Gwangju Institute of
Science and Technology

School of Information and Communications

INFONET Journal Club 2013

Information & System Group

2014. 2. 3

Advisor: Prof. Heung-No Lee

Editor: Seungchan Lee, Haeung Choi

Gwangju Institute of
Science and Technology

BCI CS INFONET

- Table of Contents -

		processing and information theory. We applied this CS technique to WSN which consists of a large number of wireless sensor nodes and a central fusion center (FC). This CS based signal acquisition and compression is done by a simple linear projection at each sensor node. Then, each sensor transmits the compressed samples to the FC. The FC which collects the compressed signals from the sensors jointly reconstructs the signals in polynomial time using a signal recovery algorithm. The distributed sensors observe similar event in designated region. Therefore, the observed signals have considerable correlation each other. We pay some effort in modeling correlation between the signals acquired from the sensors. After modeling the correlated signals, we propose POMP (Phased-OMP) which can recover any type of correlated signals stably and effectively. We introduce the idea of our proposed algorithm in detail and then compare the reconstruction performance of POMP with previous algorithms
5	2012-12-13	Multiuser Cooperative Diversity Through Network Coding Based on Classical Coding Theory
	Jintaek Seong	To increase the diversity order of cooperative wireless communication systems without sacrificing the system's rate, they propose the generalized dynamic-network code (GDNC). They showed that the problem of designing network codes that maximize the diversity order is related to that of designing optimal linear block codes, in the Hamming distance sense, over a nonbinary finite fields.
6	2013-01-24	A fast approach for over-complete sparse decomposition based on smoothed L0 norm
	Oliver	This paper proposes a fast algorithm for overcomplete sparse decomposition.The algorithm is derived by directly minimizing the L0 norm after smoothening. Hence, the algorithm is named as smoothed L0 (SL0) algorithm. The authors demonstrate that their algorithm is 2-3 orders of magnitude faster than the state-of-the-art interior point solvers with same (or better) accuracy.
7	2013-01-10	Statistical physics-based re construction in compressed sensing
	Jaewook Kang	In this report, the author introduces a expectation maximization (EM) based belief propagation algorithm (BP) for sparse recovery, named EM-BP. The algorithm have been mainly devised by Krzakala et al. from ParisTech in France. The properties of EM-BP are as given below:1) It is A low-computation approach to sparse recovery, 2) It works well without the prior knowledge of the signal, 3) It overcomes the l1 phase transition given by Donoho and Tanner under the noiseless setup, 4) It is further improved in conjunction with seeding matrices (or spatial coupling matrices). The main purpose of this report regenerates a precise description of EM-BP derivation from the reference paper. It might be very helpful

		for understanding of EM-BP algorithm, and an answer for such a question: How and why does the algorithm work ? Therefore, we will focus on the explanation of 1) and 2) in the properties, and just show the result of the paper with respect to that of 3) and 4).
8	2013-02-07	Performance Analysis of Iterative Decoding Algorithms with Memory over Memoryless Channels
	Jeongmin	In this work, they propose a model for iterative decoding algorithms with memory which covers successive relaxation (SR) version of belief propagation and differential decoding with binary message passing (DD-BMP) algorithms. Based on this model, they derive a Bayesian network for iterative algorithms with memory over memoryless channels and use this representation to analyze the performance of the algorithms using density evolution.
9	2013-02-14	Faster STORM using compressed sensing
	Eunseok	In super-resolution microscopy methods based on single-molecule switching, the rate of accumulating single-molecule activation events often limits the time resolution. Here we developed a sparse-signal recovery technique using compressed sensing to analyze images with highly overlapping fluorescent spots. This method allows an activated fluorophore density an order of magnitude higher than what conventional single-molecule fitting methods can handle. Using this method, we demonstrated imaging microtubule dynamics in living cells with a time resolution of 3 s.
10	2013-02-21	A Node-Based Time Slot Assignment Algorithm for STDMA Wireless Mesh Networks
	Muhammad Asif Raza	In this paper authors present a link capacity model for spatial time-division multiple access (STDMA) mesh networks. It makes use of a simplified transmission model that also considers channel fading. The model then forms the basis of a node-based slot-assignment and scheduling algorithm. This algorithm enables the user to exploit multiuser diversity that results in optimizes network throughput. The presented algorithm shows significant improvement in the throughput when compared with existing slot-assignment methods
11	2013-02-28	A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration
	Hwanchol Jang	In this paper, the authors introduces TwIST algorithms, exhibiting much faster convergence rate than IST for ill-conditioned problems. For a vast class of nonquadratic convex regularizers, they show that TwIST converges to a minimizer of the objective function, for a given range of values of its parameters.
12	2013-03-21	Scaling Up MIMO: Opportunities and challenges with very large arrays

	Woongbi	Very large MIMO systems, also known as massive MIMO, multiuser MIMO systems, or large-scale antennas systems is an emerging research area in antenna systems, electronics, and wireless communication systems. A base station with an antenna array serves a multiplicity of single-antenna terminals. In this presentation, the fundamental principle of massive MIMO technology and several issues are introduced.
13	2013-03-28	Shrinkage methods to linear regression problems
	Jaewook	In this chapter, we are interested in the linear regression with shrinkage methods. The shrinkage method have got attention to solve the problem of linear systems y=Ax because the method provides very flexible solver from dense signals to sparse signals. First, we will introduce basic concept of two shrinkage methods in the linear regression, Ridge and Lasso. Then, we move the focus to sparse recovery with Lasso and its variants for different problem setting such as Fused lasso and Elastic net.
14	2013-04-11	Efficient Design and Decoding of Polar Codes
	Jeongmin	Polar codes are shown to be instances of both generalized concatenated codes and multilevel codes. It is shown that the performance of a polar code can be improved by representing it as a multilevel code and applying the multistage decoding algorithm with maximum likelihood decoding of outer codes. Additional performance improvement is obtained by replacing polar outer codes with other ones with better error correction performance. In some cases this also results in complexity reduction. It is shown that Gaussian approximation for density evolution enables one to accurately predict the performance of polar codes and concatenated codes based on them.
15	2013-04-18	Aliasing-Free Wideband Beamforming Using Sparse Signal Representation
	J. Oliver	This paper considers the use of sparse signal representation for the wideband direction of arrival (DOA) or angle of arrival estimation problem. In particular, this paper discusses about the two ambiguities, namely, spatial and algebraic aliasing that arise in wideband-DOA. The authors of the paper suggest procedures to avoid the aliasing using multiple measurement vector and multiple dictionaries.
16	2013-05-09	Turbo Reconstruction of Structured Sparse Signal
	Hyeongho	This paper considers the reconstruction of structured-sparse signals from noisy linear observations. In particular, the support of the signal coefficients is parameterized by hidden binary pattern, and a structured probabilistic prior (e.g., Markov random chain/field/tree) is assumed on the pattern. Exact inference is discussed and an approximate inference scheme, based on loopy belief propagation (BP), is proposed. The proposed scheme iterates between

		exploitation of the observation-structure and exploitation of the pattern-structure, and is closely related to noncoherent turbo equalization, as used in digital communication receivers. An algorithm that exploits the observation structure is then detailed based on approximate message passing ideas.
17	2013-05-16	Compressive fluorescence microscopy for biological hyperspectral imaging
	Eunseok	The mathematical theory of compressed sensing (CS) asserts that one can acquire signals from measurements whose rate is much lower than the total bandwidth. Whereas the CS theory is now well developed, challenges concerning hardware implementations of CS-based acquisition devices—especially in optics—have only started being addressed. This paper presents an implementation of compressive sensing in fluorescence microscopy and its applications to biomedical imaging. Our CS microscope combines a dynamic structured wide-field illumination and a fast and sensitive single-point fluorescence detection to enable reconstructions of images of fluorescent beads, cells, and tissues with undersampling ratios (between the number of pixels and number of measurements) up to 32. We further demonstrate a hyperspectral mode and record images with 128 spectral channels and undersampling ratios up to 64, illustrating the potential benefits of CS acquisition for higher-dimensional signals, which typically exhibits extreme redundancy. Altogether, our results emphasize the interest of CS schemes for acquisition at a significantly reduced rate and point to some remaining challenges for CS fluorescence microscopy.
18	2013-05-30	Simplified Relay Selection and Power Allocation in Cooperative Cognitive Radio Systems
	Muhammad Asif	In this paper authors propose solution of a combined problem; relay selection and power allocation to secondary users under the constraint of limited interference to primary users in cognitive radio (CR) system. Objective of the joint problem was to maximize system throughput. A high complexity optimal solution and a low complexity suboptimal solution are proposed. The presented solutions show over 50% improvement in system throughput.
19	2013-06-13	Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit
	Sangjun	This paper is to show that a sufficient condition on the number of measurements for a successful greedy algorithm called Orthogonal Matching Pursuit. We understand how the authors of this paper derive their sufficient condition.
20	2013-06-27	Link Status Monitoring Using Network Coding
	Jin-Taek	This work has presented a novel approach to link status monitoring based on a deterministic approach that exploits linear network coding at the internal nodes in a network. The key problem of

		identifiability for such approaches was highlighted and various insights provided regarding this concept. New sufficient conditions were derived for successfully identifying a congested link in any logical network, and tradeoffs between length of training slots and size of the network coding alphabet established.
21	2013-07-11	A sparse signal reconstruction perspective for source localization with sensor arrays
	J.Oliver	In this paper, the authors present a source localization method based on sparse representation of sensor measurements. In particular, they use SVD of the data matrix obtained from the sensors to summarize the multiple measurements. The SVD summarized data is then sparsely represented in order to detect the sources. The authors also proposed grid refinement in order to mitigate the effects of limiting estimates to a grid of spatial locations. They demonstrate the superior resolution ability with limited time samples of their method over the existing methods via various experiments
22	2013-07-18	Enhancing Iterative Decoding of Cyclic LDPC Codes Using Their Automorphism Groups
	Jeongmin	In this paper they focus on cyclic LDPC codes defined by a circulant parity-check matrix and consider two known subgroups of the automorphism group of a cyclic code. For the large class of idempotent-based cyclic LDPC codes in the literature, they show that the two subgroups only provide equivalent parity-check matrices and thus cannot be harnessed for iterative decoding. Towards exploiting the automorphism group of a code, they propose a new class of cyclic LDPC codes based on pseudo-cyclic MDS codes with two information symbols,for which nonequivalent parity-check matrices are obtained. Simulation results show that for our constructed codes of short lengths, the automorphism group can significantly enhance the iterative decoding performance
23	2013-07-25	Multiuser detection of sparsely spread CDMA
	Jaewook	Abstract: This paper has discussed about design and analysis of multiuser detection (MUD) using sparsely spread CDMA systems. The objective of the MUD problem is how to detect multiple user signals simultaneously at the low computational cost. The main obstacle is multiple-access interference (MAI). These multiple user signals are interference for each user detection one another. The MAI problem arise in most CDMA systems, and optimal detection in such systems requires exponentially growing computation as the number of user increases. But a good news is that the simultaneous users in time is very few. Therefore, this paper investigates a suboptimal MUD detection using sparse CDMA systems. The key idea of the proposed system is to encode the transmitted waveforms using sparse spread CDMA codes and detect the signal using a linear-complexity belief propagation (BP) algorithm. We summarize the contributions of this work is following:

		- Description the sparse CDMA system - Properties of the sparsly spread CDMA codes for the convergence of the BP algorithm - Design of the BP algorithm for the MUD problem - Asymptotic analysis of performance of the BP algorithm based MUD detection In this report, we aim to sketch the key point of each contribution of this paper.
24	2013-08-01	Compressive Sensing for Spread Spectrum Receivers
	Hyeongho	This paper investigates the use of Compressive Sensing(CS) in a general Code Division Multiple Access (CDMA) receiver. They show that when using spread spectrum codes in the signal domain, the CS measurement matrix may be simplified. Furthermore, they numerically evaluate the proposed receiver in terms of bit error rate under different signal to noise ratio conditions and compare it with other receiver structures.
25	2013-08-08	Active illumination single-pixel camera based on compressive sensing
	Eunseok	This paper is organized as follows. After a brief introduction (Section 1), some mathematical back- ground essential to the understanding of CS is shown (Section 2). Then, in Section 3, CS is presented along with some of its principal properties. Section 4 explains why the ℓ1-norm is such a good option for com- pressive sensing. Some insights about the robustness of CS in the presence of noise are given in Section 5. Next, in Section 6, the single-pixel camera developed at Rice University is discussed. Subsequently, the innovative active illumination single-pixel camera developed in the scope of the current work is described. Following that, experimental results from the single-pixel cameras are presented. In the end, the main conclusions of this work are exposed.
26	2013-08-29	Resource Allocation in Cognitive Radio Relay Networks
	Muhammad Asif	In this paper authors formulate the problem of Resource Allocation (RA) in Cognitive Radio (CR) networks with relay stations. The problem takes into account the issues like: fluctuations of usable spectrum resource, channel quality variations caused by frequency selectivity, and interference caused by different transmit power levels. They propose easy to implement heuristic algorithms. The simulation results reveal that presented solutions show good proportional fairness among CR users and improvement in system throughput by power control.
27	2013-09-16	Multipath Matching Pursuit
	Hwanchol Jang	In this paper, they propose an algorithm referred to as multipath matching pursuit that investigates multiple promising candidates to recover sparse signals from compressed measurements. Their method is inspired by the fact that the problem to find the candidate

		that minimizes the residual is readily modeled as a combinatoric tree search problem and the greedy search strategy is a good fit for solving this problem. In the empirical results as well as the restricted isometry property (RIP) based performance guarantee, they show that the proposed MMP algorithm is effective in reconstructing original sparse signals for both noiseless and noisy scenarios.
28	2013-10-14	The Exact support recovery of sparse signals with noise via orthogonal matching pursuit
	Oliver	This letter derives sufficient conditions for the OMP to recover the support set of a sparse vector from noise corrupted measurements. In particular, the conditions are given in terms of the minimum absolute values of the signal amplitudes. That is, if the minimum values of the non-zero coefficient of the signal satisfies certain conditions then OMP guarantees exact support recovery.
29	2013-10-21	Hierarchical and High-Girth QC LDPC codes
	Jeongmin Ryu	They present an approach to designing capacity approaching high-girth low-density parity-check (LDPC) codes that are friendly to hardware implementation, and compatible with some desired input code structure defined using a protograph. The approach is based on a mapping of any class of codes defined using a protograph into a family of hierarchical quasi- cyclic (HQC) LDPC codes. Next, they present a girth-maximizing algorithm that optimizes the degrees of freedom within the family of codes to yield a high-girth HQC LDPC code, subject to bounds imposed by the fact that HQC codes are still quasi-cyclic. Finally, they discuss how certain characteristics of a code protograph will lead to inevitable short cycles and show that these short cycles can be eliminated using a "squashing" procedure that results in a high-girth QC LDPC code.
30	2013-10-26	Ultra-Wideband Compressed Sensing : Channel Estimation
	JuSung Kang	In this paper, they have introduced two novel ultra-wideband (UWB) channel estimation approaches based on compressive sensing (CS). The proposed approach relies on the fact that transmitting an ultra-short pulse through a multipath UWB channel leads to a received UWB signal that can be approximated by a linear combination of a few atoms from a pre-defined dictionary which means sparse representation of the received signal. The key in the proposed approach is in the design of a dictionary of parameterized waveforms (atoms) that closely matches the information-carrying pulse shape leading thus to higher energy compaction and sparse representation, and, therefore higher probability for CS reconstruction. In the first approach, the CS reconstruction capabilities are exploited to recover the composite pulse-multipath channel from a reduced set of random projections. This reconstructed signal is subsequently used as a referent template in a correlator-based detector. In the second approach, from a set of random projections of the received pilot signal, the Matching Pursuit

		algorithm is used to identify the strongest atoms in the projected signal that are related to the strongest propagation paths that composite the multipath UWB channel.
31	2013-11-18	Missing-Area Reconstruction in Multispectral Images Under a Compressive Sensing Perspective
	Hyeongho Baek	The intent of this paper is to propose new methods for the reconstruction of areas obscured by clods. They are based on compressive sensing theory, which allows finding sparse signal representations in underdetermined linear equation systems.
32	2013-11-25	Power and Channel Allocation for Cooperative Relay in Cognitive Radio Networks
	Muhammad asif	In this paper authors mention that cognitive radio relay channels can be divided into three categories: direct, dual-hop, and relay channels. The relay node involves both dual-hop and relay diversity transmission. They develop power and channel allocation approaches for cooperative relay networks. They also develop a low complexity approach that can obtain most of the benefits from power and channel allocation with minor performance loss.
33	2013-12-09	Robust Compressive Data Gathering in Wireless Sensor Networks
	Jongmok Shin	In this paper, authors investigate the impact of outlying sensor readings and broken links on high-fidelity data gathering, and propose approaches based on the compressive sensing theory to identify outlying sensor readings and derive the corresponding accurate values, and to infer broken links in Wireless Sensor Networks.

A Node-Based Time Slot Assignment Algorithm for STDMA Wireless Mesh Networks

Authors: W. Chen, and Chin-Tau Lea

Publication: IEEE Trans. Veh. Tech., Jan. 2013
Speaker: Asif Raza

Short summary: In this paper authors present a link capacity model for spatial time-division multiple access (STDMA) mesh networks. It makes use of a simplified transmission model that also considers channel fading. The model then forms the basis of a node-based slot-assignment and scheduling algorithm. This algorithm enables the user to exploit multiuser diversity that results in optimizes network throughput. The presented algorithm shows significant improvement in the throughput when compared with existing slot-assignment methods.

I. Introduction

In STDMA network the transmission time of a channel is divided into slots where multiple slots constitute a frame. These slots are assigned to potential users of the network. The goal of slot assignment scheme is to maximize network throughput. Existing assignment algorithms in STDMA make use of simplified transmission model which do not consider the time-varying fading behavior of a wireless channel. This results in slot wastage when link is in deep fade. The slot is also wasted if scheduled link has no traffic to transmit. This degrades the STDMA network throughput. Therefore a dynamic slot-assignment with that should exploit multiuser diversity is required. However sheer complexity involved in coordinating with all nodes and generating scheduling map in a reasonable time makes this approach impractical. In order to fix these issues the authors present a node-based slot-assignment scheme in which scheduling in each slot is done for nodes not for links. Their contributions include:

- Defining link capacity: a model that includes channel fading. It ensures that whichever link is used by a node will not change the interference profiles on the links selected by other users.
- Node-based time-slot assignment and scheduling algorithms.

II. System Model

Wireless STDMA mesh network with fixed routers.

Transmissions are organized in frames.

Synchronization among nodes provided through GPS.

Set of nodes are identified and assigned to a slot for their transmission.

Each node maintains a separate queue for each outgoing link and performs scheduling without coordination with other nodes.

Multiprotocol Label Switching (MPLS) multipath routing is used for routing however packets are transmitted in sequence.

Adaptive modulation and time varying fading channels are considered. It is also assumed that wireless channels undergo slow fading. Due to fading channel an instant channel gain will be fed back to transmitter. The duration for feedback is no longer than coherence time (the time for which channel conditions remain same)

Adaptive modulation is implemented that each data packet can be fragmented into multiple segments and each segment can be transmitted in with lowest data rate. If high data rate is available then multiple segments can be transmitted per slot duration.

III. Link Capacity Modeling

Each node has multiple links and it can exploit multiuser diversity i.e. different links have different traffic and fading conditions. A channel model is presented that includes shadowing and slow fading.

A. Signal to interference and noise ratio (SINR) Formulation

$h_{r,t}$: Channel response function from transmitter 't' and receiver 'r'

x_t: Signal from 't'

$I_r, n_I, t_i^{'}$: Set of transmitters causing interference to 'r', number of transmitters and i[th] transmitter in I_r respectively. Power control is not considered therefore transmission power of 't' is $pt = E\left(\left|x_t\right|^2\right)$. Let n_0 be thermal noise with power equal to k then received power at 'r' is

$$y_{r,t} = h_{r,t}x_t + \sum_{i=1}^{n_I} h_{r,t_i^{'}} x_{t_i^{'}} + n_0 \qquad (1)$$

SINR at receiver 'r' is expressed as:

$$\gamma_{r,t} = \frac{\left|h_{r,t}x_t\right|^2}{\sum_{i=1}^{n_I}\left|h_{r,t_i^{'}}x_{t_i^{'}}\right|^2 + \kappa} = \frac{\left|h_{r,t}\right|^2 p_t}{\sum_{i=1}^{n_I}\left|h_{r,t_i^{'}}\right|^2 p_{t_i^{'}} + \kappa} = \frac{s_0}{\sum_{i=1}^{n_I} s_i + \kappa} \qquad (2)$$

Here $s_0 = \left|h_{r,t}\right|^2 p_t$ *and* $s_i = \left|h_{r,t_i^{'}}\right|^2 p_{t_i^{'}}$. The Channel response function consists of three parts:

- Path loss
- Shadowing
- Fading

$$h_{r,t} = \sqrt{l_{r,t}^{-\alpha} 10^{\frac{f_{r,t}}{10}} \pi_{r,t}} \qquad (3)$$

Where $l_{r,t}$ is distance between 't' and 'r', $\alpha \mapsto [2-4]$ (constant), $10^{\frac{f_{r,t}}{10}}$ is shadowing effect and it is modeled as a log-normal distributed random variable. $\pi_{r,t}$ is fading effect and it is defined as complex Gaussian RV with mean and variance equal to 0 and 1 respectively. PDF of s_0 and s_i are defined as:

$$p_{s_0}(\alpha_0) = \tfrac{1}{\rho_0} e^{-\frac{\alpha_0}{\rho_0}} \qquad (4a)$$

$$p_{s_i}(\alpha_i) = \tfrac{1}{\rho_i} e^{-\frac{\alpha_i}{\rho_i}} \qquad (4b)$$

Here $\rho_0 = E(s_0) = l_{r,t}^{-\alpha} 10^{\frac{f_{r,t}}{10}} p_t$ and $\rho_i = E(s_i) = l_{r,t_i'}^{-\alpha} 10^{\frac{f_{r,t_i'}}{10}} p_{t_i'}$

B. PDF of SINR

Case 1: no interference is observed by receiver 'r' i.e. (I_r=0, n_I=0) then PDF of $\gamma_{r,t}$, is defined as: let $\delta = \gamma_0 / \kappa$

$$p_{\gamma_{r,t}}(z) = p_{s_0}(z\kappa) = \tfrac{1}{\delta} e^{-\frac{z}{\delta}} \qquad (5)$$

Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr(\gamma_{r,t} \geq w) = \int_w^\infty p_{\gamma_{r,t}}(z)dz = e^{-\frac{w}{\delta}} \qquad (6)$$

Case 2: unit interference is observed by 'r' i.e. (I_r>0, n_I=1) then PDF of term (s_i+k i.e. denominator of equ.2) is defined as:

$$p_{s_1+\kappa}(v) = \tfrac{1}{\sigma_1} e^{-\frac{v-\kappa}{\sigma_1}} \qquad (7)$$

Finally PDF of $\gamma_{r,t}$ is defined as:

$$p_{\gamma_{r,t}}(z) = \int_\kappa^\infty v p_{s_0}(vz) p_{s_1+\kappa}(v)dv = \int_\kappa^\infty \tfrac{v}{\sigma_0} e^{-\frac{vz}{\sigma_0}} \tfrac{1}{\sigma_1} e^{-\frac{v-\kappa}{\sigma_1}} dv = \tfrac{\kappa + \frac{1}{u}}{\sigma_0 \sigma_1 u} e^{\left(\frac{\kappa}{\sigma_1} - u\kappa\right)} \qquad (8)$$

Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr(\gamma_{r,t} \geq w) = \int_w^\infty p_{\gamma_{r,t}}(z)dz = \frac{\sigma_0}{\sigma_0 + w\sigma_1} e^{-\frac{w}{\delta}} \qquad (9)$$

Case 3: more than one interferers are present in I_r i.e. ($n_I > 1$) then PDF of $\sum_{i=1}^{n_I} s_i$ can be defined as:

$$p_I(v) = P_{\sum_{i=1}^{n_I} s_i}(v) \otimes p_{s_{n_I}}(v) = \sum_{i=1}^{n_I} \frac{b_i}{\sigma_i} e^{-\frac{v}{\sigma_i}} \qquad (10)$$

Here $b_i = \prod_{j=1, j\neq i}^{n_I} \left(\sigma_i / \sigma_i - \sigma_j \right) and \sum_{i=1}^{n_I} b_i = 1$. The PDF of term ($s_i$+k i.e. denominator of equ.2) is defined as: $p_{I+\kappa}(v) = p_I(v-\kappa)$. Finally the PDF of $\gamma_{r,t}$:

$$p_{\gamma_{r,t}}(z) = \int_{\kappa}^{\infty} v p_{s_0}(vz) p_{I+\kappa}(v) dv = \sum_{i=1}^{n_I} d_i \left(\frac{\kappa}{q_i} + \frac{\kappa}{q_i^2} \right) e^{-\kappa q_i} \qquad (11)$$

Here $q_i = \left(z / \sigma_0 \right) + \left(1 / \sigma_i \right)$ and $d_i = \left(b_i / \sigma_0 \sigma_i \right) e^{-\kappa q_i}$. Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr\left(\gamma_{r,t} \geq w \right) = \int_{w}^{\infty} p_{\gamma_{r,t}}(z) dz = \sigma_0 e^{-\frac{w}{\delta}} \sum_{i=1}^{n_I} \frac{b_i}{\sigma_0 + \sigma_i w} \qquad (12)$$

Finally Link Capacity can then be determined as:

$$c_{r,t}(I_r) = \sum_{i=1}^{\xi-1} c_i \cdot Pr\left(\gamma_{thr}^{i} \leq \gamma_{r,t} < \gamma_{thr}^{i+1} \right) + c_{\xi} \cdot Pr\left(\gamma_{r,t} \geq \gamma_{thr}^{\xi} \right)$$

Where $c_{r,t}(I_r)$ is average data rate between 't' and 'r', given interference set I_r $c_{r,t}(I_r))$ $c_{r,t}(I_r)$

IV. Proposed Time-Slot Assignment Algorithm

TDMA frame consists of a fixed number of slots is considered. The set of transmitting links that are activated in a given slot is called a **link pattern**, and the set of nodes activated in a given slot is called a **node pattern**.

A. Formulation of Node-Based Time-Slot Algorithm

Notations:

V: set of nodes

E: set of links

NP: Node Pattern

$tx_e, rx_e; e \in E$ transmitter and the receiver of link e, respectively,

$E_{s,p} = \{ e \mid e \in E, p \in s, s \in NP, tx_e = p \}$ set of links that can be used at node p, where p∈s i.e p is activated in node pattern *s*);

μ_s = portion of time that is assigned to node pattern (s) in a frame, where,

$\sum_{s \in NP} \mu_s = 1$

$\left\{ò_{s,p,e} \mid e \in E_{s,p}\right\}$:portion of time that is assigned to each link of node p in node pattern s

F: set of flows in the system; where flow defines all traffic that belongs to (S, D) pair

h_f : traffic demand for flow f, where $f \in F$

S_f : source of flow f

D_f : destination of flow f

$x_{f,e}$: percentage of traffic that flow f passes through link e,

Calculations

Link congestion: it is total amount of traffic routed through the link 'e' over its average capacity (c_e) i.e. $r_e = \left(\sum_{f \in F} \frac{x_{f,e}}{c_e} \right)$ where link capacity (data rate between transmitter 't' and receiver 'r' is defined as: $c_e = \sum_{\{s \mid s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} ò_{s,p,e}$.

Thus network congestion ratio 'r' is the maximum of all link congestion ratios, i.e. $r = \max_{e \in E} r_e$

Optimal node-based slot assignment scheme is one which minimizes congestion 'r':

$$\min \quad r \qquad (13a)$$

$$s.t \quad \frac{\sum_{f \in F} x_{f,e} h_f}{\sum_{\{e \mid s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} ò_{s,p,e}} \le r \quad (13b)$$

$$\sum_{\{e \mid e \in E_{s,p}\}} ò_{s,p,e} \le \mu_s \qquad (13c)$$

$$\sum_{\{e \mid p \in s, e \in E_{s,p}, rx_e = q\}} ò_{s,p,e} \le \mu_s \qquad (13d)$$

$$\sum_{s \in NP} \mu_s = 1 \qquad (13e)$$

$$\mu_s \ge 0, ò_{s,p,e} \ge 0, r \ge 0 \qquad (13f)$$

$$\sum_{\{e \mid tx_e = v\}} x_{f,e} - \sum_{\{e \mid rx_e = v\}} x_{f,e} = 0 \quad (13g)$$

$$\sum_{\{e \mid tx_e = S_f\}} x_{f,e} - \sum_{\{e \mid rx_e = S_f\}} x_{f,e} = 1 (13h)$$

$$x_{f,e} \ge 0 \qquad (13i)$$

Problem 13 is the optimization problem, whose purpose is to find the set of $\grave{o}_{s,p,e}$ that will lead to the optimal objective function. Constraint 13c represents that in node pattern s, for any node p∈s, p can transmit to only one node at one time. 13d ensures that a node q can receive from only one node at one time while q ∉ s and p∈s. 13f and 13i ensures non-negativity constraints. Constraint 13b is non-linear therefore $\mu_s r$ and $\grave{o}_{s,p,e} r$ are replaced by $\theta_s \; and \; \rho_{s,p,e}$ respectively. Therefore final formulation is defined as:

$$\min \quad \sum_{s \in NP} \theta_s \qquad (14a)$$

$$s.t \quad \sum_{f \in F} x_{f,e} h_f \le \sum_{\{e | s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} \grave{o}_{s,p,e} \qquad (14b)$$

$$\sum_{\{e | e \in E_{s,p}\}} \rho_{s,p,e} \le \theta_s \qquad (14c)$$

$$\sum_{\{e | p \in s, e \in E_{s,p}, rx_e = q\}} \rho_{s,p,e} \le \theta_s \qquad (14d)$$

$$\theta_s \ge 0, \rho_{s,p,e} \ge 0 \qquad (14e)$$

$$\sum_{\{e | tx_e = v\}} x_{f,e} - \sum_{\{e | rx_e = v\}} x_{f,e} = 0 \qquad (14f)$$

$$\sum_{\{e | tx_e = S_f\}} x_{f,e} - \sum_{\{e | rx_e = S_f\}} x_{f,e} = 1 \qquad (14g)$$

$$x_{f,e} \ge 0 \qquad (14h)$$

Authors describe that the presented formulation can handle scheduling of node patterns by using Linear Programming approach. However for link based approach, listing all link patterns does not work by using LP formulation. Therefore column generation method is used to tackle the problem.

B. Frame Construction and Throughput Loss due to Frame Quantization

Frame is constructed as: $n_f = \sum_{s \in NP} \left[z \mu_s \right]$ here z is frame length and function [x] rounds 'x' to nearest integer.

The frame quantization will change the portion of time assigned to all patterns (μ_s). Therefore parameters like minimum congestion ratio r_z, the optimal link capacities (c_e) and the routing scheme $x_{f,e}$

will change. These parameters need to be recomputed as follows. let z_s be number of slots assigned to node pattern 's' in a frame.

$$\min \quad \varsigma_z \qquad (15a)$$

$$s.t \quad \sum_{f\in F} y_{f,e} h_f \le \sum_{\{s|s\in NP, p\in s, e\in E_{s,p}\}} c_{s,e} \grave{o}_{s,p,e} \qquad (15b)$$

$$\sum_{\{e|e\in E_{s,p}\}} \grave{o}_{s,p,e} \le \frac{z_s}{\sum_{s\in NP} z_s} \qquad (15c)$$

$$\sum_{\{e|p\in s, e\in E_{s,p}, rx_e=q\}} \grave{o}_{s,p,e} \le \frac{z_s}{\sum_{s\in NP} z_s} \qquad (15d)$$

$$\sum_{\{e|tx_e=v\}} y_{f,e} - \sum_{\{e|rx_e=v\}} y_{f,e} = 0 \qquad (15e)$$

$$\sum_{\{e|tx_e=S_f\}} y_{f,e} - \sum_{\{e|rx_e=S_f\}} y_{f,e} = h_f \varsigma_z \qquad (15f)$$

$$y_{f,e} \ge 0,\ \grave{o}_{s,p,e} \ge 0 \qquad (15g)$$

Here $\varsigma_z = 1/r_z, \mu_s = \left(z_s / \sum_{s\in NP} z_s \right), y_{f,e} = \left(x_{f,e} / r_z \right) = x_{f,e}\varsigma_z$

C. Column Generation Method

Column generation is an algorithm for solving large LP problems. Most of the variables are usually non-basic and assume zero values in the optimal solution, only a subset of variables are needed for solving the problem. Column generation method considers only the variables which have potential to improve the objective function. It splits the problem into master problem and subproblem. Master problem is the original problem with subset of variables being considered. In subproblem it uses duality approach to select new variables to be added to master problem to improve its result.

Master Problem: it is same as defined in problem 14 except that NP is replaced with NP' (subset of NP which is feasible for 14). Solution of master problem shall provide a routing and slot-assignment scheme.

Subproblem: is a new problem created to identify a new node pattern to add to master problem and it is defined as:

$$\min_{s\in\{NP\setminus NP'\}} rp_s \qquad (15)$$

Here rp_s is reduced cost of node pattern 's' in the column generation algorithm and it is optimal value of following problem:

$$\max\left(1-\sum_{\{p|p\in s\}}\omega_{s,p}-\sum_{\{q|p\in s,e\in E_{s,p},rx_e=q\}}\tau_{s,q}\right) \quad (16)$$
$$s.t \qquad \omega_{s,p}+\tau_{s,p}-\phi_e c_{s,e}\geq 0$$

Here $\omega_{s,p}\ and\ \tau_{s,q}$ are variables that are associated with the transmitter p and the receiver q in node patterns s. Well the question is which node pattern should be included into NP' ?

According to duality theory if master problem is optimal then rp_s is always non-negative for any pattern in NP. The node patterns with negative rp_s can improve the result if they are added into NP'. So algorithm will iterate between two phases until no more patterns can be added to NP'.

Algorithm steps are defined as follows:

Step 1: Set node pattern A= ϕ and rp_A =0,

Step 2: Identify $A^c = v$ and compute $rp_{A'_v}$ for node pattern $A'_v\, s.t\, A'_v = \{A, v\}$.

Step 3: select v from A^c with minimum $rp_{A'_v}$ and compute rp_A of A.

Step 4: If $rp_{A'_v} \geq rp_A$, node v will be deleted from A^c and add it A.

Step 5: If $A^c \neq \phi$ stop else go to step 3.

V. SCHEDULING ALGORITHMS

Two scheduling algorithms are proposed in which each node will locally schedule its link transmissions without inter-node coordination and without disturbing interference profiles of other nodes.

A. Scheme 1

Every node 't' in node pattern 's' assigns a transmission probability to every link associated with 't'. The set of transmission probabilities is then defined as:

$$P_{s,t}=\left\{p_{s,t,e}\,|\,e\in E_{s,t}, p_{s,t,e}=\frac{\partial_{s,t,e}}{\mu_s}\right\} \qquad (17)$$

The region [0,1] is then divided into subregions, one for each link in $|E_{s,t}|$, and length of regions is set according to $P_{s,t}$. The algorithm works as follows; Suppose a node pattern 's' is activated in slot x. Each node $t \in s$ will generate a RV w, uniformly distributed within [0, 1]. The node will then schedule

link into which subregion w falls. If selected link (e) is not usable (either due to fading or no traffic) the scheduler will check link next to 'e' one by one until a usable link is found.

B. Scheme 2

Scheme 1 does not consider link quality while scheduling the links. Therefore authors presented another scheduling mechanism.

Selection criteria = (queue length * link capacity)

Each node maintains two queues for each of its link:

1) a real data queue to store packets and

2) A shadow queue for scheduling.

These queues of link 'e' whose transmitter can be activated in slot 'x', are defined as:

$$q_e(x) = q_e(x-1) + a_e(x) - d_e(x)$$
$$q'_e(x) = q'_e(x-1) + a'_e(x) - d'_e(x)$$

Here $q_e(x)$ and $q'_e(x)$ are lengths of the real queue and shadow queue respectively. $a_e(x)$, $a'_e(x)$, $d_e(x)$ and $d'_e(x)$ are the number of arrivals and departures for the two queues in 'x', respectively. In shadow queue the term $a'_e(x)$ is defined as: $a'_e(x) = \left(1 + \gamma_x\right)\sum_{t=0}^{x} a_e(t)$ i.e. it is used to smooth the incoming traffic from source or previous hop.

Packets departing from link 'e' are defined as: $d_e(x) = \min\left\{\tilde{c}_e(x), q_e(x)\right\}$. Here $\tilde{c}_e(x)$ is instant capacity of link 'e' in slot 'x'. Thus scheduling, in slot 'x', the scheduler in node $t \in s$ will select the link from all its associated links with a maximum value of $q'_e(x)\tilde{c}_e(x)$. In doing so, it tries to strike the optimal balance between link quality and traffic backlog.

VI. Simulation and Results

A. Simulation Environment and Settings

Linear optimization toolbox of MATLAB is used for proposed routing and slot-assignment algorithm. C++ program is then used to inspect maximum achievable throughput for different scheduling schemes.

The physical-layer parameters are summarized as follows:

- Transmission power: 20 dBm.
- Thermal noise:−90 dBm.
- Path loss(α):3.5.
- Variance of shadow fading: 4 dBm.

• Minimal distance of two nodes: 15 m.

• Slot duration: 0.22 ms.

• Frame size: 100 slots. So frame length = 22 ms.

The mapping between the following data rates and SINR threshold is summarized as follows.

• 54 Mb/s: 24.56 dBm.

• 48 Mb/s: 24.05 dBm.

• 36 Mb/s: 18.80 dBm.

• 24 Mb/s: 17.04 dBm.

• 18 Mb/s: 10.79 dBm.

• 12 Mb/s: 9.03 dBm.

• 9 Mb/s: 7.78 dBm.

• 6 Mb/s: 6.02 dBm.

Network Topology: two networks 15-node and 30-node with two gateway nodes and three gateway nodes are considered, respectively.

The traffic load of each flow is assumed to be the same i.e.,h_f =1Mb/s,

Throughput loss due to Frame Quantization:

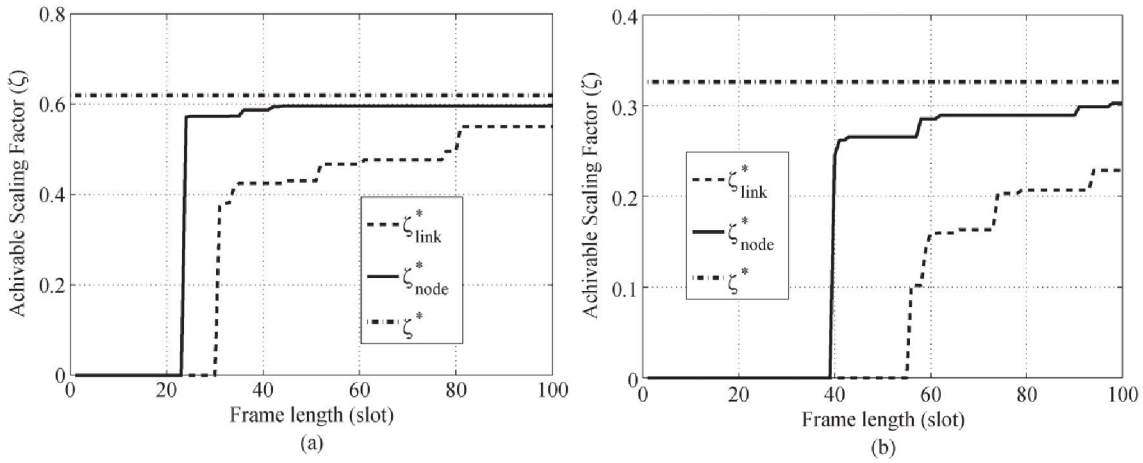


Fig. 1. Achievable throughput after frame generation for (a) 15- and (b) 30-node networks.

The solid (ζ^*_{node}), dashed (ζ^*_{link}) and dashed–dotted (ζ^*) lines indicate the achievable throughput in node-based, link based and before frame construction (I.e. upper bound on throughput) respectively. The flat area represents the range where the performance does not improve. Note that (ζ^*_{node}) and (ζ^*_{link}) are function of 'z' and are not always monotonically increasing due to the quantization involved in the process, and small oscillation occurs within a short range of z. This is why, in Fig. 1(a) and (b), the curves move up in steps.

TABLE I
PERFORMANCE COMPARISONS FOR THE LINK- AND NODE-BASED SCHEMES

N	ζ^*	ζ^*_{node} (ζ^*_{link})	n_{use}	$\zeta^*_{exp,1}$	$\zeta^*_{exp,2}$	$\zeta^*_{det,1}$	$\zeta^*_{det,2}$
15 (Node)	0.62	0.60	17	0.62	0.77	0.62	0.76
15 (Link)	0.62	0.55	26	NA	NA	NA	NA
30 (Node)	0.33	0.30	30	0.31	0.39	0.31	0.39
30 (Link)	0.33	0.23	54	NA	NA	NA	NA

From the table it is clear that ζ^*_{node} approaches ζ^* much faster than ζ^*_{link}. Moreover difference in throughput between ζ^*_{node} and ζ^*_{link} is also significant as shown in table 1.

The optimal scaling factors of the ζ^*_{node} for schemes 1 and 2 under the Poisson and a deterministic arrival process are denoted as shown by $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{exp},2}$, $\zeta^*_{\text{det},1}$, $\zeta^*_{\text{det},2}$ respectively.

{the ζ^*_{node} is derived from problem 14 and it does not include multi-user diversity gain. Therefore, it can be viewed as a lower bound of the two proposed scheduling schemes 1 and 2 As shown in the Table I.

It is also clear from the table that , both (posisson and deterministic arrival rates) $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{det},1}$ are only slightly larger than ζ^*_{node} for the 15- and 30-node networks. The difference is only about 3%. This is because scheme 1 tries to follow $\grave{o}_{s,p,e}$ i.e. portion of time that is assigned to each link of node p in node pattern s and does not select a link with the best quality. However, the situation is different in scheme 2, because link quality is part of the selection criteria. With scheme 2, $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{det},1}$ are about 26% larger than ζ^*_{node} for the 15-node network and 30% larger for the 30-node network.

On the Recovery Limit of Sparse Signal Using Orthogonal Matching Pursuit

Authors: Jian Wang and Byonghyo Shim

Publication: IEEE T. Signal Processing, SEPT.2012
Speaker: Sangjun Park

Short summary: In the paper, the authors give a sufficient condition of the Orthogonal Matching Pursuit (OMP) algorithm. In [2], Wakin and Davenport insisted that OMP can reconstruct any K sparse signal if $\delta_{K+1} < 1/\left(3\sqrt{K}\right)$, where δ_K is the restricted isometry constant. However, in this talk, an improved sufficient condition that guarantees the perfect recovery of OMP is presented

I. FINAL SUMMARY OF THE PAPER

a. A strategy of the proof of Theorem 1.

1) We aim to find a condition such that the OMP algorithm selects a correct index in the first iteration.

=> We need to show that $\min_{i\in\mathcal{I}} \left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right| > \max_{j\notin\mathcal{I}} \left|\langle \mathbf{a}_j, \mathbf{y}\rangle\right|$. (e.g., see from (7) to (10).)

2) Let us suppose that the initial k iterations of the OMP algorithm are successful, and that $\mathcal{T}^k$ is the estimated support set after the initial k iterations. Now, the OMP algorithm selects a correct index, which belongs to $\mathcal{I}\setminus\mathcal{T}^k$, in the $k+1$ iteration.

=> Clearly, $\mathcal{T}^k \subseteq \mathcal{I}$, therefore $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} = \mathbf{P}^{\perp}_{\mathcal{T}^k}\mathbf{y} \in span\left(\mathbf{A}_{\mathcal{I}}\right)$ can be considered as a linear combination of the K columns of $\mathbf{A}_{\mathcal{I}}$. Thus, $\mathbf{r}^k = \mathbf{A}\mathbf{b}$, where $\|\mathbf{b}\|_0 \le K$, and $\mathrm{supp}(\mathbf{b}) \subseteq \mathrm{supp}(\mathbf{x}) = \mathcal{I}$.

=> Again, we find a condition such that the OMP algorithm selects a correct index t^{k+1} which belongs to $\mathrm{supp}(\mathbf{b})$.

=> Furthermore, for any $i \in \mathcal{T}^k$, we have $\left|\langle \mathbf{r}^k, \mathbf{a}_i\rangle\right| = 0$. Thus, $t^{k+1} \in \mathcal{I}\setminus\mathcal{T}^k$.

3) Thus, we conclude that the OMP algorithm can reconstruct K sparse signal provided that the condition of 1) and the condition of 2) are satisfied

b. Comparison between the result by the authors in this paper and the result by Davenport and Wakin.

According to the authors, the improvement is possible due to 1) contradiction based construction of the success condition in the first iteration ($\min_{i\in\mathcal{I}}\left|\langle \mathbf{a}_i,\mathbf{y}\rangle\right| > \max_{j\notin\mathcal{I}}\left|\langle \mathbf{a}_j,\mathbf{y}\rangle\right|$) , and 2) observation that the residual in the general iteration preserves the sparsity level of the input signal. ($\mathbf{r}^k = \mathbf{Ab}$, where $\|\mathbf{b}\|_0 \le K$, and $\mathrm{supp}(\mathbf{b}) \subseteq \mathrm{supp}(\mathbf{x}) = \mathcal{I}$).

In fact, the authors again improved the result by Davenport and Wakin.

The more detailed explanations are referred to the paper.

c. Future Works

1) Can we apply the techniques, which are used in the proof, to find a sufficient condition of an algorithm based from the OMP algorithm? For example, the SOMP algorithm selects a index i such as $\arg\max_i \left\|\mathbf{a}_i^T\mathbf{R}^{(k)}\right\|_q$, where $\mathbf{R}^{(k)} = \left[\mathbf{r}_1^{(k)} \quad \cdots \quad \mathbf{r}_S^{(k)}\right]$, $\mathbf{r}_i^{(k)} = \mathbf{y}_i - \mathbf{A}_{T^k}\hat{\mathbf{x}}_{i,T^k}$, and $q = 1 \text{ or } 2$. Can we find a condition such that the SOMP algorithm selects a correct index?

II. History of sufficient conditions of the OMP algorithm

In the below table 1, sufficient conditions that the OMP algorithm reconstructs a K spars signal from a set of linear measurements $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{A} \in \Re^{M\times N}$ ($N > M$), are given.

Year	A sufficient condition
2007[1]	$\mu < 1/(2K-1)$
2010[2]	$\delta_{K+1} < 1/\left(3\sqrt{K}\right)$

Besides, there are many theoretical papers which analyze algorithms based on the OMP algorithm. In here, it is not scope of this seminar. Therefore, we do not care about them.

III. System Model

Let us consider the below equation:

$$\mathbf{y} = \mathbf{Ax}, \tag{1}$$

where $\mathbf{A} \in \Re^{M\times N}$ ($N > M$), and $\mathbf{x} \in \Re^N$ is a K sparse signal, and $\mathbf{y} \in \Re^M$ is a set of linear measurements. The smallest constant δ_K called “the restricted isometry constant” satisfies

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1+\delta_K)\|\mathbf{x}\|_2^2 \tag{2}$$

for any K sparse signal $\mathbf{x}$.

IV. Main Results

A. Improved Recovery Bound of the OMP algorithm

Theorem 1: For any K sparse signal $\mathbf{x}$, the OMP algorithm perfectly reconstructs $\mathbf{x}$ from $\mathbf{y}$ if the isometry constant δ_{K+1} satisfies

$$\delta_{K+1} < \frac{1}{\sqrt{K}+1}. \tag{3}$$

In this talk, we try to understand a proof of Theorem 1.

Before we study the proof, let us consider whether the OMP algorithm perfectly reconstructs $\mathbf{x}$ or not if $\delta_{K+1} = 1/\sqrt{K}$.

B. The OMP algorithm can fail under $\delta_{K+1} = 1/\sqrt{K}$.

Example 1: Let us consider the problem of reconstructing a K sparse signal $\mathbf{x} \in \Re^{K+1}$ such as $x_{K+1} = 0$, and $x_i = 1$ for $i = 1, \cdots, K$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$, where

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & & \vdots \\ \vdots & & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix} \in \Re^{(K+1)\times(K+1)}.$$

Obviously, all the Eigen values of $\mathbf{A}^T\mathbf{A}$ are $\lambda_1 = \lambda_2 = \cdots = \lambda_K = 1-b$, and $\lambda_{K+1} = 1+Kb$. (See Example 1 on Appendix). When we assume $b = -1/\left(K\sqrt{K}\right)$, $\mathbf{A}^T\mathbf{A}$ becomes

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & -1/\left(K\sqrt{K}\right) & \cdots & -1/\left(K\sqrt{K}\right) \\ -1/\left(K\sqrt{K}\right) & 1 & & \vdots \\ \vdots & & \ddots & -1/\left(K\sqrt{K}\right) \\ -1/\left(K\sqrt{K}\right) & \cdots & -1/\left(K\sqrt{K}\right) & 1 \end{bmatrix} \in \Re^{(K+1)\times(K+1)}, \tag{4}$$

and the smallest and biggest Eigen values are

$$\lambda_{\min} = 1-1/\sqrt{K}, \text{ and } \lambda_{\max} = 1+1/\left(K\sqrt{K}\right).$$

Therefore, we have $\delta_{K+1} = 1/\sqrt{K}$ (In fact, all the Eigen values of $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$ must be contained in the interval $\left[1-\delta_{|\mathcal{S}|}, 1+\delta_{|\mathcal{S}|}\right]$, Thus, $\delta_{K+1} = \max\left\{\lambda_{\max}\left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)-1, 1-\lambda_{\min}\left(\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}\right)\right\}$). Now, we investigate a quantity $\left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right|$ for $i = 1, \cdots, K+1$. For the OMP algorithm to reconstructs $\mathbf{x}$, $\left|\langle \mathbf{a}_{K+1}, \mathbf{y}\rangle\right|$ must be less than any $\left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right|$ for $i = 1, \cdots, K$. This is reason that we investigate the quantities. First, for $i \in \{1, \cdots, K\}$, we have

$$\begin{aligned} \left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right| &\overset{(a)}{=} \left|\langle \mathbf{a}_i, \mathbf{A}\mathbf{x}\rangle\right| \\ &\overset{(b)}{=} \left|\left\langle \mathbf{A}^T \mathbf{a}_i, \mathbf{x}\right\rangle\right| \\ &\overset{(c)}{=} 1 - \frac{K-1}{K\sqrt{K}}, \end{aligned} \tag{5}$$

where (a) from the fact $\mathbf{y} = \mathbf{A}\mathbf{x}$, (b) from the fact $\langle \mathbf{a}_i, \mathbf{A}\mathbf{x}\rangle = \mathbf{a}_i^T \mathbf{A}\mathbf{x} = \mathbf{x}^T \mathbf{A}^T \mathbf{a}_i = \left\langle \mathbf{x}, \mathbf{A}^T \mathbf{a}_i\right\rangle$, and (c) from the fact that $\mathbf{A}^T \mathbf{a}_i$ is the i^{th} column of $\mathbf{A}^T \mathbf{A}$ presented in (4), and $\mathbf{x}$ such as $x_{K+1} = 0$, and $x_i = 1$ for $i = 1, \cdots, K$. Second, for $i = K+1$, we have

$$\begin{aligned} \left|\langle \mathbf{a}_{K+1}, \mathbf{y}\rangle\right| &= \left|\langle \mathbf{a}_{K+1}, \mathbf{A}\mathbf{x}\rangle\right| \\ &= \left|\left\langle \mathbf{A}^T \mathbf{a}_{K+1}, \mathbf{x}\right\rangle\right| \\ &= \frac{1}{\sqrt{K}}. \end{aligned} \tag{6}$$

Obviously, the OMP algorithm must fail in the first iteration if an inequality $\left|\langle \mathbf{a}_{K+1}, \mathbf{y}\rangle\right| \geq \left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right|$ for all $i \in \{1, \cdots, K\}$. The inequity becomes

$$\frac{1}{\sqrt{K}} \geq 1 - \frac{K-1}{K\sqrt{K}}$$

which is always true if $K = 2$. Thus, the OMP algorithm in the first iteration selects an incorrect index.

V. Proof of Theorem 1

A. Notations

The below notations will be used throughout the rest of this presentation. $\mathcal{T} = \text{supp}(\mathbf{x}) := \{i | x_i \neq 0\}$ is the set of indices corresponding to non-zero coefficients of $\mathbf{x}$. $|\mathcal{T}|$ is the cardinality of $\mathcal{T}$, and $\mathcal{T} \setminus \mathcal{I}$ is the set of elements belonging to $\mathcal{T}$ but not to $\mathcal{I}$. $\mathbf{A}_{\mathcal{T}} \in \Re^{M \times |\mathcal{T}|}$ is a sub-matrix of $\mathbf{A}$ which contains columns

corresponding to indices of $\mathcal{T}$. $\mathbf{x}_{\mathcal{T}} \in \Re^{|\mathcal{T}|}$ is a restriction of $\mathbf{x}$ to the elements indexed by $\mathcal{T}$. $span(\mathbf{A}_{\mathcal{T}})$ is the span of columns in $\mathbf{A}_{\mathcal{T}}$, $\mathbf{A}_{\mathcal{T}}^T$ is the transpose of $\mathbf{A}_{\mathcal{T}}$, and $\mathbf{A}_{\mathcal{T}}^\dagger = \left(\mathbf{A}_{\mathcal{T}}^T \mathbf{A}_{\mathcal{T}}\right)^{-1} \mathbf{A}_{\mathcal{T}}^T$ is the pseudo inverse of $\mathbf{A}_{\mathcal{T}}$. $\mathbf{P}_{\mathcal{T}} = \mathbf{A}_{\mathcal{T}} \mathbf{A}_{\mathcal{T}}^\dagger$ is the orthogonal projection onto $span(\mathbf{A}_{\mathcal{T}})$, and $\mathbf{P}_{\mathcal{T}}^\perp = \mathbf{I} - \mathbf{P}_{\mathcal{T}}$ is the orthogonal projection onto the orthogonal complement of $span(\mathbf{A}_{\mathcal{T}})$.

B. Lemmas

We need the below lemmas to prove Theorem 1.

Lemma 1: For a set $\mathcal{I}$, if $\delta_{|\mathcal{I}|} < 1$, then

$$\left(1 - \delta_{|\mathcal{I}|}\right) \|\mathbf{v}\|_2 \le \left\| \mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{I}} \mathbf{v}_{\mathcal{I}} \right\|_2 \le \left(1 + \delta_{|\mathcal{I}|}\right) \|\mathbf{v}\|_2$$

holds for any $\mathbf{v}$ supported on $\mathcal{I}$.

Lemma 2: For disjoint sets $\mathcal{I}, \mathcal{J}$, if $\delta_{|\mathcal{I}|+|\mathcal{J}|} < 1$, then

$$\left\| \mathbf{A}_{\mathcal{I}}^T \mathbf{A} \mathbf{v} \right\|_2 = \left\| \mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{J}} \mathbf{v}_{\mathcal{J}} \right\|_2 \le \delta_{|\mathcal{I}|+|\mathcal{J}|} \|\mathbf{v}\|_2$$

holds for any $\mathbf{v}$ supported on $\mathcal{J}$.

Lemma 3: If the sensing matrix satisfies the RIP of both orders K_1 and K_2, then $\delta_{K_1} \le \delta_{K_2}$ for any $K_1 \le K_2$

All proofs of the above lemmas are given in [3].

C. Proof of Theorem 1

1) We provide a condition under which the OMP algorithm selects a correct index in the first iteration. 2) We show that the residual in the general iteration preservers the sparsity of a K sparse signal. 3) The condition for the first iteration can be extended to the general iteration. 4) Theorem 1 is established from the conditions. The statements are an overall strategy of Proof of Theorem1.

First, we need investigate the condition when the OMP algorithm selects a correct index in the first iteration. Let us denote t^k be the index of the column maximally correlated with the residual $\mathbf{r}^{k-1}$. In the first iteration, we have

$$t^1 = \arg\max_i \left| \left\langle \mathbf{a}_i, \mathbf{r}^0 \right\rangle \right| = \arg\max_i \left| \left\langle \mathbf{a}_i, \mathbf{y} \right\rangle \right|. \quad (7)$$

Now, let us suppose that t^1 always belong to the support set $\mathcal{I}$ of $\mathbf{x}$. From (7), we have

$$\begin{aligned}\left\|\left\langle \mathbf{a}_{t^1}, \mathbf{y}\right\rangle\right\| &= \left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{y}\right\|_\infty \\ &\overset{(a)}{\geq} \frac{1}{\sqrt{K}}\left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{y}\right\|_2 \\ &\overset{(b)}{\geq} \frac{1}{\sqrt{K}}\left(1-\delta_K\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2,\end{aligned} \tag{8}$$

where (a) from the norm inequalities, and (b) from the fact that $y = \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}$ and Lemma 1.Suppose that t^1 does not belong to the support set $\mathcal{I}$, then

$$\begin{aligned}\left\|\left\langle \mathbf{a}_{t^1}, \mathbf{y}\right\rangle\right\| &= \left\|\mathbf{a}_{t^1}^T \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}\right\| \\ &\overset{(a)}{\leq} \left(1-\delta_{K+1}\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2,\end{aligned} \tag{9}$$

where (a) from Lemma 2. Clearly, t^1 must belong to the support set $\mathcal{I}$. Thus, if

$$\frac{1}{\sqrt{K}}\left(1-\delta_K\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2 > \left(1-\delta_{K+1}\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2 \tag{10}$$

then, the OMP algorithm selects a correct index in the first iteration. The equation (10) becomes $\sqrt{K}\delta_{K+1} + \delta_K < 1$. From Lemma 3, the inequality becomes $\sqrt{K}\delta_{K+1} + \delta_{K+1} < 1$ which leads to

$$\delta_{K+1} < \frac{1}{\sqrt{K}+1} \tag{11}$$

In short, if (11) is true, then the OMP algorithm always selects a correct index in the first iteration.

Now, we investigate a condition such that the OMP algorithm selects a correct index in the $(k+1)^{\text{th}}$ iteration. Let us suppose that initial k iterations of the OMP algorithm are successful. Namely, $\mathcal{T}^k = \left\{t^1, \cdots, t^k\right\} \in \mathcal{I}$. Then, $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} \in span\left(\mathbf{A}_{\mathcal{I}}\right)$ because $\mathbf{y} = \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}$ and $\mathbf{A}_{\mathcal{T}^k}$ is a sub-matrix of $\mathbf{A}_{\mathcal{I}}$. Thus, $\mathbf{r}^k$ can be expressed as $\mathbf{r}^k = \mathbf{A}\mathbf{x}^k$ (i.e., $\mathbf{r}^k$ is a linear combination of the K columns of $\mathbf{A}_{\mathcal{I}}$), where the support set of $\mathbf{x}^k$ belongs to the support set of $\mathbf{x}$. If the OMP algorithm selects a correct index belonging to the support set of $\mathbf{x}^k$, then the OMP algorithm also selects a correct index belonging to the support set of $\mathbf{x}$. Clearly, if $\sqrt{K}\delta_{K+1} + \delta_{K+1} < 1$ is satisfied, then the OMP algorithm success in the $(k+1)^{\text{th}}$ iteration.

Last, we need to show that the index t^{k+1} selected at the $(k+1)^{\text{th}}$ iteration of the OMP algorithm does not belong to T^k. First, we have $\hat{\mathbf{x}}_{\mathcal{T}^k} = \mathbf{A}_{\mathcal{T}^k}^{\dagger}\mathbf{y}$, and $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} = \mathbf{P}_{\mathcal{T}^k}^{\perp}\mathbf{y}$. Second, for all $i \in \mathcal{T}^k$, we have

$$\begin{aligned}\left\langle \mathbf{a}_i, \mathbf{r}^k \right\rangle &= \left\langle \mathbf{a}_i, \mathbf{y} - \mathbf{A}_{\mathcal{T}^k} \hat{\mathbf{x}}_{\mathcal{T}^k} \right\rangle \\ &= \left\langle \mathbf{a}_i, \mathbf{y} \right\rangle - \left\{ \mathbf{a}_i, \mathbf{A}_{\mathcal{T}^k} \hat{\mathbf{x}}_{\mathcal{T}^k} \right\} \\ &= \mathbf{a}_i^T A_{\mathcal{I}} x_{\mathcal{I}} - \mathbf{a}_i^T \mathbf{A}_{\mathcal{T}^k} \mathbf{A}_{\mathcal{T}^k}^{\dagger} \mathbf{y} \\ &= 0.\end{aligned}$$

Therefore, we conclude that $\mathbf{r}^k$ is orthogonal to the columns $\mathbf{a}_i$ for all $i \in T^k$. It leads to $t^{k+1} \notin \mathcal{T}^k$. Furthermore, if $\mathbf{r}^k \neq \mathbf{0}$ and $\mathbf{r}^k \in span\left(\mathbf{A}_{\mathcal{I}}\right)$, then there exists $i \in \mathcal{I}$ such as $\left\langle \mathbf{a}_i, \mathbf{r}^k \right\rangle \neq 0$. Therefore, the OMP algorithm selects $i \in \mathcal{I} \setminus \mathcal{T}^k$.

Now, we apply the mathematical induction. First, we proved that the OMP algorithm selects a correct index if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Second, when we assume that the initial k iterations of the OMP algorithm are successful, the OMP algorithm selects a correct index in the $(k+1)^{\text{th}}$ iteration if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Thus, the OMP algorithm will terminate after the K^{th} iteration if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$.

VI. Discussion on Theorem 1

It is hard for us to determine δ_{K+1} from a sensing matrix because we need to examine all possible *K* sparse signal. However, the below result is known

Result [4]: If an $M \times N$ sensing matrix $\mathbf{A}$ whose entries are i.i.d. $\mathcal{N}(0,1/M)$, then $\mathbf{A}$ obeys the RIP condition $\delta_K \leq \varepsilon$ with high probability under

$$M \geq \frac{\rho K \log\left(\frac{N}{K}\right)}{\varepsilon^2} \tag{12}$$

where ρ is a positive constant. When we utilize the above inequalities, we indirectly compare the result obtained by [2].

	A sufficient condition	A sufficient condition on M
[1]	$\delta_{K+1} < 1/\left(3\sqrt{K}\right)$	$M \geq \rho 9K(K+1)\log\left(\frac{N}{K+1}\right)$

The paper	$\delta_{K+1} < 1/\left(\sqrt{K}+1\right)$	$M \geq \rho(K+1)\left(\sqrt{K}+1\right)^2 \log\frac{N}{K+1}$

Appendix

Example 1) computing all the Eigen values of $\begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix}$.

$$\begin{vmatrix} 1-\lambda & b & b \\ b & 1-\lambda & b \\ b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & 1-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda-b & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & b & 1-\lambda+b \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 1-\lambda+2b \end{vmatrix}$$

$$= (1-\lambda-b)^2(1-\lambda+2b)$$

Therefore, $\lambda_1 = \lambda_2 = 1-b$, and $\lambda_3 = 1+2b$.

$$\begin{vmatrix} 1-\lambda & b & b & b \\ b & 1-\lambda & b & b \\ b & b & 1-\lambda & b \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ b & b & 1-\lambda & b \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & b & 1-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & b & b & 1-\lambda+b \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 2b & 1-\lambda+b \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 0 & 1-\lambda+3b \end{vmatrix}$$

$$= (1-\lambda-b)^3(1-\lambda+3b)$$

Therefore, $\lambda_1 = \lambda_2 = \lambda_3 = 1-b$, and $\lambda_4 = 1+3b$

Thus, we concluded all the Eigen values of a $(K+1)\times(K+1)$ $\begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix}$ are

$\lambda_1 = \cdots = \lambda_K = 1-b,$ and $\lambda_{K+1} = 1+Kb$.

Reference

[1] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, pp. 4655 – 4666, Dec. 2007.
[2] M. A. Davenport and M. B. Wakin, “Analysis of Orthogonal Matching Pursuit using the restricted i sometry property,” IEEE Trans. Inf. Theory, vol. 56, pp. 589 – 592, 2008.
[3] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230 – 2249, May 2009.
[4] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” Construct. Approx., vol. 28, no. 3, pp. 253– 263, 2008.

Capacity of OFDM Systems over Fading Underwater Acoustic Channels

Authors: Chantri Polprasert, *Member, IEEE*, James A. Ritcey, *Fellow, IEEE*, and Milica Stojanovic, *Fellow, IEEE*
Publication: IEEE Journal of Oceanic Engineering, vol. 36, no. 4, Oct. 2011
Speaker: Zafar Iqbal

Short Summary:

This paper derives the upper and lower bounds for channel capacity of the OFDM systems over underwater acoustic channels as a function of distance between the transmitter and the receiver. The upper bound is obtained using perfect CSI at the receiver while the lower bound is obtained by assuming that the input is drawn from a PSK constellation which results in non-Gaussian distribution of the output signal and no CSI. It incorporates frequency dependent path loss at each arrival path at the receiver due to acoustic propagation. This leads the UW channel to be modeled as wide sense stationary and correlated scattering (WSS-non-US) fading channel. Results from both Rayleigh and Rician fading show a gap between the upper and lower bounds which depends, not only on the ranges and shape of the scattering function of the UW channel but also on the distance between the transmitter and the receiver.

I. INTRODUCTION

Recently, OFDM has been applied to the UWA communications and yields high data rate with strong bit error rate performance [2-5].

Time and frequency spreading are the main challenges for data transmission through UW channels. Several attempts have been made to characterize the UW channel, most of which view the UW channel as a linear time-varying channel with wide sense stationary and uncorrelated scattering (WSSUS) [8-10]. However, this approach treats the entire frequency band as a whole and neglects the frequency dependent path loss. This model is acceptable for transmissions at low bandwidth (<10 kHz) [9].

Channel capacity over WSSUS fading channel has been studied [12-15] under these assumptions: 1) no CSI is available at the transmitter or receiver, and 2) peak power constraints. It is shown that channel capacity is achieved at capacity maximizing bandwidth, which depends

on the ranges and shape of the scattering function of the fading channel. These studies are conducted over wireless fading channels which assume constant power spectral density (PSD) and AWGN noise. There has been some research on the capacity of UW channels [6, 16, 17] but all assume no fading in their UW channels.

This paper investigates the capacity of OFDM systems over the UW fading channels with no CSI at the transmitter or the receiver. The UW channel is modeled by taking into account frequency-dependent path loss. This invalidates the assumption of stationarity in frequency of the WSSUS model and leads to a frequency-dependent doubly spread (DS) fading channel characterized by the WSS-non-US [18] assumptions. The conventional WSSUS model is uncorrelated in both delay and Doppler domains but the proposed model is uncorrelated in the Doppler domain and correlated in the delay domain.

Using this channel model and assuming that the acoustic propagation and ambient noise PSD are available at both the transmitter and receiver, capacity upper and lower bounds are derived. Capacity upper bound is derived by assuming perfect CSI at the receiver, while lower bound is obtained by the mutual information rate whose input is an i.i.d. random variable and is drawn from a PSK modulation [12,19], which results in a non-Gaussian distribution of the output signal. Results are obtained for both Rayleigh and Rician fading of the UW channel. Simulation results show a gap between the upper and lower bounds which depends not only on the ranges and shape of the scattering function of the UW channel, but also on the distance between the transmitter and receiver. Results are confirmed with the scattering function obtained from the 2008 rescheduled Acoustic Communications Experiment (RACE08) experimental data.

II. OFDM System and UW Channel Model

In this section, an OFDM system model for UW acoustic communications is developed. Physical and statistical properties of the channel as well as PSD of the ambient noise are investigated and a frequency-dependent UW DS fading channel has been proposed.

A. OFDM System

A conventional CP-OFDM system is considered as shown in Fig. 1. Let $\mathbf{X}_n = \left[X_{n,0} ... X_{n,K-1} \right]^T$ and $\mathbf{Y}_n = \left[Y_{n,0} ... Y_{n,K-1} \right]^T$ be the sent and received block of data at the *n*th OFDM symbol duration, respectively.

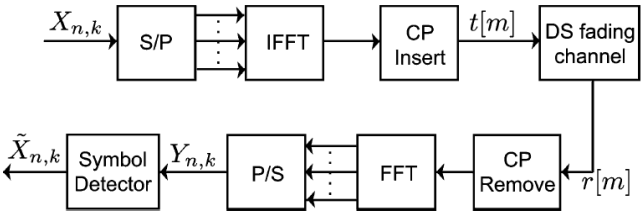


Fig. 1. System model

Assuming the guard interval L_{cp} is longer than the channel length L to avoid the interblock interference (IBI), the input/output relationship can be written as,

$$Y_{n,k} = G_{n,k}(d) X_{n,k} + N_{n,k} \tag{1}$$

where $k \in [0,...,K-1]$ is the subcarrier index and $n \in [0,...,N-1]$, while d is the distance between transmitter and receiver. $G_{n,k}(d)$ denotes the channel transfer function at the kth subcarrier. $N_{n,k}$ is the ambient noise in the ocean. This simplifies the fading effect into multiplicative coefficient, which is the basis for analysis of the UW channel in this paper. The impact of ICI is assumed to be negligible through appropriate parameter settings (Justified in App. I). For simplicity, the overall system input/output of the entire N OFDM transmissions is characterized by a vector of size $NK \times 1$, as follows.

$$\mathbf{Y} = \text{diag}(\mathbf{X})\mathbf{G}(d) + \mathbf{N} = \text{diag}(\mathbf{G}(d))\mathbf{X} + \mathbf{N} \tag{2}$$

where

$$\mathbf{Y} = \left[\mathbf{Y}_0^T ... \mathbf{Y}_{N-1}^T \right]^T \text{ and } \mathbf{Y}_n = \left[Y_{n,0} ... Y_{n,K-1} \right]^T \tag{3}$$

$$\mathbf{X} = \left[\mathbf{X}_0^T ... \mathbf{X}_{N-1}^T \right]^T \text{ and } \mathbf{X}_n = \left[X_{n,0} ... X_{n,K-1} \right]^T \tag{4}$$

$$\mathbf{N} = \left[\mathbf{N}_0^T ... \mathbf{N}_{N-1}^T \right]^T \text{ and } \mathbf{N}_n = \left[N_{n,0} ... N_{n,K-1} \right]^T \tag{5}$$

$$\mathbf{G}(d) = \left[\mathbf{G}_0^T(d) ... \mathbf{G}_{N-1}^T(d) \right]^T \text{ and } \mathbf{G}_n(d) = \left[G_{n,0}(d) ... G_{n,K-1}(d) \right]^T \tag{6}$$

B. Characterization of Approximate DS Fading Channels

UW channel is modeled using both the physical property, which is the attenuation depending on the propagation distance and bandwidth of the transmitted signal, and the statistical property for which the channel is usually assumed WSSUS.

1) Frequency-dependent Path Loss

For the signal propagated through UW medium, the attenuation or path loss, which is a function of distance and signal frequency, is a combination of geometric spreading and absorption, written as,

$$Q^2(d, f) = d^{-sp}(q^2(f))^{-d} \tag{7}$$

where d is the propagated distance in meter and f is the frequency in kilohertz. d^{-sp} represents the spreading loss and sp is the spreading factor which is set to 1.5. $q^2(f)$ is the absorption coefficient in seawater which is given by,

$$10\log(q^2(f)) = 2.49\times10^{-7} f^2 + 0.99\frac{f^2}{f^2+1.23\times10^4} + 1.48\times10^{-4}\frac{f^2}{f^2+1.522} \text{ dB/m} \tag{8}$$

Eq. (8) is calculated when the salinity S is 35 parts per thousand (ppt), gauge pressure P_a is 1 atm, temperature T=14 ^{o}C, and the relaxation frequency is 111 kHz.

2) Conventional Statistical Model

The CIR is modeled by a sum of several multipath components [9], [10]. Let $h(t, \tau)$ denote a continuous-time CIR of linear time-variant (LTV) UW channels and its corresponding transfer function $H(t, f)$ is,

$$h(t,\tau) = \sum_{i=0}^{I-1} h_i(t)\delta(\tau-\tau_i), \quad H(t, f) = \sum_{i=0}^{I-1} h_i(t)e^{-j2\pi f\tau_i} \tag{9}$$

where I is the number of arrival paths. WSSUS is commonly assumed to characterize the channel, i.e., $E\{h[t,\tau]h^*[t',\tau']\} = R_{h_c}(t-t',\tau)\delta(\tau-\tau')$ where $R_{h_c}(t-t',\tau)$ is the autocorrelation function of the delay τ between time t and t'. Its corresponding scattering function is $S_c(\tau, v) = \int R_{h_c}(\Delta t,\tau)\exp(-j2\pi\Delta tv)d\Delta t$ where $\tau \in [0, \tau_m]$. For a bandwidth of less than 10 kHz, let τ_m and f_d denote the maximum channel delay spread and 3-dB Doppler spread of $S_c(\tau,v)$, respectively.

3) Frequency-dependent DS Fading Channels

Conventionally, UW models use WSSUS properties to characterize LTV UW channels, assuming equal attenuation across all the signal bandwidth, treating the entire frequency band as

flat and neglecting frequency-dependent parameters of the individual arrival path. In reality, various factors from channel physics such as the attenuation, reflection loss, or tx/rx operating ranges influence frequency dependency on the path loss. In this paper, the impact of channel physics is limited to only the attenuation $Q^2(d_i, f)$ (7) where d_i is the propagation distance of ith delay path.

Let $\chi_{d_i}(\tau)$ denote a CIR of the ith delay path corresponding to $Q(d_i, f)$ i.e., $Q(d_i, f) = \int \chi_{d_i}(\tau)\exp(-j2\pi f\tau)d\tau$ where $Q^2(d_i, f) = Q(d_i, f)Q^*(d_i, f)$. Taking into account $\chi_{d_i}(\tau)$ yields a modified CIR, $g_d(t, \tau)$

$$g_d(t,\tau) = \sum_{i=0}^{I-1} h_i(t)\chi_{d_i}(\tau) \otimes \delta(\tau - \tau_i) \tag{10}$$

$$\begin{aligned} G_d(t, f) &= \int g_d(t,\tau)e^{-j2\pi\tau f}d\tau \\ &= \sum_{i=0}^{I-1} h_i(t)Q(d_i, f)e^{-j2\pi\tau_i f} \\ &\approx Q(d0, f)\sum_{i=0}^{I-1} h_i(t)e^{-j2\pi\tau_i f} \quad \because Q(d, f) \approx Q(d_{I-1}, f) \\ &= Q(d, f)H(t, f) \end{aligned} \tag{11}$$

d_0 is the distance between transmitter and receiver and the subscript of d is neglected for simplicity. Hence the modified CIR is

$$g_d(t,\tau) = \chi_d(\tau) \otimes \sum_{i=0}^{I-1} h_i(t)\delta(\tau - \tau_i) \tag{12}$$

From the sampling theorem, the T_s-spaced discrete time CIR is,

$$\begin{aligned} g_{0,d}[m, p_l] &= \int_{\tau} g_d(mT_s, \tau)\,\mathrm{sinc}(B\tau - p_l)d\tau \\ &\approx g_d(mT_s, p_lT_s) \quad \because B \text{ is large} \\ &= \chi_d[l] \otimes h_0[m, p_l] \end{aligned} \tag{13}$$

where B=1/T_s. From (13) the channel transfer function can be written as

$$
\begin{aligned}
G_{n,k}(d) &= \sum_{m'=0}^{K-1}\left(\frac{1}{K}\sum_{l=0}^{L-1} g_{0,d}^{n}\left[m'+L_{cp},l\right]\sum_{p=0}^{K-1} e^{j2\pi p(m'-l)/K}\right)\times e^{-j2\pi m'k/K} \\
&= \sum_{p=0}^{K-1}\sum_{l=0}^{L-1}\left(\frac{1}{K}\sum_{m'=0}^{K-1}\left(\chi_d[l]\otimes h_0^n\left[m'+L_{cp},l\right]\right)\times e^{j2\pi m'(p-k)/K}\right)e^{-j2\pi pl/K} \\
&\approx \sum_{p=0}^{K-1}\sum_{l=0}^{L-1} G_{n,l}^{d}[0]e^{-j2\pi kl/K} \\
&= Q(d,f_k)H_{n,k}
\end{aligned}
\tag{14}
$$

where $f_k = f_c + k/(KT_s)$ and f_c is the center frequency corresponding to the zeroth subcarrier. $g_{0,d}^{n}\left[m',l\right] = g_{0,d}\left[nN_s + p_0 + m', p_l\right]$ and $h^n[m',l] = h_0\left[nN_s + p_0 + m', p_l\right]$ where $N_s = L_{cp} + K$ is the OFDM symbol length and p_0 is the arrival time of the first arrival path. Eq. (14) is derived under negligible ICI. Moreover,

$$
\begin{aligned}
H_{n,k} &= \sum_{l=0}^{L-1}\left(\frac{1}{K}\sum_{m'=0}^{K-1} h_0^n\left[m'+L_{cp},l\right]\right)e^{-j2\pi lk/K} \\
&= \sum_{l=0}^{L-1} h_{n,l}[0]e^{-j2\pi lk/K}
\end{aligned}
\tag{15}
$$

$$
Q(d,f_k) = \sum_{l=0}^{L-1}\chi_d[l]e^{-j2\pi lk/K} \tag{16}
$$

and

$$
G_{n,l}^{d}[p-k] = \frac{1}{K}\sum_{m'=0}^{K-1}\left(\chi_d[l]\otimes h_0^n\left[m'+L_{cp},l\right]\right)e^{j2\pi m'(p-k)/K} \tag{17}
$$

$G_{n,k}(d)$ is the fading gain encountered by the signal transmitted on the kth subcarrier. $Q(d, f_k)$ is assumed constant within a subcarrier with center frequency f_k. $H_{n,k}[0]$ is the approximate CIR. Eq. (14) simplifies the transfer function of frequency-dependent UW DS channel into a multiplication of the attenuation $Q(d,f_k)$ and statistical part $H_{n,k}$ governed by the scattering function $S[l, \lambda]$. Assuming $S[l,\lambda] \approx S_c\left(lT_s, \lambda/T_b\right)$ when the variation of $h_0[m, p_l]$ within $T_b(T_b=N_sT_s)$ is negligible [32]. T_b is the OFDM symbol interval $\lambda\in$[-0.5,0.5]. Its range (L, λ_d) is related to (τ,f_d) of $S_c(\tau,v)$ through $L = \lceil \tau_m B \rceil$ and $\lambda_d=f_dT_b$. This leads $G_{n,k}(d)$ to be a WSS but non-US fading channel [18].

$$
E\left\{G_{n',k'}(d)G_{n,k}^{*}(d)\right\} = Q(d,f_k)Q^{*}(d,f_{k'})R_H\left[n-n',k-k'\right] \tag{18}
$$

where $R_H\left[n-n',k-k'\right] = E\left[H_{n,k}H_{n,k}^{*}\right]$. Compared to the conventional WSSUS model (uncorrelated in both delay and Doppler domains), the proposed model is still uncorrelated in

Doppler but correlated in the delay domain because of attenuation. To be consistent, a vector form for $G_{n,k}(d)$ from (14) is

$$\mathbf{G}_n(d) = \mathbf{Q}(d)\mathbf{H}_n \tag{19}$$

where $\mathbf{Q}(d) = \mathrm{diag}\left(\left[Q(d, f_0)...Q(d, f_{K-1})\right]\right)$ and $\mathbf{H}_n = \left[H_{n,0}...H_{n,K-1}\right]^T$ from (15). Fig 2 shows a realization of $\left|G_{n,k}(d)\right|^2$ when d=5 and 20 km. $H_{n,k}$ is assumed zero-mean complex Gaussian random variable with exponentially decaying PDP with 20-dB power difference between the first and last paths. Transmit bandwidth is 51.2 kHz. Channel delay length is 5 ms which corresponds to L=256. The number of subcarriers K is 512. We can see that the propagation distance and signal frequency have a significant impact on the realization of $\left|G_{n,k}(d)\right|^2$.

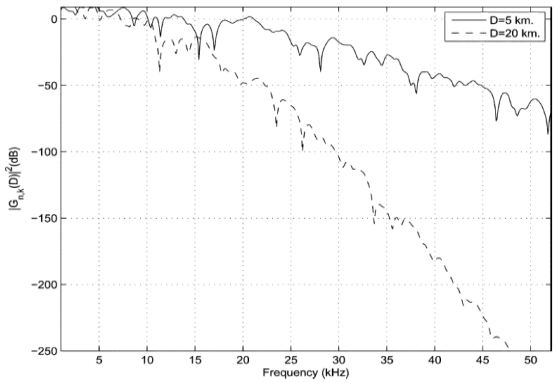


Fig. 2. Impact of attenuation on CIR

C. Ambient Noise

$N_{n,k}$ in (1) is assumed the ambient noise in the ocean which consists of four sources [6]: turbulence $A_t(f)$, shipping $A_s(f)$, waves $A_w(f)$, and thermal noise $A_{th}(f)$, described by Gaussian statistics with a continuous PSD in dBre/μPa per hertz,

$$\begin{aligned} A_t(f) &= 17 - 30\log f \\ A_s(f) &= 40 + 20(s - 0.5) + 26\log f - 60\log(f + 0.03) \\ A_w(f) &= 50 + 7.5\sqrt{w} + 20\log f - 40\log(f + 0.4) \\ A_{th}(f) &= -15 + 20\log f \end{aligned} \tag{20}$$

where f is the frequency in kilohertz, $s \in [0,1]$ is the shipping activity, w is the wind speed in meters per second, and overall noise PSD is

$$A(f) = 10\log\left(10^{A_t(f)/10} + 10^{A_s(f)/10} + 10^{A_w(f)/10} + 10^{A_{th}(f)/10}\right) \tag{21}$$

III. CAPACITY OF THE UW CHANNELS

The upper bound $U_c(d)$ and lower bound $L_c(d)$ are derived as a function of distance d between the transmitter and receiver. The capacity $C(d)$ is given in bits per second by

$$C(d) = \lim_{N\to\infty} \frac{1}{NT_b} \sup_{p(\mathbf{X})} I(\mathbf{Y};\mathbf{X}) \tag{22}$$

where the maximization is over the set $p(\mathbf{X})$ of all input distributions that satisfy a given average-power constraint. $U_c(d)$ is obtained when the input vector follows a join complex Gaussian distribution. $L_c(d)$ is obtained under imperfect CSI whose reduction from $U_c(d)$ comes from limited mutual information from PSK constellation and the MMSE prediction error related to channel uncertainty [12], [14]. This bounding technique is used in [12] for wireless fading channels while this paper uses it for UW channels. The bounds are derived under the following assumptions:

- Information of attenuation (7) and ambient noise PSD (21) of UW channels are available at both the transmitter and receiver.
- For statistical part [$H_{n,k}$ (15)] of UW channels, its approximate CIR $h_{n,l}[0]$ is assumed a WSSUS random process with variance σ_l^2 where $\sum_l \sigma_l^2 = E\left[\left|H_{n,k}\right|^2\right] = 1$. Rayleigh and Rician fading are also considered. A scattering function which characterizes $H_{n,k}$ is available at the receiver.
- The noise vector $\mathbf{N} \sim \mathcal{CN}\left(\mathbf{0}, diag(\mathbf{A})\right)$. Where $\mathbf{A} = \left[\mathbf{A}_0 ... \mathbf{A}_{N-1}\right]^T$ and $\mathbf{A}_n = \left[A(f_0), ..., A(f_{K-1})\right]^T$.
- The impact of ICI is negligible compared to $A(f_k)$.

Let F denote the subcarrier spacing and $B=KF$, the signal bandwidth. P is the signal transmit power in dBre/μPa.

A. Upper Bound $U_c(d)$

To bound $\sup_{p(\mathbf{X})} I(\mathbf{Y};\mathbf{X})$, we use the chain rule $I(\mathbf{Y};\mathbf{X}) = I(\mathbf{Y};\mathbf{X},\mathbf{G}(d)) - I(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X})$. The output vector $\mathbf{Y}$ depends on the input vector $\mathbf{X}$ through $\mathbf{b}$=daig($\mathbf{X}$)$\mathbf{G}(d)$, so $I(\mathbf{Y};\mathbf{X},\mathbf{G}(d))=I(\mathbf{Y};\mathbf{b})$. The upper bound of I(Y;b) is achieved when the input $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I} + \mathbf{R}_\mathbf{X}(d) \odot \mathbf{R}_\mathbf{G}(d))$. Where

$\mathbf{R}_{\mathbf{G}}(d) = E[\mathbf{G}(d)\mathbf{G}^H(d)]$. $\mathbf{R}_{\mathbf{X}}(d) = \text{diag}\left(\left[\mathbf{R}_{\mathbf{X}_0}(d),...,\mathbf{R}_{\mathbf{X}_{N-1}}(d)\right]\right)$ where

$\mathbf{R}_{\mathbf{X}_n}(d) = \text{diag}\left(\left[\sigma_x^2(d, f_0)...\sigma_x^2(d, f_{K-1})\right]\right)$ and $\sigma_x^2(d, f_k) = E\left[\left|X_{n,k}\right|^2\right]$.

The upper bound $U_c(d)$ is [27],

$$\begin{aligned} C(d) &\le \lim_{N\to\infty} \frac{1}{NT_b} \sup_{\mathbf{R}_{\mathbf{X}}(d)} \log\det\left(\mathbf{I} + \left(\mathbf{R}_{\mathbf{X}}(d) \odot \mathbf{R}_{\mathbf{G}}(d)\right)\text{diag}(\mathbf{A})^{-1}\right) \\ &\le \lim_{N\to\infty} \frac{1}{NT_b} \sup_{\mathbf{R}_{\mathbf{X}}(d)} \sum_{n=0}^{N-1}\sum_{k=0}^{K-1} \log\left(1 + E\left[\left|X_{n,k}\right|^2\right]\frac{Q^2(d, f_k)}{A(f_k)}\right) \\ &= \frac{1}{T_b}\sum_{k=0}^{K-1} \sup_{\sigma_x^2(d,f_k)} \log\left(1 + \sigma_x^2(d, f_k)\frac{Q^2(d, f_k)}{A(f_k)}\right) \\ &= U_c(d) \end{aligned} \tag{23}$$

where the inequality follows from Hadamard's inequality [11]. This result is similar to [6] which is the capacity of time-invariant UW channels but is scaled by a factor of FT_b which is greater than 1 to avoid IBI. $\sigma_x^2(d, f_k)$ is subject to the source power constraint

$$F\sum_{k=0}^{K-1}\sigma_x^2(d, f_k) = P \tag{24}$$

U_c is obtained when energy allocation across all subcarriers satisfies

$$\sigma_x^2(d, f_k) = \begin{cases} \max\left(Th - \left(\frac{A(f_k)}{Q(d,f_k)^2}\right), 0\right) & f_k \in B \\ 0 & otherwise \end{cases} \tag{25}$$

where Th is chosen so that (24) is satisfied according to the water-filling algorithm [11].

B. Lower Bound $L_c(d)$ over Rayleigh Fading Channels

For lower bound, channel fading statistics are assumed available at the receiver, not the transmitter. Our results show, for the first time, that decrease in $L_c(d)$ depends not only on the channel variations but also on the propagation distance d between the transmitter and receiver.

Consider $I(\mathbf{Y};\mathbf{X})$ where each entry of $\mathbf{X}$, $X_{n,k}$ is an i.i.d r.v. drawn from PSK modulation whose amplitude $\left|X_{n,k}\right| = \sigma_x$ and phase $\angle X_{n,k}$ has a uniform discrete distribution across a circle. $I(\mathbf{Y};\mathbf{X})$ can be written as,

$$\begin{aligned} I(\mathbf{Y};\mathbf{X}) &= I\left(\mathbf{Y};\mathbf{X},\mathbf{G}(d)\right) - I\left(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X}\right) \\ &\ge I\left(\mathbf{Y};\mathbf{X}\,|\,\mathbf{G}(d)\right) - I\left(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X}\right) \end{aligned} \tag{26}$$

The inequality is due to the non-negativity property of mutual information. Exact calculation of the mutual information is infeasible due the non-Gaussian distribution of **Y** [28]. Note that [29]

$$I(\mathbf{Y};\mathbf{X}\,|\,\mathbf{G}(d)) = N\ I(\mathbf{Y}_N;\mathbf{X}_N\,|\,\mathbf{G}_N(d)) \tag{27}$$

where $\mathbf{G}_N(d)=\mathbf{G}_n(d)$, $\mathbf{X}_N=\mathbf{X}_n$, and $\mathbf{Y}_N=\mathbf{Y}_n$ since the input $X_{n,k}$ has an i.i.d. distribution and every block of the channel coefficients $\mathbf{G}_n(d)$ has the same distribution. $\sigma_x^2(d, f_k)$ is set according to (25) under constraint (24) and apply it to $I(\mathbf{Y}_N;\mathbf{X}_N\,|\,\mathbf{G}_N(d))$. This water-filling policy is suboptimal for PSK constellation [30]. $I(\mathbf{Y};\mathbf{G}(d)|\mathbf{X})$ is calculated in App. II which yields

$$I(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X}) = \sum_{n=0}^{N-1}\log\det\left(\mathbf{I}+\mathbf{B}_n(d)\,\mathrm{diag}\left(\mathbf{S}(d)\right)\right) \tag{28}$$

where $\mathbf{S}(d)$ is the $K\times 1$ vector whose kth entry is $\sigma_x^2(d, f_k)Q^2(d, f_k)/A(f_k)$. $\mathbf{B}_n(d)$ is the linear MMSE prediction error matrix which depends on both the transmission distance d and channel variation $R_H[m,k]$. Substituting (27) and (28) into (26), the mutual information is

$$I(\mathbf{X};\mathbf{Y}) \ge N\ I\left(\mathbf{Y}_N;\mathbf{X}_N\,|\,\mathbf{G}_N(d)\right) - \sum_{n=0}^{N-1}\log\det\left(\mathbf{I}+\mathbf{B}_n(d)\,\mathrm{diag}\left(\mathbf{S}(d)\right)\right) \tag{29}$$

Finally, the lower bound $L_c(d)$ of the capacity $C(d)$ can be written as,

$$\begin{aligned} C(d) &\ge \lim_{N\to\infty}\frac{1}{NT_b}I(\mathbf{Y};\mathbf{X}) \\ &\ge \frac{1}{T_b}I\left(\mathbf{Y}_\infty;\mathbf{X}_\infty\,|\,\mathbf{G}_\infty(d)\right) - \frac{1}{T_b}\log\det\left(\mathbf{I}+\mathbf{B}_\infty(d)\,\mathrm{diag}\left(\mathbf{S}(d)\right)\right) \\ &= L_c(d) \end{aligned} \tag{30}$$

where $\mathbf{B}_\infty(d)$ is calculated given infinite past channel symbols. From (30), unlike [12] and [19], channel scattering function is not explicit but lies within $\mathbf{B}_\infty(d)$.

C. Lower Bound $L_c(d)$ over Rician Fading Channels

Let ρ denote a Rician fading parameter which is the ratio of the fixed to a scatter part. ρ is assumed independent of the transmission distance d and identical for every delay path. The approximate CIR $h_{n,l}[0]$ of the lth path is modeled as

$$h_{n,l}[0] = \sigma_l\left(A_l e^{j\phi_l} + s_{n,l}\right) \tag{31}$$

$$\text{where } \rho = \frac{A_l^2}{E\left[\left|s_{n,l}\right|^2\right]} \qquad \text{and } A_l^2 = \frac{\rho}{\rho+1}, E\left[\left|s_{n,l}\right|^2\right] = \frac{1}{\rho+1} \tag{32}$$

$E\left[\left|h_{n,l}[0]\right|^2\right] = \sigma_l^2$. ϕ_l is assumed uniformly distributed from –π to π and uncorrelated across different delay paths. For $h(\mathbf{Y})$, using (31), $H_{n,k}$ is

$$H_{n,k} = \sum_{l=0}^{L-1} \sigma_l A_l e^{j\phi_l} e^{-j2\pi lk/K} + \sum_{l=0}^{L-1} \sigma_l s_{n,l} e^{-j2\pi lk/K} \tag{33}$$

From (33), sum of scatter part follows $\mathcal{CN}\left(0,1/(\rho+1)\right)$. This causes $H_{n,k} \sim \mathcal{CN}\left(D_k,1/(\rho+1)\right)$ where $Dk = \sum_{l=0}^{L-1} \sigma_l A_l e^{j\phi_l} e^{-j2\pi lk/K}$.

For $h(\mathbf{Y}|\mathbf{X})$, we assume that the receiver can successfully track the fixed part $A_l e^{j\phi_l}$ and the autocorrelation function of the approximate CIR is

$$E\left[h_{n,l}[0]h_{n',l'}^*[0]\right] = \left(A_l^2 + R_s\left[n-n',l\right]\right)\sigma_l^2\delta\left[l-l'\right] \tag{34}$$

Where $R_s\left[n-n',l\right] = E\left[s_{n,l}s_{n',l}^*\right]$. Apply (34) to calculate $\mathbf{B}_n(d)$ and obtain $h(\mathbf{Y}|\mathbf{X})$.

IV. Simulation results

The UW fading channel is modeled by two parts, attenuation and statistical as explained earlier. The delay profile is assumed exponentially decaying whose maximum delay spread τ_m is set where the first and last arrival paths have 10-dB power difference. The range of the Doppler profile scattering function is determined by f_d, the 3-dB bandwidth of the frequency response. For $A(f_k)$, the shipping activity s = 0.5 and wind speed w = 10 m/s. OFDM symbols are transmitter at frequency beyond 1 KHz. Energy allocation across transmit bandwidth $B_c(d)$ is implemented using (25) subject to power constraint (24). P=145 dBre/μPa and Rayleigh fading is assumed unless stated otherwise.

1) Limitations due to the ICI

Because of the attenuation, the variance of ICI is frequency dependent. This model assumes the ICI variance is negligible compared to that of the ambient noise. In simulation, the ICI variance is limited to at least 3 dB lower than ambient noise variance. The ICI variance depends

on attenuation, $\sigma_x^2(f_k)$, and shape of the scattering function. Two scattering functions, AR-1 and uniform scattering are considered whose 3 dB bandwidth is equal to λ_d. Let $S_1[l, \lambda]$ and $S_2[l, \lambda]$ denote these scattering functions of $h_{n,l}[0]$, respectively, given by

$$S_1[l,\lambda] = \frac{\sigma_l^2}{\left|1-\alpha_l e^{-j2\pi\lambda}\right|^2},\ \lambda \in [-0.5, 0.5]$$
$$S_2[l,\lambda] = \begin{cases} \frac{\sigma_l^2}{2\lambda_d}, & |\lambda| \le \lambda_d \\ 0, & \lambda_d < |\lambda| \le 0.5 \end{cases} \tag{35}$$

These scattering functions are assumed unchanged over the transmission ranges of interest. Fig. 3 displays variance of the ICI at their widest spread of both scattering functions when d=5 km such that its variance is at least 3 dB lower than that of the ambient noise for most of the transmission bandwidth. For the AR-1 model, τ_m =1 ms and f_d =1 Hz. For the uniform model, τ_m=5 ms and f_d =7 Hz. We notice that the 3-dB gap is violated when signal bandwidth is greater than 31 kHz. These account for only 0.39% of the total signal energy and have negligible impact on the capacity as justified in Appendix I.

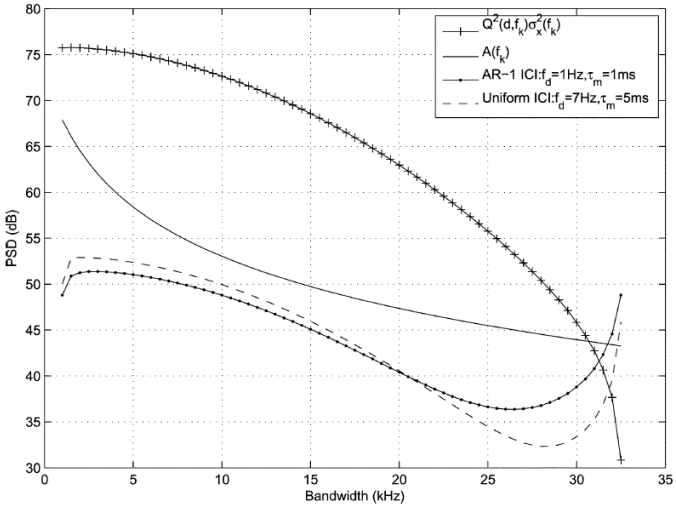


Fig. 3. PSD of the received signal, ambient noise, and the ICI variance at d=5 km

2) Impact of Signal Bandwidth

From Fig. 4, we can see that both $U_c(d)$ and $L_c(d)$ increase as a function of signal bandwidth B and remain fixed when B is greater than a certain value. We define this value as the capacity-maximizing bandwidth $B_c(d)$ which is a signal bandwidth that maximizes both $L_c(d)$ and $U_c(d)$. The gap beyond $B_c(d)$ is rather wide due to the limited mutual information that can be conveyed by the PSK constellation.

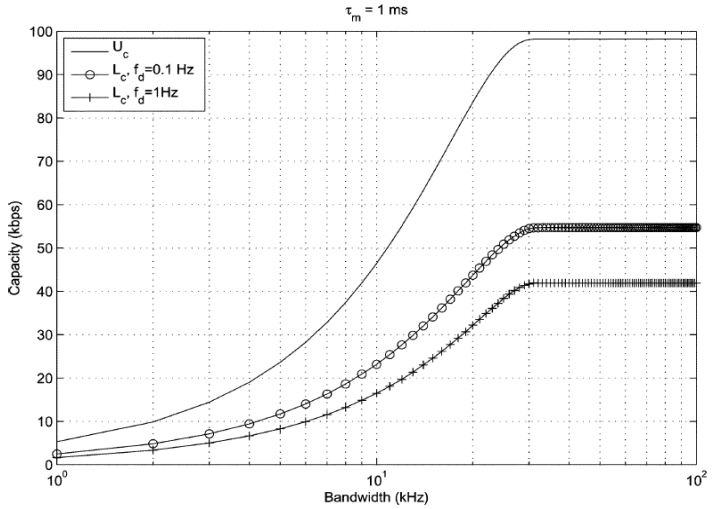


Fig. 4. $U_c(d)$ and $L_c(d)$ versus bandwidth for AR-1 model at d=5km

3) Impact of Ranges and Shape of the Scattering Function

Figs. 5 and 6 show the impact of the ranges of $(f_{d,}\tau_m)$ on $L_c(d)$ over the distance for $S_1[l, \lambda]$ and $S_2[l, \lambda]$, respectively.

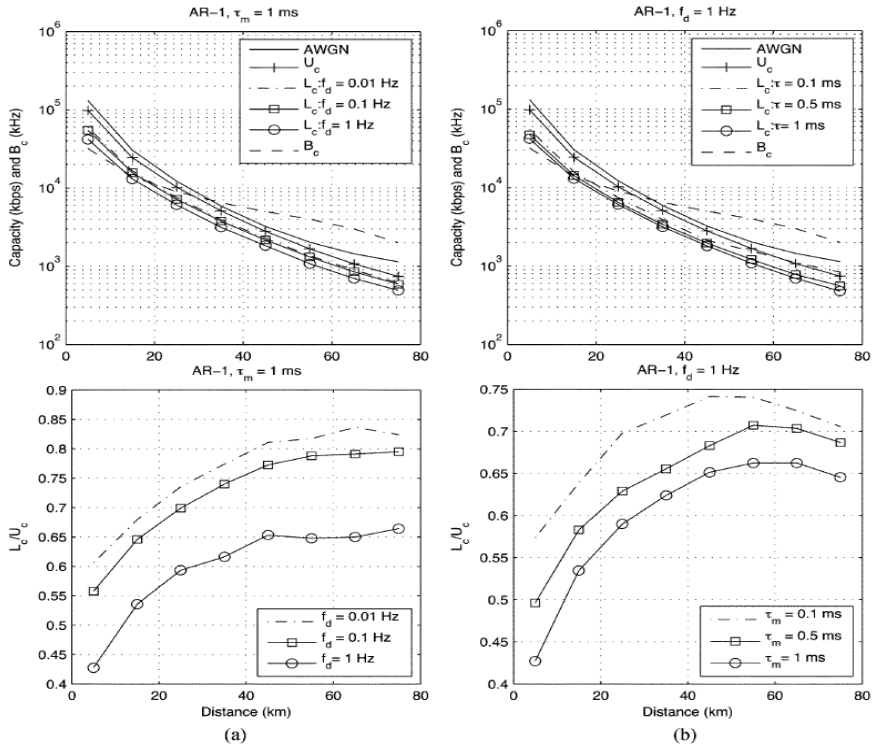


Fig. 5. Impact of (a) Doppler spread and (b) delay spread on $L_c(d)$ for AR-1 scattering function

As expected, the ratio between $L_c(d)$ and $U_c(d)$ increases as either f_d or τ_m increases. This is due to the higher prediction error influenced by stronger channel variations.

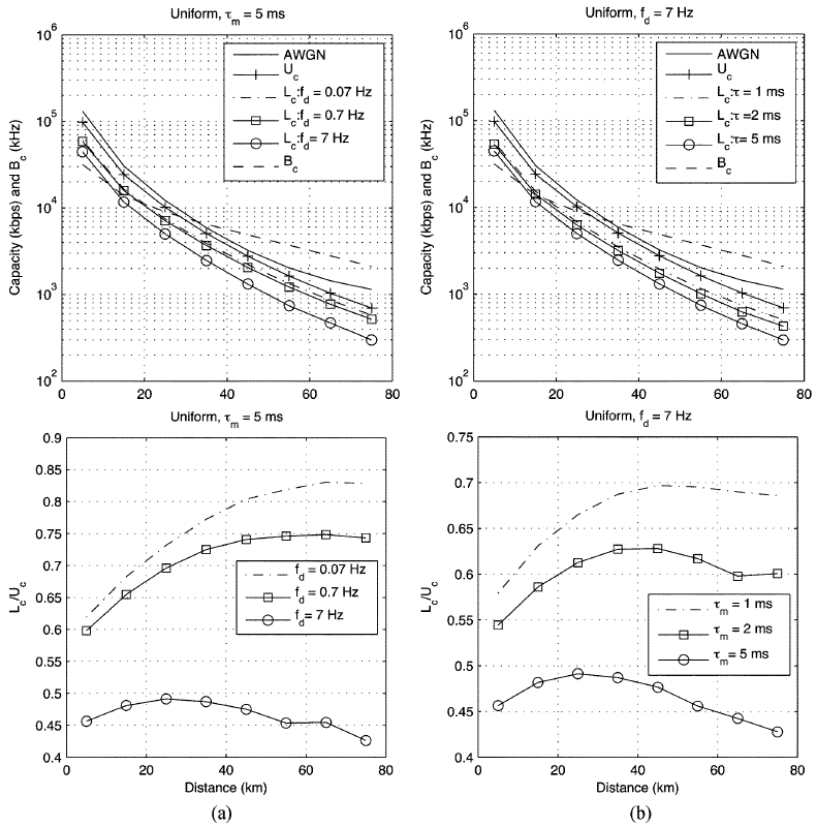


Fig. 6. Impact of (a) Doppler spread and (b) delay spread on $L_c(d)$ for uniform scattering

The impact of the shape of the scattering function is compared in Fig. 7 when f_d=1 Hz and τ_m=1 ms. We set F=500 kHz and T_b=15 ms. From the figure, $L_c(d)$ from $S_1[l, \lambda]$ is lower than that of $S_2[l, \lambda]$ as shown in Fig. 7(a). Fig. 7(b) shows the ratio of $L_c(d)$/ $U_c(d)$.

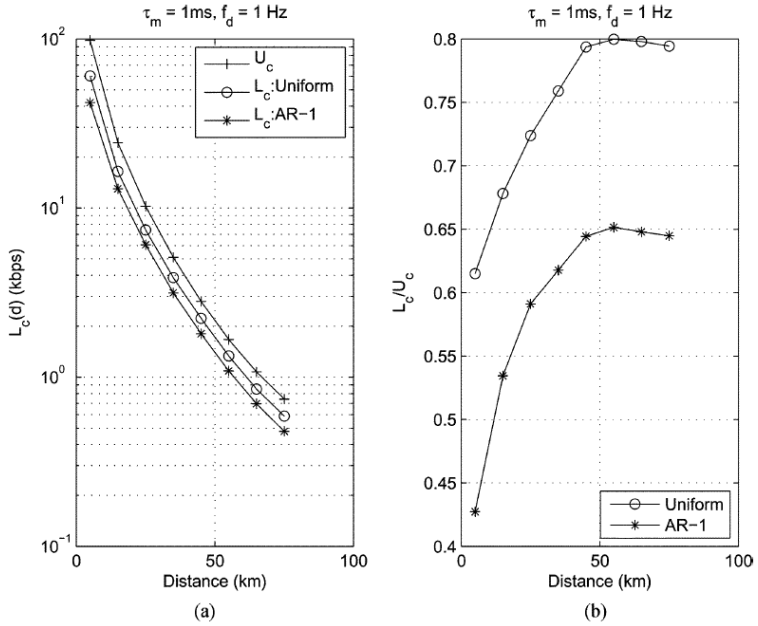


Fig. 7. Impact of the shape of scattering function on (a) $L_c(d)$ and (b) $L_c(d)$/ $U_c(d)$

4) Impact of Over Rician Fading Channels

The Ricean fading parameter ρ is set to -5, 0, 5, and 10 dB, identical for every path and independent of the distance. The Doppler spread profile of the scatter part (34) is assumed uniformly distributed. The fixed part is perfectly known at the receiver. From Fig. 8, the gap between the upper and lower bounds decreases as ρ increases which is due to the reduced power in the scatter part of the channel.

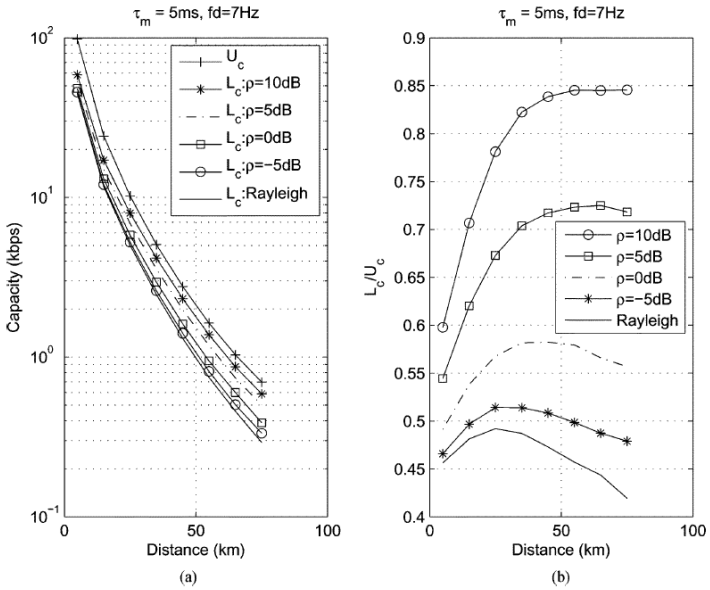


Fig. 8. $U_c(d)$ and $L_c(d)$ to the channel capacity over (a) Rician fading channel, (b) $L_c(d)/ U_c(d)$. Uniform Doppler spread profile

5) Impact of the Transmission Distance

From Figs. 5 and 6, both $L_c(d)$ and $U_c(d)$ decrease at longer distance owing to strong channel attenuation which determines $B_c(d)$. The gap at a short transmission distance is due to the energy wasted because of the PSK constellation while the gap at a very long distance is due to the higher prediction error because of the stronger attenuation.

6) Impact of Transmit Power

Fig. 9 shows the impact of transmit power on $L_c(d)$ and $B_c(d)$ for AR-1 scattering. A significant decrease in $L_c(d)$ and $B_c(d)$ occurs especially at long distance. This shows that for data transmission at low power, a short distance or multiple short hops across the transducers are preferred to one long transmission.

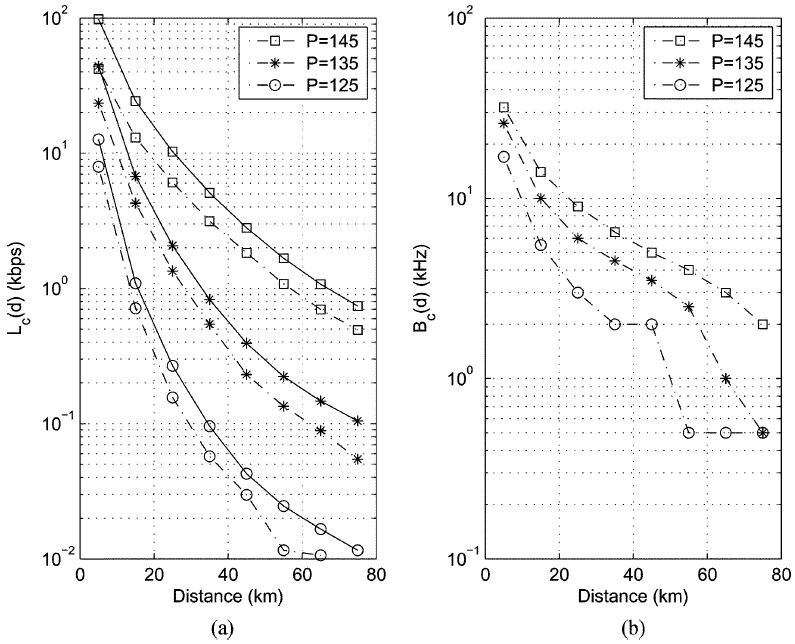


Fig. 9. Impact of transmit power on (a) $L_c(d)$ and (b) $B_c(d)$ for AR-1 scattering

V. EXPERIMENTAL DATA

The capacity of OFDM systems is investigated using the scattering function from real UW environments measured from the RACE08 experiment. Data is selected from the receiving arrays which are 1000 m from the transducer. The array is a 12-element vertical array with 12-cm spacing between elements. 8-PSK signals are upsampled by a factor of ten and filtered by a square root raised-cosine filter with a rolloff factor 0.25. A block of data which contain 64 data symbols are transmitted every 28.7 ms. A guard period is inserted between blocks to avoid the IBI. The bandwidth is 4.8 kHz at 12-kHz carrier frequency. Fig. 10(a) shows a contour plot of the estimates of the scattering function and Fig. 10(b) shows their corresponding PDP of process I–IV obtained from four different measurement periods.

Fig. 11(a) shows $L_c(d)$ and $U_c(d)$ from process I–IV over a range of the distance. Their corresponding $L_c(d)$ / $U_c(d)$ are displayed in Fig. 11(b). From the results, process II yields the best performance while process IV yields the worst. This is due to high Doppler spread at the dominant arrival paths in process IV while process II experiences smallest Doppler spread for almost every arrival path as shown in Fig. 10. Processes I and III exhibit similar results although process III is slightly worse since more dominant paths experience stronger Doppler spread compared to process I.

