Electromagnetic-informed generative models for passive RF sensing and perception of body motions

Stefano Savazzi\textsuperscript{1}, Federica Fieramosca\textsuperscript{2}, Sanaz Kianoush\textsuperscript{1}, Michele D’Amico\textsuperscript{2}, and Vittorio Rampa\textsuperscript{1}

\textsuperscript{1}Consiglio Nazionale delle Ricerche, IEIIT institute
\textsuperscript{2}DEIB, Politecnico di Milano

March 06, 2024
Electromagnetic-informed generative models for passive RF sensing and perception of body motions

Stefano Savazzi1, Federica Fieramosca2, Sanaz Kianoush1, Michele D’Amico2, Vittorio Rampa1
1 Consiglio Nazionale delle Ricerche, IEIIT institute, Piazza Leonardo da Vinci 32, I-20133, Milano, Italy.
2 DEIB, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133, Milano, Italy.

Abstract—Electromagnetic (EM) body models predict the impact of human presence and motions on the Radio-Frequency (RF) field originated from wireless devices nearby. Despite their accuracy, EM models are time-consuming methods which prevent their adoption in strict real-time computational imaging problems and Bayesian estimation, such as passive localization, RF tomography, and holography. Physics-informed Generative Neural Network (GNN) models have recently attracted a lot of attention thanks to their potential to reproduce a process by incorporating relevant physical laws and constraints. They can be used to simulate or reconstruct missing data or samples, reproducing EM propagation effects, approximated EM fields and learn a physics-informed data distribution, i.e., a Bayesian prior. The paper discusses two popular techniques, namely variational auto-encoders (VAEs) and generative adversarial networks (GANs), and their adaptations to incorporate relevant EM body diffraction concepts. Proposed EM-informed generative models are verified against classical EM tools driven by diffraction theory and validated on real data. Physics-informed generative machine learning represents a multidisciplinary research area weaving together physical/EM modelling, signal processing and artificial intelligence (AI): the paper explores emerging opportunities of GNN tools targeting real-time passive RF sensing in multiple-input multiple-output (MIMO) communication systems. Proposed generative tools are designed, implemented and verified on resource constrained wireless devices being members of a Wireless Local Area Network (WLAN).

Index Terms—EM body models, generative models, variational autoencoders, generative adversarial networks, radio tomography, integrated sensing and communication, localization.

I. INTRODUCTION

PASSIVE radio sensing employs stray ambient radio signals from Radio Frequency (RF) devices to detect, locate, and track people that do not carry any electronic device, namely device-free [1]- [4]. In line with the Communication while Sensing paradigm [4], these tools provide seamless detection capabilities, while performing wireless communications. Radio signals encode a view of all moving/fixed objects traversed by their propagation, while several data analytic tools, such as Bayesian [5], [6] and machine learning approaches [7] can be used to decode this information, typically by large-scale processing of radio signals exchanged by different devices.

Most of emerging approaches proposed for solving the radio sensing problem require an approximated knowledge of a physical-informed (prior) model to interpret the effects of human subjects on radio propagation. The perturbative effects of the radio signals induced by the presence or movements of human bodies can be interpreted using electromagnetic (EM) propagation theory considerations [8]. These have paved the way to several physical and statistical models for passive radio sensing, which exploit full wave approaches, ray tracing, moving point scattering [9] and diffraction theory [10]- [13]. The body-induced perturbations that impair the radio channel, can be thus acquired, measured, and processed using model-based methods to estimate location and track target information. A general EM model for the prediction of body-induced effects on propagation is still under scrutiny [14]. On the other hand, current models are too complex to be of practical use for real-
Time sensing scenarios [16], although they can be used for pre-deployment assessment [20].

Physics-informed generative machine learning [21] is an emerging field in different application contexts ranging from imaging, EM field computation, to Bayesian estimation for inverse problems. For example, generative deep neural networks (GNN) can be trained to produce observations drawn from a distribution which reflects the complex underlying physics of the environment under study, or rather reproduce approximate fields in an almost negligible time compared with classical numerical methods [30], [31]. For the first time, the paper discusses the adoption of GNN models designed to reproduce as close as possible the effects of body movements on EM propagation, considering varying size, position and orientation/posture of the body, multiple antenna (MIMO) setups and different physical properties of the radio link(s). Proposed physics-informed GNN models are trained with samples obtained from EM models based on diffraction theory [12], [13], [16], under different environment configurations. Generative models are validated using both EM field simulations and measurements. GNN tools discussed in this paper are based on variational auto-encoders (VAEs) [18] and generative adversarial networks (GANs) [17], [19]. Opportunities and limitations of each proposed approach are discussed and compared in several case studies targeting the perception of body motions and passive RF sensing applications.

A. Generative models for reproducing EM body effects

RF sensing, depicted in Fig. 1, generally targets an inverse problem where the goal is to isolate or extract the EM effects (Eθ) of the human body(ies) from noisy measurements St of the RF radiation observed at time t. The human subject(s) are characterized by an unknown state θ, such as location, size/shape, orientation [5], [12], [13], which is recovered from the underlying reconstructed data Eθ. Under Bayesian formulation, the objective is to maximize the a posteriori distribution p(Eθ|St)

\[
p(E_\theta|S_t) = \frac{p(S_t|E_\theta) \cdot p(E_\theta)}{p(S_t)}
\]  

(1)

of the (unknown) body-induced effects Eθ, given the measurements St. Body effects are generally represented as EM field samples Eθ, however they can be also be evaluated in terms of body-induced RSS attenuations Aθ, as in the example of Fig. 1, or channel state information (CSI), Cθ. Measurements are hardware-specific: they can be in the form of received power, Received Signal Strength (RSS), or base-band channel impulse response [4] obtained from devices equipped with either single or multiple antenna frontends. Observations St are perturbed by the body movements according to a prior-distribution, p(Eθ), which models the expected effects of the body (or the target) in state θ as the result of the propagation of the reflected, scattered, and diffracted EM waves. Maximum A-Posteriori (MAP) solution to the inverse problem (1) allows to extract the most likely body effects \( \hat{E}_\theta, \hat{\theta} \) = argmaxp(Eθ|St), from which it is possible to recover the subject state \( \hat{\theta} \) and any feature (\( \vartheta \in \hat{\theta} \)) of interest, e.g. the body position, orientation, or size.

The Bayesian approach (1) for solving the radio sensing problem requires the knowledge of the RF measurement model, or the likelihood function p(St|Eθ), and the prior distribution p(Eθ), namely a model to interpret the body-induced EM effects as a function of the subject state θ. While the likelihood term depends on the data collection process as well as on the impairments introduced by the measurement instrument or the environment, the prior distribution p(Eθ) is usually hard to model as it often requires full wave EM approaches, or approximated solutions, which are in many cases too time-consuming to be of practical use for real-time sensing scenarios [16].

The paper proposes the use of EM-informed GNN tools designed to learn the prior distribution p(Eθ) under different configurations of the target(s), and generate RF signals ranging from body-induced attenuations to EM field samples. The off-line training of the proposed generative neural networks is optimized to match the distribution p(Eθ) using (few) examples obtained from an EM body model which is based on the scalar diffraction theory [12], [13]. Accurate learning of the prior p(Eθ) allows the generative model to reproduce body effects under target or link configurations which might be unseen during the training phase, or rather impossible to reproduce through traditional EM field computing methods based on numerical integration. Generative modeling is also well-suited for real-time target tracking implementations as it does not need an ad-hoc generation of model samples, that might require intensive computations depending on the target size, number, and the environment (walls, floor, ceiling, and other obstacles).

B. Related works and contributions

Physics-informed generative models use machine learning (ML) for computing approximate EM fields or physical processes. Although still in their infancy, a small body of existing works related to this problem does exist. A ML model is proposed in [27] to obtain an approximation of the EM electromagnetic field in a cavity with an arbitrary spatial dielectric permittivity distribution. The model is shown to be one order of magnitude faster than similar finite-difference frequency-domain simulations, suggesting possible applications in inverse problems. In [28] a neural solver for Poisson’s equation is proposed using a purely-convolutional neural network structure. An approach for solving partial differential equations (PDEs) using neural networks (NNs) has recently emerged [29], where a physics-based loss function is constructed to improve NN training. Compared to traditional EM field computing methods based on numerical integration and/or mesh-based methods, an attractive feature of physics-informed models based on deep NN (DNN) implementations, is that they could break the curse of dimensionality [25]. In addition, once trained, DNNs can solve an EM problem in an almost negligible time in comparison to classical numerical methods [30], [31]. Finally, generation accuracy and training
time can be improved by incorporating a small amount of “labeled” data or EM field measurements (if available) during the training process.

