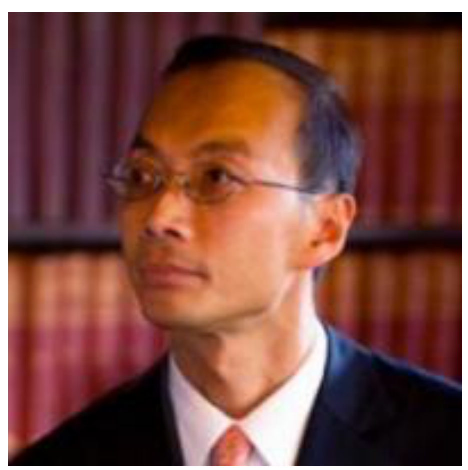
Fundamental Limits in Asynchronous Communication

Aslan Tchamkerten
Telecom ParisTech



V. Chandar	D. Tse	G. Wornell	L. Li	G. Caire
D.E. Shaw	Stanford	MIT	TPT	TU Berlin

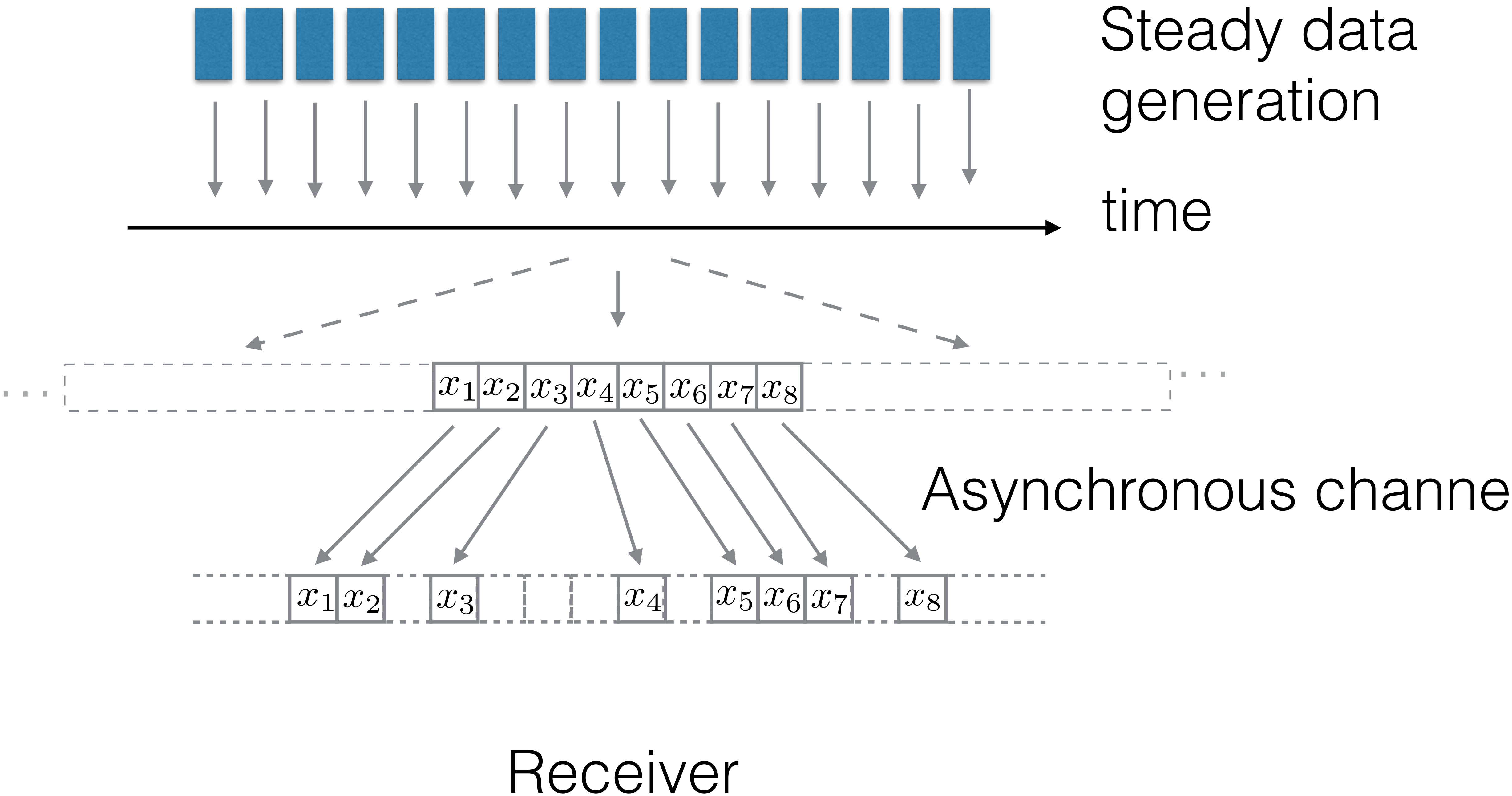
What is asynchronism?

"When timing of information transmission is unknown to the receiver."

Why important?

Because virtually all communication systems are a priori asynchronous.

Channel burstiness



Example: internet.

Channel burstiness: models

Insertion-deletion-substitution channel (Dobrushin 1967)

$$W(\mathbf{y}|x), \quad x \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}^\star$$

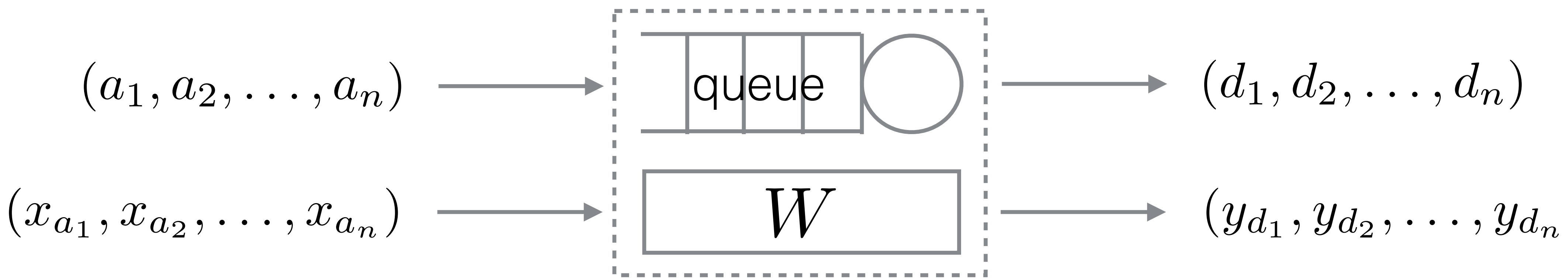
Channel coding theorem exists:

$$C = \lim_{n \to \infty} \frac{1}{n} \max_{X^n} I(X^n; \mathbf{Y}^n)$$

but no known single letter expression, even for the purely deletion channel (Kanoria-Montanari 2013).

assumes that positions are known

Timing channels (Anantharam-Verdù 1996)

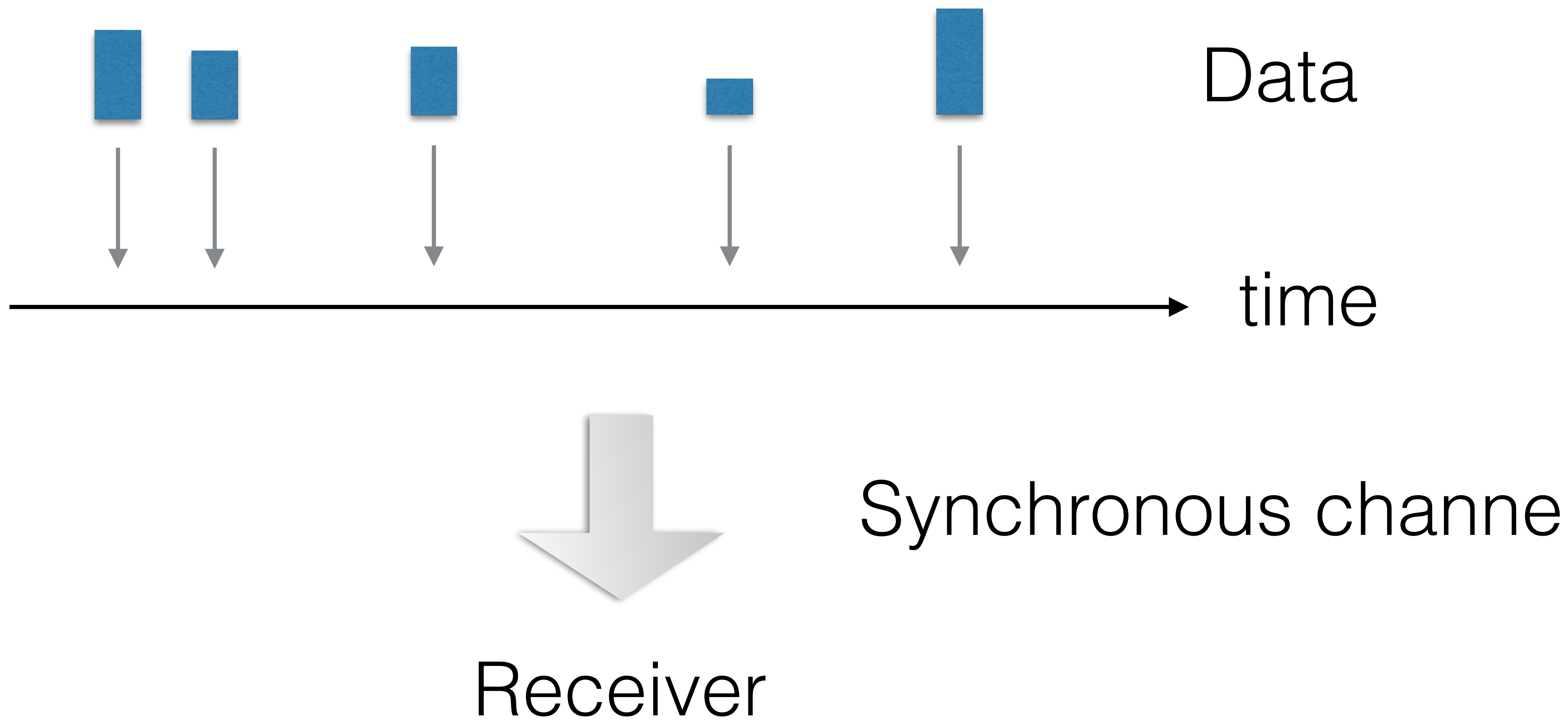


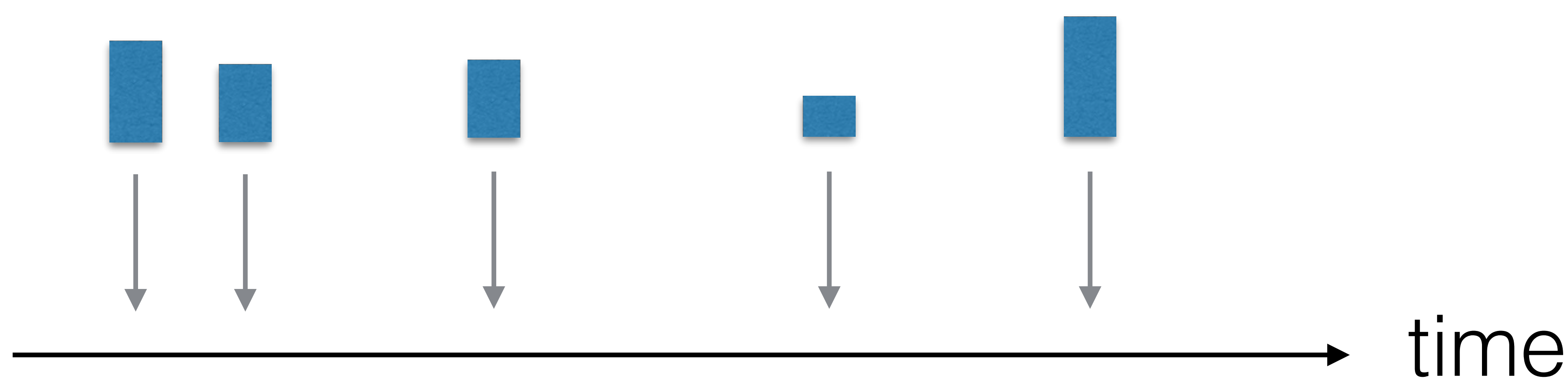
Timing information embedded into $a_{i+1} - a_i$ and $d_{i+1} - d_i$.

Well understood for exponential service, less so for other service times.

this tutorial will focus mostly on one type of asynchronism, here we briefly review other setups

Source burstiness

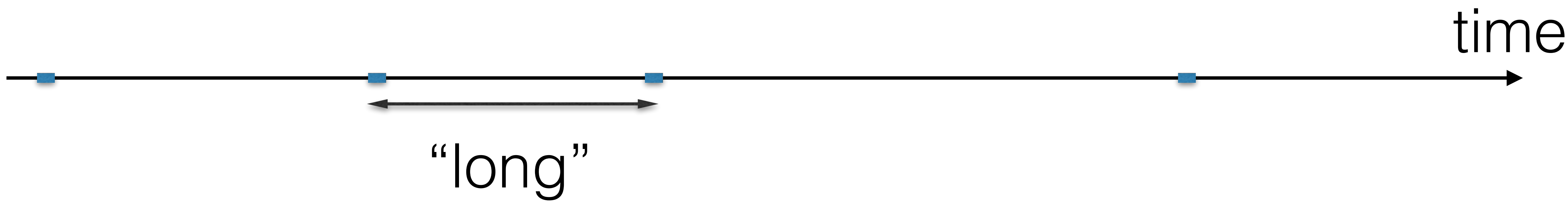




Small amounts of data because

- sources produces small amounts of data
- limited resources for transmission (energy harvesting systems, Ulukus et al. 2015)

This tutorial: source burstiness



- Packets sent once in a "long" while
- Fundamental limits (bits/channel use or /joule)?
- Efficient communication strategies?

Roadmap

- Preliminaries
- Detection
- Capacity & capacity per unit cost
- Receiver sampling constraint
- Finite length analysis

Preliminaries

Channel

- $W = \{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ DMC
- X input random variable to channel W
- Y output random variable to channel W

Hence if $X \sim P$ then $(X,Y) \sim PW = P(\cdot)W(\cdot|\cdot)$

Notation

$$W_a = W(\cdot|a) \quad Y_a \sim W_a$$

$$D(Y_a||Y_b) = D(W(\cdot|a)||W(\cdot|b)) = \sum_y W(y|a) \log \frac{W(y|a)}{W(y|b)}$$

Typicality

- empirical distribution: $\hat{P}_{x^n}(a) = \frac{1}{n}|\{i : x_i = a\}|$
- typical sequence: $x^n \in T(V)$ if $||\hat{P}_{x^n} - V||_1 = \left\{ \begin{array}{c} o(1) \\ \omega(1/\sqrt{n}) \end{array} \right.$
- if X^n i.i.d. P then $Pr(X^n \in T(P)) = 1 - o(1)$

$$Pr(X^n \in T(V)) \doteq 2^{-nD(V||P)}$$

where $D(V||P) = \sum_x V(x) \log \frac{V(x)}{P(x)}$

$$a_n \doteq b_n \equiv \log \frac{a_n}{b_n} \stackrel{n\to\infty}{\longrightarrow} 0$$

Typicality (cont.)

Hence, if (X^n, Y^n) i.i.d. PW

then $Pr((X^n, Y^n) \in T(J)) \doteq 2^{-nD(J||PW)}$

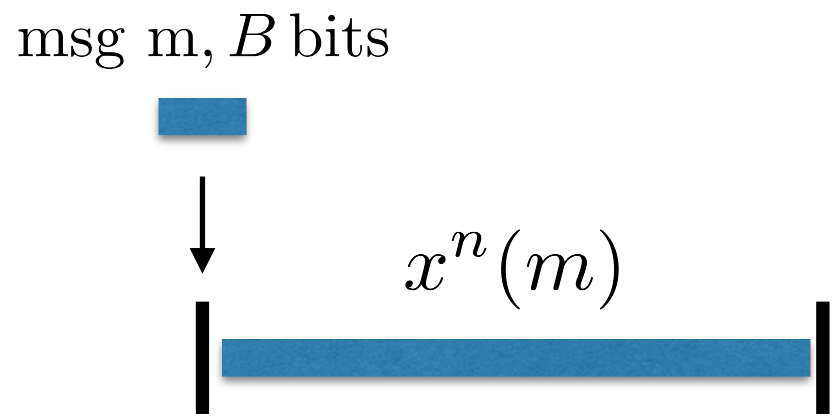
and if (X^n, Y^n) i.i.d. PW_a

then $Pr((X^n, Y^n) \in T(J)) \doteq 2^{-nD(J||P,W_a)}$

where

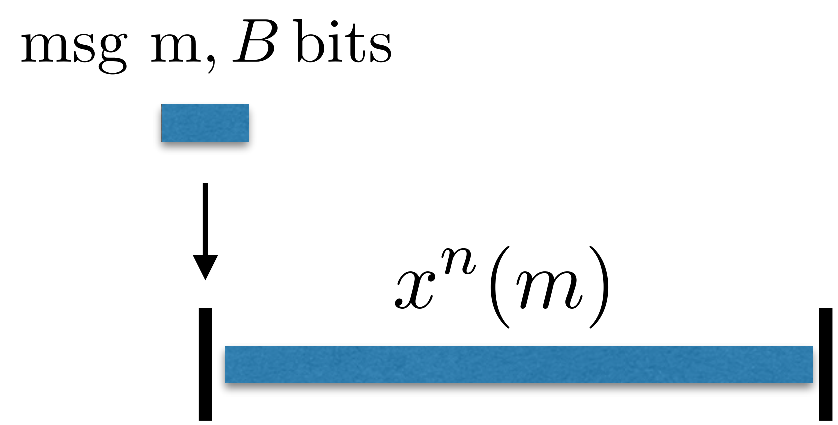
$$D(J||P, W_a) = \sum_{x,y} J(x,y) \log \frac{J(x,y)}{P(x)W(y|a)}$$

Energy-limited synchronous communication



- Input cost: $x \mapsto k(x) \in \mathbb{R}_+$
- Encoding: $m \mapsto x^n(m)$
- Cost for sending m: $k(x^n(m)) \stackrel{\text{def}}{=} \sum_{i=1}^{n} k(x_i(m))$
- Channel: $x \longrightarrow Y \sim W(\cdot|x)$
- Decoding: $\hat{m}(Y^n)$

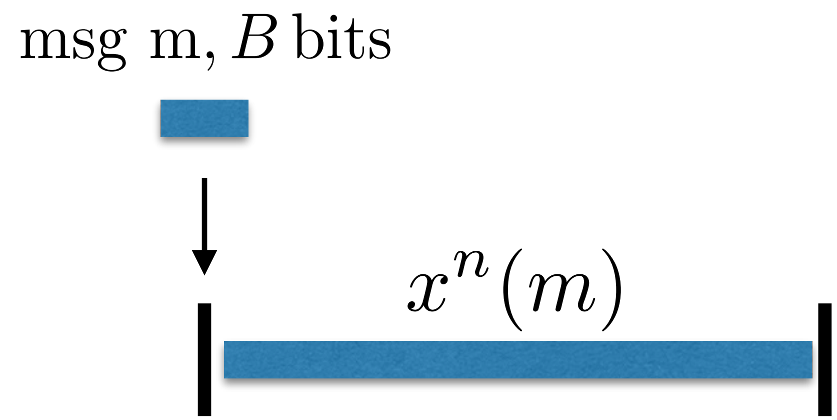
Energy-limited synchronous communication



$$\mathbf{R} = \frac{B}{\max_m k(x^n(m))} \text{ bits/joule}$$

$$\mathbf{R} \text{ achievable if } Pr(\hat{m} \neq m) \overset{n\to\infty}{\longrightarrow} 0$$

$$\mathbf{C}_{\text{sync.}} = \sup\{\text{achievable } \mathbf{R}\}$$



Theorem (Gallager, Verdú, late 80's)

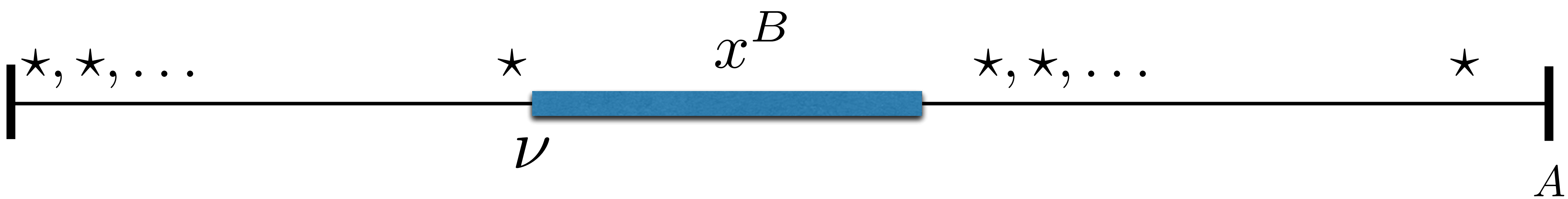
$$\mathbf{C}_{\text{sync.}} = \max_X \frac{I(X;Y)}{\mathbb{E}k(X)}$$

If there exists a zero cost symbol

$$\mathbf{C}_{\text{sync.}} = \max_x \frac{D(Y_x||Y_0)}{k(x)}$$

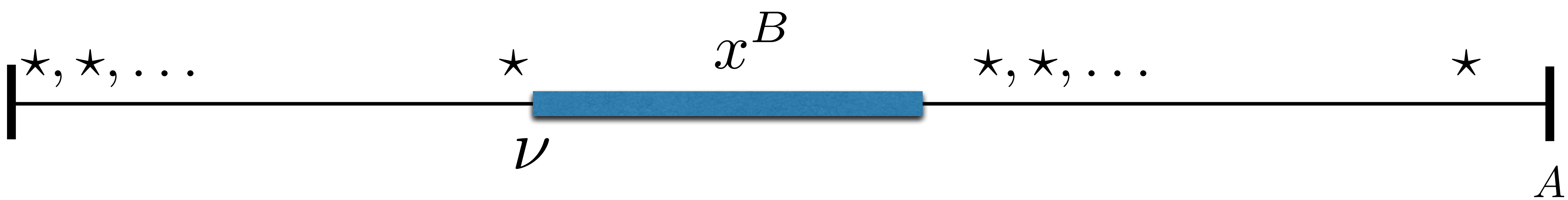
Detection

Detection



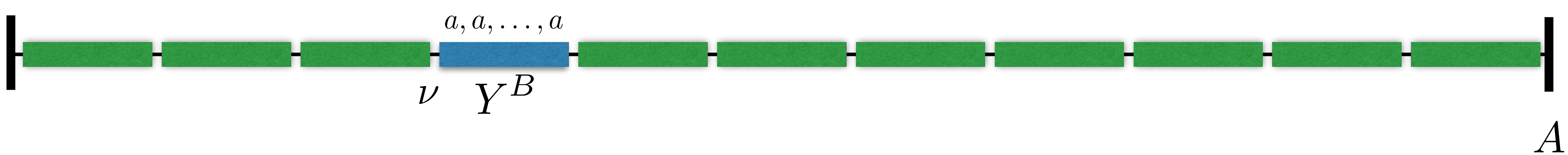
- Single message x^B
- Sent at random time $\nu \sim \{1, 2, \ldots, A\}$ across W
- Receiver: $Y^B \sim W_{x^B}$ over $\{\nu, \nu+1, \cdots, \nu+B-1\}$

  and $Y \sim W_\star$ otherwise.



