Discrete Transparent Boundary Conditions for Parabolic Systems

Andrea Zisowsky *,1 and Matthias Ehrhardt 1

Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, D–10623 Berlin, Germany

Abstract

In this work we construct and analyse *transparent boundary conditions* (TBCs) for general *systems of parabolic equations*. These TBCs are constructed for the fully discrete scheme (θ–method, finite differences), in order to maintain unconditional stability of the scheme and to avoid numerical reflections. The *discrete transparent boundary conditions* (DTBCs) are discrete convolutions in time and are constructed using the solution of the $\mathcal{Z}$–transformed *exterior problem*. We will analyse the numerical error of these convolution coefficients caused by the inverse $\mathcal{Z}$–transformation. Since the DTBCs are non–local in time and thus very costly to evaluate, we present approximate DTBCs of a sum–of–exponentials form that allow for a fast calculation of the boundary terms. Finally, we will use our approximate DTBCs for an example of a *fluid stochastic Petri net* and present numerical results.

Key words: parabolic systems, unbounded domains, discrete transparent boundary condition, finite difference method

1 Introduction

In this work we consider the numerical solution of *parabolic systems* posed on an unbounded domain. Therefore the computational domain must be restricted by introducing *artificial boundary conditions*. These artificial BCs are

* Corresponding author.
Email addresses: zisowsky@math.tu-berlin.de (Andrea Zisowsky), ehrhardt@math.tu-berlin.de (Matthias Ehrhardt).
URLs: http://www.math.tu-berlin.de/~zisowsky/ (Andrea Zisowsky), http://www.math.tu-berlin.de/~ehrhardt/ (Matthias Ehrhardt).
[1] Supported by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin.

called *transparent boundary conditions* (TBCs), if the solution on the whole–space (restricted to the computational domain) is equal to the solution with the artificial BCs. The artificial boundary splits the problem into three parts: the interesting interior problem and a left and right exterior problem. For constant coefficients the exterior problems can be solved explicitly by the Laplace method. Assuming (spatial) C^1-continuity of the solution at the artificial boundaries yields the TBC as a Dirichlet–to–Neumann map. An ad-hoc discretisation of these continuous TBCs can destroy the stability of the employed numerical scheme for the PDE and induce numerical reflections. To avoid this, we derive *discrete TBCs* (DTBCs) for the fully discretised PDE. The procedure is analogous to the continuous case and uses the $\mathcal{Z}$-transformation. The inverse Laplace/$\mathcal{Z}$-transformation yields a convolution in time. Hence, the perfectly *exact* BC is non-local in time and therefore very costly for long–time simulations. To reduce the numerical effort, we introduce *approximate DTBCs*. Since the inverse $\mathcal{Z}$-transformation must be accomplished numerically for systems, an additional small numerical error is induced.

For *scalar parabolic equations* research results are already advanced (cf. [1, Chap. 2]) and DTBCs give outstanding results. In [2] Halpern developed a family of artificial boundary conditions for the linear convection–diffusion equation with small diffusion. This work was generalised by Lohéac in [3,4] to the case of a spatial dependent diffusion coefficient. Halpern and Rauch derived in [5] absorbing boundary conditions with variable coefficients, curved artificial boundary and arbitrary convection. The numerical study of this conditions were carried out in [6] by Dubach. Lill generalised in [7] the approach of Engquist and Majda [8] to boundary conditions for a convection–diffusion equation and drops the standard assumption that the initial data is compactly supported inside the computational domain. However, the derived $\mathcal{Z}$–transformed boundary conditions were approximated in order to get local–in–time artificial boundary conditions. In [1, Chap. 2] DTBCs for a general class of finite difference discretisations of a scalar parabolic equation were constructed such that the overall scheme is unconditionally stable and as accurate as the discretised whole–space problem.

For *parabolic systems* there are only few works in this direction (e.g. [9,10] and a special 2×2 model problem was treated in [11]) and to the authors' knowledge none for general parabolic systems. Such vector–valued parabolic equations have a broad range of applications. E.g. they arise in the linearised Navier-Stokes equations [11,7], energy-transport models in semiconductor modelling [12], in mathematical biology, e.g. the dispersal of species [13] or at the analysis of second order fluid stochastic Petri nets (FSPNs) [14] to investigate performance and reliability of models for e.g. software systems [15]. In this last mentioned application field we will give a numerical example.

2 The Transparent Boundary Conditions

For the vector $\mathbf{u} \in \mathbb{R}^d$ we consider the general parabolic system

$$\mathbf{u}_t = \frac{\partial}{\partial x}(\mathbf{A}(x,t)\mathbf{u}_x) + \mathbf{M}(x,t)\mathbf{u}_x + \mathbf{V}(x,t)\mathbf{u}, \qquad x \in \mathbb{R},\ t > 0, \tag{1}$$

where $\mathbf{A}$, $\mathbf{M}$ and $\mathbf{V}$ are real–valued $d \times d$–matrices. We use the following definition of a *parabolic system*:

Definition 1 ([16]) *The system* (1) *is called* parabolic *in* $0 \leq t \leq T$ *if there is a constant* $\delta > 0$ *such that for all* $x \in \mathbb{R}$, $0 \leq t \leq T$ *and for all eigenvalues* κ *of the matrix* $\mathbf{A}$ *holds*

$$\kappa \geq \delta > 0. \tag{2}$$

We will now start to derive the analytic TBCs for the parabolic system (1). In the scalar case the Laplace transformed equation in the exterior domain can be solved explicitly. Afterwards the solution is inverse transformed explicitly, thus yielding the analytic TBCs (cf. [1, Chap. 2]). For systems of equations a Laplace transformation yields a system of ordinary differential equations, that can be reduced to first order. Then the solution of this system can be given in terms of its eigenvalues and eigenvectors. We will prove, that half of the eigenvalues have positive real parts and thus yield solutions increasing for $x \to \infty$; the other half has negative real parts, yielding decreasing solutions. Demanding that the part of the increasing solutions in the right exterior domain vanishes, leads to the transformed right TBC (and analogously for the left TBC). However, for systems the inverse Laplace transform in general cannot be calculated explicitly. Nevertheless, we will present the derivation of the Laplace transformed TBC and show when it exists.

We consider the system (1) in the bounded (computational) domain $[x_L, x_R]$ together with TBCs at $x = x_L$ and $x = x_R$. We will denote the constant parameter matrices in the left and right exterior problem by a superscript L and R respectively, when we need to distinguish between the boundaries. But since the derivation for the left and right TBC is analogous, we focus on the right boundary and omit the superscript R until needed. The TBC at $x = x_R$ is constructed by considering (1) with constant coefficients for $x > x_R$, the so called *right exterior problem*

$$\mathbf{u}_t = \mathbf{A}\mathbf{u}_{xx} + \mathbf{M}\mathbf{u}_x + \mathbf{V}\mathbf{u}, \qquad x > x_R, \tag{3}$$

where the matrices $\mathbf{A} = \mathbf{A}^R$, $\mathbf{M} = \mathbf{M}^R$ and $\mathbf{V} = \mathbf{V}^R$ are constant in x and t. The parabolicity condition (2) then reads:

$$\kappa > 0, \qquad \text{for all eigenvalues } \kappa \text{ of } \mathbf{A}.$$

Thus, we will restrict our considerations to *positive definite* matrices $\mathbf{A}$.

To derive the TBC we make the *basic assumption* that the initial data $\mathbf{u}(x, 0)$ is supported inside the bounded domain $[x_L, x_R]$. We note that a strategy to overcome this restriction could be found in [1, Chap. 1]. We now use the Laplace transformation given by

$$\hat{\mathbf{u}}(x, s) = \int_0^\infty e^{-st}\mathbf{u}(x, t)dt\,, \qquad s = \alpha + i\xi, \quad \alpha > 0, \quad \xi \in \mathbb{R},$$

and obtain from (3) the *transformed right exterior problem*

$$(s\mathbf{I} - \mathbf{V})\,\hat{\mathbf{u}} = \mathbf{A}\hat{\mathbf{u}}_{xx} + \mathbf{M}\hat{\mathbf{u}}_x, \qquad x > x_R. \tag{4}$$

Reducing the order of the differential equation to first order we obtain a system in $\left(\begin{smallmatrix}\hat{\mathbf{u}} & \hat{\mathbf{u}}_x\end{smallmatrix}\right)^T$:

$$\begin{pmatrix}\hat{\mathbf{u}}\\ \hat{\mathbf{u}}_x\end{pmatrix}_x = \begin{pmatrix}\mathbf{0} & \mathbf{I}\\ \mathbf{A}^{-1}(s\mathbf{I} - \mathbf{V}) & -\mathbf{A}^{-1}\mathbf{M}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{u}}\\ \hat{\mathbf{u}}_x\end{pmatrix} = \mathbf{C}\begin{pmatrix}\hat{\mathbf{u}}\\ \hat{\mathbf{u}}_x\end{pmatrix}, \qquad x \geq x_R. \tag{5}$$

