UNIT 4- MULTIPATH MITIGATION TECHNIQUES

Equalization techniques fall into two broad categories: linear and nonlinear. The linear techniques are generally the simplest to implement and to understand conceptually. However, linear equalization techniques typically suffer from more noise enhancement than nonlinear equalizers and hence are not used in most wireless applications.

Among nonlinear equalization techniques, decision-feedback equalization (DFE) is the most common because it is fairly simple to implement and usually performs well. However, on channels with low SNR, the DFE suffers from error propagation when bits are decoded in error, leading to poor performance. The optimal equalization technique is maximum likelihood sequence estimation (MLSE). Unfortunately, the complexity of this technique grows exponentially with the length of the delay spread, so it is impractical on most channels of interest. However, the performance of the MLSE is often used as an upper bound on performance for other equalization techniques.

Figure below summarizes the different equalizer types along with their corresponding structures and tap updating algorithms.

Equalizers can also be categorized as symbol-by-symbol (SBS) or sequence estimators (SEs). SBS equalizers remove ISI from each symbol and then detect each symbol individually.

All linear equalizers in Figure (as well as the DFE) are SBS equalizers. Sequence estimators detect sequences of symbols, so the effect of ISI is part of the estimation process.

Maximum likelihood sequence estimation is the optimal form of sequence detection, but it is highly complex.

Linear and nonlinear equalizers are typically implemented using a transversal or lattice structure. The transversal structure is a filter with $N-1$ delay elements and N taps featuring tunable complex weights. The lattice filter uses a more complex recursive structure . In exchange for this increased complexity relative to transversal structures, lattice structures often have better numerical stability and convergence properties and greater flexibility in changing their length. This chapter will focus on transversal structures; details on lattice structures and their performance relative to transversal structures can be found.

In addition to the equalizer type and structure, adaptive equalizers require algorithms for updating the filter tap coefficients during training and tracking. Many algorithms have been developed over the years for this purpose. These algorithms generally incorporate trade-offs between complexity, convergence rate, and numerical stability. In the remainder of this chapter – after discussing conditions for ISI-free transmission – we will examine the different equalizer types, their structures, and their update algorithms in more detail. Linear Equalizers If *F_(f)* is not flat, we can use the equalizer *H*eq*(z)* in Figure to reduce ISI. In this section we assume a linear equalizer implemented via a 2*L* +1 = *N*-tap transversal filter:

$$H_{eq}(z) = \sum_{i=-L}^{L} w_i z^{-i}.$$

The length of the equalizer *N* is typically dictated by implementation considerations, since a large *N* entails more complexity and delay. Causal linear equalizers have *wi* = 0 when*i* < 0. For a given equalizer size *N*, the equalizer design must specify (i) the tap weights

{*wi*}*Li*=−*L*

for a given channel frequency response and (ii) the algorithm for updating these tap weights as the channel varies. Recall that our performance metric in wireless systems is probability of error (or outage probability), so for a given channel the optimal choice of equalizer coefficients would be the coefficients that minimize probability of error. Unfortunately, it is extremely difficult to optimize the {*wi*} with respect to this criterion. Since we cannot directly optimize for our desired

performance metric, we must instead use an indirect optimization that balances ISI mitigation with the prevention of noise enhancement, as discussed with regard to the preceding simple analog example. We now describe two linear equalizers: the zero-forcing (ZF) equalizer and the minimum mean-square error (MMSE) equalizer. The former equalizer cancels all ISI but can lead to considerable noise enhancement. The latter technique minimizes the expected mean-squared error between the transmitted symbol and the symbol detected at the equalizer output, thereby providing a better balance between ISI mitigation and noise enhancement. Because of this more favorable balance, MMSE equalizers
tend to have better BER performance than equalizers using the ZF algorithm.

Zero-Forcing (ZF) Equalizers

The samples $\{y_n\}$ input to the equalizer can be represented based on the discretized combined equivalent lowpass impulse response

$$f(t) = h(t) * g^*(-t)$$

as

$$Y(z) = D(z)F(z) + N_g(z),$$

where $N_g(z)$ is the z-transform of the noise samples at the output of the matched filter $G^*_m(1/z^*)$ and

$$F(z) = H(z)G^*_m\left(\frac{1}{z^*}\right) = \sum_n f(nT_s)z^{-n}.$$

The zero-forcing equalizer removes all ISI introduced in the composite response $f(t)$. From above equation we see that the equalizer to accomplish this is given by

$$H_{ZF}(z) = \frac{1}{F(z)}.$$

This is the discrete-time equivalent lowpass equalizer of the analog equalizer described before, and it suffers from the same noise enhancement properties.

Specifically, the power spectrum *N(z)* of the noise samples at the equalizer output is given by

$$N(z) = N_0|G_m^*(1/z^*)|^2|H_{ZF}(z)|^2 = \frac{N_0|G_m^*(1/z^*)|^2}{|F(z)|^2}$$
$$= \frac{N_0|G_m^*(1/z^*)|^2}{|H(z)|^2|G_m^*(1/z^*)|^2} = \frac{N_0}{|H(z)|^2}.$$

We see from above equation that if the channel *H(z)* is sharply attenuated at any frequency within the signal bandwidth of interest – as is common on frequency-selective fading channels – the noise power will be significantly increased. This motivates an equalizer design that better optimizes between ISI mitigation and noise enhancement. One such equalizer is the MMSE equalizer, described in the next section.

The ZF equalizer defined by *H*ZF*(z)* = 1/*F(z)* may not be implementable as a finite– impulse response (FIR) filter. Specifically, it may not be possible to find a finite set of coefficients *w*–*L*, . . . ,*wL* such that

$$w_{-L}z^L + \cdots + w_L z^{-L} = \frac{1}{F(z)}.$$

In this case we find the set of coefficients {*wi*} that best approximates the zero-forcing equalizer. Note that this is not straightforward because the approximation must be valid for all values of *z*. There are many ways we can make this approximation. One technique is to represent *H*ZF*(z)* as an infinite–impulse response (IIR) filter, 1/*F(z)* = _∞ *i*=–∞ *ciz*–*i* , and then set *wi* = *ci* . It can be shown that this minimizes

$$\left|\frac{1}{F(z)} - (w_{-L}z^L + \cdots + w_L z^{-L})\right|^2$$

at *z* = *ejω*. Alternatively, the tap weights can be set to minimize the peak distortion (worstcase ISI). Finding the tap weights to minimize peak distortion is a convex optimization problem that can be solved by standard techniques – for example, the method of steepest descent.

