Lateral confinement of high-impedance surface-waves through reinforcement learning

M.E. Morocho-Cayamcela and W. Lim✉

The authors present a model-free policy-based reinforcement learning model that introduces perturbations on the pattern of a metasurface. The objective is to learn a policy that changes the size of the patches, and therefore the impedance in the sides of an artificially structured material. The proposed iterative model assigns the highest reward when the patch sizes allow the transmission along a constrained path and penalties when the patch sizes make the surface wave radiate to the sides of the metamaterial. After convergence, the proposed model learns an optimal patch pattern that achieves lateral confinement along the metasurface. Simulation results show that the proposed learned-pattern can effectively guide the electromagnetic wave through a metasurface, maintaining its instantaneous eigenstate when the homogeneity is perturbed. Moreover, the pattern learned to prevent reflections by changing the patch sizes adiabatically. The reflection coefficient $S_{1,2}$ shows that most of the power gets transferred from the source to the destination with the proposed design.

Introduction: Metasurfaces can be engineered to contain surface waves (SWs) in a homogeneous path, where patches of sub-wavelength size are engraved on a high-frequency grounded laminate [1]. SW waveguides (SWGs) are characterised for propagating SWs along a confined path without radiation [2]. To avoid losses due to propagation in undesirable directions, dissimilar patch sizes representing different values of refractive-index are required. These metasurfaces are formed by highly subwavelength inclusions embedded in a host medium, leading to homogenised permittivity and permeability values not available in nature. Specifically, to achieve surface guiding, a negative refractive index is of interest. These inexpensive surfaces are engineered to have high performance with low-loss and low-dispersion. The effect of perturbing the homogeneity of the patches has been exploited to design planar lenses and leaky-wave antennas [3]. Among the emerging technologies for future 5G/B5G networks, the potential of SWG has been pointed by several authors to create extremely large aperture arrays, capable of trap, guide, and leak signals [4, 5]

In this paper, we focus on the task of guiding SW in a metamaterial along a confined path. For this effect, we propose to (i) design a metasurface with uniform patches to trap the wave along the surface. (ii) Alter the uniformity of the patches according to a policy that assigns a reward when the signal is transmitted from the desired source to the desired destination, and a penalty when the signal propagates to the sides. Our model-free policy-based reinforcement learning (RL) method, enables to learn an optimal design for an SWG by maximising the reward iteratively.

High-impedance metasurface design: To produce a high-impedance surface on a high-frequency material, a periodical grid of perfectly electric conductor made by an array of patches with equal width and length, have to be etched on one of the faces of the printed circuit board. We first design the metasurface for a required attenuation constant α_z, which fixes the value of l_p, D, w and the surface impedance X_s. Fig. 1*a* illustrates a single patch with periodicity D, patch length l_p, gap of $w = D - l_p$, height of substrate h', height of metallic patch h_p, and height of the ground conductor h_g.

We adopt the design constraints of $D = \lambda/10$ and $w = D/10$, according to the results obtained in [6] for 10 GHZ high-impedance metasurfaces. The high-frequency laminate used for the design is Rogers RT/Duroid 5880, which has a relative permittivity $\epsilon_r = 2.20$, dissipation factor $\tan\delta = 9 \times 10^{-4}$, laminate height $h' = 1.54$ and metal height $h_p = 0.07$. Using $n_{eq} = \sqrt{1 + (\alpha_z/k)^2} > 1$, with $k = 2\pi/\lambda$, we estimate the value of the equivalent refractive index n_{eq}, where k is the wavenumber. From here, X_s can be obtained from $n_{eq} = \sqrt{1 + (X_s/Z_0)^2}$, with $Z_0 = \sqrt{\mu_0/\epsilon_0}$ as the characteristic impedance in free space (FS), estimated as the square root of the rate of the permeability of FS (Henrys/m) to the permittivity of FS (Farads/m). For a dielectric dependant value δ in the metasurface, one can solve (1) for δ and obtain a value in function of the α_z