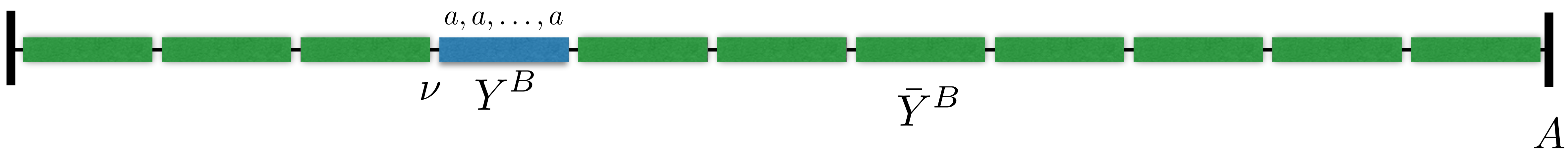
- Given Y^{A+B-1} produce estimate $\hat{\nu}(Y^{A+B-1})$ of ν
- Largest A for which $Pr(\hat{\nu} \neq \nu) \stackrel{B\to\infty}{\longrightarrow} 0$?
- Optimal x^B ?

A natural scheme



- Suppose slotted model, i.e., receiver knows $\nu \bmod B$
- Suppose $x^B = a, a, \ldots, a$
- Declare $\hat{\nu}$ such that $Y_{\hat{\nu}}^{\hat{\nu}+B-1}$ is W_a-typical

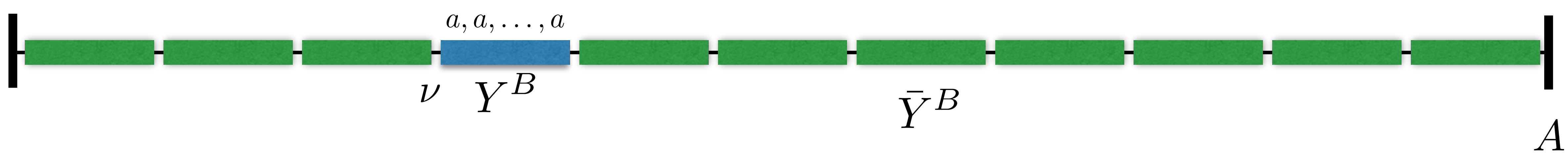
Analysis



$$P(\hat{\nu} \neq \nu) \leq P(\text{miss detection}) + P(\text{false-alarm})$$
$$= P(Y_{\nu}^{\nu+B-1} \notin T(W_a)) + (A/B)P(\bar{Y}^B \in T(W_a))$$

where $\bar{Y}^B$ is i.i.d. $\sim W_\star$

Analysis



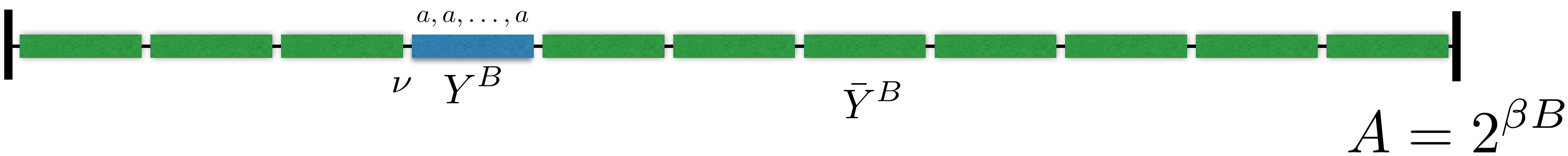
$$P(Y_\nu^{\nu+B-1} \notin T(W_a)) = o(1)$$

$$Pr(\bar{Y}^B \in T(W_a)) \doteq 2^{-BD(Y_a||Y_\star)}$$

$$\Rightarrow (A/B)P(\bar{Y}^B \in T(W_a)) = o(1)$$

$$\text{if } A = 2^{\beta B} \text{ with } \beta < D(Y_a||Y_\star)$$

Analysis



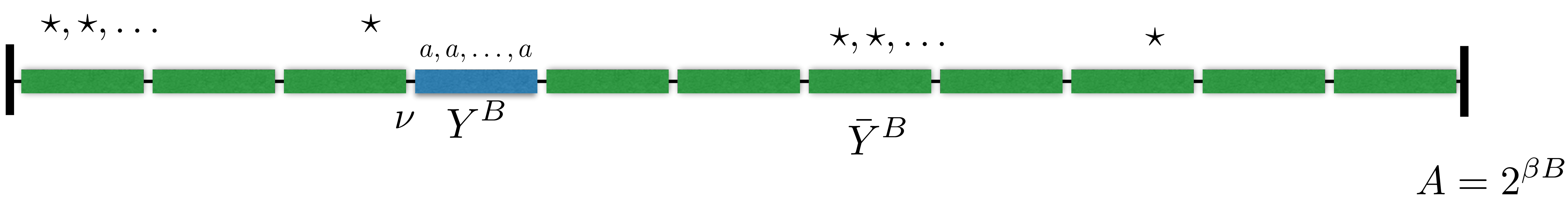
Hence $Pr(\hat{\nu} \neq \nu) = o(1)$ as $B \to \infty$

if $A = 2^{\beta B}$ with $\beta < D(Y_a || Y_\star)$

$$\text{largest } \beta : \beta_o = \max_a D(Y_a || Y_\star)$$

$$\text{which corresponds to } \bar{a} = \arg\max_a D(Y_a || Y_\star)$$

All observations?



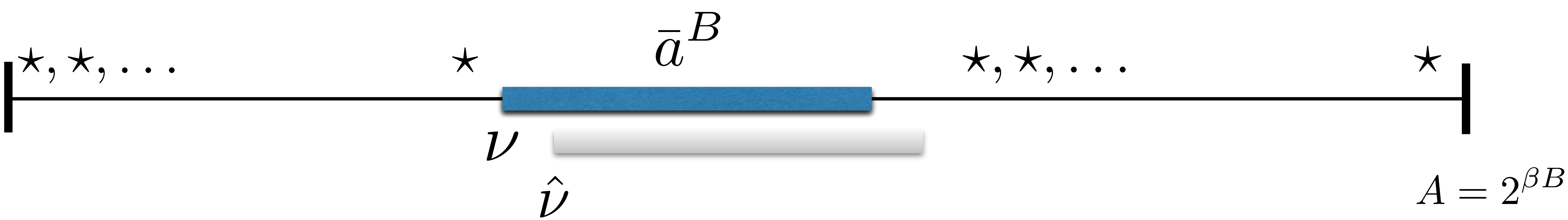
Sliding window sequential detection:

$$\tau = \min\{t \geq 1 : Y_t^{t+B-1} \in T(W_{\bar{a}})\}$$

achieves same performance as non-sequential.

Non-slotted model ?

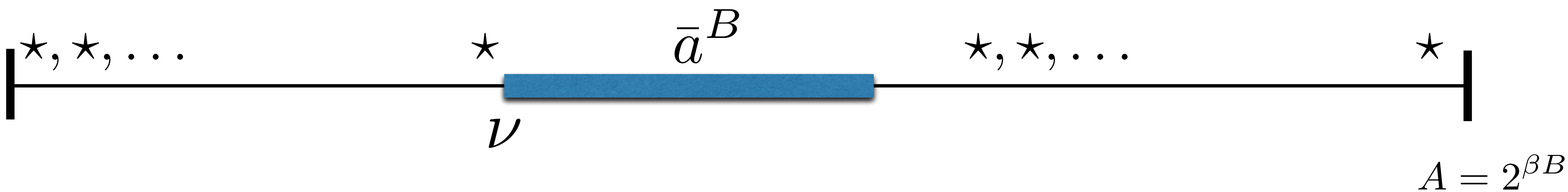
Slotted model guarantees “pure hypothesis”



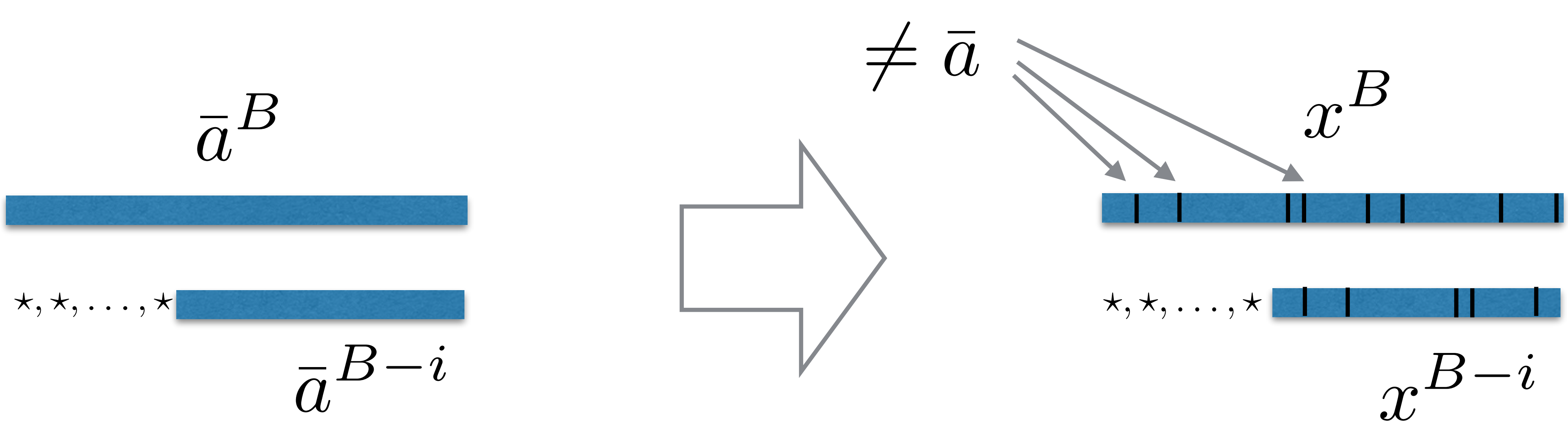
If non-slotted

$$Pr(|\hat{\nu} - \nu| \geq 1) = \Theta(1)\ !$$

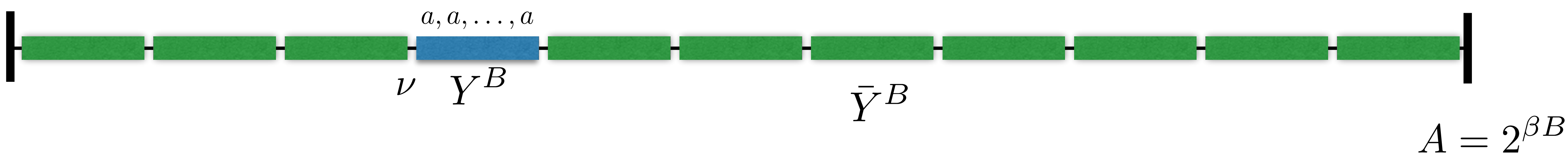
Non-slotted model?



Modify $\bar{a}^B$ slightly to get low autocorrelation;
Hamming distance with any of its shift is $\Theta(B)$.

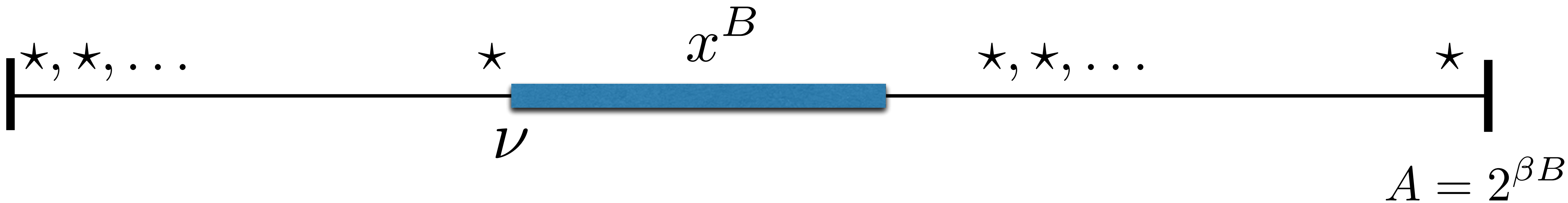


$\beta > \beta_o$: converse (intuition)



- Assume $x^B = a^B$
- Reveal $\nu \bmod B$
- Typicality decoding $\approx$ MAP (can be shown)
- (blue): $Pr(Y^B \in T(W_a)) = 1 - o(1)$
- (green): $Pr(\bar{Y}^B \in T(W_a)) \doteq 2^{-BD(Y_a||Y_\star)}$
- $Pr(\text{false-alarm}) \doteq 2^{-B(D(Y_a||Y_\star)-\beta)}$
- Non constant x^B, no gain (can be shown).

Summary



Theorem (Chandar-T-Wornell 2008):

$$Pr(\hat{\nu} \neq \nu) = o(1) \text{ as } B \to \infty$$

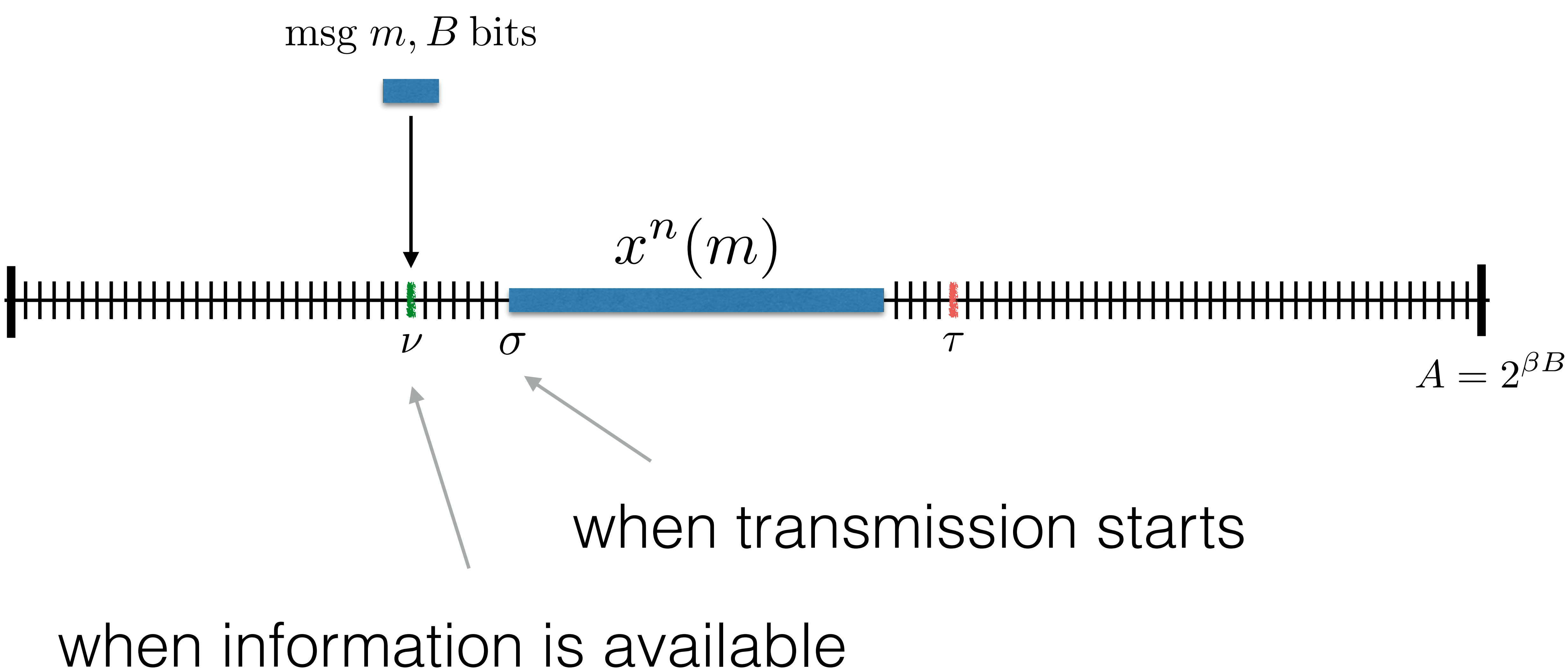
$$\text{iff } A = 2^{\beta B} \text{ with } \beta < \max_a D(Y_a || Y_\star)$$

Optimal rule:

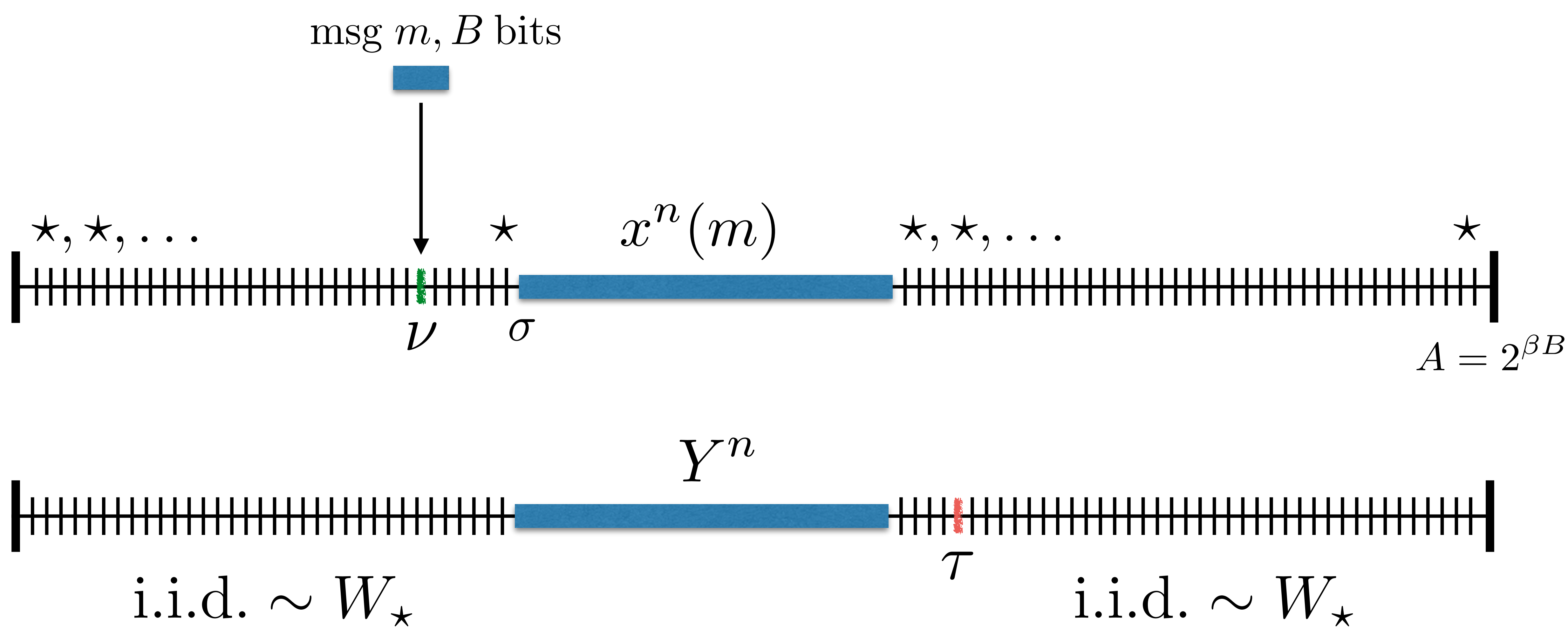
$$\tau = \min\{t \geq 1 : Y_t^{t+B-1} \in T(W_{x^B})\}$$

