

Fading Models and Metrics for Contemporary Wireless Systems

Nihar Jindal
University of Minnesota
Minneapolis, MN 55455
Email: nihar@umn.edu

Angel Lozano
Universitat Pompeu Fabra (UPF)
Barcelona 08018, Spain
Email: angel.lozano@upf.edu

Abstract—The purpose of this paper is to examine (i) some of the models commonly used to represent fading, and (ii) the information-theoretic metrics most commonly used to evaluate performance over those models. We raise the question of whether these models and metrics remain adequate in light of the advances that wireless systems have undergone over the last two decades. Weaknesses are pointed out, and ideas on possible fixes are put forth.

I. MOTIVATION

Multipath fading is one of the most essential attributes of wireless channels and, as such, the characterization of its impact on fundamental communication limits has been the object of much scrutiny (cf. [1] and references therein). This characterization is difficult because the physical mechanisms that underlie fading are intricate, and thus not easily modeled in a way that facilitates analysis and insights. Recognizing this, a few canonical settings that offer a compromise between realism and tractability have become established over time. These settings, involving models for the fading, operational assumptions, and metrics, have served the research community very well for years. The question addressed in this paper is whether these settings remain adequate in light of the rapid advances experienced by wireless systems in their characteristics and operating conditions.

II. CANONICAL SETTINGS

In the settings object of our interest, fading is modeled as a stochastic process. The marginal modeling is not problematic: the central limit theorem applied to the sum of a large number of multipath components yields a theoretical Rayleigh distribution for the amplitude, and experimental measurements have repeatedly confirmed the validity of this distribution [2].

It is the modeling of the fading dynamics that presents the most complications. From an information-theoretic vantage, what is of essence is the selectivity of the fading over each codeword and the two most common canonical settings idealize this selectivity in limiting senses:

- *Ergodic setting.* The fading varies ergodically over the span of each codeword.

- *Quasi-static setting.* The fading is fixed over the span of each codeword, and varies only from codeword to codeword.

Both these settings represent an admittedly drastic simplification of reality, deemed necessary in order to enable rigorous mathematical treatment. We next elaborate on these two settings, with emphasis on the operational assumptions that have to do with channel-state information (CSI) availability at the transmitter.

A. Ergodic Setting

In this setting, each codeword is subject to the entire distribution of the fading. This entails a high degree of selectivity, which can be modeled as continuous but more often is modeled simply as a succession of blocks [1]. While artificial, a block-fading structure further simplifies the setting while retaining its essence.

The ergodic setting has a well-defined capacity in the Shannon sense [1], the so-called “ergodic capacity”, which entails an expectation over the fading distribution. The ergodic capacity can be achieved by transmitting at some fixed rate regardless of whether the transmitter is deprived of CSI [3] or in possession of it [4].

It must be noted that—under equivalent power constraints—it is immaterial whether, in addition to time, the fading exhibits selectivity also in frequency and/or space.

While analytically convenient, the ergodic setting is frequently dismissed as inadequate on the grounds of the latency constraints of many applications. Implicit in this association is the tacit assumption that the signals are narrowband.

B. Quasi-Static Setting

In this setting, each codeword is subject to a single realization of the fading process. The model thus conforms to a block-fading structure where each codeword encompasses only one such block.

For many fading distributions, including Rayleigh, the Shannon capacity of this setting is strictly zero [1]. The relevant metric is then the outage probability, i.e., the probability that a given transmission rate is not supported because of fading [3]. In the case of an outage, the

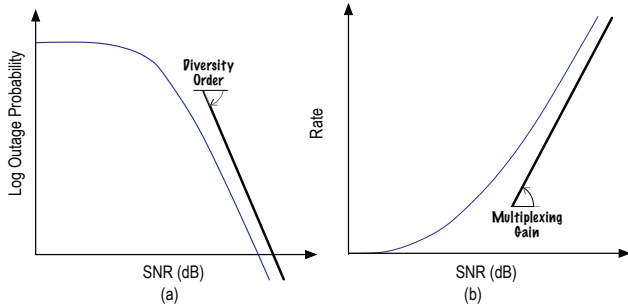


Fig. 1. (a) Diversity order as the asymptotic slope of the outage probability (in log scale) v. SNR (in dB). (b) Multiplexing gain as the asymptotic slope of the rate v. SNR (in dB).

error probability does not decay exponentially with the codeword length, whereas otherwise it does, and thus—with minimally long codewords—the errors are fully dominated by outage events and consequently we can identify the error probability and the outage probability.