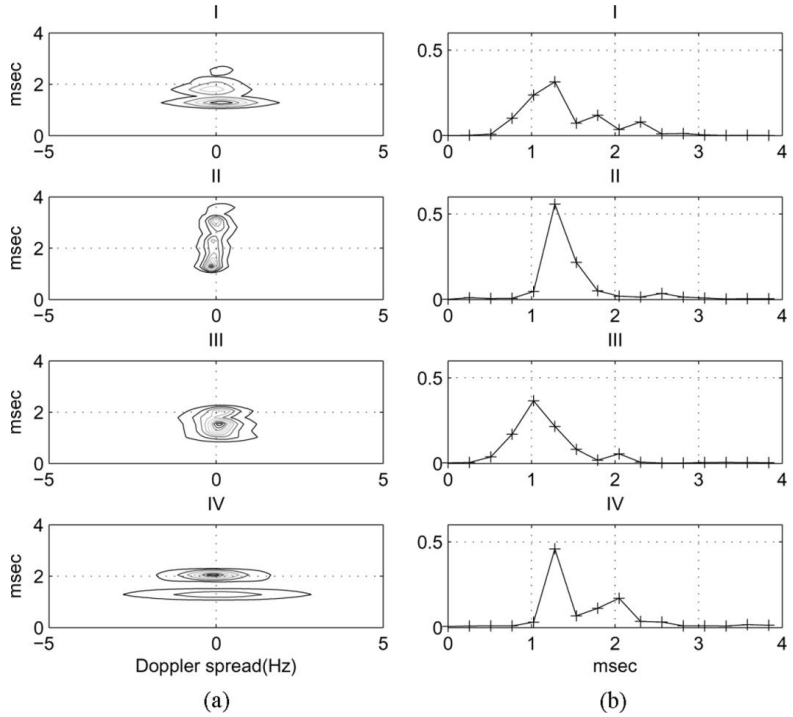


Fig. 10. (a) Scattering function estimates and (b) corresponding normalized PDP

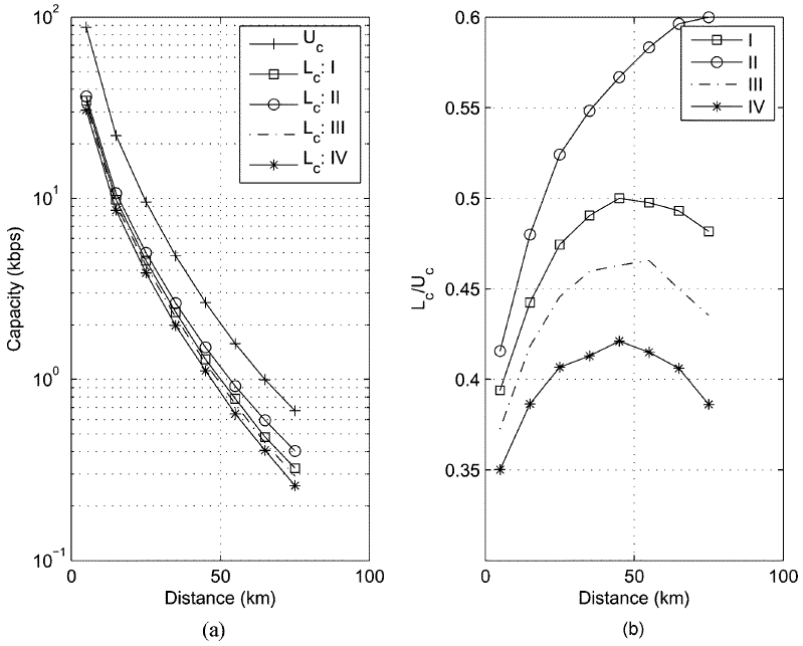


Fig. 11. (a) $L_c(d)$ $U_c(d)$ and (b) corresponding $L_c(d)/ U_c(d)$ over experimental UW fading channels

VI. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

Appendix

A. ICI Justification

To investigate the ICI impact, a simulation is run assuming that the ICI behaves as an independent complex Gaussian r.v. Therefore, the total noise accumulated in the simulation is the ICI plus the ambient noise. From (1), by including the ICI, the received signal can be written as

$$\begin{aligned} Y_{n,k} &= G_{n,k}(d)X_{n,k} + C_{n,k} + N_{n,k} \\ &= G_{n,k}(d)X_{n,k} + Z_{n,k} \end{aligned} \tag{36}$$

where $Z_{n,k}$ is the complex Gaussian noise consisting of the ambient and ICI noise whose variance is $E\left[\left|C_{n,k}\right|^2\right] + A(f_k)$. Using this assumption, Fig. 12 shows the $U_c(d)$ and $L_c(d)$ bounds at 5 km distance between transmitter and receiver. This distance gives highest ICI variance since longer distance means higher attenuation resulting in lower ICI.

In conclusion, it is shown that by taking into account the ICI as an additive complex Gaussian noise, $U_c(d)$ is reduced by at most 5.89% while $L_c(d)$ is reduced by at most 3.03%. This reduction is quite small and has little impact on the overall performance, and justifies our ICI setting.

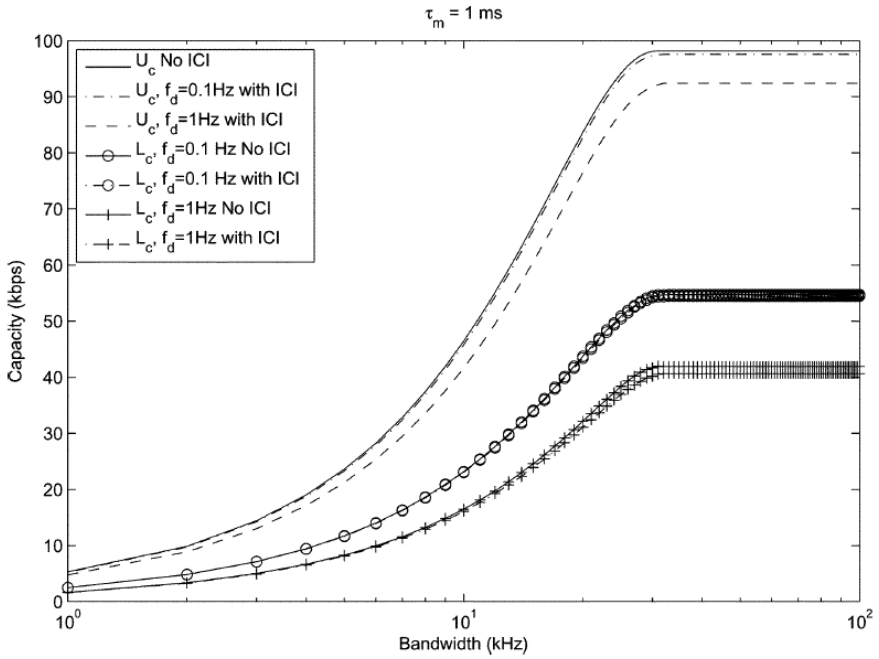


Fig. 12. Impact of ICI on $L_c(d)$ and $U_c(d)$ for AR-1 scattering function

B. I(Y;G(d)|X) Derivation

To calculate $I(\mathbf{Y};\mathbf{G}(d)|\mathbf{X})$, use the chain rule of differential entropy [11],

$$\begin{aligned} I\left(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X}\right) &= h\left(\mathbf{Y}\,|\,\mathbf{X}\right) - h\left(\mathbf{Y}\,|\,\mathbf{G}(d),\mathbf{X}\right) \\ &= h\left(\mathbf{Y}_0,\mathbf{Y}_1,...,\mathbf{Y}_{N-1}\,|\,\mathbf{X}\right) - N\sum_{k=0}^{K-1}\log\left(\pi eA(f_k)\right) \\ &= \sum_{n=0}^{N-1} h\left(\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right) - N\sum_{k=0}^{K-1}\log\left(\pi eA(f_k)\right) \end{aligned} \tag{37}$$

where

$$h\left(\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right) \triangleq \log\left(\left(\pi e\right)^K \det\left(\mathrm{Cov}\left[\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right]\right)\right) \tag{38}$$

To calculate $\mathrm{Cov}\left[\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right]$, we begin with mean

$$\begin{aligned} E\left[\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right] &= \mathrm{diag}(\mathbf{X}_n)E\left[\mathbf{G}_n(d)\,|\,\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right] \\ &= \left(\mathrm{diag}\left(\mathbf{X}_n\right)\mathbf{Q}(d)\right)\hat{\mathbf{H}}_n(d) \end{aligned} \tag{39}$$

Where (39) is obtained using (2) and (19). $\hat{\mathbf{H}}_n(d) = E\left[\mathbf{H}_n(d)\,|\,\mathbf{Y}_0,...,\mathbf{Y}_{n-1},\mathbf{X}\right]$ is the MMSE channel estimate given the current and past detected symbols and can be written as the 1-step output of the linear $K\times K$ MIMO predictor filter of length J

$$\hat{\mathbf{H}}_n(d) = \sum_{j=1}^{J} E_j(d)\tilde{\mathbf{H}}_{n-j}(d) \tag{40}$$

$\tilde{\mathbf{H}}_n(d) = \left[\tilde{H}_{n,0}(d)...\tilde{H}_{n,K-1}(d)\right]^T$ and $\mathbf{E}_j(d)$ is the predictor coefficient of size $K\times K$. With (1), the observation $\tilde{H}_{n,K-1}(d)$ is obtained by

$$\tilde{H}_{n,k}(d) = \frac{Y_{n,k}}{X_{n,k}Q(d,f_k)} = H_{n,k} + \frac{N_{n,k}}{X_{n,k}Q(d,f_k)} \tag{41}$$

Then from (39), $\mathrm{Cov}\left[\mathbf{Y}_n\,|\,\mathbf{Y}_0...\mathbf{Y}_{n-1},\mathbf{X}\right]$ is

$$\begin{aligned} &\mathrm{Cov}\left[\mathbf{Y}_n\,|\,\mathbf{Y}_0.s\mathbf{Y}_{n-1},\mathbf{X}\right] \\ &\triangleq E\left[\left(\mathbf{Y}_n - \left(\mathrm{daig}(\mathbf{X}_n)\mathbf{Q}(d)\right)\hat{\mathbf{H}}_n(d)\right)\times\left(\mathbf{Y}_n - \left(\mathrm{daig}(\mathbf{X}_n)\mathbf{Q}(d)\right)\hat{\mathbf{H}}_n(d)\right)^H\right] \\ &= \left(\mathrm{daig}(\mathbf{X}_n)\mathbf{Q}(d)\right)\mathbf{B}_n(d)\left(\mathrm{daig}(\mathbf{X}_n^*)\mathbf{Q}(d)\right) + \mathrm{diag}(\mathbf{A}_n) \end{aligned} \tag{42}$$

Where $\mathbf{B}_n(d)$ is the linear MMSE prediction error matrix obtained using the orthogonality principles.

Substituting (42) into (38) and into (37), $I(\mathbf{Y};\mathbf{G}(d)|\mathbf{X})$ is given as

$$\mathbf{I}(\mathbf{Y};\mathbf{G}(d)\,|\,\mathbf{X}) = \sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \mathrm{diag}\left(\mathbf{X}_n\right)\mathbf{B}_n(d)\,\mathrm{diag}\left(\mathbf{X}_n^*\right)\mathrm{diag}(\mathbf{A}_n)^{-1}\right)$$

$$= \sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \mathbf{B}_n(d)\,\mathrm{diag}\left(\mathbf{X}_n^*\right)\mathrm{diag}(\mathbf{A}_n)^{-1}\,\mathrm{diag}\left(\mathbf{X}_n\right)\right) \quad \because \det(\mathbf{I}+\mathbf{XY}) = \det(\mathbf{I}+\mathbf{YX}) \tag{43}$$

$$= \sum_{n=0}^{N-1} \log\det\left(\mathbf{I} + \mathbf{B}_n(d)\,\mathrm{diag}\left(\mathbf{S}(d)\right)\right)$$

The kth entry of the $K\times 1$ vector $\mathbf{S}(d)$ is $\sigma_x^2(d, f_k)Q^2(d, f_k) / A(f_k)$.

References

[1] Same as paper.

[2]

Master Thesis:
Compressive sensing and its application in Wireless Sensor Network & Correlated Signal Recovery Method

Abstract

In this paper, we consider the application of compressive sensing (CS) in wireless sensor networks (WSNs). CS is a signal acquisition and compression framework recently developed in the field of signal processing and information theory. We applied this CS technique to WSN which consists of a large number of wireless sensor nodes and a central fusion center (FC). This CS based signal acquisition and compression is done by a simple linear projection at each sensor node. Then, each sensor transmits the compressed samples to the FC. The FC which collects the compressed signals from the sensors jointly reconstructs the signals in polynomial time using a signal recovery algorithm. The distributed sensors observe similar event in designated region. Therefore, the observed signals have considerable correlation each other. We make some effort in modeling correlation between the signals acquired from the sensors and analyze the component in observed signals. After modeling the correlated signals, we propose POMP (Phased-OMP) which can recover any type of correlated signals stably and effectively. We introduce the idea of our proposed algorithm in detail and then compare the reconstruction performance of POMP with previous algorithms ReMBo, MEM, SOMP, etc.

Jae-Gun Choi

Contents

List of Figures

List of Tables

1. Introduction

In this paper, we discuss the application of a new compression technique called compressive sensing (CS) in wireless sensor networks (WSNs). The objective of a WSN which we assume in this paper is to collect information about events occurring in a region of interest. This WSN consists of a large number of wireless sensor nodes and a central fusion center (FC). The sensor nodes are spatially distributed over the said region to acquire physical signals such as sound, temperature, wind speed, pressure, and seismic vibrations. After sensing, they transmit the measured signals to the FC. In this paper, we focus on the role of the FC which is to recover the transmitted signals in their original waveforms for further processing. By doing so, the FC can produce a global picture that illustrates the event occurring in the sensed region. Each sensor uses its onboard battery for sensing activities and makes reports to FC via wireless transmissions. Thus, limited power at the sensor nodes is the key problem to be resolved in the said WSN.

CS is a signal acquisition and compression framework recently developed in the field of signal processing and information theory [1],[2]. Donoho [1] says that "The Shannon–Nyquist sampling rate may lead to too many samples; probably not all of them are necessary to reconstruct the given signal. Therefore, compression may become necessary prior to storage or transmission." According to Baraniuk [3], CS provides a new method of acquiring compressible signals at a rate significantly below the Nyquist rate. This method employs non-adaptive linear projections that preserve the signal's structure; the compressed signal is then reconstructed from these projections using an optimization process.

We applied this CS technique to WSN. One of our aims in this paper is to determine whether the CS can be used as a useful framework for the aforementioned WSN to compress and acquire signals and

save transmittal and computational power at the sensor node. This CS based signal acquisition and compression is done by a simple linear projection at each sensor node. Then, each sensor transmits the compressed samples to the FC; the FC which collects the compressed signals from the sensors jointly reconstructs the signal in polynomial time using a signal recovery algorithm. Illustrating this process in detail throughout this chapter, we check to see if CS can become an effective, efficient strategy to be employed in WSNs, especially for those with low-quality, inexpensive sensors.

The distributed sensors observe similar event in designated region. Therefore, the observed signals have considerable correlation each other. In this paper, as we assume a scenario in which a WSN is used for signal acquisition, we intend to pay some effort in modeling correlation between the signals acquired from the sensors. Then, we divide the correlated signals to three parts for example, common sparsity, innovation sparsity, and total sparsity. Those terminologies give more easy understanding to solve multiple measurement vector (MMV) modeled from WSN structure.

If we will use the correlated information to recover signals transmitted from each sensor, its reconstruction performance will increase over that not using correlated information. We demonstrated this assumption by showing a simulation result. Finally, we proposed advanced algorithm to recovery the correlated signals effectively. The proposed algorithm is called phased advanced orthogonal matching pursuit (POMP). POMP has better performance about reconstruction probability than previous algorithms, for examples, SOMP, ReMBo etc. We will introduce the idea of our proposed algorithm in detail and then compare the reconstruction performance of our algorithms with previous algorithms

2. Wireless sensor network

2.1. Network structure

We consider a WSN consisting of a large number of wireless sensor nodes and one FC (**Figure 1**). The wireless sensor nodes are spatially distributed over a region of interest and observe physical changes such as those in sound, temperature, pressure, or seismic vibrations. If a specific event occurs in a region of distributed sensors, each sensor makes local observations of the physical phenomenon as the result of this event taking place. An example of sensor network applications is area monitoring to detect forest fires. A network of sensor nodes can be installed in a forest to detect when a fire breaks out. The nodes can be equipped with sensors to measure temperature, humidity, and the gases produced by fires in trees or vegetation [7]. Other examples include military and security applications. Military applications vary from monitoring soldiers in the field, to tracking vehicles or enemy movement. Sensors attached to soldiers, vehicles and equipment can gather information about their condition and location to help planning activities on the battlefield. Seismic, acoustic and video sensors can be deployed to monitor critical terrain and approach routes; reconnaissance of enemy terrain and forces can be carried out [8].

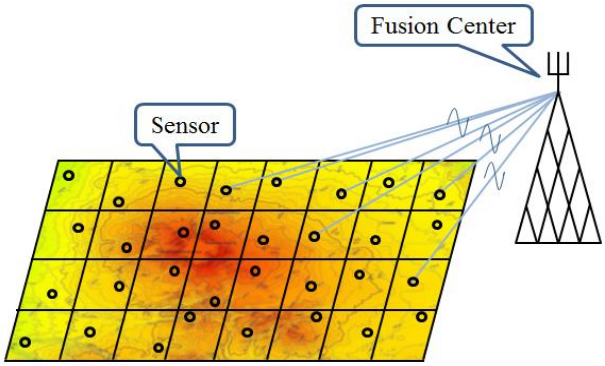


Figure 1. Wireless Sensor Network (WSN)

After sensors observe an event taking place in a distributed region, they convert the sensed information into a digital signal and transmit the digitized signal to the FC. Finally, the FC assembles the data transmitted by all the sensors and decodes the original information. The decoded information at the FC provides a global picture of events occurring in the region of interest. Therefore, we assume that the objective of the sensor network is to determine accurately and rapidly reconstruct transmitted information and reconstruct the original signal.

We discuss the resource limitations of WSNs in the next section.

2.2. Resource limitations in WSNs

In this section, we describe the assumptions made in the sensor network we are interested in. We assume that the sensors are distributed and supposed to communicate with the FC through a wireless channel. Because each sensor is important components of WSN which observes event, they should typically be deployed in a large volume over the region of interest. Therefore, they are usually designed to be inexpensive and small. For that reason, each sensor operates on an onboard battery which is not rechargeable at all; thus, for simplicity, the hardware implementation of sensor nodes can provide only limited computational performance, bandwidth, and transmission power. As a result of limitations on the hardware implementation in sensor nodes, the FC has powerful computation performance and plentiful energy which naturally performs most of the complex computations.

Under the limited conditions stated above for a WSN, CS can substantially reduce the data volume to be transmitted at each sensor node. With the new method, it is possible to compress the original signal using only $O\left(k\log\left(n/k\right)\right)$ samples without going through many complex signal processing steps. These signals can be recovered successfully at the FC. All these are done under the CS framework. As the result, the consumption of power for transmission of signal contents at each sensor can be significantly reduced thanks to decreased data volume. Moreover, this data reduction comes without utilizing complex signal processing. Namely, the sensor nodes can compress the signal while not spending any power for running complex compression algorithms onboard.

We discuss the new technique CS in the next section and check how CS can get the advantages like data reduction and simple data compression.

3. Compressive sensing (Literature survey)

In a conventional communication system, an analog-to-digital converter based on the Shannon–Nyquist sampling theorem is used to convert analog signals to digital signals. The theorem says that if a signal is sampled at a rate twice, or higher, the maximum frequency of the signal, the original signal can be exactly recovered from the samples. Once the sampled signals are obtained over a fixed duration of time, a conventional compression scheme can be used to compress them. Because the sampled signals often have substantial redundancy, compression is possible. Several compression schemes follow this approach, e.g., the MP3 and JPEG formats for audio or image data. However, conventional compression in a digital system is sometimes inefficient because it requires unnecessary signal processing stages, for example, retaining all of the sampled signals in one location before data compression. According to Donoho [1], the CS framework, as shown in **Figure 2**, can bypass these intermediate steps, and thus provides a light weight signal acquisition apparatus which is suitable for those sensor nodes in our WSN.

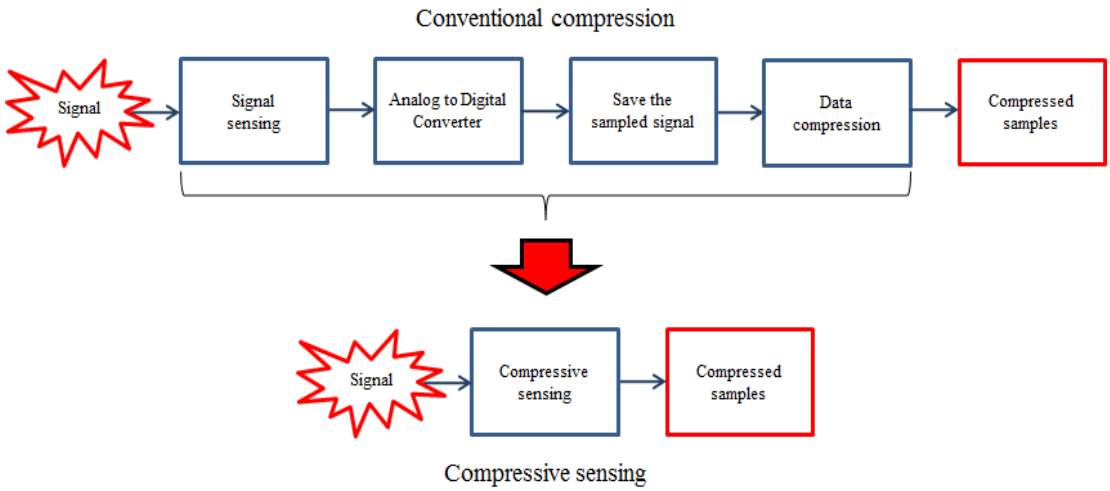


Figure 2. Conventional compression and compressive sensing

The CS provides a direct method which acquires compressed samples without going through the intermediate stages of conventional compression. Thus, CS provides a much simpler signal acquisition solution. In addition, the CS provides several recovery routines which the original signal can be regenerated perfectly from the compressed samples.

3.1 Theoretical background

Let a real-valued column vector $\mathbf{s}$ be a signal to be acquired. Let it be represented by

$$\mathbf{s} = \Psi\mathbf{x} \tag{1}$$

,where $\mathbf{x}$ and $\mathbf{s} \in \mathbf{R}^n$, and $\mathbf{x}$ is also a real-valued column vector. The matrix $\Psi \in \mathbf{R}^{n\times n}$ is an orthonormal basis, i.e., $\Psi^T\Psi = \Psi\Psi^T = I_n$, the identity matrix of size $\mathbf{R}^{n\times n}$. The signal $\mathbf{s}$ is called k-sparse if it can be represented as a linear combination of only k columns of Ψ, i.e., only the k components of the vector $\mathbf{x}$ are nonzero as represented Eq.**Error! Reference source not found.** .

.

$$\mathbf{s} = \sum_{i=1}^{n} x_i\psi_i, \text{ where } \psi_i \text{ is a column vector of } \Psi. \tag{2}$$

A signal is called compressible if it has only a few significant (large in magnitude) components and a greater number of insignificant (close to zero) components. The compressive measurements $\mathbf{y}$ (compressed samples) are obtained via linear projections as follows (**Figure 3**):

$$\mathbf{y} = \Phi\mathbf{s} = \Phi\Psi\mathbf{x} = \mathbf{A}\mathbf{x} \tag{3}$$

where the measurement vector is $\mathbf{y} \in \mathbf{R}^m$, with $m < n$, and the measurement matrix $\mathbf{A} \in \mathbf{R}^{m\times n}$. Our goal is to recover $\mathbf{x}$ from the measurement vector $\mathbf{y}$. We note that Eq. **Error! Reference source not found.** is an underdetermined system because it has fewer equations than unknowns; thus, it does not have a unique solution in general. However, the theory of CS asserts

that, if the vector $\mathbf{x}$ is sufficiently sparse, an underdetermined system is guaranteed with high probability to have a unique solution.

In this section, we discuss the basics of CS in more detail.

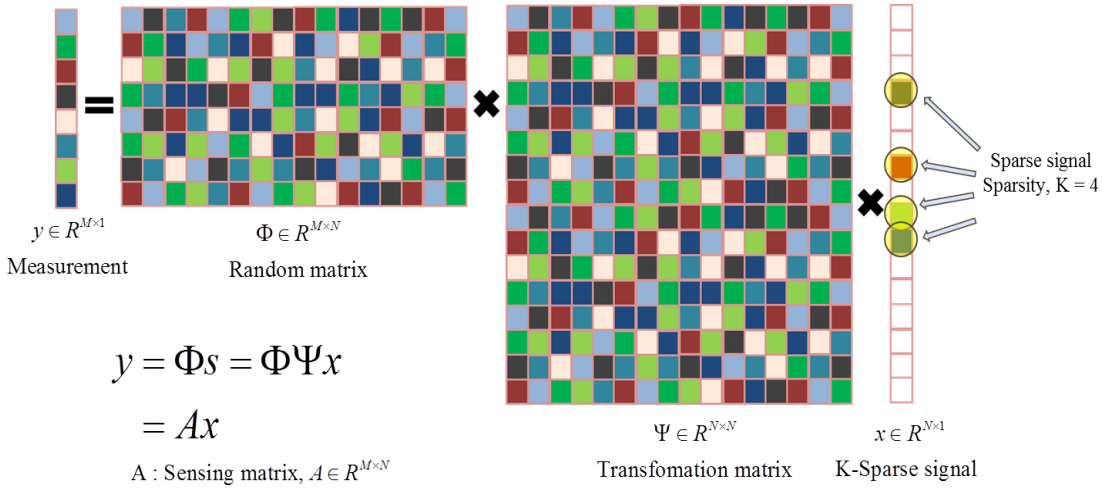


Figure 3. The summary of compressive sensing

i) k -sparse signal $\mathbf{x}$ in orthonormal basis

The k -sparse signal, $\mathbf{s}$ in Eq. **Error! Reference source not found.**, has k nonzero components in $\mathbf{x}$. The matrix Ψ is, again, an orthonormal basis, i.e., $\Psi^T\Psi = \Psi\Psi^T = I_n$, the identity matrix of size $\mathbf{R}^{n\times n}$.

ii) Measurement vector $\mathbf{y}$ and underdetermined system

The sensing matrices are Φ and $\mathbf{A}$ in Eq. **Error! Reference source not found.**, where its dimension $\mathbf{R}^{m\times n}$, $m < n$. When m is closer to k than n is, sufficient conditions for good signal recovery are satisfied. Then a compression effect exists. Note that Eq.

Error! Reference source not found. appears to be an ill-conditioned equation. That is, the number of unknowns n is larger than m the number of equations, $m < n$. However, if $\mathbf{x}$ is k-sparse and the locations of the k nonzero elements are known, the problem can be solved provided $m \geq k$. We can form a simplified equation by deleting all those columns and elements corresponding to the zero-elements, as follows:

$$\mathbf{y} = \mathbf{A}_{\kappa}\mathbf{x}_{\kappa} \tag{4}$$

where $\kappa \in \{1, 2, \cdots, n\}$ is the support set, which is the collection of indices corresponding to the nonzero elements of x. Note that the support set κ can be any size-k subset of the full index set, $\{1, 2, 3, ..., n\}$. Eq. **Error! Reference source not found.** has the unique solution $\mathbf{x}_{\kappa}$ if the columns of $\mathbf{A}_{\kappa}$ are linearly independent. The solution can be found using pseudo inverse easily as Eq. **Error! Reference source not found.**

$$\mathbf{x}_{\kappa} = \left(\mathbf{A}_{\kappa}^{T}\mathbf{A}_{\kappa}\right)^{-1}\mathbf{A}_{\kappa}^{T}\mathbf{y} \tag{5}$$

Thus, if the support set κ can be found, the problem is easy to solve provided the columns are linearly independent.

iii) Incoherence condition

The incoherence condition is that the rows of Φ should be incoherent to the columns of Ψ. If the rows of Φ are coherent to the columns of Ψ, the matrix $\mathbf{A}$ cannot be a good sensing matrix. In the extreme case, we can show a matrix $\mathbf{A}$ having m rows of Φ that are the first m columns of Ψ.

$$\mathbf{A} = \Phi\Psi = \Psi^{T}_{(1:m,:)}\Psi = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \tag{6}$$

If $\mathbf{A}$ of Eq. **Error! Reference source not found.** is used as sensing matrix, the compressed measurement vector $\mathbf{y}$ captures only the first m elements of the vector $\mathbf{x}$, and the rest of the information contained in $\mathbf{x}$ is completely lost.

iv) Designing a sensing matrix Φ

One choice for designing a sensing matrix Φ is Gaussian. Under this choice, the sensing matrix Φ is designed as a Gaussian, i.e., matrix elements are independent and identically distributed Gaussian samples. This choice is deemed good since a Gaussian sensing matrix satisfies the incoherence condition with high probability for any choice of orthonormal basis Ψ. This randomly generated matrix acts as a random projection operator on the signal vector $\mathbf{x}$. Such a random projection matrix needs not depend on specific knowledge about the source signals. Moreover, random projections have the following advantages in the application to sensor networks [5].

1) Universal incoherence: Random matrices Φ can be combined with all conventional sparsity basis Ψ, and with high probability sparse signals can be recovered by an L_1 minimum algorithms from the measurements $\mathbf{y}$.

2) Data independence: The construction of a random matrix does not depend on any prior knowledge of the data. Therefore, given an explicit random number generator, only the sensors and the fusion center are required to agree on a single random seed for generating the same random matrices of any dimension.

3) Robustness: Transmission of randomly projected coefficients is robust to packet loss in the network. Even if part of the elements in measurement $\mathbf{y}$ is lost, the receiver can still recover the sparse signal, at the cost of lower accuracy.

3.2 System equations

We knew the method how to find a unique solution of CS problem in previous section. In this section, we discuss various equations which are handled in CS theory as single measurement vector (SMV) and multiple measurement vector (MMV). The SMV is a basic equation for CS. It is expressed as Eq. **Error! Reference source not found.**. Many CS paper about this SMV problem is researched in [Ref],[Ref].

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & & a_{2,n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} \quad (7)$$

Otherwise, the MMV has multiple measurement vectors and sparse matrix as Eq. **Error! Reference source not found.**. The sparse vector in each SMV results in MMV. It has much unknowns compared with SMV. The many number of unknowns may make the MMV to be solved hard. To solve this equation effectively, some algorithms are proposed as SOMP, ReMBo, M-FOCUSS. If each column of sparse matrix $\mathbf{X}$ has similar support set, the priori information about support location can be used to get exact solution easily.

$$\mathbf{Y} = \mathbf{AX}$$

$$\begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,J} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,J} \\ & & & \\ y_{m,1} & y_{m,2} & \cdots & y_{m,J} \end{bmatrix}_{1,} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & & a_{2,n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,J} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,J} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ & & & \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,J} \end{bmatrix}_{1} \quad (8)$$

The MMV equation can have the more number of equations by transforming Eq. **Error! Reference source not found.**. It means that the MMV equation has more information to solve underdetermined equation. The modified MMV equation is expressed as below Eq. **Error! Reference source not found.**. Furthermore, infinite measurement vector (IMV) consists of an infinite set of jointly sparse vectors.

$$\mathbf{y}_j = \mathbf{A}_j \mathbf{x}_j, \text{ where } j \in \{1, 2, ..., J\}$$

$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ \\ y_{m-1,1} \\ y_{m,1} \end{bmatrix}_1 = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & & a_{2,n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & \cdots & a_{m,n} \end{bmatrix}_1 \begin{bmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ \vdots \\ \\ \vdots \\ x_{n,1} \end{bmatrix}_1 ,\cdots, \begin{bmatrix} y_{1,J} \\ y_{2,J} \\ \vdots \\ \\ y_{m-1,J} \\ y_{m,J} \end{bmatrix}_J = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & & a_{2,n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & \cdots & a_{m,n} \end{bmatrix}_J \begin{bmatrix} x_{1,J} \\ x_{2,J} \\ \vdots \\ \vdots \\ \\ \vdots \\ x_{n,J} \end{bmatrix}_J$$

$$\mathbf{Y} = \mathbf{A}_j \mathbf{X}, \text{ where } j \in \{1, 2, ..., J\}$$

$$\Rightarrow \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,J} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,J} \\ & & & \\ y_{m,1} & y_{m,2} & \cdots & y_{m,J} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & & a_{2,n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & \cdots & a_{m,n} \end{bmatrix}_j \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,J} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,J} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ & & & \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,J} \end{bmatrix} \quad (9)$$

Therefore, we can draw the relationship among the SMV, MMV and IMV as **Figure 4.** As the **Figure 4** shows, the MMV includes all of the SMV. It means that the MMV has all the information of the SMV. Therefore, if we solve the MMV equation exactly, it results the solution of each SMV also.

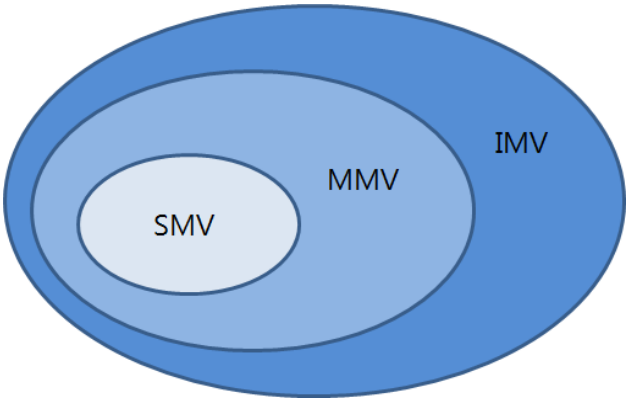


Figure 4. The relationship among SMV, MMV and IMV

3.3. Unique solution condition of SMV and MMV

In CS, a core problem is to find a unique solution for an underdetermined equation. This problem is related to the signal reconstruction algorithm, which takes the measurement vector $\mathbf{y}$ as an input and the k-sparse vector $\mathbf{x}$ as an output. To solve an underdetermined problem, we consider minimization criteria using different norms such as the L_2, L_1, and L_0 norms. The L_p norm of a vector $\mathbf{x}$ of length n is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad p > 0 \tag{10}$$

.

Although we can define the L_2 and L_1 norms as $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$, respectively, using the definition of L_p norm, L_0 norm cannot be defined this way. The L_0 norm is a pseudo-norm that counts the number of nonzero components in a vector as defined by Donoho and Elad [6]. Using this definition of norms, we will discuss the minimization problem to get solution $\mathbf{x}$.

i) The minimization problem in SMV

1) L_2 norm minimization in SMV

$$\begin{aligned} (L_2)\ \hat{\mathbf{x}} &= \arg\min \|\mathbf{x}\|_2 \quad \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \text{ where } \mathbf{A} \in \mathrm{R}^{m \times n},\ rank(\mathbf{A}) = m \\ &= \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T \right)^{-1} \mathbf{y} \end{aligned} \tag{11}$$

However, this conventional solution yields a non-sparse solution, so it is not appropriate as a solution to the CS problem. Thus, we do not consider this method for finding solution.

2) L_0 norm minimization in SMV

$$(L_0) \text{ Minimize } \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \text{ where } \mathbf{A} \in \mathrm{R}^{m \times n},\ rank(\mathbf{A}) = m \tag{12}$$

The L_0 norm of a vector is, by definition, the number of nonzero elements in the vector. In the CS literature, it is known that the L_0 norm problem can be solved by examining all the possible cases. Since this process involves a combinatorial search for all possible $\binom{n}{k}$ support sets, it is an NP-complete problem. Thus, we cannot solve it within polynomial time. Therefore, we consider L_1 norm minimization as an alternative. In literature [Ref], the unique solution of the L_0 minimization is known as following,

$$k < \frac{\mathrm{spark}(\mathbf{A})}{2} \tag{13}$$

The $\mathrm{spark}(\mathbf{A})$ is the smallest number n such that there exists a set of n columns in $\mathbf{A}$ which are linearly dependent. In summary, if the above equation is satisfied, then the unique solution of the Eq. **Error! Reference source not found.** is guaranteed.

3) L_1 norm minimization in SMV

$$(L_1) \text{ Minimize } \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \text{where } \mathbf{A} \in \mathrm{R}^{m \times n},\ rank(\mathbf{A}) = m \tag{14}$$

This L_1 norm minimization can be considered as a relaxed version of the L_0 problem. Fortunately, the L_1 problem is a convex optimization problem and in fact can be recast as a linear programming problem. For example, it can be solved by an interior point method. Many effective algorithms have been developed to solve the minimum L_1 problem, and it will be considered later in this chapter. Here, we aim to study the sufficient conditions under which Eq. **Error! Reference source not found.** and **Error! Reference source not found.** have unique solutions. We provide a theorem related to this issue.

L_0 / L_1 **equivalence condition in SMV:**
Let $\mathbf{A} \in \mathbf{R}^{m \times n}$ be a matrix with a maximum correlation definition μ, $\mu(\mathbf{A}) = \max_{i \neq j} \left\| \langle \mathbf{a}_i, \mathbf{a}_j \rangle \right\|$, where $\mathbf{a}_i$ is the i th column vector of $\mathbf{A}$ with $i = 1, 2, ..., n$, and $\mathbf{x}$ is a k -sparse signal. Then, if $k < \frac{1}{2}\left(1 + \frac{1}{\mu}\right)$ is satisfied, then the solution of L_1 coincides with that of L_0 [6] .

Table 1. L_0 / L_1 Equivalence condition.

ii) The minimization problem in MMV

To get the unique solution of MMV, it can be considered similar method with that of SMV. We introduce theorems from references [Ref]. To explain the uniqueness condition for MMV, we introduce the following definitions $R(\mathbf{X})$ and $relax(\mathbf{X})$.

$$R(\mathbf{X}) = \left\| m(\mathbf{x}_i)_{n \times 1} \right\|_0 \tag{15}$$

where $x_i \in R^L$ is the transpose of the ith row of matrix $\mathbf{X}$, ie $\mathbf{X} = [x_1, x_2, ..., x_n]^T$, $m(\cdot)$ is any vector norm in R^L. Therefore, $R(\mathbf{X})$ is the number of rows which have nonzero element in matrix $\mathbf{X}$. When norm of $m(\mathbf{x}_i)_{n\times 1}$ is one, then it is defined as $relax(\mathbf{X})$.

$$relax(\mathbf{X}) = \left\| \left(m(\mathbf{x}_i) \right)_{n\times 1} \right\|_1 \tag{16}$$

1) L_0 norm minimization in MMV

$$(L_0) \quad \text{Minimize } R(\mathbf{X}) = \left\| m(\mathbf{x}_i)_{n\times 1} \right\|_0 \quad \text{subject to} \quad \mathbf{A} \in \mathbb{R}^{m\times n}, \text{with } rank(\mathbf{A}) = m \tag{17}$$

In literature [Ref], the MMV unique solution of the L_0 minimization is known as following,

$$R(\mathbf{X}) < \frac{spark(\mathbf{A}) - 1 + rank\left(Cols(\mathbf{Y})\right)}{2} \tag{18}$$

The $rank\left(Cols(\mathbf{Y})\right)$ is the column rank of matrix $\mathbf{Y}$. If the above equation is satisfied, then the unique solution of the Eq. **Error! Reference source not found.** is guaranteed.

2) L_1 norm minimization in MMV

$$(L_1) \quad \text{Minimize } relax(\mathbf{X}) = \left\| m(\mathbf{x}_i)_{n\times 1} \right\|_1 \quad \text{subject to} \quad \mathbf{A} \in \mathbb{R}^{m\times n}, \text{with } rank(\mathbf{A}) = m \tag{19}$$

A sufficient condition to be the unique solution to 2) of MMV is that

$$\left\| \mathbf{A}_S^{\dagger} \mathbf{A}_j \right\|_1 < 1, \ \forall j \notin S \tag{20}$$

.

$\mathbf{A}_S$ is reduced matrix of $\mathbf{A}$ corresponding to indices from support location of $R(\mathbf{X})$. So, we can write $\mathbf{Y} = \mathbf{A}_S \mathbf{X}_S$, where matrix $\mathbf{X}_S$ is made by nonzero rows of $\mathbf{X}$. $\mathbf{A}_S$ is of full column rank. $\mathbf{A}_S^{\dagger}$ is pseudo-inverse which is defined by $\mathbf{A}_S^{\dagger} = \left(\mathbf{A}_S^T \mathbf{A}_S \right)^{-1} \mathbf{A}_S^T$. Because $\mathbf{A}_S$ is of full column rank, the generalized inverse is well defined. The above is the Exact Recovery Condition (ERC) in Tropp's "Greed is good: Algorithmic results for sparse approximation"

L_0 / L_1 **equivalence condition in MMV:**
If $R(\mathbf{X}) < \frac{spark(\mathbf{A})}{2}$ is satisfied, then the solution of L_1 in MMV coincides with that of L_0 [Ref] .

Table 1. L_0 / L_1 Equivalence condition.

4. Compressive sensing and its application in WSN

4.1 The usefulness of CS in WSNs

In this section, we provide a brief comparison of using CS and using the conventional compression in a WSN. This comparison illustrates why CS could be a useful solution for WSNs.

i) Sensor network scheme with conventional compression

For a conventional sensor system, the distributed sensors observe physical changes in designed area. Since each sensor observes similar physical changes, the signals observed from each sensor have much correlation. The correlated signal can be compressed for reducing data. The conventional compression for WSN requires exchanging information between distributed sensors in order to exploit inter-sensor correlation. Such a transmission strategy makes the network system complex below **Figure 9**.

The conventional compression needs to get together redundant data for compression as **Figure 10**. At the collection point, joint compression can be made and compressed information can be sent to the FC. This option has a couple drawbacks. First, gathering the samples from all the sensors and jointly compressing them cause a transmission delay. Second, a lot of onboard power should be spent at the collaboration point. Third, each sensor should be collocated so that the transmitted information can be gathered at collaboration location.

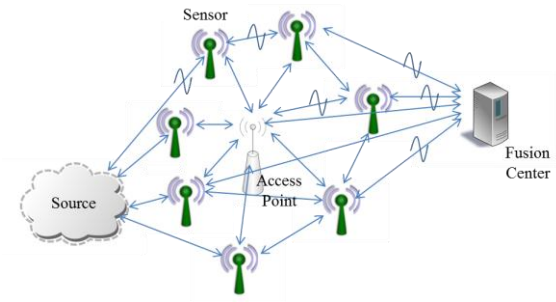


Figure 9. Conventional sensor network structure

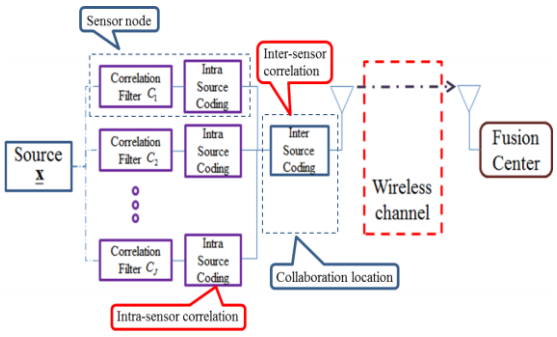


Figure 10. Conventional sensor network structure

Now, we may suppose that the joint compression is not aimed at and each sensor compresses the signal on its own. First, the data reduction effect with this approach will be limited because inter-sensor correlation is not exploited at all. The total volume of the independently compressed data is much larger than that of jointly compressed data. This may produce a large traffic volume in the WSN and a large amount of transmission power will be wasted from the sensor nodes which transmit essentially the same information to the FC. Thus, this is an inefficient strategy as well.

ii) Sensor network scheme with compressive sensing

In contrast to the conventional schemes considered in the previous paragraph, the CS method aims to acquire compressed samples directly. If a high-dimensional observation vector $\mathbf{x}$ exhibits sparsity in a certain domain (by exploiting intra-sensor correlation), CS provides the *direct method* for signal compression as discussed in **Figure 2**. To compress the high-dimensional signal $\mathbf{x}$ into a low-dimensional signal $\mathbf{y}$, as Eq. **Error! Reference source not found.**, it uses a simple matrix multiplication with an $m \times n$ projection matrix $\mathbf{A}_j$, $j \in \{1,2,...J\}$, where j is the sensor index, as depicted in **Figure 12**.

In the CS-based sensor network scheme, each sensor compresses the observed signals using a simple linear projection and transmits the compressed samples to the FC. Then, the FC can jointly reconstruct the received signals (by exploiting inter-sensor correlation) using one of the CS algorithms. Therefore, each sensor does not need to communicate with its neighboring sensors for joint compression. Our method is distributed compression without having the sensors to talk to each other; only the joint recovery at the FC is needed. Thus, no intermediate stages are required which are to gather all of the samples at a single location and carry out compression aiming to exploiting inter-sensor correlation. This free of intermediate stages allow us to reduce time delay significantly as well. Therefore, if the original

data are compressed by CS, each sensor node produces much smaller traffic volume which can be transmitted to the FC at a much lower transmission power and with a smaller time delay. The CS sensor network structure applied for WSN is as below (**Figure 11**). You can check the simplicity of transmission strategy of CS based WSN compared with conventional network.

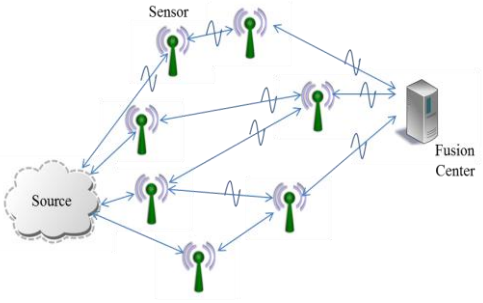


Figure 11. CS sensor network scheme

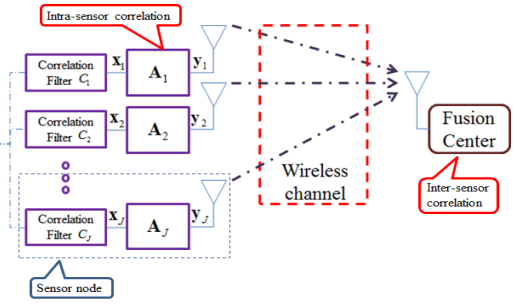


Figure 12. CS sensor network structure

4.2 Distributed compressive sensing

Each sensor can observe only the local part of an entire physical phenomenon, and a certain event of interest is measured by one or more sensors. Therefore, the sensed signals are often partially correlated. These measured signals have two distinct correlations: intra-sensor correlation and inter-sensor correlation. Intra-sensor correlation exists in the signals observed by each sensor. Once a high-dimensional sensed signal has a sparse representation in a certain domain, we can reduce its size by using CS. This process exploits the intra-sensor correlation. In contrast, inter-sensor correlation exists between the signals sensed by different sensors. By exploiting inter-sensor correlation, further reduction in transmitted signals can be made.

These two correlations can be exploited to improve the system performance. As the number of sensors in a region becomes dense, each sensor has a strongly correlated signal that is similar to that of neighboring sensors. In contrast, if we decrease the density of sensors distributed in a given region, the sensed signals will obviously be more weakly correlated with each other. In this section, we discuss two strategies for transmitting signals in a multi-sensor CS-based system. One strategy uses only intra-sensor correlation, and the other uses both types of correlation. We illustrate that CS-based system in WSN exploits the inter-sensor correlation more effectively and simply than that of conventional sensor network.

i) Exploiting only intra-sensor correlation

In **Figure 13**, each sensor observes the source signal and independently compresses it to a low-dimensional signal. After compression, each sensor transmits the compressed signal to the FC. Without exploiting inter-sensor correlation between transmitted signals, the FC recovers these signals separately. In this case, even if there exists correlation among the sensed signals, because only intra-sensor

correlation is exploited, we cannot gain any advantages from joint recovery. This method has the following characteristics:

1) Independent compression and transmission at each sensor

2) Signal recovery by exploiting only intra-sensor correlation at the FC

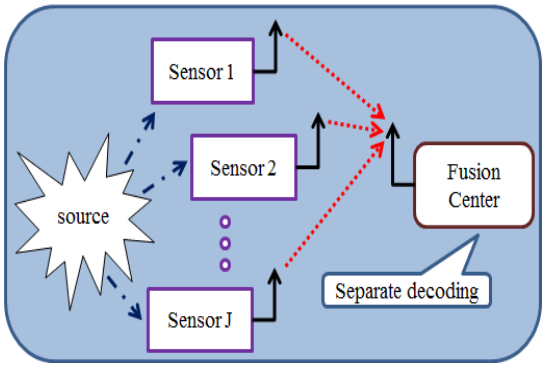
ii) Exploiting both intra- and inter-sensor correlation

Figure 14 shows the same process as in situation *i*) above, except that the FC exploits the inter-sensor correlation among sensed signals at signal reconstruction stage. In conventional sensor network system as shown in **Figure 10**, the sensor nodes communicate with their neighboring sensors to take advantage of joint compression by exploiting inter-sensor correlation. However, in the CS-based system, a stage for exploiting inter-sensor correlation is achieved at FC. It means that if inter-sensor correlation exists within the sensed signals, and the FC can exploit it. This is done with sensors communicating with the FC but not among the sensors themselves. We refer to this communication strategy as the Distributed Compressive Sensing (DCS). Exploitation of inter-sensor correlation should be manifested with the reduction of the measurement size m of matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$, where $\mathbf{y} = \mathbf{A}\mathbf{x}$, required for good single recovery. The characteristics of our DCS sensor network are:

1) Independent compression and transmission at each sensor

2) Exploitation of inter-sensor signal correlation with the joint recovery scheme at the FC

3) Variation of the per sensor CS measurements to manipulate the level of signal correlation



F i g u r e .1 B n t r a - s e n s o r c o r r e l

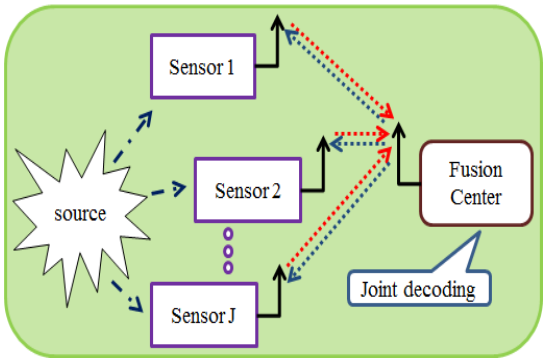
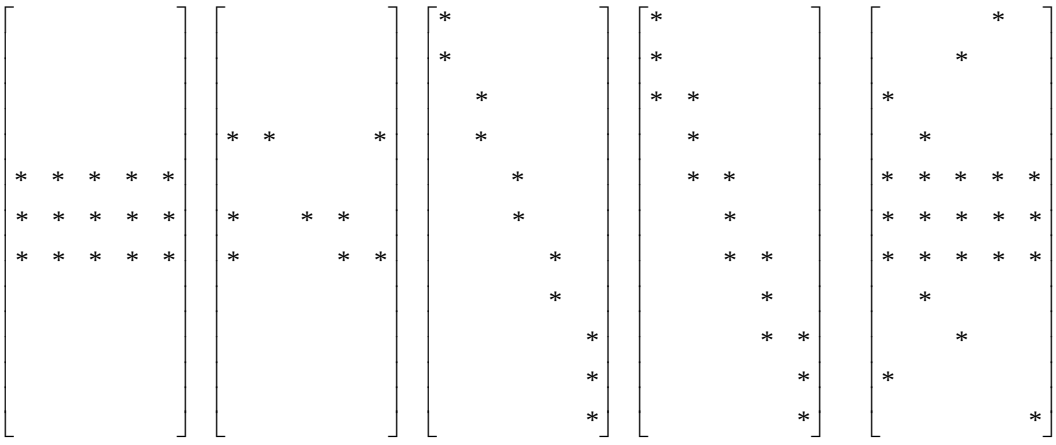


Figure 14. Intra/Inter-sensor correlation scheme

4.3. Correlated signal models

We assumed the WSN which consists of large number of sensor with a built-in CS and one fusion center. Because of the feature of considered WSN, the observed signals have inter-sensor correlation. We can model this WSN as Eq. **Error! Reference source not found.** or Eq. **Error! Reference source not found.**. Those two equations have sparse signal matrix **X** which consists of signals transmitted from each sensor.

In this section, we introduce how the signal matrix with different degrees of correlation can be generated as sparse signal models. The sparse signal matrix in WSN has correlated properties. The degree of sparseness which is called the sparsity, is proportional to the amount of correlation. More correlated signal means sparser in terms of intra-sensor correlation. In addition, inter-sensor signal correlation can be modeled *i*) by the degree of overlaps in the support sets of any two sparse signals, and *ii*) by the correlation of non-zero signal values. By using those two properties, we can model correlated sparse signal matrix **X** as below examples **Figure 15**.



*: Unknown nonzero value

Figure 15. The examples of correlated signals

We can divide those correlated signals in **Figure 15** as three components; common sparsity part, innovation sparsity part, and total sparsity part. The common sparsity part has one more nonzero value in the row of sparse signal matrix **X**. The Innovation sparsity part has only one nonzero value in the row of signal matrix. Lastly, the total sparsity part is the total number of rows which have nonzero elements. The common sparsity is a correlated part. Therefore, if we find the location of common part, we can also use it to solve another SMV. The innovation sparsity is a uncorrelated part. Even if we find the location of innovation part, we cannot use it to solve other SMV equations. Finally, the total sparsity is related with the degree of correlation among observed signals. We re-expressed the correlated signals as following **Figure 16** by using three terminologies mentioned.

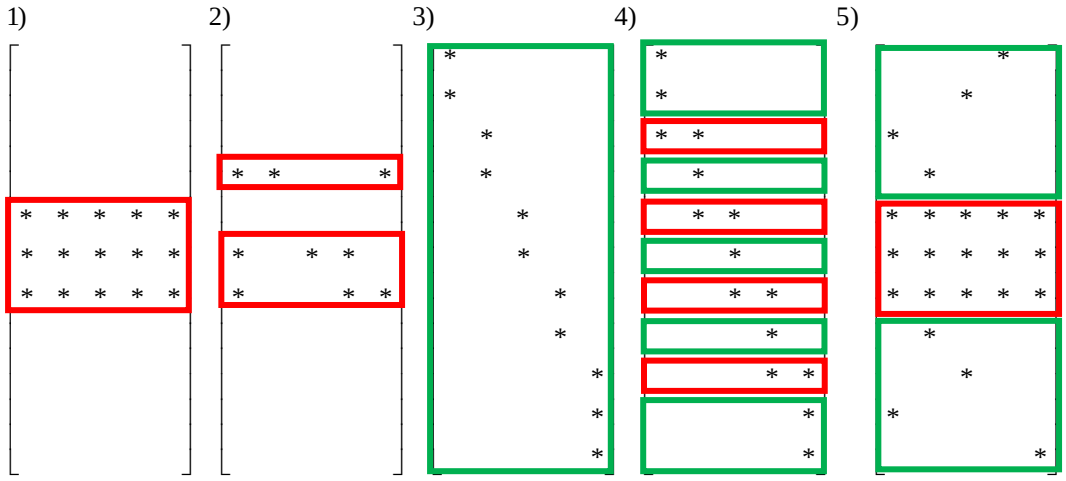


Figure 16. The components of correalted signals

The first and second correlated signals of **Figure 16** have only common sparsity part. The third signal consists of only innovation sparsity. Therefore, there is no common part which has one

more nonzero element in same row. The 4), 5) signals have both common part and innovation part. If we know the prior information of the heuristic signal **X** about support location, we can use it to find solution effectively. We can use those correlated properties to recover signals transmitted from each sensor, and its reconstruction performance will increase over that not using correlated information. We discuss the ideas for recovering those correlated signals in next section.

5. The recovery ideas for correlated signals

5.1 Joint decoding and separate decoding

We discussed the correlated signals which consist of common, innovation, and total parts in the previous section. The understanding of various correlated signal models gives more clues to get solution. In this section, we argue ideas using correlated information to get solution effectively.

Some specific correlated signals also are handled in [Ref],[Ref]. In those references, the correlation signals are referred to as JSM-1 (joint signal model) or JSM-2 depending on the correlation type. In JSM-1, all of the signals share exactly the same common nonzero components that have the same values, whereas each signal also independently has different nonzero components, which is called innovation. In JSM-2, it shares same support location that has different value. Those two signals is expressed below, **Figure 17**. In [Ref],[Ref], they proposed methods which find the solution of the correlated signals consisting of those specific pattern in Eq. **Error! Reference source not found.**.

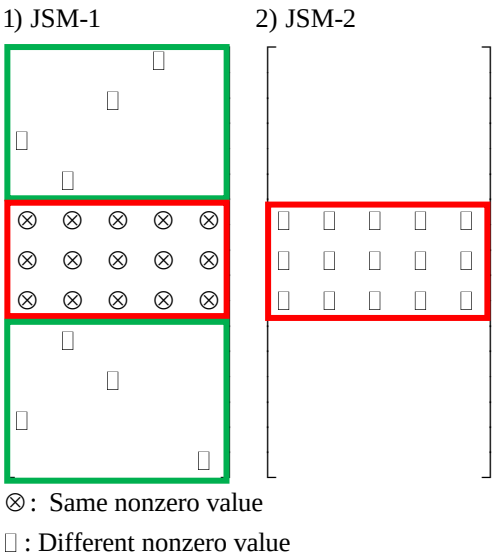


Figure 17. Joint signal models, JSM-1, JSM-2

The JSM-1 is expressed as

$$\mathbf{x}_j = \mathbf{z}_c + \mathbf{z}_j,\ j \in \{1,2,...,J\},\ j \text{ is the index of the sensors} \tag{21}$$

where $\|\mathbf{z}_c\|_0 = k_c$ (Common part), and $\|\mathbf{z}_j\|_0 = k_j$ (Innovation part) in each sensed signal. Obviously, $\mathbf{z}_c$ appears in all the columns of the correlated signals. It can be recognized as the inter-sensor correlation. We note that the intra-sensor correlation is that all of the signals are sparse. The j th sensor transmits $\mathbf{y}_j = \mathbf{A}_j\mathbf{x}_j$ to the FC. After all the sensed signals are transmitted to the FC, the FC aims to recover all the signals. Because inter- sensor correlation exists in the sensed signals, we can obtain several benefits by using the correlated information in the transmitted signals. For ease of explanation, suppose that the WSN contains J sensors, and its sensed signal follows JSM-1 pattern. Then, the FC can exploit both intra- and inter- sensor correlation by solving Eq. **Error! Reference source not found.** as described below.

i) Joint recovery scheme for JSM-1 (Modified equation method)

The sensed signals from j sensors can be expressed as follows.

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{z}_c + \mathbf{z}_1 \in \mathbf{R}^n \\ \mathbf{x}_2 &= \mathbf{z}_c + \mathbf{z}_2 \in \mathbf{R}^n \\ &\vdots \\ \mathbf{x}_J &= \mathbf{z}_c + \mathbf{z}_J \in \mathbf{R}^n \end{aligned},$$

where the sparsity of vectors $\mathbf{z}_c$ and $\mathbf{z}_j$ are k_c and k_j, respectively and each sensor has same spasity $k = k_c + k_j$. Then, the transmitted signal $\mathbf{y}_j$ can be divided into two parts as follows.

$$\mathbf{y}_j = \mathbf{A}_j(\mathbf{z}_c + \mathbf{z}_j) = \mathbf{A}_j\mathbf{z}_c + \mathbf{A}_j\mathbf{z}_j$$

If the FC received all the signals transmitted from J sensors, it then concatenates the used sensing matrix and received signal using Eq. **Error! Reference source not found.**. Therefore, the sensed signal in JSM-1 is transformed into **Figure 18**. This idea is handled in [Ref],[Ref].

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_J \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_3 & \mathbf{0} & \mathbf{0} & \mathbf{A}_3 & \mathbf{0} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{A}_J & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_J \end{bmatrix} \begin{bmatrix} \mathbf{z}_c \\ \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \\ \vdots \\ \mathbf{z}_J \end{bmatrix} \tag{22}$$

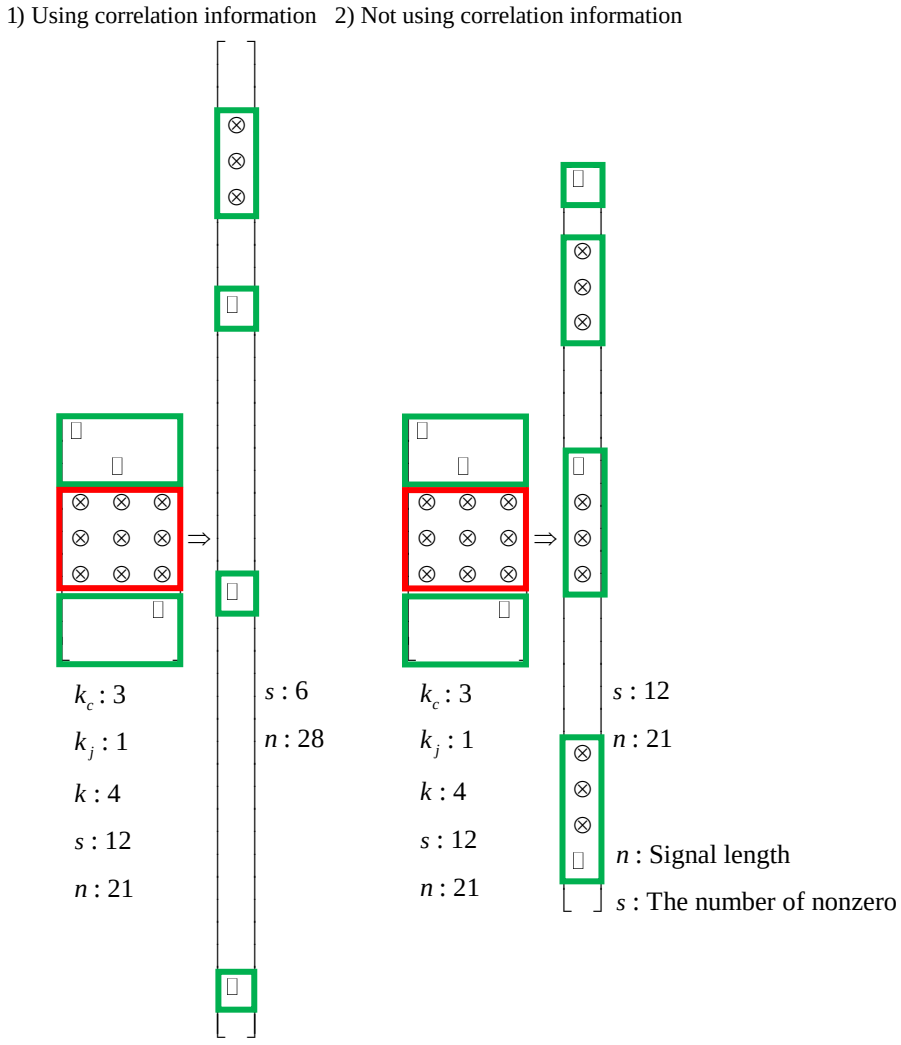


Figure 18. Concatenating JSM-1 to a column signal

Because JSM-1 shares common part $\mathbf{z}_c$ in the equation, we can reduce the number of nonzero value as 1) of **Figure 18**. In conclusion, the total number of nonzero in matrix $\mathbf{X}$ is 12, but in transforming equation, it is 6 only. Thus, the total number of nonzero, s is reduced from $J \times \left(k_c + k_j \right)$ to $k_c + \left(J \times k_j \right)$. The total number of sparsity affects the probability of exact reconstruction. By solving this equation, the FC can take advantage of exploiting inter-sensor correlation. However, if the FC recovers the received signals independently without using any correlation information, separate recovery is done. Even if the sensed signals are correlated, separate recovery offers no advantages for signal reconstruction because it does not exploit inter- sensor correlation.

ii) Separate recovery scheme for JSM-1

Even if a common correlated element exists in the sensed signals, separate recovery does not use that correlation information as before example. Therefore, the received signals are recovered as follows and its concatenated signal is express as 2) of **Figure 18**.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_J \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_J \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_J \end{bmatrix} \tag{23}$$

To solve Eq. **Error! Reference source not found.** and **Error! Reference source not found.**, we use the primal-dual interior point method (PDIP) in **Appendix 7.1**, which is an L_1 minimization algorithm, and compare the results of the two types of recovery, joint decoding and separate decoding respectively. Using the comparison results, we can confirm that the measurement size required for perfect reconstruction is smaller for joint recovery than for separate recovery.

Now, we introduce JSM-2, which is simpler than JSM-1. All the signal coefficients are different, but their indices for nonzero components are the same. Suppose that there exist two signals, $\mathbf{x}_1$ and $\mathbf{x}_2$. The i th coefficient for $\mathbf{x}_1$ is nonzero if and only if the i th coefficient for $\mathbf{x}_2$ is nonzero. This property represents inter-sensor correlation, because if we know the support set for $\mathbf{x}_1$, then we automatically know the support set for $\mathbf{x}_2$.

iii) Joint and separate decoding scheme for JSM-2

The prior inter-correlation becomes relevant when the number of sensors is more than two. To get the advantages of exploiting inter-sensor correlation about JSM-2, we should solve the Eq. **Error! Reference source not found.** and Eq. **Error! Reference source not found.** jointly. Like the FC in JSM-1, the FC in JSM-2 can exploit the fact that the support set is shared. By solving the MMV jointly, we obtain several benefits as high reconstruction probability on same number of measurement. If we solve those two equations separately, but not jointly, it is separate recovery. As an algorithm for solving the equation of the JSM-2 signal, we use a simultaneous OMP (SOMP) modified from an OMP algorithm for joint decoding and apply OMP for separate decoding. These algorithms are introduced in **Appendix 7.2** and **7.3** correspondingly.

iv) Joint vs. separate recovery performance for JSM-1 and JSM-2

Now, we compare the results of joint recovery and separate recovery. In joint recovery, if a correlation exists between the signals observed from the distributed sensors, the FC can use the correlated information to recover the transmitted signals. In separate recovery, correlated information is not used regardless of whether a correlation pattern exists between the observed signals. In **Figure 19**, solid lines were obtained from joint reconstructions, whereas dotted lines are the results of separate reconstructions.

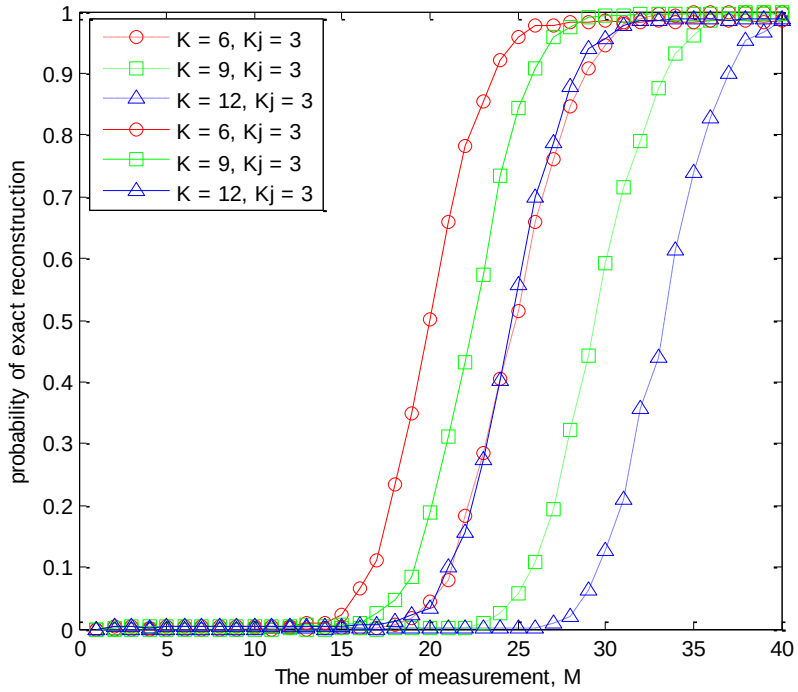


Figure 19. Joint (solid line) and separate (dotted line) reconstruction using PDIP algorithm for JSM-1. System parameters: $N = 50$, $J = 2$. The benefits of joint reconstruction depend on the common sparsity k_c.

When we use separate reconstruction, we cannot obtain any benefits from correlated information. However, when we use joint reconstruction, we can reduce the measurement size. For example, in **Figure 20**, the required number of measurements is almost 40 (dashed line and circles, $k = 6$) for perfect reconstruction when we use separate reconstruction. On the other hand, when we use joint reconstruction, it decreases to around 30 (solid line and circles, $k = 6$). Furthermore, as the common sparsity increases, the performance gap increases. For example, when the common sparsity is 9, joint reconstruction has a 90% probability of recovering all the signals at $m = 30$. However, the probability that separate reconstruction can recover all the signals is only 70%. **Figure 19** also shows that joint reconstruction is superior to separate reconstruction. For example, we need at least 30 measurements for reliable recovery

using separate reconstruction. However, we merely need at least 25 measurements for reliable recovery using joint reconstruction.

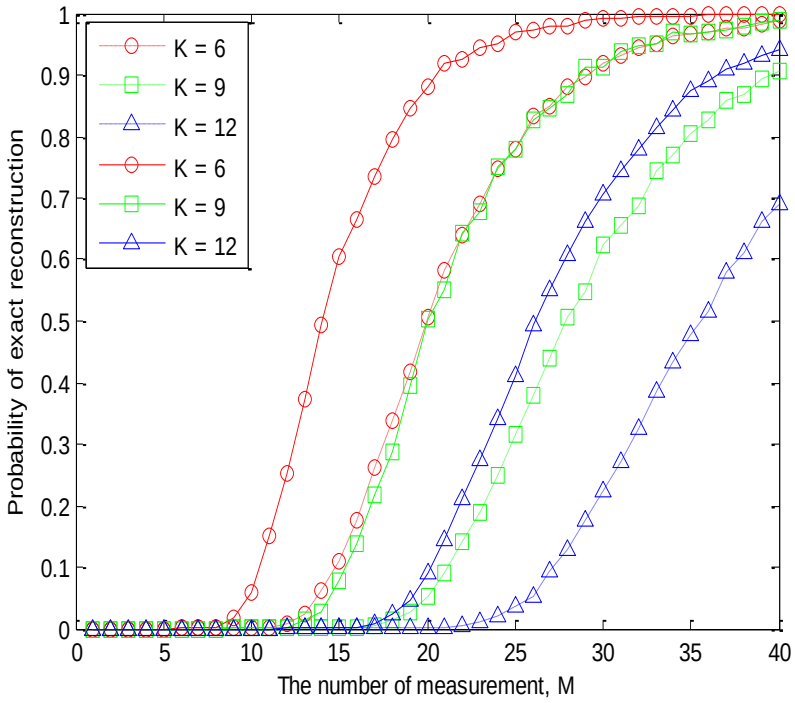


Figure 20. Joint (solid line) and separate (dotted line) reconstruction using SOMP for **JSM-2**. System parameters: $N = 50$, $J = 2$. Joint reconstruction has a higher probability of success than separate reconstruction.

5.2 Phased-Orthogonal matching pursuit (POMP)

In previous section, we discussed joint decoding methods for specific correlated signal, JSM-1 and JSM-2. The joint decoding method for JSM-1 cannot apply for JSM-2 which shares same support location only, since JSM-2 does not have same nonzero value. Therefore, if we use same idea for JSM-2, we cannot get the advantages of exploiting inter-sensor correlation. In reverse, SOMP, joint decoding algorithm for JSM-2, cannot apply for JSM-1 which has a large number of innovation sparsity. If we use SOMP algorithm for JSM-1, it may not find solution exactly. In summary, those two methods cannot apply all of the correlated signals which have various correlated pattern. To get exact solution of various correlated signals, we proposed joint decoding algorithm. The proposed algorithm is called phased orthogonal matching pursuit (POMP). POMP has better performance about the exact reconstruction probability of correlated signals than previous algorithms, for examples, PDIP, SOMP, ReMBo, etc. We will introduce the idea of our proposed algorithm in detail and then compare the reconstruction performance of our algorithms with previous algorithms.

i) Previous algorithm for MMV

1) One-step greedy algorithm

Figure 21 plots the probability of success in recovering the support set by using the one-step greedy algorithm (OSGA). OSGA finds common support location by using method described in **Table 9**, for the JSM-2 signal. In comparison with other greedy algorithms, it finds all the nonzero location at once so its performance is lower than that of SOMP.

The result of **Figure 21** suggests that the number of required measurements decreases for the same probability of exact reconstruction as the number of sensors increases. The OSGA works for a small

number of measurements m if the number of sensors is sufficiently large. Therefore, if many distributed sensors observe a correlated signal, each sensor is enough to send only a small number of compressed signals to achieve perfect reconstruction probability. Consequently, the transmission power of each sensor can be reduced because only the traffic volume required for exact reconstruction, which decreases significantly, must be transmitted. However, OSGA works poorly when there are fewer sensors, so it is not good method finding correlated signals. The OSGA is described in more details in **Error! Reference source not found.**.

The one-step greedy algorithm (OSGA):

1. Make greedy choice: Given all of the measurements, compute the test statistics

$$\varepsilon_n = \frac{1}{J}\sum_{j=1}^{J}\left\langle y_j, \phi_{j,n}\right\rangle^2$$

for $n \in \{1,2,...,N\}$ and estimate the common coefficient support set by

$\hat{\Omega} = \{n$ having one of the k largest $\varepsilon_n\}$.

Table 9. The one-step greedy algorithm (OSGA).

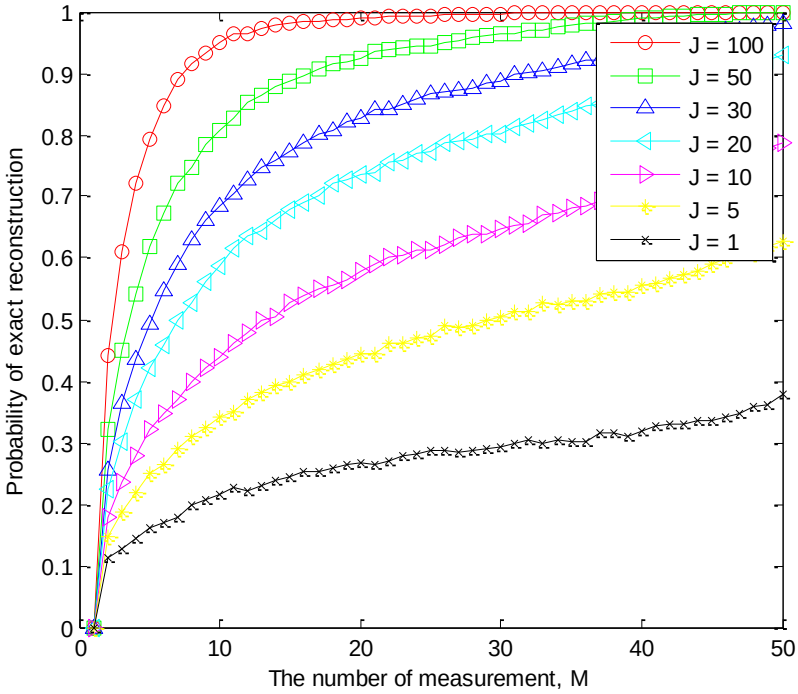
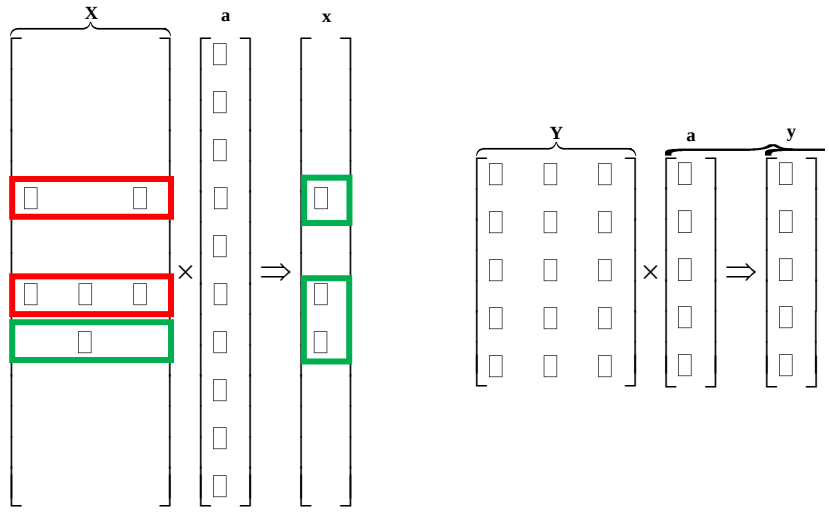


Figure 21. Reconstruction using OSGA for **JSM-2**. Experimental probability of success in recovering the support set in **JSM-2**. Signal length and sparsity are $n = 50$ and $k = 5$, respectively.

2) Reduced and boost algorithm

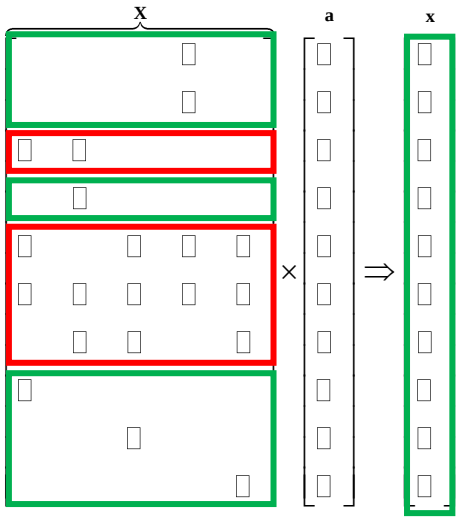
The Reduced and boost algorithm (ReMBo) is introduced in [Ref]. They reduce the correlated signal matrix $\mathbf{X}$ to one column signal and then solve reduced SMV problem by using greedy or gradient algorithm. The algorithm is summarized in Appendix 7.4. To get solution of MMV, ReMBo makes $\mathbf{Y} = \mathbf{AX}$ to $\mathbf{y} = \mathbf{Ax}$ by multiplying randomly generated vector $\mathbf{a}$ as **Figure 22**.



□ : Different value

Figure 22. Reduce MMV to SMV in ReMBo

The reduced SMV can be solved by using any one of SMV algorithm and then ReMBo algorithm saves the support location of SMV solution. From the information of support set, they can get the exact solution of matrix $\mathbf{X}$. This algorithm has easy stage for understanding and its algorithm speed is fast and effective. However, it is possible to apply only Eq. **Error! Reference source not found.** not **Error! Reference source not found.**, and if the original matrix $\mathbf{X}$ has much number of distributed innovation part as **Figure 23**, it cannot find solution. In the case of **Figure 23**, it makes the reduced signal $\mathbf{x}$ which is not sparse. The transformed equation $\mathbf{y} = \mathbf{A}\mathbf{x}$ from the signal of **Figure 23** cannot be solved. Therefore, it has limitations to apply various correlated signals.



□ : Different value

Figure 23. The limition of ReMBo algorithm

ii) Phased-OMP algorithm

Until now, we discussed various methods for recovering correlated signals. They are modified equation method, SOMP, OSGA, ReMBo. Those methods can apply only specific correlated signals as JSM-1, JSM-2 which have fixed pattern. If those methods are applied to other various signal models, it would not work properly. Thus, we proposed one method for any kind of correlated signals. It is called to phased-OMP (POMP). In this section, we explain how the algorithm works to recover various correlated signals. To help understanding algorithm, we will use the terminologies which are mentioned in previous section.

1) Basic idea of correlated signal recovery

We already talked about the unique solution of SMV problem before. The condition satisfied for solving SMV equation is $k < \frac{spark(A)}{2}$. It is proved in [Ref]. We used this proof as an idea for making

our proposed algorithm. If the total sparsity made from matrix X satisfies $T < \frac{spark(\mathbf{A})}{2}$, it guarantees each column has unique solution. Otherwise, even though each column satisfies $k < \frac{spark(A)}{2}$, it doesn't mean that $T < \frac{spark(\mathbf{A})}{2}$ is satisfied due to distributed innovation part

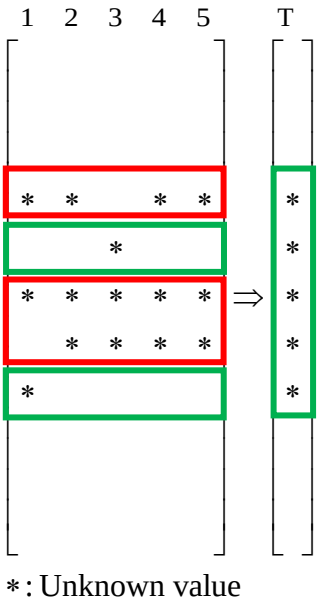


Figure 24. Total sparsity of MMV equation

Consider this specific signal which has only common sparsity and its total sparsity satisfy $T < \frac{spark(\mathbf{A})}{2}$. Because each column in MMV also satisfies $k < \frac{spark(A)}{2}$, and then we can get unique solution and its support location by using separate decoding. From the support set information of first SMV problem, we can also solve next SMV problem easily by using pseudo-inverse, if the next support set also has same support set. Therefore, it is important to know common part information since exact common part can reduce calculation time and complexity.

The idea for proposed algorithm follows. Common sparsity is a correlated part. Therefore, if we find the location of common part, we can also use it to solve another SMV. Innovation sparsity is an uncorrelated part. Therefore, even if we know the location of innovation part, we cannot use it to recover the innovation of other signals. In SMV problem, if we want to find the unique solution, each SMV should satisfy the unique condition. However, if we use correlated sparsity information in MMV equation, each SMV problem can get more guarantee for exact solution by using correlated information.

To use correlated information in POMP, we will find the common support location at first by using joint decoding. Then, by using separate decoding, we will find the remaining support location for each SMV problem. Therefore, we use the specific characters of both the joint decoding and separate decoding for effective reconstruction. Although the proposed algorithm works with easily understanding, its performance is better than previous methods (Modified equation method, POMP, SOMP) for correlated signal reconstruction. In addition, it doesn't be related with the number of total sparsity.

2) Pseudo-code of POMP

POMP uses the method which finds support set at every iteration. It is similar with the OMP method about finding support set, but we applied many ideas different with original OMP. We illustrate the pseudo code of POMP algorithm as following table.

Input	Output
A $m \times n$ measurement matrix $\mathbf{A}_j$ A $m-$dimensional data vector $\mathbf{y}_j$ The sparsity level k of the ideal signal The estimate number of common sparisty C Stop conditon ε	An estimate $\hat{\mathbf{x}}_j$ in $\mathbf{R}^n$ for the ideal signal. A set $\Lambda_{j,k}$ containing k elements from $\{1,...,n\}$ An $m-$dimensional approximation $\hat{\mathbf{y}}_{j,k}$ of the data $\mathbf{y}_j$ An $m-$dimensional residual $\mathbf{r}_{j,k} = \mathbf{y}_j - \hat{\mathbf{y}}_{j,k}$

Table 4. Inputs and outputs of SOMP algorithm.

The POMP algorithm:

Phase 1: For find common sparsity

1. Initialize:

Let the residual matrix be $\mathbf{r}_{j,0} = \mathbf{y}_{j,0}$. The sparse set $\Lambda_{j,0} = \{\}$, and iteration number $t = 1$.

2. Find the common sparsity index $\lambda_{j,t}$ for each j:

$\lambda_{j,t} = \arg\max_{i=1,...,n} \sum_{j=1}^{J} \left| \left\langle \mathbf{r}_{j,t-1}, \mathbf{a}_{j,i} \right\rangle \right|$. The $\mathbf{a}_{j,i}$ is the i th column vector of matrix $\mathbf{A}_j$.

3. Update set:

$\Lambda_{j,t} = \Lambda_{j,t-1} \cup \{\lambda_{j,t}\}$.

4. Signal estimate:

$\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right) = \mathbf{A}_{j,\Lambda_t}^{\dagger} \mathbf{y}_j$ and $\mathbf{x}_{j,t}\left(\Lambda_{j,t}^{C}\right) = \mathbf{0}$, where $\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)$ is the set of elements whose indices are corresponding to the sparse set.

5. Get new residual:

$\hat{\mathbf{y}}_{j,t} = \mathbf{A}_{j,t}\mathbf{x}_{j,t}$, $\mathbf{r}_{j,t} = \mathbf{y}_j - \hat{\mathbf{y}}_{j,t}$.

6. Increment t:

Increase iteration number $t = t+1$, and

if $t < C$ return to **Step 2 of Phase 1**

otherwise, $t > C$ go to **Phase 2**

Phase 2: For find innovation sparsity for each j

7. Find the index $\lambda_{j,t}$ **for each** j **:**

$\lambda_{j,t} = \arg\max_{i=1,\ldots,n} \left| \left\langle \mathbf{r}_{j,t-1}, \mathbf{a}_{j,i} \right\rangle \right|$ for every j. The $\mathbf{a}_{j,i}$ is the i th column vector of matrix $\mathbf{A}_j$.

8. Update set:

$\Lambda_{j,t} = \Lambda_{j,t-1} \cup \left\{ \lambda_{j,t} \right\}$ for each j.

9. Signal estimate:

$\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right) = \mathbf{A}_{j,\Lambda_t}^{\dagger} \mathbf{y}_j$ and $\mathbf{x}_{j,t}\left(\Lambda_{j,t}^{C}\right) = \mathbf{0}$, where $\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)$ is the set of elements whose indices are corresponding to the sparse set.

10. Get new residual:

$\hat{\mathbf{y}}_{j,t} = \mathbf{A}_{j,t}\mathbf{x}_{j,t}$, $\mathbf{r}_{j,t} = \mathbf{y}_j - \hat{\mathbf{y}}_{j,t}$.

11. Increment t **:**

Increase iteration number $t = t + 1$, and

return to **Step 7 of phase 2** if $\sum_{j=1}^{J} \left\| \mathbf{y}_j - \mathbf{A}_j \mathbf{x}_j \right\|_2 > \varepsilon$

otherwise stop the algorithm.

Table 5. POMP algorithm.

For **Phase 1**, it is a stage for finding common sparsity Because the common sparsity is the correlated part of the signal matrix $\mathbf{X}$, we use joint decoding method for finding the location of common part. The joint decoding method is able to find support location successfully. We already knew the advantages of joint decoding from the comparison of joint decoding and separate decoding in **Section 5.1**. Therefore, if it is possible to use joint decoding for MMV equation, we should use it for advantages about signal reconstruction. It results in better performance for solving MMV equation. According to pseudo-code of POMP, it finds the location of common part in **Phase 1** and memorizes the index as **Figure 26**. After the stage of **Phase 1** is finished, POMP algorithm tries to find the

remaining support set by separate decoding. Due to separate decoding of **Phase 2** for remaining support set, POMP can find the missed common sparsity in previous stage.

We draw **Figure 26** which expresses the movement of POMP algorithm. The nonzero values in red box are common sparsity which is exploited in **Phase 1** and then the remaining nonzero values in green box can be exploited in **Phase 2**. The index of row having nonzero values is added to Λ at every iteration until the criteria $\sum_{j=1}^{J}\left\|\mathbf{y}_j - \mathbf{A}_j\mathbf{x}_j\right\|_2 \le \varepsilon$ is satisfied. After finishing the movement of POMP, we can get the original solution by using pseudo-inverse based on the estimated support set.

Support set in **Phase 1**:

j = 1, {3, 5, 6, 7}

j = 2, {3, 5, 6, 7}

j = 3, {3, 5, 6, 7}

j = 4, {3, 5, 6, 7}

j = 5, {3, 5, 6, 7}

Support set in **Phase 2**:

j = 1, {3, 5, 6, 7, 8}

j = 2, {3, 4, 5, 6, 7}

j = 3, {3, 5, 6, 7, 9}

j = 4, {1, 2, 3, 5, 6, 7}

j = 5, {3, 5, 6, 7, 10}

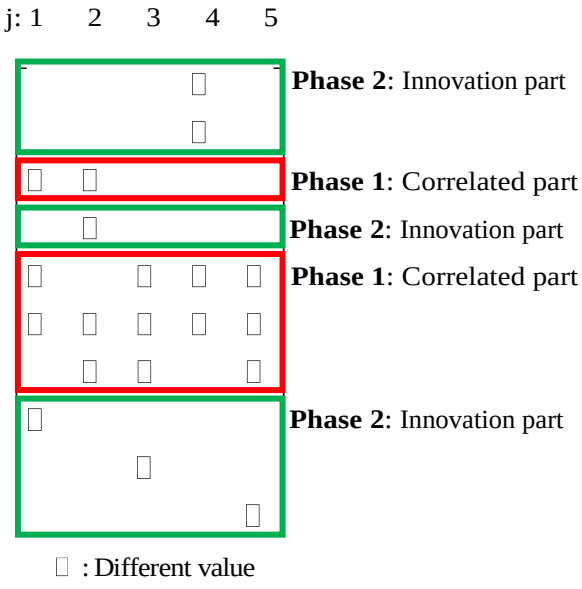


Figure 25. The movement of POMP algorithm

Moore-Penrose pseudo inverse:

If we define that Λ_j is the support set of the j th column in matrix $\mathbf{X}$, we can reduce the sensing matrix $\mathbf{A}_j$ to $\mathbf{A}_{\Lambda_j}$ corresponding to the nonzero elements of $\mathbf{x}_j$. If the columns of the reduced matrix $\mathbf{A}_{\Lambda_j}$ are linearly independent, Moore-Penrose pseudo inverse equation is accepted.

$$\left(\mathbf{A}_{\Lambda_j}\right)^{\dagger}\mathbf{A}_{\Lambda_j}=\mathbf{I}\text{, where }\ \mathbf{A}_{\Lambda_j}^{\dagger}=\left(\mathbf{A}_{\Lambda_j}^{T}\mathbf{A}_{\Lambda_j}\right)^{-1}\mathbf{A}_{\Lambda_j}^{T}$$

Therefore, if we know the support set and the reduced matrix $\mathbf{A}_{\Lambda_j}$ are linearly independent, then the original signal $\mathbf{x}_j$ can be found by using pseudo-inverse.

$$\mathbf{x}_{\Lambda_j}=\left(\mathbf{A}_{\Lambda_j}^{T}\mathbf{A}_{\Lambda_j}\right)^{-1}\mathbf{A}_{\Lambda_j}^{T}\mathbf{y}$$

Table 9. **Moore-Penrose** pseudo inverse.

5.3 The properties of POMP algorithm

i) The advantages of using prior correlation information

If we know the prior information of correlated signal like the number of common sparstiy, innovation sparsity, or the distribution of support location, we can use that information for signal recovery. If we know the number of common sparsity as prior information, we can choice parameter C as the number of iteration used for finding common part exactly. To select the number of iteration exactly in POMP affects the reconstruction performance as **Figure 25**. The parameters of simulation are $N=150$, $C=10$, $I=10$. Even though the value of estimated C is not exact correct as red and or blue, its performance is stable. However, it requires much number of measurements for perfect signal recovery.

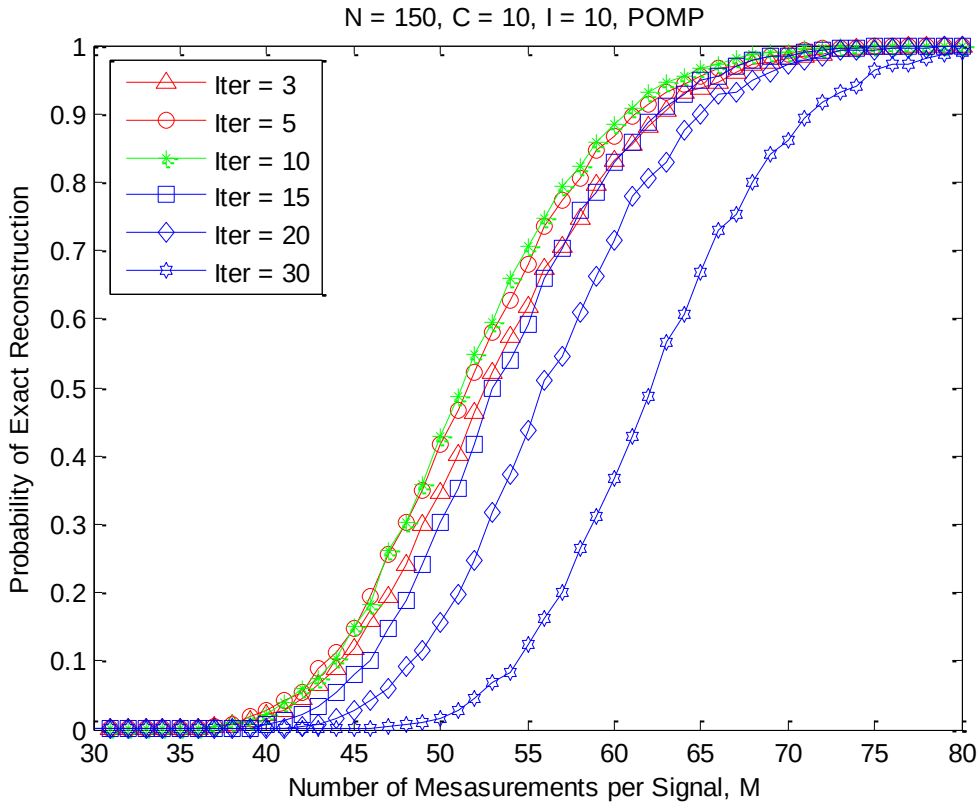


Figure 25. The advantages of using prior information

ii) The complexity of POMP

1) The complexity of Moore-Penrose pseudo inverse

We express how POMP works for MMV equation throughout pseudo-code. According to POMP algorithm, it requires pseudo-inverse calculation to get unique solution when we know the support set of original signal. In general, the calculation of matrix multiplication and inverse matrix require high complexity respectively $O\left(n^2\right)$ and $O\left(n^3\right)$ respectively. However, we already know that the number of sparsity k is very short in comparison with the length of signal n as $k << m < n$ so, the complexity of those calculations is simple as below **Table**.

The complexity of pseudo-inverse:
By using pseudo-inverse, we can get the nonzero values corresponding to support set. The pseudo inverse complexity is not high. From the relation $k << m < n$, we can get the complexity.
$\underset{k\times m\ m\times k}{\left(A_S^T A_S\right)^{-1}} \underset{k\times m\ m\times 1}{A_S^T y} = x_S$
Multiplication: $k\times m\times k + k\times k\times m + k\times m\times 1 \approx 2k^2 m + km \approx O\left(mk^2\right)$
Inverse: $O\left(k^3\right)$. Therefore, total:$O\left(k^3\right)+O\left(mk^2\right)\approx O\left(mk^2\right)$ $\because k << m$

Table 9. Pseudo-inverse in MMV.

2) The complexity of POMP algorithm in terms of sparsity.

POMP algorithm has two stages for find support location which consists of common sparsity and innovation sparsity. In this section, we analyze the complexity of POMP algorithm related with sparsity and the number of sensors J . We assumed that the observed signal has both the C number of common sparstiy and the I number of innovation sparsity. In **Phase 1**, it will find the C number of sparsity and it requires such as calculations, for examples $\lambda_{j,t} = \underset{i=1,\ldots,n}{\arg\max} \sum_{j=1}^{J} \left|\left\langle \mathbf{r}_{j,t-1}, \mathbf{a}_{j,i} \right\rangle\right|$ and

$\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)=\mathbf{A}_{j,\Lambda_t}^{\dagger}\mathbf{y}_j$. After finishing the calculations of **Phase 1**, it starts the calculation of **Phase 2** for finding innovation sparsity. We consider $\lambda_{j,t}=\arg\max_{i=1,\ldots,n}\left|\left\langle \mathbf{r}_{j,t-1},\mathbf{a}_{j,i}\right\rangle\right|$ and $\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)=\mathbf{A}_{j,\Lambda_t}^{\dagger}\mathbf{y}_j$ in **Phase 2**. The complexity of POMP algorithm is below.

The complexity of POMP

Phase 1

Considered parameter: signal length n, measurement m, sparsity k, the number of sensors J, common sparstiy C, innovation sparsity I. We already know the relationship $(C\approx I)<k<<m<n$

1) $\lambda_{j,t}=\arg\max_{i=1,\ldots,n}\underbrace{\sum_{j=1}^{J}\left|\left\langle \underset{1\times m}{\mathbf{r}_{j,t-1}},\underset{m\times 1}{\mathbf{a}_{j,i}}\right\rangle\right|}_{J\text{ times summation}}$

Inner product and sigma summation: $O(m)+O(J)\approx O(m)$ ∵ In general $J<m$

By $J\times n$ iteration: $JnO(m)=O(Jnm)$

2) $\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)=\mathbf{A}_{j,\Lambda_t}^{\dagger}\mathbf{y}_j$

Pseudo-inverse: $O\left(mk^2\right)$

3) The number of iteration: C

Therefore, $C\left(O(Jmn)+O\left(mk^2\right)\right)\approx O(CJmn)$ ∵ In general $k^2<n$

In conclusion, the complexity of **Phase 1** is $O(CJmn)$

Phase 2

1) $\lambda_{j,t}=\arg\max_{i=1,\ldots,n}\left|\left\langle \mathbf{r}_{j,t-1},\mathbf{a}_{j,i}\right\rangle\right|$

Inner product : $O(m)$

By $J\times n$ iteration: $JnO(m)=O(Jnm)$

2) $\mathbf{x}_{j,t}\left(\Lambda_{j,t}\right)=\mathbf{A}_{j,\Lambda_t}^{\dagger}\mathbf{y}_j$

Pseudo-inverse: $O(mk^2)$

3) The number of iteration: I

Therefore, $I\left(O(Jmn)+O(mk^2)\right)\approx O(IJmn)$ $\because$ In general $k^2 < n$

In conclusion, the complexity of **Phase 2** is $O(IJmn)$

By **Phase 1** + **Phase 2**,

In conclusion, the complexity of POMP is $O(CJmn)+O(IJmn)=O\left(Jmn(C+I)\right)\approx O(Jmnk)$.

The complexity of POMP algorithm is affected by the parameters J, m, n, k.

Table 9. The complexity of POMP algorithm

iii) The recovery condition of POMP

6. Performance evaluation

In previous section, we already discuss the various correlated signals which are handled with references. For distinguishing those correlated signals, we named it correlated signal model (CSM) as following **Figure 27** and then we solved MMV equation **Error! Reference source not found.** and **Error! Reference source not found.** by using algorithms like modified equation method (MEM), SOMP, ReMBo, POMP. All of the algorithms are handled in previous section.

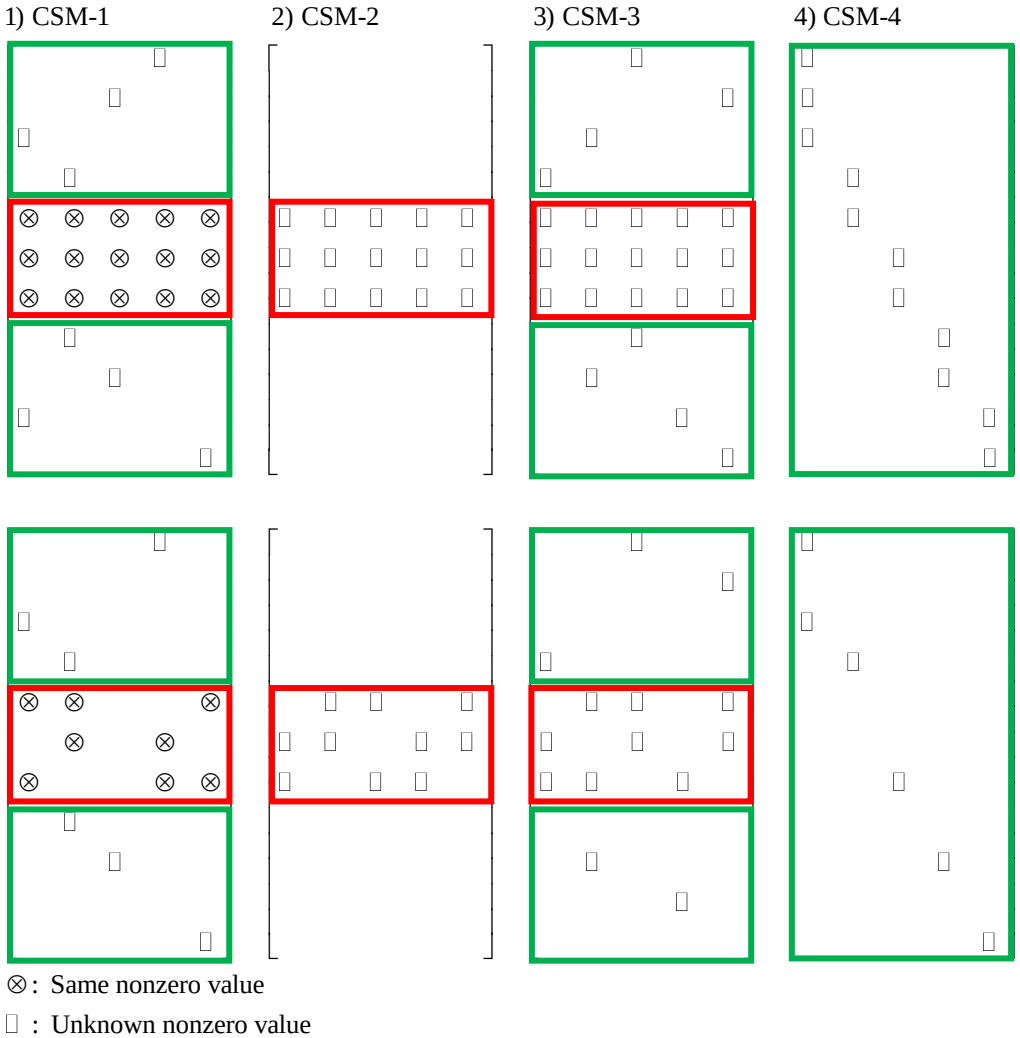
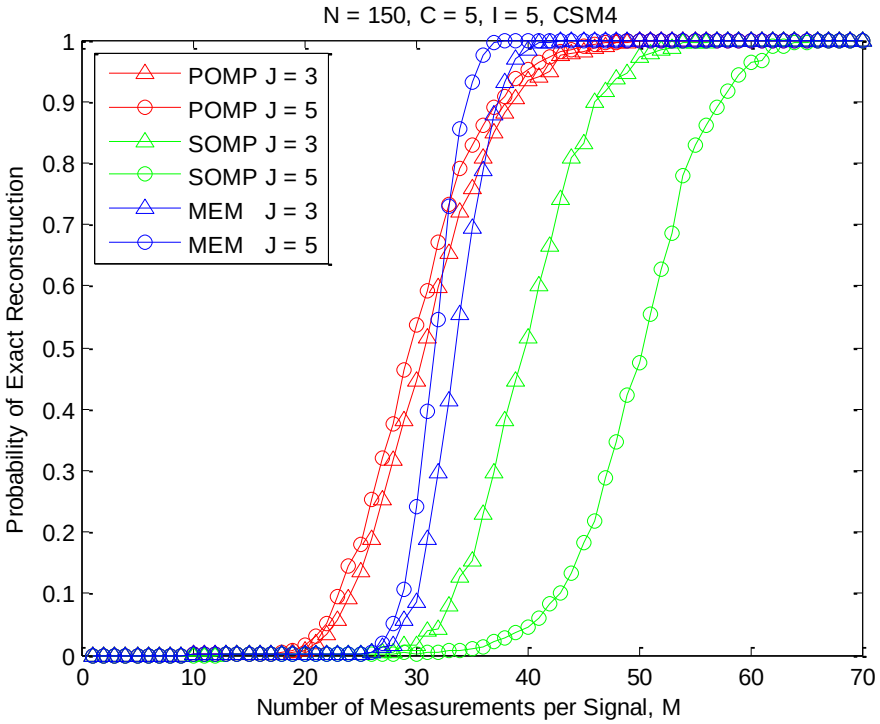


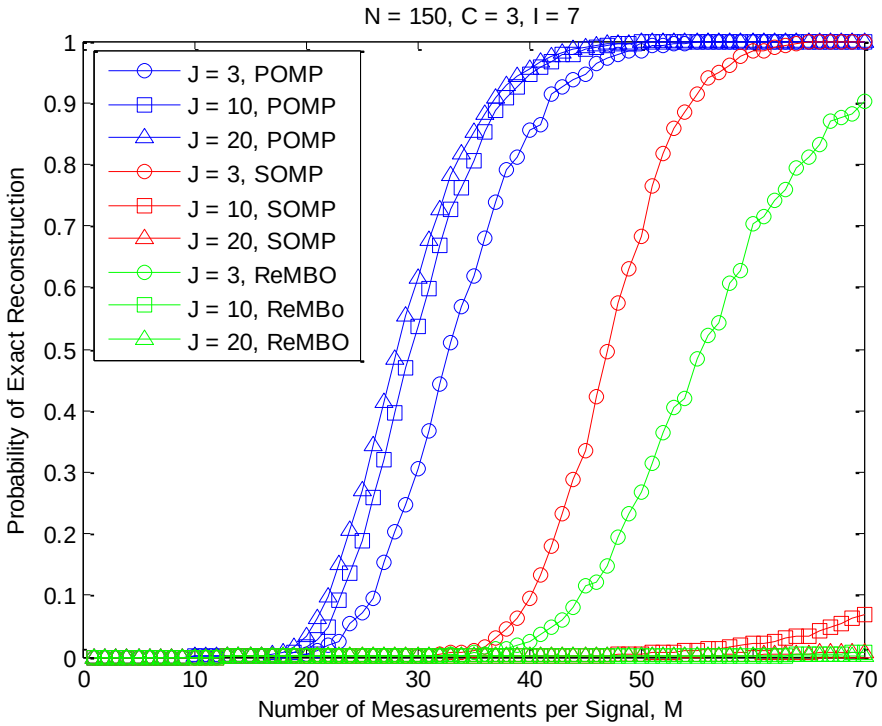
Figure 27. Correlated signal model (CSM)

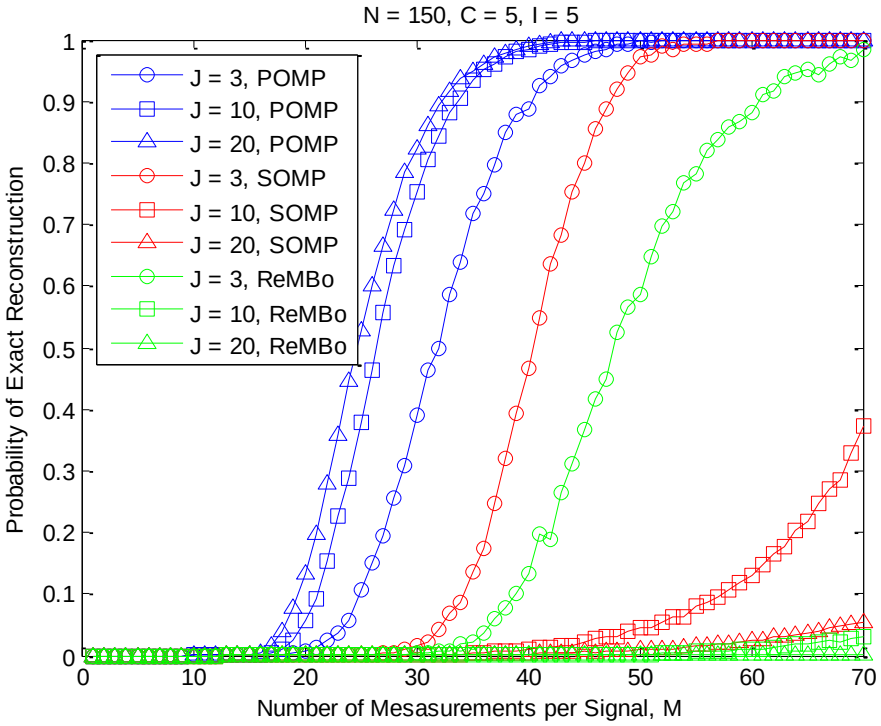
In **Figure 27**, we define correlated signal model which has various kinds of pattern. Most of the signal pattern can be defined by our CSM. CSM-1 and CSM-2 is similar with JSM-1 and JSM-2 which are handled in [Ref] but we consider the case that has some vacancies in common part and innovation part is also same signal model and we changed the definition of innovation part. In our case, innovation sparsity exist only one nonzero value in same row in different with [Ref]. CSM-1 and CSM-3 have common part and innovation part together but CSM-1 has same value for common part. CSM-2 has only common part and CSM-4 has only innovation part. Other case which does not exist in **Figure 27** will not be considered in this paper. Now, we simulate the performance POMP algorithm compared with other methods as ReMbo, SOMP, MEM.

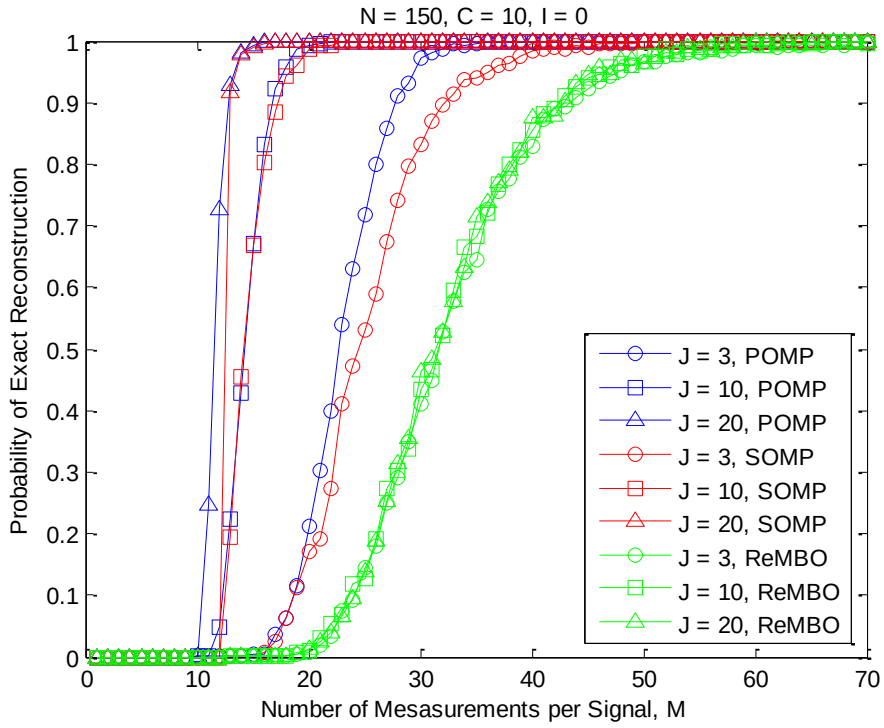
1) CSM-1 and Different matrix **A**

We generated the CSM-1 which has common and innovation part. Its common part has same values. We observed the reconstruction performance for MEM, SOMP, and POMP algorithm when the signal has CSM-1 pattern and each sensing matrix is different.


