

Received September 9, 2021, accepted September 19, 2021, date of publication September 24, 2021, date of current version October 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3115596

Wireless Fingerprinting Localization in Smart Environments Using Reconfigurable Intelligent Surfaces

CAM LY NGUYEN¹, ORESTIS GEORGIU², (Senior Member, IEEE),
GABRIELE GRADONI³, (Member, IEEE), AND MARCO DI RENZO⁴, (Fellow, IEEE)

¹Wireless System Laboratory, Corporate Research and Development Center, Toshiba Corporation, Kawasaki 212-8582, Japan

²Department of Electrical and Computer Engineering, University of Cyprus, 1678 Nicosia, Cyprus

³Department of Electrical and Electronic Engineering, School of Mathematical Sciences, University of Nottingham, University Park NG72RD, U.K.

⁴Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 Rue Joliot-Curie, 91192 Gif-sur-Yvette, France

Corresponding author: Cam Ly Nguyen (camly268@gmail.com)

This work was supported in part by Vingroup Joint Stock Company (Vingroup JSC), Vingroup, and in part by Vingroup Innovation Foundation under Grant VINIF.2020.DA09. The work of Orestis Georgiou was supported by the European Union's Horizon 2020 Research and Innovation Program through the Marie Skłodowska-Curie Project eNergy nEutral Wireless SENSor Networks (NEWSENS), under Grant 787180. The work of Gabriele Gradoni and Marco Di Renzo was supported by the EU H2020 Reconfigurable Intelligent Sustainable Environments for 6G Wireless Networks (RISE-6G) Project under Grant 101017011. The work of Gabriele Gradoni was supported by the Royal Society under Grant INF\R2\192066.

ABSTRACT Reconfigurable Intelligent Surfaces (RISs) promise improved, secure, and more efficient wireless communications. One less understood aspect relates to the benefits of RIS towards wireless localization and positioning of mobile users and devices. In this paper we propose and demonstrate two practical solutions that exploit the diversity offered by RIS-enhanced indoor environments and to select RIS state configurations that generate easily differentiable radio maps for use with wireless fingerprinting localization estimators. Specifically, we first investigate supervised learning feature selection methods to prune the large state space of the RIS, thus reducing complexity and enhancing localization accuracy and device position acquisition time. We then analytically derive noise correlated heuristics that can further reduce the computational complexity of our proposed solution. Finally, we validate and benchmark our proposed solutions through accurate end-to-end models and computer simulations while demonstrating an average localization accuracy improvement of about 33%. Our explorations thus demonstrate how and why accuracy improvements are achieved and also hint towards how these can be further enhanced in practical localization settings while utilizing more than one RIS.

INDEX TERMS Wireless localization, RIS, 6G, RSSI, fingerprinting, feature selection.

I. INTRODUCTION

Reconfigurable intelligent surface (RIS) technology has given rise to the concept of “smart radio environments” [1] thus unlocking the engineering of the wireless propagation environment itself - a key stepping stone towards the sixth generation (6G) mobile network vision. In a 6G architecture, it is envisaged that the propagation environment itself will be available as a service to improve network performance [2] through RISs which essentially operate to controllably backscatter electromagnetic (EM) signals originating from a network of traditional access points (APs). The backscattering at each RIS is achieved via an array of N metallic elements, e.g.,

The associate editor coordinating the review of this manuscript and approving it for publication was Fakhru Alam¹.

dipoles or patches, which are themselves loaded with tunable circuitry. The RIS and the AP can thus operate in unison to create a tailored EM field that optimizes specific signal characteristics at a dynamic target, e.g., a mobile user (MU) terminal while for example mitigating blockages in a non-line of sight (NLOS) setting. One of the many advances enabled by RIS-enhanced smart radio environments is improved wireless localization [3].

Improved radio localization and positioning of MUs and other internet of things (IoT) devices using RIS capabilities is an important and promising research direction. In current cellular communications for example, the FCC mandate requires network operators to locate those calling 911 to within certain accuracy requirements (50m horizontally and ± 3 m vertically). The need for even more accurate localization is

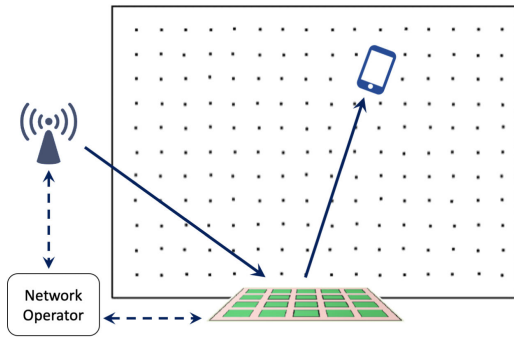


FIGURE 1. System model for RIS-assisted localization. It consists of an AP and a RIS connected to a network operator. A MU attempts to self-localize using radio maps generated jointly by the AP and RIS and sampled along a grid.

echoed but also amplified by many other position-related services, such as, logistics, smart factories, smart cities, autonomous vessels, vehicles, and localized sensing. Therefore, the wealth of wireless localization methods that have been developed and are widely deployed through 4G and now mm-wave 5G frequencies and network densification (broadly classified as fingerprint database, angle-based, range-based, and range-free methods [4]), are now being re-imagined in a 6G RIS-enhanced smart environment setting; a rich and active research field [5]. In addition to improved localization accuracy, RIS-enhanced wireless localization promises to reduce AP deployment costs, signal processing overheads, and to create an unprecedented capillary network for distributed sensing and computation.

This paper focuses on RIS-enhanced wireless fingerprinting localization (WFL) and proposes two practical solutions for improving localization accuracy while reducing computational complexity. One solution is based on machine learning and training data, while the other one is based on correlated noise heuristic estimators. The performance of these two solutions are benchmarked through accurate computer simulations and discussed. Our main contributions are:

- We propose, detail and validate two practical WFL solutions in RIS-enhanced indoor environments that can improve localization accuracy by 33% while also significantly reducing the RIS configuration state space thus improving localization complexity and delay.
- We demonstrate that localization heuristics that account for signal noise correlations can perform just as good as a computationally more expensive supervised learning approach.
- The proposed solutions are generalizable and scalable and do not require large amounts of training data, nor do they require an offline fingerprint measurement campaign.
- Our results demonstrate the different sources of WFL localization errors and also highlight the importance of using accurate simulation tools, especially in RIS-enhanced environments where exotic EM

wave phenomena (e.g., reflections, beamforming, and scattering) control the radio map characteristics.

The paper is structured as follows: in Sec. II we summarize recent works on wireless localization with a focus on RIS-enhanced environments; in Sec. III we present our system model and problem statement while also introducing the notation that we will use throughout the rest of the paper; in Sec. IV we provide an overview of the end-to-end method used to simulate and generate different radio maps as a result of different RIS states; in Sec. V we detail our first proposed solution for selecting a reduce RIS state set leveraging a supervised machine learning framework; in Sec. VI we detail two heuristic state selection solutions, a naive one and a more sophisticated one leveraging noise correlations; in Sec. VII present several computer simulation experiments that validate and benchmark the proposed solutions; and finally in Sec. VIII we summarize our results, discuss their implications and suggest future research questions.

II. BACKGROUND AND MOTIVATION

Traditional WFL approaches are both practical and fairly accurate as they use readily available received signal strength information (RSSI) at the MU while mitigating the effect of wireless signal fluctuations. Research in WFL is quite mature and the technology has been widely deployed and tested [6]. Typically, in WFL applications a dense spatial database of RSSI measurements (i.e., the fingerprint) is constructed during an offline phase usually along a grid of L locations. Then, during an online phase, real-time RSSI measurements are compared with the offline database and matched to estimate the MU location [7].

The standard approach towards enhancing WFL accuracy is to deploy multiple APs thus improving the dimensionality of the fingerprint at the expense of added infrastructure and signal processing costs along with the high cost of training data collection and high computational cost of the subsequent training process. While some of these issues can be mitigated for example through crowdsourcing techniques [8], the use of existing Wi-Fi infrastructure [9], or through additional data points obtained through ray-tracing simulators [10] WFL challenges and potential problems like the high computational costs and relatively unstable positioning accuracy remain unresolved.

In this paper we aim to mitigate some of the shortcomings of traditional WFL systems through the use of smart environments and in particular the RIS ability to flexibly alter the EM radio map environment. Namely, we propose and demonstrate that a single RIS can effectively replace the requirement of multiple APs by improving the dimensionality and diversity of the wireless fingerprint. However, while the RIS can inject a very large number of degrees of freedom into the fingerprint, this induces 1) significant computational costs to the WFL matching algorithm, and 2) significant time-delays towards the creation of the fingerprint since the RIS and MU need some time to transmit and process every single new EM radio map realized by the RIS mask.

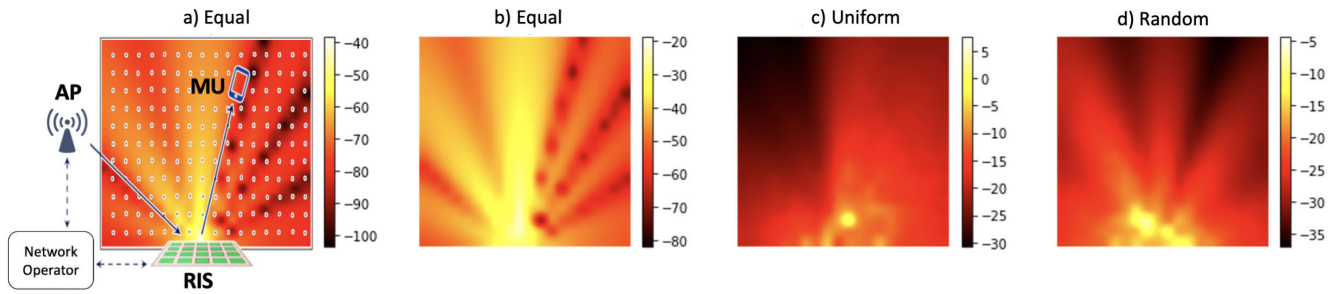


FIGURE 2. a)-d) Examples of four Radio Maps in noise-free environments under different configurations of RIS. The modeled EM radio map corresponds to a $20 \times 20 \text{ m}^2$ environment, using a 2.4 GHz transmitting AP and RIS comprising of $N = 16$ elements. The heat maps represent the RSSI on a dBm scale. a) and b) apply two different uniformly increasing impedance values to the RIS elements thus creating different beam-steering radio maps, c) applies random impedances chosen from a uniform distribution, and d) applies a constant impedance thus simply reflecting the incoming radiation from the AP.

