Terahertz Transmit Beamforming with 1-Bit DACs and ADCs

Rasoul Nikbakht, Angel Lozano
Universitat Pompeu Fabra (UPF)
08018 Barcelona, Spain.
Email: {rasoul.nikbakht, angel.lozano}@upf.edu

***Abstract*—This paper tackles the problem of transmit beamforming with 1-bit digital-to-analog and analog-to-digital converters. While, at frequencies above 100 GHz, such 1-bit converters are instrumental to restrain the power consumption and enable transmissions spanning many gigahertz of bandwidth, they fundamentally alter the nature of the communication problem. Transmit beamforming, a key task when communicating at these high frequencies, then amounts to identifying the quartet of channel-dependent transmit vectors that maximizes the mutual information. This problem, which becomes unwieldy for even modest numbers of antennas, is herein tackled by means of an unsupervised learning approach that proves effective for very large arrays.**

I. Introduction

The next frontier in the quest for fresh spectrum over which to communicate wirelessly is the terahertz band, broadly taken to be 100 GHz–10 THz [1]. Although there are reasons why this band remains largely unexplored, some of the long-standing obstacles look increasingly surmountable [2]. Because of the lack of diffraction and atmospheric attenuation, propagation is predominantly line-of-sight (LOS) and short-range, but that is compatible with a number of emerging applications.

A major challenge to ultrabroadband communication at terahertz frequencies is the power consumption associated with high-resolution digital-to-analog (DAC) and analog-to-digital (ADC) conversion at transmitter and receiver, respectively. Precisely, high-resolution DACs go hand in hand with highly linear power amplifiers whose efficiency is very poor [3]. In turn, the ADC power consumption grows linearly with the bandwidth and exponentially with the number of resolution bits [4].

The crux of the power consumption issue is therefore the resolution of the converters, and the natural solution is to lower that resolution. Taken to the limit, this leads to 1-bit DACs and ADCs, which do drastically curb the power consumption, at the expense of an exceedingly nonlinear behavior that severely distorts the signals. There is extensive literature on transmission strategies and the ensuing performance with 1-bit ADCs but full-resolution DACs (see [5]–[8] and references therein), and a smaller but growing body of work that considers 1-bit converters at both ends [9]–[18].

Because of the extremely high omnidirectional pathloss at terahertz frequencies, antenna arrays are instrumental and a central problem is that of transmit beamforming. In the face of 1-bit DACs and ADCs, the beamforming problem amounts to the identification of the most appropriate quantized transmit vectors for each channel realization. From such general starting point, it was proposed in [11] to determining those beamforming vectors on the basis of minimizing the uncoded bit error probability. More fundamentally, though, the beamforming vectors should be determined on the basis of maximizing the mutual information [17]. At low SNR, this is tantamount to maximizing the received power, and efficient ways have been put forth to identify the corresponding vectors [19]. More generally, though, this is a binary optimization that entails a burdensome comparison over all possible vectors spawn by the transmit array, the number of which grows exponentially with the number of antennas.

To tame this problem and enable beamforming with large arrays, we advocate a learning-based technique and seek to approximate the mapping between channel realizations and transmit vectors by means of a neural network (NN). And, to circumvent the need for labelled training data, which would require solving the problem we seek to overcome in the first place, we espouse an unsupervised form of learning [20].

II. Signal and Channel Models

A. Signal Model

Consider a transmitter equipped with N_{t} antennas and 1-bit DACs per complex dimension. The receiver, which features one antenna with a 1-bit ADC per complex dimension, observes

$$y = \operatorname{sgn}\left(\sqrt{\frac{\mathsf{SNR}}{2N_{\mathrm{t}}}}\,\boldsymbol{h}\boldsymbol{x} + \boldsymbol{z}\right) \tag{1}$$

where the sign function applies separately to the real and imaginary parts of each entry, such that $y \in \{\pm 1 \pm \mathrm{j}\}$, while $\boldsymbol{h}$ is the $1 \times N_{\mathrm{t}}$ channel row vector normalized to have unit-variance entries, $z \sim \mathcal{N}_{\mathbb{C}}(0,1)$ is the noise, and SNR is the signal-to-noise ratio per receive antenna in the absence of beamforming. The $N_{\mathrm{t}} \times 1$ transmit column vector $\boldsymbol{x}$ has entries $x_n \in \{\pm 1 \pm \mathrm{j}\}$ for $n = 1, \ldots, N_{\mathrm{t}}$.

