Decision–Feedback Equalisation with Fixed–Lag Smoothing in Nonlinear Channels

by

David Wayne Bartel

Bachelor of Applied Science (Exploration Geophysics),
Curtin University of Technology, 1982
Graduate Diploma of Applicable Mathematics,
Murdoch University, 1988
Master of Mathematical Sciences (Signal and Information Processing),
University of Adelaide, 1999

Thesis submitted for the degree of

Doctor of Philosophy

in

Electrical and Electronic Engineering
University of Adelaide

2006

Chapter 1

Introduction

This thesis is concerned with the problem of effective information recovery in digital communication and storage systems. In a modern magnetic hard drive, for example, information is encoded as a stream of binary digits (bits) and written to magnetic platters with read–write heads. In digital communication systems, the information–bearing bit stream is used to modify the parameters of an electrical, electromagnetic or acoustic wave that propagates to a receiver. In all such digital systems, one is concerned with being able to accurately recover the information stream. Bits that are in error at a receiver can lead to disastrous consequences, such as mistaking the message 'attack at dawn' for 'attack at dusk,' or flipping a sign bit in some computer code within an aircraft's navigation system, leading to a fatal crash.

When echoes occur within a digital information channel, as in shallow underwater–acoustic communication, delayed and distorted copies of a signal are combined at a receiver. These multipath copies may cause intersymbol interference (ISI), whereby one copy of a signal overlaps one or more other copies in time. To combat this interference engineers commonly implement a decision–feedback equaliser (DFE), which attempts to estimate and remove echoes from the primary signal. One of the undesirable shortcomings of DFEs, however, is the presence of feedback errors, which can lead to lengthy error bursts at the output of the receiver logic. It is consequently desirable to seek to improve the performance of DFE algorithms.

In a recent paper, Perreau *et al.* introduced a DFE algorithm that incorporated fixed–lag smoothing [52]. According to Anderson, '... at high signal to noise ratios, smoothing gives greater improvement over filtering than at low signal to noise ratios' [4]. Since the ordinary DFE algorithm (without fixed–lag smoothing) uses filtering only, the fixed–lag variant DFE of Perreau *et al.* thus offered promise. Drawing on an earlier study [5], Anderson notes that with P_S and P_F being the mean square error in estimating a signal with a smoother and a filter, respectively, a minimum bound exists for the ratio P_S/P_F as a function of signal to noise ratio. It is important to observe that this

bound indicates the *potential* improvement of smoothing over filtering, and not the improvement guaranteed to hold for any particular system being modelled.

In this thesis, we provide a clear derivation of the fixed–lag smoothing variant of the DFE, termed the FLSDFE, and show via simulations that it has the potential to improve the equalisation performance over that of the standard DFE. We show that this improvement is robust across benign and difficult channels, but that there are caveats on its use owing to the complicated nonlinear nature of the FLSDFE algorithm on linear and nonlinear channels.

Some motivation for studying the performance of the combination of fixed–lag smoothing, decision–feedback equalisation and nonlinear channels may be had from the words of Stuart Anderson, a radar physicist at Australia's Defence Science and Technology Organisation. Writing in the August 2005 edition of the internal DSTO newsletter *Connections*, Anderson was quoted as saying that exploring the underlying radar physics 'in more depth than seems necessary at the time' was a key factor in the development of enhanced long–range radar surveillance capabilities. We thus approach the current study in the same light, with a view to moving beyond the assumptions inherent in linear channel models of an essentially nonlinear world.

The remainder of this thesis is organized as follows. In section 2 we derive the FLSDFE algorithm within the context of two classes of nonlinear digital communication channel. The first class covers M–ary PSK and QAM signalling, $M \geq 4$, and uses a complex–valued Volterra series with terms of odd order only, owing to a 'bandpass' restriction. The second class covers BPSK signalling, which is a special case of MPSK signalling where we set $M = 2$. For BPSK channels, the input–output model is a real–valued Volterra series, and for contrast we relax the bandpass restriction to allow terms of both odd and even order. For each class of channel, we derive the algebraic form of the FLSDFE estimator, as well as the discrete probability distribution function.

In chapter 3 we discuss state space models, where the states are vectors containing current and past symbols, together with their filtered estimates. These models play an important role in describing the transient and limiting behaviour of the FLSDFE output sequence. The largest state space model is termed the atomic model, and this contains the current and previous symbols, up to the channel memory, together with the filtered estimates of all of these quantities. Although this model captures the transient dynamics of the FLSDFE exactly, it is too unwieldy to use for channels of long memory.

To obviate this exponential complexity, aggregated state space models were found, based upon a computer–algebraic proof that used the theory of set partitions and Bell numbers. We showed that there exists a unique optimal aggregation of the atomic state space model that has fewer states than the atomic model, yet is capable of modelling the transient dynamics of the atomic model exactly. A slightly larger aggregate model was subsequently found that lent itself naturally to the study of the dynamics of error recovery. This suboptimal model, too, was a Finite–State Markov Process, and thus represented the transient dynamics of the atomic model exactly.

We consider the problem of deriving the discrete probability distribution of the error recovery time for general channels, based upon the atomic, optimal and suboptimal state space models. No closed form solutions are found, in general, even for the simplest case of linear BPSK channels of arbitrary memory length. One of the complicating factors is that the number of terms in the formulae for the recovery time seemed to be described by linear recurrence relations related to the Fibonacci recurrence. No general closed form solutions to these recurrence relations seem to be known, at least to the author's knowledge. Nonetheless, we show interesting connections with the theory of higher-order linear recurrence relations, and with the theory of constrained and unconstrained integer partitions.

In chapter 4 we study the performance of the FLSDFE algorithm on a set of six channels. The first two are derived from physics–based modelling of underwater acoustic propagation in a shallow ocean. To demonstrate the reliability of acoustic ray tracing in underwater communications simulation modelling, a comparison of predicted transmission loss was made using three alternative propagation models, one written by the author (HANKEL). We show that the ray tracing technique agrees very well with the other three solutions at a typical communications frequency of 1 kHz.

Based on the ray tracing technique we derived two approximate underwater acoustic channel models in the form of normalized finite–impulse response channels. These were based upon the computed set of rays between a fixed source and receiver, and taking into account the time delays, ray amplitudes and phases. The underwater channel models were subsequently used to simulate BPSK transmission and the FLSDFE algorithm was exercised. Some interesting results were produced from the underwater channels.

To explore these results further, four generic channels were constructed to expose some of the pathological behaviours of the FLSDFE algorithm. Specifically, two linear and

two cubic channels were devised, and each had a 'minimum phase' and a 'nonminimum phase' variant. We showed via simulation the existence of pseudo–resonance peaks in the curves of state transition versus input signal–to–noise ratio, using a published state–space model [24].

Finally, in chapter 5 we illustrate some variants of the FLSDFE algorithm that have the potential to accelerate error recovery, by feeding back smoothed outputs. The form of the FLSDFE discussed in this thesis did not feed back smoothed outputs.

1.1 Literature Review

A great deal of research has been published in recent decades on decision–feedback equalisation, owing to the importance of reducing error rates and of understanding and controlling the phenomenon of error bursts [1–3,7–11,13,14,16,19–24,28,29,34,37,39–42,45,46,51,52,60,61]. O'Reilly and de Oliveira Duarte, for example, investigated bounds on error propagation statistics using a state–space model [51]. Shortly afterwards, Kennedy and Anderson deduced some results for error recovery times of DFEs on noiseless linear FIR channels [39], and derived tight bounds for the probability of an error on a noisy linear FIR channel [40].

Beaulieu, Choy and Altekar continued to produce refined estimates of error bounds for the probability of a decision error, and bounds on the mean and variance of error recovery times, throughout the 1990s [2,3,8,20–24]. All of these results were still limited to linear FIR channels. Interestingly, very little work seems to have been done on the autocorrelation properties of sequences of errors, with the recent paper of Beaulieu and Choy being about the only contribution [10]. Even this study was of limited scope, being restricted to the case of a one–tap linear FIR equaliser. Choy and Beaulieu introduced a state–space model that provided improved bounds for error recovery times, although the formulae that were presented applied strictly to the case of stationary state–transition probabilities [24]. By contrast, this thesis provides partial results for error recovery times using the instantaneous state–transition probabilities and 'exact' state–space models.

Some recent effort has been invested in analysing the statistics of error propagation in decision–feedback equalisers on nonlinear channels. In particular, Tsimbinos and White investigated a second–order Volterra channel, and showed that the error probability and recovery times could be computed based upon the previous results of

Beaulieu and co–workers for the linear FIR channels [61]. Although this work addressed nonlinear channels, the effect of fixed–lag smoothing on decision–feedback equalisation was not addressed.

Perreau *et al.* introduced a fixed–lag smoothing scheme to a decision–feedback equaliser, which the FLSDFE algorithm of this thesis generalizes [52]. In this introductory work, the FLSDFE was studied only for linear FIR channels. This thesis extends that study to the case of nonlinear channels described by Volterra series.

1.2 Original Contributions

The major contributions of this thesis are summarized below.

- Applied the FLSDFE to communications systems with baseband input–output relationship in the form of a Volterra series. This combination, of the equalisation of a nonlinear channel with the FLSDFE algorithm, is novel and is discussed in detail in chapter 2.
- Provided novel connections in chapter 3 between state space models involved in error recovery time studies and some topics from 'pure' mathematics, such as the Fibonacci sequence (and some of its generalizations), integer and set partitions, Bell numbers and restricted growth strings.
- Demonstrated in chapter 4 the existence of 'resonance' phenomena (the existence of an optimum signal–to–noise ratio) within the context of the single–distinct–errors state space model of Choy and Beaulieu [24], which imply that under some situations the *addition* of extra noise may actually lead to shorter mean error burst lengths during channel equalisation.
- Proposed two alternative FLSDFE algorithms in chapter 5 that incorporate the feedback of smoothed outputs, which may accelerate error recovery.

In greater detail, more specific contributions of this thesis are as follows.

- Derived the explicit equations of the FLSDFE algorithm for either MPSK or MQAM signalling on a channel described by a third–order Volterra series, in section 2.3 and appendices A–D.

- In section 2.3.3 showed that the probability distribution of the FLSDFE output cannot be explicitly derived for the case of MPSK and MQAM signalling on a Volterra channel, due to its complexity.

- Derived the explicit equations of the FLSDFE algorithm for BPSK signalling on a channel described by a third–order Volterra series, in section 2.4 and appendices E–G.

- Derived in section 2.4.3 the probability distribution of the FLSDFE estimator for the case of BPSK signalling on a Volterra channel.

- Proposed an *atomic* state space model in section 3.2.1 for the case of FLSDFE operation in filtering–only mode (a smoothing lag of zero), based on a similar model of Kennedy and Anderson [39].

- Provided a novel formulation of the state transition probability matrix for the filtering–only atomic state space model, in section 3.2.3.

- Derived novel results (3.117)–(3.119) for the probability distribution of the error recovery time for the case of BPSK signalling on a memoryless linear FIR channel.

- Raised proposition 3.2.1, concerning a postulated connection between the Horadam $(0, 1, 4, 2)$ sequence (a generalization of the Fibonacci sequence) and the number of distinct paths to error recovery for the case of BPSK signalling on a channel of memory $N = 1$.

- Raised proposition 3.2.2 which postulates a connection between the Fibonacci sequence and the number of distinct classes of error recovery path for the case of BPSK signalling on a channel of memory $N = 1$.

- Raised propositions 3.2.3 and 3.2.4, which generalize propositions 3.2.1 and 3.2.2 to the case of channels with memory $N \geq 0$, and which concern specific generalizations of the Fibonacci sequence.

- Demonstrated a connection between the atomic state space models of section 3.2.1 and the theory of constrained and unconstrained integer partitions, in section 3.2.10.

- Devised an atomic state space model for the case of FLSDFE operation in smoothing–only mode (a nonzero smoothing lag) in section 3.3.1.

- Introduced a formulation of the state transition probability matrix for the smoothing–only atomic state space model in section 3.3.3. This is applicable for the case of BPSK signalling on a general Volterra channel.

- Provided a novel demonstration in section 3.4 of the applicability of the theory of Bell numbers, restricted growth strings and set partitions to the discovery of aggregated state space models—those which satisfy the finite–state Markov process property.

- Demonstrated a previously unreported connection between the single–distinct–errors state space model of Choy and Beaulieu [24], which is a steady–state model, and the transient atomic state space models of section 3.2.1. Specifically, proposition 3.4.1 postulates the existence of a transient aggregated state space model that can be derived from the Choy and Beaulieu model by a judicious set partition of the Choy and Beaulieu states.

- Produced an exact wavenumber–integral model, `HANKEL`, for the rigorous computation of the complex acoustic pressure field in a horizontally–stratified ocean–acoustic environment. Showed via examples in section 4.2.1 that `HANKEL` is of benchmark quality.

- Demonstrated by example in section 4.2.2 that the simpler approach of ray–tracing is adequate for the physics component of an underwater acoustic communications system simulation, at typical frequencies of 1 kHz or more.

- Showed the existence of 'resonance' phenomena (that is, the existence of an optimum SNR) with the use of the FLSDFE in two simulations of underwater acoustic communication, using realistic ocean–acoustic parameters. These phenomena appear not to have been reported previously, and occur in the context of the Choy and Beaulieu single–distinct–errors state space model.

- Illustrated the effect that the resonance phenomena have on error recovery in section 4.3.4, particularly demonstrating that the addition of noise may actually improve the time to recovery.

Chapter 2

The FLSDFE Algorithm

2.1 Introduction

This chapter provides an alternative and extended derivation of a signal processing algorithm introduced in 2000 by Perreau *et al.* [52]. That algorithm was not named in the original paper, so for convenience throughout this thesis we shall refer to it as the FLSDFE (**F**ixed–**L**ag–**S**moothing **D**ecision–**F**eedback–**E**qualiser). The FLSDFE is a non-linear channel equaliser, 'blind' in the sense of not requiring training symbols. It was introduced as having a computational complexity linear in the channel memory; and as having a potential advantage over the class of ordinary decision–feedback equalisers (DFEs) in its use of fixed–lag smoothing [4], which delays the estimation of a given message symbol until more information about that symbol is available.

Perreau *et al.* developed the FLSDFE algorithm with reference to linear FIR (finite–impulse response) channels only [59]. In this chapter we show how to apply the FLSDFE to the wider class of Volterra channels, whose input–output relationship is described by a Volterra series [13, 19, 54, 55, 60, 61]. In the process of this extension, we provide an alternative algorithm development to that used by Perreau *et al.* [52].

We illustrate the development of the FLSDFE algorithm using two specific models of a nonlinear SISO (Single–Input, Single–Output) digital communication system. Section 2.3 treats the general MPSK and MQAM models, and section 2.4 treats the simpler BPSK model. For each system model, we give the explicit form of the estimator $\hat{X}_{t-n|t}$ of random baseband message symbol X_{t-n} at lag $n \in \{0, \ldots, N\}$, where t is the current time index and N is the channel memory.

Sections 2.3.3 and 2.4.3 give new results on the probability distribution function of $\hat{X}_{t-n|t}$, which was not discussed in Perreau *et al.* [52]. For the BPSK channel model, the explicit distribution of $\hat{X}_{t-n|t}$ is provided in section 2.4.3. As for the general MPSK and

MQAM channel models of section 2.3.3, we show the difficulties involved in deriving the distribution of $\hat{X}_{t-n|t}$, and leave the problem open.

2.2 Narrowband signal models

We briefly review some aspects of the statistical theory of communication that provide context for the development of the FLSDFE algorithm. This material is covered amply by an extensive literature, and good starting points are [6, 12, 44, 47, 53, 57].

Suppose we transmit a *narrowband signal* such as a fixed–amplitude underwater communications tone $A\cos\omega_0 t$, of angular frequency $\omega_0 = 2\pi f_0$ radians per second. We wish to use this *carrier wave* for transmitting digital information, and let us do this, say, using *binary–phase–shift keying* (BPSK) [57]. To effect BPSK modulation, we simply multiply the narrowband carrier signal $A\cos\omega_0 t$ by $+1$ to transmit a logical 0; and -1 to transmit a logical 1 (or *vice versa*). We do this multiplication during a *bit period* $t \in [t_k, t_{k+1})$, say, for $k \in \{1, \ldots, K\}$, where K is the number of bits in the message. The transmit waveform is then modelled as $s(t)$, where

$$s(t) = \bigcup_{k\in\{1,\ldots,K\}} s_k(t), \tag{2.1}$$

with

$$s_k(t) = \begin{cases} x_k A\cos\omega_0 t, & \text{if } t \in [t_k, t_{k+1}) \\ 0, & \text{if } t \notin [t_k, t_{k+1}), \end{cases} \tag{2.2}$$

and $x_k \in \{-1, 1\}$, $k \in \{1, \ldots, K\}$. Note that the signal model (2.1)–(2.2) is merely a convenient theoretical device, and the transmitted pressure waveform may be a distorted version of this, owing to limitations and practical constraints imposed by the sonar hardware.

Note that the time index t in the above discussion is continuous time, whereas the time index k in x_k is discrete time. We use discrete time throughout sections 2.3 and 2.4, since therein we deal exclusively with baseband sequences.

The underwater communications scenario described above typically involves complexities that require complicated signal processing to accurately and effectively extract x_k. Echoes, for example, caused by the superposition of delayed and distorted copies of $s(t)$ at the receiver hydrophone, may result in *intersymbol interference* [12, 44, 53, 57],

the mitigation of which is the subject of this thesis. We shall ignore these difficulties for now, and concentrate on a simpler reference model through which we can more easily introduce the notion of the *envelope* process.

Following the treatment in Andrews [6], we suppose that at the output of a receiver we have the continuous–time waveform $V(t)$, a random process, which is given by

$$V(t) = A\cos\omega_0 t + Z(t), \tag{2.3}$$

where $A\cos\omega_0 t$ is a sinusoidal signal and $Z(t)$ is *narrowband noise*, defined as

$$Z(t) = \mathcal{P}(t)\cos\omega_0 t - \mathcal{Q}(t)\sin\omega_0 t. \tag{2.4}$$

We suppose that the receiver implements a narrowband filter, centred at $f_0 = \omega_0/(2\pi)$, with a bandwidth $B \ll f_0$, and that the receiver is effectively a linear, memoryless device. (Note that this contrasts with the assumptions made in sections 2.3 and 2.4, where we effectively assume that the receiver is a nonlinear device with memory.)

In (2.4), $\mathcal{P}(t)$ and $\mathcal{Q}(t)$ are assumed to be independent Gaussian random processes, both with zero mean and a variance of σ^2. Use (2.4) to rewrite (2.3):

$$\begin{aligned} V(t) &= [A + \mathcal{P}(t)]\cos\omega_0 t - \mathcal{Q}(t)\sin\omega_0 t \\ &= R(t)\cos\left[\omega_0 t + \Theta(t)\right], \end{aligned} \tag{2.5}$$

where

$$R(t) = \sqrt{[A + \mathcal{P}(t)]^2 + \mathcal{Q}^2(t)}, \quad \text{and} \tag{2.6}$$

$$\Theta(t) = \tan^{-1}\left(\frac{\mathcal{Q}(t)}{A + \mathcal{P}(t)}\right). \tag{2.7}$$

We term $R(t)$ the *envelope* process and $\Theta(t)$ the *phase* process [6]. It is the envelope that we refer to in sections 2.3 and 2.4.

From Andrews [6], we have the following results for the simple receiver model of (2.3) and (2.4). The random processes $R(t)$ and $\Theta(t)$ have the joint distribution

$$f_{R,\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2}\exp\left[-\frac{\left(r^2 + A^2 - 2Ar\cos\theta\right)}{2\sigma^2}\right]; \tag{2.8}$$

$R(t)$ has the marginal *Rice–Nakagami* distribution

$$f_R(r) = \frac{r}{\sigma^2}\exp\left[-\frac{\left(r^2 + A^2\right)}{2\sigma^2}\right] I_0\left(\frac{Ar}{\sigma^2}\right), \quad r > 0; \tag{2.9}$$

and $\Theta(t)$ has the marginal distribution

$$f_\Theta(\theta) = \frac{1}{2\pi} e^{-s} \left\{ 1 + \sqrt{\pi s}\, e^{s\cos^2\theta} \cos\theta \left[1 + \operatorname{erf}\left(\sqrt{s} \cos\theta \right) \right] \right\}, \quad -\pi < \theta < \pi, \tag{2.10}$$

where I_0 is a modified Bessel function of the first kind, defined for $x \in \mathbb{R}$ by

$$I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{x\cos\theta} d\theta, \tag{2.11}$$

(see also (2.37) from [62], which gives a somewhat different definition for $I_0(z)$) and where erf is the error function, which for $x \in \mathbb{R}$ is given by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{2.12}$$

The quantity $s = A^2/2\sigma^2$ in (2.10) is defined as the *signal–to–noise ratio* (SNR) of the power of the sinusoidal signal to that of the noise.

In sections 2.3 and 2.4, we assume a more general class of channel models than the linear, memoryless model just used for the illustration of the envelope process. Although the formulae (2.8)–(2.10) do not carry over to the more general class of model, the envelope process idea does, and we use that in the rest of this chapter.

2.3 MPSK and MQAM signalling, $M \geq 4$

In this section we derive the FLSDFE estimator, $\hat{X}_{t-n|t}$, in the context of the nonlinear narrowband communication system model of Cheng and Powers [19]. Explicit formulae are given in appendixes A–D for a third–order Volterra channel model, assuming either M–ary phase–shift keying (MPSK) or M–ary quadrature amplitude modulation (MQAM) channel inputs [57], with $M \geq 4$.

Following the problem definition and solution in sections 2.3.1 and 2.3.2, respectively, we discuss the probability distribution function of $\hat{X}_{t-n|t}$ in section 2.3.3. The distribution of $\hat{X}_{t-n|t}$ was not discussed in the original work [52], and the material in 2.3.3 is new. Using QPSK for illustration, we show the considerable difficulties involved in deriving the distribution of $\hat{X}_{t-n|t}$ for general MPSK and MQAM channels, for $M \geq 4$, and we leave this as an open problem. (Section 2.4.3 shows that if we restrict ourselves to the special case of BPSK signalling, which is MPSK with $M = 2$, we *can* derive the explicit formula for the distribution in terms of the error function.)

2.3.1 Problem definition

Following Cheng and Powers [19], assume that stochastic processes $\{X_t\}$ and $\{Y_t\}$ represent the complex–valued *envelopes* (input and output, respectively) of a narrowband SISO (Single–Input, Single–Output) communication system, where $t \in \mathbb{Z}$ is discrete–time. Specifically, we suppose that random variable X_t models a random state in some state–space $\mathcal{A}_M$ [50], the *alphabet* of M–ary signalling [57]. We will regard $\mathcal{A}_M$ as the *signal constellation* of either MPSK or MQAM inputs.

For MPSK inputs, we have

$$\mathcal{A}_M = \left\{ \exp\left(\frac{2\pi m i}{M}\right) : \ m \ \in \{0, \ldots, M-1\} \right\}, \tag{2.13}$$

with $i^2 = -1$ and $M \in \{2^k : k \in \mathbb{Z}^+\}$. Setting $M = 4$, for instance, we have the *quaternary*–phase–shift–keying (QPSK) alphabet $\mathcal{A}_4 = \{1, i, -1, -i\}$. (With $M = 2$, we have the real–valued *binary*–phase–shift–keying (BPSK) alphabet $\mathcal{A}_2 = \{-1, 1\}$, discussed in section 2.4.) For MQAM inputs, various signal constellations are possible, such as rectangular 16–QAM. See section 9.9 of Sklar [57] for some examples.

Suppose X_t and Y_t are related by the Volterra series model [19]

$$\begin{aligned} Y_t = & \sum_{k_1=0}^{N_1} \tilde{h}_1(k_1) X_{t-k_1} + \sum_{k_1=0}^{N_3} \sum_{k_2=0}^{N_3} \sum_{k_3=0}^{N_3} \tilde{h}_3(k_1, k_2, k_3) X_{t-k_1} X_{t-k_2} X^*_{t-k_3} \\ & + \sum_{k_1=0}^{N_5} \cdots \sum_{k_5=0}^{N_5} \tilde{h}_5(k_1, \ldots, k_5) X_{t-k_1} X_{t-k_2} X_{t-k_3} X^*_{t-k_4} X^*_{t-k_5} \\ & + \cdots + \sum_{k_1=0}^{N_{2p+1}} \cdots \sum_{k_{2p+1}=0}^{N_{2p+1}} \tilde{h}_{2p+1}(k_1, \ldots, k_{2p+1}) \\ & \times X_{t-k_1} \ldots X_{t-k_{p+1}} X^*_{t-k_{p+2}} \ldots X^*_{t-k_{2p+1}} + V_t, \end{aligned} \tag{2.14}$$

of *(total) order* $2p+1$, where $p \in \mathbb{Z}^*$. In (2.14), $N_1, N_3, \ldots, N_{2p+1} \in \mathbb{Z}^*$ are called the *memories* of the Volterra *terms* (single– or multiple–sums) of *orders* $1, 3, \ldots, 2p+1$, respectively. The coefficients $\tilde{h}_1, \tilde{h}_3, \ldots, \tilde{h}_{2p+1}$ are called *Volterra kernels* [11, 13, 14, 19, 54, 55]; and $\{V_t\}$ is an *iid* (independent and identically distributed), zero–mean, complex–valued, additive Gaussian noise process representing measurement noise.

We will assume, without loss of generality, a single memory $N \in \mathbb{Z}^*$ for all Volterra terms in (2.14). This can easily be accommodated in the model (2.14) by setting $N = \max\{N_1, N_3, \ldots, N_{2p+1}\}$ and enforcing $\tilde{h}_{2j+1}(k_1, \ldots, k_{2j+1}) = 0$ whenever any index $k_1, \ldots, k_{2j+1} \in \{0, \ldots, N\}$ exceeds memory N_{2j+1}, for $j \in \{0, \ldots, p\}$.

As observed by Cheng and Powers [19], we can use symmetry in the product terms $X_{t-k_1}X_{t-k_2}X^*_{t-k_3}$, $X_{t-k_1}X_{t-k_2}X_{t-k_3}X^*_{t-k_4}X^*_{t-k_5}$, and so on, to produce a model with fewer kernel coefficients. In doing so, we merely re–define the coefficients. With the single memory term N, and making use of this symmetry and kernel re–definition, we thus reduce (2.14) to the more economical form

$$\begin{aligned} Y_t = &\sum_{i_1=0}^{N} h_1(i_1)X_{t-i_1} + \sum_{i_1=0}^{N}\sum_{i_2=i_1}^{N}\sum_{j_1=0}^{N} h_3(i_1,i_2,j_1)X_{t-i_1}X_{t-i_2}X^*_{t-j_1} \\ &+ \sum_{i_1=0}^{N}\sum_{i_2=i_1}^{N}\sum_{i_3=i_2}^{N}\sum_{j_1=0}^{N}\sum_{j_2=j_1}^{N} h_5(i_1,i_2,i_3,j_1,j_2)X_{t-i_1}X_{t-i_2}X_{t-i_3}X^*_{t-j_1}X^*_{t-j_2} \\ &+ \cdots + \sum_{i_1=0}^{N}\sum_{i_2=i_1}^{N}\cdots\sum_{i_{p+1}=i_p}^{N}\sum_{j_1=0}^{N}\sum_{j_2=j_1}^{N}\cdots\sum_{j_p=j_{p-1}}^{N} h_{2p+1}(i_1,\ldots,i_{p+1},j_1,\ldots,j_p) \\ &\times X_{t-i_1}\ldots X_{t-i_{p+1}}X^*_{t-j_1}\ldots X^*_{t-j_p} + V_t. \end{aligned} \tag{2.15}$$

We note from (2.15) that Y_t is a function of the variates $X_t, \ldots, X_{t-N}$ and V_t. For a given channel memory N and with fixed Volterra kernels h_1, h_3, $\ldots$, h_{2p+1}, the probability density function $f_{Y_t}(y_t)$ of Y_t is thus a mapping from $\mathcal{A}_M^{N+1} \times \mathbb{C}$ to $\mathbb{C}$.

The problem we will be concerned with is then as follows. Given samples $y_t, y_{t-1}, \ldots, y_{t-N}$ of measurement process $\{Y_t\}$, as defined in the Volterra channel model (2.15), what was the message symbol x_{t-n} transmitted at the lagged time $t-n$, where t is the current time index and $n \in \{0,\ldots,N\}$ is a discrete time lag? Specifically, we will restrict our attention to the case where the model (2.15) is assumed to be exact and where we know all of the model parameters (the model order p, the channel memory N, the Volterra kernels $h_1, h_3, \ldots, h_{2p+1}$, the signalling alphabet $\mathcal{A}_M$, the *a priori* symbol probability distribution $\mathbb{P}(X_t = x_t)$, and the variance σ_v^2 of noise process $\{V_t\}$). Further, we will assume these model parameters to be time–invariant.

In practice, we would need to estimate some or all of these parameters, some of which might be time–varying. In the original work on the FLSDFE, for example, Perreau *et al.* [52] used the *Expectation–Maximization* (EM) algorithm to estimate the time–varying kernel coefficients for a linear FIR channel of known memory N.

2.3.2 Derivation of FLSDFE formulae

Referring to (2.15), we noted above that Y_t is a function of $X_t, \ldots, X_{t-N}$. Thus Y_t is a function of symbol X_{t-n} at lag $n \in \{0,\ldots,N\}$. Moreover, keeping n fixed, we observe

from the model (2.15) that *each* output variate $Y_t, \ldots, Y_{t-n}$ is a function of X_{t-n}. We illustrate this in table 2.1 below for $N = 6$ and maximum lag $n = 6$. The leftmost column gives the output variate Y_{t-k}, followed by the variates upon which each Y_{t-k} depends, where $k \in \{0, \ldots, n\}$ is an auxiliary lag index.

Table 2.1. Channel outputs $\{Y_t\}$ and dependent variables $\{X_t\}$ and $\{V_t\}$ for a SISO channel of memory $N = 6$ and an FLSDFE smoothing lag of $n = 6$.

output	input symbols							noise
Y_{t-6}	X_{t-6}	X_{t-7}	X_{t-8}	X_{t-9}	X_{t-10}	X_{t-11}	X_{t-12}	V_{t-6}
Y_{t-5}	X_{t-5}	X_{t-6}	X_{t-7}	X_{t-8}	X_{t-9}	X_{t-10}	X_{t-11}	V_{t-5}
Y_{t-4}	X_{t-4}	X_{t-5}	X_{t-6}	X_{t-7}	X_{t-8}	X_{t-9}	X_{t-10}	V_{t-4}
Y_{t-3}	X_{t-3}	X_{t-4}	X_{t-5}	X_{t-6}	X_{t-7}	X_{t-8}	X_{t-9}	V_{t-3}
Y_{t-2}	X_{t-2}	X_{t-3}	X_{t-4}	X_{t-5}	X_{t-6}	X_{t-7}	X_{t-8}	V_{t-2}
Y_{t-1}	X_{t-1}	X_{t-2}	X_{t-3}	X_{t-4}	X_{t-5}	X_{t-6}	X_{t-7}	V_{t-1}
Y_t	X_t	X_{t-1}	X_{t-2}	X_{t-3}	X_{t-4}	X_{t-5}	X_{t-6}	V_t

Table 2.1 highlights the variable X_{t-n} for $n = 6$ and shows the manner in which symbol X_{t-n} is 'seen by the channel' [52] at not only the current time t but at the previous times $t-1, \ldots, t-n$ as well. We also note the presence of the independent, uncorrelated, zero–mean Gaussian variates V_{t-k}, $k \in \{0, \ldots, n\}$. Since these variates are *additive* in the model (2.15), this suggests the use of a *least squares* [38] method for the development of an estimator of the unknown lagged symbol X_{t-n}. This is the method followed by Perreau *et al.* in the original development of the FLSDFE algorithm [52], and the method we will follow in this thesis.

