Closed-loop, Human-AI–driven optimization of geological hydrogen production

Abhishek Soni[1], Merrill K. Chiang[1], Bachu Sravan Kumar[1], Nicky Evans[2,3], Jakob Nielsen[1], Haihui Lan[1], Tonio Buonasissi[2], Iwnetim I. Abate[1,†]

[1]Department of Materials Science and Engineering, MIT, Cambridge, MA 02139, USA
[2]Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA
[3]Research Laboratory for Electronics, MIT, Cambridge, MA 02139, USA
Correspondence to: Iwnetim Abate Email: iabate@mit.edu

1. Introduction

Geological hydrogen (H_2) produced from iron-rich rocks via serpentinization represents a promising, low-carbon alternative to fossil fuels.[1-3] However, natural serpentinization reaction rates are slow, poorly controlled, and difficult to optimize due to the coupling of multiple reaction parameters.[4-5] Here we report a closed-loop, high-throughput experimental framework that integrates domain knowledge with batch Bayesian optimization (BO) to accelerate geological H_2 production (Fig. 1). By systematically exploring a nine-dimensional parameter space, we achieved H_2 concentrations up to two orders of magnitude higher than baseline rock–water reactions (Fig. 2). Experiments using model brucite systems identify moderate temperatures, bicarbonate-rich fluids, and elevated catalyst and surfactant loadings as key drivers of enhanced hydrogen yield. Multimodal surface characterization reveals mechanistic links between the optimized conditions and accelerated redox Fe^{2+} to Fe^{3+} kinetics (Fig. 3). Together, these results demonstrate that AI-guided experimentation can accelerate the geological H_2 production and establish a scalable pathway for clean energy generation, using Earth as a chemical factory.

2. Substantial section

2.1 Related work

Serpentinization is a complex multi-step process[6,7] where the final step involves the redox reaction that creates natural hydrogen (Fig. 1A). It happened through the reduction of water upon oxidation of Fe^{2+}-containing brucite in ultramafic rocks (see Fig. 1 and equation 1)[8]:

$$2Fe(OH)_2\ (\text{brucite}) + 2H_2O \rightarrow Fe_3O_4\ (\text{magnetite}) + H_2 + 4H_2O \quad (1)$$

To optimize geological H_2 production in a controlled manner, we initiated our experiments using a brucite model system, $Fe(OH)_2$, which represents the final step in serpentinization (Equation 1) and forms via the redox reaction between aqueous $FeCl_2$ and NaOH. Previous study has quantified H_2 production in this model system primarily using design-of-experiments–based approaches.[8] In contrast, we systematically explored a nine-dimensional parameter space hypothesized to influence reaction kinetics, solute chemistry, and the reaction environment (see Table 1).

2.2 Figures and tables

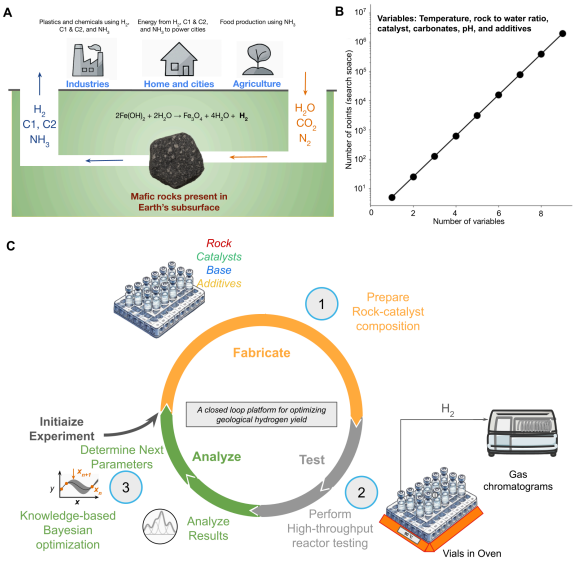


Fig. 1. Closed-loop, AI-driven optimization framework for geological hydrogen production, illustrating (A) serpentinization-enabled conversion of water and additives into value-added products, (B) the exponential growth of the multi-parameter experimental search space, and (C) the integrated high-throughput platform coupled with BO for accelerated discovery in a nine-dimensional chemical space.

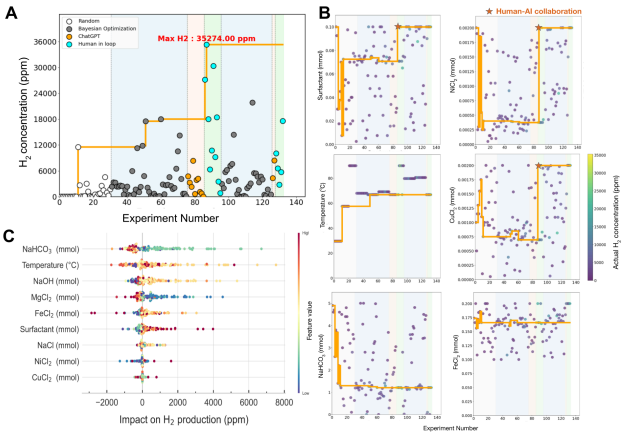


Fig. 2. (A) Human–AI closed-loop optimization rapidly enhances geological hydrogen production (100 times more than the control sample containing only $Fe(OH)_2$), showing accelerated yield improvement over successive experiments, (B) adaptive evolution of key reaction parameters, and (C) SHAP feature-importance analysis identifying dominant drivers of H_2 generation.

