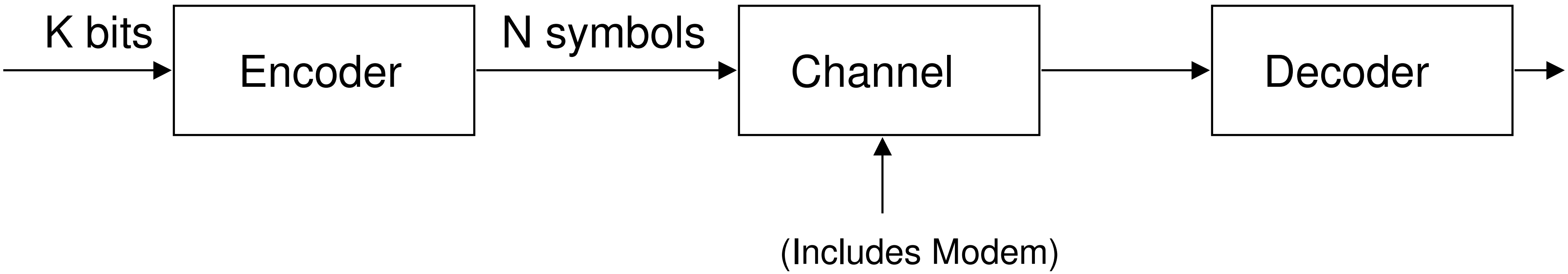
Shannon Theory and Turbo Decoding

Andrew J. Viterbi
Vice Chairman
QUALCOMM Incorporated

Shannon Day
May 18, 1998

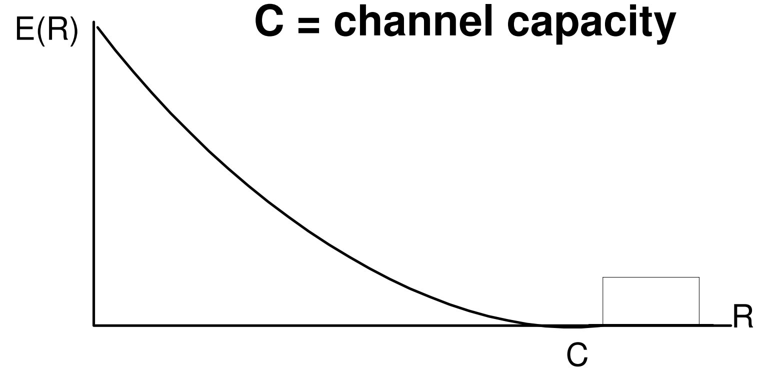
Shannon Channel Coding Fundamentals



For any memoryless channel -- such as the additive wideband Gaussian noise (AWGN) channel -- there exists a block code which encodes K bits into N symbols (Rate R=K/N bits/symbol) for which error probability is bounded by

$$\boxed{P_E < \exp[-NE(R)] \text{ where } E(R) > 0 \text{ for all } R < C}$$

C = channel capacity



C and E(R) depend only on the <u>channel</u> parameters

Central Problem of Channel Coding Theory for Past 50 Years

Encoding is <u>Easy</u>: Most codes are good provided they are <u>long</u> enough. Average P_E over all possible codes $\overline{P}_E < \exp[-NE(R)]$

Decoding is <u>Hard</u>: Exponentially increasingly with length, N. Decoder Complexity $\sim 2^K = 2^{RN}$

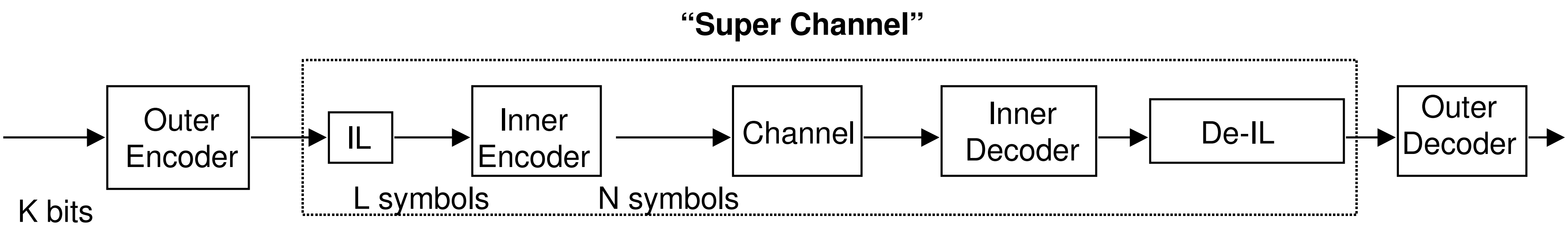
Some Partial Facilitators:

<u>Algebraic Block Codes</u>: Ease decoding, but suited primarily to hard decisions and do not nearly approach capacity.

<u>Convolutional Codes</u>: Performance-complexity tradeoff improved.

<u>Concatenated Codes</u>: "Divide and Conquer" Separate encoding and decoding into two or more levels.

Concatenated Codes "Divide and Conquer"



- **First proposed by Forney over 30 years ago.**
- **Interleaving and De-interleaving break up errors out of Inner Decoder which tend to be Bursty.**
- **Divide complexity between inner and outer decoding - complexity considerably reduced (Reed-Solomon outer code and convolutional inner code used by NASA, ESA, DSB.)**
- **<u>But</u> performance suffers because inner decoder makes "0" or "1" (HARD) decisions, destroying information usable by outer decoder.**

Concatenated Code Performance Improvements

Soft Decision Convolutional Decoder
(Inner Decoder generates Soft-Decision Output)

SOVA (Hagenauer): 1989
MAP (BJCR): 1974

MAP is superior: Soft decision is log - likelihood function for each symbol (output from inner and input to outer decoder)

$$\ln \frac{P(u_k = +1 | \tilde{y})}{P(u_k = -1 | \tilde{y})} \qquad \begin{array}{l} \text{"0"} \rightarrow +1 \\ \text{"1"} \rightarrow -1 \end{array}$$

$\tilde{y}$ **is entire sequence of received (soft-input) values**

Simplified MAP Decoder

Suppose Inner Code is Convolutionally Generated Block Code (Terminated to the "all-zero" state);

$$L_k = \ln \frac{P(u_k = +1 | \underset{\sim}{y})}{P(u_k = -1 | \underset{\sim}{y})} = \ln \frac{\sum_{\underset{\sim}{u}:u_k=+1} p(\underset{\sim}{u}, \underset{\sim}{y})}{\sum_{\underset{\sim}{u}:u_k=-1} p(\underset{\sim}{u}, \underset{\sim}{y})} = \ln \sum_{\underset{\sim}{u}:u_k=+1} e^{M(\underset{\sim}{u},\underset{\sim}{y})} - \ln \sum_{\underset{\sim}{u}:u_k=-1} e^{M(\underset{\sim}{u},\underset{\sim}{y})}$$

