THE UNIVERSITY *of* WISCONSIN

# Objectives of talk

# Agenda

**Analyzing decoding error for a bounded information decoder: regimes of interest**

Error exponents of ML decoders

Non-asymptotic analysis of ML decoder & Normal approximation

# Agenda

**Analyzing decoding error for a bounded**

# Motivating question

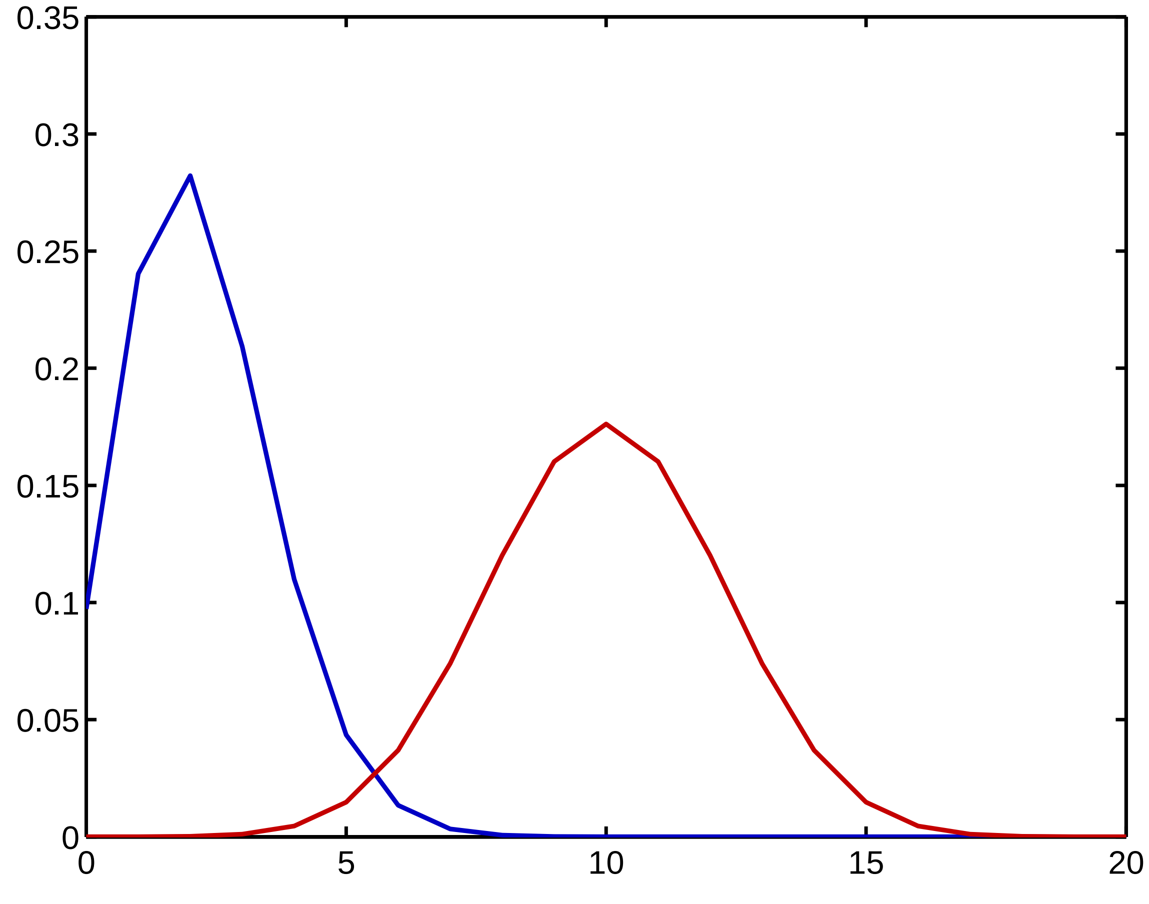
Want to analyze ML decoding

# A few manipulations

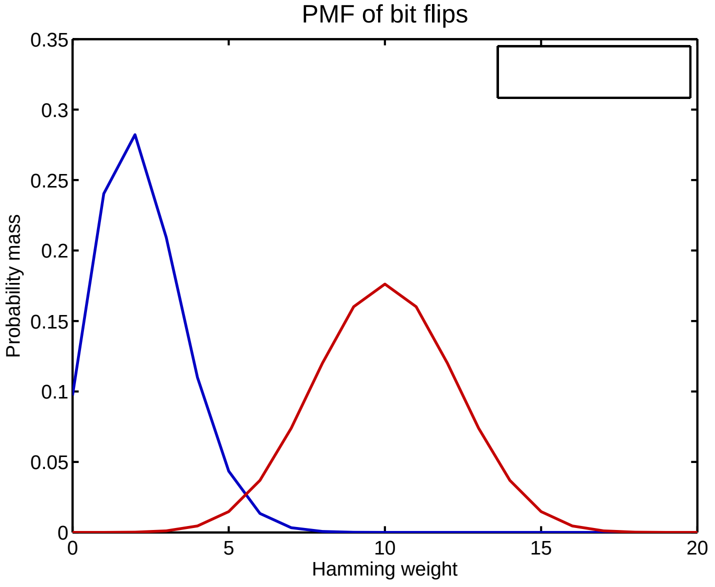
# Standard random codebook ensemble analysis

# Bounded information = bounded distance decoder

“re-center” analysis around observation

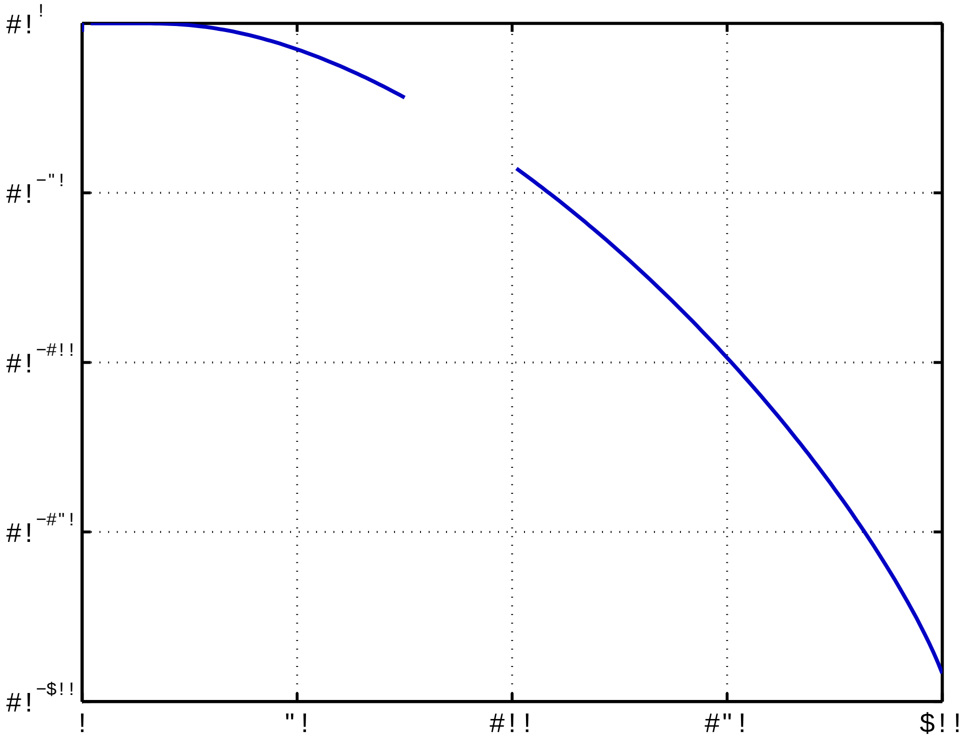


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20


PMF of bit flips
Probability mass
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
Hamming weight

