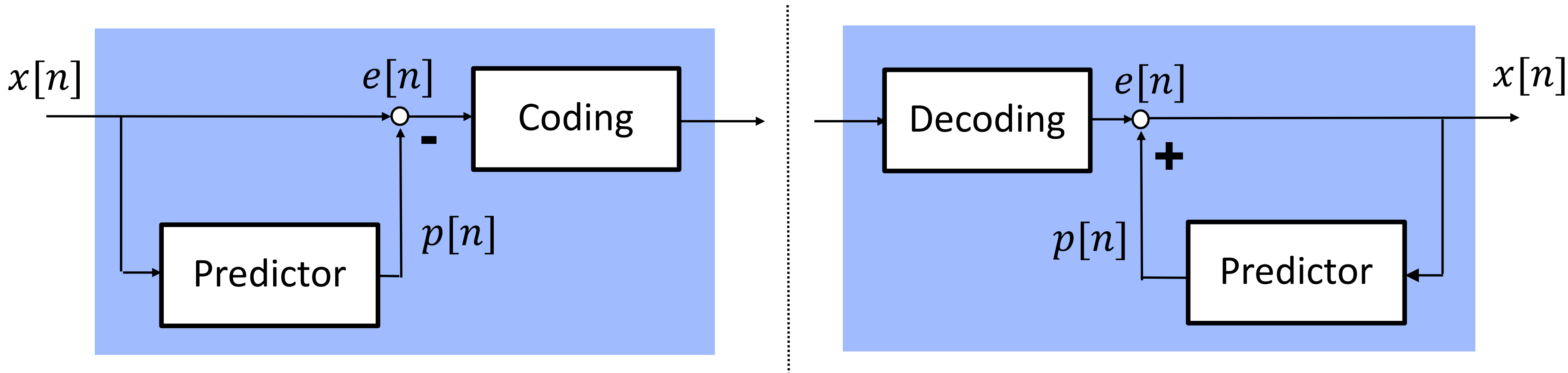
6 Predictive Coding

6.1 Lossless Predictive Coding



Prediction signal $p[n]$ is calculated from $x[n]$ using previous samples

- Prediction error $e[n]$ has reduced statistical dependencies between adjacent samples
- Entropy coding of prediction error typically under assumption of iid property

Receiver can reconstruct $x[n]$ from $e[n]$ without loss

Linear Predictor

Linear predictor: weighted superposition of past N samples

$$p[n] = \sum_{i=1}^{N} a_i \, x[n-i]$$

Predictor design: minimize MSE between original signal and its prediction

$$\mathrm{E}\{e^2[n]\} = \mathrm{E}\{(x[n]-p[n])^2\} = \mathrm{E}\left\{\left(x[n] - \sum_{i=1}^{N} a_i x[n-i]\right)^2\right\}$$

by setting

$$\frac{\partial}{\partial a_i}\mathrm{E}\{e^2[n]\} = 0 \quad \text{for} \quad i = 1,2,\dots,N$$

Minimization using $\varphi_{xx}[k] = \mathrm{E}\{x[n+k]x[n]\}$ yields **Wiener-Hopf equation**

$$\underbrace{\begin{bmatrix} \varphi_{xx}[1] \\ \varphi_{xx}[2] \\ \vdots \\ \varphi_{xx}[N] \end{bmatrix}}_{\boldsymbol{\varphi}_{xx}} = \underbrace{\begin{bmatrix} \varphi_{xx}[0] & \varphi_{xx}[1] & \cdots & \varphi_{xx}[N-1] \\ \varphi_{xx}[1] & \varphi_{xx}[0] & \cdots & r_{xx}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{xx}[N-1] & \varphi_{xx}[N-2] & \cdots & \varphi_{xx}[0] \end{bmatrix}}_{\boldsymbol{\Phi}_{xx}} \cdot \underbrace{\begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[N] \end{bmatrix}}_{\boldsymbol{a}}$$

which are solved by Toeplitz matrix inversion: $\boldsymbol{a} = \boldsymbol{\Phi}_{xx}^{-1}\boldsymbol{\varphi}_{xx}$

Linear Predictor (cont.)

