MDL-PAR

- Minimum Description Length Piecewise Autoregressive Predictor
  - sequentialize multidimensional signals
    - adaptive filter order and support shape

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2011

Adaptive Sequential Prediction of Multidimensional Signals With Applications to Lossless Image Coding

Xiaolin Wu, *Fellow, IEEE*, Guangtao Zhai, *Member, IEEE*, Xiaokang Yang, and Wenjun Zhang, *Senior Member, IEEE*

Linear Prediction

$$\chi^K(x_0) \;=\; (x_1, x_2, \ldots, x_K)$$

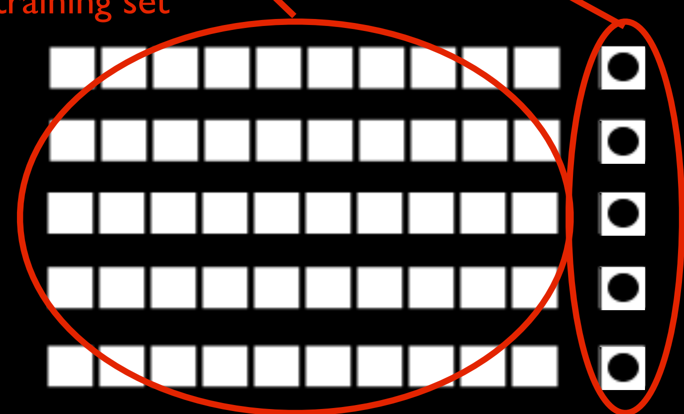
$$\mathbf{a} = \arg\min_{\alpha \in \Re^K} E_{\chi^K(x) \in S} \left\| x - \chi^K(x)\alpha^T \right\|_\ell$$

filter coefficients

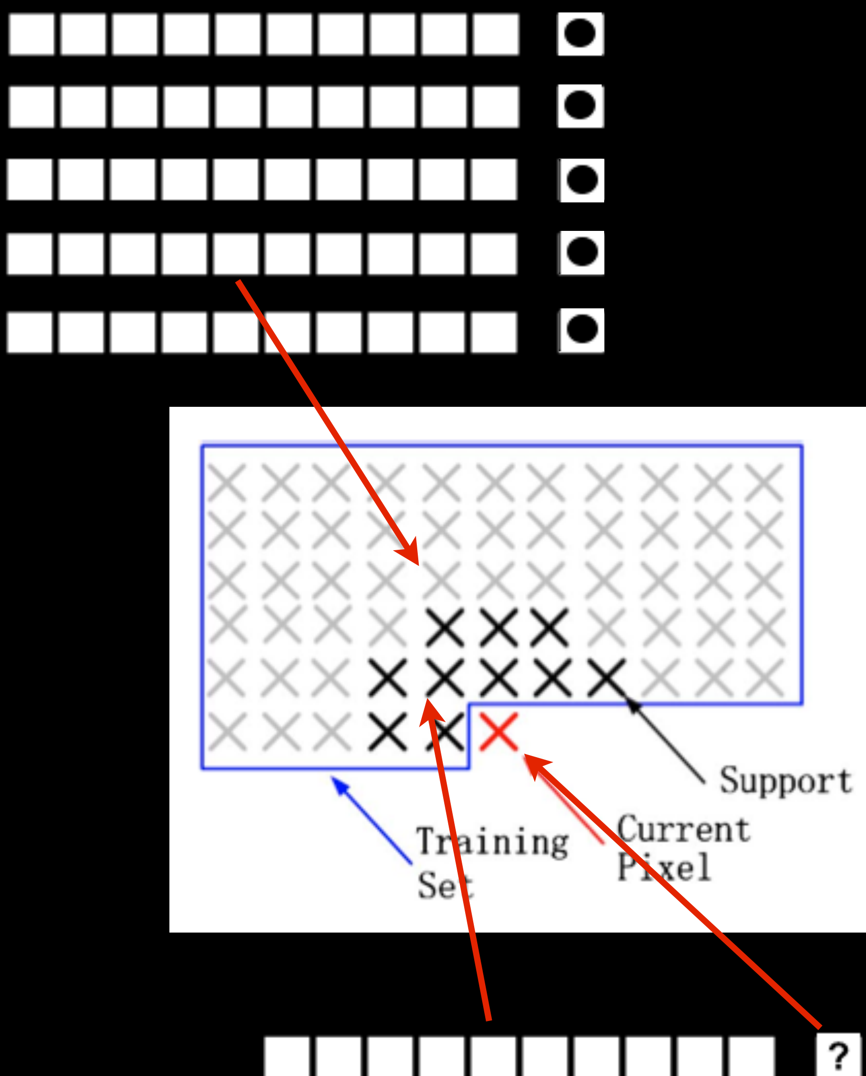
$$\mathbf{a} = (a_1, a_2, \ldots, a_K)$$

