The Findability of Microkinetic Parameters by Heterogeneous Chemical Reaction Neural Networks (hCRNNs)

Hannes Stagge[a], Robert Güttel[a,*]

[a]*Institute of Chemical Engineering, Ulm University, Albert-Einstein-Allee 11, 89081, Ulm, Germany*

Abstract

Finding microkinetic parameters for heterogeneously catalyzed processes with conventional methods is a challenging task. Recently, the use of artificial neural networks has been described as a promising and flexible tool for kinetic parameter estimation. In this work, an extension to the methodology of chemical reaction neural networks (CRNNs) to heterogeneously catalyzed reaction networks (hCRNNs) is proposed. The developed network architecture encapsulates physically interpretable layers for the Arrhenius expression, coverage dependency, and power-law terms encountered in a typical microkinetic model and accounts for possible reversibility of all elementary step reactions in the mechanism. Thus, it is fully interpretable and acts as a drop-in replacement for a conventional kinetic expression.

The methodology is further examined on a prototypical heterogeneously catalyzed reaction mechanism under transient conditions and various operational and kinetic regimes. This work offers a framework for quantifying network errors and interpreting its predictions as well as a systematic overview assessing its ability to identify kinetic parameters. It is found that kinetic behavior is generally described very well by the network. Additionally, kinetic discovery is possible for the fastest reaction in the mechanism, if observed. A link between the results and the transient regime is established. With this, the design of suitable hCRNNs training strategies becomes possible.

Keywords: Kinetic Discovery, Micokinetic Modeling, Neural Networks, Chemical Reaction Neural Networks, Transient Kinetics

*Corresponding author

September 25, 2024

Nomenclature

Latin Symbols

A	$\mathrm{m}^m\,\mathrm{mol}^{-n}\,\mathrm{s}^{-1}$	pre-exponential factor (unit varies)
$b_j, \boldsymbol{b}$	a.u.	bias (vector notation in bold)
$c_i, \mathbf{c}$	$\mathrm{mol}\,\mathrm{m}^{-2}/\mathrm{mol}\,\mathrm{m}^{-2}$	concentration
E_A	$\mathrm{kJ}\,\mathrm{mol}^{-1}$	activation energy
k	$\mathrm{m}^m\,\mathrm{mol}^{-n}\,\mathrm{s}^{-1}$	rate constant (unit varies)
M	$\mathrm{g}\,\mathrm{mol}^{-1}$	molar mass
m	1	number of species
n	1	number of reactions
p	bar	pressure
R	$\mathrm{J}\,\mathrm{mol}^{-1}\,\mathrm{K}^{-1}$	universal gas constant
$r_j, \boldsymbol{r}$	$\mathrm{mol}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$	reaction rate
S	1	sticking coefficient
$\dot{s}_i, \dot{\boldsymbol{s}}$	$\mathrm{mol}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$	formation rate
T	K	temperature
$w_{i,j}, \boldsymbol{w}$	a.u.	weight
$x_i, \boldsymbol{x}$	a.u.	network input
$y_i, \boldsymbol{y}$	a.u.	network output

Greek Symbols

β	1	temperature exponent
Γ_0	$\mathrm{mol}\,\mathrm{m}^{-2}$	surface site density
Δ	–	difference
ϵ	$\mathrm{mol}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$	(overall) loss, error, MAE
ϵ_I	1	integer loss
ε	$\mathrm{kJ}\,\mathrm{mol}^{-1}$	coverage dependency parameter
$\Theta, \boldsymbol{\Theta}$	1	surface coverage (vector notation in bold)
μ	1	coverage dependency parameter
ν	1	stoichiometric coefficients
ν'	1	reaction orders
σ	–	nonlinear activation function

Sub- and Superscripts

ads	adsorption
G	gas phase
i	species index
j	reaction index
ref	reference
S	surface

Acronyms

aNN	artificial neural network
CRNN	chemical reaction neural network
DFT	density functional theory
hCRNN	heterogeneous chemical reaction neural network
LHHW	Langmuir-Hinshelwood-Hougen-Watson
MAE	mean absolute error
MKM	microkinetic model
PTK	periodic transient kinetics method
RDS	rate-determining step

1. Introduction

1.1. The microkinetic formalism

For simulation of industrially relevant chemical processes, source terms of species are critical since they set apart reaction processes from thermal or mechanical separation processes. Generally, the source term for species i ,$\dot{s}_i$, in $\mathrm{mol\,m^{-2}}$ can be written in accordance to their stoichiometric coefficients, $\nu_{i,j}$, and the rate of reaction j, r_j, as eq. (1).

$$\dot{s}_i = \sum_{j=1}^{n} \nu_{i,j} r_j \tag{1}$$

The much debated challenge is describing these formation rates in terms of the current reaction conditions. This functional dependency is referred to as the reaction *kinetics*. However, it is often not trivially accessible, how many reaction steps make up a seemingly simple reaction. This matter is further complicated by the fact that heterogeneously catalyzed reactions only take place at a phase boundary where interactions between bulk phase and catalytic phase (and oftentimes even supposedly inert phases like catalyst support [1]) occur. This results in the fact that the unknown number of reaction steps might be (and more often than not *is*) interconnected in an unknown way by an unknown number of intermediate species at the phase boundary. Each of these reaction rates are in turn described by a set of unknown kinetic parameters linking the reaction conditions to the observed rate. Equation (2) is often used to model the reaction rate for a generic surface reaction j [2].

$$r_j = \underbrace{A_j T^{\beta_j} \exp\left(-\frac{E_{\mathrm{A},j}}{RT}\right)}_{\text{Arrhenius}} \cdot \underbrace{\exp\left(\sum_{i=1}^{m_\mathrm{S}} (\mu_{i,j} \ln \theta_i) + \sum_{i=1}^{m_\mathrm{S}} \left(\frac{\varepsilon_{i,j}\Theta_i}{RT}\right)\right)}_{\text{Cov. Dep.}} \cdot \underbrace{\prod_{i=1}^{m} c_i^{\nu'_{i,j}}}_{\text{Power Law}} \tag{2}$$

Before investigating possible solutions to the parameter fitting problem at hand, the full complexity of eq. (2) has to be broken down into its distinct parts. Firstly, concentration (c_i) dependency of reaction j is conventionally covered by power-law expressions of reactant i and its respective reaction orders, $\nu'_{i,j}$. Temperature (T) dependency is introduced into eq. (2) by an Arrhenius-type expression. The most often used standard parameterization (form eq. (2), hereafter referred to as *standard*) contains only the pre-exponential constant, A_j and activation energy, $E_{\mathrm{A},j}$. However, a temperature exponent, β, may be included for temperature correction of the pre-exponential factor [3]. Note that at a fixed temperature, pre-exponential constant, A, and activation energy, E_A, can not be determined independently of each other. Correlation between the two values show even when classically calculated by regression of temperature-rate data pairs – an effect known as *compensation effect* [4]. It has been shown before that a slight reparameterization of the Arrhenius expression with respect to an arbitrary reference temperature, T^ref, yields enhanced interpretability of kinetic constants [5, 6]. Hence, eq. (3) is implemented in favor of the standard Arrhenius expression in

eq. (2) though they are equivalent and parameters can be transformed between them. Equation (3) will be referred to as the *reparameterized Arrhenius equation* [6].

$$k_j = A_j T^{\beta_j} \exp\left(-\frac{E_{\mathrm{A},j}}{R}\left(\frac{1}{T} - \frac{1}{T^{\mathrm{ref}}}\right)\right) \tag{3}$$

Coverage (Θ_i) dependency of surface reactions is known to significantly impact the performance of microkinetic models (MKMs) and is usually incorporated into the Arrhenius expression by a correction term, $\varepsilon_{i,j}$, influencing the activation barrier of reaction j at high coverages of surface species i [2, 7, 8]. The desorption of CO from Ni surfaces, for example, is known to be encouraged at high coverages [9, 10]. The additional parameter $\mu_{i,j}$ resembling a pseudo power-law expression for surface coverage dependency is rarely tabulated for experimentally or computationally derived kinetics and thus omitted. In other words, we assume $\mu_{i,j} = 0$ throughout this work. It must be noted that this coverage dependency formulation in general applies for all possible surface reactions, adsorptions, and desorptions. However, only the adsorbed surface species are considered when calculating the expression in eq. (2) as indicated by summing up to m_{S}, the number of surface species [2].
Special care is often taken when considering adsorption. Since adsorption kinetics can be described from a statistical thermodynamics viewpoint, they are oftentimes parameterized using the sticking coefficient, S_j, following the formalism depicted in eq. (4)[2].

$$k_j^{\mathrm{ads}} = \frac{S_j}{(\Gamma_0)^{\mu}} \sqrt{\frac{RT}{2\ \pi\ M_i^{\mathrm{ads}}}} \tag{4}$$

$$S_j = S_{0,j}\ T^{\beta_j} \exp\left(-\frac{E_{\mathrm{A},j}}{RT}\right) \tag{5}$$

Herein k_j^{ads} denotes the rate coefficient of reaction j if it is an adsorption, μ is the number of reactant molecules and sites partaking in the adsorption ($\mu = 2$ in most cases e.g. simple adsorption), M_i^{ads} is the molar mass of the gas-phase molecule adsorbed in that step and Γ_0 denotes the overall active site density of the catalytic material. It can be shown easily that above parameterization is consistent to the overall formalism presented in eqs. (1) to (3) and can be rewritten accordingly. For more information see appendix SI A.
One can easily see that the number of unknowns in this system quickly explodes even for simple systems [11] leading to the development of simpler (almost surrogate) kinetic expressions. The most commonly used kinetic expressions are of the Langmuir-Hinshelwood-Hougen-Watson (LHHW)-type. They derive from assuming a simple surface reaction network of adsorption and reaction steps under the assumption of equilibrated and rate-determining steps (RDSs). Therefore, they are analytically closed, obey the respective adsorption and reaction equilibria and reproduce kinetic behavior near steady-state [12].
Unfortunately, it has been shown repeatedly that the approximation of surface processes by steady-state kinetics does not provide sufficient detail to derive insights into the system behavior as it undergoes transient changes [13, 14, 15]. To encapsulate these dynamic processes MKMs that contain all the aforementioned complexity are sought.

