

# Information theory and coding

## نظرية المعلومات و الترميز

مدرسة المقرر

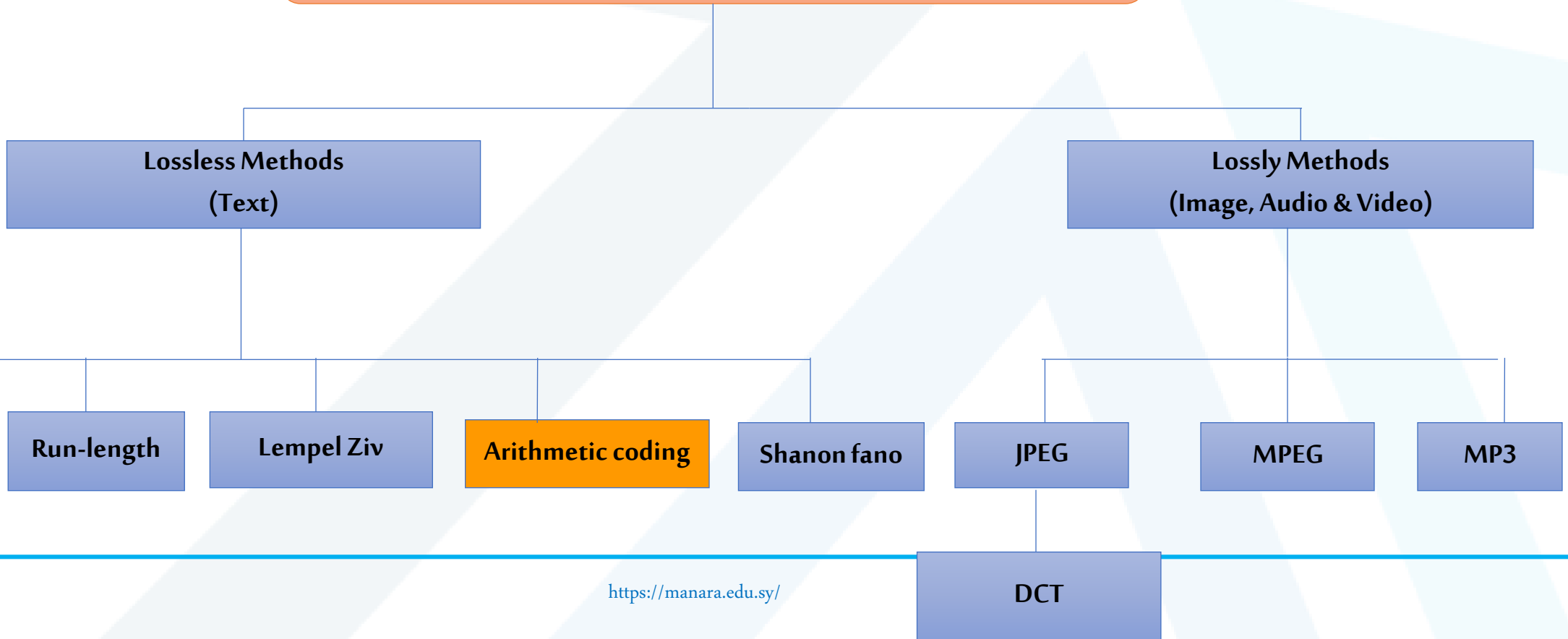
د. بشري علي معلا

## مفردات المحاضرة

الترميز الحسابي ➤ Arithmetic Code

## طرائق ضغط البيانات

### Data Compression Methods



## تعريف الترميز الحسابي Arithmetic Coding

- ✓ أحد أنواع ترميز الانتروبيا، وهو يمثل تقنية ترميز دون خسارة.
- ✓ لا يستخدم رقماً محدداً من البتات لكل حرف، بل ينتج رمزاً وحيداً (عددًا حقيقياً) لعدة محارف.
- ✓ ويحقق بذلك ضغطاً مثالياً، ونسبة ضغط أعلى من ترميز هوفمان.
- ✓ يستخدم بشكل واسع في ضغط النصوص.
- ✓ تقوم فكرة هذا الترميز على أخذ مجال يبدأ بالصفراً وأكبر قيمة هي الواحد، ويقسم هذا المجال حسب الاحتمالات التكرارية للحروف أو للقيم أو المعلومات المراد ترميزها ويستخدم في JPEG2000 and Mpeg4 .

