Códigos Corretores de Erros

E J Nascimento

Universidade Federal do Vale do São Francisco

www.univasf.edu.br/~edmar.nascimento

August 7, 2019

Roteiro

Introdução

- Para proteger a informação dos erros causados pelo canal de comunicação, é necessário utilizar códigos corretores de erros (CCE)
- Os CCE inserem redundância na mensagem transmitida
- Os CCE podem ser de dois tipos: códigos de bloco e códigos convolucionais
  - Códigos de bloco são sem memória e transformam k dígitos em uma palavra código de n dígitos ($n > k$)
  - Códigos convolucionais possuem memória, sendo os n dígitos de saída dependentes da entrada k e de uma memória (dados anteriores)

Introdução

- O teorema da codificação de canais ruidosos de Shannon estabelece para um canal com capacidade C, existem códigos com taxa $R < C$ tal que a probabilidade de erro (após a decodificação de máxima verossimilhança)

$$P_e \leq 2^{-nE_b(R)}$$

- Em que $E_b(R)$ representa a energia de bit em função da taxa R
- Assim, a probabilidade de erro pode se tornar arbitrariamente pequena mantendo-se a taxa R constante e fazendo-se o comprimento da palavra código n aumentar
- Do ponto de vista prático, a decodificação de códigos de comprimento elevado é mais complexa
- Códigos Turbo e LDPC são os que alcançam os melhores resultados atualmente

Introdução

- Um código (n, k) é aquele em que k dígitos de informação são transmitidos por uma palavra código de n dígitos
- Os dígitos não são necessariamente binários
- A taxa do código (n, k) é definida como $R = k/n$
- Tanto os dígitos de informação, quanto as palavras códigos podem ser organizadas em vetores
  - $\mathbf{d}$ é um vetor de dimensão k
  - $\mathbf{c}$ é um vetor de dimensão n
- A distância de Hamming entre dois vetores $\mathbf{c}$ e $\mathbf{c}'$ corresponde ao número de elementos em que eles diferem
- Uma esfera de Hamming de raio t engloba todos os vetores $\mathbf{c}'$ que estão a uma distância de Hamming t da palavra código $\mathbf{c}$

Introdução

- Um CCE pode corrigir até t erros se a distância de Hamming mínima entre as palavras código é dada por

$$d_{min} = 2t + 1$$

- Essa condição é necessária para que as esferas de Hamming não se interceptem
- Através de técnicas de combinatória é possível encontrar o número de esferas de Hamming para um código (n, k)
- Se $m = n - k$ (número de dígitos de redundância), então o limitante de Hamming é dado por

$$2^m \geq \sum_{j=0}^{t} \binom{n}{j}$$

Introdução

- O limitante de Hamming é uma condição necessária (mas não suficiente) para que um código (n, k) possa corrigir t erros
- Se a igualdade é verificada, o código é dito ser perfeito
- Para $t = 1$ ($d_{min} = 3$), códigos perfeitos são chamados de códigos de Hamming
- Nesse caso, tem-se que

$$\begin{aligned} 2^m &= \sum_{j=0}^{1} \binom{n}{j} = 1 + n \\ n &= 2^m - 1 \end{aligned}$$

- Códigos de Hamming com $m \geq 3$ podem ser escritos como $(n, k, d) = (2^m - 1, 2^m - 1 - m, m)$

Introdução

- O código de Hamming $(7, 4, 3)$ permite corrigir um erro e possui taxa $4/7 = 0,57$
- Se o objetivo for apenas detectar o erro e não corrigir, os requisitos de distância mínima são menores
- Um código permite a detecção de t erros se a sua distância mínima é

$$d_{min} = t + 1$$

- O estudo dos CCE ficará limitado aqui aos códigos binários
  - Nos demais casos é necessário um formalismo com corpos finitos (Galois)
- No caso binário, as operações serão realizadas módulo-2 (adição como uma operação XOR)

Códigos de Bloco

- As palavras de dados e palavras código são representadas respectivamente pelos vetores

$$\begin{aligned} \mathbf{d} &= (d_1, d_2, \cdots, d_k) \\ \mathbf{c} &= (c_1, c_2, \cdots, c_n) \end{aligned}$$

- A codificação de um código de bloco linear é realizada pela operação $\mathbf{c} = \mathbf{d}.\mathbf{G}$, em que $\mathbf{G}$ é a matriz geradora do código com dimensão $k \times n$
- Um código é sistemático quando $c_1 = d_1, \cdots, c_k = d_k$ e os demais termos $c_{k+1}, \cdots, c_n$ são funções de $\mathbf{d}$
- Para um código sistemático, os $m = n - k$ elementos $c_{k+1}, \cdots, c_n$ são chamados de dígitos de paridade

Códigos de Bloco

- A matriz geradora de um código sistemático é escrita como

$$\mathbf{G} = [\mathbf{I}_{k\times k}\ \mathbf{P}_{k\times m}]$$

- Para esse tipo de código, as palavras código são dadas por

$$\begin{aligned}\mathbf{c} &= \mathbf{dG} = \mathbf{d}[\mathbf{I}_k\ \mathbf{P}] \\ &= [\mathbf{d}\ \mathbf{dP}] = [\mathbf{d}\ \mathbf{c}_P]\end{aligned}$$

- Em que $\mathbf{c}_P = \mathbf{dP}$ são conhecidos como dígitos de paridade
- O peso de uma palavra código corresponde ao número de uns na palavra código
- A distância de Hamming e o peso estão relacionados por

