УДК 004.85:004.94:544.478:544.723

ОПТИМИЗАЦИЯ РЕАКЦИИ ОКИСЛЕНИЯ УГАРНОГО ГАЗА НА ПОВЕРХНОСТИ НАНОЧАСТИЦ ПАЛЛАДИЯ МЕТОДОМ МАШИННОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

© 2023 г. М. С. Лифарь[a, b], А. А. Терещенко[a, *], А. Н. Булгаков[a], А. А. Гуда[a, **], С. А. Гуда[a, b], А. В. Солдатов[a]

[a] Международный исследовательский институт интеллектуальных материалов, Южный федеральный университет, Ростов-на-Дону, 344090 Россия

[b] Институт математики, механики и компьютерных наук им. И.И. Воровича, Южный федеральный университет, Ростов-на-Дону, 344058 Россия

*e-mail: tereshch1@gmail.com

**e-mail: guda@sfedu.ru

Поступила в редакцию 17.06.2022 г.
После доработки 22.08.2022 г.
Принята к публикации 22.08.2022 г.

Выход продуктов реакции зависит от взаимодействия между процессами на поверхности катализатора: адсорбции, активации, десорбции и других. Эти процессы, в свою очередь, зависят от величин потоков реакционных смесей, температуры и давления. В стационарных условиях активные центры на поверхности могут быть отравлены побочными продуктами реакции или заблокированы избытком адсорбированных молекул реагентов. Динамический контроль параметров реакции учитывает изменения свойств поверхности и соответствующим образом регулирует температуру, скорости потоков и другие параметры. Применен алгоритм обучения с подкреплением для управления реакцией окисления угарного газа CO на поверхности наночастиц палладия. Алгоритм был натренирован максимизировать скорость производства углекислого газа на основе информации о величинах потоков CO, O_2 и CO_2 на каждом временно́м шаге. Был выбран алгоритм градиентной политики с непрерывным пространством действий, и расширены наблюдения за скоростями потока на несколько последовательных временны́х шагов, что позволило получить набор нестационарных решений. Максимальный выход продукта достигается при периодическом изменении газовых потоков, обеспечивающем баланс между доступными центрами адсорбции и концентрацией активированных интермедиатов. Эта методология открывает перспективы для оптимизации каталитических реакций в нестационарных условиях.

Ключевые слова: машинное обучение, обучение с подкреплением, катализаторы, наночастицы палладия, адсорбция, монооксид углерода.

DOI: 10.31857/S1028096023030081, **EDN:** LMOGVS

ВВЕДЕНИЕ

Наночастицы благородных металлов, в частности палладия, – известные катализаторы множества химических реакций окисления [1] и восстановления [2]. Их каталитическая активность помимо размера частиц [3] и материала подложки [4] также во многом определяется и формой наночастиц [5–8]. В стационарных условиях протекания реакции поверхность катализатора может деградировать за счет формирования оксидов [9], карбидов [10] или адсорбированных молекул с большой энергией связи [11].

Для замедления процесса формирования вторичных фаз на поверхности и повышения активности катализатора условия протекания реакции можно изменять динамически [12]. Например, при проведении каталитического циклирования на отдельной стадии цикла поверхность катализатора может быть принудительно очищена от нежелательных продуктов и, таким образом, осуществлена ее регенерация для лучшего заселения желательными реагентами на других стадиях [13]. Использование внешних возмущений может привести к большему преимуществу динамического режима по сравнению со стационарным, что используется для разработки нестационарных реакторов с лучшими характеристиками. В частности, модуляции реакции монооксида углерода на поверхности нанокатализаторов благородных ме-

таллов представляют особый интерес, и их интенсивно изучают: это не только распространенная модельная реакция для фундаментальных исследований в области гетерогенного катализа [14–18], но и широко используемый в настоящее время процесс в автомобильных каталитических конвертерах. Там трехходовые катализаторы периодически подвергаются окислительному (воздушный реактор) и восстановительному (топливный реактор) режимам, и такое циклирование позволяет резко увеличить эффективность итоговой конверсии CO в CO_2 [19–21].

