IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 9, 2021, accepted July 31, 2021, date of publication August 16, 2021, date of current version August 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3104997

Optimum Averaging of Superimposed Training Schemes in OFDM Under Realistic Time-Variant Channels

IGNASI PIQUÉ MUNTANÉ, (Graduate Student Member, IEEE), AND M. JULIA FERNÁNDEZ-GETINO GARCÍA, (Member, IEEE)
Department of Signal Theory and Communications, Carlos III University of Madrid, 28911 Leganés, Madrid, Spain

Corresponding author: Ignasi Piqué Muntané (ignasip@tsc.uc3m.es)

This work was supported by the Spanish National Project Hybrid Terrestrial/Satellite Air Interface for 5G and Beyond - Areas of Difficult Access (TERESA-ADA) [Ministerio de Economía y Competitividad (MINECO)/Agencia Estatal de Investigación (AEI)/Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea (UE)] under Grant TEC2017-90093-C3-2-R.

ABSTRACT The current global bandwidth shortage in orthogonal frequency division multiplexing (OFDM)-based systems motivates the use of more spectrally efficient techniques. Superimposed training (ST) is a candidate in this regard because it exhibits no information rate loss. Additionally, it is very flexible to deploy and it requires low computational cost. However, data symbols sent together with training sequences cause an intrinsic interference. Previous studies, based on an oversimplified channel (a quasi-static channel model) have solved this interference by averaging the received signal over the coherence time. In this paper, the mean square error (MSE) of the channel estimation is minimized in a realistic time-variant scenario. The optimization problem is stated and theoretical derivations are presented to attain the optimum amount of OFDM symbols to be averaged. The derived optimal value for averaging is dependent on the signal-to-noise ratio (SNR) and it provides a better MSE, of up to two orders of magnitude, than the amount given by the coherence time. Moreover, in most cases, the optimal number of OFDM symbols for averaging is much shorter, about 90% reduction of the coherence time, thus it provides a decrease of the system delay. Therefore, these results match the goal of improving performance in terms of channel estimation error while getting even better energy efficiency, and reducing delays.

INDEX TERMS OFDM, superimposed training, time-variant channel, channel estimation, least squares, optimization, averaging.

I. INTRODUCTION

Wireless communications have become more demanding and complex with each generation, and the fifth generation (5G) expects to improve on the capabilities of the preceding ones by several orders of magnitude, either in data traffic, lower latencies or multi-connectivity. Last but not least, it espouses the green goal of reducing energy consumption requirements [1]. Furthermore, the global bandwidth shortage motivates the deployment of underexploited transmission ranges like millimeter wave (mmWave) regime [2], [3]. However, if the utilisation of these bandwidths is not efficient and with proper techniques or suitable waveforms, the problem will persist.

The associate editor coordinating the review of this manuscript and approving it for publication was Chen Chen.

Orthogonal frequency division multiplexing (OFDM) has become the most appealing scheme for multiple reasons, e.g., it provides high data rates with efficient management of the frequency resources, it requires low computational cost, and is robust in front of multipath scenarios, among others [4], [5]. Nevertheless, it usually requires an accurate channel estimation in order to compute a coherent demodulation. The common approach is to use pilot-symbol assisted modulation (PSAM) techniques which set aside some OFDM symbols to allocate training sequences [6]. Several compressed sensing (CS) techniques also proved to obtain accurate channel estimations [7], particularly when tracking time-variable underwater accoustic (UWA) channel models [8]. In any case, this exclusiveness in training resources is aggravated in high mobility scenarios where time variability of the channel requires a high density of tracking pilots. Besides, massive multiple-input multiple-output (MIMO) candidate

schemes hamper the deployment of non-overlapping training sequences [9], [10].

Therefore, superimposed training (ST) may become an attractive solution because these techniques superimpose the pilot symbols over the data stream. Thus, ST better exploits the bandwidth efficiency, there is no loss of information rate, and it is easier to avoid pilot contamination issues. Then, the channel estimation can be performed using techniques like maximum likelihood (ML), minimum mean square error (MMSE) or least squares (LS), as well as their linear simplifications or more accurate iterative methods [6], [11]–[15]. Finally, the pilot contribution is usually retrieved after equalization so the original data symbols can be recovered. The main drawback of these schemes is that the channel estimation deteriorates because the superimposed data symbols introduce an additional interference [16].

ST schemes tend to be very flexible since different pilot configurations may be set over the OFDM grid. In [6], it was proven that the optimum MSE under ML and MMSE channel estimation techniques is obtained when equally spaced superimposed pilots are deployed. In this case, those positions which do not have a superimposed pilot may either employ or interpolate the channel estimation from nearest-neighbour positions.

Moreover, ST techniques are known to be remarkably easy to deploy over other schemes. For example, it may be employed when refining the frame synchronization [17], in cooperative relay scenarios [14], as well as in unmanned aerial vehicles (UAV) assisted systems [18]. Adaptation to other waveforms like filter bank multicarrier (FBMC) [19] or to other technologies such as non-orthogonal multiple-access (NOMA) [20], visible light communications (VLC) [21], MIMO systems with sparse channels [15] or with many access devices [22] is feasible, too.

The contribution of [11] conceives a trade-off between the MSE and the throughput when the channel estimation is performed by LS and MMSE methods. It also computes the optimum power of the superimposed pilot in a closed-form solution. Following this analysis, [12] optimizes the data interference by introducing an additional control parameter that improves the MSE and capacity performances. In all the above works, before performing the LS channel estimation, the first order statistics of the received signal, i.e. the mean, is empirically computed. Actually, this averaging of the signal is mandatory in order to ensure a valid channel estimation and data detection (see Section III). Additionally, it helps to reduce the noise and the interference giving more accurate channel estimations.

In terms of channel modeling, most studies define the channel to behave under quasi-static constraints, which mean that for a specific number of time symbols, whose duration is given by the coherence time, the channel coefficients remain completely constant. This simplification is usually employed for two reasons: firstly, the averaging of the received signal over these specific symbols guarantees a better estimation because the channel has not changed, and secondly, the simpler model allows closed-form MSE expressions or their respective optimization solutions to be obtained.

Despite that, more realistic time-variant channel models have been presented in the literature [23]. The classical approach is described in [24] and [25] where densely scattered rays are received with an omni-directional antenna and the spread in frequency domain is dependent on the relative speed between the user equipment (UE) and the base station (BS). Later, [26] proposed a finite-impulse response (FIR) filter that correlates independent Gaussian random variables into a realistic time variability similarly to [24].

Several works have attempted to discuss the implementation of ST in more realistic channel models. In [27], the coefficients of the channel were defined as Fourier basis expansions which were then approximated by a truncated discrete Fourier basis. Other approaches implemented the time-varying models of [28] which expressed the channel with discrete prolate spheroidal basis, particularly with Slepian sequences as the basis functions, [29]. Even though, both studies were more realistic than previous works that considered quasi-static constraints, they did not optimize the averaging since the expressions of the channel became too complex.

In this paper, for the first time to the authors' knowledge, the optimal averaging in ST schemes under realistic time-varying channels that follow a simplified version of [26], is studied. Hence, the correlation of the channel is no longer quasi-static nor is the channel created using a basis expansion approach. In this scenario, a closed-form expression for the MSE of the LS channel estimation is obtained. Then, a classical differential calculus optimization is performed and the optimum number of averages can be computed by solving a transcendent equation. It is shown that the quasi-static approach overestimates properties that highly deteriorate the theoretical results whereas the proposed channel model matches the realistic one. Finally, the consideration of averaging the quasi-static period, given by the coherence time, is disapproved and an optimal number of time averages are computed for a specific signal-to-noise ratio (SNR). Moreover, it is found that using the optimal values determined by these derivations outperforms previous works in terms of MSE of the channel estimation and even leads to a reduction in delay.

The remainder of this paper is organized as follows: in Section II the system model is described, and the proposed channel model is compared with other state-of-the-art approaches. In Section III, the analytical MSE of the LS channel estimation is computed, first under quasi-static constraints and then under the proposed channel model. Section IV provides the optimum number of averages in time, where a second derivative test gives a rule of thumb to prove the minimum condition. Section V shows the simulation results and validates the MSE expressions. Finally, in Section VI some conclusions are pointed out. Also, some computations are detailed in Appendices A-D.