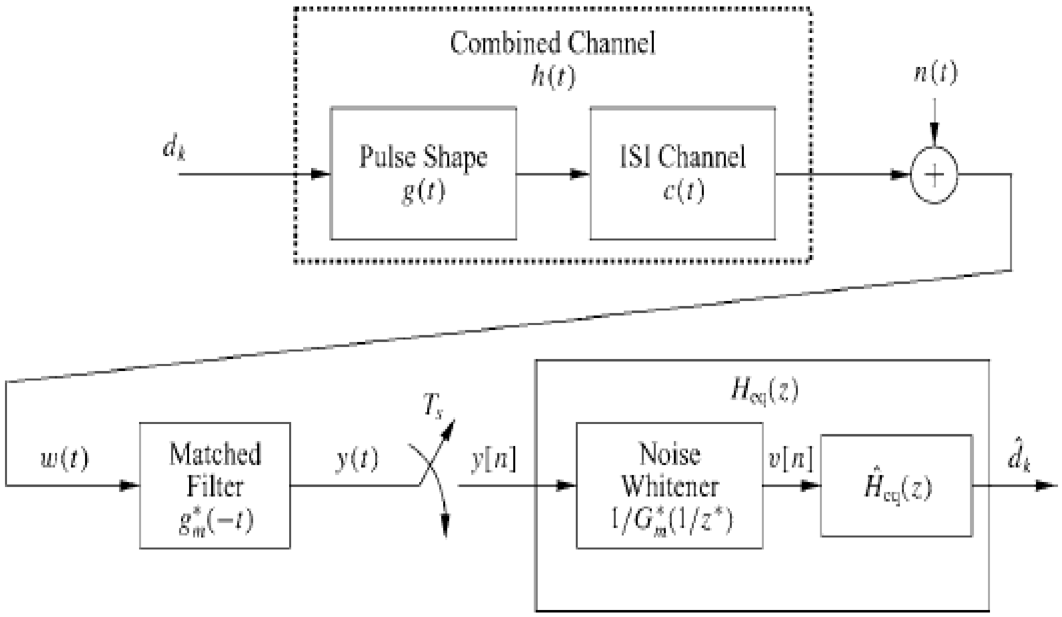
Minimum Mean-Square Error (MMSE) Equalizers

In MMSE equalization, the goal of the equalizer design is to minimize the average meansquare error (MSE) between the transmitted symbol *dk* and its estimate ˆ *dk* at the output of the equalizer. In other words, the {*wi*} are chosen to minimize **E**[*dk*− ˆ *dk*]2. Since the MMSE equalizer is linear, its output ˆ *dk* is a linear combination of the input samples *y*[*k*]:

$$\hat{d}_k = \sum_{i=-L}^{L} w_i y[k-i].$$

As such, finding the optimal filter coefficients {*wi*} becomes a standard problem in linear estimation. In fact, if the noise input to the equalizer is white then we have a standard Weiner filtering problem. However, because of the matched filter *g∗m(−t)* at the receiver front end, the noise input to the equalizer is not white; rather, it is colored with power spectrum *N*0|*G∗m(*1*/z∗)*|2. Therefore, in order to apply known techniques for optimal linear estimation, we expand the filter *H*eq*(z)* into two components – a noise-whitening component and an ISI-removal component ˆ*H*eq*(z)* – as shown in Figure.

The purpose of the noise-whitening filter (as the name indicates) is to whiten the noise so that the noise component output from this filter has a constant power spectrum. Since the



MMSE equalizer with noise-whitening filter.

noise input to this filter has power spectrum $N_0|G^*_m(1/z^*)|^2$, the appropriate noise-whitening filter is $1/G^*_m(1/z^*)$. The noise power spectrum at the output of the noise-whitening filter is then $N_0|G^*_m(1/z^*)|^2/|G^*_m(1/z^*)|^2 = N_0$. Note that the filter $1/G^*_m(1/z^*)$ is not the only filter that will whiten the noise, and another noise-whitening filter with more desirable properties (like stability) may be chosen.

It might seem odd at first to introduce the matched filter $g^*_m(-t)$ at the receiver front end only to cancel its effect in the equalizer. however, that the matched filter is meant to maximize the SNR prior to sampling. By removing the effect of this matched filter via noise whitening after sampling, we merely simplify the design of $\hat{H}_{eq}(z)$ to minimize MSE. In fact, if the noise-whitening filter does not yield optimal performance then its effect would be cancelled by the $\hat{H}_{eq}(z)$ filter design, as we shall see in the case of IIR MMSE equalizers.

We assume that the filter $\hat{H}_{eq}(z)$, with input v_n, is a linear filter with $N = 2L + 1$ taps:

$$\hat{H}_{\text{eq}}(z) = \sum_{i=-L}^{L} w_i z^{-i}.$$

Our goal is to design the filter coefficients $\{w_i\}$ so as to minimize $\mathbf{E}[d_k - \hat{d}_k]^2$. This is the same goal as for the total filter $H_{eq}(z)$ – we just added the noise-whitening filter to make solving for these coefficients simpler. Define $\mathbf{v}^T = (v[k+L], v[k+L-1], \ldots, v[k-L]) = (v_{k+L}, v_{k+L-1}, \ldots, v_{k-L})$ as the row vector of inputs to the filter $\hat{H}_{eq}(z)$ used to obtain the filter output $\hat{d}_k$ and define $\mathbf{w}^T = (w_{-L}, \ldots, w_L)$ as the row vector of filter coefficients. Then

$$\hat{d}_k = \mathbf{w}^T\mathbf{v} = \mathbf{v}^T\mathbf{w}.$$

Thus, we want to minimize the mean-square error

$$J = \mathrm{E}[d_k - \hat{d}_k]^2 = \mathrm{E}[\mathbf{w}^T\mathbf{v}\mathbf{v}^H\mathbf{w}^* - 2\,\mathrm{Re}\{\mathbf{v}^H\mathbf{w}^* d_k\} + |d_k|^2].$$

Define $\mathbf{M_v} = \mathbf{E}[\mathbf{v}\mathbf{v}^H]$ and $\mathbf{v_d} = \mathbf{E}[\mathbf{v}^H d_k]$. The matrix $\mathbf{M_v}$ is an $N \times N$ Hermitian matrix,

and $\mathbf{v_d}$ is a length-N row vector. Assume $\mathbf{E}|d_k|^2 = 1$. Then the MSE J is

$$J = \mathbf{w}^T\mathbf{M}_\mathbf{v}\mathbf{w}^* - 2\,\mathrm{Re}\{\mathbf{v_d}\mathbf{w}^*\} + 1.$$

LINEAR EQUALIZERS

We obtain the optimal tap vector $\mathbf{w}$ by setting the gradient $\nabla_\mathbf{w} J = 0$ and solving for $\mathbf{w}$. the gradient is given by

$$\nabla_\mathbf{w} J = \left(\frac{\partial J}{\partial w_{-L}}, \ldots, \frac{\partial J}{\partial w_L}\right) = 2\mathbf{w}^T\mathbf{M}_\mathbf{v} - 2\mathbf{v_d}.$$

Setting this to zero yields $\mathbf{w}^T\mathbf{M_v} = \mathbf{v_d}$ or, equivalently, that the optimal tap weights are given by

$$\mathbf{w}_{\text{opt}} = (\mathbf{M}_\mathbf{v}^T)^{-1}\mathbf{v}_\mathbf{d}^T.$$

Note that solving for **w**opt requires a matrix inversion with respect to the filter inputs. Thus, the complexity of this computation is quite high, typically on the order of *N*2 to *N*3 operations.

