Chapter 9

Error-Correcting Constrained Coding

In this chapter, we consider codes that have combined error-correction and constrained properties. We begin with a discussion of error mechanisms in recording systems and the corresponding error types observed. We then discuss associated metrics imposed on constrained systems—primarily the Hamming, Lee, and Euclidean metrics—and we survey the literature on bounds and code constructions. In addition, we consider two important classes of combined error-correction/constrained codes: spectral null codes and forbidden list codes.

9.1 Error-mechanisms in recording channels

Magnetic recording systems using peak detection, as described in Chapter 1 of this chapter, are subject to three predominant types of errors at the peak detector output. The most frequently observed error is referred to as a *bitshift error*, where a pair of recorded symbols 01 is detected as 10 (a *left bitshift*), or the pair 10 is detected as 01 (a *right bitshift*). Another commonly occurring error type is called a *drop-out error* or, sometimes, a *missing-bit error*, where a recorded symbol 1 is detected as a 0. Less frequently, a *drop-in error* or *extra-bit error* results in the detection of a recorded 0 as a 1. It is convenient to refer to the drop-in and drop-out errors as *substitution errors*.

Hamming-metric constrained codes are most pertinent in recording channels that behave like a binary symmetric channel, in which drop-in and drop-out errors occur with equal probability. However, there are alternative models of interest that suggest the use of codes designed with other criteria in mind beyond optimization of minimum Hamming distance. Among these models, the two that have received the most attention are: the asymmetric channel—where only drop-in errors or drop-out errors, but not both, are encountered; and the bitshift channel—where a symbol 1 is allowed to shift position by up to a prespecified number of positions.

Another error type we will consider is a *synchronization error*, resulting in an insertion or deletion of a symbol 0 in the detected symbol sequence. In practical digital recording systems on disks and tape, this type of error can have catastrophic consequences with regard to recovery of information that follows the synchronization loss. As a result, recording devices use synchronization and clock generation techniques in conjunction with code constraints, such as the k constraint in RLL codes for peak detection and the $\mathbf{G}$ constraint in PRML $(0, \mathbf{G}/\mathbf{I})$ codes, to effectively preclude such errors. Nevertheless, RLL-constrained synchronization-error-correcting codes have some intrinsic coding-theoretic interest, and we will discuss them below. Codes capable of correcting more than one insertion and deletion error may also be used to protect against bitshift errors, which result from the insertion and deletion of 0's on either side of a 1. The edit distance, or Levenshtein metric, and the Lee metric arise naturally in the context of synchronization errors.

In recording systems using partial-response with some form of sequence detection, exemplified by the PRML system described in Chapter 1, the maximum-likelihood detector tends to generate burst errors whose specific characteristics can be determined from the error events associated with the underlying trellis structure. We will briefly survey various trellis-coded modulation approaches for PRML that yield codes which combine $(0, \mathbf{G}/\mathbf{I})$ constraints with enhanced minimum Euclidean distance.

In practice, constrained codes must limit error propagation. Sliding-block decoders of the most frequently used (d, k)-RLL codes and PRML $(0, \mathbf{G}/\mathbf{I})$ codes typically will propagate a single detector error into a burst of length no more than eight bits. For example, the maximum error propagation of the industry standard $(2, 7)$-RLL and $(1, 7)$-RLL codes are four bits and five bits, respectively, and the PRML $(0, 4/4)$ code limits errors to a single byte. The conventional practice in digital recording devices is to detect and correct such errors by use of an outer error-correcting code, such as a Fire code, interleaved Reed-Solomon code, or a modification of such a code.

9.2 Gilbert-Varshamov-type lower bounds

9.2.1 Classical bound for the Hamming metric

There are several error metrics that arise in the context of digital recording using constrained sequences. For substitution-type errors and bitshift errors, possibly propagated into burst errors by the modulation decoder, it is natural to consider error-correcting codes based upon the *Hamming metric*. It is therefore of interest to investigate Hamming distance properties of constrained sequences.

The Gilbert-Varshamov bound provides for unconstrained sequences over a finite alphabet Σ a lower bound on the size of codes with prespecified minimum Hamming distance. In this section we present bounds of the Gilbert-Varshamov type and apply them to the class of runlength-limited binary sequences.

Let Σ denote a finite alphabet of size $|\Sigma|$ and denote the Hamming distance between two words $\mathbf{w}, \mathbf{w}' \in \Sigma^q$ by $\Delta_{\text{Hamming}}(\mathbf{w}, \mathbf{w}')$. For a word $\mathbf{w} \in \Sigma^q$, let $\mathcal{B}_{\Sigma^q}(\mathbf{w}; r)$ be the Hamming sphere of radius r in Σ^q centered at $\mathbf{w}$, that is,

$$\mathcal{B}_{\Sigma^q}(\mathbf{w}; r) = \{\mathbf{w}' : \Delta_{\text{Hamming}}(\mathbf{w}, \mathbf{w}') \leq r\} .$$

Let $\mathcal{V}_{\Sigma^q}(\mathbf{w}; r)$ be the cardinality or *volume* of the sphere $\mathcal{B}_{\Sigma^q}(\mathbf{w}; r)$. This quantity is $\sum_{i=0}^{r} \binom{q}{i} (|\Sigma|-1)^i$, independent of the center word $\mathbf{w}$, so we will use the shorthand notation $\mathcal{V}_{\Sigma^q}(r)$.

The Gilbert-Varshamov bound provides a lower bound on the achievable cardinality M of a subset of Σ^q with minimum Hamming distance at least d. We will refer to such a subset as a $(\Sigma^q, M, \mathsf{d})$-code.

Theorem 9.1 *There exists a* $(\Sigma^q, M, \mathsf{d})$*-code with*

$$M \geq \frac{|\Sigma|^q}{\mathcal{V}_{\Sigma^q}(\mathsf{d}-1)} .$$

For future reference, we recall that the proof of this bound is obtained by iteratively selecting the lth codeword $\mathbf{w}_l$ in the code from the complement of the union of Hamming spheres of radius $q-1$ centered at the previously selected codewords, $\Sigma^q - \cup_{i=1}^{l-1} \mathcal{B}_{\Sigma^q}(\mathbf{w}_i; \mathsf{d}-1)$. Continuing this procedure until the union of spheres exhausts Σ^q, an $(\Sigma^q, M, \mathsf{d})$-code is obtained whose size M satisfies the claimed inequality. □

Let $\delta = \mathsf{d}/q$ denote the *relative minimum distance* and let $H(\delta; z) = -\delta \cdot \log \delta - (1 - \delta) \cdot \log(1-\delta) + \delta \cdot \log(z-1)$, for $0 \leq \delta \leq 1 - (1/z)$, be a z-ary generalization of the entropy function. The Gilbert-Varshamov bound in terms of the rate R of the resulting code can be expressed as

$$R = \frac{\log M}{q} \geq \log|\Sigma| - \frac{\log \mathcal{V}_{\Sigma^q}(\mathsf{d}-1)}{q} \geq \log|\Sigma| - H(\delta; |\Sigma|)$$

(for the last inequality, we refer the reader to [Berl84, pp. 300–301].

9.2.2 Hamming-metric bound for constrained systems

Any generalization of the Gilbert-Varshamov bound to a constrained system S must take into account that the volumes of Hamming spheres in $S \cap \Sigma^q$ are not necessarily independent of the specified centers. Before deriving such bounds, we require a few more definitions. Let

X denote an arbitrary subset of Σ^q. For a word $\mathbf{w} \in X$, we define the Hamming sphere of radius r in X by

$$\mathcal{B}_X(\mathbf{w}; r) = \mathcal{B}_{\Sigma^q}(\mathbf{w}; r) \cap X \ .$$

The maximum volume of the spheres of radius r in X is

$$\mathcal{V}_{X,\max}(r) = \max_{\mathbf{w} \in X} |\mathcal{B}_X(\mathbf{w}; r)| \ ,$$

and the average volume of spheres of radius r in X is given by

$$\mathcal{V}_X(r) = \frac{1}{|X|} \sum_{\mathbf{w} \in X} |\mathcal{B}_X(\mathbf{w}; r)| \ .$$

We also define the set $\mathcal{B}_X(r)$ of pairs $(\mathbf{w}, \mathbf{w}')$ of words in X at distance no greater than r,

$$\mathcal{B}_X(r) = \{(\mathbf{w}, \mathbf{w}') \ : \ \Delta_{\text{Hamming}}(\mathbf{w}, \mathbf{w}') \le r\} \ .$$

Note that $|\mathcal{B}_X(r)| = \mathcal{V}_X(r) \cdot |X|$. Finally, we define an (X, M, d)-code to be a $(\Sigma^q, M, \mathsf{d})$-code that is a subset of X.

