Abstract

When the mechanisms of fading channels were first modeled in the 1950s and 1960s, the ideas were primarily applied to over-the-horizon communications covering a wide range of frequency bands. The 3–30 MHz high-frequency (HF) band is used for ionospheric communications, and the 300 MHz–3 GHz ultra-high-frequency (UHF) and 3–30 GHz super-high-frequency (SHF) bands are used for tropospheric scatter. Although the fading effects in a mobile radio system are somewhat different than those in ionospheric and tropospheric channels, the early models are still quite useful to help characterize fading effects in mobile digital communication systems. This tutorial addresses Rayleigh fading, primarily in the UHF band, that affects mobile systems such as cellular and personal communication systems (PCS). Part I of the tutorial itemizes the fundamental fading manifestations and types of degradation. Part II will focus on methods to mitigate the degradation.

Rayleigh Fading Channels in Mobile Digital Communication Systems

Part I: Characterization

Bernard Sklar, Communications Engineering Services

In the study of communication systems the classical (ideal) additive white Gaussian noise (AWGN) channel, with statistically independent Gaussian noise samples corrupting data samples free of intersymbol interference (ISI), is the usual starting point for understanding basic performance relationships. The primary source of performance degradation is thermal noise generated in the receiver. Often, external interference received by the antenna is more significant than thermal noise. This external interference can sometimes be characterized as having a broadband spectrum and is quantified by a parameter called antenna temperature [1]. The thermal noise usually has a flat power spectral density over the signal band and a zero-mean Gaussian voltage probability density function (pdf). When modeling practical systems, the next step is the introduction of bandlimiting filters. The filter in the transmitter usually serves to satisfy some regulatory requirement on spectral containment. The filter in the receiver often serves the purpose of a classical “matched filter” [2] to the signal bandwidth. Due to the bandlimiting and phase-distortion properties of filters, special signal design and equalization techniques may be required to mitigate the filter-induced ISI.

If a radio channel’s propagating characteristics are not specified, one usually infers that the signal attenuation versus distance behaves as if propagation takes place over ideal free space. The model of free space treats the region between the transmit and receive antennas as being free of all objects that might absorb or reflect radio frequency (RF) energy. It also assumes that, within this region, the atmosphere behaves as a perfectly uniform and nonabsorbing medium. Furthermore, the earth is treated as being infinitely far away from the propagating signal (or, equivalently, as having a reflection coefficient that is negligible).Basically, in this idealized free-space model, the attenuation of RF energy between the transmitter and receiver behaves according to an inverse-square law. The received power expressed in terms of transmitted power is attenuated by a factor, \( L_s(d) \), where this factor is called path loss or free space loss. When the receiving antenna is isotropic, this factor is expressed as [1]:

\[
L_s(d) = \left( \frac{4\pi d}{\lambda} \right)^2
\]

In Eq. 1, \( d \) is the distance between the transmitter and the receiver, and \( \lambda \) is the wavelength of the propagating signal. For this case of idealized propagation, received signal power is very predictable.

For most practical channels, where signal propagation takes place in the atmosphere and near the ground, the free-space propagation model is inadequate to describe the channel and predict system performance. In a wireless mobile communication system, a signal can travel from transmitter to receiver over multiple reflective paths; this phenomenon is referred to as multipath propagation. The effect can cause fluctuations in the received signal’s amplitude, phase, and angle of arrival, giving rise to the terminology multipath fading. Another name, scintillation, which originated in radio astronomy, is used to describe the multipath fading caused by physical changes in the propagating medium, such as variations in the density of ions in the ionospheric layers that reflect high-frequency (HF) radio signals. Both names, fading and scintillation, refer to a signal’s random fluctuations or fading due to multipath propagation. The main difference is that scintillation involves mechanisms (e.g., ions) that are much smaller than a wavelength. The end-to-end modeling and design of systems that mitigate the effects of fading are usually more challenging than those whose sole source of performance degradation is AWGN.

Mobile Radio Propagation: Large-Scale Fading and Small-Scale Fading

Fig. 1 represents an overview of fading channel manifestations. It starts with two types of fading effects that characterize mobile communications: large-scale and small-scale fading. Large-scale fading represents the average signal power attenuation or path loss due to motion over large areas. In
Fig. 1, the large-scale fading manifestation is shown in blocks 1, 2, and 3. This phenomenon is affected by prominent terrain contours (hills, forests, billboards, clumps of buildings, etc.) between the transmitter and receiver. The receiver is often represented as being “shadowed” by such prominetes. The statistics of large-scale fading provide a way of computing an estimate of path loss as a function of distance. This is described in terms of a mean-path loss (nth power law) and a log-normally distributed variation about the mean. Small-scale fading refers to the dramatic changes in signal amplitude and phase that can be experienced as a result of small changes (as small as a half-wavelength) in the spatial separation between a receiver and transmitter. As indicated in Fig. 1, blocks 4, 5, and 6, small-scale fading manifests itself in two mechanisms, namely, time-spreading of the signal (or signal dispersion) and time-variant behavior of the channel. For mobile radio applications, the channel is time-variant because motion between the transmitter and receiver results in propagation path changes. The rate of change of these propagation conditions accounts for the fading rapidity (rate of change of the fading impairments). Small-scale fading is also called Rayleigh fading because if the multiple reflective paths are large in number and there is no line-of-sight signal component, the envelope of the received signal is statistically described by a Rayleigh pdf. When there is a dominant nonfading signal component present, such as a line-of-sight propagation path, the small-scale fading envelope is described by a Rician pdf [3]. A mobile radio roaming over a large area must process signals that experience both types of fading: small-scale fading superimposed on large-scale fading.