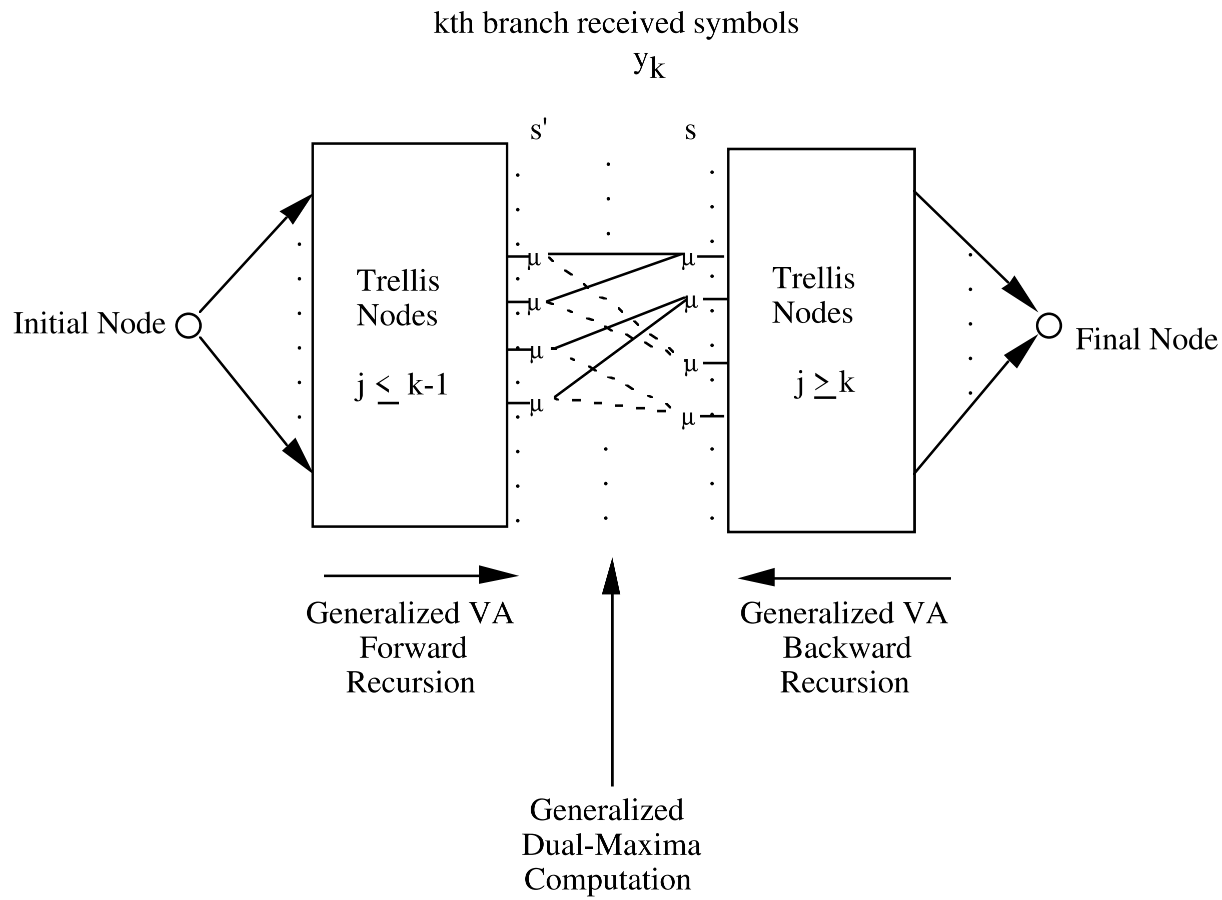
$$\text{where} \quad M(\underset{\sim}{u}, \underset{\sim}{y}) \overset{\Delta}{=} \ln p(\underset{\sim}{u}, \underset{\sim}{y})$$

$$L_k \approx \underset{\underset{\sim}{u}:u_k=+1}{\mathrm{Max}} M(\underset{\sim}{u}, \underset{\sim}{y}) - \underset{\underset{\sim}{u}:u_k=-1}{\mathrm{Max}} M(\underset{\sim}{u}, \underset{\sim}{y})$$

$$\boxed{\begin{aligned} &\ln(e^x + e^y) \\ &= \mathrm{Max}(x, y) + \ln\left(1 + e^{-|y-x|}\right) \\ &\overset{\Delta}{=} \mathrm{Max}^*(x, y) \\ &L_k = \mathrm{Max}^* M(\;) - \mathrm{Max}^* M(\;) \end{aligned}}$$

DUAL MAXIMA RULE

Simplified MAP Algorithm



$$a_k(s) = \max_{s'} \left[a_{k-1}(s') + c_k(s', s) \right]; \qquad b_{j-1}(s') = \max_{s} \left[b_j(s) + c_j(s', s) \right];$$

$$c_k(s', s) \overset{\Delta}{=} \ln p(y_k | s', s) + \ln P(s | s')$$

$$L_k \approx \max_{s', s: u_k = +1} \left[a_{k-1}(s') + c_k(s', s) + b_k(s) \right] - \max_{s', s: u_k = -1} \left[a_{k-1}(s') + c_k(s', s) + b_k(s) \right]$$

"Turbo Decoding" Soft Iterative Decoding of Concatenated Codes

Berrou et al (1993)	"Parallel Concatenated-Turbo" Codes
Benedetto et al (1996)	"Serial Concatenated" Codes

Decode <u>Inner</u> Code generating <u>Soft-decision Outputs</u> by MAP Algorithm.

Use them as Inputs to Outer code, but <u>also</u>

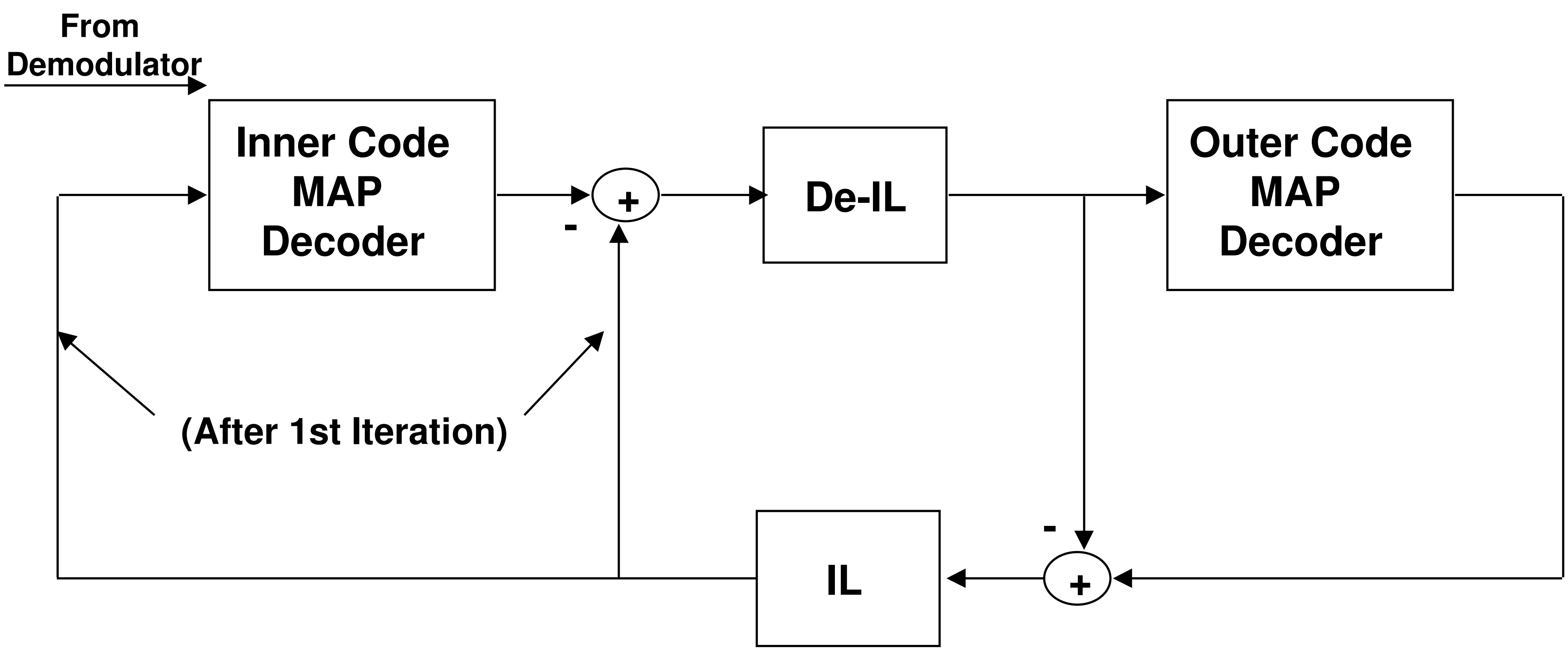
generate <u>Soft-decision Outputs</u> from Outer Code;

