Interleaver design for short block length Turbo codes

H. R. Sadjadpour, M. Salehi, N. J. A. Sloane, and G. Nebe

Abstract— **The performance of a Turbo code depends on the interleaver design. There are two major criteria that can be considered in the design of an interleaver. Distance spectrum properties of the code and the correlation of the extrinsic information with the input data are the two major criteria in designing an interleaver. This paper describes a new interleaver design based on these two criteria. Simulation results compare the new interleaver design to different existing interleavers. The distance spectrum properties of the Turbo code are compared for different interleaver choices. A new solution to the interleaver edge effects is proposed here.**

I. Introduction

TURBO codes[1] have an impressive near Shannon limit error correcting performance. This superior performance of Turbo codes compared to convolutional codes is only achievable when the length of the interleaver is very large in the order of several thousand bits. For large block size interleaver, most of random interleavers can perform well and it is not necessary to make any effort in the design of the interleaver. However, when the length of the interleaver is short, the performance of the Turbo codes degrades substantially up to a point that its bit error rate (BER) performance is worse than the conventional convolutional codes. In many applications such as voice, delay is an important issue in choosing a coding scheme. For these applications, large block size interleavers used in Turbo codes are not appropriate. Therefore, it is necessary to design an interleaver for Turbo code that demonstrates good BER performance. For these applications, appropriate choice of interleaver is important to take advantage of coding gains that Turbo code can offer. several authors have suggested interleaver design for Turbo codes suitable for short interleaver size [2-5].

There are two major criteria that can be considered in the design of an interleaver. These two criteria are 1)distance spectrum properties of the code, and 2)the correlation between the soft output of each decoder corresponding to its parity bits and information input data sequence. The second criterion is sometimes referred to as iterative decoding suitability (IDS)[2] criterion which is a measure of the effectiveness of iterative decoding algorithm and the fact that if the extrinsic information are less correlated to the input information data sequence, then the BER performance of iterative decoding algorithm of Turbo code improves.

The performance of the Turbo codes at low BER is mainly dominated by the minimum effective free distance (d_{min}). It is possible to design an interleaver that increases this d_{min}. The noise floor that occurs at moderate signal-to-noise ratios (SNR) is the result of small d_{min} [6]. The noise floor can be lowered by increasing either the interleaver size or the d_{min} which the later can be achieved by appropriate choice of interleaver. In our approach, increasing d_{min} is a condition in designing the interleaver.

Performance evaluation of the Turbo codes is usually based on the assumption that the receiver is a maximum likelihood (ML) decoder. However, Turbo codes utilize an iterative decoding algorithm such as BCJR decoding (MAP) algorithm [7]. This approach is sub-optimal compared to ML decoding algorithm. The iterative decoding algirthm performs better if the information that is sent to each decoder from the other decoder is less correlated to the input information data sequence. Hokfelt et. al. [3] has suggested an iterative decoding suitability (IDS) criterion that can be used in designing the interleaver. In the proposed interleaver design, we have recommended the use of IDS criterion with some modifications to the equations.

Trellis termination of Turbo codes is critical specially when the interleaver is designed to maximize d_{min}. If this problem is not addressed properly in the design of the code, it can cause a very small value for the d_{min} which subsequently will degrade the performance of the Turbo code. Some papers address this problem in the design of the interleaver [8-10]. A simple design criterion can be added to interleaver disign which will take care of this problem.

The paper is organized as follows. In section II the general assumptions are described and S-random interleaver[11] is described. Our approach is based on S-random interleaver.

II. Problem Statements

An interleaver π is a permutation that maps N input data sequence $d_1, d_2, \ldots, d_N$ into the same sequence of data with a new order, i.e., $i \rightarrow \pi(i)$. If the input data sequence is $\mathbf{d} = (d_1, d_2, \ldots, d_N)$, then $\mathbf{d}P$ is the permuted data sequence where P is the interleaving matrix with only one nonzero element equal to one in each column and row. Every interleaver has a corresponding de-interleaver (π^{-1}) that acts on the interleaved data sequence and puts them back into its original order. The de-interleaving matrix is simply the transpose matrix of interleaving matrix, i.e., P^T.