There are three basic mechanisms that impact signal propagation in a mobile communication system. They are reflection, diffraction, and scattering [3]:
- Reflection occurs when a propagating electromagnetic wave impinges on a smooth surface with very large dimensions compared to the RF signal wavelength (\( \lambda \)).
- Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a dense body with large dimensions compared to \( \lambda \), causing secondary waves to be formed behind the obstructing body. Diffraction is a phenomenon that accounts for RF energy traveling from transmitter to receiver without a line-of-sight path between the two. It is often termed shadowing because the diffracted field can reach the receiver even when shadowed by an impenetrable obstruction.
- Scattering occurs when a radio wave impinges on either a large rough surface or any surface whose dimensions are on the order of \( \lambda \) or less, causing the reflected energy to spread out (scatter) in all directions. In an urban environment, typical signal obstructions that yield scattering are lampposts, street signs, and foliage.

Figure 1 may serve as a table of contents for the sections that follow. We will examine the two manifestations of small-scale fading: signal time-spreading (signal dispersion) and the time-variant nature of the channel. These examinations will take place in two domains, time and frequency, as indicated in Fig. 1, blocks 7, 10, 13, and 16. For signal dispersion, we categorize the fading degradation types as frequency-selective or frequency-nonselective (flat), as listed in blocks 8, 9, 11, and 12. For the time-variant manifestation, we categorize the fading degradation types as fast- or slow-fading, as listed in blocks 14, 15, 17, and 18. The labels indicating Fourier transforms and duals will be explained later.

Figure 2 illustrates the various contributions that must be considered when estimating path loss for a link budget analysis in a cellular application [4]. These contributions are:
- Mean path loss as a function of distance due to large-scale fading
- Near-worst-case variations about the mean path loss (typically 6–10 dB), or large-scale fading margin
- Near-worst-case Rayleigh or small-scale fading margin (typically 20–30 dB)

In Fig. 2, the annotations “@ 1–2%” indicate a suggested area (probability) under the tail of each pdf as a design goal.
Hence, the amount of margin indicated is intended to provide adequate received signal power for approximately 98–99 percent of each type of fading variation (large- and small-scale).

A received signal, \( r(t) \), is generally described in terms of a transmitted signal \( s(t) \) convolved with the impulse response of the channel \( h_c(t) \). Neglecting the degradation due to noise, we write

\[
r(t) = s(t) * h_c(t),
\]

where \( * \) denotes convolution. In the case of mobile radios, \( r(t) \) can be partitioned in terms of two component random variables, as follows [5]:

\[
r(t) = m(t) \times n_0(t),
\]

where \( m(t) \) is called the large-scale-fading component, and \( n_0(t) \) is called the small-scale-fading component. \( m(t) \) is sometimes referred to as the local mean or log-normal fading because the magnitude of \( m(t) \) is described by a log-normal pdf (or, equivalently, the magnitude measured in decibels has a Gaussian pdf). \( n_0(t) \) is sometimes referred to as multipath or Rayleigh fading. Figure 3 illustrates the relationship between large-scale and small-scale fading. In Fig. 3a, received signal power \( r(t) \) versus antenna displacement (typically in units of wavelength) is plotted, for the case of a mobile radio. Small-scale fading superimposed on large-scale fading can be readily identified. The typical antenna displacement between the small-scale signal nulls is approximately a half wavelength. In Fig. 3b, the large scale fading or local mean, \( m(t) \), has been removed in order to view the small-scale fading, \( n_0(t) \), about some average constant power.

In the sections that follow, we enumerate some of the details regarding the statistics and mechanisms of large-scale and small-scale fading.

**Large-Scale Fading: Path-Loss Mean and Standard Deviation**

For the mobile radio application, Okumura [6] made some of the earlier comprehensive path-loss measurements for a wide range of antenna heights and coverage distances. Hata [7] transformed Okumura’s data into parametric formulas. For the mobile radio application, the mean path loss, \( T_p(d) \), as a function of distance, \( d \), between the transmitter and receiver is proportional to an nth power of \( d \) relative to a reference distance \( d_0 \) [3].

\[
T_p(d) = \frac{d^n}{d_0^n}
\]

(4)

\( T_p(d) \) is often stated in decibels, as shown below.

\[
T_p(d) \text{ (dB)} = L_s(d_0) \text{ (dB)} + 10 \log\left(\frac{d}{d_0}\right)
\]

(5)

The reference distance \( d_0 \) corresponds to a point located in the far field of the antenna. Typically, the value of \( d_0 \) is taken to be 1 km for large cells, 100 m for microcells, and 1 m for indoor channels. \( T_p(d) \) is the average path loss (over a multitude of different sites) for a given value of \( d \). Linear regression for a minimum mean-squared estimate (MMSE) fit of \( T_p(d) \) versus \( d \) on a log-log scale (for distances greater than \( d_0 \)) yields a straight line with a slope equal to \( 10n \) dB/decade. The value of the exponent \( n \) depends on the frequency, antenna heights, and propagation environment. In free space, \( n = 2 \), as seen in Eq. 1. In the presence of a very strong guided wave phenomenon (like urban streets), \( n \) can be lower than 2. When obstructions are present, \( n \) is larger. The path loss \( L_s(d_0) \) to the reference point at a distance \( d_0 \) from the transmitter is typically found through field measurements or calculated using the free-space path loss given by Eq. 1. Figure 4 shows a scatter plot of path loss versus distance for measurements made in several German cities [8]. Here, the path loss has been measured relative to the free-space reference measurement at \( d_0 = 100 \) m. Also shown are straight-line fits to various exponent values.