From (2.15) we have for Y_{t-k} the expression

$$\begin{aligned} Y_{t-k} = & \sum_{i_1=0}^{N} h_1(i_1) X_{t-k-i_1} + \sum_{i_1=0}^{N} \sum_{i_2=i_1}^{N} \sum_{j_1=0}^{N} h_3(i_1, i_2, j_1) X_{t-k-i_1} X_{t-k-i_2} X^*_{t-k-j_1} \\ & + \sum_{i_1=0}^{N} \sum_{i_2=i_1}^{N} \sum_{i_3=i_2}^{N} \sum_{j_1=0}^{N} \sum_{j_2=j_1}^{N} h_5(i_1, i_2, i_3, j_1, j_2) \\ & \times X_{t-k-i_1} X_{t-k-i_2} X_{t-k-i_3} X^*_{t-k-j_1} X^*_{t-k-j_2} + \ldots \\ & + \sum_{i_1=0}^{N} \sum_{i_2=i_1}^{N} \cdots \sum_{i_{p+1}=i_p}^{N} \sum_{j_1=0}^{N} \sum_{j_2=j_1}^{N} \cdots \sum_{j_p=j_{p-1}}^{N} h_{2p+1}(i_1, \ldots, i_{p+1}, j_1, \ldots, j_p) \\ & \times X_{t-k-i_1} \ldots X_{t-k-i_{p+1}} X^*_{t-k-j_1} \ldots X^*_{t-k-j_p} + V_{t-k}. \end{aligned} \tag{2.16}$$

The FLSDFE algorithm aims to provide an estimate of the symbol X_{t-n}, at smoothing lag $n \in \{0, \ldots, N\}$, given the measurements $Y_t, \ldots, Y_{t-n}$. With this in mind, we isolate those terms in (2.16) at time index $t-n$. The first–order sum in (2.16) decomposes in the following way:

$$\sum_{i_1=0}^{N} h_1(i_1) X_{t-k-i_1} = \sum_{i_1=0}^{n-k-1} h_1(i_1) X_{t-k-i_1} + h_1(n-k) X_{t-n} + \sum_{i_1=n-k+1}^{N} h_1(i_1) X_{t-k-i_1}. \tag{2.17}$$

Denote the three terms on the right–hand–side of (2.17) as $Y_{t-k}^{(<)}$, $Y_{t-k}^{(=)}$ and $Y_{t-k}^{(>)}$, respectively. We use superscripts $(<)$, $(=)$ and $(>)$, respectively, to indicate the conditions $i_1 < n-k$, $i_1 = n-k$ and $i_1 > n-k$ that are satisfied by the terms. To be specific, we have

$$\sum_{i_1=0}^{N} h_1(i_1) X_{t-k-i_1} = Y_{t-k}^{(<)} + Y_{t-k}^{(=)} + Y_{t-k}^{(>)}, \tag{2.18}$$

where

$$Y_{t-k}^{(<)} = \sum_{i_1=0}^{n-k-1} h_1(i_1) X_{t-k-i_1}, \tag{2.19}$$

$$Y_{t-k}^{(=)} = h_1(n-k) X_{t-n}, \tag{2.20}$$

and

$$Y_{t-k}^{(>)} = \sum_{i_1=n-k+1}^{N} h_1(i_1) X_{t-k-i_1}. \tag{2.21}$$

The third– and higher–order sums in (2.16) may be partitioned similarly, working along each dimension independently. Appendix A lists the terms in the decomposition of (2.16) up to third–order, that is, for $p = 1$.

Observe that the model (2.16) is of the form

$$Y_{t-k} = \sum_{l=0}^{p+1} \sum_{m=0}^{p} A_{t-k}^{(l,m)} X_{t-n}^{l} \left(X_{t-n}^{*}\right)^{m} + V_{t-k}, \tag{2.22}$$

where $A_{t-k}^{(l,m)}$ are coefficient functions that do not depend on the symbol X_{t-n}. Appendix B gives the explicit expressions for $A_{t-k}^{(l,m)}$ for a third–order Volterra series model. For the first–order ($p = 0$) model, (2.22) reads

$$Y_{t-k} = A_{t-k}^{(0,0)} + A_{t-k}^{(1,0)} X_{t-n} + V_{t-k}, \tag{2.23}$$

with coefficient functions given in appendix B and equations (2.19)–(2.21) as

$$A_{t-k}^{(0,0)} = \sum_{i_1=0}^{n-k-1} h_1(i_1) X_{t-k-i_1} + \sum_{i_1=n-k+1}^{N} h_1(i_1) X_{t-k-i_1}, \tag{2.24}$$

and

$$A_{t-k}^{(1,0)} = h_1(n-k). \tag{2.25}$$

Suppose we have available the (filtered) estimators $\hat{X}_{t-k-k_i|t-k-k_i}$ for all symbols X_{t-k-k_i}, where $k_i \in \{0,\ldots,N\} \setminus \{n-k\}$ and $i \in \{1,\ldots,2p+1\}$. Substitute $\hat{X}_{t-k-k_i|t-k-k_i}$ for X_{t-k-k_i} in the coefficient functions $A_{t-k}^{(l,m)}$ of (2.22) for all $k \in \{0,\ldots,n\}$, $l \in \{0,\ldots,p+1\}$ and $m \in \{0,\ldots,p\}$. This yields estimators $\hat{A}_{t-k|t-k}^{(l,m)}$ for $A_{t-k}^{(l,m)}$. We thus have the fixed–lag smoothing least–squares estimator $\hat{X}_{t-n|t}$ of lagged message symbol X_{t-n} given by

$$\hat{X}_{t-n|t} = \underset{X_{t-n}\in\mathcal{A}_M}{\arg\min} \sum_{k=0}^{n} \left| \hat{V}_{t-k|t-k} \right|^2, \tag{2.26}$$

where $\hat{V}_{t-k|t-k}$ is an estimator of V_{t-k}, obtained by rearrangement of (2.22) as

$$\hat{V}_{t-k|t-k} = Y_{t-k} - \sum_{l=0}^{p+1} \sum_{m=0}^{p} \hat{A}_{t-k|t-k}^{(l,m)} X_{t-n}^{l} \left(X_{t-n}^{*}\right)^{m}. \tag{2.27}$$

For the first–order ($p = 0$) model, (2.27) reads, from (2.23),

$$\hat{V}_{t-k|t-k} = Y_{t-k} - \hat{A}_{t-k|t-k}^{(0,0)} - \hat{A}_{t-k|t-k}^{(1,0)} X_{t-n}. \tag{2.28}$$

Appendix C gives the explicit forms of estimators $\hat{A}_{t-k|t-k}^{(l,m)}$ up to third–order, based upon the formulae in appendix B, with the substitution of filtered estimators $\hat{X}_{t|t}$ for X_t. For $p = 0$, for instance, (2.28) reads

$$\begin{aligned} \hat{V}_{t-k|t-k} = Y_{t-k} &- \sum_{i_1=0}^{n-k-1} h_1(i_1) \hat{X}_{t-k-i_1|t-k-i_1} \\ &- \sum_{i_1=n-k+1}^{N} h_1(i_1) \hat{X}_{t-k-i_1|t-k-i_1} - h_1(n-k) X_{t-n}. \end{aligned} \tag{2.29}$$

As can be observed from the form of (2.26), one source of estimators $\hat{X}_{t|t}$ of symbols X_t is obtained merely by setting $n = 0$, corresponding to filtering (a smoothing lag of zero). The FLSDFE algorithm assumes that the filtered estimators discussed above are

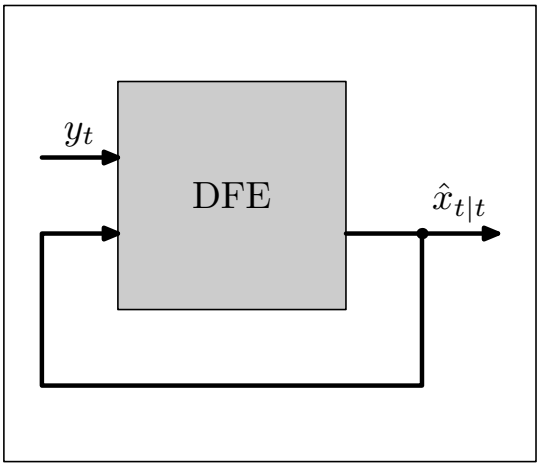

Figure 2.1. Basic block diagram of a decision–feedback equaliser (DFE) [57].

obtained as past and present zero–lag outputs of the FLSDFE algorithm itself. This is the DF (decision–feedback) part of the FLSDFE. Smoothed outputs, at lags $n > 0$, are not fed back.

In figure 2.1 below we give the basic structure of a DFE [57], and in figure 2.2 we expand this in more detail. Of note in figure 2.2 is the feedback nonlinearity (labelled 'quantizer'), which classes the DFE as a nonlinear filter, even on linear channels. Figure 2.3 shows the basic block diagram view of the FLSDFE, where we note the additional output (labelled $\hat{x}_{t-n|t}$) over that of the DFE in figure 2.1. Perreau *et al.* [52] provide a more detailed block diagram of the FLSDFE in their figure 1, with reference to a linear FIR channel model.

2.3.3 Probability distribution function of FLSDFE estimator

We wish to determine the probability distribution function of the FLSDFE output variable $\hat{X}_{t-n|t}$ defined by (2.26), where the domain of $\hat{X}_{t-n|t}$ is the discrete signalling alphabet $\mathcal{A}_M$. To this end, use (2.22) to substitute for Y_{t-k} in (2.27) and rewrite (2.26) as

$$\begin{aligned}\hat{X}_{t-n|t} &= \underset{X_{t-n}\in\mathcal{A}_M}{\arg\min} \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \left(A_{t-k}^{(l,m)} - \hat{A}_{t-k|t-k}^{(l,m)} \right) X_{t-n}^{l} \left(X_{t-n}^{*}\right)^{m} + V_{t-k} \right|^2 \\ &= \underset{X_{t-n}\in\mathcal{A}_M}{\arg\min} \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{A}_{t-k|t-k}^{(l,m)} X_{t-n}^{l} \left(X_{t-n}^{*}\right)^{m} + V_{t-k} \right|^2, \end{aligned} \tag{2.30}$$

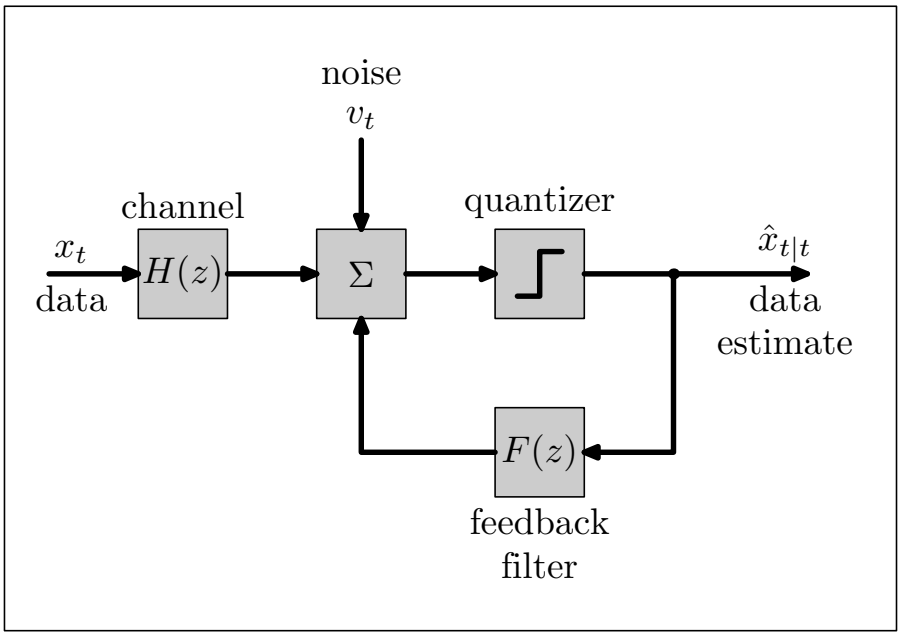

Figure 2.2. The DFE in more detail.

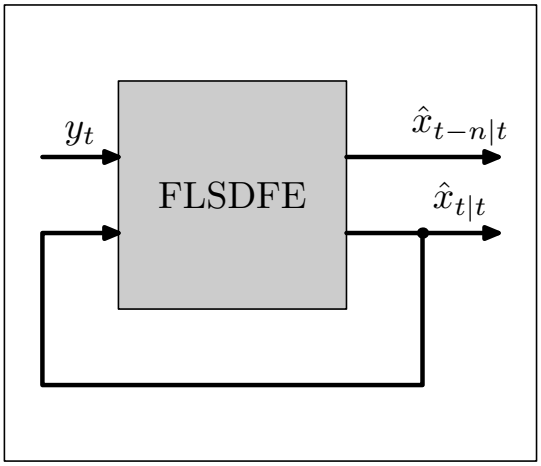

Figure 2.3. Basic block diagram of the FLSDFE.

where, for $k \in \{0,\ldots,n\}$, $l \in \{0,\ldots,p+1\}$ and $m \in \{0,\ldots,p\}$, we have

$$\tilde{A}^{(l,m)}_{t-k|t-k} = A^{(l,m)}_{t-k} - \hat{A}^{(l,m)}_{t-k|t-k}. \tag{2.31}$$

Appendix D gives explicit formulae for the terms $\tilde{A}^{(l,m)}_{t-k|t-k}$ for a third–order channel, wherein we note the presence of *generalized* decision–feedback errors [1–3,7–10,24,39–41,52], terms such as

$$X_{t-k-i_1}X_{t-k-i_2}X^*_{t-k-j_1} - \hat{X}_{t-k-i_1|t-k-i_1}\hat{X}_{t-k-i_2|t-k-i_2}\hat{X}^*_{t-k-j_1|t-k-j_1}. \tag{2.32}$$

From (2.30) and the illustrative formulae in appendix D we see that $\hat{X}_{t-n|t}$ is a function of the discrete variates $X_t, \ldots, X_{t-N-n}, \hat{X}_{t|t}, \ldots, \hat{X}_{t-n+1|t-n+1}, \hat{X}_{t-n-1|t-n-1}, \ldots, \hat{X}_{t-N-n|t-N-n}$ and the continuous variates $V_t, \ldots, V_{t-n}$. With fixed sample values $x_t, \ldots, x_{t-N-n}, \hat{x}_{t|t}, \ldots, \hat{x}_{t-n+1|t-n+1}, \hat{x}_{t-n-1|t-n-1}, \ldots, \hat{x}_{t-N-n|t-N-n}$ of the corresponding discrete variates we have the *conditional* distribution of $\hat{X}_{t-n|t}$:

$$\begin{aligned}\mathbb{P}(\hat{X}_{t-n|t} = \hat{x}_{t-n|t} | X_t = x_t \cap \cdots \cap X_{t-N-n} = x_{t-N-n} \cap \hat{X}_{t|t} = \hat{x}_{t|t} \\ \cap \cdots \cap \hat{X}_{t-n+1|t-n+1} = \hat{x}_{t-n+1|t-n+1} \cap \hat{X}_{t-n-1|t-n-1} = \hat{x}_{t-n-1|t-n-1} \\ \cap \cdots \cap \hat{X}_{t-N-n|t-N-n} = \hat{x}_{t-N-n|t-N-n}),\end{aligned} \tag{2.33}$$

which is a mapping from the domain $\mathbb{C}^{n+1}$ of $V_t, \ldots, V_{t-n}$ to the domain $\mathcal{A}_M$ of $\hat{X}_{t-n|t}$. We seek the form of (2.33) for the MPSK and MQAM signalling model (2.15).

With reference to (2.30), define the random variable Q as

$$Q = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} x_{t-n}^{l} \left(x_{t-n}^{*}\right)^{m} + V_{t-k} \right|^2, \tag{2.34}$$

where $\tilde{a}_{t-k|t-k}^{(l,m)}$ is a sample value of $\tilde{A}_{t-k|t-k}^{(l,m)}$, for $k \in \{0, \ldots, n\}$, $l \in \{0, \ldots, p+1\}$ and $m \in \{0, \ldots, p\}$. We note in (2.34) the explicit functional dependence of Q on the sample value $x_{t-n} \in \mathcal{A}_M$, and we write $Q = Q(x_{t-n})$ to emphasize this. There are M different variates $Q(x_{t-n})$ specified by (2.34), one for each different sample $x_{t-n} \in \mathcal{A}_M$.

Let us assign a unique integer $j \in \{1, \ldots, M\}$ to each $x_{t-n} \in \mathcal{A}_M$, in no particular order, producing from (2.34) the set of random variates Q_j, $j \in \{1, \ldots, M\}$. We see that Q_j is a noncentral gamma variate [64], with the probability density function

$$f_{Q_j}(q) = \frac{1}{2\sigma_v^2} \left(\frac{q}{\lambda_j}\right)^{n/2} \exp\left(-\frac{(\lambda_j + q)}{2\sigma_v^2}\right) I_n\left(\frac{\sqrt{q\lambda_j}}{\sigma_v^2}\right), \quad q \geq 0, \tag{2.35}$$

with noncentrality parameter λ_j given by

$$\lambda_j = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} x_{t-n}^{l} \left(x_{t-n}^{*}\right)^{m} \right|^2. \tag{2.36}$$

The function $I_n(.)$ in (2.35) is the modified Bessel function of the first kind [62], given for integer n and complex z by

$$I_n(z) = \sum_{k=0}^{\infty} \frac{(\frac{z}{2})^{n+2k}}{k!(n+k)!} = \frac{1}{\pi} \int_0^{\pi} e^{z\cos\theta} \cos(n\theta) d\theta. \tag{2.37}$$

Arrange $Q_1, \ldots, Q_M$ in order of magnitude and write as $Q_{(1)} \leq \cdots \leq Q_{(M)}$, where $Q_{(i)}$ denotes the i^{th} *order statistic*, $i \in \{1, \ldots, M\}$ [27]. From the form of (2.34) we observe that $Q_1, \ldots, Q_M$ are neither identically distributed nor independent, although $V_t, \ldots, V_{t-n}$ *are* independent and identically distributed (*iid*) complex Gaussian variates, with a common mean of 0 and a common variance of σ_v^2. The cumulative density function (cdf) $F_{1:M}(q)$ of the minimum order statistic $Q_{(1)}$ of the dependent variates $Q_1, \ldots, Q_M$ is given by [27]

$$F_{1:M}(q) = \sum_{j=1}^{M} (-1)^{j-1} \sum_{1 \leq i_{j+1} < \cdots < i_M \leq M} F_{j:j}^{(i_{j+1}, \ldots, i_M)}(q), \quad q \geq 0, \tag{2.38}$$

where the subscript notation $i : n$ in (2.38) means that the associated cdf is that of the i^{th} order statistic $T_{(i)}$ out of n samples $T_1, \ldots, T_n$, say. Thus, by $F_{1:M}(q)$ we mean the cdf of the minimum order statistic $Q_{(1)}$ of the set of variates $Q_1, \ldots, Q_M$. In the right–hand side of (2.38), the superscript notation $(i_{j+1}, \ldots, i_M)$ indicates that the cdf $F_{j:j}^{(i_{j+1}, \ldots, i_M)}(q)$ is that of the maximum order statistic $Q_{(i_j)}$ of the set of variates $Q_{i_1}, \ldots, Q_{i_j}$, which is a sample of size j drawn without replacement from $Q_1, \ldots, Q_M$. By $(i_{j+1}, \ldots, i_M)$ we mean that the $M - j$ variates $Q_{i_{j+1}}, \ldots, Q_{i_M}$ have been *dropped* from the full set $\{Q_{i_1}, \ldots, Q_{i_j}, Q_{i_{j+1}}, \ldots, Q_{i_M}\}$, leaving only $Q_{i_1}, \ldots, Q_{i_j}$ for consideration. We note that $i_1, \ldots, i_j, i_{j+1}, \ldots, i_M$ is a permutation of $1, \ldots, M$.

To illustrate the application of (2.38) consider the case of QPSK signalling, with baseband alphabet $\mathcal{A}_4 = \{1, i, -1, -i\}$ as obtained from (2.13) with $M = 4$. Map the integers $j \in \{1, \ldots, 4\}$ to the symbols in the set $\mathcal{A}_4$ in the order listed, so that the assignment $j = 1$ maps to symbol $x_{t-n} = 1$; $j = 2$ maps to symbol $x_{t-n} = i$; $j = 3$ maps to symbol $x_{t-n} = -1$; and $j = 4$ maps to symbol $x_{t-n} = -i$.

We thus have the dependent noncentral gamma variates $Q_1, \ldots, Q_4$, defined from (2.34) as

$$Q_1 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} + V_{t-k} \right|^2, \tag{2.39}$$

$$Q_2 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^m i^{l+m} + V_{t-k} \right|^2, \tag{2.40}$$

$$Q_3 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^{l+m} + V_{t-k} \right|^2, \tag{2.41}$$

and

$$Q_4 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^l i^{l+m} + V_{t-k} \right|^2, \tag{2.42}$$

with pdf $f_{Q_j}(q)$ in (2.35) for $j \in \{1, \ldots, 4\}$ and with noncentrality parameters $\lambda_1, \ldots, \lambda_4$ given from (2.36) as

$$\lambda_1 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} \right|^2, \tag{2.43}$$

$$\lambda_2 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^m i^{l+m} \right|^2, \tag{2.44}$$

$$\lambda_3 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^{l+m} \right|^2, \tag{2.45}$$

and

$$\lambda_4 = \sum_{k=0}^{n} \left| \sum_{l=0}^{p+1} \sum_{m=0}^{p} \tilde{a}_{t-k|t-k}^{(l,m)} (-1)^l i^{l+m} \right|^2. \tag{2.46}$$

From (2.38) the cumulative density function $F_{1:4}(q)$ of the minimum order statistic $Q_{(1)}$ of the variates $Q_1, \ldots, Q_4$ is thus given by

$$\begin{aligned} F_{1:4}(q) &= \sum_{j=1}^{4} (-1)^{j-1} \sum_{1 \leq i_{j+1} < \cdots < i_4 \leq 4} F_{j:j}^{(i_{j+1}, \ldots, i_4)}(q) \\ &= \sum_{1 \leq i_2 < i_3 < i_4 \leq 4} F_{1:1}^{(i_2,i_3,i_4)}(q) - \sum_{1 \leq i_3 < i_4 \leq 4} F_{2:2}^{(i_3,i_4)}(q) + \sum_{1 \leq i_4 \leq 4} F_{3:3}^{(i_4)}(q) - F_{4:4}(q) \\ &= F_{1:1}^{(1,2,3)}(q) + F_{1:1}^{(1,2,4)}(q) + F_{1:1}^{(1,3,4)}(q) + F_{1:1}^{(2,3,4)}(q) - F_{2:2}^{(1,2)}(q) - F_{2:2}^{(1,3)}(q) \\ &\quad - F_{2:2}^{(1,4)}(q) - F_{2:2}^{(2,3)}(q) - F_{2:2}^{(2,4)}(q) - F_{2:2}^{(3,4)}(q) + F_{3:3}^{(1)}(q) + F_{3:3}^{(2)}(q) \\ &\quad + F_{3:3}^{(3)}(q) + F_{3:3}^{(4)}(q) - F_{4:4}(q). \end{aligned} \tag{2.47}$$

The parenthetical superscripts in (2.47) indicate variates that have been dropped. Thus $F_{1:1}^{(1,2,3)}(q)$ refers to the maximum order statistic of the set of variates $\{Q_1, \ldots, Q_4\}$ after Q_1, Q_2 and Q_3 have been dropped. In other words, $F_{1:1}^{(1,2,3)}(q)$ reduces to the cdf of Q_4 itself, namely the function

$$F_{Q_4}(q) = \int_0^q f_{Q_4}(q')dq', \tag{2.48}$$

where $f_{Q_4}(q)$ is from (2.35) with $j = 4$. To take another example from (2.47), by $F_{2:2}^{(1,3)}(q)$ we mean the cdf of the maximum order statistic of the variates Q_2 and Q_4, which are the remaining variates after removal of Q_1 and Q_3 from the set $\{Q_1, \ldots, Q_4\}$.

There are considerable complications arising from the dependence of the variables $Q_1, \ldots, Q_M$, as the QPSK example above attests. For QPSK signalling, (2.47) has 15 distinct cdf terms, of which 11 involve integrals of multivariate density functions. In general, for MPSK or MQAM signalling, the expansion of (2.38) produces $2^M - 1$ distinct cdf terms, which is one less than the number of subsets of $\{Q_1, \ldots, Q_M\}$. We subtract one, as we do not consider the empty subset $\emptyset$ of $\{Q_1, \ldots, Q_M\}$ in (2.38).

We now return to the problem of deriving the probability distribution function (2.33) of the FLSDFE estimator $\hat{X}_{t-n|t}$ of (2.30). Comparing (2.30) and (2.34), we see that we can write (2.30) in the form

$$\hat{X}_{t-n|t} = \underset{x_{t-n} \in \mathcal{A}_M}{\arg} \left(Q_{(1)} \right), \tag{2.49}$$

where $Q_{(1)}$ is the minimum order statistic of $Q_1, \ldots, Q_M$, with the cumulative density function $F_{1:M}(q)$ given by (2.38).

The arg mapping in (2.49) is from the continuous interval $[\,0, \infty)$ to the discrete complex alphabet $\mathcal{A}_M$. Aside from the complications illustrated in (2.47), the nonlinearity of the mapping (2.49) makes it difficult to derive the form of the probability distribution function, and we leave this as an unsolved problem. (The next section deals with the special case of BPSK signalling, though, for which we *can* derive the explicit form of the probability distribution function of the FLSDFE estimator in terms of error functions.)

2.4 BPSK signalling

Binary Phase–Shift Keying (BPSK) is a special case of M–ary Phase–Shift Keying (MPSK), which was the subject of section 2.3. Setting $M = 2$ in (2.13), we see that the alphabet of BPSK signalling is $\mathcal{A}_2 = \{-1, 1\}$. In this section we derive the FLSDFE estimator $\hat{X}_{t-n|t}$ for the case of BPSK inputs to a channel described by a Volterra series with real–valued kernel coefficients.

For our channel model, we draw upon a recent paper involving the use of decision–feedback equalisation in nonlinear channels [61]. The extension we add to that work is the use of the FLSDFE algorithm, instead of the ordinary DFE. In addition, in section 2.4.3, we derive the probability distribution function of the FLSDFE estimator $\hat{X}_{t-n|t}$ of baseband message symbol X_{t-n} at fixed lag $n \in \{0, \ldots, N\}$. The results in section 2.4.3 are new.

2.4.1 Problem definition

We assume the following SISO baseband channel model [61]:

$$Y_t = \sum_{k_1=0}^{N} h_1(k_1) X_{t-k_1} + \sum_{k_1=0}^{N} \sum_{k_2=k_1}^{N} h_2(k_1,k_2) X_{t-k_1} X_{t-k_2} + \dots + \sum_{k_1=0}^{N} \sum_{k_2=k_1}^{N} \cdots \sum_{k_p=k_{p-1}}^{N} h_p(k_1,\dots,k_p) X_{t-k_1} \dots X_{t-k_p} + V_t, \tag{2.50}$$

where all quantities are real–valued, and the domain of X_t is the BPSK signalling alphabet $\mathcal{A}_2 = \{-1, 1\}$, as obtained from (2.13) with $M = 2$. The additive noise process $\{V_t\}$ is assumed to be iid, zero–mean and Gaussian, with variance σ_v^2. We assume the model (2.50) to be time–invariant, with all of its parameters known. Given a set of contiguous channel outputs $\{Y_t, \dots, Y_{t-n}\}$, we wish to estimate the unknown message symbol X_{t-n} at the fixed discrete–time lag $n \in \{0, \dots, N\}$.

2.4.2 Derivation of FLSDFE formulae

Consider the output Y_{t-k} from the model (2.50), where $k \in \{0, ..., n\}$ is an auxiliary lag index and $n \in \{0, ..., N\}$ is the fixed lag of the unknown symbol X_{t-n}:

$$Y_{t-k} = \sum_{k_1=0}^{N} h_1(k_1) X_{t-k-k_1} + \sum_{k_1=0}^{N} \sum_{k_2=k_1}^{N} h_2(k_1,k_2) X_{t-k-k_1} X_{t-k-k_2} + \dots + \sum_{k_1=0}^{N} \sum_{k_2=k_1}^{N} \cdots \sum_{k_p=k_{p-1}}^{N} h_p(k_1,\dots,k_p) X_{t-k-k_1} \dots X_{t-k-k_p} + V_{t-k}. \tag{2.51}$$

Following section 2.3.2, as the first step in deriving the FLSDFE estimator $\hat{X}_{t-n|t}$ of X_{t-n}, we isolate terms in (2.51) that contain X_{t-n} or its powers. Appendix E gives the expansion of (2.51) to third order $(p = 3)$.

Since $X_t \in \{-1, 1\}$, observe that for $m \in \mathbb{Z}^+$, $X_t^{2m} = 1$ and $X_t^{2m+1} = X_t$. Thus we may write (2.51) in the simpler form

$$Y_{t-k} = A_{t-k} + B_{t-k} X_{t-n} + V_{t-k}, \tag{2.52}$$

where B_{t-k} is obtained by collecting together all the reduced terms in (2.51) containing X_{t-n}, and A_{t-k} is the remainder, being the sum of those reduced terms in (2.51) that do not contain X_{t-n}. Appendix F gives the terms A_{t-k} and B_{t-k} to third order.

Analogously to the development in section 2.3.2, we now use a least–squares criterion [38] to obtain the FLSDFE estimator $\hat{X}_{t-n|t}$ of lagged symbol X_{t-n}. Following Perreau *et al.* [52] we have

$$\begin{aligned}\hat{X}_{t-n|t} &= \underset{X_{t-n}\in\mathcal{A}_2}{\arg\min} \sum_{k=0}^{n} \left(Y_{t-k} - \hat{A}_{t-k|t-k} - \hat{B}_{t-k|t-k}X_{t-n}\right)^2 \\ &= \underset{X_{t-n}\in\{-1,1\}}{\arg\min} \sum_{k=0}^{n} \Big[\left(Y_{t-k} - \hat{A}_{t-k|t-k}\right)^2 + \left(\hat{B}_{t-k|t-k}X_{t-n}\right)^2 \\ &\quad - 2\left(Y_{t-k} - \hat{A}_{t-k|t-k}\right)\hat{B}_{t-k|t-k}X_{t-n}\Big] \\ &= \operatorname{sign}\left\{\sum_{k=0}^{n}\left[\left(Y_{t-k} - \hat{A}_{t-k|t-k}\right)\hat{B}_{t-k|t-k}\right]\right\}, \end{aligned} \tag{2.53}$$

where $\operatorname{sign}(x)$ is the *signum* function, defined for $x \in \mathbb{R}$ by

$$\operatorname{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{otherwise.} \end{cases} \tag{2.54}$$

In (2.53) the random variables $\hat{A}_{t-k|t-k}$ and $\hat{B}_{t-k|t-k}$ are (filtered) estimators of $A_{t-k|t-k}$ and $B_{t-k|t-k}$, respectively. We merely take the corresponding formulae for A_{t-k} and B_{t-k}, and substitute filtered estimates $\hat{X}_{t-k-\delta|t-k-\delta}$ for $X_{t-k-\delta}$, where $k \in \{0,\ldots,n\}$ and $\delta \in \{0,\ldots,N\} \setminus \{n-k\}$, with $n \in \{0,\ldots,N\}$. Appendix G gives the explicit formulae for $\hat{A}_{t-k|t-k}$ and $\hat{B}_{t-k|t-k}$ to third order, following the analogous formulae in appendix F for A_{t-k} and B_{t-k}.

2.4.3 Probability distribution function of FLSDFE estimator

Use (2.52) to replace Y_{t-k} in (2.53), and rearrange (2.53) as

$$\hat{X}_{t-n|t} = \operatorname{sign}\left\{\sum_{k=0}^{n}\left[\hat{B}_{t-k|t-k}\left(A_{t-k} - \hat{A}_{t-k|t-k} + B_{t-k}X_{t-n}\right) + \hat{B}_{t-k|t-k}V_{t-k}\right]\right\}. \tag{2.55}$$

We note from (2.55) that $\hat{X}_{t-n|t}$ is a function of the random variables A_{t-k}, B_{t-k}, $\hat{A}_{t-k|t-k}$, $\hat{B}_{t-k|t-k}$, X_{t-n} and V_{t-k}, where $k \in \{0,\ldots,n\}$ and $n \in \{0,\ldots,N\}$. Further, as shown in appendixes F and G for a third–order channel model, we know that A_{t-k} and B_{t-k} are functions of X_t, …, X_{t-n+1}, X_{t-n-1}, …, X_{t-N-n}; while $\hat{A}_{t-k|t-k}$ and $\hat{B}_{t-k|t-k}$ are functions of $\hat{X}_{t|t}$, …, $\hat{X}_{t-n+1|t-n+1}$, $\hat{X}_{t-n-1|t-n-1}$, …, $\hat{X}_{t-N-n|t-N-n}$.