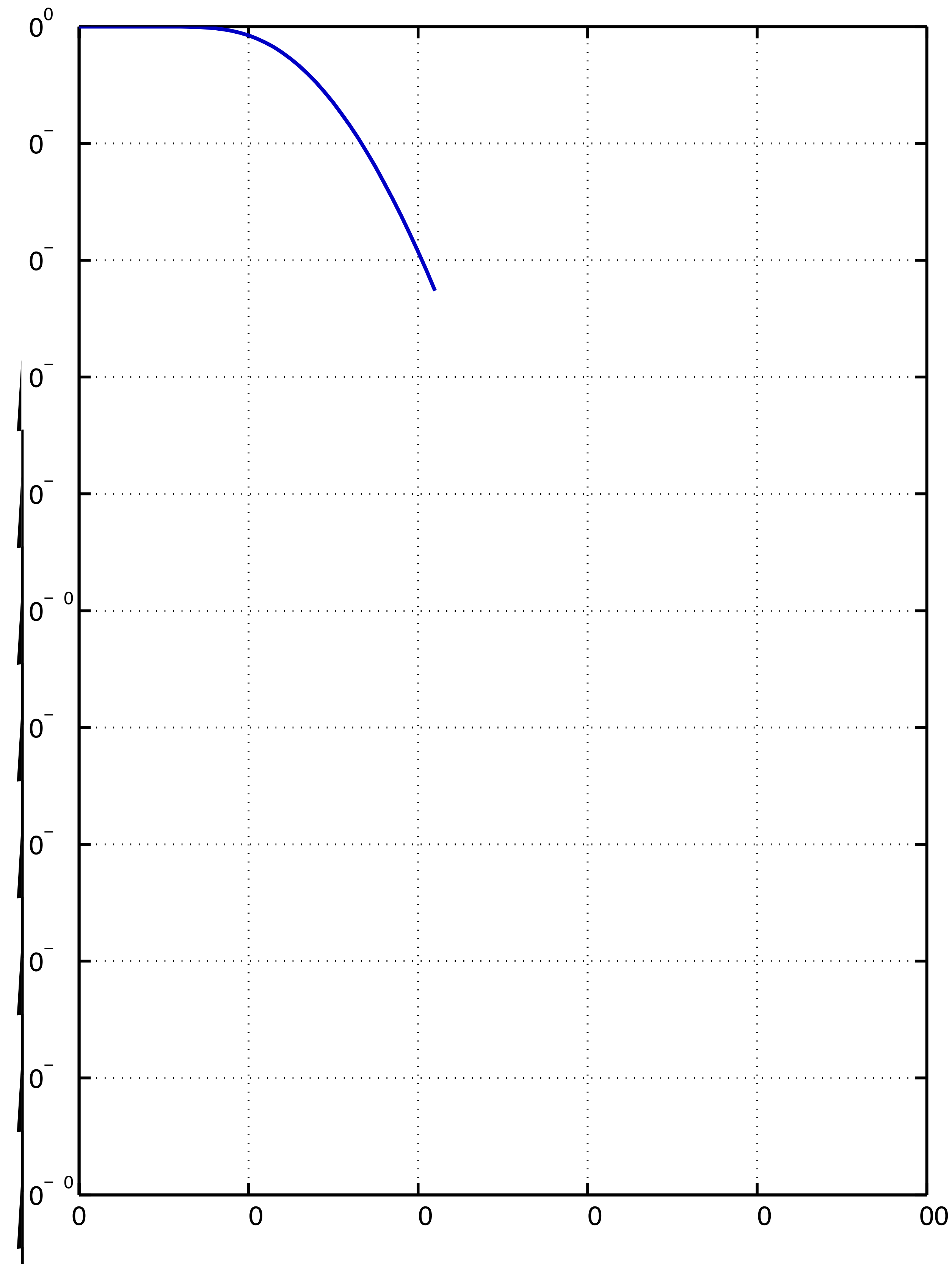
But, what we really care about is that $\mathrm{wt}_H(\tilde{\mathbf{X}}_0)$ is too big, or $\mathrm{wt}_H(\tilde{\mathbf{X}}_1)$ is too small. So, more useful to examine CDFs, rather than PMFs.

# Plot on log-scale to see better



zoom in on upper left

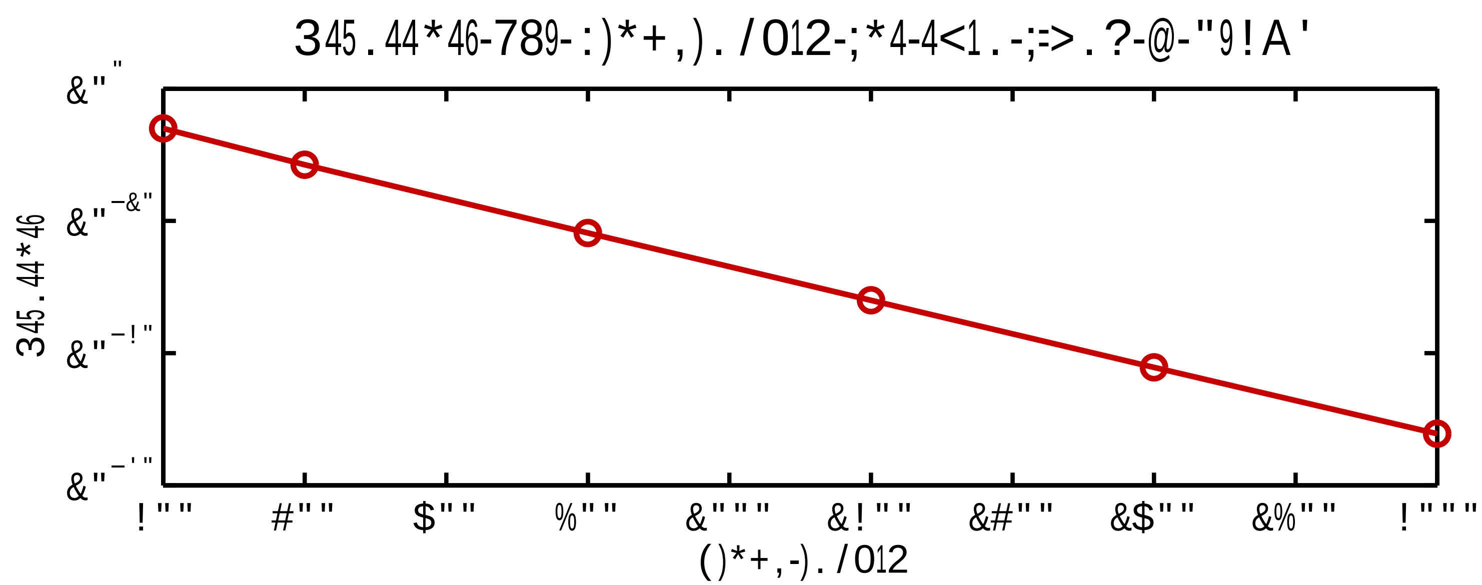
if you
(i) use this decoder, and
(ii) target an error rate of $10^{-3}$, then