## سلبيات الترميز الحسابي (AC Disadvantages)

- ✓ سرعته المنخفضة.
- ✓ لا تبدأ عملية فك الترميز حتى يتم استقبال كامل كلمة الترميز.
- ✓ عند وجود بت خاطئ ضمن كلمة الترميز هذه، تصبح الرسالة تالفة بمجملها.
- ✓ وجود حدود لدقة الأرقام التي يمكن أن ترمز (عدد الأرقام بعد الفاصلة).
- ✓ نظراً للعوائق السابقة، قلما يستخدم الترميز الحسابي في تطبيقات الزمن الحقيقي لنقل الصوت والصورة.

## آلية العمل

- ✓ نخصص لكل رمز مجال (interval).
- ✓ يقسم كل مجال لعدة مجالات جزئية والتي يكون حجمها متناسباً مع الاحتمال الحالي للرموز.
- ✓ يشكل المجال الجزئي للرمز الوارد الذي رمز أولاً مجالاً أساسياً من أجل الرمز التالي، وهكذا.
- ✓ الخرج هو المجال المتعلق بالرمز الأخير.

## خطوات الترميز

➤ لترميز الرمز  $S$  حيث أن الرموز مرقمة من 1 إلى  $n$  و الرمز  $i$  يملك الاحتمال  $\text{Pr}[i]$

- $\text{low\_bound} = \sum_{i=1}^{s-1} \text{Pr}[i]$
  - $\text{high\_bound} = \sum_{i=1}^s \text{Pr}[i]$
- } تستخدم للحصول على احتمال حدود المجالات في أول مرحلة
- 
- $\text{range} = \text{high} - \text{low}$
  - $\text{low} = \text{low} + \text{range} * \text{low\_bound}$
  - $\text{high} = \text{low} + \text{range} * \text{high\_bound}$
- } تستخدم للحصول على احتمال المراحل التالية حتى الفواصل في الوصول إلى السلسلة المطلوبة

## مثال

قم بترميز الرسالة  $a_1a_2a_3a_3a_4$  باستخدام الترميز الحسابي. إذا علمت أن احتمال تكرار المحارف هو كالآتي:

$a_1$	$a_2$	$a_3$	$a_4$
0.2	0.2	0.4	0.2

الحل:

١- ننظر احتمال تكرار المحارف المختلفة و نرتبها

$a_3$	$a_1$	$a_2$	$a_4$
0.4	0.2	0.2	0.2

٢- ترمز السلسلة بتقسيم الفاصل  $[0,1]$  ونخصص لكل محرف فاصل يتناسب مع عدد مرات ظهوره في السلسلة.  
( حيث مجموع احتمال التكرار هو ١ ) بحيث نرتبها تصاعدياً من الأسفل إلى الأعلى.



## ➤ حيث يتم الحساب وفق العلاقة:

$$\text{low} = \text{low} + \text{range} * \text{low\_bound}$$

$$\text{range} = \text{high} - \text{low} \text{ (للمجال الموسع)}$$

• هنا المجال الموسع هو مجال a1 فيكون:

