Digital Signal Processing 120 (2022) 103308

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Iteratively reweighted super-resolution channel estimation in hardware-impaired hybrid-precoded MM-wave massive MIMO systems using dual SVD and Marquardt's global search

Waheed Moses Audu *, Olutayo Oyeyemi Oyerinde

School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, 2000, Gauteng, South Africa

ARTICLE INFO

Article history:
Available online 10 November 2021

Keywords:
Hardware-impaired
Hybrid-precoded
Millimeter-wave
Massive MIMO
Dual SVD
Marquardt's search method

ABSTRACT

In this paper, a three-dimensional geometric channel model characterized by channel gain, angle of arrival, and departure (AoA/AoD) is used to develop a non-ideal and more realistic system representation that accounts for additional signal perturbations other than additive white Gaussian noise (AWGN). Hinged on this, compressive sensing (CS) based channel estimation technique is proposed for millimeter-wave (MM-Wave) massive multiple-input multiple-output (MIMO) systems with inherent challenges due to hardware impairment (HI). The proposed estimator named dual singular value decomposition (SVD) and Marquardt, abbreviated as DSM estimator aims at tolerating complexity and improving channel estimation accuracy at reasonable trade-offs in the system. The normalized mean square error (NMSE) performance of the proposed channel estimation scheme, with modified Marquardt's global minimum search, achieves higher accuracy at below 0 dB (LoS)/-3 dB (NLoS) signal-to-noise ratio (SNR) for uniform linear array (ULA) configuration and performs better for uniform planar array (UPA) configuration. A low-rank range of unit difference is adopted for matrix decomposition twice (dual SVD) per SNR in ULA/UPA setting. The proposed estimator's result shows early convergence during the simulation and a linearly scaled complexity in comparison with one of the earlier proposed schemes in literature in particular the super-resolution channel estimation with gradient descent optimization.

1. Introduction

The problem of spectrum crunch and an exponential increase in mobile data traffic before the year 2020 necessitated among others the safe exploration into optimal gains of the spectrum chunk at mm-wave frequencies together with large antenna array configuration at both mobile terminal (MT) and base station (BS) for fifth-generation (5G) implementation [1–3]. As a result, hybrid analog/digital precoding with channel estimation algorithms have been proposed to enable spectral efficient (SE) and energy-efficient (EE) realization in practical mm-wave massive multiple-input multiple-output (MIMO) systems [3–5]. A follow-up on the work in [4] sparked interest in the uplink channel estimation regime. This involves the development of accurate, low-power, and low-complexity channel estimation techniques for the BS where large portions of signal processing are expected to be handled. The work documented in [6] focused on the robustness of hardware imperfection such as phase noise (PN), quantization error (QE), and nonlinearity (NL) to uplink massive MIMO systems. Links to an extensive literature survey on modeling, the joint description of multiplicative phase-drifts ($\mathbf{D}_{\vartheta_B}(t)$) from PN that leads to inter-carrier interference (ICI) which is an additive distortion noise ($\mathbf{e}_B(t)$) in orthogonal frequency division multiplexing (OFDM) system and noise amplification (NA) from thermal noise $\mathbf{n}_B(t)$ were provided. The modulation scheme (single-carrier, OFDM, or filter bank multi-carrier) was found to determine the mapping of circuit imperfection to the three categories considered. The transmitter hardware impairment (HI) was, however, not included. The joint effect of PN, mixer in-phase and quadrature (IQ) imbalance, and power amplifier (PA) NL in various fading channels for ultra-wideband (UWB) OFDM systems based on IEEE 802.15.3.c standard using 60 GHz mm-wave band was presented in [7]. However, no unified system model representation incorporating all known impairments was exemplified.

The effect of HI on the performance of channel estimation (CE) algorithms using residual transmitter HI ($\mathbf{e}_M$) modeled as additive distortion noise was analyzed in [8], somewhat similar to that of [9]. For ease of analysis, transmitter HI was assumed uncorrelated with the transmitted signals and neglected receiver HI ($\mathbf{e}_B$) to be treated as additive noise ($\mathbf{n}$). The HI effect is also prevalent in relay systems for different application areas under various channel con-

* Corresponding author.
E-mail address: waheed.audu1@students.wits.ac.za (W.M. Audu).

https://doi.org/10.1016/j.dsp.2021.103308
1051-2004/

ditions but limited to resolving individual or selected contributing factors [10–15]. Independent and joint effect of PN, IQ imbalance, and nonlinearities (NL) was shown in [16], the details of the mapping are available in [17] where distinctions are made between transceiver multiplicative phase-drifts ($\mathbf{D}_{\vartheta_M}(t)$, $\mathbf{D}_{\vartheta_B}(t)$) and their respective additive distortions ($\mathbf{e}_M, \mathbf{e}_B$). The impact of HI was acknowledged in [18] with numerical results for the optimal level of HIs in uplink multi-user single-antenna massive MIMO Systems and its corresponding Analog to Digital Converter (ADC) resolution. The transmitter HI was however not considered.

It is important to note that hardware impairment refers to uncalibrated imperfections or residual errors after calibration ([6], page 4356). Modelling these impairments and developing digital compensation algorithms for mm-wave [19,20] and massive MIMO [21] is an active research area particularly for OFDM systems [22]. The relative significance of these impairments depends on device characteristics, specific system implementation, and configuration [23]. The complicated nature of accounting for all nonlinearities led to the consideration of the most challenging types. Where hardware impairments (HIs) are used, they are often analyzed individually ([17], [24–27]), partially [28,29], fully combined in effect, or assumed to have been compensated [16] leaving the uncompensated as the residual ([6], [30]). In other words, the ideal hardware is one with assumed compensated algorithms that accounts for only additive white Gaussian noise (AWGN), the non-ideal hardware accounts for any/all of the impairments while residual is used if few (one, two, or the main) and not all impairments are considered ([15]: page 1137, [21]: page 7113-7114). The popular use of the Gaussian model for its representation is based on experimental findings in [31] referenced by [32]. The central limit theory motivates this claim allowing the additive distortion noise to describe aggregate (combined) effects of many residual HIs [21]. Distortion is introduced due to decreased hardware or power cost. The proportionality coefficient in the Gaussian HI model is said to generally increase with signal power and channel gain. The use of the covariance matrix of the signal, similar to [8] compares with that of [21]. The additive distortion model was verified empirically to be accurate for systems that apply compensation algorithms to mitigate main HIs ([21], page 7131). The model is known to represent main HI effects such as quantization errors in automatic-gain-controlled ADC, ICI induced by PN, leakage from subcarrier under I/Q imbalance, and amplitude-to-amplitude (AM-AM) NL in the power amplifier. The multiplicative phase noise effect (to be considered as refinement) cited as a likely major challenge to mm-wave is known to be compensated by reduced symbol time derived from increased bandwidth. It can be incorporated in the channel model [33] or system model [6].

The mm-wave channel redundancy motivates formulating a compressed sensing (CS) based channel estimation problem for sparse signal reconstruction. A comprehensive comparison of the three algorithmic approaches to pilot-aided and CS-based channel estimation problems was carried out in [8]. The paper showed that convex relaxation algorithms, iterative reweight (IR) based, have the highest computational complexity compared to their greedy iterative (GI) and Bayesian inference (BI) counterparts that are quite high. The received signal-to-noise ratio (SNR) and transmitter's hardware impairments were identified causes for the IR-based. Despite the attractive low power performance of the GI algorithms, its angle quantization effect leading to poor estimation accuracy cannot be easily traded off in mm-wave systems. The severity will likely increase with node-to-node proximity. Moreover, the assumption that sparsity is known may not be too realistic [34]. Besides, a similar hardware impairment (HI) effect characterized by amplifier non-linearity, in-phase, and quadrature (IQ) mismatch, direct current (DC) offset, and phase noise (PN) occur at the receiver relative to the transmitter ([7], [24], [35]). In agreement with the first remark of [8], there is, therefore, a need to reduce the computational complexity without significantly decreasing the accuracy. Added to that, is the requirement of a more realistic system representation by accounting for the omitted receiver HI. Despite the limiting complexity of the basis pursuit (BP) algorithm, it remains an attractive source of accurate channel estimates. This propelled further research work by [36], where a global minimum using Newton's method is obtained in contrast to the likelihood of a local minimum via gradient descent method [34]. The Newton method uses the first and second approximate derivatives of the objective function for optimality. The drawback of the Newton method is however the poor performance when the initial angle is not close to the real angle of interest [37], that is, the true angle.

In this work, the methodology of [34] is adopted to harness the accuracy gain of IR over the GI method that super-resolves the angle quantization problem observed when the true angle of departure and arrival (AoD/AoA) is off the uniform distributed grid points. A low complexity method is proposed. The proposed method systematically provides almost exact information on the sparsity level which can further be used to determine the initial grid points close to the real angles. A proposition is made for the use of Marquidt's criterion [38–40] to guarantee global optimality regardless of search direction as opposed to direction-dependent Newton's method [37] used by [36] on the basis that the gradient descent method of [34] may arrive at a local minimum. The iterative reweight super-resolution channel estimation in hardware-impaired hybrid-precoded mm-wave massive MIMO systems using dual singular value decomposition (SVD) and Marquardt's global search, where errors are large, an error penalty is suggested to keep the search tractable. Furthermore, a more realistic system model that incorporates the effect of HI on transceiver sub-systems is proposed. Simulation results are provided to test the system performance using normalized mean square error (NMSE), spectral efficiency (SE), and computation speed as metrics of evaluation.

The major contributions of this paper are as follows. The derivations of system model that incorporate transceiver hardware impairment. Followed by the incorporation of dual SVD into the estimation scheme of [34]. Three versions of the dual SVD based estimation scheme are presented as the proposed systematic method (PSM) 1 to 3. The achievable performance of the proposed scheme as well as the corresponding computational complexities are summarized.

The rest of this paper is organized as follows. In the next section, the proposed system model with transmitter and receiver HI is established. Section three presents the proposed dual SVD initialization and Marquardt's optimization algorithm. Section four shows sets of simulation results for discussion. Section five presents a summary of the work.