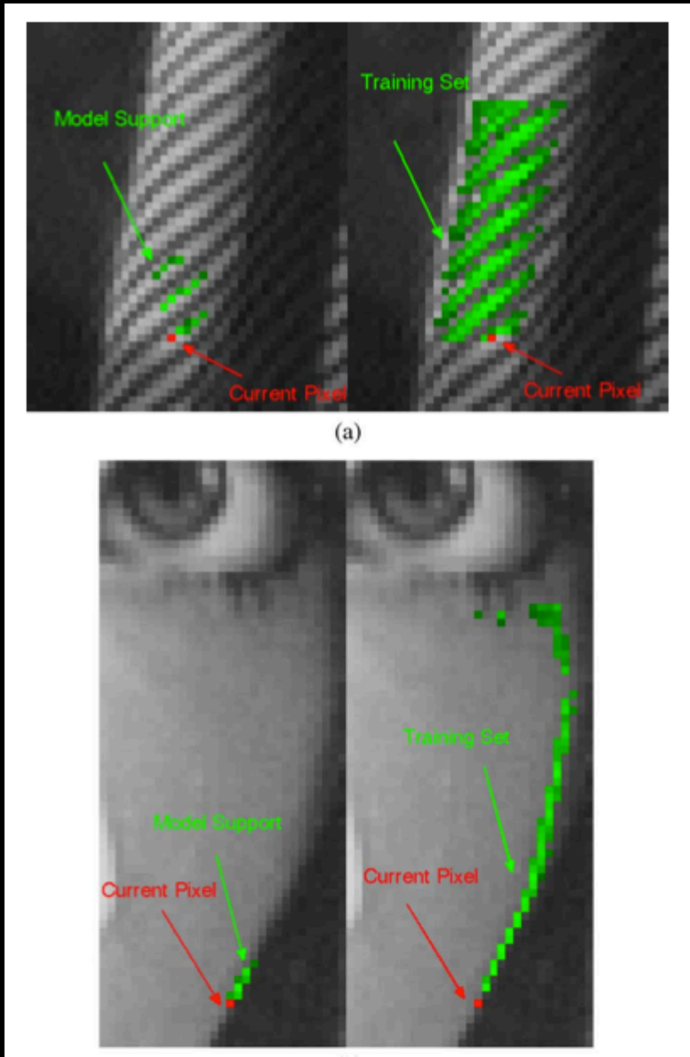
prediction value

$$\hat{x}_0 = \chi^K(x_0)\mathbf{a}^T$$



2D Linear Prediction

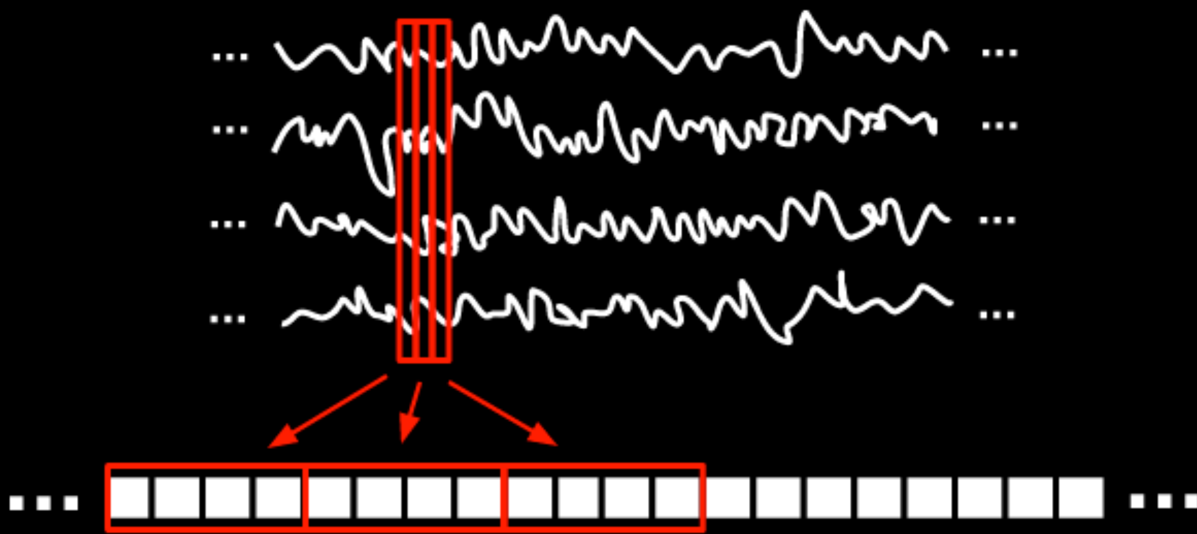