These time-delays can be considered a weakness because the MU device needs to take several RSSI measurements from different RIS configuration settings in order to benefit from different propagation conditions. This process not only takes time but also prolonged measurements may affect localization accuracy for example due to the device motion and change of its location. Our aim is to therefore reduce the dimensionality of the RIS induced fingerprint by efficiently selecting a smaller RIS configuration set that will maximize WFL accuracy.

Despite much interest in RIS-enhanced smart radio environments, e.g., for alleviating multi-path fading, mitigating blockages in non-line-of-sight (NLOS) settings [11], boosting multi-user downlink rates [12], enhancing MIMO diversity and throughput gains [13], maximizing wireless power transfer [14], or improving energy efficiency [15], very little work has been done with respect to wireless localization.

Del Hougne et al, [16] introduced the idea of using the additional configurational degrees of freedom offered by RIS to ink wave fingerprints into the received signal waveform for indoor localization purposes, and demonstrated the concept experimentally. Hu *et al.* [17] calculated lower bounds for point to point positioning accuracy. Huang *et al.* [18] described a DNN-based method for online wireless configuration of the RIS based on fingerprint localization estimates that beam-steer onto the MU thus optimising its RSSI. He *et al.* [19] studied the theoretical performance limits of a single anchor MIMO system using a path loss LOS model while also evaluating the impact of the number of RIS elements, and further proposed adaptive phase shifter designs based on hierarchical codebooks [20]. Ma *et al.* [21] a general model for UWB-aided RIS-assisted indoor positioning was developed where it was also recognized that single AP and single RIS arrangements can deliver significant accuracy improvements and cost reductions to multi-AP indoor localization solutions. Wymeersch *et al.* [3] analyzed a RIS-aided downlink positioning problem from a Fisher Information perspective which the RIS can then use to select the ‘best’ RIS configuration that minimizes MU location uncertainty. And finally, Zhang *et al.* [22], [23] proposed methods that modify

the fingerprint radio map and improves localization accuracy by making RSSI values at adjacent data set locations have significant differences.

Building on the above ideas, this paper proposes two novel practical solutions that exploit the diversity offered by RIS to WFL settings. For simplicity we will consider a basic indoor environment composed of an AP transmitter connected to a RIS through a network operator as shown in Fig. 1. While the figure shows just one MU, the algorithms proposed scale to serve multiple MUs simultaneously.

Our first proposed solution adapts off-the-shelf supervised learning feature selection (SL-FS) methods to find a near-optimal minimal set of RIS configuration. Note that there are $q!/(M!(q-M)!)$ number of ways of configuration selection, thus finding the optimal set of RIS configuration is time-consuming and resource expensive. To speed up the computational search we employ Genetic Algorithms, thus converging towards a near-optimal RIS configuration set for WFL. This method will act as a target benchmark for our second solution which is based on heuristic arguments and unlike the machine learning approach does not require training data. More specifically, we will design and test two heuristic state selection (HSS) solutions. Our naive heuristic (HSS-1) will seek to create maximally differentiable radio maps (and corresponding RIS configurations) as measured by the L^2 norm. In contrast, our noise-correlated heuristic (HSS-2) will re-scale differentiability between adjacent coordinates of the radio maps according to noise strength spatial correlations between candidate positions. We will benchmark all three WFL algorithms (SL-FS, HSS-1, and HSS-2) to the scenario where no pruning of the radio map is applied, i.e., when RIS configurations are simply chosen at random. Finally, we will discuss extensions and generalizations of our work.

Regarding generalizability and scalability, the arsenal of tools already available for use in WFL applications, such as signal denoising, back-end filtering, probabilistic positioning, fingerprint database clustering, etc. can directly be applied to our proposed solutions which are agnostic and therefore compatible to such further enhancements. Importantly and unlike all other WFL publications in the literature

to date, we do not use a pathloss based channel propagation model or fading model (e.g., Friis and Rayleigh) because such models cannot accurately capture the EM coupling effects caused by the RIS elements (e.g., beamform, scatter, null) that can completely change the radio map spatial power distribution. Instead, we use a recently proposed end-to-end model [24] based on impedance coupling of thin wire antennas which allows us to accurately test and benchmark our proposed solutions while also gaining realistic engineering insights into the design capabilities and challenges related to RIS-enhanced WFL methods.

III. SYSTEM MODEL, NOTATION, AND PROBLEM STATEMENT

In typical WFL approaches that use multiple APs, each AP contributes towards one fingerprint. Increasing the number of APs generates a longer fingerprint vector that generally improves the localization accuracy via radio map differentiability and robustness [25]. There have been multiple attempts at reducing the dimension of fingerprint through AP selection [26], [27]. In a RIS-assisted environment, however, using just one AP and the RIS, multiple fingerprints can be created through configuring the RIS in different ways therefore saving infrastructure space and deployment costs [21]. Fig. 2 illustrates four representative example of radio maps corresponding to four different configurations of the RIS in an indoor space generated using our simulation approach described in Sec. IV [24]. Thus, by changing the configuration of the RIS, i.e., the load impedance of the dipoles thus applying a phase shift to the reflected EM waves, one can get a much more diverse set of radio maps, i.e., a high dimensional fingerprint by using just one AP and one RIS.

In the most basic setting, we consider a transmitter (e.g., an AP), a RIS, and a receiver (e.g., a MU) situated within a 2D domain $\mathcal{V} \subset \mathbb{R}^2$, and assume that the RIS and the AP are connected to a network operator that can also control the RIS configuration (see Fig. 1). Our model is intentionally simple but generic enough to generalize to multi-AP and multi-RIS in future studies. Each RIS configuration will generate a different radio map (see Fig. 2.a). The RIS is usually made up of N quasi-passive tuneable elements, often modeled as cylindrical thin wires of perfectly conducting patches, arranged periodically across a grounded dielectric substrate. In Fig. 1 the RIS consists of 4×5 elements for illustration purposes only. Due to hardware limitations, the complex values (amplitude and phase-shifts) applied by the N load impedances of the RIS reflecting elements are usually quantized into D discrete phase values between 0 and 2π [28]. Thus, the RIS can be electronically controlled into any one of $Q = D^N$ possible configurations. Note that Q is usually a very large number. For example, Dai *et al.* [29] have recently experimentally built and tested a RIS operating at 2.3 GHz with $N = 256$ and $D = 4$, leading to $Q = 4^{256}$ possible RIS configurations.

Specifically, in our simulations presented in Sec. VII we have $N = 16$ and $D = 200$, resulting in $Q = 200^{16}$

TABLE 1. List of all frequently used notations found in the paper.

Symbol	Meaning
N	No. of tuneable elements per RIS
D	No. of discrete phase values of a RIS element
Q	No. of possible RIS configurations
M	No. of best RIS configurations to be chosen
L	No. of WFL reference points
l	Index of a reference point
t	Time between successive wireless messages
x	True position of the MU
y	Estimated position of the MU by WFL
\mathcal{V}	Domain of interest of the MU
R	Online RSSI measurement value
\mathbf{R}	Online RSSI fingerprint vector of length M
X	Noise added to RSSI values
σ	Standard deviation of noise
\hat{R}	Offline RSSI measurement value
$\hat{\mathbf{R}}$	Offline RSSI fingerprint vector of length M
$\hat{\mathcal{R}}$	Offline RSSI database matrix of size $L \times M$
ϵ	Localization error

possible RIS configurations. Each of these configurations will generate a slightly different radio map however there is likely to be a huge amount of redundancy and it is possible to choose $M \ll Q$ representative RIS configurations that can reduce complexity while maximizing localization accuracy.

Problem Statement: How to choose those M configurations efficiently thereby reducing the dimensionality of the fingerprint while retaining high WFL accuracy is a key open problem which we will address in the following sections. We remark that similar questions have been considered in the context of data rates and communication quality, but not for wireless localization [30]. Table 1 summarizes all the symbols and terminology that are frequently used in the paper.

For simplicity we assume a basic localization request protocol that can support multiple MUs simultaneously. More secure and advanced protocols are beyond the scope of this paper however the results that we will present are easily generalizable as we discuss in Sec. VIII.

When a WFL request is sent by a MU, the network operator instructs the AP to transmit a sequenced burst of $M \ll Q$ wireless messages periodically every t milliseconds (e.g., every $t = 100$ ms). Meanwhile, the network operator also configures the RIS by re-setting the load impedances periodically every t ms, such that the m -th transmitted message corresponds to the m -th configuration of the RIS ($1 \leq m \leq M$). The MU located at some unknown position $x \in \mathcal{V}$ also receives the M messages from the transmitter and calculates their RSSI values to form an RSSI vector $\mathbf{R}(x) = [R_1(x), R_2(x), \dots, R_M(x)] + [X_1, X_2, \dots, X_M]$, where we have explicitly separated out the noise-less RSSI values R_k and noisy part X_k of the measured RSSI for ease of notation. The MU then compares $\mathbf{R}(x)$ to a fingerprinting database of L radio maps $\hat{\mathcal{R}} = [\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2, \dots, \hat{\mathbf{R}}_L]$, one for each sampled location in \mathcal{V} . Each of the L radio maps is composed of a noise-less and noisy part $\hat{\mathbf{R}}_l = [\hat{R}_{l1}, \hat{R}_{l2}, \dots, \hat{R}_{lM}] + [\hat{X}_{l1}, \hat{X}_{l2}, \dots, \hat{X}_{lM}]$ corresponding to one

of the sampled location coordinates $y_l \in \mathcal{V}$, $l \in [1, L]$ (usually along a rectilinear grid as in Fig. 2) obtained either through computer simulations (e.g., as described in Sec. IV), or during an offline fingerprint measurement campaign and communicated to the MU by the AP. Note that for clarity of notation, we discriminate between the online (real-time) measurement $\mathbf{R}(x)$ at $x \in \mathcal{V}$ and the WFL database $\hat{\mathcal{R}}$ through the use of an over hat ($\hat{\cdot}$) symbol.