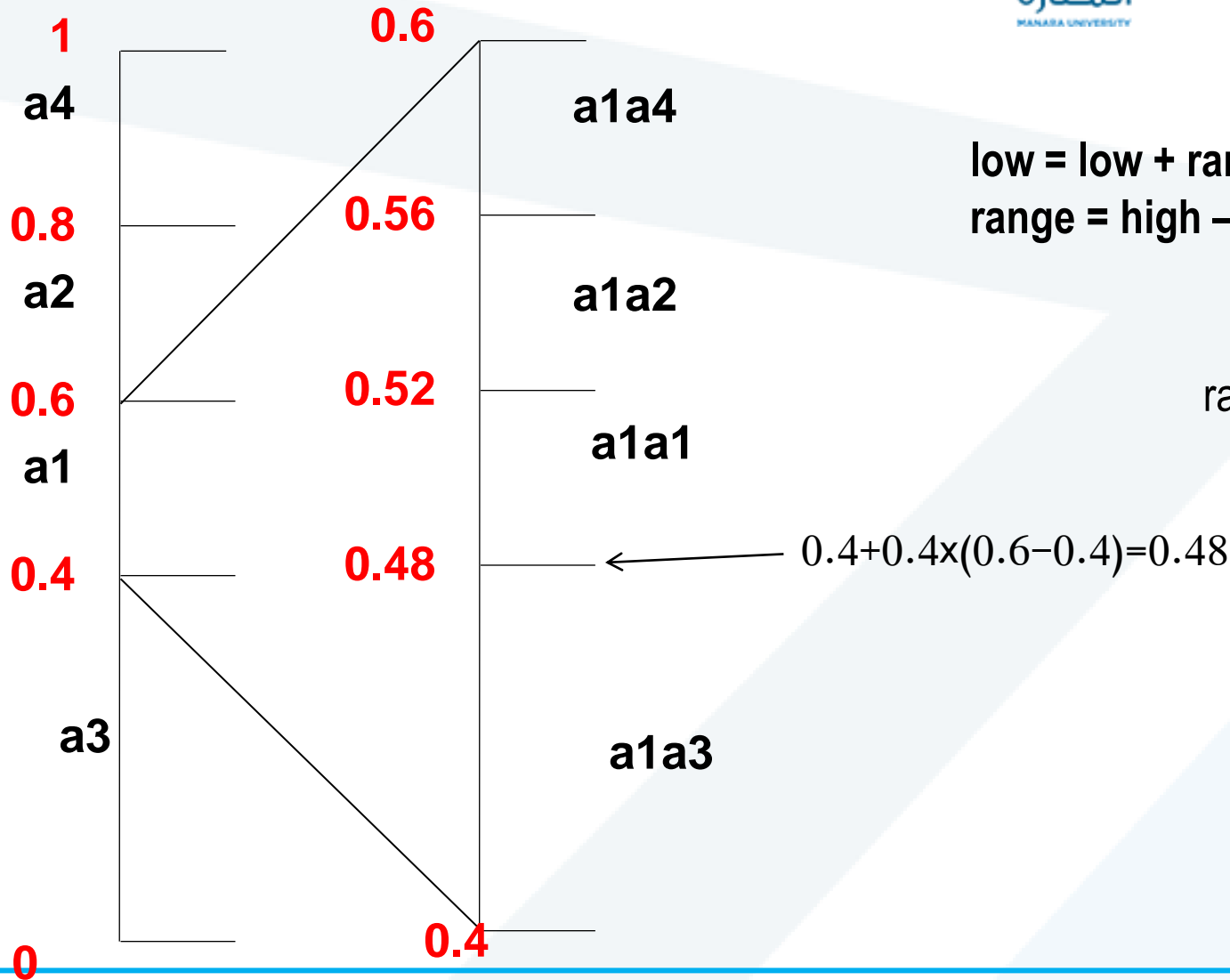
$$\text{range1} = \text{high} - \text{low} = 0.6 - 0.4 = 0.2$$

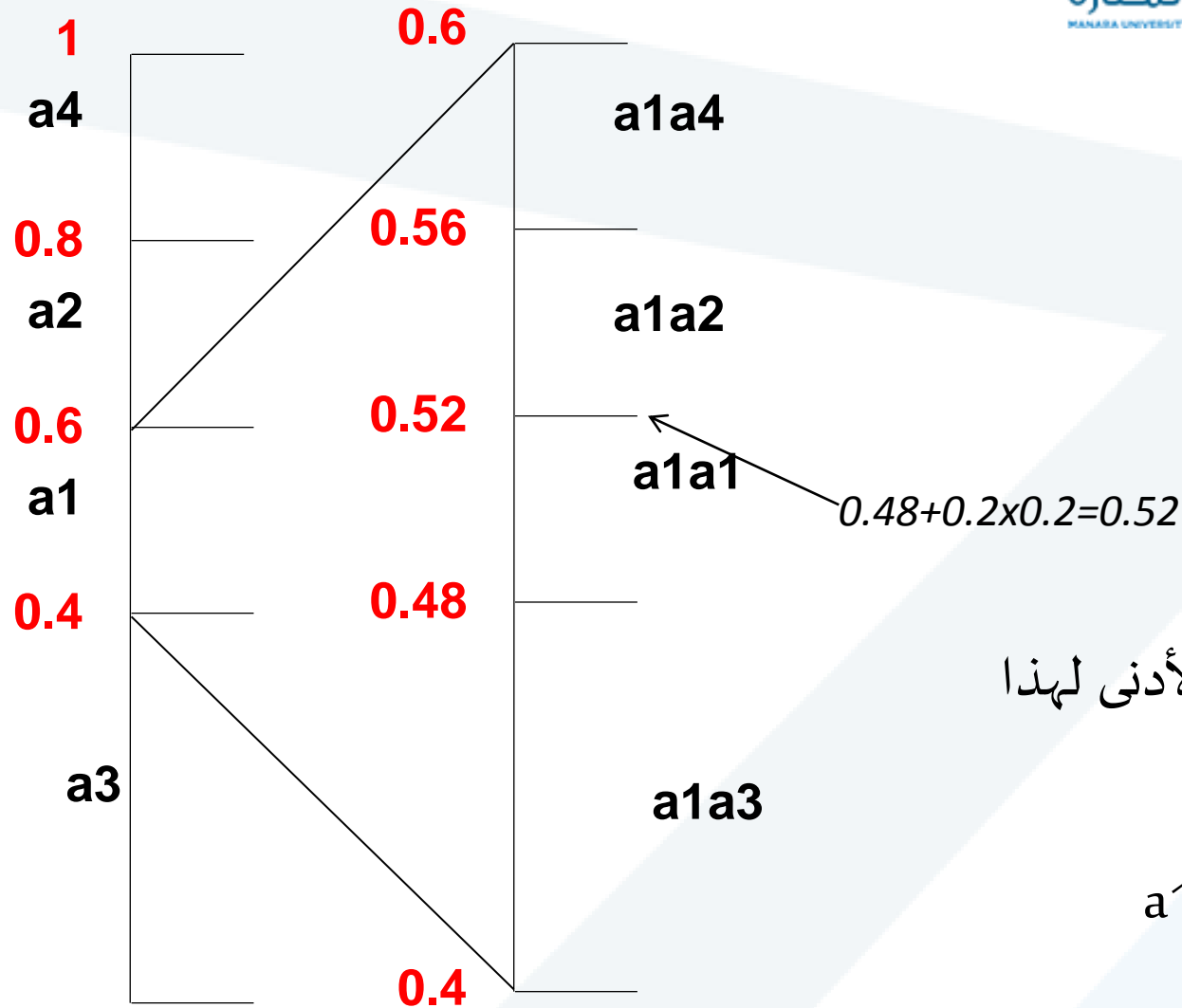
• هذه القيمة تبقى ذاتها من أجل عملية التوسيع هذه.

• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

$$\begin{aligned} \text{low} &= \text{low} + \text{range} * \text{low\_bound} \\ &= 0.4 + 0.2 * 0.4 = 0.48 \end{aligned}$$

حيث low\_bound هو المجال الأصلي للرمز هنا هو مجال a3





• هنا المجال الموسع هو ذاته لذا قيمة الـ  $range=0.2$

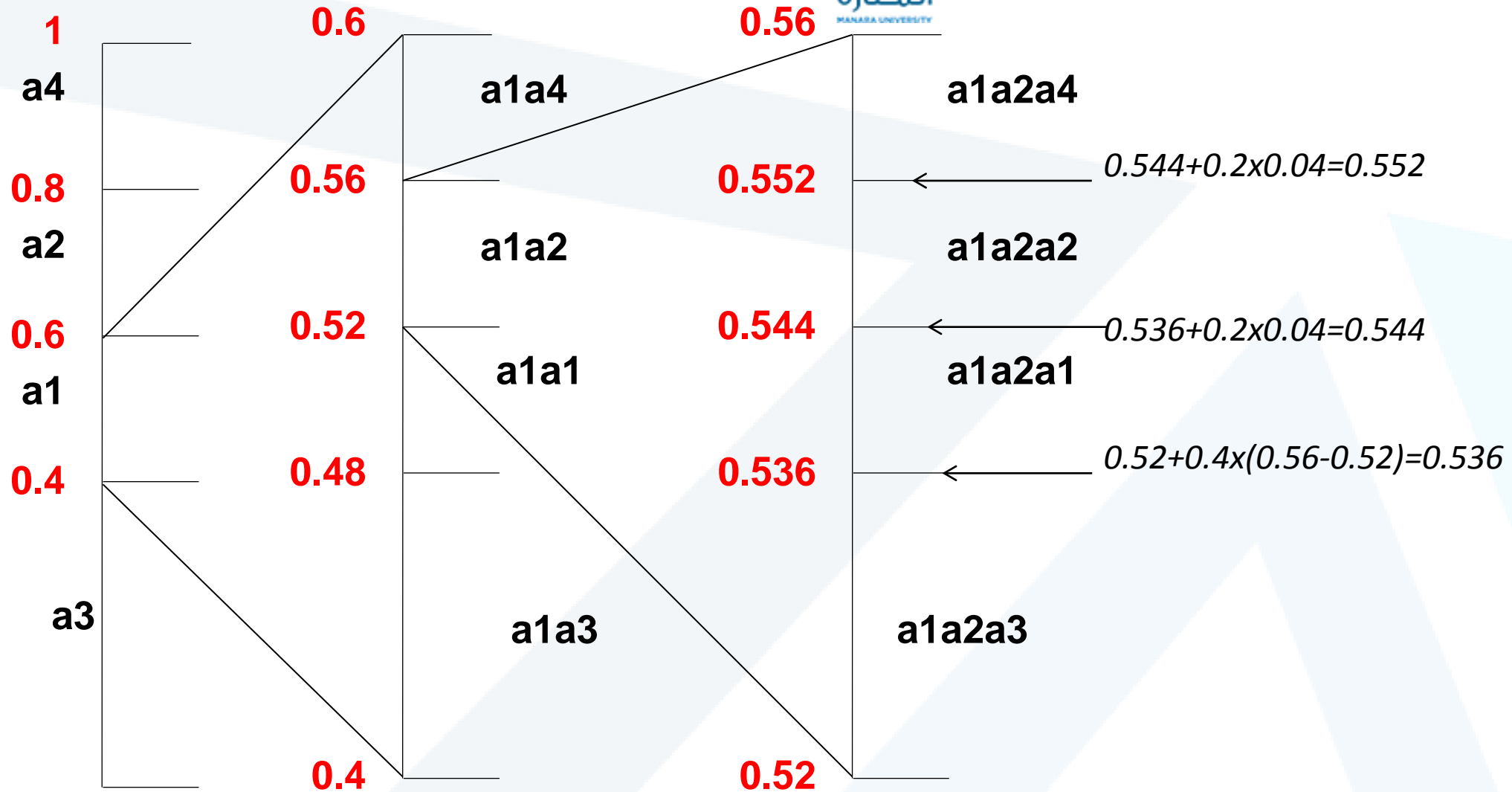
• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