We now transform $\mathbf{C}$ into Jordan form with $\mathbf{C} = \mathbf{PJP}^{-1}$, where $\mathbf{P}^{-1}$ contains the left eigenvectors in rows. We sort the Jordan blocks in $\mathbf{J}$ with respect to an increasing real part of the corresponding eigenvalue. Thus $\mathbf{J}$ can be written as $\mathbf{J} = \left(\begin{smallmatrix}\mathbf{J}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{J}_2\end{smallmatrix}\right)$, where $\mathbf{J}_1$ holds all Jordan blocks to eigenvalues with negative real parts and $\mathbf{J}_2$ those with positive real parts. Due to the following Thm. 2 $\mathbf{J}_1$ and $\mathbf{J}_2$ are $d \times d$-matrices. With $\mathbf{P}^{-1} = \left(\begin{smallmatrix}\mathbf{P}_1 & \mathbf{P}_2\\ \mathbf{P}_3 & \mathbf{P}_4\end{smallmatrix}\right)$ equation (5) can be written as

$$\mathbf{P}\begin{pmatrix}\hat{\mathbf{u}}\\ \hat{\mathbf{u}}_x\end{pmatrix}_x = \begin{pmatrix}\mathbf{J}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{J}_2\end{pmatrix}\begin{pmatrix}\mathbf{P}_1\hat{\mathbf{u}} + \mathbf{P}_2\hat{\mathbf{u}}_x\\ \mathbf{P}_3\hat{\mathbf{u}} + \mathbf{P}_4\hat{\mathbf{u}}_x\end{pmatrix}.$$

Obviously, solution components due to the upper equation decrease for $x \to \infty$ and solution components due to the lower equation increase. Extinguishing increasing solutions at the right boundary yields the *transformed right TBC*

$$\mathbf{P}_3\hat{\mathbf{u}} + \mathbf{P}_4\hat{\mathbf{u}}_x = \mathbf{0}, \qquad x = x_R. \tag{6}$$

We will now assert, that the number of eigenvalues associated to increasing and decreasing solutions is equal:

Theorem 2 (Splitting Theorem) *For the eigenvalues λ_j, $j = 1, \ldots, 2d$ of the matrix $\mathbf{C}$ in* (5) *with* $\mathrm{Re}\,(\lambda_1) \leq \mathrm{Re}\,(\lambda_2) \leq \cdots \leq \mathrm{Re}\,(\lambda_{2d})$ *holds*

$$\begin{aligned}
&\mathrm{Re}\,(\lambda_j) < 0, \qquad && j = 1, \ldots, d\\
&\mathrm{Re}\,(\lambda_j) > 0, \qquad && j = n+1, \ldots, 2d\,,
\end{aligned}$$

if $\mathrm{Re}\,(s)$ *is sufficiently large.*

PROOF. We will first show, that there is no purely imaginary eigenvalue λ of $\mathbf{C}$. Therefore, we use the ansatz $\hat{\mathbf{u}} = e^{\lambda x}\mathbf{u}_0$ in (4), which yields

$$\lambda^2 \bar{\mathbf{u}}_0^T \mathbf{A}\mathbf{u}_0 + \lambda \bar{\mathbf{u}}_0^T \mathbf{M}\mathbf{u}_0 + \bar{\mathbf{u}}_0^T \mathbf{V}\mathbf{u}_0 - s|\mathbf{u}_0|^2 = 0. \tag{7}$$

We assume $\lambda = i\beta$, $\beta \in \mathbb{R}$ and consider real parts. Since $\mathbf{A}$ is positive definite, it holds $\lambda^2 \bar{\mathbf{u}}_0^T \mathbf{A}\mathbf{u}_0 < 0$ and thus, if the condition

$$\mathrm{Re}\left(\bar{\mathbf{u}}_0^T \mathbf{V}\mathbf{u}_0 + \lambda \bar{\mathbf{u}}_0^T \mathbf{M}\mathbf{u}_0 - s|\mathbf{u}_0|^2\right) < 0$$

holds, (7) is a contradiction. But since $\frac{\mathbf{V}+\mathbf{V}^T}{2} - \beta\,\mathrm{Im}\left(\frac{\mathbf{M}-\mathbf{M}^T}{2}\right) - \mathrm{Re}\,(s)\mathbf{I}$ is negative definite for $\mathrm{Re}\,(s)$ sufficiently large, this condition is true.

Now, instead of $\mathbf{M}$ consider $\epsilon\mathbf{M}$ in (4) with $\epsilon \in [0,1]$. For $\epsilon = 0$ equation (4) is invariant for $x \to -x$ and thus the number of increasing and decreasing solutions, i.e. the number of eigenvalues of $\mathbf{C}$ with positive and negative real parts must be the same. Now, for ϵ from zero to one, the eigenvalues of $\mathbf{C}$ are continuously depending on ϵ and there exists no purely imaginary eigenvalue for any $\epsilon \in [0,1]$. Thus, for $\epsilon = 1$, still d eigenvalues have positive and d have negative real part. $\square$

If $\mathbf{P}_4$ is regular the TBC (6) can be written in *Dirichlet-to-Neumann form*

$$\hat{\mathbf{u}}_x = \mathbf{D}\hat{\mathbf{u}}, \tag{8}$$

for $\mathbf{D} = \mathbf{P}_4^{-1}\mathbf{P}_3$. The regularity of these matrices is not clear in general and must be asserted for a chosen problem.

An ad-hoc discretisation of this TBC (6) (after a numerical inverse Laplace transformation) can destroy the numerical stability of the employed finite difference scheme and induce unphysical numerical reflections. Therefore, we will derive a discrete version of the TBCs on a completely discrete level.

3 The Discrete Transparent Boundary Conditions

In this section we derive DTBCs for a full discretisation of the whole–space problem (1). For the discretisation we choose a uniform grid with the step sizes Δx in space and Δt in time: $x_j = x_L + j\Delta x, t_n = n\Delta t$ with $j \in \mathbb{Z}$, $n \in \mathbb{N}_0$. We use a general θ-method in time and central differences for the first and second spatial derivatives. With the abbreviation $u_{s,j}^{n+\theta} = (1-\theta)u_{s,j}^n + \theta u_{s,j}^{n+1}$ the discrete system reads

$$\frac{h^2}{k}(\mathbf{u}_j^{n+1} - \mathbf{u}_j^n) = \Delta^+(\mathbf{A}\Delta^-\mathbf{u}_j^{n+\theta}) + \frac{h}{2}\mathbf{M}(\Delta^+ + \Delta^-)\mathbf{u}_j^{n+\theta} + h^2\mathbf{V}\mathbf{u}_j^{n+\theta}\,, j \in \mathbb{Z}. \tag{9}$$

For the scalar parabolic equation Ehrhardt [17,1] derived a DTBC, which is reflection–free on the discrete level and conserves the stability properties of the whole–space θ-scheme. The DTBC has the form of a discrete convolution and the convolution coefficients can be obtained easily by a three-term recurrence formula. Here, our strategy is to mimic the derivation of Sec. 2 on a purely discrete level: To derive the DTBC for (9) we solve the $\mathcal{Z}$–transformed system of difference equations in the exterior domain. Then all its solutions are determined by eigenvalues and eigenvectors, which can be distinguished into decaying and increasing solutions by the absolute value of the involved eigenvalue. We obtain the DTBC by using the fact that the exterior solution decays for $|j| \to \infty$.

We focus again on the right exterior domain $j \geq J$ ($x_J = x_R$); the parameter matrices are constant and the discrete scheme (9) simplifies to

$$\frac{h^2}{k}(\mathbf{u}_j^{n+1} - \mathbf{u}_j^n) = \mathbf{A}\Delta^+\Delta^-\mathbf{u}_j^{n+\theta} + \frac{h}{2}\mathbf{M}(\Delta^+ + \Delta^-)\mathbf{u}_j^{n+\theta} + h^2\mathbf{V}\mathbf{u}_j^{n+\theta}, \qquad (10)$$

for $j \geq J$. Here $\mathbf{A} = \mathbf{A}_R, \mathbf{M} = \mathbf{M}_R$ and $\mathbf{V} = \mathbf{V}_R$ are constant matrices and Δ^+, Δ^- denote the usual forward and backward difference operators. Again, we assume for the initial data $\mathbf{u}_j^0 = 0$ for $j \geq J-1$. Then the $\mathcal{Z}$–transformation

$$\mathcal{Z}\{\mathbf{u}_j^n\} = \hat{\mathbf{u}}_j(z) := \sum_{n=0}^{\infty} z^{-n}\mathbf{u}_j^n\,, \qquad z \in \mathbb{C},\ |z| > \mathcal{R},$$

