Channel Estimation for Full-Duplex RIS-assisted HAPS Backhauling with Graph Attention Networks

Kürşat Tekbıyık*, Güneş Karabulut Kurt*, Chongwen Huang†, Ali Rıza Ekti‡, Halim Yanikomeroglu§

*Department of Electronics and Communication Engineering, İstanbul Technical University, İstanbul, Turkey
†College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China
‡Department of Electrical–Electronics Engineering, Balıkesir University, Balıkesir, Turkey
§Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada
Emails: {tekbiyik, gkurt}@itu.edu.tr, chongwenhuang@zju.edu.cn, arekti@balikesir.edu.tr, halim@sce.carleton.ca

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500697

***Abstract*—In this paper, graph attention network (GAT) is firstly utilized for the channel estimation. In accordance with the 6G expectations, we consider a high-altitude platform station (HAPS) mounted reconfigurable intelligent surface-assisted two-way communications and obtain a low overhead and a high normalized mean square error performance. The performance of the proposed method is investigated on the two-way backhauling link over the RIS-integrated HAPS. The simulation results denote that the GAT estimator overperforms the least square in full-duplex channel estimation. Contrary to the previously introduced methods, GAT at one of the nodes can separately estimate the cascaded channel coefficients. Thus, there is no need to use time division duplex mode during pilot signaling in full-duplex communication. Moreover, it is shown that the GAT estimator is robust to hardware imperfections and changes in small scale fading characteristics even if the training data do not include all these variations.**

***Index Terms*—Reconfigurable intelligent surfaces, channel estimation, graph attention networks, high-altitude platform station systems.**

I. Introduction

As the state-of-the-art reflective surface, reconfigurable intelligent surfaces (RISs) pave the way for a low-cost promising technology improving the spectral efficiency and energy efficiency in wireless networks [1, 2]. Since RISs are able to manipulate the amplitude and/or phase of the impinging signal, they can substantially nullify the randomness of the propagation medium. However, to do these, a RIS-assisted wireless communication system strictly requires high-quality channel state information (CSI). Moreover, channel estimation is getting much more complex for RIS-assisted two-way communications, which has been proposed in [3–5] recently. Also, since the coefficients of channels (i.e., core network (CN) to RIS and RIS to base station (BS)) must be acquired for each channel induced by meta-atoms, the increasing number of elements significantly raises the overhead of the communication and leads to decrease in the efficiency. This study focuses on developing a channel estimation method for RIS-assisted full-duplex communications without the need to activate time division duplex (TDD) mode during channel estimation as well as providing lightweight overhead.

In this study, we will focus on a 6G compliant scenario, specifically look into vertical heterogeneous network (V-HetNet) architecture. The V-HetNet is an emerging network topology including geostationary and low-earth orbit satellites, and high-altitude platform station (HAPS) systems along with terrestrial communication links to serve a large number of small cells with the goals of ubiquitous connectivity and user-centric communication [6]. The HAPS systems, network nodes operating at an altitude of about 20 km in the stratosphere, are the key enablers of the V-HetNets. Due to the properties of the stratosphere, a HAPS can remain in an almost stationary position and can provide ubiquitous connectivity [7]. One of the main uses of HAPS systems is to serve backhauling due to the high cost of fiber optic infrastructure [8, 9]. Many reflective surfaces can be deployed on a HAPS thanks to their large surface area in order to enable high capacity communication [10].

Considering the variations in the channel characteristics observed by the receiving nodes in HAPS communication due to scatterers, and clouds, the cascaded channel estimation in full-duplex RIS-assisted HAPS communication is challenging and requires the novel approaches which are robust to changes in the channel characteristics as well as hardware impairments. Below, we provide a solution to this problem by proposing the graph attention network (GAT) channel estimator.

A. Related Works

A few works (e.g., [3–5]) investigate RIS-assisted two-way communications. In [3], the phase shifts are optimized to maximize the minimum signal-to-noise ratio (SNR). The phase shifts and source precoders are jointly optimized in order to maximize the system sum rate in full-duplex multiple-input multiple-output (MIMO) communications [4, 5]. They reveal the upper bound performances of the proposed systems because they assume that channels are perfectly predicted, which is not the case in practice. One should note that channels need to be estimated precisely in order to perform self-interference cancellation in full-duplex communication. The majority of channel estimation methods are not suitable for RIS-assisted full-duplex communications owing to their high computational cost and/or expense of more overhead.