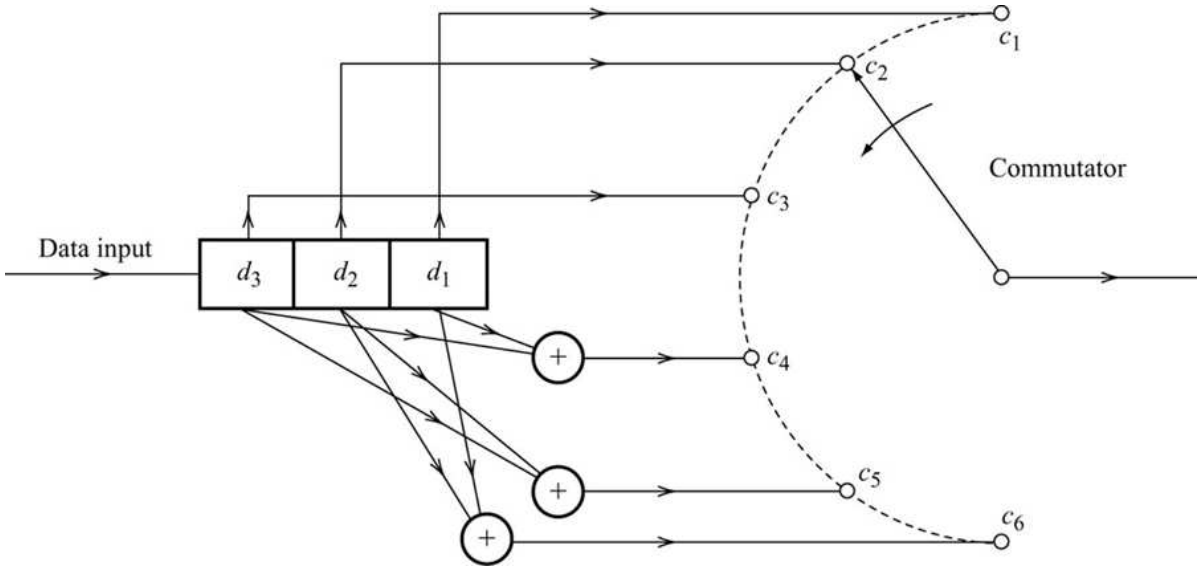
$$d(\mathbf{c}_a, \mathbf{c}_b) = \text{peso}(\mathbf{c}_a \oplus \mathbf{c}_b)$$

Códigos de Bloco

- Exemplo: Escrever as palavras código do código $(6, 3)$ com matriz geradora dada por

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

- Verifique a distância mínima do código e a sua capacidade de correção



Propriedades dos Códigos Lineares

- Um código de bloco é linear se para duas palavras código quaisquer $\mathbf{c}_a$ e $\mathbf{c}_b$, a palavra código $\mathbf{c}_a \oplus \mathbf{c}_b$ pertence ao código
- A palavra código $\mathbf{00}\cdots\mathbf{00}$ pertence a todo código linear
- A distância mínima é igual ao peso mínimo
- A decodificação de um código linear pode ser feita através do cálculo da síndrome do erro
- Para isso, verifica-se que

$$\mathbf{d}.\mathbf{P} \oplus \mathbf{c}_p = \underbrace{[\mathbf{d} \quad \mathbf{c}_p]}_{\mathbf{c}} \begin{bmatrix} \mathbf{P} \\ \mathbf{I}_m \end{bmatrix} = 0$$

$$\mathbf{c}.\mathbf{H}^T = 0, \quad \mathbf{H}^T = \begin{bmatrix} \mathbf{P} \\ \mathbf{I}_m \end{bmatrix}, \text{ ou } \mathbf{H} = [\mathbf{P}^T \quad \mathbf{I}_m]$$

Propriedades dos Códigos Lineares

- A matriz **H** é chamada de matriz de (checagem) paridade
- Se uma palavra **c** pertence ao código, então a relação descrita anteriormente é verificada
- Se devido ao ruído, a palavra recebida for dada por $\mathbf{r} = \mathbf{c} \oplus \mathbf{e}$, então

$$\mathbf{s} = \mathbf{r}.\mathbf{H}^T = \mathbf{e}.\mathbf{H}^T$$

- **s** é chamada de síndrome e caracteriza o erro **e**
- A correspondência do erro e a síndrome não é um-a-um (2^n padrões de erro para 2^k síndromes)
- Na decodificação, o padrão de erro escolhido é o de menor peso (mais provável)

$$\mathbf{c} = \mathbf{r} \oplus \mathbf{e}_{min}$$

Decodificação de Códigos Lineares

- Exemplo: Determinar para o código $(6, 3)$ com matriz geradora dada por

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

- A sua matriz de paridade $\mathbf{H}^T$, a sua tabela de síndromes e a palavra código correspondente ao vetor recebido $\mathbf{r} = (100011)$

Códigos de Hamming

- A construção de bons códigos não segue uma receita única
- Existem diversas construções que exploram propriedades das matrizes geradoras e de paridade
- O código de Hamming $(7, 4)$ é construído da seguinte forma
  - Cada padrão de erro de um bit corresponde a uma síndrome distinta
  - Cada uma das sete linhas da matriz de paridade deve ser distinta das demais
  - Assim, $\mathbf{s} = \mathbf{e}\mathbf{H}^T$ corresponde a uma linha da matriz $\mathbf{H}^T$ para erros de peso 1
  - O ordenamento das linhas de $\mathbf{H}^T$ pode variar, mas a forma sistemática é preferível

Códigos de Hamming

- Uma possível construção para $\mathbf{H}^T$ e $\mathbf{G}$ correspondente é dada por

$$\mathbf{H}^T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P} \\ \mathbf{I}_m \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_k & \mathbf{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Códigos Cíclicos

- Códigos cíclicos são uma subclasse de códigos lineares em que os procedimentos de codificação e cálculo de síndrome podem ser implementados facilmente com o uso de registradores de deslocamento
- Em um código cíclico, o deslocamento lateral dos elementos de uma palavra código resulta em outra palavra código

$$\begin{aligned} \mathbf{c} &= (c_1, c_2, \cdots, c_{n-1}, c_n) \\ \mathbf{c}' &= (c_2, c_3, \cdots, c_n, c_1) \\ \mathbf{c}^{(i)} &= (c_{i+1}, c_{i+2}, \cdots, c_n, c_1, c_2, \cdots, c_i) \end{aligned}$$

