Scattering Delay Network Simulator of Coupled Volume Acoustics

Timuçin Berk Atalay, Zühre Sü Gül, Enzo De Sena, *Senior Member, IEEE*, Zoran Cvetković, *Senior Member, IEEE*, Hüseyin Hacıhabiboğlu, *Senior Member, IEEE*

***Abstract*—Artificial reverberators provide a computationally viable alternative to full-scale room acoustics simulation methods for deployment in interactive, immersive systems. Scattering delay network (SDN) is an artificial reverberator that allows direct parametric control over the geometry of a simulated cuboid enclosure, as well as the directional characteristics of the simulated sound sources and microphones. This paper extends the concept of SDN reverberators to multiple enclosures coupled via an aperture. The extension allows independent control of the acoustical properties of the coupled enclosures and the size of the connecting aperture. Transfer functions of the coupled-volume SDN are derived. The effectiveness of the proposed method is evaluated in terms of rendered energy decay curves in comparison to full-scale ray-tracing models and scale model measurements.**

***Index Terms*—Artificial reverberators, room acoustics simulation, delay networks, coupled volume acoustics, double-sloped decay**

I. Introduction

A number of room acoustic models have been proposed over the past six decades [1], [2]. Wave-based models such as finite-difference time domain (FDTD) [3], digital waveguide meshes (DWM) [4], and finite element methods (FEM) [5] provide the highest physical accuracy but carry a significant computational cost. Recent advancements allowed real-time operation on GPU platforms, but this is still costly and typically limited in terms of operating bandwidth and room sizes [6]–[8]. Wave-based models thus remain largely unsuitable for interactive real-time applications, especially in cases where most computational resources are allocated to video rendering, e.g. in video gaming or AR/VR applications.

Geometrical-acoustics methods, such as ray-tracing [9], beam-tracing [10], and the image method [11] approximate wavefronts using acoustic rays, resulting in significant computational complexity reductions. However, these models generate an impulse response and thus still involve a convolution, which even on its own is a costly operation, especially in the presence of multiple sources and stereo or multichannel outputs.

In the context of applications with tight constraints regarding computational cost, a typical approach is to render accurately only the direct path and early specular reflections, whilst generating late reverberation using low-complexity artificial reverberators [1], [2], [12] such as feedback delay networks (FDN) [13]–[16] or digital waveguide networks (DWN) [17], [18]. FDNs and DWNs have been shown to exhibit certain isomorphism relationships [19], and both are recursive digital filters with an extremely low computational complexity.

While FDNs as proposed by Jot and Chaigne aimed at generating a natural-sounding reverberation with given acoustical properties e.g. reverberation time, significant work has gone in the direction of linking FDNs more explicitly to physical properties of room acoustics [16], [19]–[23]. Most recently, Alary *et al.* [21] extended FDNs to model non-uniform direction-dependent decay time. Schlecht and Habets [16] proposed methods for the optimal design of feedback matrices in spatialised FDNs. Wendt *et al.* [22] proposed a spatialised binaural simulator based on an FDN/image-source-method [11] hybrid taking into account spatial properties of the room. Particularly relevant in the context of this paper is the work of Das and Abel [23], who recently modified FDNs to model coupled volumes.

Scattering delay network (SDN) is an artificial reverberator that combines elements of geometric room acoustics with computationally efficient delay networks designed to provide spatialised reverberation corresponding to a given room geometry, wall absorption characteristics, and source/observer positions [24]–[26]. Although a similar reverberator has been proposed previously in a simpler form [27], SDN's difference and contribution is that it can accurately provide the perceptually most important features of the room response, such as the direct path and first-order reflections, along with a gracefully degrading accuracy of higher-order reflections. SDN, when coupled with HRTF personalization, was shown to provide subjective localization and externalization performance on par with room simulation using measured higher-order Ambisonics acoustic impulse responses [28]. Moreover, SDN was shown to provide levels of naturalness and pleasantness that are comparable to full-scale room acoustic simulators [29]. These properties make it particularly suitable for immersive interactive applications, as explored recently in the context of augmented reality [30]–[33].

An important limitation of the original SDN model is that

Manuscript received December 22, 2021.

The work reported in this paper is supported by the Turkish Scientific and Technological Research Council (TUBITAK) Research Grant 119E254 "Audio Signal Processing for Six Degrees of Freedom (6DOF) Immersive Media" (2019-2021) and in part by EPSRC grant EP/V002554/1. Timucin Berk Atalay and Hüseyin Hacıhabiboğlu are with METU Spatial Audio Research Group (SPARG), Middle East Technical University, Graduate School of Informatics (e- mail: e173621@metu.edu.tr, hhuseyin@metu.edu.tr). Zühre Sü Gül is with Department of Architecture, Bilkent University (e-mail: zuhre@bilkent.edu.tr). Enzo De Sena is with Institute of Sound Recording (IOSR), University of Surrey (e-mail: e.desena@surrey.ac.uk). Zoran Cvetković is with Department of Engineering, King's College London (e-mail: zoran.cvetkovic@kcl.ac.uk).

Digital Object Identifier 10.1109/TASL.2021.XYZXYZ

it is restricted to cuboid rooms, whilst extensions to more complex geometries, such as coupled spaces that are connected through an aperture such as a door or a window, have not been addressed yet. Coupled rooms exhibit distinctive energy decay curves (EDCs) characterized by multi-rate profiles, that depend on the decay characteristics of the individual volumes, as well as the size of the aperture between them [34]. Coupled volumes are prominent architectural design elements in concert halls and opera houses [35] and they also frequently appear in virtual immersive environments. Hence, extending the SDN framework to coupled spaces is relevant in a range of domains.

In this paper, we propose the coupled-volume SDN (CV-SDN) as an extension to the original SDN concept that allows the simulation of the acoustics of coupled volumes and a parametric control of the size of the aperture between them. We present an objective evaluation of the CV-SDN reverberators in comparison with room acoustics simulation using ODEON, a state-of-the-art ray tracing based geometric room acoustics simulator, as well as measurements made using a scale model.

The paper is organized as follows. Relevant aspects of sound energy decay in coupled rooms and the original SDN concept are reviewed in Sec. II. In Sec. III, the concept of the proposed CV-SDN reverberator is introduced and its transfer functions is analyzed in detail. Room impulse responses and their acoustic parameters obtained using the proposed CV-SDN model are compared to those obtained from a scale model and the ray tracing software in Sec. IV. In Sec. V, the computational complexity of the CV-SDN is examined and compared to overlap-add convolution. Finally, Sec. VI concludes the paper.

II. Background

A. Sound Energy Decay in Coupled Rooms

A coupled-volume system is defined as two or more spaces conjoined via a common, acoustically transparent surface, known as the *coupling aperture*. If the sound level decays at different rates in the coupled spaces, there will be excess energy in one of the spaces in comparison with the other. This leads to an energy transfer from the room with the surplus energy to the room with the faster energy decay. As a result, the room impulse responses measured at receiver positions inside the less reverberant room exhibit a multi-rate sound energy decay [36]. This effect can be observed in various venues such as concert halls that incorporate reverberation chambers, factories, office spaces, atria, and places of worship. Designers are attracted by the coupled volume concept in concert halls because it presents a compromise between the competing acoustic conditions for both spaciousness and clarity.

Research into acoustics of coupled volume spaces resulted in models rooted in statistical theory [37], [38], wave theory [39]–[42], statistical energy analysis [43]–[45], ray-based geometrical acoustics [46]–[48], diffusion equation model [49]–[51] and surface coupling [52]. The effects of source and receiver positions, the architectural variables such as aperture size, position, shape as well as the volume and absorption characteristics of individual volumes concerning the degree of acoustical coupling have also been investigated in different studies [23], [53]–[57]. Perception of coupled room acoustics has also been studied and findings suggest that the just noticable difference in aperture size is approximately 10% [58].

B. Decay Parameter Estimations

Quantifying the degree of room acoustics coupling and relevant energy decay parameters has been a persistent challenge and various metrics have been proposed. In the validation of the proposed coupled-volume SDN model, we employ a multiple-decay-rate model selection approach based on Bayesian information criterion to select the most parsimonious model. This approach is an efficient validated method for estimating the key characteristics of multiple-slope sound energy decays [59].

For a discrete-time room impulse response $h(n)$, $0 \leq n \leq K$, where K is sufficiently large, the normalised tail energy is defined as:

$$\mathcal{D}(k) = \sum_{n=k}^{K} |h(n)|^2 / \sum_{n=0}^{K} |h(n)|^2, \tag{1}$$

where $k = 0 \cdots K$ is a sample index representing the lag from the start of the impulse response sequence. Schroeder's decay curve, also known as energy decay curve (EDC), defined as the normalised tail energy of a discrete-time room impulse response expressed in the logarithmic scale,

$$\mathcal{E}(k) = 10 \log_{10} \mathcal{D}(k) \text{ (dB)} \tag{2}$$

is of central importance in the analysis of the decay characteristics of sound in rooms.

