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

Klea Ziu[†], Ruben Solozabal[†], Srinivas Rangarajan[‡], Martin Takáč[†]

† Mohamed bin Zayed University of Artificial Intelligence, Masdar City, UAE
‡ Chemical & Biomolecular Engineering Department, Lehigh University, Bethlehem, PA 18015, USA

E-mail: Srinivas Rangarajan <srr516@lehigh.edu>, Martin Takáč <Takac.MT@gmail.com>

Abstract.
In this work, we develop a deep neural network model of reaction rate of oxidative coupling of methane from published high throughput experimental catalysis data. The 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 $Mn - Na_2WO_4/SiO_2$ at different temperatures and to identify new catalysts and combination 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 $M1(M2)_{1-2}M3Ox/support$ (where M1, M2, and M3 are metals) revealed that a reactor configuration comprising of two sequential catalyst beds lead 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 of previously studied metals but with several new permutations) revealed multiple catalyst choices with enhanced yields of C2 products. This study shows the value of learning a deep neural network model of 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 at new process conditions or catalyst compositions, and ultimately proposing new catalytic materials [1, 2, 3]. Several recent works show that

datasets comprising of thousands of combinations of catalysts and process 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, 9, 10, 11]. Most of these models, however, compute the overall conversion, yield, or selectivity; such models do not take into account domain constraints and may not provide any insights 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; the learning process, however, 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 this model is more robust to extrapolations. Therefore, we apply this model to (1) understand how the composition of the reactants and products changes along the reactor and, thereby, the potential reaction network connections, and (2) 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_2H_6) and ethylene (C_2H_4) [12, 13, 14, 15, 16]. A challenge in this chemistry is that unselective reactions can lead to the production of carbon monoxide (CO) and carbon dioxide (CO_2); therefore, developing catalysts that are active and selective to 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 overall conversion of the reactant and selectivities of the products for a range of catalysts (often belonging to a similar class of materials) at different process 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 publicly available high throughput experimentation dataset [4]; the model was formulated to ensure that the atom balance constraints of a plausible reaction

network was satisfied, i.e. it was informed by chemistry. The model allowed us to compute apparent activation barriers and orders which in turn indicated that they were consistent with both 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 the simple plug flow reactor model that can describe the reactor set up in these experiments. In particular, we can say:

$$\frac{dF_i}{d\Omega} = r_i \tag{1}$$

where F_i, Ω, 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

$$\frac{dF_i}{d\tau} = r_i F_0 \tag{2}$$

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 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, $F_{out,i}$, is given by

$$F_{out,i} = F_{in,i} + \int_0^{CT} r_i F_0 d\tau \tag{3}$$

where $\tau = CT$ is the total contact (or residence) time, and $F_{in,i}$ 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:

$$F_{out,i} = F_{in,i} + \sum_{j=1}^{N} \tilde{r}_{i,j} \Delta\tau \tag{4}$$

where $\tilde{r}_{i,j}$ is the average value of $r_i F_0$ in the jth element and $\Delta\tau = \frac{CT}{N}$. Effectively, we seek to learn $\tilde{r}_{i,j}\Delta\tau$ for a discrete element j from the high throughput data (that contains $F_{in,i}$, $F_{out,i}$ information for various catalysts and process conditions) such 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; in this section, therefore, we briefly discuss the application of deep learning in systems governed by differential equations in general.

Partial differential equations (PDEs) are extensively used for modeling the underlying behavior of complex dynamical systems as well as their adaptability. Several studies [18, 19, 20, 21] have employed deep learning to learn the governing differential equations from data or 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. 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 by multiple FD-Blocks similar to residual blocks introduced in ResNet [23], that are stack sequentially in order to produce an estimated solution of the PDE at $t + \delta t$ given a solution at t. For a constant input/output shape, FD-Net uses the same amount of FD-Filters through all convolutional layers. In contrast from 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 in recovering 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 into a recurrent cell for full response prediction of nonlinear multi degrees of freedom (MDOF) 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 the above mentioned methods [18, 22, 19, 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] outperform other architectures in learning differential equations from data due to their inherited similarity to the Euler's method.

The key difference of our work is exploiting such efforts in the field of catalysis using high throughput experimentation data. Further, 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 the figure 1. The input layer consists of seven neurons which represent the flows of seven species in this reaction system: (1) the inert carrier gas argon (F_{Ar}), (2) the main reactants, viz., methane (F_{CH_4}) and oxygen (F_{O_2}), (3) the desired C_2 products, namely, ethane ($F_{C_2H_4}$) and ethylene ($F_{C_2H_6}$), and (4) the chief C_1 byproducts, viz., carbon monoxide (F_{CO}) and carbon dioxide (F_{CO_2}). These seven neurons represents the values of the flows before entering the reactor at $\tau = 0$, while the output values indicate the amount of flows leaving the reactor after a given contact time (CT) that species spend inside the reactor. The inlet values of products flows, i.e., $F_{C_2H_4}$, $F_{C_2H_6}$, F_{CO} and F_{CO_2} are respectively set to *zero* at $\tau = 0$. Apart from the input features, we introduce the global features which are *temperature* $T(°C)$ and the catalyst descriptors (see next subsection). The input flows are concatenated with global features and fed to the module of the deep neural network as shown in figure 1. Our model consists of modules which are identical deep neural network blocks, arranged sequentially, so the output of one module becomes the input of the 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 $\Delta\tau$ timestep. We employ identity mapping x in our model, to build a deeper neural network and avoid vanishing gradients while training the model. The identity mapping does not have any parameter, while the function $F(x)$ is the residual mapping 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 $y = F(x) + x$ are given as input to the successive module. Using this architecture for the model helps to save the history of how the flows of the given species change inside the model for a given timestep (check suplementary information for more details). To interpret this architecture physically, one can consider each block to represent a discrete elementss of the numerical integration in equation 4 or that each block is effectively a local, smaller, stirred tank reactor model.

