



**HAL**  
open science

# Efficient Deep Unfolding for SISO-OFDM Channel Estimation

Baptiste Chatelier, Luc Le Magoarou, Getachew Redieteb

► **To cite this version:**

Baptiste Chatelier, Luc Le Magoarou, Getachew Redieteb. Efficient Deep Unfolding for SISO-OFDM Channel Estimation. IEEE international conference on communications (ICC), May 2023, Roma, Italy. hal-03810183

**HAL Id: hal-03810183**

**<https://hal.science/hal-03810183>**

Submitted on 11 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Efficient Deep Unfolding for SISO-OFDM Channel Estimation

Baptiste Chatelier<sup>†‡</sup>, Luc Le Magoarou<sup>†‡</sup>, Getachew Redieteb<sup>§‡</sup>  
<sup>†</sup>Univ Rennes, INSA Rennes, CNRS, IETR-UMR 6164, Rennes, France  
<sup>‡</sup>b<>com, Rennes, France  
<sup>§</sup>Orange Innovation, Rennes, France  
baptiste.chatelier@insa-rennes.fr

**Abstract**—In modern communication systems, channel state information is of paramount importance to achieve capacity. It is then crucial to accurately estimate the channel. It is possible to perform SISO-OFDM channel estimation using sparse recovery techniques. However, this approach relies on the use of a physical wave propagation model to build a dictionary, which requires perfect knowledge of the system’s parameters. In this paper, an unfolded neural network is used to lighten this constraint. Its architecture, based on a sparse recovery algorithm, allows SISO-OFDM channel estimation even if the system’s parameters are not perfectly known. Indeed, its unsupervised online learning allows to learn the system’s imperfections in order to enhance the estimation performance. The practicality of the proposed method is improved with respect to the state of the art in two aspects: constrained dictionaries are introduced in order to reduce sample complexity and hierarchical search within dictionaries is proposed in order to reduce time complexity. Finally, the performance of the proposed unfolded network is evaluated and compared to several baselines using realistic channel data, showing the great potential of the approach.

**Index Terms**—Deep Unfolding, Frugal AI, SISO-OFDM channel estimation, Sparse recovery

## I. INTRODUCTION

Recently, machine learning techniques have emerged as a promising solution in many wireless communications areas [1], [2] such as beamforming [3], [4], channel charting [5]–[7] or channel mapping [8], [9]. The Deep Unfolding machine learning approach [10]–[13] is very promising as it profits from both the controlled complexity of classical signal processing approaches and the flexibility of machine learning techniques. Indeed, this approach considers that iterative algorithms can be unfolded as neural networks, that can be trained, where each layer represents one iteration of the algorithm.

In the field of channel estimation, classical statistical estimation techniques such as the Least Squares (LS) or the Minimum Mean Squared Error (MMSE) have been extensively used in the past. However, those techniques suffer from several drawbacks: on one hand, for the LS estimator, there is a huge Mean Squared Error (MSE), on the other hand, for the MMSE estimator, there is a huge computational complexity. Other approaches have been envisioned to counter those drawbacks: one of them is the usage of *sparse recovery* algorithms. It is well known that propagation channels are dominated by a few strong propagation paths: the channel is then said to be *sparse*. It has been shown that one can use

that sparsity notion to propose channel estimation strategies relying on sparse recovery algorithms such as Matching Pursuit (MP) [14]–[17]. However, those techniques rely on the knowledge of a physical wave propagation model, and it has been shown in [17] that a small uncertainty on system parameters could lead to high estimation performance losses.

**Contributions and related work.** In [17], the Deep Unfolding approach was used to solve this performance loss issue, with the unfolded network mpNet, allowing to learn the physical imperfections of the system so as to enhance the estimation performance. However, this approach yields a high number of parameters to learn, leading to both a high sample complexity and a high time complexity. In this paper, mpNet is adapted to constrained dictionaries, which allows to drastically reduce sample complexity. Moreover, hierarchical search over dictionaries is introduced in order to reduce time complexity. Furthermore, the proposed method is illustrated on a Single Input Single Output Orthogonal Frequency Division Multiplexing (SISO-OFDM) system here instead of a Multi-User Multiple Input Multiple Output (MU-MIMO) system as in [17]. However, note that the contributions apply equally to both types of systems.

The rest of the paper is organized as follows, Section II presents the SISO-OFDM channel estimation problem, the used physical wave propagation model, and the sparse recovery approach. Section III presents the unfolded neural network and the two contributions of this paper: the sample complexity reduction strategy, and the time complexity reduction strategy. Section IV presents simulation results. Finally, Section V concludes the paper and gives perspectives.

