

# Exploiting Statistical Dependence for Compression

---

- Joint entropy
- Statistical dependence among color components
- Conditional entropy and bit-rate for lossless coding of sources with memory
- Image conditional entropy measurements
- Cross entropy
- Run-length coding
- Facsimile compression standards



# Joint entropy

- Consider random vectors (with finite-alphabet components)

$$\mathbf{X} = (X_0, X_1, \dots, X_{m-1})$$

- Entropy

$$H(\mathbf{X}) \triangleq E[-\log_2 f_{\mathbf{X}}(\mathbf{X})] = E[h_{\mathbf{X}}(\mathbf{X})]$$

- Long-hand  $H(\mathbf{X}) \equiv H(X_0, X_1, \dots, X_{m-1})$

$$= -\sum_{x_0} \sum_{x_1} \dots \sum_{x_{m-1}} f_{\mathbf{X}}(x_0, x_1, \dots, x_{m-1}) \log_2 f_{\mathbf{X}}(x_0, x_1, \dots, x_{m-1})$$

- Shannon's Noiseless Source Coding Theorem: Consider a “vector source” generating i.i.d. random vectors  $\mathbf{X}$ . Joint entropy  $H(\mathbf{X})$  is achievable lower bound for bit-rate for encoding  $\mathbf{X}$ .



# Joint entropy and statistical dependence

- Theorem

$$H(X_0, X_1, X_2, \dots, X_{m-1}) \leq H(X_0) + H(X_1) + \dots + H(X_{m-1})$$

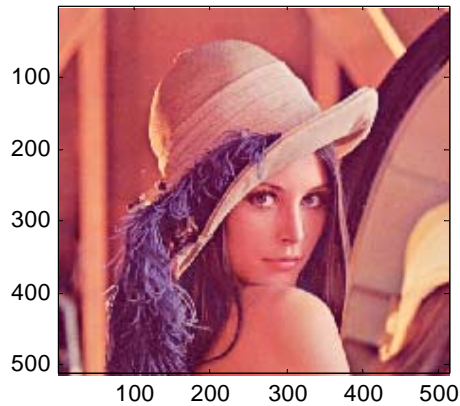
Equality for statistical independence of  $X_0, X_1, X_2, \dots, X_{m-1}$

- Exploiting statistical dependence lowers bit-rate
- Statistically independent components can be coded separately without loss in efficiency

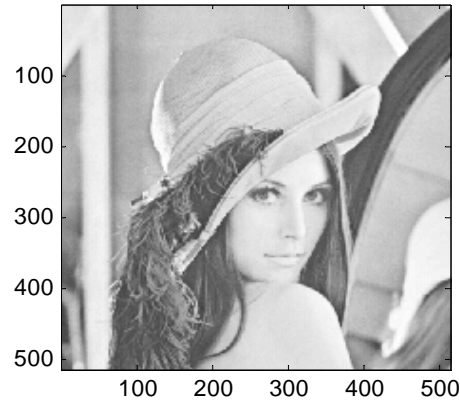


# Color components

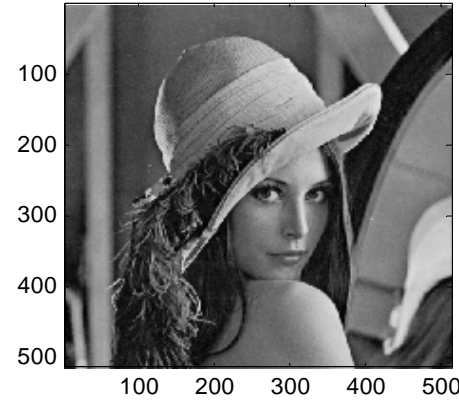
Standard Image 'Lena'



