Coding Under Observation Constraints

Manish Bhardwaj and Anantha Chandrakasan

Digital Integrated Circuits and Systems Group
Microsystems Technology Laboratory
Massachusetts Institute of Technology

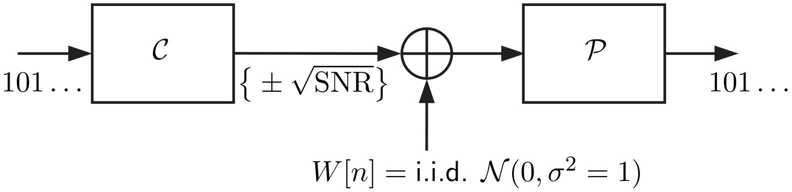
Allerton Conference on
Communication, Control, and Computing, 2007

Outline

- Problem definition and motivation
- Uncoded case
- Block coding: Asymptotic case
- Block coding: "Practical" case
- Summary

A Sampling of Prototypical Problems

Error Correcting Codes



Definition

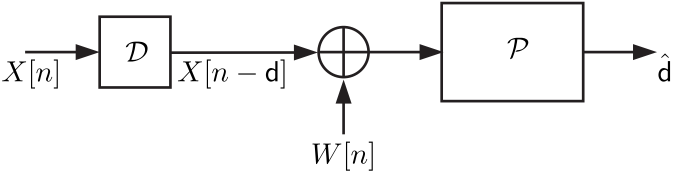
Rx-Cost(Procedure $\mathcal{P}$) $=$ Number of times $\mathcal{P}$ observes channel

Problem I: Coding

Given SNR and $p_{\text{bit-error}}$ minimize the cost per information bit (c_{bit})

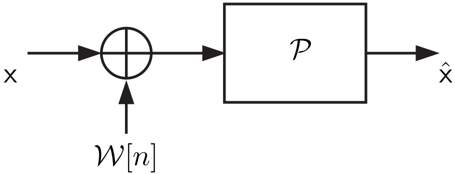
A Sampling of Prototypical Problems

Synchronization, SNR estimation



Problem II: Synchronization

Minimize cost of classifying d s.t. $\Pr(\hat{\mathsf{d}} \neq \mathsf{d}) \leq \epsilon_0$

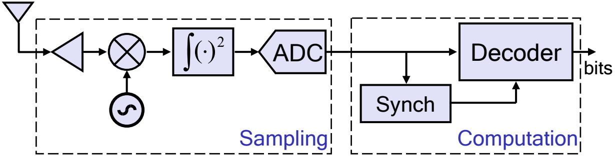


Problem III: Amplitude Estimation (in unknown noise variance)

Minimize cost of estimating x s.t. $\mathrm{E}\left[(\hat{\mathsf{x}}-\mathsf{x})^2\right] \leq \sigma_0^2$

An Application

Energy Conscious, Short Range and Low Data Rate UWB Systems



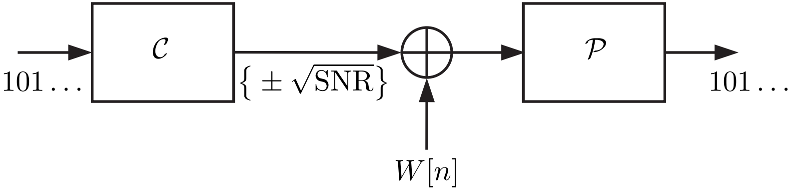
When is our cost model applicable in such systems?

- $\mathcal{E}_{\text{tx}} \ll \mathcal{E}_{\text{rx}}$
- $\mathcal{E}_{\text{rx}}$ must be dominated by sampling, not computation
- Available link-margin

Short Range, Low Data Rate UWB

- Impulse signals, BW$\geq$500 MHz, centered in 3-10 GHz
- UWB regulatory limits have led to $\mathcal{E}_{\text{tx}} \ll \mathcal{E}_{\text{rx}}$
- Recent receiver in our lab demonstrates very low overhead
- Evidence of willingness to trade-off margin for energy