## II. PROBLEM FORMULATION

### A. System model

In this paper, a SISO-OFDM scenario with  $N$  subcarriers in the uplink is considered. Let  $\mathbf{g} \in \mathbb{C}^N$  denote the antenna gain vector representing the Base Station (BS) antenna gain at each subcarrier frequency, and  $\mathbf{f} \in \mathbb{R}^N$  denote a frequency vector containing the different subcarrier frequencies. Finally, a multipath channel with  $L_p$  propagation paths is considered.

Let  $\mathbf{x} \in \mathbb{C}^N$  be an LS channel estimate of the channel vector  $\mathbf{h} \in \mathbb{C}^N$ . This estimate is noisy as it possesses a

residual estimation error:

$$\mathbf{x} = \mathbf{h} + \mathbf{n}, \quad (1)$$

where  $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$ . Its Signal to Noise Ratio (SNR) can be computed as:

$$\text{SNR}_{\text{in}} = \frac{\|\mathbf{h}\|_2^2}{N\sigma^2}. \quad (2)$$

The goal of this paper is to denoise this LS channel estimate using sparse recovery techniques and a physical model.

### B. Physical model

For a given subcarrier  $f_k$ , the channel coefficient can be written as:

$$h_k = \sum_{l=1}^{L_p} \alpha_l g_k e^{-j2\pi f_k \tau_l}, \quad (3)$$

where  $\alpha_l$  and  $\tau_l$  respectively represent the complex attenuation coefficient and the propagation delay for the  $l^{\text{th}}$  path.

Under this model, the channel vector  $\mathbf{h} \in \mathbb{C}^N$  can be expressed as:

$$\mathbf{h} = \sum_{l=1}^{L_p} \alpha_l \begin{bmatrix} g_1 e^{-j2\pi f_1 \tau_l} \\ \vdots \\ g_N e^{-j2\pi f_N \tau_l} \end{bmatrix} = \sum_{l=1}^{L_p} \alpha_l \boldsymbol{\psi}(\tau_l), \quad (4)$$

where  $\boldsymbol{\psi}(\tau_l)$  is the *frequency response vector* (FRV) for the  $l^{\text{th}}$  path. We can see that the channel vector can be defined as a linear combination of FRVs. The linear combination is said to be sparse when  $L_p$  is small.

### C. Sparse recovery approach

As the channel can be written as a linear combination of FRVs, it is possible to construct a fixed matrix  $\boldsymbol{\Psi} \in \mathbb{C}^{N \times A}$  with  $A$  FRV columns, which amounts to the discretization of the delays, and a projection vector  $\mathbf{u} \in \mathbb{C}^A$  such that the channel could be estimated as:

$$\hat{\mathbf{h}} = \boldsymbol{\Psi} \mathbf{u}. \quad (5)$$

For the rest of this paper, the FRV matrix  $\boldsymbol{\Psi}$  will be called *dictionary* and its columns will be called *atoms*, i.e. each atom will represent a FRV corresponding to a specific delay. As the channel vector can be seen as a linear combination of a few FRVs, the channel vector is said to be sparse in a dictionary of FRVs. For a fixed dictionary  $\boldsymbol{\Psi}$ , the sparse recovery optimization problem is:

$$\begin{aligned} \min_{\mathbf{u}} \quad & \|\mathbf{h} - \boldsymbol{\Psi} \mathbf{u}\|_2^2 \\ \text{s.t.} \quad & \|\mathbf{u}\|_0 \leq A \end{aligned} \quad (6)$$

The previous optimization problem makes sense only if the dictionary is correct, i.e. a sparse linear combination of dictionary columns can represent a channel vector. If this is not the case, the channel estimate won't be trustworthy. It is possible to find the optimal dictionary as:

$$\boldsymbol{\Psi}^* = \arg \min_{\boldsymbol{\Psi}} \mathbb{E}_{\mathbf{h}} \left[ \min_{\mathbf{u}, \|\mathbf{u}\|_0 \leq A} \|\mathbf{h} - \boldsymbol{\Psi} \mathbf{u}\|_2^2 \right]. \quad (7)$$

Solving Eq. (6) for a fixed FRV dictionary is possible through sparse recovery algorithm such as the MP algorithm [18]: this is the approach followed in this paper, as in [17]. The goal of this algorithm is to find the most correlated FRVs, i.e. atoms, of the dictionary with the noisy channel estimate and subtract their projections. After  $L_p$  iterations of the algorithm, in a perfect scenario, as all the corresponding FRVs will have been subtracted to the residual, the residual will only be composed of noise.

