Optimal Training Design for Over-the-Air Polynomial Power Amplifier Model Estimation

*François Rottenberg, *Thomas Feys, †Nuutti Tervo
*KU Leuven, ESAT-WaveCore, Ghent Technology Campus, 9000 Ghent, Belgium
†Centre for Wireless Communications - Radio Technologies (CWC-RT), University of Oulu, Oulu FI-90014, Finland

***Abstract*—The current evolution towards a massive number of antennas and a large variety of transceiver architectures forces to revisit the conventional techniques used to improve the fundamental power amplifier (PA) linearity-efficiency trade-off. Most of the digital linearization techniques rely on PA measurements using a dedicated feedback receiver. However, in modern systems with large amount of RF chains and high carrier frequency, dedicated receiver per RF chain is costly and complex to implement. This issue can be addressed by measuring PAs over the air, but in that case, this extra signalling is sharing resources with the actual data transmission. In this paper, we look at the problem from an estimation theory point of view so as to minimize pilot overhead while optimizing estimation performance. We show that conventional results in the mathematical statistics community can be used. We find the least squares (LS) optimal training design, minimizing the maximal mean squared error (MSE) of the reconstructed PA response over its whole input range. As compared to uniform training, simulations demonstrate a factor 10 reduction of the maximal MSE for a $L = 7$ PA polynomial order. Using prior information, the LMMSE estimator can achieve an additional gain of a factor up to 300 at low signal-to-noise ratio (SNR).**

***Index Terms*—Optimal training, power amplifier, calibration.**

I. Introduction

The power amplifiers (PAs) account for a large part of energy consumption of base stations [1]. Their operating regime faces a fundamental trade-off between linearity and efficiency [2]. While linearity is desirable to avoid signal degradation and out-of-band emission, driving the PA closer to saturation improves its efficiency while the signal suffers from more nonlinearity. This is particularly challenging in latest broadband technologies where the combination of multicarrier transmission and precoding over to multiple users, leads to a high peak-to-average power ratio, requiring a large linear range of the PA [3].

A. State-of-the-Art

A widely used solution to improve the linearity-efficiency trade-off is the use of digital pre-distortion (DPD) before the PA to linearize its response and allow to drive it closer to saturation [4]. A recently proposed solution is the use of distortion-aware precoding such as the Z3RO precoder [5], [6]. These techniques require the PA response to be properly estimated. A conventional approach is to use a feedback loop after the PA within the transmitter chain [7]. Unfortunately, in massive MIMO and/or mm-wave systems, this is not always possible or desirable. Replicating such a feedback mechanism per RF chain is expensive in terms of hardware. An interesting alternative is to train the DPD using the over-the-air (OTA) wireless link with either near-field probes [8], far-field observation antennas or the entire radio link [9], [10]. As this training signalling uses the wireless channel, it is important to minimize the signalling overhead to avoid penalizing data transmission.

To the best of our knowledge, despite this rich literature on array linearization, there is a lack of an estimation-theoretical approach of optimal training design to estimate PA model parameters. Moreover, when using a dedicated feedback loop directly after the PA, observations are typically obtained continuously at a very high signal-to-noise ratio (SNR) and noise is generally neglected [7]. In contrast, OTA training suffers from channel attenuation, multipath propagation, and limited SNR due to the lower received signal strength. Hence, the training design should be optimized to maximize performance while minimizing the signalling overhead.

B. Contributions

In this paper, we formalize the noisy OTA estimation problem and rederive the least squares (LS) and linear minimum mean squared error (LMMSE) estimators for a nonlinear memoryless polynomial PA model. We derive the LS optimal training in the sense of minimizing the generalized variance (determinant) of the error covariance matrix. To do this, we show that the problem can be viewed as a D-optimal design, well known in mathematical statistics. Interestingly, the optimal design also minimizes the maximal reconstruction mean squared error (MSE) when predicting the PA response over the entire PA input range. The optimal training design can also be used for the LMMSE estimator but is only optimal at high SNR. Simulation results demonstrate the superiority of the optimal design for the LS estimator. We also show, that if enough prior information is available on the PA response, the LMMSE estimator can provide a significant gain at low SNR, especially if the phase of the channel is known (coherent case).

Notations: Vectors and matrices are by bold lowercase and uppercase letters $\mathbf{a}$ and $\mathbf{A}$. Superscripts *, ^{T} and ^{H} stand for conjugate, transpose and Hermitian transpose. The symbols $\mathbb{E}(.)$, $\angle(.)$, $\|\mathbf{A}\|$ and $|\mathbf{A}|$ denote the expectation, phase, Frobenius norm and determinant. $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2.