A random interleaver is a permutation of N integers that for each i, there is a corresponding random interger $\pi(i)$, without repetition. For large values of N, most random interleavers perform well. However, when the interleaver block size decreases, the performance of the Turbo codes degrades substatially up to a point that its BER performance is worse than the convolutional codes with similar computational complexity.

H. Sadjadpour and N. Sloane are with the AT&T Labs-research, Florham Park, NJ. E-mail: sadjadpour@att.com .

M. Salehi is with the department of electrical engineering, Northeastern University .

G. Nebe is with the department of mathematical science, ???? University .

An S-random interleaver is a semi-random interleaver that performs almost better than any known interleaver. Each randomly selected integer is compared to S previously selected random integers. If the distance between this integer and previously selected random integers is greater than S, then it is selected. Otherwise, a new random integer will be chosen and the above condition is tested. This process repeats untill all N integers are selected in this random order. Computer simulation results have shown that if $S \leq \sqrt{\frac{N}{2}}$, then this process will converge. This interleaver design assures that the short cycle events are avoided. Short cycle event occurs when two bits are close to each other before and after interleaving.

A new interleaver design was recently proposed based on the performance of iterative decoding algorithm in Turbo codes [2]. Turbo codes utilize an iterative decoding algorithm based on MAP algorithm or any algorithm that can provide soft output. At each decoding step, some information related to the parity bits of one decoder is fed into the next decoder together with the systematic data sequence and the parity bits corresponding to that decoder. Figure 1 demonstrates the iterative decoding scheme for Turbo codes. The inputs to each decoder are input data sequence, d_k, the parity bits y_k^1 or y_k^2, and the logarithm of likelihood ratio (LLR) associated to the parity bits from the other decoder (W_k^1 or W_k^2). All these inputs are utilized by MAP decoder to create three outputs corresponding to the weighted versions of these inputs. In figure 1, $\hat{d}_k$ at the output of the first decoder represents the weighted version of input data sequence, d_k. Also d_n in the figure demonstrates the fact that the input data sequence is fed into the second decoder after interleaving. The input to each decoder from the other decoder is used as a priori probability in the next decoding step. These information will be more effective in the performance of iterative decoding if it is less correlated to the input data sequence (or interleaved input data sequence). Therefore, it is reasonable to use this criterion for designing the interleaver. For large block size interleavers, most random interleavers provide a low correlation between W_k^i and input data sequence, d_k. The correlation coefficients, $r^1_{W^1_{k_1},d_{k_2}}$, is defined as the correlation coefficient between $W^1_{k_1}$ and d_{k_2}. It has been shown [2] that $r^1_{W^1_{k_1},d_{k_2}}$ can be analytically approximated as

$$\hat{r}^1_{W^1_{k_1},d_{k_2}} = \begin{cases} a\exp^{-c|k_1-k_2|} & \text{if } k_1 \neq k_2 \\ 0 & \text{if } k_1 = k_2 \end{cases} \quad (1)$$

a and c are two constants that depend on the encoder feedback and feedforward polynomials. The correlation coefficients at the output of the second decoder, $\hat{r}^2_{\mathbf{W}^2,\mathbf{d}}$, is approximated as

$$\hat{\mathbf{r}}^2_{\mathbf{W}^2,\mathbf{d}} = \frac{1}{2}\hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} P (I + \hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}}) \quad (2)$$

where the two terms in the right hand side of (2) correspond to the correlation coefficients between $\mathbf{W}^2$ and the two input data, i.e., $\mathbf{W}^1$ and $\mathbf{d}$ [2]. Similar correlation coefficients can be computed regarding the de-interleaver. The correlation matrix corresponding to de-interleaver, $\hat{\mathbf{r}'}^2_{\mathbf{W}^2,\mathbf{d}}$, is the same as (2) except P is substituted by P^T.