Notation: For simplicity, the analysis is element-kind and non-vectorial notation is required, except in the Appendices, which are self-contained; $(\cdot)^*$ denotes complex conjugate, $\lfloor\cdot\rceil$ and $\lceil\cdot\rceil$ represent the nearest and the next integer value operation, respectively, $\mathbb{E}\{\cdot\}$ is the expectation, $\mathbb{E}_m\{\cdot\}$ is the mean over m and Var $\{\cdot\}$ refers to the variance.

II. SYSTEM MODEL

In this analysis, a ST scheme is implemented in an single-input single-output (SISO) downlink (DL) scenario and a superimposed pilot is added in every position of the OFDM grid.

For ease of notation, a single antenna system is defined because the scope of this work is to analyze the channel estimation stage, where the LS does not involve any spatial processing and each link estimation is treated independently.

A. TRANSMITTER

An OFDM system with K subcarriers is deployed and every time-frequency resource contains data and a superimposed pilot. In the frequency domain, the complete symbol x for the k-th subcarrier and the m-th time symbol is,

$$x_m^k = s_m^k + c_m^k, \quad \begin{cases} \forall\, k \in [0, \cdots, K-1] \\ \forall\, m \in \mathbb{N}_0 \end{cases} \tag{1}$$

where c_m^k and s_m^k are pilots and data symbols, respectively, and $\mathbb{N}_0$ is the set of non-negative integers.

Also, the power of s_m^k and c_m^k, defined as P_s and P_c, respectively, are related as follows,

$$P = P_s + P_c \begin{cases} P_s = (1-\beta)P \\ P_c = \beta P \end{cases} \tag{2}$$

with P being the total power of x_m^k and β the power allocation factor, a parameter that determines the amount of pilot power within the total power. This parameter is bounded between 0 and 1, where 0 would represent a system without pilots and 1 would imply a system without data.

After creating the symbols, these are introduced into the transmission scheme of the OFDM system where some guardbands are added in order to prevent cross-talking between adjacent systems, a K-point inverse fast Fourier transform (IFFT) converts the signal into the time domain and a cyclic prefix (CP) is added in order to prevent inter-carrier and inter-symbol interference (ICI and ISI, respectively).

B. RECEIVER

It is assumed that the temporal signal is correctly received with perfect synchronization and the receiver removes the CP, then a K-point fast Fourier transform (FFT) converts the signal back to frequency domain and the guardbands are discarded. The received symbol y for the k-th subcarrier and m-th time symbol is,

$$y_m^k = H_m^k x_m^k + w_m^k = H_m^k s_m^k + H_m^k c_m^k + w_m^k \tag{3}$$

where H_m^k is the channel coefficient and w_m^k is the additive white Gaussian noise (AWGN) of the system, both in the frequency domain, too. Their statistics follow $H_m^k \sim \mathcal{CN}\left(0, \frac{\sigma_h^2}{K}\right)$ and $w_m^k \sim \mathcal{CN}\left(0, \sigma_w^2\right)$, with σ_h^2 being the power of the channel ($\sigma_h^2 = 1$, by convention) and σ_w^2 the noise power.

This expression shows that the received signal is composed of the transmitted symbol with the interference of the superimposed pilot, both modified by the channel, and the white Gaussian noise. In this case, the amount of data interference is proportional to the ratio of powers between pilots and data, which is controlled by β.

Then, the ST scheme estimates the channel and employs the estimation to retrieve the original data by removing the pilot contribution and equalizing the signal.

Finally, the SNR is defined as

$$\rho_{\text{SNR}} = \frac{P\sigma_h^2}{\sigma_w^2}. \tag{4}$$

C. CHANNEL MODEL

Previous studies of ST defined a quasi-static channel model which guaranteed that for a specific number of symbols, determined by the coherence time, the channel was completely constant. However, the coherence time or coherence symbols are a rough estimation in which the channel remains fairly constant. Better simplifications may be used that consider a smooth evolution of the channel coefficients.

This correlation in time can be implementated by filtering independently each component of the randomly generated channel. As a result, the k-th frequency coefficient of the channel has the following correlation between the m-th and the m'-th time symbols,

$$\mathbb{E}\left\{\left(H_m^k\right)^* H_{m'}^k\right\} = \frac{\sigma_h^2}{K}\rho_t\left(\gamma\,\Delta m\right), \quad \Delta m = \left|m - m'\right|, \quad \gamma = 2\pi\frac{f_d}{\Delta f}\left(1 + \frac{L_{cp}}{K}\right) \tag{5}$$

where $\rho_t(\cdot)$ is the correlation profile of the channel model, γ is a constant that considers the impact of the OFDM scheme, Δf is the subcarrier spacing, L_{cp} is the number of samples of CP and f_d is the Doppler frequency of the scenario, which is computed as $f_d = \frac{v}{c} f_c$, with v being the relative speed between the UE and the BS, c the speed of light constant and f_c the carrier frequency in which the signal is modulated. The proof of (5) from the temporal domain definition is shown in Appendix A.

The possible expressions of $\rho_t(\cdot)$ are shown in Table 1. In [25], a commonly used channel correlation, referred as Clarke's model, is proposed. However, a more realistic approach is defined in [26], where J_0 and $J_{0.25}$ are the 0-th and fractional 0.25-th order Bessel function of the first kind, $\Gamma(\cdot)$ is the Gamma function, A is a normalization factor and

w_H is a Hamming window implemented in order to reduce the Gibbs effect.

Nevertheless, these expressions are intractable when computing analytical solutions. As a consequence, in order to obtain closed-form expressions of the MSE, a further simplification is preferred using the unnormalized sinc (·) function.

In order to verify the validity of the proposed model, in Figure 1 the different models of Table 1 are shown when the mobility of the UE relative to the BS is 30 km/h.

TABLE 1. Summary of correlation profile implementations $\rho_t(\gamma\Delta m)$.

Clarke's model [25]	$J_0(\gamma\Delta m)$
Realistic model [26]	$A^{1/2}\,2^{1/4}\,\Gamma\left(\frac{3}{4}\right)w_H(\Delta m)(\gamma\Delta m)^{-1/4}J_{0.25}(\gamma\Delta m)$
Proposed model	$\mathrm{sinc}(\gamma\Delta m)$

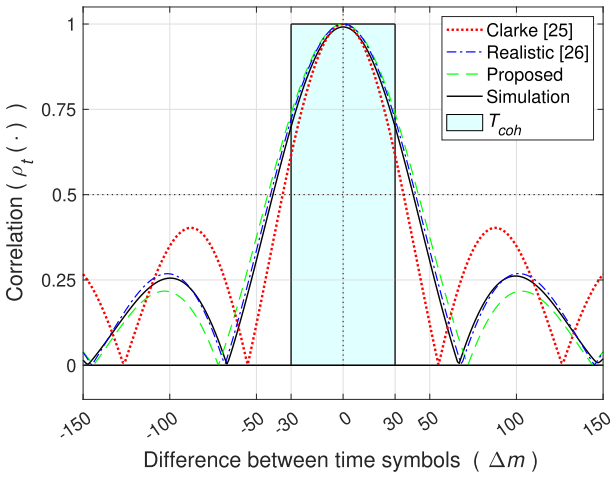

FIGURE 1. Channel correlation models of a UE moving at 30 km/h.

The system considered in the simulations implements the 5G New Radio (NR) numerology [30]. The modulation of the signal is performed in the mmWave regime with a carrier frequency of 28 GHz, and a subcarrier spacing of 120 kHz. The resultant CP overhead is about 7% of the symbol length. This configuration yields a coherence time of 60 symbols (light blue area) when it is computed with the rule of thumb $T_{coh} = \frac{0.423}{f_d}$ [31].