A straightforward application of the Gilbert-Varshamov construction yields the following result.

Lemma 9.2 *Let X be a subset of Σ^q and d be a positive integer. Then, there exists an (X, M, d)-code with*

$$M \ge \frac{|X|}{\mathcal{V}_{X,\max}(\mathsf{d}-1)} \ .$$

The following generalization of the Gilbert-Varshamov bound, first proved by Kolesnik and Krachkovsky [KolK91], is the basis for the more refined bounds derived later in the section. It provides a bound based upon the *average* volume of spheres, rather than the *maximum* volume, as was used in Lemma 9.2.

Lemma 9.3 *Let X be a subset of Σ^q and d be a positive integer. Then, there exists an (X, M, d)-code with*

$$M \ge \frac{|X|}{4\mathcal{V}_X(\mathsf{d}-1)} = \frac{|X|^2}{4|\mathcal{B}_X(\mathsf{d}-1)|} \ .$$

Proof. Consider the subset X' of words $\mathbf{w} \in X$ whose Hamming spheres of radius $\mathsf{d}-1$ satisfy $|\mathcal{B}_X(\mathbf{w}; \mathsf{d}-1)| \le 2\mathcal{V}_X(\mathsf{d}-1)$. The subset X' must then satisfy $|X'| \ge |X|/2$. If we iteratively select codewords from X', following the procedure used in the derivation of the Gilbert-Varshamov bound, we obtain an (X, M, d)-code, where

$$M \ge \frac{|X'|}{2\mathcal{V}_X(\mathsf{d}-1)} \ge \frac{\frac{1}{2}|X|}{2\mathcal{V}_X(\mathsf{d}-1)} = \frac{|X|^2}{4|\mathcal{B}_X(\mathsf{d}-1)|} \ ,$$

as desired. $\square$

In general, neither of the bounds in the preceding two lemmas is strictly superior to the other, as observed by Gu and Fuja [GuF93]. However, using an analysis of a new code search algorithm—dubbed the "altruistic algorithm" to distinguish it from the "greedy algorithm" that lies at the heart of the standard Gilbert-Varshamov form of bound—they eliminated the factor of 4 in the denominator of the bound in Lemma 9.3. This improved lower bound, stated below as Lemma 9.4, is always at least as good as the bound in Lemma 9.2, and a strict improvement over Lemma 9.3.

The key element of the improved code search algorithm is that, at each codeword selection step, the remaining potential codeword with the largest number of remaining neighbors at distance $\mathsf{d}-1$ or less takes itself out of consideration. As noted in [GuF93], a similar approach was developed independently by Ytrehus [Yt91a], who applied it to compute bounds for runlength-limited codes with various error detection and correction capabilities [Yt91b].

Lemma 9.4 *Let X be a subset of Σ^q and d be a positive integer. Then, there exists an (X, M, d)-code with*

$$M \geq \frac{|X|}{\mathcal{V}_X(\mathsf{d}-1)} = \frac{|X|^2}{|\mathcal{B}_X(\mathsf{d}-1)|} .$$

Kolesnik and Krachkovsky [KolK91] applied Lemma 9.3 to sets X consisting of words of length q in runlength-limited and charge constrained systems. Their asymptotic lower bound was based upon an estimate of the average volume of constrained Hamming spheres, whose centers ranged over all of $S \cap \Sigma^q$. Their estimate made use of a generating function for pairwise q-block distances in these families of constrained systems.

9.2.3 Improved Hamming-metric bounds

Marcus and Roth [MR92] found improved bounds by considering subsets X of $S \cap \Sigma^q$ where additional constraints, depending upon the designed relative minimum distance δ, are imposed upon the frequency of occurrence of code symbols $w \in \Sigma$. We now discuss the derivation of these bounds.

Let S be a constrained system over Σ presented by an irreducible deterministic graph $G = (V, E, L)$. Denote by $\Delta(G)$ the set of all stationary Markov chains on G (see Section 3.5). The entropy of $\mathcal{P} \in \Delta(G)$ is denoted by $\mathsf{H}(\mathcal{P})$.

Given a stationary Markov chain $\mathcal{P} \in \Delta(G)$, along with a vector of real-valued functions $\mathbf{f} = (f_1\ f_2\ \ldots\ f_t) : E_G \to \mathbb{R}^t$, we denote by $\mathsf{E}_{\mathcal{P}}(\mathbf{f})$ the expected value of $\mathbf{f}$ with respect to $\mathcal{P}$:

$$\mathsf{E}_{\mathcal{P}}(\mathbf{f}) = \sum_{e \in E_G} \mathcal{P}(e)\mathbf{f}(e) .$$

For a subset $W = \{w_1, w_2, \ldots, w_t\}$ of Σ, we define the vector indicator function $\mathcal{I}_W : E_G \to \mathbb{R}^t$ by $\mathcal{I}_W = (\mathcal{I}_{w_1}, \mathcal{I}_{w_2}, \ldots, \mathcal{I}_{w_t})$, where $\mathcal{I}_w : E_G \to \mathbb{R}$ is the indicator function for a symbol $w \in \Sigma$:

$$\mathcal{I}_w(e) = \begin{cases} 1 & \text{if } L_G(e) = w \\ 0 & \text{otherwise} \end{cases} .$$

Let $G \times G$ denote the labeled product graph defined by $V_{G\times G} = V_G \times V_G = \{(u, u') : u, u' \in V_G\}$ and $E_{G\times G} = E_G \times E_G$. There is an edge (e, e') in $G \times G$ from state (u, u') to state (v, v') whenever e is an edge in G from state u to state u' and e' is an edge in G from state v to v'. The labeling on $G \times G$ is defined by $L_{G\times G}(e, e') = (L_G(e), L_G(e'))$. We define on $E_{G\times G}$ the coordinate indicator functions $\mathcal{I}_W^{(1)}$ and $\mathcal{I}_W^{(2)}$, given by $\mathcal{I}_W^{(1)}((e, e')) = \mathcal{I}_W(e)$ and $\mathcal{I}_W^{(2)}((e, e')) = \mathcal{I}_W(e')$. Finally, we define the coordinate distance function $\mathcal{D} : E_{G\times G} \to \mathbb{R}$ by

$$\mathcal{D}((e, e')) = \begin{cases} 1 & \text{if } L_G(e) \neq L_G(e') \\ 0 & \text{otherwise} \end{cases} .$$

For a given symbol subset W of size t and a vector $\mathbf{p} \in [0, 1]^t$, we now define the quantities

$$\mathcal{S}_W(\mathbf{p}) = \sup_{\substack{\mathcal{P} \in \Delta(G) \\ \mathsf{E}_P(\mathcal{I}_W) = \mathbf{p}}} \mathsf{H}(\mathcal{P})$$

and

$$\mathcal{T}_W(\mathbf{p}, \delta) = \sup_{\substack{\mathcal{P}' \in \Delta(G\times G) \\ \mathsf{E}_{\mathcal{P}'}(\mathcal{I}_W^{(i)}) = \mathbf{p}\,,\ i=1,2 \\ \mathsf{E}_{\mathcal{P}'}(\mathcal{D}) \in [0,\delta]}} \mathsf{H}(\mathcal{P}') .$$

Finally, to concisely state the bounds, we introduce the following function of the relative designed distance δ:

$$R_W(\delta) = \sup_{\mathbf{p}\in[0,1]^t} \{ 2\mathcal{S}_W(\mathbf{p}) - \mathcal{T}_W(\mathbf{p}, \delta) \} .$$

The following theorem is proved in [MR92]. It is obtained by application of Lemma 9.3 to the words in $S \cap \Sigma^q$ generated by cycles in G starting and ending at a specified state $u \in G$, with frequency of occurrence of the symbols $w_i \in W$ given approximately by p_i, for $i = 1, 2, \ldots, t$.

Theorem 9.5 *Let S be a constrained system over Σ presented by a primitive deterministic graph, let $\delta > 0$, and let W be a subset of Σ of size t. Then there exist $(S\cap\Sigma^q, M, \delta q)$-codes satisfying*

$$\frac{\log M}{q} \geq R_W(\delta) - o(1) ,$$

where $o(1)$ stands for a term that goes to zero as q tends to infinity.

