CHANNEL ESTIMATION FOR MASSIVE MIMO : A SEMIBLIND ALGORITHM EXPLOITING QAM STRUCTURE

Baki Berkay Yilmaz [1] *and Alper T. Erdogan*[2]

[1]Georgia Institute of Technology, Atlanta, USA
[2]Koc University, Istanbul, Turkey

ABSTRACT

We introduce a new channel matrix estimation algorithm for Massive MIMO systems to reduce the required pilot symbols. The proposed method is based on Maximum A Posteriori estimation where the density of QAM transmission symbols are approximated with continuous uniform pdf. Under this simplification, joint channel source estimation problem can be posed as an optimization problem whose objective is quadratic in each channel and source symbol matrices, separately. Also, the source symbols are constrained to lie in an ℓ_∞-norm ball. The resulting framework serves as the channel estimation counterpart of the recently introduced compressed training based adaptive equalization framework. Numerical examples demonstrate that the proposed approach significantly reduces the required pilot length to achieve desired bit error rate performance.

***Index Terms*—** Massive MIMO, Channel Estimation, Compressed Training,

1. INTRODUCTION

Massive Multiple Input Multiple Output (MIMO) targets to increase wireless system throughput by employing hundreds, even thousands, of antennas at the base stations [1–4]. The user terminals are multiplexed by their spatial signatures enabling simultaneous use of precious bandwidth resources .

A critical factor for the performance of Massive MIMO systems is the adaptive acquisition of the channel state information which is used by the receiver to separate multiple co-channel users. Its accuracy will determine the interference suppression performance, and therefore, the overall effective capacity/quality of the links.

In Time Division Duplexed (TDD) Massive MIMO scheme, each user terminal appends training (pilot) symbols to the end of their packets. Base station receiver uses these training symbols to estimate the channel and/or its inverse. The estimation error is clearly inversely proportional to the number of training symbols. However, in a mobile environment with relatively short channel coherence times, the pilot interval can not be held arbitrarily long, which causes a serious limitation on the accuracy of the estimates.

The optimal training length for Massive MIMO has been investigated in [5,6]. The authors show that the capacity maximizing choice for the number of pilot symbols is equal to the number of user terminals K. This number is also the minimum length to obtain (over)determined equations to estimate channel based on pilot signals. Therefore, this result can be interpreted as the optimal pilot length is equivalent to the minimum possible number of training symbols. The minimum value being equal to K is based on the assumption that the receiver uses only training symbols to make a supervised estimate of the channel matrix. We can further reduce the required training length by utilizing the information symbols outside the training region for the adaptation process.

Along this direction, Nayebi-Rao recently proposed a semi-blind scheme for channel estimation that exploits data symbols [7, 8]. This approach is based on some Gaussian Mixture assumptions on the data symbol prior. Although their focus is for a setup, in which training length is greater than or equal to K, the authors demonstrate that the proposed algorithm provides significant improvement in reducing channel estimation error relative to Maximum Likelihood channel estimate based solely on training symbols.

As another alternative, so called “compressed training” is proposed as a semi-blind equalization framework [9]. This scheme exploits the special rectangular QAM constellation structure of the information symbols as an unsupervised resource. As an important result, in the Multiple Input Multiple Output (MIMO) setting, the required training length can be reduced to the order of $\log(K)$ instead of K [10].

In this article, we propose a Maximum A Posteriori Estimation framework based on alternating minimization algorithm. The proposed framework is “channel estimation” counterpart of the compressed training based adaptive equalization approach and reduces the required pilot sequence length significantly below the K limit, which is on par with the pilot length requirement of the compressed training equalization approach.

Similar to this work, there are many approaches for joint source-channel estimation in the literature. In [11], source-channel and soft-bit source decoders are combined to quantify extrinsic value of bit source coding. Moreover, a priori information about the channel inputs is updated based on the iterative algorithm at every transmission. Other approaches exploit redundancies due to a posteriori probability of bit indexes and their softbit-source decoding [12–14]. Liang et al. [15] proposed to utilize superimposed pilots instead of preamble inputs to decrease the delay and throughput of wideband WLAN by estimating the variance of channel estimation error. However, none of them approximate distribution of transmission signals as continuous uniform random variables. This approximation leads to minimization of a quadratic objective function with respect to channel and data matrices under an ℓ_∞-norm constraint on source symbols. Also, these approaches do not target to reduce the required pilot sequence length significantly below the K limit. We should underline that this reduction is achieved under no assumption on the sparsity of channel, which can be additionally exploited to achieve further improvement.

