Accelerating Iterative Detection for Spatially Coupled Systems by Collaborative Training

Keigo Takeuchi, *Member, IEEE*

Abstract—This letter proposes a novel method for accelerating iterative detection for spatially coupled (SC) systems. An SC system is constructed by one-dimensional coupling of many subsystems, which are classified into training and propagation parts. An irregular structure is introduced into the subsystems in the training part so that information in that part can be detected successfully. The obtained reliable information may spread over the whole system via the subsystems in the propagation part. In order to allow the subsystems in the training part to collaborate, *shortcuts* between them are created to accelerate iterative detection for that part. As an example of SC systems, SC code-division multiple-access (CDMA) systems are considered. Density Evolution for the SC CDMA systems shows that the proposed method can provide a significant reduction in the number of iterations for highly loaded systems.

Index Terms—spatial coupling, code-division multiple-access, small world, belief propagation, density evolution.

I. INTRODUCTION

SPATIAL coupling has been proved to improve the belief-propagation (BP) performance of conventional low-density parity-check (LDPC) codes up to the corresponding maximum-a-posteriori (MAP) performance [1]. This phenomenon has been observed in many other problems, such as code-division multiple-access (CDMA) systems [2]–[4], compressed sensing [5]–[7], and models in statistical physics [8]. See [3], [9] for a theoretical treatment of general systems.

A spatially coupled (SC) system is constructed as a one-dimensional chain of L *large* subsystems. A slightly irregular structure, which results in a rate loss, is introduced at both ends of the chain so that information at both ends can be detected successfully. BP detection consists of two stages: training and propagation stages. In the training stage, information at both ends is first detected by utilizing the irregularity. We refer to positions at which irregularity is imposed as training positions. In the propagation stage, on the other hand, the reliable information at both ends propagates toward the center of the chain. *Order* preserved in each large subsystem enables no error propagation. Since the rate loss due to the irregularity at both ends vanishes as $L \to \infty$, the best performance is achieved[1] in that limit. This implies that a long chain should

[1]Reliable information fails to propagate with a finite probability when the size of each subsystem is finite. Consequently, an infinitely long chain may not be best for finite-sized systems.

be used to approach the MAP performance.

When a long chain is used, many iterations are required for propagating reliable information over the whole chain. The purpose of this letter is to propose a novel method of coupling for accelerating the convergence of iterations. A remarkable effort oriented in the same direction was made by Truhachev et al.: They proposed to divide a long chain into several short chains and to re-connect the divided chains in a ladder-like [10] or loop-like [11] structure. The connecting positions, along with the end positions, correspond to the training positions. Reliable information at the connecting positions spreads over the chains simultaneously. As a result, the propagation of reliable information over the whole system is accelerated compared to one long chain.

In this letter, we shall accelerate iterative detection in the training stage by creating *shortcuts* between distant training positions. Reliable information at each training position propagates to the other training positions through the shortcuts, whereas there is no collaboration between distant training positions for conventional SC systems. This collaborative training accelerates iterative detection in the training stage.

This letter is organized as follows: After summarizing the notation and terminology used in this letter, in Section II we focus on SC CDMA systems as an example of SC systems, and explain how to connect training positions. Section III presents the density-evolution (DE) analysis of BP detection. In Section IV, comparisons between collaborative training and non-collaborative training are made in terms of the number of iterations. Section V concludes this letter.

For integers l and L, $(l)_L$ is equal to $l + iL$ for an integer i such that $0 \leq l + iL \leq L - 1$. The set of consecutive integers $\{i, i+1, \ldots, j\}$ $(i < j)$ is written as $[i : j]$. The vector $\mathbf{1}_n$ denotes the n-dimensional vector whose elements are all one. The Q-function $Q(\cdot)$ is the upper tail probability of the standard real Gaussian distribution. The real Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$ is written as $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$. In a graph, the degree of a node is defined as the number of edges connected to the node. The distance between two nodes is the number of edges in the shortest path that connects the two nodes.

II. SYSTEM MODEL

A. Spatially Coupled CDMA Systems

We consider synchronous K-user SC CDMA systems over the real additive white Gaussian noise (AWGN) channel as an example of SC systems. Note that it is possible to apply the

idea in this letter to the other SC systems. For simplicity in presentation, *dense* spreading sequences are used, whereas the *dense* limit of sparse spreading sequences should be taken for rigorous DE [12], [13]. See [3] for the details.