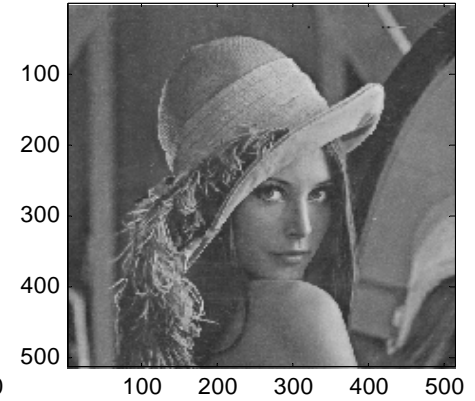
Red R



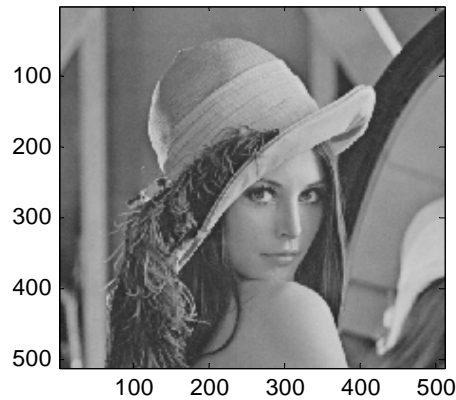
Green G



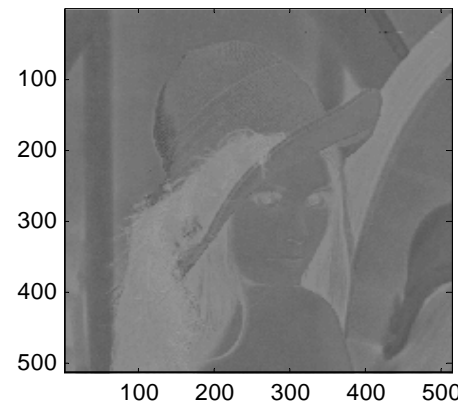
Blue B



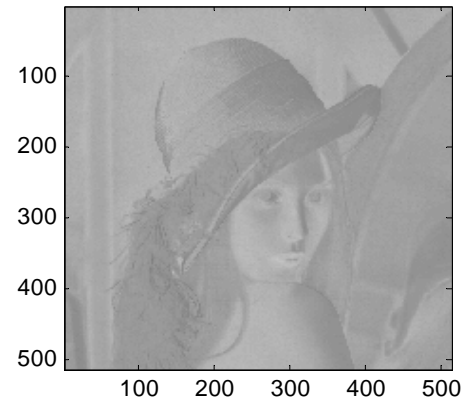
Luminance Y



Chrominance Cb



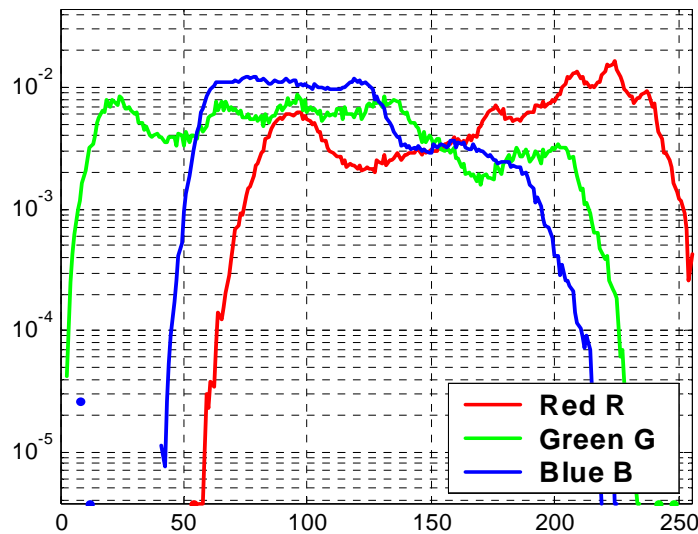
Chrominance Cr



# First order statistics of color components

- Histograms (relative number of occurrences) used in lieu of PMFs

Standard Lena Image - PMF of RGB Pixel Values

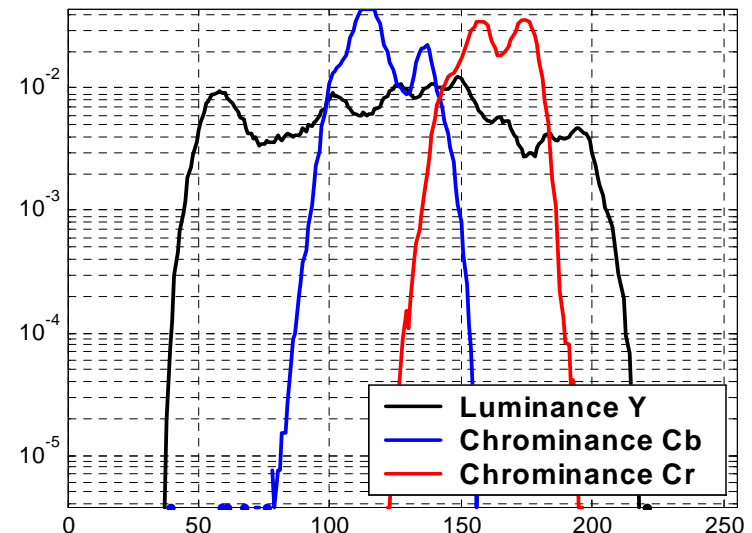


$$H(R)=7.25 \text{ bpp}$$

$$H(G)=7.59 \text{ bpp}$$

$$H(B)=6.97 \text{ bpp}$$

Standard Lena Image - PMF of YCbCr Pixel Values



$$H(Y)=7.23 \text{ bpp}$$

$$H(Cb)=5.47 \text{ bpp}$$

$$H(Cr)=5.42 \text{ bpp}$$

- Image 'Lena', 512 x 512 pixels, 8 bits per color components



# Statistical dependence among color components

- Image: 'Lena', 512 x 512 pixels, 8 bits per color component

<i>Fixed length</i>	3 x 8 = 24 bpp
$H(Y,Cb,Cr)$	15.01 bpp
$H(Y)+H(Cb)+H(Cr)$	18.12 bpp
$\Delta H$	3.11 bpp

