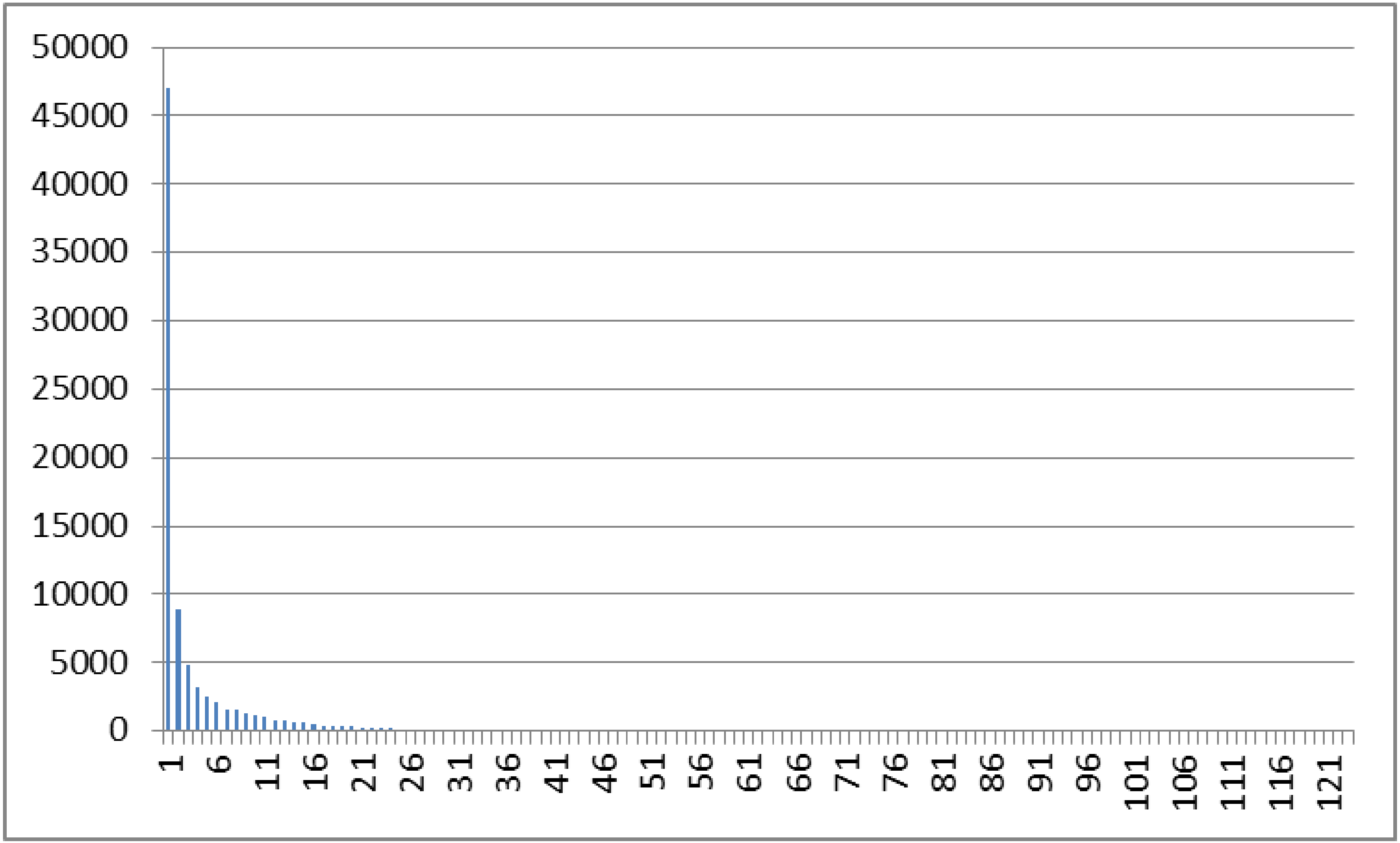
VLC for Integers

20CS6025 Data Encoding

Yizong Cheng

1-20-15

Entropy = 2.68 after BW



Coding Integers

- When smaller integers occurs more often, we want to use fewer bits to represent them in a variable-length self-delimiting code.
- Best: minimum-length binary code.
  - $\log_2 N$ for integer $N > 0$
- Problem: not self-delimiting.
- Challenge: Find variable-length codewords for integers so that length of codeword for N is very close to $\log_2 N$.