- Palavras código podem ser representadas por polinômios de grau $n-1$ com coefientes binários, ou seja

$$\begin{aligned} c(x) &= c_1 x^{n-1} + c_2 x^{n-2} + \cdots + c_n \\ c^{(i)}(x) &= c_{i+1} x^{n-1} + c_{i+2} x^{n-2} + \cdots + c_n x^i + c_1 x^{i-1} + \cdots + c_i \end{aligned}$$

Códigos Cíclicos

- Em um código cíclico, o resto da divisão de $x^i c(x)$ por $x^n + 1$ é $c^{(i)}(x)$
- Nas operações, a subtração é equivalente à soma (ou exclusivo)
- Considerando um código (n, k) em que as palavras código e de dados podem ser representadas pelos polinômios

$$\begin{aligned} c(x) &= c_1 x^{n-1} + c_2 x^{n-2} + \cdots + c_n \\ d(x) &= d_1 x^{k-1} + d_2 x^{k-2} + \cdots + d_k \end{aligned}$$

- O código pode ser gerado através da operação $c(x) = d(x)g(x)$, em que $g(x)$ é chamado de polinômio gerador do código
- Se $g(x)$ é um polinômio de grau $n - k$ e um fator de $x^n + 1$, então $g(x)$ é um polinômio que gera um código linear cíclico (n, k)

Códigos Cíclicos

- Exemplo: Obtenha um polinômio gerador para um código cíclico $(7, 4)$ e obtenha as palavras código para os vetores de dados (1010), (1111), (0001) e (1000)
- Solução: Verificar que $x^7 + 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1)$ e que $x^3 + x^2 + 1$ ou $x^3 + x + 1$ podem ser polinômios geradores
- Este código não está na forma sistemática

Códigos Cíclicos

- Para o código estar na forma sistemática, ele deve ser escrito como

$$\begin{aligned} c(x) &= x^{n-k}d(x) + \rho(x) \\ \rho(x) &= \text{Resto } \frac{x^{n-k}d(x)}{g(x)} \end{aligned}$$

- Exemplo: Para o código cíclico $(7, 4)$ com gerador $x^3 + x^2 + 1$ codifique de modo sistemático o vetor de dados (1010)
- Construa a matriz geradora do código notando que as linhas da matriz representam as palavras código referentes a (1000), (0100), (0010) e (0001) nesta ordem
- Construir o restante do código e verificar as suas propriedades cíclicas

Decodificação de Códigos Cíclicos

- Em um código cíclico, $c(x)$ é um múltiplo de $g(x)$, ou seja, $c(x)$ é divisível por $g(x)$
- Dado um polinômio $r(x)$ referente a palavra recebida $\mathbf{r}$, um erro é detectável se $r(x)$ não é divisível por $g(x)$, assim

$$\begin{aligned} \frac{r(x)}{g(x)} &= m_1(x) + \frac{s(x)}{g(x)} \\ s(x) &= \text{Resto } \frac{r(x)}{g(x)} \end{aligned}$$

- $s(x)$ é um polinômio de síndrome de grau menor ou igual a $n-k-1$
- Se $r(x) = c(x) + e(x)$, então

$$s(x) = \text{Resto } \frac{e(x)}{g(x)}$$

Decodificação de Códigos Cíclicos

- Como nos demais códigos lineares, uma síndrome pode corresponder a vários padrões de erro
- O erro mais provável é o de menor peso
- Em uma tabela de síndromes, os $2^{(n-k)}$ valores de síndrome são associados aos padrões de erro de menor peso
- Exemplo: Para o código cíclico $(7, 4)$ sistemático com gerador $x^3 + x^2 + 1$, construa a tabela de síndromes e determine os vetores de dados para $\mathbf{r} = (1101101)$, $\mathbf{r} = (0101000)$ e $\mathbf{r} = (0001100)$

Decodificação de Códigos Cíclicos

- Os códigos cíclicos são exemplos de códigos lineares cuja implementação em hardware é bastante simples
- Existem construções de códigos cíclicos que possuem propriedades adicionais que podem ser exploradas
- Códigos cíclicos de grande importância são o BCH (Bose-Chaudhuri-Hocquenghen) e Reed-Solomon
  - Usados em CDs, DVDs, TV digital, etc.
  - O projeto desses códigos envolve a determinação do polinômio gerador
  - Algoritmos eficientes permitem localizar o erro (resolver a equação da síndrome)
  - Em códigos não binários ainda é necessário determinar o tipo de erro
- Códigos CRC (Check Redundancy Check) são também exemplos de códigos cíclicos
  - Empregados em vários protocolos de redes de computadores

Efeito da Correção de Erro

- O uso de CCE melhora em geral o desempenho dos sistemas digitais
  - A probabilidade de erro do sistema codificado é menor do que a do sistema não codificado para um dado valor de $E_b/\mathcal{N}$
  - Usar CCE é mais eficiente que simplesmente aumentar a potência de transmissão
- Um código (n, k) que corrige t erros codifica k dígitos de informação em n dígitos
- Para comparar os sistemas não codificado e codificado, considera-se:
  - A potência transmitida é a mesma em ambos os casos
  - Os k dígitos de informação levam o mesmo tempo de transmissão em ambos os casos
- Assim a taxa de bit R_b do sistema não codificado é k/n vezes a do sistema codificado
- A largura de banda do sistema não codificado é k/n vezes a do sistema codificado

Efeito da Correção de Erro

- A energia total para transmitir n dígitos no sistema codificado é a mesma para transmitir k dígitos no sistema não codificado
  - No sistema codificado, E_b deve ser escalado por k/n
- Com essas hipóteses, pode-se mostrar que

$$P_{ec} \simeq \binom{n-1}{t}(P_{bc})^{t+1}, \quad P_{bc} \ll 1$$

- Em que P_{ec} é a probabilidade de erro após a correção dos erros e P_{bc} é a probabilidade de erro da sequência codificada antes da correção
- A expressão de P_{bc} é obtida a partir das fórmulas convencionais de probabilidade de erro substituindo-se E_b por $(k/n)E_b$