0^0
0^−
0^−
0^−
0^−
0^− 0
0^−
0^−
0^−
0^−
0^− 0
0
0
0
0
0
00

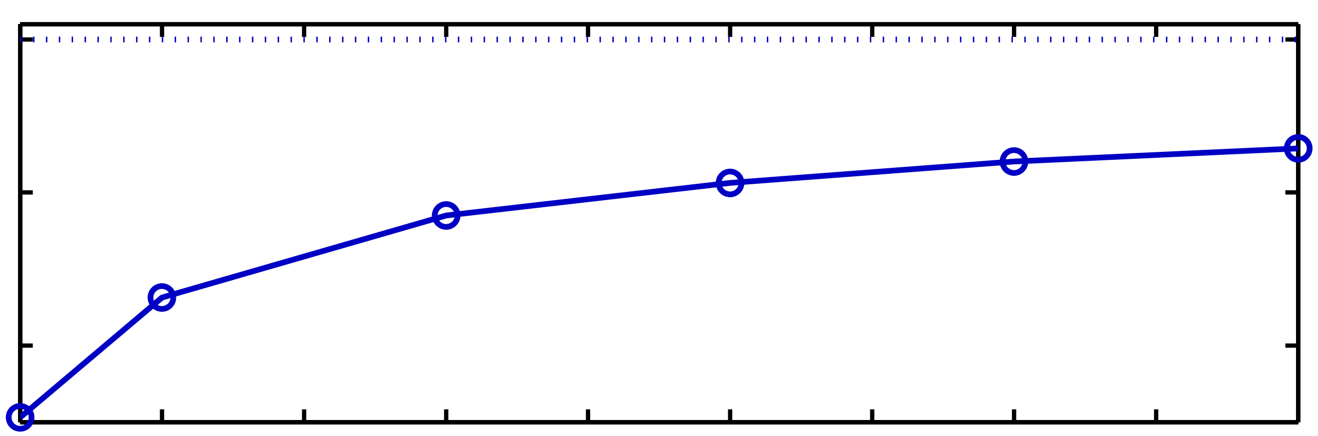
# Remainder of talk: two chunks

Chunk 1: Error exponent analysis of ML decoding



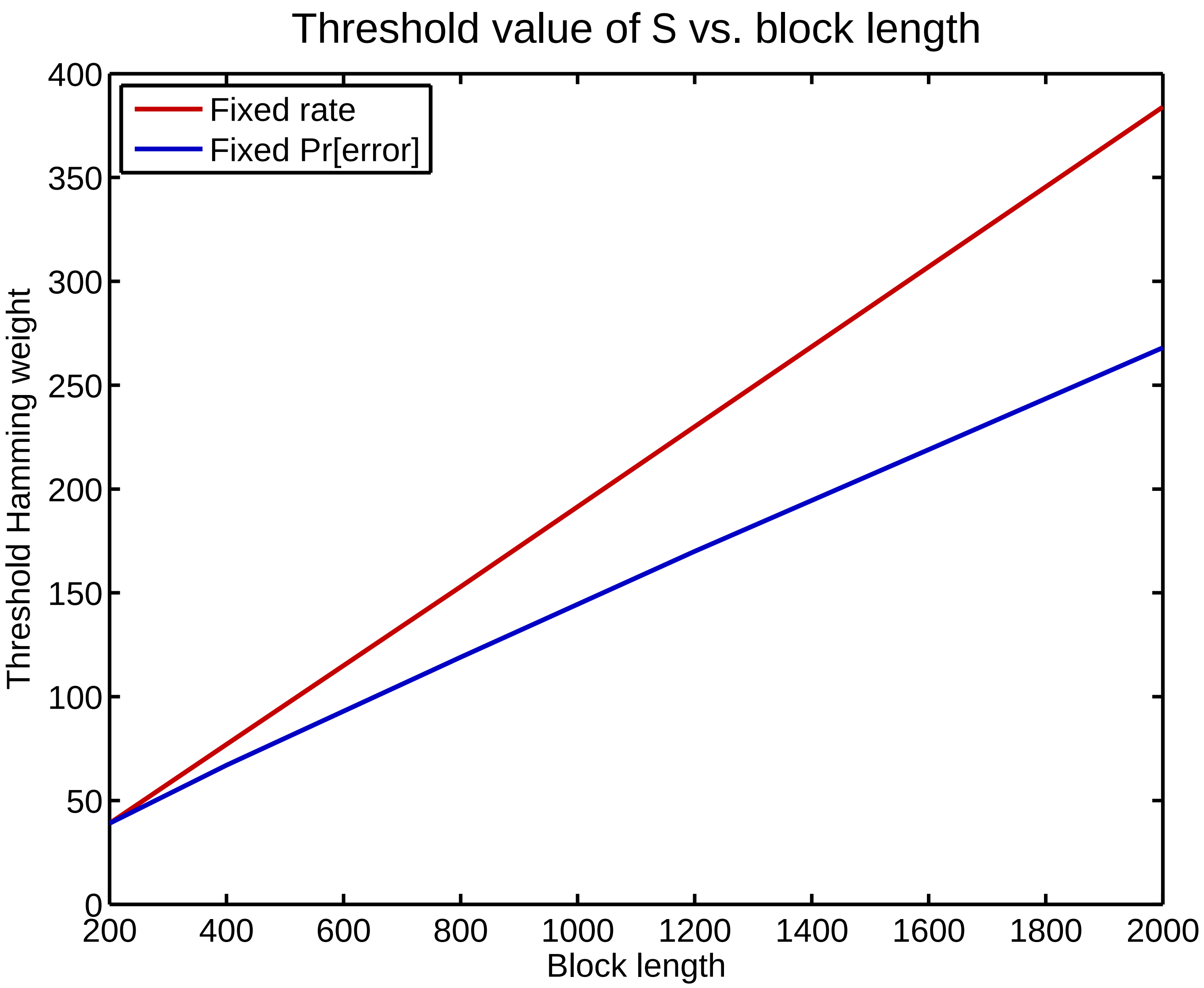
Slope (magnitude) of error decay on a log plot is the “error exponent”. Here it is about 0.0434.  Can it be improved?

Chunk 2: How fast approach capacity using ML decoding

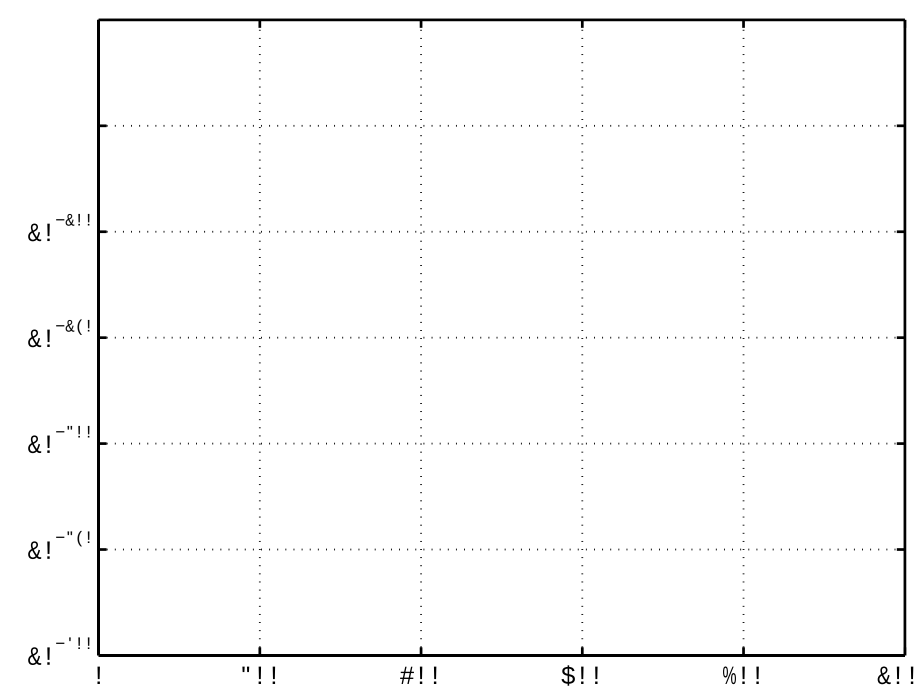


Asymptotes to capacity as n increases.  But, how long to get close?  How does the rate approach capacity? Can you approach faster than in plot to left?

