Data Compression
Lecture 6

Dr. Mona M.Soliman

Faculty of Computers and Artificial Intelligence

Cairo University

Fall 2022

Arithmetic Coding

Drawback of Huffman Coding

The main drawback of Huffman scheme is that it has problems when there is a symbol with very high probability

Minimum Codeword Length for any Symbol is "ONE" bit, there is no "Half Bit" or "Quarter Bit"

Arithmetic coding completely bypasses the idea of replacing an input symbol with a specific code.

Instead, *it takes a stream of input symbols and replaces it with a single floating point number*

The longer and more complex the message, the more bits are needed to represents the output number

Arithmetic Coding Algorithm

- Find the probability for each Symbol,
- Calculate Low_Range, High_Range of Symbol
- Divide the initial subinterval based on the distribution

  (Low_Range, High_Range of Symbols)
- The first Symbol is coded by extracting the subinterval corresponding to it, and subdivide again based on the same relative distribution
- Repeat last step for each Symbol until a final interval is determined
- Any value within this subinterval can be used to represent the sequence of Symbols

Example(1) Arithmetic Coding

Given a long Sequence of characters A, B, and C

the probabilities of Characters are

- P(A)=0.8
- P(B)=0.02
- P(C)= 0.18
- Compress the following part of the sequence "**ACBA**" using Arithmetic Coding

Arithmetic Coding

Rules To solve

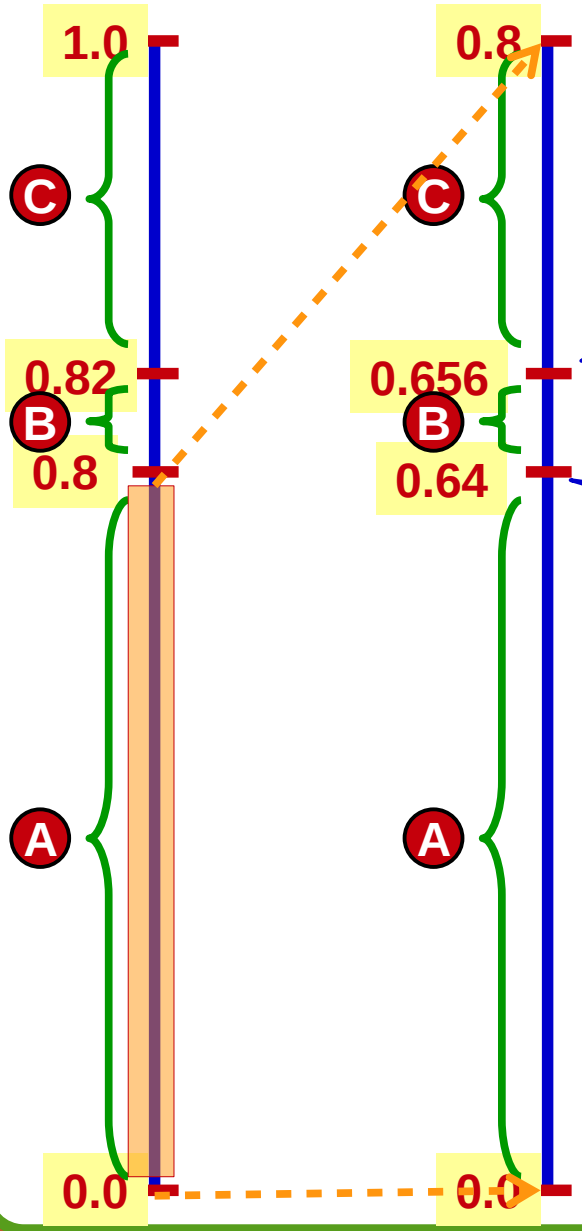
Low_range(A)=0
High_Range(A)=0.8

Low_range(B)=0.8
High_Range(B)=0.82

Low_range(C)=0.82
High_Range(C)=1.0

Lower (Symbol) = Lower+ Range * Low_Range(Symbol)
Upper (Symbol)= Lower+ Range * High_Range(Symbol)

Example(1) Arithmetic Coding



Compress Symbol "A"

