A deep neural network for oxidative coupling of methane trained on high-throughput experimental data

Klea Ziu 1, Ruben Solozabal 1, Srinivas Rangarajan 2∗ and Martin Takáč 1

1 Mohamed bin Zayed University of Artificial Intelligence, Masdar City, United Arab Emirates
2 Chemical and Biomolecular Engineering Department, Lehigh University, Bethlehem, PA 18015, United States of America
∗ Author to whom any correspondence should be addressed.

Keywords: deep neural networks, oxidative coupling of methane, high-throughput experiments, kinetic model, data-driven catalyst design, residual neural network

Abstract

In this work, we develop a deep neural network model for the reaction rate of oxidative coupling of methane from published high-throughput experimental catalysis data. A neural network is formulated so that the rate model satisfies the plug flow reactor design equation. The model is then employed to understand the variation of reactant and product composition within the reactor for the reference catalyst Mn–Na₂WO₄/SiO₂ at different temperatures and to identify new catalysts and combinations of known catalysts that would increase yield and selectivity relative to the reference catalyst. The model revealed that methane is converted in the first half of the catalyst bed, while the second part largely consolidates the products (i.e. increases ethylene to ethane ratio). A screening study of ≥3400 combinations of pairs of previously studied catalysts of the form M₁(M₂)₁−₂M₃Oₓ/support (where M₁, M₂ and M₃ are metals) revealed that a reactor configuration comprising two sequential catalyst beds leads to synergistic effects resulting in increased yield of C₂ compared to the reference catalyst at identical conditions and contact time. Finally, an expanded screening study of 7400 combinations (comprising previously studied metals but with several new permutations) revealed multiple catalyst choices with enhanced yields of C₂ products. This study demonstrates the value of learning a deep neural network model for the instantaneous reaction rate directly from high-throughput data and represents a first step in constraining a data-driven reaction model to satisfy domain information.

1. Introduction

Recent decades have witnessed a burgeoning growth in catalysis data, enabling the application of ‘catalysis informatics’, viz. using tools from data science and machine learning to infer the underlying reaction mechanism, making quick predictions of conversion and selectivity under new processing conditions or catalyst compositions, and ultimately proposing new catalytic materials [1–3]. Several recent works show that data sets comprising thousands of combinations of catalysts and processing conditions, either taken from high-throughput experimentation [4, 5] or via data collation from the literature [6, 7], can be used to build machine learned models that can then be interrogated to find novel combinations of catalysts [8–11]. However, most of these models compute the overall conversion, yield or selectivity. They do not take into account domain constraints and may not provide any insight into the chemistry or the variations along the reactor. In a plug (or a packed bed) flow reactor, the overall conversion or selectivity depends on how the rate (and therefore the flow of the reactants and products) changes along the reactor. In principle, experimental data comprising only the overall conversion and selectivity at the end of the reactor can still be employed to learn a data-driven model of the instantaneous reaction rate at any point in the reactor.; However, the learning process must take into account that the net conversion depends on the integration of the instantaneous rate along the reactor (catalyst bed).
In this work, we develop a deep neural network model of the reaction rate, for the first time, using high-throughput experimental data. Our formulation ensures that the reaction rate model satisfies the reactor design equation. As a result, we posit that this model is more robust to extrapolations. Therefore, we apply this model to (a) understand how the composition of the reactants and products changes along the reactor and, thereby, the potential reaction network connections, and (b) predict yields for new catalysts (or catalyst combinations) and conditions. We specifically apply this approach to the problem of oxidative coupling of methane (OCM), which is a reaction to convert methane, prevalent in natural gas, to two valuable products, viz. ethane (C$_2$H$_6$) and ethylene (C$_2$H$_4$) [12–16]. A challenge in this chemistry is that non-selective reactions can lead to the production of carbon monoxide (CO) and carbon dioxide (CO$_2$). Therefore, developing catalysts that are active and selective in C$_2$ products over C$_1$ oxides is key to deploying this technology. Developing a data-driven model for the reaction rate from high-throughput data is a first step to this end.

2. Background and related work

2.1. Reactor design equation

High-throughput experimental studies typically collect the overall conversion of the reactants and selectivities of the products for a range of catalysts (often belonging to a similar class of materials) under different processing conditions (temperature, inlet composition and pressure). Since only the inlet and the effluent are measured, data-driven models are typically trained to predict the effluent composition (from which conversion and selectivities are obtained) given the inlet conditions and the catalyst information. In our previous work [17], we developed a neural network model for this chemistry based on a publicly available high-throughput experimentation data set [4]. The model was formulated to ensure that the atom balance constraints of a plausible reaction network were satisfied, i.e. it was informed by chemistry. The model allows us to compute apparent activation barriers and orders, which in turn indicate that they are consistent with both the Langmuir Hinshelwood model and a Mars van Krevelen model, possibly because of a more complex underlying mechanism. Our goal here is to leverage this experimental data to learn, instead, the local rate of the reaction (and thereby the composition) at any point within the reactor. To do this, we begin by invoking a simple plug flow reactor model that can describe the reactor set up in these experiments. In particular, we show that,

