CHAPTER

Mode Coupling and its Impact on Spatially Multiplexed Systems

11

Keang-Po Ho[a] and Joseph M. Kahn[b]

[a]*Silicon Image, Sunnyvale, CA 94085, USA*

[b]*E.L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA*

11.1 INTRODUCTION

Recent developments in spatially multiplexed optical communication systems demand a deeper understanding of mode coupling effects in fibers. Spatial multiplexing is being considered for long-haul systems using coherent detection [1–6] or short-range systems using direct detection [7–9]. It increases transmission capacity by multiplexing several data signals in the cores of multicore fibers (MCFs) or in the modes of multimode fibers (MMFs), in which case, it is often called mode-division multiplexing (MDM). Index perturbations in fibers, whether intended or not, can induce coupling between signals in different modes, and can cause propagating fields to evolve randomly. Mode coupling may be classified as weak or strong, depending on whether the total system length is comparable to, or much longer than, a length scale over which propagating fields remain correlated. Mode coupling can affect MDM systems in several important ways.

First, mode coupling, whether occurring in transmission fibers [10,11] or in modal (de)multiplexers [12], leads to crosstalk between spatially multiplexed signals. In direct-detection systems, mode coupling must either be avoided by careful design of all these components, or mitigated by adaptive optical signal processing [13–15]. In systems using coherent detection, any linear crosstalk between modes can be compensated fully by multi-input multi-output (MIMO) digital signal processing (DSP) [1–4], but DSP complexity increases with an increasing number of modes. If mode coupling can be restricted to occur only within mode groups having nearly degenerate propagation constants [7,8], then DSP complexity may be reduced by processing each mode group separately [1].

Second, mode coupling substantially affects the end-to-end group delay (GD) spread of a system [10,16,17], which substantially affects the complexity of MIMO DSP. The GD spread determines the temporal memory required in MIMO time-domain equalization (TDE) of single-carrier modulation [1,18–20], while it determines the fast Fourier transform (FFT) block length in MIMO frequency-domain equalization (FDE) of single-carrier modulation [18] or in orthogonal

Optical Fiber Telecommunications VIB. http://dx.doi.org/10.1016/B978-0-12-396960-6.00011-0

frequency-division multiplexing (OFDM) [2,18]. When using FDE or OFDM with an optimized FFT block length, the DSP complexity per two-dimensional information symbol depends very weakly on the GD spread [18]. Nevertheless, the overall DSP circuit size and convergence time of an adaptive FDE scale roughly linearly with the GD spread or FFT block length [18], and unless strong mode coupling is used to significantly reduce the GD spread, long-haul spatially multiplexed systems may not be technically feasible.

Third, transmission fibers [21,22] and inline optical amplifiers [23–25] can introduce mode-dependent loss or gain, which we collectively refer to as MDL. MDL causes random variations of the powers of signals propagating in different modes. These power variations may affect the various frequency components of each signal differently, and may change over time. Like multipath fading in a MIMO wireless system [26,27], these power variations cause MIMO system capacity to become a random variable. As a result, the mean capacity may be reduced. Moreover, at a given point in time, the instantaneous capacity may drop below the transmission rate, causing system outage [28,29]. Strong mode coupling can reduce the power variations and the associated capacity fluctuations. In conjunction with modal dispersion (MD), mode coupling creates frequency diversity that can further reduce the capacity fluctuations and thus reduce the outage probability [30].

Mode coupling can be described by field coupling models [31, ch. 3], which account for complex-valued modal electric field amplitudes, or by power coupling models [31, ch. 5], a simplified description that accounts only for real-valued modal powers. Early MMF systems used incoherent light emitting diode sources, and power coupling models were used widely to describe important properties, including steady-state modal power distributions and fiber impulse responses [32–34]. Most recent MMF systems use lasers, but power coupling models are still used to describe important effects, including reduced GD spreads in plastic MMFs [16] and a filling-in of the impulse response [35]. By contrast, virtually all practical SMF systems have used laser sources. The study of random birefringence and mode coupling in SMF, which leads to polarization-mode dispersion (PMD), has always used field coupling models, which predict the existence of principal states of polarization (PSPs) [36–38]. PSPs are polarization states that undergo minimal dispersion, and which form the basis of optical compensation of PMD in direct-detection SMF systems [39,40]. In recent years, field coupling models have been applied to MMF, predicting minimally dispersive principal modes (PMs) [17,41], which are the basis for optical compensation of MD in direct-detection MMF systems [13,14]. With heightened recent interest in spatial multiplexing, field coupling models have been applied to study crosstalk in MCFs [42,43], the statistics of coupled GDs in MMF [10,19], and the statistics and system impact of coupled MDL in MMF [28–30].

In this chapter, we review mode coupling effects, how they are modeled, their effect on key fiber properties such as MD and MDL, and their impact on the performance and complexity of MDM systems. The remainder of this chapter is as follows. In Section 11.2, we review modes in optical fibers and the physical sources of coupling between modes. We then compare field- and power-coupling models,

present a matrix propagation model used throughout the chapter, and discuss regimes of weak and strong mode coupling. In Section 11.3, we study MD in the weak- and strong-coupling regimes, especially the statistics of strongly coupled GDs and the GD spread. In Section 11.4, we study MDL, describing its statistics in the strong-coupling regime for both narrow- and wide-band systems. In Section 11.5, we briefly describe MDM in direct-detection systems. In Section 11.6, we discuss long-haul MDM systems using coherent detection, addressing the impact of MDL on average capacity and outage probability, and the effect of MD on DSP complexity.

11.2 MODES AND MODE COUPLING IN OPTICAL FIBERS

In this section, we discuss mode coupling, including its physical origins, models used to describe it, and regimes of weak and strong coupling. We begin the section by a brief review of modes in optical fibers.

11.2.1 Modes in optical fibers

Optical fibers are cylindrical waveguides comprising one (or possibly several) cores surrounded by a cladding having a slightly lower refractive index, as shown in Figure 11.1 [44–47]. A mode is a solution of a wave equation describing a field distribution that propagates in a fiber without changing, except for an overall scaling that describes amplitude and phase changes [44–49]. Throughout this chapter, we assume that a fiber supports D propagating modes, including spatial and polarization degrees of freedom. In fibers of circular cross section, in the limit of weak guidance (small index difference between core and cladding), the total number of propagating modes can

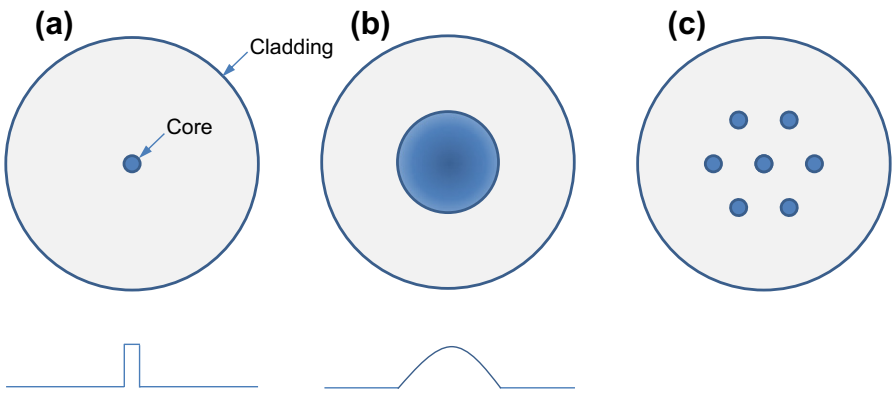

FIGURE 11.1 Different fiber types: cross sections (upper row) and typical refractive index profiles (lower row): (a) step-index single-mode fiber, (b) graded-index multimode fiber (MMF), and (c) multicore fiber.

assume a value $D = 2, 6, 10, 12, 16, 20, 24, 30,\ldots$ [49,50, Table 2.2]. In these weakly guiding fibers, there exist groups of modes having degenerate propagation constants or modal indices (including both phase and group indices) [8,49,50, Table 2.2]. When the refractive index difference between core and cladding is non-zero, the degeneracies within each mode group are generally broken slightly. Nevertheless, in typical glass fibers, the propagation coefficient differences within a mode group are usually small compared to the propagation coefficient differences between mode groups.

When the core radius and core-cladding index difference are sufficiently small, a fiber supports propagation of only one spatial mode in two polarizations ($D = 2$), and is called an SMF. An SMF having a step-index profile is shown in Figure 11.1a; other index profiles are also possible [48,50, Section 3.5]. Signals propagating in SMF are subject to chromatic dispersion (CD) arising both from material dispersion and waveguide dispersion [48,51,52], and are subject to PMD arising from random coupled birefringence [36–40], but are not subject to MD. In SMF, a change of the input launch conditions affects the launched power, but does not affect the spatial distribution at the fiber output, provided the fiber is sufficiently long (typically at least 1 m).

When the core radius and core-cladding index difference becomes sufficiently large, a fiber supports multiple spatial modes, and is called an MMF. Assuming the number of propagating modes is large, it is given approximately by $D \approx (2\pi/\lambda)^2 \int_0^R [n^2(r) - n_c^2]^{1/2} r\, dr$, where $n(r)$ describes the radial variation of the core index, n_c is the cladding index, R is the core radius, and λ is the wavelength [47]. Typically, the number of propagating modes is roughly proportional to the core-cladding index difference times $(R/\lambda)^2$.

MMF is subject to MD, whereby different spatial modes propagate with different modal group indices or GDs. MD causes signal distortion in direct-detection links. MD is severe in fibers with step-index profiles, while it is reduced substantially in fibers using nearly parabolic graded-index profiles, as shown in Figure 11.1b. Conventional glass MMFs for data centers or local-area networks have core diameters of 50 or 62.5 μm and support over one hundred modes at an 850-nm wavelength. Various optical multimode categories OM1 through OM4 [53] support increasing bandwidth × distance products, with bit rates of 10 Gb/s over distances up to 500 m. In an MMF, the output spatial distribution and the impulse response depend on the input launch conditions. In particular, excitation of higher-order modes can be enhanced by launching a beam with a transverse offset from the center axis [7,32,33,54] or with an angular offset [54]. Plastic optical fibers having core diameters of 0.5–1 mm [55,56] are used in short-range links. In these plastic fibers, index inhomogeneity causes strong mode coupling, which significantly reduces the GD spread and increases the supportable bit rate [16,57,58]. However, high losses in plastic optical fibers (33 dB/km is the minimum value given in [56]) limit link distances to under 100 m [56].

The plurality of modes propagating in MMF has long been viewed as a negative effect that limits the bit rate × distance product [33,53]. Attempts have been made to exploit multiple modes to increase capacity in short-range direct-detection links [7–9], but have not yet become practical, in part, because they did not address the

problem of MD. More recently, MDM has yielded capacity increases in long-haul systems using coherent detection [1–4], by exploiting MIMO signal processing to compensate both MD and modal crosstalk. Analogously, multipath fading in wireless communications was long considered a negative effect [59, chs. 13–14,60, Section 5.5.1.2], but multipath is now exploited to increase system throughput in MIMO wireless systems [26,27].

As an alternative to MDM in MMF, spatial multiplexing can be performed in a MCF [5,6], as shown in Figure 11.1c. A MCF can be designed for very small crosstalk between cores [61,62], making it analogous to an array of SMFs embedded in the same cladding. Alternatively, a MCF can be designed with strong coupling between cores [42,43], making it equivalent to a special type of MMF in which super-modes propagate [63]. Our treatment of coupled modes in MMF is applicable to super-modes in strongly coupled MCF, where the total number of modes D is equal to twice the number of cores (assuming single-mode cores).

11.2.2 Mode coupling and its origins

In an ideal fiber, modes propagate without cross-coupling. In a real fiber, perturbations, whether intended or unintended, can induce coupling between spatial and/or polarization modes. Throughout this chapter, we consider only coupling between forward-propagating modes, since it has a dominant effect on the system properties of interest, including MD and MDL.

In multimode transmission fibers, unintended mode coupling can arise from several sources. These include manufacturing variations causing non-circularity of the core, roughness at the core-cladding boundary, variations in the core radius, or variations in the index profile in graded-index fibers. They also include stresses induced by the jacket [64,65], or by thermal mismatches between glasses of different compositions. Finally, mode coupling can arise from micro-bending, macro-bending, or twists [66–68].

Most unintended random perturbations of transmission fibers are having longitudinal power spectra that are lowpass, e.g. scaling as $\langle |F(\Delta\beta)|^2 \rangle \propto \Delta\beta^{-4}$ to $\Delta\beta^{-8}$ [11]. This notation is defined in Sections 11.2.3.1 and 11.2.3.2. As explained there, such lowpass perturbations can strongly couple modes having nearly equal propagation constants (small $\Delta\beta$), while they weakly couple modes having highly unequal propagation constants (large $\Delta\beta$). Therefore, in glass MMFs, nearly degenerate polarization modes or spatial modes within the same group are fully coupled after distances of order 300 m [32,33]. By contrast, even in fibers with low GD spread, spatial modes in different groups are only partially coupled after distances of order 100 km [1,69,70]. Hence, if full, strong mode coupling is desired, it may be necessary to introduce intentional perturbations to the fiber having longitudinal power spectra designed to couple different mode groups. These would be analogous to the spinning used to reduce the differential GD in SMF with PMD [48,71–74]. In Section 11.6, it is shown that strong mode coupling may be required to ensure the practicality of long-haul MDM systems using coherent detection.

In long-haul MDM systems, MDL in transmission fibers is likely to be weak [1], so MDL is expected to arise mainly from inline optical amplifiers [23–25]. In order to minimize the impact of MDL from inline amplifiers, it is sufficient to have a correlation length equal to the amplifier spacing [28]. If mode coupling in transmission fibers proves insufficient, a mode scrambler can be placed in each amplification node. Mode scramblers can be implemented using micro-bending [66–68] or by splicing together unlike fiber types [75,76].

In long-haul MDM systems, a modal multiplexer need not map each input data stream to a distinct uncoupled mode, but need only provide an approximately unitary mapping from the input data streams to those modes, which can imply significant mode coupling [12,77] (analogous considerations apply to a demultiplexer). Because such coupling occurs only at the beginning or end of a link, it cannot by itself reduce the correlation length below the total link length, i.e. it is not sufficient to achieve strong mode coupling.

11.2.3 Mode coupling models

Mode coupling can be described by field coupling models [31, chs. 3–4,78], which account for complex-valued modal electric fields, or by power coupling models [31, ch. 5,11,79,80], a simplified description that accounts only for real-valued modal powers. Both classes of models describe a coupling of energy between modes and how the coupling depends on the mode fields and on a perturbations inducing mode coupling. The transverse dependences of the mode fields and perturbation determine selection rules, while the longitudinal dependence of the perturbation governs phase matching.

Field coupling models, since they describe complex field amplitudes, can describe changes to the eigenmodes and corresponding eigenvalues of certain operators of interest that are caused by mode coupling. In the case of PMD in SMF, field coupling models describe how the PSPs and their GDs are affected by mode coupling [36–40]. PSPs form the basis for optical PMD compensation [39,40]. Likewise, in MMF, field coupling models describe PMs and how their GDs are affected by mode coupling [10,13,17,41], and explain an observed polarization dependence of the impulse response [13,81]. PMs form the basis for optical signal processing to avoid or compensate MD [13,15]. As described below, field coupling models can describe various other important properties of MMF and MDM systems.

Power coupling models also describe an exchange of energy among modes and how it depends on the mode fields and on perturbations inducing mode coupling [11,33,79]. Because they describe coupling as a diffusion process (non-negative real modal powers are coupled by non-negative real coefficients), these models do not describe the changes in eigenmodes and their associated eigenvalues caused by mode coupling. Likewise, power coupling models do not describe changes of eigenmodes or coupling coefficients with optical frequency [30,82].

Both field coupling and power coupling models appeared early in the study of MMF [31]. Early MMF systems used spatially and temporally incoherent light

emitting diodes, and power coupling models were used to describe steady-state power distributions [11] and impulse responses [32,33]. Although more recent MMF systems use lasers, power coupling models have been used to explain reduced GD spreads in plastic fibers [16,57] and a filling-in of impulse responses in glass fibers [35].

Our brief review of field and power coupling models follows that in [31]. For simplicity, we expand propagating fields in a basis of ideal modes of an unperturbed fiber [31, Section 3.2]. This approach is suitable for perturbations of refractive index or geometry that represent small deviations from an ideal fiber. An alternate approach is to expand propagating fields in terms of local normal modes, which are eigenmodes that depend on the local refractive index and geometry [31, Section 3.3]. While more complicated, this approach is suitable for geometrical perturbations, such as bends or tapers, which represent large deviations from an ideal fiber, provided they vary slowly along the fiber's length. Also for simplicity, we assume weak guidance and assume the fields can be described as purely transverse. While including longitudinal components generally improves accuracy in detailed calculations, by ignoring them, we are able to obtain correct first-order insights in many problems. Important exceptions are known, e.g. longitudinal components are required to model coupling between orthogonal polarizations caused by elliptical core deformations [31, Section 4.5].

11.2.3.1 Field coupling models

We assume that light propagates along the z direction and let (x, y) define a transverse plane. The (square of the) unperturbed refractive index profile is $n_0^2(x, y)$, which is assumed to be independent of z. By solving a wave equation that depends on $n_0^2(x, y)$, we obtain a set of D orthonormal ideal propagating modal fields $\mathbf{E}_\mu(x, y), \mu = 1, \ldots, D$, which have propagation constants $\beta_\mu, \mu = 1, \ldots, D$. In this basis, any propagating field can be expanded as $\mathbf{E}(x, y, z) = \sum_{\mu=1}^{D} A_\mu(z)\mathbf{E}_\mu(x, y)$, where $A_\mu(z), \mu = 1, \ldots, D$ are complex-valued coefficients describing the amplitude and phase of each propagating mode. The quantities $\mathbf{E}(x, y, z), \mathbf{E}_\mu(x, y)$ and $A_\mu(z), \mu = 1, \ldots, D$ are all functions of angular frequency ω, but we suppress the frequency dependence for simplicity throughout this section.

If the fiber's refractive index profile (or geometry) is perturbed, since the $\mathbf{E}_\mu(x, y)$ are eigenmodes of the unperturbed fiber, they are coupled by the perturbation. If loss is neglected, propagation and coupling are described by the field coupling equations [31]:

$$\frac{dA_\mu}{dz} = -j\beta_\mu A_\mu + \sum_{\nu \neq \mu} C_{\mu\nu}(z) A_\nu \quad \mu = 1, \ldots, D. \tag{11.1}$$

On the right-hand side, the first term describes uncoupled propagation, and the second term describes coupling. As an alternative, the propagating field may be expanded as $\mathbf{E}(x, y) = \sum_\mu A_\mu(z)\mathbf{E}_\mu(x, y) \exp(-j\beta_\mu z)$, which leads to coupling equations similar to (11.1) without the first term of the right-hand side [78].

In order to illustrate some key principles governing spatial mode coupling, we assume the refractive index is modified by a small perturbation that can be factored to separate the transverse and longitudinal dependences. The (square of the) total index becomes:

$$n^2(x,y,z) = n_0^2(x,y) + \delta n^2(x,y) f(z). \tag{11.2}$$

Assuming a perturbation of the form (11.2), on the right-hand side of (11.1), the complex-valued field coupling coefficient is proportional to:

$$C_{\mu\nu}(z) \propto \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta n^2(x,y)\mathbf{E}_\mu^*(x,y)\cdot\mathbf{E}_\nu(x,y)dxdy\cdot f(z), \tag{11.3}$$

where the proportionality constant is given in [31, Section 3.2]. The first factor in (11.3) is an overlap integral describing how the transverse dependence of the index perturbation couples different modes depending on their field distributions, and it determines selection rules for mode coupling. For example, a perturbation independent of (x, y) cannot couple modes of opposite parities, while a perturbation depending linearly on x or y can couple only modes of opposite parities. It is important to realize that the form of the perturbation (11.2) is hardly general. Being isotropic, it cannot couple spatial modes in orthogonal polarizations. A perturbation of the dielectric tensor is required to describe birefringence caused by stress anisotropy, which is a major cause of polarization coupling in nominally circular fibers.

In order to explain the effect of the second factor $f(z)$ in (11.3), it is instructive to solve the coupled-mode Eq. (11.1). Instead of trying to solve them in general, we examine two special cases [31], which yield considerable insight.

As a first example, we assume that at $z=0$, only one mode ν is excited, and that as the modes propagate, the other modes $\mu \neq \nu$ are only weakly excited:

$$A_\nu(0) = 1, \quad A_\mu(0) = 0, \quad |A_\mu(z)| \ll |A_\nu(z)|, \quad \mu \neq \nu. \tag{11.4}$$

After propagating to $z = L$, the amplitude of mode $\mu \neq \nu$ is:

$$A_\mu(L) \approx e^{-j\beta_\mu L}\int_0^L C_{\mu\nu}(z)e^{-j(\beta_\nu-\beta_\mu)z}dz. \tag{11.5}$$

The right-hand side of (11.5) comprises two factors. The first describes the propagation phase. The second describes how the longitudinal dependence of the index perturbation couples different modes depending on their propagation constants, and defines conditions for phase matching. The integral in (11.5) is the Fourier transform of $C_{\mu\nu}(z)$ given by (11.3), which is proportional to the Fourier transform of $f(z)$, denoted as $F(\Delta\beta)$. The Fourier transform is computed over the interval $(0, L)$ and evaluated at $\Delta\beta = \beta_\nu - \beta_\mu$, the difference between the two propagation constants. As a consequence, nearly degenerate modes (small $\Delta\beta$) may be coupled efficiently by a lowpass $f(z)$, whereas strongly non-degenerate modes (large $\Delta\beta$) are coupled efficiently only by a broadband or resonant $f(z)$ that has significant Fourier components at the large value of $\Delta\beta$. Analogous principles are used in the design of spin profiles

for manufacturing low-PMD SMF [73,74]. This example has assumed that the index perturbation is perfectly separable and that only one mode is excited strongly. Nevertheless, even when these conditions are not satisfied strictly, the dependence of pairwise mode coupling implied by the second factor in (11.5) often remains valid qualitatively. Note that the coupling coefficient in (11.5) also includes the integral over (x, y) defined in (11.3), and hence is subject to the selection rules described above.

As a second example, we consider a system with two modes and a perturbation independent of z which, according to (11.3), induces a constant coupling coefficient. The field coupling Eq. (11.1) becomes:

$$\frac{dA_1}{dz} = -j\beta_1 A_1 + C_{12}A_2, \frac{dA_2}{dz} = -j\beta_2 A_2 - C_{12}^* A_1. \tag{11.6}$$

Assuming initial conditions $A_1(0) = 1$ and $A_2(0) = 0$, the solution becomes

$$A_1(z) = \exp\left(-j\frac{\beta_1+\beta_2}{2}z\right)\left(\cos\gamma z - j\frac{\Delta\beta}{2\gamma}\sin\gamma z\right), \tag{11.7}$$

$$A_2(z) = \frac{C_{12}}{\gamma}\exp\left(-j\frac{\beta_1+\beta_2}{2}z\right)\sin\gamma z, \tag{11.8}$$

where

$$\gamma^2 = \left(\frac{\Delta\beta}{2}\right)^2 + |C_{12}|^2 \quad \text{and} \quad \Delta\beta = \beta_1 - \beta_2.$$

The solutions (11.7) and (11.8) describe fields propagating with an average propagation constant $\frac{1}{2}(\beta_1 + \beta_2)$. A fraction of the total energy is coupled back and forth between the two modes, as described by the factors $\cos\gamma z$ and $\sin\gamma z$. The fraction of coupled energy depends on the ratio $|C_{12}/\Delta\beta|$, becoming small for $|C_{12}| \ll \Delta\beta$, and approaching unity for $|C_{12}| \gg \Delta\beta$. This observation helps explain how even weak perturbations can fully couple nearly degenerate modes, such as the two polarizations in SMF [83].

11.2.3.2 Power coupling models

For many purposes, such as computing the redistribution of average power among modes [11,79,80], it is sufficient to describe evolution of the modal powers $P_\mu(z) = \left\langle \left|A_\mu(z)\right|^2 \right\rangle, \mu = 1, \ldots, D$, which are non-negative and real. The brackets $\langle\,\rangle$ denote an ensemble average. Their evolution is described by the power coupling equations [31, ch. 5]:

$$\frac{dP_\mu}{dz} = -\alpha_\mu P_\mu + \sum_{\nu\neq\mu} h_{\mu\nu}\left(P_\nu - P_\mu\right), \quad \mu = 1, \ldots, D. \tag{11.9}$$

On the right-hand side, the first term describes loss by power attenuation coefficients α_μ, and the second term describes coupling by non-negative real coefficients $h_{\mu\nu}$. Let us assume a perturbed index profile of the form (11.2), except

that the longitudinal dependence $f(z)$ is generally considered a stationary random process. The coupling coefficients are:

$$h_{\mu\nu} = \left\langle \left| \int_0^L C_{\mu\nu}(z)\, e^{-j(\beta_\nu - \beta_\mu)z} dz \right|^2 \right\rangle, \tag{11.10}$$

which is the power spectrum of the coupling coefficient $C_{\mu\nu}(z)$ (11.3), which is proportional to the power spectrum of $f(z)$, denoted by $\langle |F(\Delta\beta)|^2 \rangle$. The power spectrum is evaluated at $\Delta\beta = \beta_\mu - \beta_\nu$, the difference between the two propagation constants. This describes phase matching conditions, and can be interpreted as a stochastic version of (11.5), which involves the Fourier transform of $C_{\mu\nu}(z)$ or $f(z)$. Note that the coupling coefficients (11.10) also include the squared modulus of the integral over (x, y) defined in (11.3), and hence is subject to selection rules, as described above.

In the simple case of two modes, the power coupling equations become

$$\frac{dP_1}{dz} = -\alpha P_1 + h_{12}(P_2 - P_1), \frac{dP_2}{dz} = -\alpha P_2 + h_{12}(P_1 - P_2), \tag{11.11}$$

where we have assumed equal attenuation coefficients for the two modes. If the initial conditions are $P_1(0) = 1$ and $P_2(0) = 0$, the solution is

$$\begin{aligned} P_1(z) &= \tfrac{1}{2}\exp(-\alpha z)\left[1 + \exp(-2h_{12}z)\right], \\ P_2(z) &= \tfrac{1}{2}\exp(-\alpha z)\left[1 - \exp(-2h_{12}z)\right]. \end{aligned} \tag{11.12}$$

Unlike the field coupling result for a fixed, uniform perturbation, for any non-zero coupling coefficient, given sufficient propagation distance, the optical power distribution in the two modes always becomes uniform. This apparent inconsistency can be resolved by observing that, in practice, even a nominally fixed, uniform perturbation is subject to random fluctuations (especially of optical phase) that vary over time and space. Hence, in modeling field coupling, a long fiber should be subdivided into short segments over which the perturbation is very nearly uniform, and modeled for an ensemble of different random phase shifts between the segments. When this procedure is carried out, the field coupling model reproduces the ensemble-average result obtained using the power coupling model.

In the remainder of this chapter, only field coupling models are used to analyze MDM systems, as these models can account for the evolution of complex-valued modal amplitudes, enabling study of how mode coupling affects the eigenmodes and eigenvalues of coupled systems, such as the GDs or gains/losses of coupled modes. Power coupling models can provide partial explanations of some relevant phenomena (e.g. in Section 11.3.1), but it is not clear how power coupling models can be extended to provide the quantitative predictions required for MDM systems.

11.2.3.3 Matrix propagation model

As mentioned above, a field at angular frequency ω propagating in a fiber can be represented as $\mathbf{E}(x, y, z, \omega) = \sum_{\mu=1}^{D} A_\mu(z, \omega)\mathbf{E}_\mu(x, y, \omega)$, where the fields $\mathbf{E}_\mu(x, y, \omega)$, $\mu = 1, \ldots, D$, are the orthonormal eigenmodes of an unperturbed

fiber (this expression remains valid even if the fields are not strictly transverse). The eigenmodes $\mathbf{E}_\mu(x, y, \omega)$ are fixed, so it is convenient to represent the field by a vector $\mathbf{A}(z, \omega) = (A_1(z, \omega), \ldots, A_D(z, \omega))^T$, which we often write simply as $\mathbf{A}(\omega)$. Ignoring noise for now, linear propagation through an optical system can generally be described by a matrix equation:

$$\mathbf{A}^{(\text{out})}(\omega) = \mathbf{M}(\omega)\mathbf{A}^{(\text{in})}(\omega), \tag{11.13}$$

where $\mathbf{A}^{(\text{in})}(\omega)$ and $\mathbf{A}^{(\text{out})}(\omega)$ describe the input and output fields and $\mathbf{M}(\omega)$ is a propagation operator, which is described by a $D \times D$ matrix.

In preparation for the treatment of MD and MDL in Sections 11.3 and 11.4, we discuss how to model a cascade of K passive fiber sections enumerated by $k = 1, \ldots, K$. Assume the kth section has length $L^{(k)}$ which, for now, has no specific relationship to the mode coupling correlation length. Uncoupled propagation of the modes in the kth section is described by the input-output relationship

$$A_\mu^{(\text{out})} = \exp\left[-\frac{\alpha_\mu^{(k)}}{2}L^{(k)} - j\beta_\mu^{(k)}(\omega)L^{(k)}\right]A_\mu^{(\text{in})}, \quad \mu = 1, \ldots, D, \tag{11.14}$$

where $\alpha_\mu^{(k)}$ are the power attenuation coefficients, which may be mode-dependent [21,22] but are assumed independent of frequency over the bandwidth of interest, and $\beta_\mu^{(k)}(\omega)$ are the propagation constants, which are generally mode-dependent [31, chs. 1,44,46,47,84].

To describe uncoupled power gain in the kth section, we define the mode-averaged attenuation constant $\bar{\alpha}^{(k)} = \frac{1}{D}\sum_\mu \alpha_\mu^{(k)}$. We quantify MDL by the uncoupled gains $g_\mu^{(k)} = -\left(\alpha_\mu^{(k)} - \bar{\alpha}^{(k)}\right)L^{(k)}, \mu = 1, \ldots, D$, where $\sum_\mu g_\mu^{(k)} = 0$. The vector

$$\mathbf{g}^{(k)} = \left(g_1^{(k)}, g_2^{(k)}, \ldots, g_D^{(k)}\right) \tag{11.15}$$

describes the uncoupled MDL in the kth section.

To describe uncoupled MD and CD in the kth section, we perform a Taylor series expansion of the propagation constant $\beta_\mu^{(k)}(\omega)$ and keep the terms linear and quadratic in ω. MD is described by the uncoupled GDs $\beta_{1,\mu}^{(k)}L^{(k)}, \mu = 1, \ldots, D$, where $\beta_{1,\mu}^{(k)} = d\beta_\mu^{(k)}(\omega)/d\omega$. The mode-averaged uncoupled GD is $\bar{\beta}_1^{(k)}L^{(k)} = \frac{L^{(k)}}{D}\sum_\mu \beta_{1,\mu}^{(k)}$. We quantify MD by the uncoupled GDs $\tau_\mu^{(k)} = \left(\beta_{1,\mu}^{(k)} - \bar{\beta}_1^{(k)}\right)L^{(k)}, \mu = 1, \ldots, D$, where $\sum_\mu \tau_\mu^{(k)} = 0$. The vector

$$\boldsymbol{\tau}^{(k)} = \left(\tau_1^{(k)}, \tau_2^{(k)}, \ldots, \tau_D^{(k)}\right) \tag{11.16}$$

describes the uncoupled modal GDs in the kth section.