0+ (0.8-0)*0.82

0+ (0.8-0)*0.8

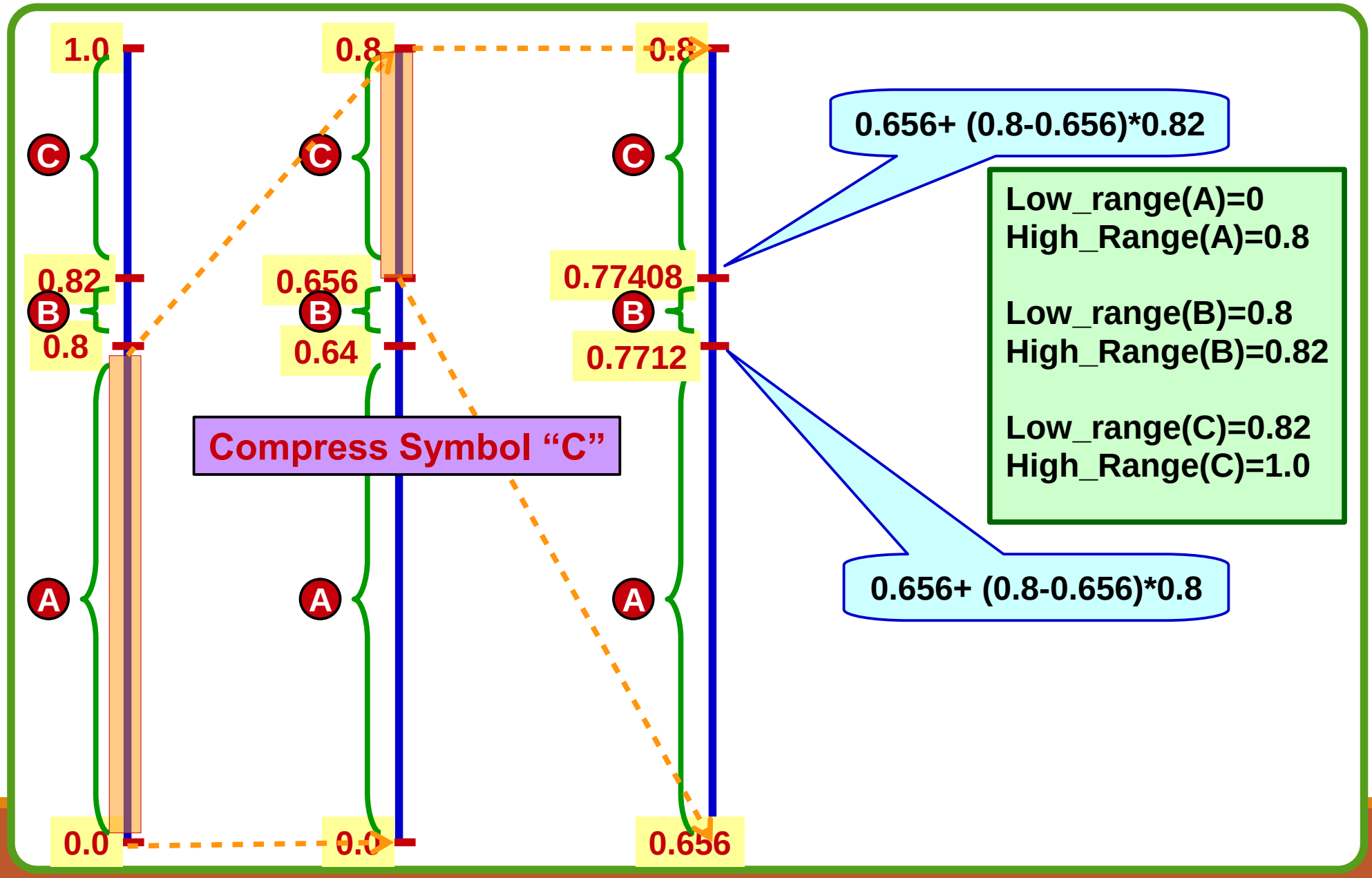
Low_range(A)=0
High_Range(A)=0.8

Low_range(B)=0.8
High_Range(B)=0.82