II. Power Amplifier Model Estimation

Without loss of generality, we consider a single PA related to a transmit antenna, which response needs to be estimated. To do this, over-the-air calibration is considered based on an observational receiver and a feedback loop. The framework can be straightforwardly extended to multiple transmit antennas by using orthogonal transmit sequences and decoupling the

problem per PA. A series of N pilots are sent, denoted by $s_n,\ n = 0, ..., N-1$. The channel is assumed narrowband and constant during training. The received signal is then

$$r_n = hf(s_n) + w_n$$

where w_n is additive noise and h is the complex channel coefficient. The function $f()$ is the PA transfer function to be estimated. It is here considered to follow a nonlinear quasi-memoryless polynomial model

$$f(s) = \sum_{l=1}^{L} \beta_l s|s|^{l-1}$$

where L is the model order and $\beta_l \in \mathbb{C}$ are the polynomial coefficients to be estimated. One can note the presence of even-order nonlinear terms, which are often neglected when considering bandpass signals. Still, motivated by [11], we include them to improve both modelling accuracy while using low order polynomials, with better numerical properties and also making the following optimization problem more standard. Without further assumptions, the problem has $L+1$ complex unknowns including the L beta coefficients plus the channel coefficient h. In practice, channel estimation will be used at the receiver so that the equivalent linear channel gain $h\beta_1$ will be estimated. We are interested in the relation of higher order terms with respect to this one. To solve this ambiguity, we set $h = 1$, which we will rediscuss in Section IV. We thus have

$$r_n = \sum_{l=1}^{L} \beta_l s_n|s_n|^{l-1} + w_n.$$

In vector form, this gives a linear observation model

$$\mathbf{r} = \mathbf{\Phi}\boldsymbol{\beta} + \mathbf{w}$$

where $\mathbf{r} = (r_0, ..., r_{N-1})^T$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_L)^T \in \mathbb{C}^{L\times 1}$, $\mathbf{w} = (w_0, ..., w_{N-1})^T$ and

$$\mathbf{\Phi} = \begin{pmatrix} s_0 & \dots & s_0|s_0|^{L-1} \\ \vdots & \ddots & \vdots \\ s_{N-1} & \dots & s_{N-1}|s_{N-1}|^{L-1} \end{pmatrix} \in \mathbb{C}^{N\times L}.$$

In the literature, other polynomial bases have been considered, e.g., orthogonal bases for given distributions of the input signals [12]. Generally considering another basis $\mathbf{\Psi} \in \mathbb{C}^{N\times L}$, we can write $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{U}$ where $\mathbf{U} \in \mathbb{C}^{L\times L}$ is a full rank matrix. We would then have $\mathbf{r} = \mathbf{\Psi}\boldsymbol{\alpha} + \mathbf{w}$ where $\boldsymbol{\alpha} = \mathbf{U}\boldsymbol{\beta}$.

A. Least Squares Estimator

The classical LS estimate is given by [7], [13]

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{r} - \mathbf{\Phi}\boldsymbol{\beta}\|^2 = (\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\mathbf{\Phi}^H\mathbf{r}.$$

We here consider the noise samples w_n identically and independently complex circularly symmetric distributed with zero mean and variance σ^2 so that the LS estimate corresponds to the maximum likelihood (ML) estimate [13]. The error covariance matrix is

$$\mathbf{C}_{\mathrm{LS}} = \sigma^2(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}.$$

We can use the LS estimate to predict, *i.e.*, reconstruct, the PA response at $\tilde{N}$ values $\tilde{s}_n, n = 0, ..., \tilde{N}-1$ as $\tilde{\mathbf{\Phi}}\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ with

$$\tilde{\mathbf{\Phi}} = \begin{pmatrix} \tilde{s}_0 & \dots & \tilde{s}_0|\tilde{s}_0|^{L-1} \\ \vdots & \ddots & \vdots \\ \tilde{s}_{\tilde{N}-1} & \dots & \tilde{s}_{\tilde{N}-1}|\tilde{s}_{\tilde{N}-1}|^{L-1} \end{pmatrix} \in \mathbb{C}^{\tilde{N}\times L}.$$