($\mathcal{R}$ denotes the radius of convergence) transforms (10) to

$$\frac{h^2}{k}\frac{z-1}{\theta z + 1 - \theta}\hat{\mathbf{u}}_j = \mathbf{A}\Delta^+\Delta^-\hat{\mathbf{u}}_j + \frac{h}{2}\mathbf{M}(\Delta^+ + \Delta^-)\hat{\mathbf{u}}_j + h^2\mathbf{V}\hat{\mathbf{u}}_j, \qquad (11)$$

for $j \geq J$. Now we reduce the system of difference equations to first order

$$\begin{pmatrix} \frac{h}{2}\mathbf{M} & \mathbf{A} \\ -\mathbf{I} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \Delta^+\hat{\mathbf{u}}_j \\ \Delta^+\Delta^-\hat{\mathbf{u}}_j \end{pmatrix} = \begin{pmatrix} \frac{h^2}{k}\frac{z-1}{\theta z+1-\theta}\,I - h^2\mathbf{V} & -\frac{h}{2}\mathbf{M} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_j \\ \Delta^-\hat{\mathbf{u}}_j \end{pmatrix},$$

or with the abbreviations $\mathbf{T}^+ := \mathbf{A} + \frac{h}{2}\mathbf{M}$ and $\mathbf{T}^- := \mathbf{A} - \frac{h}{2}\mathbf{M}$

$$\begin{aligned} \Delta^+ \begin{pmatrix} \hat{\mathbf{u}}_j \\ \Delta^-\hat{\mathbf{u}}_j \end{pmatrix} &= \begin{pmatrix} (\mathbf{T}^+)^{-1}\left[\frac{h^2}{k}\frac{z-1}{\theta z+1-\theta}\,I - h^2\mathbf{V}\right] & (\mathbf{T}^+)^{-1}\mathbf{T}^- \\ (\mathbf{T}^+)^{-1}\left[\frac{h^2}{k}\frac{z-1}{\theta z+1-\theta}\,I - h^2\mathbf{V}\right] & (\mathbf{T}^+)^{-1}\mathbf{T}^- - \mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_j \\ \Delta^-\hat{\mathbf{u}}_j \end{pmatrix} \\ &= \widetilde{\mathbf{C}} \begin{pmatrix} \hat{\mathbf{u}}_j \\ \Delta^-\hat{\mathbf{u}}_j \end{pmatrix}. \end{aligned} \qquad (12)$$

We claim, that $\mathbf{T}^+$ and $\mathbf{T}^-$ are positive definite matrices, which can be ensured by a sufficiently small space step size h.

We decompose the Jordan form $\mathbf{J}$ of $\tilde{\mathbf{C}} + \mathbf{I}$ in two blocks $\mathbf{J} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 \end{pmatrix}$, where $\mathbf{J}_1$ holds the eigenvalues with an absolute value smaller than one, $\mathbf{J}_2$ those with an absolute value larger than one. Then (12) reads with the matrix $\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_3 & \mathbf{P}_4 \end{pmatrix}$ of left (possibly generalised) eigenvectors

$$\mathbf{P} \begin{pmatrix} \Delta^+ \hat{\mathbf{u}}_j \\ \Delta^+ \Delta^- \hat{\mathbf{u}}_j \end{pmatrix} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 \end{pmatrix} \begin{pmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_3 & \mathbf{P}_4 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{u}}_j \\ \Delta^- \hat{\mathbf{u}}_j \end{pmatrix}.$$

The eigenvalues in $\mathbf{J}_2$ yield for $j \to \infty$ increasing solutions. Therefore the right *transformed DTBC* reads

$$\mathbf{P}_3 \hat{\mathbf{u}}_J + \mathbf{P}_4 \Delta^- \hat{\mathbf{u}}_J = \mathbf{0}. \tag{13}$$

The transformed DTBC (13) can be written in Dirichlet-to-Neumann form if $\mathbf{P}_4$ is regular

$$\Delta^- \hat{\mathbf{u}}_J = \widehat{\mathbf{D}} \hat{\mathbf{u}}_J,$$

where $\widehat{\mathbf{D}} = -\mathbf{P}_4^{-1} \mathbf{P}_3$. After an inverse $\mathcal{Z}$–transformation the *right DTBC* reads

$$\mathbf{u}_J^{n+1} - \mathbf{u}_{J-1}^{n+1} - \mathbf{D}^0 \mathbf{u}_J^{n+1} = \sum_{m=1}^{n} \mathbf{D}^{n+1-m} \mathbf{u}_J^m. \tag{14}$$

Ehrhardt showed in [1, Chap. 2] for a scalar parabolic equation that the imaginary parts of the convolution coefficients $\{d^n\}$ were not decaying but oscillating. Therefore he introduced the summed coefficients $\{s^n := d^n + d^{n-1}\}$, which decay like $O(n^{-3/2})$ and hence avoid subtractive cancellation in the evaluation of the convolution. For our coefficient matrices $\{\mathbf{D}^n\}$ it seems difficult to rigorously prove the asymptotic behaviour, but empirically the situation is similar to the scalar case: only the *summed coefficients* $\{\mathbf{S}^n := \mathbf{D}^n + \mathbf{D}^{n-1}\}$, $n \geq 1$, $\mathbf{S}^0 := \mathbf{D}^0$ decay. The *right DTBC* then reads

$$\mathbf{u}_J^{n+1} - \mathbf{u}_{J-1}^{n+1} - \mathbf{S}^0 \mathbf{u}_J^{n+1} = \sum_{m=1}^{n} \mathbf{S}^{n+1-m} \mathbf{u}_J^m - \mathbf{u}_J^n + \mathbf{u}_{J-1}^n\,. \tag{15}$$

Now we will justify the splitting of the eigenvalues:

Theorem 3 (Discrete Splitting Theorem) *Of the $2d$ eigenvalues of $\widetilde{\mathbf{C}} + \mathbf{I}$ d have an absolute value strictly larger and d have an absolute value strictly smaller than one, if $\frac{1}{2} \leq \theta \leq 1$, $|z| > 1$, h sufficiently small and either k sufficiently small or $(\mathbf{V} + \mathbf{V}^T)/2$ negative definite,.*

PROOF. The proof is analogous to that of Thm. 2. We will show, that no eigenvalue λ of $\tilde{\mathbf{C}} + \mathbf{I}$ with an absolute value of one exists. As in the continuous case, equation (11) is invariant for $j \to -j$ for $\mathbf{M} = \mathbf{0}$ and a continuity argument proves the splitting.

To investigate the absolute value of the eigenvalues of $\widetilde{\mathbf{C}} + \mathbf{I}$ we insert the ansatz $\hat{\mathbf{u}}_j = \lambda^j \hat{\mathbf{u}}_0$ in (11)

$$\lambda^2 \mathbf{T}^+ \hat{\mathbf{u}}_0 + \mathbf{T}^- \hat{\mathbf{u}}_0 = \lambda \left(\mathbf{T}^+ + \mathbf{T}^- - h^2 \mathbf{V} + \frac{h^2}{k} \frac{z-1}{\theta z + 1 - \theta} \mathbf{I} \right) \hat{\mathbf{u}}_0. \tag{16}$$

We assume $|\lambda| = 1$, consider absolute values of (16) and use the triangle inequality after multiplication with $\bar{\hat{\mathbf{u}}}_0^T$ from the left

$$\bar{\hat{\mathbf{u}}}_0^T (\mathbf{T}^+ + \mathbf{T}^-) \hat{\mathbf{u}}_0 \geq \left| \bar{\hat{\mathbf{u}}}_0^T (\mathbf{T}^+ + \mathbf{T}^-) \hat{\mathbf{u}}_0 - h^2 \bar{\hat{\mathbf{u}}}_0^T \mathbf{V} \hat{\mathbf{u}}_0 + \frac{h^2}{k} \frac{z-1}{\theta z + 1 - \theta} |\hat{\mathbf{u}}_0|^2 \right|,$$

where $\bar{\hat{\mathbf{u}}}_0^T (\mathbf{T}^+ + \mathbf{T}^-) \hat{\mathbf{u}}_0$ is a positive real value. But the absolute value on the r.h.s. is strictly larger than $\bar{\hat{\mathbf{u}}}_0^T (\mathbf{T}^+ + \mathbf{T}^-) \hat{\mathbf{u}}_0$, if

$$\mathrm{Re} \left(-\bar{\hat{\mathbf{u}}}_0^T \mathbf{V} \hat{\mathbf{u}}_0 + \frac{1}{k} \frac{z-1}{\theta z + 1 - \theta} |\hat{\mathbf{u}}_0|^2 \right) > 0,$$

which is a contradiction. The real part of the z-depending term can be written as

$$\mathrm{Re} \left(r \frac{z-1}{\theta z + 1 - \theta} \right) = \frac{h^2}{k} \frac{1}{\theta} \frac{(2 - \frac{1}{\theta}) \left[|z|^2 - \mathrm{Re}\,(z) \right] + \left(\frac{1}{\theta} - 1 \right) \left[|z|^2 - 1 \right]}{|z + \frac{1-\theta}{\theta}|^2}, \tag{17}$$

and thus for k sufficiently small (or for negative definite matrices $\frac{\mathbf{V}+\mathbf{V}^T}{2}$), $\frac{1}{2} \leq \theta \leq 1$ and $|z| > 1$ there exists no eigenvalue with absolute value one and the eigenvalues divide into two equal groups. □

Remark 4 *We used the central difference to discretise the first spatial derivative, since this is possible for any matrix* $\mathbf{M}$. *If* M *is diagonalisable, it can be advantageous to use an upwind discretisation for the convection term. The upwind matrices* $\mathbf{R}$ *and* $\mathbf{I} - \mathbf{R}$ *are determined from* $\mathbf{M}_{diag}$ *the diagonalised* $\mathbf{M} = \mathbf{S}^{-1} \mathbf{M}_{diag} \mathbf{S}$. *This changes the matrices* $\mathbf{T}^+$ *and* $\mathbf{T}^-$ *into*

$$\mathbf{T}^+ = \mathbf{A} + h \mathbf{S}^{-1} \mathbf{M}_{\mathrm{diag}} \mathbf{R} \mathbf{S}$$

and

$$\mathbf{T}^- = \mathbf{A} - h \mathbf{S}^{-1} \mathbf{M}_{\mathrm{diag}} (\mathbf{I} - \mathbf{R}) \mathbf{S},$$

which still must be claimed to be positive definite.