Module. The module consists of one embedding layer, where input flows (7 neurons) are concatenated with global features (64 neurons) in a total of 71 neurons which are fed forward to the first hidden linear layer. The module has four hidden linear layers each is made up of 64 neurons, chosen after extensive evaluation of different structures.

Scaled Exponential Linear Unit(SELU) [26] is used as activation function which is given by

$$selu(z) = \lambda z, \quad if \; z \geq 0 \tag{5}$$

$$selu(z) = \lambda\alpha(\exp(z) - 1) \quad if \; z < 0, \tag{6}$$

where $\lambda \approx 1.05070$ and $\alpha \approx 1.673263$.

Adam optimizer with default learning rate of 0.01 is used to optimize the model. The dataset is split randomly on training set (80%) and 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:

$$L = \frac{1}{n}\sum_{i=1}^{n}[(F_{Ar}^i - \hat{F}_{Ar}^i)^2 + (F_{CH_4}^i - \hat{F}_{CH_4}^i)^2 + (F_{O_2}^i - \hat{F}_{O_2}^i)^2 + (F_{C_2H_4}^i -$$
$$-\hat{F}_{C_2H_4}^i)^2 + (F_{C_2H_6}^i - \hat{F}_{C_2H_6}^i)^2 + (F_{CO}^i - \hat{F}_{CO}^i)^2 + (F_{CO_2}^i - \hat{F}_{CO_2}^i)^2] \tag{7}$$

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, CH_4, O_2, C_2H_6, C_2H_4, CO and CO_2) in moles of the molecule per unit time and $\hat{F}$ are the predicted flows form the model.

Flows. The dataset has the inlet flow rates of methane ($F_{CH_4,0}$), oxygen ($F_{O_2,0}$) and argon ($F_{Ar,0}$) as well as the yields of ethane ($y_{C_2H_4}$), ethylene ($y_{C_2H_6}$), carbon monoxide (y_{CO}) and carbon dioxide (y_{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, indeed, are sufficient to carry out atom balance at any point:

$$2CH_4 + 0.5O_2 \rightarrow C_2H_6 + H_2O \tag{8}$$

$$C_2H_6 + 0.5O_2 \rightarrow C_2H_4 + H_2O \tag{9}$$

$$CH_4 + 1.5O_2 \rightarrow CO + 2H_2O \tag{10}$$

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O \tag{11}$$

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.

$$F_{Ar} = F_{Ar,0} \tag{12}$$

$$F_{CH_4} = F_{CH_4,0}(1 - (y_{C_2H_4} + y_{C_2H_6} + y_{CO} + y_{CO_2})) \tag{13}$$

$$F_{O_2} = F_{O_2,0} - F_{CH_4,0}[0.25(y_{C_2H_6}) + 0.25(y_{C_2H_4}) + 1.5(y_{CO}) + 2(y_{CO_2})] \tag{14}$$

$$F_{C_2H_6} = 0.5F_{CH_4,0} * (y_{C_2H_6}) \tag{15}$$

$$F_{C_2H_4} = 0.5 F_{CH_4,0} * (y_{C_2H_4}) \tag{16}$$

$$F_{CO} = F_{CH_4,0} * (y_{CO}) \tag{17}$$

$$F_{CO_2} = F_{CH_4,0} * (y_{CO_2}) \tag{18}$$

$$X_{CH_4} = y_{C_2H_4} + y_{C_2H_6} + y_{CO} + y_{CO_2}, \tag{19}$$

where X_{CH_4} is the methane conversion.
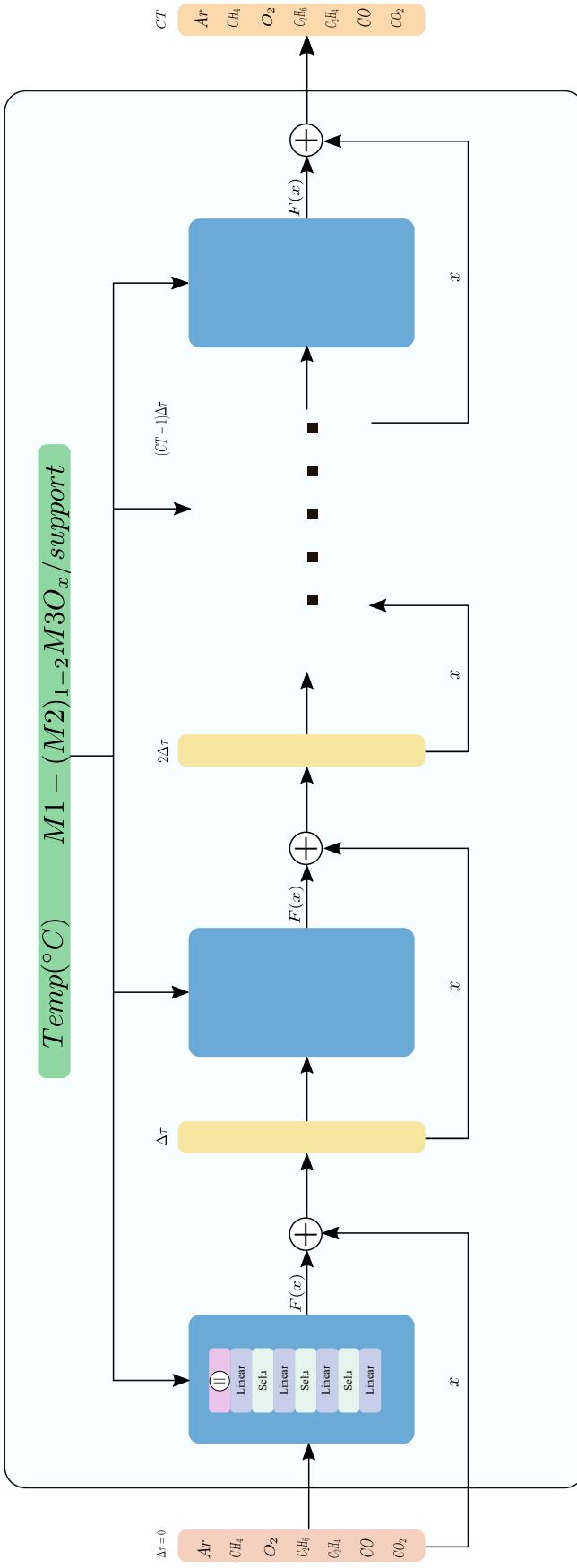
Figure 1: The architecture of OCM model. ▢ denotes the input of our model. It consists of seven input flows F_{Ar}, F_{CH_4}, F_{O_2}, $F_{C_2H_4}, F_{C_2H_6}$, F_{CO} and F_{CO_2}. ▪ represents the global features that are temperature $T(°)$ and catalyst's descriptors ($M1 - (M2)_{1-2}M3O_x/support$)). The 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 ($\oplus$) are given as input for the next module as $y = F(x) + x$. At δt the output of the first module, together with the global features are fed to the second module. This happens recursively for a given δt which is proportional to the amount of time the products spend inside the reactor. ■ are the flows predicted by our model at time-step T.