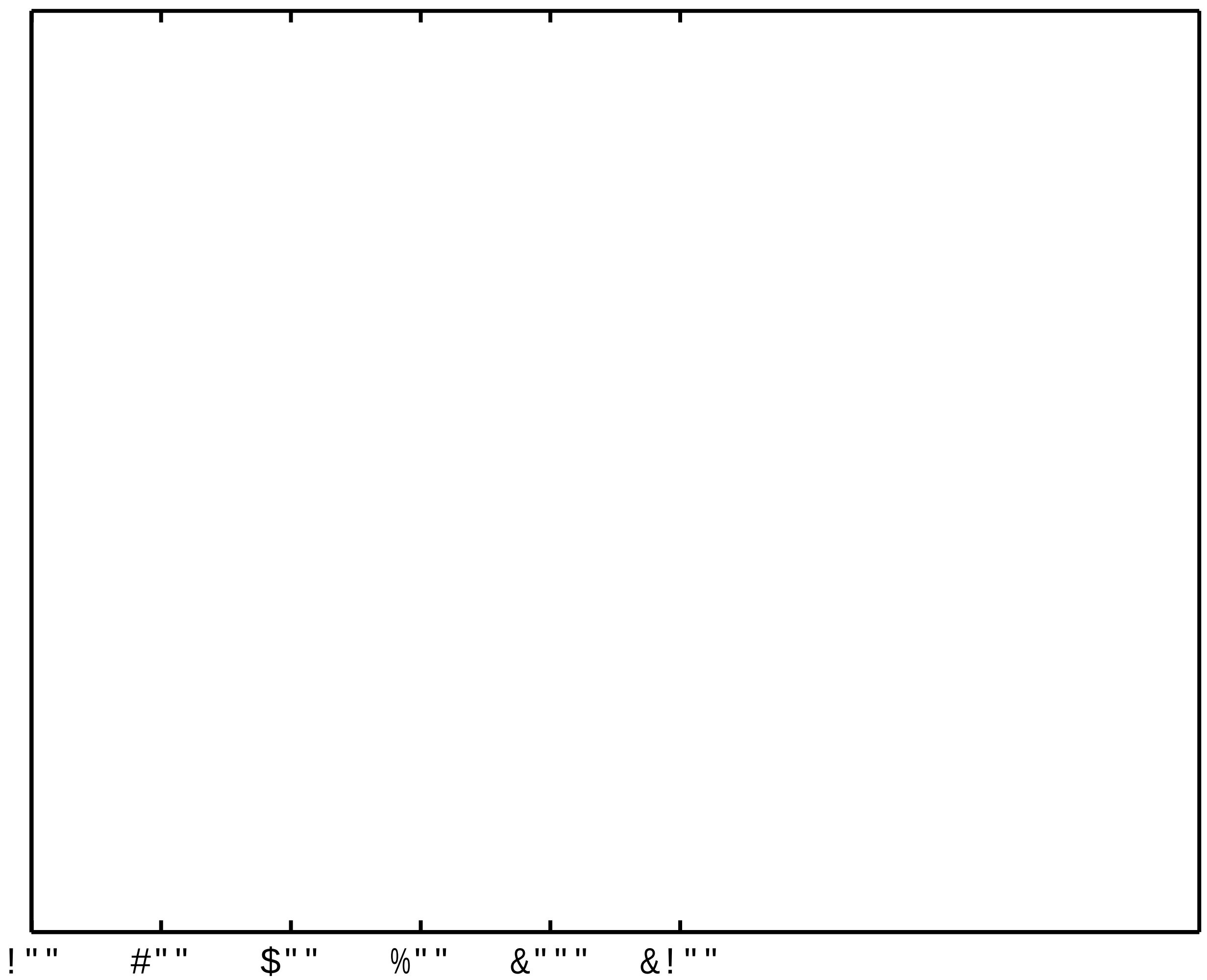
$\Delta = 384$



Threshold value of S vs. block length
Fixed rate
Fixed Pr[error]
Threshold Hamming weight
Block length
0
50
100
150
200
250
300
350
400
200
400
600
800
1000
1200
1400
1600
1800
2000



# “Excess” Hamming weight: linear vs. square-root



# Recap: what we have learned

# Agenda

Analyzing decoding error for a bounded information decoder: regimes of interest

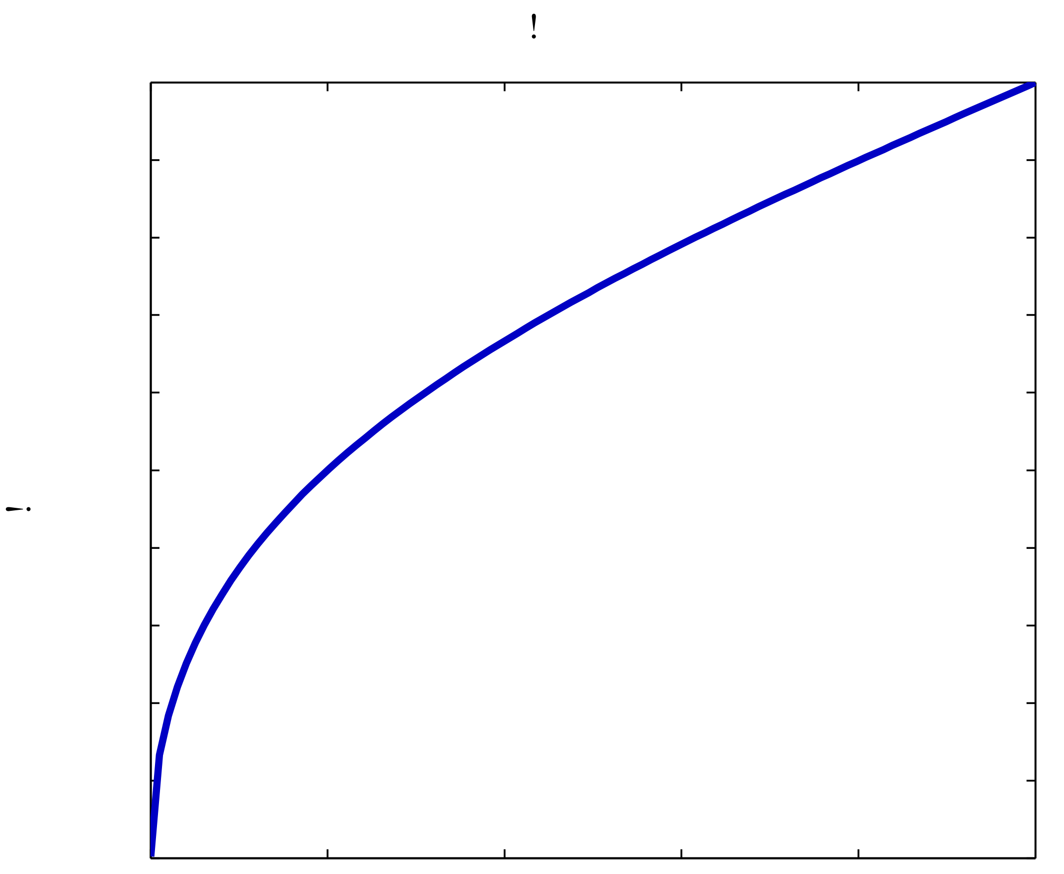
**Error exponents of ML decoders**

Non-asymptotic analysis of ML decoder & Normal approximation

# Chernoff tail bounds

# Combining and analyzing gives following result

# Analysis splits into high- and low-rate regimes



# Analysis of low rate regime

At sufficiently low rates, in particular if $R < D(\delta_{crit} \quad 0.5)$ then the

# Plot of exponent as a function of rate

# Generalization to arbitrary DMCs

We've seen $E_r(R)$ for the BSC is: (N.B. at start picked input distribution $P_x$ to be *Bern(0.5)*)

$$E_r(R) = min_{\delta \in [0,1]} D(\delta \| p) + |D(\delta \| 0.5) - R|^+$$
$$= min_{\delta \in [0,1]} D(\delta \| p) + |(1 - H_B(\delta)) - R|^+$$

worst case
channel behavior

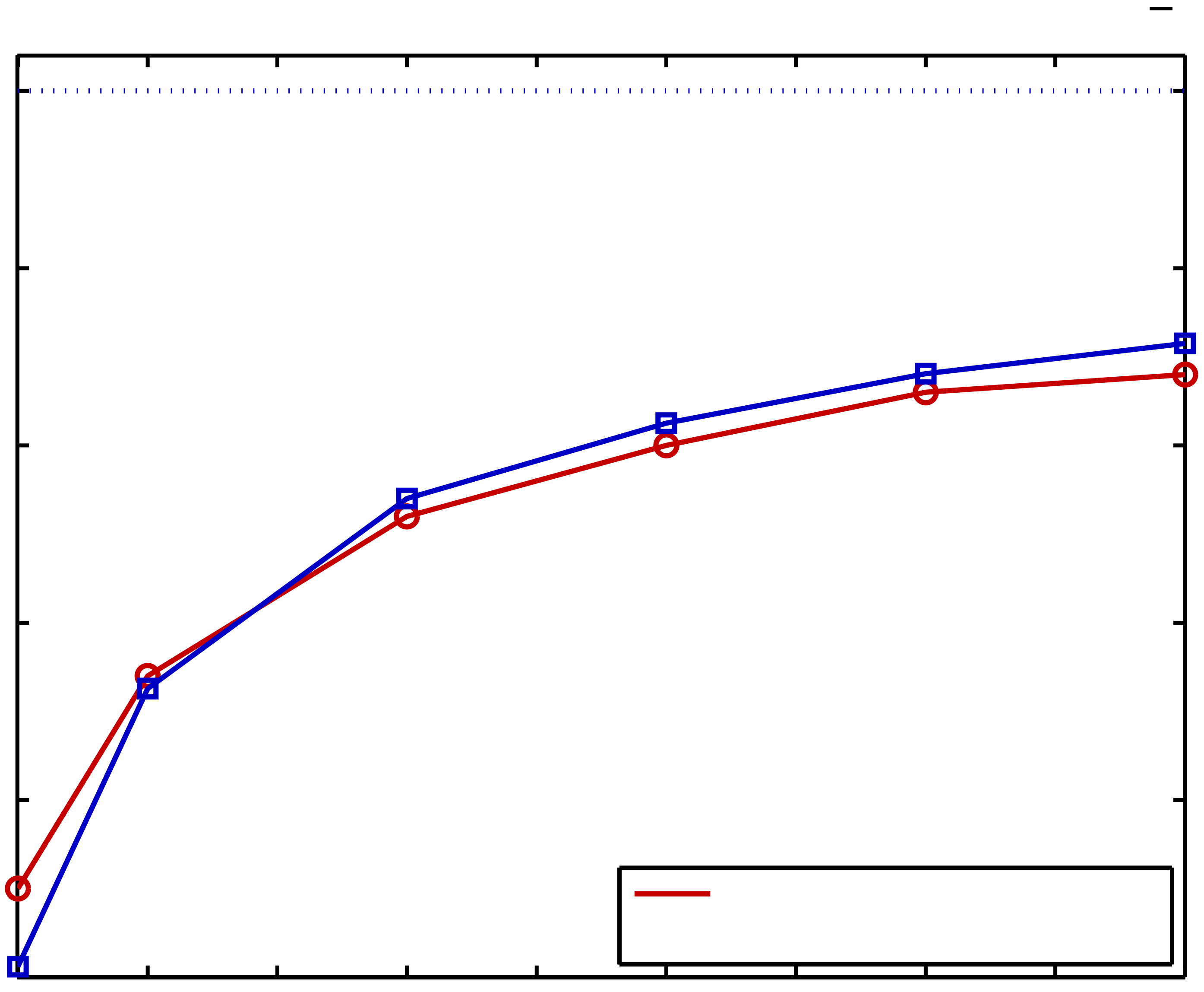
mutual info realized
across channel in worst case