feed these back to Inner Code and repeat -- typically 8 to 16 iterations.

Result is <u>almost</u> <u>equivalent</u> to optimum (maximum likelihood)

decoding of composite (overall) code! [Why?: Theory Incomplete]

Iterative Decoding Algorithm (Berrou-Benedetto)



All Inputs and Outputs are Soft Decisions

Error Probability Performance

For any linear block code of 2K codewords $\underset{\sim}{x}_0, \underset{\sim}{x}_1 \cdots \underset{\sim}{x}_{2^K-1}$ and N symbols used over a memoryless channel (e.g. AWGN) with outputs $y_1, y_2, \ldots y_N$, maximum likelihood decoder achieves

$$P_E = \Pr\left[\bigcup_{j=1}^{2^K-1} \left\{M(\underset{\sim}{x}_j) > M(\underset{\sim}{x}_0)\right\} \Big| \underset{\sim}{y}\right]$$

where $M(\underset{\sim}{x}_j)$ is decoder metric for jth codeword, which is function of $\underset{\sim}{y}$.

"Uniform Error Property": error probability same no matter which codeword sent

Let d_j be weight of $\underset{\sim}{x}_j$

(Hamming distance between $\underset{\sim}{x}_j$ and $\underset{\sim}{x}_0$) $\left[\underset{\sim}{x}_0 = + + + \cdots +\right]$

Union Bound

$$P_E < \sum_{j=1}^{2^K-1} \Pr\left[M\left(\underset{\sim}{x}_j\right) > M\left(\underset{\sim}{x}_0\right) \Big| \underset{\sim}{y}\right]$$

For AWGN, $\Pr\left[M(\underset{\sim}{x}_j) > M(\underset{\sim}{x}_0) \Big| \underset{\sim}{y}\right] < e^{-\frac{E_s}{N_0} d_j} = Z^{d_j}$ where $Z = e^{-E_s/N_0}$

$$P_E < \sum_{j=1}^{2^K-1} Z^{d_j}\Big|_{z=e^{-E_s/N_0}}$$

$$P_E < \sum_{j=1}^{2^K-1} Z^{d_j} \Big|_{Z=e^{-E_s/N_0}}$$

Suppose there are K_d codewords of weight $d \leq N$

Then $$P_E < \sum_{d=d_{min}}^{N} K_d Z^d$$

Define Weight Enumeration Function (WEF):

$$\boxed{T(Z) \stackrel{\Delta}{=} \sum_{d=d_{min}}^{N} K_d Z^d \quad \Rightarrow \quad P_E < T(Z)\Big|_{Z=e^{-E_s/N_0}}}$$

Also useful Input-Output Weight Enumeration Function (IOWEF)

$$T(W,Z) = \sum_{d=d_{min}}^{N} \sum_{k=1}^{K} T_{kd} \, W^k Z^d$$

where T_{kd} = Number of codewords of output weight d and input weight k.

Then $T(Z) = T(1,Z)$

since $\sum_{k=1}^{K} T_{kd} = K_d$

Computation of T(W,Z) and T(Z) (General)

Example

Figure
(Viterbi: CDMA)

State Transition Matrix [C(W,Z)]

$$\begin{bmatrix} 1 & 0 & Z^2 & 0 \\ WZ^2 & 0 & W & 0 \\ 0 & Z & 0 & Z \\ 0 & WZ & 0 & WZ \end{bmatrix}$$

$$\begin{bmatrix} S^{(00)}(n) \\ S^{(01)}(n) \\ S^{(10)}(n) \\ S^{(11)}(n) \end{bmatrix} = [C(w,z)] \begin{bmatrix} S^{(00)}(n-1) \\ S^{(01)}(n-1) \\ S^{(10)}(n-1) \\ S^{(11)}(n-1) \end{bmatrix}$$

$$\begin{bmatrix} S^{(00)}(K) \\ S^{(01)}(K) \\ S^{(10)}(K) \\ S^{(11)}(K) \end{bmatrix} = [C(w,z)]^K \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$T(W,Z) = S^{(00)}(K;W,Z)$$

(Rather than matrix multiplication in 2 variables, can use recursion relations in 1 variable)

Computation of $T_c(W,Z)$ for Concatenated Codes

If outer code has $$T_0(W,Z') = \sum_{k=1}^{K} \sum_{\ell=1}^{L} A_{k\ell} W^k (Z')^\ell$$

and inner code has $$T_i(W',Z) = \sum_{\ell=1}^{L} \sum_{n=1}^{N} B_{\ell n} (W')^\ell Z^j$$