As in section 2.3.3, substitute sample values x_t, ..., x_{t-N-n}, $\hat{x}_{t|t}$, ..., $\hat{x}_{t-n+1|t-n+1}$, $\hat{x}_{t-n-1|t-n-1}$, ..., $\hat{x}_{t-N-n|t-N-n}$ for the discrete variables X_t, ..., X_{t-N-n}, $\hat{X}_{t|t}$, ..., $\hat{X}_{t-n+1|t-n+1}$, $\hat{X}_{t-n-1|t-n-1}$, ..., $\hat{X}_{t-N-n|t-N-n}$, respectively, in the expressions for A_{t-k} and B_{t-k}, $\hat{A}_{t-k|t-k}$, $\hat{B}_{t-k|t-k}$. We thus obtain corresponding sample values a_{t-k}, b_{t-k}, $\hat{a}_{t-k|t-k}$, $\hat{b}_{t-k|t-k}$ and x_{t-n} of variables A_{t-k}, B_{t-k}, $\hat{A}_{t-k|t-k}$, $\hat{B}_{t-k|t-k}$ and X_{t-n}, respectively, where $k \in \{0, \ldots, n\}$ and $n \in \{0, \ldots, N\}$.

We noted in section 2.4.1 that the additive noise process $\{V_t\}$ in the channel model (2.50) is iid, zero–mean and Gaussian, with variance σ_v^2. With the sample values discussed above, then, it follows that the argument of the signum function in (2.55) is also Gaussian, T say, with mean μ_T and variance σ_T^2, where

$$T = \sum_{k=0}^{n} \left[\hat{b}_{t-k|t-k} \left(a_{t-k} - \hat{a}_{t-k|t-k} + b_{t-k} x_{t-n} \right) + \hat{b}_{t-k|t-k} V_{t-k} \right], \tag{2.56}$$

$$\mu_T = \sum_{k=0}^{n} \left[\hat{b}_{t-k|t-k} \left(a_{t-k} - \hat{a}_{t-k|t-k} + b_{t-k} x_{t-n} \right) \right], \tag{2.57}$$

and

$$\sigma_T^2 = \sigma_v^2 \sum_{k=0}^{n} \hat{b}_{t-k|t-k}^2. \tag{2.58}$$

It is easy to show now that the (conditional) probability distribution function of the FLSDFE estimator $\hat{X}_{t-n|t}$ for the symbol X_{t-n} in the case of BPSK signalling is

$$\begin{aligned} &\mathbb{P}(\hat{X}_{t-n|t} = \hat{x}_{t-n|t} | X_t = x_t \cap \cdots \cap X_{t-N-n} = x_{t-N-n} \cap \hat{X}_{t|t} = \hat{x}_{t|t} \\ &\quad \cap \cdots \cap \hat{X}_{t-n+1|t-n+1} = \hat{x}_{t-n+1|t-n+1} \cap \hat{X}_{t-n-1|t-n-1} = \hat{x}_{t-n-1|t-n-1} \\ &\quad \cap \cdots \cap \hat{X}_{t-N-n|t-N-n} = \hat{x}_{t-N-n|t-N-n}) \\ &= \frac{1}{2} \left(1 + \hat{x}_{t-n|t} \operatorname{sign}(\mu_T) \operatorname{erf} \left(\frac{|\mu_T|}{\sqrt{2}\sigma_T} \right) \right), \end{aligned} \tag{2.59}$$

with the domain of $\hat{X}_{t-n|t}$ being $\mathcal{A}_2 = \{-1, 1\}$, and where μ_T and σ_T are given in (2.57) and (2.58), respectively. The function $\operatorname{erf}(.)$ is the error function of (2.12).

Through the difference term $A_{t-k} - \hat{A}_{t-k|t-k}$ in (2.55), we observe the presence of simple errors of the form $X_{t-k-k_1} - \hat{X}_{t-k-k_1|t-k-k_1}$, and generalized errors of the form $X_{t-k-k_1} X_{t-k-k_2} - \hat{X}_{t-k-k_1|t-k-k_1} \hat{X}_{t-k-k_2|t-k-k_2}$. This is apparent from a comparison of matching portions of the equations in appendixes F and G. Note, however, that there is no corresponding difference term $B_{t-k} - \hat{B}_{t-k|t-k}$ in (2.55). This is one way in which the results for a general Volterra channel differ from the linear FIR channel models treated in [2, 3, 7–10, 24, 39–41, 52].

The result (2.59) is new, and was not discussed by Perreau *et al.* [52] during their introduction of the FLSDFE algorithm; nor later by Tsimbinos and White [61], when studying DFE error propagation on Volterra channels with the model (2.50). In this section we have extended the earlier results [2,3,7–10,24,39–41,52,61] to cover the incorporation of fixed–lag smoothing into a DFE on channels described by the Volterra input–output model (2.50), a generalization of the class of linear FIR models.

2.5 Conclusion

In this chapter we have provided an alternative and extended derivation of the signal processing algorithm introduced by Perreau *et al.* [52], which is called the FLSDFE algorithm in this thesis. In sections 2.3 and 2.4 we derived the specific FLSDFE formulae for two classes of nonlinear digital communication system model. In section 2.3 we derived the FLSDFE formulae that are applicable for MPSK and MQAM signalling on the Volterra channel model of Cheng and Powers [19]; and in section 2.4 we derived the specific FLSDFE formulae that are applicable for BPSK signalling on the Volterra channel model of Tsimbinos and White [61].

When introducing the FLSDFE algorithm, Perreau *et al.* [52] did not discuss the probability distribution of $\hat{X}_{t-n|t}$. In section 2.3.3 we showed that the derivation of the distribution function for the MPSK and MQAM signalling model is a difficult problem. There are three obstacles to this derivation. Firstly, from (2.30) and appendix D, we see the complicated nonlinear form of the dependence of the random variable $\hat{X}_{t-n|t}$ upon the symbols X_{t-n} and $X_{t-k-\delta}$, the estimators $\hat{X}_{t-k-\delta|t-k-\delta}$, and the additive noise variates V_{t-k}, where $k \in \{0,\ldots,n\}$, $n \in \{0,\ldots,N\}$ and $\delta \in \{0,\ldots,N\} \setminus \{n-k\}$. Secondly, there is the inherent complication of dealing with order statistics of dependent random variables that do not have a common probability density function. A third obstacle is the nonlinear mapping associated with the arg function in (2.30), which is a transformation from $\mathcal{A}_M^{2(N+n)+1} \times \mathbb{C}^{n+1}$ to $\mathcal{A}_M$, as $\hat{X}_{t-n|t}$ depends on $X_t,\ldots,X_{t-N-n}$, $\hat{X}_{t|t},\ldots,\hat{X}_{t-n+1|t-n+1}$, $\hat{X}_{t-n-1|t-n-1},\ldots,\hat{X}_{t-N-n|t-N-n}$ and $V_t,\ldots,V_{t-n}$. Accordingly, we leave the derivation of the conditional distribution function (2.33) as unsolved for the MPSK and MQAM model of section 2.3.

Conversely, in section 2.4.3, we provided in (2.59) the explicit distribution function of $\hat{X}_{t-n|t}$ in the simpler case of BPSK signalling.

The results of sections 2.3.3 and 2.4.3 on the distribution of $\hat{X}_{t-n|t}$ are new, extending earlier results [2,3,7–10,24,39–41,52,61] to cover the incorporation of fixed–lag smoothing into a DFE on channels described by the Volterra input–output models (2.15) and (2.50), generalizations of the class of linear FIR models.

Sections 2.3 and 2.4 may also be used as a guide for developing specific FLSDFE formulae for use with other nonlinear channel models, such as the following general rational model [18]:

$$Y_t = \frac{F_t\left(X_t, \ldots, X_{t-N}, Y_{t-1}, \ldots, Y_{t-N}, V_{t-1}, \ldots, V_{t-N}\right)}{G_t\left(X_t, \ldots, X_{t-N}, Y_{t-1}, \ldots, Y_{t-N}, V_{t-1}, \ldots, V_{t-N}\right)} + V_t, \tag{2.60}$$

where F_t and G_t are time–varying polynomial functions of the variates X_t, ..., X_{t-N}, Y_{t-1}, ..., Y_{t-N}, V_{t-1}, ..., V_{t-N}. This model can potentially represent a vast range of nonlinear SISO systems encountered in practice.

Chapter 3

State Space Models

Chapter 2 introduced the FLSDFE algorithm of Perreau *et al.* [52], and discussed the (static) probability distribution of FLSDFE output variable $\hat{X}_{t-n|t}$, via examples, in sections 2.3.3 and 2.4.3. In this chapter we turn now to an analysis of the *dynamics* of FLSDFE error propagation, drawing upon recent work on the related dynamics of DFE error propagation [1–3, 7–10, 16, 20–24, 28, 29, 37, 39–42, 45, 48, 51, 52, 61]. To this end, we construct state space models based on the theory of *finite state Markov processes* (FSMPs) [16, 28, 29, 39, 48, 50, 51].

Novel connections are drawn in this chapter between the state space models and various topics from pure mathematics, such as the Fibonacci series and some of its generalizations, Bell numbers, set partitions and the theory of integer partitions. Section 3.1 introduces some *atomic* state space models, which are the most 'fine–grained', and which describe the exact transient dynamics of error propagation. In section 3.2 the specific state space model for the case of FLSDFE operation in filtering–only mode ($N \geq 0$ and $n = 0$) is studied. Using these atomic models, interesting connections are made between error recovery times, Fibonacci series (and certain of their generalizations) and integer partitions. The analogous atomic state space model for the smoothing–only case ($N \geq 1$ and $n \in \{1, \ldots, N\}$) is studied in section 3.3.

Although the atomic models are exact, their size grows exponentially with channel memory N, and there is a need for more compact models. Section 3.4 discusses aggregated models, which have lower computational complexity. The search for an 'optimal' aggregation provides interesting and novel connections to the theory of set partitions, Bell numbers and restricted growth strings, discussed in sections 3.4.2 and 3.4.3. By *optimal* aggregation we mean one with the fewest states, but which exactly models the transient dynamics of the underlying atomic state space model [39]. In section 3.4.4 we give the unique optimal aggregate state space model for the atomic state space model described in section 3.2.5, which was for BPSK signalling on a linear FIR channel of memory $N = 1$.

From this unique optimum, a suboptimum aggregated model is derived (section 3.4.5), one with a relatively small number of aggregated states that also preserves the exact transient dynamics of the underlying atomic state space model. This suboptimal model is related to that of Choy and Beaulieu [24], yet theirs is not strictly a finite–state Markov process, and applies only to the steady–state limit. Finally, in section 3.4.6 we discuss Choy and Beaulieu's single–distinct–errors model, which provides a useful metric (a steady–state formula for error recovery time) for use in chapter 4, in connection with the discovery of 'resonances'.

3.1 Atomic State Space Models—Introduction

Sections 3.2 and 3.3 introduce some state space models that are useful in exactly modelling the transient dynamics of the input and output variables of the FLSDFE algorithm. The FLSDFE output $\hat{x}_{t|t}$ depends upon the input symbols $x_t, \ldots, x_{t-N}$, as well as the previous filtered symbol estimates $\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}$ and the random additive noise sample v_t. Thinking naïvely, it makes sense to incorporate *all* of the variables $x_t, \ldots, x_{t-N}, \hat{x}_{t|t}, \ldots, \hat{x}_{t-N|t-N}$ into one large state space vector, and track the stochastic dynamics of that vector over time. Likewise, a naïve state space model for the FLSDFE output $\hat{x}_{t-n|t}$, for $n \in \{1, \ldots, N\}$, can be constructed from a state vector containing all of the variates $x_t, \ldots, x_{t-N-n}, \hat{x}_{t|t}, \ldots, \hat{x}_{t-N-n|t-N-n}$ and $\hat{x}_{t-n|t}$, and tracking that vector over time, with stochastic variation driven by the additive noise process $\{v_t\}$. This is the essence of the *atomic* state space models discussed in this thesis, which were based on similar models of Kennedy and Anderson [39].

Section 3.2.1 introduces an atomic state space model that is useful for all channel memories $N \geq 0$ and for the case of filtering only ($n = 0$). We show that the model is a first–order Markov process, discuss its connection with a similar model of Kennedy and Anderson [39], and illustrate a lexicographical ordering system in section 3.2.2, based on the use of BPSK signalling. Whereas Kennedy and Anderson restricted their attention to FIR channels alone, we extend our focus to the Volterra models of section 2.4.

A novel formulation for the state transition probability matrix is subsequently developed in section 3.2.3, and illustrated with BPSK signalling on linear FIR channels of memory $N = 0$ and $N = 1$, in sections 3.2.4 and 3.2.5, respectively. The general formulation applies to the restricted case of BPSK signalling, but is applicable to Volterra

channels in general. Adapting the formulation to other signalling schemes, such as QPSK, would present no real difficulties.

No general form was derived for the stationary distribution vector, even for the simple case of BPSK signalling, although we illustrate the specific results in section 3.2.6 for the linear FIR examples of sections 3.2.4 and 3.2.5. We leave the development of the general form of the stationary distribution vector, for filtering–only operation of the FLSDFE, as an open problem.

Error recovery for the filtering–only state–space model is the subject of the remainder of section 3.2.

Section 3.2.7 gives the probability distribution of the recovery time R_0 for the case of BPSK signalling on a linear FIR channel of memory $N = 0$. We also present expressions for the mean and variance of R_0. These results are new.

Section 3.2.8 investigates the recovery time for the BPSK communications system modelled in section 3.2.5, which is for a linear FIR channel of memory $N = 1$. Using the atomic state space model directly, rather than the aggregated states of Choy and Beaulieu [24], we attempt to compute the explicit form of the probability distribution of the recovery time. In the process, we demonstrate evidence for an interesting connection with the Fibonacci sequence and a generalization of it, the Horadam (0,1,4,2) sequence. This connection is not proved but forms the subject of propositions 3.2.1 and 3.2.2. An explicit closed form of the probability distribution of the error recovery time random variable R_0 was not derived for the linear FIR, BPSK, $N = 1$ case, and we leave its derivation as an open problem.

Section 3.2.9 generalizes the example in section 3.2.8 to the case of BPSK signalling on a linear FIR channel of memory $N \geq 0$. We raise propositions as generalizations of the above–mentioned propositions 3.2.1 and 3.2.2, respectively. These, if true, connect the calculation of the probability distribution of R_0 to generalizations of the Fibonacci sequence. Note that Choy and Beaulieu give a formula for R_0 (used in chapter 4) but their result is based on their steady–state state space model, which is not an FSMP; the calculations in sections 3.2.8 and 3.2.9 are based directly upon the *transient* dynamics of the atomic state space models of this thesis. These transient results appear to be novel in the DFE error recovery literature.

In section 3.2.10 some interesting connections are drawn between the theory of integer partitions and the steady–state state space models of Choy and Beaulieu [24].

We illustrate a calculation of the probability of the error recovery time R_0 being exactly 10 time steps, using a communications system model involving BPSK signalling on a general Volterra channel of memory $N = 4$, and show how this calculation involves the enumeration of the constrained partitions of the numeral 5. Since the constrained partition number $P(n,k)$ has no general closed form for $k > 4$, it seems there may be no closed form for the probability distribution of R_0, in general, although Choy and Beaulieu give a formula for the *mean* of R_0. The connection with integer partitions appears not to have appeared before in the DFE error recovery time literature [8,9,20–22,24,39,41,51,52,61].

Section 3.3 parallels section 3.2 in the development of an atomic state space model suitable for use with a nonzero FLSDFE smoothing lag (the smoothing–only case), discussion of a lexicographical state–ordering system for BPSK signalling, and the development of the general form of the state transition matrix, covered in sections 3.3.1, 3.3.2 and 3.3.3, respectively. These results are new, and were not presented in the introductory work on the FLSDFE algorithm [52]. No analysis is provided of the smoothing–only state space model, as was done for the filtering–only model of section 3.2. We leave this as future work, and anticipate interesting stochastic dynamics, especially if smoothed FLSDFE outputs $\hat{x}_{t-n|t}$ are fed back.

3.2 Atomic State Space Models (Filtering only)

3.2.1 Model definition

For $N \geq 0$ and $n = 0$ the FLSDFE algorithm reduces to a DFE, and we have the filtered output $\hat{X}_{t|t}$. Based on a similar model used by Kennedy and Anderson [39], we introduce a filtering–only atomic state space model with state vector $\mathbf{S}_{t|t}$, where

$$\mathbf{S}_{t|t} = \begin{bmatrix} X_t \\ \vdots \\ X_{t-N} \\ \hat{X}_{t|t} \\ \vdots \\ \hat{X}_{t-N|t-N} \end{bmatrix}. \tag{3.1}$$

The state vectors in Kennedy and Anderson's model did not include the current symbol X_t or the corresponding estimate $\hat{X}_{t|t}$. These are included in (3.1) to give the complete set of variates upon which $\hat{X}_{t|t}$ depends, together with $\hat{X}_{t|t}$ itself.

Appendix H shows that for $m \in \mathbb{Z}^+$ we have the state transition probability

$$\begin{aligned}
&\mathbb{P}(\mathbf{S}_{t|t} = \mathbf{s}_{t|t} | \mathbf{S}_{t-1|t-1} = \mathbf{s}_{t-1|t-1} \cap \cdots \cap \mathbf{S}_{t-m|t-m} = \mathbf{s}_{t-m|t-m}) \\
&= \mathbb{P}(\mathbf{S}_{t|t} = \mathbf{s}_{t|t} | \mathbf{S}_{t-1|t-1} = \mathbf{s}_{t-1|t-1}) \\
&= \mathbb{P}(X_t = x_t) \mathbb{P}(\hat{X}_{t|t} = \hat{x}_{t|t} | X_t = x_t \cap \cdots \cap X_{t-N} = x_{t-N} \\
&\quad \cap \hat{X}_{t-1|t-1} = \hat{x}_{t-1|t-1} \cap \cdots \cap \hat{X}_{t-N|t-N} = \hat{x}_{t-N|t-N}).
\end{aligned} \tag{3.2}$$

The first two lines of (3.2) show that $\mathbf{S}_{t|t}$ is a first–order Markov process [50]. The term $\mathbb{P}(X_t = x_t)$ on the third line of (3.2) gives the *a priori* source probability of the symbol X_t; and the following term is the conditional probability of $\hat{X}_{t|t}$, given fixed values $x_t, \ldots, x_{t-N}$ and $\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}$ for all of the variables $X_t, \ldots, X_{t-N}$ and $\hat{X}_{t-1|t-1}, \ldots, \hat{X}_{t-N|t-N}$ upon which $\hat{X}_{t|t}$ depends (apart from the additive noise V_t).

3.2.2 Lexicographical ordering of states (BPSK, $N \geq 0$)

In order to write down the state transition probability matrix for the state space model (3.1)–(3.2), we first need to order the states $\mathbf{s}_{t|t}$. To illustrate, consider a lexicographical ordering that is suitable for BPSK signalling [39]. Using i for previous state $\mathbf{s}_{t-1|t-1}$ and j for current state $\mathbf{s}_{t|t}$, we have the explicit mapping rules

$$\begin{aligned}
i = &\frac{(x_{t-1}+1)}{2} 2^{2N+1} + \cdots + \frac{(x_{t-N}+1)}{2} 2^{N+2} + \frac{(x_{t-N-1}+1)}{2} 2^{N+1} \\
&+ \frac{(\hat{x}_{t-1|t-1}+1)}{2} 2^N + \cdots + \frac{(\hat{x}_{t-N|t-N}+1)}{2} 2^1 + \frac{(\hat{x}_{t-N-1|t-N-1}+1)}{2} 2^0,
\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}
j = &\frac{(x_t+1)}{2} 2^{2N+1} + \frac{(x_{t-1}+1)}{2} 2^{2N} + \cdots + \frac{(x_{t-N}+1)}{2} 2^{N+1} \\
&\frac{(\hat{x}_{t|t}+1)}{2} 2^N + \frac{(\hat{x}_{t-1|t-1}+1)}{2} 2^{N-1} + \cdots + \frac{(\hat{x}_{t-N|t-N}+1)}{2} 2^0.
\end{aligned} \tag{3.4}$$

Note that $(u+1)/2 \in \{0, 1\}$ for $u \in \{-1, 1\}$, so that (3.3) and (3.4) are simply binary–to–decimal conversions.

To illustrate, consider a channel of memory $N = 3$. We say that the state vector

$$\mathbf{s}_{t-1|t-1} = \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ x_{t-4} \\ \hat{x}_{t-1|t-1} \\ \hat{x}_{t-2|t-2} \\ \hat{x}_{t-3|t-3} \\ \hat{x}_{t-4|t-4} \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \tag{3.5}$$

is 'stored' in the 8–bit word i as shown in table 3.1, with bit 0 of i being the *least significant bit* (LSB) and bit 7 being the *most significant bit* (MSB).

Table 3.1. Illustration of storage of state vector $\mathbf{s}_{t-1|t-1}$ of (3.5) in the word i.

element of $\mathbf{s}_{t-1\|t-1}$	BPSK message symbol	bit position in word i	bit value
x_{t-1}	-1	7	0
x_{t-2}	1	6	1
x_{t-3}	1	5	1
x_{t-4}	-1	4	0
$\hat{x}_{t-1\|t-1}$	1	3	1
$\hat{x}_{t-2\|t-2}$	1	2	1
$\hat{x}_{t-3\|t-3}$	1	1	1
$\hat{x}_{t-4\|t-4}$	-1	0	0

Likewise, we say that the state vector

$$\mathbf{s}_{t|t} = \begin{bmatrix} x_{t} \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ \hat{x}_{t|t} \\ \hat{x}_{t-1|t-1} \\ \hat{x}_{t-2|t-2} \\ \hat{x}_{t-3|t-3} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{3.6}$$

is stored in word j, as shown in table 3.2.

Table 3.2. Illustration of storage of state vector $\mathbf{s}_{t|t}$ of (3.6) in the word j.

element of $\mathbf{s}_{t\|t}$	BPSK message symbol	bit position in word j	bit value
x_t	-1	7	0
x_{t-1}	-1	6	0
x_{t-2}	1	5	1
x_{t-3}	1	4	1
$\hat{x}_{t\|t}$	-1	3	0
$\hat{x}_{t-1\|t-1}$	1	2	1
$\hat{x}_{t-2\|t-2}$	1	1	1
$\hat{x}_{t-3\|t-3}$	1	0	1

The leftmost column in tables 3.1 and 3.2 gives the symbols 'stored' at the indicated bit positions of words i and j, respectively. The second column of each table shows the particular value taken by each element of $\mathbf{s}_{t-1|t-1}$ or $\mathbf{s}_{t|t}$, chosen from the BPSK alphabet $\mathcal{A}_2 = \{-1, 1\}$. The fourth column of each table shows how the symbol or estimate values map to the set of base–2 (binary) digits $\{0, 1\}$ through (3.3) and (3.4). For the example given above, we note that the decimal values of the binary words i and j evaluate to

$$\begin{aligned} i &= 0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 \\ &= 64 + 32 + 8 + 4 + 2 \\ &= 110, \quad \text{and} \end{aligned} \tag{3.7}$$

$$\begin{aligned} j &= 0 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 32 + 16 + 4 + 2 + 1 \\ &= 55. \end{aligned} \tag{3.8}$$

Observe that $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ in the above example have common elements x_{t-1}, x_{t-2}, x_{t-3}, $\hat{x}_{t-1|t-1}$, $\hat{x}_{t-2|t-2}$ and $\hat{x}_{t-3|t-3}$. The only valid state transitions $\mathbf{s}_{t-1|t-1}$ to $\mathbf{s}_{t|t}$, therefore, are those that preserve the values of these shared quantities. Transitions that do not satisfy this condition are impossible, and they result in structural zeros in the state transition probability matrix.

For general $N \geq 0$, the BPSK lexicographical ordering system in (3.3) and (3.4) is illustrated in table 3.3 below.

Table 3.3. Storage of consecutive state vectors $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ in words i and j, for the case of BPSK signalling, arbitrary channel memory $N \geq 0$, and filtering only ($n = 0$).

bit position in i	element in $\mathbf{s}_{t-1\|t-1}$	bit position in j	element in $\mathbf{s}_{t\|t}$
$2N+1$	x_{t-1}	$2N+1$	x_t
$\vdots$	$\vdots$	$2N$	x_{t-1}
$N+2$	x_{t-N}	$\vdots$	$\vdots$
$N+1$	x_{t-N-1}	$N+1$	x_{t-N}
N	$\hat{x}_{t-1\|t-1}$	N	$\hat{x}_{t\|t}$
$\vdots$	$\vdots$	$N-1$	$\hat{x}_{t-1\|t-1}$
1	$\hat{x}_{t-N\|t-N}$	$\vdots$	$\vdots$
0	$\hat{x}_{t-N-1\|t-N-1}$	0	$\hat{x}_{t-N\|t-N}$

The oldest elements in $\mathbf{s}_{t-1|t-1}$, namely x_{t-N-1} and $\hat{x}_{t-N-1|t-N-1}$, are contained within a cell that has a blue background in table 3.3, and occupy bit positions $N+1$ and 0 in word i, respectively. (For brevity throughout this thesis, we will shorten phrases such as 'contained within a cell that has a blue background' to something simpler, such as 'coloured blue'.) The elements in common between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, namely $x_{t-1}, \ldots, x_{t-N}$, $\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}$, are coloured grey in both words i and j. In word i, these common elements occupy bit positions $2N+1, \ldots, N+2, N, \ldots, 1$ respectively. In word j the elements have been shifted to the new positions $2N, \ldots, N+1, N-1, \ldots, 0$ respectively, overwriting the old elements x_{t-N-1} and $\hat{x}_{t-N-1|t-N-1}$. Finally, new elements x_t and $\hat{x}_{t|t}$ are placed into positions $2N+1$ and N of $\mathbf{s}_{t|t}$, respectively, shown coloured yellow in table 3.3.

3.2.3 State transition probability matrix (BPSK, $N \geq 0$)

In this section we give the state transition probability matrix P for the simple case of filtering only and BPSK signalling. We will use the lexicographical ordering scheme of (3.3) and (3.4), for which P has the sparse, hierarchical structure given below. The general form of P that follows does not appear to have been reported previously in the literature on the statics and dynamics of (FLS)DFE error propagation [1–3,7–10,16,20–24,28,29,34,37,39–42,45,51,52,61].

From (3.1) we see that for BPSK signalling there are 4^{N+1} atomic states $\mathbf{s}_{t|t}$, so that P will be of size $4^{N+1} \times 4^{N+1}$. Label the rows and columns of P with the indices i and j

of (3.3) and (3.4). Let the top left element of P have row and column indices $i = 0$ and $j = 0$, respectively; and let the bottom right element of P have row and column indices $i = 4^{N+1} - 1$ and $j = 4^{N+1} - 1$, respectively. In general, the element in row $i + 1$ and column $j + 1$ has row and column indices i and j, respectively. Thus P is of the form $P = (p_{i,j})$, with elements $p_{i,j}$, where $i, j \in \{0, \ldots, 4^{N+1} - 1\}$:

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,\zeta} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,\zeta} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\zeta,0} & p_{\zeta,1} & \cdots & p_{\zeta,\zeta} \end{bmatrix}, \quad \zeta = 4^{N+1} - 1. \tag{3.9}$$

From table 3.3, observe that x_t is stored in the most significant bit (MSB) of j, where j is the column index of P. With $j \in \{0, \ldots, 4^{N+1} - 1\}$, the MSB of j will be 0 for $j \in \{0, \ldots, 2^{2N+1} - 1\}$, and 1 for $j \in \{2^{2N+1}, \ldots, 4^{N+1} - 1\}$. That is, $x_t = -1$ in the left half of P, and $x_t = 1$ in the right half of P, and so we have a natural division of P into two equal–sized halves:

$$P = \begin{bmatrix} P_A & P_B \end{bmatrix}, \tag{3.10}$$

where P_A and P_B contain the transition probabilities in which $x_t = -1$ and $x_t = 1$, respectively. Note that P_A and P_B are both of size $4^{N+1} \times 2^{2N+1}$.

Further, we can partition P_A and P_B by observing that $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ have the common elements $x_{t-1}, \ldots, x_{t-N}$. This is shown clearly in table 3.3. Observe that the common elements $x_{t-1}, \ldots, x_{t-N}$ are stored in the N most significant bits of row index i, whereas they are stored in the N most significant bits *following* the MSB of column index j. Taking into account the 2^N possible values of the N–tuple $(x_{t-1}, \ldots, x_{t-N})$, and observing the requirement to have matching N–tuples $(x_{t-1}, \ldots, x_{t-N})$ between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, we find that P_A and P_B are sparse, with the block matrix forms

$$P_A = \begin{bmatrix} P_A^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_A^{(1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_A^{(\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_A^{(\zeta)} \end{bmatrix}, \tag{3.11}$$

and

$$P_B = \begin{bmatrix} P_B^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_B^{(1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_B^{(\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_B^{(\zeta)} \end{bmatrix}, \tag{3.12}$$

with the abbreviation

$$\zeta = 2^N - 1, \tag{3.13}$$

where $\mathbf{0}$ is a matrix of structural zeros, of size $2^{N+2} \times 2^{N+1}$. Each block $P_A^{(k)}$ and $P_B^{(k)}$ in (3.11) and (3.12) is also of size $2^{N+2} \times 2^{N+1}$, where $k \in \{0, \ldots, 2^N - 1\}$.

A further decomposition of each $P_A^{(k)}$ and $P_B^{(k)}$ results from observing that the 'old' element x_{t-N-1} of $\mathbf{s}_{t-1|t-1}$ is redundant, since the state transition probability (3.2) does not depend on it. Note from table 3.3 that x_{t-N-1} is stored in bit position $N+1$ of $\mathbf{s}_{t-1|t-1}$ and does not appear in $\mathbf{s}_{t|t}$. Since there is the same transition probability irrespective of the value of x_{t-N-1}, each $2^{N+2} \times 2^{N+1}$ matrix $P_A^{(k)}$ and $P_B^{(k)}$ in (3.11) and (3.12) splits horizontally into two identical matrices of size $2^{N+1} \times 2^{N+1}$, so that for $k \in \{0, \ldots, 2^N - 1\}$ we have

$$P_A^{(k)} = \begin{bmatrix} P_C^{(k)} \\ P_C^{(k)} \end{bmatrix}, \quad \text{and} \tag{3.14}$$

$$P_B^{(k)} = \begin{bmatrix} P_D^{(k)} \\ P_D^{(k)} \end{bmatrix}. \tag{3.15}$$

For a fixed $k \in \{0, \ldots, 2^N - 1\}$ we have the same fixed N–tuple $(x_{t-1}, \ldots, x_{t-N})$ in both (3.14) and (3.15), as discussed earlier. Moreover, the upper matrix of each pair in the right–hand side of both (3.14) and (3.15) corresponds to the additional choice $x_{t-N-1} = -1$, while the lower pair is for $x_{t-N-1} = 1$. This is in accordance with the bit position of x_{t-N-1} in row index word i.

Proceeding with the decomposition of P, note from table 3.3 that for each fixed value of symbol x_t, N–tuple $(x_{t-1}, \ldots, x_{t-N})$ and symbol x_{t-N-1}, we have a choice of two values of estimate $\hat{x}_{t|t}$. Observe that $\hat{x}_{t|t}$ is stored in bit position N of j.

For $k \in \{0, \ldots, 2^N - 1\}$ we can therefore vertically divide $P_C^{(k)}$ and $P_D^{(k)}$ of (3.14) and (3.15), analogously to the partitioning of P in (3.10):

$$P_C^{(k)} = \begin{bmatrix} P_E^{(k)} & P_F^{(k)} \end{bmatrix} \quad \text{and} \tag{3.16}$$

$$P_D^{(k)} = \begin{bmatrix} P_G^{(k)} & P_H^{(k)} \end{bmatrix}, \tag{3.17}$$

where the left and right matrices in each pair correspond to the choices $\hat{x}_{t|t} = -1$ and $\hat{x}_{t|t} = 1$, respectively, and where $P_E^{(k)}$, $P_F^{(k)}$, $P_G^{(k)}$ and $P_H^{(k)}$ are each of size $2^{N+1} \times 2^N$. From (3.16) and (3.17), we now expand (3.14) and (3.15) to give

$$P_A^{(k)} = \begin{bmatrix} P_E^{(k)} & P_F^{(k)} \\ P_E^{(k)} & P_F^{(k)} \end{bmatrix} \quad \text{and} \tag{3.18}$$

$$P_B^{(k)} = \begin{bmatrix} P_G^{(k)} & P_H^{(k)} \\ P_G^{(k)} & P_H^{(k)} \end{bmatrix}, \tag{3.19}$$

where $k \in \{0, \ldots, 2^N - 1\}$.