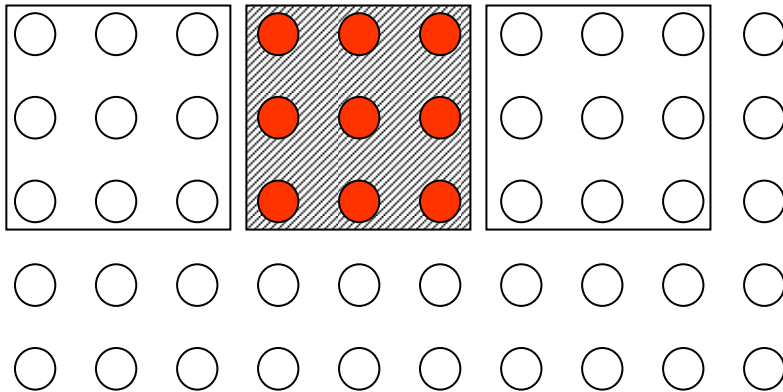
<i>Fixed length</i>	3 x 8 = 24 bpp
$H(R,G,B)$	16.84 bpp
$H(R)+H(G)+H(B)$	21.82 bpp
$\Delta H$	4.98 bpp

- Statistical dependence among R, G, B is stronger.
- If joint sources Y, Cb, Cr or R,G,B are treated as sources with memory, the possible gain by joint coding is much smaller.

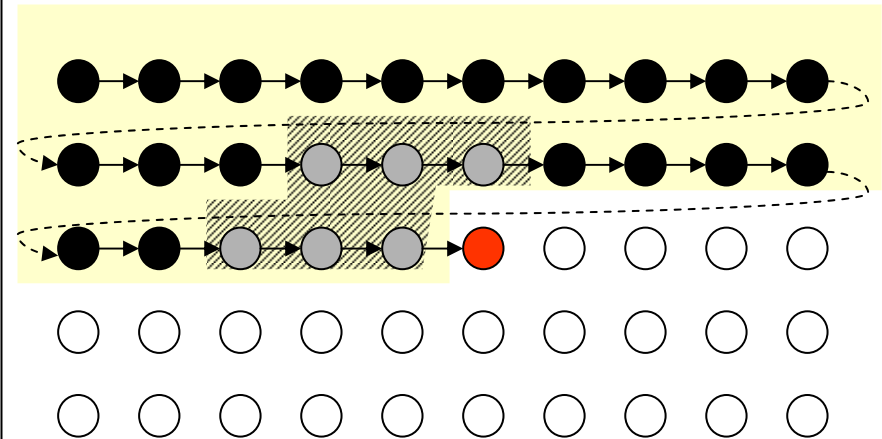


# Exploiting Statistical Dependence Among Neighboring Pixels

- Break image into blocks
- Interpret each block as vector
- Block-by-block entropy coding
- Statistical dependencies across block boundaries neglected



- Sliding window across image
- Interpret image as source with memory
- Perform conditional entropy coding



# Markov random processes

- Consider discrete random process  $\{X_n\}$  with memory
- Stationarity

$$f_{\mathbf{X}_{i:i+m}} = f_{\mathbf{X}_{0:m}} \quad \text{for all } i, m \in \mathbb{Z}, m \geq 0$$

$(m+1)$ -dimensional random vector,  
Components  $0 \dots m$  of infinite-  
dimensional random vector

- Random process with finite memory (“Markov- $p$  process”)

$$f_{X_m | \mathbf{X}_{0:(m-1)}} = f_{X_m | \mathbf{X}_{(m-p):(m-1)}}$$

- Order- $p$  conditional PMF describes stationary Markov process fully



# Conditional entropy

- Consider two finite-alphabet r.v.  $X$  and  $Y$

$$\begin{aligned} H(X|Y) &\triangleq E[-\log_2 f_{X|Y}(x, y)] \equiv -\sum_y \sum_x f_{X,Y}(x, y) \log_2 f_{X|Y}(x, y) \\ &= -\sum_y f_Y(y) \sum_x f_{X|Y}(x, y) \log_2 f_{X|Y}(x, y) \end{aligned}$$