1.2. Finding microkinetic parameters

As described above, there are many parameters to be determined when identifying a MKM. This process of parameter identification is of critical importance for the success of the endeavor. Unfortunately, conventional parameter fitting methods often struggle in multi-parameter settings making brute force identification of kinetic parameters unfeasible [16, 17]. The use of first-principle data obtained by computational chemistry like density functional theory (DFT) might eliminate the need for parameter fitting routines but comes at extremely high computational costs making it equally unfeasible to apply to complex multi-component systems [12, 11]. Thus, many of the currently employed approaches attempt to reduce the parameters (and thereby the required computational time) as far as possible. The different ways this is achieved separates the

authors: While some start from the most complex possible mechanism and systematically reduce complexity by employing knowledge about discovered RDS on the fly [7, 18], others take the inverse approach and start at simple models. After obtaining reasonable agreement with experimental data, further complexity is added to describe deviations only where needed [16]. Others again circumvent expensive DFT-calculation and use shortcut methods estimating kinetic parameters only to dynamically refine them where needed [11]. All three approaches are viable but still require some amount of parameter fitting to scale results from DFT or approximate them.
Though progress towards reliable global minimization algorithms is being made, at the time of writing this article no catch-all solution exists. Chacko et al. [17] recently revisited already published work and were able to enhance prediction accuracy by fitting kinetic parameters through basin hopping, a minimization algorithm based on local exploitation with simultaneous global exploration steps [17, 10]. This success once again underlines the importance of accurate parameter identification when attempting to combine first-principle with experimental data by MKMs.
From an industrial modeling standpoint however, the identification of physically accurate parameters might not even be necessary. For process design a reasonably accurate rate prediction is oftentimes enough. This realization and the advent of cheap computational power lead to the development of *surrogate* machine-learning driven models reproducing kinetic behavior over a wide range of conditions. These range from black-box-style deep neural networks [19] over physically-informed semi-interpretable kinetic expressions [20, 21, 22] into fully interpretable digital twins of physical and chemical processes. Especially the latter are of great interest since they allow for the accurate prediction of reaction rates with the added benefit of gaining insight into fundamental kinetic properties albeit from a phenomenological perspective. The chemical reaction neural network (CRNN) formalism as an example for such a digital twin will be the foundation of this work.

1.3. CRNNs

The CRNN formalism introduced by Ji and Deng [23] takes advantage of the superior parameter fitting capabilities of artificial neural networks (aNNs) for identifying kinetic constants. The underlying structure makes use of the similarity between the parameterization of the rate constant eq. (6) in log-scale, the stoichiometric relation eq. (1) and the general mathematical formulation of a single layer in multilayer aNNs. By construction, CRNNs obey the above mentioned well-known physicochemical laws of chemical kinetics for homogeneous reaction systems. This not only guarantees consistency with physical intuition but also ensures interpretability of the network discovered constants. As for most known machine learning algorithms, results are obtained by *training* – the process of fitting the network parameters to match the network output to supplied training data. The sufficiently trained CRNN is a ready to use, drop-in replacement for conventional kinetic expressions able to reproduce kinetic behavior and yields not only the usable kinetic expression but also guesses for the kinetic parameters. Applications of this concept have been described for a plethora of nontrivial reaction networks in conjunction with different reactor and analysis setups with great success [23, 24, 25].
Further improvements upon the neural network enabled differential equation solver allows to couple the results to the reactor scale more robustly [26]. In terms of network construction, parameter reduction was achieved by introduction of mass conserving stoichiometry layers through identifying the null space of the stoichiometric matrix [27]. The use of Bayesian inference has been demonstrated to further allow for uncertainty quantification of the retained parameters [28].

1.4. Objectives of the hCRNN framework

Despite its potential, the applicability of CRNNs has not yet been demonstrated for heterogeneous catalysis and its inherent problems like different time scales for sorption processes and reactions or the presence of many short-lived or hard to measure intermediates. Furthermore, although contained within the CRNN framework, the reversibility of reactions has never been explicitly described and thus no clear distinction between forward and backward reactions can be made in the original formalism. In regards to these challenges, the heterogeneous chemical reaction neural network (hCRNN), an extension to the homogeneous

case, is presented to address them. Consequently, the constructed framework will be systematically evaluated for plausible kinetics encountered in heterogeneous catalysis. Preliminary experiments, which employed the new formalism in conjunction with complex kinetics, suggested the existence of an intrinsic link between the kinetic properties of the mechanism under consideration and the training success of the hCRNN, which is to be investigated in this work. Thus, the focus of this work is to unveil the possibilities and limitations of parameter identification of hCRNNs for highly transient generic heterogeneous reaction networks. As will be shown hereafter, aforementioned features of heterogeneous reaction systems strongly effect the interpretability of the obtained results.

2. HCRNN Construction

Looking back at eq. (1) and eq. (2) rephrased in logarithmic scaling, reveals the similarity of eqs. (1) and (6) to the general mathematical formulation of a single aNN layer depicted in eq. (7).

$$r_j = \exp\left(\underbrace{\ln A_j + \beta_j \ln T - \frac{E_{\mathrm{A},j}}{R}\left(\frac{1}{T} - \frac{1}{T^{\mathrm{ref}}}\right)}_{\text{Arrhenius}} + \underbrace{\sum_{i=1}^{m_S} \frac{\varepsilon_{i,j}\Theta_i}{RT}}_{\text{Cov. Dep.}} + \underbrace{\sum_{i=1}^{m} \nu'_{i,j} \ln\left(c_i\right)}_{\text{Power Law}}\right) \tag{6}$$

$$y_{l,j} = \sigma\left(\sum_i w_{l,i,j} x_i + b_{l,j}\right) \tag{7}$$

While earlier works lumped much of the complexity into one single layer [23], this work aims to develop a more strongly separated network architecture in the open-source deep learning framework PyTorch[29]. This leads to the use of PyTorch's parameter groups that bundle similar parameters into distinct units and is highly beneficial, since manual handling of the larger number of parameters impedes interpretability. In other words, different parts of any generic microkinetic model get their own dedicated layers. This is also apparent from the grouping of stoichiometric coefficients, reaction orders and the parameters for the Arrhenius- and surface coverage dependency expressions in eqs. (1) and (6). This compartmentalization empowers efficient handling, interpretation and manipulation of those parameters. An overview of the proposed network architecture is depicted in fig. 1. This shows the general hCRNN structure as a parameterized function $\mathcal{N}(\boldsymbol{w}_l, \boldsymbol{b}_l, \boldsymbol{\beta}) : (\boldsymbol{c}, T) \mapsto \dot{\boldsymbol{s}}_i$ as well as the breakdown into dedicated layers. The single layer expressions resemble the terms in eq. (2):

$$\text{Power Law} \qquad P_j(\mathbf{c}) = \sum_{i=1}^{m} w_{P,i,j} \ln\left(c_i\right) \tag{8}$$

$$\text{Arrhenius} \qquad A_j(T) = b_{A,j} + \beta_j \ln\left(T\right) + w_{A,j}\left(\frac{1}{T} - \frac{1}{T^{\mathrm{ref}}}\right) \tag{9}$$

$$\text{Cov.Dep.} \qquad C_j(T, \boldsymbol{\Theta}) = \sum_{i=1}^{m_S} w_{C,i,j}\Theta_i \frac{1}{T} \tag{10}$$

$$\Theta_i = \frac{c_i^{\mathrm{S}}}{\sum_i c_i^{\mathrm{S}}} \tag{11}$$

By comparing, one can identify the weights ($\boldsymbol{w}$), biases ($\boldsymbol{b}$) and additional parameters ($\boldsymbol{\beta}$) of the layers (abbreviated as A, C and P) with the kinetic constants (or simple expressions containing them). Thus, the

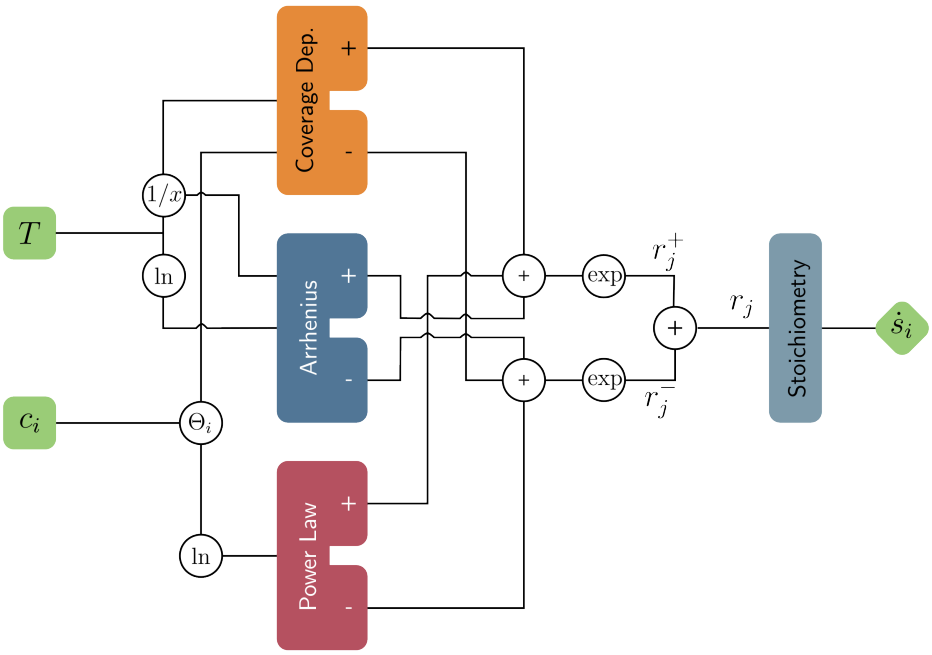


Figure 1: Proposed network architecture. Each colored box represents a fully connected linear layer (denoted A, C and P in eqs. (8) to (10) and eq. (12)) with sub-layers for forward and reverse rate indicated.

overall rate expression can be written as follows:

$$r_j = \exp\left(A_j(T) + C_j(T, \boldsymbol{\Theta}) + P_j(\mathbf{c})\right). \tag{12}$$

Logarithmic scaling is used on the inputs of concentration and temperature [23]. Additionally inverse scaling is applied to the temperature [23]. Surface coverages are calculated from the passed in concentrations for all species marked as present on the surface (compare eq. (11)).

