FORUM ACUSTICUM EURONOISE 2025

# CROSS-SPECTRAL ANALYSIS OF HIGHER-ORDER AMBISONIC SIGNALS FOR INVESTIGATING THE SPATIAL AND STATISTICAL PROPERTIES OF CAPTURED SOUND FIELDS

**Nara Hahn*** **Filippo Maria Fazi**
**Mihai Orita** **Bogdan Bacila** **Philip A. Nelson**

Institute of Sound and Vibration Research, University of Southampton, United Kingdom

## ABSTRACT

The analysis of spatial and statistical properties of a captured sound field is of significant interest in various fields, including room acoustics, spatial sound capture and reproduction. In this paper, we investigate these properties based on higher-order Ambisonic (HOA) signals. A cross-spectral matrix is constructed, where the diagonal and off-diagonal entries respectively represent the power spectral densities and the cross-spectral densities of the HOA signals. Through theoretical development, we demonstrate that the numerical properties of the cross-spectral matrix, such as its rank and eigenvalue distribution, are determined by the spatial distribution of incident sound waves and the correlation between source signals. The analysis is performed for simulated and real sound fields, considering various source distributions and Ambisonic orders.

**Keywords:** *cross-spectral analysis, covariance matrix, higher-order Ambisonics, B-format signals*

## 1. INTRODUCTION

Microphone array techniques are widely used in various acoustical applications, such as room acoustics analysis [1, 2], echo cancellation [3], and spatial sound capture [4, 5]. Recent advances in microphone arrays with higher-order directivity, combined with sophisticated signal processing techniques, have significantly enhanced the spatial resolution of such analyses. Spherical microphone arrays have gained particular attention due to their uniform spatial resolution [5]. The sound fields captured by these arrays are typically represented using spherical harmonic coefficients, which are inherently compatible with higher-order Ambisonics [4, 6]. A range of signal processing methods have been developed in the spherical harmonic domain. Notably, subspace-based techniques, such as those relying on singular value decomposition or eigenvalue decomposition, have been applied for tasks such as direction-of-arrival (DOA) estimation [7] and diffuseness estimation [8–10].

Quantifying the diffuseness of a sound field is of particular interest in applications where the estimated diffuseness is further used for the subsequent processing stages, e.g. echo cancellation and spatial encoding of captured sound fields [3, 11]. A sound field with directional diffuseness is typically characterised by two main properties: the isotropy of the incoming waves and the incoherence of their phases [12–15]. Spatial sound reproduction systems attempt to recreate these conditions and can even enhance the perceived diffuseness by manipulating the inter-channel correlation of the source signals, a process known as decorrelation [16–18]. In [8], the eigenvalue distribution of the covariance matrix in the Ambisonic domain was used to quantify the degree of diffuseness in a sound field. The spatial coherence was investigated in [19–21] for non-coincident microphones with first- and higher-order directivity patterns.

In this paper, we further investigate the eigenvalue distribution of the covariance matrix for sound fields with

**Corresponding author*: nara.hahn@soton.ac.uk.

different source correlation and source distributions. Our approach closely resembles that of [8], but the analysis is conducted in the short-time Fourier transform (STFT) domain rather than in the time domain. This paper is structured as follows. Section 2 reviews the Ambisonic encoding of plane waves in a free field and sound fields captured by microphone arrays. The covariance matrix of HOA signals is introduced in Sec. 3, where the influence of source correlation and the source positions are investigated for various scenarios. It is demonstrated that the eigenvalue behaviours of the source covariance matrices can be predicted by using the Marchenko-Pastur law. Section 4 then reports a case study where the presented analysis framework is applied to sound field reproduction scenarios with different loudspeaker arrangements. The paper concludes with potential directions for future studies in Sec. 5.

## 2. AMBISONIC ENCODING

We consider a sound field which comprises of $\mathcal{Q}$ plane waves whose propagation directions are denoted by unit vectors,

$$\hat{\boldsymbol{k}}_q = [\sin\vartheta_q \cos\varphi_q, \sin\vartheta_q \sin\varphi_q, \cos\vartheta_q]^{\mathrm{T}} \tag{1}$$

for $q = 1, \ldots, \mathcal{Q}$, where $\vartheta_q \in [0, \pi]$ and $\varphi_q \in [-\pi, \pi)$ denote the colatitude and azimuth angles, respectively. Each plane wave carries source signal with Fourier transform of $S_q(\omega)$. The angular frequency in $\mathrm{rad/s}$ is denoted by $\omega = 2\pi f$ where $f$ is the frequency in Hz.