Computation of the quantities $\mathcal{S}_W(\mathbf{p})$ and $\mathcal{T}_W(\mathbf{p},\delta)$ requires the solution of a constrained optimization problem in which the objective function $\mathcal{P} \to \mathsf{H}(\mathcal{P})$ is concave, and the constraints are linear. The theory of convex duality based upon Lagrange multipliers provides a method to translate the problem into an unconstrained optimization with a convex objective function [MR92].

In order to reformulate the problem, we need to introduce a vector-valued matrix function that generalizes the adjacency matrix A_G. For a function $\mathbf{f} : E_G \to \mathbb{R}^t$ and $\mathbf{x} \in \mathbb{R}^t$, let $A_{G;\mathbf{f}}(\mathbf{x})$ be the matrix defined by

$$\Big(A_{G;\mathbf{f}}(\mathbf{x}) \Big)_{u,v} = \sum_{e\,:\,\sigma(e)=u,\,\tau(e)=v} 2^{-\mathbf{x}\cdot\mathbf{f}(e)}$$

We remark that for any function $\mathbf{f}$, the matrix $A_{G;\mathbf{f}}(\mathbf{0})$ is precisely the adjacency matrix of G.

The following lemma is the main tool in translating the constrained optimization problem to a more tractable form. It is a consequence of standard results in the theory of convex duality.

Lemma 9.6 *Let G and $\mathbf{f}$ be as above. Let $\mathbf{g} : E_G \to \mathbb{R}^l$, and define $\boldsymbol{\psi} = [\mathbf{f},\mathbf{g}] : E_G \to \mathbb{R}^{t+l}$. Then, for any $\mathbf{r} \in \mathbb{R}^t$ and $\mathbf{s} \in \mathbb{R}^l$,*

$$\sup_{\substack{\mathcal{P}\in\Delta(G)\\ \mathsf{E}_{\mathcal{P}}(\mathbf{f})=\mathbf{r}\\ \mathsf{E}_{\mathcal{P}}(\mathbf{g})\le\mathbf{s}}} \mathsf{H}(\mathcal{P}) = \inf_{\substack{\mathbf{x}\in\mathbb{R}^t\\ \mathbf{z}\in(\mathbb{R}^+)^l}} \Big\{ \mathbf{x}\cdot\mathbf{r} + \mathbf{z}\cdot\mathbf{s} + \log\lambda(A_{G;\boldsymbol{\psi}}(\mathbf{x},\mathbf{z})) \Big\} \; .$$

Applying Lemma 9.6 to Theorem 9.5, we can derive dual formulas for the lower bounds $R_W(\delta)$, for a specified symbol set W and relative minimum distance δ. For the case where W consists of a single symbol $w \in \Sigma$, the resulting formula is particularly tractable. To express it succinctly, we define

$$\mathcal{J}_w = \Big[\mathcal{I}^{(1)}_{\{w\}} + \mathcal{I}^{(2)}_{\{w\}}, \mathcal{D}\Big] : E_{G\times G} \to \mathbb{R}^2$$

and, to simplify notation, $\mathcal{S}_w(p) = \mathcal{S}_{\{w\}}\big((p)\big)$, $\mathcal{T}_w(p,\delta) = \mathcal{T}_{\{w\}}\big((p),\delta\big)$, and $R_w(\delta) = R_{\{w\}}(\delta)$.

The lower bound on attainable rates follows from the following theorem.

Theorem 9.7 *Let S be a constrained system over Σ presented by a primitive graph G, let $\delta > 0$, and let $w \in \Sigma$. Then*

a)

$$\mathcal{S}_w(p) = \inf_{x\in\mathbb{R}} \{px + \log\lambda(A_{G;\mathcal{I}_w}(x))\} \; ;$$

b)

$$\mathcal{T}_w(p,\delta) \;=\; \inf_{x\in\mathbb{R}, z\in\mathbb{R}^+} \{2px \;+\; \delta z \;+\; \log\lambda(A_{G\times G;\mathcal{J}_w}(x,z))\} \;;$$

c)

$$\begin{aligned} R_w(\delta) = \sup_{p\in[0,1]} \Big\{ & 2\inf_{x\in\mathbb{R}} \big\{px + \log\lambda(A_{G;\mathcal{J}_w}(x))\big\} \\ & - \inf_{x\in\mathbb{R},\, z\in\mathbb{R}^+} \big\{2px + \delta z + \log\lambda(A_{G\times G;\mathcal{J}_w}(x,z))\big\} \Big\} \;. \end{aligned}$$

In particular, if $\mathcal{P} \in \Delta(G)$ has maximal entropy rate

$$\mathsf{H}(\mathcal{P}) = \sup_{\mathcal{P}'\in\Delta(G)} \mathsf{H}(\mathcal{P}') \;,$$

and the symbol probability p equals $\mathsf{E}_{\mathcal{P}}(\mathcal{I}_w)$, then

$$\mathcal{S}_w(p) = \log\lambda(A_G)$$

and, setting $x = 0$ in part b) of Theorem 9.7,

$$\mathcal{T}_w(p,\delta) \le \inf_{z\in\mathbb{R}^+} \{\delta z + \log\lambda(A_{G\times G;\mathcal{J}_w}(0,z))\} \;.$$

From Theorem 9.5 and part c) of Theorem 9.7, we recover the lower bound of Kolesnik and Krachkovsky.

Corollary 9.8

$$\frac{\log M}{q} \ge R_{KK}(\delta) - o(1) \;,$$

where

$$R_{KK}(\delta) = 2\log\lambda(A_G) - \inf_{z\in\mathbb{R}^+} \{\delta z + \log\lambda(A_{G\times G;\mathcal{J}_w}(0,z))\} \;.$$

Better lower bounds can be obtained by prescribing the frequency of occurrence of words $\mathbf{w}$ of arbitrary length, rather than only symbols. See [MR92] for more details.

Example 9.1 For the $(0,1)$-RLL constrained system, consider the cases where $W = \{11\}$ and $W = \{111\}$, with corresponding lower bounds R_{11} and R_{111}. It is not difficult to see that $R_{11}(\delta)$ must equal $R_1(\delta)$. Table 9.1 from [MR92] gives the values of $R_{KK}(\delta)$, $R_1(\delta)$, and $R_{111}(\delta)$ for selected values of δ. □

We remark that, in some circumstances, one might assign to each edge $e \in E_G$ a *cost* associated to its use in a path generating a sequence in S. Lower bounds on the rate of codes into S with specified relative minimum distance δ and *average cost constraint* have been derived by Winick and Yang [WY93] and Khayrallah and Neuhoff [KN96].

δ	$R_{KK}(\delta)$	$R_1(\delta)$	$R_{111}(\delta)$
0.00	0.6942	0.6942	0.6942
0.05	0.4492	0.4504	0.4507
0.10	0.3055	0.3096	0.3109
0.15	0.2014	0.2094	0.2119
0.20	0.1241	0.1361	0.1399
0.25	0.0679	0.0831	0.0877
0.30	0.0295	0.0461	0.0506
0.35	0.0073	0.0218	0.0254
0.40	0	0.0077	0.0097
0.45	0	0.0013	0.0016
0.50	0	0	0

Table 9.1: Attainable rates for $(0,1)$-RLL constrained system.

9.3 Towards sphere-packing upper bounds

In comparison to lower bounds, much less is known about upper bounds on the size of block codes for constrained systems. We describe here a general technique introduced by Abdel-Ghaffar and Weber in [AW91]. Let S be a constrained system over Σ and let X be a nonempty subset of Σ^q. For a word $\mathbf{w} \in \Sigma^q$, denote by $\mathcal{B}_X(\mathbf{w};t)$ the set of words $\mathbf{w}' \in X$ which are at distance t or less from $\mathbf{w}$ according to some distance measure $\Delta(\mathbf{w},\mathbf{w}')$. If $\mathcal{C}$ is an $(S \cap \Sigma^q, M, \mathsf{d} = 2t{+}1)$ code, then, by the sphere-packing bound, we must have

$$\sum_{\mathbf{w} \in \mathcal{C}} |\mathcal{B}_X(\mathbf{w};t)| \le |X| \tag{9.1}$$

for any nonempty subset $X \subseteq \Sigma^q$. In the conventional sphere-packing bound, the subset X is taken to be the whole set Σ^q. Improved bounds may be obtained by taking X to be a proper subset of Σ^q. Specifically, define

$$N(S,X;i) = |\{\mathbf{w} \in S \cap \Sigma^q \,:\, |\mathcal{B}_X(\mathbf{w};t)| = i\}| \,.$$

Now, if X is contained in $\bigcup_{\mathbf{w} \in S \cap \Sigma^q} \mathcal{B}_{\Sigma^q}(\mathbf{w};t)$, then

$$\sum_{i=0}^{|X|} i N(S,X;i) \ge |X| \,,$$

so there exists an integer j, $1 \le j \le |X|$, such that

$$\sum_{i=0}^{j-1} i N(S,X;i) < |X| \,,$$

and

$$\sum_{i=0}^{j} i N(S,X;i) \ge |X| \,.$$

Abdel-Ghaffar and Weber [AW91] used these inequalities to establish the following upper bound on the code cardinality.

