

# Performance Limits of Compressive Sensing Channel Estimation in Dense Cloud RAN

Stelios Stefanatos and Gerhard Wunder

Heisenberg Communications and Information Theory Group,

Freie Universität Berlin,

Takustr. 9, D-14195 Berlin, Germany,

Email: {stelios.stefanatos, g.wunder}@fu-berlin.de

**Abstract**—Towards reducing the training signaling overhead in large scale and dense cloud radio access networks (CRAN), various approaches have been proposed based on the channel sparsification assumption, namely, only a small subset of the deployed remote radio heads (RRHs) are of significance to any user in the system. Motivated by the potential of compressive sensing (CS) techniques in this setting, this paper provides a rigorous description of the performance limits of many practical CS algorithms by considering the performance of the, so called, oracle estimator, which knows *a priori* which RRHs are of significance but not their corresponding channel values. By using tools from stochastic geometry, a closed form analytical expression of the oracle estimator performance is obtained, averaged over distribution of RRH positions and channel statistics. Apart from a bound on practical CS algorithms, the analysis provides important design insights, e.g., on how the training sequence length affects performance, and identifies the operational conditions where the channel sparsification assumption is valid. It is shown that the latter is true only in operational conditions with sufficiently large path loss exponents.

## I. INTRODUCTION

Cloud radio access network (CRAN) is considered as one of the enablers of future cellular networks [1], [2]. In CRAN, multiple low-complexity, low-cost remote radio heads (RRHs) are distributed over the network coverage area and are all connected to a central, cloud-based baseband unit whose task is to jointly process the signals received from or send to the user equipment (UEs) in the system. This centralized-by-design network architecture is, in principle, able to realize the vision of large-scale, multi-cell cooperative networks [3]. However, a major challenge towards this goal is the need for *global* channel state information (CSI) [2], [3], i.e., estimation of the quality of all channel links among the RRHs and the UEs. With a large number of RRHs, the standard training procedure based on orthogonal pilot sequences results in an unacceptably large overhead [4].

Recently, various research efforts have been made towards reducing the training overhead in CRAN. All these works are based on the premise that, in a large-scale CRAN deployment, only a small subset of the RRHs will have significant contribution to the downlink received energy by any UE (similar consideration holds for the uplink as well). Therefore, an (artificial) *channel sparsification* [5], [6] is assumed at the UE side, which considers only the strongest RRH-to-UE links for channel estimation purposes and ignores the remaining links.

This approach effectively allows for reduced training overhead using small length pilot sequences. In [4], the problem of identifying the subset of strongest RRH channels is investigated, assuming *a priori* knowledge of large-scale fading (i.e., RRH-to-UE distances) without, however, providing an explicit channel estimation scheme. RRH-to-UE distances are also assumed known in [7] where a locally orthogonal training scheme is proposed, with short-length training sequences obtained as the solution of a graph-based problem, which, however, depends on the positions of RRHs and UEs. Another approach is considered in [8], [9], where CSI is treated as a sparse vector and estimation algorithms inspired by the compressive sensing (CS) framework [10] are applied. Good estimation performance with small length training sequences is demonstrated only via simulations and for certain system setups. All these works provide valuable insights on the effectiveness of small-length training sequences, however, the *operational conditions*, e.g., propagation losses and small-scale fading statistics, under which the channel sparsification assumption is applicable are not clear.

In this paper, the performance potential of applying CS techniques towards reducing downlink training overhead in large-scale CRAN deployments is investigated. In particular, for a CS-motivated training sequence design that is independent of number and positions of RRHs and UEs, the performance of the, so called, oracle estimator is analytically obtained. The oracle estimator has *a priori* knowledge of the set of RRHs with the strongest links, but not their channel values, and its performance serves as a optimistic estimate for the performance of many practical CS estimation algorithms (that have no *a priori* information) [11], [12].

By employing tools from stochastic geometry, a closed-form upper bound of the oracle estimator mean squared error performance is obtained, that is averaged over the distribution of RRH positions and channel statistics, and is *independent* of the UE position within the CRAN deployment. The bound expression is tight for large scale, dense CRAN deployments and provides insights on how estimation performance is affected by (a) system *design* parameters, e.g., training sequences length and number of (strongest) channels to be estimated, and (b) *operational conditions*, e.g., RRH density and channel statistics. One of the main takeaways of the analysis is that the channel sparsification assumption is not valid for operational

conditions with small path loss exponent, suggesting that training overhead reduction is not possible in these cases, irrespective of the approach to do so (not necessarily CS-based).

*Notation:*  $[N]$  denotes the set  $\{1, 2, \dots, N\}$ .  $(\cdot)^T, (\cdot)^H$  denote transpose and Hermitian transpose, respectively.  $|\cdot|$  will be used to denote the modulus of a complex number, or the (Lebesgue) measure of a set, depending on context. The Frobenius norm of a vector (possibly infinite-dimensional) will be denoted as  $\|\cdot\|$ . The  $n$ -th element of an  $N$ -dimensional vector  $\mathbf{x}$  is denoted as  $[\mathbf{x}]_n, n \in [N]$ , and  $\mathbf{I}$  is the identity matrix of appropriate dimensions.  $\mathbf{X}_{\mathcal{B}}$  ( $\mathbf{x}_{\mathcal{B}}$ ) is the matrix (vector) obtained by considering only the columns (elements) indicated by the set  $\mathcal{B}$ .

