Two Proofs of Pinsker's Conjecture Concerning Arbitrarily Varying Channels
Rudolf Ahlswede and Ning Cai

Abstract—At the 1990 IEEE Information Theory workshop, M. S. Pinsker conjectured the following theorem: For an arbitrary varying channel (AVC) every rate below the random code capacity is achievable with deterministic list codes of constant list size, if the average error criterion is used. Two proofs of this theorem are given.

Index Terms—Arbitrarily varying channels, list codes.

I. PROOF OF THEOREM BY THE ELIMINATION TECHNIQUE [1]

An AVC is defined here by a sequence $A^\infty = \{P(x) : x \in \mathcal{X}\}$ of sets of transmission probabilities, where for a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a finite set of stochastic $[w_i(x) : x \in \mathcal{X}]$-matrices,

$$P(y^n|u^n, x^n) = \prod_{i=1}^{n} w_i(y_i|x_i,u_i),$$

(1)

for all $x^n = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n = \prod \mathcal{X}$, for all $y^n \in \mathcal{Y}^n$, and for all $s^n \in \mathcal{S}^n$.

From a random code $(n, N)$ of the AVC with average error less than $e^{-n}$ one can produce for any $i \in (0,1)$ by random selection of $L$ (used here instead of the letter $R$ in [1]) deterministic codes in the original ensemble and by randomization with the uniform distribution $\mu^s$ over those codes a new random code, say

$$\left((u_i, D_i) : 1 \leq i \leq N; l = 1, \ldots, L, \mu^s \right).$$

It was shown in [1], (inequality (4.3), that this random selection fails to lead for the new code and a fixed $s^n \in \mathcal{S}^n$ to an average error probability less than $\lambda$ with a probability smaller

Manuscript received September 19, 1990; revised April 5, 1991. This work was presented at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 24-29, 1991.

The authors are with the Fakultät für Mathematik, Universität Bielefeld, Postfach 8011, D-3350 Bielefeld 1, Germany.

IEEE Log Number 9102263.
than
\[ e^{-\alpha L}(1 + e^{-\epsilon\alpha L})^L, \quad \text{for any } \alpha > 0. \]

It suffices now to choose \( \alpha \) and \( L \) in such a way that
\[ |\mathcal{A}|e^{-\alpha L}(1 + e^{-\epsilon\alpha L})^L < 1. \quad (2) \]

The original choice in [1] was \( \alpha = 2 \) and \( L = n^2 \). We show now how to keep \( L \) constant.

Clearly, for \( \alpha = \epsilon n \) the left side in (2) is bounded from above by
\[ \exp(n \log |\mathcal{A}| - n\epsilon L + L) \]
and this is smaller than 1 for
\[ L > \left( \epsilon L - \frac{1}{\epsilon n} \right) \log |\mathcal{A}|, \quad (3) \]
and therefore, any constant
\[ L > (\epsilon L - \frac{1}{\epsilon n}) \log |\mathcal{A}| \quad (4) \]
is good for \( n \) sufficiently large.

Now our random code can be modified to the deterministic list code
\[ \{(u_i, D_i^i); (i, i) \in \{1, 2, \ldots, N\} \times \{1, \ldots, L\} \}. \]
It has list size \( L_a \), a length \( N - L \), even greater than the original code, and an average error probability less than \( \lambda \), because
\[ \frac{1}{L_a} \log |\mathcal{A}| \sum_{i=1}^{L_a} P(D_i^i | u_i | s^* \rangle) \geq 1 - \lambda, \quad \text{for } s^* \in \mathcal{A}. \]

**Remark 1:** We are curious to know whether the random code capacity can be obtained with a universal list length \( L_a \) that is a function of \( |\mathcal{A}| \) alone, that is, it does not depend on the error exponent \( \epsilon \) (which is a function of the rate), \( \lambda \), and on the class of matrices, in particular on \( |\mathcal{A}| \).

**Remark 2:** The present and also the following approach yield the bound in (4). Both approaches are based on modifications of certain classical \( (L = 1) \) codes. By a clever direct selection of the list codes one should do better. The ideas of [3] are promising.

