Physics-Informed Coding for Sound Field Estimation in Wireless Acoustic Sensor Networks

Matthias Blochberger[1], Jan Østergaard[2], Filip Elvander[3], Marc Moonen[1], Toon van Waterschoot[1]
[1]*STADIUS, Dept. of Electrical Engineering (ESAT), KU Leuven*, Leuven, Belgium
[2]*Dept. of Electronic Systems, Aalborg University*, Aalborg, Denmark
[3]*Dept. of Information and Communications Engineering, Aalto University*, Espoo, Finland

***Abstract*—We introduce a latent coding approach for distributed sound field estimation in wireless acoustic sensor networks (WASNs) that leverages physics-based acoustic models and priors. Instead of transmitting raw microphone signals within the WASN, we propose transmitting compressed latent representations that encode information about the sound field, based on known acoustic models such as the plane wave (PW) decomposition and the boundary element method (BEM). This leads to multiple benefits, such as scalability through distributed aggregation, greater flexibility in network topology, and improved communication efficiency. During the training phase, a global latent prior is learned from data collected at a set of measurement positions distributed across the sound field. This prior is subsequently used locally at each node to condition the physics-informed mapping from signal space to latent space, thereby enabling efficient communication of compressed latent representations between WASN nodes. The approach employs a reverse water-filling strategy to assign bit rates to latent-space coordinates via linear coupling maps that map latent coordinates to estimation error. Simulations demonstrate the performance of the proposed approach using rate-distortion functions and the trade-off between estimation performance and allocated rate under an imposed target distortion.**

***Index Terms*—sound field estimation, wireless acoustic sensor networks, latent coding, physics-informed models, rate-distortion theory**

I. Introduction

Accurate sound field estimation in wireless acoustic sensor networks (WASNs) is challenged by two coupled constraints: the physical structure of sound propagation and the communication budget available to exchange information across distributed sensor nodes. There is a rich literature on sound field representation and estimation using physics-informed models, e.g., plane wave (PW) expansions [1] and boundary element method (BEM) [2] that exploit the underlying Helmholtz equation. Classical approaches [3]–[5] assume centralized access to raw measurements and fixed spatial bases. As a result, scalability with respect to sensor layouts, room geometry, and bandwidth remains limited. We propose a physics-informed latent coding framework that compresses and communicates modal information rather than raw microphone signals. The combination of acoustic modeling with information theory was pioneered in [6] framing the sound field in a room as the plenacoustic function, and deriving sampling and interpolation conditions.

During a one-time training phase, a global latent prior is learned from representative sound fields that have been mapped through Helmholtz-consistent feature operators, such as Herglotz/PW dictionaries and boundary-integral/BEM operators [1], [2]. The learned prior shapes a low-rank latent basis whose coordinates correspond with physically meaningful degrees of freedom while remaining agnostic to the eventual sensor and estimation layouts. At run time, each node projects its local measurement onto this shared latent space via the corresponding row of the physics-informed map, producing additive local summaries that can be aggregated with lightweight distributed averaging/consensus [7]. A regularized decoder then reconstructs the field at arbitrary estimation positions. To manage communication rate, we explicitly quantize in the latent space and couple quantization variance to estimation error through a linear sensitivity map derived from the estimator, which yields a reverse water-filling-style bit allocation across latent coordinates and frequencies guided by the estimator covariance under a Gaussian model [8]. This allocation provides rate targets that can be split across nodes (e.g., uniformly or topology-weighted) without retraining or re-factoring the latent space.

Rather than introducing a new representation, we focus on how a PW/BEM-shaped, potentially low-rank latent space can be exploited for distributed sound field estimation and rate control in WASNs. We show that the local latent representations are additive across nodes, so simple quantized consensus recovers the aggregate needed for estimation without centralizing raw signals. We link per-dimension latent quantization to estimation error via a linear sensitivity map of the regularized estimator and pose a convex geometric program that implements reverse water-filling across modes and frequencies under a Gaussian model—yielding rate targets that can be split across nodes. Finally, simulations in a reverberant room compare PW vs. BEM at different parameter combinations, reporting rate-distortion (RD) curves and normalized mean square error (NMSE) to illustrate the benefits of exchanging latent coordinates.

This research work was carried out at the ESAT Laboratory of KU Leuven, and has received funding from KU Leuven internal funds (C14/21/075), Research Foundation Flanders (FWO projects S005525N and G0A0424N), the Flemish Government (AI Research Program), and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.