## II. SYSTEM MODEL AND PERFORMANCE METRICS

A dense CRAN deployment is considered employing  $N_{\text{RRH}}$  single-antenna RRHs whose positions are independently and uniformly distributed over a large deployment area  $\mathcal{A} \subset \mathbb{R}^2$ , resulting in an RRH *density* (i.e., average number of RRHs per unit area of deployment) equal to  $\lambda \triangleq N_{\text{RRH}}/|\mathcal{A}|$ . Let  $\Phi \subset \mathbb{R}^2$  denote the set of RRH positions and  $h_x \in \mathbb{C}$ , denote the baseband, flat-fading channel on a single OFDM subcarrier of the link between an RRH positioned at  $x \in \Phi$  and a UE located at an arbitrary position  $x_u \in \mathbb{R}^2$ . This channel is commonly modeled as [6], [7], [13]

$$h_x = c_x \|x - x_u\|^{-\alpha/2}, x \in \Phi, \quad (1)$$

where  $\alpha > 2$  is the (deterministic) *path loss exponent*, and  $c_x \in \mathbb{C}$  represents small and/or large scale fading effects. The variables  $\{c_x\}_{x \in \Phi}$  are assumed to be independent, identically distributed (i.i.d.) with  $\mathbb{E}(c_x) = 0$  and  $\mathbb{E}(|c_x|^{4/\alpha}) < \infty$ . The seemingly random last assumption is only a technical one that will be employed later in the analysis and is actually satisfied by many models used in practice for the statistics of  $c_x$  such as Rayleigh and lognormal fading.

Towards acquiring downlink CSI, each RRH transmits a signature training (pilot) sequence that is known by each UE in the system. It is assumed that the transmission of training sequences is done synchronously among RRHs and lasts for  $N_p$  symbols. Let  $\mathbf{p}_x \in \mathbb{C}^{N_p}$  denote the column vector of  $N_p$  pilot symbols transmitted by the RRH positioned at  $x \in \Phi$  with a normalized average transmit power equal to  $\frac{1}{N_p} \mathbb{E}(\mathbf{p}_x^H \mathbf{p}_x) = 1$ , for all  $x \in \Phi$ . The received signal  $\mathbf{y} \in \mathbb{C}^{N_p}$  of the UE at  $x_u$  during the training period equals

$$\mathbf{y} = \sum_{x \in \Phi} h_x \mathbf{p}_x + \mathbf{w} = \mathbf{P} \mathbf{h} + \mathbf{w} \quad (2)$$

where  $\mathbf{h} \in \mathbb{C}^{N_{\text{RRH}}}$  is a vector containing the  $N_{\text{RRH}}$  channel values in some arbitrary order,  $\mathbf{P} \in \mathbb{C}^{N_p \times N_{\text{RRH}}}$  is the *pilot matrix* whose columns are the RRH training sequences with the same ordering as the elements of  $\mathbf{h}$ , and  $\mathbf{w} \in \mathbb{C}^{N_p}$  represents noise, consisting of i.i.d. complex Gaussian elements of zero mean and variance  $\sigma_w^2$ .

Under the conventional, orthogonal training sequence approach, i.e., when it holds  $\mathbf{P}^H \mathbf{P} = N_p \mathbf{I}$ , global CSI can be

acquired at the UE by simply computing  $\frac{1}{N_p} \mathbf{P}^H \mathbf{y}$ . However, the orthogonality condition requires  $N_p \geq N_{\text{RRH}}$ , which results in an unacceptable pilot overhead when  $N_{\text{RRH}} \gg 1$ . This observation motivates the investigation of channel estimation procedures under the condition  $N_p < N_{\text{RRH}}$ , i.e., with small-length training sequences, that are *necessarily non-orthogonal*. Even though this approach does not allow for a reliable estimate of the global CSI (less observations than unknowns), it is motivated by the channel sparsification assumption [6], namely, that out of the  $N_{\text{RRH}}$  elements of  $\mathbf{h}$ , only  $s \ll N_{\text{RRH}}$  are of significant modulus, whereas the others can be safely considered as equal to zero. This, in turn, motivates the use of CS techniques that achieve reliable estimates with  $N_p = \mathcal{O}(s)$  [10].

Unfortunately, the performance of CS estimation algorithms is difficult to describe accurately. For this reason, a commonly used approach is to lower bound their performance by considering the performance of the, so called, *oracle estimator* [12], which is essentially the standard least squares (LS) channel estimator of the  $s$  largest-modulus elements of  $\mathbf{h}$  under the assumption that their positions within  $\mathbf{h}$  are known. In addition to providing a performance bound, the oracle estimator performance will provide insights on how accurate the channel sparsification assumption is under various operational conditions, which is critical for the effectiveness of any short-length training approach (not necessarily based on CS techniques).