$$\frac{dF_i}{d\tau} = r_i,$$

where $F_i$, $\Omega$ and $r_i$ are the molar flow (moles per unit time) of species $i$, the total moles of catalytic sites distributed uniformly (both radially and axially) within the reactor, and the rate of production of species $i$ (positive for production, negative for consumption, with units of moles produced per mole of catalyst sites per unit time). This equation can be rewritten as,

$$\frac{dF_i}{d\tau} = r_i F_0,$$

where $F_0$ is the initial molar flow rate of the reactant (a reference, such as methane in this case) and $\tau = \frac{\Omega}{F_0}$ is the contact time (CT) of the reactor (roughly representing the average time the reactant molecule spends in the reactor passing through the catalyst bed, also known as the residence time). Therefore, the outlet flow rate of species $i$, $F_{\text{out},i}$, is given by,

$$F_{\text{out},i} = F_{\text{in},i} + \int_0^{\text{CT}} r_i F_0 d\tau,$$

where $\tau = \text{CT}$ is the total contact (or residence) time and $F_{\text{in},i}$ is the molar inlet flow rate of species $i$.

Since a plug flow (or tubular packed bed) reactor can be written as an infinite sum of differential continuous stirred tank reactors, to a first approximation, we can rewrite the equation above as a discrete summation:

$$F_{\text{out},i} = F_{\text{in},i} + \sum_{j=1}^{N} \bar{r}_{i,j} \Delta \tau,$$

where $\bar{r}_{i,j}$ is the average value of $r_i F_0$ in the $j$th element and $\Delta \tau = \frac{\text{CT}}{N}$. Effectively, we seek to learn $\bar{r}_{i,j} \Delta \tau$ for a discrete element $j$ from the high-throughput data (that contain $F_{\text{in},i}, F_{\text{out},i}$ information for various catalysts and processing conditions) so that equation (4) is satisfied.
2.2. Using deep neural networks for differential equations

Deep neural networks have never been reported for handling differential equations in the context of catalysis. Therefore, in this section, we briefly discuss the application of deep learning in systems governed by differential equations in general.

Partial differential equations (PDEs) have been extensively used to model the underlying behavior of complex dynamical systems as well as their adaptability. Several studies [18–21] have employed deep learning to learn the governing differential equations from data or to solve a set of known differential equations. In [18, 22], to approximate the unknown nonlinear responses of diffusion and convection processes, the authors introduced PDE-Net, which was influenced by Wavelet theory. PDE-Net [18, 22] is a deep feed-forward network that operates on the principle of learning differential operators by learning convolution kernels (filters) and then using neural networks to approximate unknown nonlinear responses. The PDE-Net structure, made of δt blocks, resembles residual blocks of ResNet [23].

FD-Net [19] is a finite difference inspired network that uses only a few trainable parameters to learn the underlying governing PDEs from trajectory data and iteratively estimate future dynamical behavior. FD-Net is made up of multiple FD-Blocks similar to residual blocks introduced in ResNet [23] that are stacked sequentially in order to produce an estimated solution of the PDE at t + δt given a solution at t. For a constant input/output shape, FD-Net uses the same number of FD-Filters across all convolutional layers. In contrast to our approach, FD-Net establishes the parameters of each layer without bias terms, as well as the outputs of the layers without using nonlinear activation functions, in order to capture the behavior of linear equations.

The hybrid NN-PDE [24] model demonstrates how neural networks can be used to learn partial PDE solver completion in the setting of reactive flows of laminar flames. When compared to a purely data-driven method, the incomplete PDE description assists the neural network model to recover the target simulation with much higher accuracy. The authors employ a fully convolutional neural network model consisting of ResBlocks with skip connections, analogous to residual blocks [23], to overcome the problem of vanishing gradients.

DynNet [25] is a network motivated by implicit equation of motion solvers in a recurrent cell for full response prediction of nonlinear multi degrees of freedom systems. The architecture of the model is based on ResNet [23], and the residual block is the only component of the network capable of learning the nonlinear behavior of the dynamic system.

All of the above-mentioned methods [18, 19, 22, 24, 25] use different approaches to learn PDEs from data, but all of them, including our approach, are based on residual blocks. The main reason for this is that ResNet [23] outperforms other architectures in learning differential equations from data due to its inherited similarity to the Euler method. The key difference of our work is exploiting such efforts in the field of catalysis using high-throughput experimentation data. Furthermore, as we discuss below, the ResNet architecture naturally allows for representing the summation in equation (4).