Wu et al., 2011

Data Conditioning

Serialization



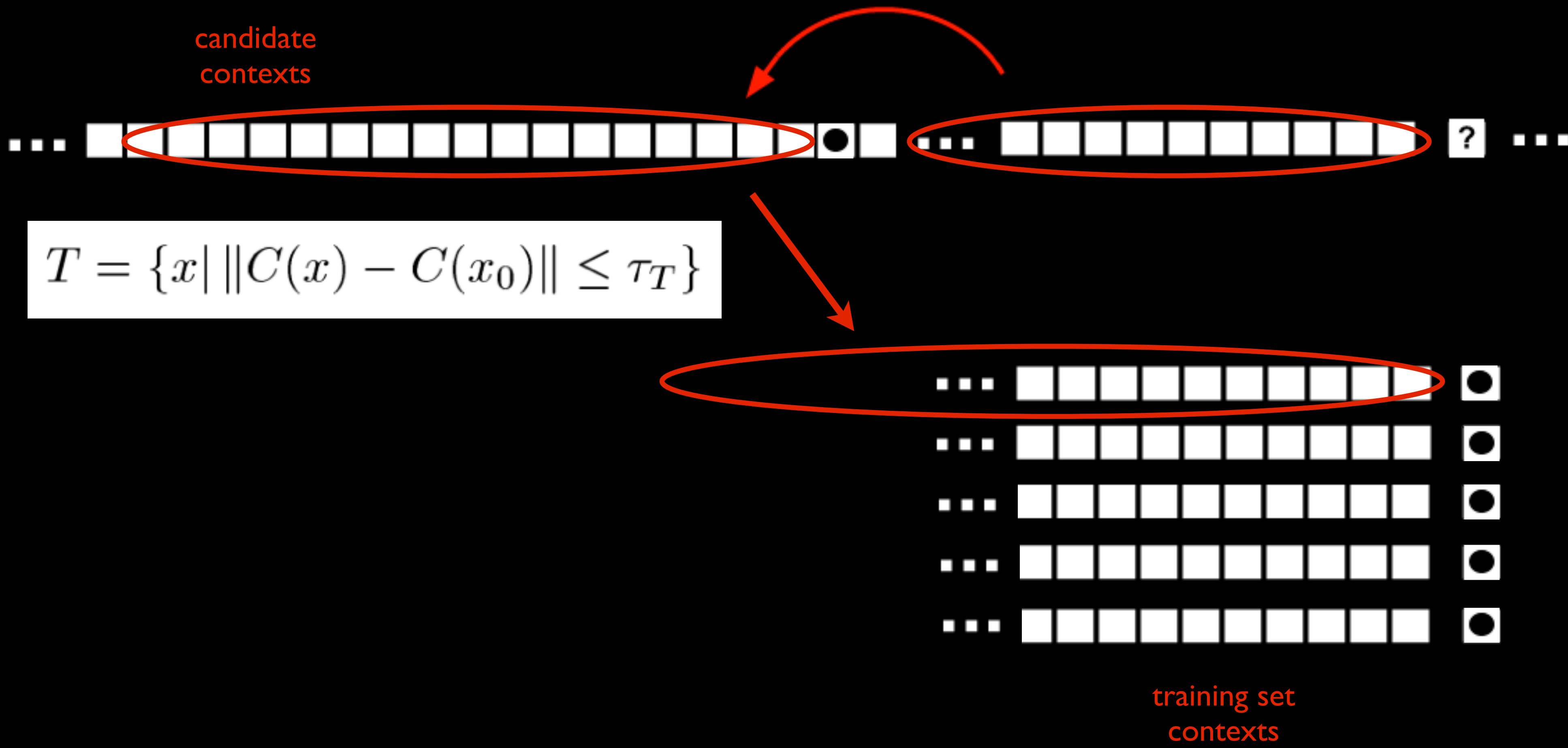
Dynamic Range Compression

(-32768 — 32767) —> (0 — 255)

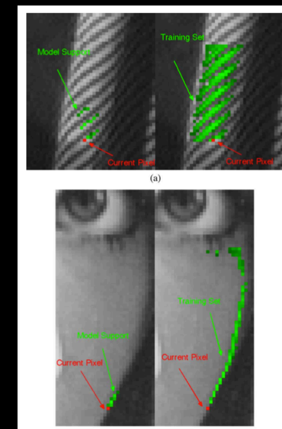
Context Selection*

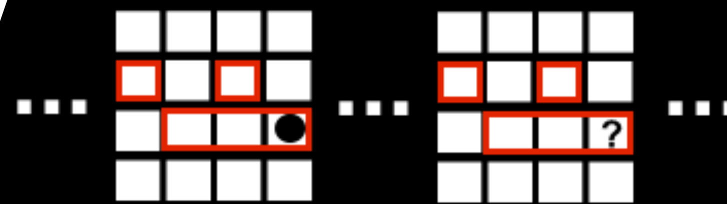
$$C(x) = (c_1(x), c_2(x), \ldots, c_M(x))$$