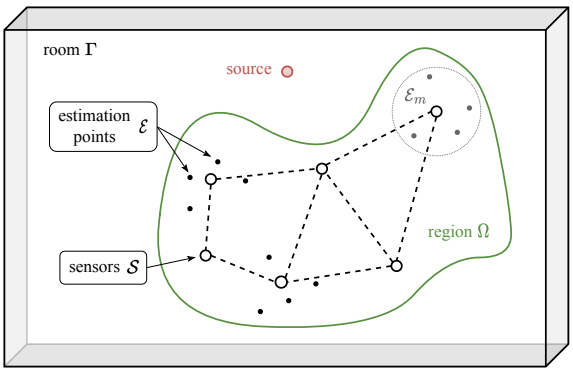

Fig. 1. Illustration of the problem setup with a WASN in a room Γ, showing a sound source, the source-free region Ω, sensor network with positions $\mathcal{S}$, as well as estimation positions $\mathcal{E}$ (indicated is also a node-local subset $\mathcal{E}_m \subseteq \mathcal{E}$).

II. Model

Consider a WASN deployed in a reverberant room $\Gamma \subset \mathbb{R}^3$ for the task of sound field estimation within a source-free region of interest $\Omega \subset \Gamma$ [1], [2], [4]. Let this WASN consist of S sensor nodes, each equipped with a microphone, with positions $\mathcal{S} = \{\mathbf{p}_1, \dots, \mathbf{p}_S\} \subset \Omega$. The goal is to estimate the sound field at a set of estimation positions $\mathcal{E} = \{\mathbf{q}_1, \dots, \mathbf{q}_E\} \subset \Omega$, where each node might only be interested in a subset $\mathcal{E}_m \subseteq \mathcal{E}$ of the estimation positions. The setup is illustrated in Figure 1.

A. Measurement Model

Let the sound field generated at one position within the room by the sound source and the room boundary scattering be denoted by $u_k : \Gamma \to \mathbb{C}$ for the wave number $k = \frac{\omega}{c}$, with angular frequency ω and speed of sound c. Each sensor node m measures the sound field at its position with coordinate vector $\mathbf{p}_m$, corrupted by additive noise. For each wave number $k \in \mathcal{K}$ within the set of wave numbers of interest, i.e., a frequency band of interest, the complex measurement model of the sound field is therefore denoted by $y_k(\mathbf{p}_m) = u_k(\mathbf{p}_m) + \varepsilon_{k,m}$ for $m = 1, \dots, S$, where $\varepsilon_{k,m}$ is the measurement noise at node m with variance σ_ε^2. We omit the subscript k in the following for brevity; the model applies to each frequency independently. For the full set of sensor positions, we can write the measurement model in vector form as

$$\mathbf{y}_\mathcal{S} = \mathbf{u}_\mathcal{S} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{CN}(0, \sigma_\varepsilon^2 \mathbf{I}) \tag{1}$$

where $\mathbf{y}_\mathcal{S} = [y(\mathbf{p}_1), \dots, y(\mathbf{p}_S)]^\top$ and $\mathbf{u}_\mathcal{S} = [u(\mathbf{p}_1), \dots, u(\mathbf{p}_S)]^\top$, and the measurement noise $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_S]^\top$, which is complex, circularly symmetric Gaussian noise that is spatially white.

To accurately reconstruct the sound field at the estimation positions, it is necessary to leverage the underlying physics of sound propagation, which is governed by the Helmholtz equation [1], [2]. For this, two well-established physics-based methods, both representing solutions to the Helmholtz equation, are considered: the PW method and the BEM. Through these methods, we can construct physics-informed feature maps that map any position in the source-free region to a latent space that represents a physics-consistent decomposition of the sound field. These methods are well-established in literature, and we will state the resulting maps without detailed derivations, referring the reader to the relevant literature [1], [2], [4], [5] for details.

Plane wave (PW) map: Choose a set of P PW directions $\{\hat{\boldsymbol{\zeta}}_p\}_{p=1}^P$ on the unit sphere and a set of A positions $\{\mathbf{q}_a\}_{a=1}^A$ in the source-free region, and define the PW map as

$$\left[\boldsymbol{\Phi}_\mathcal{A}^{\mathrm{PW}}\right]_{a,p} \coloneqq e^{jk\hat{\boldsymbol{\zeta}}_p^\top \mathbf{q}_a}. \tag{2}$$

We use $[\cdot]_{a,p}$ to denote the matrix element indexed by a and p. This is the discrete Herglotz representation of Helmholtz solutions as a superposition of PWs [1], which has been used, e.g., in [3], [4]. The latent space coordinates in this case correspond to the weights of the PWs.