3. Methods

3.1. Deep neural network to capture the local rate within the reactor

A deep neural network resembling the reactor model (equation (4)) was developed, as shown in figure 1. The input layer consists of seven neurons, which represent the flows of seven species in this reaction system: (a) the inert carrier gas argon (F_{N_2}), (2) the main reactants, viz. methane (F_{CH_4}) and oxygen (F_{O_2}), (c) the desired C_2 products, namely, ethane (F_{C_2H_6}) and ethylene (F_{C_2H_4}), and (d) the chief C_1 byproducts, viz. carbon monoxide (F_{CO}) and carbon dioxide (F_{CO_2}). These seven neurons represent the values of the flows before entering the reactor at η = 0, while the output values indicate the number of flows leaving the reactor after a given CT that species spend inside the reactor. The inlet values of product flows, i.e. F_{C_2H_2}, F_{C_2H_6}, F_{CO} and F_{CO_2}, are respectively set to zero at η = 0. Apart from the input features, we introduce global features that are temperature T(°C) and catalyst descriptors (see next subsection). The input flow is concatenated with global features and fed to the module of the deep neural network, as shown in figure 1. Our model consists of modules that are identical deep neural network blocks, arranged sequentially, so that the output of one module becomes the input of a successive module (akin to the summation in equation (4) or the physical inference of approximating a tubular flow reactor as a series of continuous stirred tank reactors). The modules are identical, as each subsequent module has the same weights as before. Each module corresponds to a Δτ timestep. We employ identity mapping x in our model, to build a deeper neural network to avoid vanishing gradients while training the model. The identity mapping does not have any parameter, while the function F(x) is the residual mapping that the model should learn. The residual is the amount by which the model’s prediction must be adjusted to match the actual result. The combination of input x and F(x) as y = F(x) + x is given as input to the successive module. Using this architecture for the model helps to save the history of how the flow of the given species changes inside the model for a given timestep (check
Figure 1. Architecture of the OCM model. □ denotes the input of our model. It consists of seven input flows $F_{Ar}$, $F_{CH_4}$, $F_{O_2}$, $F_{C_2H_4}$, $F_{C_2H_6}$, $F_{CO}$ and $F_{CO_2}$. ■ represents the global features that are temperature $T(\degree)$ and catalyst descriptors ($M1-(M2,-M3O_x/support)$). Input flows are concatenated (||) with global features and fed to the module □ of the deep neural network. Function $F(x)$ together with the input $x$ (□) are given as input for the next module as $y = F(x) + x$. At $\delta t$ the output of the first module, together with the global features are fed to the second module. This happens recursively for a given $\delta t$ that is proportional to the amount of time the products spend inside the reactor. □ are the flows predicted by our model at time-step $T$. 
supplementary information for more details). To interpret this architecture physically, one can consider each block to represent a discrete element of the numerical integration in equation (4) or that each block is effectively a local, smaller, stirred tank reactor model.

3.1.1. Module
The module consists of an embedding layer, where input flows (7 neurons) are concatenated with global features (64 neurons) out of a total of 71 neurons that are fed forward to the first hidden linear layer. The module has four hidden linear layers, each made up of 64 neurons, chosen after extensive evaluation of different structures. Scaled exponential linear unit (SELU) [26] is used as an activation function, which is given by,

\[
\text{selu}(z) = \lambda z, \quad \text{if } z \geq 0,
\]

\[
\text{selu}(z) = \lambda \alpha (\exp(z) - 1), \quad \text{if } z < 0,
\]

where \(\lambda \approx 1.05070\) and \(\alpha \approx 1.673263\).

An Adam optimizer with a default learning rate of 0.01 is used to optimize the model. The data set is split randomly on the training set (80%) and the testing set (20%). The training is done on batches with size 64. We use mean squared error (MSE) to evaluate the loss of the model, which is given by,

\[
L = \frac{1}{n} \sum_{i=1}^{n} \left( (F_{\text{Ar}}^{i} - \hat{F}_{\text{Ar}}^{i})^2 + (F_{\text{CH}_4}^{i} - \hat{F}_{\text{CH}_4}^{i})^2 + (F_{\text{O}_2}^{i} - \hat{F}_{\text{O}_2}^{i})^2 + (F_{\text{C}_2\text{H}_6}^{i} - \hat{F}_{\text{C}_2\text{H}_6}^{i})^2 + (F_{\text{CO}}^{i} - \hat{F}_{\text{CO}}^{i})^2 + (F_{\text{CO}_2}^{i} - \hat{F}_{\text{CO}_2}^{i})^2 \right),
\]

where \(n\) denotes the number of data points used to train the model, \(F\) denotes the flows of the seven species pertaining to this chemistry (Ar, \(\text{CH}_4\), \(\text{O}_2\), \(\text{C}_2\text{H}_6\), \(\text{C}_2\text{H}_4\), \(\text{CO}\) and \(\text{CO}_2\)) in moles of the molecule per unit time and \(\hat{F}\) are the predicted flows from the model.