- Conditional entropy  $H(X|Y)$  is average additional information, if  $Y$  is already known

$$\begin{aligned} H(X, Y) &= E[-\log_2 f_{X,Y}(X, Y)] \\ &= E[-\log_2 (f_Y(Y) f_{X|Y}(X, Y))] \\ &= E[-\log_2 f_Y(Y)] + E[-\log_2 f_{X|Y}(X, Y)] \\ &= H(Y) + H(X|Y) \end{aligned}$$



# Conditional entropy (cont.)

- Independent random variables  $X$  and  $Y$

$$\begin{aligned} H(X | Y) &= E \left[ -\log_2 f_{X|Y}(X, Y) \right] \\ &= E \left[ -\log_2 f_X(X) \right] = H(X) \end{aligned}$$

- Let  $\mathbf{Y}'$  be a random vector containing any subset of elements of another random vector  $\mathbf{Y}$

$$H(X | \mathbf{Y}) \leq H(X | \mathbf{Y}')$$

Equality, iff  $X$  is conditionally independent from elements missing from  $Y$ .

Hence, extra prior information can only reduce uncertainty.



# Bit-rate bound for Markov random process

- Consider discrete, stationary Markov- $p$  random process, characterized by conditional PMF  $f_{X_p | \mathbf{X}_{0:(p-1)}}$
- Shannon's Noiseless Source Coding Theorem: achievable lower bound for bit-rate

$$R \geq H(X_p | \mathbf{X}_{0:(p-1)})$$


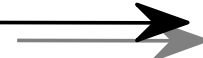
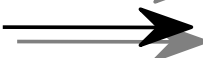
- Important to exploit statistical dependence

$$H(X_p | \mathbf{X}_{0:(p-1)}) \leq H(X_p)$$



# Switched variable length coding

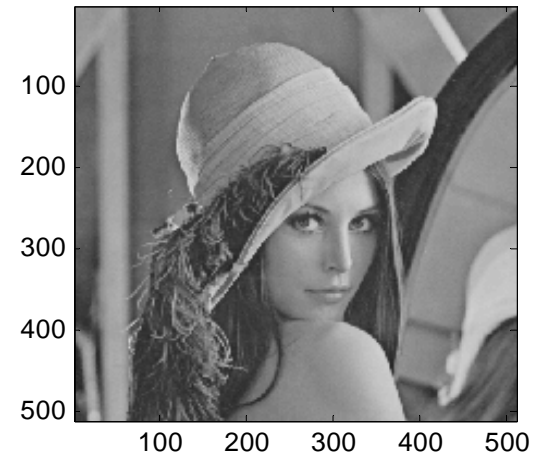
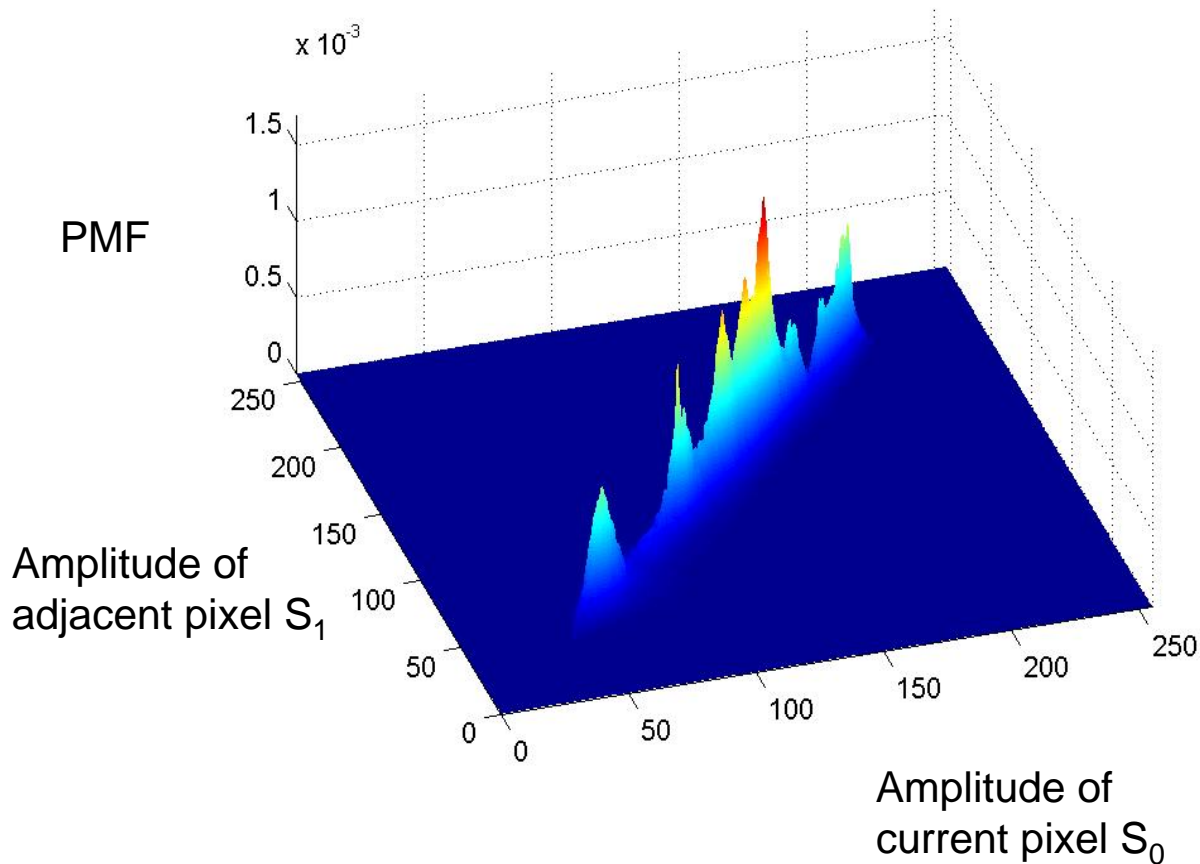
- Average code word length can approach  $H(X_p | \mathbf{X}_{0:(p-1)})$ , e.g., with a switched Huffman code
- Number of states for 8-bit pixels