$$T = \{x | \, \|C(x) - C(x_0)\| \leq \tau_T\}$$

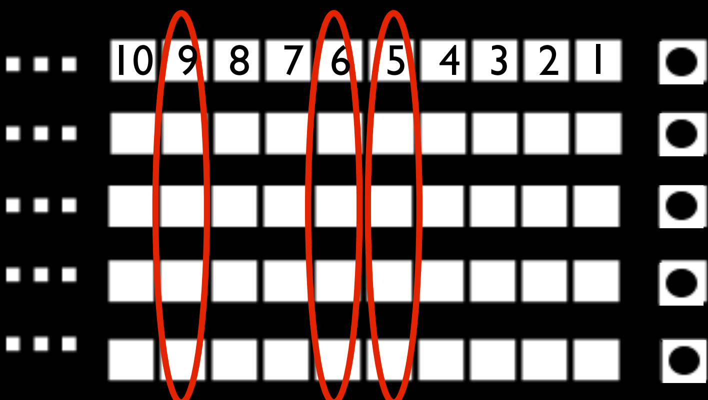


Training Set Selection





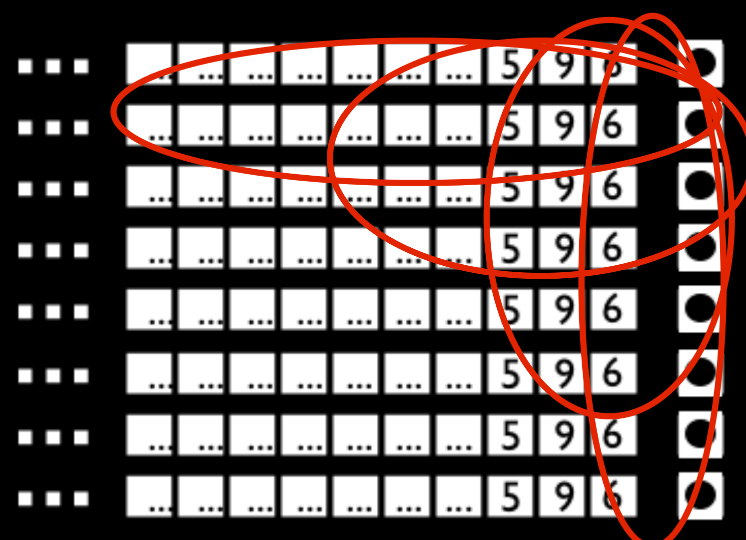
$$\hat{\rho}_{0,m} = \frac{|T| \sum_{x \in T} c_m(x) x - \sum_{x \in T} x \sum_{x \in T} c_m(x)}{\sqrt{|T| \sum_{x \in T} x^2 - \left(\sum_{x \in T} x\right)^2} \sqrt{|T| \sum_{x \in T} c_m(x)^2 - \left(\sum_{x \in T} c_m(x)\right)^2}}$$



$$S_k = \{x \,|\, \|\chi^k(x) - \chi^k(x_0)\| \leq \tau_S, x \in S_{k-1}\} \qquad k > 1$$
$$S_1 = T$$

$$S_1 \supseteq S_2 \supseteq \cdots \supseteq S_k \supseteq \cdots$$

nested training sets



$$\chi^K(x_0) = \{x_1, x_2, \cdots, x_K \,|\, |\rho_{0,1}| \geq |\rho_{0,2}| \geq \cdots \geq |\rho_{0,K}|\}$$

nested support shapes

Prediction*/Compression

$$\mathbf{a} = \arg\min_{\alpha \in \Re^K} E_{\chi^K(x) \in S} \left\| x - \chi^K(x)\alpha^T \right\|_\ell$$

MDL objective function

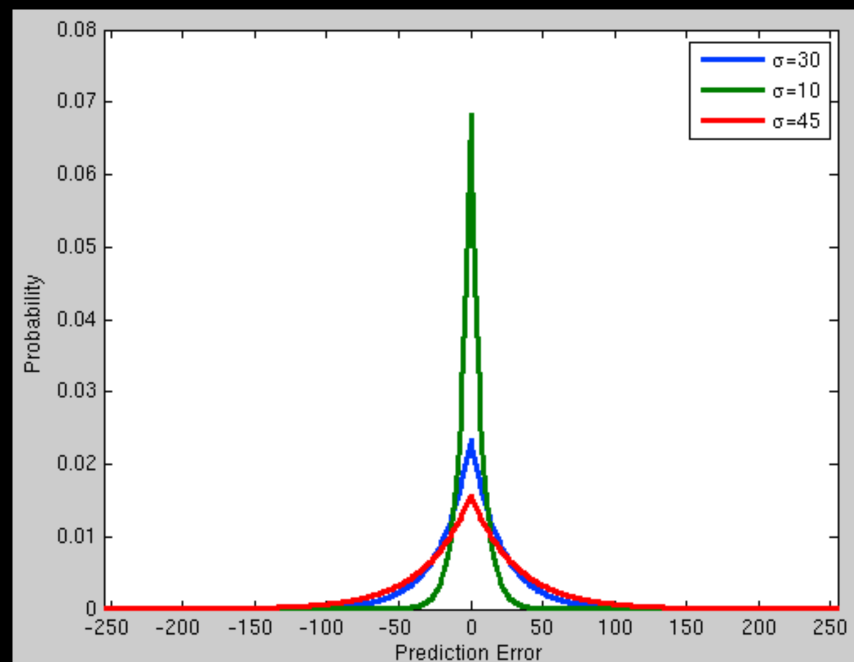
$$L(\chi^k, S) = H(\chi^k, S) + \frac{k}{2|S|} \log |S|$$