To ensure stoichiometric consistency splitting of forward and backward reactions into two separate sub-layers according to Barwey and Raman [30] is employed. By using the notion that $\nu^+_{i,j} = -\nu^-_{i,j}$ and assuming elementary steps for all surface, adsorption, and desorption reactions (i.e. $\nu'_{i,j} = -\nu_{i,j}$ for reactants and $\nu'_{i,j} = 0$ for products) one can derive a formulation relying only on the stoichiometric coefficients of the forward reactions to accurately describe both the reaction orders and the stoichiometry of forward *and* backward reactions at the same time. To our knowledge the proposed extension is novel to the CRNN methodology and extends usability to equilibrium reactions without doubling the amount of network parameters by singularly including both the forward and backward reaction. These common and well documented assumptions aid in reducing the number of trainable parameters in the power-law and stoichiometry layer from 4 $(m \times n)$ for the fully explicit notation down to $(m \times n)$ for the elementary step notation. During preliminary testing, this splitting has significantly sped up training and increased interpretability of the obtained results for reaction systems as small as $m = 2$ and $n = 2$. Recently, attempts have been made to further decrease the number of parameters by using the fact that atoms have to be conserved during any reaction [27]. Examining this methodology is however, not the focus of this work.

This construction leaves only one major hyperparameter up to the users choice: The number of neurons in each layer (indicated by the index j) which corresponds directly to the number of elementary reactions sought for in the microkinetic model [23]. Thus, even this a priori undetermined hyperparameter gains physical interpretability and can be used to reason about the reaction system at hand. Note that generally hyperparameters are not available through training but have to be chosen beforehand according to some heuristics. In our specific case, an educated guess towards the correct choice of the number of neurons can

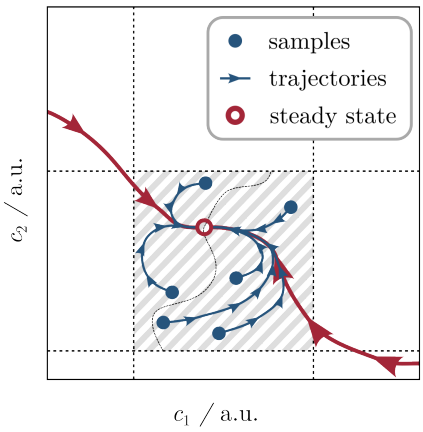


Figure 2: Schematic of the sampling strategy. Different samples within the valid state space develop along trajectories far from steady-state.

be made by prior insight into or literature knowledge about the investigated system.

3. Training Data Generation

The preparation of meaningful training data is of utmost importance for the training success of any machine learning project including physically motivated aNNs [31]. It has to be sampled from a wide yet physically feasible field of possible conditions. The data collection will proof to influence training success heavily, as is reported in literature [32]. In context of the here presented hCRNNs the goal is to predict species formation rates, $\dot{s}_i$, from data containing gas and surface phase compositions, c_i^{G} and c_i^{S} (or simply c_i in short), and temperature, T. The latter three are hereafter referred to as a *state* and directly sampled while the formation rates are state-dependent. The sampling strategy is schematically depicted in fig. 2: From a range of possible states a number of points is chosen uniformly [32]. Those states are likely to be distant from equilibrium or steady-state (compare trajectories in fig. 2) and thus show strongly kinetic behavior. The proposed sampling strategy is only concerned with the initial reaction rates at these conditions and contain no temporal information about the overall progression towards steady-state. This kind of data is referred to as being *differential* for it does not contain any information about its history. We refer to this setup as the *maximally transient* regime because it allows for observation of kinetic effects far from steady-state.
Specifically, a two phase system containing a catalytic surface in contact with a gas phase is considered and studied. Transport limitations are neglected. For this system, a single system state is generated by sampling gas and surface phase compositions (x_i and Θ_i) according to a flat Dirichlet distribution [33]. Concentration measures are subsequently derived from the surface site density, Γ_0, and the gas phase pressure, p, and temperature, T, according to

$$c_i^{\mathrm{G}} = x_i \frac{p}{RT} \tag{13}$$

$$c_i^{\mathrm{S}} = \Theta_i \Gamma_0. \tag{14}$$

Temperature (and pressure, respectively) is either taken as a fixed value or sampled according to the uniform distribution, U, given by a mean value ($\overline{T}$) and a half-range (ΔT) as $U\big(\overline{T} - \Delta T, \overline{T} + \Delta T\big)$. To ensure reproducibility, all random number generators are seeded before sampling.
For each thus generated state the formation rate of species i, $\dot{s}_i$, is calculated assuming a ground truth reaction mechanism. Hence, one set of $\dot{s}_i$ is derived from a sampled set of x_i and Θ_i.
In this work the focus is set on a generic yet universal Langmuir-Hinshelwood type mechanism of eight species present both in the gas and surface phase partaking in four reversible reactions. An additional inert species

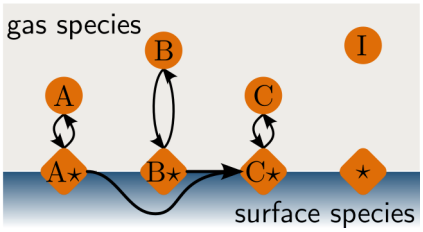


Figure 3: Schematic of the chosen reaction mechanism.

Table 1: Reaction mechanism and kinetic parameters of the prototypical Langmuir-Hinshelwood kinetics. †Value varied for study of reaction rate regime. ‡Value varied for study of coverage dependency.

	reaction			A_j s^{-1}	$E_{A,j}$ kJ mol^{-1}	$\varepsilon_{i,j}$ kJ mol^{-1}
R1	$A + \star$	$\longrightarrow$	$A\star$	8	0	0
	$A\star$	$\longrightarrow$	$A + \star$	2×10^{6}	90	$-40\Theta_A$ ‡
R2	$B + \star$	$\longrightarrow$	$B\star$	30	0	0
	$B\star$	$\longrightarrow$	$B + \star$	3×10^{6}	80	0
R3	$C + \star$	$\longrightarrow$	$C\star$	5×10^{-2}	0	0
	$C\star$	$\longrightarrow$	$C + \star$	5×10^{4}	40	0
R4	$A\star + B\star$	$\longrightarrow$	$C\star$	6×10^{13} †	70	0

is added. The mechanism is illustrated in fig. 3 and the chosen ground truth kinetic constants tabulated in table 1 in full complexity. Adsorptions are taken to be non-activated. Desorption of $A\star$ is promoted at high coverages by itself, loosely inspired by aforementioned self-interactions of CO on Ni surfaces [10, 9]. Product readsorption is highly discouraged but not impossible. The rate constants have been constructed to expect reactions to proceed with roughly the same net rate at stoichiometric conditions and $T = 550\,\mathrm{K}$.
To investigate the influence of kinetic parameters isolated from each other, multiple simplified and varied mechanisms are derived from the prototype described in table 1 by, e.g, varying the surface reaction rate of R4. An exhaustive overview over all generated mechanisms can be taken from appendix SI B.
All samples generated with the same operational and mechanistic settings make up a data set and are used for training. Through combination of different operational and mechanistic settings multiple datasets can be dynamically created. The default data set contains 10 000 individual states. 20 % of a data set is used as holdout (or test) data not supplied during network training.