Boundary element method (BEM) map: The boundary integral representation of the sound field at $\mathbf{q}_a \in \Omega$ is given by [2]

$$u(\mathbf{q}_a) = \int_{\partial\Omega} \Big(G(\mathbf{q}_a, \mathbf{r}) \partial_\mathbf{n} u(\mathbf{r}) - \partial_\mathbf{n} G(\mathbf{q}_a, \mathbf{r}) u(\mathbf{r}) \Big) \mathrm{d}\mathbf{r}, \tag{3}$$

with the free-field Green's function $G(\mathbf{q}, \mathbf{r}) = e^{jk\|\mathbf{q}-\mathbf{r}\|_2}/(4\pi\|\mathbf{q} - \mathbf{r}\|_2)$, and where $\partial_\mathbf{n}$ denotes the normal derivative at the boundary point $\mathbf{r}$ on the region boundary $\partial\Omega$. For numerical implementation, the boundary integral is discretized using boundary elements [2]. Using single and double layer potentials on the discretized boundary $\partial\Omega$, represented by boundary elements $\{(\mathbf{r}_p, \mathbf{n}_p, w_p)\}_{p=1}^{P'}$ (position, outward normal, quadrature weight), the discretized operators are [2]

$$[\mathbf{L}_\mathcal{A}]_{a,p} = w_p G(\mathbf{q}_a, \mathbf{r}_p), \quad [\mathbf{M}_\mathcal{A}]_{a,p} = w_p \partial_{\mathbf{n}_p} G(\mathbf{q}_a, \mathbf{r}_p). \tag{4}$$

With these operators, we define the double-layer BEM map as

$$\boldsymbol{\Phi}_\mathcal{A}^{\mathrm{BEM}} \coloneqq \begin{bmatrix} \mathbf{M}_\mathcal{A} & -\mathbf{L}_\mathcal{A} \end{bmatrix}. \tag{5}$$

The latent space coordinates in this case correspond to monopole and dipole strengths for BEM at the boundary element positions. Note that the latent space for BEM has dimension $P = 2P'$.

B. Latent Representation

The established methods for sound field estimation using these physics-based maps explicitly or implicitly assume access to the full set of measurements $\mathbf{y}_\mathcal{S}$ at a central processing unit. Here we show how the physics-informed latent representation conveniently decomposes additively across nodes, enabling distributed aggregation. For this, let us restate the measurement model (1) as

$$\mathbf{y}_\mathcal{S} = \mathbf{F}_\mathcal{S} \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{6}$$

where $\boldsymbol{\eta} \in \mathbb{C}^L$ are the latent variables and

$$\mathbf{F}_\mathcal{S} = \boldsymbol{\Phi}_\mathcal{S} \mathbf{B} \in \mathbb{C}^{S \times L} \tag{7}$$

is the physics-informed map corresponding to the sensor positions, shaped with additional information in the form of a learned latent prior, represented by the matrix $\mathbf{B} \in \mathbb{C}^{P \times L}$. The learned latent prior may introduce low-rank structure when

$L \leq P$, i.e., reduce the number of latent coordinates, and we will discuss this prior in more detail later.

To reconstruct the sound field at the estimation positions, apply ridge regression in the latent representation

$$\hat{\mathbf{u}}_{\mathcal{E}} = \mathbf{F}_{\mathcal{E}}\hat{\boldsymbol{\eta}}, \quad \hat{\boldsymbol{\eta}} = \mathbf{H}_{\mathcal{S}}\boldsymbol{\psi}, \tag{8}$$

where $\boldsymbol{\psi} = \mathbf{F}_{\mathcal{S}}^{\mathrm{H}}\mathbf{y}_{\mathcal{S}}$ is the projection of the measurements onto the latent space and $\mathbf{H}_{\mathcal{S}} = \left(\mathbf{G}_{\mathcal{S}} + \sigma_{\varepsilon}^2\mathbf{I}_L\right)^{-1}$ with $\mathbf{G}_{\mathcal{S}} = \mathbf{F}_{\mathcal{S}}^{\mathrm{H}}\mathbf{F}_{\mathcal{S}}$ being the Gram matrix. The vector $\hat{\boldsymbol{\eta}}$ contains the estimated latent variables, and $\hat{\mathbf{u}}_{\mathcal{E}}$ is the reconstructed sound field at the estimation positions. The latent variable $\boldsymbol{\psi}$ can be written as a sum of contributions from each node as

