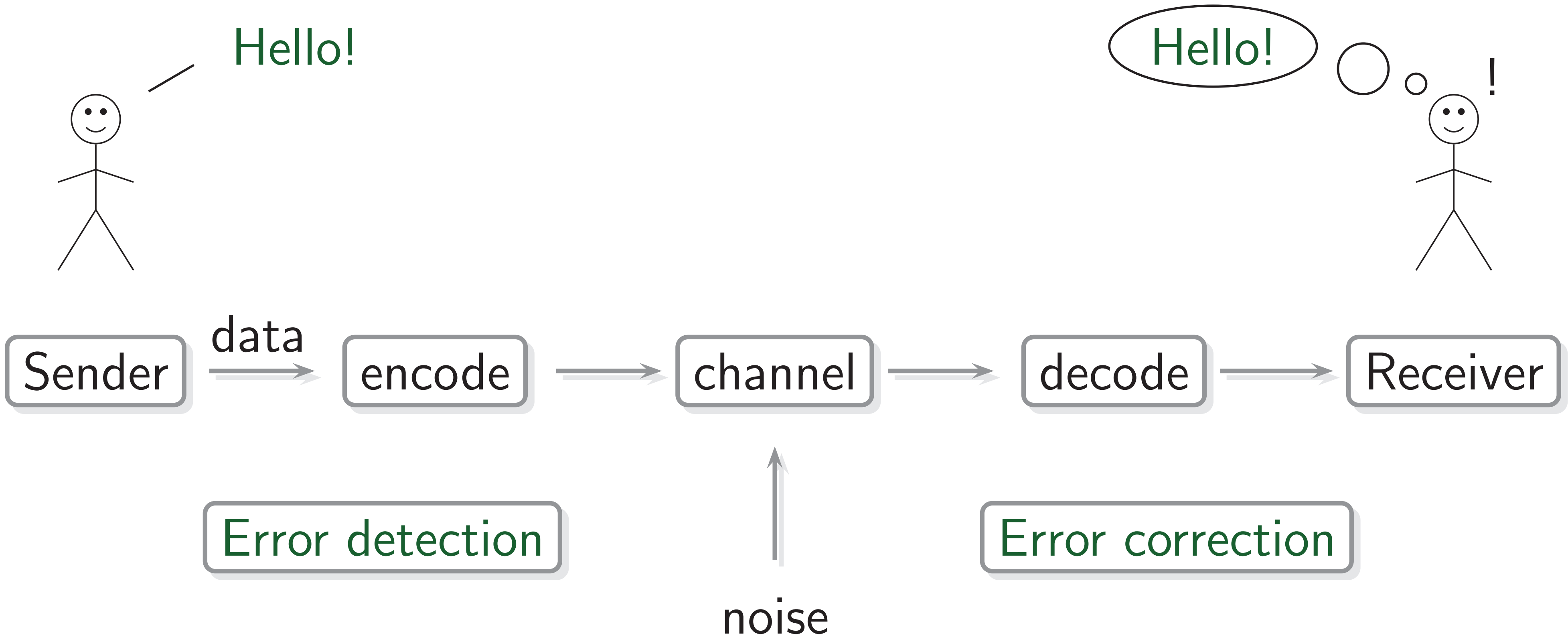
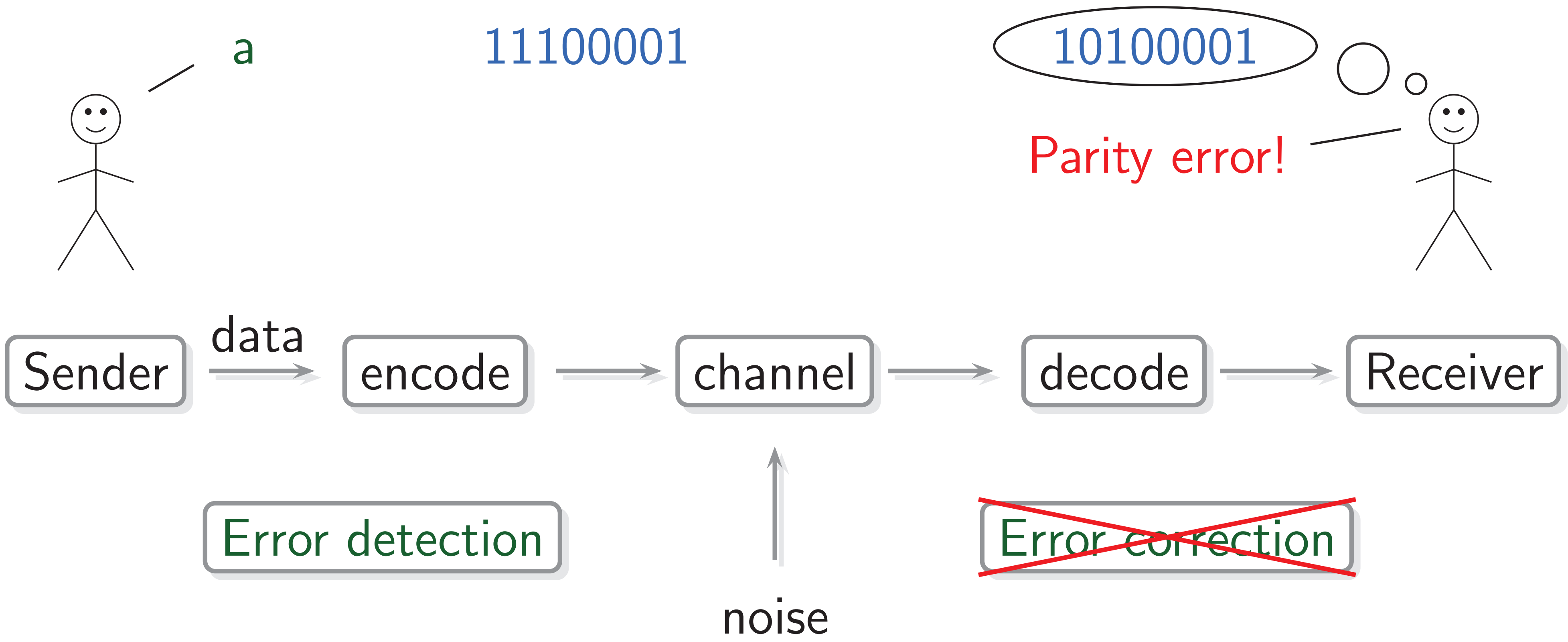
Chapter 2: Error Detection and Correction Codes