$p = 1$		256 states
$p = 2$		65,536 states
$p = 3$		16,777,216 states



# Statistical Dependence of Adjacent Pixels

Histogram of two horizontally adjacent pixels  
( 'Lena', 512 x 512 pixels, 8 bpp)



# Calculating Entropies for an Individual Image

- Assume that a particular image has its own entropy code – implies some overhead to send code book
- Obtain probabilities from the image histogram
- Joint entropy for two adjacent pixels (bound for joint coding of 2x1 blocks)

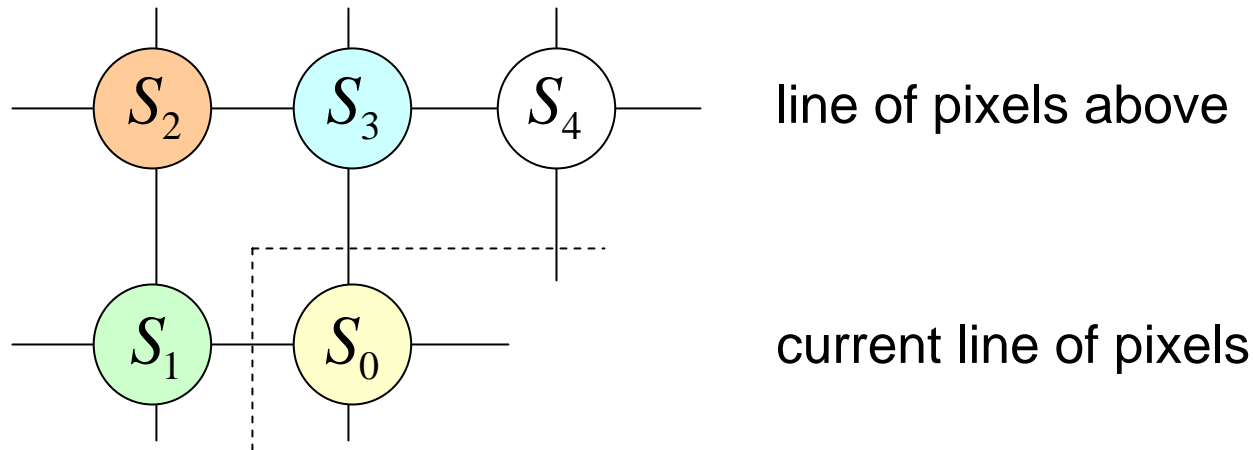
$$H(S_0, S_1) = - \sum_{s_0} \sum_{s_1} f_{S_0, S_1}(s_0, s_1) \log_2 f_{S_0, S_1}(s_0, s_1)$$

- First-order conditional entropy (bound for conditional coding using previous pixel)

$$H(S_0 | S_1) = - \sum_{s_0} \sum_{s_1} f_{S_0, S_1}(s_0, s_1) \log_2 f_{S_0 | S_1}(s_0 | s_1)$$