4. Training Process

During training, deviation between network prediction ($\hat{\dot{\mathbf{s}}}$) and ground truth value ($\dot{\mathbf{s}}$) is calculated according to the L^1-norm [34] (more commonly referred to as MAE) (compare eq. (15)) as implemented in PyTorch[29].

$$\epsilon = L^1(\dot{\mathbf{s}} - \hat{\dot{\mathbf{s}}}) = \frac{1}{n}\sum_{i=0}^{n} |\dot{s}_i - \hat{\dot{s}}_i| \tag{15}$$

Error minimization is conducted using the well known backpropagation algorithm coupled with the ADAM-optimizer for network parameter optimization [35]. The initial learning rate was chosen to be $\mathrm{lr} = 4\times10^{-3}$ with a custom decay schedule reducing it by a factor of $\lambda = 0.72$ whenever no significant training success is achieved over 80 epochs. Identification of suitable network and training hyperparameters was conducted preliminary on test problems using Bayesian optimization to minimize the loss on the holdout data set. Training proceeds for 8000 epochs after which no significant reduction of loss was observed for most experiments conducted.
After the training epochs, one test step is performed on the holdout data set. During this step, L^1-loss is calculated as well. However, since interpreting L^1-loss in terms of prediction accuracy is difficult, another

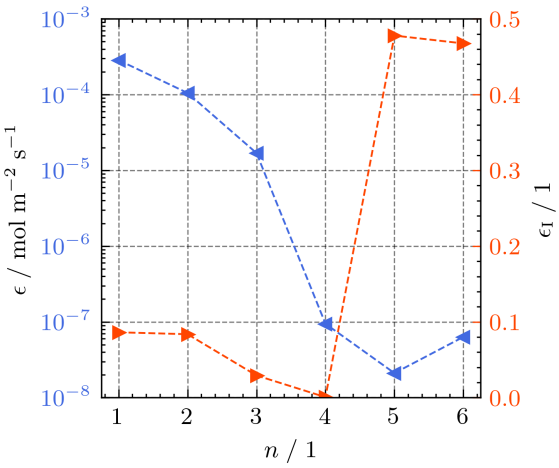


Figure 4: Loss, ϵ, (blue, left ordinate) and integer loss, ϵ_I, (orange, right ordinate) for the trained hCRNN at $T = 550\,\mathrm{K}$ under variation of n. The ground truth mechanism contains 4 reactions and is shown in table 1. Stoichiometric coefficients for the point at $n = 4$ match the ground truth stoichiometry as tabulated in table 2.

measure on the *network parameters* is proposed.

$$\epsilon_\mathrm{I} = \max_{i,j} |\lfloor \nu_{i,j} \rceil - \nu_{i,j}| \tag{16}$$

Equation (16) will be referred to as *integer loss* and yields a value $\epsilon_\mathrm{I} \in [0, 0.5]$ representing the maximum difference of any network identified stoichiometric coefficient from the nearest integer (rounding to the nearest integer is denoted using $\lfloor \cdot \rceil$). With this metric one can asses how well elementary step assumption is obeyed by the identified parameters without knowing the ground truth constants as such. Note that no parameter or error measure used in training or testing is derived from knowing ground truth information. The proposed training strategy is designed specifically to formulate heuristics for applying the procedure to data *without* knowledge of the underlying mechanism or kinetic constants.

5. Results

5.1. Isothermal Data

Figure 4 shows loss, ϵ, and integer loss, ϵ_I, for hCRNNs with a differing number of presumed elementary reaction steps, n. As described above, this hyperparameter has to be chosen by the user and is critical for the network to identify the reactions pathways within the mechanism. To select a fitting value, n is linearly increased during a hyperparameter search and the error metrics of the trained network are observed during screening. Generally, small losses, ϵ, indicate overall low errors for formation rate predictions, while low integer losses, ϵ_I, indicate very little deviation from integer stoichiometric coefficients and thereby the elementary-step assumption.

For the chosen reaction network, the loss significantly decreases between $n = 3$ and $n = 5$ while the integer loss shows the opposite behavior. Focusing only on overall loss might lead the reader to choose the local (possibly global) minimum at $n = 5$ and consequently assume 5 elementary reactions, neglecting the fact that decrease in loss from $n = 4$ to 5 is far less significant than that from $n = 3$ to 4. At the same time the strong rise in integer loss ϵ_I indicates the presence of at least one row with a stoichiometric coefficient far from the next integer (for explicit results the reader is referred to appendix SI C). Experience shows that this typically coincides with completely chaotic stoichiometric coefficients in that row.

Both facts combined reveal the overall tendency repeatable throughout all conducted experiments: If too

Table 2: Stoichiometric coefficients, $\nu_{i,j}$, of forward elementary reaction steps predicted by the hCRNN at $p = 1\,\mathrm{bar}$ and $T = 550\,\mathrm{K}$.[1]

		$\nu_{\mathrm{A}\star}$	$\nu_{\mathrm{B}\star}$	$\nu_{\mathrm{C}\star}$	$\nu_{\star}$	ν_{A}	ν_{B}	ν_{C}	ν_{I}
R1	$\mathrm{A} + \star \rightleftharpoons \mathrm{A}\star$	1.00	0.00	0.00	−1.00	−1.00	0.00	0.00	0.00
R2	$\mathrm{B} + \star \rightleftharpoons \mathrm{B}\star$	0.00	1.00	0.00	−1.00	0.00	−1.00	0.00	0.00
R3	$\mathrm{C} + \star \rightleftharpoons \mathrm{C}\star$	0.00	0.00	1.00	−1.00	0.00	0.00	−1.00	0.00
R4	$\mathrm{A}\star + \mathrm{B}\star \longrightarrow \mathrm{C}\star + \star$	−1.00	−1.00	1.00	1.00	0.00	0.00	0.00	0.00

few elementary steps are presumed the network fails to capture the expected kinetic behavior (high loss, ϵ); the overall kinetic model does not contain enough parameters. Choosing too many elementary steps might depict the overall kinetic behavior well but leads to no insights into the stoichiometry of the mechanism; the overall kinetic model is overparameterized and at least one row of the stoichiometric matrix does contain parameters without physical meaning.
For the present case, prediction of the observed formation rates for the species at given (variable) gas phase compositions and constant temperature of $T = 550\,\mathrm{K}$ is possible for a number of presumed reactions of $n = 4$ with a low loss of $\epsilon = 9.448 \times 10^{-8}\,\mathrm{mol\,s^{-1}\,m^{-2}}$. At the same time stoichiometric discovery is still possible as the integer loss is virtually zero. The effect can be seen in table 2: The network is able to identify the stoichiometry of all 4 elementary reactions in remarkable agreement with the ground truth mechanism.
It is apparent that no additional mechanistic information is gained by assuming more reactions than present in the ground truth mechanism and that knowledge about the integerness of the stoichiometric coefficients can be used as a tool to reject kinetically feasible but stoichiometrically overparameterized mechanisms proposed by hCRNNs.

5.2. Influence of rate regime

Most reaction systems undergo a change of the reaction rate regime when different reaction conditions are present. Therefore, the reaction rate constant of the surface reaction, k_{S}, is used as a proxy to investigate the reaction system from the surface reaction being rate limiting to quasi-equilibrated. Since the reaction network was constructed to be approximately equally fast on all reaction routes at an arbitrarily chosen reference state at $T^{\mathrm{ref}} = 550\,\mathrm{K}$, the resulting k_{S} is chosen as a reference and denoted with the superscript ref as well (eq. (17)).

$$k_{\mathrm{S}}^{\mathrm{ref}} = A_{0,4}^{+} \exp\left(-\frac{E_{\mathrm{A},4}^{+}}{RT^{\mathrm{ref}}}\right) \tag{17}$$

This means for $k_{\mathrm{S}} \ll k_{\mathrm{S}}^{\mathrm{ref}}$ the surface reaction is rate limiting while for $k_{\mathrm{S}} \gg k_{\mathrm{S}}^{\mathrm{ref}}$ the surface reaction is quasi-equilibrated. Figure 5 reveals an overall linearly increasing loss, ϵ, with increasing k_{S} (especially for $n = 3$ and $n = 4$) which is explained by the chosen error measure. Assuming constant relative errors in rate prediction, the chosen L^1-norm computes errors proportional to the predicted rates. Since the rate of the surface reaction is purposefully increased, it is expected to see a linear rise in losses with increasing reaction rate constant. Apart from this trend, a distinct bifurcation of losses, ϵ, for $n = 3$ and $n = 4$ with a minimum at $k_{\mathrm{S}} = 1 \times 10^{7}\,\mathrm{m^3\,mol^{-1}\,s^{-1}}$ for $n = 4$ is apparent. This minimum unsurprisingly corresponds to the identifiable ground truth mechanism at the point of reference $k_{\mathrm{S}}^{\mathrm{ref}}$ where all reactions are equally fast by construction.
The overall loss, ϵ, for $n = 3$ and $n = 4$ to the left of the bifurcation in fig. 5 (slow surface reactions, region I) behaves very similarly, while the integer loss, ϵ_{I}, simultaneously is maximal for $n = 4$ and $n = 5$. Hence, the low loss, ϵ, suggest the existence of three or four reactions, while the high integer loss, ϵ_{I}, actively contradicts four or more reactions. Consequently, $n = 3$ reactions suffice to describe kinetic behavior while retaining kinetic and stoichiometric information in region I. This observation might seem contradictory to

[1] Reaction indices have been rearranged to match order and sign in the ground truth mechanism for simplicity in all tables and figures.

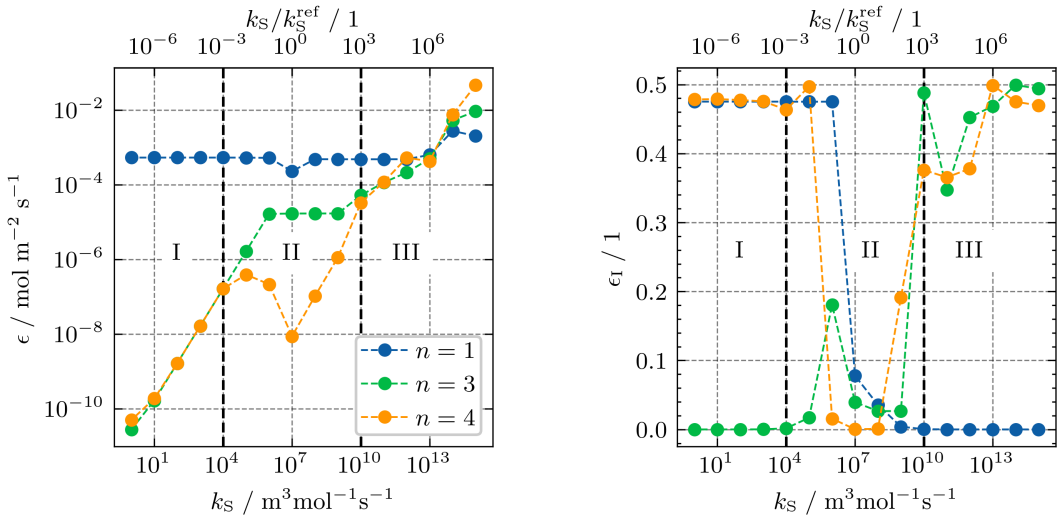


Figure 5: Loss, ϵ, (a) and integer loss, ϵ_I, (b) for the trained hCRNN at variable surface reaction rate constants k_S under variation of n.

the expected behavior assuming four reactions present in the ground truth mechanism. The unidentified ("missing") reaction is tied to the limited observability of slow processes in the chosen maximally transient regime. Under these conditions, the formation of adsorbed species by surface reaction is far outpaced by the formation via sorption processes. Thus, identifying the three main sorption reactions R1, R2 and R3 is sufficient for describing overall species formation rates.