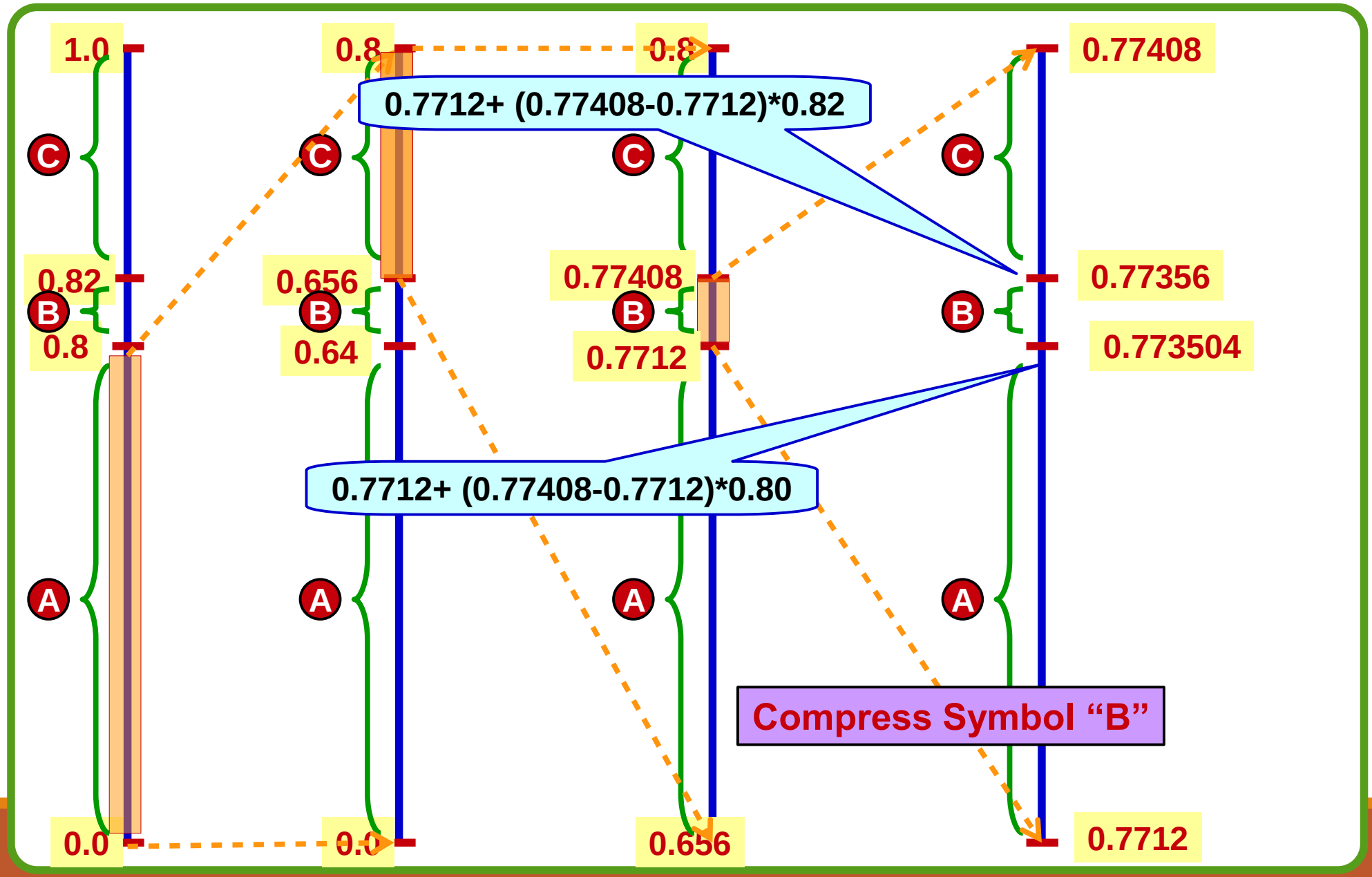
Low_range(C)=0.82
High_Range(C)=1.0

Lower (Symbol) = Lower+ Range * Low_Range(Symbol)
Upper (Symbol)= Lower+ Range * High_Range(Symbol)