The organization of the article is as follows: Section 2 describes the Massive MIMO data model. The proposed MAP based channel estimation approach is introduced in Section 3. Numerical examples to illustrate the performance of the proposed algorithm are provided in Section 4. Finally, Section 5 is the conclusion.

2. MULTIUSER MIMO DATA MODEL

We assume the following uplink multiuser MIMO scenario:

- There are K user terminals (with single antenna).
- The packet length is Γ, where the first τ_D symbols are data symbols. The remaining $\tau_T = \Gamma - \tau_D$ symbols are reserved for training.
- $\{s_k(n) \in Q_\beta : n \in \{1, \ldots, \Gamma\}\}$ represent the transmission packet of the k^{th} user, where $k \in \{1, \ldots, K\}$. Here $Q_\beta = \{a_R + i a_I : a_R, a_I \in P_{\sqrt{\beta}}\}$ is the rectangular $\beta-$QAM constellation set, where $P_{\sqrt{\beta}} = \{-\sqrt{\beta}+1, -\sqrt{\beta}+3, \ldots, \sqrt{\beta}-3, \sqrt{\beta}-1\}$.
- The source vector at time instant n is defined as $\boldsymbol{s}(n) = \left[\begin{array}{cccc} s_1(n) & s_2(n) & \ldots & s_K(n) \end{array}\right]^T$.
- $\boldsymbol{S} \in \mathbb{C}^{K \times \Gamma}$ represents the uplink transmission packet, which is given by

$$\boldsymbol{S} = \left[\begin{array}{cccc} \boldsymbol{s}(1) & \boldsymbol{s}(2) & \ldots & \boldsymbol{s}(n) \end{array}\right], \tag{1}$$

  and which can be partitioned as

$$\boldsymbol{S} = \left[\begin{array}{cc} \boldsymbol{S}_D & \boldsymbol{S}_T \end{array}\right], \tag{2}$$

  where $\boldsymbol{S}_D \in \mathbb{C}^{K \times \tau_D}$ is the data symbols matrix and $\boldsymbol{S} \in \mathbb{C}^{K \times \tau_T}$ is the training symbols matrix.

- There are M base station antennas. The receiver samples at antenna-l is represented by $\{y_l(n) : n \in \{1, \ldots, \Gamma\}\}$. The received packet matrix at the base station is defined as $\boldsymbol{Y} \in \mathbb{C}^{M \times \Gamma}$ which is also partitioned into data and training regions as follows:

$$\boldsymbol{Y} = \left[\begin{array}{cc} \boldsymbol{Y}_D & \boldsymbol{Y}_T \end{array}\right]. \tag{3}$$

- The linear mapping between user terminals and base station antennas is represented by $\boldsymbol{H} \in \mathbb{C}^{M \times K}$, where we assume flat fading scenario, potentially corresponding to a single OFDM channel in the frequency selective case.
- The mapping between transmitted and received packets is given by

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{V}, \tag{4}$$

  where $\boldsymbol{V} \in \mathbb{C}^{M \times \Gamma}$ is the receiver noise. We can partition the noise matrix as

$$\boldsymbol{V} = \left[\begin{array}{cccc} \boldsymbol{v}(1) & \boldsymbol{v}(2) & \ldots & \boldsymbol{v}(\Gamma) \end{array}\right], \tag{5}$$

  where $\boldsymbol{v}(n)$ is the noise vector at time instant n.