As Figure 1 shows, the validity of the proposed sinc (·) model (green dashed line) is proven since it coincides with the realistic model of [26] (blue dash-dotted line), at least within the coherence time regime. It is true that these expressions do not perfectly match in highly uncorrelated time symbols, e.g. $\Delta m = 100$, however, since the optimum number of averages will not be within these ranges, a perfect overlap is therefore not crucial. Moreover, a numerical simulation of a channel created from filtering independent Gaussian random variables with the correlation weights of the realistic model of [26] (solid black line) is plotted. Indeed, this simulated curve matches the correlation profile employed in its filtering stage (the realistic model), hence, it matches the proposed model, too. Surprisingly, both curves differ substantially from the channel model of [25] (red dotted line), especially at the coherence time edges. Obviously, the state-of-the-art quasi-static approach (light blue area) is the worst overestimated simplification.

Finally, for the sake of simplicity, since the time and frequency domain effects can be modeled independently, it will be assumed that the channel is flat in the frequency domain. Otherwise, an additional analysis could be derived discussing the inaccuracy of the estimation due to the uncorrelatedness of the subcarriers. Analogously to the analysis of this paper in the time domain, the correlation profile in the frequency domain would be linked to an additional error. For this reason, since the scope of this study is to analyze a time-variant channel, i.e. the correlation in the time domain, all the previous channel's notation is considered to be k-th subcarrier-independent and the k index may be omitted. The rest of the manuscript is consistent with this assumption.

III. CHANNEL ESTIMATION PERFORMANCE ANALYSIS

For LS channel estimation in ST, the received signal must be averaged in order to avoid signal cancellation when the supersimposed pilot is retrieved [11], [12]. Also, this averaging computation improves the quality of the signal by reducing noise and interference.

In previous works, the averaging was performed for all the coherence symbols of the channel. The estimation was accurate since it was assumed that the channel coefficients remained constant during the coherence time. However, in reality they evolve smoothly and averaging more symbols than the optimal value may induce an additional error. In this section, the complete MSE is computed, and the optimum number of averages is discussed in extreme scenario situations.

Moreover, since the channel is modeled to be flat fading in the frequency domain, a bidimensional averaging may be performed which allows noise and interference to be reduced more drastically. The following constants $N_f \in [1, K]$ and $N_t \in \mathbb{N}_1$ are introduced, with $\mathbb{N}_1$ being the set of positive integers, which represent the number of subcarriers and time symbols that are going to be averaged, respectively. Then, the averaged received signal is computed with the subgrid of $\left[N_f \times N_t\right]$ symbols,

$$\bar{y} = \sum_{m'=0}^{N_t-1}\sum_{k'=0}^{N_f-1}\frac{1}{N_tN_f}y_{m'}^{k'}. \tag{6}$$

Also, following the system model of [11], [12], pilot symbols are set index-independent ($c_{m'}^{k'} = c$) within the subgrid. The LS computation is defined from the following cost function [32],

$$J_{cost}(H_m) = \left|\frac{c^*}{\beta P}\bar{y} - H_m\right|^2 \rightarrow \hat{H}_{\mathrm{LS}} = \frac{c^*}{\beta P}\bar{y} \tag{7}$$

where $\hat{H}_{\mathrm{LS}}$ is the channel estimation that yields the LS to be later employed in the $\left[N_f \times N_t\right]$ subgrid of symbols. The time dependency in the cost function and in the channel estimation are discussed in the following section in terms of MSE.

A. QUASI-STATIC SCENARIO

The statistical properties of the channel estimation error, $\Delta H = H - \hat{H}$, when N_f subcarriers and N_t time symbols are averaged in a quasi-static channel, are

$$\mathbb{E}\left\{\Delta H\right\} = 0, \tag{8}$$

$$\mathrm{Var}\left\{\Delta H\right\} = \Psi_0 = \frac{1}{N_t N_f \beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right), \tag{9}$$

as [11] derived. The MSE of the estimation in this scenario is defined as Ψ_0 and it will be used in the following analysis.

Since the channel is assumed to be constant within N_f subcarriers and N_t time symbols, the error becomes index-independent. Apart from that, it basically depends on the power allocation and the number of averages, where averaging as much as possible in any domain reduces the error by a compounded inverse factor.

Equation (9) may be interpreted as the sum of the weighted channel power per subcarrier and the noise power. The first term shows how the data interferes with the channel estimation and the second term represents the noise contribution related to the signal power.

If $\beta = 1$ the MSE is equivalent to the PSAM scheme, there is no superimposed data interference and the error is only due to the AWGN, yet the spectral efficiency is nulled because there is no data transmission. Instead, if $\beta = 0$, $\Psi_0 \to \infty$ because the fact of not deploying pilots at all prevents any kind of channel estimation.

B. TIME-VARIANT SCENARIO

As shown in (9), the averaging process drastically reduces the MSE if the channel is quasi-static. However, the channel model presented in Section II-C introduces an additional error due to its variability that must be taken into account in the MSE expression.

Considering this time-variant channel condition, the error becomes time dependent, $\Delta H_m = H_m - \hat{H}$, and after some computations (see Appendix B), it is found that the mean is $\mathbb{E}\left\{\Delta H_m\right\} = 0$, and the MSE, $\Psi_v(m)$, is given by (10), as shown at the bottom of the next page.

It is shown that β does not penalize supplementarily the $\Psi_v(m)$, instead its effect is the same as in (9).

On the other hand, the temporal variability of the channel introduces an additional error related to the integration of $\rho_t(\cdot)$, which shows how the correlation profile of the channel impacts on the MSE computation. Actually, this integration is split into two sums where the first one represents the complete integration and the second one behaves like a correction term dependent on the m-th time symbol, as $\Psi_v(m)$ is defined. The reason behind this dependence is proved by the accuracy of the averaged estimation as opposed to the real m-th time channel coefficient. In general, by the law of large numbers, middle positions like $m = \lceil N_t/2 \rceil$ will be more similar to the averaged estimation than extreme positions, i.e. $m = 0$ or $m = N_t - 1$.

In order to obtain an overall time-independent MSE and get rid of the temporal variable, the mean over m is computed in (11), as shown at the bottom of the next page, and the complete error is defined as Ψ_v.

Then, if the channel is defined as quasi-static during N_t symbols $\left(\rho_t(\gamma\Delta m) = 1,\ \forall \Delta m \in [0, \cdots, N_t - 1]\right)$, Ψ_v is reduced to Ψ_0, which verifies the consistency with (9). In this quasi-static case, averaging as much as possible while the channel remains constant reduces the error of the estimation, as stated before.

Instead, if an extreme time-varying scenario is emulated and a Kronecker delta function is defined as the correlation $\left(\rho_t(\gamma\Delta m) = 0 \text{ for } \Delta m \neq 0 \text{ and } \rho_t(\gamma\Delta m) = 1 \text{ for } \Delta m = 0\right)$, the MSE becomes

$$\Psi_v^{extreme} = \frac{\sigma_h^2}{K}\left(1 - \frac{1}{N_t}\right) + \Psi_0. \tag{12}$$

This expression shows that any averaging performed in an extremely fast fading channel will introduce an additional error since the coefficients will have already changed. The optimum solution would be $N_t = 1$ which reduces the expression into Ψ_0. Nonetheless, in this scenario it is impossible to reduce the noise or the interference by averaging in the time domain.

IV. MSE OPTIMIZATION VIA TIME AVERAGING

In the previous Section III-B the complete expression for MSE was computed and the optimum number of time averages in extreme scenarios was provided. In this section, the optimum number of averages in time (N_t) is derived for any time variability by performing a classical calculus optimization approach [33].

In relation to the optimization of other variables, it is assumed that:

- $N_f^{opt} = K$ because the channel is considered to be flat fading in the frequency domain and its impact is inversely proportional in Ψ_0.
- β^{opt} will not be discussed and the common defined values by academia will be employed ($\beta = 0.2$) [6], [11]–[13], [15], [22]. Usually, the optimization of β requires further analysis related to the throughput which is not covered in this paper. Also, many simulations with different β values were performed over the course of this study and the conclusions described in Section V may be extrapolated because the behaviour of the MSE is similar.

