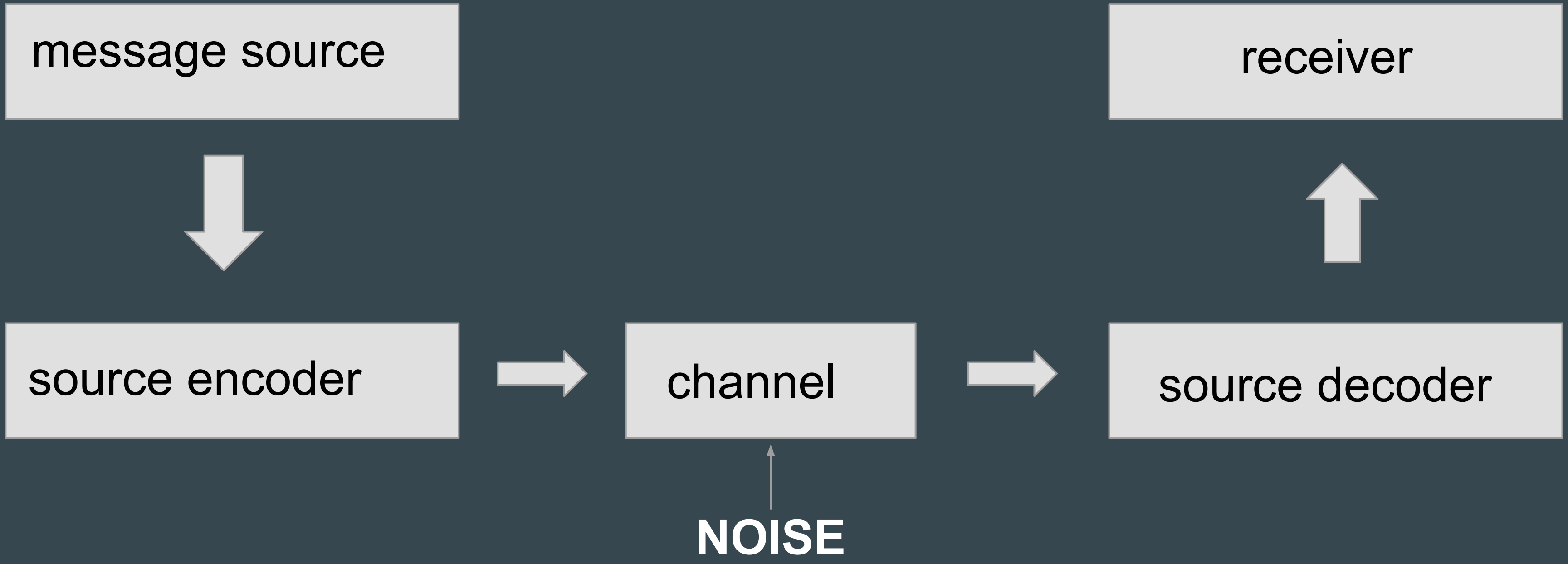
Coding Theory

Kaman Phamdo
Mentor: Sean Ballentine
December 9, 2015

What is a code?

- A code converts information into another representation
- Used for communication through a channel
- How computers communicate
- Encoding
- Decoding

What is a code?