Transmission over L symbol periods is considered. Let N_l and $\bar{N} = L^{-1} \sum_{l=0}^{L-1} N_l$ denote the spreading factor in symbol period l and the average spreading factor, respectively. The received vector $\boldsymbol{y} = (\boldsymbol{y}_0^{\mathrm{T}}, \ldots, \boldsymbol{y}_{L-1}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{L\bar{N}}$ is given by

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}). \tag{1}$$

In (1), $\boldsymbol{x} = (\boldsymbol{x}_0^{\mathrm{T}}, \ldots, \boldsymbol{x}_{L-1}^{\mathrm{T}})^{\mathrm{T}} \in \{1, -1\}^{LK}$ denotes the data symbol vector with binary phase shift keying (BPSK), in which $\boldsymbol{x}_m \in \{1, -1\}^K$ consists of the mth data symbols of all users. Furthermore, $\boldsymbol{S}$ represents the $L\bar{N} \times LK$ spreading matrix, constructed as

$$\boldsymbol{S} = \begin{bmatrix} b_{0,0}\boldsymbol{S}_{0,0} & \cdots & b_{0,L-1}\boldsymbol{S}_{0,L-1} \\ \vdots & & \vdots \\ b_{L-1,0}\boldsymbol{S}_{L-1,0} & \cdots & b_{L-1,L-1}\boldsymbol{S}_{L-1,L-1} \end{bmatrix}. \tag{2}$$

In (2), $\{\boldsymbol{S}_{l,m}\}$ are independent $N_l \times K$ matrices that have independent entries taking $\pm 1/\sqrt{N_l}$ with probability $1/2$. Furthermore, $b_{l,m} \in \mathbb{R}$ is the (l, m)-element of the base matrix $\boldsymbol{B} \in \mathbb{R}^{L \times L}$, which characterizes spatial coupling. Imposing $\sum_{l=0}^{L-1} b_{l,m}^2 = 1$ for all m normalizes the average power used for transmitting each data symbol.

The set $\mathcal{L} = [0 : L-1]$ of all symbol periods is decomposed into the training phase $\mathcal{T} \subset \mathcal{L}$ and the propagation phase $\mathcal{P} \subset \mathcal{L}$, which are disjoint subsets of $\mathcal{L}$. Let N_{tr} and N denote the spreading factors in the training and propagation phases, respectively. Large N_{tr} is assumed to obtain reliable estimates of the data symbols transmitted in the training phase. The average load $\bar{\alpha}$ is defined as

$$\bar{\alpha} = \frac{LK}{|\mathcal{T}|N_{\mathrm{tr}} + |\mathcal{P}|N} = \left\{ \alpha_{\mathrm{tr}}^{-1} \frac{\tau}{L} + \alpha^{-1} \left(1 - \frac{\tau}{L}\right) \right\}^{-1}, \tag{3}$$

with $\tau = |\mathcal{T}|$. In (3), $\alpha_{\mathrm{tr}} = K/N_{\mathrm{tr}}$ and $\alpha = K/N$ denote the loads in the training and propagation phases, respectively. Under the BPSK assumption the average load (3) is equal to the average sum rate, and tends toward α as $L \to \infty$ and $\tau/L \to 0$.

B. Spatial Coupling & Collaborative Training

In one-dimensional or circular coupling with coupling width W, a circulant matrix is used as the base matrix $\boldsymbol{B}$:

$$b_{l,m} = \frac{1}{\sqrt{2W+1}} b_{(l-m)_L}, \tag{4}$$

with $(b_0, \ldots, b_{L-1}) = (\boldsymbol{1}_{W+1}^{\mathrm{T}}, \boldsymbol{0}, \boldsymbol{1}_W^{\mathrm{T}})$. This regular base matrix is written as $\boldsymbol{B}_{L,W}^{(\mathrm{reg})}$. In order to present collaborative training, we shall introduce a graph representation of the base matrix $\boldsymbol{B}$.

The $L \times L$ base matrix $\boldsymbol{B}$ can be represented by a bipartite graph that consists of L factor nodes and L variable nodes. The factor nodes shown by squares correspond to the row indices of $\boldsymbol{B}$, whereas the variable nodes represented by circles are associated with the column indices. If the element $b_{l,m}$ is non-zero, there is an edge between factor node l and variable

Fig. 1. Regular bipartite graph with $L = 32$ and $W = 2$.

node m. Otherwise, there is no edge between the two nodes. In this letter, we refer to the bipartite graph that represents the regular base matrix $\boldsymbol{B}_{L,W}^{(\mathrm{reg})}$ as the regular bipartite graph $\boldsymbol{B}_{L,W}^{(\mathrm{reg})}$. See Fig. 1 for the regular bipartite graph $\boldsymbol{B}_{32,2}^{(\mathrm{reg})}$.