To start with, since the variable to find its optimum (N_t) is discrete, in order to find candidates of minimums by differentiation it is necessary to extend it to the continuous domain. This extension has been considered to be equivalent to the discrete expression (11) since the selected correlation profile function is bounded and differentiable. Also, the equivalence

has been empirically observed in the following Section V. By doing so, the new variable to be optimized is defined as $N_t \rightarrow n_t \in \mathbb{R}_{>0}$ and discrete operations like summations are substituted by integrals. The discrete MSE expression $\Psi_v(N_t)$ is now extended and renamed as the continuous expression $\psi_v(n_t)$.

Then, the channel model proposed in Section II-C, whose integral is already defined as the *Sine Integral* with the following properties [34], is employed

$$\mathrm{Si}\,(x) \;=\; \int_0^x \frac{\sin(t)}{t}\,dt, \tag{13}$$

$$\int_a^b \mathrm{Si}\,(x)\;dx \;=\; \cos(x) + x\,\mathrm{Si}\,(x)\Big|_a^b. \tag{14}$$

Under these assumptions, after employing (13) and (14), Ψ_v is rewritten as (15), as shown at the bottom of the next page, (see Appendix C for detailed calculations).

In general, $\Psi_v(N_t)$ (or $\psi_v(n_t)$) shows an error floor when $N_t, n_t \rightarrow \infty$, however, there are occasions where there is a local minimum before rising into the error floor. With the closed form expression (15) in continuous domain, it is possible to find the minimum candidates by taking the first derivative and computing its roots.

After applying the fundamental theorem of calculus in (13) and the following trigonometric identity,

$$\sin^2(x) = \frac{1}{2}\Big(1 - \cos(2x)\Big), \tag{16}$$

the derivative respect to n_t of (15) is shown in (17), as shown at the bottom of the next page.

Then, their roots may be obtained by solving the following transcendental equation,

$$\mathrm{Si}\left(\gamma\, n_t^{opt}\right) - \gamma\, n_t^{opt}\;\mathrm{sinc}^2\left(\frac{\gamma}{2}\, n_t^{opt}\right) - \frac{\gamma\left(1+\frac{K}{\rho_{\mathrm{SNR}}}-\beta\right)}{2\,N_f\,\beta} = 0 \tag{18}$$

which gives the optimum averaging $N_t^{opt} = \left\lfloor n_t^{opt} \right\rceil$.

In order to verify that the candidate is a minimum, the second derivative test must be fulfilled. In (19), as shown at the bottom of the next page, the second derivative is computed. Since the expression evaluated in n_t^{opt} satisfies (18), the curvature of $\psi_v(n_t)$ is simplified by the following relation (see Appendix D),

$$2\cos\left(\gamma\, n_t^{opt}\right) + \gamma\, n_t^{opt}\sin\left(\gamma\, n_t^{opt}\right) - 2 \underset{\text{maximum}}{\overset{\text{minimum}}{\lessgtr}} 0. \tag{20}$$

Then, the optimum averaging may be obtained by solving (18) and may be verified by checking (20). In case of (18) not having a solution, it means that there is not a local minimum before the error floor and the optimum averaging may be obtained by averaging as much as possible (see Section V).

V. NUMERICAL RESULTS

In this section, several simulation results are shown which graphically prove the consistency of the analysis. First of all, the MSE equations including the time-variant channel effect are plotted against the quasi-static approach. Then, the same curves are compared with the expression that gives the optimum number of averages. After, the MSE is studied in scenarios with different speeds (pedestrian, vehicular and ultra-high-speed mobility) and the SNR dependence on the optimum number of averages is exposed. $N_f = K = 512$ has been considered for all these simulations except for the last ones, where realistic 5G-NR scenarios are emulated and $N_f \neq K$ since the transmission schemes set aside several subcarriers as guardbands. In these simulations, the empirical MSE of the channel estimation is compared with the analytical expressions. Last but not least, all the presented results are focused on the MSE of the channel estimation because under the same power allocation factor β and symbol modulation scheme, the bit error rate (BER) performance of the superimposed symbols will improve as long as the MSE is enhanced [11]–[13].

To start with, in Figure 2 a comparison of the analytical MSE performance, given by (11), between the quasi-static approach and the proposed realistic time-variant channel model is shown. The speed of the scenario is 30 km/h and the number of coherence symbols is $N_c = 60$ (computed in Section II-C and plotted in Figure 1). The curves show low and high SNR regimes, $\rho_{\mathrm{SNR}} = 0$ and 20 dB, respectively.

As can be seen, the state-of-the-art MSE given by the quasi-static model [11], [12] decreases until N_c since the channel correlation is modeled as a binary square function with a step width equivalent to the coherence time. On the other hand, in realistic time-variant scenarios the MSE does not match that performance at all. Actually, in high SNR regimes ($\rho_{\mathrm{SNR}} = 20$dB), averaging N_c symbols in a realistic system yields an MSE that is one order of magnitude worse than the optimum. Also, when averaging N_c symbols, the true MSE obtained in a realistic time-variant scenario (solid blue line) differs by almost two orders of magnitude from the

$$\mathrm{Var}\left\{\Delta H_m\right\} = \Psi_v(m) = \frac{\sigma_h^2}{K}\left(\frac{1}{(N_t)^2}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t\left(\gamma\,|m_1-m_2|\right) - \frac{2}{N_t}\sum_{m_1=0}^{N_t-1}\rho_t\left(\gamma\,|m_1-m|\right) + 1\right) + \Psi_0 \tag{10}$$

$$\Psi_v = \mathbb{E}_m\left\{\Psi_v(m)\right\} = \frac{\sigma_h^2}{K}\left(1 - \frac{1}{(N_t)^2}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t\left(\gamma\,|m_1-m_2|\right)\right) + \Psi_0 \tag{11}$$

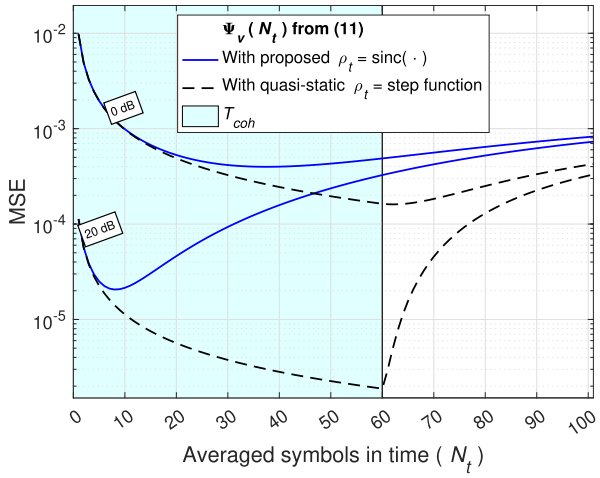

FIGURE 2. MSE with proposed correlation and quasi-static correlation (relative speed 30 km/h).

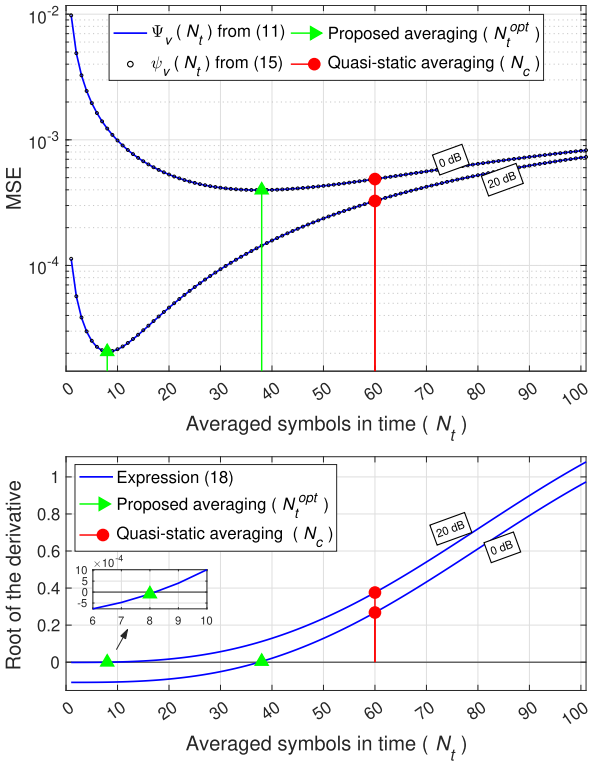

FIGURE 3. Equivalence of (11) and (15) (top); Verification of (18) for the optimum averaging computation (bottom) (relative speed 30 km/h).

expected value of the quasi-static model (black dashed line). This shows the inaccuracy of the state-of-the-art approach, given by the quasi-statical assumption.

