Orthogonal or Superimposed Pilots? A Rate-Efficient Channel Estimation Strategy for Stationary MIMO Fading Channels

Asyhari, A. T. & Ten Brink, S.

DOI 10.1109/TWC.2017.2665467
ISSN 1536-1276

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Orthogonal or Superimposed Pilots? A Rate-Efficient Channel Estimation Strategy for Stationary MIMO Fading Channels

A. Taufiq Asyhari, *Member, IEEE*, and Stephan ten Brink, *Senior Member, IEEE*

***Abstract*— This paper considers channel estimation for multiple-input multiple-output (MIMO) channels and revisits two competing concepts of including training data into the transmit signal, namely, orthogonal pilot (OP) that periodically transmits alternating pilot-data symbols, and superimposed pilot (SP) that overlays pilot-data symbols over time. We investigate rates achievable by both schemes when the channel undergoes time-selective bandlimited fading and analyze their behaviors with respect to the MIMO dimension and fading speed. By incorporating the multiple-antenna factors, we demonstrate that the widely known trend in which the OP is superior to the SP in the regimes of high signal-to-noise ratio (SNR) and slow fading, and vice versa, does not hold in general. As the number of transmit antennas (n_t) increases, the range of operable fading speeds for the OP is significantly narrowed due to limited time resources for channel estimation and insufficient fading samples, which results in the SP being competitive in wider speed and SNR ranges. For a sufficiently small n_t, we demonstrate that as the fading variation becomes slower, the estimation quality for the SP can be superior to that for the OP. In this case, the SP outperforms the OP in the slow-fading regime due to full utilization of time for data transmission.**

***Index Terms*— Achievable rates, Doppler frequency, generalized mutual information, MIMO, multiple antennas, orthogonal pilots, pilot-aided channel estimation, superimposed pilots.**

I. Introduction

ESTIMATING channel state information (CSI) is indispensable for multiple-input multiple-output (MIMO) wireless communication systems to facilitate reliable transmission. This task can be accomplished using observations from known training sequences (also referred to as pilots [2], [3]), received data symbols (blind methods [4]) or a combination of both (semi-blind methods [4]). Although each method has its own merits, pilot-aided channel estimation has arguably been the widely used technique in most wireless standards (see [5]–[7]), which is likely due to the ability of providing satisfactory estimates at a low complexity under various channel models.

In this paper, we study pilot-aided channel estimation in stationary MIMO fading channels and investigate the effects of fading dynamics to the effective selection of emitted pilot patterns. More specifically, we focus our attention to two competing approaches for transmitting pilots. The first and possibly the most popular approach is *orthogonal pilot (OP)* transmission as considered in [8]–[10], where several time instants are exclusively reserved for transmitting pilots and no data transmission is permitted at those reserved slots. The second approach is *superimposed pilot (SP)* transmission, where pilot symbols are transmitted at the same time instants as the data symbols [3], [11].

A. Previous Works

Most works on pilot-aided channel estimation focused on the signal processing aspects such as mean-squared error (MSE) and bit-error rate (BER) [2], [3], [12]. An earlier work by Cavers [2] underlined the advantage of the OP by showing that despite a loss due to exclusive transmission of pilots, the OP is able to produce a reliable channel estimate at sufficiently high signal-to-noise ratio (SNR), which in turn provides a reasonably good BER performance. Hoeher and Tufvesson [3] demonstrated that the loss in the OP scheme may be significant in fast fading channels due to frequent emittance of pilots and thus suggested that the SP scheme may have a competitive advantage. This insight is further confirmed by Dong *et al.* [12]. Assuming a Gauss-Markov fading process, they showed that in terms of MSE and BER, superimposed pilots can outperform orthogonal pilots for fast-fading channels and/or low SNR [12]. A more comprehensive survey on those works and related studies can be found in [13].

Fundamental insights on pilot-aided channel estimation can be deduced from information-theoretic studies. Most previous studies such as [14], [15], and [9], [10], [16] focused on the information-theoretic analysis of the OP scheme. It has been shown in [10] and [16] that the OP scheme is an attractive choice in the high SNR regime as it achieves the same or nearly the same rate growth as the channel capacity in this regime. At low SNR, it seems clear that relying on orthogonal pilots for channel estimation may not achieve near-capacity performance [10].

Systematic information-theoretic comparisons on the OP and SP schemes were considered by Coldrey and Bohlin [17]

Manuscript received April 11, 2016; revised August 3, 2016 and November 12, 2016; accepted November 24, 2016. Date of publication March 17, 2017; date of current version May 8, 2017. This paper was presented in part at the 2014 International Symposium on Wireless Communication Systems, Barcelona, Spain, August 2014 [1]. The associate editor coordinating the review of this paper and approving it for publication was Y. Zeng.

A. T. Asyhari is with Centre for Electronic Warfare, Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA, UK (e-mail: taufiq-a@ieee.org).

S. ten Brink is with the Institute of Telecommunications, University of Stuttgart, 70569 Stuttgart, Germany (e-mail: tenbrink@inue.uni-stuttgart.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2665467

and Wang *et al.* [18]. In those works, the authors considered a block-fading channel where highly correlated fading coefficients in several adjacent symbols are assumed to be constant and lowly correlated fading coefficients are assumed to be independent. Such block-fading simplifications allow the authors to derive a lower bound to the instantaneous mutual information. By comparing the lower bound for both schemes, they showed that the OP performs better at high SNR while the SP is superior at low SNR.

The high-SNR superiority of the OP due to spatial multiplexing gain is subject to question in the context of the recent interests in large-scale MIMO systems. Largely based on the block-fading channel, references [19] and [20] highlighted the significant dimension cost of channel estimation if the orthogonality of pilots is to be maintained across users and cells. It is therefore intuitive that the MIMO dimension plays an important role in the performance comparison between the OP and SP, and its interplay with the fading dynamics deserves a close investigation.

Block-fading channels used in the existing studies of the OP [19]–[21] as well as rate comparisons between the OP and SP [17], [18] oversimplify modeling the fading dynamics, by underestimating the dependency among the blocks. This model inherently facilitates channel estimation from channel outputs within a block and thus cannot fully capture the effects of time-correlation of the fading.

B. Contributions of Our Work

In this work we exclude the block-fading simplifications and precisely specify the fading memory from symbols to symbols in the analysis as a basis for computing the information rates. More specifically, we focus on a stationary ergodic fading process where the fading dynamics is characterized by a bandlimited power spectral density (PSD). While such a model has been well studied for the OP (see [9], [10], [14]–[16]), limited efforts have been devoted to investigate—from the reliable transmission rate perspective—the same channel model for the SP. One of these few efforts is our preliminary work [1], which provides systematic information-theoretic comparisons between the OP and SP, and demonstrates some similar trends (with respect to the SNR and fading variation) to those observed in the signal-processing results [3], [12]. A main limitation of all these works is the fact that the training period of the OP is always restricted to the inverse of twice the fading bandwidth (proportional to the *coherence time* [22]) and thus limited insights can be drawn from the comparisons.

Building upon all of these existing results, we remove the restriction on the training period of the OP and analyze transmission rates achievable by the OP and SP schemes for any ranges of SNR and fading bandwidth. Removing this restriction allows for more comprehensive understanding of the rate behavior at fast fading where the mobility of devices leads to channel variation that is faster than the training period. By utilizing the framework of *generalized mutual information (GMI)* [23] for stationary MIMO fading channels, we underline that the desirable feature of the OP is interference-free fading observation whereas the strengths of the SP are given by having not only full utilization of time instants for data transmission, but also more fading observations (i.e., the *observation gain*) than the OP for channel estimation.

Our in-depth analysis reveals the following new insights. While interference-free fading observation in the OP facilitates a very accurate fading estimate at high SNR—which in turn leads to a high-SNR logarithmic growth of the rate with appropriate scaling due to multiple antennas as reported in [24] and [16]—, such a rate behavior holds true only if sufficient fading samples can be retained. This sufficiency cannot always be guaranteed in bandlimited MIMO channels because, e.g., fading varies faster than the frequency of the pilot emittance or the MIMO dimension is too large to estimate (i.e., the training period is too short to accommodate training for a large number of transmit antennas). Therefore, in addition to its already-known low-SNR superiority [17], [18], the SP with its efficient time-utilization for transmitting data can also be superior to the OP when the latter cannot maintain enough fading samples. We further demonstrate the inherent attribute of observation gain that reveals a new superiority regime for the SP. For a small transmit dimension, the observation gain enables the SP to produce a reliable fading estimate in the slow-fading regime, which, coupled with efficient time-utilization, makes it superior to the OP in terms of rates.

These new insights provide more comprehensive understanding on the complex interplay among achievable rates, MIMO dimension and fading bandwidth in determining the choice between the OP and SP. Moreover, they can unlock the potentials of the SP in assisting data communication using MIMO technology. Some numerical results exemplify the superiority of the SP in slow-fading single-input multiple-output (SIMO) channels (scenarios of slow mobility: static to typical walking speeds in mobile communications), fast-fading MIMO channels (scenarios of fast mobility: typical train speeds) and across a wide range of the SNR at a sufficiently fast fading variation. Those new insights and results constitute a more complete picture of the rate behaviors in time-varying wireless channels than the existing understanding of the SP's low-SNR and fast-fading superiority for the block-fading [17], [18].

The rest of the paper is organized as follows. Section II describes our channel model. Section III describes the OP scheme and its corresponding achievable rate. Section IV explains the SP scheme alongside its achievable rate. Section V specifically compares the two schemes using a rectangular fading PSD to reveal some important design parameters of the schemes. Section VI provides concluding remarks on the main contribution of the paper.

II. Channel Model

We consider a discrete-time MIMO channel with n_{t} transmit antennas and n_{r} receive antennas. For a given time-k n_{t}-dimensional channel input vector $\boldsymbol{X}_k = \boldsymbol{x}_k$, the channel output at time instant k is given by

$$\boldsymbol{Y}_k = \sqrt{\mathsf{SNR}}\,\mathbb{H}_k \boldsymbol{x}_k + \boldsymbol{Z}_k. \tag{1}$$

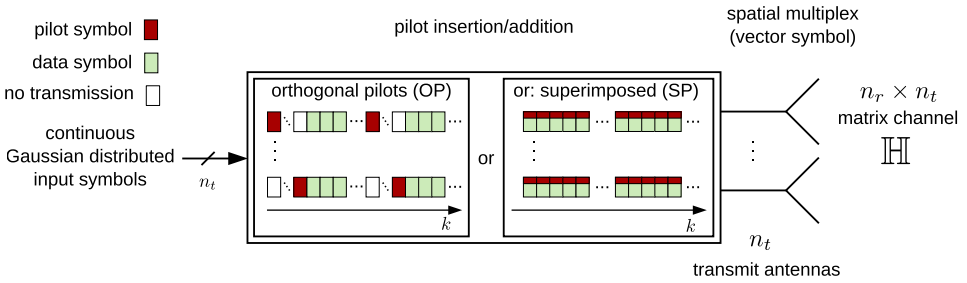

Fig. 1. System model that captures insertion of orthogonal and superimposed pilots.

The channel input $\boldsymbol{X}_k$ is assumed to satisfy the two constraints

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left[\|\boldsymbol{X}_k\|^2\right] = 1, \tag{2}$$

$$\Pr\left\{|X_k(t)|^2 > \rho_{\text{th}}\right\} \le e^{-\rho_{\text{th}}}, \quad \rho_{\text{th}} \ge 1 \tag{3}$$

for all $t = 1, \ldots, n_{\text{t}}$, where $X_k(t)$ denotes the channel input symbol transmitted from the t-th transmit antenna. The constraint (2) corresponds to average power constraint, which is commonly used in the analysis of fading channels. The constraint (3) is applied to limit the peakiness of the input signals with parameter ρ_{th}.[1]

The fading process $\{\mathbb{H}_k, k \in \mathbb{Z}\}$ consists of $n_{\text{t}} \cdot n_{\text{r}}$ independent and identically distributed (i.i.d.) processes $\{H_k(r,t), k \in \mathbb{Z}\}$, $r = 1, \ldots, n_{\text{r}}$, $t = 1, \ldots, n_{\text{t}}$. Each $\{H_k(r,t), k \in \mathbb{Z}\}$ has a bandlimited PSD $f_H(\lambda)$, $-1/2 \le \lambda \le 1/2$, with bandwidth $\lambda_D < 1/2$. The fading PSD is related to the autocorrelation function $A_H(\cdot)$ as

$$A_H(m) \triangleq \mathsf{E}\left[H_{k+m}(r,t)H_k^*(r,t)\right] = \int_{-1/2}^{1/2} e^{\imath 2\pi m\lambda} f_H(\lambda) d\lambda. \tag{4}$$

The noise $\{\boldsymbol{Z}_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. n_{r}-variate Gaussian with zero mean and identity covariance matrix. This noise assumption and equation (2) imply that the average SNR per receive antenna is given by SNR. We assume that $\{\mathbb{H}_k, k \in \mathbb{Z}\}$ and $\{\boldsymbol{Z}_k, k \in \mathbb{Z}\}$ are independent and that their joint law does not depends on $\{\boldsymbol{x}_k, k \in \mathbb{Z}\}$.