**II. PROOF OF THEOREM BY A REFINEMENT OF THE ROBUSTIFICATION TECHNIQUE OF [2]***

As in [2, Section 5] we start with a code for an associated compound channel and build from these a random code for the AVC via permutations. Here this is done so that \( L \), defined as in (4), permutations suffice. Then we pass over to the list code as before. For the compound channel with class of matrices \( \mathcal{F} \)
\[ \mathcal{F} = \left\{ \sum_{x} w(x | s) Q(s); Q \in \mathcal{Q}(\mathcal{A}) \right\}, \]
the coding theorem says that any rate \( R \) below the random code capacity for the AVC, that is,
\[ C = \max_{P \in \mathcal{Q}(\mathcal{A})} \min_{s \in \mathcal{F}} I(P, \tilde{w}), \]
is achievable with deterministic codes of an error probability less than \( \exp(-\epsilon R) \), where \( \epsilon(R) > 0 \). Clearly, this code meets the same error bound for the compound channel with class of matrices \( \mathcal{F}_n \subset \mathcal{F} \), where
\[ \mathcal{F}_n = \left\{ \sum_{x} w(x | s) Q(s); Q \in \mathcal{Q}_n \right\}, \quad (5) \]
\[ \mathcal{R}_n = \left\{ Q \in \mathcal{Q}(\mathcal{A}); Q(s) \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\} \right\} \quad \text{for all } s \in \mathcal{A}. \quad (6) \]

With Chebyshev’s inequality we derive, from the inequality
\[ \sum_{s \in \mathcal{G}} \frac{1}{n} P(D_i^i | u_i | s^*) \sum_{i=1}^{N} Q(s_i) \leq e^{-\epsilon n}, \quad \text{for } Q \in \mathcal{Q}_n, \]
that for any \( \lambda > 0 \) and \( \gamma > 0 \) there are subsets \( \mathcal{G} \subset \mathcal{G}_n \) of the set of strict \( Q \)-typical sequences in \( \mathcal{A} \), with the properties
\[ \sum_{s \in \mathcal{G}} P(D_i^i | u_i | s^*) \leq \gamma, \quad \text{for } s \in \mathcal{A}. \quad (7) \]

Consider now the symmetric group (the set of all permutations) \( \Sigma_n \) acting on \( \{1, 2, \ldots, n\} \). We then define for \( s^* \in \mathcal{A} \), \( A \subset \mathcal{A} \) and \( \pi \in \Sigma_n \),
\[ \pi(s^*) = \pi(s_1, \ldots, s_n) = (s_{\pi(n_1), \ldots, s_{\pi(n)}}, \ldots), \quad (8) \]
\[ \pi(A) = \{ \pi(s^*); s^* \in A \}. \quad (9) \]

The desired result is an immediate consequence of the following fact.

**Lemma:** Suppose that for a family of sets \( \{B_p; P \in \mathcal{Q}_n\} \) with \( B_p \subset \mathcal{G}_n \) and for some \( \gamma > 0 \)
\[ |B_p| \geq |\mathcal{G}_n|(1 - e^{-\gamma n}), \]
then for every \( \delta > 0 \) and integer
\[ \frac{\log |\mathcal{A}|}{\delta n}, \quad (10) \]
there are permutations \( \{\pi_i\}_{i=1}^{L} \) with
\[ |\{\pi_i(s^*); i = 1, 2, \ldots, L\} \cap B_p| \geq L(1 - \delta), \]
for all \( P \in \mathcal{Q}_n \), and all \( s^* \in \mathcal{A}_n \), if \( n \) is larger than a suitable \( n(\delta, \gamma) \). (If we choose \( L > 2(\log |\mathcal{A}|/\delta n) \), then \( n(\delta, \gamma) = [2/\delta] \) does it.)

**Proof:** Let \( \{\pi_i\}_{i=1}^{L} \) be random permutations taking values in \( \Sigma_n \) according to the uniform distribution on \( \Sigma_n \). Then \( Pr(s \neq \pi s) = 1/n! \) for \( s \in \Sigma_n \) and all \( s^*, \pi(s^*) \in \mathcal{G}_n \). \( \pi_i(s^*) \subseteq B_p \) if
\[ \Pr(B(s^*, \pi s)) = \frac{\sum_{i=1}^{L} \left( \frac{L!}{|B_p|} \right) \left( \frac{|B_p|}{|\mathcal{G}_n|} \right)^{L-1} \leq 2 \left( \frac{|B_p|}{|\mathcal{G}_n|} \right)^{|B_p|} \leq e^{-\epsilon \log |\mathcal{G}_n|} \leq e^{-\gamma n(\delta, \gamma)}. \]

Therefore, the probability that for all \( s^* \in \mathcal{A}_n \), \( B(s^*, \pi s) \) occurs exceeds \( 1 - e^{-\epsilon \log |\mathcal{G}_n|} \). Clearly, this quantity is positive for \( L \) as specified in (11) and \( n \) large.