Efeito da Correção de Erro

- Tomando como exemplo a modulação BPSK, a probabilidade de erro sem codificação é dada por

$$P_{eu} = Q\left(\sqrt{\frac{2E_b}{\mathcal{N}}}\right)$$

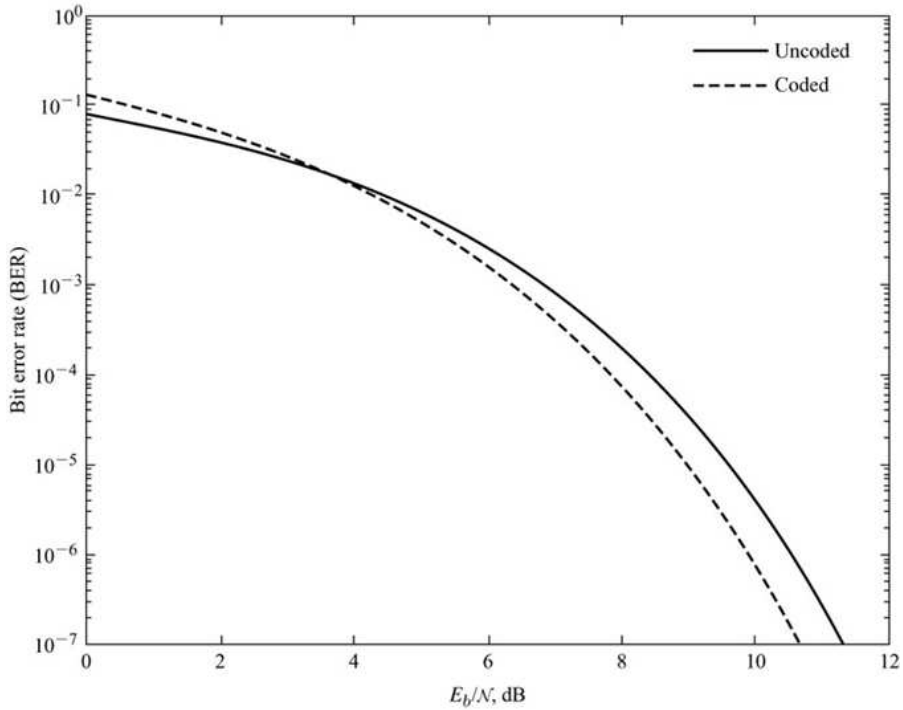
- Com o uso do código, tem-se que

$$P_{bc} = Q\left(\sqrt{\frac{2kE_b}{n\mathcal{N}}}\right)$$

$$P_{ec} = \binom{n-1}{t}\left[Q\left(\sqrt{\frac{2kE_b}{n\mathcal{N}}}\right)\right]^{t+1}$$

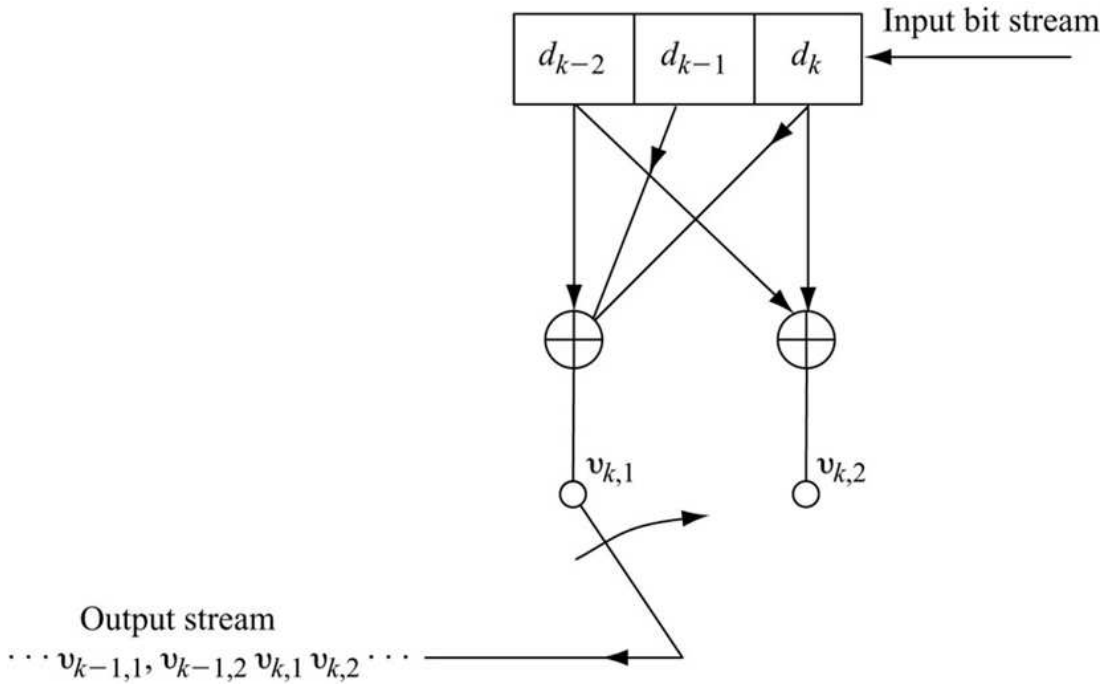
Efeito da Correção de Erro

- Considerando um código $(7, 4)$ com capacidade de correção de um erro



Códigos Convolucionais

- Códigos convolucionais diferem dos códigos de bloco pela presença de elementos de memória
  - Os bits codificados dependem da entrada atual e de entradas anteriores
- Códigos convolucionais costumam trabalhar com valores de k e n baixos, ao contrário dos códigos de bloco