Let  $\mathcal{S} \subseteq [N_{\text{RRH}}]$  denote the set of  $s \triangleq |\mathcal{S}| \geq 1$  indices of largest-modulus elements in  $\mathbf{h}$ , which are known by the oracle estimator. The value of  $s$  is treated here as a free design parameter to be optimized in the following. Note that it is not clear at this point of the discussion how to optimally choose  $s$  due to the conflicting requirements of small-length training sequences and accurate global CSI, suggesting small and large values for  $s$ , respectively. Assuming that  $\mathbf{P}_{\mathcal{S}}$  is full column rank (which implies  $N_p \geq s$ ), the LS estimate of  $\mathbf{h}_{\mathcal{S}}$  equals

$$\begin{aligned} \hat{\mathbf{h}}_{\mathcal{S}} &\triangleq (\mathbf{P}_{\mathcal{S}}^H \mathbf{P}_{\mathcal{S}})^{-1} \mathbf{P}_{\mathcal{S}}^H \mathbf{y} \\ &= \mathbf{h}_{\mathcal{S}} + (\mathbf{P}_{\mathcal{S}}^H \mathbf{P}_{\mathcal{S}})^{-1} \mathbf{P}_{\mathcal{S}}^H (\mathbf{P}_{\bar{\mathcal{S}}} \mathbf{h}_{\bar{\mathcal{S}}} + \mathbf{w}), \end{aligned} \quad (3)$$

where  $\bar{\mathcal{S}} \triangleq [N_{\text{RRH}}] \setminus \mathcal{S}$  denotes the indices of elements of  $\mathbf{h}$  that are not estimated. Note that since  $\mathbf{P}$  consists of non-orthogonal columns, the product  $\mathbf{P}_{\mathcal{S}}^H \mathbf{P}_{\bar{\mathcal{S}}}$  in (3) does not vanish, therefore, the pilot transmissions from RRHs corresponding to  $\bar{\mathcal{S}}$  act as *interference* for the channel estimation procedure.

It directly follows from the channel statistics assumptions that the error term of the LS estimate appearing on the right hand side (RHS) of (3) is zero mean, i.e., the LS estimate is unbiased, irrespective of the choice of  $\mathcal{S}$  and  $\mathbf{P}$ . However, the latter do affect the *average* mean squared error ( $\text{MSE}_{\text{av}}$ ) of the estimate, which provides a measure of the estimation accuracy for each element of  $\hat{\mathbf{h}}_{\mathcal{S}}$  and is defined as

$$\begin{aligned} \text{MSE}_{\text{av}} &\triangleq \frac{1}{s} \mathbb{E} \left( \|\mathbf{h}_{\mathcal{S}} - \hat{\mathbf{h}}_{\mathcal{S}}\|^2 \right) \\ &= \frac{1}{s} \mathbb{E} \left( \left\| (\mathbf{P}_{\mathcal{S}}^H \mathbf{P}_{\mathcal{S}})^{-1} \mathbf{P}_{\mathcal{S}}^H (\mathbf{P}_{\bar{\mathcal{S}}} \mathbf{h}_{\bar{\mathcal{S}}} + \mathbf{w}) \right\|^2 \right). \end{aligned}$$

Another related metric for evaluating the channel estimation performance, of particular importance in a CRAN setting, is the *total* mean squared error ( $\text{MSE}_{\text{tot}}$ ), defined as

$$\begin{aligned}\text{MSE}_{\text{tot}} &\triangleq \mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2) + \mathbb{E}(\|\mathbf{h}_{\mathcal{S}} - \hat{\mathbf{h}}_{\mathcal{S}}\|^2) \\ &= \mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2) + s\text{MSE}_{\text{av}}.\end{aligned}\quad (4)$$

$\text{MSE}_{\text{tot}}$  provides an indication of how well the LS estimate  $\hat{\mathbf{h}}_{\mathcal{S}}$  (of  $s$  elements) can be used to approximate the *complete* channel vector  $\mathbf{h}$  (of  $N_{\text{RRH}}$  elements) under the channel sparsification assumption, i.e., by treating the elements of  $\mathbf{h}_{\bar{\mathcal{S}}}$  as zeros.

### III. MEAN SQUARED ERROR PERFORMANCE OF THE ORACLE ESTIMATOR

The MSE performance of the oracle channel estimate depends on multiple parameters, namely,

- positions  $\Phi$  of the RRHs;
- position  $x_u$  of the UE in consideration;
- path loss exponent  $\alpha$ ;
- channel fading statistics;
- number  $s$  of estimated channels;
- pilot matrix  $\mathbf{P}$ .

From this set of parameters, only the last two are *design* parameters that should, in principle, be optimized given the values of the remaining parameters describing *operational conditions*. However, such a design approach is of little practical interest as it is dependent on the UE position, whereas practical systems employ a fixed set of training sequences that (hopefully) results in a good channel estimate at the UE side irrespective of its position.

Towards such a training sequence design and analysis, the following assumption on the pilot matrix will be employed.

**Assumption.** *The elements of the pilot matrix  $\mathbf{P}$  are generated as i.i.d. real-valued Gaussian variables of zero mean and variance 1.*

Although this design choice is not necessarily optimal, it is motivated by noting that, in the observation model of (2),  $\mathbf{P}$  acts as a *sensing* matrix in a standard CS estimation setting. It is well known that a sensing matrix generated as per the previous assumption is a good design choice [10]. Note that  $N_p$  is not specified in the above assumption, i.e., it is a free design parameter to be optimized in the following.