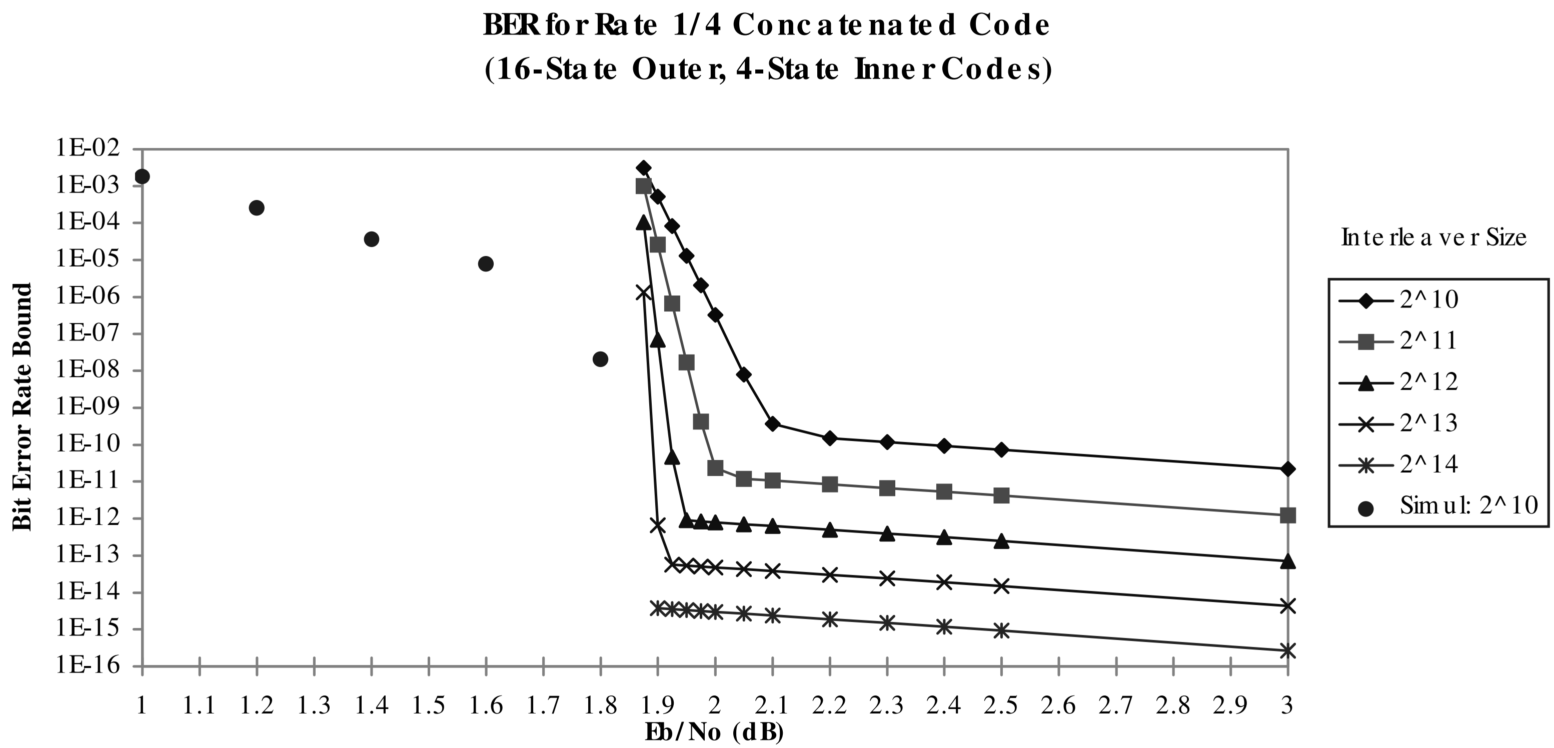
Then for concatenated code, averaged over ensemble of Interleavers of length L,

$$T_c(W,Z) = \sum_{\ell=1}^{L} \frac{A_\ell(W) B_\ell(Z)}{\binom{L}{\ell}}$$

where $A_\ell(W) = \sum_{k=1}^{K} A_{k\ell} W^k$ and $B_\ell(Z) = \sum_{n=1}^{N} B_{\ell n} Z^j$

$$P_E^{(C)} < T_c(1,Z) = \sum_{\ell=1}^{L} \frac{A_\ell(1) B_\ell(Z)}{\binom{L}{\ell}} \Bigg|_{Z=e^{-E_S/N_0}}$$

<u>Union Bound</u>



But Union Bound cuts off at E_b/N_0 corresponding to "computational Cutoff Rate" R_0 (1.9dB) well above value of E_b/N_0 corresponding to Capacity (-0.8dB)

Improved Union Bound (Gallager [1963])

Define "Rate-Weight" Function $\boxed{r(\delta) \triangleq \ln(K_d)/N}$

where K_d = No. of codewords of weight d

$\boxed{\delta \triangleq d/N}$ Normalized weight (or Hamming distance)

$$P_E < \sum_{d_{min}}^{d_{max}} P(d)$$

$$P(d) = \exp[-NE(\delta)]$$

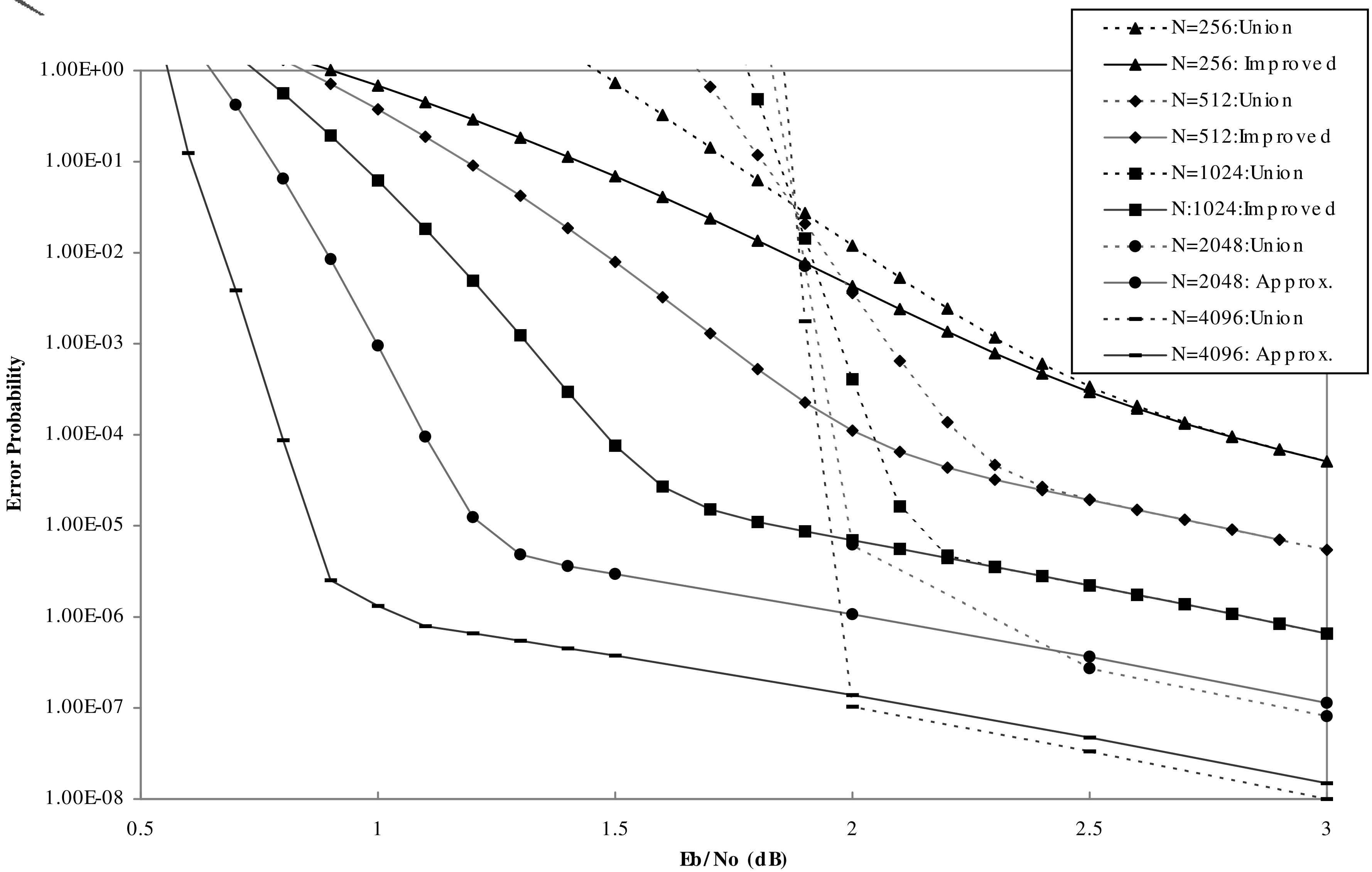
$$E(\delta) = \underset{0<\rho\le 1}{\mathrm{Max}} \left\{-\rho\left[r(\delta) + \delta \ln h(\rho) + (1-\delta)\ln g(\rho)\right] - (1-\rho)\ln[h(\rho) + g(\rho)]\right\}$$