Then $\mathbf{V}_{k_1}$ is defined as

$$\mathbf{V}_{k_1} = \frac{1}{N-1}\sum_{k_2=1}^{N}(\hat{r}^2_{W^2_{k_1},d_{k_2}} - \bar{\hat{r}}^2_{W^2_{k_1},d_{k_2}}) \quad (3)$$

where

$$\bar{\hat{r}}^2_{W^2_{k_1},d_{k_2}} = \frac{1}{N}\sum_{k_1=1}^{N}\hat{r}^2_{W^2_{k_1},d_{k_2}} \quad (4)$$

V'_{k_1} is defined in a similar way using $\hat{\mathbf{r}'}^2_{\mathbf{W}^2,\mathbf{d}}$. The iterative decoding suitability (IDS) measure is then defined as

$$IDS = \frac{1}{2N}\sum_{k_1=1}^{N}(V_{k_1} + V'_{k_1}) \quad (5)$$

A low value of IDS is an indication that correlation properties between $\mathbf{W}^1$ and $\mathbf{d}$ are equally spread along the data sequence of length N. An interleaver design based on the IDS condition is suggested in [12].

III. 2-step S-random Interleaver Design

A new interleaver design, 2-step S-random interleaver, is presented here based on the S-random interleaver that described earlier. The 2-step S-random interleaver is designed under the constraint to increase the minimum effective free distance of the Turbo code while decreasing the correlation properties between the information input data sequence, d_k, and W_k^i. Hokfelt et. al [2,12] presented IDS criterion to evaluate this correlation properties. The two vectors that are used for the computation of IDS, namely V_{k_1} and V'_{k_1}, are very similar criterion and for almost all interleavers, it is sufficient to only use V_{k_1}. However, we can define a new criterion as decreasing the correlation after the third decoding step, the correlation between feedback extrinsic information and from second decoder and input information data sequence. In this regard, the new $V^{'(new)}_{k_1}$ can be defined as

$$\begin{aligned} \hat{\mathbf{r}'}^2_{\mathbf{W}^2,\mathbf{d}} &= \frac{1}{2}\hat{\mathbf{r}}^2_{\mathbf{W}^2,\mathbf{d}} P^T (I + \hat{\mathbf{r}}^2_{\mathbf{W}^2,\mathbf{d}}) \\ &= \frac{1}{4}(\hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} + \hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} P \hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} P^T) \\ &\times \ (I + \frac{1}{2}\hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} P + \frac{1}{2}\hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}} P \hat{\mathbf{r}}^1_{\mathbf{W}^1,\mathbf{d}}) \end{aligned} \quad (6)$$

$V^{'(new)}_{k_1}$ can be defined similar to (3) based on (6). The new iterative decoding suitability (IDS_1) is then defined as

$$IDS_1 = \frac{1}{2N}\sum_{k_1=1}^{N}(V_{k_1} + V^{'(new)}_{k_1}) \quad (7)$$

A small value for IDS_1 only gaurantees that the correlation properties are spread equally throughout the information input data sequence. However, this criterion does not attempt to reduce the power of correlation coefficients, i.e.,

$(\hat{\mathbf{r}}^2_{\mathbf{W}^2,\mathbf{d}})^2$ and $(\hat{\mathbf{r}'}^2_{\mathbf{W}^2,\mathbf{d}})^2$. Therefore, we recommend to add this criterion to the iterative decoding suitability.

$$IDS_2 = \frac{1}{2N^2}\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}((\hat{\mathbf{r}}^2_{\mathbf{W}^2,\mathbf{d}})^2 + (\hat{\mathbf{r}'}^2_{\mathbf{W}^2,\mathbf{d}})^2) \quad (8)$$

A simple approach to define the new IDS ($IDS_{(new)}$) criterion based on IDS_1 and IDS_2 can be defined as the average of these two values. It is not clear that this approach is necessarily optimum.

$$IDS_{(new)} = \frac{1}{2}(IDS_1 + IDS_2) \quad (9)$$

We have decided to use (9) as one of the conditions for optimization of the interleaver.