For a Gaussian pilot matrix, a preliminary characterization of the MSE, averaged over all possible pilot matrix realizations can be obtained.

**Lemma 1.** *With Gaussian training sequences, the average mean squared error performance of the oracle estimator considering the  $s \leq N_p - 4$  largest-modulus elements of  $\mathbf{h}$  equals*

$$\text{MSE}_{\text{av}} = \frac{\mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2) + \sigma_w^2}{N_p - s - 1}, \quad (5)$$

whereas the total mean squared error equals

$$\text{MSE}_{\text{tot}} = \frac{(N_p - 1)\mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2) + \sigma_w^2}{N_p - s - 1}. \quad (6)$$

*Proof:* Direct application of [12, Theorem 1] for the LS estimate of  $\mathbf{h}_{\mathcal{S}}$  based on the observation  $\mathbf{y} = \mathbf{P}_{\mathcal{S}}\mathbf{h}_{\mathcal{S}} + (\mathbf{P}_{\bar{\mathcal{S}}}\mathbf{h}_{\bar{\mathcal{S}}} + \mathbf{w})$  yields

$$\text{MSE}_{\text{av}} = \frac{\mathbb{E}(\|\mathbf{P}_{\bar{\mathcal{S}}}\mathbf{h}_{\bar{\mathcal{S}}} + \mathbf{w}\|^2)}{N_p(N_p - s - 1)} = \frac{\mathbb{E}(\|\mathbf{P}_{\bar{\mathcal{S}}}\mathbf{h}_{\bar{\mathcal{S}}}\|^2) + N_p\sigma_w^2}{N_p(N_p - s - 1)}.$$

Noting that  $\mathbb{E}(\|\mathbf{P}_{\bar{\mathcal{S}}}\mathbf{h}_{\bar{\mathcal{S}}}\|^2) = \mathbb{E}(\mathbf{h}_{\bar{\mathcal{S}}}^H \mathbb{E}(\mathbf{P}_{\bar{\mathcal{S}}}^H \mathbf{P}_{\bar{\mathcal{S}}}) \mathbf{h}_{\bar{\mathcal{S}}}) = N_p \mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2)$ , since  $\mathbb{E}(\mathbf{P}_{\bar{\mathcal{S}}}^H \mathbf{P}_{\bar{\mathcal{S}}}) = N_p \mathbf{I}$  by construction, leads to (5). Substituting (5) into (4) leads to (6). ■

*Remark:* The upper bound  $N_p - 4$  for  $s$  appearing in Lemma 1 is a technical one [12] and has no effect on the design, as it will be shown that values of  $s$  far smaller than this bound are of interest.

The expectation appearing on the RHS of (5) is over the statistics of  $\mathbf{h}_{\bar{\mathcal{S}}}$  for a given UE position  $x_u \in \mathbb{R}^2$ . In order to obtain a robust (i.e., worst case) design that is independent of the UE position, it is natural to consider the case where  $x_u$  lies “in the middle” of the RRH deployment area since, in that case, the effect of the interference is expected to be greater, compared to the case when the UE lies on the edge of the RRH deployment area. However, it is not clear which is this “middle” position and how it depends on the RRH deployment area  $\mathcal{A}$ . In order to avoid these issues, the following approximation on the distribution of RRH positions will be considered.

**Approximation.** *The positions of the RRHs are distributed over all  $\mathbb{R}^2$  as a homogeneous Poisson point process (HPPP)  $\Phi \subset \mathbb{R}^2$  of density  $\lambda$ .*

This approximation can be viewed as extending the original RRH deployment area  $\mathcal{A}$  to the whole  $\mathbb{R}^2$ , thus implying an infinite number of RRHs (instead of  $N_{\text{RRH}}$ ), however, with the same density as the actual deployment. Clearly, this approximation leads to a pessimistic performance analysis, since, for a given  $\mathcal{S}$ , the power of the interference term in (3) statistically increases for any UE position. However, this approximation is expected to be accurate under sufficiently large  $N_{\text{RRH}}$  and/or  $\alpha$  for a UE lying “in the middle” of the actual deployment. Note that, due to the stationarity of the HPPP [13], the statistics of  $\mathbf{h}$  (over fading and RRH positions) are independent of the UE position, hence, it will be assumed for convenience that  $x_u = (0, 0)$ , i.e., the UE in consideration lies at the origin of the plane.

Under the HPPP approximation, a closed-form expression for the term  $\mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2)$  appearing in (5) and (6) can be obtained.

**Proposition 2.** *Under the HPPP approximation for the RRH distribution, it holds*

$$\mathbb{E}(\|\mathbf{h}_{\bar{\mathcal{S}}}\|^2) = \frac{2(\pi\lambda\mathbb{E}(|c_x|^{4/\alpha}))^{\frac{\alpha}{2}}\Gamma(s+1-\frac{\alpha}{2})}{(s-1)!(\alpha-2)}, \quad (7)$$

for all  $s > \frac{\alpha}{2} - 1$ , where  $\Gamma(t) \triangleq \int_0^\infty z^{t-1}e^{-z}dz$  is the Gamma function.