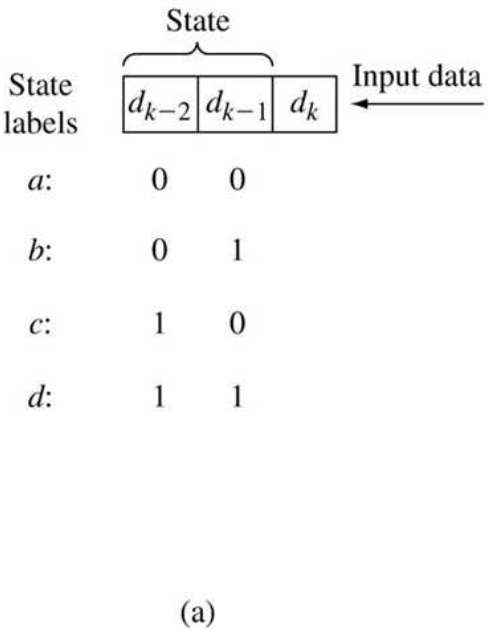


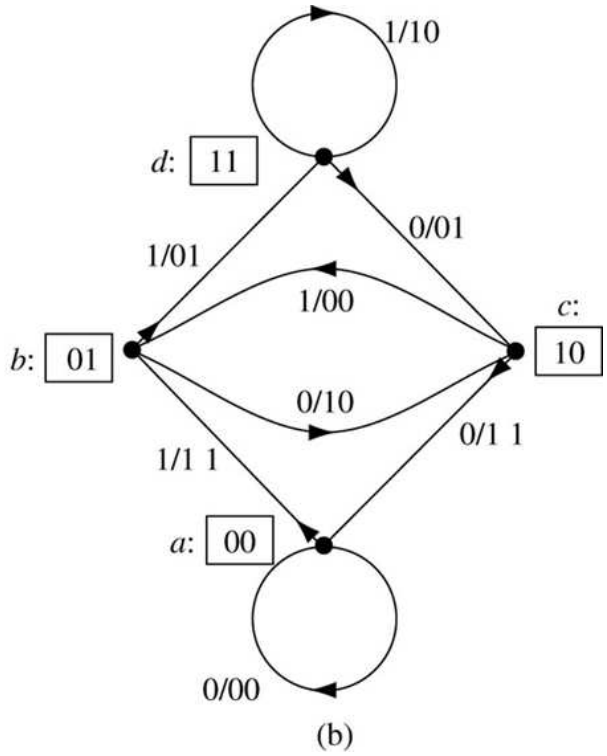
Códigos Convolucionais

- Na figura é mostrado um codificador convolucional com restrição de comprimento $N = 3$ (número de registradores de deslocamento) e $I = 2$ somadores
- A sequência de saída correspondente à entrada 11010 é 11010100101100
- Incialmente, os registradores possuem o valor 0
- Para finalizar, $N - 1$ zeros são inseridos a fim de esvaziar os registradores
- Tem-se $n = (N + k - 1)I$ dígitos na saída para k dígitos na entrada
- Na prática, $k \gg N$, o que implica em uma taxa aproximada de $1/I$

Códigos Convolucionais

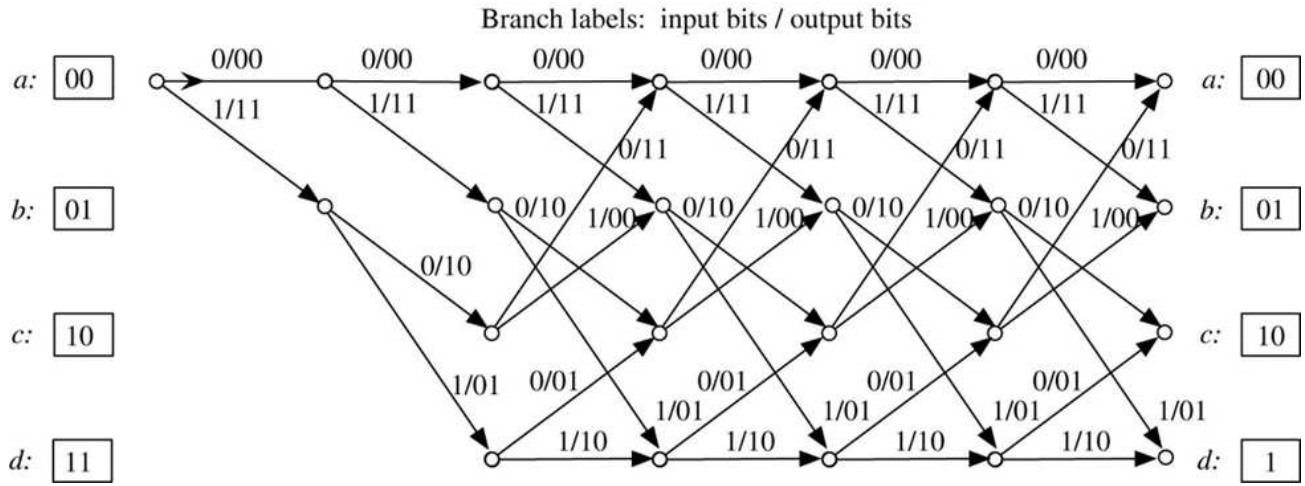
- O comportamento dinâmico de um codificador convolucional é caracterizado por um diagrama de estados ou por uma treliça
  - No codificador do exemplo, o estado é representado pelos valores possíveis assumidos pelos dois registradores da esquerda