4 The Sum-of-Exponentials Ansatz and the Fast Evaluation of the Convolution-type Boundary Condition

In order to reduce the numerical effort of the boundary convolution (15), it is necessary to make some suitable approximation. We will use the approach of [18] to approximate the coefficients $\widetilde{s}^n_{s,l}$ of the convolution matrix $\mathbf{S}^n$ by the *sum-of-exponentials* ansatz and show a method to evaluate the discrete convolution with the approximated convolution coefficients $\widetilde{a}^n_{s,l}$ efficiently.

4.1 The Sum-of-Exponentials Ansatz

The approximation has to be done for each element in $\mathbf{S}$ separately. We use for each $s, \tau = 1, \dots, d$ the following ansatz

$$\widetilde{s}^n_{s,\tau} \approx \widetilde{a}^n_{s,\tau} := \begin{cases} \widetilde{s}^n_{s,\tau}, & n = 0, \dots, \nu - 1 \\ \sum\limits_{l=1}^{L(s,\tau)} g_{s,\tau,l}\, h^{-n}_{s,\tau,l}, & n = \nu, \nu + 1, \dots \end{cases}, \tag{19}$$

where $L(s, \tau) \in \mathbb{N}$ and $\nu \geq 0$ are tuneable parameters. The approximation quality of this sum-of-exponentials ansatz depends on $L(s, \tau)$, ν and the sets $\{g_{s,\tau,l}\}$ and $\{h_{s,\tau,l}\}$ for all $s, \tau = 1, \dots, d$.

In the following we present a method to calculate these sets for given $L(s, \tau)$ and ν. We consider the formal power series

$$f_{s,\tau}(x) := \widetilde{s}^\nu_{s,\tau} + \widetilde{s}^{\nu+1}_{s,\tau} x + \widetilde{s}^{\nu+2}_{s,\tau} x^2 + \dots, \qquad \text{for } |x| \leq 1\,. \tag{20}$$

If the Padé approximation of (20) $\tilde{f}_{s,\tau}(x) := \frac{n^{(L(s,\tau)-1)}_{s,\tau}(x)}{d^{(L(s,\tau))}_{s,\tau}(x)}$ exists (where the numerator and the denominator are polynomials of degree $L(s, \tau) - 1$ and $L(s, \tau)$ respectively), then its Taylor series $\tilde{f}_{s,\tau}(x) = \widetilde{a}^\nu_{s,\tau} + \widetilde{a}^{\nu+1}_{s,\tau} x + \widetilde{a}^{\nu+2}_{s,\tau} x^2 + \dots$ satisfies the conditions $\widetilde{a}^n_{s,\tau} = \widetilde{s}^n_{s,\tau}$ for $n = \nu, \nu+1, \dots, 2L(s, \tau)+\nu-1$ according to the definition of the Padé approximation rule.

We now explain, how to compute the coefficient sets $\{g_{s,\tau,l}\}$ and $\{h_{s,\tau,l}\}$.

Theorem 5 ([18], Theorem 3.1.) *Let $d^{L(s,\tau)}_{s,\tau}$ have $L(s, \tau)$ simple roots $h_{s,\tau,l}$ with $|h_{s,\tau,l}| > 1$, $l = 1, \dots, L(s, \tau)$. Then*

$$\widetilde{a}^n_{s,\tau} = \sum_{l=1}^{L(s,\tau)} g_{s,\tau,l}\, h^{-n}_{s,\tau,l}\,, \qquad n = \nu, \nu + 1, \dots,$$

where

$$g_{s,\tau,l} := -\frac{n_{s,\tau}^{(L(s,\tau)-1)}(h_{s,\tau,l})}{\left(d_{s,\tau}^{(L(s,\tau))}\right)'(h_{s,\tau,l})}\, h_{s,\tau,l}^{\nu-1} \neq 0\,, \qquad l = 1,\ldots,L(s,\tau)\,.$$

Remark 6 *The asymptotic decay of the* $\widetilde{a}_{s,\tau}^{n}$ *is exponential. This is due to the sum-of-exponentials ansatz* (19) *and the assumption* $|h_{s,\tau,l}| > 1$, $l = 1,\ldots,L(s,\tau)$.

The above analysis permits us to give the following description of the approximation to the convolution coefficients by the representation (19) if we use a $[L(s,\tau)-1|L(s,\tau)]$ Padé approximant to (20): the first $2L(s,\tau)+\nu-1$ coefficients are reproduced exactly; however, the asymptotic behaviour of $\widetilde{s}_{s,\tau}^{n}$ and $\widetilde{a}_{s,\tau}^{n}$ (as $n \to \infty$) differs strongly (algebraic versus exponential decay).

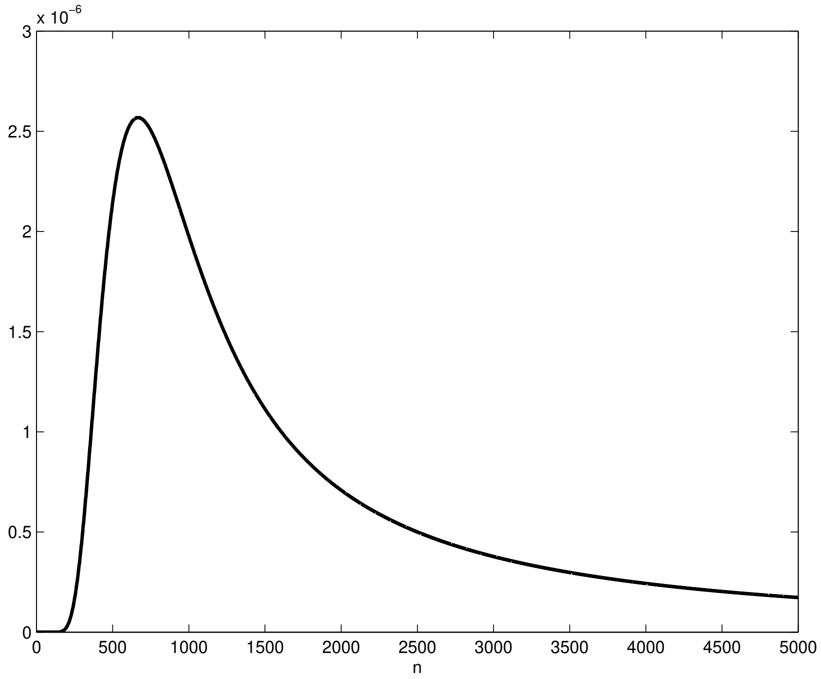

Fig. 1. Error $|\widetilde{s}_{1,1}^{n} - \widetilde{a}_{1,1}^{n}|$ versus n.

We note that the Padé approximation must be performed with high precision ($2L(s,\tau)-1$ digits mantissa length) to avoid a 'nearly breakdown' by ill conditioned steps in the Lanczos algorithm. If such problems still occur or if one root of the denominator is smaller than 1 in absolute value, the orders of the numerator and denominator polynomials are successively reduced. In our numerical test case (see Sec. 6) we started with $L(s,\tau) \equiv 30$, $\nu = 2$ and the finally reached values of $L(s,\tau)$ were between 26 and 30. Figure 1 shows the error $|\widetilde{s}_{1,1}^{n} - \widetilde{a}_{1,1}^{n}|$ versus n for the first diagonal element with the parameters taken from the numerical example of Sec. 6. Clearly, the error increases significantly for $n > 2L(s,\tau)+1$ but remains of moderate size for large values of n.

4.2 The Fast Evaluation of the Approximate Convolution

Now we describe the fast evaluation of the discrete approximate convolution. Let us consider the approximation (19) of the discrete convolution kernel appearing in the DTBC (15). With these "exponential" coefficients the convolution

$$C_{s,\tau}^{(n+1)}(u) := \sum_{m=1}^{n+1-\nu} \widetilde{a}_{s,\tau}^{n+1-m}\, u_{\tau,J}^{m}\,, \ \text{with}\ \widetilde{a}_{s,\tau}^{n} := \sum_{l=1}^{L(s,\tau)} g_{s,\tau,l}\, h_{s,\tau,l}^{-n},\ n=\nu,\nu+1,\dots,$$

$|h_{s,\tau,l}| > 1$, of a discrete function $u_{\tau,J}^{m}$, $m = 1,2,\dots$, with the kernel coefficients $\widetilde{a}_{s,\tau}^{n}$, can be calculated by recurrence formulas, and this will reduce the numerical effort significantly. A straightforward calculation yields:

Theorem 7 ([18], Theorem 4.1.) *The value $C_{s,\tau}^{(n+1)}(u)$ for $n \geq \nu - 1$ is represented by*

$$C_{s,\tau}^{(n+1)}(u) = \sum_{l=1}^{L(s,\tau)} C_{s,\tau,l}^{(n+1)}(u) \tag{21}$$

can be calculated efficiently by a simple recurrence formula:

$$\begin{aligned} C_{s,\tau,l}^{(n+1)}(u) &= h_{s,\tau,l}^{-1}\, C_{s,\tau,l}^{(n)} + g_{s,\tau,l}\, h_{s,\tau,l}^{-\nu}\, u_{\tau,J}^{n+1-\nu}\,, \quad n=\nu-1,\nu,\dots \\ C_{s,\tau,l}^{(\nu)}(u) &\equiv 0\,. \end{aligned} \tag{22}$$

4.3 Summary of the Proposed Method to Evaluate Approximate DTBCs

(1) For each s,τ choose $L(s,\tau)$ and ν and calculate numerically the exact convolution coefficients $\widetilde{s}_{s,\tau}^{n}$ for $n = 0,\dots,2L(s,\tau)+\nu-1$.
(2) For each s,τ use the Padé approximation for the Taylor series with $\widetilde{a}_{s,\tau}^{n} = \widetilde{s}_{s,\tau}^{n}$, for $n=\nu,\nu+1,\dots,2L(s,\tau)+\nu-1$ to calculate the sets $\{g_{s,\tau,l}\}$ and $\{h_{s,\tau,l}\}$ for all $s,\tau = 1,\dots,d$ according to Theorem 5.
(3) Implement the recurrence formulas (21), (22) to calculate the approximate convolutions.