Continuing, note from table 3.3 the occurrence of common elements $\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}$ in $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$. Taking into account the 2^N possible values of the N–tuple

$$\left(\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}\right),$$

and observing the requirement to have matching N–tuples $\left(\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}\right)$ between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, we find that for each $k \in \{0, \ldots, 2^N - 1\}$, the blocks $P_E^{(k)}$, $P_F^{(k)}$, $P_G^{(k)}$ and $P_H^{(k)}$ in (3.18) and (3.19) are themselves sparse, and of the form

$$P_E^{(k)} = \begin{bmatrix} P_E^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_E^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_E^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_E^{(k,\zeta)} \end{bmatrix}, \tag{3.20}$$

$$P_F^{(k)} = \begin{bmatrix} P_F^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_F^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_F^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_F^{(k,\zeta)} \end{bmatrix}, \tag{3.21}$$

$$P_G^{(k)} = \begin{bmatrix} P_G^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_G^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_G^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_G^{(k,\zeta)} \end{bmatrix}, \quad \text{and} \tag{3.22}$$

$$P_H^{(k)} = \begin{bmatrix} P_H^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_H^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_H^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_H^{(k,\zeta)} \end{bmatrix}, \tag{3.23}$$

with $\zeta = 2^N - 1$, where each block matrix in (3.20)–(3.23) is now of size 2×1.

The final step in the decomposition of P results from observing that the 'old' element $\hat{x}_{t-N-1|t-N-1}$ of $\mathbf{s}_{t-1|t-1}$ is redundant, since the state transition probability (3.2) does not depend on it. Note from table 3.3 that $\hat{x}_{t-N-1|t-N-1}$ is stored in the least significant bit position 0 of $\mathbf{s}_{t-1|t-1}$ and does not appear in $\mathbf{s}_{t|t}$. Since there is the same transition probability no matter the value of $\hat{x}_{t-N-1|t-N-1}$, each 2×1 matrix $P_E^{(k,l)}$, $P_F^{(k,l)}$, $P_G^{(k,l)}$ and $P_H^{(k,l)}$ in (3.20)–(3.23) splits horizontally into two identical scalars, where $k \in \{0, \ldots, 2^N - 1\}$ and $l \in \{0, \ldots, 2^N - 1\}$. With this in mind, for $k \in \{0, \ldots, 2^N - 1\}$ rewrite (3.20)–(3.23) as follows:

$$P_E^{(k)} = \begin{bmatrix} p_E^{(k,0)} & 0 & \cdots & 0 & 0 \\ p_E^{(k,0)} & 0 & \cdots & 0 & 0 \\ 0 & p_E^{(k,1)} & \cdots & 0 & 0 \\ 0 & p_E^{(k,1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_E^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & p_E^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & 0 & p_E^{(k,\zeta)} \\ 0 & 0 & \cdots & 0 & p_E^{(k,\zeta)} \end{bmatrix}, \tag{3.24}$$

$$P_F^{(k)} = \begin{bmatrix} p_F^{(k,0)} & 0 & \cdots & 0 & 0 \\ p_F^{(k,0)} & 0 & \cdots & 0 & 0 \\ 0 & p_F^{(k,1)} & \cdots & 0 & 0 \\ 0 & p_F^{(k,1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_F^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & p_F^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & 0 & p_F^{(k,\zeta)} \\ 0 & 0 & \cdots & 0 & p_F^{(k,\zeta)} \end{bmatrix}, \tag{3.25}$$

$$P_G^{(k)} = \begin{bmatrix} p_G^{(k,0)} & 0 & \cdots & 0 & 0 \\ p_G^{(k,0)} & 0 & \cdots & 0 & 0 \\ 0 & p_G^{(k,1)} & \cdots & 0 & 0 \\ 0 & p_G^{(k,1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_G^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & p_G^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & 0 & p_G^{(k,\zeta)} \\ 0 & 0 & \cdots & 0 & p_G^{(k,\zeta)} \end{bmatrix}, \quad \text{and} \tag{3.26}$$

$$P_H^{(k)} = \begin{bmatrix} p_H^{(k,0)} & 0 & \cdots & 0 & 0 \\ p_H^{(k,0)} & 0 & \cdots & 0 & 0 \\ 0 & p_H^{(k,1)} & \cdots & 0 & 0 \\ 0 & p_H^{(k,1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & p_H^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & p_H^{(k,\zeta-1)} & 0 \\ 0 & 0 & \cdots & 0 & p_H^{(k,\zeta)} \\ 0 & 0 & \cdots & 0 & p_H^{(k,\zeta)} \end{bmatrix}, \tag{3.27}$$

with $\zeta = 2^N - 1$, where each element is now a scalar transition probability.

Since P is a stochastic matrix, we have the requirements

$$0 \leq p_E^{(k,l)} \leq 1, \tag{3.28}$$

$$0 \leq p_F^{(k,l)} \leq 1, \tag{3.29}$$

$$0 \leq p_G^{(k,l)} \leq 1, \tag{3.30}$$

$$0 \leq p_H^{(k,l)} \leq 1, \quad \text{and} \tag{3.31}$$

$$p_E^{(k,l)} + p_F^{(k,l)} + p_G^{(k,l)} + p_H^{(k,l)} = 1, \tag{3.32}$$

where $k, l \in \{0, \ldots, 2^N - 1\}$.

To recap, the state transition probability matrix P for the filtering–only atomic state space model in (3.1) and (3.2) is given by (3.10), where P_A and P_B are given by (3.11) and (3.12), with the further decompositions in (3.18), (3.19) and (3.24)–(3.27). This decomposition appears to be novel in the (FLS)DFE literature.

Two simple examples of these results will now be given, to illustrate the explicit form of P. These examples will be referred to later in this chapter. Both involve BPSK signalling

over a linear FIR channel. The first example problem involves a memoryless channel, with $N = 0$; the second involves a channel of minimum memory $N = 1$. Other cases could have been considered as examples, such as channels of longer memory, or QPSK signalling over a nonlinear Volterra channel, but the algebraic manipulations would be more involved and not necessarily more informative.

3.2.4 State transition probability matrix (BPSK, $N = 0$)

For the first example, with zero memory, we have from (2.50) the channel model

$$Y_t = h_1(0)X_t + V_t; \tag{3.33}$$

from (2.53) we get the DFE estimator

$$\hat{X}_{t|t} = \text{sign}\left\{h_1(0)Y_t\right\}; \tag{3.34}$$

and from (3.2) and (2.59) we get the state transition probability

$$\mathbb{P}(\mathbf{S}_{t|t} = \mathbf{s}_{t|t}|\mathbf{S}_{t-1|t-1} = \mathbf{s}_{t-1|t-1}) = \frac{1}{4}\left(1 + \hat{x}_{t|t}\,\text{sign}(\mu_T)\,\text{erf}\left(\frac{|\mu_T|}{\sqrt{2}\sigma_T}\right)\right), \tag{3.35}$$

assuming equiprobable source symbols X_t, that is, $\mathbb{P}(X_t = x_t) = \frac{1}{2}$ for $x_t \in \{-1, 1\}$.

In (3.35), μ_T and σ_T are given from (2.57) and (2.58) by

$$\mu_T = h_1^2(0)x_t, \quad \text{and} \tag{3.36}$$

$$\sigma_T^2 = h_1^2(0)\sigma_v^2, \tag{3.37}$$

where σ_v^2 is the variance of V_t in the channel model (3.33).

The atomic state vector $\mathbf{S}_{t|t}$ is given from (3.1) as the 2–vector

$$\mathbf{S}_{t|t} = \begin{bmatrix} X_t \\ \hat{X}_{t|t} \end{bmatrix}. \tag{3.38}$$

With this model, the mapping rules (3.3) and (3.4) for $N = 0$ are

$$i = \frac{(x_{t-1} + 1)}{2}2^1 + \frac{(\hat{x}_{t-1|t-1} + 1)}{2}2^0, \tag{3.39}$$

and

$$j = \frac{(x_t + 1)}{2}2^1 + \frac{(\hat{x}_{t|t} + 1)}{2}2^0, \tag{3.40}$$

as illustrated in table 3.4 below, which is a specialization of table 3.3.

Table 3.4. Storage of consecutive state vectors $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ in words i and j, for the case of BPSK signalling, channel memory $N = 0$ (memoryless), and FLSDFE smoothing lag $n = 0$ (filtering only).

bit position in i	element in $\mathbf{s}_{t-1\|t-1}$	bit position in j	element in $\mathbf{s}_{t\|t}$
1	x_{t-1}	1	x_t
0	$\hat{x}_{t-1\|t-1}$	0	$\hat{x}_{t\|t}$

At time $t-1$ there are $2^2 = 4$ state vectors $\mathbf{s}_{t-1|t-1}$, indexed by $i \in \{0,1,2,3\}$. Similarly, at time t, there are 4 state vectors $\mathbf{s}_{t|t}$, indexed by $j \in \{0,1,2,3\}$. We note that there are no common elements between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, and so P will contain no structural zeros, unlike the general case discussed above.

Rows of P are referenced sequentially by $i \in \{0,1,2,3\}$, with the top row having index $i = 0$, while the bottom row has index $i = 3$. Likewise, columns of P are referenced sequentially by $j \in \{0,1,2,3\}$, with the leftmost column having index $j = 0$, while the rightmost column has index $j = 3$.

From (3.10) we have $P = [P_A \; P_B]$, where the 4×2 matrices P_A and P_B are decomposed in (3.11) and (3.12) as

$$P_A = \left[P_A^{(0)}\right], \tag{3.41}$$

and

$$P_B = \left[P_B^{(0)}\right]. \tag{3.42}$$

Further, $P_A^{(0)}$ and $P_B^{(0)}$ are given from (3.18) and (3.19) by

$$P_A = \begin{bmatrix} P_E^{(0)} & P_F^{(0)} \\ P_E^{(0)} & P_F^{(0)} \end{bmatrix} \quad \text{and} \tag{3.43}$$

$$P_B = \begin{bmatrix} P_G^{(0)} & P_H^{(0)} \\ P_G^{(0)} & P_H^{(0)} \end{bmatrix}, \tag{3.44}$$

where each submatrix $P_E^{(0)}, \ldots, P_H^{(0)}$ is of size 2×1. From (3.24)–(3.27) and the model (3.33)–(3.40) we have $P_E^{(0)}, \ldots, P_H^{(0)}$ given explicitly by

$$P_E^{(0)} = \frac{1}{4}\begin{bmatrix} 1+a \\ 1+a \end{bmatrix}, \tag{3.45}$$

$$P_F^{(0)} = \frac{1}{4}\begin{bmatrix} 1-a \\ 1-a \end{bmatrix}, \tag{3.46}$$

$$P_G^{(0)} = \frac{1}{4}\begin{bmatrix} 1-a \\ 1-a \end{bmatrix}, \quad \text{and} \tag{3.47}$$

$$P_H^{(0)} = \frac{1}{4}\begin{bmatrix} 1+a \\ 1+a \end{bmatrix}, \tag{3.48}$$

with

$$a = \operatorname{erf}\left(\frac{|h_1(0)|}{\sqrt{2}\sigma_v}\right). \tag{3.49}$$

For the simple case of $N = 0$, we observe from the above formulae that P is composed of identical rows, since the transition probability (3.35) is independent of the previous state $\mathbf{s}_{t-1|t-1}$. Indeed, putting the pieces together, we have

$$P = \frac{1}{4}\begin{bmatrix} 1+a & 1-a & 1-a & 1+a \\ 1+a & 1-a & 1-a & 1+a \\ 1+a & 1-a & 1-a & 1+a \\ 1+a & 1-a & 1-a & 1+a \end{bmatrix}. \tag{3.50}$$

3.2.5 State transition probability matrix (BPSK, $N = 1$)

The second example problem is for a channel memory $N = 1$. From (2.50) we have the channel model [52]

$$Y_t = h_1(0)X_t + h_1(1)X_{t-1} + V_t; \tag{3.51}$$

from (2.53) we get the DFE estimator

$$\hat{X}_{t|t} = \operatorname{sign}\left\{\left(Y_t - h_1(1)\hat{X}_{t-1|t-1}\right)h_1(0)\right\}; \tag{3.52}$$

and from (3.2) and (2.59) we get the state transition probability

$$\mathbb{P}(\mathbf{S}_{t|t} = \mathbf{s}_{t|t}|\mathbf{S}_{t-1|t-1} = \mathbf{s}_{t-1|t-1}) = \frac{1}{4}\left(1 + \hat{x}_{t|t}\operatorname{sign}(\mu_T)\operatorname{erf}\left(\frac{|\mu_T|}{\sqrt{2}\sigma_T}\right)\right), \tag{3.53}$$

assuming equiprobable source symbols X_t, that is, $\mathbb{P}(X_t = x_t) = \frac{1}{2}$ for $x_t \in \{-1, 1\}$.

In (3.53), μ_T and σ_T are given from (2.57) and (2.58) by

$$\mu_T = h_1^2(0)x_t + h_1(0)h_1(1)\left(x_{t-1} - \hat{x}_{t-1|t-1}\right), \quad \text{and} \tag{3.54}$$

$$\sigma_T^2 = h_1^2(0)\sigma_v^2, \tag{3.55}$$

where σ_v^2 is the variance of V_t in the channel model (3.51).

The atomic state vector $\mathbf{S}_{t|t}$ is given from (3.1) as the 4–vector

$$\mathbf{S}_{t|t} = \begin{bmatrix} X_t \\ X_{t-1} \\ \hat{X}_{t|t} \\ \hat{X}_{t-1|t-1} \end{bmatrix}. \tag{3.56}$$

With this model, the mapping rules (3.3) and (3.4) for $N = 1$ are

$$i = \frac{(x_{t-1}+1)}{2}2^3 + \frac{(x_{t-2}+1)}{2}2^2 + \frac{(\hat{x}_{t-1|t-1}+1)}{2}2^1 + \frac{(\hat{x}_{t-2|t-2}+1)}{2}2^0, \tag{3.57}$$

and

$$j = \frac{(x_t+1)}{2}2^3 + \frac{(x_{t-1}+1)}{2}2^2 + \frac{(\hat{x}_{t|t}+1)}{2}2^1 + \frac{(\hat{x}_{t-1|t-1}+1)}{2}2^0, \tag{3.58}$$

as illustrated in table 3.5 below, a specialization of table 3.3.

Table 3.5. Storage of consecutive state vectors $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ in words i and j, for the case of BPSK signalling, channel memory $N = 1$, and FLSDFE smoothing lag $n = 0$.

bit position in i	element in $\mathbf{s}_{t-1\|t-1}$	bit position in j	element in $\mathbf{s}_{t\|t}$
3	x_{t-1}	3	x_t
2	x_{t-2}	2	x_{t-1}
1	$\hat{x}_{t-1\|t-1}$	1	$\hat{x}_{t\|t}$
0	$\hat{x}_{t-2\|t-2}$	0	$\hat{x}_{t-1\|t-1}$

At time $t-1$ there are $2^4 = 16$ state vectors $\mathbf{s}_{t-1|t-1}$, indexed by $i \in \{0, \ldots, 15\}$. Similarly, at time t, there are 16 state vectors $\mathbf{s}_{t|t}$, indexed by $j \in \{0, \ldots, 15\}$. We note from table 3.5 that there are two shared values between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, namely, x_{t-1} and $\hat{x}_{t-1|t-1}$, and this commonality means P will be sparse, with structural zeros.

Rows of P are referenced sequentially by $i \in \{0, \ldots, 15\}$, with the top row having index $i = 0$, while the bottom row has index $i = 15$. Likewise, columns of P are referenced

sequentially by $j \in \{0, \ldots, 15\}$, with the leftmost column having index $j = 0$, while the rightmost column has index $j = 15$.

From (3.10) we have $P = [P_A \; P_B]$, where P_A and P_B are given below:

$$P_A = \begin{bmatrix} P_E^{(0)} & P_F^{(0)} & \mathbf{0} & \mathbf{0} \\ P_E^{(0)} & P_F^{(0)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_E^{(1)} & P_F^{(1)} \\ \mathbf{0} & \mathbf{0} & P_E^{(1)} & P_F^{(1)} \end{bmatrix} \quad \text{and} \tag{3.59}$$

$$P_B = \begin{bmatrix} P_G^{(0)} & P_H^{(0)} & \mathbf{0} & \mathbf{0} \\ P_G^{(0)} & P_H^{(0)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_G^{(1)} & P_H^{(1)} \\ \mathbf{0} & \mathbf{0} & P_G^{(1)} & P_H^{(1)} \end{bmatrix}. \tag{3.60}$$

Each submatrix $P_E^{(0)}, \ldots, P_H^{(1)}$ and $\mathbf{0}$ is of size 4×2, with $\mathbf{0}$ being a matrix of structural zeros. From (3.24)–(3.27) and the model (3.51)–(3.58) we have $P_E^{(0)}, \ldots, P_H^{(1)}$ given explicitly by

$$P_E^{(0)} = \frac{1}{4} \begin{bmatrix} 1+a & 0 \\ 1+a & 0 \\ 0 & 1+b \\ 0 & 1+b \end{bmatrix}, \tag{3.61}$$

$$P_F^{(0)} = \frac{1}{4} \begin{bmatrix} 1-a & 0 \\ 1-a & 0 \\ 0 & 1-b \\ 0 & 1-b \end{bmatrix}, \tag{3.62}$$

$$P_E^{(1)} = \frac{1}{4} \begin{bmatrix} 1+c & 0 \\ 1+c & 0 \\ 0 & 1+a \\ 0 & 1+a \end{bmatrix}, \tag{3.63}$$

$$P_F^{(1)} = \frac{1}{4} \begin{bmatrix} 1-c & 0 \\ 1-c & 0 \\ 0 & 1-a \\ 0 & 1-a \end{bmatrix}, \tag{3.64}$$

$$P_G^{(0)} = \frac{1}{4}\begin{bmatrix} 1-a & 0 \\ 1-a & 0 \\ 0 & 1-c \\ 0 & 1-c \end{bmatrix}, \tag{3.65}$$

$$P_H^{(0)} = \frac{1}{4}\begin{bmatrix} 1+a & 0 \\ 1+a & 0 \\ 0 & 1+c \\ 0 & 1+c \end{bmatrix}, \tag{3.66}$$

$$P_G^{(1)} = \frac{1}{4}\begin{bmatrix} 1-b & 0 \\ 1-b & 0 \\ 0 & 1-a \\ 0 & 1-a \end{bmatrix}, \quad \text{and} \tag{3.67}$$

$$P_H^{(1)} = \frac{1}{4}\begin{bmatrix} 1+b & 0 \\ 1+b & 0 \\ 0 & 1+a \\ 0 & 1+a \end{bmatrix}, \tag{3.68}$$

where a was given earlier in (3.49), and where b and c are the constants

$$b = \text{sign}\left\{h_1(0)\left(h_1(0)+2h_1(1)\right)\right\} \text{erf}\left(\frac{|h_1(0)+2h_1(1)|}{\sqrt{2}\sigma_v}\right), \quad \text{and} \tag{3.69}$$

$$c = \text{sign}\left\{h_1(0)\left(h_1(0)-2h_1(1)\right)\right\} \text{erf}\left(\frac{|h_1(0)-2h_1(1)|}{\sqrt{2}\sigma_v}\right). \tag{3.70}$$

Introduce the abbreviations

$$a^+ = \frac{1}{4}(1+a), \tag{3.71}$$

$$a^- = \frac{1}{4}(1-a), \tag{3.72}$$

$$b^+ = \frac{1}{4}(1+b), \tag{3.73}$$

$$b^- = \frac{1}{4}(1-b), \tag{3.74}$$

$$c^+ = \frac{1}{4}(1+c), \quad \text{and} \tag{3.75}$$

$$c^- = \frac{1}{4}(1-c). \tag{3.76}$$

Then for the case $N = 1$ we have P given explicitly as follows, putting together the pieces in (3.10), (3.49) and (3.59)–(3.76):

$$P = \begin{bmatrix}
a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ & 0 & 0 & 0 & 0 & 0 \\
a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ & 0 & 0 & 0 & 0 & 0 \\
0 & b^+ & 0 & b^- & 0 & 0 & 0 & 0 & 0 & c^- & 0 & c^+ & 0 & 0 & 0 & 0 \\
0 & b^+ & 0 & b^- & 0 & 0 & 0 & 0 & 0 & c^- & 0 & c^+ & 0 & 0 & 0 & 0 \\
a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ & 0 & 0 & 0 & 0 & 0 \\
a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ & 0 & 0 & 0 & 0 & 0 \\
0 & b^+ & 0 & b^- & 0 & 0 & 0 & 0 & 0 & c^- & 0 & c^+ & 0 & 0 & 0 & 0 \\
0 & b^+ & 0 & b^- & 0 & 0 & 0 & 0 & 0 & c^- & 0 & c^+ & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c^+ & 0 & c^- & 0 & 0 & 0 & 0 & 0 & b^- & 0 & b^+ & 0 \\
0 & 0 & 0 & 0 & c^+ & 0 & c^- & 0 & 0 & 0 & 0 & 0 & b^- & 0 & b^+ & 0 \\
0 & 0 & 0 & 0 & 0 & a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ \\
0 & 0 & 0 & 0 & 0 & a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ \\
0 & 0 & 0 & 0 & c^+ & 0 & c^- & 0 & 0 & 0 & 0 & 0 & b^- & 0 & b^+ & 0 \\
0 & 0 & 0 & 0 & c^+ & 0 & c^- & 0 & 0 & 0 & 0 & 0 & b^- & 0 & b^+ & 0 \\
0 & 0 & 0 & 0 & 0 & a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+ \\
0 & 0 & 0 & 0 & 0 & a^+ & 0 & a^- & 0 & 0 & 0 & 0 & 0 & a^- & 0 & a^+
\end{bmatrix}. \tag{3.77}$$

3.2.6 Stationary distribution vector

A problem of interest in state space modelling is the determination of the vector π_∞ of limiting state probabilities, which is the solution to the matrix–vector equation [26, 50]

$$P^T \pi_\infty = \pi_\infty, \tag{3.78}$$

where P is the state transition probability matrix. No general solution to (3.78) was found for the filtering–only atomic state space model, due to the intricate structure of P. We illustrate explicit solutions to the simple cases discussed above, though.

For the memoryless example, with $N = 0$, we have from (3.50) and (3.78) the matrix–vector equation

$$\frac{1}{4}\begin{bmatrix}
1+a & 1+a & 1+a & 1+a \\
1-a & 1-a & 1-a & 1-a \\
1-a & 1-a & 1-a & 1-a \\
1+a & 1+a & 1+a & 1+a
\end{bmatrix}\begin{bmatrix}
\pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3
\end{bmatrix} = \begin{bmatrix}
\pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3
\end{bmatrix}, \tag{3.79}$$

with the solution

$$\pi_\infty = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1+a \\ 1-a \\ 1-a \\ 1+a \end{bmatrix}, \tag{3.80}$$

where a is given in (3.49). For $j \in \{0,1,2,3\}$, π_j in (3.80) is the asymptotic probability of being in state j, where the state–to–integer mapping (3.40) is detailed in table 3.6 below.

Table 3.6. Mapping of state vector $\mathbf{s}_{t|t}$ to state index j, for $N = 0$ and $n = 0$.

elements of $\mathbf{s}_{t\|t}$		state index
x_t	$\hat{x}_{t\|t}$	j
-1	-1	0
-1	1	1
1	-1	2
1	1	3

From this we obtain, for example, the asymptotic probability of an error as

$$P_e = \pi_1 + \pi_2 = \frac{1}{2}(1-a) = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{|h_1(0)|}{\sqrt{2}\sigma_v}\right)\right). \tag{3.81}$$

For the second example problem, with $N = 1$, from (3.71)–(3.78) we have the following set of linear equations in the unknowns $\pi_0, \ldots, \pi_{15}$, the components of π_∞:

$$a^+(\pi_0 + \pi_1 + \pi_4 + \pi_5) = \pi_0, \tag{3.82}$$
$$b^+(\pi_2 + \pi_3 + \pi_6 + \pi_7) = \pi_1, \tag{3.83}$$
$$a^-(\pi_0 + \pi_1 + \pi_4 + \pi_5) = \pi_2, \tag{3.84}$$
$$b^-(\pi_2 + \pi_3 + \pi_6 + \pi_7) = \pi_3, \tag{3.85}$$
$$c^+(\pi_8 + \pi_9 + \pi_{12} + \pi_{13}) = \pi_4, \tag{3.86}$$
$$a^+(\pi_{10} + \pi_{11} + \pi_{14} + \pi_{15}) = \pi_5, \tag{3.87}$$
$$c^-(\pi_8 + \pi_9 + \pi_{12} + \pi_{13}) = \pi_6, \tag{3.88}$$
$$a^-(\pi_{10} + \pi_{11} + \pi_{14} + \pi_{15}) = \pi_7, \tag{3.89}$$
$$a^-(\pi_0 + \pi_1 + \pi_4 + \pi_5) = \pi_8, \tag{3.90}$$

$$c^{-}\left(\pi_2+\pi_3+\pi_6+\pi_7\right)=\pi_9, \tag{3.91}$$
$$a^{+}\left(\pi_0+\pi_1+\pi_4+\pi_5\right)=\pi_{10}, \tag{3.92}$$
$$c^{+}\left(\pi_2+\pi_3+\pi_6+\pi_7\right)=\pi_{11}, \tag{3.93}$$
$$b^{-}\left(\pi_8+\pi_9+\pi_{12}+\pi_{13}\right)=\pi_{12}, \tag{3.94}$$
$$a^{-}\left(\pi_{10}+\pi_{11}+\pi_{14}+\pi_{15}\right)=\pi_{13}, \tag{3.95}$$
$$b^{+}\left(\pi_8+\pi_9+\pi_{12}+\pi_{13}\right)=\pi_{14}, \quad \text{and} \tag{3.96}$$
$$a^{+}\left(\pi_{10}+\pi_{11}+\pi_{14}+\pi_{15}\right)=\pi_{15}. \tag{3.97}$$

Solving (3.82)–(3.97) for $\pi_0,\ldots,\pi_{15}$ we obtain

$$\pi_0=\kappa\left(1+a\right)\left(2+b+c\right), \tag{3.98}$$
$$\pi_1=2\kappa\left(1-a\right)\left(1+b\right), \tag{3.99}$$
$$\pi_2=\kappa\left(1-a\right)\left(2+b+c\right), \tag{3.100}$$
$$\pi_3=2\kappa\left(1-a\right)\left(1-b\right), \tag{3.101}$$
$$\pi_4=2\kappa\left(1-a\right)\left(1+c\right), \tag{3.102}$$
$$\pi_5=\pi_0, \tag{3.103}$$
$$\pi_6=2\kappa\left(1-a\right)\left(1-c\right), \tag{3.104}$$
$$\pi_7=\pi_2, \quad \text{and} \tag{3.105}$$
$$\pi_k=\pi_{15-k}, \quad k\in\{8,\ldots,15\}, \tag{3.106}$$

with

$$\kappa=\frac{1}{2\left(4+b+c-2a\right)}, \tag{3.107}$$

and where a, b and c are given in (3.49), (3.69) and (3.70), respectively.

For $j\in\{0,\ldots,15\}$, π_j in (3.98)–(3.106) is the asymptotic probability of being in state j, where the states $\mathbf{s}_{t|t}$ are given in table 3.7 below.

Table 3.7. Elements of state vector $\mathbf{s}_{t|t}$ and associated state indices for $N = 1$.

elements of $\mathbf{s}_{t\|t}$				state index
x_t	x_{t-1}	$\hat{x}_{t\|t}$	$\hat{x}_{t-1\|t-1}$	j
-1	-1	-1	-1	0
-1	-1	-1	1	1
-1	-1	1	-1	2
-1	-1	1	1	3
-1	1	-1	-1	4
-1	1	-1	1	5
-1	1	1	-1	6
-1	1	1	1	7
1	-1	-1	-1	8
1	-1	-1	1	9
1	-1	1	-1	10
1	-1	1	1	11
1	1	-1	-1	12
1	1	-1	1	13
1	1	1	-1	14
1	1	1	1	15

From this we obtain the limiting probability that the state vector $\mathbf{s}_{t|t}$ will contain two consecutive errors ($\hat{x}_{t|t} \neq x_t$ and $\hat{x}_{t-1|t-1} \neq x_{t-1}$):

$$\pi_3 + \pi_6 + \pi_9 + \pi_{12} = \frac{(1-a)(2-b-c)}{2(4+b+c-2a)}. \tag{3.108}$$

Similarly, the asymptotic probability of no errors is given by

$$\pi_0 + \pi_5 + \pi_{10} + \pi_{15} = \frac{(1+a)(2+b+c)}{2(4+b+c-2a)}; \tag{3.109}$$

the probability of a previous error and no current error is

$$\pi_1 + \pi_4 + \pi_{11} + \pi_{14} = \frac{(1-a)(2+b+c)}{2(4+b+c-2a)}; \tag{3.110}$$

and the probability of a current error and no previous error is

$$\pi_2 + \pi_7 + \pi_8 + \pi_{13} = \frac{(1-a)(2+b+c)}{2(4+b+c-2a)}. \tag{3.111}$$

The problem of obtaining a general analytic solution to (3.78) for the filtering–only atomic state space models must be left open, but we note that for any given channel model the solution can be worked 'by hand', since (3.78) is a set of linear equations, amenable to solution by standard techniques of linear algebra [49].

3.2.7 Error recovery time (BPSK, $N = 0$)

A further problem of interest is the determination of the time to recover from an error [8, 9, 20–22, 24, 39–41, 50, 52, 61]. We will apply Choy and Beaulieu's definition of error recovery time R_0, which is the first passage time to an 'error–free' state from an 'error–just–occurred' state [24]. In the process, some new results are derived, such as the mean and variance formulae (3.117) and (3.119); connections between the atomic state space models and some generalizations of the Fibonacci series; and the application of the theory of integer partitions to recovery time calculations within the context of the steady–state aggregated model of Choy and Beaulieu [24].

To illustrate, consider the first example problem given in section 3.2.4, this being the case of BPSK signalling on a linear FIR channel, for filtering only and $N = 0$. Recall the previous results in (3.33)–(3.40), and the state transition probability matrix (3.50), with reference to the definition of parameter a in (3.49). Consecutive states $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$ are stored in the words i and j, respectively, where $i, j \in \{0, 1, 2, 3\}$. Recall the state–to–integer mapping in table 3.4.

Figure 3.1 is a state transition diagram for the simple case $N = 0$, with circles denoting states and arrows showing possible transitions. *Error–free* states are coloured blue, and *error–just–occurred* states are yellow. By an error–free state $\mathbf{s}_{t|t}$, we mean one in which $\hat{x}_{t|t} = x_t$; an error–just–occurred state has $\hat{x}_{t|t} \neq x_t$.

From Choy and Beaulieu's definition, the error recovery time R_0 for this example would be the first passage time from any error–just–occurred state (yellow) to any error–free state (blue). Figure 3.1 shows that it is possible to make one or more consecutive filtering errors, where $\hat{x}_{t|t} \neq x_t$, such as the sequence of transitions $1 \to 1 \to 2 \to 2$. The minimum value of R_0 is 1, but all values $r_0 \in \{1, 2, \dots\}$ are possible. We will compute the probability distribution, mean and variance of R_0.