For a single-volume room with a simple shape and major reflecting surfaces having substantially similar absorptive characteristics, the impulse response decays exponentially [60], and hence the corresponding EDC is substantially a linearly decreasing sequence. The reverberation time, RT_{60}, can be calculated directly from the EDC as the time at which the total cumulative energy decays to 60 dB below its initial value. Typically, the noise floor in measured room impulse responses prevents a reliable direct calculation of RT_{60}. Instead, the reverberation time is estimated using the slope of the EDC curve as:

$$\mathrm{T}_L = \frac{-60T}{L_2 - L_1}(n_{L_2} - n_{L_1}) \tag{3}$$

where n_{L_i} is the time index such that $\mathcal{E}(n_{L_i}) \approx L_i$, $L = L_2 - L_1$, and T is the sampling period. Typically, $L_2 = -25$, and selecting $L_1 = -5$ gives T_{20} whereas selecting $L_2 = -35$ gives T_{30} which are both estimates of RT_{60}. The decision on which of these two estimates to use in a specific case depends on the relative level of the noise floor.

In a system of coupled volumes, each volume has its distinct exponential decay. When combined, these rooms form a decay characteristic that is typically a weighted sum of exponential components that overall exhibit either a single or a multi-rate decay depending on the acoustical properties of the individual volumes. Hence, a system of coupled volumes requires a different approach than a single volume to characterize its acoustics.

An approach based on a Bayesian analysis of the Schroeder decay curve has been shown in several studies to be an effective and reliable method for characterizing sound energy decay in coupled spaces [59], [61]–[63]. The decay profile of a coupled-volume system is characterized by several decay profiles and their ordinate intercepts. Bayesian model-based parameter estimation aims to determine the decay profiles by selecting the most parsimonious multi-rate model that fits the decay curve [59][1].

A parametric decay profile model describing the normalized tail energy, $\beta(\mathbf{A}_S, \mathbf{T}_S, k)$ is given as:

$$\mathcal{D}(k) = \beta(\mathbf{A}_S, \mathbf{T}_S, k) + \eta(k) \tag{4}$$

where $\eta(k)$ is a sample-wise deviation of measured data from the model. This parametric model is defined as:

$$\begin{aligned}\beta_S &\triangleq \beta(\mathbf{A}_S, \mathbf{T}_S, k) \\ &= A_0(K-k) + \sum_{s=1}^{S} A_s \left[e^{\frac{-13.8kT}{T_s}} - e^{\frac{-13.8KT}{T_s}} \right]\end{aligned} \tag{5}$$

where S is the number of exponential decay terms, also termed the *decay order*, the vectors $\mathbf{A}_\mathrm{S} = [A_0,\ A_1, \cdots A_\mathrm{S}]^T$ and $\mathbf{T}_\mathrm{S} = [T_1,\ T_2, \cdots T_\mathrm{S}]^T$ are the amplitudes and decay times associated with the individual exponential decay terms, and T is the sampling period. The first term on the right hand side, $A_0(K-k)$, represents the background noise. The vectors, $\mathbf{A}_\mathrm{S}$ and $\mathbf{T}_\mathrm{S}$ in combination, comprise the parameters of the decay profile to be estimated. The measured values of the normalized tail energy in (1) can be cast as a $(K+1)\times 1$ vector $\mathbf{d} = [\mathcal{D}(0), \mathcal{D}(1), \cdots \mathcal{D}(K)]^T$.

A decay profile model can be fit to measured data by calculating the decay parameter vectors that maximize the likelihood:

$$\mathcal{L}\left(\mathbf{A}, \mathbf{T}\right) = p(\mathbf{d} \mid \mathbf{A}, \mathbf{T}, \beta_\mathrm{S}, I). \tag{6}$$

where $p(\cdot)$ is a multidimensional Gaussian distribution as described in [63], and I represents the prior probability. For a given decay order, $\mathrm{S} = 1 \cdots \mathrm{S}_{\max}$, where $\mathrm{S}_{\max}$ is the maximum decay order to be considered, the decay parameter vectors that specify the decay profile are selected as those that maximize the likelihood such that $\{\hat{\mathbf{A}}_\mathrm{S}, \hat{\mathbf{T}}_\mathrm{S}\} = \operatorname{argmax} \mathcal{L}\left(\mathbf{A}_\mathrm{S}, \mathbf{T}_\mathrm{S}\right)$.

Once models with different orders are obtained this way, the model that minimizes the Bayesian information criterion (BIC):

$$\mathrm{BIC}(\mathrm{S}) = (2\mathrm{S}+1)\ln K - 2\ln \mathcal{L}(\hat{\mathbf{A}}_\mathrm{S}, \hat{\mathbf{T}}_\mathrm{S}) \tag{7}$$

is selected as the most parsimonious model. Notice that BIC promotes parsimony by penalizing higher order models. Further details of the Bayesian approach used in the estimation of the decay profile is beyond the scope of this paper; interested readers are referred to [59].

[1]While it might have been possible to fit a parametric curve to the EDC, the employed decay parameter estimation approach is guaranteed to provide the model with the lowest order, and is therefore preferred.

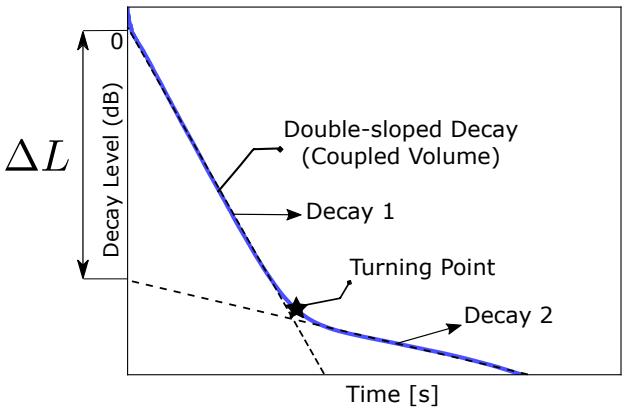

Fig. 1. Diagram showing a hypothetical double-slope EDC with a level difference of ΔL. The two decay terms (dashed lines) with the early decay labeled as Decay 1 and late decay as Decay 2. The *turning point* when the slope changes is shown as a point (star).

C. Coupled systems of two volumes

Two rooms coupled via an aperture can, depending on their acoustic properties, have an EDC with either one or more slopes [64]. For the cases where source and microphone are placed in different volumes, with the coupled volume being more reverberant and the microphone to be located in the coupled room, EDC is likely to exhibit a single slope [65]. With a single slope, reverberation time, RT_{60} can be measured to analyze the EDC. The slope is determined by the room that is more reverberant. On the other hand, the setting where the source and the microphone are in the same volume may result in more than one slope. This requires describing the decay profile model and therefore the EDC, using two parameters related to the levels, A_1 and A_2, and decay times, T_1 and T_2 of the constituent exponential components in the parametric model.

The relative amplitude of the two decay components in a double slope decay profile is evaluated by means of their logarithmic ratio, namely level difference, $\Delta L = 10\log_{10}(A_1/A_2)$ (dB). This quantity parameterizes the double-slope characteristics by quantifying the degree of energy feedback [65].

Another parameter related to double-slope decay characteristics is the so-called *turning point*, which specifies the level of the EDC at the lag from the start of the impulse response at which the decay characteristics of the initially non-dominant room starts to dominate the EDC. The knee point on the actual EDC and the intersection point of individual decay elements obtained via Bayesian approach will, in general, not exactly coincide. Therefore, the turning point is calculated as the point on the EDC closest to the intersection point of the two decay lines. We will compare results obtained via CV-SDN with those from a room acoustics simulator and the scale model measurements using the following parameters: ΔL, *turning point*, first and second decay times T_1 and T_2. These parameters are illustrated in Fig. 1. The hypothetical EDC shown in the figure is smoothed out for clearer illustration of the parameters.

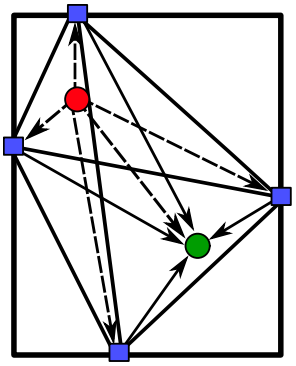

Fig. 2. 2D diagram indicating the network topology of basic SDN. Red and green circles indicate source and microphone nodes. Blue squares indicate wall nodes.

D. Scattering Delay Networks

Scattering delay network (SDN) is a fully connected sparse digital waveguide network (DWN) [66] designed for simulating the sound field of an acoustic source in a cuboid space. This is done by approximating reflective surfaces using scattering junctions positioned at points of incidence of the first order reflections. This construction allows for the exact simulation of first-order specular reflections in terms of their time delays and amplitudes, combined with a gracefully degrading accuracy of the simulation of second- and higher-order reflections. We will henceforth refer to the original SDN model as the basic SDN to differentiate it from the coupled-volume SDN proposed in this paper.

The basic SDN model comprises source and microphone nodes and six scattering nodes, one for each wall, floor and ceiling. Fig. 2 shows a simplified 2D diagram of the basic SDN structure. The positions of the source (red circle) and microphone (green circle) nodes determine the positions of the scattering nodes (blue squares) on the simulated reflecting surfaces. Dashed lines represent the unidirectional connections originating from the sound source. Solid black lines between nodes represent the bidirectional connections between scattering nodes, which carry out lossless scattering operation followed by a filtering (or a frequency independent gain) operation to model wall absorption. Finally, solid black lines extending out from the scattering nodes to the microphone node are unidirectional connections. Basic SDN was shown to exhibit echo build up density profiles and reverberation times very close to the image method (IM) [11] and in agreement with Sabine [67] and Norris-Eyring [68] equations.