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 having the general formula $M1-(M2)_{1-2}M3Ox/support$. M1 is mainly a variety of transition metals or lanthanide elements. Molybdenum and tungsten are the two possible choices for M3 and 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_{CH_4}) and yields (and selectivity) of all main carbon-containing products, viz. ethane (C_2H_4), ethylene (C_2H6), carbon monoxide (CO), and carbon dioxide (CO_2), are included in the dataset. Through data preprocessing we discovered that not all data points satisfied carbon and oxygen balance. As a result we filtered out the data points that had a negative computed oxygen flow at the outlet. The methane conversion is calculated as sum of the yields given on the dataset as shown in (Eq.19). The remaining dataset comprises 9271 data points, which are sufficient for further data-driven analysis. We represent each data point by descriptors, including catalyst composition and process conditions. The moles of metals M1, M2, and M3 per unit gram of support describe the catalyst's composition, while the support is represented as integer value. Reaction condition's descriptors are temperature $Temp(°C)$, argon flow (F_{Ar}), methane flow (F_{CH_4}) and oxygen flow (F_{O_2}). All the data points are grouped by the contact time (CT) of 0.75 , 0.50 and 0.38 (seconds) given on the dataset.

4. Results and discussion

4.1. Performance of the neural network model

Mean absolute error (MAE), the root mean square error (RMSE) and (R_2) score on the testing dataset are shown in Table 1. The parity plots are in the supporting information (Figure S1). As we can note, the R_2 score for all the predicted yields and conversion is higher than 0.85, indicating that our deep neural network model is able to capture the relation 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 for a given CT, as well as the effect of various material components on the reaction rate. We note that we use four blocks to train data points with CT = 0.38s, five blocks for data points with CT= 0.5s and seven blocks for CT= 0.75s (i.e., each block roughly corresponds to 0.1s; 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 used this model to analyze one of the most promising OCM catalysts, $Mn-Na_2WO4/SiO_2$, which has attracted a lot of attention due to its relatively high stability and selectivity for C_2 products, viz. ethane (C_2H_4) and ethylene (C_2H_6)[13].

Table 1: Results of the model on testing dataset. Mean absolute error(MAE), Root mean squared error(RMSE) and R_2 score which represents the coefficient of determination.

Predicted variables	MAE	RMSE	R_2 score
Methane conversion (x_{CH_4})	1.6754	2.53	0.93
Ethylene yield $(y_{C_2H_4})$	0.7056	1.02	0.89
Ethane yield $(y_{C_2H_6})$	0.3006	0.44	0.87
Carbon monoxide yield (y_{CO})	1.2940	1.91	0.88
Carbon dioxide yield (y_{CO_2})	0.6637	1.15	0.88

   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 are not all 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 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 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). This nonlinear behavior of the yields is clearly due to the complicated network of reactions involved in oxidative coupling.

   Figure 3 plots C2 selectivity ($\sigma = 100(y_{C_2H_4} + y_{C_2H_6})/X_{CH_4}$)) and the ratio of ethylene to total C2 ($\rho = 100 y_{C_2H_4}/(y_{C_2H_4} + y_{C_2H_6})$)) within the reactor for the reference catalyst. At 750 °C, we can note that the C2 selectivity remains largely constant at around 55% although there is a small increase in the middle of the reactor (and a subsequent decrease). On the other hand, the ratio of ethylene to C2 keeps increasing along the reactor to reach 62%, indicating that more ethylene is formed via oxidative dehydrogenation of ethane along the reactor. Therefore, while the bulk of the conversion of methane to C$_2$ occurs early on, the second half of the reactor enables consolidation of the products and, thereby, the production of more olefin. Similar plots for other catalysts are shown in the supporting information (Figure S4-9); it can be noted that ρ keeps increasing in all cases while σ often remains constant or drops (the drop specifically indicates potential oxidation of C2 products). While these observations offer mechanistic insights, we can further posit that inclusion of a dehydrogenation catalyst that is not

poisoned by gas phase byproducts of OCM at the end of the reactor may further increase selectivity to the olefin.

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 considered three case studies.

First, considering $Mn - Na_2WO_4/SiO_2$ as a reference catalyst, we investigated the local sensitivity of the catalyst to changes in the metal (M1, M2, and M3) and the support, changed one at a time. Note that not all combinations considered in the local sensitivity analysis were actually experimentally examined. The reaction condition is: 1123.15 K(850°C), 10.5 ml/min methane CH_4 flow rate, 0.75 ml/min oxygen O_2 flow, 10 ml/min Argon Ar, and CT = 0.1 (i.e. one neural network block). For this setting, the reference methane conversion X_{CH_4} 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 significant negative effect on methane conversion and C2 yield, indicating that Na is a the local optimal choice. Finally, changing the support to TiO₂ seems to improve both conversion of methane and the yield of C2 products.