The prominent channel estimation methods proposed for unidirectional RIS-assisted communications are mentioned below, and the main drawbacks that are possibly experienced during channel estimation even in TDD mode are discussed. Some of the existing methods assume that only a single meta-atom is active at a given time period [11, 12]. This method is called on-off state control, which is time-consuming for a massive number of meta-atoms and allows utilizing the small portion of the elements due to a few active elements at a time even though utilizes deep learning (DL) [12]. Furthermore, this method hampers the synchronization of meta-atoms since the channel coefficients probably change when the last one is estimated. Therefore, the synchronization is spoiled by this switching delay. At this point, an extra algorithm or method is required to recover the synchronization error between meta-atoms. Obviously, this new recovery algorithm would increase the computational complexity and time delay. As RISs are composed of massive passive scattering elements, they are not able to estimate the channel coefficients on their own. Some methods, such as given by [13], propose channel estimation using RISs with active elements at the cost of destroying this attractive feature of RISs. In this approach, RISs are used with active sensors, which are equipped with baseband signal processing units for channel estimation. On the other hand, the channel estimation methods proposed in [14] with massive element RISs do not seem feasible since their computational complexity is proportional to the cube of the number of RIS elements (i.e., $\mathcal{O}(N^3)$). Frankly, these methods are not capable to acquire both channel states in RISs-assisted full-duplex communications. To the best knowledge of the authors, there is not any channel estimation method directly focusing on full-duplex communications, yet. Furthermore, unlike existing deep learning methods, we incorporate the GATs to minimize the computational complexity [15] and increase the learning rate by handling unseen nodes within the proposed RIS-assisted full-duplex communication system.

Even though fiber optic communications are conventionally utilized in backhaul connectivity, building fiber optic infrastructure for small cells is mostly costly solution [16]. Recently, HAPS systems have been proposed to deal with the cost of backhaul connectivities [9]. Therefore, we consider a two-way RIS-assisted HAPS backhauling by considering the prominent features of HAPS backhauling [17] and RIS-enhanced two-way communications [3] for an example application of the proposed GAT channel estimator.

B. Contributions

The main contributions of this study are two-fold and can be summarized as follows:

- To the best of the authors' knowledge, this study firstly investigates a channel estimation method performing at only one of the nodes which transmit their data over full-duplex wireless communication links in order to acquire the coefficients for both channels as illustrated in Fig. 1. In virtue of the proposed channel estimation method, there is no need to switch half-duplex instead of full-duplex when estimating channels. Another crucial point is that this method does not need an on-off state control, which is a time-consuming approach in the channel estimation method proposed in the previous studies owing to the capability of acquiring all channel coefficients regarding the elements of RIS.
- This study considers GAT in channel estimation (even in wireless communications) for the first time. Owing to the attention mechanism in GAT, the proposed system is robust against changes in channel parameters even if the network is trained under different and better channel conditions.

Note that the proposed method can be easily revised to be used for multi-user and/or MIMO RIS-assisted wireless communication scenarios. For instance, it is thought that it may be sufficient to expand the label vector to include channels related to each antenna for channel estimation in RIS-assisted MIMO. It should be highlighted that the method can be also utilized in half-duplex systems by limiting the label vector with only single channel coefficients. Moreover, it should be emphasized that the proposed channel estimation method can be also applied to the channel estimation problems other than RIS-assisted HAPS communications.

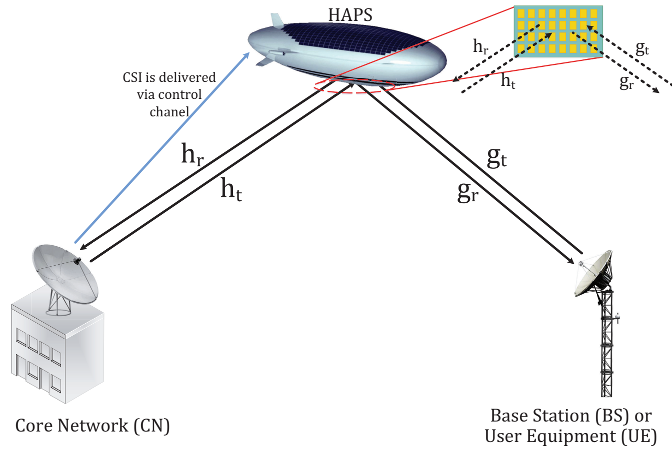

Fig. 1. Full-duplex backhaul communication from BS to CN supported by RIS-assisted HAPS.

C. Outline

The remainder of this work first addresses the mathematical background and channel estimation for RIS-assisted full-duplex communication in Section II. Section III provides the basic mathematical background for GATs. Next, in Section IV, the GAT channel estimator and the channel estimation methodology are introduced. Section V presents the numerical and simulation results with discussions regarding the proposed GAT estimator and least square (LS). Finally, Section VI concludes this study.