Notation: For the rest of this paper, bold upper and lower case symbolize matrices ($E.g.\mathbf{V}, \mathbf{W}$) and vectors ($E.g.\mathbf{v}, \mathbf{w}$), respectively. The transpose and Hermitian transpose operators are denoted as $(.)^T$ and $(.)^H$, respectively. The $diag(\mathbf{v})$ is the vector to matrix transformation with elements of $\mathbf{v}$ on its diagonal while $vec_d(\mathbf{V})$ is the matrix to vector transform of diagonal entries in $\mathbf{V}$. The trace of $\mathbf{W}$ written as $tr(\mathbf{W})$ is the sum of elements on its diagonal. $E[.]$ denotes expectation of [.]. $\mathbb{C}\mathcal{N}(v, w)$ is a complex normal random variable with scalar mean v and variance w. If $\mathbb{R}\mathcal{U}(v, w)$, implies real uniform random variable within an interval $[v, w]$. The set of complex and real matrix or vector of dimension $N_v - by - N_w$ is $\in \mathbb{C}^{(N_v \times N_w)}$ and $\in \mathbb{R}^{(N_v \times N_w)}$, respectively. The Kronecker product of $\mathbf{V}$ and $\mathbf{W}$ is $\mathbf{V} \otimes \mathbf{W}$.

2. System model

Consider Fig. 1 as a single cell consisting of a base station (B-for short) with N_B antennas receive transmit pilot sequences

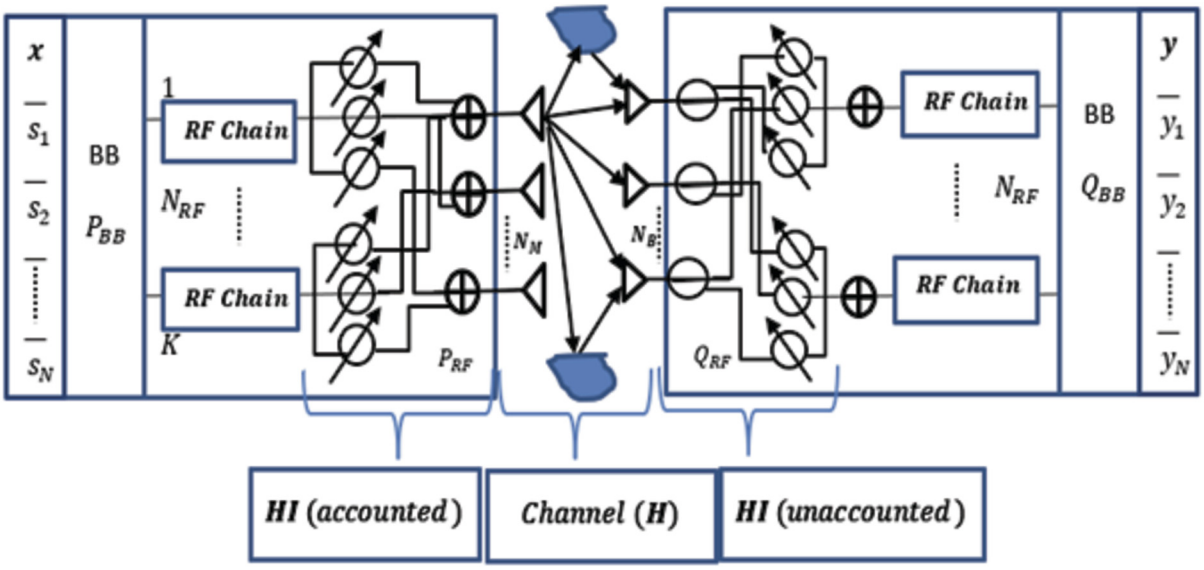


Fig. 1. Millimeter Wave massive MIMO system.

$(s_1, s_2, s_3, \ldots, s_N)$ of size N_s in parallel as vector $\mathbf{x} \in \mathbb{C}^{(N_s \times 1)}$ from the mobile terminal (M-for short) with N_M antennas via baseband (digital) precoder $\mathbf{P}_{BB} \in \mathbb{C}^{(N_{RF} \times N_s)}$. The precoder encodes the symbol vector $\mathbf{x}$ to match the limited number of radio frequency (RF) chains. Each RF chain at the transmitter does digital to analog conversion (DAC) followed by amplification and mixing with a carrier frequency. Analog phase shifters or RF precoder $\mathbf{P}_{RF} \in \mathbb{C}^{(N_M \times N_{RF})}$ is further used to drive the signals to the transmitting antenna segment. The combined transmitter precoding matrix is represented as $\mathbf{P} = \mathbf{P}_{RF}\mathbf{P}_{BB} \in \mathbb{C}^{(N_M \times N_s)}$. Transmitted signal from N_M antenna elements, $\mathbf{s} = \mathbf{P}\mathbf{x} \in \mathbb{C}^{(N_M \times 1)}$ satisfies $E[\mathbf{x}\mathbf{x}^H] = \rho_T N_s^{-1}\mathbf{I}_{N_s}$, where ρ_T is the total power of transmitted symbols and $\mathbf{I}$ is an identity matrix. Since pilot power and pattern are design criteria affecting the performance of sparse channel estimation let $N_s : N_{RF} < N_s < N_M$. For an assumed evenly placed pilot symbols and equal transmit power, the receiver combines all signals from the receiver antenna with RF and baseband combiners in the form of $\mathbf{Q} \in \mathbb{C}^{(N_B \times N_s)}$ where $\mathbf{Q}$ is the product of $\mathbf{Q}_{RF} \in \mathbb{C}^{(N_B \times N_{RF})}$ and $\mathbf{Q}_{BB} \in \mathbb{C}^{(N_{RF} \times N_s)}$. For each set of transmitting pilot sequence $\mathbf{s}_q (1 \leq q \leq N_s)$, T-time slots are used to get $N_B - by - T$ received pilot sequence $\mathbf{y}_{q,T}$. By adopting the Saleh-Valenzuela $(S-V)$ model found to be almost unchanged within channel coherence time, the training overhead is TN_s, and the combining matrix $\mathbf{Q}_{N_B \times N_s}$ is used to get $N_B - by - 1$ received pilot sequence at the t^{th} time slot [34]. The received signal at base station can then be expressed as

$$\mathbf{y}_{q,t} = [\mathbf{Q}^H\mathbf{H}(\mathbf{P}\mathbf{x} + \mathbf{e}_M) + \mathbf{Q}^H\mathbf{n} + \mathbf{Q}^H\mathbf{e}_B] \tag{1}$$

Equation (1) can be written as:

$$\mathbf{y}_{q,t} = [\mathbf{Q}^H\mathbf{H}(\mathbf{s}_q + \mathbf{e}_M) + \mathbf{Q}^H(\mathbf{n} + \mathbf{e}_B)], \tag{2}$$

where $\mathbf{H} \in \mathbb{C}^{(N_B \times N_M)}$ is the channel matrix, $\mathbf{s}_q = \mathbf{P}_{N_M \times N_s}\mathbf{x}_{N_s \times 1} \in \mathbb{C}^{(N_M \times 1)}$, and the noise vector $\mathbf{n} = \mathbf{n}_{N_B \times 1} \sim \mathbb{C}\mathcal{N}(0, \sigma_n^2)$. The transmitter and receiver hardware impairment (HI) vectors are represented by $\mathbf{e}_M$ and $\mathbf{e}_B$, respectively. In [8] only the transmitter HI was accounted for, however, in reality, a similar effect occurs at the receiver ([7], [24], [35]). Four major sources of these impairments are considered in this work. The first is amplifier nonlinearity from the non-linear power amplifier with the distortive effect of high power on message signals that can cause intermodulation noise. This would be more critical in NOMA systems that use power discrimination for user signal detection during successive interference cancellation [35]. Users with low power allocation will experience greater levels of signal degradation than those of higher power. The second is the in-phase and quadrature (I/Q) imbalance. It is the amplitude and phase mismatch between I and Q branches of a received signal meant to have equal amplitude and phase shift of 90^0. This results in imperfect signal rejection and overall performance degradation. The third is the direct current (DC) offset that occurs when the local oscillator (LO) signal leaks up to the RF port of the mixer during mixing. The result is a DC component and a component twice the bandpass signal frequency. This unwanted DC component can reduce the effective dynamic range of the analog-to-digital converter (ADC). Finally, the phase noise (PN) arising from the unstable nature of the LO introduces random phase rotation to the received signal when it is down-converted. This results in a wider than necessary oscillator spectrum. The wealth of knowledge available in the pieces of literature reviewed is drawn to present a unified system model representation as follows. From equation (2), after the T-time slot, the received signal:

$$\mathbf{Y} = \mathbf{Q}^H\mathbf{D}_{\vartheta_M}\mathbf{H}(\mathbf{S} + \mathbf{E}_M) + \mathbf{Q}^H\mathbf{D}_{\vartheta_B}(\mathbf{N} + \mathbf{E}_B), \tag{3}$$

where $\mathbf{Y}$ is $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ and vector $\mathbf{y}_T \in \mathbb{C}^{(N_B \times 1)}$, with the combiner $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_T] : \mathbf{Q}_T \in \mathbb{C}^{(N_B \times N_s)}$, for a total transmit signal matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_T]$, and noise matrix $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_T] : \mathbf{n}_T = \mathbf{n}_{N_B \times 1}$. Let the multiplicative phase-drifts at transmitter and receiver be $\mathbf{D}_{\vartheta_M} = \mathbf{D}_{\vartheta_B}$. Such that $\mathbf{D}_{\vartheta_B} \triangleq diag(e^{(-i\vartheta_{B1}(t))}, \ldots, e^{(-i\vartheta_{BN_B}(t))})$ according to [6] which describes a Wiener process $\vartheta_{BN}(t) \sim \mathcal{N}(\vartheta_{BN}(t-1), \delta), N = 1, \ldots, N_B$ with $\delta = 4\pi^2 f_c^2 T_s \beth$ at symbol time T_s, carrier frequency f_c, and Local Oscillator Constant (LOC) $\beth = 10^{-17}$. Suppose $\vartheta_{BN}(t) = 0$ then $\mathbf{D}_{\vartheta_M}$ and $\mathbf{D}_{\vartheta_B}$ will become identity matrix $\mathbf{I}_{N_B}$. Then equation (3) becomes:

$$\mathbf{Y} = \mathbf{Q}^H\mathbf{H}(\mathbf{S} + \mathbf{E}_M) + \mathbf{Q}^H(\mathbf{N} + \mathbf{E}_B), \tag{4}$$