*Proof:* See Appendix A. ■

*Remark:* The restriction  $s > \frac{\alpha}{2} - 1$  appearing in Prop. 2 is only a minor one since, as will be shown in the following, much greater values are of interest.

By combining Lemma 1 with Prop. 2, a closed form upper bound for the oracle estimator MSE directly follows.

**Proposition 3.** *Consider a CRAN deployment with an arbitrary number of RRHs independently and uniformly distributed over a large deployment area employing Gaussian training sequences of length  $N_p$ . For any UE in the system, the  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  of the oracle estimate considering the  $s \in (\frac{\alpha}{2} - 1, N_p - 4]$  largest-modulus elements of  $\mathbf{h}$  is upper bounded as*

$$\text{MSE} < \beta \frac{2(\pi\lambda\mathbb{E}(|c_x|^{4/\alpha}))^{\frac{\alpha}{2}} \Gamma(s+1-\frac{\alpha}{2})}{(N_p-s-1)(\alpha-2)(s-1)!} + \gamma \frac{\sigma_w^2}{N_p-s-1}, \quad (8)$$

with  $\{\beta = 1, \gamma = 1\}$  and  $\{\beta = N_p - 1, \gamma = s\}$ , for  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$ , respectively.

The following remarks regarding the upper bound of the MSE are in order:

- It is independent of the number  $N_{\text{RRH}}$  of actually deployed RRHs. This is due to the infinite number of RRHs implied by the HPPP approximation.
- Its dependence on the channel fading statistics is only via the moment  $\mathbb{E}(|c_x|^{4/\alpha})$ , i.e., fading statistics with the same moment are indistinguishable in terms of estimation performance.
- For  $\alpha \rightarrow 2$ , it grows unbounded, irrespective of the values of the other parameters, since the sum power of the interfering signals becomes very large.
- When the noise effect is negligible (i.e.,  $\sigma_w^2 = 0$ ), it grows as  $\mathcal{O}(\lambda^{\alpha/2})$ , implying that RRH densification (i.e., increasing  $\lambda$ ) under conditions with high path loss exponent has more impact on the estimation performance than under small path loss exponents.
- For  $N_p \rightarrow \infty$  it scales as  $\mathcal{O}(1/N_p)$  and  $\mathcal{O}(1)$  for  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$ , respectively, implying that, for a fixed  $s$ , increasing  $N_p$  is beneficial only in terms of accuracy of  $\hat{\mathbf{h}}_S$  but has no effect on the accuracy of the complete channel vector estimate since, by default, the elements of  $\hat{\mathbf{h}}_S$  are set to zero irrespective of  $N_p$ .
- Under conventional, orthogonal training, it is easy to see that  $\text{MSE}_{\text{av}} = \sigma_w^2/N_p$ , irrespective of  $s$ . In contrast, the upper bound MSE expression for the non-orthogonal case indicates that there exists a value of  $s$  that achieves minimum MSE by optimally balancing between two contradicting requirements: (a) consideration of a small  $s$ , suggested by estimation theory in a multiple parameter estimation setting in the presence of noise/interference, and (b) consideration of a large  $s$  in order to reduce both the number and interference power of ignored RRHs.

In regards to the last remark, the closed form expression provided by Prop. 3 allows for an efficient optimization with respect to (w.r.t.)  $s$ , without the need to resort in time-consuming simulations. In particular, for the case of  $\sigma_w^2 = 0$

(i.e., no noise), a closed form asymptotic characterization of the optimal  $s$  is available.

**Proposition 4.** *With  $\sigma_w^2 = 0$ , the upper bounds for  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  of the oracle estimator are minimized for the same value of  $s$ , which is asymptotically equal to (for  $N_p \rightarrow \infty$ )*

$$s^* \sim \frac{(\alpha-2)(N_p-1)}{\alpha}, \quad (9)$$

resulting in the asymptotic upper bound for the minimum MSE

$$\text{MSE}^* < \beta \left( \frac{\alpha\pi\mathbb{E}(|c_x|^{4/\alpha})}{(\alpha-2)} \cdot \frac{\lambda}{N_p-1} \right)^{\alpha/2}, \quad (10)$$

with  $\beta = 1$  and  $\beta = N_p - 1$ , for  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$ , respectively.

*Proof:* With  $\sigma_w^2 = 0$ , it directly follows from Prop. 3 that the optimal  $s$  for both  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  is the one maximizing the term  $\frac{\Gamma(s+1-\frac{\alpha}{2})}{(N_p-s-1)(s-1)!}$ . Numerical examination of this term indicates that the optimal  $s$  is an increasing function of  $N_p$ , which, in turn, suggests that for  $N_p \gg 1$  it will also be  $s^* \gg 1$ . Noting that  $\frac{\Gamma(s+1-\frac{\alpha}{2})}{(s-1)!} \sim s^{1-\alpha/2}$  for  $s \gg 1$ ,  $s^*$  can be found as the minimizer of  $s^{1-\alpha/2}/(N_p-s-1)$ , which is given by (9). Substituting the latter into the (8) results in (10). ■