By using the plane wave expansion, the sound field can be parametrised by plane wave direction, $\hat{\boldsymbol{k}} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^{\mathrm{T}}$, defined similarly as (1). In a free-field, the plane wave density function reads [4]

$$\bar{S}(\hat{\boldsymbol{k}}, \omega) = \sum_{q=1}^{\mathcal{Q}} \delta(\hat{\boldsymbol{k}} - \hat{\boldsymbol{k}}_q), \tag{2}$$

where $\delta(\hat{\boldsymbol{k}} - \hat{\boldsymbol{k}}_q) \coloneqq \delta(\phi - \varphi_q)\delta(\cos\theta - \cos\vartheta_q)$. The plane wave density function can be approximated by a truncated spherical harmonic expansion [6],

$$\bar{S}(\hat{\boldsymbol{k}}, \omega) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{S}_{n,m}(\omega)\, Y_{nm}(\hat{\boldsymbol{k}}). \tag{3}$$

We consider real-valued spherical harmonics, denoted by $Y_{nm}(\hat{\boldsymbol{k}}) \coloneqq Y_{nm}(\theta, \phi)$ with $n$ and $m$ denoting the order and degree, respectively. In 3-dimensional Ambisonics, the expansion coefficients $\tilde{S}_{nm}(\omega)$ are referred to as the B-format signals, and the truncation order $N$ as the Ambisonic order. For a given order $N$, the number of B-format channels are $\mathcal{L} = (N+1)^2$. Following the Ambisonic channel number (ACN) convention, the B-format channel index $l = 0, 1, \ldots$ is related to $n$ and $m$ by $l = n^2 + n + m$ [4].

The spatial encoding of the $\mathcal{Q}$ plane waves in a free field into B-format reads

$$\tilde{S}_{nm}(\omega) = \sum_{q=1}^{\mathcal{Q}} S_q(\omega)\, Y_{nm}(\hat{\boldsymbol{k}}_q), \tag{4}$$

which can be expressed in matrix form as

$$\tilde{\boldsymbol{s}} = \boldsymbol{Y}\boldsymbol{s}, \tag{5}$$

where $\tilde{\boldsymbol{s}}$ is the B-format signal vector ($\mathcal{L} \times 1$),

$$\tilde{\boldsymbol{s}} = \begin{bmatrix} \tilde{S}_{0,0} & \tilde{S}_{1,-1} & \tilde{S}_{1,0} & \tilde{S}_{1,1} & \ldots & \tilde{S}_{N,N} \end{bmatrix}^{\mathrm{T}}, \tag{6}$$

$\boldsymbol{Y}$ the matrix of spherical harmonics ($\mathcal{L} \times \mathcal{Q}$),

$$\boldsymbol{Y} = \begin{bmatrix} Y_{0,0}(\hat{\boldsymbol{k}}_1) & \ldots & Y_{0,0}(\hat{\boldsymbol{k}}_{\mathcal{Q}}) \\ \vdots & \ddots & \vdots \\ Y_{N,N}(\hat{\boldsymbol{k}}_1) & \ldots & Y_{N,N}(\hat{\boldsymbol{k}}_{\mathcal{Q}}) \end{bmatrix}, \tag{7}$$

and $\boldsymbol{s}$ the source signal vector ($\mathcal{Q} \times 1$),

$$\boldsymbol{s} = \begin{bmatrix} S_1(\omega) & \ldots & S_{\mathcal{Q}}(\omega) \end{bmatrix}^{\mathrm{T}}. \tag{8}$$

The vectors $\boldsymbol{s}$ and $\tilde{\boldsymbol{s}}$ are defined for each frequency. The frequency-dependency is suppressed for brevity.

The B-format signals can also be obtained from a captured sound field. For spherical microphone arrays, the encoding process can be expressed as [5]

$$\tilde{\boldsymbol{s}} = \boldsymbol{B}(\boldsymbol{Y}_{\mathrm{mic}}^{\mathrm{T}})^{\dagger}\boldsymbol{H}\boldsymbol{s}. \tag{9}$$

The transfer functions $H_{q,\mu}(\omega)$ from the sources to $\mathcal{M}$ microphones are represented by $\boldsymbol{H}$ ($\mathcal{M} \times \mathcal{Q}$),

$$\boldsymbol{H} = \begin{bmatrix} H_{1,1}(\omega) & \ldots & H_{\mathcal{Q},1}(\omega) \\ \vdots & \ddots & \vdots \\ H_{1,\mathcal{M}}(\omega) & \ldots & H_{\mathcal{Q},\mathcal{M}}(\omega) \end{bmatrix}. \tag{10}$$

$(\boldsymbol{Y}_{\mathrm{mic}}^{\mathrm{T}})^{\dagger}$ is the pseudo-inverse of $\boldsymbol{Y}_{\mathrm{mic}}^{\mathrm{T}}$ which has the size of $\mathcal{M} \times \mathcal{L}$ and defined as

$$\boldsymbol{Y}_{\mathrm{mic}}^{\mathrm{T}} = \begin{bmatrix} Y_{0,0}(\hat{\boldsymbol{x}}_1) & \ldots & Y_{N,N}(\hat{\boldsymbol{x}}_1) \\ \vdots & \ddots & \vdots \\ Y_{0,0}(\hat{\boldsymbol{x}}_{\mathcal{M}}) & \ldots & Y_{N,N}(\hat{\boldsymbol{x}}_{\mathcal{M}}) \end{bmatrix}, \tag{11}$$

where $\hat{\boldsymbol{x}}_\mu, \mu = 1, \ldots, \mathcal{M}$ represent the angular positions of the microphones. Finally, an equalisation (called radial filtering) is performed by the diagonal matrix $\boldsymbol{B}$, [4, Ch. 6]

$$\boldsymbol{B} = \operatorname{diag}\left(b_0(\tfrac{\omega}{c}r) \quad \ldots \quad b_N(\tfrac{\omega}{c}r)\right), \tag{12}$$

where $b_n(\frac{\omega}{c}r)$ denotes the respective filters that depend on the order $n$ and the array radius $r$.

