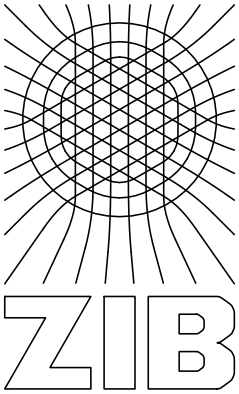
Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany



PATRICK GELSS[1] , SEBASTIAN MATERA[1]
AND CHRISTOF SCHÜTTE[1,2]

[1]*Department of Mathematics and Computer Science, Freie Universität Berlin, Germany*

[2]*Zuse Institute Berlin, Germany*

Solving the master equation without kinetic Monte Carlo: tensor train approximations for a CO oxidation model

ZIB-Report 15–41 (August 2015)

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Solving the master equation without kinetic Monte Carlo: tensor train approximations for a CO oxidation model

Patrick Gelß, Sebastian Matera, Christof Schütte

Dept. of Math. and Comp. Science, FU Berlin, Arnimallee 6, D-14195, Germany

Abstract

In multiscale models of heterogeneous catalysis, one crucial point is the solution of a Markovian master equation describing the stochastic reaction kinetics. This usually is too high-dimensional to be solved with standard numerical techniques and one has to rely on sampling approaches based on the kinetic Monte Carlo method. In this study we break the *curse of dimensionality* for the direct solution of the Markovian master equation by exploiting the Tensor Train Format for this purpose. The performance of the approach is demonstrated on a first principles based, reduced model for the CO oxidation on the $RuO_2(110)$ surface. We investigate the complexity for increasing system size and for various reaction conditions. The advantage over the stochastic simulation approach is illustrated by a problem with increased stiffness.

Keywords: heterogeneous catalysis, master equation, kinetic Monte Carlo, tensor decompositions, tensor train format, alternating linear scheme

2010 MSC: 15A69, 65L05, 65L15, 80A30

1. Introduction

The world's growing demand for more efficient energy and materials exploration or sustainable energy conversion challenges current chemical processing.

Email addresses: p.gelss@fu-berlin.de (Patrick Gelß), matera@math.fu-berlin.de (Sebastian Matera), schuette@mi.fu-berlin.de (Christof Schütte)

Particularly, the field of heterogeneous catalysis is a key technology and examples, where heterogeneous catalysis plays or is expected to play an important role, range from the artificial photo-synthesis [1] to the more classic automotive exhaust gas cleaning [2]. New catalyst materials need to be developed to meet these challenges. A desirable rational design based on a microscopic understanding of the surface chemistry would allow for a more systematic and faster progress into this direction [3]. A central aspect is an understanding of the interplay of the elementary surface reactions making up the catalytic cycle. The prevalent approach to model this interplay is phenomenological microkinetics [4], where a set of rate equations is fitted to experimental, mostly macrokinetic data. Because of the short-comings of this approach with respect to the interpretability in microscopic terms, novel strategies have been established during the last years, which build the reactivity model from bottom up [5, 6]. That is the models are based on (sub-)nanoscale observations and first-principles (i. e. parameter free) simulations. In the end, one arrives at a continuous time, discrete state Markov jump model on the lattice of adsorption sites on the catalyst surface. In these models each jump corresponds to the execution of a particular reaction event changing the occupation of a set of particular sites.

In practice, the state space of these models is very high-dimensional and the corresponding master equation cannot be solved by classical numerical methods due to the *curse of dimensionality*, i.e. the number of unknowns in the master equation grows exponentially with the dimension. The simplest approach to that problem is the *Mean-Field Approximation* (MFA), which yields the classical rate equations. While this set of ordinary differential equations (ODEs) can effectively be solved by implicit time stepping schemes, the physical approximation introduces an uncontrollable error and might lead to qualitatively wrong findings [7, 8]. The only way to achieve numerical (i.e. tunable) accuracy for these problems and to circumvent the curse of dimensionality is the use of kinetic Monte Carlo (kMC, also termed stochastic simulation) methods [9, 10, 11]. The methods simulate trajectories of the stochastic process and the targeted expected values, such as average surface coverage and reactivity, are

then estimated by statistical averaging. The drawback of these methods are a generally slow convergence, the need of multiple realizations or long simulations for stationary problems to ensure sufficient sampling. Further, they severely suffer from stiff problems and the case when rare events are important, which is less problematic for the rate equation based approach. Thus, there is a need for methods, which can overcome the correlation problem of the MFA on the one hand and the stiffness problem of the sampling approach on the other hand, while maintaining the good scaling with dimensionality of both approaches. Attempts into that direction are generalized phenomenological approaches, where the MFA is lifted and additional ODEs for the key spatial correlation are proposed, based on a different closure scheme than the MFA [12, 13]. Problems arise for the identification of the key correlations and the possibly huge number of these.

Because of the limitations of the aforementioned approaches, we follow a different route and try to numerically approximate the solution of the master equation directly. As already stated, this makes it impossible to use standard linear algebra routines, in our showcase below we deal with up to nearly 10^{50} unknowns. The idea is to exploit that the probability distribution is a high order tensor, where the individual entries are not completely random but have a certain structure, for instance we expect that the correlation decays for distant lattice sites. For such tensors, a number of low parametric representations have been developed during the last years, the so-called low rank approximations [14], such as the *Tensor Train Format* (TT-format) [15, 16, 17], which we will employ in this study. The appealing feature of this method is that every tensor can be arbitrarily closely approximated by just increasing the rank, i.e. the parameter space. In every linear sub-problem resulting from the time discretization, we thus approximate the probability distribution with such a representation and increase the rank if necessary. In the context of stochastic processes, the performance of the low-rank approaches has been studied for the *Chemical Master Equations* (CMEs) for well mixed systems [18, 19, 20] or stochastic queuing problems [21, 22]. However, such approaches have not been tested for models

describing heterogeneous catalysis.

As a prototypical showcase, we consider the reduced version of the established model for the CO oxidation at $RuO_2(110)$ [5], which is a popular fruit-fly system in the theoretical study of heterogeneous catalysis [13, 23, 24]. Still, it is derived fully from first-principles and reproduces experimental findings [25, 26] reasonably well. As it is a rather common feature for such models, the problem is very stiff and shows strong correlation [8].

The paper is organized as follows. In Section 2 we give a short introduction of the Markovian modeling of surface reactions and detail the employed model. In Section 3 we explain the Tensor Train format and derive an expression of the master equation with help of tensor products. We close the section by giving a brief overview on how to compute the stationary and time dependent distributions using the *Alternating Linear Scheme* (ALS) [27]. In Section 4, we demonstrate the capabilities of the approach by testing the scaling of the computational complexity with increasing dimensionality and benchmark its results against kinetic Monte Carlo simulations for varying CO partial pressure above the surface. Finally, we compare both approaches for their performance when artificially increasing the stiffness of the problem. Section 5 summarizes our findings and we provide our concluding remarks.

2. Markovian Master Equation

2.1. General Description

On the time scale of molecular motion, chemical reactions are rare transitions from one meta-stable basin to another. Due to rapid motion within each basin, the system has most likely forgotten from which meta-stable state it came from, before the next transition takes place. Thus, the coarse-grained dynamics, only considering the sequence of meta-stable states, can be modeled as Markov jump processes. The probability distribution $P(X,t)$ for being in the state X at time t then obeys a *Markovian Master Equation* (MME) [28]

$$\frac{\partial}{\partial t}P(X,t) = \sum_Y W(X|Y)P(Y,t) - \sum_Y W(Y|X)P(X,t), \tag{1}$$

where $W(Y|X)$ is the transition rate (in units of frequency) to go from state X to state Y. For heterogeneous catalysis, the meta-stable states can often be mapped on the lattice of adsorption sites, where molecules can bind. If we have N different kind of adsorbates $S_1, \ldots, S_N$ and d adsorptions sites, the state can be identified by a vector $X = (x_1, x_2, ..., x_d) \in \{1, \ldots, N\}^d$. Here x_ν denotes the current occupation of site $\nu \in \{1, \ldots, d\}$, i.e. which species is adsorbed on this site. Of course, there are in principle $(N^d)^2$ different entries in the transition matrix $W(Y|X)$. This number is significantly lowered as I) reaction events are local, i.e. an event will only be affected and change the occupations in the vicinity of a particular site, II) the typically considered surfaces have a translational invariance with respect to a shift by a surface unit cell. So while there might be different kinds of adsorptions sites, these will be arranged as a repeating pattern on the surface. So the number of distinct $W(Y|X)$ will be normally very small compared to the size of the state space. The values of the transition rates for reactions events which differ only by a symmetry operation will be called rate constants in the following. The transition rates depend parametrically on the local reaction conditions, e.g. partial pressures in the gas phase and temperature. Usually, we are interested in the limit of very large lattices, as the typical lattice spacing is very small (a few ångström) compared to the size of the catalyst. Therefore, periodic boundary will be employed to mimic an infinite system.