CD includes contributions from both material dispersion and waveguide dispersion. As wavelength increases, any mode has more power in the cladding, where the refractive index is lower, so waveguide dispersion is always negative. Waveguide dispersion is enhanced for modes near cutoff [85–87]. When all modes

are well above cutoff, which is desirable for low loss and good bend tolerance, mode-dependent CD is reduced [88].

In the kth section, CD is described by the parameters

$$\beta_{2,\mu}^{(k)} = d^2\beta_\mu^{(k)}(\omega)/d\omega^2 = \left(-\lambda^2/2\pi c\right) D_\mu^{(k)}, \quad \mu = 1, \ldots, D,$$

where $D_\mu^{(k)} = d\left(1/v_{g,\mu}^{(k)}\right)/d\lambda$ is the dispersion coefficient of mode μ. The mode-average CD parameter is $\bar{\beta}_2^{(k)} = \frac{1}{D}\sum_\mu \beta_{2,\mu}^{(k)}$, and mode-dependent CD is described by $\Delta\beta_{2,\mu}^{(k)} = \beta_{2,\mu}^{(k)} - \bar{\beta}_2^{(k)}$, where $\sum_\mu \Delta\beta_{2,\mu}^{(k)} = 0$.

Including MDL, MD, and mode-dependent CD, uncoupled propagation in the kth section is described by the diagonal matrix:

$$\Lambda^{(k)}(\omega) = \begin{pmatrix} e^{\frac{1}{2}g_1^{(k)} - j\omega\tau_1^{(k)} - \frac{j}{2}\omega^2\Delta\beta_{2,1}L^{(k)}} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}g_D^{(k)} - j\omega\tau_D^{(k)} - \frac{j}{2}\omega^2\Delta\beta_{2,D}L^{(k)}} \end{pmatrix}. \tag{11.17}$$

If it is necessary to include mode-averaged gain and dispersion, the diagonal matrix (11.17) can be multiplied by a constant factor:

$$\exp\left(-\frac{1}{2}\bar{\alpha}^{(k)}L^{(k)} - j\omega\bar{\beta}_1^{(k)}L^{(k)} - \frac{j}{2}\omega^2\bar{\beta}_2 L^{(k)}\right). \tag{11.18}$$

Including mode coupling, propagation in the kth section is modeled by a product of three $D \times D$ matrices:

$$\mathbf{M}^{(k)}(\omega) = \mathbf{V}^{(k)}\mathbf{\Lambda}^{(k)}(\omega)\mathbf{U}^{(k)*}, \quad k = 1, \ldots, K, \tag{11.19}$$

where * denotes Hermitian transpose. The $D \times D$ matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ are frequency-independent unitary matrices representing mode coupling at the input and output of the kth section, respectively.

Finally, propagation through a cascade of K fiber sections is represented by a total propagation matrix $\mathbf{M}^{(t)}(\omega)$, which is the product of K matrices [10,17,19,28–30]:

$$\mathbf{M}^{(t)}(\omega) = \mathbf{M}^{(K)}(\omega)\mathbf{M}^{(K-1)}(\omega)\cdots\mathbf{M}^{(2)}(\omega)\mathbf{M}^{(1)}(\omega). \tag{11.20}$$

The matrix model (11.19) and (11.20) can describe signal propagation through any cascade of linear elements, as illustrated in Figure 11.2. For example, an optical amplifier can be described by a section in which $\bar{\alpha}^{(k)} < 0$, such that the factor $\exp\left(-\bar{\alpha}^{(k)}L^{(k)}\right) > 1$ describes the mode-averaged power gain. (De)multiplexers with or without crosstalk can be described easily.

Addition of noise by optical amplifiers cannot be described by the matrix model (11.17) and (11.19), and is discussed in Section 11.4.5.

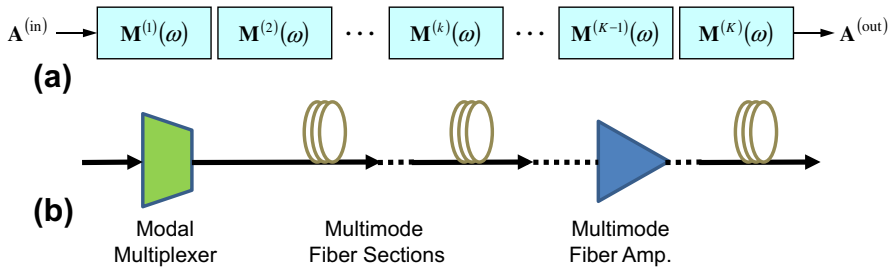

FIGURE 11.2 (a) Matrix model for an MMF system described by the product of a cascade of matrices. (b) Each matrix in the cascade may represent a section of MMF, a modal multiplexer or demultiplexer, a multimode optical amplifier or other components.

11.2.3.4 Regimes of mode coupling

Different regimes of mode coupling are familiar from the study of randomly coupled birefringence and PMD [37,89], and apply as well to mode coupling in MMF [16,17,82]. These regimes are especially relevant in studying key effects such as MD (Section 11.3) or MDL (Section 11.4), where operators describing MD or MDL are defined, whose eigenvalues describe physical properties of interest, i.e. coupled modal GDs or coupled modal gains.

As stated above, modal fields are assumed to be strongly correlated over distances less than or equal to a correlation length, and weakly correlated over distances far larger than the correlation length. This correlation length is a generalization of the polarization correlation length defined in the study of PMD [83,90–92].

In the weak-coupling regime, the correlation length is comparable to, or slightly shorter than, the total system length. In this regime, signal propagation can be modeled using a small number of sections K, where each section should be slightly longer than the correlation length. If a larger number of sections K is employed, the unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ defined in (11.19) should be correlated to reflect correlation between sections. In the weak-coupling regime, the spread of eigenvalues describing quantities of interest, such as modal GDs or modal gains, scales linearly with the number of sections K or the total system length $L_t = \sum_k L^{(k)}$, and each coupled eigenmode is a linear combination of a small number of uncoupled modes.

In the strong-coupling regime, the correlation length is far shorter than the total system length. In this regime, signal propagation must be modeled using a large number of sections K, where each section should be slightly longer than the correlation length. The unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k-1)}$ should be statistically independent to ensure independence between sections. In the strong-coupling regime, the spread of eigenvalues describing quantities of interest, such as modal GDs or modal gains, scales with the square-root of the number of sections K or the square-root of the total system length L_t, and each coupled eigenmode is a linear combination of many uncoupled modes.

These regimes of mode coupling are further illustrated and compared in Section 11.3.1.

11.3 MODAL DISPERSION

In coherent MDM systems, MD poses no fundamental performance limitations. MIMO equalizer complexity increases with the GD spread [18–20], but if laser phase noise is sufficiently small and the channel does not change too rapidly, MD can be compensated with no penalty. To date, most long-haul coherent MDM systems [1–4,70] have operated in the weak-coupling regime. But as transmission distances and the number of multiplexed modes are increased, equalizer complexity is likely to become prohibitive unless strong mode coupling is exploited to reduce the GD spread.

In this section, field coupling models are used to study the effect of mode coupling on MD. PMs, which are coupled modes having well-defined GDs, are described. Of greatest relevance to long-haul MDM systems, the statistics of the GDs of the PMs in the strong-coupling regime are analyzed.

11.3.1 Coupled modal dispersion

The effect of mode coupling on MD was studied for the two-mode case in the context of PMD in SMF [36–40,89–91], where coupled polarization states having well-defined GDs were called PSPs. The theory was extended to the multimode case in [41], where coupled modes having well-defined GDs were called PMs. Following [41], we assume MDL is negligible (the combined impact of MD and MDL is studied in Sections 11.4.6 and 11.6.2). We further neglect mode-averaged loss or gain to simplify the notation here, although that assumption was not made in [38,41]. In this case, the total propagation operator defined in (11.20) becomes unitary: $\mathbf{M}^{(t)*}(\omega)\mathbf{M}^{(t)}(\omega) = \mathbf{I}$, where $\mathbf{I}$ is a $D \times D$ identity matrix. We restrict attention to optical signals occupying a narrow bandwidth near angular frequency ω. We define a set of D input PMs as a set of input field patterns defined by vectors $\mathbf{A}_\mu^{\mathrm{PM,(in)}}, \mu = 1, \ldots, D$. These input PMs are linear combinations of the uncoupled modes that have special properties. If an input PM is launched, the field at the fiber output is described by the corresponding output PM $\mathbf{A}_\mu^{\mathrm{PM,(out)}} = \mathbf{M}^{(t)}(\omega)\mathbf{A}_\mu^{\mathrm{PM,(in)}}$. Each of the input PMs is defined such that if we fix the input field pattern to be $\mathbf{A}_\mu^{\mathrm{PM,(in)}}$ and vary ω slightly, the output field pattern $\mathbf{A}_\mu^{\mathrm{PM,(out)}}$ remains unchanged to first order in ω. One can define a Hermitian GD operator:

$$\mathbf{G} = j\mathbf{M}_\omega^{(t)}\mathbf{M}^{(t)*}(\omega), \tag{11.21}$$

where $\mathbf{M}_\omega^{(t)} = d\mathbf{M}^{(t)}(\omega)/d\omega$ denotes differentiation with respect to ω. As shown in [41], the D input PMs are mutually orthogonal eigenmodes of the GD operator whose eigenvalues are the coupled GDs:

$$\mathbf{G}\mathbf{A}_\mu^{\mathrm{PM,(in)}} = \tau_\mu \mathbf{A}_\mu^{\mathrm{PM,(in)}}, \quad \mu = 1, \ldots, D. \tag{11.22}$$

In general, each input PM differs from the corresponding output PM, since the operators $\mathbf{M}^{(t)}$ and $\mathbf{G}$ do not commute.

For signals occupying a sufficiently narrow bandwidth near angular frequency ω, the overall input-output relationship of a fiber can be expressed as:

$$\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}\boldsymbol{\Lambda}^{(t)}(\omega)\mathbf{U}^{(t)*}, \tag{11.23}$$

where $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ are unitary matrices that are independent of frequency (to first order), whose columns represent the input PMs and the output PMs, respectively, in the basis of ideal modes. The matrix

$$\boldsymbol{\Lambda}^{(t)}(\omega) = \begin{pmatrix} e^{-j\omega\tau_1} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega\tau_D} \end{pmatrix} \tag{11.24}$$

describes propagation of the PMs. The matrix (11.24) describes no crosstalk, since it is diagonal, and describes differential delay but no distortion, since its phase depends linearly on ω. Many factorizations of the form (11.23) are possible mathematically, but this choice is unique in yielding frequency-independent $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ and a diagonal $\boldsymbol{\Lambda}^{(t)}(\omega)$ with well-defined GDs.

In [17], the effect of spatial- and polarization-mode coupling on the modal GDs in graded-index MMF was studied. A fiber of fixed total length L_t was divided into a fixed number of sections K, and each section was given a random curvature to induce spatial-mode coupling and birefringence. Successive sections were rotated by a random angle to induce polarization-mode coupling. By changing the standard deviation (STD) of the curvature σ_κ, different regimes of mode coupling can be obtained. Figure 11.3 shows the coupled GDs of the PMs in a 1-km-long, graded-index MMF supporting $D = 110$ modes in two polarizations. In the weak-coupling regime ($\sigma_\kappa < 0.3$), modes form groups having GD degeneracies of 2, 4, 6, …, and the GD spread scales linearly with K or L. In the medium-coupling regime ($0.3 < \sigma_\kappa < 2$), these GD degeneracies are broken, and each coupled PM is a linear combination of a small number of uncoupled modes. In the strong-coupling regime ($2 < \sigma_\kappa < 20$), the GD spread is reduced substantially, scaling in proportion to $\sqrt{K}$ or $\sqrt{L_t}$, and each coupled PM is a linear combination of many uncoupled modes. The reduction of the GD spread, while not commonly observed in glass MMFs of 1-km length, is analogous to that observed in plastic MMF. The regimes of mode coupling were explained previously using a power coupling model in [16]. The power coupling model, while intuitively appealing, may not be capable of providing the quantitative predictions required for MDM systems.

The field pattern of each PM varies with optical frequency ω, and is correlated over a coherence bandwidth that depends, in principle, on the uncoupled modes and GDs and the mode coupling [30,82]. In the absence of coupling and MDL, the PMs are essentially equivalent to the ideal modes [41], so the coherence bandwidth is extremely large, of the order of THz. In [17], the coherence bandwidths of the PMs

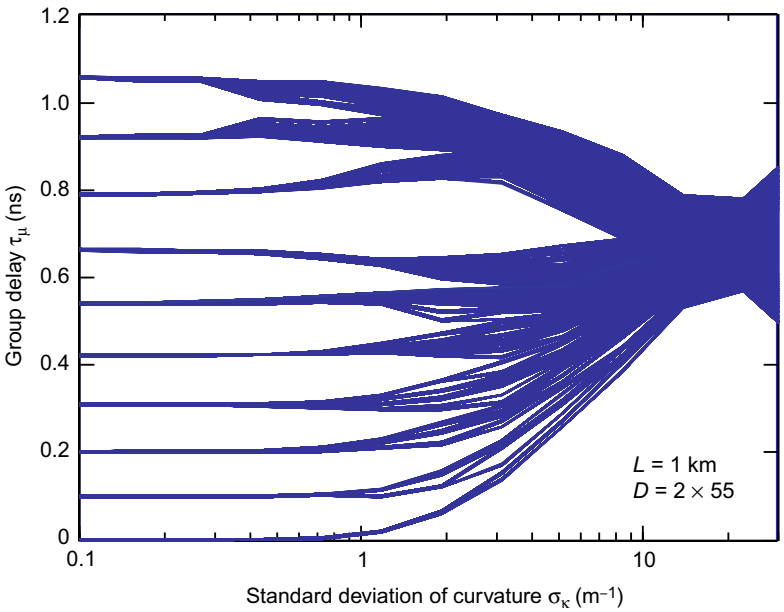

FIGURE 11.3 PM group delays (GDs) versus the standard deviation (STD) of curvature determining the mode coupling strength for a MMF of 1-km length, supporting $D = 110$ propagating modes. The GDs exhibit regimes of weak, moderate, and strong mode coupling. Adapted from [17].

in graded-index MMF were studied. For a 1-km fiber with weak coupling, coherence bandwidths were found to be about 300 GHz for a parabolic index profile, and smaller than 10 GHz for a non-parabolic profile. In the strong-coupling regime, following arguments similar to those given in Section 11.4.6 for coupled MDL, the coherence bandwidths of PMs should be of the order of $1/\sigma_{\text{gd}}$, where σ_{gd} is the overall STD of coupled GD, given by (11.45).

When a modulated signal occupies a bandwidth that is not small compared to the coherence bandwidth, higher-order MD effects occur [30,82]. The overall input-output relationship becomes more complicated than (11.23); in particular, the phase of the propagation operator is no longer a linear function of ω, as in (11.24). Higher-order MD effects include a filling-in between peaks in the pulse response of a fiber (observed in [35]), a nonlinear relationship between input and output intensity waveforms, polarization- and spatial mode-dependent CD, and enhanced depolarization of modulated signals, all of which have analogs in higher-order PMD [93,94]. Higher-order MD limits the signal bandwidth over which frequency-independent optical signal processing can avoid distortion from MD in direct-detection links [95] or avoid crosstalk in direct-detection MDM systems (see Section 11.5).

In long-haul coherent MDM systems, MD poses no fundamental performance limitations. The MIMO equalizer complexity increases with the GD spread due to MD [18–20], but in systems with sufficiently small laser phase noise and sufficiently slow channel variation over time, an arbitrarily large GD spread and arbitrary-order MD can be compensated with no penalty. Most long-haul coherent MDM systems demonstrated to date [1–4] have operated in the weak-coupling regime. Nevertheless, as shown below, strong coupling reduces GD spread and thus minimizes MIMO signal processing complexity, and it mitigates MDL. Since strong coupling is beneficial for long-haul coherent MDM systems, we analyze the statistics of strongly coupled GDs in the remainder of this section, and we analyze the statistics of strongly coupled MDL in Section 11.4.

11.3.2 Group delay statistics in strong-coupling regime

In the strong-coupling regime, MD in an MMF can be modeled by random matrices, and the distribution of the GDs can be characterized by studying the statistics of the eigenvalues of those random matrices [10,96].

11.3.2.1 Zero-trace Gaussian unitary ensemble

In the strong coupling regime, in the matrix model of Section 11.2.3.3, a MMF is modeled using many independent sections. Using the diagonal matrix (11.17) and neglecting MDL and mode-dependent CD, uncoupled propagation in the kth section can be modeled by the diagonal matrix:

$$\mathbf{\Lambda}^{(k)}(\omega) = \begin{pmatrix} e^{-j\omega\tau_1^{(k)}} & & 0 \\ & \ddots & \\ 0 & & e^{-j\omega\tau_D^{(k)}} \end{pmatrix}, \tag{11.25}$$

where the vector $\boldsymbol{\tau}^{(k)}$ (11.16) describes the uncoupled modal GDs.

In the absence of MDL, $\mathbf{\Lambda}^{(k)}(\omega)$, $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ are all unitary matrices, such that $\mathbf{M}^{(k)}(\omega)\mathbf{M}^{(k)*}(\omega) = \mathbf{I}$. With strong random mode coupling, following the discussion of Section 11.2.3.4, $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ can be assumed to be independent random unitary matrices, such that both input and output eigenvectors are independently oriented from one section to the next.

The model here is valid regardless of whether or not the modal delay vectors in each section $\boldsymbol{\tau}^{(k)} = \left(\tau_1^{(k)}, \tau_2^{(k)}, \ldots, \tau_D^{(k)}\right), k = 1, \ldots, K$, (11.16) have the same statistical properties. The vector $\boldsymbol{\tau}^{(k)}$ may even be a deterministic vector, identical for each section. As in Section 11.2.3.3 and without loss of generality, we assume that $\sum_\mu \tau_\mu^{(k)} = 0$, ignoring the mode-averaged delay of each section that does not lead to MD.

In the kth section, the input-output relationship is

$$\mathbf{A}_k^{(\text{out})} = \mathbf{M}^{(k)}(\omega)\mathbf{A}_k^{(\text{in})}$$

and $\mathbf{A}_k^{(\text{in})} = \mathbf{M}^{(k)*}(\omega)\mathbf{A}_k^{(\text{out})}$. Using (11.22), the coupled GDs correspond to the eigenvalues of the GD operator $\mathbf{G}$ given by (11.21). With only a single section, we may verify that

$$\mathbf{G}^{(k)} = j\mathbf{M}_\omega^{(k)}\mathbf{M}^{(k)*} = \mathbf{V}^{(k)}\mathbf{T}^{(k)}\mathbf{V}^{(k)*}, \tag{11.26}$$

where $\mathbf{T}^{(k)} = \text{diag}\left[\tau_1^{(k)}, \tau_2^{(k)}, \ldots, \tau_D^{(k)}\right]$ is a diagonal matrix of the uncoupled GDs in the kth section. With the constraint $\sum_\mu \tau_\mu^{(k)} = 0$, we have $\text{tr}\left(T^{(k)}\right) = \text{tr}\left(j\mathbf{M}_\omega^{(k)}\mathbf{M}^{(k)*}\right) = 0$. The GD operator matrices $\mathbf{G}^{(k)} = j\mathbf{M}_\omega^{(k)}\mathbf{M}^{(k)*}$, $k = 1, \ldots, K$, are all Hermitian matrices with real eigenvalues given by $\mathbf{T}^{(k)}$.

Considering the overall propagation matrix $\mathbf{M}^{(t)}$, the overall PMs and their GDs correspond to the eigenvectors and eigenvalues of the overall GD operator $\mathbf{G} = j\mathbf{M}_\omega^{(t)}\mathbf{M}^{(t)*}(\omega)$. Because of the chain rule

$$\mathbf{M}_\omega^{(t)} = \mathbf{M}^{(K)}\cdots\mathbf{M}^{(2)}\mathbf{M}_\omega^{(1)} + \mathbf{M}^{(K)}\cdots\mathbf{M}_\omega^{(2)}\mathbf{M}^{(1)} + \cdots + \mathbf{M}_\omega^{(K)}\cdots\mathbf{M}^{(2)}\mathbf{M}^{(1)}, \tag{11.27}$$

we obtain

$$\begin{aligned}\mathbf{G} = {} & j\mathbf{M}_\omega^{(K)}\mathbf{M}^{(K)*} + j\mathbf{M}^{(K)}\mathbf{M}^{(K-1)}\mathbf{M}_\omega^{(K-1)*}\mathbf{M}^{(K)*} + \\ & \cdots + \mathbf{M}^{(K)}\cdots\mathbf{M}^{(2)}\mathbf{M}_\omega^{(1)}\mathbf{M}^{(1)*}\mathbf{M}^{(2)*}\cdots\mathbf{M}^{(K)*}.\end{aligned} \tag{11.28}$$

The overall matrix $\mathbf{G}$ given by (11.28) is the summation of K random matrices. All those K random matrices are statistically identical to those of (11.26). Their eigenvectors are independent of each other. The first matrix $\mathbf{M}_\omega^{(K)}\mathbf{M}^{(K)*}$ has the same eigenvectors as $\mathbf{V}^{(K)}$. The second matrix $j\mathbf{M}^{(K)}\mathbf{M}^{(K-1)}\mathbf{M}_\omega^{(K-1)*}\mathbf{M}^{(K)*}$ has eigenvectors from $\mathbf{M}^{(K)}\mathbf{V}^{(K-1)}$. Both matrices $\mathbf{V}^{(K)}$ and $\mathbf{M}^{(K)}\mathbf{V}^{(K-1)}$ are unitary matrices that are obviously independent of each other. The matrix $\mathbf{M}^{(K)}\mathbf{V}^{(K-1)}$ is frequency-dependent. All K random matrices in (11.28) are independent, owing to the different directions of their independent eigenvectors. Excepting the first matrix $j\mathbf{M}_\omega^{(K)}\mathbf{M}^{(K)*}$, all other matrices are frequency-dependent, due to the frequency-dependent factor in $\mathbf{\Lambda}^{(k)}(\omega)$ (11.25).

The matrix elements of $\mathbf{G}$, $g_{\mu,\nu}$, $\mu, \nu = 1, \ldots, D$, are the summation of K independent random variables. When the number of independent sections K is large, the matrix $\mathbf{G}$ is a random matrix whose entries are Gaussian random variables from the Central Limit Theorem. Because all K component matrices in (11.28) are Hermitian, $\mathbf{G}$ is also a Hermitian matrix. The diagonal elements $g_{\mu,\mu}$, $\mu = 1, \ldots, D$, are all real Gaussian random variables. All off-diagonal elements $g_{\mu,\nu}, \mu \neq \nu$, $\mu, \nu = 1, \ldots, D$, are complex Gaussian random variables with independent real and imaginary parts.

If the D vectors in $\mathbf{V}^{(k)}$ are assumed to be independent of each other, it can be shown that the variance of the matrix elements is

$$\sigma_g^2 = \frac{1}{D^2}\sum_{k=1}^{K}\left\|\boldsymbol{\tau}^{(k)}\right\|^2 = \frac{1}{D}\sum_{k=1}^{K}\sigma_{\tau^{(k)}}^2, \tag{11.29}$$

where $\sigma^2_{\tau_{(k)}}$, $k = 1\ldots, K$, are variances of the GDs in each section. If all K sections have statistically identical GD profiles, we have

$$\sigma_g^2 = \frac{K}{D}\sigma_\tau^2, \tag{11.30}$$

where σ_τ^2 are the GD variances in all sections.

The D vectors in $\mathbf{V}^{(k)}$ are not independent of each other, however, as the Dth vector is determined by other $D-1$ vectors due to the unitarity condition. The elements of the vector $\boldsymbol{\tau}^{(k)}$ also sum to zero due to the zero-trace condition. The variance of matrix elements σ_g^2 is an averaged variance in which the diagonal elements have smaller variance than the non-diagonal elements.

In random matrix theory, the matrix $\mathbf{G}$ is described as zero-trace (or traceless) Gaussian unitary ensemble. Typically, a Gaussian unitary ensemble does not have any constraint aside from the variance of its matrix elements. However, in (11.28), all matrix components have zero-trace so that

$$\mathrm{tr}(\mathbf{G}) = 0. \tag{11.31}$$

In the strong mode coupling regime, the GDs in a MMF are statistically described by the eigenvalues of the zero-trace Gaussian unitary ensemble.

Assuming a Gaussian unitary ensemble of $\mathbf{A}$ without trace constraint, the diagonal elements are real Gaussian random variables with variance σ_a^2. The off-diagonal elements of $\mathbf{A}$ are complex Gaussian random variables with variance σ_a^2. Equivalently, the real and imaginary parts of the off-diagonal elements are independent of each other with variance $\sigma_a^2/2$. The corresponding zero-trace Gaussian unitary ensemble is equivalently the Gaussian random matrix $\mathbf{A} - \mathrm{tr}(\mathbf{A})\mathbf{I}/D$. The off-diagonal elements have variance σ_a^2 but the diagonal elements have variance $\sigma_a^2(1 - 1/D)$. The variance of all matrix elements has an arithmetic average $\sigma_a^2(1 - 1/D^2)$.

For the zero-trace Gaussian unitary ensemble $\mathbf{G}$, numerical simulation shows that the diagonal elements have equal variance $\sigma_g^2(1 - 1/D)$ and the off-diagonal elements have equal variance $\sigma_g^2\,[1 + 1/D/(D - 1)]$, and the average variance is σ_g^2. If the zero-trace Gaussian unitary ensemble $\mathbf{G}$ is related to a Gaussian unitary ensemble $\mathbf{A}$ via $\mathbf{G} = \mathbf{A} - \mathrm{tr}(\mathbf{A})\mathbf{I}/D$, the Gaussian unitary ensemble $\mathbf{A}$ without trace constraint should have a variance $\sigma_g^2 D^2/(D^2 - 1)$. This variance correction is required because the statistics of zero-trace Gaussian unitary ensemble are always derived based on the corresponding Gaussian unitary ensemble without trace constraint.

11.3.2.2 Group delay distribution

In the regime of strong mode coupling, at each single frequency, the PMs and their GDs are given by the eigenvectors and eigenvalues of the zero-trace Gaussian unitary ensemble (11.28). As above, the zero-trace Gaussian unitary ensemble $\mathbf{G}$ may be related to a Gaussian unitary ensemble without trace constraint $\mathbf{A}$ as $\mathbf{G} = \mathbf{A} - \mathrm{tr}(\mathbf{A})\mathbf{I}/D$. Without loss of generality, after normalization, the elements of $\mathbf{A}$ may be assumed to be zero-mean independent identically distributed Gaussian

random variables with variance $\sigma_a^2 = 1/2$, similar to the classic normalization of [96, Section 3.3]. The diagonal elements of $\mathbf{A}$ are real with a variance of $\sigma_a^2 = 1/2$. The off-diagonal elements of $\mathbf{A}$ are complex Gaussian distributed with independent real and imaginary parts, each having a variance of 1/4.

The joint distribution for a Gaussian unitary ensemble without the zero-trace constraint is well known. The unordered joint probability density function of the eigenvalues of a $D \times D$ Gaussian unitary ensemble is [96, Section 3.3,97]:

$$p_{\mathrm{nc}}(x) = C_{D2}^{-1} \prod_{D \geqslant \mu \geqslant \nu \geqslant 1} (x_\mu - x_\nu)^2 \exp\left(-\sum_{\mu=1}^{D} x_\mu^2\right), \tag{11.32}$$

where

$$C_{D2} = \frac{\pi^{D/2}}{2^{D(D-1)/2}} \prod_{n=1}^{D} n!.$$

With suitable normalization, the random variables $x_\mu, \mu = 1, \ldots, D$, yield the GDs of the MMF with strong mode coupling. With the zero-trace constraint (11.31), the joint probability density function becomes

$$p_{\mathrm{zt}}(x) = \tilde{C}_{D2}^{-1} \delta\left(\sum_{\mu=1}^{D} x_\mu\right) \prod_{D \geqslant \mu \geqslant \nu \geqslant 1} (x_\mu - x_\nu)^2 \exp\left(-\sum_{\mu=1}^{D} x_\mu^2\right). \tag{11.33}$$

The constant $\tilde{C}_{D2}$ is not the same as C_{D2} and is found equal to $\tilde{C}_{D2} = C_{D2}/\sqrt{\pi D}$ later in this section.

The eigenvalue distribution for zero-trace Gaussian unitary ensemble $\mathbf{G}$ is given by

$$p_D(x_1) = \tilde{C}_{D2}^{-1} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \delta\left(\sum_{\mu=1}^{D} x_\mu\right) \prod_{D \geqslant \mu \geqslant \nu \geqslant 1} (x_\mu - x_\nu)^2 \exp\left(-\sum_{\mu=1}^{D} x_\mu^2\right) dx_2 \cdots dx_D. \tag{11.34}$$

The $D \times D$ Vandermonde determinant gives [98, Section 4.6]

$$\det\left[x_\mu^{\nu-1}\right]_{\mu,\nu=1,2,\ldots,D} = \prod_{D \geqslant \mu \geqslant \nu \geqslant 1} (x_\mu - x_\nu),$$

where $\det[\cdot]$ denotes a determinant. Following the method of [96, Section 6.2], and directly from the properties of determinants, the Vandermonde determinant can be expressed in terms of Hermite polynomials as

$$\begin{aligned} \det\left[x_\mu^{\nu-1}\right]_{\mu,\nu=1,2,\ldots,D} &= \det\left[\frac{1}{2^{\nu-1}} H_{\nu-1}(x_\mu)\right]_{\mu,\nu=1,2,\ldots,D} \\ &= \det\left[\frac{1}{2^{\nu-1}} H_{\nu-1}(x_\mu + c_\nu)\right]_{\mu,\nu=1,2,\ldots,D}, \end{aligned} \tag{11.35}$$

where $H_n(x)$ are Hermite polynomials and c_v are constants. The leading terms in the entries of the determinants (11.35) are all x_μ^{v-1}. Using the Hermite polynomials, we obtain

$$p_{\text{nc}}(x) = \frac{1}{D!} \det[K_D(x_\mu, x_v)]_{\mu,v=1,2,\ldots,D}, \tag{11.36}$$

where

$$K_D(x,y) = \sum_{n=0}^{D-1} \frac{1}{2^n n! \sqrt{\pi}} H_n(x) H_n(y) \exp\left(-\frac{x^2+y^2}{2}\right). \tag{11.37}$$

Without trace constraint, the eigenvalue distribution is given by $K_D(x,x)/D$ and $K_D(x,x)$ is called the correlation function [96, Section 6.2].