Second, having looked at local sensitivities of the reference catalyst, we considered 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 $59 \cdot 59 = 3481$ combinations. The reaction conditions are essentially same as before (except for CT): 1123.15 K(850°C), 10.5 ml/min methane CH_4 flow rate, 0.75 ml/min oxygen O_2 flow, 10 ml/min 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 $Mn-Na_2WO_4/SiO_2$ 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 higher yield than the reference. The table 2 shows the top performing candidates (vis-a-vis the reference at the bottom) in terms of the yield of the C_2 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 $Mn - Na_2WO_4/SiO_2$ with regard to C2 yield (e.g. $Mn - Na_2WO_4/SiCnf$), 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 2, viz. $Mn-Na_2WO_4/TiO_2+Mn-Na_2WO_4/SiCnf$, 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 $\frac{y_{C2}}{X_{CH_4}}$) is 75%, which is lower than that of the second catalyst and the combined bed (both about 83%). That is, combining a more active catalyst with one that is more selective leads to better performing catalytic system than using either catalyst would allow. A similar explanation can be provided for the other combinations in Table 2). This promising strategy of mixing catalysts to exploit synergies need further experimental validation. The complete list of 3481 catalysts considered here is given in the supporting information.

Table 2: Top performing catalyst combinations for a two-bed catalyst system (the reference catalyst at the end provides a baseline). The operating condition is: 1123.15 K(850°C), 10.5 ml/min methane (CH_4) flow rate, 0.75 ml/min oxygen (O_2) flow, and 10 ml/min Argon (Ar) and for each catlayst (for a total contact time of 0.2). For easy comparison, the values of the C_2 yield (y_{C2}) and methane conversion (X_{CH_4}) for each of the catalysts are reported at the same conditions and CT= 0.2

| Catalyst 1 | Catalyst 2 | y_{C2} | X_{CH_4} | Catalyst 1 | | Catalyst 2 | |
				y_{C2}	X_{CH_4}	y_{C2}	X_{CH_4}
$Mn-Na_2WO_4/TiO_2$	$Mn-Na_2WO_4/SiCnf$	9.72	11.65	9.47	12.68	9.16	11.07
$Mn-Na_2WO_4/TiO_2$	$Mn-Na_2WO_4/SiC$	9.56	12.02	9.47	12.68	8.54	10.14
$Mn-Na_2WO_4/TiO_2$	$Mn-Na_2WO_4/SiO_2$	9.54	12.31	9.47	12.68	8.72	11.51
$Ti-NaWO_4/SiO_2$	$Mn-Na_2WO_4/SiCnf$	9.41	11.85	9.04	11.55	9.16	11.07
$Mn-Na_2WO_4/SiO_2$	$Mn-Na_2WO_4/SiO_2$	8.72	11.51	-	-	-	-

Third, we expanded our search for optimal catalysts to compositions outside those considered in the high throughput data set but using the same space of materials for M1-M3 and the support. Specifically, we considered all permutations of catalysts within our space of metals and supports ($19_{M1} \cdot 10_{M2} \cdot 3_{M3} \cdot 13_{support} = 7410$), keeping the same relative concentration of M1, M2, and M3 as the reference catalyst. We evaluated their predicted conversion and C2 yield at the same operating condition 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 there are some catalysts that 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 $Nd-Na_2WO_4/SiC$) have C2 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. The complete list of 7410 catalysts considered here is given in the supporting information.

Table 3: Top performing catalysts from the expanded single catalyst search (the reference catalyst is listed at the end). The operating condition: 1123.15 K($850°C$), 10.5 ml/min methane (CH_4) flow rate, 0.75 ml/min oxygen (O_2) flow, and 10 ml/min Argon (Ar) and CT $= 0.1$(i.e., 1 prediction block)

Catalyst	C_2 yield	X_{CH_4}
$Ti - Na_2WO_4/SiC$	6.18	7.23
$Ti - KWO_4/SiC$	5.97	7.27
$Nd - Na_2WO_4/SiC$	5.96	6.84
$Mn - KWO_4/TiO_2$	5.95	7.47
$Ti - KWO_4/TiO_2$	5.93	8.09
$Mn - Na_2WO_4/SiO_2$	5.54	6.44

(a)

(b)

(c)

(d)

