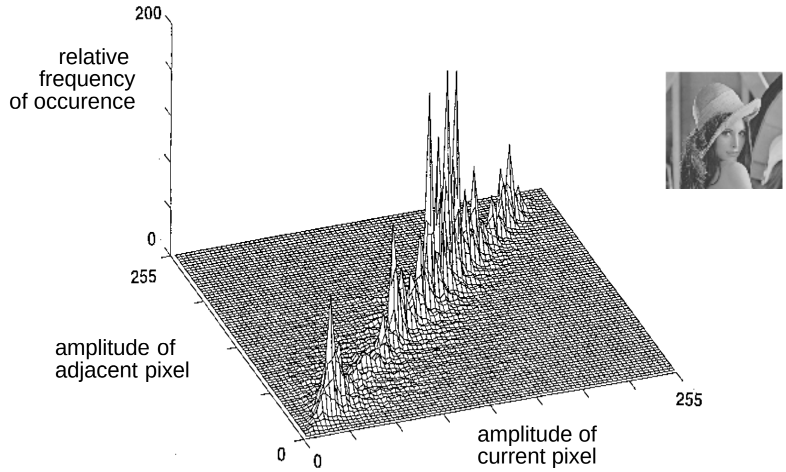
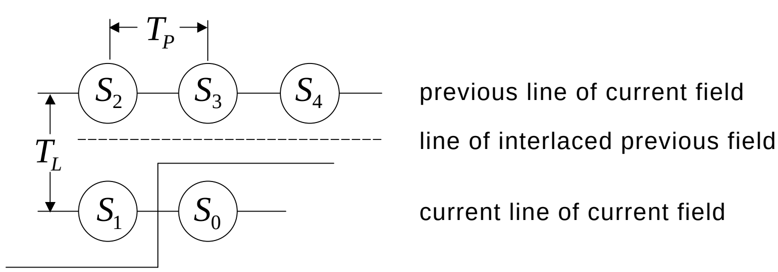
Second order statistics for luminance signal Y



Histogram of two horizontally adjacent pels
(picture: female head-and-shoulder view)

Conditional entropy of the video signal



component	T_P [ns]	$H(S_0)$	$H(S_0 \mid S_1)$	$H(S_0 \mid S_3)$
Y	100	7.34	4.66	4.85
R-Y	500	5.57	3.76	2.96
B-Y	500	5.24	3.75	2.93

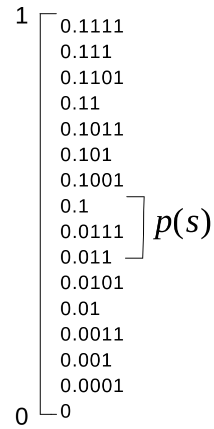
Averages in bit/sample. 3 EBU test slides, 3 SMPTE test slides, uniform quantization to 256 levels.

Arithmetic coding I

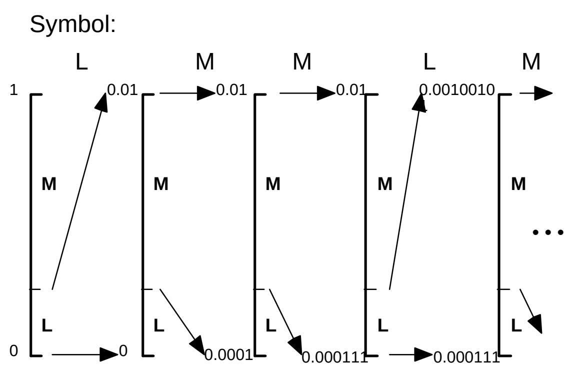
- Universal entropy coding algorithm for strings
- Representation of a string by a subinterval of the unit interval [0,1)
- Width of the subinterval is approximately equal to the probability of the string $p(s)$
- Interval of width $p(s)$ is guaranteed to contain one number that can be represented by *b* binary digits, with

$$-\log(p(s))+1 \leq b < -\log(p(s))+2$$

- Each interval can be represented by a number which needs 1 to 2 bits more than the ideal code word length



Arithmetic Coding II



$p(M) = (0.11)_b$ $p(L) = (0.01)_b$

- Requires multiplications and additions with (potentially) very long wordlength.
- Universal coding: probabilities can be changed on the fly.
- Decoder: compare to subinterval boundaries bit by bit.