$$\boldsymbol{\psi} = \mathbf{F}_{\mathcal{S}}^{\mathrm{H}}\mathbf{y}_{\mathcal{S}} = \sum_{m=1}^{S} \underbrace{\mathbf{F}_m^{\mathrm{H}} y(\mathbf{p}_m)}_{\boldsymbol{\psi}_m}, \tag{9}$$

where $\mathbf{F}_m \in \mathbb{C}^{1\times L}$ is the physics-informed map corresponding to the position of node m, i.e., the m-th row of $\mathbf{F}_{\mathcal{S}}$. This is essentially the encoder, which maps the local measurement at node m to a local latent representation $\boldsymbol{\psi}_m$. We can define a decoder as $\mathbf{D} = \mathbf{F}_{\mathcal{E}}\mathbf{H}_{\mathcal{S}}$, then the estimator can be written as $\hat{\mathbf{u}}_{\mathcal{E}} = \mathbf{D}\boldsymbol{\psi}$. Note that locally at each node, only a subset of estimation positions may be of interest, therefore a local decoder $\mathbf{D}_m$ can be defined for each node m that only reconstructs for a set of estimation positions $\mathcal{E}_m \subseteq \mathcal{E}$.

Using this theoretical setup, we will now introduce the proposed framework for physics-informed coding in WASNs.

III. Method

Instead of transmitting raw signal data $y(\mathbf{p}_m)$ to other nodes or a fusion center, we propose to communicate in the latent space defined in Section II, i.e., nodes will collaborate to compute a physics-informed global model of the sound field and communicate compressed latent information and subsequently reconstruct the sound field at estimation positions. While this increases the number of scalars that need to be transmitted per node, it allows topology independent communication and estimation. We make the following assumptions for the proposed method:

- There is a training phase where data is collected at measurement positions and a central unit learns the global latent prior $\mathbf{B}$ from the data.
- Each node has knowledge of the global latent prior $\mathbf{B}$.
- Each node can compute the physics-based map $\mathbf{F}_m$.
- Each node has the decoder $\mathbf{D}_m$ available. Either via consensus on the Gram matrix $\mathbf{G}_{\mathcal{S}} = \sum_m \mathbf{F}_m^{\mathrm{H}}\mathbf{F}_m$ or learned from the training data.
- Communication links are instantaneous and error-free.

In a training phase, the global latent prior is learned from data collected at a set of measurement positions distributed in the sound field. Given training positions $\mathcal{T} = \{\mathbf{q}_t\}_{t=1}^{T} \subset \Omega$ and sound field measurements $y(\mathbf{q}_t)$, compute $\boldsymbol{\Phi}_{\mathcal{T}}$ (PW or BEM), then form sample covariances and learn $\mathbf{B}$ via ridge regression and truncated singular value decomposition (SVD) to a target rank L. This encodes the global prior $\boldsymbol{\Sigma} \approx \mathbf{B}\mathbf{B}^H$ used then at arbitrary $\mathcal{S}, \mathcal{E}$. If required, the decoder can be learned from the training data as well, e.g., via ridge regression $\min \|\mathbf{y}_{\mathcal{E}_m} - \mathbf{D}_m\mathbf{F}_{\mathcal{T}}\mathbf{y}_{\mathcal{T}}\|_2^2 + \lambda\|\mathbf{D}_m\|_F^2$, if the estimation set is contained in the training set, $\mathcal{E}_m \subset \mathcal{T}$.

In order to reconstruct the sound field at the estimation positions, the nodes need to collaboratively compute $\boldsymbol{\psi} = \sum_m \boldsymbol{\psi}_m$, which can be achieved via distributed consensus [7] for example. We show how the physics-informed latent representation allows for distributed aggregation, how to allocate quantization noise across latent coordinates from a global perspective and how this affects the estimation performance, rather than proposing a specific consensus algorithm. We will later give an example of how this can be used in a simple distributed averaging scheme.

