

Correction to “Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs”

Lucas N. Ribeiro, Stefan Schwarz, *Senior Member, IEEE*,
Markus Rupp, *Fellow, IEEE*, André L. F. de Almeida, *Senior Member, IEEE*

The authors of [1] were requested to clarify a concern regarding the proposed additive quantization noise (AQN) model. The concern is related to a square-root term present in the proposed model that is not seen in the AQN model adopted in other works, such as [2]. Upon reviewing the paper to resolve this concern, the authors noticed that the assumptions made on the proposed AQN model prevent the use of a certain formula used to calculate the quantization distortion term ρ_b . The proposed AQN model relies on a power conservation assumption that does not necessarily lead to optimality in the mean-square error (MSE) sense. Nevertheless, the formula used to calculate the quantization distortion term ρ_b in [1] assumes an MSE-optimal quantizer. Therefore, the application of that formula is not correct.

This erratum aims to provide a detailed explanation of the assumptions behind the proposed AQN model, and discusses the formula used to calculate ρ_b . It is argued that, although the formula’s assumption is not satisfied, its use is still reasonable under the considered circumstances.

I. THE ADDITIVE QUANTIZATION NOISE MODEL

Let the input and the output of a b -bit quantizer $Q_b(\cdot)$ be denoted by u and $p = Q_b(u)$, respectively. Without loss of generality, the input u is assumed to have zero mean. The AQN model for the quantizer is given by

$$p = Q_b(u) = Bu + e, \quad (1)$$

where B represents a linear scaling factor, and e a zero-mean distortion term. According to [3, Section 5.3], AQN models are defined generally and then “carried farther by making specific assumptions as to the statistical nature of the quantization noise so that the model can be handled analytically”. In this sense, consider a quantization function described by the following assumptions:

- 1) The input power $P_u = \mathbb{E}[|u|^2]$ is conserved, i.e., it is equal to the output power $P_p = \mathbb{E}[|p|^2]$.
- 2) The additive quantization noise e and the input u are uncorrelated, i.e., $\mathbb{E}[ue^*] = \mathbb{E}[eu^*] = 0$.

L. N. Ribeiro was affiliated with the Federal University of Ceará, and the Institute of Telecommunications, TU Wien, when the original paper was published.

S. Schwarz, and M. Rupp are with the Institute of Telecommunications, TU Wien, Vienna, Austria.

A. L. F. de Almeida is with the Wireless Telecommunications Research Group (GTEL), Federal University of Ceará, Fortaleza, Brazil.

- 3) The power of the distortion term e is given by $P_e = \rho_b P_u$, where ρ_b is a proportionality constant that measures quantization distortion.

Assumption 1 can be motivated by companding, a quantization technique that involves compressing the input signal’s dynamic range and expanding it after quantization, such that the signal power is conserved. Furthermore, this assumption is quite useful when formulating precoding optimization problems, because it ensures that quantization does not change the power level set by the precoding filters. Assumption 2 is a classic assumption in additive quantization models [3], [4], [5], and Assumption 3 ensures that the input and distortion power terms are proportional. In fact, ρ_b can also be seen as the inverse signal-to-quantization-noise ratio for a resolution of b bits.

According to the Bussgang decomposition [4], the scaling factor B is uniquely defined for each choice of non-linear function. With Assumptions 1–3, the scaling factor B can now be uniquely determined. Consider the output power

$$P_p = \mathbb{E}[|p|^2] = B^2 P_u + P_e. \quad (2)$$

From Assumption 3, Equation (2) can be further expressed as

$$P_p = B^2 P_u + \rho_b P_u = (B^2 + \rho_b) P_u. \quad (3)$$

Based on Assumption 1, it is inferred that $B^2 + \rho_b = 1$. Since B is a gain term, only the positive root of this equation is relevant, and, then the AQN model (1) is written as

$$p = Q_b(u) = \sqrt{1 - \rho_b} u + e. \quad (4)$$

The AQN model introduced in [2], [6] is written slightly different:

$$p = K_b(u) = (1 - \rho_b)u + e. \quad (5)$$

Instead of enforcing power conservation (Assumption 1), this model assumes a quantizer $K_b(\cdot)$ that has the following property $\mathbb{E}[u|K_b(u)] = K_b(u)$, i.e., the quantization interval is represented by its mean value [4]. As demonstrated in [2, Appendix C-1], this property leads to $B = 1 - \rho_b$.

A. MSE Optimality

The property $\mathbb{E}[u|K_b(u)] = K_b(u)$ assumed by the quantizer $K_b(\cdot)$ in (5) is necessary to MSE optimality [3, Section 6.2]. The proposed AQN model (4) does not consider this property, and, therefore, is not necessarily optimal in the MSE

sense. This is relevant for modeling the quantization distortion term ρ_b . In [1], this term is calculated as

$$\rho_b = \frac{\pi\sqrt{3}}{2} 2^{-2b} \quad (6)$$

as in [5]. This formula, however, applies to optimal non-uniform quantizers, and there are no guarantees that (4) satisfies this condition. Nevertheless, (6) can be regarded as a *best case* distortion term, which would then motivate its use in [1]. In any case, it would be necessary to fit ρ_b to the quantization distortion of relevant digital-to-analog converters in order to get more realistic models.