"compression code length"

"compression model size"

$$H(\chi^k, S_k) = - \sum_{-A \le d \le A} P(d) \log P(d)$$

$$P(d) = \begin{cases} 1 - e^{-\frac{1}{\sqrt{2}\sigma}}, & d = 0 \\ \frac{1}{2}\left(e^{-\frac{|d|-0.5}{\sigma/\sqrt{2}}} - e^{-\frac{|d|+0.5}{\sigma/\sqrt{2}}}\right), & 0 < |d| < A \\ \frac{1}{2} e^{-\frac{|d|-0.5}{\sigma/\sqrt{2}}}, & |d| = A. \end{cases}$$



$$\hat{x}_0 = \chi^K(x_0)\mathbf{a}^T$$

prediction error

$$x_0 - \hat{x}_0 = d$$

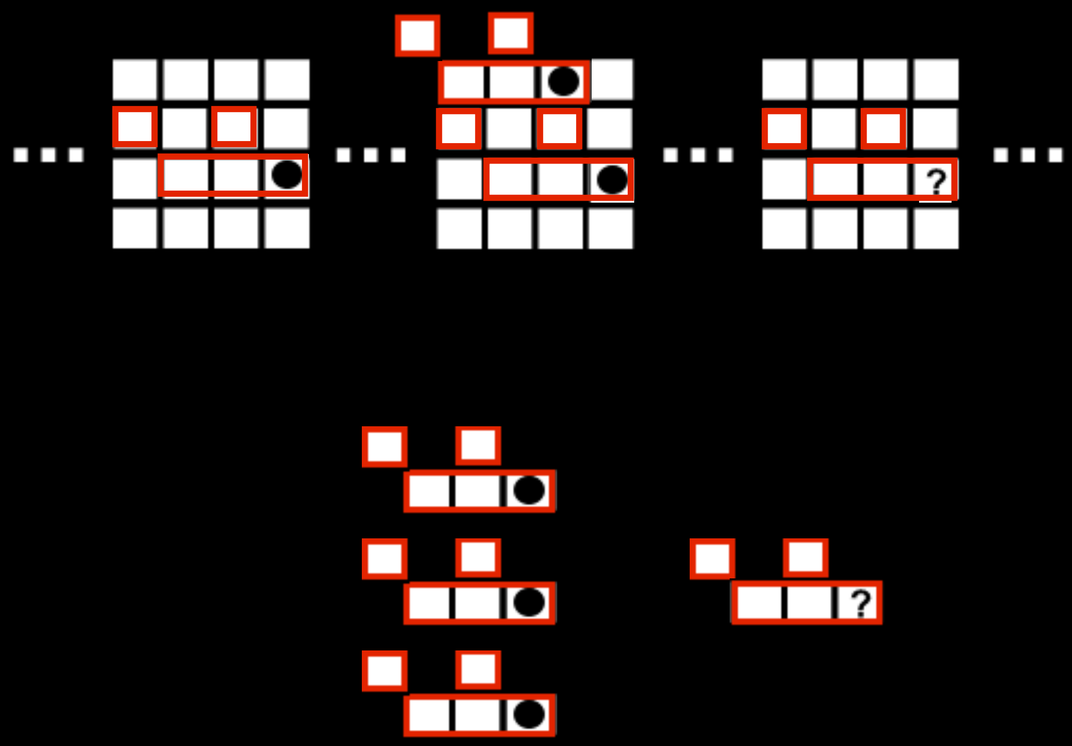
Minimum Description Length (MDL) Principle