Because X is an integer vector with entries $x_\nu \in \{1, \ldots, N\}$, the state space is finite and the probability distribution can be regarded as a tensor $\mathbf{P}_t \in \mathbb{R}^{N \times \ldots \times N}$ (d times). The number of elements of a tensor grows exponentially with the order d, i.e. $O(N^d)$. Due to this so called curse of dimensionality, storing a d-dimensional tensor and calculating on it may be infeasible for growing d. Therefore, we require special representations of tensors, and, for the problem presented next, we expect the TT-format to perform particularly well. For the later reformulation of the master equation in the TT-format, we rewrite the

MME,

$$\frac{\partial}{\partial t}P(X,t) = \sum_{\mu=1}^{M} a_\mu(X-\xi_\mu)P(X-\xi_\mu,t) - a_\mu(X)P(X,t). \tag{2}$$

Here, the sum runs now over all $M = O(N^d)$ allowed reactions events $\mathrm{R}_1, \ldots, \mathrm{R}_M$. We denote the net changes in the state vector X caused by a single firing of R_μ by the vector $\xi_\mu \in \mathbb{Z}^d$. The so-called *reaction propensity* a_μ is given by

$$a_\mu(X) = W(X+\xi_\mu|X). \tag{3}$$

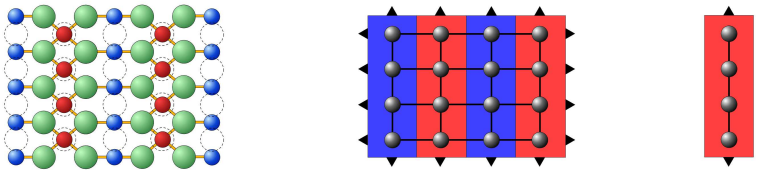
Note that a_μ is only non-zero, if X and $X+\xi_\mu$ are both in $\{1,\ldots,N\}^d$ and X complies with the requirements that R_μ may fire. Hence, formula 2 contains all possible states from which X can be reached and all states that can be reached from X by a single firing of one of the elementary reaction events.

On this abstract level, formula 2 has a same structure as the CME [29]. Still the differences are notable. Despite that here we have no closed formula for the propensities, the state coordinates x_ν further do not represent numbers of molecules, instead they denote the current occupation of the sites. Increasing the system size does not correspond to an increment of N but to an increment of the dimensionality d, and the large scale limit cannot be obtained by the classical system size expansion [28]. The advantage is that we work spatially resolved and corresponding spatial correlation drops out as a simulation result, differently to CMEs where this information must enter the expression for the propensities.

2.2. Reduced model for the CO Oxidation at RuO_2

In this paper, we consider the established microkinetic model for the CO oxidation at $\mathrm{RuO_2}(110)$ by Reuter and Scheffler [5]. The full model lives on a rectangular lattice with alternating columns of so-called *bridge* sites (br) and *coordinatively unsaturated* sites (cus). In figure 1a a top view of the surface is shown. Figure 1b displays its lattice representation, where bridge sites are underlayed in blue and cus sites in red. Each site may be in three different states

(1 = empty, 2 = O-covered, 3 = CO-covered). The possible events are unimolecular adsorption/desorption of CO on either br or cus, dissociative oxygen adsorption on two neighboring sites of any kind and the corresponding reverse processes, diffusion of adsorbed CO/O to a neighboring site, and the formation of gaseous CO_2 from adsorbed CO and O on neighboring sites. There are no lateral interactions between different sites at the surface and each elementary reaction event changes only the occupation of one site ν or of two neighboring active sites ν and $\tilde{\nu}$. In other words, the corresponding propensities depend only on the occupation of one or two sites. Details on how the transition rates entering this model have been estimated from first-principles using density-functional can be found in ref. [5].



(a) (b) (c)

Figure 1: (a) Top view of the $RuO_2(110)$ surface showing the two prominent adsorption sites, bridge sites between the ruthenium atoms in blue and cus sites on the ruthenium atoms in red. (b) 2D lattice model of the coarse-grained surface composed of alternating rows of bridge and cus sites as used in refs. [5, 8]. (c) 1D lattice model completely composed of cus sites.

It has been found that the chemical kinetics predominantly take place only on the cus sites [8, 30]. We therefore restrict to a reduced model, where all reaction involving bridge sites are omitted. We have compared this reduced model with the full one and found both to agree quantitatively for the very most of the considered reaction conditions. Only in very narrow window we observed small deviations, which never are qualitatively different. With the bridge sites turned off, the reduced model consists of non-communicating columns of cus-sites, thus the problem becomes one-dimensional. In practice, we consider a ring of d cus sites. Table 1 summarizes the elementary reactions of the reduced model and

the specific values for the rate constants for the gas phase conditions employed in this study, which are identical to those employed in ref. [8]. In detail, these are a fixed O_2 pressure of $p_{O_2} = 1$ atm, fixed temperature $T = 600$ K, and varying CO pressure $p_{CO} \in [10^{-4}, 10^2]$ atm. As the rate of CO adsorption depends linearly on the CO partial pressure, this range corresponds to $k_{CO}^{\mathrm{Ad}} \in [10^4, 10^{10}]\,\mathrm{s}^{-1}$.

Adsorption							
$\mathrm{R}_{\mathrm{O_2}}^{\mathrm{Ad}}$	:	$\varnothing_\nu + \varnothing_{\tilde{\nu}}$	$\rightarrow$	$\mathrm{O}_\nu + \mathrm{O}_{\tilde{\nu}}$	, $k_{\mathrm{O_2}}^{\mathrm{Ad}}$	=	$9.7 \cdot 10^7 s^{-1}$
$\mathrm{R}_{\mathrm{CO}}^{\mathrm{Ad}}$	:	$\varnothing_\nu$	$\rightarrow$	CO_ν	, $k_{\mathrm{CO}}^{\mathrm{Ad}}$	=	$10^4 - 10^{10} s^{-1}$
Desorption							
$\mathrm{R}_{\mathrm{O_2}}^{\mathrm{De}}$	:	$\mathrm{O}_\nu + \mathrm{O}_{\tilde{\nu}}$	$\rightarrow$	$\varnothing_\nu + \varnothing_{\tilde{\nu}}$	, $k_{\mathrm{O_2}}^{\mathrm{De}}$	=	$2.8 \cdot 10^1 s^{-1}$
$\mathrm{R}_{\mathrm{CO}}^{\mathrm{De}}$	:	CO_ν	$\rightarrow$	$\varnothing_\nu$	, $k_{\mathrm{CO}}^{\mathrm{De}}$	=	$9.2 \cdot 10^6 s^{-1}$
$\mathrm{R}_{\mathrm{CO_2}}^{\mathrm{De}}$	:	$\mathrm{CO}_\nu + \mathrm{O}_{\tilde{\nu}}$	$\rightarrow$	$\varnothing_\nu + \varnothing_{\tilde{\nu}}$	, $k_{\mathrm{CO_2}}^{\mathrm{De}}$	=	$1.7 \cdot 10^5 s^{-1}$
Diffusion							
$\mathrm{R}_{\mathrm{O}}^{\mathrm{Diff}}$	:	$\mathrm{O}_\nu + \varnothing_{\tilde{\nu}}$	$\rightarrow$	$\varnothing_\nu + \mathrm{O}_{\tilde{\nu}}$	, $k_{\mathrm{O}}^{\mathrm{Diff}}$	=	$0.5 s^{-1}$
$\mathrm{R}_{\mathrm{CO}}^{\mathrm{Diff}}$	:	$\mathrm{CO}_\nu + \varnothing_{\tilde{\nu}}$	$\rightarrow$	$\varnothing_\nu + \mathrm{CO}_{\tilde{\nu}}$	, $k_{\mathrm{CO}}^{\mathrm{Diff}}$	=	$6.6 \cdot 10^{-2} s^{-1}$

Table 1: Elementary reaction steps on the cus sites together with their corresponding rate constants, see ref. [5] for details. The reactions are defined on two neighboring sites ν and $\tilde{\nu}$, except for adsorption and desorption of CO, these reactions are defined only on site ν.

3. Tensor Train Format

3.1. Definition and Notation

Tensors are multidimensional generalizations of matrices represented by arrays $T \in \mathbb{R}^{n_1 \times \ldots \times n_d}$, where $n_k \in \mathbb{N}\backslash\{0\}$ for $k = 1, \ldots, d$. The different dimensions n_k of the array are called *modes* and the total number of modes d is the *order* of the tensor. If we want to clarify that all possible entries of a mode are selected we use colons, for example the notation

$$(T)_{x_1,:,x_3,:,x_5,\ldots,x_d} \tag{4}$$

defines a matrix $\tilde{T} \in \mathbb{R}^{n_2 \times n_4}$ where $(\tilde{T})_{ij} = (T)_{x_1,i,x_3,j,x_5,\ldots,x_d}$.

In order to break the curse of dimensionality, we will rely on low parametric representations of tensors. For further details on tensor decompositions and operations we refer to ref. [31].

A tensor $T \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ of order d is called a *rank-one-tensor*, if it can be written as the outer product of d vectors, i.e.,

$$T = T^{(1)} \otimes \ldots \otimes T^{(d)}, \tag{5}$$

where $T^{(i)} \in \mathbb{R}^{n_i}$ for $i = 1, \ldots, d$. Using the above definition, the concept of tensor decompositions was already introduced in 1927 by Hitchcock [32], who presented the idea expressing a tensor as the sum of a finite number of rank-one-tensors. This *polyadic* or *canonical decomposition* [33] is outlined in the following definition.