Since $\boldsymbol{x}$ and y are discrete-valued, (1) embodies, for each given $\boldsymbol{h}$, a discrete memoryless channel with $4^{N_{\mathrm{t}}} \times 4$ transition probabilities. These transition probabilities are determined by [14]

$$p_{y|\boldsymbol{x}} = p_{\Re\{y\}|\boldsymbol{x}}\, p_{\Im\{y\}|\boldsymbol{x}}, \tag{2}$$

where the factorization follows from the independence of the real and imaginary noise components. Each such noise component has variance $1/2$, hence

$$p_{\Re\{y\}|\boldsymbol{x}}(1|\mathsf{x}) = \Pr\left[\sqrt{\frac{\mathsf{SNR}}{2N_{\rm t}}}\Re\{\boldsymbol{h}\mathsf{x} + z\} > 0\right] \tag{3}$$

$$= \Pr\left[\Re\{z\} > -\sqrt{\frac{\mathsf{SNR}}{2N_{\rm t}}}\Re\{\boldsymbol{h}\mathsf{x}\}\right] \tag{4}$$

$$= Q\left(-\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\boldsymbol{x}\}\right) \tag{5}$$

where

$$Q(\xi) = \frac{1}{\sqrt{2\pi}}\int_{\xi}^{\infty} e^{-u^2/2}\,{\rm d}u \tag{6}$$

is the Gaussian Q-function. Similarly,

$$p_{\Re\{y\}|\boldsymbol{x}}(-1|\mathsf{x}) = Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\boldsymbol{x}\}\right). \tag{7}$$

From (5) and (7), we can write

$$p_{\Re\{y\}|\boldsymbol{x}}(\Re\{\mathsf{y}\}|\mathsf{x}) = Q\left(-\Re\{\mathsf{y}\}\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\mathsf{x}\}\right) \tag{8}$$

and, mirroring it for the imaginary part,

$$p_{y|\boldsymbol{x}}(\mathsf{y},\mathsf{x}) = Q\left(-\Re\{\mathsf{y}\}\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\mathsf{x}\}\right) \cdot Q\left(-\Im\{\mathsf{y}\}\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Im\{\boldsymbol{h}\mathsf{x}\}\right). \tag{9}$$

The transition probabilities correspond to (9) evaluated for the 4 possible values of y and the $4^{N_{\rm t}}$ values of $\boldsymbol{x}$. If $\boldsymbol{h}$ is known, these transition probabilities can be readily computed. Conversely, if the transition probabilities are known, $\boldsymbol{h}$ can be deduced.

B. Channel Model

For the sake of completeness, we consider the two extremes in terms of the distribution of $\boldsymbol{h}$:

1) IID Rayleigh-faded entries.
2) LOS propagation with planar wavefronts, whereby the entries of $\boldsymbol{h}$ are governed by a few geometric parameters. For instance, with a uniform linear array (ULA) having antenna spacing $d_{\rm t}$, and under the premise of planar wavefronts,

$$h_n = e^{-{\rm j}\frac{2\pi}{\lambda} n d_{\rm t}\cos\theta} \qquad n = 1,\ldots,N_{\rm t} \tag{10}$$

where θ is the angle spanned by the transmit-receive direction and the ULA while λ denotes the wavelength.