Often it is more convenient to express the performance in terms of rate, and the term “outage capacity” is used to denote the highest rate supported with a given outage probability.

CSI at the transmitter is habitually not contemplated in the quasi-static setting. This setting is easily associated with indoor/pedestrian scenarios and, generally speaking, it is regarded by many as more relevant than its ergodic counterpart to modern systems and applications.

Unlike in the ergodic setting, frequency or space selectivity within a given temporal fading block can be greatly beneficial in a quasi-static setting [3]. The study of this benefit within the context of the tradeoff between transmission rate and outage probability has given rise to an entire new framework. Since the main obstacle in evaluating the tradeoff between transmission rate and outage probability lies in the lack of convenient expressions describing their interplay, proxy metrics related to the transmission rate and the outage probability have been introduced over the years and simpler, more tractable tradeoffs have been established between them. These metrics are the “diversity order” (as proxy for the outage probability) and the “multiplexing gain” (as proxy for the rate), and the tradeoff between them is the DMT (Diversity-Multiplexing Tradeoff) [5]. Denoting by SNR the average signal-to-noise ratio, the diversity order quantifies the asymptotic slope of the outage-vs-SNR relationship, in log-log scale, while the multiplexing gain quantifies the asymptotic slope of the rate-vs-SNR relationship, in log scale (cf. Fig. 1). The formulation of the DMT, which specifies the largest diversity order for a given multiplexing gain and vice versa, has spurred a large volume of work whose goal is to describe the behavior of all sorts of transmission schemes in terms of diversity-vs-multiplexing.

III. A CONTEMPORARY PERSPECTIVE

Summarizing the contents of the previous section, a vast majority of the analyses of fundamental communication limits on wireless channels rely on:

- The ergodic setting to represent fading that is fast relative to the length of a codeword.
- The quasi-static setting to represent, predominantly through the diversity and multiplexing metrics and the DMT framework, fading that is slow relative to the length of a codeword.

The transmission rate is usually fixed, which reflects well the operating conditions of early generations of wireless systems.

These canonical settings have been in use for years and are by now deeply ingrained. They are seldom questioned, but mostly just invoked and applied. Wireless systems, however, have evolved greatly since the time when these settings were defined. They have made link adaptation a norm, grown wideband, adopted packet switching and scheduling, and embraced ARQ and HARQ (hybrid ARQ), among other advances. In this section, we shall examine the canonical settings through the lens of a contemporary system design.

A. Link Adaptation

One of the defining hallmarks of modern wireless systems is their adaptivity. The transmit rate, in particular, is matched to the fading whenever timely CSI can be had at transmitter, i.e., in slow fading. This fundamentally changes the nature of the communication problem: fading-induced outages are essentially eliminated.

Adaptivity thus brings about an incongruence between slow fading and a fixed transmit rate. If the fading is slow, then link adaptation should be modeled. Hence, the canonical quasi-static setting with a fixed transmission rate can no longer represent the slow-fading regime. Rather, the correct model for this regime is an unfaded channel whose instantaneous signal-to-noise ratio is dictated by the fading distribution. If the fading is ergodic over time, then the average rate over a sufficiently long interval coincides with the expectation over such distribution, which in turn equals the ergodic capacity. Notice that here each codeword does not stretch over many fades, as in the canonical ergodic setting, but essentially over a single fading realization as illustrated using block-fading in Fig. 2. The ergodic capacity is thus a measure, in this case, of the average rate achieved by a user over many codewords. Although the significance of such average rate could be questioned from the standpoint of applications that require a minimum instantaneous rate, subsequent features later in this section will ease this concern. Paradoxically then, it is the ergodic rather than the outage capacity that emerges as an informative metric for the slow-fading regime.

Fast fading, on the other hand, precludes the availability of timely CSI at the transmitter, and thus of link

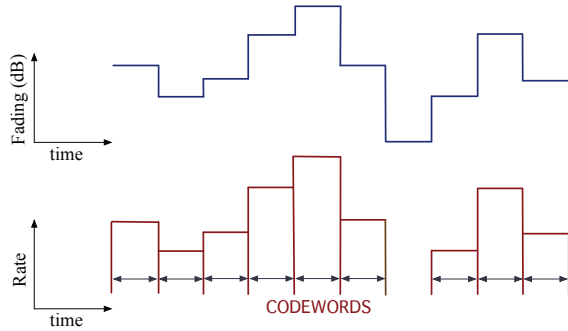


Fig. 2. Succession of codewords with rates matched to the channel via link adaptation. The average rate over time equals the ergodic capacity.

adaptation. In this case the rate can only be matched to the fading statistics, and outages do occur.

