

# Generalized Forms of the Kraft Inequality for Finite-State Encoders

**Neri Merhav**

The Viterbi Faculty of Electrical and Computer Engineering  
Technion – Israel Institute of Technology, Haifa, Israel

*Entropy Special Issue Webinar on  
Information Theory and Data Compression  
May 27, 2026*

# Outline

- 1 Background & Motivation
- 2 The Kraft Matrix and Basic GKI
- 3 Stronger Results for Irreducible Encoders
- 4 Converse Bounds
- 5 Extensions
- 6 Summary

# The Classical Kraft–McMillan Inequality

## Classical Kraft Inequality (CKI)

A uniquely decodable (UD) binary code with codeword lengths  $\{L(x)\}_{x \in \mathcal{X}}$  satisfies:

$$\sum_{x \in \mathcal{X}} 2^{-L(x)} \leq 1.$$

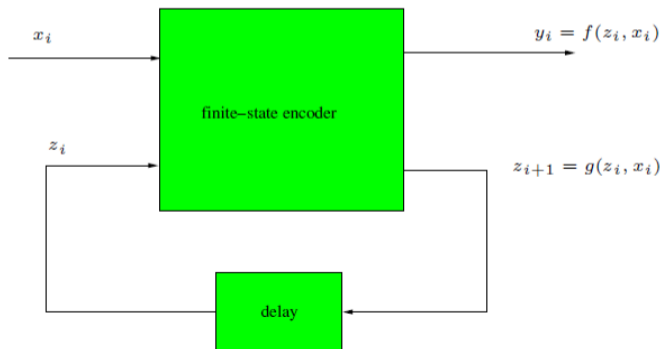
**Why it is so central:**

- Necessary *and* sufficient condition for the existence of a UD code
- The key tool in proving the source coding converse: rate  $\geq$  entropy
- Connects combinatorial code structure to analytical bounds elegantly

**Natural question:** Can we generalize this to encoders with **memory**?

# Finite-State Encoders

- $\mathcal{X}$ : source alphabet of size  $\alpha$ ;
- $\mathcal{Y}$ : output alphabet = set of variable-length binary strings (including 'null');
- $\mathcal{Z}$ : set of states of size  $s$ ;
- $f : \mathcal{Z} \times \mathcal{X} \rightarrow \mathcal{Y}$ : output function (variable-length binary strings);
- $g : \mathcal{Z} \times \mathcal{X} \rightarrow \mathcal{Z}$ : next-state function.



## Finite-State Encoders (Cont'd)

Notation:  $f(z_i, x_i^{i+\ell}) = y_i^{i+\ell}$ ;  $g(z_i, x_i^{i+\ell}) = z_{i+\ell+1}$

### Information Losslessness (IL)

Given any initial state  $z_i$ , block length  $n$ , and input  $x_i^{i+n}$ , the triplet  $(z_i, f(z_i, x_i^{i+n}), g(z_i, x_i^{i+n}))$  *uniquely* determines  $x_i^{i+n}$ .

**Challenge:** Different states produce outputs of different lengths  $\Rightarrow$  no single Kraft sum captures feasibility. A *matrix* generalization is needed.

# Prior Work and Its Limitations

## Ziv & Lempel GKI (1978)

For every IL encoder with  $s$  states and every block length  $\ell$ :

$$\sum_{x^\ell \in \mathcal{X}^\ell} 2^{-\min_z L[f(z, x^\ell)]} \leq s^2 \left[ 1 + \log \left( 1 + \frac{\alpha^\ell}{s^2} \right) \right]$$

**Two main weaknesses:**

- 1. Does not recover the CKI for  $s = 1$ :** the right-hand side is  $1 + \log(1 + \alpha^\ell) > 1$ , and the inequality is always strict regardless of  $\ell$
- 2. Super-alphabet formulation:** tight only asymptotically in  $\ell$ , whereas the CKI is stated directly at the single-symbol level

**Our goal:** A GKI that (i) reduces *exactly* to the CKI for  $s = 1$ , and (ii) is posed at the single-symbol/state level with no block extensions.