There are a number of established algorithms available that compare and match the offline and online measurements, e.g., probabilistic, neural networks, nearest neighbors, etc. [25]. To illustrate their basic principle one can consider the offline radio map database $\hat{\mathcal{R}}$ and compare its vector entries to the measured RSSI vector $\mathbf{R}(x)$ through a permuted Pearson's correlation coefficient

$$l^* = \operatorname{argmax}_{l \in [1, L]} \left(\max_{k \in [1, M]} \frac{\operatorname{cov}(\mathbf{R}(x), \boldsymbol{\pi}^k \hat{\mathbf{R}}_l)}{\sigma_{\mathbf{R}(x)} \sigma_{\boldsymbol{\pi}^k \hat{\mathbf{R}}_l}} \right) \quad (1)$$

where $\boldsymbol{\pi}^k$ is an $M \times M$ permutation matrix that cycles the elements of $\hat{\mathbf{R}}_l$ by k positions to the left. Thus, in (1), the inner max operation finds the largest Pearson's correlation coefficient when comparing the online measured RSSI vector $\mathbf{R}(x)$ to a permutation of the l_{th} offline radio map RSSI vector $\boldsymbol{\pi}^k \hat{\mathbf{R}}_l$, while the outer argmax operation returns the location l^* with the most similar RSSI vector, thus identifying the most likely location of the MU. Note that the Pearson's correlation coefficient is equivalent to the cosine similarity metric of centered (zero mean) vectors. Alternatively, if the measured and stored RSSI vectors do not need any sorting and permuting then one can directly test their similarity and find the MU's most likely location l^* by minimizing the vector difference between the measured RSSI values $\mathbf{R}(x)$ and the stored WFL database one $\hat{\mathbf{R}}_l$ at location l

$$l^* = \operatorname{argmin}_{l \in [1, L]} \left\| \mathbf{R}(x) - \hat{\mathbf{R}}_l \right\| \quad (2)$$

The k nearest neighbors (k -NN) algorithm generalizes (2) slightly by taking a weighted spatial average of the k most similar locations therefore introducing a layer of robustness. Equation (2) describes a deterministic matching algorithm that is robust, easy to understand and implement, and is therefore the most commonly used in WFL literature. While other distance metrics have been shown to improve localization accuracy, we note that their performance will depend on the specific setup and signal characteristics (see comprehensive study [31]) Other matching algorithms for positioning include support vector machines (SVM) and neural networks (NN) for statistical learning from a dataset of fingerprints. Both these techniques are similar in principle to pattern recognition, a mature but computationally demanding field of research.

A. LOCALIZATION ACCURACY

The error in the estimated location of x using the estimator in equation (2) can be calculated though $\varepsilon(x, y_{l^*}) = |x - y_{l^*}|$. Here, x is the true location of the MU, while y_{l^*} is the

estimated location that uses the wireless fingerprint data and the matching algorithm estimator. Averaging over all possible locations of $x \in \mathcal{V}$ provides us with a neat way of representing the absolute expected error in our estimator

$$\mathcal{E} = \frac{1}{|\mathcal{V}|} \int_{\mathcal{V}} |x - y_{l^*}| dx \quad (3)$$

or its root mean squared (RMS) error

$$\mathcal{E}_{\text{RMS}} = \sqrt{\frac{1}{|\mathcal{V}|} \int_{\mathcal{V}} |x - y_{l^*}|^2 dx} \quad (4)$$

Inspecting equation (3) and (4), one can see that the localization error can vary with the location of x , the number of sample locations L , the noise power σ^2 , the performance of the estimator that returns l^* , and the RIS configurations that generate the different radio maps $\hat{\mathcal{R}}$ being compared. Treating the average error as a statistical observable, one can also explore other interesting observables such as the cumulative distribution function (CDF) of the RMS error and its variance, both of which can provide insightful information about the chosen WFL method and its performance. Further standard methodologies used to evaluate indoor localization systems can be found in the ISO/IEC 18305:2016 International Standard, which defines a complete framework for performing Tests and Evaluation of localization and tracking systems [32].

B. LOCALIZATION COMPLEXITY AND DELAY

Further inspection of Equation (2) indicates that the computational complexity of those WFL estimators grows with L and M . Importantly, the process of acquiring the radio map database is a labor intensive and time-consuming effort which grows with L . Also, the time needed for the AP and RIS to transmit and reconfigure the EM spatial distribution grows like $M \times t$, where t is the time interval needed for a RIS configuration update. It is therefore desirable to reduce M and L wherever possible while maintaining high levels of WFL accuracy. Reducing M will lead to a proportional reduction in any time-delays related to a WFL estimation of a MU.

RIS-aided WFL accuracy and performance therefore generally depends on the design of the fingerprinting algorithm employed and the specifications of the RIS and AP being used. While WFL algorithms have been well-studied in the past, the reconfigurability property offered by RIS has introduced many new possibilities to the space, enabling the design of new fingerprinting algorithms and new RIS optimization techniques.

The main aim of this paper is therefore to propose and evaluate practical methods for efficiently selecting the $M \ll Q$ best RIS configurations while also reducing L such that \mathcal{E}_{RMS} is minimized.

IV. RADIO MAP GENERATION

We will use the impedance based model developed in [24] to obtain simulated RSSI values at the MU. While the full power of this model is not required for our simplified WFL setting

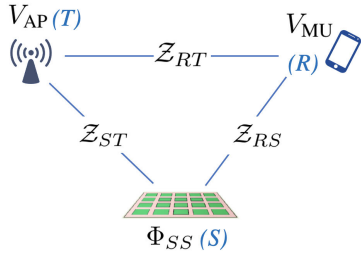


FIGURE 3. Ent-to-End model configuration based on impedance matrices.

(see Fig. 1) we believe it is useful to summarize the assumptions made and also how this model is appropriate towards accurately simulating RSSI values for radio maps in the presence of different RIS configurations. Firstly, we assume that the wireless link in Fig. 2a) is realized with a single-input single-output (SISO) system. Furthermore, we let V_{AP} be the voltage signal given as an input to the AP, and V_{MU} the voltage signal received as an output at the MU. Therefore, an end-to-end (E2E) model between an AP and a MU located at $x \in \mathcal{V}$ can be expressed as a linear complex-valued relation between the two voltage signals as

$$V_{MU} = \mathcal{H}_{E2E} V_{AP}. \quad (5)$$

Fig. 3 shows the configuration modeled by (5), the single antenna transmitter T models the AP, the single antenna receiver R models the MU, and the multi-element surface S models the RIS. A detailed discussion in [24, Corollary 1.] shows that the channel transfer function between AP and MU through the RIS can be found by an order reduction procedure, which yields

$$\mathcal{H}_{E2E} = \mathcal{Y}_0 (\mathcal{Z}_{RT} - \mathcal{Z}_{RS} \Phi_{SS} \mathcal{Z}_{ST}), \quad (6)$$

where \mathcal{Y}_0 includes mismatching factors at the transmitter and receiver ports, and

$$\Phi_{SS} = (\mathcal{Z}_{SS} + \mathcal{Z}_{RIS})^{-1}, \quad (7)$$

is a rank- N matrix expressing the configuration state of the RIS (S), where \mathcal{Z}_{RIS} is a diagonal matrix containing the N ‘tunable’ load impedances terminating at the RIS elements, and \mathcal{Z}_{SS} is a full rank matrix containing the self (diagonal part) and mutual (off-diagonal part) impedances of the RIS elements when their ports are left in open circuit (i.e., when no loads are connected to the element terminals). We remark that the entries of \mathcal{Z}_{RS} and \mathcal{Z}_{ST} in (6) have the meaning of channel gains between the AP (T) or MU (R) antenna and RIS (S) elements. Furthermore, it is worth noticing that the joint amplitude-phase unit cell control is fully captured by the matrix Φ_{SS} , which entails the exact model of the control circuitry via \mathcal{Z}_{RIS} . In particular, for an impedance based transfer function, a practical reflection coefficient has been developed in the model of [33, Sec. III.], where an equivalent circuit has been used to construct a wave-based reflection model. The reflection phase profile across RIS

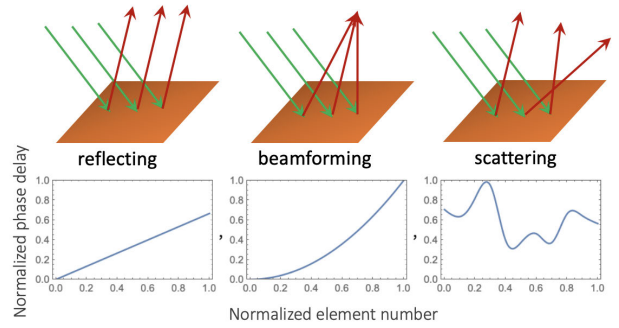


FIGURE 4. Simplified illustrations of some of the different effects a RIS can have on an incoming planar EM wave. In addition to these, the RIS can refract, absorb, polarize, split and collimate incident EM radiation [5]. These effects can be achieved through the effective application of a phase shift mask to different parts of the reflected wave.

elements is configured as a linear phase gradient when the RIS performs a reflection; as a quadratic phase gradient when the RIS performs focusing/beamforming; and as a random phase profile when the RIS performs diffusive scattering. A qualitative representation of those three functionalities is depicted in Fig. 4.

The received voltage at the MU forms the basis to calculate the RSSI at spatial position x or $y_l \in \mathcal{V}$ within the sampling grid used for fingerprinting as $\text{RSSI} = |\mathcal{H}_{E2E}|^2$. The noiseless RSSI value in dBm within the grid position $l \in [1, L]$ and for the selected RIS mask configuration $m \in [1, M]$ is

$$\hat{R}_{lm} = 30 + 10 \log_{10} \text{RSSI}_{lm} \quad (\text{dBm}). \quad (8)$$

Equation (8) therefore takes as input details about the AP (position, transmit power and frequency), the RIS (position and impedance configuration) and outputs the predicted noiseless RSSI at a hypothetical MU located somewhere in \mathcal{V} .

As indicated in the previous section III, we can include a noisy component X of different powers to the received RSSI at the MU. For simplicity, our simulations will only consider additive white Gaussian noise (AWGN) with $X \sim N(0, \sigma^2)$. Varying the noise power σ^2 provides a means for testing the robustness of our proposed localization schemes in Sec. VII. Besides additive noise, rich multi-path fading for indoor environments can be included into the RSSI through the impedance formalism via the random coupling model (RCM) [34]. Including such effects would increase the accuracy of our model and simulations since.