The path loss versus distance expressed in Eq. 5 is an average, and therefore not adequate to describe any particular setting or signal path. It is necessary to provide for variations about the mean since the environment of different sites may be quite different for similar transmitter-receiver separations. Figure 4 illustrates that path-loss variations can be quite large. Measurements have shown that for any value of \( d \), the path loss \( L_p(d) \) is a random variable having a log-normal distribution about the mean distant-dependent value \( T_p(d) \) [9]. Thus, path loss \( L_p(d) \) can be expressed in terms of \( T_p(d) \) plus a random variable \( X_\sigma \), as follows [3]:

\[
L_p(d) \text{ (dB)} = L_s(d_0) \text{ (dB)} + 10 \log\left(\frac{d}{d_0}\right) + X_\sigma \text{ (dB)}
\]

(6)

where \( X_\sigma \) denotes a zero-mean Gaussian random variable (in decibels) with standard deviation \( \sigma \) (also in decibels). \( X_\sigma \) is site- and distance-dependent. The choice of a value for \( X_\sigma \) is
often based on measurements; it is not unusual for it to take on values as high as 6–10 dB or greater. Thus, the parameters needed to statistically describe path loss due to large-scale fading for an arbitrary location with a specific transmitter-receiver separation are:

- The reference distance \( d_0 \)
- The path-loss exponent \( n \)
- The standard deviation \( \sigma \) of \( X_d \)

There are several good references dealing with the measurement and estimation of propagation path loss for many different applications and configurations [3, 7–11].

**Small-Scale Fading: Statistics and Mechanisms**

When the received signal is made up of multiple reflective rays plus a significant line-of-sight (nonfaded) component, the envelope amplitude due to small-scale fading has a Ricean pdf, and is referred to as Ricean fading [3]. The nonfaded component is called the specular component. As the amplitude of the specular component approaches zero, the Ricean pdf approaches a Rayleigh pdf, expressed as

\[
p(r) = \begin{cases} 
\frac{r}{\sigma^2} \exp \left[ -\frac{r^2}{2\sigma^2} \right] & \text{for } r \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]  

(7)

where \( r \) is the envelope amplitude of the received signal, and \( 2\sigma^2 \) is the pre-detection mean power of the multipath signal. The Rayleigh faded component is sometimes called the random or scatter or diffuse component. The Rayleigh pdf results from having no specular component of the signal; thus, for a single link it represents the pdf associated with the worst case of fading per mean received signal power. For the remainder of this article, it will be assumed that loss of signal-to-noise ratio (SNR) due to fading follows the Rayleigh model described. It will also be assumed that the propagating signal is in the ultra-high-frequency (UHF) band, encompassing present-day cellular and personal communications services (PCS) frequency allocations — nominally 1 GHz and 2 GHz, respectively.

As indicated in Fig. 1, blocks 4, 5, and 6, small-scale fading manifests itself in two mechanisms:

- Time-spreading of the underlying digital pulses within the signal
- A time-variant behavior of the channel due to motion (e.g., a receive antenna on a moving platform)

Figure 5 illustrates the consequences of both manifestations by showing the response of a multipath channel to a narrow pulse versus delay, as a function of antenna position (or time, assuming a constant velocity of motion). In Fig. 5, we distinguish between two different time references — delay time \( \tau \) and transmission or observation time \( t \). Delay time refers to the time-spreading manifestation which results from the fading channel’s nonoptimum impulse response. The transmission time, however, is related to the antenna’s motion or spatial changes, accounting for propagation path changes that are perceived as the channel’s time-variant behavior. Note that for constant velocity, as assumed in Fig. 5, either antenna position or transmission time can be used to illustrate this time-variant behavior. Figures 5a–5c show the sequence of received pulse-power profiles as the antenna moves through a succession of equally spaced positions. Here, the interval between antenna positions is 0.4 \( \lambda \) [12], where \( \lambda \) is the wavelength of the carrier frequency. For each of the three cases shown, the response pattern differs significantly in the delay time of the largest signal component, the number of signal copies, their magnitudes, and the total received power (area) in the received power profile. Figure 6 summarizes these two small-scale fading mechanisms, the two domains (time or time-delay and frequency or Doppler shift) for viewing each mechanism, and the degradation categories each mechanism can exhibit. Note that any mechanism characterized in the time domain can be characterized equally well in the frequency domain. Hence, as outlined in Fig. 6, the time-spreading mechanism will be characterized in the time-delay domain as a multipath delay spread, and in the frequency domain as a channel coherence bandwidth. Similarly, the time-variant mechanism will be characterized in the time domain as a channel coherence time, and in the Doppler-shift (frequency) domain as a channel fading rate or Doppler spread. These mechanisms and
their associated degradation categories will be examined in greater detail in the sections that follow.

**Signal Time-Spreading Viewed in the Time-Delay Domain: Fig. 1, Block 7 — The Multipath Intensity Profile**

A simple way to model the fading phenomenon was introduced by Bello [13] in 1963; he proposed the notion of wide-sense stationary uncorrelated scattering (WSSUS). The model treats signal variations arriving with different delays as uncorrelated. It can be shown [4, 13] that such a channel is effectively WSS in both the time and frequency domains. With such a model of a fading channel, Bello was able to define functions that apply for all time and all frequencies. For the mobile channel, Figure 7 contains four functions that make up this model [4, 13–16]. We will examine these functions, starting with Fig. 7a and proceeding counterclockwise toward Fig. 7d.