Definition 1. *A tensor $T \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ is said to be in canonical format, if*

$$\begin{aligned} T &= \sum_{k=1}^{r} T_k^{(1)} \otimes \ldots \otimes T_k^{(d)} \\ &\Leftrightarrow \\ (T)_{x_1,\ldots,x_d} &= \sum_{k=1}^{r} \left(T_k^{(1)}\right)_{x_1} \cdot \ldots \cdot \left(T_k^{(d)}\right)_{x_d}, \end{aligned} \tag{6}$$

with $T_k^{(l)} \in \mathbb{R}^{n_l}$ for $k = 1, \ldots, r$ and $l = 1, \ldots, d$, where r is called the rank *of the decomposition.*

With help of the canonical format we can reduce the storage consumption of an order d tensor which now can be estimated as $O(r \cdot n \cdot d)$, where $n \in \mathbb{N}$ is the maximum of all mode sizes. This means that we do not have to deal with an exponential dependence on the order d anymore. Anyway, algorithms for the computation of best approximations as we need them to find solutions of master equations are not robust, since canonical tensors with bounded rank r do not form a manifold and the optimization problems are ill-posed [34]. Thus, we have to use different formats in order to apply time-stepping schemes or eigenvalue solvers to the master equation 2. A promising candidate is the TT-format which was developed by Oseledets and Tyrtyshnikov in 2009, see refs. [15, 16].

Definition 2. *A tensor $T \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ is said to be in the TT-format, if*

$$\begin{aligned} T &= \sum_{k_0=1}^{r_0} \cdots \sum_{k_d=1}^{r_d} (T^{(1)})_{k_0,:,k_1} \otimes \ldots \otimes (T^{(d)})_{k_{d-1},:,k_d} \\ &\Leftrightarrow \\ (T)_{x_1,\ldots,x_d} &= (T^{(1)})_{:,x_1,:} \cdot \ldots \cdot (T^{(d)})_{:,x_d,:}, \end{aligned} \tag{7}$$

where $T^{(i)} \in \mathbb{R}^{r_{i-1} \times n_i \times r_i}$, $i = 1, \ldots, d$, are called the TT-cores *and the numbers r_i are called the* TT-ranks. *It is $r_0 = r_d = 1$.*

Above, we draw on the notation mentioned in eq. 4. The TT-ranks determine the storage consumption of a tensor train and have a strong influence on the possible complexity, i.e. the capability of representing a given tensor as a tensor train. In physics community the above format is known as *Matrix Product State* (MPS) representation introduced as the ground state of the AKLT model in 1987 [35].

Operators $\mathbb{T}$ involved in the presented approach can be seen as multilevel square matrices with pairs of modes, i.e. $\mathbb{T} \in \mathbb{R}^{(n_1 \times n_1) \times \ldots \times (n_d \times n_d)}$. Comparable to eq. 7, $\mathbb{T}$ is in the TT-format if

$$\begin{aligned} \mathbb{T} &= \sum_{i_0=1}^{r_0} \cdots \sum_{i_d=1}^{r_d} (\mathbb{T}^{(1)})_{i_0,:,:,i_1} \otimes \ldots \otimes (\mathbb{T}^{(d)})_{i_{d-1},:,:,i_d} \\ &\Leftrightarrow \\ (\mathbb{T})_{x_1,y_1,\ldots,x_d,y_d} &= (\mathbb{T}^{(1)})_{:,x_1,y_1,:} \cdot \ldots \cdot (\mathbb{T}^{(d))}]_{:,x_d,y_d,:}. \end{aligned} \tag{8}$$

The storage consumption of the TT-format is estimated as $O(r^2 \cdot n \cdot d)$ for eq. 7 and $O(r^2 \cdot n^2 \cdot d)$ for eq. 8, where n is the maximum mode size and r the maximum TT-rank. Another important property is the ensured existence of a best approximation with bounded TT-ranks [36, 37]. Hence, that format is stable in the sense that we can compute quasi-optimal approximations [16, 38].

For a better understanding, it is advantageous to visualize the TT-cores as 2-dimensional arrays containing vectors (or matrices in case of an operator) as

elements:

$$\begin{bmatrix} \left(T^{(k)}\right)_{1,:,1} & \cdots & \left(T^{(k)}\right)_{1,:,r_k} \\ \vdots & & \vdots \\ \left(T^{(k)}\right)_{r_{k-1},:,1} & \cdots & \left(T^{(k)}\right)_{r_{k-1},:,r_k} \end{bmatrix}. \tag{9}$$

We then utilize the following notation for a TT-decomposition.

$$\begin{aligned} T = \begin{bmatrix} \left(T^{(1)}\right)_{1,:,1} \cdots \left(T^{(1)}\right)_{1,:,r_1} \end{bmatrix} \otimes \begin{bmatrix} \left(T^{(2)}\right)_{1,:,1} & \cdots & \left(T^{(2)}\right)_{1,:,r_2} \\ \vdots & & \vdots \\ \left(T^{(2)}\right)_{r_1,:,1} & \cdots & \left(T^{(2)}\right)_{r_1,:,r_2} \end{bmatrix} \otimes \dots \\ \dots \otimes \begin{bmatrix} \left(T^{(d-1)}\right)_{1,:,1} & \cdots & \left(T^{(d-1)}\right)_{1,:,r_{d-1}} \\ \vdots & & \vdots \\ \left(T^{(d-1)}\right)_{r_{d-2},:,1} & \cdots & \left(T^{(d-1)}\right)_{r_{d-2},:,r_{d-1}} \end{bmatrix} \otimes \begin{bmatrix} \left(T^{(d)}\right)_{1,:,1} \\ \vdots \\ \left(T^{(d)}\right)_{r_{d-1},:,1} \end{bmatrix}. \end{aligned} \tag{10}$$

One can see that the above operation is a generalization of the matrix multiplication, where the matrices contain vectors as elements instead of scalar values. Just like multiplying two matrices, we compute the outer products of the according elements and then summarize over the columns and rows, respectively. Later, we will make use of the notation in eq. 10 to give a compact expression of the considered operator for our model.

3.2. TT-Representation of the MME

Let M be the number of all reaction events involving one or two sites. For a state $X = (x_1, \dots, x_d) \in \{1, 2, 3\}^d$ each propensity a_μ, $\mu = 1, \dots, M$, as defined in eq. 3 is decomposable in a product of functions:

$$a_\mu(X) = a_\mu^{(1)}(x_1) \cdot \ldots \cdot a_\mu^{(d)}(x_d). \tag{11}$$

This is the case for classical CMEs, see ref. [9]. But eq. 11 also holds for the considered CO oxidation problem and thus we can write the propensities a_μ as rank-one-tensors $\mathbf{a}_\mu \in \mathbb{R}^{3\times\dots\times 3}$ such that

$$a_\mu(X) = (\mathbf{a}_\mu)_{x_1,\dots,x_d} = (\mathbf{a}_\mu^{(1)})_{x_1} \cdot \ldots \cdot (\mathbf{a}_\mu^{(d)})_{x_d} \tag{12}$$

with $\mathbf{a}_\mu^{(i)} \in \mathbb{R}^3$ for $i = 1, \ldots, d$. Equation 12 implies that $\mathbf{a}_\mu = \mathbf{a}_\mu^{(1)} \otimes \ldots \otimes \mathbf{a}_\mu^{(d)}$. Furthermore, we can express each propensity function according to an reaction event between two neighboring sites ν and $\nu + 1$ as

$$\mathbf{a}_\mu = k_\mu \cdot (e \otimes \ldots \otimes e \otimes v \otimes \tilde{v} \otimes e \otimes \ldots \otimes e)\,, \tag{13}$$

where v corresponds to site ν, $\tilde{v}$ to site $\nu+1$ and $e = (1,1,1)^T$. The constant k_μ is the rate constant of reaction R_μ. Table 2 gives an overview on the components v and $\tilde{v}$.

Adsorption											
O_2	:	v	=	$(1,0,0)^{\mathrm{T}}$,	$\tilde{v}$	=	$(1,0,0)^{\mathrm{T}}$	, CO :	v	=	$(1,0,0)^{\mathrm{T}}$, $\tilde{v}$ = $(1,1,1)^{\mathrm{T}}$
Desorption											
O_2	:	v	=	$(0,1,0)^{\mathrm{T}}$,	$\tilde{v}$	=	$(0,1,0)^{\mathrm{T}}$	, CO :	v	=	$(0,0,1)^{\mathrm{T}}$, $\tilde{v}$ = $(1,1,1)^{\mathrm{T}}$
CO_2	:	v	=	$(0,0,1)^{\mathrm{T}}$,	$\tilde{v}$	=	$(0,1,0)^{\mathrm{T}}$	or	v	=	$(0,1,0)^{\mathrm{T}}$, $\tilde{v}$ = $(0,0,1)^{\mathrm{T}}$
Diffusion											
O	:	v	=	$(0,1,0)^{\mathrm{T}}$,	$\tilde{v}$	=	$(1,0,0)^{\mathrm{T}}$	or	v	=	$(1,0,0)^{\mathrm{T}}$, $\tilde{v}$ = $(0,1,0)^{\mathrm{T}}$
CO	:	v	=	$(0,0,1)^{\mathrm{T}}$,	$\tilde{v}$	=	$(1,0,0)^{\mathrm{T}}$	or	v	=	$(1,0,0)^{\mathrm{T}}$, $\tilde{v}$ = $(0,0,1)^{\mathrm{T}}$

Table 2: Components of propensity tensors $\mathbf{a}_\mu$, $\mu = 1, \ldots, M$. Each of the associated reactions R_μ corresponds to one of the reactions given in Table 1.