---

### Algorithm 1 MP algorithm for channel denoising

---

**Input:** Dictionary  $\boldsymbol{\Psi}$ , Noisy LS estimate  $\mathbf{x}$ , Tolerance level  $\epsilon$

1: Initialize the residual:  $\mathbf{r} \leftarrow \mathbf{x}$

2: **while**  $\|\mathbf{r}\|_2^2 > \epsilon$  **do**

3: Find the most correlated atom:  $i^* \leftarrow \arg \max_i \left| \boldsymbol{\psi}_i^H \mathbf{r} \right|$

4: Update the residual:  $\mathbf{r} \leftarrow \mathbf{r} - \boldsymbol{\psi}_{i^*} \boldsymbol{\psi}_{i^*}^H \mathbf{r}$

5: **end while**

**Output:**  $\hat{\mathbf{h}} \leftarrow \mathbf{x} - \mathbf{r}$  (Denoised LS estimate)

---

### D. Hardware impairments

In reality, the FRVs are not exactly known as there exists a lot of hardware imperfections related to frequency generation/acquisition. More specifically, Carrier Frequency Offset (CFO) and Sampling Clock Offset (SCO) can offset the subcarrier frequencies. Let  $\tilde{f}_i$  be the *nominal*  $i^{\text{th}}$  subcarrier frequency, i.e. the non-offsetted subcarrier frequency, and  $f_i$  be the *real*  $i^{\text{th}}$  subcarrier frequency, i.e. the potentially offsetted subcarrier frequency. For the CFO, the frequency offset  $\delta f$  is common to all subcarriers:

$$\forall i \in \left[ -\frac{N}{2}, \frac{N}{2} \right], f_i = \tilde{f}_i + \delta f. \quad (8)$$

For the SCO, it has been shown in [19], [20] that the frequency offset is dependent on the subcarrier index, on the oscillator part per million (ppm) value  $\xi$ , and on the subcarrier spacing  $\Delta f$ :

$$\forall i \in \left[ -\frac{N}{2}, \frac{N}{2} \right], f_i = \tilde{f}_i + i\delta f = \tilde{f}_i + i\xi\Delta f. \quad (9)$$

Moreover, we consider an antenna gain imperfection. Let  $\tilde{g}_i$  be the nominal antenna gain, and  $g_i$  be the real antenna gain for the  $i^{\text{th}}$  subcarrier. We obtain:

$$\forall i \in \left[ -\frac{N}{2}, \frac{N}{2} \right], g_i = \tilde{g}_i + n_{g_i}, \quad (10)$$

with  $n_{g_i} \sim \mathcal{N}(0, \sigma_g^2)$ .

For the rest of this paper, we will only consider SCO and antenna gain imperfections. We can then define the notions of *nominal* and *real dictionaries*. The nominal dictionary  $\boldsymbol{\Psi}$  will represent the dictionary constructed from the nominal knowledge of the system parameters, which is unaware of impairments. On the other hand, the real dictionary will represent the dictionary constructed from the perfect

knowledge of the system parameters, i.e. with full knowledge of the SCO and antenna gain imperfections. The proposed method consists in initializing a neural network using the nominal dictionary, and approaching the performance one would get using the real dictionary via learning.

### III. PROPOSED APPROACH

#### A. mpNet architecture

Using the same approach as in [17], we propose to use the unfolded neural network mpNet to achieve SISO-OFDM channel estimation while learning the physical imperfections. The network is the unfolded version of the MP iterative algorithm and is presented in Algorithm 2.

---

#### Algorithm 2 Forward pass of mpNet [17]

---

**Input:** Weight matrix  $\mathbf{W} \in \mathbb{C}^{N \times A}$ , Noisy LS estimate  $\mathbf{x} \in \mathbb{C}^N$ , Noise variance  $\sigma^2$ , Number of subcarriers  $N$

- 1:  $\mathbf{r} \leftarrow \mathbf{x}$
- 2:  $\epsilon \leftarrow \frac{\sigma^2 N}{\|\mathbf{x}\|_2^2}$
- 3: **while**  $\|\mathbf{r}\|_2^2 > \epsilon$  **do**
- 4:    $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{WHT}_1(\mathbf{W}^H \mathbf{r})$
- 5: **end while**

**Output:**  $\hat{\mathbf{h}} \leftarrow \mathbf{x} - \mathbf{r}$  (Denoised LS estimate)

---

$\mathbf{HT}_1$  is a non-linear hard-thresholding operator which keeps only its input of greatest modulus and sets all the others to zero. The weight matrix  $\mathbf{W}$  is initialized with the nominal dictionary  $\tilde{\Psi}$ . This thoughtful initialization allows to reduce the cost function convergence time in comparison to a random initialization. The goal is to learn the optimal dictionary through gradient descent in the unfolded network. In brief, this network has two goals:

- The forward pass of mpNet performs the channel estimation/denoising operation as the architecture of mpNet is the unfolded MP algorithm. Each layer will estimate one path of the channel.
- The backward pass of mpNet performs the dictionary learning. As the weight matrix of mpNet is initialized with the nominal dictionary, the goal of the backward pass is to update the weight matrix so that it tends towards the real dictionary.