N = 150, C = 5, I = 5, CSM4
POMP J = 3
POMP J = 5
SOMP J = 3
SOMP J = 5
MEM J = 3
MEM J = 5
Probability of Exact Reconstruction
Number of Mesasurements per Signal, M
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70


N = 150, C = 3, I = 7
J = 3, POMP
J = 10, POMP
J = 20, POMP
J = 3, SOMP
J = 10, SOMP
J = 20, SOMP
J = 3, ReMBO
J = 10, ReMBo
J = 20, ReMBO
Probability of Exact Reconstruction
Number of Mesasurements per Signal, M
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70


N = 150, C = 5, I = 5
Probability of Exact Reconstruction
Number of Mesasurements per Signal, M
J = 3, POMP
J = 10, POMP
J = 20, POMP
J = 3, SOMP
J = 10, SOMP
J = 20, SOMP
J = 3, ReMBo
J = 10, ReMBo
J = 20, ReMBo


N = 150, C = 10, I = 0
Probability of Exact Reconstruction
Number of Mesasurements per Signal, M
J = 3, POMP
J = 10, POMP
J = 20, POMP
J = 3, SOMP
J = 10, SOMP
J = 20, SOMP
J = 3, ReMBO
J = 10, ReMBO
J = 20, ReMBO

7. Conclusion

In this chapter, we discussed the application of compressive sensing (CS) for wireless sensor networks (WSNs). We assumed a WSN consisting of spatially distributed sensors and one fusion center (FC). The sensor nodes take signal samples and pass their acquired signal samples to the FC. When the FC receives the transmitted data from the sensor nodes, it aims to recover the original signal waveforms, for later identification of the events possibly occurring in the sensed region. (Section 2.1)

We discussed that CS is the possible solution which provides simpler signal acquisition and compression. CS is suitable for the wireless sensor networks since it allows removal of intermediate stages such as sampling the signal and gathering the sampled signals at one collaboration point which would usually be the case in a conventional compression scheme. Using CS, the amount of signal samples that need to be transferred to the FC from the sensors can be significantly reduced. This may lead to reduction of power consumption at the sensor nodes, which was discussed in Section 4.1. In summary, each sensor with CS can save power by not needing to run complex compression operations on board and by cutting down signal transmissions.

Distributed sensors usually observe a single globally occurring event and thus the observed signals are often correlated with each other. We considered two types of correlations: intra- and inter-sensor signal correlation. We provided the sparse signal models which encompass both types of correlation in Sections 4.2 and 4.3.

The FC receives the compressed signals from the sensors. The FC then recovers the original signal waveforms from the compressed signals using a CS recovery algorithm. We considered two types of algorithms. One is a greedy algorithm type, which includes the orthogonal matching pursuit (OMP) and the simultaneous orthogonal matching pursuit (SOMP) algorithms, discussed in Section

Error! Reference source not found.. The other is a gradient type for which we used the primal-dual interior point (PDIP) method, in Section **Error! Reference source not found.**.

Finally, we presented simulations results in which the CS based WSN system parameters such as the number of measurements, the sparsity, and the signal length were varied. We discussed the use of a joint recovery scheme at the FC. A CS recovery algorithm is referred to as the joint recovery scheme when it utilizes inter-sensor signal correlation as well. In contrast, when the inter-sensor signal correlation is not utilized, it is referred to as the separate recovery scheme. In the joint recovery scheme, inter-sensor signal correlation information is incorporated in the formation of recovery equation as shown Eq. **Error! Reference source not found.** and **Error! Reference source not found.**. In the separate recovery scheme, a sensor signal recovery is done individually and independently from the recovery of other sensor signals. We compared the results of the joint recovery with those of the separate recovery scheme. We have shown that correlation information can be exploited and the number of measurements needed for exact reconstruction can be significantly reduced as shown in Figure 14. It means that the traffic volume transmitted from the sensors to the FC can decrease significantly without degrading the quality of the recovery performance. (Section **Error! Reference source not found.**)

We have shown that the CS is an efficient and effective signal acquisition and sampling framework for WSN which can be used to save transmittal and computational power significantly at the sensor node. This CS based signal acquisition and compression scheme is very simple, so it is suitable for inexpensive sensors. The number of compressed samples required for transmission from each sensor to the FC is significantly small, which makes it perfect for sensors whose operational power is drawn from onboard battery. Finally, the joint CS recovery at the FC exploits signal correlation and enables Distributed Compressive Sensing.

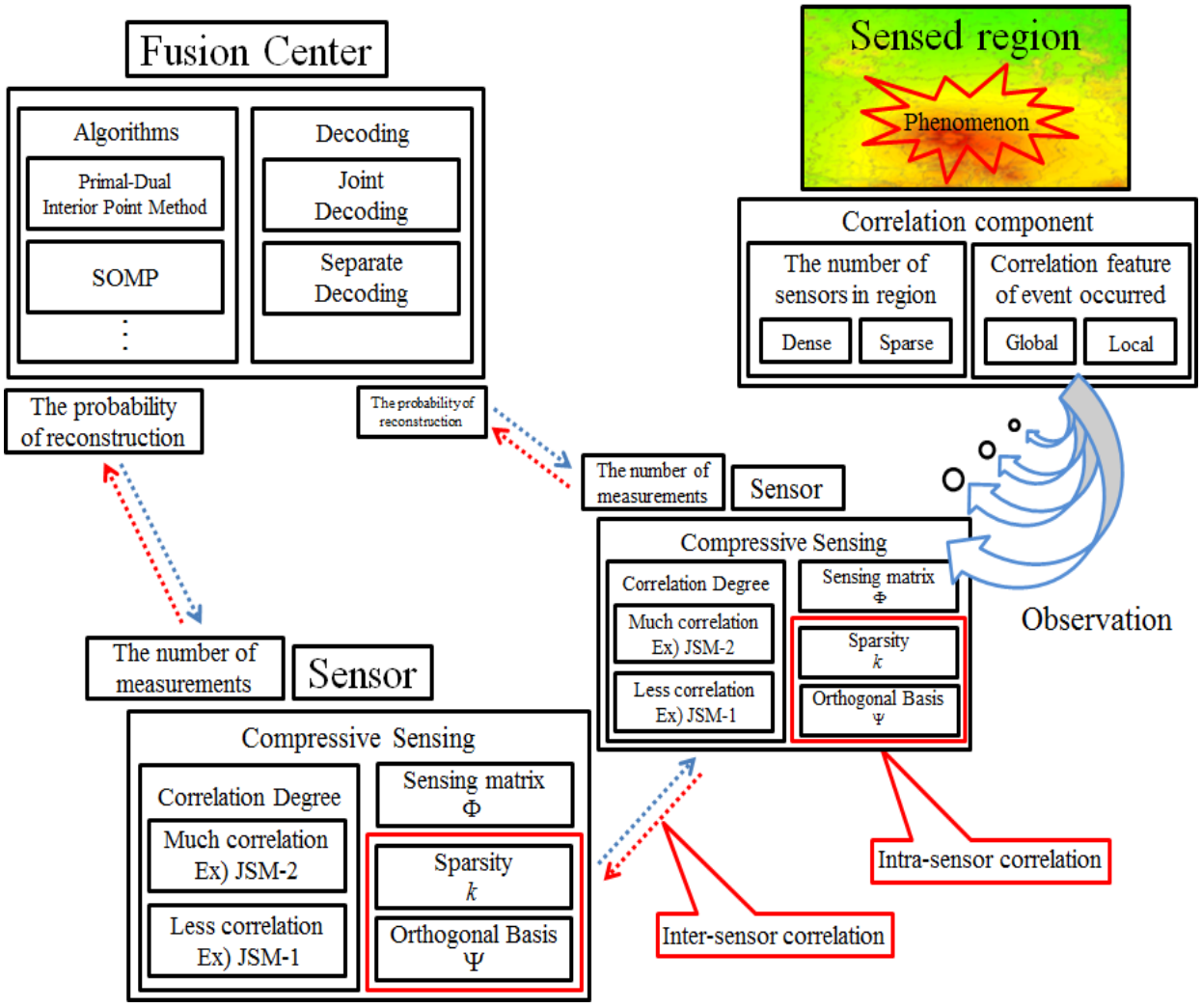


Figure 16. Summary of CS application in WSN

8. Appendix

8.1. Primal-dual interior point method (PDIP)

The L_1 minimization in Eq. **Error! Reference source not found.** can be recast as linear programming. Here we examine this relationship. Clearly, the L_1 minimization problem in Eq. **Error! Reference source not found.** is not linear programming because its cost function is not linear. However, by using a new variable, we can transform it to linear programming. Thus, the problem that we want to solve is

$$\begin{aligned} &\min_{(x,u)} \sum_i u_i \\ &\text{subject to} \\ &\forall_i \left| x(i) \right| \le u_i \\ &\mathbf{Ax} = \mathbf{b} \end{aligned} \tag{24}$$

The solution of the above equation is equal to the solution of the L_1 minimization problem. Many approaches to solving Eq. **Error! Reference source not found.** have been studied and developed. Here, we discuss the primal-dual interior point (PDIP) method, which is an example of gradient-type algorithms. First, we have the Lagrangian function of Eq. **Error! Reference source not found.**, as follows:

$$L(\mathbf{t}, \boldsymbol{\lambda}, \mathbf{v}) = \begin{bmatrix} \mathbf{0}_1^{\mathrm{T}} & \mathbf{1}^{\mathrm{T}} \end{bmatrix} \mathbf{t} + \mathbf{v}^{\mathrm{T}} \left(\begin{bmatrix} \mathbf{A} & \mathbf{0}_2 \end{bmatrix} \mathbf{t} - \mathbf{b} \right) + \boldsymbol{\lambda}^{\mathrm{T}} \left(\begin{bmatrix} \mathbf{e} & -\mathbf{e} \\ -\mathbf{e} & -\mathbf{e} \end{bmatrix} \mathbf{t} \right) \tag{25}$$

where $\mathbf{e}$ is the $n\times n$ identity matrix, $\mathbf{0}_1$ is the zero vector, $\mathbf{0}_2$ is the $m\times n$ zero vector, and $\mathbf{1}$ is the $n\times 1$ vector whose elements are all one, $\mathbf{t}:=\begin{bmatrix}\mathbf{x}\\ \mathbf{u}\end{bmatrix}\in\mathbf{R}^{2n\times 1}$, $\mathbf{v}\in\mathbf{R}^{m\times 1}$, and $\boldsymbol{\lambda}\in\mathbf{R}^{2n\times 1}\geq 0$. From the Lagrangian function, we have several KKT conditions,

$$\begin{aligned}
&\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}+\begin{bmatrix}\mathbf{A}^{\mathrm{T}}\\ \mathbf{0}^{\mathrm{T}}\end{bmatrix}\mathbf{v}^{*}+\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\boldsymbol{\lambda}^{*}=\mathbf{0}_3\\
&\begin{bmatrix}\mathbf{A} & \mathbf{0}_2\end{bmatrix}\mathbf{t}^{*}-\mathbf{b}=\mathbf{0}_4\\
&\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\mathbf{t}^{*}\leq\mathbf{0}_1\\
&\left(\boldsymbol{\lambda}^{*}\right)^{\mathrm{T}}\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\mathbf{t}^{*}=0,\boldsymbol{\lambda}^{*}\geq\mathbf{0}_3
\end{aligned}\qquad(26)$$

where $\mathbf{0}_3$ is the $2n\times 1$ zero vector, and $\mathbf{0}_4$ is the $m\times 1$ zero vector. The main point of the PDIP is to seek the point $\left(\mathbf{t}^{*},\boldsymbol{\lambda}^{*},\mathbf{v}^{*}\right)$ that satisfies the above KKT conditions. This is achieved by defining a mapping function $\mathrm{F}\left(\mathbf{t},\boldsymbol{\lambda},\mathbf{v}\right):\mathbf{R}^{(2n+m)\times 1}\rightarrow\mathbf{R}^{(2n+m)\times 1}$, which is

$$\mathrm{F}\left(\mathbf{t},\boldsymbol{\lambda},\mathbf{v}\right)=\begin{bmatrix}\begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}+\begin{bmatrix}\mathbf{A}^{\mathrm{T}}\\ \mathbf{0}^{\mathrm{T}}\end{bmatrix}\mathbf{v}+\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\boldsymbol{\lambda}\\ \left(\boldsymbol{\lambda}^{*}\right)^{\mathrm{T}}\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\mathbf{t}^{*}\\ \begin{bmatrix}\mathbf{A} & \mathbf{0}_2\end{bmatrix}\mathbf{t}-\mathbf{b}\end{bmatrix}=\mathbf{0}_4\in\mathbf{R}^{(2n+m)\times 1},\begin{bmatrix}\mathbf{e} & -\mathbf{e}\\ -\mathbf{e} & -\mathbf{e}\end{bmatrix}\mathbf{t}^{*}\leq\mathbf{0}_1,\boldsymbol{\lambda}^{*}\geq\mathbf{0}_3\qquad(27)$$

where $\mathbf{0}_4$ is the $(2n+1)\times 1$ zero vector. Now, we would like to find the point $(\mathbf{t}^*, \boldsymbol{\lambda}^*, \mathbf{v}^*)$ satisfying $\mathrm{F}(\mathbf{t}^*, \boldsymbol{\lambda}^*, \mathbf{v}^*) = \mathbf{0}_4$. Here, we use a linear approximation method. From the Taylor expansions of the function $\mathrm{F}(\mathbf{t}, \boldsymbol{\lambda}, \mathbf{v})$, we have

$$\mathrm{F}(\mathbf{t}+\Delta\mathbf{t}, \boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}, \mathbf{v}+\Delta\mathbf{v}) \approx \mathrm{F}(\mathbf{t}, \boldsymbol{\lambda}, \mathbf{v}) + \nabla_{(\mathbf{t},\boldsymbol{\lambda},\mathbf{v})}\mathrm{F}(\mathbf{t}, \boldsymbol{\lambda}, \mathbf{v}) \begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{v} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} \tag{28}$$

Thus, solving the above equations yields the direction $(\Delta\mathbf{t}, \Delta\mathbf{v}, \Delta\boldsymbol{\lambda})$. Next, we seek the proper step length along the direction that does not violate $\begin{bmatrix} \mathbf{e} & -\mathbf{e} \\ -\mathbf{e} & -\mathbf{e} \end{bmatrix}\mathbf{t}^* \le \mathbf{0}_1$ and $\boldsymbol{\lambda}^* \ge \mathbf{0}_3$. The pseudo code for the PDIP algorithm is shown in **Table 6**.

The primal-dual interior point method algorithm:

1. Initialize:

Choose $\mathbf{v}^0 \in \mathbf{R}^{m\times 1}$, $\boldsymbol{\lambda}^0 \ge \mathbf{0}_3$, and $\mathbf{t}^0 = \begin{bmatrix} \mathbf{x}^0 & \mathbf{u}^0 \end{bmatrix}^{\mathrm{T}}$, where $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$, and $\mathbf{u}^0 = |\mathbf{x}^0| + \alpha|\mathbf{x}^0|$ and iteration number $k = 1$. (The $\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the Moore-Penrose pseudo-inverse of $\mathbf{A}$ and $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$.)

2. Find the direction vectors $(\Delta\mathbf{t}, \Delta\mathbf{v}, \Delta\boldsymbol{\lambda})$:

$$\begin{bmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{v} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = -\left[\nabla_{(\mathbf{t}^k, \boldsymbol{\lambda}^k, \mathbf{v}^k)}\mathrm{F}(\mathbf{t}^k, \boldsymbol{\lambda}^k, \mathbf{v}^k)\right]^{-1} \mathrm{F}(\mathbf{t}^k, \boldsymbol{\lambda}^k, \mathbf{v}^k).$$

3. Find the proper step length:

Choose the largest α satisfying $\left\|\mathrm{F}(\mathbf{t}^k + \alpha, \boldsymbol{\lambda}^k + \alpha, \mathbf{v}^k + \alpha)\right\|_2^2 \le \left\|\mathrm{F}(\mathbf{t}^k, \boldsymbol{\lambda}^k, \mathbf{v}^k)\right\|_2^2$.

4. Update parameters:

$\mathbf{t}^{k+1} = \mathbf{t}^k + \alpha\Delta\mathbf{t}$, $\mathbf{v}^{k+1} = \mathbf{v}^k + \alpha\Delta\mathbf{v}$, $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha\Delta\boldsymbol{\lambda}$.

5. Update the signal:

$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{t}[1:n]$.

6. Increment the iteration number k **:**

Increase iteration number $k = k + 1$, and return to Step 2 if $\left\| \mathbf{y} - \mathbf{A}\mathbf{x}^k \right\|_2^2 > eps$.

Table 6. Primal-dual interior point method algorithm.

8.2 Orthogonal matching pursuit (OMP)

The orthogonal matching pursuit (OMP) is a famous greedy-type algorithm **Error! Reference source not found.**. OMP produces a solution within k steps because it adds one index to the sparse set Λ at each iteration. The strategy of OMP is outlined in **Tables 2** and **3**.

Input	Output
A $m \times n$ measurement matrix $\mathbf{A}$ A $m-$dimensional data vector $\mathbf{y}$ The sparsity level k of the ideal signal	An estimate $\hat{\mathbf{x}}$ in $\mathbf{R}^n$ for the ideal signal. A set Λ_k containing k elements from $\{1,...,n\}$ An $m-$dimensional approximation $\hat{\mathbf{y}}_k$ of the data $\mathbf{y}$ An $m-$dimensional residual $\mathbf{r}_k = \mathbf{y} - \hat{\mathbf{y}}_k$

Table 2. Inputs and outputs of OMP algorithm.

The OMP algorithm:
1. Initialize: Let the residual vector be $\mathbf{r}_0 = \mathbf{y}$, the sparse set $\Lambda_0 = \{\}$, and iteration number $t = 1$.
2. Find the index λ_t : $\lambda_t = \arg\max_{i=1,...,n} \left\| \langle \mathbf{r}_{t-1}, \mathbf{a}_i \rangle \right\|$. The $\mathbf{a}_i$ is the i th column vector of matrix $\mathbf{A}$.
3. Update set: $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$.
4. Signal estimate: $\mathbf{x}_t(\Lambda_t) = \mathbf{A}_{\Lambda_t}^{\dagger}\mathbf{y}$ and $\mathbf{x}_t(\Lambda_t^C) = \mathbf{0}$, where $\mathbf{x}_t(\Lambda_t)$ is the set of elements whose indices are corresponding to the sparse set.
5. Get new residual: $\hat{\mathbf{y}}_t = \mathbf{A}_t\mathbf{x}_t$, $\mathbf{r}_t = \mathbf{y} - \hat{\mathbf{y}}_t$.
6. Increment t : Increase iteration number $t = t+1$, and return to Step 2 if $t < k$.

Table 3. OMP algorithm.

Let us examine the above OMP algorithm. In step 2, OMP selects one index that has a dominant impact on the residual vector $\mathbf{r}$. Then, in step 3, the selected index is added to the sparse set, and the sub matrix $\mathbf{A}_{\Lambda_t}$ is constructed by collecting the column vectors of $\mathbf{A}$ corresponding to the indices of the sparse set Λ_t. OMP estimates the signal components corresponding to the indices of the

sparse set and updates the residual vector by removing the estimated signal components in steps 4 and 5, respectively. Finally, OMP finishes its procedures when the cardinality of the sparse set is k.

OMP is a greedy-type algorithm because it selects the one index regarded as the optimal decision at each iteration. Thus, its performance is dominated by its ability to find the sparse set exactly. If the sparse set is not correctly reconstructed, OMP's solution could be wrong. Because OMP is very easy to understand, a couple of modified algorithms based on OMP have been designed and developed. For further information on the OMP algorithm and its modifications, interested readers are referred to two papers [13][14].

8.3 Simultaneous orthogonal matching pursuit (SOMP)

We introduce another greedy-type algorithm based on OMP as an example: simultaneous orthogonal matching pursuit (SOMP) **Error! Reference source not found.**. This greedy algorithm has been proposed for treating multiple measurement vectors for **JSM-2** when the sparse locations of all sensed signals are the same. Namely, SOMP algorithm handles multiple measurements $\mathbf{y}_j$ as an input, when j is the index of distributed sensors, $j \in \{1,2,...,J\}$. In a later section, we use this algorithm to recover **JSM-2**. The pseudo code for SOMP is shown in **Table 4** and **5**.

Input	Output
A $m \times n$ measurement matrix $\mathbf{A}_j$ A $m-$dimensional data vector $\mathbf{y}_j$ The sparsity level k of the ideal signal	An estimate $\hat{\mathbf{x}}_j$ in $\mathbf{R}^n$ for the ideal signal. A set Λ_k containing k elements from $\{1,...,n\}$ An $m-$dimensional approximation $\hat{\mathbf{y}}_{j,k}$ of the data $\mathbf{y}_j$ An $m-$dimensional residual $\mathbf{r}_{j,k} = \mathbf{y}_j - \hat{\mathbf{y}}_{j,k}$

Table 4. Inputs and outputs of SOMP algorithm.

The SOMP algorithm:

1. Initialize:

Let the residual matrix be $\mathbf{r}_{j,0} = \mathbf{y}_{j,0}$. The sparse set $\Lambda_0 = \{\}$, and iteration number $t = 1$.

2. Find the index λ_t : $\lambda_t = \arg\max_{i=1,...,n} \sum_{j=1}^{J} \left| \left\langle \mathbf{r}_{j,t-1}, \mathbf{a}_{j,i} \right\rangle \right|$.

The $\mathbf{a}_{j,i}$ is the i th column vector of matrix $\mathbf{A}_j$.

3. Update set: $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$.

4. Signal estimate: $\mathbf{x}_{j,t}(\Lambda_t) = \mathbf{A}_{j,\Lambda_t}^{\dagger} \mathbf{y}_j$ and $\mathbf{x}_{j,t}(\Lambda_t^C) = \mathbf{0}$, where $\mathbf{x}_{j,t}(\Lambda_t)$ is the set of elements whose indices are corresponding to the sparse set.

5. Get new residual: $\hat{\mathbf{y}}_{j,t} = \mathbf{A}_{j,t} \mathbf{x}_{j,t}$, $\mathbf{r}_{j,t} = \mathbf{y}_j - \hat{\mathbf{y}}_{j,t}$.

6. Increment t : Increase iteration number $t = t+1$, and return to Step 2 if $t < k$.

Table 5. SOMP algorithm.

8.4 Reduce and boost (ReMBo)

ReMBo algorithm is for recovering correlated signals. The authors in [Ref] insisted that the algorithm improves the recovery probability of any suboptimal methods for signal matrix $\mathbf{X}$. Its idea is simple and effective. They transformed the matrix $\mathbf{X}$ to a single vector $\mathbf{x}$ and do $\mathbf{Y}$ to a single measurement vector $\mathbf{y}$. After modifying MMV equation to SMV, they apply any algorithm for SMV. We attached ReMBo algorithm from [Ref].

Input	Output
A $m \times n$ measurement matrix $\mathbf{A}_j$	An estimate $\hat{\mathbf{x}}_j$ in $\mathbf{R}^n$ for the ideal signal.
A $m-$dimensional data vector $\mathbf{y}_j$	Support set $\hat{S}$
The sparsity level k of the ideal signal	flag

Table 4. Inputs and outputs of ReMBo algorithm.

The ReMBo algorithm:

Control parameters : k **,** ε **,** Maxiters

1. Initialize:

Set iter = 1, flag = false.

2. while (iter $\leq$ Maxiter) and (flag is false) **do**

Draw a random vector $\mathbf{a}$ of length j according to randomly generated distribution.

$\mathbf{y} = \mathbf{Aa}$

Solve $\mathbf{y} = \mathbf{Ax}$ using SMV algorithm and save the solution $\mathbf{x}$.

$\hat{S} = I(\mathbf{x})$

If $\left(\left|\hat{S}\right| \leq K\right)$ and $\left(\|\mathbf{y} - \mathbf{Ax}\|_2 \leq \varepsilon\right)$ **then**

flag = true

else

flag = false

end if

Construct $\mathbf{X}$ using $\hat{S}$ and pseudo inverse

iter = iter + 1

end while

return $\mathbf{X}$, $\hat{S}$, flag

Table 5. ReMBo algorithm.

9. Reference

[1] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006

[2] D. L. Donoho and J. Tanner, “Precise undersampling theorems,” *Proc. IEEE*, vol. 98. pp. 913-924, May 2010.

[3] R. G. Baraniuk, “Lecture notes: Compressed sensing,” *IEEE Signal Process. Mag.*, pp. 118-121, July 2007.

[4] J. Romberg, “Imaging via compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14-20, March 2008.

[5] A. Y. Yang, M. Gastpar, R. Bajcsy, and S. S. Sastry, "Distributed sensor perception via sparse representation," to appear in *Proc. IEEE*.

[6] D. L. Donoho and M. Elad, “Maximal sparsity representation via ℓ_1 minimization,” *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197-2202, March 4, 2003.

[7] Jie Chen and Xiaoming Huo, “Theoretical results on sparse representations of multiple measurement vectors”, *IEEE Signal Process.*, pp. 4634-4643, Dec. 2006.

[8] J. Solobera, “Detecting forest fires using wireless sensor networks with Waspmote.”

[9] A. Hac, “*Wireless Sensor Network Designs*,” John Wiley & Sons, Ltd., 2003.

[10] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, “An information theoretic approach to distributed compressed sensing,” in *Proc. 43rd Allerton Conf. Comm., Control, Comput.*, Sept. 2005.

[11] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, “Distributed compressed sensing of jointly sparse signals,” *Asilomar Conf. on Signals, Systems and Computers*, pp. 1537-1541, 2005.

[12] M. Mishali and Y. C. Eldar, “Reduce and boost: Recovering arbitrary sets of jointly sparse vectors,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692-4702, 2008.

[13] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inform. Theor.*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[14] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Simultaneous sparse approximation via greedy pursuit,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 725, pp. v/72I-v/724, 2005.

[15] M. E. Davies and Y. C. Eldar, “Rank awareness in joint sparse recovery,” *Arxiv preprint arXiv:1004.4529*, 2010.

[16] E. Candes and J. Romberg, Caltech, L_1 -Magic: Recovery of sparse signals via convex programming, Oct 2005.

Multiuser Cooperative Diversity Through Network Coding Based on Classical Coding Theory

Authors: J. Rebelatto, B.Uchoa-Filho, Y. Li, B. Vucetic

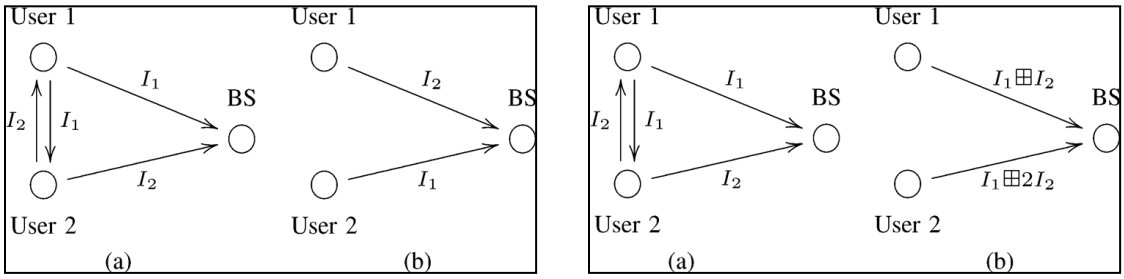
Publication: IEEE Trans. Sig. Processing, Feb. 2012.

Speaker: Jin-Taek Seong

Short summary: To increase the diversity order of cooperative wireless communication systems without sacrificing the system's rate, they propose the generalized dynamic-network code (GDNC). They showed that the problem of designing network codes that maximize the diversity order is related to that of designing optimal linear block codes, in the Hamming distance sense over finite fields.

I. INTRODUCTION

In the network coding schemes, each user linearly combines the messages coming from other users, and generates a new message, and then forwards it to the destination.



- Each user broadcasts its own information
- With the routing protocol, each user transmits its partner's information after decoding,
- With network coding schemes, each user transmits a linear combination regenerated.

The contributions of this paper are:

i) they investigate another relationship between network codes and classical error-correcting codes. They explain on the dynamic-network coding (DNC) scheme by first recognizing the associated network code design problem as equivalent to that

of designing linear block codes over for erasure correction. In particular, for perfect interuser channels, we note that the diversity order equals the minimum Hamming distance of the block code, so the network transfer matrix should correspond to the generator matrix of an optimal block code under the Hamming metric. The Singleton upper bound for the minimum Hamming distance of a linear block code appears as a natural limit to the diversity order, and this bound is achieved with a sufficiently large field size. The codes that achieve the Singleton bound are called *maximum distance separable* (MDS) codes.

ii) Regarding the GDNC network code design, they show that if a generator matrix of a MDS code is used as the GDNC network code, the maximum diversity order is guaranteed. They also show that a much better tradeoff between rate and diversity order can be achieved, e.g., it is possible to improve both rate and diversity order over the DNC scheme.

II. System Model

A. System Model

The received baseband codeword at User i at time t is given by

$$\mathbf{y}_{j,i,t} = h_{j,i,t}\mathbf{x}_{j,t} + \mathbf{n}_{j,i,t} \tag{1}$$

where $j \in \{1,...,M\}$ represents the transmit user index and $i \in \{0,1,...M\}$ the receive user index (0 corresponds to the BS).

The mutual information $I_{j,i,t}$ between $\mathbf{x}_{j,t}$ and $\mathbf{y}_{j,i,t}$ is

$$I_{j,i,t} = \frac{1}{M}\log_2\left(1 + \left|h_{j,i,t}\right|^2 SNR\right) \tag{2}$$

where the factor $1/M$ follows form the division of the channel's resources among the M users.

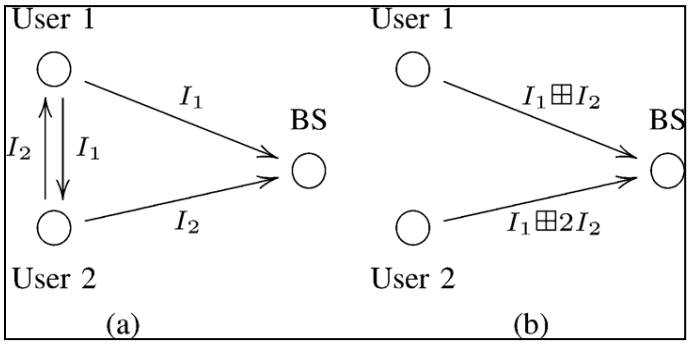
For Rayleigh fading, the outage probability is calculated as

$$P_e = \Pr\left\{\left|h_{j,i,t}\right|^2 < g\right\} = 1 - e^{-g} \approx g \tag{3}$$

The diversity order D is defined as

$$D \triangleq \lim_{SNR\to\infty} \frac{-\log P_0}{\log SNR} \tag{4}$$

B. Binary vs. Nonbinary Network Coded Cooperation



The network codes for both schemes are

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}_{\text{Binary}} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{5}$$

and

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}_{4-ary} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \tag{6}$$

For the binary and 4-ary network coding schemes, the general and exact form of the outage probability of the I_1 message at the BS are obtained as follows [1]

$$P_{o,4-ary} \approx \frac{A_1}{P_1^2 P_2} + \frac{A_2}{P_1 P_2^2} + \frac{A_3}{P_1^3} \tag{7}$$

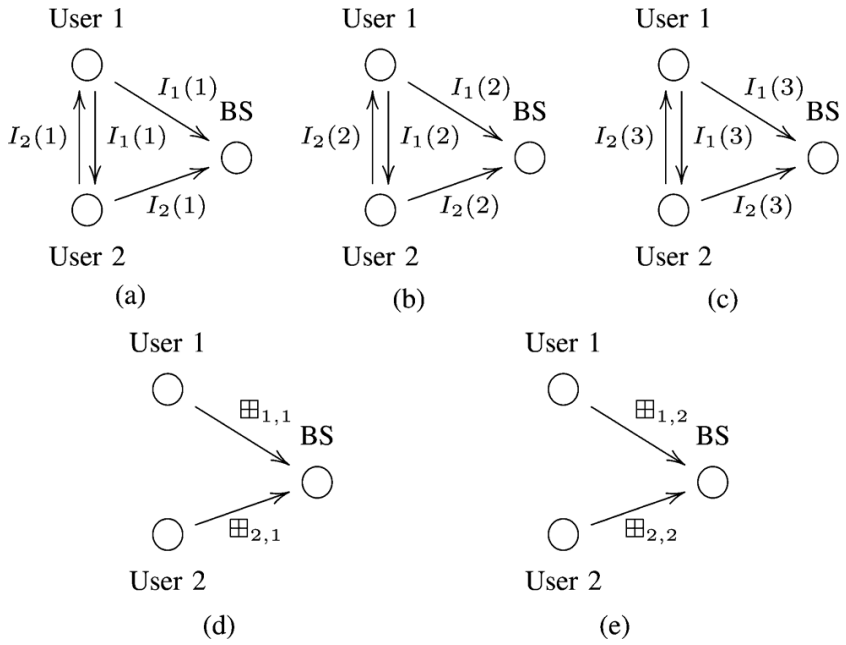
and

$$P_{o,binary} \approx \frac{B_1}{P_1 P_2} + \frac{B_2}{P_1^2 P_2} + \frac{B_3}{P_1 P_2^2} + \frac{B_4}{P_1^3} \tag{8}$$

where the constants A_c and B_c are determined by the variances of channel gains and transmission rates.

III. GENERALIZED DYNAMIC-NETWORK CODES

In the DNC scheme [2], [3], the diversity order is related to the minimum number of correctly received packets at the BS from which the information packets from all users can be recovered. A packet which is not received correctly may be thought of as an erasure, and is discarded by the receiver. The receiver's ability to recover the information packets from the non-erased packets is thus equivalent to the erasure correction capability of the associated (network) block code. It is well-known that the transmitted codeword of a linear block code with minimum Hamming distance can be recovered if no more than of its positions have been erased by the channel.



For example: Each user broadcasts three packets of its own in the broadcast phase, and then each user transmits two nonbinary linear combinations (of the six previously broadcasted packets) over in the cooperative phase, where is an integer greater than zero. The receiver collects the 10 packets, which can be seen as a codeword of a systematic 6/10 linear block code.

In this case, the outage probability of $I_1(1)$ message can be approximately derived,

$$P_{o,1} = P_e P_{f,1} + (1-P_e) P_{p,1} \approx P_e^4 \tag{9}$$

where the outage probability $P_{f,1} \approx P_e^3$ is from fact that the 3 outage events (direct transmission and two parity messages) occur when User 2 cannot correctly decode $I_1(1)$, the outage probability $P_{p,1} \approx P_e^5$ is obtained from that when User 2 can decode $I_1(1)$, one direct

transmission plus four parity messages are erased in the worst case. We can also verify that the outage probability is dominated by the term related to the interuser channel being in outage, when User 2 cannot help User 1.

The GDNC overall rate is given by

$$R_{GNDC} = \frac{k_1 M}{k_1 M + k_2 M} = \frac{k_1}{k_1 + k_2} \tag{10}$$

In general, for a rate k/n linear block code, the minimum Hamming distance is upper bounded by the Singleton bound,

$$d_{\min} \le n - k + 1 \tag{11}$$

Thus, we can see that the diversity order of the GDNC scheme is upper bounded by

$$D_{GDNC} \le k_2 M + 1 \tag{12}$$

However, due to outages in interuser channels, this upper bound cannot be achieved.

IV. On the Network Code Design

Theorem 1: The diversity order of the GDNC scheme for an appropriately designed network code with sufficiently large field size is $D_{GDNC} = M + k_2$.

Proof: Let $D_{j,t} \subseteq \{1,..,M\}$ be the index set corresponding the users that correctly decoded the information packet $I_j(t)$. We define a new set $\mathcal{D}_{j,t}(I)$ as the set of all messages correctly decoded by the users in $D_{j,t}$ in the broadcast phase, including $I_j(t)$ itself. There are at least $|\mathcal{D}_{j,t}(I)| + |D_{j,t}| k_2$ packets containing messages of $\mathcal{D}_{j,t}(I)$. For a fixed $D_{j,t}$, the message $I_j(t)$ is declared and erased at the BS only if the direct transmission $I_j(t)$ and the at least $|D_{j,t}| k_2$ out of the remaining $|\mathcal{D}_{j,t}(I)| + |D_{j,t}| k_2 - 1$ received packets are not correctly decoded by the BS which occurs with probability

$$
\begin{aligned}
P_{o,j}\left(\bar{D}_{j,t}\right) &= P_e\left[\sum_{F=\left|D_{j,t}\right|k_2}^{\left|\mathcal{D}_{j,t}(I)\right|+\left|D_{j,t}\right|k_2-1}\binom{\left|\mathcal{D}_{j,t}(I)\right|+\left|D_{j,t}\right|k_2-1}{F}P_e^F\left(1-P_e\right)^{\left(\left|\mathcal{D}_{j,t}(I)\right|+\left|D_{j,t}\right|k_2-1\right)-F}\right] \\
&\approx P_e\left[\binom{\left|\mathcal{D}_{j,t}(I)\right|+\left|D_{j,t}\right|k_2-1}{\left|D_{j,t}\right|k_2}P_e^{\left|D_{j,t}\right|k_2}\right] \\
&= \gamma\left(k_1,k_2,\bar{D}_{j,t}\right)P_e^{\left(M-\left|\bar{D}_{j,t}\right|\right)k_2+1}, \quad \gamma\left(k_1,k_2,\bar{D}_{j,t}\right)=\binom{\left|\mathcal{D}_{j,t}(I)\right|+\left|D_{j,t}\right|k_2-1}{\left|D_{j,t}\right|k_2}
\end{aligned} \tag{13}
$$

The outage probability of the information message $I_j(t)$ is given by

$$
\begin{aligned}
P_{o,j} &= \sum_{\bar{D}_{j,t}} P_e^{\left|\bar{D}_{j,t}\right|}\left(1-P_e\right)^{(M-1)-\left|\bar{D}_{j,t}\right|}P_{o,j}\left(\bar{D}_{j,t}\right) \\
&\approx \sum_{\bar{D}_{j,t}} P_e^{\left(M-\left|\bar{D}_{j,t}\right|\right)k_2+\left|\bar{D}_{j,t}\right|+1}\gamma\left(k_1,k_2,\left|\bar{D}_{j,t}\right|\right) \\
&\approx \binom{M-1}{\left|\bar{D}_{j,t}\right|^*}P_e^{\left(M-\left|\bar{D}_{j,t}\right|^*\right)k_2+\left|\bar{D}_{j,t}\right|^*+1}\gamma\left(k_1,k_2,\left|\bar{D}_{j,t}\right|^*\right)
\end{aligned} \tag{14}
$$

For $k_2 \geq 2$, $\left|\bar{D}_{j,t}\right|^* = M-1$ since the lowest exponent achieves. Thus, the exponent of (14) is $M+k_2$.

Theorem 2: An $(n,k,d_{\min})$ code $\mathcal{C}$ with generator matrix $\mathbf{G}=\left[\mathbf{I}\middle|\mathbf{P}\right]$, is minimum distance separable (MDS) if and only if every square submatrix of $\mathbf{P}$ is nonsingular.

We consider network codes in the light of classical coding theory. Let $\mathcal{C}$ be an $(n,k,d_{\min})$ linear block code over GF(q) with systematic generator matrix $\mathbf{G}$ given by

$$
\mathbf{G}=\left[\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & p_{1,1} & p_{1,2} & \cdots & p_{1,n-k} \\
0 & 1 & \cdots & 0 & p_{2,1} & p_{2,2} & \cdots & p_{2,n-k} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & p_{k,1} & p_{k,2} & \cdots & p_{k,n-k}
\end{array}\right]=\left[\mathbf{I}_k\middle|\mathbf{P}_{k\times n-k}\right] \tag{15}
$$

A. A Theory of Faulty Generator Matrices

We study some properties of new block codes obtained from systematic MDS codes by zeroing some entries of its generator matrix, and we refer to the obtained generator matrix as *faulty generator matrix*.

Theorem 3: Let $\mathcal{C}$ be an $(n,k,d_{\min})$ MDS code with systematic generator matrix $\mathbf{G}=\left[\mathbf{I}\middle|\mathbf{P}\right]$. The replacement by zeros of δ entries in any row of the matrix $\mathbf{P}$ gives rise to $(n,k,d_{\min}-\delta)$ code $\mathcal{C}$.

Theorem 4: If a systematic generator matrix of a MDS code $\mathcal{C}$ with minimum Hamming distance $d_{\min}=Mk_2+1$ is used as a transfer matrix of the GDNC scheme, the diversity order $D_{GDNC}=M+k_2$ is guaranteed.

Proof: One fault corresponds to one interuser channel being in outage. When that happens, the user's receiver cannot correctly decode its partner's information, so, when forming k_2 linear combinations to generate its k_2 parity-check packets, this user replaces this erroneous packet with an all-zero packet, or equivalently, sets to zero the k_2 coefficients associated with this partner. This amounts to replacing by zeros the k_2 corresponding entries (in same row) of the parity matrix $\mathbf{P}$, i.e., $|A|=k_2$. Since each user knows its own information, k_2 entries in each row of $\mathbf{P}$ are immune to faults, while the other $k_2(M-1)$ entries are subject to faults. In the worst scenario, when all the possible faults happen, the generator matrix takes the form

$$\mathbf{G}=\left[\begin{array}{c|ccc} I & P_1 & & \\ & & \ddots & \\ & & & P_M \end{array}\right] \tag{16}$$

where the $(k_1\times k_2)$ submatrix P_i contains the immune entries associated with User i. From Theorem 2, we know that every submatrix of P_i is nonsingular. Thus, the least minimum Hamming distance of a block code obtained from the original MDS code $\mathcal{C}$ due to the occurrence of faults is k_2+1. Nevertheless, the same minimum Hamming distance can be

achieved with a much lower number of faults. We can see that the minimum number of faults for a code with minimum distance k_2+1 is $M-1$, when all of these faults occur in the same row of $\mathbf{P}$, for example.

For all the possible minimum distances in the range $k_2+1 \le d_{\min} \le Mk_2+1$, a sufficient condition for the worst possible scenario (the lowest number of faults that result in this minimum distance) is when all the faults are located in the same row of $\mathbf{P}$. Thus, we can observe that the larger the number of faults in a given row (and consequently the lower the minimum distance of the resulting code), the lower the composite minimum distance. This assures that the code with minimum distance k_2+1 is the one that generates the least composite minimum distance, with is then given by, see the Appendix in detail,

$$\begin{aligned} d_{\min}^{comp} &= \min_{B_\chi \in \mathcal{B}} \left\{ d_{\min}\left(\mathcal{C}\left(B_\chi\right)\right) + \left|B_\chi\right| / \left|A\right| \right\} \\ &= \left(k_2+1\right) + \left(M-1\right) \\ &= M + k_2 \end{aligned} \tag{17}$$

It is easy to see the connection between the two terms in the composed minimum distance and the exponents of P_e.

V. Simulation Results

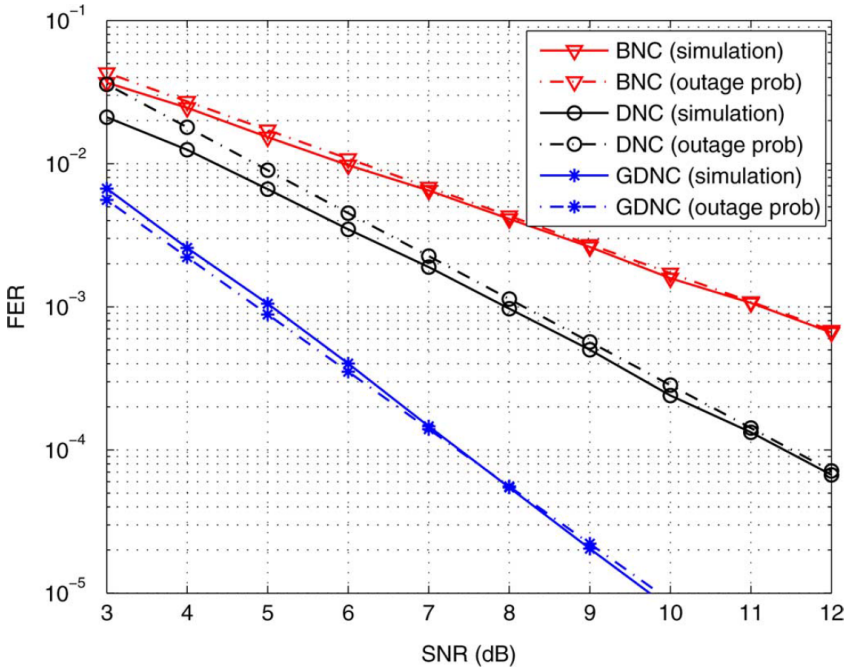


Fig. 5. FER versus SNR (dB) of a 2-user system with the BNC scheme, DNC scheme [in GF(4)] and the proposed GDNC scheme (with $k_1 = k_2 = 2$ and in GF(8), according to Table I), all with rate 1/2.

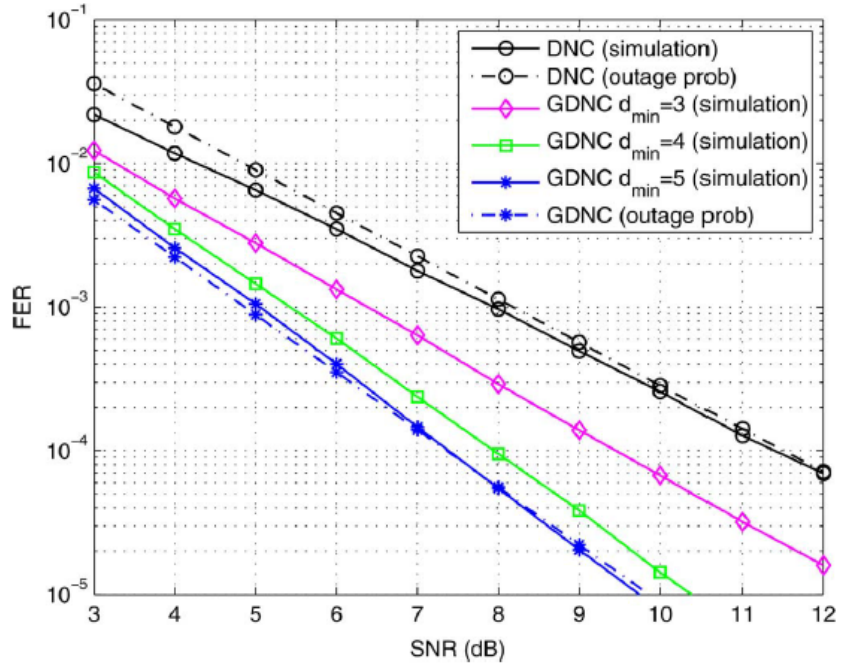


Fig. 6. FER versus SNR (dB) of a 2-user system with the DNC scheme [in GF(4)] and the proposed GDNC scheme (with $k_1 = k_2 = 2$ and in GF(8), according to Tables I and III), all with rate 1/2.

TABLE I

NETWORK CODES OBTAINED FROM RS CODES FOR $M = 2$ USERS

k_1	k_2	Code rate	Field size q	Parity matrix $\mathbf{P}$
1	1	2/4	4	$\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$
1	2	2/6	8	$\begin{bmatrix} 3 & 5 & 1 & 7 \\ 2 & 4 & 3 & 6 \end{bmatrix}$
2	1	4/6	8	Obtained from the 4/8 code by puncturing the last 2 columns
2	2	4/8	8	$\begin{bmatrix} 3 & 7 & 3 & 6 \\ 5 & 7 & 7 & 4 \\ 2 & 4 & 6 & 1 \\ 5 & 5 & 3 & 2 \end{bmatrix}$
2	3	4/10	16	Obtained from the 4/12 code by puncturing the last 2 columns
2	4	4/12	16	$\begin{bmatrix} 14 & 5 & 7 & 10 & 4 & 10 & 8 & 9 \\ 12 & 4 & 14 & 5 & 12 & 12 & 14 & 7 \\ 13 & 6 & 12 & 15 & 8 & 1 & 4 & 10 \\ 14 & 6 & 4 & 1 & 1 & 6 & 3 & 5 \end{bmatrix}$
3	2	6/10	16	Obtained from the 6/12 code by puncturing the last 2 columns
3	3	6/12	16	$\begin{bmatrix} 11 & 2 & 4 & 6 & 14 & 12 \\ 1 & 11 & 13 & 10 & 14 & 10 \\ 2 & 4 & 2 & 10 & 5 & 9 \\ 6 & 13 & 12 & 11 & 8 & 12 \\ 4 & 12 & 12 & 2 & 6 & 6 \\ 11 & 13 & 10 & 14 & 10 & 4 \end{bmatrix}$

VI. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

VII. APPENDIX

When a set of entries of the generator marix $\mathbf{P}$ of the MDS code $\mathcal{C}$ is replace by zeros, the new code $\mathcal{C}'$ produced is no longer MDS. Let $A=\{(a_1,b_1),(a_2,b_2),...\}$ be a subset of entries of $\mathbf{P}$ that are replaced by zeros, where (a,b) is the entry in row a and column b of the matrix $\mathbf{P}$. We call A as *a fault*. Let $\mathcal{A}=\{A_0,A_1,...,A_{f-1}\}$ be a collection of f faults. We consider that two different faults A_i and A_j cannot contain a common entry of the matrix $\mathbf{P}$. That is, $A_i \cap A_j = \varnothing$. It is also considered that every fault has fixed cardinality, i.e., $|A_i|=|A| \forall i$.

Let $\chi=(\chi_0,\chi_1,...,\chi_{f-1})$ be the binary indicator vector associated with the occurrence of faults, where

$$\begin{cases} \chi_i = 1, & \text{if } A_i \text{ occurs} \\ \chi_i = 0, & \text{if } A_i \text{ does not occurs} \end{cases} \tag{18}$$

For a nonnegative integer i, let $b(i)$ denote the binary (vector) representation of i. We denote the collection of all possible combinations of faults by $\mathcal{B}=\left\{B_{b(0)},B_{b(1)},...,B_{b\left(2^f-1\right)}\right\}$, with $B_\chi=\left\{(a,b)\middle|(a,b)\in\bigcup_{i:\chi_i=1}A_i\right\}$. Each event $\mathcal{B}_\chi$, which consists of the occurrence of $|\mathcal{B}_\chi|/|A|$, gives rise to a new generator matrix of a block code $\mathcal{C}(B_\chi)$ with minimum Hamming distance $d_{\min}(\mathcal{C}(B_\chi))$. We also define the *minimum composite distance* of the code $\mathcal{C}(B_\chi)$ as

$$d_{\min}^{comp}(\mathcal{C}(B_\chi)) \triangleq d_{\min}(\mathcal{C}(B_\chi)) + |B_\chi|/|A| \tag{19}$$

which is composed of its minimum Hamming distance $d_{\min}(\mathcal{C}(B_\chi))$ plus a "compensation" term related to the number of faults in the combination B_χ.

A parameter of fundamental importance to indicate the performance of a MDS code subject to a set of faulty generator matrices is the least minimum composite distance of any possible

combination of faults. In particular, given a MDS code $\mathcal{C}$ and a collection of f faults $\mathcal{A}$, this distance is defined as

$$d_{\min}^{comp}\left(\mathcal{C},\mathcal{A}\right) \triangleq \min_{B_\chi \in \mathcal{B}} \left\{ d_{\min}^{comp}\left(\mathcal{C}\left(B_\chi\right)\right)\right\} \tag{20}$$

where, when there is no confusion, it is simply called minimum composite distance.

Reference

[1] J.T. Seong and H.N. Lee, "4-ary Network Coding for Two Nodes in Cooperative Wireless Networks: Exact Outage Probability and Coverage Expansion," To appear EURASIP Wirel. Comm. Networking.

[2] M. Xiao and M. Skoglund, "M-user cooperative wireless communications based on nonbinary network codes," in *Proc. IEEE Inf. Theory Workshop (ITW'09), Jun. 2009, pp. 316–320.*

[3] M. Xiao and M. Skoglund, "Multiple-user cooperative communications based on linear network coding," *IEEE Trans. Commun., vol. 58,* no. 12, pp. 3345–3351, Dec. 2010

A fast approach for overcomplete sparse decomposition based on smoothed L0 norm

Authors: H. Mohimani, M. B. Zadeh, C. Jutten
Publication: IEEE Trans. on Sig. Process, Jan.2009
Speaker: Oliver

Short summary:

This paper proposes a fast algorithm for overcomplete sparse decomposition. The algorithm is derived by directly minimizing the L0 norm after smoothing. Hence, the algorithm is named as smoothed L0 (SL0) algorithm. The authors demonstrate that their algorithm is 2-3 orders of magnitude faster than the state-of-the-art interior point solvers with same (or better) accuracy.

I. INTRODUCTION

- To introduce the algorithm the authors have used the context of source component analysis (SCA). SCA is a method to achieve separation of sparse sources.
- Suppose that *m* source signals are recorded by a set of *n* sensors each of which records a combination of all sources. In linear instantaneous (noiseless) model, it is assumed that $x(t) = As(t)$ in which $s(t)$ and $x(t)$ are $m \times 1$ and $n \times 1$ vectors of source and recorded signals, respectively, and $A \in R^{n \times m}$ is a mixing matrix.
- The goal of blind source separation (BSS) is then to find $s(t)$ only by observing $x(t)$. The general BSS problem is impossible for the case $n<m$. However, if the sources are sparse then this problem can be solved (using L1 minimization).
- We have the problem of finding sparse solutions of the undetermined system of linear equations (USLE) $As=x$. To obtain the sparsest solution of $As=x$, we may search for a solution with minimal L0 norm. (Intractable problem, sensitive to noise)
- Hence, researchers consider L1 approaches such as basis pursuit (BP), LP-norm approaches such as IRLS, and greedy approaches such as matching pursuit (MP).
- In this paper, authors present an approach for solving USLE by direct minimization of the L0 norm after smoothing (approximating with smooth functions).
- Performance of the algorithm is equal to (or better than) the interior point based algorithms with 2 to 3 orders of magnitude faster.

algorithm	total time (sec)	MSE
SL0	0.227	$5.53\,e-5$
LP (ℓ_1-magic)	30.1	$2.31\,e-4$
FOCUSS	20.6	$6.45\,e-4$

II. APPROACH

- L0 norm of a vector $x \in R^m$ is a *discontinuous* function of that vector.

$$\|x\|_0 = \sum_{i=1}^{m} I_i \qquad I_i = \begin{cases} 1 & if\ x_i \neq 0 \\ 0 & if\ x_i = 0 \end{cases}$$

- The idea then is to approximate the discontinuous function with a continuous function. The continuous function has a parameter (say σ) that determines the quality of the approximation.
- For example, consider the (one-variable) family of Gaussian functions

$$f_\sigma(s) \triangleq \exp\left(\frac{-s^2}{2\sigma^2}\right)$$

and note that

$$\lim_{\sigma \to 0} f_\sigma(s) = \begin{cases} 1 & if\ s = 0 \\ 0 & if\ s \neq 0 \end{cases}$$

or approximately

$$f_\sigma(s) \approx \begin{cases} 1 & if\ |s| \ll \sigma \\ 0 & if\ |s| \gg \sigma \end{cases}$$

- Now define

$$F_\sigma(s) = \sum_{i=1}^{m} f_\sigma(s_i)$$
$$\|s\|_0 \approx m - F_\sigma(s)$$

- For small values of σ, the approximation tends to equality. Hence, we can define the minimum L0 norm solution by maximizing $F_\sigma(s)$.
- The value of σ determines how smooth the function $F_\sigma(s)$ is: the larger value of σ, the smoother $F_\sigma(s)$ (but worse approximation to L0-norm); and the smaller value of σ, closer the behavior of $F_\sigma(s)$ to L0-norm.
- However, for smaller values of σ, $F_\sigma(s)$ is highly non-smooth and contains a lot of local maxima, and hence its maximization is not easy. On the other hand, for larger values of σ, $F_\sigma(s)$ is smoother and contains less local maxima (in fact, no local maxima for large σ).
- "Basic idea": In order to find an s that maximizes $F_\sigma(s)$, the authors start with maximum σ. For this maximum σ, they find the maximizer of $F_\sigma(s)$. Then they decrease σ and again find the maximizer of $F_\sigma(s)$.
- They claim that eventually this process (decreasing σ and maximizing $F_\sigma(s)$) results in the maximization of $F_\sigma(s)$ or equivalently minimization of the L0 norm.

- Other family of functions that approximates the Kronecker delta functions like family of triangular functions,

$$f_\sigma(s) = \begin{cases} 1, & \text{if } |s| \geq \sigma \\ \frac{(\sigma+s)}{\sigma}, & \text{if } -\sigma \leq s \leq 0 \\ \frac{(\sigma-s)}{\sigma}, & \text{if } 0 \leq s \leq \sigma \end{cases}$$

- or truncated hyperbolic functions

$$f_\sigma(s) = \begin{cases} 1, & \text{if } |s| \geq \sigma \\ 1 - \left(\frac{s}{\sigma}\right)^2, & \text{if } |s| \leq \sigma \end{cases}$$

- or functions of the form

$$f_\sigma(s) = \frac{\sigma^2}{(s^2 + \sigma^2)}.$$

- For sufficiently large values of σ the maximizer of $F_\sigma(s)$ subject to $As=x$ is the minimum L2-norm solution, i.e., $\hat{s} = A^T \left(AA^T\right)^{-1} x$.

  Justification of the statement for Gaussian family:

  We want to maximize $F_\sigma(s) = \sum_{i=1}^{m} f_\sigma(s_i) = \sum_{i=1}^{m} e^{\left(-s_i^2 / 2\sigma^2\right)}$ subject to $As=x$

  The Lagrangian is $F_\sigma(s) - \lambda^T(As - x)$. Differentiating the Lagrangian w.r.t s and λ and setting the result to zero gives the following KKT systems of $m+n$ non-linear equations of $m+n$ unknowns.

$$\left[s_1 e^{-s_1^2/2\sigma^2}, s_2 e^{-s_2^2/2\sigma^2}, \cdots, s_m e^{-s_m^2/2\sigma^2} \right]^T - A^T \lambda_1 = 0 \qquad \lambda_1 = -\sigma^2 \lambda$$

$$As - x = 0$$

  Now, let us look at this problem: $\min \|s\|_2^2$ *s.t.* $As = x$. Again using Lagrange multipliers this minimization results in the system of equations

$$\left[s_1, s_2, \cdots, s_m\right]^T - A^T \lambda = 0$$

$$As - x = 0$$

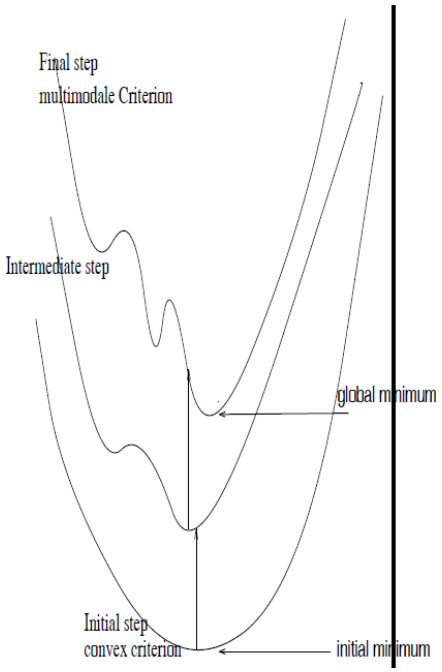
  Authors claim that for a larger σ these two systems of equations are identical and hence the maximizer of $F_\sigma(s)$ for larger σ is the minimum L2 norm solution.

- Hence, the authors start with large σ and maximize the corresponding $F_\sigma(s)$. They then decrease σ and repeat the maximization of $F_\sigma(s)$ again. They repeat the process for a few sequences of σ and shown that the subsequent maximization of $F_\sigma(s)$ leads to L0 solution. Their algorithm is based on the principles of graduated non-convexity.

  From Wikipedia:

- Graduated non-convexity is a global optimization technique that attempts to solve a difficult optimization problem by initially solving a greatly simplified problem, and progressively transforming that problem (while optimizing) until it is equivalent to the difficult optimization problem.

- Graduated optimization is an improvement to hill climbing that enables a hill climber to avoid settling into local optima. It breaks a difficult optimization problem into a sequence of optimization problems, such that the first problem in the sequence is convex (or nearly convex), the solution to each problem gives a good starting point to the next problem in the sequence, and the last problem in the sequence is the difficult optimization problem that it ultimately seeks to solve. Often, graduated optimization gives better results than simple hill climbing [1].



III. THE SL0 ALGORITHM

The algorithm consists of 2 loops: the outer and inner loops.

In the outer loop, they vary the values of σ.

In the inner loop, for a given σ they use a steepest ascent algorithm for maximizing $F_\sigma(s)$.

- Initialization:
  1) Let $\hat{\mathbf{s}}_0$ be equal to the minimum ℓ^2 norm solution of $\mathbf{As} = \mathbf{x}$, obtained by pseudo-inverse of $\mathbf{A}$.
  2) Choose a suitable decreasing sequence for σ, $[\sigma_1 \dots \sigma_J]$ (see Remarks 5 and 6 of the text).
- For $j = 1, \dots, J$:
  1) Let $\sigma = \sigma_j$.
  2) Maximize (approximately) the function F_σ on the feasible set $\mathcal{S} = \{\mathbf{s} \mid \mathbf{As} = \mathbf{x}\}$ using L iterations of the steepest ascent algorithm (followed by projection onto the feasible set):
     - Initialization: $\mathbf{s} = \hat{\mathbf{s}}_{j-1}$.
     - For $\ell = 1 \dots L$ (loop L times):
       a) Let $\boldsymbol{\delta} \triangleq [s_1 \exp(-s_1^2/2\sigma^2), \dots, s_n \exp(-s_n^2/2\sigma^2)]^T$.
       b) Let $\mathbf{s} \leftarrow \mathbf{s} - \mu\boldsymbol{\delta}$ (where μ is a small positive constant).
       c) Project $\mathbf{s}$ back onto the feasible set $\mathcal{S}$:
       $$\mathbf{s} \leftarrow \mathbf{s} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}(\mathbf{As} - \mathbf{x}).$$
  3) Set $\hat{\mathbf{s}}_j = \mathbf{s}$.
- Final answer is $\hat{\mathbf{s}} = \hat{\mathbf{s}}_J$.

Fig. 1. Final SL0 algorithm.

Remark 1: The internal loop is repeated for a fixed, say L, and small number of times. In other words, we do not wait for the steepest ascent algorithm (SAA) to converge. That is, we do not need the exact maximizer of $F_\sigma(s)$. We just need to enter the region near the (global) maximizer of for escaping from its local maximizer.

Remark 2: The SAA consists of update step $s \leftarrow s + \mu_j \nabla F_\sigma(s)$. Here, μ_j s are step size parameter and they should be chosen such that for decreasing values of σ, μ_j should be smaller. In the algorithm, they let $\mu_j = \mu\sigma^2$ for some constant μ. Then,

$$s \leftarrow s + \mu_j \nabla F_\sigma(s) = s - \mu\delta \text{ where } \delta = -\sigma^2 \nabla F_\sigma = \left[s_1 e^{-s_1^2/2\sigma^2}, s_2 e^{-s_2^2/2\sigma^2}, \cdots, s_m e^{-s_m^2/2\sigma^2} \right]^T.$$

Remark 3: Each iteration of the inner loop consists of gradient ascent step, followed by a projection step.

If we are looking for a suitable large μ (to reduce the required number of iterations), a suitable choice is to make the algorithm to force all those values of s_i satisfying $|s_i| \le \sigma$ toward zero. For this aim, we should have $\mu \exp\left(\frac{-s_i^2}{2\sigma^2}\right) \approx 1$, and because $\exp\left(\frac{-s_i^2}{2\sigma^2}\right) \le 1$ for $|s_i| \le \sigma$, the choice $\mu \ge 1$ seems reasonable.

Remark 4: The algorithm may work by initializing to an arbitrary solution. However, as discussion before the best initial value of is the minimum L2 solution. In fact, calculating minimum L2norm is one of the earliest approaches used for estimating the sparsest solution called the method of frames [4].

Remark 5: After initiating with minimum L2 solution, the next value for σ may be chosen about two to four times of the maximum absolute value of the obtained solution ($\max_i |s_i|$). For example, if we take $\sigma > 4\max_i |s_i|$, then $\exp\left(\frac{-s_i^2}{2\sigma^2}\right) \geq 0.96 \approx 1 \;\; i = 1, 2, \cdots, m$. This value of σ acts virtually like infinity for all values of s_i.

Remark 6: The term $F_\sigma(s)$ ($\|s\|_0 \approx m - F_\sigma(s)$) simply counts the number of zero components of *s*. However, instead of hard-thresholding that is "zero $\equiv |s_i| < \sigma$" and "non-zero $\equiv |s_i| > \sigma$" $f_\sigma(s) \approx \begin{cases} 1 & if \; |s| \ll \sigma \\ 0 & if \; |s| \gg \sigma \end{cases}$ uses a soft-thresholding, for which σ is a rough threshold.

Remark 7:

- If ***s*** is an exactly *K*-sparse signal, then σ can be decreased to arbitrarily small values. In fact, in this case, the minimum value of σ is determined by the desired accuracy (as will be discussed in Theorem 1).
- If ***s*** is an approximately *K*-sparse signal (say the source vector is noisy), then the smallest σ should be about one to two times of (a rough estimation of) the standard deviation of the noise (in the source vector). This is because, while σ is in this range, $f_\sigma(s)$ shows that the cost function treats small (noisy) samples as zeros (i.e., for which $f_\sigma(s) \approx 1$).
- However, below this range, the algorithm tries to 'learn' these noisy values, and moves away from the true answer. (According to the previous remark, the soft threshold should be such that all these noisy samples be considered zero).
- Restricting σ to be above the standard deviation of the noise, provides the robustness of this approach to noisy sources, which was one of the difficulties in using the exact L0 norm.

IV. Analysis of the Algorithm

A. Convergence analysis

In this section, we try to answer two questions for the noiseless case (the noisy case will be considered in Section IV-C):

a) Does the SL0 solution converges to the actual minimizer of the L0 norm?
b) If yes, how much should we decrease σ to achieve a desired accuracy?

Assuming the maximization of $F_\sigma(s)$ for fixed σ is perfectly done (and we obtain the maximizer s^σ). The authors show that the sequence of 'global' maximizers of $F_\sigma(s)$'s will converge to the sparsest solution. That is we need to prove

$$\lim_{\sigma \to 0} s^\sigma = s^0$$

For proving the above statement the authors have introduced three intermediate results via lemmas.

Lemma 1: Assume a matrix $A \in R^{n \times m}$ has the property that all of its $n \times n$ sub-matrices are invertible, which is called the unique representation property (URP) in [3]. For any $\forall s \in N(A)$ if the *m-n* elements of ***s*** have absolute values less than $\alpha \Rightarrow \|s\| \le \beta$.

Proof: We have to show that

$\forall\, \beta > 0, \quad \exists \alpha > 0, \quad s.t \quad \forall s \in N(A)$:
m-n elements of ***s*** have absolute values less than $\alpha \Rightarrow \|s\| \le \beta$ $\quad\quad \|.\|$ stands for L2 norm

Let $s \in N(A)$ and assume that the absolute values of at least *m-n* elements of it are smaller than α. Let I_α be the set of all indices *i*, for which $|s_i| > \alpha$.Consequently, $|I_\alpha| \le n$. Then we write

$$\sum_{i=1}^{m} s_i a_i = 0 \Rightarrow \sum_{i \in I_\alpha} s_i a_i + \sum_{i \notin I_\alpha} s_i a_i = 0$$

$$\Rightarrow \left\| \sum_{i \in I_\alpha} s_i a_i \right\| = \left\| \sum_{i \notin I_\alpha} s_i a_i \right\| \le \sum_{i \notin I_\alpha} \|s_i a_i\|$$

$$= \sum_{i \notin I_\alpha} |s_i| \|a_i\|$$

$$\le (m - |I_\alpha|)\alpha \le m\alpha$$

Let $\hat{A}$ be the submatrix of *A* containing only those columns of *A* that are indexed by the elements of I_α. Thus, $\hat{A}$ has at most *n* columns, and the columns of $\hat{A}$ are linearly independent, because of the URP of *A*. Therefore, there exists a left inverse $\hat{A}^{-1}$ for $\hat{A}$. Let $\bar{s}$ and $\tilde{s}$ denote those sub-vectors of ***s*** which are, and which are not indexed by I_α, respectively. Then

$$\left(\sum_{i\in I_\alpha} s_i a_i\right) = \hat{A}\bar{s} \;\Rightarrow\; \bar{s} = \hat{A}^{-1}\left(\sum_{i\in I_\alpha} s_i a_i\right)$$

$$\|\bar{s}\| = \left\|\hat{A}^{-1}\left(\sum_{i\in I_\alpha} s_i a_i\right)\right\|$$

$$\|\bar{s}\| \le \left\|\hat{A}^{-1}\right\|\left\|\sum_{i\in I_\alpha} s_i a_i\right\| = \left\|\hat{A}^{-1}\right\| m\alpha$$

$$\|\tilde{s}\| \le \sum_{i\notin I_\alpha} |s_i| \le (m - |I_\alpha|)\alpha \le m\alpha$$

$$\|s\| \le \|\bar{s}\| + \|\tilde{s}\| \le \left(1 + \left\|\hat{A}^{-1}\right\|\right) m\alpha$$

Now, let M be the set of all submatrices $\hat{A}$ of *A*, consisting of at most *n* columns of *A*. Then, M is clearly a finite set (in fact $|\mathsf{M}| < 2^m$).

Let $M = \max\left\{\left\|\hat{A}^{-1}\right\| \,\middle|\, \hat{A} \in \mathsf{M}\right\}$ then

$$\|s\| \le \left(1 + \left\|\hat{A}^{-1}\right\|\right) m\alpha \le (1 + M)\, m\alpha.$$

M is a constant and its value depends only on the matrix *A*. Thus, for each β it is suffice to choose $\alpha = \beta / m(M+1)$

<u>*Corollary 1:*</u> If $A \in R^{n\times m}$ satisfies the URP, and $s \in N(A)$ has at most *n* elements with absolute values greater than α, then $\|s\| < (1 + M)\, m\alpha$.

<u>*Lemma 2:*</u>

Let a function $f_\sigma(s)$ have the properties $f_\sigma(0) = 1$ and $\forall s \;\; 0 \le f_\sigma(s) \le 1$, and let $F_\sigma(s) = \sum_{i=1}^{m} f_\sigma(s_i)$. Assume *A* satisfies the URP, and let $S = \{s \mid As = x\}$. Assume that there exists a (sparse) solution $s^0 \in S$ for which $\|s\|_0 = k \le \frac{n}{2}$ (such a sparse solution is unique). Then, if for a solution $\hat{s} = \left(\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_m\right)^T \in S$, $F_\sigma(\hat{s}) \ge m - (n-k)$ and if $\alpha > 0$ is chosen such that the $\hat{s}_i$'s with absolute values greater than α satisfy $f_\sigma(\hat{s}_i) \le 1/m$, then $\left\|\hat{s} - s^0\right\| < (M+1)\, m\alpha$

<u>*Proof:*</u> Let I_α be the set of all indices *i*, for which $|\hat{s}_i| > \alpha$, then

$$
\begin{aligned}
F_\alpha(\hat{s}) &= \sum_{i=1}^{m} f_\alpha(\hat{s}_i) \\
&= \underbrace{\sum_{i\in I_\alpha} \underbrace{f_\alpha(\hat{s}_i)}_{\le (1/m)}}_{\le m.(1/m)=1} + \underbrace{\sum_{i\notin I_\alpha} \underbrace{f_\alpha(\hat{s}_i)}_{\le 1}}_{\le m-|I_\alpha|} \\
&\le 1 + m - |I_\alpha|
\end{aligned}
$$

We assume that we have chosen $f_\sigma(s)$ such that $F_\sigma(\hat{s}) \ge m-(n-k)$. (We prove this next)

Now, we get $m-(n-k) \le F_\sigma(\hat{s}) \le 1+m-|I_\alpha|$, from which we can get $|I_\alpha| \le n-k$.

As a result, at most *n-k* elements of $\hat{s}$ have absolute values greater than α.

Since s^0 has exactly *k* non-zero elements, we conclude that $\hat{s}-s^0$ has at most (*n-k*)+*k*=*n* elements with absolute values greater than α.

Moreover, $\hat{s}-s^0 \in N(A)$ and hence by Corollary 1 we have $\|\hat{s}-s^0\| < (M+1)m\alpha$.

<u>*Corollary 2:*</u> For the Gaussian family $f_\sigma(s) \triangleq \exp\left(\frac{-s^2}{2\sigma^2}\right)$, if $F_\sigma(\hat{s}) \ge m-(n-k)$ holds for a solution $\hat{s}$, then

$$\|\hat{s}-s^0\| < (M+1)m\sigma\sqrt{2\ln m}$$

<u>*Proof:*</u>

For the Gaussian family $f_\sigma(s) \triangleq \exp\left(\frac{-s^2}{2\sigma^2}\right)$, α required for lemma 2 can be chosen as $\alpha = \sigma\sqrt{2\ln m}$. Because for $|\hat{s}_i| > \sigma\sqrt{2\ln m}$,

$$f_\sigma(\hat{s}_i) = \exp\left(\frac{-\hat{s}_i^2}{2\sigma^2}\right) < \exp\left(\frac{-2\sigma^2\ln m}{2\sigma^2}\right) = \frac{1}{m}$$

Moreover, Gaussian family satisfies the other condition required in lemma 2.

<u>*Lemma 3:*</u> Let f_σ, F_σ, S and s^0 be as in Lemma 2, and let s^σ be the maximizer of $F_\sigma(s)$ on S, then s^σ satisfies $F_\sigma(\hat{s}) \ge m-(n-k)$.

<u>*Proof:*</u> We write

$$
\begin{aligned}
F_\sigma(s^\sigma) &\ge F_\sigma(s^0) \\
&\ge m-k \\
&\ge m-(n-k) \quad (\because\ k \le \tfrac{n}{2})
\end{aligned}
$$

Note that Lemma 3 and Corollary 2 prove together that for the Gaussian family $f_\sigma(s) \triangleq \exp\left(\frac{-s^2}{2\sigma^2}\right)$ $\arg\max_{As=x} F_\sigma(s) \to s^0$ as $\sigma \to 0$. This result can however be stated for a larger class of functions, as done in Theorem 1 (next page).

Theorem 1: Consider a family of univariate functions f_σ, indexed $\sigma \in R^+$ satisfying the following set of conditions:

1. $\lim_{\sigma \to 0} f_\sigma(s) = 0; \quad \forall s \neq 0$
2. $f_\sigma(0) = 1; \quad \forall \sigma \in R^+$
3. $0 \leq f_\sigma(s) \leq 1; \quad \forall \sigma \in R^+ \quad s \in R$
4. For each positive values of ν and α, there exists $\sigma_0 \in R^+$ that satisfies

$$|s| > \alpha \;\Rightarrow\; f_\sigma(s) < \nu; \quad \forall \sigma < \sigma_0 \qquad (1)$$

Assume *A* satisfies the URP, and let F_σ, *S* and s^0 be defined as in Lemma 2, and $s^\sigma = (s_1^\sigma, s_2^\sigma, \cdots, s_m^\sigma)^T \in S$ be the maximizer of F_σ on *S*. Then:

$$\lim_{\sigma \to 0} s^\sigma = s^0$$

Working definition of limit of a sequence

We say that $\lim_{n \to \infty} a_n = L$ if we can make a_n as close to *L* as we want for sufficiently large *n*.

Precise definition of the limit

We say that $\lim_{n \to \infty} a_n = L$ if for some positive error term ε the distance of the sequence at *n* from *L* must be less than the allowed error ε, that is, $|a_n - L| < \varepsilon$. But, it is important to remember that it is not enough that our sequence does converge once or twice; it must be within the error for all values from some point onwards, that is, $|a_n - L| < \varepsilon, \quad \forall n > N$.

Analytically, $\forall \varepsilon > 0, \quad \exists N \in \mathbb{N}, \quad \forall n > N \quad |a_n - L| < \varepsilon.$

Proof: To prove $\lim_{\sigma \to 0} s^\sigma = s^0$, we have to show that

$$\forall \beta > 0, \quad \exists \sigma_0 > 0, \quad \forall \sigma < \sigma_0 \qquad \left\| s^\sigma - s^0 \right\| < \beta. \qquad (2)$$

For each $\beta > 0$, let $\alpha = \beta / m(M+1)$. Then for this α and $\nu = 1/m$, condition 4 of the theorem gives a σ_0 for which the (1) holds. We show that this is the σ_0 we were seeking for in (2).

Note that $\forall \sigma < \sigma_0$, (1) states that for s_i^σ's with absolute values greater than α we have $f_\sigma(s_i^\sigma) \leq 1/m$. Moreover, Lemma 3 states that s^σ satisfies $F_\sigma(s^\sigma) \geq m - (n-k)$. Consequently, all the conditions of Lemma 2 have been satisfied, and hence it implies that $\left\| s^\sigma - s^0 \right\| < (M+1)m\alpha = \beta$.

Remark 1: The Gaussian family $f_\sigma(s) \triangleq \exp\left(\frac{-s^2}{2\sigma^2}\right)$ satisfies conditions 1 through 4 of Theorem 1. Other Families of functions also satisfy the conditions of Theorem 1.

Remark 2: Using Corollary 2, where using Gaussian family, to ensure an arbitrary accuracy in estimation of the sparse solution s^0, it suffices to choose

$$\sigma < \frac{\beta}{m(M+1)\sqrt{2\ln m}}$$ and do the optimization of $F_\sigma(s)$ subject to $A\boldsymbol{s}=\boldsymbol{x}$.

Remark 3: Consider the set of solutions in $\hat{s}^\sigma$ in *S*, which might not be the absolute maxima of functions F_σ on *S*, but satisfy the condition

$$F_\sigma(\hat{s}^\sigma) \geq m-(n-k)$$

By following a similar approach to the proof of Theorem 1, it can be proved that $\lim_{\sigma\to 0} s^\sigma = s^0$. In other words, for the steepest ascent (internal loop), it is not necessary to reach the absolute maximum. It is enough reach a solution in which is F_σ large.

Remark 4: The previous remark proposes another version of SL0 in which there is no need to set a parameter L: Repeat the internal loop until $F_\sigma(s)$ exceeds *m*-*n*/2[the worst case of the limit given by $F_\sigma(\hat{s}) \geq m-(n-k)$] or $m-(n-k)$ if *k* is known a priori. The advantage of such a version is that if it converges, then it is guaranteed that the estimation error is bounded as $\left\|\hat{s}-s^0\right\| < (M+1)m\sigma\sqrt{2\ln m}$, in which σ is replaced with σ_J, the last element of the sequence of σ.

It has, however, two disadvantages: first, it slows down the algorithm because exceeding the limit $m-(n-k)$ for each σ is not necessary (it is just sufficient); and second, because of the possibility that the algorithm runs into an infinite loop because $F_\sigma(s)$ cannot exceed this limit (this occurs if the chosen sequence of σ has not been resulted in escaping from local maxima).

Remark 5: As another consequence, Lemma 1 provides an upper bound on the estimation error $\left\|\hat{s}-s^0\right\|$, only by having an estimation $\hat{s}$ (which satisfies $A\hat{s}=x$): Begin by sorting the elements of $\hat{s}$ in descending order and α let be the absolute value of the $\lfloor \frac{n}{2} \rfloor$+1'th element. Since s^0 has at most *n*/2 nonzero elements, $\hat{s}-s^0$ has at most *n* elements with absolute values greater than α. Moreover, $\hat{s}-s^0 \in N(A)$ and hence Corollary 1 implies that $\left\|\hat{s}-s^0\right\| < (M+1)m\alpha$. This result is consistent with the heuristic that "if $\hat{s}$ has at most *n*/2 'large' components, the uniqueness of the sparsest solution insures that $\hat{s}$ is close to the true solution."

B. Relation to minimum norm 2 solution

Kindly refer the paper for proof of $\lim_{\sigma\to\infty} \arg\max_{As=x} F_\sigma(s) = \hat{s}$, where $\hat{s}$ is the minimum L2 norm solution.

C. Noisy case

As shown in the proof of Theorem 1 (noiseless case), a smaller value of σ results in a more accurate solution and it is possible to achieve solutions as accurate as desired by choosing small enough values of σ. However, this is not the case in the presence of additive noise, that is, if x=As+n. In fact, the noise power bounds the maximum achievable accuracy.

V. Numerical Results

The performance of the SL0 algorithm is experimentally verified and is compared with BP (FOCUSS) and LP (L1 magic). The effects of the parameters, sparsity, noise, and dimension on the performance are also experimentally discussed (Please refer the paper).

In experiments, sparse sources are artificially created using a Bernoulli–Gaussian model: each source is "active" with probability p, and is "inactive" with probability (1-p). If it is active, each sample is a zero-mean Gaussian random variable with variance σ^2_{on}; if it is not active, each sample is a zero-mean Gaussian random variable with variance σ^2_{off}, where $\sigma^2_{\text{off}} \ll \sigma^2_{\text{on}}$.

Each column of the mixing matrix is randomly generated using the normal distribution and then is normalized to unity.

To evaluate the estimation quality, signal-to-noise ratio (SNR) and mean-square error (MSE) are used. SNR (in dB) is defined as $20\log\left(\frac{\|s\|}{\|s-\hat{s}\|}\right)$ and MSE as $\frac{1}{m}\|s-\hat{s}\|^2$

TABLE I
Progress of SL0 for a Problem With $m = 1000$, $n = 400$ and $k = 100$ $(p = 0.1)$

itr. #	σ	MSE	SNR (dB)
1	1	$4.84\,e-2$	2.82
2	0.5	$2.02\,e-2$	5.19
3	0.2	$4.96\,e-3$	11.59
4	0.1	$2.30\,e-3$	16.44
5	0.05	$5.83\,e-4$	20.69
6	0.02	$1.17\,e-4$	28.62
7	0.01	$5.53\,e-5$	30.85
algorithm	**total time (sec)**	**MSE**	**SNR (dB)**
SL0	0.227	$5.53\,e-5$	30.85
LP (ℓ_1-magic)	30.1	$2.31\,e-4$	25.65
FOCUSS	20.6	$6.45\,e-4$	20.93

VI. CONCLUSIONS

In this paper, authors showed that the smoothed L0 norm can be used for finding sparse solutions of an USLE. They also showed that the smoothed version of the L0 norm results in an algorithm which is faster than the state-of-the-art algorithms based on minimizing the L1 norm.

Moreover, this smoothing solves the problem of high sensitivity of L0 norm to noise. In another point of view, the smoothed L0 provides a smooth measure of sparsity.

The basic idea of the paper was justified by both theoretical (convergence in both noiseless and noisy case, relation to the L2 norm solution) and experimental analysis of the algorithm.

Appendix (Not available in the paper)

The following are taken from [5]

Consider the following problem of the Euclidean orthogonal projection of a point to an affine set: For the given $A \in R^{m \times n}$, $b \in R^m$ and $p \in R^n$, find a vector $x^* \in R^n$ satisfying

$$Ax^* = b$$
$$\left\| p - x^* \right\| = \min_{Ax=b} \left\| p - x \right\|$$

The solution to the above problem exits and it is unique and it is

$$x^* = p - A^+ Ap + A^+ b \qquad [x^* = p - A^+(Ap - b)]$$
$$= \left[I - A^+ A \right] p + A^+ b$$
$$x^* = P_{N(A)} p + A^+ b$$

Reference

[1] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[2] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," in *Signal Process.*, 2001, vol. 81, pp. 2353–2362.

[3] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS, A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientif. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[5] J. Ding, "Perturbation analysis for the projection of a point into the affine set," *Linear Algebra and its applications*, vol. 191, pp. 199-212, 1993.

Expectation-Maximization Belief Propagation for Sparse Recovery I: Algorithm Construction

Main ref.: F. Krzakala *et al.*,"Statistical physics-based re construction in compressed sensing," *Phys. Rev X*, May, 2012 [1]

Presenter: Jaewook Kang

In Proceeding of CS Journal Club in GIST

Date: Nov. 2012, updated at Jan, 2013

I. Introduction

In this report, the author introduces a expectation maximization (EM) based belief propagation algorithm (BP) for sparse recovery, named EM-BP. The algorithm have been mainly devised by Krzakala *et al.* from ParisTech in France [1]. The properties of EM-BP are as given below:

1) It is A low-computation approach to sparse recovery,
2) It works well without the prior knowledge of the signal,
3) It overcomes the l_1 phase transition given by Donoho and Tanner [11] under the noiseless setup,
4) It is further improved in conjunction with seeding matrices (or spatial coupling matrices).

The main purpose of this report regenerates a precise description of EM-BP algorithm construction from the reference paper [1]. It might be very helpful for understanding of EM-BP algorithm, and an answer for such a question: How and why does the algorithm work ? Therefore, we will focus on the explanation of 1) and 2) in the properties, and just show the result of the paper with respect to that of 3) and 4).

In addition to EM-BP, the belief propagation approach to the sparse recovery problem has been widely investigated in [2],[3],[4],[5],[6],[7].

II. Problem setup

In the sparse recovery problem, the aim is to recovery a sparse signal $\mathbf{X} \in \mathbb{R}^N$ whose elements have nonzero value independently each other, with a probability rate q called sparsity rate. Therefore, the q determines the density of signal $\mathbf{X}$. Then, the algorithm performs the recovery from the measurements $\mathbf{Y} \in \mathbb{R}^M$, given as

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{N}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a fat measurement matrix with $M < N$, and $\mathbf{N} \in \mathbb{R}^{M}$ denotes a additive Gaussian noise vector following $\mathcal{N}(0, \mathbf{I}\sigma_N^2)$.

III. Algorithm Construction of EM-BP

Krzakala *et al.* has taken a probabilistic approach to devise EM-BP. From the Bayesian point of view, the posterior density of the signal $\mathbf{X}$ is represented in the form of $\text{Posterior} = \text{Prior} \times \frac{\text{Likelihood}}{\text{Evidence}}$ as

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi}) = f_{\mathbf{X}}(\mathbf{x}|\mathbf{\Phi}) \times \frac{f_{\mathbf{Y}}(\mathbf{y}|\mathbf{\Phi}, \mathbf{X})}{f_{\mathbf{Y}}(\mathbf{y}|\mathbf{\Phi})}. \tag{2}$$

Then, using the knowledge of $\mathbf{z}$ and $\mathbf{\Phi}$, the signal posterior is given as

$$f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi}) = \frac{1}{C} f_{\mathbf{X}}(\mathbf{x}) \times \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left[-\frac{1}{2\sigma_{N_j}^2} (y_j - \textstyle\sum_{i=1}^{N} \phi_{ji} x_i)^2 \right], \tag{3}$$

where C is a normalization constant for $\int f_{\mathbf{X}}(\mathbf{x}|\mathbf{y}, \mathbf{\Phi})\, d\mathbf{x} = 1$. In addition, we consider a mixture type prior density function represented as

$$f_{\mathbf{X}}(\mathbf{x}) := \prod_{i=1}^{N} [(1-q)\delta_0 + q\theta(x_i)], \tag{4}$$

where $\theta(x_i)$ is a Gaussian PDF with mean $\bar{x}$ and variance σ_X^2.

Exact finding of the signal posterior is computationally infeasible. Therefore, researchers have employed BP as a standard approach to approximate the signal posterior where BP finds marginal posterior density of each signal element X_i. In addition, Guo *et al.* showed that the marginal posterior finding is exact if the matrix $\mathbf{\Phi}$ is a sparse matrix and $N \to \infty$ [8],[9],[10]. BP seeks the signal posterior by iteratively exchanging probabilistic messages over the signal elements, where the messages are classically described as

Measurement to signal (MtS) message :

$$m_{j \to i}(x_i) := \frac{1}{C_{j \to i}} \int_{\{x_k\}_{k \neq i}} \prod_{k \neq i} m_{k \to j}(x_i) \times \exp\left[-\frac{1}{2\sigma_N^2} (\sum_{k \neq i} \phi_{jk} x_k + \phi_{ji} x_i - y_j)^2 \right] \left(\prod_{k \neq i} dx_i \right), \tag{5}$$

Signal to measurement (StM) message :

$$m_{i \to j}(x_i) := \frac{1}{Z_{i \to j}} [(1-q)\delta_0 + q\theta(x_i)] \times \prod_{k \neq j} m_{k \to i}(x_i), \tag{6}$$

where $C_{j \to i}$ and $Z_{i \to j}$ are normalization constants to make the messages as PDFs. Then, the marginal posterior approximately is obtained as

$$f_{X_i}(x|\,\mathbf{y}, \mathbf{\Phi}) \overset{\text{BP}}{=} \frac{1}{C_i} [(1-q)\delta_{x_i} + q\theta(x_i)] \times \prod_{k} m_{k \to i}(x_i). \tag{7}$$

However, the message update rule in (5) and (6) is practically intractable because each BP message is probability density function (PDF). Therefore, we need to convert the density-passing procedure to a parameter-passing procedure using some relaxation techniques.

Using Hubbard-Stratonovich transformation (HST) from spin glass theory which is

$$\exp\left(-\frac{w^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\int \exp\left(-\frac{\lambda^2}{2\sigma^2}+\frac{iw\lambda}{\sigma^2}\right)d\lambda, \tag{8}$$

the exponent in (5) can be rewritten as

$$\exp\left[-\frac{1}{2\sigma_{N_j}^2}(y_j-\sum\nolimits_{i=1}^{N}\phi_{ji}x_i)^2\right] = \exp\left[\underbrace{-\frac{\left(\sum\limits_{k\neq i}\phi_{jk}x_k\right)^2}{2\sigma_{N_j}^2}}_{\text{Here, HST applied}} - \frac{\sum\limits_{k\neq i}\phi_{jk}x_k(\phi_{ji}x_i-y_j)}{\sigma_{N_j}^2} - \frac{(\phi_{ji}x_i-y_j)^2}{2\sigma_{N_j}^2}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma_{N_j}^2}}\int_{\lambda}\exp\left(-\frac{\lambda^2}{2\sigma_{N_j}^2}+\frac{\sum\limits_{k\neq i}\phi_{jk}x_k(\phi_{ji}x_i-y_j+i\lambda)}{\sigma_{N_j}^2}-\frac{(\phi_{ji}x_i-y_j)^2}{2\sigma_{N_j}^2}\right)d\lambda. \tag{9}$$

By applying (9) to (5), we have

$$\begin{aligned} m_{j\to i}(x_i) = {} & \frac{\exp\left(-\frac{(\phi_{ji}x_i-y_j)^2}{2\sigma_{N_j}^2}\right)}{C_{j\to i}\sqrt{2\pi\sigma_{N_j}^2}}\int_{\lambda}\exp\left(-\frac{\lambda^2}{2\sigma_{N_j}^2}\right) \\ & \times\left\{\int_{\{x_k\}_{k\neq i}}\prod_{k\neq i}m_{k\to j}(x_i)\exp\left(\frac{\sum\limits_{k\neq i}\phi_{jk}x_k(\phi_{ji}x_i-y_j+i\lambda)}{\sigma_{N_j}^2}\right)\prod_{k\neq i}dx_i\right\}d\lambda \end{aligned} \tag{10}$$

In (10), we observe that the integration over $\{x_k\}_{k\neq i}$ can be decomposed into integration over each scalar x_k. In addition, the integration over scalar x_k takes the form of the moment generating function. Therefore,

$$\begin{aligned} m_{j\to i}(x_i) &= \frac{\exp\left(-\frac{(\phi_{ji}x_i-y_j)^2}{2\sigma_{N_j}^2}\right)}{C_{j\to i}\sqrt{2\pi\sigma_{N_j}^2}}\int_{\lambda}\exp\left(-\frac{\lambda^2}{2\sigma_{N_j}^2}\right)\times\prod_{k\neq i}\left\{\int_{\{x_k\}_{k\neq i}}m_{k\to j}(x_i)\exp\left(\frac{x_k\phi_{jk}(\phi_{ji}x_i-y_j+i\lambda)}{\sigma_{N_j}^2}\right)dx_i\right\}d\lambda \\ &= \frac{\exp\left(-\frac{(\phi_{ji}x_i-y_j)^2}{2\sigma_{N_j}^2}\right)}{C_{j\to i}\sqrt{2\pi\sigma_{N_j}^2}}\int_{\lambda}\exp\left(-\frac{\lambda^2}{2\sigma_{N_j}^2}\right)\times\prod_{k\neq i}\mathbf{E}_{X_k}\left[\exp\left(\frac{x_k\phi_{jk}(\phi_{ji}x_i-y_j+i\lambda)}{\sigma_{N_j}^2}\right)\right]d\lambda \end{aligned} \tag{11}$$

By assuming that each scalar X_k is Gaussian distributed during the BP-iteration with mean $\mu_{i\to j}$ and

variance $\sigma^2_{i\to j}$, we can approximate the MtS message expression as

$$m_{j\to i}(x_i) \approx \frac{\exp\left(-\frac{(\phi_{ji}x_i - y_j)^2}{2\sigma^2_{N_j}}\right)}{C_{j\to i}\sqrt{2\pi\sigma^2_{N_j}}} \times \\ \times \int_{\lambda} \exp\left(-\frac{\lambda^2}{2\sigma^2_{N_j}}\right) \prod_{k\neq i} \exp\left(\frac{\mu_{k\to j}\phi_{jk}(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}} + \frac{\sigma^2_{k\to j}}{2}\left(\frac{\phi_{jk}(\phi_{ji}x_i - y_j + i\lambda)}{\sigma^2_{N_j}}\right)^2\right) d\lambda. \tag{12}$$

By evaluating the Gaussian integration over λ, the expression in (12) becomes

$$m_{j\to i}(x_i) \simeq \frac{\sqrt{A_{j\to i}/2\pi}}{\phi_{ji}C_{j\to i}} \times \exp\left(-\frac{x_i^2}{2}A_{j\to i} + x_i B_{j\to i} + \frac{B^2_{j\to i}}{2A_{j\to i}}\right). \tag{13}$$

where

$$A_{j\to i} := \frac{\phi^2_{ji}}{\sigma^2_{N_j} + \sum\limits_{k\neq j}\sigma^2_{k\to j}\phi^2_{jk}}, \tag{14}$$

$$B_{j\to i} := \frac{\phi_{ji}(y_j - \sum\limits_{k\neq j}\mu_{k\to j}\phi_{jk})}{\sigma^2_{N_j} + \sum\limits_{k\neq j}\sigma^2_{k\to j}\phi^2_{jk}}. \tag{15}$$

Then, the expression of the StM message is rewritten as

$$m_{i\to j}(x_i) := \frac{1}{\widetilde{Z}_{i\to j}}[(1-q)\delta_0 + q\theta(x_i)] \times \exp\left(-\frac{x_i^2}{2}\sum_{k\neq j}A_{k\to i} + x_i\sum_{k\neq j}B_{k\to i} + \frac{1}{2}\frac{\left(\sum\limits_{k\neq j}B_{k\to j}\right)^2}{\sum\limits_{k\neq j}A_{k\to j}}\right), \tag{16}$$

where we use an approximation $\sum\limits_{k\neq j}B^2_{k\to j} \approx \left(\sum\limits_{k\neq j}B_{k\to j}\right)^2$. The exponent can be rewritten as

$$-\frac{x_i^2}{2}\sum_{k\neq j}A_{k\to i} + x_i\sum_{k\neq j}B_{k\to i} + \frac{1}{2}\frac{\left(\sum\limits_{k\neq j}B_{k\to j}\right)^2}{\sum\limits_{k\neq j}A_{k\to j}}$$

$$= -\frac{1}{2\frac{1}{\sum\limits_{k\neq j}A_{k\to i}}}\left(x_i^2 - 2\frac{\sum\limits_{k\neq j}B_{k\to i}}{\sum\limits_{k\neq j}A_{k\to i}} + \left(\frac{\sum\limits_{k\neq j}B_{k\to i}}{\sum\limits_{k\neq j}A_{k\to i}}\right)^2\right) = -\frac{\left(x_i - \frac{\sum\limits_{k\neq j}B_{k\to i}}{\sum\limits_{k\neq j}A_{k\to i}}\right)^2}{2\frac{1}{\sum\limits_{k\neq j}A_{k\to i}}} \tag{17}$$

Hence, equations (14) and (15) together with (19) fully describe the iterative BP-process. We define two variable given as

$$\Sigma_i^2 := \frac{1}{\sum\limits_{k\neq j} A_{k\to i}},\ R_i := \frac{\sum\limits_{k\neq j} B_{k\to i}}{\sum\limits_{k\neq j} A_{k\to i}}, \tag{18}$$

Using the notations, we rewrite the expression of the StM message given as Then, the expression of the StM message is rewritten as

$$m_{i\to j}(x_i) := \frac{1}{\widetilde{Z}_{i\to j}}[(1-q)\delta_0 + q\theta(x_i)] \times \exp\left(-\frac{(x_i-R_i)^2}{2\Sigma_i^2}\right), \tag{19}$$

Then, the mean $\mu_{k\to j}$ and variance $\sigma_{k\to j}^2$ of the StM message are calculated as

$$\begin{aligned}
\mu_{i\to j} &:= \int\limits_{X_i} x_i m_{i\to j}(x_i)dx_i \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \int\limits_{X_i} x_i\theta(x_i)\exp\left(-\frac{(x_i-R_i)^2}{2\Sigma_i^2}\right)dx_i \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \times \frac{\Sigma_i(\overline{x}\Sigma_i^2 + R\sigma_X^2)}{(\Sigma_i^2+\sigma_X^2)^{3/2}} \exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right),
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
\sigma_{i\to j}^2 &:= \int\limits_{X_i} x_i^2 m_{i\to j}(x_i)dx_i - \mu_{i\to j}^2 \\
&= \frac{q}{Z(\Sigma_i^2, R_i)} \int\limits_{X_i} x_i^2\theta(x_i)\exp\left(-\frac{(x_i-R_i)^2}{2\Sigma_i^2}\right)dx_i - \mu_{i\to j}^2 \\
&= \frac{q(1-q)\exp\left(-\frac{R_i^2}{2\Sigma_i^2} - \frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right)\frac{\Sigma_i}{(\Sigma_i^2+\sigma_X^2)^{5/2}}\left(\sigma_X^2\Sigma_i^2(\Sigma_i^2+\sigma_X^2) + (\overline{x}\Sigma_i^2 + R\sigma_X^2)^2\right)}{Z(\Sigma_i^2, R_i)^2} \\
&+ \frac{q^2\exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right)\frac{\sigma_X^2\Sigma_i^4}{(\Sigma_i^2+\sigma_X^2)^2}}{Z(\Sigma_i^2, R_i)^2},
\end{aligned} \tag{21}$$

where the normalization constant is

$$\begin{aligned}
Z(\Sigma_i^2, R_i) :&= (1-q)\int\limits_{X_i} \delta_0 \exp\left(-\frac{(x_i-R_i)^2}{2\Sigma_i^2}\right)dx_i + q\int\limits_{X_i} \theta(x_i)\exp\left(-\frac{(x_i-R_i)^2}{2\Sigma_i^2}\right)dx_i \\
&= (1-q)\exp\left(-\frac{R_i^2}{2\Sigma_i^2}\right) + q\frac{\Sigma_i}{\sqrt{\Sigma_i^2+\sigma_X^2}}\exp\left(-\frac{(R-\overline{x})^2}{2(\Sigma_i^2+\sigma_X^2)}\right).
\end{aligned} \tag{22}$$

The authors stated that the parameters $\bar{x}$, σ_X^2, and q of the prior density $f_{\mathbf{X}}(\mathbf{X})$ can be learned and updated at every iteration. A statistical approach for the parameter learning is the use of EM. For the object

function in EM, they used Bethe free-entropy. It is known that BP algorithm is constructed by applying Lagrange multipliers to Bethe entropy [12]. Therefore, the fixed point of BP-iteration corresponds to the stationary points of the Bethe free-entropy minimization, in the signal posterior finding problems. For details about the relationship between Bethe free-entropy and BP, please see Yedidia's paper.

The Bethe entropy is defined as

$$H_{\text{Bethe}} := -\sum_{i}^{N} H(Z_{x_i}) - \sum_{j}^{M} H(Z_{y_j}) + \sum_{j}^{M} \sum_{i \in N(j)} H(Z_{x_i}), \tag{23}$$

where the concept of free-entropy, defined as $H(Z) := \log Z$, is used and Z_{x_i} and Z_{y_j} are an approximated marginal partition function of x, that is,

$$Z_{x_i} = \int [(1-q)\delta_{x_i} + q\theta(x_i)] \times \prod_{j} m_{j \to i}(x_i) dx_i, \tag{24}$$

$$Z_{y_j} = \int \prod_{i} m_{i \to j}(x_i) \times \exp\left[-\frac{1}{2\sigma_N^2} (\sum_{i} \phi_{ji} x_i - y_j)^2 \right] \prod_{i} (dx_i)\,. \tag{25}$$

Thus, the parameters $(\bar{x}, \sigma_X, q)$ are learned by seeking the stationary point of the Bethe free-entropy function given in (23). We update the parameter for the prior knowledge from

$$\overline{x} = \frac{\sum_{i} \mu_i}{Nq} \tag{26}$$

$$\sigma_X^2 = \frac{\sum_{i} (\sigma_i^2 + \mu_i^2)}{Nq} - \overline{x}^2 \tag{27}$$

$$q = \frac{\sum_i \frac{1/\sigma_X^2 + \sum_j A_{j \to i}}{\sum_j B_{j \to i} + \overline{x}/\sigma_X^2} \mu_i}{\sum_i \left(1 - q + \frac{q}{\sigma_X \sqrt{1/\sigma_X^2 + \sum_j A_{j \to i}}} \exp\left(\frac{(\sum_j B_{j \to i} + \overline{x}/\sigma_X^2)^2}{2(1/\sigma_X^2 + \sum_j A_{j \to i})} - \frac{\overline{x}^2}{2\sigma_X^2} \right) \right)^{-1}}. \tag{28}$$

I implemented the EM-BP algorithm using the equations of (14), (15), (20), (21), (22), (26), (27) in C language. I did not update the sparsity rate q in the BP-iteration. The performance is not working well as shown in Fig.1. I need to check my implementation by translating the code to MATLAB. I think the EM update not much improve the performance. So, we need to modify the update rule to elementwise update rule like SuPrEM Algorithm.

REFERENCES

[1] F. Krzakala, M. Mezard, F. Sausset, Y. Sun, L. Zdeborova, "Statistical physics-based reconstruction in compressed sensing," *Phys. Rev X*, No. 021005, May, 2012

[2] J. Kang, H.-N. Lee, and K. Kim, "Detection-directed sparse estimation using bayesian hypothesis test and belief propagation," submitted to *IEEE Trans. Signal Process.*, 2012

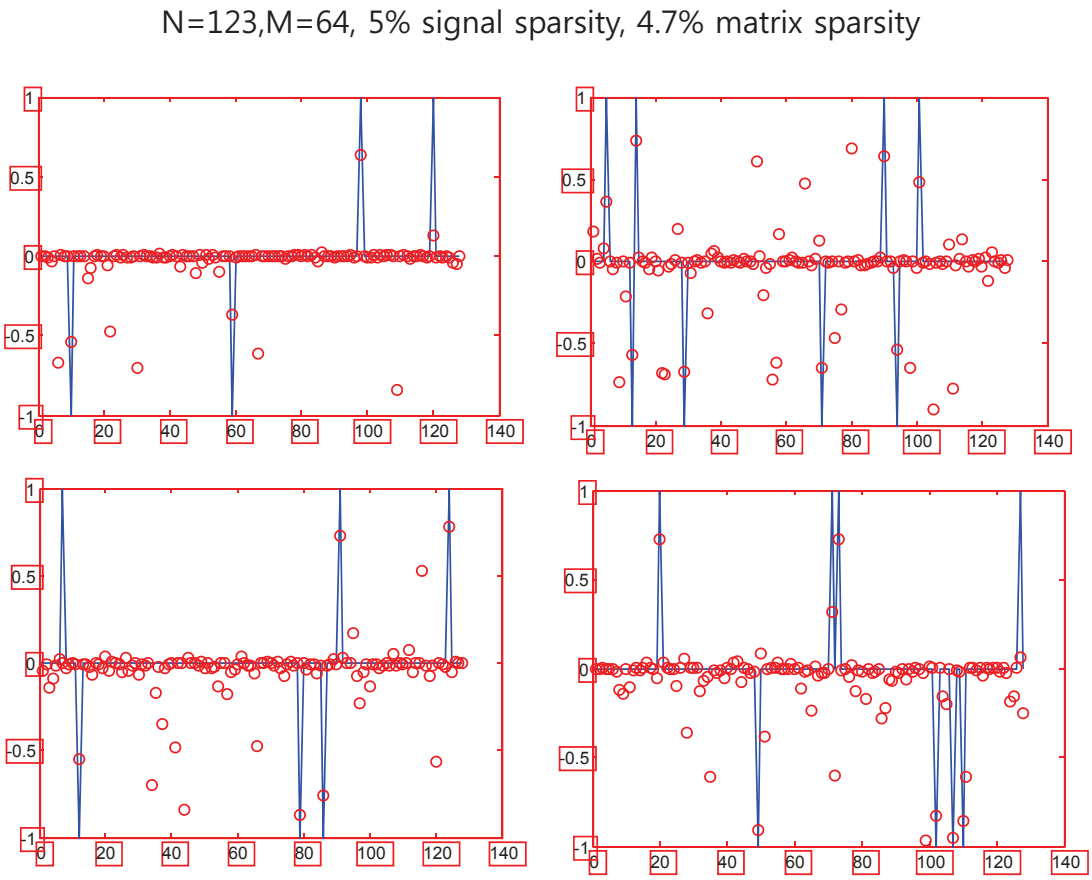


Fig. 1. Some simulation results of EM-BP when $N = 128, M = 64, q = 0.05$, $\sigma_X = 1$, and $L = 3$.

[3] D. Baron, S. Sarvotham, and R. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269-280, Jan. 2010.

[4] X. Tan and J. Li, "Computationally efficient sparse Bayesian learning via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2010-2021, Apr. 2010.

[5] M. Akcakaya, J. Park, and V. Tarokh, "A coding theory approach to noisy compressive sensing using low density frame," *IEEE Trans. Signal Process.,* vol. 59, no. 12, pp. 5369-5379, Nov. 2011.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing,", *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914-18919, Nov. 2009.

[7] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *Proc. in IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 2168-2172, Aug. 2011.

[8] D. Guo and S. Verdu, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983-2010, Jun. 2005.

[9] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," *Proc. in*

IEEE Inform. Theory Workshop, pp. 194-198, Chengdu, China, Oct. 2006.

[10] D. Guo and C.-C. Wang, "Random sparse linear systems observed via arbitrary channels: a decoupling principle,"*Proc. in IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 946-950, Nice, France, June. 2007.

[11] D.L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceeding of the IEEE*, vol. 98, issue 6, pp. 913-924, June, 2010.

[12] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Construcing Free-Energy Approximations and Generalized Belief Propagation Algorithm," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2282-2312, July. 2005.

Performance Analysis of Iterative Decoding Algorithms with Memory over Memoryless Channels

Authors: Emil Janulewics, Amir H. Banihashemi

Publication: IEEE T. Comm, Dec 2012
Speaker: Jeong-Min Ryu

Short summary:

In this work, they propose a model for iterative decoding algorithms with **memory** which covers successive relaxation (SR) version of belief propagation and differential decoding with binary message passing (DD-BMP) algorithms as special cases. Based on this model, they derive a Bayesian network for iterative algorithms with memory over memoryless channels and use this representation to analyze the performance of the algorithms using density evolution.

I. INTRODUCTION

Iterative decoding algorithm → Decoding algorithm of LDPC codes

Low-density parity-check (LDPC) codes are known to have good performance when decoded with iterative decoding algorithms, also known as message-passing algorithms.

Density Evolution → An analytical tool of LDPC codes

An analytical tool called *density evolution* can be used to find the threshold of a particular code ensemble under a given iterative decoding algorithm. The threshold is an asymptotic measure of performance and is defined as the worst channel parameter (e.g., largest noise variance) for which the probability of error still converges to zero as the number of iterations tends to infinity

Density Evolution → A technique for constructing irregular LDPC codes

Density evolution is also a powerful technique for constructing irregular LDPC codes through the optimization of the degree distributions.

All message-passing algorithms analyzed by density Evolution → Memoryless

To the best of our knowledge however, all the message-passing algorithms analyzed by density evolution in the literature are memoryless, i.e., the output message of a variable node (check node) at iteration l is only a function of the input messages to that node at iteration l (l -1) and also of the initial message of the channel in the case of variable nodes.

Iterative decoding algorithms with memory → Exist

There exist however a number of iterative decoding algorithms, such as successive relaxation (SR) variants of BP and MS and DD-BMP (differential decoding with binary message-passing), that have memory.

The presence of memory in algorithms → Improves the performance but makes the density evolution analysis much more complex.

In this paper, they develop the framework for the density evolution analysis of iterative extrinsic message-passing algorithms with memory which includes DD-BMP and SR algorithms.

They employ the *Bayesian network representation* via a directed acyclic graph (DAG), to capture the dependences among different messages and memory contents in a space with two dimensions: iteration l and the depth of the decoding tree d.

Independent

Incoming messages to a node along different edges

Dependencies

A message passed along a given edge at iteration l

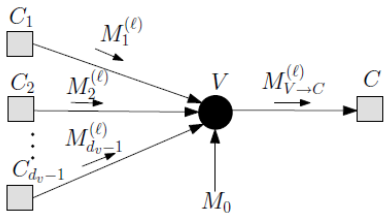
All the messages passed along that edge at previous iteration $l'<l$.

Such dependencies cause the complexity of density evolution to grow at least exponentially with l. They derive the density evolution equations and use techniques to make them tractable.

II. ITERATIVE DECODING ALGORITHMS WITH MEMORY

A. General Model

1) Memoryless decoding algorithm



(a) Memoryless

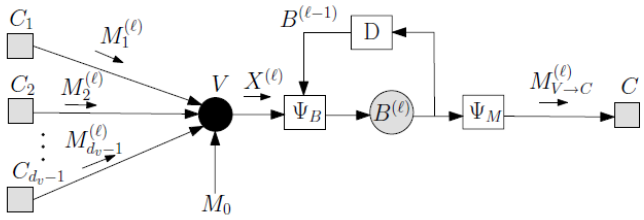
The figure (a) shows a snapshot of the Tanner graph of an LDPC code at iteration l for a memoryless decoding algorithm, where variable and check nodes are represented by circles and squares, respectively.

Under cycle-free assumption and based on the principle of extrinsic message passing,

- Incoming messages $M_1^{(l)},\ldots,M_{d_v-1}^{(l)}$ to node V are independent of each other and of the channel message M_0.
- The outgoing message $M_{V\to C}^{(l)}$ of node V to node C at iteration l is a function of d_v-1 i.i.d. random variables and the channel message.
- The outgoing message of a check node is a function of d_c-1 i.i.d. random variables corresponding to the extrinsic incoming message.

→ The distribution of a function of independent random variables is relatively easy to find since the joint distribution of these variables is the product of their marginal distribution. In this case, one can recursively derive the distribution of messages at iteration l as a deterministic function of the distribution at iteration l-1 with a complexity that is independent of l.

2) Memory decoding algorithm



(b) With Memory

Similar to figure (a), we have a set of i.i.d. extrinsic incoming messages to a variable node *V*.

The outgoing message from V: $X^{(l)}$

Memory units: Ψ_B, Ψ_M, $B^{(l)}$, D.

$B^{(l)}$ is updated by $\Psi_B\left(X^{(l)}, B^{(l-1)}\right)$, where Ψ_B is a deterministic function of $X^{(l)}$ and the content of the memory at iteration *l*-1.