3.1.2. Flows
The data set contains the inlet flow rates of methane \((F_{\text{CH}_4,0})\), oxygen \((F_{\text{O}_2,0})\) and argon \((F_{\text{Ar},0})\) as well as the yields of ethane \((y_{\text{C}_2\text{H}_6})\), ethylene \((y_{\text{C}_2\text{H}_4})\), carbon monoxide \((y_{\text{CO}})\) and carbon dioxide \((y_{\text{CO}_2})\). The yields of the carbon-containing products are defined as the percentage of carbon atoms of methane converted to that molecule (resulting in \(19\)). We assume that the OCM chemistry is captured by the following overall reactions, which are sufficient to carry out atom balance at any point:

\[
2\text{CH}_4 + 0.5\text{O}_2 \rightarrow \text{C}_2\text{H}_6 + \text{H}_2\text{O},
\]

\[
\text{C}_2\text{H}_6 + 0.5\text{O}_2 \rightarrow \text{C}_2\text{H}_4 + \text{H}_2\text{O},
\]

\[
\text{CH}_4 + 1.5\text{O}_2 \rightarrow \text{CO} + 2\text{H}_2\text{O},
\]

\[
\text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O}.
\]

At a given time-step in the model, the flow rate of the different species is computed based on basic carbon and oxygen balance emanating from these reactions:

\[
F_{\text{Ar}} = F_{\text{Ar},0},
\]

\[
F_{\text{CH}_4} = F_{\text{CH}_4,0}(1 - (y_{\text{C}_2\text{H}_6} + y_{\text{C}_2\text{H}_4} + y_{\text{CO}} + y_{\text{CO}_2})),
\]

\[
F_{\text{O}_2} = F_{\text{O}_2,0} - F_{\text{CH}_4,0}[0.25(y_{\text{C}_2\text{H}_6}) + 0.25(y_{\text{C}_2\text{H}_4}) + 1.5(y_{\text{CO}}) + 2(y_{\text{CO}_2})],
\]

\[
F_{\text{C}_2\text{H}_6} = 0.5F_{\text{CH}_4,0} \times (y_{\text{C}_2\text{H}_6}),
\]

\[
F_{\text{C}_2\text{H}_4} = 0.5F_{\text{CH}_4,0} \times (y_{\text{C}_2\text{H}_4}).
\]
\[ F_{\text{CO}} = F_{\text{CH}_4, \alpha} * (y_{\text{CO}}), \]  
(17)

\[ F_{\text{CO}_2} = F_{\text{CH}_4, \alpha} * (y_{\text{CO}_2}), \]  
(18)

\[ X_{\text{CH}_4} = y_{\text{C}_2\text{H}_6} + y_{\text{C}_2\text{H}_4} + y_{\text{CO}} + y_{\text{CO}_2}, \]  
(19)

where \( X_{\text{CH}_4} \) is the methane conversion.

3.2. Data and descriptors

We used a high-throughput OCM dataset \([4]\), which consists of 12,708 data points encompassing 59 catalysts and support combinations. The catalyst is a supported mixed metal oxide consisting of three metals or fewer and has the general formula \( \text{M1}-(\text{M2})_{1-2}\text{M3O}_x/\text{support} \). \( \text{M1} \) is mainly a variety of transition metals and lanthanide elements. Molybdenum and tungsten are the two possible choices for \( \text{M3} \), and \( \text{M2} \) is any metal that can form a molybdate or tungstate; a variety of commonly used supports are also considered in the catalyst space. Methane conversion (\( X_{\text{CH}_4} \)) and yields (and selectivity) of all main carbon-containing products, viz. ethane (\( \text{C}_2\text{H}_6 \)), ethylene (\( \text{C}_2\text{H}_4 \)), carbon monoxide (\( \text{CO} \)), and carbon dioxide (\( \text{CO}_2 \)) are included in the data set. Through data preprocessing we discovered that not all data points satisfy carbon and oxygen balance. Consequently, we filtered out the data points that had a negatively computed oxygen flow at the outlet. The methane conversion is calculated as the sum of the yields given in the dataset, as shown in equation (19). The remaining data set comprises 9,271 data points, which is sufficient for further data-driven analysis. We represent each data point by descriptors, including catalyst composition and processing conditions. The moles of metals \( \text{M1}, \text{M2} \) and \( \text{M3} \) per unit gram of support describe the catalyst composition, while the support is represented as integer values. Reaction condition descriptors are temperature \( \text{Temp} \) (\( ^\circ\text{C} \)), argon flow (\( F_{\text{Ar}} \)), methane flow (\( F_{\text{CH}_4} \)) and oxygen flow (\( F_{\text{O}_2} \)). All the data points are grouped by the CT of 0.75, 0.50 and 0.38 (seconds) given in the dataset.