Theorem 9.9 *Let $\mathcal{C}$ be an $(S \cap \Sigma^q, M, \mathsf{d} = 2t+1)$ code and let X be a nonempty subset of $\bigcup_{\mathbf{w} \in S \cap \Sigma^q} \mathcal{B}_{\Sigma^q}(\mathbf{w}; t)$. Then*

$$|\mathcal{C}| \leq \sum_{i=0}^{j-1} N(S, X; i) + \left\lfloor \frac{|X| - \sum_{i=0}^{j-1} iN(S, X; i)}{j} \right\rfloor .$$

Proof. If $|\mathcal{C}| \leq \sum_{i=0}^{j-1} N(S, X; i)$, then we are done already. So, we may assume that $|\mathcal{C}| > \sum_{i=0}^{j-1} N(S, X; i)$. Divide $\mathcal{C}$ into two subsets $\mathcal{C}_1, \mathcal{C}_2$ where $\mathcal{C}_1$ consists of the $\sum_{i=0}^{j-1} N(S, X; i)$ elements $\mathbf{w}$ of $\mathcal{C}$ with the smallest $|\mathcal{B}_X(\mathbf{w}; t)|$. Then

$$\sum_{i=0}^{j-1} iN(S, X; i) \leq \sum_{\mathbf{w} \in \mathcal{C}_1} |\mathcal{B}_X(\mathbf{w}; t)|,$$

and

$$j(|\mathcal{C}| - \sum_{i=0}^{j-1} N(S, X; i)) \leq \sum_{\mathbf{w} \in \mathcal{C}_2} |\mathcal{B}_X(\mathbf{w}; t)|.$$

Now, use the preceding two inequalities to lower bound the left-hand side of inequality (9.1):

$$\sum_{i=0}^{j-1} iN(S, X; i) + j(|\mathcal{C}| - \sum_{i=0}^{j-1} N(S, X; i)) \leq \sum_{\mathbf{w} \in \mathcal{C}} |\mathcal{B}_X(\mathbf{w}; t)| \leq |X|.$$

The theorem follows from this. □

For the special case of bitshift errors, Abdel-Ghaffar and Weber obtain in [AW91] upper bounds on single-bitshift correcting codes $\mathcal{C}$ for (d, k)-RLL constrained systems S as follows. First, partition every code $\mathcal{C} \subseteq S \cap \Sigma^q$ into constant-weight subsets $\mathcal{C} = \cup_{\mathsf{w}} \mathcal{C}_{\mathsf{w}}$, such that each element of $\mathcal{C}_{\mathsf{w}}$ has Hamming weight w; then apply Theorem 9.9 to the subsets $\mathcal{C}_{\mathsf{w}}$, for suitably chosen sets X. Table 9.2 shows results for selected RLL constraints and codeword lengths.

Constructions of codes for channels with substitution, asymmetric, and bitshift errors, as well as bounds on the *maximum* cardinality of such codes of fixed length, have been addressed by numerous other authors, for example Blaum [Blaum91]; Ferreira and Lin [FL91]; Fredrickson and Wolf [FW64]; Immink [Imm91]; Kolesnik and Krachkovsky [KolK94]; Kuznetsov and Vinck [KuV93a], [KuV93b]; Lee and Wolf [Lee88], [LW87], [LW89]; Patapoutian and Kumar [PK92]; Shamai and Zehavi [SZ91]; and Ytrehus [Yt91a], [Yt91b].

q	$(d,k)=(2,7)$	$(d,k)=(3,10)$	$(d,k)=(4,12)$	$(d,k)=(5,15)$
3	1			
4	1	1		
5	1	1	1	
6	2	1	1	1
7	2	2	1	1
8	3	3	2	1
9	4	2	2	2
10	5	3	3	2
11	8	5	2	2
12	10	6	3	3
13	14	7	5	3
14	18	9	5	3
15	26	13	7	5
16	35	16	8	6
17	48	21	11	7
18	68	29	14	9
19	91	38	18	11
20	126	49	22	13
21	176	63	28	16
22	239	84	36	21
23	329	110	46	25
24	455	147	57	32
25	627	194	73	40
26	877	255	93	49
27	1204	335	117	61
28	1670	440	151	75
29	2302	581	193	95
30	3206	774	244	117
31	4464	1024	311	143
32	6182	1356	396	179

Table 9.2: Upper bounds on sizes of (d,k)-RLL constrained single shift-error correcting codes of length $3 \le q \le 32$.

9.4 Distance properties of spectral-null codes

Finally, we mention that spectral-null constrained codes—in particular, dc-free codes—with Hamming error-correction capability have received considerable attention. See, for example, Barg and Litsyn [BL91]; Blaum and van Tilborg [TiBl89]; Blaum, Litsyn, Buskens, and van Tilborg [BLBT93]; Calderbank, Herro, and Telang [CHT89]; Cohen and Litsyn [CL91]; Etzion [Etz90]; Ferreira [Fe84]; Roth [Roth93]; Roth, Siegel, and Vardy [RSV94]; Waldman and Nisenbaum [WN95]. Spectral-null codes also have inherent Hamming-distance properties, as shown by Immink and Beenker [ImmB87]. They considered codes over the alphabet $\{+1, -1\}$ in which the order-m moment of every codeword $\mathbf{x} = (x_1\ x_2\ \ldots\ x_\mathsf{n})$ vanishes for $m = 0, 1, \ldots, K-1$, i.e.,

$$\sum_{i=1}^{\mathsf{n}} i^m x_i = 0, \qquad m = 0, 1, \ldots, K-1 .$$

They referred to a code with this property as a code with order-$(K-1)$ zero-disparity. For each codeword $\mathbf{x}$, the discrete Fourier transform, given by $\Phi_\mathbf{x}(f) = \sum_{\ell=1}^{\mathsf{n}} x_\ell e^{-\jmath 2\pi f \ell}$, where $\jmath = \sqrt{-1}$, satisfies

$$\left. \frac{d^m \Phi_\mathbf{x}(f)}{df^m} \right|_{f=0} = 0 \qquad \text{for} \quad m = 0, 1, \ldots, K-1 .$$

This implies by part 8 of Problem 3.35 that the power spectral density of the ensemble of sequences generated by randomly concatenating codewords vanishes at $f = 0$, along with its order-ℓ derivatives for $\ell = 1, 2, \ldots, 2K-1$. A code, or more generally a constraint, with this property is said to have an *order-K spectral-null at $f = 0$*.

The following theorem, from [ImmB87], provides a lower bound on the minimum Hamming distance of a code with spectral null at $f = 0$.

Theorem 9.10 *Let $\mathcal{C}$ be a code with order-K spectral null at $f = 0$. Let $\mathbf{x}, \mathbf{y}$ be distinct codewords in $\mathcal{C}$. Then, their Hamming distance satisfies*

$$\Delta_{\text{Hamming}}(\mathbf{x}, \mathbf{y}) \geq 2K .$$

This result will play a role in the subsequent discussion of codes for the Lee and Euclidean metrics.

9.5 Synchronization/bitshift error correction

Synchronization errors, resulting from the insertion or deletion of symbols, and coding methods for protection against such errors have been the subject of numerous investigations. The *edit distance*, introduced by Levenshtein and often referred to as the *Levenshtein metric*, is particularly appropriate in this setting, as it measures the minimum number of symbol insertions and deletions required to derive one finite-length sequence from another. The reader interested in codes based upon the Levenshtein metric is referred to Bours [Bours94]; Iizuka, Hasahara, and Namahawa [IKN80]; Kruskal [Krusk83]; Levenshtein [Lev65], [Lev67], [Lev71], [Lev91]; Levenshtein and Vinck [LV93]; Tanaka and Kasai [TK76]; Tenengolts [Ten76], [Ten84]; and Ullman [U66], [U67].