Substituting in these optimal tap weights, we obtain the minimum mean-square error as

$$J_{\min} = 1 - \mathbf{v_d} \mathbf{M}_{\mathbf{v}}^{-1} \mathbf{v}_{\mathbf{d}}^{H}.$$

Solving for ˆ*H* eq*(z)*, we obtain

$$\hat{H}_{\mathrm{eq}}(z) = \frac{G_m^*(1/z^*)}{F(z) + N_0}.$$

Since the MMSE equalizer consists of the noise-whitening filter 1*/G∗m(*1*/z∗)* plus the ISIremoval component ˆ*H* eq*(z)*, it follows that the full MMSE equalizer (when it is not restricted to be finite length) becomes

$$H_{\mathrm{eq}}(z) = \frac{\hat{H}_{\mathrm{eq}}(z)}{G_m^*(1/z^*)} = \frac{1}{F(z) + N_0}.$$

There are three interesting things to note about this result. First of all, the ideal infinite length MMSE equalizer cancels out the noise-whitening filter. Second, this infinite-length equalizer is identical to the ZF filter except for the noise term *N*0, so in the absence of noise the two equalizers are equivalent. Finally, this ideal equalizer design clearly shows a balance between inverting the channel and noise enhancement: if *F(z)* is highly attenuated at some frequency, then the noise term *N*0 in the denominator prevents the noise from being significantly enhanced by the equalizer. Yet at frequencies where the noise power spectral density *N*0 is small compared to the composite channel *F(z)*, the equalizer effectively inverts *F(z)*. For the equalize it can be shown that the minimum MSE can be expressed in terms of the folded spectrum *F_(f)* as

$$J_{\min} = T_s \int_{-.5/T_s}^{.5/T_s} \frac{N_0}{F_\Sigma(f) + N_0}\, df.$$

Maximum Likelihood Sequence Estimation

Maximum-likelihood sequence estimation avoids the problem of noise enhancement because it doesn't use an equalizing filter: instead it estimates the sequence of transmitted symbols. The structure of the MLSE is the same as in Figure except that the equalizer $H_{eq}(z)$ and decision device are replaced by the MLSE algorithm. Given the combined pulse-shaping filter and channel response $h(t)$, the MLSE algorithm chooses the input sequence $\{d_k\}$ that maximizes the likelihood of the received signal $w(t)$. We now investigate this algorithm in more detail.

Using a Gram–Schmidt orthonormalization procedure, we can express $w(t)$ on a time interval $[0, LT_s]$ as

$$w(t) = \sum_{n=1}^{N} w_n \phi_n(t),$$

where $\{\varphi_n(t)\}$ form a complete set of orthonormal basis functions. The number N of functions in this set is a function of the channel memory, since $w(t)$ on $[0, LT_s]$ depends on $d_0, \ldots, d_L$. With this expansion we have

$$w_n = \sum_{k=-\infty}^{\infty} d_k h_{nk} + \nu_n = \sum_{k=0}^{L} d_k h_{nk} + \nu_n,$$

$$h_{nk} = \int_0^{LT_s} h(t - kT_s)\phi_n^*(t)\, dt$$

$$\nu_n = \int_0^{LT_s} n(t)\phi_n^*(t)\, dt.$$

MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION

The ν_n are complex Gaussian random variables with mean zero and covariance $\mathbf{E}[\nu_n^* \nu_m] = N_0\delta[n - m]$. Thus, $\mathbf{w}_N = (w_1, \ldots, w_N)$

has a multivariate Gaussian distribution:

$$p(\mathbf{w}^N \mid d^L, h(t)) = \prod_{n=1}^{N} \left[\frac{1}{\pi N_0} \exp\left[-\frac{1}{N_0} \left| w_n - \sum_{k=0}^{L} d_k h_{nk} \right|^2 \right] \right].$$

$$\begin{aligned}
\hat{d}^L &= \arg\max[\log p(\mathbf{w}^N \mid d^L, h(t))] \\
&= \arg\max\left[-\sum_{n=1}^{N} \left| w_n - \sum_k d_k h_{nk} \right|^2 \right] \\
&= \arg\max\left[-\sum_{n=1}^{N} |w_n|^2 + \sum_{n=1}^{N} \left(w_n^* \sum_k d_k h_{nk} + w_n \sum_k d_k^* h_{nk}^* \right) \right. \\
&\qquad \left. - \sum_{n=1}^{N} \left(\sum_k d_k h_{nk} \right) \left(\sum_m d_m^* h_{nm}^* \right) \right] \\
&= \arg\max\left[2\,\mathrm{Re}\left\{ \sum_k d_k^* \sum_{n=1}^{N} w_n h_{nk}^* \right\} - \sum_k \sum_m d_k d_m^* \sum_{n=1}^{N} h_{nk} h_{nm}^* \right].
\end{aligned}$$

Decision-Feedback Equalization

The DFE consists of a feedforward filter *W(z)* with the received sequence as input (similar to the linear equalizer) followed by a feedback filter *V(z)* with the previously detected sequence as input. The DFE structure is shown in Figure below. In effect, the DFE determines the ISI contribution from the detected symbols {ˆˆ*dk*} by passing them through a feedback filter that approximates the composite channel *F(z)* convolved with the feedforward filter *W(z)*. The resulting ISI is then subtracted from the incoming symbols. Since the feedback filter *V(z)* in Figure sits in a feedback loop, it must be strictly causal or else the system is unstable. The feedback filter of the DFE does not suffer from noise enhancement because it estimates the channel frequency response rather than its inverse. For channels with deep spectral nulls, DFEs generally perform much better than linear equalizers.

Assuming that *W(z)* has *N*1 +1 taps and *V(z)* has *N*2 taps, we can write the DFE output as

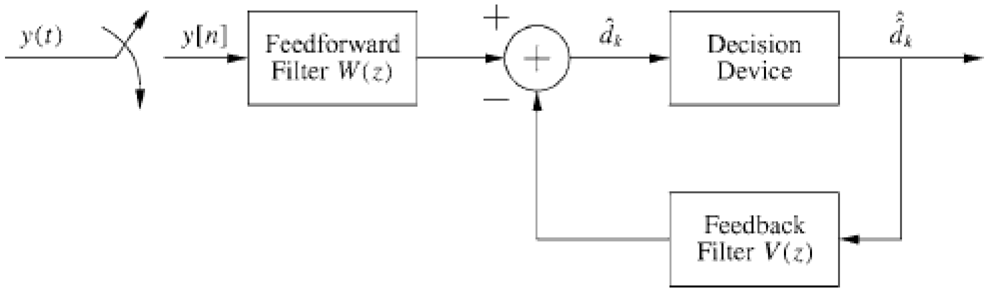


Fig: Decision-feedback equalizer structure.

$$\hat{d}_k = \sum_{i=-N_1}^{0} w_i y[k-i] - \sum_{i=1}^{N_2} v_i \hat{\hat{d}}_{k-i}.$$

Adaptive Equalizers: Training and Tracking

All of the equalizers described so far are designed based on a known value of the combined impulse response $h(t) = g(t) * c(t)$. Since the channel $c(t)$ in generally not known when the receiver is designed, the equalizer must be tunable so it can adjust to different values of $c(t)$. Moreover, since in wireless channels $c(t) = c(\tau, t)$ will change over time, the system must periodically estimate the channel $c(t)$ and update the equalizer coefficients accordingly. This process is called *equalizer training* or *adaptive equalization* . The equalizer can also use the detected data to adjust the equalizer coefficients, a process known as *equalizer tracking.* However, *blind equalizers* do not use training: they learn the channel response via the detected data only.