Coding Under Rx Cost Constraints



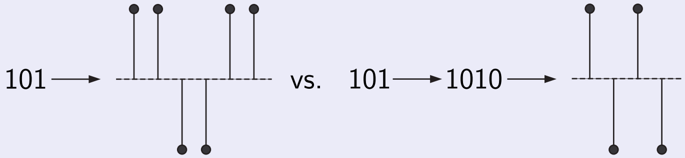
Definitions

- A (n, k) code $\mathcal{C}$
  - $r_{\mathcal{C}} = k/n$
  - $E_b/N_0 = \text{SNR}/(2r_{\mathcal{C}})$
  - $\gamma(p_{\mathsf{b}}) = E_b(\mathcal{U}, p_{\mathsf{b}})/E_b(\mathcal{C}, p_{\mathsf{b}})$
- $r_{\text{rx}} = k/\text{E}[\mathsf{N}(\mathcal{P}, k)]$
  - $\mathsf{N}(\mathcal{P}, k)$: Number of observations required to decode k bits
- $r_{\text{rx}} = r_{\mathcal{C}}$ for traditional systems
- $c_{\mathsf{bit}} = \text{SNR}/(2r_{\text{rx}}) \geq (2^{2r_{\text{rx}}} - 1)/(2r_{\text{rx}})$

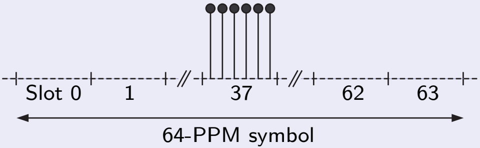
Rx Cost Under Traditional Coding

Example: (4,3) Parity Check Code

- SNR: 10.5 dB, $\gamma(10^{-6}) = 1.5$ (1.7 dB)



Example: (64,6) Orthogonal Code (PPM)



- 64-PPM: Half the E_b but 32x the observation time!
- Coding gain not always synonymous with Rx cost reduction

Simplest Case: No Coding

What $\mathcal{P}$ Minimizes Cost?

Sequential Detector $\mathcal{P}$

Specified via two rules: Termination and Decision

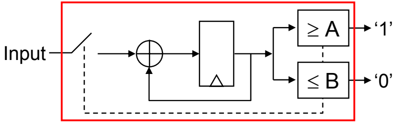
Wald's Sequential Probability Ratio Test (SPRT)

$$\ell(\mathbf{y}_n) = \sum_{i=1}^{n} \ell(\mathrm{y}_i) = \sum_{i=1}^{n} \ln \frac{p(\mathrm{y}_i \mid \theta_1)}{p(\mathrm{y}_i \mid \theta_2)} \begin{cases} \geq A & \text{Accept } \theta_1 \\ \leq B & \text{Accept } \theta_2 \\ \in [B, A] & \text{Continue sampling} \end{cases}$$

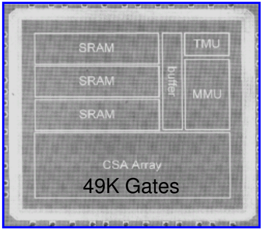
SPRT is optimal

If $\mathcal{P}'$ realizes $(\varepsilon_1' \leq \varepsilon_1, \varepsilon_2' \leq \varepsilon_2)$ then $\mathrm{E}_1[\mathrm{N}'] \geq \mathrm{E}_1[\mathrm{N}]$ and $\mathrm{E}_2[\mathrm{N}'] \geq \mathrm{E}_2[\mathrm{N}]$

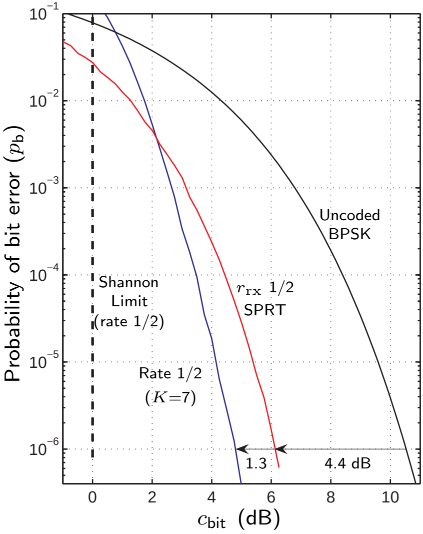
Performance of a SPRT Detector



SPRT Decoder



Rate 1/2, K=7 Viterbi ASIC
Optimized for Low-Power



SPRT: Asymptotic Performance

Asymptotic Behavior $(\mathrm{E}[\mathsf{N}] \to \infty, \varepsilon \to 0, r_{\mathrm{rx}} \to 0)$

$$\mathrm{E}_i[\mathsf{N}] \approx \frac{-\ln \varepsilon_i}{\mathcal{D}(\mathrm{P}_i \parallel \mathrm{P}_j)}$$

Asymptotic ‘Sequential’ Gain

Binary AWGN Channel:

$$\gamma_{\mathsf{s}} = \frac{\mathcal{D}(\mathcal{N}(-a, \sigma^2) \parallel \mathcal{N}(a, \sigma^2))}{\mathcal{D}(\mathcal{N}(0, \sigma^2) \parallel \mathcal{N}(a, \sigma^2))} = 4 \equiv 6 \text{ dB}$$