In this work, we focus on the *multiplexing* transmission mode of the multiple-antenna systems where independent data streams can be spatially multiplexed over the MIMO channel. Depending on the design of pilot transmission, the channel input vector $\boldsymbol{x}_k, k \in \mathbb{Z}$ can represent either a data vector or a pilot vector or a combination of both. A data vector, which is an element of a codeword, is used to convey a message whereas a pilot vector is used to facilitate channel estimation. Suppose that $\mathcal{M} = \{1, \ldots, \mathsf{M}\}$ is the set of all possible messages. The time-k data vector conveying message m, i.e., $\bar{\boldsymbol{x}}_k(m)$, is an element of the length-n codeword $\bar{\boldsymbol{x}}_1(m), \ldots, \bar{\boldsymbol{x}}_n(m)$. Note that the codeword is drawn i.i.d. from an n_{t}-variate zero-mean complex Gaussian distribution. On the other hand, a pilot vector consists of training symbols that are known at the receiver in order to extract the information about the channel. A transmission rate (in nats per channel use) is defined by

$$R \triangleq \frac{\log \mathsf{M}}{n} \tag{5}$$

and is said to be achievable if the probability of decoding error tends to zero as the codeword length n tends to infinity. Clearly, such an achievable rate depends on how reliable the transmission scheme is. In this work, we consider analysis of achievable rates for the two transmission schemes, namely the OP scheme in Section III and the SP scheme in Section IV.

[1]The role of the constraint in (3) is restricting the peakiness of deterministic signals (such as pilot symbols), but at the same time allowing the use of widely-used Gaussian constellations in the rate analysis.

III. Orthogonal Pilot (OP) Scheme

A. Transmission Scheme

For the OP scheme as illustrated in Fig. 1, data and pilot symbols are transmitted at different time instants. The data vector comes from the codeword $\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_n$, whose entries are drawn i.i.d. from n_{t}-variate Gaussian random vectors with zero mean and covariance matrix $\frac{\rho_d}{n_{\text{t}}}\mathsf{I}_{n_{\text{t}}}$. Herein $\rho_d\mathsf{SNR}$ is the average data SNR. Pilot symbols are periodically emitted every L time instants for channel training. Herein the interval L is also known as (*a.k.a.*) the *training period*. In order to estimate fading coefficients from n_{t} transmit antennas for the OP scheme, we require any set of n_{t} orthogonal pilot vectors to ensure sufficient channel observations. Due to its favorable performance in terms of channel estimation error as reported in [24, Sec. VI] and [10], we particularly adopt the pilot orthogonality in the space/time domain, i.e., by allowing only a single transmit antenna to emit a pilot symbol at a given time instant. This approach implies that a pilot vector to estimate fading from transmit antenna t, $t \in \{1, \ldots, n_{\text{t}}\}$, is given by $\boldsymbol{p}_k$, where $p_k(t) = \sqrt{\rho_p}$ and $p_k(t') = 0$ for $t' \ne t$, $k \in \mathcal{P}$. Herein ρ_p is the fraction of power allocated to pilot symbols and $\mathcal{P}$ denotes the set of time indices for pilot transmission. To estimate a complete entries of fading matrix $\mathbb{H}_k$, n_{t} pilot vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n_{\text{t}}}$ are thus required to be transmitted.

In terms of power fractions ρ_p and ρ_d, the average power constraint (2) can be re-written as

$$\rho_p \frac{n_{\text{t}}}{L} + \rho_d \left(1 - \frac{n_{\text{t}}}{L}\right) = 1. \tag{6}$$

Both ρ_p and ρ_d also have to satisfy the peakiness constraint (3), which can be expressed as

$$\rho_p \le \rho_{\text{th}}, \quad \text{and} \quad \rho_d \le n_{\text{t}}. \tag{7}$$

At the receiver, a decoder consisting of a channel estimator and a data detector is considered. Let $\mathcal{D}$ denotes the set of time indices for data transmission. The channel estimator considers the output $\boldsymbol{Y}_k$, $k \in \mathcal{P}$ to obtain a channel estimate for a transmit-receive antenna pair (r,t) as

$$\hat{H}_{\mathrm{o},k}(r,t) = \sum_{k' \in \mathcal{P}} a_{k,k'}(r,t) Y_{k'}(r), \quad k \in \mathcal{D} \tag{8}$$

where subscript o indicates results with orthogonal pilots. The channel estimation error is defined by the difference between the actual fading and its corresponding estimate, i.e.,

$$E_{\mathrm{o},k}(r,t) = H_k(r,t) - \hat{H}_{\mathrm{o},k}(r,t). \tag{9}$$

On the right-hand side (RHS) of (8), the linear coefficients $a_{k'}(r,t)$ are chosen to minimize the mean-squared channel estimation error. Such an estimator is commonly known as linear minimum mean-squared error (LMMSE) estimator. When the fading $H_k(r,t)$ is Gaussian, the LMMSE estimator achieves the globally minimum mean-squared error (MMSE) [25].

Due to stationary assumption on the fading, it can be shown that irrespective of r and k, the mean-squared error (MSE) for the channel estimator (8), namely

$$\epsilon^2_{\mathrm{o},k}(r,t) = \mathsf{E}\left[\left| H_k(r,t) - \hat{H}_{\mathrm{o},k}(r,t)\right|^2\right], \tag{10}$$

admits a general expression [9], [26]

$$\epsilon^2_{\mathrm{o},\ell}(r,t) = 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}\,|f_{\ell-t+1}(\lambda)|^2}{\mathsf{SNR} f_0(\lambda) + 1} d\lambda \tag{11}$$

where $\ell \triangleq k \bmod L$ denotes the remainder of k/L. Here $f_\ell(\cdot)$ is given by

$$f_\ell(\lambda) \triangleq \frac{1}{L} \sum_{\nu=0}^{L-1} \bar{f}_H\left(\frac{\lambda-\nu}{L}\right) e^{\imath 2\pi \ell \frac{\lambda-\nu}{L}}, \quad \ell = 0, \ldots, L-1 \tag{12}$$

and $\bar{f}_H(\cdot)$ is the periodic function of period $[-1/2, 1/2)$ that coincides with $f_H(\lambda)$ for $-1/2 \le \lambda \le 1/2$.

Once the fading estimates $\{\hat{\mathbb{H}}_{\mathrm{o},k},\ k \in \mathcal{D}\}$ (which consist of matrix entries $\{\hat{H}_{\mathrm{o},k}(r,t),\ k \in \mathcal{D}\}$) are obtained, the channel estimator forwards them to the data detector. Based on the realizations of channel outputs $\{\boldsymbol{y}_k,\ k \in \mathcal{D}\}$ and fading estimates $\{\hat{\mathsf{H}}_{\mathrm{o},k},\ k \in \mathcal{D}\}$, the data detector outputs the message $\hat{m}$ using a nearest neighbor decoding rule

$$\hat{m} = \arg\min_{m \in \mathcal{M}} \sum_{k \in \mathcal{D}} \left\| \boldsymbol{y}_k - \sqrt{\mathsf{SNR}}\, \hat{\mathsf{H}}_{\mathrm{o},k} \bar{\boldsymbol{x}}_k(m) \right\|^2. \tag{13}$$

B. Achievable Rate

Based on the channel model in Section II, we analyze the achievable rate for the scheme in Section III-A using the GMI [23]. Under a fixed decoding rule, the GMI characterizes the largest transmission rate below which the ensemble-average error probability of i.i.d. Gaussian codebooks vanishes as the codeword length increases.

For a communication scheme that utilizes OP-aided channel estimation and nearest neighbor decoding, the achievable rate has been partly derived in [16] when the training period is constrained to $L \le \frac{1}{2\lambda_D}$ (*a.k.a.* no-aliasing condition). The following proposition generalizes the result in [16] by removing the constraint $L \le \frac{1}{2\lambda_D}$.

Proposition 1: Consider the communication scheme in Section III-A. Nearest neighbor decoding together with OP-aided channel estimation achieves a rate

$$R_{\mathrm{o}} = \frac{1}{L} \sum_{\ell=n_{\mathrm{t}}}^{L-1} \mathsf{E}\left[\log_2 \det\left(\mathsf{I}_{n_{\mathrm{r}}} + \frac{\rho_d \mathsf{SNR}}{n_{\mathrm{t}} + \rho_d \mathsf{SNR}\, \epsilon^2_{\mathrm{o},*}} \hat{\mathbb{H}}_{\mathrm{o},\ell} \hat{\mathbb{H}}^\dagger_{\mathrm{o},\ell} \right) \right] \tag{14}$$

where $\hat{\mathbb{H}}_{\mathrm{o},\ell}$ is a channel estimate matrix whose (r,t)-th entry is given by (8), $\epsilon^2_{\mathrm{o},*}$ is defined by

$$\epsilon^2_{\mathrm{o},*} = \max_{\ell \in \{n_{\mathrm{t}}, \ldots, L-1\}} \sum_{t=1}^{n_{\mathrm{t}}} \epsilon^2_{\mathrm{o},\ell}(r,t) \tag{15}$$

and the MSE $\epsilon^2_{\mathrm{o},\ell}(r,t)$ has been given in (11). Subject to the constraints in (6) and (7), R_{o} can be further optimized over the values of ρ_d and ρ_p.

Proof: See Appendix A-A. ∎

The main aim of OP-aided channel estimation is to produce a good quality of estimates as measured by a small value of $\epsilon^2_{\mathrm{o},*}$ in (15). Two basic principles underpin the OP scheme:

- Interference-free fading samples (apart from the background noise) by placing pilot symbols orthogonally from one another and from data symbols;
- Sufficient fading samples by frequent emittance of pilot symbols as reflected by the training interval L.

In achieving this aim, however, inevitable rate-loss occurs due to exclusive transmission of pilot symbols (i.e., over n_{t} time instants per training interval) that effectively reduces time instants for data transmission. As captured by Proposition 1, this rate-loss (*a.k.a.* the dimension cost of the OP scheme) linearly scales with n_{t}, the transmit dimension of the MIMO channel.

The severity of the rate-loss in the OP scheme is further determined by the length of the training interval L. The larger the value of L, the smaller the rate-loss we incur. Depending on how large L with respect to the inverse of twice the Doppler bandwidth, two distinct behaviors of R_{o} can be observed.

1) No-Aliasing Case, $L \le \frac{1}{2\lambda_D}$: This is a common restriction to satisfy sufficiency of fading samples in order to produce reliable estimates [9], [27]. More specifically, if $L \le \frac{1}{2\lambda_D}$ holds, then the MSE (11) is identical for all ℓ, r, t (the estimation error becomes wide-sense stationary) and the achievable rate (14) can simply be expressed as [16], [24]

$$R_{\mathrm{o}} = \frac{L - n_{\mathrm{t}}}{L} \mathsf{E}\left[\log_2 \det\left(\mathsf{I}_{n_{\mathrm{r}}} + \mathsf{SNR}_{\mathrm{oef}} \tilde{\mathbb{H}}_{\mathrm{o}} \tilde{\mathbb{H}}^\dagger_{\mathrm{o}} \right) \right] \tag{16}$$

where $\tilde{\mathbb{H}}_{\mathrm{o}} \triangleq \frac{\hat{\mathbb{H}}_{\mathrm{o},1}}{1 - \epsilon^2_{\mathrm{o},*}/n_{\mathrm{t}}}$ has unit variance entries, and

$$\mathsf{SNR}_{\mathrm{oef}} \triangleq \frac{\rho_d \mathsf{SNR}\left(1 - \epsilon^2_{\mathrm{o},*}/n_{\mathrm{t}}\right)}{n_{\mathrm{t}} + \rho_d \mathsf{SNR}\, \epsilon^2_{\mathrm{o},*}}, \tag{17}$$

$$\epsilon^2_{\mathrm{o},*} = n_{\mathrm{t}} \times \left(1 - \int_{-1/2}^{1/2} \frac{\rho_p \mathsf{SNR}[f_H(\lambda)]^2}{\rho_p \mathsf{SNR} f_H(\lambda) + L} d\lambda \right). \tag{18}$$

No-aliasing ensures that at high SNR, the MSE (18) vanishes as $\mathcal{O}(\mathsf{SNR}^{-1})$ and the rate R_{o} grows logarithmically with SNR as captured by [16]

$$R_{\mathrm{o}} \approx \left[1 - \frac{n_{\mathrm{t}}}{L}\right] \min(n_{\mathrm{t}}, n_{\mathrm{r}}) \log_2 \mathsf{SNR}. \tag{19}$$

This approximation shows that reliable estimates $\hat{\mathbb{H}}_{\mathrm{o},\ell}$ due to vanishing MSE in the OP scheme has brought out the *coherent MIMO multiplexing gain* [28] of $\min(n_{\mathrm{t}}, n_{\mathrm{r}})$. Consequently, the effective multiplexing gain is the coherent MIMO multiplexing gain linearly scaled by a pre-factor of $(1 - n_{\mathrm{t}}/L)$, which is the fraction of time for data transmission.

