinated multipoint transmission, and heterogeneous networks. His general research interests are in the areas of wireless communication systems, statistical signal processing, queuing theory, and information theory with emphasis on delay-sensitive applications over wireless systems.

In 2011, Dr. Liu was selected as one of the New Faces of Engineering 2011 by the Diversity Council of the National Engineers Week Foundation. He was selected as an Air Force Summer Faculty Fellow by the U.S. Air Force Office of Scientific Research in 2013. He is also a Chair or a member of the technical program committees of various international conferences and symposiums. He is currently serving as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Yang Yi (S'06–M'12) received the B.S. and M.S. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2005, respectively, and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, TX, USA, in 2009.

She is an Assistant Professor with the Department of Electrical Engineering and Computer Science, the University of Kansas, Lawrence, KS, USA. Prior to joining UMKC, she worked as a Research Engineer with Freescale, IBM, Intel, and Texas Instruments. Her research interests include 3-D integrated-circuit design and analysis and low-power circuit design with advanced nano-technologies for high-speed wireless systems.

Jean-François Chamberland (S'98–M'04–SM'09) received the B.Eng. degree from McGill University, Montreal, QC, Canada, in 1998; the M.S. degree from Cornell University, Ithaca, NY, USA, in 2000; and the Ph.D. degree from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2004, all in electrical engineering.

In 2004, he joined Texas A&M University, College Station, TX, USA, where he is currently an Associate Professor with the Department of Electrical and Computer Engineering. His research interests include communication systems, queuing theory, detection and estimation, and statistical signal processing.

Dr. Chamberland was the recipient of a Young Author Best Paper Award from the IEEE Signal Processing Society in 2006 and a Faculty Early Career Development (CAREER) Award from the National Science Foundation in 2008.

Jianzhong (Charlie) Zhang (S'96–M'02–SM'09) received the Ph.D. degree from the University of Wisconsin-Madison, Madison, WI, USA.

He is currently a Senior Director with and the Head of the Wireless Communications Laboratory, Samsung Research America, Dallas, TX, USA, where he leads technology development, prototyping, and standardization for beyond 4G and 5G wireless systems. From August 2009 to August 2013, he served as the Vice-Chairman of the 3rd Generation Partnership Project (3GPP) radio access network 1 (RAN1) working group and led the development of Long Term Evolution (LTE) and LTE Advanced technologies such as 3-D channel modeling, uplink-multiple-input–multiple-output and coordinated multipoint, and carrier aggregation for time-division-LTE. Before joining Samsung, he was with Motorola from 2006 to 2007, working on 3GPP High-Speed Packet data Access (HSPA) standards, and with Nokia Research Center from 2001 to 2006, working on IEEE 802.16e (WiMAX) standard and Enhanced Data rates for GSM Evolution (EDGE)/code-division multiple-access receiver algorithms.