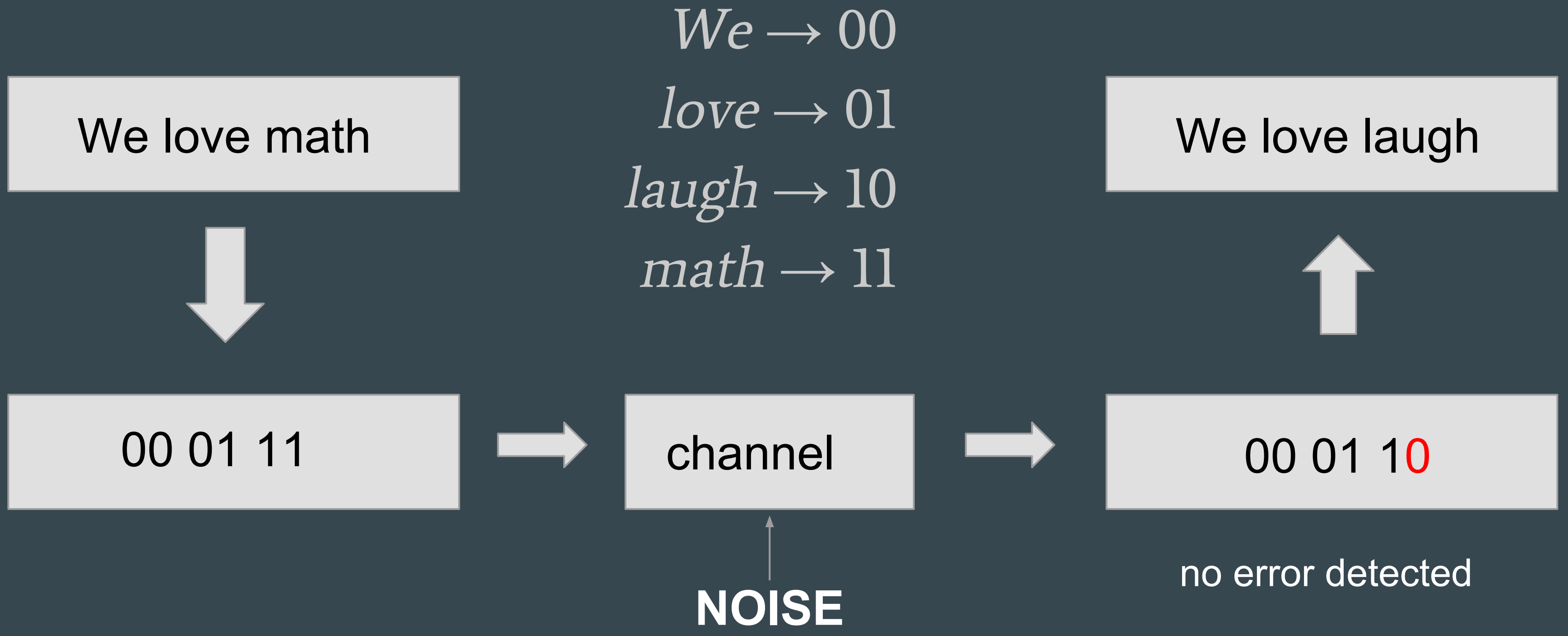


Example 1

$$\begin{aligned} We &\to 00 \\ love &\to 01 \\ laugh &\to 10 \\ math &\to 11 \end{aligned}$$

Suppose we wanted to send the message "*We love math*"...

Example 1



We → 00
love → 01
laugh → 10
math → 11
We love math
00 01 11
channel
NOISE
00 01 10
We love laugh
no error detected

Error-Detecting Codes

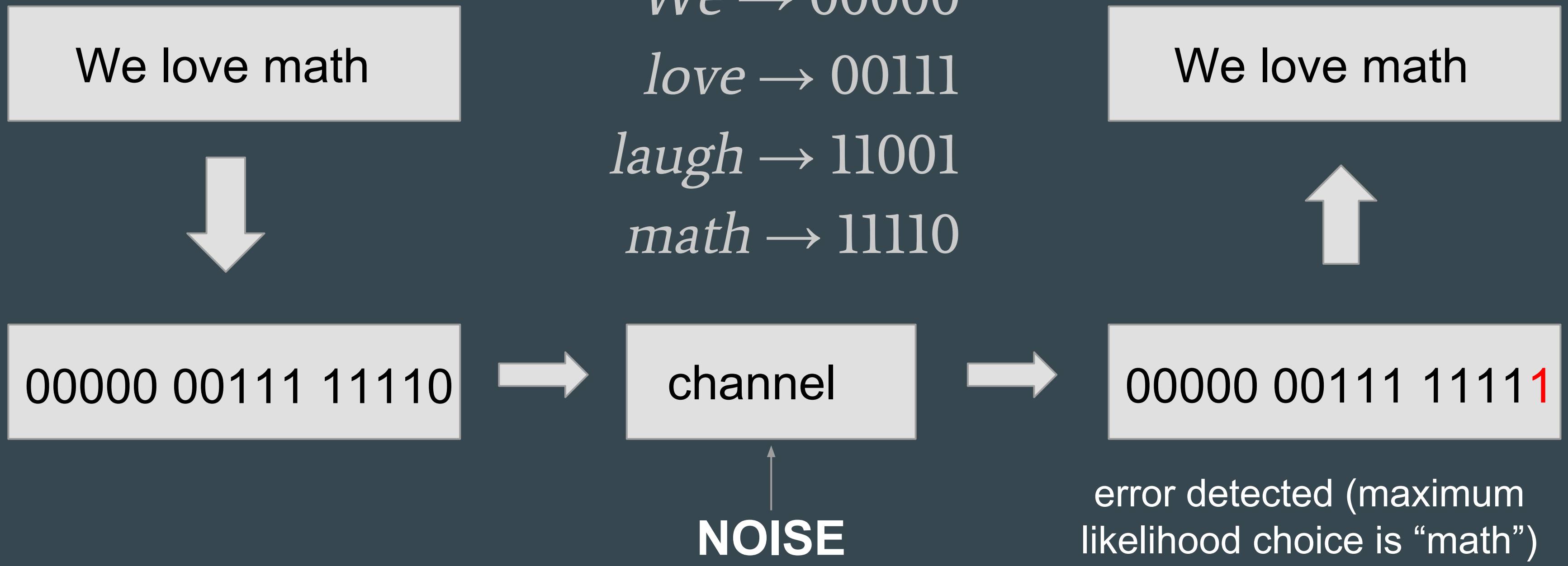
- ISBN (book numbers) - a 10-digit code used to uniquely identify a book
- Last digit is a check digit used for error detection
- Error-detecting but not error-correcting