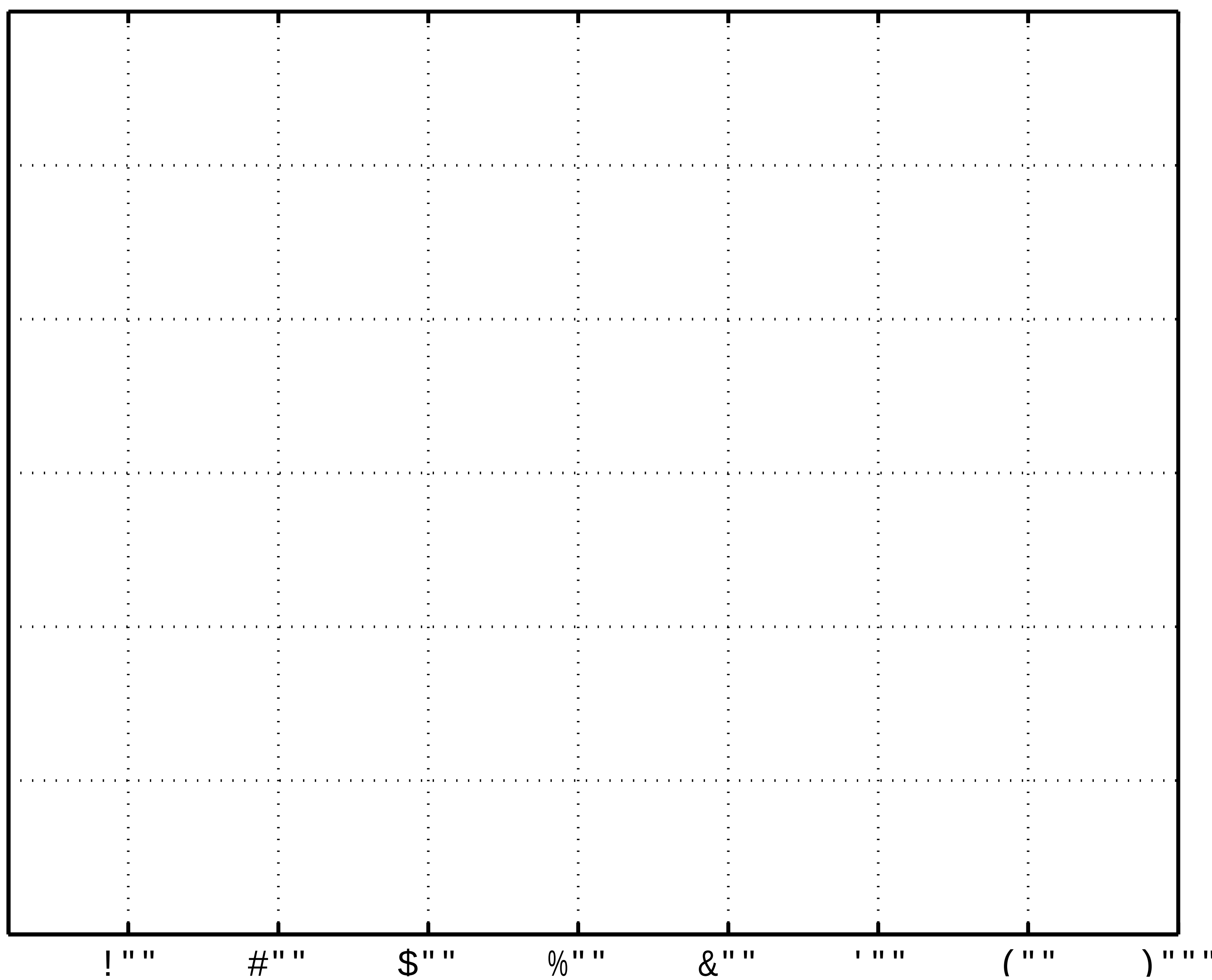
For general DMCs more complicated, but underlying idea is the same:

$$E_r(R) = max_{P_X} min_{V_{Y|X}} D(P_X V_{Y|X} \| P_X P_{Y|X}) + |I(P_X V_{Y|X}) - R|^+$$

worst case
channel behavior

mutual info realized
across channel in worst case





# Agenda

Analyzing decoding error for a bounded information decoder: regimes of interest

Error exponents of ML decoders

**Non-asymptotic analysis of ML decoder & Normal approximation**

# We branch the ML analysis down a different route

# We branch the ML analysis down a different route

# We branch the ML analysis down a different route

# Random coding union (RCU) bound

Remarks:

This bound is called the “random coding union bound” (RCU) in Polyanskiy-Poor-Verdú (PPV ’10)

Here we condition on the distance between the observation and Tx c.w. to get the coupling, taking the expectation of the conditional probability that some other codeword happens to be closer to the observation.