We also associate the probabilities $P(X,t)$, $X = (x_1, \ldots, x_d)$, with a tensor, i.e. we utilize a tensor $\mathbf{P}(t) \in \mathbb{R}^{3\times\ldots\times 3}$ that is defined by

$$(\mathbf{P}(t))_{x_1,\ldots,x_d} = P(X,t). \tag{14}$$

Assuming $\mathbf{P}(t)$ to be a rank-one-tensor, see eq. 5, leads to the well-known rate equation expressions of chemical kinetics. For the CO oxidation at $RuO_2(110)$ these equations have been described in ref. [7].

For $\mu = 1, \ldots, M$ and $\nu = 1, \ldots, d$, $S(k)$ denotes a shift matrix in $\mathbb{R}^{3\times 3}$, which is given by $(S(k))_{i,j} := \delta_{j-i,k}$, where $\delta_{j-i,k}$ represents the Kronecker delta. $S(0)$ denotes the identity matrix.

Definition 3. *The* multidimensional shift operators $\mathbb{S}_\mu$ *and* $\mathbb{S}_0$ *are defined as*

$$\mathbb{S}_\mu = S(-\xi_\mu(1)) \otimes \ldots \otimes S(-\xi_\mu(d)) \tag{15}$$

and

$$\mathbb{I} = \mathbb{S}_0 = S(0) \otimes \ldots \otimes S(0). \tag{16}$$

With help of the definitions above, we can give a more convenient counterpart of eq. 2:

$$\frac{\partial}{\partial t}\mathbf{P}(t) = \left(\sum_{\mu=1}^{M} (\mathbb{S}_\mu - \mathbb{I}) \cdot \mathrm{diag}(\mathbf{a}_\mu) \right) \cdot \mathbf{P}(t), \tag{17}$$

where $\mathrm{diag}(\mathbf{a}_\mu)$ denotes the outer product of matrices containing the entries of $\mathbf{a}_\mu^{(1)}, \ldots, \mathbf{a}_\mu^{(d)}$ as diagonals, i.e.

$$\mathrm{diag}(\mathbf{a}_\mu) = \mathrm{diag}(\mathbf{a}_\mu^{(1)}) \otimes \ldots \otimes \mathrm{diag}(\mathbf{a}_\mu^{(d)}), \tag{18}$$

for $\mu = 1, \ldots, M$. A proof that this notation leads back to the master equation given in eq. 2 can be found in the appendix (A).

Henceforth, we will write the right hand side of eq. 17 as $\mathbb{L}\mathbf{P}(t)$, where $\mathbb{L}$ denotes the linear operator defined as

$$\mathbb{L} = \sum_{\mu=1}^{M} (\mathbb{S}_\mu - \mathbb{I}) \cdot \mathrm{diag}(\mathbf{a}_\mu). \tag{19}$$

Using formula 19, we could directly give an exact canonical decomposition of the operator of the master equation 2, but we need the operator to be in the TT-format. Instead of converting the operator directly into the TT-format, we look for a systematic TT-decomposition of $\mathbb{L}$ in order to keep the ranks small. Using the notation introduced in eq. 9 and eq. 10, and exploiting the translational symmetry, we come to the following representation for the operators of the 1D model:

$$\mathbb{L} = \begin{bmatrix} A & B & I & B \end{bmatrix} \otimes \begin{bmatrix} I & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ A & B & I & 0 \\ 0 & 0 & 0 & J \end{bmatrix} \otimes \ldots \otimes \begin{bmatrix} I & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ A & B & I & 0 \\ 0 & 0 & 0 & J \end{bmatrix} \otimes \begin{bmatrix} I \\ C \\ A \\ C \end{bmatrix}, \tag{20}$$

with core elements $A \in \mathbb{R}^{1\times3\times3\times1}, B \in \mathbb{R}^{1\times3\times3\times7}, C \in \mathbb{R}^{7\times3\times3\times1}, I \in \mathbb{R}^{1\times3\times3\times1}$, and $J \in \mathbb{R}^{7\times3\times3\times7}$. An exact description for the core elements in eq. 20 is given in the appendix (B).

3.3. Computing Stationary Distributions

After we have transformed the master equation into the TT-format, we now want to present ways how to compute stationary and time dependent distributions. Finding the stationary distribution corresponds to setting the time derivative in the master equation to zero and solving the resulting linear problem. However, this is degenerate as the trivial solution always solves it. But we can reformulate the problem as an eigenvalue problem

$$(\mathbb{I} + \mathbb{L})\,\mathbf{P} = \mathbf{P}, \tag{21}$$

assuming that there exists a unique tensor $\mathbf{P} \in \mathbb{R}^{3\times\cdots\times 3}$ which satisfies eq. 21 and $\|P\|_1 = 1$. If we target on time dependent problems, we have to introduce a suitable discretization. As te MME is stiff, implicit discretizations are best choice. Simpliest is the implicit Euler method, which, for a linear ordinary differential equation as the MME, requires to solve linear systems in the form of

$$(\mathbb{I} - \tau\mathbb{L})\,\mathbf{P}_{k+1} = \mathbf{P}_k, \tag{22}$$

where $\tau \in \mathbb{R}^+$ denotes the step size. It is $\mathbf{P}_k \in \mathbb{R}^{3\times\cdots\times 3}$ for $k = 0, \ldots, s-1$ with $s \in \mathbb{N}$ the number of steps. We will use this method also to determine stationary distributions. As we are then not interested in a high temporal accuracy, the step size τ will be used to control the condition of the linear systems 22. As a truncation criterion in the iterative solution, we will employ the residual error

$$Err_1(k) = \frac{\|(\mathbb{I} - \tau\mathbb{L})\,\mathbf{P}_k - \mathbf{P}_{k-1}\|_2}{\|\mathbf{P}_{k-1}\|_2} \tag{23}$$

If we need a higher order in accuracy, we will also apply the implicit trapezoidal rule for transient processes. The linear systems in each iteration step then become

$$\left(\mathbb{I} - \frac{\tau}{2}\mathbb{L}\right)\mathbf{P}_{k+1} = \left(\mathbb{I} + \frac{\tau}{2}\mathbb{L}\right)\mathbf{P}_k, \tag{24}$$

and we use the truncation criterion

$$Err_2(k) = \frac{\left\|\left(\mathbb{I} - \frac{\tau}{2}\mathbb{L}\right)\mathbf{P}_k - \left(\mathbb{I} + \frac{\tau}{2}\mathbb{L}\right)\mathbf{P}_{k-1}\right\|_2}{\left\|\left(\mathbb{I} + \frac{\tau}{2}\mathbb{L}\right)\mathbf{P}_{k-1}\right\|_2} \tag{25}$$

In order to obtain the solutions of eqs. 21, 22 and 24, we apply the *Alternating Linear Scheme* (ALS) [27]. The ALS as well as the modified version MALS are closely related to the *Density Matrix Renormalization Group* (DMRG), which is a numerical variational technique used in quantum physics [39]. The main idea of the ALS is to start with a tensor in TT-format as initial guess and fix all TT-cores except the first. We then solve a lower dimensional linear system or a lower dimensional eigenvalue problem, respectively, to optimize that core. Subsequently, we fix all cores from the previous iterations and proceed to the next core. After an optimized core is computed, a QR decomposition is applied to a certain unfolding of the core. The folding of the orthonormal part builds then the new core and the non-orthonormal part is shifted to the next core which is optimized in the next step of the ALS. This procedure is consecutively performed in both directions of the TT-cores and is repeated several times until desired accuracy is reached.

Even though only locally convergence has been proven [40] and the ALS was particularly developed for linear systems with symmetric positive definite operators on the left hand side, we observed highly accurate approximations of the solutions for the non-symmetric problems 21, 22 and 24. However, the above mentioned properties for the ALS might explain why we could not employ one large time step in the Euler discretization for obtaining the stationary distribution, when the problem becomes dominated by the non-symmetric operator $\mathbb{L}$. For other studies dealing with non-symmetric systems we refer to refs. [41], [42] and [43].

4. Results

As in the preceding work [8], we focus on a set of gas-phase conditions, which are representative for so-called *in situ* experiments, i. e. close to the operation conditions for a catalyst in a realistic scenario. These conditions are defined by a fixed O_2 pressure $p_{O_2} = 1$ atm, fixed temperature $T = 600$K, and varying CO pressure, 10^{-4} atm $\leq p_{CO} \leq 10^2$ atm.

The central objective of heterogeneous catalysis, is the stationary behavior of the catalyst. Therefore, our experiments in Sections 4.1 and 4.2 focus on computing the stationary distribution. In Section 4.3 we consider the transient behavior and demonstrate the advantage of the TT-approach for stiff systems.