In Fig. 7a a multipath-intensity profile, \( S(\tau) \) versus time delay \( \tau \), is plotted. Knowledge of \( S(\tau) \) helps answer the question “For a transmitted impulse, how does the average received power vary as a function of time delay, \( \tau \)?” The term “time delay” is used to refer to the excess delay. It represents the signal’s propagation delay that exceeds the delay of the first signal arrival at the receiver. For a typical wireless radio channel, the received signal usually consists of several discrete multipath components, sometimes referred to as fingers. For some channels, such as the tropospheric scatter channel, received signals are often seen as a continuum of multipath components [14, 16]. For making measurements of the multipath intensity profile, wideband signals (impulses or spread spectrum) need to be used [16]. For a single transmitted impulse, the time, \( T_m \), between the first and last received component represents the maximum excess delay, during which the multipath signal power falls to some threshold level below that of the strongest component. The threshold level might be chosen at 10 or 20 dB below the level of the strongest component. Note that for an ideal system (zero excess delay), the function \( S(\tau) \) would consist of an ideal impulse with weight equal to the total average received signal power.

**Degradation Categories Due to Signal Time-Spreading Viewed in the Time-Delay Domain**

In a fading channel, the relationship between maximum excess delay time, \( T_m \), and symbol time, \( T_s \), can be viewed in terms of two different degradation categories, frequency-selective fading and frequency nonselective or flat fading, as indicated in Fig. 1, blocks 8 and 9, and Fig. 6. A channel is said to exhibit frequency-selective fading if \( T_m < T_s \). This condition occurs whenever the received multipath components of a symbol extend beyond the symbol’s time duration. Such multipath dispersion of the signal yields the same kind of ISI distortion caused by an electronic filter. In fact, another name for this category of fading degradation is channel-induced ISI.

In the case of frequency-selective fading, mitigating the distortion is possible because many of the multipath components are resolvable by the receiver. In Part II of this tutorial, several such mitigation techniques are described.

A channel is said to exhibit frequency nonselective or flat fading if \( T_m > T_s \). In this case, all the received multipath components of a symbol arrive within the symbol time duration; hence, the components are not resolvable. Here, there is no channel-induced ISI distortion, since the signal time spreading does not result in significant overlap among neighboring received symbols. There is still performance degradation since the unresolvable phasor components can add up destructively to yield a substantial reduction in SNR. Also, signals that are classified as exhibiting flat fading can sometimes experience frequency-selective distortion. This will be explained later when viewing degradation in the frequency domain, where the phenomenon is more easily described. For loss in SNR due to flat fading, the mitigation technique called for is to improve the received SNR (or reduce the required SNR). For digital systems, introducing some form of signal diversity and using error-correction coding is the most efficient way to accomplish this.

**Signal Time-Spreading Viewed in the Frequency Domain: Fig. 1, Block 10 — The Spaced-Frequency Correlation Function**

A completely analogous characterization of signal dispersion can begin in the frequency domain. In Fig. 7b is seen the function \( |R(\Delta f)| \), designated a spaced-frequency correlation function; it is the Fourier transform of \( S(\tau) \). \( R(\Delta f) \) represents the correlation between the channel’s response to two signals as a function of the frequency difference between the two signals. It can be thought of as the channel’s frequency transfer function. Therefore, the time-spreading manifestation can be viewed as if it were the result of a filtering process. Knowledge of \( R(\Delta f) \) helps answer the question “What is the correlation between received signals that are spaced in frequency \( \Delta f \)?
$= f_1 - f_2$? $R(\Delta f)$ can be measured by transmitting a pair of sinusoids separated in frequency by $\Delta f$, cross-correlating the two separately received signals, and repeating the process many times with ever-larger separation $\Delta f$. Therefore, the measurement of $R(\Delta f)$ can be made with a sinusoid that is swept in frequency across the band of interest (a wideband signal). The coherence bandwidth, $f_0$, is a statistical measure of the range of frequencies over which the channel passes all spectral components with approximately equal gain and linear phase. Thus, the coherence bandwidth represents a frequency range over which frequency components have a strong potential for amplitude correlation. That is, a signal's spectral components in that range are affected by the channel in a similar manner as, for example, exhibiting fading or no fading. Note that $f_0$ and $T_m$ are reciprocally related (within a multiplicative constant). As an approximation, it is possible to say that

$$f_0 \approx \frac{1}{T_m}$$  \hspace{1cm} (8)

The maximum excess delay, $T_m$, is not necessarily the best indicator of how any given system will perform on a channel because different channels with the same value of $T_m$ can exhibit very different profiles of signal intensity over the delay span. A more useful measurement of delay spread is most often characterized in terms of the root mean squared (rms) delay spread, $\sigma_t$, where

$$\sigma_t = \sqrt{\tau^2 - \bar{\tau}^2}$$  \hspace{1cm} (9)

$\bar{\tau}$ is the mean excess delay, $(\bar{\tau})^2$ is the mean squared, $\tau^2$ is the second moment, and $\sigma_t$ is the square root of the second central moment of $S(\tau)$ [3].

An exact relationship between coherence bandwidth and delay spread does not exist, and must be derived from signal analysis (usually using Fourier techniques) of actual signal dispersion measurements in particular channels. Several approximate relationships have been described. If coherence bandwidth is defined as the frequency interval over which the channel's complex frequency transfer function has a correlation of at least 0.9, the coherence bandwidth is approximately [17]

$$f_0 = \frac{1}{50\sigma_t}$$  \hspace{1cm} (10)

For the case of a mobile radio, an array of radially uniformly spaced scatterers, all with equal-magnitude reflection coefficients but independent, randomly occurring reflection phase angles [18, 19] is generally accepted as a useful model for urban surroundings. This model is referred to as the dense-scatterer channel model. With the use of such a model, coherence bandwidth has similarly been defined [18], for a bandwidth interval over which the channel's complex frequency transfer function has a correlation of at least 0.5, to be

$$f_0 = \frac{0.276}{\sigma_t}$$  \hspace{1cm} (11)