II. RIS-assisted Full-Duplex Communications

A considered two-way backhaul network including two end nodes, namely CN and a BS[1], supported by a RIS-assisted HAPS is denoted in Fig. 1. Each end node transmits their information at the same time via the RIS which has N passive elements. Due to long range between nodes and possible obstacles, it can be assumed that the direct link between nodes is negligibly weak. Therefore, the direct link is ignored in this study. Furthermore, the channels CN-to-RIS and RIS-to-BS can be assumed reciprocal if the nodes transmit within a coherence interval and the antennas on the nodes are placed closely. Under the reciprocity assumption, it can be said that $\mathbf{h_t} = \mathbf{h_r} = \mathbf{h}$ and $\mathbf{g_t} = \mathbf{g_r} = \mathbf{g}$, where $\mathbf{h}$ and $\mathbf{g}$ are channel coefficient vectors such that $\mathbf{h} = [h_1, h_2, \ldots, h_N]$ $\mathbf{g} = [g_1, g_2, \ldots, g_N]$. Each fading coefficient for the wireless channel between CN and the n-th element of RIS is denoted $h_n = \alpha_n \mathrm{e}^{-j\varphi_n}$. Similarly, the channel coefficient for the second channel is $g_n = \beta_n \mathrm{e}^{-j\psi_n}$. For both ground-to-HAPS and HAPS-to-ground channels, the magnitudes of channel coefficients can be assumed to follow the Rician distribution [18]. As each node transmits concurrently, the received signal at the node CN can be given as

$$\begin{aligned} y_1(t) =& \sqrt{P_2}\left(\sum_{n=1}^{N} g_n \kappa \mathrm{e}^{j(\theta_n - \psi_n - \varphi_n)} h_n\right) s_2(t) + e_1(t) \\ &+ \sqrt{P_1}\left(\sum_{n=1}^{L} h_n \kappa \mathrm{e}^{j(\theta_n - 2\varphi_n)} h_n\right) s_1(t) + w_1(t), \end{aligned} \quad (1)$$

[1]This node can also be considered as a user equipment.

where the first term is desired signal; κ and θ_n denote amplitude gain and the adjustable phase shift at n-th element of RIS, respectively. $\sqrt{P_1}$ and $\sqrt{P_2}$ stand for the received powers of the CN and BS, respectively. The nodes CN and BS simultaneously transmit symbols $s_1(t)$ and $s_2(t)$, which are the pilot symbols in this study. $e_1(t)$ and $w_1(t)$ are the residual loop interference and additive white Gaussian noise (AWGN) at CN; can be assumed to be distributed with $\mathcal{CN}\left(0, \sigma_{e_1}^2\right)$ [19] and $\mathcal{CN}\left(0, \sigma_{w_1}^2\right)$, respectively. For the simplicity, the phase shifts created by the RIS can be denoted with a diagonal matrix, $\mathbf{\Theta} \in \mathbb{C}^{N \times N}$, as

$$\mathbf{\Theta} = \operatorname{diag}\left\{\kappa \mathrm{e}^{j\theta_1}, \ldots, \kappa \mathrm{e}^{j\theta_N}\right\}. \tag{2}$$

Then, (1) can be rewritten as

$$y_1(t) = \sqrt{P_2}\mathbf{h}^{\mathrm{T}}\mathbf{\Theta}\mathbf{g}s_2(t) + \sqrt{P_1}\mathbf{h}^{\mathrm{T}}\mathbf{\Theta}\mathbf{h}s_1(t) + e_1(t) + w_1(t), \tag{3}$$

where $\mathbf{h}^{\mathrm{T}}\mathbf{\Theta}\mathbf{h}s_1(t)$ is the self-interference. Similar to $y_1(t)$, the received signal at BS is given as

$$y_2(t) = \sqrt{P_1}\mathbf{g}^{\mathrm{T}}\mathbf{\Theta}\mathbf{h}s_1(t) + \sqrt{P_2}\mathbf{g}^{\mathrm{T}}\mathbf{\Theta}\mathbf{g}s_2(t) + e_2(t) + w_2(t). \tag{4}$$

To express the received signal in terms of channel estimation terminology better, the received pilot symbols at CN for the transmitted pilot symbols, $\mathbf{s}_1 \in \mathbb{C}^{1\times M}$, can be restated as follows

$$\mathbf{y_1} = \sqrt{P_2}\mathbf{g}\mathbf{\Theta}\mathbf{h}\mathbf{s_2} + \sqrt{P_1}\mathbf{h}^{\mathrm{T}}\mathbf{\Theta}\mathbf{h}\mathbf{s_1} + \mathbf{e_1} + \mathbf{w_1}, \tag{5}$$

where, $\mathbf{y_1}$, $\mathbf{e_1}$, $\mathbf{w_1} \in \mathbb{C}^{1\times M}$ and $\mathbf{h}^{\mathrm{T}}, \mathbf{g}^{\mathrm{T}} \in \mathbb{C}^{1\times N}$. Regarding the received pilot symbols, the instantaneous SNR related to the estimation of $\mathbf{h}$ at CN can be given as

$$\gamma_{\mathbf{1},\mathbf{h}} = \frac{|\sqrt{P_2}\mathbf{g}\mathbf{\Theta}\mathbf{h}\mathbf{s_2}|^{\circ 2}}{|\mathbf{e_1}|^{\circ 2} + |\mathbf{w_1}|^{\circ 2}}, \tag{6}$$

where $(\cdot)^{\circ 2}$ denotes the Hadamard square. The LS estimation for $\mathbf{h}$ is defined as