where the transmitter and receiver additive distortions $(\mathbf{E}_M, \mathbf{E}_B)$ are Gaussian of the form suggested by [8] and [13] such that for uniform linear array (ULA) it is proposed that the following

be used. $\mathbf{E}_M$ is $[\mathbf{e}_{M,1},\ldots,\mathbf{e}_{M,T}]:\mathbf{e}_M\in\mathbb{C}\mathcal{N}(0,k_M vec_d(\mathbf{D}_e)),\mathbf{D}_e\triangleq E[\mathbf{s}\mathbf{s}^H]$. $\mathbf{E}_B=[\mathbf{e}_{B,1},\ldots,\mathbf{e}_{B,T}]:\mathbf{e}_B\in\mathbb{C}\mathcal{N}(0,k_B\mathbf{e}_M)$, and $k_B,k_M\in[0,1]$. The values k_M and k_B represent the levels of degradation assumed to increase at the base station for uplink. In uniform planar array (UPA) configuration, $k_M=k_M^{ULA}\sqrt{(\mu_M)}$ and $k_B=k_B^{ULA}\sqrt{(\mu_B)}$, where $\mu_i:i\in[M,B]$ is the mutual coherence. The $\mathbf{e}_i$ can be modeled as complex Gaussian or uniformly distributed $\mathbb{C}\mathcal{U}(0,k_i vec_d(\mathbf{D}_e)),i\in[M,B]$, to depict the equal likelihood of the symbol attenuation over the channel $\mathbf{H}$, where,

$$\mathbf{H}=\sqrt{\frac{N_B N_M}{\rho}}\sum_{\ell=1}^{L}\alpha_l\mathbf{a}_M(\theta_l)\mathbf{a}_B^H(\phi_l)[4] \tag{5}$$

Here the average path-loss between transceivers is denoted as ρ, N_B and N_M are the number of antenna elements at the base station and mobile terminal, respectively. From Fig. 2, for one user, the complex gain of the ℓth path is α_ℓ, the variables θ_ℓ and ϕ_ℓ both $\in[0,2\pi)$ are ℓth path's elevation and azimuth angles. Angles of departure (AoD) from the transmitter (M) and angle of arrival (AoA) at the receiver (B) are $\mathbf{a}_M$ and $\mathbf{a}_B$. The antenna array steering or response vectors are $\mathbf{a}_M(\theta_\ell)$ and $\mathbf{a}_B(\phi_\ell)$. The choice of this model is a result of signal scattering between transmitter and geometrically located receivers. The antenna array steering/response vectors $\mathbf{a}(\theta_\ell)$ for N Uniform Linear Array (ULA) is determined by one angle [34]:

$$\mathbf{a}(\theta_\ell)=\frac{[1,e^{-i(\frac{2\pi d\sin\theta_\ell}{\lambda})},\ldots,e^{-i(\frac{2\pi d(N-1)\sin\theta_\ell}{\lambda})}]^T}{\sqrt{N}} \tag{6}$$

For $N_i\times N_{ii}$ Uniform Planner Array (UPA), the expression is [34]:

$$\mathbf{a}(\phi_\ell,\theta_\ell)=\frac{[1,e^{-i(\frac{2\pi d\sin\phi_\ell\sin\theta_\ell}{\lambda})},\ldots,e^{-i(\frac{2\pi d(N_i-1)\sin\phi_\ell\sin\theta_\ell}{\lambda})}]^T}{\sqrt{(N_i N_{ii})}}\otimes[1,e^{-i(\frac{2\pi d\cos\theta_\ell}{\lambda})},\ldots,e^{-i(\frac{2\pi d(N_{ii}-1)\cos\theta_\ell}{\lambda})}]^T \tag{7}$$

Let $z_\ell=\alpha_\ell\sqrt{\frac{N_B N_M}{\rho}}:\mathbf{z}=[z_1,z_2,\ldots,z_L]^T$, then the channel model similar to the one in [34] can be written as:

$$\mathbf{H}=\sum_{\ell=1}^{L}z_\ell\mathbf{a}_M(\theta_l)\mathbf{a}_B^H(\phi_l) \tag{8}$$

This is a narrow band frequency flat model. The matrix form of (8):

$$\mathbf{H}=\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H, \tag{9}$$

where

$$\begin{aligned}\mathbf{D}\in\mathbb{C}^{L\times L}&=\text{diagonal matrix with non zero element } z_\ell\\&\quad\text{written as } diag(\mathbf{z}_\ell)\\\mathbf{A}_B\in\mathbf{C}^{N_B\times L}&=[\mathbf{a}(\theta_{B,1}),\ldots,\mathbf{a}(\theta_{B,L})]\text{ for ULA or}[\mathbf{a}(\phi_{B,1},\theta_{B,1}),\\&\quad\ldots,\mathbf{a}(\phi_{B,L},\theta_{B,L})]\text{ for UPA}\\\mathbf{A}_M\in\mathbf{C}^{N_M\times L}&=[\mathbf{a}(\theta_{M,1}),\ldots,\mathbf{a}(\theta_{M,L})]\text{ - ULA and}[\mathbf{a}(\phi_{M,1},\theta_{M,1}),\\&\quad\ldots,\mathbf{a}(\phi_{M,L},\phi_{M,L})]\text{ - UPA}\end{aligned}$$

Then $\mathbf{A}_B$ and $\mathbf{A}_M$ are normalized spacial angles of the form $\mathbf{A}_B$ equal to $\mathbf{A}_B(\mathrm{a}_B)$ and $\mathbf{A}_M$ equal to $\mathbf{A}_M(\mathrm{a}_M)$. The multipath channel exhibits sparsity as only a few dominant signals are received per antenna per position of the transmitter relative to the receiver. This causes the physical channel to be modeled as a discrete virtual angular entity [8]. Discretized points result in an off-grid problem where points outside the quantized regions are not accounted for. Despite the assumption that the AoAs and AoDs are taken from a uniform grid of size $G\gg L$ such that $\phi_\ell,\theta_\ell\in\{0,\frac{2\pi}{G},\ldots,\frac{2\pi(G-1)}{G}\}$, the choice of G will either increase quantization error (when small) or increase computational complexity (when large) [8]. A typical choice of $G>(N_M+N_B)$ is observed in [41]. Therefore, channel quantization error $\mathbf{E}_Q$ is fitted into the channel model as

$$\mathbf{H}=\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H+\mathbf{E}_Q \tag{10}$$

The channel estimation (CE) problem can now be approached using the compressed sensing (CS) method. In [34] the equation (4) was evaluated without considering the hardware impairments, the dual effect of combining matrix on the noise component, and the quantization error. This reduced (4) combined with (8) to

$$\mathbf{Y}=\mathbf{Q}^H\mathbf{H}\mathbf{S}+\mathbf{N} \tag{11}$$

In [8] from the equation (4) the receiver HI was not considered leading to a partial combination of (4) and (9) as

$$\mathbf{Y}=\mathbf{Q}^H(\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H+\mathbf{E}_Q)(\mathbf{S}+\mathbf{E}_M)+\mathbf{Q}^H\mathbf{N} \tag{12}$$

By a full combination of (4) and (5), a more comprehensive and encompassing system representation is achieved as

$$\begin{aligned}&\mathbf{Y}=\mathbf{Q}^H(\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H+\mathbf{E}_Q)(\mathbf{S}+\mathbf{E}_M)+\mathbf{Q}^H\mathbf{N}+\mathbf{Q}^H\mathbf{E}_B\\&\mathbf{Y}=\mathbf{Q}^H(\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H+\mathbf{E}_Q)(\mathbf{S}+\mathbf{E}_M)+\mathbf{Q}^H(\mathbf{N}+\mathbf{E}_B)\\&\mathbf{Y}=\mathbf{Q}^H(\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H(\mathbf{S}+\mathbf{E}_M))+\mathbf{Q}^H(\mathbf{E}_Q(\mathbf{S}+\mathbf{E}_M))+\mathbf{Q}^H(\mathbf{N}+\mathbf{E}_B)\end{aligned} \tag{13}$$

3. Proposed channel estimation techniques

The IR-based channel estimation can be categorized into the initialization and optimization stages. Singular value decomposition (SVD) is used for dimension reduction to obtain the number of multipath and quantized angles without losing relevant information about the signals. This is followed by a search for the real angles for optimized system performance. The iterative reweight is assumed to make quantization error zero. The right-hand side of equation (13) can then be written as

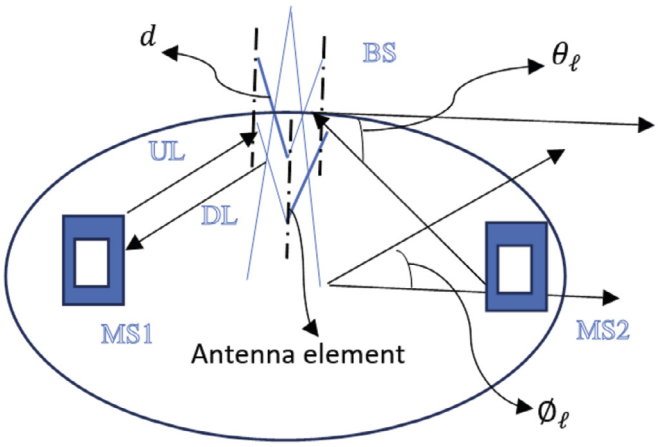
$$\mathbf{Y}=\mathbf{Q}^H(\mathbf{A}_B\mathbf{D}\mathbf{A}_M^H(\mathbf{S}+\mathbf{E}_M))+\mathbf{Q}^H(\mathbf{N}+\mathbf{E}_B), \tag{14}$$

where $\mathbf{D}$ from equation (9) shapes the size of the AoA ($\mathbf{A}_B$) and the AoD ($\mathbf{A}_M$) matrix to form the desired channel estimate ($\hat{\mathbf{H}}$). The objective is therefore to minimize the non-zero elements of $\mathbf{D}$ subject to the Frobenius norm of the difference between the real and estimated observation at a given error tolerance. That is