# Conditional entropy for an image



component	$H(S_0)$	$H(S_0   S_1)$	$H(S_0   S_3)$	$H(S_0   S_2)$
Y	7.23	4.67	4.32	4.86
Cb	5.47	3.80	3.58	3.85
Cr	5.42	3.69	3.55	3.82

Image: 'Lena', 512 x 512 pixels, 8 bpp

All values in bpp



# Cross Entropy

- Assume that an entropy code is used which is optimal for probabilities  $g_{s_0, s_1}(s_0, s_1)$  rather than the “true” probabilities
- Obtain “true” image probabilities  $f_{s_0, s_1}(s_0, s_1) \neq g_{s_0, s_1}(s_0, s_1)$  from histogram
- Joint cross entropy for two adjacent pixels (bound for joint coding of 2x1 blocks)

$$H_{cross}(f_{s_0, s_1} \parallel g_{s_0, s_1}) = - \sum_{s_0} \sum_{s_1} f_{s_0, s_1}(s_0, s_1) \log_2 g_{s_0, s_1}(s_0, s_1)$$

- First-order conditional cross entropy (bound for conditional coding using previous pixel)

$$H_{cross}(f_{s_0|s_1} \parallel g_{s_0|s_1}) = - \sum_{s_0} \sum_{s_1} f_{s_0, s_1}(s_0, s_1) \log_2 g_{s_0|s_1}(s_0, s_1)$$

- Kullback-Leibler distance (increase due to not using “true” probabilities)

$$D_{KL}(f \parallel g) = H_{cross}(f \parallel g) - H_{cross}(f \parallel f) \geq 0$$



# Run-length coding

- For sources that emit “runs” of identical symbols
- Replace a sequence  $\{x_n\}$  by a shorter sequence of symbol pairs  $\{a_k, r_k\}$  such that

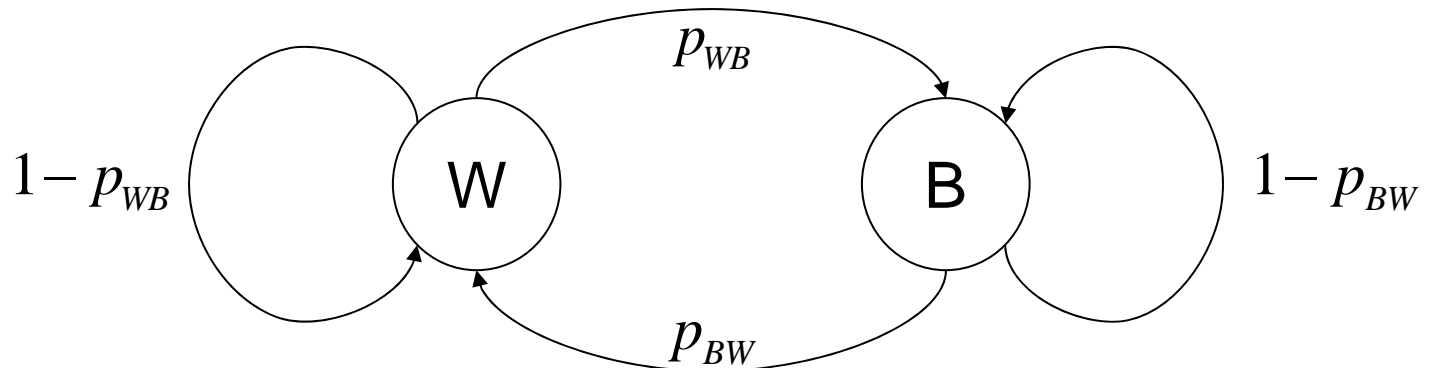
$$x_n = a_k \quad \text{for all } n : \sum_{j=1}^{k-1} r_j < n \leq \sum_{j=1}^k r_j$$

- Entropy coding (e.g., Huffman coding) of new symbol pairs  $\{a_k, r_k\}$
- For binary images,  $a_k$  can be omitted.



# Statistical model for binary images of line-art

- Markov-1 model [*Capon, 1959*]



- State probabilities

$$\Pr\{W\} = 1 - \Pr\{B\} = \frac{p_{BW}}{p_{BW} + p_{WB}}$$

- Run-length distributions

$$\Pr\{k \text{ successive white pixels}\} = (1 - p_{WB})^{k-1} p_{WB} \quad \text{for } k = 1, 2, 3, \dots$$

$$\Pr\{k \text{ successive black pixels}\} = (1 - p_{BW})^{k-1} p_{BW} \quad \text{for } k = 1, 2, 3, \dots$$