2) Aliasing Case, $L > \frac{1}{2\lambda_D}$: As the mobile terminals can move at a very high speed, aliasing may be unavoidable, which results in insufficient fading samples. Under this condition, the MSE (11) can be lower-bounded by [26]

$$\epsilon^2_{\mathrm{o},\ell}(r,t) \geq \frac{2\left[1 - \cos\left(\frac{2\pi[\ell - t + 1]}{L}\right)\right]}{L^2} \times \int_{-1/2}^{1/2} \frac{\rho_p \mathsf{SNR} \bar{f}_H\left(\frac{\lambda}{L}\right) \bar{f}_H\left(\frac{\lambda - 1}{L}\right)}{\rho_p \mathsf{SNR} f_0(\lambda) + 1} d\lambda \tag{20}$$

where we have recalled $\bar{f}_H(\lambda)$ as the periodic function in (12). Due to overlapping of $\bar{f}_H\left(\frac{\lambda}{L}\right)$ and $\bar{f}_H\left(\frac{\lambda-1}{L}\right)$ within the interval of $\lambda \in [-1/2, 1/2)$, it can be shown that the RHS of (20) is bounded away from zero. As such, the channel estimates are no longer reliable, which result in a bounded rate R_{o} at high SNR.

In order to fully operate within no aliasing boundary, existing works (see [9], [14]–[16]) commonly employed the *OP with adaptive L* (OPAL) where L is a function of Doppler bandwidth, given by $L = L^* = \lfloor 1/(2\lambda_D) \rfloor$. This choice of L^* is well-founded from the perspectives of maximizing R_{o} under no aliasing constraint [9], achieving the best known multiplexing gain of noncoherent MIMO channels [16], [22], [29] and providing foundation for a widely-celebrated block-fading model [10], [24], [30]. However, it seems to be agreed that the desirable attributes of the OPAL can only be realized at high SNR and a conservative range of antenna sizes. Specifically, the vanishing MSE of $\mathcal{O}(\mathsf{SNR}^{-1})$ at high SNR would be less beneficial at low SNR. Furthermore, the dependency of $L^* = \lfloor 1/(2\lambda_D) \rfloor$ limits the supported MIMO dimension. As captured by (19), i.e., if $n_{\mathrm{t}} > 1/(2\lambda_D)$, then the OPAL cannot achieve any positive R_{o} as the training period cannot support complete channel estimation. This is problematic when aiming to reap the MIMO gain $\min(n_{\mathrm{t}}, n_{\mathrm{r}})$ using massive transmit and receive antennas.

In some wireless standards, the *OP with a fixed L* (OPFL) is preferred and intended to operate for a wide-range of mobile speeds [31], [32]. In this case, the number of transmit antennas may no longer be limited by the fading speed, but at the same time, aliasing may occur. A smaller value of L leads to a wider no-aliasing range of λ_D at the expense of a larger linear rate loss that is proportional to $1 - n_{\mathrm{t}}/L$, and vice-versa.

IV. Superimposed Pilot (SP) Scheme

A. Transmission Scheme

For the SP scheme as illustrated in Fig. 1, pilot symbols are transmitted at the same time instant as data symbols, i.e., the channel input vector at time k is given by

$$\boldsymbol{x}_k = \boldsymbol{p} + \bar{\boldsymbol{x}}_k, \quad k \in \mathbb{Z} \tag{21}$$

where we have recalled $\bar{\boldsymbol{x}}_k$ as the time-k data vector and $\boldsymbol{p}$ denotes the pilot vector that superimposes data symbols. For simplicity, we assume that the pilot vector $\boldsymbol{p}$ is time-invariant and its entries are identical for all transmit antennas and given by $p(t) = \sqrt{\frac{\varrho_p}{n_{\mathrm{t}}}}$, $t \in \{1, \ldots, n_{\mathrm{t}}\}$. The data vectors $\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_n$ are drawn i.i.d. from n_{t}-variate complex Gaussian distribution with zero mean and covariance matrix $\frac{\varrho_d}{n_{\mathrm{t}}} \mathsf{I}_{n_{\mathrm{t}}}$. We recall that the channel input vector $\boldsymbol{x}_k$ satisfies the average power constraint (2), under which the pilot power fraction ϱ_p and data power fraction ϱ_d need to satisfy

$$\varrho_p + \varrho_d = 1. \tag{22}$$

The peakiness constraint (3) is automatically satisfied as $\varrho_p \leq 1$ and $\varrho_d \leq 1$ to validate (22).

Similarly to the OP scheme, at the receiver we employ a decoder that performs separate channel estimation and data detection. Note that for superimposed pilots, the sets of time indices for pilot transmission ($\mathcal{P}$) and for data transmission ($\mathcal{D}$) coincide, i.e., $\mathcal{P} = \mathcal{D}$. Due to the stationarity of the fading channel, the channel input and the additive noise, the time-k channel estimate $\hat{\mathbb{H}}_{\mathrm{s},k}$ can justifiably be implemented by a time invariant function of the observation, i.e.,

$$\hat{\mathsf{H}}_{\mathrm{s},k} = f(\{\boldsymbol{y}_{k'}\}_{k' \in \mathcal{P}}) \tag{23}$$

where subscript $_{\mathrm{s}}$ denotes any result with superimposed pilots. The optimal function on the RHS of (23) that minimizes the MSE—defined by $\mathsf{E}[\|\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\|^2]$—is given by

$$\hat{\mathsf{H}}_{\mathrm{s},k} = \mathsf{E}\left[\mathbb{H}_k | \{\boldsymbol{Y}_{k'} = \boldsymbol{y}_{k'}, k' \in \mathcal{P}\}\right]. \tag{24}$$

Evaluating the expectation in (24) is not only intractable due to data interference in the observation $\boldsymbol{Y}_{k'}$ for all $k' \in \mathcal{P}$, but also leading to undesirable characteristics (such as correlation among the channel estimate $\hat{\mathbb{H}}_{\mathrm{s},k}$, data $\bar{\boldsymbol{X}}_k$ and noise $\boldsymbol{Z}_k$) for mutual information and achievable rate evaluation (see the discussion in Appendix A-B and also in [33] and [34]). In order to circumvent those difficulties, we consider a suboptimal linear channel estimator that instead of considering observations at all $k' \in \mathcal{P}$, only takes into account those at $k + 2k' - 1$ for $k' \in \mathcal{P}$. More specifically, assuming the large codeword length ($n \to \infty$) for the achievable rate analysis, the channel estimator produces a time-k channel estimate for a transmit-receive antenna pair (r, t) as[2]

$$\hat{H}_{\mathrm{s},k}(r,t) = \sum_{k'=-\infty}^{\infty} b_{k'}(r,t) Y_{k+2k'-1}(r), \tag{25}$$

[2]A more precise analysis follows a similar treatment for orthogonal pilots [16], [35], which uses guard bands of size T before and after the main transmission block where T scales with n but at a sub-linear growth. The interpolator to produce fading estimates can then be defined over a window of $2T$ superimposed pilots. As $n \to \infty$, we recover the interpolator in (25).

where the coefficients $b_{k'}(r,t)$ are chosen to minimize the following scalar MSE

$$\mathsf{E}\left[|E_{\mathrm{s},k}(r,t)|^2\right] = \mathsf{E}\left[|H_k(r,t) - \hat{H}_{\mathrm{s},k}(r,t)|^2\right]. \quad (26)$$

We refer to the estimator (25) as a *single-gap linear interpolator*, motivated by the fact that consecutive observation timings, e.g., $k+2k'-1$ and $k+2(k'+1)-1$, are separated by a single time instant to ensure no correlation among the time-k triplet $(\hat{\mathbb{H}}_{\mathrm{s},k}, \bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k)$. Note that for each (r,t), the estimation suffers from extra interferences (in addition to additive noise) due to data symbols and pilot symbols from $t' \neq t$ (i.e., inter-antenna pilot interference).

As derived in Appendix B, the minimum MSE defined by (26) admits the following expression

$$\epsilon^2_{\mathrm{s},k}(r,t) = 1 - \int_{-1/2}^{1/2} \frac{\varrho_p \mathsf{SNR}\,|f_1(\lambda)|^2}{n_{\mathrm{t}}\varrho_p \mathsf{SNR}\, f_0(\lambda) + n_{\mathrm{t}}\varrho_d \mathsf{SNR} + n_{\mathrm{t}}} d\lambda \quad (27)$$

where $f_l(\lambda)$, $l = 0, 1$ is given by

$$f_l(\lambda) = \frac{1}{2}\sum_{\nu=0}^{1} \bar{f}_H\left(\frac{\lambda-\nu}{2}\right) e^{\imath 2\pi l \frac{\lambda-\nu}{2}}. \quad (28)$$

The expression on the RHS of (27) captures contributing factors that determine the accuracy of channel estimation, i.e., the PSD function, the data power $\varrho_d \mathsf{SNR}$, the pilot power $\varrho_p \mathsf{SNR}$, the number of transmit antennas n_{t} and the additive noise.

Upon obtaining the fading estimates $\{\hat{\mathbb{H}}_{\mathrm{s},k} = \hat{\mathsf{H}}_{\mathrm{s},k},\ k \in \mathbb{Z}\}$, where the (r,t)-th entry of $\hat{\mathbb{H}}_{\mathrm{s},k}$ is given by (25), the channel estimator feeds them to the data detector, which will then employ a nearest neighbor decoding rule to decide the message output. The decoding rule considers the realizations of channel outputs $\boldsymbol{y}_k$, fading estimates $\hat{\mathsf{H}}_{\mathrm{s},k}$ and the pilot vector $\boldsymbol{p}$ to select the message $\hat{m}$ such that

$$\hat{m} = \arg\min_{m\in\mathcal{M}} \sum_{k\in\mathbb{Z}} \left\| \boldsymbol{y}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{\mathrm{s},k}\,(\bar{\boldsymbol{x}}_k(m) + \boldsymbol{p}) \right\|^2. \quad (29)$$

B. Achievable Rate

To analyze the achievable rate for the SP scheme, we compute the GMI corresponding to the setup in Section IV-A. The resulting achievable rate is given in the following.

Proposition 2: Consider the communication scheme in Section IV-A with the channel estimate matrix $\hat{\mathbb{H}}_{\mathrm{s},k}$. Nearest neighbor decoding with SP-aided channel estimation achieves a rate

$$R_{\mathrm{s}} = \mathsf{E}\left[\log_2\ \det\left(\mathsf{I}_{n_{\mathrm{r}}} + \mathsf{SNR}_{\mathrm{sef}}\tilde{\mathbb{H}}_{\mathrm{s}}\tilde{\mathbb{H}}_{\mathrm{s}}^{\dagger}\right)\right] \quad (30)$$

where $\tilde{\mathbb{H}}_{\mathrm{s}}$ is the normalized channel estimate matrix $\tilde{\mathbb{H}}_{\mathrm{s}} \triangleq \frac{\hat{\mathbb{H}}_{\mathrm{s},k}}{1-\epsilon^2_{\mathrm{s}}}$ with zero-mean unit-variance entries, and

$$\mathsf{SNR}_{\mathrm{sef}} \triangleq \frac{\varrho_d \mathsf{SNR}(1-\epsilon^2_{\mathrm{s}})}{n_{\mathrm{t}} + n_{\mathrm{t}}\mathsf{SNR}\,\epsilon^2_{\mathrm{s}}}, \quad (31)$$

with the MSE ϵ^2_{s} equal to the RHS of (27), i.e,

$$\epsilon^2_{\mathrm{s}} = 1 - \int_{-1/2}^{1/2} \frac{\varrho_p \mathsf{SNR}\,|f_1(\lambda)|^2}{n_{\mathrm{t}}\varrho_p \mathsf{SNR}\, f_0(\lambda) + n_{\mathrm{t}}\varrho_d \mathsf{SNR} + n_{\mathrm{t}}} d\lambda. \quad (32)$$

The rate R_{s} can be further optimized over the values of ϱ_d and ϱ_p subject to the constraint (22).