Поиск оптимальных реакционных условий является сложной задачей даже для стационарных режимов и значительно усложняется в случае динамически изменяющихся параметров. Одним из перспективных подходов для решения данной задачи является использование машинного обучения [22, 23] и, в частности, обучения с подкреплением (известного в зарубежной литературе как Reinforcement Learning) [24–26] для прогнозирования наилучшего динамического режима и условий протекания каталитических реакций. Недавно в [27] обосновали концепцию использования машинного обучения с подкреплением для оптимизации выхода водорода в реакции частичного окисления метана. С этой целью они обучили агентов *Q*-обучения [28, 29] и градиента глубокой детерминированной политики [30] для прогнозирования производства водорода путем регулирования температуры, давления, скорости потока и состава подложки в смоделированном реакторе идеального вытеснения. Авторы [31] продемонстрировали применимость машинного обучения с подкреплением в сочетании с прогностическим контролем экономической модели для производства оксида этилена. Также, используя глубокое обучение с подкреплением, ранее оптимизировали различные микрокапельные химические реакции, например синтез изохинолина, замещенного хинолина и рибозофосфата [32]. В результате данный подход позволил уточнить оптимальные экспериментальные условия, что дало возможность увеличить скорость протекания реакций. В настоящей работе проведена оптимизация параметров реакции окисления CO, протекающей на поверхности наночастиц Pd, с использованием подхода машинного обучения с подкреплением.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Алгоритм обучения с подкреплением требует много пробных шагов для обучения. Испытания выполняют последовательно для разных значений параметров, выбираемых алгоритмом, и таким образом покрывают важные области пространства параметров реакции. Этот подход отличается от машинного обучения с учителем (известного в зарубежной литературе как Supervised Machine Learning), когда весь набор испытаний передается алгоритму пользователем. В этом разделе опишем математическую модель реакции окисления CO, основанную на системе дифференциальных уравнений, которая служит средой для обучения алгоритма.

Схема реакции окисления CO на Pd может быть описана уравнением:

$$2CO + O_2 \rightarrow 2CO_2. \quad (1)$$

Окисление CO на наночастицах Pd можно разделить на четыре элементарных этапа по механизму Ленгмюра–Хиншельвуда [33, 34]:

$$CO + Pd^* \xrightarrow{k_1} CO_{Pd}, \quad (2)$$

$$CO_{Pd} \xrightarrow{k_2} CO + Pd^*, \quad (3)$$

$$O_2 + 2Pd^* \xrightarrow{k_3} 2O_{Pd}, \quad (4)$$

$$CO_{Pd} + O_{Pd} \xrightarrow{k_4} CO_2 + 2Pd^*, \quad (5)$$

где Pd* – свободный центр адсорбции, а k_i – константа скорости стадии i = 1–4. Значения констант скорости определили, используя следующие формулы:

$$k_1 = \frac{F_{CO}}{n_{Pd}}, \quad (6)$$

$$k_2 = \nu_2 \exp\left(-\frac{E_2}{k_B T}\right), \quad (7)$$

$$k_3 = \frac{F_{O_2}}{n_{Pd}}, \quad (8)$$

$$k_4 = \nu_4 \exp\left(-\frac{E_4}{k_B T}\right), \quad (9)$$

где F_{CO} и F_{O_2} – потоки CO и O_2 соответственно, n_{Pd} – поверхностная плотность атомов Pd (n_{Pd} = = 1.53 × 10^{19} м$^{-2}$), ν_i – частотный фактор элементарной реакции i ($\nu_2 = 10^{15}$ с$^{-1}$ и $\nu_4 = 10^{7.9}$ с$^{-1}$), E_i – энергия активации элементарной реакции i (E_2 = 136 кДж · моль$^{-1}$ и E_4 = 59 кДж · моль$^{-1}$), k_B – константа Больцмана (k_B = 0.008314463 кДж · · моль$^{-1}$ · K^{-1}), T – температура образца (440 K). Для константы скорости уравнения (4) в выражение (8) были дополнительно введены поправки на деградацию и регенерацию поверхности катализатора с течением времени:

$$k_4' = \begin{cases} k_4 + k_4 V_d \left(0.25 - \dfrac{P_{O_2}}{P_{CO}}\right)\Delta t, & \dfrac{P_{O_2}}{P_{CO}} > 0.25 \\ k_4 + (1 - k_4) V_r \left(0.25 - \dfrac{P_{O_2}}{P_{CO}}\right)\Delta t, & \dfrac{P_{O_2}}{P_{CO}} \le 0.25. \end{cases} \quad (10)$$

где V_d – параметр, описывающий скорость деградации поверхности ($V_d = 0.01$ с$^{-1}$), Δt – временно́й интервал, в ходе которого происходит деградация или регенерация, P_{CO} и P_{O_2} – парциальные давления CO и O_2, V_r – параметр, описывающий скорость регенерации поверхности ($V_r = 1.5$ с$^{-1}$). В случае, если доля CO в смеси была мала (отношение потоков $\frac{P_{O_2}}{P_{CO}} \geq 2$), в формуле (9) $\frac{P_{O_2}}{P_{CO}}$ принимали равным двум, чтобы ограничить скорость деградации катализатора. В данной модели и был использован тот факт, что изменение активности может быть связано, например, с ростом доли примесной фазы на поверхности палладия [9]. Намеренно было выбрано соотношение $P_{O_2} : P_{CO} = 1 : 4$ в качестве пограничного (где катализатор не подвергается деградации или регенерации), поскольку оно далеко от стехиометрического соотношения $P_{O_2} : P_{CO} = 1 : 2$, что позволяет получить нестационарные решения. Величину потока определяли с использованием парциальных давлений соответствующих газов по формулам:

$$F_{CO} = \frac{P_{CO}}{\sqrt{\frac{2\pi M_{CO}}{N_A k_B T}}}, \quad (11)$$

$$F_{O_2} = \frac{P_{O_2}}{\sqrt{\frac{2\pi M_{O_2}}{N_A k_B T}}}, \quad (12)$$

где P_{CO} и P_{O_2} – парциальные давления CO и O_2, M_{CO} и M_{O_2} – молекулярные массы CO и O_2 ($M_{CO} = 28 \times 10^{-3}$ кг · моль$^{-1}$ и $M_{O_2} = 32 \times 10^{-3}$ кг · моль$^{-1}$), N_A – постоянная Авогадро. На основе этой модели можно ввести уравнения для покрытия поверхности молекулами монооксида углерода и атомами кислорода:

$$\frac{d\theta_{CO}}{dt} = k_1 S_{CO} - k_2 \theta_{CO} - k_4 \theta_{CO}\theta_O, \quad (13)$$

$$\frac{d\theta_O}{dt} = 2k_3 S_{O_2} - k_4 \theta_{CO}\theta_O, \quad (14)$$

где θ_{CO} и θ_O – покрытия поверхности Pd молекулами CO и атомами кислорода, а S_{CO} и S_{O_2} – соответствующие коэффициенты прилипания (известные в зарубежной литературе как sticking coefficients). Зависимости коэффициентов прилипания S_{CO} and S_{O_2} от покрытия и температурных эффектов определяли по формулам:

$$S_{CO} = S_{CO}^0 \left(1 - \frac{\theta_{CO}}{\theta_{CO}^{max}} - C_T \frac{\theta_O}{\theta_O^{max}}\right), \quad (15)$$

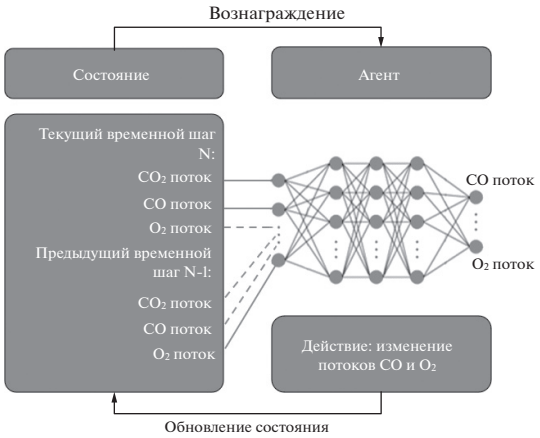

Рис. 1. Схема работы алгоритма.

$$S_{O_2} = \begin{cases} S_{O_2}^0 \left(1 - \frac{\theta_{CO}}{\theta_{CO}^{max}} - \frac{\theta_O}{\theta_O^{max}}\right)^2, & 1 - \frac{\theta_{CO}}{\theta_{CO}^{max}} - \frac{\theta_O}{\theta_O^{max}} \geq 0 \\ 0; & 1 - \frac{\theta_{CO}}{\theta_{CO}^{max}} - \frac{\theta_O}{\theta_O^{max}} < 0, \end{cases} \quad (16)$$

где S_{CO}^0 и $S_{O_2}^0$ – начальные коэффициенты прилипания при нулевом покрытии соответственно, $S_{CO}^0 = 0.96$ и от температуры не зависит, $S_{O_2}^0 = (0.78–7.4) \times 10^{-4} T$, где T – температура образца, $\theta_{CO}^{max} = 0.5$ и $\theta_O^{max} = 0.25$, C_T был принят равным 0.3. Параметры модели были взяты из [35]. Итоговую скорость формирования CO_2 (r_{CO_2}) определяли по формуле:

$$r_{CO_2} = k_4' \theta_{CO}\theta_O. \quad (17)$$

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Ключевыми параметра алгоритма обучения с подкреплением являются среда и агент. Агент выполняет действия на основе политики, задаваемой нейронной сетью, которая получает на вход значения наблюдаемых параметров из среды и на выходе выдает действия. Был использован алгоритм градиентной политики Vanilla Policy Gradient (VPG) с непрерывными значениями действий. Политика алгоритма обновлялась итерационно в ходе обучения на модели, построенной на основе дифференциальных уравнений. Целью обучения была максимизация интегрального значения награды в изменяющихся условиях среды. Модель и обучение были запрограммированы на языке Python с использованием библиотеки Tensorforce и фреймворка Tensorflow. При оптимизации коэффициентов нейронной сети использовали скорость обучения 0.001. На входы нейронной