With zero-trace constraint, the joint density becomes

$$p_{\text{zt}}(x) = \frac{C_{D2}}{\tilde{C}_{D2}} \delta\left(\sum_{\mu=1}^{D} x_\mu\right) \det[k_D(x_\mu, x_v)]_{\mu,v=1,2,\ldots,D}. \tag{11.38}$$

Similar to [99,100] but using Fourier instead of Laplace transform,

$$p_{\text{zt}}(x) = \frac{C_{D2}}{2\pi \tilde{C}_{D2} D!} \int_{-\infty}^{+\infty} \exp\left(i\omega \sum_{\mu=1}^{D} x_\mu\right) \det[K_D(x_\mu, x_v)]_{\mu,v=1,2,\ldots,D}\, d\omega. \tag{11.39}$$

With some algebra, we obtain

$$p_{\text{zt}}(x) = \frac{C_{D2}}{2\pi \tilde{C}_{D2} D} \int_{-\infty}^{+\infty} \exp\left(-\frac{D}{4}\omega^2\right) \det\left[K_D\left(x_\mu + \frac{i\omega}{2}, x_v + \frac{i\omega}{2}\right)\right]_{\mu,v=1,2,\ldots,D} d\omega. \tag{11.40}$$

In the above expression, the argument inside the Hermite polynomial changes from x_μ to $x_\mu + i\omega/2$ using the relation (11.35).

Similar to the method used to find the eigenvalue distribution, the probability density of the eigenvalues is given by

$$p_D(x) = \frac{C_{D2}}{2\pi \tilde{C}_{D2} D} \int_{-\infty}^{+\infty} K_D\left(x + \frac{i\omega}{2}, x + \frac{i\omega}{2}\right) \exp\left(-\frac{D}{4}\omega^2\right) d\omega \tag{11.41}$$

or

$$p_D(x) = \frac{C_{D2}}{2\pi \tilde{C}_{D2} D} \int_{-\infty}^{+\infty} \exp\left(-\frac{D}{4}\omega^2\right) \sum_{n=0}^{D-1} \frac{1}{\sqrt{\pi} 2^n n!} H_n^2\left(x + \frac{i\omega}{2}\right) d\omega. \tag{11.42}$$

We may first integrate (11.42) over x before integrating over ω to obtain $\tilde{C}_{D2} = C_{D2}/\sqrt{\pi D}$. The integration (11.42) becomes

$$p_D(x) = \frac{1}{2\sqrt{\pi D}} \int_{-\infty}^{+\infty} \exp\left(-\frac{D}{4}\omega^2 - \left(x + \frac{i\omega}{2}\right)^2\right) \sum_{n=0}^{D-1} \frac{1}{\sqrt{\pi} 2^n n!} H_n^2\left(x + \frac{i\omega}{2}\right) d\omega. \tag{11.43}$$

Using the substitution $s = i\omega$, the integration is very similar to the Mellin inversion formula for the Laplace transform [101] with an integration from $x - i\infty$ to $x + i\infty$. Based on the Laplace transform, we obtain

$$p_D(x) = \frac{\exp\left(-\frac{D}{D-1}x^2\right)}{\sqrt{\pi D(D-1)}} \sum_{n=0}^{D-1} \frac{1}{2^n n!} H_n^2\left(\frac{t}{2\sqrt{D-1}}\right)\Bigg|_{t^k \leftarrow (-1)^k H_k\left(\frac{Dx}{\sqrt{D-1}}\right)}. \tag{11.44}$$

In the probability density (11.44), the summation gives a $2(D-1)$-degree polynomial in t. The power t^k is algebraically substituted by the Hermite polynomial $(-1)^k H_k\left(Dx/\sqrt{D-1}\right)$.

The GD distribution (11.44) is valid for MMFs with various numbers of modes. The GD has zero mean and a variance $\frac{1}{2}(D - 1/D)$. Figure 11.4 compares $Dp_D(x)$ with the correlation function $K_D(x, x)$. Scaling the distribution (11.44) by a factor D enables comparison to $K_D(x, x)$. Figure 11.4 is plotted for $D = 2$, 6, 12, 20, 30, and 64 modes. As the distribution is always symmetric with respect to the center, Figure 11.4 shows only the positive side of the curves.

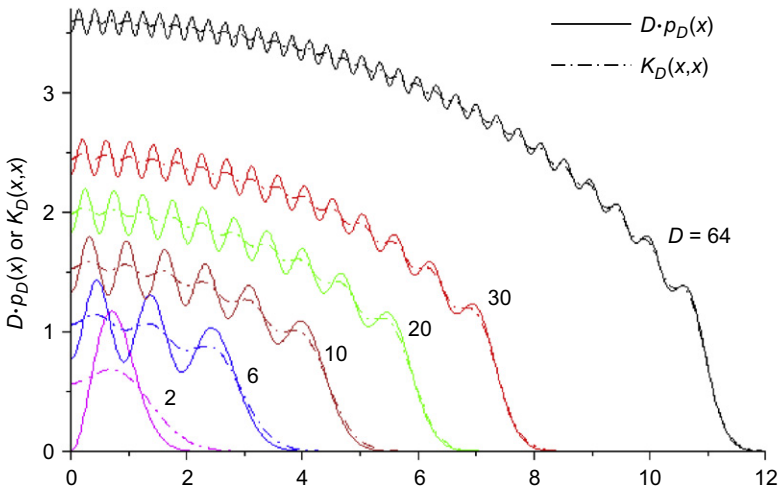

FIGURE 11.4 Probability density of the normalized GD (11.44) scaled by D, $Dp_D(x)$ shown as solid curves for $D = 2$, 6, 12, 20, 30, and 64. The correlation functions $K_D(x, x)$ given by (11.37) are shown as dash-dotted curves for comparison.

Especially for large number of modes, the general curves of (11.44) are similar to the case without trace-constraint of $K_D(x,x)$. The number of peaks is the same as the number of modes D. Each peak corresponds to values of x where individual eigenvalues tend to be concentrated. There are more ripples with zero-trace constraint (11.31) as compared to the case without the constraint. Beyond the theory of this section, further theoretical study finds that the mean of each eigenvalue x_μ, assumed ordered, is not changed by the zero-trace constraint (11.31). However, the variance of the eigenvalues x_μ is always reduced by $\frac{1}{2}D^{-1}$. The variance reduction yields more localized eigenvalues.

As the probability density function (11.44) is derived from a Gaussian unitary ensemble in which all matrix elements have a variance of 1/2, if the elements of zero-trace Gaussian unitary ensemble $\mathbf{G}$ (11.28) have average variance σ_g^2, the Gaussian unitary ensemble $\mathbf{A}$ without zero-trace constraint has variance $\sigma_g^2 D^2/(D^2-1)$. Combining all normalization factors together, given a zero-trace Gaussian unitary ensemble $\mathbf{G}$ (11.28) with average variance σ_g^2, the GD variance is $\sigma_{\rm gd}^2 = D\sigma_g^2$. Assuming all fiber sections have identical statistical properties and using (11.30) for σ_g^2, the overall GD variance is

$$\sigma_{\rm gd}^2 = K\sigma_\tau^2. \tag{11.45}$$

Regardless of number of modes, the overall STD of GD among modes is always $\sigma_{\rm gd} = \sqrt{K}\sigma_\tau$ and increases with the square-root of the number of sections. Equivalently, the overall GD increases with the square-root of system length. This result is consistent with the similar theory of PMD [36,37,89] and the results of [16,17] with strong mode coupling. Note that this is only relevant for fibers with strong mode coupling. As mentioned above, strong mode coupling is desirable to reduce the GD spread caused by MD.

The matrix $\mathbf{G}$ (11.28) is the summation of independent random matrices. The number of those random matrices is the same as the number of independent sections. Equivalently, the summation (11.28) may be considered the concatenation relationship for MD in MMF. For the case in which the number of independent sections in the fiber link is small, the matrix $\mathbf{G}$ may be calculated numerically to find the empirical GD distribution.

11.3.2.3 Few-mode fibers

Two-mode fiber is the simplest case, and may correspond to the two polarization modes in a SMF, i.e. the well-known PMD problem [36–38]. The purpose here is not to derive new properties of PMD, but to verify that the general random matrix model is applicable to PMD.

Using the density (11.44) for two-mode fiber, we obtain

$$p_2(x) = \sqrt{\frac{2}{\pi}} 4x^2 \exp(-2x^2). \tag{11.46}$$

As in the PMD literature, we define $x_{1,2} = \pm\tau/2$ with τ as the differential GD with probability density

$$p_2(\tau) = \sqrt{\frac{2}{\pi}}\tau^2 \exp\left(-\frac{\tau^2}{2}\right), \quad \tau \geqslant 0, \tag{11.47}$$

which is the well-known Maxwellian distribution with normalized mean differential GD $\bar{\tau} = 2\sqrt{2/\pi} \approx 1.60$. Random matrix models specialized to the two-mode case were used to derive the Maxwellian distribution in [37,89,102].

For fibers with up to eight modes, direct integration of (11.34) has been used to find the GD distribution analytically, as given in [10,28].

Figure 11.5 compares the analytical distribution from (11.44) with simulation results for fibers with $D = 6$ modes. The simulation is conducted similar to that in [10]. The fiber has $K = 256$ independent sections. In each section, the six modes are chosen to have deterministic delays of $\pm\tau$, three of positive sign and three of negative sign, where $\tau = 1/\sqrt{K}$, to ensure that the overall STD of GD is unity, as from (11.45). This particular choice may not correspond to a physical fiber, but simplifies the simulation. The random unitary matrices $\mathbf{U}^{(k)}$, and $\mathbf{V}^{(k)}$, $k = 1, \ldots, K$, are first initialized by 6×6 random complex Gaussian matrices and then converted to unitary matrices using the Gram-Schmidt process [98, Section 5.2.8]. All sections have independent unitary matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$. A total of 300,000 eigenvalues are used in the curves shown in Figure 11.5.

In Figure 11.5, the simulation results show excellent agreement with the analytical eigenvalue probability density $p_6(x)$. Although the modes in each section have only two GDs, with strong mode coupling, a probability density function having six peaks is obtained. In the strong-coupling regime, similar results would be obtained using

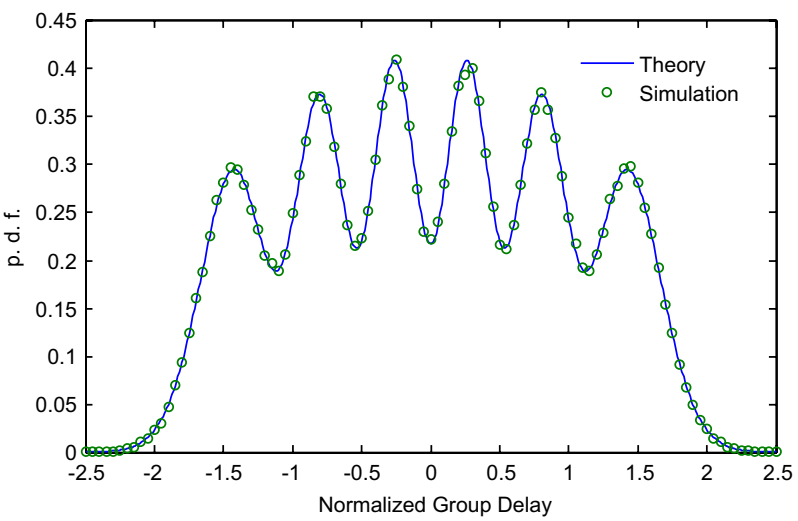

FIGURE 11.5 Probability density of GDs in a fiber with $D = 6$ modes, with GD variance normalized to unity. The curve represents the theoretical probability density (11.44) and the circles are from simulation.

any uncoupled GD vector in each section, provided that the six uncoupled GDs sum to zero and have a variance of $1/K$. To provide a more physically realistic example, four modes may have modal delays of $+\tau_1$ and the other two modes may have modal delays of $-2\tau_1$ with $\tau_1 = 1/\sqrt{2K}$. Similar results are obtained if the GDs in each section are, for example, $\pm\tau_2$, where τ_2 is random and follows a distribution with second moment $1/K$.

The simulation results in Figure 11.5 verify that zero-trace Gaussian unitary ensemble can be used to model the GD of few-mode fibers.

11.3.2.4 Many-mode fibers

With a large number of modes, a Gaussian unitary ensemble without the zero-trace constraint is described by a semicircle distribution with radius $\sqrt{2D}$ [96, Section 4.2]. With the normalization used in Sections 11.3.2.2 and 11.3.2.3, the variance of the eigenvalues is $D/2$. This semicircle law was first derived by Wigner for large random matrices [103,104]. The Wigner semicircle law is universally valid for many different types of large random matrices [105,106].

A Gaussian unitary ensemble, even with the zero-trace constraint, should follow the semicircle distribution [107,108]. As an alternative to considering $\mathbf{G}$ as a Gaussian unitary ensemble, a more straightforward derivation may use the Central Limit Theorem for free random variables.

In free probability theory, free random variables are equivalent to statistically independent large random matrices [109,110]. The Central Limit Theorem for the summation of free random variables gives the semicircle distribution [109,111]. The matrix $\mathbf{G}$ (11.28) is the summation of many independent random matrices.

The Central Limit Theorem for free random variables states the following: let $\boldsymbol{X}_{\mathrm{k}}$, $k = 1, \ldots, K$, be identically distributed independent zero-mean free random variables with unit variance, the summation

$$\mathbf{Y}_K = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \cdots \mathbf{X}_K}{\sqrt{K}} \tag{11.48}$$

is described by semicircle distribution with radius of two and unit variance:

$$p_Y(r) = \begin{cases} \frac{1}{2\pi}\sqrt{4 - r^2} & |r| < 2, \\ 0 & \text{otherwise}, \end{cases} \tag{11.49}$$

as K approaches infinity.

In the Central Limit Theorem of free random variables, when free random variables are represented by large random matrices, the distribution of the free random variables is equivalent to the distribution of the eigenvalues of the random matrices. When the theorem is applied to $\mathbf{G}$ given by (11.28), if the variance of the zero-mean GD per section is σ_τ^2 for all K sections, the eigenvalues of $\mathbf{G}$ are described by a semicircle distribution with radius $2\sqrt{K}\sigma_\tau$ and variance $K\sigma_\tau^2$. Equivalently, the GD of the MMF has a semicircle distribution with variance $K\sigma_\tau^2$. The normalization used in this section based on the eigenvalues of $\mathbf{X}_k$ and $\mathbf{Y}_K$ in (11.48) is customary in free probability theory. However, the normalization used in Section 11.3.2.2 is based on the matrix elements of $\mathbf{G}$, similar to that in Mehta [96].

Figure 11.6 compares the simulated marginal probability density of GD in fibers having $D = 16$, 64, and 512 modes to the semicircle distribution. In Figure 11.6a, the probability density for $D = 16$ modes from (11.44) is also shown for comparison. Each MMF is comprised of $K = 256$ sections. In each section, the uncoupled GDs are deterministic, with the first $D/2$ modes with a delay of τ and the other $D/2$ modes with a delay of $-\tau$. For normalization purposes, $\tau = 1/\sqrt{K}$ is chosen to facilitate comparison with a semicircle distribution with radius 2 and unit variance. The simulated curves are obtained from 1,600,000 and 640,000 and 102,400 eigenvalues for $D = 16$ and 64 and 512, respectively. The model here is valid as long as the component matrices in (11.28) may be modeled as free random variables [111].

In Figure 11.6, the simulated distributions match the semicircle distribution well for $D = 64$ and 512 modes. For a fiber having $D = 16$ modes, the distribution is close to a semicircle distribution, but has an obvious periodic structure with 16 peaks, consistent with the probability density (11.44). The ripples become less obvious as D increases from 16 to 64 to 512. Upon close examination of the curve for $D = 64$, the ripples seem periodic, consistent with the ripples in Figure 11.4.

The GD relationship $\sigma_{\text{gd}}^2 = K\sigma_\tau^2$ (11.45) remains valid when the number of modes D is very large. With a large number of modes, the relationship (11.45) can be derived directly from free probability theory.

The semicircle distribution, which describes the GDs in fibers with an infinite number of modes has strict upper and lower limits, so such fibers have a strictly bounded GD spread, given by $4\sigma_{\text{gd}}$. In fibers with large but finite number of modes D, it will be sufficient to accommodate a GD spread just slightly larger than $4\sigma_{\text{gd}}$. In the next section, the statistics of the GD spread of fibers with a finite number of modes is studied.

11.3.3 Statistics of group delay spread

As shown in Section 11.6.3.3, the complexity of MIMO equalization depends on the GD spread of the fiber [18,19]. This GD spread is equivalent to the difference between the maximum and minimum eigenvalues of a zero-trace Gaussian unitary ensemble. Comparing $\mathbf{G}$ and $\mathbf{A}$ related by $\mathbf{G} = \mathbf{A} - \text{tr}(\mathbf{A})\mathbf{I}/D$, the difference between maximum and minimum eigenvalues is the same for both $\mathbf{G}$ and $\mathbf{A}$ with and without zero-trace constraint. The GD spread may be studied by the Gaussian unitary ensemble without trace constraint. As discussed earlier, the average variance of the matrix elements of $\mathbf{A}$ is a factor of $D^2/(D^2 - 1)$ larger than that of $\mathbf{G}$. The difference between the maximum and minimum eigenvalues for a Gaussian unitary ensemble $\mathbf{A}$ without trace constraint is studied here.

Many works have studied the maximum eigenvalue of Gaussian unitary ensemble, especially when the dimension of $\mathbf{G}$ approaches infinity. For many-mode fibers, the GD spread is approximately $4\sigma_{\text{gd}}$, as mentioned in Section 11.3.2.4.

For a finite-dimensional Gaussian unitary ensemble, the GD spread may be studied using the Fredholm determinant [96, Section 20.1], which was first derived to study integral equations. The solution of an integral equation gives numerical

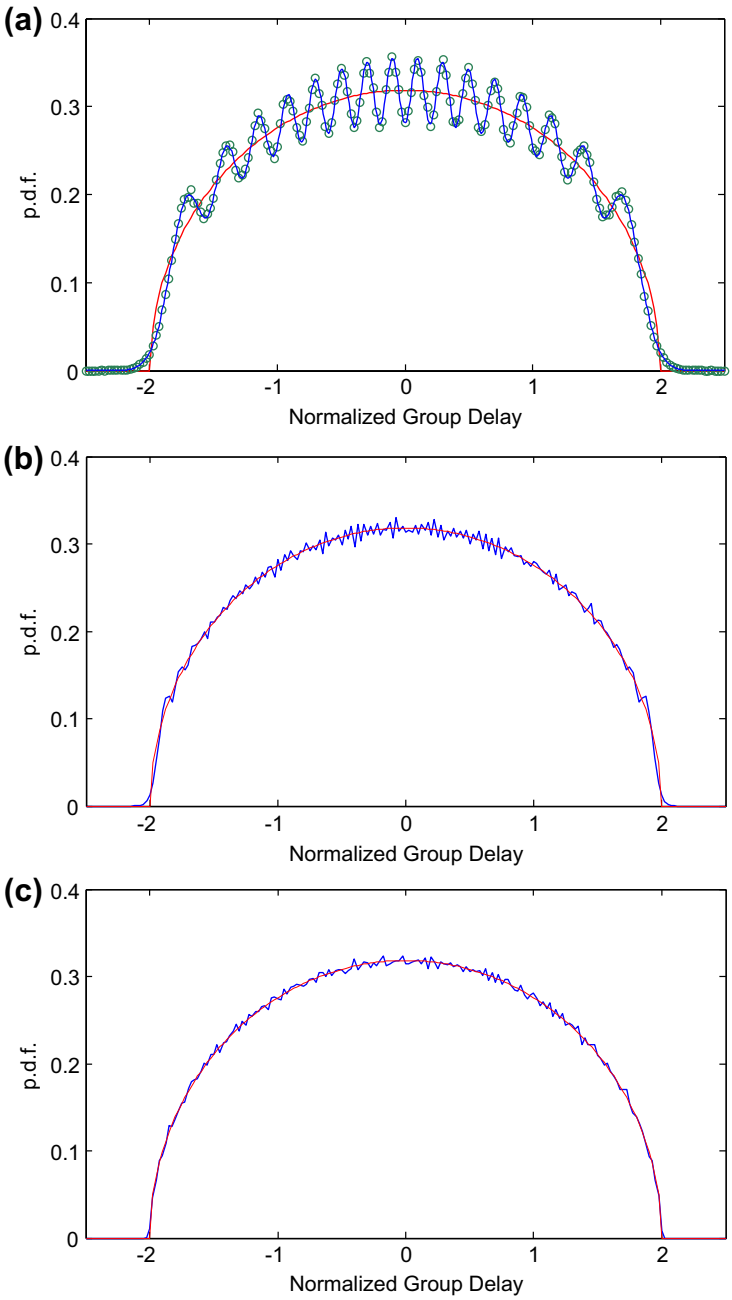

FIGURE 11.6 GD distribution compared to semicircle distribution. (a) $D = 16$: solid line with ripples from (11.44), circles from simulation. (b) $D = 64$: solid line with ripples from simulation. (c) $D = 512$: solid line with ripples from simulation. Adapted from [10].

values of the Fredholm determinant [112,113]. The joint probability density for the maximum and minimum eigenvalues is given by [114]

$$F^{(D)}(x, y) = \Pr(\lambda_{\max} \leqslant x, \lambda_{\min} \geqslant y) = E_2(0, J) = \det[1 - K_D(x, y)|_J], \tag{11.50}$$

where $J = (-\infty, y) \bigcup (x, +\infty)$ specifies a region, $E_2\,(0, J)$ denotes no eigenvalue in the region J, and $\det\left[1 - K_D(x, y)|_J\right]$ is the Fredholm determinant for the kernel $K_D(x, y)$ given by (11.37). The corresponding integral equation involves the kernel $K_D(x, y)$ integrated over the complement of J. The function $F^{(D)}(x, y)$ can be found by numerically solving the corresponding integral equation [113,114].

The GD spread distribution has a cumulative distribution of

$$\Pr(\lambda_{\max} - \lambda_{\min} \leqslant x) = -\int_{-\infty}^{+\infty} \frac{\partial F^{(D)}(t, y)}{\partial y}\bigg|_{t=y+x} dy. \tag{11.51}$$

Figure 11.7 shows the complementary cumulative distribution $\Pr\,(\lambda_{\max} - \lambda_{\min} > x)$ for fibers with $D = 6$ and 12 modes. In the curves of Figure 11.7, the overall GD

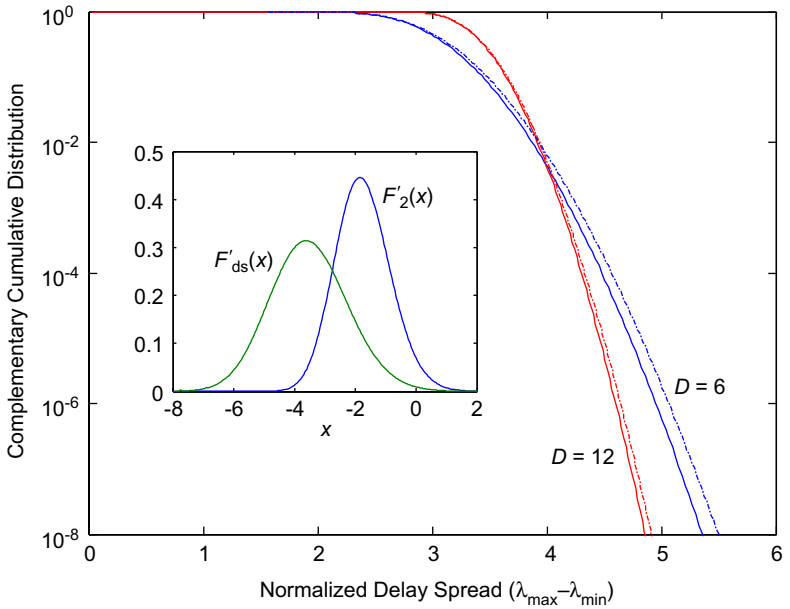

FIGURE 11.7 Complementary cumulative distribution of the GD for fibers with $D = 6$ and $D = 12$ modes, with GD variance normalized to unity. The solid curves are obtained by evaluating the Fredholm determinant (11.51) using a toolbox from [114]. The dash-dotted curves are derived using the approximation (11.54). The inset shows the Tracy-Widom distributions $F_2'(x)$ and the GD spread distribution $F_{ds}'(x)$.

variance σ_{gd}^2 is normalized to unity. Figure 11.7 is consistent with the empirical simulations in [18]. The complementary cumulative probability $\Pr(\lambda_{max} - \lambda_{min} > x)$ corresponds to the event that the normalized GD spread of the fiber exceeds a value x, which typically describes the normalized temporal memory length of an equalizer for compensating MD. Here, $\Pr(\lambda_{max} - \lambda_{min} > x)$ is referred to here as the "GD spread outage probability," although system outage does not necessarily occur when the fiber GD spread exceeds the equalizer length.

For many-mode fibers, the distribution of the GD spread can be approximated very accurately. For very large matrices, the maximum eigenvalue is described by the celebrated Tracy-Widom distribution [115,116]. For the Gaussian unitary ensemble described by the distribution (11.32), the largest eigenvalue is approximately $\sqrt{2D}$. The Tracy-Widom distribution for a Gaussian unitary ensemble is given by

$$\lim_{D\to\infty} \Pr\left(\frac{\lambda_{max} - \sqrt{2D}}{2^{-1/2}D^{-1/6}} \le x\right) = F_2(x), \tag{11.52}$$

where λ_{max} denotes the maximum eigenvalue and $F_2(x)$ is the cumulative density function for the Tracy-Widom distribution, with numerical values given in [113,114].

For many-mode fibers, as shown in [117], the maximum and minimum eigenvalues are independent of each other. The GD spread for a large Gaussian unitary ensemble is given by

$$\lim_{D\to\infty} \Pr\left(\frac{\lambda_{max} - \lambda_{min} - 2\sqrt{2D}}{2^{-1/2}D^{-1/6}} \leqslant x\right) = F_{ds}(x) = \int_{-\infty}^{+\infty} F_2(x-t)\frac{d}{dt}F_2(t)dt \tag{11.53}$$

as the summation of two independent Tracy-Widom random variables. The inset of Figure 11.7 shows the Tracy-Widom distribution $F_2'(x)$ and the corresponding GD spread distribution $F_{ds}'(x)$.

Figure 11.7 also shows the approximate complementary cumulative distribution computed using (11.53):

$$\Pr(\lambda_{max} - \lambda_{min} > x) \approx 1 - F_{ds}\left[D^{2/3}(x\sqrt{1-D^{-2}} - 4)\right] \tag{11.54}$$

for fibers with $D = 6$ and 12 modes with GD variance normalized to unity. The approximation (11.54) follows directly from (11.53) for all values of D and takes into account that the GD variance for the statistics corresponding to (11.53) is $\frac{1}{2}(D - 1/D)$. The GD spread outage probabilities in Figure 11.7 govern the temporal memory length required for MD equalization. For typical outage probabilities in the range of 10^{-4}–10^{-6}, the equalizer length should be about 4 to 6 times σ_{gd}, the STD of GD.

The approximation based on the sum of two independent Tracy-Widom random variables always over-estimates the GD spread outage probability, but is sufficiently accurate for most engineering purposes for $D \geqslant 12$. The approximation (11.54) shows that as $D \to \infty$, the delay spread is upper-bounded by four times the STD

of GD. In this limit, the GD follows a Wigner semicircle distribution from Section 11.3.2.4, which has a finite support equal to four times the STD.

The Tracy-Widom distribution has a mean of -1.77, and the normalized GD spread has a mean of approximately

$$\frac{4 - 3.54D^{-2/3}}{\sqrt{1 - D^{-2}}}. \tag{11.55}$$

The mean GD spread (11.55) approaches 4 as D increases. The Tracy-Widom distribution has an STD of 0.902 and the normalized GD spread has an STD of approximately

$$\frac{1.28D^{-2/3}}{\sqrt{1 - D^{-2}}}. \tag{11.56}$$

The GD spread STD (11.56) decreases to zero as $D^{-2/3}$ as D increases.

In Section 11.6.3.3 below, the GD spread outage probabilities in Figure 11.7 are used in quantifying the complexity of equalizers for compensating MD.

11.4 MODE-DEPENDENT LOSS AND GAIN

Multimode transmission fibers [21,22] or passive components may introduce loss that is mode-dependent, while multimode optical amplifiers [23–25] may introduce mode-dependent gain. Throughout this chapter, these effects are referred to collectively as MDL. MDL is likely to be relatively unimportant in short-range systems without amplifiers, but long-haul MDM systems with many cascaded amplifiers may accumulate significant amounts of MDL. Unlike MD, MDL is fundamentally a performance-limiting factor. In the extreme case, MDL is equivalent to a reduction in the number of propagating modes, leading to a proportional decrease in overall data rate or channel capacity.

Like the MD studied in Section 11.3, MDL can be studied using the matrix propagation model of Section 11.2.3.3. The overall propagation matrix $\mathbf{M}^{(t)}$ (11.20) describes coupled propagation of signals from input to output in an MMF, while $\mathbf{M}^{(t)*}$ describes propagation from output to input. We can define a round-trip propagation operator $\mathbf{M}^{(t)*}\mathbf{M}^{(t)}$ whose eigenmodes are the electric field profiles that, when propagated from the input to the output and then back to the input, remain unchanged except for a scaling describing gain or loss. The eigenvalues of $\mathbf{M}^{(t)*}\mathbf{M}^{(t)}$ describe the power gains or losses of these eigenmodes, and the square-root of the eigenvalues describe the scaling of the electric field amplitudes. MDL can obviously be defined equivalently in terms of the operator $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$. Since its eigenvalues describe modal power gains, we may refer to $\mathbf{M}^{(t)*}\mathbf{M}^{(t)}$ as a modal gain operator.

For polarization-dependent loss (PDL) in SMF ($D = 2$), the two eigenmodes are the modes having maximum and minimum gains. For MDL in MMF, the eigenmodes represent directions along which the gain is an extremum, which is either a maximum or minimum when approached from different orientations. For $D = 3$, MDL may be

visualized in a three-dimensional space using a triaxial ellipsoid with three unequal semi-axes, where the distance from the origin to a point on the surface represents modal gain. The directions of maximal and minimal gain are obvious. The triaxial ellipsoid has an extremum, which can be either a maximum or minimum, depending on the direction from which it is approached. This extreme point corresponds to the third eigenmode. Without MDL, the ellipsoid degenerates to a sphere, and any three orthogonal directions may be chosen as eigenmodes.

Like the GD spread caused by MD, the variation of loss caused by MDL is minimized by strong mode coupling. In the remainder of this section, we study the statistics of MDL in the strong-coupling regime. In this regime, the propagation operator of an MMF can be represented by the product of many independent random matrices of the form (11.19) which converges to a limiting form whose statistics depend only on the number of modes and the accumulated MDL. In this regime, the accumulated MDL increases with the square-root of the number of independent matrices or the square-root of the total system length. However, the overall MDL exhibits a nonlinear dependence on the accumulated MDL.

11.4.1 Statistics of strongly coupled mode-dependent gains and losses

Based on the Central Limit Theorem for the product of random matrices [28], MDL follows a limiting distribution that depends only on the number of modes and a single additional parameter. In the regime of small MDL, the following two results are valid.

Proposition I: In the strong-coupling regime, when the overall MDL is small, the distribution of the overall MDL (measured in units of the logarithm of power gain or decibels) is identical to the eigenvalue distribution of zero-trace Gaussian unitary ensemble.