III. Transmit Beamforming

The set of $4^{N_{\rm t}}$ possible transmit vectors $\boldsymbol{x}$ can be partitioned into $4^{N_{\rm t}-1}$ quartets, each containing four vectors and being invariant under a $90°$ phase rotation of all the entries: from any vector in the quartet, the rest are obtained by repeatedly multiplying by j. Since a $90°$ phase rotation of $\boldsymbol{x}$ propagates as a $90°$ phase rotation of $\boldsymbol{h}\boldsymbol{x}$, and the added noise z is rotationally invariant, the vectors within each transmit quartet are statistically equivalent and they should thus have the same probability of being transmitted so as to convey the maximum amount of information; this intuition is formalized in [17, lemma 1].

By the same token, for every transmit vector giving rise to a specific value of y there are three rotated transmit vectors (the other members of the quartet) that give rise to the other possible values of y with equal probability. Consequently, and irrespective of the channel realization, y takes the four values $\pm 1 \pm {\rm j}$ equiprobably; again, this intuition is formalized in [17].

Let $\boldsymbol{h}$ be known by both transmitter and receiver, and let $\mathcal{H}(\cdot)$ stand for entropy. The mutual information between $\boldsymbol{x}$ and y for a given $\boldsymbol{h}$ satisfies

$$I(\boldsymbol{x};y|\boldsymbol{h}) = \mathcal{H}(y|\boldsymbol{h}) - \mathcal{H}(y|\boldsymbol{x},\boldsymbol{h}) \tag{11}$$

$$= 2 - \mathcal{H}(y|\boldsymbol{x},\boldsymbol{h}) \tag{12}$$

where (12) follows from the equiprobability of the four values of y. Denoting by p_k the probability of transmitting the kth quartet, $\boldsymbol{x} \in \{\mathsf{x}_k, {\rm j}\mathsf{x}_k, -\mathsf{x}_k, -{\rm j}\mathsf{x}_k\}$,

$$I(\boldsymbol{x};y|\boldsymbol{h}) = 2 - \sum_{k=1}^{4^{N_{\rm t}-1}} \frac{p_k}{4}\sum_{i=0}^{3}\mathcal{H}(y|\boldsymbol{x} = {\rm j}^i\mathsf{x}_k,\boldsymbol{h}) \tag{13}$$

$$= 2 - \sum_{k=1}^{4^{N_{\rm t}-1}} p_k\,\mathcal{H}(y|\boldsymbol{x} = \mathsf{x}_k,\boldsymbol{h}) \tag{14}$$

$$= 2 - \sum_{k=1}^{4^{N_{\rm t}-1}} p_k\Big[\mathcal{H}\big(\Re\{y\}|\boldsymbol{x} = \mathsf{x}_k,\boldsymbol{h}\big) + \mathcal{H}\big(\Im\{y\}|\boldsymbol{x} = \mathsf{x}_k,\boldsymbol{h}\big)\Big] \tag{15}$$

where (14) follows from the equiprobability and statistical equivalence of the vectors in each quartet k. Given $\boldsymbol{x}$ and $\boldsymbol{h}$, $\Re\{y\}$ and $\Im\{y\}$ are binary random variables whose respective probabilities of being ± 1, recalling (9), are[1]

$$Q\left(\pm\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\boldsymbol{x}\}\right) \tag{16}$$

and

$$Q\left(\pm\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Im\{\boldsymbol{h}\boldsymbol{x}\}\right). \tag{17}$$

Hence, for arbitrary quartet probabilities $\{p_k\}$,

$$I(\boldsymbol{x};y|\boldsymbol{h}) = 2 - \sum_{k=1}^{4^{N_{\rm t}-1}} p_k\left[\mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\mathsf{x}_k\}\right)\right)\right.$$

[1] $Q(-\xi) = 1 - Q(\xi)$.

$$+\mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Im\{\boldsymbol{h}\mathbf{x}_k\}\right)\right)\Bigg]. \quad (18)$$

where

$$\mathcal{H}_{\rm b}(p) = -p\log_2 p - (1-p)\log_2(1-p) \quad (19)$$

is the binary entropy function.