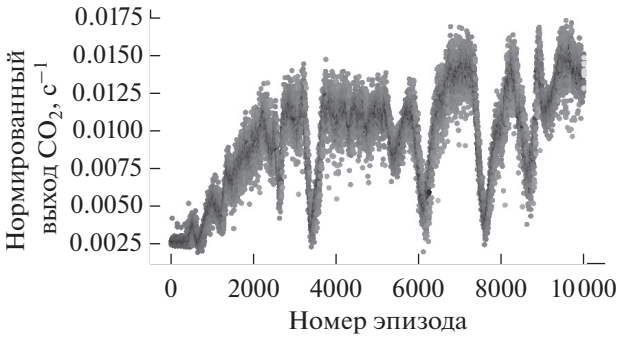

Рис. 2. Суммарный выход CO_2, деленный на длину эпизода, как функция от номера эпизода.

сети подавали нормированные значения потоков газов, приведенные к диапазону [0; 1]. Схема работы алгоритма показана на рис. 1.

Алгоритм VPG обучения с подкреплением относится к классу алгоритмов on-policy, т.е. алгоритмов, которые оптимизируют политику агента, базируясь только на информации, полученной в ходе использования текущей политики. Политика – функция, определяющая, какое действие будет выбирать алгоритм исходя из текущего состояния системы. Обучение агента происходит в течение большого числа эпизодов обучения. Алгоритм в течение эпизода длительностью 500 с применяет политику, запрограммированную в виде нейронной сети. Задача по нахождению агентом оптимальной политики сводится к оптимизации параметров политики – весов нейронной сети. Действие агента в настоящей работе заключается в изменении потоков CO и O_2 каждые 10 с произвольным образом в пределах заданного диапазона. На вход нейронной сети подают нормированные значения потоков CO, O_2 и CO_2 на текущем шаге и (опционально) значения соответствующих потоков на одном или двух предыдущих шагах обучения. На выходе считывали новые значения потоков CO и O_2. По окончанию эпохи обучения коэффициенты нейронной сети обновляются согласно правилу:

$$\theta_{k+1} = \theta_k + \alpha \nabla_\theta \left(J\left(\pi_\theta\right)\right)\Big|_{\theta k}, \qquad (18)$$

где θ – веса нейросети, π_θ – политика агента, $J\left(\pi_\theta\right)$ – ожидаемая кумулятивная награда для данной политики π_θ. Градиент в формуле (18) носит название градиента политики. Награда J вычисляется как интегральный выход молекул CO_2 реакции окисления $CO + O_2$. На рис. 2 показан пример обучения алгоритма на протяжении 10000 эпизодов. Наблюдается немонотонное увеличение награды, получаемой алгоритмом за одну эпоху. Выбор способа награждения и спецификации состояния и действия определяют качество и скорость обучения алгоритма.

Чтобы оценить эффективность решений, найденных алгоритмом обучения с подкреплением, было проведено сравнение их с решениями, найденными иным методом. Рассматривали суммарный выход CO_2 за эпизод как функцию от потоков $O_2(t)$ и $CO(t)$: $R = R(O_2(t), CO(t))$. Далее рассматривали только стационарные решения, т.е. такие, что $O_2(t) = O_2 = \text{const}$, $CO(t) = CO = \text{const}$. Затем максимизировали R как функцию двух вещественных переменных, используя метод Нелдера–Мида, или симплекс-метод [36]. Полученное стационарное решение использовали для сравнения с решением, найденным обучением с подкреплением.

В табл. 1 собраны результаты тренировки в течение 10000 эпизодов для разных комбинаций. Как видно, наилучшее качество обучения было достигнуто при использовании информации о потоках газов на протяжении трех последних шагов по времени. После тренировки алгоритм можно применять для оптимального управления экспериментальной установкой синтеза. На рис. 3 показаны результаты найденной политики для уравнений (1)–(3) со случайными начальными условиями, полученными с помощью оптимиза-

Таблица 1. Выбор состояния для тренировки алгоритма и интегральный выход продуктов реакции за одну эпоху

Состояние	Выход продуктов реакции за эпоху, отн. ед.
Потоки газов на текущем шаге по времени: $Obs_0 = (CO, O_2, CO_2)$	8.54
Потоки газов на текущем и предыдущем шаге по времени: Obs_0, Obs_1	9.35
Потоки газов на трех последующих шагах по времени: Obs_0, Obs_1, Obs_2	10.16