For computing stationary distributions, as one option we use formulation 21 and apply the ALS as an eigensolver. If we employ the time propagation schemes for this purpose and no approximation of the stationary distribution is obtained before, we chose a homogeneous surface as initial state, i.e. $\mathbf{P}_0 = e_i \otimes \ldots \otimes e_i$, where e_i is the ith vector of the standard basis of $\mathbb{R}^3$. We set the initial guess of the ALS for the first occurring linear system in TT-format to an uniformly distributed tensor. Subsequently, the computed state after time step τ is used as initial guess for determining the next probability distribution.

For benchmarking, we compare the results from the TT approximation with kinetic Monte Carlo simulations, when applicable. For these, we employ the KMOS package [11], which implements the *Variable Step Size Method* [10] and an $O(1)$ update rule per time steps. In other words, the implemented methodology is linear in system size (i.e. the dimension d) for estimating expected values, as the number of time steps needed to simulate a certain time interval increases linearly with the system size.

All experiments with the TT-format, were performed on a Windows 7 64-bit machine with 16 GB RAM and an Intel Core i5-4200U processor with a clock speed of 1.6 GHz and a cache size of 3 MB. The algorithms were implemented in MATLAB R2013b using a compound of cell arrays and multidimensional matrices for tensors in the TT-format.

4.1. Scaling with system size

When dealing with high dimensional problems, a very important aspect is how does the computational costs behave for increasing dimensionality. We examine this dependence by increasing the number of sites and measuring the CPU time needed to approximate the stationary distribution for $p_{\mathrm{CO}} = 1$ atm ($k_{\mathrm{CO}}^{\mathrm{Ad}} = 10^8 s^{-1}$), by applying the implicit Euler method. We start with a fully

O-covered surface and a uniform tensor train with rank 10 as initial guess, which turned out to be sufficient for calculating the stationary distributions for orders d up to 100. That is, the largest linear system we consider has $3^{100} \approx 10^{48}$ unknowns, which is far beyond the capabilities of current and future supercomputers using classical methods. We compute 20 steps, starting with a step size of 10^{-10} and doubling the step size after each step. The residual errors given in eq. 23 of all linear systems is smaller than 0.1, which implies a sufficient accuracy of every solution computed by the ALS. The derivatives of the last distributions for each value of d are bounded by 1 indicating that the obtained distribution is close to the stationary state since $\sum_Y W(X|Y)P(Y,t) \approx \sum_Y W(Y|X)P(X,t)$ for all $X \in \{1,2,3\}^d$, see eq. 1. Figure 2 shows the CPU times for the entire computations and the average time needed for solving one linear equation system given in the TT-format.

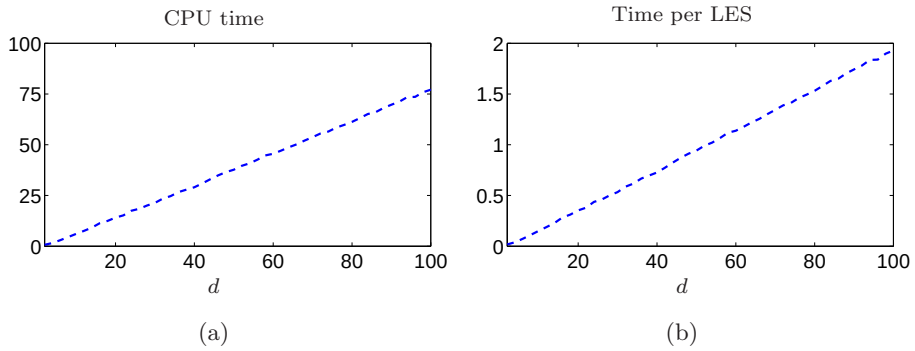


Figure 2: (a) CPU time in seconds for computing the stationary distribution over order d. (b) Average time in seconds for solving one linear equation system over d. Note that the CPU time also includes the computation of the errors 23.

The computational time for one linear equation system increases linearly with the number of sites, see Figure 2b. This reflects in the CPU time for the whole problem 2a. So we achieve desirable linear scaling with dimensionality, i.e. probably the best one can achieve without imposing symmetry, even with Monte Carlo methods. With the computational effort of the ALS only depending linearly on the the number of cores [27] and the fact that the same low rank suits

for all dimensions, this leads to the linear dependence. So the TT-format is able to effectively encode the spatial correlation for the considered one-dimensional problem. Or, in other words, it encodes the physical intuition that, from a certain size on, a twice as large system is simply the "sum" of two smaller systems.

4.2. Varying the CO Pressure

The central quantities describing the efficiency of the catalyst are the so-called turn-over frequencies (TOF), which measure how often a reaction is executed per unit time and unit surface (usually per site). For the model at hand, the TOF for the CO + O reaction is given by

$$\mathrm{TOF} = \frac{k^{\mathrm{De}}_{\mathrm{CO_2}}}{d} \sum_{\nu\in\{1,\ldots,d\}} \sum_{\tilde{\nu}\in\{\nu-1,\nu+1\}} P(\mathrm{CO\ on\ }\nu\ \&\ \mathrm{O\ on\ }\tilde{\nu}), \tag{26}$$

where $P(\mathrm{CO\ on\ }\nu\ \&\ \mathrm{O\ on\ }\tilde{\nu})$ denotes the probability to find a CO molecule on site ν and an oxygen atom on site $\tilde{\nu}$ Of course, the TOF is sensitively dependent on how many CO molecules and oxygen atoms are adsorbed on the surface. Therefore the coverages, i.e. the average number of adsorbates of a certain kind divided by the total number of sites, are of high interest, too. For the considered translational invariant problem, the coverages are simply given by

$$\{\ P(\varnothing\ \mathrm{on}\ \nu),\ P(\mathrm{O\ on}\ \nu),\ P(\mathrm{CO\ on}\ \nu)\ \} \tag{27}$$

Both, the TOF and the coverages, depend on the reaction conditions. We therefore repeat the study from ref. [8], i.e. we vary the CO partial pressure from 10^{-4} atm to 10^{2} atm, and keep temperature and oxygen pressure fixed to 600 K and 1 atm, respectively. We compare the results with kMC simulations time averaged over 10^9 steps after the simulations have been relaxed to steady state. For both methods, we choose a chain with 20 sites.

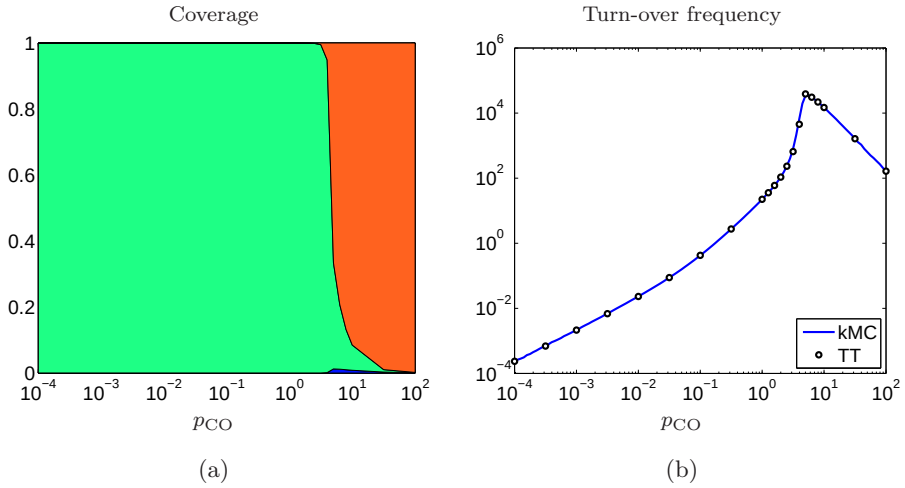
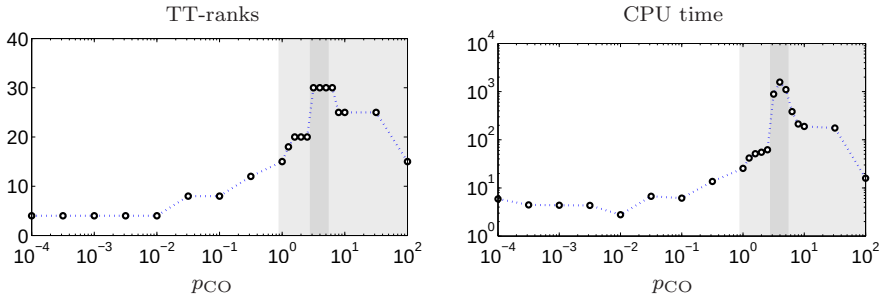


Figure 3: (a) Mean coverage of the cus sites over CO pressure p_{CO}. Blue: empty, green: O-covered, orange: CO-covered. (b) Turn-over frequency of the catalyst over p_{CO} in . Blue line: TOF obtained with kMC, black circles: values computed with tensor train approach

Figures 3a and 3b show an area plot of the coverage and the TOF over the CO pressure interval, respectively. As in the literature, we find three characteristic regimes: An almost fully O-covered surface at low CO pressures, an almost fully CO-covered surface at high CO pressure and an intermediate pressure regime where both species exist with non-vanishing probabilities. The fraction of empty sites is always small with at most 1.2% between the O-poisoned and CO-poisoned regions. The TOF increases for growing p_{CO} until it reaches its maximum at around 5 atm. As to be expected from the literature [8], the peak of the TOF is located in the intermediate regime, were both reactants for this reaction are available on the surface in appreciable amount. Comparing the TOF obtained by kMC (blue line) and the tensor train approach (black circles), we see an almost perfect agreement. To our knowledge, there exist no other probability based method proposed so far, which is able to produce such an accuracy while keeping the computational effort tractable (compare e.g. refs [13, 24]).