- Golomb code for geometric distribution



# Measured parameters for Capon model

Document	Weather Map	Printed Text
$\Pr\{W\}$	0.887	0.935
$\Pr\{B\}$	0.113	0.065
$p_{WB}$	0.027	0.024
$p_{BW}$	0.214	0.347
$H(X_i   X_{i-1})$	0.241 bpp	0.215 bpp

[Kunt, 1974]



# Facsimile compression standards

- Standards by the ITU-T (formerly CCITT)
  - T.4 (Group 3)
    - used by all fax machines over PSTN (public switched telephone network)
    - 1-d modified Huffman code (MH) or 2-d MMR code optional
  - T.6 (Group 4): fax over digital networks (e.g., ISDN)
    - Always 2-d MMR code
    - Less error-resilient than Group 3
- Picture formats
  - Horizontal resolution: 1728 pixels/line (1664 active) → 8.05 pixel/mm
  - Vertical resolution
    - Standard mode: 3.85 lines/mm (978 lines/page)
    - Fine mode: 7.7 lines/mm (1956 lines/page)
    - Very-fine mode: 15.4 lines/mm (3912 lines/pages)



# Group 3 fax: modified Huffman code

- Lengths of runs of white pixels and black pixels encoded within a scan line
- Each run represented as

$$\text{white runs: } r_w = 64 \times r_{w/\text{make-up}} + r_{w/\text{term}} \quad \text{black runs: } r_b = 64 \times r_{b/\text{make-up}} + r_{b/\text{term}}$$

- “Make-up” and “termination” run-lengths encoded independently
- Two separate Huffman code tables for white and black runs (“make-up” and “termination” codes) based on the statistics of 8 representative documents
- Shortest code words (2 bits) for black runs of length 2 and 3
- Shortest code words (4 bits) for white runs of lengths 2 . . . 7
- Same code tables also used as part of Group 4 fax standard (MMR)
- Special EOL codeword for each line, 6x EOL as end of page



# 2-d fax coding

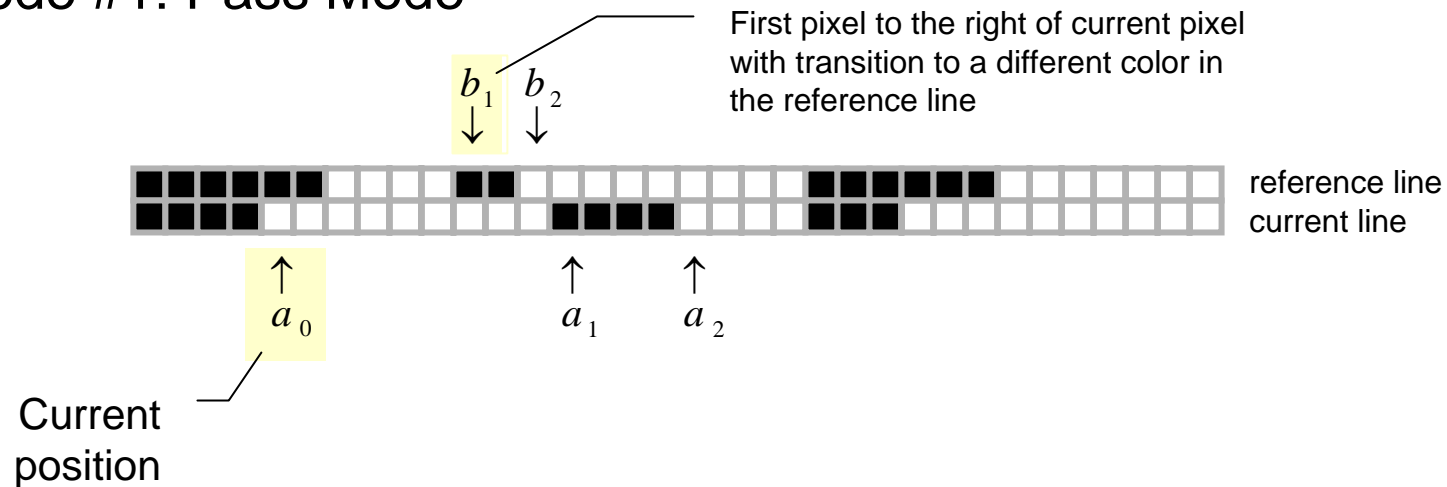
---

- MMR – Modified modified READ (relative address designate)
- Optional enhancement for Group 3, mandatory for Group 4
- For error resilience, Group 3 standard requires that at most  $K-1$  lines are encoded 2-d, with  $K=2$  for standard resolution and  $K=4$  for fine resolution.