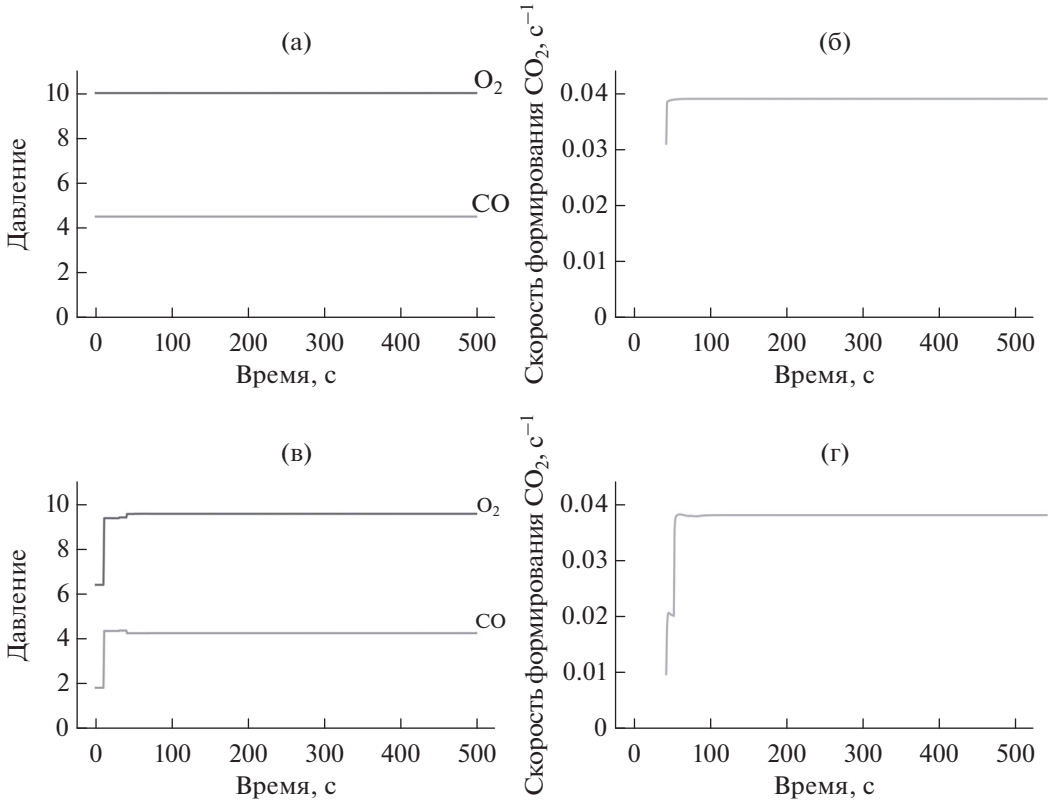

Рис. 3. Решение: а, б – стационарное, найденное методом оптимизации; в, г – полученное с помощью алгоритма обучения с подкреплением. В обоих случаях использованы модели без учета деградации поверхности катализатора. Показаны задаваемые потоки CO и O_2 (предсказанные политики) (а, в) и соответствующий им выход CO_2 (б, г).

ции (рис. 3а), и политика, найденная алгоритмом обучения с подкреплением (рис. 3б) соответственно с использованием модели без учета деградации.

Из рисунков видно, что при отсутствии в модели учета деградации катализатора алгоритм находит стационарное решение с фиксированными потоками газов, близкое к найденному оптимизацией, с соотношением потоков CO : O_2 около 0.446. Схождение алгоритма к постоянному решению можно объяснить тем, что доля поверхности, доступной для протекания реакции CO + O_2, остается неизменной в процессе синтеза и близкой к оптимальной. В данном случае периодическая реактивация поверхности катализатора повышенным потоком кислорода не приведет к увеличению выхода CO_2.

На рис. 4 приведены решения, полученные для моделей с учетом деградации: стационарное решение, найденное с помощью оптимизации (рис. 4а), и решение, найденное алгоритмом обучения с подкреплением (рис. 4б). Сравнение выхода CO_2 показывает, что для модели с учетом деградации алгоритм смог найти динамическое решение, которое обеспечивает значительно больший выход CO_2.

При длительном нахождении катализатора палладия в атмосфере с неравновесным содержанием кислорода или монооксида углерода может образовываться фаза карбида или оксида, которая ухудшает каталитические свойства поверхности. В частности, возможно уменьшение доли поверхности, доступной для протекания реакции CO + O_2, и увеличение энергетического барьера этой реакции. Это объясняет, почему алгоритм предпочел периодическое уменьшение потоков CO и O_2 в противофазе, при котором примесная фаза на поверхности не успевает сформироваться и разрушается в избытке кислорода, вследствие чего интегральный выход CO_2 увеличивается.