The resulting estimate is unbiased with error covariance matrix

$$\tilde{\mathbf{C}}_{\mathrm{LS}} = \sigma^2\tilde{\mathbf{\Phi}}(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\tilde{\mathbf{\Phi}}^H. \quad (1)$$

B. Linear Minimum Mean Squared Estimator

How to choose the polynomial order L? Intuitively, increasing it would only help as it would allow a better fit. Unfortunately, this is not the case for the conventional LS estimator due to three reasons. i) It may suffer from numerical instability if a high polynomial order L is considered [7], [12]. In other words, the matrix inverse $(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}$ becomes ill conditioned. ii) As will be shown in Corol. 1, the prediction MSE grows with L. Intuitively, more coefficients need to be estimated leading to less noise averaging. iii) Moreover, the minimum number of pilots N has to grow with L, since the matrix $\mathbf{\Phi}$ needs at least L pilots with different magnitude to be full rank so that matrix $\mathbf{\Phi}^H\mathbf{\Phi}$ can be inverted.

This requirement of $N \geq L$ pilots relates to the fact that the LS estimator gives the same importance to all polynomial coefficients β_l. In practice though, some, especially higher order ones, may be negligible. This knowledge can help to regularize the problem inversion (even if $N \leq L$). To do this, we leverage the prior information we have about the PA transfer function. Indeed, while its exact form is not known, its general behaviour can be generally assumed to be known, *e.g.*, its linear gain or saturation level. A LMMSE provides an attractive solution, which only requires the first and second-order statistical moments. We now assume that the vector $\boldsymbol{\beta}$ is random with mean $\bar{\boldsymbol{\beta}}$ and covariance matrix $\mathbf{C}_{\boldsymbol{\beta}}$. In Section IV, we discuss two cases of prior information depending if the phase of the channel gain h is known (coherent) or unknown (noncoherent). The LMMSE estimator is then

$$\hat{\boldsymbol{\beta}}_{\mathrm{LMMSE}} = \bar{\boldsymbol{\beta}} + (\mathbf{\Phi}^H\mathbf{\Phi} + \sigma^2\mathbf{C}_{\boldsymbol{\beta}}^{-1})^{-1}\mathbf{\Phi}^H(\mathbf{r} - \mathbf{\Phi}\bar{\boldsymbol{\beta}})$$

and the error covariance matrix is

$$\mathbf{C}_{\mathrm{LMMSE}} = \sigma^2(\mathbf{\Phi}^H\mathbf{\Phi} + \sigma^2\mathbf{C}_{\boldsymbol{\beta}}^{-1})^{-1}.$$

Note that the matrix to invert is always full rank, even if $N \leq L$. Moreover, given the positive definite nature of $\mathbf{C}_{\boldsymbol{\beta}}$ we have $\mathbf{C}_{\mathrm{LS}} \succeq \mathbf{C}_{\mathrm{LMMSE}}$, which implies better performance of the LMMSE estimator. Again, we can predict the PA response at $\tilde{N}$ values by $\tilde{\mathbf{\Phi}}\hat{\boldsymbol{\beta}}_{\mathrm{LMMSE}}$ with an error covariance matrix

$$\tilde{\mathbf{C}}_{\mathrm{LMMSE}} = \sigma^2\tilde{\mathbf{\Phi}}(\mathbf{\Phi}^H\mathbf{\Phi} + \sigma^2\mathbf{C}_{\boldsymbol{\beta}}^{-1})^{-1}\tilde{\mathbf{\Phi}}^H.$$

Proposition 1. The basis choice has no impact on the prediction performance related to the LS and LMMSE estimators, *i.e.*, $\tilde{\mathbf{C}}_{\mathrm{LS}}$ and $\tilde{\mathbf{C}}_{\mathrm{LMMSE}}$ do not depend on the basis $\mathbf{\Psi}$.

Proof. We here conduct the proof in the LMMSE case. The LS case can be found by particularization to the case $\mathbf{C}_{\boldsymbol{\beta}}^{-1} =$