During training, the coefficients of the equalizer are updated at time k based on a known training sequence $[d_{k-M}, \ldots, d_k]$ that has been sent over the channel. The length $M+1$ of the training sequence depends on the number of equalizer coefficients that must be determined and the convergence speed of the training algorithm. Note that the equalizer must be retrained when the channel decorrelates – that is, at least every T_c seconds, where T_c is the channel coherence time. Thus, if the training algorithm is slow relative to the channel coherence time then the channel may change before the equalizer can learn the channel.

Specifically, if $(M + 1)T_s > T_c$ then the channel will decorrelate before the equalizer has finished training. In this case equalization is not an effective countermeasure for ISI, and some other technique (e.g., multicarrier modulation or CDMA) is needed.

Principle of Diversity

For Additive White Gaussian Noise (AWGN) channels, such an approach can be quite reasonable: the Bit Error Rate (BER) decreases exponentially as the Signal-to-Noise Ratio (SNR) increases, and a 10-dB SNR leads to BERs on the order of 10−4. However, in Rayleigh fading the BER decreases only linearly with the SNR. We thus would need an SNR on the order of 40 dB in order to achieve a 10−4 BER, which is clearly unpractical. The reason for this different performance is the fading of the channel: the BER is mostly determined by the probability of channel attenuation being large, and thus of the instantaneous SNR being low. A way to improve the BER is thus to change the effective channel statistics – i.e., to make sure that the SNR has a smaller probability of being low. Diversity is a way to achieve this. The principle of diversity is to ensure that the same information reaches the receiver (RX) on statistically independent channels. Consider the simple case of an RX with two antennas. The antennas are assumed to be far enough from each other that small-scale fading is independent at the two antennas. The RX always chooses the antenna that has instantaneously larger receive power.1 As the signals are statistically independent, the probability that both antennas are in a fading dip *simultaneously* is low – certainly lower than the probability that one antenna is in a fading dip. The diversity thus changes the SNR statistics at the detector input.

Microdiversity

As mentioned in the introduction, the basic principle of diversity is that the RX has multiple copies of the transmit signal, where each of the copies goes through a statistically independent channel. This section describes different ways of obtaining these statistically independent copies. We concentrate on methods that can be used to combat small-scale fading, which are therefore called "microdiversity." The five most common methods are as follows:

1. *Spatial diversity*: several antenna elements separated in space.
2. *Temporal diversity*: transmission of the transmit signal at different times.
3. *Frequency diversity*: transmission of the signal on different frequencies.
4. *Angular diversity*: multiple antennas (with or without spatial separation) with different antenna patterns.
5. *Polarization diversity*: multiple antennas with different polarizations (e.g., vertical and horizontal).

When we speak of antenna diversity, we imply that there are multiple antennas at the *receiver*.

Only in Section will we discuss how multiple *transmit* antennas can be exploited to improve performance. The following important equation will come in handy: Consider the correlation coefficient of

two signals that have a temporal separation τ and a frequency separation $f1 - f2$. As shown in Appendix 13.A (see *www.wiley.com/go/molisch*), the correlation coefficient is

$$\rho_{xy} = \frac{J_0^2(k_0 v \tau)}{1 + (2\pi)^2 S_\tau^2 (f_2 - f_1)^2}$$

Spatial Diversity

Spatial diversity is the oldest and simplest form of diversity. Despite (or because) of this, it is also the most widely used. The transmit signal is received at several antenna elements, and the signals from these antennas are then further processed according to the principles that will be described. But, irrespective of the processing method, performance is influenced by correlation of the signals between the antenna elements. A large correlation between signals at antenna elements is undesirable, as it decreases the effectiveness of diversity. A first important step in designing diversity antennas is thus to establish a relationship between antenna spacing and the correlation coefficient. This relationship is different for BS antennas and MS antennas, and thus will be treated separately.

1. *MS in cellular and cordless systems*: it is a standard assumption that waves are incident from all directions at the MS. Thus, points of constructive and destructive interference of Multi approximately $\lambda/4$ apart. This Path Components (MPCs) – i.e., points where we have high and low received power, respectively – are spaced is therefore the distance that is required for decorrelation of received signals *BS in cordless systems and WLANs*: in a first approximation, the angular distribution of incident radiation at indoor BSs is also uniform – i.e., radiation is incident with equal strength from all directions. Therefore, the same rules apply as for MSs.

3. *BSs in cellular systems*: for a cellular BS, the assumption of uniform directions of incidence is no longer valid. Interacting Objects (IOs) are typically concentrated around the MS . Since all waves are incident essentially from one direction, the correlation coefficient (for a given distance between antenna elements *da*) is much higher. Expressed differently,

the antenna spacing required to obtain sufficient decorrelation increases. To get an intuitive insight, we start with the simple case when there are only two MPCs whose wave vectors are at an angle α with respect to each other.

It is obvious that the distance between the maxima and minima of the interference pattern is larger the smaller α is. For very small α, the connection line between antenna elements lies on a “ridge” of the interference pattern and antenna elements are completely correlated. Numerical evaluations of the correlation coefficient as a function of antenna spacing are shown in Figure below . The first column shows the results for rectangular angular power spectra; the results for Gaussian distributions are shown in the second column. We can see that antenna spacing has to be on the order of 2–20 wavelengths for angular spreads between 1◦ and 5◦ in order to achieve decorrelation. We also find that it is mostly rms angular spread that determines the required antenna spacing, while the shape of the angular power spectrum has only a minor influence.

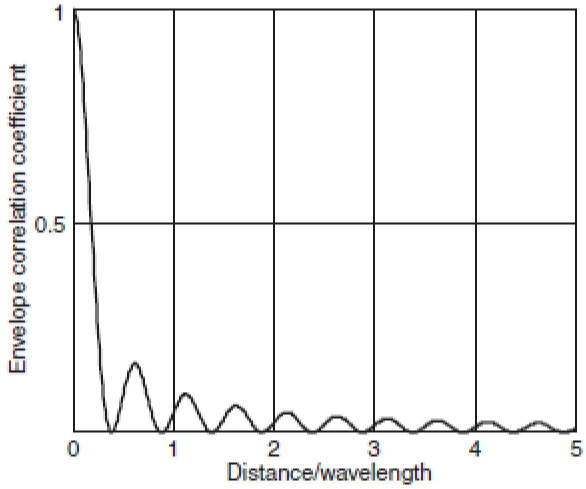


Figure Envelope correlation coefficient as a function of antenna separation.