V. SUPERVISED LEARNING APPROACH

To choose a good RIS configuration set that would perform well both in terms of accuracy (i.e., small \mathcal{E}_{RMS}) and complexity (i.e., small M), we propose to leverage off-the-shelf supervised learning (SL) tools to train and thus inform this selection process through a data-driven training phase. Fig. 5 illustrates the whole process of RIS-used localization and evaluation, while the gray rounded rectangle in the figure demonstrates the proposed RIS configuration set selection based on supervised learning approach. The proposed

Algorithm 1 Supervised Learning Based Feature Selection**Output:** Final configuration subset M

- 1: **Training data collection:** Collect q RSSI values corresponding to q RIS configurations at every grid point $l \in [1, L]$.
- 2: **Divide** the training data set consisting of L data points into train data set consisting of L_u data points and validation data set consisting of L_v data points using k -fold method. Note that $L_u + L_v = L$.
- 3: **Initialize** each individual in the population. {An individual corresponds to a set of selected configurations}.
- 4: **while** The termination condition of the iteration is not satisfied **do**
- 5: **for each** Individual Ind in the population **do**
- 6: **Train the location estimator** over training data set using RIS configuration subset corresponding to Ind
- 7: **for each** Element e in validation data set **do**
- 8: **Estimate** the position P_e of e through trained location estimator.
- 9: **Calculate** the localization error $\mathcal{E}(e) = \|P_e - \hat{P}_e\|$ {where \hat{P}_e is the ground truth of the position of e }.
- 10: **end for**
- 11: $\mathcal{E}(Ind) \leftarrow$ Mean of all $\mathcal{E}(e)$ {Localization error}
- 12: $\mathcal{F}(Ind) \leftarrow \mathcal{E}(Ind) + c \times \max(0, |Ind|_1 - M)$ {Fitness value of Ind , see Equation (9)}
- 13: Update the population using Selection, Crossover and Mutation processes
- 14: **end for**
- 15: **end while**

method is detailed further in pseudo-code format in Alg. 1 to aid towards its understanding and potential reproducibility.

First, to map the problem at hand into the realm of SL we consider each RIS configuration as a feature and thus aim to select a set of M features out of a superset of Q possible RIS configurations. Recognizing that the set of all possible RIS configurations Q is often too large of a set, we first construct a smaller representative set q with $|q| = q \ll Q$ chosen through a pseudo-random sampling strategy. This process is called Coarse configuration selection shown in Fig. 5 and detailed in Sec. VII. Using the now reduced RIS configuration set q the corresponding radio maps are created via (8) which will be used as a training *database* (see Fig. 5). In the *Search Algorithm* phase, a supervised learning approach is applied to first create sub groups of $M \ll q$ radio maps, assign each of them with a fitness objective score based on their *accuracy*, and attempt WFL using different *location estimators* (e.g., k -NN, NN, RF); these will be numerically simulated in Sec. VII. To speed up the computational search, a Genetic Algorithm based wrapper is applied to help converge towards a near optimal radio map feature selection whilst avoiding to test all qC_M possible RIS combinations.

For completeness, we describe the implementation of each of these processes.

1) TRAINING DATA COLLECTION

Since the set of all possible RIS configurations Q is often too large of a set, we first try to approximate it by constructing a smaller set q with $|q| = q \ll Q$ chosen at random but diverse enough to enhance diversified fingerprint. Namely, q is created by assigning load impedance values to the RIS dipoles, e.g., chosen from different finite support random distributions or that follow some specific pattern, for example representing a reflection, refraction, beamforming, or a diffuse scattering as illustrated in Fig. 4. In the training phase, we use these q configurations to collect the corresponding q RSSI radio maps and save into an $L \times q$ database.

2) SEARCH ALGORITHM

We aim to identify a subset of $M \ll q$ configurations that leverages the effectiveness of the fingerprinting algorithm. Since a configuration can be considered as a feature, we employ a feature selection method to this end. Among FS approaches, a wrapper approach enhances accuracy because the optimal feature subset is compatible with the specific biases and heuristics of the learning algorithms [35]. Among wrapper methods, Genetic Algorithm based Feature selection algorithm is practical and outperforms other methods in many data sets [36]. We therefore use Genetic Algorithm to search the best RIS configuration subset. We design the Search Algorithm as follows.

2A) Representation of individuals: Each individual, which is encoded by a q -bit binary vector represents a subset of RIS configurations. Bit 1 means that the corresponding configuration is being selected, while bit 0 means the opposite.

2B) Fitness function: The genetic algorithm is designed to minimize the localization error while ensuring the number of selected configurations does not exceed M . To this end, we define the fitness function as follows

$$\mathcal{F}(Ind) = \mathcal{E}(Ind) + c \times \max(0, |Ind|_1 - M) \quad (9)$$

where Ind stands for individual and is a q -bit binary feature vector, $\mathcal{E}(Ind)$ returns the localization error that is calculated by an Accuracy Evaluation process. In (9) the term $c \times \max(0, |Ind|_1 - M)$ is a penalty value, where c is a large positive constant, and $|Ind|_1$ returns the number of selected configurations. The penalty is thus 0 only if the number of selected configurations is less than or equals to M . The genetic algorithm selects the individuals with the smallest fitness.

2C) Selection: We use an elitism selection operator where a small portion of the best individuals from the previous generation is brought forward to the next generation.

2D) Crossover: We use a uniform crossover operator with each bit chosen from either parent with equal probability.

2F) Mutation: An individual has a probability to mutate. We set this probability low enough to inherit good individuals

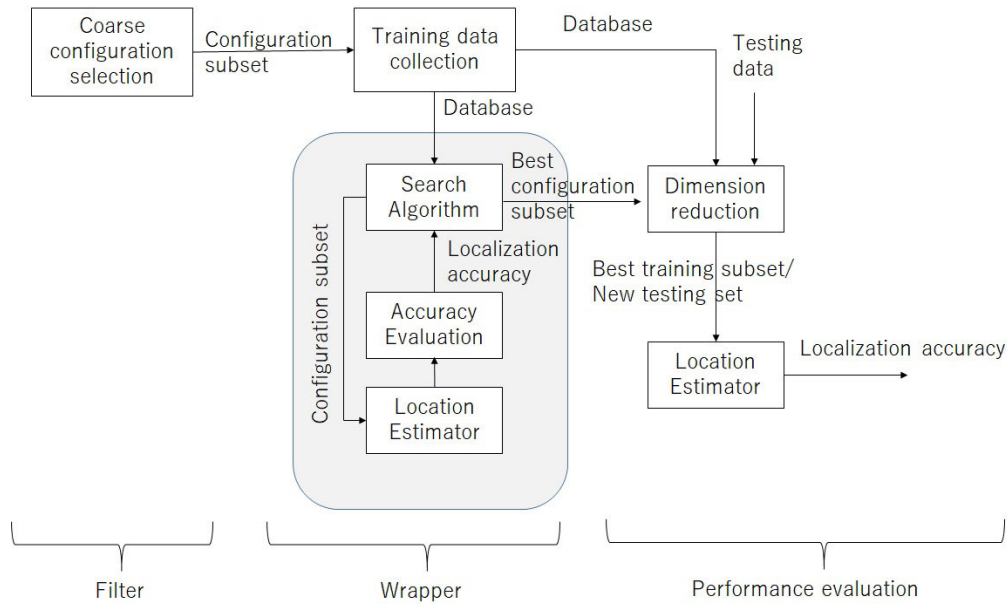


FIGURE 5. Framework of RIS-enhanced wireless fingerprinting localization method using supervised learning-based feature selection techniques.

from the previous generation. Each bit in a chosen individual is flipped with a probability of $1/q$.

2G) Termination: The feature selection process terminates when the number of generations reaches a threshold or the fitness value remains unchanged for a given number of iterations.

3) LOCATION ESTIMATOR

A location estimator can be any regressor which attempts to determine the relationship between one outcome variable, which is the position of MU in this problem, and other known variables, which are RSSI values. We have selected the best regressor from established well-known ones by evaluating their performance using the training database. We first chose three algorithms: k nearest neighbours (k -NN), which enhance accuracy in RSSI-based fingerprinting, neural networks (NN), which performs well in many applications, and Random Forest (RF), which is an ensemble method known as a strong learner.

4) ACCURACY EVALUATION

We use a K -fold cross validation process to evaluate the localization accuracy of the three location estimators (k -NN, NN and RF) using the selected RIS configurations. Namely, we divide the database of radio maps into two subsets: a training subset consisting of L_u data points, and a validation subset consisting of L_v data points such that $L_u + L_v = L$. The location estimator uses the training set to train itself and then applies the trained model on the validation subset. The localization error is then calculated as the mean Euclidean distance between the ground truth position and the estimated position of all items in the validation subset.

The proposed supervised learning-based feature selection process produces the best subset consisting of M configurations. It is combined with the RSSI database constructed through the training data collection process to produce the best training subset used with a location estimator (see Fig.5).

VI. HEURISTIC STATE SELECTION OF RIS CONFIGURATIONS

In this section we will construct two heuristic state selection (HSS) estimators to help compute a good RIS configuration set of size $M \ll Q$ without the need for a training database while only using the radio simulation tools developed in Sec. IV and [24]. We refer to the first estimator as the naive estimator (HSS-1) since it simply attempts to choose RIS combinations which produce maximally different radio maps. We refer to the second estimator as the correlated noise estimator (HSS-2) since it scales the difference metric according to spatial noise correlations.

A. HSS-1: NAIVE SELECTION

A direct approach towards maximizing WFL accuracy is to require the average error over the L different sample locations to be minimized. Moreover, motivated by pattern recognition and image distance metrics, we propose a heuristic that maximises the distance between radio map fingerprints thereby increasing their differentiability and removing any redundancies. For example, if the set of all possible RIS configurations was just the $q = 4$ radio maps shown in Fig. 2, and our algorithm wanted to reduce this set by choosing the $M = 2$ best ones, i.e., the two which would produce high WFL accuracy, then the RIS state corresponding to Figs. 2a), and b), should not be chosen together because they are too similar thus

offering little diversity gain to the fingerprint set. On the other hand, Figs. 2 b) and c) appear to be most different, therefore removing redundancy and making this reduced set a strong candidate for accurate WFL. Note that one would not be able to get such diverse radio map profiles without using the RIS end-to-end model defined in the previous section since each of these maps are produced through the application of different phase shift mask profiles such as the ones illustrated in Fig. 4.

While the human brain is good at choosing similar or different images, we would like to automate, and remove bias from such decision making and enable scalability as both q , M and qC_M become large; a mathematical formulation of similarity and differentiability is needed. To that end, we write down a simple method for choosing the most different RIS subset

$$\mathbf{M}^* = \operatorname{argmax}_{\mathbf{M} \subset \mathbf{q}} \frac{1}{L} \sum_{l=1}^L \Delta(\hat{\mathbf{R}}_l) \quad (10)$$

Equation (10) tries to select the best set \mathbf{M}^* of RIS configurations $\mathbf{M} \subset \mathbf{q}$ which on average maximizes $\Delta(\hat{\mathbf{R}}_l)$ which is a dissimilarity function, i.e., a distance metric between a set of M RSSI values at location l defined in Sec. III. The simplest image distance metric we can use is the Euclidean L^2 norm

$$\Delta(\hat{\mathbf{R}}_l) = \frac{2}{M(M-1)} \sum_{m \neq n} \|\hat{R}_{lm} + \hat{X}_{lm} - \hat{R}_{ln} - \hat{X}_{ln}\|^2 \quad (11)$$

thus giving us a heuristic state selection (HSS-1) method for the RIS. While equation (11) is simple to understand and implement, the dominant issue that we can anticipate with HSS-1 is that this approach suffers from outlier bias. If just one out of the L locations has a very high difference metric, then the whole radio map gets a high Δ score but would not perform well in practice and the average WFL error \mathcal{E}_{RMS} would likely be very high. Equation (11) is one of many possible distance metrics used in image comparison algorithms. Others include the Manhattan distance, mutual information variation, gradient correlation, normalized entropy, etc. An investigation into which one works best, why, and when is beyond the scope of this paper. Instead, we will attempt to model spatial correlations in image distance metrics scaled by the AGWN experienced by the MU when reading RSSI values.

