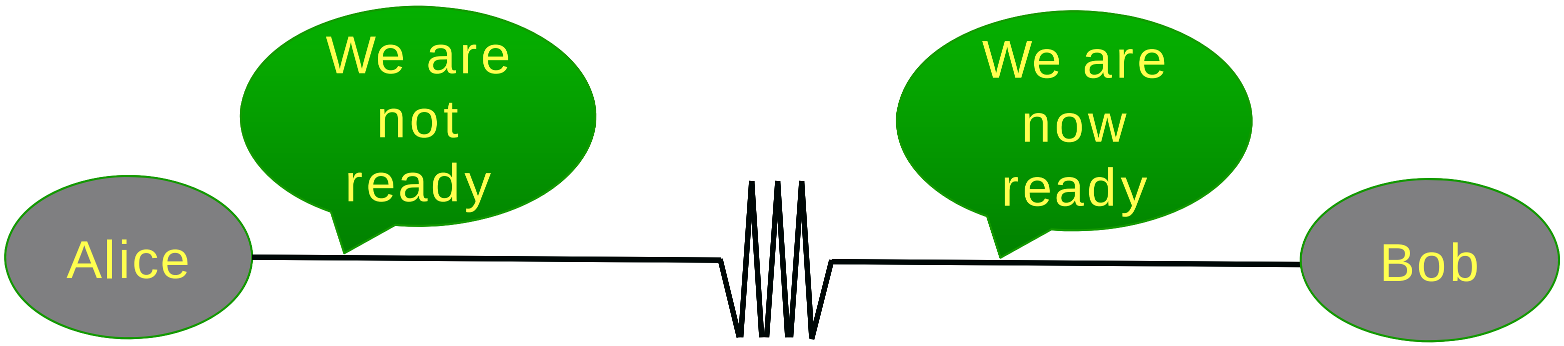
Reliable Meaningful Communication

Madhu Sudan
Microsoft, Cambridge, USA

Reliable Communication?

- Problem from the 1940s: Advent of digital age.



- Communication media are always noisy
  - But digital information less tolerant to noise!

Coding by Repetition

- Can repeat (every letter of) message to improve reliability:

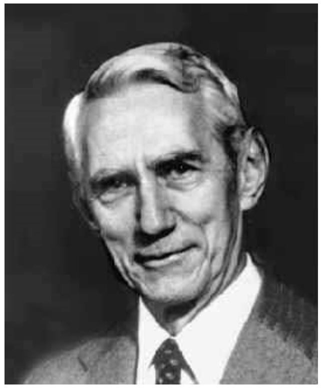
WWW EEE   AAA RRR EEE   NNN OOO WWW ...

↓

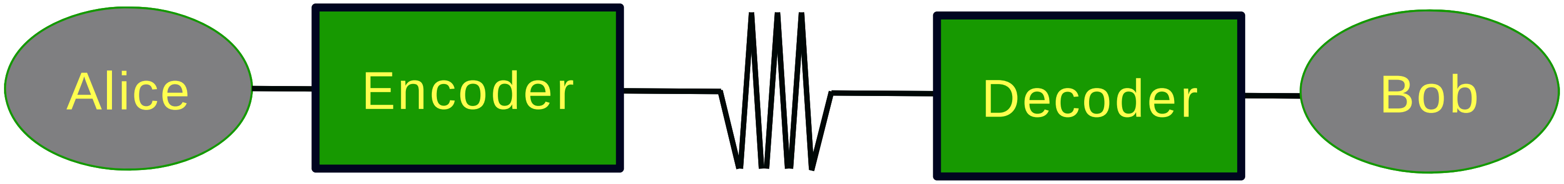
WXW EEA   ARA SSR EEE   NMN OOP WWW ...

- Calculations:
  - t repetitions $\Rightarrow$ Prob. Single symbol corrupted $\approx 2^{-t}$
  - To transmit k symbols, choose $t \approx \log k$
  - Rate of transmission $= \frac{1}{\log k} \to 0$ as $k \to \infty$
  - Belief (pre-1940s): Rate of *any* scheme $\to 0$ as $k \to \infty$

Shannon's Theory [1948]



- Sender "Encodes" before transmitting
- Receiver "Decodes" after receiving



- Encoder/Decoder arbitrary functions.

$$E: \{0,1\}^k \to \{0,1\}^n$$
$$D: \{0,1\}^n \to \{0,1\}^k$$

  - Rate = $\frac{k}{n}$;
  - Requirement: $m = D(E(m) + \text{error})$ w. high prob.
- What are the best E, D (with highest Rate)?