As we described earlier, S-random interleavers avoid short cycle events. This criterion, guarantees that two bits close to each other before interleaving, will have a minimum distance of S after interleaving. More precisely, for information data sequence i and j, $\pi(i)$ and $\pi(j)$ represent their interleaved location in the permuted data sequence. S-random interleaver will guarantee that if $|i-j| \leq S$, then $|\pi(i) - \pi(j)| > S$. In the extreme case, if $\pi(j) = j$ and the above condition satisfies, $j \rightarrow \pi(j)$ is a valid assignment for the S-random interleaver. This situation can degrade the performance of the iterative decoding of the Turbo codes. The larger the distance between j and $\pi(j)$, the smaller the correlation between the extrinsic information from the second decoder and input information data sequence. Based on the above argument, we have decided to introduce an additional constraint, S_2, which is defined as the minimum permissible distance between j and $\pi(j)$ for all $j = 1, 2, \ldots, N$.

Unlike [12] that the interleaver design is based on IDS criterion, we design our interleaver in two stages. In the first stage, we design an interleaver that satisfies S-random criterion together with the S_2 condition. In the second stage, we try to increase the minimum effective free distance [13] of the Turbo code by utilizing the $IDS_{(new)}$ condition. The design is as follows:

Step 1: Design an interleaver of length N that for all integer values of i choose its corresponding $\pi(i)$ without repetition and having the following two properties:

- For all i and j if $|i-j| \leq S_1 \Longrightarrow |\pi(i)\pi(j)| > S_1$.
- For all i and $\pi(i)$, then $|i - \pi(i)| > S_2$.

Step 2: Choose a pre-determined minimum effective free distance code [13] d_{min1}. Find all input data sequences of length N and weight $\leq w_{det}$ such that $d_{min} < d_{min1}$. Suppose one of these input data sequences of length N and weight w_1 has the following non-zero interleaver pairs $(i_1, \pi(i_1)), (i_2, \pi(i_2)), \ldots, (i_{w_1}, \pi(i_{w_1}))$ with $d_{min,w_1} < d_{min1}$. Compute $IDS_{(new)}$ based on (9). Set $j = i_1 + 1$ and find the pair $(j, \pi(j))$. Interchange the interleaver pairs $(i_1, \pi(i_1))$ and $(j, \pi(j))$ to create a new interleaver, i.e., $(i_1, \pi(j))$ and $(j, \pi(i_1))$. Compute the new IDS, $IDS'_{(new)}$, based on the new interleaver design. if $IDS'_{(new)} \leq IDS_{(new)}$, the new interleaver design will replace the previous one. Otherwise, set $j = j + 1$ and continue this search until a new interleaver with smaller $IDS'_{(new)}$ is found. Repeat this operation for all input data sequences that have $d_{min} < d_{min1}$. After the search ends, go back to step 2 again and find all input data sequences of weight w_{det} or less with $d_{min} < d_{min1}$. Continue this step untill it converges and there is not any input data sequence of weight w_{det} or less with $d_{min} < d_{min1}$. Obviously if d_{min1} is chosen a large value, the second step may never converge. In this case, a smaller value for d_{min1} should be selected.

It is well known [13-14] that the feedback polynomial used for each recursive systematic convolutional code as constituent encoder should be selected a primitive polynomial. Primitive polynomials used for Turbo codes exhibits better spectrum distance properties. Next section describes a theorem to find all weight w_{det} input data sequences that are divisible by a primitive polynomial.

IV. Polynomials Divisible by a Primitive Polynomial

Let $R = GF(2)[X]$ be the ring of polynomials with binary coefficients, and let $p(X) \in R$ be a primitive irreducible polynomial of degree $m > 1$. We wish to determine all the polynomials $f(X) \in R$ which have low weight and are divisible by $p(X)$. (The weight of a polynomial is the number of nonzero terms.)