The adjacency matrix A associated with the directed graph in figure 3.1 is

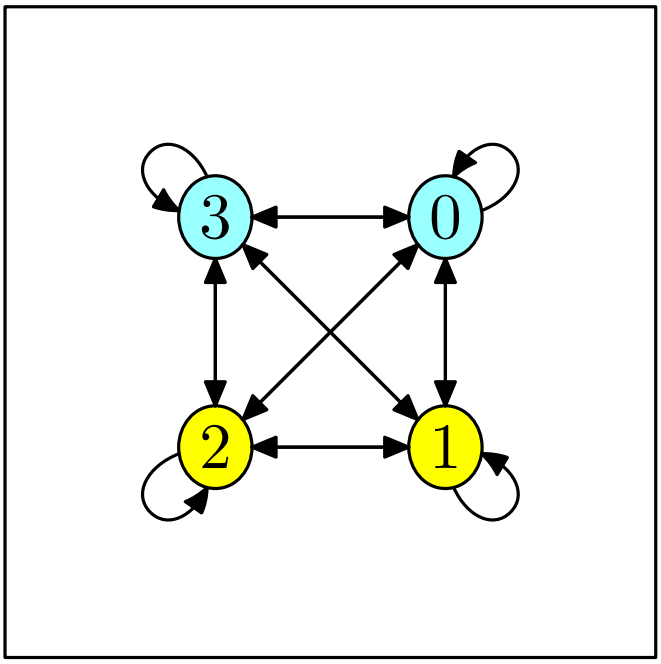

Figure 3.1. Valid atomic state transitions for BPSK channels with memory $N = 0$. Yellow states are 'error–just–occurred' states, for which $\hat{x}_{t|t} \neq x_t$; blue states are 'error–free' states, for which $\hat{x}_{t|t} = x_t$.

$$A = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \tag{3.112}$$

where entry a_{ij} shows the number of paths leading from state i to state j [62]. For the purpose of error recovery time calculation, we make the blue states in figure 3.1 *absorbing* or *terminal* states, by disallowing transitions from those states once reached [50]. With this modification, we have the revised graph given in figure 3.2.

With this change, we have the modified adjacency matrix A^*, where

$$A^* = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.113}$$

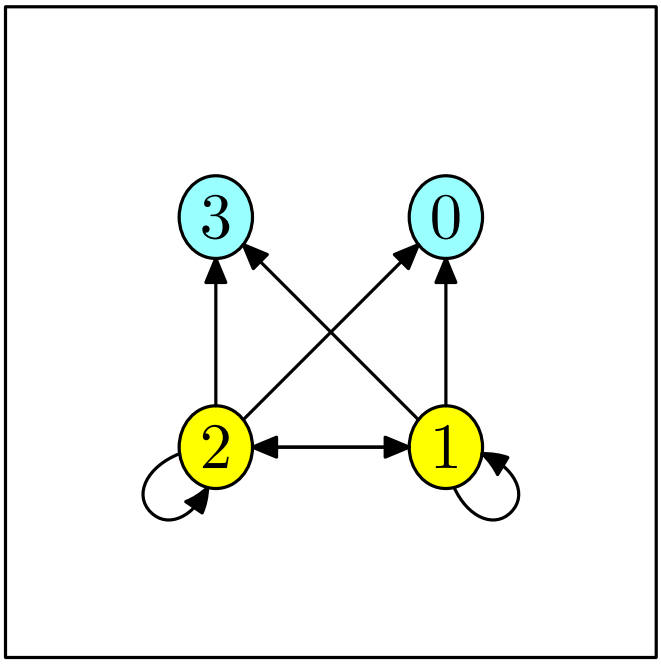

Figure 3.2. Valid atomic state transitions for BPSK channels with memory $N = 0$. This figure is similar to 3.1, except that we have made the error–free (blue) states terminal states, by preventing transitions from them.

The k–th power of A^* gives the number of paths, of exactly k steps, from state i to state j [17]. With A^* in (3.113), for $k \in \{1, 2, \dots\}$ we have

$$(A^*)^k = 2^{k-1} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.114}$$

We can now derive the probability distribution of R_0. From (3.49) and (3.50) we find that the probability of going from any of the yellow states to any of the blue states in figure 3.2 in a single step is the same value $\frac{1}{4}(1+a)$ for each pair of starting states $i \in \{1,2\}$ and finishing states $j \in \{0,3\}$. Similarly, the probability of going from any yellow state $i \in \{1,2\}$ to any yellow state $j \in \{1,2\}$ in a single step is $\frac{1}{4}(1-a)$. Assume equal *a priori* probabilities $\frac{1}{2}$ of being in states 1 and 2. With $p_{i,j}$ being the entries of the state transition probability matrix P in (3.50), where $i, j \in \{0,1,2,3\}$, by

direct calculation we have

$$\begin{aligned}\mathbb{P}(R_0 = 1) &= \frac{1}{2} \times p_{1,0} + \frac{1}{2} \times p_{1,3} + \frac{1}{2} \times p_{2,0} + \frac{1}{2} \times p_{2,3} \\ &= \frac{1}{2} \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1+a) \\ &= \frac{1}{2}(1+a), \end{aligned} \tag{3.115}$$

and

$$\begin{aligned}\mathbb{P}(R_0 = 2) &= \frac{1}{2} \times p_{1,1} \times p_{1,0} + \frac{1}{2} \times p_{1,2} \times p_{2,0} + \frac{1}{2} \times p_{1,1} \times p_{1,3} + \frac{1}{2} \times p_{1,2} \times p_{2,3} \\ &\quad + \frac{1}{2} \times p_{2,1} \times p_{1,0} + \frac{1}{2} \times p_{2,2} \times p_{2,0} + \frac{1}{2} \times p_{2,1} \times p_{1,3} + \frac{1}{2} \times p_{2,2} \times p_{2,3} \\ &= \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) \\ &\quad + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) \\ &\quad + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) \\ &\quad + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) + \frac{1}{2} \times \frac{1}{4}(1-a) \times \frac{1}{4}(1+a) \\ &= \frac{1}{4}(1-a)(1+a). \end{aligned} \tag{3.116}$$

In general, for $r_0 \in \{1, 2, \dots\}$, we have the probability distribution

$$\mathbb{P}(R_0 = r_0) = \frac{1}{2^{r_0}}(1-a)^{r_0-1}(1+a), \tag{3.117}$$

where a is given in (3.49). Using Maple$^{\text{TM}}$ we obtain for the mean and variance of the recovery time statistic R_0 the following results:

$$\begin{aligned}\mathbb{E}(R_0) &= \sum_{r_0=1}^{\infty} r_0 \mathbb{P}\left(R_0 = r_0\right) \\ &= \sum_{r_0=1}^{\infty} r_0 \frac{1}{2^{r_0}} \left(1-a\right)^{r_0-1} \left(1+a\right) \\ &= \frac{2}{1+a}, \end{aligned} \tag{3.118}$$

and

$$\begin{aligned}\mathbb{E}\left((R_0-\mathbb{E}(R_0))^2\right) &= \sum_{r_0=1}^{\infty}\left(r_0-\frac{2}{1+a}\right)^2\mathbb{P}(R_0=r_0)\\ &= \sum_{r_0=1}^{\infty}\left(r_0-\frac{2}{1+a}\right)^2 r_0\frac{1}{2^{r_0}}(1-a)^{r_0-1}(1+a)\\ &= \frac{1}{2}\left(1-\frac{2}{1+a}\right)^2(1+a)\\ &\quad\times {}_3F_2\left(1,\frac{2a}{1+a},\frac{2a}{1+a};\frac{-1+a}{1+a},\frac{-1+a}{1+a};\frac{1}{2}(1-a)\right), \quad (3.119)\end{aligned}$$

where ${}_3F_2(.)$ is a *generalized hypergeometric function* [62].

The results in (3.117)–(3.119) are new, not appearing in the main literature on error recovery times of DFEs [8–10, 20–22, 24, 28, 29, 39–41, 51, 52, 61].

3.2.8 Error recovery time (BPSK, $N = 1$)

For the second example, given in section 3.2.5, this being the case of BPSK signalling on a linear FIR channel of memory $N = 1$, and filtering only ($n = 0$), recall the channel model (3.51); the DFE estimator (3.52); the state transition probability (3.53), together with the definitions (3.54) and (3.55); the mapping rules (3.57) and (3.58) and table 3.5; and the state transition probability matrix (3.77), with reference to (3.71)–(3.76), and the quantities a, b and c of (3.49), (3.69) and (3.70), respectively.

We have $2^4 = 16$ state vectors, indexed by $i, j \in \{0, \ldots, 15\}$, where i is the index of a 'from' state $\mathbf{s}_{t-1|t-1}$ and j is the index of a 'to' state $\mathbf{s}_{t|t}$. Of note are the two shared values between $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$, namely, x_{t-1} and $\hat{x}_{t-1|t-1}$.

Following Choy and Beaulieu [24], partition the set of states $\mathbf{s}_{t|t} \in \mathcal{A}_2^4$ into mutually disjoint subsets Φ_0, Φ_1 and Γ. Γ is the set of 'error–free' states $\mathbf{s}_{t|t}$, those in which there are no filtering errors, so that $\hat{x}_{t|t} = x_t$ and $\hat{x}_{t-1|t-1} = x_{t-1}$. Φ_0 is the set of 'error–just–occurred' states, for which $\hat{x}_{t|t} \neq x_t$ and $\hat{x}_{t-1|t-1} \in \mathcal{A}_2$ (we don't care about the values of $\hat{x}_{t-1|t-1}$ and x_{t-1}). Lastly, Φ_1 is the set of 'partial–recovery' states, for which $\hat{x}_{t|t} = x_t$ and $\hat{x}_{t-1|t-1} \neq x_{t-1}$.

Table 3.8 below enumerates the state vector $\mathbf{s}_{t|t}$, showing the elements x_t, x_{t-1}, $\hat{x}_{t|t}$ and $\hat{x}_{t-1|t-1}$, the state indices $j \in \{0, \ldots, 15\}$ from the state–to–integer mapping rule (3.58), and the subset Φ_0, Φ_1 or Γ to which the state belongs. The quantities $e_{t|t}$ and $e_{t-1|t-1}$

are current and previous filtering errors, defined by

$$e_{t|t} = x_t - \hat{x}_{t|t}. \tag{3.120}$$

Table 3.8. Allocation of atomic states $\mathbf{s}_{t|t}$ to subsets Φ_0, Φ_1 and Γ for $N = 1$ and $n = 0$. $\mathbb{S}_t$ denotes an element of $\{\Phi_0, \Phi_1, \Gamma\}$, following Choy and Beaulieu [24].

NOTE:
This table is included on page 57
of the print copy of the thesis held in
the University of Adelaide Library.

From this table a directed graph of one–step transitions is provided in figure 3.3 below, which uses state indices to label the states. For convenience and brevity we shall use state indices in place of the states themselves, and write

$$\Gamma = \{0, 5, 10, 15\}, \tag{3.121}$$

$$\Phi_0 = \{2, 3, 6, 7, 8, 9, 12, 13\}, \quad \text{and} \tag{3.122}$$

$$\Phi_1 = \{1, 4, 11, 14\}, \tag{3.123}$$

with $\Gamma \cup \Phi_0 \cup \Phi_1 = \{0, \ldots, 15\}$.

Following the definition of Choy and Beaulieu [24], error recovery begins at any state in Φ_0 and ends at any state in Γ. We are interested in the recovery time statistic R_0, defined

as the number of discrete–time steps required to *first* reach a state in Γ, commencing from a state in Φ_0.

The adjacency matrix associated with the directed graph of figure 3.3 is given by

$$A = \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{bmatrix}. \tag{3.124}$$

The mirror symmetry of figure 3.3 about the dashed line is manifested in A through

$$a_{i,j} = a_{15-i,15-j}, \tag{3.125}$$

where $a_{i,j}$ is the element of A in row i and column j, with $i, j \in \{0, \ldots, 15\}$. This relation follows from figure 3.3 by exchanging states k and $15 - k$, where $k \in \{0, \ldots, 15\}$.

Since we are interested in the recovery time, we modify figure 3.3 so that all states in Γ are terminal. The modified directed graph is given in figure 3.4.

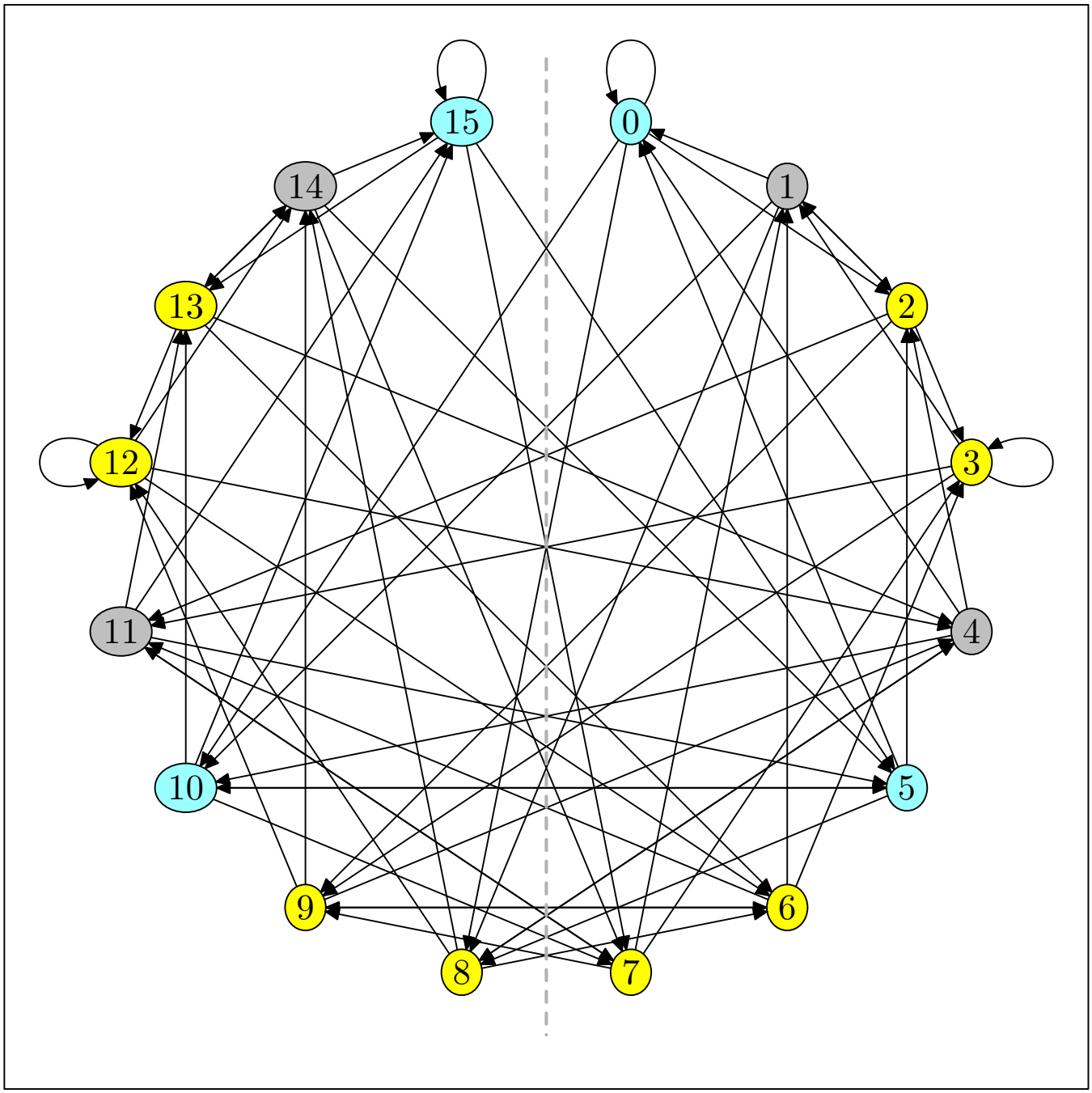

Figure 3.3. Atomic state transitions for BPSK channels with memory $N = 1$. Blue states $\mathbf{s}_{t|t}$ are 'error–free' states, with $\hat{x}_{t|t} = x_t$ and $\hat{x}_{t-1|t-1} = x_{t-1}$; yellow states are 'error–just–occurred' states, with $\hat{x}_{t|t} \neq x_t$; and grey states are intermediate states, with $\hat{x}_{t|t} = x_t$ and $\hat{x}_{t-1|t=1} \neq x_{t-1}$. Ignoring the state indexes $0, \ldots, 15$, note that the directed graph has topological mirror symmetry about the vertical dashed line.

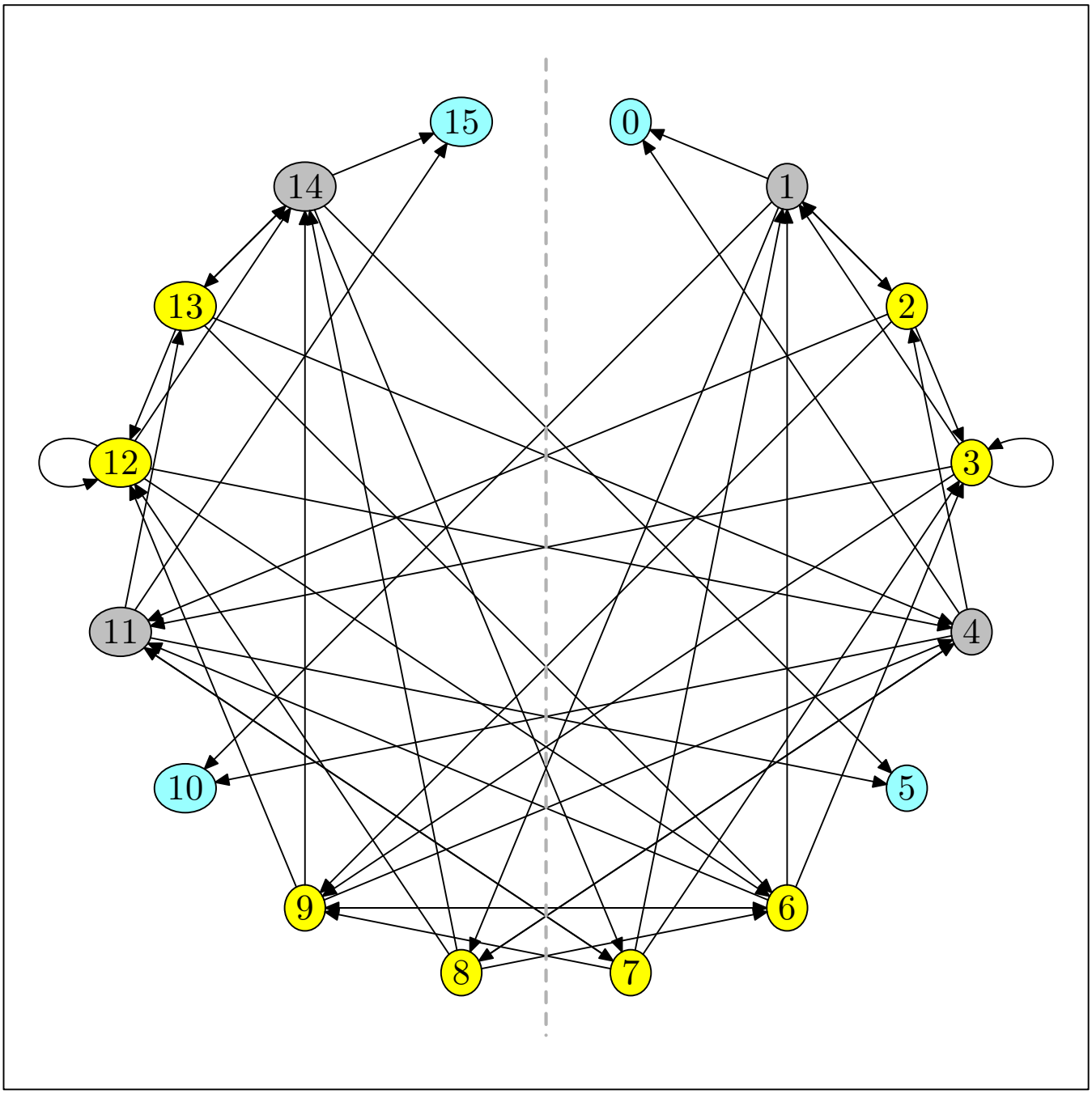

Figure 3.4. Atomic state transitions for BPSK channels with memory $N = 1$. This is a modification of figure 3.3, where states in Γ are now terminal states.

The adjacency matrix associated with the modified directed graph in figure 3.4 is now A^*, derived from (3.124) by zeroing out rows of A with row index $i \in \Gamma$, *viz.*:

$$A^* = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\underline{\mathbf{0}} & 1 & 0 & 1 & 0 & \underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} \\
\underline{\mathbf{0}} & 1 & 0 & 1 & 0 & \underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\underline{\mathbf{0}} & 1 & 0 & 1 & 0 & \underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} \\
\underline{\mathbf{0}} & 1 & 0 & 1 & 0 & \underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} \\
\underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} & 0 & 1 & 0 & 1 & \underline{\mathbf{0}} \\
\underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} & 0 & 1 & 0 & 1 & \underline{\mathbf{0}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
\underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} & 0 & 1 & 0 & 1 & \underline{\mathbf{0}} \\
\underline{\mathbf{0}} & 0 & 0 & 0 & 1 & \underline{\mathbf{0}} & 1 & 0 & 0 & 0 & \underline{\mathbf{0}} & 0 & 1 & 0 & 1 & \underline{\mathbf{0}} \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}. \tag{3.126}$$

The underlined entries $\underline{\mathbf{0}}$ in (3.126) mark the start and end states $i \in \Phi_0$ and $j \in \Gamma$ that are of interest in calculation of the recovery time R_0. We note that each marked entry is 0, meaning that there are no 1–step paths from any (yellow) state $i \in \Phi_0$ to any (blue) state $j \in \Gamma$, as figure 3.4 clearly shows. Thus we have the first term in the probability distribution of R_0 for $N = 1$ given by

$$\mathbb{P}(R_0 = 1) = 0. \tag{3.127}$$

Taking the square of A^* gives a matrix of the number of 2–paths that begin in state i and end in state j [17]. Carrying out the calculation gives

$$
(A^*)^2 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
\underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} & 1 & 1 & 1 & 1 & \underline{\mathbf{1}} \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}. \tag{3.128}
$$

Each underlined entry $\underline{\mathbf{1}}$ in (3.128) shows that there is a single path of 2 steps connecting the corresponding states $i \in \Phi_0$ and $j \in \Gamma$, such as $2 \to 11 \to 15$. Assuming the *a priori* probability of being in each of the eight starting states $i \in \Phi_0$ is $\frac{1}{8}$, and referring to the state transition probabilities in (3.77), with (3.49), (3.69), (3.70) and (3.71)–(3.76), we compute the second term $\mathbb{P}(R_0 = 2)$ in the probability distribution of R_0 for $N = 1$ as follows:

$$
\mathbb{P}(R_0 = 2) = \frac{1}{8}(1 + a)(2 + b + c). \tag{3.129}
$$

Taking higher powers of A^* in (3.126), we find that the underlined entries all have the same value, as was the case in (3.128), meaning that there is the same number of k–paths between any pair of starting and ending states $i \in \Phi_0$ and $j \in \Gamma$, respectively, where $k \in \{1, 2, \ldots\}$. Indeed, carrying out the operations, we note that the sequence of identical entries for matrix powers $k \in \{1, 2, \ldots\}$ is

$$
0, 1, 2, 8, 24, 80, 256, 832, 2688, 8704, 28160, 91136, 294912, 954368, 3088384, \tag{3.130}
$$

and so on.

By entering the partial sequence (3.130) into `Superseeker` [58], a computer search of published integer sequences revealed a match with the Horadam $(0, 1, 4, 2)$ sequence [33], given by the recursion

$$a_{k+2} = 2a_{k+1} + 4a_k, \tag{3.131}$$

where $k \geq 1$, with the initial conditions $a_1 = 0$ and $a_2 = 1$. Horadam's sequence is a generalization of the Fibonacci sequence, which satisfies the recurrence relation

$$b_{k+2} = b_{k+1} + b_k, \tag{3.132}$$

with the same initial conditions $b_1 = 0$ and $b_2 = 1$.

Generalizing the observation made above, we have:

Proposition 3.2.1. *For any pair of starting and finishing states $i \in \Phi_0$ and $j \in \Gamma$, respectively, a_k in* (3.131) *gives the number of k–paths to error recovery for the atomic state space model in figure 3.4.*

By direct calculation, the next few terms in the distribution of R_0 for $N = 1$ are

$$\mathbb{P}(R_0 = 3) = \frac{1}{32}(1+a)(2+b+c)(2-b-c), \tag{3.133}$$

$$\begin{aligned}\mathbb{P}(R_0 = 4) = \frac{1}{128}&(1+a)(2+b+c) \\ &\times (b^2 - 2ba + 2bc - 2b - 2ac - 4a + 8 + c^2 - 2c),\end{aligned} \tag{3.134}$$

and

$$\begin{aligned}\mathbb{P}(R_0 = 5) = \frac{1}{512}&(1+a)(2+b+c)(2-b-c) \\ &\times (b^2 - 4ba + 2bc - 4ac - 8a + 12 + c^2).\end{aligned} \tag{3.135}$$

No closed–form expression was obtained for general $r_0 \in \{2, 3, \dots\}$ with $N = 1$. Rather, we have the following series expressions for the terms in the probability distribution of R_0, with π_{k_1} being the *a priori* probability of initial state $k_1 \in \Phi_0$, and where p_{k_i,k_j} is the probability of making a 1–step transition from state k_i to state k_j:

$$\mathbb{P}(R_0 = 2) = \sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_1} \sum_{k_3 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3}, \tag{3.136}$$

$$\mathbb{P}(R_0 = 3) = \sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_0} \sum_{k_3 \in \Phi_1} \sum_{k_4 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4}, \tag{3.137}$$

$$\begin{aligned}\mathbb{P}(R_0 = 4) = &\sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_0} \sum_{k_3 \in \Phi_0} \sum_{k_4 \in \Phi_1} \sum_{k_5 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4} p_{k_4,k_5} \\ &+ \sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_1} \sum_{k_3 \in \Phi_0} \sum_{k_4 \in \Phi_1} \sum_{k_5 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4} p_{k_4,k_5},\end{aligned} \tag{3.138}$$

$$\begin{aligned}\mathbb{P}(R_0 = 5) = &\sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_0} \sum_{k_3 \in \Phi_0} \sum_{k_4 \in \Phi_0} \sum_{k_5 \in \Phi_1} \sum_{k_6 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4} p_{k_4,k_5} p_{k_5,k_6} \\ &+ \sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_0} \sum_{k_3 \in \Phi_1} \sum_{k_4 \in \Phi_0} \sum_{k_5 \in \Phi_1} \sum_{k_6 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4} p_{k_4,k_5} p_{k_5,k_6} \\ &+ \sum_{k_1 \in \Phi_0} \sum_{k_2 \in \Phi_1} \sum_{k_3 \in \Phi_0} \sum_{k_4 \in \Phi_0} \sum_{k_5 \in \Phi_1} \sum_{k_6 \in \Gamma} \pi_{k_1} p_{k_1,k_2} p_{k_2,k_3} p_{k_3,k_4} p_{k_4,k_5} p_{k_5,k_6},\end{aligned} \tag{3.139}$$

and so on, with the number of terms in (3.136)–(3.139) being the first few terms 1, 1, 2, 3 of the Fibonacci series [62]. This pattern was observed to follow for more terms than those listed above, and we formalize this observation as follows:

Proposition 3.2.2. *The number of multi–sum terms in the series beginning* (3.136)–(3.139) *is given by the Fibonacci sequence 1, 1, 2, 3, 5, 8, ...*

Redraw figure 3.4 by grouping atomic states according to (3.121)–(3.123). This gives figure 3.5, shown below.

Figure 3.5 ignores the individual states $\{0, \ldots, 15\}$ and focuses instead on the *classes* to which those states belong. We see from the figure that there is a single path of length 2 connecting a state in class Φ_0 (the error–just–occurred states) and a state in class Γ (the error–free states). This gives the single triple sum in (3.136), which is a sum over the individual states in classes Φ_0, Φ_1 and Γ, conditioned by the *a priori* probability of being in each starting state k_1 in Φ_0. We can collectively label all the individual state 2–paths, such as $2 \to 1 \to 0$ and $6 \to 11 \to 15$, as $\Phi_0 \to \Phi_1 \to \Gamma$.

Similarly, from figure 3.5 we observe that there is a single 3–step path from Φ_0 to Γ, which we denote by the shorthand $\Phi_0 \to \Phi_0 \to \Phi_1 \to \Gamma$. This stands for the *collection* of all atomic 3–paths that start in a state $k_1 \in \Phi_0$, transition to a second (possibly the same) state $k_2 \in \Phi_0$, then transition to a state $k_3 \in \Phi_1$, and finally end in a state $k_4 \in \Gamma$; an example of this class of error–recovery sequence is $12 \to 6 \to 11 \to 15$.

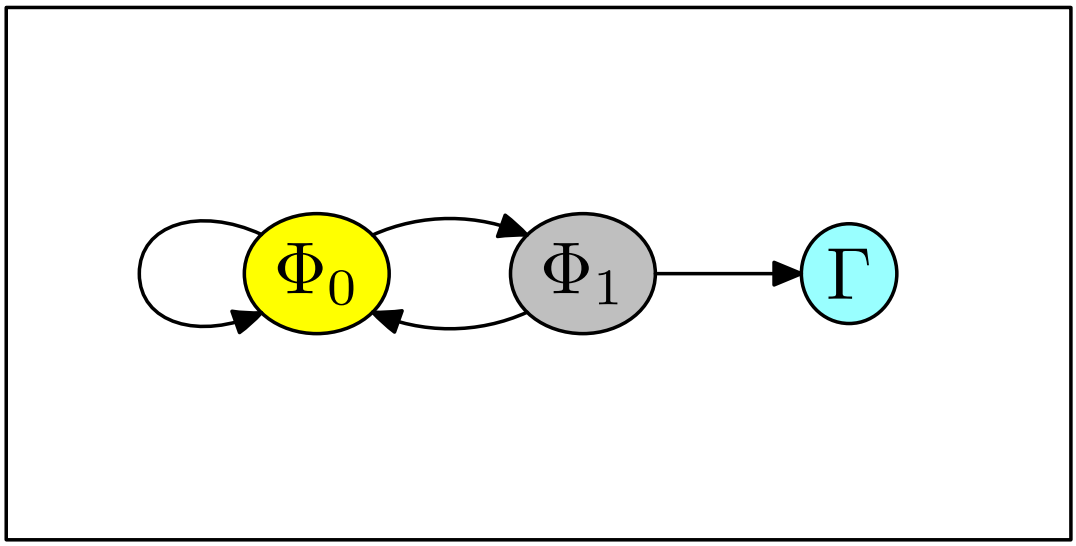

Figure 3.5. Figure 3.4 with the grouping of atomic states given by (3.121)–(3.123).

The single quadruple sum in (3.137) covers the totality of all 3–step paths described by $\Phi_0 \to \Phi_0 \to \Phi_1 \to \Gamma$.

Considering 4–paths now, we note from figure 3.5 that there are *two* possible classes of error–recovery sequences, *viz.* $\Phi_0 \to \Phi_0 \to \Phi_0 \to \Phi_1 \to \Gamma$ and $\Phi_0 \to \Phi_1 \to \Phi_0 \to \Phi_1 \to \Gamma$. Hence equation (3.138) has two distinct five–sum terms, one covering each class of atomic state path sequence. Continuing the analysis, we observe the Fibonacci sequence as given in proposition (3.2.2).