As it will be shown in the next section, the asymptotic expressions given in Prop. 4 are very accurate even for moderate values of  $N_p$ . Interestingly, the above result shows that when interference is the main source of estimation error, the optimal number of estimated channels is increasing with  $\alpha$  and/or  $N_p$  and is the same for both average and total MSE performance metrics. However, even with optimal  $s$ , the MSE performance severely degrades when  $\alpha \rightarrow 2$ , indicating a fundamental performance limitation of the non-orthogonal training approach under sufficiently small  $\alpha$  as the channel sparsification assumption in that regime is not applicable. For  $\alpha > 2$ , the optimal upper bound scales as  $\mathcal{O}(N_p^{-\alpha/2})$  and  $\mathcal{O}(N_p^{1-\alpha/2})$  for  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$ , respectively, showing that increasing  $N_p$  does help in improving both average and total estimation performance, with a more prominent effect for larger values of  $\alpha$ . This suggests that for operational conditions with sufficiently large  $\alpha$ , good oracle estimator performance is possible with  $N_p \ll N_{\text{RRH}}$ . This, in turn, implies that the channel sparsification assumption is valid in the large  $\alpha$  regime, which can be exploited to reduce training signaling overhead by, e.g., use of CS estimation algorithms.

#### IV. NUMERICAL RESULTS

1) *Comparison of upper bound on  $\text{MSE}_{\text{av}}$  with simulated performance:* A CRAN deployment of  $N_{\text{RRH}} = 500$  RRHs with a density  $\lambda = 1$ , uniformly distributed over a rectangular area  $\mathcal{A}$  was simulated. The channel fading variables  $c_x$  where generated as i.i.d. complex Gaussians of zero mean and variance 1 (Rayleigh fading), and a training sequence length of  $N_p = 81$  was considered with Gaussian pilot symbols, as described in the previous section. Note that this training

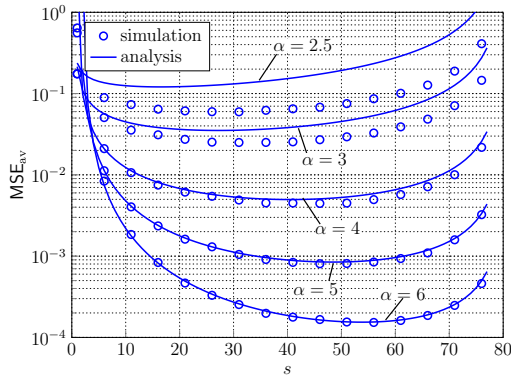


Fig. 1.  $\text{MSE}_{\text{av}}$  vs  $s$  (number of estimated channels) of the oracle estimator. Solid lines corresponds to the bound of Prop. 3 and markers are obtained by simulation.

sequence length corresponds to an 83.8% overhead reduction compared to the minimum requirements under orthogonal training. Figure 1 shows the  $\text{MSE}_{\text{av}}$  of the oracle estimator for a UE located at the center of  $\mathcal{A}$ , as a function of the number  $s$  of estimated channels, for various values of  $\alpha$  and with  $\sigma_w^2 = 0$ . The  $\text{MSE}_{\text{av}}$  values were obtained by averaging over multiple independent realizations of RRH positions and channels. In addition, the upper bound of the  $\text{MSE}_{\text{av}}$ , given in Prop. 3 (with  $\mathbb{E}(|c_x|^{4/\alpha}) = \Gamma(1 + 2/\alpha)$ ), is also shown.

It can be seen that the MSE performance depends strongly on  $\alpha$ , with larger values being beneficial due to reduced interference power, even though the power of the strongest channels is also reduced in this case as well. This provides a strong motivation for reduced length training sequences under operational conditions with large  $\alpha$  (e.g., in urban areas). In addition, the existence of an optimal value of  $s$ , depending on  $\alpha$ , is clear. Note that the analytical upper bound provides an excellent indicator of the performance for  $\alpha \geq 4$  but becomes increasingly loose for values of  $\alpha \rightarrow 2$ . However, the bound is still able to follow the actual performance trends rather closely.

2) *Optimal number of estimated channels*: Figure 2 shows the optimal value of  $s$  ( $s^*$ ) that minimizes the analytical upper bounds for the  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  given in Prop. 3, as a function of  $\alpha$  and for various values of  $N_p$ . A case with no noise ( $\sigma_w^2 = 0$ ) and a case with noise ( $\sigma_w^2 = 0.1$ ) is depicted. As was observed in Prop. 4, the optimal  $s$  is the same for both  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  with no noise. For that case, the asymptotic expression of (9) is also depicted, which can be seen to be an excellent indicator of  $s^*$ , even for moderate values of  $N_p$ . Interestingly, when noise is present, the value of  $s^*$ , even though increasing with  $N_p$ , depicts a non monotonic behavior w.r.t.  $\alpha$ . In particular, for the small  $\alpha$  regime,  $s^*$  increases with  $\alpha$  since the interference power is reduced allowing for the reliable estimation of more (strong) channels. However, beyond a certain value of  $\alpha$ , most of the channel powers become comparable to the noise level, therefore, it is beneficial to concentrate the estimation efforts on a limited set of strongest channels.