The incoming message $M_{V\to C}^{(l)}$ to node *C* from *V* is obtained by $\Psi_M\left(B^{(l)}\right)$

Note that while the message $X^{(l)}$ is a function of independent random variables, the outgoing messages, $M_{V\to C}^{(l)}$, is a function of dependent random variables $X^{(l)}$ and $B^{(l-1)}$.

Our focus will be on the link from variable nodes to check nodes and on finding the distribution of $B^{(l)}$ and $M_{V\to C}^{(l)}$.

B. SR and DD-BMP Algorithms

1) SR Algorithms: Any standard memoryless iterative algorithm, such as BP or MS, can be turned into an SR algorithm by proper introduction of memory. SR algorithms can be performed in different message domains. In this work, we assume log-likelihood ratio (LLR) domain for messages (SRLLR). Based on the model of Fig. 1(b), the SR version is defined by the following variable node map

$$\begin{aligned} B^{(l)} &= \Psi_B\left(B^{(l-1)}, X^{(l)}\right) = (1-\beta)B^{(l-1)} + \beta X^{(l)}, \\ M_{V\to C}^{(l)} &= \Psi_M\left(B^{(l)}\right) = B^{(l)}, \end{aligned} \tag{1}$$

where $X^{(l)} = \Psi_V\left(M_0, M_1^{(l)}, ..., M_{d_v-1}^{(l)}\right)$. In (1), β is called the *relaxation factor*, and can be optimized for the best performance. The optimal value of β is usually in the interval (0, 1).

2) DD-BMP: Differential decoding with binary message passing (DD-BMP) was introduced as an attractive alternative to purely hard-decision algorithms. This algorithm combines the simplicity of binary message-passing with the good performance of soft-decision algorithms, where the soft information is stored in edge- or node-based memories. In the former case, studied in this paper, the variable node map, following the model of Fig. 1(b), is defined by

$$\begin{aligned} B^{(l)} &= \Psi_B\left(B^{(l-1)}, X^{(l)}\right) = B^{(l-1)} + X^{(l)}, \\ M_{V\to C}^{(l)} &= \Psi_M\left(B^{(l)}\right) = \mathrm{sgn}_r\left(B^{(l)}\right), \end{aligned} \tag{2}$$

where $\mathrm{sgn}_r(x) = 1$ for $x > 0$, and $= -1$ for $x < 0$. For $x = 0$, $\mathrm{sgn}_r(x)$ takes +1 or -1 randomly with equal probability. In (2), $X^{(l)} = \Psi_V\left(M_0, M_1^{(l)}, ..., M_{d_v-1}^{(l)}\right)$, which for the BIAWGN channel reduces to $X^{(l)} = \sum_{i=1}^{d_v-1} M_i^{(l)}$.

Both the variable and the check node operations (particularly the latter) are simpler for DD-BMP compared to BP and MS algorithms.

C. Symmetry of the Decoder and Error Probability

The analysis of iterative decoders is greatly simplified assuming that both the channel and the decoder are symmetric.

In particular, the variable node symmetry condition has some implications on the choices of the mappings Ψ_B and Ψ_M: Ψ_B should be sign inversion invariant, and $\Psi_M(-x) = -\Psi_M(x)$. As it can be seen in (1) and (2), both conditions are satisfied for SRLLR and DD-BMP algorithms.

With both the channel and the decoder being symmetric, we can assume, without loss of generality, that the all-zero codeword is transmitted. In this case, the average fraction of incorrect messages passed at iteration *l* from variable nodes to check nodes is calculated by

$$P_e^{(l)} = P\left(B^{(l)} < 0\right) + \frac{1}{2} P\left(B^{(l)} = 0\right). \quad (3)$$

We refer to $P_e^{(l)}$ in (3) as the probability of bit error at iteration *l*.

III. Bayesian Network Representation of Iterative Decoding Algorithms with Memory

A. Bayesian Networks and Conditional Independence

In this work, they use a Bayesian network to represent the dependencies among different messages and memory contents of an iterative algorithm with memory.

The *conditional independence* between two sets of random variables $\mathcal{X}$ and $\mathcal{Y}$ given a third set $\mathcal{Z}$ is defined by

$$\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z} \Leftrightarrow \mathbb{P}\left(\mathcal{X}, \mathcal{Y} \mid \mathcal{Z}\right) = \mathbb{P}\left(\mathcal{X} \mid \mathcal{Z}\right) \mathbb{P}\left(\mathcal{Y} \mid \mathcal{Z}\right)$$

where $\mathbb{P}\left(\mathcal{X}\right)$ and $\mathbb{P}\left(\mathcal{X} \mid \mathcal{Z}\right)$ are the marginal distribution of $\mathcal{X}$ and the conditional distribution of $\mathcal{X}$ given $\mathcal{Z}$, respectively.

B. Bayesian Networks of Iterative Decoders with Memory

Based on the principle of extrinsic message-passing, one can see that $X^{(l)}$ is a deterministic function of $B_1^{(l-1)}$ and M_0. Moreover, as it can be seen in Fig. 1(b), $B^{(l)}$ is a function of $B^{(l-1)}$ and $X^{(l)}$. In addition, $B_i^{(l)}$ depends on $B_i^{(l-1)}$ and $B_{i+1}^{(l-1)}$.

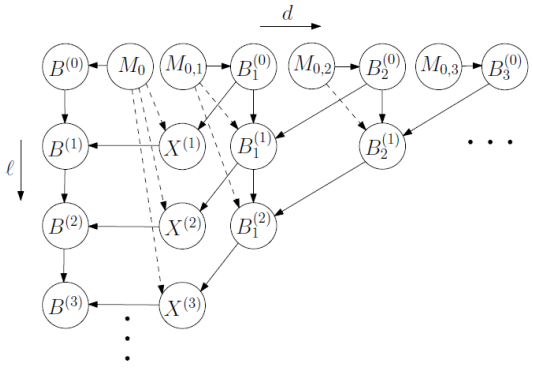


Fig. 3. Bayesian network based on the model of Fig. 1(b).

IV. DENSITY EVOLUTION

Based on (3), to obtain the error probability $P_e^{(l)}$ at each iteration l, we need to compute $\mathbb{P}\left(B^{(k)}\right)$.

A efficient approach is to compute $\mathbb{P}\left(B^{(k)}\right)$ is:

$$P\left(B^{(l)}=b\right)=\sum_{(b',x)\in S_B\times S_X : b=\Psi_B(b',x)} P\left(B^{(l-1)}=b', X^{(l)}=x\right),\ \forall b\in S_B\ .$$

where S_B is sample space of $B^{(l)}$ and $\mathbb{P}\left(B^{(l-1)}, X^{(l)}\right)$ is

$$\begin{aligned}\mathbb{P}\left(B^{(l-1)}, X^{(l)}\right)&=\sum_{x^{(1\to l-1)}}\mathbb{P}\left(B^{(l-1)}\middle| X^{(1\to l-1)}\right)\mathbb{P}\left(X^{(l)}\middle| X^{(1\to l-1)}\right)\mathbb{P}\left(X^{(1\to l-1)}\right)\\&=\sum_{x^{(1\to l-1)}}\mathbb{P}\left(B^{(l-1)}\middle| X^{(1\to l-1)}\right)\mathbb{P}\left(X^{(1\to l)}\right)\end{aligned}$$

where $X^{(1\to k)}=\left\{X^{(1)}, X^{(2)},\cdots, X^{(k)}\right\}$ for $k\geq 1$.

- Calculation of $\mathbb{P}\left(B^{(l)}\middle| X^{(1\to l)}\right)$

$$\begin{aligned}\mathbb{P}\left(B^{(l)}\middle| X^{(1\to l)}\right)&=\sum_{b^{(l-1)}}\mathbb{P}\left(B^{(l)}\middle| X^{(1\to l)}, B^{(l-1)}\right)\mathbb{P}\left(B^{(l-1)}\middle| X^{(1\to l)}\right)\\&=\sum_{b^{(l-1)}}\mathbb{P}\left(B^{(l)}\middle| X^{(l)}, B^{(l-1)}\right)\mathbb{P}\left(B^{(l-1)}\middle| X^{(1\to l-1)}\right),\ \ l\geq 2\end{aligned}$$

- Calculation of $\mathbb{P}\left(X^{(1\to l)}\right)$

The variables $M_i^{(l)}$ are i.i.d.. Since $X^{(l)}=\Psi_v\left(M_1^{(l)}, M_2^{(l)},\ldots, M_{d_v-1}^{(l)}\right)$, $X^{(l)}$ is conditionally independent of all other random variable given $M_{1\to d_v-1}^{(l)}$. We thus have

$$\begin{aligned}\mathbb{P}\left(X^{(1\to l)}\right)&=\sum_{m_{1\to d_v-1}^{(1\to l)}}\mathbb{P}\left(X^{(1\to l)}\middle| M_{1\to d_v-1}^{(1\to l)}\right)\mathbb{P}\left(M_{1\to d_v-1}^{(1\to l)}\right)\\&=\sum_{m_{1\to d_v-1}^{(1\to l)}}\prod_{i=1}^{l}\mathbb{P}\left(X^{(i)}\middle| M_{1\to d_v-1}^{(i)}\right)\prod_{j=1}^{d_v-1}\mathbb{P}\left(M_j^{(1\to l)}\right)\end{aligned}\qquad (4)$$

V. Memory Truncation

A. Main Idea

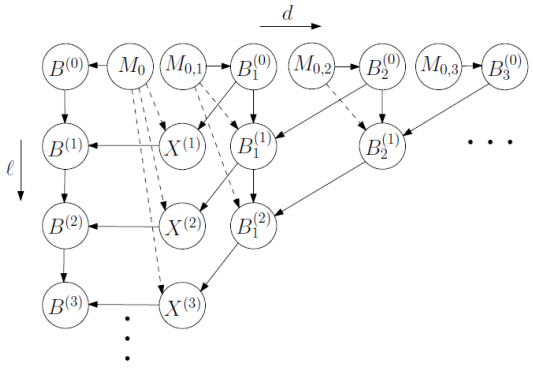


Fig. 3. Bayesian network based on the model of Fig. 1(b).

To explain the approximation, we consider the calculation of the joint distribution $\mathbb{P}\left(B^{(l-1)}\middle|X^{(l)}\right)$. This computation uses the fact that $B^{(l-1)} \perp X^{(l)} \mid X^{(1\to l-1)}$. The problem lies in the fact that the size of the sample space of the conditioning set $X^{(1\to l-1)}$ grows exponentially with *l*. Now consider making the following approximation:

$$B^{(l-1)} \perp X^{(l)} \mid X^{(l-n+1\to l-1)}, n \geq 2 . \quad (5)$$

Regardless of *l*, the conditioning set in (5) will always have a sample space with size $\left|S_X\right|^{n-1}$. Consider the sequence $B^{(0\to l)}$. This sequence, in general, is not a Markov process of some finite order. The memory truncation approximates $B^{(0\to l)}$ by a Markov process of order *n*, $n \in \mathbb{N}$. This is represented by the following:

$$\mathbb{P}\left(B^{(n+k)}\middle|B^{(n+k-1)},\ldots,B^{(0)}\right) \approx \mathbb{P}\left(B^{(n+k)}\middle|B^{(n+k-1)},\ldots,B^{(k)}\right) (6)$$

and corresponds to removing the edges between $X^{(i)}$ and $B^{(i)}$ for $i = 1,...,k$, in the Bayesian network of Fig.3.

We refer to the approximation of (6) as memory truncation of order $n\left(MT^n\right)$.

B. Analysis

We consider a memory truncation of order n, and assume that we have already calculated (approximated) $\mathbb{P}\left(B^{(k)}\right)$, $k \geq n$. For $k = 1, ..., n$, we have the following distributions available:

- $\mathbb{P}\left(B^{(k-1)}, X^{(k)}\right)$
- $\mathbb{P}\left(B^{(k-1)} \middle| X^{(k-n+1 \to k-1)}\right)$
- $\mathbb{P}\left(X^{(k-n+1 \to k)}\right)$

We now derive $\mathbb{P}\left(B^{(k+1)}\right)$. To perform this, we will use the calculation of the joint distribution $\mathbb{P}\left(B^{(k)}, X^{(k+1)}\right)$.

$$\begin{aligned}\mathbb{P}\left(B^{(k)}, X^{(k+1)}\right) &= \sum_{x^{(k-n+2 \to k)}} \mathbb{P}\left(B^{(k)} \mid X^{(k-n+2 \to k)}\right) \mathbb{P}\left(X^{(k+1)} \mid X^{(k-n+2 \to k)}\right) \mathbb{P}\left(X^{(k-n+2 \to k)}\right) \\ &= \sum_{x^{(k-n+2 \to k)}} \mathbb{P}\left(B^{(k)} \mid X^{(k-n+2 \to k)}\right) \mathbb{P}\left(X^{(k-n+2 \to k+1)}\right)\end{aligned} \tag{7}$$

In (7), the distribution $\mathbb{P}\left(B^{(k)} \mid X^{(k-n+2 \to k)}\right)$ is calculated by

$$\mathbb{P}\left(B^{(k)} \mid X^{(k-n+2 \to k)}\right) = \sum_{b^{(k-1)}} \mathbb{P}\left(B^{(k)} \mid X^{(k)}, B^{(k-1)}\right) \mathbb{P}\left(B^{(k-1)} \mid X^{(k-n+2 \to k)}\right),$$

where

$$\begin{aligned}&\mathbb{P}\left(B^{(k-1)} \mid X^{(k-n+2 \to k)}\right) \\ &= \sum_{x^{(k-n+1)}} \mathbb{P}\left(B^{(k-1)} \mid X^{(k-n+1 \to k)}\right) \mathbb{P}\left(X^{(k-n+1)} \mid X^{(k-n+2 \to k)}\right) \\ &= \frac{1}{\mathbb{P}\left(X^{(k-n+2 \to k)}\right)} \sum_{x^{(k-n+1)}} \mathbb{P}\left(B^{(k-1)} \mid X^{(k-n+1 \to k-1)}\right) \mathbb{P}\left(X^{(k-n+1 \to k)}\right) \quad \because \mathbb{P}\left(X^{(k-n+1)} \mid X^{(k-n+2 \to k)}\right) = \frac{\mathbb{P}\left(X^{(k-n+1 \to k)}\right)}{\mathbb{P}\left(X^{(k-n+2 \to k)}\right)}\end{aligned}$$

VI. Simulation Results

In general, the accuracy of $\mathbb{P}\left(B^{(l)}\right)$ increases with increasing the memory truncation order *n*, and so does the complexity. It is however expected that after increasing *n* beyond a certain order n_0, the accuracy improvement would be negligible. The goal is thus to find n_0.

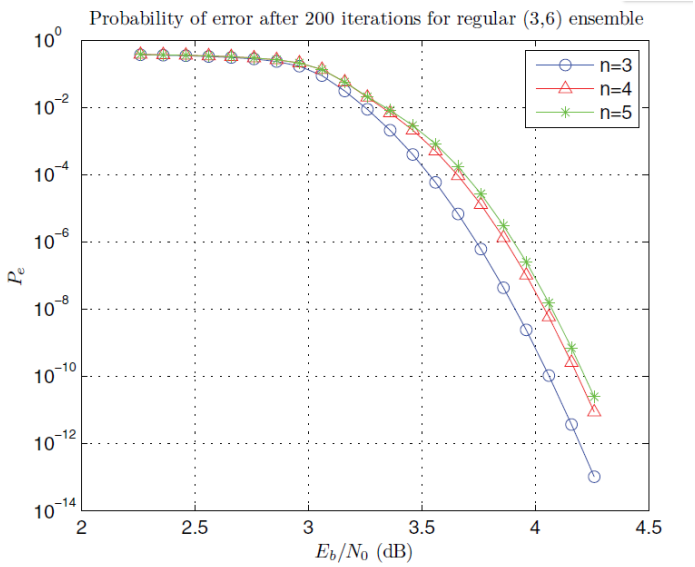


Fig. 4. P_e after 200 iterations vs. E_b/N_0 for the (3,6) ensemble with memory truncation orders 3, 4, and 5.

In Fig. 4, we have shown $P_e^{(l)}$ of the (3, 6) LDPC code ensemble for $l = 200$ vs. E_b / N_0 for different values of memory truncation order *n*. The curves demonstrate a convergence behavior as *n* is increased. In particular, the two curves for *n* = 4 and *n* = 5 are very close. We have also tried a number of other ensembles and observed a similar trend

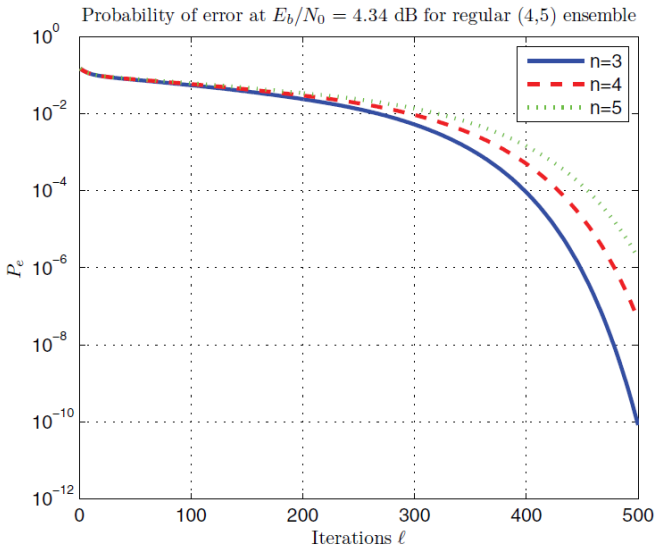


Fig. 5. $P_e^{(\ell)}$ of the (4,5) ensemble vs. ℓ at $E_b/N_0 = 4.34$ dB (above threshold) for memory truncation orders 3, 4, and 5.

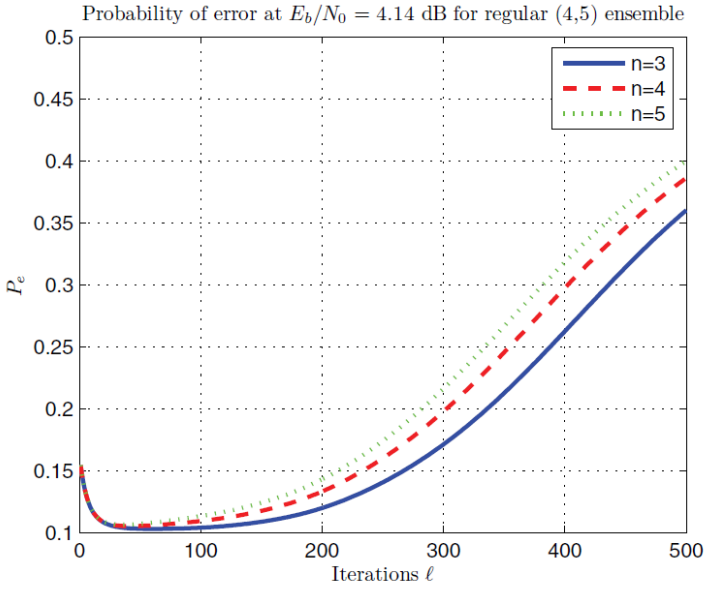


Fig. 6. $P_e^{(\ell)}$ of the (4,5) ensemble vs. ℓ at $E_b/N_0 = 4.14$ dB (below threshold) for memory truncation orders 3, 4, and 5.

For the ensemble of (4, 5) codes, we have plotted $P_e^{(l)}$ vs. l for truncation orders 3, 4 and 5, and for E_b / N_0 values 4.34 dB and 4.14 dB in Figures 5 and 6, respectively. The figures suggest that the ensemble threshold is between the two SNR values.

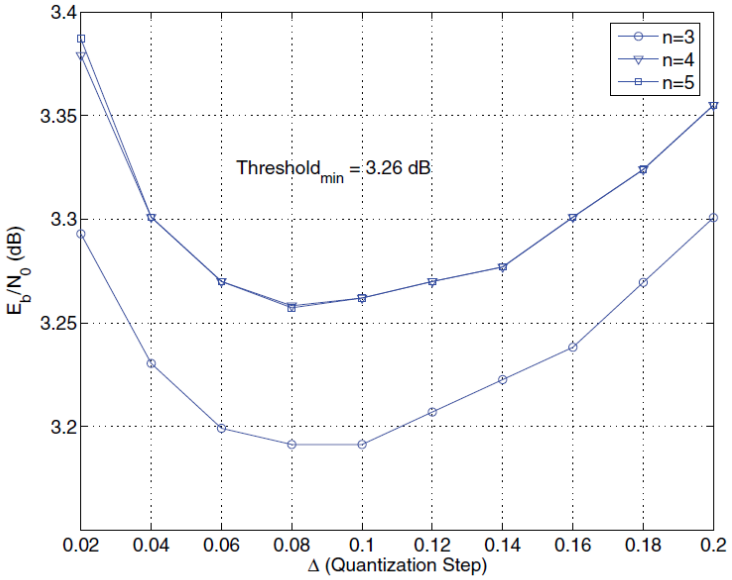


Fig. 7. Threshold values of the $(3, 6)$ ensemble vs. Δ for different memory truncation orders $(q = 8)$.

To clearly see the effect of memory truncation on the calculated thresholds, in Fig. 7, we show the threshold values of the (3, 6) ensemble for different memory truncation orders *n*. The thresholds for each truncation order are plotted versus the quantization step Δ for q = 8.the calculated thresholds for n = 4 and n = 5 are practically identical for different values of Δ. From Fig. 7, the optimal threshold of the (3, 6) ensemble (as a function of Δ) is seen to be about 3.26 dB. Based on the above results, in the following, we use n_0 = 4 to derive the thresholds. In all cases, we use q = 8 and the optimal value of Δ that minimizes the threshold.

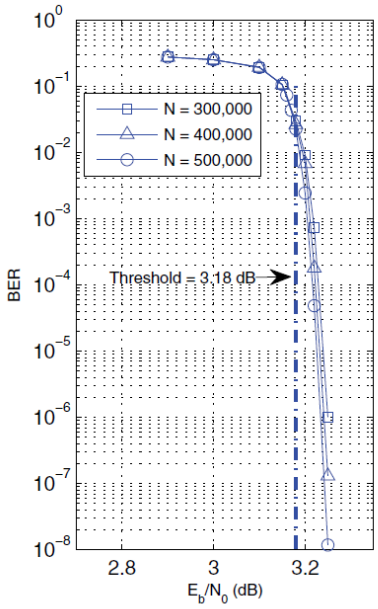


Fig. 8. BER curves of $(5, 10)$ LDPC codes with block lengths 300,000; 400,000; and 500,000 decoded by DD-BMP, and the corresponding threshold $(q = 8)$.

To verify the calculated threshold, for the (5, 10) ensemble, we compare the performance of randomly constructed (5, 10) codes of large block length (N = 300, 000; 400, 000 and 500, 000) with the threshold value of the ensemble (3.18 dB) in Fig. 8.

TABLE I
THRESHOLDS OF BP AND DD-BMP FOR REGULAR ENSEMBLES WITH $d_v = 4$ AND $5 \leq d_c \leq 18$

d_c	$rate$	$\left(\frac{E_b}{N_0}\right)_{BP}$ (dB)	$\left(\frac{E_b}{N_0}\right)_{DD-BMP}$ (dB)	Gap (dB)
5	1/5	2.57	4.24	1.67
6	1/3	1.73	3.18	1.45
7	3/7	1.27	2.88	1.61
8	1/2	1.59	2.80	1.21
9	5/9	1.67	2.82	1.15
10	3/5	1.78	2.87	1.09
11	7/11	1.89	2.93	1.04
12	4/6	1.99	2.99	1.00
13	9/13	2.10	3.06	0.96
14	5/7	2.20	3.13	0.93
15	11/15	2.29	3.20	0.91
16	3/4	2.39	3.27	0.88
17	13/17	2.47	3.33	0.86
18	7/8	2.55	3.40	0.85

TABLE II
THRESHOLDS OF BP AND DD-BMP FOR REGULAR ENSEMBLES WITH $d_v = 6$ AND $7 \leq d_c \leq 18$

d_c	$rate$	$\left(\frac{E_b}{N_0}\right)_{BP}$ (dB)	$\left(\frac{E_b}{N_0}\right)_{DD-BMP}$ (dB)	Gap (dB)
7	1/7	5.14	6.46	1.32
8	1/4	3.52	4.71	1.19
9	1/3	2.90	4.00	1.10
10	2/5	2.61	3.63	1.02
11	5/11	2.47	3.45	0.98
12	1/2	2.41	3.34	0.93
13	7/13	2.39	3.26	0.87
14	6/7	2.40	3.24	0.84
15	3/5	2.43	3.24	0.81
16	5/8	2.46	3.24	0.78
17	11/17	2.50	3.26	0.76
18	2/3	2.54	3.28	0.74

These results show that for a fixed dv, the threshold gap between DD-BMP and BP decreases with increasing the rate. As it can be seen, at higher rates the performance gap is less than 1 dB. In comparison with MS, for codes with larger degrees, DD-BMP outperforms MS

TABLE III
THRESHOLDS OF MS AND DD-BMP FOR RATE-1/2 ENSEMBLES

d_v	d_c	$\left(\frac{E_b}{N_0}\right)_{MS}$ (dB)	$\left(\frac{E_b}{N_0}\right)_{DD-BMP}$ (dB)	Gap (dB)
3	6	1.70	3.26	1.56
4	8	2.49	2.80	0.31
5	10	3.09	3.18	0.09
6	12	3.54	3.34	-0.21
7	14	3.91	3.61	-0.30

These results show that by increasing the degrees, the performance gap between MS and DD-BMP, which is to the advantage of MS for smaller degrees, disappears and then reverses to the advantage of DD-BMP.

These performance results for DD-BMP are impressive considering that both the check node operations and the message-passing for DDBMP are much simpler than those of BP and MS. They also demonstrate the potential of iterative decoding algorithms with memory in achieving better performance/complexity tradeoffs compared to memoryless algorithms.

Faster STORM using compressed sensing

Lei Zhu et al.

Nature method. (2012.04)

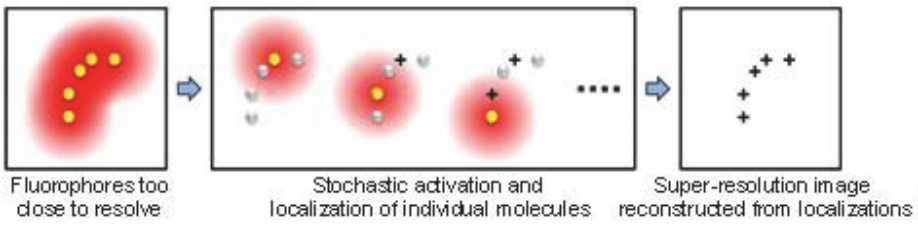
Presenter : Eunseok Jung

GIST, Dept. of Mechatronics , Bioscopy Lab.

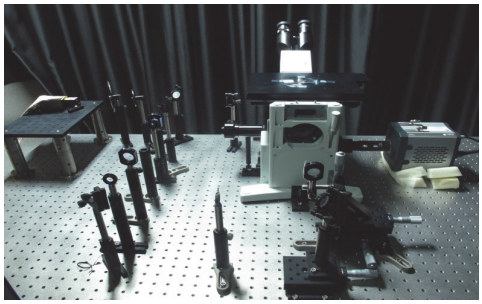
Gwangju Institute of Science and Technology

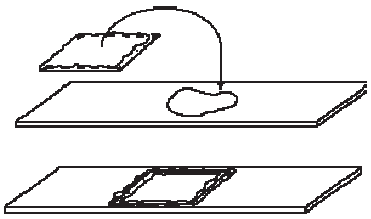
Background

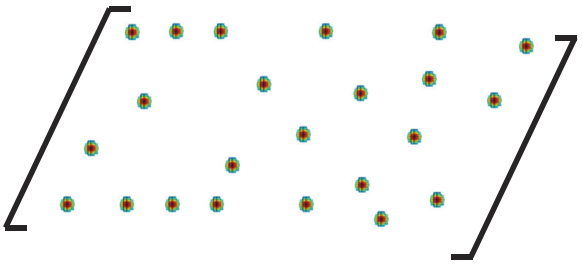
- STORM (Stochastic optical reconstruction microscopy)
  - STORM is a super-resolution optical microscopy technique based on stochastic switching of single-molecule fluorescence signal.
  - STORM utilizes fluorescent probes that can switch between fluorescent and dark states so that in every snapshot, only a small, optically resolvable fraction of the fluorophores is detected.
  - This enables determining their positions with high precision from the center positions of the fluorescent spots.



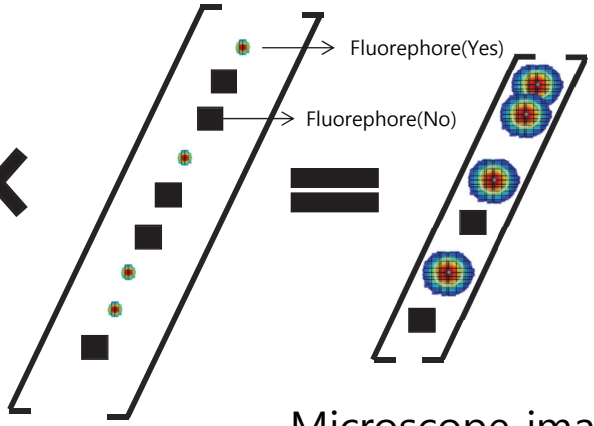
Cited form : http://huanglab.ucsf.edu







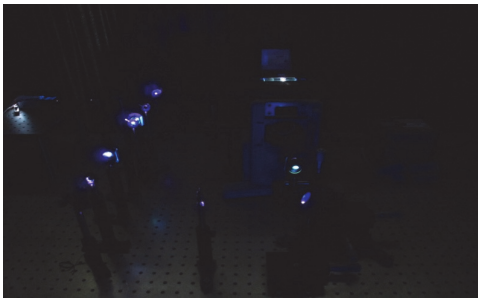
Matrix(**A**)
Optic system - LTI

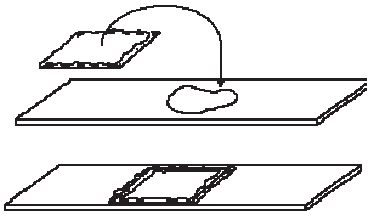


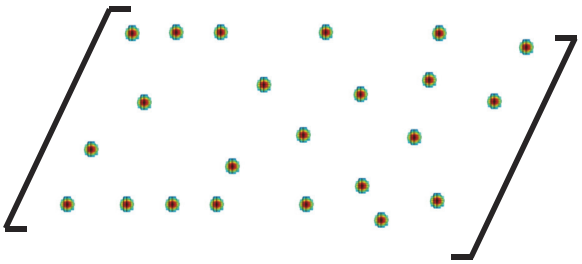
Original
Image(**X**)

Microscope image
-Blurred picture

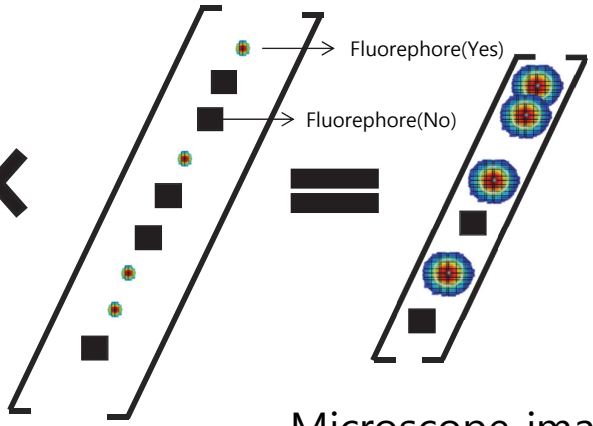
From : http://www.ruf.rice.edu/







Matrix(**A**)
Optic system - LTI



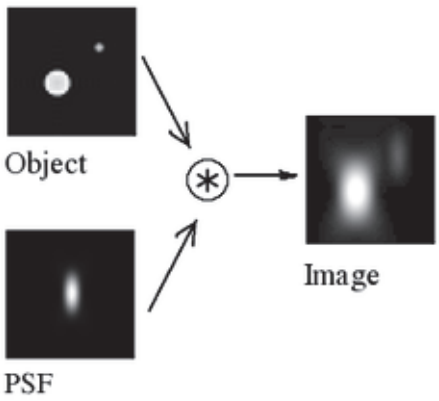
Original
Image(**X**)

Microscope image
-Blurred picture

From : http://www.ruf.rice.edu/

Background

- Point spread function(PSF)
  - The **point spread function** (**PSF**) describes the response of an imaging system to a point source or point object.
  - A more general term for the PSF is a system's impulse response, the PSF being the impulse response of a focused optical system.



From : http://en.wikipedia.org

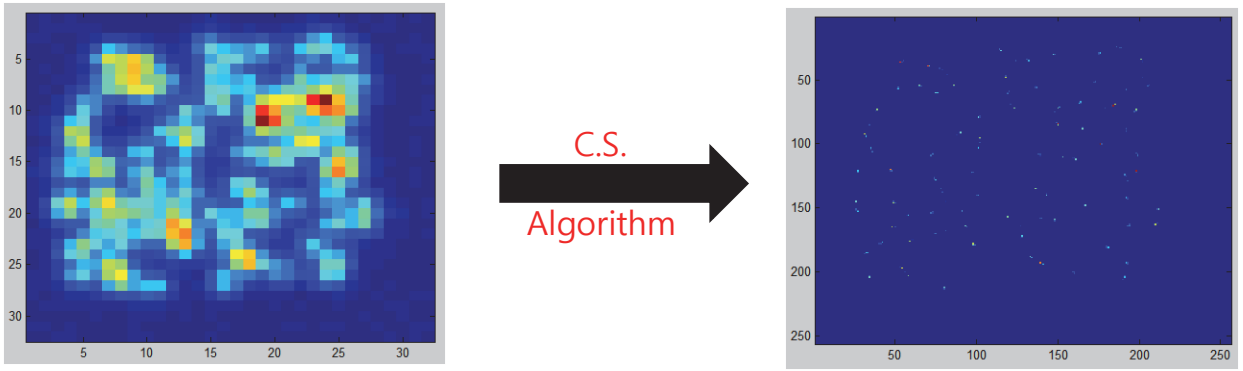
Introduction & Motivation

- Super resolution microscope overcome traditional optical microscope limit.(According to Abbe's theory : $\frac{Light\ wave\ length}{2} \approx 200\text{nm}$)
- The super resolution microscope and fluorescence technique make spatial resolution closer to the molecular scale.(Approximately : 30nm)
- **Now we can see a ten nanometer scale cell structure.**

From : Imaging Intracellular Fluorescent Proteins at Nanometer Resolution, E. Beitzig, science, 2006

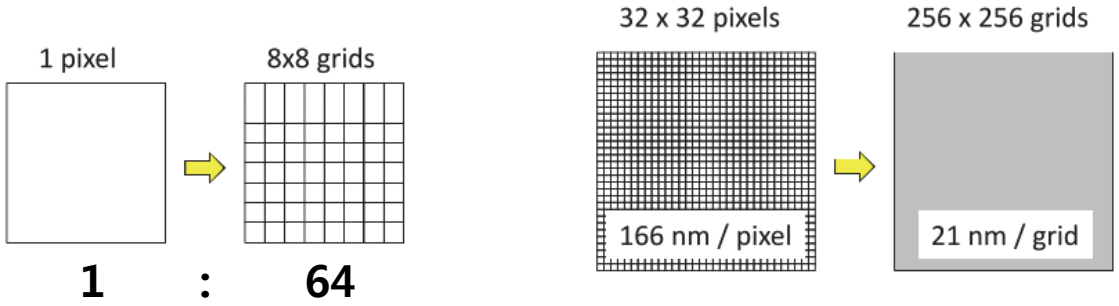
Introduction & Motivation

- Benefit of compressed sensing in fluorescence imaging.



Oversampling

- For increasing spatial resolution, the author used grid method.

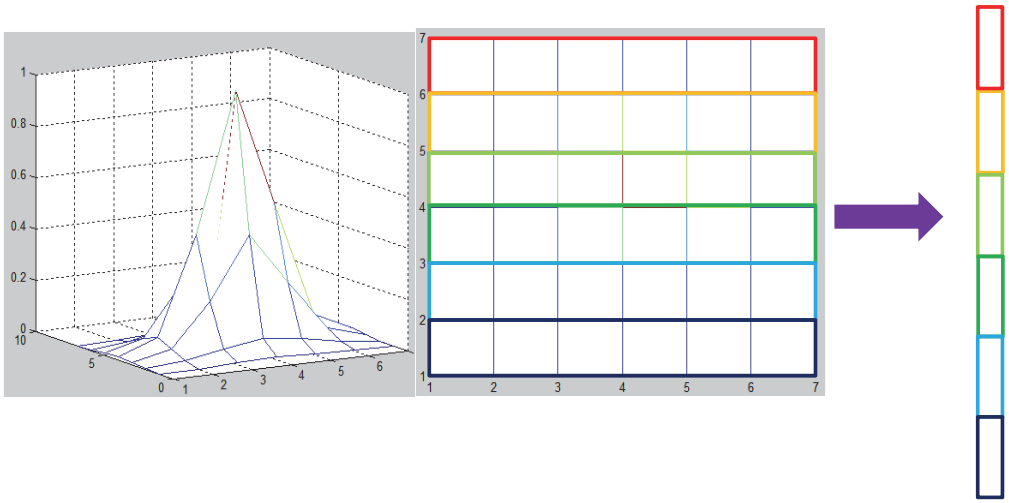


- According to the nyquist theorem, we need 2 times more sampling for reconstruct singal. In here, 21nm*2=42nm.

- Finally, the reconstruction image can get a 42nm spatial resolution.

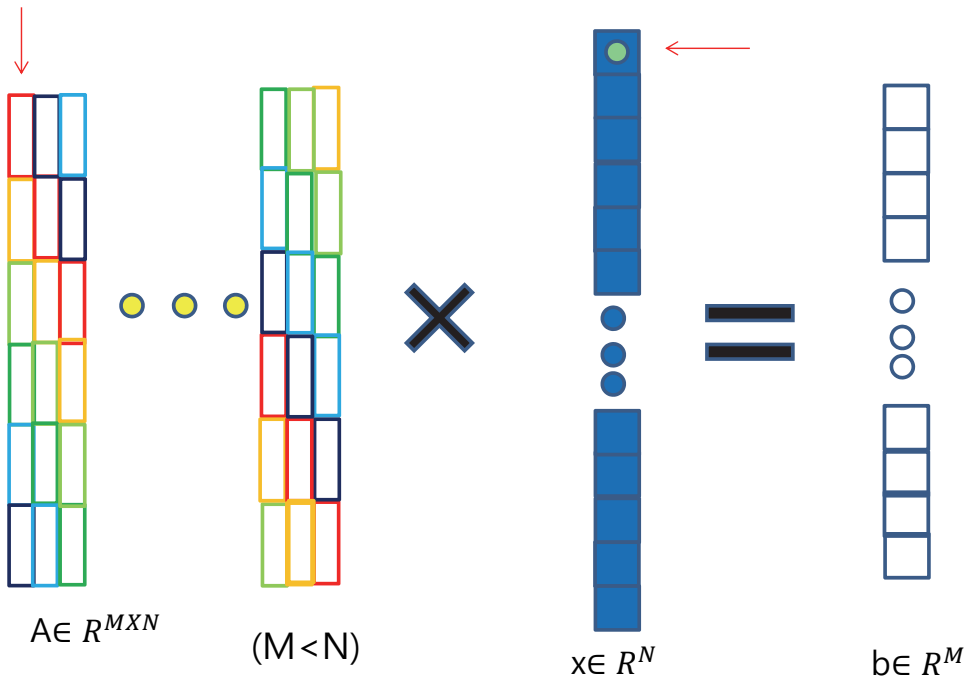
Measurement matrix (1)

- Measure the PSF of optic system.
- Fit the Gaussian function.



- Optic system is LTI system.

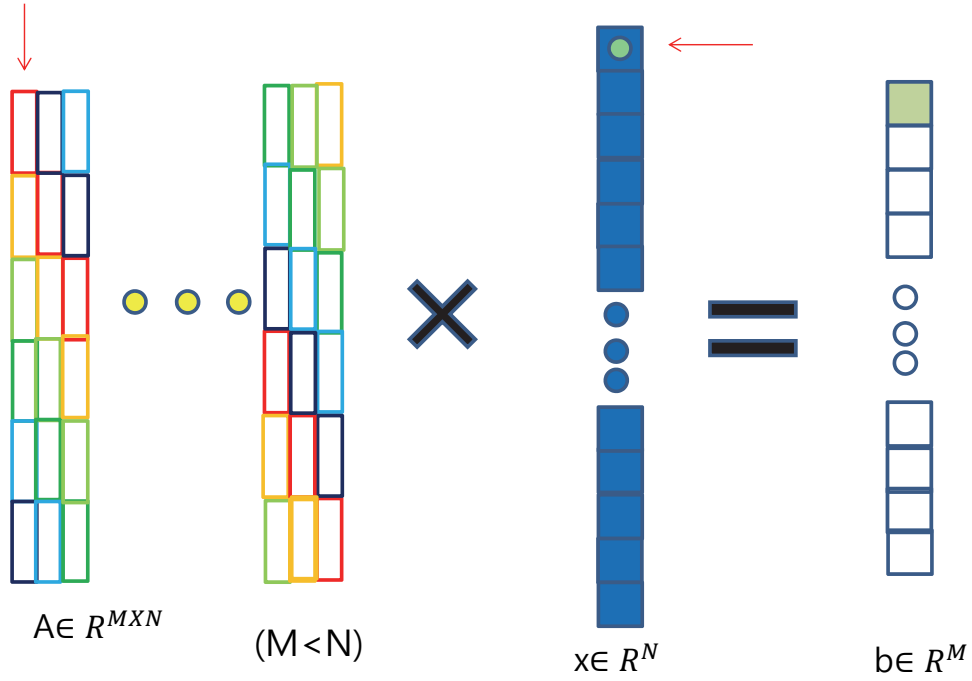
Measurement matrix (2)



A = Measurement matrix **x** = Reconstruct image

Measurement matrix (3)

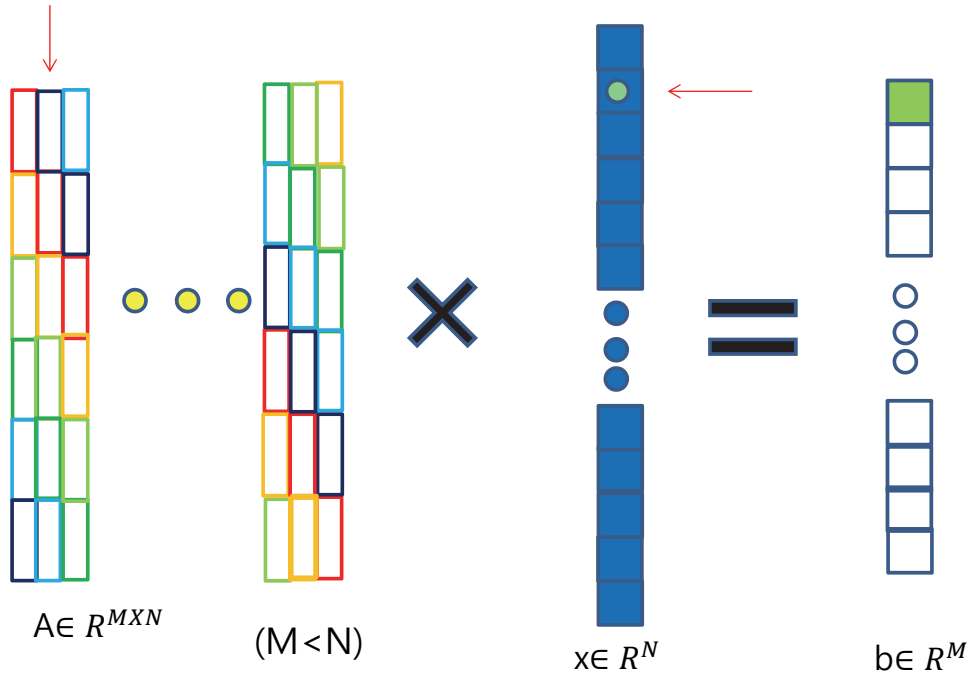
- Measurement matrix A



A = Measurement matrix **x** = Reconstruct image

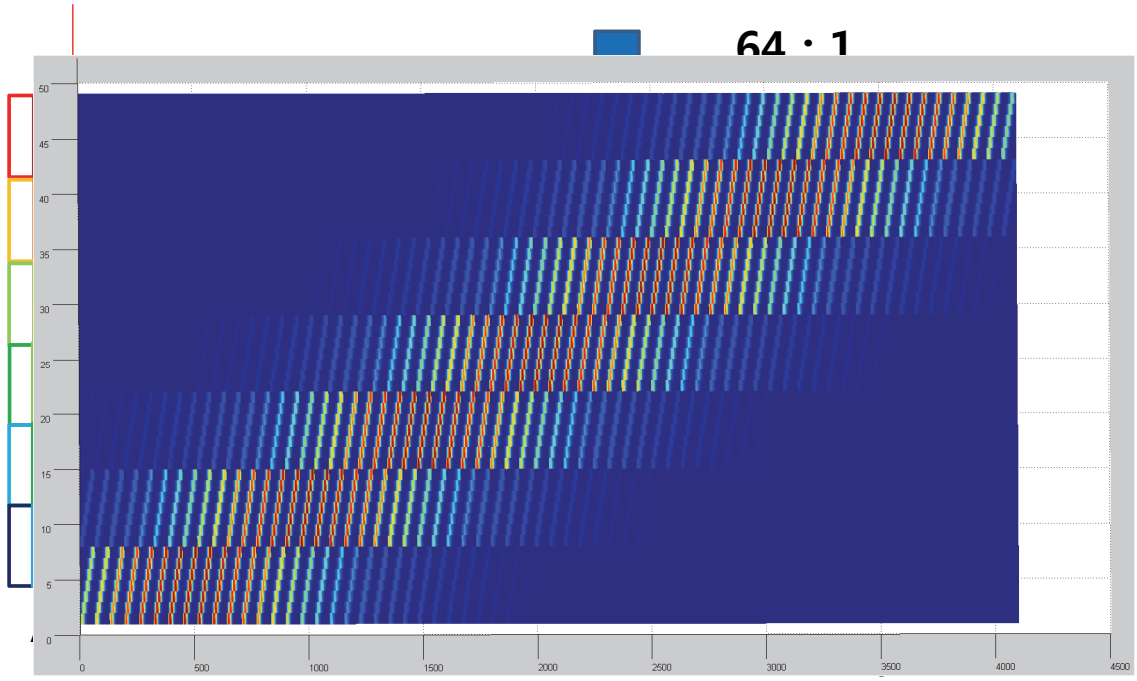
Measurement matrix (4)

- Measurement matrix A

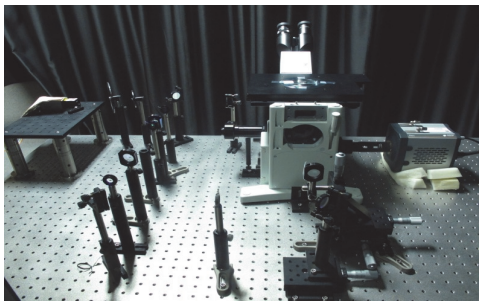


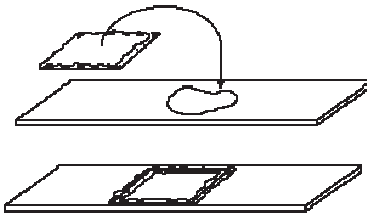
A = Measurement matrix **x** = Reconstruct image

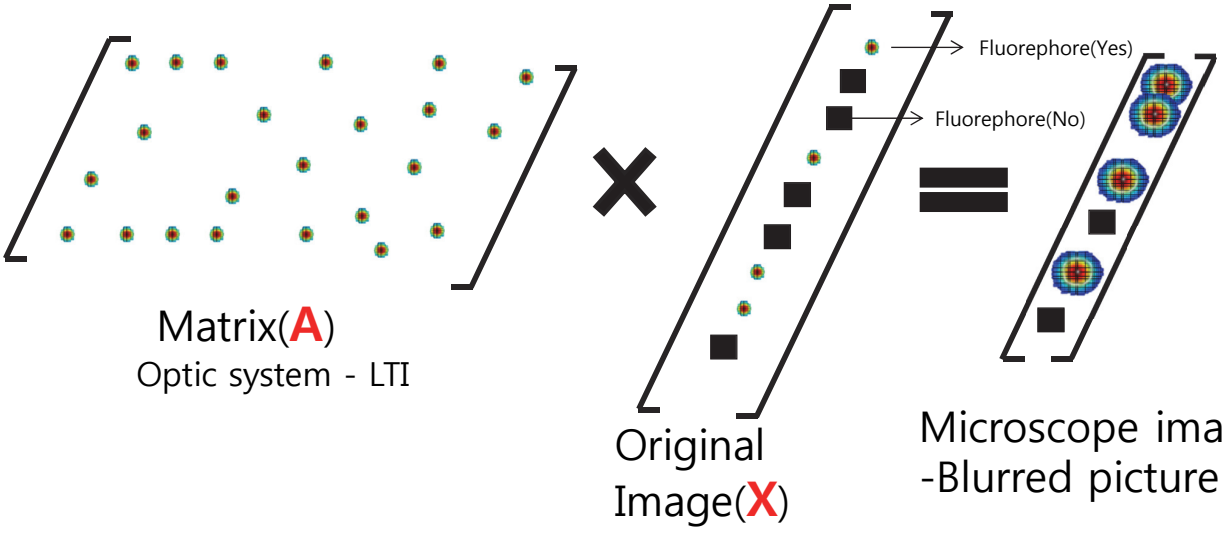
Measurement matrix (5)



A = Measurement matrix **x** = Reconstruct image







Matrix(**A**)
Optic system - LTI

Original
Image(**X**)

Microscope image
-Blurred picture

From : http://www.ruf.rice.edu/

Reconstruction

$$\text{Minimize: } c^{\mathrm{T}}\boldsymbol{x}$$

$$\text{Subject to: } \boldsymbol{x}_i \geq 0 \text{ and } \| \boldsymbol{Ax} - \boldsymbol{b} \|_2 \leq \varepsilon \cdot (\Sigma \boldsymbol{b}_j)^{1/2}$$

- The matrix **A** is determined by the point-spread function (PSF) of the imaging system.
- The i^{th} column of **A** corresponds to the acquired raw image if only one molecule emits fluoroscopic photons at the position index i of **x.**
- The weight vector **c** is to account for the difference of the total contribution to the camera image from one fluorescent molecule at different locations.
- The value of the i^{th} element of **c** equals the summation of the i^{th} column of **A.**
- The minimization term $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ is equivalent to a weighted L1 norm of **x** because **x** is non-negative.

Reconstruction

$$\text{Minimize: } c^{\mathrm{T}}\boldsymbol{x}$$

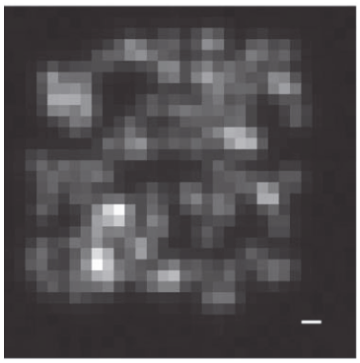
$$\text{Subject to: } \boldsymbol{x}_i \geq 0 \text{ and } \| \boldsymbol{Ax} - \boldsymbol{b} \|_2 \leq \varepsilon \cdot (\Sigma \boldsymbol{b}_j)^{1/2}$$

```
b = img_raw(:);

n = len;

cvx_begin
    variable x(n)
    minimize(c*x)
    subject to
        x >= 0;
        norm( A * x - b, 2 )<=eps;
cvx_end

img_est = x;
```

- CVX converts the above problem to an SOCP (Second-order cone programming), and solves it.

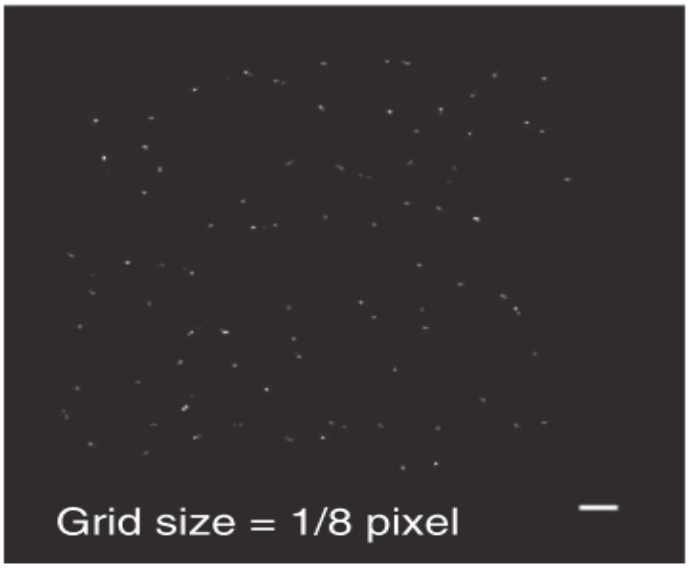
Results

100 molecules in simulation



C.S.
Algorithm

Compressed sensing

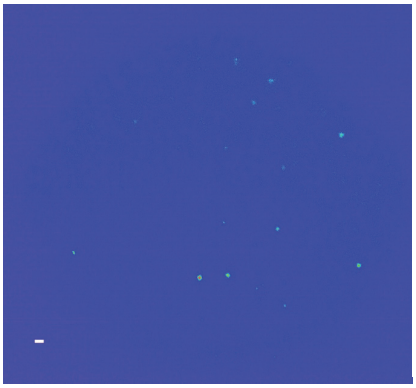


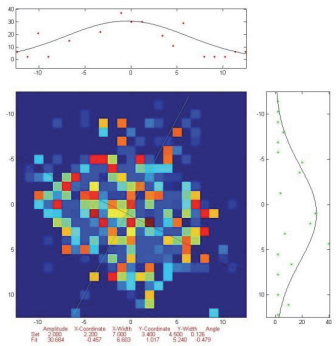
Scale bar : 300nm
FOV : 4um*4um
Pixel : 32*32

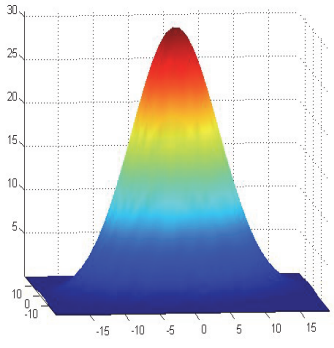
Scale bar : 300nm
FOV : 4um*4um
Pixel : 256*256
The pixel size of 166nm, the 21nm grid size should be able to support a final image resolution of 42nm.

Results

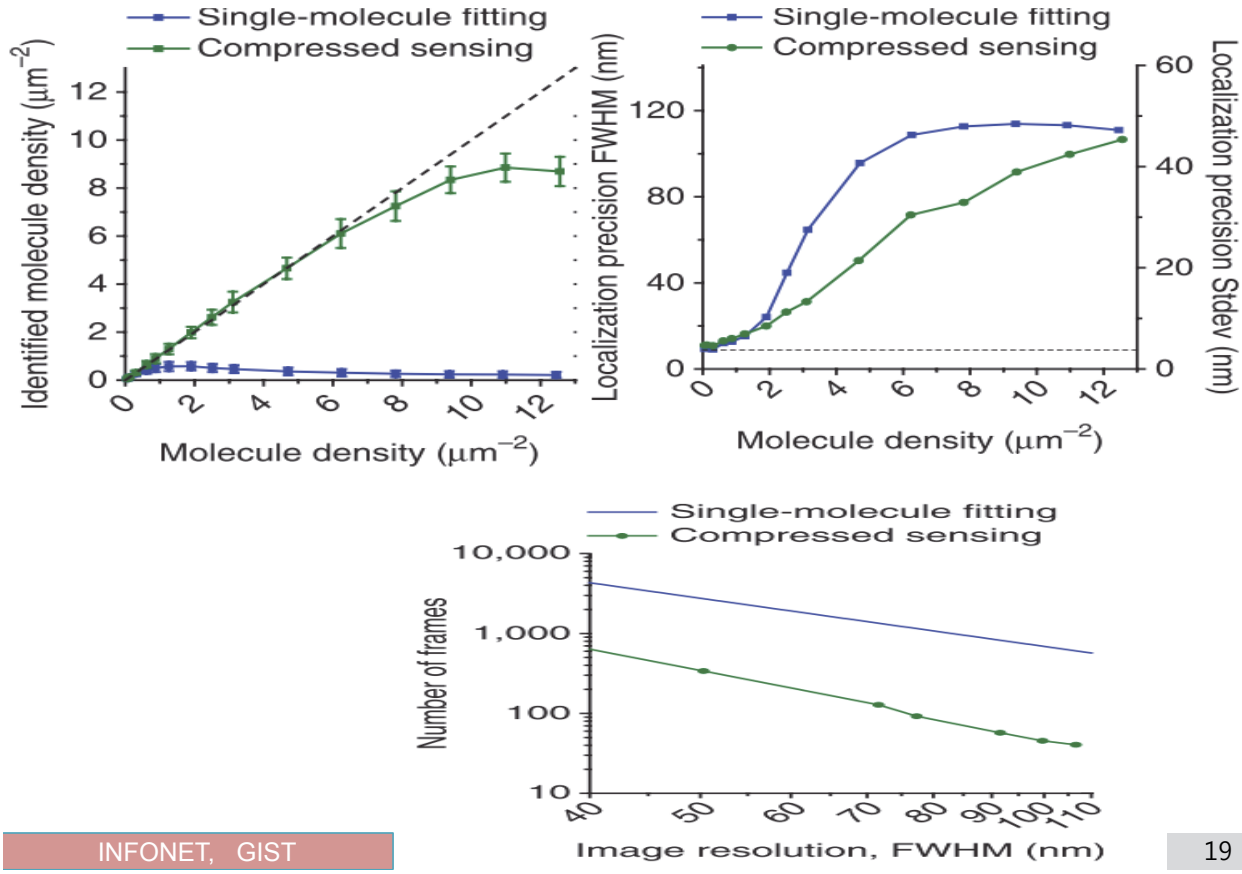
- Conventional fitting



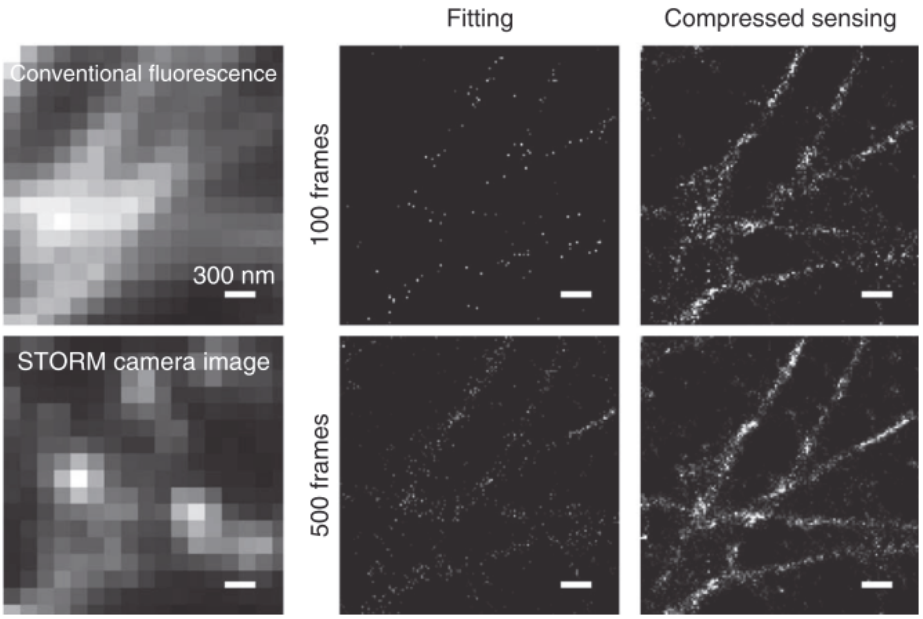




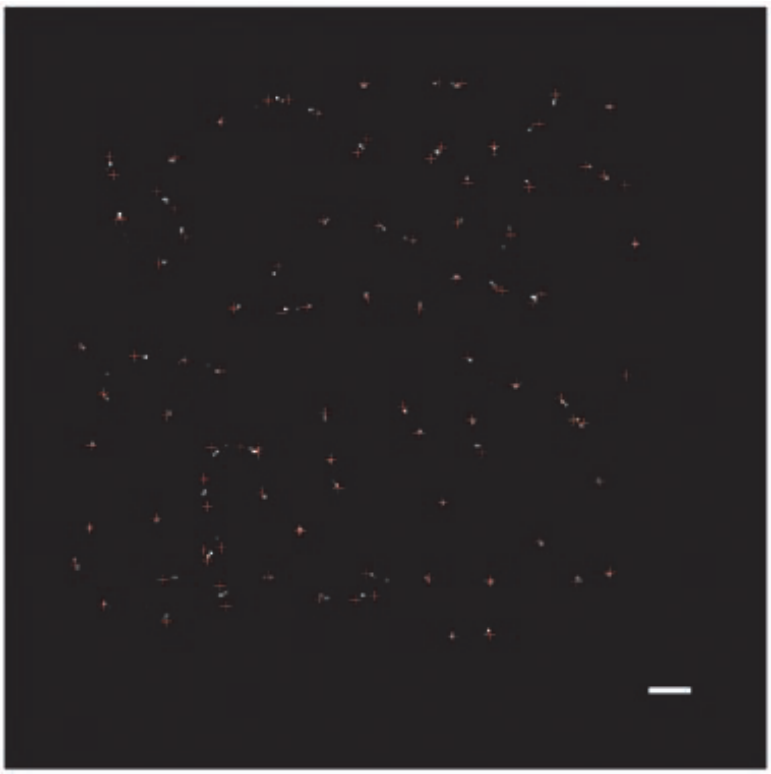
Results



Results



Data Analysis and Discussions



- The red-cross is the reconstructed image using compressed sensing
- The white one is fluorescence position (original image).
- The maximum difference between two positions is 60nm.
- Scale bar is 300nm.

Conclusions

- Conclusion

1) This journal has potential benefit. It was first step of compressed sensing with super resolution microscope.
2) It had not been impossible to taking a living cell image without compressed sensing.
3) Now, it could be done with compressed sensing.
4) The author spent 3sec for taking living cell photo. And they get a same result when they spent over 30sec. They decrease experiment time 10times more.
5) According to this, now we can get a living cell image.

Improvement

- Algorithm

1) If we make measurement matrix A using difference method, we can decrease error and also increase spatial resolution.

2) The other reconstruction method, instead of SOCP, can decrease error.

- Hard ware

1) The author used EM-CCD camera, it is very expensive detect device.

2) But if we can make same result using cheap detect device like a sCMOS or CMOS, the system cost is more cheaper than before.

Journal Club Meeting, Feb. 14, 2013

Thank you

A Node-Based Time Slot Assignment Algorithm for STDMA Wireless Mesh Networks

Authors: W. Chen, and Chin-Tau Lea

Publication: IEEE Trans. Veh. Tech., Jan. 2013
Speaker: Asif Raza

Short summary: In this paper authors present a link capacity model for spatial time-division multiple access (STDMA) mesh networks. It makes use of a simplified transmission model that also considers channel fading. The model then forms the basis of a node-based slot-assignment and scheduling algorithm. This algorithm enables the user to exploit multiuser diversity that results in optimizes network throughput. The presented algorithm shows significant improvement in the throughput when compared with existing slot-assignment methods.

I. Introduction

In STDMA network the transmission time of a channel is divided into slots where multiple slots constitute a frame. These slots are assigned to potential users of the network. The goal of slot assignment scheme is to maximize network throughput. Existing assignment algorithms in STDMA make use of simplified transmission model which do not consider the time-varying fading behavior of a wireless channel. This results in slot wastage when link is in deep fade. The slot is also wasted if scheduled link has no traffic to transmit. This degrades the STDMA network throughput. Therefore a dynamic slot-assignment with that should exploit multiuser diversity is required. However sheer complexity involved in coordinating with all nodes and generating scheduling map in a reasonable time makes this approach impractical. In order to fix these issues the authors present a node-based slot-assignment scheme in which scheduling in each slot is done for nodes not for links. Their contributions include:

- Defining link capacity: a model that includes channel fading. It ensures that whichever link is used by a node will not change the interference profiles on the links selected by other users.
- Node-based time-slot assignment and scheduling algorithms.

II. System Model

Wireless STDMA mesh network with fixed routers.

Transmissions are organized in frames.

Synchronization among nodes provided through GPS.

Set of nodes are identified and assigned to a slot for their transmission.

Each node maintains a separate queue for each outgoing link and performs scheduling without coordination with other nodes.

Multiprotocol Label Switching (MPLS) multipath routing is used for routing however packets are transmitted in sequence.

Adaptive modulation and time varying fading channels are considered. It is also assumed that wireless channels undergo slow fading. Due to fading channel an instant channel gain will be fed back to transmitter. The duration for feedback is no longer than coherence time (the time for which channel conditions remain same)

Adaptive modulation is implemented that each data packet can be fragmented into multiple segments and each segment can be transmitted in with lowest data rate. If high data rate is available then multiple segments can be transmitted per slot duration.

III. Link Capacity Modeling

Each node has multiple links and it can exploit multiuser diversity i.e. different links have different traffic and fading conditions. A channel model is presented that includes shadowing and slow fading.

A. Signal to interference and noise ratio (SINR) Formulation

$h_{r,t}$: Channel response function from transmitter 't' and receiver 'r'

x_t: Signal from 't'

$I_r, n_I, t_i^{'}$: Set of transmitters causing interference to 'r', number of transmitters and i[th] transmitter in I_r respectively. Power control is not considered therefore transmission power of 't' is $pt = E\left(\left|x_t\right|^2\right)$. Let n_0 be thermal noise with power equal to k then received power at 'r' is

$$y_{r,t} = h_{r,t}x_t + \sum_{i=1}^{n_I} h_{r,t_i^{'}} x_{t_i^{'}} + n_0 \qquad (1)$$

SINR at receiver 'r' is expressed as:

$$\gamma_{r,t} = \frac{\left|h_{r,t}x_t\right|^2}{\sum_{i=1}^{n_I}\left|h_{r,t_i^{'}}x_{t_i^{'}}\right|^2 + \kappa} = \frac{\left|h_{r,t}\right|^2 p_t}{\sum_{i=1}^{n_I}\left|h_{r,t_i^{'}}\right|^2 p_{t_i^{'}} + \kappa} = \frac{s_0}{\sum_{i=1}^{n_I} s_i + \kappa} \qquad (2)$$

Here $s_0 = \left|h_{r,t}\right|^2 p_t$ *and* $s_i = \left|h_{r,t_i^{'}}\right|^2 p_{t_i^{'}}$. The Channel response function consists of three parts:

- Path loss
- Shadowing
- Fading

$$h_{r,t} = \sqrt{l_{r,t}^{-\alpha} 10^{\frac{f_{r,t}}{10}} \pi_{r,t}} \qquad (3)$$

Where $l_{r,t}$ is distance between 't' and 'r', $\alpha \mapsto [2-4]$ (constant), $10^{\frac{f_{r,t}}{10}}$ is shadowing effect and it is modeled as a log-normal distributed random variable. $\pi_{r,t}$ is fading effect and it is defined as complex Gaussian RV with mean and variance equal to 0 and 1 respectively. PDF of s_0 and s_i are defined as:

$$p_{s_0}(\alpha_0) = \tfrac{1}{\rho_0} e^{-\frac{\alpha_0}{\rho_0}} \qquad (4a)$$

$$p_{s_i}(\alpha_i) = \tfrac{1}{\rho_i} e^{-\frac{\alpha_i}{\rho_i}} \qquad (4b)$$

Here $\rho_0 = E(s_0) = l_{r,t}^{-\alpha} 10^{\frac{f_{r,t}}{10}} p_t$ and $\rho_i = E(s_i) = l_{r,t_i'}^{-\alpha} 10^{\frac{f_{r,t_i'}}{10}} p_{t_i'}$

B. PDF of SINR

Case 1: no interference is observed by receiver 'r' i.e. (I_r=0, n_I=0) then PDF of $\gamma_{r,t}$, is defined as: let $\delta = \gamma_0 / \kappa$

$$p_{\gamma_{r,t}}(z) = p_{s_0}(z\kappa) = \tfrac{1}{\delta} e^{-\frac{z}{\delta}} \qquad (5)$$

Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr(\gamma_{r,t} \geq w) = \int_w^\infty p_{\gamma_{r,t}}(z)dz = e^{-\frac{w}{\delta}} \qquad (6)$$

Case 2: unit interference is observed by 'r' i.e. (I_r>0, n_I=1) then PDF of term (s_i+k i.e. denominator of equ.2) is defined as:

$$p_{s_1+\kappa}(v) = \tfrac{1}{\sigma_1} e^{-\frac{v-\kappa}{\sigma_1}} \qquad (7)$$

Finally PDF of $\gamma_{r,t}$ is defined as:

$$p_{\gamma_{r,t}}(z) = \int_\kappa^\infty v p_{s_0}(vz) p_{s_1+\kappa}(v)dv = \int_\kappa^\infty \tfrac{v}{\sigma_0} e^{-\frac{vz}{\sigma_0}} \tfrac{1}{\sigma_1} e^{-\frac{v-\kappa}{\sigma_1}} dv = \tfrac{\kappa + \frac{1}{u}}{\sigma_0 \sigma_1 u} e^{\left(\frac{\kappa}{\sigma_1} - u\kappa\right)} \qquad (8)$$

Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr(\gamma_{r,t} \geq w) = \int_w^\infty p_{\gamma_{r,t}}(z)dz = \frac{\sigma_0}{\sigma_0 + w\sigma_1} e^{-\frac{w}{\delta}} \qquad (9)$$

Case 3: more than one interferers are present in I_r i.e. ($n_I > 1$) then PDF of $\sum_{i=1}^{n_I} s_i$ can be defined as:

$$p_I(v) = P_{\sum_{i=1}^{n_I} s_i}(v) \otimes p_{s_{n_I}}(v) = \sum_{i=1}^{n_I} \frac{b_i}{\sigma_i} e^{-\frac{v}{\sigma_i}} \qquad (10)$$

Here $b_i = \prod_{j=1, j\neq i}^{n_I} \left(\sigma_i / \sigma_i - \sigma_j \right) and \sum_{i=1}^{n_I} b_i = 1$. The PDF of term ($s_i$+k i.e. denominator of equ.2) is defined as: $p_{I+\kappa}(v) = p_I(v-\kappa)$. Finally the PDF of $\gamma_{r,t}$:

$$p_{\gamma_{r,t}}(z) = \int_{\kappa}^{\infty} v p_{s_0}(vz) p_{I+\kappa}(v) dv = \sum_{i=1}^{n_I} d_i \left(\frac{\kappa}{q_i} + \frac{\kappa}{q_i^2} \right) e^{-\kappa q_i} \qquad (11)$$

Here $q_i = \left(z / \sigma_0 \right) + \left(1 / \sigma_i \right)$ and $d_i = \left(b_i / \sigma_0 \sigma_i \right) e^{-\kappa q_i}$. Probability that $\gamma_{r,t}$ is smaller than w is defined as:

$$Pr\left(\gamma_{r,t} \geq w \right) = \int_{w}^{\infty} p_{\gamma_{r,t}}(z) dz = \sigma_0 e^{-\frac{w}{\delta}} \sum_{i=1}^{n_I} \frac{b_i}{\sigma_0 + \sigma_i w} \qquad (12)$$

Finally Link Capacity can then be determined as:

$$c_{r,t}(I_r) = \sum_{i=1}^{\xi-1} c_i \cdot Pr\left(\gamma_{thr}^{i} \leq \gamma_{r,t} < \gamma_{thr}^{i+1} \right) + c_{\xi} \cdot Pr\left(\gamma_{r,t} \geq \gamma_{thr}^{\xi} \right)$$

Where $c_{r,t}(I_r)$ is average data rate between 't' and 'r', given interference set I_r $c_{r,t}(I_r))$ $c_{r,t}(I_r)$

IV. Proposed Time-Slot Assignment Algorithm

TDMA frame consists of a fixed number of slots is considered. The set of transmitting links that are activated in a given slot is called a **link pattern**, and the set of nodes activated in a given slot is called a **node pattern**.

A. Formulation of Node-Based Time-Slot Algorithm

Notations:

V: set of nodes

E: set of links

NP: Node Pattern

$tx_e, rx_e; e \in E$ transmitter and the receiver of link e, respectively,

$E_{s,p} = \{ e \mid e \in E, p \in s, s \in NP, tx_e = p \}$ set of links that can be used at node p, where p∈s i.e p is activated in node pattern *s*);

μ_s = portion of time that is assigned to node pattern (s) in a frame, where,

$\sum_{s \in NP} \mu_s = 1$

$\left\{ò_{s,p,e} \mid e \in E_{s,p}\right\}$:portion of time that is assigned to each link of node p in node pattern s

F: set of flows in the system; where flow defines all traffic that belongs to (S, D) pair

h_f : traffic demand for flow f, where $f \in F$

S_f : source of flow f

D_f : destination of flow f

$x_{f,e}$: percentage of traffic that flow f passes through link e,

Calculations

Link congestion: it is total amount of traffic routed through the link 'e' over its average capacity (c_e) i.e. $r_e = \left(\sum_{f \in F} \frac{x_{f,e}}{c_e} \right)$ where link capacity (data rate between transmitter 't' and receiver 'r' is defined as: $c_e = \sum_{\{s \mid s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} ò_{s,p,e}$.

Thus network congestion ratio 'r' is the maximum of all link congestion ratios, i.e. $r = \max_{e \in E} r_e$

Optimal node-based slot assignment scheme is one which minimizes congestion 'r':

$$\min \quad r \qquad (13a)$$

$$s.t \quad \frac{\sum_{f \in F} x_{f,e} h_f}{\sum_{\{e \mid s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} ò_{s,p,e}} \le r \quad (13b)$$

$$\sum_{\{e \mid e \in E_{s,p}\}} ò_{s,p,e} \le \mu_s \qquad (13c)$$

$$\sum_{\{e \mid p \in s, e \in E_{s,p}, rx_e = q\}} ò_{s,p,e} \le \mu_s \qquad (13d)$$

$$\sum_{s \in NP} \mu_s = 1 \qquad (13e)$$

$$\mu_s \ge 0, ò_{s,p,e} \ge 0, r \ge 0 \qquad (13f)$$

$$\sum_{\{e \mid tx_e = v\}} x_{f,e} - \sum_{\{e \mid rx_e = v\}} x_{f,e} = 0 \quad (13g)$$

$$\sum_{\{e \mid tx_e = S_f\}} x_{f,e} - \sum_{\{e \mid rx_e = S_f\}} x_{f,e} = 1 (13h)$$

$$x_{f,e} \ge 0 \qquad (13i)$$

Problem 13 is the optimization problem, whose purpose is to find the set of $\grave{o}_{s,p,e}$ that will lead to the optimal objective function. Constraint 13c represents that in node pattern s, for any node p∈s, p can transmit to only one node at one time. 13d ensures that a node q can receive from only one node at one time while q ∉ s and p∈s. 13f and 13i ensures non-negativity constraints. Constraint 13b is non-linear therefore $\mu_s r$ and $\grave{o}_{s,p,e} r$ are replaced by $\theta_s \; and \; \rho_{s,p,e}$ respectively. Therefore final formulation is defined as:

$$\min \quad \sum_{s \in NP} \theta_s \qquad (14a)$$

$$s.t \quad \sum_{f \in F} x_{f,e} h_f \le \sum_{\{e | s \in NP, p \in s, e \in E_{s,p}\}} c_{s,e} \grave{o}_{s,p,e} \qquad (14b)$$

$$\sum_{\{e | e \in E_{s,p}\}} \rho_{s,p,e} \le \theta_s \qquad (14c)$$

$$\sum_{\{e | p \in s, e \in E_{s,p}, rx_e = q\}} \rho_{s,p,e} \le \theta_s \qquad (14d)$$

$$\theta_s \ge 0, \rho_{s,p,e} \ge 0 \qquad (14e)$$

$$\sum_{\{e | tx_e = v\}} x_{f,e} - \sum_{\{e | rx_e = v\}} x_{f,e} = 0 \qquad (14f)$$

$$\sum_{\{e | tx_e = S_f\}} x_{f,e} - \sum_{\{e | rx_e = S_f\}} x_{f,e} = 1 \qquad (14g)$$

$$x_{f,e} \ge 0 \qquad (14h)$$

Authors describe that the presented formulation can handle scheduling of node patterns by using Linear Programming approach. However for link based approach, listing all link patterns does not work by using LP formulation. Therefore column generation method is used to tackle the problem.

B. Frame Construction and Throughput Loss due to Frame Quantization

Frame is constructed as: $n_f = \sum_{s \in NP} \left[z \mu_s \right]$ here z is frame length and function [x] rounds 'x' to nearest integer.

The frame quantization will change the portion of time assigned to all patterns (μ_s). Therefore parameters like minimum congestion ratio r_z, the optimal link capacities (c_e) and the routing scheme $x_{f,e}$

will change. These parameters need to be recomputed as follows. let z_s be number of slots assigned to node pattern 's' in a frame.

$$\min \quad \varsigma_z \qquad (15a)$$

$$s.t \quad \sum_{f\in F} y_{f,e} h_f \le \sum_{\{s|s\in NP, p\in s, e\in E_{s,p}\}} c_{s,e} \grave{o}_{s,p,e} \qquad (15b)$$

$$\sum_{\{e|e\in E_{s,p}\}} \grave{o}_{s,p,e} \le \frac{z_s}{\sum_{s\in NP} z_s} \qquad (15c)$$

$$\sum_{\{e|p\in s, e\in E_{s,p}, rx_e=q\}} \grave{o}_{s,p,e} \le \frac{z_s}{\sum_{s\in NP} z_s} \qquad (15d)$$

$$\sum_{\{e|tx_e=v\}} y_{f,e} - \sum_{\{e|rx_e=v\}} y_{f,e} = 0 \qquad (15e)$$

$$\sum_{\{e|tx_e=S_f\}} y_{f,e} - \sum_{\{e|rx_e=S_f\}} y_{f,e} = h_f \varsigma_z \qquad (15f)$$

$$y_{f,e} \ge 0,\ \grave{o}_{s,p,e} \ge 0 \qquad (15g)$$

Here $\varsigma_z = 1/r_z, \mu_s = \left(z_s / \sum_{s\in NP} z_s \right), y_{f,e} = \left(x_{f,e} / r_z \right) = x_{f,e}\varsigma_z$

C. Column Generation Method

Column generation is an algorithm for solving large LP problems. Most of the variables are usually non-basic and assume zero values in the optimal solution, only a subset of variables are needed for solving the problem. Column generation method considers only the variables which have potential to improve the objective function. It splits the problem into master problem and subproblem. Master problem is the original problem with subset of variables being considered. In subproblem it uses duality approach to select new variables to be added to master problem to improve its result.

Master Problem: it is same as defined in problem 14 except that NP is replaced with NP' (subset of NP which is feasible for 14). Solution of master problem shall provide a routing and slot-assignment scheme.

Subproblem: is a new problem created to identify a new node pattern to add to master problem and it is defined as:

$$\min_{s\in\{NP\setminus NP'\}} rp_s \qquad (15)$$

Here rp_s is reduced cost of node pattern 's' in the column generation algorithm and it is optimal value of following problem:

$$\max\left(1-\sum_{\{p|p\in s\}}\omega_{s,p}-\sum_{\{q|p\in s,e\in E_{s,p},rx_e=q\}}\tau_{s,q}\right) \quad (16)$$
$$s.t \qquad \omega_{s,p}+\tau_{s,p}-\phi_e c_{s,e}\geq 0$$

Here $\omega_{s,p}\ and\ \tau_{s,q}$ are variables that are associated with the transmitter p and the receiver q in node patterns s. Well the question is which node pattern should be included into NP' ?

According to duality theory if master problem is optimal then rp_s is always non-negative for any pattern in NP. The node patterns with negative rp_s can improve the result if they are added into NP'. So algorithm will iterate between two phases until no more patterns can be added to NP'.

Algorithm steps are defined as follows:

Step 1: Set node pattern A= ϕ and rp_A =0,

Step 2: Identify $A^c = v$ and compute $rp_{A'_v}$ for node pattern $A'_v\, s.t\, A'_v = \{A, v\}$.

Step 3: select v from A^c with minimum $rp_{A'_v}$ and compute rp_A of A.

Step 4: If $rp_{A'_v} \geq rp_A$, node v will be deleted from A^c and add it A.

Step 5: If $A^c \neq \phi$ stop else go to step 3.

V. SCHEDULING ALGORITHMS

Two scheduling algorithms are proposed in which each node will locally schedule its link transmissions without inter-node coordination and without disturbing interference profiles of other nodes.

A. Scheme 1

Every node 't' in node pattern 's' assigns a transmission probability to every link associated with 't'. The set of transmission probabilities is then defined as:

$$P_{s,t}=\left\{p_{s,t,e}\,|\,e\in E_{s,t}, p_{s,t,e}=\frac{\partial_{s,t,e}}{\mu_s}\right\} \qquad (17)$$

The region [0,1] is then divided into subregions, one for each link in $|E_{s,t}|$, and length of regions is set according to $P_{s,t}$. The algorithm works as follows; Suppose a node pattern 's' is activated in slot x. Each node $t \in s$ will generate a RV w, uniformly distributed within [0, 1]. The node will then schedule

link into which subregion w falls. If selected link (e) is not usable (either due to fading or no traffic) the scheduler will check link next to 'e' one by one until a usable link is found.

B. Scheme 2

Scheme 1 does not consider link quality while scheduling the links. Therefore authors presented another scheduling mechanism.

Selection criteria = (queue length * link capacity)

Each node maintains two queues for each of its link:

1) a real data queue to store packets and

2) A shadow queue for scheduling.

These queues of link 'e' whose transmitter can be activated in slot 'x', are defined as:

$$q_e(x) = q_e(x-1) + a_e(x) - d_e(x)$$
$$q'_e(x) = q'_e(x-1) + a'_e(x) - d'_e(x)$$

Here $q_e(x)$ and $q'_e(x)$ are lengths of the real queue and shadow queue respectively. $a_e(x)$, $a'_e(x)$, $d_e(x)$ and $d'_e(x)$ are the number of arrivals and departures for the two queues in 'x', respectively. In shadow queue the term $a'_e(x)$ is defined as: $a'_e(x) = \left(1 + \gamma_x\right)\sum_{t=0}^{x} a_e(t)$ i.e. it is used to smooth the incoming traffic from source or previous hop.

Packets departing from link 'e' are defined as: $d_e(x) = \min\left\{\tilde{c}_e(x), q_e(x)\right\}$. Here $\tilde{c}_e(x)$ is instant capacity of link 'e' in slot 'x'. Thus scheduling, in slot 'x', the scheduler in node $t \in s$ will select the link from all its associated links with a maximum value of $q'_e(x)\tilde{c}_e(x)$. In doing so, it tries to strike the optimal balance between link quality and traffic backlog.

VI. Simulation and Results

A. Simulation Environment and Settings

Linear optimization toolbox of MATLAB is used for proposed routing and slot-assignment algorithm. C++ program is then used to inspect maximum achievable throughput for different scheduling schemes.

The physical-layer parameters are summarized as follows:

- Transmission power: 20 dBm.
- Thermal noise:−90 dBm.
- Path loss(α):3.5.
- Variance of shadow fading: 4 dBm.

• Minimal distance of two nodes: 15 m.

• Slot duration: 0.22 ms.

• Frame size: 100 slots. So frame length = 22 ms.

The mapping between the following data rates and SINR threshold is summarized as follows.

• 54 Mb/s: 24.56 dBm.

• 48 Mb/s: 24.05 dBm.

• 36 Mb/s: 18.80 dBm.

• 24 Mb/s: 17.04 dBm.

• 18 Mb/s: 10.79 dBm.

• 12 Mb/s: 9.03 dBm.

• 9 Mb/s: 7.78 dBm.

• 6 Mb/s: 6.02 dBm.

Network Topology: two networks 15-node and 30-node with two gateway nodes and three gateway nodes are considered, respectively.

The traffic load of each flow is assumed to be the same i.e.,h_f =1Mb/s,

Throughput loss due to Frame Quantization:

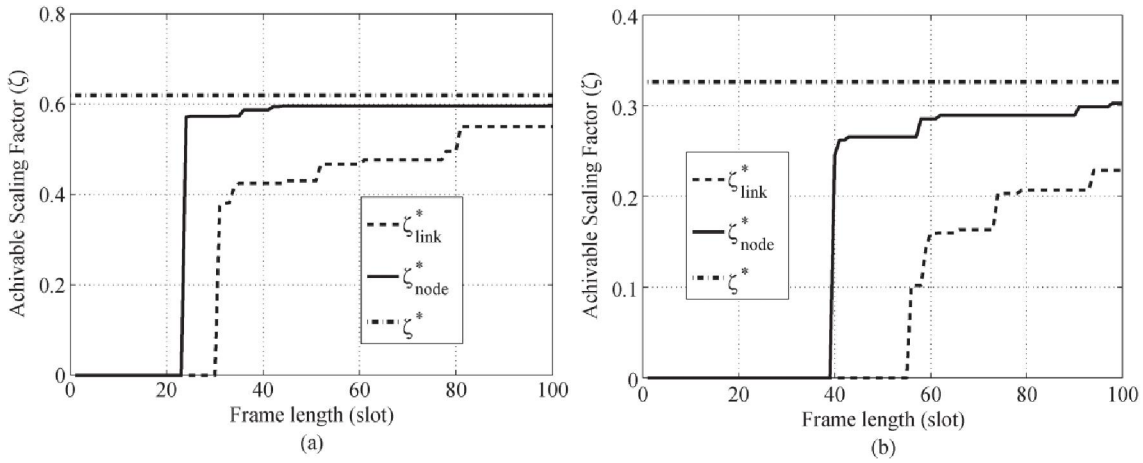


Fig. 1. Achievable throughput after frame generation for (a) 15- and (b) 30-node networks.

The solid (ζ^*_{node}), dashed (ζ^*_{link}) and dashed–dotted (ζ^*) lines indicate the achievable throughput in node-based, link based and before frame construction (I.e. upper bound on throughput) respectively. The flat area represents the range where the performance does not improve. Note that (ζ^*_{node}) and (ζ^*_{link}) are function of 'z' and are not always monotonically increasing due to the quantization involved in the process, and small oscillation occurs within a short range of z. This is why, in Fig. 1(a) and (b), the curves move up in steps.

TABLE I
PERFORMANCE COMPARISONS FOR THE LINK- AND NODE-BASED SCHEMES

N	ζ^*	ζ^*_{node} (ζ^*_{link})	n_{use}	$\zeta^*_{exp,1}$	$\zeta^*_{exp,2}$	$\zeta^*_{det,1}$	$\zeta^*_{det,2}$
15 (Node)	0.62	0.60	17	0.62	0.77	0.62	0.76
15 (Link)	0.62	0.55	26	NA	NA	NA	NA
30 (Node)	0.33	0.30	30	0.31	0.39	0.31	0.39
30 (Link)	0.33	0.23	54	NA	NA	NA	NA

From the table it is clear that ζ^*_{node} approaches ζ^* much faster than ζ^*_{link}. Moreover difference in throughput between ζ^*_{node} and ζ^*_{link} is also significant as shown in table 1.

The optimal scaling factors of the ζ^*_{node} for schemes 1 and 2 under the Poisson and a deterministic arrival process are denoted as shown by $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{exp},2}$, $\zeta^*_{\text{det},1}$, $\zeta^*_{\text{det},2}$ respectively.

{the ζ^*_{node} is derived from problem 14 and it does not include multi-user diversity gain. Therefore, it can be viewed as a lower bound of the two proposed scheduling schemes 1 and 2 As shown in the Table I.

It is also clear from the table that , both (posisson and deterministic arrival rates) $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{det},1}$ are only slightly larger than ζ^*_{node} for the 15- and 30-node networks. The difference is only about 3%. This is because scheme 1 tries to follow $\grave{o}_{s,p,e}$ i.e. portion of time that is assigned to each link of node p in node pattern s and does not select a link with the best quality. However, the situation is different in scheme 2, because link quality is part of the selection criteria. With scheme 2, $\zeta^*_{\text{exp},1}$, $\zeta^*_{\text{det},1}$ are about 26% larger than ζ^*_{node} for the 15-node network and 30% larger for the 30-node network.

A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration

J. M. Bioucas-Dias, M. A. T. Figueiredo

IEEE transactions on image processing 2007.

Summary:

- ✓ It happens that the convergence rate of IST algorithms depends heavily on the linear observation operator, becoming very sow when this operator is ill-conditioned or ill-posed.
- ✓ In this paper, the authors introduce two-step IST (TwIST) algorithms, exhibiting much faster convergence rate than IST for ill-conditioned problems. They showed that TWIST converges to a minimizer of the objective function, for a given range of values of its parameters.

I. Introduction

Many approaches to LIPs define a solution $\hat{\mathbf{x}}$ as a minimizer of a convex objective function *f*:

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{K}\mathbf{x}\|^2 + \lambda\Phi(\mathbf{x}) \quad (1)$$

In a regularization framework, minimizing *f* is seen as a way of overcoming the ill-conditioned, or singular, nature of **K** which precludes inverting it. In this context, Φ is called the regularizer and λ the regularization parameter.

The current state-of-the-art regularizers for image restoration are nondifferentiable.

Examples of such choices are total-variation (TV) regularization and wavelet-based regularization.

The nondifferentiable nature of *f*, together with the huge dimension of its argument, place its minimization beyond the reach of standard off-the-shelf optimization methods.

Contribution

This paper is strictly concerned with algorithms for minimizing (1).

This paper introduces a new class of iterative schemes, bringing together the best of IRS and IST. For ill-conditioned (but invertible) linear observation operators, they prove (linear) convergence of TwIST to minima of the objective function f, for a certain range of the algorithm parameters, and derive bounds for the convergence factor.

II. Regularizers and Denoising

Denoising with convex regularizers

Denoising problems are LIPs in which **K** is the identity, **Kx**=**x**.

$$f_{den} = \frac{1}{2} d_y^2 + \lambda\Phi \text{ . where } \quad d_y^2 = \left\|\mathbf{y} - \mathbf{x}\right\|^2 \tag{2}.$$

With several standard assumptions about the regularizer $\Phi : \chi \to R$ (convex, lower semi-continuous (lsc), proper), its minimizer is unique; refer to Theorem 5 and Theorem 7 in appendix I. This allows defining the denoising function.

$$\mathbf{\Psi}_\lambda(\mathbf{y}) = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} d_y^2(\mathbf{x}) + \lambda\Phi(\mathbf{x}) \right\} \tag{3}.$$

Denoising with l-Homogeneous Regularizers

Let $\Upsilon(\chi)$ denote the set of functions $\Phi : \chi \to R$ that are convex, lsc, proper, and phd-1.

An important recent result states that denoising with regularizers from $\Upsilon(\chi)$ corresponds to the residual of the projection onto a convex set, as formalized in the following theorem.

Theorem 1: If $\Phi \in \Upsilon(\chi)$, then the denoising function $\mathbf{\Psi}_\lambda$ defined in (3) is given by

$$\mathbf{\Psi}_\lambda(\mathbf{y}) = \mathbf{y} - P_{\lambda C}(\mathbf{y}) \tag{4}$$

Where $C \subset \chi$ is a closed convex set depending on the regularizer Φ, and $P_A : \chi \to \chi$ denotes the orthogonal projection operator onto the convex set $A \subset \chi$.

Total variation

$\Phi_{iTV}, \Phi_{niTV} \in \Upsilon(R^m)$

Weighted lp norm

$$\Phi_{l_w^p}(\mathbf{x}) = \left\|\mathbf{x}\right\|_{p,w} = \left(\sum_i w_i \left| x_i \right|^p \right)^{1/p}$$

Being a norm, $\Phi_{l_w^p}$ clearly belongs to Υ.

The denoising function $\mathbf{\Psi}_\lambda$ under a $\Phi_{l_w^p}$ regularizer cannot be optained in a closed form, except in some particular cases, the most notable of which is p=1; in this case, $\mathbf{\Psi}_\lambda$ is the well known soft-thresholding function, that is $\mathbf{\Psi}_\lambda(\mathbf{z}) = \hat{\mathbf{x}}$ with

$$\hat{x}_i = sign(z_i) \max\left\{0, \left|z_i\right| - \lambda w_i\right\} \tag{8}.$$

Orthogonal representations

$$f(\mathbf{x}) = \frac{1}{2} d_{\mathbf{y}}^2(\mathbf{HWx}) + \lambda \Phi_{l_w^p}(\mathbf{x}) \quad (9).$$

III. Existence and uniqueness of solutions

Proposition 1: Let $f : X \to \bar{R}$ be defined as in (1), where operator **K** is linear and bounded, and Φ is a proper, lsc, convex function. Let G denote the set of minimizers of *f*. Then:

i) If Φ is coercive, then G is nonempty;

ii) If Φ is strictly convex or **K** is injective, then G contains at most one element;

iii) If **K** is bounded below, then G contains exactly one element.

Application of Proposition 1 to the several regularization function.

Weighted l-p norm and its p-th power

If all the weights are strictly positive, both are coercive; this ensures existence of minimizers of *f*. If K is injective, the minimizer is unique; otherwise, the minimizer is unique with $\Phi_{l_w^p}^p$, with *p*>1 (which is strictly convex).

Finite-dimensional cases.

Injectivity of K is sufficient to guarantee existence and uniqueness of the solution (under any convex regularizer, strictly or not, coercive or not). It is because any finite-dimensional injective operator is bounded below.

IV. Previous algorithms

Iterative shrinkage/thresholding (IST)

$$\mathbf{x}_{t+1} = (1-\beta)\mathbf{x}_t + \beta \boldsymbol{\Psi}_\lambda(\mathbf{x}_t + \mathbf{K}^T(\mathbf{y} - \mathbf{K}\mathbf{x}_t)) \quad (13)$$

Each iteration of the IST algorithm only involves sums, matrix-vector products by **K** and $\mathbf{K}^T$, and the application of the denoising operation $\boldsymbol{\Psi}_\lambda$.

Theorem 2: Let f be given by (1), where Φ is convex (and lsc) and $\|\mathbf{K}\|_2^2 < 2$. Let *G* be nonempty. Fix some $\mathbf{x}_1$ and let the sequence be produced by (13), with $\beta \in [0,1]$. Then the sequence converges to a point $\mathbf{x} \in G$.

Iterative Re-Weighted Shrinkage (IRS)

The IRS algorithm was specifically designed for wavelet-based problems of the form (9), where W contains an orthogonal or redundant wavelet basis and the regularizer is not necessarily a weighted lp norm.

$$\mathbf{x}_{t+1} = \text{solution}\left\{\mathbf{A}_t\mathbf{x} = \mathbf{b}\right\} = \text{solution}\left\{(\lambda\mathbf{D}_t + \mathbf{K}^T\mathbf{K})\mathbf{x} = \mathbf{K}^T\mathbf{y}\right\}$$

$\mathbf{D}_t$ is a diagonal matrix (of non-negative elements) that depends on $\mathbf{x}_t$ and Φ. Observe that matrix $\mathbf{D}_t$ shrinks the components of $\mathbf{x}_{t+1}$, thus the term iterative reweighted shrinkage.

The huge size of $\mathbf{A}_t = (\lambda\mathbf{D}_t + \mathbf{K}^T\mathbf{K})$ forces the use of iterative methods to implement. This is done with a two-step stationary iterative method, which we will next briefly review.

Two-Step Methods for Linear Systems

Considering the linear system $\mathbf{Ax} = \mathbf{b}$, with **A** positive definite; define a so-called splitting of **A** as **A=C-R**, such that **C** is positive definite and easy to invert (e.g., a diagonal matrix). A stationary two-step iterative method (TwSIM) for solving **Ax=b** is defined as

Two-step iterative method (TwSIM)

$$\mathbf{x}_1 = \mathbf{x}_0 + \beta_0\mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$$

$$\mathbf{x}_{t+1} = (1-\alpha)\mathbf{x}_{t-1} + \alpha\mathbf{x}_t + \beta\mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_t) \qquad (15).$$

Theorem 3: Let $\{x_t, t \in \mathbb{N}\}$ be the sequence produced by (15), with arbitrary $\mathbf{x}_0$. Let λ_1 and λ_m denote the smallest and largest eigenvalues of matrix $\mathbf{C}^{-1}\mathbf{A}$, and $k = \lambda_1 / \lambda_m$. Then, $\{x_t, t \in \mathbb{N}\}$ converges to the solution of Ax=b if and only if $0 < \alpha < 2$ and $0 < \beta < 2\alpha / \lambda_m$. The optimal asymptotic convergence factor is $\rho = (1-\sqrt{k}) / (1+\sqrt{k})$.

Comparing IST with IRS

For ill conditioned systems, IRS is much faster than IST. On the other hand, when noise is the main factor, and the observation operator is not too ill-conditioned, IST outperforms IRS because it uses a closed-form denoising step in each iteration.

V. Two-Step IST (TwIST)

The TwIST method aims at keeping the good denoising performance of the IST scheme, while still being able to handle ill-posed problems as efficiently as the IRS algorithm.

Taking $\mathbf{C} = \mathbf{I} + \lambda\mathbf{D}_t$ and $\mathbf{R} = \mathbf{I} - \mathbf{K}^T\mathbf{K}$ in the splitting **A=C-R** of matrix $\mathbf{A} = \lambda\mathbf{D}_t + \mathbf{K}^T\mathbf{K}$, the two-step iteration (15) for the linear system $\mathbf{Ax} = \mathbf{K}^T\mathbf{y}$ becomes

$$\mathbf{x}_{t+1} = (1-\alpha)\mathbf{x}_{t-1} + (\alpha-\beta)\mathbf{x}_t + \beta\mathbf{C}^{-1}(\mathbf{x}_t + \mathbf{K}^T(\mathbf{y} - \mathbf{K}\mathbf{x}_t)) \quad (16)$$

Equation (13) can be obtained from (16) by setting α =1 and replacing the multiplication by matrix $\mathbf{C}^{-1}$ by the denoising operator $\mathbf{\Psi}_\lambda$. This similarity suggests a two-step version of IST (TwIST) as

$$\mathbf{x}_1 = \mathbf{\Gamma}_\lambda(\mathbf{x}_0) \quad (17)$$

$$\mathbf{x}_{t+1} = (1-\alpha)\mathbf{x}_{t-1} + (\alpha-\beta)\mathbf{x}_t + \beta\mathbf{\Gamma}_\lambda(\mathbf{x}_t) \quad (18)$$

for $t \geq 1$, where $\mathbf{\Gamma}_\lambda : R^m \to R^m$ is defined as

$$\mathbf{\Gamma}_\lambda(\mathbf{x}) = \mathbf{\Psi}_\lambda(\mathbf{x} + \mathbf{K}^T(\mathbf{y} - \mathbf{K}\mathbf{x})) \quad (19).$$

A key observation is that TwIST and IST all have the same fixed points.

VI. Experimental Results

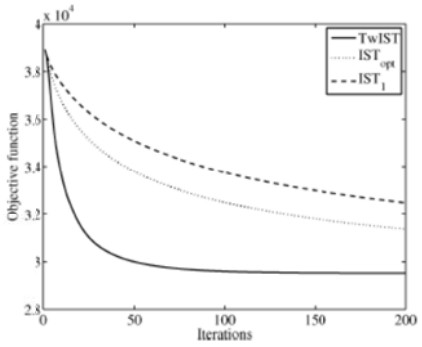


Fig. 1. TV-based deconvolution in a severely ill-conditioned problem (experiment 1). Evolution of the objective function $f(\mathbf{x}_t)$ produced by TwIST, $\mathrm{IST}_{\mathrm{opt}}$, and IST_1.

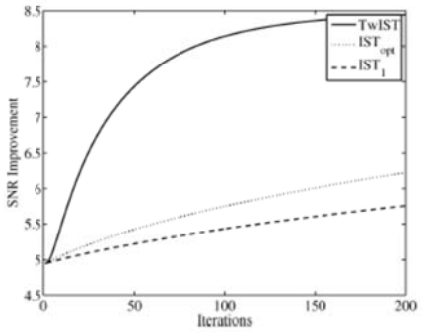


Fig. 2. TV-based deconvolution in a severely ill-conditioned problem (experiment 1). Evolution of the SNR improvement (ISNR) produced by TwIST, $\mathrm{IST}_{\mathrm{opt}}$, and IST_1.

INFONET, GIST
Journal Club (2013. 03. 21)

Scaling Up MIMO: Opportunities and challenges with very large arrays

Authors: Fredrik Rusek, Thomas L. Marzetta, et al.
Publication: IEEE Signal Processing Magazine
Speaker: Woongbi Lee

Short summary: Very large MIMO systems is an emerging research area in antenna systems, electronics, and wireless communication systems. A base station with an antenna array serves a multiplicity of single-antenna terminals. In this presentation, the fundamental principle of massive MIMO technology and several issues are introduced.

I. INTRODUCTION

Multiple-Input, Multiple-Output (MIMO) technology is becoming mature, and incorporated into emerging wireless broadband standard like LTE. Basically, the more antennas the transmitter/receiver is equipped with, and the more degrees of freedom that the propagation channel can provide, the better performance in terms of data rate or link reliability. However, MIMO technology requires increased complexity of the hardware, the complexity and energy consumption of the signal processing, and the physical space for accommodating antennas including rents of real estate.

Today, as mobile data traffic exponentially increases, further capacity enhancement is needed. As a solution for the high capacity demand, Massive MIMO (very large MIMO, Large-Scale Antenna System, Full Dimension MIMO) technology has been widely studied for last few years. Massive MIMO adopts hundreds of antennas at base station (BS) serving a much smaller number of terminals. The number of terminals that can be simultaneously served is limited, not by the number of antennas, but rather by inability to acquire channel-state information for an unlimited number of terminals. With an unlimited number of antennas, the transmit power can be made arbitrarily small and the uncorrelated interference and noise can be vanished. But, the performance is limited by pilot contamination.

This paper approaches to Massive MIMO according to three directions: Information-theoretic performance limit, and antennas and propagation aspects of large MIMO, and transmit and receive schemes.

II. Information Theory for Very Large MIMO Arrays

According to the noisy-channel coding theorem in information theory, for any communication link, there is a capacity or achievable rate, such that for any transmission rate less than the capacity, there exists a coding scheme that makes the error-rate arbitrarily small.

A. Point-to-point MIMO

1) Channel model

Transmitter has an array of n_t antennas and a receiver has an array of n_r antennas. The simplest narrowband memoryless channel has the following mathematical description,

$$\mathbf{x} = \sqrt{\rho}\mathbf{G}\mathbf{s} + \mathbf{w}$$

where $\boldsymbol{s}$ is the n_t component vector of transmitted signals, $\boldsymbol{x}$ is the n_r component vector of received signals, $\boldsymbol{G}$ is the $n_r \times n_t$ propagation matrix of complex-valued channel coefficients, and $\boldsymbol{w}$ is the n_r component vector of receiver noise. The components of the additive noise vector are i.i.d. zero mean and unit-variance circular-symmetric complex-Gaussian random variables ($CN(0,1)$). The scalar ρ is a measure of the Signal-to-Noise Ratio (SNR) of the link.

2) Achievable rate

With the assumption that the receiver has perfect knowledge of the channel matrix, $\boldsymbol{G}$, the mutual information between the input and the output of the point-to-point MIMO channel is

$$C = I(x;s) = \log_2 \det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{G}\mathbf{G}^H\right)$$

where $\mathbf{I}_{n_r}$ denotes the $n_r \times n_r$ identity matrix. The propagation matrix can be decomposed by

$$\mathbf{G} = \mathbf{\Phi}\mathbf{D}_v\mathbf{\Psi}^H,$$

where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are unitary matrices of dimension $n_r \times n_r$ and $n_t \times n_t$ respectively, and $\mathbf{D}_\nu$ is a $n_r \times n_t$ diagonal matrix whose diagonal elements are the singular values, $\left\{\nu_1, \nu_2, \cdots, \nu_{\min(n_t, n_r)}\right\}$. The achievable rate can be written as

$$C = \sum_{l=1}^{\min(n_t, n_r)} \log_2\left(1 + \frac{\rho \nu_l^2}{n_t}\right)$$

With the decomposed propagation matrix,

$$\sum_{l=1}^{\min(n_t, n_r)} \nu_l^2 = \mathrm{Tr}\left(\mathbf{G}\mathbf{G}^H\right)$$

where "Tr" denotes "trace". There can exist two extreme cases: the worst case when all except one of the singular values are equal to zero and the best case when all of the $\min\left(n_t, n_r\right)$ singular values are equal. The two cases bound the achievable rate as follows,

$$\log_2\left(1 + \frac{\rho \cdot \mathrm{Tr}\left(\mathbf{G}\mathbf{G}^H\right)}{n_t}\right) \le C \le \min\left(n_t, n_r\right) \cdot \log_2\left(1 + \frac{\rho \cdot \mathrm{Tr}\left(\mathbf{G}\mathbf{G}^H\right)}{n_t \min\left(n_t, n_r\right)}\right)$$

The rank-1 (worst) case occurs either for compact arrays under Line-of-Sight (LOS) propagation conditions such that the transmit array cannot resolve individual elements of the receive array and vice-versa, or under extreme keyhole propagation conditions. The equal singular value (best) case is approached when the entries of the propagation matrix are IID random variables. Under favorable propagation conditions and a high SNR, the achievable rate is proportional to the smaller of the number of transmit and receive antennas.

3) Limiting cases

a) Low SNRs

Low SNRs can be experience by terminals at the edge of a cell. For low SNRs, only beamforming gains are important and the achievable rate becomes

$$C_{\rho \to 0} \approx \frac{\rho \cdot \mathrm{Tr}\left(\mathbf{G}\mathbf{G}^H\right)}{n_t \ln 2}$$

$$\approx \frac{\rho n_r}{\ln 2}$$

which is independent of n_t, and thus, even under the most favorable propagation conditions the multiplexing gains are lost, and multiple transmit antennas are of no value under low SNRs.

b) Number of transmit antennas grow large

It is assumed that the row-vectors of the propagation matrix are asymptotically orthogonal. Then,

$$\left(\frac{\mathbf{G}\mathbf{G}^H}{n_t}\right)_{n_t>>n_r} \approx \mathbf{I}_{n_r}$$

and the achievable rate becomes

$$\begin{aligned} C_{n_t>>n_r} &\approx \log_2 \det\left(\mathbf{I}_{n_r} + \rho\cdot\mathbf{I}_{n_r}\right) \\ &= n_r\cdot\log_2\left(1+\rho\right) \end{aligned}$$

c) Number of receive antennas grow large

It is also assumed that the column-vectors of the propagation matrix are asymptotically orthogonal,

$$\left(\frac{\mathbf{G}^H\mathbf{G}}{n_r}\right)_{n_r>>n_t} \approx \mathbf{I}_{n_t}$$

and the achievable rate becomes

$$\begin{aligned} C_{n_r>>n_t} &= \log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\cdot\mathbf{G}^H\mathbf{G}\right) \\ &\approx n_t\cdot\log_2\left(1+\frac{\rho n_r}{n_t}\right) \end{aligned}$$

Thus, a large number of transmit or receive antennas, combined with asymptotic orthogonality of the propagation vectors (i.i.d. complex Gaussian), can increase the achievable rate. Extra receive antennas can compensate for a low SNR and restore multiplexing gains.

B. Multi-user MIMO

Multi-user MIMO consists of an array of *M* antennas and *K* autonomous terminals. We assume that each terminal has only one antenna. Multi-user MIMO differs from point-to-point MIMO in two respects: first, the terminals are typically separated by many wavelengths, and second, the terminals cannot collaborate among themselves.

1) Propagation

We assume Time Division Duplex (TDD), so the reverse link propagation matrix is merely the transpose of the forward link propagation matrix. Assumption on the TDD comes from the need to acquire channel state-information between extreme numbers of service antennas and much smaller numbers of terminals. The propagation matrix, $\mathbf{G} \in \mathbb{R}^{M \times K}$, can be decomposed by

$$\mathbf{G} = \mathbf{H}\mathbf{D}_\beta^{1/2}$$

where $\mathbf{H} \in \mathbb{R}^{M \times K}$ represents small scale fading and $\mathbf{D}_\beta^{1/2} \in \mathbb{R}^{K \times K}$ whose diagonal elements constitute a $K \times 1$ vector, and β is large scale fading coefficients. By assumption, the antenna array is sufficiently compact that all of the propagation paths for a particular terminal are subject to the same large scale fading.

For multi-user MIMO with large arrays, the number of antennas greatly exceeds the number of terminals. Under the most favorable propagation conditions the column-vectors of the propagation matrix are asymptotically orthogonal,

$$\begin{aligned}\left(\frac{\mathbf{G}^H\mathbf{G}}{M}\right)_{M>>K} &= \mathbf{D}_\beta^{1/2}\left(\frac{\mathbf{H}^H\mathbf{H}}{M}\right)_{M>>K}\mathbf{D}_\beta^{1/2} \\ &\approx \mathbf{D}_\beta\end{aligned}$$

2) Reverse link

On the reverse link, for each channel use, the K terminals collectively transmit a $K \times 1$ vector of QAM symbols, $\mathbf{q}_r$, and the antenna array receives a $M \times 1$ vector, $\mathbf{x}_r$,

$$\mathbf{x}_r = \sqrt{\rho_r}\mathbf{G}\mathbf{q}_r + \mathbf{w}_r$$

Under the assumption that the columns of the propagation matrix are nearly orthogonal, i.e., $\mathbf{G}^H\mathbf{G} = M \cdot \mathbf{D}_\beta$, the base station could process its received signal by a matched-filter (MF),

$$\begin{aligned}\mathbf{G}^H\mathbf{x}_r &= \sqrt{\rho_r}\mathbf{G}^H\mathbf{G}\mathbf{q}_r + \mathbf{G}^H\mathbf{w}_r \\ &\approx M\sqrt{\rho_r}\mathbf{D}_\beta\mathbf{q}_r + \mathbf{G}^H\mathbf{w}_r\end{aligned}$$

3) Forward link

For each use of the channel the base station transmits a $M \times 1$ vector, $\mathbf{s}_f$, through its M antennas, and the K terminals collectively receive a $K \times 1$, $\mathbf{x}_f$,

$$\mathbf{x}_f = \sqrt{\rho_f}\mathbf{G}^T\mathbf{s}_f + \mathbf{w}_f$$

III. ANTENNA AND PROPAGATION ASPECTS OF VERY LARGE MIMO

The performance of all types of MIMO systems strongly depends on properties of the antenna arrays and the propagation environment in which the system is operating. With well separated ideal antenna elements, in a sufficiently complex propagation environment and without directivity and mutual coupling, each additional antenna element in the array adds another degree of freedom that can be used by the system. But, in reality, the antenna elements are not ideal, they are not always well separated, and the propagation environment may not be complex enough to offer the large number of degrees of freedom that a large antenna array could exploit. These practical issues are presented in this section.