# The Kraft Matrix

## Definition

For an FS encoder with  $s$  states, define the  $s \times s$  nonnegative matrix  $\mathbf{K}$  by

$$K_{zz'} = \sum_{x: g(z,x)=z'} 2^{-L[f(z,x)]}, \quad (z, z') \in \mathcal{Z}^2.$$

## Intuition:

- $K_{zz'}$  accumulates the “Kraft weights” of all single-step transitions  $z \rightarrow z'$
- **Single-symbol level:** defined directly via  $f$  and  $g$ , no block extensions
- **Power formula:**  $(\mathbf{K}^\ell)_{zz'} = \sum_{\{x^\ell: g(z,x^\ell)=z'\}} 2^{-L[f(z,x^\ell)]}$  (the Kraft sum for  $\ell$ -blocks)
- **For  $s = 1$ :**  $\mathbf{K}$  is the scalar  $\sum_x 2^{-L(x)}$  — the ordinary Kraft sum

Since  $\mathbf{K} \geq 0$ , its spectral radius  $\rho(\mathbf{K})$  is itself an eigenvalue (Perron–Frobenius).

# Theorem 1: The Basic GKI

## Theorem 1

For every information-lossless FS encoder,

$$\varrho(\mathbf{K}) \leq 1.$$

### Intuition behind the proof:

- The IL property forces an injective mapping from input strings to codewords for each fixed pair of start/end states
- By a standard counting argument (at most  $2^l$  strings of total code-length  $l$ ), each entry  $(\mathbf{K}^\ell)_{zz'}$  can grow *at most linearly* in  $l$
- If  $\varrho(\mathbf{K}) > 1$ , the Perron–Frobenius theorem would force *exponential* growth of  $\mathbf{K}^\ell$
- **Contradiction** — so  $\varrho(\mathbf{K}) \leq 1$

### Two key properties:

- ✓ For  $s = 1$ :  $\varrho(\mathbf{K}) = \sum_x 2^{-L(x)} \leq 1$  — *exactly* the CKI
- ✓ Defined and verifiable at the single-symbol/state level

## Example: A 2-Block Code as a 3-State FS Encoder

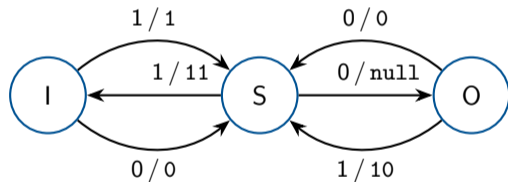
Binary source; block code of length 2:  $00 \rightarrow 0$ ,  $01 \rightarrow 10$ ,  $10 \rightarrow 110$ ,  $11 \rightarrow 111$ .

**States:** S (start), O (1st bit=0), I (1st bit=1)

Key feature: entering state I, the encoder immediately outputs 11 (the common prefix of 110 and 111).

**Kraft matrix** (order S, O, I):

$$\mathbf{K} = \begin{pmatrix} 0 & 1 & \frac{1}{4} \\ \frac{3}{4} & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$



Eigenvalues:  $\{1, 0, -1\} \Rightarrow \rho(\mathbf{K}) = 1 \checkmark$

Note: row sum of row S exceeds 1, yet  $\rho(\mathbf{K}) \leq 1$ .

The CKI need *not* hold per state.

## Definition

An FS encoder is **irreducible** if every state is reachable from every state in a finite number of steps (equivalently, the matrix  $\mathbf{K}$  is irreducible).

### Why irreducibility is the natural assumption:

- In a reducible encoder, once the machine leaves a set of transient states it *never returns*
- Asymptotically, only the recurrent (irreducible) component is active
- **Reducible encoders offer no asymptotic advantage over irreducible ones,** even for individual sequences
- So without loss of generality one may always work with irreducible encoders

### Richer toolbox for irreducible $\mathbf{K} \geq 0$ :

- The **Perron–Frobenius theorem** applies in its full (irreducible) form
- The **Collatz–Wielandt formulas** give min-max and max-min characterizations of  $\varrho(\mathbf{K})$ , leading to several equivalent and explicit forms of the GKI

# Equivalent GKI Forms for Irreducible Encoders

From the Collatz–Wielandt min-max and max-min formulas applied to  $\varrho(\mathbf{K}^\ell) \leq 1$ :

$\forall \mathbf{w} \in \mathcal{W}_+, \exists z$  with  $w_z > 0$  such that:

$$\sum_{z' \in \mathcal{Z}} w_{z'} \sum_{\{x^\ell: g(z, x^\ell) = z'\}} 2^{-L[f(z, x^\ell)]} \leq w_z$$

$\exists \mathbf{w} \in \mathcal{W}_+$  such that  $\forall z$  with  $w_z > 0$ :

$$\sum_{z' \in \mathcal{Z}} w_{z'} \sum_{\{x^\ell: g(z, x^\ell) = z'\}} 2^{-L[f(z, x^\ell)]} \leq w_z$$

Applying the former with  $\mathbf{w} = \mathbf{1}$ :  $\exists z \in \mathcal{Z} \sum_{x^\ell \in \mathcal{X}^\ell} 2^{-L[f(z, x^\ell)]} \leq 1$ .