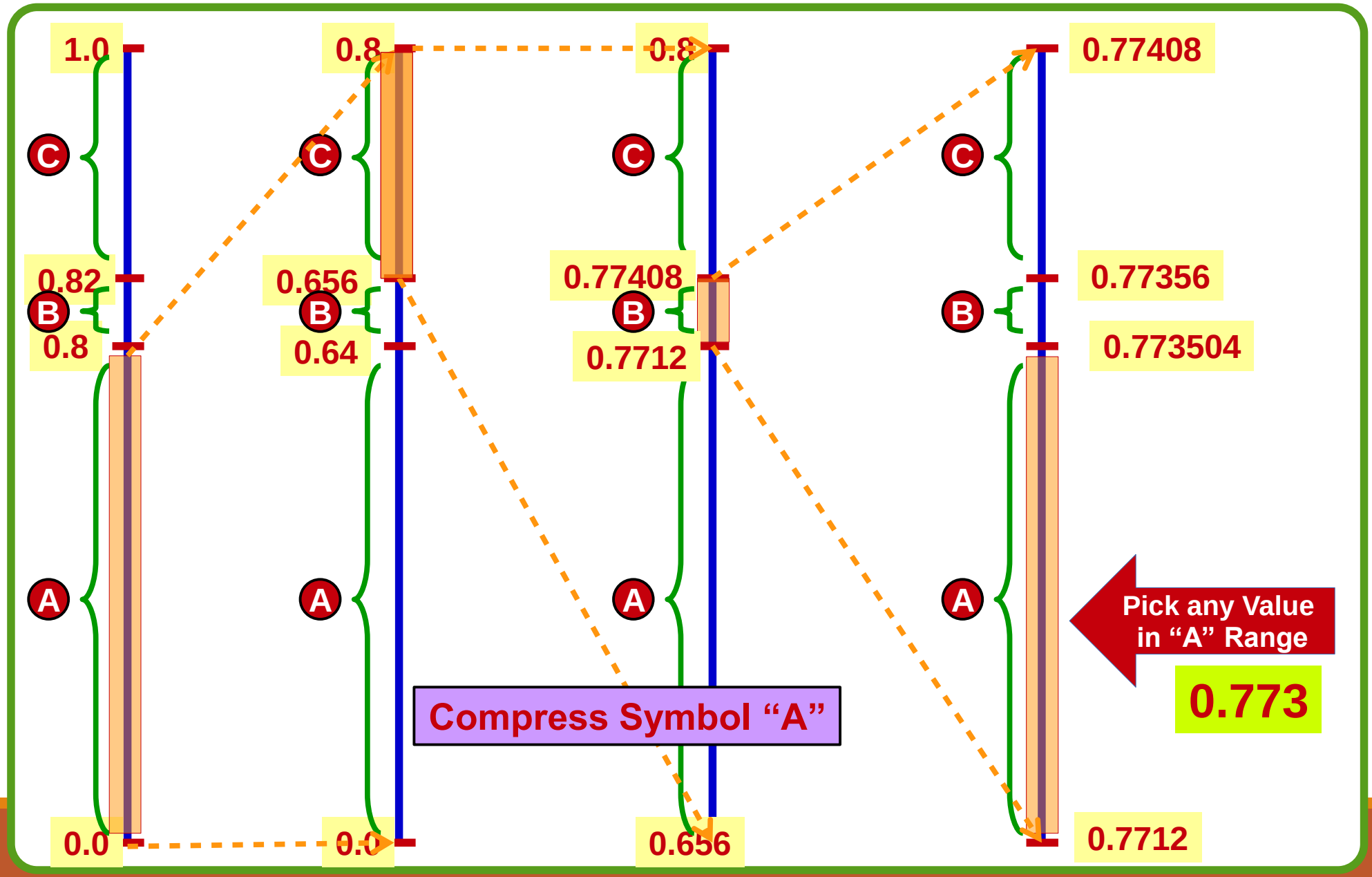
Example(1) Arithmetic Coding



Example(1) Arithmetic Coding



Example(1) Arithmetic Coding



Example(1) Arithmetic Coding

Lower (Symbol) = Lower+ Range * Low_Range(Symbol)
Upper (Symbol)= Lower+ Range * High_Range(Symbol)

Method 1: Using Equations

Low_range(A)=0
High_Range(A)=0.8

Low_range(B)=0.8
High_Range(B)=0.82

Low_range(C)=0.82
High_Range(C)=1.0

First Symbol is "A"

Lower(A)=0

Upper(A)=0.8

Second Symbol is "C"

Lower(C)=0+ (0.8-0)*0.82=0.656

Upper(C)=0+ (0.8- 0)*1= 0.8

Third Symbol is "B"

Lower (B)= 0.656+ (0.8-0.656)*0.8=0.7712

Upper(B)=0.656+(0.8-0.656)*0.82=0.77408

Fourth Symbol is "A"