(a) (b)

Figure 4: (a) TT-ranks used for computing the stationary distribution over p_{CO}. (b) CPU time in seconds needed for computation. Used methods are depicted by the different background shadings. White: treatment as eigenvalue problem, dark gray: application of implicit Euler method to the master equation, light gray: combination of both methods.

For several of the chosen parameter values it is possible to obtain approximations of the steady state with satisfying accuracy by applying the ALS to the eigenvalue problem and then normalizing the result, so that all entries sum up to 1. In case that method does not lead to a tensor sufficiently close to the stationary distribution, we simulate the system by employing the implicit Euler method to the differential equation 17 using the tensor train computed by the ALS eigensolver as initial state and doubling the step size after each time step. If even that combination fails, we fall back to applying the implicit Euler method with an empty surface as initial state and a experimentally determined set of step sizes. Figure 4a display the TT-ranks used for computation of the stationary distributions and the methods which were applied. Figure 4b shows the corresponding CPU times needed for the computation. For CO pressures below 1 atm, where we treat the task of finding the stationary distribution as an eigenvalue problem, the TT-ranks as well as the CPU times are comparably small. Also for the remaining CO pressure values outside the intermediate region we can compute the approximations of the steady states in short time using the eigenvalue problem approximation according to eq. 21 as initial condition for the implicit Euler method as given in eq. 22. On the one hand, combining an eigenvalue problem and the implicit Euler method leads to more accurate

results, on the other hand, it enables us to keep the computational times small in comparison to the calculations in the intermediate regime where the prefixing of an eigenvalue problem yields to no advantage. Overall, Figure 4 shows that the TT-ranks and CPU times increase as p_{CO} gets closer to the intermediate regime. Viewed algebraically, the full tensor of the stationary distribution is dense since the probabilities of a large amount of surface arrangements do not vanish. Thus, the tensor ranks of a TT-approximation have to be increased in order to catch the complexity of the probability distribution and to achieve accurate approximations. In contrast to that, tensor trains of rank 4 are sufficient to reproduce the stationary distributions for the smallest values of p_{CO} since the probability at these points is concentrated in a rather small number of possible surface configurations.

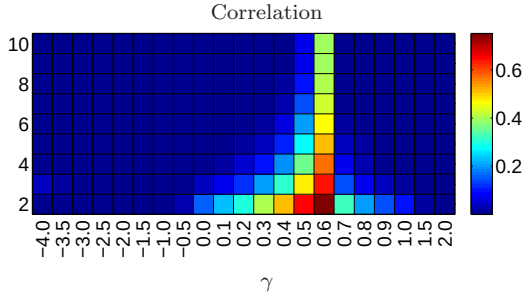


Figure 5: Values of the correlation function $\tilde{C}$ for distances $l = 2, \ldots, 10$ over exponents γ with $p_{\mathrm{CO}} = 10^{\gamma}$ atm.

To rationalize these findings, we investigate the correlation lengths of the cus sites for the different values of the CO pressure. We quantify these by calculating the measure of total spatial correlation

$$C(l) = \frac{1}{9} \sum_{i=1}^{3} \sum_{j=1}^{3} \left| P(x_1 = i \;\&\; x_{1+l} = j) - P(x_1 = i) P(x_{1+l} = j) \right|, \qquad (28)$$

for distances $l \in \{1, \ldots, 10\}$ and then normalizing it by defining $\tilde{C}(l) = C(l)/C(1)$. Comparing Figure 5 with Figure 4, we see that there is a close connection between the computational effort for calculating the stationary distribution and the correlation length in the considered pressure interval. If the correlation

length is negligibly small, we are able to keep the TT-ranks and computational time at a small level. For values of the CO pressure where the calculation of the steady states is more expensive, also the correlation length increases. This fact leads us to the conclusion that, if the dynamics of sites which are further away correlate, we need higher TT-ranks for the approximations such that the tensor product approach provides accurate results with help of the methods 21, 22 and the implementation of the ALS.

4.3. Increasing the Oxygen Desorption Rate

So far, we have demonstrated that for the one-dimensional problem we are considering in this study the tensor train approach shows a comparable performance to kMC. We now want to demonstrate its advantage over kMC, when dealing with stiff problems. In a kMC simulation, stiffness has a negative effect, if there is a number of processes which are very fast and get executed again and again. Then the system will get stuck fluctuating between only small number of states and sufficient sampling of the state space will only be possible with a very high number of Monte Carlo steps. Or, equivalently, these states are very short lived and reaching a certain final time will require also an increasing number of step if stiffness is increased.

For this we consider the case with p_{CO} set to 10^{-4} atm. Under these conditions, the surface is almost fully O-covered, and the dominant processes are oxygen ad- and desorption. As the desorption has a very low rate constant, stiffness in the above sense is increased for larger $k_{\mathrm{O_2}}^{\mathrm{De}}$. We therefore multiply the original parameter $k_{\mathrm{O_2}}^{\mathrm{De}}$ by constants λ between 1 and 10^6, compute the solution up to 1 second and compare the CPU time for this time-dependent problem between kMC and the TT-approach. We consider 10 sites and start with an empty surface and step size $\tau = 10^e$, $e = -11$. After each 10 steps we increase the exponent e by 1 such that we compute 110 steps with step sizes from 10^-11 to 10^{-1}. This time we repeat the ALS three times for each linear system. On the first time intervals the trapezoidal rule is applied, after 40 steps we switch to the implicit Euler method in order to approach the stationary distribution since

it leads to more adequate results for larger step sizes. We choose the TT-ranks such that we can keep the relative errors of both methods below 10%, i.e. the error of the trapezoidal rule given in eq. 25 and the error of the implicit Euler method given in eq. 23 should stay below 0.1. When the implicit Euler reaches the best approximation $\mathbf{P}_{\mathrm{STAT}}$ of the stationary state before the algorithm is finished, the relative error increases naturally since the iterates almost do not alter anymore. Then we require that the derivative of the current distribution is small enough, that is, after at most 110 steps all derivatives $\mathbb{L}\mathbf{P}_{\mathrm{STAT}}$ satisfy $\|\mathbb{L}\mathbf{P}_{\mathrm{STAT}}\|_2 \leq 1$.

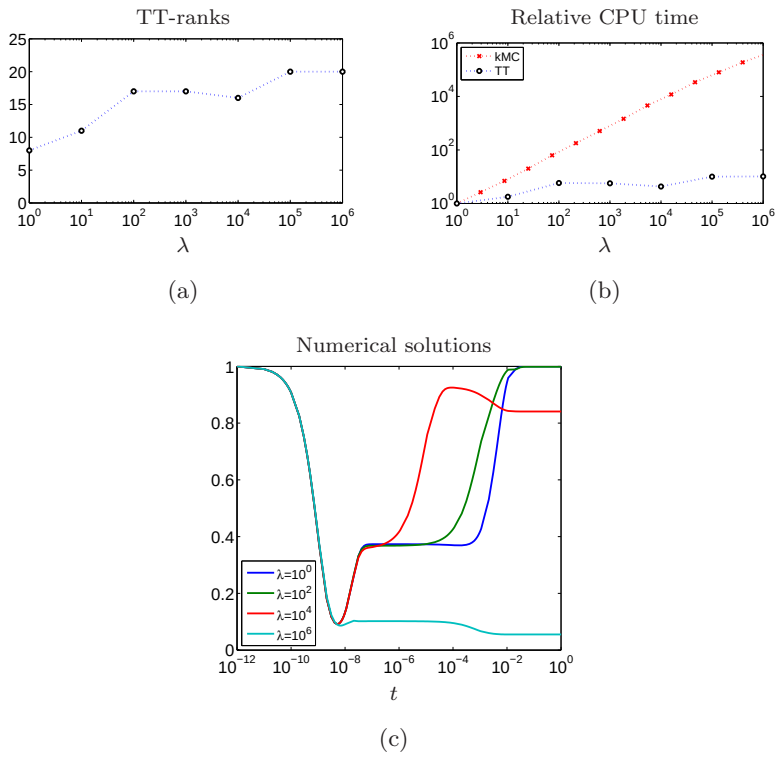


Figure 6: (a) TT-ranks used for computing the stationary distribution. (b) Relative CPU time needed for computation using kMC and TT-aproach. (c) Euclidian norm of the numerical solution.