Moreover, the value  $\epsilon$  in Algorithm 2 corresponds to a stopping criterion called  $\text{SC}_2$  in [17]:

$$\text{SC}_2 : \|\mathbf{r}\|_2^2 \leq \frac{\sigma^2}{\|\mathbf{x}\|_2^2} N. \quad (11)$$

This stopping criterion allows the network to have an adaptive number of layers for given noise characteristics: in other words, SNR adaptability. However, it imposes the knowledge of the noise variance  $\sigma^2$  at the BS side.

#### B. Reducing sample complexity: constrained dictionary

As we explained above, the learning parameters are the components of the weight matrix  $\mathbf{W} \in \mathbb{C}^{N \times A}$ . As those components are complex, each entry of  $\mathbf{W}$  possesses both

a magnitude and a phase: as a result there are  $2NA$  learning parameters. In the experimental setup that will be presented in Section IV, we will consider  $N = 256$  subcarriers, and  $A = 990$  atoms. In that setup, there are  $2NA = 506,880$  learning parameters. The first important contribution of this paper is to reduce this number by using a *constrained* dictionary.

We know that the proposed SISO-OFDM system only considered imperfections that impacted the gains and frequencies of the dictionary. Moreover, gain and frequencies are expected to be independent of the propagation delays, meaning that they will be common for each atom of the dictionary. We can then propose the following definition for the constrained dictionary  $\tilde{\Psi} \in \mathbb{C}^{N \times A}$ :

$$\tilde{\Psi} = \begin{bmatrix} g_1 e^{-j2\pi(f_1 - \frac{N}{2}\delta f)\tau_1} & \dots & g_1 e^{-j2\pi(f_1 - \frac{N}{2}\delta f)\tau_A} \\ \vdots & \vdots & \vdots \\ g_N e^{-j2\pi(f_N + \frac{N}{2}\delta f)\tau_1} & \dots & g_N e^{-j2\pi(f_N + \frac{N}{2}\delta f)\tau_A} \end{bmatrix} \quad (12)$$

Each column of the dictionary still represents an FRV, however each entry of this constrained matrix is no more learnable: the only learnable parameters are the complex antenna gains  $\mathbf{g}$  and the SCO frequency offset  $\delta f$ . From  $2NA$  parameters, we only have  $2N + 1$  learning parameters left. This is interesting as *the number of learning parameters is now independent of the number of atoms*. If we take the previous numerical example, we go from  $2NA = 506,880$  to only  $2N + 1 = 513$  learning parameters.

One of the other benefits of this constrained dictionary approach is its sample complexity, i.e. the number of required samples to achieve convergence of the network's cost function. Indeed, for the classical mpNet architecture, as each component of the weight matrix was independent, only the atoms selected in the forward pass were updated during backpropagation. For the proposed approach, as the frequency offset and gains are common over all atoms, the selection and update of one atom will result in the update of all other atoms. Consequently, the cost function convergence will be quicker with this constrained dictionary approach than with the classical mpNet architecture, as shown in the experimental part of this paper.

#### C. Reducing time complexity: hierarchical search

A quick complexity analysis allows us to state that the time complexity of the classical mpNet approach is  $\mathcal{O}(DNA)$  where  $D$  represents the number of layers of the mpNet network. If we take the previously considered scenario with  $D = 10$  layers, we obtain at least  $DNA = 2,534,400$  arithmetic operations for the estimation of one SISO-OFDM channel. In order to reduce this time complexity, we propose a new unfolded network architecture based on hierarchical atom search.