(e)
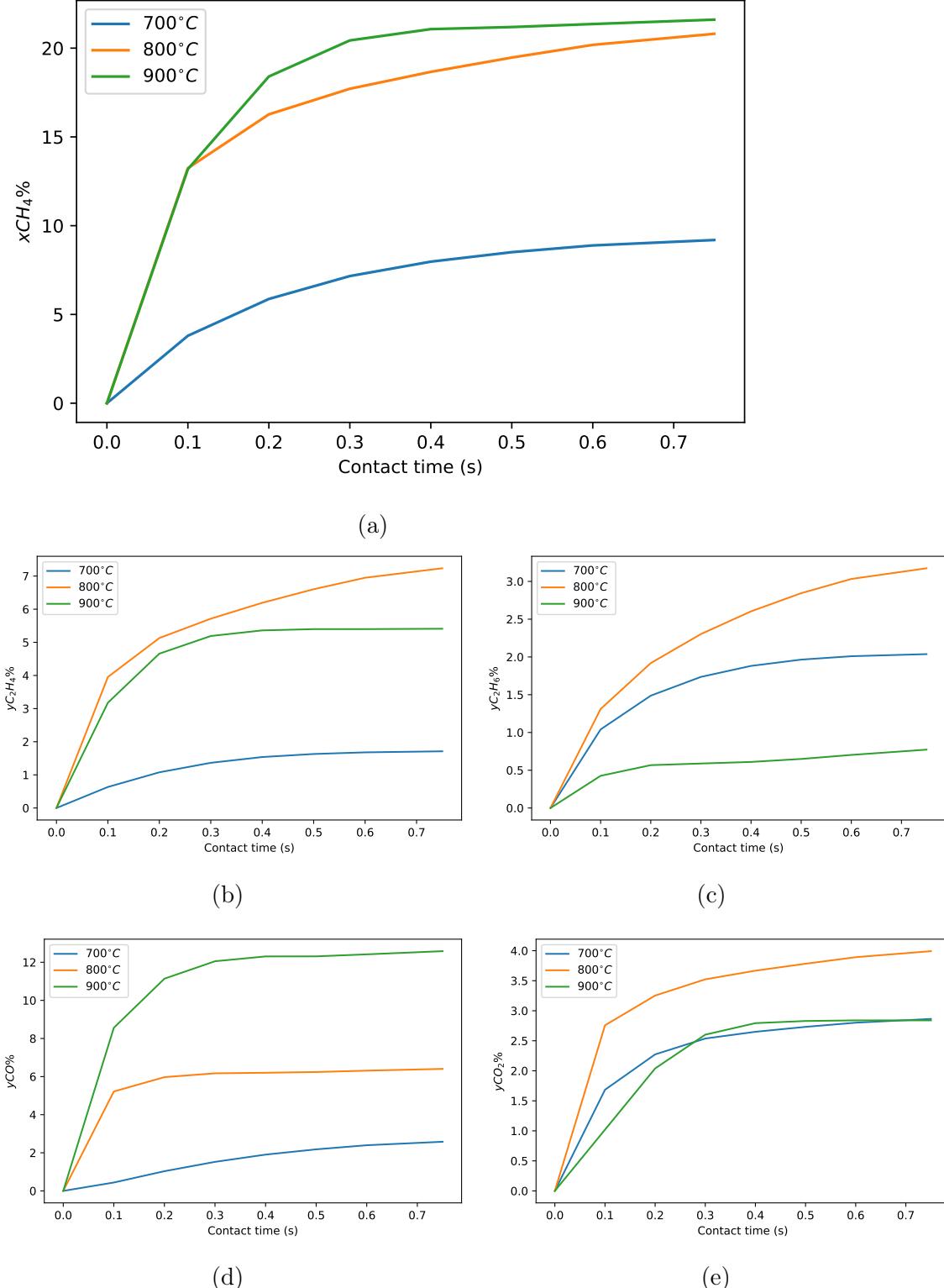
Figure 2: **(a)** Conversion of methane X_{CH4}, **(b)** Ethylene yield $y_{C_2H_4}$, **(c)** Ethane yield $y_{C_2H_6}$, **(d)** Carbon monoxide yield y_{CO} and **(e)** Carbon dioxide yield y_{CO_2} for three different temperatures($700°C$, $800°C$, $900°C$) on the reference catalyst ($Mn - Na_2WO4/SiO_2$). The operating conditions for all the above temperatures are 4.8 ml/min methane (CH_4) flow rate, 1.2 ml/min oxygen(O_2) flow and 4 ml/min Argon (Ar). CT =0.75 in all cases(7 residual blocks for each temperature).
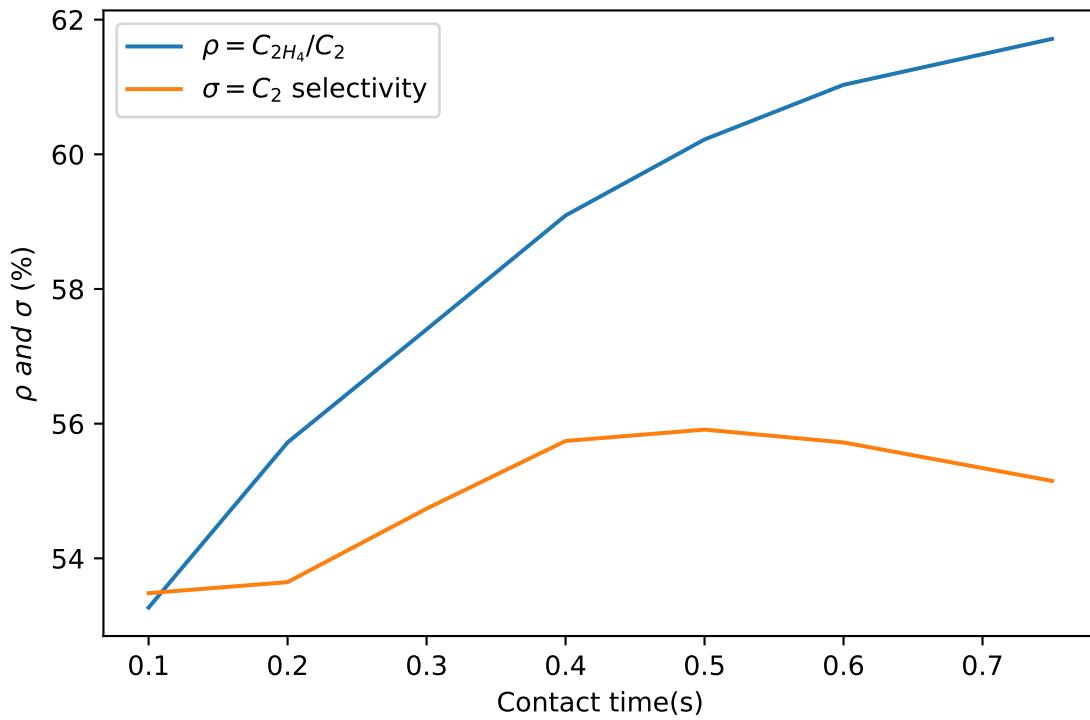
Figure 3: Ratio C_2H_4/C_2 and C_2 selectivity (both in percentage) at temperature $750°C$ for the reference catalyst ($Mn - Na_2WO4/SiO_2$).For the given temperature the operating conditions are 4.8 ml/min methane (CH_4) flow rate, 1.2 ml/min oxygen(O_2) flow and 4 ml/min Argon (Ar). CT = 0.75 in all cases(7 residual blocks).
