ModPhy: System Design for Real-time Modular Sound Synthesis with Physically-Modeled Objects

Zehao Wang
UC San Diego
zehaowang@ucsd.edu

Han Zhang
UC San Diego
haz074@ucsd.edu

Yifan Guo
UC San Diego
yig020@ucsd.edu

ABSTRACT

We introduce ModPhy, a system aiming for real-time modular sound synthesis with physically-modeled objects realized as Max/MSP and Pure Data externals. With the accelerated implementation of the finite-difference time-domain simulation for partial-differential-equation-based physical models and a hierarchical design of control parameters, these physically-modeled objects can be efficiently incorporated with an abstract sound synthesis paradigm for flexible sound design while keeping their real-world physical meaning. The unity of oscillators and audio effects and the advantages of those physical models' controllability provide a unified way for sound design and music creation and they are also in line with the philosophy of some contemporary composition schools. With the implemented Max/MSP and Pure Data objects, a number of potential music applications are discussed. We believe such a system will become a powerful toolkit for modern composers and sound artists.

1. INTRODUCTION

Real-time sound synthesis has become an indispensable requirement in contemporary music production, empowering musicians to create and manipulate sounds instantaneously during live performances. Abstract sound synthesis techniques, such as frequency modulation (FM)[1], offer users the ability to manipulate a predefined set of parameters that abstractly define how sounds are shaped for their desired sonic. In addition to abstract synthesis techniques, physical models have revolutionized sound synthesis by simulating the behavior of acoustic instruments. This approach provides a different avenue for creating expressive sounds. Recent advancements in both theoretical understanding and algorithm design[2, 3, 4, 5, 6] have significantly accelerated sound synthesis based on numerical solutions of partial differential equations (PDEs), which have made sound synthesis based on such methods much better and faster than ever before.

One notable distinction between PDE-based models and other synthesis techniques is controlling sound by describing the physical characteristics beforehand or during the generation process, rather than relying solely on post-processing. However, a significant challenge in physical modeling sound synthesis relates to the control parameters used for these models. Typically, physical models incorporate tens of raw parameters, while still needing to adhere to stability constraints. This poses a difficulty in effectively utilizing a limited set of parameter combinations within extensive feasible domains to manipulate the resulting sound which is an important aspect of sound synthesis.

Exploring the integration of physical models and abstract sound synthesis for sound design is a significant topic of interest. However, one major challenge lies in the disparity between the raw parameters used in physical synthesis and the common dimensions utilized in sound design. While it is important to preserve the dimensions of control parameters, it is equally crucial to consider both the controllability and representativeness of the resulting sound. By adopting a modular system that combines physical models with the abstract synthesis paradigm, physically-modeled objects can function as modules, interacting with other objects connected through different means that describe the sound. Such a modular way enables users to create unique sonic architectures by combining and connecting various modules, such as oscillators, filters, and envelopes, in diverse configurations, and the advantages of different techniques can be harnessed, providing a flexible and expressive platform for sound designers: physical models will not be limited to only "mimicking" existing musical instruments , they can also be treated as relatively complex processes or oscillations with meaningful physical properties to be incorporated into the entire sound design workflow, especially in modular computer music systems like Max/MSP (Max) and Pure Data (Pd).

In this paper, we introduce ModPhy, a system for real-time modular sound synthesis with physically-modeled objects in Max/Pd. We first present the underlying physics of modeled systems and the fast implementation approach of finite-difference time-domain (FDTD) methods, a particular method for numerical simulation of PDEs. Then we briefly introduce our system design of ModPhy, including the hierarchy of control parameters, and the interaction between modules. Finally, we discuss our inspiration from contemporary composers and potential music applications of ModPhy.

2. TECHNICAL SETUP

2.1 Modeling vibrations using PDEs

The underlying physics of vibrating bodies can be modeled by partial differential equations, especially the wave equation which is widely used in musical acoustics[7]. In this

1-D	Mass-spring systems	$-\omega^2\mathbf{1}$
	Strings	$c^2\frac{\partial^2}{\partial x^2}$
	Tubes	$c^2\frac{1}{S(x)}\frac{\partial}{\partial x}\left(S(x)\frac{\partial}{\partial x}\right)$
	Bars	$-\kappa^2\frac{\partial^4}{\partial x^4}$
2-D	Membranes	$c^2\Delta$
	Plates	$-\kappa^2\Delta\Delta$