3. PROPOSED APPROACH

We propose a Bayesian approach that exploits the special rectangular constellation structure for QAM sources. For this stochastic estimation approach, we model

- The receiver noise is complex Additive White Gaussian Noise (AWGN), with independent samples in spatial and time dimensions. In other words,

$$\boldsymbol{v}(n) \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \sigma_v^2 \boldsymbol{I}), \tag{6}$$

$$E(\boldsymbol{v}(m)\boldsymbol{v}(n)^H) = \sigma_v^2 \boldsymbol{I} \delta_{m-n}. \tag{7}$$

- The channel matrix $\boldsymbol{H}$ has i.i.d. complex Gaussian entries, i.e.,

$$H_{ij} \sim \mathbb{C}\mathcal{N}(0, \sigma_h^2), \tag{8}$$

$$E(H_{ij} H_{kl}^*) = \sigma_h^2 \delta_{i-k} \delta_{j-l}. \tag{9}$$

- QAM information symbols $s_k(n), n \in \{1, \ldots, \tau_D\}$ are discrete valued with probability mass function (pmf),

$$p_{s_k(n)}(a) = \frac{1}{\beta}, \quad a \in Q_\beta. \tag{10}$$

  The tractability of the algorithm derivation and analysis is affected since QAM symbols have a discrete distribution. Although, approaches based on Finite Alphabet are always an option, we follow an alternative route to ease this difficulty. The main relieving assumption is to assume that transmission symbols are continuous uniform over the convex hull of Q_β. Accordingly, the probability density function (pdf) of $s_k(n)$ becomes a continuous complex uniform, i.e.,

$$\mathbb{R}e\{s_k(n)\}, \mathbb{I}m\{s_k(n)\} \sim \mathcal{U}[-\sqrt{\beta}+1, \sqrt{\beta}-1], \tag{11}$$

  where $\mathbb{R}e\{s_k(n)\}$ and $\mathbb{I}m\{s_k(n)\}$ are independent.

Based on this stochastic construction, we formulate the joint Maximum A Posteriori (MAP) estimation problem for the channel $\boldsymbol{H}$ and the data symbols $\boldsymbol{S}_D$ as

$$\underset{\boldsymbol{H},\boldsymbol{S}_D}{\text{maximize}} \log(f(\boldsymbol{H},\boldsymbol{S}_D|\boldsymbol{Y};\boldsymbol{S}_T)) \tag{12}$$

where $f(\cdot)$ is the probability density function. By the application of the Bayes rule, the objective function can be rewritten as

$$\log(f(\boldsymbol{H},\boldsymbol{S}_D|\boldsymbol{Y};\boldsymbol{S}_T)) = \log(f(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{S}_D;\boldsymbol{S}_T)) + \log(f(\boldsymbol{H}) + \log(\boldsymbol{S}_D)) - \log(f(\boldsymbol{Y})). \tag{13}$$

Regarding the components of the expression above:

- The first term is equivalent to

$$\log(f(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{S}_D;\boldsymbol{S}_T)) = -\Gamma\log(\sqrt{2\pi}\sigma_v) - \frac{1}{2\sigma_v^2}(\|\boldsymbol{Y}_T-\boldsymbol{H}\boldsymbol{S}_T\|_F^2 + \|\boldsymbol{Y}_D-\boldsymbol{H}\boldsymbol{S}_D\|_F^2), \tag{14}$$

- The second term is equivalent to

$$\log(f(\boldsymbol{H})) = -MK\log(\sqrt{2\pi}\sigma_h) - \frac{1}{2\sigma_h^2}\|\boldsymbol{H}\|_F^2, \tag{15}$$

- The third term is

$$\log(f(\boldsymbol{S}_D)) = \begin{cases} -2\tau_D\log(2(\sqrt{\beta}-1)), & \left\|\begin{bmatrix} vec(\mathbb{R}e\{\boldsymbol{S}_D\}) \\ vec(\mathbb{I}m\{\boldsymbol{S}_D\})\end{bmatrix}\right\|_\infty \leq \sqrt{\beta}-1 \\ -\infty, & \text{otherwise.}\end{cases}$$

Ignoring the constant terms, the optimization setting in (12) is equivalent to

MAP Optimization Setting

$$\begin{aligned} \underset{\boldsymbol{H},\boldsymbol{S}_D}{\text{minimize}} \quad & \|\boldsymbol{Y}_T-\boldsymbol{H}\boldsymbol{S}_T\|_F^2 + \|\boldsymbol{Y}_D-\boldsymbol{H}\boldsymbol{S}_D\|_F^2 \\ & + \tfrac{\sigma_v^2}{\sigma_h^2}\|\boldsymbol{H}\|_F^2, \\ \text{subject to} \quad & \left\|\begin{bmatrix} vec(\mathbb{R}e\{\boldsymbol{S}_D\}) \\ vec(\mathbb{I}m\{\boldsymbol{S}_D\})\end{bmatrix}\right\|_\infty \leq \sqrt{\beta}-1. \end{aligned}$$

Although this optimization setting is non-convex, the algorithmic recipe can be produced based on the fact that it is convex in $\boldsymbol{H}$ or $\boldsymbol{S}_D$ given the other argument is fixed.