$\mathbf{0}$ (infinite variance of a priori information). We can redo the previous derivation for a general basis $\mathbf{\Psi}$. We then have $\mathbf{\Psi} = \mathbf{\Phi U}$ and $\tilde{\mathbf{\Psi}} = \tilde{\mathbf{\Phi}}\mathbf{U}$. Since $\boldsymbol{\alpha} = \mathbf{U}\boldsymbol{\beta}$, the random vector $\boldsymbol{\alpha}$ has a covariance matrix $\mathbf{C}_{\boldsymbol{\alpha}} = \mathbf{U}\mathbf{C}_{\boldsymbol{\beta}}\mathbf{U}^H$. The error covariance matrix then becomes

$$\begin{aligned}\tilde{\mathbf{C}}_{\text{LMMSE}} &= \sigma^2\tilde{\mathbf{\Psi}}(\mathbf{\Psi}^H\mathbf{\Psi} + \sigma^2\mathbf{C}_{\boldsymbol{\alpha}}^{-1})^{-1}\tilde{\mathbf{\Psi}}^H \\ &= \sigma^2\tilde{\mathbf{\Phi}}\mathbf{U}(\mathbf{U}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{U} + \sigma^2\mathbf{C}_{\boldsymbol{\alpha}}^{-1})^{-1}\mathbf{U}^H\tilde{\mathbf{\Phi}}^H \\ &= \sigma^2\tilde{\mathbf{\Phi}}(\mathbf{\Phi}^H\mathbf{\Phi} + \sigma^2\mathbf{C}_{\boldsymbol{\beta}}^{-1})^{-1}\tilde{\mathbf{\Phi}}^H\end{aligned}$$

where we used the fact that the square matrix $\mathbf{U}$ is full rank and is thus invertible. This completes the proof. □

III. Optimal Training Design

We now consider the optimal design of the pilot symbols s_n, $n = 0, ..., N-1$ or equivalently the matrix $\mathbf{\Phi}$. As a practical constraint, we consider that the PA input power can be limited by a max level. Therefore, we add the constraint $|s_n| \leq 1$, *i.e.*, a unit power constraint which can be generalized straightforwardly to any power $P_{\max}$ by scaling training points s_n by $\sqrt{P_{\max}}$.

In theoretical infinite precision, Prop. 1 shows that all bases give same performance. Hence, the basis choice has no impact on the optimal training design minimizing $\tilde{\mathbf{C}}_{\text{LS/LMMSE}}$. In practice, when implementing the estimators in finite precision, some bases will perform better and the conventional $\mathbf{\Phi}$ should be avoided. Indeed, its high order terms $|s|^l$ go to zero fast (as $O(|s|^l)$) as $|s| \to 0$ while other bases can keep a decay in $O(|s|)$, giving them increased robustness (reduced impact of quantization error) [12]. We now give a lemma that will prove useful for optimal design.

Lemma 1. The phases of the pilot signals $\angle s_n$ have no impact on the error covariance matrices $\mathbf{C}_{\text{LS/LMMSE}}$ and $\tilde{\mathbf{C}}_{\text{LS/LMMSE}}$. Hence, they can be chosen arbitrarily.

Proof. The error covariance matrices depend on the matrix

$$\mathbf{\Phi}^H\mathbf{\Phi} = \begin{pmatrix} \sum_n |s_n|^2 & \dots & \sum_n |s_n|^{L+1} \\ \vdots & \ddots & \vdots \\ \sum_n |s_n|^{L+1} & \dots & \sum_n |s_n|^{2L} \end{pmatrix}$$

which only depends on pilot amplitudes, not phases. □

A. Least Squares Estimator

We now consider the design of $\mathbf{\Phi}$ to minimize the generalized variance, *i.e.*, the determinant, of the LS estimate error covariance matrix

$$\min_{\substack{s_0,\dots,s_{N-1} \\ 0\leq|s_n|\leq 1,\ \forall n}} |\mathbf{C}_{\text{LS}}| = \sigma^{2L}|(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}|. \tag{2}$$

Theorem 1. If N is a multiple of L, the optimal training design that solves (2) divides the N pilots s_n in L sets of N/L pilots. The amplitude of the pilots in each set corresponds to one of the L support points (roots) t_l that solve

$$(1-t)P_L'(2t-1) = 0 \tag{3}$$

where $P_L'(t)$ is the derivative of the L-th degree Legendre polynomial $P_L(t)$.[1] The phase of the pilots can be chosen arbitrarily.