Binary Symmetric Channel:

$$\gamma_{\mathsf{s}} = \frac{\mathcal{D}(\mathcal{B}(p) \parallel \mathcal{B}(1-p))}{\mathcal{D}(\mathcal{B}(\frac{1}{2}) \parallel \mathcal{B}(1-p))} = \begin{cases} 2 & p \to 0 \\ 4 & p \to \frac{1}{2} \end{cases}$$

Applying Sequential Detection to Decoding

Information Bits					Parity Bits				
b_0	b_1	$\cdots$	b_{k-2}	b_{k-1}	c_0	c_1	$\cdots$	c_{t-2}	c_{t-1}
b_0	b_1	$\cdots$	b_{k-2}	b_{k-1}	c_0	c_1	$\cdots$	c_{t-2}	c_{t-1}
b_0	b_1	$\cdots$	b_{k-2}	b_{k-1}	c_0	c_1	$\cdots$	c_{t-2}	c_{t-1}
b_0	b_1	$\cdots$	b_{k-2}	b_{k-1}	c_0	c_1	$\cdots$	c_{t-2}	c_{t-1}

$\vdots$

Problem

What $\mathcal{P}$ minimizes p_{b} given $\mathcal{C}$, r_{rx} and SNR?

Optimum Experiment Design

Problem

Given a set of experiments $\mathscr{E}$ device a strategy that minimizes $\mathrm{E}_i[\mathsf{N}]$ subject to $\mathrm{P}_i[\hat{\theta}(\mathbf{y}_\mathsf{N}) \neq \theta_i] \leq \bar{\varepsilon}_i$

- Definitions

$$\mathrm{P}_{\theta;e} := p(\mathsf{y}_e \mid \theta; e)$$
$$\mathcal{D}_e(\theta_i, \theta_j) := \mathcal{D}(\mathrm{P}_{\theta_i;e} \parallel \mathrm{P}_{\theta_j;e})$$

Chernoff's 'Procedure A' is Asymptotically Optimum

I. Calculate ML estimate $\hat{\theta}$

II. Pick $e_{t+1} = \arg\sup_{e \in \mathscr{E}^*} \inf_{\theta \neq \hat{\theta}} \mathcal{D}_e(\hat{\theta}, \theta)$

III. Terminate when $\ell(\hat{\theta}) - \max_{\theta \neq \hat{\theta}} \ell(\theta) > a$

$$\mathrm{E}_i[\mathsf{N}] \sim \frac{-\ln \varepsilon_i}{\mathcal{D}_i} \text{ where } \mathcal{D}_i := \sup_{e \in \mathscr{E}^*} \inf_{\theta_j \neq \theta_i} \mathcal{D}_e(\theta_i, \theta_j)$$

An Example of Procedure A

Detecting a Biased Coin

Problem

What is the quickest way to detect a biased coin in a heap of 3 coins?

Strategy

- Let θ_1 be our best guess so far. Then the distance matrix is,

	e_1	e_2	e_3
θ_2	a	b	0
θ_3	a	0	b

- $a = \mathcal{D}(\mathcal{B}(p) \parallel \mathcal{B}(\frac{1}{2}))$, $b = \mathcal{D}(\mathcal{B}(\frac{1}{2}) \parallel \mathcal{B}(p))$
- If $a > \frac{b}{2}$ toss 1 again, else toss 2 and 3 with equal probability.

E.g. a (0.1,0.9) coin

- $a = 0.36$, $b = 0.51 \Rightarrow$ toss the best guess again.
- $\approx$ 38 tosses to get to a reliability of 10^{-6}.

Decoding via Procedure A

Optimum $\mathcal{P}$ for Decoding $\mathcal{C}(n,d)$ as $r_{\mathrm{rx}} \to 0$

- $\theta_{i=1,2,\ldots,M}$ correspond to codewords c_i
- $\mathscr{E} = \{e_{i=1,2,\ldots,n}\}$, e_i corresponds to observing i^{th} coded bit

Distance Matrix $\boldsymbol{D}^{(1)}$ under $\hat{\theta} = \theta_1$

	e_1	e_2	e_3	$\cdots$	e_{n-1}	e_n
θ_2	$*$	$*$	0	$\cdots$	$*$	0
θ_3	0	$*$	$*$	$\cdots$	$*$	0
$\vdots$				$\ddots$		
θ_M	0	0	$*$	$\cdots$	0	$*$