Generative models for RF propagation characterization with applications to communications and localization are also emerging [32]. For example, [33] discussed a convolutional encoder-decoder structure that can be trained to reproduce the results of a ray-tracer, encoding also physics-based information of an indoor environment. A ML-assisted channel modeling approach is proposed in [34] to generate site-specific mmWave channel characteristics. The model is shown to improve the generalization capabilities of conventional physical–statistical models when adopted to reproduce complex network configurations. A multibranch GAN (MBGAN) has been recently proposed for radar signal processing to synthesize data that reproduces the body-induced EM diffraction effects observed by wireless receivers equipped with single and multiple antennas, as well as the array response in typical MIMO configurations, i.e., designed for joint communication and sensing paradigms. In Sect. V applications of the proposed generative modelling approach are proposed targeting passive localization applications. Conclusions and future activities are finally discussed in Sect. VI.

II. EM BODY MODELS AND BAYESIAN PRIORS

In this section, we discuss relevant diffraction-based EM body models to reproduce the effects $E_\theta$ of body movements on the RF field, considering also body-induced RF attenuations $A_\theta$ and CSI $C_\theta$, as special cases. First, in Sect. II-A we briefly recall the body models proposed in [12] for a single link scenario using scalar diffraction theory considerations. Next, in Sect. II-B we consider a receiver equipped with an array, i.e. Uniform Linear Array (ULA), of $L = 2M + 1$ isotropic receiver antennas, and adapt the model to represent all the considered body effects, now relative to each radio link $\ell$, namely $E_\theta = [E_{\ell,\theta}]_{\ell=1}^{L}$, or equivalently $A_\theta = [A_{\ell,\theta}]_{\ell=1}^{L}$ for RSS and $C_\theta = [C_{\ell,\theta}]_{\ell=1}^{L}$ for CSI effects. A Bayesian prior formulation which is underpinned by the proposed EM body models is finally considered in Sect. II-C.

In what follows, we will always assume that the monitored target is in the Fraunhofer’s region of both transmitting (TX) and receiving (RX) antennas for all the considered links $\ell$. Extension to multi-target scenarios can be also inferred according to [13].

A. Diffraction models for body-induced attenuations

As depicted in Fig. 1, we assume that the length of the radio link is given by $d$ while $h$ is its height from the floor. The effects of floor, walls, ceiling or other obstacles are not considered. However, with some effort, these obstacles can be included, as shown in [37]. The scalar diffraction theory assumes that the 3-D shape of the human body is modeled as a 2-D rectangular absorbing sheet $S$ [12] with height $h_S$ and traversal size that changes according to a 3D cylinder view, with max. and min. traversal sizes $w_{S,1}$, $w_{S,2}$, respectively. The target has nominal position coordinates $p = [x, y]$, w.r.t. the TX position, which is defined by the projection of its barycenter on the horizontal plane that includes the Line-of-Sight (LoS). The 2-D target might be also rotated of an angle $\varphi$ with respect to the LoS direction. The body/subject “state” $\theta$ is characterized by an ensemble of body features collected into the vector $\theta := \{p, \varphi, h_S, w_{S,1}, w_{S,2}\}$.

A distribution of Huygens’ sources of elementary area $dS$ is located on the absorbing sheet $S$, so that the electric field $E_\theta$ at the receiver [12] is obtained by subtracting the contribution of the obstructed Huygens’ sources from the electric field $E_0$ of the free-space scenario (with no target in the link area):

$$E_\theta = E_0 - \int_S dE,$$

with time omitted to simplify the reasoning. According to [12], equation (2) can be rewritten in terms of the field ratio $C_\theta = \frac{E_\theta^2}{E_0^2}$:

$$C_\theta = 1 - \frac{d}{\lambda} \int_S \frac{1}{r_{12}} \exp \left\{ -\frac{2\pi}{\lambda} (r_1 + r_2 - d) \right\} d\xi_2 d\xi_3,$$

where $r_{12}$ is the distance between any two points on the absorbing sheet $S$. Integrating over the absorbing sheet $S$, we obtain

$$C_\theta = 1 - \frac{d}{\lambda} \int_S \int_S \frac{1}{r_{12}} \exp \left\{ -\frac{2\pi}{\lambda} (r_1 + r_2 - d) \right\} d\xi_2 d\xi_3.$$
where \( \lambda \) is the wavelength. Notice that each elementary source \( dS = d\xi_2 d\xi_3 \) has distance \( r_1 \) and \( r_2 \) from the TX and the RX, respectively which depends on the relative coordinates \( p \). In what follows, we refer to the field ratio \( C_\theta \) as an indicator of the body-induced channel response corresponding to the subject in state \( \theta \), herein referred to Channel State Information (CSI). Equation (3) can be adapted to take into account of non-uniform transmit and/or receiving antenna patterns.

**B. Multiple antenna configurations**

We now consider a Uniform Linear Array (ULA) with links ordered as \(-M \leq \ell \leq M\) and being RX\(\ell\) the receiver node for corresponding link \( \ell \). The central antenna is indicated by the index \( \ell = 0 \). As shown in Fig. 2, each of the \( \ell \)-th antenna RX\(\ell\) of the array is uniformly deployed at mutual distance \( \triangle \) along a segment orthogonal to the Line-of-Sight (LoS) at distance \( d \) from the TX and horizontally placed w.r.t. the floor. Ignoring mutual antenna coupling (approximately valid for \( \Delta > \lambda/4 \)), the CSI observed on the \( \ell \)-th antenna of the array corresponds to the ratio of the electric fields \( C_{\ell,\theta} = \frac{E_\ell}{E_{\ell,0}} \); therefore, using (3) it is:

\[
C_{\ell,\theta} = \frac{1 - j \frac{d_\ell}{\lambda} \int_S \frac{1}{r_1,\ell r_2,\ell} \cdot \exp \left\{ -j \frac{2\pi}{\lambda} (r_1,\ell + r_2,\ell - d_\ell) \right\} d\xi_2 d\xi_3,}
\]

where \( E_{\ell,0} \) is the EM field received by the same RX\(\ell\) node in the reference condition, i.e., the free-space scenario characterized by the absence of any target in the link area. The term \( d_\ell \) indicates the distance of the \( \ell \)-th antenna RX\(\ell\) of the array from the TX while \( d_1,\ell \) and \( d_2,\ell \) are the distances of the projection point \( O_\ell \) (of the barycenter P of the 2-D surface \( S \)) from the TX and RX\(\ell\) nodes. Likewise, \( r_1,\ell \) and \( r_2,\ell \) are the distances of the generic elementary area \( dS \) of the target \( S \) from the TX and RX\(\ell\), respectively. Notice that, for \( M = 0 \), equation (4) reduces to the single-antenna case (3), where RX\(\ell\) coincides with the RX antenna at distance \( d = d_\ell \) from the TX.