Proof. In light of Lemma 1, we can simplify the optimization problem by restricting s_n to be fully real and positive. After optimization, a phase can be reinserted if desired, without impacting performance. The fully real problem then has the form of a standard polynomial regression. A large body of literature exists on so-called optimal design of experiments [14]. The specific problem of minimizing the determinant is referred to as a D-optimal design in the mathematical statistics community. Two key specificities here are that i) no intercept is considered and ii) pilots should belong to the domain $[0, 1]$. These constraints are considered in [15, Theor. 1], which we tailored to the formalism of this paper. □

Since 1 always is a root of (3), it is optimal to set unit magnitude to N/L pilots. For other sets of pilots, roots of $P_L'(2t-1)$ should be found. For orders $L \leq 10$, closed-form expressions can be found easily. Indeed, the roots of Legendre polynomial derivatives $P_L'(t)$ come in positive/negative pairs and there is always one in zero for even L, *i.e.*, one root in $1/2$ for $P_L'(2t-1)$. In any case, these roots can be pre-computed.

The optimal design of Theorem 1 provides an additional global optimality guarantee on the total regression range. To introduce it, let us first consider the PA predicted response for a given input value $\tilde{s}$. Defining $\tilde{\boldsymbol{\phi}}(\tilde{s}) = (\tilde{s}, ..., \tilde{s}|\tilde{s}|^{L-1})^T$, the prediction error variance or MSE is then given by (1) particularized for $\tilde{N} = 1$

$$\text{MSE}_{\text{LS}}(\tilde{s}) = \sigma^2\tilde{\boldsymbol{\phi}}(\tilde{s})^H(\mathbf{\Phi}^H\mathbf{\Phi})^{-1}\tilde{\boldsymbol{\phi}}(\tilde{s})$$

and the maximal MSE over the regression range, for a given training matrix $\mathbf{\Phi}$, is defined as

$$d_{\text{LS}}(\mathbf{\Phi}) = \max_{\tilde{s},|\tilde{s}|\leq 1} \text{MSE}_{\text{LS}}(\tilde{s}).$$

Corollary 1 (Minimax MSE)**.** The optimal design of Theor. 1 also minimizes the maximal prediction MSE over the PA input range, *i.e.*, it is the solution of

$$\min_{\substack{s_0,\dots,s_{N-1} \\ 0\leq|s_n|\leq 1,\ \forall n}} d_{\text{LS}}(\mathbf{\Phi})$$

achieving $d_{\text{LS}}(\mathbf{\Phi}) = \sigma^2 L/N$ and thus bounding the prediction error variance as $\text{MSE}_{\text{LS}}(\tilde{s}) \leq \sigma^2 L/N$.

Proof. The proof directly comes from the Kiefer-Wolfowitz equivalence theorem [14], [16], particularized to our case. □

B. Linear Minimum Mean Squared Estimator

Unfortunately, we do not have the expression of the optimal training matrix in the LMMSE case. We can still put forward certain positive aspects about using the LS optimal design for the LMMSE estimator. In general, given that $\mathbf{C}_{\text{LS}} \succeq \mathbf{C}_{\text{LMMSE}}$, we also have that $|\mathbf{C}_{\text{LS}}| \geq |\mathbf{C}_{\text{LMMSE}}|$. Hence minimizing $|\mathbf{C}_{\text{LS}}|$ corresponds to minimizing an upper bound on

[1] $P_L(2t-1)$ is the shifted Legendre polynomial to the interval $[0, 1]$.

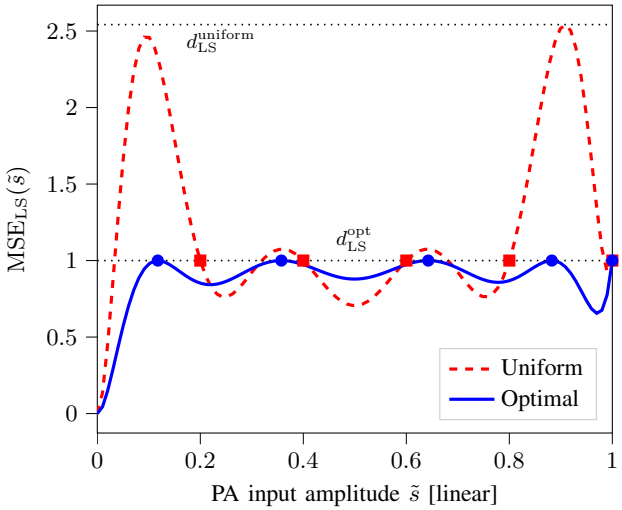

Fig. 1: LS prediction (reconstruction) MSE for a uniform and optimal allocation for $L = N = 5$ and $\sigma^2 = 1$. Pilot locations are represented by rectangles and circles respectively.