$$\hat{\mathbf{h}} = \left(\mathbf{S}^{\mathrm{T}}\mathbf{S}\right)^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{y_1}^{\mathrm{T}}, \tag{7}$$

where

$$\mathbf{S} = \begin{bmatrix} s_1(1) & s_1(1) & \ldots & s_1(1) \\ \vdots & \ddots & & \vdots \\ s_1(M) & s_1(M) & \ldots & s_1(M) \end{bmatrix}. \tag{8}$$

normalized mean square error (NMSE) is given as

$$\mathrm{NMSE} = \frac{\|\mathbf{h} - \hat{\mathbf{h}}\|_2^2}{\|\mathbf{h}\|_2^2}, \tag{9}$$

where $\|\cdot\|_2$ represents the Euclidean norm.

III. The Graph Attention Networks

While conventional DL methods such as convolutional neural networks are successful on data that exhibit a grid-like structure, they cannot show high performance on data in the irregular domain. Graph neural networks can produce state-of-the-art solutions for problems involving data that do not have a grid-like structure. In addition, the attention mechanism provides a very suitable structure for inductive learning, so that the trained network can be generalized over unobserved graphs. Therefore, GATs can be considered as an emerging healer in channel estimation where the observed data frequently changes because of the random nature of the propagation medium. The mathematical background of GATs is detailed below.

A set of P nodes which is an input to the graph attention layer (GAL) is defined as $\vartheta = \left\{\vec{\vartheta}_1, \vec{\vartheta}_2, \ldots, \vec{\vartheta}_P\right\}$, $\vec{\vartheta}_i \in \mathbb{R}^F$, where F denotes the number of features for each node. GAL outputs a set of node features as $\vartheta' = \left\{\vec{\vartheta}'_1, \vec{\vartheta}'_2, \ldots, \vec{\vartheta}'_P\right\}$, $\vec{\vartheta}'_i \in \mathbb{R}^{F'}$, where F' denotes the number of features for each output nodes and might have different cardinality than the former. A shared linear transformation parameterized by the weight matrix, $\mathbf{W} \in \mathbb{R}^{F\times F'}$, is applied to each node to transform input properties to higher-level properties. Thereafter, the self-attention on the nodes is investigated by an attention mechanism $a : \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$, which computes the attention coefficients as

$$c_{ij} = a\left(\mathbf{W}\vec{\vartheta}_i, \mathbf{W}\vec{\vartheta}_j\right), \tag{10}$$

where c_{ij} is found for only the j-th node which has neighborhood of i-th node in the graph. The attention coefficients show the importance of features of j-th node on the i-th node. To make the coefficients comparable in a different neighborhood, they are then normalized using the softmax function as follows [20]

$$\alpha_{ij} = \mathrm{softmax}_j\left(c_{ij}\right) = \frac{\exp\left(c_{ij}\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(c_{ik}\right)}, \tag{11}$$

where $\mathcal{N}_i$ denotes the neighborhood of the i-th node. The normalized coefficients, α_{ij}, are computed by the attention mechanism a as [15]

$$\alpha_{ij} = \frac{\exp\left(\mathrm{ReLU}\left(\mathbf{a}^{\top}\left[(\mathbf{X}\mathbf{W})_i \| (\mathbf{X}\mathbf{W})_j\right]\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\mathrm{ReLU}\left(\mathbf{a}^{\top}\left[(\mathbf{X}\mathbf{W})_i \| (\mathbf{X}\mathbf{W})_k\right]\right)\right)}, \tag{12}$$

where $\mathbf{a} \in \mathbb{R}^{2F'}$ and $\mathbf{X} \in \mathbb{R}^{P\times F}$ denote attention kernel and node features, respectively. Finally, the convolution over the graph network is performed as

$$\mathbf{Z} = \alpha\mathbf{X}\mathbf{W} + \mathbf{b}, \tag{13}$$

where $\mathbf{b}$ is the trainable bias vector. This layer accepts inputs which are the node attributes matrix $\mathbf{X} \in \mathbb{R}^{P\times F}$, binary adjacency matrix $\mathbf{A} \in \{0,1\}^{P\times P}$, and edge attributes matrix $\mathbf{E} \in \mathbb{R}^{P\times P\times S}$.

As stated in [21], a pooling layer is required to generalize graph convolution networks. Furthermore, to reduce the number of representations, a pooling layer is employed. Thus, the pooling layer enables the network to avoid overfitting. Since this study utilizes global attention pooling, we do not refer to other graph pooling layers here. The global attention pooling layer computes

$$\mathbf{X}' = \sum_{i=1}^{P}\left(\sigma\left(\mathbf{X}\mathbf{W}_1 + \mathbf{b}_1\right) \odot \left(\mathbf{X}\mathbf{W}_2 + \mathbf{b}_2\right)\right)_i, \tag{14}$$

where σ and $\odot$ are the sigmoid function and the broadcasted elementwise product, respectively.