The principle of the hierarchical atom search approach is presented in Fig. 1. The classical approach consists in finding the most correlated atom in a dictionary of  $A$  different atoms (represented as circles in Fig. 1):  $A$  different correlations are carried out. Our hierarchical approach consists in iteratively finding the most correlated atom in the dictionary. As in the SISO-OFDM scenario the atoms are

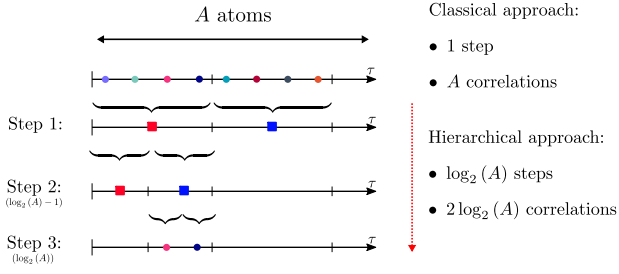


Fig. 1: Hierarchical atom search

parameterized with a propagation delay, the dictionary covers an interval in the delay domain. In the first step, we propose to construct two *meta-atoms* (represented as squares in Fig. 1) whose correlation responses separately cover the dictionary-associated delay interval. We carry out the correlation with the two meta-atoms and choose the most correlated one. We then create two new meta-atoms that cover the delay interval associated with the most correlated meta-atom of the previous step. This process is repeated until the  $\log_2(A)$ <sup>th</sup> step. In the delay interval covered by the most correlated atom chosen in the  $\log_2(A) - 1$ <sup>th</sup> step, there are only 2 atoms from the original dictionary: we compute their correlations and then find the most correlated atom of the dictionary. In other words, rather than finding the most correlated atom in 1 step but with  $A$  correlations, we find the most correlated atom in  $\log_2(A)$  steps but with only  $2\log_2(A)$  correlations.

The challenge of this approach is to propose a meta-atom structure that allows its correlation response to cover a specific delay interval. We can prove that using a sinc modulated FRV as a meta-atom allows this.

A Fourier Transform FT from the frequency domain to the delay domain is defined as:

$$x(\tau) = \text{FT}[x(f)] = \int_{\mathbb{R}} x(f) e^{-j2\pi f\tau} df. \quad (13)$$

The correlation of a pure FRV of delay  $\tau$ ,  $e(f, \tau)$ , and a generic meta-atom  $\psi_i(f)$  is considered. The bandwidth, central frequency, and subcarrier spacing of the system will respectively be denoted as  $\gamma$ ,  $f_0$ , and  $\Delta f$ . The correlation response can be computed as:

$$\chi_{\psi_i}(\tau) = \langle e(f, \tau), \psi_i(f) \rangle \quad (14)$$

$$= \int_{\mathbb{R}} \psi_i^*(f) g(f) e^{-j2\pi f\tau} \Pi_{\gamma}(f - f_0) \text{III}_{\Delta f}(f) df. \quad (15)$$

The windowing  $\Pi_{\gamma}(f - f_0)$  allows to represent the band-limited nature of the FRV and meta-atom, while the Dirac comb  $\text{III}_{\Delta f}(f)$  allows the sampling at the subcarrier frequencies. We make the assumption that the antenna gains are quasi-constant over the bandwidth so that  $g(f) = \alpha \Pi_{\gamma}(f - f_0)$ . In order to simplify the following equations, we will deliberately omit the constant  $\alpha$ . We then obtain:

$$\chi_{\psi_i}(\tau) = \text{FT}[\psi_i^*(f) \Pi_{\gamma}(f - f_0) \text{III}_{\Delta f}(f)]. \quad (16)$$

One can observe in Eq. (16) that the correlation response can be computed as a Fourier Transform. As we want it to be a constant over a certain delay interval (a rect function), it is then easy to see that using sinc modulated FRVs allows to fulfill that requirement. More specifically, we define the meta-atom  $\psi_i(f)$  as:

$$\psi_i(f) = \text{sinc}(\varpi f) e^{-j2\pi f\tau_i}. \quad (17)$$

The correlation response then becomes:

$$\begin{aligned} \chi_{\psi_i}(\tau) &= \text{FT}[\text{sinc}(\varpi f) e^{j2\pi f\tau_i} \Pi_{\gamma}(f - f_0) \text{III}_{\Delta f}(f)] \\ &= \frac{1}{\varpi} \Pi_{\varpi}(\tau - \tau_i) \otimes \gamma \text{sinc}(\gamma\tau) e^{-j2\pi f_0\tau} \otimes \frac{1}{\Delta f} \text{III}_{\frac{1}{\Delta f}}(f). \end{aligned} \quad (18)$$

One can see in Eq. (18) that the proposed meta-atom expression allows to obtain a rectangular waveform in the delay domain,  $\varpi$  allowing the rect width control and  $\tau_i$  the rect central position control. This can be seen in Fig. 2 where we vary the values of  $\varpi$  and  $\tau_i$ .