Proposition II: In the strong-coupling regime, when the overall MDL is small, the STD of the overall MDL σ_{mdl} depends solely on the square-root of the accumulated MDL variance ξ (often referred to simply as accumulated MDL) via:

$$\sigma_{\text{mdl}} = \xi\sqrt{1+\frac{1}{12}\xi^2}. \tag{11.57}$$

If the MMF comprises K independent, statistically identical sections, each with MDL variance σ_g^2, the accumulated MDL is $\xi = \sqrt{K}\sigma_g$. In (11.57), σ_{mdl} and ξ are measured in units of the logarithm of power gain. Quantities measured in units of log power gain can be converted to decibels by multiplying by $\gamma = 10/\ln 10 \approx 4.34$, i.e. $\sigma_{\text{mdl}}(\text{dB}) = \gamma\sigma_{\text{mdl}}$ (log power gain). Unless noted otherwise, all expressions in this chapter assume that both ξ and σ_{mdl} are expressed in log power gain units.

Proposition I describes the shape of the MDL distribution, which is the same as (11.44) or Figure 11.4, whose variance depends on the variance of a Gaussian unitary ensemble used to describe MDL, just as such an ensemble was used to describe MD in Section 11.3. MDL has the same statistics as MD.

Propositions I and II have not been proven rigorously. Using numerical simulation, we have found that in the small-MDL region, the shape of the overall MDL distribution essentially depends only on the dimension of the zero-trace Gaussian unitary ensemble, which is the number of modes. For systems with overall MDL in the range of practical interest, $\sigma_{\text{mdl}} \leqslant 12$ dB or $\xi = \sqrt{K}\sigma_g \leqslant 10$ dB, the overall MDL (measured in decibels or log power gain) has a distribution very close to the eigenvalue distribution of zero-trace Gaussian unitary ensemble.

In Proposition II, the STD of the overall MDL depends solely on the accumulated MDL $\xi = \sqrt{K}\sigma_g$. For MD in MMF (see Section 11.3) or for PMD in SMF [36,37,89], the overall modal GD spread depends *linearly* on $\sqrt{K}\sigma_\tau$ (11.45). For MDL in MMF, the overall MDL depends *nonlinearly* on $\xi = \sqrt{K}\sigma_g$, as described by (11.57). The nonlinear relationship in Proposition II cannot be proven rigorously in the general case, but related results can be derived analytically in the limit of many modes. Proposition II has been tested by comparing (11.57) to numerical simulations. The approximation (11.57) is highly accurate for two-mode fibers with $\xi = \sqrt{K}\sigma_g$ up to 10 dB, and the region of its validity increases with an increasing number of modes.

In the region with $\xi = \sqrt{K}\sigma_g$ larger than 10 dB, which is beyond the range of practical interest, the overall MDL measured in log power gain units has the same statistical properties as the eigenvalues of the summation of two matrices [118]

$$\xi\mathbf{G} + \kappa_D \xi^2 \mathbf{F}, \tag{11.58}$$

where $\mathbf{G}$ is a zero-trace Gaussian unitary ensemble similar to that used to describe MD in Section 11.3, except it has unit eigenvalue variance, $\mathbf{F}$ is a deterministic uniform matrix, and $\kappa_D = \frac{1}{2}D/(1+D)$ is a constant lying between 1/3 and 1/2, depending on number of modes D. The uniform matrix $\mathbf{F}$ has its eigenvalues deterministically and uniformly distributed between ± 1. For $D = 2$, an example is $\mathbf{F} = \text{diag}[1,-1]$ or any other Hermitian matrix having the same eigenvalues. From the theory of MD in Section 11.3.2.1, the summation (11.58) represents the concatenation of two MMFs: the first with strong mode coupling represented by $\mathbf{G}$ and the second with deterministic and uniform MD represented by $\mathbf{F}$.

Comparing (11.57) with (11.58), the Gaussian unitary ensemble $\mathbf{G}$ gives the linear term ξ (or the 1 inside the square-root) to the overall MDL. With $\kappa_\infty = 1/2$, the uniform matrix $\mathbf{F}$ gives the nonlinear factor ($\xi^2/12$ inside the square-root). To a certain extent, Propositions I and II approximate the zero-trace uniform matrix $\mathbf{F}$ using a zero-trace Gaussian unitary ensemble. The approximation in Propositions I and II is sufficiently accurate for practical purposes.

Using $\kappa_D = \frac{1}{2}D/(1+D)$ and (11.58), the accuracy of Proposition II may be improved slightly for few-mode fibers. The improved Proposition II is

$$\sigma_{\text{mdl}} = \xi\sqrt{1 + \frac{\xi^2}{12(1-D^{-2})}}, \tag{11.59}$$

which gives more accurate values of overall MDL for $D = 2$, 3 but gives results similar to (11.57) for $D \geqslant 4$. The improvement in (11.59) is useful only in the large-MDL regime.

11.4.2 Model for mode-dependent loss and gain

In the strong-coupling regime, a MMF is divided into K independent sections, with each section modeled as a random matrix, similar to that in Section 11.3. The length of each section should be at least equal to the correlation length, such that each section can be considered independent of the others [10,28–30].

The fiber system is described by a product of matrices of the form (11.20) with each matrix given by the product of three matrices of the form (11.19). Mode-dependent CD is neglected. Initially, in modeling MDL at one frequency, the frequency dependence of MDL is neglected, so in the diagonal matrix (11.17), the modal delay vector (11.16) is set to zero, and only the modal gain vector $\mathbf{g}^{(k)} = \left(g_1^{(k)}, g_2^{(k)}, \ldots, g_D^{(k)}\right)$ (11.15) is non-zero. Hence, uncoupled propagation in the kth section is described by

$$\Lambda^{(k)} = \begin{pmatrix} e^{\frac{1}{2}g_1^{(k)}} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}g_D^{(k)}} \end{pmatrix} \tag{11.60}$$

In the overall propagation operator $\mathbf{M}^{(t)}$ given by (11.20), overall MDL is described by the singular values of $\mathbf{M}^{(t)}$ or, equivalently, by the eigenvalues of $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$ or $\mathbf{M}^{(t)*}\mathbf{M}^{(t)}$, which are both Hermitian matrices. Mathematically, the eigenvalues of $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$ are the squares of the singular values of $\mathbf{M}^{(t)}$. The singular values of $\mathbf{M}^{(t)}$ describe electric field gains, while the eigenvalues of $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$ describe power gains. As described earlier, $\mathbf{M}^{(t)*}\mathbf{M}^{(t)}$ or $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$ can be considered a round-trip propagation or modal gain operator.

As in MIMO wireless systems [26,27], at any single frequency, using singular value decomposition (SVD), the overall matrix $\mathbf{M}^{(t)}$ can be decomposed into D spatial channels:

$$\mathbf{M}^{(t)} = \mathbf{V}^{(t)}\mathbf{\Lambda}^{(t)}\mathbf{U}^{(t)*}, \tag{11.61}$$

where $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ are input and output unitary beam-forming matrices, and we have defined

$$\Lambda^{(t)} = \begin{pmatrix} e^{\frac{1}{2}g_1^{(t)}} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}g_D^{(t)}} \end{pmatrix} \tag{11.62}$$

Here, $\mathbf{g}^{(t)} = \left(g_1^{(t)}, g_2^{(t)}, \ldots, g_D^{(t)}\right)$ is a vector of the logarithms of the eigenvalues of $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$, which quantifies the overall MDL of the MIMO system. Our goal here is to study the statistics of the overall MDL described by $\mathbf{g}^{(t)}$.

In order to describe MDL properly, we need to explain the difference between accumulated MDL and overall MDL. Accumulated MDL refers to the sum of the uncoupled MDL values in all K sections comprising a fiber. The variance of the accumulated MDL is:

$$\xi^2 = \sigma_{g^{(1)}}^2 + \sigma_{g^{(2)}}^2 + \cdots + \sigma_{g^{(k)}}^2, \tag{11.63}$$

where $\sigma^2_{g(k)}, k = 1, \cdots, K$, are the variances of the uncoupled MDL vectors in the individual sections. The accumulated MDL is similar to the variance of the elements of the MD operator in (11.27). If the individual sections are statistically identical, we have $\xi = \sqrt{K}\sigma_g$, where σ_g^2 is the variance of the uncoupled MDL in each section. This accumulated MDL $\xi = \sqrt{K}\sigma_g$ increases with the square-root of number of sections or the square-root of the total system length.

Overall MDL refers to the end-to-end coupled MDL of a fiber comprising K independent sections. The overall MDL is computed from the gain vector $\mathbf{g}^{(t)}$ that appears in (11.62). The gains in $\mathbf{g}^{(t)}$ are assumed to be ordered as $g_1^{(t)} \geqslant g_2^{(t)} \geqslant \cdots \geqslant g_D^{(t)}$, and are assumed to sum to zero: $g_1^{(t)} + g_2^{(t)} + \cdots + g_D^{(t)} = 0$. When taken together, the D elements of $\mathbf{g}^{(t)}$ have zero mean; equivalently, an element $g_\mu^{(t)}$ chosen randomly from $\mathbf{g}^{(t)}$ has zero mean.

Two statistical parameters are especially useful for characterizing MDL: the STD of overall MDL σ_{mdl} and the mean of the maximum MDL difference $\left\langle g_1^{(t)} - g_D^{(t)} \right\rangle$. The maximum MDL difference is similar to the GD spread that is the difference between the maximum and minimum eigenvalues in Section 11.3.3; equivalently it is the condition number of the overall matrix $\mathbf{M}^{(t)}$ (11.20) expressed on a logarithmic scale. The mean of the maximum MDL difference, $\left\langle g_1^{(t)} - g_D^{(t)} \right\rangle$, quantifies the gain difference between the strongest and weakest modes. In a two-mode fiber, $\left\langle g_1^{(t)} - g_2^{(t)} \right\rangle$ is commonly referred to as the mean PDL.

The STD of overall MDL is computed over all D elements of the gain vector $\mathbf{g}^{(t)}$; equivalently, it is the square-root of $\sigma^2_{\text{mdl}} = \left\langle \left(g_\mu^{(t)} \right)^2 \right\rangle$, where $g_\mu^{(t)}$ is an element chosen randomly from $\mathbf{g}^{(t)}$. The STD of overall MDL σ_{mdl} has the same units as the gains in $\mathbf{g}^{(t)}$.

The accumulated MDL variance ξ^2, given by (11.63), does not equal the variance of overall MDL because of the nonlinearity inherent in Proposition II, given by (11.57). Much of the complexity of PDL and MDL arises from the nonlinearity of (11.57).

11.4.3 Properties of the product of random matrices

Characterizing the statistics of the singular values of $\mathbf{M}^{(t)}$, or those of the eigenvalues of $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$, is the key to understanding the performance of MDM systems, as in MIMO wireless systems [26,27]. The analysis here is complicated by the fact that $\mathbf{M}^{(t)}$ is the product of random matrices by (11.20), rather than the sum of random matrices (11.28). Since the early works [119,120], there have been many studies on the statistics of the products of random matrices, but most addressed the Lyapunov exponent of the products [120–122]. We are interested here in the statistics of the eigenvalues of a product of matrices, not the Lyapunov exponent.

Because matrix multiplication is not commutative, i.e. **AB** is not generally equal to **BA**, even for square matrices, the logarithm of the product of two matrices,

$\log \mathbf{AB}$, is not equal to the sum of $\log \mathbf{A}$ and $\log \mathbf{B}$. Unlike the product of positive random variables (that do commute), which has its central limit as the log-normal distribution, the product of positive-definite random matrices (those with positive eigenvalues) does not generally have its central limit as the exponent of a Gaussian unitary ensemble. The Central Limit Theorem for the summation of random matrices is not necessary helpful for the understanding of the products of random matrices.

For any matrix $\mathbf{X}$ and a very small number δ, we have $\log(\mathbf{I} + \delta\mathbf{X}) \approx \delta\mathbf{X}$, where $\mathbf{I} + \delta\mathbf{X}$ is intended to describe a matrix $\mathbf{M}^{(k)}$ when the gain vector $\mathbf{g}^{(k)}$ has small norm. If both matrices $\mathbf{A}$ and $\mathbf{B}$ are positive-definite and both $\log \mathbf{A}$ and $\log \mathbf{B}$ are small, $\log \mathbf{AB} \approx \log \mathbf{A} + \log \mathbf{B}$. As an approximation, the product of positive-definite random matrices with small logarithm has a central limit as the exponential of a Gaussian ensemble. When applied to the overall product matrix $\mathbf{M}^{(t)}$, if all gain vectors $\mathbf{g}^{(k)}$ are small, the matrix $\mathbf{M}^{(t)}$ is the exponential of the Gaussian ensemble. The approximation used here is similar to the results of Berger [123]. This small-gain approximation yields Proposition I, but not Proposition II. If we made the approximation $\log(\mathbf{I} + \delta\mathbf{X}) \approx \delta\mathbf{X}$, Proposition II would be a linear relationship, unlike the nonlinear relationship of both (11.57) and (11.59). Of course, the approximation (11.57) is linear when ξ is far less than unity. Numerical simulation shows that Proposition I remains valid for ξ up to 10 dB (about 2.3 in log power gain units), a regime in which, equivalently speaking, $\log \mathbf{A}$ and $\log \mathbf{B}$ are not very small. The nonlinearity in (11.57) may be understood as related to the second-order term in the approximation $\log(\mathbf{I} + \delta\mathbf{X}) \approx \delta\mathbf{X} - \frac{1}{2}\delta^2\mathbf{X}^2$. The factor of 1/2 in this approximation is difficult to relate to the factor of 1/12 in (11.57), however. Also, the more accurate model has the factor 1/12 related to the matrix dimension in (11.59).

For random matrices of the form (11.19), at a single frequency, we are interested in the statistics of the singular values of the product (11.20) in the decomposition (11.61). Approximate results were obtained for the 2×2 case in the study of PDL in SMF [124,125], and for very large matrices in the study of free random variables [109,110]. In both cases, for positive-definite matrices $\mathbf{A}$ and $\mathbf{B}$, the product $\mathbf{AB}$ may be interpreted as $\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2*}$, similar to free probability theory. The repetition of $\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2*}$ with $\mathbf{A} = \mathbf{M}^{(2)}\mathbf{M}^{(2)*}$ and $\mathbf{B} = \mathbf{M}^{(1)}\mathbf{M}^{(1)*}$ from $k = 2$ to K yields $\mathbf{M}^{(t)}\mathbf{M}^{(t)*}$.

The simplest possible case, a two-mode fiber modeled using 2×2 matrices, describes PDL in SMF, where approximate analytical results were derived for the small-PDL regime [124] and extended to the large-PDL regime [126]. Both [124,125] showed that for low PDL, the PDL (measured in decibels or log power gain units) has a Maxwellian distribution, the same as the distribution of the GD in SMF with PMD [89,102]. As shown in (11.47) of Section 11.3.2.3, the Maxwellian distribution is the eigenvalue distribution for a zero-trace 2×2 unitary Gaussian ensemble. The zero-trace 2×2 unitary Gaussian ensemble may be obtained by the summation of many zero-trace 2×2 Hermitian matrices. Numerical simulations confirm that Propositions I and II are correct for an accumulated MDL ξ smaller than 10 dB.

Using the notation here, the exact PDL distribution [126] is

$$p_2(x) = 3\sqrt{\frac{6}{\pi}}\frac{x\sinh x}{\xi^3}\exp\left(-\frac{3x^2}{2\xi^2}-\frac{\xi^2}{6}\right), \quad x \geqslant 0, \tag{11.64}$$

which can be approximated accurately by a Maxwellian distribution even in the high-PDL regime.

The exact distribution (11.64) is a non-central chi distribution with three degrees of freedom [127]. The Maxwellian distribution is the "central" chi distribution with three degrees of freedom [127]. The chi distribution is not as well known as the chi-square distribution [59, Section 2.3]. In $D = 2$, PDL has a non-central Maxwellian distribution but with very specific noncentrality parameter of $\xi/3$.

The exact PDL distribution (11.64) is the same as that of the exact MDL model (11.58) with $\kappa_2 = 1/3$, corresponding to concatenation of a random Maxwellian distributed PMD with root-mean-square differential GD (DGD) ξ and deterministic PMD with DGD $\xi^2/3$ that is analyzed in [102]. The large-PDL model of [126] is identical to the exact model (11.58) [118]. The exact model gives $\sigma_{\text{mdl}} = \xi\sqrt{1+\xi^2/9}$, the same as (11.59) for $D = 2$, which is very close to the approximation (11.57), supporting Proposition II.

From the above discussion of the PDL, the comparison of the results of [124,125] with that in Section 11.3.2.3 supports Proposition I. By comparing (11.64) from [126] with the similar distribution derived from (11.58), the PDL of SMF is exactly the same as the exact model of (11.58), which reduces to Propositions I and II in the small-MDL regime.

Free random variables are equivalent to large matrices of the form (11.19) [109,110], and the statistics of free random variables are the statistical properties of the eigenvalues of the large matrices. The Central Limit Theorem for the summation of independent free random variables gives a semicircle distribution from Section 11.3.2.4, similar to the distribution of the eigenvalues of a large class of large random matrices [105,106]. In free probability theory, the semicircle distribution serves a function analogous to the normal distribution, which is the central limit for the summation of random variables in traditional probability theory [128, Section 5.10.4].

In traditional probability theory, the product of independent positive random variables has a central limit as the lognormal distribution. The lognormal distribution can model shadowing in wireless systems [60, Section 11.3.2.9] or distortion by stimulated Raman scattering in optical communication systems [129]. As shown in [130], the log-semicircle distribution is found to be the central limit of the product of positive free random variables if the free random variables have small variance; in the notation used here, this corresponds to ξ, defined in (11.63) having a small value.

From the results of [130], equivalently speaking, when the MDL is small and in the strong-coupling regime, the overall MDL has a log-semicircle distribution. The modal GDs follow a semicircle distribution in the limit of a large number of modes in Section 11.3.2.4. MDL (measured in decibels or log power gain) has the same

distribution as the modal GDs in the limit of a large number of modes. Expression (11.58) is basically equivalent to the results in [130], which were derived from the free probability theory.

For MMF with larger number of modes, the overall MDL from (11.57) is the same as the more exact model from (11.58). We may conclude that Proposition II is valid in both small and large MDL regime for MMF with large number of modes.

11.4.4 Numerical simulations of mode-dependent loss and gain

Numerical simulation verifies the approximation in Propositions I and II and the more exact model (11.58). PDL in two-mode fiber has well-known exact results in [126], consistent with Propositions I and II in the small-PDL regime and described exactly by the model of (11.58) is all regimes. For many-mode fibers, analytical results from free-probability theory are consistent with the exact model of (11.58) and Proposition II in all regimes, and Proposition I in the low-MDL regime. Results for six-mode fibers are shown to match theory in Section 11.4.4.2.

11.4.4.1 Many-mode fibers

For fibers with a large number of modes, the modal GDs have the same statistical properties as the eigenvalues of a large zero-trace Gaussian unitary ensemble from Section 11.3.2.4. The eigenvalues of large random matrices have a semicircle distribution, as shown by Wigner [103,104] and verified numerically in Figure 11.6. The summation of free random variables also gives a semicircle distribution, as explained in Section 11.3.2.4.

Figure 11.8a shows the simulated eigenvalue distribution of a fiber with 64 modes. Similar to the simulation in Figure 11.6b, the fiber has $K = 256$ sections and all unitary matrices are generated by the method described in Section 11.3.2.3 or [28]. Each curve of Figure 11.8a is constructed using 64,000 eigenvalues. Each MMF section has uncoupled modal gains $g_\mu^{(k)} = \pm\alpha$, where $\alpha = \sigma_g$. The first 32 modes have gains of $+\alpha$, and the last 32 modes have gains of $-\alpha$, such that the gains sum to zero. The x-axis of each curve in Figure 11.8a is normalized by the simulated STD of the overall MDL σ_{mdl}.

In Figure 11.8a, the simulated eigenvalue distribution is very close to the semicircle distribution up to $\xi = 15$ dB. The exact model (11.58) is also shown in Figure 11.8a as dashed lines. The distribution given by (11.58) is not available analytically for very large ξ but the moment generating function is given by [118, 130]

$$\exp\left(\frac{1}{2}\xi^2 s\right)\, {}_1F_1(1-s;\, 2;\, -\xi^2 s), \tag{11.65}$$

where ${}_1F_1(a, b; z)$ is the confluent hypergeometric function. The curves of Figure 11.8a are shown as the inverse Fourier transform of (11.65) with $s = i\omega$.

For ξ less than 15 dB, there is no significant difference between the semicircle distribution and the exact model. Even for ξ up to 20 dB, the simulation results are

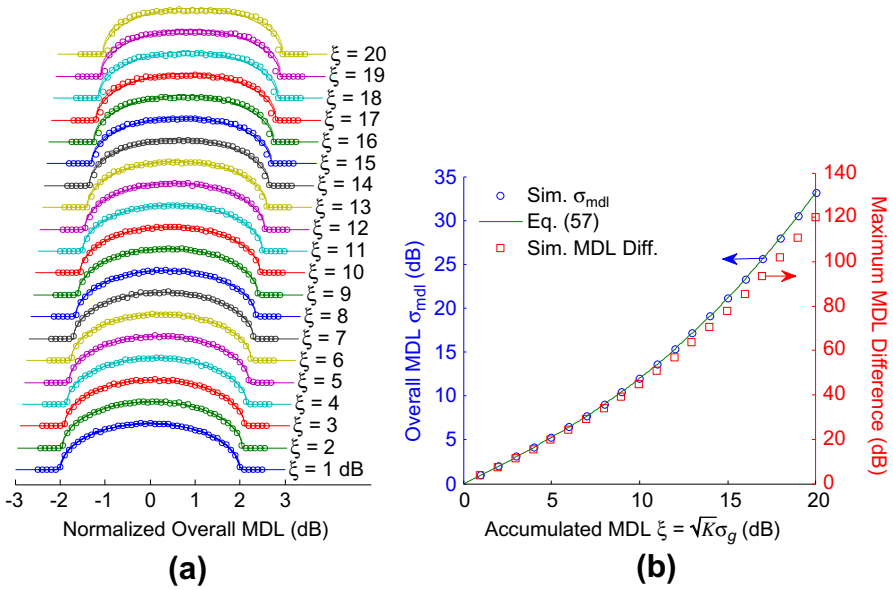

FIGURE 11.8 Mode-dependent loss (MDL) in fiber with $D = 64$ modes. (a) Comparing simulated distribution of the overall MDL (circles) with the semicircle distribution (solid curves) and exact model from (11.65) (dashed curves). The *x*-axis of each curve is normalized by the simulated STD of the overall MDL σ_{mdl}. (b) Comparing simulated STD of the overall MDL (circles) to approximation (11.57) (solid curves). The simulated mean maximum MDL difference is also shown (squares). Adapted from [28].

closer to the exact model but the difference between the exact and semicircle distribution is small.

Figure 11.8b compares the simulated STD of the overall MDL σ_{mdl} as a function of $\xi = \sqrt{K}\sigma_g$ to the approximation for σ_{mdl} given by (11.57) for a fiber with $D = 64$ modes. The approximation (11.57) agrees with simulated results within 0.01 dB for ξ up to 10 dB. For ξ from 10 to 20 dB, the overall MDL approximation (11.57) is always smaller than the simulated results and the discrepancy between the simulations and the approximation (11.57) increases to 0.15 dB. The discrepancy may arise from our simulating matrices of size $D = 64$ instead of infinitely large matrices. The discrepancy may also be caused by numerical uncertainty. Nevertheless, the approximation (11.57) can be considered highly accurate.

Figure 11.8b also shows the simulated maximum MDL difference, which is the mean of the maximum gain difference $\left\langle g_1^{(t)} - g_{64}^{(t)} \right\rangle$. In the range of ξ up to 15 dB, where the simulated distribution is close to the semicircle distribution (as seen in Figure 11.8a), the STD of overall MDL σ_{mdl} is as high as 33.4 dB and the maximum MDL difference is as high as 81 dB. Practical MDM systems should have MDL well below those values.

We conclude that for fibers with 64 modes, Proposition I is valid for ξ up to 15 dB, corresponding to σ_{mdl} up to 21.2 dB. The overall MDL approximation (11.57) from Proposition II is valid for values of ξ up to 20 dB. Over the range of validity of Proposition II, there is no observable difference between the exact model (11.58) and simulation results.

11.4.4.2 Six-mode fibers

The previous section confirms numerically that Propositions I and II are valid for fibers with $D = 64$ modes. The theoretical comparisons in Section 11.4.3 also found that Propositions I and II are valid for PDL in SMF, i.e. $D = 2$ modes. The eigenvalues of small zero-trace Gaussian unitary ensembles can be derived analytically by direct integration [10,28] or by algebraic substitution, as in (11.44).

Figure 11.9 compares results simulated for $D = 6$ modes with the approximate distribution (11.44), shown by solid lines. The approximate distribution (11.44) according to Proposition I is scaled by the overall MDL (11.57) according to Proposition II. The exact model (11.58), with distribution given by [118], is also shown by dashed lines. The MMF has $K = 256$ sections. Each curve is constructed using 300,000 eigenvalues.

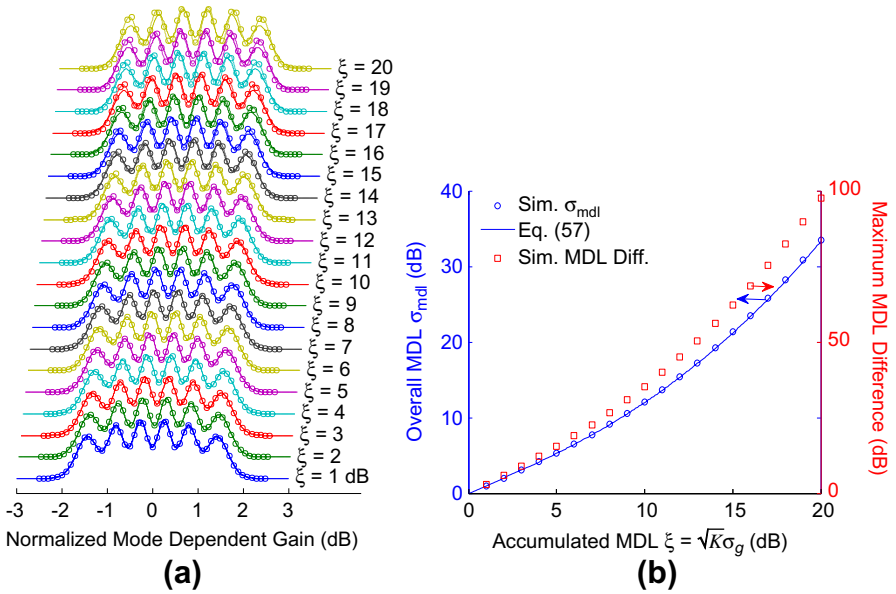

FIGURE 11.9 MDL in MMF with $D = 6$ modes. (a) Comparing simulated distribution of the overall MDL (circles) with the six-peak distribution from Figure 11.5 (solid curves). The x-axis of each curve is normalized by the simulated STD of the overall MDL σ_{mdl}. Also shown as dashed curves are the distributions derived from the exact model (11.58). (b) Comparing simulated STD of the overall MDL (circles) to approximation (11.57) (solid curves). The simulated mean maximum MDL difference is also shown (squares).

In Figure 11.9a, for a fiber with $D = 6$ modes, we observe that the simulated distribution matches the probability density (11.44) until $\xi = 13$ dB. The exact model (11.58) is valid up to $\xi = 20$ dB. As explained in Section 11.3, the number of peaks is the same as the number of modes, and each peak corresponds to values where the gain is concentrated.

In Figure 11.9b, for a fiber with $D = 6$ modes, we observe that the approximation for σ_{mdl} given by (11.57) is always smaller than the simulated results. The discrepancy is up to 0.05 dB for ξ up to 10 dB, but increases to 0.20 dB for ξ up to 20 dB. The exact model should provide a more accurate value of overall MDL, but the difference is already very small.

The numerical simulations of Figure 11.8 and Figure 11.9 show that Propositions I and II are valid in the small-MDL regime. With Proposition II and to a large extent even the exact model (11.58), the overall MDL depends solely on the accumulated MDL ξ. In strong-coupling regime, similar to the MD relationship (11.45), MDL depends on the square-root of the number of independent sections or the square-root of system length. With strong coupling, due to averaging effects, both MD and MDL depend on the square-root of the system length, i.e. strong coupling reduces both MD and MDL.

11.4.5 Spatial whiteness of received noise

In multimode systems using optical amplifiers, the amplified spontaneous emission noises generated in different uncoupled modes should be statistically independent. The MDL described in the previous sections affects both signal and noise, however, potentially making the noise spatially non-white. In system simulation using the matrix cascade of (11.20), it is straightforward to include multiple noise sources to capture this spatial correlation, as in [29]. It is helpful to observe, however, that if the number of noise sources is very large, the noises in different modes should have nearly the same power, and should be nearly uncorrelated with each other. This spatial whiteness is explained here.

Consider a system comprising K spans. For simplicity, suppose the Kth (last) fiber section contains a noise source, and that in the local uncoupled modes of the Kth section, it contributes independent noises with powers $\sigma^2_{K,\mu}, \mu = 1, \ldots, D$. At the fiber output, the electric fields from this noise source are described by a vector $\mathbf{V}^{(K)}\text{diag}\left[\sigma_{K,1}, \sigma_{K,2}, \ldots, \sigma_{K,D}\right]\mathbf{n}$, where $\mathbf{n}$ is a D-dimensional Gaussian noise vector having independent identically distributed (i.i.d.) elements with unit variance. At the fiber output, the noise correlation matrix is:

$$\mathbf{V}^{(K)}\text{diag}\left\lfloor\sigma^2_{K,1}, \sigma^2_{K,2}, \ldots, \sigma^2_{K,D}\right\rfloor\mathbf{V}^{(K)^*}. \tag{11.66}$$

For a given realization of the random unitary matrix $\mathbf{V}^{(K)}$, the output noises may be neither independent nor identically distributed. Taking an expectation over all random matrices $\mathbf{V}^{(K)}$ yields:

$$\left\langle\mathbf{V}^{(K)}\text{diag}\left[\sigma^2_{K,1}, \sigma^2_{K,2}, \ldots, \sigma^2_{K,D}\right]\mathbf{V}^{(K)^*}\right\rangle = \frac{\sigma^2_{K,1} + \sigma^2_{K,2} + \cdots + \sigma^2_{K,D}}{D}\mathbf{I}, \tag{11.67}$$

which is a constant times the identity matrix, and hence describes a noise vector with i.i.d. elements.

Considering a noise source in the kth fiber section, its output noise contribution is

$$\mathbf{M}^{(k)} \cdots \mathbf{M}^{(k+1)} \mathbf{V}^{(k)} \text{diag}\left[\sigma_{k,1}, \sigma_{k,2}, \ldots, \sigma_{k,D}\right] \mathbf{n}.$$

The SVD of $\mathbf{M}^{(K)} \cdots \mathbf{M}^{(k+1)} \mathbf{V}^{(k)} \text{diag}\left[\sigma_{k,1}, \sigma_{k,2}, \ldots, \sigma_{k,D}\right]$ may be assumed to yield $\tilde{\mathbf{V}}_n^{(k)} \tilde{\Lambda}_n^{(k)} \tilde{\mathbf{U}}_n^{(k)}$, with output noise correlation matrix equal to $\tilde{\mathbf{V}}_n^{(k)} (\tilde{\Lambda}_n^{(k)})^2 \tilde{\mathbf{V}}_n^{(k)*}$, which is of the same form as (11.66), with $\tilde{\mathbf{V}}_n^{(k)}$ independent of $\mathbf{V}^{(K)}$.