Information transmission

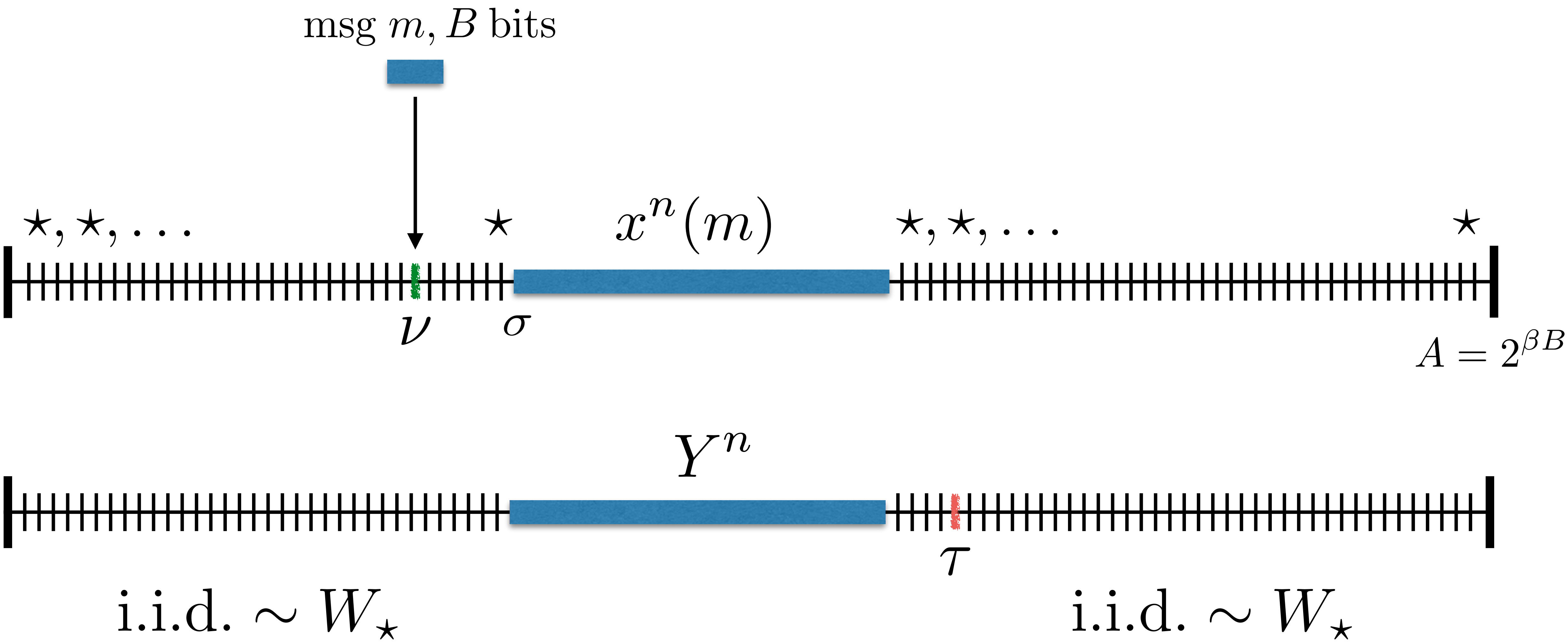
Bursty communication model



Information transmission



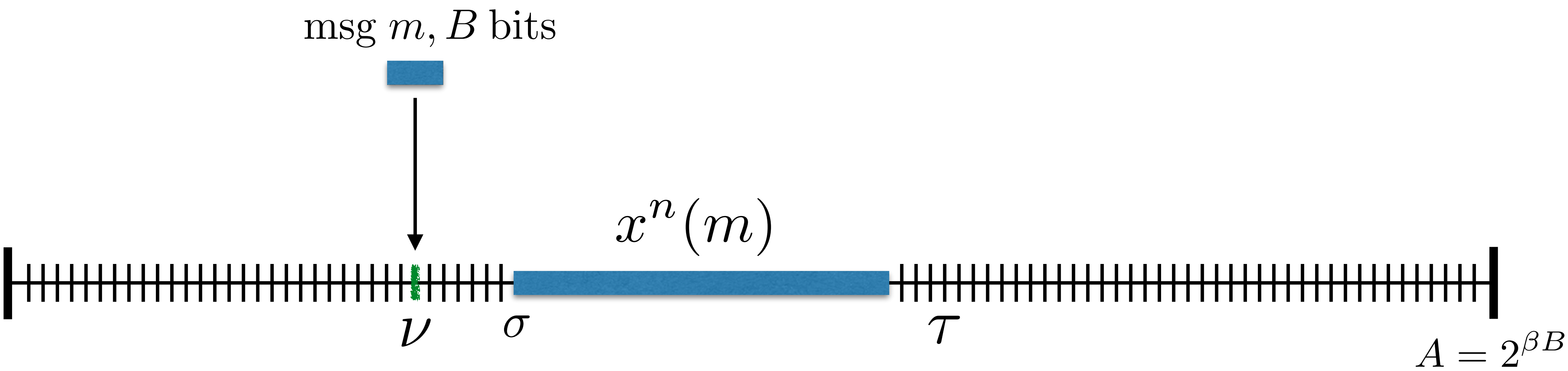
Delay constraint (for a meaningful problem):
given d we want $Pr(\tau - \nu \leq d) = 1 - o(1)$.



$$C(A, d)?$$

How does it compare with $C = \max_X I(X;Y)$?

Small delay capacity

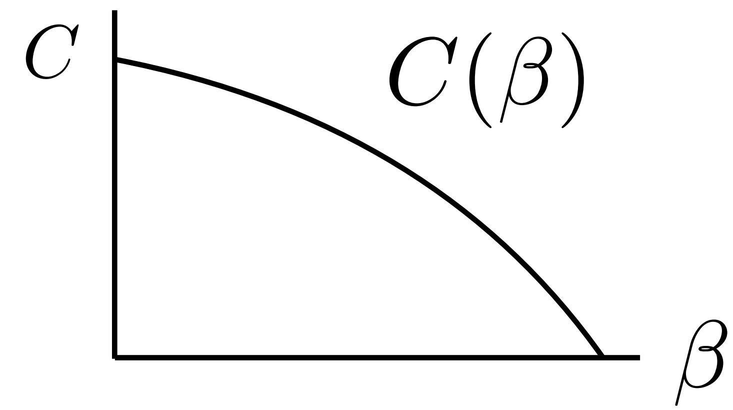


Theorem (Chandar-T-Tse 2013):

- $$C(\beta, d = 2^{o(B)}) = \max_X \min \left\{ I(X;Y), \frac{D(XY||X, Y_\star)}{1+\beta} \right\}$$
- $$d_{\min}(\beta) = \frac{B}{\max_{X \in \mathcal{P}_\beta} I(X;Y)}$$

where $\mathcal{P}_\beta$ is the set of inputs that achieve $C(\beta, \delta = 0)$.

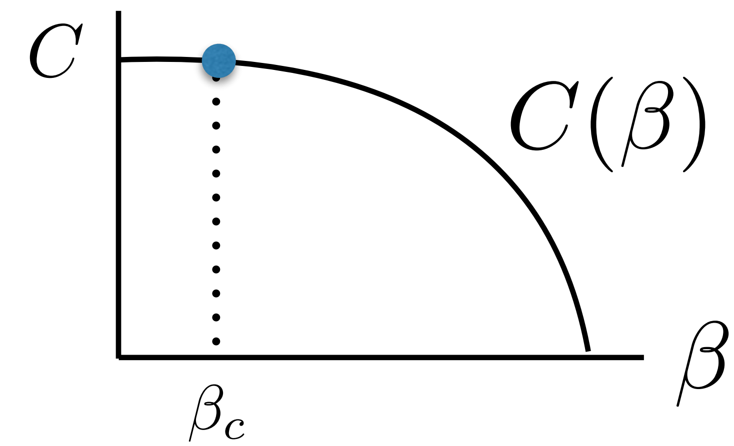
Remark



$$C(\beta, d = 2^{o(B)}) = \max_X \min \left\{ I(X;Y), \frac{D(XY \| X, Y_\star)}{1+\beta} \right\}$$

$$\Rightarrow C(\beta, d = 2^{o(B)}) \le C$$

Remark (cont.)



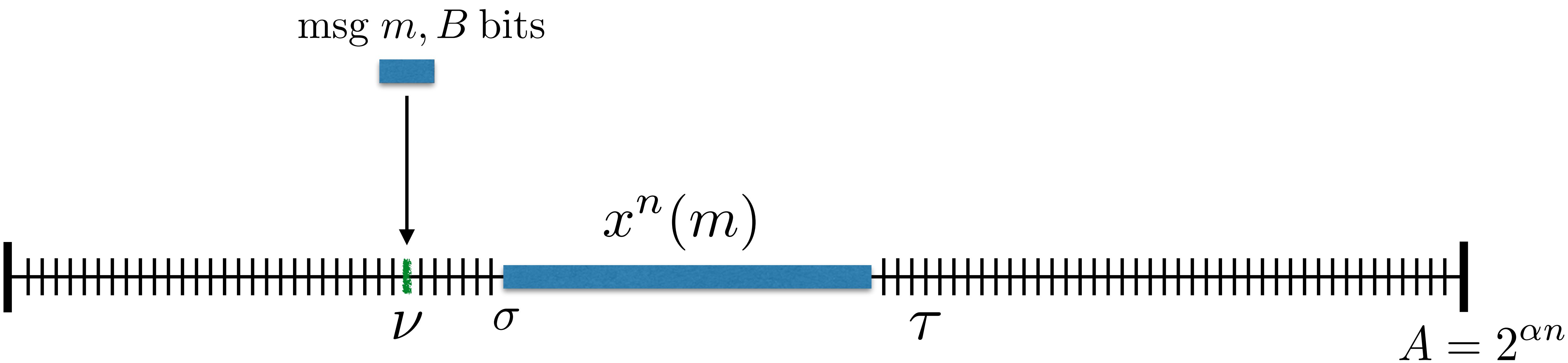
$$C(\beta, d = 2^{o(B)}) = \max_X \min \left\{ I(X;Y), \frac{D(XY\|X, Y_\star)}{1+\beta} \right\}$$

$$\beta_c = \sup\{\beta : C(\beta) = C\} > 0$$

iff (unique) capacity achieving output $\tilde{Y}$ of sync channel differs from $Y_\star$ since

$$\begin{aligned} D(XY\|X, Y_\star) &= I(X;Y) + D(\tilde{Y}\|Y_\star) \\ &= C + \underbrace{D(\tilde{Y}\|Y_\star)}_{>0} \end{aligned}$$

Small delay capacity: alternative scaling

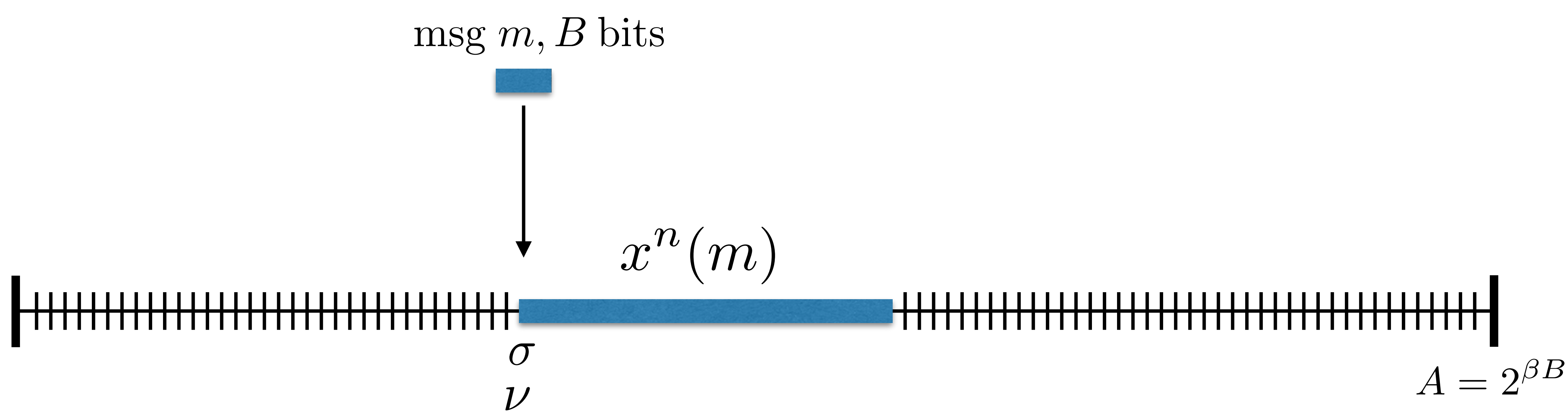


$$A = 2^{\beta B} = 2^{\alpha n} \Leftrightarrow \alpha = \beta R$$

Theorem (Chandar-T-Tse 2013):

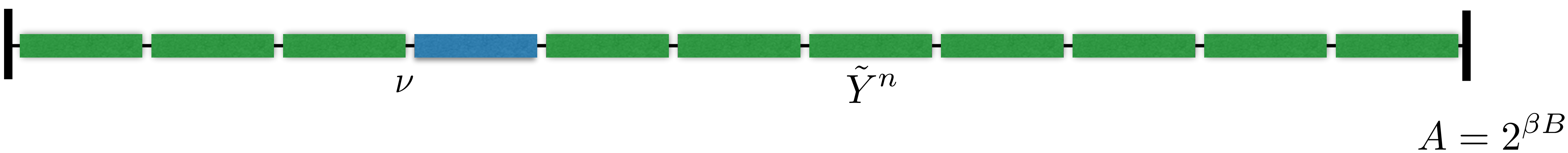
- $$C(\alpha, d = 2^{o(n)}) = \max_{X : D(Y \| Y_\star) \geq \alpha} I(X;Y)$$
- $$d_{\min}(\alpha) = n$$

Achievability



- $\{x^n(m)\}$ i.i.d. P, $\sigma = \nu$
- $\tau = \min\{t \geq 1 : \exists m \text{ s.t. } (x^n(m), Y_{t-n+1}^t) \in T(PW)\}$

Simplified analysis (slotted)



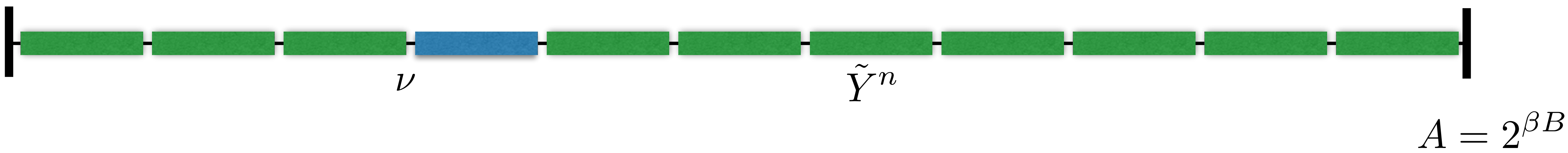
Assume $m = 1$ is sent

$$P((X^n(1), Y^n) \in T(PW)) = 1 - o(1)$$

$$P((X^n(2), Y^n) \in T(PW)) \doteq 2^{-nD(XY||X,Y)}$$

$$P((X^n(2), \tilde{Y}^n) \in T(PW)) \doteq 2^{-nD(XY||X,Y_\star)}$$

Simplified analysis (slotted)



Hence $Pr(\text{error}) \to 0$ if

$$2^{nR} 2^{-nD(XY\|X,Y)} = o(1)$$

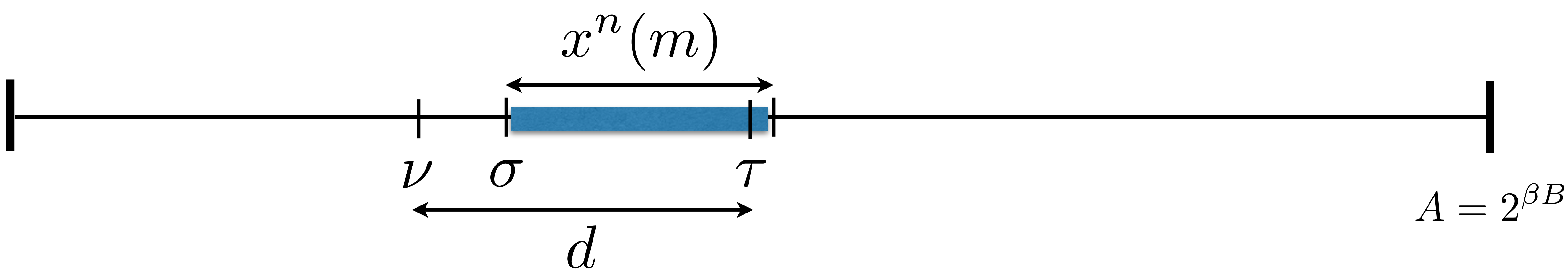
$$(A/n) 2^{nR} 2^{-nD(XY\|X,Y_\star)} \doteq 2^{-n(D(XY\|X,Y_\star) - R(1+\beta))} = o(1)$$

i.e. $R < D(XY\|X,Y) = I(X;Y)$

$R(1+\beta) < D(XY\|X,Y_\star)$ #

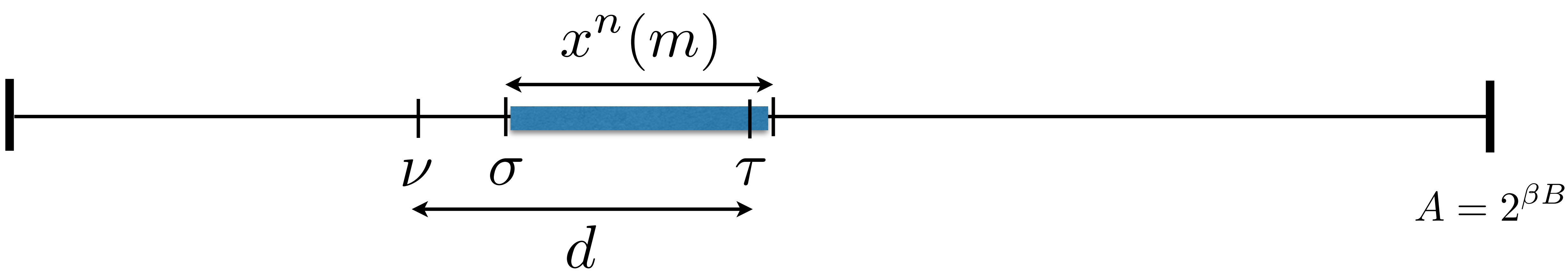
$$\Rightarrow C(\beta, \delta = 0) \geq \max_X \min \left\{ I(X;Y), \frac{D(XY\|X,Y_\star)}{1+\beta} \right\}$$

Converse: intuition (small delay, $\delta = 0$)

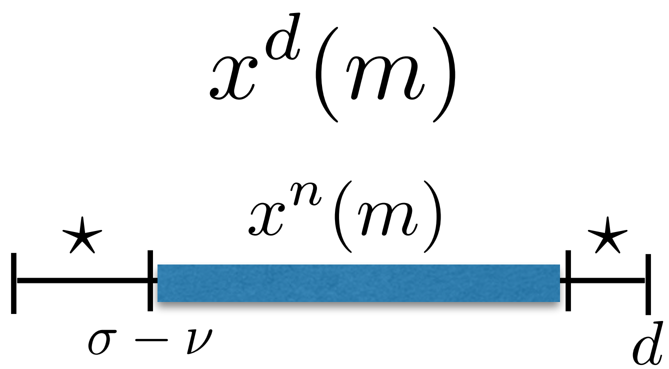


- Effective output process: pure noise from $\nu + d$ onwards
- Differs from real output process when $\sigma + n > \tau + d$
- Reveal effective output process to receiver

Converse: intuition ($\delta = 0$)

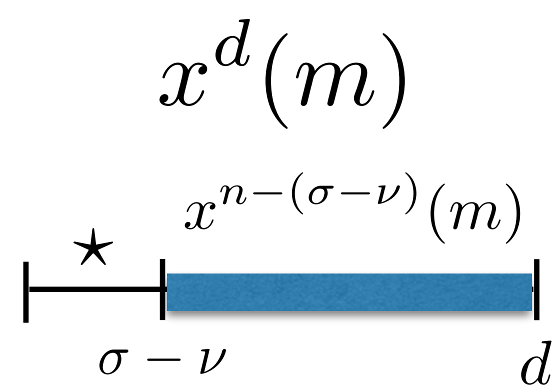


Consider extended codewords: symbols sent from ν to $\nu + d$



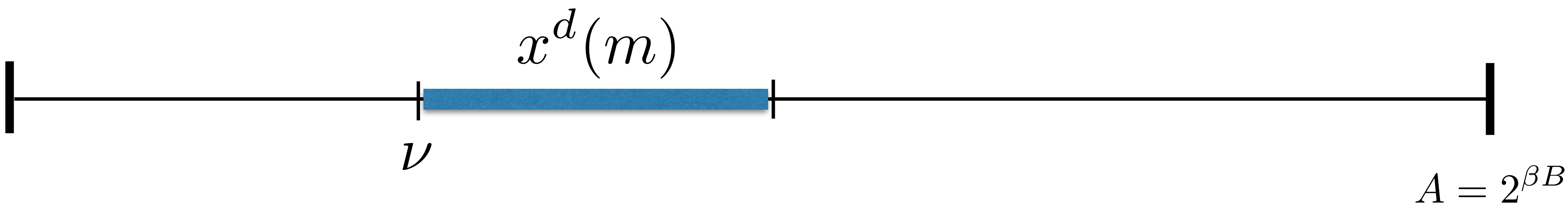
$$\nu + d \geq \sigma + n$$

or



$$\nu + d < \sigma + n$$

Converse: intuition ($\delta = 0$)



Without essential loss of generality assume $\{x^d(m)\}$ is of constant composition P.

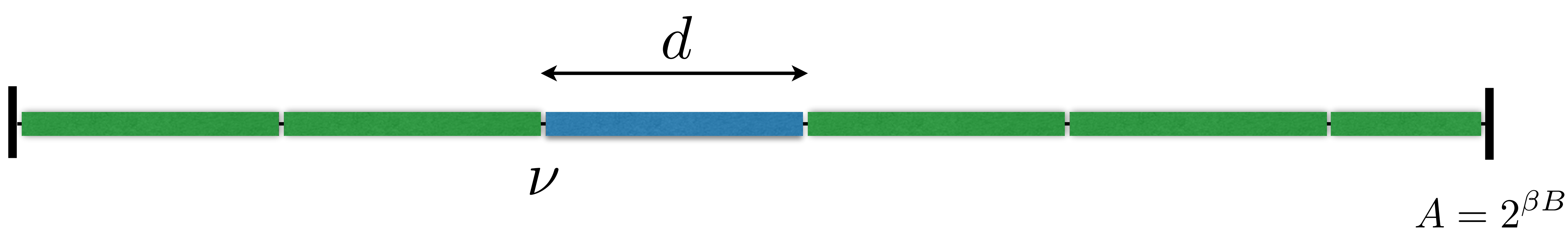
From synchronous communication:

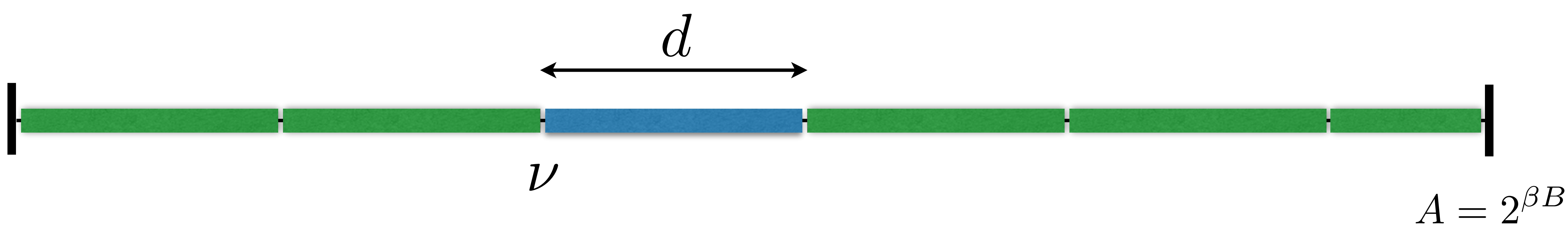
$$\tilde{R} = \frac{\log M}{d} \leq I(X;Y) \quad \text{with } X \sim P$$

Converse (cont.)