To establish this, we need to quantify the effect of quantization noise on the estimation performance. Note that the following analysis is from a global network-wide perspective. Let us define the mean-square error (MSE) incurred on the estimator due to quantization noise being the quantity which we aim to keep below a certain distortion budget,

$$\mathbb{E}\left[\|\delta\hat{\mathbf{u}}_{\mathcal{E}}\|_2^2\right] \leq D, \quad D > 0, \tag{10}$$

while minimizing the total communication rate across nodes.

The covariance of the estimator $\hat{\mathbf{u}}_{\mathcal{E}}$ (8) can be shown to be

$$\begin{aligned}\boldsymbol{\Sigma}_{\hat{\mathbf{u}}_{\mathcal{E}}} &= \mathbf{F}_{\mathcal{E}}\mathbf{H}_{\mathcal{S}}\mathbf{G}_{\mathcal{S}}\mathbf{F}_{\mathcal{E}}^{\mathrm{H}} \\ &= \mathbf{Q}\,\mathrm{diag}\left([\mu_1, \ldots, \mu_E]\right)\mathbf{Q}^{\mathrm{H}},\end{aligned} \tag{11}$$

where $\mathbf{Q}$ is the eigenvector matrix and μ_j are the eigenvalues of the covariance of the estimator.

We work under the assumption that the modes μ_j of the estimator covariance are independent, which might not be strictly true, however, it provides a useful approximation. Using (8) we can use a first-order approximation to map additive quantization noise $\delta\boldsymbol{\psi}$ to the estimator by

$$\delta\hat{\mathbf{u}}_{\mathcal{E}} \approx \mathbf{F}_{\mathcal{E}}\mathbf{H}_{\mathcal{S}}\,\delta\boldsymbol{\psi}. \tag{12}$$

Since the MSE (10) is invariant under unitary transformations, we can project the error onto the orthogonal modes of (11),

$$\mathbf{Q}^{\mathrm{H}}\,\delta\hat{\mathbf{u}}_{\mathcal{E}} \approx \mathbf{Q}^{\mathrm{H}}\mathbf{F}_{\mathcal{E}}\mathbf{H}_{\mathcal{S}}\,\delta\boldsymbol{\psi} =: \mathbf{T}\,\delta\boldsymbol{\psi}, \tag{13}$$

which gives us a linear map $\mathbf{T} \in \mathbb{C}^{E\times L}$ from quantization noise in the latent space to estimator modes μ_j, i.e., estimation error. Assume that each element $[\delta\boldsymbol{\psi}]_\ell$, $\ell = 1, \ldots, L$, of the quantization noise has uniform distribution with variance σ_ℓ^2, and is quantized with an independent scalar quantizer (for real and imaginary part). Then we can write the MSE as

$$\mathbb{E}\left[\|\delta\hat{\mathbf{u}}_{\mathcal{E}}\|_2^2\right] = \mathrm{tr}\big(\mathbf{T}\,\mathbb{E}\left[\delta\boldsymbol{\psi}\,\delta\boldsymbol{\psi}^{\mathrm{H}}\right]\mathbf{T}^{\mathrm{H}}\big) = \sum_{j,\ell}^{E,L} \left|[\mathbf{T}]_{j,\ell}\right|^2\sigma_\ell^2. \tag{14}$$

We define the coupling matrix $\mathbf{A} \in \mathbb{R}^{E\times L}$ with elements $[\mathbf{A}]_{j,\ell} = \left|[\mathbf{T}]_{j,\ell}\right|^2$, i.e., the squared magnitudes of the coupling coefficients between the quantization noise variances σ_ℓ^2 and the estimator MSE. The goal is to find quantization noise variances σ_ℓ^2 such that the total rate is minimized (reduced) given the distortion constraint (10). Define the vectors of variances and distortion targets as $\mathbf{z} := \left[\sigma_1^2, \ldots, \sigma_L^2\right]$, and

$\mathbf{d} \coloneqq [d_1, \ldots, d_E]$. The standard view on rate-minimization is to maximize the high-rate proxy $\sum_\ell \log \sigma_\ell^2$. Since log is strictly increasing and the domain of the problem is $\sigma_\ell^2 > 0$, maximizing $\prod_\ell \sigma_\ell^2$ is equivalent [9]. The variance-allocation or rate-minimization problem can therefore be cast as the geometric program (GP),

$$\begin{aligned} \max_{\sigma_\ell^2 > 0} \quad & \prod_{\ell=1}^{L} \sigma_\ell^2 \\ \text{s.t.} \quad & \mathbf{A}\mathbf{z} \leq \mathbf{d}, \end{aligned} \tag{15}$$

which can be solved efficiently by specialized interior-point routines such as ECOS [10] or SCS [11].