When dealing with synchronization errors (insertions and deletions of 0's) in (d,k)-RLL constrained systems, it is convenient to represent a constrained sequence as a sequence of "runs," where a run corresponds to a symbol 1 along with the subsequent string of contiguous symbols 0 preceding the next consecutive symbol 1. We associate to each run a positive integer called the "runlength" representing the number of symbols in the run. As an example, the $(1,7)$-RLL sequence 10100000001000100 corresponds to the sequence of runs with runlengths 2, 8, 4, 3.

Let $\mathbf{w}$ be a (d,k)-constrained sequence with n runs and corresponding runlength sequence $\mathbf{s} = s_1, s_2, \ldots, s_{\mathsf{n}}$. Insertion of e symbols 0 in the jth run of $\mathbf{w}$ generates the sequence with runlengths $\mathbf{s}' = s_1, \ldots, s_j + e, s_{j+1}, \ldots, s_{\mathsf{n}}$, while deletion of e symbols 0 from run j generates the sequence of runlengths $\mathbf{s}' = s_1, \ldots, s_j - e, s_{j+1}, \ldots, s_{\mathsf{n}}$. (In the latter, e cannot exceed s_j.) An *e-synchronization error* denotes such a pattern of e insertions or deletions occurring within a single run. Note, also, that a bitshift error, or more generally, an e-bitshift error consisting of e left-bitshift errors or e right-bitshift errors occurring at the boundary between two adjacent runs, may be viewed as a pair of e-synchronization errors in consecutive runs, one being an insertion error, the other a deletion error.

This "runlength"-oriented viewpoint has been used in the design of RLL codes capable of detecting and correcting bitshift and synchronization errors. Hilden, Howe, and Weldon [Hild91] constructed a class of variable length codes, named Shift-Error-Correcting Modulation (SECM) codes, capable of correcting up to some prespecified number of random e-bitshift errors, for a preselected shift-error size e. The runlengths are regarded as elements of a finite alphabet F whose size, usually taken to be an odd prime integer, satisfies $k - d + 1 \geq |F| \geq 2e + 1$. The binary information string is viewed as a sequence of k runs $\mathbf{r} = r_1, r_2, \ldots, r_{\mathsf{k}}$, satisfying (d,k) constraints, with runlengths $\mathbf{s} = s_1, s_2, \ldots, s_{\mathsf{k}}$. The sequence of transition positions $\mathbf{t} = t_1, t_2, \ldots, t_{\mathsf{k}}$ is then defined by:

$$t_j = \sum_{i=1}^{j} s_i \pmod{|F|}, \quad \text{for} \quad j = 1, 2, \ldots, \mathsf{k}.$$

These values are then applied to a systematic encoder for an $[\mathsf{n}, \mathsf{k}, \mathsf{d}]$ BCH code over a finite field F of prime size, yielding parity symbols $t_{\mathsf{k}+1}, t_{\mathsf{k}+2}, \ldots, t_{\mathsf{n}}$. A (d,k)-constrained binary codeword is then generated by appending to the original information runs the sequence of parity runs $r_{\mathsf{k}+1}, r_{\mathsf{k}+2}, \ldots, r_{\mathsf{n}}$ whose runlengths $s_{\mathsf{k}+1}, s_{\mathsf{k}+2}, \ldots, s_{\mathsf{n}}$ satisfy

$$d + 1 \leq s_j < d + 1 + |F|, \quad \text{for} \quad j = \mathsf{k}{+}1, \mathsf{k}{+}2, \ldots, \mathsf{n}$$

and

$$t_j = \sum_{i=1}^{j} s_i \pmod{|F|}\,, \quad \text{for} \quad j = \mathsf{k}{+}1, \mathsf{k}{+}2, \ldots, \mathsf{n}\,.$$

In particular, for $|F| = 2t{+}1$, the resulting code may be used to correct up to t random 1-bitshift errors, where t is the designed error-correcting capability of the BCH code. Note that a similar construction provides for correction of random e-synchronization errors by encoding the runlengths themselves, rather than the transition positions. The interpretation of bitshift and, more generally, synchronization errors in terms of their effect on runlengths leads naturally to the consideration of another metric, the *Lee metric*.

The *Lee distance* $\Delta_{\text{Lee}}(x, y)$ of two symbols x, y in a finite field F of prime size is the smallest absolute value of any integer congruent to the difference $x - y$ modulo $|F|$. For vectors $\mathbf{x}$, $\mathbf{y}$ in F^{n}, the Lee distance $\Delta_{\text{Lee}}(\mathbf{x}, \mathbf{y})$ is the sum of the component-wise Lee distances. The *Lee weight* $w_{\text{Lee}}(\mathbf{x})$ of a vector $\mathbf{x}$ is simply $\Delta_{\text{Lee}}(\mathbf{x}, \mathbf{0})$, where $\mathbf{0}$ denotes the all-zero vector of length n.

Among the families of codes for the Lee-metric are the well-known negacyclic codes introduced by Berlekamp [Berl84, Ch. 9], the family of cyclic codes devised by Chiang and Wolf [CW71], and the Lee-metric BCH codes investigated by Roth and Siegel [RS92], [RS94].

All of these Lee-metric code constructions have the property that the redundancy required for correction of a Lee-metric error vector of weight t is approximately t symbols. In contrast, codes designed for the Hamming metric require approximately $2t$ check symbols to correct t random Hamming errors. In a recording channel subject to e-synchronization errors and e-bitshift errors, where the predominant errors correspond to small values of e, one might anticipate reduced overhead using a Lee-metric coding solution. This observation was made independently by Roth and Siegel [RS94], Saitoh [Sai93a], [Sai93b], and Bours [Bours94] (see also Davydov [Dav93] and Kabatiansky, Davydov, and Vinck [KDV92]), who have proposed a variety of constrained code constructions based on the Lee-metric, and have derived bounds on the efficiency of these constructions, as we now describe.

For bitshift-error correction, Saitoh proposed a construction yielding codes with fixed binary symbol length. He showed that the construction is asymptotically optimal with respect to a Hamming bound on the redundancy for single-bitshift error-correcting (d, k)-RLL codes.

The construction of Saitoh requires that the codewords begin with a symbol 1 and end with at least d symbols 0. The codewords will have a fixed number of runs and, consequently, a variable length in terms of binary symbols. The codewords are defined as follows. If the runlengths are denoted s_i, $i = 0, 1, \ldots, \mathsf{n}$, the sequence of runlengths s_i, for even values of i, comprise a codeword in a single-error correcting code over the Lee metric. The sequence of runlengths s_i, for $i \equiv 3 \pmod 4$, comprise a codeword in a single error-detecting code for the Lee metric. It is evident that, in the presence of a single bitshift error, the Lee-metric single error-correcting code will ensure correct determination of the runlengths s_i for even values of

i, indicating if the erroneous runlength, say s_{2j}, suffered an insertion or deletion of a symbol 0. The Lee-metric error-detecting code will then complete the decoding by determining if the corresponding deletion or insertion applies to runlength s_{2j-1} or s_{2j+1}.

In the broader context of synchronization errors, Roth and Siegel described and analyzed a construction of (d,k)-RLL codes for detection and correction of such errors as an application of a class of Lee-metric BCH codes [RS92]. The shortened BCH code of length n over a finite prime field F, denoted $\mathcal{C}(\mathsf{n}, r; F)$, is characterized by the parity-check matrix

$$H(\mathsf{n}, r; F) \equiv \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \beta_1 & \beta_2 & \ldots & \beta_{\mathsf{n}} \\ \beta_1^2 & \beta_2^2 & \ldots & \beta_{\mathsf{n}}^2 \\ \vdots & \vdots & \ldots & \vdots \\ \beta_1^{r-1} & \beta_2^{r-1} & \ldots & \beta_{\mathsf{n}}^{r-1} \end{pmatrix},$$

where $(\beta_1 \; \beta_2 \; \ldots \; \beta_{\mathsf{n}})$ is the *locator vector*, consisting of distinct nonzero elements of the smallest h-dimensional extension field F_h of F of size greater than n.