$(* = \mathcal{D}_b := \mathcal{D}(\mathrm{P}_0 \parallel \mathrm{P}_1))$

Linear Program to Determine $\mathcal{D}_i$

$$\max \ \mathcal{D}_i \ \text{ subject to } \ \boldsymbol{D}^{(i)} \cdot \boldsymbol{\lambda} \geq \mathcal{D}_i$$

Solutions for Classes of Codes

Linear, Cyclic, Constant Projection

Observations

- Equiprobable sampling achieves $\mathcal{D}_i = \frac{d_i}{n} \cdot \mathcal{D}_b$
- Two classes (not necessarily linear) where this is optimum
    I. Cyclic Codes
    II. 'Constant Weight Projection' Codes

Conjecture: Equiprobable sampling over *some* deletion is optimum

Use the coordinate deletion q that maximizes $\left(\frac{d}{n}\right)_{|q}$

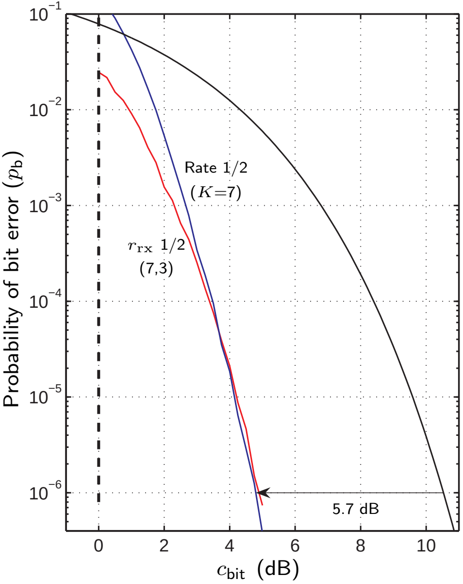
Asymptotic Gain from Equiprobable Sampling

$$\varepsilon \sim e^{-\mathrm{E}[\mathsf{N}]\frac{d}{n}\mathcal{D}_b} = e^{-\mathrm{E}[\mathsf{N}_b]k\frac{d}{n}\frac{\mathcal{D}_b}{\mathcal{D}_{\mathsf{unc}}}\mathcal{D}_{\mathsf{unc}}} = e^{-\mathrm{E}[\mathsf{N}_b]\gamma_{\mathsf{c}}\gamma_{\mathsf{s}}\mathcal{D}_{\mathsf{unc}}}$$

A Simple Example

(7,3,4) Dual Hamming Code

- $\mathcal{P}_\mathrm{s}$ leads to a gain of 0.9 dB
- $\mathcal{P}_*$ offers further 0.4 dB gain
- Achieves Rx cost of 802.15.4a
  - 15.4a: $\mathrm{RS}_6(53,43)$ outer, and rate 1/2, K=3 inner code
- (3,2) with r_rx=1/4 using $\mathcal{P}^*$ also matches 15.4a
- Cf. Yamamotoh et al.
  - (2,1,5) with decision feedback
- Differences
  - Decision vs. perfect feedback
  - Throw away vs. aggregate



Non-Asymptotic Strategies

Key Drawback of Procedure A

- Asymptotic argument ignores initial phase when $\hat{\theta}$ is poor
- Alternate strategies factor in confidence in estimates

Alternate Design Strategies

Let $\Pi_j := \Pr\big[\theta = \theta_j \mid \{\mathsf{y}\}\big]$, then pick e_{t+1} as

$$\arg\max_{e \in \mathscr{E}} \sum_j \Pi_j \mathcal{D}_e(\hat{\theta}, \theta_j) \qquad \text{[Proc B]}$$

$$\arg\max_{e \in \mathscr{E}} \sum_j \sum_k \Pi_j \Pi_k \left[\mathcal{D}_e(\theta_j, \theta_k) + \mathcal{D}_e(\theta_k, \theta_j)\right] \qquad \text{[Box-Hill]}$$

Applications to Decoding

- Notation

$$\Pi_j := \Pr\big[\mathbf{x} = \boldsymbol{x}_j \mid \{\mathrm{y}\}\big] \qquad j = 1, 2, \ldots, M$$

$$\pi_i := \Pr\big[\mathrm{x}_i = 1 \mid \{\mathrm{y}\}\big] = \sum\nolimits_{j \in J_i(1)} \Pi_j \qquad i = 1, 2, \ldots, n$$