CSI data \( (C_{\ell,\theta}) \) and corresponding excess RSS attenuations \( (A_{\ell,\theta} = -10 \log_{10} |C_{\ell,\theta}|^2) \) represent the body effects on the raw RF samples for link \( \ell \). For the considered multiple antenna configuration, these are organized into the vectors:

\[
A_\theta = \begin{bmatrix} A_{\ell,\theta} = -10 \log_{10} |C_{\ell,\theta}|^2 \end{bmatrix} \quad \ell = 1,\ldots,L, \\
C_\theta = \begin{bmatrix} C_{\ell,\theta} \end{bmatrix} \quad \ell = 1,\ldots,L.
\]

In addition to the CSI terms \( C_\theta \), the EM field \( E_{\ell,\theta} \) observed on link \( \ell \) can be also re-arranged according to (4) as:

\[
E_{\ell,\theta} = \frac{d_0}{d_\ell} \exp \left\{ -j \frac{2\pi}{\lambda} (d_\ell - d_0) \right\} \int_S \frac{1}{r_1,\ell r_2,\ell} \cdot \exp \left\{ -j \frac{2\pi}{\lambda} (r_1,\ell + r_2,\ell - d_\ell) \right\} d\xi_2 d\xi_3,
\]

where \( E_{\ell,0} \) is the electric field received by the central antenna of the array of index \( \ell = 0 \). Using (6) and (4), the EM field for each considered link can be obtained as:

\[
E_\theta = \begin{bmatrix} E_{\ell,\theta} \end{bmatrix} = C_{\ell,\theta} \times \frac{d_0}{d_\ell} \exp \left\{ -j \frac{2\pi}{\lambda} (d_\ell - d) \right\} \quad \ell = 1,\ldots,L.
\]

**C. Bayesian prior modelling**

Diffraction models (5) and (7) can be used to predict the attenuations \( A_\theta \), the CSI \( C_\theta \), or directly compute the EM field \( E_\theta \) as caused by a body described by features \( \theta \). When it comes to practice, imperfect knowledge of the environment, small, involuntary, body movements, or changing configurations of the propagation environment, make these effects hard to obtain with an acceptable level of accuracy. In addition, in RF sensing applications, we are often interested to recover a subset of the body features, e.g. the subject locations \( p \) or the obstruction size \( (h_S, w_{S,1}, w_{S,2}) \), while leaving the others partially (or fully) unknown. In the following, we thus propose an informative prior modeling approach that accounts for uncertainties in body features based on Bayesian probability theory considerations. We assume that the EM body effects \( E_\theta \) are obtained for random instances of body features \( \theta \), which follow a probability function \( p(\theta|\theta_k) \). In other words, the body induced EM field \( E_\theta \) is sampled from a distribution \( p(E_\theta) \) defined as:

\[
p(E_\theta|\theta_k) = E_{\theta \sim p(\theta|\theta_k)},
\]

where the function \( p(\theta|\theta_k) \) models the uncertainty with respect to the **nominal body features** \( \theta_k \).

Some examples are proposed in the following to clarify the approach. First, consider the problem of generating body-induced RSS attenuations \( A_\theta \) for a subject located at some (nominal) position \( \theta_k = p_k \), which is typical in passive localization [4]. Involuntary movements as the result of the complex structure of the human body makes the nominal target location \( p_k \) be subject to an error in the \( \xi_1 \) and \( \xi_2 \) directions in the order of \( \Delta = 5 \pm 10 \) cm [13]. Considering that body-induced attenuations \( A_\theta \) are strongly influenced by these movements, the prior (8) should adequately take this uncertainty into account. Replacing the wavefield \( E_\theta \) with
III. EM-INFORMED GENERATIVE NEURAL NETWORKS TOOLS

The generative models considered in this section are able to reproduce body-induced EM effects \( E_\theta \) as sampled from the prior distribution \( p(E_\theta | \theta_k) \) in (8) which is conditioned on the input body features \( \theta_k \) (conditional prior). As shown in Fig. 3, the generation process is implemented by a “decoder” (VAE), or a “generator” (GAN), both parameterized by neural network (NN) parameters \( \mathbf{W}_D \) and \( \mathbf{W}_G \), respectively. These neural networks map the input latent space \( z \sim p_Z(z) \) of size \( Z \) (\( z \in \mathbb{R}^{Z \times 1} \)), into the output space \( \hat{E}_\theta \sim p_{\text{gen}}(E_\theta | \theta_k) \) where the generated distribution is set to reproduce the targeted EM model, namely \( p_{\text{gen}}(E_\theta | \theta_k) \propto p(E_\theta | \theta_k) \) for selectable inputs \( \theta_k \). As shown in Fig. 3, the NN parameters \( \mathbf{W}_D \) and \( \mathbf{W}_G \) constitute the generation models and are trained separately to reproduce body-induced attenuations \( A_\theta \), CSI \( C_\theta \) or responses \( E_\theta \), respectively.

In this paper, we limit our focus on the body feature set \( \theta_k = \{p_k, \varphi_k, h_S, w_{S,1}, w_{S,2}\} \) so to generate body effects for varying locations \( p_k \), orientations \( -\pi/2 \leq \varphi_k \leq \pi/2 \), and sizes \( h_S, w_{S,1}, w_{S,2} \) of the target. Although more complex approaches are possible, the generation of samples from the conditional prior \( p(A_\theta | \theta_k) \) is complex enough to make a full EM simulation unfeasible, thus motivating the use of GNN methods. Besides, both the decoder (VAE) and the generator

---

1 with angular resolution of \( \pi/200 \), which is sufficient for accurate reproduction of complex body trajectories.
(GAN) can be trained to reproduce body-induced attenuations conditioned on features $\theta_k$ that are unknown at training time, i.e., unvisited subject locations, orientations or target sizes. This useful property is also analyzed in the following.

Below we discuss VAE and GAN model architectures [17, 19] to reproduce EM effects $E_b$. Models are adapted by conditioning the generative process on input body features $\theta_k$ that can be chosen at run-time. The tools are referred to as conditional-VAE (C-VAE) and unbalanced conditional-GAN (UC-GAN) [40].

### A. Conditional variational autoencoder (C-VAE)

As depicted in Fig. 3(a)(c), the C-VAE model uses an encoder $Q(z|\theta_k; W_E)$, parameterized by NN parameters $W_E$, which learns the latent space $p_Z(z|\theta_k) \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for inputs $\theta_k$. Latent space is multivariate Gaussian distributed with mean $\mu_k$ and standard deviation $\sigma_k$ parameters (other choices are not investigated here). The encoder is trained using samples of (true) body-induced attenuations $E_b$ obtained from the EM model (16) and the corresponding features $\theta_k$. Model training is further discussed in Sect. IV. The decoder produces a distribution $\mathbb{E}_\theta^{\text{VAE}} \sim \mathcal{P}_{\text{gen}}^{\text{VAE}}(\theta_k)$

$$p_{\text{gen}}^{\text{VAE}}(\theta_k) = \int p_{\text{gen}}^{\text{VAE}}(\theta_k|z; W_D) p_z(z|\theta_k) dz, \quad (9)$$

which is the marginalization of the conditional probability $p_{\text{gen}}^{\text{VAE}}(\theta_k|z; W_D)$ function of the NN parameters $W_D$. The goal is to maximize the likelihood bound called Evidence Lower BOund (ELBO) $\mathcal{L}_{\text{ELBO}}$ described in [39]: omitting dependency on parameters $W_E$ and $W_D$, it is

$$\mathcal{L}_{\text{ELBO}} = \ell_k - \beta \cdot D_{\text{KL}} [Q(z|\theta_k)||p_z(z|\theta_k)]. \quad (10)$$

The first term $\ell_k = \mathbb{E}_{z \sim Q(\cdot)} \log p_{\text{gen}}^{\text{VAE}}(\theta_k|z)$ is the log-likelihood function, while the second one is the Kullback-Leibler (KL) divergence $D_{\text{KL}}$ [41] between the encoder output and the input latent space. The ELBO metric (10) is then averaged over the input training features $\theta_k$, as described in Sect. IV.

The maximization of the likelihood $\ell_k$ makes the generated samples $\mathbb{E}_\theta^{\text{VAE}}$ more correlated to the latent variables $z$, which typically cause the model to be more deterministic. On the other hand, the number of latent variables $Z$ as well as the ELBO weight term $\beta > 0$ can be tuned to increase the contribution of the KL divergence between the posterior and the prior to the total ELBO and thus increase the randomness of generated samples. Targeting passive localization applications, in Sect. IV, we will show that both terms $(Z, \beta)$ can be optimized to improve the generation process, also to account for measurements $S_t$ affected by noise and multipath interference.

### B. Unbalanced Conditional GAN (UC-GAN)

As depicted in Fig. 3(b)(d) GAN training is formulated as a min-max problem that can be interpreted as an adversarial game with two players: the “discriminator” $D(\mathbb{E}_\theta, \mathbb{E}_G; W_D) \in [0, 1]$, namely a binary classifier which tries to improve the detection of “fake” body-induced attenuation samples, and the generator which is designed to fool the discriminator. The generator produces $\mathbb{E}_G^{\text{GAN}} \sim \mathcal{P}_{\text{gen}}^{\text{GAN}}(\theta_k; W_G)$ with

$$p_{\text{gen}}^{\text{GAN}}(\theta_k) = \int p_G(z)p_{\text{gen}}^{\text{GAN}}(\theta_k|z; W_G)dz \quad (11)$$

now driven by $p_G(z)$ $\sim \mathcal{N}(0, I)$. The goal is now to minimize the statistical distance (Jensen-Shannon divergence, based on [41]) between $p_G(\theta_k)$ and $p_{\text{gen}}(\theta_k)$: this corresponds to maximizing the discriminator loss while minimizing the generator one. More details can be found in [19]. For the considered problem, we adopted an unbalanced implementation [40] which pre-trains the generator using the parameters $W_D$ of the C-VAE decoder (9). This prevents the faster convergence of the discriminator at early epochs which could cause the generator to not converge.