Variance of prediction error

$$\sigma_e^2 = \sigma_x^2 - \sum_{i=1}^{N} a_i \varphi_{xx}[i]$$

$$= \sigma_x^2 - \boldsymbol{\varphi}_{xx}^T \boldsymbol{a} = \sigma_x^2 - \underbrace{\boldsymbol{\varphi}_{xx}^T \boldsymbol{\Phi}_{xx}^{-1} \boldsymbol{\varphi}_{xx}}_{\geq 0}$$

σ_x^2: variance of source signal

Prediction gain of linear prediction $G = 10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2}$ [dB]

Practical aspects in predictive coding

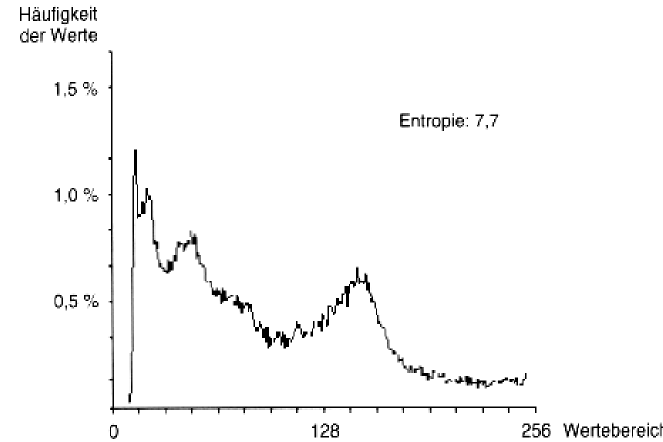
- Prediction error PMF for images in case of linear prediction is typically Laplacian, i.e. first order entropy is reduced such that simple VLC can profit
- Minimization of prediction error entropy instead of MSE leads to similar results
- Sensitivity to transmission errors, remedy by periodical PCM refresh
- Initialization of predictor (e.g. for first row in a picture) by mean value, zero in case of Laplacian PMF