Decoding delay implies the receiver can locate codeword

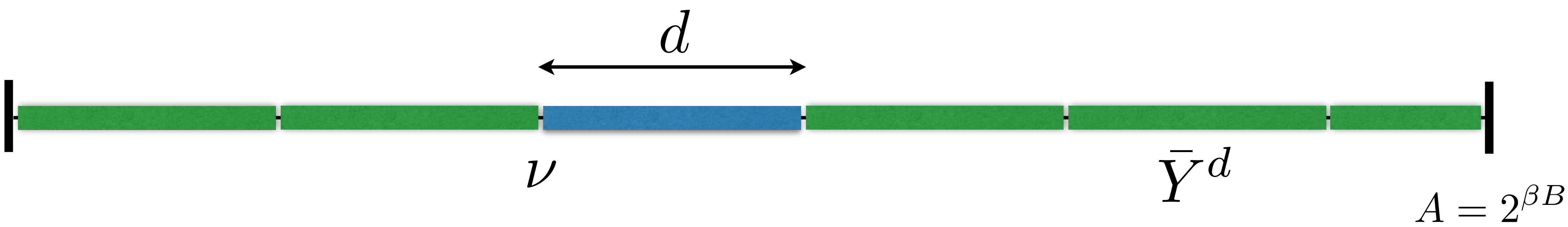
- Reveal effective output process to receiver
- Reveal $\nu \bmod d$
- Ask receiver to detect and isolate sent codeword





Typic-decoding is essentially optimal (can be shown).

Typic-decoding error probability



Assuming $m = 1$ is sent:

$$P((X^d(1), Y^d) \in T(PW)) = 1 - o(1)$$

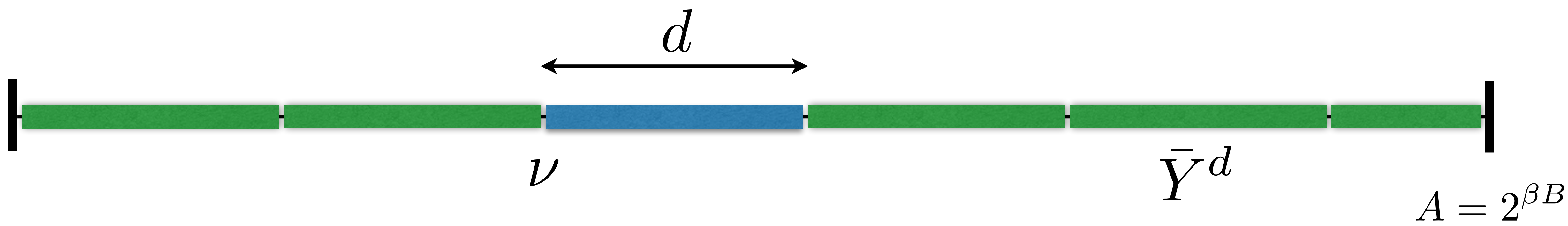
$$P((X^d(2), Y^d) \in T(PW)) \doteq 2^{-dD(XY||X,Y)}$$

$$P((X^d(2), \bar{Y}^d) \in T(PW)) \doteq 2^{-dD(XY||X,Y_\star)}$$

Union bound over msg: $\tilde{R} \leq D(XY||X,Y) = I(X;Y)$

Union bound over msg/slots: $\tilde{R}(1+\beta) \leq D(XY||X,Y_\star)$

Typic-decoding error probability



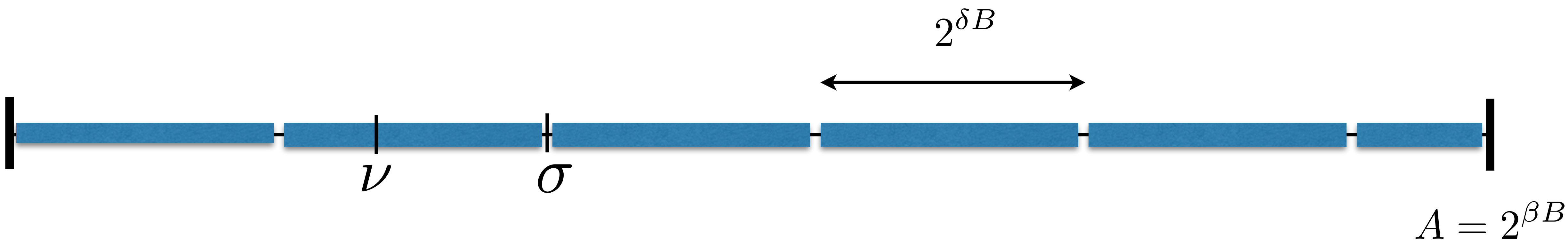
$$\tilde{R} \leq D(XY||X,Y) = I(X;Y)$$

$$\tilde{R}(1+\beta) \leq D(XY||X,Y_\star)$$

$$\Leftrightarrow \tilde{R} \leq \min\left\{ I(X;Y), \frac{D(XY||X,Y_\star)}{1+\beta} \right\}$$

$$\Rightarrow C(\beta, \delta = 0) \leq \max_X \min\left\{ I(X;Y), \frac{D(XY||X,Y_\star)}{1+\beta} \right\} \quad \#$$

Large delay capacity $d = 2^{\delta B}$



Theorem (Chandar-T-Tse 2013):

$$C(\beta, \delta) = C(\beta - \delta, 0) \text{ for any } \beta \geq \delta > 0$$

Proof: reduce asynchronism by delaying transmission

Energy to transmit asynchronously

Capacity per unit cost

Theorem (Chandar-T-Tse 2013):

- $$\mathbf{C}(\beta, \delta = 0) = \max_X \min \left\{ \frac{I(X;Y)}{\mathbf{E}k(X)}, \frac{D(XY \| X, Y_\star)}{(1+\beta)\mathbf{E}k(X)} \right\}$$
- $$d_{\min}(\beta) = \frac{B}{\max_{X \in \mathcal{P}_\beta} I(X;Y)}$$

where $\mathcal{P}_\beta$ is the set of inputs that achieve $\mathbf{C}(\beta, \delta = 0)$.

- $\mathbf{C}(\beta, \delta) = \mathbf{C}(\beta - \delta, 0) \quad \beta \geq \delta \geq 0$

Proof: follow capacity arguments.

A natural case: $k(\star) = 0$

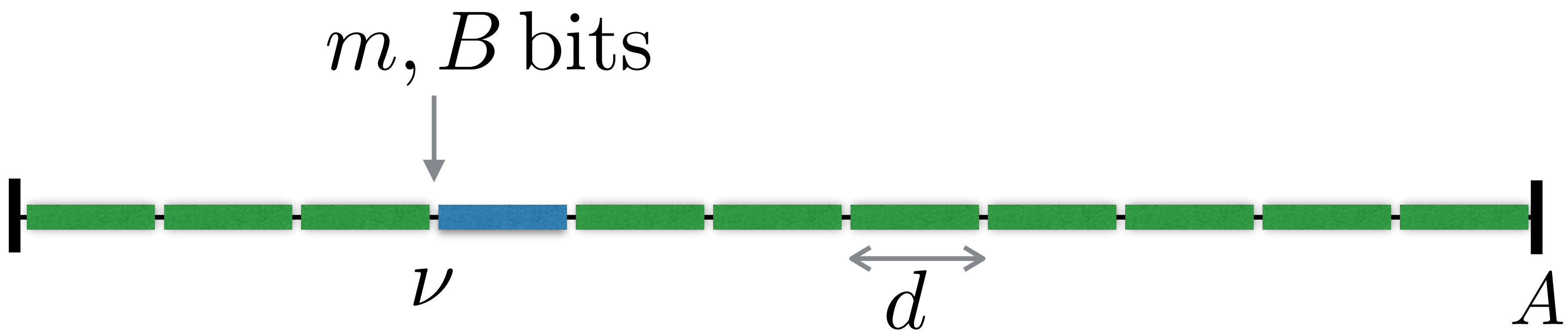
$$\mathbf{C}(\beta, \delta = 0) = \frac{1}{1+\beta} \max_x \frac{D(Y_x || Y_\star)}{k(x)}$$
$$= \frac{1}{1+\beta} \mathbf{C}$$

Example: Gaussian channel

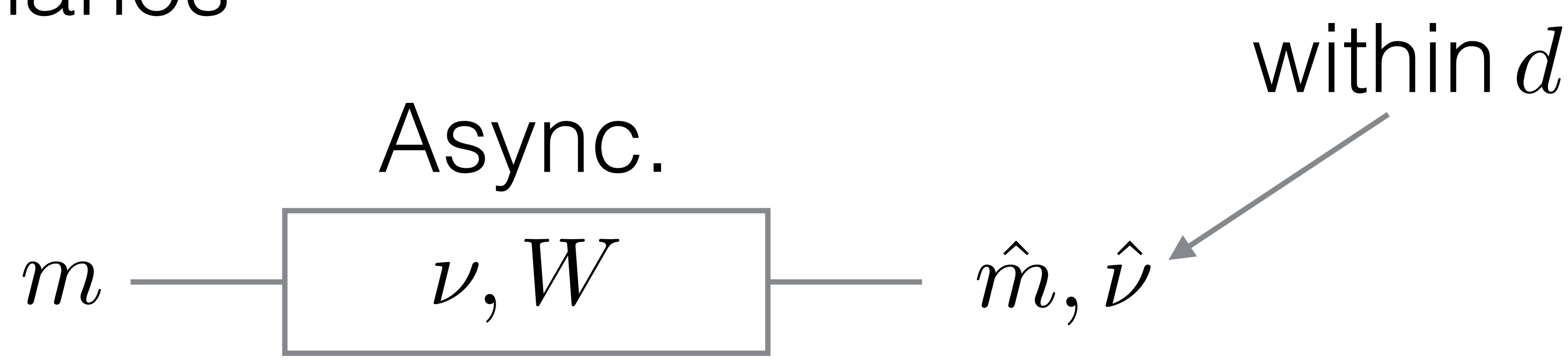
- $x \to x + Z(0, N_o/2)$
  $k(x) = x^2, \; \star = 0$

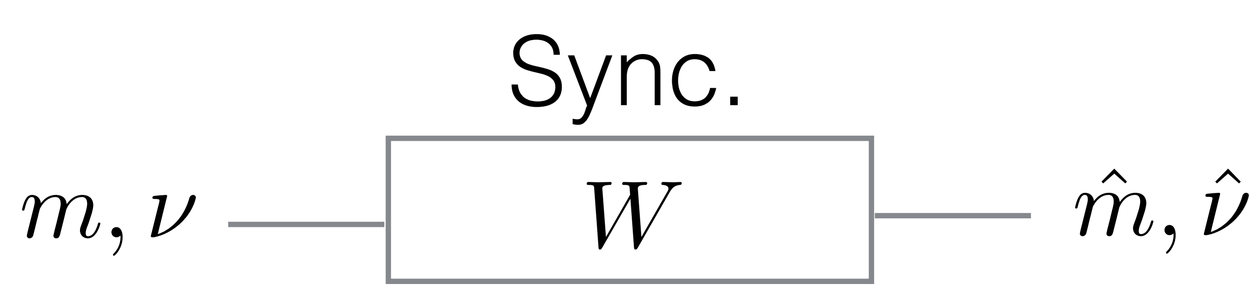
$$\mathbf{C}(\beta) = \frac{\log_2 e}{N_o(1+\beta)}$$

Alternative "proof" for $\mathbf{C}(\beta, \delta = 0)$ when $k(\star) = 0$



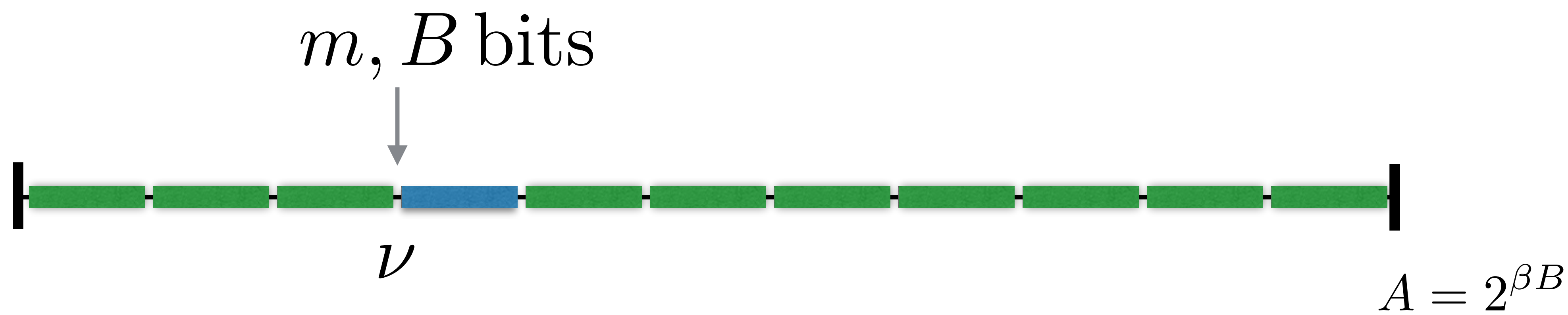
Equivalent scenarios





PPM

Alternative “proof” for $\mathbf{C}(\beta)$ when $k(\star) = 0$



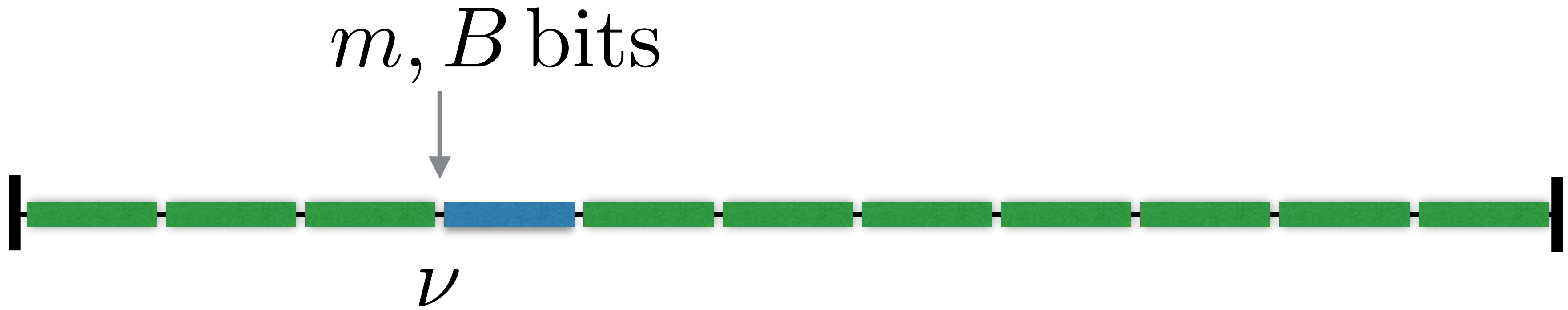
Since $k(\star) = 0$:

cost to transmit B bits asynchronously

$$=$$

cost to transmit $\log(A/d) + B \doteq \beta B + B$ bits synchronously via PPM.

Alternative “proof” for $\mathbf{C}(\beta)$ when $k(\star) = 0$



Hence,

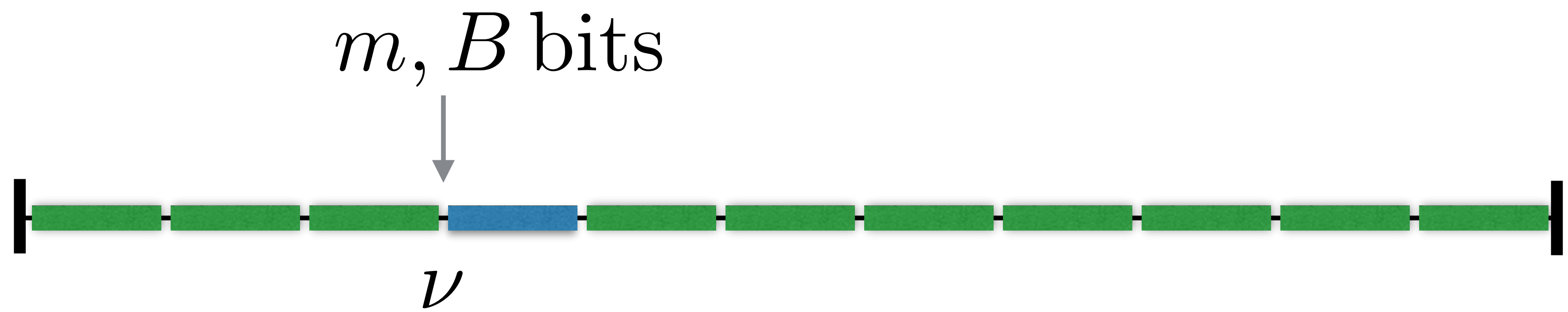
minimum cost to transmit 1 bit asynchronously

$$\geq$$

minimum cost to transmit $1 + \beta$ bits synchronously, i.e.

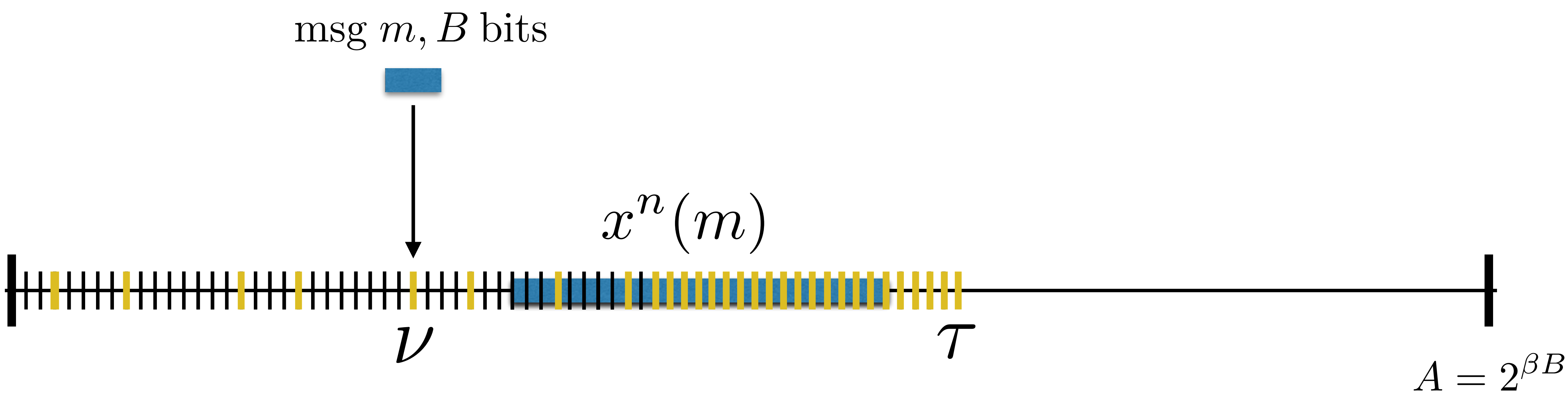
$$\frac{1}{\mathbf{C}(\beta)} \geq (1+\beta)\frac{1}{\mathbf{C}} \Leftrightarrow (1+\beta)\mathbf{C}(\beta) \leq \mathbf{C}$$

Alternative "proof" for $\mathbf{C}(\beta)$ when $k(\star) = 0$



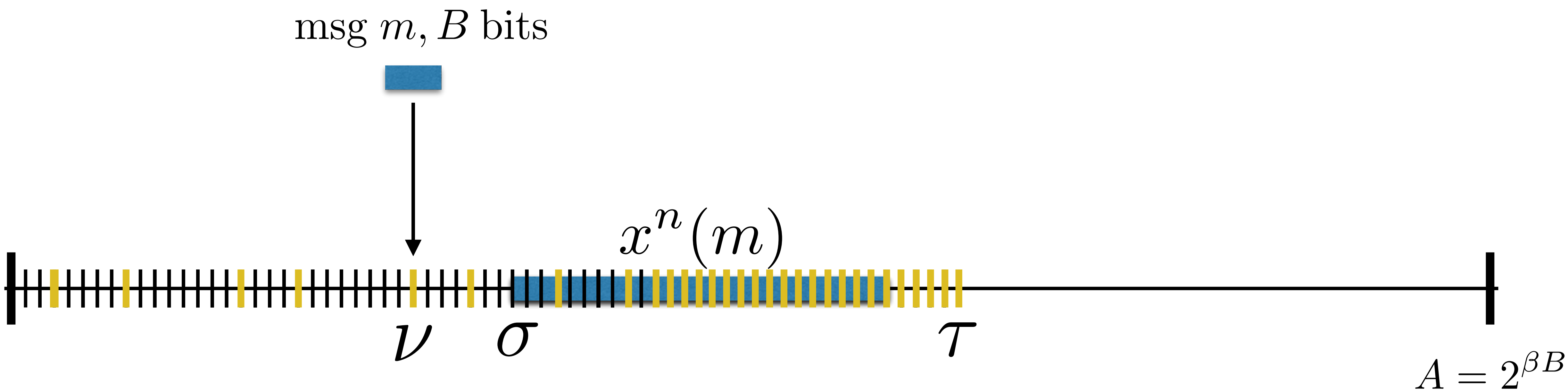
Converse: optimality of PPM $\Rightarrow (1+\beta)\mathbf{C}(\beta) \geq \mathbf{C}(\beta)$ #

Energy to transmit and <u>receive</u> asynchronously



- Transmission cost: $k(x)$
- Reception cost: sampling rate

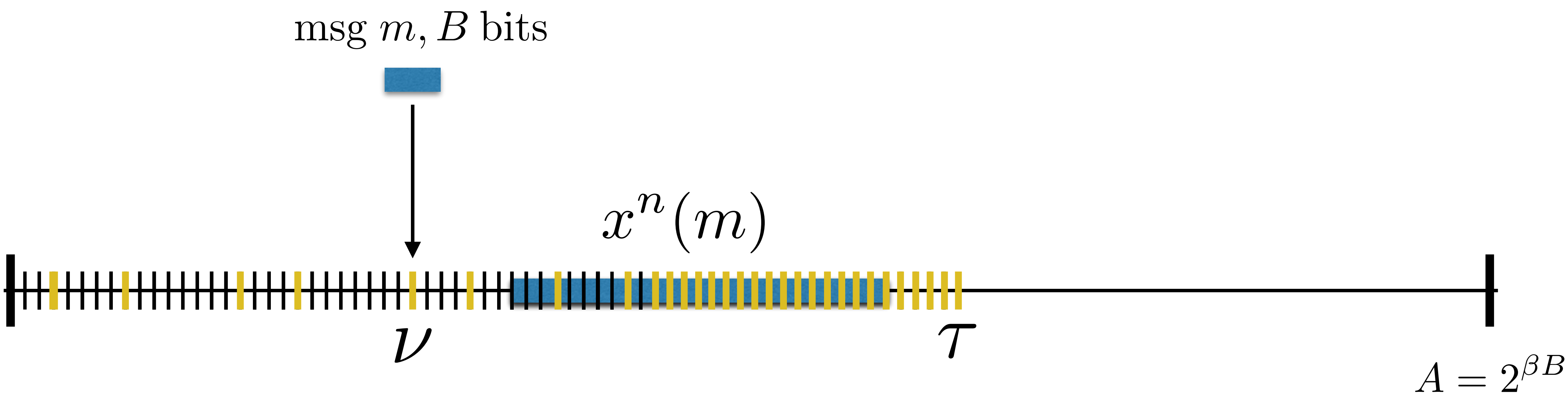
Motivation: power at receivers scales $\approx$ linearly with ρ.