A similar argument can be made in case of the surface reaction being fast in comparison to the sorption processes (high values of k_S, region III). In this regime, loss, ϵ, and integer loss, ϵ_I, eventually ($k_S > 1 \times 10^{13}\,\mathrm{m^3\,mol^{-1}\,s^{-1}}$) predict the existence of only one reaction step since both are lowest for $n = 1$. Again, species formation rates are dominated by a subset of fast reactions, the surface reaction in this case, and only one elementary reaction step is enough to describe observed data by our criteria. Only the intermediate regime (region II at $k_S \approx 1 \times 10^4 - 1 \times 10^{10}\,\mathrm{m^3\,mol^{-1}\,s^{-1}}$) is better explained by the set of all four reactions as shown by the simultaneously low losses, ϵ and ϵ_I. A visual representation of what it means for the reaction system to be dominated by a subset of reactions can be taken from fig. 6: For both region I and III the logarithmic difference in reaction rate far exceeds 4, meaning the slowest reactions is more than 10 000 times slower than the fastest one – a rather difficult case for observing kinetic behavior by any means. The region of kinetic discovery (II) aligns well with the minimum indicating that maximal kinetic discovery seems feasible in the range where the logarithmic rate difference of the reactions is < 4; or in other words, if most of the reactions happen on similar time scales. It should be mentioned again that the presented method *does* in fact represent the dominating kinetic behavior of the reaction system whether it is dominated by a subset of fast reactions or not. In certain cases this just leads to a loss in kinetic information – quite similar to the description of kinetics using macrokinetic expressions like the LHHW.

5.3. Pressure Influence

Variation of operating pressure reveals the dependency of reaction rate with total concentration expressed in the power-law in eq. (2). Because of the elementary step assumption, reaction orders, $\nu'_{i,j}$, and stoichiometric coefficients, $\nu_{i,j}$, are directly coupled and thus observing reaction order influence simultaneously enables stoichiometric discovery and thus carries important kinetic information. To demonstrate these effects, numerical experiments at varying pressures are conducted. Once again, fig. 7 shows very distinct regions separated into near ambient ($p \leq 10\,\mathrm{bar}$) and high pressures ($p > 10\,\mathrm{bar}$).

Since adsorption reaction rates are proportional to component partial pressure the adsorptions of reactants

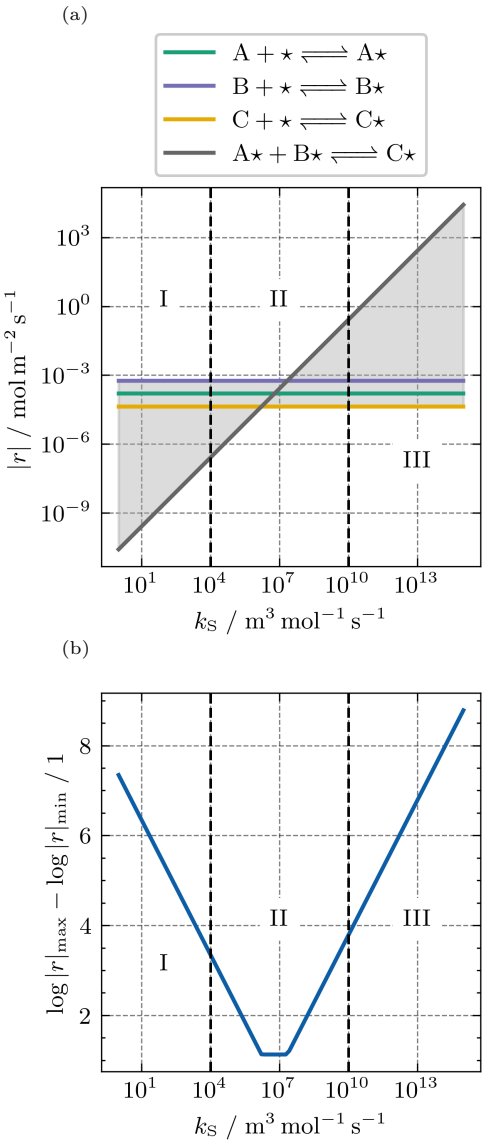
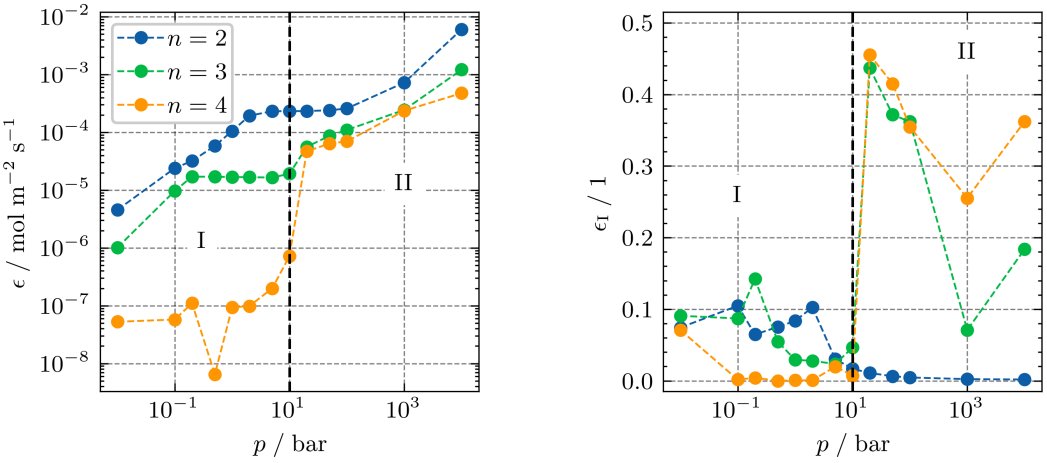


Figure 6: (a) Absolute net rate of elementary reaction steps with varying surface reaction rate constant k_S. (b) Logarithmic difference between fastest and slowest reaction rate observed. Median rates for all training samples depicted.



(a) (b)

Figure 7: Loss, ϵ, (a) and integer loss, ϵ_I, (b) for the trained hCRNN at variable pressure, p, under variation of n. Stoichiometric coefficients for $n = 4$ and $p = 100\,\mathrm{bar}$ tabulated in table 3.

A, B and C dominate the reaction network for high pressures (region II) and can be easily identified. This is discernible from approximately equal losses, ϵ, for $n = 3$ and $n = 4$ in this range, while the integer loss for $n = 3$ tends towards lower values. This again supports the idea of the fast reaction being found preferably over the slower ones.

Table 3 shows the stoichiometric coefficients for the $n = 4$ reactions identified by the hCRNN for an exemplary experiment at $p = 100\,\mathrm{bar}$. Note the deviation in R3 and R4 influencing integer loss. From that it seems that the adsorption of C (reaction R3) is not identified entirely by the hCRNN on its own. Since product adsorption, although possible, is constructed to be severely hindered even at elevated pressures it gets overshadowed by the much faster adsorptions of A and B. However, stoichiometry is reasonably close to the correct coefficients for R3 and stoichiometrically consistent for R4. To illustrate this further, consider fig. 8. It is apparent that the logarithmic difference in rates strongly increases with higher pressures, meaning stoichiometric discovery can not be guaranteed. However, note that the adsorption of C still takes place significantly fast, when compared to the surface reaction, yet can not quite compete with reactant adsorption of A and B. Nonetheless, its influence is not entirely negligible resulting in a slight increase in loss, ϵ, for $n = 2$ when compared to $n = 3$ (compare fig. 7a, region II). This in term results in a partial discovery of the product readsorption process at higher pressures.

For lower pressures (region I) adsorption rates decrease with decreasing pressure while desorption and surface reaction rates remain constant (compare fig. 8, region I). Thus, for low pressures the sorption equilibrium eventually shifts towards desorbed species with net formation rates dominated by the desorption processes, which are independent of pressure. This leads to the identification of $n = 4$ reactions even to the lowest pressure sampled. Compare again the low and nearly constant overall and integer loss in fig. 7 for $n = 4$ for low pressures. The stoichiometric consistency of the described hCRNN allows for the identification of all four steps, since the sign (direction) of the reaction step is irrelevant for its stoichiometric discovery by construction, meaning the forward pathway can be discerned from the observed backward reactions.

5.4. Temperature Influence

So far, no information regarding the influence of temperature on the reaction rate (the Arrhenius part of eq. (2)) could be extracted from the conducted experiments. To investigate kinetic discovery of activation barriers, a data set containing temperatures sampled from a range ($T^\mathrm{ref} \pm \Delta T$) is hereafter used for training.