B. HSS-2: CORRELATED NOISE SELECTION

Our starting point is equation (4). Using the law of total probability (4) can be expanded in terms of marginal probabilities

$$\mathcal{E}_{\text{RMS}} = \sqrt{\frac{1}{|\mathcal{V}|} \int_{\mathcal{V}} \sum_{l=1}^L |x - y_l|^2 \times \mathbb{P}(l = l^* | x) dx} \quad (12)$$

that can be interpreted as a double average, where the internal sum over all locations $l \in [1, L]$ gives the weighted average of the localization error, weighted by the probability that location l is indeed the nearest neighbor l^* to x .

Next, we assume that when L is large enough the grid sample points are so densely packed that the localization error is less affected by the inter-grid distance and more affected by the estimation accuracy of l^* captured in (12) by $\mathbb{P}(l = l^* | x)$. We can therefore discretize the integral in equation (12) and rewrite it as

$$\mathcal{E}_{\text{RMS}} \approx \sqrt{\frac{1}{L} \sum_{m=1}^L \sum_{l=1}^L |x_m - y_l|^2 \times \mathbb{P}(l = l^* | x_m)} \quad (13)$$

where we have restricted $x \in \mathcal{V}$ onto the grid of L sample locations which we denote by $x_m, m \in [1, L]$.

Equation (13) provides an opening for us to further analyze $\mathbb{P}(l = l^* | x_m)$ and engineer the selection of RIS radio maps $\hat{\mathbf{R}}$. To that end, we attempt to approximate it by the probability that the RSSI dissimilarity between $\hat{\mathbf{R}}_l$ and $\mathbf{R}(x_m)$ is lower than or equal to the RSSI dissimilarity between \mathbf{R}_{l^*} and $\mathbf{R}(x_m)$

$$\begin{aligned} \mathbb{P}(l = l^* | x_m) &\approx \mathbb{P}(|\hat{\mathbf{R}}_l - \mathbf{R}(x_m)|^2 \leq |\hat{\mathbf{R}}_{l^*} - \mathbf{R}(x_m)|^2) \\ &= \mathbb{P}(|\hat{\mathbf{R}}_l - \mathbf{R}(x_m)|^2 \leq |\hat{\mathbf{R}}(x_m) - \mathbf{R}(x_m)|^2) \\ &= \mathbb{P}\left(\sum_{k=1}^M (\hat{R}_{lk} + \hat{X}_{lk} - R_k(x_m) - X_k)^2 \leq \sum_{k=1}^M (\hat{X}_k - X_k)^2\right) \end{aligned} \quad (14)$$

where in the right hand side of the inequality in the second line we identify the location of the MU x_m with the correct estimator location sample l^* due to the discretization of the sample grid, noting however that the noisy RSSI values of $\hat{\mathbf{R}}(x_m)$ (offline) and $\mathbf{R}(x_m)$ (online) can indeed be different due to the AWGN which is independently calculated and incorporated in them. The intuition here is that in a noiseless idealised environment $\hat{\mathbf{R}}_{l^*} = \mathbf{R}(x_m)$ and thus the only sample location $l \in [1, L]$ where the condition in the first line of (14) is met is when $y_l = y_{l^*} = x_m$. In the presence of uncorrelated AWGN noise however, this condition may fail to other locations $l \neq l^*$, e.g., because $\hat{\mathbf{R}}(x_m)$ and $\mathbf{R}(x_m)$ have experienced significantly different noise levels, or $\mathbf{R}(x_m)$ experiences noise levels that sway its RSSI readings to resemble that of a different location $l \in [1, L] \setminus \{l^*\}$.

We can now use our assumption that the AWGN is an independent identically distributed random variable (i.i.d.) that follows a Gaussian zero-mean distribution with σ^2 variance to simplify equation (14) which can be rewritten as

$$\begin{aligned} \mathbb{P}(l = l^* | x_m) &\approx \mathbb{P}\left(\sum_{k=1}^M (\hat{R}_{lk} - R_k(x_m))^2 \right. \\ &\leq \sum_{k=1}^M 2(\hat{R}_{lk} - R_k(x_m))(\hat{X}_{lk} - X_k) \\ &\quad \left. + (\hat{X}_{lk} - X_k)^2 - (\hat{X}_k - X_k)^2\right) \end{aligned} \quad (15)$$

Observe that the distribution of each bracket in the right hand side of the inequality can be simplified. Firstly, the difference of two Gaussian $Y_k = (\hat{X}_{lk} - X_k)$ is also a zero-mean

Gaussian r.v. with variance $2\sigma^2$ and the sum of M of those becomes

$$W = \sum_{k=1}^M 2(\hat{R}_{lk} - R_k(x_m))Y_k \sim \mathcal{N}(0, 8\sigma^2 \sum_{k=1}^M (\hat{R}_{lk} - R_k(x_m))^2) \quad (16)$$

Similarly, since the sum of M squares of Gaussian r.v.s is Chi-squared distributed χ_M^2 with M degrees of freedom

$$\begin{aligned} \sum_{k=1}^M (\hat{X}_k - X_k)^2 &= Z_1 \sim 2\sigma^2 \chi_M^2 \\ \sum_{k=1}^M (\hat{X}_k - X_k)^2 &= Z_2 \sim 2\sigma^2 \chi_M^2 \end{aligned} \quad (17)$$

Finally, the difference between two correlated Chi-square random variables $Z_1 - Z_2$, each with M degrees of freedom, follows a Variance-Gamma distribution with zero mean and $4M\sigma^2$ variance [37]. Comparing this to the variance in (16), we therefore choose to ignore the contribution of the last line in (15) to arrive at a much simplified expression which is now appropriately scaled by the probability error

$$\begin{aligned} \mathbb{P}(l = l^* | x_m) &\approx \mathbb{P}\left(\sum_{k=1}^M (\hat{R}_{lk} - R_k(x_m))^2 \leq W\right) \\ &= \frac{1}{2} \operatorname{erfc}\left[\frac{1}{4\sigma} \sqrt{\sum_{k=1}^M (\hat{R}_{lk} - R_k(x_m))^2}\right] \end{aligned} \quad (18)$$

where $\operatorname{erfc}(x)$ is the complementary error function. Substituting back into (13) we can now approximately calculate the expected RMS error \mathcal{E}_{RMS} as a function of the offline radio maps \mathcal{R} and the online RSSI measurements at $x_m \in \mathcal{V}$

$$\sqrt{\frac{1}{2L} \sum_{m=1}^L \sum_{l=1}^L |x_m - y_l|^2 \times \operatorname{erfc}\left[\frac{1}{4\sigma} \sqrt{\sum_{k=1}^M (\hat{R}_{lk} - R(x_m)_k)^2}\right]} \quad (19)$$

Equation (19) essentially takes a weighted average of the Euclidean distance between each pair of sample locations x_m and y_l weighted by the probability that a localization error is made which is itself exponentially related to the relative dissimilarity between the respective noiseless RSSI measurements and scaled by the noise power. Note that for noisy environments (i.e., when $\sigma \gg 1$) the weighted probability goes to zero as one would expect, while in a noiseless environment (i.e., when $\sigma = 0$) we have that $\mathbb{P}(l = l^* | x_m) = 1$.

It follows from the derivations above, that to minimize the total localization error one should aim to design or select the M RSSI radio maps which make up the noiseless part of the WFL fingerprint $\hat{\mathcal{R}}$

$$\mathbf{M}^* = \operatorname{argmin}_{M \subset q} \mathcal{E}_{\text{RMS}} \quad (20)$$

where \mathcal{E}_{RMS} in (20) is calculated through equation (19).

C. HSS COMPUTATION FRAMEWORK

Fig. 6 illustrates the whole process of RIS-used localization using HSS configuration selection method. the proposed HSS-based configuration selection is demonstrated in the gray rounded rectangle and Alg. 2, and is demonstrated in further detail as follows.

Algorithm 2 Heuristic State Selection

Output: Final configuration subset \mathbf{M}

- 1: **Ideal Radio Map Generation:** Generate q radio maps corresponding to q configurations, using Equation (8).
 - 2: **Search Algorithm:**
 - 3: Find configuration $c \in q$ such that the expected RMS error estimated by Equation (19) is minimum
 - 4: $\mathbf{M} \leftarrow \{c\}$
 - 5: $q \leftarrow q \setminus \{c\}$
 - 6: **for** $i \leftarrow 2$ to M **do**
 - 7: Find configuration $c \in q$ such that $\mathbf{M} \cup \{c\}$ having smallest expected RMS error due to Equation (19)
 - 8: $\mathbf{M} \leftarrow \mathbf{M} \cup \{c\}$
 - 9: $q \leftarrow q \setminus \{c\}$
 - 10: **end for**
-

1) IDEAL RADIO MAP GENERATION

Similarly to the SL-FS method for training data collection described in Sec. V, since the set of all possible RIS configurations \mathbf{Q} is often too large of a set, we can apply a pseudo random sampling to arrive at a smaller more manageable set q . Then, for each candidate RIS configuration, we generate the corresponding radio map using (8) and store everything in $L \times q$ sized database. These radio maps will then be fed to the *Configuration Selection* process to output a further reduced set which is optimized according to the chosen heuristic HSS-1 or HSS-2 method.

2) SEARCH ALGORITHM

Although the search space of equation (20) is quite large with ${}^qC_M = \frac{q!}{M!(q-M)!}$ possible RIS configurations, it can be rapidly explored through a Greedy Algorithm or a meta heuristic algorithm. In our simulations reported in Sec. VII we have used a Greedy Algorithm since it is simpler to implement. The Greedy configuration selection starts off with a candidate set q , and then selects the best RIS configuration and its corresponding radio map assuming that $M = 1$. Then while holding the selected radio map, it goes through the remaining set of candidate maps (of size $q - 1$) to select the best pair, then triplet, etc. until it has the desired M -tuple. While this is not necessarily the optimal M -tuple solution, the algorithm is simple to implement and also quite fast to compute with a complexity of just $\mathcal{O}(qM)$ which is much smaller compared to other meta-heuristic methods or a brute force search of the whole space. Moreover, the simulation results shown in Sec. VII demonstrate that a Greedy HSS can significantly enhance RIS-aided WFL accuracy.