Proof: See Appendix A-B. ∎

The main attribute of the SP scheme is full utilization of time instants for both data and pilot symbols. This ensures that the pre-factor of $\log_2(\cdot)$ in (30) is unity and the MIMO transmit dimension n_{t} is not restricted by the frequency of pilot emittance. At the same time, however, the resulting channel estimate is less reliable due to data and inter-antenna pilot interference—in addition to the receiver noise—as illustrated in Fig. 1 and from the MSE (32), i.e., equation (33), as shown at the top of the next page.

Although multiple transmit and receive antennas shall offer the benefit of multiplexing gain at a sufficiently high SNR (see [28], [36]), such a benefit does not seem to materialize for the SP rate R_{s} due to unfavorable scaling of $\mathsf{SNR}_{\mathrm{sef}}$ with transmit dimension n_{t}. By analyzing (31) and (32), we can see that increasing $n_{\mathrm{t}} \to \infty$ yields $n_{\mathrm{t}} + n_{\mathrm{t}}\mathsf{SNR}\epsilon^2_{\mathrm{s}} \to \infty$ and $1-\epsilon^2_{\mathrm{s}} \downarrow 0$, which in turn lead to $\mathsf{SNR}_{\mathrm{sef}} \downarrow 0$ and $R_{\mathrm{s}} \downarrow 0$. To verify that the MIMO multiplexing gain is in fact not attainable, we show in the following that R_{s} is bounded at high SNR. Since the partial derivatives of $\mathsf{SNR}_{\mathrm{sef}}$ with respect to ϱ_p and ϱ_p are both non-negative, $\mathsf{SNR}_{\mathrm{sef}}$ is a non-decreasing function of ϱ_p and ϱ_d. For a scheme with power constraint (22), namely $\varrho_p + \varrho_d = 1$, we can then upper-bound R_{s} by setting $\varrho_p = \varrho_d = 1$ to the RHS of (31) to yield

$$\begin{aligned} R_{\mathrm{s}} &\leq \mathsf{E}\left[\log_2 \det\left(\mathsf{I}_{n_{\mathrm{r}}} + \frac{\mathsf{SNR}\cdot(1-g_1)\cdot\tilde{\mathbb{H}}_{\mathrm{s}}\tilde{\mathbb{H}}_{\mathrm{s}}^{\dagger}}{n_{\mathrm{t}} + n_{\mathrm{t}}\mathsf{SNR}\cdot g_1}\right)\right] \quad (34)\\ &\triangleq \mathsf{E}\left[Q(\mathsf{SNR}, \tilde{\mathbb{H}}_{\mathrm{s}})\right]. \end{aligned}$$

where $g_1 = 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}|f_1(\lambda)|^2}{n_{\mathrm{t}}\mathsf{SNR}f_0(\lambda)+n_{\mathrm{t}}\mathsf{SNR}+n_{\mathrm{t}}} d\lambda$. For a given $\tilde{\mathbb{H}}_{\mathrm{s}} = \tilde{\mathsf{H}}_{\mathrm{s}}$, the function $\mathsf{SNR} \mapsto Q(\mathsf{SNR}, \tilde{\mathsf{H}}_{\mathrm{s}})$ is monotonously non-decreasing in SNR [36]. Therefore, applying the Monotone Convergence Theorem [37, Th. 1.26] yields

$$\begin{aligned} \lim_{\mathsf{SNR}\to\infty} R_{\mathrm{s}} &\leq \mathsf{E}\left[\lim_{\mathsf{SNR}\to\infty} Q(\mathsf{SNR}, \tilde{\mathbb{H}}_{\mathrm{s}})\right] \quad (35)\\ &= \mathsf{E}\left[\log_2\ \det\left(\mathsf{I}_{n_{\mathrm{r}}} + g_2\cdot\tilde{\mathbb{H}}_{\mathrm{s}}\tilde{\mathbb{H}}_{\mathrm{s}}^{\dagger}\right)\right] \quad (36) \end{aligned}$$

where $g_2 = \frac{\int_{-1/2}^{1/2} \frac{|f_1(\lambda)|^2}{n_{\mathrm{t}}f_0(\lambda)+n_{\mathrm{t}}} d\lambda}{n_{\mathrm{t}}\left(\int_{-1/2}^{1/2} \frac{n_{\mathrm{t}}[f_0(\lambda)|^2 - |f_1(\lambda)|^2 + n_{\mathrm{t}}f_0(\lambda)}{n_{\mathrm{t}}f_0(\lambda)+n_{\mathrm{t}}} d\lambda\right)}$ is SNR-independent, which implies that R_{s} is bounded and its multiplexing gain is strictly zero. This is valid irrespective of λ_D.

While the SP scheme seems unattractive at high SNR due to more noisy channel estimates, the absence of linear rate-loss in the expression (30) can be its desirable attribute at low SNR where any channel estimator will likely produce unreliable estimates. In terms of the single-gap linear estimator for stationary fading channels (25), the benefit of the SP scheme is further strengthen by the fact that the estimator has approximately $L/2$ times more observations than that of the OP scheme. Such a benefit can be referred to as the *observation gain*. We can deduce the significance of this gain to compensate more noisy observations, by comparing the mean-squared errors (MSEs) $\epsilon^2_{\mathrm{o},\ell}(r,t)$ and $\epsilon^2_{\mathrm{s},k}(r,t)$ for $n_{\mathrm{t}} = 1$

$$\epsilon_{\mathrm{s}}^2 = 1 - \int_{-1/2}^{1/2} \frac{\frac{\varrho_p}{n_{\mathrm{t}}}\mathsf{SNR}\,|f_1(\lambda)|^2\,d\lambda}{\frac{\varrho_p}{n_{\mathrm{t}}}\mathsf{SNR}\,f_0(\lambda) + \underbrace{(n_{\mathrm{t}}-1)\frac{\varrho_p}{n_{\mathrm{t}}}\mathsf{SNR}\,f_0(\lambda)}_{\text{inter-antenna pilot interference}} + \underbrace{\varrho_d\mathsf{SNR}+1}_{\substack{\text{data symbols}\\ \text{plus noise}}}} \tag{33}$$

and $2 \le L \le 1/2\lambda_D$, i.e.,

$$\epsilon_{\mathrm{o},\ell}^2(r,1) = 1 - \int_{-1/2}^{1/2} \frac{\rho_p\mathsf{SNR}[f_H(\lambda)]^2}{\rho_p\mathsf{SNR}\,f_H(\lambda)+L}d\lambda, \tag{37}$$

$$\epsilon_{\mathrm{s},k}^2(r,1) = 1 - \int_{-1/2}^{1/2} \frac{\varrho_p\mathsf{SNR}\,[f_H(\lambda)]^2}{\varrho_p\mathsf{SNR}\,f_H(\lambda)+2(\varrho_d\mathsf{SNR}+1)}d\lambda. \tag{38}$$

Assuming ϱ_p does not significantly differ from ρ_p, the denominators in the integrands of (37) and (38) indicate that the competitive advantage offered by the SP can be measured by the ratio

$$\eta = \frac{2\varrho_d\mathsf{SNR}+1}{L}. \tag{39}$$

If $\eta < 1$, then the SP provides a cleaner channel estimate than the OP and vice-versa. Therefore, we can clearly see justification of the superiority of the SP at low SNR for $L \ge 2$. Contradictorily, at high SNR with $\varrho_d > 0$, we have $\eta \gg 1$ and the OP definitely produces a better quality of channel estimates. The ratio η may also depend on λ_D when the OPAL is employed. By replacing L on the RHS of (39) with $L^* = \lfloor 1/(2\lambda_D)\rfloor$, we see that as $\lambda_D \downarrow 0$, η will eventually be less than one, which further implies that the SP can be superior at a sufficiently slow fading speed.

Note that aliasing for the SP only occurs at an extremely fast fading speed, i.e., $\lambda_D > 1/4$, due to the single-gap interpolator (25) that undersamples the fading process by a factor of two. Unlike the OP in Section III, however, aliasing or non-aliasing condition does not seem to drastically change the behavior of R_{s} at both high and low SNR. The R_{s} expression (30) is valid for any λ_D and the channel estimation error $\{\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\}_k$ remains a wide-sense stationary process, irrespective of whether aliasing occurs (see Appendix B).

V. Analysis and Results for Rectangular Fading PSD

In this section we give a specific analysis on the behavior of the rates R_{o} and R_{s} under a fading process with a rectangular PSD (*a.k.a.* ideal low-pass filter Doppler spectrum [9]), i.e.,

$$f_H(\lambda) = \begin{cases} \frac{1}{2\lambda_D}, & |\lambda| \le \lambda_D \\ 0, & \text{otherwise.} \end{cases} \tag{40}$$

This PSD simplifies numerical computations of R_{o} and R_{s} for any $\lambda_D \in [0, 1/2]$, which are based on the channel estimators in (8) and (25), respectively.

A. Optimizing R_{o} and R_{s} in the Case of no Aliasing

Recall from Sections III and IV that for the two pilot-aided schemes, the pilot and data power fractions are optimized subject to the constraints (2) and (3). In the case of no aliasing, the values of $(\rho_d,\ \rho_p)$ that maximize R_{o} and the values of $(\varrho_d,\ \varrho_p)$ that maximize R_{s} can be analytically determined as follows.

The OP scheme is aliasing-free whenever $L \le \frac{1}{2\lambda_D}$. In such a case, the MSE $\epsilon_{\mathrm{o},\ell}^2(r,t)$ in (11) and the effective $\mathsf{SNR}_{\mathrm{oef}}$ in (17) with $f_H(\lambda)$ in (40) can be expressed as

$$\epsilon_{\mathrm{o},\ell}^2(r,t) = \frac{2\lambda_D L}{\rho_p\mathsf{SNR}+2\lambda_D L}, \tag{41}$$

$$\mathsf{SNR}_{\mathrm{oef}} = \frac{\rho_d\mathsf{SNR}\cdot\frac{\rho_p\mathsf{SNR}}{\rho_p\mathsf{SNR}+2\lambda_D L}}{n_{\mathrm{t}}+\rho_d\mathsf{SNR}\,n_{\mathrm{t}}\cdot\frac{2\lambda_D L}{\rho_p\mathsf{SNR}+2\lambda_D L}}. \tag{42}$$

Based on (16), the values of $(\rho_d,\ \rho_p)$ satisfying the two constraints (6) and (7) that maximize $\mathsf{SNR}_{\mathrm{oef}}$ also maximize R_{o} due to the monotonicity of $\log_2\det(\cdot)$ function. Since the partial derivatives of $\mathsf{SNR}_{\mathrm{oef}}$ with respect to ρ_p and ρ_d are non-negative, $\mathsf{SNR}_{\mathrm{oef}}$ is a non-decreasing function of each individual ρ_p and ρ_d. Thus, in the case of the peakiness constraint (7) only without the average constraint (6), we have the optimal values of $\rho_p = \rho_{\mathrm{th}}$ and $\rho_d = n_{\mathrm{t}}$. Combining this result with the optimal power fractions under the average constraint (6) given in [9], [24], we obtain the optimal ρ_p and ρ_d satisfying both (6) and (7) as

$$\rho_p^* = \min\left\{\rho_{\mathrm{th}},\ \frac{L}{n_{\mathrm{t}}}\cdot\frac{\sqrt{1-\frac{\mathsf{SNR}\cdot(L-n_{\mathrm{t}}-2\lambda_D L n_{\mathrm{t}})}{(L-n_{\mathrm{t}})\cdot(\mathsf{SNR}+2\lambda_D n_{\mathrm{t}})}}}{1+\sqrt{1-\frac{\mathsf{SNR}\cdot(L-n_{\mathrm{t}}-2\lambda_D L n_{\mathrm{t}})}{(L-n_{\mathrm{t}})\cdot(\mathsf{SNR}+2\lambda_D n_{\mathrm{t}})}}}\right\}, \tag{43}$$

$$\rho_d^* = \min\left\{n_{\mathrm{t}},\ \frac{L}{L-n_{\mathrm{t}}}\cdot\frac{1}{1+\sqrt{1-\frac{\mathsf{SNR}\cdot(L-n_{\mathrm{t}}-2\lambda_D L n_{\mathrm{t}})}{(L-n_{\mathrm{t}})\cdot(\mathsf{SNR}+2\lambda_D n_{\mathrm{t}})}}}\right\}. \tag{44}$$