Unlike perceptually motivated reverberators such as Jot's reverberator [69], one-to-one correspondence with the room geometry allows a trivial selection of the delay line lengths in the SDN room simulator to model the propagation delay between two points. Since SDN allows selecting specific absorption characteristics of walls, it affords complete control over the reverberation characteristics of the simulated room. Moreover, SDN can simulate the directional properties of sound sources and microphones as well as their movement with six degrees of freedom.

The transfer function of the basic SDN for a rectangular enclosure has the following form [24]:

$$H_{\text{SDN}}(z) = \bar{g} z^{-D_{\text{sm}}} + \frac{1}{5}\mathbf{k}_{\text{m}}(z)\mathbf{F}(z)\mathbf{k}_{\text{s}}(z) \tag{8}$$

where the individual terms are specified below using the following notation:

- $\mathbf{x}_{\text{s}}$ and $\mathbf{x}_{\text{m}}$ are positions of the source and microphone nodes, respectively,
- $\{\mathbf{x}_i\}_{i=1\cdots 6}$ are the positions of the scattering nodes,
- $\mathbf{\Gamma}_{\text{s}}(\boldsymbol{\theta})$ and $\mathbf{\Gamma}_{\text{m}}(\boldsymbol{\theta})$ are the directivity patterns of the source and microphone, respectively,
- $\boldsymbol{\theta}_{\text{si}}$ and $\boldsymbol{\theta}_{\text{im}}$ are the angles between the acoustic axes of the source node and the microphone node, respectively, and the i-th scattering node,
- $D_{\text{sm}} = \|\mathbf{x}_{\text{s}} - \mathbf{x}_{\text{m}}\| \cdot F_s/c$ is the delay between the source and microphone nodes.
- F_s is the sampling frequency and c is the speed of sound,

The source-model adaptor vector, $\mathbf{k}_{\text{s}}(z)$, and the model-microphone adaptor vector, $\mathbf{k}_{\text{m}}(z)$, are given as:

$$\mathbf{k}_{\text{s}}(z) = \mathbf{G}_{\text{s}}\mathbf{D}_{\text{s}}(z)\boldsymbol{\gamma}_{\text{s}}, \tag{9}$$

$$\mathbf{k}_{\text{m}}(z) = \boldsymbol{\gamma}_{\text{m}}^T\mathbf{D}_{\text{m}}(z)\mathbf{G}_{\text{m}}, \tag{10}$$

where the individual factors are specified below using the Kronecker product of two matrices $\otimes$:

- $\boldsymbol{\gamma}_{\text{s}} = \mathcal{G}_{\text{s}} \otimes \mathbf{1}_{5\times 1}$ with $\mathcal{G}_{\text{s}} = [\mathbf{\Gamma}_{\text{s}}(\boldsymbol{\theta}_{s1}), \cdots, \mathbf{\Gamma}_s(\boldsymbol{\theta}_{s6})]^T$ and $\boldsymbol{\gamma}_{\text{m}} = \mathcal{G}_{\text{m}} \otimes \mathbf{1}_{5\times 1}$ with $\mathcal{G}_{\text{m}} = [\mathbf{\Gamma}_{\text{m}}(\boldsymbol{\theta}_{1\text{m}}), \cdots, \mathbf{\Gamma}_{\text{m}}(\boldsymbol{\theta}_{6\text{m}})]^T$ comprise source to node and node to microphone directivity gains, respectively
- $\mathbf{D}_{\text{s}} = \mathbf{d}_{\text{s}} \otimes \mathbf{I}_{5\times 5}$ where $\mathbf{d}_{\text{s}} = \text{diag}(z^{-D_{\text{s},1}}, \cdots, z^{-D_{\text{s},6}})$ with $D_{\text{s},i} = \|\mathbf{x}_{\text{s}} - \mathbf{x}_i\| \cdot F_s/c$, and $\mathbf{D}_{\text{m}} = \mathbf{d}_{\text{m}} \otimes \mathbf{I}_{5\times 5}$ where $\mathbf{d}_{\text{m}} = \text{diag}(z^{-D_{1,\text{m}}}, \cdots, z^{-D_{6,\text{m}}})$ with $D_{i,\text{m}} = \|\mathbf{x}_i - \mathbf{x}_{\text{m}}\| \cdot F_s/c$ are the source-node and node-microphone delay matrices,
- $\mathbf{G}_{\text{s}} = \mathbf{g}_{\text{s}} \otimes \mathbf{I}_{5\times 5}$ and $\mathbf{G}_{\text{m}} = \mathbf{g}_{\text{m}} \otimes \mathbf{I}_{5\times 5}$ with $\mathbf{g}_{\text{s}} = \text{diag}(g_{s1}, \cdots g_{s6})$ and $\mathbf{g}_{\text{m}} = \text{diag}(g_{1\text{m}}, \cdots g_{6\text{m}})$ are gain matrices that depend on the relative physical positions of the nodes. The gains comprising these matrices emulate the distance attenuation due to spherical spreading and are given as $g_{sk} = (\|\mathbf{x}_{\text{s}} - \mathbf{x}_i\|)^{-1}$ and $g_{k\text{m}} = \left(1 + \frac{\|\mathbf{x}_i - \mathbf{x}_{\text{m}}\|}{\|\mathbf{x}_{\text{s}} - \mathbf{x}_i\|}\right)^{-1}$, product of which gives the correct distance attenuation for the given source and microphone positions.

The component $\mathbf{F}(z)$ has the form

$$\mathbf{F}(z) = [\mathbf{H}(z^{-1}) - \mathbf{\Sigma}\mathbf{P}\mathbf{D}_f(z)]^{-1},$$

where

- $\mathbf{\Sigma}$ is the block diagonal scattering matrix:

$$\mathbf{\Sigma} = \mathbf{I}_{5\times 5} \otimes \mathbf{S}, \tag{11}$$

where

- $\mathbf{P}$ is a permutation matrix specified by the connectivity of the scattering nodes,
- $\mathbf{D}_f(z)$ is the feedback delay matrix:

$$\mathbf{D}_f(z) = \text{diag}(z^{-D_{1,2}}, \cdots, z^{-D_{6,5}}), \tag{12}$$

where $D_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\| \cdot F_s/c$ is the delay between two scattering nodes, i and j, positioned at $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\mathbf{S} = \frac{2}{5}\mathbf{1}_{5\times5} - \mathbf{I}_{5\times5}$ is the scattering matrix. This specific scattering matrix can be shown to be unilossless, i.e. it results in lossless operation regardless of the length of the feedback delay lines [14].

- $\mathbf{H}(z)$ is the reflectance matrix:

$$\mathbf{H}(z) = \mathbf{H}_w(z) \otimes \mathbf{I}_{5\times5}, \tag{13}$$

with $\mathbf{H}_w(z) = \mathrm{diag}(H_1(z), \cdots, H_6(z))$ where subscript w is used to associate the reflectance matrix with the walls of the enclosure and $H_i(z)$ is the reflectance transfer function of the i-th wall determined by the frequency-dependent absorption characteristics of that wall.

Finally, the direct path attenuation is set as:

$$\overline{g} = g_{\mathrm{sm}}\mathbf{\Gamma}_{\mathrm{s}}(\boldsymbol{\theta}_{\mathrm{sm}})\mathbf{\Gamma}_{\mathrm{m}}(\boldsymbol{\theta}_{\mathrm{ms}}), \tag{14}$$

where $g_{\mathrm{sm}} = \|\mathbf{x}_{\mathrm{s}} - \mathbf{x}_{\mathrm{m}}\|^{-1}$ is the direct path gain.

III. Coupled-Volume Acoustics Modelling using Scattering Delay Networks

In this section, the description of coupled-volume SDN model, which is the main contribution of this paper, is given in detail.

A. Coupled-Volume SDN (CV-SDN)

There are two possible settings for CV-SDN that depend on the source and the receiver positions. In the first setting, both the source and the receiver are in the same volume (the *main* volume) that is connected to a second volume (the *coupled* volume). This is referred to as the *same volume setting*. In the second setting, referred to here as the *different volume setting*, the source and the microphone are in different volumes. The former is more challenging from the perspective of room acoustics rendering, since in that case the underlying EDC is more likely to exhibit a double-slope behaviour which is not trivial to simulate.

For CV-SDN, each transmission from a source to a receiving end is modeled with a single SDN network to ensure that early reflections are modeled accurately like basic SDN. Therefore, the proposed CV-SDN model consists of five SDN components in the same volume setting and four SDN components in the different volume setting. The sound energy flow between the two volumes is modelled using the concept of *aperture node*. The aperture node acts either as a source node, a microphone node, or both, depending on the direction of energy transfer in a particular CV-SDN component. The individual SDN components are illustrated in Fig. 3 and their physical meanings are described below.

1) $\boldsymbol{H}^{(1)}_{\mathbf{sm}}(\boldsymbol{z})$: The component renders the sound field of the main volume between the source and the microphone, as illustrated in Fig. 3(a) by the blue scattering nodes. This component only exists for the same volume setting and it is very similar to the basic SDN, apart from a change in the absorption coefficient of the wall that contains the aperture node, as will be explained in the next section.