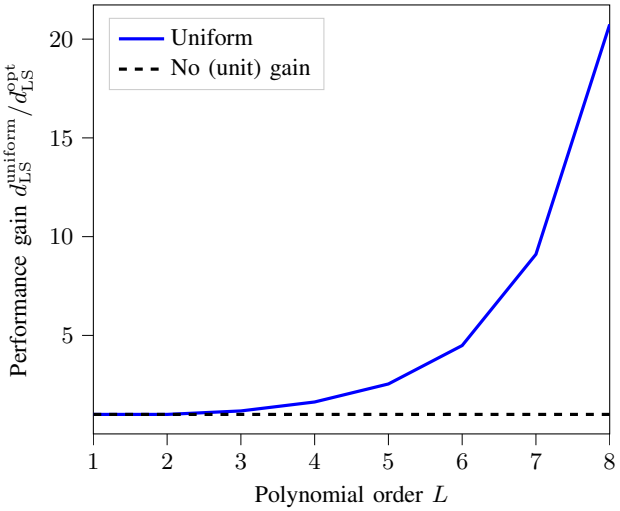

Fig. 2: Performance gain of using the optimal versus uniform allocation in terms of ratio of maximal prediction MSE.

$|\mathbf{C}_{\text{LMMSE}}|$, which is generally beneficial. Given the improved performance, we also have the guarantee that the maximal prediction MSE $d_{\text{LMMSE}}(\mathbf{\Phi})$ is lower or equal than $\sigma^2 L/N$. Moreover, given that the matrix $\mathbf{\Phi}^H\mathbf{\Phi}$ grows with N, it makes sense to normalize it by $1/N$ so that the error covariance matrix becomes

$$\tilde{\mathbf{C}}_{\text{LMMSE}} = \frac{\sigma^2}{N}\tilde{\mathbf{\Phi}}\left(\frac{1}{N}\mathbf{\Phi}^H\mathbf{\Phi} + \frac{\sigma^2}{N}\mathbf{C}_{\boldsymbol{\beta}}^{-1}\right)^{-1}\tilde{\mathbf{\Phi}}^H.$$

It is then clear that, as $\sigma^2/N \to 0$, the two estimators (and their performance) $\hat{\boldsymbol{\beta}}_{\text{LMMSE}}$ and $\hat{\boldsymbol{\beta}}_{\text{LS}}$ converge provided that $\mathbf{\Phi}$ is full rank (thus $N \geq L$). In other words, when the noise variance is small and/or the number of pilots is large, the prior information can be neglected, which intuitively makes sense [17]. This implies that the optimal designs of Section III-A are asymptotically optimal in the LMMSE case as $\sigma^2/N \to 0$.

IV. Simulation Results

In the following, as a benchmark, a uniform pilot allocation is considered, defined as $s_n = 1/N + n/N$ for $n = 0, ..., N-1$.

A. Least Squares Estimator

Fig. 1 illustrates the result of Corol. 1 for a fifth polynomial order ($L = 5$), $\sigma^2 = 1$ and $N = L = 5$ pilots. The optimal one performs sensibly better than a naive uniform allocation. Quantitatively, the maximal MSE is more than two times reduced by the optimal allocation.

For $L = 1$ and $L = 2$, the optimal support points are $[1]$ and $[1/2, 1]$ and the uniform allocation coincides with the optimal one. As L increases, the performance gap between the uniform allocation and the optimal one gets larger. Fig. 2 illustrates this by plotting the ratio of the maximal prediction MSE d_{LS} (maximal MSE value in Fig. 1) for $L = N$. As a reminder, from Corol. 1, we simply have $d_{\text{LS}} = \sigma^2$ for $N = L$ when using the optimal allocation, *i.e.*, it remains constant. In other words, the alternations of the blue continuous curve in Fig. 1 will never be over 1, even as L increases. On the other hand, for the uniform one, the symmetric peaks on the left and right of the graphs will continue to increase as L grows. This explains the large improvement observed in Fig. 2.