Alternating projection based algorithms for non-convex settings have long been used for their superior empirical performance in applications such as dictionary learning [16], despite the fact that their optimality has long been debated. Recently there has been significant advance in the theoretical findings regarding the global convergence proofs of such algorithms [17, 18]. Therefore, we can adopt alternating minimization strategies which are successfully employed especially for non-convex programs in machine learning such as sparse coding.

Algorithm 1 Alternating Minimization Based Channel Estimation Algorithm.

1: (Initialize) $\boldsymbol{H}^{(0)} \leftarrow \boldsymbol{Y}_T\boldsymbol{S}_T^*(\boldsymbol{S}_T\boldsymbol{S}_T^* + \frac{\sigma_v^2}{\sigma_h^2}\boldsymbol{I})^{-1}$, $i \leftarrow 1$,
2: (Update $\boldsymbol{S}_D$) $\boldsymbol{S}_D^{(i)} \leftarrow \mathcal{P}_{\ell_\infty}\{\boldsymbol{H}^{(i-1)^\dagger}\boldsymbol{Y}_D\}$,
3: (Update $\boldsymbol{H}$) $\boldsymbol{H}^{(i)} \leftarrow (\boldsymbol{Y}_D\boldsymbol{S}_D^{(i)^*} + \boldsymbol{Y}_T\boldsymbol{S}_T^*)(\boldsymbol{S}_D^{(i)}\boldsymbol{S}_D^{(i)^*} + \boldsymbol{S}_T\boldsymbol{S}_T^* + \frac{\sigma_v^2}{\sigma_h^2}\boldsymbol{I})^{-1}$,
4: If *Stopping Condition* is False then Set $i \leftarrow i+1$ and Go to Step 2.

The proposed alternating minimization based algorithm is shown in (**Algorithm 1**). It starts with initializing $\boldsymbol{H}$ based on the training signals, $\boldsymbol{S}_T$, only (Step 1). Then it iterates between updating data symbols $\boldsymbol{S}_D$ and the channel matrix $\boldsymbol{H}$. Step 2 is the iteration for $\boldsymbol{S}_D$ which corresponds to finding the minimum of the quadratic cost above, with $\boldsymbol{H}$ fixed and then projecting the result to the ℓ_∞-norm boxes for the real and imaginary components. Therefore, the projection operator $\mathcal{P}_{\ell_\infty}$ in Step 2 is elementwise clipping operation applied to real and imaginary parts:

$$[\mathcal{P}_{\ell_\infty}\{\boldsymbol{S}_D\}]_{k,l} = C_{\sqrt{\beta}-1}(\mathbb{R}e\{[\boldsymbol{S}_D]_{k,l}\}) + iC_{\sqrt{\beta}-1}(\mathbb{I}m\{[\boldsymbol{S}_D]_{k,l}\})$$

where $C_{\sqrt{\beta}-1}(x)$ is the clipping function given by

$$C_{\sqrt{\beta}-1}(x) = \begin{cases} x & |x| \leq \sqrt{\beta}-1, \\ sign(x)(\sqrt{\beta}-1) & \text{otherwise.}\end{cases} \tag{16}$$

Step 3 correspond to the minimization of the quadratic cost function in MAP Optimization Setting with respect to $\boldsymbol{H}$ given that $\boldsymbol{S}_D$ is fixed. Step 2 and Step 3 are repeated in sequence till a stopping condition (based on the number of iterations and/or the cost function improvement) is achieved.

4. NUMERICAL EXPERIMENTS

In this section, we present some numerical experimental results for the proposed approach. We assume a scenario with $M = 500$ base station antennas and $K = 40$ users. We assume that the packet length is equal to $\Gamma = 300$. Data symbols and training symbols are based on 4-QAM constellation.