Note from (5) and (9) that the spatial encoding from source signals to B-format signals is described by a matrix multiplication, $\boldsymbol{A}_{\text{model}} = \boldsymbol{Y}$ for model-based encoding and $\boldsymbol{A}_{\text{data}} = \boldsymbol{B}(\boldsymbol{Y}_{\text{mic}}^{\mathrm{T}})^{\dagger}\boldsymbol{H}$ for data-based encoding. The B-format signals obtained from a microphone array exhibit a limited Ambisonic order mainly because of the finite number of microphones. A regularisation is typically used in the radial filter design, cf. (11), in order to avoid excessive boost at high orders. These limit the operating frequency range in which all Ambisonic orders can be fully used. The components outside this range exhibit reduced Ambisonic order and spatial aliasing artefacts.

## 3. COVARIANCE MATRIX OF HOA SIGNALS

The covariance matrix of B-format signals $\tilde{\boldsymbol{s}}$ is defined as

$$\boldsymbol{C}_{\tilde{s}\tilde{s}} := \mathbb{E}\left\{(\tilde{\boldsymbol{s}} - \mathbb{E}(\tilde{\boldsymbol{s}}))(\tilde{\boldsymbol{s}} - \mathbb{E}(\tilde{\boldsymbol{s}}))^{\mathrm{H}}\right\}, \tag{13}$$

where $\mathbb{E}(\cdot)$ denotes the statistical expectation and $(\cdot)^{\mathrm{H}}$ the Hermitian transpose. The diagonal terms in $\boldsymbol{C}_{\tilde{s}\tilde{s}}$ represent the auto-spectra of the individual B-format signals, whereas the off-diagonal terms represent the cross-spectra of B-format signal pairs. An eigenvalue decomposition of the covariance matrix is performed

$$\boldsymbol{C}_{\tilde{s}\tilde{s}} = \boldsymbol{Q}_{\tilde{s}}\boldsymbol{\Lambda}_{\tilde{s}}\boldsymbol{Q}_{\tilde{s}}^{-1} = \boldsymbol{Q}_{\tilde{s}}\boldsymbol{\Lambda}_{\tilde{s}}\boldsymbol{Q}_{\tilde{s}}^{\mathrm{H}}, \tag{14}$$

where the columns of $\boldsymbol{Q}_{\tilde{s}}$ are the eigenvectors and the diagonal entries of $\boldsymbol{\Lambda}_{\tilde{s}}$ are the corresponding eigenvalues,

$$\boldsymbol{\Lambda}_{\tilde{s}} = \operatorname{diag}\left(\lambda_{\tilde{s}}^{(1)} \quad \ldots \quad \lambda_{\tilde{s}}^{(\mathcal{L})}\right). \tag{15}$$

The second equality in (14) follows from the fact that $\boldsymbol{C}_{\tilde{s}\tilde{s}}$ is positive semi-definite, which leads to $\boldsymbol{Q}_{\tilde{s}}^{-1} = \boldsymbol{Q}_{\tilde{s}}^{\mathrm{H}}$ [22].

The covariance matrix can also be expressed in terms of the spatial encoding matrix $\boldsymbol{Y}$ and the source signals $\boldsymbol{s}$. Substituting (5) into (13) yields

$$\begin{aligned}
\boldsymbol{C}_{\tilde{s}\tilde{s}} &= \boldsymbol{Y}\,\mathbb{E}\left\{(\boldsymbol{s} - \mathbb{E}(\boldsymbol{s}))(\boldsymbol{s} - \mathbb{E}(\boldsymbol{s}))^{\mathrm{H}}\right\}\boldsymbol{Y}^{\mathrm{H}} && (16)\\
&= \boldsymbol{Y}\boldsymbol{C}_{ss}\boldsymbol{Y}^{\mathrm{H}} && (17)\\
&= \boldsymbol{Y}\boldsymbol{Q}_s\boldsymbol{\Lambda}_s\boldsymbol{Q}_s^{-1}\boldsymbol{Y}^{\mathrm{H}} && (18)\\
&= \boldsymbol{Y}\boldsymbol{Q}_s\boldsymbol{\Lambda}_s\boldsymbol{Q}_s^{\mathrm{H}}\boldsymbol{Y}^{\mathrm{H}}. && (19)
\end{aligned}$$

The deterministic matrices, $\boldsymbol{Y}$ and $\boldsymbol{Y}^{\mathrm{H}}$ are factored out from $\mathbb{E}(\cdot)$ in the first equality (16). Eigenvalue decomposition of the source covariance matrix is performed in (17), $\boldsymbol{C}_{ss} = \mathbb{E}\left\{(\boldsymbol{s} - \mathbb{E}(\boldsymbol{s}))(\boldsymbol{s} - \mathbb{E}(\boldsymbol{s}))^{\mathrm{H}}\right\}$. $\boldsymbol{Q}_s$ and $\boldsymbol{\Lambda}_s$ denote the eigenvector matrix and the eigenvalue matrix, respectively, with

$$\boldsymbol{\Lambda}_s = \operatorname{diag}\left(\lambda_s^{(1)} \quad \ldots \quad \lambda_s^{(\mathcal{Q})}\right). \tag{20}$$

The fourth equality (19) again follows from the fact that $\boldsymbol{C}_{ss}$ is positive semi-definite.