Unary Code

- The unary code of integer N is N 1’s followed by a 0.
- Levenstein 1968, Elias 1975. α code: (N-1) 0’s followed by a 1 for coding positive integers.
- Optimal when probability Pr(N) < Pr(N + 1) / 2.
- Base 1 representation of integers.
  - 0, 10, 110, 1110, 11110, 111110
  - 1, 01, 001, 0001, 00001, 000001
  - 1, 2, 3, 4, 5, 6

Binary Code

- Base 2 representation of integers.
- Requires extras length information.
- Also called Elias β code
  - 0, 1, 10, 11, 100, 101, 110, 111, 1000
  - 0, 1, 2, 3, 4, 5, 6, 7, 8
  - 1, 10, 11, 100, 101, 110, 111, 1000
  - 1, 2, 3, 4, 5, 6, 7, 8
- Observation: leading 1 is unnecessary
  - , 0, 1, 00, 01, 10, 11, 000 for 1,2,3,4,5,6,7,8

γ Code

- This is a way to make binary code self-delimiting.
- For a binary codeword of k numeric bits, we add k 0's followed by a 1.
- The codewords for 1,2,3,4,5,6,7,8 are
  - 1, 010, 011, 00100, 00101, 00110, 00111, 0001000.
- Length is encoded as unary.

Peter Elias

$1=2^0+0=1$	$10=2^3+2=0001\mathit{010}$
$2=2^1+0=01\mathit{0}$	$11=2^3+3=0001\mathit{011}$
$3=2^1+1=01\mathit{1}$	$12=2^3+4=0001\mathit{100}$
$4=2^2+0=001\mathit{00}$	$13=2^3+5=0001\mathit{101}$
$5=2^2+1=001\mathit{01}$	$14=2^3+6=0001\mathit{110}$
$6=2^2+2=001\mathit{10}$	$15=2^3+7=0001\mathit{111}$
$7=2^2+3=001\mathit{11}$	$16=2^4+0=00001\mathit{0000}$
$8=2^3+0=0001\mathit{000}$	$17=2^4+1=00001\mathit{0001}$
$9=2^3+1=0001\mathit{001}$	$18=2^4+2=00001\mathit{0010}$

Table 3.10: 18 Elias Gamma Codes.

MIT Professor Peter Elias dies at 78; Was computer science pioneer

Robert J. Sales, News Office

December 10, 2001

Share

CAMBRIDGE, Mass. -- MIT Professor Emeritus Peter Elias, a pioneer in the field of computer science, died of Creutzfeld-Jakob Disease last Friday in the living room of his Cambridge home, surrounded by his family. He was 78 years old.

"Many of us will remember Peter as a quiet and unassuming colleague who contributed a great deal to education and research at MIT," said Victor W. Zue, director of the Laboratory for Computer Science. "He was one of the most energetic emeritus professors I know -- coming to work almost every day and continuing to advise undergraduate students."

"In addition to being a distinguished researcher and educator, Peter was one of MIT's great citizens," said Professor John V. Guttag, head of the Department of Electrical Engineering and Computer Science (EECS). "He will be sorely missed."

Professor Elias was born on Nov. 23, 1923 in Brunswick, N.J., the son of an engineer in Thomas A. Edison's laboratory. He was a member of the MIT faculty from 1953 to 1991, at which time he assumed emeritus rank and became a senior lecturer.

Professor Elias attended Swarthmore College for two years before transferring to MIT in