A. Spatial focus with more antennas

The field strength is not necessarily focused in the direction of the intended receiver, but rather to a geographical point where the incoming multipath components add up constructively. As a technique for focusing transmitted energy to a specific location, Time Reversal (TR) has drawn attention, where the transmitted signal is a time-reversed replica of the channel impulse response. In this paper, the Time-Reversal Beam Forming (TRBF) is considered.

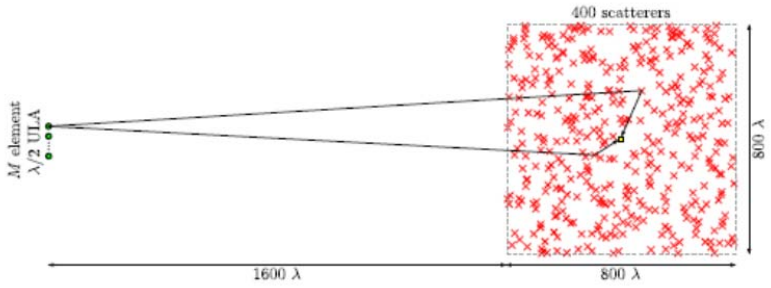


Figure 1. Geometry of the simulated dense scattering environment.

Figure 1 shows a simple geometrical channel model. The channel is composed of 400 uniformly distributed scatterers in a square of dimension $800\lambda \times 800\lambda$, where λ is the signal wavelength. The broadside direction of the *M*-element Uniform Linear Array (ULA) with adjacent element spacing $d = \lambda / 2$ is pointing towards the center of the scatterer area. This model creates a field strength that varies rapidly over the geographical area, typical of

small-scale fading. With a complex enough scattering environment and a sufficiently large element spacing in the transmit array, the filed strength resulting from different elements in the transmit array can be seen as independent.

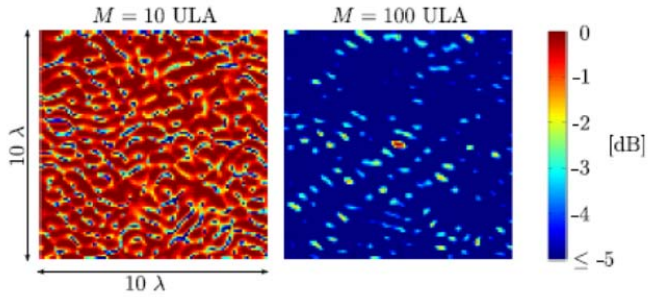
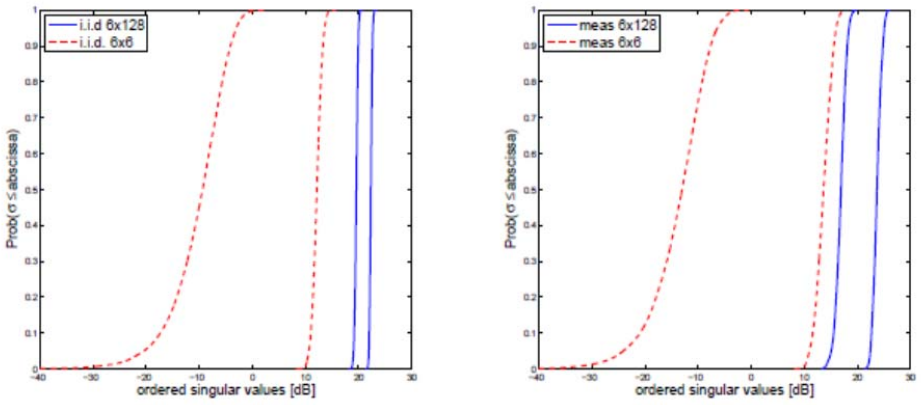


Figure 2. Normalized filed strength in a $10\lambda \times 10\lambda$ area

Figure 2 shows the resulting normalized field strength in a small $10\lambda \times 10\lambda$ environment around the receiver to which we focus the transmitted signal (using MF precoding), for ULAs with $d = \lambda/2$ of size $M = 10$ and $M = 100$ elements. Figure 2 illustrates two important properties of the spatial MF precoding: (i) that the field strength can be focused to a point rather than in a certain direction and (ii) that more antennas improve the ability to focus energy to a certain point, which leads to less interference between spatially separated users.

B. Antenna aspects

Massive MIMO relies to a large extent on a property of the radio environment called *favorable propagation*. Favorable propagation means that propagation channel responses from the base station to different terminals are sufficiently different. One way of quantifying how different the channel responses to different terminals are, is to look at the spread between the smallest and largest singular value of the channel. Figure 3(a) shows this for a computer simulated "i.i.d." channel. The figure shows the cumulative density function for the smallest respectively the largest singular value for two cases: A conventional array of 6 elements serving 6 terminals (red curves), and a massive array of 128 elements serving 6 terminals (blue curves)



(a) Theory. (b) Measured results.

Figure 3. Singular value spread of massive MIMO channels

For the 6-element array, the singular value spread is about 30 dB, meaning that 1000 times more power would be required to serve all six terminals, as compared to the power required to serve just one of them. With the massive array, the gap is less than 3 dB.

In real channel implementation, measurements were conducted using an indoor 128-antenna base station consisting of four stacked double polarized 16 element circular patch arrays as shown in Fig. 4. Three of the terminals are indoors at various positions and 3 users are outdoors. The measurements were performed at 2.6 GHz with a bandwidth of 50 MHz, and the results were averaged over this bandwidth and over a physical displacement of 10 meters.

Figure 4. Massive MIMO antenna used in measurements

The blue curves in Fig 3. (b) show the corresponding singular value distributions. It is striking how well reality resembles the ideal case in Fig. 3. (a). The spread between the smallest and the largest singular value is a bit larger than for the ideal case, but the probability that the spread

exceeds 10 dB is negligible. As a reference, Fig. 3(b) also shows the result when only 6 of the 128 elements are activates (red curves). Overall, there is compelling evidence that the assumption on favorable propagation that underpin massive MIMO are substantially valid in practice.

IV. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

In the antenna implementation, they compared system of 6-antenna array and 6 terminals with single antenna and system of 128-antenna array and 6 terminals with single antenna. They used $f_c[\text{cycle}/\text{sec}] = 2.6\times10^9 \approx 3.0\times10^9$, $c\left[\text{m}/\text{sec}\right] = 3.0\times10^8$. Thus, the wavelength $\lambda\left[\text{m}/\text{cycle}\right] = \frac{c}{f_c} = 0.1$. According to the wavelength, if they used $d = \lambda/2$, then the distance between patched antennas is $d = \lambda/2 = 50\text{cm}$.

Appendix

Reference

[1]

Shrinkage Methods to Sparse Linear Regression

Main ref.:T. Hastie, R. Tibshirani, and J. Friedman,

The Elements of Statistical Learning 2nd Edition, Springer 2008 [1]

Presenter: Jaewook Kang

CS Journal Club in GIST, Apr. 2013

This report will be appeared in Ph.D. dissertation of Jaewook Kang

Abstract

In this report, we introduce linear regression approaches using shrinkage methods. The shrinkage method have got attention to solve the problem of linear systems $\mathbf{y} = \mathbf{A}\mathbf{x}$ because the method enables us to obtain the solution with lower variance than the conventional least square estimator having the minimum variance unbiasedness. First we will introduce basic concept of two shrinkage methods in the linear regression, *ridge and lasso*. Then, we move our focus to problems of the Lasso variants such as *Fused lasso* and *Elastic-net*. For the discussion in this report, we have partially referred to the chapter 3 of the book [1].

I. Introduction

A linear regression problem starts from an assumption that the corresponding regression function $\underline{Y} = f(\underline{X})$ is linear where $\underline{Y} \in \mathbb{R}^M$ is a measurement vector generated by the function $f(\cdot)$ given a vector $\underline{X} \in \mathbb{R}^N$. This assumption allows us to describe the function $f(\cdot)$ using a linear projection, given by

$$\underline{Y} = \mathbf{A}\underline{X} \in \mathbb{R}^M, \tag{1}$$

where a measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ specifies the linear relation between $\underline{Y}$ and $\underline{X}$. In such a regression problem, a typical aim is to estimate the unknown vector $\underline{X}$ from a set of known inputs or training data $(Y_1, \underline{a}_{\text{1st-row}})$ ... $(Y_M, \underline{a}_{\text{Mst-row}})$ where $\underline{a}_{\text{jth-row}} = [a_{j1}, a_{j2}, ..., a_{jN}]$ denotes the j-th row vector of the matrix $\mathbf{A}$. In addition, as before, we confine our focus to the linear regression problems which is underdetermined $(M < N)$ such that there exists infinitely many solutions for $\underline{X}$.

The most standard approach to the linear regression problems is *least square estimation* (LSE). LSE obtains its estimate by solving the following optimization problem, given by

$$\begin{aligned}(P_{\text{LSE}}) : \underline{\widehat{X}}_{\text{LSE}} &= \arg\min_{\underline{X}} \|\underline{Y} - \mathbf{A}\underline{X}\|_2^2 \\ &= \arg\min_{\underline{X}} \sum_{j=1}^{M} \left(Y_j - \sum_{i=1}^{N} a_{ji} X_i \right)^2 . \end{aligned} \tag{2}$$

It is well known that the LSE solver obtains an estimate $\underline{\widehat{X}}_{\text{LSE}}$ by projecting the measurement vector $\underline{Y}$ to a subspace of $\mathbb{R}^M$ spanned by the column vectors of the matrix $\mathbf{A}$. Namely, the minimization task in (2) chooses $\underline{\widehat{X}}_{\text{LSE}}$ which makes the vector difference $\underline{Y} - \mathbf{A}\underline{\widehat{X}}_{\text{LSE}}$ to be orthogonal to the subspace. Such a LSE solution can be represented as a linear function of the measurement vector, *i.e.*,

$$\underline{\widehat{X}}_{LSE} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\underline{Y}. \tag{3}$$

The popularity of LSE is originated from the Gauss-Markov theorem, one of the famous results in statistics. The Gauss-Markov theorem states that the LSE solver provides the smallest variance among all linear unbiased estimators. Let $\underline{\widetilde{X}}$ denote an unbiased linear estimate, *i.e.*, $E\left[\underline{\widetilde{X}}\right] = \underline{X}$. The mean squared error (MSE) of $\underline{\widetilde{X}}$ is calculated as

$$\begin{aligned}\text{MSE}\left(\underline{\widetilde{X}}\right) :=& E\left[\left(\underline{\widehat{X}} - \underline{X}\right)^2\right] = \text{Var}\left(\underline{\widetilde{X}}\right) + \left(E\left[\underline{\widehat{X}}\right] - \underline{X}\right)^2 \\ =& \text{Var}\left(\underline{\widetilde{X}}\right). \end{aligned} \tag{4}$$

Then, the Gauss-Markov theorem shows that

$$\text{Var}\left(\underline{\widehat{X}}_{LSE}\right) \leq \text{Var}\left(\underline{\widetilde{X}}\right) \tag{5}$$

for any other unbiased linear estimate $\underline{\widetilde{X}}$ (We omit the proof here. please refer to [1]).

However, there may exist a biased estimator which can offer smaller MSE then the LSE solver. Such an estimator would provide a significant reduction in MSE at the expense of losing the unbiasedness [1],[3]. This is one motivation to use the shrinkage method to linear regression problem. The shrinkage methods is a biased estimation approach to impose a penalty to the optimization setting of (2). If the imposed penalty can properly catch the characteristic of the target unknown $\underline{X}$, the shrinkage methods greatly improve the estimation accuracy. In this report, we first introduce two types of the most well-known shrinkage methods, *ridge* and *lasso*, by partially referring to [1],[3]. Then, we extend our discussion to shrinkage methods which estimates the vector $\underline{X}$ having piecewise smooth or approximately sparse property.

II. SHRINKAGE WITH RIDGE PENALTY

In ridge regression, the elements of $\underline{X}$ are shrunk by imposing a penalty on the energy of $\underline{X}$ [2]. Therefore, the ridge penalty takes a quadratic form of $\underline{X}$, leading to the following optimization setup

$$\begin{aligned}(P_{\text{Ridge}}) : \underline{\widehat{X}}_{\text{Ridge}} =& \arg\min_{\underline{X}} \|\underline{Y} - \mathbf{A}\underline{X}\|_2^2 + \lambda \|\underline{X}\|_2^2 \\ =& \sum_{j=1}^{M} \left(y_j - \sum_{i=1}^{N} a_{ji} X_i \right)^2 + \lambda \sum_{i=1}^{N} X_i^2, \end{aligned} \tag{6}$$

where $\lambda \geq 0$ denote a parameter to control the amount of ridge shrinkage. Note that by applying the quadratic penalty $\|\underline{X}\|_2^2 = \underline{X}^T \underline{X}$, the ridge estimation can be represented as a closed form function of $\underline{Y}$ if $M \geq N$, given by

$$\underline{\widehat{X}}_{\text{Ridge}} = (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I}_N)^{-1} \mathbf{A}^T \underline{Y}. \tag{7}$$

This imposition of the ridge penalty adds a positive constant to the diagonal $\mathbf{A}^T\mathbf{A}$ in (7) before inversion. It is noteworthy that this addition makes the regression problem nonsingular even when $\mathbf{A}^T\mathbf{A}$ does not have full rank. Namely, the ridge solution is necessarily unique regardless of the condition of the matrix $\mathbf{A}$. This is a strong motivation to use ridge regression.

Ridge regression shrinks the coordinate of $\underline{\widehat{X}}_{\text{Ridge}}$ according to the singular value of the matrix $\mathbf{A}$. The singular value decomposition (SVD) of $\mathbf{A}$ has the form

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{8}$$

where $\mathbf{U} \in \mathbb{R}^{M \times N}$ and $\mathbf{V} \in \mathbb{R}^{N \times N}$ are orthogonal matrices, and $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with singular values $d_1 \geq d_2 \geq ... \geq d_N \geq 0$ of $\mathbf{A}$. By applying SVD to the ridge solution, we can efficiently compute a ridge estimate $\underline{\widehat{X}}_{\text{Ridge}}$ associated with the orthonormal basis $\mathbf{U}$ and $\mathbf{V}$, as LSE does using the QR decomposition

$$\begin{aligned}\underline{\widehat{X}}_{\text{Ridge}} =& (\mathbf{A}^T\mathbf{A} + \lambda \mathbf{I}_N)^{-1} \mathbf{A}^T \underline{Y} \\ =& \mathbf{V}(\mathbf{D}^2 + \lambda \mathbf{I}_N)^{-1} \mathbf{D}\mathbf{U}^T \underline{Y} \\ =& \sum_{i=1}^{N} \underline{v}_i \frac{d_i}{d_i^2 + \lambda} \underline{u}_i^T \underline{Y}, \end{aligned} \tag{9}$$

where the $\underline{u}_i \in \mathbb{R}^M$ and $\underline{v}_i \in \mathbb{R}^N$ are the column vectors of $\mathbf{U}$ and $\mathbf{V}$ respectively. In (9), ridge regression shrinks the elements of $\underline{\widehat{X}}_{\text{Ridge}}$ by the factors $d_i/(d_i^2 + \lambda)$. This means that a greater amount of shrinkage is applied to the elements of $\underline{\widehat{X}}_{\text{Ridge}}$ associated with $\underline{v}_i$ having smaller singular values d_i. Namely, ridge

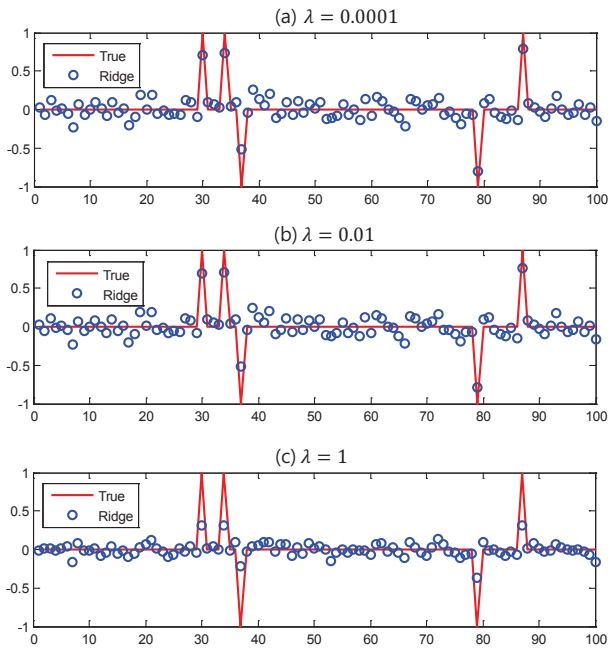


Fig. 1. Sparse estimation via ridge estimator with different λ where $N = 100, M = 70, K = 5$

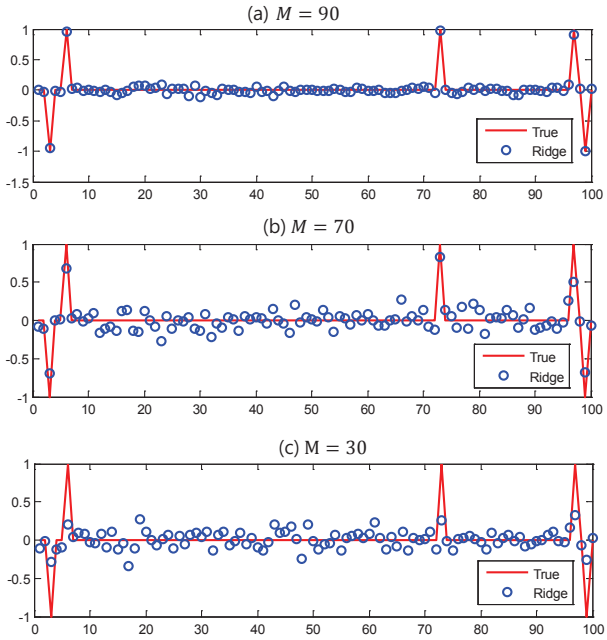


Fig. 2. Sparse estimation via ridge estimator with different M where $N = 100, K = 5, \lambda = 0.001$

regression shrinks together the correlated elements of $\underline{X}$ with respect to $\underline{v}_i$ if the direction of $\underline{v}_i$ has small energy in the column space of $\mathbf{A}$.

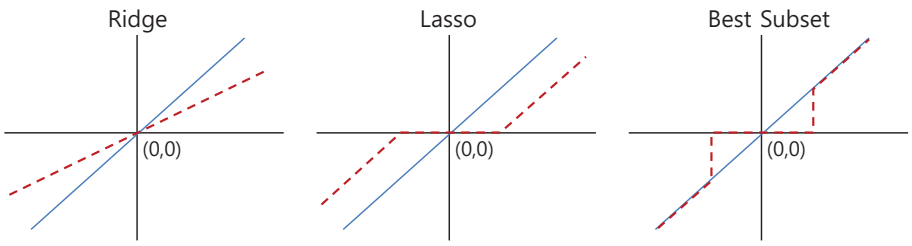


Fig. 3. Shrinkage characteristic of Ridge, Lasso and Best subset selection where the orthonomal matrix $\mathbf{A}$ is assumed. In addition, the blue solid line in the figure is the 45° line to show the LSE solution as a reference (The figure is borrowed from Table 3.4 of [1]).

The ridge penalty can be used to estimate sparse vectors $\underline{X} \in \mathbb{R}^N$ in undetermined systems $\underline{Y} = \mathbf{A}\underline{X} \in \mathbb{R}^M$. In order to apply the expression of (7), we need an augmented matrix $\mathbf{A}' \in \mathbb{R}^{N\times N}$ which additionally includes $N - M$ zero rows from $\mathbf{A} \in \mathbb{R}^{M\times N}$. Let us consider sparse vectors $\underline{X}$ which contains K nonzero signed elements having unit magnitude. The ridge penalty shrinks the elements of $\underline{X}$ with respect to non-principal basis of $\mathbf{A}'$. Hence, the ridge regression enables the K largest elements, which are most related to the principal basis of $\mathbf{A}'$, to have exceptionally large magnitude.

We examine the ridge regression on the parameter $N = 100, K = 5$ with standard Gaussian matrix $a_{ji} \in \mathbf{A} \sim \mathcal{N}(0, 1/M)$. Fig.1 shows that the ridge estimation can finds the K largest elements of $\underline{X}$ with appropriately chosen λ. Another example is shown in Fig.2 where we show the behavior of ridge regression according to the number of M. We note in Fig.2 that the magnitude of the K largest elements of $\widehat{\underline{X}}$ becomes smaller as M decreases. This means that for clear distinction of the K largest elements, the ridge method requires M close to N. In addition, we know from Fig.1 and Fig.2 that the ridge solver cannot exactly fit the nonzero elements of $\underline{X}$.

III. Shrinkage with Lasso Penalty

The main characteristic of lasso is that the elements of $\underline{X}$ are shrunk by imposing a L_1-norm penalty of $\underline{X}$ [3]. Namely, the penalty in the lasso setup takes an absolute sum of $\underline{X}$, *i.e.*, $\|\underline{X}\|_1 = \sum_{i=1}^{N} |X_i|$. Following that, its optimization setup is represented as

$$\begin{aligned}(P_{\text{Lasso}}) : \widehat{\underline{X}}_{\text{Lasso}} &= \arg\min_{\underline{X}} \|\underline{Y} - \mathbf{A}\underline{X}\|_2^2 + \lambda \|\underline{X}\|_1 \\ &= \sum_{j=1}^{M} \left(y_j - \sum_{i=1}^{N} a_{ji} X_i \right)^2 + \lambda \sum_{i=1}^{N} |X_i|, \end{aligned} \tag{10}$$

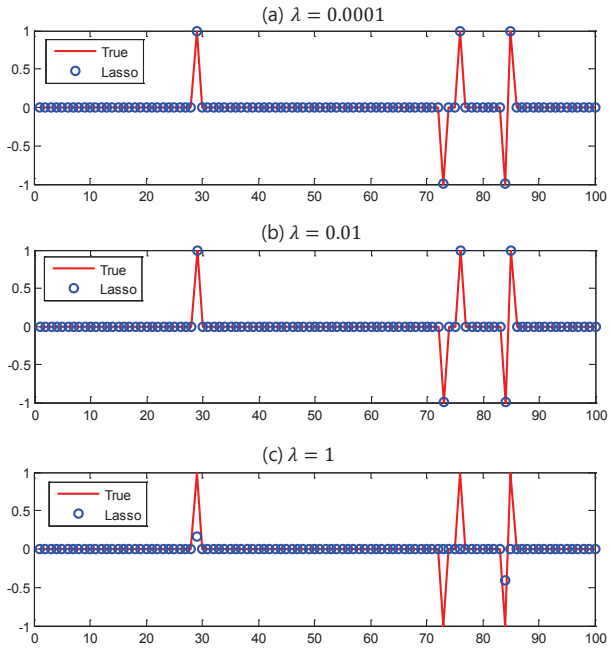


Fig. 4. Sparse estimation via lasso estimator with different λ where $N = 100, M = 70, K = 5$

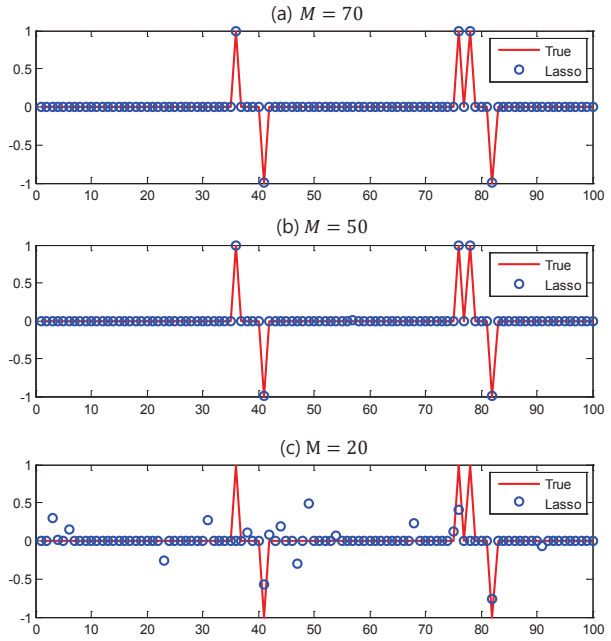


Fig. 5. Sparse estimation via lasso estimator with different M where $N = 100, K = 5, \lambda = 0.001$

where λ is a parameter to control the amount of lasso shrinkage. The larger λ leads to the stronger shrinkage against the elements of $\underline{X}$. When $\lambda = 0$ the solution is equivalent to the LSE solution. This L_1 penalty generates the solutions of (10) nonlinear with respect to the measurement vector $\underline{Y}$; therefore, there is no closed form solver as in ridge regression. The lasso solver can be implemented via a quadratic programming. In addition, the *LARs* algorithm is well known as a computationally efficient algorithm for the lasso solution [4].

To see the detail of the lasso behavior, we investigate the *Karush-Kuhn-Tucker* (KKT) condition with the Lagrangian $\mathcal{L}(\underline{X}, \lambda)$ of the setup in (10).

$$\begin{aligned}
&1) \text{ Stationarity: } \nabla_{\underline{X}} \mathcal{L}(\underline{X}, \lambda) = \mathbf{G}\underline{X} - \mathbf{A}^T\underline{Y} + \lambda \underline{B} = 0, \\
&2) \text{ Dual feasibility: } \lambda \geq 0, \\
&3) \text{ Prime feasibility: } \nabla_{\lambda} \mathcal{L}(\underline{X}, \lambda) = \|\underline{X}\|_1 \leq 0, \\
&4) \text{ Complementary slackness for strong duality: } \lambda \|\underline{X}\|_1 = 0,
\end{aligned} \tag{11}$$

where we define a Gram matrix $\mathbf{G} := \mathbf{A}^T\mathbf{A}$ and

$$\underline{B} := \nabla_{\underline{X}} \|\underline{X}\|_1 = \left[\frac{\partial \sum |X_i|}{\partial X_1}, \frac{\partial \sum |X_i|}{\partial X_2}, ..., \frac{\partial \sum |X_i|}{\partial X_N} \right]. \tag{12}$$

Since $\sum |X_i|$ is not differentiable, we apply the concept of sub-differential to $\frac{\partial \sum |X_i|}{\partial X_1}$. Then, each element of $\underline{B}$ is given by

$$B_i = \frac{\partial \sum |X_i|}{\partial X_1} := \begin{cases} \text{sign}(X_i) & \text{if } |X_i| \geq \lambda \\ B_i \in [-1, 1] & \text{if } |X_i| < \lambda \end{cases}. \tag{13}$$

We note the stationarity condition in (11), which can be rewritten as

$$\mathbf{A}^T\underline{Y} - \lambda \underline{B} = \mathbf{G}\underline{X}. \tag{14}$$

Insight about the lasso shrinkage can be obtained by assuming that the matrix $\mathbf{A}$ is orthonomal, *i.e.*, $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. By applying the orthonomal assumption to (14), we have

$$\underline{\widehat{X}}_{\text{Lasso}} = \mathbf{A}^T\underline{Y} - \lambda \underline{B}. \tag{15}$$

Then, the expression in (15) can be represented by a soft thresholding function with the parameter λ [5], *i.e.*,

$$\underline{\widehat{X}}_{\text{Lasso}} = \eta\left(\mathbf{A}^T\underline{Y}; \lambda\right), \tag{16}$$

where the thresholding function $\eta(\tau;\lambda)$ is elementwisely defined as

$$\eta\left(\tau_i;\lambda\right) = \begin{cases} \tau_i - \lambda & \text{if } \tau \geq \lambda_i, \\ \tau_i + \lambda & \text{if } \tau \leq -\lambda_i, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

We know from (17) that lasso shrinks the elements of $\underline{X}$ according to their magnitude. For the comparison purpose, we also consider ridge estimate with the orthonomal matrix $\mathbf{A}$, given by

$$\underline{\widehat{X}}_{\text{Ridge}} = \frac{1}{1+\lambda}\mathbf{A}^T\underline{Y}. \tag{18}$$

Differently from the lasso case, the ridge estimate is obtained with a proportional shrinkage $\frac{1}{1+\lambda}$ (radically the proportional shrinkage of ridge is determined by singular values of $\mathbf{A}$). We borrow Fig.3 from the reference book (the figure in Table 3.4 of [1]) to depict the shrinkage characteristic of ridge and lasso, compared to the best subset selection which is an optimal estimator to find the $K(\leq M)$ largest elements of $\underline{X} \in \mathbb{R}^N$. Fig.3 explicitly shows the difference among those three estimators.

We examine the lasso solver to estimate the signed K-sparse vectors $\underline{X} \in \mathbb{R}^N$ from the undetermined system $\underline{Y} = \mathbf{A}\underline{X} \in \mathbb{R}^M$, as in the ridge regression. Fig.4 shows that the lasso solver perfectly finds the K largest elements with appropriate λ. We note in Fig.4 that the lasso estimate of the case $\lambda = 1$ does not fit to the true of $\underline{X}$ because in this case, the lasso penalty shrinks the elements too much. Fig.5 shows the lasso recovery of $\underline{X}$ for a variety of the number of measurements M. In the figure, we see that the lasso solver finds an accurate solution when $M = 50, 70$, but fails in the estimation when $M = 20$.

IV. Variants of Lasso

A. Elastic-Net for Approximately Sparse Signal

We can generalize the ridge and the lasso penalty by using the concept of L_p-norm, *i.e.*,

$$(P_{L_p}) : \underline{\widehat{X}}_{L_p} = \arg\min_{\underline{X}} \sum_{j=1}^{M}\left(y_j - \sum_{i=1}^{N} a_{ji}X_i\right)^2 + \lambda\sum_{i=1}^{N}|X_i|^p, \tag{19}$$

for $p \geq 0$, where the case $p = 0$ corresponds to the best subset selection which is non-convex; $p = 2$ corresponds to ridge regression which is convex; $p = 1$ is the lasso case which has the smallest p such that the problem is convex. Value of $p \in (1, 2)$ suggests a compromise between the lasso and ridge regression. If p is closer to 1, the solver has the ability to put small elements close to zero which is the nature of the lasso solver, If p is closer to 2, the solver more tends to shrink signal elements associated with the singular values of $\mathbf{A}$ which is the nature of ridge regression.

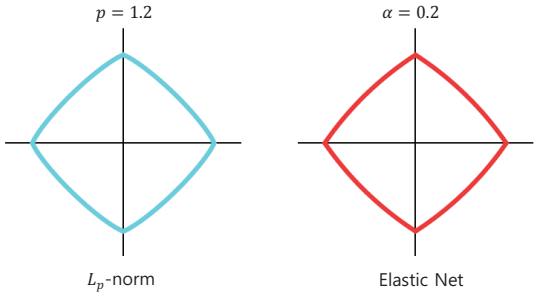


Fig. 6. Contours of the L_p penalty for $p = 1.2$ (left plot) and the elastic-net penalty ($\alpha = 0.2$) (right plot) (The figure is borrowed from Figure 3.13 of [1].)

Elastic-net, proposed by Zou and Hastie, introduced a different compromise between ridge and lasso [7]. The elastic-net selects the largest elements like lasso, and shrinks the remaining small elements like ridge, using a mixture penalty. Therefore, the elastic-net solver is useful for approximately sparse signals whose small elements are originally not exactly zero. The optimization setting of the elastic-net solver is given by

$$\begin{aligned}(\mathrm{P_{EN}}): \ \underline{\widehat{X}}_{EN} = \arg\min_{\underline{X}} \|\underline{Y} - \mathbf{A}\underline{X}\|_2^2 \\ + \lambda \left(\alpha \|\underline{X}\|_2^2 + (1-\alpha)\|\underline{X}\|_1\right),\end{aligned} \tag{20}$$

where α is a mixing rate of the mixture penalty. We borrow Fig.6 from the book (Figure 3.13 of [1]). This figure compares contours of the L_p norm penalty with $p = 1.2$ and the mixture penalty with $\alpha = 0.2$. It is very difficult to distinguish those two penalties by eyes. Although those two are visually very similar, there exists a fundamental difference. The elastic-net has sharp (non-differentiable) corners such that it can put the elements exactly zero, whereas the L_p penalty does not [7]. Likewise with lasso, the elastic-net can be solved via quadratic programming, and the *LARs-EN* algorithm was introduced as a LARs type algorithm to solve the elastic-net problem by Zou and Hastie [7]

We compare the elastic-net solver to the lasso solver in Fig.7 where the problem size is $N = 100, M = 70$. For the comparison, we test an approximately sparse signal generated from *i.i.d* two-state Gaussian mixture density, *i.e.*,

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{N} q\mathcal{N}(x_i; 0, \sigma_{X_1}^2) + (1-q)\mathcal{N}(x_i; 0, \sigma_{X_0}^2), \tag{21}$$

with $q = 0.07, \sigma_{X_1} = 0.05, \sigma_{X_1} = 1$. We set the elastic-net parameters $\alpha = 0.4$, and we use $\lambda = 0.001$ for the lasso and elastic-net both. In Fig.7, the elastic-net with appropriately calibrated parameters α, λ

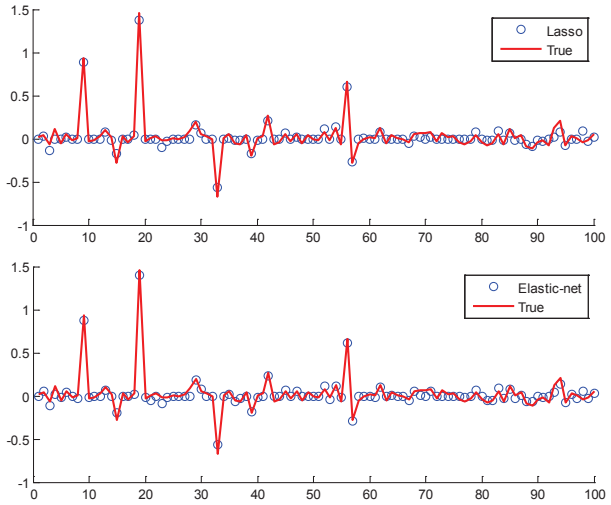


Fig. 7. Approximately sparse signal estimation ($N = 100, M = 70, \lambda = 0.001$) via lasso (Upper plot), and via the elastic-net (bottom plot) where the MSE of lasso estimate is 0.0509, while that of the elastin-net is 0.0359 in this example.

surely improves the estimation accuracy from lasso although it is very hard to be distinguished by eyes. Indeed, the MSE of lasso estimate is 0.0509, while that of the elastin-net is 0.0359 in this example.

B. Fused Lasso for Piecewise Smooth Signals

The use of various types of penalties enables us to solve the $\underline{Y} = \mathbf{A}\underline{X}$ problem adaptively to the characteristic of the signal $\underline{X}$. The *fused lasso* is one of such solvers to find piecewise smooth signals. The fused lasso solves the problem given by

$$\begin{aligned}(\mathrm{P_{FL}}): \ \underline{\widehat{X}}_{FL} = \arg\min_{\underline{X}} &\|\underline{Y} - \mathbf{A}\underline{X}\|_2^2 \\ &+ \lambda\left(\alpha\sum_{i=2}^{N}|X_i - X_{i-1}| + (1-\alpha)\|\underline{X}\|_1\right),\end{aligned} \tag{22}$$

where the difference penalty, $\sum_{i=2}^{N}|X_i - X_{i-1}|$, enforces the estimate $\underline{\widehat{X}}_{FL}$ to be piecewise smooth by considering the order of the features. Namely, the fuse lasso encourages both sparsity of the signal values and sparsity of difference between adjacent elements. Fig.8 shows contour plot of the fused lasso penalty compared to that of the lasso penalty. As shown in Fig.8, the fused lasso penalty has asymmetric contour owing the difference penalty, and it becomes severe as α increases. This asymmetricity of the fused lasso encourages the smoothness of the signal.

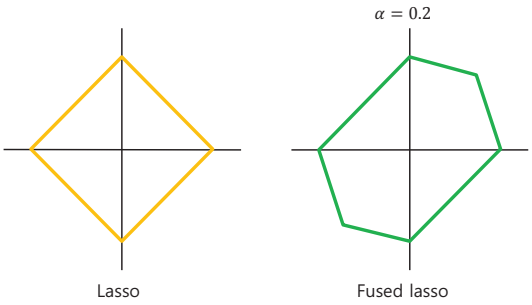


Fig. 8. Contours of the Lasso penalty (left plot) and the fuse lasso penalty ($\alpha = 0.2$) (right plot)

We show an example of the piecewise smooth signal recovery using a fused lasso solver in Fig.9, where measurements $\underline{Y}$ is generated from a piecewise smooth signal $\underline{X}$ with $N = 100, M = 50$ using a standard Gaussian matrix $\mathbf{A}$. This example shows that the piecewise smooth signal can be recovered via the fused lasso as α increases although the signal $\underline{X}$ itself is not sparse. We also checked that the signal can be recovered even from $M = 25$ measurements when $\alpha = 0.9$. In the figure, the case of $\alpha = 0$ is noteworthy because the case is equivalent to the conventional lasso case. This case informs us that such a piecewise smooth recovery is not successful via the normal lasso solver.

V. Conclusive Remarks

We have discussed about shrinkage method to solve the linear system $\underline{Y} = \mathbf{A}\underline{X}$. Estimation through such a method has smaller MSE than LSE at the expense of losing the unbiasedness. Ridge regression is one of the shrinkage methods applying a penalty on the energy of $\underline{X}$. This ridge penalty makes the solver to shrinks together the correlated elements of $\underline{X}$ with respect to the matrix $\mathbf{A}$. Estimation accuracy of ridge is not satisfied for the K sparse signal estimation because the ridge solver cannot exactly fit the nonzero elements of $\underline{X}$. We also have introduced the lasso solver which imposes L_1-norm penalty of $\underline{X}$. The lasso solver shrinks the elements of $\underline{X}$ according to their magnitude, performing the shrinkage as a soft thresholding function. The estimation accuracy of lasso is very good for K sparse signals by putting the nonzero elements of $\underline{X}$ exactly to zero. Elastic-net solver is a compromise of ridge and lasso using a mixture penalty. This solver is useful for approximately sparse signals whose small elements are not exactly zero. The fused lasso solver was devised to find piecewise smooth signals. Imposing of difference penalty, which reflects the order of signal features, enables us to estimate the piecewise smooth signal effectively.

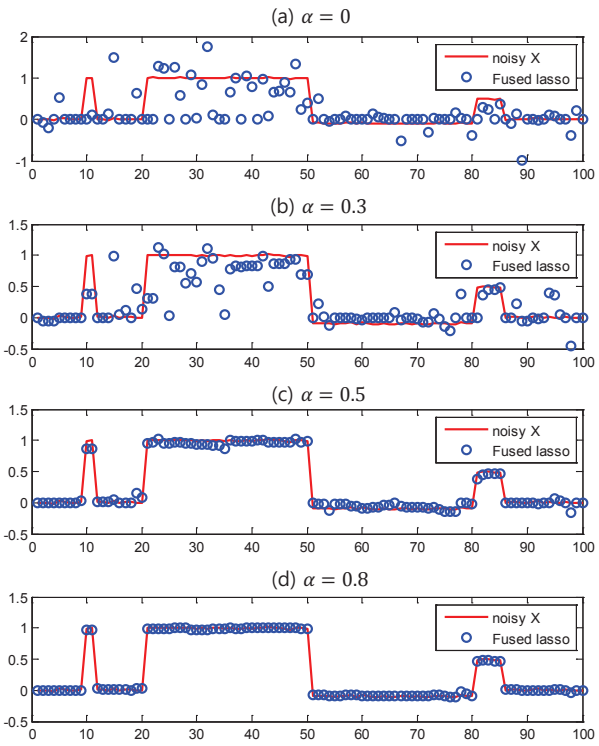


Fig. 9. Piecewise smooth signal estimation via fused lasso for a variety of α when $N = 100, M = 50, \lambda = 0.01$

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning 2nd Edition,* Springer 2008

[2] E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, issue 1, 1970.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statisti. Soc., Ser. B*, vol. 58, no. 1, pp. 267-288, 1996.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407-499, 2004.

[5] D. L. Donoho, "De-nosing by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613-627, May. 1995.

[6] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso, *J. R. Statist. Soc. Ser. B*, vol.67, pp. 91-108, 2005.

[7] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Statist. Soc. Ser. B*, vol. 67 pp. 301-320, 2005.

Efficient Design and Decoding of Polar Codes

Authors: Peter Trifonov

Publication: IEEE T. Comm, Nov 2012
Speaker: Jeong-Min Ryu

Short summary: Polar codes are shown to be instances of both **generalized concatenated codes** and **multilevel codes**. It is shown that the performance of a polar code can be improved by representing it as a multilevel code and applying the multistage decoding algorithm with maximum likelihood decoding of outer codes. Additional performance improvement is obtained by replacing polar outer codes with other ones with better error correction performance. In some cases this also results in complexity reduction. It is shown that Gaussian approximation for density evolution enables one to accurately predict the performance of polar codes and concatenated codes based on them.

I. Introduction

The practical performance of polar codes under the successive cancellation (SC) decoding reported up to now turns out to be worse than that of LDPC and Turbo codes.

This paper demonstrates

1) Polar codes can be efficiently constructed using Gaussian approximation for density evolution.

2) It is shown that polar codes can be treated in the framework of multilevel coding. This enables one to improve the performance of polar codes by considering them as multilevel or, equivalently, generalized concatenated (GCC) ones, and using block-wise near-maximum-likelihood decoding of outer codes. In some cases this results also in reduced decoding complexity.

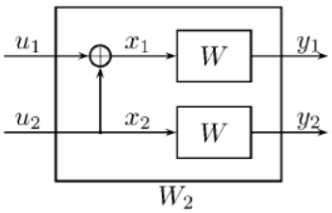
3) A simple algorithm for construction of GCC with inner polar codes.

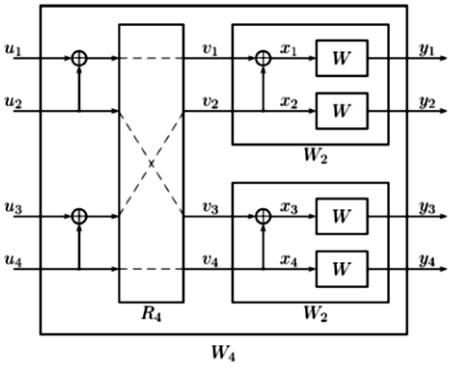
II. BACKGROUND

A. Polar codes

Consider a binary input output symmetric memoryless channel with output probability density function $W(y|x)$, $y \in Y$, $x \in \mathbb{F}_2$. It can be transformed into a vector channel given by $W_n(y_1^n|u_1^n) = W^n(y_1^n|u_1^n G_n)$, where $W^n(y_1^n|x_1^n) = \prod_{i=1}^n W(y_i|x_i)$, $G_n = B_s F^{\otimes s}$, $n = 2^s$, $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, $\otimes s$ denotes s-times Kronecker product of a matrix with itself, and B_s is a $2^s \times 2^s$ bit reversal permutation matrix.

For example)





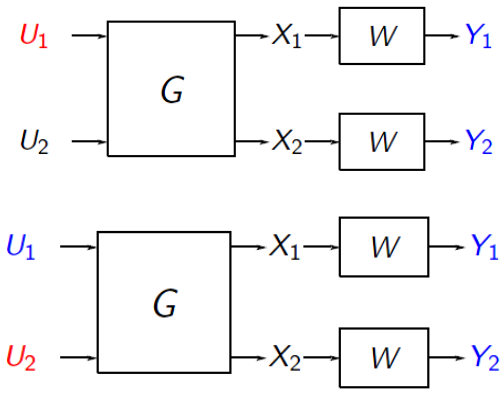
$W_2(y_1^2|u_1^2) = W^2(y_1^2|u_1^2 G_2)$ $\qquad$ $W_4(y_1^4|u_1^4) = W_2(y_1^2|u_1 \oplus u_2, u_3 \oplus u_4)\, W_2^2(y_3^4|u_3, u_4)$

The vector channel can be further decomposed into equivalent subchannels

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \sum_{u_{i+1}^N} \frac{1}{2^{N-1}} W_N(y_1^N|u_1^N).$$

For example)

According to this, we can write $(W, W) \mapsto (W_2^{(1)}, W_2^{(2)})$ for any given B-DMC W.



$$W_2^{(1)}(y_1^2|u_1) \triangleq \sum_{u_2} \frac{1}{2} W_2(y_1^2|u_1^2)$$
$$= \sum_{u_2} \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2)$$

$$W_2^{(2)}(y_1^2, u_1|u_2) \triangleq \frac{1}{2} W_2(y_1^2|u_1^2)$$
$$= \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2)$$

Given y_1^n and estimates u_1^{i-1} of u_1^{i-1}, the SC decoding algorithm attempts to estimate u_i. This can be implemented by computing the following log-likelihood ratios $L_n^{(i)}\left(y_1^n, u_1^{i-1}\right) = \log \frac{W_n^{(i)}\left(y_1^n, u_1^{i-1} \mid u_i = 0\right)}{W_n^{(i)}\left(y_1^n, u_1^{i-1} \mid u_i = 1\right)}$:

$$L_n^{(2i-1)}\left(y_1^n, u_1^{2i-2}\right) = 2\tanh^{-1}\left(\tanh\left(L_{n/2}^{(i)}\left(y_1^{n/2}, u_{1,e}^{2i-2} \oplus u_{1,o}^{2i-2}\right)/2\right)\tanh\left(L_{n/2}^{(i)}\left(y_{n/2+1}^n, u_{1,e}^{2i-2}\right)/2\right)\right), \quad (1)$$

$$L_n^{(2i)}\left(y_1^n, u_1^{2i-1}\right) = L_{n/2}^{(i)}\left(y_{n/2+1}^n, u_{1,e}^{2i-2}\right) + (-1)^{u_{2i-1}} L_{n/2}^{(i)}\left(y_1^{n/2}, u_{1,e}^{2i-2} \oplus u_{1,o}^{2i-2}\right) \quad (2)$$

where $u_{1,e}^{i}, u_{1,o}^{i}$ are subvectors of u_1^i with even and odd indices, respectively, and $L_1^{(i)}\left(y_i\right) = \log \frac{W\left(y_i \mid 0\right)}{W\left(y_i \mid 1\right)}$.

III. Design of Polar Codes Based on Gaussian Approximation

The main drawback of the polar code construction method based on density evolution is its high computational complexity. The most practically important case corresponds to the AWGN channel. In this scenario, $L_1^{(i)}\left(y_i\right) \sim N\left(\frac{2}{\sigma^2}, \frac{4}{\sigma^4}\right)$, provided that the all-zero codeword is transmitted.

The value given by (1)-(2) can be considered as Gaussian random variables with $\boldsymbol{D}\left[L_n^{(i)}\right] = 2\boldsymbol{E}\left[L_n^{(i)}\right]$, where $\boldsymbol{E}$ and $\boldsymbol{D}$ are the mean and variance, respectively. This enable one to compute only the expected value of $L_n^{(i)}$, drastically reducing thus the complexity. In the case of polar codes this approach reduces to

$$\boldsymbol{E}\left[L_n^{(2i-1)}\right] = \phi^{-1}\left(1 - \left(1 - \phi\left(\boldsymbol{E}\left[L_{n/2}^{(i)}\right]\right)\right)^2\right), \quad (3)$$

$$\boldsymbol{E}\left[L_n^{(2i)}\right] = 2\boldsymbol{E}\left[L_{n/2}^{(i)}\right] \quad (4)$$

where

$$\phi(x)=\begin{cases}1-\dfrac{1}{\sqrt{4\pi x}}\displaystyle\int_{-\infty}^{\infty}\tanh\frac{u}{2}e^{-\frac{(u-x)^2}{4x}}dx, & x>0\\ 1 & ,\quad x=0.\end{cases}$$

The error probability for each subchannel is given by

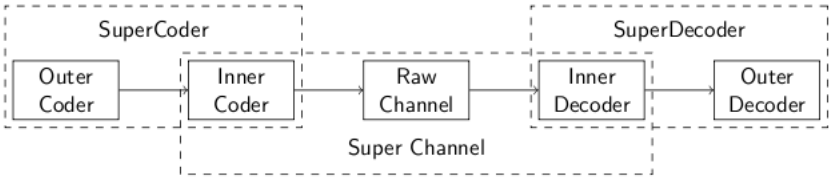
$$\pi_i \approx Q\left(\sqrt{\boldsymbol{E}\left[L_n^{(i)}\right]/2}\right),\ 1\le i\le n.$$

IV. Decomposition of Polar Codes

The overall performance of a polar code is dominated by the performance of the worst subchannel. The proposed approach avoids this problem by performing joint decoding over a number of subchannels.

A. Generalized concatenated polar codes

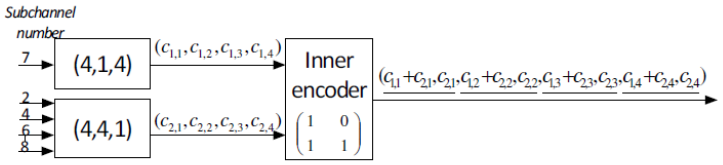
The recursive structure of polar codes enables one to consider them as GCC. Namely, the generator matrix of a polar code can be represented as $G=AF^{\otimes s}=A\left(F^{\otimes(s-l)}\otimes F^{\otimes l}\right)$, where $F=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$ and A is a full-rank matrix with at most one non-zero element in each column.



Inner code encoding: Inner codes $\mathbb{C}_i$ of length $n=2^l$ is generated by rows $i,\ldots,2^l$ of matrix $B_lF^{\otimes l}$.

Outer code encoding: The generator matrix of the $\left(1+R(i,l)\right)$-th outer code C_i is obtained by taking rows $1+R(j,s-l)$ of $F^{\otimes(s-l)}$, such that row $1+R\left(i2^{s-l}+j,s\right)$ of $F^{\otimes s}$ is included into the generator matrix of the original polar code, where $0\le i<2^l$, $0\le j<2^{s-l}$, and

$$R\left(\sum_{j=0}^{m-1}2^j i_j,m\right)=\sum_{j=0}^{m-1}2^j i_{m-1-j},\quad i_j\in\{0,1\}.$$



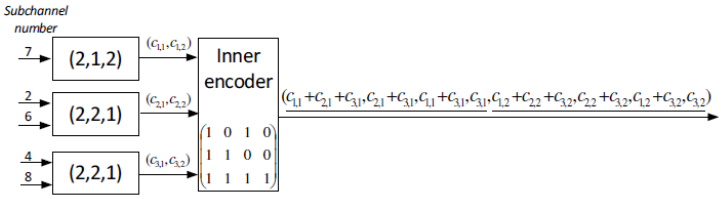
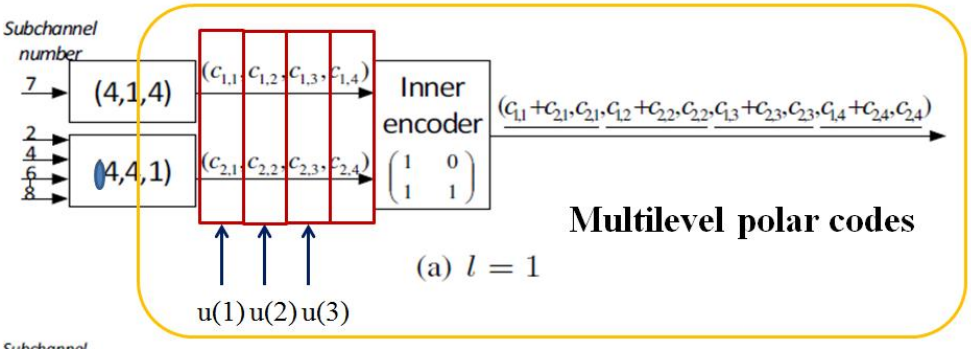


Fig. 3. Representation of $(8, 5, 2)$ polar code as GCC.

B. Multilevel polar codes

In the context of polar codes, signal constellation A is given by 2^n binary n-vectors $a(u)$, which can be obtained as $a(u) = uB_lF^{\otimes l}$, $u \in GF(2)^n$, where $n = 2^l$. This constellation is recursively partitioned into subsets $A(u_1^i)$ by fixing the values of $u_1, ..., u_i$. The elements of u are obtained as codeword symbols of outer codes C_i of length $N = 2^{s-l}$. That is, one can construct N vectors $u^{(j)} = (c_{1,j}, ..., c_{n,j})$, $1 \le j \le N$, where $(c_{i,1}, ..., c_{i,N}) \in C_i$, $1 \le i \le n$ and obtain a multilevel codeword $(u^{(1)}B_lF^{\otimes l}, ..., u^{(N)}B_lF^{\otimes l})$.



The multilevel polar codes can be decoded by multistage decoding algorithm.

V. Concatenated Codes Based on Polar Codes

The performance of a polar code under the multistage decoding with block-wise maximum-likelihood decoding of outer codes can be improved by changing the set of frozen bits. Furthermore, if the algorithm used to perform block-wise decoding of outer codes does not take into account their structure, one can use any linear block code with

suitable parameters, not necessary polar, as C_i. This enables one to employ outer codes with better error correction performance.

A. Capacity rule

The rate R_i of C_i should be chosen equal to the capacity C_i of the *i*-th subchannel of the multilevel code, which is induced by matrix $B_l F^{\otimes l}$. According to [10], one obtains

$$C_i = I\left(y_1^n; u_i \mid u_1^{i-1}\right) = E_{u_1^{i-1}}\left[C\left(A\left(u_1^{i-1}\right)\right)\right] - E_{u_1^i}\left[C\left(A\left(u_1^i\right)\right)\right]$$

where

$$C(B) = \int_{R^n} \sum_{a \in B} \frac{W^n\left(y_1^n \mid a\right)}{|B|} \log_2 \left(\frac{|B| W^n\left(y_1^n \mid a\right)}{\sum_{b \in B} W^n\left(y_1^n \mid b\right)} \right) dy_1^n$$

is the capacity when using the subset B of $\mathbb{F}_2^n$ for transmission over the vector channel $W^n\left(y_1^n \mid x_1^n\right)$. In the case of binary input memoryless output symmetric channels, one can drop the expectation operator to obtain $C_i = C\left(A^{(i-1)}\right) - C\left(A^{(i)}\right)$, where $A^{(i)} = A\underbrace{(0,\ldots,0)}_{i \text{ times}}$.

It can be seen that the latter set is a linear block code C_i generated by $l-i$ last rows of $B_l F^{\otimes l}$. The expression can be further simplified to

$$C\left(A^{(i)}\right) = \int_{R^N} \prod_{j=1}^{N} W\left(y_j \mid 0\right) \log_2 \left(\frac{|C_i| \prod_{j=1}^{N} W\left(y_j \mid 0\right)}{\sum_{b \in B} \prod_{j=1}^{N} W\left(y_j \mid b_j\right)} \right) dy_1^N .$$

Hence, the capacity of the *i*-th subchannel of the multilevel polar code can be computed as

$$C_i = \int_{R^N} \prod_{j=1}^{N} W\left(y_j \mid 0\right) \log_2 \left(\frac{2 \sum_{b \in C_{i+1}} \prod_{j=1}^{N} W\left(y_j \mid b_j\right)}{\sum_{b \in C_i} \prod_{j=1}^{N} W\left(y_j \mid b_j\right)} \right) dy_1^N . \quad (5)$$

Obviously, employing this rule results in a capacity achieving concatenated code, provided that the outer codes can achieve the capacity too. However, evaluating (5) seems to be a difficult task.

B. Equal error probability rule

The probability of incorrect decoding of a binary linear block code C can be obtained as

$$p_e \le \sum_{j=d}^{N} A_j Q\left(\sqrt{\frac{\boldsymbol{E}\left[L_i\right]}{2} j}\right)$$

where A_i are weight spectrum coefficients of code C, and d is its minimum distance. Since it is in general difficult to obtain code weight spectrum, and union bound is known to be not tight in the low-SNR region, one can use simulations to obtain a performance curve for the case of AWGN channel and some fixed (probably, non-ML) decoding algorithm, and use least squares fitting to find suitable α and δ, so that the decoding error probability is given by

$$p_e(m) \approx \alpha Q\left(\sqrt{\frac{\boldsymbol{E}\left[L_i\right]}{2} \delta}\right).$$

Assume now that the outer codes C_i are selected from some family of error-correcting codes (not necessary polar) of length N. Let K_t, D_t and $P_t(m)$ be the dimension, minimum distance and decoding error probability function for the t-th code, respectively, where m is the expected value of LLR.

Figure 4 presents a simple algorithm for construction of a generalized concatenated (multilevel) code of rate R according to the equal error probability rule. The algorithm employs the bisection method to approximately solve the equation $\sum_{i=1}^{2^l} K(i,P) = RN2^l$, where $K(i,P)$ is the maximum dimension of a code capable of achieving error probability P at the i-th subchannel. The parameter ε is a sufficiently small constant, which affects the precision of the obtained estimate for P. The code is optimized for the case of AWGN channel with noise variance σ^2. The algorithm returns the dimensions of optimal codes for each level, as well as an estimate for the decoding error probability for each code.

CODEOPTIMIZATION(σ, R, N, l)

1 $\mathbf{E}[L_1^{(1)}] \leftarrow 2/\sigma^2$
2 Compute $m_i = \mathbf{E}[L_{2^l}^{(i)}], 1 \le i \le 2^l$ via (5)–(6)
3 $P' \leftarrow 1; P'' \leftarrow 0$
4 **while** $P' - P'' > \epsilon P'$
5 **do** $\tilde{P} \leftarrow (P' + P'')/2$
6 $t_i \leftarrow \arg\max_{t:P_t(m_i) \le \tilde{P}} K_t, 1 \le i \le 2^l$
7 $K \leftarrow \sum_{i=1}^{2^l} K_{t_i}$
8 **if** $K < RN2^l$
9 **then** $P'' \leftarrow \tilde{P}$
10 **else** $P' = \tilde{P}$
11 **return** $(K_{t_1}, \ldots, K_{t_{2^l}}), \tilde{P}$

The SC/multistage decoder produces an error if decoding of any of the component codes is incorrect. Therefore, the overall error probability of the GCC can be computed as

$$\begin{aligned} P &= 1 - P\{C_1, \ldots, C_n\} \\ &= 1 - P\{C_1\} P\{C_2 \mid C_1\} \cdots P\{C_n \mid C_1, \ldots, C_{n-1}\} \\ &\approx 1 - \prod_{i=1}^{n} \left(1 - P_{t_i}(m_i)\right) \approx 1 - \left(1 - P\right)^n \end{aligned}$$

where C_i denotes the event of correct decoding of the outer code at the i-th level, P is the quantity computed by the above algorithm, and t_i is the index of the code selected for the i-th subchannel. This expression enables semi-analytic prediction of the performance of the concatenated code, based on the available performance results for component outer codes.

C. Decoding complexity

One can use any suitable algorithm to implement soft-decision decoding of outer codes in the GCC obtained either by decomposing a polar code, or constructed explicitly using the algorithm in Figure 4. Box-and-match algorithm is one of the most efficient methods to perform near maximum likelihood decoding of short linear block codes [20]. Its worstcase complexity for the case of (N, K) code with order t reprocessing is given by $O\left((N-K)K^t\right) = O\left(N^{t+1}\right)$, although in practice it turns out to be much more efficient. Decoding of a concatenated code of length $v = Nn$ involves decoding of N inner codes using the SC algorithm, and decoding of n outer codes. Therefore the overall complexity is given by $O\left(N^{t+1} n C_b + Nn \log n C_s\right)$, where C_b and C_s are some factors which reflect the cost of elementary operations performed by these algorithms.

VI. Numerical Results

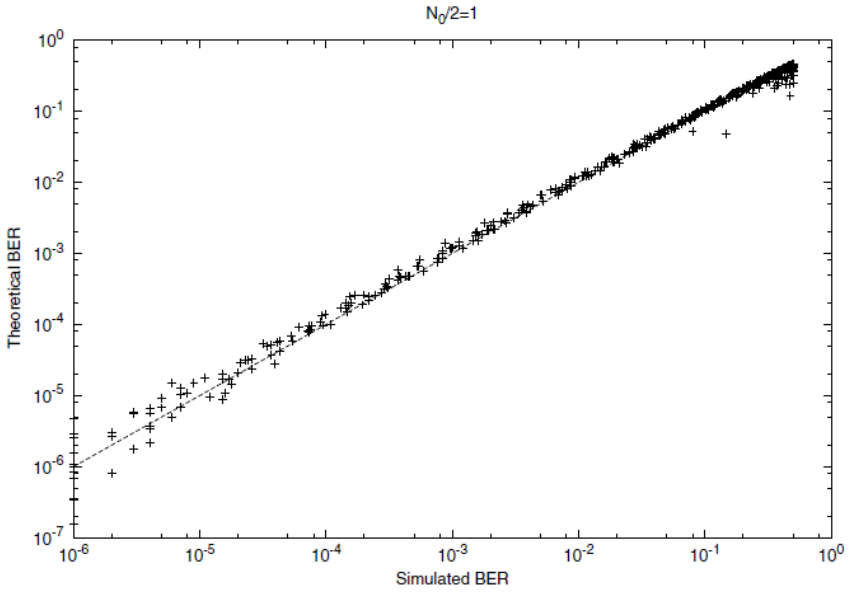


Fig. 5. Accuracy of Gaussian approximation.

Figure 5 presents simulation results illustrating the accuracy of bit error rate analysis based on the Gaussian approximation.

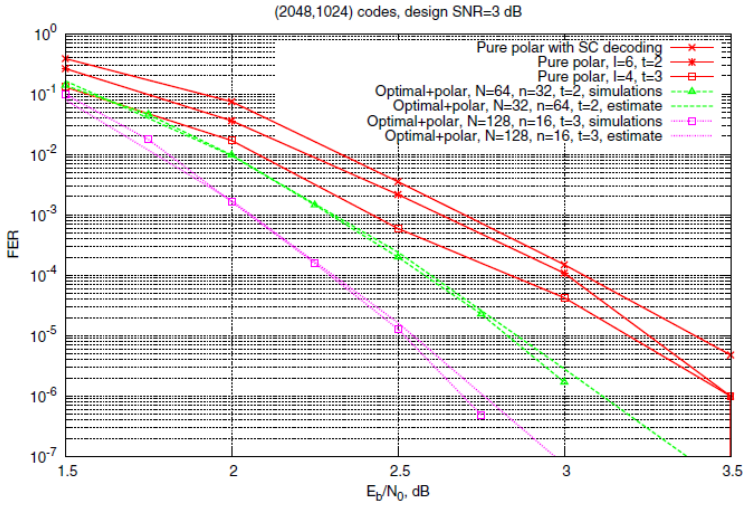


Fig. 6. Performance of polar and concatenated codes.

Figure 6 presents the performance of polar codes of length 2048 designed using the Gaussian approximation method for the case of AWGN channel with Eb/N0 = 3 dB. For multistage decoding, degree *l* decomposition of the original polar code was performed, and box-and-match algorithm with order *t* reprocessing was used for decoding of outer polar codes.

It can be seen that block-wise decoding of outer codes provides up to 0.25 dB performance gain compared to SC decoding. Higher values of N do not provide any noticeable performance improvement. The figure presents also the performance of GCC based on inner polar codes and outer optimal linear block codes with multistage decoding. It can be seen that increasing the length of outer codes provides additional 0.5 dB performance gain. This is due to much higher minimum distance of optimal codes compared to polar codes of the same length, obtained by decomposing the polar code of length Nn.

CS Journal Club, Apr. 18, 2013

Aliasing-Free Wideband Beamforming Using Sparse Signal Representation

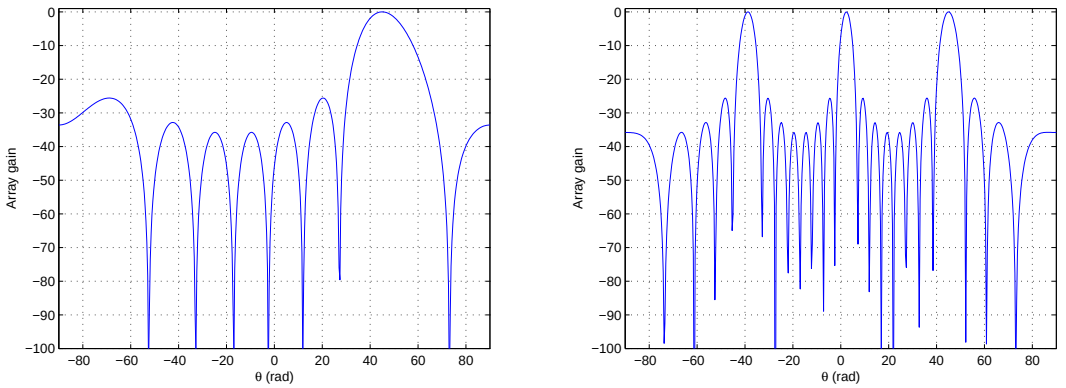
Authors: Zijian Tang, Gerrit Balcquiere, and Geert Leus
Journal: IEEE Trans. on. Sign. Proc. July, 2011.
Presenter: J. Oliver

Abstract

This paper considers the use of sparse signal representation for the wideband direction of arrival (DOA) or angle of arrival estimation problem. In particular, this paper discusses about the two ambiguities, namely, spatial and algebraic aliasing that arise in wideband-DOA. The authors of the paper suggest procedures to avoid the aliasing using multiple measurement vector and multiple dictionaries.

Introduction and Background

- A beamformer is a processor used in conjunction with an array of sensors to provide spatial filtering. The sensor array collects spatial samples of propagating wave fields, which can be processed by the beamformer.
- The objective of a beamformer is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals. A beamformer thus performs spatial filtering to separate signals that have overlapping frequency content but originate from different spatial locations.
- Estimating the spatial locations (or directions) is a well-known problem in array signal processing.
- Three major DOA estimation techniques are 1. Classical methods (Delay-sum beamformer, MVDR) 2. Subspace methods (MUSIC, ESPRIT) 3. ML-based methods
- This paper discusses about beamforming and in particular wide-band beamforming.
- DOA estimation by beamforming can be subjected to ambiguity called spatial aliasing [1].
- Spatial aliasing occurs when the spacing, d, between the sensors is larger than half of the apparent wavelength, that is, $d > \lambda/2$ (See Fig. 1)
- We note from the figures, the resolution increases as d increases, but spatial aliasing also increases.
- This paper discusses how to avoid spatial aliasing (if there is any) in a wideband setting.
- Various authors [2-5] have studied sparse representation (SSR) for narrowband DOA estimation in various contexts. In [2], CS is applied to reduce the ADC sampling rate, in [3,4] it is used to improve angle resolution, in [5] it is used to reduce hardware complexity. All these works assume spatial aliasing is not present.
- However, in SSR based methods aliasing (or ambiguity) comes not only from spatial aliasing, but also from the over-completeness of the dictionary (algebraic aliasing).
- This paper discusses, how to avoid both spatial and algebraic aliasing.



(a) Array tuned to 45 deg with $d = 0.5\lambda$ (b) Array tuned to 45 deg with $d = 1.5\lambda$

Figure 1: Illustration of spatial aliasing

- In summary, the spatial aliasing can be avoided by using multiple dictionaries and the robustness to algebraic alaising can be obtained by using multiple measurement vectors.

Data model

- A uniform linear array (ULA) comprised of N channels indexed by which are equally spaced on a line with spacing d. It receives signals radiated from Q point sources.
- The signal at each channel after time-sampling is partitioned into P segments, where for each segment, K frequency subbands are computed by e.g., a filter bank or the discrete Fourier transform (DFT).
- Let $S_{q,k}(p)$ denote the kth subband (frequency) coefficient computed for the pth segment of the signal that is radiated from the Qth target; similarly, let $y_{n,k}(p)$ denote the kth subband (frequency) coefficient for the pth segment of the signal received at the nth channel.
- With narrow-band assumption, the received signal at the nth sensor at the kth DFT bin is given [1] by

$$y_{n,k}(p) = \sum_{q=0}^{Q-1} e^{j2\pi f_k \frac{d}{c} n \sin\theta_{m_q}} \; S_{q,k}(p) \tag{1}$$

- The aim of this paper is to estimate the target DOAs $\{\theta_0, \theta_1, \cdots, \theta_{Q-1}\}$
- The matrix-vector form of Eqn. (1) is

$$\boldsymbol{y}_{k,p} = \sum_{q=0}^{Q-1} \boldsymbol{a}_{k,m_q} \; S_{q,k}(p) = A_k \boldsymbol{S}_{k,p} \tag{2}$$

where $\boldsymbol{a}_{k,m_q} = \left[1, e^{j2\pi f_k \frac{d}{c} 1 \sin\theta_{m_q}}, \cdots, e^{j2\pi f_k \frac{d}{c}(N-1)\sin\theta_{m_q}}\right]^T$ is called array response vector and A_k steering matrix.

Assumption 1: The array response vectors corresponding to different targets are mutually independent.

Classical beamforming

- Classical beamforming (in this paper, delay-sum beamformer) sets the beamformer coefficients corresponding to a single target angle.
- For example, if the beamformer wants to listen to angle $\theta = 30\,\text{deg}$, then it sets its coefficient vector as $\left[1, e^{j2\pi f_k \frac{d}{c} 1 \sin\frac{\pi}{6}}, \cdots, e^{j2\pi f_k \frac{d}{c}(N-1)\sin\frac{\pi}{6}}\right]^T$ and forms the product $\boldsymbol{a}_{k,m}^H \boldsymbol{y}_{k,p}$
- The angle domain is divided into M points $\Theta = \{\theta_0, \cdots, \theta_{M-1}\}$
- In many applications, such as sonar, a range (time)-bearing(angle) image is desired which can be made by repeating the above procedure for all the subbands; the outputs are then combined and transformed back into the time domain by means of e.g., an inverse Fourier transform.
- In the end, the signal for the pth segment at the mth angle in the range-bearing image $I(p,m)$ can be computed as

$$I(p,m) = \left| \frac{1}{N} \sum_{k=0}^{K-1} \boldsymbol{a}_{k,m}^H \boldsymbol{y}_{q,k}(p) e^{j2\pi f_k p} \right|^2 \tag{3}$$

- $I(p,m)$ can be interpreted as the power of the output of a spatial-temporal filter steered to the direction θ_m. The DOAs are estimated by seeking those θ_m whose corresponding values in $\sum_{p=0}^{P-1} I(p,m)$ are the largest.
- The delay-sum beamformer is subject to spatial aliasing. That is, when the spacing d is larger than apparent wavelength, it is possible to find another $\theta_{m'} \neq \theta_m$ such that for an arbitrary integer j

$$f_k \frac{d}{c} \sin\theta_m = f_k \frac{d}{c} \sin\theta_{m'} + j \tag{4}$$

holds and thus $\boldsymbol{a}_{k,m} = \boldsymbol{a}_{k,m'}$, which gives multiple peaks in the range-bearing image $I(p,m)$.

DOA estimation via SSR

Problem formulation

- Divide the whole angle search range into a fine grid $\Theta = \{\theta_0, \theta_1, \cdots, \theta_{M-1}\}$.
- Each θ_m corresponds to a certain array response vector $\boldsymbol{a}_{k,m}$, which depends on f_k.
- Construct $N \times M$ steering matrix $A_k = [\boldsymbol{a}_{k,0}, \cdots, \boldsymbol{a}_{k,M-1}]$ (dictionary)
- Assumption 2: The DOAs of the targets $\{\theta_{m_0}, \theta_{m_1}, \cdots, \theta_{m_{Q-1}}\} \in \Theta \quad \Omega = \{m_0, m_1, \cdots, m_{Q-1}\}$
- Data model : $\boldsymbol{y}_{k,p} = A_k \boldsymbol{x}_{k,p}$ $\boldsymbol{x}_{k,p}$ is a Q-sparse signal ($Q < N$)
- We have P such snapshots (measurement vector), then we can form $Y_k = A_k X_k$ Y_k is $N \times P$, A_k is $N \times M$ and X_k is $M \times P$
- Assume that the DOAs during the span of P snapshots remain unchanged, then the columns of X_k share a common sparsity.
- Let $\mathcal{R}(A)$ denote an operation that collects the indexes of all the nonzero rows of a matrix A. $\mathcal{R}(X_k) = \Omega$ and $|\mathcal{R}(X_k)| = Q$.
- With these notations, we can formulate the sparse recovery problem as

$$\min_{\widehat{X}_k} \; |\mathcal{R}(\widehat{X}_k)| \quad \textit{subject to} \quad Y_k = A_k X_k \tag{5}$$

Aliasing Suppression

- As mentioned earlier, spatial aliasing occurs if d is larger than half of the apparent wavelength, which leads to similar columns in the steering matrix
- In classical beamforming, the DOAs are sought by steering a beamformer to different potential angles.
- However, the SSR-based method recovers X_k first and then estimates the DOAs by locating the rows of X_k that contain dominant entries.
- The over-completeness of the SSR dictionary gives rise to non-unique solutions and thus ambiguity in DOA estimation, which is termed as algebraic aliasing.
- Algebraic aliasing is essentially related to the "goodness" of the sensing matrices (steering matrix) for the DOA recovery.

Proposition 1: Under Assumption 1, if the number of targets Q and channels N satisfy

$$N > 2Q - rank(Y_k) \tag{6}$$

then the SSR-based method will not suffer from algebraic aliasing.

Proof: Algebraic aliasing will not exist if we can find unique solution $\widehat{X}_k$ satisfying $Y_k = A_k\widehat{X}_k$. This is only possible if the Kruskal-rank of A_k is larger than $2Q - rank(Y_k)$ [6, Theorem 2.4]. Since A_k is a Vandermonde matrix, whose Kruskal-rank is equal to its rank, N.

Kruskal-rank (or k-rank) of a matrix A is defined as the largest integer r for which every set of r columns of A is linearly independent.

Remarks:

- If $rank(Y_k) = 1$ $(P = 1)$, then $Q < N/2$, that is we can discriminate at most $N/2$ targets.
- On the other side, $rank(Y_k) \leq rank(X_k) \leq Q$, suggests that $Q < N$.
- Thus, using multiple measurement vectors the authors argue that it is possible to counter the algebraic aliasing.

So far, we have concentrated on the data model for a single frequency f_m. We can obtain different measurements and different dictionaries $A_k \neq A_l$ if we use different frequency $f_k \neq f_l$. We will next show that using multiple dictionaries enables us to eliminate spatial aliasing.

- Let Γ_k denote the support of all possible DOA solutions for the k-th dictionary
  $\Gamma_k = \{\mathcal{R}(\widehat{X}_k^{(0)}), \mathcal{R}(\widehat{X}_k^{(1)}), \cdots\}$
- Spatial aliasing is frequency-dependent, which means that for different center frequencies, the resulting ambiguity will not (completely) overlap. Therefore, we can imagine that if we solve Eqn. (5) for several frequencies: $f_0, f_1, \cdots, f_{K-1}$ and combine the solutions in a judicious way, the ambiguity due to spatial aliasing will at least be reduced.

Theorem 1: With Proposition 1 met, if there exist at least two dictionaries, whose corresponding frequencies, say f_k and f_l , satisfy

$$0 < |f_k - f_l| < \frac{c}{2d} \tag{7}$$

then the intersection of the solution support related to different dictionaries will contain exclusively the target DOAs, i.e.,

$$\bigcap_k \Gamma_k = \Omega \tag{8}$$

Proof: With proposition 1 satisfied, we can exclude the ambiguity due to algebraic aliasing and need to focus only on spatial aliasing. Let us proceed with a counter-example. Suppose θ_m is one of the target angles and $\theta_m \neq \theta_{m'}$ is spatial aliasing contained in both dictionaries corresponding to f_k and f_l , which implies that $\{\theta_m, \theta_{m'}\}$ belongs to both Γ_k and Γ_l. In accordance with Eqn. (4) we then have

$$f_k \frac{d}{c} \sin\theta_m - f_k \frac{d}{c} \sin\theta_{m'} = j_1$$

$$f_l \frac{d}{c} \sin\theta_m - f_l \frac{d}{c} \sin\theta_{m'} = j_2$$

$$f_k \frac{d}{c} \sin\theta_m - f_k \frac{d}{c} \sin\theta_{m'} - f_l \frac{d}{c} \sin\theta_m + f_l \frac{d}{c} \sin\theta_{m'} = j_1 - j_2 = j_3$$

where j_1, j_2, j_3 are integers and j_1, j_2 are not equal to 0. Using trignometric identities, the above equations can be written as

$$-2 f_k \frac{d}{c} \sin\frac{\theta_m - \theta_{m'}}{2} \cos\frac{\theta_m + \theta_{m'}}{2} = j_1 \tag{9}$$

$$-2 f_l \frac{d}{c} \sin\frac{\theta_m - \theta_{m'}}{2} \cos\frac{\theta_m + \theta_{m'}}{2} = j_2 \tag{10}$$

$$-2 (f_k - f_l) \frac{d}{c} \sin\frac{\theta_m - \theta_{m'}}{2} \cos\frac{\theta_m + \theta_{m'}}{2} = j3 \tag{11}$$

Since $0 < |f_k - f_l| < \frac{c}{2d}$, it is only possible for Eqn. (11) to hold if the integer $j_3 = j_1 - j_2 = 0$. On the other hand, from the above equations we know that j_1 and j_2 are not zero and they cannot be equal. Therefore, the angle $\theta_{m'}$ cannot be contained simultaneously in Γ_k and Γ_l, which concludes the proof. A judicious choice of frequencies can not only prevent spatial aliasing, but also enhance the performance in a noisy environment.

Aliasing-Free SSR Recovery

Based on the analysis in the previous section, the authors formulate the following multi-dictionary (MD) joint optimization problem with the joint-sparsity constraint:

$$\min_{\widehat{X}_k} \ |\mathcal{R}(\widehat{X}_k)| \quad for\ k = 0, 1, \cdots, K-1$$

$$subject\ to \quad Y_k = A_k X_k, \quad and \ \ \mathcal{R}(\widehat{X}_k) = \mathcal{R}(\widehat{X}_l)\ for\ k \neq l \tag{12}$$

whose solution will be free from any ambiguity under Theorem 1.

The authors have not proposed any new algorithm. They have used OMP in their simulations.

Numerical Examples

1. The authors demonstrate their approach using synthetic and real data.
2. For both cases, they considered ULA with $N = 16$ hydrophones, with a spacing of $d = 0.06\ m$. The speed of the signal wave is assumed to be $c = 1500\ m/s$
3. In the synthetic data they consider two sinusoids $Q = 2$ with frequencies $f_0 = 25$ kHz and $f_1 = 35$ kHz. The DOA are $\{35°, 39°\}$. The search gird is defined as $\Theta = \{-90°, -89.75°, \cdots, 90°\}$. With P=100 snapshots, each dictionary has a dimension of 16×720

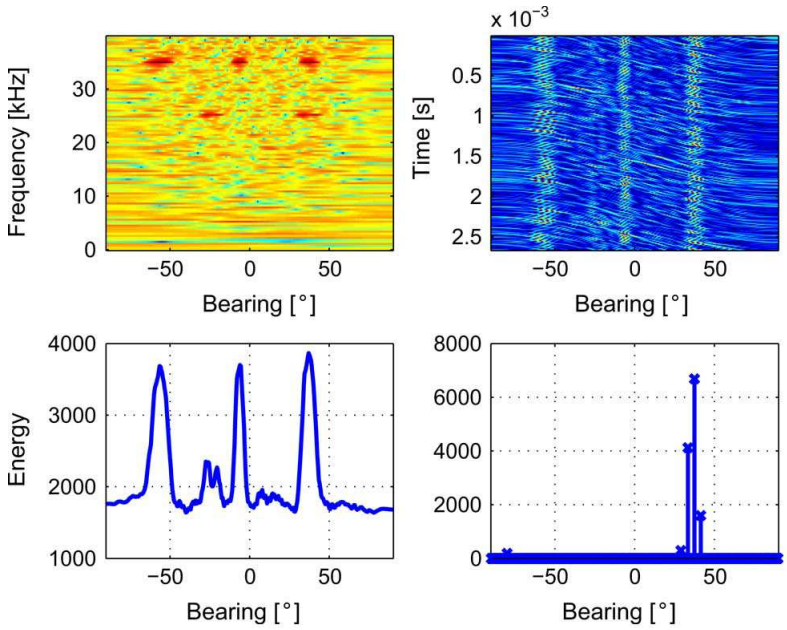


Figure 2: Comparison of classical beamforming with the proposed method. Upper-left subplot: the frequency-bearing image after beamforming; upper-right subplot: the time-bearing image after beamforming; lower-left subplot: the integrated energy of the time-bearing image; lower-right subplot: the result yielded by the proposed method.

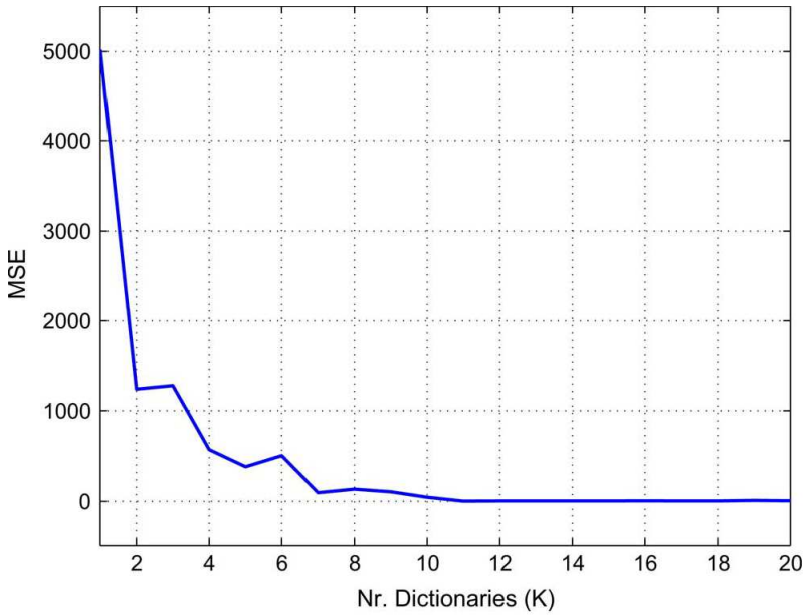


Figure 3: MSE performance against the number of utilized frequencies

4. They halt OMP after 5 iterations. MSE $= \frac{1}{N_s} \sum_{q=0}^{Q-1} (\widehat{\theta}_m - \theta_m)$

1. In the real data experiment, the direction of the divers has to be estimated based on their exhaling sound. Two divers who are 150 m away from the hydrophone are considered. The received signals are from 52° and 60°, respectively.

2. We can see that the frequencies lower than 10 kHz are completely useless for DOA estimation: the diver signal is subdued by the ambient noise dominated by the ship traffic in the harbor.

3. In the midfrequency range (between 10 and 12.5kHz), where the hydrophone array is not subject to aliasing, there is a strong interference signal at a direction around −40°, which possibly comes from a

departing ship blowing the horn.

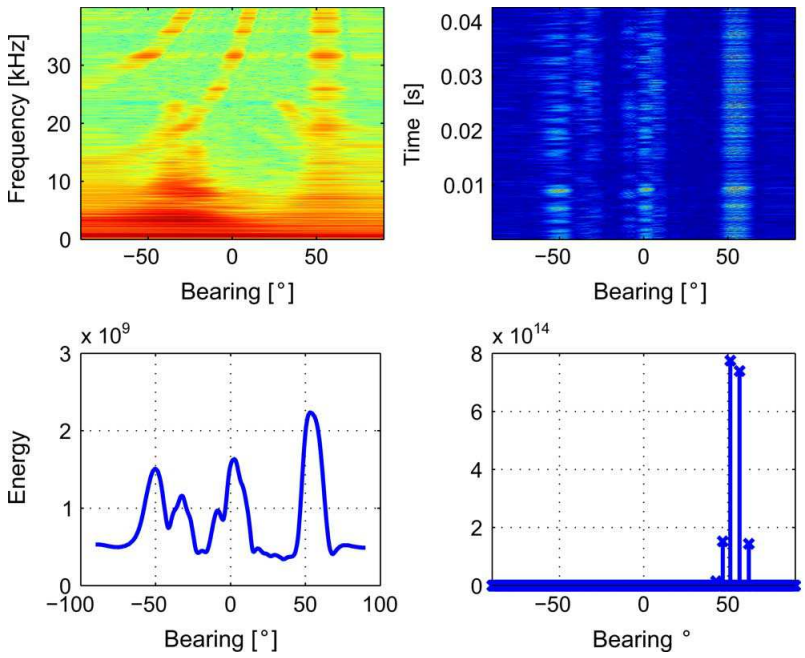


Figure 4: Comparison of classical beamforming with the proposed method for the diver signal. Upper-left subplot: the frequency-bearing image after beamforming; upper-right subplot: the time-bearing image after beamforming (only signals above 25 kHz are taken); lower-left subplot: the integrated energy of the time-bearing image; lower-right subplot: the result yielded by the proposed method.

In Summary, the authors have applied sparse signal reconstruction for DOA estimation (for ULA). They formed an MD optimization problem with joint sparsity constraints. They show how to avoid ambiguities (spatial and algebraic) by using multiple dictionaries and multiple measurement vectors, respectively. They have demonstrated their findings through synthetic and real-life examples.

References

[1] H. L. van Trees, Optimum Array Processing (Detection, Estimation and Modulation Theory, Part IV). New York: Wiley, 2002.

[2] V. Cevher, A. C. Grbz, J. H. McClellan, and R. Chellappa, Compressive wireless arrays for bearing estimation of sparse sources in angle domain, in Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Mar. 2008, pp. 24972500.

[3] J. J. Fuchs, On the application of the global matched filter to DOA estimation with uniform circular arrays, IEEE Trans. Signal Process., vol. 49, no. 4, pp. 702709, Apr. 2001.

[4] D. Malioutov, M. Cetin, and A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process., vol. 53, no. 8, pp. 30103022, Aug. 2005.

[5] Y. Wang, G. Leus, and A. Pandharipande, Direction estimation using compressive sampling array processing, presented at the IEEE Workshop on Statistical Signal Processing, Cardiff, U.K., 2009.

[6] J. Chen and X. Huo, Theoretical results on sparse representations of multiple-measurement vectors, IEEE Trans. Signal Process., vol. 54, no. 12, pp. 46344643, Dec. 2006.

[7] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, Distributed compressive sensing, 2009. Available: http://arxiv.org/abs/0901.3403

Turbo Reconstruction of Structured Sparse Signals

Authors: Philip Schniter

Publication: Information Sciences and System (CISS), 2010

Speaker: Hyeong-ho Baek

Short summary**: This paper considers the reconstruction of structured-sparse signals from noisy linear observations. In particular, the support of the signal coefficients is parameterized by hidden binary pattern, and a structured probabilistic prior (e.g., Markov random chain/field/tree) is assumed on the pattern. Exact inference is discussed and an approximate inference scheme, based on loopy belief propagation (BP), is proposed. The proposed scheme iterates between exploitation of the observation-structure and exploitation of the pattern-structure, and is closely related to noncoherent turbo equalization, as used in digital communication receivers. An algorithm that exploits the observation structure is then detailed based on approximate message passing ideas.**

I. INTRODUCTION

The main objective is to estimate the sparse signal $\mathbf{x} \in \mathbb{C}^N$ from the noisy linear measurements $\mathbf{y} \in \mathbb{C}^M$,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is a known matrix and $\mathbf{w} \in \mathbb{C}^M$ is additive noise, often modeled as circular white Gaussian, i.e., $\mathbf{w} \sim CN(0, \sigma^2 \mathbf{I})$. By "sparse," we mean that the signal has only a few (say K, where $K << N$) non-zero coefficients.

In many cases of interest, the system of equations in (1) is underdetermined, i.e., $M << N$, so that, even in the noiseless case, there is no unique inverse. However, when $\mathbf{x}$ is known to be sparse, it is possible to accurately reconstruct $\mathbf{x}$ from $\mathbf{y}$ if the columns of A are sufficiently incoherent. For various sparse reconstruction algorithms, including convex- optimization-based,

greedy, and iterative thresholding algorithms, there exist elegant bounds on reconstruction error that hold when $\mathbf{A}$ satisfies a certain restricted isometry property (RIP). In many applications, however, the signal $\mathbf{x}$ has structure beyond simple sparsity. For example, the wavelet transform coefficients of natural scenes are not only approximately sparse, but also exhibit persistence across scales, which manifests as correlation within the sparsity pattern. Many other forms of structure in the sparsity pattern are also possible, and so we desire a powerful and flexible approach to modeling and exploiting such structure.

In this paper, we take a probabilistic approach to modeling sparsity structure, allowing the use of, e.g., Markov chain (MC), Markov random field (MRF), and Markov tree (MT) models [2]. Such models have been previously exploited for sparse reconstruction, but only to a limited extent. For example, [3] and [4] proposed Monte-Carlo-based [5] sparse reconstruction algorithms using MRF and MT models, respectively, and [6] and [7] proposed to iterate matching- pursuit with MAP pattern detection based on MRF and MT models, respectively. Monte-Carlo algorithms, while flexible, are typically regarded as computationally too expensive for many problems of interest. Matching-pursuit algorithms are typically much faster, but the schemes in [6], [7] are ad hoc. We attack the problem of reconstructing structured-sparse signals through the framework of belief propagation (BP) [8]. While BP has been successfully used to recover unstructured sparse signals (e.g., [9], [10]), we believe that its application to structured sparse signals is novel. As we shall see, the BP framework suggests an iterative approach, where sparsity pattern beliefs are exchanged between two blocks, one exploiting observation structure and the other exploiting pattern structure. In this regard, our scheme resembles turbo equalization from digital communications [11], where bit beliefs are exchanged between a soft equalizer and a soft decoder. Our two blocks are themselves naturally implemented using BP, and we detail a particularly efficient algorithm based on the approximate message passing (AMP) framework recently proposed by Donoho, Maleki, and Montanari [10].

II. Signal Model

Our structured-sparse signal model uses hidden binary indicators $\{s_n\}_{n=1}^N$, where $s_n \in \{0,1\}$. In particular, $s_n = 1$ indicates that the signal coefficient x_n is active while $s_n = 0$ indicates that

x_n is inactive. Assuming that the active signal coefficients are independently but non-identically distributed, we can write

$$p(x_n \mid s_n) = s_n q_n(x_n) + (1 - s_n)\delta(x_n) \tag{2},$$

Where $q_n(\cdot)$ denotes the pdf of x_n, when active, and $\delta(\cdot)$ denotes the Dirac delta. We refer to $\mathbf{s} = [s_1, s_2, \cdots, s_N]^T \in \{0,1\}^N$ as the sparsity pattern, and model structure in $\mathbf{s}$ through an assumed prior pmf $p(\mathbf{s})$.

III. Turbo Inference

Our primary goal is estimating the structured-sparse signal $\mathbf{x}$ given the observations $\mathbf{y} = \mathbf{y}_0$ in model (1). In particular, we are interested in computing minimum mean-squared error(MMSE) estimates of $\{x_n\}$.

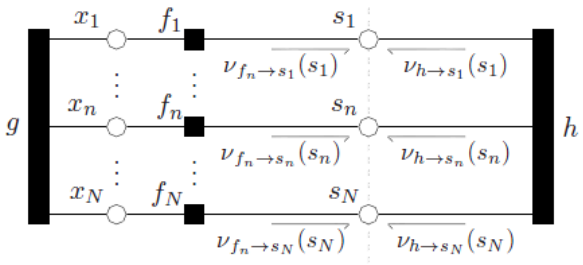


Figure 1 Factor graph of posterior $p(\mathbf{x}, \mathbf{s} \mid \mathbf{y} = \mathbf{y}_0)$. The boxes represent factor nodes and the circles represent variable nodes. Dashed line partitions the factor graph into two sub-graphs

A. Exact inference

The estimation task is facilitated by the following factorization of the posterior pdf shown by the factor graph in Fig. 1.

$$p(\mathbf{x}, \mathbf{s} \mid \mathbf{y} = \mathbf{y}_0) \propto p(\mathbf{y} = \mathbf{y}_0 \mid \mathbf{x}, \mathbf{s})\, p(\mathbf{x}, \mathbf{s}) = \underbrace{p(\mathbf{s})}_{\triangleq h(\mathbf{s})} \underbrace{p(\mathbf{y} = \mathbf{y}_0 \mid \mathbf{x})}_{\triangleq g(\mathbf{x})} \prod_{n=1}^{N} \underbrace{p(x_n \mid s_n)}_{\triangleq f_n(x_n, s_n)} \tag{3}$$

We use $\propto$ to denote equality after scaling to unit area.

The MMSE estimate of x_n is given by the mean of the marginal posterior $p(x_n \mid \mathbf{y} = \mathbf{y}_0)$, which can be written as

$$p(x_n \mid \mathbf{y}=\mathbf{y}_0) = \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} p(\mathbf{x},\mathbf{s} \mid \mathbf{y}=\mathbf{y}_0) = \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} \frac{p(\mathbf{x},\mathbf{s},\mathbf{y}=\mathbf{y}_0)}{p(\mathbf{y}=\mathbf{y}_0)} \tag{4}$$

$$\propto \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} p(\mathbf{x},\mathbf{s},\mathbf{y}=\mathbf{y}_0) = \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} p(\mathbf{s} \mid \mathbf{x},\mathbf{y}=\mathbf{y}_0)\, p(\mathbf{y}=\mathbf{y}_0 \mid \mathbf{x})\, p(\mathbf{x}) \tag{5}$$

$$= \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} p(\mathbf{s} \mid \mathbf{x})\, p(\mathbf{y}=\mathbf{y}_0 \mid \mathbf{x})\, p(\mathbf{x}) = \sum_{\mathbf{s}\in\{0,1\}^N} \int_{\mathbf{x}_{-n}} g(\mathbf{x})\, p(\mathbf{x} \mid \mathbf{s})\, p(\mathbf{s}) \tag{6}$$

$$= \sum_{s_n=0}^{1} f_n(x_n, s_n)\, p(s_n) \int_{\mathbf{x}_{-n}} g(\mathbf{x}) \prod_{q\neq n} \sum_{s_q=0}^{1} f_q(x_q, s_q) \sum_{\mathbf{s}_{-n,q}\in\{0,1\}^{N-2}} p(\mathbf{s}_{-n} \mid s_n) \tag{7}$$

Where $\mathbf{s}_{-n}$ denotes vector $\mathbf{s}$ with the n^{th} element omitted, and $\mathbf{s}_{-n,q}$ denotes $\mathbf{s}$ with both the n^{th} and q^{th} elements omitted. Writing $p(\mathbf{s}_{-n} \mid s_n) = p(\mathbf{s}_{-n,q} \mid s_q, s_n)\, p(s_q \mid s_n)$, the last summation in (7) reduces to $p(s_q \mid s_n)$, giving

$$p(x_n \mid \mathbf{y}=\mathbf{y}_0) \propto v_{f_n \to x_n}(x_n)\, v_{g \to x_n}(x_n) \tag{8}$$

$$v_{f_n \to x_n}(x_n) \triangleq \sum_{s_n=0}^{1} f_n(x_n, s_n)\, p(s_n) \tag{9}$$

$$v_{g \to x_n}(x_n) \triangleq \int_{\mathbf{x}_{-n}} g(\mathbf{x}) \prod_{q\neq n} \sum_{s_q=0}^{1} f_q(x_q, s_q)\, p(s_q \mid s_n) \tag{10}$$

B. Implementing the Message Passes

Whereas exact posterior calculation via (8)-(10) is computationally prohibitive for typical problem sizes, approximate calculation can be efficiently accomplished using message passing. Using the framework of BP, the functions $v_{f_n \to x_n}(\cdot)$ and $v_{g \to x_n}(\cdot)$ can be approximated.

$$v^{(t)}_{f_n \to x_n}(x_n) \propto \sum_{s_n=0}^{1} f_n(x_n, s_n)\, v^{(t)}_{s_n \to f_n}(s_n) \tag{11}$$

$$v^{(t)}_{g \to x_n}(x_n) \propto \int_{\mathbf{x}_{-n}} g(\mathbf{x}) \prod_{q\neq n} \underbrace{\sum_{s_q=0}^{1} f_q(x_q, s_q)\, v^{(t)}_{s_q \to f_q}(s_q)}_{v_{f_q \to x_q}(x_n) = v_{x_q \to g}(x_n)} \tag{12}$$

Which depend on the other messages

$$v^{(t)}_{s_n \to f_n}(s_n) = v^{(t)}_{h \to s_n}(s_n) \propto \sum_{\mathbf{s}_{-n}\in\{0,1\}^{N-1}} h(\mathbf{s}) \prod_{q\neq n} \underbrace{v^{(t-1)}_{s_q \to h}(s_q)}_{= v^{(t-1)}_{f_q \to s_q}(s_q)} \tag{13}$$

$$v^{(t)}_{f_n \to s_n}(s_n) \propto \int_{x_n} f_n(x_n, s_n) \underbrace{v^{(t)}_{x_n \to f_n}(x_n)}_{=v^{(t)}_{g \to f_n}(x_n)} \tag{14}$$

We use the superscript-(t) to denote iteration. These messages can then be combined for marginal inference:

$$p^{(t)}(x_n \mid \mathbf{y} = \mathbf{y}_0) \propto v^{(t)}_{f_n \to x_n}(x_n) v^{(t)}_{g \to x_n}(x_n) \tag{15}$$

$$p^{(t)}(s_n \mid \mathbf{y} = \mathbf{y}_0) \propto v^{(t)}_{f_n \to s_n}(s_n) v^{(t)}_{h \to s_n}(s_n) \tag{16}$$

Where $p^{(t)}$ denotes the iteration-t approximation to the pdf.

We now partition our factor graph into the two sub-graphs separated by the dashed line in Fig.1. The message $\left\{v^{(t)}_{f_n \to s_n}(\cdot)\right\}_{n=1}^{N}$ form the outputs of the left sub-graph and the inputs to the right one, while the messages $\left\{v^{(t)}_{h \to s_n}(\cdot)\right\}_{n=1}^{N}$ form the outputs of the right sub-graph and the inputs to the left one. From this, we can interpret the BP scheme as iterationg between two blocks, one which performs inference on the left sub-graph (which models structure in the observation) and the other which performs inference on the right sub-graph (which models structure in the sparsity pattern), with message-passing between blocks.

We will henceforth refer to inference on the left sub-graph of Fig.1 as "sparsity pattern equalization" (SPE) and inference on the right sub-graph as "sparsity pattern decoding" (SPD). We now formally decouple these subtasks and represent each of them using a separate factor graph, as in Fig. 2. For this, we define two additional t^{th} iteration constraint functions,

$$h^{(t)}_n(s_n) \triangleq v^{(t)}_{h \to s_n}(s_n) \tag{17}$$

$$d^{(t)}_n(s_n) \triangleq v^{(t-1)}_{f_n \to s_n}(s_n) \tag{18}$$

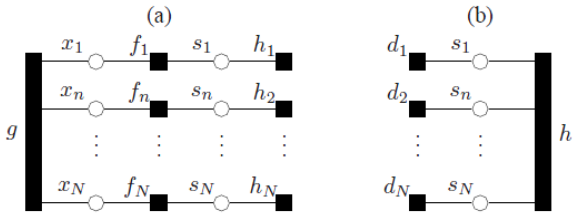


Figure 2 Decoupling of partitioned factor graph from Fig. 1 into
(a) sparsity pattern equalization and (b) sparsity pattern decoding.

IV. SPARSITY PATTERN EQUALIZATION

Below we outline a BP-based technique that follows the "approximate message passing" (AMP) framework recently proposed by Donoho, Maleki, and Montanari. Since we focus on a single iteration t, we suppress the superscript-(t) notation on messages in this section.

For BP-based SPE, we expand the g node in Fig. 2(a), yielding the loopy factor graph in Fig.3, with constraints

$$g_m(x) \triangleq CN(y_m; a_m^H \mathbf{x}, \sigma^2) \tag{19}$$

Where a_m^H denotes the m^{th} row of $\mathbf{A}$. Noting that SPE will require several iterations of message passing between nodes $\{g_m\}$ and $\{x_n\}$, we will henceforth use $v^i_{x_n \to g_m}$ and $v^i_{g_m \to x_n}$ to denote the SPE-iteration-i messages. In addition, we will assume Gaussian active-coefficients, i.e.,

$$q_n(x_n) = CN(x_n; 0, \sigma_n^2) \tag{20}$$

We use λ_n to abbreviate $h_n(1)$, the prior probability of $s_n = 1$ assumed by SPE. Thus, the coefficient is Bernoulli-Gaussian, with the form

$$v_{f_n \to x_n}(x_n) = \lambda_n CN(x_n; 0, \sigma_n^2) + (1-\lambda_n)\delta(x_n) \tag{21}$$

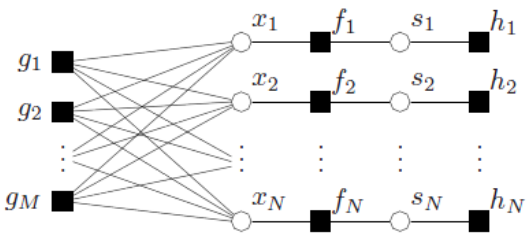


Figure 3 Factor graph for BP-based implementation of SPE

A. BP approximation via the large-system limit

Exact calculation of $v^i_{g_m \to x_n}(x_n)$ would involve the iteration of 2^{N-1} terms, which is cleary impractical. However, in the large system limit(i.e., $M, N \to \infty$ with M / N fixed), the central limit theorem motivates the treatment of $v^i_{g_m \to x_n}(x_n)$ as Gaussian. In this case, it is sufficient to parameterize the inputs to g_m via

$$\mu^i_{nm} \triangleq \int_{x_n} x_n v^i_{x_n \to g_m}(x_n) \tag{22}$$

$$v^i_{nm} \triangleq \int_{x_n} (x_n - \mu^i_{nm})^2 v^i_{x_n \to g_m}(x_n) \tag{23}$$

Which yields outputs from g_m that take the form

$$v^i_{g_m \to x_n}(x_n) \propto CN(A_{mn} x_n; z^i_{mn}, c^i_{mn}) \tag{24}$$

$$z^i_{mn} \triangleq y_m - \sum_{q \neq n} A_{mq} \mu^i_{qm} \tag{25}$$

$$c^i_{mn} \triangleq \sigma^2 + \sum_{q \neq n} |A_{mq}|^2 \mu^i_{qm} \tag{26}$$

From (22), (23), we see that μ^{i+1}_{nm} and v^{i+1}_{nm} are then determined by the mean and variance, respectively of the pdf

$$v^{i+1}_{x_n \to g_m}(x_n) \propto v_{f_n \to x_n}(x_n) \prod_{l \neq m} v^i_{g_l \to x_n}(x_n) \tag{27}$$

Using following equation

$$\prod_q CN(x; \mu_q, v_q) \propto CN\left(x; \frac{\sum_q \frac{\mu_q}{v_q}}{\sum_q \frac{1}{v_q}}, \frac{1}{\sum_q \frac{1}{v_q}} \right) \tag{28}$$

The product term in (27) reduces to

$$CN\left(x_n; \frac{\sum_{l\neq m} A_{\ln}^* z_{\ln}^i / c_{\ln}^i}{\sum_{l\neq m} |A_{\ln}|^2 / c_{\ln}^i}, \frac{1}{\sum_{l\neq m} |A_{\ln}|^2 / c_{\ln}^i} \right) \tag{29}$$

And so, under the large-system-limit approximations

$$c_{\ln}^i \approx c_n^i \triangleq \frac{1}{M} \sum_{m=1}^M c_{mn}^i \tag{30}$$

And $\sum_{l\neq m} |A_{\ln}|^2 \approx \sum_{l=1}^M |A_{\ln}| = 1$, (27) simplifies to

$$v_{x_n \to g_m}^{i+1}(x_n) \propto (\lambda_n CN(x_n; 0, \sigma_n^2) + (1-\lambda_n)\delta(x_n)) \times CN(x_n; \sum_{l\neq m} A_{\ln}^* z_{\ln}^i, c_n^i) \tag{31}$$

Applying (28) to (31), we find, after some algebra, that

$$\mu_{nm}^{i+1} = \alpha_n(c_n^i)\theta_{nm}^i / (1+\gamma_{nm}^i) \tag{32}$$

$$v_{nm}^{i+1} = \gamma_{nm}^i |\mu_{nm}^{i+1}|^2 + \mu_{nm}^{i+1} c_n^i / \theta_{nm}^i \tag{33}$$

$$\theta_{nm}^i \triangleq \sum_{l\neq m} A_{\ln}^* z_{\ln}^i \tag{34}$$

$$\gamma_{nm}^i \triangleq \beta_n(c_n^i) \exp(-\zeta_n(c_n^i) |\theta_{nm}^i|^2) \tag{35}$$

Where $\alpha_n(c) \triangleq \frac{\sigma_n^2}{c+\sigma_n^2}$, $\beta_n(c) \triangleq \frac{1-\lambda_n}{\lambda_n} \frac{c+\sigma_n^2}{c}$, $\zeta_n(c) \triangleq \frac{\sigma_n^2}{c(c+\sigma_n^2)}$.

The i^{th} SPE iteration yields the x_n-posterior approximation

$$p^i(x_n | \mathbf{y} = \mathbf{y}_0) \propto v_{f_n \to x_n}(x_n) \prod_{l=1}^M v_{g_{l\to x_n}}^{i-1}(x_n) \tag{36}$$

The mean and variance of (36) constitute the MMSE estimate of x_n and its MSE. Nothing that (35) differs from (27) only in the inclusion of the m^{th} product term.

B. Approximate message passing

The approximate BP algorithm outlined updates $O(NM)$ variables per iteration. When N and M are large, the resulting complexity may be undesirably high, motivating us to find a simpler scheme.

Recently, Donoho, Maleki proposed AMP algorithms that greatly simplify BP algorithms of the form outlined by tracking only $O(N)$ variables. Using AMP, we find that

$$\theta_n^i = \sum_{m=1}^{M} A_{mn}^* z_m^i + \mu_n^i \tag{37}$$

$$\mu_n^{i+1} = F_n(\theta_n^i; c^i) \tag{38}$$

$$v_n^{i+1} = G_n(\theta_n^i; c^i) \tag{39}$$

$$c^{i+1} = \sigma^2 + \frac{1}{M}\sum_{n=1}^{N} v_n^{i+1} \tag{40}$$

$$z_m^{i+1} = y_m - \sum_{n=1}^{N} A_{mn}\mu_n^{i+1} + \frac{z_m^i}{M}\sum_{n=1}^{N} F_n^{'}(\theta_n^i; c^i) \tag{41}$$

Above, $F_n(.;.)$, $G_n(.;.)$, and $F_n^{'}(.;.)$ are nonlinear functions that depend on the coefficient prior. We chose the Bernoulli-Gaussian prior. Thus, the nonlinear functions take the following form

$$F_n(\theta; c) = \frac{\alpha_n(c)}{1+\beta_n(c)e^{-\zeta_n(c)|\theta|^2}}\theta \tag{42}$$

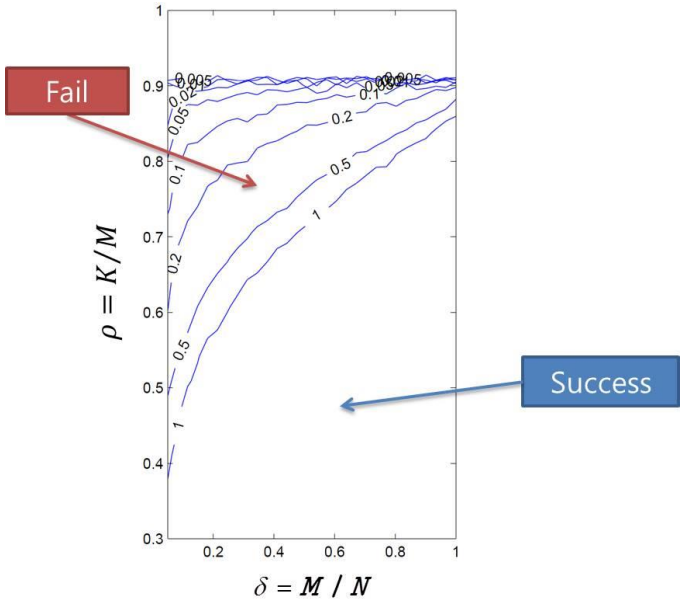
$$G_n(\theta; c) = \beta_n(c)e^{-\zeta_n(c)|\theta|^2} \mid F_n(\theta; c)\mid^2 + \frac{c}{\theta}F_n(\theta; c) \tag{43}$$

$$F_n^{'}(\theta; c) = \frac{\alpha_n(c)}{1+\beta_n(c)e^{-\zeta_n(c)|\theta|^2}} \times \left[1 + \frac{\zeta_n(c)\mid\theta\mid^2}{1+(\beta_n(c)e^{-\zeta_n(c)|\theta|^2})^{-1}}\right] \tag{44}$$

V. Numerical Results

Numerical experiments were conducted for the observation model (1), where the elements of $\mathbf{A}$ were independently drawn from a $CN\left(0, \frac{1}{M}\right)$ distribution and where the signal coefficients were generated via $p(x_n \mid s_n) = s_n CN(x_n; 0,1) + (1-s_n)\delta(x_n)$ using Markov chain-generated binary sprsity pattern $\{s_n\}$. We set $\gamma \in (0,1]$ called the Markov independence parameter. Note

that, as γ increases, the pattern becomes less correlated, with $\gamma = 1$ corresponding to an i.i.d pattern.



VI. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

Appendix

Reference

Compressive fluorescence microscopy for biological and hyperspectral imaging

Vincent Studer et al.

PNAS. (2012.12)

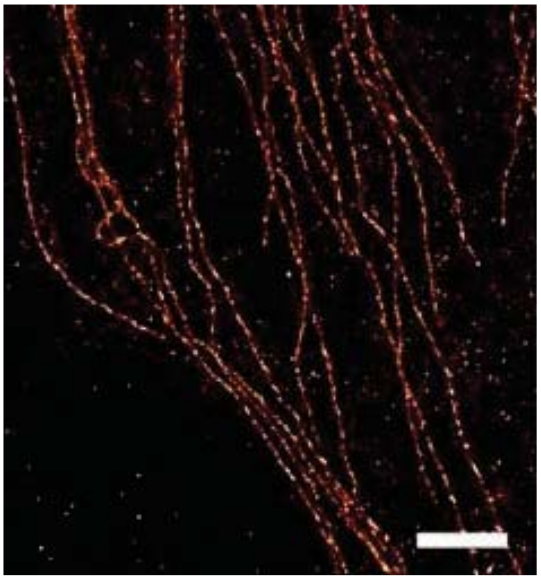
Presenter : Eunseok Jung

GIST, Dept. of Mechatronics , BiO-scopy Lab.

Gwangju Institute of Science and Technology

Background

- Fluorescence protein
  - Fluorescence microscopy use fluorescence protein.
  - Fluorescence protein can help to see molecule structure or phenomenon.