IV. Full-Duplex Channel Estimation

In this section, we introduce a channel estimation procedure using GAT for two-way backhaul over RIS-assisted HAPS as illustrated in Fig. 1. As aforementioned, it is required that channel estimation should be carried out without switching

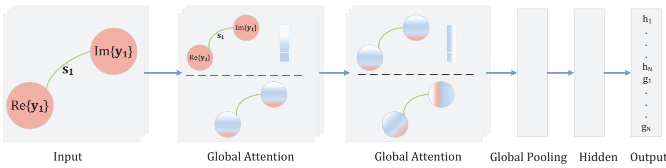

Fig. 2. The illustration for the proposed graph attention network including two consecutive graph attention networks and global attention pooling. The real and imaginary parts of received signal, $\mathbf{y_1}$, are assigned to attributes of two nodes. The edge attributes are set as the pilot symbols, $\mathbf{s_1}$.

TABLE I
SUMMARY OF THE DATASET PARAMETERS EMPLOYED DURING TRAINING AND TEST.

Parameters	**Training**	**Test**
PN Polynomial	x^4+x^2+1	x^4+x^2+1
Modulation	BPSK	BPSK
# of Samples per SNR	1000	500
SNR (dB)	-30:2:0	-30:2:10
K	10	0, 4, 8, 10, 12
M	16, 32, 64, 128	16, 32, 64, 128
N	128, 256, 512, 1024	128, 256, 512, 1024
ϵ	0	0, 1e-1, 1e-2, 1e-3

RIS elements on-off and be robust against hardware imperfections and serious fluctuations in channel characteristics. GATs can be generalized to completely unobserved graphs during the training [15]; therefore, it is an appropriate solution to the channel estimation problem. Considering the random nature of the wireless propagation medium and that RIS cannot manipulate random behavior during the channel estimation, a generalizable model or procedure is required for accurate channel estimation in order to avoid performance degradation in estimation when channel characteristics significantly change. In consequence, GATs are utilized in this study to obtain all channel parameters regarding RIS elements in a one-shot that is to say without consecutive on-off switch for each element on RIS.

A. Training Dataset Generation

Instead of using TDD mode during pilot signaling, we consider remaining two-way communications when channel estimation is performed. Hence, we can define the problem as estimation of $\mathbf{h}$ and $\mathbf{g}$ from $\mathbf{y_1}$ at node CN. Similary, the same problem can be defined for the node BS; however, we consider only $\mathbf{y_1}$ for the estimation of $\mathbf{h}$ and $\mathbf{g}$.

Firstly, the M-length pilot symbols are created as pseudo-noise (PN) sequence utilizing the polynomial given by x^4+x^2+1. Utilizing PN sequence also enables the nodes and HAPS to make synchronous, which is demanded by a proper wireless communications. In this study, $\mathbf{s_1}$ and $\mathbf{s_2}$ are selected same. Then, this pilot symbols are received at CN after reflected from RIS with N meta-atoms. P_1 and P_2 are chosen as unit power. During pilot signaling, all elements of RIS can switch on without any phase shift, scilicet unitary phase shift matrix, $\mathbf{\Theta}$. In other word, $\kappa=1$. As the scatterers which are close to ground stations are required to be considered [22], we assume that both channels depicted in Fig. 1 follow Rician fading with $K=10$. By using these parameters, the training dataset is generated. The input regarding received signal, $\mathbf{X}$, is created as follows:

$$\mathbf{X}=\left[\mathrm{Re}\{\mathbf{y_1}\};\,\mathrm{Im}\{\mathbf{y_1}\}\right]. \tag{15}$$

TABLE II
THE PROPOSED GAT LAYOUT AND PARAMETERS.

Layers		**Dimensions**
Inputs	$\mathbf{X}$	$2\times M$
	$\mathbf{A}$	2×2
	$\mathbf{E}$	$2\times 2\times M$
Labels	$\mathbf{y}$	$2N\times 1$
Graph Attention 1		2×128
Graph Attention 2		2×32
Global Attention Pool		128
Dense		$2N$
Parameters		**Values**
Activation		ReLU
Optimizer		Adam
Loss		MSE
Learning Rate		1e-3
L_2 Regularization		5e-4

The adjacency matrix for the graph network denoting the real and imaginary parts of channel coefficient is given as

$$\mathbf{A}=\begin{bmatrix}0&1\\1&0\end{bmatrix}, \tag{16}$$

which implies that two nodes are connected with a single edge as illustrated in Fig. 2. The known pilot symbols are assigned to weight of the edge for the k-th nonzero element of adjacency matrix as

$$\mathbf{E_k}=\mathbf{s_1},\ \mathbf{E}\in\mathbb{C}^{2\times 2\times M}. \tag{17}$$

Also, the label vector related to these inputs is defined as

$$\mathbf{y}=\left[h_1,h_2,\cdots,h_N,g_1,g_2,\cdots,g_N\right]^{\mathrm{T}}. \tag{18}$$

The training dataset includes 1000 input samples for each SNR value in between -30 dB and 0 dB with 2 dB step. The chosen SNR interval allows proper RIS-assisted communications in terms of bit error rate as seen in [1]. The training dataset totally consists of 16000 input samples for each M, N, and SNR values. The dataset is divided into two parts, which are used for training and validation with the rate of $4:1$. The parameters which are used during the training dataset generation are summarized in Table I.