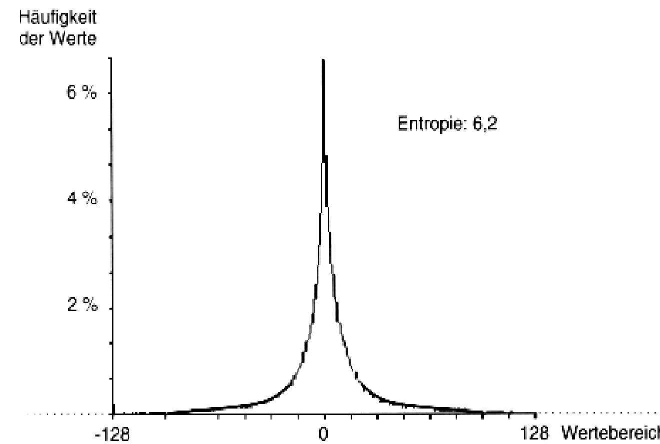
Example for Prediction Error

$x_1 \rightarrow x_0$

1D scan line prediction

Original image and prediction error

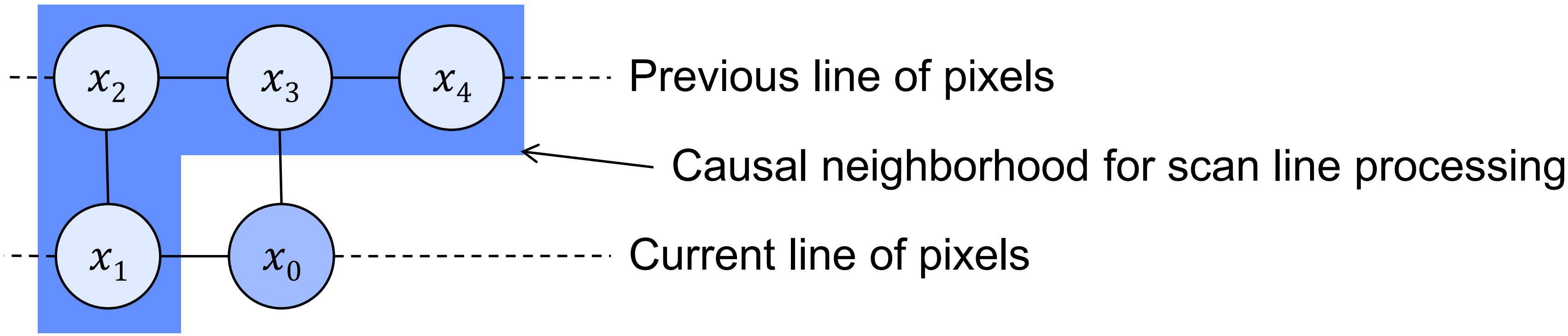




Histogram of original and prediction error

6.2 Optimum 2D Linear Predictor

Image signals can take advantage of 2D prediction:



⇨ Wiener-Hopf equations are solved for 2D signal $x[m, n]$ in causal neighbor-hood, e.g. by reordering pixels into 1D vector $x_0, x_1, x_2, \ldots$

Example: image “Lena”, 512×512 pixels, 8 bpp

⇨ Optimal minimum MSE 2D linear predictor using neighborhood x_1, x_2, and x_3

a_1	a_2	a_3	$H(X)$	$H(E)$
0.59	-0.43	0.83	7.23 bit	4.30 bit

X has approx. uniform PMF, prediction error E is approx. Laplacian distributed

Demo 6a „Spatial Prediction“

6.3 JPEG-LS Lossless Compression Standard

Formal notation: ISO/IEC 14495-1 and ITU-T T.87

- Based on LOCO-I („low complexity compression of images“) from HP Labs
- Significantly differs from lossless mode of original JPEG

Principle: predictive coding with non-linear predictor

- Context-adaptive Golomb coding of prediction error
- 365 different 2D coding contexts based on pixel differences on causal neighborhood
- Switchable to 1D run-length coding