Table 1: $\mathcal{L}$ for different vibrating objects, where $\mathbf{1}$ is the identity operator and $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator; $S(x)$ is the cross-sectional area of the tube at position x; ω, c, κ are scaled parameters derived from physical parameters mentioned in the last tier mentioned in Sec 3.2.

paper, we mainly consider several different types of vibrating objects shown in Table 1, whose equations share the following general form of second-order systems in time:

$$\frac{\partial^2 u}{\partial t^2} = \mathcal{L}u + f, \tag{1}$$

where u denotes the displacement (like strings) or pressure (like tubes) at each position of the object, $\mathcal{L}$ is a differential operator corresponding to the object's vibration mechanism which has both spatial and temporal differential operators, and f is an external force like excitation for the use of oscillators/instruments or an input waveform for the use of audio effects like reverb. For the sake of brevity, here we only show these systems in their simplest form, in our implementation, we also consider models like systems with loss [1] , stiff strings, or nonlinear plate models, where extra computational costs are required.

2.2 Simulating vibrations by FDTD methods

FDTD methods have a long history in scientific computing[9] and there are a large number of papers applying these methods for sound synthesis[8]. The basic idea is to discretize both temporal and spatial domains to grids u_l^n on a given region [2] where n denotes the time step and l denotes the spatial coordinate , and then derive the corresponding update rules using difference operators as approximations for differential operators. For the discretization of temporal domain, $\frac{\partial^2 u}{\partial t^2}$ in Eq. (1) is approximated by $\left.\frac{\partial^2 u}{\partial t^2}\right|_{t=nk} \approx \frac{u^{n+1}-2u^n+u^{n-1}}{k^2}$, i.e., the forward Euler method, where k is the equally spaced time interval. For other temporal/spatial differential operators in $\mathcal{L}$, we can also apply the forward Euler method or backward/central Euler method for approximation which are well established in various textbooks[9, 8]. Thus, the numerical scheme of Eq. (1) should have the following matrix form

$$\hat{\mathbf{u}}^{n+1} = \mathbf{B}\hat{\mathbf{u}}^n + \mathbf{C}\hat{\mathbf{u}}^{n-1} + \mathbf{f^n}, \tag{2}$$

where $\mathbf{B}$ and $\mathbf{C}$ are sparse matrices derived from the discrete operators and boundary conditions, and $\mathbf{f^n}$ denotes a function of discretized external force and extra terms which may include nonlinearities. Note that some external forces may depend on u or its derivatives, like $f(t) = \delta(x-x_i)F(t)\phi\left(\frac{\partial u}{\partial t}(x_i) - v(t)\right)$ for bow-string interaction where δ is the Dirac function, F and v is the force and velocity of the bowing motion, and ϕ is some friction parameter. In this paper, implementations of several simple external forces or interactions like raised cosine, bow/hammer-string interaction, mallet/stick excitation are provided [3] .

To guarantee the numerical stability of FDTD-based simulation, each system needs to satisfy some stability conditions derived from frequency-domain analysis or von Neumann analysis[8]. Recent studies about invariant energy quadratization and scalar auxiliary variable methods also provide another approach to derive the numerical scheme of nonlinear systems like nonlinear string vibrations and nonlinear plate vibrations[2] with relatively weak requirements for stability conditions compared to the previous simple FDTD schemes. In general, the stability condition can be given as $h \geq \psi(k;\theta)$, where h is the equally-spaced spatial interval for each axis and $\psi(k)$ is a monotonic increasing function w.r.t. k given θ which is the parameter set of the corresponding PDE.

2.3 Variable conversion for other synthesis approaches

In order to interact with other synthesis approaches like digital waveguide methods (DWG)[10], variable conversion might be required. Considering the equivalence between DWG and FDTD methods[11], the K/W converter between DWG's wave variables (u^+, u^-) and FDTD's Kirchhoff variables (u) can thus be given as follows[10],

$$\begin{bmatrix} W^+ \\ z^{-1}W \end{bmatrix} = \begin{bmatrix} 1 & -z^{-2} \\ 1 & (1-z^{-2}) \end{bmatrix} \begin{bmatrix} z^{-1}K \\ W^- \end{bmatrix}, \tag{3}$$

where K is the Kirchhoff variable in FDTD, and $W = W^+ + W^-$ is the wave variable in DWG. This converter can be implemented as a standalone object in Max and Pd to connect objects using these two types of variables in real-time, which will be discussed in Sec 3.3 and Sec 3.4.