- Transmitter: $x^n(m)$ at time σ
- Receiver: sampling, stop, decode

$$\rho = \frac{\text{number of } | \text{ until } \tau}{\tau}$$

Receiver operates at rate 30% if $Pr(\rho \leq 30\%) = 1 - o(1)$



$$\mathbf{C}_{\text{async.}}(\beta, \rho)?$$

Theorem (non-adaptive sampling, T-Chandar-Caire 2014):

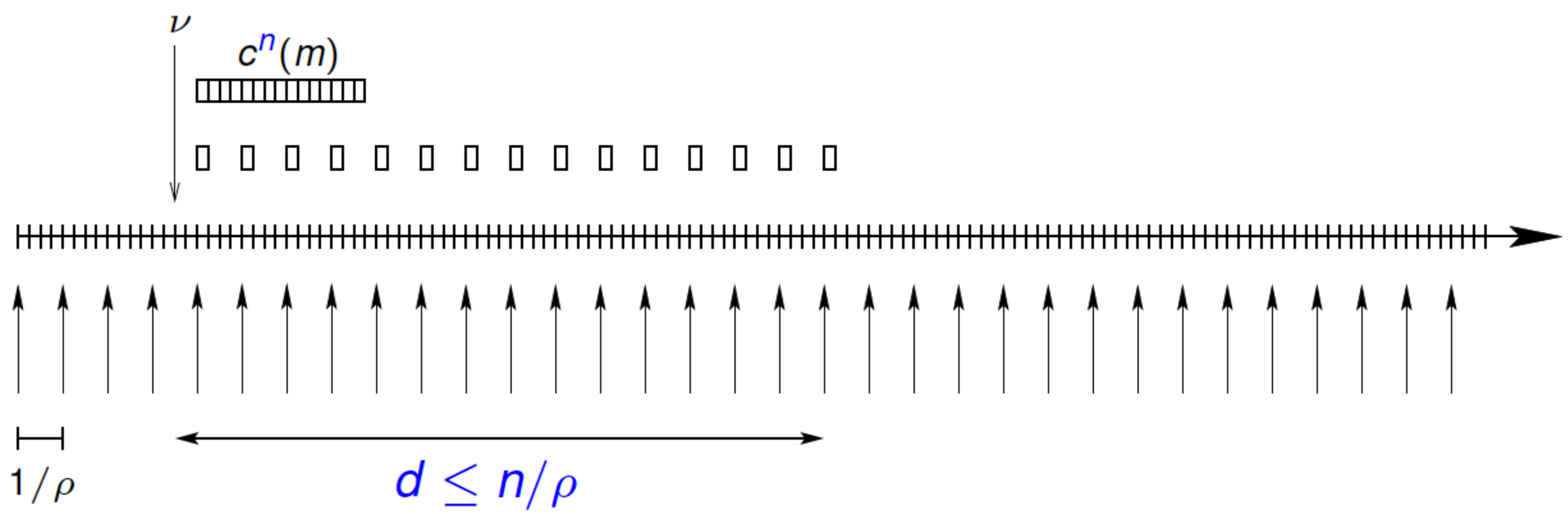
For any $\beta > 0$ and $0 < \rho < 1$

- $$\mathbf{C}_{\mathrm{async.}}(\beta, \rho) = \mathbf{C}_{\mathrm{async.}}(\beta, 1)$$

  $$d_{\min}(\beta, \rho) = \frac{1}{\rho} d_{\min}(\beta, 1)$$

Achievability (asynchronism reduction)

- consider coding scheme under full sampling
- stretch codewords by $1/\rho$, sample each $1/\rho$ output



Converse

Non-adaptive sampling budget ρA implies delay $1/\rho$.
(change of measure argument).

Theorem (adaptive sampling, T-Chandar-Caire 2014):

For any $\beta \geq 0$ and $0 < \rho \leq 1$

- $\mathbf{C}_{\text{async.}}(\beta, \rho) = \mathbf{C}_{\text{async.}}(\beta, 1)$

  $d_{\min}(\beta, \rho) = d_{\min}(\beta, 1)(1 + o(1))$

Minimum input energy,

minimum delay,

arbitrarily small sampling rate.

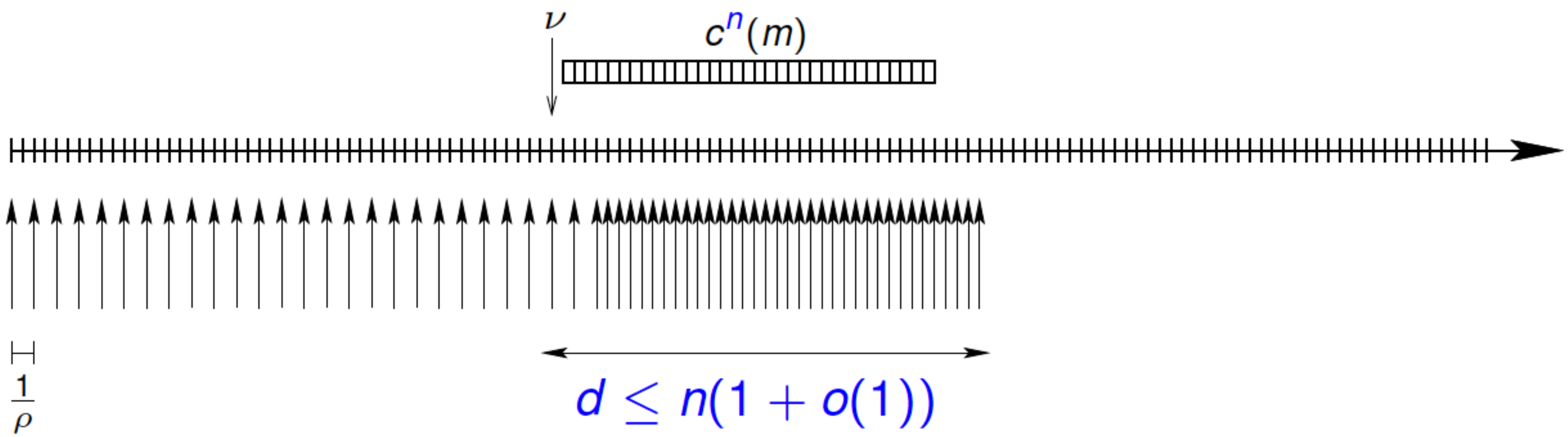
Achievability

Transmitter: random coding, $\sigma = \nu$ ✓

Receiver:
- decoding function ✓
- stopping rule
- sampling strategy

?

Two-phase scheme



- Sparse sampling: at multiples of $1/\rho$ compute empirical distribution $\hat{P}$ of the last $\rho\log(n)$ samples.
- If $\mathbb{P}_\star(T(\hat{P})) \leq \varepsilon$ switch to full sampling for period n
- At the end of the full sampling period, typic-decode or return to sparse mode if no codeword is found.

Theorem (adaptive sampling, T-Chandar-Caire 2014):

For any $\beta \geq 0$ and $0 < \rho \leq 1$

- $\mathbf{C}_{\text{async.}}(\beta, \rho) = \mathbf{C}_{\text{async.}}(\beta, 1)$

  $d_{\min}(\beta, \rho) = d_{\min}(\beta, 1)(1 + o(1))$

No loss for constant sampling rate but what if

$\rho \to 0$?

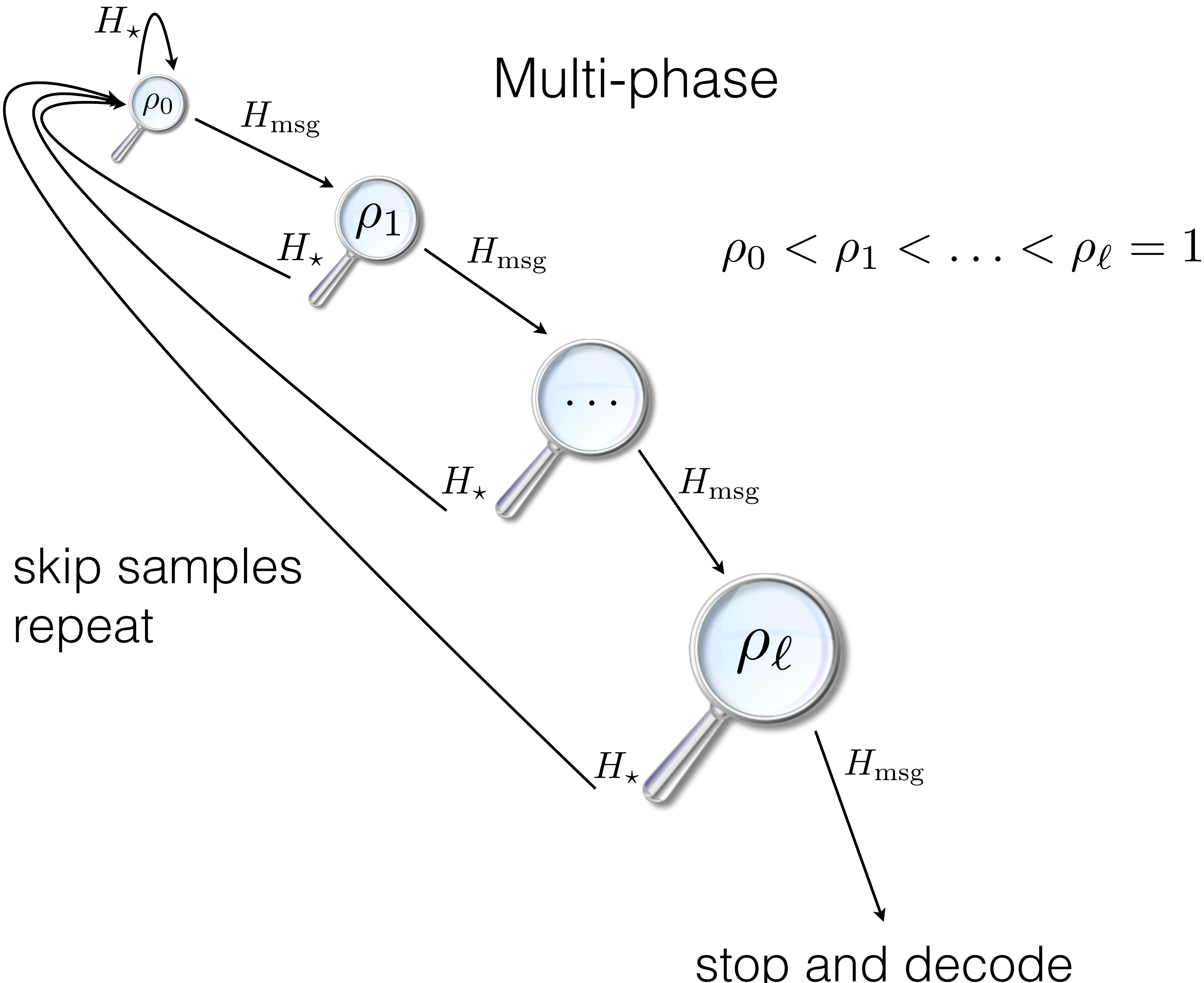
Theorem (min,min,min Chandar-T 2016):

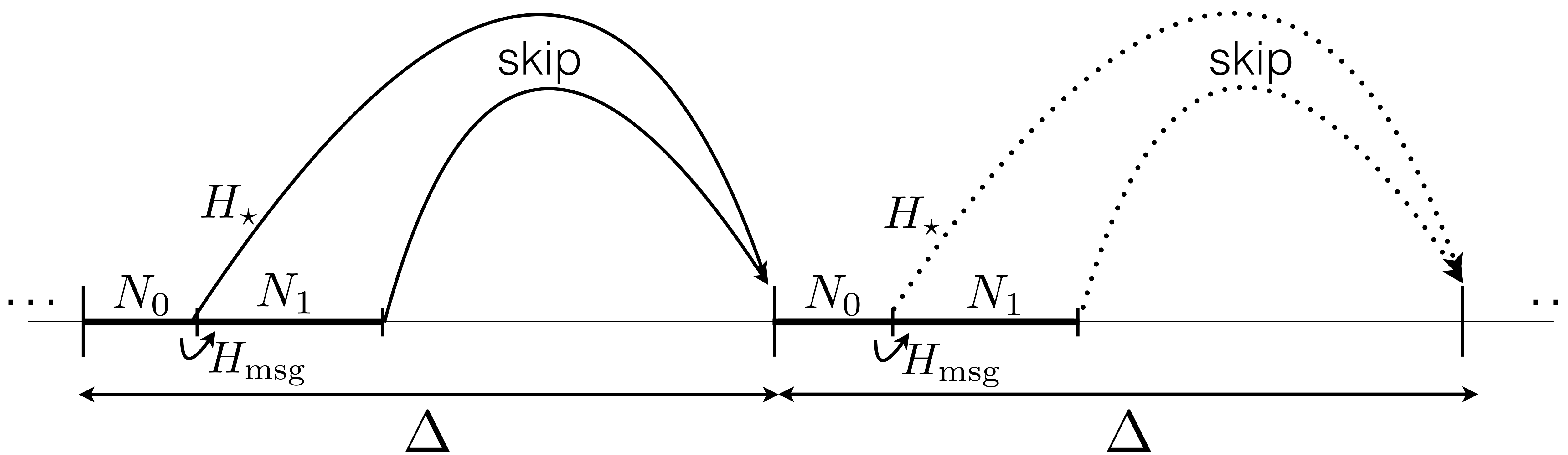
For any $\beta \geq 0$ and $0 < \rho \leq 1$

- if $\rho = \omega(1/B)$ $\quad\quad$ $\mathbf{C}_{\text{async.}}(\beta, \rho) = \mathbf{C}_{\text{async.}}(\beta, 1)$

  $d_{\min}(\beta, \rho) = d_{\min}(\beta, 1)(1 + o(1))$

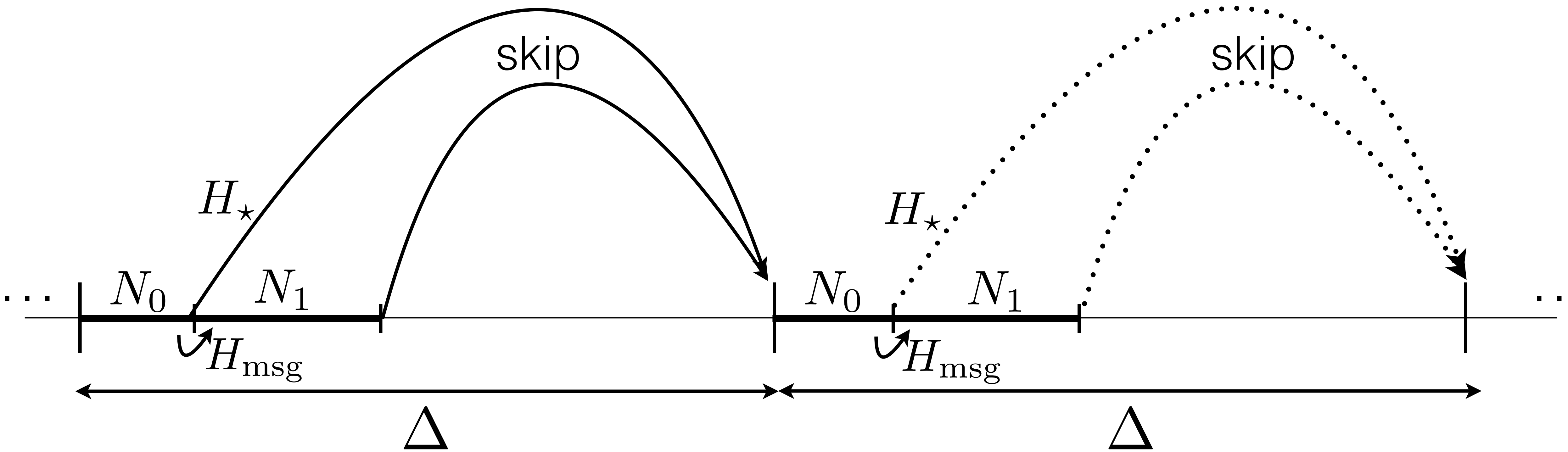
- if $\rho = o(1/B)$ $\quad\quad$ unreliable communication


Multi-phase
$\rho_0 < \rho_1 < \ldots < \rho_\ell = 1$
$H_\star$
ρ_0
H_{msg}
ρ_1
$H_\star$
H_{msg}
$\ldots$
$H_\star$
H_{msg}
ρ_ℓ
$H_\star$
H_{msg}
skip samples
repeat
stop and decode



At each s-instant $\{t = j \cdot \Delta, j \in \mathbb{N}\}$

- test $N_0, N_1, \ldots, N_\ell$ samples
- if a test $\rightarrow H_\star$ skip samples until next s-instant, repeat
- if ℓ consecutive test $\rightarrow H_{\text{msg}}$, decode



Parameters can be chosen such that

$$\rho = \omega(1/B)$$

Theorem (min,min,min Chandar-T 2016):

For any $\beta \geq 0$ and $0 < \rho \leq 1$

- if $\rho = \omega(1/B)$ $\quad$ $\mathbf{C}_{\mathrm{async.}}(\beta,\rho) = \mathbf{C}_{\mathrm{async.}}(\beta,1)$

  $d_{\min}(\beta,\rho) = d_{\min}(\beta,1)(1+o(1))$

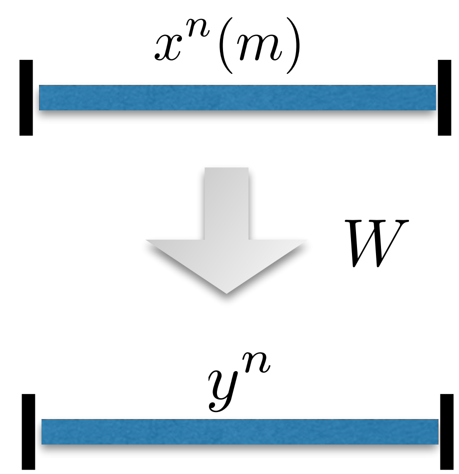
- if $\rho = o(1/B)$ $\quad$ unreliable communication

Artefacts of asymptotic analysis?