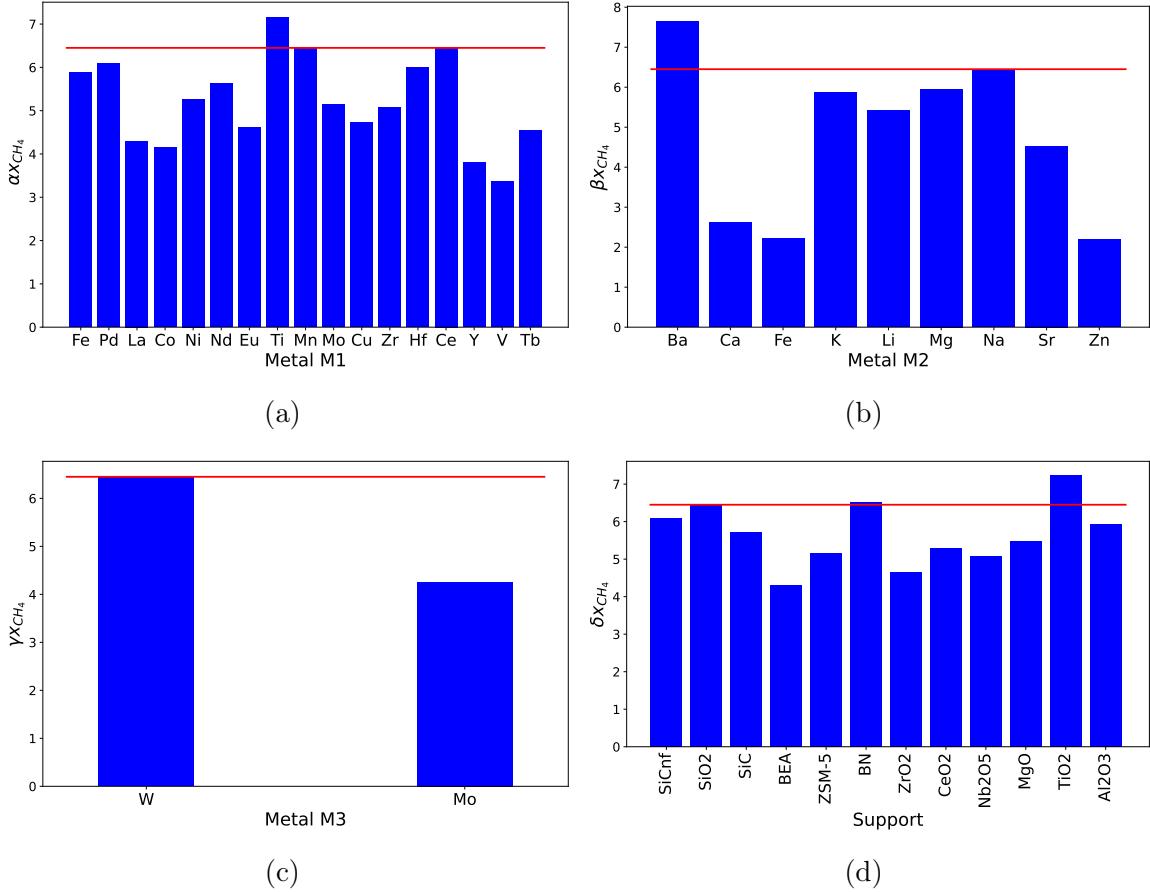
Figure 4: The effect of varying metals M1,M2,M3 and support individually on the methane conversion. The red line represents the reference point ($Mn-Na2WO4/SiO2$). The reaction condition is: 1123.15 K($850°C$), 10.5 ml/min methane CH_4 flow rate, 0.75 ml/min oxygen O_2 flow, 10 ml/min Argon Ar, and CT = 0.1 (i.e. one neural network block).
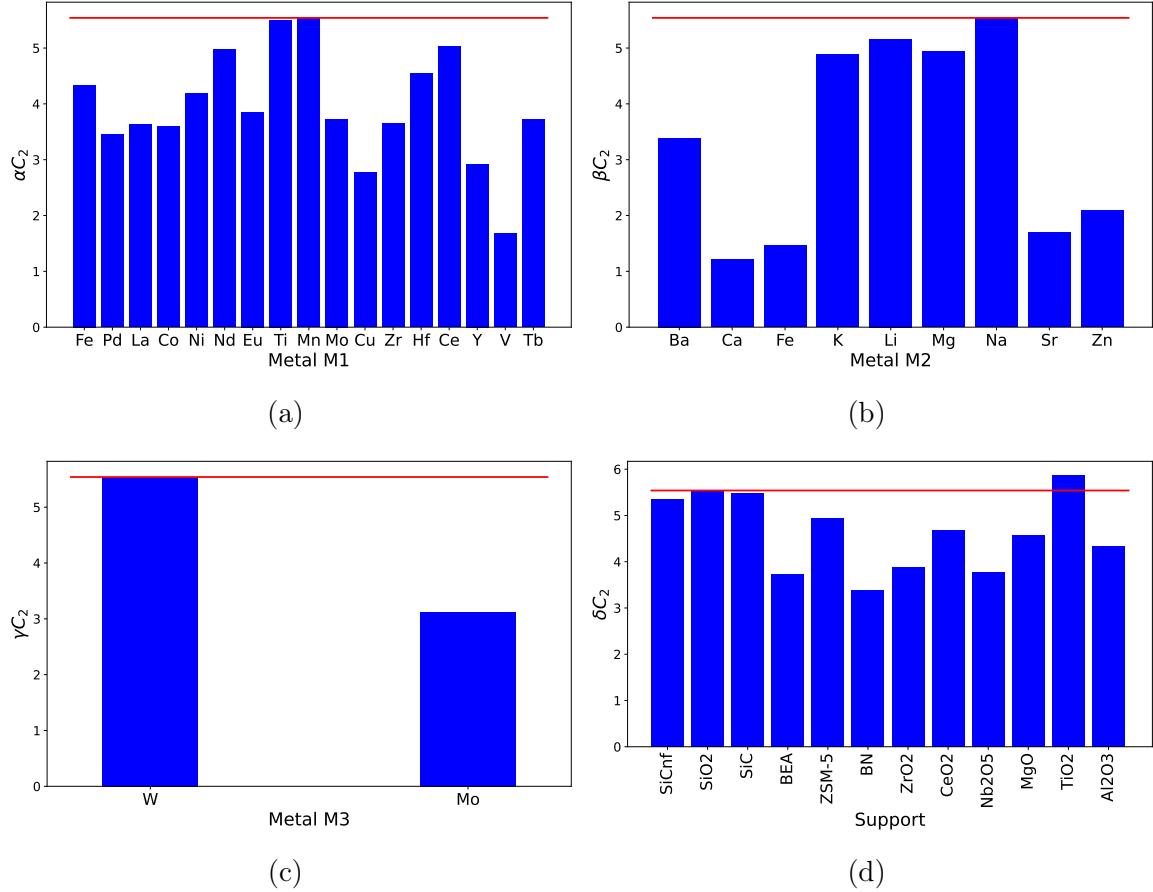
Figure 5: The effect of varying metals M1,M2,M3 and support individually on the C_2 yield. The red line represents the reference point ($Mn - Na_2WO4/SiO_2$). n.a means that there is no metal or support.
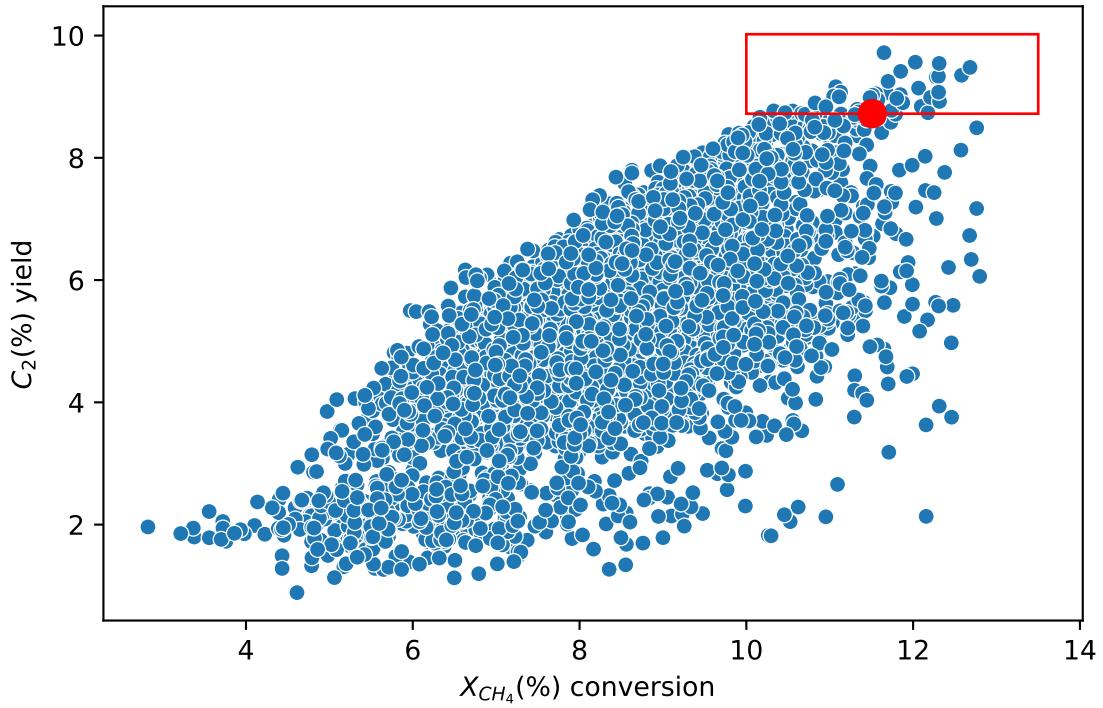
Figure 6: Scatter plot of predicted conversion and yield of all combinations of the two-catalyst beds using only those catalysts considered in the high throughput data