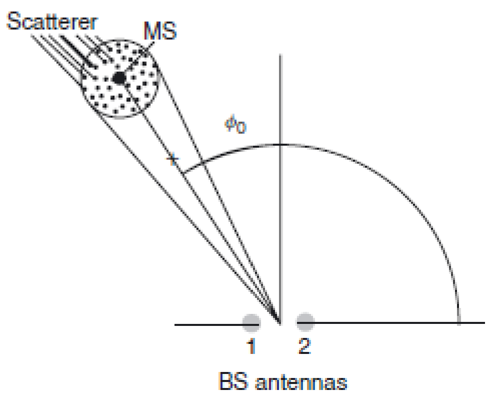


Figure Scatterers concentrated around the mobile station.

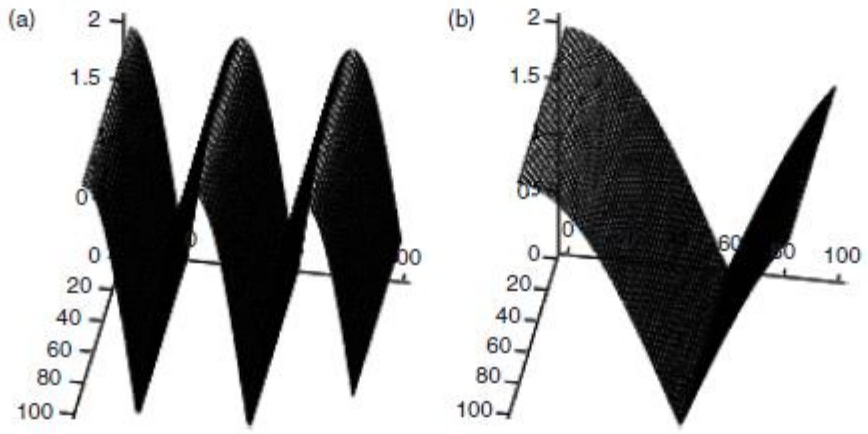


Figure Interference pattern of two waves with 45° (a) and 15° (b) angular separation.

Temporal Diversity

As the wireless propagation channel is time variant, signals that are received at different times are uncorrelated. For "sufficient" decorrelation, the temporal distance must be at least $1/(2\nu_{max})$, where ν_{max} is the maximum Doppler frequency. In a static channel, where neither transmitter (TX), RX, nor the IOs are moving, the channel state is the same at all times. Such a situation can occur, e.g., for WLANs. In such a case, the correlation coefficient is $\rho = 1$ for all time intervals, and temporal diversity is useless.

Temporal diversity can be realized in different ways:

1. *Repetition coding*: this is the simplest form. The signal is repeated several times, where the repetition intervals are long enough to achieve decorrelation. This obviously achieves diversity, but is also highly bandwidth inefficient. Spectral efficiency decreases by a factor that is equal to the number of repetitions.

2. *Automatic Repeat reQuest (ARQ)*: here, the RX sends a message to the TX to indicate whether it received the data with sufficient quality .If this is not the case, then the transmission is repeated (after a wait period that achieves decorrelation). The spectral efficiency of ARQ is better than that of repetition coding, since it

requires multiple transmissions only when the first transmission occurs in a bad fading state, while for repetition coding, retransmissions occur always. On the downside, ARQ requires a feedback channel.

3. *Combination of interleaving and coding*: a more advanced version of repetition coding is forward error correction coding with interleaving. The different symbols of a codeword are transmitted at different times, which increases the probability that at least some of them arrive with a good SNR. The transmitted codeword can then be reconstructed.

Frequency Diversity

In frequency diversity, the same signal is transmitted at two (or more) different frequencies. If these frequencies are spaced apart by more than the coherence bandwidth of the channel, then their fading is approximately independent, and the probability is low that the signal is in a deep fade at both frequencies simultaneously. For an exponential PDP, the correlation between two frequencies can be obtained from Eq. (13.4) by setting the numerator to unity as the signals at the two frequencies occur at the same time. Thus

$$\rho = \frac{1}{1 + (2\pi)^2 S_\tau^2 (f_2 - f_1)^2}$$

This again confirms that the two signals have to be at least one coherence bandwidth apart from each other. Figure shows ρ as a function of the spacing between the two frequencies. For a more general discussion of frequency correlation would greatly decrease spectral efficiency. Rather, information is spread over a large bandwidth, so that small parts of the information are conveyed by different frequency components. The RX can then sum over the different frequencies to recover the original information. This spreading can be done by different methods:

• *Compressing the information in time*: – i.e., sending short bursts that each occupy a large bandwidth – TDMA.

• *Code Division Multiple Access* (*CDMA*):

• Multicarrier CDMA and coded orthogonal frequency division multiplexing.

• *Frequency hopping in conjunction with coding*: different parts of a codeword are transmitted on different carrier frequencies.

These methods allow the transmission of information without wasting bandwidth. For the moment, we just stress that the use of frequency diversity requires the channel to be frequency selective. In other words, frequency diversity (delay dispersion) can be exploited by the system to make it more robust, and decrease the effects of fading. This seems to be a contradiction to the results of Chapter where we had shown that frequency selectivity leads to an *increase* of the BER, and even an error floor. The reason for this discrepancy is that Chapter considered a very simple RX that takes no measures to combat (or exploit) the effects of frequency selectivity.

Angle Diversity

A fading dip is created when MPCs, which usually come from different directions, interfere destructively. If some of these waves are attenuated or eliminated, then the location of fading dips changes

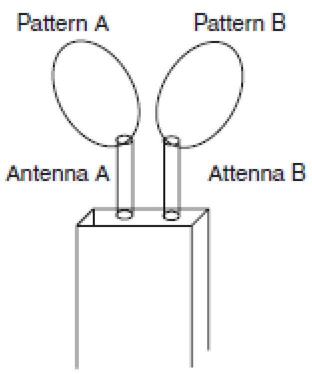


Figure Angle diversity for closely spaced antennas.

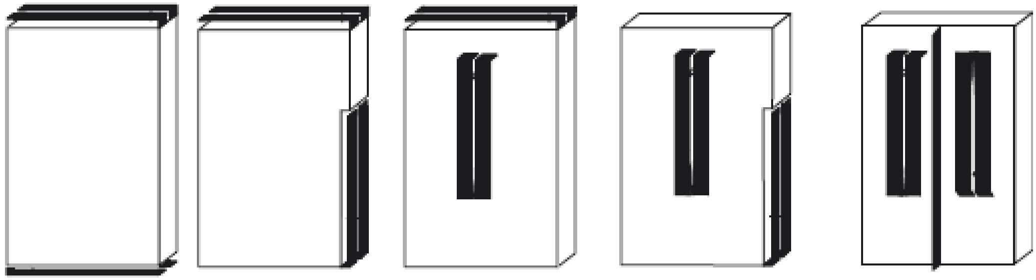


Figure Configurations of diversity antennas at a mobile station.

