IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Strictly Uniform Exponential Decay of the Mixed-FEM Discretization for the Wave Equation with Boundary Dissipation

David C. Del Rey Fernández
Department of Applied Mathematics
University of Waterloo
Waterloo, ON, Canada
ddelreyfernandez@uwaterloo.ca

Luis A. Mora
Department of Applied Mathematics
University of Waterloo
Waterloo, ON, Canada
lmora@uwaterloo.ca

Kirsten Morris
Department of Applied Mathematics
University of Waterloo
Waterloo, ON, Canada
kmorris@uwaterloo.ca

***Abstract*—Uniform preservation of stability in approximations of wave equations is a long-standing issue. In this paper, a one-dimensional wave equation with a partially reflective boundary is approximated using a first-order mixed finite element method. The multiplier method is used to prove that the approximated systems are exponentially stable with a decay rate independent of the mesh size. Upper bounds on the exponential decay are obtained in terms of the physical parameters.**

I. Introduction

Generally, partial differential equations (PDEs) need to be approximated in order to design controllers and estimators. There are many results in the numerical analysis literature on the approximation of PDEs; however, most numerical analyses are concerned with simulation, not controller design which introduces different requirements for a satisfactory approximation. One issue is that spurious high-frequency eigenvalues can occur (see for example, [1], [2]). While not significant in simulation, they negatively affect controller design. Importantly, a uniform stability margin, or more generally uniform stabilizability, is a necessary condition for controller and estimator convergence; see [3] for an overview of existing results.

For systems where the dynamics are governed by an analytic semigroup, higher modes are associated with a greater rate of decay and also less energy. Additionally, the dynamics result in a smoothing effect on the solution. Approximation by a finite-dimensional system is relatively straightforward, and most Galerkin methods yield uniformly stable approximations.

The issue of approximation is much more difficult for systems that do not involve an analytic semigroup, even those that are exponentially stable. In this paper, we consider the wave equation in one space dimension with a partially reflective boundary. Although one of the simpler PDEs, it illustrates the key issues with the approximation of systems lacking dissipation that leads to an analytic semigroup. The original PDE is exponentially stable, with the spectrum lying on a vertical line. The minimum distance between the real part of the eigenvalues and the imaginary axis is the *stability margin.* Finite differences generally fail to preserve the stability margin; e.g. [1], [4], and may even fail to be uniformly stabilizable [5]. Many Galerkin-type approaches, such as standard finite-elements, also fail to preserve the stability margin; see, for instance, [1]. One approach to obtain a uniform decay rate is artificial viscosity [6] but this method adds a term to compensate for the stability error that may be difficult to generalize.

Many PDEs, including the propagation of electrical or acoustic waves, can be derived using Hamilton's principle which provides information about the energy of the system. An extension of Hamilton's principle to systems with inputs and outputs, the port-Hamiltonian (PH) framework [7]–[9], is a method for handling boundary conditions and interconnected systems. The wave equation studied here is an example of a PH system. Discretization methods in the PH framework are focused on the preservation of the PH structure, so that the approximated model conserves the energy dissipativity property. However, this does not guarantee the preservation of a stability margin.

In this paper, we prove that a first-order mixed finite-element method (MFEM) yields uniformly exponentially stable approximations. The scheme uses linear splines as basis functions and piecewise constant splines for test functions. It is equivalent to a finite-volume method and also to a structure-preserving PH discretization in [10]. The approximation method differs from that in [1], [2] in several respects. First, those references use a first-order form in terms of the states $(w, \partial_t w)$ where w is deflection while here the energy realization (p, q) where $p = \partial_t w$ and $q = \partial_x w$ is used. Furthermore, in [1] the term "mixed" refers to using different basis functions for approximating $\partial_t w$ and for w. In [2] the same bases for $\partial_t w$ and w are used as in [1] with unity parameters, but the test functions are different. The resulting approximated wave equation with homogeneous boundary conditions is shown in [2] to be uniformly observable with measurements of both w and $\partial_t w$ at a boundary. In contrast, the dissipative boundary condition considered here corresponds to a single

This work was supported by NSERC and the Faculty of Mathematics, University of Waterloo.

observation. The second-order form of the wave equation with unity parameters is also approximated in [11] but with a finite-difference method. This method yields a decay rate similar to ours in the case of unity parameters.

The decay rate of the approximated state can be determined by computing the eigenvalues of the system matrix of the discretization. Obtaining an analytic solution for the eigenvalues of a matrix with an arbitrary size is challenging. To avoid eigenvalues computations, we use the multiplier approach to obtain a bound for the exponential decay rate of the approximated system. Specifically, the wave equation is considered with physical parameters and also a general multiplier function. The approach presented in this paper demonstrates the utility of the (discrete) multiplier approach and represents a concrete step towards extension to more general problems.

II. CONTINUOUS SYSTEM

Consider the wave equation, in first-order form, on the interval $x \in [a, b]$ of length $\ell = b - a$, with state variables $q(x, t)$ and $p(x, t)$ and physical parameters $\tau > 0$ and $\rho > 0$

$$\partial_t q(x,t) = \frac{-1}{\rho}\partial_x p(x,t),\ \partial_t p(x,t) = -\tau \partial_x q(x,t) \tag{1}$$

with boundary conditions

$$p(a,t) = 0, \quad \tau q(b,t) - \beta \frac{p(b,t)}{\rho} = 0, \tag{2}$$

and energy given by

$$\mathcal{E}(t) = \frac{1}{2}\int_a^b \tau (q(x,t))^2 + \frac{(p(x,t))^2}{\rho} \mathrm{d}x. \tag{3}$$

Differentiating $\mathcal{E}$ along trajectories results in $\dot{\mathcal{E}}(t) = -\frac{\beta}{\rho^2}(p(b,t))^2$, thus, the energy is non-increasing. However, this does not imply exponential stability. While for this simple system, the eigenvalues can be calculated and used to prove exponential stability (e.g. [3, Eg. 3.30]), for more complex systems such an approach is generally infeasible.