Finite length analysis

($k(x) = 1$)

Synchronous communication



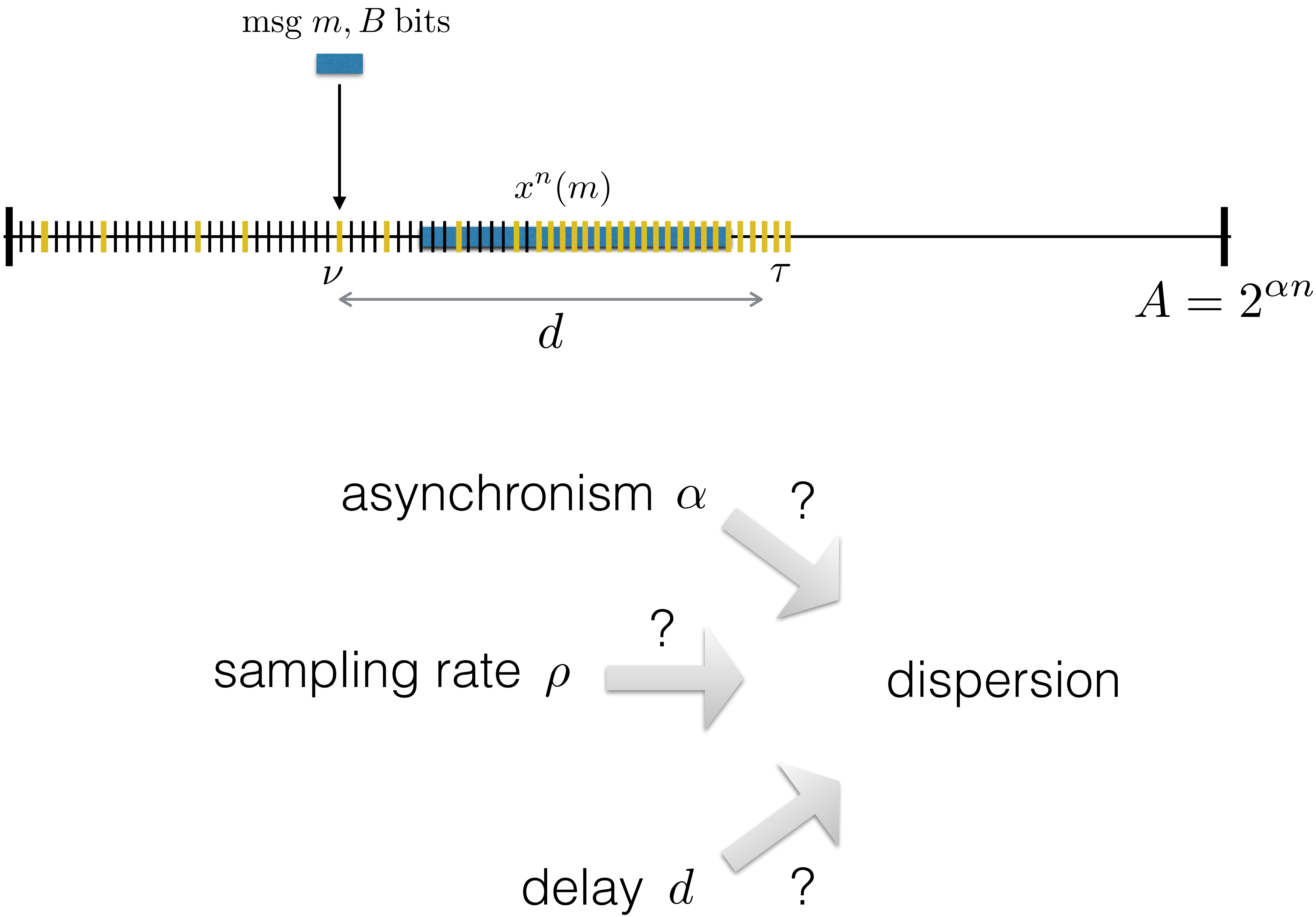
$$Pr(\hat{m} \neq m) \leq \varepsilon$$

Theorem (Strassen 1962, Polyanskiy et al. 2010)

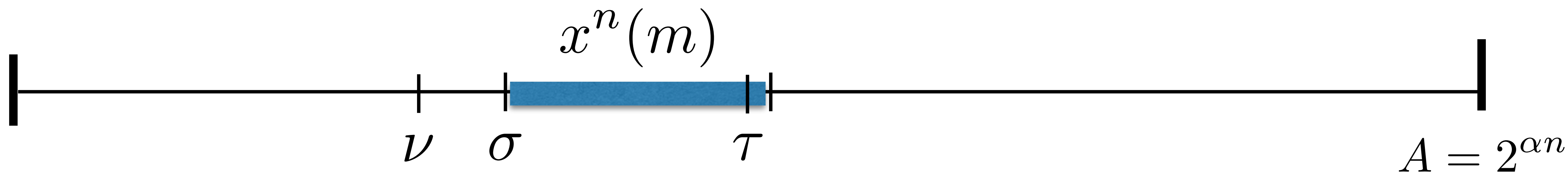
$$B^*(n, \varepsilon) = nC - \sqrt{n\mathcal{V}_\varepsilon(W)}Q^{-1}(\varepsilon) + O(\log n)$$

largest message size

dispersion, “gap to capacity”



msg m, B bits
$x^n(m)$
ν
τ
d
$A = 2^{\alpha n}$
asynchronism α
?
sampling rate ρ
?
dispersion
delay d
?

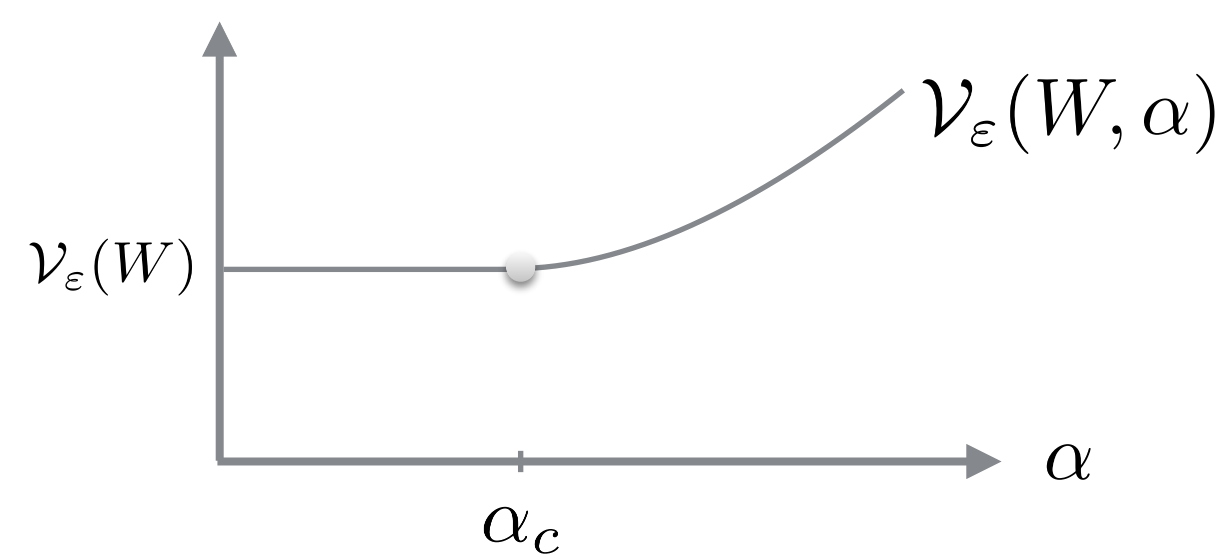


Recall $C(\alpha) = \max_{X: D(Y||Y_\star) \geq \alpha} I(X;Y)$

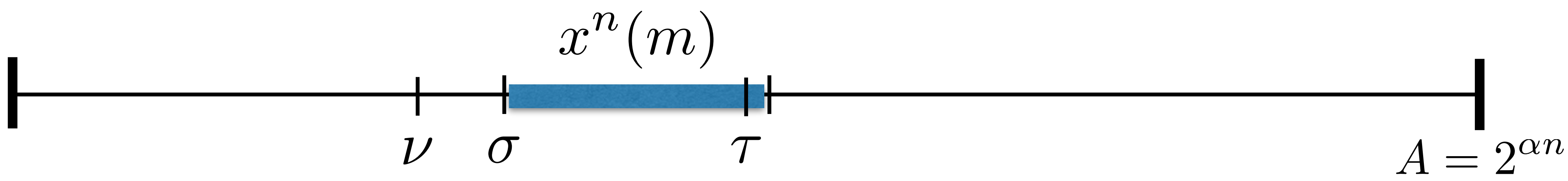
$$d_{\min}(\alpha) = n$$

Definition:

$\mathcal{V}_\varepsilon(W, \alpha)$ dispersion of synchronous channel W under input constraint $D(Y||Y\star) \geq \alpha.$



Minimum delay: $d_n = d_{\min} = n$



Theorem (full sampling, Polyanskiy 2013, Li-T. 2017)

$$B^*(n, \varepsilon, \alpha, \rho_n = 1) = nC(\alpha) - \sqrt{n\mathcal{V}_\varepsilon(W,\alpha)}Q^{-1}(\varepsilon) + O(\log n)$$

Theorem (sparse sampling, Li-T. 2017)

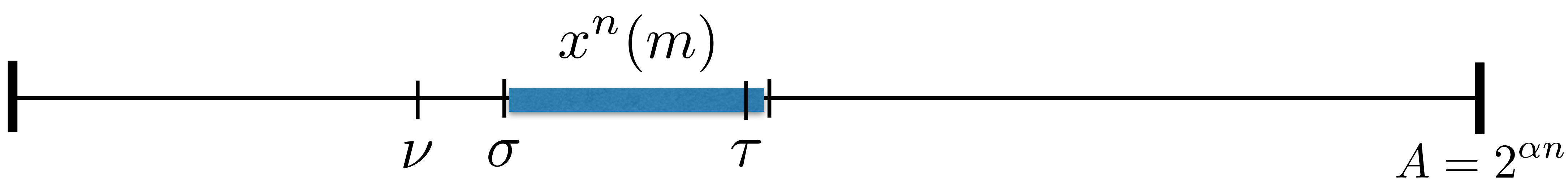
$$B^*(n, \varepsilon, \alpha, \omega(1/\sqrt{n})) = nC(\alpha) - \sqrt{n\mathcal{V}_\varepsilon(W,\alpha)}Q^{-1}(\varepsilon) + o(\sqrt{n})$$

$$B^*(n, \varepsilon, \alpha, \omega(1/n)) = nC(\alpha) - \Theta(1/\rho_n) + O(\sqrt{n})$$

↑

$\&\ O(1/\sqrt{n})$

Close to minimum delay: $d_n = n(1+o(1))$

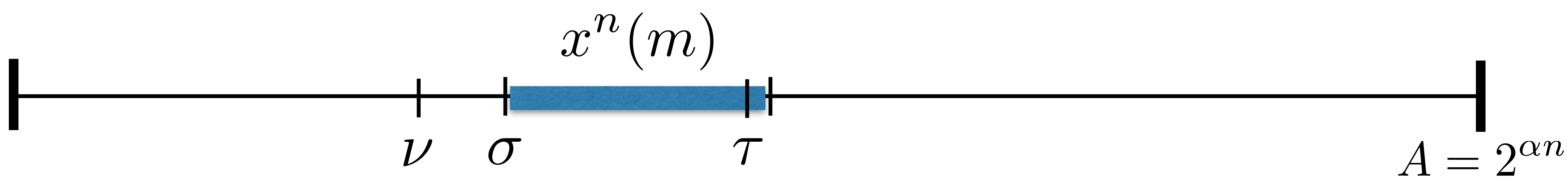


Theorem (sparse sampling, Li-T. 2017)

$$B^*(n, \varepsilon, \alpha, \omega(1/n)) = nC(\alpha) - \sqrt{n\mathcal{V}_\varepsilon(W,\alpha)}Q^{-1}(\varepsilon) + O(\log n)$$

minimum sampling rate & $d = n(1+o(1))$ $\underset{\text{dispersion}}{\equiv}$ minimum delay & $\rho_n = \omega(1/\sqrt{n})$

Intuition: $d_n = d_{\min} = n$



$$B^*(n, \varepsilon, \alpha, \omega(1/n)) = nC(\alpha) - \Theta(1/\rho_n) + O(\sqrt{n})$$

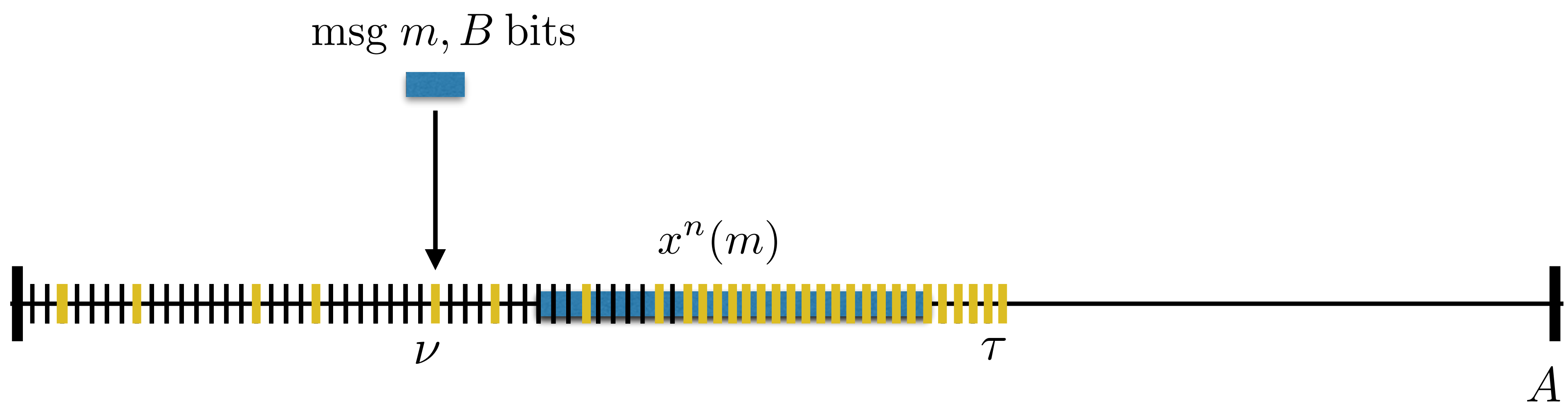
- At sampling rate ρ we miss $\Theta(1/\rho)$ symbols of the sent codeword.

- So if $\rho = o(1/\sqrt{n})$ decoder will miss $\omega(\sqrt{n})$

  of the sent codeword.

- Codes of length $n - \omega(\sqrt{n})$ have $\omega(\sqrt{n})$ order synchronous dispersion.

Summary



- Capacity, capacity per unit cost
- Tradeoffs between rate, delay, sampling rate
- Finite length analysis

Open issues

- Bit asynchronism: capacity per unit cost?
- Random/multiple access: bursty interference (Chandar-T.-Caire 2014, Shahi et. al. 2016, Farkas-Kói 2016)
- Asynchronous networks (Shomorony et al. 2014, Gallager 1976)

Bibliography

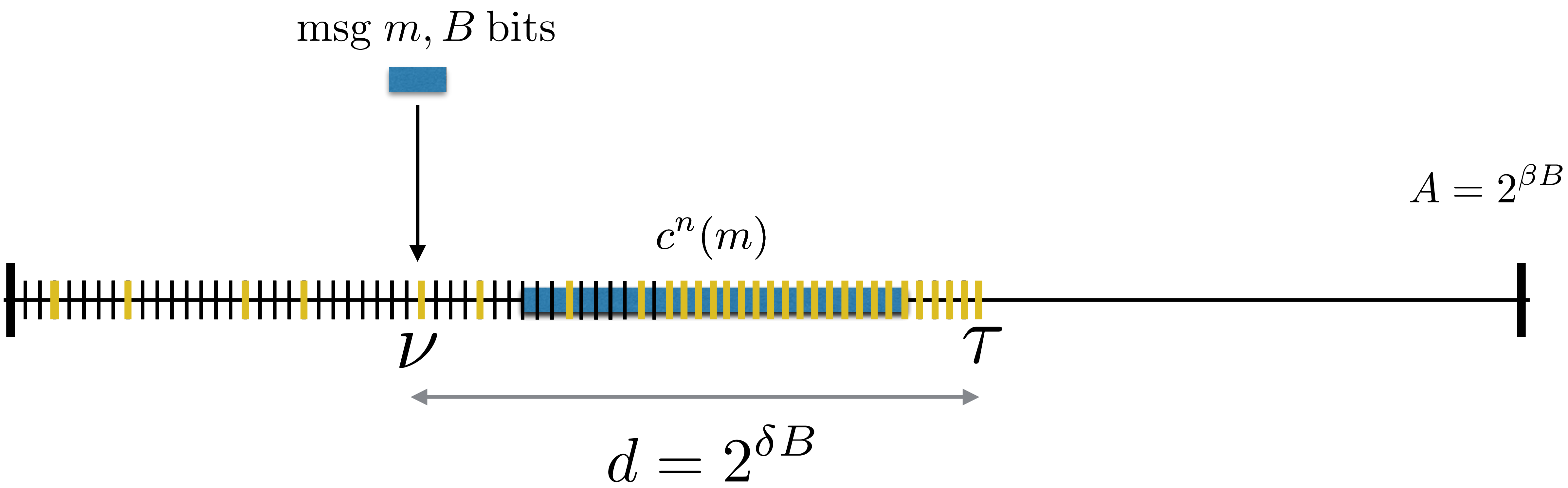
1. Dobrushin, Roland L'vovich. "Shannon's theorems for channels with synchronization errors." *Problemy Peredachi Informatsii* 3.4 (1967): 18-36.
2. Kanoria, Yashodhan, and Andrea Montanari. "Optimal coding for the binary deletion channel with small deletion probability." IEEE Transactions on Information Theory 59.10 (2013): 6192-6219.
3. Anantharam, Venkat, and Sergio Verdu. "Bits through queues." IEEE Transactions on Information Theory 42.1 (1996): 4-18.
4. Ulukus, Sennur, et al. "Energy harvesting wireless communications: A review of recent advances." IEEE Journal on Selected Areas in Communications 33.3 (2015): 360-381.
5. Verdu, Sergio. "On channel capacity per unit cost." IEEE Transactions on Information Theory 36.5 (1990): 1019-1030.
6. LIDS-P-1714, November 1987. Gallager, R. G., "Energy Limited Channels: Coding, Multiaccess, and Spread Spectrum", M.I.T., Laboratory for Information and Decision Systems, Report
7. Chandar, Venkat, Aslan Tchamkerten, and Gregory Wornell. "Optimal sequential frame synchronization." IEEE Transactions on Information Theory 54.8 (2008): 3725-3728.
8. Chandar, Venkat, Aslan Tchamkerten, and David Tse. "Asynchronous capacity per unit cost." IEEE Transactions on Information Theory 59.3 (2013): 1213-1226.
9. Tchamkerten, Aslan, Venkat Chandar, and Giuseppe Caire. "Energy and sampling constrained asynchronous communication." IEEE Transactions on Information Theory 60.12 (2014): 7686-7697.
10. Chandar, Venkat, and Aslan Tchamkerten. "How to Quickly Detect a Change while Sleeping Almost all the Time." arXiv preprint arXiv:1501.05930 (2015).
11. Strassen, Volker. "Asymptotische abschätzungen in Shannons informationstheorie." Trans. Third Prague Conf. Inf. Theory. 1962.
12. Polyanskiy, Yury, H. Vincent Poor, and Sergio Verdú. "Channel coding rate in the finite blocklength regime." IEEE Transactions on Information Theory 56.5 (2010): 2307-2359.
13. Polyanskiy, Yury. "Asynchronous communication: Exact synchronization, universality, and dispersion." IEEE Transactions on Information Theory 59.3 (2013): 1256-1270.

Bibliography

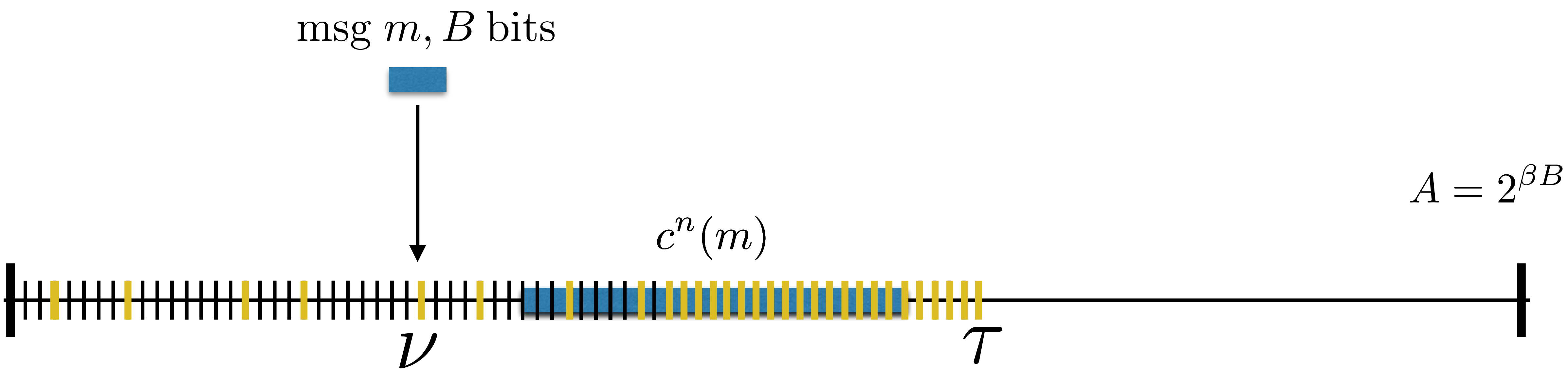
14. Li Longguang and Aslan Tchamkerten. “Infinite dispersion in bursty communication." ISIT 2017.
15. Shahi, Sara, Daniela Tuninetti, and Natasha Devroye. "On the capacity of strong asynchronous multiple access channels with a large number of users." Information Theory (ISIT), 2016 IEEE International Symposium on. IEEE, 2016.
16. Farkas Lóránt, Tamás Kói. "Universal Random Access Error Exponents for Codebooks of Different Word-Lengths." arXiv preprint arXiv:1607.01935 (2016).
17. Shomorony, Ilan, et al. "Diamond networks with bursty traffic: Bounds on the minimum energy-per-bit." IEEE Transactions on Information Theory 60.1 (2014): 461-493.
18. Gallager, Robert. "Basic limits on protocol information in data communication networks." IEEE Transactions on Information Theory 22.4 (1976): 385-398.