Polarization Diversity

Horizontally and vertically polarized MPCs propagate differently in a wireless channel,6 as the reflection and diffraction processes depend on polarization. Even if the transmit antenna only sends signals with a single polarization, the propagation effects in the channel lead to depolarization so that both polarizations arrive at the RX. The fading of signals with different polarizations is statistically independent. Thus, receiving both polarizations using a dual-polarized antenna, and processing the signals separately, offers diversity. This diversity can be obtained without any requirement for a minimum distance between antenna elements. Let us now consider more closely the situation where the transmit signal is vertically polarized, while the signal is received in both vertical and horizontal polarization. In that case, fading of the two received signals is independent, but the average received signal strength in the two diversity branches is *not* identical. Depending on the environment, the horizontal (i.e., cross-polarized) component is some 3–20 dB weaker than the vertical (co-polarized) component. As we will see later on, this has an important impact on the effectiveness of the diversity scheme. Various antenna arrangements have been proposed in order to mitigate this problem.

It has also been claimed that the diversity order that can be achieved with polarization diversity is up to 6: three possible components of the E-field and three components of the H-field can all be exploited.

However, propagation characteristics as well as practical considerations prevent a full exploitation of that diversity order especially for outdoor situations.

This is usually not a serious restriction for diversity systems, as we will see later on that going from diversity order 1 (i.e., no diversity) to diversity order 2 gives larger benefits than increasing the diversity order from 2 to higher values. However, it is an important issue for Multiple Input Multiple Output (MIMO) systems.

Macrodiversity

The previous section described diversity methods that combat small-scale fading – i.e., the fading created by interference of MPCs. However, not all of these diversity methods are suitable for combating large-scale fading, which is created by shadowing effects. Shadowing is almost independent of transmit frequency and polarization, so that frequency diversity or polarization diversity are not effective. Spatial diversity (or equivalently, temporal diversity with moving TX/RX) can be used, but we have to keep in mind that the correlation distances for large-scale fading are on the order of tens or hundreds of meters. In other words, if there is a hill between the TX and RX, adding antennas on either the BS or the MS does not help to eliminate the shadowing caused by this hill. Rather, we should use a separate base station (BS2) that is placed in such a way that the hill is not in the connection line between the MS and BS2. This in turn implies a large distance between BS1 and BS2, which gives rise to the word *macrodiversity*. The simplest method for macrodiversity is the use of *on-frequency repeaters* that receive the signal and retransmit an amplified version of it. *Simulcast* is very similar to this approach; the same signal is transmitted

simultaneously from different BSs. In cellular applications the two BSs should be synchronized, and transmit the signals intended for a specific user in such a way that the two waves arrive at the RX almost simultaneously (timing advance) can only be obtained if the runtimes from the two BSs to the MS are known. Generally speaking, it is desirable that the synchronization error is no larger than the delay dispersion that the RX can handle. Especially critical are RXs in regions where the strengths of the signals from the two BSs are approximately equal.
Simulcast is also widely used for broadcast applications, especially digital TV. In this case, the exact synchronization of all possible RXs is not possible – each RX would require a different timing advance from the TXs. A disadvantage of simulcast is the large amount of signaling information that has to be carried on landlines. Synchronization information as well as transmit data have to be transported on landlines (or microwave links) to the BSs. This used to be a serious problem in the early days of digital mobile telephony, but the current wide availability of fiber-optic links has made this less of an issue.
The use of on-frequency repeaters is simpler than that of simulcast, as no synchronization is required. On the other hand, delay dispersion is larger, because (i) the runtime from BS to repeater, and repeater to MS is larger (compared with the runtime from a second BS), and (ii) the repeater itself introduces additional delays due to the group delays of electronic components, filters, etc.

Combination of Signals

Now we turn our attention to the question of how to use diversity signals in a way that improves the total quality of the signal that is to be detected. To simplify the notation, we speak here only about the combination of signals from different antenna signals at the RX. However, the mathematical methods remain valid for other types of diversity signals as well. In general, we can distinguish two ways of exploiting signals from the multiple diversity branches:

1. ***Selection diversity***, where the "best" signal copy is selected and processed (demodulated and decoded), while all other copies are discarded. There are different criteria for what constitutes the "best" signal.

2. ***Combining diversity,*** where all copies of the signal are combined (before or after the demodulator), and the combined signal is decoded. Again, there are different algorithms for combination of the signals. Combining diversity leads to better performance, as all available information is exploited. On the downside, it requires a more complex RX than selection diversity. In most RXs, all processing is done in the baseband. Thus, an RX with combining diversity needs to downconvert all available signals, and combine them appropriately in the baseband. Thus, it requires Nr antenna elements as well as Nr complete Radio Frequency (RF) (downconversion) chains. An RX with selection diversity requires only *one* RF chain, as it processes only a single received signal at a time. In the following, we give a more detailed description of selection (combination) criteria and algorithms. We assume that different signal copies undergo statistically independent fading – this greatly simplifies the discussion of both the intuitive explanations and the mathematics of the signal combination. In these considerations, we also have to keep in mind that the gain of multiple antennas is due to two effects**:** ***diversity gain*** **and** ***beamforming gain***. Diversity gain reflects the fact that it is improbable that several antenna elements are in a fading dip simultaneously; the probability for very low signal levels is thus decreased by the use of multiple antenna elements. Beamforming gain reflects the fact that (for combining diversity) the combiner performs an averaging over the noise at different antennas. Thus, even if the signal levels at all antenna elements are identical *Selection Diversity*

Received-Signal-Strength-Indication-Driven Diversity

In this method, the RX selects the signal with the largest instantaneous power (or *Received Signal Strength Indication – RSSI*), and processes it further. This method requires Nr antenna elements, Nr RSSI sensors, and a Nr-to-1 multiplexer (switch), but only one RF chain (see Figure above). The method allows simple tracking of the selection criterion even in fast-fading channels. Thus, we can switch to a better antenna as soon as the RSSI becomes higher there. 1. If the BER is determined by noise, then RSSI-driven diversity is the best of all the selection diversity methods, as maximization of the RSSI also maximizes the SNR.

2. If the BER is determined by co-channel interference, then RSSI is no longer a good selection criterion. High receive power can be caused by a high level of interference, such that the RSSI criterion makes the system select branches with a low signal-to-interference ratio. This is especially critical when interference is caused mainly by one dominant interferer – a situation that is typical for Frequency Division Multiple Access (FDMA) or TDMA systems.

3. Similarly, RSSI-driven diversity is suboptimum if the errors are caused by the frequency selectivity of the channel. RSSI-driven diversity can still be a reasonable approximation.