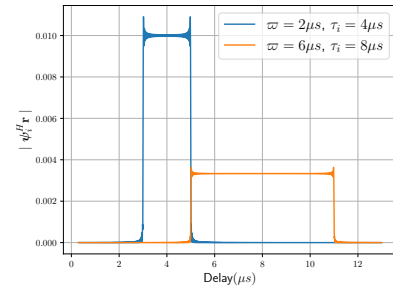


Fig. 2: Influence of  $\varpi$  and  $\tau_i$  on the correlation waveform

For two meta-atoms, the time complexity of the proposed hierarchical search can be evaluated as  $\mathcal{O}(2\log_2(A))$ . If we consider the partition of  $A > 1$  atoms into  $n \in \mathbb{N}$ ,  $n > 1$  intervals, the time complexity of the associated hierarchical search is  $\mathcal{O}(n \log_n(A))$ . Minimizing the time complexity of this hierarchical approach is then equivalent to solving Eq. (19):

$$n^* = \arg \min_{n \in \mathbb{N}, n > 1} n \log_n(A). \quad (19)$$

It can be easily shown that solving Eq. (19) gives us  $n^* = 3$ .

A quick complexity analysis of all the unfolded networks presented so far is presented in Table I. We compare the classical mpNet as it was proposed in [17], mpNet with a constrained dictionary (C. mpNet), and mpNet with a constrained dictionary and hierarchical atom search with 3 meta-atoms (H.C.3 mpNet). As we stated earlier, the constrained dictionary approach allows to reduce the sample complexity and the hierarchical search approach was developed to reduce the time complexity. It is then interesting to merge the constrained and hierarchical search approach, as it allows to reduce both the sample and time complexities. For  $D = 10$  layers,  $N = 256$  subcarriers and  $A = 990$  atoms, the time complexity of the forward pass goes from 2,534,400 arithmetic operations for the classical mpNet to around only 48,220 arithmetic operations for the H.C.3 mpNet, which is a fifty-fold reduction.

		mpNet	C. mpNet	H.C.3 mpNet
Time complexity	Forward	$\mathcal{O}(DNA)$	$\mathcal{O}(DNA)$	$\mathcal{O}(DN3 \log_3(A))$
	Backward	$\mathcal{O}(DN)$	$\mathcal{O}(DN)$	$\mathcal{O}(DN)$
Sample complexity		+++	+	+

TABLE I: Unfolded network complexity analysis

#### IV. EXPERIMENTAL RESULTS

We propose to validate our contributions using realistic channel data, namely the ‘‘O1’’ outdoor DeepMIMO dataset [21]. We consider a central frequency of 3.4GHz, and a bandwidth of 50MHz. Furthermore, we generate the channels with  $N = 256$  subcarriers. The dictionary is composed of  $A = 990$  atoms. We consider up to  $L_p = 10$  propagation paths. Regarding the physical imperfections, the oscillator ppm value for the SCO is  $\xi = 40$ ppm, and we consider two gain noise variance scenarios:  $\sigma_g^2 = 0.36$  (very high gain noise) and  $\sigma_g^2 = 0.09$  (high gain noise). We consider two SNR scenarios:  $\text{SNR}_{\text{in}} = 10$ dB and  $\text{SNR}_{\text{in}} = 5$ dB. In order to compute the stopping criterion for the network, we assume that the noise variance is known at the BS side.

We consider the same minibatch online learning strategy as in [17]. The network will see batches of 10 channels, and the estimation performance will be evaluated on 2000 test channels. More specifically, regarding the DeepMIMO parameters, we consider the BS represented with a red triangle in Fig. 3, and User Equipments (UEs) randomly generated in the blue zone. We will use the Adam optimizer [22] for the learning parameter update. The evaluation metric will be the Normalized Mean Squared Error (NMSE), defined as  $\mathbb{E} \left[ \frac{\|\hat{\mathbf{h}} - \mathbf{h}\|_2^2}{\|\mathbf{h}\|_2^2} \right]$ .

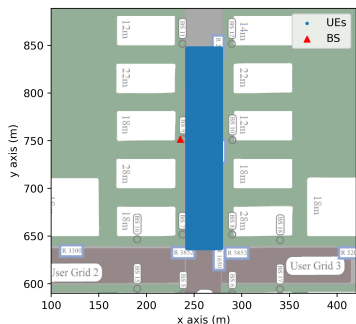


Fig. 3: DeepMIMO channel generation configuration

We propose to evaluate the performance of the classical network (mpNet), the constrained dictionary mpNet (C. mpNet), the hierarchical search with 2 meta-atoms and constrained dictionary mpNet (H.C.2 mpNet) and the one with 3 meta-atoms (H.C.3 mpNet). We will compare those network performances against the LS estimator, the MP algorithm with nominal and real dictionaries, and the MMSE estimator with a substituted channel covariance matrix using

the Low-Rank Approximation (LRA) presented in [23]. We use the substituted covariance matrix, as, in reality, estimating the covariance matrix for each UE is too complex as it requires tracking each UE channel during several frames.