2.4 Fast implementation for real-time synthesis

To implement these simulations as Max or Pd externals running on personal computers, efficient implementation approaches on CPUs are needed. Webb and Bilbao[3] show the limits of several systems' real-time simulation mentioned in Sec 2.1, and (multi-thread) single instruction, multiple data (SIMD) parallelization using instructions like advanced vector extensions (AVX) has the best performance . The basic idea of SIMD parallelization is to rewrite/unroll the matrix-form update rules (2) to individual update rules for each grid point in a matrix-free way, and then perform the operations for a consecutive series of grid points in parallel using SIMD instructions. Using these acceleration strategies, Wang et al.[5] shows that only 0.025 seconds are needed to run a Pd external for a 1-second linear plate sound synthesis at 44.1Khz; with fast algorithms and implementations for sparse linear system solution[4], a 1-second 44.1kHz simulation single nonlinear plate with a

[1] For example, in the form of $\frac{\partial}{\partial t}$ (frequency-independent) or $\Delta\frac{\partial}{\partial t}$ (frequency-dependent)[7, 8]. In most cases of sound synthesis, the order of temporal differential operators in $\mathcal{L}$ is not greater than 2.

[2] In this paper, we only consider rectangular regions for 2-D systems.

[3] Note that some excitation mechanisms are coupled with the systems, which cannot be explicitly added as extra terms, but can be implemented as a separate coupled model which is beyond the scope of this paper.

discretization of 13×27 only takes 0.15 seconds[6]. Those results show the possibility to run a number of physically-modeled objects simultaneously in a modular system.

3. SYSTEM DESIGN

3.1 Oscillator as reverberator, reverberator as oscillator: to think physical

The presented synthesis approach illustrates one crucial insight: the oscillator and the reverberator or other audio effects could be combined into a single entity. Essentially, both are responses of systems with respective inputs: a short impulse results in the creation of an acoustic sound, while a more extended waveform induces a wet output with the reverb of the object. Thanks to this principle, the control parameters are no longer limited by their conservative paradigm. Parameters for these two objects thus become interchangeable, which provides the creative works with more flexibility. Most importantly, due to the nature of physical models, one can easily imagine the real-time counterparts of the modeled objects. It could inspire sound design by putting a sound source into a physical object and controlling the reverb with its physical parameters, or reversely synthesizing a sound with terms of reverberation like decay time and room size. Not limited to mimicking real-world objects, the capability of manipulating the identity to provide a wide sounding scale allows artists to find the border between the abstract and concrete form of sound perception and provoke communication between the real and virtual. Moreover, the unity of these two types of objects also benefits the design of ModPhy to be more efficient and controllable, which could also lead to a unified paradigm for music creation and sound design[12].

3.2 Hierarchical design of control parameters

Here we briefly discuss our design of control parameters. We design a hierarchy of control parameters from the raw physical parameters at the bottom to upper-level parameters that are more semantically connected to the way that people understand a sound and its vibration. In the following three-tier structure:

(1) materials, vibration parameters, sound descriptors
(2) stiffness, flexural rigidity, ...
(3) Young's modulus, density, thickness, size, ...

we use only the bottom layer to implement the numerical calculation. The inputs of this most fundamental object are specifications and dimensions that are directly utilized in the physical model, such as Young's modulus, material density, size, etc. However, these parameters are not practically convenient to use, so we added two more layers on top of it. The layer in the middle comprises some parameters to describe the material of the sounding object, such as stiffness and flexural rigidity. This layer serves as a translator from the physical quantities to a level of interpretation that could be connected to the sound with the imagination of the object. The highest-level parameters have the closest association with real-world analogs. In this layer, users can manipulate the parameters of the chosen materials or parameters that describe the process of wave propagation on the object. The formation of the propagation involves multiple factors including the material itself and the boundary condition, therefore it is accomplished by adjusting the two lower layers. The last set of parameters in this layer is the sound descriptors. These descriptors are not as quantified as the timbre descriptors but are qualitatively aligned with the sound quality. This design makes the whole system scalable and also opens up opportunities for customization according to the users' control demands.

3.3 As Max or Pd objects

The hierarchical paradigm with real-time inputs and outputs will be natural for Max MSP and Pd users, and the implementation can also benefit from the design. In PureData, for instance, the bottom layer is realized as the parameters of a DSP object which contains all the computations according to update rules described in Sec 2.2 and can be excited by indicator messages like a bang or DSP objects of excitation or outputs from other synthesis modules. Those DSP objects (both vibrating objects and excitation mechanisms) are FDTD simulations of systems in Sec 2.1 which are implemented with acceleration strategies presented in Sec 2.4, and they support at most 8-channel outputs from single/scanned positions. An example of a bowed string with the external force mentioned in Sec 2.2 can be realized by inserting one outlet of string$\sim$ at a given bowing position to a moving-difference high-pass filter divided by the temporal interval for velocity approximation, then inserting its outlet to bow$\sim$'s inlet, finally inserting bow$\sim$'s outlet to string$\sim$'s inlet. The rest of ModPhy is simply realized as Pd abstractions with inputs and outputs.