The distortion targets $\mathbf{d}$ are determined from the total distortion budget D via reverse water-filling [8], [12]. We re-introduce the frequency index k to account for multiple frequencies within the allocated budget D, coupling the distortion targets across frequencies. The distortion targets are determined by choosing a water level θ such that

$$d_{jk} = \min(\mu_{jk}, \theta), \quad \sum_{k \in \mathcal{K}} \sum_{j=1}^{E} d_{jk} = D. \tag{16}$$

The standard expression for the Gaussian RD function [8] is

$$R(D) = \sum_{k \in \mathcal{K}} \sum_{j=1}^{E} \frac{1}{2} \max\left(\log_2 \frac{\mu_{jk}}{d_{jk}}, 0\right), \tag{17}$$

with per-mode power μ_{jk} and distortion d_{jk}. Note that (15) can still be solved for each frequency independently. The maximization problem (15) yields variances $\sigma_{\ell k}^{2\star}$ that minimize the total rate given the distortion targets under our assumptions.

Let us give an example of how this can be used in a simple distributed averaging scheme: given the optimal variances $\sigma_{\ell k}^{2\star}$, which are computed from a global perspective, we can assign rates to nodes by allocating a portion of the variance to each node, e.g., equally, $\sigma_{\ell k m}^{2\star} = \sigma_{\ell k}^{2\star}/M$ or based on some other criterion such as a network topology-based weight similar to [7]. The corresponding quantizer steps, using the standard approximation of the variance of quantization noise, choosing the same step for real and imaginary parts $\sigma^2 \approx (\Delta_{\mathrm{Re}}^2 + \Delta_{\mathrm{Im}}^2)/12$, with $\Delta_{\mathrm{Re}} = \Delta_{\mathrm{Im}} = \Delta$ is $\Delta_{\ell k m}^{2\star} \approx 6\sigma_{\ell k m}^{2\star}$. And then each node can quantize its contribution $\boldsymbol{\psi}_m$ to the latent representation with the corresponding quantizer step before communicating it to neighboring nodes for aggregation.

IV. Numerical Experiments

We consider a box-shaped room of size $5.0\,\mathrm{m} \times 4.0\,\mathrm{m} \times 3.0\,\mathrm{m}$. Let us consider a cubic volume of interest Ω with side length of $1\,\mathrm{m}$ within the room, centered at $(1.0, 1.0, 1.5)$ m. We generate the sound field by placing a monopole spectrally white source at $(4.0, 1.0, 1.5)$ m and using the randomized image source method (RISM) [13] to compute the sound field at the positions of interest. We consider frequencies from $100\,\mathrm{Hz}$ to $1000\,\mathrm{Hz}$ in steps of $100\,\mathrm{Hz}$. The speed of sound is $c = 343\,\mathrm{m/s}$ and the medium density $\rho = 1.21\,\mathrm{kg/m^3}$. The latent elements are P weights of the PW directions distributed on the unit sphere using a Fibonacci lattice and P triangular boundary elements discretizing an ellipsoid enclosing Ω.

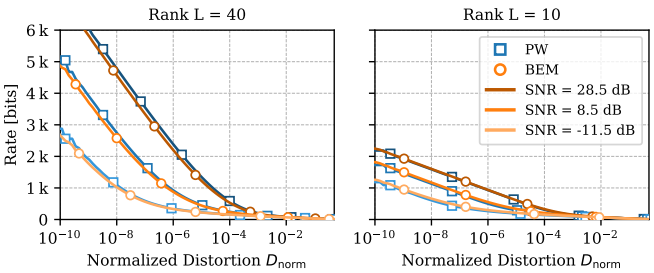

Fig. 2. Theoretical RD curves (17) with normalized distortion and total rate for physics-informed latent maps at chosen noise levels and target ranks L.