set. The red box encompasses those catalysts with higher C2 yield compared to the reference $Mn - Na_2WO_4/SiO_2$ (red point). The operating specs are: 1123.15 K(850°C), 10.5 ml/min methane (CH_4) flow rate, 0.75 ml/min oxygen (O_2) flow, and 10 ml/min Argon (Ar) and CT = 0.2 in total for both catalysts (1 block for each catlayst).
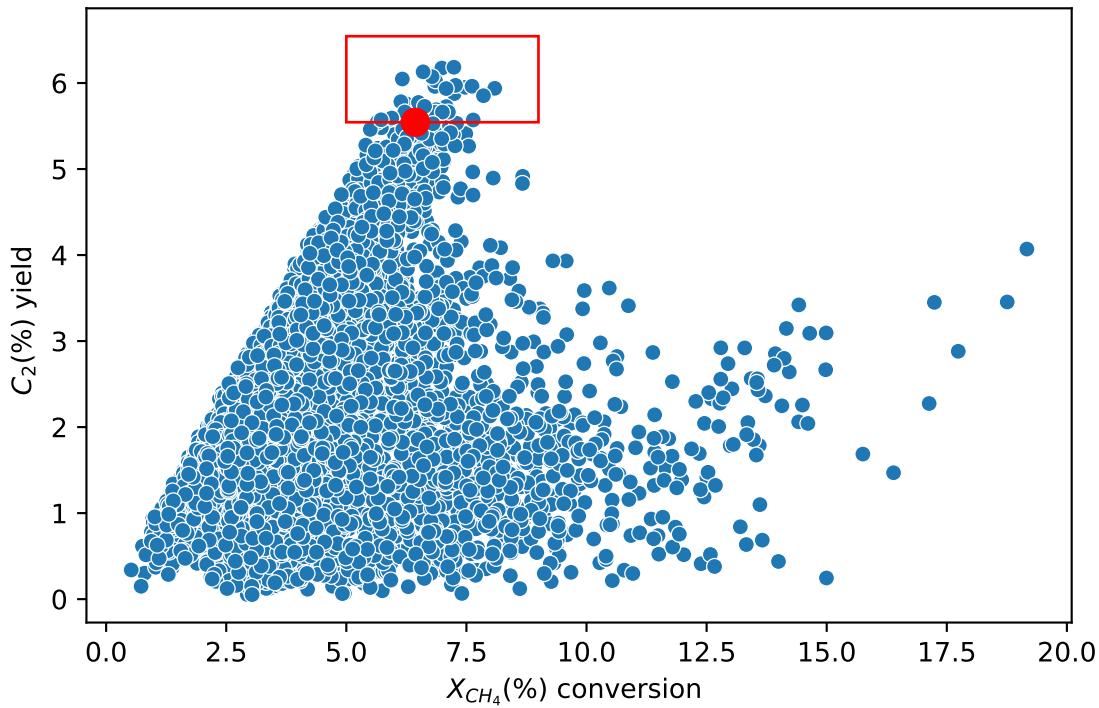
Figure 7: A scatter plot of the predicted C2 yield and methane conversion of all possible catalyst combinations but using only those metals and supports considered in the original high throughput experimentation data set. The reference catalyst, Mn-Na_2WO_4/SiO_2, is marked in red and the red box shown encompasses the space of higher-yielding catalysts. The operating specs are: 1123.15 K($850°C$), 10.5 ml/min methane (CH_4) flow rate, 0.75 ml/min oxygen (O_2) flow, and 10 ml/min Argon (Ar) and CT = 0.1.