Полученные результаты согласуются с экспериментальными данными. Как было показано ранее [37–39], переход от статических условий к динамическим может значительно ускорить протекание реакции. Например, в работе [40], описывающей окисление CO на Pd/Al_2O_3, периодическое переключение подачи между CO/N_2 и O_2/N_2 позволило добиться усредненной скорости

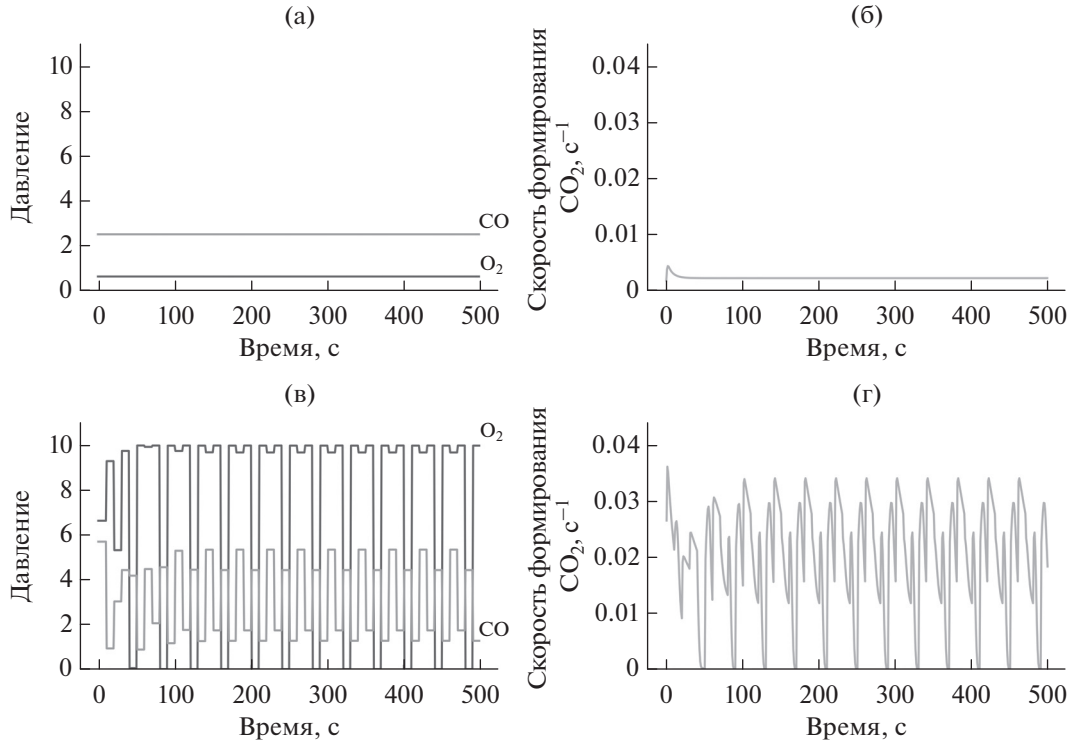

Рис. 4. Решение: а, б – стационарное, найденное методом оптимизации; в, г – динамическое, полученное с помощью алгоритма обучения с подкреплением. В обоих случаях использованы модели с учетом деградации поверхности катализатора. Показаны предсказанные политики (а, в) и выход CO_2 (б, г).

реакции, которая более чем в 40 раз превышала максимально достижимую скорость в стационарном режиме.

ЗАКЛЮЧЕНИЕ

Алгоритм обучения с подкреплением был применен для исследования пространства рабочих параметров реакции окисления CO. Агент градиента политики VPG получал на вход потоки CO, O_2 и CO_2 на текущем и предыдущем временны́х шагах и предсказывал оптимальные потоки CO и O_2 на следующем шаге. Обучение алгоритма проводилось на модели реакции окисления монооксида углерода на поверхности палладия как без учета, так и с учетом деградации поверхности в неравновесных потоках реагентов. В результате исследования были получены стационарные политики и политики периодического переключения. Максимальный выход продукта был достигнут при использовании модели с учетом деградации катализатора при периодическом изменении газовых потоков, обеспечивающих баланс между доступными адсорбционными участками и концентрацией активированных промежуточных продуктов. Продемонстрированный подход может быть расширен для оптимизации многих других промышленно значимых реакций и каталитических систем.

БЛАГОДАРНОСТИ

Исследование выполнено при финансовой поддержке Минобрнауки России (Соглашение № 075-15-2021-1363).