Arithmetic Coding III

Multiplication-free algorithm

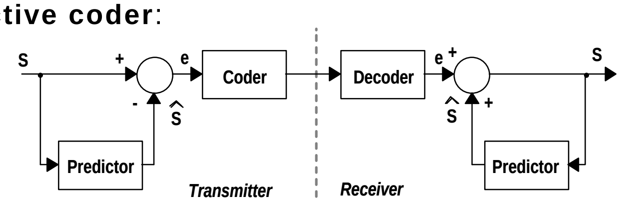
- Normalize interval width to $0.75 \leq A < 1.5$ by binary shift after each symbol
- Approximation $p(L) \leq p(M)$:

$$Ap(L) \approx p(L)$$
$$Ap(M) = A(1 - p(L)) \approx A - p(L)$$

- Multiplication-free algorithm for non-binary alphabets (Rissanen & Mohiuddin, 1989)

Redundancy reduction by prediction

- Much simpler than a conditional Huffman coder is a **predictive coder**:



- Principle of the predictive coder:
  - S_0 is the current sample of the original signal S.
  - $\hat{S}_0$ is a prediction for S_0 calculated from previous samples $S_1, S_2, ..., S_N$.
  - e is the prediction error, with greatly reduced statistical dependencies between adjacent samples.
  - The receiver can reconstruct S without loss.

Entropy and variance of the prediction error

- Approximation of the entropy of the prediction error *e*

standard deviation of *e*; constant that depends on the shape of the pdf

$$H(e) \approx \log \frac{s_e}{\Delta} + C \quad \text{for} \quad \Delta << s_e$$

quantization step size

- Shape constant *C:*

  Gaussian pdf: $C = 2.047$ bit Laplacian pdf: $C = 1.943$ bit

- With linear prediction of video signals the prediction error pdf is typically Laplacian.
- Minimization of prediction error variance vs. entropy typically leads to very similar results.

Statistical optimization of a nonlinear predictor

- Prediction obtained from previous samples

$$\underline{S} = (S_1, S_2, ..., S_N)$$

- Variational problem

$$E\{[S_0 - f(\underline{S})]^2\} \longrightarrow \min$$

- Solution:

$$\hat{S} = f(\underline{S}) = E\{S_0 \mid \underline{S}\} = \sum_{S_0} S_0 P(S_0 \mid \underline{S})$$

- Solution can be stored in table of size 2^{8N} for 8-bit representation of $S_1, S_2, ..., S_N$.

Statistical optimization of a linear predictor I

- Assume zero-mean: $E\{S\}=0$
- Linear predictor:

$$\hat{S}_0 = a_1 S_1 + a_2 S_2 + \ldots + a_N S_N$$

- Variance of the prediction error:

$$s_e^2 = \sum_{i=0}^{N}\sum_{j=0}^{N} a_i a_j R_{ij} \quad \text{with } a_0 = -1$$

- Autocorrelation matrix can be measured for a given signal.

$$R_{ij} = E\{S_i S_j\}$$

Statistical optimization of a linear predictor II

- Minimization of the prediction error variance with

$$\frac{\partial s_e^2}{\partial a_i} = 0 \quad \text{for all } i = 1,\ldots,N$$

  Leads to "orthogonality principle":

$$E\{eS_i\} = 0 \quad \text{for all } i = 1,\ldots,N$$

- Optimum prediction coefficients $a_1, a_2, \ldots, a_N$ obtained by inverting *N*x*N* autocorrelation matrix R_{ij}
- Orthogonality principle implies decorrelation of the prediction error:

$$E\{e_0 e_i\} = E\{e_0\}E\{e_i\} = 0 \quad \text{for } i = 1,2,\ldots,N$$

Statistical optimization of a linear predictor III

- If the signal is not zero-mean, i.e. $E\{S\} \neq 0$

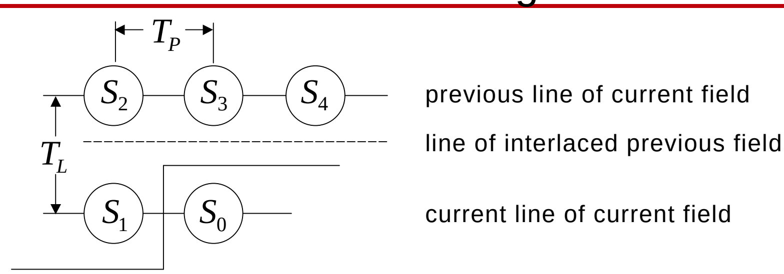
$$\hat{S}_0 = a_{DC} + a_1 S_1 + ... + a_N S_N \text{ with } a_{DC} = E\{S\}(1 - \sum_{i=1}^{N} a_i)$$