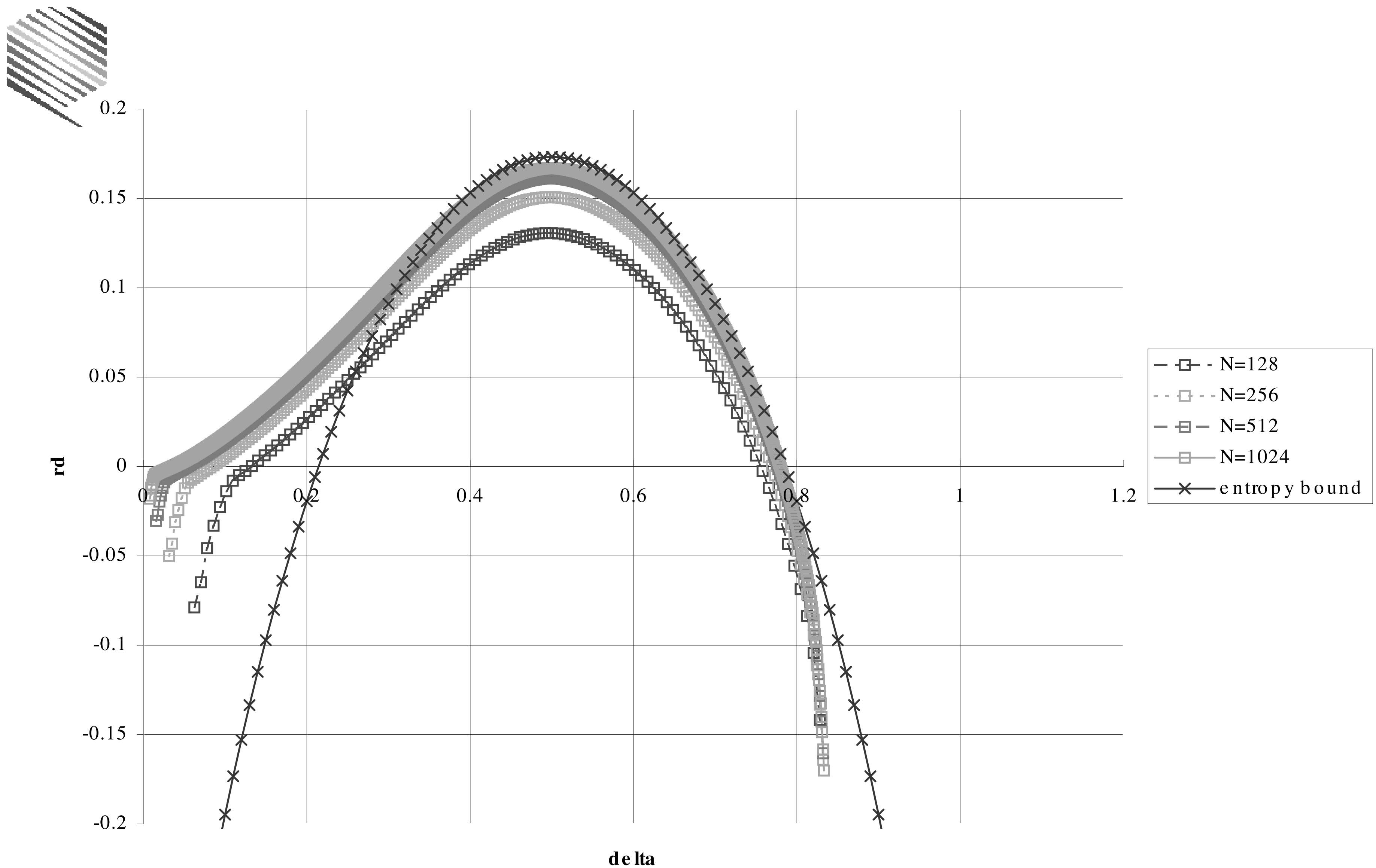
where $$h(\rho) = \int_{-\infty}^{\infty} \left[p_0(y)^{1/1+\rho} + p_0(-y)^{1/1+\rho}\right]^{-(1-\rho)} p_0(y)^{1/1+\rho} p_0(-y)^{1/1+\rho} dy$$

$$g(\rho) = \int_{-\infty}^{\infty} \left[p_0(y)^{1/1+\rho} + p_0(-y)^{1/1+\rho}\right]^{-(1-\rho)} p_0(y)^{2/1+\rho} dy$$

AWGN : $$p_0(y) \triangleq p(y|x_0) = \exp\left[-\left(y - \sqrt{2E_s/N_0}\right)^2/2\right]/\sqrt{2\pi}$$

Error Probability Bounds: 4-State Inner; 16-State Outer Concatenated Codes





0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
rd
0
0.2
0.4
0.6
0.8
1
1.2
delta
N=128
N=256
N=512
N=1024
entropy bound

Note that with "random coding" we replace K_d by

$$\bar{K}_d = 2^K \binom{N}{d} 2^{-N}$$

$$\bar{r}(\delta) \triangleq \frac{1}{N} \ln(\bar{K}_d) = \frac{1}{N}\left[(K-N)\ln 2 + \binom{N}{d}\right]$$