*At least one initial state yields Kraft sum  $\leq 1$ .*

## Theorem 2: Kraft Sums Are Uniformly Bounded

### Theorem 2

Let  $\mathbf{K}$  be an *irreducible* Kraft matrix and  $L_{\max} = \max_{z,x} L[f(z,x)]$ . For all states  $z, z' \in \mathcal{Z}$  and every  $n \geq 1$ :

$$(\mathbf{K}^n)_{zz'} = \sum_{\{x^n: g(z,x^n)=z'\}} 2^{-L[f(z,x^n)]} \leq 2^{(s-1)L_{\max}}.$$

Summing over destination states:

$$\sum_{x^n \in \mathcal{X}^n} 2^{-L[f(z,x^n)]} \leq s \cdot 2^{(s-1)L_{\max}} \quad \forall z \in \mathcal{Z}.$$

### Why this is much stronger than Theorem 1:

- General (reducible) bound from Thm. 1:  $s(1 + nL_{\max})$  — **linear** in  $n$
- Irreducible bound (Thm. 2):  $s \cdot 2^{(s-1)L_{\max}}$  — **constant** in  $n$
- **For**  $s = 1$ : bound is  $2^0 = 1$  — *exactly* the CKI, as it should be

**Intuition:** Irreducibility forces the encoder to revisit every state, so Kraft mass cannot accumulate indefinitely along any direction.

# Compression Rate: Lower Bounds

Applying Theorem 2 and Jensen's inequality to any joint distribution  $P(z, x^\ell)$  gives the following converse bound on the compression rate  $R$ :

## Converse Bound on Rate (irreducible encoders)

$$R \geq \sup_{\ell \geq 1} \left\{ H(X_\ell | X^{\ell-1}) - \frac{2 \log s + (s-1)L_{\max}}{\ell} \right\}$$

**Why this improves on Ziv–Lempel:**

	Correction term	Decay rate
Our bound (Thm. 2, irreducible)	$[2 \log s + (s-1)L_{\max}]/\ell$	$O(1/\ell)$
Ziv–Lempel / reducible bound	$\log[s^2 \log(1 + \alpha^\ell/s^2)]/\ell$	$O(\log \ell/\ell)$

A faster-decaying correction means the bound becomes tight *sooner*, giving a sharper converse at finite block lengths. An analogous bound holds for individual sequences in terms of LZ complexity  $c(x^n)$ .

# Prediction: Lower Bound on Loss

**Setting:** Irreducible FS predictor  $(u, v)$  with  $q$  states;

$$\hat{x}_{i+1} = u(\sigma_i, x_i); \quad \sigma_{i+1} = v(\sigma_i, x_i).$$

Prediction error:  $\varepsilon_i = x_i - \hat{x}_i$ .

Prediction performance criterion:  $\sum_i \varrho(x_i - \hat{x}_i)$ .

**Key idea — predictive coding bridge:**

- 1 Map  $x^k \rightarrow \varepsilon^k$ , compress  $\varepsilon^k$  with  $L(x^k) = L(\varepsilon^k) = -\log P(\varepsilon^k)$ , where  $P(\varepsilon^k) \propto \prod_{i=1}^k e^{-\varrho(\varepsilon_i)/\theta}$ .
- 2 The corresponding encoder has  $q \cdot \alpha^k$  states; apply the GKI for a *lower* bound on  $L(x^n)$ .
- 3 Contrast  $L(x^k)$  with the GKI lower bound optimized over  $\theta$ .