The mutual information in (18) is maximized by assigning transmission probability 1 to the quartet with the smallest sum of binary entropies; by means of the four vectors in that quartet, a scalar 2-bit symbol can be conveyed to the receiver. This gives, for channel $\boldsymbol{h}$, a spectral efficiency of

$$\mathcal{I}(\mathsf{SNR},\boldsymbol{h}) = 2 - \min_k \Bigg[\mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\mathbf{x}_k\}\right)\right) + \mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Im\{\boldsymbol{h}\mathbf{x}_k\}\right)\right)\Bigg]. \quad (20)$$

If $\boldsymbol{h}$ varies in time and/or frequency and the coding takes place over a sufficiently broad range of variations, then what is operationally relevant is the ergodic spectral efficiency [21]

$$\mathcal{I}(\mathsf{SNR}) = \mathbb{E}_{\boldsymbol{h}}\big[\mathcal{I}(\mathsf{SNR},\boldsymbol{h})\big]. \quad (21)$$

Alternatively, if the channel is information stable, i.e., stable over each codeword transmission, then $\mathcal{I}(\mathsf{SNR},\boldsymbol{h})$ has itself operational significance and $\mathcal{I}(\mathsf{SNR})$ should be interpreted as the average spectral efficiency over the settings described by the distribution of $\boldsymbol{h}$.

IV. Unsupervised Learning Approach

The determination of the optimum quartet is a binary optimization that entails selecting, for each channel $\boldsymbol{h}$, one of $4^{N_{\rm t}-1}$ possibilities. This problem can be interpreted as a parametric optimization, with $\boldsymbol{h}$ being the parameter, the $4^{N_{\rm t}-1}$ possible transmit quartets being the search space, and the optimum such quartet being the solution for the specific instance of the optimization associated with $\boldsymbol{h}$. An unsupervised learning approach for parametric optimization is developed in [20] for implementation on a feedforward NN. Applied to the problem at hand, this approach entails:

- Considering $\boldsymbol{h}$ as the input to the NN.
- Defining a loss function based on the objective of identifying the optimum quartet.
- Iteratively updating the weights so as to minimize that loss.

No labeled data is required for the training and, after convergence, the NN approximates the mapping between $\boldsymbol{h}$ and the optimum transmit quartet.

A. Loss Function

The nature of the search space at hand, consisting of $N_{\rm t}$-dimensional vectors with binary real and imaginary parts, is incompatible with the gradient back-propagation required to update the NN weights. To skirt this hurdle, we relax the search space into that of $N_{\rm t}$-dimensional complex vectors and incorporate to the loss function a term that favors solutions with near-binary real and imaginary parts. Precisely, the loss function we minimize is, recalling (20),

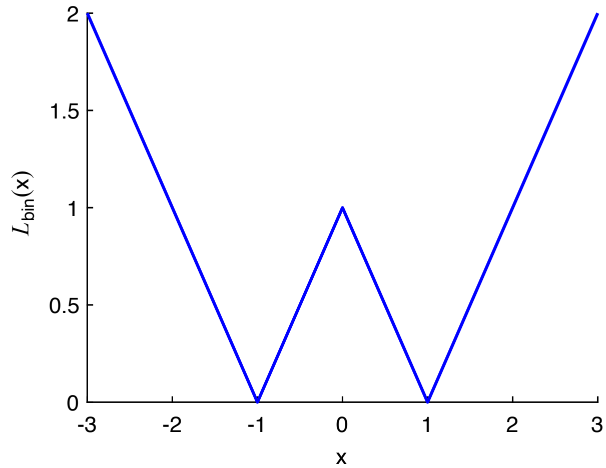

Fig. 1: Loss term penalizing solutions as they deviate from having binary real and imaginary parts.