4. Results and discussion

4.1. Performance of the neural network model

Mean absolute error (MAE), root MSE (RMSE) and the \( R_2 \) score on the testing data set are shown in table 1. The parity plots are in the supporting information (figure S1). We note that the \( R_2 \) score for all the predicted yields and conversions is higher than 0.85, indicating that our deep neural network model is able to capture the relationship between descriptors sufficiently. The benefit of developing this model is that it can be used to evaluate the reaction kinetics of a catalyst of interest in a given CT, as well as the effect of various material components on the reaction rate. Note that we use four blocks to train data points with \( \text{CT} = 0.38 \) s, five blocks for data points with \( \text{CT} = 0.5 \) s and seven blocks for \( \text{CT} = 0.75 \) s (i.e. each block roughly corresponds to 0.1 s; there is a small discrepancy for the smallest and largest Cts but this does not seem to substantially affect our model performance).

4.2. Understanding intra-reactor variations

We then use this model to analyze one of the most promising OCM catalysts, \( \text{Mn}–\text{Na}_2\text{WO}_4/\text{SiO}_2 \), which has attracted a lot of attention due to its relatively high stability and selectivity for \( \text{C}_2 \) products, viz. ethane (\( \text{C}_2\text{H}_6 \)) and ethylene (\( \text{C}_2\text{H}_4 \)).

Figure 2 shows the variation of conversion of methane and the yields of the different products along the reactor for this reference catalyst at various temperatures (figures S2/S3 show the flows of these species on two different catalysts). Clearly, the majority of the conversion occurs in the first one-third of the reactor. Therefore, the system likely reaches near equilibrium early on. The conversion of methane increases with temperature as expected, while the behavior of the yield curves is not entirely monotonous. For instance, the yield of ethane increases going from 700°C to 800°C. However, it drops significantly upon increasing the temperature to 900°C, indicating that ethane is getting further converted. On the other hand, while the ethylene yield also increases from 700°C to 800°C and then drops at the higher temperature, the reduction is rather small indicating that some of the ethane is undergoing oxidative dehydrogenation to produce ethylene. The oxidation product CO increases substantially with increase in temperature, indicating clearly that methane gets converted into CO directly or indirectly (i.e. via combustion of ethane and ethylene). The formation of carbon dioxide increases from 700°C to 800°C but then drops at higher temperature perhaps because of a decrease in oxygen concentration in the reactor (consumed due to the production of CO and ethylene). The nonlinear behavior of the yields is clearly due to the complicated network of reactions involved in oxidative coupling.
shows the top-performing candidates (vis-a-vis the reference). We consider the effect of varying the metals and the support of the catalyst on methane conversion. A scatter plot case of only having Mn–Na 0.1 of each catalyst (i.e. two neural network blocks, one for each catalyst one after the other). We consider the second catalyst the combined bed (both about 83%). That is, combining a more active catalyst with one that is more selective leads to a better performing catalytic system than using either catalyst would allow. A significant negative effect on methane conversion and C₂ yield for these combinations. The red box encompasses all combinations involving the reference catalyst that can improve the yield compared to the case with only one catalyst. Table 4.3. Catalyst screening The advantage of a neural network model for the rate is that we can use the model to consider the effect of changing the catalyst and operating conditions on the product yield and selectivity. To this end, we consider three case studies. First, considering Mn–Na₂WO₄/SiO₂ as a reference catalyst, we investigated the local sensitivity of the catalyst to changes in the metal (M₁, M₂ and M₃) and the support, which changed one at a time. Note that not all combinations considered in the local sensitivity analysis were experimentally examined. The reaction conditions are 1123.15 K (850°C), 10.5 ml min⁻¹ methane (CH₄) flow rate, 0.75 ml min⁻¹ oxygen (O₂) flow, 10 ml min⁻¹ argon (Ar), and CT = 0.1 (i.e. one neural network block). In this setting, the reference methane conversion X(CH₄) is 6.4% and the C₂ yield is 5.5%, with the rest being oxidation products. Figures 4 and 5 show the effect of varying the metals and the support of the catalyst on methane conversion and C₂ yield, respectively. It can be seen that the reference catalyst is pretty robust to local variations. However, our results indicate that changing M₁ from Mn to Ti will slightly improve methane conversion but will not affect C₂ yield (indicating that the selectivity will be lower). Changing M₂ will have a significant negative effect on methane conversion and C₂ yield, indicating that Na is the local optimal choice. Finally, changing the support to TiO₂ seems to improve both conversion of methane and yield of C₂ products. Second, having looked at the local sensitivities of the reference catalyst, we consider the effect of having two catalyst beds in a reactor to identify potential cases of synergies between catalysts. Only 59 choices considered in the high-throughput data set were considered. Therefore, we looked at about 59 × 59 = 3481 combinations. The reaction conditions are essentially the same as before (except for CT): 1123.15 K (850°C), 10.5 ml min⁻¹ methane (CH₄) flow rate, 0.75 ml min⁻¹ oxygen (O₂) flow, 10 ml min⁻¹ argon (Ar), and CT = 0.1 of each catalyst (i.e. two neural network blocks, one for each catalyst one after the other). We consider the case of only having Mn–Na₂WO₄/SiO₂ catalyst with CT = 0.2 as the reference. Figure 6 shows a scatter plot of the predicted conversion and C₂ yield for these combinations. The red box encompasses all combinations that lead to a higher yield than the reference. Table 4.3.1 shows the top-performing candidates (vis-a-vis the reference at the bottom) in terms of the yield of the C₂ products. Clearly, there are a few combinations involving the reference catalyst that can improve the yield compared to the case with only one catalyst. Indeed, while some of the catalysts individually perform worse than Mn–Na₂WO₄/SiO₂ with regard to C₂ yield (e.g. Mn–Na₂WO₄/SiCnf), they can, in combination with other catalysts, actually lead to improved yields than the reference; this indicates potential reaction engineering optimizations with exiting catalysts. To understand this synergy, we report the C₂ yield and methane conversion for each of the individual catalysts at CT = 0.2. For the first combination shown in table 4.3.1, viz. Mn–Na₂WO₄/TiO₂+Mn–Na₂WO₄/SiCnf, the methane conversion of the first catalyst (12.68%) is higher than that of the second catalyst (11.07%) and the combined system (11.65%). However, its selectivity (i.e. the ratio \( \frac{\sigma}{X(CH_4)} \)) is 75%, which is lower than that of the second catalyst the combined bed (both about 83%). That is, combining a more active catalyst with one that is more selective leads to a better performing catalytic system than using either catalyst would allow. A
Figure 2. (a) Conversion of methane \( X_{\text{CH}_4} \), (b) ethylene yield \( y_{\text{C}_2\text{H}_4} \), (c) ethane yield \( y_{\text{C}_2\text{H}_6} \), (d) carbon monoxide yield \( y_{\text{CO}} \) and (e) carbon dioxide yield \( y_{\text{CO}_2} \) for three different temperatures (700 \(^\circ\)C, 800 \(^\circ\)C, 900 \(^\circ\)C) on the reference catalyst (Mn–Na\(_2\)WO\(_4\)/SiO\(_2\)). Operating conditions for all the above temperatures are 4.8 ml min\(^{-1}\) methane (CH\(_4\)) flow rate, 1.2 ml min\(^{-1}\) oxygen (O\(_2\)) flow and 4 ml min\(^{-1}\) argon (Ar). CT = 0.75 in all cases (seven residual blocks for each temperature).