Typical scenarios

	Message at transmitter			Receiver		Performance
	arrival time	size	delay	sampling	decision time	
Continuous	fixed	fixed	fixed	fixed	fixed	asymptotic
Bursty	random	random	variable	adaptive	adaptive	finite

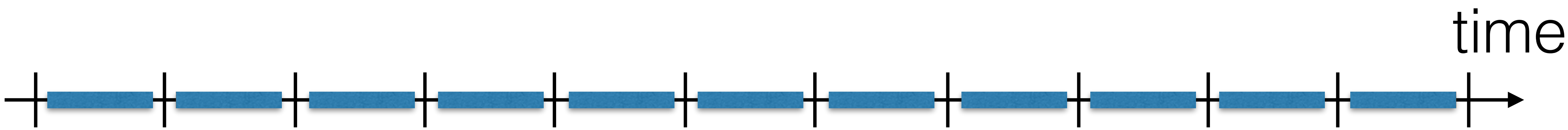


$$\mathbf{C}_{\text{async.}}(\beta, \delta, \rho)?$$



$$\mathbf{C}_{\text{async.}}(\beta, \rho)?$$

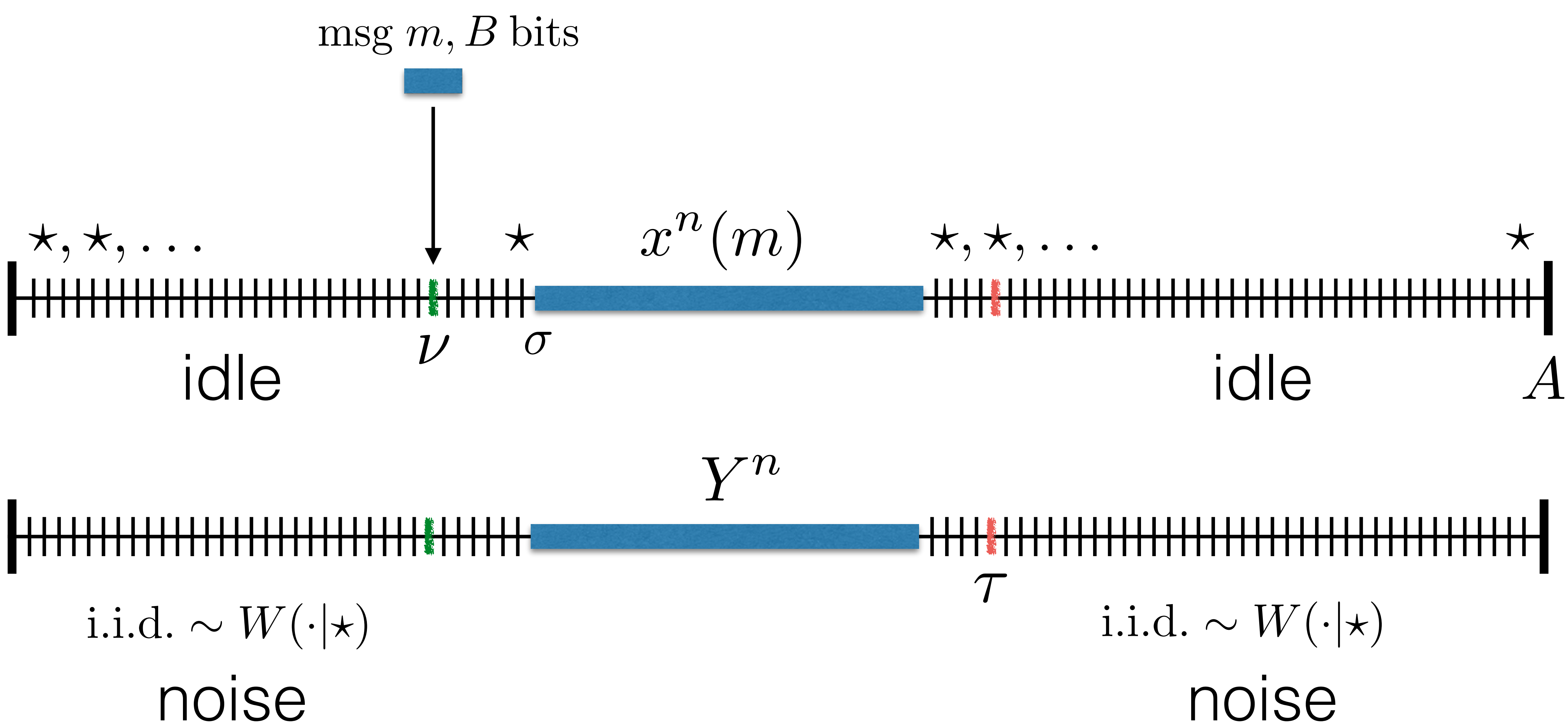
Synchronous communication



- Voice, video
- Packets sent contiguously
- Negligible cost to acquire synchronization

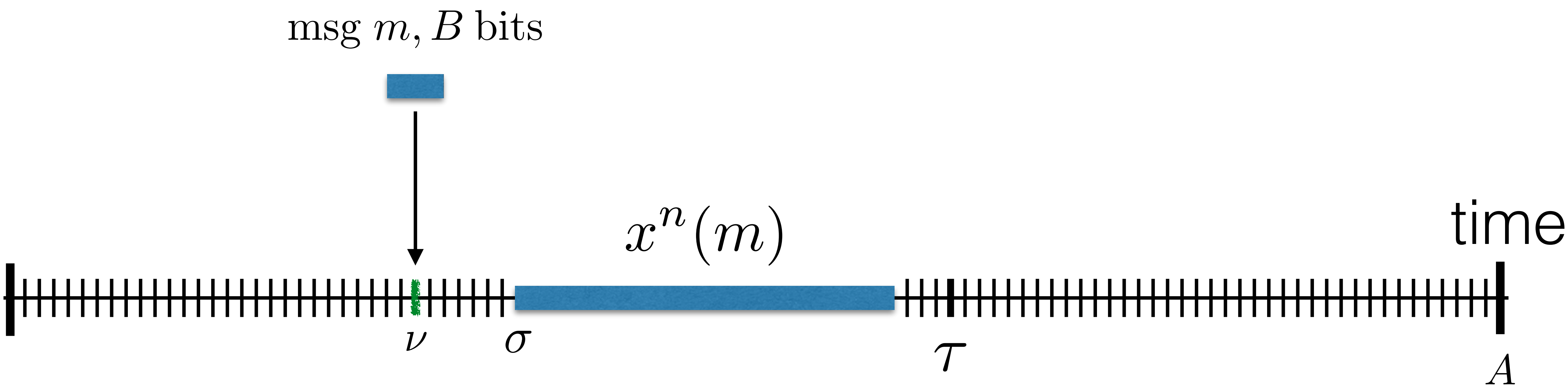
How does asynchronism impact capacity per unit cost?

Bursty communication model



τ : stopping time with respect to $Y_1, Y_2, \ldots$

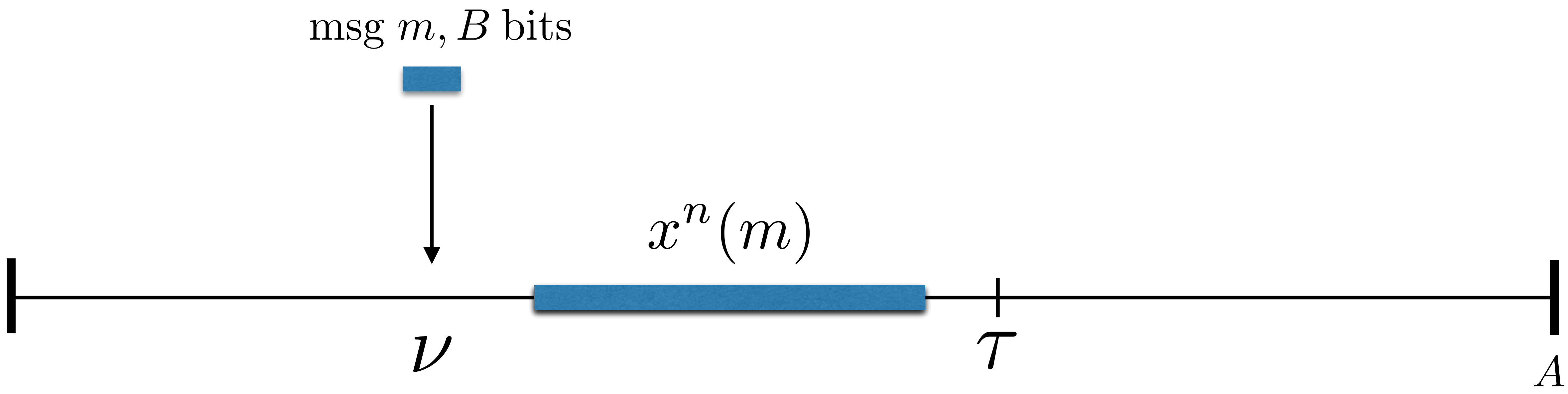
Bursty communication model



A:level of asynchronism

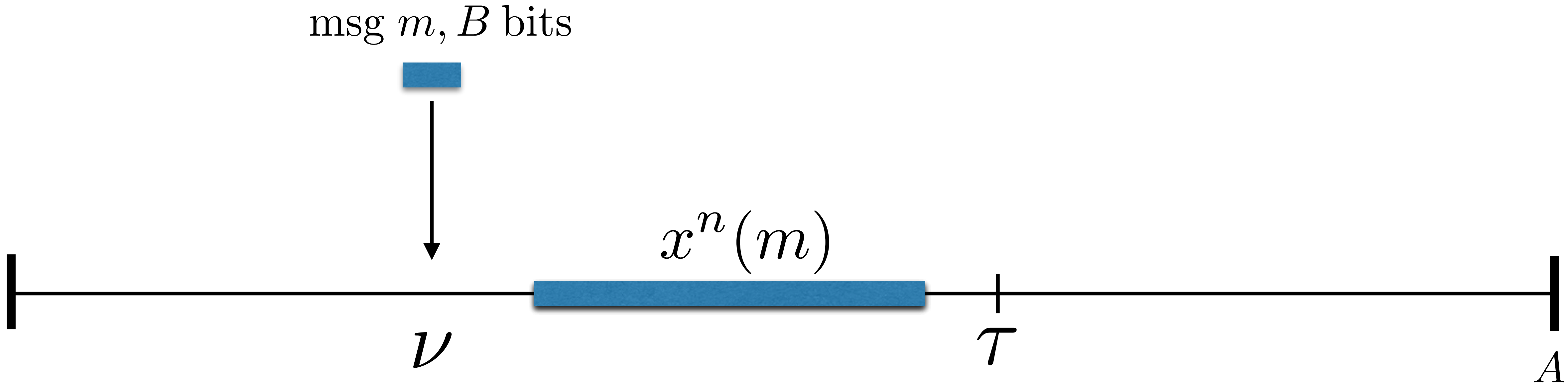
σ:transmission start time

One message sent over the $[1, A]$



- Given A
- Find $\mathbf{C}_{\text{async.}}(A) = \sup\{\text{achievable } \mathbf{R}\}$
-

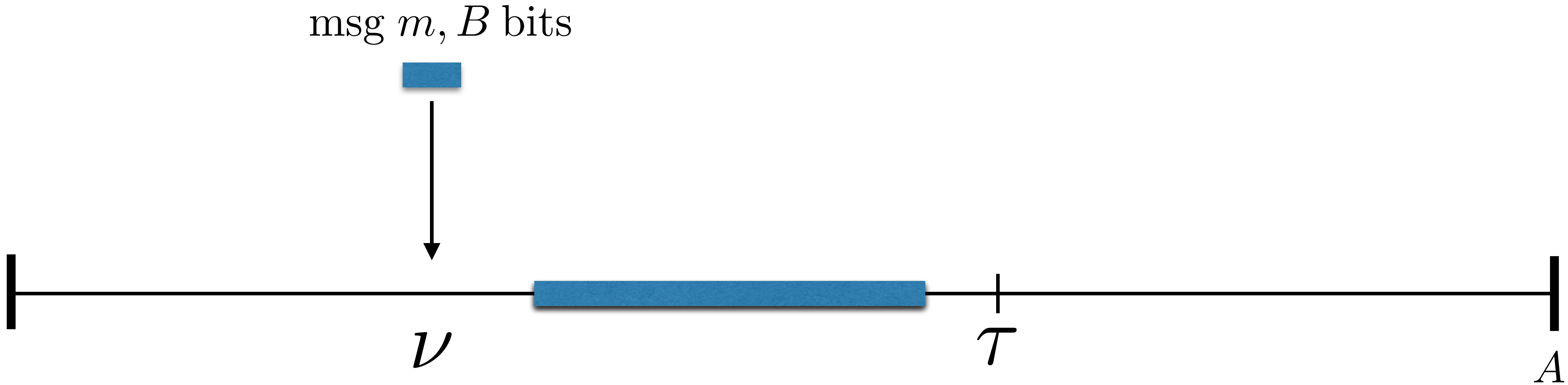
A natural scheme



- Random code $\{x^n(m)\}$ i.i.d. P
- Sequential typicality:

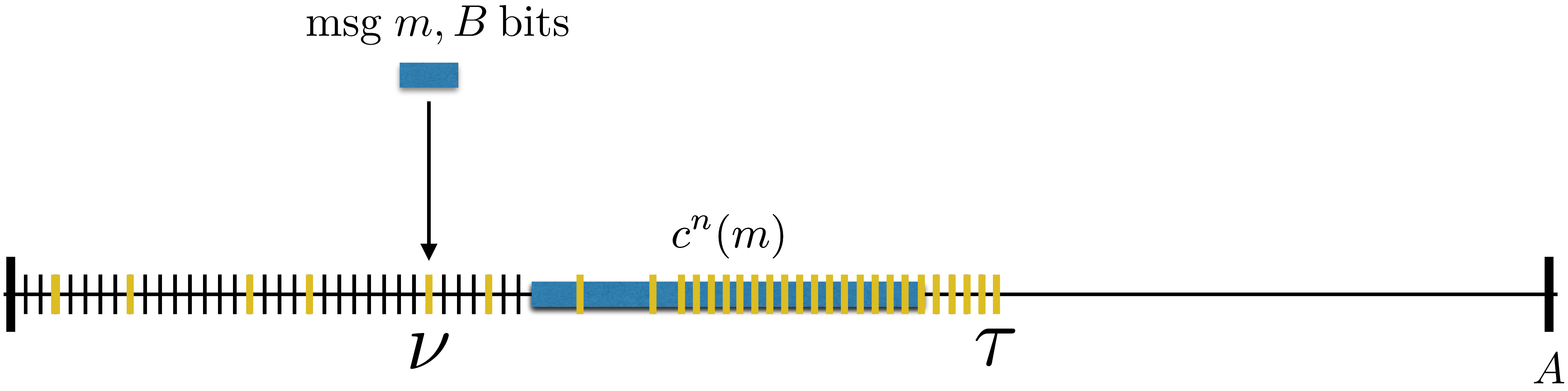
$$\tau = \inf\{t \geq 1 : (x^n(m), y_{t-n+1}^t) \; PW\text{-typical for some } m\}$$

Analysis



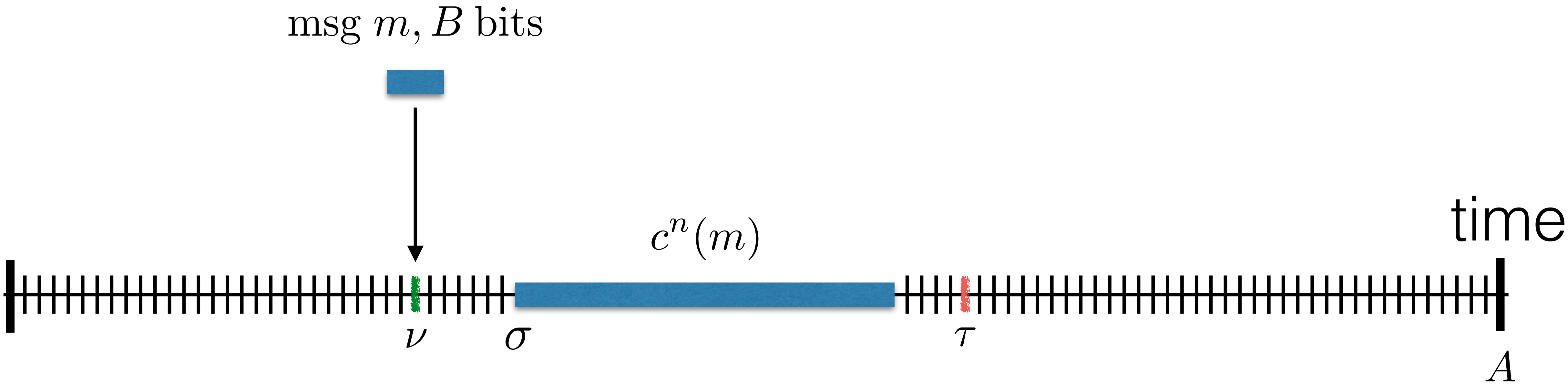
- Random code $\{x^n(m)\}$ i.i.d. P
- Sequential typicality:

$$\tau = \inf\{t \geq 1 : (x^n(m), y_{t-n+1}^t) \; PW\text{-typical for some } m\}$$

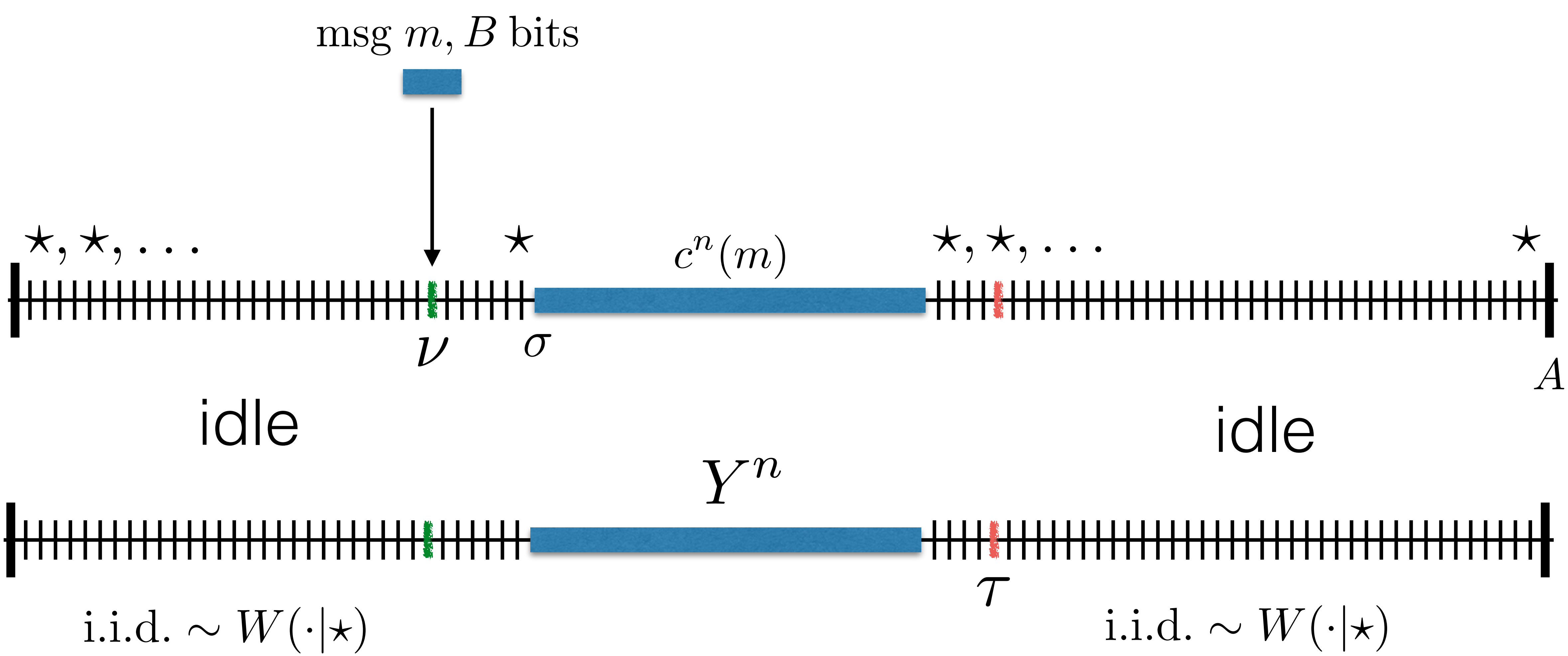


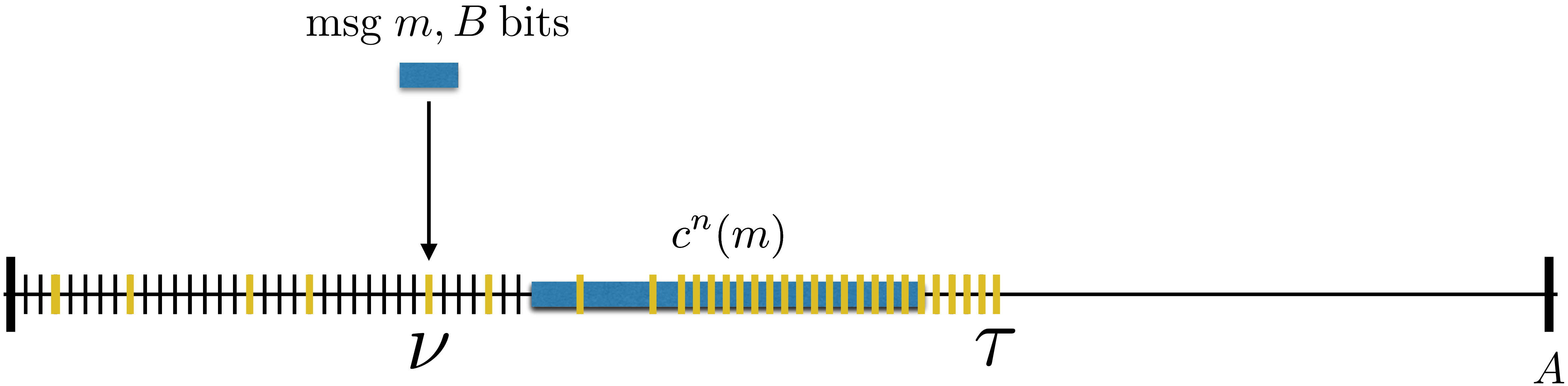
$$\mathbf{C}_{\text{async.}}(A, \rho)?$$