The SP scheme has a wider margin of λ_D that is free from aliasing than the OP scheme, i.e., $\lambda_D \le 1/4$. In this case, the integral in the MSE expression (27) for rectangular PSD (40) can be analytically evaluated to yield

$$\epsilon_{\mathrm{s}}^2 = \epsilon_{\mathrm{s},k}^2(r,t) = \frac{(n_{\mathrm{t}}-1)\varrho_p\mathsf{SNR}+4\lambda_D(n_{\mathrm{t}}\varrho_d\mathsf{SNR}+n_{\mathrm{t}})}{n_{\mathrm{t}}\varrho_p\mathsf{SNR}+4\lambda_D(n_{\mathrm{t}}\varrho_d\mathsf{SNR}+n_{\mathrm{t}})}, \tag{45}$$

$$\mathsf{SNR}_{\mathrm{sef}} = \frac{\varrho_d\mathsf{SNR}(1-\epsilon_{\mathrm{s}}^2)}{n_{\mathrm{t}}+n_{\mathrm{t}}\mathsf{SNR}\,\epsilon_{\mathrm{s}}^2}. \tag{46}$$

Similarly to the OP scheme, it suffices to find the values of ϱ_p and ϱ_d that maximize $\mathsf{SNR}_{\mathrm{ef}}$ in order to optimize R_{s}. By substituting $\varrho_p = 1-\varrho_d$ according to the constraint (22), and subsequently deriving $\mathsf{SNR}_{\mathrm{sef}}$ with respect to ϱ_d and equating to zero as a standard convex optimization approach, it can be shown that the optimal values of $(\varrho_d,\ \varrho_p)$ that

maximize (46) are given by

$$\varrho_d = \frac{1}{\sqrt{1+\frac{4\lambda_D n_{\mathrm{t}}{}^2\mathsf{SNR}^2+4\lambda_D n_{\mathrm{t}}{}^2\mathsf{SNR}-n_{\mathrm{t}}{}^2\mathsf{SNR}-n_{\mathrm{t}}\mathsf{SNR}^2(n_{\mathrm{t}}-1)}{4\lambda_D n_{\mathrm{t}}{}^2+4\lambda_D n_{\mathrm{t}}{}^2\mathsf{SNR}+n_{\mathrm{t}}{}^2\mathsf{SNR}+n_{\mathrm{t}}\mathsf{SNR}^2(n_{\mathrm{t}}-1)}}+1}, \tag{47}$$

$$\varrho_p = 1-\varrho_d. \tag{48}$$

At high SNR, the optimal ϱ_d and ϱ_p generally depend on both n_{t} and λ_D. Interestingly, as the SNR tends to zero, the transmit power has to be equally divided for both data and pilot symbols, i.e., $\varrho_d = \varrho_p = 1/2$ for optimal performance.

Remark that from the practical perspectives, the Doppler bandwidth λ_D in the PSD (40) can be associated with the speed of the mobile device, carrier frequency (f_c) and the coherence bandwidth of the channel (W_c) [22]. For a given f_c and W_c, λ_D is directly proportional to the mobile speed. In the following comparisons, we thus associate slow-fading with slow mobility (static to typical walking speeds) and fast-fading with fast mobility (typical bullet train speeds).

B. Behaviors of Channel Estimation Errors and Rates

In the following we have a closer look on the channel estimation errors and rates for the OP and SP schemes. We show the non-trivial dependency of those performance metrics with the transmit dimension and highlight that the existing trend drawn from existing works [3], [12], i.e., fast- (slow-) fading superiority of the SP (OP), is only a special case of this work. Note from Section III-B that two possible types of OP can be designed based on specification of the training interval L. The first one, which is common in the literature, is the OPAL where L is adapted according to $L = L^* = \lfloor 1/(2\lambda_D)\rfloor$ in order to avoid spectrum aliasing. The second one, which is more desirable in some wireless standards, is the OPFL where L is fixed to a constant and intended to operate under a diverse range of λ_D. For example, in [31], [32], L is chosen to be $L = 7$ that guarantees no aliasing for the OPFL up to $\lambda_D = 0.07$. This approximates to the vehicle speed of 150 km/h when the delay spread is 20 μs and $f_c = 5$ GHz, following from the calculation in [22].

Based on the aliasing-free MSE expressions $\epsilon^2_{\mathrm{o},\ell}(r,t)$ in (41) and $\epsilon^2_{\mathrm{s},k}(r,t)$ in (45), we can partly identify the range of λ_D under which the OP and SP are superior to each other. Specifically, for fixed n_{t}, SNR and pilot/data power fractions, we have the following.

- *MSE comparison between the OPAL and SP:* Since it holds for the OPAL that $c_0 = \frac{1}{2} \le 2\lambda_D L \le c_1 = 1$, by replacing $2\lambda_D L$ with c_0 in $\epsilon^2_{\mathrm{o},\ell}(r,t)$ of (41) and comparing it with $\epsilon^2_{\mathrm{s},k}(r,t)$, we have that the SP is always superior for

$$\lambda_D \le \frac{\varrho_p(c_0-[n_{\mathrm{t}}-1]\rho_p\mathsf{SNR})}{4\rho_p(n_{\mathrm{t}}\varrho_d\mathsf{SNR}+n_{\mathrm{t}})}, \quad \lambda_D \ge 0, \tag{49}$$

and the OPAL is always superior for

$$\lambda_D \ge \frac{\varrho_p(c_1-[n_{\mathrm{t}}-1]\rho_p\mathsf{SNR})}{4\rho_p(n_{\mathrm{t}}\varrho_d\mathsf{SNR}+n_{\mathrm{t}})}, \quad \lambda_D \ge 0. \tag{50}$$

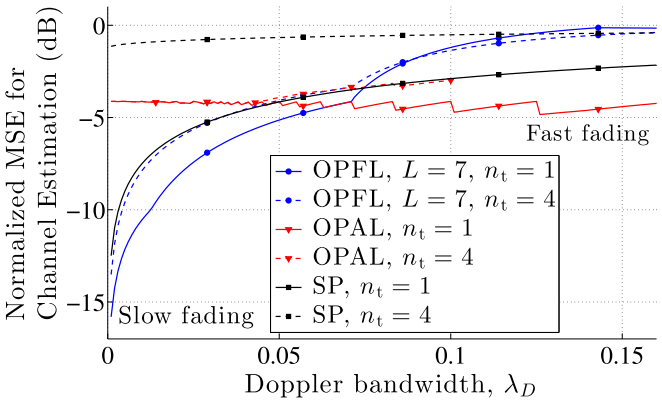

Fig. 2. Normalized MSE against the Doppler bandwidth λ_D for SNR = 1 dB and peakiness parameter $\rho_{\mathrm{th}} = 1$ dB.

- *MSE comparison between the OPFL and SP:* The region of SP's superiority below $\lambda_D \le \frac{1}{2L}$ is given by

$$\lambda_D > \frac{(n_{\mathrm{t}}-1)\rho_p\varrho_p\mathsf{SNR}}{2L\varrho_p-4n_{\mathrm{t}}\rho_p\varrho_d\mathsf{SNR}-4n_{\mathrm{t}}\rho_p}, \quad \text{if} \quad L > \frac{2n_{\mathrm{t}}\rho_p\varrho_d\mathsf{SNR}+2n_{\mathrm{t}}\rho_p}{\varrho_p}. \tag{51}$$

  Otherwise, the SP's superiority may only occur in the aliasing regime where $\lambda_D > \frac{1}{2L}$.

The interval (49) shows that in comparison to the OPAL, the SP tends to be superior at a small value of λ_D. On the other hand, the range (51) indicates that in comparison to the OPFL, the SP may be superior at a typically large value of λ_D (potentially in the aliasing region).

If the pilot and data powers are optimized with respect to λ_D (see, e.g., equations (43) and (48)), then solving the MSE superiority regimes of the OP and SP generally involves solving the parametric polynomial inequalities of degrees eight (when ρ_{th} in (43) does not dominate) and three (when ρ_{th} dominates) where the polynomial parameters depend on n_{t} and SNR. In this case, a complete analytical characterization of those superiority regimes is challenging as it is widely-known from Abel-Ruffini theorem [38] that for polynomials of degrees five and above, there exists no algebraic solution of the roots in terms of polynomial parameters.

In order to understand the behaviors of the MSE when the pilot and data powers are optimized, we invoke numerical evaluation as illustrated in Fig. 2. Herein the normalized MSE is defined by $\epsilon^2_{\mathrm{o},*}/n_{\mathrm{t}}$ where $\epsilon^2_{\mathrm{o},*}$ is given by (15) for the OP and by $n_{\mathrm{t}}{}^{-1}\sum_{t=1}^{n_{\mathrm{t}}}\epsilon^2_{\mathrm{s}}(r,t)$ where $\epsilon^2_{\mathrm{s}}(r,t)$ is given by (27) for the SP. We consider channels with $n_{\mathrm{r}} = 4$ receive antennas and two different transmit antennas, which we refer to the first as the MIMO channel ($n_{\mathrm{t}} = 4$) and the second as the SIMO channel ($n_{\mathrm{t}} = 1$).

For the MIMO case at sufficiently fast fading we observe that, in terms of the normalized MSE, the SP is competitive with the OPFL but inferior to the OPAL (for only up to $\lambda_D = 0.1$). In this regime, the OPFL looses its channel tracking capability due to insufficient fading samples (i.e., aliasing for $L > \frac{1}{2\lambda_D}$). On the other hand, the OPAL maintains its superiority to the SP due to the capability of retaining

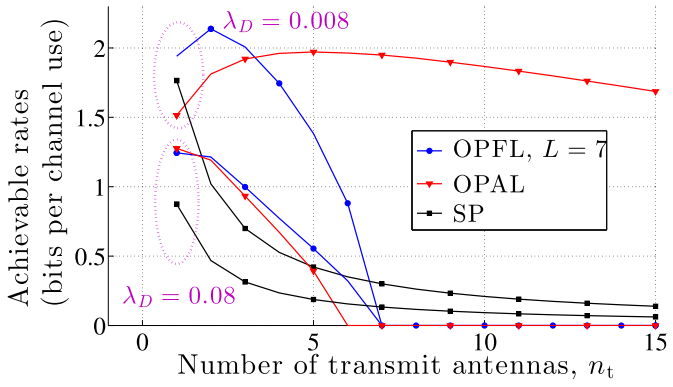

Fig. 3. Achievable rates versus n_t for SNR $= 1$ dB, the peakiness parameter $\rho_{th} = 1$ dB and $n_r = 4$.

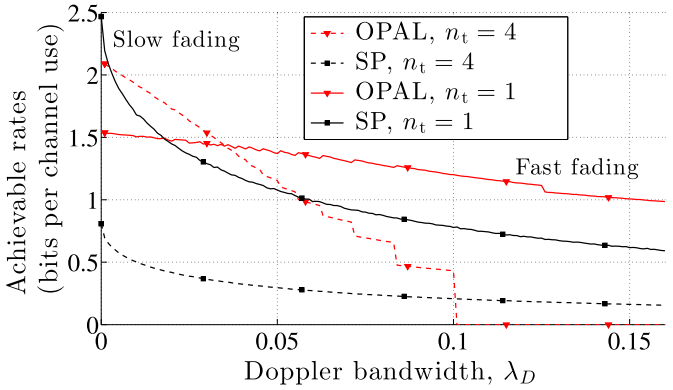

Fig. 4. Achievable rates versus the Doppler bandwidth λ_D for SNR $= 1$ dB and the peakiness parameter $\rho_{th} = 1$ dB. The number of receive antennas is set to be $n_r = 4$.

sufficient fading samples up to $\lambda_D = \frac{1}{2(n_t+1)} = 0.1$. Above $\lambda_D \geq \frac{1}{2(n_t+1)}$, the OPAL is inoperable because the MIMO dimension is too large to estimate, i.e., when $n_t > L^* = \lfloor 1/2\lambda_D \rfloor$. The SP can thus offer a wider operation range than and superior performance to the OPAL in the fast fading regime. In the slow-fading regime, the SP appears to be inferior to both OPFL and OPAL due to the dominant inter-antenna pilot interference (33) that scales up with an increase of n_t. This is confirmed from the MSE in (45). I.e., since the function $\frac{a+x}{b+x}$ is monotonously non-decreasing in x for $a \leq b$, the MSE for the SP can in fact be lower-bounded as

$$\epsilon^2_{s,k}(r,t) = \frac{(n_t-1)\varrho_p \mathsf{SNR} + 4\lambda_D(n_t\varrho_d \mathsf{SNR} + n_t)}{n_t\varrho_p \mathsf{SNR} + 4\lambda_D(n_t\varrho_d \mathsf{SNR} + n_t)} \geq \frac{n_t - 1}{n_t}, \quad (52)$$

which is bounded away from zero for $n_t > 1$.

