Syndrome decoding: example

An $(8,4)$ binary linear block code $\mathcal{C}$ is defined by systematic matrices:

$$H = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{array}\right] \Longrightarrow G = \left[\begin{array}{cccc|cccc} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Consider two possible messages:

$$\mathbf{m}_1 = [0\,1\,1\,0] \qquad \mathbf{m}_2 = [1\,0\,1\,1]$$
$$\mathbf{c}_1 = [0\,1\,1\,0\,0\,1\,1\,0] \qquad \mathbf{c}_2 = [0\,1\,0\,0\,1\,0\,1\,1]$$

Suppose error pattern $\mathbf{e} = [0\,0\,0\,0\,0\,1\,0\,0]$ is added to both codewords.

$$\mathbf{r}_1 = [0\,1\,1\,0\,0\,0\,1\,0] \qquad \mathbf{r}_2 = [0\,1\,0\,0\,1\,1\,1\,1]$$
$$\mathbf{s}_1 = [1\,0\,1\,1] \qquad \mathbf{s}_2 = [1\,0\,1\,1]$$

Both syndromes equal column 6 of H, so decoder corrects bit 6.

$\mathcal{C}$ is an expanded Hamming code with weight enumerator $A(x) = 1 + 14x^4 + x^8$.

Standard array

Syndrome table decoding can also be described using the *standard array*.

The *standard array* of a group code $\mathcal{C}$ is the coset decomposition of F^n with respect to the subgroup $\mathcal{C}$.

0	c_2	c_3	$\cdots$	c_M
e_2	$c_2 + e_2$	$c_3 + e_2$	$\cdots$	$c_M + e_2$
e_3	$c_2 + e_3$	$c_3 + e_3$	$\cdots$	$c_M + e_3$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
e_N	$c_2 + e_N$	$c_3 + e_N$	$\cdots$	$c_M + e_N$

- The first row is the code $\mathcal{C}$, with the zero vector in the first column.
- Every other row is a coset.
- The n-tuple in the first column of a row is called the *coset leader*.

  We usually choose the coset leader to be the most plausible error pattern, e.g., the error pattern of smallest weight.

Standard array: example

The systematic generator and parity-check matrices for a $(6, 3)$ LBC are

$$G = \left[\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \implies H = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right]$$

The standard array has 6 coset leaders of weight 1 and one of weight 2.

000000	001110	010101	011011	100011	101101	110110	111000
000001	001111	010100	011010	100010	101100	110111	111001
000010	001100	010111	011001	100001	101111	110100	111010
000100	001010	010001	011111	100111	101001	110010	111100
001000	000110	011101	010011	101011	100101	111110	110000
010000	011110	000101	001011	110011	111101	100110	101000
100000	101110	110101	111011	000011	001101	010110	011000
001001	000111	011100	010010	101010	100100	111111	110001

See http://www.stanford.edu/class/ee387/src/stdarray.pl for the short Perl script that generates the above standard array. This code is a *shortened* Hamming code.

Standard array: decoding

An (n, k) LBC over $\mathrm{GF}(Q)$ has $M = Q^k$ codewords.

Every n-tuple appears exactly once in the standard array. Therefore the number of rows N satisfies

$$MN = Q^n \implies N = Q^{n-k}.$$

All vectors in a row of the standard array have the same syndrome.