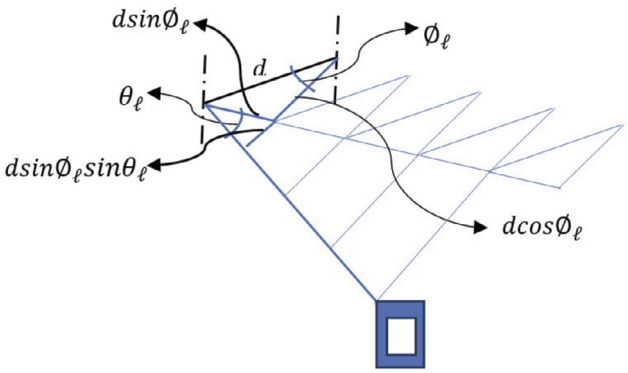
$$\begin{aligned}&\min_{\hat{\mathbf{z}},\hat{\mathbf{a}_M},\hat{\mathbf{a}_B}}\quad\|\hat{\mathbf{z}}\|_0\\&\text{subject to}\quad\|\mathbf{Y}-\mathbf{Q}^H\hat{\mathbf{H}}(\mathbf{S}+\mathbf{E}_M)\|_F\le\varepsilon,\end{aligned} \tag{15}$$

3.1. Initialization by dual singular value decomposition (SVD)

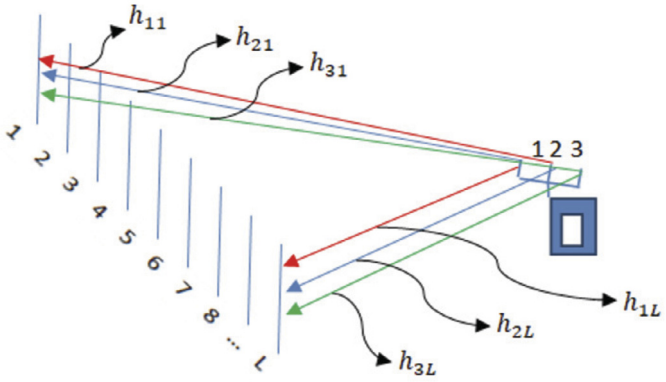
While the paper [34] that used gradient descent (GD) arrives at the unknown number of paths L by assuming an initial number N_{in} in the range $[N_{in},N_{in}+5]$ during the SVD throughout the iteration process, it is proposed that for a given signal-to-noise ratio (SNR) the exact or a better-estimated number of path N_e can be obtained. Beginning with an application of a threshold to the eigenvalue, discard the pivot value and count the number of non-zeros. The pivot value, according to [42], is discarded for being a strong disturbing signal. Subsequently, $[N_e,N_e+n]$ for $n=1,2,\ldots,5$ paths are used in the iterative reweight process. This is expected to reduce the run time (in seconds) and improve convergence to the local/global minimum.



(a) Single cell with azimuth and elevation AoD/AoA



(b) Resolution of azimuth and elevation angles



(c) Channel coefficient along L multiple paths for one user

Fig. 2. (a)-(c) Mobile to Base station interaction in a Time-Division Duplexed (TDD) mode.

3.2. Optimization by Marquardt's global search

To solve the problem of local minimum the use of the Hessian matrix in Marquardt's (M_Q) method [40] is proposed for comparison with the gradient descent (GD) method when the number of multiple paths L is unknown and SVD initialization is used. The proposed alternative, Marquardt's method [38], [39], [40], is a combination of Cauchy's and Newton's method. Details of the derivation for optimization are provided in Appendix A. For the sake of brevity $\frac{\partial^2 \zeta_{opt}}{\partial \mathbf{a}_{M,\ell}{}^2}$ and $\frac{\partial^2 \zeta_{opt}}{\partial \mathbf{a}_{M,\ell} \partial \mathbf{a}_{B,\ell}}$ can be obtained by swapping denominator terms. The first and second partial derivatives of the equation are obtained with (22)-(25) to form the Hessian matrix ($\mathcal{H}$) so that the true angle estimates are as follows.

$$\begin{cases} \hat{\mathbf{a}}_B^{(j+1)} = \hat{\mathbf{a}}_B^{(j)} - (\mathcal{H}^{(j)} + \lambda^{(j)}\mathbf{I})^{-1} \nabla_{\mathbf{a}_{B,\ell}} \zeta_{opt}^{(j)}(\hat{\mathbf{a}}_B^{(j+1)}, \hat{\mathbf{a}}_M^{(j+1)}) \\ \hat{\mathbf{a}}_M^{(j+1)} = \hat{\mathbf{a}}_M^{(j)} - (\mathcal{H}^{(j)} + \lambda^{(j)}\mathbf{I})^{-1} \nabla_{\mathbf{a}_{B,\ell}} \zeta_{opt}^{(j)}(\hat{\mathbf{a}}_B^{(j+1)}, \hat{\mathbf{a}}_M^{(j+1)}) \end{cases} \tag{16}$$

Algorithm 1: DSM-based channel estimator

Input1: The noisy received signal $\mathbf{Y}$, precoded transmit pilot signals $\mathbf{S}_q(\mathbf{S}+\mathbf{E}_M)$, combining matrix $\mathbf{Q}$, pruning threshold z_{th}, termination threshold ε_{th}

1. Let $\hat{\mathbf{H}}_p = \mathbf{Q}(\mathbf{Y}-\varepsilon_{th})\mathbf{S}_q^H$
2. Normalize $\hat{\mathbf{H}}_p = \hat{\mathbf{H}}_{norm}$
3. Do $[\mathbf{U},\mathbf{\Sigma},\mathbf{V}] = SVD(\hat{\mathbf{H}}_{norm}) - 1^{st}$
4. Discard $\mathbf{U}_{i,j}, \mathbf{\Sigma}_{i,j}, \mathbf{V}_{i,j}$ for $i=1, j=1$ and find $\|\mathbf{\Sigma}_{i+1,j+1} > 2z_{th}\|_0$ as $N_{in} = N_e$
5. Initialize $\hat{\mathbf{a}}_B^{(0)}, \hat{\mathbf{a}}_M^{(0)}, \hat{\mathbf{z}}^{(0)}$ by boundaries or by N_e to get $\hat{\mathbf{H}}^{(0)}$
6. ${\mathbf{R}_{es}}^{(k)} = \mathbf{Y} - \mathbf{Q}^H\hat{\mathbf{H}}^{(k)}\mathbf{S}_q : k = 0, 1, 2, \ldots$
7. $[\mathbf{U},\mathbf{\Sigma},\mathbf{V}] = SVD({\mathbf{R}_{es}}^{(k)}) - 2^{nd}$
8. **for** $k = 1, 2, \ldots, iter$
9. **for** $\ell = 1, 2, \ldots, N_e$ **do**
10. $(\hat{\phi}_B^{(1)}, \hat{\theta}_B^{(1)}) = \underset{(\phi_B,\theta_B)\in\Omega_B}{\operatorname{argmax}} \quad \mathbf{u}_\ell^H\mathbf{Q}^H\mathbf{a}_B(\phi_B,\theta_B)$
11. $(\hat{\phi}_M^{(1)}, \hat{\theta}_M^{(1)}) = \underset{(\phi_M,\theta_M)\in\Omega_M}{\operatorname{argmax}} \quad \mathbf{v}_\ell^H\mathbf{S}_q^H\mathbf{a}_M(\phi_M,\theta_M)$

end for(in 9)

Output1:The quantized AoAs/AoDs estimates of the N_e paths

Input2:The effective initial on-grid AoA/AoDs $\hat{\mathbf{a}}_B^{(1)}, \hat{\mathbf{a}}_M^{(1)}$ from **Output1**

12. Initialize $\hat{\mathbf{z}}^{(1)} = \mathbf{z}_{opt}(\hat{\mathbf{a}}_B^{(1)}, \hat{\mathbf{a}}_M^{(1)})$ by (20)
13. **Repeat the following**.
14. Update $\mathfrak{K}$ by $\mathfrak{K} = \max(\|{\mathbf{R}_{es}}^{(j)}\|_F^2, \mathfrak{K}_{max})$
15. Construct the function $\wr_{opt}^{(j)}(\mathbf{a}_B, \mathbf{a}_M)$ by (21)
16. Search for new angle estimates $\hat{\mathbf{a}}_B^{(j+1)}, \hat{\mathbf{a}}_M^{(j+1)}$ by (16)
17. Estimate the new path gains $\hat{\mathbf{z}}^{(j+1)}$ by (20)
18. Prune path ℓ with $\mathbf{z}_{th}$ if $\hat{\mathbf{z}}^{(j+1)} < \mathbf{z}_{th}$
19. Until $j^{(last)}$ or $\|{\mathbf{R}_{es}}^{(j)}\|_F^2 < \varepsilon_{th}$
20. $\hat{\mathbf{a}}_B = \hat{\mathbf{a}}_B^{(last)}$, $\hat{\mathbf{a}}_M = \hat{\mathbf{a}}_M^{(last)}$, $\hat{\mathbf{z}} = \hat{\mathbf{z}}^{(last)}$

end for(in 8)

Output2:The estimated AoAs/AoDs and path gains.

$$\text{where } \text{人}^{(j+1)} = \begin{cases} \frac{\text{人}^{(j)}}{2}, & \wr_{opt}^{(j+1)} < \wr_{opt}^{(j)} \\ 2\text{人}^{(j)}, & \wr_{opt}^{(j+1)} > \wr_{opt}^{(j)} \end{cases},$$

$$\mathcal{H}^{(j)} \triangleq \nabla^2\wr_{opt}^{(j)} = \begin{bmatrix} \frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{M,\ell}{}^2} & \frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{M,\ell}\partial\mathbf{a}_{B,\ell}} \\ \frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{B,\ell}\partial\mathbf{a}_{M,\ell}} & \frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{B,\ell}{}^2} \end{bmatrix},$$

3.3. Proposed dual SVD and Marquardt's (DSM) optimization channel estimator

The proposed estimator captured in Algorithm 1 can be evaluated in two compact stages. The first stage, consisting of the initialization using pseudo-estimate $\hat{\mathbf{H}}_p$, obtains an estimated number of paths from input 1 to output 1. The second stage, consisting of the optimization, estimates the channel with output 1 as input 2 in an iterative re-weight paradigm to output 2.