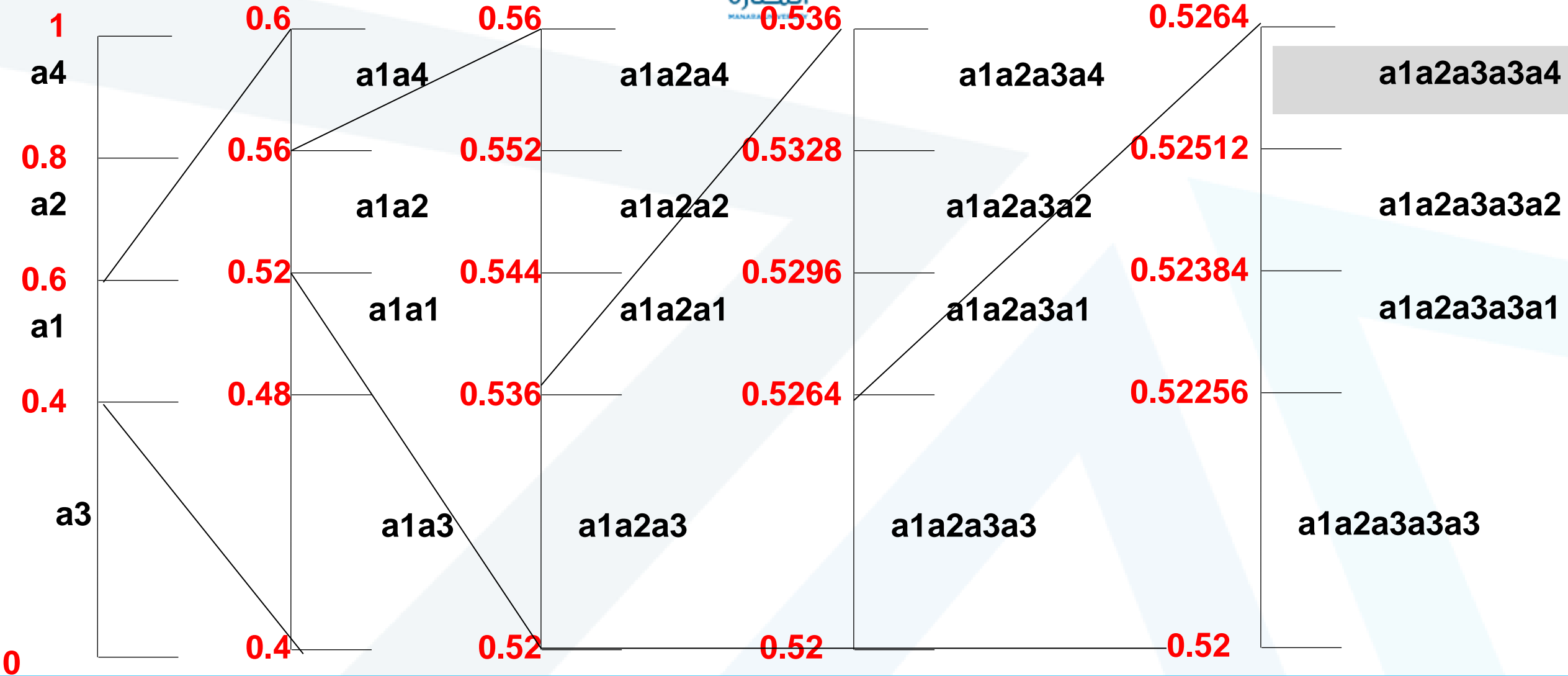
$$\begin{aligned} low &= low + range * low\_bound \\ &= 0.48 + 0.2 * 0.2 = 0.52 \end{aligned}$$