Figure 9 shows loss, ϵ, and integer loss, ϵ_I, after network training on the nonisothermal data set comparing both the standard (eq. (2)) as well as the reparameterized (eq. (3)) Arrhenius expression. The difference in

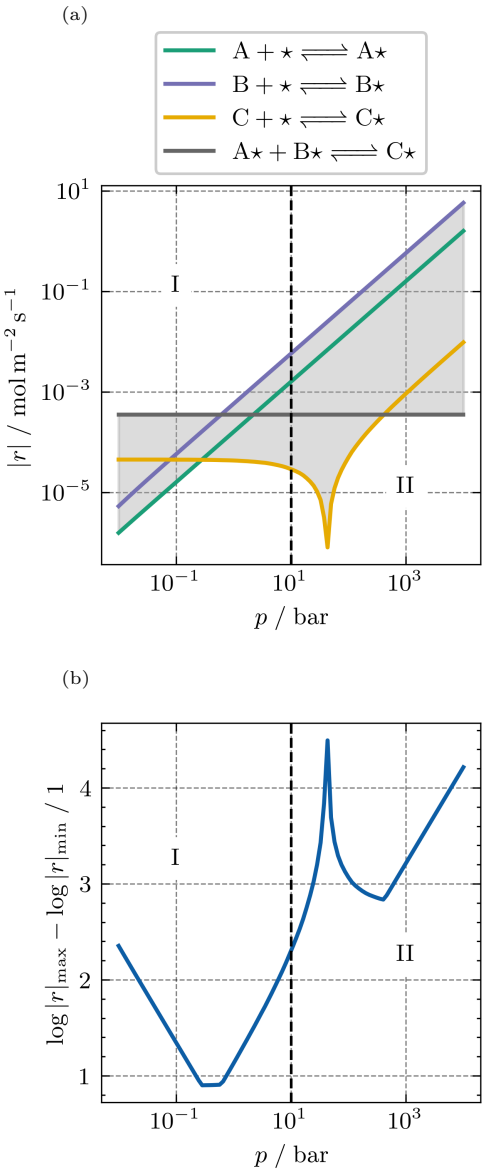


Figure 8: (a) Absolute net rate of elementary reaction steps with varying pressure, p. (b): Logarithmic difference between fastest and slowest reaction rate observed.

Table 3: Stoichiometric coefficients of forward elementary reaction steps predicted by the CRNN with $n = 4$ at $p = 100\,\mathrm{bar}$ and $T = 550\,\mathrm{K}$.

		$\nu_{\mathrm{A}\star}$	$\nu_{\mathrm{B}\star}$	$\nu_{\mathrm{C}\star}$	$\nu_{\star}$	ν_{A}	ν_{B}	ν_{C}	ν_{I}
R1	$\mathrm{A} + \star \rightleftharpoons \mathrm{A}\star$	1.00	0.00	0.00	−1.00	−1.00	0.00	0.00	0.00
R2	$\mathrm{B} + \star \rightleftharpoons \mathrm{B}\star$	0.00	1.00	0.00	−1.00	0.00	−1.00	0.00	0.00
R3	$\mathrm{C} + \star \rightleftharpoons \mathrm{C}\star$	0.00	0.00	0.84	−1.00	0.00	0.00	−0.97	0.00
R4	$\mathrm{A}\star + \mathrm{B}\star \longrightarrow \mathrm{C}\star + \star$	−0.35	−0.35	0.32	0.30	0.00	0.00	0.00	0.00

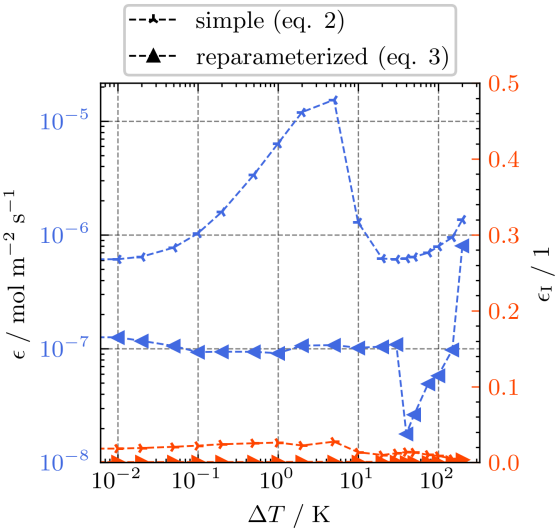
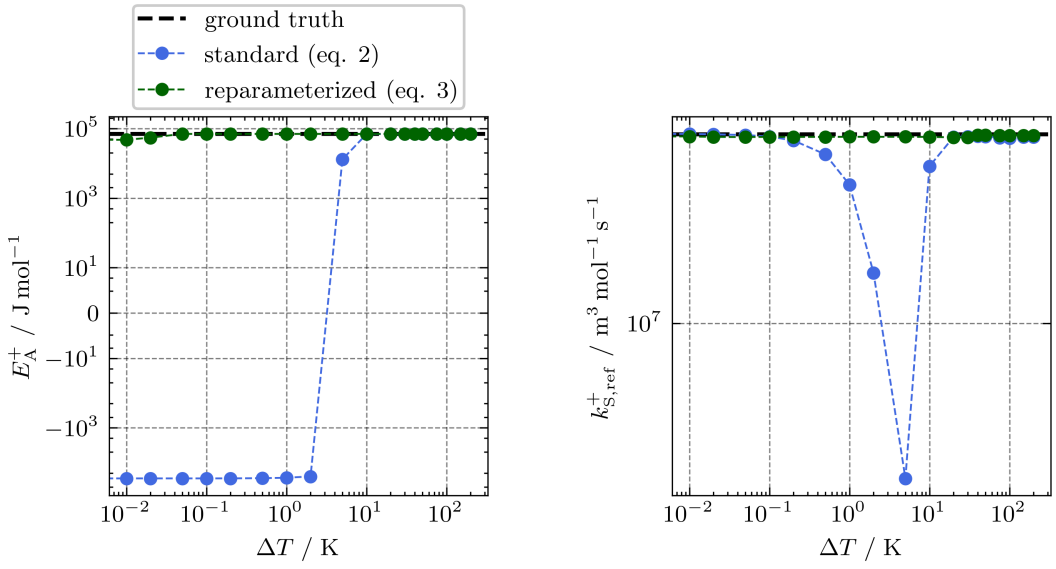


Figure 9: Loss, ϵ, (blue, left ordinate) and integer loss, ϵ_I, (orange, right ordinate) for the trained hCRNN with data uniformly sampled from the interval $\left[T^\mathrm{ref} - \Delta T, T^\mathrm{ref} + \Delta T\right]$. Depicted are the standard (eq. (2)) and the reparameterized (eq. (3)) Arrhenius expression.



(a) (b)

Figure 10: hCRNN identified parameters in comparison to the ground truth (ideal) values of (a) the activation energy, E_A^+, and (b) the rate constant, $k^+_{\mathrm{S},T^\mathrm{ref}}$, evaluated at $T^\mathrm{ref} = 550\,\mathrm{K}$. Both values are calculated for the forward reaction of R4 for variable training temperature ranges (ΔT) and contrasted for both described parameterization methods.

loss, ϵ, between both parameterizations is striking. With only minor deviations from integer stoichiometric coefficients (compare ϵ_I), stoichiometry can be discovered over the whole training range with both methods. With the stoichiometry recovered for all sampled temperature ranges and constant integer losses, ϵ_I, (compare fig. 9), the difference in overall loss can solely be attributed to the (miss)identification of the pre-exponential constants, A, and activation energies, E_A, differing between both parameterizations. From these results it is apparent that the reparameterized expression performs better in terms of minimizing error alone.
Making use of the physically informed network architecture, the network predicted surface reaction rate constant at T^ref, $k^+_\mathrm{S,ref}$, is calculated for each ΔT (eq. (18)) and depicted in fig. 10 alongside the network-identified absolute value of the activation energy, E_A.

$$k^+_\mathrm{S,ref} = k^+_\mathrm{S}\left(T^\mathrm{ref} = 550\,\mathrm{K}\right) \tag{18}$$

From fig. 10 it is quite apparent that the reparameterized Arrhenius equation significantly outperforms the standard one not only in terms of overall loss ϵ but also in interpretability of the recovered results: Figure 10a shows that, while the standard formulation correctly discovers activation energies, E_A, only for larger temperature ranges, ΔT, the reparameterized formulation yields good agreement over all but the smallest temperature ranges ($\Delta T < 0.1\,\mathrm{K}$). Figure 10b in contrast supports the overall reparameterization idea of referring the constants to a reference state: The rate constant at reference conditions, $k^+_\mathrm{S,ref}$, is inherently always identified using eq. (3) as can be seen from the constant and near ideal course. The difference between their values obtained by the reparameterized and the standard expression is particularly striking. For a detailed discussion of the results for the eq. (2) parameterization, the reader is referred to appendix SI D.
These results suggest to use the parameterization according to eq. (3) and show that this way, robust discovery of activation energies is feasible over experimentally relevant temperature ranges.