In the first set of experiments, we investigate the channel estimation performance $\frac{\|\boldsymbol{H}-\hat{\boldsymbol{H}}\|_F^2}{KM}$ as a function of the training length, for different receiver SNR values. For this purpose, we compare the performance of our algorithm with the following:

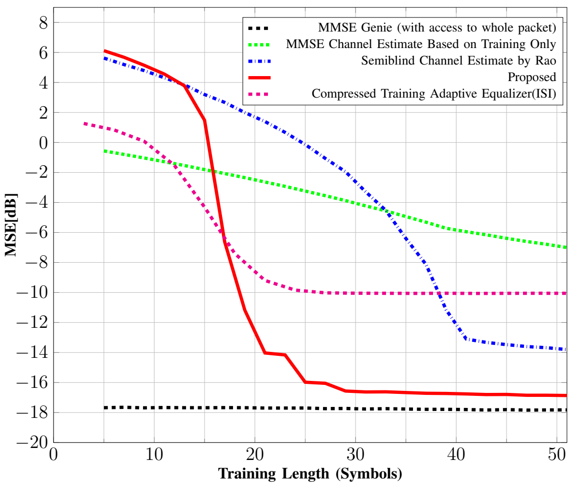

Fig. 1. Channel estimation Mean Square Error (MSE) as a function of training length for 0dB receiver SNR level.

- As a benchmark, we consider MMSE estimate of the channel if we unrealistically assumed that the receiver has full access to the whole transmission packet symbols (data and training). This "Genie aided" case sets the ultimate performance lower bound.
- MMSE channel estimator based on training symbols only,
- Semi-blind channel estimate proposed in [8], by Nayebi and Rao.

Fig. 1 shows the results of these simulations for 0 dB receiver SNR level. We observe a phase transition-like behavior for the proposed algorithm's estimation error curve around training length of $15 - 20$. Its performance is better than both training only and the semi-blind algorithm by Nayebi and Rao. In the same figure, we include the Inter-Symbol-Interference (ISI) level for the compressed training equalizer proposed in [19]. Comparison of this curve with the MSE corresponding to the proposed approach, we can observe that their phase transitions occur around the same training lengths.

Fig. 2 repeats this comparison for a higher level (15dB) receiver SNR level. The phase transition property, around $L_T = 15$ symbols, and the performance advantage of the proposed algorithm is even more pronounced. Furthermore, we can see that the location of this phase transition is very similar to compressed training adaptive equalizer's phase transition location for the ISI performance.

As the final group of experiments, we evaluated Bit Error Rate (BER) versus receiver branch SNR performance for the same scenario. We assume that the receiver employs a Maximum Likelihood (ML) receiver, using the received symbols as well as the channel estimate information. As a performance benchmark, we also included the performance of the ML receiver with the perfect Channel State Information (CSI). For MMSE channel estimator using training symbols only and for

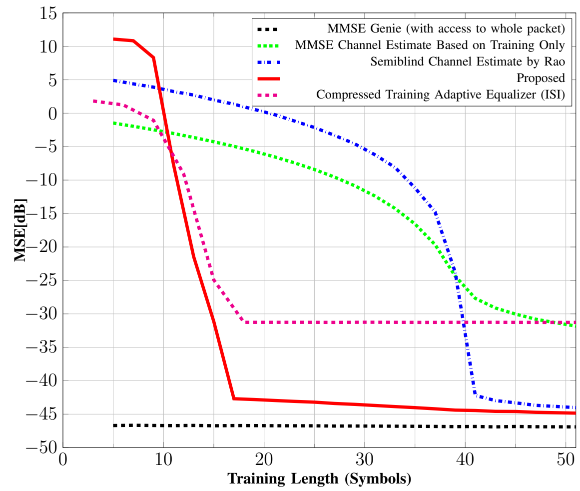

Fig. 2. Channel estimation Mean Square Error (MSE) as a function of training length for 15dB receiver SNR level.

the Semi-blind channel estimator, we assumed the number of training symbols as $L_T = 40$. According to the results shown in Fig. 3, we can see that among all adaptive ML approaches, closest performance to ML with perfect CSI benchmark is obtained by the proposed approach. Even if the training length is reduced to 30, the proposed approach still performs better.