Figures 6a and 6b show the TT-ranks and the CPU times over λ. For better clarity, we show relative CPU times, i.e. we normalize with the CPU times

for the case $\lambda = 1$. We see that the relative computational time of the kMC approach increases linearly with lambda. In contrast to that, the TT approach shows only a very small dependence on the stiffness parameter λ in terms of performance. This is mostly because we have to increase the TT-ranks for larger λ. The rise of the TT-ranks originates in the system's dynamics altering with growing λ which is recognizable in Figure 6c where the euclidian norms of the numerical solutions over time $t \leq 1\,\mathrm{s}$ are shown. As one can see by the norm of the final distribution for $\lambda = 1$, the solution converges to an almost single state, i.e. the probability distribution in concentrated in the state of an fully O-covered surface. Due to the higher desorption rates, oxygen atoms dwell for shorter times letting the probability distribution spread over various surface configurations. For the cases where the fast processes do not alter the probability distribution, we expect no impact on the CPU time, for instance equilibrated pairs of forward and backward processes multiplied by the same factor, where the terms containing λ would cancel in the master equation.

5. Conclusion and Outlook

We have investigated the benefit of tensor trains to solve the master equations appearing in the correlated dynamics at catalytic surfaces on the example of a reduced microkinetic model for the CO oxidation at RuO_2. For this model, we derived an exact tensor decomposition of the operator and described how to compute stationary and time dependent distributions using a formulation as an eigenvalue problem or applying implicit time propagation schemes and the Alternating Linear Scheme.

We demonstrated that linear complexity in the system's dimensionality can be achieved. Benchmarking with highly accurate kinetic Monte Carlo simulations, we showed that the approach provides numerical accuracy over a large range of input parameters for the model. The approach showed superior behavior above kinetic Monte Carlo for a sequence problems with increasing stiffness, where its computational complexity grew at much lower rate.

However, at the current state of research, it is non-trivial to derive a TT-decomposition for a given reaction network. Experimenting with various step sizes for the implicit methods, it became apparent that the convergence of the ALS highly depends on the time step. Reasons for the effect of ALS failing to solve a linear system at first attempt may be the stiffness of the surface model, and thereby bad conditioning of the linear system for large time steps, as well as the fact that the requirements of the ALS are not complied, i.e. the considered operators are non-symmetric.

The tensor trains based approaches to Markov processes are a promising route for the treatment of catalytic surface kinetics. The here presented one-dimensional example was a first step into this direction. Future research will include the implementation of refined step size adaption schemes and the extension of the approach for two-dimensional problems. Other promising directions are improvement/development of tensor structured solvers for non-symmetric operators and corresponding preconditioners.

6. Acknowledgments

The Berlin Mathematical School, the Einstein Center for Mathematics and the Matheon Research Center are acknowledged for financial support.

References

[1] Y. Qu, X. Duan, Progress, challenge and perspective of heterogeneous photocatalysts, Chemical Society Reviews 42 (7) (2013) 2568–2580. doi:10.1039/C2CS35355E.

[2] H. Gandhi, G. Graham, R. McCabe, Automotive exhaust catalysis, Journal of Catalysis 216 (1-2) (2003) 433–442. doi:10.1016/S0021-9517(02)00067-2.

[3] J. K. Norskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, Towards the computational design of solid catalysts, Nature Chemistry 1 (2009) 37–46. doi:10.1038/nchem.121.

[4] O. Deutschmann, S. Tischer, Numerical simulation of catalytic reactors by molecular-based models, Model Based Parameter Estimation, Contributions in Mathematical and Computational Sciences 4 (2013) 227–250. doi:10.1007/978-3-642-30367-8_11.

[5] K. Reuter, M. Scheffler, First-principles kinetic Monte Carlo simulations for heterogeneous catalysis: Application to the CO oxidation at RuO_2(110), Physical Review B 73 (4) (2006) 045433. doi:10.1103/PhysRevB.73.045433.

[6] M. Stamatakis, D. G. Vlachos, Unraveling the complexity of catalytic reactions via kinetic monte carlo simulation: Current status and frontiers, ACS Catalysis 2 (12) (2012) 2648–2663. doi:10.1021/cs3005709.

[7] B. Temel, H. Meskine, K. Reuter, M. Scheffler, H. Metiu, Does phenomenological kinetics provide an adequate description of heterogeneous catalytic reactions?, Journal of Chemical Physics 126 (20) (2007) 204711. doi:10.1063/1.2741556.

[8] S. Matera, H. Meskine, K. Reuter, Adlayer inhomogeneity without lateral interactions: Rationalizing correlation effects in CO oxidation at

RuO_2(110) with first-principles kinetic Monte Carlo, Journal of Chemical Physics 134 (064713). doi:10.1063/1.3553258.

[9] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, Journal of Computational Physics 22 (4) (1976) 403–434. doi:10.1016/0021-9991(76)90041-3.

[10] A. P. J. Jansen, An introduction to Monte Carlo simulations of surface reactions (2003). arXiv:cond-mat/0303028v1.

[11] M. J. Hoffmann, S. Matera, K. Reuter, kmos: A lattice kinetic monte carlo framework, Computer Physics Communications 185 (7) (2014) 2138–2150. doi:10.1016/j.cpc.2014.04.003.

[12] J. Mai, V. N. Kuzovkov, W. von Niessen, A general stochastic model for the description of surface reaction systems, Physica A: Statistical Mechanics and its Applications 203 (2) (1994) 298–315. doi:10.1016/0378-4371(94)90158-9.

[13] G. J. Herschlag, S. Mitran, G. Lin, A consistent hierarchy of generalized kinetic equation approximations to the master equation applied to surface catalysis, The Journal of Chemical Physics 142 (23) (2015) 234703. doi:10.1063/1.4922515.

[14] L. Grasedyck, D. Kressner, C. Tobler, A literature survey of low-rank tensor approximation techniques, GAMM-Mitteilungen 36 (1) (2013) 53–78. doi:10.1002/gamm.201310004.

[15] I. V. Oseledets, A new tensor decomposition, Doklady Mathematics 80 (1) (2009) 495–496. doi:10.1134/S1064562409040115.

[16] I. V. Oseledets, E. E. Tyrtyshnikov, Breaking the curse of dimensionality, or how to use svd in many dimensions, SIAM Journal on Scientific Computing 31 (5) (2009) 3744–3759. doi:10.1137/090748330.

[17] I. V. Oseledets, Tensor-train decomposition, SIAM Journal on Scientific Computing 33 (5) (2011) 2295–2317. doi:10.1137/090752286.

[18] T. Jahnke, W. Huisinga, A dynamical low-rank approach to the chemical master equation, Bulletin of Mathematical Biology 70 (8) (2008) 2283–2302. doi:10.1007/s11538-008-9346-x.

[19] V. Kazeev, M. Khammash, M. Nip, C. Schwab, Direct solution of the chemical master equation using quantized tensor trains, PLoS Comput Biol 10 (3) (2014) e1003359. doi:10.1371/journal.pcbi.1003359.

[20] S. Dolgov, B. Khoromskij, Simultaneous state-time approximation of the chemical master equation using tensor product formats, Numerical Linear Algebra with Applications 22 (2) (2015) 197–219. doi:10.1002/nla.1942.

[21] P. Buchholz, Product form approximations for communicating markov processes, Performance Evaluation 67 (9) (2010) 797–815. doi:10.1016/j.peva.2009.12.005.

[22] D. Kressner, F. Macedo, Low-rank tensor methods for communicating markov processes, Quantitative Evaluation of Systems, Lecture Notes in Computer Science 8657 (2014) 25–40. doi:10.1007/978-3-319-10696-0_4.

[23] S. Matera, K. Reuter, When atomic-scale resolution is not enough: Spatial effects on in situ model catalyst studies, Journal of Catalysis 295 (2012) 261–268. doi:doi:10.1016/j.jcat.2012.08.020.

[24] D. J. Liu, J. W. Evans, Transitions between strongly correlated and random steady-states for catalytic CO-oxidation on surfaces at high-pressure, The Journal of Chemical Physics 142 (13) (2015) 134703. doi:10.1063/1.4916380.

[25] K. Reuter, D. Frenkel, M. Scheffler, The steady state of heterogeneous catalysis, studied by first-principles statistical mechanics, Physical Review Letters 93 (11) (2004) 116105. doi:10.1103/PhysRevLett.93.116105.

[26] M. Rieger, J. Rogal, K. Reuter, Effect of surface nanostructure on temperature programmed reaction spectroscopy: First-principles kinetic monte carlo simulations of CO oxidation at RuO_2(110), Physical Review Letters 100 (1) (2008) 016105. doi:10.1103/PhysRevLett.100.016105.

[27] S. Holtz, T. Rohwedder, R. Schneider, The alternating linear scheme for tensor optimization in the tensor train format, SIAM Journal on Scientific Computing 34 (2) (2012) A683–A713. doi:10.1137/100818893.

[28] N. G. van Kampen, Stochastic processes in physics and chemistry (third edition), North-Holland Personal Library, Elsevier B.V., 2007. doi:10.1016/B978-044452965-7/50008-8.

[29] D. T. Gillespie, A rigorous derivation of the chemical master equation, Physica A 188 (1-3) (1992) 404–425. doi:10.1016/0378-4371(92)90283-V.