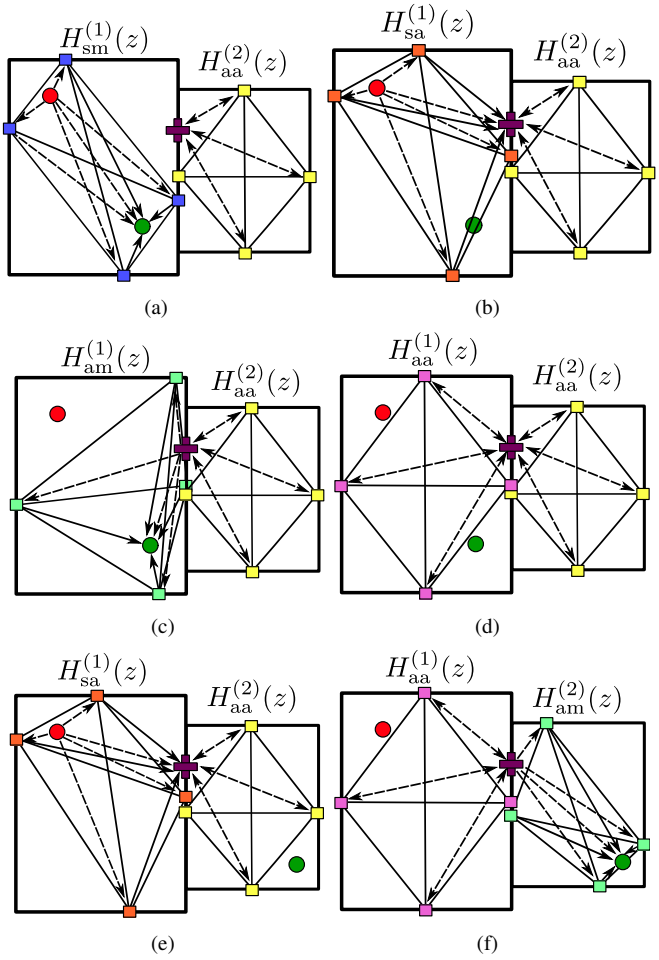

Fig. 3. 2D diagrams illustrating individual SDN components for both settings of CV-SDN with purple cross to represent aperture node, annotated using corresponding transfer functions. Same volume CV-SDN represented with: (a) source-to-microphone SDN in the main volume (blue network), (b) source-to-aperture SDN in the main volume (orange network), (c) aperture-to-microphone SDN in the main volume (green network), and (d) aperture-to-aperture SDN in the main volume (pink network) with aperture-to-aperture connection in the coupled volume in all drawings (yellow network). Different volume CV-SDN represented with: (e) source-to-aperture with aperture-to-aperture SDN, and (f) aperture-to-aperture with aperture-to-microphone SDN in the main and coupled volumes.

2) $\boldsymbol{H}^{(1)}_{\mathbf{sa}}(\boldsymbol{z})$: The component simulates the sound field of the source in the main volume that is transferred to the coupled volume, and it is therefore the SDN network of the source in the main volume with the aperture node as the microphone (receiver). It is illustrated in Figs. 3(b) and 3(e) by orange scattering nodes in the same and different volume settings.

3) $\boldsymbol{H}^{(2)}_{\mathbf{aa}}(\boldsymbol{z})$: The third component is the response of the coupled volume in which the aperture acts both as the source and the microphone node. This component is represented by the yellow scattering nodes in Fig. 3.

4) $\boldsymbol{H}^{(1)}_{\mathbf{aa}}(\boldsymbol{z})$: The fourth component is the response of the main volume to the sound field of the coupled volume that is transferred back to the coupled volume, and it is the SDN of the main volume with the aperture node acting both as the source node and the microphone node. This component is illustrated in Figs. 3(d) and 3(f) by the pink scattering nodes.

5) $\boldsymbol{H}^{(q)}_{\mathbf{am}}(\boldsymbol{z})$: The fifth component is the SDN network of

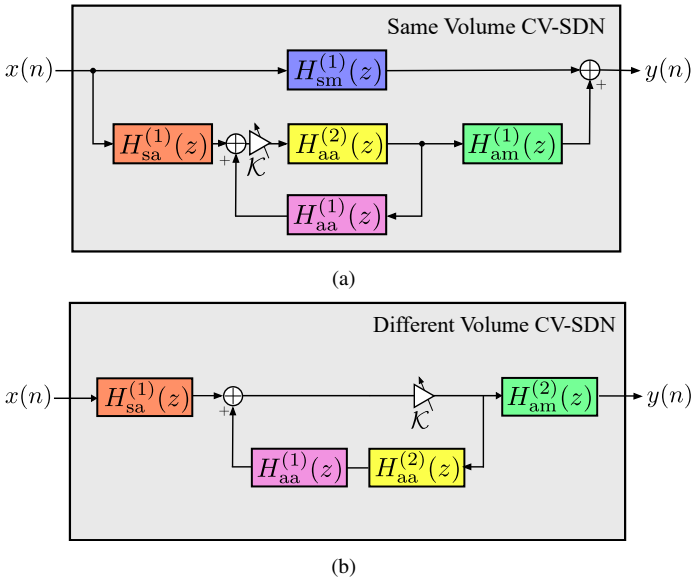

Fig. 4. System diagrams of coupled-volume SDN showing the individual SDN components (a) showing the setting where the source and the microphone are in the same volume with i) the source-to-microphone component in the main volume ($H_{\text{sm}}^{(1)}(z)$), ii) source-to-aperture component ($H_{\text{sa}}^{(1)}(z)$), iii) aperture-to-aperture components of the main and the coupled volumes ($H_{\text{aa}}^{(1)}(z)$ and $H_{\text{aa}}^{(2)}(z)$), and iv) aperture-to-microphone component ($H_{\text{am}}^{(1)}(z)$), and (b) showing the setting where the source and the microphone are in the different volumes with i) source-to-aperture component ($H_{\text{sa}}^{(1)}(z)$) in the main volume), ii) aperture-to-aperture components of the main and the coupled volumes ($H_{\text{aa}}^{(1)}(z)$ and $H_{\text{aa}}^{(2)}(z)$), and iii) aperture-to-microphone component ($H_{\text{am}}^{(2)}(z)$) in the coupled volume. All blocks are colored by the scattering nodes of SDN components as depicted in Fig. 3. $\mathcal{K}$ is the coupling coefficient that models the acoustic energy transmission between the volumes.

the main ($q = 1$) or the coupled ($q = 2$) volume between the aperture as the source node and the microphone. It is illustrated in Figs. 3(c) and 3(f) by the green scattering nodes for the same and differemt volume settings, respectively.

A block diagram showing the transfer function of the proposed CV-SDN for both settings are shown in Fig. 4, where each block represents one of the SDN components shown in Fig. 3.

B. Transfer functions of CV-SDN

Source-to-microphone, $H_{\text{sm}}^{(1)}(z)$. This component is used only in the same volume setting and is identical to a basic SDN except for an adjustment of one of the terms of the reflectance matrix due to the aperture. In contrast with the basic SDN, the reflectance matrix is $\mathbf{H}_w^{(1)}(z) = \text{diag}(H_1^{(1)}(z), \cdots, H_{cw}^{(1)}(z), \cdots, H_6^{(1)}(z))$ where $H_{cw}^{(1)}(z)$ is the adjusted reflectance term for the scattering node emulating the wall where the aperture is located. This adjustment will be discussed later. Overall, the transfer function of this component is:

$$H_{\text{sm}}^{(1)}(z) = \overline{g}_{\text{sm}} z^{-D_{\text{sm}}} + \frac{1}{5}\mathbf{k}_{\text{m}}(z)\mathbf{F}_{\text{sm}}^{(1)}(z)\mathbf{k}_{\text{s}}(z). \tag{15}$$

where $\mathbf{F}_{\text{sm}}^{(1)}(z) = [\mathbf{H}_w^{(1)}(z) - \mathbf{\Sigma P D}_{\text{sm,f}}^{(1)}(z)]^{-1}$.

Source-to-aperture, $H_{\text{sa}}^{(1)}(z)$. The main differences of this component from a basic SDN appear in (i) the direct path from the source to the aperture, and (ii) the reflectance matrix. In this component, the aperture node is modelled as an omnidirectional microphone. Therefore the direct path gain includes only the source directivity and distance attenuation terms such that $\overline{g}_{\text{sa}} = g_{\text{sa}}\mathbf{\Gamma}_{\text{s}}(\boldsymbol{\theta}_{\text{sa}})$ where $g_{\text{sa}} = (\|\mathbf{x}_{\text{s}} - \mathbf{x}_{\text{a}}\|)^{-1}$, $\mathbf{x}_{\text{a}}$ is the position of the aperture. The reflectance matrix is modified as in the source-to-microphone component, and the aperture adaptor matrix is defined as $\mathbf{k}_{\text{as}} = \mathbf{I}_{\text{as}}\mathbf{G}_{\text{sa}}\mathbf{D}_{\text{sa}}\mathbf{1}_{6\times 1}$ with $\mathbf{I}_{\text{as}} = \text{diag}(1 - \delta_{i\text{a}}, \cdots, 1 - \delta_{6\text{a}})$ where $\delta_{i\text{a}}$ is an indicator function which is 1 only if aperture and the connected scattering node, i, are on the same boundary of the volume and zero otherwise. Overall, the transfer function of this component is:

$$H_{\text{sa}}^{(1)}(z) = \overline{g}_{\text{sa}} z^{-D_{\text{sa}}} + \frac{1}{5}\mathbf{k}_{\text{a}}^{(1)}(z)\mathbf{F}_{\text{sa}}^{(1)}(z)\mathbf{k}_{\text{as}}(z) \tag{16}$$

where $\mathbf{F}_{\text{sa}}^{(1)}(z) = [\mathbf{H}_w^{(1)}(z) - \mathbf{\Sigma P D}_{\text{sa,f}}^{(1)}(z)]^{-1}$ and $\mathbf{k}_{\text{a}}^{(1)}(z) = \mathbf{1}_{6\times 1}^T \mathbf{D}_{\text{a}}^{(1)}\mathbf{G}_{\text{a}}^{(1)}$.