The ionospheric effects community employs the following definition [20]:

$$f_0 = \frac{1}{2\sigma_t}$$  \hspace{1cm} (12)

A more popular approximation of $f_0$ corresponding to a bandwidth interval having a correlation of at least 0.5 is [3]

$$f_0 = \frac{1}{50\sigma_t}$$  \hspace{1cm} (13)

### Degradation Categories Due to Signal Time Spreading Viewed in the Frequency Domain

A channel is referred to as frequency-selective if $f_0 < 1/T_e \otimes W$, where the symbol rate, $1/T_e$, is nominally taken to be equal to the signal bandwidth $W$. In practice, $W$ may differ from $1/T_e$ due to system filtering or data modulation type (quaternary phase shift keying, QPSK, minimum shift keying, MSK, etc.) [21]. Frequency-selective fading distortion occurs whenever a signal's spectral components are not all affected equally by the channel. Some of the signal's spectral components, falling outside the coherence bandwidth, will be affected differently (independently) compared to those components contained within the coherence bandwidth. This occurs whenever $f_0 < W$ and is illustrated in Fig. 8a.

Frequency-nonselective or flat fading degradation occurs whenever $f_0 > W$. Hence, all of the signal's spectral components will be affected by the channel in a similar manner (e.g., fading or no fading); this is illustrated in Fig. 8b. Flat-fading does not introduce channel-induced ISI distortion, but performance degradation can still be expected due to loss in SNR whenever the signal is fading. In order to avoid channel-induced ISI distortion, the channel is required to exhibit flat fading by ensuring that
Typical Examples of Flat Fading and Frequency-Selective Fading

Figure 9 shows some examples of flat fading and frequency-selective fading for a direct-sequence spread-spectrum (DS/SS) system [20, 22]. In Fig. 9, there are three plots of the output of a pseudo-noise (PN) code correlator versus delay as a function of time (transmission or observation time). Each amplitude versus delay plot is akin to $S(t)$ versus $\tau$ in Fig. 7a. The key difference is that the amplitudes shown in Fig. 9 represent the output of a correlator; hence the waveforms are a function not only of the impulse response of the channel, but also of the impulse response of the correlator. The delay time is expressed in units of chip durations (chips), where the chip is defined as the spread-spectrum minimal-duration keying element. For each plot, the observation time is shown on an axis perpendicular to the amplitude versus time-delay plane. Figure 9 is drawn from a satellite-to-ground communications link exhibiting scintillation because of atmospheric disturbances. However, Fig. 9 is still a useful illustration of three different channel conditions that might apply to a mobile radio situation. A mobile radio that moves along the observation-time axis is affected by changing multipath profiles along the route, as seen in the figure. The scale along the observation-time axis is also in units of chips. In Fig. 9a, the signal dispersion (one “finger” of return) is on the order of a chip time duration, $T_{ch}$. In a typical DS/SS system, the spread-spectrum signal bandwidth is approximately equal to $1/T_{ch}$; hence, the normalized coherence bandwidth $f_0T_{ch}$ of approximately unity in Fig. 9a implies that the coherence bandwidth is about equal to the spread-spectrum bandwidth. This describes a channel that can be called frequency-nonselective or slightly frequency-selective. In Fig. 9b, where $f_0T_{ch}$ = 0.25, the signal dispersion is more pronounced. There is definite interchip interference, and the coherence bandwidth is approximately equal to 25 percent of the spread-spectrum bandwidth. In Fig. 9c, where $f_0T_{ch}$ = 0.1, the signal dispersion is even more pronounced, with greater interchip interference effects, and the coherence bandwidth is approximately equal to 10 percent of the spread-spectrum bandwidth. The channels of Figs. 9b and 9c can be categorized as moderately and highly frequency-selective, respectively, with respect to the basic signaling element, the chip. In Part II of this article, we show that a DS/SS system operating over a frequency-selective channel at the chip level does not necessarily experience frequency-selective distortion at the symbol level.

Time Variance Viewed in the Time Domain: Fig. 1, Block 13 — The Spaced-Time Correlation Function

Until now, we have described signal dispersion and coherence bandwidth, parameters that describe the channel’s time-spreading properties in a local area. However, they do not offer information about the time-varying nature of the channel caused by relative motion between a transmitter and receiver, or by movement of objects within the channel. For mobile radio applications, the channel is time variant because motion between the transmitter and receiver results in propagation path changes. Thus, for a transmitted continuous wave (CW) signal, as a result of such motion, the radio receiver sees variations in the signal’s amplitude and phase. Assuming that all scatterers...
making up the channel are stationary, whenever motion ceases, the amplitude and phase of the received signal remains constant; that is, the channel appears to be time-invariant. Whenever motion begins again, the channel once again appears time-variant. Since the channel characteristics are dependent on the positions of the transmitter and receiver, time variance in this case is equivalent to spatial variance.