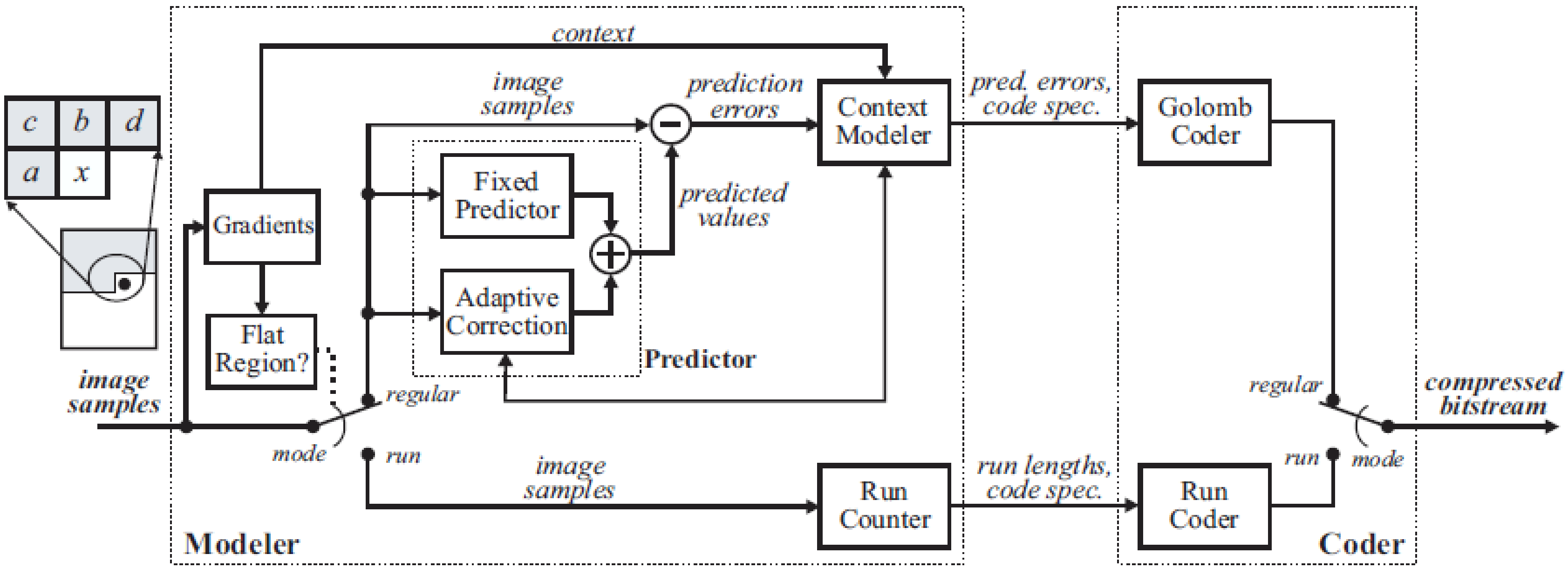
Near-lossless mode extension

- Allowing maximum deviation of p grey levels from original sample x

$$\|x - \hat{x}\|_\infty = \max|x - \hat{x}| \leq p$$

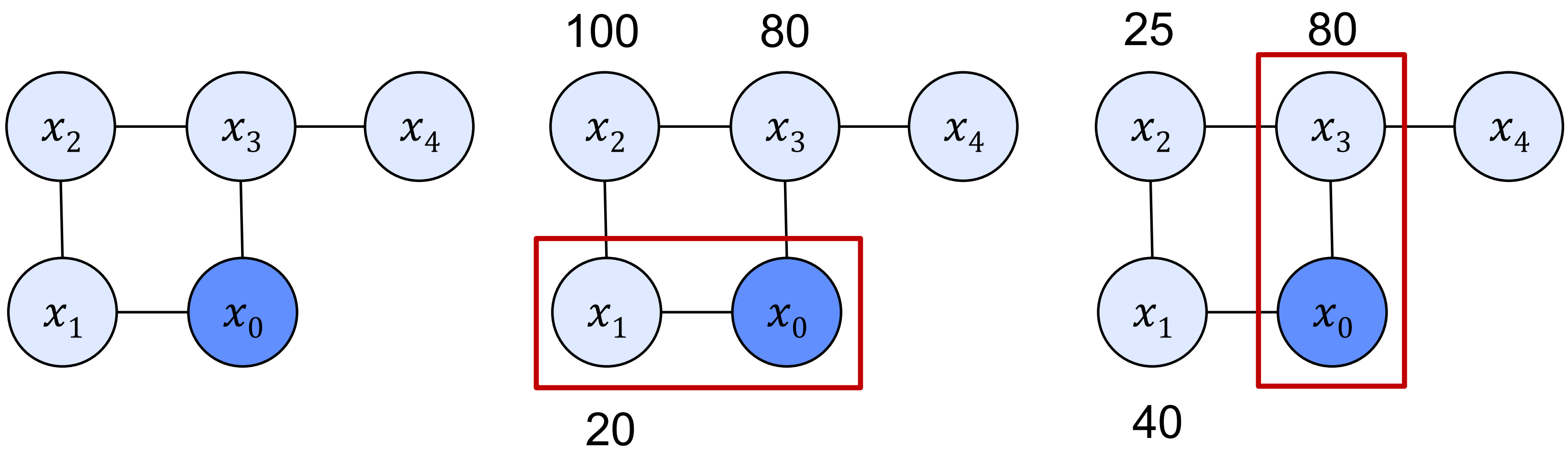
JPEG-LS Block Diagram

Block diagram using 2D spatial prediction and mode decision



[Source: Weinberger, Seroussi, Sapiro, IEEE Transactions on Image Processing, Vol. 9, August 2000]