As an ensemble of collaborative training, we use an ensemble based on a small-world (SW) network [14], which is a model for elucidating the so-called SW phenomenon such as "six degrees of separation." SW networks are highly clustered, as regular networks are, *and* have short path lengths between any two nodes, as random networks do. These properties of SW networks are suitable for accelerating iterative detection in the training stage. In this letter, this SW-network-based ensemble is referred to as "SW ensemble."

Let $\mathcal{C}_W^{(\mathrm{reg})}(m) \subset \mathcal{L}$ denote the set of node m and nodes with distance 2 from node m in a regular bipartite graph with coupling width W. For example, $\mathcal{C}_2^{(\mathrm{reg})}(0)$ for variable nodes is shown by the gray nodes in Fig. 1. The SW ensemble is obtained by modifying a regular bipartite graph at positions included in equally spaced c clusters $\{\mathcal{C}_W^{(\mathrm{reg})}(iL/c)\}_{i=0}^{c-1}$. One could add nodes with more than two distances from node m into $\mathcal{C}_W^{(\mathrm{reg})}(m)$. However, such a generalization would increase the number of instances included in the SW ensemble. In this letter, only nodes with distance 2 from node m are included into $\mathcal{C}_W^{(\mathrm{reg})}(m)$.

1) Let $i = 0$ and generate the regular bipartite graph $\boldsymbol{B}_{L,W}^{(\mathrm{reg})}$.
2) Repeat the following for all variable nodes $m \in \mathcal{C}_W^{(\mathrm{reg})}(iL/c)$ in the ith cluster: With probability p, re-connect each edge that is connected to variable node m to a factor node $l \in \cup_{j \neq i} \mathcal{C}_W^{(\mathrm{reg})}(jL/c)$ in the other clusters uniformly and randomly.
3) If $i = c-1$, terminate the algorithm. Otherwise, go back to Step 2) after $i := i + 1$.

The probability p controls a tradeoff between the *clustering* property and the short-path-length property: The SW coupling with $p = 0$ reduces to the highly-clustered regular coupling, whereas it with $p = 1$ does to the random coupling with short path lengths. Thus, moderate p should be used to obtain instances with the two properties.

In bipartite graphs obtained by the algorithm above, different factor nodes may have different degrees, whereas all variable nodes have degree $2W + 1$. We propose a heuristic

Fig. 2. A bipartite graph obtained from the $(64, 2, 0.1, 2, 14)$-SW ensemble. The black factor nodes show the nodes corresponding to the training phases. The variable nodes $m \notin C_2^{(\mathrm{reg})}(0) \cup C_2^{(\mathrm{reg})}(32)$ and the associated factor nodes are omitted, since they can be restored uniquely.

algorithm for determining the training phase $\mathcal{T}$ with size $\tau = |\mathcal{T}|$. The proposed algorithm assigns factor nodes with largest degrees to the training phase. Factor nodes with large degrees serve as hubs from which reliable information spreads over variable nodes.

1) Let $\tilde{\tau} = \tau$, $\mathcal{T} = \emptyset$, and $d = d_{\max}$, with $d_{\max}$ denoting the maximum degree of the factor nodes.
2) If the number of factor nodes with degree d is greater than or equal to $\tilde{\tau}$, pick up $\tilde{\tau}$ factor nodes from all factor nodes with degree d uniformly and randomly, add the corresponding indices into $\mathcal{T}$, and terminate the algorithm. Otherwise, go to the next step.
3) Add the indices $\mathcal{L}_d$ that represent all factor nodes with degree d into $\mathcal{T}$, and go back to Step 2) after $\tilde{\tau} := \tilde{\tau} - |\mathcal{L}_d|$ and $d := d - 1$.

We refer to the ensemble[2] that consists of all instances obtained by the two algorithms above as the (L, W, p, c, τ)-SW ensemble. See Fig. 2 for an instance obtained from the SW ensemble.