Shannon's Theorem

- If every bit is flipped with probability p
  - Rate $\rightarrow 1 - H(p)$ can be achieved.

$$H(p) \triangleq p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

  - This is best possible.
  - Examples:
    - $p = 0 \Rightarrow Rate = 1$
    - $p = \frac{1}{2} \Rightarrow Rate = 0$
    - Monotone decreasing for $p \in (0, \frac{1}{2})$
    - Positive rate for $p = 0.4999$; even if $k \rightarrow \infty$

Challenges post-Shannon

- Encoding/Decoding functions not "constructive".
  - Shannon picked E at random, D brute force.
  - Consequence:
    - D takes time $\sim 2^k$ to compute (on a computer).
    - E takes time 2^{2^k} to find!
- Algorithmic challenge:
  - Find E, D more explicitly.
  - Both should take time $\sim k, k^2, k^3$ … to compute

Explicit Codes: Reed-Solomon Code

- Messages = Coefficients of Polynomials.
  - Example:
    - Message = $(100{,}23{,}45{,}76)$
    - Think of polynomial $p(x) = 100 + 23x + 45x^2 + 76x^3$
    - Encoding: $(p(1), p(2), p(3), p(4), \ldots, p(n))$
    - First four values suffice, rest is redundancy!
  - (Easy) Facts:
    - Any k values suffice where k = length of message.
    - Can handle $n - k$ erasures or $(n - k)/2$ errors.
    - Explicit encoding ✓
    - Efficient decoding? [Peterson 1960]

More Errors? List Decoding

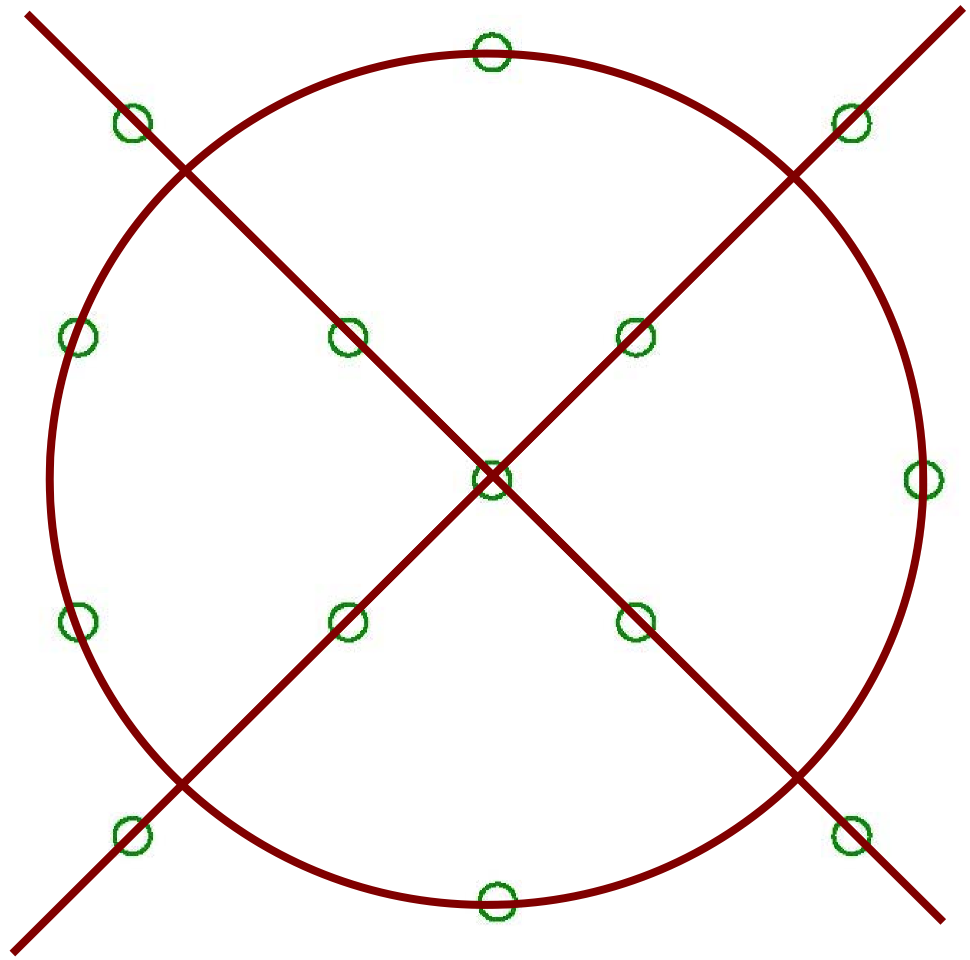
- Why was $(n-k)/2$ the limit for #errors?
  - $\frac{n}{2}$ is clearly a limit – right?
    - First half = evaluations of p_1
    - Second half = evaluations of p_2
    - What is the right message: p_1 or p_2?
- $\frac{n}{2}$ (even $\frac{n-k}{2}$) is the limit for "unique" answer.
- List-decoding: Generalized notion of decoding.
  - Report (small) list of possible messages.
  - Decoding "successful" if list contains the message polynomial.