- “The minimum description length (MDL) principle is a formalization of Occam's Razor in which the best hypothesis for a given set of data is the one that leads to the best compression of the data.” (Wikipedia)
- "[Occam's Razor] is often summarized as 'other things being equal, a simpler explanation is better than a more complex one.' […] The razor asserts that one should proceed to simpler theories until simplicity can be traded for greater explanatory power." (Wikipedia)

$$L(\chi^k, S) = H(\chi^k, S) + \frac{k}{2|S|} \log |S|$$

Wu et al., 2011

When your context is smaller than your “frame”...



Why haven't I set the context to a larger value?

$$\lim_{dim \to \infty} \frac{dist_{max} - dist_{min}}{dist_{min}} \to 0$$

$$\lim_{k \to \infty} \frac{Variance[d(p,q)]}{Expected[d(p,q)]} = 0$$

(In a naive or poorly constructed space!)

(And the algorithm gets sloooower...)

Performance

- Context selection [O(n) for a single scan in naive case]
  - scanning
  - distance calculation
  - sorting
- filter coefficient calculation [large matrix multiplication = size of training set, k]
- probability distribution construction [dynamic range / number of symbols]
- disk I/O [online vs batch writing]
- memory management [allocation and deallocation, stack vs heap, preallocation]

(Bottlenecks, and how close together are they?)

Algorithm Parameters

$$C(x) = (c_1(x), c_2(x), \ldots, c_M(x))$$

context size

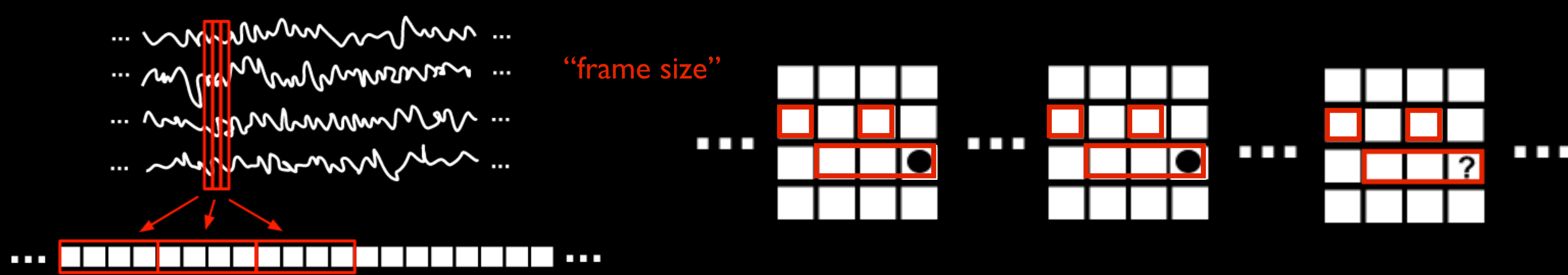
$$T = \{x \mid \|C(x) - C(x_0)\| \le \tau_T\}$$