The numerical properties of $\boldsymbol{C}_{\tilde{s}\tilde{s}}$ are determined by the correlation between the source signals and the spatial distribution of the plane wave directions. While the covariance matrix of a single plane wave will be rank-one, and exhibits a wide spread in eigenvalues, the covariance matrix for a larger number of isotropic and uncorrelated plane waves with equal power will be full-rank and have equal eigenvalues [8]. In the following, we investigate the influence of source signals and source distribution on the resulting covariance matrix.

### 3.1 Source covariance matrix

In the case of an ideally diffuse field, the source directions are isotropic, and the source signals are uncorrelated and have equal power. The source signal covariance matrix $\boldsymbol{C}_{ss}$, cf. (18), thus will be an identity matrix scaled by the signal power, and the eigenvalues will be equal, i.e., $\lambda_s^{(1)} = \ldots = \lambda_s^{(\mathcal{Q})} = \lambda_s$.

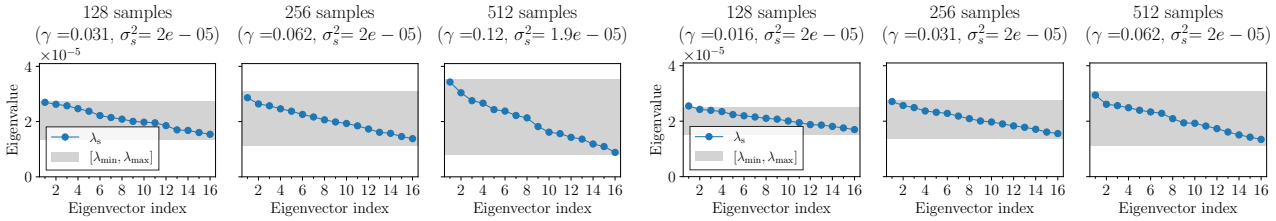
In practice, however, the covariance matrix is commonly obtained by approximating the statistical expectation $\mathbb{E}(\cdot)$ by a sample average under the ergodicity assumption [23],

$$\boldsymbol{C}_{ss} \approx \frac{1}{\mathcal{J}-1}\sum_{j=0}^{\mathcal{J}-1}\left(\boldsymbol{s}(j,\kappa) - \boldsymbol{\mu}_s(\kappa)\right)\left(\boldsymbol{s}(j,\kappa) - \boldsymbol{\mu}_s(\kappa)\right)^{\mathrm{H}}, \tag{21}$$

where the entries of the vector $[\boldsymbol{s}(j,\kappa)]_q = S_q(j,\kappa)$ are the short-time Fourier spectra of the source signals at time step $j$ and frequency index $\kappa$. The expected value of the spectrum at each frequency is also approximated by the sample mean,

$$\boldsymbol{\mu}_s(\kappa) = \frac{1}{\mathcal{J}}\sum_{j=0}^{\mathcal{J}-1}\boldsymbol{s}(j,\kappa). \tag{22}$$

Since the estimated source signal covariance matrix (21) is based on a short-time spectral analysis, the covariance




(a) Rectangular window (no overlap)

(b) Hanning window (50 % overlap)

**Figure 1**: Eigenvalues (filled circles) of the source signal covariance matrices $\boldsymbol{C}_{ss}$ for 16 uncorrelated white noise signals $\sim \mathcal{N}(0,1)$ with length of $2^{16}$ samples ($\approx$ 1.37 s) at a sampling frequency of $f_s = 48$ kHz. Different frame lengths (128, 256, and 512 samples) and analysis windows are considered: (a) rectangular window without overlap and (b) Hanning window with 50 % overlap. The covariance matrix is computed from the short-time Fourier transforms, cf. (21). The results for 1 kHz are shown. The shaded area depicts the range of the eigenvalues $[\lambda_{\text{min}}, \lambda_{\text{max}}]$, estimated by the Marchenko-Pastur law (23). The ratio $\gamma = \frac{\mathcal{Q}}{\mathcal{J}}$ and the variance of the short-time spectra $\sigma_s^2$ are shown on the top of each figure.

of any signal pair is generally non-zero and the eigenvalues are no longer equal. Moreover, the distribution of the eigenvalues is largely dependent on the short-time Fourier analysis parameters, such as the frame length and the number of frames for averaging.

The Marchenko-Pastur law describes the properties of the eigenvalues of a covariance matrix for signals with independent and identical distributions [24]. Let us denote the power of each signal at a given frequency bin as $\sigma_s^2$, and the ratio between the number of sources $\mathcal{Q}$ and the number of samples $\mathcal{J}$ for averaging, cf. (21), as $\gamma = \frac{\mathcal{Q}}{\mathcal{J}}$. The empirical eigenvalues are known to follow the Marchenko-Pastur distribution as $\mathcal{Q}, \mathcal{J} \to \infty$. The mean and variance of the distribution are $\sigma_s^2$ and $\sigma_s^4 \cdot \gamma$, respectively. Additionally, the eigenvalues cluster within a finite support,

$$\lambda_s^{(q)} \in [\sigma_s^2 \cdot (1 - \sqrt{\gamma})^2, \sigma_s^2 \cdot (1 + \sqrt{\gamma}^2)]. \qquad (23)$$

As the sample size (number of frames) increases ($\mathcal{J} \to \infty$), the variance of the distribution tends to zero and the eigenvalue distribution becomes asymptotically uniform.