5.5. Coverage dependency

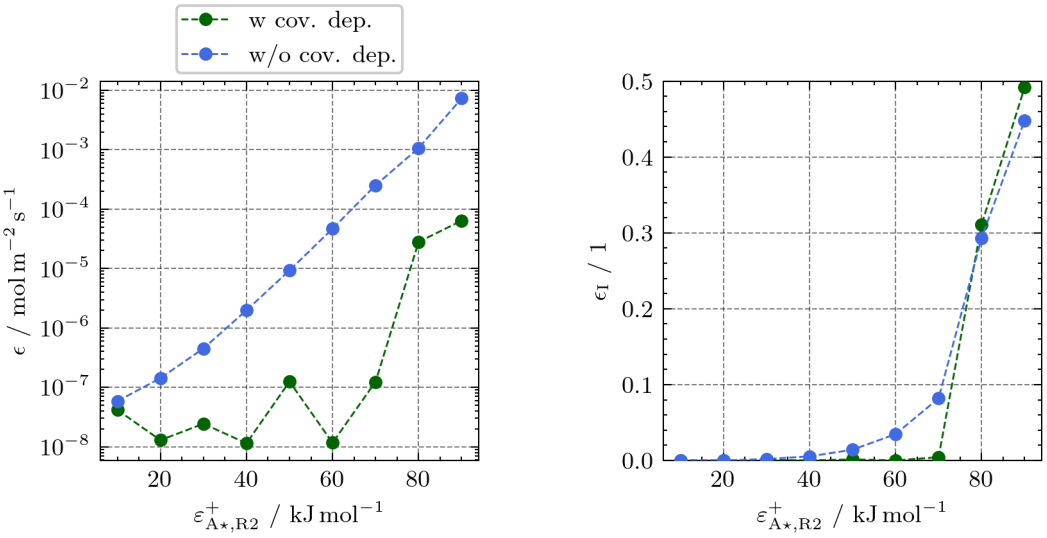
As discussed above, the presence of surface intermediates is known to have an influence on sorption reactions. In the following it is briefly demonstrated that coverage dependency can be depicted with the proposed network architecture, yet detailed analysis is omitted for brevity since superimposing effects of stoichiometry, reversibility and temperature hinder the intuitive explanation of surface coverage dependency. Moving on, the parameter $\varepsilon^+_\mathrm{A\star,R2}$ (change in activation energy of reaction R2 with increasing coverage of A$\star$) is varied. The generated mechanism is investigated with a hCRNN containing *no* coverage dependency layer (w/o cov. dep.) and one that does contain the coverage dependency layer (w cov. dep.). As can be seen from fig. 11b, integer loss ϵ_I is nearly identical and small for both models and thus promises stoichiometric discovery for low values of $\varepsilon^+_\mathrm{A\star,R2}$ regardless of coverage dependency. However, overall loss ϵ in fig. 11a is significantly lower for the coverage dependent hCRNN indicating that the inclusion of a coverage dependency layer is beneficial, if the observed mechanism contains even just one coverage dependent reaction. The hCRNN-predicted reaction rate constants can be calculated at arbitrary reference conditions for comparison as follows:

$$k^-_\mathrm{1,ref} = k^-_1\left(T^\mathrm{ref} = 550\,\mathrm{K}, \boldsymbol{\Theta}^\mathrm{ref}\right) \tag{19}$$

Again, a reference temperature of $T^\mathrm{ref} = 550\,\mathrm{K}$ is chosen. Additionally, two different reference coverages are defined as $\Theta^\mathrm{ref}_\mathrm{A\star,0} = 0$ and $\Theta^\mathrm{ref}_\mathrm{A\star,1} = 1$ respectively. Tabulating the hCRNN-predicted reaction rate constants at different fractional coverages of A$\star$ (compare fig. 12) reveals that the hCRNN with coverage dependence layer does indeed recover the correct kinetic dependence (for full tabulation see appendix SI E) for describing reaction rates at full and zero coverage. In comparison, excluding it from the hCRNN yields rate predictions not representative of the coverage state of the surface and wrong by multiple orders of magnitude.

6. Conclusion

This works offers an extension of the CRNN methodology to heterogeneous reaction networks including reversible and coverage dependent reactions called hCRNN. The demonstrated construction offers clearly



(a) (b)

Figure 11: Loss, ϵ, (a) and integer loss, ϵ_{I}, (b) for a hCRNN with *no* coverage dependency layer (w/o cov. dep.) and one with a coverage dependency layer (w cov. dep). Temperature is sampled uniformly from $[500\,\mathrm{K}, 600\,\mathrm{K}]$.

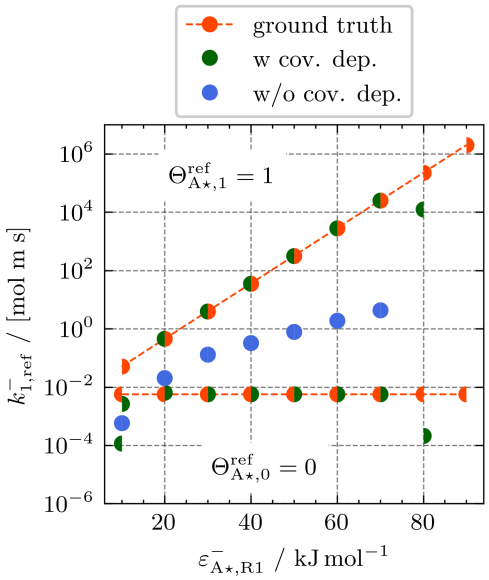


Figure 12: Rate constant $k^{-}_{T^{\mathrm{ref}}}$ at reference temperature $T^{\mathrm{ref}} = 550\,\mathrm{K}$ and coverages $\Theta = 0$ and $\Theta = 1$ comparing a hCRNN with *no* coverage dependency layer (w/o cov. dep.) and one with a coverage dependency layer (w cov. dep.).

separated and interpretable network layers directly corresponding to terms in the general kinetic expression. As a consequence, fitting arbitrarily complex observed kinetic behavior is possible while at the same time ensuring that the recovered kinetic obeys fundamental physicochemical principles of chemical kinetics. This contrasts the hCRNN against black box models and allows for the discovery of kinetic parameters. Since the hCRNN works the same way as a conventional kinetic expression, usage in common chemical engineering applications, i.e, reactor scale simulations is trivially accessible. The fully trained model can function as a drop-in replacement for conventional kinetic expressions:

$$\dot{s}_i = \mathrm{hCRNN}\left(c_i, T\right) \tag{20}$$

The above discussed strategy for setting up a training workflow can be used to gradually increase model complexity and maximize training process. Although the demonstration of the introduced method for a generic yet universal heterogeneous mechanism suggests the potential for extension to all elementary step reaction mechanisms, it is evident that interpretation of results for complex mechanisms requires a detailed analysis. Despite the promise of easy interpretation of CRNNs, this work also shows that discovery for complex cases is not trivial and still requires substantial domain knowledge.

Nonetheless, it was demonstrated that kinetic discovery for heterogeneous mechanisms is possible and can yield good approximations for pre-exponential constants, activation energies and even coverage dependency parameters. The so discovered kinetic model obeys thermodynamic and kinetic restrictions without any prior knowledge of the mechanism. In it's current state, the hCRNN only fulfills these restrictions by training, i.e., thermodynamically and kinetically consistent data must be supplied during training. For example, the conservation of mass is not enforced by the network construction but implicitly satisfied by observation of mass conserving processes. The conservation of mass for the retained mechanism can be checked after the training, but requires the stoichiometric compositions of the observed species. Similar reasoning applies to thermodynamic consistency: Although reversibility is assured to be captured when observed, the correctness of equilibrium constants depends on the presence of correct equilibria in the training data.

From this, it is reasonable to assume that training data plays a critical role in interpreting the results of the method. This work is only concerned with the maximally transient regime because it is conceptually interesting. However, the authors acknowledge that this type of data is difficult to access experimentally for a few reasons: Firstly, extremely fast, yet still precise analytics would be needed to correctly resolve the formation and consumption dynamics in maximally transient regime. Furthermore, since the only way to reliably load a catalytic surface is by bringing it into contact with a gas phase and advancing it towards or even into chemical equilibrium, the start-up of such a reactor would be tedious and skewed towards coverages close to equilibrium. Lastly, even if analytics and experimental setup were ideal, instrumentation is oftentimes not: switching times of valves and nonideal residence time distributions in pipes make highly transient kinetic measurements experimentally hard to realize. Nonetheless, the described sampling method was specifically designed to elucidate the dynamic processes far from equilibrium. Indeed, we can conclude that the fastest (dominating) kinetic behavior is well recovered under all circumstances by the hCRNN in the chosen maximally transient regime. If the mechanism contains very fast steps, these will be identified in favor of slower ones all the while dynamic behavior is well retained. The number of reactions needed for describing the kinetic behavior depends on their speed (processes that dominate the overall kinetics are captured while additional slow ones may not be correctly identified and consequently omitted from the construction). Thus, best kinetic discovery is possible, if all reactions are equally fast.

This in terms means that slow processes are not observable through the chosen sampling strategy. In contrast, in a steady-state experiment, the slowest step limits productivity and dictates overall reactor behavior. Thus, in a typical steady-state catalytic experiment, only the slowest reaction can be observed, while all the others are (assumed to be) in quasi-equilibrium. This can be formalized with the concept of the RDS. In that regard, the observability of fast reactions in the maximally transient regime is the counterpart of the observability of rate determining reactions in a steady-state experiment.

Full kinetic discovery by hCRNNs might consequently be achievable by observing different regimes: While this work demonstrates that the use of maximally transient data favors identification of fast reactions, theory promises the discovery of slow reactions from steady-state data. Whether or not a well-chosen sampling strategy is capable of exposing both fast *and* slow reactions to the hCRNN remains unclear up to now, and

answering this question is left as part for future work. Such a strategy might be based on observing varying time horizons of a reaction system evolving from its initial state into steady-state or equilibrium. This might mean observing the reaction system at different residence times (for continuous systems) or physical times (for transient systems) in the systematic way detailed throughout this work. This kind of experimentation, in contrast to maximally transient experiments, is achievable by, e.g., the periodic transient kinetics method (PTK) and has been demonstrated before [36]. Ultimately, understanding the behavior of hCRNNs in the different rate regimes might allow for the design of a training strategy tailored to the reaction system at hand and consequently enable kinetic discovery even for heterogeneous systems.

Author contributions: CRediT

Hannes Stagge: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Data curation, Visualization, Software, Writing original draft. **Robert Güttel:** Conceptualization, Writing review & editing, Ressources, Supervision.

Supporting Information

Supporting information is available and contains details regarding the mechanisms used in this study, further mathematical analysis and detailed training results (PDF).

Data Availability

The data supporting the findings and conclusions of this study are available within the article. Extra data is available from the corresponding author upon request.