Hence, a word $\mathbf{x} = (x_1 \; x_2 \; \ldots \; x_{\mathsf{n}}) \in F^{\mathsf{n}}$ is in $\mathcal{C}(\mathsf{n}, r; F)$ if and only if it satisfies the following r parity-check equations over F_h:

$$\sum_{i=1}^{\mathsf{n}} x_i \beta_i^m = 0 \;, \quad \text{for} \quad m = 0, 1, \ldots, r-1 \;.$$

The following theorem provides a lower bound on the minimum Lee distance of $\mathcal{C}(\mathsf{n}, r; F)$, denoted $\mathsf{d}_{\text{Lee}}(\mathsf{n}, r; F)$.

Theorem 9.11

$$\mathsf{d}_{\text{Lee}}(\mathsf{n}, r; F) \geq \begin{cases} 2r & \text{for } r \leq (|F|-1)/2 \\ |F| & \text{for } (|F|+1)/2 \leq r < |F| \end{cases} .$$

This bound follows from Newton's identities [ImmB87],[KS91a] and can be regarded, in a way, as the analogue of the BCH lower bound $r{+}1$ on the minimum Hamming distance of $\mathcal{C}(\mathsf{n}, r; F)$, although the proof of the $2r$ lower bound is slightly more complicated. For $r \geq |F|$ we can bound $\mathsf{d}_{\text{Lee}}(\mathsf{n}, r; F)$ from below by the minimum Hamming distance $r{+}1$.

The $2r$ lower bound does not hold in general for all values of r; however, it does hold for all r in the base-field case $\mathsf{n} \leq |F|{-}1$. The $2r$ lower bound for the base-field case takes the following form.

Theorem 9.12 *For $r \leq n \leq |F|{-}1$,*

$$\mathsf{d}_{\text{Lee}}(\mathsf{n}, r; F) \geq 2r \;.$$

The *primitive* case corresponds to codes $\mathcal{C}(\mathsf{n}, r; F)$ for which $\mathsf{n} = |F|^h - 1$. The redundancy of such codes is known to be bounded from above by $1 + (r-1)h$. This bound, along with the following lower bound derived by a sphere-packing argument, combine to show that the primitive codes are near-optimal for sufficiently small values of r.

Lemma 9.13 (Sphere-packing bound, Golomb and Welch [GoW68], [GoW70]) *A code over a finite prime field F of length n, size M, and minimum Lee distance $\geq 2r-1$ for some $r \leq (|F|+1)/2$ must satisfy the inequality*

$$M \cdot \sum_{i=0}^{r-1} 2^i \binom{n}{i} \binom{r-1}{i} \leq |F|^{\mathsf{n}} .$$

Theorem 9.14 *A code over a finite prime field F of length n, size M, and minimum Lee distance $\geq 2r-1$ for some $r \leq (|F|+1)/2$ must satisfy the inequality*

$$(r-1)\left(\log_{|F|}(\mathsf{n}-r+2) - \log_{|F|}(r-1)\right) \leq \mathsf{n} \;-\; \log_{|F|} M .$$

Proof. By Lemma 9.13 we have

$$\frac{(\mathsf{n}-r+2)^{r-1}}{(r-1)^{r-1}} \cdot 2^{r-1} \leq |F|^{\mathsf{n}}/M .$$

The theorem now follows by taking the logarithm to base $|F|$ of both sides of this inequality.□

The construction of synchronization-error correcting codes based upon the Lee-metric BCH codes is as follows. Given constraints (d, k), we choose $|F| \leq k-d+1$. We regard every run of length s in the (d, k)-constrained information sequence as an element $(s-d-1) \pmod{|F|}$ of F, and use a systematic encoder for $\mathcal{C}(\mathsf{n}, r; F)$ to compute the corresponding check symbols in F. Each check symbol a, in turn, is associated with a run of length $\overline{a} + d + 1$, where $\overline{a}$ is the smallest nonnegative integer such that $a = \overline{a} \cdot 1$, where 1 stands for the multiplicative unity in F. The code $\mathcal{C}(\mathsf{n}, r; F)$, with $r \leq (|F|-1)/2$ and $\mathsf{n} \leq |F|^h - 1$ can simultaneously correct b bitshift errors and s non-bitshift synchronization errors whenever $2b + s < r$. (Observe that, when counting errors, an e-bitshift error is counted as e bitshift errors; this applies respectively also to synchronization errors. Also, bitshift or synchronization errors may create runlengths that violate the (d, k)-constraint. In such a case we can mark the illegal runlength as an erasure rather than an error.) The redundancy required will be no more than $1 + (r-1)h$ symbols from the alphabet F, and we recall that Theorem 9.14 indicates the near-optimality of the Lee-metric primitive BCH codes $\mathcal{C}(|F|^h - 1, r; F)$, for values $r \ll |F|^h - 1$.

Example 9.2 Two typical choices for parameters (d, k) are $(1, 7)$ and $(2, 8)$, both satisfying $k - d + 1 = 7$. Setting $|F| = 7$ and $r = 3$, we obtain a family of codes for these constraints, based upon $\mathcal{C}(\mathsf{n}, 3; 7)$, that can correct any error pattern of Lee weight 2 (and detect error patterns of Lee weight 3). In particular, the codes will correct one single-bitshift (1-bitshift) error or any other combination of two insertions/deletions of symbols 0. For $\mathsf{n} \leq |F|^h - 1$, the required redundancy is no more than $1 + 2h$ symbols. □

As mentioned above, the class of Hamming-metric SECM codes are directed primarily toward the situation when only bitshift-type errors occur. The constructions based upon Lee-metric codes can be modified to improve their efficiency in this type of error environment by recording, instead of the nominal codeword $\mathbf{x} = (x_1\ x_2\ \ldots\ x_{\mathsf{n}})$, the differentially precoded word $\mathbf{y} = (y_1\ y_2\ \ldots\ y_{\mathsf{n}})$ defined by $y_1 = x_1$ and $y_i = x_i - x_{i-1}$ for $2 \leq i \leq \mathsf{n}$, where all operations are taken modulo $|F|$. If $\mathbf{y}$ is recorded, and no bitshift errors occur, the original word $\mathbf{x}$ is reconstructed by an "integration" operation:

$$x_i = \sum_{j=1}^{i} y_j \ .$$

If, however, an e-bitshift error occurs at the boundary between runs j and $j+1$ of $\mathbf{y}$, the integration operation converts the error into an e-synchronization error in run j of $\mathbf{x}$. In other words, the original bitshift error pattern of Lee weight $2e$ is converted into a synchronization error pattern of Lee weight e. In order to ensure the correctness of the first run y_1, it suffices to require that the code contain the all-one word $(1\ 1\ \ldots\ 1)$ and all of its multiples.

For the Lee-metric BCH codes, this construction provides the capability to correct up to $r-1$ bitshift errors and detect up to r bitshift errors, when $2r < |F| \leq k-d+1$. The construction can be extended to the base-field case as well.

Example 9.3 Let $|F| = 7$ and $r = 3$ as in the previous example. The construction above will generate codes with length n a multiple of 7. For $\mathsf{n} = 7$, the redundancy is $1+(r-1) = 3$ runs; for $\mathsf{n} = 14, 21, \ldots, 49$ the redundancy is $1+2(r-1) = 5$ runs; for $\mathsf{n} = 56, 63, \ldots, 343$ the redundancy is $1 + 3(r-1) = 7$ runs. All of these codes will correct up to two single-bitshift errors or one double-bitshift (2-bitshift) error. By way of comparison, in [Hild91] Hilden et al. describe SECM codes of lengths 26, 80, and 242 for correcting two single-bitshift errors, requiring redundancy of 7, 9, and 11 runs, respectively. These SECM codes do not handle double-bitshift errors. □

Example 9.4 As $|F|$ increases, so does the discrepancy in the number of check symbols (runs) compared to the SECM codes in [Hild91]. For $|F| = 11$, suitable for representing $(d, k) = (1, 11)$ for example, and $r = 5$, the Lee-metric BCH code with $\mathsf{n} = 11$ requires 5 check symbols; for $\mathsf{n} = 22, 33, \ldots, 121$, the redundancy is 9 symbols; for $\mathsf{n} = 132, 143, \ldots, 1331$ the redundancy will be 13 symbols. These codes will correct up to four single-bitshift errors; two single-bitshift and one double-bitshift errors; or two double-bitshift errors. The codes presented in [Hild91] for correcting up to four single-bitshift errors have lengths 26, 80, and 242 and require redundancy of 16, 21, and 26, respectively. □

Bours [Bours94] provided a construction of synchronization-error correcting RLL codes with *fixed length* over the binary alphabet that also relies on an underlying Lee-metric code. He did not require the underlying code to be a Lee-metric BCH code, however, and thereby avoided having the error-correction capability limited by the code alphabet size.