Cited form : http://huanglab.ucsf.edu, Wikipedia, http://www.microscopyu.com, http://smb.snu.ac.kr/tools_updating.htm

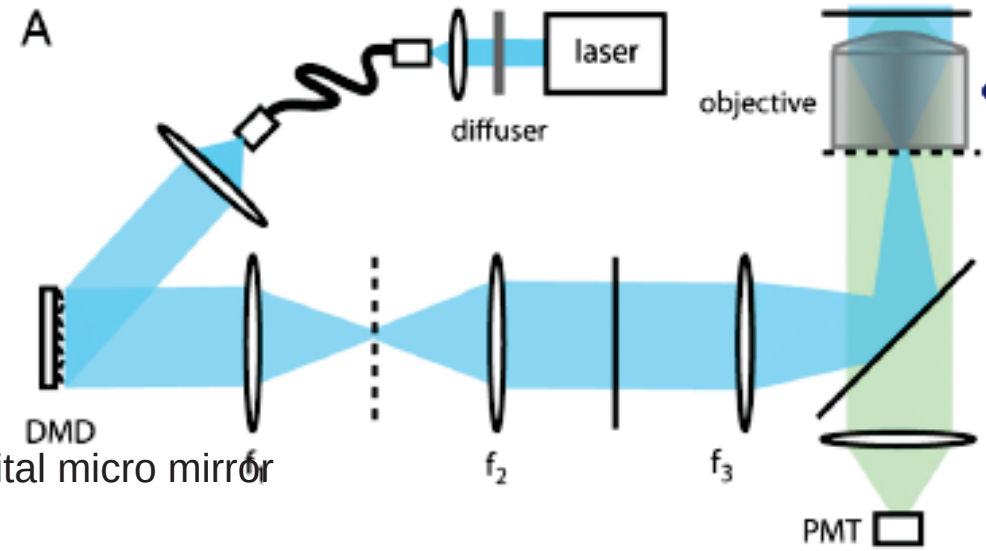
Introduction & Motivation

- CS algorithm make up fluorescence microscope major drawback

1. CS help to imaging in diffusing media.
2. CS can decrease experiment time.
3. If we use CS algorithm, it doesn't need expensive CCD camera.

From : Imaging Intracellular Fluorescent Proteins at Nanometer Resolution, E. Beitzig, science, 2006

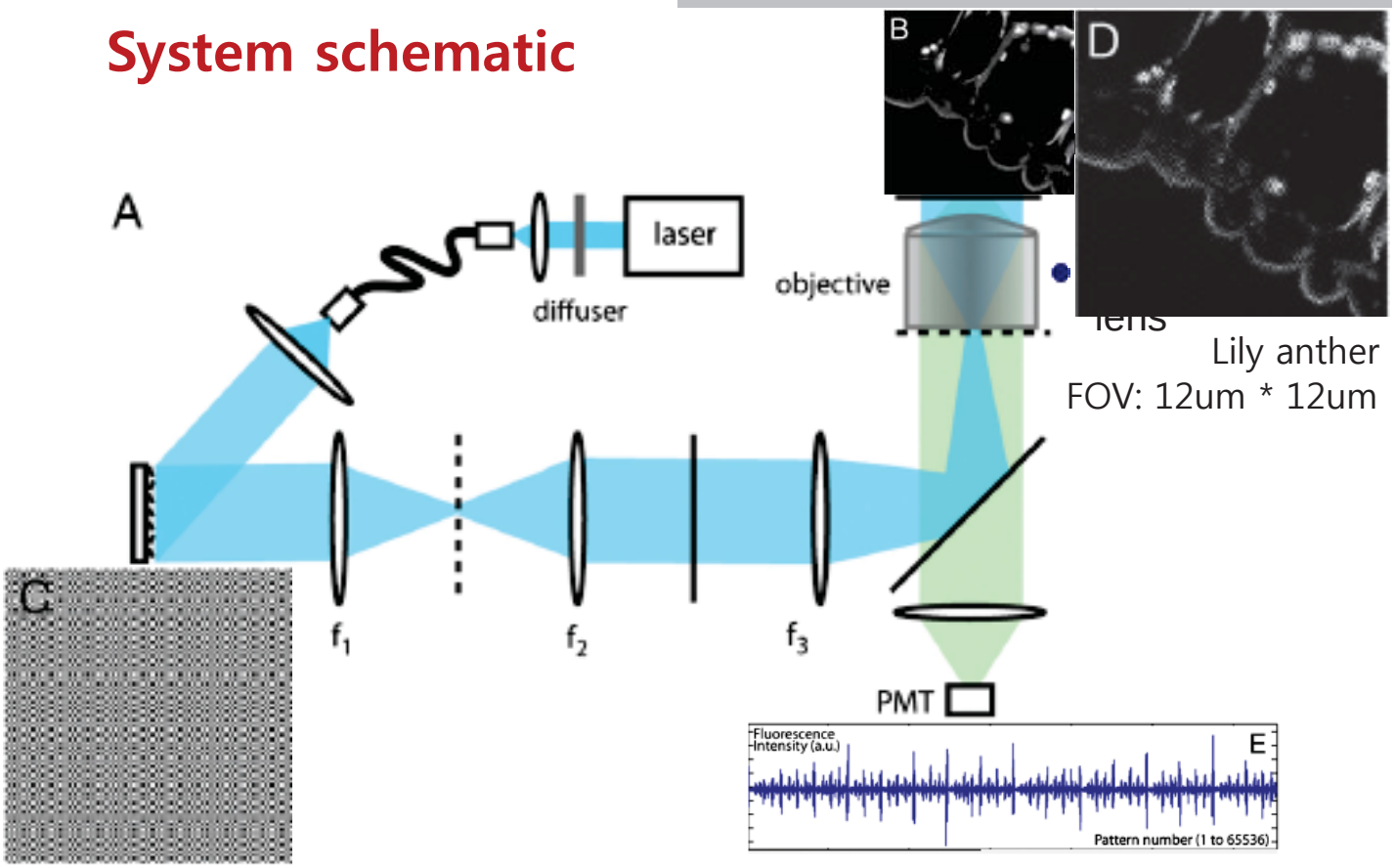
System schematic



- Sample plane
- 60x objective lens
- Digital micro mirror
- Photo multiplier tube(Detector)

From : Imaging Intracellular Fluorescent Proteins at Nanometer Resolution, E. Beitzig, science, 2006

System schematic



From : Imaging Intracellular Fluorescent Proteins at Nanometer Resolution, E. Beitzig, science, 2006

Reconstruction

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{W}^T\mathbf{x}\|_{\ell_1} \text{ subject to } \|\mathbf{y}-\mathbf{\Phi x}\|_{\ell_2} \le \epsilon$$

- Recovering the signal x from acquired signal by solving the optimization problem.
- The acquired signal is noisy, it is better to relax the constraints into

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{W}^T\mathbf{x}\|_{\ell_1} + \frac{\alpha}{2}\|\mathbf{y}-\mathbf{\Phi x}\|_{\ell_2}^2$$

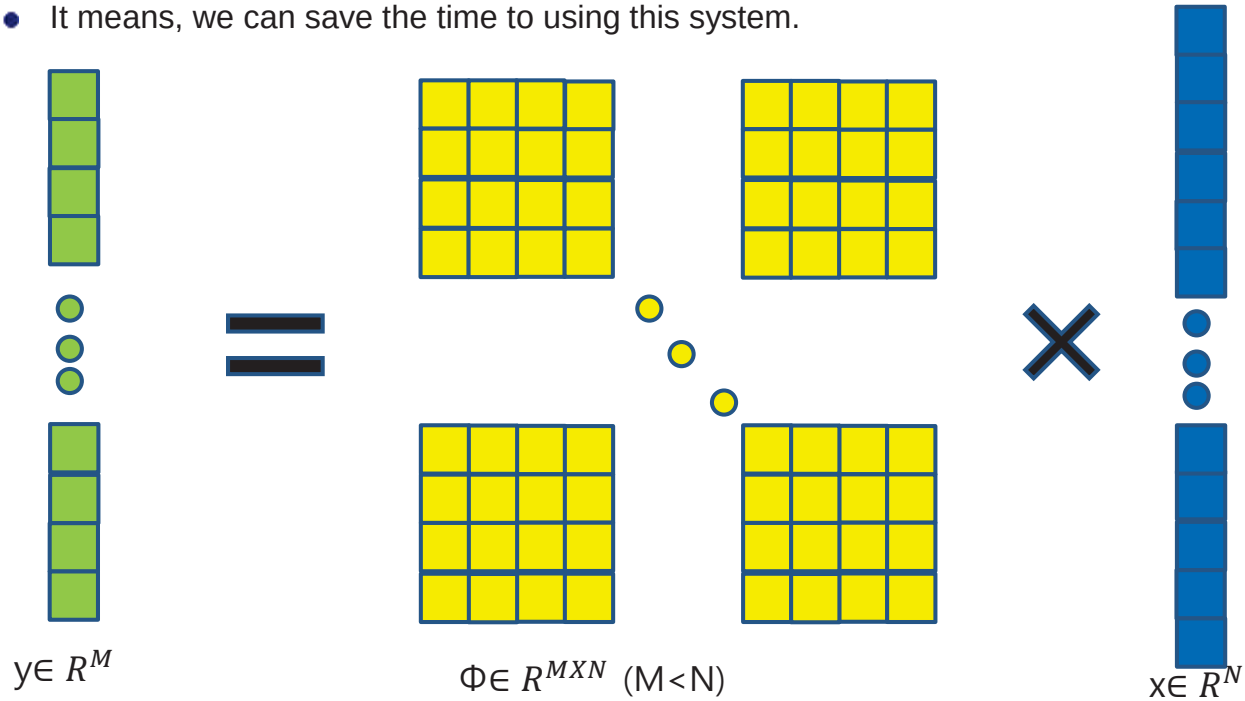
- **W** will be either an orthonormal basis(e.g., Dirac basis) or an overcomplete signal representation(e.g., undecimated wavelet frame or curvelet frame).
- $\alpha(\epsilon)$ is chosen empirically depending on the noise level.

Reconstruction

- M=N/Under sampling-ratio.
- Under sampling ratio = 8, 16, 32, 64,...

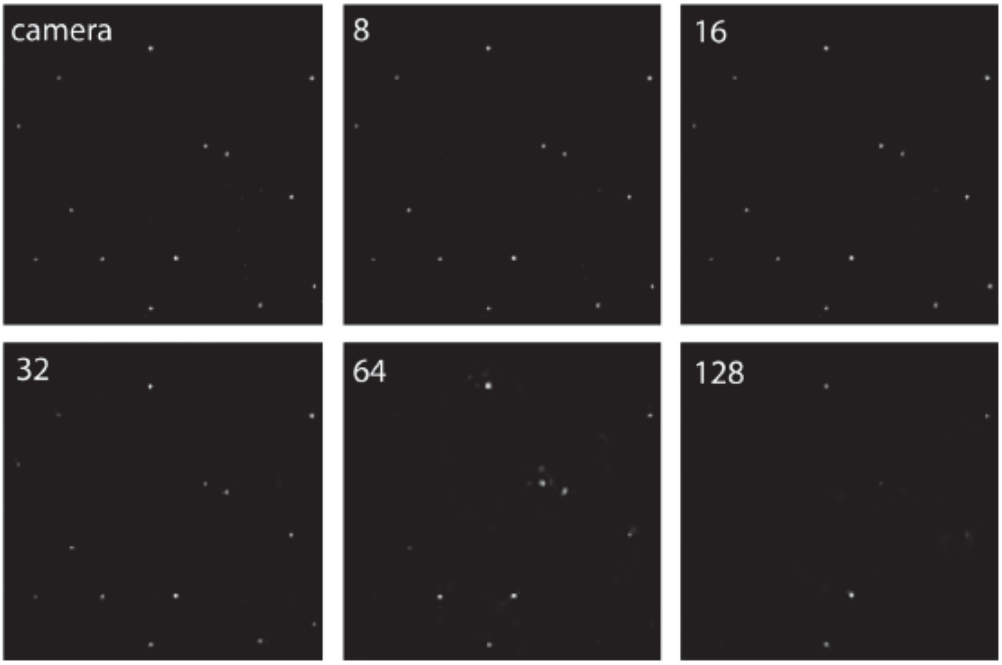
  => M=8192, 4096, 2048,…
- It means, we can save the time to using this system.

N=256*256,
It was fixed by pixel

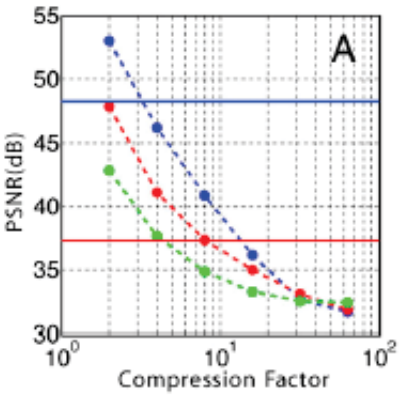


y∈ R^M Φ∈ R^{MXN} (M<N) x∈ R^N

Results

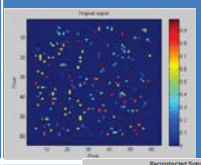
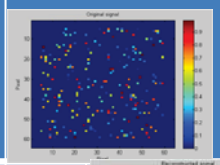
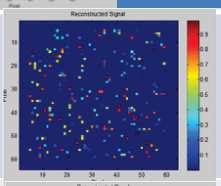
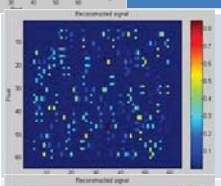
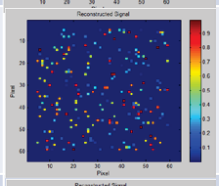
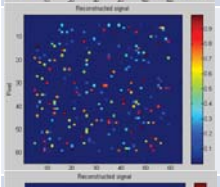
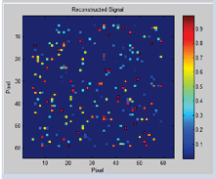
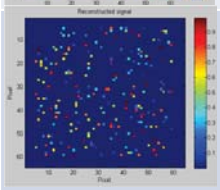
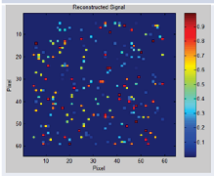
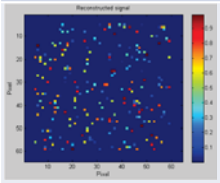


- Top left to bottom right: camera snapshot and reconstructed 256-by-256 bead images for values of the undersampling ratio equal to 8, 16, 32, 64, and 128.
- FOV: 6um * 6um



- Nominal illumination level(blue) and for the same level reduced by a factor 10(red) and a factor of 100(green). Solid lines correspond to the PSNR in raster scan for the same surfacic illumination(Blue: I, Red: I/10).

Simulation

		Random matrix	Hadamard matrix
			
Rank : 1000	PSNR(dB)	78 	20 
Rank : 2000	PSNR(dB)	87 	31 
Rank : 3000	PSNR(dB)	88 	96 
Rank : 4000	PSNR(dB)	90 	96 

Discussion

- Conclusion

1. Reconstruct result is affected by measurement matrix.
   =>If we can make measurement matrix well, the reconstructed image will get high resolution image.

2. Fluorescence microscopy imaging is possible in diffusing media.

Thank you

Simplified Relay Selection and Power Allocation in Cooperative Cognitive Radio Systems

Authors: Liying Li et. al.

Publication: IEEE Trans. Wireless Comm., Jan. 2011

Speaker: Asif Raza

Short summary: In this paper authors propose solution of a combined problem; relay selection and power allocation to secondary users under the constraint of limited interference to primary users in cognitive radio (CR) system. Objective of the joint problem was to maximize system throughput. A high complexity optimal solution and a low complexity suboptimal solution are proposed. The presented solutions show over 50% improvement in system throughput.

I. INTRODUCTION

Cooperative technique for spectrum sensing and sharing in CR networks has been investigated in the literature. It can obtain spatial diversity and combat detrimental effects of wireless channels however it has some limitations associated. For example, while doing relay selection and resource allocation one must also consider spectrum efficiency and interference limitation as well. Authors in this paper consider these combined issues i.e. relay selection and power allocation with interference limitation.

II. SYSTEM DESCRIPTION

In order to have effective cooperation following decisions must be made prior to cooperation:

- When to cooperate
- To whom cooperate with
- What resources to share and how to share?

These decisions are basis of relay selection and power allocation problem.

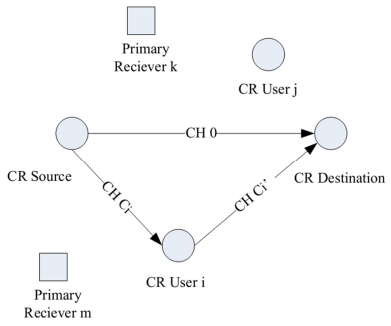


Figure 1: The structure of a cooperative CR network

A simple three-node relay system where each CR user can only help one CR transceiver pair is shown in the figure. The source node transmits data to the destination and the relay simultaneously using orthogonal channels, Channel (CH) 0 and CH C_i, respectively. The relay node forwards scaled version of the received signal from CH C_i to the destination node using CH C_i'.

Power modeling:

Existing relay selection schemes does not consider interference issue. In order to prevent primary users from interference the transmission powers on channels (CH 0), CH C_i and CH C_i' must satisfy:

$$\begin{aligned} P_{1,i}\left|h_{s,p,d}\right|^2 &\le I_1 \\ P_{2,i}\left|h_{s,p,i}\right|^2 &\le I_2 \qquad (1) \\ P_{3,i}\left|h_{i,p}\right|^2 &\le I_3 \end{aligned}$$

Where $h_{s,p,i}$ and $h_{s,p,d}$ are channel gain between CR source and primary users of CH C_i and (CH 0) respectively while $h_{i,p}$ is channel gain between CR source and primary user of channel CH C_i'. I_1, I_2 and I_3 are acceptable interference powers of primary users over channel CH 0, CH C_i and CH C_i' respectively. The overall transmission power of CR source and relay nodes are limited as:

$$P_{1,i} + P_{2,i} \le P_{total} \; and \; P_{3,i} \le P_3$$

The channels under consideration i.e. CH 0, CH C_i and CH C_i' are $\mathcal{N}(0, \sigma^2)$ with known channel gains at CR source and CR relay nodes. If ith CR user is relay node then signal and noise powers at destination from relay is:

$$P_{s,i} = \frac{P_{3,i}P_{2,i}\left|h_{s,i}\right|^2\left|h_{i,d}\right|^2}{P_{2,i}\left|h_{s,i}\right|^2+\sigma^2}$$

$$P_{n,i} = \left(\frac{P_{3,i}\left|h_{i,d}\right|^2}{P_{2,i}\left|h_{s,i}\right|^2+\sigma^2}+1\right)\sigma^2$$

$$SNR_i = \frac{P_{s,i}}{P_{n,i}}$$

Where $h_{s,i}$ and $h_{i,d}$ are channel gains between source and relay and relay and destination nodes respectively and SNR_i is SNR value at destination from ith relay channel.

Throughput Calculation:

The system throughput for ith relay is:

$$T_i\left(P_{1,i},P_{2,i},P_{3,i}\right)=\left(1-\alpha\right)\log 2\left(1+\frac{P_{1,i}\left|h_{s,d}\right|^2}{\sigma^2}\right)+\left(1-\alpha\right)\log 2\left(1+\frac{P_{3,i}\left|h_{i,d}\right|^2 P_{2,i}\left|h_{s,i}\right|^2}{\left(P_{3,i}\left|h_{i,d}\right|^2+P_{2,i}\left|h_{s,i}\right|^2+\sigma^2\right)\sigma^2}\right) \quad (2)$$

Where α is mis-detection probability of spectrum sensing.

This equation tells us that data will be lost if interference happens.

III. ALGORITHM DEVELOPMENT

Optimal and suboptimal algorithms are developed for power allocation and relay selection problem.

A. Optimal Approach

Optimization problem is formulated as:

$$\left(P_{1,i}^*,P_{2,i}^*,P_{3,i}^*\right)=\arg\max_{P_{1,i},P_{2,i},P_{3,i}} T_i\left(P_{1,i},P_{2,i},P_{3,i}\right) \quad (3a)$$

$$i=\arg\max_i T_i\left(P_{1,i}^*,P_{2,i}^*,P_{3,i}^*\right) \quad (3b)$$

subject to

$$P_{1,i}+P_{2,i}\le P_{total} \quad (3c)$$

$$P_{3,i}\le P_3 \quad (3d)$$

$$0\le P_{1,i}\left|h_{s,p,d}\right|^2\le I_1 \quad (3e)$$

$$0\le P_{2,i}\left|h_{s,p,i}\right|^2\le I_2 \quad (3f)$$

$$0\le P_{3,i}\left|h_{i,p}\right|^2\le I_3 \quad (3g)$$

Lagrange multiplier is then used to divide the problem into subproblems and then obtain solution for subproblems:

$$L\left(P_{1,i},P_{2,i},\lambda_1,\lambda_2,\lambda_3\right)=-\left(1-\alpha\right)\log_2\left(1+\frac{P_{1,i}\left|h_{s,d}\right|^2}{\sigma^2}\right)-\left(1-\alpha\right)\log_2\left(1+\frac{P_{3,i}\left|h_{i,d}\right|^2 P_{2,i}\left|h_{s,i}\right|^2}{\left(P_{3,i}\left|h_{i,d}\right|^2+P_{2,i}\left|h_{s,i}\right|^2+\sigma^2\right)\sigma^2}\right)$$
$$+\lambda_1\left(P_{1,i}+P_{2,i}-P_{total}\right)+\lambda_2\left(P_{1,i}\left|h_{s,p,d}\right|^2-I_1\right)+\lambda_3\left(P_{2,i}\left|h_{s,p,i}\right|^2-I_2\right) \quad (4)$$

According to Karush-Kuhn-Tucker conditions:

$$0\le P_{1,i},\ 0\le P_{2,i} \text{ and } \lambda_i\ge 0\ \ \forall i$$

$$\lambda_1\left(P_{1,i}+P_{2,i}-P_{total}\right)=0,$$
$$\lambda_2\left(P_{1,i}\left|h_{s,p,d}\right|^2-I_1\right)=0,$$
$$\lambda_3\left(P_{2,i}\left|h_{s,p,i}\right|^2-I_2\right)=0$$
$$and$$
$$\frac{\partial L}{\partial P_{1,i}}=0\,and\,\frac{\partial L}{\partial P_{2,i}}=0$$

Solving them by using dual-domain and sub-gradient method we get solution for Lagrangian dual variables as:

$$\lambda_1^{(n+1)}=\left[\lambda_1^n+\mu^n\left(P_{1,i}^n\left(\lambda_1^n,\lambda_2^n,\lambda_3^n\right)+P_{2,i}^n\left(\lambda_1^n,\lambda_2^n,\lambda_3^n\right)-P_{total}\right)\right]^+,$$
$$\lambda_2^{(n+1)}=\left[\lambda_2^n+\mu^n\left(P_{1,i}^n\left(\lambda_1^n,\lambda_2^n,\lambda_3^n\right)\left|h_{s,p,d}\right|^2-I_1\right)\right]^+,$$
$$\lambda_3^{(n+1)}=\left[\lambda_3^n+\mu^n\left(P_{2,i}^n\left(\lambda_1^n,\lambda_2^n,\lambda_3^n\right)\left|h_{s,p,i}\right|^2-I_2\right)\right]^+$$

Here 'μ^n' is sequence of scalar step-sizes. Once we get $\lambda_i,\forall i$ we can calculate $P_{1,i}$ and $P_{2,i}$ as follows;

$$P_{1,i}=\left[\frac{1-\alpha}{\left(\lambda_1+\lambda_2\right)\left|h_{s,p,d}\right|^2\ln 2}-\frac{\sigma^2}{\left|h_{s,d}\right|^2}\right]^+,$$
$$P_{2,i}=\left[\frac{\sqrt{P_{3,i}^2\left|h_{i,d}\right|^4+4P_{3,i}\left|h_{i,d}\right|^2\left|h_{s,i}\right|^2K}}{2\left|h_{s,i}\right|^2}-\frac{P_{3,i}\left|h_{i,d}\right|^2+2\sigma^2}{2\left|h_{s,i}\right|^2}\right]^+$$

Where $K=\frac{1-\alpha}{\left(\lambda_1+\lambda_3\right)\left|h_{s,p,i}\right|^2\ln 2}$ and $\left[\cdot\right]^+=\max\left(\cdot,0\right)$

By using these values we get optimal power allocation $P_{1,i}^*,P_{2,i}^*$. Note that $P_{3,i}^*=\min\left\{\frac{I_3}{\left|h_{i,p}\right|^2},P_3\right\}$.

Finally system throughput T_i^* when i[th] CR node acts as relay is then calculated from equation …

B. Sub-Optimal Approach

Joint relay selection and power allocation problem provides optimal throughput yet it is quite complex algorithm. A low complexity, sub-optimal version of the problem can be defined as follows:

Transmission power constraints of CR nodes are defined as:

$$P_{1,i} = \frac{I_1}{\left|h_{s,p,d}\right|^2}$$

$$P_{2,i} = \frac{I_2}{\left|h_{s,p,i}\right|^2}$$

In order to calculate system throughput by using equation 2 we need to choose relay as:

$$i = \arg\max_i \frac{P_{3,i}\left|h_{s,i}\right|^2\left|h_{i,d}\right|^2}{P_{3,i}\left|h_{i,d}\right|^2\left|h_{s,p,i}\right|^2 + I_2\left|h_{s,i}\right|^2 + \sigma^2\left|h_{s,p,i}\right|^2}$$

The optimal power limit and other constraints in (2) are taken as total power limit.

IV. Simulation Results

Parameters:

Interference limits:= $I_1 = I_2 = I_3 = 0.1mW$, Path loss exponent= 4, channel of unit bandwidth is used. Channel fading follows Rayleigh distribution with $\sigma = 6dB$

Results:

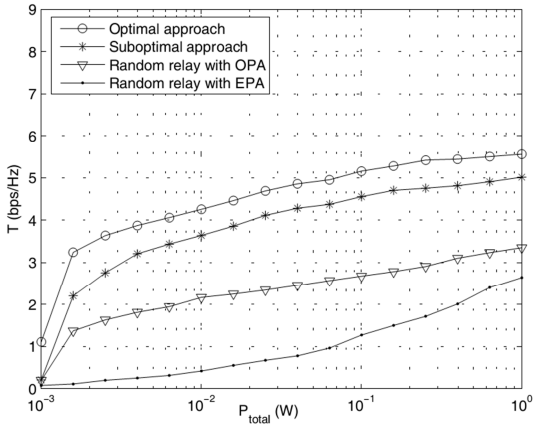


Fig. 2. System throughput versus transmission power limit of CR source

Transmission power limit of relay node is P_3 = 0.5W and no. of candidate relay nodes = 20 From figure we see proposed scheme achieves about 50% throughput achievement over optimal power allocation (OPA) and equal power allocation (EPA) schemes. Comparing sub-optimal scheme with optimal scheme we see only about 15% degradation in throughput is observed. Moreover in low P_{total} region the system throughput increase rapidly however for high P_{total} region the growth is restricted due to interference limits.

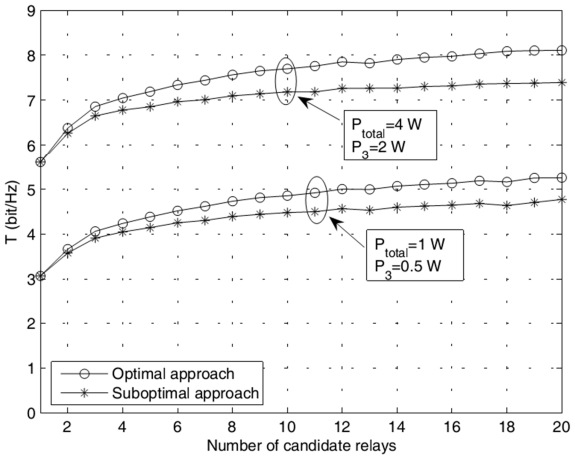


Fig. 3. System throughput versus number of candidate relays.

Figure 3 shows that gap between optimal and sub-optimal schemes is small well when the number of candidate users is small. This shows that sub-optimal approach performs well when number of relays are small.

Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit

J. A. Tropp and A. C. Gilbert

IEEE Trans. on Inform. Theory

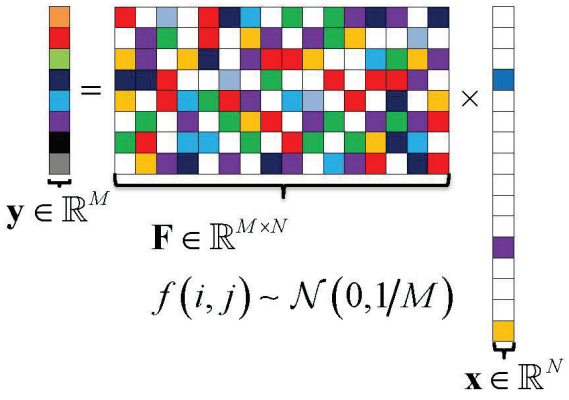
Presenter : Sangjun Park

GIST, Dept. of Information and Communication, INFONET Lab.

Gwangju Institute of Science and Technology

Questions and System Model

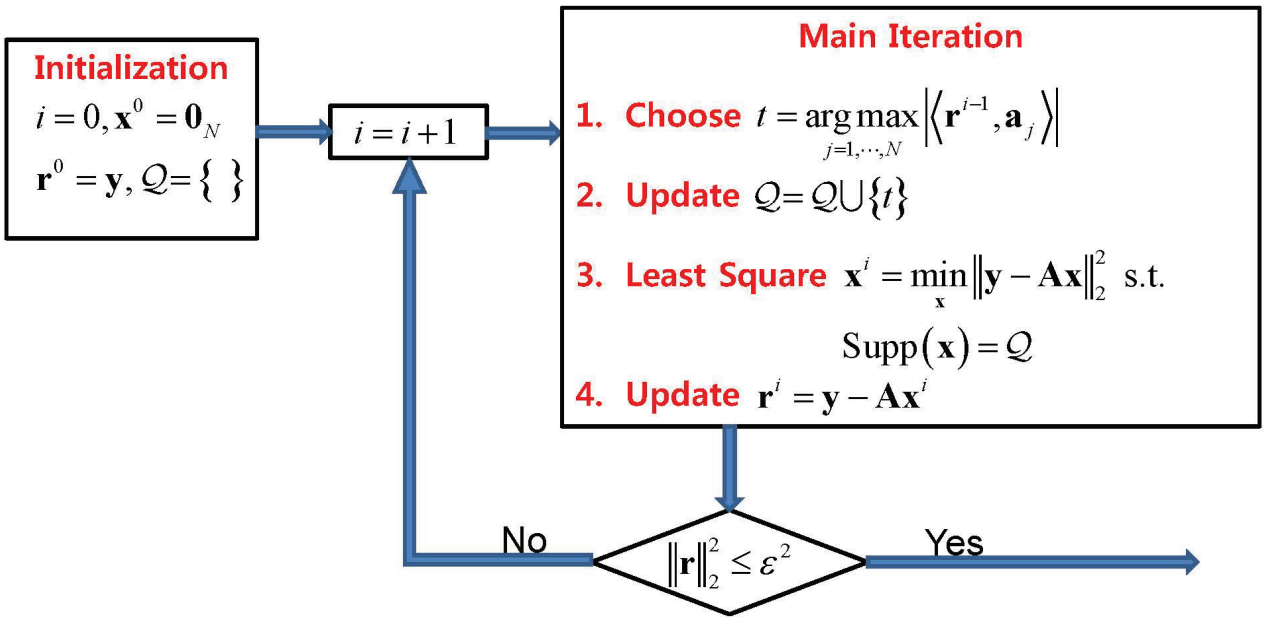
- Let us suppose that we aim to find the support set of a sparse vector by using OMP.
- Then, what is a sufficient condition for successful OMP?



Orthogonal Matching Pursuit

- OMP finds one index at a time for approximating the solution of

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \le \varepsilon^2$$



Sufficient conditions for successful OMP

- There are many papers that report sufficient conditions for successful OMP.

Year	A sufficient condition	Types
2004	$\mu < 1/(2K-1)$	Deterministic
2010	$\delta_{K+1} < 1/(3\sqrt{K})$	Deterministic
2012	$\delta_{K+1} < 1/(\sqrt{K}+1)$	Deterministic
This paper	$M=\Omega(K\log(N))$	Probabilistic

2007: J. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. On. Inform. Theory
2010: M. A. Davenport, M. B. Wakin, "Analysis of Orthogonal Matching Pursuit Using the Restricted Isometry Property", IEEE Trans. On. Inform. Theory
2012: J. Wang and B. Shim, "On the recovery Limit of Sparse Signals Using Orthogonal Matching Pursuit", IEEE Trans. Signal Processing Letter

The short overview of the paper [2012]

- To derive their sufficient condition, the authors considered the event that OMP correctly selects index j at the i^{th} iteration.
- The event occurs if $\min_{t\in\mathcal{I}} \left\|\langle \mathbf{a}_t, \mathbf{y}\rangle\right\|_2 > \max_{t\notin\mathcal{I}} \left\|\langle \mathbf{a}_t, \mathbf{y}\rangle\right\|_2$.
- They have shown that the left term is lower bounded by

$$\min_{t\in\mathcal{I}} \left\|\langle \mathbf{a}_t, \mathbf{y}\rangle\right\|_2 \geq \frac{1}{\sqrt{K}}(1-\delta_K)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2.$$

- Also, they have shown that the right term is upper bounded by

$$\max_{t\notin\mathcal{I}} \left\|\langle \mathbf{a}_t, \mathbf{y}\rangle\right\|_2 \leq (1-\delta_{K+1})\left\|\mathbf{x}_{\mathcal{I}}\right\|_2.$$

- Then, they have derived their sufficient condition from the two bounds.

The main Theorem

- (OMP with Admissible Measurement matrix.) Fix $\delta \in (0,1)$, and choose $M = \Omega(Klog(N/\delta))$. Suppose that $\mathbf{x}$ is an arbitrary $K-$sparse vector in $\mathcal{R}^N$, and draw a random $M \times N$ admissible measurement matrix $\mathbf{A}$ independent from the vector. Given the measurement vector $\mathbf{y} = \mathbf{Ax}$. Then, OMP can reconstruct the support set with probability exceeding $1-\delta$.

Admissible Measurement Matrices

- An admissible measurement matrix for K −sparse vectors in $\mathcal{R}^N$ is an $M \times N$ random matrix $\mathbf{A}$ with four properties.

(M0) Independence : The columns of $\mathbf{A}$ are stochastically independent.

(M1) Normalization : $\mathbb{E}\left[\left\|\mathbf{a}_i\right\|_2^2\right] = 1 \text{ for } j = 1, \cdots, N.$

(M2) Joint correlation : Let $\{\mathbf{u}^t\}$ be a sequence of vectors whose l_2 norms do not exceed one. Let $\mathbf{a}$ be a column of $\mathbf{A}$ that is independent from $\{\mathbf{u}^t\}$. Then,

$$\mathbb{P}\left\{\max_t \left|\left\langle \mathbf{a}, \mathbf{u}^t \right\rangle\right| \le \varepsilon\right\} \ge 1 - 2K\exp\left(-c\varepsilon^2 M\right)$$

(M3) Smallest singular value: Given an $M \times K$ submatrix $\boldsymbol{Z}$ from $\mathbf{A}$, the largest singular value $\sigma_{min}(\mathbf{Z})$ satisfies $\mathbb{P}\left\{\sigma_{\min}\left(\mathbf{Z}\right) \ge 0.5\right\} \ge 1 - \exp\left(-cM\right)$

The proof of the main Theorem-1

- First, let us define the greedy ratio at the l^{th} iteration:

$$\rho\left(\mathbf{r}^l\right) := \frac{\max_{i \notin \mathcal{I}} \left|\left\langle \mathbf{r}^l, \mathbf{a}_i \right\rangle\right|}{\max_{i \in \mathcal{I}} \left|\left\langle \mathbf{r}^l, \mathbf{a}_i \right\rangle\right|}$$

- OMP correctly selects an index belonging to the support set if $\rho\left(\mathbf{r}^l\right) < 1.$

- OMP correctly reconstructs the support set when the event $E_{succ} := \max_{l \le K} \rho\left(\mathbf{r}^l\right) < 1$ occurs

- We aim to obtain the probability

$$\mathbb{P}\left\{E_{succ}\right\} := \mathbb{P}\left\{\max_{l \le K} \rho\left(\mathbf{r}^l\right) < 1\right\}$$

$$\ge \mathbb{P}\left\{\max_{l \le K} \rho\left(\mathbf{r}^l\right) < 1 \cap \sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right) \ge 0.5\right\}$$

- Owing to (M3), we can solve LS within the K^{th} iterations.

The proof of the main Theorem-2

- Continuously, we aim to consider the probability

$$\mathbb{P}\left\{\max_{l\le K}\rho\left(\mathbf{r}^l\right)<1\middle|\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5\right\}$$

- For this end, we consider the greedy ratio at the l^{th} iteration. Then, we have

$$\rho\left(\mathbf{r}^l\right)=\frac{\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{r}^l,\mathbf{a}_i\right\rangle\right|}{\max_{i\in\mathcal{I}}\left|\left\langle\mathbf{r}^l,\mathbf{a}_i\right\rangle\right|}=\frac{\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{r}^l,\mathbf{a}_i\right\rangle\right|}{\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_\infty}\le\frac{\sqrt{K}\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{r}^l,\mathbf{a}_i\right\rangle\right|}{\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2}$$

- Now, we simplify the upper bound of the greedy ratio. First, let us define $\mathbf{r}^l:=\mathbf{u}^l\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2/0.5$. Then, the upper bound becomes

$$\frac{\sqrt{K}\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{r}^l,\mathbf{a}_i\right\rangle\right|}{\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2}=\frac{\sqrt{K}\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{u}^l\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2/0.5,\mathbf{a}_i\right\rangle\right|}{\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2}$$

$$=2\sqrt{K}\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{u}^l,\mathbf{a}_i\right\rangle\right|.$$

The proof of the main Theorem-3

- Owing to M3, we have $\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2/\left\|\mathbf{r}^l\right\|_2\ge\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5.$

- Then, we can show that the l^2 norm of the vector $\mathbf{u}^l$ is always less than one.

$$\mathbf{u}^l=0.5\mathbf{r}^l/\left\|\mathbf{A}_{\mathcal{I}}^T\mathbf{r}^l\right\|_2\le\mathbf{r}^l/\left\|\mathbf{r}^l\right\|_2$$

- Now, we have

$$\mathbb{P}\left\{\max_{l\le K}\rho\left(\mathbf{r}^l\right)<1\middle|\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5\right\}\ge\mathbb{P}\left\{\max_{l\le K}2\sqrt{K}\max_{i\notin\mathcal{I}}\left|\left\langle\mathbf{u}^l,\mathbf{a}_i\right\rangle\right|<1\middle|\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5\right\}$$

$$=\mathbb{P}\left\{\max_{i\notin\mathcal{I}}\max_{l\le K}\left|\left\langle\mathbf{u}^l,\mathbf{a}_i\right\rangle\right|<\frac{1}{2\sqrt{K}}\middle|\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5\right\}$$

$$\ge\prod_{i\notin\mathcal{I}}\mathbb{P}\left\{\max_{l\le K}\left|\left\langle\mathbf{u}^l,\mathbf{a}_i\right\rangle\right|<\frac{1}{2\sqrt{K}}\middle|\sigma_{\min}\left(\mathbf{A}_{\mathcal{I}}\right)\ge 0.5\right\}$$

$$\ge\left[1-2K\exp\left(-cM/(4K)\right)\right]^{N-K}$$

The proof of the main Theorem-4

- In addition, we have $\mathbb{P}\{\sigma_{\min}(\mathbf{Z}) \geq 0.5\} \geq 1-\exp(-cM)$.
- Thus, we finally obtain

$$\mathbb{P}\{E_{succ}\} \geq \left[1-2K\exp\left(-cM/(4K)\right)\right]^{N-K}\left[1-\exp(-cM)\right].$$

- To simplify the lower bound, we apply the inequality $(1-x)^n \geq 1-kn$ for $n \geq 1$ and $x \leq 1$. Then, for $K(N-K) \leq N^2/4$, we have

$$\mathbb{P}\{E_{succ}\} \geq 1-2K(N-K)\exp\left(-cM/(4K)\right)-\exp(-cM).$$

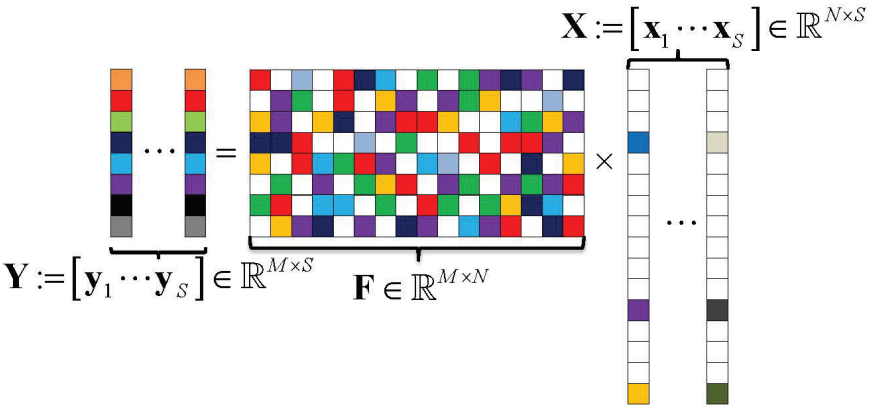
- By again simplifying the above lower bound, we have

$$\mathbb{P}\{E_{succ}\} \geq 1-N^2\exp(-cM/K).$$

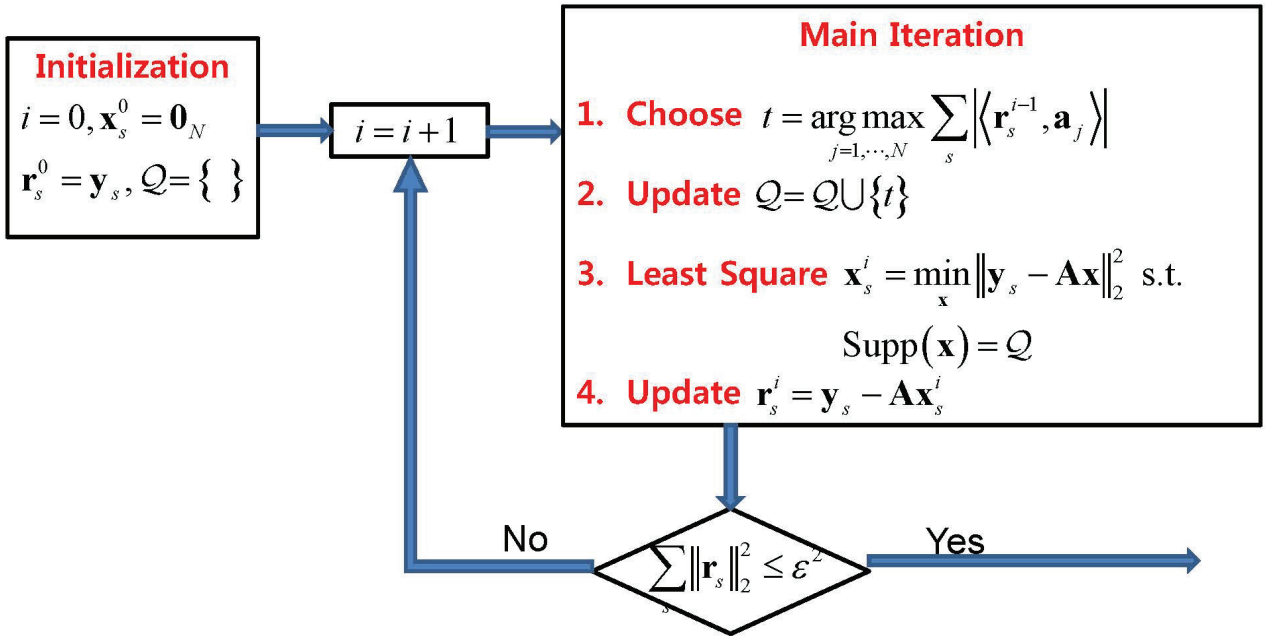
- Finally, we can see that the choice $M = \Omega(Klog(N/\delta))$ is sufficient to reduce the failure probability below δ.

New researches problems

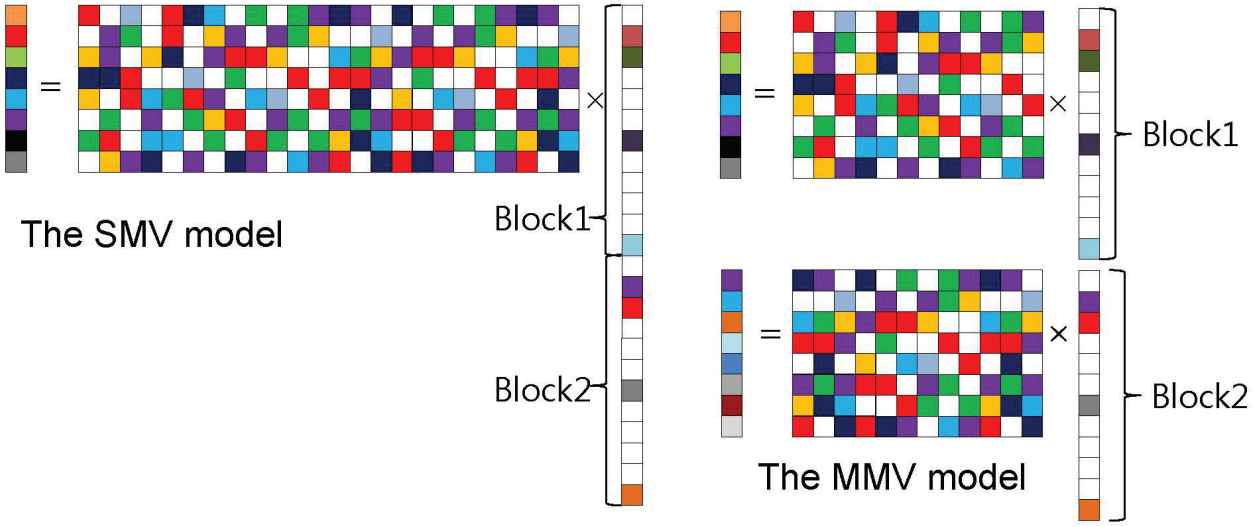
1. Can we establish a sufficient condition for Simultaneously Orthogonal Matching Pursuit?



Simultaneously Orthogonal Matching Pursuit



New researches problems



2. Let M_1 be the number of measurements in the SMV model when OMP is exploited. Let M_2 be the total number of measurements in the MMV model when SOMP is exploited. What is the relation between M_1 and M_2?

Link Status Monitoring Using Network Coding

M. H. Firooz et al.

To appear IEEE/ACM Trans. on Networking

Presenter : Jin-Taek Seong

GIST, Dept. of Information and Communications, INFONET Lab.

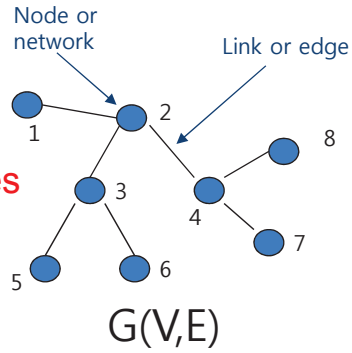
Gwangju Institute of Science and Technology

Outline

- **Network Tomography**
  - Introduction (Network Monitoring)
  - Approaches:
    - Deterministic vs. Stochastic
    - Active vs Passive
  - Challenges: Overhead, Identifiability
- **Network Coding**
  - Applications to network monitoring: new method
  - Optimization : speed/complexity tradeoffs

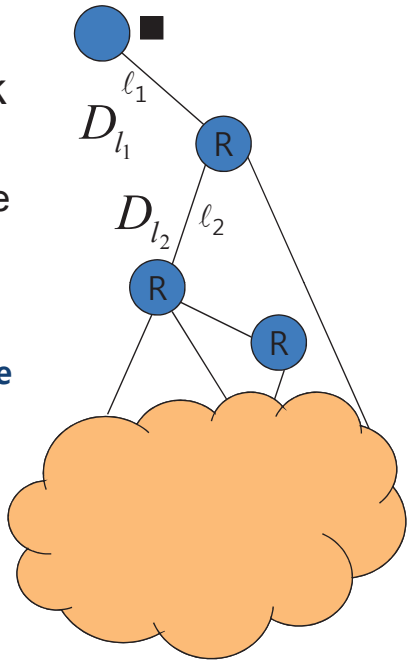
Network Tomography

- Networks: **set of nodes, links modeled as graph G(V,E)**
- Network monitoring
  - Involves collection of network performance statistics (link delay, link loss or failure status)
  - Important for QoS guarantees (media streaming, interactive video applications)
- Challenges
  - Choice of appropriate measurement techniques and algorithms



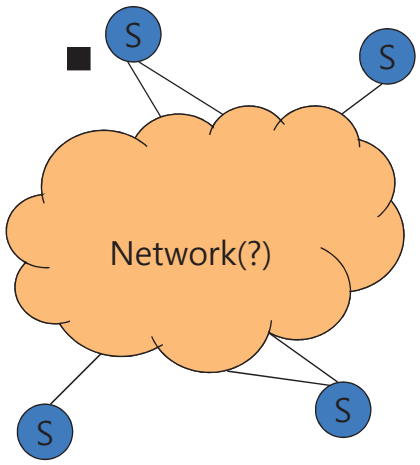
Measurement Methods

- **Node-oriented**: These methods are based on cooperation among network nodes, e.g., ping or traceroute
  - Using Ping, round trip delay to every node can be measured.
  - Uses Internet control message protocol (ICMP) packets
    - **Many routers do NOT respond to these packets**
  - Many service providers do not own the entire network



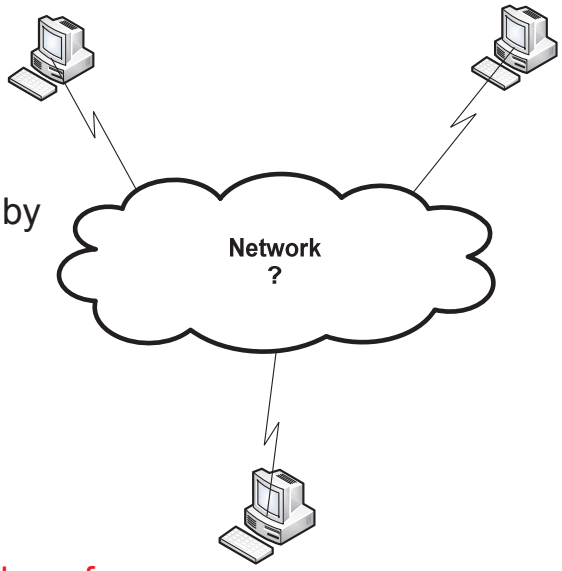
Measurement Methods

- **Edge-oriented**: Access is available to all nodes at the edge only (and not to any in the interior)
  - Does not require exchanging special control messages between interior nodes
  - ***Inverse problem***: estimate *link level* status from end-to-end (path level) measurements

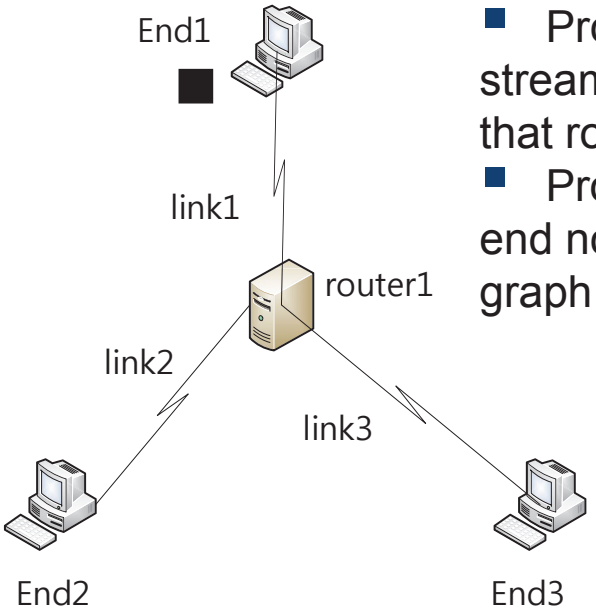


Measurement Methods

- **Active** (sending probe packets)
  - Adds overhead to normal data traffic by introducing new control packets
- **Passive** (insitu traffic analysis)
  - No overhead; temporal and spatial dependence might bias measurement
- Considered method: *edge-oriented, active* network tomography
  - Given a network, and a limited number of end hosts, **when** can we infer failure status of the links?



End-to-End Probing



- Probes are inserted into a data stream, and end-to-end properties on that route measured.
- Probes are exchanged between end nodes using routing matrix of the graph

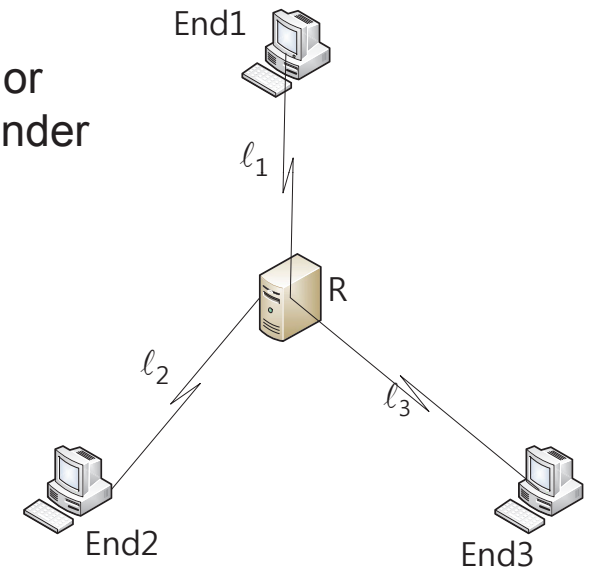
Routing matrix A

	$link1$	$link2$	$link3$
$End1 \rightarrow End2$	1	1	0
$End1 \rightarrow End3$	1	0	1
$End2 \rightarrow End3$	0	1	1

End-to-End Probes

- Routing matrix relates link attribute to route attribute
- For some parameters like delay or path loss, this relation is linear under some assumptions

$$\begin{bmatrix} D_{End1\rightarrow End2} \\ D_{End1\rightarrow End3} \\ D_{End2\rightarrow End3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} D_{l_1} \\ D_{l_2} \\ D_{l_3} \end{bmatrix}$$



Deterministic

- Link attributes (e.g. delay) are considered unknown, constant
- Goal: estimate constants
- Link attributes are typically time varying

→ method is suitable for periods of local '*stationarity*'

Stochastic

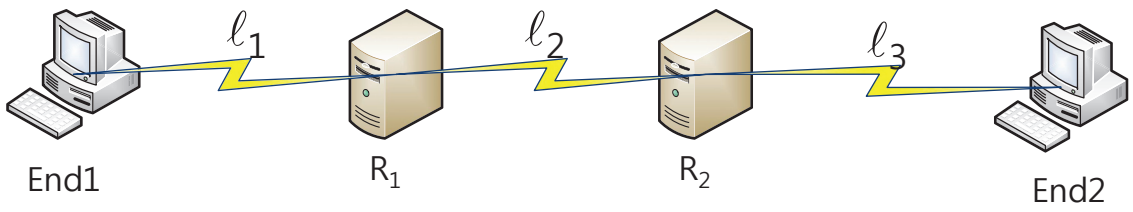
- Link attribute specified by a suitable probability distribution
  - e.g. link delay follows a Gaussian distribution
- Estimation problem: unknown model parameters

based on path observation in the presence of additive noise

Deterministic vs. Stochastic Methods

- Stochastic
  - Bayesian - requires a prior distribution
    - incorrect choice leads to biases in the estimates
  - More computationally intensive
- Deterministic
  - Lower complexity but suffers from generic *identifiability* (will be discussed later) problems

Link Failure Model

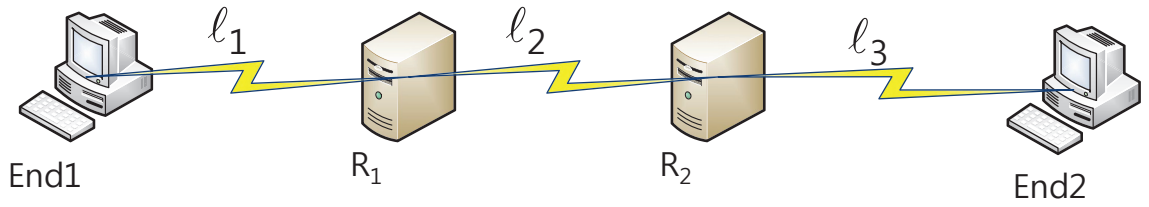


Define an indicator function for status of each link

$$x_{l_i} = \begin{cases} 0 & l_i \text{ is ok} \\ 1 & l_i \text{ is congested} \end{cases}$$

$$y_{end1 \to end2} = \begin{cases} 0 & \text{all of } l_1, l_2, l_3 \text{ is ok} \\ 1 & o.w. \end{cases}$$

Binary Deterministic Model



$$y_{end1 \to end2} = x_{l_1} \; or \; x_{l_2} \; or \; x_{l_3}$$

$$y = Ax$$

A: N-by-M binary routing matrix

x: M-by-1 binary vector, the status of each link

y: N-by-1 binary vector, the status of each path (measurements)

Failure Monitoring

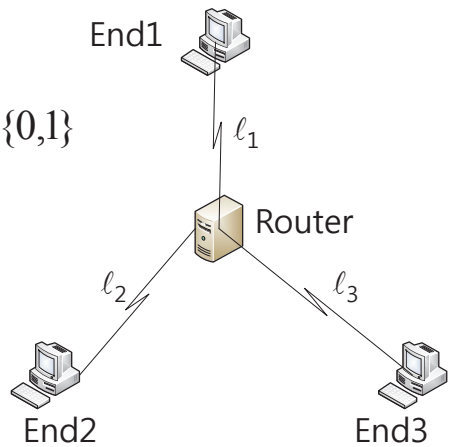
- Network G(V,E) with set of paths *P*
- **x, y** are binary vectors
- A path is congested if at least one of its links is congested

$$\mathbf{x} \in \{0,1\}^{|\mathbf{E}|}, \mathbf{y} \in \{0,1\}^{|\mathbf{P}|}$$

$$\begin{matrix} & & l_1 & l_2 & l_3 \end{matrix}$$

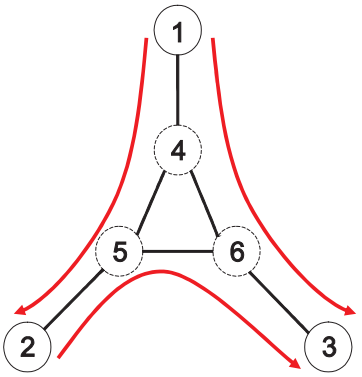
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{matrix} End1 \to End2 \\ End1 \to End3 \\ End2 \to End3 \end{matrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{l_1} \\ x_{l_2} \\ x_{l_3} \end{bmatrix}, \quad x_{l_1} \in \{0,1\}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_{l_1}(OR)x_{l_2} \\ x_{l_1}(OR)x_{l_3} \\ x_{l_2}(OR)x_{l_3} \end{bmatrix}$$



Identifiability y = Ax

- **Problem:** Estimate **x** from **y** with
  - **A *(N-by-M)*** : binary routing matrix
  - **x *(M-by-1)*** : binary link failure status
  - **y *(N-by-1)*** : end-to-end measurements

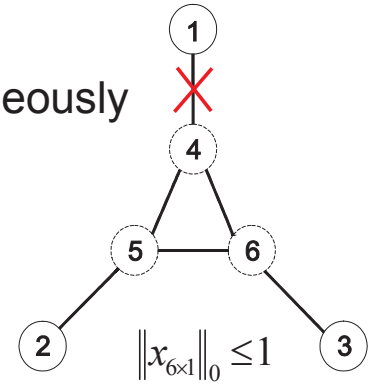


6 links, 3 End-to-End routes ➔ M=6, N=3

- **Identifiability**: a network is identifiable if **y = Ax** has a ***unique solution***
  - Usually, *M* (# of links in network) >> *N* (# of measurements), so network is generically NOT identifiable.

Identifiability: Binary Model

- Solution: limit (maximum) number of failed links inside the network
  - Suppose at most *k* links can fail simultaneously



Only one link can be congested

- Definition: *k-Identifiability*
  - Network is k-identifiable if

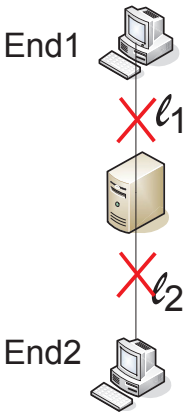
$$\left\|\mathbf{x}_{|E|\times1}\right\|_0 \le k$$

$$\forall \mathbf{x}_1, \mathbf{x}_2 \text{ s.t.} \left\|\mathbf{x}_1\right\|_0 \le k, \left\|\mathbf{x}_2\right\|_0 \le k,\ \mathbf{x}_1 \ne \mathbf{x}_2 \Rightarrow \mathbf{A}\mathbf{x}_1 \ne \mathbf{A}\mathbf{x}_2$$

- From end-to-end observation it is possible to uniquely identify up to ***k*** congested links

1-Identifiability

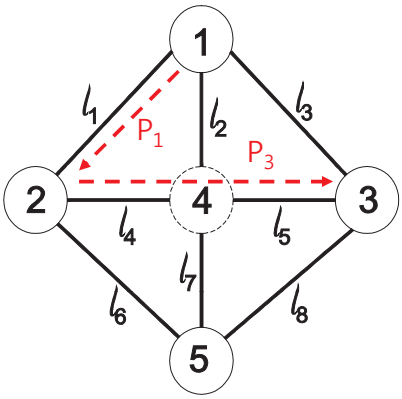
- *A network with an intermediate degree two node is **not** 1-identifiable*
  - ✓ If path End1→End2 is congested, it is impossible to determine which link among l_1 and l_2 is congested .
- Necessary but not sufficient!



$$x_{l_1} = 1 \Rightarrow y_{End1 \to End2} = 1$$
$$x_{l_2} = 1 \Rightarrow y_{End1 \to End2} = 1$$

k=1 Identifiability

- **1-identifiability Theorem**: End-to-End probe based measurements can detect a unique congested link in a network *if and only if* there are ***no two identical columns*** in the network routing matrix



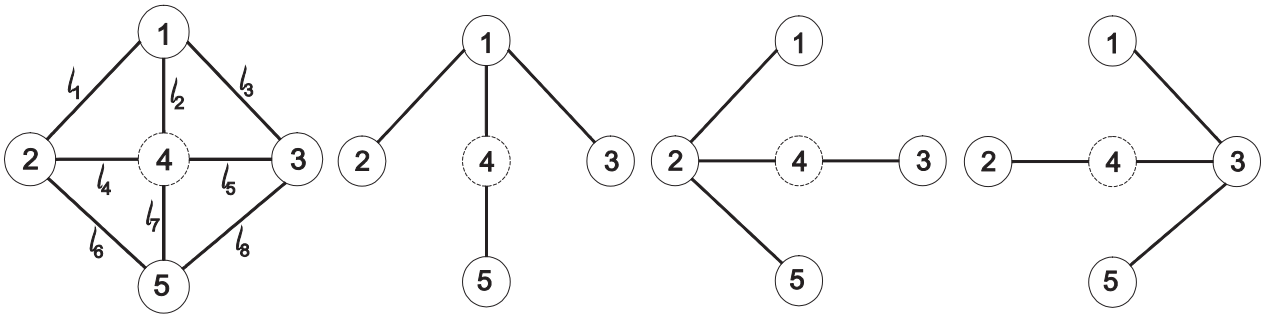
$$\begin{array}{c} P_1 \\ \\ \\ P_3 \\ \\ \\ \end{array} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

k- identifiability

- **k-identifiability Theorem**: End-to-End probe based measurements can detect a unique congested link in a network *only if* there are ***no k+1 dependent columns*** in the network routing matrix

Shortest Path Routing Revisited

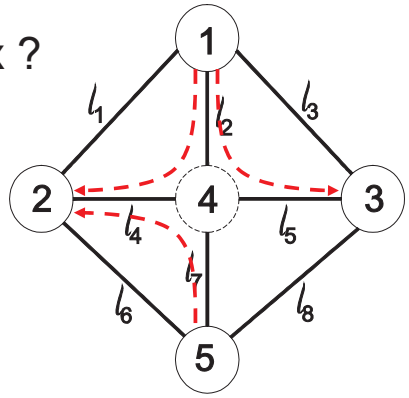
- Packets are sent on shortest path between two end nodes
  - sub-graphs = tree starting from a boundary (source) node
    - ➢ Node 4 has two degrees in all graphs
- But node 4 has 4 degrees in the original network

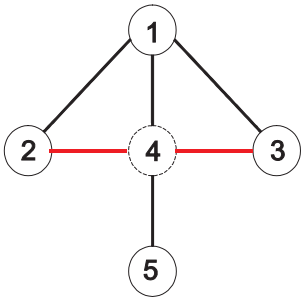


Revisiting Shortest Path Routing

- What if we could change routing matrix ?

  Example: in place of shortest path routing, route packets through longer paths, e.g. $n_1 \rightarrow \ell_2 \rightarrow \ell_4 \rightarrow n_2$
  - Now network is 1-identifiable !
- Intrinsic limitation for end-to-end measurement methods based on shortest path routes
  - probes transmitted along such paths contain only *minimum information*





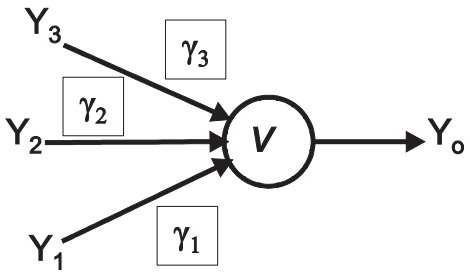
Solution

- Look to exchange probes between boundary nodes via other (non-shortest) paths?
- Changing the routing tables violates tomography assumption
- Use **Network Coding**; exploit broadcast nature of network coding, a transmitted probe will traverse almost every path between two boundary nodes

Linear Network Coding

- Network Coding is a coding at layer three
- The coding is conducted over the finite field F_u, $u=2^q$
- Each coded symbol can be represented by *q*-bits within an IP layer frame
- Signal ***Y***(j) on an outgoing link *j* of node *v* is a linear combination of signals ***Y***(i) on incoming link *i* of *v*:
  - We assume there is no process generated at node *v*

$$Y(j) = \sum_{\{l:d(l)=v\}} \gamma_l Y(l)$$



Received Symbols

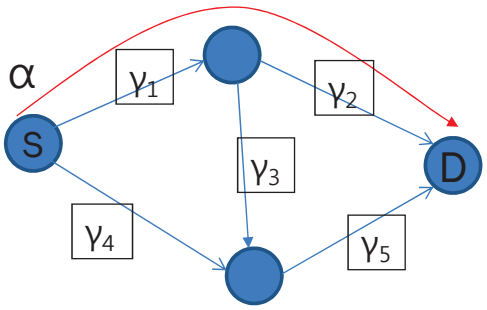
- $\boldsymbol{P^i}$: *i*-th route from source to destination
- Source sends α over $\boldsymbol{P^i}$

$$y = \alpha \prod_{l \in P^i} \gamma_l = \alpha \beta_i(G), \quad \alpha \in F_{2^q}$$

$$\beta_i(G) = \prod_{l \in P^i} \gamma_l \quad \text{Path NC Coef.}$$

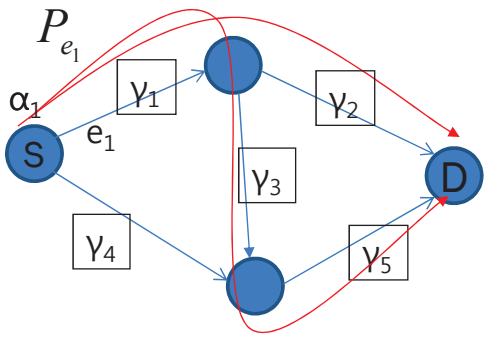
- β_i depends on topology *G* hence $\beta_i(G)$

$$y = \alpha\gamma_1\gamma_2 = \alpha\beta_1(G)$$



Received Symbols: Linear Model

- e_k one of source outgoing links
- P_{ek} : collection of all paths between source and destination starts at the k-th outgoing edge e_k
- Source sends α_k over e_k. By superposition destination receives



$$y = \alpha_k \sum_{P^i \in P_{e_k}} \prod_{l \in P^i} \gamma_l = \alpha_k \sum_{i=1}^{|P_{e_k}|} \beta_{i,e_k}(G)$$

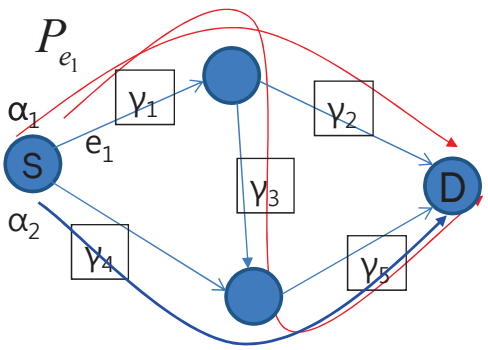
$$y = \alpha_1(\gamma_1\gamma_2 + \gamma_1\gamma_3\gamma_5) = \alpha_1(\beta_{1,e_1} + \beta_{2,e_1})$$

Received Symbols: Linear Model

- Source sends out symbols α_k over e_k using superposition once more

$$y = \sum_{k=1}^{K} \alpha_k \sum_{i=1}^{|P_{e_k}|} \beta_{i,e_k}(G)$$

- In vector format: $y=\boldsymbol{\alpha}^t\boldsymbol{\beta}(G)$
- $\boldsymbol{\beta}(G)$ is total network coding vector



$$y = \alpha_1(\gamma_1\gamma_2 + \gamma_1\gamma_3\gamma_5) + \alpha_2\gamma_4\gamma_5$$

Received Symbols: Linear Model

- Source sends symbols in *M* consecutive time slots:

$$y_{M\times 1} = A_{M\times N}\beta(G)_{N\times 1}$$

$$\beta(G)_{N\times 1} = \Big[\underbrace{\beta_{1,e_1} \quad \beta_{2,e_1} \quad \cdots \quad \beta_{N_1,e_1}}_{P_{e_1}} \Big| \underbrace{\beta_{1,e_1} \quad \cdots \quad \beta_{N_2,e_2}}_{P_{e_2}} \Big| \cdots \Big| \underbrace{\beta_{N_K,e_K}}_{P_{e_K}}\Big]^t$$

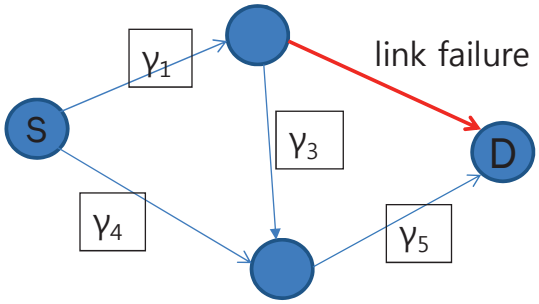
$$N = \Big|\bigcup_{i=1}^{K} \mathcal{P}_{e_i}\Big|$$

$$A_{M\times N} = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,|\mathcal{P}_{e1}|} & \cdots & \alpha_{1,N} \\ \alpha_{2,1} & \cdots & \alpha_{2,|\mathcal{P}_{e1}|} & \cdots & \alpha_{2,N} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \alpha_{M,1} & \cdots & \alpha_{M,|\mathcal{P}_{e1}|} & \cdots & \alpha_{M,N} \end{bmatrix}$$

A: consisting *K* distinct columns

Link Failure Model

- If a link is severely congested, packets are significantly delayed and assumed lost at the destination
- We model the network with link *l* in congestion state by its *edge deleted subgraph* denoted by $G_l(V,E_l)$



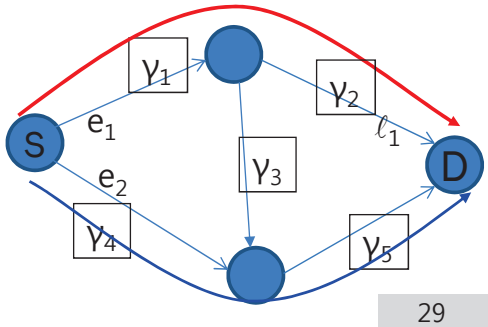
Link Failure Model

- Total network coding vector of $G_l(V;E_l)$, $\beta(G_l)$ is different from $\beta(G)$

$$\beta_{i,e_k}(G_l) = \begin{cases} \beta_{i,e_k}(G) & if\ l \notin P^i_{e_k}(d) \\ 0 & o.w. \end{cases}$$

- If the congested link doesn't belong to *i*-th path from source to destination, P^i, it will not affect packets going through those paths
  - It is zero otherwise

$$\beta_1(G) = \gamma_1\gamma_2 \quad \beta_2(G) = \gamma_4\gamma_5 \quad \Rightarrow \quad \beta_1(G_{l_1}) = 0 \quad \beta_2(G_{l_1}) = \beta_2(G)$$



Link Failure Model

- Training sequence is **A**
- $\mathbf{y}^l$: vector of symbols observed at the destination in *M* time slots with link l congested

$$y^l_{M\times 1} = A_{M\times N}\beta(G_l)_{N\times 1}$$

- Potential for identifying: received symbols change uniquely in response to link congestion

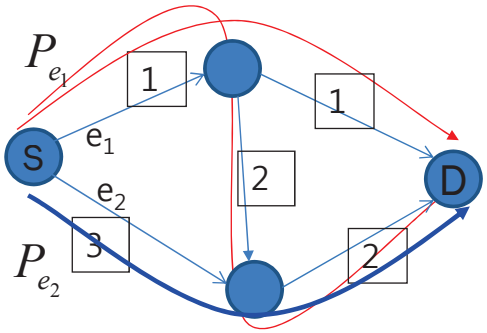
$$y_{M\times 1} \neq y^l_{M\times 1}$$
$$y^{l_1}_{M\times 1} \neq y^{l_2}_{M\times 1}$$

Example

$$\beta(G) = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

$$\beta_{1,e_1} = 1 \times 1 = 1$$
$$\beta_{2,e_1} = 1 \times 2 \times 2 = 3$$
$$\beta_{2,e_2} = 3 \times 2 = 1$$

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 3 & 3 \end{bmatrix}$$



	--	e_1	e_2	ℓ_1	ℓ_2	ℓ_3
1st time slot	0	2	2	3	1	1
2nd time slot	2	3	1	0	1	3

Received symbols corresponding a single link failure

Theorem 2: Sufficient Conditions

- If Rank(**A**)= deg(S), and
  - for all P_{ek} set of paths between source and destination starting at e_k

$$\sum_{j=1}^{|P_{e_k}|} \xi_j \beta_{j,e_i} = 0 \Leftrightarrow \xi_j = 0 \, \forall j$$ (more next slide)

then

$$A\beta(G) \neq A\beta(G_l) \quad \forall l \notin E$$
$$A\beta(G_{l_1}) \neq A\beta(G_{l_2}) \quad \forall l_1, l_2 \notin E$$

Theorem 2

- Condition $\sum_{j=1}^{|P_{ek}|} \xi_j \beta_{j,e_i} = 0 \Leftrightarrow \xi_j = 0 \; \forall j$ means
  - For a set of paths having e_k in common, P_{ek} , NC coefficient of the paths are independent !

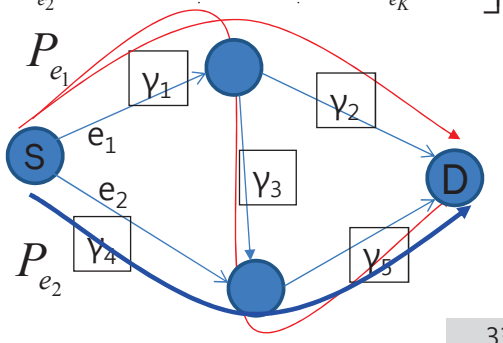
$$\left[\underbrace{\beta_{1,e_1} \quad \beta_{2,e_1} \quad \cdots \quad \beta_{N_1,e_1}}_{P_{e_1}} \;\middle|\; \underbrace{\beta_{1,e_1} \quad \cdots \quad \beta_{N_2,e_2}}_{P_{e_2}} \;\middle|\; \cdots \;\middle|\; \underbrace{\beta_{N_K,e_K}}_{P_{e_K}} \right]^t$$

Independent

Independent

$$\left.\begin{aligned} \beta_{1,e_1} &= \gamma_1\gamma_2 \\ \beta_{2,e_1} &= \gamma_1\gamma_3\gamma_5 \end{aligned}\right\} \text{independent}$$

$$\beta_{1,e_2} = \gamma_4\gamma_5$$



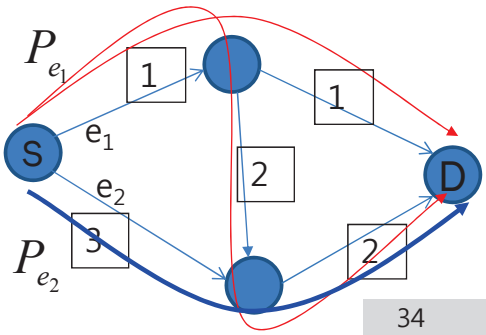
Example

❖ $\beta(G) = \begin{bmatrix} 1 \\ 3 \\ \hline 1 \end{bmatrix}$ Independent

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 3 & 3 \end{bmatrix} \quad Rank(A) = 2 = \deg(S)$$

$$\beta_{1,e_1} = 1 \times 1 = 1$$
$$\beta_{2,e_1} = 1 \times 2 \times 2 = 3$$
$$\beta_{2,e_2} = 3 \times 2 = 1$$

	--	e_1	e_2	ℓ_1	ℓ_2	ℓ_3
1st time slot	0	2	2	3	1	1
2nd time slot	2	3	1	0	1	3



Complexity/Speed

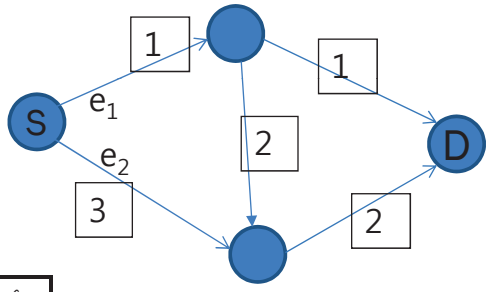
- First condition of Theorem 2:

$$\text{Rank}(A_{M\times N}) = \deg(S) \;\; \text{implies} \;\; M \geq \deg(S)$$

  - In previous example M=2=deg(S)
- Number of time slots: at least the number of outgoing links of source
- Is it possible to decrease number of time slots? ➔ faster monitoring
- Possible by increasing number of bits in LNC coeff. ➔ more complexity

Example

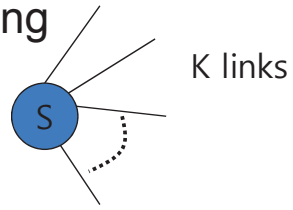
- q=3
- A=[1 1 4]



	--	e_1	e_2	ℓ_1	ℓ_2	ℓ_3
1st time slot	6	4	2	5	7	1

Theorem 3: Complexity/Speed tradeoff

- $N_i=|P^i|$
- *q* bits per symbol are used in network coding
- *M* number of (desired) time slots
- Let Z={1,2,…,K}
- *K* degree of source
- $\boldsymbol{Z}_M$: collection of all partitions of *Z* with size *M*



$$Z_M = \{\{H_1, H_2, ..., H_M\} \mid \bigcup_{i=1}^{M} H_i = Z, H_i \cap H_j = \Phi\}$$

- *K*=3, *M*=2 ➔ *Z*={1,2,3}
- Z_M={ {{1,2},{3}} , {{1,3},{2}} , {{2,3},{3}} }

Theorem 3: Complexity/speed tradeoff

- Network is 1-identifiable if

$$q \geq \min_{\{H_i, i=1,...,M\} \in Z_M} \max_i \sum_{j \in H_i} N_j$$

Rank(**A**)=M

Theorem 3 provides a tradeoff between number of time slots for training sequence (speed of the method) and size of network coding coefficient (complexity) to make a network *G(V; E)* identifiable.

CS Journal Club, July 11, 2013

A sparse signal reconstruction perspective for source localization with sensor arrays

Authors: Dmitry Malioutov, Mujdat Cetin, and Alan S. Willsky
Journal: IEEE Trans. on. Sign. Proc. Aug., 2005.
Presenter: J. Oliver

Abstract

In this paper, the authors present a source localization method based on sparse representation of sensor measurements. In particular, they use SVD of the data matrix obtained from the sensors to summarize the multiple measurements. The SVD summarized data is then sparsely represented in order to detect the sources. The authors also proposed grid refinement in order to mitigate the effects of limiting estimates to a grid of spatial locations. They demonstrate the superior resolution ability with limited time samples of their method over the existing methods via various experiments.

Introduction and Background

- Source localization methods deal with finding the closely spaced sources in presence of considerable noise.
- Many advanced techniques for the localization of sources achieve super-resolution by exploiting the presence of a small number of sources. For example, the key component of the MUSIC method is the assumption of a low-dimensional signal subspace.
- Estimating the spatial locations (or directions) is a well-known problem in array signal processing.
- Three major source estimation techniques are 1. Classical methods (beamformer, MVDR) 2. Subspace methods (MUSIC, ESPRIT) 3. ML-based methods (deterministic and stochastic).
- Beamforming is simple but its resolution is limited. Subspace methods achieve super resolution, provided SNR is moderately high and sources are not strongly correlated and the number of snapshots (measurement vectors) are sufficient. ML techniques are superior than the subspace methods but require accurate initialization for global convergence.
- By turning to the sparse signal representation framework, the authors are able to achieve super-resolution without the need for a good initialization, without a large number of time samples, and with lower sensitivity to SNR and to correlation of the sources.
- The authors have developed the method for narrowband case and discussed in brief how it can be used for wideband source localization.
- Prior research has established sparse signal representation as a valuable tool for signal processing, but its application to source localization has been developed only for very limited scenarios. For example, [1, 2] is concerned with source localization in the beam-space domain, under the assumption that the sources are uncorrelated, and that a large number of time samples is available.

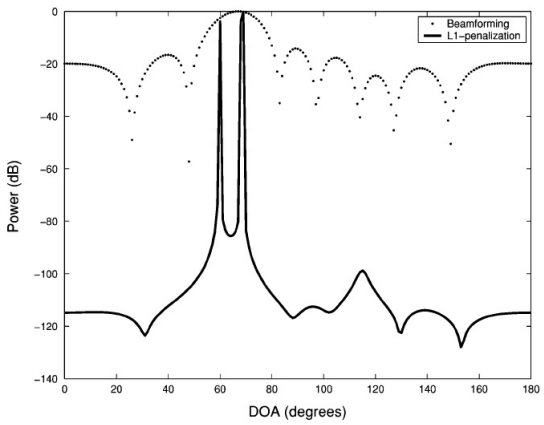


Fig. 1. Single sample source localization with ℓ_1. Spatial spectra of two sources with DOAs of 60° and 70° (SNR = 20 dB).

- In its most basic form, the problem of sparse signal representation in overcomplete bases asks to find the sparsest signal x to satisfy $y = Ax$, where $A \in C^{M\times N}$ is an overcomplete basis, i.e., $M < N$.Without the sparsity prior on x, the problem $y = Ax$ is ill-posed and has infinitely many solutions. Additional information that x should be sufficiently sparse allows one to get rid of the ill-posedness.

Source localization framework

- The goal of the source localization is to find locations of sources of wavefields that impinge on an array of sensors that are seperated by a distance less than or equal to $\lambda/2$
- Consider K narrowband signals $u_k(t), k \in \{1, 2, \cdots, K\}$, arriving at an array of M sensors, after being corrupted by additive noise $n_m(t)$, resulting in sensor outputs $y_m(t), m \in \{1, 2, \cdots, M\}$. After demodulation, the vector form of the received signal is

$$\boldsymbol{y}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{u}(t) + \boldsymbol{n}(t), \quad t \in \{t_1, \cdots, t_T\} \tag{1}$$

- $\boldsymbol{A}(\boldsymbol{\theta})$ is array manifold matrix. The $(m, k)^{th}$ element $\boldsymbol{A}$ contains the delay and gain information from the kth source (at location θ_k) to the mth sensor. The column, $a(\theta_k)$, of $\boldsymbol{A}$ are called steering vectors and is given by $a(\theta_k) = \left[e^{j\frac{2\pi}{\lambda}1\sin\theta_k}, e^{j\frac{2\pi}{\lambda}2\sin\theta_k}, \cdots, e^{j\frac{2\pi}{\lambda}M\sin\theta_k}\right]^T$
- Any source localization method aims to find the unknown locations of the sources $\theta_k, \forall k$, given $\boldsymbol{y}(t)$ and $\boldsymbol{A}$.
- We note that finding $\boldsymbol{\theta}$ is a non-linear estimation problem.

Sparse representation for a single time sample, that is, $T = 1$

- To cast a sparse representation problem, the authors introduce an overcomplete representation of $\boldsymbol{A}$ in terms of all possible source locations.
- Let $\{\tilde{\theta}_1, \tilde{\theta}_2, \cdots, \tilde{\theta}_N\}$ be a sampling grid of all source locations of interest.
- The number of potential sources N will typically be much greater than the number of actual sources K and the number of sensors M.

- A matrix composed of steering vectors corresponding to each potential source location as its columns constitute an over-complete dictionary, that is, $\boldsymbol{A} = \left[a(\tilde{\theta}_1), a(\tilde{\theta}_2), \cdots, a(\tilde{\theta}_N)\right]$. We note that $\boldsymbol{A}$ is known and does not depend on the actual source locations.

- The signal vector is $\boldsymbol{s}(t)$ with the nth element $s_n(t) = u_k(t)$ if the source k comes from θ_n for some k and zero otherwise. For $T = 1$, then the source localization problem reduces to

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{n} \tag{2}$$

- In effect, this overcomplete representation allows us to exchange the problem of parameter estimation of $\boldsymbol{\theta}$ for the problem of sparse spectrum estimation of $\boldsymbol{s}$.

- With the key assumption that the source numbers are less, the underlying spatial spectrum is sparse (i.e., has only a few nonzero elements), and hence we can solve this inverse problem via l_1 methodology, $min\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{s}\|_2^2 + \lambda\|\boldsymbol{s}\|_1$

- The data for the model is complex-valued; hence, neither linear nor quadratic programming can be used for numerical optimization. Instead, the authors adopt an SOC programming framework and find $\boldsymbol{s}$. Once $\boldsymbol{s}$ is found, the estimates of the source locations correspond to the locations of the peaks in $\boldsymbol{s}$.

Source location with multiple time samples and $l_1 - SVD$

- Source localization with multiple snapshots from potentially correlated sources is of greater practical importance.

- When we bring time into the picture, the overcomplete representation is easily extended and it has the following form:

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t), \quad t \in \{t_1, t_2, \cdots, t_T\} \tag{3}$$

Single and Joint inverse problem

- The first thought that comes to mind when we switch from one time sample to several time samples is to solve each problem indexed by separately. In that case, we would have a set of solutions $\hat{\boldsymbol{s}}(t)$.

- If the sources are moving fast, then the evolution of the sources is of interest, and the approach is suitable for displaying it.

- When the sources are stationary over several time samples, then it is preferable to combine the independent estimates to get one representative estimate of source locations from them, for example, by averaging or by clustering.

- Now, we consider a simple approach that uses different time samples together. Let $\boldsymbol{Y} = [\boldsymbol{y}(t_1), \boldsymbol{y}(t_2), \cdots, \boldsymbol{y}(t_T)]$, and define $\boldsymbol{S}$ and $\boldsymbol{N}$ similarly. Then, we have

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N} \tag{4}$$

- We note that the matrix $\boldsymbol{S}$ is parametrized temporally and spatially, but sparsity only has to be enforced in time not in space.

- To accommodate this issue in the optimization problem, the authors first compute the l_2 norm of all time-samples of a particular space index of s, that is, $\boldsymbol{s}_i^{l_2} = \|[s_i(t_1), s_i(t_2), \cdots, s_i(t_T)]\|_2$.

- Then the authors minimize the l_1 norm of $\boldsymbol{s}^{l_2} = \left[s_1^{l_2}, s_2^{l_2}, \cdots, s_N^{l_2}\right]$. Now the problem becomes

$$min\|\boldsymbol{Y} - \boldsymbol{AS}\|_f^2 + \lambda\|\boldsymbol{s}^{l_2}\|_1 \tag{5}$$

- Note in Eqn. (5), the optimization is performed over the matrix $\boldsymbol{S}$ and once the estimate of $\boldsymbol{S}$ is computed the peaks of $\boldsymbol{S}$ provide the source locations.
- The main drawback of this technique is its computational cost. The size of the inverse problem increases linearly with T, and the computational effort required to solve it increases superlinearly with T. In order to alleviate this, the authors propose a SVD based solution.

l_1- SVD

- To reduce both the computational complexity and the sensitivity to noise, the authors propose to use the SVD of the $M \times T$data matrix $\boldsymbol{Y}$.
- The idea is to decompose the data matrix into the signal and noise subspaces.
- With the signal subspace, mold the problem as multiple-vector sparse spectrum estimation problem similar to Eqn. (4).
- Without noise on the sensors, the set of vectors of $\boldsymbol{Y}$ would lie in a K-dimensional subspace.
- If we can relate the basis of this K-dimensional subspace (set of K vectors) to the source matrix $\boldsymbol{S}$, then we can just keep K vectors (instead of T) for the estimation problem.
- Take the SVD $\boldsymbol{Y} = \boldsymbol{ULV'}$ and form a $M \times K$ dimensional matrix $\boldsymbol{Y}_{sv}$ as $\boldsymbol{Y}_{sv} = \boldsymbol{YVD}_k$, where $\boldsymbol{D}_k$ is an $T \times K$ matrix given as $\boldsymbol{D}_k = [\boldsymbol{I}_K \boldsymbol{0'}]$
- Now $\boldsymbol{Y}_{sv}$ can be written as

$$\begin{aligned} \boldsymbol{Y}_{sv} &= \boldsymbol{YVD}_k \\ &= (\boldsymbol{AS} + \boldsymbol{N})\boldsymbol{VD}_k \\ &= \boldsymbol{ASVD}_k + \boldsymbol{NVD}_k \\ &= \boldsymbol{AS}_{sv} + \boldsymbol{N}_{sv} \end{aligned} \tag{6}$$

- We note that the sparsity structure of $\boldsymbol{S}$ is retained in $\boldsymbol{S}_{sv}$.
- Considering the k-th column of Eqn. (6) we have

$$\boldsymbol{y}^{sv}(k) = \boldsymbol{A}\boldsymbol{s}^{sv}(k) + \boldsymbol{n}^{sv}(k), \quad k = 1, 2, \cdots, K \tag{7}$$

  This is exactly the same form as multiple-vector model in Eqn. (3), expect that indexing is by singular vector, k.
- By bringing SVD, the problem size is reduced from T to K. This reduction is substantial, because in typical situations $K \ll T$.
- Now in the matrix $\boldsymbol{S}_{sv}$, the sparsity is along the spatial domain and not in the singular vector domain.
- To accommodate the true sparsity in the minimization problem, the authors define $\tilde{s}_i^{l_2} = \|[s_i^{sv}(1), s_i^{sv}(2), \cdots, s_i^{sv}(K)\|_2$. The sparsity of the $N \times 1$ vector $\tilde{s}_i^{l_2}$ is the sparsity of the spatial spectrum, which can be found by minimizing

$$\|\boldsymbol{Y}_{sv} - \boldsymbol{AS}_{sv}\|_f^2 + \lambda\|\tilde{\boldsymbol{s}}^{l_2}\|_1 \tag{8}$$

- In this paper, the authors have solved the above problem using SOC programming (see paper for details)

Multi-resolution grid refinement

- Thus far, in this paper, the estimates of the source locations are confined to a grid.
- We cannot make the grid very fine uniformly since this would increase the computational complexity and also the columns of $\boldsymbol{A}$ becomes more linearly dependent.
- Hence, the authors explore the idea of adaptively refining the grid in order to achieve better precision
- Instead of having a universally fine grid, we make the grid fine only around the regions where sources are present.
- This requires an approximate knowledge of the locations of the sources, which can be obtained by using a coarse grid first.

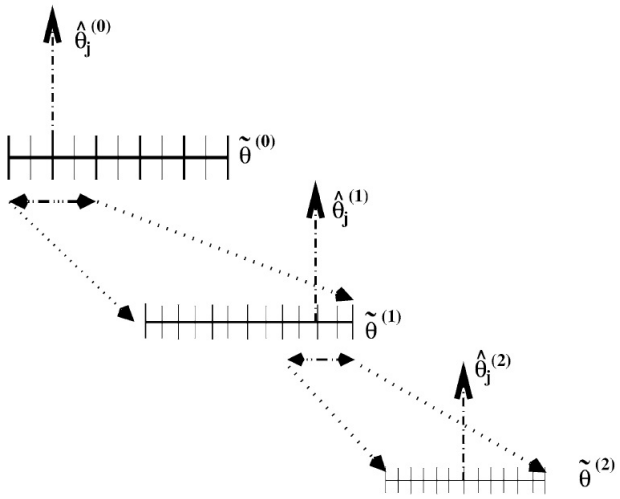


Fig. 3. Illustration of grid refinement.

- The grid refinement algorithm goes like this

1. Create a rough grid of potential source locations $\tilde{\theta}^{(0)}$, for $i = 1, 2, \cdots, N$. Set $r = 0$.
2. Form $\boldsymbol{A}_r = \boldsymbol{A}(\tilde{\boldsymbol{\theta}}^{(r)})$, where $\tilde{\boldsymbol{\theta}}^{(r)} = \left[\tilde{\theta}_1^{(r)}, \tilde{\theta}_2^{(r)}, \cdots, \tilde{\theta}_N^{(r)}\right]$. Use the SOC minimization to find the estimates of the source locations and set $r = r + 1$.
3. Get a refined grid $\tilde{\boldsymbol{\theta}}^{(r)}$ around the locations of the peak, $\hat{\theta}_j^{(r-1)}$ (explained below).
4. Return to step 2, until the grid is fine enough.

- There are many ways of refining the grid; the authors have chosen a simple equispaced grid refinement.
- Suppose at step r, we have a uniform grid with spacing δ_r. Also, we have an estimate $\hat{\theta}_j^{(r)}$

- Pick an interval around the jth detected source with two grid spacing on either side, that is, $[\hat{\theta}_j^{(r)} - 2\delta_r, \hat{\theta}_j^{(r)} + 2\delta_r]$, for $j = 1, 2, \cdots, K$.
- In the intervals around the peak, select a new grid whose spacing is a fraction of the old one $\delta_{r+1} = \delta_r/\gamma$

Simulation results

- The authors consider $M = 8$ sensors separated by half a wavelength. $K = 2 \quad (62°, 65°), T = 200, N = 180$.

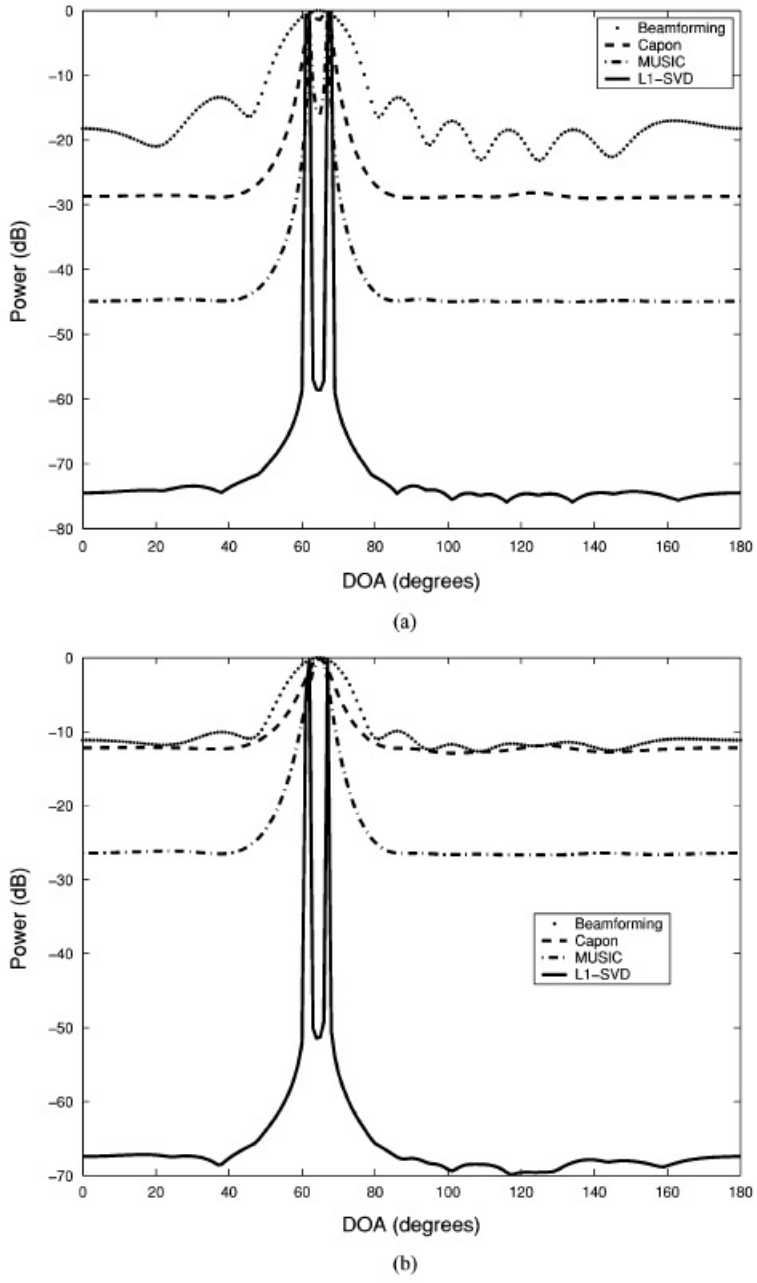


Fig. 4. (a) and (b). Spatial spectra for beamforming, Capon's method, MUSIC, and the proposed method (ℓ_1-SVD) for uncorrelated sources. DOAs: 62° and 67°. Top: SNR = 10 dB. Bottom: SNR = 0 dB.

- For correlated sources, the result is as follows

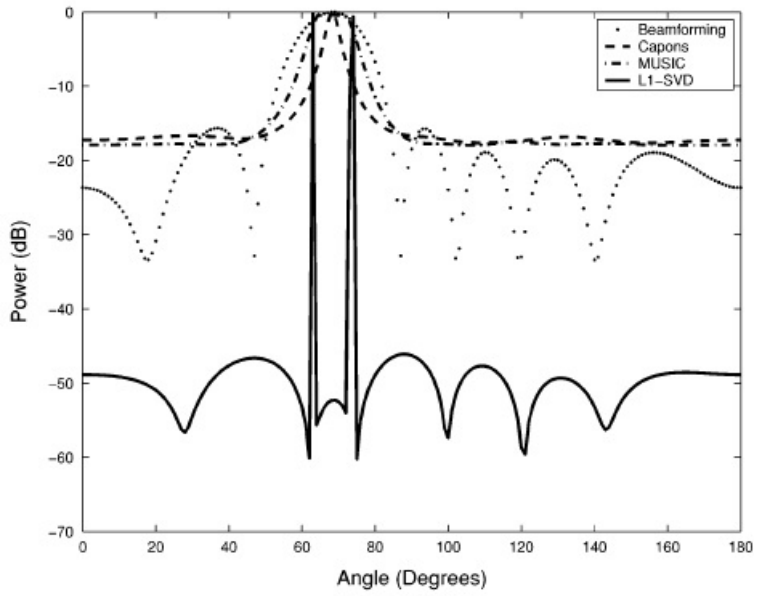


Fig. 5. Spectra for correlated sources. SNR = 20 dB. DOAs: 63° and 73°.

summary

In this paper, the authors have proposed a source location estimation based on sparse representation. The SVD of the sensor measurements summarizes the large chunk of data which is then used as a model for identifying the sources. This method is applicable for both narrow and wideband beamforming. The authors have also presented a grid refinement method in order to obtain fine estimates. The advantages of the proposed method include superior resolution ability with limited time samples for both correlated and uncorrelated sources.

References

[1] J. J. Fuchs, "Linear programming in spectral estimation. Application to array processing," in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 6, 1996, pp. 3161-3164.

[2] J. J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," IEEE Trans. Signal Process., vol. 49, no. 4, pp. 702-709, Apr. 2001.

Enhancing Iterative Decoding of Cyclic LDPC Codes Using Their Automorphism Groups

Authors: Chao Chen, Baoming Bai, Xinquan Yang, Li Li, Yang Yang
Publication: IEEE T. Comm, June 2013
Speaker: Jeong-Min Ryu

Short summary: For cyclic LDPC codes, they propose to use their automorphism groups to improve the iterative decoding performance. Three types of iterative decoders are devised to take advantage of the code's automor-phism group. Towards exploiting the automorphism group of a code, they propose a new class of cyclic LDPC codes based on pseudo-cyclic MDS codes with two information symbols, for which nonequivalent parity-check matrices are obtained. Simulation results show that for their constructed codes of short lengths, the automorphism group can significantly enhance the iterative decoding performance.

I. INTRODUCTION

- **The use of automorphism group for classical codes**

Most classical codes are defined by high-density parity-check (HDPC) matrices, whose Tanner graphs have a large number of short cycles.

→ Iterative decoding performs rather poorly for these codes.

→To mitigate the deleterious effect of short cycles, Jiang and Narayanan [3] and Kothiyal et al. [4] proposed adaptive versions of iterative decoding, respectively.

→As a result, the performance was greatly improved. However, a significant increase in decoding complexity was incurred.

Classical codes are known to have a very rich algebraic structure.

→ To overcome the adverse effect of short cycles while maintaining a reasonable complexity, the automorphism group, as a code structure, was exploited for iterative decoding

→ For HDPC and moderate-density parity-check (MDPC) codes, the automorphism group aided iterative decoding techniques are applied.

- **In this paper,**

1) they apply automorphism group aided iterative decoding techniques to cyclic LDPC codes.

2) For a cyclic code, two particular subgroups of the automorphism group are well known. They show that for a large class of cyclic LDPC codes [15]-[18], [20], the two subgroups of the automorphism group belong to the same equivalence class and thus cannot be harnessed for iterative decoding.

3) They present a class of cyclic LDPC codes for which the automorphism group can be exploited for iterative decoding.

II. How to Use the Automorphism Group of a Code in Iterative Decoding

A. The Automorphism Group of a Code

Definition: Let C be a binary linear block code of length N. The set of coordinate permutations that map C to itself forms a group under the composition operation. This group is called the automorphism group of C, denoted by Aut(C)

For a permutation $\pi \in \mathrm{Aut}(C)$, let π^{-1} denote its inverse. From the definition we know that for any $c = (c_0, c_1, ..., c_{N-1}) \in C$, $\pi c = \left(c_{\pi^{-1}(0)}, c_{\pi^{-1}(1)}, ..., c_{\pi^{-1}(N-1)}\right) \in C$.

Let $C^{\perp}$ denote the dual code of C, then the following property holds.

Property 1: $\mathrm{Aut}(C) = \mathrm{Aut}(C^{\perp})$.

Property 2: For any $\pi \in \mathrm{Aut}(C)$ and a parity check matrix H of C, πH also forms a parity check matrix of C.

Property 3: For a binary cyclic code with odd length N, the automorphism group contains the following two subgroups:

S_0: The set of permutations $\tau^0, \tau^1, ..., \tau^{N-1}$, where $\tau^k : j \to (j+k) \mod N$.

S_1: The set of permutations $\zeta^0, \zeta^1, ..., \zeta^{N-1}$, where $\zeta^k : j \to (2^k \cdot j) \mod N$ and m_1 is the smallest positive integer such that $2^{m_1} \equiv 1 \mod N$.

B. Two Perspectives and Their Equivalence

Using the automorphism group of a code for decoding has a long history. In the early 1960s, MacWilliams devised a hard-decision decoding procedure, called the *permutation decoding* [14]. Recently, the code's automorphism group was brought into use in the soft-decision iterative decoding of HDPC codes [5]−[9] and MDPC codes [10]. Here, they review two possible perspectives involved and show their equivalence.

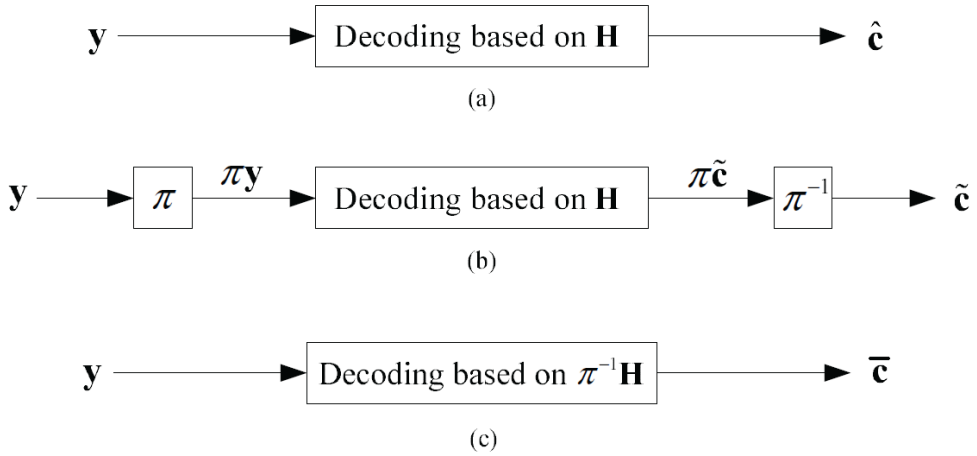


Fig. 1. Iterative decodings of the received sequence based on the SPA with flooding schedule.

(a) Assume that the BPSK signaling is used over the AWGN channel. Let $c \in C$ be the transmitted codeword and $x \in \{\pm 1\}^N$ the corresponding modulated sequence. Then the received signal sequence is given by

$$y = x + n,$$

where n contains N i.i.d. Gaussian noise samples with zero mean and variance σ^2.

(b) Applying a permuation $\pi \in \mathrm{Aut}(C)$, we have

$$\pi y = \pi x + \pi n.$$

Fact 1: In Fig. 1(a) and Fig. 1(b), the outputs $\hat{c}$ and $\tilde{c}$ are not necessarily equal. It is possible that only one of $\hat{c}$ and $\tilde{c}$ is the transmitted codeword.

Fact 2: In Fig. 1(a) and Fig. 1(c), the outputs $\hat{c}$ and $\bar{c}$ are not necessarily equal. It is possible that only one of $\hat{c}$ and $\bar{c}$ is the transmitted codeword.

Fact 3: In Fig. 1(b) and Fig. 1(c), the outputs $\tilde{c}$ and $\bar{c}$ are equal.

C. A partition of the Automorphism Group

We call two parity-check matrices *equivalent* if they can be obtained from each other through row permutations; otherwise, we call them *nonequivalent*. Since the flooding schedule is assumed, we further have

Fact 4: In Fig. 1(a) and Fig. 1(c), if H and πH are equivalent, then the outputs $\hat{c}$ and $\bar{c}$ are equal.

Let $\pi_1, \pi_2 \in Aut(C)$, then π_1 and π_2 belong to the same equivalence class if and only if $\pi_1 H$ and $\pi_2 H$ are equivalent. Note that the partition depends on the selection of H. For a given H, we can construct the same number of nonequivalent parity-check matrices as that of equivalence classes.

D. Design of Three Types of Iterative Decoders

Definition: Let the automorphism group of a code be partitioned based on a given H, a d-order diversity set is a set of d permutations that belong to different equivalence classes.

We choose a d-order diversity set $\{\pi_l : \pi_l \in Aut(C), l = 0,1,...,d-1\}$. Denote by L_y the log-likelihood ratio vector (LLRV) computed from y. Below, they present three types of iterative decoders that use the diversity set in different manners.

- **Decoder-1**: the diversity set is used in a *serial* manner. The decoding procedure is shown in Algorithm 1

Algorithm 1 The Decoding Procedure of Decoder-1

1: **for** $0 \le l \le d-1$ **do**
2: Perform at most I iterations of $\pi_l \mathbf{L_y}$ based on $\mathbf{H}$: if a codeword $\mathbf{c}$ is obtained, stop decoding and output $\pi_l^{-1}\mathbf{c}$.
3: **if** $l = d-1$ and no codeword has been obtained **then**
4: Output the hard-decision of $\mathbf{L_y}$.
5: **end if**
6: **end for**

- **Decoder-2**: the diversity set is used in a *periodic* manner. The decoding procedure is shown in Algorithm 2. In line 5, an inner iteration refers to one time updating of all check nodes and variable nodes of H. For $d = 2$, the decoder works in a Turbo manner [1]. But there are two main differences: 1) The message passing out of a component decoder in the preceding iteration is not subtracted from the a priori information passing to this component decoder in the current iteration; 2) Soft information exchanged between the two component decoders is not limited to information bits.

Algorithm 2 The Decoding Procedure of Decoder-2

1: $\mathbf{s} \leftarrow \mathbf{L_y}$.
2: **for** $0 \le i \le I-1$ **do**
3: **for** $0 \le l \le d-1$ **do**
4: $\mathbf{s} \leftarrow \pi_l \mathbf{s}$.
5: Perform one inner iteration of $\mathbf{s}$ based on $\mathbf{H}$ and update $\mathbf{s}$ as the *a posteriori* LLRV.
6: $\mathbf{s} \leftarrow \pi_l^{-1} \mathbf{s}$.
7: **end for**
8: Use $\mathbf{s}$ to make a hard-decision: if a codeword $\mathbf{c}$ is obtained, stop decoding and output $\mathbf{c}$.
9: **if** $i = I-1$ and no codeword has been obtained **then**
10: Output the hard-decision of $\mathbf{L_y}$.
11: **end if**
12: **end for**

Decoder-3: the diversity set is used in a *parallel* manner. Define $H_l \triangleq \pi_l^{-1} H$. Then by concatenating H_l, we form an augmented parity-check matrix

$$H_{aug} = \begin{bmatrix} H_0 \\ H_1 \\ \vdots \\ H_{d-1} \end{bmatrix}.$$

The decoder performs the SPA with flooding schedule on this highly redundant parity-check matrix, with the maximum number of iterations I.

III. A NEW CONSTRUCTION

We define an $(l\cdot c)\times(l\cdot c)$ binary matrix as

$$B=\begin{bmatrix} A_0 & A_1 & \cdots & A_{c-2} & A_{c-1} \\ A_{c-1}^{(1)} & A_0 & \cdots & A_{c-3} & A_{c-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_2^{(1)} & A_3^{(1)} & \cdots & A_0 & A_1 \\ A_1^{(1)} & A_2^{(1)} & \cdots & A_{c-1}^{(1)} & A_0 \end{bmatrix},$$

where each submatrix is an $l\times l$ circulant and the zeroth row of $A_i^{(1)}$ is the first row of A_i. Define a permutation π as

$$\pi: j\to(j \bmod l)\cdot c+\lfloor j/l\rfloor,\ j=0,1,...,l\cdot c-1. \tag{1}$$

Theorem 4: If we perform row and column permutations on *B*, both using the permutation π given in (1), then we obtain a circulant matrix.

They summarize the construction procedure as follows.

Step 1: Choose a nonzero codeword from an (*n*, 2) pseudo-cyclic MDS code with $a=\alpha$.

Step 2: Use the codeword and its pseudo-cyclic shifts to construct the base matrix W'.