A. Training

For the given acoustic scenario, we consider a training set of $T = 100$ random positions $\{\mathbf{q}_t\}_{t=1}^T \subset \Omega$ on a uniform grid with small random perturbations ($0.05\,\mathrm{m}$) for better conditioning. We simulate 1000 random sound fields for source positions placed randomly around position $(4.0, 1.0, 1.5)$ m with random displacements drawn from a Gaussian distribution with standard deviation of $0.5\,\mathrm{m}$ along each coordinate axis. We compute the physics-based maps $\boldsymbol{\Phi}_{\mathcal{T}}$ for the training positions $\mathcal{T} = \{\mathbf{q}_t\}_{t=1}^T$ using both PW and BEM with $P = 320$ elements. Then, we compute the prior $\mathbf{B}$ using ridge regression with regularization parameter $\gamma = 10^{-4}$ and a noise variance of $\sigma_\varepsilon^2 = 10^{-12}$ (noiseless) and truncating the SVD to target ranks $L \in \{10, 20, 30, 40\}$.

B. Rate-Distortion and Sound Field Estimation

We first consider the theoretical RD curves (17) for physics-informed latent maps with noise levels ($\sigma_\varepsilon^2 \in \{1^{-6}, 1^{-4}, 1^{-2}\}$), corresponding to signal-to-noise ratios (SNRs) of $\{28.5, 8.5, -11.5\}\,\mathrm{dB}$. We place $S = 40$ sensor nodes and $E = 80$ estimation positions randomly in Ω. The number of elements in the physics-based maps is $P = 320$ for both PW and boundary elements BEM, and we vary $L \in \{10, 40\}$. Figure 2 shows the mean of 50 random realizations. Shown is the total distortion normalized by the total power in the modes

$$D_{\mathrm{norm}} = \frac{\sum_{j,k} d_{jk}}{\sum_{j,k} \mu_{jk}}, \tag{18}$$

and the total rate in bits defined by (17). We observe the rate-distortion trade-off, where the rate increases with decreasing distortion. For lower SNRs (higher σ_ε^2), the rate is lower for a given distortion, as expected, as well as for lower target rank L.

We then consider the sound field estimation performance of the proposed method with quantization using the distortion-allocation strategy described in Section III. Consider the same setup as above with $S = 40$ and $E = 80$, and consider $L \in \{10, 20, 30, 40\}$. We compute the NMSE and the relative error to the unquantized case over all estimation positions at each frequency k as

$$\mathrm{NMSE} = \frac{\mathbb{E}\left[\|\hat{\mathbf{u}}_{\mathcal{E}} - \mathbf{u}_{\mathcal{E}}\|_2^2\right]}{\mathbb{E}\left[\|\mathbf{u}_{\mathcal{E}}\|_2^2\right]}, \quad \mathrm{RNMSE} = \frac{\mathrm{NMSE}_{D>0}}{\mathrm{NMSE}_{D=0}}, \tag{19}$$

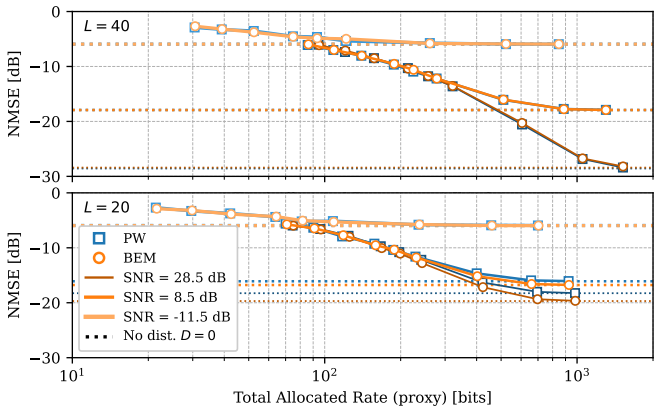

Fig. 3. NMSE of the sound field estimation at the estimation positions for different target ranks $L \in \{20, 40\}$ and noise levels σ_ε^2 with the proposed distortion-allocation strategy over allocated rate R_{total}. Dotted lines indicate the unquantized NMSE of the estimator for each noise level.

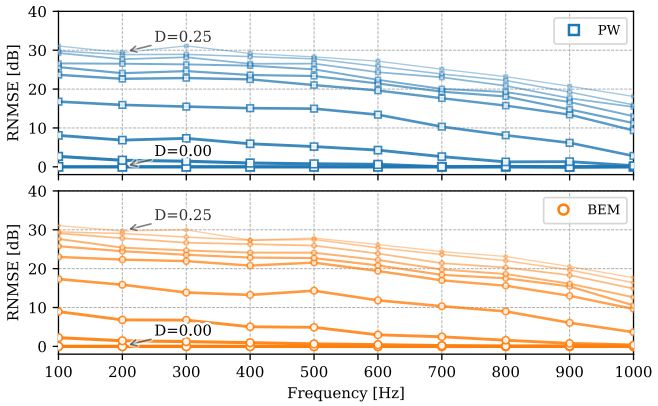

Fig. 4. Relative NMSE per frequency of the sound field estimation task at the estimation positions for $L = 40$ and noise level $\sigma_\varepsilon^2 = 1^{-6}$ showing the effect of distortion $D \in \{1^{-5}, \ldots, 2.5^{-1}\}$ on estimation error across frequencies.