For all noise and hardware scenarios, it is possible to remark in Fig. 4 that dictionary learning is achieved as there is convergence of the NMSE performances. One can remark that at the end of the training, the proposed networks (C. mpNet, H.C.2 and H.C.3 mpNet) always outperform the LS and LRA-MMSE approaches.

One can see that, as exposed in Section III-B, the convergence of the constrained versions of mpNet (C. mpNet and H.C. mpNet) is quicker than the convergence of the classical mpNet. For all scenarios, only approximately 400 channels are needed to achieve convergence. Moreover, we can see that, after convergence, the estimation performances of the constrained networks are better than the one of the classical mpNet. This can be explained by the bias-variance decomposition of the estimation NMSE. In fact, the C./H.C. mpNet can be seen as an estimator with fewer parameters than the classical mpNet, so its variance will be lower than the classical mpNet one. As the dictionary structure is chosen wisely, the C./H.C. mpNet has the same bias as the classical mpNet. As a result, the estimation NMSE is lower for both the C. and H.C. mpNet architectures than for the classical one.

One can also remark in Fig. 4c that, if the hardware imperfections are too important, the model-based MP algorithm does not perform well. In that particular case, using the MP approach with nominal dictionary is worse than using LS.

It is noticeable that H.C.2 mpNet offer slightly worse performance than the C. mpNet. This can be explained by the hierarchical search approach: in the classical approach, the algorithm finds the most correlated atom of a dictionary with a residual. For the hierarchical approach, the algorithm will select the delay interval that has the most energy. As a result, for a fixed noisy LS estimate, the two approaches may not select the same atoms, leading to different channel estimation performance. It is then interesting to study the impact of the number of meta-atoms, as increasing the number of windows in the delay domain can improve the estimation performance for the hierarchical approach. This can be seen with the NMSE performance of H.C.3 mpNet: this network offers slightly better performance than the H.C.2 mpNet, while presenting a lower time complexity as shown in Table I.

It has been shown in Section III-C that the hierarchical approach with 3 meta-atoms was optimal in terms of time complexity. This can be seen in Table II where an execution time comparison of the MP algorithm to its hierarchical versions with 2 and 3 meta-atoms (MP H.2 and MP H.3) is proposed. One can see that the MP H.3 outperforms both the MP H.2 and classical MP algorithm: its mean execution time is up to a hundred time quicker than the classical MP.

#### V. CONCLUSION

In this paper, the unfolded neural network mpNet [17] was enhanced in two ways. Firstly, constrained dictionaries were

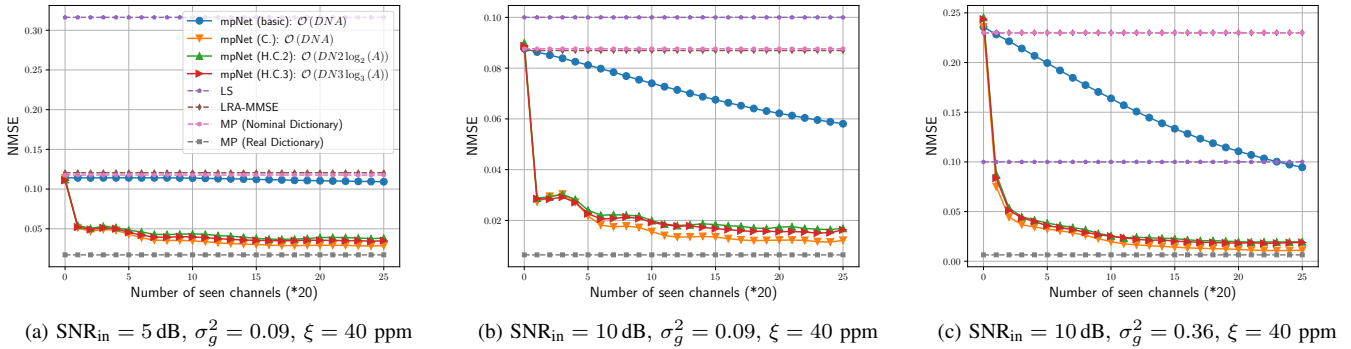


Fig. 4: Channel estimation performance for various SNRs and model imperfections

	MP	MP H.2	MP H.3	
$\mathbb{E}[t_{\text{exec}}]$ (s)	$A = 990$	0.048	0.014	<b>0.013</b>
	$A = 5,000$	0.263	0.019	<b>0.015</b>
	$A = 50,000$	2.866	0.024	<b>0.020</b>