Lower (A)= 0.7712+ (0.77408-0.7712)*0=0.7712

Upper(A)=0.7712+(0.77408-0.7712)*0.8=0.773504

Pick any Value in "A" Range → **0.773**

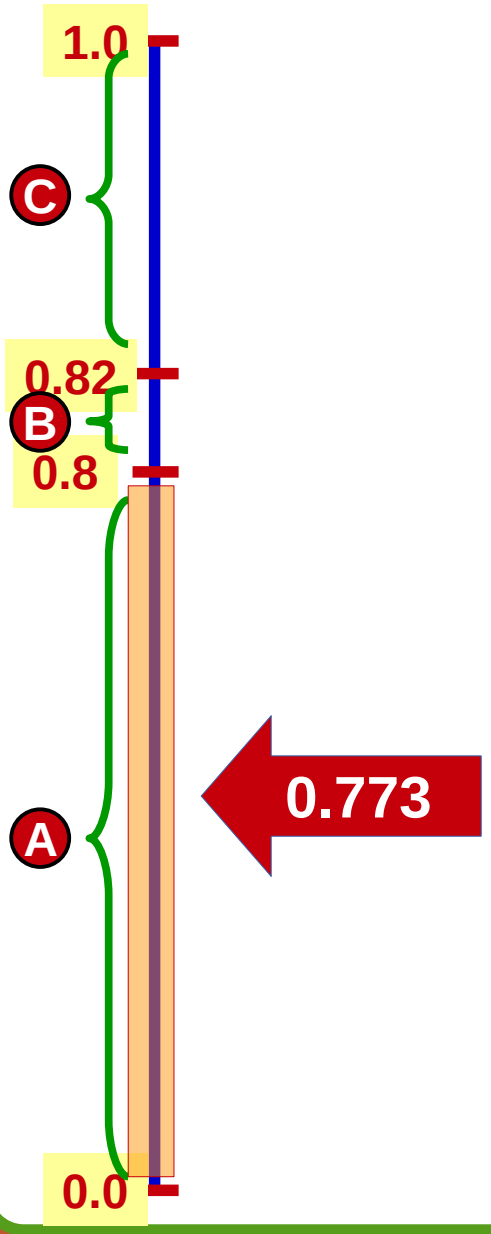
Decoding Algorithm

Algorithm Arithmetic Decoding

```
/*
Input: code: binary code
  Low_Range(Symbol):Lower Range of Symbol (in
Accumulative Probabilities Scale)
  High_Range(Symbol):Upper Range of Symbol (in
Accumulative Probabilities Scale)
Output: The decoded message
*/
 value = convert2decimal(code);
 Do {
  find a symbol s so that
```

Example(1) Arithmetic Coding
Decoding

Method 1: Using Graph



Compression Code = 0.773
0.0 <0.773 <0.8
First Symbol is "A"