$$X_s(w, k_t) = \frac{\zeta k_{z1} \tan(h' k_{z1})}{k\epsilon_r - k(\epsilon_r + 1)\delta k_{z1} \tan(h' k_{z1})}, \quad (1)$$

where δ can be expressed as $\delta = (D/\pi)\ln[1/\sin(\pi w/2D)]$. The value of δ, is used to calculate the gap w, in function of n_{eq} with $w = (2D/\pi)\arcsin(1/\exp(\delta\pi/D))$ [7]. From here, the values of δ, k_z, and X_s can be derived. These values trap the SW in the metasurface, but they also determine a fixed size for the patches as well as the gap between them.

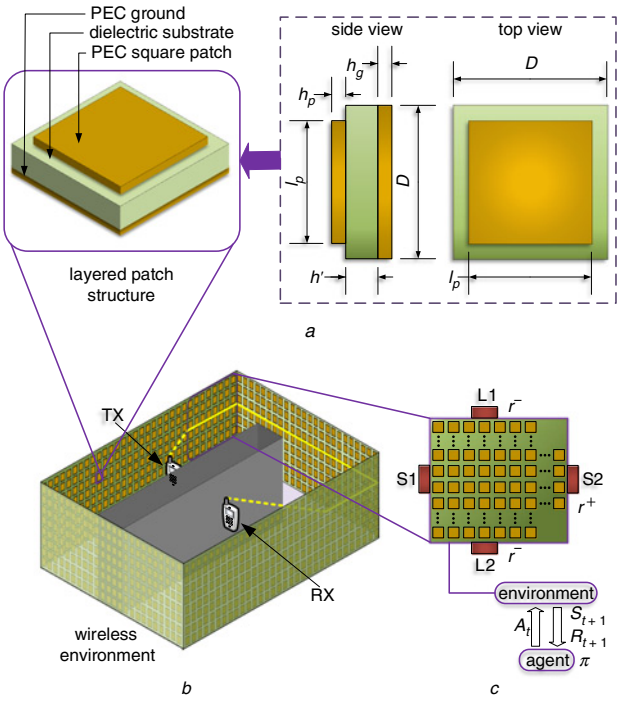

Fig. 1 *Metamaterial structure, applicability, and RL model*

a Dimensions of the metallic unit-cell patch on a high-frequency grounded laminate

b Wireless environment showing an application of metamaterial SWGs

c Agent takes an action A_t on the environment and receives a reward R_{t+1} at state S_{t+1}

RL-based lateral confinement: For a grid of patches to be considered an SWG, the metasurface must guide the SW along a constrained path formed by a variable-impedance patch array. Fig. 1*b* shows a potential application of SWG in a wireless environment (assuming beamforming leaky-wave antennas and coupling in the corners). We propose to confine the SW in-homogeneous tapered metasurface reactance by altering the value of l_p in the patches. We introduce a differential value Δ_{l_p} and update the value $l_p \leftarrow l_p \pm \Delta_{l_p}$, with $l_p \leq D - w$. For training our model, a standard WG16 waveguide port (S1) is used as excitation with a frequency range of 10 GHz (Fig. 1*c*). To ensure the matching between the SWG and the radiation port structures, we set the alteration over distances of the order of a guided wavelength in width, so one mode blends into the next without suffering reflections. An agent can take one action $a \in A_t = \{l_p \leftarrow l_p + \Delta_{l_p}, l_p \leftarrow l_p - \Delta_{l_p}\}$ to change the size of each patch in the metasurface. The EM-wave is trapped in the metamaterial environment and propagates in all directions, receiving a reward R_t at state S_t. To achieve lateral confinement, the total reward is defined as the sum of the rewards in the waveguides L1, L2 and S2, as follows:

$$R = \begin{cases} r^- & \text{for reaching lateral waveguides}(L1, L2), \\ r^+ & \text{for reaching the goalwaveguide } (S2), \end{cases} \quad (2)$$

with $r^- \ll r^+$. The agent learns an optimal value for Δ_{l_p} for every patch in the metasurface matrix $\boldsymbol{M}$ by maximising the cumulative reward. For the following iterations, the agent receives a reward R_{t+1} and the state changes to S_{t+1}. If we let the mathematical problem be represented as a Markov decision process, we can describe the task with a set of actions $\boldsymbol{A}$, a set of states $\boldsymbol{S}$, a matrix with the transition probability $\boldsymbol{P}$ and the set of rewards $\boldsymbol{R}$. For every iteration t, the agent observes the state s_t before taking an action a_t. Next, the state shifts into s_{t+1} and the agent gets a reward R_t. We iterate the later process (i.e. observing the state s_t, taking an action a_t, and receiving a reward R_t), until our agent maximises the cumulative sum of rewards over a period of

time t. This maximisation formulates an optimal patch design policy $\pi(s_t) \in A$ [8]. In our patch design model, the maximisation of the cumulative reward implies the minimisation of the propagation of the SW to the sides of the metasurface. We define a state-value function (3) that evaluates the value of a certain metasurface patters state, under the design policy π as follows:

$$V^{\pi}(s) = E\left[\sum_{t=0}^{\infty} \gamma^{t} R(s_t, \pi(s_t)) | s_0 = s\right], \tag{3}$$

where $E[\cdot]$ represents the expectation operator and $\gamma \in [0, 1)$ is a discount factor. Therefore, the optimal SWG design policy π^*, can be obtained following the Bellman's optimality criterion as follows:

$$V^{*}(s) = \max_{a \in A} [R(s, a) + \gamma \sum_{s'} P_{s,a}(s') V(s')], \tag{4}$$

where the transition probability from s to s' when action a is taken, is represented as $P_{s,a}(s')$. Our proposal estimates this value $P_{s,a}(s')$ that changes the patch sizes as a Q-learning task [9]. For the SWG design policy π, the Q-value that maps the dimension-state of each row of patches to the action of increasing or reducing its size (s,a), is defined as the expected discounted reward of taking the action a in the per-row dimension state s, according to the design policy π (5). In the same way, the optimal Q-value Q^* (6), can be estimated if we ponder the optimal value $V^*(s')$. By setting (6), the value of $V^*(s)$ can be superseded by $\max_{a \in A} Q^*(s, a)$ (7).

$$Q^{\pi}(s, a) = R(s, a) + \gamma \sum_{s'} P_{s,a}(s') V^{\pi}(s'), \tag{5}$$

$$Q^{*}(s, a) = R(s, a) + \gamma \sum_{s'} P_{s,a}(s') V^{*}(s'), \tag{6}$$

As reported by [10], with a learning rate of α the Q-values can be expressed as

$$\begin{aligned} Q_{t+1}(s, a) =& Q_t(s, a) + \alpha(R(s, a)) \\ &+ \alpha\left(\gamma \max_{a' \in A} Q_t(s', a') - Q_t(s, a)\right). \end{aligned} \tag{7}$$

The agent learns the optimal Q-values by iterating over (7), obtaining the state s_t, receiving the reward R_t as expressed in (2), and changing the patch size of every row by selecting the associated action a_t at each time t. The convergence of the iteration (7) produces an optimal value V^* and an optimal SWG design policy π^*. We assume a ϵ-greedy strategy to determine an action for every iteration.