When the total number of noise sources is very large, by the law of large numbers [128, Section 7.4], the overall noise correlation matrix converges to a form similar to (11.66). The law of large numbers is applicable provided the number of noise sources is large and the $\tilde{\mathbf{V}}_n^{(k)}$ for each k indexing a noise source are independent, random unitary matrices. Based on the law of large numbers, at the fiber output, the noises in the different modes are i.i.d. Ideally, the receiver uses $\mathbf{V}^{(t)*}$ to diagonalize the channel, as in (11.61). After diagonalization, provided the number of noise sources is large and the $\mathbf{V}^{(t)*} \tilde{\mathbf{V}}_n^{(k)}$ for all values of k indexing noise sources are independent, random unitary matrices, the law of large numbers is applicable, and the noises in the different diagonal spatial channels are i.i.d. Because of the relationship between $\mathbf{V}^{(t)}$, $\tilde{\mathbf{V}}_n^{(k)}$, and the individual $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$, the matrices $\mathbf{V}^{(t)*} \tilde{\mathbf{V}}_n^{(k)}$ are random unitary matrices, except in special cases.

Typically, the law of large numbers is concerned with averages. The overall noise correlation matrix is a summation over the correlation matrices corresponding to all the independent noise sources. The normalized cross-correlation of the noise is characterized by the ratios between its off-diagonal elements and its diagonal elements that are variances proportional to the number of noise sources. Thus, the normalized cross-correlation is implicitly an "average," to which the law of large numbers is applicable.

Figure 11.10 shows simulations quantifying the spatial non-whiteness of the output noise as a function of $1/\sqrt{K}$, where K is the number of noise sources. These simulations, for $D = 8$ modes, are similar to those in [28]. All K noise sources are spatially white, i.e. they contribute i.i.d. noises with equal variance in all D uncoupled modes. The accumulated MDL $\xi = \sqrt{K}\sigma_g$ is held constant at either 5 or 10 dB.

Given a realization of the $\mathbf{M}^{(k)}, k = 1, \ldots, K$, described by (11.19) with $\mathbf{\Lambda}^{(k)}$ given by (11.60) and a realization of the K noise sources, a realization of the D output noises is obtained; the absolute square of its discrete Fourier transform yields a realization of a spatial noise spectrum. Taking an ensemble average of the spatial noise spectrum over realizations of the noise sources yields a spatial spectral distribution, which quantifies the spatial whiteness of the output noise. If the output noises are uncorrelated (and thus i.i.d., since they are jointly Gaussian), the spatial spectral distribution is white, whereas if the noises are correlated, the spatial spectral distribution is non-white. The spatial spectral distribution is also the discrete Fourier transform of the spatial autocorrelation sequence of the D output noises [128, Section 9.3.4]. The procedure for calculating the spatial spectral distribution from the spatial autocorrelation sequence is similar to the serial correlation test for randomness described in [131, Section 3.3.2.K]. Figure 11.10

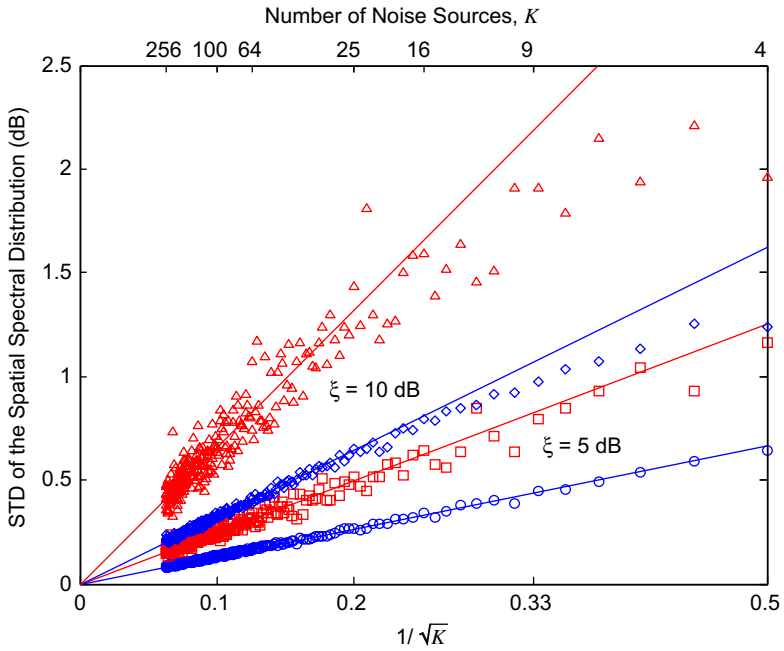

FIGURE 11.10 Spatial non-whiteness of the output noise, quantified by the STD of the spatial spectral distribution as a function of $1/\sqrt{K}$, where K is the number of noise sources, for $D = 8$ modes. The markers are simulation results, and the solid lines are fitted numerically. Upper and lower data sets for the corresponding accumulated MDL represent the maximum and mean of the STD over 100 random realizations at each value of K. The upper two data sets are for accumulated MDL $\xi = 10$ dB and the lower two are for accumulated MDL $\xi = 5$ dB. Adapted from [28]. (For interpretation of the color in this figure, the reader is referred to the web version of this book.)

quantifies non-whiteness by the STD of the spatial spectral distribution, presenting both the worst-case (maximum) and mean of the STD among 100 realizations of $\mathbf{M}^{(k)}$ for each value of K. The number of noise sources K ranges from 4 to 256.

In Figure 11.10, the spatial spectral distribution is observed to approach a white spectrum as K increases, with the STDs decreasing with the square-root of K, as given by the law of large numbers [128, Section 5.10.3]. The straight lines in Figure 11.10 indicate the best-fit slope of the STDs as function of $1/\sqrt{K}$. For different values of accumulated MDL $\xi = \sqrt{K}\sigma_g$, the slope is found to be approximately proportional to the value of overall MDL σ_{mdl} (11.57). While the simulated worst-case STD at any particular value of K is not statistically significant, slopes fitted to the worst-case STDs are clearly about twice as large as those fitted to the mean STDs. If a mean STD of 0.5 dB is desired, K needs to be larger than 8 and 42 for $\xi = 5$ and 10 dB, respectively. If a worst-case STD of 0.5 dB is desired, K needs to be larger than 25 and 160 for $\xi = 5$ and 10 dB, respectively. In any case, the spatial spectral distribution exhibits far smaller variations than the signal power variations caused by MDL.

Figure 11.10 confirms that spatial whiteness is a good assumption for systems with many noise sources, such as the long-haul systems considered in this chapter. This assumption may not be accurate for systems with a small number of noise sources, however.

11.4.6 Frequency-dependent mode-dependent loss and gain

In general, MDL depends on frequency if the diagonal matrix in each section, $\mathbf{\Lambda}^{(k)}(\omega)$ (11.17), includes both non-zero MDL vector $\mathbf{g}^{(k)}$ (11.15) and non-zero modal GDs vector $\boldsymbol{\tau}^{(k)}$ (11.16). The overall propagation matrix $\mathbf{M}^{(t)}(\omega)$, given by (11.20), is frequency-dependent. This frequency dependence of MDL gives frequency diversity. Frequency-dependent MDL has a coherence bandwidth that should be inversely proportional to the STD of GD, σ_{gd}. If MDM signals occupy a bandwidth far larger than the MDL coherence bandwidth, because of statistical averaging, signals at frequencies with large MDL are complemented by signals at frequencies with small MDL. This frequency diversity is described here, and its system implications are explained in Section 11.6.2.

Similar to MIMO wireless systems [26,27], at any single frequency, using SVD, the overall matrix $\mathbf{M}^{(t)}(\omega)$ can be decomposed into D spatial channels:

$$\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}(\omega)\mathbf{\Lambda}^{(t)}(\omega)\mathbf{U}^{(t)}(\omega)^*, \tag{11.68}$$

where $\mathbf{U}^{(t)}(\omega)$ and $\mathbf{V}^{(t)}(\omega)$ are frequency-dependent input and output unitary beamforming matrices, respectively, and

$$\mathbf{\Lambda}^{(t)}(\omega) = \begin{pmatrix} e^{\frac{1}{2}\mathrm{g}_1^{(t)}(\omega)} & & 0 \\ & \ddots & \\ 0 & & e^{\frac{1}{2}\mathrm{g}_D^{(t)}(\omega)} \end{pmatrix}. \tag{11.69}$$

Here, $\mathbf{g}^{(t)}(\omega) = \left(g_1^{(t)}(\omega), g_2^{(t)}(\omega), \ldots, g_D^{(t)}(\omega)\right)$ is a frequency-dependent vector of the logarithms of the eigenvalues of $\mathbf{M}^{(t)}(\omega)\mathbf{M}^{(t)}(\omega)^*$, which quantifies the overall MDL of an MDM system. The decomposition (11.68) is the same as that of (11.61) at each individual frequency and the diagonal matrix (11.69) is the similar to (11.62).

The MDL given by the SVD (11.68) is frequency-dependent in general. In the special case that there is no MD, such that $\sigma_{\mathrm{gd}} = \sqrt{K}\sigma_\tau$ is equal to zero, the MDL is independent of frequency as assumed in Section 11.4.3. Assuming non-zero σ_{gd}, the correlation of the MDL at two frequencies depends on the frequency separation. If the frequency separation is small, the phase factors for the uncoupled modes appearing in (11.17) are similar, leading to similar MDL values at the two frequencies. If the frequency separation is large, the values of $\mathbf{M}^{(t)}(\omega)$ at the two frequencies are independent, leading to independent MDL at the two frequencies.

Considering the simplest case of two modes, Figure 11.11 illustrates the frequency dependence of MDL in the regimes of small and large GD spread, quantified by σ_{gd},

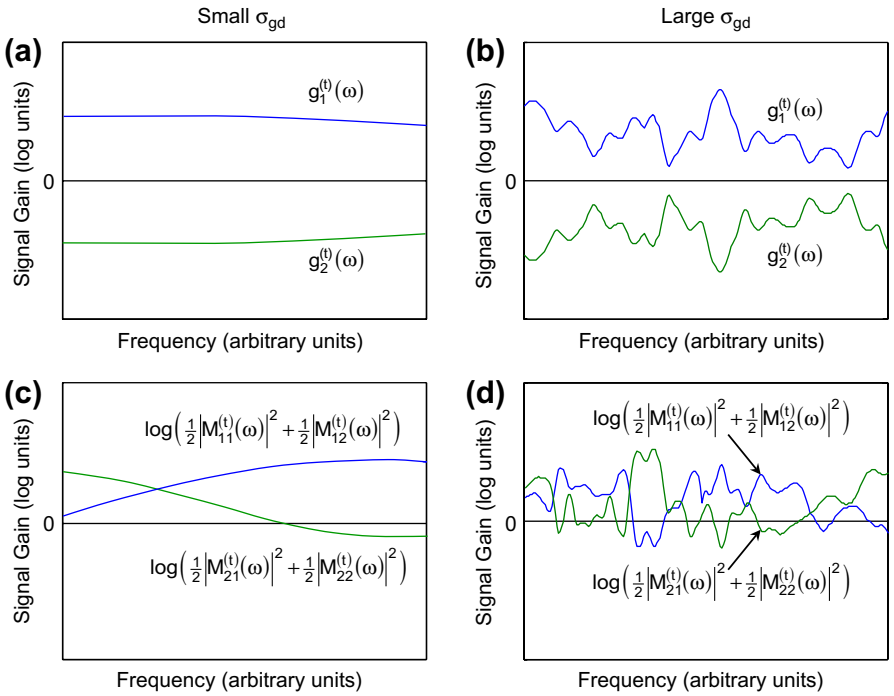

FIGURE 11.11 Frequency dependence of the MDL in a two-mode fiber for (a) small σ_{gd} and (b) large σ_{gd}, where σ_{gd} is the STD of GD. Output powers of signals launched into two orthogonal reference modes for (c) small σ_{gd} and (d) large σ_{gd}. Adapted from [30].

the STD of GD. Over the frequency range shown, the gains $g_1^{(t)}(\omega)$ and $g_2^{(t)}(\omega)$ (and thus the MDL) vary slowly for small σ_{gd} and rapidly for large σ_{gd}, as shown in Figure 11.11a and b, respectively. For signals launched into two orthogonal reference modes, the output powers (in logarithmic units) are

$$\log\left(\frac{1}{2}\left|M_{11}^{(t)}(\omega)\right|^2+\frac{1}{2}\left|M_{12}^{(t)}(\omega)\right|^2\right) \text{ and } \log\left(\frac{1}{2}\left|M_{21}^{(t)}(\omega)\right|^2+\frac{1}{2}\left|M_{22}^{(t)}(\omega)\right|^2\right),$$

respectively. Over frequency, these output powers vary slowly for small σ_{gd} and rapidly for large σ_{gd}, as shown in Figure 11.11c and d, respectively. For MDM signals spanning the frequency range shown, Figure 11.11c and d would correspond to regimes of low diversity order and moderate-to-high diversity order, respectively.

The correlation properties of MDL should depend on the normalized frequency separation $\vartheta = \Delta\omega\sigma_{gd}/2\pi$, where $\Delta\omega$ is the angular frequency separation. For small normalized frequency separation, $\vartheta \ll 1$, the MDLs at the two frequencies are identical, while for large normalized frequency separation, $\vartheta \gg 1$, the MDLs at the two

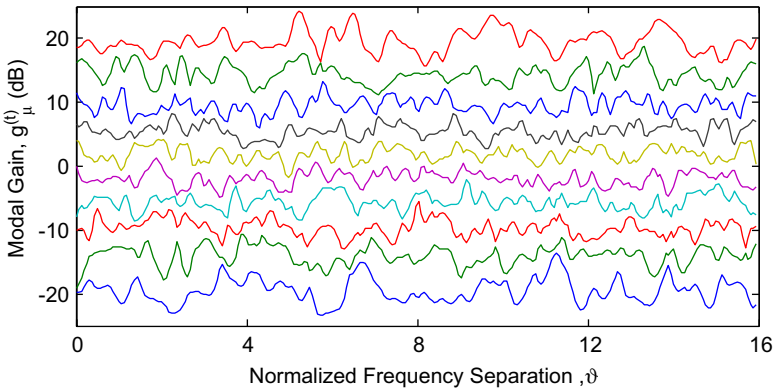

FIGURE 11.12 Modal gains, $g_\mu^{(t)}\mu = 1, \ldots, D$, as a function of normalized frequency separation for a fiber with $D = 10$ modes. Adapted from [30].

frequencies are independent. The coherence bandwidth of MDL should be of the same order as the reciprocal of the overall STD of GD, $1/\sigma_{\text{gd}}$; hence, the normalized coherence bandwidth should be of order unity.

Figure 11.12 shows simulations of the gain vector $\mathbf{g}^{(t)}(\omega)$ defined in (11.69) as a function of normalized frequency separation ϑ. The MMF has $D = 10$ modes and an accumulated MDL $\xi = 10\,\text{dB}$. The MMF comprises $K = 256$ statistically identical sections, the same as in Section 11.3.2. The gain vector in each section $\mathbf{g}^{(k)}$ is the same as that in Section 11.4.4. The GD vector $\boldsymbol{\tau}^{(k)}$ in each section is generated as a Gaussian random vector whose entries sum to zero, using the method described in the Appendix of [28]. Each curve in Figure 11.12 corresponds to one of the elements of the vector $\mathbf{g}^{(t)}(\omega)$ as a function of normalized frequency separation ϑ. The x-axis of Figure 11.12 is the normalized frequency separation with respect to the first frequency.

Figure 11.12 illustrates how the correlation of the MDL depends on frequency separation, similar to Figure 11.11a and b. The gain of each mode is a smooth, continuous curve, so each modal gain is highly correlated for small frequency separations. Conversely, each modal gain is uncorrelated for large frequency separations. Figure 11.12 also shows that the highest and lowest modal gains are subject to larger variations than the intermediate modal gains because the outer peaks of the probability density function exhibit a larger spread than the inner peaks.

Figure 11.13 shows the correlation coefficients of the elements of the modal gain vector $\mathbf{g}^{(t)}(\omega)$ as a function of normalized frequency separation. The simulation parameters are the same as in Figure 11.12, but the correlation coefficients are calculated with 23,000 realizations of modal gain curves, each similar to Figure 11.12. In Figure 11.13, the correlation coefficient is calculated for each gain coefficient after conversion to a decibel scale. In Figure 11.13, the 10 curves are observed to cluster into five pairs, which are for the gain coefficients $g_\mu^{(t)}$ and $g_{D-\mu+1}^{(t)}, \mu = 1, \ldots, 5$.

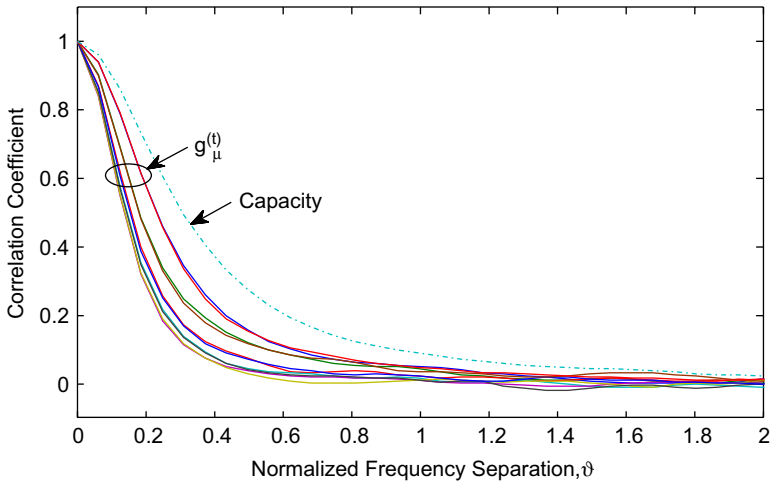

FIGURE 11.13 Correlation coefficients of modal gains, $g_\mu^{(t)}$ $\mu = 1, \ldots, D$, as a function of normalized frequency separation for a fiber with $D = 10$ modes. Also shown is the correlation coefficient of the average channel capacity, which is used in Section 11.6.2, assuming a SNR of 20 dB and assuming CSI is not available at the transmitter. Adapted from [30].

The correlation coefficients are observed to decrease with an increase of μ. Referring to Figure 11.12, the highest and lowest two curves ($\mu = 1, g_1^{(t)}, g_{10}^{(t)}$) exhibit the largest (and similar) correlation over frequency, while the middle two curves ($\mu = 5, g_5^{(t)}, g_6^{(t)}$) exhibit the smallest (and similar) correlation over frequency.

In Figure 11.13, it is difficult to uniquely define a single coherence bandwidth for all the modal gains, because they decay at different rates, and do not decay fully to zero at large normalized frequency separation (this may arise, at least in part, from numerical errors). Considering the highest and lowest gains with the largest correlations, the normalized one-sided coherence bandwidths are 0.25 or 0.67 for correlation coefficients of 50% or 10%, respectively. At 10% correlation coefficient, the normalized one-sided coherence bandwidth ranges from 0.32 to 0.67 for the different gain coefficients. At a normalized frequency separation of unity, the correlation coefficients range from 0 to 4.7% for the different gain coefficients.

Figures 11.11–11.13 illustrate the frequency dependence of MDL. For wide-band MDM systems, frequencies having large MDL may be statistically averaged with frequencies having small MDL, at a frequency separation beyond the coherence bandwidth. The system benefits of this frequency diversity are quantified in Section 11.6.2.

Because the MDL is frequency-dependent, the potential exists for the received noise power spectrum to become non-white over frequency. Using arguments similar to those in Section 11.4.5 above, it can be shown that as the number of noise sources K becomes large, the received noise becomes white over frequency.

11.5 DIRECT-DETECTION MODE-DIVISION MULTIPLEXING

In short-range links, MDM using direct detection may be possible. Signals may be intensity-modulated or modulated by differential phase-shift keying. The entire end-to-end system must be designed with great care to avoid crosstalk between multiplexed signals and signal distortion by MD, or else these effects must be avoided by adaptive optical signal processing. If signal distortion and crosstalk are allowed to occur, they cannot be compensated fully by linear electrical signal processing, because of nonlinearity caused by higher-order MD [30,82], as explained in Section 11.3.1.

Direct-detection MDM has been proposed numerous times since the early 1980s. Some of the first proposals were based on multiplexing signals at different angles [132] or at different radial offsets [7], and attempted to exploit the relatively weak coupling between different mode groups over short distances, especially in step-index fibers with high GD differences between different mode groups. More recent works have pursued various approaches to multiplexing signals into weakly coupled mode groups [8,9,14,15,133,134]. As of this writing, however, we are not aware of any direct-detection MDM systems that have achieved significantly higher throughput than conventional MMF links.

In systems where MDL is negligible or small, MDM using optical signal processing to exploit PMs [135] can avoid crosstalk and signal distortion from MD in principle. Figure 11.14 shows a conceptual link design, assuming a single wavelength channel. To explain the operating principle, assuming the signal spans a narrow band near angular frequency ω, we can represent signal propagation by $\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}\boldsymbol{\Lambda}^{(t)}(\omega)\,\mathbf{U}^{(t)*}$ (11.23). Recall that the columns of $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)}$ describe the input and output PMs, respectively. In Figure 11.14, a modal multiplexer at the transmitter maps input signals into different input PMs, implementing $\mathbf{U}^{(t)}$, while

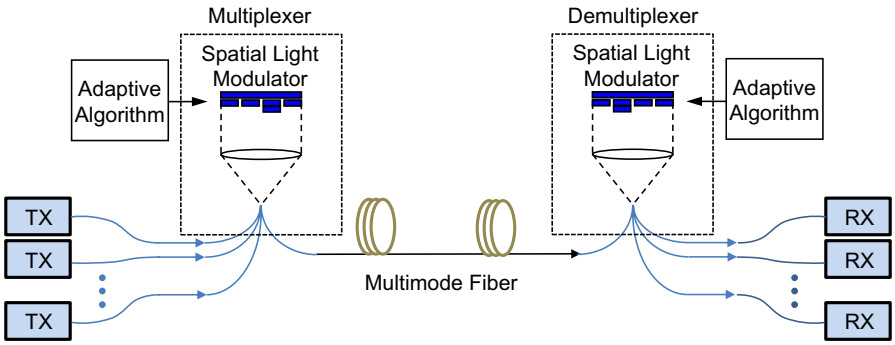

FIGURE 11.14 Schematic direct-detection mode-division multiplexed (MDM) system that multiplexes transmitted signals in input PMs, and demultiplexes received signals from output PMs. Adaptive optical components are used in both multiplexer and demultiplexer, with spatial light modulators shown as an example.

a modal demultiplexer at the receiver maps output signals from the corresponding output PMs, implementing $\mathbf{V}^{(t)*}$. Propagation of the multiplexed signals is described by $\mathbf{\Lambda}^{(t)}$, given by (11.24), which causes no crosstalk and no distortion, as explained in Section 11.3.1. By performing only memoryless optical signal processing, we have avoided signal distortion, obviating the need for signal processing that has memory. If the coherence bandwidth of the PMs (see Section 11.3.1) is sufficiently large, it may be possible to use the modal multiplexer and demultiplexer for multiple wavelength-division-multiplexed channels, similar to [136].

The modal multiplexer and demultiplexer in Figure 11.14 must be adjusted to track changes in the PMs caused by environmental perturbations. Spatial light modulators (SLMs) [13–15,133–137] or photonic integrated circuits [138–140] can be employed, but some algorithm is needed to determine the device settings. If the fiber propagation operator is known, the SVD $\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}\mathbf{\Lambda}^{(t)}(\omega)\,\mathbf{U}^{(t)*}$ given by (11.23) may be computed, and then the input and output SLMs can be set to implement $\mathbf{U}^{(t)}$ and $\mathbf{V}^{(t)*}$, respectively. In practice, it may be difficult to measure the propagation operator and compute the SVD, so it may be preferable to adjust the multiplexer and demultiplexer using an adaptive algorithm that optimizes their settings based on measurement of eye openings, signal-to-noise-plus-interference ratios, or some other metrics. Standard adaptive algorithms [141] are not readily applicable. An adaptive algorithm for setting the SLMs is described in [135]. Adaptation of the modal multiplexer at the transmitter can be problematic in long-haul systems, where the round-trip signal propagation delay can be tens of milliseconds, which is longer than the time scale on which fiber properties can change.

An important outstanding question concerns how to implement the adaptive multiplexer and demultiplexer so as to minimize loss and crosstalk. For multiplexing (or demultiplexing) D modes, one can use D separate SLM regions to track the input (or output) PMs, and use beamsplitters for signal combining (or splitting). A multiplexer or demultiplexer would have a loss of at least $1/D$, for a combined loss of at least $1/D^2$. Alternatively, these devices can be implemented using the principle of a multiplexed correlator [15] or photonic integrated circuit [138–140], but it is not clear how the resulting loss and crosstalk scale with the number of multiplexed modes D.

Higher-order MD, described in Section 11.3.1 above, poses a fundamental limitation to direct-detection MDM systems. Although PMs are independent of frequency to first order, their frequency dependence can become important for wideband signals or long-haul propagation [30,82]. When $\mathbf{U}^{(t)}$ or $\mathbf{V}^{(t)}$ become frequency-dependent, SLMs or similar frequency-independent devices cannot realize them, and it may be too difficult or costly to implement frequency-dependent adaptive optics. Higher-order MD can cause nonlinear signal distortion and crosstalk that cannot be compensated effectively by linear electrical signal processing after a direct-detection receiver.

MDL poses another fundamental limitation to direct-detection MDM systems. Even in short links without optical amplifiers, MDL can be caused by offset connectors. MDL can make the signal-to-noise ratios different for various multiplexed signals. When signals are multiplexed into orthogonal input PMs, MDL can cause the

received signals to become non-orthogonal, making it impossible to simultaneously avoid MD and crosstalk [142].

For $D = 2$, Figure 11.14 corresponds to polarization multiplexing in SMF by adaptively multiplexing into input PSPs and demultiplexing from output PSPs, which is fundamentally limited by higher-order PMD [143,144] and by PDL. To avoid the difficulties associated with an adaptive transmitter, it is far easier to multiplex signals into any pair of orthogonal polarization states, and use adaptive optical PMD compensation at the receiver prior to polarization demultiplexing (as a further simplification, PMD compensation may be replaced by simple polarization tracking, making the system subject to degradation by PMD [145–147]). By contrast, for $D > 2$, using a fixed orthogonal launch is probably not practical, because receiver-based adaptive optical compensation of MD becomes increasingly complicated as D increases.

11.6 COHERENT MODE-DIVISION MULTIPLEXING

As described in Section 11.5 above, in direct-detection MDM, unless the entire end-to-end system can be designed to achieve minimal crosstalk and distortion, it may be necessary to multiplex into input PMs at the transmitter and demultiplex from output PMs at the receiver using adaptive optics. Assuming frequency-independent multiplexers and demultiplexers, the usable signal bandwidth should not exceed the coherence bandwidth of the PMs, or higher-order MD will cause nonlinear distortion and crosstalk.

By contrast, in coherent MDM, the transmitter needs only multiplex signals into some orthogonal set of modes, and the receiver needs only to project the received signal onto some orthogonal set of modes [12,77] and perform dual-quadrature detection of each mode, providing sufficient statistics for signals to be separated using MIMO DSP [1–4,12,18,20]. In a coherent system, an arbitrarily large GD spread and arbitrary-order MD can be compensated with no penalty, provided laser phase noise is sufficiently small and the channel does not change too rapidly.

A coherent system may use either single- or multi-carrier modulation [1,2,18,148]. Multi-carrier modulation is typically implemented using OFDM, as we assume throughout this section. To compare these options briefly, single-carrier modulation is expected to incur less degradation from fiber nonlinearity, and typically requires a simpler transmitter but slightly more complex DSP at the receiver. OFDM is expected to incur more degradation from fiber nonlinearity and requires more complex DSP at the transmitter, but uses a simpler receiver architecture with lower computational complexity.

In this section, we address important issues governing the complexity and performance of coherent MDM systems, using the analyses of MD and MDL from Sections 11.3 and 11.4. In Section 11.6.1, we study how MDL can reduce average channel capacity. In Section 11.6.2, we show that strong mode coupling and MD lead to frequency diversity that can reduce outage probability. In Section 11.6.3, we discuss MIMO DSP methods for multi- or single-carrier MDM systems, and study how MD affects the implementation complexity of MIMO DSP.

11.6.1 Average channel capacity of narrowband systems

Long-haul MDM systems are assumed here to use inline optical amplification and coherent detection. The dominant noise at the receiver is assumed to arise from amplified spontaneous emission. The received noise power spectral density is assumed to be the same in each mode, as was assumed in wireless communications [26,27]. This assumption of spatially white noise was justified analytically and verified numerically in Section 11.4.5, but differs from the assumption made in [29]. The signal-to-noise ratio (SNR) ρ_t is defined as the received signal power (total over all D modes) divided by the received noise power (per mode). Throughout this subsection, the channel capacity at a single frequency is considered, assuming the signal has a very narrow bandwidth, as in [28,29].

Our definition of SNR follows the convention used in wireless MIMO systems [26] and is compatible with a conservative assumption that the total signal power in all D modes is constrained independent of D, while the noise power per mode is independent of D. Nevertheless, our results, if interpreted correctly, do not depend on this assumption in any way. For example, one might reason that the total signal power constraint scales in proportion to the number of modes D for a fixed time-averaged nonlinear phase shift, since D scales approximately in proportion to the core area from Section 11.2.1 and [47]. To accommodate such a constraint, one may scale the SNR as $\rho_t = D \cdot \rho_1$, where ρ_1 is the SNR per mode. Note, however, that both MD and CD may reduce nonlinear interactions between different modes [149–152].

In an MMF without MDL, such that all modes have equal gain, the power received in each mode is equal, and the channel capacity is equal to

$$C = D \log_2 \left(1 + \frac{\rho_t}{D}\right). \tag{11.70}$$

Throughout this chapter, channel capacity is expressed on a per-unit-bandwidth basis (i.e. in terms of spectral efficiency), and has units of b/s/Hz. When the SNR is much greater than the number of modes, the capacity (11.70) increases almost linearly with the number of modes. When the SNR is much less than the number of modes, the channel capacity is approximately proportional to SNR and independent of the number of modes. In the limit of an infinite number of modes, the capacity is given asymptotically by $C_\infty = \rho_t \log_2 e$, which is independent of the number of modes.

In MMF with MDL, channel state information (CSI) represents the information about a channel, described by a propagation operator $\mathbf{M}^{(t)}(\omega)$ (11.20), required by a transmitter to allocate transmit power optimally. CSI is defined more precisely in Section 11.6.1.2. In a coherent MDM system, the availability of CSI at the transmitter is an essential factor governing channel capacity. In the absence of CSI, an increase in MDL always leads to a decrease in capacity. In the extreme limit of MDL, the fiber supports propagation in only one mode. If CSI is available to the transmitter, only the surviving mode is used for transmission, and the channel capacity is $\log_2(1 + \rho_t)$, which is (11.70) with $D = 1$. If CSI is not available and the surviving mode is not known to the transmitter, the transmitter must allocate equal power to all

modes, and the channel capacity is $\log_2(1+\rho_t/D)$. As demonstrated below, channel capacity can be improved greatly by the availability of CSI. In some situations, when CSI is available, the channel capacity with MDL may exceed the capacity (11.70) without MDL.