Low_range(A)=0
High_Range(A)=0.8

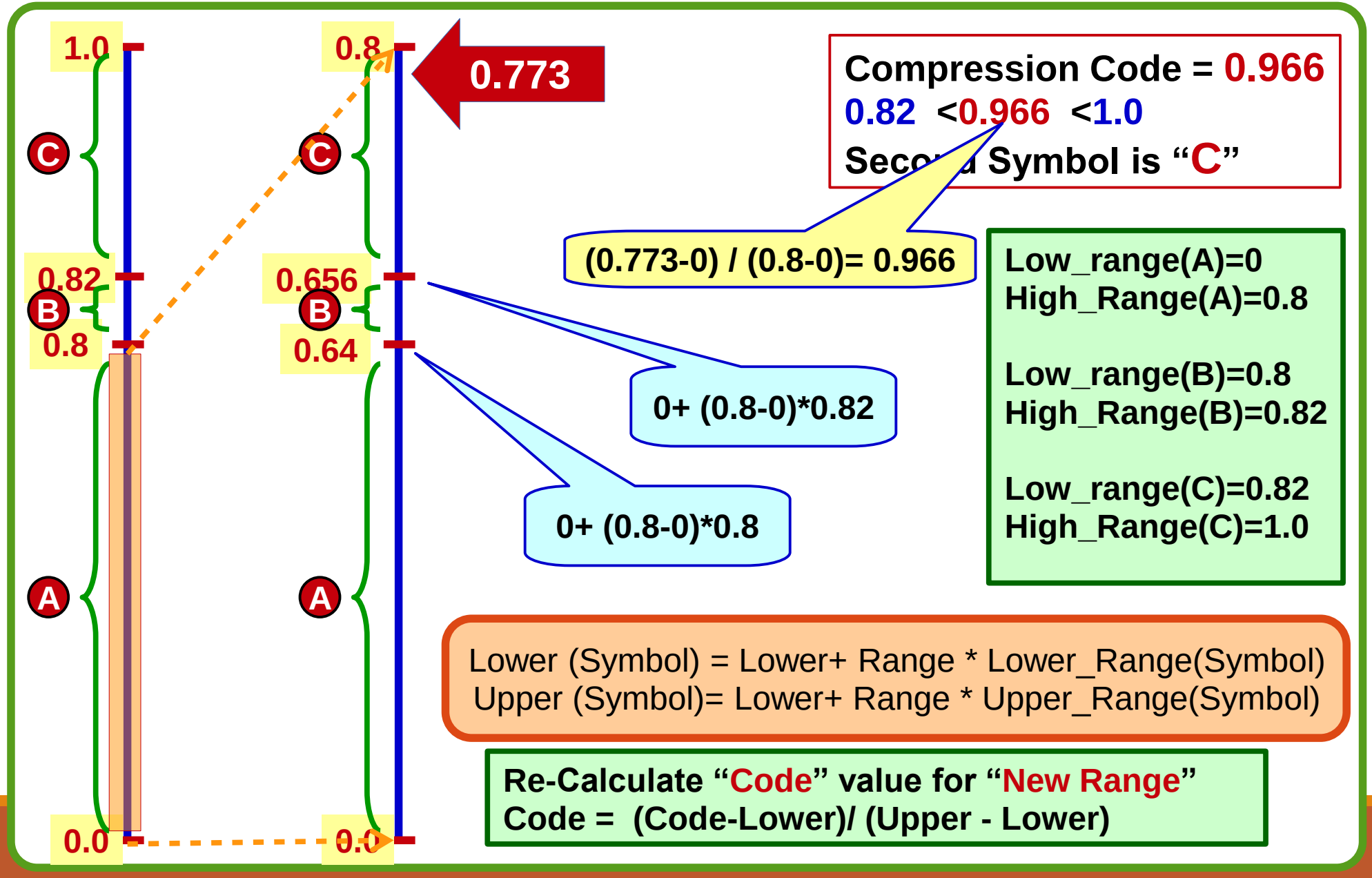
Low_range(B)=0.8
High_Range(B)=0.82

Low_range(C)=0.82
High_Range(C)=1.0

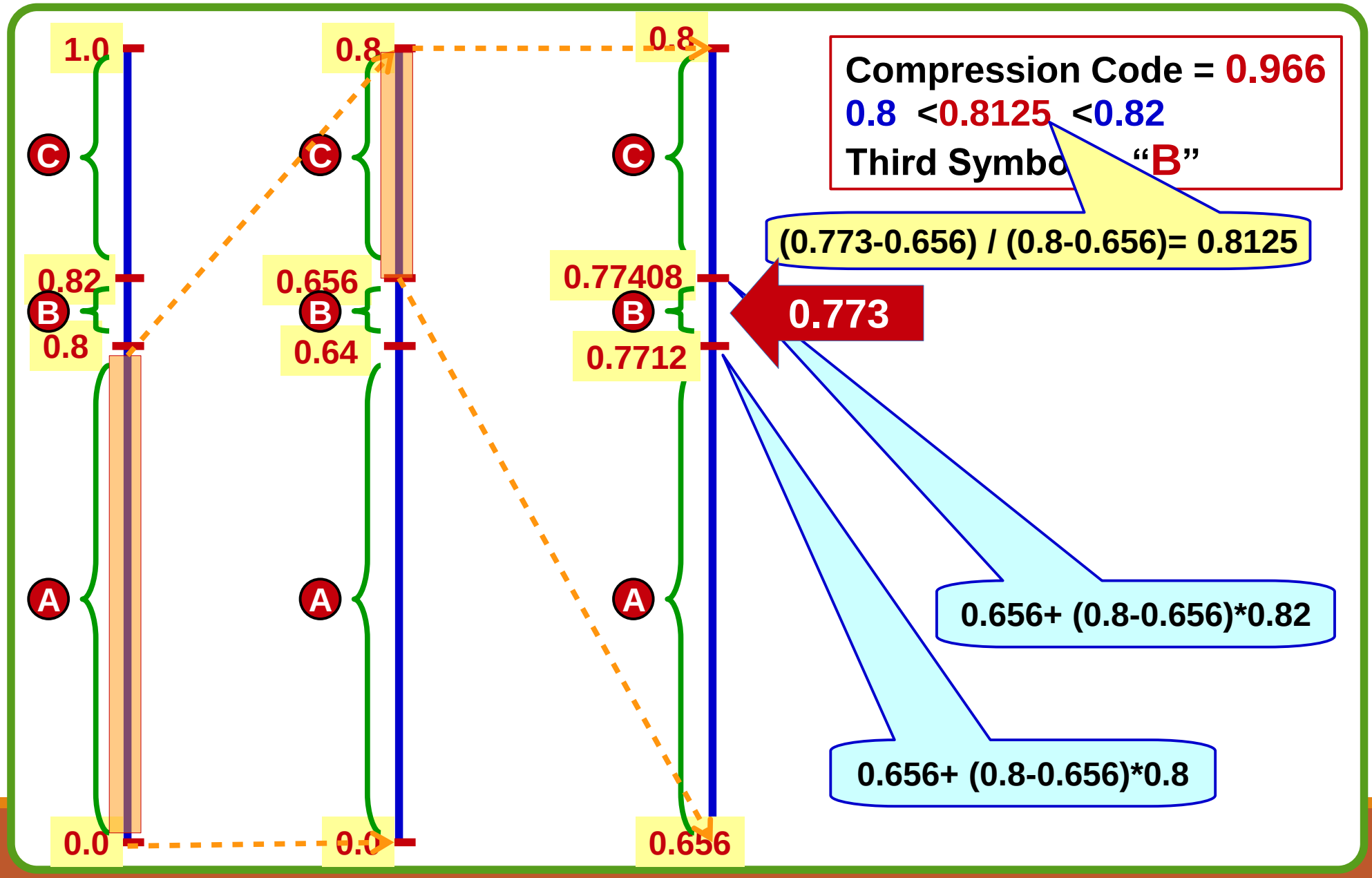
Lower (Symbol) = Lower+ Range * Lower_Range(Symbol)
Upper (Symbol)= Lower+ Range * Upper_Range(Symbol)