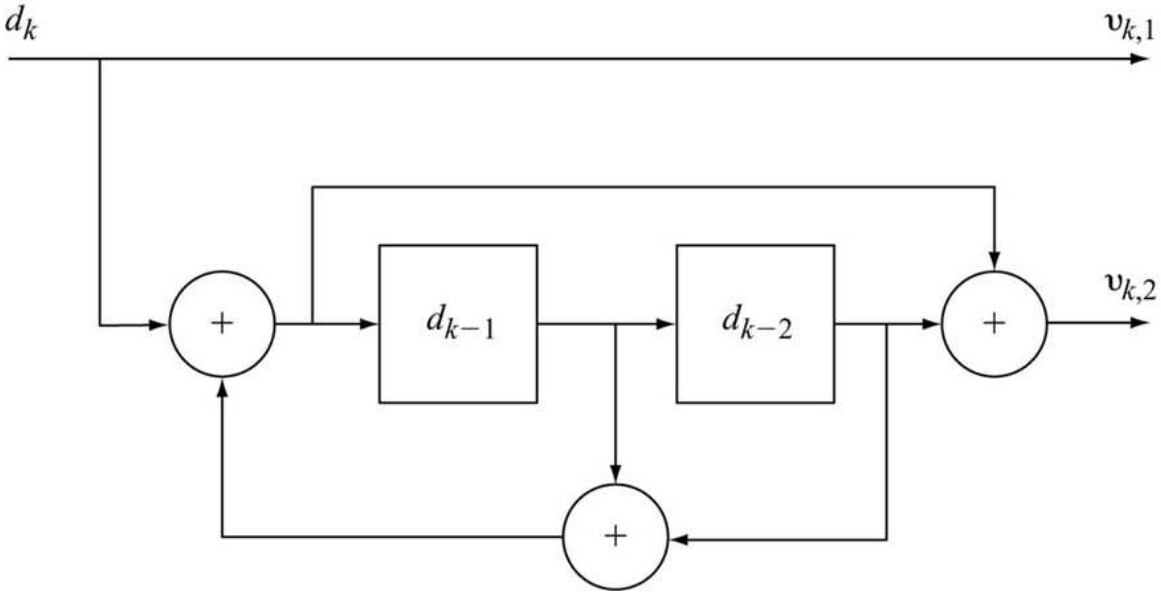
Códigos Convolucionais

- No diagrama de treliças, as transições ao longo do tempo são explicitadas
  - Para obter a sequência codificada, é suficiente seguir um caminho
  - Após um regime estacionário, a estrutura é repetitiva



Códigos Convolucionais

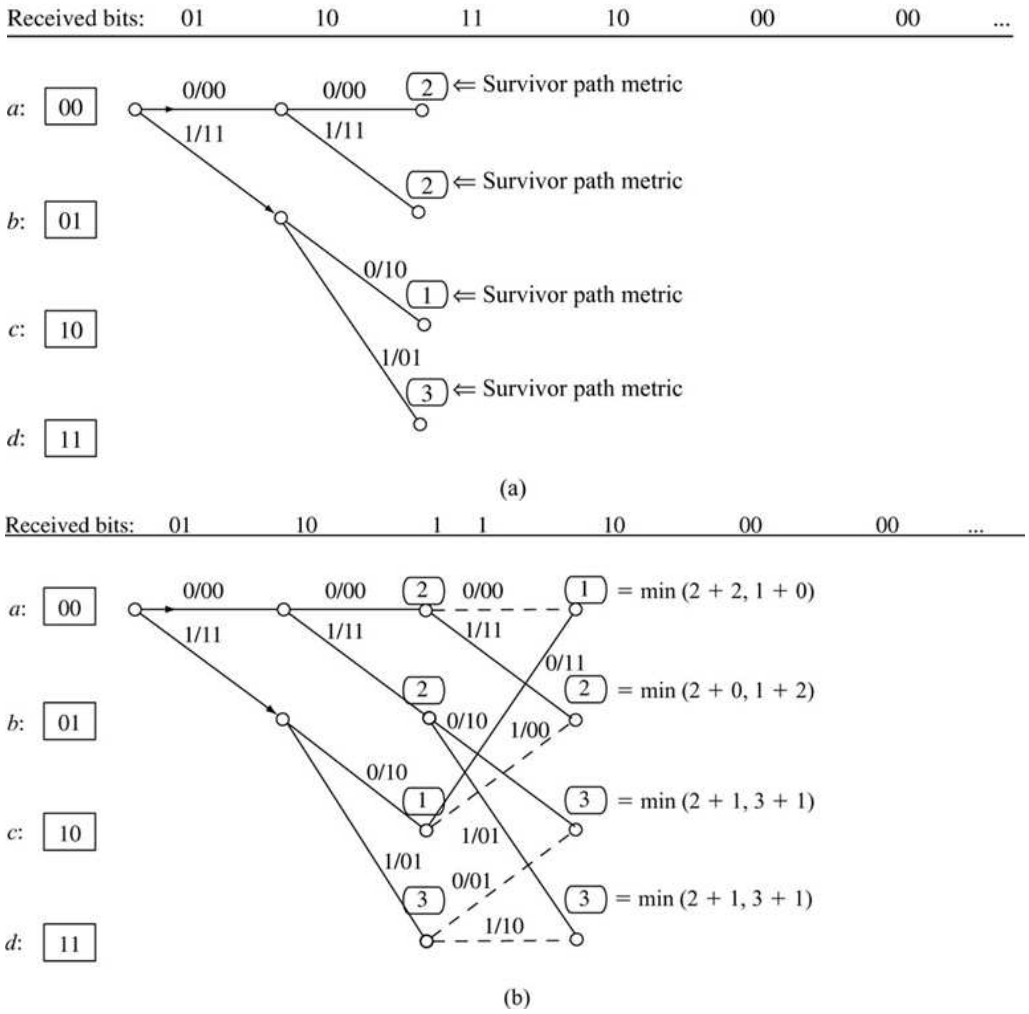
- Códigos convolucionais também podem ser codificados na forma sistemática (dados aparecendo diretamente na saída)