To interact with other modules, especially DWG-based synthesis modules, DSP objects for variable conversion like the K/W converter described in Sec 2.3 are also required.

3.4 Advantages of modularization

In a modular system, particularly in realizations using Max MSP or Pd, the flexibility to connect various elements together is readily available. This allows for not only the simulation of musical instruments using pre-built backbone blocks representing vibrating objects but also experimentation with modules employing different synthesis methodologies to observe the resulting sound. Furthermore, these physically-modeled objects can be utilized for cross-synthesis with other objects or input waveforms. It is even possible to introduce feedback from the cross-synthesized sounds back into the input of the physically-modeled objects. However, it is important to exercise caution as there is a risk of numerical instabilities that may arise when implementing such feedback loops. Being mindful of these potential instabilities is crucial to ensure the stability and reliability of the overall system.

4. MUSIC INSPIRATIONS AND APPLICATIONS

From the early 20th century, an increasing number of Western composers embarked on a quest to explore and expand the potential of musical expression through scientific and technological means. Many of their endeavors involved investigating the physical properties of sounds or utilizing the acoustic attributes of sound to give auditory form to non-musical concepts. To provide a clearer demonstration of the acoustic-composition trajectory throughout the 20th century, a selection of notable instances is presented here.

Several composers contributed to the development of this path: Edgar Varèse, a pioneer in the realm of sound, created novel sounds and performance forms while emancipating them from old restrictions by using unusual instruments, non-musical objects, or electronic sounds; Giacinto Scelsi aimed to perceive the third dimension of sound, namely its depth, through the repetitive playing of a single note on the piano; Pierre Schaeffer, drawing inspiration from the study of subjective characteristics of natural sound, developed a quantitative classification system for its features, employed innovatively both gramophone and magnetic recordings of subjective sounds to establish a novel compositional methodology; Iannis Xenakis sonified formulas and concepts from mathematics, statistics, and physics into musical events[13]; spectral composers (e.g. Gérard Grisey, Jonathan Harvey, Tristan Murail, and Julian Anderson), by valuing the inherent nature and properties of sound as a prominent organizing principle for music, sought to take advantage of emerging technologies in order to expand their compositional toolkit.

Along the path of new sound exploration, numerous compelling ideas and remarkable compositions have emerged. Nevertheless, it is crucial to acknowledge the challenges that have arisen during this nearly century-long exploration. When it comes to unraveling the higher dimensions of a single audible sound, identifying new potentials seems impossible. Consequently, it has become increasingly rare to encounter compositions in recent years that are solely dedicated to the exploration of new sounds, let alone introducing novel sound control methods. When it comes to the realm of sound design, it is indeed challenging to see additional approaches for the real-time integration of acoustic and electronic sound, apart from manipulating acoustic sound through electronic devices or resonating physical objects via speakers. Furthermore, relatively few works effectively embody non-musical concepts through sound. While works like Stockhausen's *Gruppen* and Xenakis' *Metastasis* are renowned for their musical excellence, conveying the intended ideas through sound alone may not always be so intuitive - who can actually perceive them? Many composers encounter challenges in successfully sonifying non-musical concepts, resulting from limitations in musical practice, technology, or a lack of understanding regarding the conversion of these concepts into musical practice. These limitations can hinder the process of effective sonification, leading to potential shortcomings in expressing non-musical ideas through sound.

One important thing to note is that the notion of a shared sonic experience holds significant importance for both composers and audiences. Within the Western music canon, countless examples demonstrate the effectiveness of composing in relation to the common sonic experience. Conversely, deliberately challenging and defying that experience can also be impactful, as evidenced by notable instances such as the premiere of Igor Stravinsky's *The Rite of Spring* and John Cage's *4'33"*. In the current musical landscape, the potential for surprise only through the introduction of "new sound" has diminished. The exploration of new sound has become a commonplace practice in music creation, which can sometimes result in clichéd outcomes.