Aperture-to-aperture, $H_{\text{aa}}^{(q)}(z)$. These components use the aperture node both as a source and as a microphone. However, there is no direct path component, but only the recursive part. There are two aperture-to-aperture components in the CV-SDN model, $H_{\text{aa}}^{(1)}(z)$ and $H_{\text{aa}}^{(2)}(z)$ such that:

$$H_{\text{aa}}^{(q)}(z) = \frac{1}{5}\mathbf{k}_{\text{a}}^{(q)}(z)\mathbf{F}_{\text{aa}}^{(q)}(z)\mathbf{k}_{\text{a}}^{(q),T}(z), \quad q = 1, 2 \tag{17}$$

where $\mathbf{F}_{\text{aa}}^{(q)}(z) = [\mathbf{H}_w^{(q)}(z) - \mathbf{\Sigma P D}_{\text{aa,f}}^{(q)}(z)]^{-1}$ with $\mathbf{D}_{\text{aa,f}}^{(q)}(z)$ calculated using scattering nodes positioned on the geometric centroids of bounding surfaces of the main ($q = 1$) and the coupled volume ($q = 2$), respectively.

Aperture-to-microphone, $H_{\text{am}}^{(q)}(z)$. In this component the aperture node acts as a source node. The positions of scattering nodes that are not on the same boundary with the aperture node are calculated using microphone and aperture positions with respect to the bounding surfaces of the main ($q = 1$) and the coupled volume ($q = 2$), respectively. Similarly to the source-to-aperture component, the aperture acts as an omnidirectional source resulting in a direct path gain of $\overline{g}_{\text{am}} = g_{\text{am}}\mathbf{\Gamma}_{\text{m}}(\boldsymbol{\theta}_{\text{am}})$ where $g_{\text{am}} = (\|\mathbf{x}_{\text{a}} - \mathbf{x}_{\text{m}}\|)^{-1}$. The aperture adaptor vector is given as $\mathbf{k}_{\text{a}}(z) = \mathbf{I}_{\text{a}}^{(q)}\mathbf{G}_{\text{am}}\mathbf{D}_{\text{am}}\mathbf{1}_{6\times 1}$.

The transfer function of this component is:

$$H_{\text{am}}^{(q)}(z) = \overline{g}_{\text{am}} z^{-D_{\text{am}}} + \frac{1}{5}\mathbf{k}_{\text{m}}(z)\mathbf{F}_{\text{am}}(z)\mathbf{k}_{\text{m}}(z) \tag{18}$$

where $\mathbf{F}_{\text{am}}^{(q)}(z) = [\mathbf{H}_w^{(q)}(z^{-1}) - \mathbf{\Sigma P D}_{\text{am,f}}^{(q)}(z)]^{-1}, q = 1, 2$.

Aperture size control. Acoustic energy transmission from a volume to another via an aperture is modeled using the *coupling coefficient*, $\mathcal{K} = \kappa_1\kappa_2$ where κ_1 and κ_2 model the acoustic energy flow from the main volume to the coupled volume, and from the coupled volume to the main volume, respectively. These coefficients are calculated as $\kappa_q = \sqrt{S_{\text{A}}/S_{\text{T}}^{(q)}}$, where S_{A} is the aperture surface area, $S_{\text{T}}^{(q)} = S_{\text{A}} + S_{cw}^{(q)}$ is the total area of the respective volume boundary, and $S_{cw}^{(q)}$ is area of the common wall shared between two coupled volumes as observed from the main volume ($q = 1$) or from the coupled volume ($q = 2$).

Considering that the aperture would act as an anechoic patch, the effective absorption coefficient, $\alpha_i^{(q)}(\omega)$, of the wall that contains the aperture also has to be changed accordingly. This adjustment would result in a modified absorption coefficient, $\alpha_{cw}^{(q)}(\omega)$ given as:

$$\alpha_{cw}^{(q)}(\omega) = \frac{S_{\mathrm{A}} + \alpha_i^{(q)}(\omega) S_{cw}^{(q)}}{S_{cw}^{(q)} + S_{\mathrm{A}}} \tag{19}$$

where $\alpha_i^{(q)}(\omega)$ is the frequency dependent absorption coefficient of the material of the common surface on the side of volume q. The reflectance filter for that wall can be designed to satisfy:

$$\left| H_{cw}^{(q)}(z) \right|_{z=e^{j\omega}} = \sqrt{1 - \alpha_{cw}^{(q)}(\omega)}, \tag{20}$$

which has the effect of increasing the energy decay rate and reducing the effective reverberation time.

Once these components are defined, the transfer functions for same and different volume CV-SDN models are:

$$H_{\text{CV-SDN}}(z) = H_{\text{sm}}^{(1)}(z) + \mathcal{K} \frac{H_{\text{sa}}^{(1)}(z) H_{\text{aa}}^{(2)}(z) H_{\text{am}}^{(1)}(z)}{1 - \mathcal{K} H_{\text{aa}}^{(1)}(z) H_{\text{aa}}^{(2)}(z)}, \tag{21}$$

$$H_{\text{CV-SDN}}(z) = \mathcal{K} \frac{H_{\text{sa}}^{(1)}(z) H_{\text{am}}^{(2)}(z)}{1 - \mathcal{K} H_{\text{aa}}^{(1)}(z) H_{\text{aa}}^{(2)}(z)}, \tag{22}$$

respectively, both of which follow in a straightforward manner from the block diagrams in Fig. 4.

IV. Evaluation

This section presents evaluations of CV-SDN in comparison with real measurements obtained from a scale model and simulations based on ray tracing. The scale model technique, a very early tool in room acoustics research, is frequently applied in systematic data collection and analysis [70]. In a scale model, every aspect of the problem is scaled, including the frequency ranges of the transducers used in obtaining impulse responses. The measured impulse responses are then processed to obtain room impulse responses that correspond to the room that the scale model mimics.

In this study, a set of impulse response data from a previous study [71] is used as the ground truth. The employed measurement uses a $1:8$ scale acoustic model of a two-room coupled system for varying aperture configurations. The dimension of the main and the coupled rooms were $(5.6, 4.8, 6.4)$ and $(6.8, 7.2, 7)$ in m along x,y, and z directions, respectively. The corresponding natural (uncoupled) reverberation times, at a frequency of 1 kHz were measured as 0.35 s for the main room and 1.07 s for the coupled room, which results in an absorption coefficient of 0.40 for the main room and 0.17 for the coupled room. Aperture opening is the ratio of the aperture area to the total area.. The maximum total area of the aperture was $S_A = 30.72$ m^2. Only the measurements for 7%, 15%, 30%, and 60% of the total aperture area were available for the same volume setting which corresponds to $S_A = 2.25$, 4.62, 9.18, and 18.4 m^2, respectively. Whereas for the different volume case, only the measurements for 60% of total aperture area were available. For the corresponding

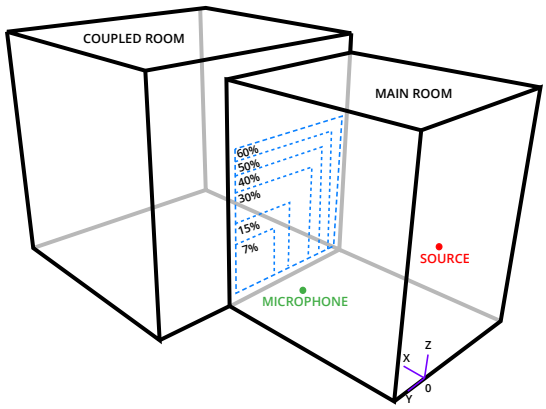

Fig. 5. Model used in the ray-tracing simulation of the coupled rooms highlighting source (S) and receiver (R) positions for the same volume setting, and different sized apertures indicated with blue dashed squares. The positions of the receiver for different volume setting are given in Table II.

aperture opening percentages, κ_1 and κ_2 values are set to $\kappa_1 = 0.27$, 0.39, 0.55, 0.63, 0.71, and 0.77 for the main volume, and $\kappa_2 = 0.21$, 0.30, 0.43, 0.49, 0.55, and 0.60 for the coupled volume. The measurements were carried out using a $1:8$ scaled mini microphone with a two-channel microphone amplifier of GRAS Power Module Type 12AA, a mini-omnidirectional dodecahedron loudspeaker amplified with Radio Shack Optimus SA-155 Integrated stereo amplifier, and a dual-channel programmable filter (Stanford Research Systems, Model SR650) to reduce the low-frequency noise components within the measurement environment.

Ray tracing simulations generated by ODEON v. 14.00 software package were used as a second reference. Simulations allowed comparisons for configurations that are not available in the scale-model measurement dataset. These additional configurations were for the 40%, and 50% aperture openings, i.e. aperture areas of $S_A = 12.25$, and 15.36 m^2, respectively. Room acoustic models with the same dimensions emulated in the scale-model measurements were generated. An omnidirectional point source and an omnidirectional receiver were placed at positions corresponding to the same positions in the measurements. The models and absorption coefficients were tuned using the Sabine equation to fit the reverberation times measured at scale-model tests when the rooms were in the decoupled condition. Fig. 5 shows a diagram of the evaluated coupled-volume system highlighting source and receiver positions as well as different sizes of the aperture.