Step 3: Use matrix dispersion on W' to obtain the QC matrix $H_{disp}(W')$.

Step 4: Apply Theorem 4 to $H_{disp}(W')$ to obtain a circulant as the parity-check matrix *H*.

For **Step 1** and **Step 2**, the form of base matrix W' is given by

$$W'=\begin{bmatrix} w_0 & w_1 & \cdots & w_{n-2} & w_{n-1} \\ \alpha w_{n-1} & w_0 & \cdots & w_{n-3} & w_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha w_2 & \alpha w_3 & \cdots & w_0 & w_1 \\ \alpha w_1 & \alpha w_2 & \cdots & \alpha w_{n-1} & w_0 \end{bmatrix}. \tag{2}$$

To construct the base matrix of the form (2), they consider using pseudo-cyclic MDS codes with two information symbols. A pseudo-cyclic code with parameter $a \in GF(q)$ has the property that for any codeword $(c_0, c_1, ..., c_{n-1})$, its pseudo-cyclic shift $(ac_{n-1}, c_0, ..., c_{n-2})$ is also a codeword. If $a = 1$, the pseudo-cyclic code reduces to a cyclic code.

For **Step3**, the Tanner graph corresponding to the matrix $H_{disp}(W)$ has no cycles of length 4 and hence has a girth of at least 6. So the matrix $H_{disp}(W)$ can serve as the parity-check matrix and gives a QC-LDPC code of length $n(q-1)$.

$$H_{disp}(W) = \begin{bmatrix} A(w_{0,0}) & A(w_{0,1}) & \cdots & A(w_{0,n-1}) \\ A(w_{1,0}) & A(w_{1,1}) & \cdots & A(w_{1,n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ A(w_{m-1,0}) & A(w_{m-1,1}) & \cdots & A(w_{m-1,n-1}) \end{bmatrix}.$$

The way to construct the matrix A is given as follow:

Let GF(q) be a finite field with q elements and a be a primitive element of GF(q). Then $\alpha^{-\infty} \triangleq 0, \alpha, \cdots, \alpha^{q-2}$ give all the elements of GF(q). For each non-zero element $\alpha^i, (0 \le i \le q-2)$, define a $(q-1)\times(q-1)$ matrix $A(\alpha^i)$ over GF(2): it is a circulant permutation matrix; the zeroth row is a $(q-1)$-tuple with weight one where the ith component is equal to one and all the other $q-2$ components are equal to zero. The matrix $A(\alpha^i)$ is referred to as the $(q-1)$-fold matrix dispersion of element ai over GF(2). The $(q-1)$-fold matrix dispersion of zero element of GF(q), $A(0)$, is defined as the $(q-1)\times(q-1)$ all-zero matrix.

IV. Simulation Results

They present the simulation results for our constructed cyclic LDPC codes. The BPSK modulated AWGN channel is assumed. In addition to the three decoders presented in Section II, they also simulated a decoder that is not assisted by the automorphism group. The decoder is called Decoder-0, which performs the SPA with flooding schedule on the defining parity-check matrix. For all these decoders, the maximum number of iterations is set to be 100.

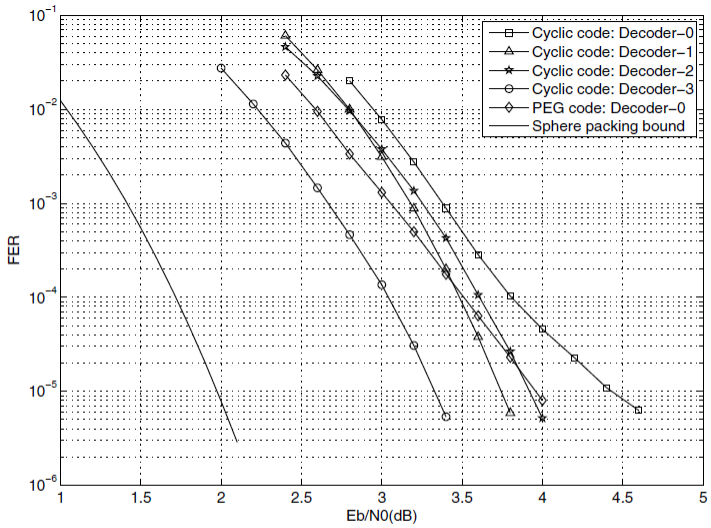


Fig. 2. Performance of the $(341, 160)$ cyclic LDPC code and the $(341, 160)$ PEG-based LDPC code.

Fig. 2 shows the FER performance of the code. The 2-order diversity set $\{\zeta^0, \zeta^1\}$ is used. For comparison, they further simulated a (341,160) LDPC code constructed using the progressive-edge-growth (PEG) algorithm [29]. The parity-check matrix of the code has column weight 3 and row weights 5 and 6. The sphere-packing bound [30] for this length and rate is also included in the figure.

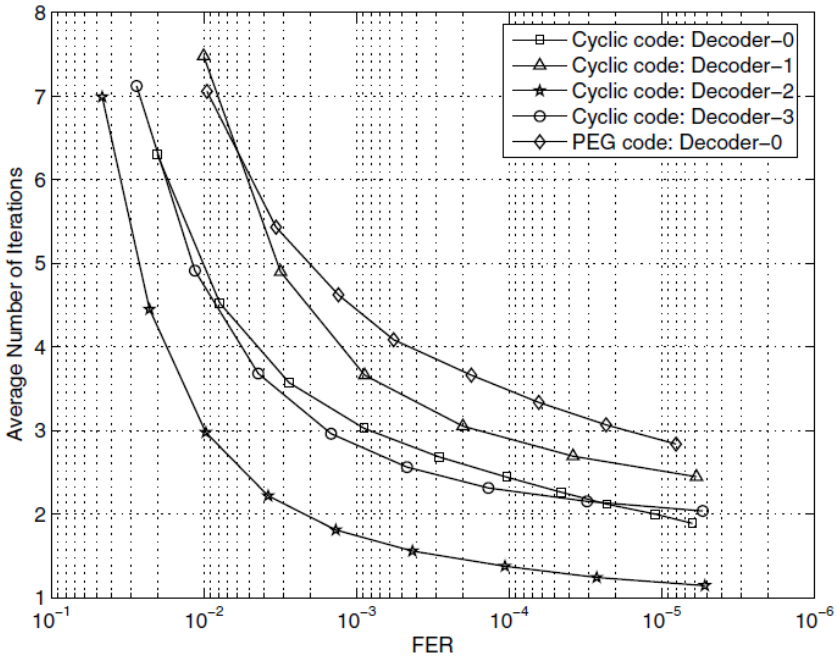


Fig. 3. Average number of iterations for all decoders.

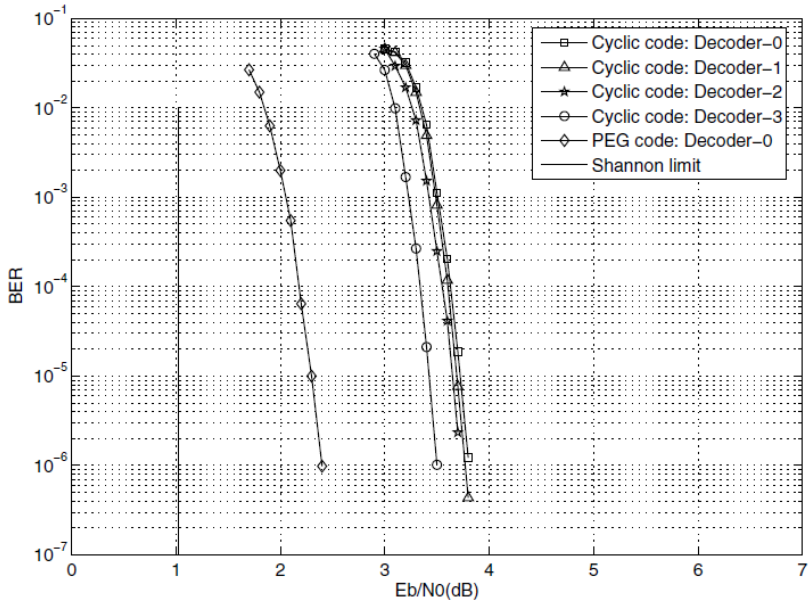


Fig. 4. Performance of the $(5461, 3612)$ cyclic LDPC code and the $(5461, 3612)$ PEG-based code.

- Like HDPC codes, the performance gain for LDPC codes seems more significant for short code lengths. This can be seen by comparing Fig. 2 and Fig. 4 (note that the diversity sets for the two codes have the same order).
- For both HDPC and LDPC with long code lengths, the automorphism group aided iterative decoding does not perform well. In fact, for long HPDC codes, it performs even worse than the hard-decision decoding.
- To obtain a noticeable performance gain, LDPC codes require a smaller diversity set than HDPC codes. This is because that LDPC codes are inherently more suited to iterative decoding than HDPC codes.

Multiuser Detection of Sparsely Spread CDMA

Main ref.: D. Guo, and C.-C. Wnag, "Multiuser detection of sparsely spread CDMA," *IEEE Journal on selected areas in communi.*, Apr, 2008 [1]

Presenter: Jaewook Kang

In Proceeding of CS Journal Club in GIST

Date: July. 2013

I. Introduction

This paper has discussed about design and analysis of multiuser detection (MUD) using sparsely spread CDMA systems. The objective of the MUD problem is how to detect multiple user signals simultaneously at the low computational cost. The main obstacle is multiple-access interference (MAI). These multiple user signals are interference for each user detection one another. The MAI problem arise in most CDMA systems, and optimal detection in such systems requires exponentially growing computation as the number of user increases. This paper investigates a suboptimal MUD detection using sparse CDMA systems. The key idea of the proposed system is to encode the transmitted waveforms using sparsely spread CDMA codes and detect the signal using a linear-complexity belief propagation (BP) algorithm. We summarize the contributions of this work is following:

1) Description of the sparse CDMA system
2) Ensemble of the sparsely spread CDMA codes
3) Design of the BP algorithm for the MUD problem
4) Asymptotic analysis of performance of the BP algorithm based MUD detection

In this report, we aim to sketch the key point of each contribution of this paper.

II. Description of the sparse CDMA system

We consider a fully-synchronous CDMA system which is able to simultaneously transmit K user signals. As shown in Fig.1, symbols X_k from the k-th user is multiplied by the spreading code $\{S_{lk}\}_{l=1}^{L}$ having code length L, being transmitted to the receiver with gain $\frac{A_k}{\sqrt{\Lambda_k}}$ for the transmit power regulation. Then, the receive observes the L channel outputs per a symbol transmission from K users, given by

$$Y_l = \sum_{k=1}^{K} S_{lk} \frac{A_k}{\sqrt{\Lambda_k}} X_k + N_l \quad \text{for } l = 1 \text{ to } L, \tag{1}$$

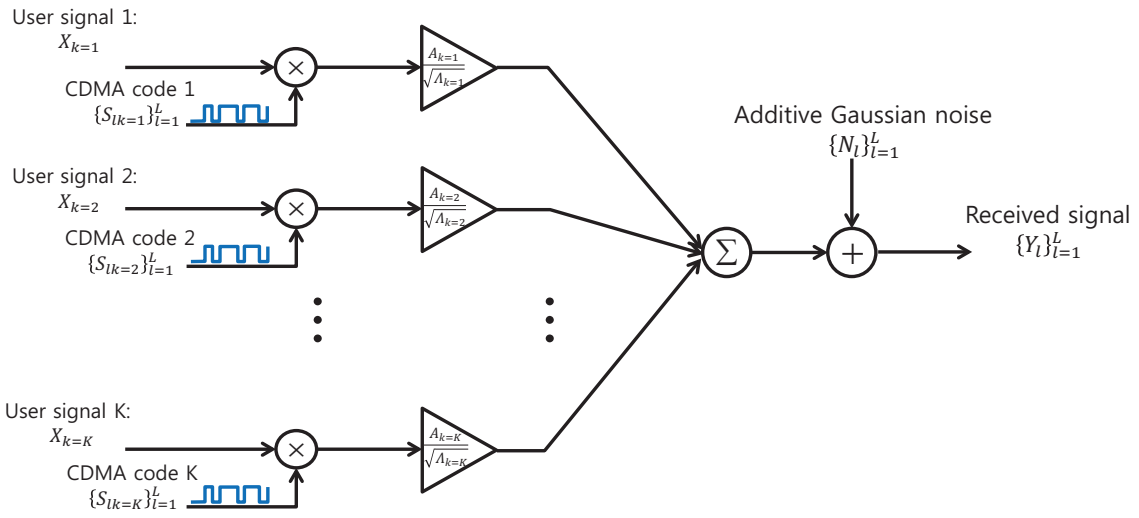


Fig. 1. System model

where we consider additive noise following the zero-mean and unit variance Gaussian distribution, i.e., $N_l \sim \mathcal{N}(0,1)$. In vector form, the expression in (1) can be represented as

$$\underline{Y} = \mathbf{S}\mathbf{A}\underline{X} + \underline{N}, \tag{2}$$

where $\underline{Y} = [Y_1, ..., Y_L] \in \mathbb{R}^L$ denotes the channel output vector, $\underline{X} = [X_1, ..., X_K] \in \mathcal{X}^K \subset \mathbb{R}^K$ is the input symbol vector, $\mathbf{S} \in \mathbb{R}^{L\times K}$ is the sparse spreading matrix, and $\mathbf{A} = diag(\frac{A_1}{\sqrt{\Lambda_1}}, \frac{A_2}{\sqrt{\Lambda_2}}, ..., \frac{A_K}{\sqrt{\Lambda_K}})$ is the gain matrix which has a diagonal form. In the system model, we additionally assume that input symbols X_k, elements of the spreading codes S_{lk} and the transmit gain A_k are i.i.d. drawn from P_X,P_S,P_A respectively. In the receiver side, the goal of the multiuser detector is to estimate the input vector $\underline{X}$ from the channel output vector $\underline{Y}$ given $\mathbf{S}, \mathbf{A}, P_X$.

III. Ensemble of the sparsely spread CDMA codes

Let $\mathbf{H} \in \{0,1\}^{L\times K}$ denote an incidence matrix of the spreading codes $\mathbf{S}$ which indicates nonzero position of the matrix $\mathbf{S}$. The authors also defined that two notation from the incidence matrix, which are

$$\text{The k-th symbol degree: } \Lambda_k = \sum_{l=1}^{L} H_{lk} \tag{3}$$

$$\text{The l-th chip degree: } \Gamma_k = \sum_{k=1}^{K} H_{lk} \tag{4}$$

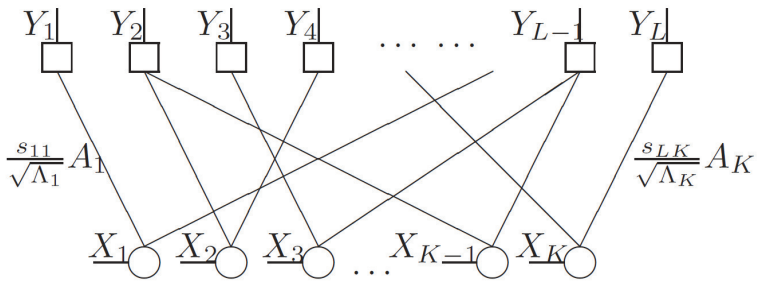


Fig. 2. Factor graphical representation of the CDMA system

Similarly, the average of the symbol and chip degree is defined as $\overline{\Lambda} = \frac{1}{K}\sum_{k=1}^{K}\Lambda_k$ and, $\overline{\Gamma} = \frac{1}{L}\sum_{l=1}^{L}\Gamma_l$, respectively. Then, the factor graph representation of the CDMA system is given in Fig.2. The authors of this paper have tried to analyze the performance of this CDMA system by assuming the following:

1) *Large-system limit*: The system size is very large, i.e., $K, L \to \infty$, and its system load remains a constant, i.e., $\beta \to K/L$.
2) *No-short-cycle*: Under the large-system-limit, the factor graph of the CDMA system does not include cycles shorter than the number of the BP iterations denoted by t.
3) *Chip-semi-regular*: Under the large-system-limit, the chip degree concentrate around their average, i.e., for every l and very small constant $\epsilon > 0$, $\lim_{K,L\to\infty} \Pr\{|\Gamma_l - \mathbf{E}\overline{\Gamma}| > \varepsilon \mathbf{E}\overline{\Gamma}\} = 0$.

Throughout this report, such CDMA system satisfying above assumption is referred to *large-sparse-system* (LSS).

IV. Design of the BP algorithm for the MUD problem

Before discussing the BP detection algorithm for the MUD system, let us summarize the important known facts the BP algorithms

1) BP basically aims to find marginal posterior PDF of each element X_k.
2) In order to reduce the complexity, BP removes the duplicated calculation with message exchanging over the graph connection.
3) *Optimality of BP*: BP provides exact inference (optimal) of the marginal PDFs if the corresponding factor graph is perfectly tree-structured.
4) *Loopy BP*: BP is well applied to graphs with cycles and provides good approximation of the marginal PDFs in practice even through the performance is suboptimal .

For the description of the BP algorithm, we define two notation for the message: *symbol-to-chip* (StC) messages denoted by $V_{k\to l}^{(t)}(x)$ and *chip-to-symbol* (CtS) messages denoted by $U_{l\to k}^{(t)}(x)$ where t is the

number of iterations. In addition, $V_k(x)$ denotes the marginal PDF of X_k. For convenience, we define a set of edge representing statistical connection over the factor graph as $\mathcal{E} := \{(l,k)|S_{lk} \neq 0\}$. Also we define ∂l (resp. ∂k) as the subset of symbols (resp. ships) which have the statistical connection to chip l (resp. symbol k), called its neighborhood. Then, the iterative BP algorithm for computing the marginal PDF of all symbols is shown in Algo.1. This iterative BP algorithm performs exact marginalization of each symbol X_k given the entire observation $\underline{Y}$ if the factor graph is cycle-free. In practical CDMA systems, however, the average node degree is always greater than 2 such that cycles are inevitable. Thus, the BP algorithm performs approximate inference by assuming that all nodes, $\{X_k\}$ and $\{Y_l\}$, are i.i.d. each other.

Algorithm 1 Iterative BP

Inputs: Channel output $\underline{Y}$, Spreading matrix $\mathbf{\Phi}$, Gain matrix $\mathbf{A}$, Prior knowledge $p_X(x)$

Outputs: Marginal PDFs $V_k(x)$ for every k

1)Initialization:

set $U^0_{l\to k}(x) = 1 \ \forall x \in \mathcal{X}$ for every $(l,k) \in \mathcal{E}$

2)Iterations:

for $t = 1$ **to** T **do**

set $V^{(t)}_{k\to l}(x) \propto p_X(x) \times \prod_{j\in\partial k\backslash l} U^{(t-1)}_{j\to k}(x)$ for every $(l,k) \in \mathcal{E}$

set $U^{(t)}_{l\to k}(x) \propto \mathbf{E}\left\{p_{Y_l|\underline{X}}(y|\underline{X})|X_k = x, V^{(t)}_{k\to l}\right\}$

$$:= \sum_{(x_i)_{\partial l\backslash k}} \exp\left[-\frac{1}{2}\left(y_l - \frac{s_{lk}a_k}{\sqrt{\Lambda_k}}x - \sum_{i\in\partial l\backslash k}\frac{s_{li}a_i}{\sqrt{\Lambda_i}}x_i\right)^2\right] \times \prod_{i\in\partial l\backslash k} V^{(t)}_{i\to l}(x_i) \text{ for every } (l,k) \in \mathcal{E}$$

end for

3)Marginal PDFs calculation:

set $V_k(x) \propto p_X(x) \prod_{j\in\partial k} U^{(T)}_{j\to k}(x)$ for every k

The LLR form of BP algorithm is simply obtained by fixing a reference point $x_0 \in \mathcal{X}$ and then defining LLR messages as

$$\text{LLR CtS message: } L^{(t)}_{l\to k}(x) := \log \frac{\mathbf{E}\left\{p_{Y_l|\underline{X}}(y|\underline{X})|X_k = x, L^{(t)}_{k\to l}\right\}}{\mathbf{E}\left\{p_{Y_l|\underline{X}}(y|\underline{X})|X_k = x_0, L^{(t)}_{k\to l}\right\}} \tag{5}$$

$$\text{LLR StC message: } L_{k\to l}^{(t)}(x) := \log p_X(x) + \sum_{j\in\partial k\setminus l} L_{j\to k}^{(t-1)}(x) \tag{6}$$

V. Asymptotic analysis of performance of BP

The key result of this paper states that

The marginal posterior computed for each symbol X_k using BP after t itertions essentially converges to the marginal posterior of a scalar Gaussian channel as the system size increases.

Now, we provide mathematical support for the statement above by stage. Let $P_{X_k}^{bp}(\cdot|\underline{Y}_k^{(t)}, \mathbf{S}, \mathbf{A})$ denote the output CDF from BP, which is approximate posterior of X_k given $\underline{Y}_k^{(t)}$. Here, $\underline{Y}_k^{(t)}$ is all observations within distance $2t-1$ to X_k on the factor graph. If X_k and Y_l is directly connected, the distance will be 1. In addition, let us introduce the canonical scalar Gaussian channel, given as

$$Z = \sqrt{g}X + N, \tag{7}$$

where $X \sim P_X$ and $N \sim \mathcal{N}(0,1)$ are independent, and g denotes the channel gain. For remainder derivation, we use $P_{X|Z;g}(\cdot|z;g)$ to denote the CDF of the posterior distribution of X given Z, according to the Gaussian channel model in (7).

***Theorem** 1 (Gaussian convergence of Marginal posterior):* Given fixed iterations t, the marginal posterior of X_k converges to that of the Gaussian channel, i.e. for every k

$$P_{X_k}^{bp}(x|\underline{Y}_k^{(t)}, \mathbf{S}, \mathbf{A}) \to P_{X|Z;g}(x|h(\underline{Y}_k^{(t)}, \mathbf{S}, \mathbf{A}); \eta^{(t)}A_k^2), \tag{8}$$

in probability under the LSS setup, where the Gaussian channel output is given as $Z = h(\underline{Y}_k^{(t)}, \mathbf{S}, \mathbf{A}) \sim \mathcal{N}(\sqrt{\eta^{(t)}}ax, 1)$, the channel gain $A_k\sqrt{\eta^{(t)}}$ is determined by the following recursion:

$$\frac{1}{\eta^{(t)}} = 1 + \beta\overline{\text{var}}\left\{AX|\sqrt{\eta^{(t-1)}}AX + N\right\}, \tag{9}$$

and

$$\overline{\text{var}}\left\{U|V\right\} := \mathbf{E}\left\{\left(U - \mathbf{E}\left\{U|V\right\}\right)^2\right\}. \tag{10}$$

Proof: The authors proved Theorem 1 by considering messages of the LLR form given in (5) and (6). Proving Theorem 1 is equivalent to showing the LLR StC message is Gaussian distributed with $X_k \sim \mathcal{N}(A_k^2\eta^{(t)}X_k, 1)$. We summarize this proof in four steps.

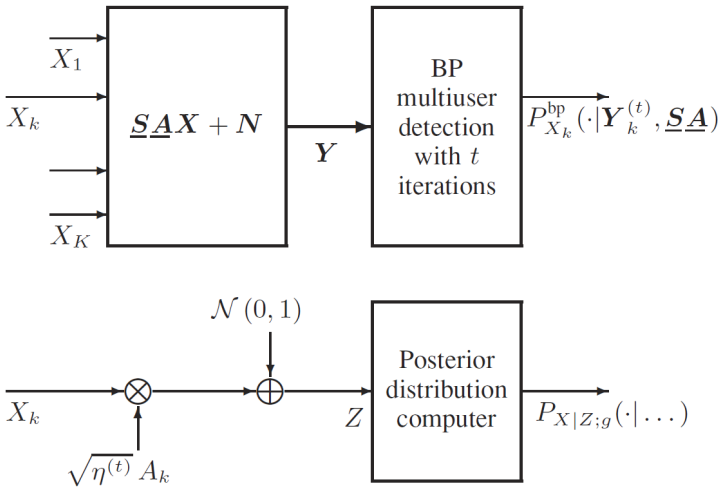


Fig. 3. Upper diagram: Multiuser channel and BP detection. Lower diagram: The asymptotically equivalent scalar Gaussian channel

Step I: The StC message is Gaussian RV by the central limit theorem (CLT): Under the no-short-cycle assumption, all CtS messages $L_{l\to k}^{(t)}$ are i.i.d. conditioned on $X_k = x_k$. From (6), by CLT, the message is a Gaussian random vector.

Step II: LLR obtained from a scalar Gaussian channel is also Gaussian distributed. Namely, for $Y = \sqrt{\gamma}X + N$, its LLR is a Gaussian RV, i.e.,

$$\log \frac{p_{Y|X}(Y|x_1)}{p_{Y|X}(Y|x_0)} = \sqrt{\gamma}(x_1 - x_0)Y - \gamma(x_1^2 - x_0^2)/2 \tag{11}$$

Step III: Calculation of mean and covariance of the StC messages given as

$$\mathbf{E}[L_{k\to l}^{(t)}(x)] = \log p_X(x) + \sum_{j\in\partial k\setminus l} \mathbf{E}[L_{j\to k}^{(t-1)}(x)]. \tag{12}$$

We first consider the mean of the CtS messages. To this end, we have

$$\begin{aligned} f(y,x) &:= \mathbf{E}\left\{p_{Y_l|\underline{X}}(y|\underline{X})|X_k = x, L_{k\to l}^{(t)}\right\} \\ &= \sum_{(x_i)_{\partial l\setminus k}} \left(\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y - \sum_{i\in\partial l\setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i - c_k x)^2 + \underbrace{\sum_{i\in\partial l\setminus k} L_{i\to l}^{(t)}(x_i)}_{=B_{x_k}} \right] \right), \end{aligned} \tag{13}$$

where we use $c_k = \frac{s_{lk}a_k}{\sqrt{\Lambda_k}}$. Then, we apply the 2nd order Taylor approximation with respect to $x = 0$ as

$$\begin{aligned} f(y,x) &\approx f(y,x=0) + f\prime(y,x=0)x + \frac{1}{2}f\prime\prime(y,x=0)x^2 \\ &= g_0(y) + g_1(y)c_k x + \frac{1}{2}g_2(y)c_k^2 x^2, \end{aligned} \tag{14}$$

where we define

$$g_0(y) := \frac{1}{\sqrt{2\pi}} \sum_{(x_i)_{\partial l \setminus k}} \exp\left[-\frac{1}{2}(y - \sum_{i \in \partial l \setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i)^2 + B_{x_k} \right] \tag{15}$$

$$g_1(y) := \frac{1}{\sqrt{2\pi}} \sum_{(x_i)_{\partial l \setminus k}} (y - \sum_{i \in \partial l \setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i) \exp\left[-\frac{1}{2}(y - \sum_{i \in \partial l \setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i)^2 + B_{x_k} \right] \tag{16}$$

$$g_2(y) := \frac{1}{\sqrt{2\pi}} \sum_{(x_i)_{\partial l \setminus k}} ((y - \sum_{i \in \partial l \setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i)^2 - 1) \exp\left[-\frac{1}{2}(y - \sum_{i \in \partial l \setminus k} \frac{s_{li}A_i}{\sqrt{\Lambda_i}} x_i)^2 + B_{x_k} \right]. \tag{17}$$

Then, from (5), the CtS LLR message is given as

$$\begin{aligned} L_{l\to k}^{(t)} &= \log \frac{g_0(y) + g_1(y)c_k x + \frac{1}{2}g_2(y)c_k^2 x^2}{g_0(y) + g_1(y)c_k x_0 + \frac{1}{2}g_2(y)c_k^2 x_0^2} \\ &\approx \frac{g_1(y)}{g_0(y)} c_k (x - x_0) + \frac{g_2(y)}{g_0(y)} c_k^2 (x^2 - x_0^2) - \frac{1}{2}\frac{g_1^2(y)}{g_0^2(y)} c_k^2 (x^2 - x_0^2), \end{aligned} \tag{18}$$

where we further apply the 2nd order Taylor approximation of $\log(x)$. The mean of the CtS message $\mathbf{E}[L_{j\to k}^{(t)}(x)]$ can be obtained by taking integration to (18) with respect to y. Then, using (12), the mean of the LLR StC message is obtained as

$$\mathbf{E}\left[L_{k\to l}^{(t)}(x) \right] = \Theta(x_k(x - x_0) - (x^2 - x_0^2)/2) \tag{19}$$

where

$$\Theta = A_k^2 \int \frac{g_1^2(y)}{g_0(y)} dy \frac{\sum_{j \in \partial k \setminus l} S_{jk}^2}{\Lambda_k}. \tag{20}$$

In (20), by law of large number, $\frac{\sum_{j \in \partial k \setminus l} S_{jk}^2}{\Lambda_k} \to 1$. Here, importantly note that the result in (19) is exactly equivalent to mean of LLR in a scalar Gaussian channel

$$\begin{aligned} X_k &= \sqrt{\Theta} X + N \\ &= A_k \sqrt{\int \frac{g_1^2(y)}{g_0(y)} dy} X + N \end{aligned} \tag{21}$$

where X_k here is a symbol obtained from the BP iteration given the observation $\underline{Y}$ (which is equivalent to Z in Th.1) and $N \sim \mathcal{N}(0,1)$ is additive noise with unit variance. Although the derivation of covariance is omitted here, they are also equivalent.

Now, we summarize the proof as following

1) LLR of StC message is asymptotically a Gaussian RV by CLT from Step I.
2) LLR of a scalar Gaussian channel is a Gaussian RV from Step II.
3) The mean and covariance of LLR StC message have the exactly same form as the LLR of the scalar Gaussian channel in (21) from Step III.
4) From (12), each individual symbol X_k via BP is also Gaussian distributed with $N \sim \mathcal{N}(\sqrt{\Theta}, 1)$.

The last piece of the proof of Theorem 1 is to quantify the corresponding SNR Θ with respect to the number of BP iterations t by showing that

$$\lim_{\substack{L,K\to\infty \\ \overline{\Gamma}\to\infty}} \int \frac{g_1^2(y)}{g_0(y)} dy = \eta^{(t)}. \tag{22}$$

But, the proof of (22) was not well explained in the paper. One thing is that one can derive the recursion in (9) by showing (22).

REFERENCES

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

Compressive Sensing for Spread Spectrum Receivers

Authors: Karsten Fyhn, Tobias L. Jensen, Torben Larsen
Publication: IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, May.2013
Speaker: Hyeongho Baek

Short summary: Compressive sensing enables the receiver to sample below the Shannon-Nyquist sampling rate, which may lead to a decrease power efficiency and production cost. This paper investigates the use of CS in a general Code Division Multiple Access (CDMA) receiver. Furthermore, they numerically evaluate the proposed receiver in terms of bit error rate under different signal to noise ratio conditions and compare it with other receiver structures.

I. INTRODUCTION

- As wireless communication devices are becoming more and more widespread and ubiquitous, the need for power efficiency and low production cost becomes paramount.
- Recently, a new concept termed CS has been attracting more and more attention in the signal processing community. If a signal is sparse in some arbitrary basis, it may be sampled at a rate lower than the Nyquist frequency.
- In the spread spectrum area, some researchers have studied the general use of CS for spread spectrum communication systems.
- In their work they apply CS to a general CDMA system. And they show that a random demodulation implementation may be used to subsample the CDMA signal, but they also develop a simplified version of the RD which performs equally well for CDMA signals but is simpler and cheaper to implement.

II. SIGNAL MODEL

Each slot contains an independent CDMA signal and the slots decoded sequentially and independently of each other.

For one slot, define a discrete QPSK baseband signal, $\mathbf{x} \in \mathbb{C}^{N\times 1}$ as:

$$\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha} \tag{1}$$

where $\boldsymbol{\Psi} \in S_{\Psi} \subset \{\pm 1\}^{N\times N}$ is an orthogonal or near orthogonal dictionary, containing spreading waveforms for transmission, S_{Ψ} is the subset of $\{\pm 1\}^{N\times N}$ that contains orthogonal or near-orthogonal dictionaries and $\boldsymbol{\alpha} \in \{\pm 1 \pm j, 0\}^{N\times 1}$ is a sparse vector, that selects which spreading waveform(s) and what QPSK constellation point(s) to send.

Each node has a unique CDMA sequence assigned, which it uses to transfer information and each node does not know which neighbors it has, but it knows all possible CDMA sequences. Note that in this signal model $\boldsymbol{\alpha}$ is defined so that all users have identical amplitude.

In cases where the number of active nodes or users in a network is smaller than the total number of possible users, the vector $\boldsymbol{\alpha}$ may be assumed sparse, which is the enabling factor for CS.

At the receiver the following signal is observed:

$$\mathbf{y} = \boldsymbol{\Theta}(\mathbf{x} + \mathbf{w}) = \boldsymbol{\Theta}\boldsymbol{\Psi}\boldsymbol{\alpha} + \boldsymbol{\Theta}\mathbf{w} \tag{2}$$

Where $\boldsymbol{\Theta}$ is a measurement matrix, which we shall treat later, and $\mathbf{w} \in \mathbb{C}^{N\times 1}$ is Additive White Gaussian Noise (AWGN). Notice here that we take into account noise folding as the noise is folded down into the compressed domain together with the signal. This makes the noise has an impact on the demodulation performance, because each time the sampling rate is reduced by one half, the Signal to Noise Ratio (SNR) is decreased by $3dB$.

A. Spread Spectrum Dictionary of Gold Sequences

In spread spectrum signals, a possible dictionary $\boldsymbol{\Psi}$ is a set of Gold sequences, as used in e.g. GPS technology. A set of Gold sequences is a special dictionary of binary sequences with very low auto and cross-correlation properties.

When using such a CDMA dictionary, the received signal must be sampled at a rate corresponding to the chip rate, where a chip is one entry in the received Gold sequences. If $\boldsymbol{\alpha}$ is

sparse the information rate of the signal is much lower and it may be possible to decrease the sampling rate by using CS.

III. Compressive Sensing

CS is novel sampling scheme, developed to lower the number of samples required to obtain some desired signal.

Denote by $\mathbf{\Theta}_{\kappa} \in \mathbb{R}^{M \times N}$ a CS measurement matrix, where $\kappa \in \mathbb{N}$ is the subsampling ratio when compared to the Nyquist rate and $M = N / \kappa$. This measurement matrix is then responsible for mapping the N-dimensional signal $\mathbf{x}$ to a M-dimensional signal $\mathbf{y}$. Normally this would make it impossible to recover the original signal, but under the assumption that $\mathbf{x}$ is sparse in some basis, it is possible to reconstruct the original signal from the sampled, M-dimensional signal $\mathbf{y}$.

A. Compressive Spread Spectrum Measurement Matrix

In most CS literature a choice of measurement matrix or structure must be made. The Random Demodulator (RD) sampling structure is one of the most well-known measurement matrix structures developed, which is well suited for practical implementation. In the RD a Pseudo-Random Noise (PRN) sequence is mixed with the received signal. Because a spread spectrum transmitter has already spread the signal before transmission, we show that the RD structure can be improved so that the mixing with a PRN sequence at the receiver may be skipped.

The proposed measurement matrix may therefore be defined similarly to the definition of the RD matrix in “Beyond Nyquist : Efficient Sampling of Sparse Bandlimited Signals”.

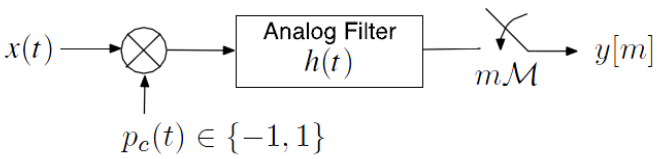


Figure 1 Pseudo-random demodulation scheme

In their work, the measurement matrix is based on two matrices, $\mathbf{D}$ and $\mathbf{H}$. First, let $\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_N \in \{\pm 1\}$ be the chipping sequence used in the RD for a signal of length N. The mapping $\mathbf{x} \to \mathbf{Dx}$ signifies the demodulation mapping with the chipping sequence, where $\mathbf{D}$ is the diagonal matrix:

$$\mathbf{D} = \begin{bmatrix} \varepsilon_0 & & & \\ & \varepsilon_1 & & \\ & & \ddots & \\ & & & \varepsilon_N \end{bmatrix}$$

Second, the $\mathbf{H}$ matrix denotes the accumulate-and-dump action performed after mixing. Let M denote the number of samples taken. Then each sample is the sum of N/M consecutive entries of the demodulated signal. An example with $M = 3$ and $N = 6$ is :

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & & & & \\ & & 1 & 1 & & \\ & & & & 1 & 1 \end{bmatrix}$$

The reason for applying a chipping sequence is to spread the signal across the frequency spectrum, so that information is aliased down into the lower frequency area, which is left untouched by the low-pass filtering.

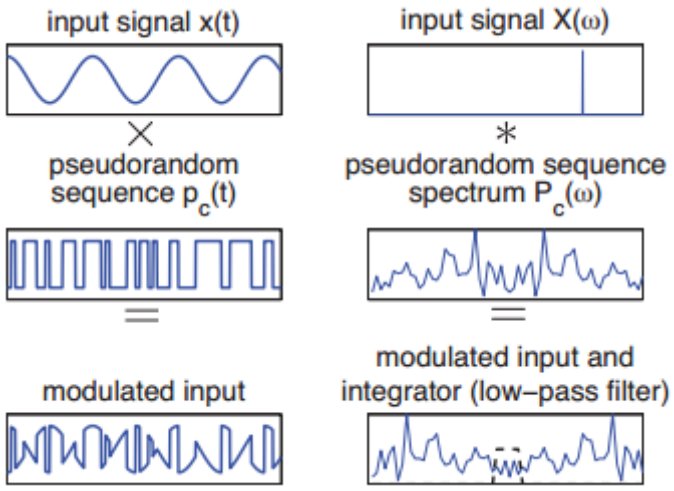


Figure 2 Action of the demodulator on a pure tone. The demodulation process multiplies the continuous-time input signal by a random sequence wave. The action of the system on a single tone is illustrated in the time domain(left) and the frequency domain (right). The dashed line indicates the frequency response of the lowpass filter.

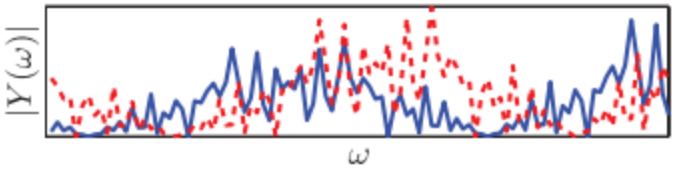


Figure 3 Signatures of two different tones. The random demodulator furnishes each frequency with a unique signature. This image enlarges the filter's passband region of the demodulator's output for two input tones (solid and dashed). The two signatures are nearly orthogonal.

In the proposed receiver this mixing is unnecessary because the signal has already been spread at the transmitter. The proposed receiver may therefore be simplified to:

$$\mathbf{y = Hx} \tag{3}$$

This is significantly simpler to implement in hardware than the RD. The use of a CDMA dictionary introduces a random-like dictionary matrix, which spreads the signal out so that each sample contains a little bit of the original information signal. Therefore, the sampling process may be rewritten as:

$$\mathbf{y = Hx = H\Psi\alpha = \Theta\alpha} \tag{4}$$

Here, the measurement matrix becomes $\mathbf{\Theta = H\Psi}$

B. Subspace Pursuit

To reconstruct the signal a reconstruction algorithm must be chosen. Many different approaches have been developed, but two main classes of reconstruction algorithms are in widespread use: l_1 minimization and greedy algorithms. Often, l_1 minimization provides the best solution, but if the matrices $\mathbf{\Psi}$ and $\mathbf{\Theta}$ are very large, it is much more efficient to use the simpler greedy algorithms. Therefore, we choose to use greedy algorithms in this work.

Recall that $\mathbf{\Theta}_\kappa$ is a measurement matrix with N columns and N/κ rows and define $\mathbf{A} = \mathbf{\Theta}_\kappa \mathbf{\Psi}$. Then we define the Subspace Pursuit algorithm as in Algorithm 1. In each algorithm iteration, the pseudo-inverse is calculated as the least-squares solution as this is less computationally demanding.

Algorithm 1 Subspace Pursuit Algorithm [3]

Input:

Sparsity S, measurement and dictionary matrices combined $\mathbf{A}$ and received, sampled signal $\mathbf{y}$

Initialization:

$T^0 = \{\text{indices of the S largest absolute magnitude entries in the vector } \mathbf{A}^T\mathbf{y}\}$

$\mathbf{y}_r^0 = \mathbf{y} - \mathbf{A}_{T^0}\mathbf{A}_{T^0}^T\mathbf{y}$

Repeat

$l \leftarrow l+1$

$T^l \leftarrow T^{l-1} \cup \{\text{indices of the S largest absolute magnitude entries in the vector } \mathbf{A}^T\mathbf{y}_r^{l-1}\}$

$T^l \leftarrow \{\text{indices of the S largest absolute magnitude entries in the vector } \mathbf{A}_{T^l}^{\dagger}\mathbf{y}\}$

$\mathbf{y}_r^l \leftarrow \mathbf{y} - \mathbf{A}_{T^l}\mathbf{A}_{T^l}^{\dagger}\mathbf{y}$

Until $\| \mathbf{y}_r^l \|_2 > \| \mathbf{y}_r^{l-1} \|_2$, $l \geq S$

To demonstrate the performance of the Subspace Pursuit algorithm with the Gold dictionary, they have performed numerical experiments to find the phase transition in the noise-less case for various choices of measurement matrices.

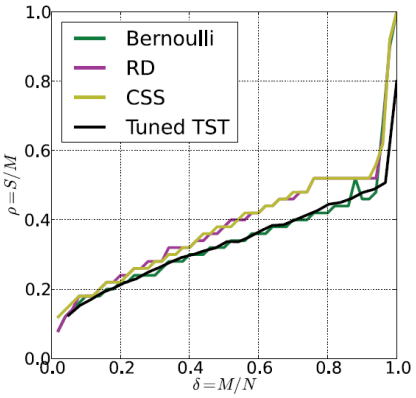


Figure 4 Phase Transition Diagrams for the three different measurement matrices (Rademacher, RD and CSS measurement matrix). The black line is the phase transition line for the Tuned Two Stage Thresholding(TST) algorithm from "Optimally tuned iterative reconstruction algorithms for compressed sensing"

IV. CONCLUSION

In this work they apply CS to a general CDMA system and they show that it is possible to use a very simple measurement scheme at the receiver side to enable subsampling of the CDMA signal.

V. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

Appendix

Reference

[1] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," IEEE Trans.Inf. Theory, vol. 55, no. 5, pp. 2230–2249, May 2009.
[2] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," IEEE J. Sel. Topics Signal Process.,vol. 4, no. 2, pp. 330–341, Apr. 2010.
[3] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230–2249, May 2009.

Active illumination single-pixel camera based on compressive sensing.

Filipe Magalhaes et al.

APPLIED OPTICS. (2011.02)

Presenter : Eunseok Jung

GIST, Dept. of Mechatronics , BiO-scopy Lab.

Gwangju Institute of Science and Technology

Contents

Background

- Active illumination

  - Direction, intensity, and pattern of illumination are controlled by commands or signals.

- Single pixel camera

  - Single pixel camera, developed originally at Rice University, is one of the paramount examples of CS.

  - It can be seen as an optical computer comprising:

    = digital micromirror device(DMD) (1024x768 micromirros)

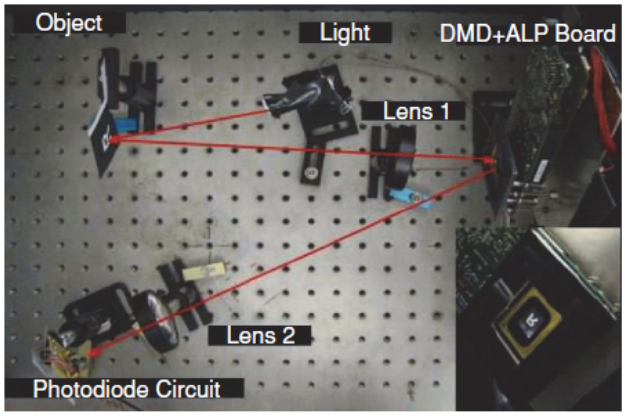
    = two lenses

    = single photodetector

    = A/D converter.

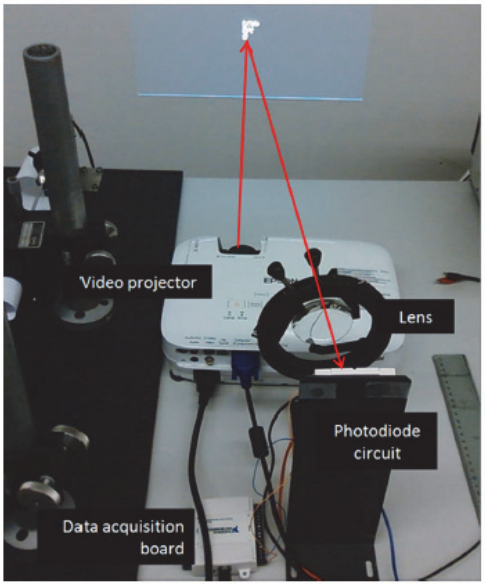
Background

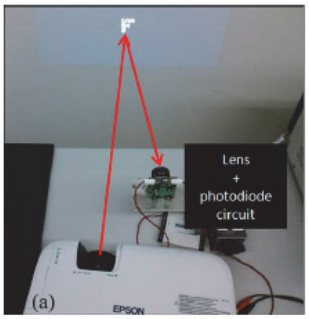
- Single pixel camera

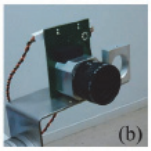


- The process which include changing DMD pattern and taking signal from PhotoDiode will be repeated until M values are acquired.
- Each of these values (output voltage of the photodiode) can be interpreted as the inner product of the desired image x with a measurement basis $\Phi(m) = 1, 2, \ldots, M$.
- The resolution of the reconstructed image is limited by the pixel arrangement of the DMD.

Active illumination Single-pixel camera







Instead of the DMD, Active illumination single-pixel camera setup used video projector to incorporate the random measurement matrix into the system.

Active illumination Single-pixel camera

- The video projector was used to project the result of the product between the image to be reconstructed and the random measurement patterns.
- Each of the output voltages of the photodiode amplifier circuit is representative for the inner product between the pattern used for that measurement and the image to be reconstructed.

Result

- Traditional Single-pixel camera(Rice university version)

Ideal image with 64*64 pixels

(a) (b)

Fig. 9. Images (64 × 64 pixels) reconstructed using the best K-term Haar wavelet approximation: (a) $K = 400$ and (b) $K = 675$ [11].

(a) (b)

Fig. 10. CS reconstruction from: (a) $M = 820$ measurements ⇒ 20% and (b) $M = 1600$ measurements ⇒ 39% [11].

Result

- Active illumination single-pixel camera

(a) (b) (c)

Fig. 14. First results obtained (32 × 32 pixels ⇒ $N = 1024$) with the active illumination single-pixel camera: (a) 205 measurements ⇒ 20% (peak signal-to-noise ratio (PSNR) = 11.08 dB), (b) 410 measurements ⇒ 40% (PSNR = 12.30 dB), and (c) 717 measurements ⇒ 70% (PSNR = 13.21 dB).

1. Sharp edge was increased when use active illumination single pixel camera.
2. This system is cheaper and more compact than Rice university one.
3. It doesn't need active illumination source.

Thank you

Resource Allocation in Cognitive Radio Relay Networks

Authors: J.C Liang et. al.

Publication: IEEE JSAC, Mar. 2013
Speaker: Asif Raza

Short summary: In this paper authors formulate the problem of Resource Allocation (RA) in Cognitive Radio (CR) networks with relay stations. The problem takes into account the issues like: fluctuations of usable spectrum resource, channel quality variations caused by frequency selectivity, and interference caused by different transmit power levels. They propose easy to implement heuristic algorithms. The simulation results reveal that presented solutions show good proportional fairness among CR users and improvement in system throughput by power control.

I. INTRODUCTION

Resource allocation in CR relay networks is considered in the paper. The relay nodes are simply Medium Access Control (MAC) repeaters.

Main challenges in designing RA algorithm are:

- Fluctuations in available spectrum resource: The number of usable resource may differ in each area. It may cause great variations in available band-width and jitters between packets. Therefore they adopt proportional fair scheduling which allocates resource to CR MSs proportionally to their capabilities such as transmission rates determined by channel quality.
- Instability of wireless channels: The frequency selectivity would cause variations of Signal-to-Noise Ratio (SNR). As a result, the quality of each channel would differ on each node. The RA algorithm needs to take into account the instability of channel quality.
- Power control and interference among nodes: the number of usable channels are determined by the activity of PR networks. With proper power control, spatial reuse can be achieved and two adjacent CR RSs can transmit over the same channel.

II. SYSTEM DESCRIPTION

The network design considered is shown in the following figure.

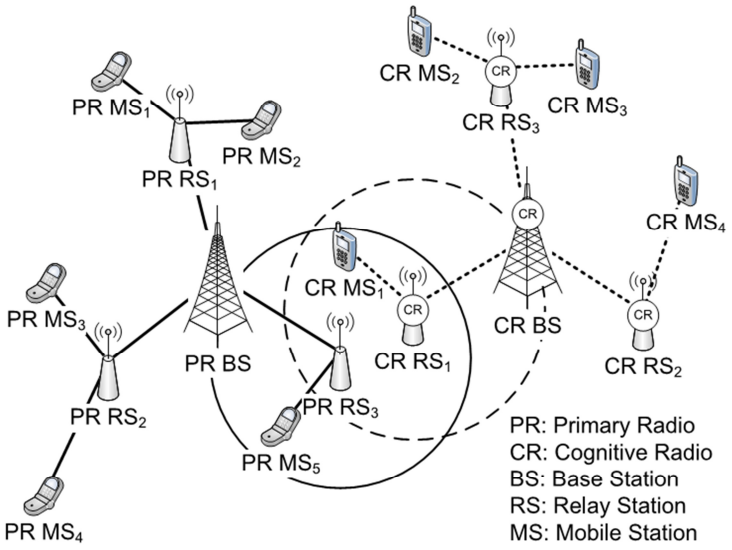


Fig. 1. An illustration of a cognitive radio relay network.

The network drawn with solid and dashed lines corresponds to Primary Radio (PR) network and Cognitive Radio (CR) network. The RA is performed by the Cognitive Radio Base Station (CR BS) in a centralized manner. The transmission of CR network is time-divided into frames and synchronized with the PR network. In each frame there are multiple time-slots not used by PR. The CR BS gathers available channels within the vicinity of each node and the quality of each channel. The CR BS then schedules the usage of frequency bands and time-slots for its downstream Cognitive Radio Relay Stations (CR RSs) and Cognitive Radio Mobile Stations (CR MSs). Each CR MS can connect to CR BS directly or through CR RS. The RA is done in downlink transmission only. The notations used are shown in table 1. The authors starts RA problem formulation by considering fixed transmit power and then extend the formulation to incorporate adjustable transmit power.

III. FIXED TRANSMIT POWER

A. *Problem Formulation:* proportional fair scheduling for cognitive relay networks with fixed transmit power in relay stations is considered. (Given parameters same as in system description section)

$$\mathrm{O}(\lambda) = \max \sum_{m \in \mathcal{M}} \frac{\lambda_m(t)}{\rho_m(t)} \text{ where } \lambda_m = \sum_{c \in \mathcal{C}} \sum_{t=0}^{N-1} r_{1_m^p}(c,t) \qquad (1)$$

Where $\lambda_m(t)$ is scheduled rate of CR MS *m* in frame *t*, and $\rho_m(t)$ is the long-term average rate of CR MS *m* until frame *t*.

TABLE I
LIST OF NOTATIONS

Notation	Description
$\mathcal{M}$	Set of CR Mobile Stations
$\mathcal{R}$	Set of CR Relay Stations
$\mathcal{R}^+$	Set of CR Relay Stations and the CR Base Station
$\mathcal{L}$	Set of all links
$\mathcal{C}$	Set of sub-channels
$\mathcal{P}$	Set of available transmit power levels
N	Number of slots in one frame
l_i^c	The set of links connecting to node i's children.
l_i^p	The link connecting to node i's parent.
$\mathrm{R}_l(c)$	The maximum sustainable rate of sub-channel c on link l in bits/sec, $l \in \mathcal{L}, c \in \mathcal{C}$
$\mathrm{r}_l(c,t)$	The actual data transmitted on link l over sub-channel c at time-slot t in bits, $l \in \mathcal{L}, c \in \mathcal{C}$
ρ_m	Long-term average rate of MS m in bits/frame
p_i	The transmit power level of CR RS i
$\mathrm{I}_l(c,t)$	Indicator variable as defined in (6)
$\mathrm{e}_{(i,j)}$	Indicator variable as defined in (7)
$\mathrm{v}_{(i,c)}$	Indicator variable as defined in (8)

$$\sum_{c\in\mathcal{C}}\sum_{t=0}^{\tau} r_{\mathrm{l}_i^p}(c,t) \geq \sum_{k\in\mathrm{l}_i^c}\sum_{c\in\mathcal{C}}\sum_{t=0}^{\tau} r_k(c,t) \;\; \forall i\in\mathcal{R},\, 0\leq\tau\leq N-1$$
$$\sum_{c\in\mathcal{C}}\sum_{t=0}^{N-1} r_{\mathrm{l}_i^p}(c,t) = \sum_{k\in\mathrm{l}_i^c}\sum_{c\in\mathcal{C}}\sum_{t=0}^{N-1} r_k(c,t) \qquad \forall i\in\mathcal{R} \tag{2}$$

$$\max_{c\in\mathcal{C}} \mathrm{I}_{\mathrm{l}_i^p}(c,t) + \max_{k\in\mathrm{l}_i^c, c\in\mathcal{C}} \mathrm{I}_k(c,t) \leq 1 \;\; \forall i\in\mathcal{R},\, 0\leq t\leq N \tag{3}$$

$$\sum_{k\in\mathrm{l}_i^c} \mathrm{I}_k(c,t) + \sum_{j\in\left\{\mathcal{R}^+\setminus i\right\}}\sum_{m\in\mathrm{l}_j^c} e_{(i,j)}\mathrm{I}_m(c,t) \leq 1 \quad \forall i\in\mathcal{R}^+,\, \forall c\in\mathcal{C},\, 0\leq t\leq N \tag{4}$$

$$r_j(c,t) \leq \mathrm{I}_j(c,t)\mathcal{R}_j(c) \quad \forall j\in\mathcal{L},\, \forall c\in\mathcal{C},\, 0\leq t\leq N \tag{5}$$

$$\max_{k\in\mathrm{l}_i^c, 0\leq t\leq N} \mathrm{I}_k(c,t) \leq v_{(i,c)} \quad \forall i\in\mathcal{R}^+,\, \forall c\in\mathcal{C} \tag{6}$$

The problem is integer programming problem which is NP-hard and computationally infeasible to solve. By relaxing the integrality constraint the computationally lighter upper bound of original problem is defined as: let $w_l(c) = \frac{1}{N}\sum_{t=0}^{N-1} \mathrm{I}_l(c,t)$

Maximize: $$\mathrm{O}(\lambda) = \sum_{m \in \mathcal{M}} \frac{\lambda_m^*}{\rho_m} \tag{1}$$

Subject to: $$\lambda_m^* = \mathrm{N} \sum_{c \in \mathcal{C}} w_{\mathrm{l}_m^p}(c) R_{\mathrm{l}_m^p}(c) \tag{2}$$

$$\sum_{c \in \mathcal{C}} w_{\mathrm{l}_i^p}(c) R_{\mathrm{l}_i^p}(c) \geq \sum_{k \in \mathrm{l}_i^c} \sum_{c \in \mathcal{C}} w_k(c) R_k(c) \qquad \forall i \in \mathcal{R} \tag{3}$$

$$\sum_{k \in \mathrm{l}_i^c} w_k(c) + \sum_{j \in \{\mathcal{R}^+ \setminus i\}} e_{(i,j)} \sum_{m \in \mathrm{l}_j^c} w_m(c) \leq \frac{2}{R+1} \forall i \in \mathcal{R}^+, \forall c \in \mathcal{C} \tag{4}$$

$$\max_{k \in \mathrm{l}_i^c} w_k(c) \leq v_{(i,c)} \qquad \forall i \in \mathcal{R}^+, \forall c \in \mathcal{C} \tag{5}$$

Algorithm 1 Greedy Algorithm for PFSCRN-FTP

Let T be the set of CR RSs that have been scheduled.
Let U be the set of CR RSs that interfere with each other.
Let M be the candidate CR MSs to be scheduled.
$M \leftarrow \phi$
for all $c \in \mathcal{C}$ **do**
  $T \leftarrow \phi$
  for all $r \in \mathcal{R}$ such that $\mathrm{v}_{(r,c)} = 1$ and $r \notin T$ **do**
    $U \leftarrow \phi$
    for all $i \in \mathcal{R}$ **do**
      if $\mathrm{e}_{(r,i)} = 1$ and $\mathrm{v}_{(i,c)} = 1$ **then**
        $T \leftarrow i, \quad U \leftarrow i$
      end if
    end for
    for all $u \in U$ **do**
      $m_{(u,c)} = \arg \max_{m \in u's\ children} \frac{\mathrm{R}_{\mathrm{l}_m^p}(c)}{\rho_m}$
      $M \leftarrow M \cup \{m_{(u,c)}\}$
    end for
  end for
end for
Sort elements in M in descending order of contribution to the objective function
Denote by $n_{(l,c)}$ the available time-slots on link l over sub-channel c
for all $m_{(u,c)} \in M$ **do**
  $D_m = \mathrm{R}_{m_{(u,c)}}(c) n_{(\mathrm{l}^p_{m_{(u,c)}}, c)}$
  $SchedData \leftarrow 0$
  while $SchedData < D_m$ and CR BS has available sub-channels **do**
    $c' \leftarrow \arg \max_{c \in \mathcal{C}\ and\ \mathrm{v}_{(BS,c)}=1} \mathrm{R}_{\mathrm{l}_u^p}(c)$
    Allocate available time-slots of sub-channel c' on link l_u^p to CR RS u
    $SchedData \leftarrow SchedData + \mathrm{R}_{\mathrm{l}_u^p}(c')$
  end while
  Allocate time-slots of sub-channel c on link $\mathrm{l}^p_{m_{(u,c)}}$ to $m_{(u,c)}$
end for

The algorithm has two parts:

- Find and schedule available sub-channels for each CR RS.
- Allocate sub-channels and time-slots of all hops in a greedy-based approach.

Resolve conflicts among interfering CR RSs and schedule their available sub-channels: The algorithm scans through all the sub-channels. For each sub-channel c, it scans through each CR RS r, and record all CR RSs that interfere with r in the set U. All of the CR RSs in the set U have access to sub-channel c, but are interfered with each other. The CR RS m with maximum value of $\frac{R_{l_m^p}(c)}{\rho_m}$ in U can access the sub-channel c at a given time.

Allocate the resource between the CR BS and the CR RSs: the CR MSs are sorted in descending order of their contributions to the objective function. For each m in the sorted order and the CR RS u that m attaches to, the amount of service that m receives is initialized to the number of bits that can be transmitted using all the available time-slots between m and u over the scheduled sub-channel c (line 23). The sub-channels between the CR BS and u are allocated in the order of transmission rate. That is, the sub-channel c' with highest rate $R_{l_u^p}(c')$ is allocated first, and then the second highest sub-channel is allocated, and so on.

IV. Variable Transmit Power

Better spectrum utilization and system throughput can be achieved with proper control of CR RS transmit power. A scheduler decides power level of a CR RS from pool of discrete power levels. The decision normally involves the interference indicator between two CR RSs, e(i, j), and the maximum sustainable rate Rl (c) on an CR RS-CR MS link l over a sub-channel c become functions of the transmit power p.

A. Problem Formulation: Let p_i denotes the transmit power level of CR RS *i*. (Given parameters are same as in fixed power transmission).

$$\text{Objective: } O(\lambda) = \max \sum_{m \in \mathcal{M}} \frac{\lambda_m(t)}{\rho_m(t)} \text{ where } \lambda_m = \sum_{c \in \mathcal{C}} \sum_{t=0}^{N-1} r_{l_m^p}(c,t) \tag{1}$$

$$\text{Subject to: } p_i \in \mathcal{P}, \forall i \in \mathcal{R} \tag{2}$$

$$\begin{aligned} &\sum_{c \in \mathcal{C}} \sum_{t=0}^{\tau} r_{l_i^p}(c,t) \geq \sum_{k \in l_i^c} \sum_{c \in \mathcal{C}} \sum_{t=0}^{\tau} r_k(c,t) \ \forall i \in \mathcal{R}, 0 \leq \tau \leq N-1 \\ &\sum_{c \in \mathcal{C}} \sum_{t=0}^{N-1} r_{l_i^p}(c,t) = \sum_{k \in l_i^c} \sum_{c \in \mathcal{C}} \sum_{t=0}^{N-1} r_k(c,t) \qquad \forall i \in \mathcal{R} \end{aligned} \tag{3}$$

$$\max_{c\in\mathcal{C}} \mathrm{I}_{1_i^p}(c,t) + \max_{k\in 1_i^c, c\in\mathcal{C}} \mathrm{I}_k(c,t) \le 1 \quad \forall i\in\mathcal{R}, 0\le t\le N \tag{4}$$

$$\sum_{k\in 1_i^c} \mathrm{I}_k(c,t) + \sum_{j\in\{\mathcal{R}^+\setminus i\}} \sum_{m\in 1_j^c} e_{(i,j)}(p_i)\mathrm{I}_m(c,t) \le 1 \qquad \forall i\in\mathcal{R}^+, \forall c\in\mathcal{C}, 0\le t\le N \tag{5}$$

$$r_j(c,t) \le \mathrm{I}_j(c,t)\mathcal{R}_j(c,p_i) \qquad \forall i\in\mathcal{R}^+, \forall j\in 1_i^c, \forall c\in\mathcal{C}, 0\le t\le N \tag{6}$$

$$\max_{k\in 1_i^c, 0\le t\le N} \mathrm{I}_k(c,t) \le v_{(i,c)} \quad \forall i\in\mathcal{R}^+, \forall c\in\mathcal{C} \tag{7}$$

The problem formulation same as fixed power transmission case except that interference indicator e(i,j)(p) between two CR RSs (i, j) and the maximum sustainable rate $R_l(c, p)$ on an CR RS and CR MS link *l* over sub-channel c are functions of the CR RS transmit power as shown in constraint 5 and 6 respectively.

B. *Proposed Greedy Algorithm for PFSCRN-ATP:* an algorithm for power level selection, shown in Algorithm 2 below, is based upon following criteria:

- Interference to other CR RSs is minimized.
- Spectrum should be fully utilized

Algorithm 2 Arg-Max Algorithm for Determining Power for CR RSs

for all $r \in \mathcal{R}$ **do**
  for all available power p **do**
    $u_{(r,p)} \leftarrow 0$
    for all $c \in \mathcal{C}$ and $\mathrm{v}_{(r,c)} = 1$ **do**
      $g_{(r,c,p)} \leftarrow \max_{m\in r's\ children} \frac{\mathrm{R}_{1_m^p}(c,p)}{\rho_m}$
      $u_{(r,p)} \leftarrow u_{(r,p)} + g_{(r,c,p)}$
    end for
  end for
end for
Let $p(r)$ be the scheduled power level of CR RS r
$p(r) \leftarrow \arg \max_{all\ power\ level\ p} u_{(r,p)}$
If there are multiple maximum $u_{(r,p)}$, set $p(r)$ as the lowest p among them

A score is calculated for each combination of (CR RS, power level)(r, p). It is calculated based upon contribution, i.e., $\frac{R_{1_m^p}(c,p)}{\rho_m}$, of m^{th} CR RM node to (1) on each available sub-channel *c* of *r*. The contribution of CR MSs is summed up. The power level for CR RS is determined then as: Priority is

given to the (r, p) combination with highest score while tie is broken by choosing the combination that can achieve the objective function with lowest power level.

After the transmit power of CR RSs are determined by using Algorithm 2, the scheduling problem is solved by using a PFSCRN-ATP as shown in Algorithm 3. It is modified version of Algorithm 1 with selected power levels.

Algorithm 3 Greedy Algorithm for PFSCRN-ATP

Let $p(r)$ be the scheduled power determined by Algorithm 2
Let T be the set of CR RSs that have been scheduled.
Let U be the candidate CR RSs to be scheduled.
Let M be the candidate CR MSs to be scheduled.
$M \leftarrow \phi$
for all $c \in \mathcal{C}$ **do**
  $T \leftarrow \phi$
  for all $r \in \mathcal{R}$ such that $\mathrm{v}_{(r,c)} = 1$ and $r \notin T$ **do**
    $U \leftarrow \phi$
    for all $i \in \mathcal{R}$ **do**
      if $\mathrm{e}_{(r,i)}(p(r)) = 1$ and $\mathrm{v}_{(i,c)} = 1$ **then**
        $T \leftarrow i, \quad U \leftarrow i$
      end if
    end for
    for all $u \in U$ **do**
      $m_{(u,c)} = \arg \max_{m \in u's\ children} \frac{\mathrm{R}_{\mathrm{l}_m^p}(c, p(u))}{\rho_m}$
      $M \leftarrow M \cup \{m_{(u,c)}\}$
    end for
  end for
end for
Denote by $n_{(l,c)}$ the available time-slots on link l over sub-channel c
for all $m_{(u,c)} \in M$ **do**
  $D_m = \mathrm{R}_{m_{(u,c)}}(c, p(u)) n_{(\mathrm{l}^p_{m_{(u,c)}}, c)}$
  $SchedData \leftarrow 0$
  while $SchedData < D_m$ and CR BS has available sub-channels **do**
    $c' \leftarrow \arg \max_{c \in \mathcal{C}\ and\ \mathrm{v}_{(BS,c)}=1} \mathrm{R}_{\mathrm{l}^p_u}(c)$
    Allocate available time-slots of sub-channel c' on link l^p_u to CR RS u
    $SchedData \leftarrow SchedData + \mathrm{R}_{\mathrm{l}^p_u c'}()$
  end while
  Allocate time-slots of sub-channel c on link $\mathrm{l}^p_{m_{(u,c)}}$ to $m_{(u,c)}$
end for

V. PERFORMANCE EVALUATION

The simulation environment of an IEEE 802.16j relay network has been constructed to evaluate the proportional fairness of proposed algorithms, and to compare the system throughput of each algorithm.

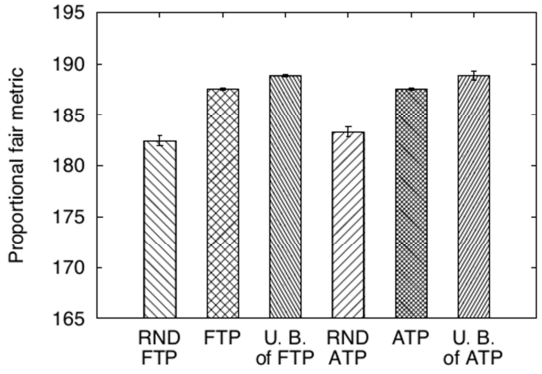
A. *Simulation Setup:* two scenarios for evaluation of performance of proposed algorithms.

Total CR BS =1 (located in center of the cell), Total CR RS = 4, Total CR MS = 20 or 40 in low system load or high system load case, Total 40 random topologies are generated,

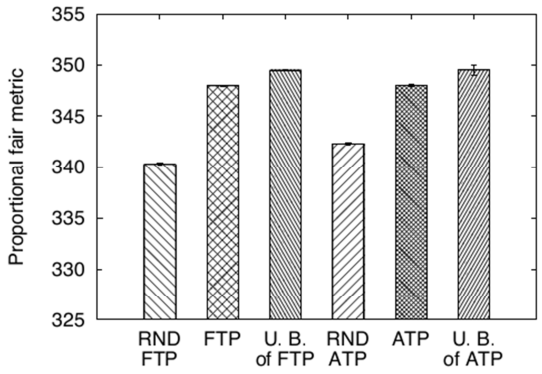
Total Frames = 2000, No of sub-channels = 64

Scenario 1 {purpose} Signal enhancement	**Scenario 2** {purpose} Range extension
CR RSs within 1200m of CR BS	CR RSs within 1500m of CR BS
CR MSs uniformly distributed in cell, with in 1800m of CR BS	CR MSs uniformly distributed with in 300m of randomly selected CR RS

B. Proportional Fairness:

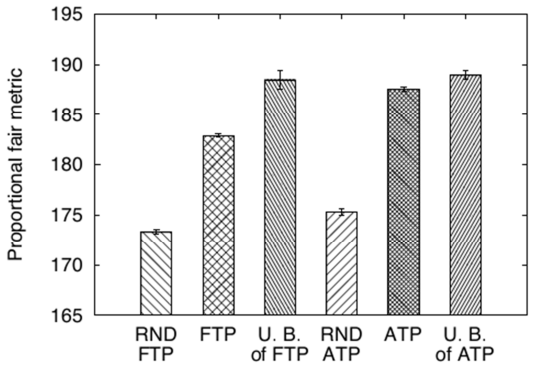


(a) Low system load

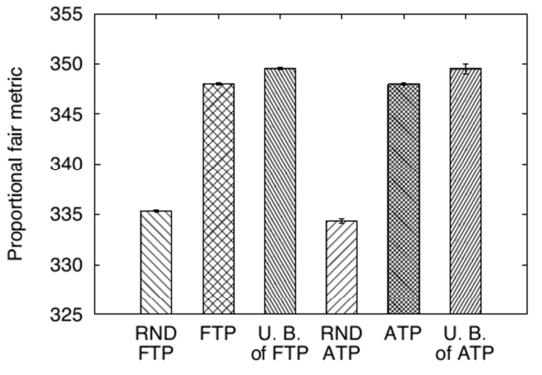


(b) High system load

Fig. 2. Proportional fair metrics of scenario 1 with 95% of confidenc interval. (RND: Random Scheduling; FTP: Fixed Transmit Power; U.B. Upper Bound; ATP: Adjustable Transmit Power.)



(a) Low system load



(b) High system load

Fig. 3. Proportional fair metrics of scenario 2 with 95% of confiden interval.(RND: Random Scheduling; FTP: Fixed Transmit Power; U.B.: Upp Bound; ATP: Adjustable Transmit Power.)

The upper bounds of FTP case in both figures are derived as that presented in III-A. The upper bounds of ATP in both figures are derived by using the power levels determined by Algorithm 2. The random scheduling is lower than proposed algorithms because in scenario 1, the CR RSs are placed closer to the CR BS. The channel condition of each CR MS would be less variant. On the contrary, in scenario 2, the CR RSs are placed far apart from each other as well as from the CR BS, and the channel condition of each CR MS varies a lot. However proposed algorithms are insensitive to the changes of scenario and system load.

C. *System Throughput:* Overall system throughput is the sum of the services that each CR MS receives during the whole simulation period.

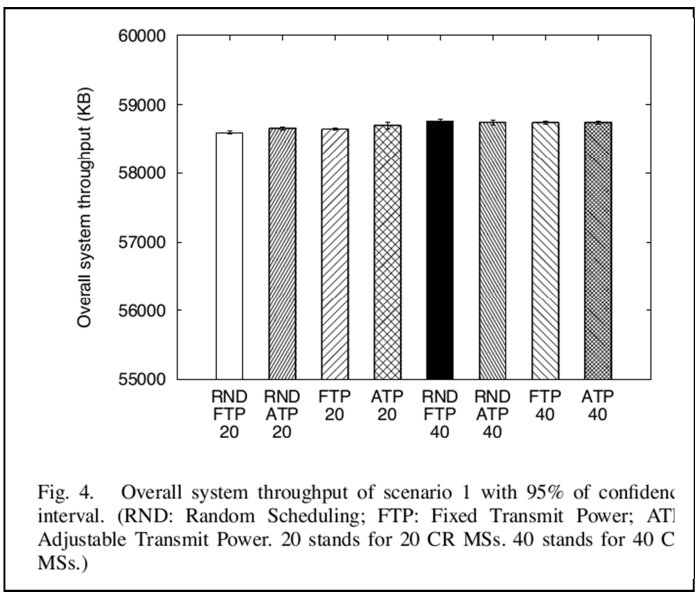


Fig. 4. Overall system throughput of scenario 1 with 95% of confidenc interval. (RND: Random Scheduling; FTP: Fixed Transmit Power; ATI Adjustable Transmit Power. 20 stands for 20 CR MSs. 40 stands for 40 C MSs.)

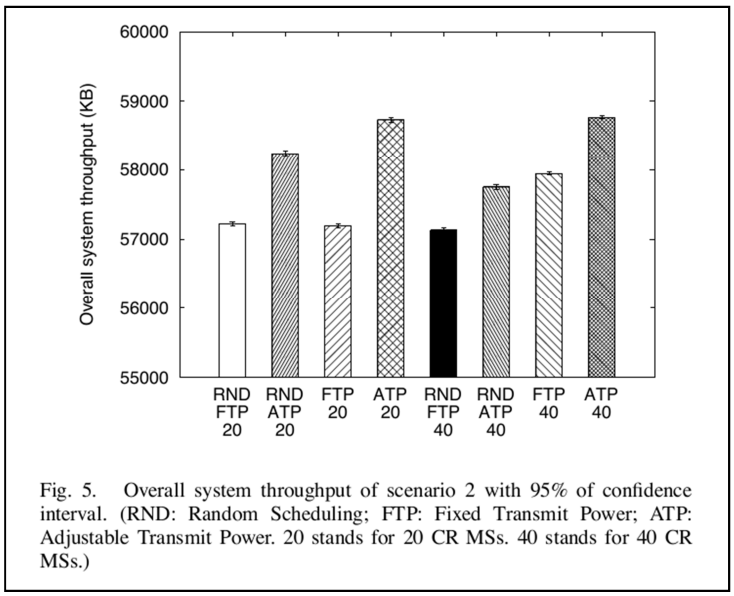


Fig. 5. Overall system throughput of scenario 2 with 95% of confidence interval. (RND: Random Scheduling; FTP: Fixed Transmit Power; ATP: Adjustable Transmit Power. 20 stands for 20 CR MSs. 40 stands for 40 CR MSs.)

In scenario 1, the overall system throughput of 20 and 40 CR-MSs does not differ much. The overall system throughput of FTP and ATP does not differ much, either. The reason is that CR RSs are placed closer to the CR BS. As a result, CR RSs will easily interfere with each other even with the lowest transmit power level. In such circumstance where cell coverage is small and CR MSs are evenly distributed, transmit power control will not help much. Also, in scenario 1, the channel condition of each CR MS does not differ much. Hence, the throughput does not differ much in all settings.
In scenario 2 the CR RSs are placed farther from the CR BS. Also, CR MSs tend to form hot-spot near the CR RSs. In systems where CR RSs are apart far enough and CR MSs are clustered in a certain extent, the interference among CR RSs will be highly controllable through adjustment of transmit power levels. Thus, as shown in Fig. 5, the performance of ATP is better than that of FTP.

D. *Algorithm Running Time:* The implemented algorithms are executed with 4 CR RSs, 40 CR MSs, 64 sub-channels, 48 time-slots per frame, and 2000 frames. The proposed FTP and ATP requires 0.38 ms and 0.75ms, respectively. however random FTP and random ATP require 0.14 ms and0.23ms, respectively. Although they are slightly less than proposed algorithms, yet proposed algorithms outperform random scheduling algorithms as shown in Figs. 2–5.

Multipath Matching Pursuit

Submitted to IEEE trans. on Information theory

Authors: S. Kwon, J. Wang, and B. Shim

Presenter: Hwanchol Jang

- Multipath is investigated rather than a single path for a greedy type of search
- In the final moment, the most promising path is chosen.
- They propose "breadth-first search" and "depth-first search" for greedy algorithm.
- They provide analysis for the performance of MMP with RIP

I. Introduction

CS

- The sparse signals $\mathbf{x} \in \mathbb{R}^n$ can be reconstructed from the compressed measurements $\mathbf{y} = \mathbf{\Phi x} \in \mathbb{R}^n$ even when the system representation is underdetermined ($m<n$), as long as the signal to be recovered is sparse (i.e., number of nonzero elements in the vector is small).

Reconstruction

1. L_0 mimimization

- K-sparse signal $\mathbf{x}$ can be accurately reconstructed using $m=2K$ measurements in a noiseless scenario [2].

2. L1 minimization

- Since ℓ_0-minimization problem is NP-hard and hence not so practical, early works focused on the reconstruction of sparse signals using ℓ_1-norm minimization technique (e.g., basis pursuit [2]).

3. Greedy search

- the greedy search approach is designed to further reduce the computational complexity of

the basis pursuit.

- In a nutshell, greedy algorithms identify the support (index set of nonzero elements) of the sparse vector x in an iterative fashion, generating a series of locally optimal updates.

OMP

- In the orthogonal matching pursuit (OMP) algorithm, the index of column that maximizes the magnitude of correlation between columns of Φ and the modified measurements (often called residual) is chosen as a new support element in each iteration.
- If at least one incorrect index is chosen in the middle of the search, the output of OMP will be simply incorrect.

II. MMP algorithm

L0 minimization

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{\Phi x} = \mathbf{y} \text{ . (1)}$$

OMP

- OMP is simple to implement and also computationally efficient
- Due to the choice of the single candidate it is very sensitive to the selection of index.
- The output of OMP will be simply wrong if an incorrect index is chosen in the middle of the search.

Multiple indices

- StOMP algorithm identifying more than one indices in each iteration was proposed. In this approach, indices whose magnitude of correlation exceeds a deliberately designed threshold are chosen [9].
- CoSaMP and SP algorithms maintaining K supports in each iteration were introduced.
- In [12], generalized OMP (gOMP), was proposed. By choosing multiple indices corresponding to N (> 1) largest correlation in magnitude in each iteration, gOMP reduces the misdetection probability at the expense of increase in the false alarm probability.

MMP

- The MMP algorithm searches *multiple promising candidates* and then chooses one minimizing the residual in the final moment.
- Due to the investigation of multiple full-blown candidates instead of partial ones, MMP improves the chance of selecting the true support.
- The effect of the random noise vector cannot be accurately judged by just looking at the partial candidate, and more importantly, incorrect decision affects subsequent decision in many greedy algorithms.
- MMP is effective in noisy scenario.

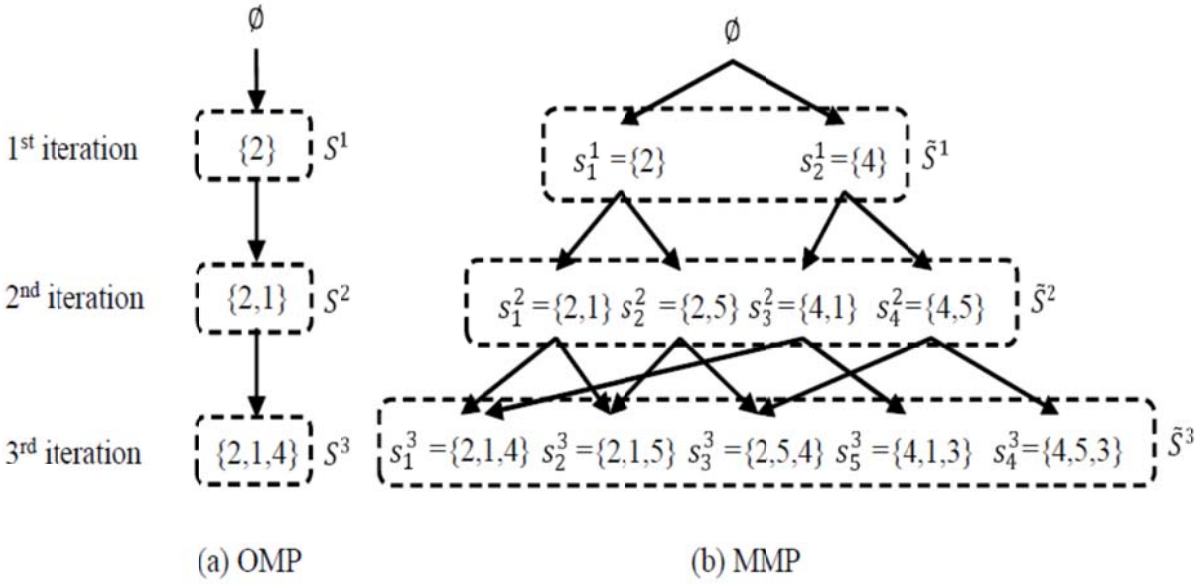


Fig. 1. Comparison between the OMP and the MMP algorithm ($L = 2$ and $K = 3$).

III. Perfect Recovery Condition for MMP

- A recovery condition under which MMP can accurately recover *K*-sparse signals in the noiseless scenario.
- two parts:
  - A condition ensuring the successful recovery in the initial iteration ($k = 1$).
  - A condition guaranteeing the success in the non-initial iteration ($k > 1$).

- By success we mean that an index of the true support T is chosen in the iteration.

RIP

- A sensing matrix $\mathbf{\Phi}$ is said to satisfy the RIP of order *K* if there exists a constant $\delta \in (0,1)$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \le \|\mathbf{\Phi x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2 \quad (2)$$

  for any *K*-sparse vector **x**.

- The minimum of all constants δ satisfying (2) is called the restricted isometry constant δ_K .

Lemma 3.1 (Monotonicity of the restricted isometry constant [1]): If the sensing matrix $\mathbf{\Phi}$ satisfies the RIP of both orders K_1 and K_2, then $\delta_{K_1} \le \delta_{K_2}$ for any $K_1 \le K_2$.

Lemma 3.2 (Consequences of RIP [1]): For $I \subset \Omega$, if $\delta_{|I|} < 1$ then for any $\mathbf{x} \in \mathbb{R}^{|I|}$,

$$(1-\delta_{|I|})\|\mathbf{x}\|_2 \le \|\mathbf{\Phi}_I{}'\mathbf{\Phi}_I\mathbf{x}\|_2 \le (1+\delta_{|I|})\|\mathbf{x}\|_2 \quad (3)$$

$$\frac{1}{1+\delta_{|I|}}\|\mathbf{x}\|_2 \le \|(\mathbf{\Phi}_I{}'\mathbf{\Phi}_I)^{-1}\mathbf{x}\|_2 \le \frac{1}{1-\delta_{|I|}}\|\mathbf{x}\|_2 \quad (4)$$

Lemma 3.3 (Lemma 2.1 in [19]): Let $I_1, I_2 \subset \Omega$ be two disjoint sets ($I_1 \cap I_2 = \varnothing$). If $\delta_{|I_1|+|I_2|} < 1$, then

$$\|\mathbf{\Phi}_{I_1}{}'\mathbf{\Phi}_{I_2}\mathbf{x}\|_2 \le \delta_{|I_1|+|I_2|}\|\mathbf{x}\|_2 \quad (5)$$

holds for any $\mathbf{x}$.

Lemma 3.4: For *m*×*n* matrix $\mathbf{\Phi}$, $\|\mathbf{\Phi}\|_2$ satisfies

$$\|\mathbf{\Phi}\|_2 = \sqrt{\lambda_{\max}(\mathbf{\Phi}'\mathbf{\Phi})} \le \sqrt{1+\delta_{\min(m,n)}} \quad (6)$$

A. Success Condition in Initial Iteration

In the first iteration, MMP computes the correlation between measurements $\mathbf{y}$ and each column ϕ_i of $\mathbf{\Phi}$ and then selects L indices whose column has largest correlation in magnitude. Let Λ be the set of L indices chosen in the first iteration, then

$$\left\|\mathbf{\Phi}'_{\Lambda}\mathbf{y}\right\|_2 = \max_{|I|=L} \sqrt{\sum_{i\in I} \left|\langle \phi_i, \mathbf{y}\rangle\right|^2}. \tag{7}$$

Following theorem provides a condition under which at least one correct index belonging to T is chosen in the first iteration.

Theorem 3.5: Suppose $\mathbf{x} \in \mathbb{R}^n$ is K-sparse signal, then among L candidates at least one contains the correct index in the first iteration of the MMP algorithm if the sensing matrix $\mathbf{\Phi}$ satisfies the RIP with

$$\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}. \tag{8}$$

Proof: From (7), we have

$$\begin{aligned}
\frac{1}{\sqrt{L}}\left\|\mathbf{\Phi}'_{\Lambda}\mathbf{y}\right\|_2 &= \frac{1}{\sqrt{L}} \max_{|I|=L} \sqrt{\sum_{i\in I} \left|\langle \phi_i, \mathbf{y}\rangle\right|^2} && (9)\\
&= \max_{|I|=L} \sqrt{\frac{1}{|I|}\sum_{i\in I} \left|\langle \phi_i, \mathbf{y}\rangle\right|^2} && (10)\\
&\geq \sqrt{\frac{1}{|T|}\sum_{i\in T} \left|\langle \phi_i, \mathbf{y}\rangle\right|^2} && (11)\\
&= \frac{1}{\sqrt{K}}\left\|\mathbf{\Phi}'_{T}\mathbf{y}\right\|_2 && (12)
\end{aligned}$$

where $|T| = K$. Since $\mathbf{y} = \mathbf{\Phi}_T\mathbf{x}_T$, we further have

$$\begin{aligned}
\left\|\mathbf{\Phi}'_{\Lambda}\mathbf{y}\right\|_2 &\geq \sqrt{\frac{L}{K}}\left\|\mathbf{\Phi}'_T\mathbf{\Phi}_T\mathbf{x}_T\right\|_2 && (13)\\
&\geq \sqrt{\frac{L}{K}}\left(1-\delta_K\right)\left\|\mathbf{x}\right\|_2 && (14)
\end{aligned}$$

where (14) is due to Lemma 3.2.

On the other hand, when an incorrect index is chosen in the first iteration (i.e., $\Lambda \cap T = \emptyset$),

$$\left\|\mathbf{\Phi}'_{\Lambda}\mathbf{y}\right\|_2 = \left\|\mathbf{\Phi}'_{\Lambda}\mathbf{\Phi}_T\mathbf{x}_T\right\|_2 \leq \delta_{K+L}\left\|\mathbf{x}\right\|_2, \tag{15}$$

where the inequality follows from Lemma 3.3. This inequality contradicts (14) if

$$\delta_{K+L} \|\mathbf{x}\|_2 < \sqrt{\frac{L}{K}} (1 - \delta_K) \|\mathbf{x}\|_2 . \tag{16}$$

In other words, under (16) at least one correct index should be chosen in the first iteration ($T_i^1 \in \Lambda$). Further, since $\delta_K \leq \delta_{K+N}$ by Lemma 3.1, (16) holds true if

$$\delta_{K+L} \|\mathbf{x}\|_2 < \sqrt{\frac{L}{K}} (1 - \delta_{K+L}) \|\mathbf{x}\|_2 . \tag{17}$$

Equivalently,

$$\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K} + \sqrt{L}}. \tag{18}$$

In summary, if $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+\sqrt{L}}$, then among L indices at least one belongs to T in the first iteration of MMP. ■

B. Success Condition in Non-initial Iterations

Now we turn to the analysis of the success condition for non-initial iterations. In the k-th iteration ($k > 1$), we focus on the candidate s_i^{k-1} whose elements are exclusively from the true support T (see Fig. 3). In short, our key finding is that at least one of L indices chosen by s_i^{k-1} is from T under $\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K}+3\sqrt{L}}$. Formal description of our finding is as follows.

Theorem 3.6: Suppose a candidate s_i^{k-1} includes indices only in T, then among L children generated from s_i^{k-1} at least one candidate chooses an index in T under

$$\delta_{K+L} < \frac{\sqrt{L}}{\sqrt{K} + 3\sqrt{L}}. \tag{19}$$

Before we proceed, we provide definitions and lemmas useful in our analysis. Let f_i be the i-th largest correlated index in magnitude between $\mathbf{r}^{k-1}$ and $\{\phi_j\}_{j \in T^C}$. That is, $f_j = \arg \max_{j \in T^C \setminus \{f_1, \ldots, f_{(j-1)}\}} \left| \left\langle \phi_j, \mathbf{r}^{k-1} \right\rangle \right|$. Let F_L be the set of these indices ($F_L = \{f_1, f_2, \cdots, f_L\}$). Also, let α_j^k be the j-th largest correlation in magnitude between the residual $\mathbf{r}^{k-1}$ associated with s_i^{k-1} and columns indexed by incorrect indices. That is,

$$\alpha_j^k = \left| \left\langle \phi_{f_j}, \mathbf{r}^{k-1} \right\rangle \right| . \tag{20}$$

Note that α_j^k are ordered in magnitude ($\alpha_1^k \geq \alpha_2^k \geq \cdots$). Finally, let β_j^k be the j-th largest correlation in magnitude between $\mathbf{r}^{k-1}$ and columns whose indices belong to $T - T_i^{k-1}$ (the set of remaining true indices). That is,

$$\beta_j^k = \left| \left\langle \phi_{\varphi(j)}, \mathbf{r}^{k-1} \right\rangle \right| \tag{21}$$

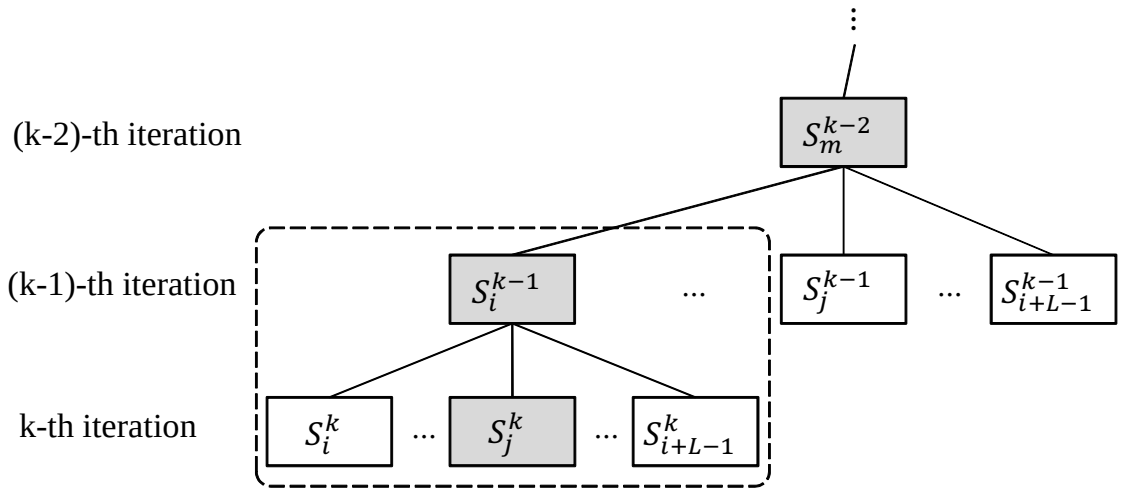


Fig. 3. Relationship between the candidates in $(k-1)$-th iteration and those in k-th iteration. Candidates inside the gray box contain elements of true support T only.

where $\varphi(j) = \arg \max_{j \in \left(T - T^{k-1}\right) \setminus \{\varphi(1), \ldots, \varphi(j-1)\}} \left|\left\langle \phi_j, \mathbf{r}^{k-1} \right\rangle\right|$. Similar to α_j^k, β_j^k are ordered in magnitude ($\beta_1^k \geq \beta_2^k \geq \cdots$). In the following lemmas, we provide the upper bound of α_L^k and lower bound of β_1^k.

Lemma 3.7: α_L^k satisfies

$$\alpha_L^k \leq \left(\delta_{L+K-k+1} + \frac{\delta_{L+k-1}\delta_K}{1-\delta_{k-1}}\right) \frac{\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2}{\sqrt{L}}. \tag{22}$$

Proof: See Appendix A. ■

Lemma 3.8: β_1^k satisfies

$$\beta_1^k \geq \left(1 - \delta_{K-k+1} - \frac{\sqrt{1+\delta_{K-k+1}}\sqrt{1+\delta_{k-1}}\delta_K}{1-\delta_{k-1}}\right) \frac{\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2}{\sqrt{K-k+1}}. \tag{23}$$

Proof: See Appendix B. ■

Proof of Theorem 3.6: From the definitions of α_j^k and β_j^k, it is clear that a (sufficient) condition under which at least one out of L indices is true in k-th iteration of MMP is

$$\alpha_L^k < \beta_1^k \tag{24}$$

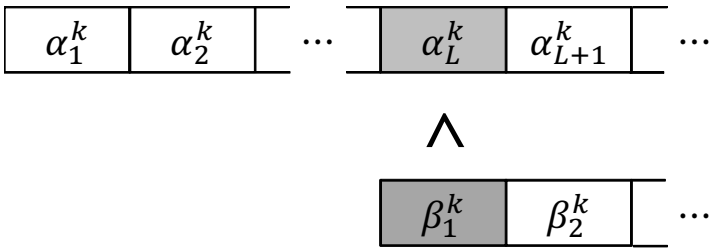


Fig. 4. Comparison between α_N^k and β_1^k. If $\beta_1^k > \alpha_N^k$, then among L indices chosen in K-iteration, at least one is from the true support T.

First, from Lemma 3.1 and 3.7, we have

$$\alpha_L^k \leq \left(\delta_{L+K-k+1} + \frac{\delta_{L+k-1}\delta_K}{1-\delta_{k-1}}\right) \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{L}} \tag{25}$$

$$\leq \left(\delta_{L+K} + \frac{\delta_{L+K}\delta_{L+K}}{1-\delta_{L+K}}\right) \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{L}} \tag{26}$$

$$= \frac{\delta_{L+K}}{1-\delta_{L+K}} \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{L}}. \tag{27}$$

Also, from Lemma 3.1 and 3.8, we have

$$\beta_1^k \geq \left(1 - \delta_{K-k+1} - \frac{\sqrt{1+\delta_{K-k+1}}\sqrt{1+\delta_{k-1}}\delta_K}{1-\delta_{k-1}}\right) \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{K-k+1}} \tag{28}$$

$$\geq \left(1 - \delta_{L+K} - \frac{(1+\delta_{L+K})\,\delta_{L+K}}{(1-\delta_{L+K})}\right) \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{K-k+1}} \tag{29}$$

$$= \frac{1-3\delta_{L+K}}{1-\delta_{L+K}} \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{K-k+1}}. \tag{30}$$

Using (24), (27), and (30), we can obtain the sufficient condition of (24) as

$$\frac{1-3\delta_{L+K}}{1-\delta_{L+K}} \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{K-k+1}} > \frac{\delta_{L+K}}{1-\delta_{L+K}} \frac{\left\|\mathbf{x}_{T-s_j^{k-1}}\right\|_2}{\sqrt{L}}. \tag{31}$$

From (31), we further have

$$\delta_{L+K} < \frac{\sqrt{L}}{\sqrt{K-k+1}+3\sqrt{L}}. \tag{32}$$

Since $\sqrt{K-k+1} < \sqrt{K}$ for $k > 1$, (32) holds under $\delta_{L+K} < \frac{\sqrt{L}}{\sqrt{K}+3\sqrt{L}}$, which completes the proof. ■

APPENDIX A
PROOF OF LEMMA 3.7

Proof: The ℓ_2-norm of the correlation $\mathbf{\Phi}'_{F_L}\mathbf{r}^{k-1}$ is expressed as

$$\begin{aligned}
\left\|\mathbf{\Phi}'_{F_L}\mathbf{r}^{k-1}\right\|_2 &= \left\|\mathbf{\Phi}'_{F_L}\mathbf{P}^{\perp}_{s_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (102)\\
&= \left\|\mathbf{\Phi}'_{F_L}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-s_j^{k-1}} - \mathbf{\Phi}'_{F_L}\mathbf{P}_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (103)\\
&\leq \left\|\mathbf{\Phi}'_{F_L}\mathbf{\Phi}_{T-s_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 + \left\|\mathbf{\Phi}'_{F_L}\mathbf{P}_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2. && (104)
\end{aligned}$$

Since F_L and $T-s_j^{k-1}$ are disjoint ($F_L \cap (T-s_j^{k-1}) = \emptyset$) and also noting that the number of correct indices in s_j^k is k by the hypothesis,

$$|F_L| + |T - s_j^{k-1}| = L + K - (k-1). \tag{105}$$

Using this together with Lemma 3.3,

$$\left\|\mathbf{\Phi}'_{F_L}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-s_j^{k-1}}\right\|_2 \leq \delta_{L+K-k+1}\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2. \tag{106}$$

Similarly, noting that $F_L \cap T_j^{k-1} = \emptyset$ and $|F_L| + |s_j^{k-1}| = L + k - 1$, we have

$$\left\|\mathbf{\Phi}'_{F_L}\mathbf{P}_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 \leq \delta_{L+k-1}\left\|\mathbf{\Phi}^{\dagger}_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 \tag{107}$$

where

$$\begin{aligned}
\left\|\mathbf{\Phi}^{\dagger}_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 &= \left\|\left(\mathbf{\Phi}'_{T_j^{k-1}}\mathbf{\Phi}_{T_j^{k-1}}\right)^{-1}\mathbf{\Phi}'_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (108)\\
&\leq \frac{1}{1-\delta_{k-1}}\left\|\mathbf{\Phi}'_{T_j^{k-1}}\mathbf{\Phi}_{T-T_j^{k-1}}\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (109)\\
&\leq \frac{\delta_{(k-1)+K-(k-1)}}{1-\delta_{k-1}}\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (110)\\
&= \frac{\delta_K}{1-\delta_{k-1}}\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2 && (111)
\end{aligned}$$

where (109) and (110) follow from Lemma 3.2 and 3.3, respectively. Since T_j^{k-1} and $T - T_j^{k-1}$ are disjoint, if the number of correct indices in T_j^{k-1} is $k-1$, then

$$\left|T_j^{k-1} \cup \left(T - T_j^{k-1}\right)\right| = (k-1) + K - (k-1). \tag{112}$$

Using (104), (106), (107), and (111), we have

$$\left\|\mathbf{\Phi}'_{F_L}\mathbf{r}^{k-1}\right\|_2 \leq \left(\delta_{L+K-k+1} + \frac{\delta_{L+k-1}\delta_K}{1-\delta_{k-1}}\right)\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2. \tag{113}$$

Using the norm inequality $\left(\|\mathbf{z}\|_1 \leq \sqrt{\|\mathbf{z}\|_0}\,\|\mathbf{z}\|_2\right)$, we further have

$$\begin{aligned}\left\|\mathbf{\Phi}'_{F_L}\mathbf{r}^{k-1}\right\|_2 &\geq \frac{1}{\sqrt{L}}\sum_{i=1}^{L}\alpha_i^k &(114)\\ &\geq \left\|\mathbf{\Phi}'_{F_L}\mathbf{r}^{k-1}\right\|_2 &(115)\\ &\geq \frac{1}{\sqrt{L}}L\alpha_L^k = \sqrt{L}\alpha_L^k &(116)\end{aligned}$$

where α_j^k is $\left|\left\langle\phi_{f_j},\mathbf{r}^{k-1}\right\rangle\right|$ [5] and $\alpha_1^k \geq \alpha_2^k \geq \cdots \geq \alpha_L^k$. Combining (113) and (116), we have

$$\left(\delta_{L+K-k+1} + \frac{\delta_{L+k-1}\delta_K}{1-\delta_{k-1}}\right)\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2 \geq \sqrt{L}\alpha_L^k, \tag{117}$$

and hence

$$\alpha_L^k \leq \left(\delta_{L+K-k+1} + \frac{\delta_{L+k-1}\delta_K}{1-\delta_{k-1}}\right)\frac{\left\|\mathbf{x}_{T-T_j^{k-1}}\right\|_2}{\sqrt{L}}. \tag{118}$$

■

APPENDIX B
PROOF OF LEMMA 3.8

Proof: Since β_1^k is the largest correlation in magnitude between $\mathbf{r}^{k-1}$ and $\{\phi_j\}_{j\in T-T_j^{k-1}}$ $\left(\left|\left\langle\phi_{\varphi(j)},\mathbf{r}^{k-1}\right\rangle\right|\right)$[6], it is clear that

$$\beta_1^k \geq \left|\left\langle\phi_j,\mathbf{r}^{k-1}\right\rangle\right| \tag{119}$$

for all $j \in T - T_j^{k-1}$, and hence

$$\begin{aligned}\beta_1^k &\geq \frac{1}{\sqrt{K-(k-1)}}\left\|\mathbf{\Phi}'_{T-T_j^{k-1}}\mathbf{r}^{k-1}\right\| &(120)\\ &= \frac{1}{\sqrt{K-k+1}}\left\|\mathbf{\Phi}'_{T-T_j^{k-1}}\mathbf{P}^{\perp}_{T_j^{k-1}}\mathbf{\Phi}\mathbf{x}\right\| &(121)\end{aligned}$$

[5] $f_j = \arg\max_{j\in T^C\setminus\{f_1,\ldots,f_{(j-1)}\}}\left|\left\langle\phi_j,\mathbf{r}^{k-1}\right\rangle\right|$

[6] $\varphi(j) = \arg\max_{j\in\left(T-T^{k-1}\right)\setminus\{\varphi(1),\ldots,\varphi(j-1)\}}\left|\left\langle\phi_j,\mathbf{r}^{k-1}\right\rangle\right|$

where (121) follows from $\mathbf{r}^{k-1} = \mathbf{y} - \mathbf{\Phi}_{T_j^{k-1}} \mathbf{\Phi}^{\dagger}_{T_j^{k-1}} \mathbf{y} = \mathbf{P}^{\perp}_{T_j^{k-1}} \mathbf{y}$. Using the triangle inequality,

$$\begin{aligned} \beta_1^k &\geq \frac{1}{\sqrt{K-k+1}} \left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{P}^{\perp}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (122) \\ &\geq \frac{\left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 - \left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2}{\sqrt{K-k+1}}. && (123) \end{aligned}$$

Since $\left| T - T_j^{k-1} \right| = K - (k-1)$,

$$\left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 \geq (1 - \delta_{K-k+1}) \left\| \mathbf{x}_{T-T_j^{k-1}} \right\| \quad (124)$$

and also

$$\begin{aligned} \left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 &\leq \left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \right\|_2 \left\| \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (125) \\ &\leq \sqrt{1+\delta_{K-k+1}} \left\| \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (126) \end{aligned}$$

where (126) follows from Lemma 3.4. Further, we have

$$\begin{aligned} & \left\| \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (127) \\ &= \left\| \mathbf{\Phi}_{T_j^{k-1}} \left(\mathbf{\Phi}'_{T_j^{k-1}} \mathbf{\Phi}_{T_j^{k-1}} \right)^{-1} \mathbf{\Phi}'_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (128) \\ &\leq \sqrt{1+\delta_{k-1}} \left\| \left(\mathbf{\Phi}'_{T_j^{k-1}} \mathbf{\Phi}_{T_j^{k-1}} \right)^{-1} \mathbf{\Phi}'_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (129) \\ &\leq \frac{\sqrt{1+\delta_{k-1}}}{1-\delta_{k-1}} \left\| \mathbf{\Phi}'_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (130) \\ &\leq \frac{\delta_{(k-1)+K-(k-1)} \sqrt{1+\delta_{k-1}}}{1-\delta_{k-1}} \left\| \mathbf{x}_{T-T_j^{k-1}} \right\|_2 && (131) \end{aligned}$$

where (129) and (130) are from the definition of RIP and Lemma 3.2. (131) follows from Lemma 3.3 and $\left| T_j^{k-1} \cup \left(T - T_j^{k-1} \right) \right| = (k-1) + K - (k-1)$ since T_j^{k-1} and $T - T_j^{k-1}$ are disjoint sets. Using (126) and (131), we obtain

$$\left\| \mathbf{\Phi}'_{T-T_j^{k-1}} \mathbf{P}_{T_j^{k-1}} \mathbf{\Phi}_{T-T_j^{k-1}} \mathbf{x}_{T-T_j^{k-1}} \right\|_2 \leq \frac{\sqrt{1+\delta_{K-k+1}} \sqrt{1+\delta_{k-1}} \delta_K}{1-\delta_{k-1}} \left\| \mathbf{x}_{T-T_j^{k-1}} \right\|_2. \quad (132)$$

Finally, by combining (123), (124) and (132), we have

$$\beta_1^k \geq \left(1 - \delta_{K-k+1} - \frac{\sqrt{1+\delta_{K-k+1}} \sqrt{1+\delta_{k-1}} \delta_K}{1-\delta_{k-1}} \right) \frac{\left\| \mathbf{x}_{T-T_j^{k-1}} \right\|_2}{\sqrt{K-k+1}}. \quad (133)$$

■

REFERENCES

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, April 2012.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007

[5] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.

[6] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.

[7] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sept. 2011.

[8] R. G. Baraniuk, M. A. Davenport, R. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, Dec. 2008.

[9] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[10] D. Needell and J. A. Tropp, "Cosamp: iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, no. 12, pp. 93–100, Dec. 2010.

[11] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[12] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[13] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, April 1967.

[14] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.

[15] B. Shim and I. Kang, "Sphere decoding with a probabilistic tree pruning," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4867–4878, Oct. 2008.

[16] B. M. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, March 2003.

[17] W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, 2013.

[18] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

[19] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, May 2008.

[20] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, pp. 317–334, June 2009.

The Exact Support Recovery of Sparse Signals with Noise via Orthogonal Matching Pursuit

Authors: Rui Wu, Wei Huang, and Di-Rong Chen

Publication: IEEE Sig. Proc. Letters, Apr. 2013
Speaker: J. Oliver

Short summary: This letter derives sufficient conditions for the OMP to recover the support set of a sparse vector from noise corrupted measurements. In particular, the conditions are given in terms of the minimum absolute values of the signal amplitudes. That is, if the minimum values of the non-zero coefficient of the signal satisfy certain bound then OMP guarantees exact support recovery**.**

I. System Model and Background

- Consider a model $\mathbf{y} = A\mathbf{x} + \mathbf{e}$, $\mathbf{y} \in R^m$ 정지 $A \in R^{m \times n}$ with $m \ll n$ and $\mathbf{x}$ is a *K*-sparse signal.
- Let A_i be the i*th* column of A and assume that $\|A_i\|_2 = 1, \quad i = 1, 2, \cdots, n$.
- Let $supp(x) = \{i \mid x_i \neq 0\}$ and $|supp(x)| \leq K$.
- The goal of OMP is to estimate the support of $\mathbf{x}$ iteratively.
- At each iteration, OMP selects a column of A that is most correlated with the current residual. OMP then updates the residual by projecting $\mathbf{y}$ onto a linear space spanned by the selected columns. The algorithm iterates until certain stopping rule is satisfied.

A. The OMP algorithm

Notations: For two sets Γ and Λ, let $\Gamma \setminus \Lambda = \{i \mid i \in \Gamma, i \notin \Lambda\}$ and $\Gamma^c = \{1, 2, \cdots, n\} \setminus \Gamma$. Let A_Γ denotes a sub-matrix whose column indices are elements of the set Γ and $\mathbf{x}_\Gamma$ denotes the elements of $\mathbf{x}$ whose indices are specified by Γ and $A_\Gamma^+ = (A_\Gamma^T A_\Gamma)^{-1} A_\Gamma^T$ represents the pseudo-inverse of A.

1. Initialize: Given A and $\mathbf{y}$, set the initial residual vector $\mathbf{r_0} = \mathbf{y}$ (that is $\mathbf{x}_0 = 0$), the initial index set as empty, $\Omega_0 = \Phi$ and the iteration counter t=1.
2. Find the index $i_t = \min_i |\langle A_i, \mathbf{r}_t \rangle|$ and update the support set estimate $\Omega_t = \Omega_{t-1} \cup i_t$
3. Estimate: $\mathbf{x}_t = A_{\Omega_t}^{+}\mathbf{y}$ and update the residual $\mathbf{r}_t = \mathbf{y} - A_{\Omega_t}\mathbf{x}_t$
4. Halt if some stopping rule is satisfied. Otherwise, set t=t+1 and return to step 2.

Stopping rule design for the OMP depends on noise. In noiseless case, (when $\mathbf{e} = 0$) the stopping rule can simply be $\mathbf{r}_t = 0$. This letter considers two types of bounded noises, namely, l_2 bounded noise, $\|\mathbf{e}\|_2 \le \epsilon_1$ and l_∞ bounded noise, $\|A^*\mathbf{e}\|_\infty \le \epsilon_2$. The stopping rules for these two noises in terms of residuals are $\|\mathbf{r}_t\|_2 \le \epsilon_1$ and $\|A^*\mathbf{r}_t\|_\infty \le \epsilon_2$, respectively. This paper also considers the case when e_i follows $\mathcal{N}(0, \sigma^2)$.

II. RIP AND A FEW ASSOCIATED LEMMAS

Two features of a sensing matrix are often used to analyze and derive the recovery performance guarantee of OMP. One is the Mutual Incoherence Property (MIP) [1] defined as $\mu = \max_{i \neq j} |\langle A_i, A_j \rangle|$. And, the other one is restricted isometry property (RIP).