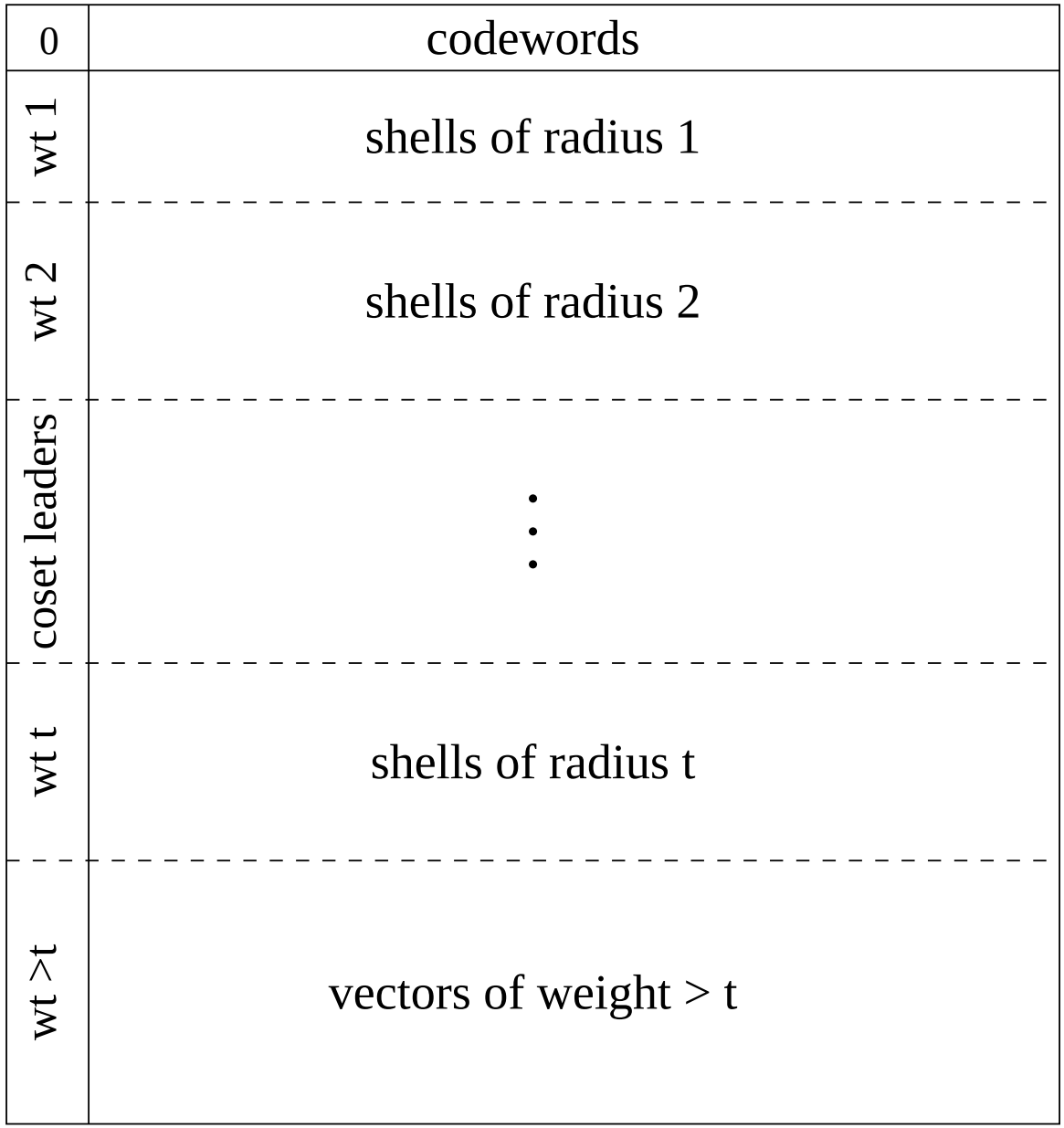
Thus there is a one-to-one correspondence between the rows of the standard array and the Q^{n-k} syndrome values.

Decoding using the standard array is simple: decode senseword $\mathbf{r}$ to the codeword at the top of the column that contains $\mathbf{r}$.

The decoder subtracts the coset leader from the received vector to obtain the estimated codeword.

The *decoding region* for a codeword is the column headed by that codeword.

Standard array and decoding regions



Bounds on minimum distance

The minimum distance of a block code is a *conservative* measure of the quality of an error control code.

- A large minimum distance guarantees reliability against random errors.
- However, a code with small minimum distance *may* be reliable, if the probability of sending codewords with nearby codewords is small.

We use minimum distance as the measure of a code's reliability because:

- A single number is easier to understand than a weight/distance distribution.
- The guaranteed error detection and correction ability are
  - detection: $e = d^* - 1$
  - correction: $t = \lfloor \frac{1}{2}(d^* - 1) \rfloor$
- Algebraic codes covered in the course are limited by minimum distance. Their decoders cannot correct more than t errors even if there is only one closest codeword.