B. Hybrid ARQ

Another trait that is central to the adaptive nature of modern systems is H-ARQ, whereby the codeword length itself is made adaptive. If the initial transmission of a codeword is not decoded successfully, a one-bit notification is relayed back to the transmitter and an additional transmission containing redundant coded symbols is sent, extending the codeword and diluting its rate. The process continues into subsequent redundancy transmissions until either the decoding succeeds or a predetermined number of transmissions is reached without successful decoding, in which case an outage is declared.

The combination of link adaptation and H-ARQ allows for a finely tuned match between the rate and the fading in the slow-fading regime. In the fast-fading regime, in turn, successive transmissions tend to experience distinct realizations of the fading and thus H-ARQ better exposes the codewords to the fading distribution. Nowadays, H-ARQ is applied to almost all data including latency-sensitive traffic such as voice telephony.

C. Wideband Signaling

In contrast with older wireless systems, which were organized into narrowband channels, most modern systems are wideband. Signals occupy many MHz of bandwidth, and the trend is for them to continue growing wider. This has two immediate consequences. First, it renders frequency selectivity a property that cannot be ignored by the models. And second, it allows for long codewords without long latency. In single-carrier transmission, this benefit is reaped directly from having shorter symbols. In OFDM transmission, it is reaped by coding partly over the frequency domain (which does not incur delay).

The OFDM approach, increasingly ubiquitous, offers additional flexibility:

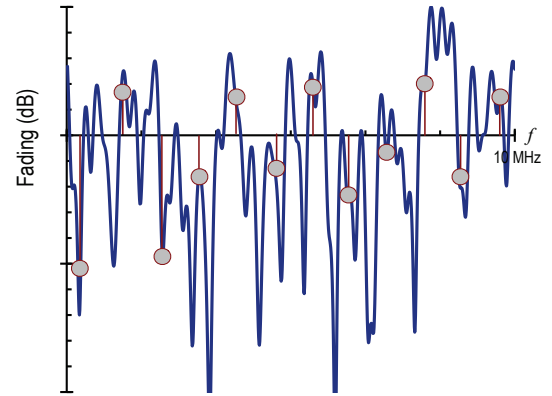


Fig. 3. Set of 12 tones constituting a basic signaling resource in LTE, interspersed over 10 MHz of bandwidth. The tones are cast on a realization of a Rayleigh-faded channel exhibiting a TU power delay profile, with an r.m.s. delay spread of $1 \mu\text{s}$.

- In slow fading, each transmission can be localized on convenient frequency resources and the rate can be matched to the fading thereon. Resources exhibiting poor fading can simply be avoided. Furthermore, the amount of bandwidth allocated to each user can be varied, altogether ensuring minimum instantaneous rates to those applications that require them. Therefore, the average rate is rendered significant even for applications with minimum rate requirements, reinforcing the adequacy of the ergodic capacity as a metric for this regime.
- In fast fading, when CSI is not available at the transmitter, each transmission can be interspersed over frequency resources sufficiently apart relative to the selectivity as to fade independently.

Example 1 In the 3GPP LTE system [6], a basic signaling resource spans 12 OFDM tones and 14 OFDM symbols in time. (This corresponds to 180 kHz and 1 ms, respectively.) In slow fading, the 12 tones are adjacent. In fast fading, they are interspersed over 10 MHz of bandwidth and thus fade independently. Shown in Fig. 3 are those interspersed tone positions cast on a realization of a Rayleigh-faded channel exhibiting a TU (typical urban) power delay profile [6].

D. Operating Point

Yet another fundamental facet of modern wireless systems is that the physical layer operates at some fixed block error probability, which depends on the specific parameters of each system but is typically around 1%. The aim of reliable communication is not given up, but the physical layer is no longer alone in ensuring it. The upper layers of the protocol stack incorporate additional ARQ mechanisms that can lower the error probability as needed, depending on the latency sensitivity of each

application. Some of the arguments that support having a fixed operating point at the physical layer are:

- With ARQ at the upper layers, it is inefficient to operate the physical layer at too small an error probability [7].
- There is little point in making the error probability on the traffic channels lower than that on the control channels, which cannot be made arbitrarily small because control messages are short and subject to tight latency constraints. As an example, the loss of a one-bit H-ARQ notification on a control channel incurs the cost of an entire packet of data in the associated traffic channel.