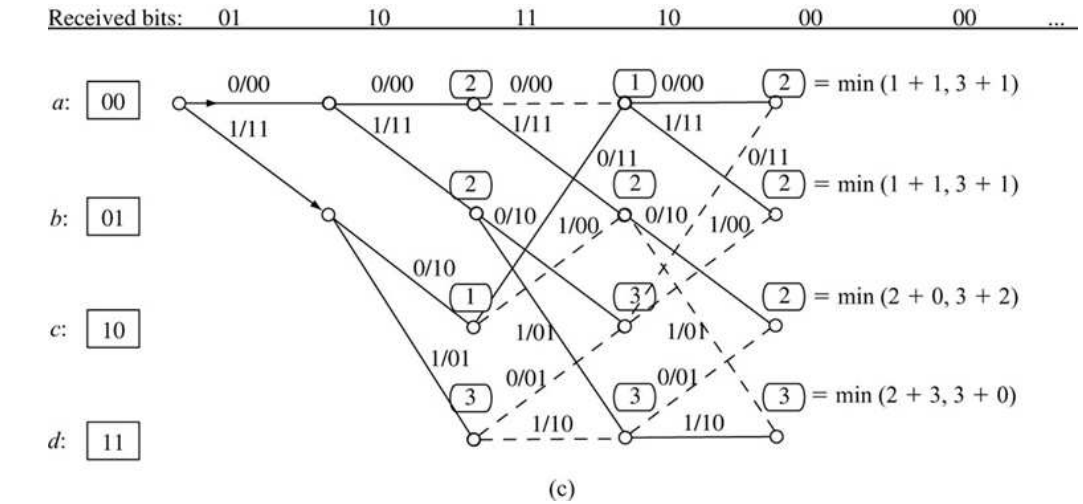
Decodificação de Códigos Convolucionais

- São usados para decodificar códigos convolucionais as seguintes técnicas
  - Algoritmo de Viterbi
  - Decodificação sequencial
  - Decoficação com realimentação (Feedback decoding)
- O algoritmo de Viterbi realiza uma decodificação de máxima verossimilhança, ou seja, ele encontra uma sequência válida (caminho na treliça) mais próxima da sequência recebida
  - Caminho com menor distância de Hamming da palavra recebida
  - Nos pontos de convergência de caminhos, só o caminho de maior peso sobrevive

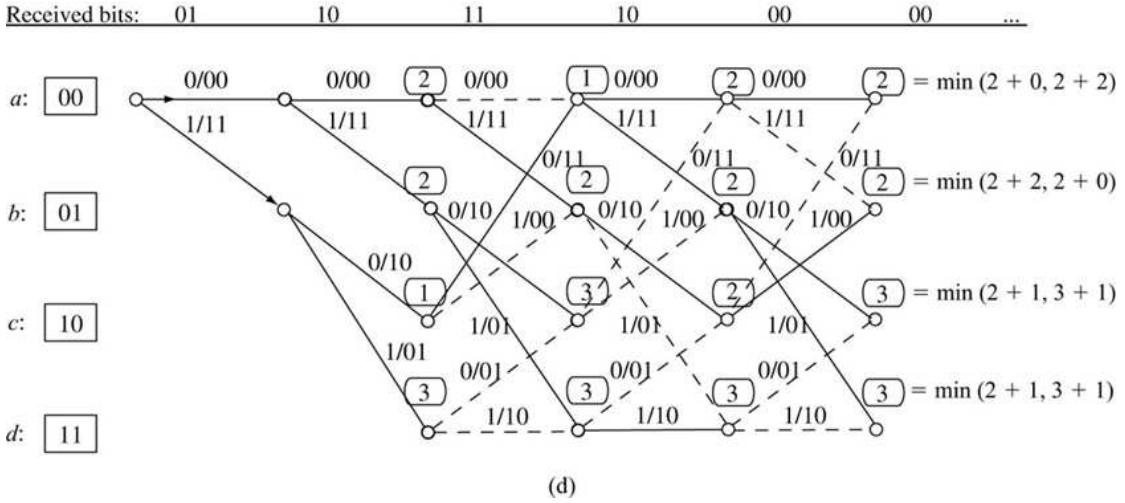
Algoritmo de Viterbi



Algoritmo de Viterbi

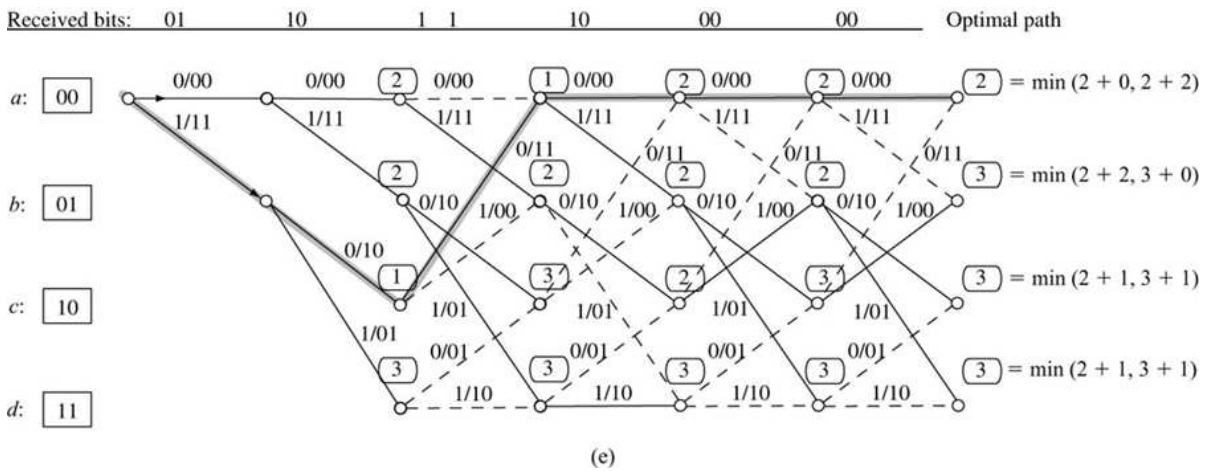


(c)



(d)