TABLE II: Average execution time for 10,000 random UEs

introduced to allow sample complexity reduction. Secondly, hierarchical atom search has been presented to reduce time complexity. The performance of the resulting network was assessed using realistic channel data with hardware impairments: it showed that our network outperforms classical approaches such as the LS estimator or the MMSE estimator with substituted covariance matrix, but also the classical mpNet network. This combined approach is really promising as it allows to enhance the estimation performance while reducing the overall complexity: this follows the efficiency philosophy of model-based deep learning [10]. Future work could for example consider the adaptation of the proposed techniques to MU-MIMO-OFDM scenarios.

## REFERENCES

- [1] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.
- [2] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, “Deep learning for wireless physical layer: Opportunities and challenges,” *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.
- [3] L. Le Magoarou, T. Yassine, S. Paquelet, and M. Crussière, “Deep learning for location based beamforming with NLOS channels,” in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2022, pp. 8812–8816.
- [4] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, “Deep learning coordinated beamforming for highly-mobile millimeter wave systems,” *IEEE Access*, vol. 6, pp. 37 328–37 348, 2018.
- [5] T. Yassine, L. Le Magoarou, S. Paquelet, and M. Crussière, “Leveraging triplet loss and nonlinear dimensionality reduction for on-the-fly channel charting,” in *IEEE Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*, 2022.
- [6] C. Studer, S. Medjkouh, E. Gönültaş, T. Goldstein, and O. Tirkkonen, “Channel charting: Locating users within the radio environment using channel state information,” *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.
- [7] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, “Triplet-based wireless channel charting: Architecture and experiments,” *IEEE J. Sel. Areas Commun.*, vol. 39, pp. 2361–2373, 2021.
- [8] L. Le Magoarou, “Similarity-based prediction for channel mapping and user positioning,” *IEEE Commun. Lett.*, vol. 25, pp. 1578–1582, 2021.
- [9] M. Alrabeiah and A. Alkhateeb, “Deep learning for TDD and FDD Massive MIMO: Mapping channels in space and frequency,” in *2019 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 1465–1470.
- [10] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, “Model-based deep learning,” *arXiv preprint arXiv:2012.08405*, Dec. 2020.
- [11] V. Monga, Y. Li, and Y. C. Eldar, “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing,” *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.
- [12] J. R. Hershey, J. L. Roux, and F. Wenginger, “Deep unfolding: Model-based inspiration of novel deep architectures,” *arXiv preprint arXiv:1409.2574*, 2014.
- [13] A. Balatsoukas-Stimming and C. Studer, “Deep unfolding for communications systems: A survey and some new directions,” in *2019 IEEE Int. Workshop on Signal Process. Syst. (SiPS)*, 2019, pp. 266–271.
- [14] T. Kang and R. A. Iltis, “Matching pursuits channel estimation for an underwater acoustic OFDM modem,” in *2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2008, pp. 5296–5299.
- [15] G. Z. Karabulut and A. Yongacoglu, “Sparse channel estimation using orthogonal matching pursuit algorithm,” in *IEEE 60th IEEE Veh. Technol. Mag. (VTC Fall)*, vol. 6, 2004, pp. 3880–3884 Vol. 6.
- [16] M. A. Khojastepour, K. Gomadam, and X. Wang, “Pilot-assisted channel estimation for MIMO OFDM systems using theory of sparse signal recovery,” in *2009 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 2693–2696.
- [17] T. Yassine and L. Le Magoarou, “mpNet: variable depth unfolded neural network for massive MIMO channel estimation,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5703–5714, 2022.
- [18] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [19] D. K. Kim, S. H. Do, H. B. Cho, H. J. Chol, and K. B. Kim, “A new joint algorithm of symbol timing recovery and sampling clock adjustment for OFDM systems,” *IEEE Trans. Consum. Electron.*, vol. 44, no. 3, pp. 1142–1149, 1998.
- [20] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, “Optimum receiver design for wireless broad-band systems using OFDM. I,” *IEEE Trans. on Commun.*, vol. 47, no. 11, pp. 1668–1677, 1999.
- [21] A. Alkhateeb, “DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications,” in *Proc. of Inf. Theory and Applications Workshop (ITA)*, San Diego, CA, Feb 2019, pp. 1–8.
- [22] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [23] V. Savaux and Y. Louët, “LMMSE channel estimation in OFDM context: a review,” *IET Signal Process.*, vol. 11, no. 2, pp. 123–134, 2017.