### C. Model training and implementation considerations

C-VAE and UC-GAN pre-trained models shown in Fig. 3 are available on-line [36] together with example codes for training on new samples, and testing, namely generating body-induced attenuations ($\mathbb{A}_\theta$) and CSI ($\mathbb{C}_\theta$), according to specific body configurations.

Considering C-VAE in Fig. 3(a)(c), the encoder model $W_E$ includes two key components: a sequence of convolutional
layers and a feed-forward network. The encoder takes as input training samples obtained from the diffraction model as well as the conditional inputs $\theta$, which describe the corresponding body features. Attenuations $A_\theta$ and EM field generation, such as CSI $C_\theta$, require a different number of convolutional layer subcomponents, which reflect the dimension of the data: 2 layers are chosen for reproducing attenuations $A_\theta$, while 3 layers are used to generate EM field samples, namely $C_\theta$ or $E_\theta$. This choice is conservative since it is critical to limit the size of trainable model parameters, while no apparent performance improvement is observable beyond this limit. The decoder $W_D$ reproduces samples of body effects as a function of customized inputs $\theta$ that are one-hot encoded before being used as input to the neural network. It shows a similar structure as the encoder, while we used transposed convolution layers, also referred to as fractionally-strided convolutions, to increase (upsample) the spatial dimensions of intermediate features, so that generated outputs respects the desired dimensions.

Considering now the UC-GAN model in Fig. 3(b),(d), the discriminator ($W_I$) and the generation ($W_C$) model structures include similar components. Following the unbalanced GAN implementation, the C-VAE decoder model parameters $W_D$ are transferred to discriminator $W_I$ at the beginning of the training stage. To simplify comparison, the outputs of both the models have the same dimension: for an assigned input $\theta_k = [p_k,h_S,w_{S,1},w_{S,2}]$, the GNN generates 201 different subject orientations in the interval $-\pi/2 \leq \varphi_k \leq \pi/2$, for all the configured physical links $L$.

Generation times of C-VAE, UC-GAN and EM diffraction models are compared in Table I considering single antenna TX and RX and a MIMO setup with $L = 81$ links. Reproduction of attenuations $A_\theta$ and EM field samples $E_\theta$ are analyzed separately. For each case, time measurements are obtained using a Jetson Nano single-board computer equipped with a quad-core ARM-Cortex-A57 SoC, 4 GB RAM, 128 CUDA cores, and a Maxwell GPU architecture which is representative of a typical resource-constrained wireless device. Note that, on average, the VAE/GAN-based generation of the EM body effects is about $\times 60 \div 100$ times faster than EM model computation, which also depends on the chosen numerical integration configurations, i.e., tiled integration method, and absolute error tolerance, target size, and antenna configuration (omnidirectional vs directional antenna radiation patterns). The generative model can be therefore used to reproduce the desired prior distribution in real-time, with sufficiently high randomness of samples, namely at least $50 \div 100$ samples per second, which is reasonable considering typical body movement speeds (max. 1m/s).

For the implemented optimizations, Table II analyzes the size of the trainable parameters, namely the model footprint, of the decoder (C-VAE) and the generator (UC-GAN). Footprints range from 1MB to 240MB, being the EM field generation $E_\theta$ more demanding in terms of memory occupation than attenuation $A_\theta$ generation. Although out of the scope of the current paper, accurate model pruning is desirable to minimize the memory footprint on resource-constrained devices [4]. In what follows we assess the ability of C-VAE and UC-GAN approaches to reproduce the EM model diffraction effects and, more generally, the effectiveness of the models in sampling from the conditional prior $p(C_\theta|\theta)$, considering the problem of reproducing body induced attenuations ($A_\theta$) and CSI ($C_\theta$), separately, for varying input features and scenarios.

IV. GENERATION OF EM BODY MODEL SAMPLES

EM simulations of body effects are obtained by using the scalar diffraction theory (Sect. II) and assuming a carrier frequency set to $f_c = 2.4$ GHz. TX and RX nodes are equipped with a uniform linear array (UL) of 9 omni-directional antennas (corresponding to $L = 81$ links) and spaced at $\Delta = \lambda/2$. The length of the central link of the array is equal to $d = 4m$ while all the links of the array are horizontally placed at height $h = 0.99m$ from the ground. The human target has also variable height $h_S$, traversal max. and min. sizes of $w_{S,1}$ and $w_{S,2}$, respectively. In these tests the target stands vertically on the floor, that is used only to support the target and does not have any EM influence on the radio links. Generative models C-VAE and UC-GAN are trained using samples of EM body diffraction as the result of different body configurations $\theta_k = [p_k,\varphi_k,h_S,w_{S,1},w_{S,2}]$, namely:

1) 75 marked locations $p_k$, $k = 1,...,K = 75$, inside the Fresnel’s area of the considered link, so that the spacing between marked positions is 0.25m along and across the LOS (for $d = 4m$);

2) 4 subject rotations $\varphi_k = [-\pi/2,-\pi/3,-\pi/6,0]$;

3) 21 different dimensions of the targets ranging from $h_S = 1.2m$ to $h_S = 2.2m$ and $w_{S,1} = 0.25m$ to $w_{S,1} = 0.55m$: note that in the reported tests semi-size dimension $w_{S,2}$ is fixed as $w_{S,2} = 0.25m$.

The generative models can be used to reproduce body-induced EM diffraction effects for any position of the target inside the Fresnel’s area of the link. Target dimensions are however subject to the limitations detailed above, which are reasonable as far as the goal is to represent a human body [13].

C-VAE method requires parameter optimization, namely optimization of the number of latent variables $Z$ and ELBO weights $\beta$, which is the goal of the first part of the analysis in Sect. IV-A. Next, in Sect. IV-B a comparison is made with UC-GAN model targeting the generation of EM attenuations samples for a MIMO configuration. GAN based models were proposed in several works [31], [33], although not optimized for the specific problem. In Sect. IV-C we consider the problem of array processing using generated EM field samples, namely the CSI $C_\theta$.

A. C-VAE latent variable optimization

Fig. 4 shows an example of C-VAE generation of diffraction model samples, namely body-induced attenuations, using varying latent variables ranging from $Z = 8$ to $Z = 32$.
Fig. 4. C-VAE generation of body-induced excess attenuation values $A_{E,\theta}$ for different movements of the target (along, across the LOS and varying orientations), and dimensions $(h_S, w_{S,1}, w_{S,2})$, with $w_{S,2} = 0.25m$. C-VAE results are shown for varying latent samples $Z$ and $\beta = 0.05$. From left to right: (a) the subject is moving along the LOS ($0.25m \leq x \leq 3.75m, y = 0$). The generated EM body excess attenuation values obtained via numerical methods are represented in dashed lines by averaging over random target orientations $-\pi/2 \leq \varphi \leq \pi/2$ and random movements in an elementary area of size $\Delta x = \Delta y = 0.1m$. (b) The subject is moving across the LOS ($x = 1m, -0.5m \leq y \leq 0.5m$). Dashed lines shows the corresponding ground-truth diffraction model samples obtained similarly as in (a). (c) The target is in position $x = 0.75m, y = 0$ and changing orientation $-\pi/2 \leq \varphi \leq 0$ while performing small movements in the same elementary area. Dashed lines shows the EM body model attenuations obtained for a subset of the subject orientations.