Because MDL is a statistical phenomenon, as explained in Section 11.4, the channel capacity by itself is a random variable. In narrowband MDM systems, it is more relevant to study outage capacity, as in [28,29], than average capacity. As shown in Section 11.6.2, it is likely that practical MDM systems will need to operate in a wideband regime in which strong mode coupling and MD provide frequency diversity, such that the overall outage capacity approaches the average channel capacity, which is independent of bandwidth (when expressed on a per-unit-bandwidth basis). With this in mind, here we study only the average channel capacity of narrowband systems.

11.6.1.1 Average channel capacity without channel state information

When CSI is not available, the transmitter allocates equal power to each mode. Given the number of modes D and a realization of the gain vector $\mathbf{g}^{(t)}$ (11.15), the instantaneous channel capacity is

$$C=\sum_{i=1}^{D}\log_2\left[1+\frac{\chi}{D}\exp\left(\mathrm{g}_i^{(t)}\right)\right]. \tag{11.71}$$

Taking the average of (11.71), the average channel capacity is:

$$C=D\int_{-\infty}^{+\infty}\log_2\left[1+\frac{\chi}{D}\exp(\sigma_{\mathrm{mdl}}x)\right]p_D(x)dx, \tag{11.72}$$

where $p_D(x)$ is the probability density defined in (11.44) with unit variance, σ_{mdl} is the STD of overall MDL, and the constant χ is determined by the constraint:

$$\chi=\frac{\rho_t}{\int_{-\infty}^{+\infty}\exp(\sigma_{\mathrm{mdl}}x)\;p_D(x)dx}. \tag{11.73}$$

If the noise power per mode is normalized to unity, the constant χ is the total transmitted power and χ/D in (11.72) is the transmitted power per mode. Because of Proposition I of Section 11.4.1, the probability density $p_D(x)$ is the same as (11.44). However, scaling is required, as the probability density given by (11.44) has a variance of $\frac{1}{2}(D-D^{-1})$ instead of unity.

From Section 11.4, the MDL (measured in units of decibels or log power gain) has zero mean. Measured on a linear scale, the overall power loss/gain of a K-section MMF, summed over all D modes, is equal to $\cosh^K\sigma_g$. This overall gain gives a Lyapunov exponent, defined in [120–122], of $\log\cosh\sigma_g$. The factor χ, given by (11.73), normalizes the overall gain to unity, as measured on a linear scale. For a system with MDL, the channel gains are not constant, but are random variables. The SNR ρ_t is the mean (statistical average) SNR, and the factor χ given by (11.73) can be interpreted as the ratio of the mean SNR to the mean gain of the channel.

The product $\chi \exp(\sigma_{\text{mdl}} x)$ in (11.72) can be interpreted as the SNR of a channel realization with normalized gain x.

Analytical expressions are available for the constant χ, given by (11.73), but are too complicated for practical calculations. There is no known analytical expression for the capacity (11.72) except in the limit of very large D. Numerical integration of (11.72) may be employed to find the channel capacity. For fibers with a large number of modes, MDL is log-semicircle distributed, and the constant χ becomes

$$\chi = \frac{\rho_t \sigma_{\text{mdl}}}{2 I_1(\sigma_{\text{mdl}})},$$

where $I_1(\cdot)$ is the modified Bessel function of the first kind. If D is far larger than $\chi \cdot \exp(2\sigma_{\text{mdl}})$, where $2\sigma_{\text{mdl}}$ is the upper limit of the semicircle distribution, the channel capacity approaches C_∞. In this limit, each mode is allocated a very small power, and the overall capacity is proportional to the total power.

Figure 11.15 shows the average capacity with MDL but without CSI, as a function of the accumulated MDL $\xi = \sqrt{K}\sigma_g$, for fibers with various numbers of modes D. As CSI is unavailable, equal power is allocated to each mode. The mean SNR is $\rho_t = 10$ dB. In Figure 11.15, the theoretical channel capacity is calculated by using (11.57) to

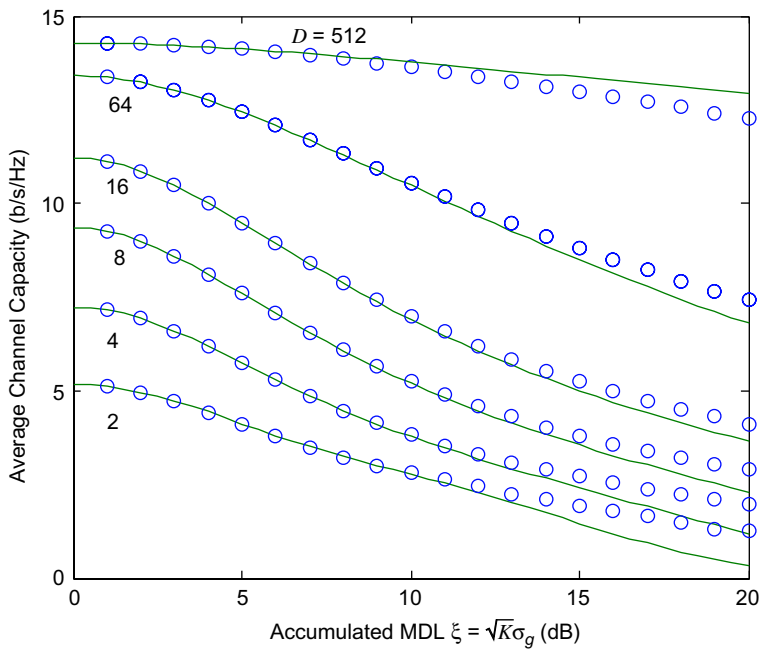

FIGURE 11.15 Average channel capacity for MDM in an MMF as a function of accumulated MDL ξ for fibers with various numbers of modes D, assuming CSI is not available. Equal power is allocated to each mode, and the SNR is ρ_t = 10 dB. Theoretical results are shown as curves and simulated results are shown as circles. Adapted from [28].

find the STD of overall MDL σ_{mdl}, and the probability distributions (11.45) (scaled to unit variance) are used in both (11.72) and (11.73) to compute the average channel capacity. The simulations use channel matrices generated by the same methods used to generate Figure 11.9.

In Figure 11.15, in the absence of CSI, average channel capacity always decreases with increasing ξ, particularly in fibers with smaller numbers of modes D. For values of ξ up to 10 dB, the average channel capacities from theory and simulation match very well. For a two-mode fiber, theoretical and simulated capacities match within 5% up to $\xi = 11$ dB. For a 512-mode fiber, theoretical and simulated capacities match within 5% up to $\xi = 19$ dB. The discrepancy between theory and simulation decreases with an increase in the number of modes.

The total launched power, equivalent to χ given by (11.73), is assumed not to change with the random realization of MDL. Satisfying this assumption requires that there be no amplifier saturation caused by MDL in each stage. For example, consider a two-mode case with unit total input power, unit noise level (in the absence of saturation), unit SNR, and gains of 0.5 and 1.5 (measured in linear units). In the absence of saturation, the output powers may vary from 0.5 to 1.5, depending on the alignment of the input signal to the eigenmodes of MDL (i.e. the modes of minimum and maximum gain). When saturation occurs, however, the maximum output powers may be limited to values less than the nominal value of 1.5, while noise levels decrease proportionally to maintain the mean SNR of unity. The model given here does not take account of these effects. If there are many optical amplifier stages or the system has a wide bandwidth, similar to the arguments of Section 11.4.5 based on the law of large numbers, a gain reduction in one stage (or at one frequency) should be compensated well by a gain enhancement in other stages (or at other frequencies). However, saturation effects for MDM systems are not well studied at this writing.

Due to statistical averaging, the outage capacity approaches the average channel capacity for a wideband system with strong mode coupling, as shown in Section 11.6.2.

11.6.1.2 Average channel capacity with channel state information

Ideally, in the operation of a MIMO system, the receiver estimates the overall channel matrix $\mathbf{M}^{(t)}$, computes the channel decomposition $\mathbf{M}^{(t)} = \mathbf{V}^{(t)}\mathbf{\Lambda}^{(t)}\mathbf{U}^{(t)*}$ (11.61), sends a description of the precoding matrix $\mathbf{U}^{(t)}$ and the gain vector $\mathbf{g}^{(t)}$ to the transmitter, and uses the received beam-forming vector $\mathbf{V}^{(t)}$ in decoding received signals [26]. The matrix $\mathbf{U}^{(t)}$ and the gain vector represent CSI that is to be fed back from the receiver to the transmitter. Transmit power and information bits are allocated to spatial channels based on the gain vector $\mathbf{g}^{(t)}$.

In a long-haul system, the feedback process described above may become impractical if the MMF changes on a time scale shorter than or comparable to the round-trip propagation delay, which can be tens of milliseconds. When feedback becomes impossible, space-time codes [153–155] or error-correction codes across the spatial channels can provide diversity. In these cases, all spatial channels are allocated equal power, yielding the capacity computed in Section 11.6.1.1. As

described below in Section 11.6.3.3, receivers for systems with or without CSI have the same architecture, apart from the means to estimate CSI to feed back to the transmitter.

When CSI is available at the transmitter, transmit power may be allocated to spatial channels in an optimal way. If the noise power per mode is normalized to unity and the overall gain vector is $\mathbf{g}^{(t)}$ known, the optimal transmit powers are given by

$$\left[\eta - e^{-g_\mu^{(t)}}\right]^+, \ \mu = 1, \ldots, D,$$

where $[\ ]^+$ denotes limiting to non-negative values, i.e. $[x]^+ = \max(0, x)$. The constant η is chosen to satisfy the total power constraint:

$$\sum_{\mu=1}^{D} \left[\eta - e^{-g_\mu^{(t)}}\right]^+ = \chi, \tag{11.74}$$

with χ given by (11.73). The average channel capacity is

$$C = \left\langle \sum_{\mu=1}^{D} \log_2 \left(1 + \left[\eta e^{g_\mu^{(t)}} - 1\right]^+\right)\right\rangle, \tag{11.75}$$

which is an expectation over random realizations of the MDL.

A brute-force method to compute the channel capacity (11.75) involves generating many random realizations of the overall matrix (11.20) and computing their singular values to evaluate the average capacity [29]. Each realization of (11.20) is the product of $3K$ matrices. Obviously, the total number of matrices can be reduced to $2K+1$ by combining $\mathbf{U}^{(k+1)}$ and $\mathbf{V}^{(k)}, k = 1, \ldots, K-1$, into single random unitary matrices.

Assuming values of MDL sufficiently small to be of practical interest, Propositions I and II are valid (see Section 11.4.1). In this regime, a more efficient method of computing (11.75) is based on Proposition I, namely, that the joint probability density for the vector $g^{(t)}$ is given by the eigenvalue distribution of a zero-trace Gaussian unitary ensemble. Hence, a zero-trace random Gaussian Hermitian matrix may be generated using the method described in the Appendix of [28], and the eigenvalues of the random matrix can be used to compute the capacity (11.75). Also, instead of generating new matrices for different values of ξ, Proposition II can be exploited. A single matrix may be generated, and the eigenvalues can be scaled using the approximation for σ_{mdl} given by (11.57).

Figure 11.16 shows the average capacity for a system with MDL and with CSI, as a function of the accumulated MDL $\xi = \sqrt{K}\sigma_g$, for fibers with various numbers of modes D. The SNR is $\rho_t = 10\,\text{dB}$, which represents a statistical average. The theoretical curves are obtained using 100,000 zero-trace random Gaussian-Hermitian matrices. Simulated results are obtained using brute-force generation of channel matrices by the same methods used for Figures 11.8 and 11.9.

In Figure 11.16, we see that for fibers with $D = 2$, 4, or 8 modes, even with CSI, the channel capacity decreases with increasing MDL. With $D = 16$ modes,

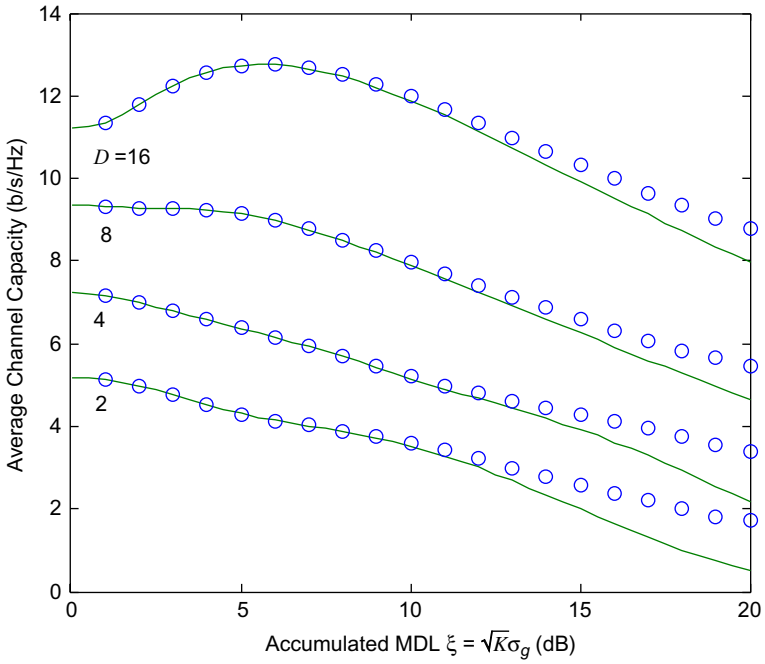

FIGURE 11.16 Average channel capacity for MDM in an MMF as a function of accumulated MDL ξ for fibers with various numbers of modes *D*. CSI is available, and transmit power is allocated optimally to each mode, and the SNR is ρ_t = 10 dB. Theoretical results are shown as curves and simulated results are shown as circles. Adapted from [28].

at small MDL, the capacity increases with increasing MDL, although it eventually decreases with a further increase in MDL. More generally, MDL can increase capacity when the SNR measured in linear units is small relative to the number of modes D. This may seem counter-intuitive, but can be made plausible by the following example for $D=2$ modes. Assume that the total transmitted power is unity, the noise power per mode is unity, and the mean SNR is $\rho_t = 1$. Hence, the mean channel gain is 0 dB, corresponding to unity in linear units. Without MDL, using (11.70), the channel capacity is $2\log_2(1+0.5) = 1.17$ b/s/Hz. Now suppose that MDL is present, and that in a particular channel realization, the modal gains are -3.01 dB and $+1.76$ dB, which correspond to 0.5 and 1.5 in linear units (the mean of these two values is unity, as in the absence of MDL). When CSI is not available, the transmitter allocates equal power of 0.5 to each mode. The capacity is $\log_2(1+0.5\cdot 0.5)+\log_2(1+1.5\cdot 0.5) = 1.13$ b/s/Hz, slightly smaller than that without MDL. When CSI is available, the transmitter may allocate all the power to the stronger mode. The capacity becomes $\log_2(1+1.5\cdot 1) = 1.32$ b/s/Hz, slightly larger than that without MDL.

In Figure 11.16, for a two-mode fiber, theoretical and simulated average capacities agree within 5% up to $\xi = 11\,\text{dB}$. For a 16-mode fiber, theoretical and simulated average capacities agree within 5% up to $\xi = 17$ dB. The discrepancy decreases with an increasing number of modes. While the more exact MDL model (11.58) gives slightly more accurate results, values of accumulated MDL $\xi > 10\,\text{dB}$ are likely too large for practical systems.

Although mathematically correct, the results of Figure 11.16 are difficult to reconcile with the saturation effects of optical amplifiers. The constraint (11.74) is nearly linear in χ, independent of η. Amplifier saturation effects may lead to interdependence between χ and η. Even for wideband systems and systems with many amplifier stages, it is not expected that amplifier saturation effects will fully average out between frequencies or stages.

In the remainder of this chapter, unless noted otherwise, we assume that CSI is not available.

11.6.2 Wideband systems and frequency diversity

In Section 11.4.6, Figures 11.11–11.13 demonstrate that MDL is frequency-dependent, and is strongly correlated only over a finite coherence bandwidth. If MDM signals occupy a bandwidth far larger than the coherence bandwidth, the outage channel capacity should approach the ensemble average channel capacity computed in Section 11.6.1 due to statistical averaging. For typical values of MDL and SNR, the coherence bandwidth of the capacity is found to be approximately equal to $1/\sigma_{\text{gd}}$, the reciprocal of the STD of GD. The difference between the average capacity and the outage capacity is found to decrease with the square-root of a diversity order that is given approximately by the ratio of the signal bandwidth to the coherence bandwidth of the channel capacity.

As explained earlier, the channel capacity for narrowband MDM systems is a random variable. Figure 11.17 shows the simulated distribution of the channel capacity of a narrowband system with $D = 10$ modes at an SNR $\rho_t = 20\,\text{dB}$, corresponding to an SNR per mode $\rho_1 = 10\,\text{dB}$, assuming CSI is not available at the transmitter. The system parameters are the same as those in Figure 11.13 of Section 11.4.6. The distribution in Figure 11.17 is constructed using about 5,900,000 channel capacity values. In Figure 11.17, the average channel capacity, near the peak of the distribution, is about 17.2 b/s/Hz, while the capacity for 10^{-3} outage probability is about 14.3 b/s/Hz. In a wideband MDM system, at any single frequency, the channel capacity has the same distribution as that in Figure 11.17.

Figure 11.13 of Section 11.4.6 shows the correlation coefficients of the modal gains versus normalized frequency separation, illustrating how the gains at nearby frequencies are highly correlated. Figure 11.13 also shows the correlation coefficient of the channel capacity. The capacity is computed as in Figure 11.17, assuming an SNR $\rho_t = 20\,\text{dB}$, no CSI available at the transmitter, and equal power allocated to all modes. The normalized one-sided coherence bandwidths of the channel capacity are 0.31 and 0.92 for correlation coefficients of 50% and 10%,

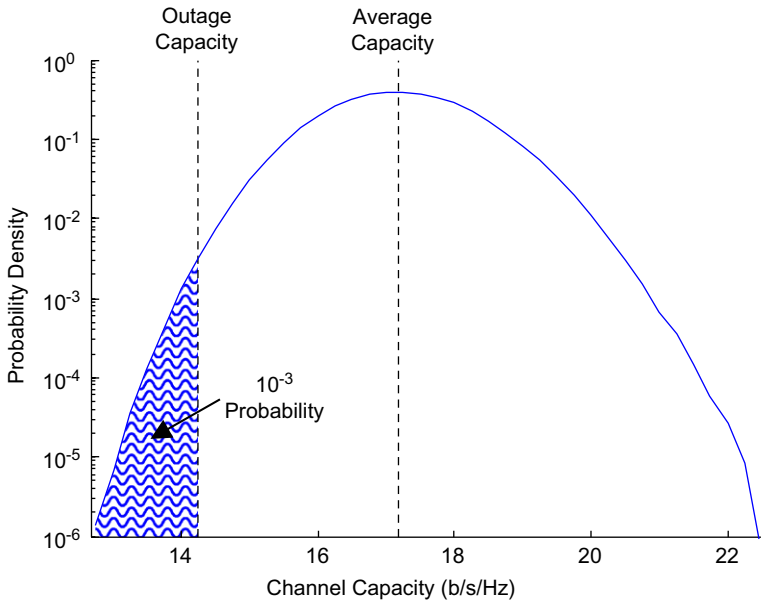

FIGURE 11.17 Distribution of channel capacity at a single frequency for an MDM system with $D = 10$ modes and an SNR of 20 dB, assuming CSI is not available. The average capacity and outage capacity at 10^{-3} outage probability are indicated. Adapted from [30].

respectively. A normalized frequency separation of unity gives a correlation coefficient of 8.8%.

If MDM signals occupy a bandwidth much greater than the coherence bandwidth of the capacity, statistical averaging over frequency should cause the outage channel capacity to approach the average channel capacity. Figure 11.18 shows the outage capacity as a function of SNR, for signals occupying different bandwidths. All parameters are as in Figure 11.13 and Figure 11.17 and the outage probability is 10^{-3}. A normalized signal bandwidth is defined as $b = B_{\text{sig}}\sigma_{\text{gd}}$, where B_{sig} is the signal bandwidth (measured in Hz). In Figure 11.18, signals occupy normalized bandwidths from $b = 0$ (a single frequency) to $b = 8$. The average capacity is also shown for comparison. Figure 11.18 shows that as the normalized bandwidth increases, the outage capacity approaches the average capacity.

Statistical averaging over frequency is a consequence of the law of large numbers. For example, consider two OFDM subchannels at frequencies whose separation far exceeds the coherence bandwidth of the capacity, e.g. at two frequencies well separated in Figure 11.12b. We assume that the subchannels have capacities C_1 and C_2; these are independent random variables following a common distribution (e.g. that in Figure 11.17). A channel comprising the two subchannels has an overall capacity $\frac{1}{2}(C_1 + C_2)$, as the capacity is computed on a per-unit-frequency basis. If

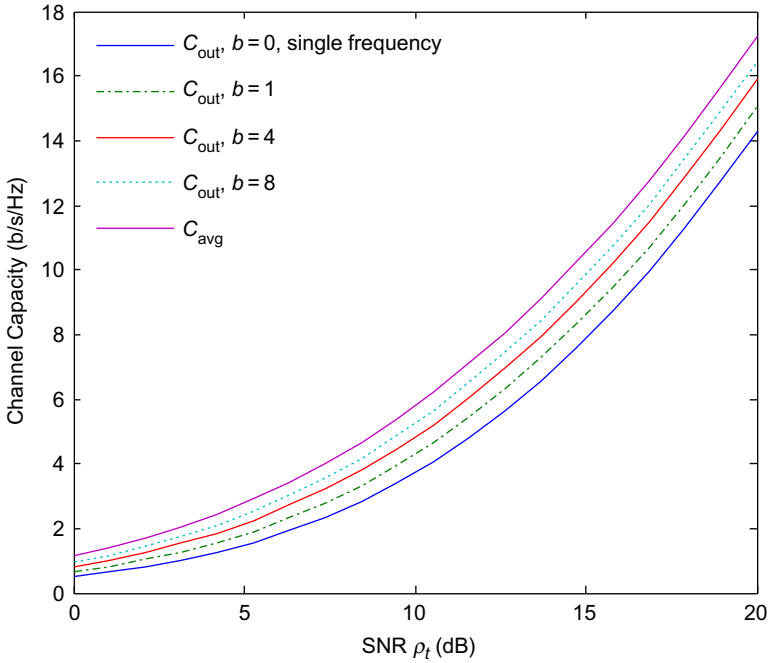

FIGURE 11.18 Outage channel capacity at 10^{-3} outage probability versus SNR for various normalized bandwidths, for an MDM system with $D = 10$ modes, assuming CSI is not available. The average channel capacity is shown for comparison. Adapted from [30].

each subchannel capacity has variance σ_C^2, the overall channel capacity has variance $\frac{1}{2}\sigma_C^2$, i.e. it is reduced by a factor of two. The channel comprising two independent subchannels has a diversity order of two. More generally, given a channel spanning a finite bandwidth B_{sig}, we define the diversity order in terms of a reduction of the variance of capacity: a diversity order equal to F_D corresponds to a reduction of the variance of capacity from σ_C^2 to σ_C^2/F_D. With this definition, the diversity order may be any real number not smaller than unity.

It would be useful to be able to estimate the diversity order F_D directly from the statistics of the frequency-dependent gain vector $\mathbf{g}^{(t)}(\omega)$, rather than having to compute the frequency-dependent capacity and characterize its statistics. The method of principal component analysis [156] may be used to calculate the diversity order directly from the frequency correlation coefficients of the channel capacity. Principal component analysis is very similar to the discrete Karhunen-Loève transform [157]. The conversion from correlation coefficients to diversity order is explained in detail in [30].

Numerical simulations of MDM systems similar to those in Figure 11.18 have been performed, with number of modes $D = 10$ and normalized bandwidth b ranging

from 0 to 16. Regardless of the normalized bandwidth b, the distribution of the channel capacity is found to retain approximately the same shape as that of Figure 11.17, but the variance is reduced from σ_C^2 to approximately σ_C^2/F_D. The mean channel capacity does not change with diversity order. The outage capacity as a function diversity order is found to follow

$$C_{\text{out},F_D} \approx C_{\text{avg}} - \frac{1}{\sqrt{F_D}}\left(C_{\text{avg}} - C_{\text{out},1}\right), \tag{11.76}$$

where $C_{\text{out},1}$ is the single-frequency outage capacity following the distribution in Figure 11.17. The relationship (11.76) is found to be independent of the outage probability, provided the outage capacities $C_{\text{out},1}$ and C_{out,F_D} refer to the same outage probability. The relationship (11.76) is found to be valid even if the shape of the distribution of capacity deviates from that shown in Figure 11.17. If the capacity distribution is assumed to be Gaussian, as in [29], the relationship between outage capacity and diversity order F_D at any particular outage probability can be computed analytically. However, when plotted on a semi-logarithmic scale, as in Figure 11.17, the capacity distribution is observed to deviate noticeably from a Gaussian distribution, i.e. it is slightly asymmetric. This non-Gaussianity is consistently observed at all SNRs and all diversity orders.

Figure 11.19 shows the outage capacity reduction ratio, defined as

$$\frac{C_{\text{avg}} - C_{\text{out},F_D}}{C_{\text{avg}} - C_{\text{out},1}}, \tag{11.77}$$

as a function of $1/\sqrt{F_D}$, where F_D is the diversity order computed using principal component analysis from the correlation coefficients of the channel capacity shown in Figure 11.13. The simulation parameters used for Figure 11.19 are the same as those of Figure 11.13 and Figure 11.18, i.e. the MDM system uses $D = 10$ modes, and CSI is not available at the transmitter. Values of the diversity order F_D are only computed for SNR $\rho_t = 20\,\text{dB}$, as in Figure 11.13, but Figure 11.19 shows values of (11.77) computed at SNR $\rho_t = 10$ and $20\,\text{dB}$, illustrating that the diversity order F_D is valid over a wide range of SNR values. The correlation coefficients in Figure 11.13 are subject to numerical error, as they never go to zero even for large frequency separations. To limit numerical error, diversity orders in Figure 11.19 are computed only using values of the correlation coefficients from Figure 11.13 that are larger than 1%.

Based on (11.76), the outage capacity reduction ratio (11.77) should approximately equal $1/\sqrt{F_D}$, and the plots in Figure 11.19 should be straight lines with unit slope. In Figure 11.19, the best-fit slope is found to be 1.06. Figure 11.19 clearly shows that $C_{\text{avg}} - C_{\text{out},F_D}$ approaches zero as the diversity order F_D increases. The observed dependence of the difference between average and outage capacities on $1/\sqrt{F_D}$ is a direct consequence of the law of large numbers. The outage and average capacities converge slowly with an increase in diversity order F_D. The diversity order F_D must be four to decrease the capacity difference to half that without diversity, and must be 100 to decrease the difference to 10% of that without diversity.

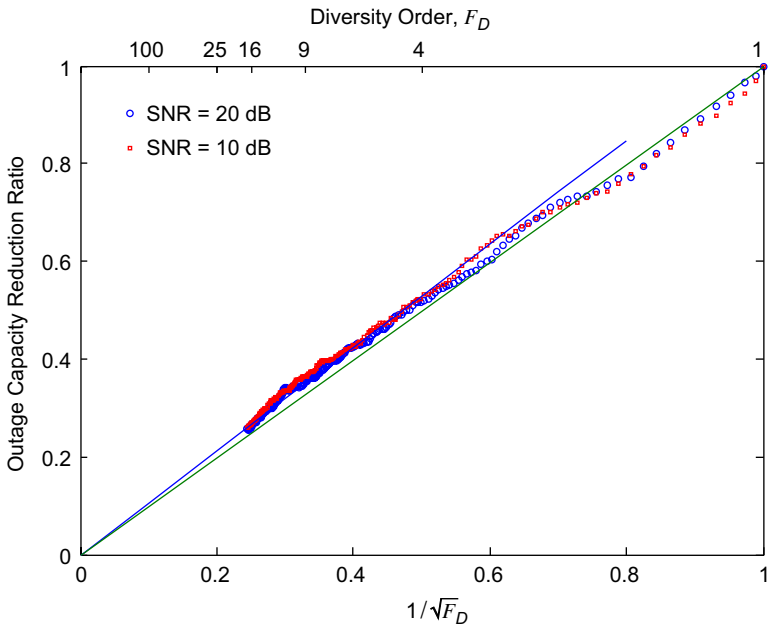

FIGURE 11.19 The outage capacity reduction ratio, given by (11.77), as a function of $1/\sqrt{F_D}$, where F_D is the diversity order. The lines have slopes of 1 (theoretical) and 1.06 (best-fit), respectively. The MDM system uses $D = 10$ modes, and CSI is not available. Adapted from [30]. (For interpretation of the color in this figure, the reader is referred to the web version of this book.)

The ratio of outage to average capacities, $C_{\text{out},1}/C_{\text{avg}}$, decreases with an increase of MDL or a reduction of SNR. In the limit of a very high SNR and an MDL smaller than the SNR, the channel capacity without CSI (11.71) is approximately equal to $D\log_2\chi/D + \log_2 e\sum_{i=1}^{D} g_i^{(t)} = D\log_2\chi/D$, which is independent of the frequency-dependent gain vector $\mathbf{g}^{(t)}(\omega)$. The average SNR ρ_t needs to be large enough that even the weakest mode has sufficiently high SNR. At high SNR, the channel capacity is independent of $\mathbf{g}^{(t)}(\omega)$ even for a system with CSI.

The combined matrix model (11.20) shows that the GD statistics do not depend on the mode-averaged CD. In the case of spatial-mode-dependent CD, the higher-order frequency dependence of the GD statistics may be modified slightly. The general approach presented here should remain valid. The STD of GD becomes a frequency-dependent $\sigma_{\text{gd}}(\omega)$ to include the effect of spatial-mode-dependent CD. The normalized frequency separation may be modified to $\vartheta = (2\pi)^{-1}\int_0^{\Delta\omega}\sigma_{\text{gd}}(\omega_0+s)ds$, where ω_0 is a reference frequency and $\Delta\omega$ is the frequency difference. The normalized bandwidth b is always the normalized frequency separation between the lowest and highest frequencies.

Numerical simulation of Figure 11.19 uses a maximum normalized bandwidth of $b = 16$. The diversity order F_D is approximately equal to the normalized bandwidth

for systems with $b \geqslant 4$. As shown in Section 11.6.3.3 below, receiver DSP complexity scales in proportion to the normalized bandwidth b. Complexity constraints may permit systems to have normalized bandwidths b up to the order of 100 or larger. Assuming practical MDL values, the outage channel capacity of a wideband system with $b = 100$ should approach the average channel capacity of a narrowband system. With such high-order frequency diversity, average channel capacity is a good metric for characterizing system performance.

11.6.3 Signal processing for mode-division-multiplexing

In coherent MDM systems, DSP is used at the receiver, and possibly at the transmitter, to perform various functions [1–4,18,20]. In the transmitter, DSP may be used for precoding, and is used to perform modulation in OFDM systems. In the receiver, DSP is used for timing synchronization, frequency and phase synchronization, MIMO channel estimation and equalization, and other functions.