similar explanation can be provided for the other combinations in table 2). This promising strategy of mixing catalysts to exploit synergies needs further experimental validation. A complete list of the 3481 catalysts considered here is given in the supporting information.

Third, we expand our search for optimal catalysts to compositions outside those considered in the high-throughput data set but use the same space of materials for M1–M3 and the support. Specifically, we consider all permutations of catalysts within our space of metals and supports
Figure 3. Ratio $C_2H_4/C_2$ and $C_2$ selectivity (both in percentage) at temperature 750$^\circ$C for the reference catalyst (Mn–Na$_2$WO$_4$/SiO$_2$). For the given temperature, the operating conditions are 4.8 ml min$^{-1}$ methane (CH$_4$) flow rate, 1.2 ml min$^{-1}$ oxygen (O$_2$) flow and 4 ml min$^{-1}$ argon (Ar). $CT = 0.75$ in all cases (seven residual blocks).

Figure 4. Effect of varying metals M1, M2, M3 and the support individually on the methane conversion. Red line represents the reference point (Mn–Na$_2$WO$_4$/SiO$_2$). Reaction conditions: 1123.15 K (850$^\circ$C), 10.5 ml min$^{-1}$ methane (CH$_4$) flow rate, 0.75 ml min$^{-1}$ oxygen (O$_2$) flow, 10 ml min$^{-1}$ argon (Ar) and $CT = 0.1$ (i.e. one neural network block).
Figure 5. Effect of varying metals M1, M2, M3 and the support individually on the C\textsubscript{2} yield. Red line represents the reference point (Mn–Na\textsubscript{2}WO\textsubscript{4}/SiO\textsubscript{2}), and n.a. means that there is no metal or support.

Figure 6. Scatter plot of predicted conversion and yield of all combinations of the two-catalyst beds using only those catalysts considered in the high-throughput data set. Red box encompasses those catalysts with higher C\textsubscript{2} yield compared to the reference Mn–Na\textsubscript{2}WO\textsubscript{4}/SiO\textsubscript{2} (red point). The operating specs are 1123.15 K (850 \degree C), 10.5 ml min\(^{-1}\) methane (CH\textsubscript{4}) flow rate, 0.75 ml min\(^{-1}\) oxygen (O\textsubscript{2}) flow, 10 ml min\(^{-1}\) argon (Ar) and CT = 0.2 in total for both catalysts (one block for each catalyst).
Table 2. Top-performing catalyst combinations for a two-bed catalyst system (the reference catalyst at the end provides a baseline). The operating conditions: 1123.15 K (850 °C), 10.5 ml min⁻¹ methane (CH₄) flow rate, 0.75 ml min⁻¹ oxygen (O₂) flow, and 10 ml min⁻¹ argon (Ar) and for each catalyst (for a total contact time of 0.2). For easy comparison, the values of the C₂ yield ($y_{C_2}$) and methane conversion ($X_{CH_4}$) for each of the catalysts are reported under the same conditions and $CT = 0.2$.