Figure 7c shows the function $R(\Delta t)$, designated the spacet ime correlation function; it is the autocorrelation function of the channel’s response to a sinusoid. This function specifies the extent to which there is correlation between the channel’s response to a sinusoid sent at time $t_1$ and the response to a similar sinusoid sent at time $t_2$, where $\Delta t = t_2 - t_1$. The coherence time, $T_\text{coh}$, is a measure of the expected time duration over which the channel’s response is essentially invariant. Earlier, we made measurements of signal dispersion and coherence bandwidth by using wideband signals. Now, to measure the time-variant nature of the channel, we use a narrowband signal [16]. To measure $R(\Delta t)$ we can transmit a single sinusoid ($\Delta t = 0$) and determine the autocorrelation function of the received signal. The function $R(\Delta t)$ and the parameter $T_\text{coh}$ provide us with knowledge about the fading rapidity of the channel. Note that for an ideal time-invariant channel (e.g., a mobile radio exhibiting no motion at all), the channel’s response would be highly correlated for all values of $\Delta t$, and $R(\Delta t)$ would be a constant function. When using the dense-scatterer channel model described earlier, with constant velocity of motion, and an unmodulated CW signal, the normalized $R(\Delta t)$ is described as [19].

$$R(\Delta t) = J_0(k\nu \Delta t)$$  \hspace{1cm} (15)

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, $\nu$ is velocity, $\nu \Delta t$ is distance traversed, and $k = 2\pi/\lambda$ is the free-space phase constant (transforming distance to radians of phase). Coherence time can be measured in terms of either time or distance traversed (assuming some fixed velocity of motion). Amoroso described such a measurement using a CW signal and a dense-scatterer channel model [18]. He measured the statistical correlation between the combination of received magnitude and phase sampled at a particular antenna location $x_0$, and the corresponding combination sampled at some displaced location $x_0 + \zeta$, with displacement measured in units of wavelength $\lambda$. For a displacement $\zeta$ of $\leq \lambda$ between two antenna locations, the combined magnitudes and phases of the received CW are statistically uncorrelated. In other words, the state of the signal at $x_0$ says nothing about the state of the signal at $x_0 + \zeta$. For a given velocity of motion, this displacement is readily transformed into units of time (coherence time).

**The Concept of Duality**

Two operators (functions, elements, or systems) are dual when the behavior of one with reference to a time-related domain (time or time-delay) is identical to the behavior of the other in reference to the corresponding frequency-related domain (frequency or Doppler shift).

In Fig. 7, we can identify functions that exhibit similar behavior across domains. These behaviors are not identical to one another in the strict mathematical sense, but for understanding the fading channel model it is still useful to refer to such functions as duals. For example, $R(\Delta t)$ in Fig. 7b, characterizing signal dispersion in the frequency domain, yields knowledge about the range of frequency over which two spectral components of a received signal have a strong potential for amplitude and phase correlation. $R(\Delta f)$ in Fig. 7c, characterizing fading rapidly in the time domain, yields knowledge about the span of time over which two received signals have a strong potential for amplitude and phase correlation. We have labeled these two correlation functions duals. This is also noted in Fig. 1 as the duality between blocks 10 and 13, and in Fig. 6 as the duality between the time-spreading mechanism in the frequency domain and the time-variant mechanism in the time domain.

**Degradation Categories Due to Time Variance Viewed in the Time Domain**

The time-variant nature of the channel or fading rapidity mechanism can be viewed in terms of two degradation categories listed in Fig. 6: fast fading and slow fading. The terminology “fast fading” is used to describe channels in which $T_\text{coh} < T_P$, where $T_\text{coh}$ is the channel coherence time and $T_P$ is the time duration of a transmission symbol. Fast fading describes a condition where the time duration in which the channel behaves in a correlated manner is short compared to the time duration of a symbol. Therefore, it can be expected that the fading character of the channel will change several times while a symbol is propagating, leading to distortion of the baseband pulse shape. Analogous to the distortion previously described as channel-induced ISI, here distortion takes place because the received signal’s components are not all highly correlated throughout time. Hence, fast fading can cause the baseband pulse to be distorted, resulting in a loss of SNR that often yields an irreducible error rate. Such distorted pulses cause synchronization problems (failure of phase-locked-loop receivers), in addition to difficulties in adequately defining a matched filter.

**Figure 9. DS/SS matched-filter output time-history examples for three levels of channel conditions, where $T_\text{coh}$ is the time duration of a chip.**
A channel is generally referred to as introducing slow fading if $T_A > T_d$. Here, the time duration that the channel behaves in a correlated manner is long compared to the time duration of a transmission symbol. Thus, one can expect the channel state to virtually remain unchanged during the time in which a symbol is transmitted. The propagating symbols will likely not suffer from the pulse distortion described above. The primary degradation in a slow-fading channel, as with flat fading, is loss in SNR.

**TIME VARIANCE VIEWED IN THE DOPPLER-SHIFT DOMAIN: FIG. 1, BLOCK 16 — THE DOPPLER POWER SPECTRUM**

A completely analogous characterization of the time-variant nature of the channel can begin in the Doppler shift (frequency) domain. Figure 7d shows a Doppler power spectral density, $S(v)$, plotted as a function of Doppler-frequency shift, $v$. For the case of the dense-scatterer model, a vertical receive antenna with constant azimuthal gain, a uniform distribution of signals arriving at all arrival angles throughout the range $(0, 2\pi)$, and an unmodulated CW signal, the signal spectrum at the antenna terminals is [19]

$$S(v) = \frac{1}{\pi f_d} \left[ 1 - \left( \frac{v}{f_d} \right)^2 \right]$$

(16)

The equality holds for frequency shifts of $v$ that are in the range $\pm f_d$ about the carrier frequency $f_c$, and would be zero outside that range. The shape of the RF Doppler spectrum described by Eq. 16 is classically bowl-shaped, as seen in Fig. 7d. Note that the spectral shape is a result of the dense-scatterer channel model. Equation 16 has been shown to match experimental data gathered for mobile radio channels [23]; however, different applications yield different spectral shapes. For example, the dense-scatterer model does not hold for the indoor radio channel; the channel model for an indoor area assumes $S(v)$ to be a flat spectrum [24].