Reed-Solomon List-Decoding Problem

- Given:
  - Parameters: n, k, t
  - Points: $(x_1, y_1), \dots, (x_n, y_n)$ in the plane
    (finite field actually)
- Find:
  - All degree k poly's that pass thru t of n points
    - i.e., all p s.t.
      - $\deg(p) < k$
      - $\#\{\, i \mid p(x_i) = y_i \,\} \geq t$
- $t \geq \frac{(n+k)}{2}$: Answer unique; [Peterson 60] finds it.
- [S. 96, Guruswami+S. '98]: $t \geq \sqrt{kn}$; small list

Decoding by example + picture [S'96]

$n = 14; k = 1; t = 5$



Algorithm idea:

- Find algebraic explanation of all points.

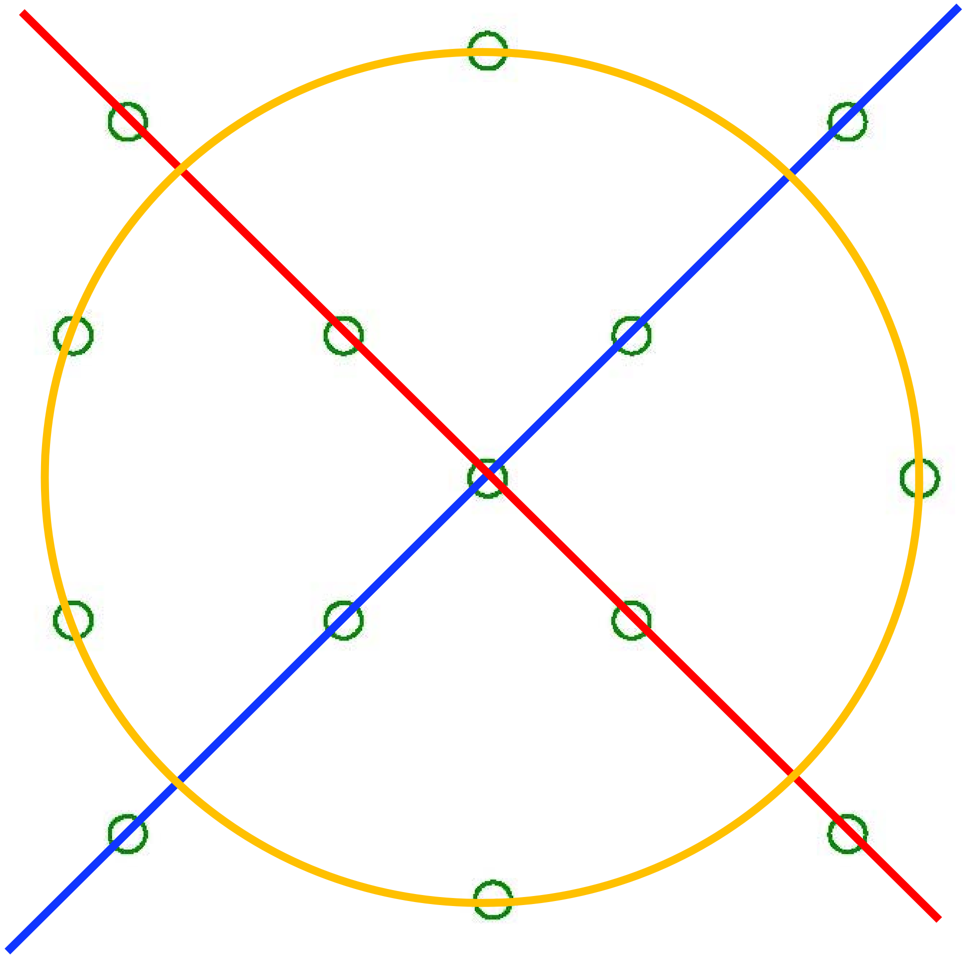
$$x^4 - y^4 - x^2 + y^2 = 0$$

- Stare at the solution ☺ (factor the polynomial)

$$(x + y)(x - y)(x^2 + y^2 - 1)$$

Decoding by example + picture [S'96]

$$n = 14; k = 1; t = 5$$



Algorithm idea:

- Find algebraic explanation of all points.

$$x^4 - y^4 - x^2 + y^2 = 0$$

- Stare at the solution ☺ (factor the polynomial)

$$(x+y)\,(x-y)\,(x^2+y^2-1)$$

Decoding Algorithm

- Fact: There is always a degree $2\sqrt{n}$ polynomial thru n points
  - Can be found in polynomial time (solving linear system).
- [80s]: Polynomials can be factored in polynomial time [Grigoriev, Kaltofen, Lenstra]
- Leads to (simple, efficient) list-decoding correcting κ fraction errors for $\kappa \to 1$

Summary and conclusions

- (Many) errors can be dealt with:
  - Pre-Shannon: vanishing fraction of errors
  - Pre-list-decoding: small constant fraction
  - Post-list-decoding: overwhelming fraction

- Future challenges?
  - Communication can overcome errors introduced by channels.
  - Can communication overcome errors in misunderstanding between sender and receiver?
    - [Goldreich,Juba,S. ‘2011]; [Juba,Kalai,Khanna,S.’2011] ....

Thank You!