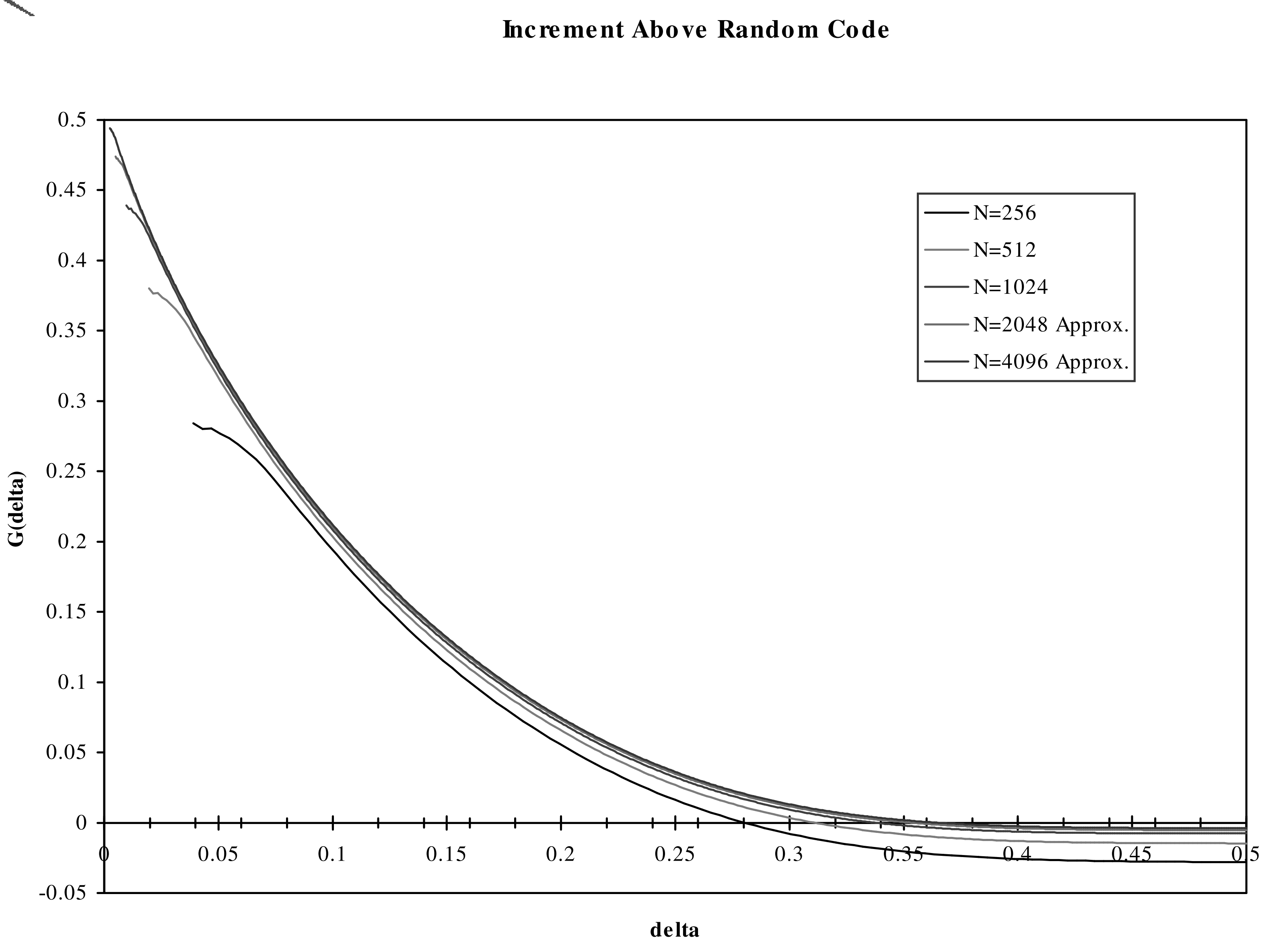
$$\approx -(1-R)\ln 2 + H(\delta) \text{ where } H(\delta) = -\delta \ln \delta - (1-\delta)\ln(1-\delta)$$

$$\text{and} \quad \delta = \frac{d}{N} \le 1$$

and with this $\bar{r}(\delta)$,

E $(\delta) > 0$ for all δ, provided R < C (thus achieves Shannon Capacity)

Define $G(\delta) \triangleq r(\delta) - \bar{r}(\delta)$, deviation from Random Coding Distribution.


Increment Above Random Code
G(delta)
delta
N=256
N=512
N=1024
N=2048 Approx.
N=4096 Approx.
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

To Relate Turbo Bound to Random Coding Bound

Can show $$E(\delta) = \underset{0<\rho\le 1}{\text{Max}} \left\{ \left[E_0(\rho) - \rho R \ln 2\right] + \rho\left[\delta \ln \frac{\delta}{\alpha(\rho)} + (1-\delta)\ln\frac{1-\delta}{1-\alpha(\rho)} - G(\delta)\right]\right\}$$

where

$$E_0(\rho) = \rho \ln 2 - \ln[g(\rho)] = \ln \int_{-\infty}^{\infty} \left[\frac{p_0(y)^{1/1+\rho}}{2}\right]^{1+\rho} dy \quad \left(\begin{array}{c}\text{Input} - \text{Binary} \\ \text{Output} - \text{Symmetric}\end{array}\right)$$

$$\alpha(\rho) = h(\rho)/[g(\rho) + h(\rho)]$$

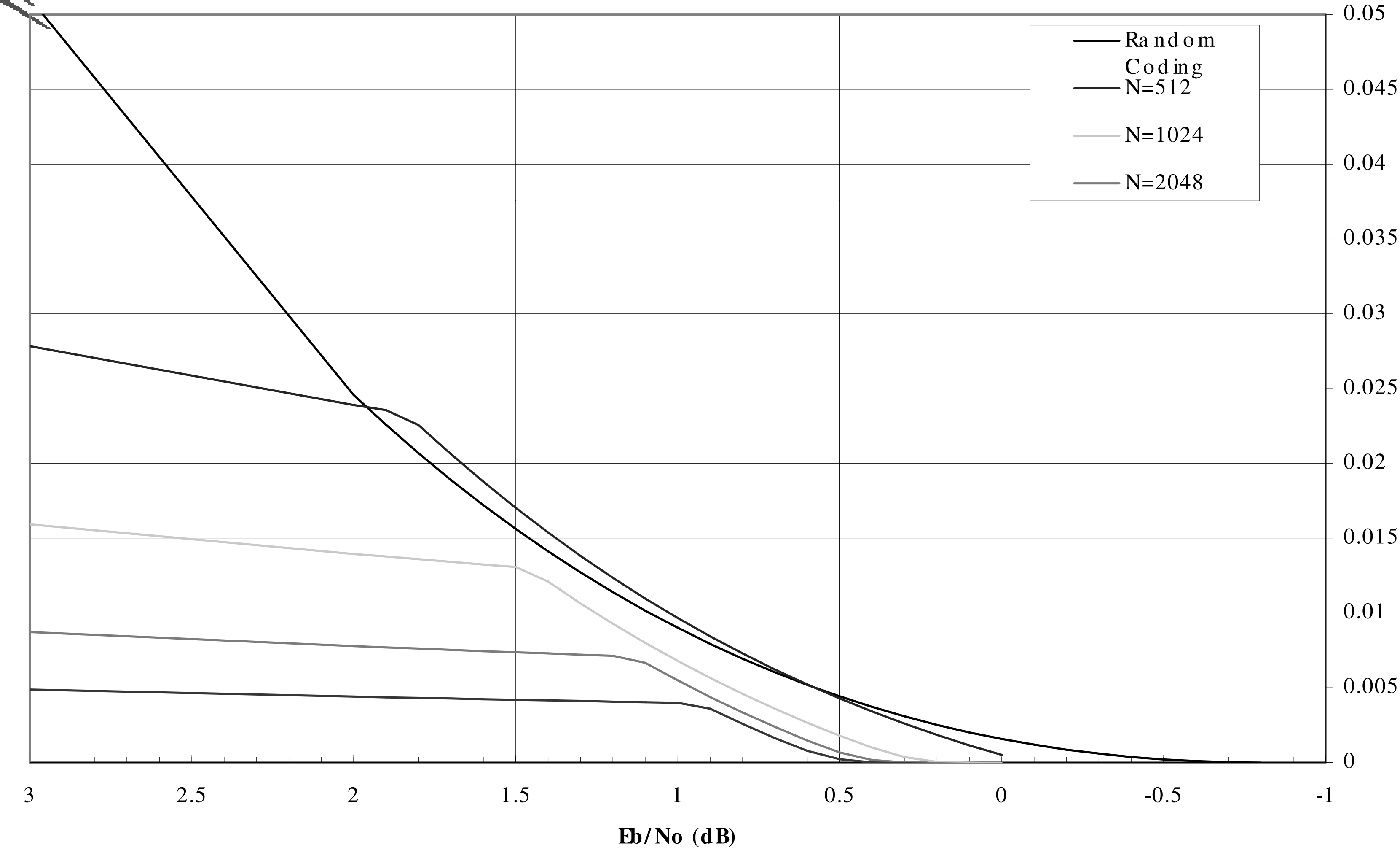
Capacity corresponds to $\underset{\delta}{\text{Min}}\ E(\delta) \to 0$