multimedia

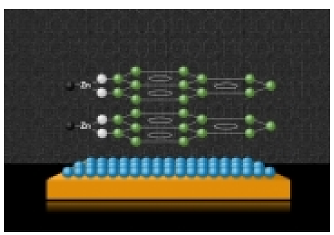
Peter Elias

tags

obituaries

today's news

Storing data in individual molecules



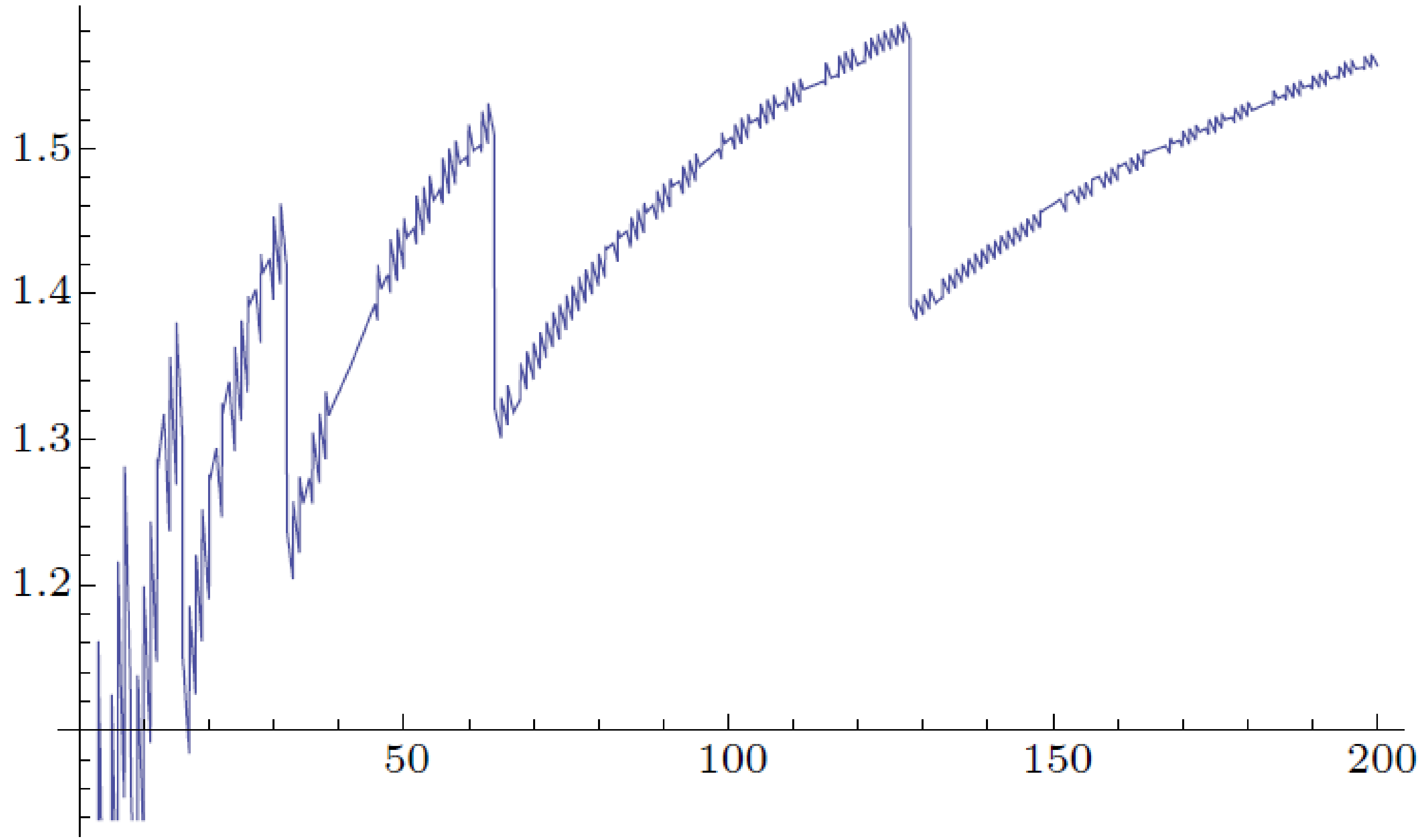
The new molecules are known as 'graphene fragments,' because they largely consist of flat sheets of carbon (which are attached to zinc atoms). That makes them easier to align during deposition, which could simplify the manufacture of molecular memories.
Graphic: Christine Daniloff/MIT

An international team of researchers demonstrates the possibility of molecular memory near room

gammacode.txt

0
100
101
11000
11001
11010
11011
1110000
1110001
1110010
1110011
1110100
1110101
1110110
1110111
111100000
111100001
111100010
111100011
111100100
111100101
111100110
111100111
111101000
111101001
111101010
111101011
111101100
111101101
111101110
111101111
11111000000
11111000001
11111000010
11111000011
11111000100

Gamma Code Length vs $\log_2 N$




```
gamma[i_] := 1. + 2 Floor[Log[2, i]];
Plot[Sum[gamma[j], {j,1,n}]/(n Ceiling[Log[2,n]]), {n,1,200}]
```

Figure 3.11: Gamma Code Versus Binary Code.

Fractional Bases

- Binary code for N has $\log_2 N$ bits.
- Unary code for N has $\log_1 N$ bits.
- γ code for integer N has $2\log_2 N$ bits, or twice as long as the binary code.
- Let $2\log_2 N = \log_b N = x$. $b=?$
- $N=b^x=2^{x/2}=(2^{1/2})^x$. $b=2^{1/2}=1.414$.

