

Aspects of Coding Theory



RET-Mathematics
Summer 2005

Definitions

- ◆ A code is a set of symbol sequences used to transmit information.
- ◆ Any one of the symbol sequences permitted in the code is a code word, and words all have the same length.
- ◆ The set of allowed symbols is called an alphabet.

ASCII Code

DOCTOR FUN presents 1990 df1990-534



"Those bastards! According to the manual, they've used a proprietary encryption scheme called 'ASCII' code!"

Copyright © 2003 David Farley, d-farley@ibiblio.org <http://ibiblio.org> <http://ibiblio.org/Dave/drfun.html>

ASCII Code

- ◆ Question: What should everyone have at the symposium?
- ◆ Intended answer: 070 085 078
- ◆ Transmitted answer: 071 085 078
- ◆ Since the difference between the words is one digit, the Hamming distance is 1.

ASCII Code

- ◆ The code provides no ability for error detection.
- ◆ (Crude) solution: double up the code, e.g. 070070 085085 078078.
- ◆ Hamming distance between words is now 2, so that an error like 070071 would at least be detected, because it is not a valid code word.

ASCII Code

- ◆ The doubled code provides no ability for error correction, however.
- ◆ Solution: Triple the words, e.g. 070070070.
- ◆ Hamming distance is now 3, and, using a nearest-neighbor algorithm, 070071070 could be corrected to 070707070.

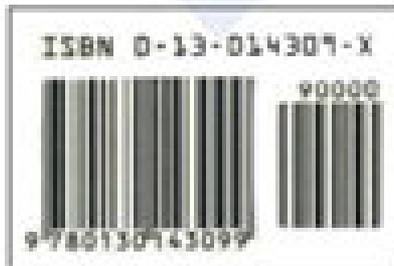
ASCII Code

- ◆ Advantages of solution
 - Allows error detection and correction.
- ◆ Disadvantages of solution
 - Increases transmission time for messages.

ISBN Numbers

You can find the
ISBN number
right here...

ISBN
number
isbn



What is an ISBN?

an ISBN is a unique 10 character
ID number given to every book.



(Back of Textbook)

ISBN Numbers

- ◆ The alphabet for ISBN numbers is the set $\{0, 1, 2, 3, 4, 5, 6, 8, 9, X\}$.
- ◆ Code words are formed by sequences of 10 symbols, e.g. 0-691-11321-1.
- ◆ The first symbol represents the language; the second group the publisher; the third group the title (randomly assigned).

ISBN Numbers

- ◆ A weighted total is calculated by multiplying each symbol by its position, e.g. 0-691-11321-1 yields $1*0+2*6+3*9+4*1+5*1+6*1+7*3+8*2+9*1+10*1=110$.
- ◆ The last symbol acts as a check digit, and is chosen so that the weighted total is divisible by 11.

ISBN Numbers

- ◆ Advantages of code
 - Allows error detection in multiple cases (incorrect check digit, swapped digits), which minimizes shipping errors.
- ◆ Disadvantages of code
 - Doesn't allow for error correction.
 - Major publishers could "run out" of code words.

Encoding/Decoding Information



Encoding/Decoding Information

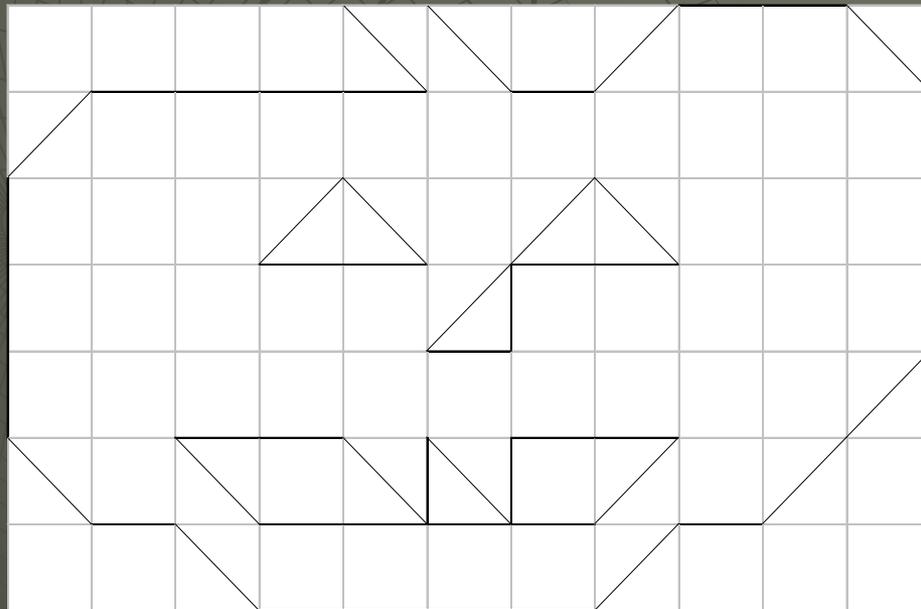
- ◆ As long as a code with a simple alphabet can be established, even complex information can be transmitted, e.g. pictures from Mars, sounds from the bottom of the ocean, or our super-secret plans to take over Notre Dame... oops.
- ◆ Carl Sagan's *Contact* included a subplot in which God encodes a message to mankind using successive ones and zeroes buried deep in the number pi.