In Fig. 7d, the sharpness and steepness of the boundaries of the Doppler spectrum are due to the sharp upper limit on the Doppler shift produced by a vehicular antenna traveling among the stationary scatterers of the dense scatterer model. The largest magnitude (infinite) of $S(v)$ occurs when the scatterer is directly ahead of the moving antenna platform or directly behind it. In that case the magnitude of the frequency shift is given by

$$f_d = \frac{V}{\lambda}$$

(17)

where $V$ is relative velocity, and $\lambda$ is the signal wavelength. $f_d$ is positive when the transmitter and receiver move toward each other, and negative when moving away from each other. For scatterers directly broadside of the moving platform the magnitude of the frequency shift is zero. The fact that Doppler components arriving at exactly 0° and 180° have an infinite power spectral density is not a problem, since the angle of arrival is continuously distributed and the probability of components arriving at exactly these angles is zero [3, 19].

$S(v)$ is the Fourier transform of $R(\Delta r)$. We know that the Fourier transform of the autocorrelation function of a time series is the magnitude squared of the Fourier transform of the original time series. Therefore, measurements can be made by simply transmitting a sinusoid (narrowband signal) and using Fourier analysis to generate the power spectrum of the received amplitude [16]. This Doppler power spectrum of the channel yields knowledge about the spectral spreading of a transmitted sinusoid (impulse in frequency) in the Doppler shift domain. As indicated in Fig. 7, $S(v)$ can be regarded as the dual of the multipath intensity profile, $S(t)$, since the latter yields knowledge about the time spreading of a transmitted impulse in the time-delay domain. This is also noted in Fig. 1 as the duality between blocks 7 and 16, and in Fig. 6 as the duality between the time-spreading mechanism in the time-delay domain and the time-variant mechanism in the Doppler-shift domain.

Knowledge of $S(v)$ allows us to glean how much spectral broadening is imposed on the signal as a function of the rate of change in the channel state. The width of the Doppler power spectrum is referred to as the spectral broadening or Doppler spread, denoted by $f_o$, and sometimes called the fading bandwidth of the channel. Equation 16 describes the Doppler frequency shift. In a typical multipath environment, the received signal arrives from several reflected paths with different path distances and different angles of arrival, and the Doppler shift of each arriving path is generally different from that of another path. The effect on the received signal is seen as a Doppler spreading or spectral broadening of the transmitted signal frequency, rather than a shift. Note that the Doppler spread, $f_o$, and the coherence time, $T_0$, are reciprocally related (within a multiplicative constant). Therefore, we show the approximate relationship between the two parameters as

$$T_0 \approx \frac{1}{f_o}$$

(18)

Hence, the Doppler spread $f_o$ or $1/T_0$ is regarded as the typical fading rate of the channel. Earlier, $T_0$ was described as the expected time duration over which the channel’s response to a sinusoid is essentially invariant. When $T_0$ is defined more precisely as the time duration over which the channel’s response to a sinusoid has a correlation greater than 0.5, the relationship between $T_0$ and $f_o$ is approximately [4]

$$T_0 = \frac{9}{16\pi f_o}$$

(19)

A popular rule of thumb is to define $T_0$ as the geometric mean of Eqs. 18 and 19. This yields

$$T_0 = \left( \frac{9}{16\pi f_o} \right)^{0.423} \approx \frac{0.423}{f_o}$$

(20)

For the case of a 900 MHz mobile radio, Fig. 10 illustrates the typical effect of Rayleigh fading on a signal's envelope.
amplitude versus time [3]. The figure shows that the distance traveled by the mobile in the time interval corresponding to two adjacent nulls (small-scale fades) is on the order of a half-wavelength ($\lambda/2$) [3]. Thus, from Fig. 10 and Eq. 17, the time (approximately the coherence time) required to traverse a distance when traveling at a constant velocity, $V$, is

$$T_0 = \frac{\lambda/2}{V} = \frac{0.5}{f_d}$$  \hspace{1cm} (21)

Thus, when the interval between fades is taken to be $\lambda/2$, as in Fig. 10, the resulting expression for $T_0$ in Eq. 21 is quite close to the rule of thumb shown in Eq. 20. Using Eq. 21 with the parameters shown in Fig. 10 (velocity = 120 km/hr and carrier frequency = 900 MHz), it is straightforward to compute that the coherence time is approximately 5 ms and the Doppler spread (channel fading rate) is approximately 100 Hz. Therefore, if this example represents a voice-grade channel with a typical transmission rate of $10^6$ symbols/s, the fading rate is considerably less than the symbol rate. Under such conditions, the channel would manifest slow-fading effects. Note that if the abscissa of Fig. 10 were labeled in units of wavelength instead of time, the figure would look the same for any radio frequency and any antenna speed.

**ANALOGY BETWEEN SPECTRAL BROADENING IN FADING CHANNELS AND SPECTRAL BROADENING IN DIGITAL SIGNAL KEYING**

Help is often needed in understanding why spectral broadening of the signal is a function of fading rate of the channel. Figure 11 uses the keying of a digital signal (e.g., amplitude shift keying or frequency shift keying) to illustrate an analogous case. Figure 11a shows a single tone, cos $2\pi f t$ ($-\infty < t < \infty$), which exists for all time characterized in the frequency domain in terms of impulses at $\pm f_c$. This frequency domain representation is ideal (i.e., zero bandwidth), since the tone is pure and neverending. In practical applications, digital signaling involves switching (keying) signals on and off at a required rate. The keying operation can be viewed as multiplying the infinite-duration tone in Fig. 11a by an ideal rectangular (switching) function in Fig. 11b. The frequency domain description of the ideal rectangular function is of the form $\sin f/\pi$. In Fig. 11c, the result of the multiplication yields a tone, cos $2\pi f t$, that is time-duration-limited in the interval $-T/2 < t < T/2$. The resulting spectrum is obtained by convolving the spectral impulses in part a with the $(\sin f)/f$ function in part b, yielding the broadened spectrum in part c. It is further seen that, if the signaling occurs at a faster rate characterized by the rectangle of shorter duration in part d, the resulting spectrum of the signal in part e exhibits greater spectral broadening. The changing state of a fading channel is somewhat analogous to the keying on and off of digital signals. The channel behaves like a switch, turning the signal “on and off.” The greater the rapidity of the change in channel state, the greater the spectral broadening of the received signals. The analogy is not exact because the on and off switching of signals may result in phase discontinuities, but the typical multipath-scatterer environment induces phase-continuous effects.