3.2.9 Error recovery time (BPSK, $N \geq 0$)

The results of section 3.2.8 generalize for BPSK signalling on channels of memory $N \geq 0$. Following Choy and Beaulieu [24], group atomic states $\mathbf{s}_{t|t}$ of (3.1) into mutually disjoint subsets $\Phi_0, \ldots, \Phi_N, \Gamma$, whose union is $\{0, \ldots, 2^{2N+2} - 1\}$. Note that these states are aggregations of atomic states, but they do not form a finite–state Markov process. This point will be discussed further in section 3.4. The convention we followed above, and that we shall maintain below, is that of speaking synonymously about a set, Φ_0 say, as being *either* a subset of states $\mathbf{s}_{t|t} \in \mathcal{A}_2^{2N+2}$ *or* a matching subset of state indices, recalling the state–to–integer mapping in (3.3) and (3.4). We have the following

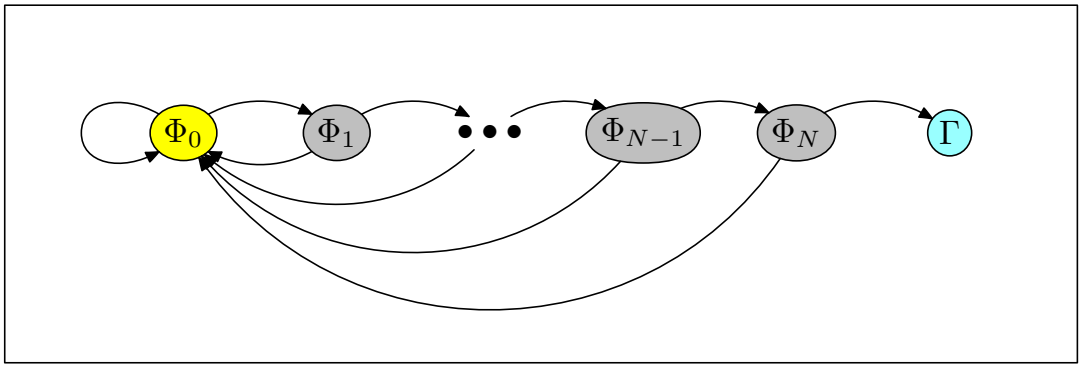

Figure 3.6. Error recovery for general BPSK channels of memory $N \geq 0$.

sets of states:

$$\Phi_0 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} \neq x_t\}, \tag{3.140}$$

$$\vdots$$

$$\Phi_k = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \ldots, \hat{x}_{t-k+1|t-k+1} = x_{t-k+1}, \hat{x}_{t-k|t-k} \neq x_{t-k}\}, \tag{3.141}$$

$$\vdots$$

$$\Phi_N = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \ldots, \hat{x}_{t-N+1|t-N+1} = x_{t-N+1}, \hat{x}_{t-N|t-N} \neq x_{t-N}\}, \tag{3.142}$$

and

$$\Gamma = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \ldots, \hat{x}_{t-N|t-N} = x_{t-N}\}. \tag{3.143}$$

Φ_k is the set of atomic states with $k \in \{0, \ldots, N\}$ consecutive error–free estimates, counting backwards from time t, and Γ is the set of 'error–free' atomic states. Φ_0 is the 'error–just–occurred' state, and $\Phi_1, \ldots, \Phi_N$ are sets of 'partial–recovery' states.

According to the definition used by Choy and Beaulieu [24], error recovery starts at an atomic state $i \in \Phi_0$ and ends at an atomic state $j \in \Gamma$, after a minimum of $N + 1$ steps. This is shown in figure 3.6, which generalizes figure 3.5.

Using a computer for rapid enumeration, it was observed that the number of distinct k–step paths from a given atomic state $i \in \Phi_0$ to a given atomic state $j \in \Gamma$ followed a recurrence relation that generalizes Horadam's (0,1,4,2) sequence (3.131). This result, obtained by entering partial sequences into `Superseeker` [58], appears to be new in the literature on DFE error propagation.

Proposition 3.2.3. *For $N \geq 0$, BPSK signalling, and filtering only ($n = 0$), the number of k–step paths from a given atomic state $i \in \Phi_0$ to a given atomic state $j \in \Gamma$ follows a recurrence*

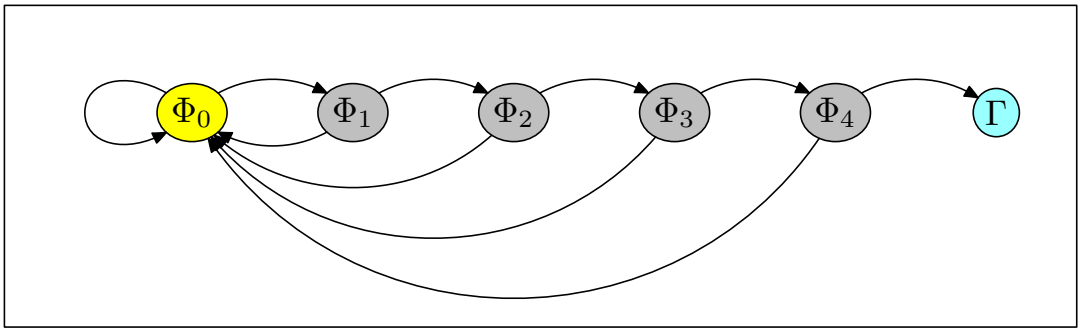

Figure 3.7. Error recovery for general BPSK channels of memory $N = 4$.

relation that generalizes Horadam's (0,1,4,2) sequence (3.131), viz.

$$a_{k+N+1} = \sum_{l=1}^{N+1} 2^l a_{k+N+1-l}, \tag{3.144}$$

with initialization conditions $a_1 = \cdots = a_N = 0$ *and* $a_{N+1} = 1$.

Moreover, it was observed that there was a generalization of the Fibonacci series for $N \geq 0$, analogous to the Fibonacci series resulting from (3.136)–(3.139) for $N = 1$:

Proposition 3.2.4. *The number of distinct multi–sum terms in the series of probabilities* $\mathbb{P}(R_0 = r_0)$ *follows a* generalized *Fibonacci series, defined by the* $(N+1)$*–term linear recurrence relation [62]*

$$b_{k+N+1} = \sum_{l=1}^{N+1} b_{k+N+1-l}, \tag{3.145}$$

with initialization conditions $b_1 = \cdots = b_N = 0$ *and* $b_{N+1} = 1$, *which is simply* (3.144) *with constants 1 instead of* 2^l.

For example, with $N = 4$, proposition 3.2.3 is that the number of k–step paths to error recovery between pairs $i \in \Phi_0$ and $j \in \Gamma$ of atomic states is given by

$$a_{k+5} = 2a_{k+4} + 4a_{k+3} + 8a_{k+2} + 16a_{k+1} + 32a_k, \tag{3.146}$$

with initialization conditions $a_1 = a_2 = a_3 = a_4 = 0$ and $a_5 = 1$, where $k \geq 1$. This sequence begins 0, 0, 0, 0, 1, 2, 8, 32, 128, 512, 1984, 7808, 30720, 120832, 475136, ... and we observe from figure 3.7 that the minimum recovery time is 5 time steps.

Continuing with this example we see that proposition 3.2.4 states that the number of distinct terms in the sums for $\mathbb{P}(R_0 = r_0)$ will be given by the generalized Fibonacci series

$$b_{k+5} = b_{k+4} + b_{k+3} + b_{k+2} + b_{k+1} + b_k, \tag{3.147}$$

with initialization conditions $b_1 = b_2 = b_3 = b_4 = 0$ and $b_5 = 1$, where $k \geq 1$. This sequence begins 0, 0, 0, 0, 1, 1, 2, 4, 8, 16, 31, 61, 120, 236, 464, ... so that the distribution of R_0 begins as $\mathbb{P}(R_0 = 1) = \cdots = \mathbb{P}(R_0 = 4) = 0$, and continues with

$$\begin{aligned}\mathbb{P}(R_0 = 5) = &\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_1}\sum_{k_3\in\Phi_2}\sum_{k_4\in\Phi_3}\sum_{k_5\in\Phi_4}\sum_{k_6\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6},\end{aligned} \tag{3.148}$$

$$\begin{aligned}\mathbb{P}(R_0 = 6) = &\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_0}\sum_{k_3\in\Phi_1}\sum_{k_4\in\Phi_2}\sum_{k_5\in\Phi_3}\sum_{k_6\in\Phi_4}\sum_{k_7\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7},\end{aligned} \tag{3.149}$$

$$\begin{aligned}\mathbb{P}(R_0 = 7) = &\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_0}\sum_{k_3\in\Phi_0}\sum_{k_4\in\Phi_1}\sum_{k_5\in\Phi_2}\sum_{k_6\in\Phi_3}\sum_{k_7\in\Phi_4}\sum_{k_8\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8}\\ &+\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_1}\sum_{k_3\in\Phi_0}\sum_{k_4\in\Phi_1}\sum_{k_5\in\Phi_2}\sum_{k_6\in\Phi_3}\sum_{k_7\in\Phi_4}\sum_{k_8\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8},\end{aligned} \tag{3.150}$$

$$\begin{aligned}\mathbb{P}(R_0 = 8) = &\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_0}\sum_{k_3\in\Phi_0}\sum_{k_4\in\Phi_0}\sum_{k_5\in\Phi_1}\sum_{k_6\in\Phi_2}\sum_{k_7\in\Phi_3}\sum_{k_8\in\Phi_4}\sum_{k_9\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8}p_{k_8,k_9}\\ &+\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_0}\sum_{k_3\in\Phi_1}\sum_{k_4\in\Phi_0}\sum_{k_5\in\Phi_1}\sum_{k_6\in\Phi_2}\sum_{k_7\in\Phi_3}\sum_{k_8\in\Phi_4}\sum_{k_9\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8}p_{k_8,k_9}\\ &+\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_1}\sum_{k_3\in\Phi_0}\sum_{k_4\in\Phi_0}\sum_{k_5\in\Phi_1}\sum_{k_6\in\Phi_2}\sum_{k_7\in\Phi_3}\sum_{k_8\in\Phi_4}\sum_{k_9\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8}p_{k_8,k_9}\\ &+\sum_{k_1\in\Phi_0}\sum_{k_2\in\Phi_1}\sum_{k_3\in\Phi_2}\sum_{k_4\in\Phi_0}\sum_{k_5\in\Phi_1}\sum_{k_6\in\Phi_2}\sum_{k_7\in\Phi_3}\sum_{k_8\in\Phi_4}\sum_{k_9\in\Gamma}\pi_{k_1}\\ &\times p_{k_1,k_2}p_{k_2,k_3}p_{k_3,k_4}p_{k_4,k_5}p_{k_5,k_6}p_{k_6,k_7}p_{k_7,k_8}p_{k_8,k_9},\end{aligned} \tag{3.151}$$

and so on.

For general channel memory $N \geq 1$, BPSK signalling and filtering only, no expressions in closed form were found for the probability distribution of the error recovery time

R_0. It is not expected that any exist, relating to the fact that there are no explicit solutions to both of the linear recurrence relations (3.144) and (3.145) for $N > 4$. Closed form solutions are known for the Fibonacci numbers ($b_{k+2} = b_{k+1} + b_k$), the tribonacci numbers (generalized Fibonacci $b_{k+3} = b_{k+2} + b_{k+1} + b_k$) and the tetranacci numbers (generalized Fibonacci $b_{k+4} = b_{k+3} + b_{k+2} + b_{k+1} + b_k$), from the roots of the polynomials $P_2(x) = 1 - x - x^2$, $P_3(x) = 1 - x - x^2 - x^3$ and $P_4(x) = 1 - x - x^2 - x^3 - x^4$, respectively [62]. For instance, an explicit formula for the elements of the tribonacci sequence $\{b_k\}$ is given for $k \in \{1, 2, \dots\}$ by

$$\left[3 \frac{\left\{ \frac{1}{3}\left(19 + 3\sqrt{33}\right)^{1/3} + \frac{1}{3}\left(19 - 3\sqrt{33}\right)^{1/3} + \frac{1}{3} \right\}^k \left(586 + 102\sqrt{33}\right)^{1/3}}{\left(586 + 102\sqrt{33}\right)^{2/3} + 4 - 2\left(586 + 102\sqrt{33}\right)^{1/3}} \right], \quad (3.152)$$

where $[x]$ denotes the nearest integer function [62].

3.2.10 Error recovery time and the theory of integer partitions

There is an interesting connection, however, between the atomic state space models of this section and the theory of *integer partitions* [62]. These connections appear not to have been made elsewhere in the literature on DFE error propagation. An integer partition is a way of writing an integer as a sum of positive integers, without respect to order, and is usually written in descending order. For example, the partitions of 5 are

$$\begin{aligned} 5 &= 5 \\ &= 4 + 1 \\ &= 3 + 2 \\ &= 3 + 1 + 1 \\ &= 2 + 2 + 1 \\ &= 2 + 1 + 1 + 1 \\ &= 1 + 1 + 1 + 1 + 1. \end{aligned} \quad (3.153)$$

Partitions are obtained as solutions to the Diophantine equation [62]

$$1j_1 + 2j_2 + 3j_3 + \cdots + nj_n = n, \quad (3.154)$$

as illustrated in table 3.9 for $n = 5$.

Table 3.9. The partitions of 5 as solutions to (3.154).

j_1	j_2	j_3	j_4	j_5
0	0	0	0	1
1	0	0	1	0
0	1	1	0	0
2	0	1	0	0
1	2	0	0	0
3	1	0	0	0
5	0	0	0	0

The *partition function* $P(n)$ is the number of unrestricted partitions of n, so that $P(5) = 7$ in the example above. Related to $P(n)$ is the function $P(n,k)$, which gives the number of integer partitions of n for which the largest is exactly k. There is a beautiful exact formula for $P(n)$, given by Rademacher in 1937 [62], *viz.*

$$P(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n) \sqrt{k} \frac{d}{dn} \left\{ \frac{\sinh\left(\frac{\pi}{k}\sqrt{\frac{2}{3}\left(n - \frac{1}{24}\right)}\right)}{\sqrt{n - \frac{1}{24}}} \right\}, \tag{3.155}$$

where

$$A_k(n) = \sum_{h=1}^{k} \delta_{\mathrm{GCD}(h,k),1} \exp\left\{ \pi i \sum_{j=1}^{k-1} \frac{j}{k} \left(\frac{hj}{k} - \left\lfloor \frac{hj}{k} \right\rfloor - \frac{1}{2} \right) - \frac{2\pi i h n}{k} \right\}, \tag{3.156}$$

$\delta_{m,n}$ is the Kronecker delta, $\mathrm{GCD}(m,n)$ is the greatest common divisor function, and $\lfloor x \rfloor$ is the floor function. Exact formulae for the related partition function $P(n,k)$ do not appear to exist, however. Rather, there is a recurrence relation

$$P(n,k) = P(n-1,k-1) + P(n-k,k), \tag{3.157}$$

with $P(n,k) = 0$ for $k > n$, $P(n,n) = 1$, and $P(n,0) = 0$. The recurrence relation (3.157) can be solved for small k to give

$$P(n,1) = 1, \tag{3.158}$$

$$P(n,2) = \left\lfloor \frac{1}{2} t_2(n) \right\rfloor = \left\lfloor \frac{1}{2} n \right\rfloor, \tag{3.159}$$

$$P(n,3) = \left\lfloor \frac{1}{12} t_3^2(n) \right\rfloor = \left\lfloor \frac{1}{12} n^2 \right\rfloor, \quad \text{and} \tag{3.160}$$

$$P(n,4) = \begin{cases} \left\lfloor \frac{1}{144} t_4^3(n) - \frac{1}{48} t_4(n) \right\rfloor & \text{for } n \text{ even} \\ \left\lfloor \frac{1}{144} t_4^3(n) - \frac{1}{12} t_4(n) \right\rfloor & \text{for } n \text{ odd,} \end{cases} \tag{3.161}$$

where

$$t_k(n) = n + \frac{1}{4}k(k-3), \tag{3.162}$$

and $[x]$ is the nearest integer function [62]. For the example partition above, we note that $P(5,2) = 2$, for instance, as there are two integer partitions of 5 with a maximum addend of exactly 2, *viz.* $5 = 2+2+1$ and $5 = 2+1+1+1$.

Exploring the connection with integer partitions, consider the set of states

$$\mathbb{S} = \{\Phi_0, \ldots, \Phi_N, \Gamma\}, \tag{3.163}$$

where $\Phi_0, \ldots, \Phi_N, \Gamma$ were given earlier by (3.140)–(3.143). Each state $\Phi_0, \ldots, \Phi_N, \Gamma$ in the set $\mathbb{S}$ is an aggregation of one or more atomic states $\mathbf{s}_{t|t}$, and we note that

$$\Phi_0 \cup \cdots \cup \Phi_N \cup \Gamma = \mathbb{S}. \tag{3.164}$$

Note once again that these states are those of Choy and Beaulieu [24], and that they do not form a finite–state Markov process (see section 3.4 for further discussion of aggregated state space models).

We can describe the dynamics of the atomic state space model by a sequence $\{\mathbb{S}_t\}$ of states from $\mathbb{S}$, although we lose some information this way, through discarding the individual atomic states that we pass through. $\{\mathbb{S}_t\}$ is not a Markov chain, whereas $\{\mathbf{s}_{t|t}\}$ is, as we show in section 3.4. Suppose that there exist the *steady–state* transition probabilities [24]

$$\alpha_i = \lim_{t\to\infty} \mathbb{P}(\mathbb{S}_t = \Phi_{i+1} | \mathbb{S}_{t-1} = \Phi_i), \quad i \in \{0, \ldots, N-1\}, \quad \text{and} \tag{3.165}$$

$$\alpha_N = \lim_{t\to\infty} \mathbb{P}(\mathbb{S}_t = \Gamma | \mathbb{S}_{t-1} = \Phi_N). \tag{3.166}$$

Figure 3.8 gives the steady–state transition diagram for a general channel of memory $N = 4$. We do not show transitions from state Γ, to emphasize its role as an absorbing state for the purpose of error recovery time calculations.

From (3.140)–(3.143) we have the state definitions for $N = 4$ given by

$$\Phi_0 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} \neq x_t\}, \tag{3.167}$$

$$\Phi_1 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \hat{x}_{t-1|t-1} \neq x_{t-1}\}, \tag{3.168}$$

$$\Phi_2 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \hat{x}_{t-1|t-1} = x_{t-1}, \hat{x}_{t-2|t-2} \neq x_{t-2}\}, \tag{3.169}$$

$$\Phi_3 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \hat{x}_{t-1|t-1} = x_{t-1}, \hat{x}_{t-2|t-2} = x_{t-2}, \hat{x}_{t-3|t-3} \neq x_{t-3}\}, \tag{3.170}$$

$$\Phi_4 = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \ldots, \hat{x}_{t-3|t-3} = x_{t-3}, \hat{x}_{t-4|t-4} \neq x_{t-4}\}, \quad \text{and} \tag{3.171}$$

$$\Gamma = \{\mathbf{s}_{t|t} : \hat{x}_{t|t} = x_t, \ldots, \hat{x}_{t-4|t-4} = x_{t-4}\}. \tag{3.172}$$

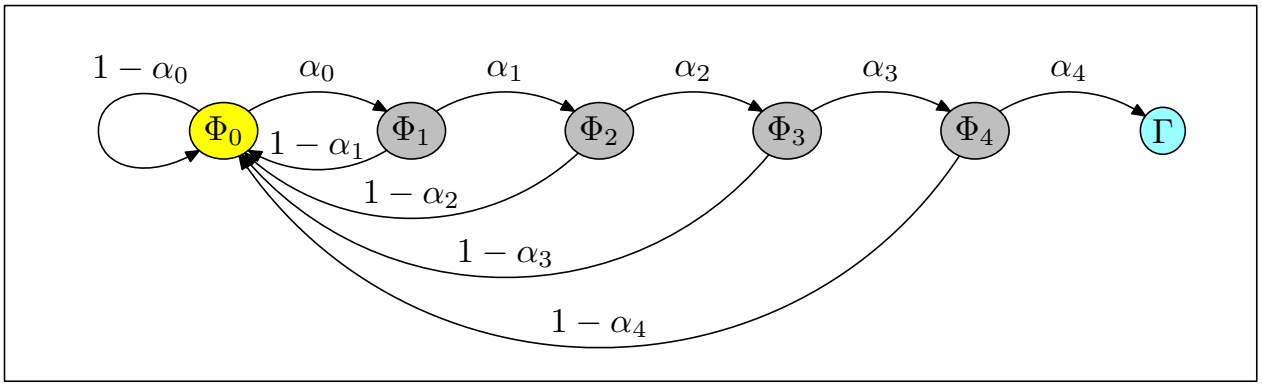

Figure 3.8. Steady–state transition diagram for Volterra communication channels using BPSK signalling and with memory $N = 4$. The aggregate states $\Phi_0, \ldots, \Phi_4, \Gamma$ are given by (3.167)–(3.172). This diagram is similar to figure 3.7 except that the probabilities α_i are *steady–state* probabilities, given by (3.165) and (3.166).

To illustrate the connection between integer partitions and the atomic state space models of this section, consider the above example with $N = 4$ and focus on error recovery sequences with a recovery time of $r_0 = 10$. Note from (3.147) with $k = 10$ that there are 16 distinct state–path sequences, as table 3.10 shows. For convenience in the discussion to follow, the sequences in table 3.10 are indexed by $l \in \{1, \ldots, 16\}$.

Table 3.10. State–path sequences from figure 3.8 that have a recovery time of $r_0 = 10$.

l	$\mathbb{S}_t$	$\mathbb{S}_{t+1}$	$\mathbb{S}_{t+2}$	$\mathbb{S}_{t+3}$	$\mathbb{S}_{t+4}$	$\mathbb{S}_{t+5}$	$\mathbb{S}_{t+6}$	$\mathbb{S}_{t+7}$	$\mathbb{S}_{t+8}$	$\mathbb{S}_{t+9}$	$\mathbb{S}_{t+10}$
1	Φ_0	Φ_0	Φ_0	Φ_0	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
2	Φ_0	Φ_0	Φ_0	Φ_0	Φ_1	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
3	Φ_0	Φ_0	Φ_0	Φ_1	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
4	Φ_0	Φ_0	Φ_0	Φ_1	Φ_2	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
5	Φ_0	Φ_0	Φ_1	Φ_0	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
6	Φ_0	Φ_0	Φ_1	Φ_0	Φ_1	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
7	Φ_0	Φ_0	Φ_1	Φ_2	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
8	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ

Continued on next page

Table 3.10 – concluded from previous page

l	$\mathbb{S}_t$	$\mathbb{S}_{t+1}$	$\mathbb{S}_{t+2}$	$\mathbb{S}_{t+3}$	$\mathbb{S}_{t+4}$	$\mathbb{S}_{t+5}$	$\mathbb{S}_{t+6}$	$\mathbb{S}_{t+7}$	$\mathbb{S}_{t+8}$	$\mathbb{S}_{t+9}$	$\mathbb{S}_{t+10}$
9	Φ_0	Φ_1	Φ_0	Φ_0	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
10	Φ_0	Φ_1	Φ_0	Φ_0	Φ_1	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
11	Φ_0	Φ_1	Φ_0	Φ_1	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
12	Φ_0	Φ_1	Φ_0	Φ_1	Φ_2	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
13	Φ_0	Φ_1	Φ_2	Φ_0	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
14	Φ_0	Φ_1	Φ_2	Φ_0	Φ_1	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
15	Φ_0	Φ_1	Φ_2	Φ_3	Φ_0	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ
16	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Φ_0	Φ_1	Φ_2	Φ_3	Φ_4	Γ

Each sequence in table 3.10 begins with an atomic state $\mathbf{s}_{t|t}$ in the class of error–just–occurred states Φ_0, and ends at an atomic state $\mathbf{s}_{t+10|t+10}$ in the class of error–free states Γ. Moreover, note that each sequence ends with the maximal–length subsequence of error–free transitions, $\Phi_0 \rightarrow \Phi_1 \rightarrow \Phi_2 \rightarrow \Phi_3 \rightarrow \Phi_4 \rightarrow \Gamma$. Prior to commencing this final error–recovery run, however, we note the occurrence of one or more filtering errors. Each error marks the end of a 'false start' with a return to Φ_0.

Use the term *cycle* to refer to a sequence of transitions that begin at Φ_0, end at Φ_0, and have an error–free run in between. Let the cycle $\Phi_0 \rightarrow \Phi_1 \rightarrow \cdots \rightarrow \Phi_{m-1} \rightarrow \Phi_m \rightarrow \Phi_0$ be denoted by c_m, where $m \in \{0, \ldots, N\}$ is the *length* of the cycle. In sequence 1, for example, we see that cycle c_0 occurs 5 times in a row. Table 3.11 below gives the cycle, or sequence of cycles, for each of the 16 state–path sequences listed in table 3.10. Each cycle sequence is read with time increasing left–to–right.

Table 3.11. Cycle(s) of the state–path sequences in table 3.10.

l	cycle(s)
1	$c_0\ c_0\ c_0\ c_0\ c_0$
2	$c_0\ c_0\ c_0\ c_1$
Continued on next page	

Table 3.11 – concluded from previous page

l	cycle(s)
3	$c_0\ c_0\ c_1\ c_0$
4	$c_0\ c_0\ c_2$
5	$c_0\ c_1\ c_0\ c_0$
6	$c_0\ c_1\ c_1$
7	$c_0\ c_2\ c_0$
8	$c_0\ c_3$
9	$c_1\ c_0\ c_0\ c_0$
10	$c_1\ c_0\ c_1$
11	$c_1\ c_1\ c_0$
12	$c_1\ c_2$
13	$c_2\ c_0\ c_0$
14	$c_2\ c_1$
15	$c_3\ c_0$
16	c_4

Observe that for $N = 4$ the minimum error–recovery time is 5 time steps, and so for $r_0 = 10$ there are 5 time steps in excess of this. Table 3.11 illustrates that these steps are taken up by permutations of one or more cycles chosen with replacement from $\{c_0, \ldots, c_N\}$. Each cycle c_m adds $m+1$ time steps to the recovery time, and the selection of cycles is constrained to have a sum of exactly 5 time steps. We can thus obtain the number of combinations of cycles in table 3.11 from the constrained partition numbers $P(5,1), \ldots, P(5,4)$ of (3.158)–(3.162):

$$P(5,1) = 1, \tag{3.173}$$

$$P(5,2) = \left\lfloor \frac{1}{2} \times 5 \right\rfloor = 2, \tag{3.174}$$

$$P(5,3) = \left\lfloor \frac{1}{12} \times 5^2 \right\rfloor = 2, \quad \text{and} \tag{3.175}$$

$$P(5,4) = \left\lfloor \frac{1}{144} (5+1)^3 - \frac{1}{12} (5+1) \right\rfloor = 1. \tag{3.176}$$

There is a single partition of 5 with a maximum addend of 1, as (3.173) shows, and that is $5 = 1+1+1+1+1$. This gives the cycle sequence $c_0c_0c_0c_0c_0$ at row $l = 1$ of table 3.11. From (3.174) we have two distinct partitions of 5 with maximum addend exactly 2, these being $5 = 2+2+1$ and $5 = 2+1+1+1$. The first partition, $5 = 2+2+1$,

gives the cycle sequences $c_0c_1c_1$, $c_1c_0c_1$ and $c_1c_1c_0$ at rows 6, 10 and 11, respectively. Observe that there are $3!/(1!\,2!) = 3$ of these cycle sequences, this being the number of visually distinct permutations of the string $c_0c_1c_1$. Similarly, there are 4 distinct permutations of $c_0c_0c_0c_1$, given in rows 2, 3, 5 and 9 of table 3.11.

Taking into account the state–path sequences in table 3.11, and the steady–state transition probabilities in figure 3.8, we have the following result for the probability of the (steady–state) recovery time R_0 being exactly 10 time steps:

$$\begin{aligned}\mathbb{P}(R_0 = 10) = \{&\alpha_0\alpha_1\alpha_2\alpha_3(1-\alpha_4) + 2(1-\alpha_0)\alpha_0\alpha_1\alpha_2(1-\alpha_3) \\ &+ 2\alpha_0^2(1-\alpha_1)\alpha_1(1-\alpha_2) + 3(1-\alpha_0)^2\alpha_0\alpha_1(1-\alpha_2) \\ &+ 3(1-\alpha_0)\alpha_0^2(1-\alpha_1)^2 + 4(1-\alpha_0)^3\alpha_0(1-\alpha_1) \\ &+ (1-\alpha_0)^5\}\prod_{i=0}^{4}\alpha_i. \end{aligned} \tag{3.177}$$

No general closed form was found for the probability distribution of R_0 using the approach outlined above, however, partly because the constrained partition function $P(n,k)$ has no closed form. Using a different approach, however, Choy and Beaulieu derived a closed form solution for R_0 for a related state space model [24].

3.3 Atomic State Space Models (Smoothing only)

Sections 3.3.1–3.3.3 parallel the treatment in sections 3.2.1–3.2.3, except that here we consider the smoothing–only case of FLSDFE operation, for channel memory $N \in \{1,2,\dots\}$ and FLSDFE smoothing lag $n \in \{1,\dots,N\}$. Interesting further work would be to explore analogous results to those in sections 3.2.4–3.2.10. Difficulties were encountered in this attempted generalization due to the fact that the FLSDFE algorithm presented in this thesis does not feed back smoothed errors $\hat{x}_{t-n|t}$, at lags $n > 0$. Although more algebraically involved, a smoothing–fed–back variant of the FLSDFE might provide a more consistent algorithm to analyze, and we leave that work for the future. More will be said on variant FLSDFE algorithms in section 5.

3.3.1 Model definition

For $N \geq 1$ and $n \in \{1,\dots,N\}$ we have the smoothed FLSDFE output $\hat{X}_{t-n|t}$. Extending the model (3.1), which was in turn based on a similar model used by Kennedy and

Anderson [39], we introduce a smoothing–only atomic state space model with state vector $\mathbf{S}_{t-n|t}$, where

$$\mathbf{S}_{t-n|t} = \begin{bmatrix} X_t \\ \vdots \\ X_{t-N-n} \\ \hat{X}_{t|t} \\ \vdots \\ \hat{X}_{t-N-n|t-N-n} \\ \hat{X}_{t-n|t} \end{bmatrix}. \tag{3.178}$$

$\hat{X}_{t-n|t}$ depends upon the symbols $X_t, \ldots, X_{t-N-n}$, and these occupy the top $N+n+1$ elements of $\mathbf{S}_{t-n|t}$. The next $N+n+1$ elements are filled with the corresponding filtered outputs $\hat{X}_{t|t}, \ldots, \hat{X}_{t-N-n|t-N-n}$. Note that $\hat{X}_{t-n|t}$ depends on all of these filtered outputs with the exception of $\hat{X}_{t-n|t-n}$. We include $\hat{X}_{t-n|t-n}$ in $\mathbf{S}_{t-n|t}$ for continuity, as $\hat{X}_{t-2n|t-n}, \ldots, \hat{X}_{t-n-1|t-1}, \hat{X}_{t-n+1|t+1}, \ldots, \hat{X}_{t-n+N|t+N}$ all depend upon $\hat{X}_{t-n|t-n}$. Lastly, in element $2N+2n+3$ is the FLSDFE smoothed output $\hat{X}_{t-n|t}$ itself.

Appendix I shows that for $m \in \mathbb{Z}^+$ the state transition probability is given by

$$\begin{aligned}
&\mathbb{P}(\mathbf{S}_{t-n|t} = \mathbf{s}_{t-n|t} | \mathbf{S}_{t-n-1|t-1} = \mathbf{s}_{t-n-1|t-1} \cap \cdots \cap \mathbf{S}_{t-n-m|t-m} = \mathbf{s}_{t-n-m|t-m}) \\
&= \mathbb{P}(\mathbf{S}_{t-n|t} = \mathbf{s}_{t-n|t} | \mathbf{S}_{t-n-1|t-1} = \mathbf{s}_{t-n-1|t-1}) \\
&= \mathbb{P}(X_t = x_t)\mathbb{P}(\hat{X}_{t|t} = \hat{x}_{t|t} | X_t = x_t \cap \cdots \cap X_{t-N} = x_{t-N} \\
&\quad \cap \hat{X}_{t-1|t-1} = \hat{x}_{t-1|t-1} \cap \cdots \cap \hat{X}_{t-N|t-N} = \hat{x}_{t-N|t-N}) \\
&\quad \times \mathbb{P}(\hat{X}_{t-n|t} = \hat{x}_{t-n|t} | X_t = x_t \cap \cdots \cap X_{t-N-n} = x_{t-N-n} \cap \hat{X}_{t|t} = \hat{x}_{t|t} \\
&\quad \cap \cdots \cap \hat{X}_{t-n+1|t-n+1} = \hat{x}_{t-n+1|t-n+1} \cap \hat{X}_{t-n-1|t-n-1} = \hat{x}_{t-n-1|t-n-1} \\
&\quad \cap \cdots \cap \hat{X}_{t-N-n|t-N-n} = \hat{x}_{t-N-n|t-N-n}).
\end{aligned} \tag{3.179}$$

The first two lines of (3.179) show that $\mathbf{S}_{t-n|t}$ is a first–order Markov process. The first probability term on the third line of (3.179), $\mathbb{P}(X_t = x_t)$, is the *a priori* probability of source symbols X_t; the second and third terms give the conditional probabilities of filtered output $\hat{X}_{t|t}$ and smoothed output $\hat{X}_{t-n|t}$, respectively.

The state space model described by the state vector definition (3.178), the state transition probability (3.179), and the FLSDFE algorithm is new. Much previous work has dealt largely with state space models based on the use of conventional DFE algorithms only (no fixed–lag smoothing); and on linear FIR channels only (unlike the Volterra channels that are treated in this thesis) [8–10, 20–24, 39–41, 51, 52, 61].