Since only the measurements for 60% aperture opening were available for the different volume setting, six different microphone positions are tested in contrast with the same volume setting, which only uses a fixed microphone and source positioning. The six microphone positions for different volume case are labeled as: J_1, J_2, I_1, I_2, H_1, and H_2 with exact positions given in Table II. The CV-SDN model uses the same room sizes, aperture areas, and source-receiver positions.

The obtained EDCs for the same volume setting are compared in Figs. 6(a) and 6(b). The figures are given for the cases of 15% and 30% aperture area ratios and show the EDC of CV-SDN, the ray tracing software (RTS) and the scale

model (SM) as well as the noise floor for the scale model. Corresponding turning points estimated using the described Bayesian analysis framework are shown in Figs. 7(a) and 7(b). It may be observed that CV-SDN closely follows the SM and RTS patterns for smaller aperture openings, and that the double-rate decay profile where the faster decay of the main volume is followed by the slower decay of the coupled volume can easily be discerned. Since CV-SDN and RTS both provide noise-free room impulse responses, noise floor is only visible for SM.

The comparison of ΔL (in dB), *turning point*, and decay times T_1 and T_2(in seconds) are given in Table I. The turning points are also shown graphically for 15% and 30% aperture opening ratios in Fig. 7.

It may be observed that as the aperture area ratio increases, ΔL decreases. We speculate that this effect is due to the system starting to behave more like a single volume than a coupled volume. It may be observed that CV-SDN provides ΔL values that agree well with the ΔL values obtained for SM and RTS for aperture openings up to 30%. The similarity of ΔL value for smaller openings is high, though the decrease in CV-SDN is not as pronounced as in RTS or SM. Specifically, ΔL values of CV-SDN in 15% and 30% cases are very close to SM that is the ground truth.

The *turning point* occurs earlier with increasing aperture area ratio as the energy transferred from the coupled room back to the main room starts to dominate the EDC earlier. This trend can be observed both in the numerical models (i.e. CV-SDN and RTS) and for the measured room impulse responses (i.e. SM). This rate of decrease is lower in CV-SDN in comparison with the SM. As evident in Fig. 7, the order of turning points of the three models changes as the aperture size increases, *i.e.* SM starts to have a turning point earlier than the CV-SDN. This is because the decrease in the turning point in CV-SDN is not as pronounced as in SM or RTS. The level at which the *turning point* occurs increases for all of the methods, with CV-SDN being closer to SM, whereas RTS has substantially higher values.

The estimated decay times T_1 and T_2 indicate that CV-SDN has comparable results to SM and RTS. While these parameters are perhaps less meaningful for specifying the behaviour of the coupled-volume system they are shown in Table I to facilitate the reproducibility of these results.

EDCs obtained for a different volume setting are shown in Fig. 8. For the given setting the aperture area ratio is 60%. EDC from SM is distorted due to the noise floor but the slope is extended to correctly show the RT_{60}. It can be observed that CV-SDN is in between SM and RTS in terms of RT_{60}, and closer to the ground truth compared to RTS in all of the measurements given in Table II. It may be observed that CV-SDN and RT both result in a single-sloped EDC, as do the SM measurements.

TABLE I
PARAMETERS OF THE COMPARED MODELS WITH DIFFERENT APERTURE RATIOS FOR SAME VOLUME SETTING. THE MISSING VALUES FOR 7% SM ARE DUE TO THE FACT THAT A SINGLE-SLOPE DECAY IS OBSERVED IN THE SCALE MODEL AT THIS APERTURE VALUE.

Aperture %	T1(s)	T2(s)	ΔL (dB)	Turning Point (s, dB)
7% CV-SDN	0.35	1.12	12.7	$(0.14, -19)$
7% RTS	0.32	0.95	15.5	$(0.14, -23)$
7% SM	0.42	–	–	–
15% CV-SDN	0.34	1.1	11.5	$(0.12, -16)$
15% RTS	0.3	0.95	12	$(0.10, -17)$
15% SM	0.36	0.91	12	$(0.14, -19)$
30% CV-SDN	0.35	1.14	10.7	$(0.12, -15)$
30% RTS	0.3	0.92	8	$(0.07, -12)$
30% SM	0.28	0.91	9.2	$(0.10, -13)$
40% CV-SDN	0.35	1.1	9.5	$(0.11, -14)$
40% RTS	0.33	0.9	6.5	$(0.07, -10)$
50% CV-SDN	0.34	1.1	9.5	$(0.10, -14)$
50% RTS	0.3	0.86	5	$(0.05, -8)$
60% CV-SDN	0.3	1.13	9.3	$(0.09, -13)$
60% RTS	0.3	0.84	2.5	$(0.03, -5)$
60% SM	0.29	0.93	6.6	$(0.08, -10)$

TABLE II
RT_{60} TIMES FOR 60% APERTURE OPENING FOR DIFFERENT VOLUME SETTING WITH CHANGING MICROPHONE POSITIONS

Position	CV-SDN	RTS	SM
J1 (6.98,6.45,1)	1.02 s	1.13 s	0.91 s
J2 (11,6.45,2)	0.98 s	1.13 s	0.88 s
I1 (6.98,4.35,1)	0.99 s	1.13 s	0.92 s
I2 (11, 4.35, 2)	0.95 s	1.15 s	0.92 s
H1 (6.98, 1.40, 1)	1.01 s	1.13 s	0.85 s
H2 (11, 1.40, 1)	0.99 s	1.15 s	0.92 s

V. COMPUTATIONAL COMPLEXITY

The number of floating-point operations per second (FLOPS) carried out by the basic SDN has been shown to be $F_s\left[2(M-1)^3+(P+2)(M-1)^2+M\right]$, where M is the number of scattering junctions ($M=6$ in the simulations of this paper), and P is the number of operations per sample required by each wall filter [26].

Inspection of Fig. 4 shows that the complexity of CV-SDN is approximately four to five times that of the basic SDN, depending on whether the source and microphone are in different or in the same volume. This translates to approximately 58 MFLOPS for the different-volume CV-SDN and 73 MFLOPS for the same-volume CV-SDN, having assumed $F_s = 44.1$ kHz and simple gains for the wall operations ($P=1$). For more elaborate wall operations, biquad filters ($P=8$) can be used to approximate the absorbent characteristics of e.g. cotton carpet, ceiling tiles, or brick walls [26]. The complexity in this case becomes 89 MFLOPS for the different-volume CV-SDN and 112 MFLOPS for the same-volume CV-SDN. The remainder of this section will refer to the same-volume case with frequency independent wall processing.

The complexity of CV-SDN is compared here to the complexity of convolution, a step required by most room acoustic models that allow simulation of connected spaces, including ray tracing algorithms. For $\mathrm{RT}_{60} = 1$ s, naïve time-domain convolution has a computational cost of approximately 3.9 GFLOPS or 53 times more than same-volume CV-SDN. If a certain amount of buffering delay is acceptable, convolution can be carried out vastly more efficiently using the overlap-add method [72]. The method involves dividing the input signal into time frames and calculating 2 FFT of length N (having assumed that the FFT of the IR also needs to be recalculated,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2022.3143697, IEEE/ACM Transactions on Audio, Speech, and Language Processing

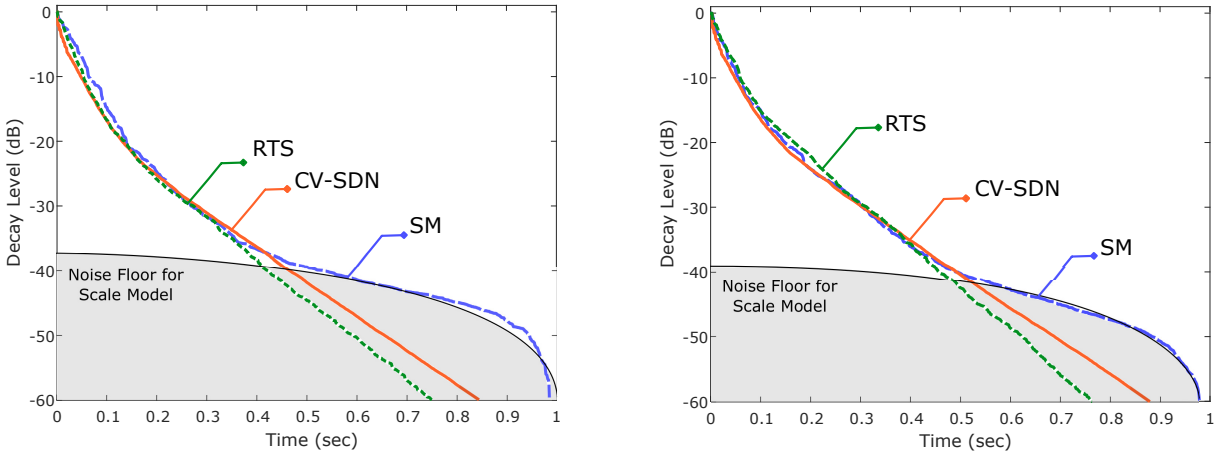

Fig. 6. Energy decay curves obtained using room impulse responses measured in the scale model (SM; blue dashed curve), simulated with ray tracing software (RTS; green dotted curve) and coupled-volume SDN (CV-SDN; orange solid curve) for the same model geometry. Different aperture sizes of (a) 15% and (b) 30% of the common wall are shown. Also shown is the noise floor for the scale model estimated using the Bayesian analysis framework as described in the text.