Therefore, adopting an approach that incorporates elements of the common or familiar sonic experience but also plays with and deviates from it can be more intriguing and persuasive as a compositional strategy. Working with ModPhy while composing or sound designing holds considerable potential in addressing longstanding issues through its distinctive insight into reconstructing sound, offering promising solutions to historical challenges in the field. ModPhy provides users with a pathway to interact with sound by manipulating the attributes of a sounding object, including material, shape, size, and more. By assuming the role of a designer of the sounding process, users can exercise control over the sound and bring their sonic ideas to fruition in a swift and efficient manner. ModPhy serves as an empowering tool, enabling users to transform their imaginative concepts in the aspect of acoustic instrument design into tangible sonic realities. When employing ModPhy, one can begin by mimicking the sound of a familiar instrument. By subtly altering factors such as its size, material, or various excitation and parameter settings, users may be able to achieve a comprehensive range of sonic characteristics within the category. When ModPhy is employed in a performance alongside acoustic instruments, it can serve as an extension of their sound, effectively transforming them into a truly innovative and powerful hyper-instrument. By integrating ModPhy's capabilities of modeling real-world physical objects, new sonic dimensions and possibilities emerge. Let's imagine a potential scenario: a percussionist, adorned with motion sensors, is playing a tremolo on a tam-tam while traversing its surface. Concurrently, the physically modeled bowing tam-tam sound produced by ModPhy comes into play. The motion sensors track the percussionist's hand positions and manipulate the excitation points, materials, and thickness of the modeled sound. The fusion of this composite auditory experience with the visual spectacle on stage creates a sonic realm that exists between reality and dream. ModPhy enhances and expands the expressive and sonic potential of the acoustic instruments, resulting in a symbiotic relationship where they collectively form a compelling and versatile hyper-instrument. In addition, taking it to a surrealistic level, users can transcend the limitations of physical reality and explore even more extraordinary possibilities. ModPhy enables users to explore their sonic imaginations with physical perspective, allowing for the creation of otherworldly and dynamically evolving soundscapes that transcend the limitations of traditional acoustic instruments, importantly, while still retaining a connection to the common sonic experience.

5. CONCLUSIONS AND DISCUSSION

In this paper, we present our design of the real-time modular sound synthesis system, ModPhy. With the implementation of physically-modeled objects and a properly designed hierarchy of control parameters, it is possible now to bring their advantages to other sound synthesis techniques in a modular form. Future directions may include analyzing feasible regions of control parameters, encapsulating the acceleration of FDTD schemes as a C/C++ library, and providing more music and sound examples.

6. REFERENCES

[1] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *Journal of the audio engineering society*, vol. 21, no. 7, pp. 526–534, 1973.

[2] S. Bilbao, M. Ducceschi, and F. Zama, "Explicit exactly energy-conserving methods for Hamiltonian systems," *Journal of Computational Physics*, vol. 472, p. 111697, 2023.

[3] C. J. Webb and S. Bilbao, "On the limits of real-time physical modelling synthesis with a modular environment," in *Proceedings of the international conference on digital audio effects*, vol. 65, 2015.

[4] Z. Wang and M. Puckette, "A Fast Algorithm for the Inversion of the Biharmonic in Plate Dynamics Applications," in *Proceedings of the 5th Stockholm Music Acoustic Conference*, Stockholm, Sweden, June 2023.

[5] Z. Wang, M. Puckette, T. Erbe, and S. Bilbao, "Real-Time Implementation of the Kirchhoff Plate Equation using Finite-Difference Time-Domain Methods on CPU," *2023 Sound and Music Computing Conference (SMC)*, 2023.

[6] S. Bilbao, C. Webb, Z. Wang, and M. Ducceschi, "Real-Time Gong Synthesis," *submitted to DAFx*, 2023.

[7] A. Chaigne and J. Kergomard, *Acoustics of musical instruments*. Springer, 2016.

[8] S. Bilbao, *Numerical sound synthesis: finite difference schemes and simulation in musical acoustics*. John Wiley & Sons, 2009.

[9] K. S. Kunz and R. J. Luebbers, *The finite difference time domain method for electromagnetics*. CRC press, 1993.

[10] M. Karjalainen and C. Erkut, "Digital waveguides versus finite difference structures: Equivalence and mixed modeling," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 7, pp. 1–12, 2004.

[11] J. O. Smith III, "On the equivalence of the digital waveguide and finite difference time domain schemes," *arXiv preprint physics/0407032*, 2004.

[12] C. Chafe, "Case studies of physical models in music composition," in *Proceedings of the 18th international congress on acoustics*, 2004.

[13] E. Childs, "ACHORRIPSIS: A SONIFICATION OF PROBABILITY DISTRIBUTIONS," 2002.