3.3.2 Lexicographical ordering of states (BPSK, $N \geq 1$)

We illustrate a lexicographical ordering of states that is suitable for use with BPSK signalling. This is an extension of the example given in 3.2 for the filtering–only case, and is based on the scheme of Kennedy and Anderson [39]. Using i for previous state $\mathbf{s}_{t-n-1|t-1}$ and j for current state $\mathbf{s}_{t-n|t}$, we have the explicit mapping rules

$$\begin{aligned} i = &\frac{(x_{t-1}+1)}{2}2^{2N+2n+2} + \cdots + \frac{(x_{t-N-n}+1)}{2}2^{N+n+3} + \frac{(x_{t-N-n-1}+1)}{2}2^{N+n+2} \\ &+ \frac{(\hat{x}_{t-1|t-1}+1)}{2}2^{N+n+1} + \cdots + \frac{(\hat{x}_{t-N-n|t-N-n}+1)}{2}2^{2} \\ &+ \frac{(\hat{x}_{t-N-n-1|t-N-n-1}+1)}{2}2^{1} + \frac{(\hat{x}_{t-n-1|t-1}+1)}{2}2^{0}, \end{aligned} \tag{3.180}$$

and

$$\begin{aligned} j = &\frac{(x_{t}+1)}{2}2^{2N+2n+2} + \frac{(x_{t-1}+1)}{2}2^{2N+2n+1} + \cdots + \frac{(x_{t-N-n}+1)}{2}2^{N+n+2} \\ &+ \frac{(\hat{x}_{t|t}+1)}{2}2^{N+n+1} + \frac{(\hat{x}_{t-1|t-1}+1)}{2}2^{N+n} + \cdots \\ &+ \frac{(\hat{x}_{t-N-n|t-N-n}+1)}{2}2^{1} + \frac{(\hat{x}_{t-n|t}+1)}{2}2^{0}. \end{aligned} \tag{3.181}$$

For general $N \geq 1$ and $n \in \{1, \ldots, N\}$, the scheme (3.180) and (3.181) is illustrated in table 3.12 below.

Table 3.12. Storage of consecutive state vectors $\mathbf{s}_{t-n-1|t-1}$ and $\mathbf{s}_{t-n|t}$ in words i and j, for the case of BPSK signalling, arbitrary channel memory $N \geq 1$, and smoothing only ($n \in \{1, \ldots, N\}$).

bit position in i	element in $\mathbf{s}_{t-n-1\|t-1}$	bit position in j	element in $\mathbf{s}_{t-n\|t}$
$2N+2n+2$	x_{t-1}	$2N+2n+2$	x_t
$\vdots$	$\vdots$	$2N+2n+1$	x_{t-1}
$N+n+3$	x_{t-N-n}	$\vdots$	$\vdots$
$N+n+2$	$x_{t-N-n-1}$	$N+n+2$	x_{t-N-n}
$N+n+1$	$\hat{x}_{t-1\|t-1}$	$N+n+1$	$\hat{x}_{t\|t}$
$\vdots$	$\vdots$	$N+n$	$\hat{x}_{t-1\|t-1}$
2	$\hat{x}_{t-N-n\|t-N-n}$	$\vdots$	$\vdots$
1	$\hat{x}_{t-N-n-1\|t-N-n-1}$	1	$\hat{x}_{t-N-n\|t-N-n}$
0	$\hat{x}_{t-n-1\|t-1}$	0	$\hat{x}_{t-n\|t}$

The oldest elements in $\mathbf{s}_{t-n-1|t-1}$, these being $x_{t-N-n-1}$, $\hat{x}_{t-N-n-1|t-N-n-1}$ and $\hat{x}_{t-n-1|t-1}$, are coloured blue, and occupy bit positions $N+n+2$, 1 and 0 in word i, respectively. The elements in common between $\mathbf{s}_{t-n-1|t-1}$ and $\mathbf{s}_{t-n|t}$, namely x_{t-1}, $\ldots$, x_{t-N-n}, $\hat{x}_{t-1|t-1}$, $\ldots$, $\hat{x}_{t-N-n|t-N-n}$, are coloured grey in both words i and j. In word i, these common elements occupy bit positions $2N+2n+2, \ldots, N+n+3, N+n+1, \ldots, 2$ respectively. In word j the elements have been shifted to the new positions $2N+2n+1$, $\ldots$, $N+n+2$, $N+n$, $\ldots$, 1 respectively, overwriting the old elements $x_{t-N-n-1}$ and $\hat{x}_{t-N-n-1|t-N-n-1}$. Finally, new elements x_t, $\hat{x}_{t|t}$ and $\hat{x}_{t-n|t}$ are placed into positions $2N+2n+2$, $N+n+1$ and 0 of $\mathbf{s}_{t-n|t}$, respectively, shown coloured yellow in the table.

3.3.3 State transition probability matrix (BPSK, $N \geq 1$)

In this section we give the state transition probability matrix P for the case of smoothing only and BPSK signalling. We will use the lexicographical ordering scheme of (3.180) and (3.181), for which P has the sparse, hierarchical structure given below. The general form of P given here is novel, and was not reported in the original work on the FLSDFE algorithm [52].

From (3.178) we see that for BPSK signalling there are $2^{2N+2n+3}$ atomic states $\mathbf{s}_{t-n|t}$, so that P will be of size $2^{2N+2n+3} \times 2^{2N+2n+3}$. Label the rows and columns of P with the

indices i and j of (3.180) and (3.181). Let the top left element of P have row and column indices $i = 0$ and $j = 0$, respectively; and let the bottom right element of P have row and column indices $i = 2^{2N+2n+3} - 1$ and $j = 2^{2N+2n+3} - 1$, respectively. Thus P is of the form $P = (p_{i,j})$, with elements $p_{i,j}$, where $i, j \in \{0, \ldots, 2^{2N+2n+3} - 1\}$:

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,\zeta} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,\zeta} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\zeta,0} & p_{\zeta,1} & \cdots & p_{\zeta,\zeta} \end{bmatrix}, \quad \zeta = 2^{2N+2n+3} - 1. \tag{3.182}$$

Observe from table 3.12 that x_t occurs only in $\mathbf{s}_{t-n|t}$, and that x_t is stored in the most significant bit of j, where j is the column index of P. With $j \in \{0, \ldots, 2^{2N+2n+3} - 1\}$, the MSB of j will be 0 for $j \in \{0, \ldots, 2^{2N+2n+2} - 1\}$, and 1 for $j \in \{2^{2N+2n+2}, \ldots, 2^{2N+2n+3} - 1\}$. That is, $x_t = -1$ in the left half of P, and $x_t = 1$ in the right half of P, and so we have a natural division of P into two equal–sized halves:

$$P = \begin{bmatrix} P_A & P_B \end{bmatrix}, \tag{3.183}$$

where P_A and P_B contain the transition probabilities in which $x_t = -1$ and $x_t = 1$, respectively. Note that P_A and P_B are both of size $2^{2N+2n+3} \times 2^{2N+2n+2}$.

Further, we can partition P_A and P_B by observing that $\mathbf{s}_{t-n-1|t-1}$ and $\mathbf{s}_{t-n|t}$ have the common elements $x_{t-1}, \ldots, x_{t-N-n}$. Observe from table 3.12 that the common elements $x_{t-1}, \ldots, x_{t-N-n}$ are stored in the $N + n$ most significant bits of row index i, whereas they are stored in the $N + n$ most significant bits *following* the MSB of column index j.

Taking into account the 2^{N+n} values of the $(N + n)$–tuple $(x_{t-1}, \ldots, x_{t-N-n})$, and observing the requirement to have matching $(N + n)$–tuples $(x_{t-1}, \ldots, x_{t-N-n})$ between $\mathbf{s}_{t-n-1|t-1}$ and $\mathbf{s}_{t-n|t}$, we find that P_A and P_B are sparse, with the block matrix forms

$$P_A = \begin{bmatrix} P_A^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_A^{(1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_A^{(\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_A^{(\zeta)} \end{bmatrix}, \tag{3.184}$$

and

$$P_B = \begin{bmatrix} P_B^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_B^{(1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_B^{(\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_B^{(\zeta)} \end{bmatrix}, \tag{3.185}$$

with $\zeta = 2^{N+n} - 1$, where $\mathbf{0}$ is a matrix of structural zeros, of size $2^{N+n+3} \times 2^{N+n+2}$. Each block $P_A^{(k)}$ and $P_B^{(k)}$ in (3.184) and (3.185), $k \in \{0, \ldots, \zeta\}$, is also of size $2^{N+n+3} \times 2^{N+n+2}$.

A further decomposition of each $P_A^{(k)}$ and $P_B^{(k)}$ results from observing that the 'old' element $x_{t-N-n-1}$ of $\mathbf{s}_{t-n-1|t-1}$ is redundant, since the state transition probability (3.179) does not depend on it. Note from table 3.12 that $x_{t-N-n-1}$ is stored in bit position $N+n+2$ of $\mathbf{s}_{t-n-1|t-1}$ and does not appear in $\mathbf{s}_{t-n|t}$.

Since the transition probability is independent of $x_{t-N-n-1}$, each $2^{N+n+3} \times 2^{N+n+2}$ matrix $P_A^{(k)}$ and $P_B^{(k)}$ in (3.184) and (3.185) splits horizontally into two identical matrices of size $2^{N+n+2} \times 2^{N+n+2}$, so that for $k \in \{0, \ldots, 2^{N+n} - 1\}$ we have

$$P_A^{(k)} = \begin{bmatrix} P_C^{(k)} \\ P_C^{(k)} \end{bmatrix}, \quad \text{and} \tag{3.186}$$

$$P_B^{(k)} = \begin{bmatrix} P_D^{(k)} \\ P_D^{(k)} \end{bmatrix}, \tag{3.187}$$

with upper and lower matrices corresponding to the choices $x_{t-N-n-1} = -1$ and $x_{t-N-n-1} = 1$, respectively.

Proceeding with the decomposition of P, note that for each fixed value of symbol x_t, $(N+n)$–tuple $(x_{t-1}, \ldots, x_{t-N-n})$ and symbol $x_{t-N-n-1}$, we have a choice of two values of estimate $\hat{x}_{t|t}$. Observe from table 3.12 that $\hat{x}_{t|t}$ is stored in bit position $N+n+1$ of j.

For $k \in \{0, \ldots, 2^{N+n} - 1\}$ we can therefore vertically divide $P_C^{(k)}$ and $P_D^{(k)}$ of (3.186) and (3.187), analogously to the partitioning of P in (3.183):

$$P_C^{(k)} = \begin{bmatrix} P_E^{(k)} & P_F^{(k)} \end{bmatrix} \quad \text{and} \tag{3.188}$$

$$P_D^{(k)} = \begin{bmatrix} P_G^{(k)} & P_H^{(k)} \end{bmatrix}, \tag{3.189}$$

where the left and right matrices in each pair correspond to the choices $\hat{x}_{t|t} = -1$ and $\hat{x}_{t|t} = 1$, respectively, and where $P_E^{(k)}$, $P_F^{(k)}$, $P_G^{(k)}$ and $P_H^{(k)}$ are each of size $2^{N+n+2} \times 2^{N+n+1}$. From (3.188) and (3.189), we now expand (3.186) and (3.187) to give

$$P_A^{(k)} = \begin{bmatrix} P_E^{(k)} & P_F^{(k)} \\ P_E^{(k)} & P_F^{(k)} \end{bmatrix} \quad \text{and} \tag{3.190}$$

$$P_B^{(k)} = \begin{bmatrix} P_G^{(k)} & P_H^{(k)} \\ P_G^{(k)} & P_H^{(k)} \end{bmatrix}, \tag{3.191}$$

where $k \in \{0, \ldots, 2^{N+n} - 1\}$.

From table 3.12 note the occurrence of common elements $\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N-n|t-N-n}$ in $\mathbf{s}_{t-1|t-1}$ and $\mathbf{s}_{t|t}$. Taking into account the 2^{N+n} possible values of the $(N+n)$–tuple

$$\left(\hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N-n|t-N-n}\right), \tag{3.192}$$

and observing the requirement to have matching $(N+n)$–tuples (3.192) between $\mathbf{s}_{t-n-1|t-1}$ and $\mathbf{s}_{t-n|t}$, we find that for each $k \in \{0, \ldots, 2^{N+n} - 1\}$, the matrices $P_E^{(k)}$, $P_F^{(k)}$, $P_G^{(k)}$ and $P_H^{(k)}$ in (3.190) and (3.191) are themselves sparse, and of the form

$$P_E^{(k)} = \begin{bmatrix} P_E^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_E^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_E^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_E^{(k,\zeta)} \end{bmatrix}, \tag{3.193}$$

$$P_F^{(k)} = \begin{bmatrix} P_F^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_F^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_F^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_F^{(k,\zeta)} \end{bmatrix}, \tag{3.194}$$

$$P_G^{(k)} = \begin{bmatrix} P_G^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_G^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_G^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_G^{(k,\zeta)} \end{bmatrix}, \quad \text{and} \tag{3.195}$$

$$P_H^{(k)} = \begin{bmatrix} P_H^{(k,0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & P_H^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & P_H^{(k,\zeta-1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & P_H^{(k,\zeta)} \end{bmatrix}, \tag{3.196}$$

with $\zeta = 2^{N+n} - 1$, where each block matrix in (3.193)–(3.196) is now of size 4×2.

Consider now the LSB of j, holding the FLSDFE smoothed output $\hat{x}_{t-n|t}$. For $k, l \in \{0, \ldots, 2^{N+n} - 1\}$, we split the matrices $P_E^{(k,l)}$, $P_F^{(k,l)}$, $P_G^{(k,l)}$ and $P_H^{(k,l)}$ of (3.193)–(3.196) vertically, with the left half of each reflecting the choice $\hat{x}_{t-n|t} = -1$ and the right the choice $\hat{x}_{t-n|t} = 1$:

$$P_E^{(k,l)} = \begin{bmatrix} P_{E_\alpha}^{(k,l)} & P_{E_\beta}^{(k,l)} \end{bmatrix}, \tag{3.197}$$

$$P_F^{(k,l)} = \begin{bmatrix} P_{F_\alpha}^{(k,l)} & P_{F_\beta}^{(k,l)} \end{bmatrix}, \tag{3.198}$$

$$P_G^{(k,l)} = \begin{bmatrix} P_{G_\alpha}^{(k,l)} & P_{G_\beta}^{(k,l)} \end{bmatrix}, \quad \text{and} \tag{3.199}$$

$$P_H^{(k,l)} = \begin{bmatrix} P_{H_\alpha}^{(k,l)} & P_{H_\beta}^{(k,l)} \end{bmatrix}, \tag{3.200}$$

where each matrix is now of size 4×1.

Substituting (3.197)–(3.200) into (3.193)–(3.196) produces the following further decompositions of $P_E^{(k)}$, $P_F^{(k)}$, $P_G^{(k)}$ and $P_H^{(k)}$, for $k \in \{0, \ldots, 2^{N+n} - 1\}$:

$$P_E^{(k)} = \begin{bmatrix} P_{E_\alpha}^{(k,0)} & P_{E_\beta}^{(k,0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_{E_\alpha}^{(k,1)} & P_{E_\beta}^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & P_{E_\alpha}^{(k,\zeta-1)} & P_{E_\beta}^{(k,\zeta-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & P_{E_\alpha}^{(k,\zeta)} & P_{E_\beta}^{(k,\zeta)} \end{bmatrix}, \tag{3.201}$$

$$P_F^{(k)} = \begin{bmatrix} P_{F_\alpha}^{(k,0)} & P_{F_\beta}^{(k,0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_{F_\alpha}^{(k,1)} & P_{F_\beta}^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & P_{F_\alpha}^{(k,\zeta-1)} & P_{F_\beta}^{(k,\zeta-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & P_{F_\alpha}^{(k,\zeta)} & P_{F_\beta}^{(k,\zeta)} \end{bmatrix}, \tag{3.202}$$

$$P_G^{(k)} = \begin{bmatrix} P_{G_\alpha}^{(k,0)} & P_{G_\beta}^{(k,0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_{G_\alpha}^{(k,1)} & P_{G_\beta}^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & P_{G_\alpha}^{(k,\zeta-1)} & P_{G_\beta}^{(k,\zeta-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & P_{G_\alpha}^{(k,\zeta)} & P_{G_\beta}^{(k,\zeta)} \end{bmatrix}, \tag{3.203}$$

and

$$P_H^{(k)} = \begin{bmatrix} P_{H_\alpha}^{(k,0)} & P_{H_\beta}^{(k,0)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & P_{H_\alpha}^{(k,1)} & P_{H_\beta}^{(k,1)} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & P_{H_\alpha}^{(k,\zeta-1)} & P_{H_\beta}^{(k,\zeta-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & P_{H_\alpha}^{(k,\zeta)} & P_{H_\beta}^{(k,\zeta)} \end{bmatrix}, \tag{3.204}$$

with $\zeta = 2^{N+n} - 1$, where each block matrix in (3.201)–(3.204) is now of size 4×1.

The final step in the decomposition of P results from observing that the 'old' elements $\hat{x}_{t-N-n-1|t-N-n-1}$ and $\hat{x}_{t-n-1|t-1}$ of $\mathbf{s}_{t-n-1|t-1}$ are redundant, since the state transition probability (3.179) does not depend on them. Note from table 3.12 that $\hat{x}_{t-N-n-1|t-N-n-1}$ and $\hat{x}_{t-n-1|t-1}$ are stored in the least significant bit positions 1 and 0, respectively, of $\mathbf{s}_{t-n-1|t-1}$, and they do not appear in $\mathbf{s}_{t-n|t}$.

As there is the same transition probability for each four values of the pair

$$\left(\hat{x}_{t-N-n-1|t-N-n-1}, \hat{x}_{t-n-1|t-1}\right), \tag{3.205}$$

for $k, l \in \{0, \ldots, 2^{N+n} - 1\}$ each of the 4×1 matrices $P_{E_\alpha}^{(k,l)}$, $P_{E_\beta}^{(k,l)}$, $P_{F_\alpha}^{(k,l)}$, $P_{F_\beta}^{(k,l)}$, $P_{G_\alpha}^{(k,l)}$, $P_{G_\beta}^{(k,l)}$, $P_{H_\alpha}^{(k,l)}$ and $P_{H_\beta}^{(k,l)}$ in (3.201)–(3.204) splits horizontally into four identical scalars. With this in mind, for $k, l \in \{0, \ldots, 2^{N+n} - 1\}$ rewrite (3.197)–(3.200) as

$$P_E^{(k,l)} = \begin{bmatrix} p_{E_\alpha}^{(k,l)} & p_{E_\beta}^{(k,l)} \\ p_{E_\alpha}^{(k,l)} & p_{E_\beta}^{(k,l)} \\ p_{E_\alpha}^{(k,l)} & p_{E_\beta}^{(k,l)} \\ p_{E_\alpha}^{(k,l)} & p_{E_\beta}^{(k,l)} \end{bmatrix}, \tag{3.206}$$

$$P_F^{(k,l)} = \begin{bmatrix} p_{F_\alpha}^{(k,l)} & p_{F_\beta}^{(k,l)} \\ p_{F_\alpha}^{(k,l)} & p_{F_\beta}^{(k,l)} \\ p_{F_\alpha}^{(k,l)} & p_{F_\beta}^{(k,l)} \\ p_{F_\alpha}^{(k,l)} & p_{F_\beta}^{(k,l)} \end{bmatrix}, \tag{3.207}$$

$$P_G^{(k,l)} = \begin{bmatrix} p_{G_\alpha}^{(k,l)} & p_{G_\beta}^{(k,l)} \\ p_{G_\alpha}^{(k,l)} & p_{G_\beta}^{(k,l)} \\ p_{G_\alpha}^{(k,l)} & p_{G_\beta}^{(k,l)} \\ p_{G_\alpha}^{(k,l)} & p_{G_\beta}^{(k,l)} \end{bmatrix}, \quad \text{and} \tag{3.208}$$

$$P_H^{(k,l)} = \begin{bmatrix} p_{H_\alpha}^{(k,l)} & p_{H_\beta}^{(k,l)} \\ p_{H_\alpha}^{(k,l)} & p_{H_\beta}^{(k,l)} \\ p_{H_\alpha}^{(k,l)} & p_{H_\beta}^{(k,l)} \\ p_{H_\alpha}^{(k,l)} & p_{H_\beta}^{(k,l)} \end{bmatrix}, \tag{3.209}$$

where each element is now a scalar transition probability.

Since P is a stochastic matrix, we have the requirements

$$0 \leq p_{E_\alpha}^{(k,l)} \leq 1, \tag{3.210}$$

$$0 \leq p_{E_\beta}^{(k,l)} \leq 1, \tag{3.211}$$

$$0 \leq p_{F_\alpha}^{(k,l)} \leq 1, \tag{3.212}$$

$$0 \leq p_{F_\beta}^{(k,l)} \leq 1, \tag{3.213}$$

$$0 \leq p_{G_\alpha}^{(k,l)} \leq 1, \tag{3.214}$$

$$0 \leq p_{G_\beta}^{(k,l)} \leq 1, \tag{3.215}$$

$$0 \leq p_{H_\alpha}^{(k,l)} \leq 1, \tag{3.216}$$

$$0 \leq p_{H_\beta}^{(k,l)} \leq 1, \quad \text{and} \tag{3.217}$$

$$p_{E_\alpha}^{(k,l)} + p_{E_\beta}^{(k,l)} + p_{F_\alpha}^{(k,l)} + p_{F_\beta}^{(k,l)} + p_{G_\alpha}^{(k,l)} + p_{G_\beta}^{(k,l)} + p_{H_\alpha}^{(k,l)} + p_{H_\beta}^{(k,l)} = 1, \tag{3.218}$$

where $k, l \in \{0, \ldots, 2^{N+n} - 1\}$.

To recap, the state transition probability matrix P for the smoothing–only atomic state space model in (3.178) and (3.179) is given by (3.183), where P_A and P_B are given by (3.184) and (3.185), with the further decompositions in (3.190), (3.191), (3.201)–(3.204) and (3.206)–(3.209). This decomposition appears to be novel in the FLSDFE literature, and was not discussed in the original work of Perreau *et al.* [52].

3.4 Aggregated state space models

3.4.1 Introduction

The state space models in sections 3.2.1 and 3.3.1 are *atomic*, in that their state vectors consist of *fundamental* elements: the message symbols $\{x_t\}$, together with their filtered and smoothed estimates, $\{\hat{x}_{t|t}\}$ and $\{\hat{x}_{t-n|t}\}$, respectively. As such, these models are the most fine–grained, but that detail comes at the cost of high computational complexity. The filtering–only BPSK example in section 3.2, for instance, has 2^{2N+2} unique states, so is unwieldy for channels having a long memory N. We therefore seek models with fewer states that are still able to exactly model the transient properties of the underlying atomic finite–state Markov process (FSMP) [39]. For clarity, we shall refer to aggregations that satisfy the FSMP property as 'exact' aggregations. The Choy and Beaulieu aggregation, for instance, is not exact [24].

We first seek *optimal* exact aggregations, those with the smallest number of aggregated states. For the $N = 1$ example of section 3.2 we show, in section 3.4.4, that there exists a unique optimal exact aggregation of the atomic states $\{0, \ldots, 15\}$, this being (3.222). The proof of this involved an exhaustive computer search, making use of the theory of *set partitions* and *Bell numbers*, introduced in section 3.4.2, and of *restricted growth strings*, introduced in section 3.4.3. The application of these topics to error propagation dynamics appears to be novel in the (FLS)DFE literature.

A *suboptimal* exact aggregation (3.223) is subsequently introduced in section 3.4.5 for the BPSK, filtering–only, linear FIR, $N = 1$ example of section 3.2.5. This particular aggregation is related in a neat way to the single–distinct–errors model of Choy and Beaulieu [24], which is a (non–exact) *steady–state* model.

Section 3.4.6 discusses the single–distinct–errors model of Choy and Beaulieu [24] more fully. This model is not strictly an FSMP, and transitions between states are marked with their steady–state probabilities only. Nevertheless, the single–distinct–errors model provides a convenient formula for the (asymptotic) error recovery time, R_0, which is put to use in chapter 4, where the concept of 'resonances' is introduced into the state space models of DFE error propagation.

3.4.2 Set partitions and Bell numbers

By *set partition* we mean the partition of a set into nonempty subsets [62]. For example, the set partitions of $\{a,b,c\}$ are $\{\{a\},\{b\},\{c\}\}$, $\{\{a,b\},\{c\}\}$, $\{\{a,c\},\{b\}\}$, $\{\{a\},\{b,c\}\}$, and $\{\{a,b,c\}\}$. The number of set partitions of a set with n elements is called a *Bell number*, denoted B_n, with $B_0 = 1$ by definition [62]. Bell numbers grow rapidly, as can be observed from the recurrence relation

$$B_{n+1} = \sum_{k=0}^{n} B_k \binom{n}{k}, \quad n \geq 0. \tag{3.219}$$

The set $\{0,\ldots,15\}$, for instance, has $B_{16} = 10,480,142,147$ set partitions, such as

$$\{\{0\},\ldots,\{15\}\}, \tag{3.220}$$

$$\{\{0,5,10,15\},\{1,4,11,14\},\{2,3,6,7,8,9,12,13\}\}, \tag{3.221}$$

$$\{\{0,1,4,5\},\{2,3,6,7\},\{8,9,12,13\},\{10,11,14,15\}\}, \quad \text{and} \tag{3.222}$$

$$\{\{0,5\},\{1,4\},\{2,3,6,7\},\{8,9,12,13\},\{10,15\},\{11,14\}\}. \tag{3.223}$$

We will refer to examples (3.220)–(3.223) later in this section. Example (3.220) is the partition associated with the second of the two filtering–only examples of section 3.2, involving BPSK signalling on a linear FIR channel of memory $N = 1$. Partition (3.221) was introduced in equations (3.121)–(3.123) in the context of error recovery, which is the subject of the work of Choy and Beaulieu [24]. We will show that partition (3.222) is the unique optimal exact aggregation of $\{0,\ldots,15\}$, in the sense that it is the partition with fewest elements that exactly models the transient properties of the atomic FSMP. Partition (3.223) also gives an exact aggregation of the atomic state space model, but it has application to the study of error recovery dynamics, and is closely related to the work of Choy and Beaulieu [24].

Of the four partitions given above, only (3.221) does not produce an exact aggregated model. Although we introduced the partition (3.221) in (3.121)–(3.123), all error recovery time calculations were done using the atomic state space model directly. Thus the sums in (3.136)–(3.139), for instance, are over atomic states, not aggregated states. In section 3.4.4, we discuss a property that must be observed for a partition to be an exact aggregation (the FSMP property [39]), and show that (3.221) is not an exact aggregation of the atomic state space model for the $N = 1$ example of section 3.2.5.

3.4.3 Restricted growth strings

A *restricted growth string* (RGS) is an integer string $a_1 a_2 \dots a_n$ associated with a given set partition, and provides a convenient notation system for the partition [62]. This is best illustrated with an example, for instance partition (3.222) of $\{0, \dots, 15\}$, containing the subsets $\mathcal{S}_0, \dots, \mathcal{S}_3$ of $\{0, \dots, 15\}$, where

$$\mathcal{S}_0 = \{0, 1, 4, 5\}, \tag{3.224}$$

$$\mathcal{S}_1 = \{2, 3, 6, 7\}, \tag{3.225}$$

$$\mathcal{S}_2 = \{8, 9, 12, 13\}, \quad \text{and} \tag{3.226}$$

$$\mathcal{S}_3 = \{10, 11, 14, 15\}. \tag{3.227}$$

The first element of the set $\{0, \dots, 15\}$, namely 0, is in subset $\mathcal{S}_0$. We indicate this fact by setting the first element in the RGS, namely a_1, to 0, where 0 is the subscript of the subset $\mathcal{S}_0$. Since $a_1 = 0$, we say that the subsets have been *ordered*. Each remaining element $a_2, \dots, a_n$ of the RGS has *restricted growth* because of the inequality

$$a_{i+1} \leq \max\{a_1, a_2, \dots, a_i\} + 1, \quad i \in \{1, 2, \dots, n-1\}. \tag{3.228}$$

The complete RGS for set partition (3.222) is 0011001122332233, as can be seen by working through the elements of $\{0, \dots, 15\}$, in left–to–right order, and noting down the subscript of the subset in which each element occurs.

Restricted growth strings were used to efficiently enumerate all 10,480,142,147 set partitions of $\{0, \dots, 15\}$. Each partition was tested to see whether it was the basis of an FSMP of aggregated states [39]. Section 3.4.4 gives the unique optimal FSMP aggregation, and section 3.4.5 gives a suboptimal FSMP aggregation that is related to the non–FSMP state space models of Choy and Beaulieu [24].

3.4.4 Optimal state aggregation

As discussed above, the atomic state space model in section 3.2.1 is fine–grained but has a high computational complexity in the channel memory N. The number of states may be reduced by aggregation of these atomic states. In particular, we seek exact aggregations, those that are FSMPs. An FSMP has exactly the same transient behaviour as the atomic state space model, but with lower computational complexity [39].

To determine whether an aggregation is an FSMP, we must check whether the following property holds, which we shall call the 'FSMP property' [39]. Consider a set

partition $\{\mathcal{S}_0, \ldots, \mathcal{S}_k\}$ of the set of states $\{0, \ldots, m\}$ of a particular state space model. Then for each $x \in \mathcal{S}_i$, we must have the same probability of making a transition from x to $\mathcal{S}_j$. This property must be satisfied for all $i, j \in \{0, \ldots, k\}$, in order for $\{\mathcal{S}_0, \ldots, \mathcal{S}_k\}$ to describe an FSMP [39].

We illustrate using the partition (3.224)–(3.227) of the set of states $\{0, \ldots, 15\}$ associated with the linear BPSK channel example of memory $N = 1$. Consider aggregated sets $\mathcal{S}_0$ and $\mathcal{S}_1$ given by (3.224) and (3.225), respectively. For each element $x \in \mathcal{S}_0$, we will compute the probability of making a transition to the *set* $\mathcal{S}_1$.

From (3.49) and (3.69)–(3.77), the probability of going from element $0 \in \mathcal{S}_0$ to set $\mathcal{S}_1$ is

$$p_{0,2} + p_{0,3} + p_{0,6} + p_{0,7} = \frac{1}{4}(1-a) + 0 + 0 + 0 = \frac{1}{4}(1-a). \tag{3.229}$$

Likewise, the probability of going from element $1 \in \mathcal{S}_0$ to set $\mathcal{S}_1$ is

$$p_{1,2} + p_{1,3} + p_{1,6} + p_{1,7} = \frac{1}{4}(1-a) + 0 + 0 + 0 = \frac{1}{4}(1-a); \tag{3.230}$$

the probability of going from element $4 \in \mathcal{S}_0$ to set $\mathcal{S}_1$ is

$$p_{4,2} + p_{4,3} + p_{4,6} + p_{4,7} = \frac{1}{4}(1-a) + 0 + 0 + 0 = \frac{1}{4}(1-a); \tag{3.231}$$

and the probability of going from element $5 \in \mathcal{S}_0$ to set $\mathcal{S}_1$ is

$$p_{5,2} + p_{5,3} + p_{5,6} + p_{5,7} = \frac{1}{4}(1-a) + 0 + 0 + 0 = \frac{1}{4}(1-a). \tag{3.232}$$

Since all probability sums are equal, we have the same probability of going from each element $x \in \{0, 1, 4, 5\}$ to the set of elements $\{2, 3, 6, 7\}$. Repeating this test for each pair $i, j \in \{\mathcal{S}_0, \ldots, \mathcal{S}_3\}$ shows that the partition (3.222) describes an FSMP.

Figure 3.9 gives the state transition diagram for this partition. It is the unique optimal FSMP aggregation of the atomic state space model for the linear BPSK channel example of memory $N = 1$. This is the only FSMP that has 4 subsets, and all other FSMPs have 5 or more subsets.