We now continue with the final steps in the second proof. Choose \( B_p = A_p \), \( \delta \in [0, 1] \) and \( \gamma \) such that \( e^{-\gamma n} \geq 1/(k\epsilon n) \). For large \( n \), \( \eta \) can be made arbitrarily close to \( \epsilon \), because \( \lambda \) is constant. By the Lemma, from (12) every \( s^* \in \mathcal{G}_n \) is contained in \( \mathcal{G}_n \).
Sequence Estimation and Synchronization from Nonsynchronized Samples

Costas N. Georghiades, Senior Member, IEEE and Marc Moeneclaey

Abstract—Recent advances in digital signal processing chips have motivated the study of algorithms for data detection and timing extraction that are easily digitally implemented. These algorithms operate on samples of the output of a matched-filter obtained asynchronously with the actual symbol timing in order to extract their timing and data estimates. A general analysis is presented of sampled receivers that handle arbitrary baseband pulse-shapes and arbitrary sampling rates. It is observed that the optimal processing of the matched-filter samples consists of digital interpolation, followed by symbol-by-symbol decoding when sampling is at (or above) the Nyquist rate, or Viterbi (sequence) decoding when sampling is below the Nyquist rate. Performance is studied through the Cramer–Rao bound on mean-square estimation error and a lower bound on error-probability.

Index Terms—Sequence estimation, synchronization, likelihood-function, sampling, Nyquist pulses.

I. INTRODUCTION

The advent of high-speed digital signal processing (DSP) chips in recent years has generated a need for receiver algorithms that operate on sampled as opposed to analog data to produce their timing and symbol estimates. A number of such algorithms that treat the problem of timing-recovery from sampled data has already appeared in the literature over the past few years [1]–[4]. These algorithms have a common theme in that they operate on samples of the matched filter output taken at some integer multiple of the symbol rate and at some arbitrary phase which may or may not be updated in time. This updating can be done by adjusting the phase of an analog sampler or by digitally interpolating between samples taken by a free-running clock.

In their classic paper, Mueller and Muller [1] study the problem of timing recovery from samples taken at the rate of one sample per symbol. The timing-recovery algorithms introduced are decision-directed and operate on baseband signals. The overall approach is to use the samples in order to estimate the timing-error, which is further used to correct the phase of the analog sampler in order to reduce the error. In a subsequent paper, Agazzi et al. [2] study timing-recovery in digital subscriber loops, and show that the wave difference method (WDM), originally introduced in [5], can be implemented with two samples per symbol. According to their proposed algorithm, even at phase lock, neither of the samples coincides with the actual decision timing and an interpolator is used to provide the decision data. The authors note that to avoid the need for an interpolator, sampling at the rate of four samples per symbol may be necessary.

Another timing-error detector of interest which operates on two samples per symbol was recently introduced by Gardner [3]. In contrast to [1], this algorithm is not decision-directed and does not assume carrier phase lock. Although similar to the algorithm introduced in [2], it has the advantage that one of the samples is taken at the data strobe and handles both baseband, as well as carrier signals. As with the other two algorithms in [1] and [2], Gardner's algorithm generates an error-signal that may subsequently be used to correct the sampling phase in the direction of reducing the timing-error, or to control a digital interpolator. The latter approach has the advantage of a fully digital implementation, and may well be the preferred way, especially at sufficiently high sampling rates when interpolation improves.

A further algorithm that does not update the sampling phase was introduced by Oerder and Meyr [4]. Their algorithm is effectively open-loop and operates on samples obtained under the command of a free-running oscillator. Although not explicitly studied, the authors assume that symbol detection is done by operating on data obtained through interpolation of the samples.

All of the previously described algorithms deal almost exclusively with timing-recovery, and assume that once timing is established, data detection can be accomplished using a slicer operating on the samples (or an interpolated value) obtained at the timing estimate. In contrast to [1]–[5], in recent work [6], [7], the problem of optimum joint timing and sequence estimation from samples obtained at the rate of one sample per symbol is investigated. The jointly optimal algorithm is based on maximizing a likelihood function for the sampled data, obtained by a free-running oscillator, over both timing and modulation sequences. The joint optimization is done naturally using the Viterbi algorithm in parallel for a number of discrete timing delays. An iterative algorithm for evaluating the likelihood function based on the expectation-maximization (EM) algorithms [8], [9] is also introduced in [10]. Results obtained indicate that such joint, open-loop, algorithms are quite robust to timing errors and can yield completely digital receiver realizations.

In another relevant paper, Falconer and Salz [11] treat the problem of joint data, delay and phase estimation for a Gauss-