| Catalyst 1 | Catalyst 2 | $y_{C_2}$ | $X_{CH_4}$ | Catalyst 1 | Catalyst 2 | $y_{C_2}$ | $X_{CH_4}$ |
|------------|------------|----------|----------|------------|------------|----------|----------|
| Mn–Na₂WO₄/TiO₂ | Mn–Na₂WO₄/SiCnf | 9.72 | 11.65 | 9.47 | 12.68 | 9.16 | 11.07 |
| Mn–Na₂WO₄/TiO₂ | Mn–Na₂WO₄/SiC | 9.56 | 12.02 | 9.47 | 12.68 | 8.54 | 10.14 |
| Mn–Na₂WO₄/TiO₂ | Mn–Na₂WO₄/SiO₂ | 9.54 | 12.31 | 9.47 | 12.68 | 8.72 | 11.51 |
| Ti–NaWO₄/SiO₂ | Mn–Na₂WO₄/SiCnf | 9.41 | 11.85 | 9.04 | 11.55 | 9.16 | 11.07 |
| Mn–Na₂WO₄/SiO₂ | Mn–Na₂WO₄/SiO₂ | 8.72 | 11.51 | — | — | — | — |

Figure 7. Scatter plot of the predicted C₂ yield and methane conversion of all possible catalyst combinations but using only those metals and supports considered in the original high-throughput experimentation data set. Reference catalyst, Mn–Na₂WO₄/SiO₂, is marked in red and the red box shown encompasses the space of higher-yielding catalysts. Operating specs: 1123.15 K (850 °C), 10.5 ml min⁻¹ methane (CH₄) flow rate, 0.75 ml min⁻¹ oxygen (O₂) flow, 10 ml min⁻¹ argon (Ar) and $CT = 0.1$.

Table 3. Top-performing catalysts from the expanded single catalyst search (the reference catalyst is listed at the end). Operating conditions: 1123.15 K (850 °C), 10.5 ml min⁻¹ methane (CH₄) flow rate, 0.75 ml min⁻¹ oxygen (O₂) flow, 10 ml min⁻¹ argon (Ar) and $CT = 0.1$ (i.e. one prediction block).

| Catalyst | C₂ yield | $X_{CH_4}$ |
|----------|----------|----------|
| Ti–Na₂WO₄/SiC | 6.18 | 7.23 |
| Ti–KWO₄/SiC | 5.97 | 7.27 |
| Nd–Na₂WO₄/SiC | 5.96 | 6.84 |
| Mn–KWO₄/TiO₂ | 5.95 | 7.47 |
| Ti–KWO₄/TiO₂ | 5.93 | 8.09 |
| Mn–Na₂WO₄/SiO₂ | 5.54 | 6.44 |

($19_{M1} \cdot 10_{M2} \cdot 3_{M3} \cdot 13_{support} = 7410$), keeping the same relative concentration of M1, M2 and M3 as the reference catalyst. We evaluated their predicted conversion and C₂ yield under the same operating conditions as above (but assuming just one catalyst and a $CT = 0.1$). Figure 7 plots the yield and conversion of all these catalyst choices. Clearly, there is substantial scatter in the data; while some catalysts seem to be significantly more active than the reference, their C₂ yields are much lower indicating that these catalysts tend to lead to overoxidation. Few catalysts showed higher C₂ yield than the reference. These are in the region bounded by the red box in figure 7, and the top five are tabulated in table 3. However, it is important to note that all of
these (except Nd–Na$_2$WO$_4$/SiC) have C$_2$ selectivity equal to or lower than the reference catalyst. Nevertheless, these are promising candidates that can be explored further in experiments. Furthermore, some combinations not containing a support or one of the metals showed high activity. However, the reliability of these predictions is unclear. A complete list of the 7410 catalysts considered here is given in the supporting information.