The optimal partition follows logically from the fact that state $\mathbf{s}_{t-1|t-1}$ has two elements that are discarded in the transition to $\mathbf{s}_{t|t}$, namely, x_{t-2} and $\hat{x}_{t-2|t-2}$, which occupy bit positions 2 and 0 in i, respectively, as shown earlier in table 3.5. The (atomic) state transition probability $\mathbb{P}(\mathbf{S}_{t|t} = \mathbf{s}_{t|t} | \mathbf{S}_{t-1|t-1} = \mathbf{s}_{t-1|t-1})$ of (3.53) does not depend on 'old' elements x_{t-2} and $\hat{x}_{t-2|t-2}$, and these are shown in blue in table 3.5. The new

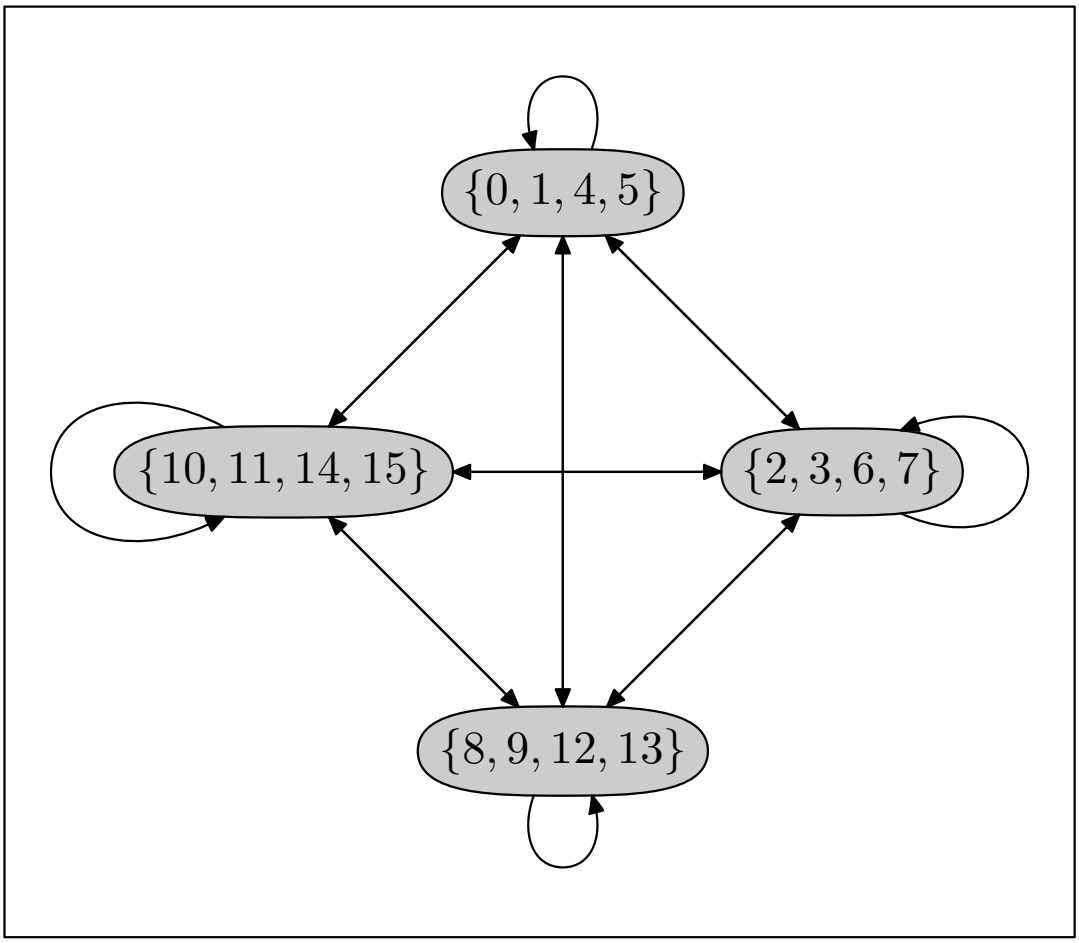

Figure 3.9. State transition diagram for the optimal FSMP for the linear BPSK channel of memory $N = 1$. The aggregated states are subsets $\mathcal{S}_0, \ldots, \mathcal{S}_3$ of the optimal partition of $\{0, \ldots, 15\}$, as given in (3.224)–(3.225).

elements x_t are $\hat{x}_{t|t}$ in table 3.5 are coloured yellow, and the elements in common are in grey.

We note that the groupings of atomic states in figure 3.9 are over all possible values of x_{t-2} and $\hat{x}_{t-2|t-2}$ for fixed values of x_{t-1} and $\hat{x}_{t-1|t-1}$. Thus the set $\{0, 1, 4, 5\}$ has $x_{t-1} = -1$ and $\hat{x}_{t-1|t-1} = -1$; the set $\{2, 3, 6, 7\}$ has $x_{t-1} = -1$ and $\hat{x}_{t-1|t-1} = 1$; the set $\{8, 9, 12, 13\}$ has $x_{t-1} = 1$ and $\hat{x}_{t-1|t-1} = -1$; and the set $\{10, 11, 14, 15\}$ has $x_{t-1} = 1$ and $\hat{x}_{t-1|t-1} = 1$.

The state transition probability matrix for the aggregated states is given by

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1+a & 1-a & 1-a & 1+a \\ 1+b & 1-b & 1-c & 1+c \\ 1+c & 1-c & 1-b & 1+b \\ 1+a & 1-a & 1-a & 1+a \end{bmatrix}, \tag{3.233}$$

where $p_{i,j}$ is the probability of making a transition from aggregate state $\mathcal{S}_i$ to aggregate state $\mathcal{S}_j$, where $i, j \in \{0, 1, 2, 3\}$. The stationary distribution vector π_∞ for the optimal FSMP is found from solving $P^T \pi_\infty = \pi_\infty$, *viz.*

$$\frac{1}{4} \begin{bmatrix} 1+a & 1+b & 1+c & 1+a \\ 1-a & 1-b & 1-c & 1-a \\ 1-a & 1-c & 1-b & 1-a \\ 1+a & 1+c & 1+b & 1+a \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}. \tag{3.234}$$

Solving (3.234), we find

$$\pi_\infty = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \frac{1}{2(4+b+c-2a)} \begin{bmatrix} 2+b+c \\ 2(1-a) \\ 2(1-a) \\ 2+b+c \end{bmatrix}. \tag{3.235}$$

The example illustrated above was for a linear FIR channel of memory $N = 1$, described by the channel model (3.51); for BPSK signalling, with alphabet $\{-1, 1\}$; and for filtering only ($n = 0$), with FLSDFE output $\hat{X}_{t|t}$ given by (3.52). A question of interest is whether there exists a unique optimum aggregated state space model for arbitrary channel memories $N \in \{0, 1, \dots\}$; general Volterra channel models; arbitrary smoothing lags $n \in \{0, \dots, N\}$; and a variety of other signalling schemes, such as QPSK or 16–QAM. In particular, we seek to know whether these aggregated models have fewer states than the set of atomic states from which they are composed.

Direct checking by the brute force method illustrated above is infeasible for most other models. Extending the example to channels of memory $N = 2$, for example, we find that there are $2^{2N+2} = 64$ states in $\{-1, 1\}^6$, and so there are B_{64} set partitions of $\{0, \dots, 63\}$ to check. From (3.219), the Bell number B_{64} is very large, given approximately by the asymptotic formula [62]

$$\frac{\ln B_{64}}{64} = \ln 64 - \ln\ln 64 - 1 + \frac{\ln\ln 64}{\ln 64} + \frac{1}{\ln 64} + \frac{1}{2}\left(\frac{\ln\ln 64}{\ln 64}\right)^2 + O\left(\frac{\ln\ln 64}{(\ln 64)^2}\right), \tag{3.236}$$

with approximate solution (ignoring the remainder) of $B_{64} \approx 1.4 \times 10^{69}$.

White *et al.* [65] discuss the notion of *lumpability*, and we repeat their definition here, using the notation of their paper:

Definition 3.4.1. *Let X_t be a n–state Markov chain taking values in the state set $\mathcal{S}$. Then X_t is said to be m–lumpable if and only if there exist nonempty disjoint subsets $\mathcal{S}_i, i = 1, \ldots, m$ such that $\mathcal{S} = \bigcup_{i=1}^{m} \mathcal{S}_i$ and for every $i = 1, \ldots, m$, $Pr\{X_t \in \mathcal{S}_i | X_{t-1} = q\}$ is independent of $q, \forall q \in \mathcal{S}_j, \forall j \neq i$.*

In their paper, White *et al.* present a finite algorithm for testing m–lumpability. We shall not discuss that work further here, but we note that it may prove useful in deriving exact and approximate FSMP aggregations of some of the more general atomic state space models that this thesis addresses.

The optimal aggregate FSMP is not convenient for obtaining the probability distribution or moments of the error recovery time R_0. Nor can we use it to compute the asymptotic probability of a bit error. This is because the aggregate states contain mixtures of the sets Φ_0, Φ_1 and Γ of (3.121)–(3.123). (Recall that Γ is the set of 'error–free' states, Φ_0 is the set of 'error–just–occurred' states, and Φ_1 is the set of 'intermediate' states.) Indeed, for the aggregate set $\{0, 1, 4, 5\}$ we note that $0, 5 \in \Gamma$, whereas $1, 4 \in \Phi_1$; and for the aggregate set $\{10, 11, 14, 15\}$ we note that $10, 15 \in \Gamma$, whereas $11, 14 \in \Phi_1$.

Although the sets Φ_0, Φ_1 and Γ do not constitute an FSMP aggregation of the atomic states $\{0, \ldots, 15\}$ for the linear BPSK channel of memory $N = 1$, we saw in section 3.2.8 that this was a convenient partition for the purpose of error recovery time calculation. Motivated by this fact, we seek an aggregation of the set $\{0, \ldots, 15\}$ that is an FSMP and is of convenient form to compute the probability distribution of R_0. Such a suboptimal aggregation exists, and is the subject of the next section.

3.4.5 A suboptimal aggregation

Using the same search program that was used in discovering the optimal exact aggregation (3.222) of the previous section, a slightly larger partition of $\{0, \ldots, 15\}$ was found that was well suited for error recovery time calculations. This suboptimal partition was given earlier as (3.223), and has the state transition diagram in figure 3.10. Note that the aggregated states are subsets (set partitions, moreover) of Φ_0, Φ_1 and Γ of (3.121)–(3.123). Thus $\{0, 5\}$ and $\{10, 15\}$ are subsets of the set of error–free states Γ;

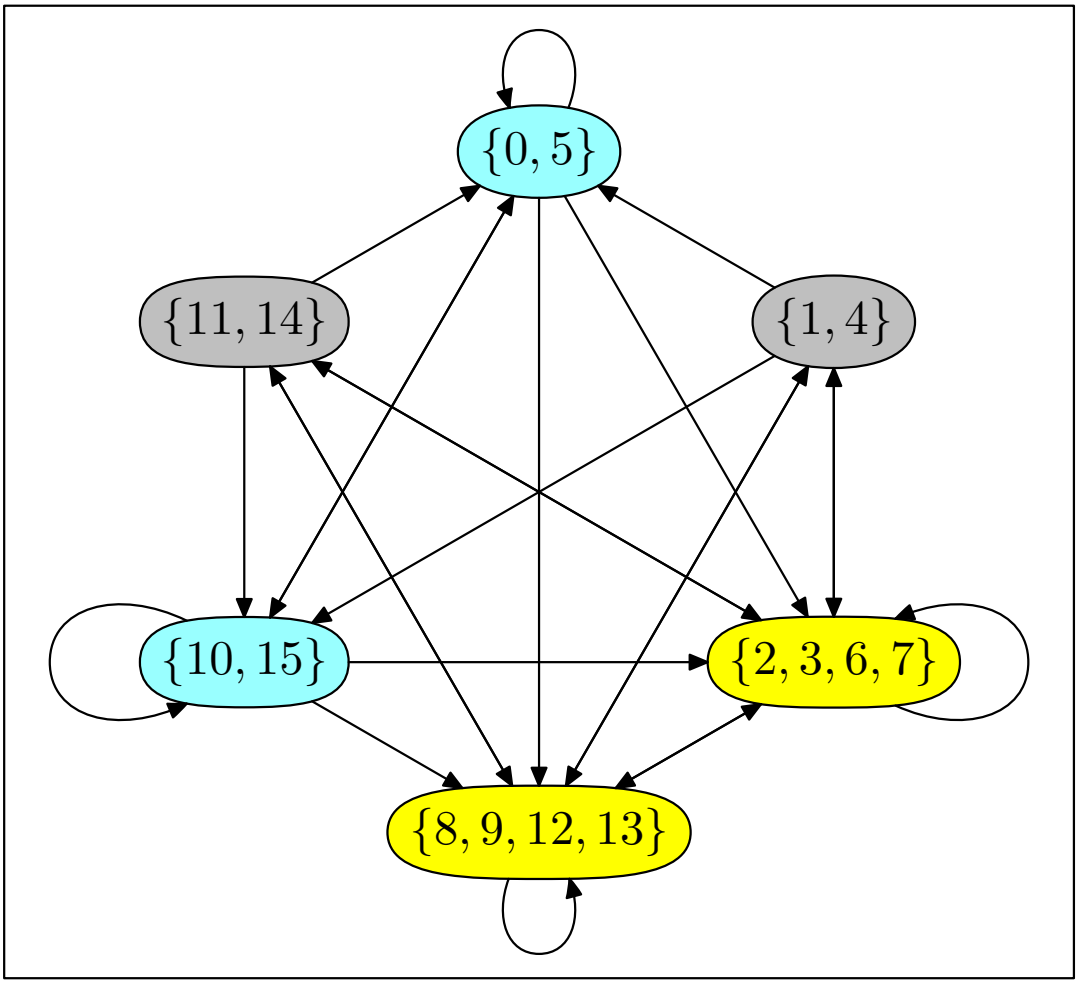

Figure 3.10. State transition diagram for a suboptimal FSMP for the linear BPSK channel of memory $N = 1$.

$\{2, 3, 6, 7\}$ and $\{8, 9, 12, 13\}$ are subsets of the set of error–just–occurred states Φ_0; and $\{1, 4\}$ and $\{11, 14\}$ are subsets of the intermediate partial–recovery states Φ_1.

Note that the diagram in figure 3.10 would apply to the wider case of a communications system model that employed BPSK signalling, with a channel memory of $N = 1$, but with a channel model of, say, a Volterra series (instead of the linear FIR channel model assumed so far). This is the case because the computer program that derived exact aggregations of the atomic states $\{0, \ldots, 15\}$ did not make any assumptions on the form of the channel model, representing state transition probabilities between atomic states as unique bits in a multibit word.

The asymptotic formula (3.236) for the Bell number B_n shows that it is impossible at present to derive by brute force the exact aggregations of atomic states $\{0, \ldots, 63\}$, since B_{64} is about 1.4×10^{69}. For communications system models with memory $N = 2$

(the next extension to the $N = 1$ case discussed in the preceding paragraph), similar diagrams to figures 3.9 and 3.10 were obtained by analogy to the $N = 1$ case. These diagrams are more complicated and are not presented here. A check was performed on paper to demonstrate that the suboptimal aggregation thus constructed of atomic states $\{0, \ldots, 63\}$ was exact (an FSMP). The suboptimal aggregation for $N = 2$ is given below as (3.243)–(3.262).

Of interest, it was observed that each of the aggregated states in the suboptimal exact aggregation for $N = 2$ was a *subset* of one of the corresponding Choy and Beaulieu states Φ_0, Φ_1, Φ_2 and Γ, which are specified in (3.140)–(3.143) by setting $N = 2$. We note once more that the Choy and Beaulieu states apply only to their *steady–state* state space model (which is not an FSMP), whereas the states in the suboptimal exact aggregation discussed here apply to the *transient* state space model, derived from the transient atomic state space model of section 3.2.1. We are thus led by the observations of the $N = 1$ and $N = 2$ cases to the following proposition.

Proposition 3.4.1. *Consider the filtering–only atomic state space model of section 3.2.1 within the context of a communications system model employing BPSK signalling and a Volterra channel of memory $N \geq 0$. An exact aggregation of the atomic state space model exists that is related to the steady–state state space model of Choy and Beaulieu, specifically their single–distinct–errors model [24]. Moreover, the states of the exact aggregation of the atomic state space model are formed by taking certain set partitions of the states $\Phi_0, \ldots, \Phi_N, \Gamma$ of the Choy and Beaulieu model.*

Thus, for example, from (3.223) the aggregated states in figure 3.10 are the subsets $\mathcal{S}_0, \ldots, \mathcal{S}_5$ of $\{0, \ldots, 15\}$, where

$$\mathcal{S}_0 = \{0, 5\}, \tag{3.237}$$

$$\mathcal{S}_1 = \{1, 4\}, \tag{3.238}$$

$$\mathcal{S}_2 = \{2, 3, 6, 7\}, \tag{3.239}$$

$$\mathcal{S}_3 = \{8, 9, 12, 13\}, \tag{3.240}$$

$$\mathcal{S}_4 = \{10, 15\}, \quad \text{and} \tag{3.241}$$

$$\mathcal{S}_5 = \{11, 14\}. \tag{3.242}$$

Thus, as noted earlier, we observe that $\{\mathcal{S}_0, \mathcal{S}_4\}$ forms a set partition of the error–free states $\Gamma = \{0, 5, 10, 15\}$ of (3.121); that $\{\mathcal{S}_2, \mathcal{S}_3\}$ forms a set partition of the error–just–occurred states $\Phi_0 = \{2, 3, 6, 7, 8, 9, 12, 13\}$ of (3.122); and that $\{\mathcal{S}_0, \mathcal{S}_4\}$ forms a set partition of the partial–error–recovery states $\Phi_1 = \{1, 4, 11, 14\}$ of (3.123).

As a second example, for the case $N = 2$ it was observed by analogy with the $N = 1$ case that the set of atomic states $\{0, \ldots, 63\}$ could be partitioned into the exact suboptimal aggregation $\mathcal{S}_0, \ldots, \mathcal{S}_{19}$, where

$$\mathcal{S}_0 = \{0, 9\}, \tag{3.243}$$
$$\mathcal{S}_1 = \{1, 8\}, \tag{3.244}$$
$$\mathcal{S}_2 = \{2, 3, 10, 11\}, \tag{3.245}$$
$$\mathcal{S}_3 = \{4, 5, 12, 13\}, \tag{3.246}$$
$$\mathcal{S}_4 = \{6, 7, 14, 15\}, \tag{3.247}$$
$$\mathcal{S}_5 = \{16, 17, 24, 25\}, \tag{3.248}$$
$$\mathcal{S}_6 = \{18, 27\}, \tag{3.249}$$
$$\mathcal{S}_7 = \{19, 26\}, \tag{3.250}$$
$$\mathcal{S}_8 = \{20, 21, 28, 29\}, \tag{3.251}$$
$$\mathcal{S}_9 = \{22, 23, 30, 31\}, \tag{3.252}$$
$$\mathcal{S}_{10} = \{32, 33, 40, 41\}, \tag{3.253}$$
$$\mathcal{S}_{11} = \{34, 35, 42, 43\}, \tag{3.254}$$
$$\mathcal{S}_{12} = \{36, 45\}, \tag{3.255}$$
$$\mathcal{S}_{13} = \{37, 44\}, \tag{3.256}$$
$$\mathcal{S}_{14} = \{38, 39, 46, 47\}, \tag{3.257}$$
$$\mathcal{S}_{15} = \{48, 49, 56, 57\}, \tag{3.258}$$
$$\mathcal{S}_{16} = \{50, 51, 58, 59\}, \tag{3.259}$$
$$\mathcal{S}_{17} = \{52, 53, 60, 61\}, \tag{3.260}$$
$$\mathcal{S}_{18} = \{54, 63\}, \quad \text{and} \tag{3.261}$$
$$\mathcal{S}_{19} = \{55, 62\}. \tag{3.262}$$

In terms of the Choy and Beaulieu states Φ_0, Φ_1, Φ_2 and Γ, we observe that proposition 3.4.1 applies in that $\{\mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_8, \mathcal{S}_9, \mathcal{S}_{10}, \mathcal{S}_{11}, \mathcal{S}_{15}, \mathcal{S}_{16}\}$ is a set partition of the error–just–occurred states Φ_0; $\{\mathcal{S}_2, \mathcal{S}_5, \mathcal{S}_{14}, \mathcal{S}_{17}\}$ is a set partition of the partial–error–recovery states Φ_1; $\{\mathcal{S}_1, \mathcal{S}_7, \mathcal{S}_{13}, \mathcal{S}_{19}\}$ is a set partition of the partial–error–recovery states Φ_2; and $\{\mathcal{S}_0, \mathcal{S}_6, \mathcal{S}_{12}, \mathcal{S}_{18}\}$ is a set partition of the error–free states Γ.

For the case $N = 1$ with $p_{i,j}$ being the transition probability from aggregate state $\mathcal{S}_i$ to aggregate state $\mathcal{S}_j$ of (3.237)–(3.242), where $i, j \in \{0, \ldots, 5\}$, we have the (aggregate)

state transition probability matrix given by

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & p_{0,4} & p_{0,5} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} & p_{1,5} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} & p_{2,5} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} & p_{3,5} \\ p_{4,0} & p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} & p_{4,5} \\ p_{5,0} & p_{5,1} & p_{5,2} & p_{5,3} & p_{5,4} & p_{5,5} \end{bmatrix}$$
$$= \frac{1}{4} \begin{bmatrix} 1+a & 0 & 1-a & 1-a & 1+a & 0 \\ 1+a & 0 & 1-a & 1-a & 1+a & 0 \\ 0 & 1+b & 1-b & 1-c & 0 & 1+c \\ 0 & 1+c & 1-c & 1-b & 0 & 1+b \\ 1+a & 0 & 1-a & 1-a & 1+a & 0 \\ 1+a & 0 & 1-a & 1-a & 1+a & 0 \end{bmatrix}. \tag{3.263}$$

The stationary distribution vector π_∞ is therefore the solution to

$$\frac{1}{4} \begin{bmatrix} 1+a & 1+a & 0 & 0 & 1+a & 1+a \\ 0 & 0 & 1+b & 1+c & 0 & 0 \\ 1-a & 1-a & 1-b & 1-c & 1-a & 1-a \\ 1-a & 1-a & 1-c & 1-b & 1-a & 1-a \\ 1+a & 1+a & 0 & 0 & 1+a & 1+a \\ 0 & 0 & 1+c & 1+b & 0 & 0 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix}, \tag{3.264}$$

which is

$$\pi_\infty = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix} = \frac{1}{4(4+b+c-2a)} \begin{bmatrix} (1+a)(2+b+c) \\ (1-a)(2+b+c) \\ 4(1-a) \\ 4(1-a) \\ (1+a)(2+b+c) \\ (1-a)(2+b+c) \end{bmatrix}. \tag{3.265}$$

We can now compute the asymptotic probability of various quantities of interest, analogous to the calculations in (3.108)–(3.111). For instance, the probability of no current or past filtering errors ($\hat{x}_{t|t} = x_t$ and $\hat{x}_{t-1|t-1} = x_{t-1}$) is

$$\pi_0 + \pi_4 = \frac{(1+a)(2+b+c)}{2(4+b+c-2a)}, \tag{3.266}$$

which matches (3.109), obtained from the atomic state space model.

The adjacency matrix for the suboptimal model is

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}. \tag{3.267}$$

As before, by zeroing out rows 0 and 4 of A, corresponding to error–free states, we have the modified matrix

$$A^* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ \underline{\mathbf{0}} & 1 & 1 & 1 & \underline{\mathbf{0}} & 1 \\ \underline{\mathbf{0}} & 1 & 1 & 1 & \underline{\mathbf{0}} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}, \tag{3.268}$$

whose k–th powers give the number of paths to error recovery. Indeed, the highlighted entries in A^* correspond to a starting state i in either of the error–just–occurred aggregate states $\mathcal{S}_2 = \{2,3,6,7\}$ or $\mathcal{S}_3 = \{8,9,12,13\}$, and a finishing state j in either of the error–free aggregate states $\mathcal{S}_0 = \{0,5\}$ or $\mathcal{S}_4 = \{10,15\}$. We observe from figure 3.10 that the marked entries of 0 in (3.268) indicate that there are no 1–paths to error recovery, as we have to pass through an intermediate state first.

As before, taking successive powers of A^*, we find that the marked entries all follow the sequence 0, 2, 4, 16, 48, 160, 512, 1664, and so on, which matches the Horadam $(0,2,4,2)$ sequence [33], given by the recursion

$$a_{k+2} = 2a_{k+1} + 4a_k, \tag{3.269}$$

where $k \geq 1$, with the initial conditions $a_1 = 0$ and $a_2 = 2$. Note that this sequence is slightly different from that found to hold for the atomic state space model, namely (3.131).

3.4.6 Choy and Beaulieu's state space models

Choy and Beaulieu were concerned with error recovery in DFEs, and they used a state space model with states defined in terms of error patterns [24]. We shall illustrate this

model using an example problem involving a (linear or nonlinear) channel of memory $N = 4$, BPSK signalling, and filtering only ($n = 0$). This is the same problem that was used to generate the (transient) state space model depicted in figure 3.7, and connected with equations (3.146)–(3.151); and it was the same problem used to generate the (steady–state) state space model depicted in figure 3.8, and connected with equations (3.167)–(3.172).

For $N = 4$, Choy and Beaulieu's single–distinct–errors model involves the states Ψ_0, $\ldots$, Ψ_{N-1}, Φ_0, $\ldots$, Φ_{N-1} and Γ, as illustrated in table 3.13 below. Note that the states in this table are different from the states Φ_0, $\ldots$, Φ_{N-1} and Γ defined in (3.167)–(3.172) and portrayed in figures 3.7 and 3.8. An E in the column headed e_{t-k} denotes a filtering error ($\hat{x}_{t-k|t-k} \neq x_{t-k}$) at time $t-k$, where $k \in \{1, 2, \ldots\}$; an O in column e_{t-k} indicates no filtering error ($\hat{x}_{t-k|t-k} = x_{t-k}$); and an X means that we are not concerned whether there was a filtering error. Moreover, an error–free run of length $N-1$ is denoted by a string of $N-1$ consecutive O symbols, colour–coded blue. The complement of this is a run of length $N-1$ in which at least one error occurred, and this is denoted in the table by a string of $N-1$ consecutive X symbols, colour–coded grey.

Table 3.13. Choy and Beaulieu's single–distinct–errors model for a channel of memory $N = 4$ (linear or nonlinear), BPSK signalling, and filtering only ($n = 0$). State labels are in the leftmost column, and the remaining columns indicate error patterns. e_t is defined as the filtering error $x_t - \hat{x}_{t|t}$; the value itself is symbolized in the table by the coloured and uncoloured labels E, O and X.

state	e_{t-1}	e_{t-2}	e_{t-3}	e_{t-4}	e_{t-5}	e_{t-6}	e_{t-7}	e_{t-8}	$\cdots$	e_{t-k}	$\cdots$
Ψ_0	E	O	O	O	X	X	X	X	$\cdots$	X	$\cdots$
Ψ_1	O	E	O	O	O	X	X	X	$\cdots$	X	$\cdots$
Ψ_2	O	O	E	O	O	O	X	X	$\cdots$	X	$\cdots$
Ψ_3	O	O	O	E	O	O	O	X	$\cdots$	X	$\cdots$

Continued on next page

Table 3.13 – concluded from previous page

state	e_{t-1}	e_{t-2}	e_{t-3}	e_{t-4}	e_{t-5}	e_{t-6}	e_{t-7}	e_{t-8}	$\cdots$	e_{t-k}	$\cdots$
Φ_0	E	X	X	X	X	X	X	X	$\cdots$	X	$\cdots$
Φ_1	O	E	X	X	X	X	X	X	$\cdots$	X	$\cdots$
Φ_2	O	O	E	X	X	X	X	X	$\cdots$	X	$\cdots$
Φ_3	O	O	O	E	X	X	X	X	$\cdots$	X	$\cdots$
Γ	O	O	O	O	X	X	X	X	$\cdots$	X	$\cdots$

Figure 3.11 shows the state transition diagram for the single–distinct–errors state space model. The steady–state transition probabilities $\alpha_0, \ldots, \alpha_{N-1}$, $\beta_0, \ldots, \beta_{N-1}$ and ε are defined as follows [24, eqs. 12a, 12b]:

$$\alpha_i = \lim_{t\to\infty} \mathbb{P}(e_t = 0 | s_t = \Phi_i), \quad i \in \{0, \ldots, N-1\}, \tag{3.270}$$

$$\beta_i = \lim_{t\to\infty} \mathbb{P}(e_t = 0 | s_t = \Psi_i), \quad i \in \{0, \ldots, N-1\}, \tag{3.271}$$

$$\varepsilon = \lim_{t\to\infty} \mathbb{P}(e_t \neq 0 | s_t = \Gamma), \tag{3.272}$$

where s_t in this context refers to a particular error state, chosen from the set of states

$$\mathbf{S} = \{\Psi_0, \ldots, \Psi_{N-1}, \Phi_0, \ldots, \Phi_{N-1}, \Gamma\}. \tag{3.273}$$

Choy and Beaulieu note that the sequence $\{s_t\}$ is not a Markov chain. They define the time to error recovery R_0 as the number of time steps required to first reach state Γ, commencing in state Ψ_0, which is the set of atomic states with $\hat{x}_{t-1|t-1} \neq x_{t-1}$ and $\hat{x}_{t-k|t-k} = x_{t-k}$ for all $k \in \{2, \ldots, N\}$. Note the two *branches*, which we will call the β (upper) branch and the α (lower) branch. Ignoring cross–overs, we may think of error recovery as travel along either branch towards Γ. In the next chapter, we will provide experimental results that show how these branches may be used to quantify the error recovery behaviour of a given model. In particular, there exist interesting 'resonance' phenomena that are associated with barriers to error recovery.

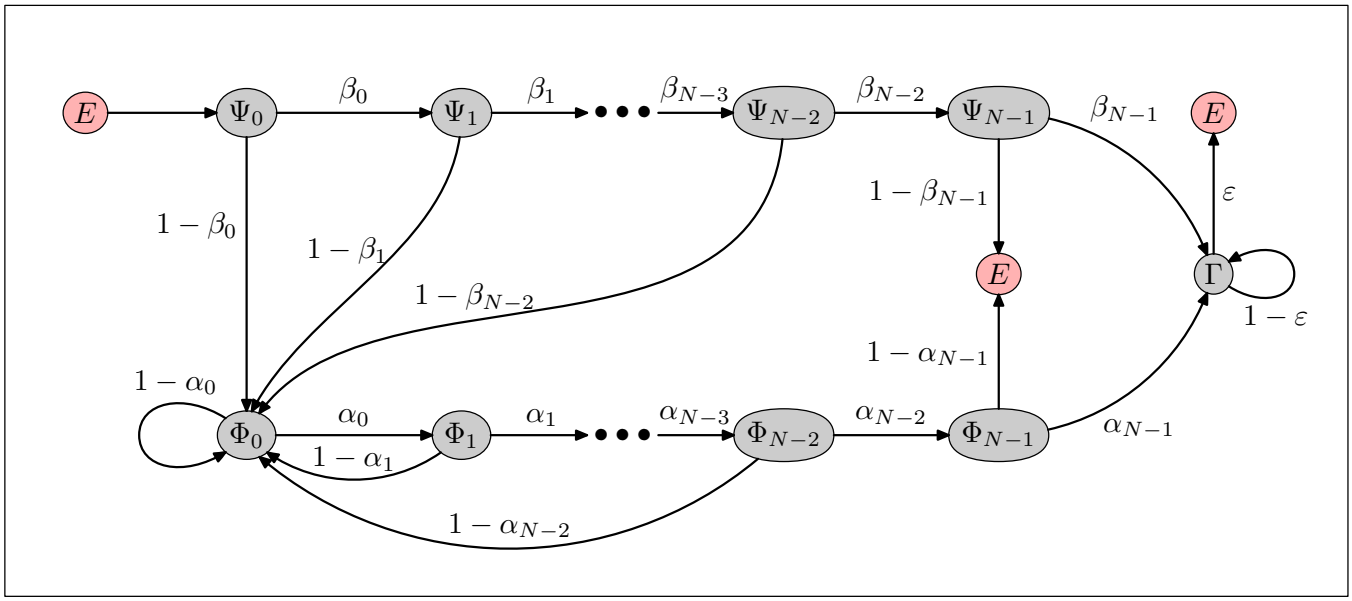

Figure 3.11. Choy and Beaulieu's single–errors state space model. The grey ovals labelled $\Psi_0, \ldots, \Psi_{N-1}, \Phi_0, \ldots, \Phi_{N-1}$ and Γ are sets of atomic states, these being $2N$–tuples $(x_{t-1}, \ldots, x_{t-N}, \hat{x}_{t-1|t-1}, \ldots, \hat{x}_{t-N|t-N}) \in \mathcal{A}_2^{2N}$ of previous symbols x_t and their FLSDFE estimates $\hat{x}_{t|t}$, up to the channel memory N.