The multiplier method [12] has emerged as a useful approach to proving exponential stability for PDEs without the need to compute eigenvalues. The approach relies on the construction of an auxiliary Lyapunov functional

$$V(t) = \mathcal{E}(t) + \varepsilon w(t) \tag{4}$$

where $w(t)$ is a functional that depends on the state variables and a multiplier function $m(x)$. The key to this analysis is the choice of $w(t)$ and the multiplier function $m(x)$. For this system, choosing multiplier function $m(x) = x - a$, setting $\varepsilon > 0$, and

$$w(t) = \int_a^b p(x,t) m(x) q(x,t) \mathrm{d}x$$

results in V decaying exponentially and thus, $\mathcal{E}(t) \le M e^{-\alpha t}\mathcal{E}(0)$, where $M = \frac{\varepsilon_0 + \varepsilon}{\varepsilon_0 - \varepsilon}$ and $\alpha = \frac{c\varepsilon\varepsilon_0}{\varepsilon + \varepsilon_0}$ for any $\varepsilon \in [0, \min\{\varepsilon_0, \varepsilon_1\}]$, with $c = \min_{x\in[a,b]} \partial_x m(x)$, $\varepsilon_0 = \frac{\min(\tau, 1/\rho)}{\ell}$ and $\varepsilon_1 = \frac{2\beta\tau}{\ell(\beta^2 + \tau\rho)}$. Details on the multiplier approach can be found [13, chap. 8] for the second-order form of the wave equation and [14] for the port-Hamiltonian form. An extension to general port-Hamiltonian systems with different multiplier functions is given in [15].

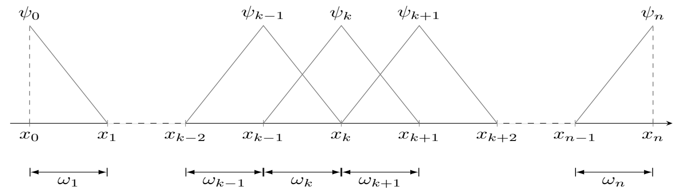

Fig. 1. Mixed finite elements basis and test functions

III. DISCRETIZED SYSTEM

Consider a uniform partition of the spatial domain $[a, b]$ into n elements of length $h = (b-a)/n = \ell/n$. Define $x_k = a + kh$ and the basis

$$\psi_k(x) = \begin{cases} \frac{1}{h}(x - x_{k-1}), & x_{k-1} \le x < x_k \\ \frac{1}{h}(x_{k+1} - x), & x_k \le x \le x_{k+1} \\ 0, & \text{otherwise.} \end{cases}$$

Then, $q(x, t)$ and $p(x, t)$ are approximated by

$$q_h(x,t) = \sum_{k=0}^{n} q_k(t)\psi_k(x),\ p_h(x,t) = \sum_{k=0}^{n} p_k(t)\psi_k(x).$$

Consider the test functions

$$\omega_k(x) = \begin{cases} \frac{1}{h}, & x_{k-1} \le x \le x_k \\ 0, & \text{otherwise} \end{cases} \quad k = 1 \dots n,$$

$$\int_0^\ell \omega_k(x) dx = 1, \quad \psi_k' = \begin{cases} \omega_k, & x_{k-1} \le x < x_k \\ -\omega_{k+1}, & x_k \le x \le x_{k+1} \\ 0, & \text{otherwise} \end{cases}.$$

The wave equation is approximated by

$$\begin{aligned} \langle \dot{q}_h(x,t), \omega_k \rangle &= -\left\langle \partial_x \left(\frac{p_h(x,t)}{\rho} \right), \omega_k \right\rangle,\ k = 1 \dots n \\ \langle \dot{p}_h(x,t), \omega_k \rangle &= -\left\langle \partial_x \left(\tau q_h(x,t) \right), \omega_k \right\rangle,\ k = 1 \dots n. \end{aligned} \tag{5}$$

This leads to equations

$$\frac{\dot{q}_k + \dot{q}_{k-1}}{2} = \frac{p_{k-1} - p_k}{\rho h}, \qquad k = 1 \dots n \tag{6}$$

$$\frac{\dot{p}_k + \dot{p}_{k-1}}{2} = \frac{\tau}{h}(q_{k-1} - q_k), \qquad k = 1 \dots n. \tag{7}$$

To satisfy the boundary conditions, set $p_0 = 0$ and $\tau q_n = (\beta/\rho)p_n$. The state vectors are $\boldsymbol{q} = \begin{bmatrix} q_0 & \dots & q_{n-1} \end{bmatrix}^\top$ and $\boldsymbol{p} = \begin{bmatrix} p_1 & \dots & p_n \end{bmatrix}^\top$. Define a lower diagonal $n \times n$ matrix L and vector of length n, $\boldsymbol{t}_r$,

$$L_{jk} = \begin{cases} 1, & k = j - 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \boldsymbol{t}_{r,j} = \begin{cases} 1, & j = n \\ 0, & \text{otherwise} \end{cases}.$$