Numerical simulations are performed for 16 uncorrelated noise signals. The signals are generated in the time domain, from a normal distribution of zero mean and unit variance $\mathcal{N}(0,1)$. The length of the signals are $2^{16}$ samples ($\approx$ 1.37 s) at a sampling frequency of $f_s = 48$ kHz. A short-time Fourier transform was performed with varying parameters:

- Frame length: 128, 256, 512 samples
- Rectangular window (0 % overlap) and Hann window (50 % overlap)

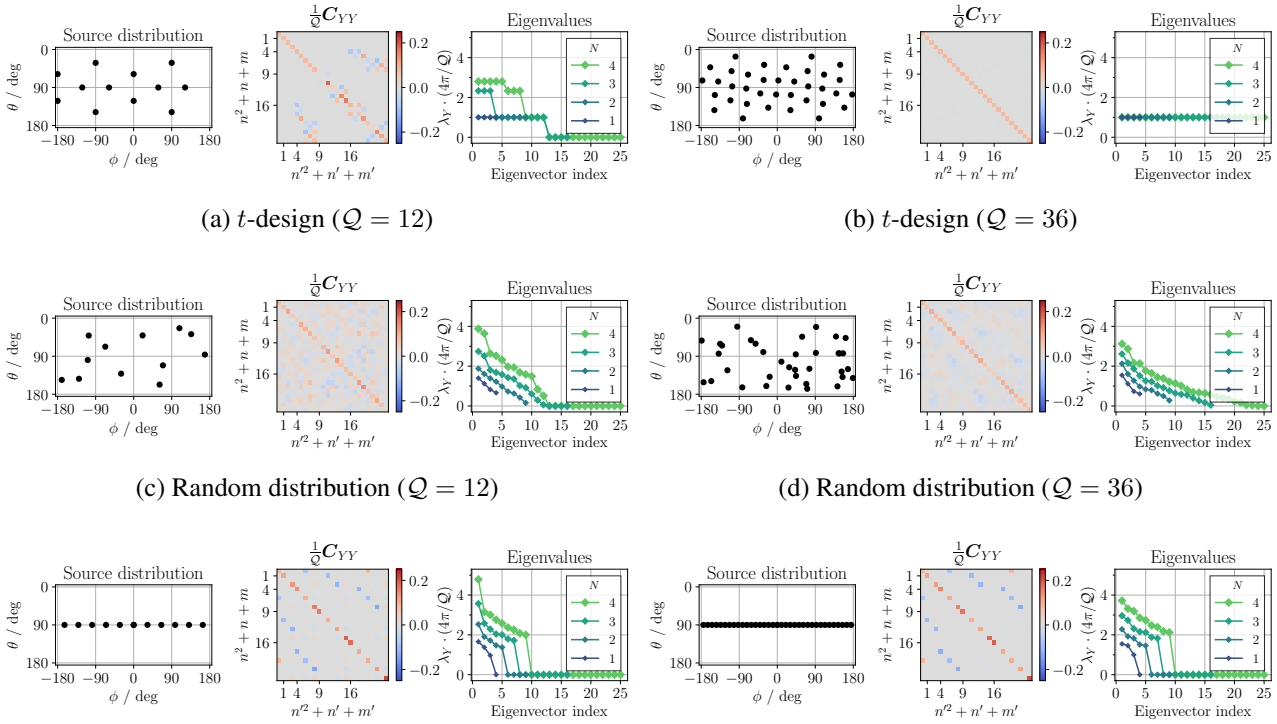
The FFT length is fixed to 1536 samples in order to have the same frequency resolution. The frame length and the overlap length determine the number of frames $\mathcal{J}$ and thus the ratio $\gamma = \frac{\mathcal{Q}}{\mathcal{J}}$.

The eigenvalues of the empirical source covariance matrix at the frequency of 1 kHz are shown in Fig. 1. The results for other frequencies are similar in nature, and are not shown here. It can be seen that the magnitude and the spread of the eigenvalues are dependent of the short-time Fourier analysis parameters. The eigenvalue range predicted by the Marchenko-Pastur law (23) is indicated by shaded gray area. The results are in good agreement with the predictions irrespective of the frame length, overlap length, and the window type.

The lower and upper bounds of the eigenvalues can potentially be used to predict the correlation of the source signals for given short-time Fourier domain signals. The more the signals are correlated, the more eigenvalues will lie beyond the limits. This is particularly useful when the covariance matrix needs to be estimated within a shorter time frame which typically leads to a wider spread in eigenvalues and the uniform assumption does not hold.