where the expectation is approximated by averaging over the 20 random sound fields. Further, we compute the total allocated rate as

$$R_{\text{total}} = \sum_{\ell,k} \log_2 \frac{\nu_{\ell k}^2}{\sigma_{\ell k}^{2\star}}, \tag{20}$$

where $\nu_{\ell k}^2$ is the variance of the quantizer input $[\psi_k]_\ell$ and $\sigma_{\ell k}^{2\star}$ is the allocated quantization variance from the distortion-allocation problem (15).

Figure 3 shows the NMSE of the sound field estimation at the estimation positions for ranks $L \in \{20, 40\}$ and noise levels $\sigma_\varepsilon^2 \in \{1^{-6}, 1^{-4}, 1^{-2}\}$ over the allocated rate R_{total}. We can observe that the NMSE decreases with increasing allocated rate, to a plateau governed by the distortion-less NMSE of the estimator. For lower SNRs (higher σ_ε^2) the NMSE increases while the allocated rate is lower, as expected. The rank L has a limited effect on the NMSE and allocated rate, however, it does lift the plateau of the NMSE for low distortion levels. It is however to note that the impact of L on the NMSE depends on the complexity of the sound field, i.e., multiple sources, many reflections, etc. This has not been explored further here, but requires further investigation.

Figure 4 shows the relative NMSE per frequency of the sound field estimation task. We can see that the effect of the distortion D on the estimation error is higher at lower frequencies, however, we can observe that both PW and BEM follow the same trend, i.e., the same frequencies are more or less affected by distortion.

V. Conclusion

We have proposed a method for physics-informed coding of sound fields in WASNs using latent space representations with learned priors. We showed how the physics-informed latent representation allows for distributed aggregation and how to allocate quantization noise across latent coordinates from a global perspective. Numerical experiments in a simulated acoustic scenario showed the rate-distortion trade-off and the effect of the proposed distortion-allocation strategy on sound field estimation.

This work opens up several avenues for future research, including relaxing assumptions made here, e.g., moving towards fully distributed coding schemes, online learning of the latent prior and decoder, and more complex acoustic scenarios. Another path to take is to consider more general distortion measures, e.g., perceptual ones to move quantization noise to less perceptually relevant frequencies, and naturally, to consider non-linear data-driven (physics-informed) latent representations.

References

[1] E. G. Williams and J. A. Mann, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Oct. 2000, vol. 108.

[2] S. Kirkup, "The Boundary Element Method in Acoustics: A Survey," *Applied Sciences*, vol. 9, no. 8, p. 1642, Apr. 2019.

[3] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel Ridge Regression with Constraint of Helmholtz Equation for Sound Field Interpolation," in *Proc. 2018 Int. Workshop Acoustic Signal Enhancement (IWAENC '18)*. Tokyo: IEEE, Sep. 2018, pp. 1–440.

[4] D. Sundstrom, F. Elvander, and A. Jakobsson, "Bayesian Sound Field Reconstruction Using Partial Boundary Information," *IEEE Trans. Audio Speech Lang. Process.*, vol. 33, pp. 4620–4631, 2025.

[5] J. Brunnström, M. B. Møller, J. Østergaard, and M. Moonen, "Bayesian sound field estimation using uncertain data," in *Proc. 2024 Int. Workshop Acoustic Signal Enhancement (IWAENC '24)*, 2024, pp. 329–333.

[6] T. Ajdler, L. Sbaiz, and M. Vetterli, "The Plenacoustic Function and Its Sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[7] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[8] T. M. Cover and J. A. Thomas, "Rate Distortion Theory," in *Elements of Information Theory*. Hoboken, NJ: Wiley-Interscience, 2001.

[9] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[10] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *Proc. 2013 Eur. Control Conf. (ECC)*. Zurich: IEEE, Jul. 2013, pp. 3071–3076.

[11] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding," *J Optim Theory Appl*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. New York: Wiley, 1991.

[13] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.