$$L(\boldsymbol{x},\boldsymbol{h},\mathsf{SNR}) = -2 + \mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Re\{\boldsymbol{h}\boldsymbol{x}\}\right)\right) + \mathcal{H}_{\rm b}\left(Q\left(\sqrt{\frac{\mathsf{SNR}}{N_{\rm t}}}\Im\{\boldsymbol{h}\boldsymbol{x}\}\right)\right) + \boldsymbol{\beta} L_{\rm bin}(\boldsymbol{x}) \quad (22)$$

where

$$L_{\rm bin}(\boldsymbol{x}) = [-\boldsymbol{x}-1]^+ + [\boldsymbol{x}+1]^+ - 2\,[\boldsymbol{x}]^+ + 2\,[\boldsymbol{x}-1]^+ \quad (23)$$

given

$$[\xi]^+ = \begin{cases} 0 & \text{for } \xi \le 0 \\ \xi & \text{for } \xi > 0 \end{cases} \quad (24)$$

as the rectified linear unit (ReLU) function, applied in (23) to the real and imaginary parts of each entry of $\boldsymbol{x}$. Depicted in Fig. 1, the term $L_{\rm bin}$ increasingly penalizes solutions as they deviate from being binary-valued, with this penalty being modulated by $\boldsymbol{\beta}$.

B. Learning Stage

A lean NN is employed, with only three layers whose numbers of neurons depending on $N_{\rm t}$ are described in Table I, along with their type of activation functions. Assembled into a complex vector, the $2N_{\rm t}$ real outputs produced by the output layer represent $\boldsymbol{x}$.

For learning purposes, $\boldsymbol{h}$ is repeatedly sampled from its distribution with a batch size of 5000, and each sample is assigned a random $\mathsf{SNR} \in [-30, 20]$ dB. The resulting tensor, concatenation of $\boldsymbol{h}$ and SNR, is fed to the NN and $L(\boldsymbol{x},\boldsymbol{h},\mathsf{SNR})$ is evaluated for the NN's output, $\boldsymbol{x}$. The NN weights, randomly initialized, are then updated using the Adam algorithm. The process is repeated 10000 times, and $\boldsymbol{\beta}$ is obtained through a cross-validation sweeping from

TABLE I: Number of neurons per layer depending on N_t. Also indicated is the type of activation function at each layer.

N_t	2–8	16–32	64–128
Input layer (ReLU activation)	50	100	500
Hidden layer (ReLU activation)	50	100	500
Output layer (linear activation)	$2N_\mathrm{t}$	$2N_\mathrm{t}$	$2N_\mathrm{t}$

TABLE II: Learning parameters depending on N_t.

N_t	2	4	8	16	32	64	128
$\boldsymbol{\beta}$	0.05	0.1	0.1	0.1	.1	0.2	0.3
τ (dB)	20	3	0	-3	-6	-9	-12

0.01 to 10; the optimized values are listed in Table II. To avoid oscillations around local optima during the weight adjustment, the learning rate—amplitude of the gradient steps—is reduced gradually from 0.001 down to 0.0001.

At very high SNR, the gradient of $Q(\cdot)$ in (22) vanishes. Likewise, very small values for that function cause numerical problems when computing the gradient of $\mathcal{H}_\mathrm{b}(\cdot)$. To circumvent these issues:

- $Q(\cdot)$ is clipped at 10^{-5} whenever its value falls within $[0, 10^{-5}]$, and it is clipped at $1-10^{-5}$ whenever its value falls within $[1-10^{-5}, 1]$.
- During learning, SNR is replaced by

$$\mathsf{SNR}_{\text{learn}} = [-\mathsf{SNR}+\tau]^+ + 0.1\,[\mathsf{SNR}-\tau]^+ \tag{25}$$

  where τ is to be set to the value beyond which the spectral efficiency saturates (see Table II). The intuition behind this modification of SNR is that the NN weights optimized for SNR $=\tau$ remain valid thereafter. Once the training is complete, the actual SNR is used for performance evaluation, as the gradient is no longer required.