Fig. 5. C-VAE generated sample probabilities $p_{VAE}(A_{\theta}|\theta_k)$ of body-induced RF attenuations for varying latent dimensions ($Z = 8, 16, 32$) and ELBO weights $\beta$, compared with samples obtained from EM body model (dashed lines). (a),(b) target standing at $x = 0.5m$ from the TX with size $h_S = 1.4m, w_{S,1} = 0.5m$. Generation with $\beta = 0.05$ (a) and $\beta = 1e^{-0.09}$ (b). (c),(d) target standing at $x = 2m$ from the TX with size $h_S = 1.4m, w_{S,1} = 0.35m$. Generation with $\beta = 0.05$ (c) and $\beta = 1e^{-0.09}$ (d). (e),(f) target with size $h_S = 1.65m, w_{S,1} = 0.65m$ standing at $x = 0.5m$ (e) and $x = 2m$ (f) from the TX. (g),(h) target with size $h_S = 2.0m, w_{S,1} = 0.65m$ standing at $x = 0.5m$ (g) and $x = 2m$ (h) from the TX. (e),(f),(g),(h): generation with $\beta = 0.05$. 

and ELBO weight $\beta = 0.05$. The subject is moving along and across the (single) radio link in specific marked locations $p_k$, as well as changing its orientation $0 \leq \varphi \leq -\pi/2$. TX and RX are equipped with a single antenna ($L = 1$). Here, we are interested in generating body attenuations $A_{\theta}^{\beta} = \mathbb{E}^\beta[f_{\theta}^A]$. To account for the uncertainties introduced by different body postures and small, i.e., involuntary, movements in the assigned location $p_k$, we report the average attenuations w.r.t. 50 generated samples from $\mathbf{A}_{\theta}^{\beta} \sim p_{\text{gen}}^\beta(A_{\theta} | \theta)$. In Fig. 4(a), the C-VAE model is used to reproduce the average attenuation samples corresponding to a subject that is moving along the LOS ($0.25m \leq x \leq 3.75m$, $y = 0$) with a step of 0.25m, namely occupying 15 marked locations, from $p_1 = [0.25m, 0]$ to $p_{15} = [3.75m, 0]$. The target has different dimensions, namely $h_S = 1.4m$, $w_{S,1} = 0.35m$ (black), $h_S = 2.0m$, $w_{S,1} = 0.65m$ (red), $h_S = 1.65m$, $w_{S,1} = 0.65m$ (green) and $w_{S,2} = 0.25m$. Also, it is changing its orientation while standing in each marked location. Fig. 4(b) shows the corresponding generated samples now featuring a subject moving across the LOS ($-0.5m \leq y \leq 0.5m$, $x = 1m$) and with same dimensions. Finally, in Fig. 4(c), the target is now fixed in position $p = [0.5m, 0]$ but uniformly changing its orientation $\varphi$ from $\varphi = -\pi/2$ to $\varphi = 0$. Generated samples are compared with the average EM body attenuations $A_{\theta}^{\beta}$ obtained from (4) and (5) via numerical methods for the same link and corresponding positions (dashed lines). The attenuations are averaged over 50 random target movements in an elementary 2D area of size $\Delta = 0.1m$ surrounding the corresponding marked positions $p_k$. Considering all the tests, we found that using $Z = 16$ latent variables constitutes a good compromise between complexity and accuracy.

In addition to average attenuation terms, Fig. 5 analyzes the distribution of the generated attenuation samples $p_{\text{gen}}^\beta(A_{\theta} | \theta)$ compared with those obtained from the EM diffusion model, for varying latent variables $Z$ and two choices of ELBO weight $\beta$, namely $\beta = 0.05$, in Fig. 5(a),(c), and $\beta = 1e^{-09}$ in Fig. 5(b),(d). Three target configurations are considered, namely $h_S = 1.4m$, $w_{S,1} = 0.35m$ in Fig. 5(a),(b),(c),(d), $h_S = 1.65m$, $w_{S,1} = 0.65m$ in Fig. 5(e),(f), and $h_S = 2.0m$, $w_{S,1} = 0.65m$ in Fig. 5(g),(h). The generated distributions $p_{\text{gen}}^\beta$ reproduce the attenuations observed with a target standing at distance 0.5m, in Fig. 5(a),(b),(c),(g), and 2m, in Fig. 5(c),(d),(f),(h), from the TX, but changing its posture in the same 2D elementary area of size $\Delta = 0.1m$ previously considered. As evident from the corresponding cases, the number of latent variables $Z$ substantially affects the generated samples, while the ELBO weight $\beta$ seems to have less evident effects. The C-VAE tool configured for $Z = 16$ (and $Z = 32$) provides a good representation of the distribution of the attenuations when compared with the EM diffusion model one. On the other hand, the C-VAE model seems to better reproduce the attenuation samples corresponding to targets placed at some distance, i.e., 2m from the TX, rather than close to TX, i.e., 0.5m. The trend is particularly evident when the C-VAE model is set to reproduce sample distribution with high variance and few training samples, as in the case for small target size ($h_S = 1.4m$, $w_{S,1} = 0.35m$). In the next section we consider the impact of such effects on passive RF sensing performance when using generated samples as prior information. Finally, the choice for $\beta = 0.05$ stands as a good compromise between the average reproduction accuracy and the reconstruction of the entire probability function which require to increase the randomness of generated samples$^3$ [18]

### B. UC-GAN and C-VAE analysis in MIMO setups

In Fig. 6 we compare the C-VAE generation tool using the optimized parameters $Z = 16$, $\beta = 0.05$ shown previously with the UC-GAN implementation described in Sect. III-B. The following analysis is of interest as it shows the behavior of two different generative systems, and compares their ability to reproduce the EM body diffraction effects. With respect to previous section, we now consider a MIMO UL array consisting of 3 antennas at the TX and RX, respectively, $L = 9$ radio links and distance $d = 4m$. The samples $\mathbf{A}_{\theta}^{\beta} = \mathbb{E}^\beta[f_{\theta}^A]_{L=1}^{L=0}$ obtained through C-VAE (solid lines) and with UC-GAN $\mathbf{A}_{\theta}^{\beta} = \mathbb{E}^\beta[f_{\theta}^A]_{L=1}^{L=0}$ (diamond markers) tools are compared with the corresponding EM body-induced attenuations $A_{\theta} = [A_{\theta}]_{L=0}$ obtained from diffraction theory (dashed lines). The subject has fixed dimensions $h_S = 2m$, $w_{S,1} = 0.55m$, $w_{S,2} = 0.25m$, and it is moving along the LOS path of the link $\ell = 5$ ($0.25m \leq x \leq 3.75m$, $y = 0$).

Table III reports a comparative analysis of C-VAE and UC-GAN generation for the same MIMO setup, in terms of Mean Squared Error (MSE) and KL divergence $D_{KL}$.

<table>
<thead>
<tr>
<th>Target $(\theta_S, x, y)$ [m]</th>
<th>MSE [dB]</th>
<th>$D_{KL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_S = 1.4 \pm 1.6$, $x \leq 0.5$, $y = 0$</td>
<td>0.28</td>
<td>0.87</td>
</tr>
<tr>
<td>$h_S = 1.4 \pm 1.6$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>0.19</td>
<td>0.46</td>
</tr>
<tr>
<td>$h_S = 1.6 \pm 1.8$, $x \leq 0.5$, $y = 0$</td>
<td>0.44</td>
<td>0.37</td>
</tr>
<tr>
<td>$h_S = 1.6 \pm 1.8$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>0.37</td>
<td>0.41</td>
</tr>
<tr>
<td>$h_S = 1.8 \pm 2$, $x \leq 0.5$, $y = 0$</td>
<td>0.14</td>
<td>0.38</td>
</tr>
<tr>
<td>$h_S = 1.8 \pm 2$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>0.31</td>
<td>0.37</td>
</tr>
<tr>
<td>$h_S = 1.4 \pm 1.6$, $x \leq 0.5$, $y = 0$</td>
<td>1.74</td>
<td>&gt;3</td>
</tr>
<tr>
<td>$h_S = 1.4 \pm 1.6$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>1.07</td>
<td>&gt;3</td>
</tr>
<tr>
<td>$h_S = 1.6 \pm 1.8$, $x \leq 0.5$, $y = 0$</td>
<td>1.43</td>
<td>&gt;3</td>
</tr>
<tr>
<td>$h_S = 1.6 \pm 1.8$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>0.74</td>
<td>&gt;3</td>
</tr>
<tr>
<td>$h_S = 1.8 \pm 2$, $x \leq 0.5$, $y = 0$</td>
<td>1.12</td>
<td>2.21</td>
</tr>
<tr>
<td>$h_S = 1.8 \pm 2$, $x = 0.5 \pm 2$, $y = 0$</td>
<td>0.41</td>
<td>2.3</td>
</tr>
</tbody>
</table>

$^3$which is equivalent to minimize the divergence between the reproduced distributions and the corresponding true ones.
Additive White Gaussian Noise (AWGN) complex vector $n$ where $H$ indicates conjugate transpose operation. The subject has dimensions $(h_B = 2m, w_{S,1} = 0.55m, w_{S,2} = 0.25m)$ and is moving along the LOS path of link $\ell = 5$ ($0.25m \leq x \leq 3.75m, y = 0$). The EM body average excess attenuation values $\Delta_{\text{EM}}$ are compared with the corresponding diffraction model samples (dashed lines). C-V AE latent variable dimension is $Z = 16$ with $\beta = 0.05$, as optimized as in Fig. 4. UC-GAN is pre-trained using C-V AE decoder model.