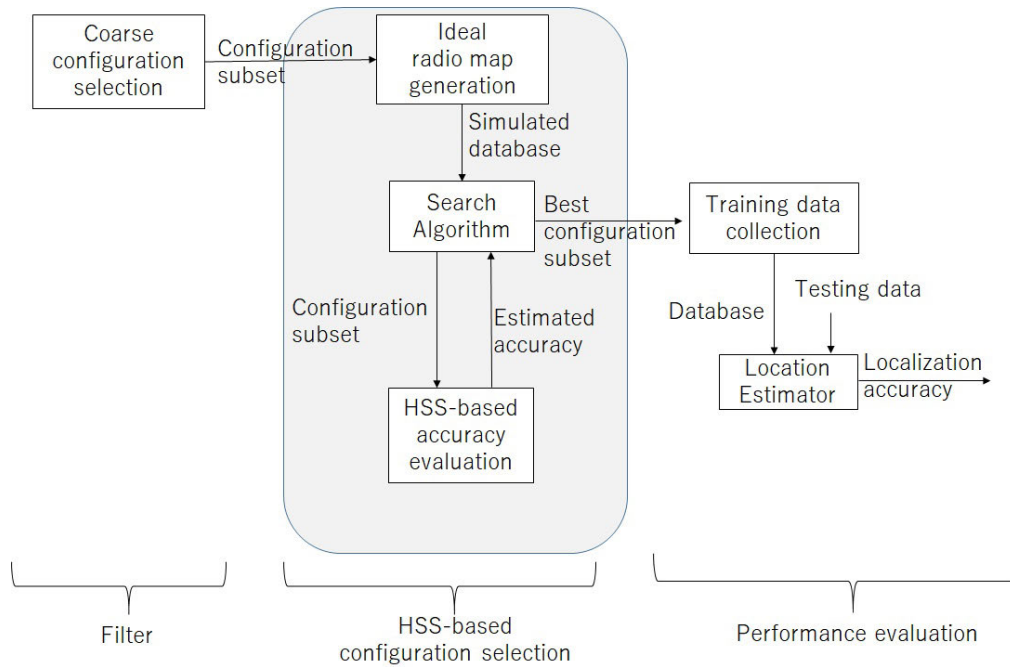


FIGURE 6. Framework of RIS-used localization method using HSS-based configuration selection.

3) HSS-BASED ACCURACY EVALUATION

A configuration subset is evaluated through Eq. (11) if using HSS-1 or Eq. (19) if using HSS-2. The estimated accuracy is fed back to the Search Algorithm to find out the best subset consisting of M RIS configurations.

The best RIS configuration subset is then used to build a database through the Training data collection process. The database is then used with a regression method to build a location estimator (see Fig.6).

VII. PERFORMANCE EVALUATION AND SIMULATIONS

In this section we perform several numerical simulations to evaluate the proposed localization techniques. We assume an indoor space of $20 \times 20 \text{ m}^2$ where the AP is just outside the top left corner of the room, and the RIS comprising of $N = 16$ equally spaced dipoles located at the middle of the bottom wall (see Fig. 1). Both AP and MU are equipped with SISO omnidirectional dipole antennas. In our simulations, the AP emits signals at frequency of 2.4 GHz, with a transmission power of 0.1 Watt as with a traditional WiFi router, while we set the number of RIS elements to $N = 16$ and their load impedance discretization to $D = 200$. A frequency and RIS element count investigation is beyond the scope of this paper, however we can reasonably expect that a higher frequency AP and larger RIS would result in a higher resolution RSSI radio map but with more attenuation from the EM source and reflector. We also suppose that there is a line of sight (LOS) between AP-RIS and RIS-MU, but no direct LOS between AP-MU, which is reasonable if the AP is in a different room

or outdoors, while the MU is attempting localization indoors. It is reasonable to expect that the case where a LOS also exists between AP and MU would result in further enhancements of WFL accuracy. Finally, a zero-mean Gaussian noise X is added to each noise-free RSSI value simulated by equation (8) with standard deviation of 3 dBm.

To compare HSS-1, HSS-2 and SL-FS, we generated $q = 50$ different RIS configurations and simulate the corresponding radio maps at $L = 100$ (sparse $2 \times 2 \text{ m}^2$ grid) and at $L = 400$ (dense $1 \times 1 \text{ m}^2$ grid) locations. The set q was chosen to include: 10 RIS configurations where the dipoles are set to the same value of the RIS dipole impedance thus emulating planar reflection or refraction, 10 configurations where the dipoles have a quadratically increasing impedance value thus emulating beamsteering, and 30 configurations in which the impedance values are randomly chosen thus emulating random (diffuse) scattering as illustrated in Fig. 4. This pseudo-random sampling process allows us to create a diverse RIS configuration set and corresponding radio maps.

We then choose the optimal set M^* following the HSS-1 (10) and HSS-2 approach (20), and then the SL-FS approach on $L_u = L/10$ randomly chosen locations.

We have performed three simulation experiments. In the first experiment, we investigate three well-known location estimators for localization using the FS and HSS-2 approach. Namely, we set $M = 15$ and using the dense sample grid with $L = 400$ we compute the cumulative distribution function (CDF) of the RMS localization error under k -Nearest Neighbors (k -NN), Neural Network (NN), and Random Forest (RF) [38].

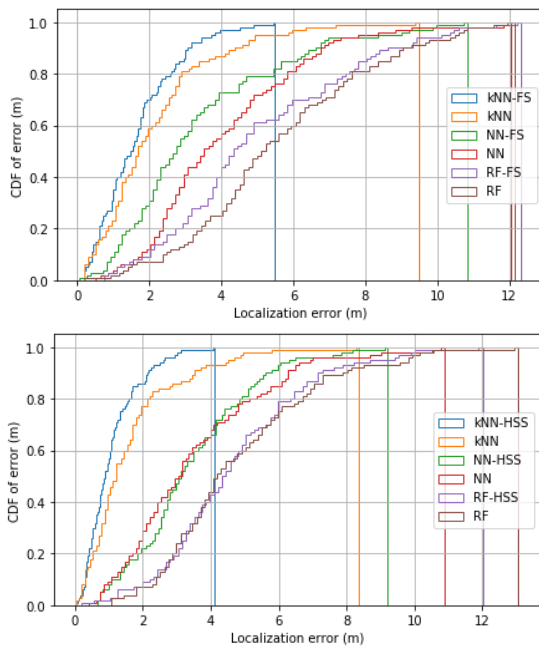


FIGURE 7. Plots of the CDF of the RMS localization error using different supervised learning methods with and without using FS (top) as described in Sec. V and HSS-2 (bottom) as described in Sec. VI.

We set the parameters for these estimators as follows but note that a systematic approach towards their optimization is beyond the scope of this. For each parameter of each estimator, we have only checked a small set of different values and selected the best one based on its localization performance accuracy. For instance, for k -NN, we choose $k = 1, 3, 5, 7$ with uniform weights or inverse distance weights. In our test sample, $k = 5$ together with weights according to the inverse of the distance between the RSSI values received from the MU and the RSSI values registered in the database performed best and are thus reported in the simulations that follow. For NN, we use a Multi-layer Perceptron (MLP) regressor using one hidden layer with 100 nodes. An activation function is set to the rectified linear unit function. The solver for weight optimization is set as Adam, while an L2 penalty parameter (Ridge Regression) and a learning rate are set with values of 0.0001 and 0.001, respectively. The maximum number of iterations is set as 10000. For the RF case, the number of trees in the forest is set to 100, the function that measures the quality of a split is set to the Gini impurity, and nodes are expanded until all leaves are pure [38].

The results of the first experiment with respect to SL-FS are shown in the top sub-figure of Fig. 7 comparing localization errors between the three WFL estimator algorithms (k -NN, NN, RF) with and without SL-FS. In this way we can benchmark the different estimators against each other but also see the individual accuracy enhancement due to the heuristic state selection procedure of Alg. 2. It is observed that SL-FS always has a positive enhancement effect since it shifts the CDF curves to the left and that k -NN has outperformed the other two by a significant margin. While this may be slightly

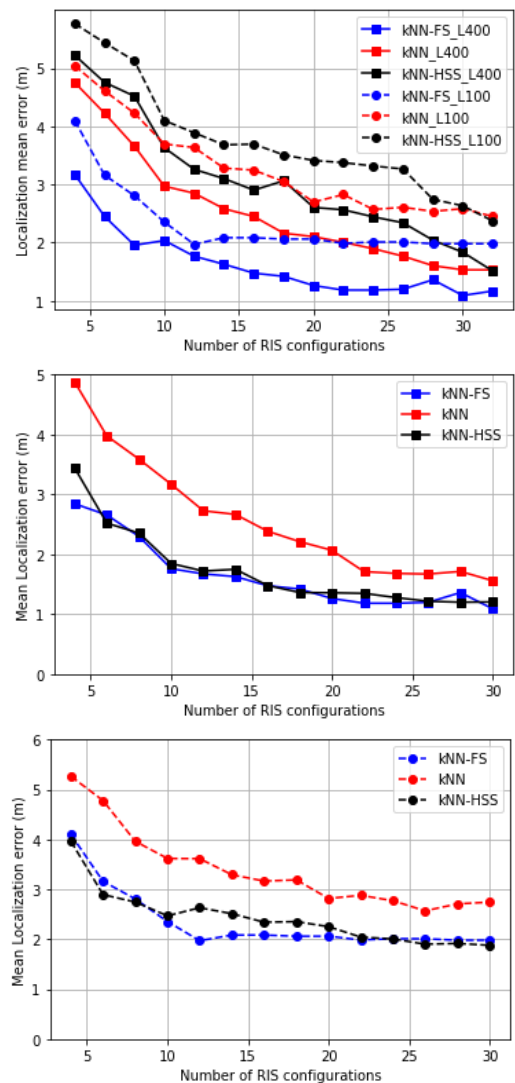


FIGURE 8. Plot of the mean RMS localization error vs. the number of RIS configurations M . The top sub-figure compares between SL-FS, HSS-1 and no state selection methods for dense ($1 \times 1 \text{ m}^2$) and sparse ($2 \times 2 \text{ m}^2$) sampling grids using a k -NN estimator with $k = 5$. The middle and bottom sub-figures compare between SL-FS, HSS-2 and no state selection methods for dense (middle) and sparse (bottom) sampling grids.

surprising, we note that other studies leveraging supervised learning based localization have also observed that k -NN can outperform other learning methods such as NN [6]. This is mainly because accurate NN and RF require a larger training dataset, which we have assumed is not available in practical localization settings. Thus, for our next two simulation experiments k -NN is chosen as the preferred approach since it is also less complex to implement and more robust to RSSI noise fluctuations [39]. Finally, the k -NN SL-FS method being computationally more demanding yet most accurate will act as the near-optimal target benchmark against which we will contrast HSS-1 and HSS-2.