Lecture 3



The codeword “Hello!” was corrupted to the (corrupted code)word “ello!”. We write this as

$$\text{Hello!} \rightsquigarrow \text{ello!}$$



Example

Extended ASCII has 8-bit binary codewords with even parity.

Suppose that

$$11100001 \rightsquigarrow 10100001$$

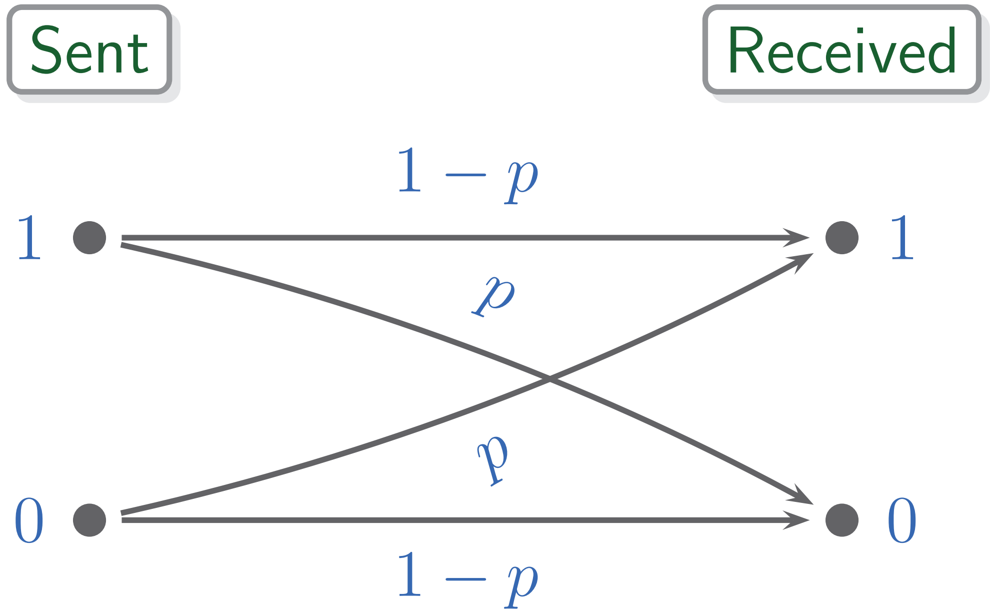
There is odd parity, so an error is detected (but cannot be corrected).