Remark 1. *The degrees of the factor nodes provide an impact on the complexity of iterative detection algorithms. One method for making a fair comparison between regular and irregular bipartite graphs would be to consider sparsely spread CDMA systems in which, for factor nodes with a large degree, small row weights are assigned to the corresponding subsystems. However, this influence on the performance vanishes in the dense limit. This argument implies that comparisons presented in this letter are not necessarily unfair in terms of the complexity.*

III. DENSITY EVOLUTION

We consider an iterative detection algorithm based on BP with the Gaussian approximation [15]. See [3] for the details. In order to evaluate the bit error rate (BER) via DE, we take the large-system limit in which K and $\{N_l\}$ tend to infinity with

[2] The ensemble with $L = 64$, $W = 2$, and $c = 2$ contains all bipartite graphs obtained by re-connection of the edges in Fig. 2. However, their occurrence probabilities are non-uniform and determined on the basis of the probability p, whereas instances contained in conventional ensembles are assumed to occur uniformly and randomly.

the ratios $\alpha_l = K/N_l$ fixed for all l. The following theorem is useful for quickly evaluating the performance of instances picked up from the SW ensemble.

Theorem 1. *In the large-system limit, the BER of each element of the mth data symbol vector $\boldsymbol{x}_m$ in iteration i is given by $Q(\sqrt{\mathrm{sir}_m(i)})$, in which $\{\mathrm{sir}_m(i)\}$ are determined via the coupled DE equations,*

$$\mathrm{sir}_m(i) = \sum_{l=0}^{L-1} \frac{b_{l,m}^2}{\sigma_l^2(i)}, \tag{5}$$

$$\sigma_l^2(i) = \sigma^2 + \alpha_l \sum_{m=0}^{L-1} b_{l,m}^2 \mathrm{MMSE}(\mathrm{sir}_m(i-1)), \tag{6}$$

with $\mathrm{sir}_m(0) = 0$ for all m. In (6), the function $\mathrm{MMSE}(x)$ denotes the minimum mean-squared error (MMSE) of the BPSK-input AWGN channel with signal-to-noise ratio (SNR) x.

Proof: The proof is a generalization of the results in [3] and is therefore omitted. ∎

When a bipartite graph of coupling and the load α_{tr} in the training phase are given, the DE equations (5) and (6) have a unique fixed-point for small SNR $1/\sigma^2$ or small load α in the propagation phase. On the other hand, there may be multiple fixed-points for large SNR and large load. The BP threshold for fixed SNR is defined as the maximum load such that the DE equations have a unique fixed-point.

Definition 1. *When α_{tr}, a bipartite graph of coupling, and SNR are fixed, the BP threshold is defined as the supremum of α_{c} such that the DE equations (5) and (6) have a unique fixed-point for all $\alpha < \alpha_{\mathrm{c}}$.*

See [3] for how to estimate the BP threshold. The operational meaning of the BP threshold is that, when α is smaller than the BP threshold, the BP-based iterative algorithm can eliminate multiple-access interference (MAI), and achieve performance close to the single-user bound. On the other hand, the system is MAI-limited when α is larger than the BP threshold. When $1/\sigma^2 = 10$ dB, the BP threshold α_{BP} for the uncoupled CDMA system is given by $\alpha_{\mathrm{BP}} \approx 1.73078$ [3]. As L and W tend to infinity with $W/L \to 0$, on the other hand, the corresponding BP threshold $\alpha_{\mathrm{BP}}^{(\mathrm{reg})}$ for the circularly coupled CDMA system tends toward the optimal threshold $\alpha_{\mathrm{MAP}} \approx 1.98267$ [3].

IV. NUMERICAL COMPARISONS

We shall make comparisons between regular coupling and the SW ensemble. Bipartite graphs obtained from the SW ensemble have distinctly different performance instance by instance for small-sized graphs. Thus, we used good instances that are obtained by examining many instances picked up from the SW ensemble with Theorem 1.

Figure 3 shows the evolution of the BERs for the bipartite graph of coupling shown in Fig. 2. We find that the data symbols in the two clusters $C_2^{(\mathrm{reg})}(0)$ and $C_2^{(\mathrm{reg})}(32)$ are detected first. Their BERs are not constant, because of the irregularity of the bipartite graph in the clusters. Then, the reliable information propagates toward the middle positions

Fig. 3. BER versus m/L for the bipartite graph of coupling obtained from the $(64, 2, 0.1, 2, 14)$-SW ensemble, shown in Fig. 2. $1/\sigma^2 = 10$ dB, $\alpha_{\mathrm{tr}} = 1.45$, and $\alpha = 1.98$.