Following the same system configuration, the equivalence between expressions (11) and (15) is displayed in Figure 3 (top). Also, the proposed optimal averaging (green triangles) is compared to the oversimplified N_c averaging employed in the state-of-the-art [11], [12] (red circles). It is found that a minimum MSE may be reached by averaging less than in previous works, since in most cases $N_t^{opt} < N_c$. Then, the benefits are twofold: a lower MSE, indeed the minimum MSE, and a shorter averaging which yields a smaller delay. Additionally, this last benefit also involves an improvement in terms of computational cost by a reduction of the number of operations in the averaging computation which requires fewer summands. The minimum values depend on the SNR of the system and they correspond to the roots of the transcendental equation (18) as Figure 3 (bottom) shows. Needless to say, both optimum values satisfy the second derivative test of (20).

Once the equivalence is proved, theoretical simulations of the rest of the section will be computed using (11) or (15) indistinctly. The optimum number of averages will be considered to satisfy (18) and (20), too.

In order to study the dependence of the SNR in the MSE, the following Figures 4–6 show the performance of the estimation in pedestrian, vehicular and ultra-high-speed mobility scenarios with different SNRs. In general, the curves with lower MSE correspond to higher SNR systems. Also, each figure plots the optimum number of averages and opposes them to N_c.

In the lowest speed scenario (Figure 4), it is evident that the optimum averaging is in general much shorter than the coherence time ($N_c = 360$ symbols). In low SNR regimes, since the noise power is important, the optimum averaging

$$\psi_v(n_t) = \frac{\sigma_h^2}{K}\left(1 - \frac{2}{(\gamma\, n_t)^2}\left(\cos(\gamma\, n_t) + \gamma\, n_t\, \mathrm{Si}(\gamma\, n_t) - 1\right)\right) + \frac{1}{n_t N_f \beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right) \tag{15}$$

$$\frac{\mathrm{d}}{\mathrm{d}n_t}\psi_v(n_t) = \frac{\sigma_h^2}{K}\left(\frac{2}{\gamma\, n_t^2}\,\mathrm{Si}(\gamma\, n_t) - \frac{2}{n_t}\,\mathrm{sinc}^2\left(\frac{\gamma}{2}\, n_t\right)\right) - \frac{1}{n_t^2 N_f \beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right) = 0 \tag{17}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}n_t^2}\psi_v(n_t) = \frac{\sigma_h^2}{K}\left(\frac{6}{n_t^2}\,\mathrm{sinc}^2\left(\frac{\gamma}{2} n_t\right) - \frac{4}{\gamma\, n_t^3}\,\mathrm{Si}(\gamma\, n_t) - \frac{2}{n_t^2}\,\mathrm{sinc}(\gamma\, n_t)\right) + \frac{2}{n_t^3 N_f \beta}\left(\frac{\sigma_h^2}{K}(1-\beta) + \frac{\sigma_w^2}{P}\right) \tag{19}$$

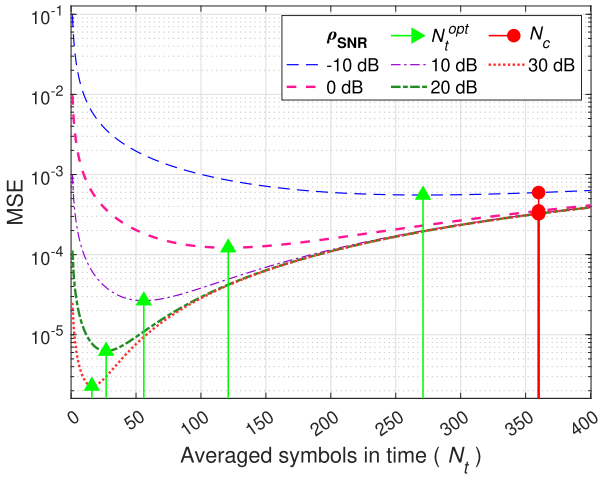

FIGURE 4. MSE of a pedestrian mobility system (relative speed 5 km/h).

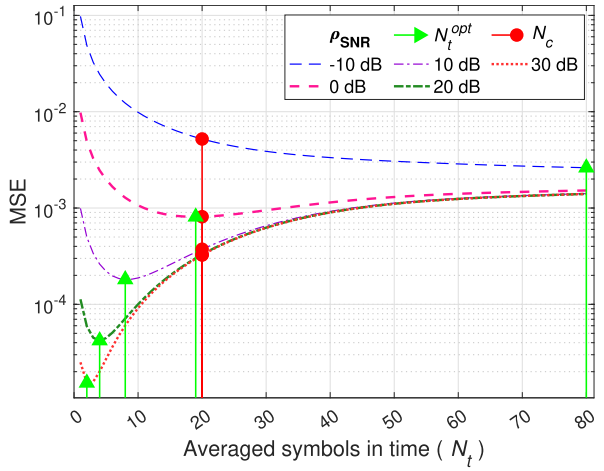

FIGURE 5. MSE of a vehicular mobility system (relative speed 90 km/h).

implies a longer duration. In this case, the main effect of the averaging is focused on reducing the noise presence. Also, the MSE does not show a deep valley and the error-floor appears directly. In contrast, in high SNR regimes, the averaging is less productive since there is not so much noise that can be reduced, and its contribution is focused on minimizing the superimposed data interference. Thereby, a local minimum appears before the asymptotic error floor. After that, the channel has changed sufficiently so the data interference blocks out any reduction of MSE by averaging.

In any case, these optimum values yield a superior MSE performance than coherence time values while offering a much lower latency, e.g., when $\rho_{\text{SNR}} = -10$ dB, the performance is the same but the latency is reduced by 24.72% ($N_t^{opt} = 271$ symbols); or when $\rho_{\text{SNR}} = 30$ dB, the performance is improved by two orders of magnitude and the latency is reduced by 95.55% ($N_t^{opt} = 16$ symbols).

The profile of the curves is similar in vehicular and ultra-high-speed mobility, Figures 5 and 6, respectively. Overall,

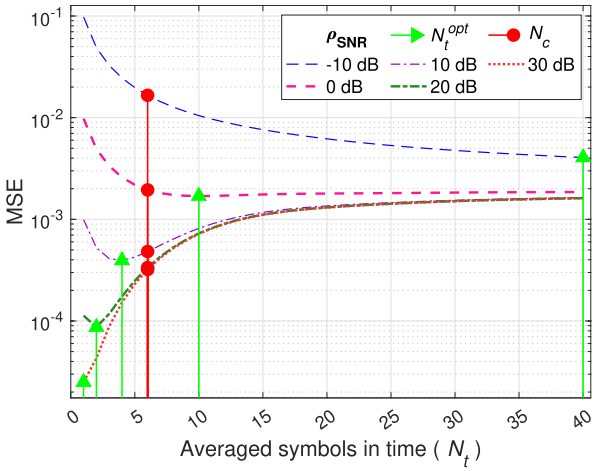

FIGURE 6. MSE of an ultra-high speed mobility system (relative speed 300 km/h).

the valleys and error-floors increase for higher time variabilities.

In high SNR regimes ($\rho_{\text{SNR}} = 30$ dB), previously made assumptions may be extrapolated. The latency reduction is almost preserved: 90% in vehicular ($N_t^{opt} = 2$ and $N_c = 20$) and 83.33% in ultra-high mobility ($N_t^{opt} = 1$ and $N_c = 6$). However, the improvement of the MSE is more modest but is still one order of magnitude.

On the other hand, in low SNR regimes, the high variability of the channel impedes the averaging to be useful when reducing noise and interference. In these situations, since the data interference over the superimposed pilot will possibly not be reduced, the best strategy is to average as much as possible even though the channel has already changed. Then, at least the noise will be diminished until reaching the MSE error floor.