4. Simulation results

The simulation set up of [34] is adopted as the benchmark for this work. This is consistent with items 1-9 of Table 1. The noise variance σ_n^2 is computed for several SNR using σ_z^2/σ_n^2. The path gain $z_\ell \sim \mathbb{C}\mathcal{N}(0,\sigma_z^2)$ which is a function of the propagation environment and the hardware impairment vector $(\mathbf{e}_M, \mathbf{e}_B)$ are assumed Gaussian as shown in section 2 for worst-case scenarios [43]. The choice of $k_B = 0.2$, from the range of values in [44], is deduced after some random trials that realize minimum NMSE. To reduce the computation time, only the main diagonal of the Hessian matrix is used. This implies that $\frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{B,\ell}\partial\mathbf{a}_{M,\ell}} = \frac{\partial^2\wr_{opt}^{(j)}}{\partial\mathbf{a}_{M,\ell}\partial\mathbf{a}_{B,\ell}} = 0$. The ULA and UPA configurations are considered under line-of-sight (LoS) and non-line-of-sight (NLoS) conditions.

Table 1
Simulation Parameters.

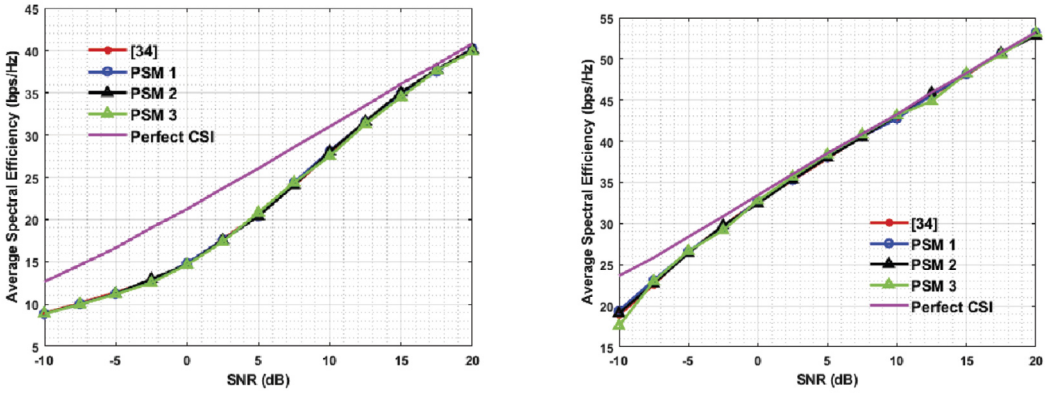
S/N	Parameters	Value
1	Number of RF chains (N_{RF})	4
2	Number of transceiver antennas (N_M/N_B)	64
3	Wavelength (λ)	1 m
4	Antenna spacing (d)	$\lambda/2$
5	Number of pilot symbols (N_s)	32
6	LoS K-factor	20 dB
7	Transmitter pilot signal ($\mathbf{S}$)	$s_{i,j} = (\rho_T/N_M)^{\frac{1}{2}} e^{-i\omega_{i,j}}$
8	Transmit power (ρ_T)	$\approx 1W$
9	Angular frequency ($\omega_{i,j}$)	$\in [0, 2\pi)$
10	Multi-path (L)	3
11	Signal-to-noise ratio (SNR)	$[-10:20]$
12	Transmitter proportionality coefficient (k_M)	0.175 [13]
13	Receiver proportionality coefficient (k_B)	0.2

Table 2 is the LoS/NLoS simulation time comparison between the arbitrary dimension reduction method [34] and the proposed systematic methods (PSM 1-3) under transmitter and receiver hardware impairment (TRHI). The PSM is tested with the gradient descent (GD) true-value search in PSM1 followed by not only predicting the sparsity but also the initial angle quantization in PSM2 for 500 Monte Carlo simulation (MCS). PSM2 offers the best time savings in ULA representing low computational complexity without compromising the spectral efficiency (SE) shown in Fig. 3a and 3b for ULA, that of UPA is similar. The result is better performing at high SNR compared to an orthogonal matching pursuit (OMP) based estimator [45]. Added to the advantage provided by the basis pursuit (BP) based estimators, it is observed that the spectral efficiency is highly degraded when any of the impairment is modeled as a function of the channel gain [45] and unchanged when modeled as a function of the signal. The Frobenius norm and vector norm are used for normalization in the ULA and UPA configurations respectively, to achieve the best NMSE performance in PSM 1-2. The reverse is the case for PSM3. The PSM3 combines sparsity prediction with Marquardt's (Ma_Q) search method ran over 200 Monte Carlo simulations (MCS). The PSM 1-2 of the UPA fails to provide the needed time savings on estimation. However, PSM3 provides a competitive solution. The overall complexity can be expressed as $\mathcal{O}(N_{S_x}N_{S_y}[N_B+N_M](N_cL)^2)$. Where $N_c = 1$ in [34], $0 < N_c < 1$ for PSM 1-2 and $1 < N_c < 2$ for PSM3 in ULA, while PSM 1-2 can be twice as complex as the others in UPA. From Table 3 analysis the number of SNRs tested is $N_{snr} = 13$ and the average of the mean predicted path (L_{ms}) for all SNRs per method can be obtained after simulation. Observed variations between Table 2 and Table 3 can be attributed to the state of the computing device.

Fig. 4a and 4b shows the NMSE performance for the four cases considered. PSM3 provides the best performance below 0 dB SNR in LoS/ULA pair and below −3.5 dB for NLoS/ULA pair respectively. However, PSM3 completely outperforms [34] and PSM 1-2 for almost all ranges of SNR considered as shown in Fig. 5a and 5b for both LoS/UPA and NLoS/UPA pairs. This is achieved with an affordable linear scale increase in complexity and the use of low mutual coherent signals. A general trend in Table 2 shows that the higher the run time the higher the likelihood of signal degradation at the higher SNR values tested. In Fig. 5a, the degraded NMSE performance for PSM2 at SNR = 20 dB is due to the estimation method which involves initial multipath and angle prediction compared to the others where only the initial multipath is predicted. In Fig. 5b, the degraded NMSE performance for PSM3 at SNR = 10 dB is due to the high sensitivity of the estimator meted by the impairment and as the noise level grows with an increase in signal strength.

Table 2
Total running time per method for antenna configuration and propagation environment.

Ref.	Method (L=3)	Pseudo estimate $\hat{H}_p$/initialization	LoS Time (sec)	NLoS Time (sec)
	ULA with Transmitter and Receiver Hardware Impairment			
[34]	GD	$[L1, L2] = [5, 10]$	73823.01	108195.74
PSM1	$Ma(L\_GD)$	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)})\ \&\ L2 = L1+1$	67058.46	86606.25
PSM2	LA_GD	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)}, \hat{\mathbf{a}}^{(0)})\ \&\ L2 = L1+1$	44420.70	70388.30
PSM3	$Ma\_Q\,(L\_M)$	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)})\ \&\ L2 = L1+1$	113741.92	198861.68
	UPA with Transmitter and Receiver Hardware Impairment			
[34]	GD	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)})\ \&\ L2 = L1+1$	171812.36	207259.25
PSM1	$Ma(L\_GD)$	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)})\ \&\ L2 = L1+1$	330238.10	261744.88
PSM2	LA_GD	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)}, \hat{\mathbf{a}}^{(0)})\ \&\ L2 = L1+1$	369581.32	310579.98
PSM3	$Ma\_Q\,(L\_M)$	$L1 = N_e : Algorithm\ 1-5\ (\hat{\mathbf{z}}^{(0)})\ \&\ L2 = L1+1$	176925.61	112546.22



(a) LoS Spectral Efficiency versus SNR for ULA (b) NLoS Spectral Efficiency versus SNR for ULA

Fig. 3. Spectral Efficiency plot for Uniform Linear Array (ULA).

Table 3
Computational Time Complexity Contrast between Proposed Methods.

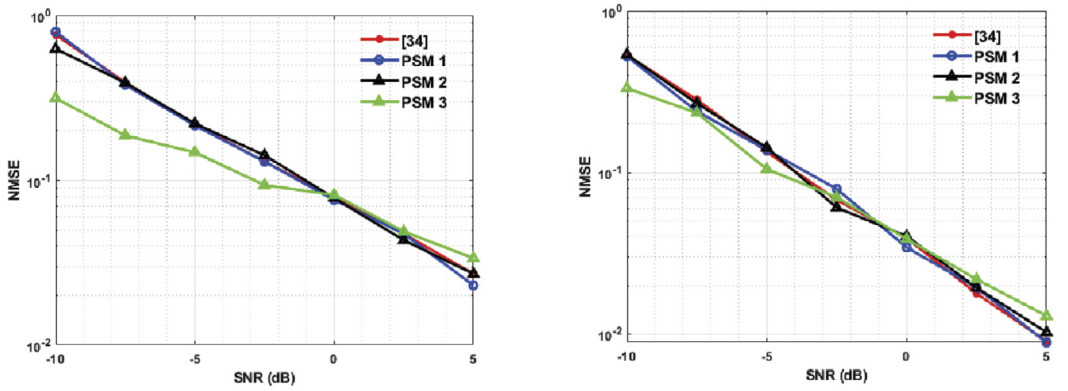
Ref.	Method (L=3)	Computational complexity	LoS	NLoS
	ULA: $(N_c L)^2 = (N_{snr} L_{ms}),\ N_{S_x}/N_{S_y} = 32,\ N_M/N_B = 64$		Numerical results	
[34]	GD	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 7,390,907,597$	$\approx 7,395,252,634$
PSM1	$Ma(L\_GD)$	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 6,552,315,494$	$\approx 6,859,023,974$
PSM2	LA_GD	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 6,687,778,406$	$\approx 7,125,093,581$
PSM3	$Ma\_Q\,(L\_M)$	$\approx \mathcal{O}(2N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 6,033,501,061$	$\approx 6,004,943,094$
	UPA: $(N_c L)^2 = (N_{snr} L_{ms}),\ N_{S_x}/N_{S_y} = 32,\ N_M/N_B = 64$		Numerical Results	
[34]	GD	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 7,449,778,586$	$\approx 7,426,519,859$
PSM1	$Ma(L\_GD)$	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 12,302,162,330$	$\approx 10,257,013,146$
PSM2	LA_GD	$\approx \mathcal{O}(N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 12,974,961,459$	$\approx 10,183,743,898$
PSM3	$Ma\_Q\,(L\_M)$	$\approx \mathcal{O}(2N_I N_{S_x} N_{S_y} [N_B + N_M] N_{snr} L_{ms})$	$\approx 4,334,813,184$	$\approx 3,211,033,313$

PSM3 has therefore traded off the algorithm complexity issues for a fewer number of iterations and accurate channel estimation.