5. Conclusion and future work

Our work differs from previous models in the literature in that we employ high-throughput experimental data to learn a deep neural network model of the reaction rate (as opposed to just the effluent composition) that is further constrained to satisfy the plug flow design equation. Using this model allowed us to understand how the conversion of the reactant and the yield of the products changes along the reactor. This allowed us to determine that, for the reference catalyst (Mn–Na$_2$WO$_4$/SiO$_2$), most of the methane that gets converted in the reactor does so in the first one-third of the reactor. Subsequently, the yield of ethylene increases, while selectivity remains about the same. This indicates that ethylene is formed from the dehydrogenation of ethane (through oxidative conversion). The combustion products can arise from both methane and C$_2$ products. Local and global catalyst screening studies point out that while the reference catalyst is locally near optimum, there are potentially other catalysts and catalyst combinations (comprising two catalyst beds) that can lead to better yields. We posit that since the rate model is learned to satisfy the basic underlying physics of the reactor, (a) it is robust enough to be used to explore catalysts beyond those considered in the experiments but within a larger space for material choices and (b) it can be employed to consider novel catalyst combinations or different processing conditions that were not considered in the experiments. This work, therefore, represents a first approach to developing domain-informed deep learning models of reaction rates from high-throughput experiments. We posit that domain information can further include physical, chemical and thermodynamic constraints that the model must also intrinsically satisfy beyond the reactor design equation.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

S R acknowledges financial support from the National Science Foundation (CBET #2045550).

ORCID iDs

Klea Ziu  https://orcid.org/0000-0001-5809-6816
Ruben Solozabal  https://orcid.org/0000-0001-5523-7492
Srinivas Rangarajan  https://orcid.org/0000-0002-6777-9421
Martin Takač  https://orcid.org/0000-0001-7455-2025

References

[1] Medford A J, Kunz M R, Ewing S M, Borders T and Fushimi R 2018 ACS Catal. 8 7403–29
[2] Toyao T, Maeno Z, Takakusagi S, Kamachi T, Takigawa I and Shimizu K I 2019 ACS Catal. 10 2260–97
[3] Takahashi K et al 2019 ChemCatChem 11 1146–52
[4] Nguyen T et al 2019 ACS Catal. 10 921–32
[5] McCullough K, Williams T, Mingle K, Jamshidi P and Luterbach J 2020 Phys. Chem. Chem. Phys. 22 11174–96
[6] Odaşlı Ç, Güney M E and Yıldırım R 2014 Int. J. Hydrog. Energy 39 5733–46
[7] Zaryalova U, Holena M, Schloegl R and Baerms M 2011 ChemCatChem 3 1935–47
[8] Mine S, Takao M, Yamaguchi T, Toyao T, Maeno Z, Hakim Siddiki S, Takakusagi S, Shimizu K I and Takigawa I 2021 ChemCatChem 13 3636–55
[9] Takahashi K, Miyazato I, Nishimura S and Ohyama J 2018 ChemCatChem 10 3223–8
[10] Nguyen T N, Nakanowatari S, Nhat Tran T P, Thakur A, Takahashi L, Takahashi K and Taniike T 2021 ACS Catal. 11 1797–809
[11] Nakanowatari S, Nguyen T N, Chikuma H, Fujibara A, Seenivasan K, Thakur A, Takahashi L, Takahashi K and Taniike T 2021 ChemCatChem 13 3262–9
[12] Sourav S, Wang Y, Kiani D, Baltrusaitis J, Fushimi R R and Wachs I E 2021 Angew. Chem., Int. Ed.
[13] Kiani D, Sourav S, Baltrusaitis J and Wachs I E 2019 ACS Catal. 9 5912–28
[14] Horn R and Schloegl R 2015 Catal. Lett. 145 23–39
[15] Galadima A and Muraza O 2016 J. Ind. Eng. Chem. 37 1–13
[16] Pak S, Qiu P and Lunsford J H 1998 J. Catal. 179 222–30
[17] Chen K, Tian H, Li B and Rangarajan S 2022 AIChE J. 68 e17584
[18] Long Z, Lu Y, Ma X and Dong B 2018 PDE-net: learning PDEs from data Int. Conf. on Machine Learning (PMLR) pp 3208–16
[19] Gulgec N S, Shi Z, Deshmukh N, Pakzad S and Takač M 2019 FD-net with auxiliary time steps: fast prediction of PDEs using Hessian-free trust-region methods Neurips 2019 Workshop (Beyond First Order Methods in ML)
[20] Sadeghi Eshkevari S, Pakzad S N, Takač M and Matarazzo T J 2020 J. Eng. Mech. 146 04020011
[21] Shi Z, Gulgec N S, Berahas A S, Pakzad S N and Takač M 2020 Finite difference neural networks: fast prediction of partial differential equations 2020 19th IEEE Int. Conf. on Machine Learning and Applications (ICMLA) (IEEE) pp 130–5
[22] Long Z, Lu Y and Dong B 2018 CoRR (arXiv:1812.04426)
[23] He K, Zhang X, Ren S and Sun J 2015 CoRR (arXiv:1512.03385)
[24] Tathawadekar N, Doan N A K, Silva C F and Thuerey N 2021 arXiv:2111.11185
[25] Eshkevari S S, Takač M, Pakzad S N and Jahani M 2021 Eng. Struct. 229 111582
[26] Klambauer G, Unterthiner T, Mayr A and Hochreiter S 2017 Self-normalizing neural networks Proc. 31st Int. Conf. on Neural Information Processing Systems (NIPS’17) (Red Hook, NY: Curran Associates Inc.) pp 972–81