A more in-depth study of the optimum strategy when selecting the number of symbols to be averaged is shown in Figure 7. The plot reflects the optimum number of averages against the number of coherence symbols for several time variabilities of the channel, given by the speeds, and different SNR values. The coherence symbols are labeled in the bottom axis, and their corresponding speeds are set in the top axis.

As is shown, almost all the proposed optimum averages fall below the black dashed line, which represents the state-of-the-art averaging, given by the coherence time. Also, in high SNR situations the reduction of latency is enhanced, which fulfills the previously stated assumptions. Conversely, in low SNR situations where the optimum values match the coherence time or higher values, the best approach is to average as much as possible if latency is not a constraint. These latter scenarios represent the cases where the MSE does not show a deep valley and the only improvement may be reached by averaging at the error floor regime.

From this figure, it can be seen the linear correlation dependence between the optimums in log-log scale. It could be interesting to fit a regression curve with several input

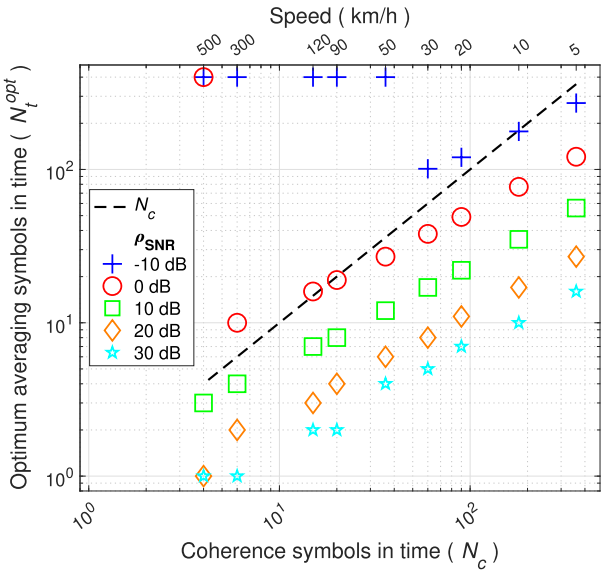

FIGURE 7. Optimum number of averages against the coherence symbols for several time-variant channels.

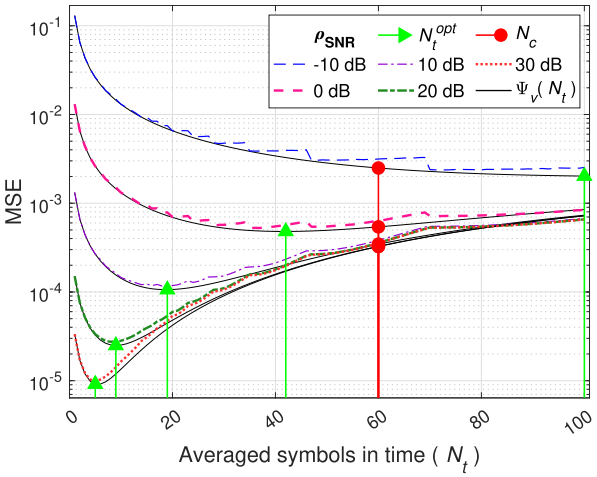

FIGURE 8. MSE curves of 5G-NR empirical simulation and theoretical analysis with a relative speed of 30 km/h.

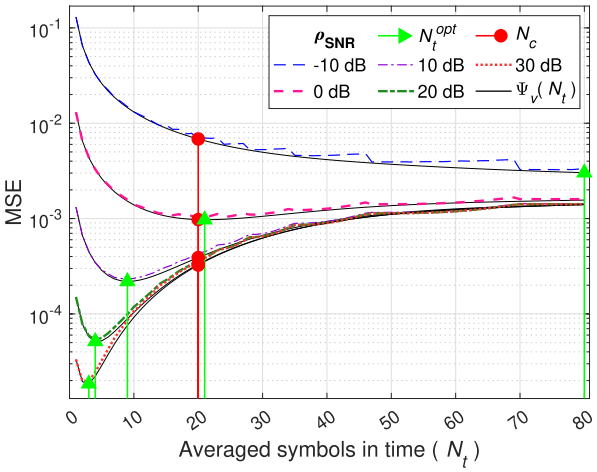

FIGURE 9. MSE curves of 5G-NR empirical simulation and theoretical analysis with a relative speed of 90 km/h.

parameters so the computation of these minimums would be done straightforwardly and any transcendental equation solution, like (18), would be avoided. In this parametric study, the robustness of the approximated model and the sensibility to non-perfect input of parameters such as the Doppler frequency or the SNR would be advised.

To illustrate the robustness of the proposed model, an example of the MSE deterioration when the input values are inaccurate, is discussed. The case study is performed for the presented model of solving transcendent equations, and uses as the benchmark the curve of $\rho_{\text{SNR}} = 10$ dB from the $v = 90$ km/h ($f_d = 2.3$ kHz) scenario of Figure 5, where the optimum averaging is $N_t^{opt} = 8$. In this case, in order to allow a worsening of less than 50% of the minimum MSE, the range of averaging values must befall within $N_t^{\text{robust}} \in [4, 16]$. Then, in order to compute this optimum value in its allowed range, considering a perfect knowledge of the SNR, the relative speed between the UE and the BS must be $v^{\text{robust}} \in [35, 350]$ km/h, which in terms of Doppler frequency is a range of $f_d^{\text{robust}} \in [0.9, 9.5]$ kHz; on the other hand, if there is a perfect knowledge of the Doppler frequency, the SNR range of possible mismatching extends to $\rho_{\text{SNR}}^{\text{robust}} \in [3, 22]$ dB. Obviously, the presented model exhibits a high robustness since the ranges of tolerance to input imprecisions are very broad relative to the resultant MSE computation worsening.

Finally, in the last Figures 8–10 a realistic 5G-NR system is simulated where the configuration of the system model is the same as in Section II-C and the speed is 30 km/h (like in Figures 1–3), 90 km/h and 300 km/h, respectively. The main difference is that the number of subcarriers averaged is no longer 512, instead $N_f = 384$. It is because the bandwidth of transmission is 50 MHz, and 128 subcarriers must be reserved for guardband. Also, the channel coefficients employed in these figures have been created by filtering independent Gaussian random variables with the correlation weights of the realistic model of [26], like the solid black line shown in Figure 1 for the scenario with a relative speed of 30 km/h.

Theoretical curves are plotted as black continuous lines while coloured dashed lines represent empirical simulations. As a matter of fact, both type of curves roughly match in every scenario, which verifies the validity of using the proposed channel model as the realistic scenario, and the optimization algorithm solution as the optimum averaging. Any discrepancy among curves may be attributed either to the difference between models or to the requirement of using more computational power so the randomness is smoothed by performing more iterations. In any case, the optimum averaging values overlap both curves.

In the end, it may be stated that the simplified correlation model assumed in the analysis is sufficiently similar to the realistic one. Thus, theoretical N_t^{opt} values match both models

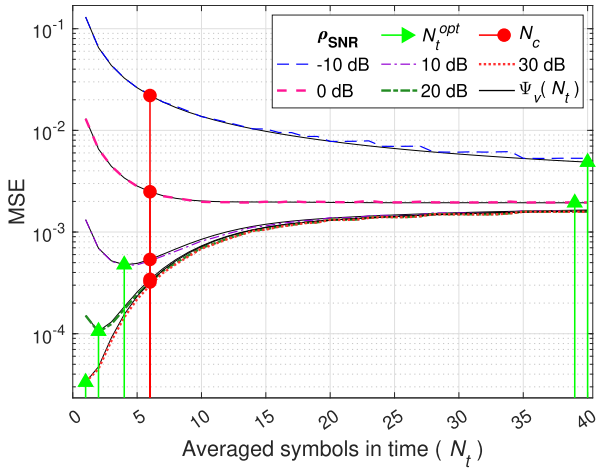

FIGURE 10. MSE curves of 5G-NR empirical simulation and theoretical analysis with a relative speed of 300 km/h.

while improvements in latency and performance are also achieved.