Choose a zero α of $p(X)$. Then α generates $GF(2^m)$ as a field. Since $p(X)$ is primitive, by definition the minimal $n > 0$ with $\alpha^n = 1$ is $n = 2^m - 1$. Note that the nonzero elements of $GF(2^m)$ are precisely the n zeros of the polynomial $X^n - 1$.

Since $p(X)$ is irreducible, a polynomial $f(X) \in R$ is divisible by $p(X)$ if and only if $f(\alpha) = 0$. If $i, j \in \mathbf{N}$ satisfy $i \equiv j \pmod{n}$, then $\alpha^j = \alpha^i$, hence $X^i + X^j$ is divisible by $p(X)$. Let T_2 be the set of polynomials $X^i + X^j \in R$ with $0 \leq i < j$, $i \equiv j \pmod{n}$. More generally, let T_{2k} $(k = 2, 3, \ldots)$ be the sum of k disjoint (i.e. all monomials are distinct) terms from T_2.

Let H be the Hamming single-error-correcting code with generator polynomial $p(X)$, and let A_w be the set of codewords of H of weight w, written in the usual way as polynomials of degree $< n$ corresponding to residue classes in $R/(X^n - 1)$. Note that A_i is empty unless $i \equiv 3$ or $0 \pmod 4$, and in particular A_1, A_2, A_5, A_6, ... are empty.

Theorem

A polynomial $f(X) \in R$ of weight w is divisible by $p(X)$ if and only if $f(X)$ can be written as

$$f(X) = g(X) + h(X)\ ,$$

where $g(X) \in T_{2i}$, $h(X) \in R$ has weight j, the terms in $g(X)$ and $h(X)$ are disjoint, $\pi(h(X)) \in A_j$, $2i + j = w$, and π means "read exponents mod n".

Proof:

"$\Leftarrow$" Let $f(X) = g(X) + h(X)$ be as in the theorem. Since $\pi(h(X))$ is divisible by $p(X)$, one has $\pi(h)(\alpha) = h(\alpha) = 0$. Therefore $g(X) \in T_{2i}$ and $h(X)$ are both

divisible by $p(X)$ and so is $f(X)$. By construction the weight of $f(X)$ is $w = 2i + j$.

"$\Rightarrow$" Let $f(X) \in R$ be divisible by $p(X)$. Then $f(X)$ can be written as $f(X) = g(X) + h(X)$, where no pair of terms in $h(X)$ has exponents that are congruent modulo n and $g(X) \in T_{2i}$ for some i, where $g(X)$ consists of pairs of terms of $f(X)$ in which the exponents are congruent modulo n. By construction $g(X)$ and hence $h(X)$ is divisible by $p(X)$, whence $\pi(h(X)) \in A_j$ for some j. Again by construction the weight of $h(X)$ is the weight of $\pi(h(X))$ and the weigth of $f(X)$ is $2i + j = w$.

Remark Note that the polynomials $g(X)$ and $h(X)$ are not necessarily unique. But one may define $g(X)$ by starting from the highest exponent of $f(X)$ and always taking the first term that fits to make the decomposition unique.

We discuss the first few values of w individually, and illustrate by taking $m = 3$, $n = 7$ and $p(X) = X^3 + X + 1$. Then H is a Hamming code of length 7, containing 7 words of weight 3, 7 of weight 4, and 1 word of weight 7.

Weight $w = 1$ No monomials are divisible by $p(X)$.

Weight $w = 2$ A weight two polynomial is divisible by $p(X)$ if and only if it is in T_2.

Examples: $1 + X^7$, $X^4 + X^{39}$.

General form: $f(X) = X^i + X^{i+7j}$, $i \geq 0, j \geq 1$.

Weight $w = 3$ A weight three polynomial is divisible by $p(X)$ if and only if it reduces to a weight 3 codeword in H when the exponents are read mod n.

Example: The 7 words in A_3 are the cyclic shifts of $p(X)$ itself. So for instance $X^{32} + X^{16} + X^8$ is divisible by $p(X)$, since it reduces to $X^4 + X^2 + X = Xp(X) \in A_3$.