5 Computation of the Convolution Coefficients by Numerical Inverse $\mathcal{Z}$–Transformation

The $\mathcal{Z}$–transformation (or in the analytical case the Laplace transformation) enables us to solve the exterior domain equations for deriving transparent boundary conditions. In the implementation the numerical inverse $\mathcal{Z}$–transformation of the convolution coefficients is a subtle problem due to its inherent instabilities.

In this section we will examine the numerical error caused by the inverse $\mathcal{Z}$–transformation, since it is the crucial point in our numerical implementation. First we shall review the inverse $\mathcal{Z}$–transformation: Assume that the $\mathcal{Z}$–transform of the series $\{\ell_n\}$: $\hat{\ell}(z) = \sum_{n=0}^{\infty} \ell_n z^{-n}$ is analytic for $|z| > \mathcal{R} \geq 0$. The coefficients are then recovered by $\ell_n = \frac{1}{2\pi \mathrm{i}} \oint_{S_\rho} \hat{\ell}(z) z^{n-1} dz$, where S_ρ denotes the circle with radius $\rho > \mathcal{R}$. With the substitution $z = \rho e^{i\varphi}$ we have

$$\ell_n = \frac{\rho^n}{2\pi} \int_0^{2\pi} \hat{\ell}\left(\rho \mathrm{e}^{\mathrm{i}\varphi}\right) \mathrm{e}^{\mathrm{i}n\varphi} d\varphi \,. \tag{23}$$

For $\rho = 1$ this shows that the (inverse) $\mathcal{Z}$–transformation is an isometry between $\{\ell_n\} \in \ell^2(\mathbb{N}_0)$ and $\hat{\ell}\Big|_{|z|=1} \in L^2(0, 2\pi)$.

For $\rho > 1$, however, the *amplification factors* ρ^n in (23) cause numerical instabilities. On the other hand, $\rho = 1$ cannot be chosen either for the application to DTBCs, due to the poor regularity of $\widehat{\mathbf{D}}(z) = -\mathbf{P}_4^{-1}\mathbf{P}_3$ on the unit circle. For the scalar parabolic equation, e.g., $\hat{d}(z)$ has two branch-points of type $\sqrt{z^2 - 1}$ (cf. [1, Chap. 2]), and hence too many quadrature points would be necessary for the numerical evaluation of (23). But $\hat{d}(z)$ is analytic for $|z| > 1$. So, one has to choose ρ as a compromise between more smoothness of $\hat{\ell}\Big|_{|z|=\rho}$ (which allows for an efficient discretisation of (23)), and growing instabilities for large values of ρ.

For the numerical inverse $\mathcal{Z}$–transformation we choose a radius r and N equidistant sampling points $z_p = r\mathrm{e}^{-\mathrm{i}p2\pi/N}$. The approximate inverse transform,

$$\ell_n^N = \frac{1}{N} r^n \sum_{p=0}^{N-1} \hat{\ell}(z_p)\, \mathrm{e}^{\mathrm{i}np\frac{2\pi}{N}} \,, \quad n = 0, \ldots N-1, \tag{24}$$

can then be calculated efficiently by an FFT. The numerical error of ℓ_n^N can be separated into ε_{approx}, the approximation error due to the finite number of sampling points, and the roundoff error ε_{round}, which is amplified by ρ^n. We shall now derive an estimate for this error. Defining $Q_{\hat{\ell}}^{\rho} = \max\limits_{0 \leq \varphi \leq 2\pi} |\hat{\ell}\left(\rho \mathrm{e}^{\mathrm{i}\varphi}\right)|$ gives the estimate

$$|\ell_n| \leq \rho^n Q_{\hat{\ell}}^{\rho} \,. \tag{25}$$

We insert the exact form of $\hat{\ell}_p = \hat{\ell}(z_p)$ into (24), change the order of summation and use the orthogonality property

$$\begin{aligned}
\ell_n^N &= \frac{1}{N} r^n \sum_{m=0}^{\infty} \ell_m r^{-m} \sum_{p=0}^{N-1} \mathrm{e}^{-\mathrm{i}mp\frac{2\pi}{N}} \mathrm{e}^{\mathrm{i}np\frac{2\pi}{N}} \\
&= \frac{1}{N} r^n \sum_{m=0}^{\infty} \ell_m r^{-m} \begin{cases} N & \text{, if } m = n + jN \,, \quad j \in \mathbb{N}_0 \\ 0 & \text{, else} \end{cases} \quad .
\end{aligned}$$

This gives $\quad \ell_n^N - \ell_n = \sum_{p=1}^{\infty} \ell_{n+pN}\, r^{-pN}$. Here, we insert inequality (25) and sum the geometric series, which yields

$$|\ell_n^N - \ell_n| \le \rho^n Q_{\hat{\ell}}^{\rho} \sum_{p=1}^{\infty} \left(\frac{\rho}{r}\right)^{pN} = \rho^n Q_{\hat{\ell}}^{\rho} \frac{\left(\frac{\rho}{r}\right)^N}{1-\left(\frac{\rho}{r}\right)^N}, \qquad \text{for } r > \rho > \mathcal{R}. \tag{26}$$

We remark that similar estimates have been derived in the application of quadrature rules to numerical integration by Lubich, which involve Fourier transformation (cf. [19]).

The other influential error is the *roundoff error* that depends on the machine accuracy ε_m and the accuracy ε in the numerical computation of $\hat{\ell}_p$. For instance, we will use $\tilde{a} = a(1+\varepsilon_m)$ as the computer representation of an exact value a. The roundoff error of the inverse $\mathcal{Z}$–transformation is calculated from equation (24). The main part results from the N fold summation of $\hat{\ell}_p$ and the exponential function:

$$\left|\tilde{\ell}_n^N - \ell_n^N\right| \le r^n \left(CN\varepsilon_m + \varepsilon\right) Q_{\hat{\ell}_p}^r .$$

Together with (26) the error is bounded by

$$|\tilde{\ell}_n^N - \ell_n| \le \rho^n Q_{\hat{\ell}}^{\rho} \frac{\left(\frac{\rho}{r}\right)^N}{1-\left(\frac{\rho}{r}\right)^N} + r^n \left((N+1)\varepsilon_m + \varepsilon\right) Q_{\hat{\ell}_p}^r + O(\varepsilon_m^2 + \varepsilon\varepsilon_m) . \tag{27}$$

We shall illustrate this error behaviour with the numerical example of Sec. 6.

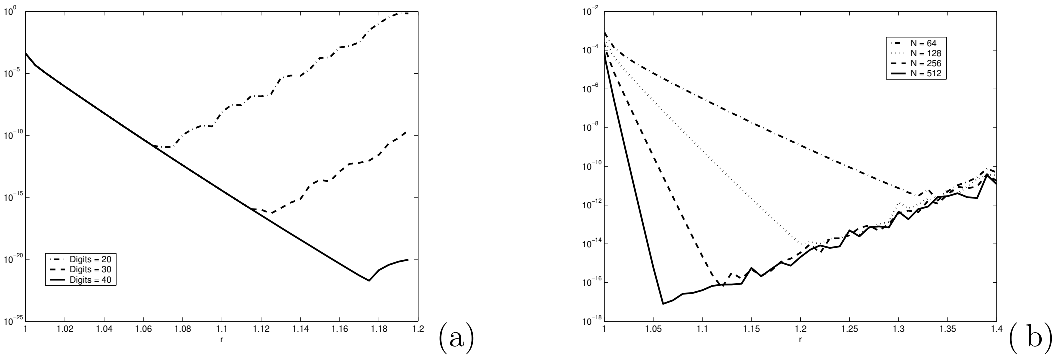

Fig. 2. Error in one element of the matrix $\mathbf{D}$ as a function of the radius r
(a) depending on the number of digits (with $N = 256$ fixed) and
(b) calculated with 20 digits precision depending on the number N of sampling points for the inverse $\mathcal{Z}$–transformation.