Extended ASCII is a single-error detection code: it cannot detect 2 errors.

It can however detect any odd number of errors.

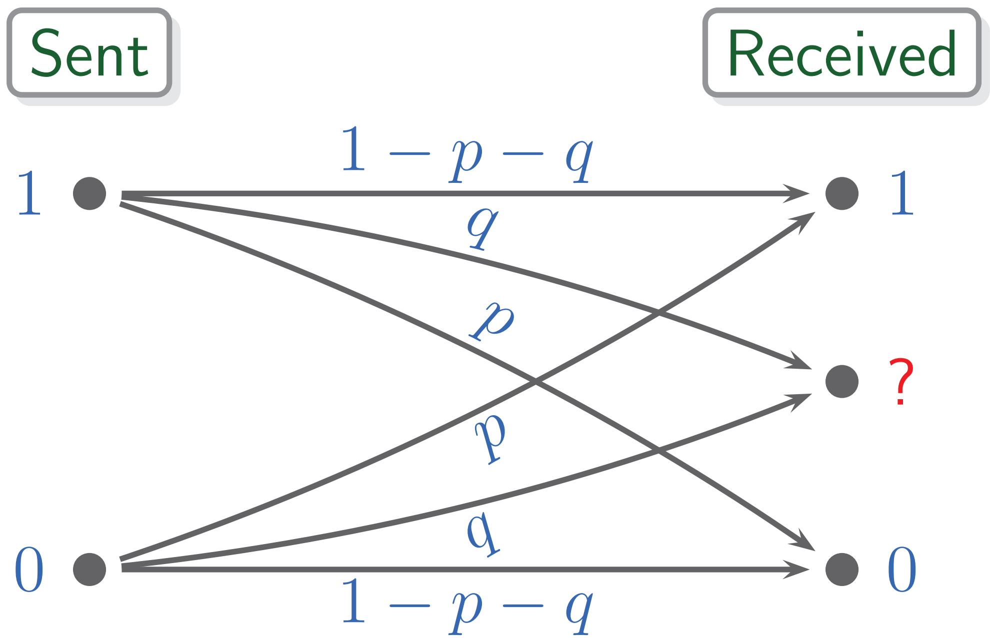
Binary Symmetric Channel (BSC)

- White noise (uniformly random noise) on binary codewords
- Each bit-error has constant probability $p = P(0 \rightsquigarrow 1) = P(1 \rightsquigarrow 0)$.
- The bit-errors are independent of each other.



Binary Symmetric Erase Channel (BSEC)

- White noise (uniformly random noise) on binary codewords
- Bits can be erased by the symbol ?
- $P(0 \rightsquigarrow 1) = P(1 \rightsquigarrow 0) = p$ and $P(0 \rightsquigarrow ?) = P(1 \rightsquigarrow ?) = q$
- The bit-errors are independent of each other and of bit position.



Extended ASCII can correct one erasure, by even parity checking.
For instance, 0101?001 corrects to 01011001 .

n-Bit Even Parity Code

- Codewords are of length n
- Each codeword has even parity (even number of 1s)

Example

The extended ASCII is an 8-bit even parity code.

n-Bit Even Parity Code

- Codewords are of length n
- Each codeword has even parity (even number of 1s)

Consider a binary symmetric channel with bit-error probability p.
If X counts the number of bit-errors, then X has binomial distribution

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for} \quad k = 0, 1, \ldots, n$$

The code detects errors whenever there are an odd number of bit errors.
The probability of undetected errors is therefore

$$P(X=2) + P(X=4) + P(X=6) + \cdots = \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2j} p^{2j} (1-p)^{n-2j}$$

Example

The extended ASCII is an 8-bit even parity code.
The probability of undetected error when

$p = 0.1$ is 0.1534
$p = 0.001$ is 2.8×10^{-5}
$p = 0.00001$ is 2.8×10^{-9}

Theorem

ISBN-10s can detect two types of errors:

- one wrong digit
- two digits swapped

Example

$\mathbf{x} = 1581691750$ is a valid ISBN.
Replacing the digit 9 by 5 gives the number $\mathbf{y} = 1581651750$.
This is not an ISBN:

$$\begin{aligned}\sum_{i=1}^{10} i y_i &= 1\cdot 1 + 2\cdot 5 + 3\cdot 8 + 4\cdot 1 + 5\cdot 6 + 6\cdot 5 + 7\cdot 1 + 8\cdot 7 + 9\cdot 5 + 10\cdot 0 \\ &= 1 + 10 + 24 + 4 + 30 + 30 + 7 + 56 + 45 + 0 \\ &\equiv 1 + 10 + 2 + 4 + 8 + 8 + 7 + 1 + 1 \pmod{11} \\ &\equiv 9 \pmod{11}\end{aligned}$$

Theorem

ISBN-10s can detect two types of errors:

- one wrong digit
- two digits swapped

Example

$\mathbf{x} = 1581691750$ is a valid ISBN.

Now swap the numbers in 69 to give us $\mathbf{z} = 1581961750$.

Let us check whether $\mathbf{z}$ is an ISBN:

$$\begin{aligned}
\sum_{i=1}^{10} i z_i &= 1 \cdot 1 + 2 \cdot 5 + 3 \cdot 8 + 4 \cdot 1 + 5 \cdot 9 + 6 \cdot 6 + 7 \cdot 1 + 8 \cdot 7 + 9 \cdot 5 + 10 \cdot 0 \\
&= 1 + 10 + 24 + 4 + 45 + 36 + 7 + 56 + 45 + 0 \\
&\equiv 1 + 10 + 2 + 4 + 1 + 3 + 7 + 1 + 1 \pmod{11} \\
&\equiv 8 \pmod{11}
\end{aligned}$$

We see that $\mathbf{z}$ is not a valid ISBN.

Theorem

ISBN-10s can detect two types of errors:

- one wrong digit
- two digits swapped

Proof

Suppose that $\mathbf{x} = x_1 x_2 \cdots x_{10}$ is sent with $S(\mathbf{x}) = \sum_{i=1}^{10} i x_i \equiv 0 \pmod{11}$ and that $\mathbf{y} = y_1 y_2 \cdots y_{10}$ is received.

$\textsc{Case}$ 1: $\mathbf{y}$ differs from $\mathbf{x}$ in one digit.
Then $y_k = x_k + m$ for some k and some $m \not\equiv 0 \pmod{11}$.
Then since $k\, y_k = k\, x_k + km$,

$$S(\mathbf{y}) = \sum_{i=1}^{10} i\, y_i = \sum_{i=1}^{10} i\, x_i + km \equiv 0 + km \equiv km \pmod{11}$$

Since $m, k \not\equiv 0 \pmod{11}$ and 11 is prime, $S(\mathbf{y}) \equiv km \not\equiv 0 \pmod{11}$.
The error has therefore been detected.

Theorem

ISBN-10s can detect two types of errors:

- one wrong digit
- two digits swapped

Proof

Suppose that $\mathbf{x} = x_1 x_2 \cdots x_{10}$ is sent with $S(\mathbf{x}) = \sum_{i=1}^{10} i x_i \equiv 0 \pmod{11}$
and that $\mathbf{y} = y_1 y_2 \cdots y_{10}$ is received.

Case 2: $\mathbf{y}$ differs from $\mathbf{x}$ in two swapped digits.
Then $y_k = x_\ell$ and $y_\ell = x_k$ for some k, ℓ.
Then since $k\, y_k = k\, x_\ell = \ell\, x_\ell + (k - \ell)\, x_\ell$
and $\ell\, y_\ell = \ell\, x_k = k\, x_k + (\ell - k)\, x_k$,

$$S(\mathbf{y}) = \sum_{i=1}^{10} i\, y_i = \sum_{i=1}^{10} i\, x_i + (k-\ell)\, x_\ell + (\ell-k)\, x_k \equiv (k-\ell)(x_\ell - x_k) \pmod{11}$$

As before, 11 is prime, so $S(\mathbf{y}) \equiv (k - \ell)(x_\ell - x_k) \not\equiv 0 \pmod{11}$.
The errors have therefore been detected in this case as well. $\square$

Theorem

ISBN-10s can detect two types of errors:

- one wrong digit
- two digits swapped

Many other common codes use check digits, like

- bar codes
- Australian Business Number (ABN)
- Tax File Number (TFN)
- credit card numbers