## Prediction Lower Bound

Let  $Z(\theta) = \sum_x e^{-\varrho(x)/\theta}$  and  $\Delta(R) = \sup_{\theta \geq 0} \theta[R - \log Z(\theta)]$ . Then:

$$\frac{1}{n} \sum_{i=1}^n \varrho(x_i - \hat{x}_i) \geq \Delta \left( \hat{H}(X_\ell | X^{\ell-1}) - \frac{2k \log(q\alpha) + (q\alpha^k - 1)L_{\max}}{\ell} - O\left(\frac{1}{k}\right) \right)$$

## Extension I: Side Information

**Setting:** Both encoder and decoder observe side information  $w_i \in \mathcal{W}$  synchronously.

A family of Kraft matrices, one per SI symbol

$$[K(w)]_{zz'} = \sum_{\{x: g(z,x,w)=z'\}} 2^{-L[f(z,x,w)]} \quad \text{For SI sequence } w^n: K(w^n) = K(w_1) \cdots K(w_n)$$

**Theorem 3: GKI with Side Information**

For an IL FS encoder with SI, the **joint spectral radius (JSR)** of the family  $\{K(w)\}_{w \in \mathcal{W}}$  satisfies:

$$\rho_{\text{JSR}} = \lim_{n \rightarrow \infty} \max_{w^n} \|K(w^n)\|^{1/n} \leq 1.$$

**Key subtlety:** individual spectral radii are not enough

Even if  $\rho(K(w)) \leq 1$  for every  $w$  individually, the JSR can be **arbitrarily large**. Example:  $A = \begin{pmatrix} \varepsilon & 1/\varepsilon \\ 0 & \varepsilon \end{pmatrix}$ ,  $B = A^\top$  have  $\rho(A) = \rho(B) = \varepsilon$  yet  $\rho(AB) \approx 1/\varepsilon^2$ .

$\Rightarrow$  The JSR is the *correct* and *necessary* quantity here.

## Extension II: Lossy Compression

**Model:** Source block  $x^\ell$  is quantized to  $\hat{x}^\ell = Q(x^\ell)$  within distortion  $D$ ; then  $\hat{x}^\ell$  is losslessly compressed by an IL FS encoder.

**Key observation:** Each  $x^\ell$  inside the distortion ball of  $\hat{x}^\ell$  contributes the *same* codeword as  $\hat{x}^\ell$ , so the lossy Kraft matrix satisfies  $K_{zz'} \leq B_\ell \cdot \hat{K}_{zz'}$ , where  $B_\ell = \max_{\hat{x}^\ell} |\{x^\ell : d(x^\ell, \hat{x}^\ell) \leq D\}|$  and  $\hat{K}$  is the lossless Kraft matrix on the reproduction alphabet.

### Lossy GKI

Since  $\varrho(\hat{K}) \leq 1$ :  $\varrho(K) \leq B_\ell$ . For additive distortion measures (method of types):

$$\varrho(K) \leq 2^{\ell \Phi(D)}, \quad \Phi(D) = \max_{\{P_{X\hat{X}}: \mathbb{E}d(X, \hat{X}) \leq D\}} H(X|\hat{X}).$$

**Interpretation:**  $\Phi(D)$  is the maximum equivocation allowed by the distortion constraint — precisely the term that inflates the Kraft budget beyond 1.

## Main Contributions

- 1 **Kraft matrix  $\mathbf{K}$ :** a single  $s \times s$  nonneg. matrix capturing all encoder structure at the single-symbol/state level
- 2 **Basic GKI (Thm. 1):**  $\varrho(\mathbf{K}) \leq 1$  is necessary for IL; exact recovery of the CKI for  $s = 1$ ; no super-alphabets
- 3 **Irreducible encoders:** several equivalent forms (Collatz–Wielandt); *uniform* boundedness of Kraft sums (Thm. 2):  $O(1)$  vs.  $O(n)$  for the general case
- 4 **Sharper converses:**  $O(1/\ell)$  correction to entropy rate vs.  $O(\log \ell/\ell)$  for Ziv–Lempel; analogous individual-sequence and prediction bounds
- 5 **Side information:** GKI via joint spectral radius  $\varrho_{\text{JSR}} \leq 1$ ; individual spectral radii are insufficient
- 6 **Lossy coding:**  $\varrho(\mathbf{K}) \leq 2^{\ell\Phi(D)}$ ; equivocation inflates the Kraft budget

**Thank you!** [merhav@technion.ac.il](mailto:merhav@technion.ac.il)