VI. CONCLUSION

In this paper, ST techniques have been studied under a more realistic approach where time variability of the channel is continuous, following a Doppler fading spectrum, instead of being a discrete quasi-static model. Also, in order to obtain closed-form solutions for the MSE of the channel estimation, a simplified channel model, yet similar to the realistic scenario simulations, is employed. With this assumption, the analytical MSE and its extended continuous version are derived, and a classical calculus optimization is performed to determine the duration of time averaging. Moreover, the candidates to be minima are verified through the second derivative test.

Simulation results have shown that ST techniques can be accurate in channel estimation performance and efficient in latency feasibility. Quasi-static channel simplifications behave poorly, and proper channel models must be used. The results reveal an intrinsic SNR dependence that may be exploited by selecting the optimum number of averages through an effective strategy. Finally, theoretical curves are plotted against realistic 5G-NR simulations and the validity and consistency of the presented results are proven.

APPENDIX A
TEMPORAL CHANNEL CORRELATION IN FREQUENCY DOMAIN

The evolution of the channel is commonly studied through the evolution of the sampled taps of the power delay profile (PDP) in the time domain. In order to obtain (5) it is necessary to transform the channel coefficients into the frequency domain.

A channel with L coefficients, which are the resolved multipath components, is defined with the $\mathbf{h}_m^{(t)}$ vector with dimensions $(L \times 1)$. The total power of the channel is σ_h^2 and the power of the l-th tap is $\sigma_{t_l}^2$, which fulfills $\sigma_h^2 = \sum_{l=0}^{L-1} \sigma_{t_l}^2$. The evolution of each component is modeled by the correlation,

$$\mathbb{E}\left\{\left[\mathbf{h}_m^{(t)}\right]_l^* \left[\mathbf{h}_{m'}^{(t)}\right]_l\right\} = \sigma_{t_l}^2 \rho_t\left(\gamma\,\Delta m\right). \tag{21}$$

Even though this article is focused on single-tap channels, the following proof satisfies any channel length.

The relation between the channel coefficients in the time domain and the frequency domain is defined by the discrete Fourier transform (DFT), which can be written using the following matrix form,

$$\mathbf{h}_m^{(f)} = \frac{1}{\sqrt{K}} \mathbf{F}_L \mathbf{h}_m^{(t)} \tag{22}$$

where $\mathbf{h}_m^{(f)} = \left[H_m^1 \cdots H_m^K\right]^T$ is a vector with $(K \times 1)$ channel coefficients in the frequency domain and $\mathbf{F}_L$ is the front $(K \times L)$ submatrix of $\mathbf{F}$ whose components are $\left[\mathbf{F}\right]_{k,l} = e^{-j\frac{2\pi}{K}kl}$.

Then, the components of the matrix product of (22) may be written in a summation form which allows (21) to be expressed in the frequency domain as follows,

$$\begin{aligned}
\mathbb{E}\left\{\left(H_m^k\right)^* H_{m'}^k\right\} &= \frac{1}{K} \sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} \mathbb{E}\left\{\left[\mathbf{h}_m^{(t)}\right]_{l_1}^* \left[\mathbf{h}_{m'}^{(t)}\right]_{l_2}\right\} \\
&\quad \cdot e^{j\frac{2\pi}{K}k(l_2-l_1)} \\
&= \frac{1}{K} \sum_{\substack{l_1=0 \\ (l_1=l_2)}}^{L-1} \mathbb{E}\left\{\left[\mathbf{h}_m^{(t)}\right]_{l_1}^* \left[\mathbf{h}_{m'}^{(t)}\right]_{l_1}\right\} \\
&\quad \{\text{cross-terms} = 0\} \\
&= \frac{1}{K} \rho_t\left(\gamma\,\Delta m\right) \sum_{l=0}^{L-1} \sigma_{t_l}^2 \\
&= \frac{\sigma_h^2}{K} \rho_t\left(\gamma\,\Delta m\right)
\end{aligned} \tag{23}$$

which gives the temporal correlation in terms of frequency domain coefficients, as (5) stated.

APPENDIX B
CHANNEL ESTIMATION ERROR IN A TIME-VARIANT CHANNEL

The channel estimation error in a time-variant channel is expressed in (24), as shown at the top of the next page. This expression shows how the true channel coefficient H_m (Ⓑ) is not perfectly estimated but approximated by the averaging (Ⓐ). If the channel is highly static, the average will be similar to the true value. Thus, the error will be reduced, and the estimation will be more accurate. However, if the mobility is high, the average will not match the true channel and the error of the estimation will increase. Moreover, the dependency on the m-th time symbol is comprehensible since the averaging is more accurate for middle symbol positions than for extreme ones.

The statistical properties show that the mean of the error, like in the quasi-static approach, is 0, whereas the MSE, Ψ_v, is computed as the mean of all the cross-products of (24).

$$\Delta H_m = \hat{H} - H_m = \underbrace{\frac{1}{N_t}\sum_{m'=0}^{N_t-1} H_{m'}}_{\textcircled{A}} \underbrace{- H_m}_{\textcircled{B}} + \underbrace{\frac{1}{N_t N_f}\frac{c^*}{\beta P}\left(\sum_{m'=0}^{N_t-1} H_{m'}\sum_{k'=0}^{N_f-1} s_{m'}^{k'} + \sum_{m'=0}^{N_t-1}\sum_{k'=0}^{N_f-1} w_{m'}^{k'}\right)}_{\textcircled{C}} \tag{24}$$

From this computation, the only cross-products that are not cancelled are the following,

$$S_1 = \textcircled{A}^* \cdot \textcircled{A} = \frac{1}{(N_t)^2}\mathbb{E}\left\{\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\left(H_{m_1}\right)^* H_{m_2}\right\}$$
$$= \frac{1}{(N_t)^2}\frac{\sigma_h^2}{K}\sum_{m_1=0}^{N_t-1}\sum_{m_2=0}^{N_t-1}\rho_t\left(\gamma\,|m_1 - m_2|\right), \tag{25}$$

$$S_2 = \textcircled{A}^* \cdot \textcircled{B} = -\frac{1}{N_t}\mathbb{E}\left\{\sum_{m_1=0}^{N_t-1}\left(H_{m_1}\right)^* H_m\right\}$$
$$= -\frac{1}{N_t}\frac{\sigma_h^2}{K}\sum_{m_1=0}^{N_t-1}\rho_t\left(|m_1 - m|\right), \tag{26}$$

$$S_3 = \textcircled{B}^* \cdot \textcircled{A} = \textcircled{A}^* \cdot \textcircled{B} = S_2, \tag{27}$$

$$S_4 = \textcircled{B}^* \cdot \textcircled{B} = \mathbb{E}\left\{\left(-H_m\right)^*\left(-H_m\right)\right\} = \frac{\sigma_h^2}{K}, \tag{28}$$

$$S_5 = \textcircled{C}^* \cdot \textcircled{C} = \Psi_0. \tag{29}$$

Finally, $\Psi_v(m) = \sum_{i=1}^{5} S_i$ as (10) showed.

APPENDIX C
CONTINUOUS MSE COMPUTATION

The expansion into the continuous domain of (11) is straightforward except for the double summation factor. In order to compute it, the double summation is transformed into the double integral,

$$\int_0^{n_t}\int_0^{n_t}\rho_t\left(\gamma\,|n_{t1} - n_{t2}|\right) dn_{t1}\, dn_{t2} \tag{30}$$

where the correlation is substituted by the even function $\rho_t(\zeta) = \mathrm{sinc}(\zeta)$ and the absolute operator of the argument can be dropped.

Then, the double integral (30) is solved after applying the following change of variables with Jacobian $1/\gamma^2$,

$$\begin{pmatrix} u \\ v \end{pmatrix} = \gamma \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} n_{t1} \\ n_{t2} \end{pmatrix}$$

which gives,

$$\int_0^{\gamma\, n_t}\int_{u-\gamma\, n_t}^{u}\frac{1}{\gamma^2}\,\mathrm{sinc}\,(v)\; dv\, du$$
$$= \frac{1}{\gamma^2}\int_0^{\gamma\, n_t}\mathrm{Si}\,(u) - \mathrm{Si}\,(u - \gamma\, n_t)\; du$$
$$= \frac{2}{\gamma^2}\Big(\cos\left(\gamma\, n_t\right) + \gamma\, n_t\,\mathrm{Si}\left(\gamma\, n_t\right) - 1\Big) \tag{31}$$

and (15) is finally computed.