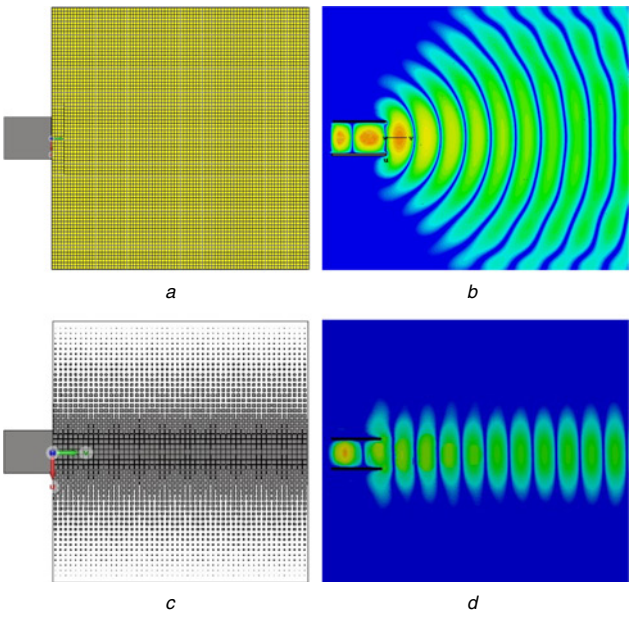

Fig. 2 *Magnetic field of the uniform patch model, and the proposed SWG structure*

a Uniform patch structure, with $D = \lambda/10$ and $w = D/10$
b Magnetic field of the uniform structure showing uncontrolled propagation
c SWG designed with the proposed model-free policy-based model
d Lateral confinement of the magnetic field from the proposed design

Results and performance evaluation: Fig. 2*a* illustrates the uniform structure metasurface and Fig. 2*b* shows that the structure traps the SW on its surface, but it propagates in all undesired directions without control. Furthermore, the reflections of the SW on the edges of the artificeially structured surface generates standing waves and destructive reflections. Fig. 2*d* shows the constrained pattern learned from the proposed policy-based system. Fig. 2*d*, confirms that our RL-based design confines the electromagnetic waves laterally, and guide the SWs along the constrained path for the designed frequency of 10 GHz, Fig. 3 proves a metasurface based on a uniform patch array fails a port-to-port power transfer test, while the proposed RL-based metasurface effectively transfers the power from port 1 to port 2. The learned design maintains the instantaneous eigenstate of the metasurface when the homogeneity is perturbed, that is, the gap between patches change gradually following the adiabatic theorem. The complexity of the value-iteration algorithm in the Q-learning problem is $O(en)$, where e represents the total number of actions and n is the size of the state space [11].

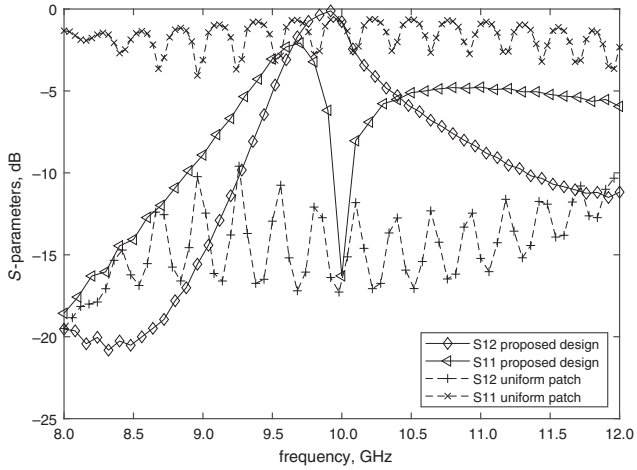

Fig. 3 *The plot of* S-*Parameters showing the reflection coefficients from excitation ports S1 to the goal port S2, for a uniform array metasurface, and the proposed RL-based pattern*

Conclusions: Simulations revealed that our design improved the SW transmission by 15 dB compared with a metamaterial with equal patch size. Our model-free policy-based model learned to gradually alter the pattern of the patch sizes to achieve a refractive index that contains the propagation of the SW in undesirable directions. Further work that can be engineered from this analysis includes the design of reconfigurable metasurfaces, software-defined metasurfaces, metasurface planar lenses, metamaterial cloaks and transformation optic devices. Nevertheless, there are still challenges that need to be addressed in the future like the complexity introduced, the interpretability of results and the time required to train the model.