B. GAT Model and Training

In Section III, the background for GATs is visited. In this section, we introduce the parameters and details related to the proposed GAT model. Two consecutive GALs are used to learn the channel parameters over the graph structure detailed before. The first one has 128 output channels while the latter has 32. In two layers, ReLU activation function is employed. Considering the dataset regarding channel estimation problem, the dimensions of inputs now become as $P=2$, $F=M$, and $S=M$. Besides GALs, a global attention pooling layer is employed in order to reduce the number of representations and thus avoid the network overfitting. To keep away the network from overfitting problem, the dropout in both GALs with the rate of 0.5 and L_2 regularization are utilized. The network is terminated by a hidden layer with $2N$ neurons. The optimizer and loss function are chosen as ADAM with the learning rate of 10^{-3} and mean square error (MSE). We utilize Spektral [23] in the implementation of graph neural network, namely GAT. The number of epochs is set to 20, but early stopping is activated if there is no improvement in the loss for 5 epochs. The parameters for the GAT and inputs are summarized in Table II.

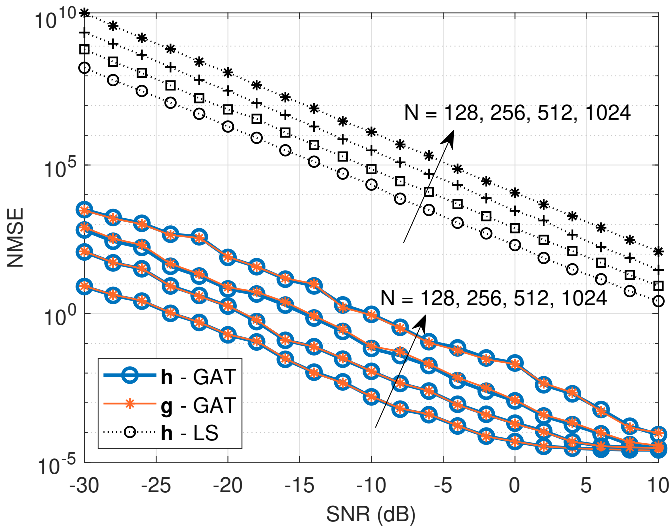

Fig. 3. NMSE performance of the proposed GAT-aided full-duplex channel estimation versus the SNRs and the number of RIS elements, N, for $M = 128$, $K = 10$ and $\epsilon = 0$. The proposed method can estimate the channel coefficients, $\mathbf{h}$ and $\mathbf{g}$, with almost the same performance at CN.

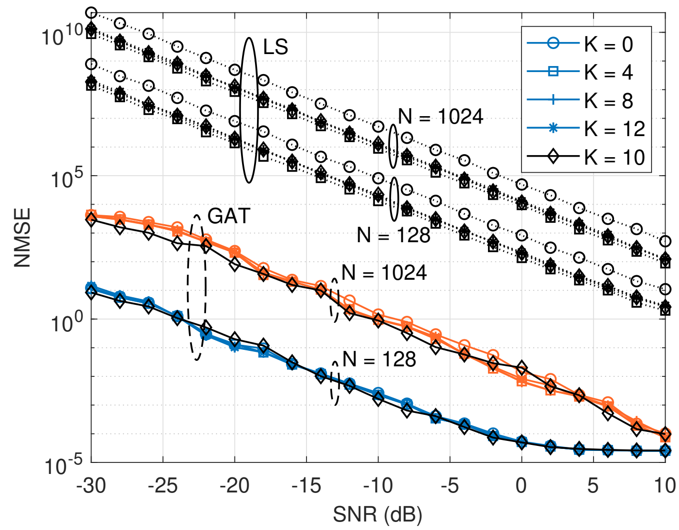

Fig. 5. NMSE performance of the proposed GAT-aided full-duplex channel estimation versus the SNRs, the number of RIS elements, N, and the Rician K-factor for $M = 128$ and $\epsilon = 0$. The proposed method is able to remain almost the same performance under the effect of changing fading.