Throughout the learning stage, the continuous nature of the values of $\boldsymbol{x}$ is respected upon evaluation of the loss function. Afterwards, for every input $\boldsymbol{h}$, the output is clipped to produce $\mathbf{x}_{k^\star}$ representing the selected quartet for that $\boldsymbol{h}$.

The pipelines for both the learning and the evaluation stages are graphically represented in Fig. 2.

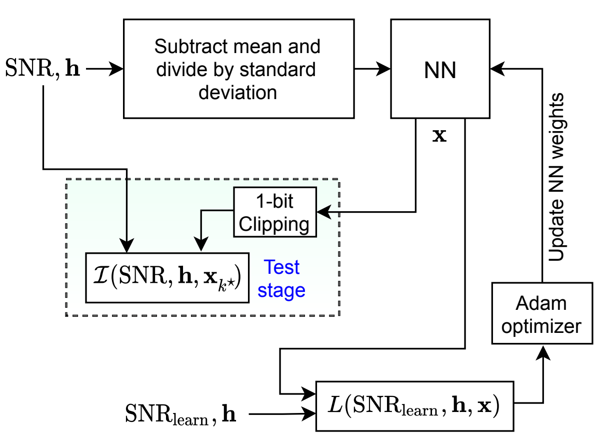

Fig. 2: Learning and evaluation pipelines.

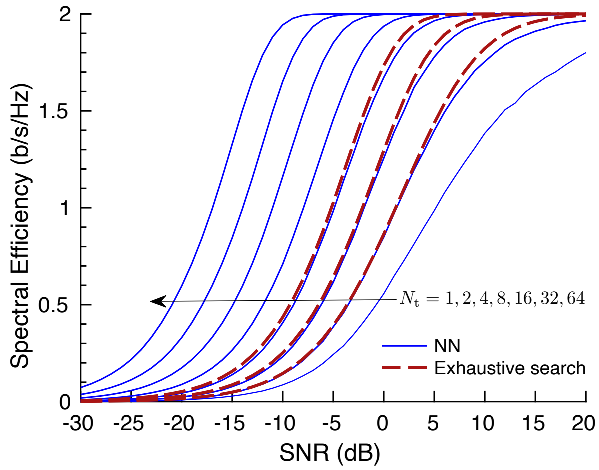

Fig. 3: Ergodic spectral efficiency vs SNR over an IID Rayleigh-faded channel. For $N_\mathrm{t}=1$, 2, 4 and 8: NN (in solid) versus exhaustive search (in dashed). For $N_\mathrm{t}=16,32,64,128$, only NN.

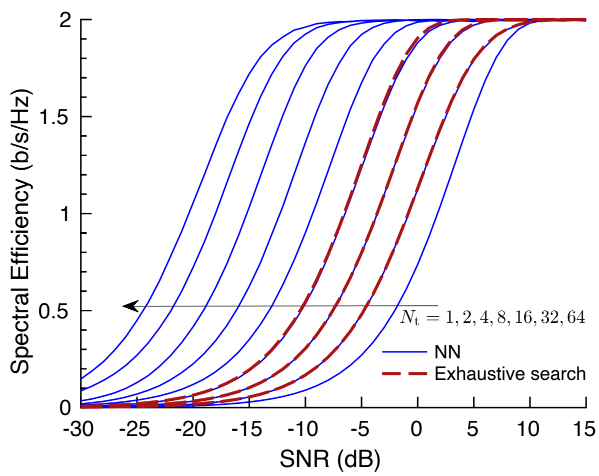

Fig. 4: Average spectral efficiency vs SNR over an LOS channel. For $N_\mathrm{t}=1$, 2, 4 8: NN (in solid) versus exhaustive search (in dashed). For $N_\mathrm{t}=16,32,64,128$, only NN. The transmit array is a ULA with $d_\mathrm{t}=\lambda/2$.