In the SIMO case ($n_t = 1$), the required dimension to estimate for the OP and the inter-pilot interference for the SP are both minimized. While the MSE trend at fast fading exhibit similar behavior to that in the MIMO case, the slow-fading behavior is unalike. The SP can be superior to the OPAL because the latter maintains just enough fading samples satisfying no aliasing criterion, which linearly scale with λ_D. For the OPAL, this results in a nearly-invariant MSE (41)—with respect to λ_D—with some minor variation due to the product of $2\lambda_D L^* = 2\lambda_D \lfloor 1/2\lambda_D \rfloor$, which strictly lies in the interval of $[1/2, 1]$. On the other hand, the SP is inferior to the OPFL in terms of the MSE. As $\lambda_D \downarrow 0$, the number of fading samples are kept the same for the OPFL, but the channel varies slower. The MSE for the OPFL thus improves with a decrease in λ_D. The absence of inter-pilot and data interference further justifies the superiority of the OPFL.

The advantages and limitations of the OP and SP in terms of achievable rates when varying the MIMO transmit dimension are illustrated in Fig. 3. The transmit dimensions n_t of both OPAL and OPFL are strictly constrained by the length of the training period L to yield positive rates. On the other hand, the full utilization of time slots for data transmission ensures that the SP can maintain a positive rate even when n_t is very large. Surprisingly, due to an improved quality of channel estimates at a slower fading speed, the SP can be superior to the OPAL at a sufficiently small n_t (see the curves for $\lambda_D = 0.008$).

Results in Figs. 2 and 3 underline the complex roles of the MIMO dimension, frequency of pilot emittance and fading speed in determining the performance of the OP and SP schemes, which in turn reveal insights that were minimally captured in the previous works such as [12], [13], [17], [18]. The premise of a superior quality of estimates in the OP does not always hold across different fading speeds and, depending on the n_t and frequency of pilot emittance, the SP can produce more reliable estimates. In the context of stationary fading channels, the SP is benefited *not only* from the full utilization of time instants for data transmission, *but also* from the observation gain captured by (39). This latter SP's strong point is particularly instrumental at slow fading speed to demonstrate its superiority to the OPAL. The substantial dimension cost of the OP can be unfavorable, especially when n_t is close to the value of the training interval.

From Figs. 2 and 3, the OPAL and OPFL have different behaviors of the MSE, which naturally affect the behaviors of the achievable rate R_o. Therefore, in order to gain more insights on their performance over SNR and λ_D, we separately compare the SP with the OPAL and with the OPFL in terms of achievable rates in the following subsections.

C. Rate Comparisons of SP and OPAL

In terms of achievable rates, we exemplify in the following that the SP can be competitive with (or, in some cases, superior to) the OPAL, not only at fast-fading as reported in [3], [12], but also at slow-fading. This is possible since in the stationary fading channels, the SP is benefited not only from full utilization of time for data transmission, but also the observation gain, cf. Section IV-B.

In Fig. 4, we plot the achievable rates against the Doppler bandwidth λ_D for both the MIMO ($n_t = 4$) and SIMO ($n_t = 1$) cases. The OPAL avoids aliasing with the adaptive $L^* = \lfloor 1/(2\lambda_D) \rfloor$ at the expense of having a Doppler-limited transmit dimension $n_t < L^*$. For a given n_t, the OPAL can only accommodate both data transmission and channel estimation up to $\lambda_D = \frac{1}{2(n_t+1)}$.

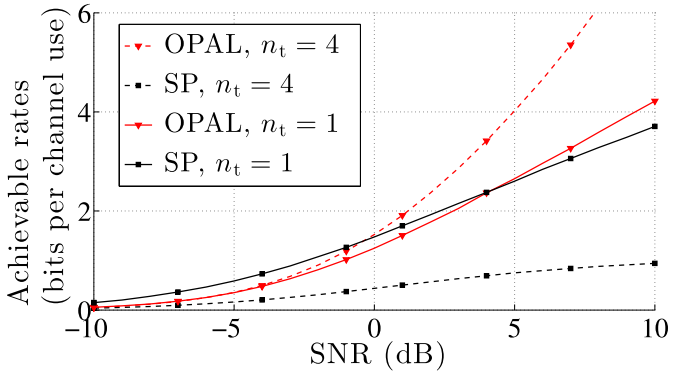

Fig. 5. Achievable rates versus the SNR with the peakiness parameter $\rho_{\text{th}} = 1$ dB and Doppler bandwidth $\lambda_D = 0.01$. The number of receive antennas is set to be $n_r = 4$.

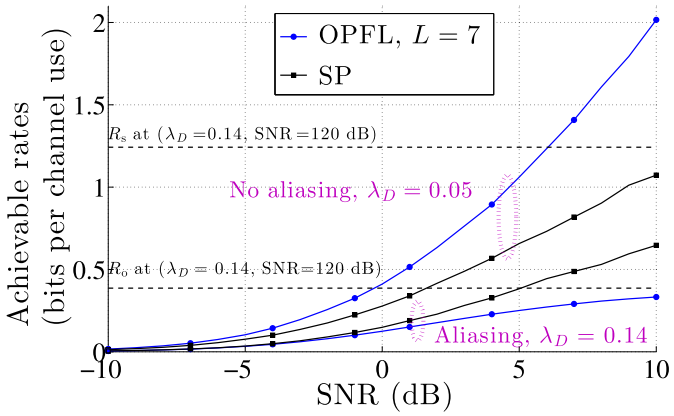

Fig. 6. Achievable rates for a SISO channel ($n_t = n_r = 1$) versus the SNR with the peakiness parameter $\rho_{\text{th}} = 1$ dB. For $\lambda_D = 0.14$, R_o and R_s at 120 dB approximate upper-bounding quantities at high SNR.

For the MIMO case, the limitation of the OPAL in the fast fading regime is clearly demonstrated by an abrupt transition around $\lambda_D = 0.1$ in Fig. 4, where its rate goes to zero. In this case, the SP is preferred due to full utilization of time for data transmission irrespective of the fading speed. The trend at slow fading is the exact opposite. In this regime, we have $n_t \ll L^* = \lfloor 1/(2\lambda_D) \rfloor$ and consequently, the more reliable estimates enable the OPAL to reap the rate gains due to multiple antennas [36] and significantly outperform the SP.

When n_t is reduced to one, i.e., the SIMO case, the operable fading speed for the OPAL extends up to $\lambda_D = \frac{1}{4}$. This implies the fast-fading superiority of the SP may only occur at an extremely fast fading variation, i.e., $\lambda_D > 0.25$. Surprisingly, the advantage of the SP over the OPAL is more instrumental at slow fading as a consequence of a better MSE as shown in Fig. 2 as well as no timing loss in terms of data transmission. Herein the better MSE is due to the observation gain of the SP. More specifically, the ratio η (which is inversely proportional to the observation gain) in (39) tends to zero as $\lambda_D \downarrow 0$ when $L^* = \lfloor 1/(2\lambda_D) \rfloor$.

Given $n_t < L^*$ for the OPAL, the trend over different values of SNR is in agreement to the widely-known signal-processing results (see [3], [12]) as demonstrated in Fig. 5. The OPAL is superior to the SP at high SNR due to the logarithmic growth of its rate, which is further amplified in the case of $n_t = 4$ due to a better MIMO multiplexing gain, cf. (19), as compared to the bounded rate, cf. (36). At low SNR, the SP is competitive with the OPAL because of the observation gain and effective time-utilization for sending data. The regime of superiority for the SP is wider for $n_t = 1$ as explained by the fact that the MSE for the SP decreases with decreasing n_t, cf. Fig. 2 and equation (52).

D. Rate Comparisons of SP and OPFL

For the purpose of comparing the SP and OPFL, we note that the OPFL can suffer from aliasing if the channel varies faster than the frequency of pilot emittance. In contrast to the well-known high-SNR superiority of the OP [1], [17], [18], we show in the following example where the SP can outperform the OPFL at high SNR where the latter experiences aliasing.

In Fig. 6, we demonstrate the achievable rates against SNR for two different values of λ_D, where one corresponds to non-aliasing ($\lambda_D = 0.05$) and the other corresponds to aliasing ($\lambda_D = 0.14$). If the OPFL does not suffer from aliasing, then the widely-known trend applies, i.e., the OPFL mostly outperforms the SP across a wide range of SNR with the SP being reasonably competitive at very small SNR values. If the OPFL suffers from aliasing, then both the OPFL and SP rates are bounded from above at high SNR, which implies that the high-SNR trend depends on their asymptotic values. The exact opposite trend (i.e., the SP outperforms the OPFL at high SNR) typically occurs for: 1) a smaller value of n_t (1 or 2) due to a significantly smaller MSE for the SP, cf. equation (52); and 2) $n_t > L$ as the MIMO dimension is too large to estimate. This opposite trend is clearly observed in Fig. 6 for the case of $n_t = 1$ and $\lambda_D = 0.14$.

VI. Conclusion

We have studied the performance of orthogonal pilot (OP)- and superimposed pilot (SP)-aided channel estimation schemes in stationary bandlimited fading channels. The analyses have revealed new insights on the complex interplay among the achievable rate, MIMO dimension, frequency of pilot emittance and SNR, which further determine the regimes of superiority for each scheme. The desirable behavior of the OP heavily relies on aliasing-free condition in order to achieve the high-SNR logarithmic growth of the rate, which can be significantly enhanced by the effective MIMO multiplexing. Alongside this advantage, there are two caveats: 1) restriction of the MIMO transmit dimension by the frequency of pilot emittance (training interval); 2) challenge of maintaining sufficient fading samples at fast fading. On the other hand, at the expense of more noisy fading samples due to data and inter-antenna pilot interference, the SP is benefited from two attributes: 1) efficient time-utilization for data transmission (the absence of linear rate-loss); 2) observation gain. Numerical results have demonstrated that these desirable attributes of the SP can be dominant in slow-fading SIMO channels, fast-fading MIMO

channels and across a wide range of the SNR when the OP suffers from spectrum aliasing.

The two schemes considered in this work represent two contrasting cases of pilot placements in the light of pilot-aided channel estimation. The new insights from the comparisons shall provide a direction to the design of hybrid schemes for stationary MIMO fading channels that take into account the superiority regimes of the OP and SP.

APPENDIX A
ACHIEVABLE RATES

A framework of computing an achievable rate in stationary fading channels has been outlined in [33], [34] using the GMI for mismatched decoders with *genie-aided* imperfect CSI. Leveraging upon the same framework, we replace the genie with specific channel estimators of the OP and SP, cf. (8) and (25), and evaluate the corresponding achievable rates.

A. Orthogonal Pilots

For $L \leq \frac{1}{2\lambda_D}$, the rate (14) has been obtained in [24], [27] using mutual information and can be similarly obtained using the GMI [33], [34]. The condition $L \leq \frac{1}{2\lambda_D}$ is instrumental in guaranteeing the wide-sense stationarity of the channel estimation error.

For any $L \geq 1$, however, the channel and its estimate are generally not stationary, but jointly cyclostationary with a cyclic period of L. Therefore, instead of using the GMI in [33], [34] directly, we apply a similar derivation by noting that the channel and its estimate are blockwise ergodic. We can then obtain the GMI (in nats/channel use) for any $L > 0$ as

$$I^{\text{gmi}} \triangleq \sup_{\theta<0} \frac{1}{L} \sum_{\ell=n_{\text{t}}}^{L-1} [\kappa_\ell(\theta) - \theta T_\ell] \tag{53}$$

where

$$\begin{aligned}\kappa_\ell(\theta) &= \mathsf{E}\left[\log\,\det\left(\mathsf{I}_{n_{\text{r}}} - \theta\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}\right)\right] \\ &\quad - \theta\mathsf{E}\left[\boldsymbol{Y}^{\dagger}\cdot\left(\mathsf{I}_{n_{\text{r}}} - \theta\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}\right)^{-1}\cdot\boldsymbol{Y}\right] \end{aligned} \tag{54}$$

$$\begin{aligned}&= \mathsf{E}\left[\log\,\det\left(\mathsf{I}_{n_{\text{r}}} - \theta\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}\right)\right] \\ &\quad - \theta\left(1+\tfrac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\textstyle\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t)\right) \\ &\quad \times \text{tr}\left\{\mathsf{E}\left[\left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t)}\right)\right.\right. \\ &\quad \left.\left. \times \left(\mathsf{I}_{n_{\text{r}}} - \theta\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}\right)^{-1}\right]\right\}, \end{aligned} \tag{55}$$

$$T_\ell = \mathsf{E}\left[\left\|\sqrt{\mathsf{SNR}}\left(\mathbb{H}_\ell - \hat{\mathbb{H}}_{\text{o},\ell}\right)\boldsymbol{X}_\ell + \boldsymbol{Z}_\ell\right\|^2\right] \tag{56}$$

$$= n_{\text{r}} + n_{\text{r}}\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\textstyle\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t). \tag{57}$$

Herein we have recalled $\hat{\mathbb{H}}_{\text{o},\ell}$ as a channel estimate matrix whose (r,t)-th entry is distributed as $\mathcal{N}_{\mathbb{C}}(0,[1-\epsilon_{\text{o},\ell}^2(r,t)])$ where $\epsilon_{\text{o},\ell}^2(r,t)$ has been given in (11).