### 3.2 Spatial covariance matrix

The distribution of source positions affects the B-format covariance matrix $\boldsymbol{C}_{\bar{s}\bar{s}}$ through the spatial encoding matrix $\boldsymbol{Y}$, cf. (5). In this section, we investigate the proper-




(a) $t$-design ($\mathcal{Q} = 12$)

(b) $t$-design ($\mathcal{Q} = 36$)

(c) Random distribution ($\mathcal{Q} = 12$)

(d) Random distribution ($\mathcal{Q} = 36$)

(e) Circular distribution ($\mathcal{Q} = 12$)

(f) Circular distribution ($\mathcal{Q} = 36$)

**Figure 2**: Spatial covariance matrices and their eigenvalues for different source distributions: (a)–(b) Nearly uniform distributions using $t$-design; (c)–(d) Random distributions; (e)–(f) Uniform circular distribution on the $xy$-plane. For each distribution type, 12 and 36 source positions are generated. Spatial encoding matrices with different Ambisonic orders $N = 1, 2, 3, 4$ are considered. The spatial covariance matrices are scaled with $\frac{1}{\mathcal{Q}}$ and the eigenvalues with $\frac{4\pi}{\mathcal{Q}}$. The eigenvalues are sorted in descending order.

ties of the spatial covariance matrix, defined as $\boldsymbol{C}_{YY} = \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}$. Note that the model-based encoding (5) is independent to frequency. The presented results thus align with the observations in [8], where time-domain signals were considered.

The spatial covariance matrix is diagonalised by using the eigenvalue decomposition as

$$\boldsymbol{C}_{YY} = \boldsymbol{Q}_Y \boldsymbol{\Lambda}_Y \boldsymbol{Q}_Y^{\mathrm{H}}, \tag{24}$$

where $\boldsymbol{Q}_Y$ and $\boldsymbol{\Lambda}_Y$ denote the eigenvector matrix and the eigenvalue matrix, respectively. The eigenvalues are denoted by $\lambda_Y^{(\eta)}$ for $\eta = 1, \ldots, \mathrm{rank}(\boldsymbol{C}_{YY})$. The rank of $\boldsymbol{C}_{YY}$, i.e. the number of non-zero eigenvalues, is determined by the number of source directions $\mathcal{Q}$, the spatial distribution of the sources, and the number of B-format channels $\mathcal{L}$.

The matrix $\boldsymbol{C}_{YY}$ resembles the orthogonality matrix for spatial sampling schemes on a sphere [5, Ch. 3]. The entries of the matrix can be regarded as the numerical integration of two spherical harmonics sampled at the source directions with unit quadrature weights,

$$\left[\boldsymbol{C}_{YY}\right]_{\ell\ell'} = \sum_{q=1}^{\mathcal{Q}} Y_{nm}(\hat{\boldsymbol{k}}_q)\, Y_{n'm'}(\hat{\boldsymbol{k}}_q), \tag{25}$$

where $\ell = n^2 + n + m + 1$ and $\ell' = n'^2 + n' + m' + 1$ are the row and column indices, respectively. A uniform (or nearly uniform) distribution can achieve orthogonality of spherical harmonic pairs up to a given order. The covariance matrix then reads $\boldsymbol{C}_{YY} = \frac{\mathcal{Q}}{4\pi}\boldsymbol{I}$ with equal eigenvalues of $\lambda_Y^{(\eta)} = \frac{\mathcal{Q}}{4\pi}$, $\forall \eta$.

The structure of the spatial covariance matrices and

the eigenvalues are depicted in Fig. 2 for different source distributions:

- Spatial distribution type: $t$-design [25], circular distribution on the $xy$-plane, and random distribution
- Number of sources $\mathcal{Q} = 12, 36$ (corresponding to degrees 5 and 8 for the $t$-design)
- Ambisonic order $N = 1, 2, 3, 4$ (corresponding to B-format channels of $\mathcal{L} = 4, 9, 16, 25$)

For a given source distribution, the entries of the spatial covariance matrix are uniquely determined independent to the Ambisonic order $N$. The covariance matrix for lower orders (e.g. $N = 1, 2, 3$) are sub-matrices of the covariance matrix for $N = 4$. However, the resulting eigenvalues generally change with the Ambisonic order.

For nearly uniform source distributions, Figs. 2(a) and 2(b), the covariance matrices are diagonal up to $l = 9$ and $l = 25$ corresponding to Ambisonic orders $N = 2$ and $N = 4$, respectively. The eigenvalues have discrete values and exhibit structured distributions. The $t$-design with degree 8 ensures the orthogonality of the spherical harmonics up to order $N = 4$. The spatial covariance matrix is thus a diagonal matrix, and all the eigenvalues are equal to $\frac{4\pi}{\mathcal{Q}}$ irrespective of the Ambisonic order $N$, cf. Fig. 2(b). For the $t$-design of degree 5, this is only the case up to $N = 2$. At higher orders, e.g. $N = 3, 4$, the covariance matrix exhibit off-diagonal entries, resulting in unequal eigenvalues.

For randomly distributed source directions, Figs. 2(c) and 2(d), the covariance matrices exhibit non-zero off-diagonal entries for all considered spherical harmonics. No clear structured distribution can be observed in the eigenvalues. The rank of the covariance matrices stays the same as for the nearly uniform distributions. However, the spread of the eigenvalues is considerably larger.