In the same figure, we include the MIMO compressed training based equalizer's ([19]) BER performance where decisions are obtained by slicing the equalizer outputs. It has slightly worse performance than the proposed semi-blind channel estimate based ML approach.

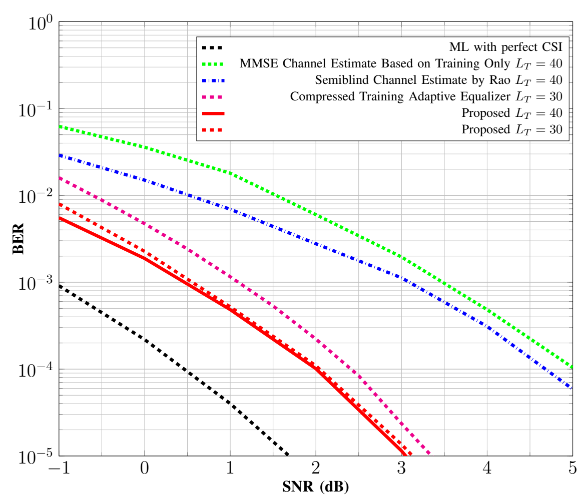

Fig. 3. Bit Error Rate performance as a function of SNR level.

5. CONCLUSION

In this article, we proposed a novel semi-blind channel estimation approach which is demonstrated to significantly reduce the required number of pilot symbols. The phase transition behavior of the channel estimation performance as a function of training length is very similar to that of the equalization performance of the compressed training based adaptation approach. These two algorithms both exploit the special QAM boundedness property. Therefore, the proposed framework can be considered as the channel estimation counterpart of the compressed training based equalization approach.

6. REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[2] Thomas L Marzetta, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[3] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[4] Lu Lu, Geoffrey Ye Li, A Lee Swindlehurst, Alexei Ashikhmin, and Rui Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[5] Hien Quoc Ngo, Michail Matthaiou, and Erik G Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, 2014.

[6] T. L. Marzetta, "How much training is required for multiuser MIMO?," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 359–363.

[7] E. Nayebi and B. D. Rao, "Semi-blind channel estimation in massive MIMO systems with different priors on data symbols," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 3879–3883.

[8] Elina Nayebi and Bhaskar D Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, 2018.

[9] Baki B Yilmaz and Alper T Erdogan, "Compressed training adaptive equalization: Algorithms and analysis," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3907–3921, 2017.

[10] Baki B Yilmaz and Alper T Erdogan, "Compressed training adaptive MIMO equalization," in *The 17th. IEEE International Workshop on Signal Processing Advances in Wireless Communications*. IEEE, July 2016.

[11] Marc Adrat, Peter Vary, and Julian Spittka, "Iterative source-channel decoder using extrinsic information from softbit-source decoding," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*. IEEE, 2001, vol. 4, pp. 2653–2656.

[12] Rainer Bauer and Joachim Hagenauer, "Iterative source/channel decoding based on a trellis representation for variable length codes," in *2000 IEEE International Symposium on Information Theory (Cat. No. 00CH37060)*. IEEE, 2000, p. 238.

[13] N Gortz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1662–1670, 2001.

[14] Jörg Kliewer and Norbert Görtz, "Iterative source-channel decoding for robust image transmission," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2002, vol. 3, pp. III–2173.

[15] Ting-Jung Liang, Wolfgang Rave, and Gerhard Fettweis, "Iterative joint channel estimation and decoding using superimposed pilots in ofdm-wlan," in *2006 IEEE International Conference on Communications*. IEEE, 2006, vol. 7, pp. 3140–3145.

[16] Bruno A Olshausen and David J Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[17] Alekh Agarwal, Animashree Anandkumar, Prateek Jain, Praneeth Netrapalli, and Rashish Tandon, "Learning sparsely used overcomplete dictionaries," in *Proceedings of Conference on Learning Theory*, 2014, pp. 123–137.

[18] Sanjeev Arora, Rong Ge, Tengyu Ma, and Ankur Moitra, "Simple, efficient, and neural algorithms for sparse coding," *Proceedings of Conference on Learning Theory*, July 2015.

[19] Baki Berkay Yilmaz and Alper T Erdogan, "Compressed training based massive mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1191–1206, 2019.