Defining matrices

$$M = \frac{1}{2}(I + L^\top) \quad \text{and} \quad D = (I - L), \tag{8}$$

the discrete system can be written

$$\begin{bmatrix} M & \frac{\beta}{2\rho\tau}\boldsymbol{t}_r\boldsymbol{t}_r^\top \\ 0 & M^\top \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{\rho h}D \\ \frac{\tau}{h}D^\top & -\frac{\beta}{\rho h}\boldsymbol{t}_r\boldsymbol{t}_r^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{bmatrix}. \tag{9}$$

The discretized energy is

$$\mathcal{E}_d(t) = \frac{h}{2\rho}\left\langle M^\top \boldsymbol{p}, M^\top \boldsymbol{p}\right\rangle + \frac{\tau h}{2}\left\langle M\boldsymbol{q} + \frac{\beta}{2\rho\tau}\boldsymbol{t}_r\boldsymbol{t}_r^\top \boldsymbol{p}, M\boldsymbol{q} + \frac{\beta}{2\rho\tau}\boldsymbol{t}_r\boldsymbol{t}_r^\top \boldsymbol{p}\right\rangle. \tag{10}$$

System (1)–(3) is a port-Hamiltonian formulation of the wave equation [9], [16]. We now show that the discrete system (9) also is a port-Hamiltonian system and, as a result, preserves the structure of the continuous systems.

Lemma III.1. *The matrices M and D defined in* (8) *satisfy the following identities:*

a) $M^\top D = DM^\top$
b) $M^{-1}D^\top = D^\top M^{-1}$
c) $\frac{1}{2}D^\top + M = I$
d) for any $K = K^\top \in \mathbb{R}^{n\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$, $\left\langle \boldsymbol{x}, \left(\frac{1}{2}KD + MK\right)\boldsymbol{x}\right\rangle = \langle \boldsymbol{x}, K\boldsymbol{x}\rangle$.

Proof. Identities (a) and (c) follow by using the definitions of M and D. Identity (a) implies that $D^\top M = MD^\top$. This is used to obtain *b*) by multiplying $M^{-1}D^\top$ by MM^{-1} on the right-hand side, which leads to $M^{-1}D^\top = M^{-1}MD^\top M^{-1} = D^\top M^{-1}$.

Finally, we have that $\left\langle \boldsymbol{x}, \left(\frac{1}{2}KD + MK\right)\boldsymbol{x}\right\rangle = \langle \boldsymbol{x}, K\boldsymbol{x}\rangle + \frac{1}{2}\left\langle \boldsymbol{x}, (L^\top K - KL)\boldsymbol{x}\right\rangle$ for any $\boldsymbol{x} \in \mathbb{R}^n$. Since $K = K^\top$, matrix $L^\top K - KL$ is skew-symmetric and $\left\langle \boldsymbol{x}, (L^\top K - KL)\boldsymbol{x}\right\rangle = 0$, completing the proof. □

These properties allow us to express system (9) as a generalized port-Hamiltonian descriptor system [17], [18], as shown in the following lemma.

Lemma III.2. *Define state* $\boldsymbol{z} = \begin{bmatrix}\boldsymbol{q}\\ \boldsymbol{p}\end{bmatrix}$, $S = \begin{bmatrix}M & \frac{\beta}{2\rho\tau}\boldsymbol{t}_r\boldsymbol{t}_r^\top\\ 0 & M^\top\end{bmatrix}$,

$$J = \frac{1}{h^2}\begin{bmatrix}0 & -DM^\top\\ M^{-1}D^\top & 0\end{bmatrix},\ Q = \begin{bmatrix}h\tau M & \frac{h\beta}{2\rho}\boldsymbol{t}_r\boldsymbol{t}_r^\top\\ 0 & \frac{h}{\rho}M^\top\end{bmatrix}$$

$$\text{and}\quad R = \begin{bmatrix}0 & 0\\ 0 & \frac{\beta}{h^2}M^{-1}\boldsymbol{t}_r\boldsymbol{t}_r^\top M^{-\top}\end{bmatrix}. \tag{11}$$

Then $S^\top Q = Q^\top S$ *and system* (9) *can be expressed as the port-Hamiltonian descriptor system*

$$S\dot{\boldsymbol{z}} = (J-R)Q\boldsymbol{z} \tag{12}$$

and energy (10) *can be written as*

$$\mathcal{E}_d(t) = \frac{1}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle. \tag{13}$$

Proof. The identity $S^\top Q = Q^\top S$ can be verified by routine manipulation, as can (13). Using (11),

$$(J-R)Q = \begin{bmatrix}0 & -\frac{1}{\rho h}D\\ \frac{\tau}{h}M^{-1}D^\top M & \frac{\beta}{\rho}M^{-1}\left(\frac{1}{2}D^\top - I\right)\boldsymbol{t}_r\boldsymbol{t}_r^\top\end{bmatrix}.$$

Then, identities *b*) and *c*) of Lemma III.1 imply that $M^{-1}D^\top M = D^\top$ and $\frac{1}{2}D^\top - I = -M$. This leads to

$$(J-R)Q = \begin{bmatrix}0 & -\frac{1}{\rho h}D\\ \frac{\tau}{h}D^\top & -\frac{\beta}{\rho}\boldsymbol{t}_r\boldsymbol{t}_r^\top\end{bmatrix}$$

completing the proof. □

The time derivative of the energy is

$$\dot{\mathcal{E}}_d(t) = \langle Q\boldsymbol{z}, S\dot{\boldsymbol{z}}\rangle = \langle Q\boldsymbol{z}, (J-R)Q\boldsymbol{z}\rangle = -\langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle = -\frac{\beta}{\rho^2}\left(p_n\right)^2.$$

Thus, this discretization method conserves the dissipation of the continuous system.