حيث  $low$  هو القيمة المحسوبة سابقاً والتي تمثل الحد الأدنى لهذا المجال

حيث  $low\_bound$  هو المجال الأصلي للرمز هنا هو مجال  $a1$

وهكذا....





## مثال

لذا يكون الترميز النهائي للرسالة هو أي رقم حقيقي ضمن المجال  $(0.52512 - 0.5264)$  الذي يتألف من ثلاث أرقام عشرية مثلاً: 0.526.

## خطوات فك الترميز

➤ يصل إلى مفكك الترميز القيمة  $V$  و طول السلسلة المرسل  $L$

➤ ويكون لدى مفكك الترميز مجالات المحارف

➤ الخطوات:

١. نبدأ من المجال  $[0,1)$

٢. نقارن قيمة  $V$  مع مجالات المحارف لتحديد المحرف الأول من السلسلة

٣. نأخذ مجال ورود المحرف الأول ونسحب المجالات ونحسب حدود المجالات الجزئية اعتماداً على القانون:

- $range = high - low$
- $low = low + range * low\_bound$
- $high = low + range * high\_bound$

٤. مع مقارنة كل قيمة يتم حسابها مع قيمة  $V$  وفي كل مرة نأخذ بالحسبان فقط المجال الذي توجد ضمنه هذه القيمة ونأخذ المحارف التي تمثل هذا المجال.

٥. تكرر الإجرائية حتى الحصول على عدد  $L$  من المحارف

## مثال

بفرض أن القيمة الواصلة إلى المستقبل هي  $V=0.62$  وأن طول السلسلة  $L=4$  ما هي السلسلة التي سيحصل عليها المستقبل باستخدام مفكك الترميز الحسابي. إذا علمت أن احتمال تكرار المحارف ومواقعها هو كالآتي:

Symbol, x	Probability, N[x]	[Q[x-1], Q[x]]
A	0.4	0.0, 0.4
B	0.3	0.4, 0.7
C	0.2	0.7, 0.9
D	0.1	0.9, 1.0

نلاحظ أن القيمة  $V=0.62$  تقع ضمن المجال  $[0.4,0.7]$  إذاً المحرف الأول هو B

• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

$$\text{low1} = \text{low} + \text{range1} * \text{low\_bound} = 0.4 + 0.3 * 0.4 = 0.52$$

$V$  خارج هذا المجال لذا نتابع الحساب

$$\text{low2} = \text{low} + \text{range1} * \text{low\_bound} = 0.52 + 0.3 * 0.3 = 0.61$$

$V$  خارج هذا المجال لذا نتابع الحساب

$$\text{low3} = \text{low} + \text{range1} * \text{low\_bound} = 0.61 + 0.3 * 0.2 = 0.67$$

$V$  داخل  $[0.61,0.67]$  ، وتكون السلسلة BC

• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

$$\text{low1} = \text{low} + \text{range2} * \text{low\_bound} = 0.61 + 0.06 * 0.4 = 0.634$$

$V$  داخل  $[0.61,0.634]$  ، وتكون السلسلة BCA



• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

$$\text{low1} = \text{low} + \text{range3} * \text{low\_bound} = 0.61 + 0.024 * 0.4 = 0.6196$$