We calculated the series $\mathbf{D}^n$ for the queueing system with different accuracies (20, 30 and 40 digits precision) and considered the solution obtained with 50 digits precision as a reference solution. We used $N = 256$ sampling points on the circle. The Euclidean norm of the error is shown in Fig. 2(a) for one of the 36 entries in the matrix $\mathbf{D}$. For all entries the error has the same behaviour:

the error decreases with growing radius, up to a r_{opt}, after which the roundoff error grows rapidly. Observe, that the y-axis of the plot is in logarithmic scale. The curves for 20, 30 and 40 digits coincide for small values of r up to the radius r^{20}_{opt}, r^{30}_{opt} respectively.

Fig. 2(a) shows the influence of the mantissa length on the accuracy of the calculation. Next, we want to show the dependence of the error on the number N of sampling points. Fig. 2(b) shows four error curves with 20 digits precision; one for N= 64, 128, 256 and 512, respectively. The Euclidean norm of the error is summed up to 64. A higher number of sampling points yields a faster decreasing error, r_{opt} becomes smaller and of course the error at r_{opt} becomes less. An influence of N on the round off error is hardly discernable. Comparing the errors at the different N-depending r_{opt} we notice that the gain of taking the double number of points gets less with increasing N. Of course the error cannot become less than the precision in the calculation of $\hat{\ell}_n$.

Since the calculation for a system is rather expensive, it is desirable to predict a radius close to r_{opt}. For the different entries in $\mathbf{D}$ the optimal radius varies only slightly - up to a difference of 0.001. We computed the matrices $\hat{\mathbf{D}}$ and $\hat{\mathbf{S}}$ with MATLAB with an accuracy of $\varepsilon = 10^{-16}$. Thus, with a radius $r = 1.018$ and $N = 2^{12}$ sampling points, we achieve an accuracy of 10^{-8}.

6 Numerical Example

As an illustrating example we consider a second order *fluid stochastic Petri Net*. *Stochastic Petri nets* (SPNs) [20,21] are a tool for describing and studying systems that model time dependent processes. Lately SPNs have been widely used for model–based performance and dependability evaluation of computer and communication systems. Due to the ever increasing complexity of these systems, the size of the state space explodes. Thus, *fluid stochastic Petri nets* (FSPNs) have gained attention to approximate these extremely large state spaces or to model continuous quantities (cf. [22–24]), because FSPNs introduce beside the discrete also a continuous sub–model — both effecting each other. The hybrid net we use here, is defined in [14] and has been used to model computer systems [15] and supervisory control systems [25]. Its transient behaviour is described by the parabolic equation

$$\frac{\partial}{\partial t}\boldsymbol{\pi}(x,t) + \frac{\partial}{\partial x}\left(\boldsymbol{M}(x)\boldsymbol{\pi}(x,t)\right) \;=\; \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(\boldsymbol{\Sigma}^2(x)\boldsymbol{\pi}(x,t)\Big) + \mathbf{Q}^T\boldsymbol{\pi}(x,t), \quad (28)$$

$x \geq x^{\min}$, $t \geq 0$, that is weakly coupled by the generator matrix $\mathbf{Q} \in \mathbb{R}^{d\times d}$, which describes the dynamics of the discrete model part. $\boldsymbol{\pi}(x,t) \in \mathbb{R}^d$ is the vector valued probability density function. The other $d \times d$–matrices are diagonal. $\boldsymbol{M}(x) = \mathrm{d}iag(\mu_1, \ldots, \mu_d)(x)$ and $\boldsymbol{\Sigma}^2(x)$ are the expectation and variance of the fluid flow. The initial boundary value problem is completed by a reflecting barrier at $x = x^{\min}$:

$$\frac{1}{2}\frac{\partial}{\partial x}\left(\boldsymbol{\Sigma}^2(x)\boldsymbol{\pi}(x,t)\right)\Big|_{x=x^{\min}} - \boldsymbol{M}(x)\boldsymbol{\pi}(x,t)\Big|_{x=x^{\min}} = \mathbf{0}, \tag{29}$$

and the *initial condition*

$$\boldsymbol{\pi}(x,0) = \delta(x - x_0)\boldsymbol{\pi}_0, \quad x \geq x^{\min}, \tag{30}$$

where δ denotes the Dirac–Delta distribution. Due to the special structure of (28) it is sensible to use a slightly changed discretisation scheme: the coupling term $\mathbf{Q}^T\boldsymbol{\pi}(x,t)$ is discretised explicitly by $\mathbf{Q}^T((1+\theta)\boldsymbol{\pi}_j^n - \theta\boldsymbol{\pi}_j^{n-1})$, where $\boldsymbol{\pi}_j^n \approx \boldsymbol{\pi}(x_j, t_n)$. Thus, the discrete system is in tridiagonal form and can be solved efficiently. $\frac{\partial}{\partial x}(\boldsymbol{M}(x)\boldsymbol{\pi}(x,t))$ is discretised via the upwind method (see Rem. 4) and a discrete maximum principle holds, that ensures $\mathbf{T}^+$ and $\mathbf{T}^-$ to be positive definite without any restriction on the step size h [26]. For this problem the proof of the discrete splitting theorem holds without any restrictions.

The discrete scheme for state s is

$$\begin{aligned}
&\frac{\pi_{s,j}^{n+1} - \pi_{s,j}^{n}}{k} + (1-\rho_{s,j})\frac{\mu_{s,j}\pi_{s,j}^{n+\theta} - \mu_{s,j-1}\pi_{s,j-1}^{n+\theta}}{h} + \rho_{s,j}\frac{\mu_{s,j+1}\pi_{s,j+1}^{n+\theta} - \mu_{s,j}\pi_{s,j}^{n+\theta}}{h} \\
&= \frac{1}{2}\frac{\sigma_{s,j-1}^2\pi_{s,j-1}^{n+\theta} - 2\sigma_{s,j}^2\pi_{s,j}^{n+\theta} + \sigma_{s,j+1}^2\pi_{s,j+1}^{n+\theta}}{h^2} + \sum_{l=1}^{d}\left((1+\theta)\,\pi_{l,j}^{n} - \theta\pi_{l,j}^{n-1}\right)q_{s,l},
\end{aligned} \tag{31}$$

where $\rho_{s,j}$ is the upwind parameter.

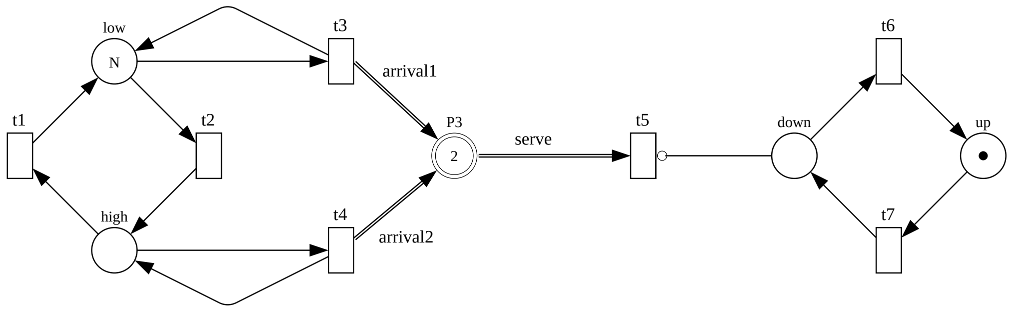

Fig. 3. FSPN of a queueing system with failure and repair

Fig. 3 shows a queueing system, that represents the model of a node in a communication network. Its buffer is approximated by the fluid place P_3. Transitions t_3 and t_4 fire with different rates clients into the system. This imitates the existence of different peak times. If there is a token in place *down*, the firing (with exponentially distributed firing time) of transition t_5 is prevented as would be the case if the server fails. The parameters *arrival1*, *arrival2* and *serve* are defined by the rates of the transitions t_3, t_4 and t_5 respectively. In Fig. 4 we present the *reduced reachability graph* of the queueing system of Fig. 3. We choose $N = 2$ to obtain a concise graph. The expressions *up* and *down* give the partial markings $\#up = 1, \#down = 0$ and $\#up = 0, \#down = 1$ respectively. As well *high*, *high–low* and *low* signify the two places in the left part of the petri net, which currently share at least one of the two tokens, that were initially in place *low*. Transition t_5 appears not in the reduced reachability

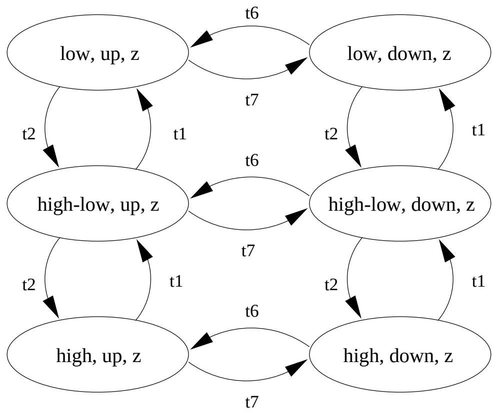

Fig. 4. Reduced reachability graph of the queueing system of Fig. 3 for $N = 2$.