The definition of the Lee metric can also be generalized in a straightforward manner to integer rings. Orlitsky described in [Or93] a nonlinear construction of codes over the ring of integers modulo 2^h for correcting any prescribed number of Lee errors. His construction is based on dividing a codeword of a binary BCH code into nonoverlapping h-tuples and regarding the latter as the Gray-code representations of the integers between 0 and $2^h - 1$.

It is also worth remarking that all of the Lee-metric codes mentioned above can be efficiently decoded algebraically.

We close the discussion of Lee-metric codes by noting that the definition of the class of Lee-metric BCH codes was motivated by a Lee-metric generalization of the result of Immink and Beenker in Theorem 9.10 to integer-valued spectral-null constraints [KS91a], [EC91].

Theorem 9.15 *Let S be a constrained system over an integer alphabet with order-K spectral null at $f = 0$, presented by a labeled graph G. Let $\mathbf{x}, \mathbf{y}$ be distinct sequences in S generated by paths in G, both of which start at a common state u and end at a common state v. Then, the Lee distance satisfies*

$$\Delta_{\mathrm{Lee}}(\mathbf{x}, \mathbf{y}) \geq 2K \ .$$

This result will play an important role in the next section in the context of Euclidean-metric codes for PRML.

When combining bitshift and synchronization errors, any bitshift error can obviously be regarded as two consecutive synchronization errors in opposite directions – one e-insertion, one e-deletion – thus reducing to the synchronization-only model of errors. However, such an approach is not optimal, and better constructions have been obtained to handle a limited number of bitshift and synchronization errors (combined). See Hod [Hod95], Kløve [Kl95], and Kuznetsov and Vinck [KuV93a], [KuV93b].

9.6 Soft-decision decoding through Euclidean metric

Let $\mathbf{x}$ and $\mathbf{y}$ be sequences of length n over the real numbers. The *squared-Euclidean distance* between these sequences, denoted $\Delta^2_{\mathrm{Euclid}}(\mathbf{x}, \mathbf{y})$ is given by

$$\Delta^2_{\mathrm{Euclid}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\mathsf{n}} (x_i - y_i)^2 \ .$$

The Euclidean metric is most relevant in channels with additive white Gaussian noise (AWGN). In particular, it is of interest in connection with the model of the magnetic recording channel as a binary input, partial-response system with AWGN. The success of trellis-coded modulation, as pioneered by Ungerboeck, in improving the reliability of memoryless channels with AWGN provided the impetus to design coding schemes for channels with memory, such as partial-response channels, in AWGN. For binary input-restricted partial-response channels suitable as models for recording channels, such as the Class-4 channel characterized by the input-output relation $y_i = x_i - x_{i-2}$, several approaches have been proposed that make use of binary convolutional codes. These approaches typically require a computer search of some kind to determine the codes that are optimal with respect to rate, Euclidean distance, and maximum-likelihood detector complexity. See, for example, Wolf and Ungerboeck [WU86]; Calderbank, Heegard, and Lee [CHL86], Hole [Hole91]; and Hole and Ytrehus [HoY94].

There is another approach, however, that relies upon the concepts and code construction techniques that have been developed in the previous chapters. The underlying idea is as follows. First, find a constrained system S, presented by a labeled graph G, that ensures a certain minimum Euclidean distance between the partial-response channel output sequences generated when channel inputs are restricted to words in S. Then, apply state-splitting (or other) methods to construct an efficient encoder from binary sequences to S. Since the graph structure $\mathcal{E}$ underlying the encoder is often more complex (in terms of number of states and interconnections) than the original graph G, use G rather than $\mathcal{E}$ as the starting point for the trellis-based Viterbi detector of the coded channel.

Karabed and Siegel [KS91a] showed that this approach can be applied to the family of constrained systems S whose spectral null frequencies coincide with those of the partial-response channel transfer function. The resulting codes are referred to as *matched-spectral-null codes*. We conclude this section with a brief summary of the results that pertain to the application of this technique to the Class-4 and related partial-response systems. We will refer to the *dicode* channel, which is characterized by the input-output relation $y_i = x_i - x_{i-1}$, and has a first-order spectral null at $f = 0$, and we remark that the Class-4 partial-response channel may be interpreted as a pair of interleaved dicode channels, one operating on inputs with even indices, the other on inputs with odd indices.

Lemma 9.16 *Let S be a constrained system over an integer alphabet with order-K spectral null at zero frequency. Let S' be the constrained system of sequences at the output of a cascade of N dicode channels, with inputs restricted to words in S. Then, S' has an order-$(K+N)$ spectral null at zero frequency.*

Noting that any lower bound on Lee distance provides a lower bound on squared-Euclidean distance for sequences over integer alphabets, we obtain from the preceding lemma and Theorem 9.15 the following lower bound on the minimum squared-Euclidean distance of a binary, matched-spectral-null coded, partial-response channel with spectral null at $f = 0$.

Theorem 9.17 *Let S be a constrained system over the alphabet $\{+1, -1\}$, with order-K spectral null at zero frequency. Let $\mathbf{x}$ and $\mathbf{y}$ be distinct sequences in S, differing in a finite number of positions. If $\mathbf{x}'$ and $\mathbf{y}'$ are the corresponding output sequences of a partial-response channel consisting of a cascade of N dicode channels, then*

$$\Delta^2_{\mathrm{Euclid}}(\mathbf{x}, \mathbf{y}) \geq 8(K + N) \ .$$

It is easy to see that the lower bound of Theorem 9.17 remains valid in the presence of J-way, symbol-wise interleaving of the constrained sequences and the partial-response channel. In particular, for the Class-4 partial-response channel (i.e., the 2-way interleaved dicode channel), the application of sequences obtained by 2-way interleaving a code having a first-order spectral null at zero frequency doubles the minimum squared-Euclidean distance at the channel output, relative to the uncoded system.

We remark that J-way interleaving of sequences with an order-K spectral null at $f = 0$ generates sequences with order-K spectral nulls at frequencies $f = r/J$, for $r = 0, 1, \ldots, J-1$ [MS87]. Thus, a 2-way interleaved, dc-free code has spectral nulls at zero frequency and at frequency $f = 1/2$, corresponding to the spectral null frequencies of the Class-4 partial-response channel.

Graph presentations for spectral null sequences are provided by *canonical diagrams*, introduced by Marcus and Siegel [MS87] for first-order spectral null constraints and then extended to high-order constraints by, among others, Monti and Pierobon [MPi89]; Karabed and Siegel [KS91a]; Eleftheriou and Cideciyan [EC91]; and Kamabe [Kam94].

Discussion of the canonical diagrams requires the notion of a labeled graph with an infinite number of states. Specifically, a *countable-state labeled graph* $G_\infty = (V, E, L)$ consists of a countably-infinite set of states V; a set of edges E, where each edge has an initial state and a terminal state, both in V, and the states in V have bounded out-degree and in-degree; and an edge labeling $L : E \to \Sigma$, where Σ is a finite alphabet.

We say that a countable-state graph G_∞ is a *period-*p *canonical diagram* for a spectral null constraint if:

1. Every finite subgraph $H \subset G_\infty$ generates a set of sequences with the prescribed spectral null constraint.
2. For any period-p graph G'_∞ that presents a system with the specified spectral null constraint, there is a label-preserving graph homomorphism of G'_∞ into G_∞, meaning a map from the edges of G'_∞ to the edges of G_∞ that preserves initial states, terminal states, and labels.

The canonical diagram G_∞ for a first-order spectral null constraint at zero frequency, with symbol alphabet $\{+1, -1\}$, is shown in Figure 9.1. As mentioned in Example 3.4,

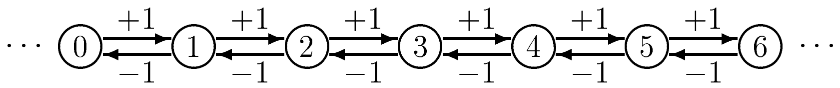

Figure 9.1: Canonical diagram for first-order spectral null at $f = 0$.

the capacity of the constrained system generated by a subgraph G_B, consisting of $B{+}1$ consecutive states and the edges with beginning and ending states among these, is given by

$$\mathsf{cap}(S(G_B)) = \log\left(2\cos\frac{\pi}{B{+}2}\right) .$$

From this expression, it follows that

$$\lim_{B\to\infty} \mathsf{cap}(S(G_B)) = 1 .$$

We pointed out in Chapter 2 that the constrained system generated by any finite subgraph of G_∞ is almost-finite-type. Applying Theorem 4.12, we see that by choosing B large enough, we can construct a sliding-block decodable finite-state encoder for the constrained system $S(G_B)$ at any prespecified rate $p : q$ with $p/q < 1$.