[30] H. Meskine, S. Matera, M. Scheffler, K. Reuter, H. Metiu, Examination of the concept of degree of rate control by first-principles kinetic monte carlo simulations, Surface Science 603 (10) (2009) 1724–1730. doi:10.1016/j.susc.2008.08.036.

[31] W. Hackbusch, Tensor Spaces and Numerical Tensor Calculus, Vol. 42 of Springer Series in Computational Mathematics, Springer, 2012. doi:10.1007/978-3-642-28027-6.

[32] F. L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, Journal of Mathematics and Physics 6 (1927) 164–189.

[33] J. D. Carroll, J. J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition, Psychometrika 35 (3) (1970) 283–319. doi:10.1007/BF02310791.

[34] V. de Silva, L.-H. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM Journal on Matrix Analysis and Applications 30 (3) (2008) 1084–1127. doi:10.1137/06066518X.

[35] I. Affleck, T. Kennedy, E. H. Lieb, H. Tasaki, Rigorous results on valence-bond ground states in antiferromagnets, Physical Review Letters 59 (7) (1987) 799–802. doi:10.1103/PhysRevLett.59.799.

[36] A. Falcó, W. Hackbusch, On minimal subspaces in tensor representations, Foundations of Computational Mathematics 12 (6) (2012) 765–803. doi:10.1007/s10208-012-9136-6.

[37] S. Holtz, T. Rohwedder, R. Schneider, On manifolds of tensors of fixed TT-rank, Numerische Mathematik 120 (4) (2012) 701–731. doi:10.1007/s00211-011-0419-7.

[38] L. Grasedyck, Hierarchical singular value decomposition of tensors, SIAM Journal on Matrix Analysis and Applications 31 (4) (2010) 2029–2054. doi:10.1137/090764189.

[39] S. R. White, Density matrix formulation for quantum renormalization groups, Physical Review Letters 69 (19) (1992) 2863–2866. doi:10.1103/PhysRevLett.69.2863.

[40] T. Rohwedder, A. Uschmajew, On local convergence of alternating schemes for optimization of convex problems in the tensor train format, SIAM Journal of Numerical Analysis 51 (2) (2013) 1134–1162. doi:10.1137/110857520.

[41] S. V. Dolgov, B. N. Khoromskij, Tensor-product approach to global time-space-parametric discretization of chemical master equation, Preprint No.68, Max-Planck-Institut für Mathematik in den NaturwissenschaftenAvailable at http://www.mis.mpg.de/preprints/2012/preprint2012{_}68.pdf [Accessed 7 July 2015].

[42] S. V. Dolgov, D. V. Savostyanov, Alternating minimal energy methods for linear systems in higher dimensions, SIAM Journal on Scientific Computing 36 (5) (2014) A2248–A2271. doi:10.1137/140953289.

[43] S. V. Dolgov, B. N. Khoromskij, I. V. Oseledets, D. V. Savostyanov, Computation of extreme eigenvalues in higher dimensions using block tensor train format, Computer Physics Communications 185 (4) (2014) 1207–1216. doi:10.1016/j.cpc.2013.12.017.

Appendix A

Theorem. *For any* $X = (x_1, \ldots, x_d)^{\mathrm{T}} \in \{1,2,3\}^d$*, we have*

$$\left(\frac{\partial}{\partial t}\mathbf{P}(t)\right)_{x_1,\ldots,x_d} = \frac{\partial}{\partial t}P(X,t).$$

Proof. *Following the definition of the multiplication, we can write*

$$\left(\frac{\partial}{\partial t}\mathbf{P}(t)\right)_{x_1,\ldots,x_d} = \left(\left(\sum_{\mu=1}^{M}(\mathbb{S}_\mu - \mathbb{S}_0)\cdot \mathrm{diag}(\mathbf{a}_\mu)\right)\cdot \mathbf{P}(t)\right)_{x_1,\ldots,x_d}$$

as

$$\sum_{\mu=1}^{M}\sum_{i_1=1}^{n_1}\cdots\sum_{i_d=1}^{n_d}\left((\mathbb{S}_\mu - \mathbb{S}_0)\cdot \mathrm{diag}(\mathbf{a}_\mu)\right)_{x_1,i_1,\ldots,x_d,i_d}\cdot(\mathbf{P}(t))_{i_1,\ldots,i_d}.$$

Furthermore, it is

$$\begin{aligned}&\left((\mathbb{S}_\mu - \mathbb{S}_0)\cdot \mathrm{diag}(\mathbf{a}_\mu)\right)_{x_1,i_1,\ldots,x_d,i_d}\\ &\quad= \sum_{j_1=1}^{n_1}\cdots\sum_{j_d=1}^{n_d}(\mathbb{S}_\mu)_{x_1,j_1,\ldots,x_d,j_d}\cdot(\mathrm{diag}(\mathbf{a}_\mu))_{j_1,i_1,\ldots,j_d,i_d}\\ &\qquad-\sum_{j_1=1}^{n_1}\cdots\sum_{j_d=1}^{n_d}(\mathbb{S}_0)_{x_1,j_1,\ldots,x_d,j_d}\cdot(\mathrm{diag}(\mathbf{a}_\mu))_{j_1,i_1,\ldots,j_d,i_d}.\end{aligned}$$

Considering the definition of the shift operators, this becomes

$$(\mathrm{diag}(\mathbf{a}_\mu))_{x_1-\xi_\mu(1),i_1,\ldots,x_d-\xi_\mu(d),i_d} - (\mathrm{diag}(\mathbf{a}_\mu))_{x_1,i_1,\ldots,x_d,i_d}.$$

Just as $a_\mu(X)$ *and* $P(X,t)$ *are set to zero if* $X \notin \{1,2,3\}^d$*, see section 2, we set* $(\mathrm{diag}(\mathbf{a}_\mu))_{x_1-\xi_\mu(1),i_1,\ldots,x_d-\xi_\mu(d),i_d} = 0$ *if* $x_k - \xi_\mu(k) \notin \{1,2,3\}$ *for a* $k \in \{1,\ldots,d\}$*. The same we do for* $(\mathbf{P}(t))_{x_1-\xi_\mu(1),\ldots,x_d-\xi_\mu(d)}$*. Due to the construction of* $\mathrm{diag}(\mathbf{a}_\mu)$*, we then get*

$$\left(\frac{\partial}{\partial t}\mathbf{P}(t)\right)_{x_1,\ldots,x_d} = \sum_{\mu=1}^{M}a_\mu(X-\xi_\mu)P(X-\xi_\mu,t) - a_\mu(X)P(X,t) = \frac{\partial}{\partial t}P(X,t).$$

□

Appendix B

Elementary matrices:

$$L = \begin{pmatrix} -k_{\mathrm{CO}}^{\mathrm{Ad}} & 0 & k_{\mathrm{CO}}^{\mathrm{De}} \\ 0 & 0 & 0 \\ k_{\mathrm{CO}}^{\mathrm{Ad}} & 0 & -k_{\mathrm{CO}}^{\mathrm{De}} \end{pmatrix}, \quad M^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$M^{(2)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M^{(3)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$M^{(4)} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad M^{(5)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$M^{(6)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad M^{(7)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$N^{(1)} = -\begin{pmatrix} k_{\mathrm{CO_2}}^{\mathrm{Ad}} & 0 & 0 \\ 0 & k_{\mathrm{O}}^{\mathrm{Diff}} & 0 \\ 0 & 0 & k_{\mathrm{CO}}^{\mathrm{Diff}} \end{pmatrix}, \quad N^{(2)} = \begin{pmatrix} 0 & k_{\mathrm{O_2}}^{\mathrm{De}} & k_{\mathrm{CO_2}}^{\mathrm{De}} \\ k_{\mathrm{O}}^{\mathrm{Diff}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$N^{(3)} = \begin{pmatrix} 0 & k_{\mathrm{CO_2}}^{\mathrm{De}} & 0 \\ 0 & 0 & 0 \\ k_{\mathrm{CO}}^{\mathrm{Diff}} & 0 & 0 \end{pmatrix}, \quad N^{(4)} = \begin{pmatrix} 0 & k_{\mathrm{O}}^{\mathrm{Diff}} & 0 \\ k_{\mathrm{O_2}}^{\mathrm{Ad}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$N^{(5)} = -\begin{pmatrix} k_{\mathrm{O}}^{\mathrm{Diff}} & 0 & 0 \\ 0 & k_{\mathrm{O_2}}^{\mathrm{De}} & 0 \\ 0 & 0 & k_{\mathrm{CO_2}}^{\mathrm{De}} \end{pmatrix}, \quad N^{(6)} = \begin{pmatrix} 0 & 0 & k_{\mathrm{CO}}^{\mathrm{Diff}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$N^{(7)} = -\begin{pmatrix} k_{\mathrm{CO}}^{\mathrm{Diff}} & 0 & 0 \\ 0 & k_{\mathrm{CO_2}}^{\mathrm{De}} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Id = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

TT-cores:

$$A = [L], \quad B = \left[M^{(1)}, \ldots, M^{(7)}\right], \quad C = \begin{bmatrix} N^{(1)} \\ \vdots \\ N^{(7)} \end{bmatrix}, \quad I = [Id], \quad J = \begin{bmatrix} Id & & \\ & \ddots & \\ & & Id \end{bmatrix}$$