B. Linear Minimum Mean Squared Error Estimator

The MSE of the LMMSE estimator $\text{MSE}_{\text{LMMSE}}$ does not scale linearly with σ^2 so that the performance gap with the LS estimator MSE_{LS} will generally depend on the SNR which we define as $\text{SNR} = P_{\max}/(N\sigma^2)$. In simulations, we set $P_{\max} = 1$ and $N = L$. To determine the a priori statistics of the LMMSE $\bar{\boldsymbol{\beta}}$ and $\mathbf{C}_{\boldsymbol{\beta}}$, we consider that the PA follows a random Rapp model without phase distortion so that

$$f(s) = \frac{Gs}{\left(1 + \left(\frac{Gs}{V_{\text{sat}}}\right)^{2S}\right)^{1/2S}}$$

where $G \sim \mathcal{N}(1, 0.01)$, $V_{\text{sat}} \sim \mathcal{N}(1, 0.01)$ and $S \sim \mathcal{N}(2, 0.1)$ are the linear gain, the saturation voltage and the smoothness, respectively. We generate $M = 100$ realizations of the PA response, for which confidence intervals are shown in Fig. 3. We then fit a $L = 7$ polynomial order to obtain coefficients $\boldsymbol{\beta}_m$ for each realization m. As can be seen in the figure, this polynomial approximation is very accurate and cannot be distinguished at first sight from the exact Rapp model response. As discussed in Section II, in practice, the channel h, assumed narrowband, will multiply these coefficients. This induces an attenuation $|h|$, which we take into account through the scaling of the noise level σ^2 but also a phase rotation $e^{\jmath\angle h}$. Depending if this phase is known, *e.g.*, from previous measurements or as part of the calibration process, or not, we consider two cases. i) **Coherent** case where it is known and can be compensated so that we can estimate the coefficients mean and covariance as $\bar{\boldsymbol{\beta}} = 1/M\sum_{m=0}^{M-1}\boldsymbol{\beta}_m$ and $\mathbf{C}_{\boldsymbol{\beta}} = 1/M\sum_{m=0}^{M-1}\boldsymbol{\beta}_m\boldsymbol{\beta}_m^H - \bar{\boldsymbol{\beta}}\bar{\boldsymbol{\beta}}^T$. ii) **Noncoherent** case where phase $\angle h$ is unknown, assumed

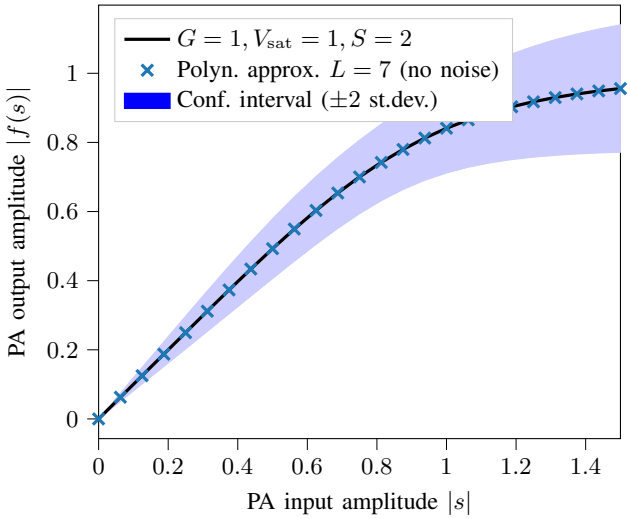

Fig. 3: Random power amplifier response over 100 realizations.

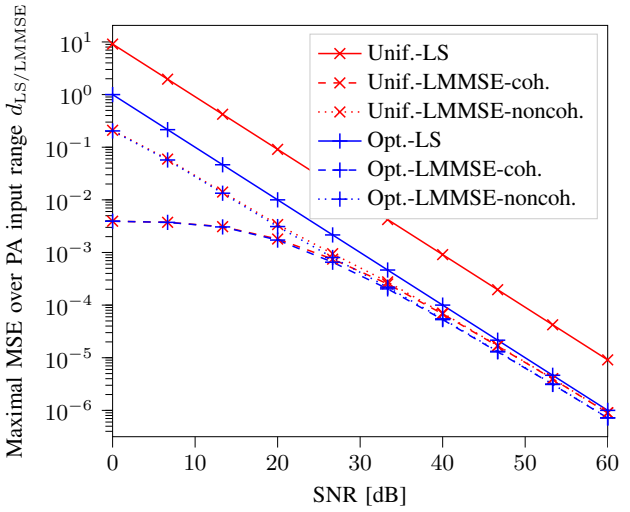

Fig. 4: Maximal prediction variance for LS/LMMSE estimators and uniform versus optimized training.

uniformly distributed in $[0, 2\pi]$, causing coefficients to have a zero mean so that $\bar{\boldsymbol{\beta}} = \mathbf{0}$ and $\mathbf{C}_{\boldsymbol{\beta}} = 1/M \sum_{m=0}^{M-1} \boldsymbol{\beta}_m \boldsymbol{\beta}_m^H$.