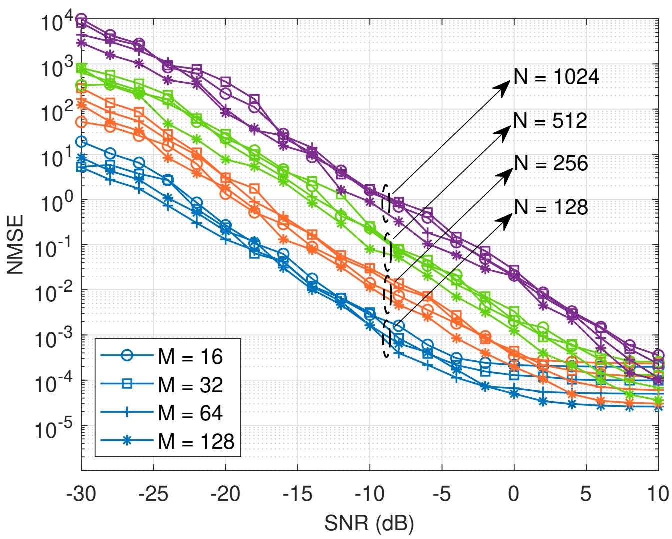

Fig. 4. NMSE performance of the proposed GAT-aided full-duplex channel estimation versus the SNRs, the number of RIS elements, N, and the number of pilot symbols, M, for $K = 10$ and $\epsilon = 0$.

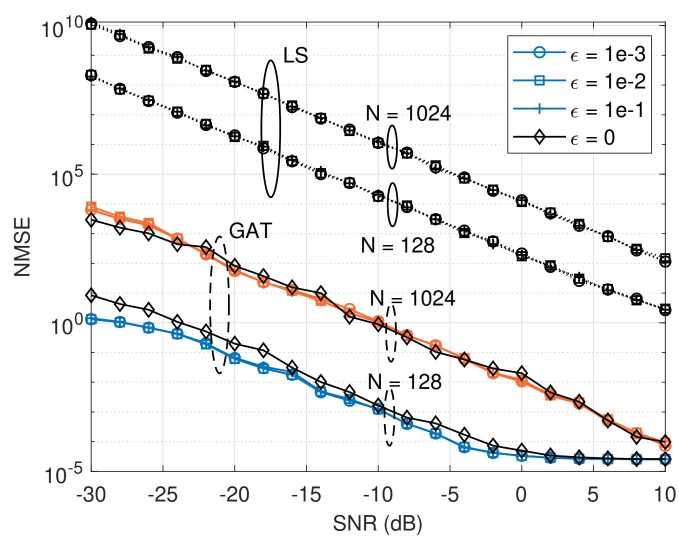

Fig. 6. NMSE performance of the proposed GAT-aided full-duplex channel estimation versus the SNRs, the number of RIS elements, N, and switching error, ϵ, for $M = 128$ and $K = 10$. The proposed method is able to remain almost the same performance under the effect of changing hardware imperfection.

V. Numerical Results and Discussions

In this section, the channel estimation performance of the GAT model, the details of which are given in the previous section, is investigated under different scenarios. Firstly, we consider the full-duplex channel estimation performances for both $\mathbf{h}$ and $\mathbf{g}$ in Fig. 3. The proposed GAT estimator outperforms LS estimation where TDD is not activated. Both methods show the same NMSE performance with about 3 dB of SNR loss when the number of RIS elements is doubled and the number of pilot symbols is kept constant (i.e. $M = 128$). Although the estimation procedure is carried out at only CN[2], Fig. 3 denotes almost the same performance for both channels, which shows that the main purpose of this study is successfully achieved. Since NMSE performance for two channel estimations are almost identical, only NMSE results for $\mathbf{h}$ are given hereafter to increase the readability of the figures. It should be noted that although SNR values between 0 and 10 dB are not included in the training set, the proposed method can make channel estimation with high performance at these SNR values.

Also, we investigate the channel estimation performance jointly for the number of pilot symbols and the number of RIS elements as seen in Fig. 4. It is observed that the increase in the number of pilot symbols does not make a significant difference in NMSE performance except for the maximum achievable performance limit at the high SNR region. The main reason behind these results is that the attention mechanism used in this DL network can focus on the most relevant part of inputs when making a decision, as stated in [15]. In the light of the results, it can be said that the GAT estimator

[2]The same GAT can be used at BS without any change in the parameters and architecture.

can reduce the overhead for channel estimation. On the other hand, increasing the number of channel coefficients required to be estimated slightly deteriorates the NMSE performance. As said before, when the number of meta-atoms is doubled, the system needs 3 dB more SNR to remain the same NMSE performance.

As the value of K is dependent on the propagation environment, for instance, it can be observed that K is lower in urbanized regions where many scatterers are usually found (vice versa in rural areas). Furthermore, it should be noted that the HAPS movement or displacement can give rise to change in the platform elevation angle; thereby, the value of K changes with the variation in the elevation angle [18]. In consequence, it is important to test the performance of the channel estimation method according to the small scale fading factor. In Fig. 5, the trained GAT model is tested under varying K values even though the network is trained with only $K = 10$. While the LS estimator's performance degrades in the case that there is no line-of-sight link, GAT estimator is able to show satisfactory NMSE performance regardless of K values. This is another finding that GATs can be successful in different inputs.