II. MIMO EXTENSION

In Section II-B of [1], an M -dimensional input vector $\mathbf{u}$ is fed to a vector quantizer $Q_b(\cdot)$, producing an M -dimensional output $\mathbf{p} = Q_b(\mathbf{u})$. It is assumed that the AQN model (4) is applied individually to each input u_m of the vector quantizer. This implies an array of M independent scalar quantizers $\{Q_{b,1}, \dots, Q_{b,M}\}$ such that its m th output is given by

$$p_m = Q_{b,m}(u_m) = \sqrt{1 - \rho_{b,m}} u_m + e_m, \quad (7)$$

for $m = 1, \dots, M$. Because each quantizer is applied independently from each other, it is reasonable to assume:

4. $\mathbb{E}[e_m e_n^*] = 0$ for all $m \neq n$.
5. $\mathbb{E}[u_m e_n^*] = 0$ for all $m \neq n$.

Now, define the scaling matrix

$$\mathbf{\Upsilon}_b = \text{Diag}(\sqrt{1 - \rho_{b,1}}, \dots, \sqrt{1 - \rho_{b,M}}). \quad (8)$$

Equation (7) can be written in matrix notation as

$$\mathbf{p} = \mathbf{\Upsilon}_b \mathbf{u} + \mathbf{e}, \quad (9)$$

where $\mathbf{p} = [p_m]$, $\mathbf{u} = [u_m]$, and $\mathbf{e} = [e_m]$ for $m = 1, \dots, M$. From Assumptions 3–5, the power of the m th distortion term is given by $P_{e,m} = \rho_{b,m} P_{u,m}$. This equation can be written in matrix notation as

$$\mathbf{R}_{ee} = \text{Diag}(P_{e,1}, \dots, P_{e,M}) \quad (10a)$$

$$= \text{Diag}(\rho_{b,1}, \dots, \rho_{b,M}) \text{Diag}(\mathbf{R}_{uu}), \quad (10b)$$

where $\mathbf{R}_{ee} = \mathbb{E}[\mathbf{e}\mathbf{e}^H]$ denotes the autocorrelation matrix of the distortion vector, and $\mathbf{R}_{uu} = \mathbb{E}[\mathbf{u}\mathbf{u}^H]$ the autocorrelation matrix of the input vector. In the case that all M quantizers undergo the same distortion ρ_b , (8) and (10) simplify to

$$\mathbf{\Upsilon}_b = \sqrt{1 - \rho_b} \mathbf{I}_M \quad (11)$$

and

$$\mathbf{R}_{ee} = \rho_b \text{Diag}(\mathbf{R}_{uu}), \quad (12)$$

respectively.

III. TRANSMIT POWER CONSTRAINT

The lossless transmit signal in [1] is written as

$$\tilde{\mathbf{x}} = \mathbf{F}_{\text{RF}} \mathbf{Q}_b(\mathbf{F}_{\text{BB}} \mathbf{s}) = \mathbf{F}_{\text{RF}} \mathbf{\Upsilon}_b \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{F}_{\text{RF}} \mathbf{e} \quad (13)$$

with $\mathbf{u} = \mathbf{F}_{\text{BB}} \mathbf{s}$ and the average transmit power constraint is formulated as

$$P_{\tilde{\mathbf{x}}} = \mathbb{E}[\|\tilde{\mathbf{x}}\|_2^2] \leq P_{\text{max}}. \quad (14)$$

The non-linear quantization in (13) turns the precoder optimization problem quite challenging. As discussed in Section III of [1], the AQN model (4) simplifies the precoder optimization based on following result.

Lemma 1. *If the baseband precoder $\mathbf{F}_{\text{BB}}$ is unitary, then $P_{\tilde{\mathbf{x}}} = \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2$.*

Proof. The power of the lossless transmit signal can be written as

$$P_{\tilde{\mathbf{x}}} = \mathbb{E}[\text{Tr}(\tilde{\mathbf{x}} \tilde{\mathbf{x}}^H)] \quad (15a)$$

$$= \text{Tr}(\mathbf{F}_{\text{RF}} \mathbf{\Upsilon}_b \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{\Upsilon}_b^H \mathbf{F}_{\text{RF}}^H) + \text{Tr}(\mathbf{F}_{\text{RF}} \mathbf{R}_{ee} \mathbf{F}_{\text{RF}}^H). \quad (15b)$$

Substituting (11) and (12) into (15) gives:

$$P_{\tilde{\mathbf{x}}} = (1 - \rho_b) \text{Tr}(\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \mathbf{F}_{\text{RF}}^H) + \rho_b \text{Tr}(\mathbf{F}_{\text{RF}} \text{Diag}(\mathbf{R}_{uu}) \mathbf{F}_{\text{RF}}^H). \quad (16)$$

As $\mathbf{F}_{\text{BB}}$ is unitary, it holds that

$$\text{Diag}(\mathbf{R}_{uu}) = \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H = \mathbf{I}_M, \quad (17)$$

and (16) can be written as

$$P_{\tilde{\mathbf{x}}} = (1 - \rho_b) \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 + \rho_b \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2. \quad \square$$

This lemma shows that, similar to conventional hybrid precoding schemes, the AQN model (4) does not change the transmit power set by the hybrid precoder if the baseband term is unitary. This result is central for the formulation of the *quantized hybrid precoding problem* in [1]. Note that the alternative AQN model (5) does not satisfy Lemma 1 in general and would lead to a complicated power expression.

ACKNOWLEDGMENT

The authors would like to thank Mengyuan Ma for kindly reaching out and raising the concern about the AQN model.

REFERENCES

- [1] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, 2018.
- [2] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 618–632, 2007.
- [3] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 1991.
- [4] O. T. Demir and E. Bjornson, "The Bussgang decomposition of nonlinear systems: Basic theory and MIMO extensions [lecture notes]," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, 2020.
- [5] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 1–5, 2012.
- [6] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. 2015 Information Theory and Applications Workshop (ITA)*, pp. 191–198, IEEE, 2015.