The results of the first experiment with respect to HSS-2 are shown in the bottom sub-figure of Fig. 7 comparing

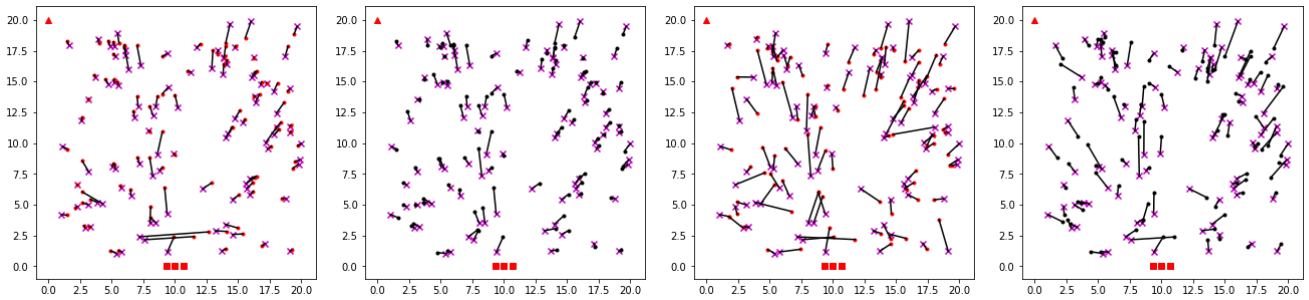


FIGURE 9. Example WFL realizations showing the estimated positions y_{l^*} of 100 MUs vs. their true positions x chosen randomly in \mathcal{V} . Magenta crosses express the ground truth positions of each MU. Red dots express estimated positions using random RIS configurations. Black dots express the estimated MU position using configurations selected through the proposed HSS method. The magenta crosses and red/black dots are connected by black line that indicates the magnitude and direction of the estimation error $e(x, y_{l^*})$. Red triangle and rectangles express the location of AP and RIS respectively. Left two sub-figures have no added noise. Right two sub-figures employ AWGN with $\sigma = 3\text{dBm}$.

localization errors between the three WFL estimator algorithms (k -NN, NN, RF) with and without HSS-2. In this way we can benchmark the different estimators against each other but also see the individual accuracy enhancement due to the noise correlated heuristic selection procedure of Alg. 2. It is observed that here too, k -NN outperformed the other two estimators (NN and RF) by a significant margin, possibly for the same reason as above (i.e., small training datasets). Importantly, we observe that HSS-2 offers a significant enhancement towards localization accuracy only in the case of k -NN. This is because HSS-2 is developed on the assumption of using k -NN (with $k = 1$) as the learner, and not some other estimator.

This can be understood through the assumptions that lead to equation (13) which is constructed under the assumption that the nearest neighbor is usually chosen as the estimated position.

In the second experiment, we investigate the k -NN localization accuracy between SL-FS, HSS-1 and HSS-2 as a function of M using both the sparse and dense sample grids, $L = 100$ and $L = 400$, respectively. This experiment will therefore provide some insight both in terms of sampling density and in terms of the performance of the proposed low-complexity heuristic state selection WFL solutions. The results for the second experiment are shown in Fig. 8. As expected, we observe that all algorithms converge towards the sampling grid-size of 1 or 2 meters with increasing resolution in M . We note that fluctuations and non-monotonicity in the localization accuracy is expected since both the Genetic and Greedy algorithms implemented in our simulations do not guarantee convergence towards an optimal fingerprint set of size M .

An interesting observation arising from Fig. 8 is that SL-FS performs the best, while HSS-1 performs the worst. In fact HSS-1 is even worse than random RIS configuration selection (see top sub-figure in Fig. 8). We ascribe this to a badly designed heuristic which suffers from outlier biases that tend to maximize the naively chosen distance metric (11). Meanwhile, we observe that the HSS-2 approach performs almost as well as the computationally expensive SL-FS method

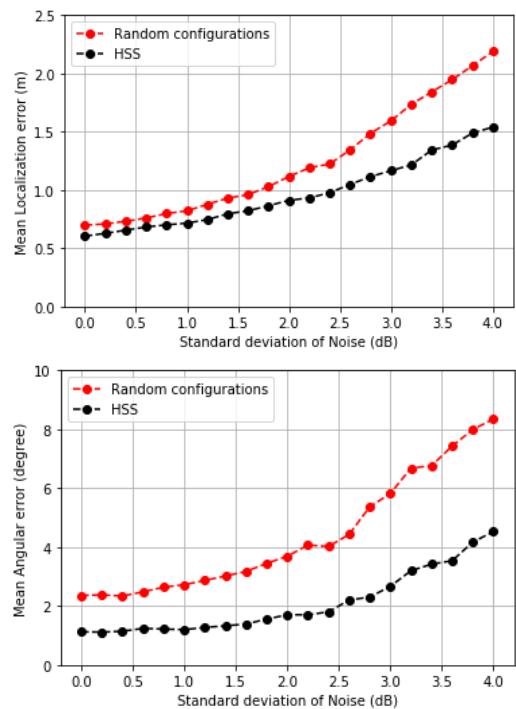


FIGURE 10. Top: RMS error vs. noise level. Bottom: RMS angular error vs. noise level. Red and black dots indicate the error using random RIS configurations and configurations selected by using the proposed HSS-2 method, respectively. Angular errors are calculated as the absolute angular deviation from the polar coordinate of the MU's position.

(see middle and bottom sub-figures in Fig. 8). We think that the good performance of HSS-2 is because the noise of the simulated RSSI data follows a Gaussian distribution, which makes the assumption in HSS-2 correct. However, this may not be true in more exotic situations with different types of correlated noise in which case the SL-FS would probably be most robust against. Importantly, we note that one can trade-off Radio Map resolution L by applying SL-FS or HSS-2, or by using a larger optimized fingerprint M . For example, a mean localization error of 2m can be achieved by having $L = 400$ (dense) grid points with $M = 20$ random RIS configurations, or by using $L = 100$ (sparse) grid points with

SL-FS and $M = 12$, thus saving both time and complexity but not sacrificing accuracy.

In the third simulation experiment, we investigate the effects of noise onto the proposed HSS-2 solution. The simulation employs a k -NN localization estimator with $k = 5$, $M = 30$ and $L = 400$. The results for the third experiment are shown in Figs. 9 and 10. In Fig. 9 we observe that the location estimation error is significantly affected by noise, especially if no RIS state selection method is applied. HSS-2 seems to be a good mitigation strategy. Further, we observe that the estimation errors are in general aligned radially outwards from the RIS location. This is to be expected due to the general reflective nature of the RIS and could be potentially much improved by including a second RIS, preferably on the adjacent wall to the existing RIS, as to provide stereo information. Fig. 10 further amplifies our observation that the use of HSS-2 for RIS configurations is very effective in combating noise and in reducing both the magnitude and the angular RMS errors in WFL.

VIII. CONCLUSION AND DISCUSSION

Reconfigurable intelligent surfaces (RISs) promise great advancements and cost savings and may play a key role in upcoming 6G wireless systems [1]. In this paper we have investigated wireless fingerprinting localization (WFL) [6] in a RIS-enhanced setting and have proposed and evaluated novel and practical localization algorithms with the main aim of reducing complexity while maximizing WFL accuracy. We have argued that while a single RIS can inject a large number of dimensions to the wireless fingerprint vector, many of them are redundant and should be removed as to avoid unnecessary processing and save time when attempting to localize. Such enhancements can reduce the cost and usability of indoor localization solutions, e.g., for way-finding, object tracking, and other location-based services.

To that end, we have proposed both machine learning and heuristic algorithms for pruning the state space of the RIS of size Q and selecting a significantly smaller subset of size M that can still result in near-optimal localization accuracy. Our machine learning approach uses a supervised learning feature selection (SL-FS) method to first train and then identify the RIS configuration set that would minimize localization error. Our implementation leveraged off-the-shelf tools such as k -NN matching estimators and Genetic Algorithms. Our two heuristic state selection approaches (HSS-1 and HSS-2) used heuristics to maximize radio map differentiability (HSS-1) or to minimize correlated noise within the localization domain and at nearby candidate locations (HSS-2). The former was shown to be a bad estimator, while the latter performed almost as good as the computationally more expensive SL-FS. Several computer simulation experiments were performed to investigate and benchmark the proposed WFL algorithms. Importantly the simulations employed a novel end-to-end model [24] based on impedance coupling of thin wire antennas thus capturing the rich scattering effects of the RIS. As seen in Fig. 8.b), HSS-2 and SL-FS can

improve WFL accuracy by about 33% as compared to a random selection of RIS configurations. Thus, we have demonstrated significant complexity reductions as well performance enhancements.

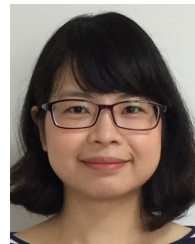
The practicality and generalizability of the proposed methods follows from the simplicity of our system model (see Fig. 1) and also from the use of off-the-shelf solutions such as the k -NN localization estimator, Genetic and Greedy Algorithm implementations. RIS-enhanced WFL can be exploited by multiple users simultaneously and does not require large amounts of training data and is therefore a scalable solution, ideal for indoor scenarios where the radio environment is not too dynamic, e.g., office or warehouse. The WFL accuracy has converged in all of our experiments to the sample grid resolution, thus suggesting that further improvements could not be achieved by higher frequency AP or larger RIS element counts N , unless the domain space is sampled more densely thus emphasizing the need for indoor radio map simulation software that can also account for the wave phenomena induced by the RIS (see Fig. 4 and c.f. [24]).

An interesting next step to our findings would be to study the effect of multiple RIS deployments and also their spatial arrangements and the effects and limitations introduced by dynamic human movement and EM blockages due to obstacles and walls. Further, having provided an initial validation of our proposed solutions that significantly reduce some of the risks and complexities associated to RIS-enhanced WFL, we would encourage followup experimental studies using real data and real deployments.