Finding the exact optimal θ for (53) is analytically challenging, particularly due to the sum over ℓ on the RHS of (53). Therefore, in order to obtain a reasonable closed-form expression, we will substitute the optimal θ with a good choice of θ similarly to [39, App. D], i.e.,

$$\theta = \frac{-1}{1+\frac{\rho_d \mathsf{SNR}}{n_{\text{t}}}\epsilon_{\text{o},*}^2}, \quad \epsilon_{\text{o},*}^2 \triangleq \max_{\ell\in\{n_{\text{t}},\ldots,L-1\}}\textstyle\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t). \tag{58}$$

We next evaluate the trace on the RHS of (55) for the value of θ in (58), i.e.,

$$\begin{aligned}\text{tr}\left\{\mathsf{E}\left[\left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t)}\right)\right.\right. \\ \left.\left. \times \left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\epsilon_{\text{o},*}^2}\right)^{-1}\right]\right\}. \end{aligned} \tag{59}$$

Since the two matrices

$$\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t)} \quad \text{and} \quad \mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\epsilon_{\text{o},*}^2}$$

are both Hermitian positive definite matrices, it can be shown using the definition of trace operator and singular value decomposition that

$$\begin{aligned}&\text{tr}\left\{\mathsf{E}\left[\left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\sum_{t=1}^{n_{\text{t}}}\epsilon_{\text{o},\ell}^2(1,t)}\right)\right.\right. \\ &\qquad \left.\left. \times \left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\epsilon_{\text{o},*}^2}\right)^{-1}\right]\right\} \\ &\geq \text{tr}\left\{\mathsf{E}\left[\left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\epsilon_{\text{o},*}^2}\right)\right.\right. \\ &\qquad \left.\left. \times \left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}\,\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}}{n_{\text{t}} + \rho_d \mathsf{SNR}\epsilon_{\text{o},*}^2}\right)^{-1}\right]\right\} = \text{tr}\left\{\mathsf{I}_{n_{\text{r}}}\right\} = n_{\text{r}}. \end{aligned} \tag{60}$$

Inserting the value of θ in (58) to the RHS of (55) and (57), combining the results with (53) and applying the inequality (60) yield a rate (in bits/channel use)

$$\begin{aligned}R_{\text{o}} &= \frac{1}{L}\sum_{\ell=n_{\text{t}}}^{L-1}\mathsf{E}\left[\log_2\,\det\left(\mathsf{I}_{n_{\text{r}}} + \frac{\rho_d \mathsf{SNR}}{n_{\text{t}} + \rho_d \mathsf{SNR}\,\epsilon_{\text{o},*}^2}\hat{\mathbb{H}}_{\text{o},\ell}\hat{\mathbb{H}}_{\text{o},\ell}^{\dagger}\right)\right] \\ &\leq I^{\text{gmi}}. \end{aligned} \tag{61}$$

Since any rate $R < I^{\text{gmi}}$ is achievable, the result (61) implies that R_{o} is a valid achievable rate.

B. Superimposed Pilots

Building upon the same framework as in [33], [34], we first consider GMI evaluation for any general time-invariant channel estimator (23), which is not necessarily the single-gap interpolator (25). Some basic setups from the channel model, input symbols, noise and channel estimator directly satisfy [34, Assumps. 1–3, 5] and [33, Assumps. 1–4, 6, 7], i.e.,

- Random coding with i.i.d. Gaussian inputs $\mathcal{N}_{\mathbb{C}^{n_\mathrm{t}}}(\mathbf{0}, \frac{\varrho_d}{n_\mathrm{t}}\mathsf{I}_{n_\mathrm{t}})$.
- Ergodic fading and noise processes. The fading, noise and input sequences are independent.
- The noise has zero mean and identity covariance matrix.
- Equation (23) implies that the channel estimate $\hat{\mathbb{H}}_{\mathrm{s},k}$ is a time-invariant function of $\{(\mathbb{H}_k, \boldsymbol{Z}_k, \boldsymbol{X}_k)\}$, which further implies the convergence of the decoder metric (29), namely

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \left\| \boldsymbol{y}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{\mathrm{s},k}\bar{\boldsymbol{x}}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{\mathrm{s},k}\boldsymbol{p}\right\|^2 = \mathsf{E}\left[\left\|\sqrt{\mathsf{SNR}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)(\bar{\boldsymbol{X}}_k + \boldsymbol{p}) + \boldsymbol{Z}_k\right\|^2\right] \quad \text{a.s.} \tag{62}$$

These key assumptions enable us to arrive at the following GMI expression for the SP (in nats/channel use)

$$I^{\mathrm{gmi}} = \sup_{\theta<0}\left[\kappa(\theta) - \theta T\right] \tag{63}$$

where the scaled cumulant moment-generating function (MGF) of the decoding metric (29) associated with incorrect codewords is given by

$$\begin{aligned}\kappa(\theta) &= \mathsf{E}\left[\log\,\det\left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)\right] \\ &\quad - \theta\,\mathsf{E}\Big[\mathrm{tr}\Big\{\left(\sqrt{\mathsf{SNR}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)\boldsymbol{p} + \sqrt{\mathsf{SNR}}\,\mathbb{H}_k\bar{\boldsymbol{X}}_k + \boldsymbol{Z}_k\right) \\ &\quad \times \left(\sqrt{\mathsf{SNR}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)\boldsymbol{p} + \sqrt{\mathsf{SNR}}\,\mathbb{H}_k\bar{\boldsymbol{X}}_k + \boldsymbol{Z}_k\right)^{\dagger} \\ &\quad \times \left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)^{-1}\Big\}\Big]\end{aligned} \tag{64}$$

and the parameter T is given by

$$T = \mathsf{E}\left[\left\|\sqrt{\mathsf{SNR}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)(\bar{\boldsymbol{X}}_k + \boldsymbol{p}) + \boldsymbol{Z}_k\right\|^2\right] \tag{65}$$

$$\begin{aligned} &= \mathrm{tr}\Big\{\mathsf{E}\left[\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)^{\dagger}\right] \\ &\qquad + \mathsf{E}\Big[\sqrt{\mathsf{SNR}}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)\bar{\boldsymbol{X}}_k \\ &\qquad \times \left(\sqrt{\mathsf{SNR}}\left[\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right]\boldsymbol{p} + \boldsymbol{Z}_k\right)^{\dagger}\Big] \\ &\qquad + \mathsf{E}\Big[\left(\sqrt{\mathsf{SNR}}\left[\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right]\boldsymbol{p} + \boldsymbol{Z}_k\right) \\ &\qquad \times \sqrt{\mathsf{SNR}}\,\bar{\boldsymbol{X}}_k^{\dagger}\left(\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right)^{\dagger}\Big] \\ &\qquad + \mathsf{E}\Big[\left(\sqrt{\mathsf{SNR}}\left[\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right]\boldsymbol{p} + \boldsymbol{Z}_k\right) \\ &\qquad \times \left(\sqrt{\mathsf{SNR}}\left[\mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}\right]\boldsymbol{p} + \boldsymbol{Z}_k\right)^{\dagger}\Big]\Big\}. \end{aligned} \tag{66}$$

Simplifying the expressions of (64) and (66) for the channel estimator (23) is generally intractable due to the estimate $\hat{\mathbb{H}}_{\mathrm{s},k}$ that depends on $\bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k$. In such an estimate, the lack of closed-form expressions of correlation of $(\hat{\mathbb{H}}_{\mathrm{s},k}, \bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k)$ and correlation of any pair from $(\hat{\mathbb{H}}_{\mathrm{s},k}, \bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k)$ conditioned on the rest prohibits further evaluation of (64) and (66).

We circumvent this problem by considering a suboptimal estimator, namely the single-gap interpolator (26) that decorrelates the time-k estimate $\hat{\mathbb{H}}_{\mathrm{s},k}$ from the time-k data and noise $\bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k$. Let $\mathbb{E}_{\mathrm{s},k} \triangleq \mathbb{H}_k - \hat{\mathbb{H}}_{\mathrm{s},k}$ be the time-k estimation error matrix, which is uncorrelated with $\hat{\mathbb{H}}_k$ from the orthogonality principle (see Appendix B). The estimator (26) ensures that the correlation of the triplet $(\hat{\mathbb{H}}_{\mathrm{s},k}, \bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k)$ and correlation of any pair from $(\hat{\mathbb{H}}_{\mathrm{s},k}, \bar{\boldsymbol{X}}_k, \boldsymbol{Z}_k)$ conditioned on the rest are all zero. We can thus simplify (64) and (66), and obtain

$$\begin{aligned}\kappa(\theta) &= -\theta\,\mathsf{E}\Big[\mathrm{tr}\Big\{\Big(\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\mathbb{H}_k\mathbb{H}_k^{\dagger} + \left[\sqrt{\mathsf{SNR}}\,\hat{\mathbb{E}}_{\mathrm{s},k}\boldsymbol{p} + \boldsymbol{Z}_k\right] \\ &\quad \times \left[\sqrt{\mathsf{SNR}}\,\hat{\mathbb{E}}_{\mathrm{s},k}\boldsymbol{p} + \boldsymbol{Z}_k\right]^{\dagger}\Big) \\ &\quad \times \left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)^{-1}\Big\}\Big] \\ &\quad + \mathsf{E}\left[\log\,\det\left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)\right]\end{aligned} \tag{67}$$

$$\begin{aligned} &= -\theta\,\mathsf{E}\Big[\mathrm{tr}\Big\{\Big(\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\mathbb{H}_k\mathbb{H}_k^{\dagger} + \sqrt{\mathsf{SNR}}\,\hat{\mathbb{E}}_{\mathrm{s},k}\,\boldsymbol{p}\boldsymbol{p}^{\dagger}\,\hat{\mathbb{E}}_{\mathrm{s},k}^{\dagger} \\ &\quad + \boldsymbol{Z}_k\boldsymbol{Z}_k^{\dagger}\Big) \times \left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)^{-1}\Big\}\Big] \\ &\quad + \mathsf{E}\left[\log\,\det\left(\mathsf{I}_{n_\mathrm{r}} - \theta\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\,\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)\right] \end{aligned} \tag{68}$$

and

$$\begin{aligned} T &= \mathsf{E}\Big[\mathrm{tr}\Big\{\frac{\varrho_d\mathsf{SNR}}{n_\mathrm{t}}\mathbb{E}_{\mathrm{s},k}\mathbb{E}_{\mathrm{s},k}^{\dagger} + \left(\sqrt{\mathsf{SNR}}\,\mathbb{E}_{\mathrm{s},k}\boldsymbol{p} + \boldsymbol{Z}_k\right) \\ &\quad \times \left(\sqrt{\mathsf{SNR}}\,\mathbb{E}_{\mathrm{s},k}\boldsymbol{p} + \boldsymbol{Z}_k\right)^{\dagger}\Big\}\Big] \end{aligned} \tag{69}$$

$$= n_\mathrm{r} + \frac{\mathsf{SNR}(\varrho_p + \varrho_d)}{n_\mathrm{t}}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\epsilon_{s,k}^2(r,t) \tag{70}$$

$$= n_\mathrm{r} + n_\mathrm{r}\mathsf{SNR}(\varrho_p + \varrho_d)\epsilon_\mathrm{s}^2 \tag{71}$$

where $\epsilon_{s,k}^2(r,t) = \epsilon_\mathrm{s}^2$ has been given in (27).