3) *Compressive sensing channel estimation*: When the UE

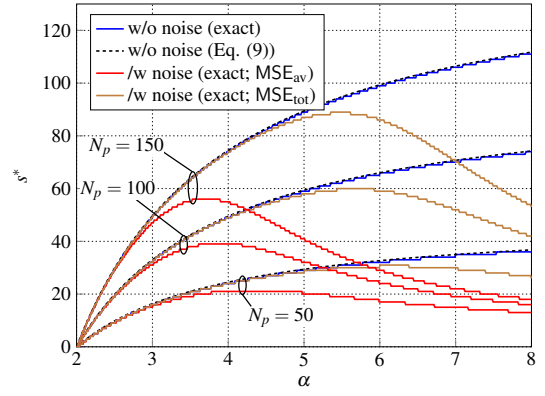


Fig. 2. Optimal  $s$  that minimizes the analytical upper bounds for the  $\text{MSE}_{\text{av}}$  and  $\text{MSE}_{\text{tot}}$  as a function of  $\alpha$  with and without noise.

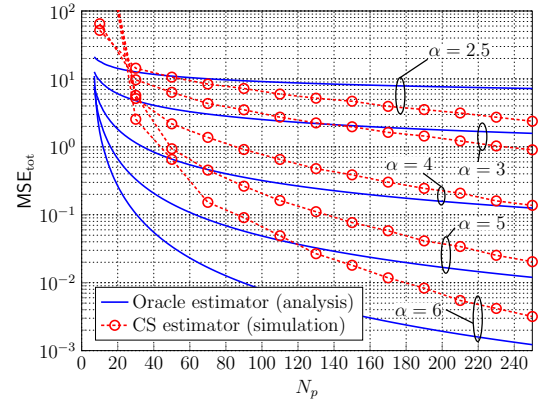


Fig. 3.  $\text{MSE}_{\text{tot}}$  for the CS-based channel estimator as a function of the training sequence length  $N_p$ . The analytical upper bound for the  $\text{MSE}_{\text{tot}}$  of the oracle estimator with optimized  $s$  is also shown.

has no *a priori* knowledge of the set of largest-modulus elements of  $\mathbf{h}$ , a CS estimation approach can be employed for estimating the complete channel vector  $\mathbf{h}$  as long as the channel sparsification assumption is valid [8], [9]. Figure 3 shows the simulated  $\text{MSE}_{\text{tot}}$  performance of the estimator applying the standard Basis Pursuit algorithm on the received signal  $\mathbf{y}$  [10], as a function of  $N_p$  and for various values of  $\alpha$ , for the exact same system setup considered in Fig. 1 ( $N_{\text{RRH}} = 500$ ,  $\sigma_w^2 = 0$ ). It can be seen that the CS estimator performance improves with increasing  $N_p$ , as expected from standard CS theory, and has similar dependence on  $\alpha$  as the oracle estimator, i.e., very good performance is achieved in the large  $\alpha$  regime even with small  $N_p$ . Figure 1 also depicts the  $\text{MSE}_{\text{tot}}$  expression of (8) with optimal  $s$ . It can be seen that it can serve as a reasonable lower bound for the CS-based estimator performance for large  $\alpha$ , whereas this is not the case for small  $\alpha$ , since, as discussed previously, the bound is not tight in this regime.

## V. CONCLUSION

This paper considered the mean squared error performance of the oracle estimator as an attempt to identify the limits

of practical CS-based techniques towards reducing downlink training signaling overhead in CRAN. Using tools from stochastic geometry, an upper bound for the oracle estimator performance was obtained in closed form, clearly demonstrating the effects of design parameters, e.g., training sequence length, and operational conditions, e.g., path loss exponent. It was shown that good estimation performance can be expected with significant training overhead reduction only under sufficiently large path loss exponent.

#### ACKNOWLEDGMENT

This work has been performed in the framework of the Horizon 2020 project ONESG (ICT-760809) receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project. The work of G. Wunder was also supported by DFG grants WU 598/7-1 and WU 598/8-1 (DFG Priority Program on Compressed Sensing).

#### APPENDIX A

##### PROOF OF PROPOSITION 2

Let  $g_s > 0$  denote the received power from the RRH with the  $s$ -th largest modulus channel gain. Expressing  $\|\mathbf{h}_S\|^2$  as

$$\|\mathbf{h}_S\|^2 = \sum_{x \in \Phi} \mathbb{I}(|c_x|^2 \|x\|^{-\alpha} < g_s) |c_x|^2 \|x\|^{-\alpha},$$

where  $\mathbb{I}(\cdot)$  is the indicator (0 – 1) function, it follows that

$$\begin{aligned} \mathbb{E}(\|\mathbf{h}_S\|^2) &\stackrel{(a)}{=} \lambda \int_{\mathbb{R}^2} |x|^{-\alpha} \mathbb{E}(\mathbb{I}(|c_x|^2 \|x\|^{-\alpha} < g_s) |c|^2) dx \\ &\stackrel{(b)}{=} \lambda 2\pi \int_0^\infty r^{1-\alpha} \mathbb{E}(\mathbb{I}(|c|^2 r^{-\alpha} < g_s) |c|^2) dr \\ &\stackrel{(c)}{=} \lambda 2\pi \mathbb{E} \left( |c|^2 \int_{(|c|^2/g_s)^{1/\alpha}}^\infty r^{1-\alpha} dr \right) \\ &= \frac{2\pi \lambda \mathbb{E}(|c|^{4/\alpha})}{\alpha - 2} \mathbb{E}(g_s^{1-2/\alpha}) \end{aligned}$$

where (a) is an application of Cambell's Theorem [13, Theorem A. 2], (b) follows by switching to polar coordinates for the integration and dropping the explicit dependence of  $c$  on  $x$  (by definition  $c_x$  is independent of  $x$ ), and (c) follows by switching the order of integration and expectation (Fubini's theorem).