For Random Coding $G(\delta) = 0 \Rightarrow \underset{0<\delta<1}{\text{Min}}\ E(\delta) = \underset{0<\rho\le 1}{\text{Min}}\ \left[E_0(\rho) - \rho R \ln 2\right]$

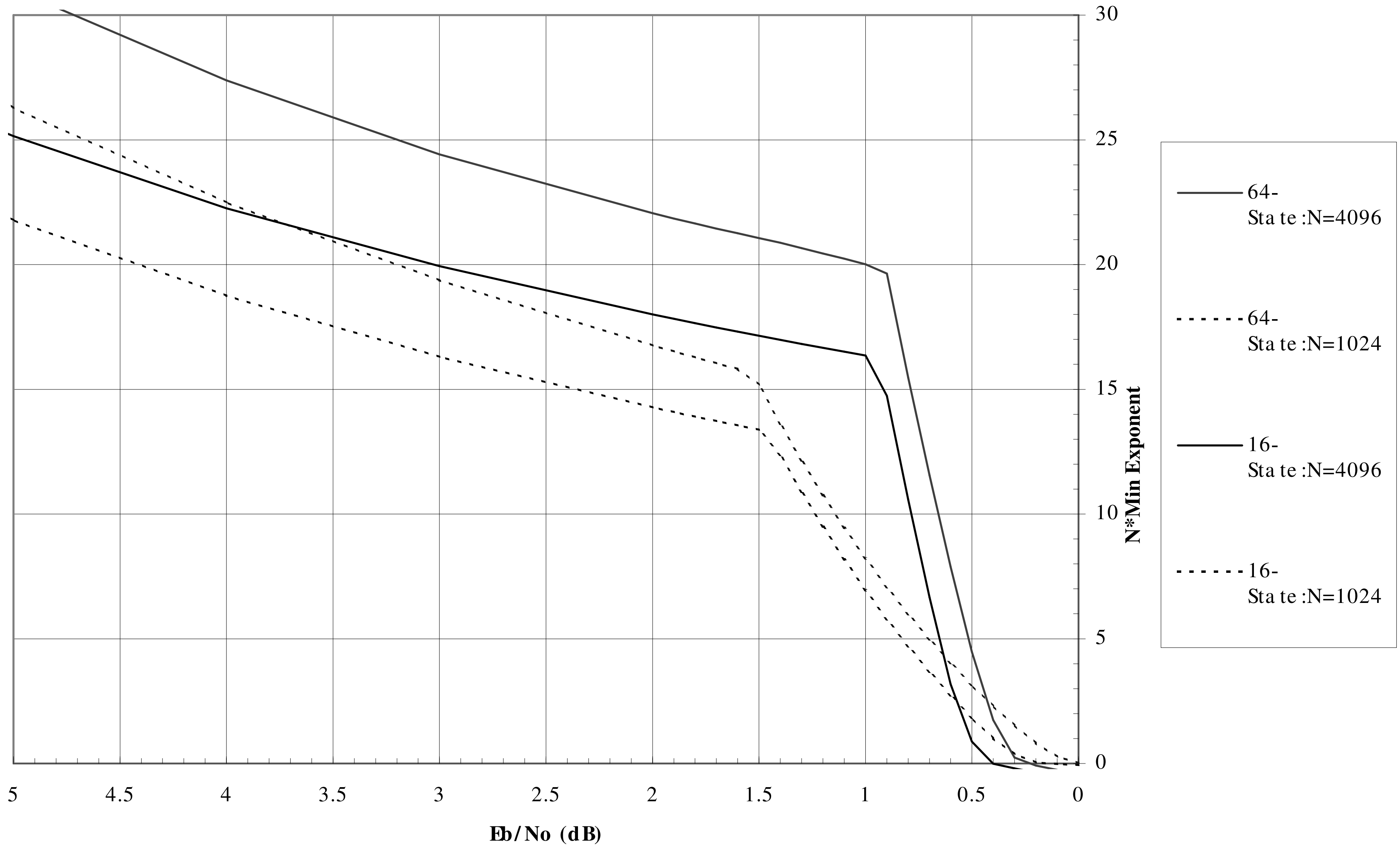
For Turbo, $\underset{\delta}{\text{Min}}\ \underset{\rho}{\text{Min}}\ \{\quad\}$ **Depends strongly on N**