Bursty communication model



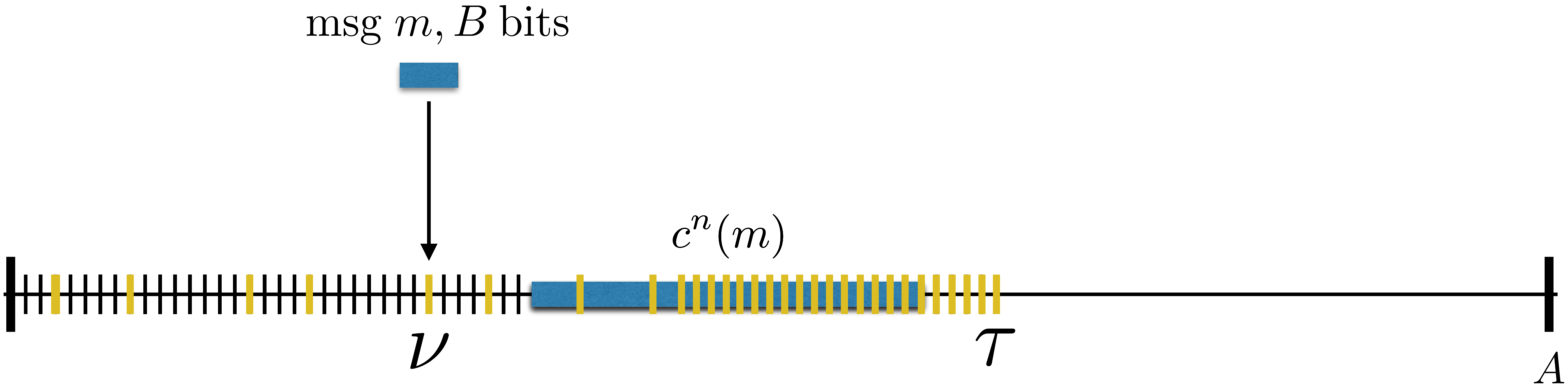
Bursty communication model





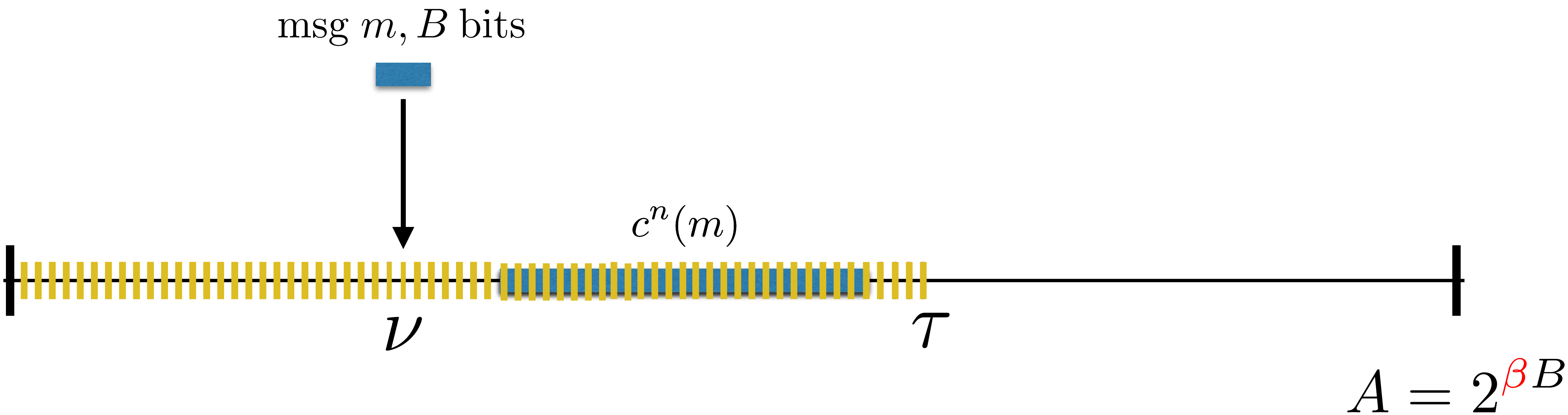
- Input cost
- Output cost: sampling rate

$$\rho = \frac{\text{number of } \textcolor{#D4B820}{|} \text{ until } \tau}{\tau}$$



$$\mathbf{C}_{\text{async.}}(A, \rho)?$$

$$\rho = 1$$

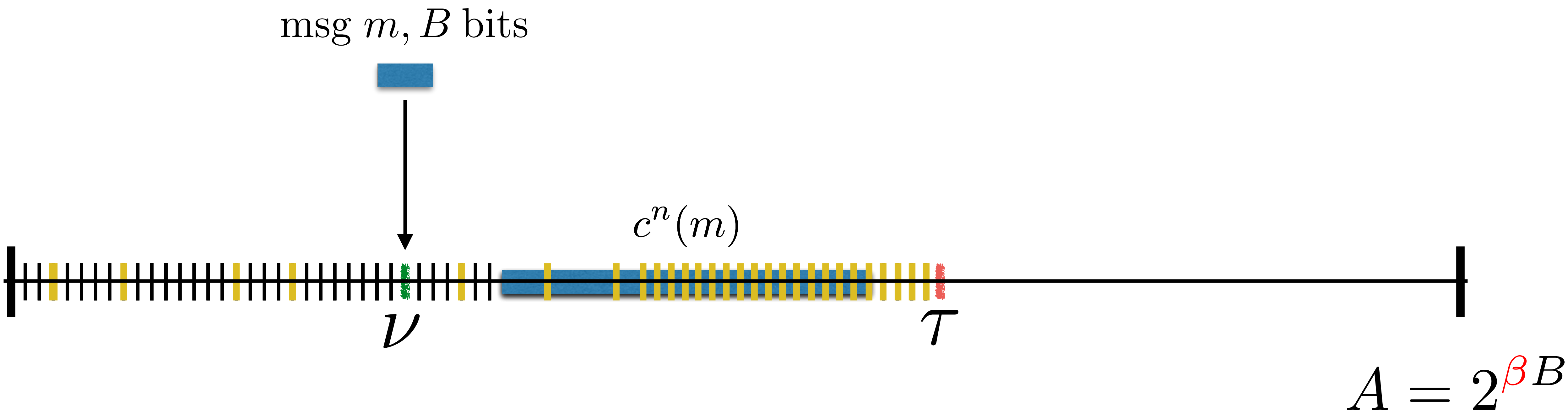


Theorem (Chandar, T., Tse 2010):

$$\mathbf{C}_{\text{async.}}(A, \rho = 1) = \max_X \min \left\{ \frac{I(X;Y)}{\mathbb{E}k(X)}, \frac{I(X;Y) + D(Y \| Y_\star)}{(1+\beta)\mathbb{E}k(X)} \right\}$$

Minimum delay $\quad d_{\min}(\beta, \rho = 1)(= \Theta(B))$

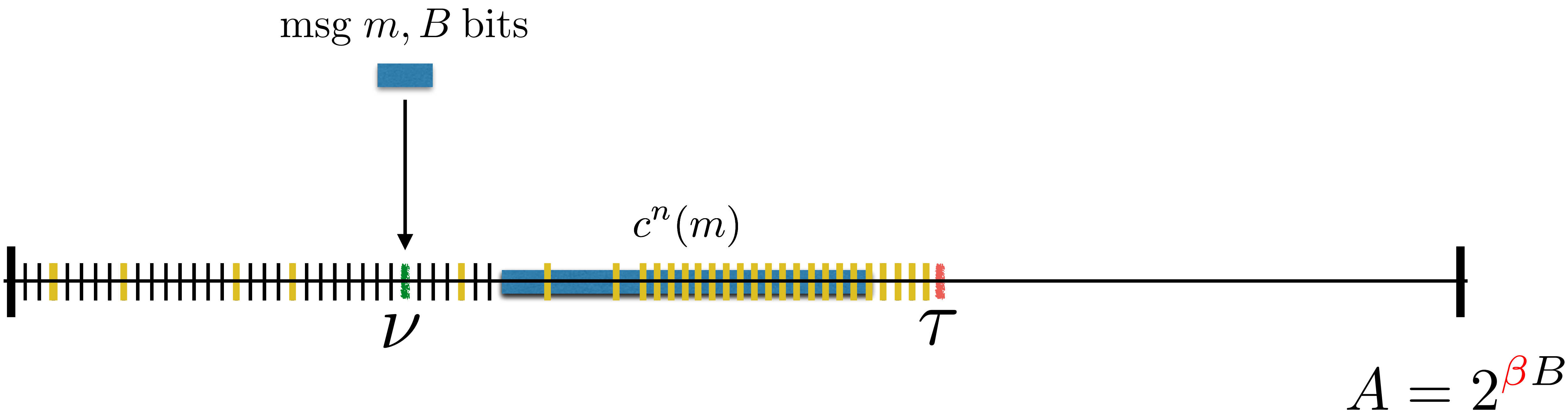
$$0 < \rho \leq 1$$



Theorem (Chandar, T., Caire 2013): for any $\beta > 0$

$$\mathbf{C}_{\text{async.}}(\beta, \rho) = \mathbf{C}_{\text{async.}}(\beta, 1)$$

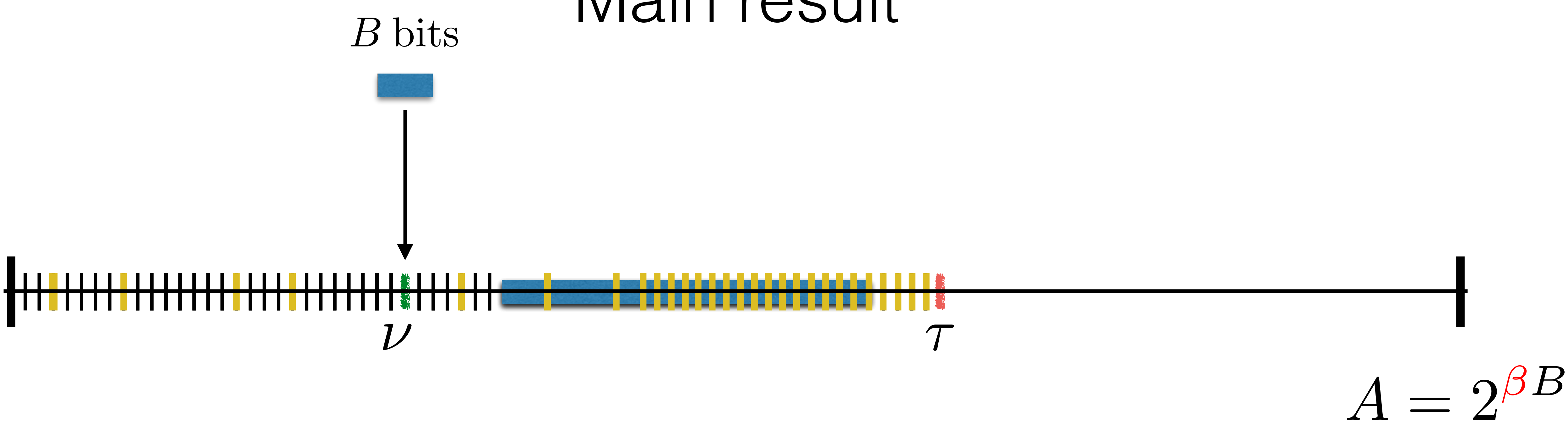
$$d_{\min}(\beta, \rho) \sim d_{\min}(\beta, \rho = 1)$$



No loss for constant sampling rate.

$$\rho \longrightarrow 0\,?$$

Main result



Theorem: for any $\beta > 0$

- if $\rho = \omega(1/B)$ $\quad$ $\mathbf{C}_{\text{async.}}(A, \rho) = \mathbf{C}_{\text{async.}}(A, 1)$

  $d_{\min}(\beta, \rho) \sim d_{\min}(\beta, \rho = 1)$

- if $\rho = o(1/B)$ $\quad$ unreliable communication

Achievability

Transmitter: random coding ✓

Receiver: decoding function ✓

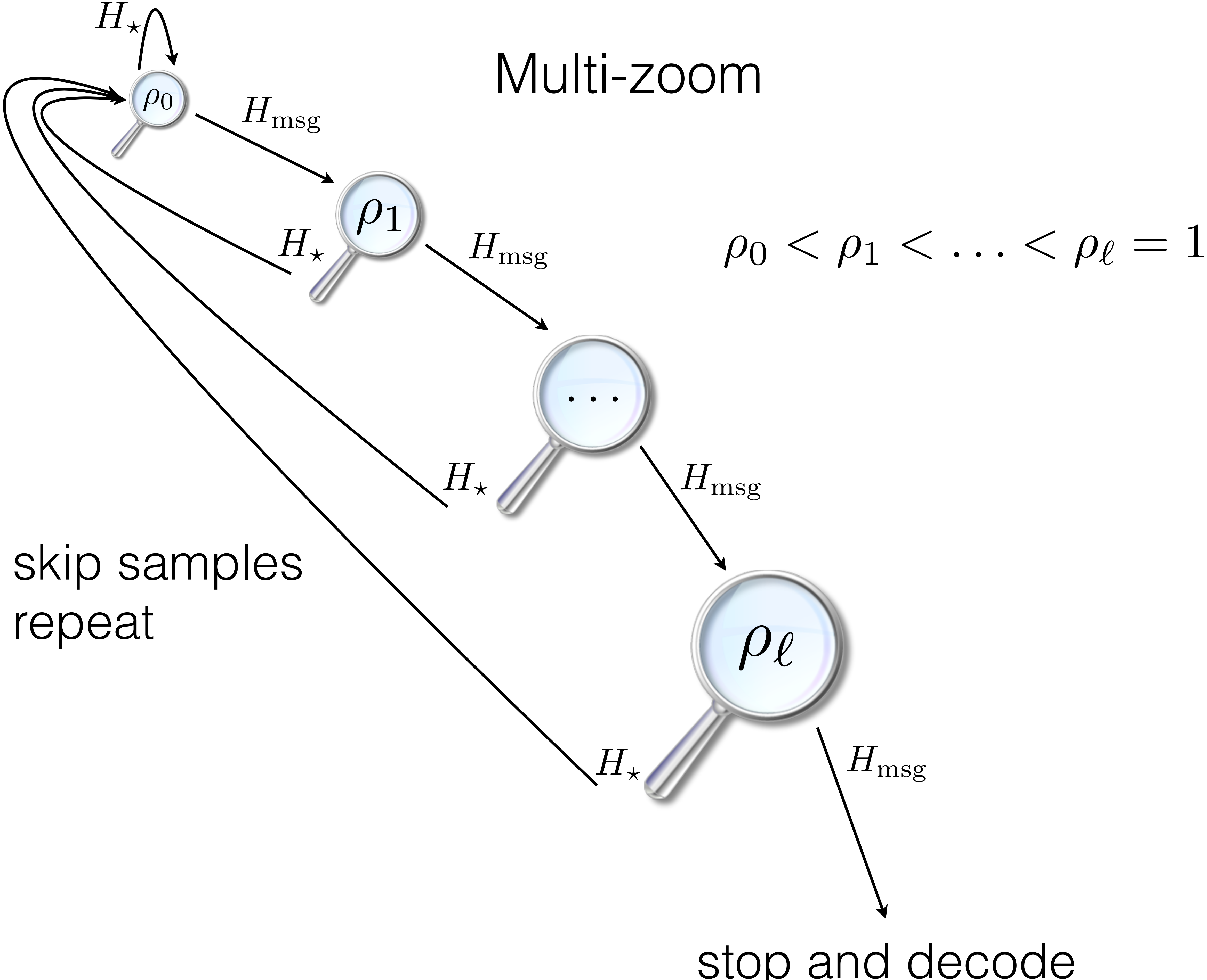
stopping rule

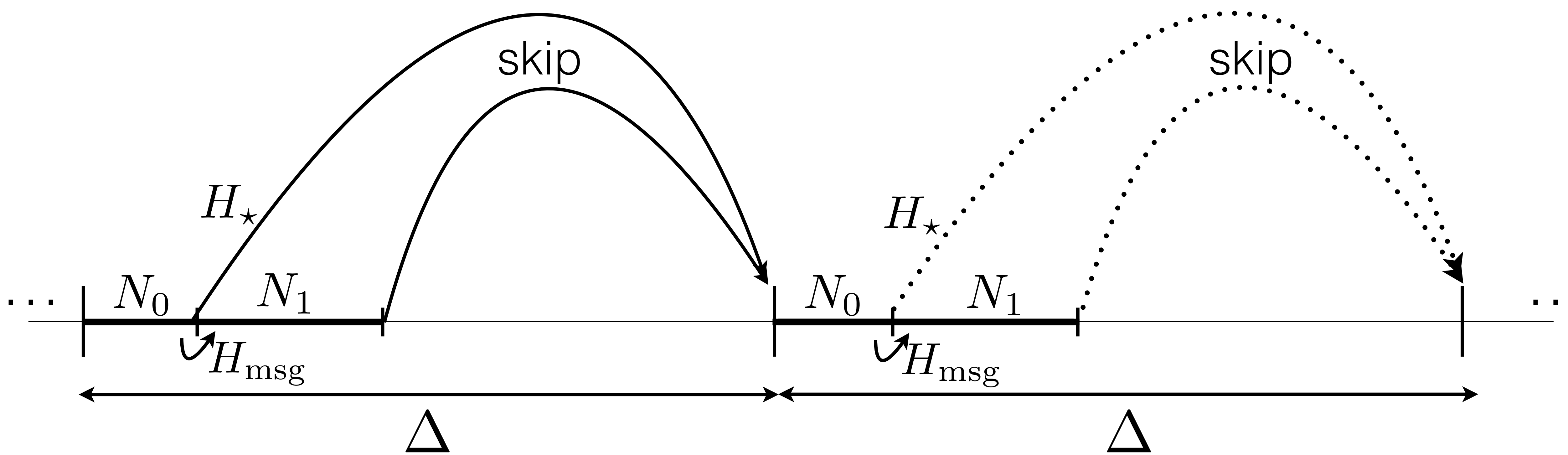
sampling strategy

?

Message detection with $\rho = \omega(1/B)$ samples?

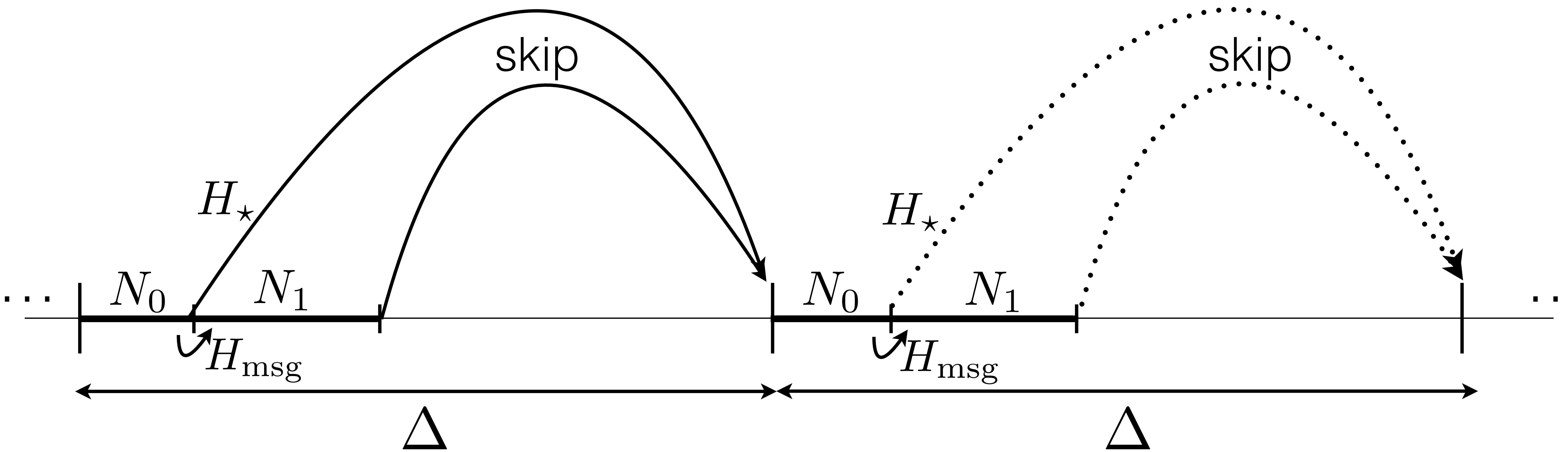
Multi-zoom


$H_\star$
ρ_0
H_{msg}
ρ_1
$H_\star$
H_{msg}
$\rho_0 < \rho_1 < \ldots < \rho_\ell = 1$
$\ldots$
$H_\star$
H_{msg}
skip samples
repeat
ρ_ℓ
$H_\star$
H_{msg}
stop and decode



At each s-instant $\{t = j \cdot \Delta, j \in \mathbb{N}\}$

- test $N_0, N_1, \ldots, N_\ell$ samples
- if a test $\rightarrow H_\star$ skip samples until s-instant, repeat
- if ℓ consecutive test $\rightarrow H_{\rm msg}$, decode

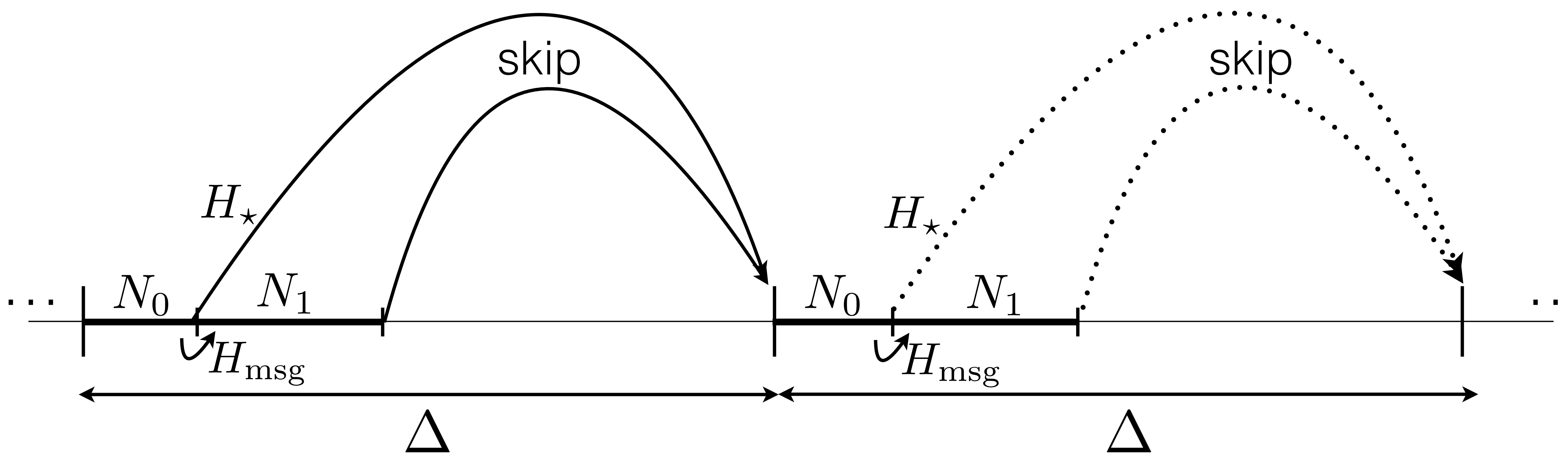


Parameters can be chosen such that

$$\rho = \omega(1/B)$$

Summary

- Fundamental tradeoffs for bursty communication
- To combat asynchronism we only need strong signals; better output sampling does not help
- Decoder can sleep almost all the time and yet be maximally efficient (asympt.).



Design tests such that first phase dominates:

$$\Rightarrow \begin{cases} \#\{\text{noise samples at s-instant}\} \sim N_0 \\ N_0 = \omega(1) \end{cases}$$

$$\Rightarrow \rho \sim \frac{\omega(1)}{\Delta}$$

But $\Delta = o(B)$ otherwise change is missed.

$$\Rightarrow \rho \sim \omega(1/B)$$