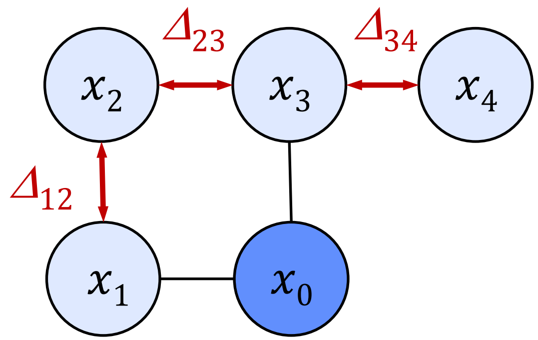
JPEG-LS Non-Linear Predictor



$$P = \begin{cases} \min(x_1, x_3) & \text{if } x_2 = \max(x_1, x_2, x_3) \\ \max(x_1, x_3) & \text{if } x_2 = \min(x_1, x_2, x_3) \\ x_1 - x_2 + x_3 & \text{else} \end{cases}$$

⇨ **Result:** horizontal or vertical directional prediction along dominant spatial image structure

JPEG-LS Context Labeling



Quantization of prediction error

- Each of three difference Δ_{ij} is quantized to 1 out of 9 different indices using default thresholds 3, 7, 21

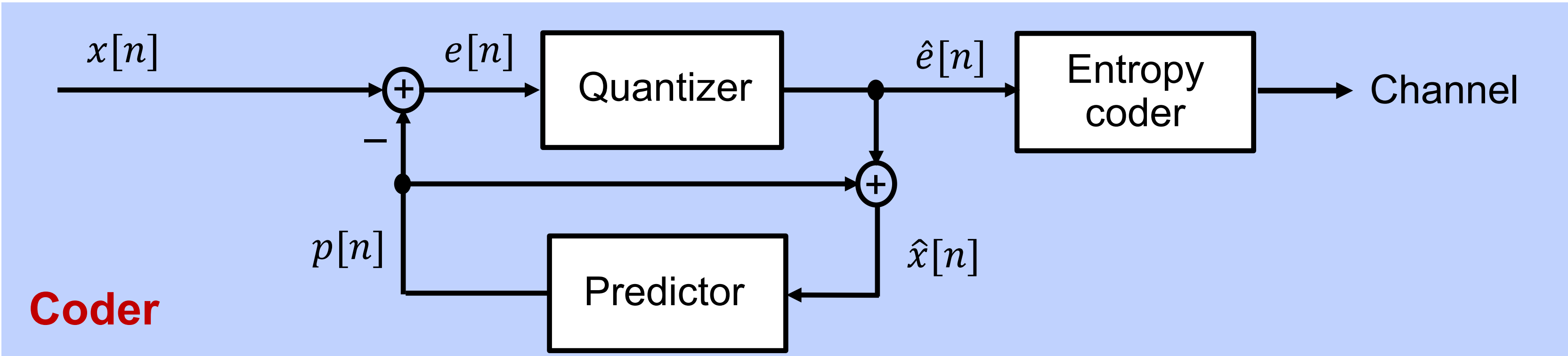
Labeling using quantized errors

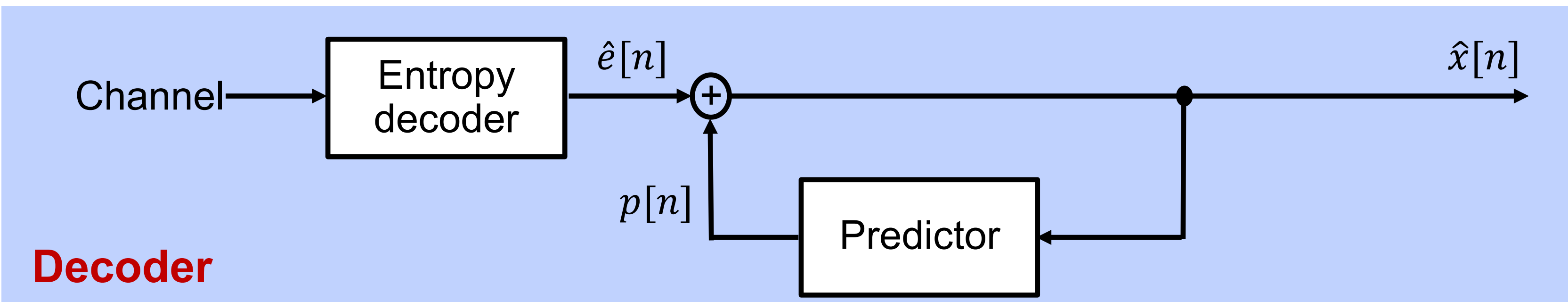
- Results in 9^3 =729 context which are reduced to 365 exploiting sign symmetries