Hamming (sphere-packing) bound

The Hamming bound for a (n, k) block code over Q-ary channel alphabet:

- A code corrects t errors iff spheres of radius t around codewords do not overlap. Therefore

$$Q^k = \text{number of codewords} \leq \frac{\text{volume of space}}{\text{volume of sphere of radius } t} = \frac{Q^n}{V(Q, n, t)},$$

  where $V(Q, n, t)$ is the "volume" (number of elements) of a sphere of radius t in Hamming space of n-tuples over a channel alphabet with Q symbols:

$$V(Q, n, t) = 1 + \binom{n}{1}(Q-1) + \binom{n}{2}(Q-1)^2 + \cdots + \binom{n}{t}(Q-1)^t$$

- Rearranging the inequality gives a lower bound on $n - k$ and thus an upper bound on rate R:

$$Q^{n-k} \geq V(Q, n, t) \implies n - k \geq \log_Q V(Q, n, t)$$

$$R \leq 1 - \frac{1}{n} \log_Q \left(1 + \binom{n}{1}(Q-1) + \binom{n}{2}(Q-1)^2 + \cdots + \binom{n}{t}(Q-1)^t \right)$$

Hamming bound: example

A wireless data packet contains 192 audio samples, 16 bits for each of two channels. The number of information bits is $192 \cdot 2 \cdot 16 = 6144$.

The communications link is a binary symmetric channel with raw error rate 10^{-3}. How many check bits are needed for reliable communication?

t	$n-k$	n	Rate	$\Pr\{> t$ errors$\}$
10	105	6249	0.983	5.4×10^{-02}
12	123	6267	0.980	1.2×10^{-02}
14	141	6285	0.978	2.2×10^{-03}
16	158	6302	0.975	3.0×10^{-04}
18	175	6319	0.972	3.5×10^{-05}
20	192	6336	0.970	3.3×10^{-06}
24	225	6369	0.965	1.8×10^{-08}
28	257	6401	0.960	5.5×10^{-11}
32	288	6432	0.955	1.0×10^{-13}

The Hamming bound shows that more than 4% redundancy is needed to achieve a reasonable bit error rate.

Other bounds on minimum distance

- Plotkin upper bound for binary linear block codes (homework exercise):

$$d^* \leq \frac{n \cdot 2^{k-1}}{2^k - 1} \implies \delta = \frac{d^*}{n} \leq \tfrac{1}{2} \text{ for large } k.$$

  $\delta = d^*/n$ is *normalized* minimum distance.

- McEliece-Rodemich-Rumsey-Welch (MRRW) upper bound.