- A matrix A satisfies RIP of order K with parameter δ_K if it is the smallest constant such that

$$(1-\delta_K)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_K)\|x\|_2^2 \tag{1}$$

  holds for any K-sparse vector $\mathbf{x}$.
- ***Lemma 1:*** Suppose that a matrix A satisfies RIP of order K. Let Γ be an index set with $|\Gamma| \le K$. Then all singular values of sub-matrix A_Γ, which are denoted by $\sigma_i(A_\Gamma)$, satisfy

$$\sqrt{1-\delta_K} \le \sigma_i(A_\Gamma) \le \sqrt{1+\delta_K} \tag{2}$$

- ***Remark 1:*** For any given matrix $B \in R^{m \times n}$, $\|B^*\|_2 = \|B\|_2$. Let $|\Gamma| \le K$, then

$$\|B_\Gamma^*\|_2 = \|B_\Gamma\|_2 = \max_i \sigma_i(B_\Gamma) \le \sqrt{1+\delta_K} \tag{3}$$

- ***Lemma 2:*** Suppose that a matrix *A* satisfies RIP of order *K*. Let Γ be an index set with $|\Gamma| \leq K$. Then all eigenvalues of matrix $A_\Gamma^* A_\Gamma$, which are denoted by $\sigma_i(A_\Gamma^* A_\Gamma)$, satisfy

$$1-\delta_K \leq \sigma_i(A_\Gamma^* A_\Gamma) \leq 1+\delta_K \tag{4}$$

- ***Lemma 3***: Suppose that a matrix *A* satisfies RIP of order *K*. Let Γ and Λ be two disjoint sets with $|supp(\Gamma) \cup supp(\Lambda)| \leq K$. Then for any vector $\mathbf{x}$ with $supp(\mathbf{x}) \subset \Lambda$, it holds that

$$\| (A_\Gamma^* A\mathbf{x})\|_2 = \| (A_\Gamma^* A_\Lambda \mathbf{x}_\Lambda)\|_2 \leq \delta_K \| \mathbf{x}\|_2 \tag{5}$$

Recovery conditions of OMP algorithm			
MIP		RIP	
Noiseless case	Noisy case	Noiseless case	Noisy case
$\mu < \frac{1}{2K-1}$ [3]	$\|x_{min}\| \geq \frac{2\sigma\sqrt{2(1+\alpha)\log N}}{1-(2K-1)\mu}$ [2]	$\delta_{K+1} < \frac{1}{\sqrt{K}+1}$ [4, 5]	$\mu < \frac{1}{2K-1}$ and $\|x_{min}\| \geq \frac{2\varepsilon_1}{1-(2K-1)\mu}$ [6] $\delta_{K+1} < \frac{1}{\sqrt{K}+3}$ and $\|x_{min}\| > \frac{2(1-\delta_{K+1})\epsilon_1}{(1-\delta_{K+1})^2 - \delta_{K+1}(1+\sqrt{K})}$ [7]

III. EXACT SUPPORT SET RECOVERY OF SPARSE SIGNALS

Condition

Let Ω be an original support set of the signal $\mathbf{x}$. Let $\mathbf{r}_{t-1}$ is the residual at the t*th* iteration, $t = 1, 2 \cdots, K$. The condition for OMP to select a correct index at t*th* iteration is

$$\boxed{\| A_{\Omega^c}^* \mathbf{r}_{t-1}\|_\infty < \| A_\Omega^* \mathbf{r}_{t-1}\|_\infty} \tag{6}$$

A. l_2 bounded noise

Theorem 1: Suppose that $\|\mathbf{e}\|_2 \le \epsilon_1$ and the matrix A satisfies condition $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Then OMP with stopping rule $\|\mathbf{r}_t\|_2 \le \epsilon_1$ will exactly recover the support Ω of K-sparse signal $\mathbf{x}$, if the minimum magnitude of nonzero elements of $\mathbf{x}$ satisfies

$$\min_{i\in\Omega} |\mathbf{x}_i| \ge \frac{(\sqrt{1+\delta_{K+1}}+1)\epsilon_1}{1-(\sqrt{K}+1)\delta_{K+1}} \tag{7}$$

Proof

- Suppose that OMP selects only correct indexes at the first t-1 iterations, then $\Omega_{t-1} \subseteq \Omega$ and the support of the solution $\mathbf{x}_{t-1}$ obtained at t-1*th* iteration is $supp(\mathbf{x}_{t-1}) \subseteq \Omega$ and $|supp(\mathbf{x}_{t-1})| \le t-1 \le K$.
- We can write the residual $\mathbf{r}_{t-1}$ as

$$\mathbf{r}_{t-1} = \mathbf{y} - A_{\Omega_{t-1}}\mathbf{x}_{t-1} = A_\Omega(\mathbf{x}-\mathbf{x}_{t-1}) + \mathbf{e} \tag{8}$$

- Our goal is to find the RHS and LHS of the condition $\|A_{\Omega^c}^*\mathbf{r}_{t-1}\|_\infty < \|A_\Omega^*\mathbf{r}_{t-1}\|_\infty$

- Let us start with the LHS, that is, $\|A_{\Omega^c}^*\mathbf{r}_{t-1}\|_\infty$

$$\begin{aligned}\|A_{\Omega^c}^*\mathbf{r}_{t-1}\|_\infty &= \|A_{\Omega^c}^*A_\Omega(\mathbf{x}-\mathbf{x}_{t-1}) + A_{\Omega^c}^*\mathbf{e}\|_\infty \\ &\le \|A_{\Omega^c}^*A_\Omega(\mathbf{x}-\mathbf{x}_{t-1})\|_\infty + \|A_{\Omega^c}^*\mathbf{e}\|_\infty \\ &= \max_{i\in\Omega^c} |\langle A_i, A_\Omega(\mathbf{x}-\mathbf{x}_{t-1})\rangle| + \max_{i\in\Omega^c} |\langle A_i, \mathbf{e}\rangle|\end{aligned} \tag{9}$$

- Now from Lemma 3 it holds for any $i \in \Omega^c$

$$|\langle A_i, A_\Omega(\mathbf{x}-\mathbf{x}_{t-1})\rangle| \le \delta_{K+1}\|(\mathbf{x}-\mathbf{x}_{t-1})\|_2$$

- Also, since $\|\mathbf{e}\| \le \epsilon_1$ *and* $\|A_i\|_2 = 1$, we have

$$|\langle A_i, \mathbf{e}\rangle| \le \|A_i\|_2\|\mathbf{e}\|_2 \le \epsilon_1$$

- Now, the LHS becomes

$$\|A_{\Omega^c}^*\mathbf{r}_{t-1}\|_\infty \le \delta_{K+1}\|(\mathbf{x}-\mathbf{x}_{t-1})\|_2 + \epsilon_1 \tag{10}$$

- Let us find the RHS, that is, $\| A_\Omega^* \mathbf{r}_{t-1} \|_\infty$. Let us recall that the residual $\mathbf{r}_{t-1}$ is orthogonal to the columns of $A_{\Omega_{t-1}}$, that is, $A_{\Omega_{t-1}}^* \mathbf{r}_{t-1} = 0$. Then

$$A_\Omega^* \mathbf{r}_{t-1} = \begin{bmatrix} A_{\Omega \setminus \Omega_{t-1}}^* \\ A_{\Omega_{t-1}}^* \end{bmatrix} \mathbf{r}_{t-1} = \begin{bmatrix} A_{\Omega \setminus \Omega_{t-1}}^* \mathbf{r}_{t-1} \\ \mathbf{0} \end{bmatrix}$$

- Thus, $A_\Omega^* \mathbf{r}_{t-1}$ has only $|\Omega \setminus \Omega_{t-1}| = K - (t-1)$ non-zero elements. By using the relation $\| \mathbf{x} \|_\infty \geq \frac{\| \mathbf{x} \|_2}{\sqrt{n}}$, we have

$$\| A_\Omega^* \mathbf{r}_{t-1} \|_\infty \geq \frac{\| A_\Omega^* \mathbf{r}_{t-1} \|_2}{\sqrt{K-(t-1)}}$$

- Now,

$$\begin{aligned} \| A_\Omega^* \mathbf{r}_{t-1} \|_2 &= \| A_\Omega^* A_\Omega (\mathbf{x} - \mathbf{x}_{t-1}) + A_\Omega^* \mathbf{e} \|_2 \\ &\geq \| A_\Omega^* A_\Omega (\mathbf{x} - \mathbf{x}_{t-1}) \|_2 - \| A_\Omega^* \mathbf{e} \|_2 \end{aligned}$$

- $\| A_\Omega^* A_\Omega (\mathbf{x} - \mathbf{x}_{t-1}) \|_2 \geq (1 - \delta_K) \| (\mathbf{x} - \mathbf{x}_{t-1}) \|_2 \geq (1 - \delta_{K+1}) \| (\mathbf{x} - \mathbf{x}_{t-1}) \|_2$ (Consequence of RIP)
- $\| A_\Omega^* \mathbf{e} \|_2 \leq \| A_\Omega^* \|_2 \| \mathbf{e} \|_2 \leq \sqrt{1+\delta_K} \epsilon_1 \leq \sqrt{1+\delta_{K+1}} \epsilon_1$
- Therefore, the RHS is lower bounded by

$$\| A_\Omega^* \mathbf{r}_{t-1} \|_\infty \geq \frac{(1-\delta_{K+1})}{\sqrt{K-(t-1)}} \| (\mathbf{x} - \mathbf{x}_{t-1}) \|_2 - \frac{\sqrt{1+\delta_{K+1}}}{\sqrt{K-(t-1)}} \epsilon_1 \qquad (11)$$

- $\| (\mathbf{x} - \mathbf{x}_{t-1}) \|_2 \geq \sqrt{K-(t-1)} \min_{i \in \Omega} |\mathbf{x}_i|$ (12)

- Using Eqns. (10), (11), and (12), we can find that for the condition (6) to be satisfied , the following inequality must hold true

$$\min_{i \in \Omega} |\mathbf{x}_i| \geq \frac{(\sqrt{1+\delta_{K+1}} + 1)\epsilon_1}{1 - (\sqrt{K}+1)\delta_{K+1}}$$

which is stated in Theorem 1.

- After all the K indexes in Ω have been identified, we find a new estimator via $\mathbf{x}_K = A^+ \mathbf{y}$. Then, the residual $\mathbf{r}_K$ obeys

$$\| \mathbf{r}_K \|_2 = \| \mathbf{y} - A_\Omega \mathbf{x}_K \|_2 \leq \| \mathbf{y} - A_\Omega \mathbf{x} \|_2 = \| \mathbf{e} \|_2 \leq \epsilon_1$$

- Therefore, OMP stops after *K* iterations during which the stopping rule is satisfied.

B. l_∞ bounded noise:

Theorem 2: Suppose that $\| A^* \mathbf{e} \|_\infty \le \epsilon_2$ and the matrix A satisfies condition $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Then OMP with stopping rule $\| A^* \mathbf{r}_t \|_\infty \le \epsilon_2$ will exactly recover the support Ω of K-sparse signal $\mathbf{x}$, if the minimum magnitude of nonzero elements of $\mathbf{x}$ satisfies

$$\min_{i \in \Omega} |\mathbf{x}_i| \ge \frac{(\sqrt{1+\delta_{K+1}}+1)\sqrt{K}\epsilon_2}{1-(\sqrt{K}+1)\delta_{K+1}}$$

C. Gaussian noise case

It is well known that when the noise in the model $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ follows iid Gaussian distribution with zero-mean and variance σ^2, then

$$P\left(\| \mathbf{e} \|_2 \le \sigma \sqrt{m + 2\sqrt{m \log m}} \right) \ge 1 - \frac{1}{m}$$

Theorem 3: Suppose that each element of the noise vector follows Gaussian with zero mean and variance σ^2 and the matrix A satisfies condition $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Then, OMP with stopping rule $\| \mathbf{r}_t \|_2 \le \sigma \sqrt{m + 2\sqrt{m \log m}}$ will exactly recover the support Ω of *K*-sparse signal $\mathbf{x}$ with probability at least 1-(1/m), if the minimum magnitude of nonzero elements of $\mathbf{x}$ satisfies

$$\min_{i \in \Omega} |\mathbf{x}_i| \ge \frac{(\sqrt{1+\delta_{K+1}}+1)\sigma\sqrt{m + 2\sqrt{m \log m}}}{1-(\sqrt{K}+1)\delta_{K+1}}$$

Remarks:

- In l_2 bounded noise case, the minimum magnitude of the *K*-sparse signal needs to be in the same order of the noise level.
- In l_∞ bounded noise case, the minimum magnitude needs to be about $\sqrt{K}$ times the noise level

- In the Gaussian case, the minimum magnitude depends on the size "*m*" of the matrix. Thus, *m* must be chosen to satisfy $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$ (for example say $m > K^2$), then the minimum magnitude need to be about *K* times of σ.

Reference

[1] D. Donoho and X. Huo, ``Uncertainty principles and ideal atomic decomposition," IEEE Trans. Info. Theory, vol. 47, no. 7, pp. 2845-2862, 2001.

[2] A. B. Heim, Y. Eldar, and M. Elad, ``Coherence-based performance guarantees for estimating a sparse vector under random noise," IEEE Trans. On Sig. Process. vol. 58, no.10, pp. 5030-5042, Oct. 2010.

[3] J. Tropp, ``Greed is good: Algorithmic results for sparse approximation,"IEEE Trans. Info. Theory, vol. 50, no. 10, pp. 2231-2242, 2004.

[4] Q. Mo and Y. Shen, ``A remark on the restricted isometry property in orthogonal matching pursuit," IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3654-3656, 2012.

[5] J. Wang and B. Shim, ``On the recovery limit of sparse signals using orthogonal matching pursuit," IEEE Trans. Signal Process., vol. 60, no. 9, pp. 4973-4976, Sept. 2012.

[6] T. Cai and L. Wang, ``Orthogonal matching pursuit for sparse signal recovery with noise," IEEE Trans. Inf. Theory, vol. 57, no. 7, pp.4680–4688, 2011.

[7] Y. Shen and S. Li, ``Sparse signals recovery from noisy measurements by orthogonal matching pursuit," Arxiv Preprint arXiv: 1105.6177, 2011.

Hierarchical and High-Girth QC LDPC codes

Authors: Yige Wang, Stark C. Draper, Jonathan S. Yedidia
Publication: IEEE T. Info.Theory, July 2013
Speaker: Jeong-Min Ryu

Short summary:

They present **an approach to designing capacity approaching high-girth low-density parity-check (LDPC) codes** that are friendly to hardware implementation, and compatible with some desired input code structure defined using a protograph. The approach is based on a mapping of any class of codes defined using a protograph into a family of hierarchical quasi-cyclic (HQC) LDPC codes. Next, they present **a girth-maximizing algorithm** that optimizes the degrees of freedom within the family of codes to yield a high-girth HQC LDPC code, subject to bounds imposed by the fact that HQC codes are still quasi-cyclic. Finally, they discuss **how certain characteristics of a code protograph will lead to inevitable short cycles** and show that **these short cycles can be eliminated** using a "squashing" procedure that results in a high-girth QC LDPC code.

(The "girth" of a code is the length of the shortest cycle in the code graph)

I. INTRODUCTION

1. The construction of LDPC codes

- Highly random graph construction
- Algebraic construction

1) Highly random graph construction

→ It can produce LDPC codes that closely approach the Shannon capacity

→ **Not easy to implement** in hardware as the irregular connections imply wiring complexity.

2) Algebraic construction

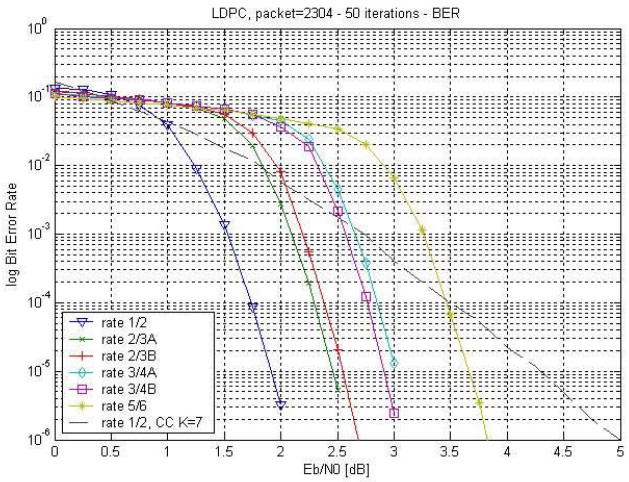
→ In actual implementations, more structured constructions have been strongly preferred

→ Quasi- cyclic LDPC (QC LDPC) codes are a particularly practical and widely used class of structured LDPC codes.

→ In view of the practicality, **they focus in this paper on the design of QC LDPC codes** that have good decoding performance

2. Optimizing the decoding performance

- Water-fall
- Error floor



1) Water-fall

→ **"**Water-fall" is a regime where the signal-to-noise (**SNR**) is relatively **low**.

→ The standard way to do that for irregular random constructions is to use "**density-evolution**" or "**EXIT chart**" techniques to obtain the degree distribution that **optimizes the code threshold in the asymptotic limit of long block lengths**

2) Error floor

→ An "error floor" in the performance curve means that **the decoding failure rate does not continue to decrease rapidly as the SNR increases**.

In this paper, they focus on how to take a code structure, such as a particular spatial-coupling structure, that has been designed to perform near the Shannon limit in the waterfall regime, and constructing a QC LDPC code with that structure that also empirically has excellent error floor performance.

II. QC LDPC CODES

- **Review of Standard QC LDPC codes**

QC LDPC codes are defined in terms of **circulant permutation matrices**. Let $I_{i,p}$ denote the circulant permutation matrix, or "cyclic shift matrix," obtained by cyclically left-shifting a $p \times p$ identity matrix by i positions, where $0 \le i \le p-1$; $I_{0,p}$ is thus the $p \times p$ identity matrix. We often suppress the dependence on p, writing I_i instead of $I_{i,p}$. As an example, if $p = 4$, then

$$I_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

An important special case of QC LDPC codes is "weight-I (J, L) regular" QC LDPC code. The parity check matrix of such a code consists of J rows and L columns of $p \times p$ cyclic shift submatrices. The submatrix in the jth row and lth column is $I_{i_{j,l}} = (I_1)^{i_{j,l}}$ and the code has blocklength $N = pL$. They abstractly represent the (j,l)th submatrix as a power of dummy variable x as $x^{i_{j,l}}$.

More generally, a QC LDPC code is represented by a polynomial parity check matrix $H(x)$ whose entries are polynomials in x:

$$H(x)=\begin{bmatrix} h_{1,1}(x) & h_{1,2}(x) & \cdots & h_{1,L}(x) \\ h_{2,1}(x) & h_{2,2}(x) & \cdots & h_{2,L}(x) \\ \vdots & & \ddots & \vdots \\ h_{J,1}(x) & h_{J,2}(x) & \cdots & h_{J,L}(x) \end{bmatrix}$$

where $h_{j,l}(x)=\sum_{s=0}^{p-1} c_s[j,l]x^s$ for $1\le j\le J, 1\le l\le L$, $c_s[j,l]\in\{0,1\}$.

Example 1: Let *C* be a length-9 QC LDPC code described by

$$\mathbf{H}=\left[\begin{array}{ccc|ccc|ccc} 1&0&0&1&0&0&1&0&0\\ 0&1&0&0&1&0&0&1&0\\ 0&0&1&0&0&1&0&0&1\\ \hline 0&0&0&1&0&0&0&1&1\\ 0&0&0&0&1&0&1&0&1\\ 0&0&0&0&0&1&1&1&0 \end{array}\right]$$

For this code, $J=2, L=3$, and $p=3$, and H can equivalently be written as

$$\mathbf{H}=\begin{bmatrix} \mathbf{I}_0 & \mathbf{I}_0 & \mathbf{I}_0 \\ \mathbf{0} & \mathbf{I}_0 & \mathbf{I}_1+\mathbf{I}_2 \end{bmatrix}.$$

The polynomial version of the parity check matrix is

$$\mathbf{H}(x)=\begin{bmatrix} x^0 & x^0 & x^0 \\ 0 & x^0 & x^1+x^2 \end{bmatrix}=\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & x^1+x^2 \end{bmatrix}.$$

For the maximum weight *M* among all polynomial entries $h_{j,l}(x)$ in $H(x)$, they call such a code a weight-*M* QC-LDPC code.

The code in Example 1 is a weight-II QC LDPC code.

III. Graphical Representations of QC LDPC Codes

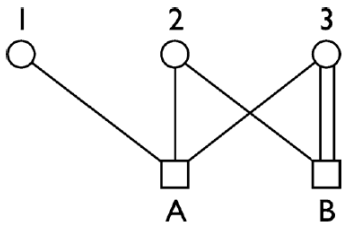


Fig. 1. A simple protograph with three types of variables and two types of checks.

A "**protograph**," as introduced by Thorpe in [30], is a template that can be used to derive a class of Tanner graphs. **Each node in a protograph represents a "type" of node** in a Tanner graph. The nodes will all be duplicated p times in the Tanner graph derived from the protograph.

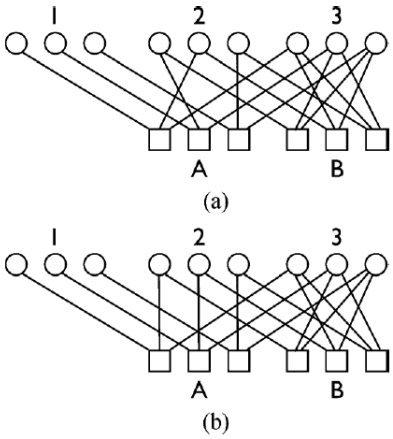


Fig. 2. Two Tanner graphs corresponding to the protograph shown in Fig. 1. The Tanner graph in (a) does not have a quasi-cyclic structure; the one in (b) does, and in fact has the parity check matrix of the QC LDPC code given in Example 1.

Fig. 2 shows two Tanner graphs derived from the protograph of Fig.1, with $p = 3$. Note that there are many possible Tanner graphs that one can construct, which correspond to a particular protograph, and they need not necessarily have a quasi-cyclic structure. The Tanner graph shown in Fig. 2(a) is not quasi- cyclic. But it is always easy to construct a quasi-cyclic version of any protograph.

Protographs can equivalently be described by an **"incidence" matrix**. An incidence matrix has a number of rows equal to the number of types of checks in the protograph and a number of

columns equal to the number of types of variables. Each entry in the incidence matrix tells you how many edges there are connecting a type of check node to a type of variable node in the protograph. For example, the incidence matrix P for the protograph in Fig.1 would be

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}.$$

- **Lifting procedure** (used to maximizes the girth of the code)

The lifting procedure is simply to replace each entry in the incidence matrix with a polynomial of weight equal to the entry.

For example, the protograph in Fig. 1, which has the incidence matrix P, can be lifted into a QC LDPC code with parity check matrix

$$H(x) = \begin{bmatrix} x^a & x^b & x^c \\ 0 & x^d & x^e + x^f \end{bmatrix},$$

where $a, b, c, d, e,$ and f are integer exponents between 0 and $p-1$, with $e \neq f$. These integer exponents **parameterize an ensemble of QC LDPC codes** all of which are liftings of (and which cover) the original protograph. In our algorithms, **they will optimize over the choice of these exponents to find a lifting that maximizes the girth of the resulting code**.

IV. Cycles in QC LDPC Codes

- How to identify cycles in QC LDPC codes from their parity check matrix
- For weight-I QC LDPC codes → For higher weight QC LDPC codes
- Review of an obstacle in constructing QC LDPC codes with good girth (The higher weight QC LDPC codes with certain characteristics are inevitable to have short cycles)
- HQC LDPC codes overcome the obstacle
- Applying a lifting transformation into HQC codes to obtain high-girth QC codes.

A. Finding Cycles in Weight-I QC LDPC codes

- **Cycle**

A cycle is a path through nodes in the Tanner graph of a code. It alternates between check and variable nodes, and starts and ends at the same node.

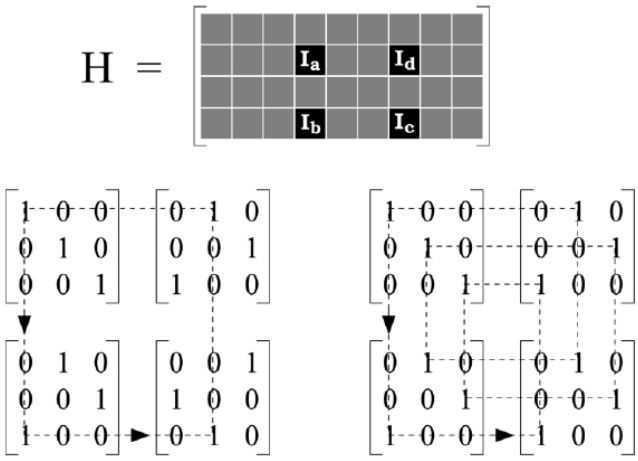


Fig. 3. A parity-check matrix and four 3×3 circulant permutation matrices ($\mathbf{I}_a$, $\mathbf{I}_b$, $\mathbf{I}_c$ and $\mathbf{I}_d$) selected from it. One set of parameters (lower left, $a = 0$, $b = 2$, $c = 1$, $d = 2$) results in a cycle of length four. An alternate set (lower right, $a = 0$, $b = c = d = 2$) results in a cycle of length twelve.

- **Condition of the cycles for weight-I QC LDPC codes**

They specify the conditions on the $\{a,b,c,d\}$ developed in [33] that result in a cycle. **Calculate an alternating sum of the shift indices associated with neighboring permutation matrices along a given path (every odd shift index is subtracted rather than added)**. For example, consider the left-hand path of Fig. 3. The sum is $-a+b-c+d$. Only if the differences sum to zero (mod-p) at the end of the path will the path return to the same variable node in the starting permutation matrix, thereby forming a cycle. For the example of Fig. 3, the condition for a length-four cycle to exist is:

$$(-a+b-c+d) \mod p = 0,$$

which is satisfied for $a=0, b=2, c=1, d=2$, but is not satisfied by $a=0, b=c=d=2$.

B. Finding Cycles in Higher Weight QC LDPC codes

Let us take the matrix $H(x)$ of Exmple 1,

$$H(x)=\begin{bmatrix} x^0 & x^0 & x^0 \\ 0 & x^0 & x^1+x^2 \end{bmatrix}.$$

Now, consider the following ordered series:

$$O=\{(1,2),(2,2),(2,3),(2,3),(2,3),(1,3)\}$$

where each pair (j,l) in O satisfies $1\le j\le J=2$ and $1\le l\le L=3$. This ordered series specifies a sequence of rectilinear moves through $H(x)$.

To specify a candidate cycle through the Tanner graph, we associate a coefficient index *s* with each pair (j,l) in O, such that $c_s[j,l]\neq 0$. They denote this series of coefficient indices by S. **The candidate cycle will be a cycle if the alternating sum of coefficient indices in *S* modulo *p* equals zero**.

In their example, consider the two following choices for the respective (ordered) sets of coefficient indices:

$$S_a=\{0,0,1,2,1,0\}$$
$$S_b=\{0,0,2,1,2,0\}.$$

Each of these choices corresponds to a cycle of length-6 through the Tanner graph of the code. The alternating sums modulo-3 can be verified to be equal to zero. Respectively, these sums are

$$(-0+0-1+2-1+0)\bmod 3=0$$
$$(-0+0-2+1-2+0)\bmod 3=0.$$

C. Invertible Cycles in Higher Weight QC LDPC codes

An important theorem proven by Smarandache and Vontobel [35] states that any weight-III QC LDPC code will inevitably contain cycles of length six. Suppose that, without loss of generality, the polynomial $h_{j,l}(x)$ is weight-III and has the form $x^a + x^b + x^c$. To see that a cycle must exist using thier notation, choose the length-six ordered series

$$O = \{(j,l),(j,l),(j,l),(j,l),(j,l),(j,l)\},$$

and choose $S = \{a,b,c,a,b,c\}$. We find that

$$(-a+b-c+a-b+c) \mod p = 0,$$

for any value of *p*.

One can also prove (see [35, Th. 17] or [27, Example 3.3]) that a parity check matrix of a weight-II QC LDPC code that contains two weight-2 polynomials in the same row or the same column will inevitably have eight-cycles. To see this, suppose the two weight-2 polynomials are in the same row j, but in two different columns $l_1 \neq l_2$. Let $h_{j,l_1} = x^a + x^b$ and $h_{j,l_2} = x^c + x^d$. Consider the length-eight ordered series

$$O = \{(j,l_1),(j,l_1),(j,l_2),(j,l_2),(j,l_1),(j,l_1),(j,l_2),(j,l_2)\}$$

and choose

$$S = \{a,b,c,d,b,a,d,c\}.$$

We again find that

$$(-a+b-c+d-b+a-d+c) \mod p = 0,$$

regardless of the value of *p*.

These inevitable six-cycles and eight-cycles appear to put serious limitations on what protographs can be converted into quasi-cyclic codes with high girth.

V. HQC LDPC CODES

- To solve the problem of invertible short cycles, they introduce HQC LDPC codes.
- An HQC LDPC code is formed from "levels" that each has a quasi-cyclic structure. The structure can be specified in two forms:
  1) Polynomial parity check matrices
  2) Tree structure
- They connect the hierarchical structure to a particular sequence of liftings of a base graph.

A. Parity Check Matrices of HQC LDPC Codes

Example 2: Consider the polynomial parity check matrix specified in (18) with $p=8$. Because the highest weight of any of the polynomial entries is 2, and because there are 12 columns in the matrix, this is a length-96 weight-II QC LDPC code

$$\mathbf{H}(x)=\left[\begin{array}{ccc|ccc|ccc|ccc} x^6 & 0 & x^1+x^7 & 0 & x & 1+x^2 & 0 & 0 & 0 & x^6 & x^3 & 1 \\ x^1+x^7 & x^6 & 0 & 1+x^2 & 0 & x & 0 & 0 & 0 & 1 & x^6 & x^3 \\ 0 & x^1+x^7 & x^6 & x & 1+x^2 & 0 & 0 & 0 & 0 & x^3 & 1 & x^6 \\ \hline 0 & x & 1+x^2 & x^6 & 0 & x^1+x^7 & x^6 & x^3 & 1 & 0 & 0 & 0 \\ 1+x^2 & 0 & x & x^1+x^7 & x^6 & 0 & 1 & x^6 & x^3 & 0 & 0 & 0 \\ x & 1+x^2 & 0 & 0 & x^1+x^7 & x^6 & x^3 & 1 & x^6 & 0 & 0 & 0 \end{array}\right] \tag{18}$$

$$\mathbf{H}(x,y)=\left[\begin{array}{c|c|c|c} x^6+(x+x^7)y & (1+x^2)y+xy^2 & 0 & x^6+y+x^3y^2 \\ \hline (1+x^2)y+xy^2 & x^6+(x+x^7)y & x^6+y+x^3y^2 & 0 \end{array}\right] \tag{19}$$

$$\mathbf{H}(x,y,z)=\left[x^6+(x+x^7)y+((1+x^2)y+xy^2)z \;\middle|\; (x^6+y+x^3y^2)z\right] \tag{20}$$

Each of the three contractions of the parity check matrix of this code into the polynomial parity check matrices represented by (18), (19), and (20), corresponds to a "level" in the hierarchy of this **three-level** HQC LDPC code.

We now present a formal definition of the family of *K*-level HQC LDPC codes which generalizes our example.

Definition 1: An HQC LDPC code with K levels is defined by a $J_{[k]} \times L_{[k]}$ multivariate polynomial parity check matrix $H(\cdot)$ in *K* variables. The entry in the *j*th row and *l*th column of $H(\cdot)$, $1 \le j \le J_{[k]}$, $1 \le l \le L_{[k]}$ is a *K*-variate polynomial $h_{j,l}(\cdot,\cdots,\cdot)$ over the *K* variables $x_{[1]},\ldots,x_{[k]}$. With these definitions, we defined the code by the $J_{[k]} \cdot L_{[k]}$ polynomials

$$h_{j,l}(x_{[1]},\ldots,x_{[K]}) = \sum_{s_K=0}^{p_{[K]}-1} \cdots \sum_{s_1=0}^{p_{[1]}-1} c_{s_1,\ldots,s_K}[j,l] \left(\prod_{k=1}^{K} x_{[k]}^{s_k} \right).$$

Example) We can rewrite the term $h_{1,1}(x,y,z)$ of (20) as

$$\mathbf{H}(x,y,z) = \left[x^6 + (x+x^7)y + ((1+x^2)y + xy^2)z \;\middle|\; (x^6 + y + x^3y^2)z \right] \tag{20}$$

$$\begin{aligned} & h_{1,1}(x_{[1]},x_{[2]},x_{[3]}) \\ &= x_{[1]}^6 + \left(x_{[1]} + x_{[1]}^7\right) x_{[2]} + \left(\left(1 + x_{[1]}^2\right) x_{[2]} + x_{[1]} x_{[2]}^2\right) x_{[3]} \\ &= \sum_{s_3=0}^{1} \sum_{s_2=0}^{2} \sum_{s_1=0}^{7} c_{s_1,s_2,s_3}[1,1] x_{[1]}^{s_1} x_{[2]}^{s_2} x_{[3]}^{s_3}, \end{aligned}$$

where all coefficients $c_{s1,s2,s3}[1,1]$ are zero except for

$$c_{6,0,0}[1,1] = c_{1,1,0}[1,1] = c_{7,1,0}[1,1] = c_{0,1,1}[1,1] = c_{2,1,1}[1,1] = c_{1,2,1}[1,1] = 1.$$

B. Tree Structure of HQC LDPC Codes

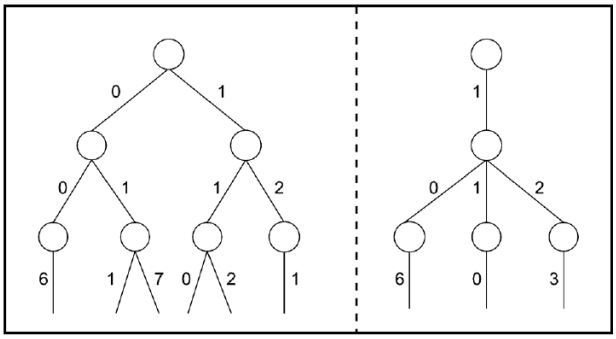


Fig. 4. Example of the tree structure of a family of three-level HQC LDPC codes. The left-hand tree is $\mathbf{T}_{1,1}$, and the right-hand tree is $\mathbf{T}_{1,2}$.

Remained contests

- **Finding cycles in HQC LDPC codes**
- **Inevitable cycles in HQC LDPC codes**

- **Proposing girth maximization using hill climbing**

- **Design of restricted two-level HQC LDPC codes (The additional "restriction" is that the weight of the first(lowest) level must be one)**
  →The restricted two-level HQC LDPC codes can considered weight-I QC LDPC codes
- **Squaring sets of trees to eliminate inevitable cycles**

- **Design of high-girth codes**

Numerical Result

In Figs. 9, 10, and 11, they plot the respective error rate performance of the three codes for the binary symmetric channel (BSC). For purposes of comparison, they plot analogous results for some randomly generated girth-6 QC LDPC codes. These codes have the same length, same rate, and same nonzero positions in the base matrix (i.e., same protograph structure) as the girth-10 and girth-8 codes to which they are compared.

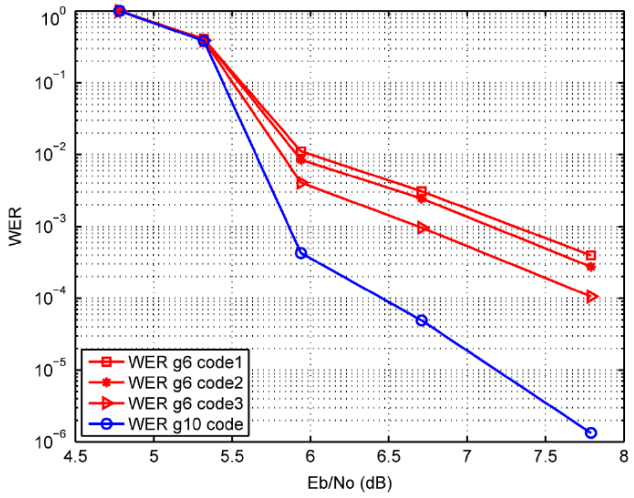


Fig. 9. Word-error rate plots of the Gallager-B algorithm for the rate-0.45, length-8000 girth-6 and girth-10 QC LDPC codes over the BSC.

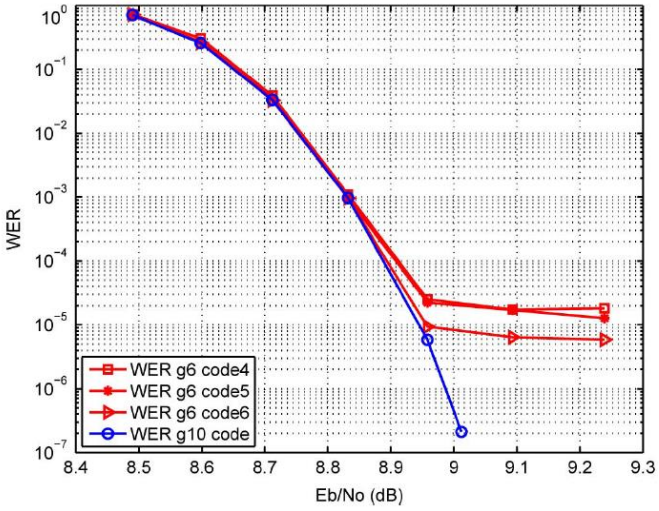


Fig. 10. Word-error rate plots of the Gallager-B algorithm for the rate-1/3, length-24000 girth-6 and girth-10 QC LDPC codes over the BSC.

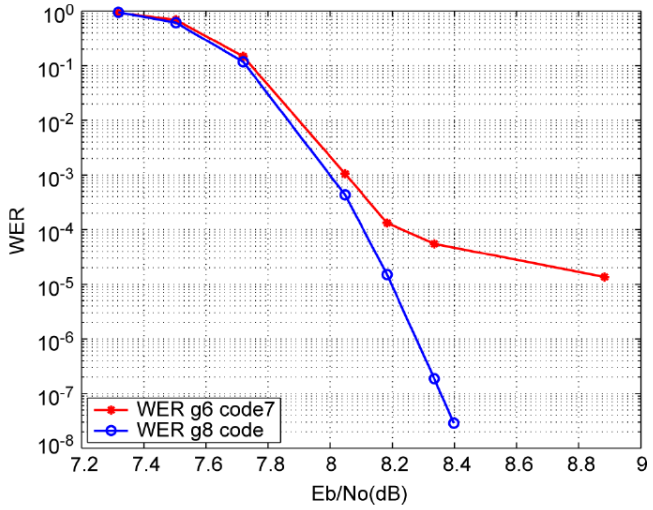


Fig. 11. Word-error rate plots of the Gallager-B algorithm for the rate-0.7, length-28000 girth-6 and girth-8 QC LDPC codes over the BSC.

- High girth or High rate → Low error floor

Ultra-Wideband Compressed Sensing : Channel Estimation

Authors: Jose L. Paredes, Member, IEEE, Gonzalo R. Arce, Fellow, IEEE, and Zhongmin Wang.
Publication: IEEE Journal of Selected Topics in Signal Processing, Octovber 2007
Speaker: Ju-Sung Kang

Short summary:

In this paper, they have introduced two novel ultra-wideband (UWB) channel estimation approaches based on compressive sensing (CS).

The proposed approach relies on the fact that transmitting an ultra-short pulse through a multipath UWB channel leads to a received UWB signal that can be approximated by a linear combination of a few atoms from a pre-defined dictionary which means sparse representation of the received signal.

The key in the proposed approach is in the design of a dictionary of parameterized waveforms (atoms) that closely matches the information-carrying pulse shape leading thus to higher energy compaction and sparse representation, and, therefore higher probability for CS reconstruction.

In the first approach, the CS reconstruction capabilities are exploited to recover the composite pulse-multipath channel from a reduced set of random projections. This reconstructed signal is subsequently used as a referent template in a correlator-based detector.

In the second approach, from a set of random projections of the received pilot signal, the Matching Pursuit algorithm is used to identify the strongest atoms in the projected signal that are related to the strongest propagation paths that composite the multipath UWB channel.

I. INTRODUCTION

1. Ultra-wideband (UWB) communications

- High bandwidth, lower-power consumption, shared spectrum resources, ranging from short-distance high-data-rate application to long-distance low-data-rate application.

- An ultra-short duration pulse is used as the elementary pulse-shaping to carry information → simplicity in the transmitter (carry-less signal), little impact on other narrowband radio system, rich in multipath diversity.

- Interference cancellation, antenna design, timing synchronization, and channel estimation. → requirement of high-speed ADC converters. : Such formiable sampling rates are not feasible with state of the art ADC technology.

- This paper focuses on this goal by casting the problem of USB channel estimation and detection into the emerging framework of CS.

2. Compressed sensing

- The remarkable result of CS reveals that with high probability, a signal, f, with a large number of data points that is M-sparse in some dictionary Ψ of basis functions, can be exactly reconstructed using only a few number of random projections of the signal onto a random basis Φ that is incoherent with Ψ.

- The number of projections is much smaller than the number of samples in the original signal leading to a reduced sampling rate and to a reduced use of ADCs resources.

3. Basic assumption

- When the short duration pulses propagate through multipath channels, the received signals remain sparse in some domain and thus CS is applicable.

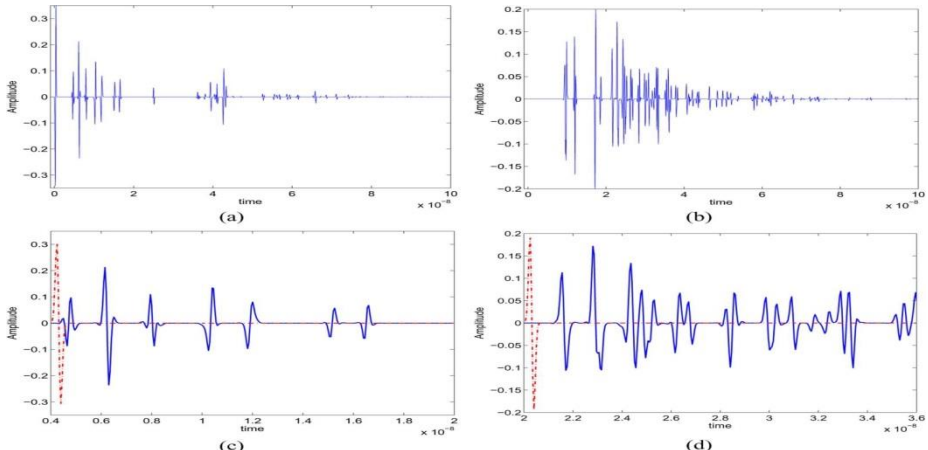


Fig. 1. Effect of UWB channel (indoor propagation in residential environments) on the transmitted pulse for two different propagation scenarios; (a) line-of-sight (LOS); (b) non-line-of-sight (NLOS); (c) zoom-in of (a); and (d) zoom-in of (b). Transmitted pulse (–.–) is also shown in (c) and (d).

- Gaussian monocycle($0.65\,\text{ns}$), IEEE 802.15.4a channel model 1 and 2(CM1, CM2).

- As depicted in above figure, the received UWB signal is composed of set of spaced clusters of the transmitted pulse which captures the statistical characteristics of multipath arivals in a UWB channel.

- It can be seen relatively long time intervals between clusters and rays where the signal takes on zero or negligible values. It is precisely this signal sparsity of the received UWB signals that is exploited in this work.

II. Ultra-Wideband Compressive Sensing.

- The sparsity of the signal can be in any domain and the number of random measurements is much smaller than the number of samples in the original signal leading to a reduced sampling rate and reduced use of ADCs resources.

1. Compressive sensing overview.

- f : N-point discrete-time representation of signal.
- y : a set of K measurements $y = \Phi f$
- Φ : $K \times N$ measurement matrix, rows are basis vectors of the space R^N
- If f is sparse, f can be written as a superposition of a small number of vector taken from a dictionary $\Psi = [\psi_1, ..., \psi_Z]$ of basis

$$f = \sum_{i=1}^{M} \theta_{l_i} \psi_{l_i} = \Psi\Theta \quad (1)$$

- $K << N$, and measurement matrix Φ is incoherent with the dictionary Ψ.
- $\Theta = [\theta_1, ..., \theta_Z]^T$ is a vector that contains M nonzeros coefficients where Z is the number of elements (atoms) in the dictionary Ψ.

- The signal f can be recovered from the solution of convex, nonquadratic optimization problem known as basis pursuit.

- But solving the optimization problem is computationally expensive and is not suitable for real-time application. So, there are more efficient recovery algorithms such as matching pursuit, orthogonal matching pursuit, and tree-based matching pursuit.

TABLE I : MATCHING PURSUIT ALGORITHM

Step A	Initialize: the residual error $e_0 = y$ the approximation $\hat{\Theta} = 0, \hat{\Theta} \in R^Z$ Interation counter $t = 1$
Step B	Select the atom in the holographic dictionary that best match the residual error. $\ell_t = \underset{i=1,2,\ldots,Z}{\arg\max} \frac{\lvert\langle e_{t-1}, v_i\rangle\rvert}{\lVert v_i \rVert}$
Step C	Update the residual error and the estimate of the coefficient for the selected vector $e_t = e_{t-1} - \frac{\langle r_{t-1}, v_{\ell_t}\rangle}{\lVert v_{\ell_t}\rVert^2} v_{\ell_t}$ $\hat{\theta}_{\ell_t} = \hat{\theta}_{\ell_t} + \frac{\langle r_{t-1}, v_{\ell_t}\rangle}{\lVert v_{\ell_t}\rVert^2}$
Step D	Check for convergence If $t < T_0$ and $\lVert e_t \rVert_2 > \epsilon \lVert y \rVert_2$ then set $t = t + 1$ and go to step B; otherwise, go to step E.
Step E	Reconstruct the signal estimate as: $\hat{f} = \Psi\hat{\Theta}$.

- MP is a computationally simple iterative greedy algorithm that tries to recover the signal by finding(in the measurement signal) the strongest component (atom of dictionary), removing it from the signal, and searching again the dictionary for the strongest atom that is presented in the residual signal.

- This procedure is iteratively repeated until the residual signal contains just insignificant information.

- Signal reconstruction is then achieved by linearly combining the set of atoms found in the measurements.

- $V = \Phi\Psi = [v_1, \ldots, v_Z]$, T_0 : maximum # of algorithm iterations, ε : the minimum energy that is left in the residual error signal.

2. Processing UWB signals Using CS.

- The received UWB signal model

$$g(t) = p(t) * h(t) = \sum_{l=0}^{L-1} \alpha_l p(t - \tau_l) \quad (3)$$

- $p(t)$: transmitting pulse, $h(t)$: noiseless UWB channel.

- We call it as composite pulse-multipath channel.

- Typically, a Gaussian pulse or its derivatives are used as $p(t)$.

- $p(t) = p_n(t) e^{-\frac{t^2}{2\sigma^2}}$, $p_n(t)$ is a polynomial of degree n that depends on the order of the derivative used.

- $h(t)$ is the impulse response of the UWB channel

$$h(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l) \quad (4)$$

- α_l : gain factor, τ_l : delay factor, L : # of propagation paths.

- In our analysis, the set of delays and gains are generated according to the models proposed by the IEEE 802.15.4a working group in [15] . But we restrict our analysis to real-valued UWB channel models where there is not pulse distortion.

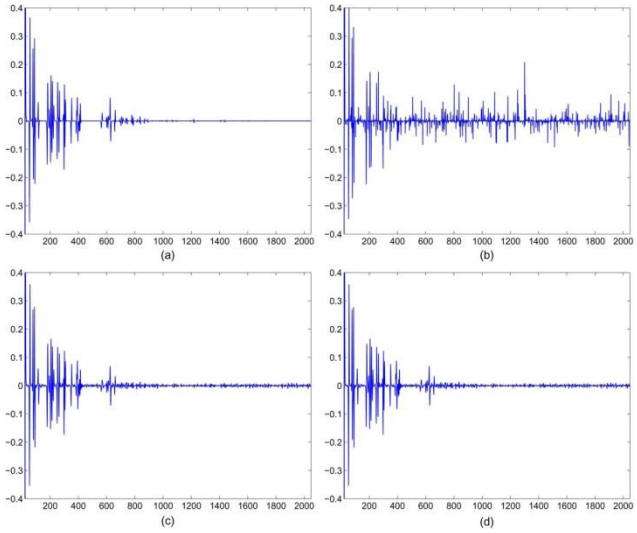


Fig. 2. (a) Received UWB signal for a realization of an indoor residential channel with LOS propagation (CM1). (b) CS reconstruction using time-sparsity model, with 500 random projections. (c) CS reconstruction using multipath diversity, with 500 random projections. (d) CS reconstruction using multipath diversity, with 250 random projections.

1) UWB signal reconstruction Using Time Sparsity Models :

- A first approach is assuming that the signal is sparse in the time domain.

- This signal model is adequate for the UWB channel in industrial environments with LOS propagation.

- $g = [g(0), g(T), ..., g((N-1)T)]^T$

- T : sampling period, N # of samples

- $\Phi \sim N(0,1)$, K*N random matrix with entries i.i.d.

- Since we are assuming sparsity in the time domain, the dictionary $\Psi = I$

- Running the MP algorithm with the $V = \Phi$ and the random projection $y = \Phi g$ yields the results show in fig.2.

- Fig. 2(a) : the 2048-point channel for a realization of an indoor residential channel with LOS propagation obtained from [15]. This is the signal targeted for reconstruction from a reduced set of random projections.

- Fig. 2(b) : the reconstructed signal obtained using 500 random measurements. Note that it fails to recover many of the signal details yielding a poor performance.

- Increasing the # of random projection means that higher sampling rate and demanding ADC resources.

- Appealing approach : to design a dictionary of parameterized waveforms where the received UWB signal can be compactly represented, increasing thus the sparsity of the underlying signal.

- This approach is motivated by the fact that the received UWB signal given by (3) can be thought of as a linear combination of the signal contributions of the various propagation paths that compose the UWB multipath channel.

2) UWB signal Reconstruction Using Multipath Diversity.

- Since CS theory relies on the fact that the underlying signal is sparse in some dictionary of basis or tight-frames, it is important to define a suitable dictionary to represent the underlying UWB signal.

- Alternatively, we can generate a new dictionary just inspecting the characteristic of the received UWB waveform.

- Since the received UWB signal is formed by scaled and delayed versions of the transmitted pulse and since the dictionary should contain elements (atoms) that can fully represent the signal of interest, it is natural to think that the elementary function to generate the atoms of the dictionary should be closely related to the pulse waveform used to covey information, i.e., the Gaussian pulse or its derivatives.

- Therefore, the dictionary is generated by shifting with minimum step Δ the generating function, $p(t)$, leading to a set of parameterized waveforms given by

$$d_j(t) = p(t - j\Delta) = p_n(t - j\Delta)e^{-\frac{(t-j\Delta)^2}{2\sigma^2}} \qquad j = 0,1,2,... \quad (5)$$

- Dictionary $D = \{d_0(t), d_1(t), ...\}$: delayed versions of the UWB transmitted pulse.

- The other definitions are same with Time Sparsity Model cases.

- $g = [g(0), g(T), ..., g((N-1)T)]^T$, T : sampling period, N # of samples, $\Phi \sim N(0,1)$, K*N random matrix with entries i.i.d.

- The MP algorithm is then applied on the random projected signal, y, and the dictionary Ψ

- Ψ is the discrete time dictionary defined by uniformly sampling the atoms of the dictionary D.

- Fig. 2(c) and (d) show the reconstructed signal using 500 and 250 random measurements, respectively. As it can be seen from Fig. 2(c) and (d), CS successfully recovers the desired signal from random projections

- Furthermore, comparing Fig. 2(b) and (c), it can be seen that reconstruction using multipath diversity outperforms reconstruction using time sparsity model

- Therefore, by building a dictionary that is closely matched to the underlying waveform, a notable performance gain is achieved in the reconstruction

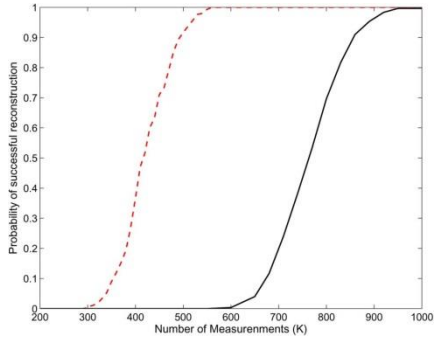


Fig. 3. Probability of success reconstruction for UWB signal for two different propagation scenarios: LOS - - - and NLOS —.

3. UWB Channel Estimation Using CS

- Consider the composite pulse-multipath channel, given by (3), where the channel parameters $\{\alpha_i, \tau_i\}_{i=1}^{L}$ related to the various propagation paths have to be estimated.

- The number of multipath components in (4) that form the UWB channel can be quite large, leading to a large time dispersion of the transmitted pulse [3].

- But only some paths have the amount of original energy. (e.g. 1160 $\rightarrow$ 70) Therefore, we limit ourselves to estimate the L_c most significant paths that composes the UWB channel impulse response

- Furthermore, the reconstruction step in the MP algorithm can be thought of as a weighted sum of the elements in the dictionary, that is $\sum_{i=1}^{Z} \theta_i d_i(t)$.

- Since each element in the dictionary is a shifted version of the transmitted pulse, it turns out that θ_i is an estimate of the path gain related to the i th propagation path.

- Furthermore, the path delay is directly determined by observing the time-location of the i th atom found in the received UWB signal.

- Let $\Theta = [\theta_1, \theta_2, ..., \theta_Z]^T$ and let let $\theta_{(k)}$ for $k = 1, 2, ..., Z$ be sorted elements of the set $\{|\theta_1|, ..., |\theta_Z|\}$. Also let $l_{(k)}$ be the index in the sparse vector of the kth sorted element. For $i = 1, 2, ..., L_C$:

$$\hat{\alpha}_i = \theta_{l_{(i)}}$$
$$\hat{\tau}_i = l_{(i)}\Delta \quad (6)$$

III. Ultrawideband Detection Based on Compressive Sensing

- Until now, we have the assumption of noiseless conditions. But we have to consider the noise and interferences.

1. UWB Signal Models.

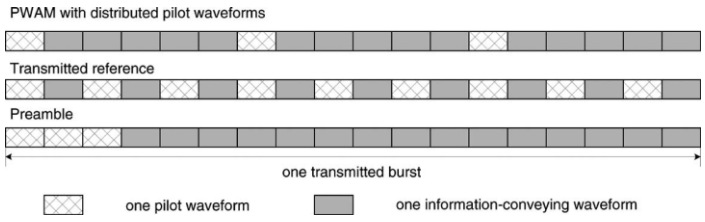


Figure. Placement of pilot waveforms for PWAM, TR, and preamble ($N_f = 3$, $N_p = 3$, $N = 18$).

- Consider a peer-to-peer UWB communication system where the k th binary information symbol is transmitted by sending N_f ultra-short pulses in the symbol interval T_s, that is [22]

$$s(t) = \sum_k b(k) \sum_{j=0}^{N_f - 1} p(t - jT_f - kT_s) \quad (7)$$

- $T_f = T_s / N_f$: frame time ; time interval between two consecutive pulses.

- $b(k) \in \{-1,1\}$: binary information symbol that modulated the amplitude of the pulse stream.

- T_p : pulse duration ; $T_p << T_f$.

- N_f nonoverlapped pulses are transmitted for each information symbol.

- The channel is static during a burst of N_s consecutive symbols. ($h(t)$ is fixed during the burst of N_s symbols).

- Let $T_f \geq \tau_{L-1} + T_p$: there is no inperpulse interference. τ_{L-1} : max delay spread of multi path channel.

- The received waveform during the first frame of the kth transmitted information symbol

$$r_f(t) = b(k) \sum_{l=0}^{L-1} \alpha_l p(t - kT_s - \tau_l) + \eta(t) \quad (8)$$

- $\eta(t)$: zero mean AWGN that models thermal noise and other interference like multi user interference.

- Since $T_f \geq \tau_{L-1} + T_p$ and the UWB channel is fixed, the received signal during the kth information symbol can be represented by periodically repeating the noiseless part of $r_f(t)$ every T_s seconds.

$$r(t) = \sum_{j=0}^{N_f - 1} r_f(t - jT_f) + \eta(t) \quad (9)$$

- Two common approaches in detection problem : correlator based detector and Rake receiver.

- In the UWB correlator-based detector, it is assumed that the channel impulse response is completely known at the receiver to define the reference template that is used in the demodulation stage.

- Likewise, for the RAKE-based receiver the channel taps $\{\alpha_{l_i}, \tau_{l_i}\}|_{i=1}^{L_r}$ related to the most significant propagation paths are assumed to be known a priori to define the set of templates for the bank of correlators and the weights for MRC [28].

- In either case, the need for UWB channel estimation arises.

- The problem of UWB channel estimation using CS under the data-aided framework. : We use N_p known pilots symbols in each packet to estimate the channel impulse response. Based on these pilots, the channel is estimated either by CS template reconstruction (Section II-B2) or CS

channel tap estimation (Section II-C). The remaining $(N_s - N_p)$ symbols that convey information are decoded based on the acquired channel characteristics.

- Under this setting, the received UWB signal (9) can be conveniently rewritten as shown in (10)

$$r(t) = \begin{cases} \sum_{k=0}^{N_w-1} b_p\left(\left\lfloor \frac{k}{N_f} \right\rfloor\right) \sum_{l=1}^{L} \alpha_l p(t - kT_f - \tau_l) + \eta(t) & for\ 0 < t < T_w \\ \sum_{k=N_w}^{(N-N_p)N_f-1} b_i\left(\left\lfloor \frac{k}{N_f} \right\rfloor\right) \sum_{l=1}^{L} \alpha_l p(t - kT_f - T_w - \tau_l) + \eta(t) & for\ T_w < t < N_s N_f T_w \end{cases} \quad (10)$$

- $N_w = N_p N_f$, T_w : time turation of the pilot waveforms.

- The received UWB signal is observed over nonoverlapped time intervals $kT_f \le t < (k+1)T_f$ for $k = 0,1,...,N_w - 1$. the received pilot waveform in a frame time is :

$$r(t) = b_p\left(\left\lfloor \frac{k}{N_f} \right\rfloor\right) \sum_{l=1}^{L} \alpha_l p(t - kT_f - \tau_l) + \eta(t) \quad (11)$$

2. CS correlator based detertor

- A first approach exploits the CS reconstruction is a correlator-based detector.

- By observing the received UWB signal in a frame-long interval and random projecting the observed signal, a noisy template can be recovered using MP algorithm. Since N_w pilot waveforms are used for channel estimation, the estimate composite pulse-multipath channel is formed by averaging over N_w noisy templates. This approach is computationally demanding as a noisy template is recovered for each received pilot waveform.

- Alternatively, the random projected signals corresponding to the received pilot waveforms can be averaged and input to the MP algorithm for template reconstruction. This latter approach requires less computation since the MP algorithm is performed just once. Furthermore, by ensemble averaging the random projected signals, the effect of AWG noise is mitigated.

- Thus, CS template reconstruction is achieved by random projecting the frame-long received signals, ensemble averaging the random projected signals, and using MP algorithm to recover an estimate of the composite pulse-multipath channel.

- Once the template has been estimated, it can be used as correlator template to enable integrate-and-dump demodulation at frame-rate sampling.

- Since each symbol is present in N_f frames, the decision statistics for the kth symbol is formed by adding up the N_f correlator output samples related to the transmitted symbol.

$$z(k) = \sum_{j=0}^{N_f-1} \int_{jT_f+kT_s}^{(j+1)T_f+kT_s} r(t) g_{cs}(t - jT_f - kT_s) dt \quad (12)$$

- $g_{cs}(t)$ is the CS estimate of the composite pulse-multipath channel.

- It can be extended to symbol-rate directly.

3. CS rake receiver

- Rake-based detectors relies on the assumption that the UWB channel parameters, path delays and path gains, related to the most significant propagation paths are known at the receiver [4], [21].

- Consider the received pilot waveform given by (11) for $k = 1, 2, ..., N_w$, where α_l and τ_l are the UWB channel taps to be estimated.

- To reduce the effect of AWGN on the estimation of the UWB channel parameters, the CS projected pilot signals are averaged to obtain a reduced-noise projected signal that is used in the MP algorithm to estimate the channel parameters as described in Section II-C.

- Thus, CS channel estimation is performed using the ensemble average of the random projections leading to a reduced computational cost and minimizing the noise effect.

- After the estimation of parameters, the CS Rake Receiver is followed.

- Let $\{\hat{\alpha}_l, \hat{\tau}_l\}|_{l=1}^{L_c}$ be the channel parameters related to the strongest paths obtained using CS channel estimation.

- The received signal, $r(t)$, is fed to a bank of L_c correlators with templates given by the atoms $p(t - \hat{\tau}_l)$ for $l = 1, 2, ..., L_c$.

- The outputs of these correlators contain the energy captured by the strongest paths and are combined via maximum ratio combining (MRC) [29] to obtain sufficient statistic for detecting the kth bit transmitted during the jth frame.

$$z_R(k,j)=\sum_{l=1}^{L_c}\hat{\alpha}_l\int_{kT_s+jT_f+\hat{\tau}_l}^{kT_s+jT_f+\hat{\tau}_l+T_p} r(t)\,p(t-kT_s-jT_f-\hat{\tau}_l)dt \quad (13)$$

- Recalling that N_f pulses are used to transmit an information symbol, the decision statistic for symbol detection is formed by summing up the MRC outputs for N_f consecutive frames.

$$\hat{b}(k)=\mathrm{sgn}\left(\sum_{j=0}^{N_f-1} z_R(k,j)\right) \quad (14)$$

IV. Simulation Results

- The Proposed CS-based detectors are compared to that of correlator detectors used in [16], [22]. : $\tilde{g}(t)=\frac{1}{N_w}\sum_{k=0}^{N_w-1} r_k(t)$ and tradition correlator (i.e. analog-template estimation followed by correlator based detector.).

- $N=10000$ symbols are transmitted.

1. BER Performance for Different Propagation Scenarios :

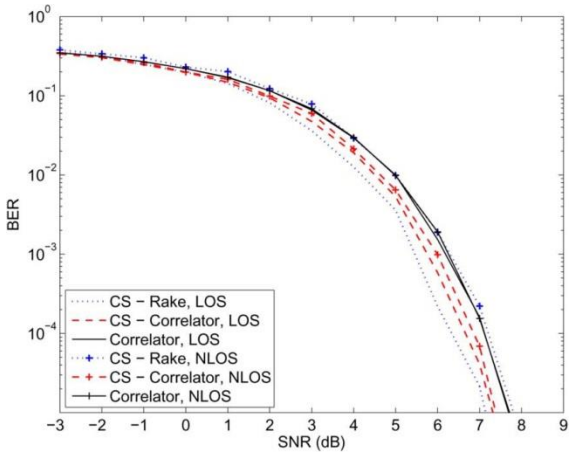


Fig. 4. Indoor residential BER performance for CS-Correlator, CS-Rake, and traditional correlator with $K/N=0.36$.

- The CS-Correlator outperforms the traditional correlator for all range of SNR.

- This shows that the reconstructed template using CS framework, $g_{cs}(t)$, is more reliable for symbol detection than the one obtained by averaging the received pilot signal, $\tilde{g}(t)$.

- This performance is expected since a denoising operation is inherently applied on the recovered signal yielding a template that is a linear combination of the transmitted pulses.

- The performance of CS-correlator for LOS channel is better than that for NLOS channel.

- This is also expected since NLOS channel introduces more multipath components than LOS channel, yielding thus a received UWB signal with less sparsity.

- CS-Rake outperforms the correlator-based detectors for LOS channel and yields competitive performance to that yielded by the traditional correlator for NLOS channel.

- As can be seen, CS-Rake degrades its performance for dense multipath channel since the CS channel estimation is unable to resolve the strongest paths among the multiple closely spaced propagation paths.

2. BER Performance for Different Number of Pilot Symbols :

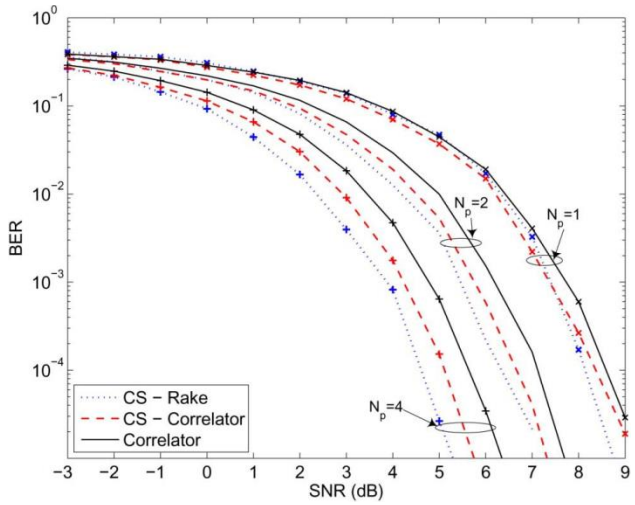


Fig. 5. BER performance for different number of pilot symbols, with $K/N = 0.36$

- Increasing the number of pilot waveforms, improvement in the channel estimation is achieved, leading to a performance gain on all the methods.

3. BER Performance for Different Number of Projections :

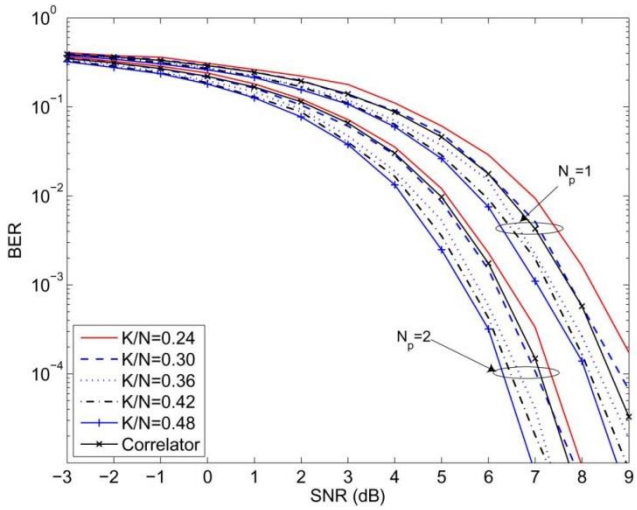


Fig. 6. BER performance for different number of projections.

- As expected, the CS-correlator's performance improves as the number of projections increases.

- More interestingly, by sampling the random projected signal at 30% of the signal's sampling rate, the CS-Correlator achieves the same performance as that yielded by the traditional correlator.

- Thus, with reduced ADC resources, the CS framework is able to reconstruct a template as good as the one obtained sampling the received UWB signal at a much higher sampling rate.

V. DISCUSSION

- What is the relation between estimated parameters and CS?

→Because of the UWB property, and the equation (3) we can use the result of MP for estimate the parameters.

- What is the value of dictionary? (form time domain to discrete domain)

→ From the uniformly sampled D, we can generate the dictionary Ψ as shifted version of pulse signal.

- How to reduce the # of samples?

→ If we think about one frame, in tradition method, we need all sample point of frame (e.g. 2048) but, by using CS, form only 250 samples, we can reconstruct the original signals.

Missing-Area Reconstruction in Multispectral Images Under a Compressive Sensing Perspective

Authors: Luca Lorenzi, Farid Melgani, Gregoire Mercier
Publication: IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, July.2013
Speaker: Hyeong ho, Baek

Short summary: The intent of this paper is to propose new methods for the reconstruction of areas obscured by clods. They are based on compressive sensing theory, which allows finding sparse signal representations in underdetermined linear equation systems.

I. INTRODUCTION

Clouds in remotely sensed imagery may or may not represent an unwanted source of noise. In case they are viewed as a noise source, several methodologies have been developed in the past in order to cope with this problem. In this paper, they will focus on the approach which attempts to remove the clouds by substituting them with cloud-free estimations.

Recently, CS has been introduced by Donoho and Candes et al. CS theory aims at recovering an unknown sparse signal from a small set of linear projections. By exploiting this new and important result, it is possible to obtain equivalent or better representations by using less information compared with traditional methods.

In this paper, they propose three novel methods to solve the problem of the reconstruction of missing data due to the presence of clouds. Given a cloud-free and a cloud-contaminated image, each of the missing measurements is recovered by applying the CS theory in which cloud-free pixels are exploited.

II. PROBLEM FORMULATION

Let us consider two multispectral images $I^{(1)}$ and $I^{(2)}$ acquired by an optical sensor at two different dates and registered over the same geographical area. Let us suppose that the two acquisitions are temporally close to each other.

We make the hypothesis that image $I^{(2)}$ is obscured by the presence of clouds. We will call cloudy area in image $I^{(2)}$ as target region $\Omega^{(2)}$ and the remaining part as source region $\Phi^{(2)}$. Image $I^{(1)}$ does not contain clouds it is supposed cloud free. Their aim is to generate a new image $I^{(2)}$ without clouds.

They assume that any pixel $x^{(1)} \in \Omega^{(1)}$ can be expressed as linear combination of pixels in region $\Phi^{(1)}$.

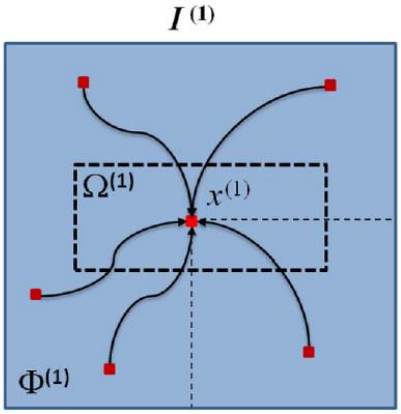


Figure 1 Illustration of the reconstruction principle

In other words, in $I^{(1)}$, we have

$$x^{(1)} = \Phi^{(1)}\alpha \qquad \forall x^{(1)} \in \Omega^{(1)} \tag{1}$$

Where α is an unknown weight vector associated with the considered pixel $x^{(1)}$. Once α is computed, if we assume that $I^{(1)}$ and $I^{(2)}$ are temporally close, so that the scene did not change in between the two observations, it will be possible to reuse the α coefficients to reconstruct the spatially corresponding pixel in the missing area $\Omega^{(2)}$.

$$\begin{aligned} &\text{from } I^{(1)}: \ \alpha = f\left(\Phi^{(1)}, x^{(1)}\right) \\ &\text{to } I^{(2)}: \ x^{(2)} = \Phi^{(2)}\alpha \end{aligned} \tag{2}$$

Where $f(\bullet)$ represents an estimation function.

III. RECONSTRUCTION VIA CS

A. CS solutions

- BP : A well-known solution for problem (1) is the BP principle. It suggests a convexificaion of the problem by using the L_1 norm. Note that, if the original signal x is sufficiently sparse, the recovery via BP is provably exact.
- OMP : One of the easiest and fastest alternative techniques is the OMP, an improved version of the MP method. MP finds the atom that has the highest correlation with the signal. It subtracts off the correlated part from the signal and then iterates the procedure on the resulting residual signal.
- BP VS OMP : In general, BP and OMP algorithms provide good performances in reconstruction problems. Nonetheless, BP is considered more powerful than OMP, since it can recover with high probability all sparse signals and is more stable. On the contrary, OMP results attractive for its fast convergence and in its ease of implementation.

B. Genetic Algorithm

GA are a part of evolutionary computation which solves optimization problems by mimicking the principles of biological evaluation.

In general, a common GA involves the following steps. First, an initial population of chromosomes is randomly generated. Then, the goodness of each chromosome is evaluated according to a predefined fitness function representing the aim of the optimization. Evaluating the fitness function allows keeping or discarding chromosomes, by using a proper rule based on the principle that, the better the fitness, the higher the chance of being selected. Once the

selection of the best chromosomes is done, the next step is devoted to to the reproduction of a new population. This is done by genetic operators such as crossover and mutation operators. All these steps are iterated until some predefined condition is satisfied. In this situation, the fitness function are given below

$$f_1 = \min \|\alpha\|_0 \tag{3}$$

$$f_2 = \min \|D\alpha - x\|^2 \tag{4}$$

IV. Experimental Results

A. *Data set Description and Setup*

- Compare the reconstructed image with the original cloud-free image.
- Two aspects : 1. The kind of ground covers obscured and 2. The size of the contaminated area.
- For the purpose of comparison, we implemented two other methods developed to reconstruct cloudy areas in images. One consists in a recent work exploiting a multiresolution inpainting (MRI), whereas the second method estimates a missing pixel by contextual multiple linear prediction (CMLP).

B. *Results*

- Contamination of Different Ground Covers : In Figure shows mask A covering a region that includes mainly an urban area, mask B obscuring an industrial zone, and mask C covering a vegetation area.

Method	Mask A			Mask B			Mask C		
	PSNR	Complexity	Time [s]	PSNR	Complexity	Time [s]	PSNR	Complexity	Time [s]
MRI	22.54	-	2856	16.05	-	2517	33.77	-	2898
CMLP	20.99	1	1	20.11	1	1	24.05	1	1
OMP	**23.96**	3	4	20.60	3	4	31.97	3	4
BP	22.22	294	66	**24.74**	168	59	30.67	301	60
GA	23.78	148	68621	23.15	95	26312	**32.01**	138	43193

In general, MRI can re return visually satisfactory results only when the missing area refers to a uniform region such as a vegetation region.

The OMP algorithm produces very sparse reconstruction solution. On the contrary, the BP algorithm selects a large number of weight coefficients. Finally, GA can be viewed as a compromise between the two previous methods. Despite the very long time needed to estimate the reconstruction model, it results sparser than BP but less parsimonious than OMP.

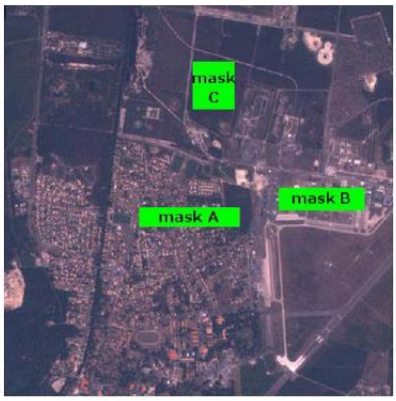


Figure 2 Masks adopted to simulate the contamination of different ground covers.

- Contamination with different size : Figure shows the three different masks adopted to simulate different increasing cloud cover sizes.

Method	Mask 1			Mask 2			Mask 3		
	PSNR	Complexity	Time [s]	PSNR	Complexity	Time [s]	PSNR	Complexity	Time [s]
MRI	24.27	-	2995	22.85	-	10176	23.82	-	22353
CMLP	24.61	1	1	24.43	1	2	25.46	1	2
OMP	26.36	3	5	26.42	3	16	27.39	3	21
BP	26.45	338	61	26.82	332	143	**28.25**	329	973
GA	**26.72**	173	69231	**27.10**	168	103342	28.15	170	259459

To get higher PSNR values, one needs to resort to CS techniques. Indeed, our implementations return better results in terms of PSNR in all the simulations. The result form this viewpoint underline the main weakness of the GA solution i.e., its expensive computational needs.

Figure 3 Masks adopted to simulate the different sizes of contamination.

V. CONCLUSION

This paper deals with the complex and important problem of removal of clouds from images. First we have shown how two common CS solutions, namely, the OMP and BP algorithms, can be formulated for a cloud-contaminated-image reconstruction problem. Then, we have proposed a solution for solving the CS problem exploiting the capabilities of GA.

The experimental results point out the superiority of the proposed methods compared to two reference methods for cloud removal. OMP has the advantage of being sparser and significantly faster than BP and GA, but it is the less robust method. And BP is much less sparse than OMP. GA represents a good compromise between the OMP and BP methods, mainly because it is more robust than OMP and more sparse than BP.

VI. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

Appendix

Reference

Power and Channel Allocation for Cooperative Relay in Cognitive Radio Networks

Authors: G. Zhao et. al.

Publication: IEEE JSTSP, Feb. 2011
Speaker: Asif Raza

Short summary: In this paper authors mention that cognitive radio relay channels can be divided into three categories: direct, dual-hop, and relay channels. The relay node involves both dual-hop and relay diversity transmission. They develop power and channel allocation approaches for cooperative relay networks. They also develop a low complexity approach that can obtain most of the benefits from power and channel allocation with minor performance loss.

I. INTRODUCTION

Resource (channel and power) allocation in CR relay networks is considered in the paper. The power and channel allocation for cooperative relay in a three-node CR network, which consists of a source, a relay, and a destination and can operate in multiple spectrum bands, is considered. In this context CR relay channels(CRRCs) can be divided into three categories as shown in Fig. 1.

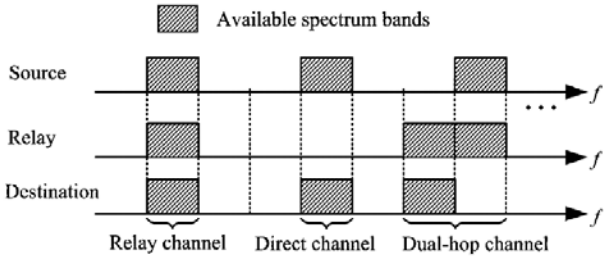


Fig. 1. CRRC in CR networks.

These relay channels have their advantages over each other. For example a dual-hop channel has a bottleneck in throughput whereas a relay channel loses half of its throughput due to its half-duplex constraint. While a CR transmitter and at its intended CR receiver using direct channel on their respective links can result in scarcity of the available spectrum bands for other users in highly congested areas. In this paper authors propose to assign the spectrum band of the relay channel to assist the transmission in dual-hop or direct channels.

Authors in this paper first introduce CRRC in a CR network with four typical spectrum bands, and then discuss power constraints for both the source and the relay. Finally they obtain end-to-end throughput of CRRC.

II. CR System Description

A. *Cooperative Relay Channel*

The network design considered is shown in the following figure.

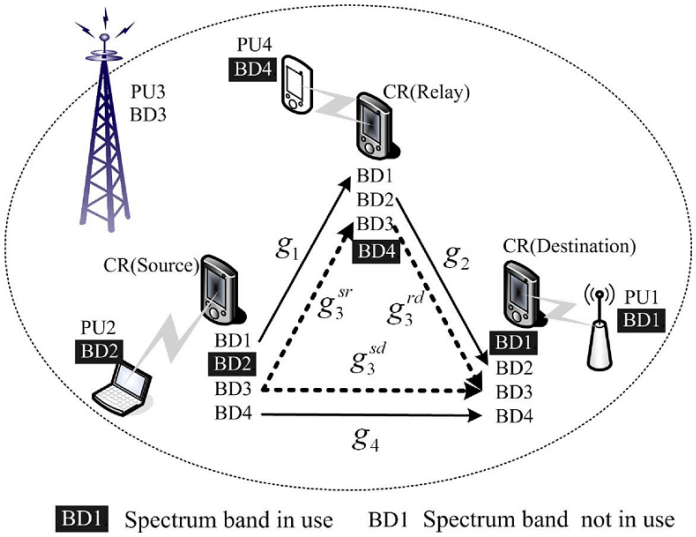


Fig. 2. System setup of cooperative relay in CR networks.

In the network setup every CR node is equipped with an omnidirectional antenna and can simultaneously sense four licensed spectrum bands, BD_i. Each of them belongs to a PU exclusively. The primary user 1 (PU1), PU2 and PU4 are using BD1, BD2 and BD4 channels respectively. However they have local effect only. The PU3 has large coverage area and effects the whole CR network. However, for example, if PU3 is not transmitting then BD3 is available to relay node. The source-relay (sr), relay-destination (rd) and source-destination (sd) links are using channel powers over BD3 as g_3^{sr}, g_3^{rd} and g_3^{sd} respectively.

B. *Transmit Power Constraint*

Let $\mathbf{P}^S = \left[p_1^S, p_2^S, p_3^S, p_4^S \right]$, $\mathbf{P}^R = \left[p_1^R, p_2^R, p_3^R, p_4^R \right]$ represents power allocation vectors for source and relay nodes over all four BDs, respectively. The power constraint is defined as:

$p_i^S \le P_{\max}$, $p_i^R \le P_{\max}$ and total power is defined as: $\sum_{i=1}^{4} p_i^S \le P_{\max}^S$, $\sum_{i=1}^{4} p_i^R \le P_{\max}^R$ where $P_{\max}^S$ and $P_{\max}^R$ maximum powers that source and relay are able to transmit.

C. *End-to-End Throughput*

End-to-end throughput on direct transmission on BD4 can be expressed as: $R_{direct} = C\left(p_4^S g_4\right)$ where g_4 is channel power over BD4. For dual-hop transmission in BD1 and BD2, both operates serially thus the end-to-end throughput is smaller of two hops, i.e., $R_{dual} = \min\left\{C\left(p_1^S g_1\right), C\left(p_2^R g_2\right)\right\}$. The throughput on relay channel is: $R_{relay} = \frac{1}{2}\min\left\{C\left(p_3^S g_3^{sr}\right), C\left(p_3^S g_3^{sd}\right) + C\left(p_3^R g_3^{rd}\right)\right\}$. The overall throughput of CRRC is given as: $R_{all}\left(\mathbf{P}^S, \mathbf{P}^R\right) = R_{direct} + R_{dual} + R_{relay}$.

III. Power and Channel Allocation

Due to complexity, the channel and power allocation is considered independently.

A. *Channel Allocation:* four possible transmission modes are defined as:

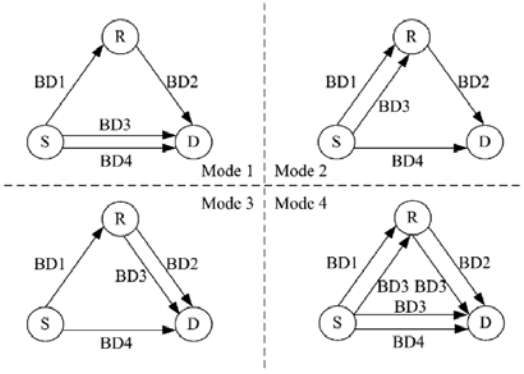


Fig. 3. Different modes of channel allocation.

Objective of channel allocation is to select proper mode to maximize overall end-to-end throughput. The throughput for each mode is defined in end-to-end throughput section. It requires power allocation for each mode.

B. Power Allocation: for first 3 modes, power allocation at relay node is defined as:

$$J_{RD} = \max_{\mathbf{P}^R}\left\{\sum_{i\in\Gamma_{RD}} C\left(p_i^R g_i\right)\right\}$$
$$= \max_{\mathbf{P}^R}\left\{\sum_{i\in\Gamma_{RD}} B\log\left(1+p_i^R g_i\right)\right\}$$

Subject to:

$$\sum_{i\in\Gamma_{RD}} p_i^R \le P_{\max}^R$$

$$p_i^R \le P_{\max}, \qquad i \in \Gamma_{RD}$$

$$p_i^R \ge 0, \quad i \in \Gamma_{RD}$$

Similarly power allocation at source is defined as:

$$J_{SD} = \max_{\mathbf{P}^S} \left\{ \sum_{i\in\Gamma_{SR}\cup\Gamma_{SD}} C\left(p_i^S g_i\right) \right\}$$

Subject to:

$$\sum_{i\in\Gamma_{SR}\cup\Gamma_{SD}} p_i^S \le P_{\max}^S$$

$$p_i^S \le P_{\max}, \qquad i \in \Gamma_{SR}\cup\Gamma_{SD}$$

$$p_i^S \ge 0, \quad i \in \Gamma_{SR}\cup\Gamma_{SD}$$

$$\sum_{i\in\Gamma_{SR}} C\left(p_i^S g_i\right) \le R^*$$

where $R^* = \sum_{i\in\Gamma_{SR}} C\left(p_{io}^S g_i\right)$ represents throughput on link from source to relay node. Similarly the throughput on link from relay to destination can be defined in similar way. Here p_{io}^S represents power allocated to source by using water-filling solution[2], $p_{io}^S = \left[\frac{B}{\ln 2(\lambda+\mu_i)} - \frac{1}{g_i}\right]$. The last constraint in above problem is non-convex.

For mode 1 and 3, the last constraint can be converted into inequality constraint as $p_1^S \le \frac{1}{g_1}\left(2^{R^*/B} - 1\right)$. The problem then becomes convex optimization problem and can be solved by using water-filling solution.

For mode 2 there are more than one spectrum bands for first hop of dual hop transmission. In this case the last constraint of the defined problem is non-convex. It is transformed into equality constraints as:

Step 1: Perform power allocation without considering the constraint and obtain the power allocation vector $\mathbf{P}^S$.

Step 2: Check whether $\mathbf{P}^S$ meets the constraint. If so, it is the power allocation vector that we need for the source. Otherwise, reduce the sum power constraint of the source $P_{\max}^S$ and perform power allocation until $\mathbf{P}^{S*}$ meets:

$R^* - \varepsilon \le \sum_{i\in\Gamma_{SD}} C\left(p_i^{S*} g_i\right) \le R^* + \varepsilon$ where $\mathbf{P}^{S*} = \left[p_1^{S*}, 0, p_3^{S*}, 0\right]$

Step 3: Obtain the inequality constraints by $p_1^S \le p_1^{S*}$ and $p_3^S \le p_3^{S*}$

For mode 4: → use SD link (direct transmission) if $g_3^{sd} > g_3^{sr}$ but if $g_3^{sd} \le g_3^{sr}$ then all three links should be used in relay diversity transmission i.e., $R_{all}\left(\mathbf{P}^S, \mathbf{P}^R\right) = R_{direct} + R_{dual} + R_{relay}$

In this case power allocation at relay is: $\mathbf{P}^R(\alpha) = \left[0, p_2^R(\alpha), p_3^R(\alpha), 0\right]$ and source needs to divided its power into three parts; direct, dual and relay diversity.

Direct transmission case; $R_{direct}(\alpha) = C\left(p_4^S g_4\right)$,

Dual-hop case; $R_{dual}(\alpha) = \min\left\{C\left(p_1^S g_1\right), C\left(p_2^R(\alpha) g_2\right)\right\}$,

Relay transmission; $R_{relay}(\alpha) = \frac{1}{2}\min\left\{C\left(p_3^S g_3^{sr}\right), C\left(p_3^S g_3^{sd}\right) + C\left(p_3^R(\alpha) g_3^{rd}\right)\right\}$

The overall end-to-end throughput can be maximized as:

$$J(\alpha) = \max_{\mathbf{P}^S}\left\{R_{direct}(\alpha) + R_{dual}(\alpha) + R_{relay}(\alpha)\right\}$$

$$\text{Subject to:} \quad \begin{aligned} &\sum_{i\in\{1,3,4\}} p_i^S \le P_{\max}^S \\ &p_i^S \le P_{\max}^S, \qquad i = 1,3,4 \\ &p_i^S \ge 0, \qquad i = 1,3,4 \\ &p_1^S \le \frac{g_2}{g_1} p_2^S(\alpha) \end{aligned}$$

- When $C\left(p_3^S g_3^{sr}\right) \le C\left(p_3^S g_3^{sd}\right) + C\left(p_3^R g_3^{dr}\right)$ then objective function becomes:

$$J(\alpha) = \max_{\mathbf{P}^S}\left\{C\left(p_1^S g_1\right) + \frac{1}{2}C\left(p_3^S g_3^{sr}\right) + C\left(p_4^S g_4\right)\right\}$$

- When $C\left(p_3^S g_3^{sr}\right) > C\left(p_3^S g_3^{sd}\right) + C\left(p_3^R g_3^{dr}\right)$ then objective function becomes:

$$J(\alpha) = \max_{\mathbf{P}^S}\left\{C\left(p_1^S g_1\right) + \frac{1}{2}\left(C\left(p_3^S g_3^{sd}\right) + C\left(p_3^R g_3^{dr}\right)\right) + C\left(p_4^S g_4\right)\right\}$$

In both of the cases, the problem is convex problem and can be solved by using water-filling solution as:

$$p_{1o}^S = \left[\frac{B}{\ln 2(\lambda + \mu_1)} - \frac{1}{g_1}\right]$$

$$p_{3o}^S = \left[\frac{B}{2\ln 2(\lambda + \mu_1)} - \frac{1}{g_3}\right]$$

$$p_{4o}^S = \left[\frac{B}{\ln 2(\lambda + \mu_4)} - \frac{1}{g_4}\right]$$

In brief, the power and channel allocation in CRRC can be summarized as follows:

• List all possible modes of the channel allocation

• Perform power allocation for each mode

• Pick the mode with the largest overall end-to-end throughput by exhaustive search.

IV. NUMERICAL RESULTS

The parameters used for evaluation are: number of CR nodes=3; number of spectrum bands = 4; spectrum bandwidth = 1 MHz; noise at CR node= -126 dBW; path loss between two CR nodes = 126 dB; maximum allowable power on each spectrum band i.e. P_{max}=3W;

A. Different Source / Relay Power Constraints

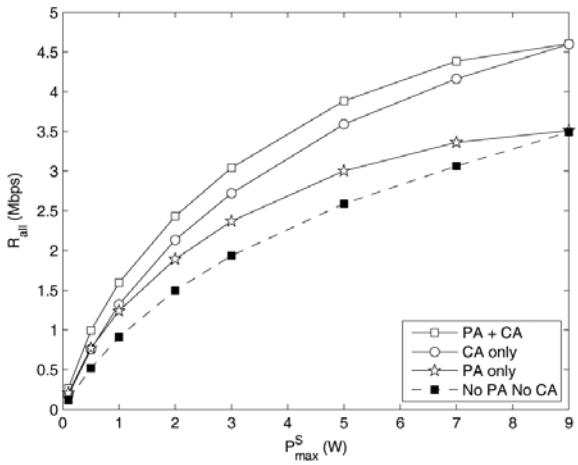


Fig. 4. End-to-end throughput versus the power constraint at the source.

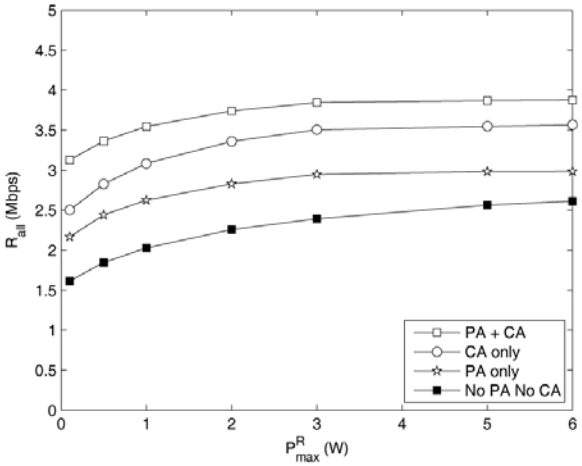


Fig. 5. End-to-end throughput versus the power constraint at the relay.

PA= Power Allocation, CA = Channel Allocation, $P^S_{\max}$ = maximum power at source, $P^R_{\max}$ = maximum power at relay, R_{all}= end-to-end throughput. "No PA No CA" is Mode 4 used as a baseline for comparison. The notable observation in Fig. 4 is that CA continue to increase throughput for increase in sum power constraint however PA can only improve throughput when $P^S_{\max} \le 9W$. This is because when the sum power constraint is large enough, the per band power constraint will limit the transmit power. {Then the source sends signals with maximum allowable transmit power on each spectrum band. This is equivalent to equal power allocation, i.e., P_{max} no power allocation.} Therefore, channel allocation is more effective than power allocation in CRRC. In Fig. 5 the throughputs of different schemes grow almost at similar scales. However when the sum power constraint is large enough, the throughput will be capped by per band power constraint.

A. Low Complexity Approach: if the CR system works in Mode 4, the relay has to conduct both *du*al-hop and relay diversity transmission, which *complicates the system. Therefore, we omit Mode 4 and only consider Modes 1, 2, and 3 for the power and channel allocation.*

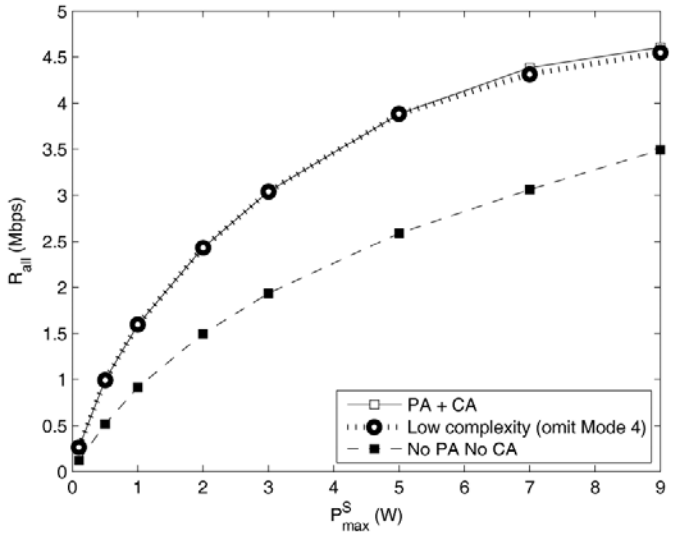


Fig. 6. Performance of the low complexity approach in the typical case with four spectrum bands.

We can find that the low complexity approach of omitting Mode 4 has similar performance to the method of considering all four modes. Furthermore, when the sum power constraint at the source is larger than 9 W, it only decreases the throughput from about 4.6 Mbps to about 4.5 Mbps compared to the scheme with power and channel allocation, i.e., about 2% performance loss.

B. Performance in Multiple Spectrum Bands:

When N independent spectrum bands are used, there are L=N/4 relay channels on average. It is shown in fig. 7 that low complexity scheme has performance close to power and channel

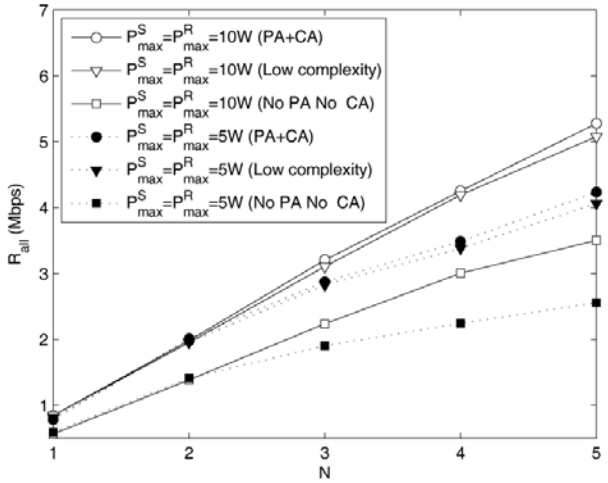


Fig. 7. Performance of the low complexity approach in the case with N spectrum bands.

allocation scheme for both 5W and 10W sum power constraints. Moreover it outperforms the scheme with no power and channel allocation in both of the power constraint cases.

INFONET, GIST
Journal Club (2013. 12. 09)

Robust Compressive Data Gathering in Wireless Sensor Networks

Authors: Yu Tang, Bowu Zhang, Tao Jing, Dengyuan Wu, Xiuzhen Chen
Publication: IEEE T. On Wireless Comms, June.2013
Speaker: Jongmok Shin

I. INTRODUCTION

-In this paper, authors address the problem of data gathering in WSNs with outlying sensor readings and broken links.

-Traditional approaches rely on in-network data compression (ex)wavelet transform, joint entropy coding and so on), which may suffer from following two major drawbacks:

1) High communication overhead(in the worst case O(N^2) single-hop transmissions are needed to collect the data from N sources)

2) Some approaches such as the distributed source coding rely on a static correlation structure, which may not be easily obtained in a dynamic environment.

-The recent breakthroughs in CS theory motivate us to investigate compressive data gathering.

-The main contributions of this paper are as follows. (Authors target the robustness of compressive data gathering: the accuracy of the recovered data at the sink is significantly affected by outlying sensor readings and broken links, whose existence makes the collected signal uncompressible)

1. Authors argue that compressible data may lose its compressiveness in WSNs due to outlying sensor readings and broken links.
2. They propose two CS based approaches, with one focusing on detecting and recovering (correcting) outlying sensor readings and the other inferring the broken links.
3. They perform an extensive simulation study to evaluate the performance of the proposed methods over various parameter settings and compared to other popular in-network compression algorithms.

II. FUNDAMENTALS OF COMPRESSIVE SENSING

Let x be a $N \times 1$ column vector in R^N. Given an $N \times N$ orthogonal basis $\Psi = [\Psi(1), \Psi(2), ..., \Psi(N)]$ with each $\Psi(i)$ being a column vector, x can be expressed by (1),

$$x = \Psi s = \sum_{i=1}^{N} s_i \Psi(i), \quad (1)$$

where s is the coefficient sequence of x in the transform domain Ψ .The signal x is k -sparse if it is a linear combination of k basis vectors. That is, only k of the s_i coefficients are nonzero and the other $(N-k)$ ones are zero. If $K \ll N$, instead of acquiring all the N values from x , CS aims to reconstruct x by taking only a small set of measurements:

$$y = \Phi x = \Phi \Psi s = As, \tag{2}$$

Where y is a $M \times 1$ vector, $k < M \ll N, \Phi \, is \, a \, M \times N$ *measurement matrix*, and A is a $M \times N$ matrix. For a $N \times 1$ vector s , it has been proved that if A holds the Restricted Isometry Property(RIP), s can be recovered with only $M \geq c \times k \log(N/k)$ measurements at an overwhelming probability through the following l_1 -minimization

$$\min |s|_{l_1} \; subject \, to \, y = As,$$

where c is a constant depending on each instance.

The definition of RIP is as follows: a matrix A obeys RIP with $(k, \delta) \; for \, \delta \in (0,1) \, if$

$$1 - \delta \leq \frac{\| Av \|_2^2}{\| v \|_2^2} \leq 1 + \delta \tag{3}$$

holds for all k -sparse vector v .

(A should project all k -sparse vector v with equal energy.)

If the measurement vector y is corrupted with noise N (AWGN),

$$y = As + N, \tag{4}$$

Then the l_1 -minimization is

$$\min |s|_{l_1} \; subject \, to \| As - y \|_{l_2} < \varepsilon, \tag{5}$$

where ε bounds the amount of noise in the data.

III. SYSTEM MODEL AND MOTIVATION

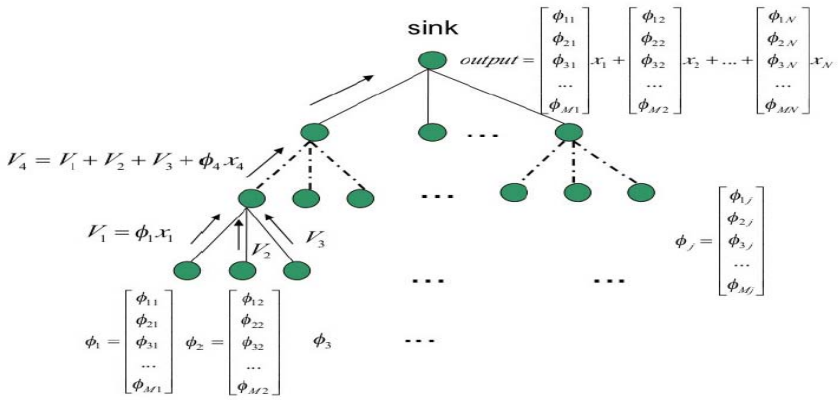


Fig.1. The architecture for compressive data gathering. Each observation at the data sink is a random projection of all sensor readings.

In Fig.1, the shortest path tree spanning all sensor nodes and rooted at the data sink is adopted to gather the readings of the whole network. The objective is to collect the original sensor readings from all sensors. If no compression policy is employed, $O(N^2)$ number of messages are required. However, compressive data gathering requires $O(MN)$ messages, with $c \times k \log(N/k) \le M \ll N$. This architecture is detailed as follows.

Let N be the number of sensors and x be the set of sensor readings forming a $N \times 1$ column vector in $\Re^N$. Each sensor reading $x_i, i \in \{1, 2, ..., N\}$, is multiplied with a vector of M random values and the resultant vectors are added together forming partial projections from the non-leaf nodes along the paths to reach the root, which computes the final M random projections of the N sensor readings. The random vectors associated with the sensor nodes constitute the column of the measurement matrix Φ . Thus the data received at the sink is:

$$y = \Phi x \tag{6}$$

This architecture can preserve high fidelity data recovery at the data sink only if the original sensor readings (vector x) are compressible. In principle this is true as the values in x are the samples of a real-world smooth signal and they are spatially correlated.

However, our sensor network is not perfect. Its sampled data x may not be compressible due to the outlying readings and broken links.

-The outlying readings: Malfunctioning sensors may report outlying readings. Since outlying readings are typically uncorrelated with each other and uncorrelated with normal sensor readings, the spartcity of the signal in its transform domain may be violated.

-Broken links: It may exist due to power depletion or sensor malfunction. In compressive data gathering, a broken link may affect multiple sensor readings.

Therefore, in this paper, authors focus on solving both problems.

IV. OUTLYING SENSOR READING IDENTIFICATION AND BROKEN LINK DETECTION

A. Outlying sensor reading identification.

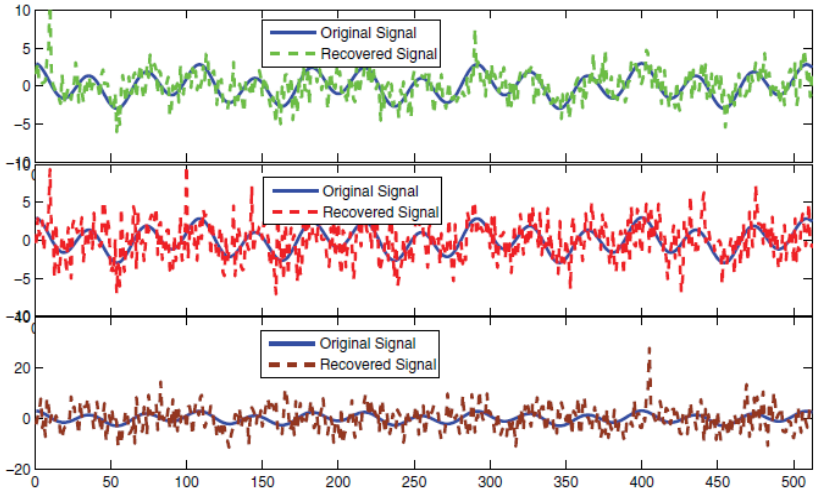


Fig.2. The recovered signal with outliers.

They randomly select sensors to be outliers and test the impact of outlying readings on compressive data gathering. Figure.2 illustrates the recovered signal when we have one, two or 100 sensors reporting outlying readings. We observe that even with only one sensor reporting an outlying reading, the recovered signal deviates from the original one significantly.

Outlying sensor reading detection is a challenging problem in compressive data gathering because the data collected at the sink are the random linear projections of the real sensor readings, which renders the popular statistics-based outlier detection algorithms inapplicable. To solve this problem, we resort to the compressive sensing theory again.

Sensor outlying is itself a sparse event in the primal domain(I, the identity matrix) of x since we assume that only a very small of sensor readings is outlying sensor readings. Therefore, outlying sensor readings can be identified based on the compressive sensing theory. Let $x_m = x + x_o$ (x_m :the vectors of the readings of all the sensors, x: normal sensor readings, x_o : outlying sensor readings)

Note that x is sparse in its transform domain Ψ , but x_o is not. However, x_o is sparse with respect to I . Therefore, we have

$$\begin{aligned} y &= \Phi x_m = \Phi x + \Phi x_o \\ &= \Phi \Psi s + \Phi I x_o = \Gamma s + \Phi x_o \\ &= (\Gamma \quad \Phi)(s \quad x_o)^T \end{aligned} \tag{7}$$

where $(s \quad x_o)^T$ can be reconstructed by any CS recovery algorithm.

B. Broken Link Detection.

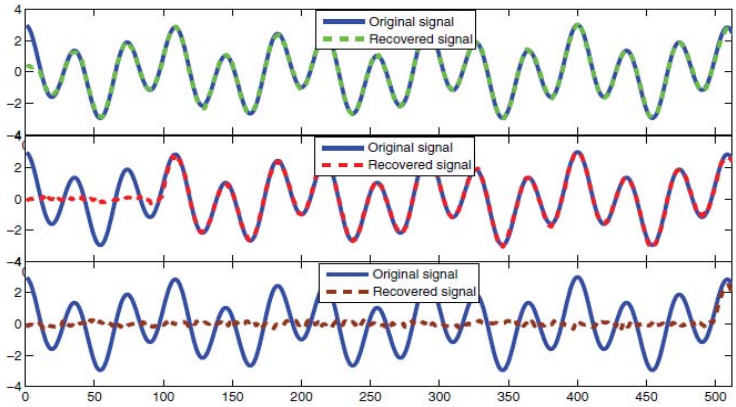


Fig.3. The recovered signal with broken links.

As shown in Fig.3, the existence of broken links makes the direct utilization of compressive sensing on x invalid.
-Dashed line of the first graph: when the link connecting the node at position 10 to the tree is broken.
-Dashed line of the second graph: when the link connecting the node at position 113 to the tree is broken.
-Dashed line of the third graph: when the link connecting the node at position 487 to the tree is broken.

In this subsection, they propose an approach to infer the broken links based on the compressive sensing theory. Let l_j be a binary variable denoting the status of the link from sensor j to its parent node in the routing tree. Then $l_j = 1$ if and only if the link from j to its parent is broken. Let $x_l = [l_1, l_2, ..., l_N]$ be the vector of the link statuses. Note that x_l is sparse in the primal domain of x since we assume that only a very small number of links are broken at any instant of time. In other words, x_l is sparse with respect to the identity matrix I, i.e., $x_l = Is_l$. Let i be the parent of j. Denote by y_l the vector of M_l observation characterizing the broken links. Each observation is the random projection of the link statuses. If i locally concludes that the link to j is broken, i simply adds a vector of M_l random values to the partial projections of x_l for broken link detection. All the random values used for the projection form the observation matrix Φ_l. At the data sink, the received value can be written as

$$y_l = \Phi_l x_l = \Phi_l I s_l \tag{8}$$

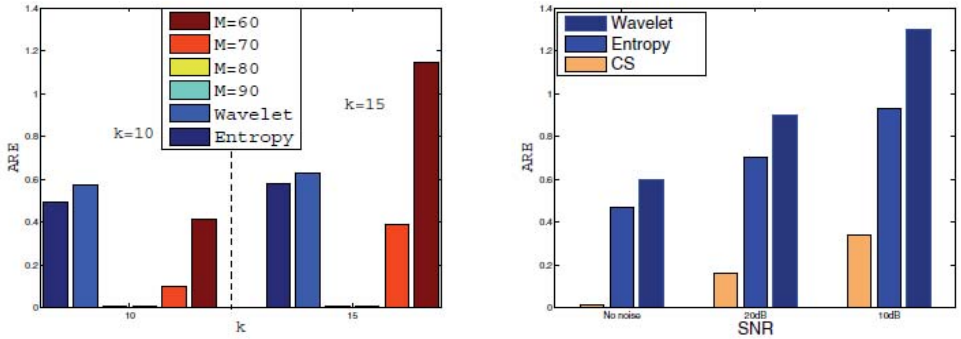
which can be solved easily based on any CS sparse recovery algorithm.

V. THE RESULT OF SIMULATION

In this section, they show the performance of robust compressive data gathering in sensor networks.
Definition: For each sensor $n_i, i \in \{1, 2, ..., N\}$, *let* x_i *and* $\hat{x}_i$ be the true and the estimated reading, respectively. The Average relative error(ARE) is defined to be the average of the ratio between the difference of the estimated reading and the true reading vs. the true reading:

$$ARE = \frac{\sum_{i=1}^{N} | x_i - \hat{x}_i | / x_i}{N} \tag{9}$$

1) Compressive Data Gathering In a Perfect Network: They first evaluate the performance of compressive data gathering when the network is perfect: neither broken links nor outlying sensor readings exist in the network.



(a) (b)

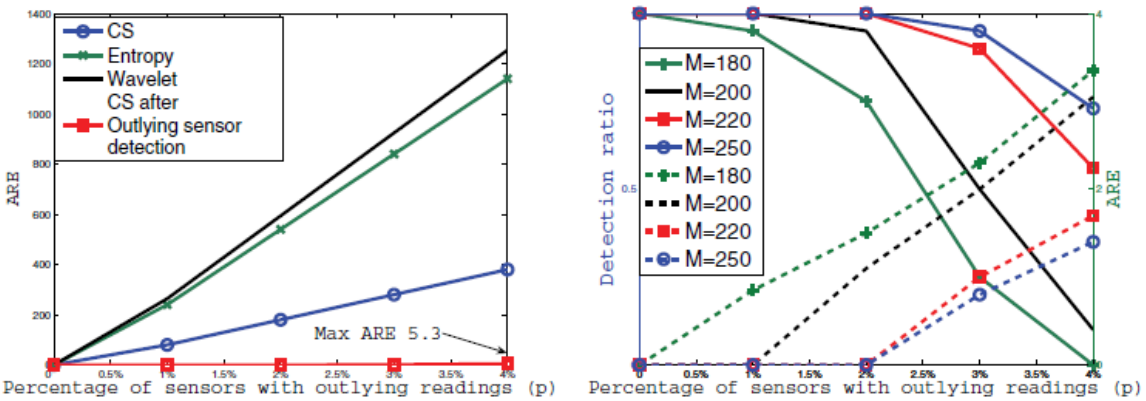
Fig.4. The ARE of the three algorithms

Fig.4(a) is without the random noise case, (b) is with the random noise case.

In Fig.4(a), the ARE increases as k increases and the ARE decreases with an increasing M, while for entropy coding and wavelet transformation, their AREs slightly increase with an increasing k. Because they are not dependent on the number of k, but the correlation among the transmitted data.

In Fig.4(b), we could notice that the ARE of compressive data gathering when noise exists is greater than that of the case when there is no noise. Overall, the compressive data gathering achieves a lower ARE than the other two methods. This is because the noise damages the original data correlation, leading to inaccurate data decompression in conventional data gathering methods.

2) Outlying reading detection and recovery: For outlying reading detection they consider the following scenario. Let p be the percentage of outlying sensors. In other words, the vector of outlying sensor readings is $o = N \times p$ sparse.



(a) (b)

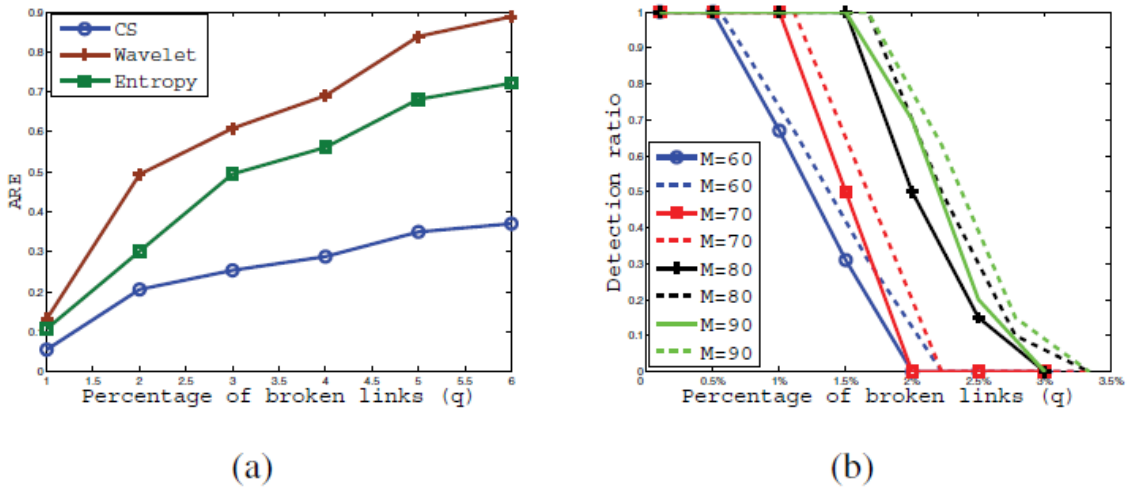
Fig.5. (a) The ARE of the three algorithms vs. p

(b) Outlying reading detection ratio vs. p (solid curves) and ARE vs. p (dashed curves)

In Fig.5(a), we could notice that ARE increases as the percentage of the sensor nodes with outlying readings increases and compressive data gathering achieves the best signal recovery performance. This is because outlying readings can severely distort the data correlation. In addition to, authors show that by removing the outlying sensor readings, the true reading can be obtained with very low recovery error.

In Fig.5(b), we could notice that the detection ratio increases and ARE decreases as M increases.

3) Broken link detection and recovery: For broken link detection we consider the following scenario. Given a percentage q, randomly choose $N \times q$ links to be broken. Let x_l be an l-sparse vector where $l = N \times q$.



(a) (b)

Fig.6. The performance of the three algorithms when there are broken links. (a) ARE vs. q (b) Broken link detection ratio vs. q.

In Fig.6(a), it is observed that the ARE increases as q increases. We also could observe that compressive data gathering achieves a much better result than the algorithms based on entropy coding and wavelet transformation. This is because broken links that are far away from the sink may not severely affect the data recovery accuracy of compressive data gathering while the impact of broken links on the algorithms based on entropy coding and wavelet transformation is the same wherever the broken links are.

In Fig.6(b), it is noticed that as an overall trend, the larger the M, the higher the detection accuracy we can achieve. And for a particular M, the larger the q, the lower the detection ratio.

VI. CONCLUSION

-In this paper, Authors investigate the problem of robust data gathering in WSNs based on the compressive sensing theory.

Firstly, propose an architecture for compressive data gathering

And then, develop two CS based methods

1. to identify outlying readings

2. to identify broken links

Finally, authors carry out an extensive simulation study and their results demonstrate that the proposed robust compressive data gathering approach outperforms other popular in-network compression algorithms.