Neglecting AWGN noise and considering the CSI defined in (4), the array response $R_\theta(\gamma)$, or the array factor, as due to a target in state $\theta$ is defined as [38]

$$R_\theta(\gamma|C_\theta) = \sum_{m=-M}^{M} w_m^* E_{\ell,0} E_{\ell,\theta} = w(\gamma)^H \cdot C_{\theta}^{-1}. \quad (13)$$

In the following analysis, we verify the ability of the proposed C-VAE model to reproduce the response $R_\theta(\gamma|C_{\text{VAE}}) \sim R_\theta(\gamma|C_\theta)$ in real-time through the generation of EM field samples $C_{\theta}^\text{VAE}$. For conventional ULA scenarios, that assume planar wavefront propagation, the steering vector $w(\gamma) = [a_m]$ for the considered array is given by [42]:

$$w = \left\{ \exp \left\{ \frac{2\pi}{\lambda} \frac{\Delta \cos \gamma}{\lambda} \right\} \right\}_{k=-M,\ldots,0,\ldots,+M}.$$ \quad (14)

where $\gamma$ is the DoA (Direction Of Arrival), namely the direction of propagation of the impinging wavefront w.r.t. the axis of the array, and $\Delta = \lambda/2$ is the inter-element antenna distance. According to (14), it is also $|w|^2 = w^H w = (2M+1)$.

Fig. 7 considers an RX-side UL array layout of $L = 9$ antennas (see Fig. 7(a)) and shows the body-induced array response $20 \log_{10} |R_\theta(\gamma)|$ as a function of the DoA $\gamma$ and for different values of the $y$ displacement of the target (w.r.t. the central LOS) and $x = 2m$. The array signal processing is set to extract the response for varying DoA $\gamma$ and is based on Fast Fourier Transform (FFT) with $N_{FFT} = 257$ points. Fig. 7(b)\footnote{In line with the setup described in (1), the generative model is now designed to reproduce the prior effects of body movements on the response of the array; therefore, it appears reasonable to neglect the effect of measurement and AWGN noise, as well as fading.}

C. Generated EM field samples for array processing

We now verify the capability of the proposed C-VAE model to generate samples of the EM field $\mathbf{\mathcal{E}}_\theta^{\text{VAE}} = \left[ \mathcal{E}_{\ell,0} \right]_{\ell=1}^L$; the goal is to reproduce the array response of conventional linear beamforming processing [42] as well as the effects of target movements on such response [38]. We consider $L = 2M + 1$ antennas $(M = 4)$ and the vector $w(\gamma) = [w_k]_{k=-M,\ldots,0,\ldots,+M} = [w_{-M}, \ldots, w_0, w_1, \ldots, w_M]^T$ of linear beamforming coefficients designed to steer the antenna array in a direction $\gamma$. The received baseband signal $r_\theta(\gamma)$ at the output of the beamforming processing is given in by:

$$r_\theta(\gamma) = \sum_{m=-M}^{M} w_m^* [E_{\ell,0} + n_\ell] = w(\gamma)^H \cdot [E_\theta + n], \quad \text{(12)}$$

where $H$ indicates conjugate transpose operation, $n_\ell$ is the Additive White Gaussian Noise (AWGN) complex vector $n = [n_{-M}, \ldots, n_0, n_1, \ldots, n_M]^T$ of size $2M + 1$, that is assumed to be spatially white with zero mean and covariance $\sigma^2 I$. 

---

Fig. 6. C-VAE vs UC-GAN generation of EM body model for a MIMO array consisting of 3 antennas at the TX and RX, respectively, and $L = 9$ radio links. The subject has dimensions $(h_B = 2m, w_{S,1} = 0.55m, w_{S,2} = 0.25m)$ and is moving along the LOS path of link $\ell = 5$ ($0.25m \leq x \leq 3.75m, y = 0$). The EM body average excess attenuation values $\Delta_{\text{EM}}$ are compared with the corresponding diffraction model samples (dashed lines). C-V AE latent variable dimension is $Z = 16$ with $\beta = 0.05$, as optimized as in Fig. 4. UC-GAN is pre-trained using C-V AE decoder model.

---

Fig. 7. C-V AE vs UC-GAN generation of EM body model for a MIMO array consisting of 3 antennas at the TX and RX, respectively, and $L = 9$ radio links. The subject has dimensions $(h_B = 2m, w_{S,1} = 0.55m, w_{S,2} = 0.25m)$ and is moving along the LOS path of link $\ell = 5$ ($0.25m \leq x \leq 3.75m, y = 0$). The EM body average excess attenuation values $\Delta_{\text{EM}}$ are compared with the corresponding diffraction model samples (dashed lines). C-V AE latent variable dimension is $Z = 16$ with $\beta = 0.05$, as optimized as in Fig. 4. UC-GAN is pre-trained using C-V AE decoder model.
Fig. 7. (a) Array layout setup; (b) Array response $R_\theta(\gamma)$ for $\gamma = 0$ and a target with dimensions $h_S = 1.64m$, $w_{S,1} = 0.55m$, $w_{S,2} = 0.25m$ and 2 positions, namely $x = 2m$, $y = -0.25m$ and $x = 2m$, $y = 0.25m$. Response is obtained using C-VAE EM field generated samples $C_\theta^{\text{VAE}} (Z = 16,$ $\beta = 0.05)$ - green/red solid lines - and compared with the array response obtained with EM Diffraction field samples (dashed lines); (c) Direction of Arrival (DoA) analysis for a subject moving across the LOS ($-0.25m \leq y \leq 0.25m$, $x = 2m$), and changing orientations randomly in the interval $-\pi/2 \leq \varphi \leq \pi/2$. Shows two responses (red and green lines) for corresponding target locations $y = -0.25m$ and $y = 0.25m$, respectively. The theoretical responses $R_\theta(\gamma|C_\theta)$ using EM diffraction are in dashed lines while solid lines represent the reproduced responses $R_\theta(\gamma|C_\theta^{\text{VAE}})$ using C-VAE generated CSI samples $C_\theta^{\text{VAE}}$. Fig. 7(c) compares the dominant DoA $\gamma_{\text{max}}$, namely the maximum response of the array. This is obtained from the EM body model $\gamma_{\text{max}} = \arg \max_\gamma R_\theta(\gamma|C_\theta)$ (blue dots) and C-VAE $\gamma_{\text{max}} = \arg \max_\gamma R_\theta(\gamma|C_\theta^{\text{VAE}})$ (red dots) corresponding to a target moving across the LOS ($-0.25m \leq y \leq 0.25m$, $x = 2m$), with speed 0.5m/s and changing orientation randomly in the interval $0 \leq \varphi \leq \pi/2$. It can be noticed that in both cases the maximum response $\gamma_{\text{max}}$ is perturbed by the presence of the subject and such alteration is well reproduced by the C-VAE model.

V. CASE STUDIES IN PASSIVE RADIO LOCALIZATION

In this section the goal is to verify the capability of the generative model to reproduce and predict the measured RF samples for an assigned propagation scenario. The EM-informed generative tool has been thus validated with measurements taken in a hall of size $6.1m \times 14.4m$ as shown in Fig. 8. Both TX and RX are equipped with directional antennas with parameters summarized in the Table of Fig. 8. A linear guide system is used to move the RX antenna and collect RF measurements on multiple links $\ell$. The target is located in $K = 75$ marked positions $p_k$, $k = 1, ..., K$, which belong to a regular 2D grid as shown in the same figure. The received power $S_t$, namely the RSS, is measured using a tracking generator enabled spectrum analyzer [44] in the $2.4 \div 2.5$GHz band, over 81 frequencies with $1.25$MHz spacing; for each frequency and target position under test, 500 consecutive time samples are acquired in 1 min. (120ms sampling time). The human target (one of the authors who volunteered) has height $h_S = 1.80m$ and traversal max. and min. sizes that can be approximated as $w_{S,1} = 0.55m$ and $w_{S,2} = 0.25m$, respectively.