The link adaptation and H-ARQ processes are calibrated so as to ensure that the system operates tightly around the target block error probability. In fast fading specifically, and recalling the identity between error probability and outage probability, a fixed operating point implies a fixed outage probability. This affects the role of diversity, which by definition quantifies the speed at which the outage probability decreases with SNR (asymptotically). While material if one were to try to drive the outage probability to zero, diversity is not as informative for fixed outage levels.

IV. NEW WISDOMS

Armed with the perspectives furnished in the previous section, the following observations can be made:

- The quasi-static setting with fixed transmission rate has manifestly lost relevance. It is precluded by link adaptation in slow fading and by wideband signalling and H-ARQ in fast fading. These facts notwithstanding, the quasi-static setting remains popular and is the source of many artifact results. By subjecting each fixed-rate codeword to the vagaries of a single fading realization, the benefits of certain transmission and reception techniques can be grossly overstated (cf. [8]).
- The ergodic setting has grown in relevance. In wideband systems, long latencies are no longer a necessary toll for a codeword to experience very substantial selectivity. Rather, the codeword can get exposed to the fading distribution largely in the frequency domain (and, increasingly, in the space domain as well).

Altogether, the picture that emerges is as follows:

- For slow-fading, the appropriate setting is an unfaded channel with instantaneous signal-to-noise ratio accorded by the fading distribution. The ergodic capacity measures the average rate on this setting. For a richer description of the rate, one can calculate its entire distribution, which coincides with the outage capacity only interpreted differently: here, it does not represent the probability that a given fixed rate is supported, but rather the percentage

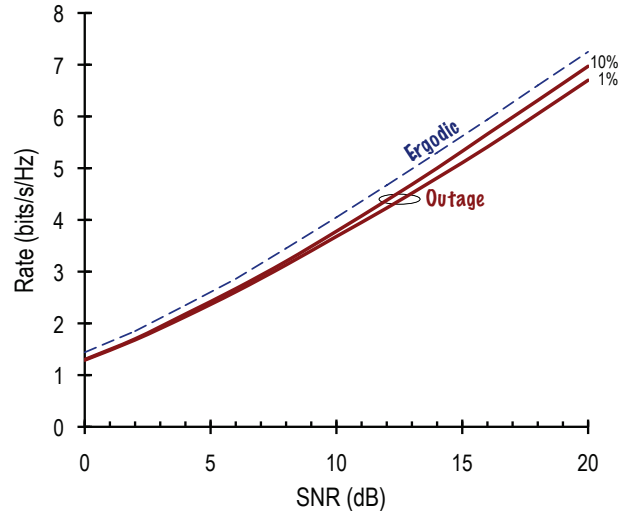


Fig. 4. In solid, 1% and 10% outage capacities on an LTE system with one transmit and two receive antennas, for SNRs of 0, 10, and 20 dB. In dashed, the corresponding ergodic capacity. The channel is Rayleigh-faded with a TU power delay profile and a Clarke-Jakes Doppler spectrum. The velocity is 100 Km/h and the carrier frequency is 2 GHz.

of time that transmission can take place—via link adaptation—at each given rate.

- For fast fading, the appropriate setting is a fading channel (continuous or block-fading) exposing each codeword to the correct number of realizations depending on the specifics of each system. The pertinent metric is the outage capacity at the desired operating point. Typically, the number of fading realizations affecting each codeword is large enough for the outage capacity at such operating point to be well approximated by the ergodic capacity.

Example 2 Consider the LTE system and the channel introduced in Example 1. The H-ARQ procedure extends each coded packet over as many as 6 rounds, each corresponding to a signaling resource and separated from the previous one by 6 ms. The channel evolves in a continuous fashion according to a Clarke-Jakes Doppler spectrum [2]. The velocity is 100 Km/h, the carrier frequency is 2 GHz and the additive noise is white Gaussian. Shown in Fig. 6 are the 1%-outage and the 10%-outage capacities, as well as the ergodic capacity, with one transmit and two receive antennas.

As anticipated, the ergodic capacity accurately approximates the outage capacity at both 1% and 10% outage levels. With additional transmit or receive antennas, this would be reinforced even further. From a computational standpoint, the agreement between the ergodic and outage capacities at virtually all outage levels is welcome news because of the added tractability of ergodic quantities. The ergodic capacity is thus seen as both relevant and analytically friendly, and this can only have positive implications in terms of deriving insights.