Figure 11.20 schematically shows key transmitter and receiver functions in an OFDM system, while Figure 11.21 shows corresponding functions in a single-carrier system, assuming an FDE-based receiver. These figures show only those functions that are relevant to the discussion in this section, and omit other important functions. For simplicity, only one channel in a wavelength-division-multiplexed system is shown. "Tx" includes digital-to-analog conversion and electrical-to-optical conversion and "Rx" includes combination of the signal and a local oscillator in an optical hybrid, optical-to-electrical conversion, and analog-to-digital conversion.

In Figures 11.20 and 11.21, the transmitter may optionally include a precoder. If CSI is available, precoding may be used with power and bit allocation (often

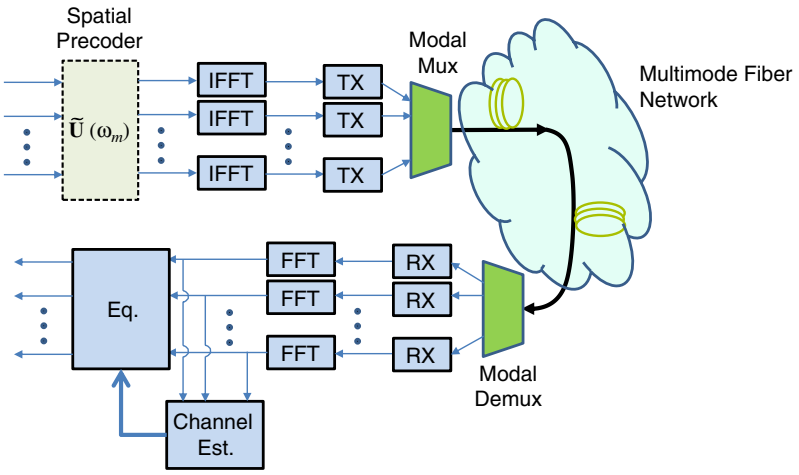

FIGURE 11.20 Major functions of OFDM transmitter and receiver in an MDM system.

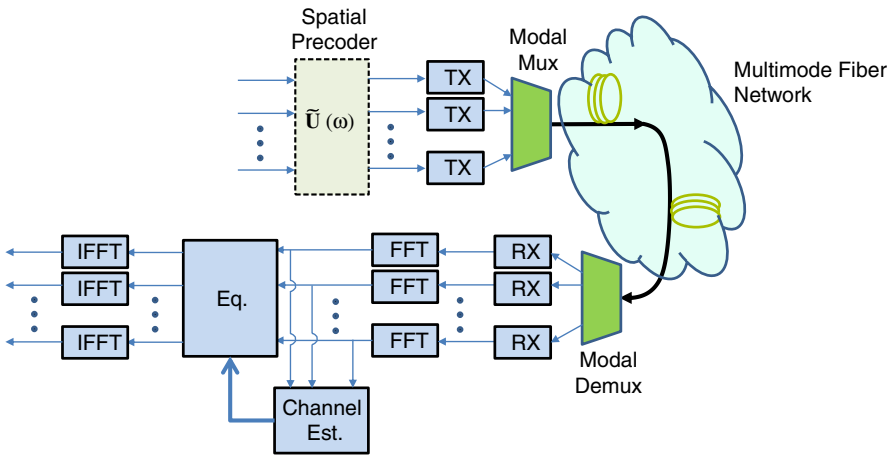

FIGURE 11.21 Major functions of single-carrier transmitter and receiver in an MDM system, assuming the receiver uses frequency-domain equalization (FDE).

known as "bit loading") to enhance performance, as discussed in Section 11.6.1.2. In this case, assuming the overall propagation matrix is decomposed using SVD to $\mathbf{M}^{(t)}(\omega) = \mathbf{V}^{(t)}(\omega)\mathbf{\Lambda}^{(t)}(\omega)\mathbf{U}^{(t)}(\omega)^*$ (11.68), the ideal precoder is $\mathbf{U}^{(t)}(\omega)$. Alternatively, precoding may be used to implement space-time coding [153–155] for spatial diversity. In general, the actual precoder employed is denoted by $\tilde{\mathbf{U}}^{(t)}(\omega)$. If no precoder is used, we can make the assignment $\tilde{\mathbf{U}}^{(t)}(\omega) = \mathbf{I}$. Even if CSI is available, the actual precoder $\tilde{\mathbf{U}}^{(t)}(\omega)$ may not be equal to $\mathbf{U}^{(t)}(\omega)$ because of estimation error in the receiver due to additive noise, quantization noise or round-off error, or because of delay in the CSI feedback and estimation process. CSI may be estimated for a group of adjacent frequencies, resulting in estimation error at some frequencies within the group.

We assume that the overall propagation matrix $\mathbf{M}^{(t)}(\omega)$ (11.20) includes the modal multiplexer and demultiplexer. Including the precoder and mode-averaged CD, and ignoring mode-averaged gain, the overall channel response becomes:

$$\mathbf{H}(\omega) = \mathbf{M}^{(t)}(\omega)\tilde{\mathbf{U}}^{(t)}(\omega)\exp\left(-j\frac{1}{2}\bar{\beta}_2(\omega)^2 L_t\right), \tag{11.78}$$

where $L_t = \sum_k L^{(k)}$ is the total system length.

11.6.3.1 Multi-carrier signals

Figure 11.20 shows key functions in an OFDM transmitter and receiver. In the transmitter, for the mth subchannel at angular frequency ω_m, the precoder is $\tilde{\mathbf{U}}^{(t)}(\omega_m)$. In the receiver, after downconverting all D spatial channels and performing FFTs of each one, each subchannel may be compensated using a memoryless one-tap

equalizer [158–160]. Assuming a minimum mean-square error (MMSE) equalizer [59, Section 15.1–2], the equalizer response in the mth subchannel is:

$$\left[\mathbf{H}^*(\omega_m)\mathbf{H}(\omega_m) + \mathbf{S}_n(\omega_m)\right]^{-1}\mathbf{H}^*(\omega_m), \tag{11.79}$$

where $\mathbf{S}_n(\omega_m)$ is the power spectrum of the sampled noise. From Section 11.4.5, $\mathbf{S}_n(\omega_m)$ is likely to be approximately proportional to $\mathbf{I}$, but may exhibit variations among subchannels. In this section, $\mathbf{S}_n(\omega_m) \propto \mathbf{I}$ is assumed for each subchannel.

Frequency-dependent MDL causes the channel response $\mathbf{H}(\omega)$ to become non-unitary over space and frequency. At the output of the equalizer (11.79), the SNR may vary widely across different spatial and frequency subchannels. Strong error-correction coding across subchannels exploits frequency diversity and provides coding gain. Many modern strong codes are decoded using soft information, typically the log-likelihood ratio [161,162], which is the logarithm of the ratio of the probabilities that a bit is equal to 1 or 0. If $\mathbf{x}(\omega_m)$ is the D-dimensional transmitted vector for the mth subchannel, the received signal is $\mathbf{y}(\omega_m) = \mathbf{H}(\omega_m)\mathbf{x}(\omega_m) + \mathbf{n}(\omega_m)$, where $\mathbf{n}(\omega_m)$ is the additive noise in the subchannel. Assume that $\mathbf{b}_l, l = 1, \ldots, M$, are all possible transmitted signal vectors, where $M = 4^D$ for quadrature phase-shift keying (QPSK) signals in each mode, for example. The log-likelihood ratio is given by

$$\log \frac{\sum_{l,\text{bit}=1} \exp\left(-\frac{\|\mathbf{y}(\omega_m)-\mathbf{H}(\omega_m)\mathbf{b}_l\|^2}{\mathbf{S}_n(\omega_m)}\right)}{\sum_{l,\text{bit}=0} \exp\left(-\frac{\|\mathbf{y}(\omega_m)-\mathbf{H}(\omega_m)\mathbf{b}_l\|^2}{\mathbf{S}_n(\omega_m)}\right)}, \tag{11.80}$$

where the numerator sums over all possible transmit vectors in which a given bit is equal to 1 and the denominator sums over those in which it is equal to 0. The log-likelihood ratio in the form (11.80) is difficult to obtain. Using the log-max approximation, only the maximum terms in both numerator and denominator are used. Including only the maximum terms and ignoring some common factors, the log-likelihood ratio becomes

$$2\left[\max_{l,\text{bit}=1} \text{Re}\left\{\frac{\mathbf{b}_l^*\mathbf{H}^*(\omega_m)\mathbf{y}(\omega_m)}{\mathbf{S}_n(\omega_m)}\right\} - \max_{l,\text{bit}=0} \text{Re}\left\{\frac{\mathbf{b}_l^*\mathbf{H}^*(\omega_m)\mathbf{y}(\omega_m)}{\mathbf{S}_n(\omega_m)}\right\}\right]. \tag{11.81}$$

In a typical system, the approximation (11.81) is very close to the exact expression (11.80). To evaluate (11.81), we only need to calculate

$$2\mathbf{S}_n^{-1}(\omega_m)\mathbf{H}^*(\omega_m)\mathbf{y}(\omega_m) \tag{11.82}$$

in each OFDM subchannel and compare it with the possible $\mathbf{b}_l$ to obtain the log-likelihood ratio. In an OFDM receiver, the equalizer just needs to multiply the received signal vector by the channel matrix (11.82), which is a very simple operation.

The possible transmitted $\mathbf{b}_l$ depends on CSI and the bit loading procedure. Because the same CSI should be available at both transmitter and receiver, the bit loading table may be calculated by the same procedure at both transmitter and receiver. However, the CSI may not be fed back to the transmitter correctly. More robust schemes require that

the bit loading table or some indicator be transmitted as side information in the OFDM signals. In addition to the overhead of this side information, the round-trip delay of the feedback may be too long for a time-varying channel. For this reason, OFDM wireless systems typically do not use bit loading. In practice, if error correction coding is used across frequency and spatial subchannels, as in coded OFDM [163,164], the extra complexity of bit loading is usually not justified. The complexity of an OFDM receiver remains nearly the same regardless of whether CSI is fed back to the transmitter.

11.6.3.2 Single-carrier signals

Figure 11.21 shows key functions in a single-carrier transmitter and receiver, assuming an FDE-based receiver. As shown in [18,20], the DSP complexity for TDE is typically far higher.

A single-carrier transmitter, like its multi-carrier counterpart, may use a precoder $\widetilde{\mathbf{U}}^{(t)}(\omega)$. The single-carrier precoder $\widetilde{\mathbf{U}}^{(t)}(\omega)$ is frequency-dependent, and its implementation is not yet widely known. The precoder $\widetilde{\mathbf{U}}^{(t)}(\omega)$ should be designed to have minimal GD spread over the signal bandwidth, allowing it to be implemented in the time domain. If the GD spread of $\widetilde{\mathbf{U}}^{(t)}(\omega)$ is too large, a frequency-domain implementation is required, similar to the FDE in a receiver. Theoretically, any precoder described by

$$\widetilde{\mathbf{U}}^{(t)}(\omega)\text{diag}\left[e^{-j\omega\tau_1^{(p)}}, e^{-j\omega\tau_2^{(p)}}, \ldots, e^{-j\omega\tau_D^{(p)}}\right] \tag{11.83}$$

for any delay vector $\boldsymbol{\tau}^{(p)} = \left(\tau_1^{(p)}, \tau_2^{(p)}, \ldots, \tau_D^{(p)}\right)$ will yield equivalent system performance. In designing the precoder, the vector $\boldsymbol{\tau}^{(p)}$ should be chosen to minimize the GD spread of (11.83).

In the receiver, all D spatial channels are optically downconverted and sampled, broken into blocks, and converted to the frequency domain using FFTs. Assuming MMSE equalization, the blocks are equalized by

$$\left[\mathbf{H}^*(\omega)\mathbf{H}(\omega) + \mathbf{S}_n(\omega)\right]^{-1}\mathbf{H}^*(\omega) \tag{11.84}$$

and converted back to time domain using IFFTs. In (11.84), $\mathbf{H}(\omega)$ is given by (11.78) and $\mathbf{S}_n(\omega)$ is the power spectrum of the sampled noise. Throughput may be maximized using overlap-add or overlap-save convolution [165,166], while receiver complexity is minimized by inserting a cyclic prefix [167] or unique word [168–170] at the transmitter. Transmission of the cyclic prefix or unique word reduces throughput and wastes energy, but the unique word can be used as pilots to aid in timing, frequency, or phase synchronization. Overlap-add or overlap-save convolution is assumed in the complexity analysis in Section 11.6.3.3.

When MDL causes the channel response $\mathbf{H}(\omega)$ to become non-unitary over space and frequency, each spatial output of the single-carrier equalizer (11.84) represents a linear combination of different frequency components, providing the averaging required for frequency diversity. In single-carrier systems, strong error-correction coding, typically decoded with soft information, such as log-likelihood ratios

analogous to (11.80), provides coding gain, but is not essential to frequency diversity, unlike OFDM systems [171].

When CSI is not available at the transmitter, in the presence of significant frequency-dependent MDL, coded OFDM generally outperforms single-carrier modulation with the linear equalizer (11.84). Performance equivalent to OFDM can be achieved using single-carrier modulation with a decision-feedback equalizer (DFE) or a Tomlinson-Harashima precoder (THP) [172,173]. Unfortunately, both DFE and THP are extremely difficult to implement at very high speeds. Of course, if MDL is small and the overall channel (11.78) is nearly unitary, DFE or THP are not required for single-carrier modulation to have the same performance as OFDM.

Note that the above comparison between single-carrier and OFDM signals is in the context of linear propagation with MDL. In the presence of fiber nonlinearity, single-carrier signals are expected to tolerate a higher launched power and achieve a higher SNR [148], probably outweighing any performance advantage of OFDM in the linear regime.

11.6.3.3 Signal processing complexity

A primary goal of MDM is reducing energy consumption per bit [174]. Hence, the computational complexity and resulting energy consumption of DSP are of critical importance. In this section, we discuss the complexity of FDE in single- or multi-carrier systems. TDE in single-carrier systems is not considered here, as its complexity is typically far higher than FDE [18,20].

The received signal is sampled at a rate $r_{\text{os}}R_s$, where r_{os} is an oversampling ratio, usually between 1.5 and 2, and R_s represents the symbol rate for single-carrier modulation or the occupied bandwidth (an effective symbol rate) for OFDM. In an optimized architecture, only a fraction $1/r_{\text{os}}$ of the frequency-domain subchannels require equalization. In the OFDM system of Figure 11.20, only a fraction $1/r_{\text{os}}$ of the subchannels are modulated with nonzero amplitudes. In the single-carrier FDE receiver of Figure 11.21, it is only necessary to compute the IFFT over the middle subcarriers, which comprise a fraction $1/r_{\text{os}}$ of the total number of subchannels. In the FDE, the IFFT also downsamples the signal by a factor $1/r_{\text{os}}$ to the symbol rate R_s.

Receiver computational complexity depends strongly on the GD spread of the equivalent channel $\mathbf{H}(\omega)$, given by (11.75). The GD spread arises mainly from CD and MD. The duration of the impulse response of CD, measured in samples, is given by [18,175,176]

$$N_{\text{cd}} = \left\lceil 2\pi \left|\bar{\beta}_2\right| L_t \left(r_{\text{os}}R_{\text{s}}\right)^2 \right\rceil, \tag{11.85}$$

where $\bar{\beta}_2$ is the mode-averaged CD coefficient described in Section 11.2.3.3, L_t is the total system length, and $\lceil x \rceil$ denotes the smallest integer larger than x. Assuming the strong-coupling regime, the duration of the impulse response of MD, measured in samples, is

$$N_{\text{md}} = \left\lceil \sigma_{\text{gd}} u_D(p) r_{\text{os}} R_{\text{s}} \right\rceil, \tag{11.86}$$

where σ_{gd} is the STD of coupled GD defined in (11.45) and $u_D(p)$ is a function of the number of modes D and a GD spread outage probability p [18], which is the probability that a random realization of the impulse response has duration longer than (11.86). The value of $u_D(p)$ is given by the complementary cumulative GD spread distribution, as shown in Figure 11.7. For suitably small values of p ($p = 10^{-6}$ is considered here), $u_D(p) = 4.94$ and 4.60 for $D = 6$ and 12, respectively, and decreases toward 4.00 as $D \to \infty$, since the support of the semicircle distribution approaches $4\sigma_{\text{gd}}$. In the duration (11.86), the product $\sigma_{\text{gd}} R_s$ approximately equals the normalized bandwidth b defined in Section 11.6.2 (equality is exact for OFDM signals).

Instead of choosing the FFT size to minimize computational complexity as in [18], we choose $N_{\text{FFT}} = 2^{\lceil 3+\log_2(N_{\text{cd}}+N_{\text{md}})\rceil}$, where the number 3 (corresponding to a factor of 8) is chosen as a compromise to reduce complexity without requiring an excessively large value of FFT size N_{FFT}. In an OFDM system, the IFFT in the transmitter has the same length as the FFT in the receiver. In a single-carrier receiver using FDE, the FFT size can be a factor r_{os} larger than the IFFT size. For simplicity, we consider each FFT (or IFFT) to require approximately $\frac{1}{2} N_{\text{FFT}} \lceil \log_2 N_{\text{FFT}} \rceil$ complex multiplications. In practice, the required number of complex multiplications depends on the FFT architecture, but the operations required per sample are always proportional to $\log_2 N_{\text{FFT}}$.

The OFDM receiver of Figure 11.20 may compensate MD and CD together using D complex multiplications, according to (11.82). Here, we assume that MD and CD are compensated separately using $D + 1$ complex multiplications. Assuming a sufficient cyclic prefix, and considering both transmitter and receiver, the required number of complex multiplications per two-dimensional symbol for OFDM is

$$\text{CM}_{\text{OFDM}} = 1 + D + r_{\text{os}} \log_2 N_{\text{FFT}}. \tag{11.87}$$

In the single-carrier receiver of Figure 11.21, the total number of complex multiplications per two-dimensional symbol is

$$\text{CM}_{\text{FDE}} = \frac{(1+D)N_{\text{FFT}} + \frac{1}{2} r_{\text{os}} N_{\text{FFT}} \log_2 N_{\text{FFT}} + \frac{1}{2} N_{\text{FFT}} \lceil \log_2(N_{\text{FFT}}/r_{\text{os}}) \rceil}{N_{\text{FFT}} - N_{\text{cd}} - N_{\text{md}} + 1}, \tag{11.88}$$

where $D(N_{\text{FFT}} - N_{\text{cd}} - N_{\text{md}} + 1)/r_{\text{os}}$ is the number of symbols processed per FFT block, assuming overlap-add or overlap-save convolution. Single-carrier FDE requires fewer FFT operations than OFDM signals, but achieves lower computational efficiency due to overlap convolution.

Figure 11.22 shows the number of complex multiplications per two-dimensional symbol as calculated by (11.88). The total system length is $L_t = 2000\,\text{km}$ and the mode-averaged CD is $\bar{\beta}_2 = -22.5\,\text{ps}^2/\text{km}$ (corresponding to 17.5 ps/nm/km at the wavelength of 1550 nm). Assuming fibers with numerical aperture NA = 0.15 and a graded-index depressed-cladding index profile, using values from [18], the STD of the uncoupled MD is

$$\left[\frac{1}{D} \sum\nolimits_{\mu} \left(\beta_{1,\mu} - \bar{\beta}_1 \right)^2 \right]^{1/2} = 277 \text{ and } 383 \text{ ps/km}$$

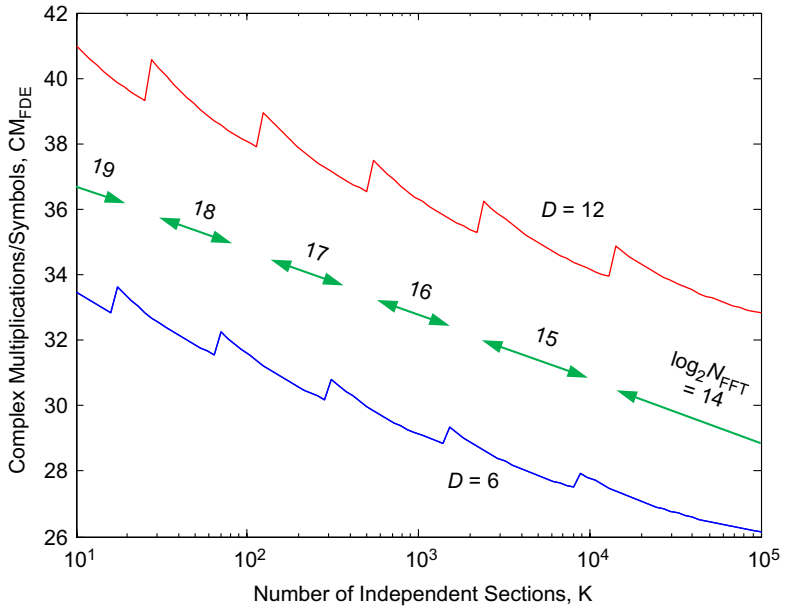

FIGURE 11.22 Number of complex multiplications per two-dimensional symbol as a function of the number of independent fiber sections for $D = 6$ and 12 modes for single-carrier modulation using FDE. Also indicated are the approximate ranges for different FFT block lengths.

for $D = 6$ and 12, respectively, corresponding to the value of σ_τ per kilometer in (11.45). The symbol rate is $R_s = 32$ GHz. An oversampling ratio $r_{os} = 1.5$ is assumed. In practice, although $r_{os} = 1.5$ cannot be achieved exactly using power-of-two FFT sizes, a value close to $r_{os} = 1.5$ such that N_{FFT}/r_{os} is an integer can be achieved using an appropriate FFT architecture. Figure 11.22 shows the number of complex multiplications per symbol as a function of the number of independent sections K. As K ranges from 10 to 10^5, the length of each section ranges from 200 km to 2 m. Over this wide range, the computational complexity is relatively insensitive to K. As K increases and the GD spread N_{md} decreases, the FFT size N_{FFT} decreases, but this reduces both the numerator and denominator of (11.88).

In Figure 11.22, the computational complexity exhibits small jumps when the FFT size N_{FFT} changes. The values of N_{FFT}, which are indicated in Figure 11.22, can be very large, which may have adverse implications for system practicality. In OFDM systems, large values of N_{FFT} can cause sensitivity to carrier frequency offset and phase noise-induced inter-carrier interference [171] and may require phase noise cancelation [177,178]. In single-carrier systems, equalization-enhanced phase noise increases with the GD spread from MD [179], similar to the effect of CD in SMF systems [180–182]. The convergence times of adaptive equalizers [18], and DSP

memory requirements, increase at least linearly with N_{FFT}, and both can become prohibitive for large values of N_{FFT}.

For the particular type of MMF considered in Figure 11.22, the above issues are mitigated if the number of independent sections K is larger than 500–1000, corresponding to a mode-coupling correlation length shorter than 2–4 km. Within this range, the MD duration N_{md} (11.86) is about 6 to 9 times CD duration N_{cd} (11.85), N_{FFT} is less than 2^{16}, and compensating MD and the CD requires less than 30 or 37 complex multiplications per symbol for $D = 6$ and 12 modes, respectively. The corresponding normalized bandwidth in this range is always larger than $b = 500$. For comparison, compensating CD only requires $N_{\mathrm{FFT}} = 2^{13}$, and about 19 complex multiplications per symbol.

As mentioned in Section 11.2.2 above, a short correlation length can be achieved by intentionally introducing perturbations in the MMF that enhance mode coupling, analogous to the spinning used to reduce DGD in SMF with PMD [48,71–74]. It would be helpful to employ MMFs having lower values of uncoupled GD spread than those considered here. As of this writing, however, these have been demonstrated only for $D = 6$ modes [1,183,184].

The overall complexity for MIMO DSP in MDM is expected to scale at least in proportion to $D \cdot N_{\mathrm{FFT}}$. Given the high symbol rates employed, careful attention to DSP architecture optimization and application-specific integrated circuit implementation is required. In particular, FFT architectures must be chosen carefully, considering the available options. These include: radix-2 FFTs based on the Cooley-Tukey algorithm [185], which are often the most straightforward; split-radix FFTs [186,187], which minimize the total number of operations (multiplications and additions); Winograd FFTs [188], which minimize the number of multiplications; and prime-factor FFTs [189,190], which can implement FFTs of flexible size while avoiding twiddle factor multiplications. Also, buffer sizes and read/write bandwidths must be considered carefully, including those for performing FFTs [191,192] and those for storing channel matrix coefficients for LLR computations (11.82) or FDE (11.84). At each frequency in (11.82) or (11.84), a total of D^2 complex-valued channel coefficients must be stored.

11.7 CONCLUSION

Fields propagating in a multimode fiber with mode coupling are strongly correlated over distances shorter than a correlation length, and are weakly correlated over longer distances. In a transmission system, mode coupling may be classified as weak or strong, depending on whether the total system length is comparable to or much longer than the correlation length. In the weak-coupling regime, MD and MDL accumulate linearly with system length, while in the strong-coupling regime they accumulate with the square-root of system length. For a given system length, both MD and MDL are minimized by strong mode coupling.

Strong mode coupling may be essential for the practicality of long-haul MDM systems, which employ inline optical amplifiers and use coherent detection and

MIMO DSP to compensate MD and to separate multiplexed signals. A system's end-to-end GD spread governs the MIMO DSP complexity. Achieving sufficiently low GD spread may require intentionally perturbing the MMF to ensure strong mode coupling, analogous to the spinning used to reduce the GD spread in SMF with PMD. MDL arising from inline amplifiers can cause random signal fluctuations analogous to multipath fading in wireless systems. This MDL can reduce average channel capacity and may cause outage in narrowband systems. Strong mode coupling can minimize MDL and maximize average capacity. In combination with sufficient MD, strong coupling leads to frequency diversity, which reduces outage probability in wideband systems, allowing outage capacity to approach average capacity.

The operator describing strongly coupled MD is equivalent to a zero-trace Gaussian unitary ensemble, and the statistics of the GDs are the same as the statistics of the eigenvalues of the ensemble. The STD of coupled GD scales as the square-root of system length. The probability density of the GDs can be calculated analytically for few-mode fibers, and approaches a Wigner semicircle distribution for many-mode fibers. The statistics of the coupled GD spread, which governs DSP complexity, can be computed using the Fredholm determinant or approximated using the Tracy-Widom distribution. Strongly coupled MDL, when described on a logarithmic scale in the low-MDL regime of practical interest, follows the same statistics as strongly coupled GDs. Accumulated MDL scales with the square-root of system length. The overall MDL is determined from the accumulated MDL by a nonlinear relationship.

Acknowledgments

The authors are grateful for the contributions of S.Ö. Arık to Section 11.6.3.3. The research of J.M. Kahn was supported, in part, by National Science Foundation Grant No. ECCS-1 101 905 and Corning, Inc.

References

[1] R. Ryf, S. Randel, A.H. Gnauck, C. Bolle, A. Sierra, S. Mumtaz, M. Esmaeelpour, E.C. Burrows, R.-J. Essiambre, P.J. Winzer, D.W. Peckham, A.H. McCurdy, R. Lingle, Mode-division multiplexing over 96 km of few-mode fiber using coherent 6×6 MIMO processing, J. Lightwave Technol. 30 (4) (2012) 521–531.

[2] A. Al Amin, A. Li, S. Chen, X. Chen, G. Gao, W. Shieh, Dual-LP_{11} mode 4×4 MIMO-OFDM transmission over a two-mode fiber, Opt. Express 19 (17) (2011) 16672–16678.

[3] S. Randel, R. Ryf, A. Sierra, P. Winzer, A.H. Gnauck, C. Bolle, R.-J. Essiambre, D.W. Peckham, A. McCurdy, R. Lingle, 6×56-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 6×6 MIMO equalization, Opt. Express 19 (17) (2011) 16697–16707.

[4] C. Koebele, M. Salsi, D. Sperti, P. Tran, P. Brindel, H. Margoyan, S. Bigo, A. Boutin, F. Verluise, P. Sillard, M. Bigot-Astruc, L. Provost, F. Cerou, G. Charlet, Two mode transmission at 2×100 Gb/s, over 40 km-long prototype few-mode fiber, using LCOS based mode multiplexer and demultiplexer, Opt. Express 19 (17) (2011) 16593–16600.

[5] B. Zhu, T.F. Taunay, M. Fishteyn, X. Liu, S. Chandrasekhar, M.F. Yan, J.M. Fini, E.M. Monberg, F.V. Dimarcello, 112-Tb/s Space-division multiplexed DWDM transmission with 14-b/s/Hz aggregate spectral efficiency over a 76.8-km seven-core fiber, Opt. Express 19 (17) (2011) 16665–16671.
[6] J. Sakaguchi, B.J. Puttnam, W. Klaus, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, K. Imamura, H. Inaba, K. Mukasa, R. Sugizaki, T. Kobayashi, M. Watanabe, 19-Core fiber transmission of 19×100×172-Gb/s SDM-WDM-PDM-QPSK signals at 305Tb/s, in: OFC '12, Paper PDP5C.
[7] S. Berdagué, P. Facq, Mode division multiplexing in optical fibers, Appl. Opt. 21 (11) (1982) 1950–1955.
[8] S. Murshid, B. Grossman, P. Narakorn, Spatial domain multiplexing: A new dimension in fiber optic multiplexing, Opt. Laser Technol. 40 (2008) 1030–1036.
[9] C.P. Tsekrekos, A.M.J. Koonen, Mode-selective spatial filtering for increasing robustness in mode group diversity multiplexing link, Opt. Lett. 32 (9) (2007) 1041–1043.
[10] K.-P. Ho, J.M. Kahn, Statistics of group delays in multimode fiber with strong mode coupling, J. Lightwave Technol. 29 (21) (2011) 3119–3128.
[11] R. Olshansky, Mode-coupling effects in graded-index optical fibers, Appl. Opt. 14 (4) (1975) 935–945.
[12] R. Ryf, M.A. Mestre, A.H. Gnauck, S. Randel, C. Schmidt, R.-J. Essiambre, P.J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, D.W. Peckham, A. McCurdy, R. Lingle, Low-loss mode coupler for mode-multiplexed transmission in few-mode fiber, in: OFC '12, Paper PDP5B.
[13] X. Shen, J.M. Kahn, M.A. Horowitz, Compensation for multimode fiber dispersion by adaptive optics, Opt. Lett. 30 (20) (2005) 2985–2987.
[14] H.S. Chen, H.P.A. van den Boom, A.M.J. Koonen, 30-Gb/s 3×3 optical mode group-division-multiplexing system with optimized joint detection, IEEE Photon. Technol. Lett. 23 (18) (2011) 1283–1285.
[15] J. Carpenter, T. Wilkinson, All optical mode-multiplexing using holography and multimode fiber couplers, J. Lightwave Technol. 30 (12) (2012) 1978–1984.
[16] A.F. Garito, J. Wang, R. Gao, Effects of random perturbations in plastic optical fibers, Science 281 (5379) (1998) 962–967.
[17] M.B. Shemirani, W. Mao, R.A. Panicker, J.M. Kahn, Principal modes in graded-index multimode fiber in presence of spatial- and polarization-mode coupling, J. Lightwave Technol. 27 (10) (2009) 1248–1261.
[18] S.Ö. Arik, D. Askarov, J.M. Kahn, Effect of mode coupling on signal processing complexity in mode-division multiplexing, J. Lightwave Technol. 31 (3) (2013) 423–431.
[19] C. Antonelli, A. Mecozzi, M. Shtaif, P.J. Winzer, Stokes-space analysis of modal dispersion in fibers with multiple mode transmission, Opt. Express 20 (11) (2012) 11718–11733.
[20] B. Inan, B. Spinnler, F. Ferreira, D. van de Borne, A. Lobato, S. Adhikari, V.A.J.M. Slieffer, M. Kuschnerov, N. Hanik, S.L. Jansen, DSP requirements for mode-division multiplexed receivers, Opt. Express 20 (9) (2012) 10859–10869.
[21] R. Olshansky, D.A. Nolan, Mode-dependent attenuation of optical fibers; Excess loss, Appl. Opt. 15 (4) (1976) 1045–1047.
[22] A.R. Mickelson, M. Eriksrud, Mode-dependent attenuation in optical fibers, J. Opt. Soc. Am. 73 (10) (1983) 1282–1290.