graph, because it is not connected to a discrete place. Its rate influences the fluid flow parameters. For $N = 2$ we get the following fluid parameters, if we enumerate the tangible states in Fig. 4 from left to right:

$$\boldsymbol{M} = \mathrm{diag}\begin{pmatrix} arrival1 - serve \\ arrival1 \\ arrival1 + arrival2 - serve \\ arrival1 + arrival2 \\ arrival2 - serve \\ arrival2 \end{pmatrix} = \mathrm{diag}\begin{pmatrix} -1.2 \\ 0.4 \\ 0.0 \\ 1.6 \\ -0.4 \\ 1.2 \end{pmatrix},$$

$$\boldsymbol{\Sigma}^2 = \mathrm{diag}\begin{pmatrix} arrival1 + serve \\ arrival1 \\ arrival1 + arrival2 + serve \\ arrival1 + arrival2 \\ arrival2 + serve \\ arrival2 \end{pmatrix} = \mathrm{diag}\begin{pmatrix} 2.0 \\ 0.4 \\ 3.2 \\ 1.6 \\ 2.8 \\ 1.2 \end{pmatrix},$$

if we choose $arrival1 = 0.4$, $arrival2 = 1.2$ and $serve = 1.6$. The generator matrix $\mathbf{Q}$ results from the the rates $\lambda_1 = 4.0$, $\lambda_2 = 5.0$, $\lambda_6 = 1.0$, $\lambda_7 = 0.25$

of the transitions (t_1, t_2, t_6 and t_7), which have exponential distribution time and are part of the discrete petri net. Thus $\mathbf{Q}$ evaluates to

$$
\mathbf{Q} = \begin{pmatrix}
-5.25 & 0.25 & 5.0 & 0.0 & 0.0 & 0.0 \\
1.0 & -6.0 & 0.0 & 5.0 & 0.0 & 0.0 \\
4.0 & 0.0 & -9.25 & 0.25 & 5.0 & 0.0 \\
0.0 & 4.0 & 1.0 & -10.0 & 0.0 & 5.0 \\
0.0 & 0.0 & 4.0 & 0.0 & -4.25 & 0.25 \\
0.0 & 0.0 & 0.0 & 4.0 & 1.0 & -5.0
\end{pmatrix} .
$$

At the beginning the system is in the state one shown in Fig. 3. Thus, the initial marking is $\boldsymbol{\pi}_0 = (1, 0, 0, 0, 0, 0)$.

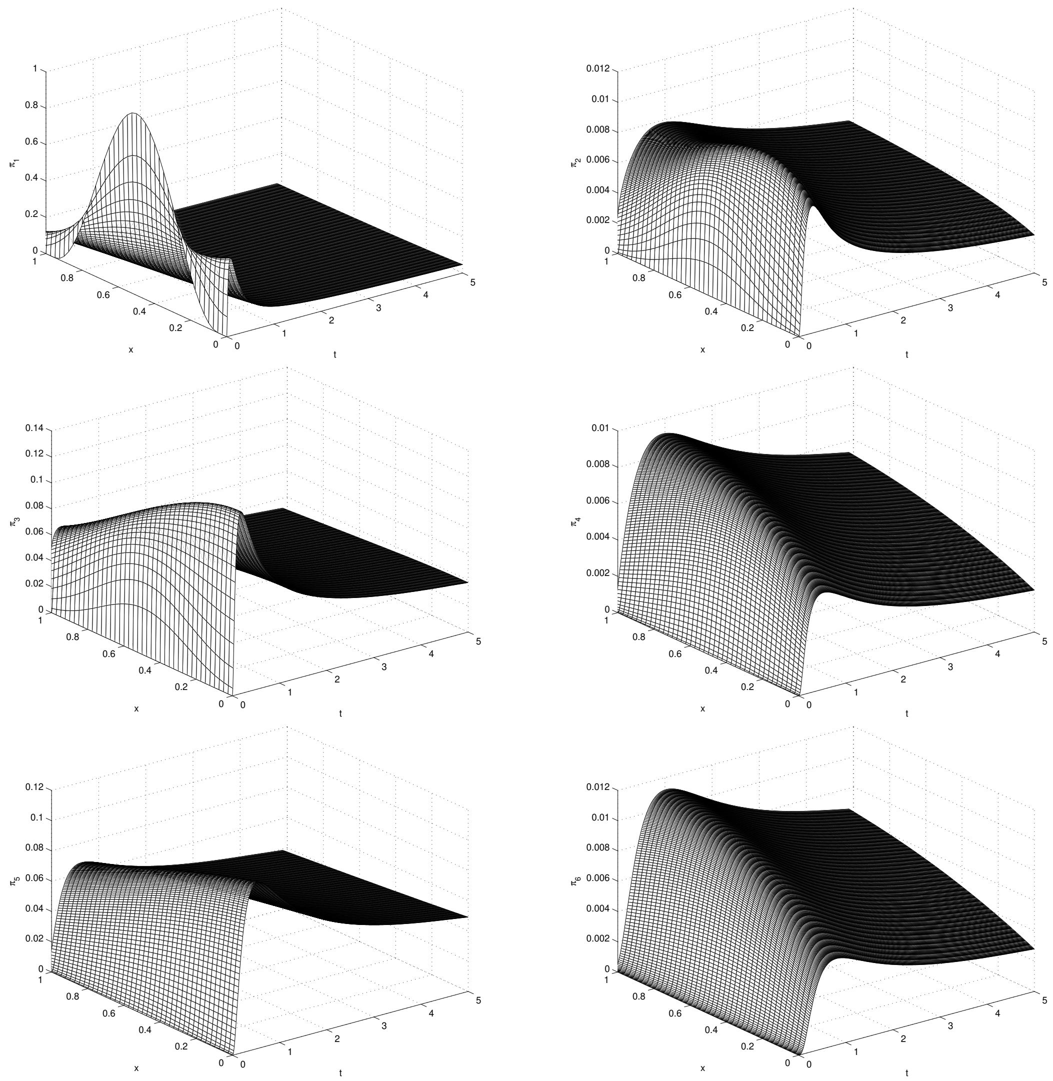

Fig. 5. The numerically calculated density $\pi_1, \ldots, \pi_6$

Fig. 5 shows the xt–diagram of the density function $\boldsymbol{\pi}$ for all six states. We observe, that the mass in the states $s = 2, 4, 6$ moves to the left, what we expected due to $\mu_2, \mu_4, \mu_6 > 0$. The mass moving to the left is interpreted as an increasing number of waiting clients in the system, which grows since for $s = 2, 4, 6$ the server fails and the petri net is in the state "down". Due to the coupling, the mass in state $s = 3$ moves to the right. μ_3 is zero, but the (by the coupling) in–flowing mass comes especially from the states $s = 1$ and $s = 5$ (see above $q_{1,3} = 5$, $q_{5,3} = 4$), which have negative μ.

In Fig. 6 we plotted the discrete l^2-error of the solution π_1^n when using the approximated DTBC of Sec. 4 with $L(s, \tau) = 30$ and $\nu = 2$. The error is comparatively big at the start of the evaluation, when most of the mass leaves the computational domain. Then the error decreases. Due to the approximation the error increases again moderately in time.

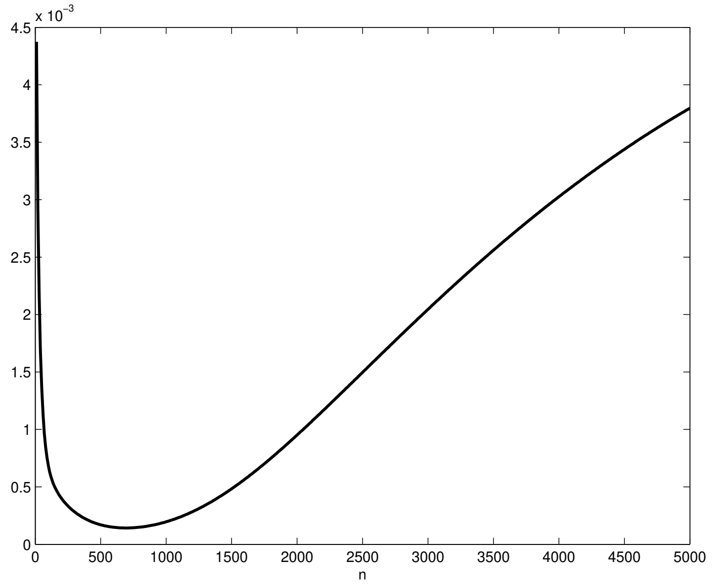

Fig. 6. l^2-error of solution π_1^n with approximate DTBC ($L \approx 30$, $\nu = 2$).

6.1 Stability

Finally, we want to check numerically the stability of the θ-scheme with DTBCs for this example. Therefore, we have to assert that the l^1-norm of the numerical solution does not grow in time and we define

$$||\boldsymbol{\pi}^n||_{l^1} := \sum_{j=0}^{J-1} \sum_{s=1}^{d} |\pi_{s,j}^n| = \sum_{j=0}^{J-1} \sum_{s=1}^{d} \pi_{s,j}^n. \tag{32}$$

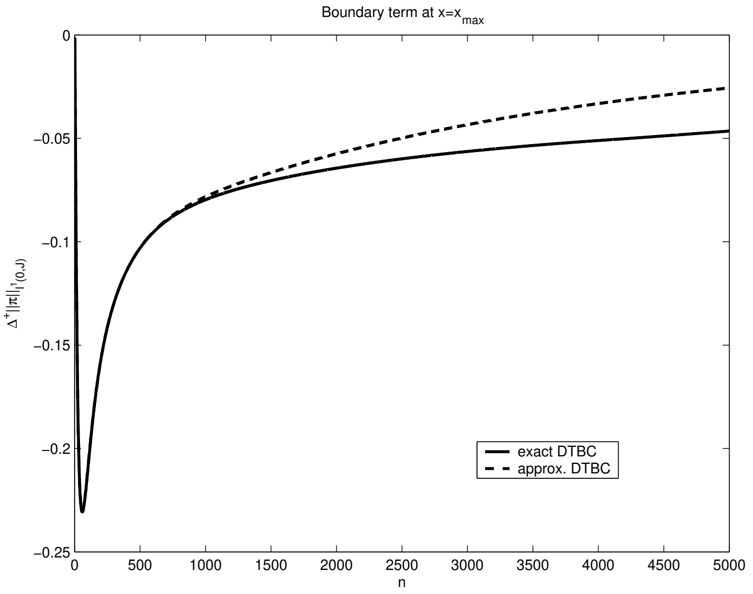

Fig. 7. The time dependent change in the l^1-norm of the solution: $\triangle_t^+ ||\boldsymbol{\pi}^n||_{l^1}$ for exact and approximate ($L \approx 30$, $\nu = 2$) DTBC.