APPENDIX D
SECOND DERIVATIVE TEST

The evaluation of (19) with the optimum candidate $\left(n_t^{opt}\right)$, which fulfills (18), is simplified as,

$$\frac{\mathrm{d}^2}{\mathrm{d}n_t^2}\psi_v\left(n_t^{opt}\right) = \frac{\sigma_h^2}{K}\frac{2}{\left(n_t^{opt}\right)^2}\left(\mathrm{sinc}^2\left(\frac{\gamma}{2}n_t^{opt}\right) - \mathrm{sinc}\left(\gamma\, n_t^{opt}\right)\right). \tag{32}$$

Considering that $n_t^{opt} > 0$, the second derivative test of (32) is equivalent to,

$$\mathrm{sinc}^2\left(\frac{\gamma}{2}n_t^{opt}\right) - \mathrm{sinc}\left(\gamma\, n_t^{opt}\right) \underset{\text{minimum}}{\overset{\text{maximum}}{\lessgtr}} 0. \tag{33}$$

Finally, after applying (16) and performing some changes, the previous inequality may be rewritten as

$$\frac{2}{\left(\gamma\, n_t^{opt}\right)^2}\left(1 - \cos\left(\gamma\, n_t^{opt}\right)\right) - \frac{\sin\left(\gamma\, n_t^{opt}\right)}{\gamma\, n_t^{opt}} \underset{\text{minimum}}{\overset{\text{maximum}}{\lessgtr}} 0 \tag{34}$$

which results in (20).

REFERENCES

[1] A. Gupta and E. R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
[3] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 3rd Quart., 2018.
[4] H.-C. Wu, "Analysis and characterization of intercarrier and interblock interferences for wireless mobile OFDM systems," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 203–210, Jun. 2006.
[5] V. Tarokh, S. M. Alamouti, V. Koivunen, and K. Kim, "MIMO-OFDM and its application," *J. Commun. Netw.*, vol. 9, no. 2, pp. 109–111, Jun. 2007.
[6] T. Cui and C. Tellambura, "Pilot symbols for channel estimation in OFDM systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 4, Dec. 2005, p. 5.
[7] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Novel compressed sensing-based channel estimation algorithm and near-optimal pilot placement scheme," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2590–2603, Apr. 2016.
[8] Y. Zhang, R. Venkatesan, O. A. Dobre, and C. Li, "Efficient estimation and prediction for sparse time-varying underwater acoustic channels," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 1112–1125, Jul. 2020.
[9] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.
[10] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3366–3370.

[11] W.-C. Huang, C.-P. Li, and H.-J. Li, "On the power allocation and system capacity of OFDM systems using superimposed training schemes," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1731–1740, May 2009.
[12] J. C. Estrada-Jiménez and M. J. .-G. García, "Partial-data superimposed training with data precoding for OFDM systems," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 234–244, Jun. 2019.
[13] S. He, J. K. Tugnait, and X. Meng, "On superimposed training for MIMO channel estimation and symbol detection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 3007–3021, May 2007.
[14] L. He, Y.-C. Wu, S. Ma, T.-S. Ng, and H. V. Poor, "Superimposed training-based channel estimation and data detection for OFDM amplify-and-forward cooperative systems under high mobility," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 274–284, Jan. 2012.
[15] B. Mansoor, S. J. Nawaz, B. Amin, S. K. Sharma, and M. N. Patwary, "Superimposed training based estimation of sparse MIMO channels for emerging wireless networks," in *Proc. 23rd Int. Conf. Telecommun. (ICT)*, May 2016, pp. 1–6.
[16] N. Chen and G. T. Zhou, "What is the price paid for superimposed training in OFDM?" in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, May 2004, p. 4.
[17] J. K. Tugnait and X. Meng, "On superimposed training for channel estimation: Performance analysis, training power allocation, and frame synchronization," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 752–765, Feb. 2006.
[18] S. Gong, S. Wang, C. Xing, S. Ma, and T. Q. S. Quek, "Robust superimposed training optimization for UAV assisted communication systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1704–1721, Mar. 2020.
[19] J. C. Estrada-Jiménez, K. Chen-Hu, M. J. F.-G. García, and A. G. Armada, "Power allocation and capacity analysis for FBMC-OQAM with superimposed training," *IEEE Access*, vol. 7, pp. 46968–46976, 2019.
[20] J. Ma, C. Liang, C. Xu, and L. Ping, "On orthogonal and superimposed pilot schemes in massive MIMO NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2696–2707, Dec. 2017.
[21] J. C. Estrada-Jiménez, B. G. Guzman, M. J. F.-G. García, and V. P. G. Jimenez, "Superimposed training-based channel estimation for MISO optical-OFDM VLC," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6161–6166, Jun. 2019.
[22] K. Zhang, W. Wang, and H. Yin, "Simultaneous channel estimation and data detection based on superimposed training for many access MIMO system in uplink," *IEEE Access*, vol. 8, pp. 123799–123812, 2020.
[23] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, Jun. 2003.
[24] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Hoboken, NJ, USA: Wiley, 1994.
[25] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, July/Aug. 1968.
[26] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology and Techniques*, 2nd ed. Norwell, MA, USA: Kluwer, 2000.
[27] X. Dai, H. Zhang, and D. Li, "Linearly time-varying channel estimation for MIMO/OFDM systems using superimposed training," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 681–693, Feb. 2010.
[28] T. Zemen and C. F. Mecklenbrauker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.
[29] J. K. Tugnait and S. He, "Multiuser/MIMO doubly selective fading channel estimation using superimposed training and slepian sequences," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1341–1354, Mar. 2010.
[30] *5G; NR; Physical Channels and Modulation*, document 3GPP, (TS) 36.331, 01 2021, Version 16.4.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138200_138299/138211/16.04.00_60/ts_138211v160400p.pdf
[31] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[33] M. Spivak, *Calculus*, 4th ed. Houston, TX, USA: Publish or Perish, 2008.
[34] K. B. Oldham, J. Myland, and J. Spanier, *An Atlas of Functions: With Equator, the Atlas Function Calculator*, 2nd ed. New York, NY, USA: Springer, 2010.

IGNASI PIQUÉ MUNTANÉ (Graduate Student Member, IEEE) received the B.S. degree in engineering physics from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2018, and the M.S. degree in multimedia and communications from the Universidad Carlos III de Madrid (UC3M), Madrid, Spain, in 2019, where he is currently pursuing the Ph.D. degree. In 2019, he joined the Communications Research Group, UC3M, and he is involved in TERESA-ADA National Project. His research interests include signal processing and channel estimation techniques for next generation wireless technologies.

M. JULIA FERNÁNDEZ-GETINO GARCÍA (Member, IEEE) received the M.Eng. and Ph.D. degrees in telecommunications engineering from the Polytechnic University of Madrid, Spain, in 1996 and 2001, respectively.

From 1996 to 2001, she held a research position with the Department of Signals, Systems, and Radiocommunications, Polytechnic University of Madrid. She visited Bell Laboratories, Murray Hill, NJ, USA, in 1998; Lund University, Sweden, during two periods in 1999 and 2000; the Politecnico di Torino, Italy, in 2003 and 2004; and Aveiro University, Portugal, in 2009 and 2010. She is currently with the Department of Signal Theory and Communications, Carlos III University of Madrid, Spain, as an Associate Professor. Her research interests include multicarrier communications, coding, and signal processing for wireless systems. She received the Best Master's Thesis Award from the Professional Association of Telecommunication Engineers of Spain, in 1998, the Best Ph.D. Thesis Award from the Professional Association of Telecommunication Engineers of Spain, in 2003, the Student Paper Award at the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), in 1999, the Certificate of Appreciation at the IEEE Vehicular Technology Conference (VTC), in 2000, the Ph.D. Extraordinary Award from the Polytechnic University of Madrid, in 2004, the Juan de la Cierva National Award from the AENA Foundation, in 2004, and the Excellence Award from Carlos III University of Madrid, in 2012, for her research career.