Holds for all block lengths

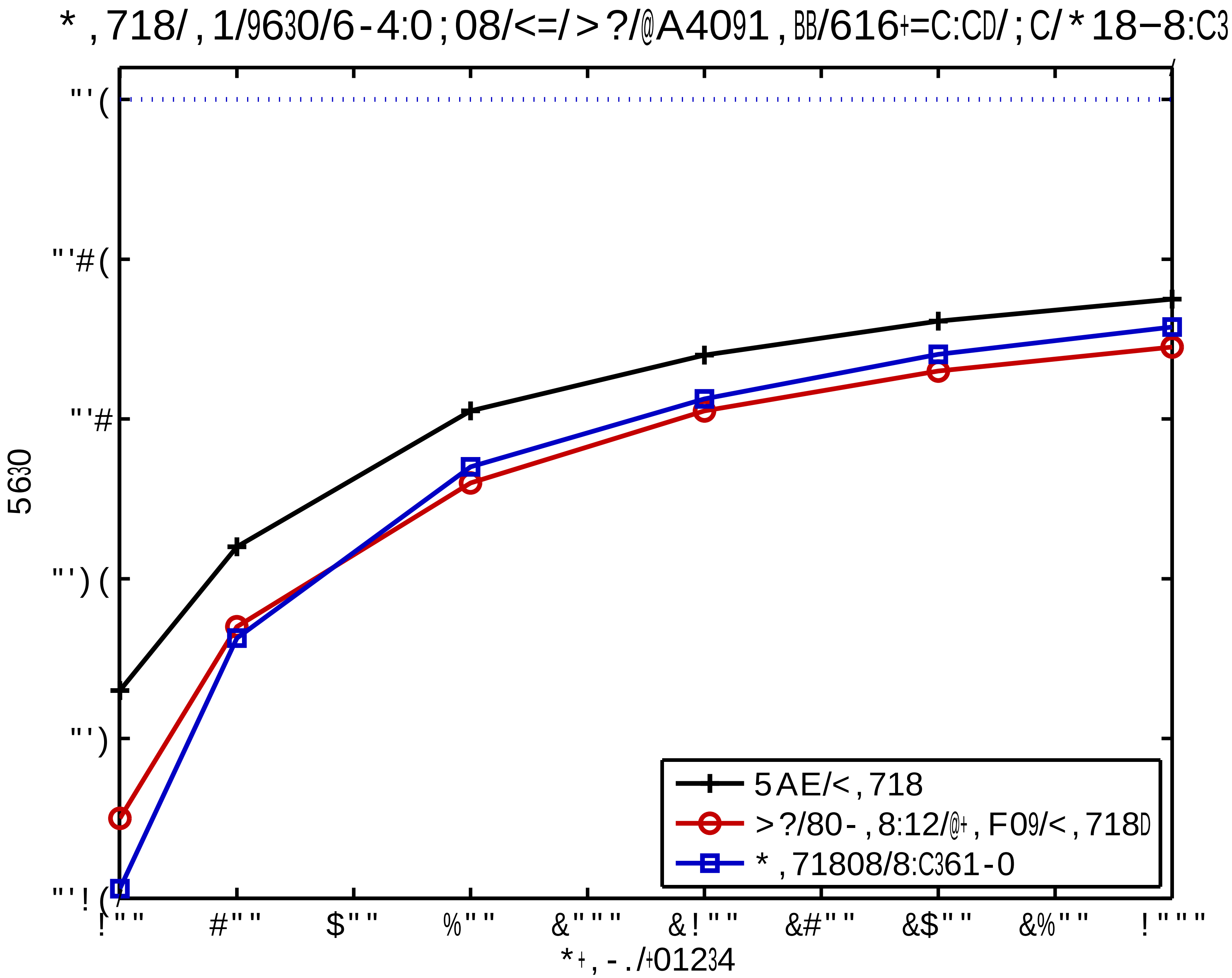
Have *not* applied a Chernoff bound

Can plot this bound for non-trivial block lengths

Aside: to plot for large *n* & *k*, it is better numerically to compute

$$\log \binom{n}{k} = \sum_{i=\max\{k,n-k\}}^{n} \log i - \sum_{i=1}^{\min\{k,n-k\}} \log i$$

# Compare: RCU, Bounded-dist, ML via Chernoff



Now we see the ML decoder (RCU bound) outperforms bounded distance.

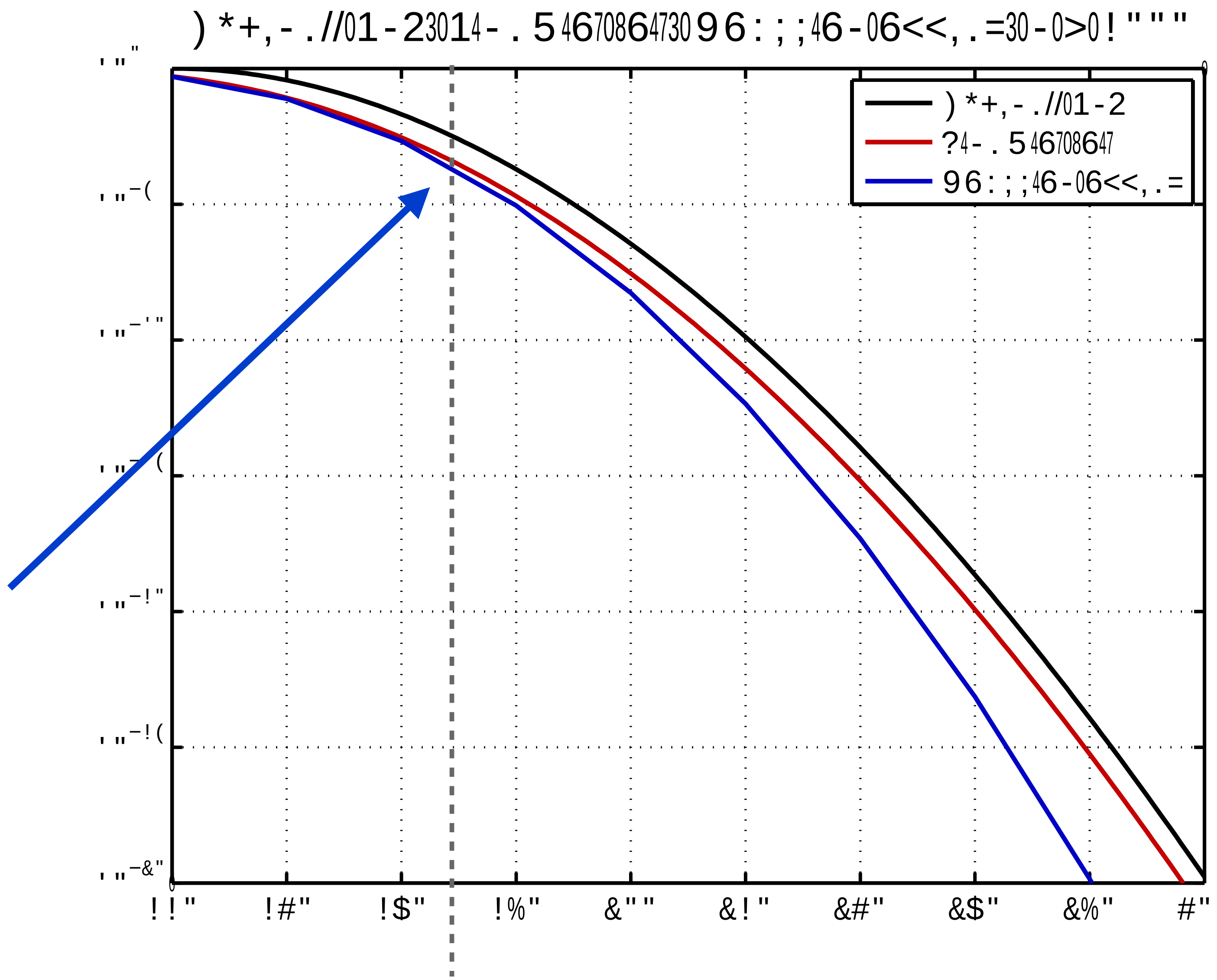
Can we characterize how we approach capacity as block length gets large?

Can we cleanly break into “outage” and “confusion” events as before?

# Recall CDF of Gaussian approximated better near mean

Will need to bound events more similar to those in the bounded distance dec.

In this regime a Gaussian approximation looks good, we need to correct the approximation to turn it into a bound



threshold 268
for Pr[err] = 0.001

# Turn Gaussian approx into a bound via Berry-Esseen

**Theorem (Berry-Esseen for i.i.d. Bernoulli r.v.)**

*For* $X_1, \ldots, X_n$ *i.i.d.* $Bern(p)$

$$\left| \Pr\left( \sum_{i=1}^{n} (X_i - \mu) \geq \lambda \sqrt{n\sigma^2} \right) - Q(\lambda) \right| \leq \frac{B}{\sqrt{n}}$$

*alternately*

$$\left| \Pr\left( \sum_{i=1}^{n} (X_i - \mu) \leq \lambda \sqrt{n\sigma^2} \right) - Q(-\lambda) \right| \leq \frac{B}{\sqrt{n}}$$

*where* $Q(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-y^2/2} dy$ *and* $\mu = p$, $\sigma^2 = p(1-p)$.

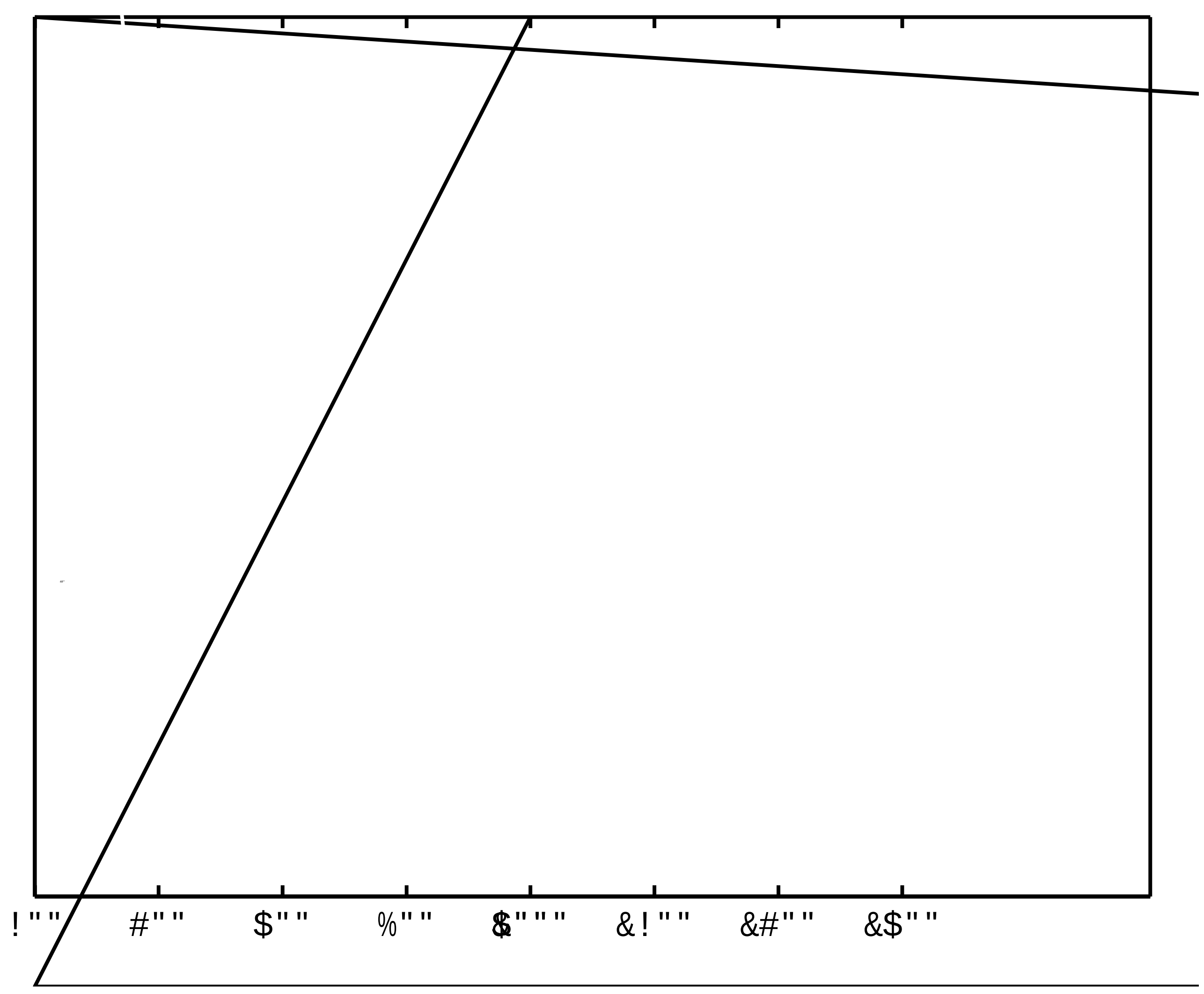
Note: (i) Bounds the absolute difference between CDF and tail of a Gaussian