Figs. 6a to 6d show the mean sparsity levels attained for both ULA and UPA at different SNR, and the respective LoS and NLoS condition. The lower the sparsity level the faster the estimation time representing lower computational complexity as described. This correlates with the run time recorded in Table 2.

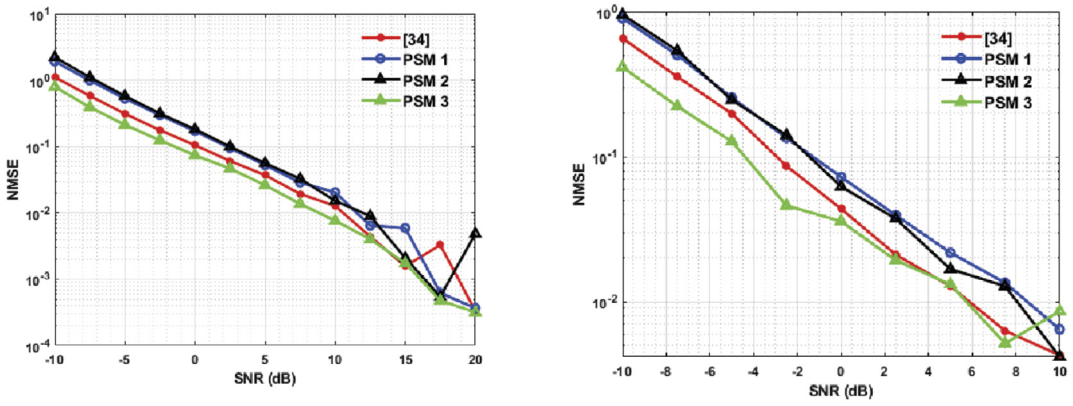
5. Conclusion

In this paper, three variants of iterative re-weight super-resolution channel estimation are suggested for low-power and high accuracy in mmWave massive MIMO systems with hybrid-precoding and transceiver hardware impairments. The hardware impairment introduces higher signal degradation in the system requiring a more sophisticated channel estimator to achieve the desired goal. A system model is first presented followed by a compressed sensing approach for channel estimation procedure in mm-wave massive MIMO. This involves dual-SVD for initialization and modified Marquardt's method of optimization. The proposed systematic method, under LoS/NLoS condition, has the best NMSE performance for ULA configuration at below $0/-3$ dB SNR and almost all SNR tested for UPA configuration. Finally, the effect of the level of impairments and that of the number of users on the NMSE would need to be investigated for the proposed system.



(a) LoS NMSE versus SNR for ULA (b) NLoS NMSE versus SNR for ULA

Fig. 4. Normalized Mean Square Error (NMSE) plot for ULA.



(a) LoS NMSE versus SNR for UPA (b) NLoS NMSE versus SNR for UPA

Fig. 5. Normalized Mean Square Error (NMSE) plot for UPA.

CRediT authorship contribution statement

Waheed Moses Audu: Investigation, Methodology, Writing – original draft, Writing – review & editing. **Olutayo Oyeyemi Oyerinde:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The financial assistance of the South Africa National Research Foundation (NRF) towards this research in terms of NRF Competitive Program (funding) for Rated Researchers with grant number 118547 and South Africa/NCBR, Poland joint science and technology research collaboration with grant number 118678 is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Appendix A. Derivation of Hessian matrix for real angle optimization

To obtain the estimate $\hat{\mathbf{H}}$ from equation (15) is the same as estimating $\mathbf{a}_B, \mathbf{a}_M$, and $\mathbf{z}$ for all L paths before substitution into equation (8). The complicated and computationally inefficient log-normal is replaced with a log-sum function based on [46]:

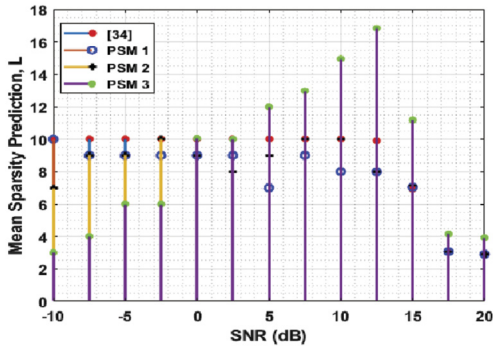
$$\begin{aligned} \underset{\mathbf{z},\mathbf{a}_M,\mathbf{a}_B}{\text{argmin}} \quad & F(z) \triangleq \sum_{\ell=1}^{L} \log\left(|z_\ell|^2 + \Delta\right) \\ \text{subject to} \quad & \|\mathbf{Y} - \mathbf{Q}^H \hat{\mathbf{H}}(\mathbf{S} + \mathbf{E}_M)\|_F \leq \varepsilon, \end{aligned} \tag{17}$$

The log-sum function is well defined for $\Delta > 0$. A regularization parameter $\mathfrak{K} > 0$ is added, for trade-off balance between data fitting and sparsity of the solution, to formulate the problem in (17) as an unconstrained optimization problem:

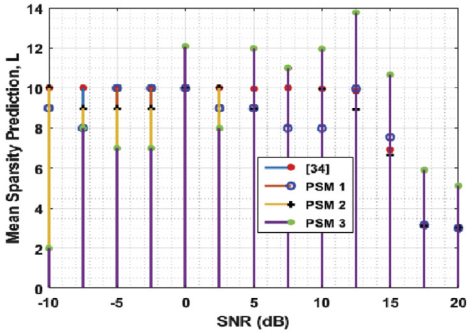
$$\begin{aligned} &\min_{\mathbf{z},\mathbf{a}_M,\mathbf{a}_B} G(\mathbf{z}, \mathbf{a}_M, \mathbf{a}_B) \\ &\triangleq \sum_{\ell=1}^{L} \log\left(|z_\ell|^2 + \Delta\right) + \mathfrak{K}\|\mathbf{Y} - \mathbf{Q}^H \hat{\mathbf{H}}(\mathbf{S} + \mathbf{E}_M)\|_F^2 \end{aligned} \tag{18}$$

The log-sum function is replaced with a surrogate function:

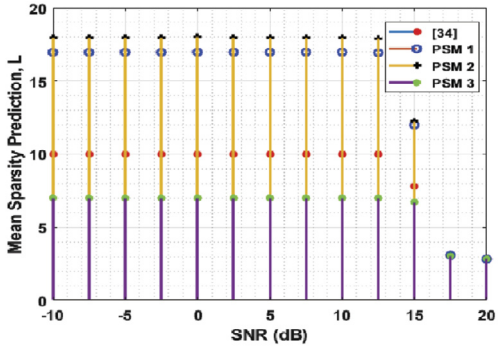
$$\min_{\mathbf{z},\mathbf{a}_M,\mathbf{a}_B} \wr^{(j)}(\mathbf{z}, \mathbf{a}_M, \mathbf{a}_B) \triangleq \mathfrak{K}^{-1} z^H \mathbf{D}_L^{(j)} z + \mathfrak{K}\|\mathbf{Y} - \mathbf{Q}^H \hat{\mathbf{H}}(\mathbf{S} + \mathbf{E}_M)\|_F^2, \tag{19}$$



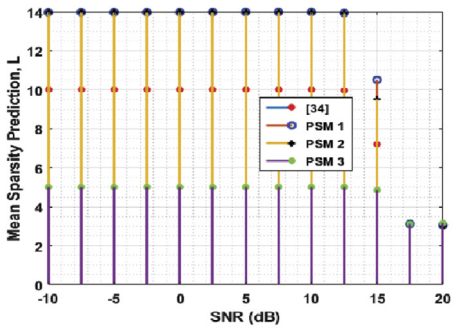
(a) LoS Mean multi-path (sparsity level, L) versus SNR for ULA



(b) NLoS Mean multi-path (sparsity level, L) versus SNR for ULA



(c) LoS Mean multi-path (sparsity level, L) versus SNR for UPA



(d) NLoS Mean multi-path (sparsity level, L) versus SNR for UPA

Fig. 6. (a)-(d) Mean multi-path prediction plot for ULA/UPA.

where $\mathbf{D}_L^{(j)} \triangleq diag(\frac{1}{|\hat{z}_1^{(j)}|^2+\Delta} \frac{1}{|\hat{z}_2^{(j)}|^2+\Delta} \dots \frac{1}{|\hat{z}_L^{(j)}|^2+\Delta})$ and $\hat{z}^{(j)}$ is the estimate of z at the j^{th} iteration. The optimal z that satisfies (19) is gotten by function expansion and equating it to zero to give:

$$\mathbf{z}_{opt}(\mathbf{a}_B, \mathbf{a}_M) = \frac{\{\sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_p^H) - \sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_q)\}}{\{\mathfrak{K}^{-1}\mathbf{D}_L + \sum_{q=1}^{N_s}(\mathbf{K}_p\mathbf{K}_p^H)\}} \tag{20}$$

This implies that $\wr_{opt}(\mathbf{z}, \mathbf{a}_B, \mathbf{a}_M)$ at $\mathbf{z} = \mathbf{z}_{opt}(\mathbf{a}_B, \mathbf{a}_M)$ can be expressed as

$$\begin{aligned}\wr_{opt}(\mathbf{a}_B, \mathbf{a}_M) = &-\frac{\{\sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{K}_p) - \sum_{q=1}^{N_s}(\mathbf{K}_p\mathbf{K}_q^H)\}}{\{\mathfrak{K}^{-1}\mathbf{D}_L + \sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_p)\}} \\ &\times \{\sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_p^H) - \sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_q)\} + \sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{y}_q) \\ &- \sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{K}_q) - \sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_q^H) + \sum_{q=1}^{N_s}(\mathbf{K}_q^H\mathbf{K}_q)\end{aligned} \tag{21}$$

Let $\wr_{opt}(\mathbf{a}_B, \mathbf{a}_M) = \wr_{opt}$, $\mathfrak{X} = \mathfrak{K}^{-1}\mathbf{D}_L + \sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_p)$, and $\mathfrak{A} = \{\sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_p^H) - \sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_q)\}$:

$$\begin{aligned}\wr_{opt} = &-\mathfrak{A}^H\mathfrak{X}^{-1}\mathfrak{A} + \sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{y}_q) - \sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{K}_q) - \sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_q^H) \\ &+ \sum_{q=1}^{N_s}(\mathbf{K}_q^H\mathbf{K}_q)\end{aligned} \tag{22}$$