IV. DISCRETE SYSTEM STABILITY

In this section, the uniform stability of the discrete system (9) is analyzed. In particular, it is demonstrated that the approximation (9) preserves uniformly the exponential stability of (1)-(2), i.e., there exist constants $M_d \geq 1$ and $\alpha_d > 0$ independent of h such that $\mathcal{E}_d(t) \leq M_d e^{-\alpha_d t}\mathcal{E}_d(0)$.

A discrete version of the multiplier approach in the analysis of the continuous system is used to prove the exponential stability of the discretized model. For this purpose, define —for some matrix W— the following Lyapunov function candidate:

$$V_d(t) = \mathcal{E}_d(t) + \frac{h}{2}\epsilon\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle. \tag{14}$$

The last term in (14) is a discrete version of the term $w(t)$ in $V(t)$ in the continuous system analysis.

The next theorem delineates the conditions on the Lyapunov function candidate and system (9) required in order to guarantee exponential stability.

Theorem IV.1. *Consider the port-Hamiltonian descriptor system* (12) *and function $V_d(t)$ as in* (14). *If there are constants $\delta < 1$, $\epsilon_0 > 0$ and $\epsilon_1 > 0$, such that*

$$h\left|\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle\right| \leq \frac{1}{\epsilon_0}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle, \tag{15}$$

$$\left\langle S\boldsymbol{z}, \left(\frac{1}{2}I + h\begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}(J-R)\right)Q\boldsymbol{z}\right\rangle \leq \frac{\delta}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle + \frac{1}{\epsilon_1}\langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle, \tag{16}$$

then the discrete energy (13) *decays exponentially. More precisely, define* $c = 1-\delta$, $\alpha_d = \frac{c\epsilon\epsilon_0}{\epsilon+\epsilon_0}$ *and* $M_d = \frac{\epsilon_0+\epsilon}{\epsilon_0-\epsilon}$. *Then, for any* $\epsilon \in [0, \min(\epsilon_0, \epsilon_1)]$,

$$\mathcal{E}_d(t) \leq M_d e^{-\alpha_d t}\mathcal{E}_d(0).$$

Proof. The proof follows the lines of that for the continuous system; see [15].

Step 1: Find conditions for which $V_d(t)$ is non-negative. Note that for all $\epsilon \geq 0$

$$\mathcal{E}_d(t) - \frac{h\epsilon}{2}\left|\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle\right| \leq V_d(t),$$

$$V_d(t) \leq \mathcal{E}_d(t) + \frac{h\epsilon}{2}\left|\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle\right|.$$

If condition (15) holds, and using the definition (10),

$$\left(1-\frac{\epsilon}{\epsilon_0}\right)\mathcal{E}_d(t) \leq V_d(t) \leq \left(1+\frac{\epsilon}{\epsilon_0}\right)\mathcal{E}_d(t). \tag{17}$$

As a consequence, $V_d(t)$ is non-negative if $0 \leq \epsilon \leq \epsilon_0$.

Step 2: Determine the conditions for which $\dot{V}_d(t)$ is non-positive and proportional to $\mathcal{E}_d(t)$.

$$\begin{aligned}\dot{V}_d(t) =& \dot{E}_d(t) + h\epsilon \left\langle S\boldsymbol{z}, \begin{bmatrix} 0 & W \\ W^\top & 0 \end{bmatrix} S\dot{\boldsymbol{z}} \right\rangle \\ =& - \langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle \\ &+ h\epsilon \left\langle S\boldsymbol{z}, \begin{bmatrix} 0 & W \\ W^\top & 0 \end{bmatrix} (J-R)Q\boldsymbol{z} \right\rangle .\end{aligned}$$

Adding and subtracting $\frac{\epsilon}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle$ gives

$$\begin{aligned}\dot{V}_d(t) =& -\frac{\epsilon}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle - \langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle \\ &+\epsilon \left\langle S\boldsymbol{z}, \left(\frac{1}{2}I + h\begin{bmatrix} 0 & W \\ W^\top & 0 \end{bmatrix}(J-R)\right) Q\boldsymbol{z} \right\rangle .\end{aligned}$$

If condition (16) holds, then

$$\begin{aligned}\dot{V}_d(t) \leq& -\frac{\epsilon}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle - \langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle \\ &+ \epsilon\left(\frac{\delta}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle + \frac{1}{\epsilon_1}\langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle\right) \\ \leq& -\frac{\epsilon(1-\delta)}{2}\langle S\boldsymbol{z}, Q\boldsymbol{z}\rangle \\ &- \left(1 - \frac{\epsilon}{\epsilon_1}\right)\langle Q\boldsymbol{z}, RQ\boldsymbol{z}\rangle .\end{aligned}$$

Thus, for any $\epsilon \leq \epsilon_1$ we have, defining $c = 1-\delta$,

$$\dot{V}_d(t) \leq -c\epsilon\mathcal{E}_d(t). \tag{18}$$

If $\delta < 1$, then $\dot{V}_d(t)$ is non-positive.