$$R \leq H\left(\tfrac{1}{2} - \sqrt{\delta(1-\delta)}\right).$$

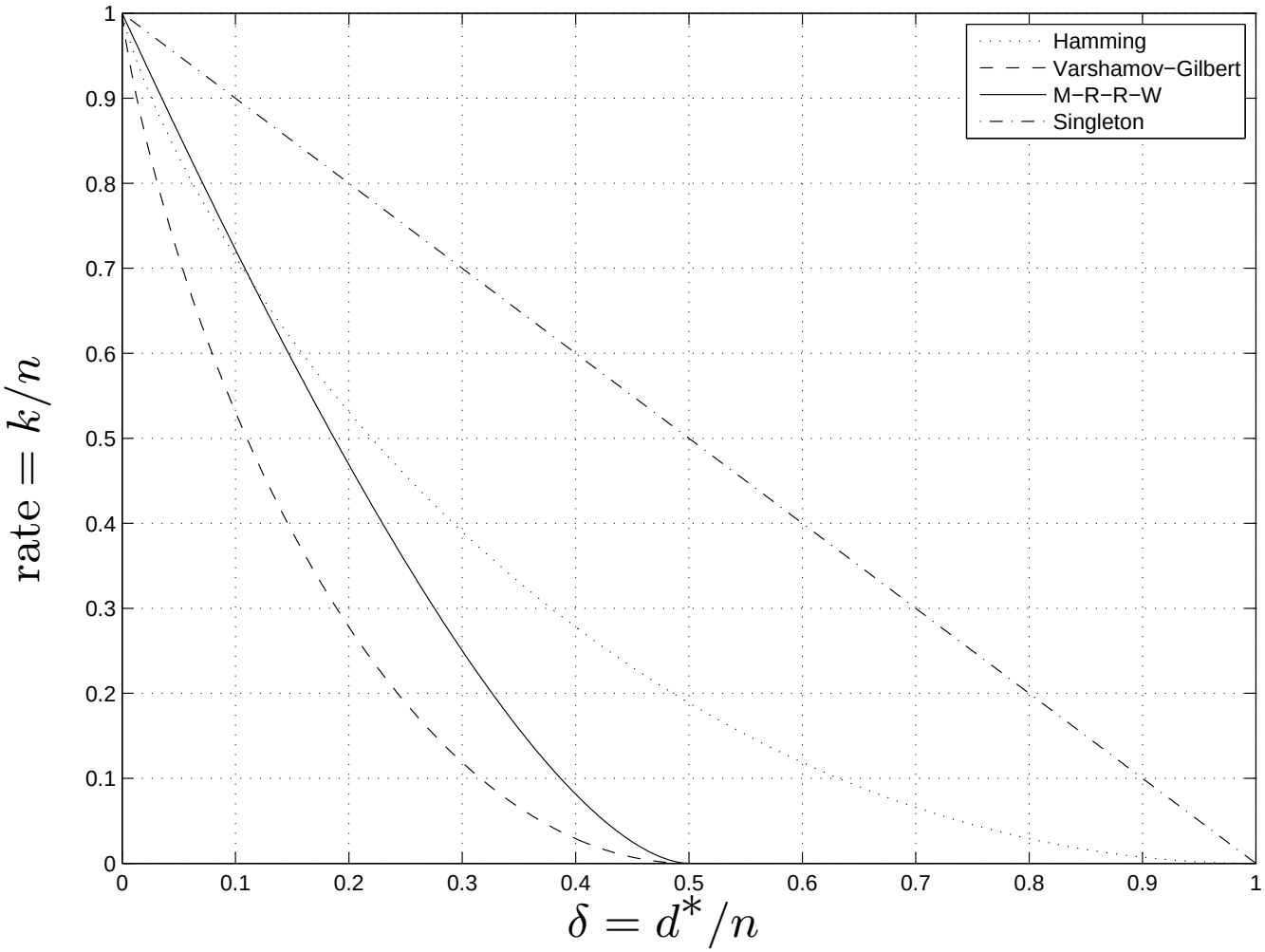
  H is binary entropy function. MRRW bound is better than Hamming or Plotkin for some ranges of δ.

- Varshamov-Gilbert *lower* bound for binary block codes. If $d^* < n/2$ there then *exists* a code with minimum distance d^* and rate R satisfying

$$R \geq 1 - \log_2 \left(\sum_{i=0}^{d-1} \binom{n}{i} \right) \approx 1 - H(d^*/n) = 1 - H(\delta).$$

  For comparison, the Hamming bound is $R \leq 1 - H(\delta/2)$.

Plots of rate vs. normalized minimum distance



The MRRW bound is stronger than Hamming bound except for high rates.

The Hamming bound is fairly tight for high rates. E.g., to correct 10 errors in 1000 bits, Hamming bound requires 78 check bits, but there exists a BCH code with 100 check bits.

Perfect codes

Definition: A block code is called *perfect* if every senseword is within distance t of exactly one codeword.

Other definitions of perfect codes:

- decoding spheres pack perfectly
- have complete bounded-distance decoders
- satisfy the Hamming bound with equality

There are only finitely many classes of perfect codes:

- Codes with no redundancy ($k = n$)
- Repetition codes with odd blocklength: $n = 2m + 1,\ k = 2m,\ t = m$
- Binary Hamming codes: $n = 2^m - 1,\ n - k = m$
- Nonbinary Hamming codes: $n = (q^m - 1)/(q - 1),\ n - k = m,\ q > 2$
- Binary Golay code: $q = 2,\ n = 23,\ k = 12,\ t = 3$
- Ternary Golay code: $q = 3,\ n = 11,\ k = 6,\ t = 2$

Golay discovered both perfect Golay codes in 1949 — a very good year for Golay.

Quasi-perfect codes

Definition: A code is *quasi-perfect* if every n-tuple

- is within distance t of *at most* one codeword, and
- is within distance $t+1$ of *at least* one codeword.

In other words, a code is quasi-perfect if

- spheres of radius t surrounding codewords do not overlap, while
- spheres of radius $t+1$ cover the space of n-tuples.

Examples of quasi-perfect codes:

- Repetition codes with even blocklength
- Expanded Hamming and Golay codes with overall parity-check bit

Exercise: Show that expurgated Hamming codes (obtained by adding an overall parity-check equation) are *not* quasi-perfect.