СПИСОК ЛИТЕРАТУРЫ

1. *Pakhare D., Spivey J.* // Chem. Soc. Rev. 2014. V. 43. № 22. P. 7813. https://doi.org/10.1039/C3CS60395D
2. *Pareek V., Bhargava A., Gupta R., Jain N., Panwar J.* // Adv. Sci. Eng. Med. 2017. V. 9. № 7. P. 527. https://doi.org/10.1166/asem.2017.2027
3. *Kinoshita K.* // J. Electrochem. Soc. 1990. V. 137. № 3. P. 845. https://doi.org/10.1149/1.2086566
4. *Rojluechai S., Chavadej S., Schwank J.W., Meeyoo V.* // Catal. Commun. 2007. V. 8. № 1. P. 57. https://doi.org/10.1016/j.catcom.2006.05.029
5. *DeSantis C.J., Peverly A.A., Peters D.G., Skrabalak S.E.* // Nano Lett. 2011. V. 11. № 5. P. 2164. https://doi.org/10.1021/nl200824p

6. *Sun C., Cao Z., Wang J., Lin L., Xie X.* // New J. Chem. 2019. V. 43. № 6. P. 2567. https://doi.org/10.1039/C8NJ05152F
7. *Vatti S.K., Ramaswamy K.K., Balasubramanaian V.* // J. Adv. Nanomat. 2017. V. 2. № 1. P. 127. https://doi.org/10.22606/jan.2017.22006
8. *Cuenya B.R.* // Thin Solid Films. 2010. V. 518. № 12. P. 3127. https://doi.org/10.1016/j.tsf.2010.01.018
9. *Schalow T., Brandt B., Laurin M., Schauermann S., Libuda J., Freund H.J.* // J. Catal. 2006. V. 242. № 1. P. 58. https://doi.org/10.1016/j.jcat.2006.05.021
10. *Skorynina A., Tereshchenko A., Usoltsev O., Bugaev A., Lomachenko K., Guda A., Groppo E., Pellegrini R., Lamberti C., Soldatov A.* // Rad. Phys. Chem. 2018. V. 175. № 1. P. 108079. https://doi.org/10.1016/j.radphyschem.2018.11.033
11. *Albers P., Pietsch J., Parker S.F.* // J. Mol. Catal. A. 2001. V. 173. № 1–2. P. 275. https://doi.org/10.1016/S1381-1169(01)00154-6
12. *Gromotka Z., Yablonsky G., Ostrovskii N., Constales D.* // Entropy. 2021. V. 23. № 7. P. 818. https://doi.org/10.3390/e23070818
13. *Armstrong C.D., Teixeira A.R.* // React. Chem. Eng. 2020. V. 5. № 12. P. 2185. https://doi.org/10.1039/D0RE00330A
14. *Cutlip M., Hawkins C., Mukesh D., Morton W., Kenney C.* // Chem. Eng. Commun. 1983. V. 22. № 5–6. P. 329. https://doi.org/10.1080/00986448308940066
15. *Vaporciyan G., Annapragada A., Gulari E.* // Chem. Eng. Sci. 1988. V. 43. № 11. P. 2957. https://doi.org/10.1016/0009-2509(88)80049-6
16. *Schwankner R., Eiswirth M., Möller P., Wetzl K., Ertl G.* // J. Chem. Phys. 1987. V. 87. № 1. P. 742. https://doi.org/10.1063/1.453572
17. *Eiswirth M., Ertl G.* // Phys. Rev. Lett. 1988. V. 60. № 15. P. 1526. https://doi.org/10.1103/PhysRevLett.60.1526
18. *Newton M.A., Ferri D., Smolentsev G., Marchionni V., Nachtegaal M.* // Nat. Commun. 2015. V. 6. № 1. P. 8675. https://doi.org/10.1038/ncomms9675
19. *Fang H., Haibin L., Zengli Z.* // Int. J. Chem. Eng. 2009. V. 2009. № 1. P. 710515. https://doi.org/10.1155/2009/710515
20. *Moghtaderi B.* // Energy Fuels. 2012. V. 26. № 1. P. 15. https://doi.org/10.1021/ef201303d
21. *Yoshida H., Kakei R., Fujiwara A., Tomita A., Miki T., Machida M.* // Top Catal. 2019. V. 62. № 1. P. 345. https://doi.org/10.1007/s11244-018-1100-5
22. *Toyao T., Maeno Z., Takakusagi S., Kamachi T., Takigawa I., Shimizu K.-I.* // ACS Catal. 2019. V. 10. № 3. P. 2260. https://doi.org/10.1021/acscatal.9b04186
23. *Segler M.H.S., Preuss M., Waller M.P.* // Nature. 2018. V. 555. № 7698. P. 604. https://doi.org/10.1038/nature25978
24. *Kaelbling L.P., Littman M.L., Moore A.W.* // J. Artif. Intell. Res. 1996. V. 4. № P. 237. https://doi.org/10.1613/jair.301
25. *Sutton R.S., Barto A.G.* Introduction to Reinforcement Learning. Cambridge: MIT Press, 1998. P. 380.
26. *Littman M.L.* // Nature. 2015. V. 521. № 7553. P. 445. https://doi.org/10.1038/nature14540
27. *Neumann M., Palkovits D.S.* // Ind. Eng. Chem. Res. 2022. V. 61. № 11. P. 3910. https://doi.org/10.1021/acs.iecr.1c04622
28. *Watkins C.J.* Learning from Delayed Rewards: PhD Thesis. Cambridge: King's Colledge, 1989. 242 p.
29. *Watkins C.J., Dayan P.* // Mach. Learn. 1992. V. 8. № 3. P. 279. https://doi.org/10.1007/BF00992698
30. *Lillicrap T.P., Hunt J.J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D.* Continuous Control with Deep Reinforcement Learning; https://arxiv.org/abs/1509.02971.pdf
31. *Alhazmi K., Albalawi F., Sarathy S.M.* // Chem. Eng. J. 2022. V. 428. №. P. 130993. https://doi.org/10.1016/j.cej.2021.130993
32. *Zhou Z., Li X., Zare R.N.* // ACS Cent. Sci. 2017. V. 3. № 12. P. 1337. https://doi.org/10.1021/acscentsci.7b00492
33. *Engel T., Ertl G.* // Elementary Steps in the Catalytic Oxidation of Carbon Monoxide on Platinum Metals. Munchen: Elsevier, 1979. P. 43.
34. *Chorkendorff I., Niemantsverdriet J.W.* // Concepts of Modern Catalysis and Kinetics. Weinheim: John Wiley & Sons, 2017. P. 66.
35. *Libuda J., Meusel I., Hoffmann J., Hartmann J., Piccolo L., Henry C., Freund H.-J.* // J. Chem. Phys. 2001. V. 114. № 10. P. 4669.
36. *Nelder J.A., Mead R.* // The Comput. J. 1965. V. 7. № 4. P. 308. https://doi.org/10.1093/comjnl/7.4.308
37. *Unni M., Hudgins R., Silveston P.* // Can. J. Chem. Eng. 1973. V. 51. № 6. P. 623. https://doi.org/10.1002/cjce.5450510601
38. *Abdul-Kareem H.K., Silveston P., Hudgins R.* // Chem. Eng. Sci. 1980. V. 35. № 10. P. 2077. https://doi.org/10.1016/0009-2509(80)85029-9
39. *Abdul-Kareem H.K., Hudgins R., Silveston P.* // Chem. Eng. Sci. 1980. V. 35. № 10. P. 2085. https://doi.org/10.1016/0009-2509(80)85030-5
40. *Zhou X., Barshad Y., Gulari E.* // Chem. Eng. Sci. 1986. V. 41. № 5. P. 1277. https://doi.org/10.1016/0009-2509(86)87100-7