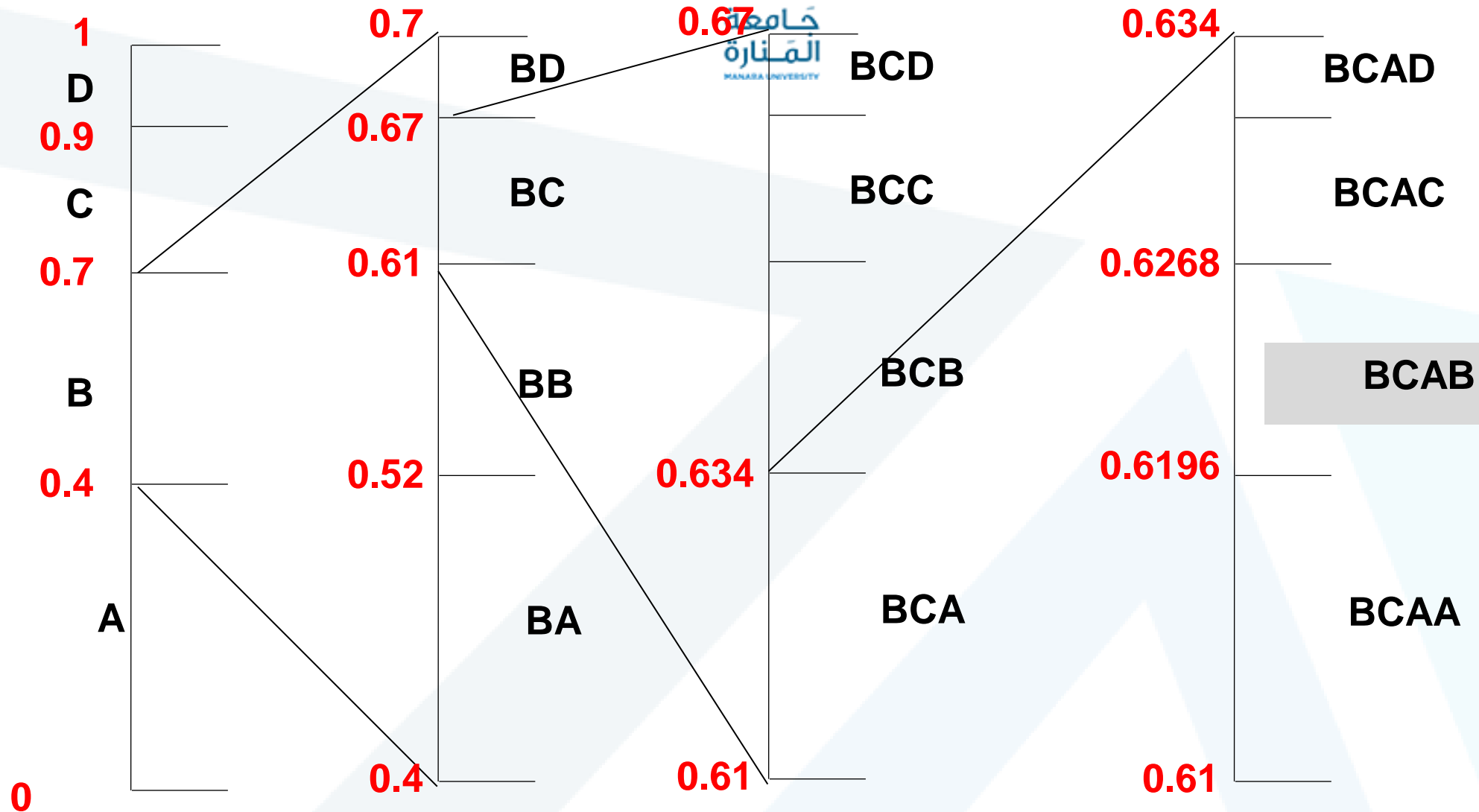
V خارج هذا المجال لذا نتابع الحساب

$$\text{low2} = \text{low} + \text{range3} * \text{low\_bound} = 0.6196 + 0.024 * 0.3 = 0.6268$$

V داخل (0.6196, 0.6268] ، وتكون السلسلة الناتجة عن فك الترميز **BCAB**



جامعة  
المنارة  
MANARA UNIVERSITY



السلسلة الناتجة عن فك الترميز هي BCAB

## مثال 2

بفرض أن القيمة الواصلة إلى المستقبل هي  $V=0.0625$  وأن طول السلسلة  $L=3$  ما هي السلسلة التي سيحصل عليها المستقبل باستخدام مفكك الترميز الحسابي. إذا علمت أن احتمال تكرار المحارف ومواقعها هو كالاتي:

Symbol	Probability
a	0.25
b	0.5
c	0.25

باستخدام نفس طريقة الحل السابقة يكون:

## مثال 2

نلاحظ أن القيمة  $V=0.0625$  تقع ضمن المجال  $[0,0.5]$  إذاً المحرف الأول هو  $b$

• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

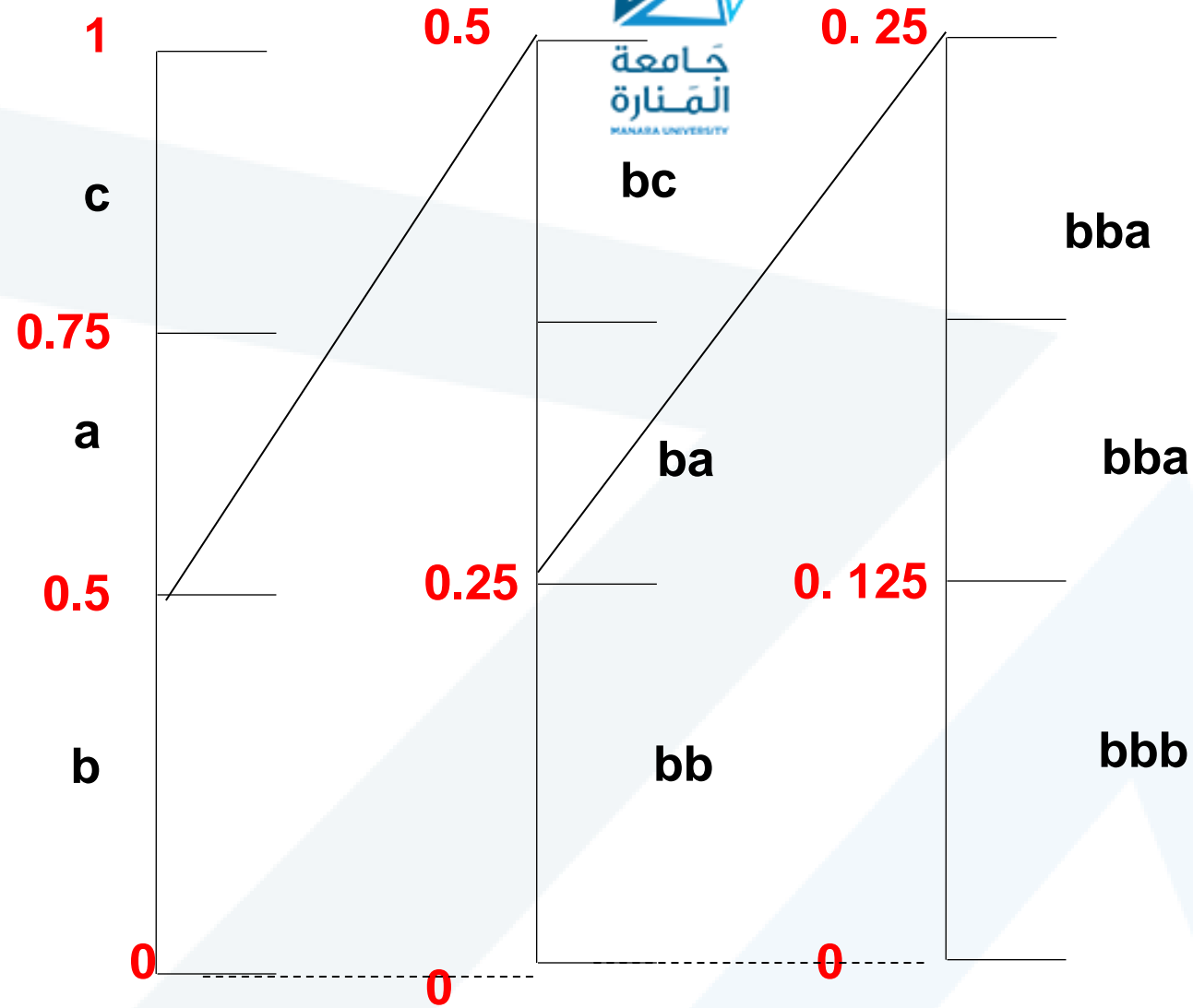
$$\text{low1} = \text{low} + \text{range1} * \text{low\_bound} = 0 + 0.5 * 0.5 = 0.25$$

$V$  داخل  $[0,0.25]$  ، وتكون السلسلة  $bb$

• فيكون الحد الأدنى للمجال الجديد يحسب كالآتي:

$$\text{low1} = \text{low} + \text{range1} * \text{low\_bound} = 0 + 0.25 * 0.5 = 0.125$$

$V$  داخل  $[0,0.125]$  ، وتكون السلسلة الناتجة عن فك الترميز هو  $bbb$



السلسلة الناتجة عن فك الترميز هي bbb