Step 3: Since V_d is non-negative and non-increasing it is a Lyapunov function if $\epsilon \in [0, \min\{\epsilon_0, \epsilon_1\}]$. Finally, using the bounds on V_d (17), $\dot{V}_d(t) \leq -\frac{c\epsilon\epsilon_0}{\epsilon+\epsilon_0}V_d(t)$ and so $V_d(t) \leq e^{-\frac{c\epsilon\epsilon_0 t}{\epsilon+\epsilon_0}}V_d(0)$. From this it can be concluded that

$$\mathcal{E}_d(t) \leq M_d e^{-\alpha_d t}$$

with M_d, α_d as in the theorem statement. □

The goal now is to find a matrix W such that the conditions of Theorem IV.1 hold. Specifically, we search for W such that the same values of α_d and M_d as in the continuous system are obtained. For this purpose, define a matrix C such that, for any $\boldsymbol{f} \in \mathbb{R}^n$ the k-th element of vector $\boldsymbol{g} = C\boldsymbol{f}$ is

$$g_k = \begin{cases} \frac{1}{2}f_2, & k=1 \\ \frac{k}{2}(f_{k+1} - f_{k-1}), & k \in [2, n-1] \\ n(f_n - f_{n-1}), & k = n \end{cases} . \tag{19}$$

If $f_k = f(x_k)$, g_k is an approximation of $(x-a)\partial_x f$ at the the mesh edges. The following lemma is needed to prove the main results of this work.

Lemma IV.2. *Let M, D and C be as defined in* (8) *and* (19), *respectively. Then, for any vectors $\boldsymbol{f} = \begin{bmatrix} f_1 & \dots & f_n\end{bmatrix}^\top \in \mathbb{R}^n$ and $\boldsymbol{g} = \begin{bmatrix} g_1 & \dots & g_n\end{bmatrix}^\top \in \mathbb{R}^n$,*

$$\frac{h\rho}{2}\|M^\top\boldsymbol{f}\|^2 + h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{f}\right\rangle \leq \frac{\ell\rho}{2}(f_n)^2 \tag{20}$$

$$\frac{\tau}{2h}\|D\boldsymbol{g}\|^2 - \frac{\tau}{h}\left\langle D^\top D\boldsymbol{g}, C\boldsymbol{g}\right\rangle - \beta\left\langle \boldsymbol{t}_r\boldsymbol{t}_r^\top\boldsymbol{f}, C\boldsymbol{g}\right\rangle \leq \frac{\beta^2\ell}{2\tau}(f_n)^2, \tag{21}$$

$$|h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{g}\right\rangle| \leq \frac{\ell h\rho^2}{2}\|M^T\boldsymbol{f}\|^2 + \frac{\ell}{2h}\|D\boldsymbol{g}\|^2. \tag{22}$$

Proof. Note that

$$\|M^\top\boldsymbol{f}\|^2 = \frac{1}{4}\left(f_1^2 + \sum_{k=2}^{n}(f_k + f_{k-1})^2\right) \tag{23}$$

$$\|D\boldsymbol{f}\|^2 = f_1^2 + \sum_{k=2}^{n}(f_k - f_{k-1})^2 \tag{24}$$

and the k-th elements of $MM^\top\boldsymbol{f}$ and $D^\top D\boldsymbol{f}$ respectively are

$$(MM^\top\boldsymbol{f})_k = \begin{cases} \frac{1}{4}(2f_1 + f_2), & k=1 \\ \frac{1}{4}(f_{k-1} + 2f_k + f_{k+1}), & k \in [2, n-1] \\ \frac{1}{4}(f_n + f_{n+1}) & k = n. \end{cases} \tag{25}$$

$$(D^\top D\boldsymbol{f})_k = \begin{cases} 2f_1 - f_2, & k=1 \\ -f_{k-1} + 2f_k - f_{k+1}, & k \in [2, n-1] \\ f_n - f_{n+1} & k = n. \end{cases} \tag{26}$$

Using (25) and (19) leads to

$$\begin{aligned} &h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{f}\right\rangle \\ &= \frac{h\rho}{8}\left[(2f_1 + f_2)f_2 + 2n(f_n + f_{n-1})(f_n - f_{n-1})\right] \\ &+ \frac{h\rho}{8}\sum_{k=2}^{n-1} k(f_{k+1} + 2f_k + f_{k-1})(f_{k+1} - f_{k-1}).\end{aligned}$$

Since

$$\begin{aligned}&\sum_{k=2}^{n-1} k(f_{k+1} + 2f_k + f_{k-1})(f_{k+1} - f_{k-1}) \\ &= n(f_n + f_{n-1})^2 - (f_2 + f_1)^2 - \sum_{k=2}^{n}(f_k + f_{k-1})^2\end{aligned}$$

the previous equation can be expressed as

$$\begin{aligned}h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{f}\right\rangle =& \frac{h\rho}{8}\left(-f_1^2 - \sum_{k=2}^{n}(f_k + f_{k-1})^2\right) \\ &- \frac{h\rho n}{8}(f_n - f_{n-1})^2 + \frac{h\rho n}{2}(f_n)^2\end{aligned}$$

Using (23) and $h = \ell/n$ we obtain

$$h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{f}\right\rangle \leq -\frac{h\rho}{2}\|M^\top\boldsymbol{f}\|^2 + \frac{\ell\rho}{2}(f_n)^2$$

leading to (20).