5. Conclusion & 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 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 ($Mn - Na_2WO4/SiO_2$), 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 C2 products. Local and global catalyst screening studies pointed out that while the reference catalyst is locally near optimum, there are potentially other catalysts and catalyst combinations (comprising of two catalyst beds) that can lead to better yields. We posit that since the rate model was learned to satisfy basic underlying physics of the reactor: (1) it is robust enough to be used to explore catalysts beyond those considered in the experiments but within the larger space of material choices and (2) it can be employed to consider novel catalyst combinations or different process conditions that were not considered in the experiments. This work, therefore, represents a first approach to developing domain informed deep learning models of reaction rate from high throughput experiments; we posit that the domain information can further include physical, chemical, and thermodynamic constraints that the model must additionally intrinsically satisfy beyond the reactor design equation.

[1] Medford A J, Kunz M R, Ewing S M, Borders T and Fushimi R 2018 *ACS Catalysis* **8** 7403–7429

[2] Toyao T, Maeno Z, Takakusagi S, Kamachi T, Takigawa I and Shimizu K i 2019 *Acs Catalysis* **10** 2260–2297

[3] Takahashi K, Takahashi L, Miyazato I, Fujima J, Tanaka Y, Uno T, Satoh H, Ohno K, Nishida M, Hirai K *et al.* 2019 *ChemCatChem* **11** 1146–1152

[4] Nguyen T, Nhat T, Takimoto K, Thakur A, Nishimura S, Ohyama J, Miyazato I, Takahashi L, Fujima J, Takahashi K and Taniike T 2019 *ACS Catalysis* **10** 921–932

[5] McCullough K, Williams T, Mingle K, Jamshidi P and Lauterbach J 2020 *Physical Chemistry Chemical Physics* **22** 11174–11196

[6] Odabaşı Ç, Günay M E and Yıldırım R 2014 *International journal of hydrogen energy* **39** 5733–5746

[7] Zavyalova U, Holena M, Schlögl R and Baerns M 2011 *ChemCatChem* **3** 1935–1947

[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** 3636–3655

[9] Takahashi K, Miyazato I, Nishimura S and Ohyama J 2018 *ChemCatChem* **10** 3223–3228

[10] Nguyen T N, Nakanowatari S, Nhat Tran T P, Thakur A, Takahashi L, Takahashi K and Taniike T 2021 *ACS Catalysis* **11** 1797–1809

[11] Nakanowatari S, Nguyen T N, Chikuma H, Fujiwara A, Seenivasan K, Thakur A, Takahashi L, Takahashi K and Taniike T 2021 *ChemCatChem* **13** 3262–3269

[12] Sourav S, Wang Y, Kiani D, Baltrusaitis J, Fushimi R R and Wachs I E 2021 *Angewandte Chemie International Edition*

[13] Kiani D, Sourav S, Baltrusaitis J and Wachs I E 2019 *ACS Catalysis* **9** 5912–5928

[14] Horn R and Schlögl R 2015 *Catalysis Letters* **145** 23–39

[15] Galadima A and Muraza O 2016 *Journal of Industrial and Engineering Chemistry* **37** 1–13

[16] Pak S, Qiu P and Lunsford J H 1998 *Journal of Catalysis* **179** 222–230

[17] Chen K, Tian H, Li B and Rangarajan S 2022 *AIChE Journal* **68** e17584

[18] Long Z, Lu Y, Ma X and Dong B 2018 Pde-net: Learning pdes from data *International Conference on Machine Learning* (PMLR) pp 3208–3216

[19] Gulgec N S, Shi Z, Deshmukh N, Pakzad S and Takáč 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, Takáč M and Matarazzo T J 2020 *Journal of Engineering Mechanics* **146** 04020011

[21] Shi Z, Gulgec N S, Berahas A S, Pakzad S N and Takáč M 2020 Finite difference neural networks: Fast prediction of partial differential equations *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE) pp 130–135

[22] Long Z, Lu Y and Dong B 2018 *CoRR* **abs/1812.04426** (*Preprint* 1812.04426) URL http://arxiv.org/abs/1812.04426

[23] He K, Zhang X, Ren S and Sun J 2015 *CoRR* **abs/1512.03385** (*Preprint* 1512.03385) URL http://arxiv.org/abs/1512.03385

[24] Tathawadekar N, Doan N A K, Silva C F and Thuerey N 2021 *arXiv preprint arXiv:2111.11185*

[25] Eshkevari S S, Takáč M, Pakzad S N and Jahani M 2021 *Engineering Structures* **229** 111582

[26] Klambauer G, Unterthiner T, Mayr A and Hochreiter S 2017 Self-normalizing neural networks *Proceedings of the 31st International Conference on Neural Information Processing Systems* NIPS'17 (Red Hook, NY, USA: Curran Associates Inc.) p 972–981 ISBN 9781510860964