# Modified modified READ algorithm

- Encode black and white run lengths relative to reference line above
- Mode #1: Pass Mode



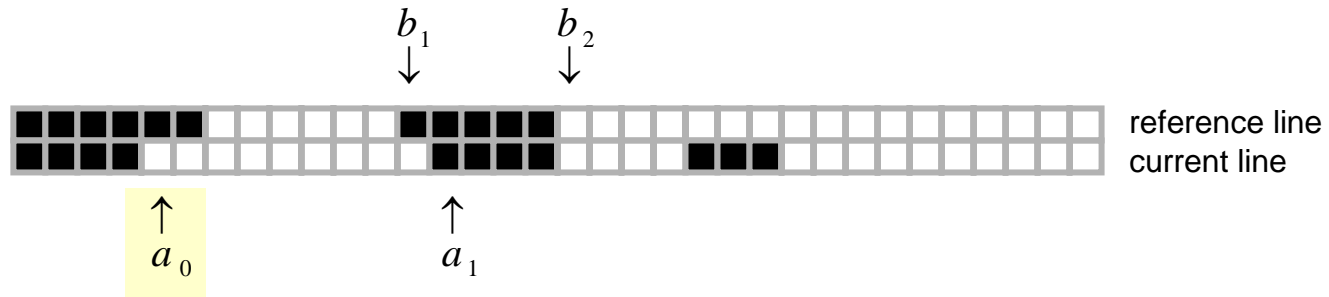
If  $b_2 < a_1$ : issue pass code word 0001

Update pointers:  $a_0 \leftarrow b_2$ , find new  $b_1, b_2$



# Modified modified READ algorithm (cont.)

- Mode #2: Vertical Mode



If  $|a_1 - b_1| \leq 3$ : encode difference  $a_1 - b_1$  using table

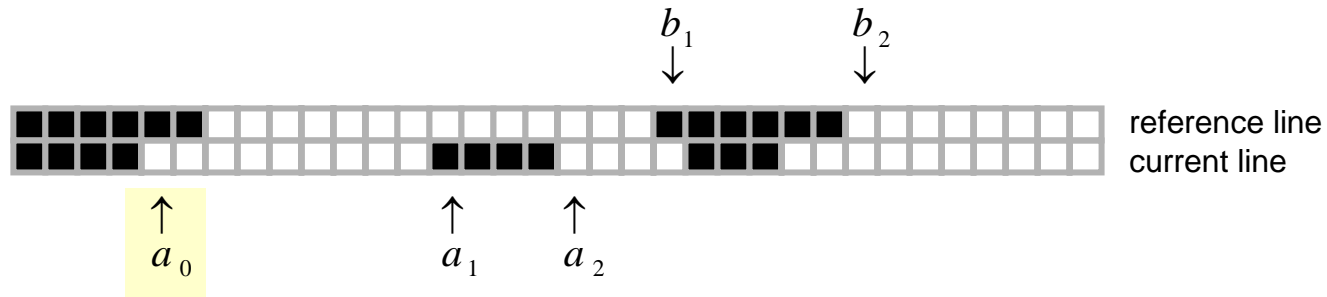
$a_1 - b_1$	-3	-2	-1	0	1	2	3
	0000011	000011	011	1	010	000010	0000010

Update pointers:  $a_0 \leftarrow a_1$ ,  $b_1 \leftarrow b_2$  etc.



# Modified modified READ algorithm (cont.)

- Mode #3: Horizontal Mode



If neither Pass Mode nor Vertical Mode:

encode next two runs 1-d, using Group 3 modified Huffman code

Update pointers:  $a_0 \leftarrow a_2$ , find new  $a_1, a_2, b_1, b_2$



# Compression efficiency of fax standards

CCITT Document Number	Standard resolution (2,052,864 pixels)		Fine resolution (4,105,728 pixels)	
	1-d MHC	2-d READ	1-d MHC	2-d READ
1	133,095	93,196	266,283	141,826
2	123,930	53,366	247,443	80,550
3	244,028	138,411	487,485	229,375
4	436,450	366,055	871,983	553,942
5	253,509	162,186	506,283	257,548
6	191,347	78,577	381,905	132,509
7	428,028	357,130	855,841	539,152
8	238,221	89,654	476,624	137,560
Average	256,076	167,322	511,731	259,058

[Yasuda, 1980]



# Reading

---

- Taubman + Marcellin, Chapter 2.3 – 2.5
- H. Meyr, H. G. Rosdolsky, T.S. Huang, Optimum Run Length Codes, IEEE Trans. Communications, vol. COM-22, no. 6, pp. 1425-1433, June. 1974.
- H. G. Musmann, D. Preuss, Comparison of Redundancy Reducing Codes for Facsimile Transmission of Documents, IEEE Trans. Communications, vol. COM-25, no. 11, pp. 1425-1433, Nov. 1977.