Re-Calculate "Code" value for "New Range"
Code = (Code-Lower)/ (Upper - Lower)

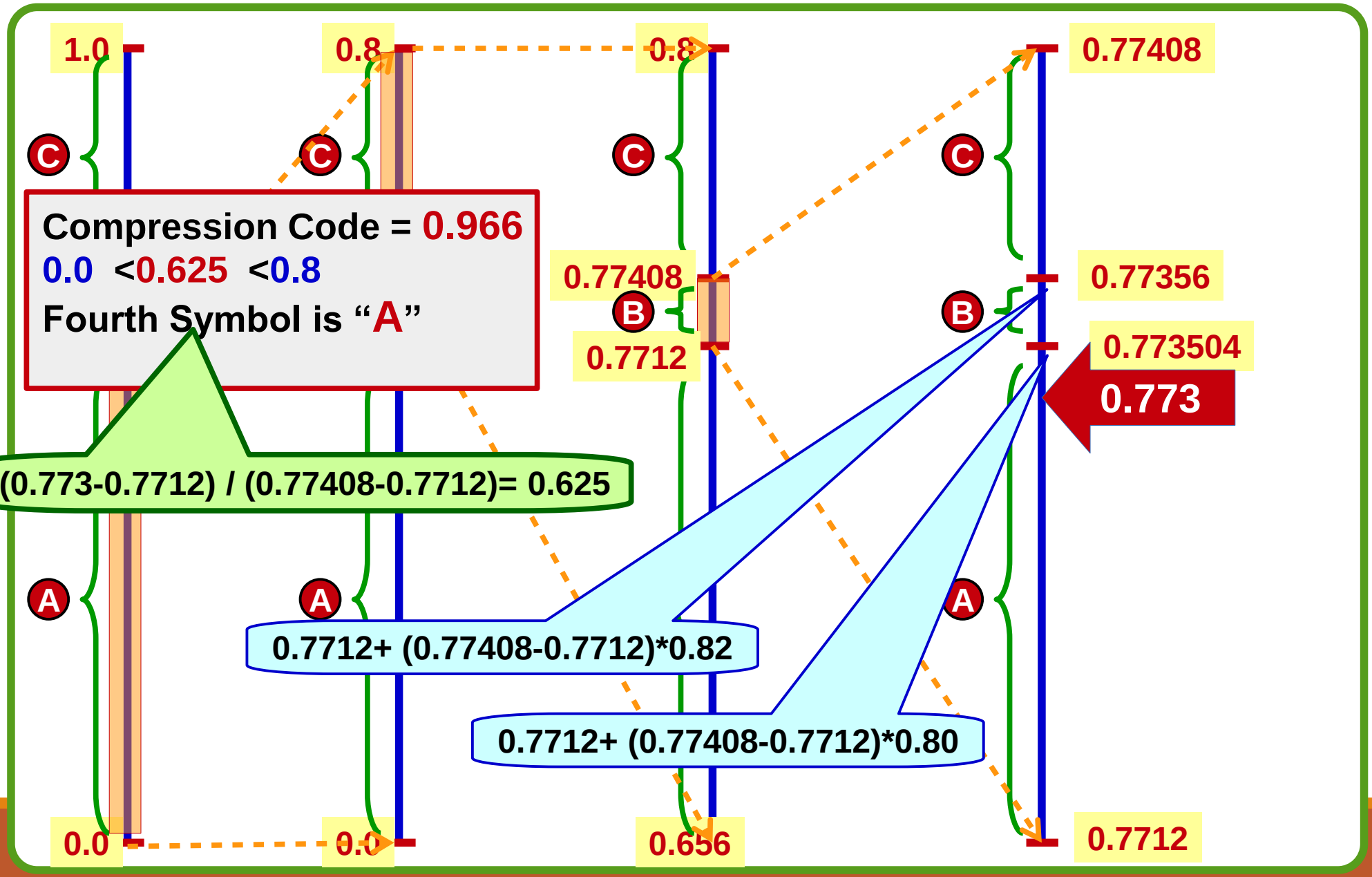
Example(1) Arithmetic Coding



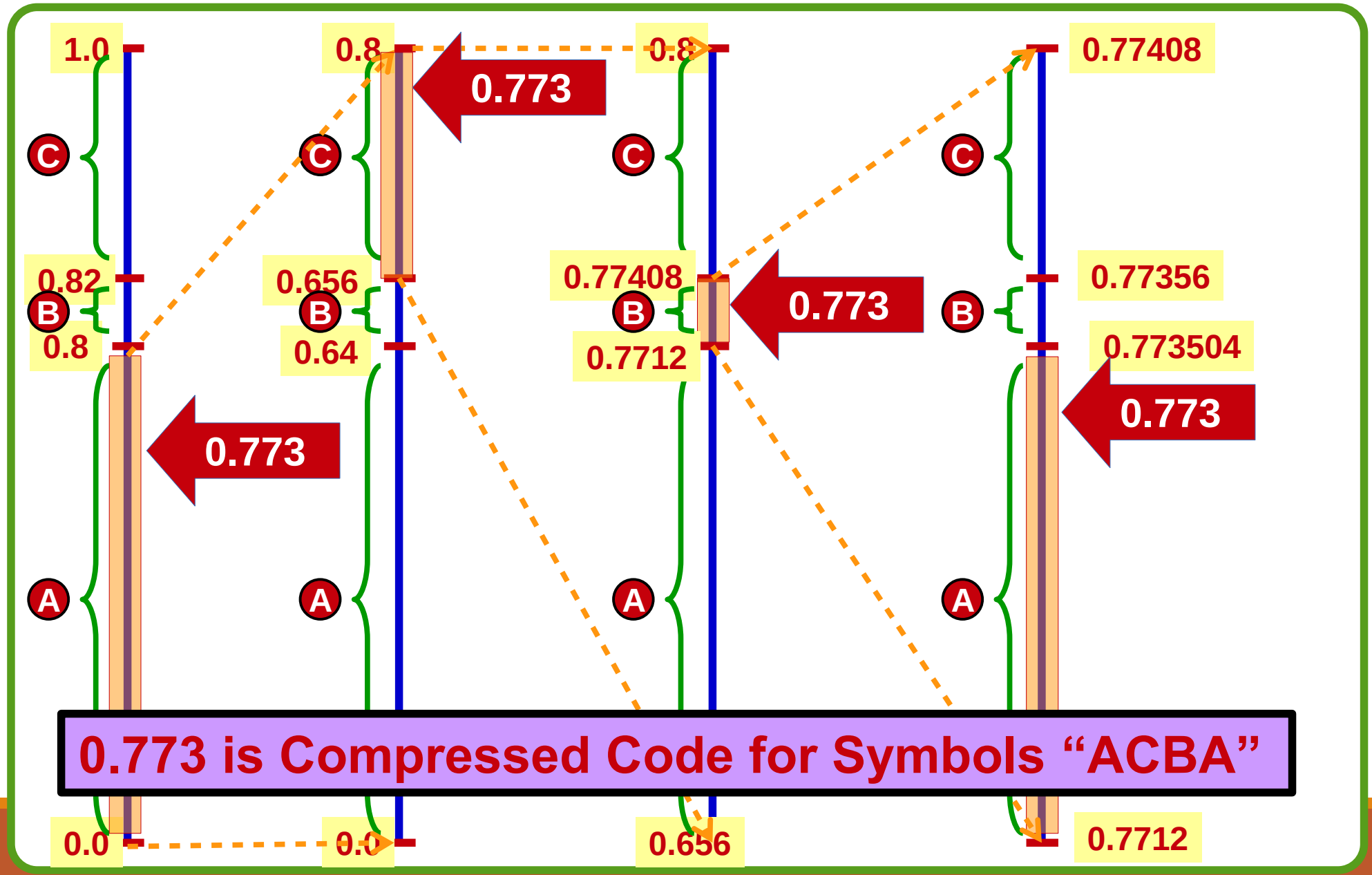
Example(1) Arithmetic Coding



Example(1) Arithmetic Coding



Example(1) Arithmetic Coding



Example(1) Arithmetic Coding (Decompression)

Method 1: Using Equations

Compression Code = 0.773

0.0 <0.773 <0.8

First Symbol is "A"

Lower(A)=0

Upper(A)=0.8

Code=(0.773-0) / (0.8-0)= 0.966

0.82 <0.966 <1.0

Second Symbol is "C"

Lower(C)=0+ (0.8-0)*0.82=0.656

Upper(C)=0+ (0.8- 0)*1= 0.8

Code= (0.733-0.656)/ (0.8 – 0.656)=0.8125

0.8 <0.8125 <0.82

Third Symbol is “B”

Lower (B)= 0.656+ (0.8-0.656)*0.8=0.7712

Upper(B)=0.656+(0.8-0.656)*0.82=0.77408

Code=(0.773-0.7712)/(0.77408-0.7712)=0.625

0.0 <0.625 <0.8

Fourth Symbol is “A”