REFERENCES

- [1] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, and J. De Rosny, "Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, 2019.
- [2] E. Calvanese Strinati, G. C. Alexandropoulos, V. Sciancalepore, M. D. Renzo, H. Wymeersch, D.-T. Phan-Huy, M. Crozzoli, R. D'Errico, E. D. Carvalho, P. Popovski, P. D. Lorenzo, L. Bastianelli, M. Belouar, J. E. Mascolo, G. Gradoni, S. Phang, G. Lerosey, and B. Denis, "Wireless environment as a service enabled by reconfigurable intelligent surfaces: The RISE-6G perspective," in *Proc. EUCNC 6G Summit*, 2021, pp. 562–567.
- [3] H. Wymeersch and B. Denis, "Beyond 5G wireless localization with reconfigurable intelligent surfaces," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
- [4] Z. Xiao and Y. Zeng, "An overview on integrated localization and communication towards 6G," 2020, *arXiv:2006.01535*. [Online]. Available: <https://arxiv.org/abs/2006.01535>
- [5] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [6] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on Wi-Fi: An overview," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, p. 135, Apr. 2017.
- [7] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2568–2599, 3rd Quart., 2017.
- [8] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 444–457, Feb. 2015.

- [9] J. Niu, B. Wang, L. Cheng, and J. J. P. C. Rodrigues, "WicLoc: An indoor localization system based on WiFi fingerprints and crowdsourcing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 3008–3013.
- [10] P.-H. Tseng, Y.-C. Chan, Y.-J. Lin, D.-B. Lin, N. Wu, and T.-M. Wang, "Ray-tracing-assisted fingerprinting based on channel impulse response measurement for indoor positioning," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 5, pp. 1032–1045, May 2017.
- [11] E. Björnson, O. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Feb. 2020.
- [12] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.
- [13] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, "Reconfigurable intelligent surface aided NOMA networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, Nov. 2020.
- [14] H. Yang, X. Yuan, J. Fang, and Y.-C. Liang, "Reconfigurable intelligent surface aided constant-envelope wireless power transfer," *IEEE Trans. Signal Process.*, vol. 69, pp. 1347–1361, 2021.
- [15] 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, Aug. 2019.
- [16] P. del Hougne, M. F. Imani, M. Fink, D. R. Smith, and G. Lerosey, "Precise localization of multiple noncooperative objects in a disordered cavity by wave front shaping," *Phys. Rev. Lett.*, vol. 121, no. 6, Aug. 2018, Art. no. 063901.
- [17] S. Hu, F. Rusek, and O. Edfors, "Cramér-rao lower bounds for positioning with large intelligent surfaces," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–6.
- [18] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.
- [19] J. He, H. Wymeersch, L. Kong, O. Silvén, and M. Juntti, "Large intelligent surface for positioning in millimeter wave MIMO systems," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC)*, May 2020, pp. 1–5.
- [20] J. He, H. Wymeersch, T. Sanguanpuak, O. Silven, and M. Juntti, "Adaptive beamforming design for mmWave RIS-aided joint localization and communication," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.
- [21] T. Ma, Y. Xiao, X. Lei, W. Xiong, and Y. Ding, "Indoor localization with reconfigurable intelligent surface," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 161–165, Jan. 2021.
- [22] H. Zhang, J. Hu, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "MetaRadar: Indoor localization by reconfigurable metamaterials," *IEEE Trans. Mobile Comput.*, early access, Dec. 14, 2020, doi: 10.1109/TMC.2020.3044603.
- [23] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "Towards ubiquitous positioning by leveraging reconfigurable intelligent surface," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 284–288, Jan. 2020.
- [24] G. Gradoni and M. Di Renzo, "End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *IEEE Wireless Commun. Lett.*, vol. 10, no. 5, pp. 938–942, May 2021.
- [25] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless RSSI fingerprinting localization," *Signal Process.*, vol. 131, pp. 235–244, Feb. 2017.
- [26] Y. Chen, Q. Yang, J. Yin, and X. Chai, "Power-efficient access-point selection for indoor location estimation," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 7, pp. 877–888, Jul. 2006.
- [27] P. Huang, H. Zhao, W. Liu, and D. Jiang, "Maps: Indoor localization algorithm based on multiple ap selection," *Mobile Netw. Appl.*, vol. 26, no. 2, pp. 1–8, 2020.
- [28] P. Xu, G. Chen, Z. Yang, and M. D. Renzo, "Reconfigurable intelligent surfaces-assisted communications with discrete phase shifts: How many quantization levels are required to achieve full diversity?" *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 358–362, Feb. 2020.
- [29] L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. Di Renzo, and C. B. Chae, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45913–45923, 2020.
- [30] H. Zhang, B. Di, L. Song, and Z. Han, "Reconfigurable intelligent surfaces assisted communications with limited phase shifts: How many phase shifts are enough?" *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4498–4502, 2020.
- [31] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte, and J. Huerta, "Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems," *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9263–9278, 2015.
- [32] *Information Technology—Real Time Locating Systems—Test and Evaluation of Localization and Tracking Systems*, Standard ISO/IEC 18305:2016, ISO/IEC JTC 1/SC 31, Standard ISO/IEC 18305, 2016.
- [33] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.
- [34] G. Gradoni, J.-H. Yeh, B. Xiao, T. M. Antonsen, S. M. Anlage, and E. Ott, "Predicting the statistics of wave transport through chaotic cavities by the random coupling model: A review and recent progress," *Wave Motion*, vol. 51, no. 4, pp. 606–621, Jun. 2014.
- [35] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature Extraction, Construction Selection*. Boston, MA, USA: Springer, 1998, pp. 117–136.
- [36] B. Oluleye, L. Armstrong, J. Leng, and D. Diepeveen, "A genetic algorithm-based feature selection," *Int. J. Electron. Commun. Comput. Eng.*, vol. 5, no. 4, pp. 899–905, 2014.
- [37] B. Klar, "A note on gamma difference distributions," *J. Stat. Comput. Simul.*, vol. 85, no. 18, pp. 3708–3715, 2015.
- [38] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
- [39] J. Zhao, "Analyzing the robustness of deep learning against adversarial examples," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 5133–5142.



CAM LY NGUYEN received the B.Sc. degree in computer science and engineering from Osaka University, in 2010, and the M.Sc. and Ph.D. degrees in computer science from The University of Tokyo, in 2019. She has been with the Corporate Research and Development Center, Toshiba Corporation, Tokyo, as a Research Engineer, with a focus on wireless networks. She has filed over ten patent applications and has authored or coauthored nine papers as the first author in leading journals and conferences of Internet and communications. Her current research interests include wireless localization, the Internet of Things, combinatorial optimization, and machine learning.



ORESTIS GEORGIU (Senior Member, IEEE) received the Ph.D. degree in mathematics from the University of Bristol, in 2012. He was a Postdoctoral Researcher with Max Planck Institute of Complex Physics, Dresden, in 2013. He then moved to industry, first working with Toshiba Telecommunications Research Laboratory, as a Senior Researcher, from 2013 to 2017, and later with Ultraleap as the Director of Research, since 2017. He is the author of five patents and over 80 academic articles published in leading journals and conferences of mathematics, physics, computer science, engineering, and medicine. He was awarded a Marie Curie Individual Fellowship hosted at the University of Cyprus, from 2019 to 2021, is the author of two best paper awards (ISWCS'2013 and AutoUI'2018). He was a recipient of the prestigious IEEE Heinrich Hertz Award in 2019.



GABRIELE GRADONI (Member, IEEE) received the Ph.D. degree in electromagnetics from the Università Politecnica delle Marche, Ancona, Italy, in 2010. He was a Visiting Researcher with the Time, Quantum, and Electromagnetics Team, National Physical Laboratory, Teddington, U.K., in 2008. From 2010 to 2013, he was a Research Associate with the Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD, USA. From 2013 to 2016, he was a

Research Fellow with the School of Mathematical Sciences, University of Nottingham, U.K. Since 2020, he has been an Associate Professor of mathematics and electrical engineering with the University of Nottingham. His research interests include probabilistic and asymptotic methods for propagation in complex wave systems, wave chaos, and metasurfaces, with applications to electromagnetic compatibility and modern wireless communication systems. He is a member of the American Physical Society, and the Italian Electromagnetics Society. He received the URSI Commission B Young Scientist Award in 2010 and 2016, the Gaetano Latmiral Prize in 2015, and the Honorable Mention IEEE TEMC Richard B. Schulz Transactions Prize Paper Award in 2020. From 2014 to 2021, he was the URSI Commission E Early Career Representative. Since 2020, he has been a Royal Society Industry Fellow at the Maxwell Centre, Cavendish Laboratory, University of Cambridge, U.K., and an Adjunct Associate Professor with the Department of Electrical and Computer Engineering, University of Illinois, Urbana Champaign, USA.



MARCO DI RENZO (Fellow, IEEE) received the Laurea (*cum laude*) and Ph.D. degrees in electrical engineering from the University of L'Aquila, Italy, in 2003 and 2007, respectively, and the Habilitation à Diriger des Recherches (D.Sc.) degree from University Paris-Sud (now Paris-Saclay University), France, in 2013. Since 2010, he has been with the French National Center for Scientific Research (CNRS), where he is currently a CNRS Research Director (Professor) with the Laboratory of Sig-

nals and Systems (L2S) of Paris-Saclay University -CNRS and CentraleSupélec, Paris, France. In Paris-Saclay University, he serves as the Coordinator for the Communications and Networks Research Area of the Laboratory of Excellence DigiCosme, and as a Member of the Admission and Evaluation Committee of the Ph.D. School on Information and Communication Technologies. He is a fellow of the U.K. Institution of Engineering and Technology (IET), an Ordinary Member of the European Academy of Sciences and Arts (EASA), and an Ordinary Member of the Academia Europaea (AE). He is also a Highly Cited Researcher. He has received several research distinctions, which include the SEE-IEEE Alain Glavieux Award, the IEEE Jack Neubauer Memorial Best Systems Paper Award, the Royal Academy of Engineering Distinguished Visiting Fellowship, the Nokia Foundation Visiting Professorship, the Fulbright Fellowship, and the 2021 EURASIP Journal on Wireless Communications and Networking Best Paper Award. He is the Editor-in-Chief of IEEE COMMUNICATIONS LETTERS and a Distinguished Speaker of the IEEE Vehicular Technology Society. From 2017 to 2020, he was a Distinguished Lecturer of the IEEE Vehicular Technology Society and the IEEE Communications Society.

...