On the other hand,

$$\begin{aligned}\frac{\tau}{h}\left\langle D^\top D\boldsymbol{g}, C\boldsymbol{g}\right\rangle =& \frac{\tau}{2h}\left((2g_1 - g_2)g_2 + 2n(g_n - g_{n-1})^2\right) \\ &+ \frac{\tau}{2h}\sum_{k=2}^{n-1} k(-g_{k-1} + 2g_k - g_{k+1})(g_{k+1} - g_{k-1}).\end{aligned}$$

The last term can be expressed as

$$\begin{aligned}&\sum_{k=2}^{n-1} k(-g_{k-1} + 2g_k - g_{k+1})(g_{k+1} - g_{k-1}) \\ &= -n(g_n - g_{n-1})^2 + (g_2 - g_1)^2 + \sum_{k=2}^{n}(g_k - g_{k-1})^2\end{aligned}$$

leading us to

$$\frac{\tau}{h}\left\langle D^\top D\boldsymbol{g}, C\boldsymbol{g}\right\rangle = \frac{\tau}{2h}\left(g_1^2+\sum_{k=2}^{n-1}(g_{k+1}-g_{k-1})^2\right) + \frac{\tau}{2h}n(g_n-g_{n-1})^2 = \frac{\tau}{2h}\|D\boldsymbol{g}\|^2+\frac{\tau n}{2h}(g_n-g_{n-1})^2$$

by identity (24). Therefore, since $\beta\left\langle \boldsymbol{t}_r\boldsymbol{t}_r^\top \boldsymbol{f}, C\boldsymbol{g}\right\rangle = \beta n f_n(g_n-g_{n-1})$ and using Young's inequality $|f_n(g_n-g_{n-1})|\le \frac{\tau}{2h\beta}(g_n-g_{n-1})^2+\frac{\beta h}{2\tau}(f_n)^2$,

$$\frac{\tau}{2h}\|D\boldsymbol{g}\|^2-\frac{\tau}{h}\left\langle D^\top D\boldsymbol{g}, C\boldsymbol{g}\right\rangle-\beta\left\langle \boldsymbol{t}_r\boldsymbol{t}_r^\top \boldsymbol{f}, C\boldsymbol{g}\right\rangle = -\frac{\tau n}{2h}(g_n-g_{n-1})^2-\beta n f_n(g_n-g_{n-1}) \le \frac{\beta^2 hn}{2\tau}(f_n)^2.$$

Since $h=\ell/n$, inequality (21) is obtained. Finally,

$$\left|h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{g}\right\rangle\right| = \left|\frac{h\rho}{8}(2f_1+f_2)g_2 + \frac{h\rho n}{4}(f_n+f_{n-1})(g_n-g_{n-1}) + \frac{h\rho}{8}\sum_{k=2}^{n-1}k(f_{k+1}+2f_k+f_{k-1})(g_{k+1}-g_{k-1})\right|.$$

Using Young's inequality $|fg|\le\frac{1}{2c}f^2+\frac{c}{2}g^2$, with an appropriate choice of the positive constant c,

$$|(2f_1+f_2)g_2| \le \frac{\ell\rho}{2}(f_1^2+(f_1+f_2)^2) + \frac{2n^2}{\ell\rho}(g_1-g_2)^2+\frac{2n^2}{\ell\rho}g_1^2$$

$$|n(f_n+f_{n-1})(g_n-g_{n-1})| \le \frac{\ell\rho}{4}(f_n+f_{n-1})^2 + \frac{n^2}{\ell\rho}(g_n-g_{n-1})^2$$

$$|k(f_{k+1}+2f_k+f_{k-1})(g_{k+1}-g_{k-1})| \le \frac{\ell\rho}{2}(f_{k+1}+f_k)^2 + \frac{\ell\rho}{2}(f_k+f_{k-1})^2+\frac{2n^2}{\ell\rho}\left((g_{k+1}-g_k)^2+(g_k-g_{k-1})^2\right).$$

Substituting these inequalities in the equation above and rearranging terms, we obtain

$$\left|h\rho\left\langle MM^\top\boldsymbol{f}, C\boldsymbol{g}\right\rangle\right| \le \frac{\ell h\rho^2}{8}\left[f_1^2+\sum_{k=2}^{n}(f_k+f_{k-1})^2\right] + \frac{hn^2}{2\ell}\left[g_1^2+\sum_{k=2}^{n}(g_k-g_{k-1})^2\right] \le \frac{\ell h\rho^2}{2}\|M^\top\boldsymbol{f}\|^2+\frac{n}{2}\|D\boldsymbol{g}\|^2,$$

leading to (22) and completing the proof. □

The next theorem is the main result of the paper and demonstrates that the discrete system (9) has a strict uniform preservation of the exponential stability.

Theorem IV.3. *Define*

$$W = -hD^{-\top}C^\top M. \tag{27}$$

Then, the inequalities (15)-(16) *hold with*

$$\delta=0,\quad \epsilon_0=\frac{\min(\tau,1/\rho)}{\ell}\quad \textit{and}\quad \epsilon_1=\frac{2\beta\tau}{\ell(\beta^2+\rho\tau)}.$$

Proof. Define $\tilde S=\begin{bmatrix}-\frac{D}{h} & 0\\ 0 & \rho M^\top\end{bmatrix}$, $\tilde{\boldsymbol{z}}=\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix}$, with $\boldsymbol{u}\in\mathbb{R}^n$ and $\boldsymbol{v}\in\mathbb{R}^n$. Then, considering the linear transformation $\mathcal{T}:\boldsymbol{z}\mapsto\tilde{\boldsymbol{z}}$ defined by

$$\tilde{\boldsymbol{z}}(t)=\mathcal{T}\boldsymbol{z}=\tilde S^{-1}S\boldsymbol{z} \tag{28}$$

we obtain that

$$\frac{h}{2}\left|\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle\right| = \frac{h}{2}\left|\left\langle \tilde S\tilde{\boldsymbol{z}}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}\tilde S\tilde{\boldsymbol{z}}\right\rangle\right| = h\rho\left|\left\langle MM^\top\boldsymbol{v}, C\boldsymbol{u}\right\rangle\right|.$$