δ and ω Codes

- γ code is binary code preceded by a unary code of its length.
- The length itself can be γ coded.
- The result is the δ code, good for larger numbers.
- So are the length of length, length of length of length,… until length is representable by 2 bits.
- The result is the ω code.

$1 = 2^0 + 0 \rightarrow \lvert L \rvert = 0 \rightarrow 1$	$10 = 2^3 + 2 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{010}$
$2 = 2^1 + 0 \rightarrow \lvert L \rvert = 1 \rightarrow 010\mathit{0}$	$11 = 2^3 + 3 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{011}$
$3 = 2^1 + 1 \rightarrow \lvert L \rvert = 1 \rightarrow 010\mathit{1}$	$12 = 2^3 + 4 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{100}$
$4 = 2^2 + 0 \rightarrow \lvert L \rvert = 2 \rightarrow 011\mathit{00}$	$13 = 2^3 + 5 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{101}$
$5 = 2^2 + 1 \rightarrow \lvert L \rvert = 2 \rightarrow 011\mathit{01}$	$14 = 2^3 + 6 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{110}$
$6 = 2^2 + 2 \rightarrow \lvert L \rvert = 2 \rightarrow 011\mathit{10}$	$15 = 2^3 + 7 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{111}$
$7 = 2^2 + 3 \rightarrow \lvert L \rvert = 2 \rightarrow 011\mathit{11}$	$16 = 2^4 + 0 \rightarrow \lvert L \rvert = 4 \rightarrow 00101\mathit{0000}$
$8 = 2^3 + 0 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{000}$	$17 = 2^4 + 1 \rightarrow \lvert L \rvert = 4 \rightarrow 00101\mathit{0001}$
$9 = 2^3 + 1 \rightarrow \lvert L \rvert = 3 \rightarrow 00100\mathit{001}$	$18 = 2^4 + 2 \rightarrow \lvert L \rvert = 4 \rightarrow 00101\mathit{0010}$

Table 3.12: 18 Elias Delta Codes.

1	0	10	11 1010 0
2	10 0	11	11 1011 0
3	11 0	12	11 1100 0
4	10 100 0	13	11 1101 0
5	10 101 0	14	11 1110 0
6	10 110 0	15	11 1111 0
7	10 111 0	16	10 100 10000 0
8	11 1000 0	17	10 100 10001 0
9	11 1001 0	18	10 100 10010 0

Table 3.13: 18 Elias Omega Codes.

1	1	1	2
2	3	4	3
3	3	4	4
4	5	5	4
5–7	5	5	5
8–15	7	8	6–7
16–31	9	9	7–8
32–63	11	10	8–10
64–88	13	11	10
100	13	11	11
1000	19	16	16
10^4	27	20	20
10^5	33	25	25
10^5	39	28	30

Table 3.14: Lengths of Three Elias Codes.

Fibonacci Numbers

- Recursively defined as
- $F_i = F_{i-2}+F_{i-1}$, with $F_1=F_2=1$
- F_{k+1}/F_k tends to the golden ratio
- $(1 + 5^{1/2})/2 = 1.618034\ldots$
  - 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89,…
- Order-3 Fibonacci numbers
- $F_i=F_{i-3}+F_{i-2}+F_{i-1}$, with $F_1=F_2=1$, $F_3=2$
  - 1, 1, 2, 4, 7, 13, 24, 44, 81, 149,…

Zeckendorf Representation

- 1, 2, 3, 5, 8, 13, 21, 34, 55, 89,…
- Any positive integer can be represented as sum of Fibonacci numbers.
  - 19=13+5+1: 100101
- Never has two adjacent 1's
- Append another 1 to delimit
- Fraenkel and Klein 1996 C^1 code
  - $C^1(19)$ = 1001011

1, 2, 3, 5, 8, 13, 21, 34, 55, 89

1	11	7	01011
2	011	8	000011
3	0011	9	100011
4	1011	10	010011
5	00011	11	001011
6	10011	12	101011

Table 3.37: Twelve Fibonacci Codes.