(ii) Can be generalized to other (and non-i.i.d.) distributions

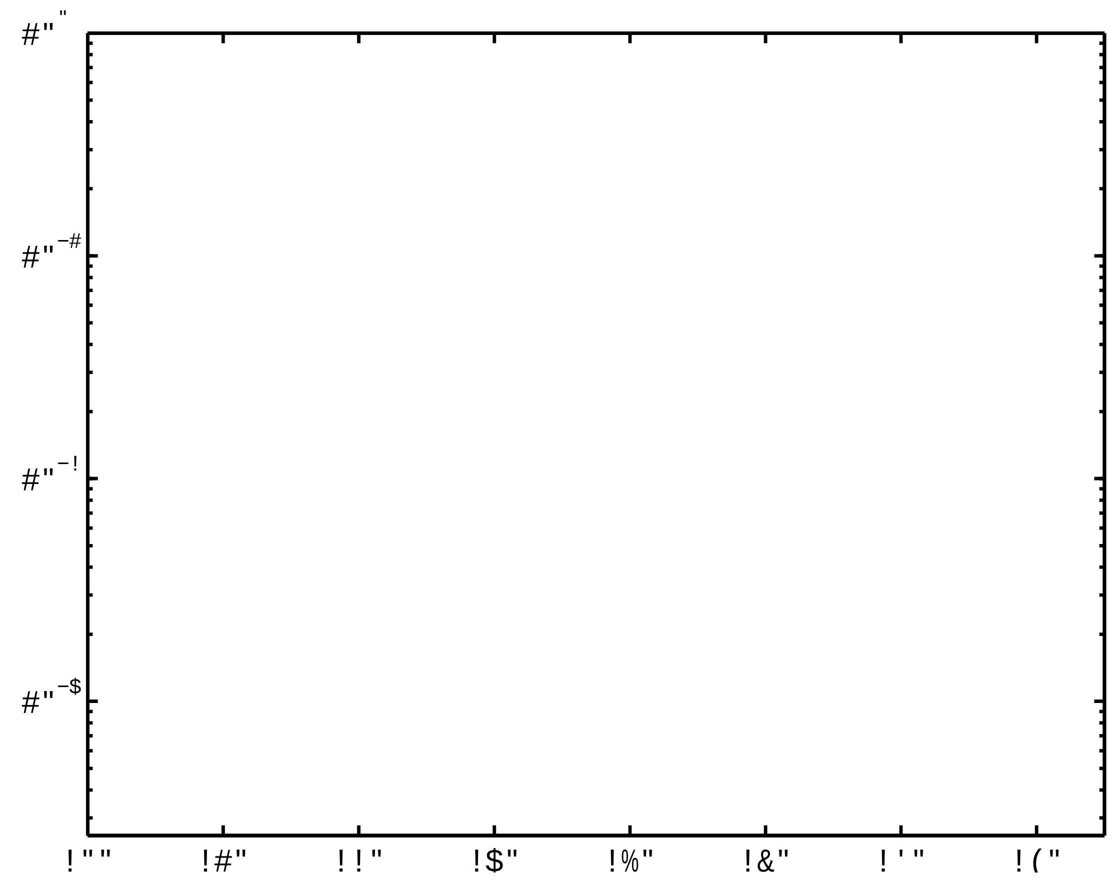
(iii) The Berry-Esseen constant "*B*" in this case is about 2.5

# As before split into “outage” and “confusion events”

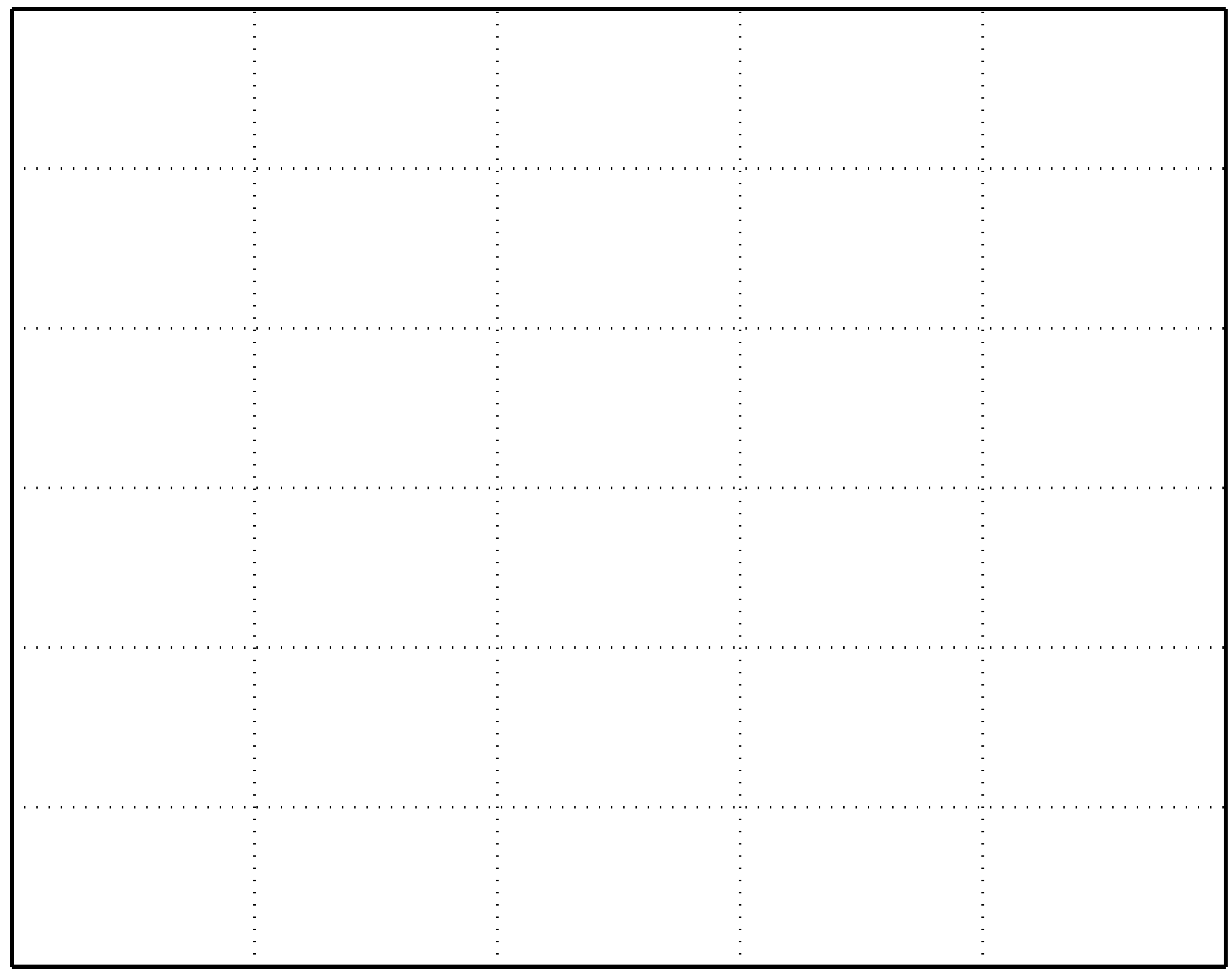
**Confusion**: some other c.w. too close