In order to obtain  $\mathbb{E}(g_s^{1-2/\alpha})$ , the probability distribution function (pdf) of  $g_s$  is pursued next. To this end, note that  $g_s$  represents the  $s$ -th largest element of the one-dimensional point process  $\Pi \triangleq \{g_x = |c_x|^2 \|x\|^{-\alpha}\}_{x \in \Phi}$  obtained by a transformation of the points of  $\Phi$ . It can be shown that  $\Pi$  is an *inhomogeneous* PPP with density function  $\lambda_\Pi(g) = \frac{2\pi}{\alpha} \lambda \mathbb{E}(|c_x|^{4/\alpha}) g^{-(1+2/\alpha)}$ ,  $g \geq 0$  [14, Theorem 4.1]. This, in turn, means that the number  $N_{>\delta}$  of points in  $\Pi$  that are greater than a value  $\delta \geq 0$  is a Poisson random variable of mean

$$\begin{aligned} \mathbb{E}(N_{>\delta}) &= \int_\delta^\infty \lambda_\Pi(g) dg \\ &= \pi \lambda \mathbb{E}(|c_x|^{4/\alpha}) \delta^{-2/\alpha}. \end{aligned}$$

Following the approach of [15], the cumulative distribution function (cdf) of  $g_s$  equals

$$\begin{aligned} \mathbb{P}(g_s < \delta) &= \mathbb{P}(N_{>\delta} < s) = \sum_{k=0}^{s-1} \mathbb{P}(N_{>\delta} = k) \\ &= \sum_{k=0}^{s-1} \frac{(\pi \lambda \mathbb{E}(|c_x|^{4/\alpha}) \delta^{-2/\alpha})^k}{k!} e^{-\pi \lambda \mathbb{E}(|c_x|^{4/\alpha}) \delta^{-2/\alpha}}, \end{aligned}$$

and differentiation of the cdf w.r.t.  $\delta$  gives the probability distribution function (pdf) of  $g_s$  as

$$f_{g_s}(\delta) = \frac{2 (\pi \lambda \mathbb{E}(|c_x|^{4/\alpha}))^s e^{-\pi \lambda \mathbb{E}(|c_x|^{4/\alpha}) \delta^{-2/\alpha}}}{\delta^{\frac{2s}{\alpha}+1} \alpha (s-1)!}, \delta \geq 0.$$

$\mathbb{E}(g_s^{1-2/\alpha})$  can now be computed as  $\int_0^\infty f_{g_s}(\delta) \delta^{1-2/\alpha} d\delta$ , which has a closed form expression for  $s > \frac{\alpha}{2} - 1$ , finally leading to (7).

#### REFERENCES

- [1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.
- [2] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, Third Quarter 2016.
- [3] J. Schreck, G. Wunder, and P. Jung, "Robust iterative interference alignment for cellular networks with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 882–894, Feb. 2015.
- [4] Y. Shi, J. Zhang, and K. Letaief, "CSI overhead reduction with stochastic beamforming for cloud radio access networks," in *Proc. of IEEE Int. Conf. on Commun. (ICC)*, Sydney, Australia, Jun. 2014.
- [5] G. Wunder, H. Boche, T. Strohmer, and P. Jung, "Sparse signal processing concepts for efficient 5G system design," *IEEE Access*, vol. 3, pp. 195–208, 2015.
- [6] C. Fan, Y. J. Zhang, and X. Yuan, "Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1881–1894, Mar. 2016.
- [7] J. Zhang, X. Yuan, and Y. J. Zhang, "Locally orthogonal training design for cloud-RANs based on graph coloring," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6426–6237, Oct. 2017.
- [8] X. Xu, X. Rao, and V. K. Lau, "Active user detection and channel estimation in uplink CRAN systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 2727–2732.
- [9] Q. He, T. Q. S. Quek, Z. Chen, and S. Li, "Compressive channel estimation and multi-user detection in CRAN," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6, Paris.
- [10] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.
- [11] W. Chen, M. Rodrigues, and I. Wassell, "On the use of unit-norm tight frames to improve the average mse performance in compressive sensing applications," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 8–11, Jan. 2012.
- [12] G. Coluccia, A. Roumy, and E. Magli, "Exact performance analysis of the oracle receiver for compressed sensing reconstruction," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014.
- [13] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Found. Trends Netw.*, vol. 3, no. 2, pp. 127–248, 2008.
- [14] S. Mukherjee, *Analytical modeling of heterogeneous cellular networks: Geometry, coverage, and capacity*. Cambridge University Press, 2014.
- [15] M. Haenggi, "On distances in uniformly random networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3584–3586, Oct. 2005.