Run-length coding mode

- Switch to run-length coding mode if all Δ_{ij} are equal to zero

6.4 Differential Pulse Code Modulation (DPCM)





Reconstruction error is solely determined by quantization error

Prediction error:	$e[n] = x[n] - p[n]$
Quantization error:	$e_q[n] = \hat{e}[n] - e[n]$
Reconstruction:	$\hat{x}[n] = \hat{e}[n] + p[n]$
Total error:	$\hat{x}[n] - x[n] = \hat{e}[n] + p[n] - x[n] = \hat{e}[n] - e[n] = e_q[n]$

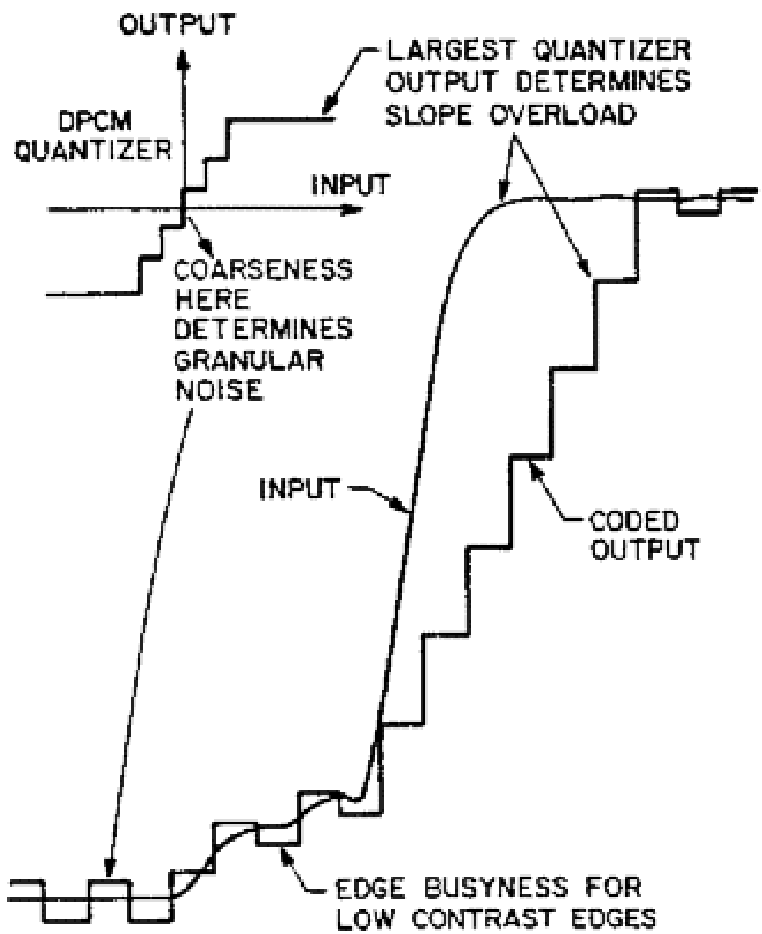
Signal Distortions in DPCM Coding

Granular noise
- Random noise in flat picture areas

Slope overload
- Blur of high-contrast edges
- Moiré patterns in periodic structures

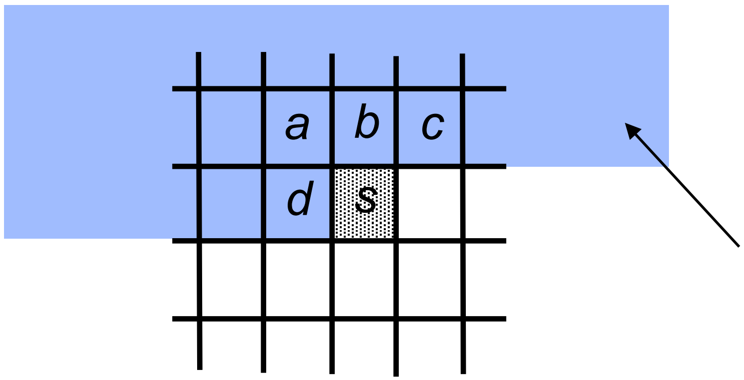
Edge busyness
- Jittery appearance of edges