V. Performance Evaluation

Presented in Fig. 3 is the ergodic spectral efficiency as a function of SNR in IID Rayleigh fading. Up to $N_\mathrm{t}=8$, we are able to evaluate (20) exhaustively, confirming the excellent performance of the learning approach. By about $N_\mathrm{t}=16$, an exhaustive search becomes prohibitive, yet the learning approach continues to function. The figure includes the spectral efficiency for N_t as high as 128, when the number of candidate quartets exceeds a staggering 10^{76}. The performance improves steadily and settles onto a 3-dB SNR reduction for every doubling of N_t, the same beamforming gain that would be attained with full-resolution DACs and ADCs [22].

A similar set of results is presented in Fig. 4 for an LOS channel. Since, as mentioned, the array orientation may change slowly enough that the LOS channel is information stable over each value of θ, the curves in Fig. 4 are

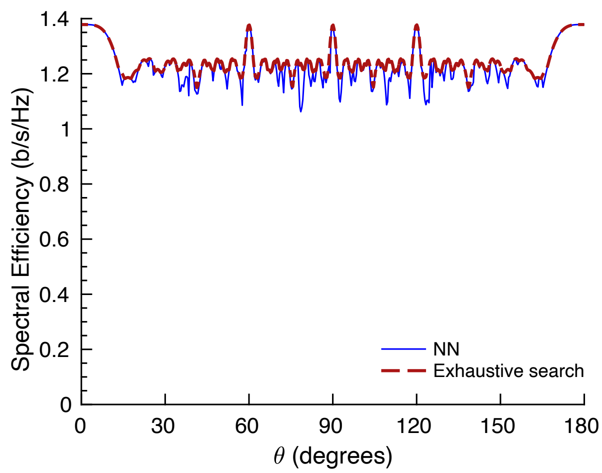

Fig. 5: Spectral efficiency vs θ for a LOS channel with $N_{\rm t} = 8$ and SNR $= -5$ dB. The transmit array is a ULA with $d_{\rm t} = \lambda/2$.

best interpreted as the average spectral efficiency over $\theta \in [0, \pi/2]$. The performance as a function of θ is then also of interest. For $N_{\rm t} = 8$ and SNR $= -5$ dB, this performance is illustrated in Fig. 5. Also in this angular fashion, the performance of the learning approach closely matches its exhaustive-search counterpart.

VI. Extension to Multiple Receive Antennas

The learning approach presented hitherto applies also for $N_{\rm r} > 1$, with the channel then being an $N_{\rm r} \times N_{\rm t}$ matrix. For rank-1 channels specifically, such matrix can be expressed as $\boldsymbol{H} = \sigma \boldsymbol{u} \boldsymbol{v}^*$ where σ is the singular value and $\boldsymbol{u}$, $\boldsymbol{v}$, are the singular vectors. The procedure laid down throughout the paper to identify the optimum transmit beamforming quartet of vectors continues to apply, only with $\boldsymbol{v}$ in place of $\boldsymbol{h}$.

VII. Conclusion

Channel estimation is an important aspect that needs to be addressed to consolidate the findings in this paper. In IID fading, $N_{\rm t}$ complex coefficients have to be estimated to obtain $\boldsymbol{h}$. However, in LOS conditions, much more prevalent at terahertz frequencies, one or two geometrical parameters suffice to reconstruct $\boldsymbol{h}$ under the premise of planar wavefronts, and at most four geometrical parameters suffice with spherical wavefronts [23]. While, at present, we are feeding the reconstructed $\boldsymbol{h}$ to the NN, follow-up work will aim at having this handful of parameters serve directly as inputs to the NN, in lieu of $\boldsymbol{h}$. This might allow for even further simplification of the NN and the learning process.

Acknowledgment

This work was supported by the European Research Council under the H2020 Framework Programme/ERC grant agreement 694974, by MINECO's Projects RTI2018-102112 and 101040, and by the ICREA Academia program.