Acknowledgments: This work was supported by the National Research Foundation of Korea (NRF) under grant 2020R1A4A101777511; and in part by the Technology Development Program (S2829065) funded by the Ministry of SMEs and Startups (MSS, Korea).

Submitted: *09 July 2020* E-first: *25 September 2020*
doi: 10.1049/el.2020.1977
One or more of the Figures in this Letter are available in colour online.

M.E. Morocho-Cayamcela (*Centro de Investigación, Desarrollo e Innovación en Sistemas Computacionales (CIDIS), Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador*)

W. Lim (*Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, Gyeongsangbuk-do, 39177, Republic of Korea*)

✉ E-mail: wansu.lim@kumoh.ac.kr

References

1 Barlow, H.M., and Brown, J.: 'Conditions for the support of surface waves at an interface between two different homogeneous media, in Radio surface waves' (Clarendon Press, UK, 1962), pp. 1–25

2 Quarfoth, R., and Sievenpiper, D.: 'Impedance surface waveguide theory and simulation'. 2011 IEEE Int. Symp. on Antennas and Propagation (APSURSI), Spokane, WA, USA, July 2011, pp. 1159–1162, doi: 10.1109/APS.2011.5996489
3 Morocho-Cayamcela, M.E., Angsanto, S.R., Lim, W., *et al*.: 'An artificially structured step-index metasurface for 10 GHz leaky waveguides and antennas'. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT) Singapore, Singapore, 22018, pp. 568–573, doi: 10.1109/WF-IoT.2018.8355195
4 Björnson, E., Sanguinetti, L., Wymeersch, H., *et al*.: 'Massive MIMO is a reality – what is next?', *Digit. Signal Process.*, 2019, **94**, pp. 3–20, doi: 10.1016/j.dsp.2019.06.007
5 Morocho-Cayamcela, M.E., Lee, H., and Lim, W.: 'Machine learning for 5G/B5G mobile and wireless communications: potential, limitations, and future directions', *IEEE Access*, 2019, **7**, pp. 137184–137206, doi: 10.1109/ACCESS.2019.2942390
6 Holloway, C.L., Mohamed, M.A., Kuester, E.F., *et al*.: 'Reflection and transmission properties of a metafilm: with an application to a controllable surface composed of resonant particles', *IEEE Trans. Electromagn. Compatib.*, 2005, **47**, (4), pp. 853–865, doi: 10.1109/TEMC.2005.853719
7 Luukkonen, O., Simovski, C., Granet, G., *et al*.: 'Simple and accurate analytical model of planar grids and high-impedance surfaces comprising metal strips or patches', *IEEE Trans. Antennas Propag.*, 2008, **56**, (6), pp. 1624–1632, doi: 10.1109/TAP.2008.923327
8 Sutton, R.S., and Barto, A.G.: 'Reinforcement learning: an introduction' (The MIT Press, London, England, 2018, 2nd edn.), ISBN: 9780262039246
9 Morocho-Cayamcela, M. E., Lee, H., and Lim, W.: 'Machine learning to improve multi-hop searching and extended wireless reachability in v2x', *IEEE Commun. Lett.*, 2020, **24**, pp. 1477–1481, doi: 10.1109/LCOMM.2020.2982887
10 Nie, J., and Haykin, S.: 'A dynamic channel assignment policy through Q-learning', *IEEE Trans. Neural Netw.*, 1999, **10**, (6), pp. 1443–1455
11 Koenig, S., and Simmons, R. G.: 'Complexity analysis of real-time reinforcement learning'. Proc. Association for the Advancement of Artificial Intelligence (AAAI), Washington, DC, USA, 1993, pp. 99–107, ISBN: 0262510715