---

**DEGRADATION CATEGORIES DUE TO TIME VARIANCE VIEWED IN THE DOPPLER SHIFT DOMAIN**

A channel is referred to as fast fading if the symbol rate, $1/T_s$ (approximately equal to the signaling rate or bandwidth $W$) is less than the fading rate, $1/T_0$ (approximately equal to $f_d$); that is, fast fading is characterized by

$$W < f_d$$ \hspace{1cm} (22a)

or

$$T_s > T_0$$ \hspace{1cm} (22b)

Conversely, a channel is referred to as slow fading if the signaling rate is greater than the fading rate. Thus, in order to avoid signal distortion caused by fast fading, the channel must be made to exhibit slow fading by ensuring that the signaling rate must exceed the channel fading rate. That is,

$$W > f_d$$ \hspace{1cm} (23a)

or

$$T_s < T_0$$ \hspace{1cm} (23b)

In Eq. 14, it was shown that due to signal dispersion, the coherence bandwidth, $f_0$, sets an upper limit on the signaling rate which can be used without suffering frequency-selective distortion. Similarly, Eq. 23 shows that due to Doppler spreading, the channel fading rate, $f_d$, sets a lower limit on the signaling rate that can be used without suffering fast fading distortion. For HF communication systems, when teletype or Morse code messages were transmitted at a low data rate, the channels were often fast fading. However, most present-day terrestrial mobile radio channels can generally be characterized as slow fading.

Equation 23 does not go far enough in describing what we desire of the channel. A better way to state the requirement for mitigating the effects of fast fading would be that we desire $W >> f_d$ (or $T_s << T_0$). If this condition is not satisfied, the random frequency modulation (FM) due to varying Doppler shifts will limit the system performance significantly. The Doppler effect yields an irreducible error rate that cannot be overcome by simply increasing $E_b/N_0$ [25]. This irreducible error rate is most pronounced for any modulation that involves switching the carrier phase. A single specular Doppler path, without scatterers, registers an instantaneous frequency

---

![Figure 11. Analogy between spectral broadening in fading and spectral broadening in keying a digital signal.](image-url)
shift, classically calculated as \( f_d = V/\lambda \). However, a combination of specular and multipath components yields a rather complex time dependence of instantaneous frequency which can cause much larger frequency swings than \( \pm V/\lambda \) when detected by an instantaneous frequency detector (a nonlinear device) [26]. Ideally, coherent demodulators that lock onto and track the information signal should suppress the effect of this FM noise and thus cancel the impact of Doppler shift. However, for large values of \( f_d \), carrier recovery becomes a problem because very wideband (relative to the data rate) phase-locked loops need to be designed. For voice-grade applications with bit error rates of \( 10^{-3} \) to \( 10^{-4} \), a large value of Doppler shift is considered to be on the order of 0.01 \( \times \) \( W \).

Therefore, to avoid fast fading distortion and the Doppler-induced irreducible error rate, the signaling rate should exceed the fading rate by a factor of 100–200 [27]. The exact factor depends on the signal modulation, receiver design, and required error rate [3, 26–29]. Davarian [29] showed that a frequency-tracking loop can help to lower, but not completely remove, the irreducible error rate in a mobile system by using differential minimum-shift keyed (DMSK) modulation.

**SUMMARY**

In Part I of this article, the major elements that contribute to fading in a communication channel have been characterized. Figure 1 was presented as a guide for the characterization of fading phenomena. Two types of fading, large-scale and small-scale, were described. Two manifestations of small-scale fading (signal dispersion and fading rapidly) were examined, and the examination involved two views, time and frequency. Two degradation categories were defined for dispersion: frequency-selective fading and flat fading. Two degradation categories were defined for fading rapidly: fast and slow. The small-scale fading degradation categories were summarized in Fig. 6. A mathematical model using correlation and power density functions was presented in Fig. 7. This model yields a nice symmetry, a kind of "poetry" to help us view the Fourier transform and duality relationships that describe the fading phenomena. In Part II, mitigation techniques for ameliorating the effects of each degradation category will be treated, and methods that have been implemented in two mobile communication systems will be described.

**REFERENCES**


**BIOGRAPHY**

Bernard Sklar [LSM] has over 40 years of experience in technical design and placement positions at Republic Aviation Corp., Hughes Aircraft, Litton Industries, and The Aerospace Corporation. At Aerospace, he helped develop the MILSTAR satellite system, and was the principal architect for EHF Satellite Data Link Standards. Currently, he is head of advanced systems at Communications Engineering Services, a consulting company he founded in 1984. He has taught engineering courses at several universities, including the University of California, Los Angeles and the University of Southern California, and has presented numerous training programs throughout the world. He has published and presented scores of technical papers. He is the recipient of the 1984 Prize Paper Award from the IEEE Communications Society for his tutorial series on digital communications, and he is the author of the book Digital Communications (Prentice Hall). He is past Chair of the Los Angeles Council IEEE Education Committee. His academic credentials include a B.S. degree in math and science from the University of Michigan, an M.S. degree in electrical engineering from the Polytechnic Institute of Brooklyn, New York, and a Ph.D. degree in engineering from the University of California, Los Angeles.