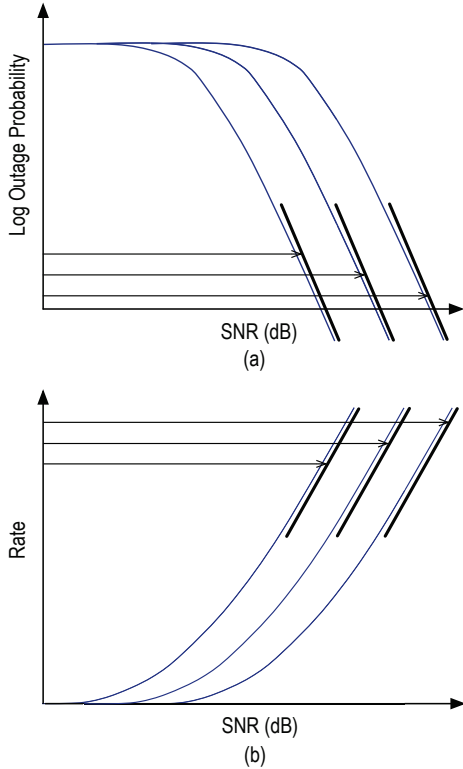


Fig. 5. (a) Outage probability (in log scale) v. SNR (in dB) for three channels with identical diversity order. (b) Rate v. SNR (in dB) for three channels with identical multiplexing gain.

V. THE DMT FRAMEWORK, REVISITED

For quasi-static settings, the DMT framework allows skirting many of the analytical hurdles associated with the computation of the outage capacity. At the same time, this framework suffers from some weaknesses, chiefly that diversity and multiplexing quantify only the (asymptotic) slope of outage and rate, respectively, as function of SNR. Arbitrarily large differences in the SNR required for a given outage probability may exist for identical diversities (cf. Fig. 5a). These differences are simply neglected. Likewise, arbitrarily large differences in the SNR required for a given rate (cf. Fig. 5b) are lost in the multiplexing gain. Therefore, there can be arbitrarily large differences between systems and channels exhibiting the same diversity and/or multiplexing. The DMT, in synthesis, expresses the tradeoff between the derivatives of the quantities of interest rather than between those quantities themselves.

Because of the above, diversity and multiplexing provide only a coarse depiction of the tradeoff between outage and rate; care must be exercised when utilizing these notions to establish comparisons between schemes or to determine absolute performance standards. Moreover, as argued earlier, the quasi-static setting has lost relevance to the slow-fading regime with the advent of

link adaptation. In fast fading, in turn, the operation at a fixed outage probability automatically resolves the tradeoff between outage and rate: A fixed outage probability is tantamount to zero diversity. Maximizing the rate at the operating outage level, and thus maximizing the multiplexing gain, should be the cardinal design principle of modern physical-layer designs [8].

VI. CONCLUSION

This paper has focused on how the canonical fading settings have been affected by the rapid evolution of wireless systems. The point at stake is not so much the level of detail in the models, necessarily limited, but rather insurance that the key features are captured and that artifacts are prevented. To be sure, the way forward proposed in Section IV does not complicate the canonical settings, but rather it simplifies them in some respects.

The discussion has centered on systems for which information-theoretic limits are relevant. Other types of systems exist that may not abide by the operational principles laid down in this paper. Wireless sensor networks, for instance, may not be adaptive, but by the same token they are not likely to feature capacity-approaching codes. In fact, they may employ little or no coding and thus the fundamental communication limits object of our attention are of little relevance to them. The control channels of a wireless system would be another example that does not conform to our observations, but for which capacity is again a metric of arguable relevance because of the very short nature of the messages. Interestingly, it is for these systems that do not operate close to capacity that the notion of diversity and the DMT framework is highly relevant, formulated in terms of error probability rather than outage probability (the two no longer coincide).

ACKNOWLEDGMENTS

This work was supported by AGAUR (Ref. 2009 SGR 70) and by the Spanish MICINN (Ref. TEC2009-13000).

REFERENCES

- [1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [2] W. C. Jakes, *Microwave Mobile Communications*. New York, IEEE Press, 1974.
- [3] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [4] A. J. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [5] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.
- [6] *UTRA-UTRAN Long Term Evolution (LTE)*, 3rd Generation Partnership Project (3GPP), Nov. 2004.
- [7] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: How reliable should the PHY be?" *Proc. of IEEE GLOBECOM'09*, Dec. 2009.
- [8] A. Lozano and N. Jindal, "Transmit diversity v. spatial multiplexing in modern MIMO systems," *IEEE Trans. Wireless Communications*, vol. 9, Jan. 2010.