Fig. 4. Average BER versus the number of iterations for $W = 2$, $1/\sigma^2 = 10$ dB, and $\alpha_{\mathrm{tr}} = 1.45$. For the regular coupling with $L = 64$ (resp. $L = 128$), 14 (resp. 28) symbol periods $\mathcal{T} = \{[61 : 63], [0 : 3], [29 : 35]\}$ (resp. $\mathcal{T} = \{[125 : 127], [0 : 3], [29 : 35], [61 : 67], [93 : 99]\}$) are assigned to the training phase. The bipartite graph of coupling shown in Fig. 2 was used for the $(64, 2, 0.1, 2, 14)$-SW ensemble.

between the two clusters. Eventually, the BERs at all positions tend to a small level of 10^{-3}.

Figure 4 presents comparisons between the regular coupling and the SW ensembles. For $\alpha = 1.9$ much smaller than the optimal threshold $\alpha_{\mathrm{MAP}} \approx 1.98267$, three methods of coupling are indistinguishable from each other. For $\alpha = 1.98$ close to the optimal threshold, on the other hand, the average BERs for the SW ensembles converge toward a small BER more quickly than that for the regular coupling. It is worth noting that the $(64, 2, 0.1, 2, 14)$-SW ensemble outperforms the $(128, 2, 0.1, 4, 28)$-SW ensemble. This implies that there is no point in using the SW ensembles with $c > 2$, or that it is difficult to find a good instance from such SW ensembles. In order to investigate the reasons of the quick convergence, we focus on the minimum BERs over all $m = 0, \dots, L-1$, shown

by the lines with pluses, which should be at a training position. We find that the minimum BERs for the SW ensembles converge more quickly than that for the regular coupling. This quick convergence in the training phase reduces the overall number of iterations. Another reason is a slight improvement of the BP threshold: The BP thresholds for the regular coupling and the $(64, 2, 0.1, 2, 14)$-SW ensemble are approximately 1.98958 ($\bar{\alpha} \approx 1.83981$) and 1.99911 ($\bar{\alpha} \approx 1.84617$), respectively.

V. CONCLUSIONS

The SW ensemble of coupling has been proposed to accelerate iterative detection in the training stage for SC CDMA systems. Instances picked up from the ensemble have direct connections between distant training positions. The DE analysis has shown that the proposed method can provide a significant reduction in the number of iterations for highly loaded systems compared to the conventional method of coupling. We conclude that irregular coupling in the training phase can accelerate iterative detection for SC systems.

REFERENCES

[1] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[2] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of BP-based CDMA multiuser detection by spatial coupling," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 1489–1493.

[3] ——, "Performance improvement of iterative multiuser detection for large sparsely-spread CDMA systems by spatial coupling," submitted to *IEEE Trans. Inf. Theory*, 2012, [Online]. Available: http://arxiv.org/abs/1206.5919.

[4] C. Schlegel and D. Truhachev, "Multiple access demodulation in the lifted signal graph with spatial coupling," in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 2989–2993.

[5] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. 48th Annual Allerton Conf. Commun. Control & Computing*, Los Alamos, USA, Sep.–Oct. 2010, pp. 347–353.

[6] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, pp. 021 005–1–18, May 2012.

[7] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 1231–1235.

[8] S. H. Hassani, N. Macris, and R. Urbanke, "Chains of mean field models," *J. Stat. Mech.*, no. 2, p. P02011, Feb. 2012.

[9] K. Takeuchi, T. Tanaka, and T. Kawabata, "A phenomenological study on threshold improvement via spatial coupling," *IEICE Trans. Fundamentals*, vol. E95-A, no. 5, pp. 974–977, May 2012.

[10] D. Truhachev, D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, Jr., "Connecting spatially coupled LDPC code chains," in *Proc. IEEE Int. Conf. Commun. (ICC 2012)*, Ottawa, Canada, Jun. 2012.

[11] ——, "Improving spatially coupled LDPC codes by connecting chains," in *Proc. 2012 IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, Jul. 2012, pp. 468–472.

[12] A. Montanari and D. N. C. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. 2006 IEEE Inf. Theory Workshop*, Punta del Este, Uruguay, Mar. 2006, pp. 160–164.

[13] D. Guo and C.-C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.

[14] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[15] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A: Math. Gen.*, vol. 36, no. 43, pp. 11 111–11 121, Oct. 2003.