A performance comparison is given in Fig. 4 showing the maximal prediction variance $d_{\text{LS/LMMSE}}(\boldsymbol{\Phi})$ (for the LS or LMMSE estimator) as a function of the SNR. A few important observations can be made. i) Using the optimized allocation of Theor. 1 provides most gain as compared to the uniform one when using the LS estimator (factor 10 in accordance to Fig. 2) while the improvement is negligible when using the LMMSE estimator. ii) At low SNR, the LMMSE estimator has a large gain as compared to the LS estimator given the prior information, especially in the coherent case where even more prior information is available: factor 300 (coherent) versus 5 (noncoherent) at 0 dB SNR. At high SNR, their performance converges to the LS one, as expected from Section IV-B. iii) Interestingly, even at 60 dB SNR, when using the uniform allocation, the LMMSE estimator has not yet converged to the LS one. It achieves a similar performance as the LS/LMMSE one with optimized training. This means that, when using the LMMSE estimator, a simpler uniform allocation would provide close-to-optimal solution and can be considered sufficient.

V. Conclusion

In this work, we considered OTA PA calibration using estimation theory. The LS and LMMSE estimators were re-derived. The optimal training in the LS case was obtained by minimizing the generalized variance of the error covariance matrix. This optimal design also minimizes the maximal MSE when reconstructing the PA over its full input range. Via simulations, we show an improvement of a factor 10 for a $L = 7$ PA order. Leveraging prior information, the LMMSE estimator provides an additional gain of a factor 5 and 300 in the noncoherent and coherent cases respectively.

Acknowledgment

This research was partially funded by 6GTandem, supported by the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096302.

References

[1] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 40–49, 2011.
[2] P. M. Lavrador *et al.*, "The linearity-efficiency compromise," *IEEE Microw. Mag.*, vol. 11, no. 5, pp. 44–58, 2010.
[3] C. Fager *et al.*, "Linearity and Efficiency in 5G Transmitters: New Techniques for Analyzing Efficiency, Linearity, and Linearization in a 5G Active Antenna Transmitter Context," *IEEE Microw. Mag.*, vol. 20, no. 5, pp. 35–49, 2019.
[4] S. C. Cripps, *RF power amplifiers for wireless communications*. Artech House, 2006, vol. 2.
[5] F. Rottenberg, G. Callebaut, and L. Van der Perre, "The Z3RO Family of Precoders Cancelling Nonlinear Power Amplification Distortion in Large Array Systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 2036–2047, 2023.
[6] T. Feys, X. Mestre, and F. Rottenberg, "Self-supervised learning of linear precoders under non-linear pa distortion for energy-efficient massive mimo systems," in *ICC 2023*, 2023, pp. 6367–6372.
[7] X. Huang, Z. Zhu, and H. Leung, *Signal processing for RF circuit impairment mitigation*. Boston: Artech house, 2014.
[8] A. Ben Ayed *et al.*, "Digital predistortion of millimeter-wave arrays using near-field based transmitter observation receivers," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 7, pp. 3713–3723, 2022.
[9] N. Tervo *et al.*, "Digital predistortion concepts for linearization of mmw phased array transmitters," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, 2019, pp. 325–329.
[10] N. Tervo, "Concepts for radiated nonlinear distortion and spatial linearization in millimeter-wave phased arrays." PhD thesis, University of Oulu, March 2022.
[11] L. Ding and G. Zhou, "Effects of Even-Order Nonlinear Terms on Power Amplifier Modeling and Predistortion Linearization," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 156–162, Jan. 2004.
[12] R. Raich, H. Qian, and G. Zhou, "Orthogonal Polynomials for Power Amplifier Modeling and Predistorter Design," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1468–1479, Sep. 2004.
[13] S. M. Kay, *Fundamentals of statistical signal processing*, ser. Prentice Hall signal processing series. Prentice-Hall PTR, 1993.
[14] F. Pukelsheim, *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.
[15] M.-N. L. Huang, F.-C. Chang, and W. K. Wong, "D-optimal designs for polynomial regression without an intercept," *Statistica Sinica*, pp. 441–458, 1995.
[16] J. Kiefer and J. Wolfowitz, "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, vol. 12, pp. 363–366, 1960.
[17] K. Chaloner and I. Verdinelli, "Bayesian Experimental Design: A Review," *Statistical Science*, vol. 10, no. 3, Aug. 1995.