Remark that (68) and (71) are similar to the scaled cumulant MGF and parameter T in [34] when the time-independent decoder-weighting matrix of $\sqrt{-\theta}\,\mathsf{I}_{n_\mathrm{r}}$, $\theta < 0$, input with covariance matrix $\frac{\varrho_d}{n_\mathrm{t}}\mathsf{I}_{n_\mathrm{t}}$ and "effective" noise vector $(\sqrt{\mathsf{SNR}}\,\mathbb{E}_{\mathrm{s},k}\boldsymbol{p} + \boldsymbol{Z}_k)$ are applied. It thus can be shown by following [34, App. B] that the optimal θ maximizing the GMI (63) is given by

$$\theta = -\frac{1}{1 + (\varrho_p + \varrho_d)\mathsf{SNR}\,\epsilon_\mathrm{s}^2} = -\frac{1}{1 + \mathsf{SNR}\,\epsilon_\mathrm{s}^2} \tag{72}$$

where the last equality is from the power constraint $\varrho_p + \varrho_d = 1$ in (2). Inserting the optimal θ to the RHS of (68), and then combining the result and (71) with (63)

yield (in bits/channel use)

$$R_{\mathrm{s}} = I^{\mathrm{gmi}} = \mathsf{E}\left[\log_2 \det\left(\mathsf{I}_{n_{\mathrm{r}}} + \frac{\varrho_d \mathsf{SNR}}{n_{\mathrm{t}} + n_{\mathrm{t}}\mathsf{SNR}\,\epsilon_{\mathrm{s}}^2}\hat{\mathbb{H}}_{\mathrm{s},k}\hat{\mathbb{H}}_{\mathrm{s},k}^{\dagger}\right)\right]. \tag{73}$$

In Proposition 2 the index k has been removed due to the stationarity assumption.

Appendix B
LMMSE of the Single-Gap Interpolator for Superimposed Pilots

Due to the symmetry of the fading processes among all the transmit-receive antenna pairs, it suffices to consider fading estimation for the antenna pair $(r, t) = (1, 1)$. Let $V_k = Y_k(1)$ be the observation at receive antenna 1 at time k. We have from the channel model (1) that

$$\begin{aligned} V_k &= Y_k(1) \\ &= \sqrt{\frac{\varrho_p}{n_{\mathrm{t}}}}\sum_{t=1}^{n_{\mathrm{t}}} H_k(1,t) + \sqrt{\frac{\varrho_d}{n_{\mathrm{t}}}}\sum_{t=1}^{n_{\mathrm{t}}} H_k(1,t)x_k(t) + Z_k. \end{aligned} \tag{74}$$

The fading estimate for transmit-receive antenna pair $(r, t) = (1, 1)$ at time k can be written as

$$\hat{H}_{s,k}(1,1) = \sum_{k'=-\infty}^{\infty} b_{k'} V_{k+2k'-1}. \tag{75}$$

In order to evaluate the minimum MSE, we first invoke the orthogonality principle [25] as

$$\mathsf{E}\left[H_k(1,1)V_{k+2\tau-1}^*\right] = \mathsf{E}\left[\hat{H}_{s,k}(1,1)V_{k+2\tau-1}^*\right] \tag{76}$$

for some integer τ. Evaluating the expectation on the LHS of (76) yields

$$\begin{aligned} \mathsf{E}\left[H_k(1,1)V_{k+2\tau-1}^*\right] &= \sqrt{\frac{\varrho_p}{n_{\mathrm{t}}}}\,\mathsf{E}\left[H_k(1,1)H_{k+2\tau-1}^*(1,1)\right] \\ &= \sqrt{\frac{\varrho_p}{n_{\mathrm{t}}}}A_H(-2\tau+1) \end{aligned} \tag{77}$$

where $A_H(\cdot)$ has been given in (4). Evaluating the expectation on the RHS of (76) leads to

$$\begin{aligned} &\mathsf{E}\left[\hat{H}_{s,k}(1,1)V_{k+2\tau-1}^*\right] \\ &= \sum_{k'=-\infty}^{\infty} b_{k'}\mathsf{E}\left[V_{k+2k'-1}V_{k+2\tau-1}^*\right] \qquad (78) \\ &= \sum_{k'=-\infty}^{\infty} b_{k'}\Big(\varrho_p\mathsf{E}\left[H_{k+2k'-1}(1,1)H_{k+2\tau-1}^*(1,1)\right] + \\ &\qquad (\rho_d+1)\delta_f[-2(\tau-k')]\Big) \qquad (79) \\ &= \sum_{k'=-\infty}^{\infty} b_{k'}\Big(\varrho_p A_H[-2(\tau-k')] + (\rho_d+1)\delta_f[-2(\tau-k')]\Big). \end{aligned} \tag{80}$$

By combining (77) and (80) with (76), we obtain

$$\begin{aligned} \sqrt{\frac{\varrho_p}{n_{\mathrm{t}}}}A_H(-2\tau+1) = \sum_{k'=-\infty}^{\infty} b_{k'}\Big(&\varrho_p A_H[-2(\tau-k')] + \\ &(\rho_d+1)\delta_f[-2(\tau-k')]\Big) \end{aligned} \tag{81}$$

which, as observed from [9], [13], [26], resembles to the orthogonality condition of the OP-aided channel estimator in Section III-A with modified effective pilot and noise power when the training period is $L = 2$. Therefore, following the steps in [9, App.], we can then obtain the MSE for the SP as

$$\epsilon_k^2(1,1) = 1 - \int_{-1/2}^{1/2} \frac{\rho_p \mathsf{SNR}\,|f_1(\lambda)|^2}{n_{\mathrm{t}}\varrho_p\mathsf{SNR}\,f_0(\lambda) + n_{\mathrm{t}}\varrho_d\mathsf{SNR} + n_{\mathrm{t}}} d\lambda \tag{82}$$

where $f_0(\lambda)$ and $f_1(\lambda)$ follows from (12) for $L = 2$, i.e.,

$$\begin{aligned} f_0(\lambda) &= \frac{1}{2}\sum_{\nu=0}^{1} \bar{f}_H\left(\frac{\lambda-\nu}{2}\right), \\ f_1(\lambda) &= \frac{1}{2}\sum_{\nu=0}^{1} \bar{f}_H\left(\frac{\lambda-\nu}{2}\right) e^{\imath 2\pi\frac{\lambda-\nu}{2}}. \end{aligned} \tag{83}$$

Due to the symmetry of fading observations, the interpolator coefficients $\{b_{k'}\}_{k'\in\mathbb{Z}}$ in (75) do not vary with k. This implies that the estimation error incurred by (75) is wide-sense stationary. This is in contrast to the estimation error for the OP, which is cyclostationary since the optimal coefficients $\{a_{k,k'}\}_{k'\in\mathbb{Z}}$ in (8) depends on k via $\ell = k \bmod L$ [16].

References

[1] A. T. Asyhari and S. ten Brink, "Channel estimation for stationary fading channels: Orthogonal versus superimposed pilots," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, p. 83.

[2] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[3] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Rio de Janeiro, Brazil, Dec. 1999, pp. 2162–2166.

[4] D. Slock and A. Medles, "Blind and semiblind MIMO channel estimation," in *Space-Time Wireless Systems*, H. Bölcskei, D. Gesbert, C. B. Papadias, and A.-J. van der Veen, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 279–301.

[5] 3GPP, "Physical channels and modulation (release 11)," 3GPP, Tech. Spec. TS 36.211 V11.0.0 (2012-09), Sep. 2012.

[6] IEEE 802.11n, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. Amendment 5: Enhancements for higher throughput," IEEE Std 802.11n-2009, Oct. 2009.

[7] ETSI, "Digital video broadcasting (DVB); frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," ETSI EN 302 755 V1.3.1 (2012-04), Apr. 2012.

[8] M. L. Moher and J. H. Lodge, "TCMP—A modulation and coding strategy for Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 9, pp. 1347–1355, Dec. 1989.

[9] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[11] T. P. Holden and K. Feher, "A spread-spectrum based synchronization technique for digital broadcast systems," *IEEE Trans. Broadcast.*, vol. 36, no. 3, pp. 185–194, Sep. 1990.

[12] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.
[13] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.
[14] W. Zhou, J. Wu, and P. Fan, "High mobility wireless communications with Doppler diversity: Fundamental performance limits," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6981–6992, Dec. 2015.
[15] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 5464–5476, Mar. 2014.
[16] A. T. Asyhari, T. Koch, and A. Guillén i Fàbregas, "Nearest neighbour decoding and pilot-aided channel estimation in stationary Gaussian flat-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Aug. 2011, pp. 2786–2790.
[17] M. Coldrey and P. Bohlin, "Training-based MIMO systems—Part I: Performance comparison," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5464–5476, Nov. 2007.
[18] G. Wang, F. Gao, and C. Tellambura, "Channel estimation with amplitude constraint: Superimposed training or conventional training?" in *Proc. Can. Workshop Inf. Theory (CWIT)*, May 2011, pp. 190–193.
[19] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
[20] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4360–4373, Aug. 2014.
[21] A. T. Asyhari and A. Guillén i Fàbregas, "MIMO block-fading channels with mismatched CSI," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7166–7185, Nov. 2014.
[22] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
[23] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
[24] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.
[25] H. V. Poor, *An Introduction to Signal Detection Estimation*, 2nd ed. New York, NY, USA: Springer-Verlag, 1994.
[26] A. T. Asyhari, T. Koch, and A. Guillén i Fàbregas. (Apr. 2014). "Nearest neighbor decoding and pilot-aided channel estimation for fading channels." [Online]. https://arxiv.org/abs/1301.1223
[27] A. Lozano, "Interplay of spectral efficiency, power and doppler spectrum for reference-signal-assisted wireless communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5020–5029, Dec. 2008.
[28] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[29] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 661–665.
[30] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[31] L.-H. Nguyen, R. Rheinschmitt, T. Wild, and S. ten Brink, "Limits of channel estimation and signal combining for multipoint cellular radio (CoMP)," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aachen, Germany, Nov. 2011, pp. 176–180.
[32] G. TR36.211, "E-UTRA Physical channels and modulation (release 10)," 3GPP, Tech. Spec. TS 36.211, 2011.
[33] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
[34] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
[35] A. T. Asyhari, T. Koch, and A. Guillén i Fàbregas, "Nearest neighbour decoding with pilot-assisted channel estimation for fading multiple-access channels," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2011, pp. 1686–1693.
[36] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
[37] R. J. Muirhead, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.
[38] J.-P. Tignnol, *Galois' Theory of Algebraic Equations*. Singapore: World Scientific, 2001.
[39] A. T. Asyhari and A. Guillén i Fàbregas, "Nearest neighbour decoding in MIMO block-fading channels with imperfect CSIR," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1483–1517, Mar. 2012.

A. Taufiq Asyhari (S'09–M'13) received the B.Eng. degree (Hons.) from NTU, Singapore, in 2007, and the Ph.D. degree in Information Engineering from the University of Cambridge, U.K., in 2012. He has been a Lecturer in Networks and Communications with Cranfield University, U.K., since February 2017, where he is currently with the Centre for EW, Information and Cyber. He previously held positions at the University of Bradford, National Chiao Tung University, and Bell Laboratories, Stuttgart, Germany. He also held visiting appointments at the University of Stuttgart–Institute of Telecommunications and the NCTU Information Theory Laboratory. His research interests are in the area of information theory, communication and coding theory, and signal processing techniques with applications to wireless and nano-molecular networks.

Dr. Asyhari is a Fellow with the Higher Education Academy, U.K. He received the Best Paper Award at the 11th IEEE–ISWCS in 2014, the Starting Grant from the National Science Council of Taiwan in 2013, and funding from the Cambridge Trust (Yousef Jameel Scholarship) in 2008–2011.

Stephan ten Brink (M'97–SM'11) has been a faculty member with the University of Stuttgart, Germany, since 2013, where he is currently the Head of the Institute of Telecommunications.

From 1995 to 1997 and 2000 to 2003, he was with Bell Laboratories, Holmdel, NJ, USA, conducting research on multiple antenna systems. From 2003 to 2010, he was with Realtek Semiconductor Corp., Irvine, CA, USA, as the Director of the Wireless ASIC Department, developing WLAN and UWB single chip MAC/PHY CMOS solutions. In 2010, he returned to Bell Laboratories, Stuttgart, Germany, as the Department Head of the Wireless Physical Layer Research Department.

Dr. ten Brink is a recipient or a co-recipient of several awards, including the IEEE Stephen O. Rice Paper Prize and the IEEE Communications Society Leonard G. Abraham Prize for contributions to channel coding and signal detection for multiple-antenna systems. He is best known for his work on iterative decoding (EXIT charts) and MIMO communications (soft sphere detection, massive MIMO).