From the structure of the canonical diagram, it is clear that any constrained system with first-order spectral null at $f = 0$ limits the number of consecutive zero samples at the output of the dicode channel. When the constrained sequences are twice-interleaved and applied to the Class-4 partial-response channel, the number of zero samples at the output is limited "globally" as well as in each of the even/odd interleaved subsequences. This condition is analogous to that achieved by the $(0, \mathbf{G/I})$ constraints for the baseline PRML system.

The subgraph G_B chosen for the code construction may be augmented to incorporate the dicode channel memory, as shown in Figure 9.2 for the case $M = 6$, providing the basis for a dynamic programming (Viterbi) detection algorithm for the coded-dicode channel with AWGN. Each state in the trellis has a label of the form v^m, where v is the state in Figure 9.1 from which it is derived, and the superscript m denotes the sign of the dicode channel memory. Just as does the uncoded dicode detector graph, represented by the trellis in Figure 1.20 of Chapter 1, the coded-dicode detector graph supports sequences that can cause potentially unbounded delays in the merging of survivor sequences and, therefore, in decoding. The spectral-null code sequences that generate these output sequences are called *quasicatastrophic sequences*, and they are characterized in the following proposition.

Proposition 9.18 *The quasicatastrophic sequences in the constrained system presented by G_B are those generated by more than one path in G_B.*

To limit the merging delay, the matched-spectral-null code is designed to avoid these sequences, and it is shown in [KS91a] that this is always possible without incurring a rate loss for any G_B, with $B \geq 3$.

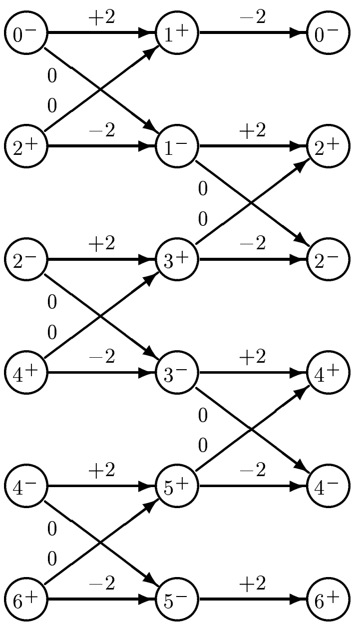

Figure 9.2: Graph underlying coded-dicode channel Viterbi detector for G_6.

Further details and developments regarding the design and application of matched-spectral-null codes to PRML systems may be found in [Shung91], [Thap92], [Thap93], [Fred94], and [Rae94].

9.7 Forbidden list codes for targeted error events

This section (which is yet to be written) will be based on results taken from Karabed-Siegel-Soljanin [KSS00].

Problems

Problem 9.1 A graph G is called *binary* if its labels are over the alphabet $\{0,1\}$. The *(Hamming) weight* of a word $\mathbf{w}$ generated by a binary graph G is the number of 1's in $\mathbf{w}$. The weight of a path in a binary graph G is the weight of the word generated by that path.

Given a binary graph G and states u and v in G, denote by $\tau^{(\ell)}_{u,v,k}$ the number of paths of length ℓ and weight k in G that originate in u and terminate in v. For states u and v in G, define the *length-ℓ weight-distribution polynomial* (of paths from u to v), in the indeterminate z, by

$$P^{(\ell)}_{u,v}(z) = \sum_{k=0}^{\ell} \tau^{(\ell)}_{u,v,k} z^k \ .$$

As an example, for the graph H in Figure 9.3, $P^{(4)}_{A,C} = z^2 + 2z^3$, since there are three paths of length 4 that originate in A and terminate in C: one path has weight 2, and the other paths each has weight 3.

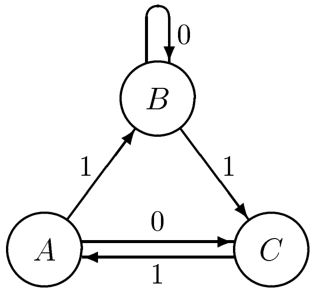

Figure 9.3: Graph H for Problem 9.1.

For a binary graph $G = (V, E, L)$, let $B_G(z)$ be the $|V| \times |V|$ matrix in the indeterminate z, where

$$(B_G(z))_{u,v} = P^{(1)}_{u,v}(z)$$

for every $u, v \in V$. Each entry in $B_G(z)$ is therefore a polynomial in z of degree at most 1.

1. Compute $B_H(z)$ for the graph in Figure 9.3.
2. For the matrix $B_H(z)$ found in 1, compute $(B_H(z))^2$ and $(B_H(z))^4$.
3. Let $B_G(z)$ be the matrix associated with a binary graph G, and let u and v be states in G. Given a positive integer ℓ, obtain an expression for the polynomial $P^{(\ell)}_{u,v}(z)$ in terms of $B_G(z)$.
4. Identify the matrix $B_G(1)$ associated with a binary graph G.
5. Identify the matrix $B_G(0)$.
6. Let G be a binary graph and let z_0 be a positive real number. Show that G is irreducible if and only if the matrix $B_G(z_0)$ is irreducible. Does this hold also when $z_0 = 0$?

Problem 9.2 Recall the definitions from Problem 9.1. Let G be a binary *lossless* graph and let u and v be states in G. For positive integers ℓ and d, denote by $J^{(\ell)}_{u,v,d}$ the number of words of length ℓ and weight $\le d$ that can be generated in G by paths that originate in u and terminate in v.

1. Show that for every $0 \leq d \leq \ell$,

$$J^{(\ell)}_{u,v,d} = \sum_{k=0}^{d} \tau^{(\ell)}_{u,v,k} .$$

2. Based on 1, show that for every real z in the range $0 < z \leq 1$,

$$J^{(\ell)}_{u,v,d} \leq \sum_{k=0}^{\ell} \tau^{(\ell)}_{u,v,k} z^{k-d}$$

and, therefore,

$$J^{(\ell)}_{u,v,d} \leq \min_{0 \leq z \leq 1} z^{-d} P^{(\ell)}_{u,v}(z) .$$

3. Based on 2 and 3, derive an upper bound on the number of words of length ℓ and weight $\leq d$ in $S(G)$, as a function of $B_G(z)$, ℓ, and d.

Problem 9.3 Recall the definitions from Problem 9.1. Let S be a constrained system presented by a deterministic binary graph G with finite memory $\mathcal{M}$. For nonnegative integers ℓ and k, denote by $Y_k^{(\ell)}$ the number of (ordered) pairs $(\mathbf{w}, \mathbf{w}')$ of words of length ℓ in S such that $\mathbf{w}$ and $\mathbf{w}'$ are at Hamming distance k; i.e., they differ on exactly k locations. Define the polynomial $Y^{(\ell)}(z)$ by

$$Y^{(\ell)}(z) = \sum_{k=0}^{\ell} Y_k^{(\ell)} z^k .$$

1. Show that $Y^{(\ell)}(0) = |S \cap \{0,1\}^\ell|$.

2. Show that $Y^{(\ell)}(1) = (Y^{(\ell)}(0))^2$.

3. Let $S_{0,1}$ denote the $(0,1)$-RLL constrained system. Show that when $S = S_{0,1}$, the polynomial $Y^{(\ell)}(z)$ can be written as

$$Y^{(\ell)}(z) = (1\ z\ z\ 1)\,(B_{G*G}(z))^{\ell-1}\,\mathbf{1} ,$$

where $\mathbf{1}$ is the all-one column vector and

$$B_{G*G}(z) = \begin{pmatrix} 1 & z & z & 1 \\ 1 & 0 & z & 0 \\ 1 & z & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

4. Compute $Y_4(z)$ explicitly for $S = S_{0,1}$.

5. Find the largest integer n for which there exists a block $(S_{0,1}^4, n)$-encoder whose codewords are at Hamming distance at least 2 from each other.

6. Generalize 3 for any constrained system S over $\{0,1\}$ with finite memory $\mathcal{M}$.