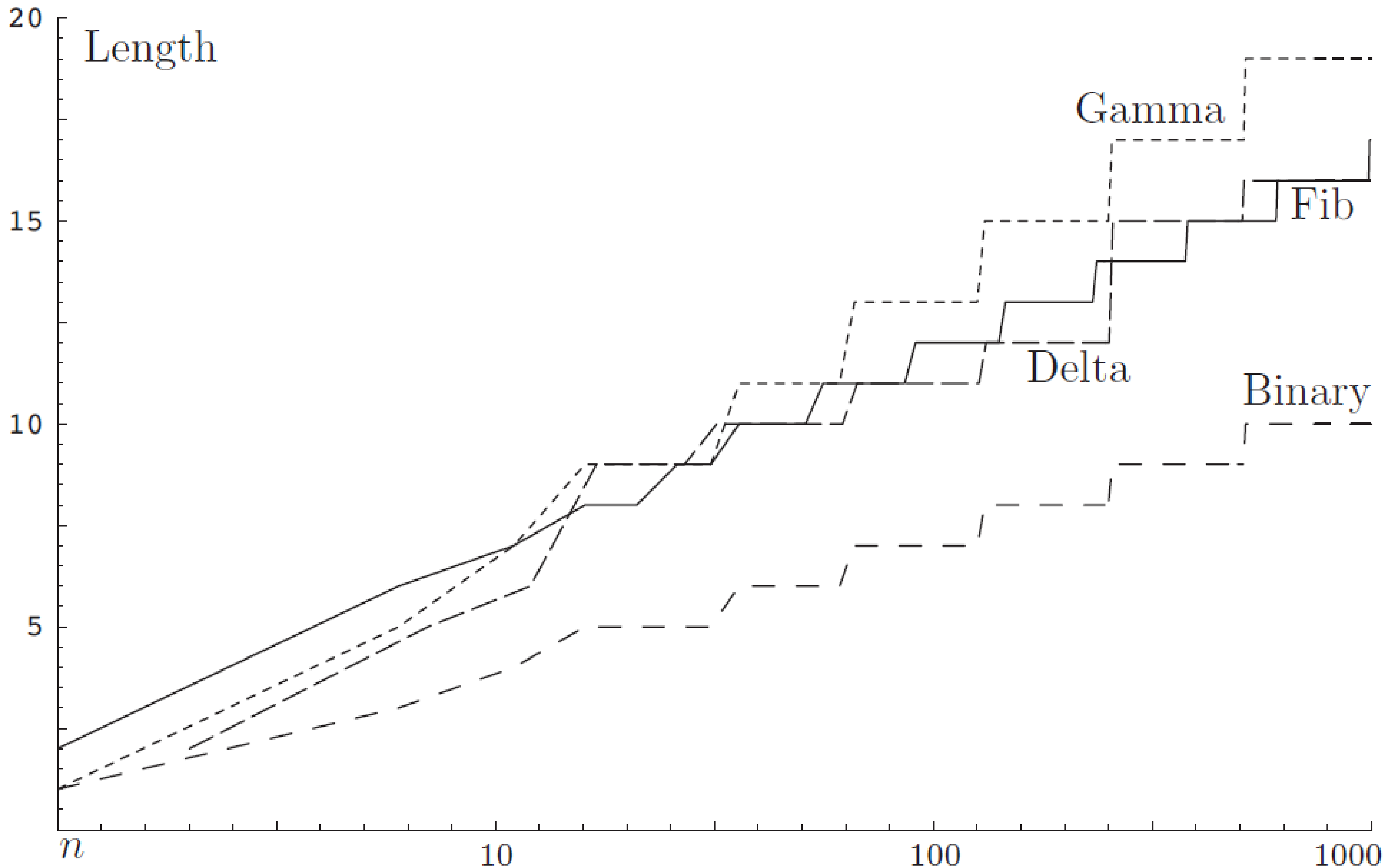
fibonaccicode.txt

11
011
0011
1011
00011
10011
01011
000011
100011
010011
001011
101011
0000011
1000011
0100011
0010011
1010011
0001011
1001011
0101011
00000011
10000011
01000011
00100011
10100011
00010011
10010011
01010011
00001011
10001011
01001011
00101011
10101011
000000011
100000011

Fractional Base for C^1

- Since the ratio $F_{i+1}/F_i = 1.618$, the length of C^1 code approaches $\log_{1.618}N$ for the representation of N.
- If order-3 Fibonacci is used, this ratio and also the base approach 1.84, really close to the binary representation of base 2.

Comparison of 4 Codes



Homework 3: due 1-26-15

- Complete H3B.java to decode Gamma coded files generated by java H3A gammacode.txt.
- Submit your source code writings and a test on a text file.