Error-Correcting Example

$$\begin{aligned} \textit{We} &\rightarrow 00000 \\ \textit{love} &\rightarrow 00111 \\ \textit{laugh} &\rightarrow 11001 \\ \textit{math} &\rightarrow 11110 \end{aligned}$$

Suppose we wanted to send the message "*We love math*" again, but this time using a longer length for code words.

Error-Correcting Example



$$We \longrightarrow 00000$$
$$love \longrightarrow 00111$$
$$laugh \longrightarrow 11001$$
$$math \longrightarrow 11110$$

Error-Correcting Codes

- Need to detect **and** correct errors due to noisy channels
- Can be more expensive and less efficient
- We want good error-correcting capabilities and transmission rates
- Coding theory examines transmission of data across noisy channels and recovery of corrupted messages

Hamming Distance

- Let **x** and **y** be words of length *n* over alphabet *A*. The Hamming distance d(**x**,**y**) is the number of places at which **x** and **y** differ.
- We can define a minimum Hamming distance for a code
- Larger minimum distance = better error-correcting capability

Linear Codes

- A linear code is an error-correcting code in which each linear combination of codewords are also in the coding alphabet
- Linear codes are vector spaces
- Easier to encode and decode
- Example: A = {000, 001, 010, 011}

Encoding Linear Codes

- Let C be a binary linear code with basis $\{r_1 \ldots r_k\}$
- C can represent 2^k pieces of information (words)
- Any codeword u can be written uniquely as: $u_1 r_1 + \ldots + u_k r_k$
- The process of representing these elements is called encoding

Decoding Linear Codes

- For non-linear codes, decoding can require exponential computing
- This is why we want linear codes to use in practice
- Nearest neighbor decoding: simple algorithm for decoding linear codes

The main coding theory problem

- Three parameters
  - d - Minimum (hamming) distance
  - n - Length of code words
  - M - Size of coding alphabet
- Given a fixed *n* and *d*, what is the largest possible size *M* that a code can achieve?
- We also examined fixing the other two parameters

Hamming Ball

- For alphabet A, a ball of radius *r* and center *u* is the set of vectors in A that have a distance ≤ r from center *u*.
- The size of a ball of radius *r* and vectors of length *n* is given by: $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{r}$ (for a binary code)

```
#r = radius, n = length
def ballsize(r, n):
    if (r == 1):
        return (1 + n)
    else:
        return fact(n)/(fact(r) * fact(n - r)) + ballsize(r - 1, n)
```

Our Approach

- Used Python to create computational algorithm
- Created a list to hold our optimal code and added 0 vector
- Generated a code that included each possible vector of at least distance d
- Continued until we had every possibility
- Kept track of best choice

More on Coding Theory

- Other possible paths:
  - Nonlinear codes
  - Nonbinary codes
- Coding Theory: A First Course - San Ling, Chaoping Xing