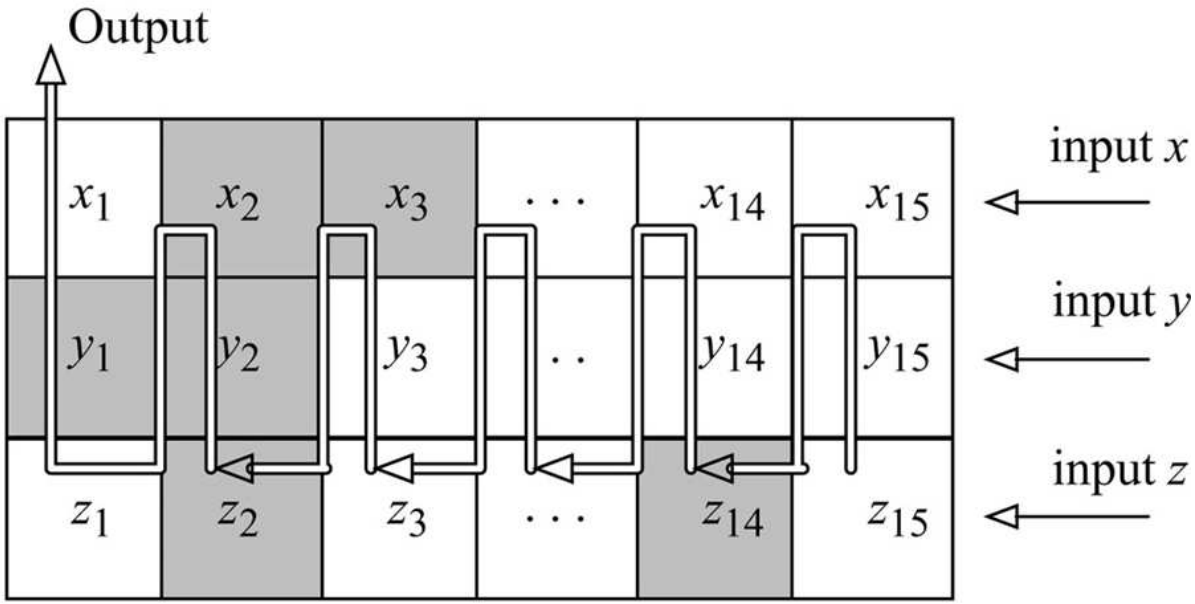
Algoritmo de Viterbi



(e)

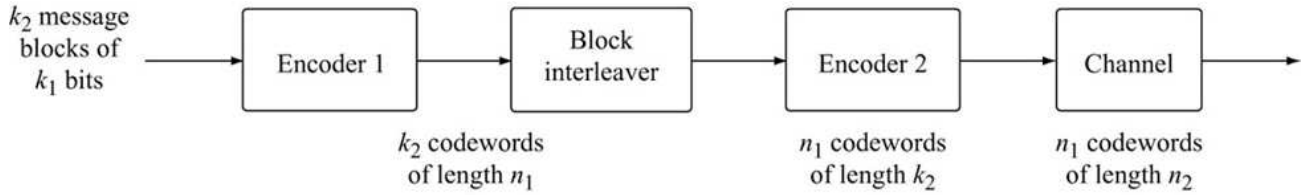
Combinação de Códigos e Entrelaçamento

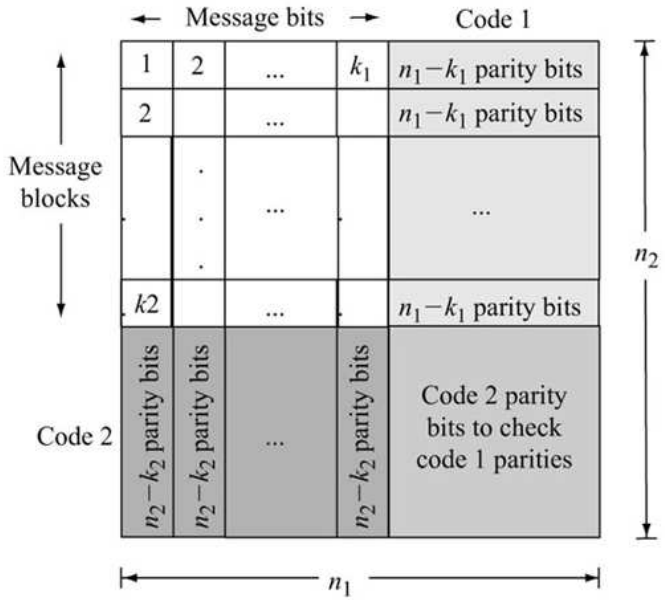
- Para aumentar o poder de correção de códigos, pode-se recorrer a várias técnicas
- Para proteger a informação de erros em rajada (sequênciais) pode-se utilizar o entrelaçamento



Combinação de Códigos e Entrelaçamento

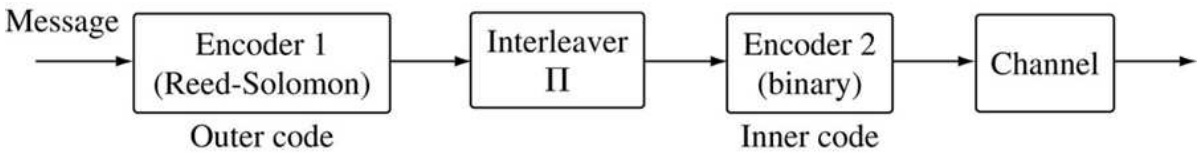
- Códigos produto com entrelaçamento





Combinação de Códigos e Entrelaçamento

- Concatenação de códigos binários e não binários
- Uma das configurações mais usadas usa um código convolucional como código interno binário



Considerações Finais

- A complexidade da utilização de um código não está na codificação e sim na decodificação
- Métodos de decodificação diferentes vão alcançar desempenhos diferentes
- O sucesso de um código depende da existência de um algoritmo de decodificação eficiente
- Atualmente, os códigos TURBO (combinação de códigos convolucionais com entralaçador) e LDPC (código de bloco com matriz de paridade esparça) estão próximos da capacidade de Shannon
- Em sistemas de comunicação, vários fatores são levados em conta ao se escolher os tipos de códigos utilizados