candidate context similarity threshold

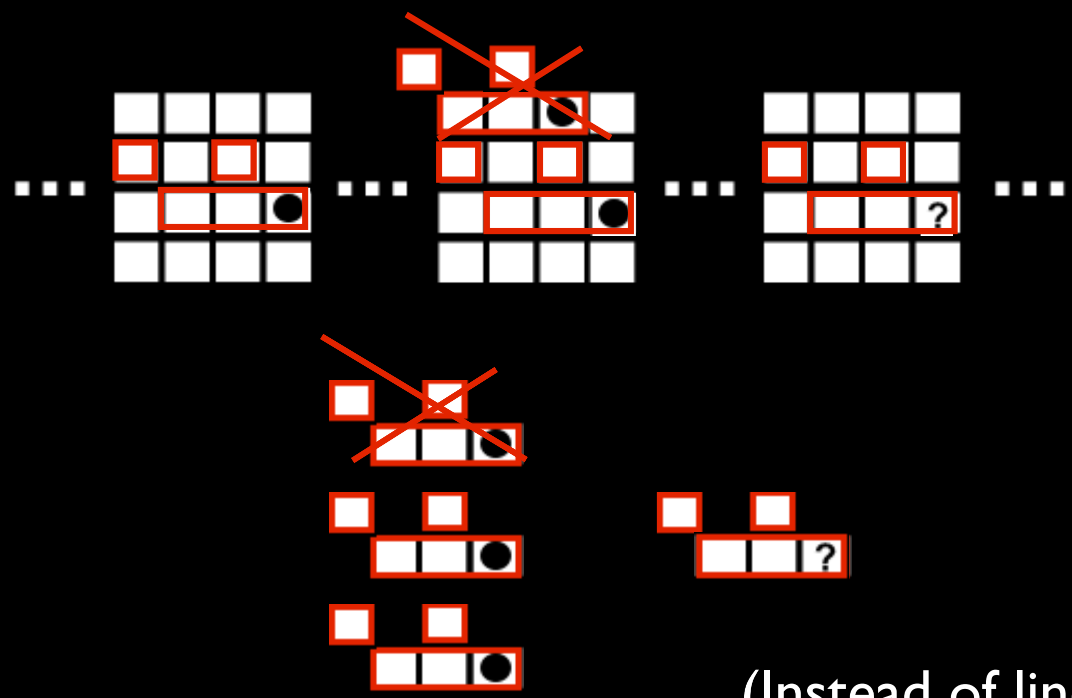
$$S_k = \{x \mid \|\chi^k(x) - \chi^k(x_0)\| \le \tau_S, x \in S_{k-1}\} \qquad k > 1$$
$$S_1 = T$$

candidate training set support threshold

Serialization

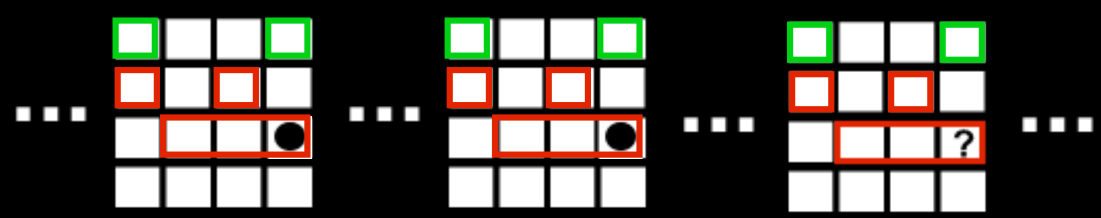


“Frame locking” (“frame size” as a new parameter)...



(Instead of linear scan, only search for candidate contexts at multiples of the frame size.)

But if context is still too small...



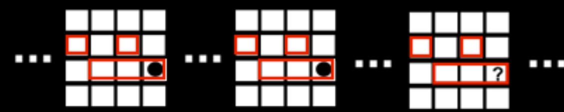
Moving forward...

- Key issues:
  - noisy context selection in high dimensions (probably)
  - performance

- Possible solution:
  - “frame locked” approximate nearest neighbor searching”
    - Better context matching
    - Faster context matching
    - Bigger contexts?

possible 50-100x speedup with better results

“frame-locked” approximate nearest neighbor searching



- a “context repository” for each
  - (predictionSymbolIndex *mod* frameSize)
- context repository is a red-black tree indexed on contexts collapsed to 1D (cf. chinese remainder theorem)
  - O(log n) vs O(n) [e.g. log2 (1,000,000) ~= 20]
- (it already works with a single context repository -- massive speedup, with slightly degraded compression)