Encoding/Decoding Information

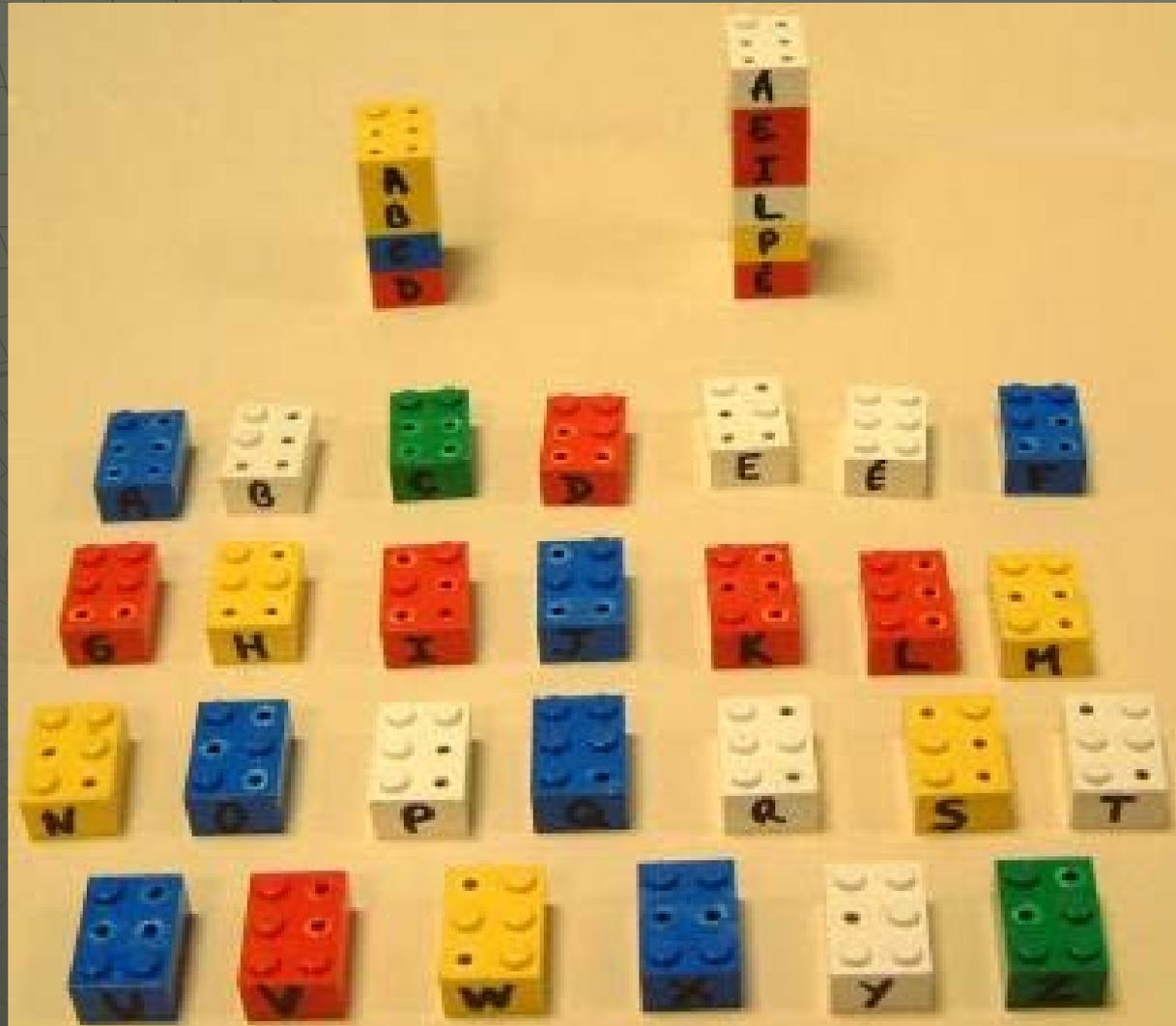
- ◆ When designing a code, a balance must be struck between transmission accuracy and encoding efficiency.
- ◆ In general, a larger Hamming distance corresponds to a less efficient encoding process.

Encoding/Decoding Information

- ◆ Transmitted message (77 digits):
00001302553255555555555555415415555
5555455555555555555555541530110254000
155554000



Braille



Braille

- ◆ The alphabet consists of specific combinations of six "1"s and "0"s to represent raised and unraised dots.
- ◆ The dots are arranged in a 3x2 matrix, but can easily be transmitted as a string of six digits.
- ◆ Some special characters indicate numbers and punctuation.

Connections to Mathematics

- ◆ Matrices
- ◆ Linear Algebra
- ◆ Modular Arithmetic
- ◆ Probability
- ◆ Networks

Participants

- ◆ Lori Bettcher
- ◆ Ben Dillon
- ◆ Terri Emerick
- ◆ Michael Largey
- ◆ Michelle McConahay
- ◆ Katy McShane
- ◆ Rich Sypel
- ◆ Bruce Urban
- ◆ Maggie Vogel

Advisers

- ◆ Margaret Doig
- ◆ Dr. Alex Hahn