Decoding Strategies

$$i^* = \arg\max_i \ \mu_i$$

$$\mu_i^{\mathsf{B}} = \Pr[\mathrm{x}_i = \neg(\hat{\mathbf{x}})_i \mid \{\mathrm{y}\}] = \pi_i \text{ if } (\hat{\mathbf{x}})_i = 0 \text{ etc.}$$

$$\mu_i^{\mathsf{BH}} = 2\pi_i(1 - \pi_i)$$

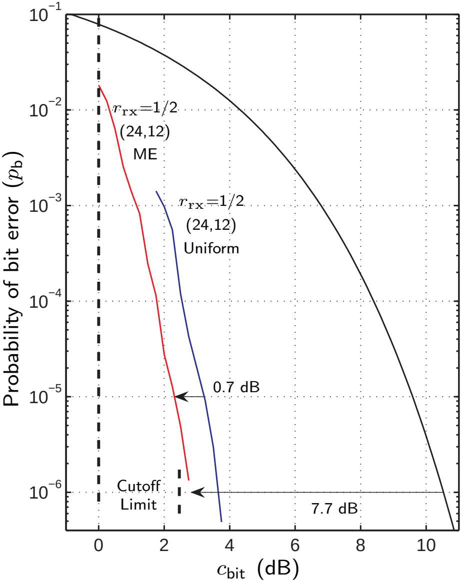
A Class of Max Entropy (ME) Designs

- Box-Hill picks bit with the maximum posterior entropy
- Proc B reduces to ME when $\widehat{\mathrm{x}_i} = (\hat{\mathbf{x}})_i \ \forall i$

Examples

Golay (24,12,8), Dual Hamming (4095,12,2048)

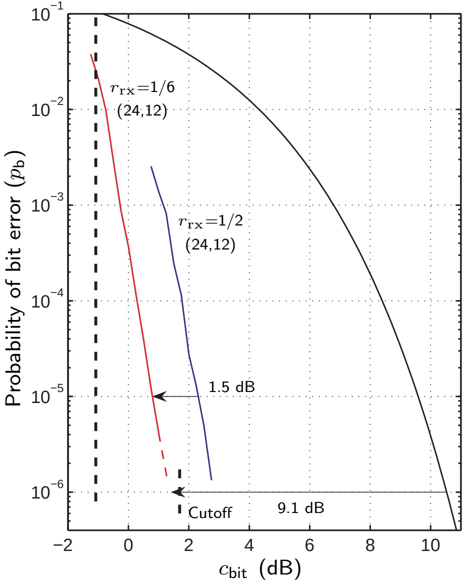
- Golay and Hamming performance identical
  - 0.3 dB away from cutoff
- $\mathcal{P}_{\text{ME}} \gg \mathcal{P}_{\text{s}}$
- Performance comparable to
  - $\text{RS}_8(160,128) + (2,1,7)$
  - LDPC/Turbo, $n \sim 500$
- Cf. Hagenaeur's RCPC
  - (3,1,7) mother code
  - n=256: At cutoff
- Cf. Rowitch & Milstein's RCPT
  - (256,231) BCH + (3,1,5) RSC
  - n=256: Better than cutoff (0.2 dB)



Performance at Lower Rates

Golay (24,12,8) at r_{rx}=1/6

- r_{rx} =1/6 gains 1.5 dB
- Surpasses cutoff limit by 0.3 dB
- Cf. RCPC
  - n=256: 2.3 dB worse
- Cf. RCPT
  - n=256: 1.8 dB worse



Decoding Algorithms

Calculating Π_j and π_i

Brute-Force

Operations per observation	MUL	ADD
Multiply $\mathbf{\Pi}$ with $e^{\pm y_{i^*}}$	M	-
Normalize $\mathbf{\Pi}$	-	$M-1$
If $\max \mathbf{\Pi} > a$ terminate	-	$M-1$
Calculate $\pi_i = \sum_{j \in J_i(1)} \Pi_j$	-	$n.M/2$
Pick $i^* = \arg\max_i \lvert\pi_i - 1/2\rvert$	-	$2M-1$

- Complexity summary
  - Storage $\Theta(M)$, Computation $\Theta(n \cdot M \cdot r_{\mathrm{rx}}^{-1})$
  - (7,3) incurs 16 MUL, 56 ADD per bit *on average*
  - (24,12) incurs 8K MUL, 48K ADD per bit!

BCJR Algorithm

- Storage and computation complexity is $\Theta(|E|)$
  - (24,12) has $|E|$=3580 and incurs about 14K MUL, 11K ADD per bit

Summary of Coding Under Observation Constraints

Conclusions

- Best design for $r_{\mathrm{rx}} \to 0$ is serial sampling
- Maximum entropy improves this significantly
- Surpass cutoff limit at rate $1/6$ with Golay $+$ ME
- But large computational costs near cutoff

Questions

- How does ME perform under non-zero Tx-rate constraints?
- How can computational complexity be reduced?