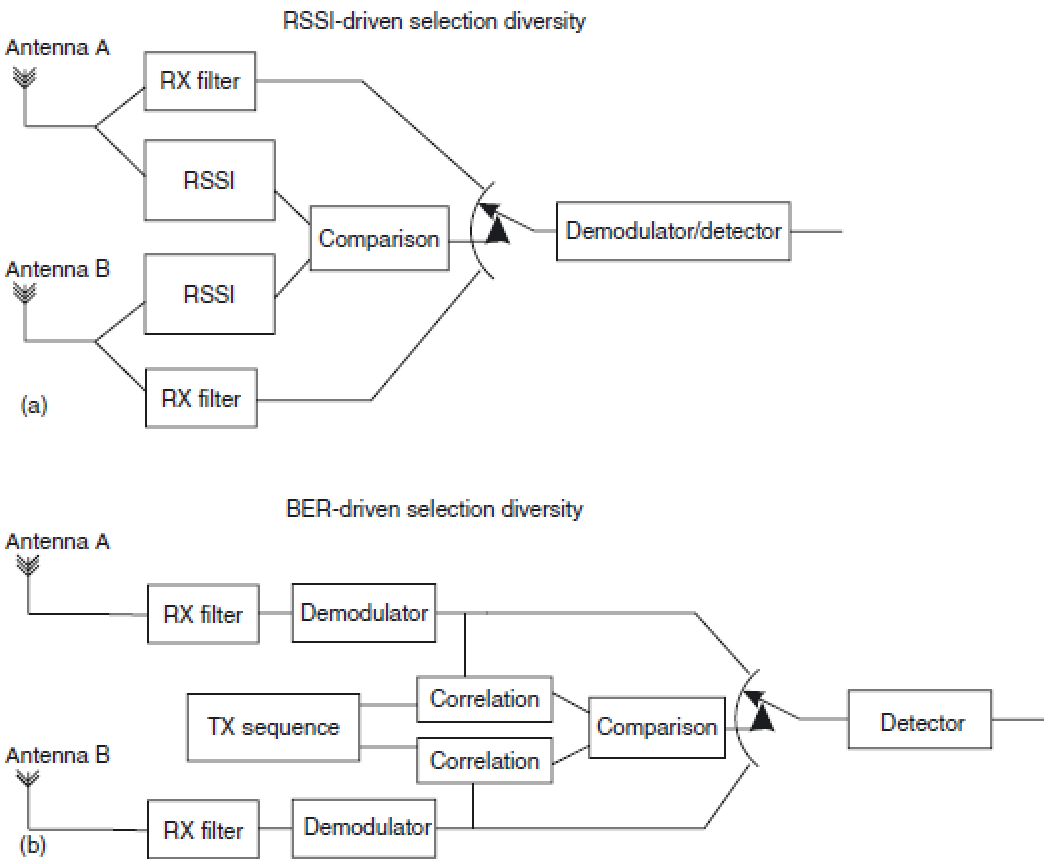


Figure Selection diversity principle: (a) Received-signal-strength-indication-controlled diversity. (b) Bit-error-rate-controlled diversity.

Combining Diversity

Basic Principle

Selection diversity wastes signal energy by discarding $(N_r - 1)$ copies of the received signal. This drawback is avoided by combining diversity, which exploits *all* available signal copies. Each signal copy is multiplied by a (complex) weight and then added up. Each complex weight w_n^* can be thought of as consisting of a phase correction,10 plus a (real) weight for the amplitude:

• Phase correction causes the *signal amplitudes* to add up, while, on the other hand, noise is added incoherently, so that *noise powers* add up.

• For amplitude weighting, two methods are widely used: *Maximum Ratio Combining* (MRC) weighs all signal copies by their amplitude. It can be shown that (using some assumptions) this is an optimum combination strategy. An alternative is *Equal Gain Combining* (EGC), where all amplitude weights are the same (in other words, there is no weighting, but just a phase correction).

Maximum Ratio Combining

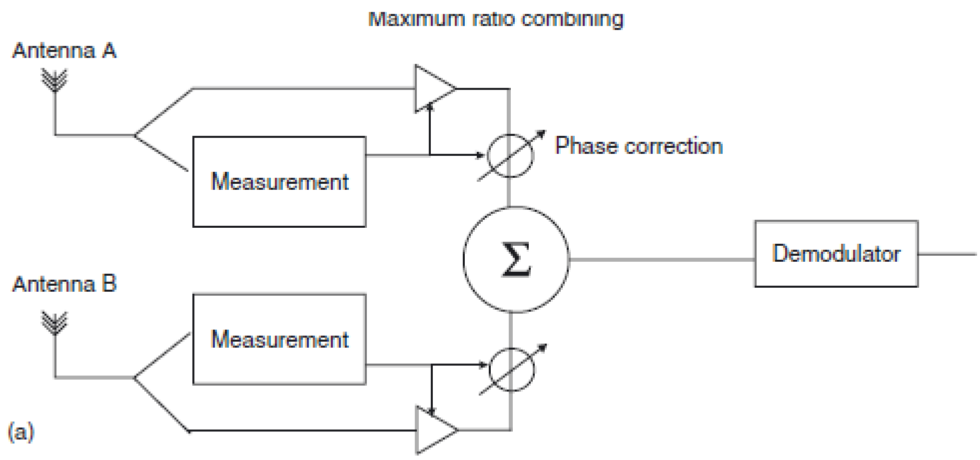
MRC compensates for the phases, and weights the signals from the different antenna branches according to their SNR. This is the optimum way of combining different diversity branches – if several assumptions are fulfilled. Let us assume a propagation channel that is slow fading and flat fading. The only disturbance is AWGN. Under these assumptions, each channel realization can be written as a time-invariant filter with impulse response:

$$h_n(\tau) = \alpha_n \delta(\tau)$$

where *αn* is the (instantaneous) gain of diversity branch *n*. This signals at the different branches are multiplied with weights *w∗n* and added up, so that the SNR becomes

$$\frac{\left|\sum_{n=1}^{N} w_n^* \alpha_n\right|^2}{P_n \sum_{n=1}^{N} |w_n|^2}$$

i.e., the signals are phase-corrected (remember that the received signals are multiplied with and weighted by the amplitude. We can then easily see that in that case the output SNR of the diversity combiner is the *sum* of the branch SNRs:



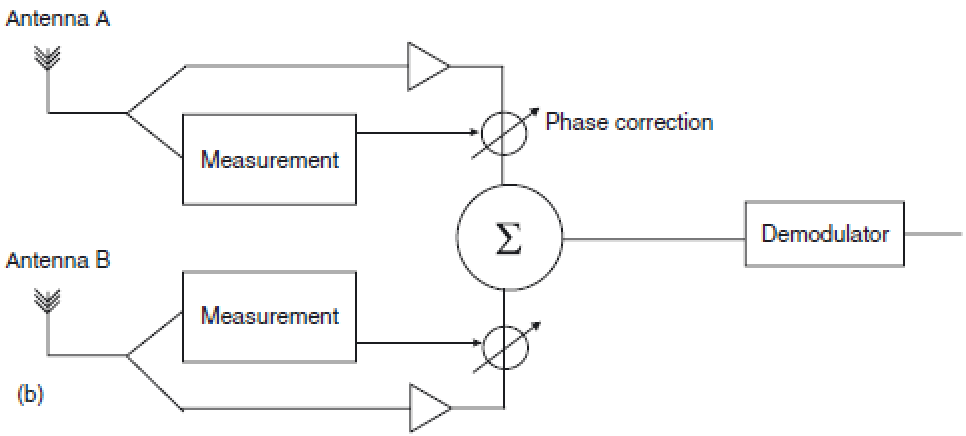


Figure Combining diversity principle: (a) maximum ratio combining, (b) equal gain combining.