Using inequality (22) of Lemma IV.2

$$h\rho\left|\left\langle MM^\top\boldsymbol{v}, C\boldsymbol{u}\right\rangle\right| \le \frac{\ell h\rho^2}{2}\|M^\top\boldsymbol{v}\|^2+\frac{\ell}{2h}\|D\boldsymbol{u}\|^2 \le \frac{\ell h}{2}\|\tilde S\tilde{\boldsymbol{z}}\|^2=\frac{\ell h}{2}\|S\boldsymbol{z}\|^2,$$

leading us to

$$\frac{h}{2}\left|\left\langle S\boldsymbol{z}, \begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}S\boldsymbol{z}\right\rangle\right| \le \frac{\ell}{2\min(\tau,1/\rho)}\left\langle S\boldsymbol{z}, Q\boldsymbol{z}\right\rangle$$

by using the inverse transformation $\mathcal{T}^{-1}:\tilde{\boldsymbol{z}}\mapsto\boldsymbol{z}$. This implies that (15) holds with $\epsilon_0=\frac{\min(\tau,1/\rho)}{\ell}$, i.e., ϵ_0 is independent of the mesh size and equal to ε_0.

Writing $Q=KS$ with $K=\begin{bmatrix}\tau h I & 0\\ 0 & \frac{h}{\rho}I\end{bmatrix}$ and using the transformation $\mathcal{T}$,

$$\begin{aligned}&\left\langle S\boldsymbol{z}, \left(\frac{I}{2}+h\begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}(J-R)\right)Q\boldsymbol{z}\right\rangle\\ &=\left\langle S\boldsymbol{z}, \left(\frac{I}{2}+h\begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}(J-R)\right)KS\boldsymbol{z}\right\rangle\\ &=\left\langle \tilde S\tilde{\boldsymbol{z}}(t), \left(\frac{I}{2}+h\begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}(J-R)\right)K\tilde S\tilde{\boldsymbol{z}}(t)\right\rangle\\ &=\frac{\tau}{2h}\|D\boldsymbol{u}\|^2+\frac{h\rho}{2}\|M^\top\boldsymbol{v}\|^2-\frac{\tau}{h}\left\langle D^\top D\boldsymbol{u}, C\boldsymbol{u}\right\rangle\\ &\quad-\beta\left\langle C\boldsymbol{u}, \boldsymbol{t}_r\boldsymbol{t}_r^\top\boldsymbol{v}\right\rangle+h\rho\left\langle MM^\top\boldsymbol{v}, C\boldsymbol{v}\right\rangle,\end{aligned}$$

and using (20) and (21)

$$\left\langle S\boldsymbol{z}, \left(\frac{I}{2}+h\begin{bmatrix}0 & W\\ W^\top & 0\end{bmatrix}(J-R)\right)Q\boldsymbol{z}\right\rangle \le \frac{\ell(\rho\tau+\beta^2)}{2\tau}\left\langle \boldsymbol{v}, \boldsymbol{t}_r\boldsymbol{t}_r^\top\boldsymbol{v}\right\rangle$$

Furthermore, notice that

$$\left\langle \boldsymbol{v}, \boldsymbol{t}_r\boldsymbol{t}_r^\top\boldsymbol{v}\right\rangle = \frac{h^2}{\beta\rho^2}\left\langle \tilde S\tilde{\boldsymbol{z}}(t), R\tilde S\tilde{\boldsymbol{z}}(t)\right\rangle = \frac{h^2}{\beta\rho^2}\left\langle S\boldsymbol{z}, RS\boldsymbol{z}\right\rangle=\frac{1}{\beta}\left\langle Q\boldsymbol{z}, RQ\boldsymbol{z}\right\rangle$$

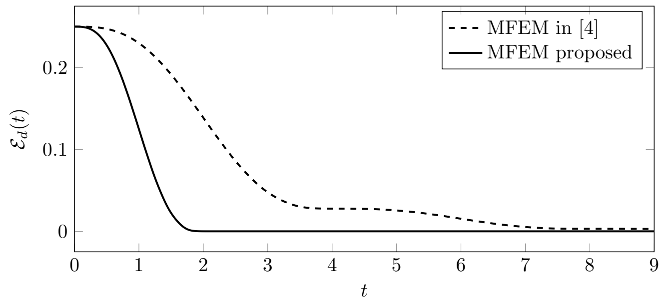

Fig. 2. Comparison of the approximated energy $\mathcal{E}_d(t)$ using the MFEM described in Section III and that in [1], both with $n = 21$ and initial conditions set such that $p(x, 0) = 0$ and $q(x, 0) = \sin\left(\frac{\pi}{2}x\right)$.

by applying $\mathcal{T}^{-1}$. As a consequence, condition (16) holds with $\delta = 0$ and $\epsilon_1 = \frac{2\beta\tau}{\ell(\beta^2+\rho\tau)}$, completing the proof. □

The discrete energy $\mathcal{E}_d(t)$ has been shown to have uniform exponential decay with parameters M_d and α_d equal to those of the continuous system.

The approximation considered here, with a re-definition of variables, is equivalent to the structure-preserving discretization in [10]. A similar discrete system is proposed in [1]. The bound on the exponential decay rate obtained is larger than here. Figure 2 compares the behavior of the two discretized energies for a normalized wave equation with unitary boundary dissipation: $\tau = \rho = \ell = \beta = 1$. With these parameters, the continuous system total energy reaches zero in finite time of 2 [19]. The MFEM used here preserves the exact decay, unlike that in [1].