We consider its difference in time and insert the discrete equation (31) using the fact that the row sum of $\mathbf{Q}$ equals zero

$$
\begin{aligned}
\frac{h^2}{k}\triangle^+ ||\boldsymbol{\pi}^n||_{l^1} \;=\;& \sum_{j=1}^{J-1}\sum_{s=1}^{d}\Big(\frac{1}{2}\triangle^+\triangle^-(\sigma_{s,j}^2\pi_{s,j}^{n+\theta}) - h\,\big\{\triangle^+(\mu_{s,j}\rho_{s,j}\pi_{s,j}^{n+\theta}) \\
&+\triangle^-(\mu_{s,j}(1-\rho_{s,j})\pi_{s,j}^{n+\theta})\big\}\Big) + \frac{h^2}{k}\sum_{s=1}^{d}(\pi_{s,0}^{n+1} - \pi_{s,0}^{n}) \\
=& \sum_{s=1}^{d}\left(\sum_{j=2}^{J} T_{s,j}^{+}\pi_{s,j}^{n+\theta} - \sum_{j=1}^{J-1}(T_{s,j}^{+} + T_{s,j}^{-})\pi_{s,j}^{n+\theta} + \sum_{j=0}^{J-2} T_{s,j}^{-}\pi_{s,j}^{n+\theta} + \frac{h^2}{k}(\pi_{s,0}^{n+1} - \pi_{s,0}^{n})\right) \\
=& \sum_{s=1}^{d}\left(-T_{s,1}^{+}\pi_{s,1}^{n+\theta} + T_{s,0}^{-}\pi_{s,0}^{n+\theta} + \frac{h^2}{k}(\pi_{s,0}^{n+1} - \pi_{s,0}^{n}) - T_{s,J-1}^{-}\pi_{s,J-1}^{n+\theta} + T_{s,J}^{+}\pi_{s,J}^{n+\theta}\right) \\
=& \sum_{s=1}^{d}\left(-T_{s,J-1}^{-}\pi_{s,J-1}^{n+\theta} + T_{s,J}^{+}\pi_{s,J}^{n+\theta}\right) \qquad (33)
\end{aligned}
$$

and use an index transformation for the first and third sum over j. The last equality is just the reflecting boundary condition. The abbreviations are $T_{s,j}^{+} = \frac{1}{2}\sigma_{s,j}^2 - h\rho_{s,j}\mu_{s,j}$ and $T_{s,j}^{-} = \frac{1}{2}\sigma_{s,j}^2 + h(1-\rho_{s,j})\mu_{s,j}$. The aim is now, to show that (33) is non–positive. But using the DTBC does not yield any estimate, because our information about properties of the convolution matrix is to small. It remains the possibility to check the sign of (33) numerically. Fig. 7 shows the time dependent change in the l^1-norm of the numerical solution using the exact and the approximate ($L \approx 30$, $\nu = 2$) DTBC. It is negative for each time step $n = 1, \ldots, 5000$. Thus, for this specific discretisation we used a stable scheme.

Conclusion

We have proposed new discrete transparent boundary conditions (DTBCs) for the numerical solution of parabolic systems on unbounded domains. Since the *exact* DTBCs are non–local in the time variable and therefore very costly for long–time simulations we reduced the numerical effort drastically by introducing a 'sum-of-exponential' approximation to the DTBCs. Finally we presented a numerical example in the application to second order fluid stochastic Petri nets.

References

[1] M. Ehrhardt, Discrete artificial boundary conditions, Ph.D. thesis, Technische Universität Berlin (2001).

[2] L. Halpern, Artificial boundary conditions for the linear advection diffusion equation, Math. Comp. 46 (1986) 425–438.

[3] J.-P. Lohéac, An artificial boundary condition for an advection–diffusion problem, Math. Methods Appl. Sci. 14 (1991) 155–175.

[4] J.-P. Lohéac, In– and out–flow artificial boundary conditions for advection–diffusion equations, Z. Angew. Math. Mech. 76 (S5) (1996) 309–310.

[5] L. Halpern, J. Rauch, Absorbing boundary conditions for diffusion equations, Num. Math. 71 (1995) 185–224.

[6] E. Dubach, Artificial boundary conditions for diffusion equations: Numerical study, J. Comput. Appl. Math. 70 (1996) 127–144.

[7] G. Lill, Diskrete Randbedingungen an künstlichen Rändern, Ph.D. thesis, Technische Hochschule Darmstadt (1992).

[8] B. Engquist, A. Majda, Radiation boundary conditions for acoustic and elastic wave calculations, Comm. Pure Appl. Math. 32 (1979) 313–357.

[9] T. Hagstrom, Asymptotic expansions and boundary conditions for time-dependent problems, SIAM J. Numer. Anal. 23 (1986) 948–958.

[10] T. Hagstrom, Asymptotic boundary conditions for dissipative waves: General theory, Math. Comput. 56 (1991) 589–606.

[11] T. Hagstrom, Open boundary conditions for a parabolic system, Math. Comput. Modelling 20 (1994) 55–68.

[12] P. Degond, A. Jüngel, P. Pietra, Numerical discretization of energy-transport models for semiconductors with nonparabolic band structure, SIAM J. Sci. Comput. 22 (3) (2000) 986–1007.

[13] P. White, J. Powell, Spatial invasion of pine beetles into lodgepole forests: a numerical approach, SIAM J. Sci. Comput. 20 (1) (1998) 164–184.

[14] K. Wolter, Performance and dependability modelling with second order fluid stochastic petri nets, Ph.D. thesis, Technische Universität Berlin, Shaker Verlag Aachen (1999).

[15] K. Wolter, A. Zisowsky, On Markov reward modelling with FSPNs, Performance Evaluation 44 (2001) 165–186.

[16] H.-O. Kreiss, J. Lorenz, Initial-Boundary Value Problems and the Navier-Stokes Equations, Vol. 136 of Pure and Applied Mathematics, Academic Press, 1989.

[17] M. Ehrhardt, Discrete transparent boundary conditions for parabolic equations, Z. Angew. Math. Mech. 77 (S2) (1997) 543–544.

[18] A. Arnold, M. Ehrhardt, I. Sofronov, Discrete transparent boundary conditions for the Schrödinger equation: Fast calculation, approximation, and stability, Comm. Math. Sci. 1 (3) (2003) 501–556.

[19] C. Lubich, Convolution Quadrature and Discretized Operational Calculus II, Numer. Math. 52 (1988) 413–425.

[20] M. Ajmone Marsan, Stochastic Petri Nets: an elementary Introduction, in: G. Rozenberg (Ed.), Advances in Petri Nets 1989, Vol. 424 of Lecture Notes in Computer Science, Springer-Verlag, 1990, pp. 1–29.

[21] M. Ajmone Marsan, G. Balbo, G. Chiola, S. Donatelli, G. Francheschinis, Modelling with Generalized Stochastic Petri Nets, John Wiley & Sons, 1995.

[22] G. Horton, V. G. Kulkarni, D. M. Nicol, K. S. Trivedi, Fluid Stochastic Petri Nets: Theory, Applications and Solution, European Journal of Operations Research 105 (1) (1998) 184–201, also published as ICASE report no. 96-5.

[23] G. Ciardo, D. Nicol, K. Trivedi, Discrete-event Simulation of Fluid Stochastic Petri Nets, in: Proc. Seventh International Workshop on Petri Nets and Performance Models - PNPM'97, IEEE-CS Press, Saint Malo, France, 1997, pp. 217–225.

[24] A. Bobbio, S. Garg, M. Gribaudo, A. Horváth, M. Sereno, M. Telek, Modeling Software Systems with Rejuvenation, Restoration and Checkpointing through Fluid Stochastic Petri Nets, in: Proc. Eighth International Workshop on Petri Nets and Performance Models - PNPM'99, Zaragoza, Spain, 1999.

[25] K. Wolter, A. Zisowsky, G. Hommel, Performance models for a hybrid reactor system, in: E. Schnieder, S. Engell (Eds.), Modelling, Analysis, and Design of Hybrid Systems, Vol. 279 of Lecture Notes in Computer Science, Springer-Verlag, 2002, pp. 193–210.

[26] A. Zisowsky, Discrete transparent boundary conditions for systems of evolution equations, Ph.D. thesis, Technische Universität Berlin (2003).