# A caveat



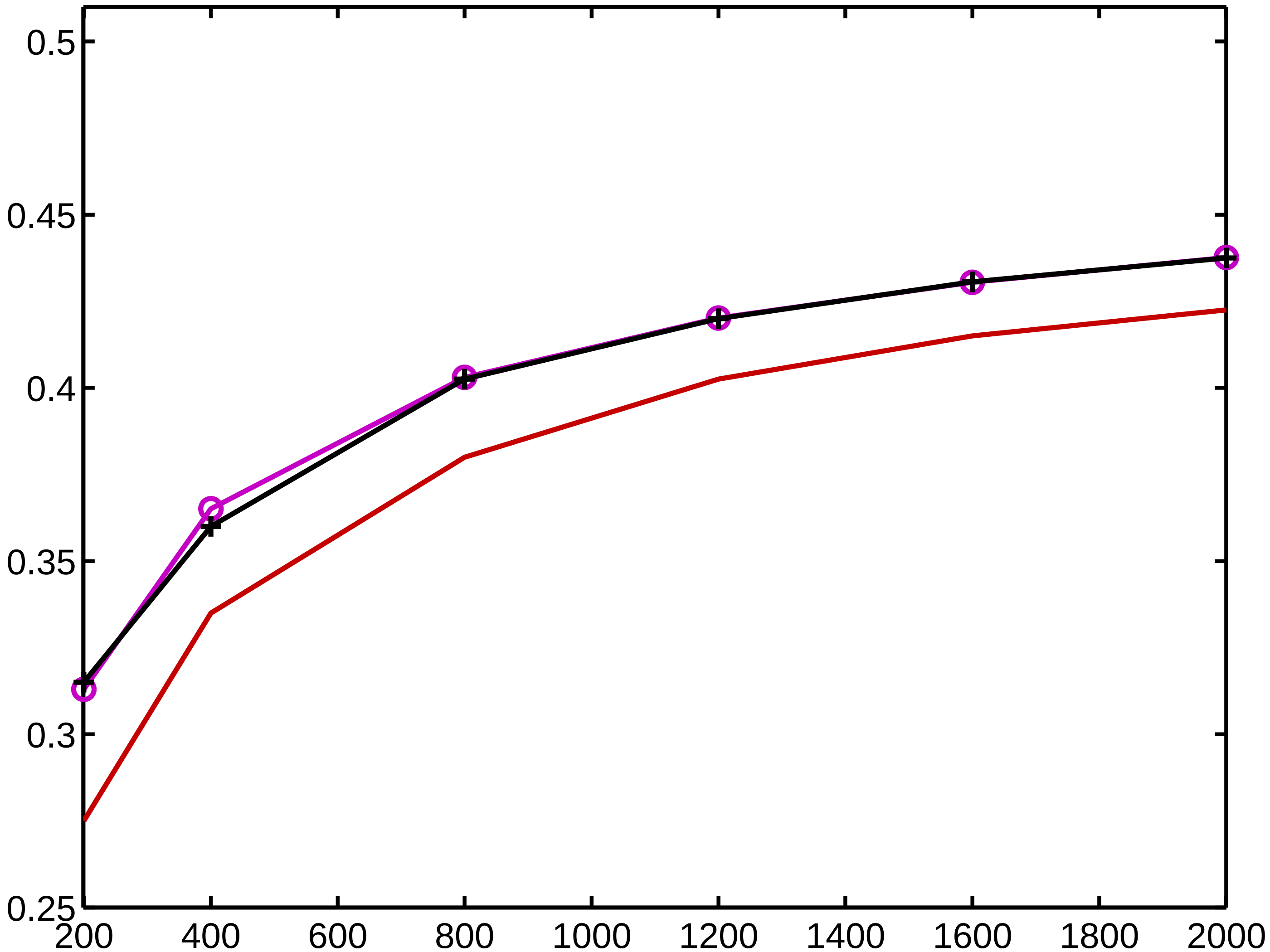
# Visualize the need to combine bounding techniques



$$2^{-n} \sum_{s=0}^{k} \binom{n}{s} \le 2$$

$$\log M \geq n - \log \binom{n}{K} - \log \frac{n-K}{n-2K}$$

# Compare approximation to earlier bounds



# Recap: three decoders

## Bounded distance

- Easy to analyze & gave us theme of analysis, split into "outage" and "confusion" events
- Identified regimes of interest: fixed rate and fixed Pr[err]
- Allowed us to see scaling of threshold

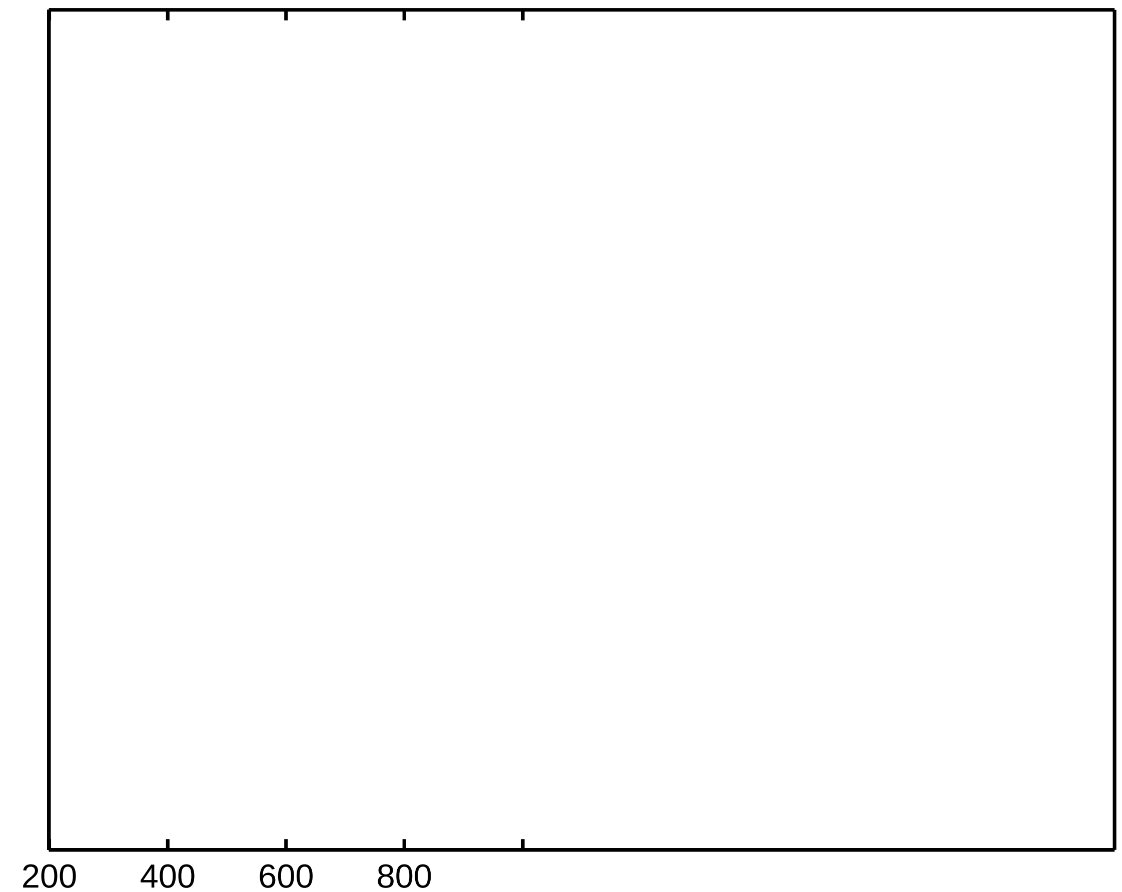
## ML analysis for fixed rate (error exponents)

- Introduced coupling between "outage" & "confusion"
- Applied Chernoff bounds to each

## ML analysis for fixed Pr[err] (non-asymptotic)

- Finer analysis of coupled event
- Got to the "normal" approximation: use Be5d edt1657cm BT 270 0 -2755.5511

# Talk objectives: recap


200
400
600
800