<table>
<thead>
<tr>
<th>Start/stop frequency</th>
<th>Power</th>
<th>Frequency spacing</th>
<th>Antenna HPBW</th>
<th>Antenna gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4GHz - 2.5GHz</td>
<td>0 dBm</td>
<td>1.25MHz</td>
<td>H (60 deg.), V (76 deg.)</td>
<td>9 dBi</td>
</tr>
</tbody>
</table>
A. Bayesian tool for RF sample generation

So far we considered the deployment of a generative model to reproduce body-induced effects such as excess attenuations and CSI. In what follows, we propose a tool that exploits the C-V AE model now to synthesize RSS samples and CSI. In what follows, we propose a tool that

\[ p_{\text{gen}}(S|t,\theta_k) \]

where \( \theta_k \) is assumed to be known or measured during calibration: \( \ell, t \) as induced by body movements. This is illustrated in Fig. 9. C-VAE generation of RSS samples \( p_{\text{gen}}(S_{\ell, t}|\theta_k) \) obtained by marginalization (15) with target size \( h_s = 2m, w_s, l = 0.35m, \) and orientation \( \varphi \) uniformly distributed in \( 0 \leq \varphi \leq \pi/4 \), to model small involuntary movements in the surrounding of the nominal position. Generated samples are compared with histograms \( p_{\text{RSS}}(S) \) obtained from true RSS measurements at 2.4GHz. Likelihood \( p(S_{\ell, t}|A_{\ell, 0}) \) are \( \mu = 1.5dB, \sigma_T = 1dB \) and \( \sigma_0 = 1dB \). The subject is standing while performing small movements around 4 nominal positions \( p_k = (x, y) \), namely \( x = 0.25m, y = 0 \) (blue color), \( x = 0.5m, y = 0 \) (violet color), \( x = 0.75m, y = 0 \) (yellow color), and outside Fresnel area \( x = 2m, y = -0.5 \) (black color).

\[ p_{\text{gen}}(S_{\ell, t}|\theta_k) = \int_{A_{\ell, 0}} p(S_{\ell, t}|A_{\ell, 0}) p_{\text{VAE}}(A_{\ell, 0}|\theta_k) dA_{\ell, 0}, \]  

with \( p_{\text{VAE}}(A_{\ell, 0}|\theta_k) \) in (9) and \( p(S_{\ell, t}|A_{\ell, 0}) \) being the likelihood function. Notice that the term \( \int_{A_{\ell, 0}} \) is replaced by numerical integration. A log-normal model is used to represent the likelihood function \( p(S_{\ell, t}|A_{\ell, 0}) \) which is defined at the generic link \( \ell, p(S_{\ell, t}|A_{\ell, 0}), \) frequency \( f \) and time \( t \) as

\[ S_{\ell, t} = \begin{cases} P_{\ell, 0} + w_{\ell, 0} & \text{free-space only} \\ P_{\ell, 0} - A_{\ell, 0, \ell} + w_{\ell, T} & \text{with target } S \end{cases} \]

where \( A_{\ell, 0} \) is the excess attenuation due to the presence of \( S \) in the first Fresnel zone of the considered link \( \ell \). 

![Fig. 9. C-VAE generation of RSS samples (\( p_{\text{gen}}(S_{\ell, t}|\theta_k) \)) obtained by marginalization (15) with target size \( h_s = 2m, w_s, l = 0.35m, \) and orientation \( \varphi \) uniformly distributed in \( 0 \leq \varphi \leq \pi/4 \), to model small involuntary movements in the surrounding of the nominal position. Generated samples are compared with histograms \( p_{\text{RSS}}(S) \) obtained from true RSS measurements at 2.4GHz. Likelihood \( p(S_{\ell, t}|A_{\ell, 0}) \) are \( \mu = 1.5dB, \sigma_T = 1dB \) and \( \sigma_0 = 1dB \). The subject is standing while performing small movements around 4 nominal positions \( p_k = (x, y) \), namely \( x = 0.25m, y = 0 \) (blue color), \( x = 0.5m, y = 0 \) (violet color), \( x = 0.75m, y = 0 \) (yellow color), and outside Fresnel area \( x = 2m, y = -0.5 \) (black color).](image)

\[ \text{Target } P_{\ell, 0} \text{ and } w_{\ell, 0} \text{ and } w_{\ell, T} \text{ model the log-normal multipath fading and the other disturbances. Noise terms are Gaussian distributed where } w_{\ell, 0} \sim N(0, \sigma_0^2), \text{ with variance } \sigma_0^2, \text{ refers to the free-space case only and } w_{\ell, T} \sim N(\Delta h_T, \sigma_T^2), \text{ with mean } \Delta h_T \text{ and variance } \sigma_T^2 = \sigma_0^2 + \Delta h_0^2, \text{ to the case with the target. } \Delta h_T \text{ and } \Delta \sigma_T^2 \geq 0 \text{ are the residual stochastic fading terms that depend on the specific scenario as shown in [12]. In the following we choose } \Delta h_T = 2dB, \Delta \sigma_T = 1dB, \sigma_T = 2dB \text{ and } \sigma_0 = 1dB \text{ [13].} \]

The Fig. 9 compares the sample probability, namely the probability mass function, \( p_{\text{gen}}(S_{\ell, t}|\theta_k) \) obtained by C-VAE generated RSS samples with corresponding histograms \( p_{\text{RSS}}(S_{\ell, t}) \) obtained from true RSS measurements at 2.4GHz.

Table IV: Average \( \varepsilon_f \) and standard deviation \( \delta_f \) error analysis between true RSS and predicted via C-VAE generative model (setup in Fig. 8).

| Target \( P_{\ell, 0} \) and \( w_{\ell, 0} \) and \( w_{\ell, T} \) model the log-normal multipath fading and the other disturbances. Noise terms are Gaussian distributed where \( w_{\ell, 0} \sim N(0, \sigma_0^2) \), with variance \( \sigma_0^2 \) refers to the free-space case only and \( w_{\ell, T} \sim N(\Delta h_T, \sigma_T^2) \), with mean \( \Delta h_T \) and variance \( \sigma_T^2 = \sigma_0^2 + \Delta h_0^2 \), to the case with the target. \Delta h_T \text{ and } \Delta \sigma_T^2 \geq 0 \text{ are the residual stochastic fading terms that depend on the specific scenario as shown in [12]. In the following we choose } \Delta h_T = 2dB, \Delta \sigma_T = 1dB, \sigma_T = 2dB \text{ and } \sigma_0 = 1dB \text{ [13].} \]

The Fig. 9 compares the sample probability, namely the probability mass function, \( p_{\text{gen}}(S_{\ell, t}|\theta_k) \) obtained by C-VAE generated RSS samples with corresponding histograms \( p_{\text{RSS}}(S_{\ell, t}) \) obtained from true RSS measurements at 2.4GHz.

5the power \( P_{\ell, 0} \) is assumed to be known or measured during calibration: the C-VAE model is used to reproduce the alterations of the path loss measure \( P_{\ell, 0} \) as induced by body movements.
The estimated position of the target relative to RX, is finally
determined in real-time, as in Fig. 8. The generated samples are again compared with RSS measurements where the target is set to move along the LOS path with a constant speed approximated as 0.25m/s. The average error $\varepsilon (p_k) = E_t [S_t \mid \theta_k] - E_t [\hat{S}_t \mid p_k]$ between the RSS values $S_t (p_k)$ predicted by the C-VAE model and the corresponding measurements $\hat{S}_t (p_k)$ are summarized in Table IV for varying target locations $p_k$ along the LOS path. The corresponding error $\delta (p_k) = \sigma (p_k) - \hat{\sigma} (p_k)$ between the true $\sigma (p_k) = \sqrt{E_t^2 [S_t \mid \theta_k] - E_t^2 [\hat{S}_t \mid p_k]}$ and predicted $\hat{\sigma} = \sqrt{E_t^2 [\hat{S}_t \mid p_k] - E_t^2 [\hat{S}_t \mid \theta_k]}$ standard deviation values are shown as well. The generative model can be effectively used to predict the true RSS values for both the considered links with average error of 2.7dB and standard deviation error of 2.9dB. On the other hand, as also observed in Sect. IV-A, the C-VAE generation tool seems to under-estimate ($\varepsilon (p_k) < 0$) the observed RSS values for target positions close to the transmitter or receiver, i.e., $x = 3m$, since it is trained using diffraction models [12].