[23] N. Bai, E. Ip, Y.-K. Huang, E. Mateo, F. Yaman, M.-J. Li, S. Bickham, S. Ten, J. Liñares, C. Montero, V. Moreno, X. Prieto, V. Tse, K.M. Chung, A.P.T. Lau, H.-Y. Tam, C. Lu, Y. Luo, G.-D. Peng, G. Li, T. Wang, Mode-division multiplexed transmission with inline few-mode fiber amplifier, Opt. Express 20 (3) (2012) 2668–2680.
[24] Y. Jung, S. Alam, Z. Li, A. Dhar, D. Giles, I.P. Giles, J.K. Sahu, F. Poletti, L. Grüner-Nielsen, D.J. Richardson, First demonstration and detailed characterization of a multimode amplifier for space division multiplexed transmission systems, Opt. Express 19 (26) (2011) B952–B957.
[25] D. Askarov, J.M. Kahn, Design of multi-mode Erbium-doped fiber amplifiers for low mode-dependent gain, in: IEEE Summer Topical on Spatial Multiplexing '12.
[26] D. Tse, P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, 2005.
[27] A.M. Tulino, S. Verdú, Random Matrix Theory and Wireless Communications, Now, 2004.
[28] K.-P. Ho, J.M. Kahn, Mode-dependent loss and gain: Statistics and effect on mode-division multiplexing, Opt. Express 19 (17) (2011) 16612–16635.
[29] P.J. Winzer, G.J. Foschini, MIMO capacities and outage probabilities in spatially multiplexed optical transport systems, Opt. Express 19 (17) (2011) 16680–16696.
[30] K.-P. Ho, J.M. Kahn, Frequency diversity in mode-division multiplexing systems, J. Lightwave Technol. 29 (24) (2011) 3719–3726.
[31] D. Marcuse, Theory of Dielectric Optical Waveguide, second ed., Academic Press, 1991.
[32] L. Raddatz, I.H. White, D.G. Cunningham, M.C. Nowell, An experimental and theoretical study of the offset launch technique for the enhancement of the bandwidth of multimode fiber links, J. Lightwave Technol. 16 (3) (1998) 324–331.
[33] P. Pepeljugoski, S.E. Golowich, A.J. Ritger, P. Kolsar, A. Risteski, Modeling and simulation of next-generation multimode fiber links, J. Lightwave Technol. 21 (5) (2003) 1242–1254.
[34] K.M. Patel, S.E. Ralph, Enhanced multimode fiber link performance using a spatially resolved receiver, IEEE Photon. Technol. Lett. 14 (3) (2002) 393–395.
[35] K. Balemarthy, A. Polley, S.E. Ralph, Electronic equalization of multikilometer 10-Gb/s multimode fiber links: mode-coupling effects, J. Lightwave Technol. 24 (12) (2006) 4885–4894.
[36] J.P. Gordon, H. Kogelnik, PMD fundamentals: polarization-mode dispersion in optical fibers, Proc. Natl. Acad. Sci. 97 (9) (2000) 4541–4550.
[37] H. Kogelnik, R.M. Jopson, L.E. Nelson, Polarization-mode dispersion, in: I. Kaminow, T. Li (Eds.), Optical Fiber Telecommunications IVB, Academic Press, 2002.
[38] C.D. Poole, R.E. Wagner, Phenomenological approach to polarization dispersion in long single-mode fibers, Electron. Lett. 22 (19) (1986) 1029–1030.
[39] R. Noé, D. Sandel, M. Yoshida-Dierolf, S. Hinz, V. Mirvoda, A. Schöpflin, C. Glingener, E. Gottwald, C. Scheerer, G. Fischer, T. Weyrauch, W. Haase, Polarization mode dispersion compensation at 10, 20, and 40 Gb/s with various optical equalizers, J. Lightwave Technol. 17 (9) (1999) 1602–1616.
[40] H. Sunnerud, C. Xie, M. Karlsson, R. Samuelsson, P.A. Andrekson, A comparison between different PMD compensation techniques, J. Lightwave Technol. 20 (3) (2002) 368–378.
[41] S. Fan, J.M. Kahn, Principal modes in multimode waveguides, Opt. Lett. 20 (2) (2006) 135–137.

[42] Y. Kokubun, M. Koshiba, Noval multi-core fiber for mode division multiplexing: proposal and design principle, IEICE Electron. Express 6 (8) (2009) 522–528.
[43] A.W. Snyder, Coupled-mode theory for optical fibers, J. Opt. Soc. Am. 62 (11) (1972) 1267–1277.
[44] E. Snitzer, Cylindrical dielectric waveguide modes, J. Opt. Soc. Am. 51 (5) (1961) 491–498.
[45] K.C. Kao, G.A. Hockham, Dielectric-fibre surface waveguides for optical frequencies, Proc. IEE 113 (7) (1966) 1151–1158.
[46] D. Gloge, Weakly guiding fibers, Appl. Opt. 10 (10) (1971) 2252–2258.
[47] D. Gloge, E.A.J. Marcatilli, Multimode theory of graded-core fibers, Bell Syst. Tech. J. 52 (9) (1973) 1563–1578.
[48] M.-J. Li, D.A. Nolan, Optical transmission fiber design evolution, J. Lightwave Technol. 26 (9) (2008) 1079–1092.
[49] A.W. Snyder, J.D. Love, Optical Waveguide Theory, Chapman & Hall, 1983.
[50] G. Keiser, Optical Fiber Communications, third ed., McGraw-Hill, 2000.
[51] D. Gloge, Dispersion in weakly guiding fibers, Appl. Opt. 10 (11) (1971) 2442–2445.
[52] C. Lin, H. Kogelnik, L.G. Cohen, Optical-pulse equalization of low-dispersion transmission in single-mode fibers in the 1.3–1.7-m spectral region, Opt. Lett. 5 (11) (1980) 476–478.
[53] ISO/IEC 11801, Information technology—Generic cabling for customer premises, 2002.
[54] M. Webster, L. Raddatz, I.H. White, D.G. Cunningham, A statistical analysis of conditioned launch for gigabit Ethernet links using multimode fiber, J. Lightwave Technol. 17 (9) (1999) 1532–1541.
[55] Y. Koike, T. Ishigure, E. Nihei, High-bandwidth graded-index polymer optical fiber, J. Lightwave Technol. 13 (7) (1995) 1475–1489.
[56] Y. Koike, S. Takahashi, Plastic optical fibers: technologies and communication links, in: I.P. Kaminow, T. Li, A.E. Willner (Eds.), Optical Fiber Telecommunications VA, Elsevier Academic, 2008 (Chapter 16).
[57] A. Polley, S.E. Ralph, Mode coupling in plastic optical fiber enables 40-Gb/s performance, IEEE Photon. Technol. Lett. 19 (16) (2007) 1254–1256.
[58] R.F. Shi, C. Koeppen, G. Jiang, J. Wang, A.F. Garito, Origin of high bandwidth performance of graded-index plastic fibers, Appl. Phys. Lett. 72 (25) (1999) 3625–3627.
[59] J.G. Proakis, M. Salehi, Digital Communications, fifth ed., McGraw-Hill, 2008.
[60] T.S. Rapport, Wireless Communications: Principals & Practice, Prentice Hall, 1996.
[61] M. Koshiba, K. Saitoh, Y. Kobubun, Heterogeneous multi-core fibers: proposal and design principle, IEICE Electron. Express 6 (2) (2009) 98–103.
[62] T. Hayashi, T. Taru, O. Shimakawa, T. Sasaki, E. Sasaoka, Design and fabrication of ultra-low crosstalk and low-loss multi-core fiber, Opt. Express 19 (17) (2011) 16576–16592.
[63] C. Xia, N. Bai, I. Ozdur, X. Zhou, G. Li, Supermodes for optical transmission, Opt. Express 19 (17) (2011) 16653–16664.
[64] K. Nagano, S. Kawakami, S. Nishida, Change of the refractive index in an optical fiber due to external forces, Appl. Opt. 17 (13) (1978) 2080–2085.
[65] Y. Mitsunaga, Y. Katsuyama, Y. Ishida, Thermal characteristics of jacketed optical fibers with initial imperfection, J. Lightwave Technol. 2 (1) (1984) 18–24.

[66] M. Tokuda, S. Seikai, K. Yoshida, N. Uchida, Measurement of baseband frequency response of multimode fibre by using a new type of mode scrambler, Electron. Lett. 13 (5) (1977) 146–147.
[67] M. Ikeda, Y. Murakami, K. Kitayama, Mode scrambler for optical fiber, Appl. Opt. 16 (4) (1977) 1045–1049.
[68] L. Su, K.S. Chiang, C. Lu, Microbend-induced mode coupling in a graded-index multimode fiber, Appl. Opt. 44 (34) (2005) 7394–7402.
[69] K. Kitayama, S. Seikai, N. Uchida, Impulse response prediction based on experimental mode coupling coefficient in a 10-km long graded-index fiber, IEEE J. Quant. Electron. 16 (3) (1980) 356–362.
[70] S. Randel, R. Ryf, A.H. Gnauck, M.A. Mestre, C. Schmidt, R.-J. Essiambre, P.J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, R. Lingle, Mode-multiplexed 6 20-GBd QPSK transmission over 1200-km DGD-compensated few-mode fiber, in: OFC ’12, Paper PDP5C.
[71] D.A. Nolan, X. Chen, M.-J. Li, Fiber with low polarization-mode dispersion, J. Lightwave Technol. 22 (4) (2004) 1066–1077.
[72] A.J. Barlow, J.J. Ramskov-Hansen, D.N. Payne, Birefringence and polarization mode-dispersion in spun single-mode fibers, Appl. Opt. 20 (17) (1981) 2962–2968.
[73] M.-J. Li, D.A. Nolan, Fiber spin-profile designs for producing fibers with low polarization mode dispersion, Opt. Lett. 23 (1998) 1659–1661.
[74] A. Galtarossa, L. Palmieri, A. Pizzinat, Optimized spinning design for low PMD fibers: An analytical approach, J. Lightwave Technol. 19 (10) (2001) 1502–1512.
[75] W.F. Love, Novel mode scrambler for use in optical-fiber bandwidth measurements, in: OFC ’79, Paper ThG2.
[76] Y. Koyamada, T. Horiguchi, M. Tokuda, N. Uchida, Theoretical analysis of optical fiber mode exciters constructed with alternate concatenation of step-index and graded-index fibers, Electron. Commun. J., pt. I 68 (3) (1985) 66–74.
[77] M. Saffman, D.Z. Anderson, Mode multiplexing and holographic demultiplexing communication channels on a multimode fiber, Opt. Lett. 16 (5) (1991) 300–302.
[78] A. Yariv, Coupled-mode theory for guided-wave optics, IEEE J. Quant. Electron. 9 (9) (1973) 919–933.
[79] D. Gloge, Optical power flow in multimode fibers, Bell Syst. Tech. J. 51 (8) (1972) 1767–1780.
[80] D. Marcuse, Losses and impulse response in parabolic index fibers with random bends, Bell Syst. Tech. J. 52 (8) (1973) 1423–1437.
[81] S.S.-H. Yam, F.-T. An, M.E. Marhic, L.G. Kazovsky, Polarization sensitivity of 40 Gb/s transmission over short-reach 62.5 m multimode fiber, in: OFC ’04, Paper FA-5.
[82] M.B. Shemirani, J.M. Kahn, Higher-order modal dispersion in graded-index multimode fiber, J. Lightwave Technol. 27 (23) (2009) 5461–5468.
[83] C.R. Menyuk, P.K.A. Wai, Polarization evolution and dispersion in fibers with spatially varying birefringence, J. Opt. Soc. Am. B 11 (7) (1994) 1288–1296.
[84] H. Kogelnik, P.J. Winzer, Modal birefringence in weakly guiding fibers, J. Lightwave Technol. 30 (14) (2012) 2240–2245.
[85] C.D. Poole, J.M. Wiesenfeld, A.R. McCormick, K.T. Nelson, Broadband dispersion compensation by using the higher-order spatial mode in a two-mode fiber, Opt. Lett. 17 (14) (1992) 985–987.

[86] A.H. Gnauck, L.D. Garrett, Y. Danziger, U. Levy, M. Tur, Dispersion and dispersion-slope compensation of NZDSF over the entire C band using higher-order-mode fibre, Electron. Lett. 36 (23) (2000) 1946–1947.
[87] A. Huttunen, P. Törmä, Optimization of dual-core and microstructure fiber geometries for dispersion compensation and large mode area, Opt. Express 13 (2) (2005) 627–635.
[88] G. Yabre, Comprehensive theory of dispersion in graded-index optical fibers, J. Lightwave Technol. 18 (3) (2000) 166–177.
[89] G.J. Foschini, C.D. Poole, Statistical theory of polarization dispersion in single mode fibers, J. Lightwave Technol. 9 (11) (1991) 1439–1456.
[90] C.D. Poole, Statistical treatment of polarization dispersion in single-mode fiber, Opt. Lett. 13 (8) (1988) 687–689.
[91] P.K.A. Wai, C.R. Menyuk, Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence, J. Lightwave Technol. 14 (2) (1996) 148–157.
[92] A. Galtarossa, L. Palmieri, M. Schiano, T. Tambosso, Measurement of biregringence correlation length in long, single-mode fibers, Opt. Lett. 26 (13) (2001) 962–964.
[93] G.J. Foschini, L.E. Nelson, R.M. Jopson, H. Kogelnik, Probability densities of second-order polarization mode dispersion including polarization dependent chromatic fiber dispersion, IEEE Photon. Technol. Lett. 12 (3) (2000) 293–295.
[94] G.J. Foschini, L.E. Nelson, R.M. Jopson, H. Kogelnik, Statistics of second-order PMD depolarization, J. Lightwave Technol. 19 (12) (2001) 1882–1886.
[95] R.A. Panicker, A.P.T. Lau, J.P. Wilde, J.M. Kahn, Experimental comparison of adaptive optics algorithms in 10-Gb/s multimode fiber transmission, J. Lightwave Technol. 27 (24) (2009) 5783–5789.
[96] M.L. Mehta, Random Matrices, third ed., Elsevier Academic, 2006.
[97] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, J. Math. Phys. 6 (3) (1965) 440–450.
[98] H.G. Golub, C.F. van Loan, Matrix Computations, third ed., Johns Hopkins, 1996.
[99] S.N. Majumdar, O. Bohigas, A. Lakshminarayan, Exact minimum eigenvalues distribution of entangled random pure state, J. Stat. Phys. 131 (1) (2008) 33–49.
[100] Y. Chen, D.-Z. Liu, D.-S. Zhou, Smallest eigenvalue distribution of the fixed-trace Laguerre beta-ensemble, J. Phys. A: Math. Theor. 43 (2010) 315303.
[101] J.L. Schiff, The Laplace Transform—Theory and Applications, Springer, 1999 (Chapter 4).
[102] M. Karlsson, Probability density functions of the differential group delay in optical fiber communication systems, J. Lightwave Technol. 19 (3) (2001) 324–331.
[103] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math. 62 (3) (1955) 548–564.
[104] E. Wigner, On the distribution of the roots of certain symmetric matrices, Ann. Math. 67 (2) (1958) 325–328.
[105] T. Tao, V.H. Vu, From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices, Bull. Am. Math. Soc. 46 (3) (2009) 337–396.
[106] L. Erdös, J. Ramíez, B. Schlein, T. Tao, V.H. Vu, H.T. Yau, Bulk universality for Wigner Hermitian matrices with subexponential decay, Math. Res. Lett. 17 (4) (2010) 667–674.
[107] F. Götze, M. Gordin, Limit correlation functions for fixed trace random matrix ensembles, Commun. Math. Phys. 281 (1) (2008) 203–229.

[108] D.-Z. Liu, D.-S. Zhou, Some universal properties for restricted trace Gaussian orthogonal, unitary and symplectic ensembles, J. Stat. Phys. 140 (2) (2010) 268–288.
[109] D. Voiculescu, K. Dykema, A. Nica, Free Random Variables, CRM Monograph Series vol. 1, American Mathematical Society, 1992.
[110] A. Nica, R. Speicher, Lectures on the Combinatorics of Free Probability, London Mathematical Society Lecture Note Series, vol. 335, Cambridge Univ. Press, 2006.
[111] D. Voiculescu, Limit laws for random matrices and free products, Invent. Math. 104 (1) (1991) 201–220.
[112] C.A. Tracy, H. Widom, Fredholm determinants, differential equations and matrix models, Commun. Math. Phys. 163 (1) (1994) 33–72.
[113] F. Bornemann, On the numerical evaluation of Fredholm determinants, Math. Comput. 79 (270) (2010) 871–915.
[114] F. Bornemann, On the numerical evaluation of distribution in random matrix theory: A review, Markov Proc. Relat. Fields 16 (4) (2010) 803–866.
[115] C.A. Tracy, H. Widom, Level-spacing distributions and the Airy kernel, Phys. Lett. B 305 (1–2) (1993) 115–118.
[116] C.A. Tracy, H. Widom, Level-spacing distributions and the Airy kernel, Commun. Math. Phys. 159 (1) (1994) 151–174.
[117] F. Bornemann, Asymptotic independence of the extreme eigenvalues of Gaussian unitary ensemble, J. Math. Phys. 51 (2010) 023514.
[118] K.-P. Ho, Exact model for mode-dependent gains and losses in multimode fiber, J. Lightwave Technol. 30 (23) (2012) 3603–3609.
[119] R. Bellman, Limit theorems for non-commutative operations I, Duke Math. J. 21 (3) (1954) 491–500.
[120] H. Furstenberg, H. Kesten, Products of random matrices, Ann. Math. Stat. 31 (2) (1960) 457–469.
[121] J.E. Cohen, C.M. Newman, The stability of large random matrices and their products, Ann. Probab. 12 (2) (1984) 283–310.
[122] A. Crisanti, G. Paladin, A. Vulpiani, Products of Random Matrices in Statistical Physics, Springer, 1993.
[123] M.A. Berger, Central limit theorem for products of random matrices, Trans. Am. Math. Soc. 285 (2) (1984) 777–803.
[124] A. Mecozzi, M. Shtaif, The statistics of polarization-dependent loss in optical communication systems, IEEE Photon. Technol. Lett. 14 (3) (2002) 313–315.
[125] P. Lu, L. Chen, X. Bao, Statistical distribution of polarization dependent loss in the presence of polarization mode dispersion in single mode fibers, IEEE Photon. Technol. Lett. 13 (5) (2001) 451–453.
[126] A. Galtarossa, L. Palmieri, The exact statistics of polarization-dependent loss in fiber-optic links, IEEE Photon. Technol. Lett. 15 (1) (2003) 57–59.
[127] P.W. Hooijmans, Coherent Optical System Design, John Wiley & Sons, 1994 (Section A.4).
[128] G.R. Grimmett, D.R. Stirzaker, Probability and Random Processes, second ed., Oxford, 1992.
[129] K.-P. Ho, Statistical properties of stimulated Raman crosstalk in WDM systems, J. Lightwave Technol. 18 (7) (2000) 915–921.
[130] K.-P. Ho, Central limits for the products of free random variables. <http://arxiv.org/abs/1101.5220>.

[131] D.E. Knuth, The Art of Computer Programming II: Seminumerical Algorithms, third ed., Addison Wesley, 1998.
[132] U. Levy, H. Kobrinsky, A.A. Friesem, Angular multiplexing for multichannel communication in a single fiber, IEEE J. Quant. Electron. QE-17 (11) (1981) 2215–2224.
[133] A. Amphawan, Holographic mode-selective launch for bandwidth enhancement in multimode fiber, Opt. Express 19 (10) (2011) 9056–9065.
[134] J. Carpenter, T.D. Wilkinson, Holographic offset launch for dynamic optimization and characterization of multimode fiber bandwidth, J. Lightwave Technol. 30 (10) (2012) 1437–1443.
[135] E. Alon, V. Stojanović, J.M. Kahn, S.P. Boyd and M.A. Horowitz, Equalization of modal dispersion in multimode fiber using spatial light modulators, in: GlobeCom '04.
[136] R.A. Panicker, J.P. Wilde, J.M. Kahn, D.F. Welch, I. Lyubomirsky, 10×10 Gb/s DWDM transmission through 2.2 km multimode fiber using adaptive optics, IEEE Photon. Technol. Lett. 19 (15) (2007) 1154–1156.
[137] B. Franz, H. Bülow, Experimental evaluation of principal mode groups as high-speed transmission channels in spatial multiplex systems, IEEE Photon. Technol. Lett. 24 (16) (2012) 1363–1365.
[138] C.R. Doerr, Proposed architecture for MIMO optical demultiplexing using photonic integration, IEEE Photon. Technol. Lett. 23 (21) (2011) 1573–1575.
[139] N.K. Fontaine, C.R. Doerr, M.A. Mestre, R. Ryf, P. Winzer, L. Buhl, Y. Sun, X. Jiang, R. Lingle, Space-division multiplexing and all-optical MIMO demultiplexing using a photonic integrated circuit, in: OFC '12, Paper PDP5B.
[140] H. Bülow, Optical-mode demultiplexing by optical MIMO filtering of spatial samples, IEEE Photon. Technol. Lett. 24 (12) (2012) 1045–1047.
[141] S.O. Haykin, Adaptive Filter Theory, fourth ed., Prentice Hall, 2001.
[142] A.J. Juarez, C.A. Bunge, S. Warm, K. Petermann, Perspectives of principal mode transmission in mode-division-multiplex operation, Opt. Express 20 (13) (2012) 13810–13823.
[143] M. Shtaif, A. Mecozzi, M. Tur, J.A. Nagel, A compensator for the effects of high-order polarization mode dispersion in optical fibers, IEEE Photon. Technol. Lett. 12 (4) (2000) 434–436.
[144] Q. Yu, L.-S. Yan, Y. Xie, M. Hauer, A.E. Willner, Higher order polarization mode dispersion compensation using a fixed time delay followed by a variable time delay, IEEE Photon. Technol. Lett. 13 (8) (2001) 863–865.
[145] S.G. Evangelides, L.F. Mollenauer, J.P. Gordon, N.S. Bergano, Polarization multiplexing with soliton, J. Lightwave Technol. 10 (1) (1992) 28–35.
[146] L. Nelson, H. Kogelnik, Coherent crosstalk impairments in polarization multiplexed transmission due to polarization mode dispersion, Opt. Express 7 (10) (2000) 350–361.
[147] Z. Wang, C. Xie, X. Ren, PMD and PDL impairments in polarization division multiplexing signals with direct detection, Opt. Express 17 (10) (2009) 7993–8004.
[148] E. Ip, J.M. Kahn, Fiber impairment compensation using coherent detection and digital signal processing, J. Lightwave Technol. 28 (4) (2010) 502–519.
[149] K. Koebele, M. Salsi, G. Charlet, S. Bigo, Nonlinear effects in mode-division-multiplexed transmission over few-mode optical fiber, IEEE Photon. Technol. Lett. 23 (18) (2011) 1316–1318.

[150] X. Chen, A. Li, G. Gao, A. Al Amin, W. Shieh, Characterization of fiber nonlinearity and analysis of its impact on link capacity limit of two-mode fibers, IEEE Photon. J. 4 (2) (2012) 455–460.
[151] A. Mecozzi, C. Antonelli, M. Shtaif, Nonlinear propagation in multi-mode fibers in the strong coupling regime, Opt. Express 20 (11) (2012) 11673–11678.
[152] S. Mumataz, R.-J. Essiambre, G.P. Agrawal, Nonlinear propagation in multimode and multicore fibers: Generalization of the Manakov equations, J. Lightwave Technol. 31 (3) (2013) 398–406.
[153] B. Vucetic, J. Yuan, Space-Time Coding, Wiley, 2003.
[154] S.M. Alamouti, A simple transmit diversity technique for wireless communications, IEEE J. Sel. Areas Commun. 16 (8) (1998) 1451–1458.
[155] V. Tarokh, H. Jafarkhani, A.R. Calderbank, Space-time block coding for wireless communications: Performance results, IEEE J. Sel. Areas Commun. 17 (3) (1999) 451–460.
[156] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.
[157] M. Vetterli, J. Kovačević, Wavelets and Subband Coding, Prentice Hall, 1995 (Section 7.1.1).
[158] Z. Tong, Q. Yang, Y. Ma, W. Shieh, 21.4 Gbit/s transmission over 200 km multimode fiber using coherent optical OFDM, Electron. Lett. 44 (23) (2008) 1373–1374.
[159] W. Shieh, H. Bao, Y. Tang, Coherent optical OFDM: theory and design, Opt. Express 16 (2) (2008) 841–859.
[160] A.J. Lowery, L.B. Du, J. Armstrong, Performance of optical OFDM in ultralong-haul WDM lightwave systems, J. Lightwave Technol. 25 (1) (2007) 131–138.
[161] T. Richardson, R. Urbanke, Modern Coding Theory, Cambridge Univ. Press, 2008.
[162] S. Lin, D.J. Costello, Error Control Coding, second ed., Prentice Hall, 2004.
[163] H. Sari, G. Karam, I. Jeanclaude, Transmission techniques for digital terrestrial TV broadcasting, IEEE Commun. Mag. (1995) 100–109 February.
[164] W.Y. Zou, Y. Wu, COFDM: An overview, IEEE Trans. Broadcast. 41 (1) (1995) 1–8.
[165] M.S. Faruk, K. Kikuchi, Adaptive frequency-domain equalization in digital coherent optical receivers, Opt. Express. 19 (13) (2011) 12789–12798.
[166] J.J. Shynk, Frequency-domain and multirate adaptive filtering, IEEE Sig. Process. Mag. 9 (1) (1992) 15–37.
[167] D. Falconer, S. Lek Ariyavisitakul, A. Benyamin-Seeyar, B. Eidson, Frequency domain equalization for single-carrier broadband wireless systems, IEEE Commun. Mag. 40 (4) (2002) 48–66.
[168] H. Witschnig, T. Mayer, A. Springer, A. Koppleer, L. Maurer, M. Huemer, R. Weigel, A different look on cyclic prefix for SC/FDE, in: IEEE PIMRC ’02.
[169] J. Coon, M. Sandell, M. Beach, J. McGeehan, Channel and noise variance estimation and tracking algorithms for unique-word based single-carrier systems, IEEE Trans. Wireless Commun. 5 (6) (2006) 1488–1496.
[170] K. Kambara, H. Nishimoto, T. Nishimura, T. Ohgane, Y. Ogawa, Subblock processing in MMSE-FDE under fast fading environments, IEEE J. Sel. Areas Commun. 26 (3) (2008) 359–365.
[171] Z. Wang, X. Ma, G.B. Giannakis, OFDM or single-carrier block transmission? IEEE Trans. Commun. 52 (3) (2004) 380–394.
[172] N. Zervos, I. Kalet, Optimized decision feedback equalization versus optimized orthogonal frequency division multiplexing for high-speed data transmission over the local cable network, in: ICC ’89.

[173] N. Benvenuto, S. Tomasin, On the comparison between OFDM and single carrier modulation with a FDE using frequency-domain feedforward filter, IEEE Trans. Commun. 50 (6) (2002) 947–955.
[174] P.J. Winzer, Energy-efficient optical transport capacity scaling through spatial multiplexing, IEEE Photon. Technol. Lett. 23 (13) (2011) 851–853.
[175] S.J. Savory, Digital coherent optical receivers: Algorithms and subsystems, IEEE J. Sel. Top. Quant. Electron. 16 (5) (2010) 1164–1179.
[176] S.J. Savory, Digital filters for coherent optical receivers, Opt. Express 16 (2) (2008) 804–817.
[177] J. Shentu, K. Pantu, J. Armstrong, Effects of phase noise on performance of OFDM systems using an ICI cancellation scheme, IEEE Trans. Broadcast. 49 (2) (2003) 221–224.
[178] S. Wu, Y. Bar-Ness, OFDM systems in the presence of phase noise: consequences and solutions, IEEE Trans. Commun. 52 (11) (2004) 1988–1996.
[179] W. Shieh, Interaction of laser phase noise with differential-mode-delay in few-mode fiber based MIMO systems, in: OFC ’12.
[180] W. Shieh, K.-P. Ho, Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing, Opt. Express 16 (20) (2008) 15718–15727.
[181] C. Xie, WDM coherent PDM-QPSK systems with and without inline optical dispersion compensation, Opt. Express 17 (6) (2009) 4815–4823.
[182] K.-P. Ho, A.P.T. Lau, W. Shieh, Equalization-enhanced phase noise induced timing jitter, Opt. Lett. 36 (4) (2011) 585–587.
[183] S. Randel, R. Ryf, A.H. Gnuack, M.A. Mestre, C. Schmidt, R.-J. Esiambre, R.J. Winzer, R. Delbue, P. Pupalaikis, A. Sureka, Y. Sun, X. Jiang, R. Lingle, Mode-multiplexed 6×20-GBd QPSK transmission over 1200-km DGD-compensated few-mode fiber, in: OFC ’12, Paper PDP5C.5.
[184] L. Grüner-Nielsen, Y. Sun, J.W. Nicholson, D. Jakobsen, R. Lingle, B. Pálsdóttir, Few mode transmission fiber with low DGD, low mode coupling and low loss, in: OFC ’12, Paper PDP5A.
[185] J.W. Cooley, J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, Math. Comput. 19 (1965) 297–301.
[186] P. Duhamel, H. Hoolmann, Split-radix FFT algorithm, Electron. Lett. 20 (1) (1984) 14–16.
[187] H.V. Sorensen, M.T. Heideman, C.S. Burrus, On computing the split-radix FFT, IEEE Trans. Acoust. Speech Signal Process. 34 (1) (1986) 152–156.
[188] S. Winograd, On computing the discrete Fourier transform, Math. Comput. 32 (141) (1978) 175–199.
[189] I.J. Good, The interaction algorithm and practical Fourier analysis, J. Roy. Stat. Soc. B 20 (1958) 361–372 addendum, vol. 22, pp. 373–375, 1960.
[190] L.H. Thomas, Using a computer to solve problems in physics, Applications of Digital Computers, Ginn, 1963.
[191] B.M. Baas, A low-power, high-performance, 1024-point FFT processor, IEEE J. Solid-States Circ. 34 (3) (1999) 380–387.
[192] Y.-W. Lin, H.-Y. Liu, C.-Y. Lee, A 1-GS/s FFT/IFFT processor for UWB applications, IEEE J. Solid-States Circ. 40 (8) (2005) 1726–1735.