The partial differential of $\wr_{opt}$ w.r.t. $\mathbf{a}_{B,\ell}$ using product rule:

$$\begin{aligned}\nabla_{\mathbf{a}_{B,\ell}}\wr_{opt} = &-(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A}^H)\mathfrak{X}^{-1}\mathfrak{A} - \mathfrak{A}^H(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1})\mathfrak{A} \\ &- \mathfrak{A}^H\mathfrak{X}^{-1}(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A}) + \nabla_{\mathbf{a}_{B,\ell}}\sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{y}_q) \\ &- \nabla_{\mathbf{a}_{B,\ell}}\sum_{q=1}^{N_s}(\mathbf{y}_q^H\mathbf{K}_q) - \nabla_{\mathbf{a}_{B,\ell}}\sum_{q=1}^{N_s}(\mathbf{y}_q\mathbf{K}_q^H) \\ &+ \nabla_{\mathbf{a}_{B,\ell}}\sum_{q=1}^{N_s}(\mathbf{K}_q^H\mathbf{K}_q)\end{aligned}$$

Only $\mathbf{K}_p = \mathbf{Q}^H\mathbf{A}_B diag(\mathbf{A}_M^H(\mathbf{S} + \mathbf{E}_M))$ and $\mathbf{K}_p^H = \mathbf{Q}\mathbf{A}_B^H diag(\mathbf{A}_M(\mathbf{S} + \mathbf{E}_M)^H)$ are functions of $\mathbf{a}_{B,\ell}$, causing partial differential of other terms to vanish.

$$\begin{aligned}\nabla_{\mathbf{a}_{B,\ell}}\wr_{opt} = &-(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A}^H)\mathfrak{X}^{-1}\mathfrak{A} - \mathfrak{A}^H(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1})\mathfrak{A} \\ &- \mathfrak{A}^H\mathfrak{X}^{-1}(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A})\end{aligned}$$

Let the second term on the right-hand side (R.H.S.) be $\mathfrak{A}^H(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1})\mathfrak{A} = -\mathfrak{A}^H\mathfrak{X}^{-2}[(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X})]\mathfrak{A}$ so that $\mathfrak{A}^H(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1})\mathfrak{A} = -\mathfrak{A}^H\mathfrak{X}^{-1}(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X})\mathfrak{X}^{-1}\mathfrak{A}$ (chain rule):

$$\nabla_{\mathbf{a}_{B,\ell}}\gamma_{opt} = -(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A}^H)\mathfrak{X}^{-1}\mathfrak{A} + \mathfrak{A}^H\mathfrak{X}^{-1}(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X})\mathfrak{X}^{-1}\mathfrak{A} - \mathfrak{A}^H\mathfrak{X}^{-1}(\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{A}), \tag{23}$$

where $\nabla_{\mathbf{a}_{B,\ell}}\mathfrak{X} \triangleq \frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}} = \frac{\partial\mathfrak{R}^{-1}\mathbf{D}_L}{\partial\mathbf{a}_{B,\ell}} + \frac{\partial\sum_{q=1}^{N_s}(\mathbf{K}_p^H\mathbf{K}_p)}{\partial\mathbf{a}_{B,\ell}}$, $\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}} = \sum_{q=1}^{N_s}\{\frac{\partial(\mathbf{K}_p^H)}{\partial\mathbf{a}_{B,\ell}}\mathbf{K}_p + \mathbf{K}_p^H\frac{\partial(\mathbf{K}_p)}{\partial\mathbf{a}_{B,\ell}}\}$, and

$$\frac{\partial\mathbf{K}_p}{\partial\mathbf{a}_{B,\ell}} \triangleq tr\{(\frac{\partial\mathbf{K}_p}{\partial\mathbf{A}_B})^T\frac{\partial\mathbf{A}_B}{\partial\mathbf{a}_{B,\ell}}\} + tr\{(\frac{\partial\mathbf{K}_p}{\partial\mathbf{A}_B^H})^T\frac{\partial\mathbf{A}_B^H}{\partial\mathbf{a}_{B,\ell}}\}[46]:$$

$$\frac{\partial(\mathbf{K}_p^H)}{\partial\mathbf{a}_{B,\ell}} = \frac{\partial(\mathbf{Q}\mathbf{A}_B^H diag(\mathbf{A}_M(\mathbf{S}+\mathbf{E}_M)^H))}{\partial\mathbf{a}_{B,\ell}} \text{ is also}$$

$$\frac{\partial(\mathbf{K}_p^H)}{\partial\mathbf{a}_{B,\ell}} = (\mathbf{Q}\frac{\partial\mathbf{A}_B^H}{\partial\mathbf{a}_{B,\ell}}diag(\mathbf{A}_M(\mathbf{S}+\mathbf{E}_M)^H)),$$

the ℓth non-zero path is simply

$$\frac{\partial(\mathbf{K}_p)}{\partial\mathbf{a}_{B,\ell}} = [0\ldots0\ \mathbf{Q}^H\frac{\partial\mathbf{a}_B(\hat{\mathbf{a}}_{B,\ell})}{\partial\mathbf{a}_{B,\ell}}\mathbf{a}_M^H(\hat{\mathbf{a}}_{M,\ell})(\mathbf{s}_q+\mathbf{e}_M)\ 0\ldots0].$$

Recall $\mathbf{A}_B = \mathbf{a}_B(\hat{\mathbf{a}}_{B,\ell})$, $\frac{\partial\mathbf{A}_B^H}{\partial\mathbf{a}_{B,\ell}} = (\frac{\partial\mathbf{A}_B}{\partial\mathbf{a}_{B,\ell}})^H$, and applies to $\mathbf{A}_M = \mathbf{a}_M(\hat{\mathbf{a}}_{M,\ell})$, $\frac{\partial\mathbf{A}_M^H}{\partial\mathbf{a}_{M,\ell}} = (\frac{\partial\mathbf{A}_M}{\partial\mathbf{a}_{M,\ell}})^H$

$$\begin{aligned}\frac{\partial^2\gamma_{opt}}{\partial\mathbf{a}_{B,\ell}{}^2} &= -\frac{\partial^2\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}{}^2}\mathfrak{X}^{-1}\mathfrak{A} + \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &- \frac{\partial^H\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}} + \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &- \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &+ \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial^2\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}{}^2}\mathfrak{X}^{-1}\mathfrak{A} - \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &+ \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}} - \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}}\\ &+ \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}} - \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial^2\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}{}^2}\end{aligned} \tag{24}$$

The first and second partial derivatives with respect to the AoA and AoD:

$$\begin{aligned}\frac{\partial^2\gamma_{opt}}{\partial\mathbf{a}_{M,\ell}\partial\mathbf{a}_{B,\ell}} &= -\frac{\partial}{\partial\mathbf{a}_{M,\ell}}(\frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}})\mathfrak{X}^{-1}\mathfrak{A} + \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &- \frac{\partial^H\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{M,\ell}} + \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &- \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &+ \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial}{\partial\mathbf{a}_{M,\ell}}(\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}})\mathfrak{X}^{-1}\mathfrak{A}\\ &- \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\mathfrak{A}\\ &- \mathfrak{A}^H\mathfrak{X}^{-1}\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{B,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}} - \frac{\partial\mathfrak{A}^H}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}}\\ &+ \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial\mathfrak{X}}{\partial\mathbf{a}_{M,\ell}}\mathfrak{X}^{-1}\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}}\\ &- \mathfrak{A}^H\mathfrak{X}^{-1}\frac{\partial}{\partial\mathbf{a}_{M,\ell}}(\frac{\partial\mathfrak{A}}{\partial\mathbf{a}_{B,\ell}})\end{aligned} \tag{25}$$

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dsp.2021.103308.

References

[1] A.L. Swindlehurst, E. Ayanoglu, P. Heydari, F. Capolino, Millimeter-wave massive MIMO: the next wireless revolution?, IEEE Commun. Mag. 52 (2014) 56–62, https://doi.org/10.1109/MCOM.2014.6894453.

[2] L. Lu, G. Ye Li, A.L. Swindlehurst, R. Zhang, An overview of massive MIMO: benefits and challenges, IEEE J. Sel. Top. Signal Process. 8 (2014) 742–758, https://doi.org/10.1109/JSTSP.2014.2317671.

[3] S. Han, Z. Xu, C. Rowell, Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G, IEEE Commun. Mag. 53 (2015) 186–194, https://doi.org/10.1109/MCOM.2015.7010533.

[4] A. Alkhateeb, O. El Ayach, G. Leus, R.W. Heath Jr., Channel estimation and hybrid precoding for millimeter wave cellular systems, IEEE J. Sel. Top. Signal Process. 8 (2014) 831–846, https://doi.org/10.1109/JSTSP.2014.2334278.

[5] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, R.W. Heath Jr., Spatially sparse precoding in millimeter wave MIMO systems, IEEE Trans. Wirel. Commun. 13 (2014) 1499–1513, https://doi.org/10.1109/TWC.2014.011714.130846.

[6] E. Björnson, M. Matthaiou, M. Debbah, Massive MIMO with non-ideal arbitrary arrays: hardware scaling laws and circuit-aware design, IEEE Trans. Wirel. Commun. 14 (2015) 4353–4368, https://doi.org/10.1109/TWC.2015.2420095.

[7] R. Gomes, A. Hammoudeh, R.F. Caldeirinha, Z. Al-Daher, T. Fernandes, J. Reis, Towards 5G: performance evaluation of 60 GHz UWB OFDM communications under both channel and RF impairments, Phys. Commun. 25 (2017) 527–538, https://doi.org/10.1016/j.phycom.2017.10.011.

[8] X. Lu, W. Yang, Y. Cai, X. Guan, Comparison of CS-based channel estimation for millimeter wave massive MIMO systems, Appl. Sci. 9 (2019) 4346, https://doi.org/10.3390/app9204346.

[9] Y. Wu, Y. Gu, Z. Wang, Channel estimation for mmWave MIMO with transmitter hardware impairments, IEEE Commun. Lett. 22 (2018) 320–323, https://doi.org/10.1109/LCOMM.2017.2769657.

[10] F. Ding, H. Wang, S. Zhang, M. Dai, Impact of residual hardware impairments on non-orthogonal multiple access based amplify-and-forward relaying networks, IEEE Access 6 (2018) 15117–15131, https://doi.org/10.1109/ACCESS.2018.2813081.