- For **Gaussian random processes** decorrelation implies statistical independence.
- Prediction gain:

$$G = \frac{1}{2}\log_2\left(\frac{\sigma_s^2}{\sigma_e^2}\right) \text{ bit } = \frac{1}{2}\log_2\left(\frac{E\{S^2\} - E^2\{S\}}{\sigma_e^2}\right) \text{ bit}$$

- Rule of thumb: 6 dB = 1 bit

Statistically optimized predictors for the luminance signal Y



<u>Constraint:</u> 3 bit wordlength of the prediction coefficients

$H(S_0)$ [bit]	Predictor			$H(e)$ [bit]	Criterion
	a_1	a_2	a_3		
7.34	7/8	-5/8	3/4	4.30	minimum variance
734	7/8	-1/2	5/8	4.29	minimum entropy

Average of 3 EBU test slides, 3 SMPTE test slides, uniform quantization to 256 levels. T_P = 100 ns.

Statistically optimized predictors for the color difference signals R-Y and B-Y

- Signal R-Y: sampling interval T_P = 500 ns

$H(S_0)$ [bit]	Predictor a_1	Predictor a_2	Predictor a_3	$H(e)$ [bit]	Criterion
5.57	5/8	-1/2	7/8	2.87	minimum variance
5.57	3/8	-1/4	7/8	2.82	minimum entropy

- Signal B-Y: sampling interval T_P = 500 ns

$H(S_0)$ [bit]	Predictor a_1	Predictor a_2	Predictor a_3	$H(e)$ [bit]	Criterion
5.24	3/8	-1/4	7/8	2.46	minimum variance
5.24	3/8	-1/4	7/8	2.46	minimum entropy

Average of 3 EBU test slides, 3 SMPTE test slides,
uniform quantization to 256 levels.

Optimizing the predictor in the frequency domain

- Power spectrum of the prediction error

$$\Phi_{ee}(w_x, w_y) = \Phi_{ss}(w_x, w_y)\,|1 - P(w_x, w_y)|^2$$

power spectrum of the video signal *S* — prediction transfer function

- Mean squared error criterion

$$s_e^2 = \frac{1}{4p^2} \iint \Phi_{ss}(w_x, w_y)\,|1 - P(w_x, w_y)|^2\, dw_x\, dw_y \longrightarrow \min$$

Constraint: causality of $P(w_x, w_y)$

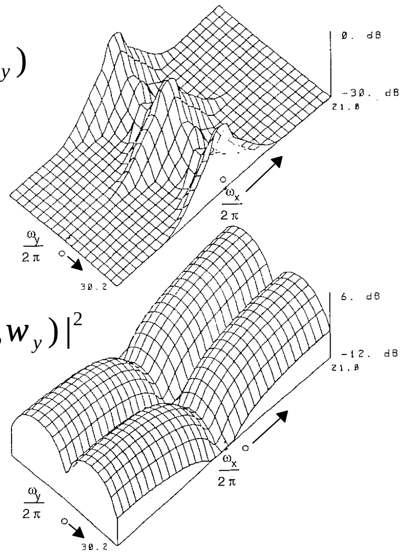
- Solution: prewhitening filter

$$|1 - P(w_x, w_y)|^2 = \frac{s_e^2}{\Phi_{ss}(w_x, w_y)}$$

Power spectrum of a luminance signal and the prewhitening filter

$\Phi_{ss}(w_x, w_y)$

$|1 - P(w_x, w_y)|^2$



- Luminance signal Y sampled at 10 MHz.
- w_x, w_y in cycles/degree (cpd) for a viewing distance of six times screen height (6H).
- Predictor coefficients:

  $a_1 = 7/8$

  $a_2 = -5/8$

  $a_3 = 3/4$

Summary: lossless compression

- Redundancy reduction exploits the properties of the signal source.
- Properties of the signal source can be described statistically.
- Entropy is the lower bound for the average code word length.
- Huffman code is optimum entropy code.
- Huffman coding: needs code table.
- Arithmetic coding is a universal coding method for encoding strings of symbols.
- Arithmetic coding does not need a code table.
- For an efficient, independent coding of symbols, statistical dependencies have to be removed.