When the sources have a circular distribution, Figs. 2(e) and 2(f), the covariance matrices have rows and columns that have only zero entries. These correspond to the spherical harmonics that have nulls on the $xy$-plane. This results in a rank deficiency of the matrix, as can be seen by the number of non-zero eigenvalues. For both $\mathcal{Q} = 12$ and 16 sources, the rank of the covariance matrix is $2N + 1$. This corresponds to the number of B-format channels in 2-dimensional Ambisonics [4].

It is worth mentioning that the normalisation of the spherical harmonics, e.g. N3D or SN3D [26, Ch. 3], has a significant impact on the distribution of the eigenvalues. If the SN3D normalisation is used, the difference between the diagonal entries of $\boldsymbol{C}_{YY}$ increases and eigenvalues have a wider spread (not shown) compared to the N3D normalisation. Since the purpose of the current work is to examine the spatial isotropy of the sound field, we employ the N3D normalisation, which yields a tighter distribution of the eigenvalues for isotropic source distributions.

## 4. REPRODUCED SOUND FIELDS

In this section, we consider sound reproduction scenarios, where multiple loudspeakers produce uncorrelated signals. Three different loudspeaker layouts are considered that are compliant with ITU-R BS.2051-3 [27]:

- 5-channel (0+5+0)
- 9-channel (4+5+0)
- 22-channel (9+10+3)

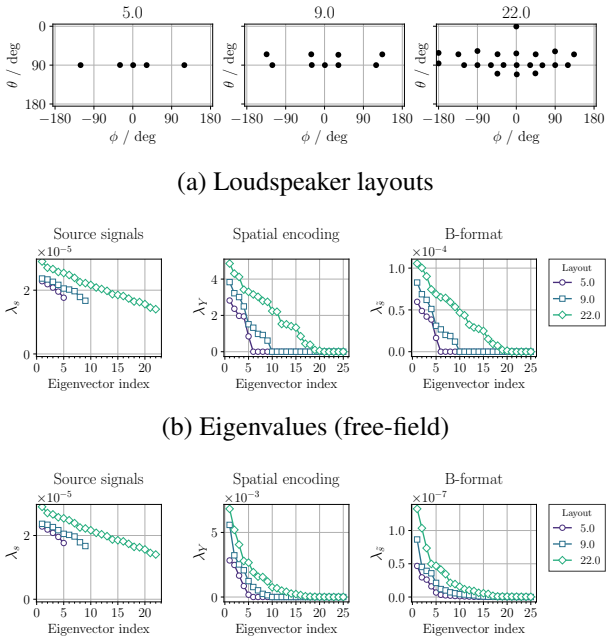
The digits in parentheses indicate the number of loudspeakers distributed across three vertical layers (upper + middle + lower). The exact positions of the loudspeakers follow the layout of the system installed in the Audio Laboratory at the University of Southampton, cf. Fig. 3(a). For each configuration, the loudspeakers are driven with uncorrelated white noise signals. The signals are generated in the time domain from a normal distribution with zero mean and unit variance. The length of the signals is $2^{16}$ samples for a sampling frequency of $f_s = 48\,\mathrm{kHz}$.

Fourth-order Ambisonic encoding is performed both for simulated and real sound fields. A free-field is assumed for the simulated case and the Ambisonic signals are obtained by using (5). The angular positions (azimuth and colatitude) of the loudspeakers are set as the plane wave directions. For the real sound field, a data-based encoding is carried out. The room impulse responses measured by a rigid spherical microphone array (mh acoustics em32) are converted into B-format signals as described by (9). A regularised radial filter (maximum magnitude boost of 30 dB) is used to equalise the effect of the microphone array [28]. Finally, a short-time Fourier analysis of the B-format signals is performed with frame length of 256 samples, Hann window with the same length, overlap length of 128 samples, and FFT length of 1536 samples. The results for 2 kHz are presented in this section.

The covariance matrices are computed for the source signals, spatial encoding, and the B-format signals as described in Sec, 2. The respective eigenvalues (denoted by $\lambda_s$, $\lambda_Y$, and $\lambda_{\tilde{s}}$) for simulated and real sound fields

are shown in Figs. 3(b) and 3(c), respectively. It can be seen that the eigenvalues of the source signal covariance matrices follow similar distribution as observed in Fig. 1. Since the same set of uncorrelated source signals are used for simulated and real cases, the eigenvalues of the signal covariance matrices are identical. The eigenvalues of the spatial covariance matrices exhibit properties somewhat between nearly uniform and random distributions, cf. Fig. 2. The rank of $\boldsymbol{C}_{YY}$ are the same as the number of loudspeakers (i.e. 5, 9, and 22). The condition number for the 22-channel layout is considerably larger than the other layouts. This is presumably due to the dense arrangement of the sources on the horizontal plane, cf. Fig. 3(a), which almost leads to a rank deficiency similar to the circular distributions, cf. Figs. 2(e) and 2(f). Incorporating the room acoustics affects the eigenvalue distribution significantly.

For the considered scenarios, the eigenvalue distribution of the B-format signals is dominated by that of the spatial encoding matrix. This is primarily because the source signals are uncorrelated and the eigenvalues of the source covariance matrix exhibit minimal spread. The impact of the source signals could be demonstrated by increasing the correlation of the source signals, which would increase the spread of the eigenvalues $\lambda_s$. Further investigations for more diverse scenarios are left as future work.