References

[1] I. F. Akyildiz, J. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Commun.*, vol. 12, pp. 16–32, 2014.

[2] H. Hamada *et al.*, "300-GHz-band 120-Gb/s wireless front-end based on InP-HEMT PAs and mixers," *IEEE J. Solid-State Circuits*, vol. 55, no. 9, pp. 2316–2335, 2020.

[3] L. Ribeiro, S. Schwarz, M. Rupp, and A. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Signal Processing*, vol. 12, no. 2, pp. 298–312, 2018.

[4] B. Murmann, "The race for the extra decibel: A brief review of current ADC performance trajectories," *IEEE Solid-State Circuits Magazine*, vol. 7, no. 3, pp. 58–66, 2015.

[5] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Int'l Conf. Wireless Commun. and Mobile Computing*, 2006, pp. 1387–1392.

[6] A. Mezghani and J. A. Nossek, "Analysis of Rayleigh-fading channels with 1-bit quantized output," in *IEEE Int'l Symp. Inform. Theory (ISIT)*, 2008, pp. 260–264.

[7] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, 2009.

[8] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Processing*, vol. 63, no. 20, pp. 5498–5512, 2015.

[9] A. Mezghani, R. Ghiat, and J. Nossek, "Transmit processing with low resolution D/A-converters," in *IEEE Int'l Conf. Electronics, Circuits and Systems-(ICECS'09)*, 2009, pp. 683–686.

[10] B. Usman, H. Jedda, A. Mezghani, and J. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *IEEE Int'l Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, 2016, pp. 3381–3385.

[11] H. Jedda, J. A. Nossek, and A. Mezghani, "Minimum BER precoding in 1-bit massive MIMO systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2016.

[12] J. Guerreiro, R. Dinis, and P. Montezuma, "Use of 1-bit digital-to-analogue converters in massive MIMO systems," *Electr. Letters*, vol. 52, no. 9, pp. 778–779, 2016.

[13] Y. Li, C. Tao, L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Commun. Letters*, vol. 21, no. 7, pp. 1669–1672, 2017.

[14] K. Gao, N. Estes, B. Hochwald, J. Chisum, and N. Laneman, "Power-performance analysis of a simple one-bit transceiver," in *Inform. Theory and Applic. Workshop (ITA)*, 2017.

[15] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, 2017.

[16] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Processing*, vol. 65, no. 17, pp. 4624–4634, 2017.

[17] Y. Nam, H. Do, Y.-S. Jeon, and N. Lee, "On the capacity of MISO channels with one-bit ADCs and DACs," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2132–2145, 2019.

[18] A. Lozano, "1-bit MIMO for terahertz channels," in *Int'l ITG Workshop on Smart Antennas*, 2021.

[19] K. Gao, J. N. Laneman, and B. Hochwald, "Beamforming with multiple one-bit wireless transceivers," in *Inform. Theory and Applic. Workshop (ITA)*, 2018.

[20] R. Nikbakht, A. Jonsson, and A. Lozano, "Unsupervised learning for parametric optimization," *IEEE Commun. Letters*, vol. 25, no. 3, pp. 678–681, 2021.

[21] A. Lozano and N. Jindal, "Are yesterday's information-theoretic fading models and performance metrics adequate for the analysis of today's wireless systems?" *IEEE Commun. Magazine*, vol. 50, no. 11, pp. 210–217, Nov. 2012.

[22] A. Tulino, A. Lozano, and S. Verdú, "MIMO capacity with channel state information at the transmitter," in *IEEE Int'l Symp. Spread Spectrum Techn. and Applic.*, 2004, pp. 22–26.

[23] H. Do, N. Lee, and A. Lozano, "Reconfigurable ULAs for line-of-sight MIMO transmission," *IEEE Trans. Wireless Commun.*, vol. 20, 2021.