Although the reflection coefficients of RIS elements are assumed as 1, which means perfect reflector, it should be noted that they cannot completely reflect the power of an incident wave. Thus, we consider the imperfection in the amplitude gain of RISs by introducing an error, ε, such that $\kappa = 1 - \epsilon$ as in [12]. The network which is trained under the ideal amplitude condition is tested for $\epsilon = 10^{-3}$, 10^{-2}, 10^{-1}. As expected, Fig. 6 presents that the LS estimator is not affected from the error in amplitude gain. The GAT provides much better channel estimation performance than LS. Moreover, its performance is not degraded when changing the amplitude gain of RIS. To speak generally, Fig. 6 denotes the robustness of the proposed channel estimation method against the hardware imperfections.

The attention mechanism in graph convolutional networks, pooling, and regularization used in the proposed GAT channel estimator help ensure robustness and high performance as shown in the results above. Noted that the proposed system can be utilized for the channel estimation in half-duplex communications as well. For example, $\mathbf{g}$ and $\mathbf{h}$ are separately estimated at CN and BS, respectively when TDD mode is activated.

VI. Concluding Remarks

In this study, we propose a channel estimation method utilizing GAT which can be generalized over unobserved graph structures in the virtue of the attention mechanism. The proposed method does not require on-off state control and can estimate the coefficients of two main channel blocks of the RIS-assisted communication system at the same time with a high performance. GAT estimator's NMSE performance is studied in the scenario full-duplex backhauling over RIS-integrated HAPS.

The simulation results reveal that the proposed method has remarkably high performance. Furthermore, thanks to the attention mechanism and graph structure, the estimator is able to maintain its performance under different channel conditions and hardware impairments which are not seen by the network during the training. Besides full-duplex channel estimation, the GAT estimator can also be employed in half-duplex communications and multi-user systems as well as MIMO systems by modifying the label vector in the training phase.

References

[1] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[2] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.

[3] S. Atapattu, R. Fan, P. Dharmawansa, G. Wang, J. Evans, and T. A. Tsiftsis, "Reconfigurable intelligent surface assisted two-way communications: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6552–6567, 2020.

[4] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, 2020.

[5] H. Shen, T. Ding, W. Xu, and C. Zhao, "Beamforming design with fast convergence for IRS-aided full-duplex communication," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2849–2853, 2020.

[6] K. Tekbiyik, A. R. Ekti, G. K. Kurt, A. Gorcin, and H. Yanikomeroglu, "A holistic investigation on terahertz propagation and channel modeling toward vertical HetNets," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 14–20, 2020.

[7] M. Tseytlin and C. Weasler, "High altitude platform stations (HAPS) — bringing connectivity to all," https://bit.ly/34ORuLx, Aug 2019, (Accessed on Oct 8, 2020).

[8] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M.-S. Alouini, "FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 218–224, 2018.

[9] G. Kurt, M. G. Khoshkholgh, S. Alfattani, A. Ibrahim, T. S. Darwish, M. S. Alam, H. Yanikomeroglu, and A. Yongacoglu, "A vision and framework for the high altitude platform station (HAPS) networks of the future," *arXiv preprint arXiv:2007.15088*, 2020.

[10] S. Alfattani, W. Jaafar, Y. Hmamouche, H. Yanikomeroglu, and A. Yongaçoglu, "Link budget analysis for reconfigurable smart surfaces in aerial platforms," *arXiv preprint arXiv:2008.12334*, 2020, (appear in IEEE Commun. Mag. on January 2021).

[11] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *Proc. IEEE ICASSP*, 2019, pp. 4659–4663.

[12] A. M. Elbir, A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Deep channel learning for large intelligent surfaces aided mm-Wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1447–1451, 2020.

[13] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *arXiv preprint arXiv:1904.10136*, 2019.

[14] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, M. Debbah *et al.*, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *arXiv preprint arXiv:2008.01459*, 2020.

[15] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2018.

[16] H. Dahrouj, A. Douik, F. Rayal, T. Y. Al-Naffouri, and M.-S. Alouini, "Cost-effective hybrid RF/FSO backhaul solution for next generation wireless systems," *IEEE Wirel. Commun.*, vol. 22, no. 5, pp. 98–104, 2015.

[17] S. Karapantazis and F. Pavlidou, "Broadband communications via high-altitude platforms: A survey," *IEEE Commun. Surveys Tuts.*, vol. 7, no. 1, pp. 2–31, 2005.

[18] E. T. Michailidis and A. G. Kanatas, "Three-dimensional HAP-MIMO channels: Modeling and analysis of space-time correlation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2232–2242, 2010.

[19] L. J. Rodriguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex AF relaying in the presence of residual self-interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, 2014.

[20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[21] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *International Conference on Machine Learning*. PMLR, 2019, pp. 3734–3743.

[22] F. Dovis, R. Fantini, M. Mondin, and P. Savi, "Small-scale fading for high-altitude platform (HAP) propagation channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 641–647, 2002.

[23] D. Grattarola and C. Alippi, "Graph neural networks in Tensorflow and Keras with Spektral," *IEEE Comput. Intell. Mag.*, vol. 16, no. 1, pp. 99–106.