Reaction of CO Oxidation on the Surface of Pd Nanoparticles: Optimization by Reinforcement Learning

M. S. Lifar[1, 2], A. A. Tereshchenko[1, *], A. N. Bulgakov[1, 2], A. A. Guda[1, **], S. A. Guda[1, 2], A. V. Soldatov[1]

[1]The Smart Materials Research Institute, Southern Federal University, Rostov-on-Don, 344090 Russia

[2]Vorovich Institute of Mathematics, Mechanics, and Computer Sciences, Southern Federal University, Rostov-on-Don, 344090 Russia

**e-mail: tereshch1@gmail.com*

***e-mail: guda@sfedu.ru*

The yield of reaction products depends on the interaction between processes on the catalyst surface: adsorption, activation, reaction, desorption, and others. These processes, in turn, depend on the magnitude of the flows of reaction mixtures, temperature, and pressure. Under stationary conditions, active sites on the surface can be poisoned by reaction by-products or blocked by an excess of adsorbed reactant molecules. Dynamic control of reaction parameters takes into account changes in surface properties and adjusts temperature, flow rates and other parameters accordingly. A reinforcement learning algorithm was applied to control the oxidation reaction of carbon monoxide CO on the surface of palladium nanoparticles. The algorithm was trained to maximize the rate of carbon dioxide production based on information about the magnitude of CO, O_2 and CO_2 fluxes at each time step. A gradient policy algorithm with a continuous action space was chosen, and observations of the flow rates were extended over several successive time steps, which made it possible to obtain a set of non-stationary solutions. The maximum yield of the product is achieved with a periodic change in gas flows, which ensures a balance between the available adsorption sites and the concentration of activated intermediates. This methodology opens up prospects for optimizing catalytic reactions under nonstationary conditions.

Keywords: machine learning, reinforcement learning, catalyst, palladium nanoparticles, adsorption, carbon monoxide.