(a) Loudspeaker layouts

(b) Eigenvalues (free-field)

(c) Eigenvalues (room)

**Figure 3**: Eigenvalue analysis for reproduced sound fields. The considered loudspeaker layouts are shown in (a). The eigenvalue distributions are shown for (b) simulated free-field and (c) real sound field.

## 5. CONCLUSION

A cross-spectral analysis was performed on higher-order Ambisonic signals. The numerical properties of the B-format covariance matrices are determined by the statistical relationship between the source signals and the spatial distribution of the source positions. The signal-dependent and spatially-dependent parts were explored in terms of the eigenvalues of the respective covariance matrices. It was shown that, the eigenvalue distribution of the source signal covariance matrix is very sensitive to short-time Fourier analysis parameters. Nevertheless, by exploiting the Marchenko-Pastur law, it is possible to reliably predict the spread of the eigenvalues in the case of uncorrelated signals. The influence of the spatial isotropy is represented by the spatial covariance matrix. The distribution of the corresponding eigenvalues is found to be dependent on the number of sources and the Ambisonic order. It was also demonstrated that certain source arrangements cause rank deficiency of the spatial covariance matrix.

The presented work is still exploratory in nature, but bears the potential to be extended towards the estimation of sound field's diffuseness, which is still an on-going research topic. It is of considerable interest to enable the prediction of associated perceptual attributes such as listener envelopment. The relationship between time integration of the auditory processing and the short-time Fourier analysis can be further explored. Incorporating the directional-dependent resolution of human spatial hearing into the model is also a possible extension of the current study.

## 6. REFERENCES

[1] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J. Acoust. Soc. Am. (JASA)*, vol. 132, no. 1, pp. 261–270, 2012.

[2] M. Nolan, M. Berzborn, and E. Fernandez-Grande, "Isotropy in decaying reverberant sound fields," *J. Acoust. Soc. Am. (JASA)*, vol. 148, no. 2, pp. 1077–1088, 2020.

[3] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer, 2001.

[4] F. Zotter and M. Frank, *Ambisonics*. Springer, 2019.

[5] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015.

[6] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[7] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, vol. 4, (Piscataway, NJ, USA), Mar. 2005.

[8] N. Epain and C. T. Jin, "Spherical harmonic signal covariance and sound field diffuseness," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 10, pp. 1796–1807, 2016.

[9] P. Massé, *Analysis, treatment, and manipulation methods for spatial room impulse responses measured with spherical microphone arrays*. PhD thesis, Sorbonne Université, Paris, France, 2022.

[10] A. Herzog, *Advances in Spatial Parameter Estimation and Signal Enhancement Using Higher-Order Ambisonics*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nuremberg, Germany, 2023.

[11] V. Pulkki, S. Delikaris-Manias, and A. Politis, *Parametric time-frequency domain spatial audio*. Wiley, 2018.

[12] L. Cremer, H. A. Müller, and T. J. Schultz, *Principles and applications of room acoustics*. 1982.

[13] L. Beranek, *Concert and opera halls: How they sound*. American Institute of Physics, 1996.

[14] H. Kuttruff, *Room acoustics*. Spon Press, 2000.

[15] C.-H. Jeong, "Diffuse sound field: challenges and misconceptions," in *Proc. 45th Int. Congr. Expo. Noise Control Eng.*, pp. 1015–1021, Aug. 2016.

[16] G. S. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Comput. Music J.*, vol. 19, no. 4, pp. 71–87, 1995.

[17] F. Zotter, M. Frank, M. Kronlachner, and J.-W. Choi, "Efficient phantom source widening and diffuseness in ambisonics," in *Proc. EAA Joint Symp. Auralization and Ambisonics*, (Berlin, Germany), Apr. 2014.

[18] S. J. Schlecht, "Frequency-dependent schroeder allpass filters," *Appl. Sci.*, vol. 10, no. 1, p. 187, 2019.

[19] A. Favrot, F. Zotter, and C. Faller, "Covariance in non-coincident tetrahedral arrays," in *Proc. Audio Eng. Soc. (AES) Conv.*, (Madrid, Spain), 2024.

[20] G. W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays: Signal Processing Techniques and Applications*, pp. 61–85, Springer, 2001.

[21] A. Politis, "Diffuse-field coherence of sensors with arbitrary directional responses," *arXiv preprint arXiv:1608.07713*, 2016.

[22] G. H. Golub and C. F. Van Loan, *Matrix computations*. The Johns Hopkins University Press, 1996.

[23] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall, 1993.

[24] V. A. Marchenko and L. A. Pastur, "Distribution of eigenvalues in certain ensembles of random matrices," *Mathematicheskii Sbornik*, vol. 72, no. 4, pp. 507–536, 1967. (in Russian).

[25] R. H. Hardin and N. J. Sloane, "McLaren's improved snub cube and other new spherical designs in three dimensions," *Discrete Comput. Geometry*, vol. 15, pp. 429–441, 1996.

[26] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, University of Paris VI, Paris, France, 2000. (in French).

[27] ITU, "ITU-R BS.2051-3: Advanced sound system for programme production," May 2022.

[28] F. Zotter, "A linear-phase filter-bank approach to process rigid spherical microphone array recordings," (Palić, Serbia), June 2018.