General form: $f(X) = X^{i+7j} + X^{i+1+7k} + X^{i+3+7l}$, $i, j, k, l \in \mathbf{Z}$, $i + 7j, i + 1 + 7k, i + 3 + 7l \geq 0$.

Weight $w = 4$ A polynomial of weight 4 is divisible by $p(X)$ if and only if it is in T_4, or it reduces to an element of A_4 when the exponents are read mod n.

Examples: $1+X^7+X^{10}+X^{17} \in T_4$, $1+X^2+X^3+X^4 \in A_4$.

Weight $w = 5$ A polynomial of weight 5 is divisible by $p(X)$ if and only if it is a disjoint sum of an element of T_2 and something which reduces to an element of A_3.

Example: $X^{15} + X^8 + X^3 + X + 1$.

Weight $w = 6$ A polynomial of weight 6 is divisible by $p(X)$ if and only if either it is in T_6, or if it is the disjoint sum of an element of T_2 and something which reduces to an element of A_4.

Examples: $X^9+X^8+X^7+X^2+X+1 \in T_6$, $X^7+X^5+X^4+X^3+X+1 \in T_2 \oplus A_4$.

V. Simulation Results

The

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, *"Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes,"* Proceeding of IEEE ICC 93, pp. 1064-1070.

[2] J. Hokfelt, O. Edfors, and T. Maseng, *"Turbo Codes: Correlated Extrinsic Information and its Impact on Iterative Decoding Performance,"* Proceeding of IEEE VTC '99, Houston, Texas.

[3] A.K. Khandani, *"Group Structure of Turbo Codes with Applications to the Interleaver Design,"* Intrenational Symposium on Information Theory, pp. 421, August 1998.

[4] O.Y. Takeshita and D.J. Costello, Jr., *"New Classes of Algebraic Interleavers for Turbo Codes,"* Intrenational Symposium on Information Theory, pp. 419, August 1998.

[5] H. Herzberg, *"Multilevel Turob Coding with Short Interleavers,"* IEEE Journal on selected areas in Communications, vol.16, no. 2, pp. 303-309, February 1998.

[6] L.C. Perez, J. Seghers, and D.J. Costello, *"A Distance Spectrum Interpretation of Turbo Codes,"* IEEE Trans. on Information Theory, vol. 42, No. 6, pp.1698-1709, November 1996.

[7] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, *"Optimum decoding of linear codes for minimizing symbol error rate,"* IEEE Trans. on Inf. Theory, vol. IT-20, pp. 284-287, Mar. 1974.

[8] W. Blackert, E. Hall, and S. Wilson, *"Turbo code termination and interleaver conditions,"* Electronics Letters, vol. 31, pp. 2082-2084, November 1995.

[9] A.S. Barbulescu and S.S. Peitrobon, *"Terminating the trellis of Turbo codes in the same state,"* Electronic Letters, vol. 31, pp. 22-23, January 1995.

[10] M.C. Reed and S.S. Peitrobon, *"Turbo code termination schemes and a novel alternative for short frames,"* Seventh IEEE Internation Symposium on Personal, Indoor, and Mobile Communications, 1996.

[11] S Dolinar and D. Divsalar, *"Weight Distribution for Turbo codes Using Random and Nonrandom Permutations,"* JPL Progress report 42-122, pp. 56-65, August 15, 1995.

[12] J. Hokfelt, O. Edfors, and T. Maseng, *"Interleaver Design for Turbo Codes Based on the Performance of Iterative Decoding,"* Proceeding of IEEE ICC '99, Vancouver, Canada.

[13] S. Benedetto and G. Montorsi, *"Design of Parallel Concatenated Convlolutional Codes, "* IEEE Trans. on Comm., vol. 44, no. 5, pp. 591-600, May 1996.

[14] L.C. Perez, J. Seghers, and D.J.Costello, *"A Distance Spectrum Interpretation of Turbo Codes,"* IEEE Trans. on Information Theory, vol. 42, no. 6, pp. 1698-1709, November 1996.