V. CONCLUSION

In this paper, we analyze a first-order MFEM for approximating the wave equation with one-sided boundary dissipation. In contrast to many other approximation schemes, this method is proven to result in a strictly uniform preservation of exponential decay. This property is important for using a scheme in the development of controller and estimator designs. Furthermore, a decay rate in terms of the original system parameters is obtained. This estimate was obtained using the multiplier approach. Numerically the decay rate appears to match the decay rate of the original PDE, proving this is a topic for future research.

The MFEM is a particular finite-volume method; it is also structure-preserving for this PH system. However, it is a first-order method. This paper demonstrates the utility of the discrete multiplier approach for proving the uniform exponential decay of an approximated system. Current work is aimed at generalizing this result to higher-order methods and to more general hyperbolic PDEs. Examples of this type are the port-Hamiltonian structure-preserving approximations in [10], [20]–[23].

REFERENCES

[1] H. T. Banks, K. Ito, and C. Wang, "Exponentially stable approximations of weakly damped wave equations," in *Estimation and Control of Distributed Parameter Systems* (W. Desch, F. Kappel, and K. Kunisch, eds.), vol. 100 of *International Series of Numerical Mathematics*, pp. 1–33, Birkhäuser Basel, 1991.

[2] C. Castro and S. Micu, "Boundary controllability of a linear semi-discrete 1-D wave equation derived from a mixed finite element method," *Numerische Mathematik*, vol. 102, pp. 413–462, jan 2006.

[3] K. A. Morris, *Controller Design for Distributed Parameter Systems*. Springer, 2020.

[4] E. Zuazua, "Propagation, observation, and control of waves approximated by finite difference methods," *SIAM Review*, vol. 47, p. 197–243, 2005.

[5] G. H. Peichl and C. Wang, "On the uniform stabilizability and the margin of stabilizability of the finite-dimensional approximations of distributed parameter systems," *J. Math. Systems Estimation and Control*, vol. 7, pp. 277–304, 1997.

[6] K. Ramdani, T. Takahashi, and M. Tucsnak, "Uniformly exponentially stable approximations for a class of second order evolution equations," *ESAIM: COCV*, vol. 13, no. 3, pp. 503–527, 2007.

[7] M. Dalsmo and A. J. van der Schaft, "On representations and integrability of mathematical structures in energy-conserving physical systems," *SIAM J. Control and Optim.*, vol. 37, no. 1, pp. 54–91, 1998.

[8] B. Jacob and H. Zwart, *Linear Port-Hamiltonian Systems on Infinite-dimensional Spaces*. Birkhäuser, 2012.

[9] A. J. van der Schaft and B. Maschke, "Hamiltonian formulation of distributed-parameter sytems with boundary energy flow," *Jour. of Geometry & Physics*, vol. 42, pp. 166–194, 2002.

[10] G. Golo, V. Talasila, A. J. van der Schaft, and B. Maschke, "Hamiltonian discretization of boundary control systems," *Automatica*, vol. 40, pp. 757–771, May 2004.

[11] J. Liu and B.-Z. Guo, "A new semidiscretized order reduction finite difference scheme for uniform approximation of one-dimensional wave equation," *SIAM J. Control Optim.*, vol. 58, no. 4, pp. 2256–2287, 2020.

[12] V. Komornik, *Exact Controllability and Stabilization, The Multiplier Method*. Research in Applied Mathematics, Jhon Wiley & Sons, 1994.

[13] M. Tucsnak and G. Weiss, *Observation and Control for Operator Semigroups*. Basel: Birkhäuser Basel, 2009.

[14] L. A. Mora and K. Morris, "Exponential decay rate bound of port-hamiltonian systems in one-dimension with boundary dissipation," in *Proceedings of the IEEE Conference on Decision and Control*, 2022.

[15] L. A. Mora and K. Morris, "Exponential decay rate of linear port-hamiltonian systems. a multiplier approach." submitted.

[16] D. Jeltsema and A. J. Van Der Schaft, "Lagrangian and Hamiltonian formulation of transmission line systems with boundary energy flow," *Reports on Mathematical Physics*, vol. 63, pp. 55–74, feb 2009.

[17] C. Beattie, V. Mehrmann, H. Xu, and H. Zwart, "Linear port-Hamiltonian descriptor systems," *Mathematics of Control, Signals, and Systems*, vol. 30, no. 4, 2018.

[18] A. van der Schaft and B. Maschke, "Generalized port-Hamiltonian DAE systems," *Systems and Control Letters*, vol. 121, pp. 31–37, 2018.

[19] S. Cox and E. Zuazua, "The rate at which energy decays in a string damped at one end," *Indiana University Mathematics Journal*, vol. 44, no. 2, pp. 545–574, 1995.

[20] G. Haine, D. Matignon, and A. Serhani, "Numerical analysis of a structure-preserving space-discretization for an anisotropic and heterogeneous boundary controlled N-dimensional wave equation as a port-Hamiltonian system," *Int. J. Numer. Anal. Model.*, vol. 20, no. 1, pp. 92–133, 2023.

[21] P. Kotyczka, B. Maschke, and L. Lefèvre, "Weak form of Stokes-Dirac structures and geometric discretization of port-Hamiltonian systems," *J. Comput. Phys.*, vol. 361, pp. 442–476, 2018.

[22] A. Macchelli, "Energy shaping of distributed parameter port-hamiltonian systems based on finite element approximation," *Systems and Control Letters*, vol. 60, pp. 579–589, 2011.

[23] V. Trenchant, H. Ramirez, Y. Le Gorrec, and P. Kotyczka, "Finite differences on staggered grids preserving the port-Hamiltonian structure with application to an acoustic duct," *Journal of Computational Physics*, vol. 373, pp. 673–697, nov 2018.