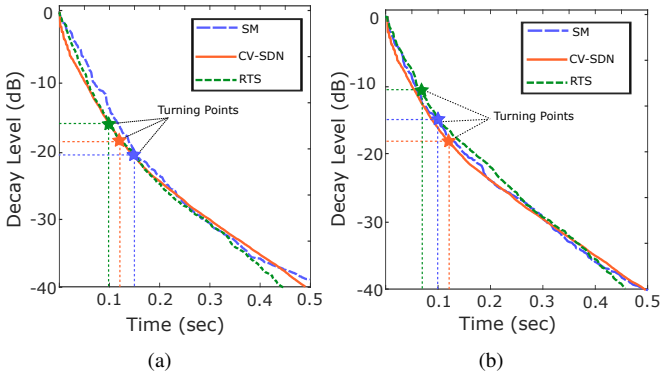

Fig. 7. Turning points estimated for room impulse responses obtained from the scale model (SM), ray tracing software (RTS), and coupled-volume SDN for an aperture size of (a) 15%, and (b) 30% of the common wall for the same volume setting.

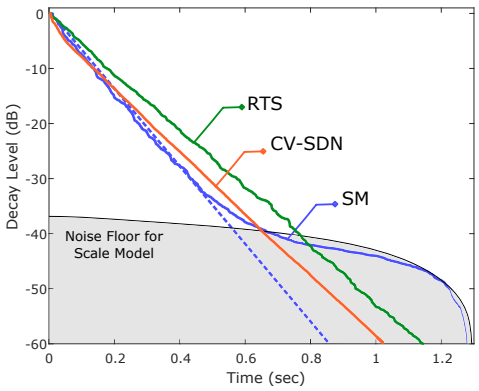

Fig. 8. Energy decay curves obtained using room impulse responses measured in the scale model (SM; blue curve), simulated with ray tracing software (RTS; green slope), and coupled-volume SDN (CV-SDN; orange slope) where the source and the microphone are placed in the different volumes with 60% aperture opening at microphone position H_1.

as would be needed for a dynamic acoustic scene), N complex multiplications, 1 inverse FFT, and $\lceil T_{60} F_s \rceil$ real additions for each time frame. In order for the circular convolution to be equal to the linear convolution, the value of N must be such that $N \geq \lceil F_s/F_r \rceil + \lceil F_s \mathrm{RT}_{60} \rceil - 1$, where F_r is the frame refresh rate (i.e. the number of time frames per second). The asymptotic cost of each FFT is $6N \log_2 N$ [72], and each complex multiplication requires 4 real multiplications and 2 additions. The computational complexity of the overlap-add method is therefore $F_r(18N \log_2 N + 6N + \lceil F_s \mathrm{RT}_{60} \rceil - 1)$ FLOPS. The choice of the frame refresh rate, F_r, depends on the acceptable amount of buffering delay, which in turn depends on the application scenario.

For real-time applications with a linked video component, the ITU standard BR.265-9 [73] recommends that buffering delay should be shorter than the maximum latency of a half-frame delay between the video and audio. Assuming a video feed running at 50 frames per second, the frame refresh rate should then be set to at least $F_r = 100$ Hz. With this setup, CV-SDN already has a lower number of FLOPS than convolution for reverberation time values beyond $\mathrm{RT}_{60} = 0.07$ s, which includes most cases of practical interest. For $\mathrm{RT}_{60} = 1$ s, same-volume CV-SDN requires nearly an order of magnitude fewer FLOPS than convolution alone, increasing to over eighteen-fold for $\mathrm{RT}_{60} = 2$ s.

Also, the reader should note that several convolution methods have been proposed for real-time auralization [74] that are more efficient than the standard overlap-add method and that have not been considered here.

It should also be noted that for geometrical and wave-based methods, the computational cost of convolution is typically negligible compared to that of generating the room impulse response. This is especially so for geometries more complex than simple cuboid rooms as considered in this paper.

Geometrical acoustics methods, for instance, require a vast number of visibility calculations, which in some cases leads to exponential complexity [75]. Wave-based models would have a greater physical accuracy but typically have an even larger computational complexity. Given the vastly different classes of room acoustic models and the respective software implementations, a more thorough comparison in terms of computational complexity is beyond the scope of this paper.

VI. Conclusions

We presented the coupled-volume SDN (CV-SDN) as an extension to the scattering delay network (SDN) models to simulate acoustics of coupled volumes. CV-SDN combines multiple SDNs that simulate individual volumes via an aperture model which allows controlling the aperture size. The transfer functions of CV-SDN is derived as a combination of modified basic SDN transfer functions.

The proposed model is tested and validated against scale model measurements and simulations obtained using a state-of-the-art ray tracing based acoustic simulation software. Three summary measures were used: level difference and turning point for cases when a double slope EDC is observed, and reverberation time when a single slope EDC is observed. CV-SDN was observed to provide simulations that closely follow the scale model, that is the ground truth, as well as the ray tracing based room acoustics simulation.

We observed that whilst the trends for acoustic parameters calculated for the room impulse responses obtained using CV-SDN follows the same trends for SM, the model's accuracy in terms of the level difference decreases for larger apertures for the same volume setting. We speculate that this decrease in accuracy might be due to the coupled volumes starting to behave like a single complex volume and that a single aperture node might no longer be an adequate model for their interactions. It is however unclear whether the relatively small differences would be audible and if a more accurate model would result in a perceptually more realistic rendition of coupled volume acoustics.

Computational complexity of CV-SDN shows that considerable savings in the number of operations compared to overlap-add convolution are possible, even without accounting for the cost of room impulse response computation.

The CV-SDN model discussed here only covers the case of two volumes which are rectangular, and whilst its extension to the case of multiple connected enclosures or complex geometries are also possible, details require further research. Similarly, whilst the proposed method did not result in stability problems for the tested cases, a formal treatment of the stability of CV-SDN also requires further research.

References

[1] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[2] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel, "More than 50 years of artificial reverberation," in *Proc. 60th AES Int. Conf.: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*, 2016.

[3] K. Kowalczyk, "Boundary and medium modelling using compact finite difference schemes in simulations of room acoustics for audio and architectural design applications," Ph.D. dissertation, Queen's University Belfast, 2010.

[4] D. T. Murphy, "Digital waveguide mesh topologies in room acoustics modelling," Ph.D. dissertation, University of York, 2000.

[5] A. Craggs, "Acoutical modeling: Finite element method," in *Handbook of Acoustics*, M. J. Crocker, Ed. Wiley-Interscience, 1998, ch. 12, pp. 149–156.

[6] R. Mehra, N. Raghuvanshi, L. Savioja, M. C. Lin, and D. Manocha, "An efficient GPU-based time domain solver for the acoustic wave equation," *Appl. Acoust.*, vol. 73, no. 2, pp. 83–94, 2012.

[7] L. Savioja, "Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics," in *Proc. 13th Int. Conf. on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010.

[8] B. Hamilton, S. Bilbao, and C. J. Webb, "Revisiting implicit finite difference schemes for 3-d room acoustics simulations on gpu," *Proc. Digital Audio Effects (DAFx), Erlangen, Germany*, pp. 41–48, 2014.

[9] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.

[10] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "3D beam tracing based on visibility lookup for interactive acoustic modeling," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 10, pp. 2262–2274, 2016.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[12] W. G. Gardner, "Reverberation algorithms," in *Applications of digital signal processing to audio and acoustics*. Springer, 2002, pp. 85–131.

[13] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," Proc. 104th AES Convention, Preprint #3030, Feb. 1991.

[14] S. J. Schlecht and E. A. Habets, "On lossless feedback delay networks," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1554–1564, 2017.

[15] S. J. Schlecht and E. Habets, "Feedback delay networks: Echo density and mixing time," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 2, pp. 374–383, 2017.

[16] S. J. Schlecht and E. A. Habets, "Sign-agnostic matrix design for spatial artificial reverberation with feedback delay networks," in *Proc. of the 2018 Int. Conf. of the Audio Eng. Soc.: Spatial Reproduction-Aesthetics and Science*, 2018.

[17] J. O. Smith, "A new approach to digital reverberation using closed waveguide networks," in *Proc. 11th International Computer Music Conference (ICMC-85)*, Burnaby, Canada, 1985, pp. 47–53.

[18] P. Huang, M. Karjalainen, and J. O. Smith, "Digital waveguide networks for room response modeling and synthesis," in *Audio Engineering Society Convention 118*. Audio Engineering Society, 2005.

[19] D. Rocchesso and J. O. Smith, "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 1, pp. 51–63, January 1997.

[20] D. Rocchesso, "The ball within the box: a sound-processing metaphor," *Computer Music Journal*, vol. 19, no. 4, pp. 47–57, 1995.

[21] B. Alary, A. Politis, S. Schlecht, and V. Välimäki, "Directional feedback delay network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762, 2019.

[22] T. Wendt, S. D. Ewert, *et al.*, *Perceptual and room acoustical evaluation of a computational efficient binaural room impulse response simulation method*, 2014.