[11] C. Deng, X. Zhao, D. Zhang, X. Li, J. Li, C.C. Cavalcante, Performance analysis of NOMA-based relaying networks with transceiver hardware impairments, KSII Trans. Int. Inf. Syst. 12 (2018) 4295–4316, https://doi.org/10.3837/tiis.2018.09.010.

[12] X. Li, J. Li, P. Takis Mathiopoulos, D. Zhang, L. Li, J. Jin, Joint impact of hardware impairments and imperfect CSI on cooperative SWIPT NOMA multi-relaying systems, in: 2018 IEEE/CIC International Conference on Communications in China (ICCC), 2018, pp. 95–99.

[13] C.B. Le, D.T. Do, M. Voznak, Exploiting impact of hardware impairments in NOMA: adaptive transmission mode in FD/HD and application in Internet-of-things, Sensors 19 (2019) 1293, https://doi.org/10.3390/s19061293.

[14] V.P. Tuan, H.Y. Kong, Impact of residual transmit RF impairments on energy harvesting relay selection systems, Int. J. Electron. 104 (2017) 928–941, https://doi.org/10.1080/00207217.2016.1253788.

[15] M. Matthaiou, A. Papadogiannis, E. Björnson, M. Debbah, Two-way relaying under the presence of relay transceiver hardware impairments, IEEE Commun. Lett. 17 (2013) 1136–1139, https://doi.org/10.1109/LCOMM.2013.042313.130191.

[16] T.C.W. Schenk, E.R. Fledderus, Rf impairments in high-rate wireless systems – understanding the impact of TX/RX-asymmetry, in: 2008 3rd International Symposium on Communications, Control and Signal Processing, 2008, pp. 117–122.

[17] T. Schenk, RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation, Springer, Netherlands, 2008.

[18] D. Verenzuela, E. Björnson, M. Matthaiou, Hardware design and optimal ADC resolution for uplink massive MIMO systems, in: 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2016, pp. 1–5.

[19] R.W. Heath Jr., N. Gonzalez-prelcic, S. Rangan, W. Roh, A. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems, IEEE J. Sel. Top. Signal Process. 10 (2016) 436–453, https://doi.org/10.1109/JSTSP.2016.2523924.

[20] G. Hueber, A.M. Niknejad, A. Pärssinen, A. Puglielli, G. LaCaille, E. Alon, B. Nikolić, H. Krishnaswamy, T. Dinc, L.A. Calderin, S. Ramakrishnan, A. Natarajan, P. Wambacq, D. Guermandi, A. Bourdoux, J. Craninckx, K. Okada, R. Wu, B. Sadhu, L. Rexberg, P. Heydari, W. Wu, R.B. Staszeweski, M. Mikhemar, Y. Zhang, P. Reynaert, H.-J. Lee, B. Sell, Millimeter-Wave Circuits for 5G and Radar, first ed., Cambridge University Press, UK, 2019.

[21] E. Björnson, J. Hoydis, M. Kountouris, M. Debbah, Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits, IEEE Trans. Inf. Theory 60 (2014) 7112–7139, https://doi.org/10.1109/TIT.2014.2354403.

[22] L. Smaini, RF Analog Impairments Modeling for Communication Systems Simulation: Application to OFDM-Based Transcievers, first ed., John Wiley & Sons, Ltd, UK, 2012.
[23] A. Balatsoukas-Stimming, A.C.M. Austin, P. Belanovic, A. Burg, Baseband and RF hardware impairments in full-duplex wireless systems: experimental characterisation and suppression, https://doi.org/10.1186/s13638-015-0350-1, arXiv: 1412.4542, 2015.
[24] J. Samuel, P. Rosson, L. Maret, C. Dehos, A. Valkanas, Impact of RF impairments in cellular wireless metropolitan area networks, in: 2008 IEEE International Symposium on Spread Spectrum Techniques and Applications, 2008, pp. 766–769.
[25] P. Robertson, S. Kaiser, Analysis of the effects of phase-noise in orthogonal frequency division multiplex (OFDM) systems, in: Proceedings IEEE International Conference on Communications ICC '95, vol. 3, 1995, pp. 1652–1657.
[26] S. Suyama, Y. Hashimoto, H. Suzuki, K. Fukawa, 60 GHz OFDM experimental system employing decision-directed phase noise compensation, in: 2012 IEEE Radio and Wireless Symposium, 2012, pp. 191–194.
[27] C. Zhang, Z. Xiao, B. Gao, D. Su, Li Jin, Robust IQ imbalance estimation and compensation via specific preamble for 60 GHz systems, in: 2013 IEEE Wireless Communications and Networking Conference (WCNC), 2013, pp. 4134–4139.
[28] L. Fan, Y. Li, M. Zhao, Joint IQ imbalance and PA nonlinearity pre-distortion for highly integrated millimeter-wave transmitters, in: 2014 IEEE Globecom Workshops (GC Wkshps), 2014, pp. 399–404.
[29] M. Abdelghany, A.A. Farid, U. Madhow, M.J.W. Rodwell, Towards all-digital mmwave massive MIMO: designing around nonlinearities, in: 2018 52nd Asilomar Conference on Signals, Systems, and Computers, 2018, pp. 1552–1557.
[30] M. Wenk, MIMO-OFDM-testbed: challenges, implementations, and measurement results, Ph.D. thesis, ETH, Zurich, Switzerland, 2010.
[31] C. Studer, M. Wenk, A. Burg, MIMO transmission with residual transmit-RF impairments, in: 2010 International ITG Workshop on Smart Antennas, WSA, 2010, pp. 189–196.
[32] X. Zhang, M. Matthaiou, M. Coldrey, E. Björnson, Impact of residual transmit RF impairments on training-based MIMO systems, IEEE Trans. Commun. 63 (2015) 2899–2911, https://doi.org/10.1109/TCOMM.2015.2432761.
[33] F. Athley, G. Durisi, U. Gustavsson, Analysis of massive MIMO with hardware impairments and different channel models, in: 2015 9th European Conference on Antennas and Propagation (EuCAP), 2015, pp. 1–5.
[34] C. Hu, L. Dai, T. Mir, Z. Gao, J. Fang, Super-resolution channel estimation for MmWave massive MIMO with hybrid precoding, IEEE Trans. Veh. Technol. 67 (2018) 8954–8958, https://doi.org/10.1109/TVT.2018.2842724.
[35] B. Selim, S. Muhaidat, P.C. Sofotasios, A. Al-Dweik, B.S. Sharif, T. Stouraitis, Radio-frequency front-end impairments: performance degradation in nonorthogonal multiple access communication systems, IEEE Veh. Technol. Mag. 14 (2019) 89–97, https://doi.org/10.1109/MVT.2018.2867646.
[36] F. Hu, L. Li, Y. Lu, L. Jin, Super-resolution channel estimation for MmWave massive MIMO: a Newtonian method, in: 2019 IEEE 19th International Conference on Communication Technology (ICCT), 2019, pp. 75–80.
[37] C.-Y. Chi, C.-C. Feng, C.-H. Chen, C.-Y. Chen, Blind Equalization and System Identification Batch Processing Algorithms, Performance and Applications, Springer-Verlag, Germany, 2006.
[38] J.Y. Fan, Y.-X. Yuan, On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption, Computing 74 (2005) 23–39, https://doi.org/10.1007/s00607-004-0083-1.
[39] G.R. Liu, X. Han, Computational Inverse Techniques in Nondestructive Evaluation, CRC Press, USA, 2003.
[40] K. Deb, Optimization for Engineering Design: Algorithms and Examples, second ed., PHI Learning Private, Limited, New Delhi, 2012.
[41] J. Lee, G.-T. Gil, Y.H. Lee, Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications, IEEE Trans. Commun. 64 (2016) 2370–2386, https://doi.org/10.1109/TCOMM.2016.2557791.
[42] L.J. Stanković, T. Thayaparan, M. Daković, Algorithm for signal decomposition by using the S-method, in: 13th European Signal Processing Conference, EUSIPCO 2005, 2005, pp. 357–360.
[43] M. Viswanathan, Wireless Communication Systems in MATLAB, second ed., Viswanathan, Mathuranathan, Singapore, 2020.
[44] N. Melatic, J. Gutierrez-Teran, E. Grass, Beamforming mmwave mimo: impact of nonideal hardware and channel state information, in: Proc. Telecommunication Forum (TELFOR), 2018, pp. 1–6.
[45] A.K. Papazafeiropoulos, G.K. Papageorgiou, O.Y. Kolawole, P. Kourtessis, S. Chatzinotas, J.M. Senior, M. Sellathurai, T. Ratnarajah, Towards the assessment of realistic hybrid precoding in millimeter wave mimo systems with hardware impairments, IET Commun. 15 (2021) 1606–1619, https://doi.org/10.1049/cmu2.12173.
[46] J. Fang, F. Wang, Y. Shen, H. Li, R.S. Blum, Super-resolution compressed sensing for line spectral estimation: an iterative reweighted approach, IEEE Trans. Signal Process. 64 (2016) 4649–4662, https://doi.org/10.1109/TSP.2016.2572041.

Waheed Moses Audu received B. Eng. degree in Electrical/Computer Engineering and an M. Eng. degree in Communication Engineering both from the Federal University of Technology (FUT), Minna, Niger State, Nigeria in 2008 and 2014 respectively. Since 2010, he has been working as an academic at the same university. He is currently a Ph.D. candidate in the School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa. He is a member of the Center for Telecommunications Access and Services (CeTAS) research group with a special interest in channel estimation and multiuser detection techniques for new generation wireless communication systems.

Olutayo Oyeyemi Oyerinde received Ph.D. degree in electronic engineering from the University of Kwazulu-Natal, Durban, South Africa, in 2011. Since 2013, he has been with the School of Electrical and Information Engineering, University of the Witwatersrand at Johannesburg, South Africa, where he is currently an Associate Professor. He is also a National Research Foundation (NRF) C2-Rated Scientist, a Registered Professional Engineer (Pr.Eng.) with the Engineering Council of South Africa (ECSA), a Registered Engineer (R.Eng) with COREN, Senior Member of Institute of Electrical and Electronics Engineer (SMIEEE), and a Corporate Member of NSE among others. He is an Associate Editor of the IEEE ACCESS journal and an Editorial Board Member of the International Journal of Sensors, Wireless Communications and Control. His current research interests include wireless communications with specific interests in 5G and beyond 5G technologies, Reconfigurable Intelligent Surface for 6G Networks, and signal processing techniques for wireless communication systems.