Funding Sources

The work was financially supported by the German Federal Environmental Foundation DBU (FKZ 20022/023), which had no involvement in this study's design, collection, analysis and interpretation of data, writing of the report and decision to submit the article for publication.

Conflicts of Interest

There are no conflicts to declare.

References

[1] A. Parastaev, V. Muravev, E. Huertas Osta, A. J. F. van Hoof, T. F. Kimpel, N. Kosinov, E. J. M. Hensen, Boosting CO_2 hydrogenation via size-dependent metal–support interactions in cobalt/ceria-based catalysts, Nat. Catal. 3 (2020) 526–533. doi:10.1038/s41929-020-0459-4.

[2] R. J. Kee, M. E. Coltrin, P. Glarborg, H. Zhu, Chemically Reacting Flow: Theory, Modeling, and Simulation, 1 ed., Wiley, 2017. doi:10.1002/9781119186304.

[3] K. J. Laidler, A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996), Pure and Applied Chemistry 68 (1996) 149–192. doi:10.1351/pac199668010149.

[4] P. J. Barrie, The mathematical origins of the kinetic compensation effect: 1. the effect of random experimental errors, Phys. Chem. Chem. Phys. 14 (2011) 318–326. doi:10.1039/C1CP22666E.

[5] M. Schwaab, J. C. Pinto, Optimum reparameterization of power function models, Chem. Eng. Sci. 63 (2008) 4631–4635. doi:10.1016/j.ces.2008.07.005.

[6] A. K. Agarwal, M. L. Brisk, Sequential experimental design for precise parameter estimation. 1. Use of reparameterization, Ind. Eng. Chem. Proc. Des. Dev. 24 (1985) 203–207. doi:10.1021/i200028a034.

[7] B. Kreitz, G. D. Wehinger, C. F. Goldsmith, T. Turek, Microkinetic Modeling of the Transient CO_2 Methanation with DFT-Based Uncertainties in a Berty Reactor, ChemCatChem 14 (2022) e202200570. doi:10.1002/cctc.202200570.

[8] M. A. Nolen, C. A. Farberow, S. Kwon, Incorporating Coverage-Dependent Reaction Barriers into First-Principles-Based Microkinetic Models: Approaches and Challenges, ACS Catal. 14 (2024) 14206–14218. doi:10.1021/acscatal.4c03358.

[9] K. Christmann, O. Schober, G. Ertl, Adsorption of CO on a Ni(111) surface, J. Chem. Phys. 60 (1974) 4719–4724. doi:10.1063/1.1680972.
[10] D. Schmider, L. Maier, O. Deutschmann, Reaction Kinetics of CO and CO_2 Methanation over Nickel, Ind. Eng. Chem. Res. 60 (2021) 5792–5805. doi:10.1021/acs.iecr.1c00389.
[11] J. E. Sutton, D. G. Vlachos, Building large microkinetic models with first-principles accuracy at reduced computational cost, Chem. Eng. Sci. 121 (2015) 190–199. doi:10.1016/j.ces.2014.09.011.
[12] A. H. Motagamwala, J. A. Dumesic, Microkinetic Modeling: A Tool for Rational Catalyst Design, Chem. Rev. 121 (2021) 1049–1076. doi:10.1021/acs.chemrev.0c00394.
[13] K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter, J.-D. Grunwaldt, Future Challenges in Heterogeneous Catalysis: Understanding Catalysts under Dynamic Reaction Conditions, ChemCatChem 9 (2017) 17–29. doi:10.1002/cctc.201600996.
[14] D. Meyer, J. Friedland, J. Schumacher, M. G. Gäßler, R. Güttel, Hydrogenation of CO/CO_2 mixtures under unsteady-state conditions: Effect of the carbon oxides on the dynamic methanation process, Chem. Eng. Sci. 250 (2022) 117405. doi:10.1016/j.ces.2021.117405.
[15] M. Langer, D. Kellermann, H. Freund, Kinetic modeling of dynamically operated heterogeneously catalyzed reactions: Microkinetic model reduction and semi-mechanistic approach on the example of the CO_2 methanation, Chem. Eng. J. 467 (2023) 143217. doi:10.1016/j.cej.2023.143217.
[16] M. Jaraíz, J. E. Rubio, L. Enríquez, R. Pinacho, J. L. López-Pérez, A. Lesarri, An Efficient Microkinetic Modeling Protocol: Start with Only the Dominant Mechanisms, Adjust All Parameters, and Build the Complete Model Incrementally, ACS Catal. 9 (2019) 4804–4809. doi:10.1021/acscatal.9b00522.
[17] R. Chacko, K. Keller, S. Tischer, A. B. Shirsath, P. Lott, S. Angeli, O. Deutschmann, Automating the Optimization of Catalytic Reaction Mechanism Parameters Using Basin-Hopping: A Proof of Concept, J. Phys. Chem. C 127 (2023) 7628–7639. doi:10.1021/acs.jpcc.2c08179.
[18] B. Kreitz, K. Sargsyan, K. Blöndal, E. J. Mazeau, R. H. West, G. D. Wehinger, T. Turek, C. F. Goldsmith, Quantifying the Impact of Parametric Uncertainty on Automatic Mechanism Generation for CO_2 Hydrogenation on Ni(111), JACS Au 1 (2021) 1656–1673. doi:10.1021/jacsau.1c00276.
[19] F. A. Döppel, M. Votsmeier, Efficient neural network models of chemical kinetics using a latent asinh rate transformation, React. Chem. Eng. 8 (2023) 2620–2631. doi:10.1039/D3RE00212H.
[20] A. Fedorov, A. Perechodjuk, D. Linke, Kinetics-constrained neural ordinary differential equations: Artificial neural network models tailored for small data to boost kinetic model development, Chem. Eng. J. 477 (2023) 146869. doi:10.1016/j.cej.2023.146869.
[21] T. Kircher, F. A. Döppel, M. Votsmeier, Global reaction neural networks with embedded stoichiometry and thermodynamics for learning kinetics from reactor data, Chem. Eng. J. 485 (2024) 149863. doi:10.1016/j.cej.2024.149863.
[22] B. L. d. O. Campos, A. O. S. da Costa, K. H. Delgado, S. Pitter, J. Sauer, E. F. d. C. Junior, Development of a surrogate artificial neural network for microkinetic modeling: Case study with methanol synthesis, React. Chem. Eng. 9 (2024) 1047–1060. doi:10.1039/D3RE00409K.
[23] W. Ji, S. Deng, Autonomous Discovery of Unknown Reaction Pathways from Data by Chemical Reaction Neural Network, J. Phys. Chem. A 125 (2021) 1082–1092. doi:10.1021/acs.jpca.0c09316.
[24] W. Ji, F. Richter, M. J. Gollner, S. Deng, Autonomous kinetic modeling of biomass pyrolysis using chemical reaction neural networks, Combust. Flame 240 (2022) 111992. doi:10.1016/j.combustflame.2022.111992.
[25] Y. Xu, Q. Chu, X. Chang, H. Wang, S. Wang, S. Xu, D. Chen, Thermal decomposition mechanism of 1,3,5-trinitroperhydro-1,3,5-triazine: Experiments and reaction kinetic modeling, Chem. Eng. Sci. 282 (2023) 119234. doi:10.1016/j.ces.2023.119234.
[26] W. Ji, W. Qiu, Z. Shi, S. Pan, S. Deng, Stiff-PINN: Physics-Informed Neural Network for Stiff Chemical Kinetics, The journal of physical chemistry. A 125 (2021) 8098–8106. doi:10.1021/acs.jpca.1c05102.
[27] F. Döppel, M. Votsmeier, Robust Mechanism Discovery with Atom Conserving Chemical Reaction Neural Networks, 2023. doi:10.26434/chemrxiv-2023-1r389.
[28] Q. Li, H. Chen, B. C. Koenig, S. Deng, Bayesian chemical reaction neural network for autonomous kinetic uncertainty quantification, Phys. Chem. Chem. Phys. 25 (2023) 3707–3717. doi:10.1039/D2CP05083H.
[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An Imperative Style, High-Performance Deep Learning Library, 2019. doi:10.48550/arXiv.1912.01703. arXiv:1912.01703.
[30] S. Barwey, V. Raman, A Neural Network Inspired Formulation of Chemical Kinetics, Energies 14 (2021) 2710. doi:10.3390/en14092710. arXiv:2008.08483.
[31] E. Komp, N. Janulaitis, S. Valleau, Progress towards machine learning reaction rate constants, Phys. Chem. Chem. Phys. 24 (2022) 2692–2705. doi:10.1039/D1CP04422B.
[32] M. Bracconi, M. Maestri, Training set design for machine learning techniques applied to the approximation of computationally intensive first-principles kinetic models, Chem. Eng. J. 400 (2020) 125469. doi:10.1016/j.cej.2020.125469.
[33] M. O. Vlad, M. Tsuchiya, P. Oefner, J. Ross, Bayesian analysis of systems with random chemical composition: Renormalization-group approach to Dirichlet distributions and the statistical theory of dilution, Phys. Rev. E 65 (2001) 011112. doi:10.1103/PhysRevE.65.011112.
[34] Q. Wang, Y. Ma, K. Zhao, Y. Tian, A Comprehensive Survey of Loss Functions in Machine Learning, Ann. Data. Sci. 9 (2022) 187–212. doi:10.1007/s40745-020-00253-5.
[35] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2017. doi:10.48550/arXiv.1412.6980. arXiv:1412.6980.

[36] D. Meyer, J. Friedland, J. Schumacher, R. Güttel, The periodic transient kinetics method for investigation of kinetic process dynamics under realistic conditions: Methanation as an example, Chem. Eng. Res. Des. 173 (2021) 253–266. doi:10.1016/j.cherd.2021.07.011.