[23] O. Das and J. S. Abel, "Grouped feedback delay networks for modeling of coupled spaces," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 486–496, 2021.

[24] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "Scattering delay network: An interactive reverberator for computer games," in *Proc. 41st Int. Conf. of the Audio Eng. Soc.: Audio for Games*, 2011.

[25] H. Hacihabiboglu, E. De Sena, and Z. Cvetkovic, "Frequency-domain scattering delay networks for simulating room acoustics in virtual environments," in *Proc. 7th Int. Conf. on Signal Image Tech. & Internet-Based Systems (SITIS-11)*, 2011, pp. 180–187.

[26] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 9, pp. 1478–1492, 2015.

[27] G. Kendall, W. Martens, D. Freed, D. Ludwig, and R. Karstens, "Image model reverberation from recirculating delays," in *Audio Engineering Society Convention 81*. Audio Engineering Society, 1986.

[28] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A minimal personalization of dynamic binaural synthesis with mixed structural modeling and scattering delay networks," in *Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP-2020)*, 2020, pp. 411–415.

[29] S. Djordjevic, H. Hacıhabiboğlu, Z. Cvetkovic, and E. De Sena, "Evaluation of the perceived naturalness of artificial reverberation algorithms," in *Proc. 148th Conv. of the Audio Eng. Soc. Convention*, May 2020.

[30] A. D. Baldwin, S. Serafin, and C. Erkut, "ScattAR: a mobile augmented reality application that uses scattering delay networks for room acoustic synthesis," in *Proc. 23rd ACM Symp. on Virtual Real. Softw. & TechnologyVirtual Reality Software and Technology*, 2017.

[31] A. Baldwin, S. Serafin, and C. Erkut, "Towards the design and evaluation of delay-based modeling of acoustic scenes in mobile augmented reality," in *IEEE 4th VR Workshop Sonic Interact. Virtual Env. (SIVE-18)*, 2018, pp. 1–5.

[32] C. Erkut, J. Holfelt, and S. Serafin, "Mobile AR in and out: Towards delay-based modeling of acoustic scenes," in *Proc. IEEE Conf. on Virtual Real. 3D User Interf. (VR-18)*, 2018, pp. 1–2.

[33] C. Yeoward, R. Shukla, R. Stewart, M. Sandler, and J. D. Reiss, "Real-time binaural room modelling for augmented reality applications," *J. Audio Eng. Soc.*, vol. 69, no. 11, pp. 818–833, 2021.

[34] H. Kuttruff, "Acoustic variables and basic relations," in *Acoustics: an introduction*. CRC Press, 2006, pp. 44–45.

[35] L. Beranek, *Concert halls and opera houses: music, acoustics, and architecture*. Springer Science & Business Media, 2012.

[36] C. F. Eyring, "Reverberation time measurements in coupled rooms," *J. Acoust. Soc. Am.*, vol. 3, no. 2A, pp. 181–206, 1931.

[37] A. H. Davis, "Vi. reverberation equations for two adjacent rooms connected by an incompletely soundproof partition," *London, Edinburgh, and Dublin Phil. Mag. J. Sci.*, vol. 50, no. 295, pp. 75–80, 1925.

[38] L. Cremer and H. A. Müller, *Principles and applications of room acoustics*. Applied Science, 1982.

[39] C. M. Harris and H. Feshbach, "On the acoustics of coupled rooms," *J. Acoust. Soc. Am.*, vol. 22, no. 5, pp. 572–578, 1950.

[40] C. Lyle, "Recommendations for estimating reverberation times in coupled spaces," *Acoust. Lett.*, vol. 5, no. 2, pp. 35–38, 1981.

[41] C. Thompson, "On the acoustics of a coupled space," *J. Acoust. Soc. Am.*, vol. 75, no. 3, pp. 707–714, 1984.

[42] M. Meissner, "Acoustic energy density distribution and sound intensity vector field inside coupled spaces," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 228–238, 2012.

[43] E. Wester and B. Mace, "A statistical analysis of acoustical energy flow in two coupled rectangular rooms," *Acta Acust. united Ac.*, vol. 84, no. 1, pp. 114–121, 1998.

[44] J. Anderson, M. Bratos-Anderson, and P. Doany, "The acoustics of a large space with a repetitive pattern of coupled rooms," *J. Sound Vib.*, vol. 208, no. 2, pp. 313–329, 1997.

[45] F. Martellotta, "Identifying acoustical coupling by measurements and prediction-models for st. peter's basilica in rome," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1175–1186, 2009.

[46] L. Nijs, G. Jansens, G. Vermeir, and M. Van der Voorden, "Absorbing surfaces in ray-tracing programs for coupled spaces," *Appl. Acoust.*, vol. 63, no. 6, pp. 611–626, 2002.

[47] J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 958–969, 2004.

[48] J. E. Summers, R. R. Torres, Y. Shimizu, and B.-I. L. Dalenbäck, "Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing," *J. Acoust. Soc. Am.*, vol. 118, no. 3, pp. 1491–1502, 2005.

[49] A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 2043–2054, 2006.

[50] P. Luizard, J.-D. Polack, and B. F. Katz, "Sound energy decay in coupled spaces using a parametric analytical solution of a diffusion equation," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2765–2776, 2014.

[51] Z. Sü Gül, "Exploration of room acoustics coupling in Hagia Sophia of istanbul for its different states," *J. Acoust. Soc. Am.*, vol. 149, no. 1, pp. 320–339, 2021.

[52] L. Du and G. Pavić, "Modeling of multiply connected sound spaces by the surface coupling approach," *J. Acoust. Soc. Am.*, vol. 146, no. 6, pp. 4273–4287, 2019.

[53] M. Ermann and M. Johnson, "Exposure and materiality of the secondary room and its impact on the impulse response of coupled-volume concert halls," *J. Sound Vib.*, vol. 284, no. 3-5, pp. 915–931, 2005.

[54] D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," *J. Acoust. Soc. Am.*, vol. 118, no. 3, pp. 1480–1490, 2005.

[55] H. Pu, X. Qiu, and J. Wang, "Different sound decay patterns and energy feedback in coupled volumes," *J. Acoust. Soc. Am.*, vol. 129, no. 4, pp. 1972–1980, 2011.

[56] D. T. Bradley and L. M. Wang, "Quantifying the double slope effect in coupled volume room systems," *Building Acoustics*, vol. 16, no. 2, pp. 105–123, 2009.

[57] S. Shi, G. Jin, B. Xiao, and Z. Liu, "Acoustic modeling and eigenanalysis of coupled rooms with a transparent coupling aperture of variable size," *Journal of Sound and Vibration*, vol. 419, pp. 352–366, 2018.

[58] P. Luizard, B. F. Katz, and C. Guastavino, "Perceptual thresholds for realistic double-slope decay reverberation in large coupled spaces," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. 75–84, 2015.

[59] N. Xiang, P. Goggans, T. Jasa, and P. Robinson, "Bayesian characterization of multiple-slope sound energy decays in coupled-volume systems," *J. Acoust. Soc. Am.*, vol. 129, no. 2, pp. 741–752, 2011.

[60] H. Kuttruff, *Room acoustics*. CRC Press, 2016.

[61] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," *J. Acoust. Soc. Am.*, vol. 110, no. 3, pp. 1415–1424, 2001.

[62] N.Xiang and P.M.Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," *J. Acoust. Soc. Am.*, vol. 113, no. 5, pp. 2685–2697, 2003.

[63] N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayeisan decay time estimation," *J. Acoust. Soc. Am.*, vol. 117, no. 6, pp. 3707–3715, 2005.

[64] Z. S. Gül, M. Çalışkan, A. Tavukçuoğlu, and N. Xiang, "Assessment of acoustical indicators in multi-domed historic structures by non-exponential energy decay analysis," *Acoustics Australia*, vol. 46, no. 2, pp. 181–192, 2018.

[65] N. Xiang, Y. Jing, and A. C. Bockman, "Investigation of acoustically coupled enclosures using a diffusion-equation model," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1187–1198, 2009.

[66] J. O. Smith, "A New Approach to Digital Reverberation Using Closed Waveguide Networks," in *Proc. 1985 Int. Computer Music Conf.*, Vancouver, Canada, 1985, pp. 47–53.

[67] W. C. Sabine, *Collected papers on acoustics*. Harvard University Press Cambridge, MA, 1921.

[68] C. F. Eyring, "Reverberation time in "dead" rooms," *J. Acoust. Soc. Am.*, vol. 1, no. 2A, pp. 217–241, 1930.

[69] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. 1992 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP-92)*, vol. 2. IEEE, 1992, pp. 221–224.

[70] N. Xiang, J. Escolano, J. M. Navarro, and Y. Jing, "Investigation on the effect of aperture sizes and receiver positions in coupled rooms," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3975–3985, 2013.

[71] Z. Su, "Systematic investigations on energy decays in acoustically coupled spaces using the scale-model technique," Master's thesis, Rensselaer Polytechnic Institute, 2006.

[72] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-time signal processing*, 2nd ed. Prentice Hall, 1999.

[73] "Recomm. BR.265-9, operating practices for the international exchange of programmes** on film for television use," International Telecommunication Union Radiocommunication Assembly, 2004.

[74] F. Wefers, *Partitioned convolution algorithms for real-time auralization*. Logos Verlag Berlin GmbH, 2015, vol. 20.

[75] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.