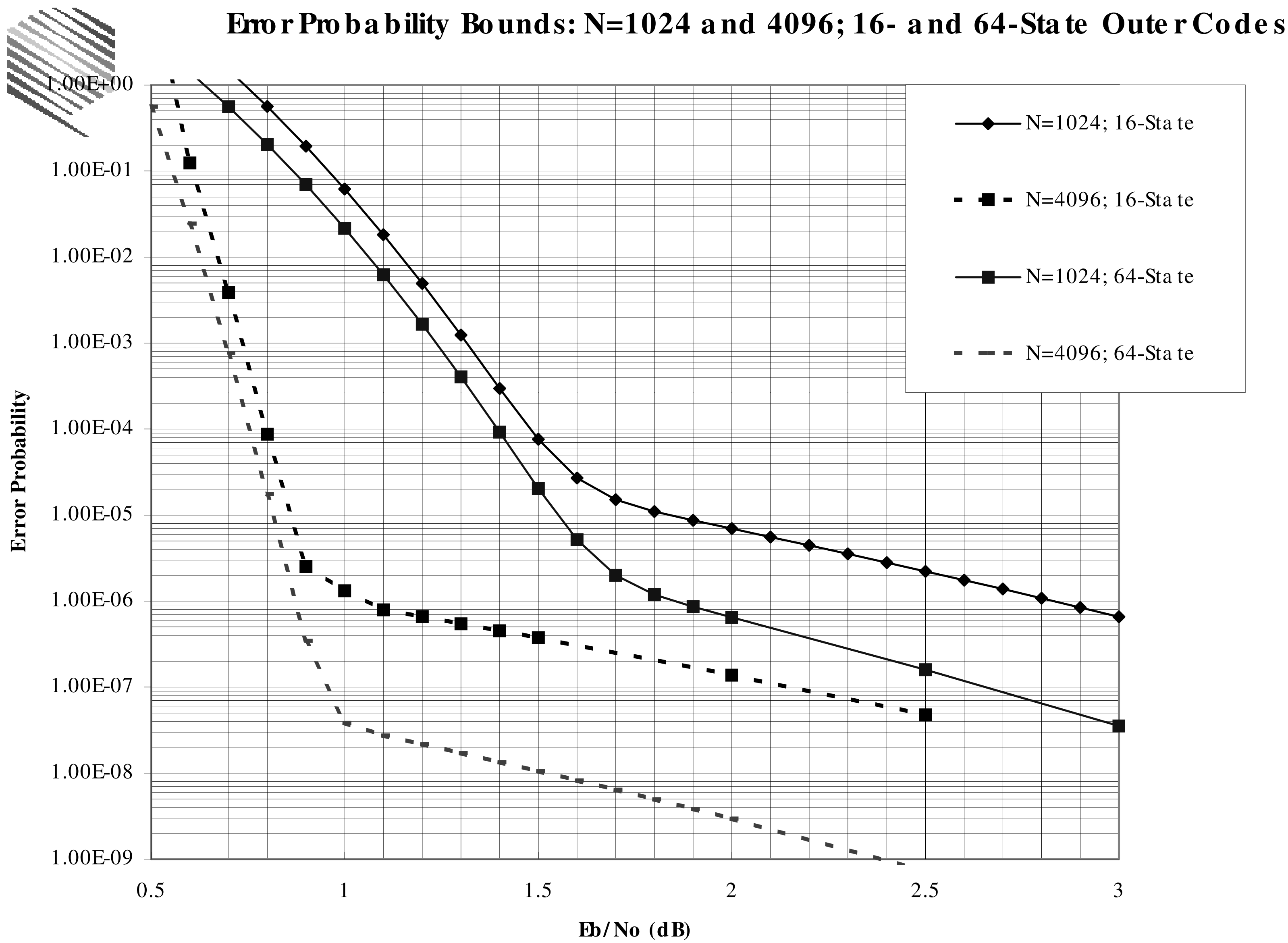
Low -δ "Tail" prevents obtaining general exponential bounds and causes "Error Floor" at high E_b/N_0.

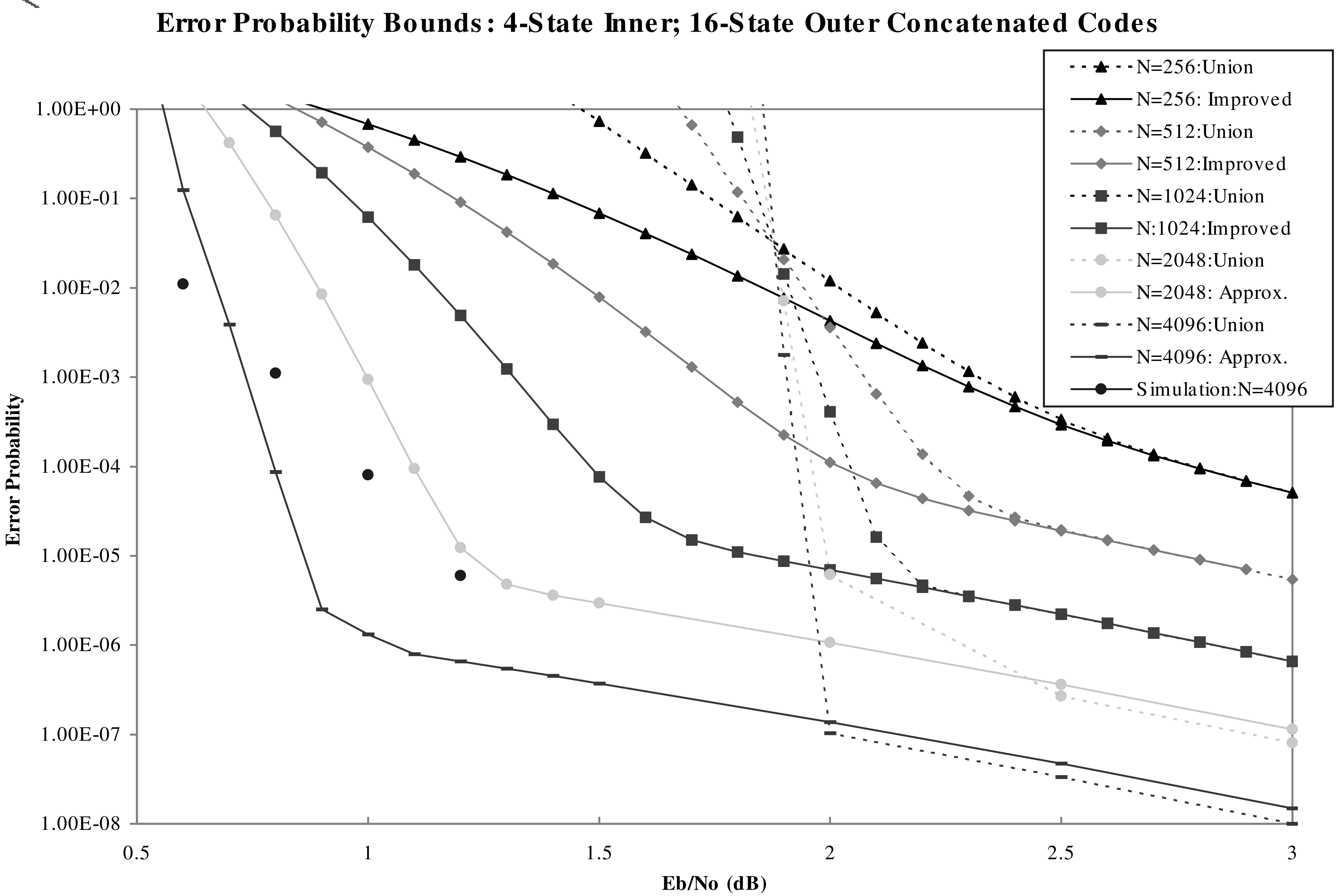
Minimum Exponent: 4-State Inner; 16-State Outer Codes



N * Min Exponent: N=1024 and 4096; 16- and 64-State Outer Codes




Error Probability Bounds: N=1024 and 4096; 16- and 64-State Outer Codes
Error Probability
1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06
1.00E-07
1.00E-08
1.00E-09
0.5
1
1.5
2
2.5
3
Eb/No (dB)
N=1024; 16-State
N=4096; 16-State
N=1024; 64-State
N=4096; 64-State



Error Probability Bounds: 4-State Inner; 16-State Outer Concatenated Codes
N=256:Union
N=256: Improved
N=512:Union
N=512:Improved
N=1024:Union
N:1024:Improved
N=2048:Union
N=2048: Approx.
N=4096:Union
N=4096: Approx.
Simulation:N=4096
1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
1.00E-05
1.00E-06
1.00E-07
1.00E-08
Error Probability
0.5
1
1.5
2
2.5
3
Eb/No (dB)