Equal Gain Combining

For EGC, we find that the SNR of the combiner output is

$$\gamma_{\mathrm{ECG}} = \frac{\left(\sum_{n=1}^{N_r} \sqrt{\gamma_n}\right)^2}{N_r}$$

where we have assumed that noise levels are the same on all diversity branches. The mean SNR of the combiner output can be found to be

$$\overline{\gamma}_{\mathrm{EGC}} = \overline{\gamma}\left(1 + (N_r - 1)\frac{\pi}{4}\right)$$

if all branches suffer from Rayleigh fading with the same mean SNR γ . Remember that we only assume here that the *mean* SNR is the same in all branches, while instantaneous branch SNRs (representing different channel realizations) can be different. It is quite remarkable that EGC performs worse than MRC by only a factor $\pi/4$ (in terms of mean SNR). The performance difference between EGC and MRC becomes bigger when mean branch SNRs are also different.

Hybrid Selection – Maximum Ratio Combining

A compromise between selection diversity and full signal combining is the so-called hybrid selection scheme, where the best L out of Nr antenna signals are chosen, downconverted, and processed. This reduces the number of required RF chains from Nr to L, and thus leads to significant savings. The savings come at the price of a (usually small) performance loss compared with the full-complexity system. The approach is called *Hybrid Selection/Maximum Ratio Combining* (H-S/MRC), or sometimes also *Generalized Selection Combining* (GSC).

It is well known that the output SNR of MRC is just the sum of the SNRs at the different receive antenna elements. For H-S/MRC, the instantaneous output SNR of H-S/MRC looks deceptively similar to MRC – namely:

$$\gamma_{\mathrm{H-S/MRC}} = \sum_{n=1}^{L} \gamma_{(n)}$$

with Nr for an MRC scheme.

The analysis of H-S/MRC based on a chosen ordering of the branches at first appears to be complicated, since the SNR statistics of the ordered branches are *not* independent. However, we can alleviate this problem by transforming ordered branch variables into a new set of random variables. It is possible to find a transformation that leads to *independently distributed* random variables (termed *virtual branch variables*). The fact that the combiner output SNR can be expressed in terms of independent identically distributed (iid) virtual branch variables enormously simplifies performance analysis of the system. For example, the derivation of Symbol Error Probability (SEP) for uncoded H-S/MRC systems, which normally would require evaluation of nested *N*-fold integrals, essentially reduces to evaluation of a *single* integral with finite limits. The mean and variance of the output SNR for H-S/MRC is thus

$$\overline{\gamma}_{\mathrm{H-S/MRC}} = L\left(1 + \sum_{n=L+1}^{N_r} \frac{1}{n}\right)\overline{\gamma}$$

$$\sigma^2_{\mathrm{H-S/MRC}} = L\left(1 + L\sum_{n=L+1}^{N_r} \frac{1}{n^2}\right)\overline{\gamma}^2$$

Error Probability in Fading Channels with Diversity Reception

In this section we determine the Symbol Error Rate (SER) in fading channels when diversity is used at the RX. We start with the case of flat-fading channels, computing the statistics of the received power and the BER. We then proceed to dispersive channels, where we analyze how diversity can mitigate the detrimental effects of dispersive channels on simple RXs. *Error Probability in Flat-Fading Channels*

Classical Computation Method

we can compute the error probability of diversity systems by averaging the conditional error probability (conditioned on a certain SNR) over the distribution of the SNR: As an example, let us compute the performance of BPSK with *N*r diversity branches with MRC. The SER of BPSK in AWGN . Let us apply this principle to the case of MRC. When inserting Eqs above, we obtain an equation that can be evaluated analytically:

$$\overline{SER} = \int_0^{\infty} pdf_{\gamma}(\gamma) SER(\gamma)\, d\gamma$$

$$SER(\gamma) = Q(\sqrt{2\gamma})$$

$$\overline{SER} = \left(\frac{1-b}{2}\right)^{N_r} \sum_{n=0}^{N_r-1} \binom{N_r - 1 + n}{n} \left(\frac{1+b}{2}\right)^n$$

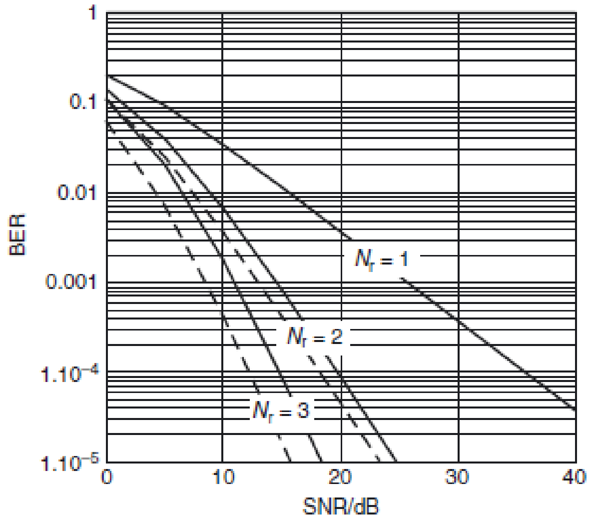


Figure Bit error rate of minimum shift keying (MSK) with received-signal-strength-indication-driven selection diversity (solid) and maximum ratio combining (dashed) as a function of the signal-to-noise ratio with N_r diversity antennas.

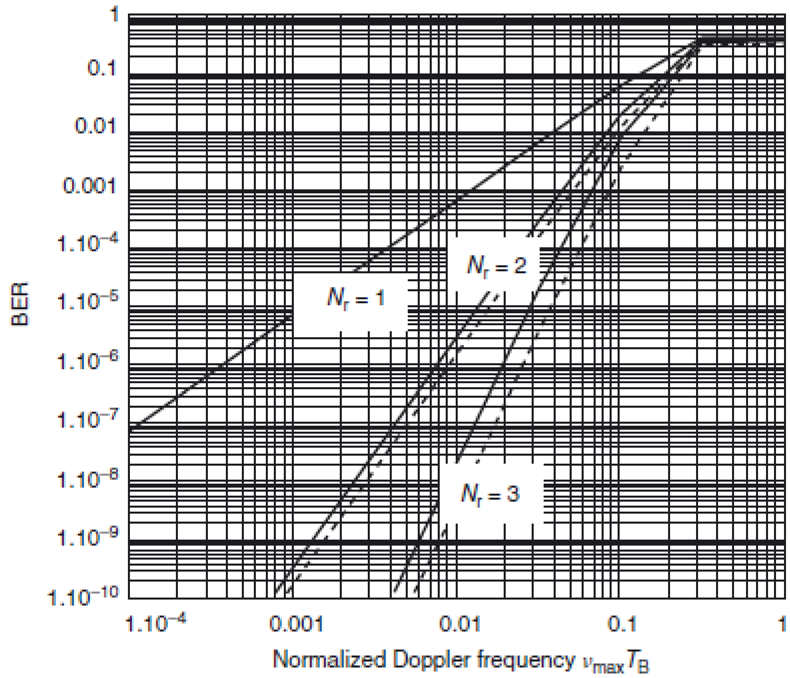


Figure . Bit error rate of MSK with received-signal-strength-indication-driven selection diversity (solid) and maximum ratio combining (dashed) as a function of the normalized Doppler frequency with N_r diversity antennas. Reproduced with permission from Molisch [2000] © Prentice Hall.