Error propagation
- Transmission errors are propagated due to prediction



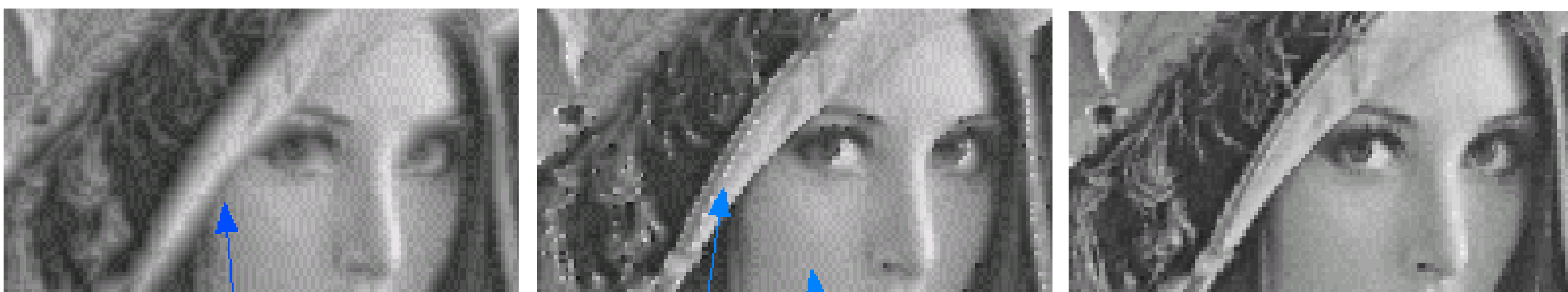
Example of Intraframe DPCM Coding

Principle: use previously transmitted 2D picture neighborhood for prediction



Causal image area used for prediction

Example: linear prediction $a = 0, b = 0.25, c = 0.25, d = 0.5$, Lloyd-Max quantizer



prediction error coding: 1bit/pixel | 2 bit/pixel | 3 bit/pixel

slope overload | *edge busyness* | *granular noise*

Demo 6b „Spatial DPCM“

Coding Gain of DPCM

High-rate distortion-rate function under assumption of MSE criterion

$$d(R) \cong \varepsilon^2 \sigma_X^2 2^{-2R}$$

In case of DPCM, only prediction error with variance σ_E^2 has to be coded:

$$d_{\mathrm{DPCM}}(R) \cong \varepsilon_E^2 \sigma_E^2 2^{-2R}$$

Prediction gain for DPCM coding defined as

$$G_{\mathrm{DPCM}} = \frac{d_{\mathrm{X}}(R)}{d_E(R)} = \frac{\varepsilon_X^2 \sigma_X^2}{\varepsilon_E^2 \sigma_E^2}$$

Scaling factor ε^2

	Shannon $D(R)$	Lloyd – Max	Entropy coded
Uniform	$\frac{6}{\pi \mathrm{e}} \cong 0.703$	1	1
Laplacian	$\frac{\mathrm{e}}{\pi} \cong 0.865$	$\frac{9}{2} = 4.5$	$\frac{\mathrm{e}^2}{6} \cong 1.232$
Gaussian	1	$\frac{\sqrt{3}\pi}{2} \cong 2.721$	$\frac{\pi \mathrm{e}}{6} \cong 1.423$

Predictive Coding - Summary

- Predictive coding widely used for lossy as well as lossless image coding
- Solve Wiener-Hopf equations for optimum linear predictor
- JPEG-LS based on 2D non-linear predictor
- DPCM coding gain: prediction reduces variance and shapes distribution
- Coding distortions are granular noise, edge busyness, and slope overload