B. Passive localization using RF generated samples

In the following, we discuss an example of passive localization where the goal is to detect the distance $d_R$ of the target from the multi-antenna RX device in real-time, as in typical radar applications. Given the RSS observations $S_t$ over the same links considered in Fig. 8, we want to recover the estimated target state $\theta_k$, in our case the position $p_k$ of the target relative to RX. Using the C-VAE generated attenuations as samples of the conditional prior distribution. We first obtain the most likely body effects $A_t (\theta_k) = \arg \max_\theta \log[p(A_t) \mid S_t, \theta_k]$ for each possible target state $\theta_k$, and the corresponding probabilities $p_A (\theta_k) = \Pr[A_t = A_t (\theta_k)]$. Using (1) and neglecting antenna correlation effects, namely $p(A_t \mid S_t, \theta_k) = \prod_\ell p(A_{\ell,t} \mid S_{\ell,t}, \theta_k)$, the probabilities $p_A$ are defined as

$$p_A (\theta_k) = \max_\theta \sum_\ell \log \left[ p(A_{\ell,t} \mid S_{\ell,t}) \cdot p_{\text{gen}} (A_{\ell,t} \mid \theta_k) \right].$$

The estimated position of the target relative to RX, is finally obtained by MAP estimation: $\hat{p}_k = \arg \max_\theta p_A (\theta_k)$, from which it is possible to make a decision about target distance $d_R = \lVert \hat{p}_k \rVert$. The proposed use case is critical in industrial scenarios where human workers operate in areas featuring increasing level of safety or privacy. Enforcing safety/privacy constraints requires the real-time monitoring and tracking of the human subject.

Using RSS samples collected from measurements, Tab. V analyzes the precision and recall probabilities:

$$p_1 (d_R) = \Pr[p_k \in L (d_R) \mid \hat{p}_k \in L (d_R)]$$
$$p_2 (d_R) = \Pr[p_k \in L (d_R) \mid \hat{p}_k \in L (d_R)],$$

with $L (d_R)$ being the region that contains the positions $p_k$ of the target at distance of $d_R$ from the RX. The recall metric measures how often the algorithm correctly identifies the target distance from all the true positive counts, while the precision indicates how often the algorithm is correct when predicting the target distance. In particular, the table analyzes precision and recall for varying distance $d_R$ from the RX and compares three cases:

1) “full calibration” scenario assumes the prior probability $p_{\text{gen}} (A_t \mid \theta_k)$ being known as obtained from measurements collected in the same environment during a calibration stage;
2) “C-VAE prior” adopts the probability $p_{\text{gen}} (A_t \mid \theta_k)$ as prior model with samples obtained using the C-VAE generator tool;
3) “Uniform prior” represents a case where no information on excess attenuation is available: the prior is replaced with a uniform probability function $U [-5dB, 15dB]$ with attenuations ranging from $-5$dB to $15$dB.

Note that scenario (1) gives the best performance, as expected; on the other hand, it requires time-consuming data collection and calibration stages which might be often unfeasible. Case (3) corresponds to the worst case scenario since no prior information on body-induced attenuations are available. Finally, case (2) uses the C-VAE tool to real-time generate samples from the prior $p_{\text{gen}} (A_t \mid \theta_k)$ and probability (17); calibration is thus not needed. From the results in the Table, the performance of “C-VAE prior” approaches the “full calibration” case, with an average drop of about 10%. The C-VAE generator is therefore able to sample from a distribution which is close to the “true” prior model.

VI. Conclusions and future activities

The paper proposed the use of EM-informed generative neural network models (GNN) to predict body-induced diffraction properties of body attenuations, while at the same time generating samples from the prior distribution of the target position. The proposed use case is critical in industrial scenarios where human workers operate in areas featuring increasing level of safety or privacy. Enforcing safety/privacy constraints requires the real-time monitoring and tracking of the human subject.

Using RSS samples collected from measurements, Tab. V analyzes the precision and recall probabilities:

$$p_1 (d_R) = \Pr[p_k \in L (d_R) \mid \hat{p}_k \in L (d_R)]$$
$$p_2 (d_R) = \Pr[p_k \in L (d_R) \mid \hat{p}_k \in L (d_R)],$$

with $L (d_R)$ being the region that contains the positions $p_k$ of the target at distance of $d_R$ from the RX. The recall metric measures how often the algorithm correctly identifies the target distance from all the true positive counts, while the precision indicates how often the algorithm is correct when predicting the target distance. In particular, the table analyzes precision and recall for varying distance $d_R$ from the RX and compares three cases:

1) “full calibration” scenario assumes the prior probability $p_{\text{gen}} (A_t \mid \theta_k)$ being known as obtained from measurements collected in the same environment during a calibration stage;
2) “C-VAE prior” adopts the probability $p_{\text{gen}} (A_t \mid \theta_k)$ as prior model with samples obtained using the C-VAE generator tool;
3) “Uniform prior” represents a case where no information on excess attenuation is available: the prior is replaced with a uniform probability function $U [-5dB, 15dB]$ with attenuations ranging from $-5$dB to $15$dB.

Note that scenario (1) gives the best performance, as expected; on the other hand, it requires time-consuming data collection and calibration stages which might be often unfeasible. Case (3) corresponds to the worst case scenario since no prior information on body-induced attenuations are available. Finally, case (2) uses the C-VAE tool to real-time generate samples from the prior $p_{\text{gen}} (A_t \mid \theta_k)$ and probability (17); calibration is thus not needed. From the results in the Table, the performance of “C-VAE prior” approaches the “full calibration” case, with an average drop of about 10%. The C-VAE generator is therefore able to sample from a distribution which is close to the “true” prior model.
tion effects with applications to passive radio sensing. A variational auto-encoder (VAE) tool, namely the Conditional VAE (C-VAE), is designed to generate samples of the targeted EM model through latent variable encoding/decoding neural network operations. Adaptations of Generative Adversarial Networks (GAN) (Unbalanced Conditional GAN) are also considered for comparative analysis. The models are designed to reproduce the effects of body-induced EM diffraction under different configurations, propagation scenarios and MIMO antenna settings. Generated samples are in the form of either received signal strength (RSS) attenuations, base-band channel state information or EM field responses. For each case, the proposed generators are based on sequences of convolutional (and de-convolutional) layers whose number varies depending on the physical quantity to be reproduced. This choice is revealed to be effective in minimizing the complexity, the generation times and the neural network model footprint which is critical for implementation in resource-constrained devices (sensors, IoT devices, robots/machines).

Targeting passive radio localization applications, we optimized the proposed GNN tools to learn an informative prior distribution as instrumental to Bayesian estimation. The generative modelling methods have been validated with RSS measurements taken in an indoor site. The proposed C-VAE approach is well-suited for real-time target tracking as it does not require intensive or ad-hoc EM computations. Besides data augmentation and Bayesian prior generation, the model is shown to provide generalization capabilities, namely it can be used to predict body-induced effects under unknown configurations, subject orientations, size or locations. Examples tailored for passive localization also reveal the possibility of optimizing the generation process so to limit the use of time-consuming calibration stages.

Although still in their infancy, we expect physics-informed GNN models to become indispensable tools for designers in different application contexts. Considering the advances in wireless communication (h5G) where dense massive connected antenna arrays, as well as millimeter-band signals with Gigahertz-scale spectral width are being now deployed in consumer devices, future radio sensing tools are expected to be paired with accurate EM modeling in high frequency bands. For example, holography methods for dynamic tracking of multiple discrete targets require the processing of large CSI tensor structures (holograms) which represent the full RF radiation field. Generative AI might be deployed to efficiently isolate signatures of moving targets and reproduce synthetic hologram recordings from limited measurement sets to foster the learning approaches. The proposed tools have been also proved to be effective in reproducing the effects of body motions in user selectable locations. As exemplified in the case study, this property is instrumental to implement privacy selectivity policies which require the radio sensing system to perform differently depending on the specific subject locations, i.e., as characterized by variable level of safety and/or privacy.

REFERENCES