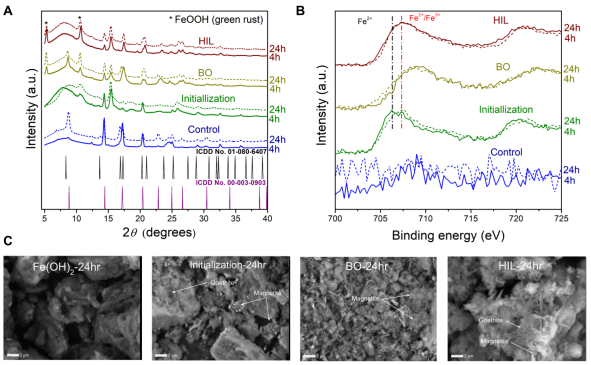


Fig. 3. (A) XRD, (B) XPS, and (C) SEM reveal accelerated iron redox kinetics and rapid formation of reactive FeOOH/magnetite-like phases under human–AI optimized conditions compared to control and non-optimized samples.

Table 1: Representative experimental conditions selected for mechanistic analysis, comparing human-in-the-loop (HIL), Bayesian-optimized (BO), and initialization (INIT) parameter sets against a control ($FeCl_2$, NaOH, and water only), with corresponding 24-hour H_2 production levels. The temperature is in Celsius while the other components are in millimoles.

Parameters	HIL (35,000 ppm)	BO (22,000 ppm)	INIT (12,000 ppm)
Temperature	66.8	80.5	57.5
$FeCl_2$	0.176	0.179	0.176
NaOH	1.155	1.452	1.155
$MgCl_2$	0.074	0.159	0.069
NaCl	4.9	5.3	4.8
$NiCl_2$	0.0020	0.0009	0.0004
$CuCl_2$	0.0020	0.0008	0.0008
$NaHCO_3$	1.21	1.23	1.30
Surfactant	0.100	0.088	0.073

2.3 Methodology and Results

Manual exploration of this nine-dimensional parameter space is infeasible. Even with five values per parameter, the combinations exceed 106 experiments to test (Fig. 1B). We addressed this challenge using our high throughput serum vial platform equipped with ML and batch-based BO (See Appendix A; Methods section). This enabled us to execute multi-dimensional experiments, analyze results, and select the next batch of experiments with a critical balance of human intervention.

Figure 2 demonstrates how initial random experiments established baseline model training, after which batch BO progressively increased H_2 concentration. Within fewer than 100 experiments, hydrogen concentrations improved stepwise, ultimately reaching a maximum of 35,274 ppm — over 100× higher than the rock–water baseline and more than threefold higher than the best initialization conditions (Fig. 2A).

While BO efficiently navigated much of the parameter space, additional human-in-the-loop (HIL) interventions proved critical for identifying underexplored regimes involving higher catalyst and surfactant loadings. These targeted adjustments enabled the discovery of the global optimum, illustrating how human chemical intuition and AI-driven exploration complement one another to accelerate convergence beyond either approach alone (Fig. 2B). To interpret the learned optimization landscape, SHAP feature-importance analysis revealed that bicarbonate concentration and temperature were the dominant drivers of hydrogen production, with peak performance occurring at moderate temperatures (~67 °C) and narrow $NaHCO_3$ optima (Fig. 2C). Secondary effects included beneficial low Mg^{2+} levels, sufficient $Fe(OH)_2$ precursor formation via balanced $FeCl_2$ and NaOH, and enhanced H_2 evolution under increased surfactant and catalyst concentrations.

Mechanistic characterization revealed that HIL conditions significantly accelerate iron redox kinetics (Fe^{2+} oxidation to Fe^{3+}) underlying hydrogen evolution (Fig. 3 and Table 1). XRD, XPS, and SEM show rapid transformation of $Fe(OH)_2$ into reactive FeOOH and magnetite-like phases within 4 hours under optimized conditions, compared to sluggish phase evolution in control and non-optimized samples (INIT and BO). These results confirm that AI-guided parameter discovery fundamentally alters reaction pathways to enhance hydrogen production.

3. Discussion and Outlook

We show that coupling domain knowledge with Bayesian optimization in a closed-loop, human–AI-guided framework enables rapid and systematic exploration of complex reaction spaces relevant to geological H_2 production. Our results highlight the necessity of data-driven, adaptive optimization strategies for geochemical systems. While the present work focused on synthetic analogs, future efforts will extend this platform to directly optimize H_2 generation from heterogeneous, compositionally complex real rock samples. In parallel, we will automate current manual steps—such as pipetting, solid handling, and sample transfer—which will enable the realization of a fully autonomous, first self-driving lab for accelerating geological H_2 production.

Acknowledgments

This work was supported by the MIT Climate Grand Challenges, ARPA-E and Bose Funding. This work was carried out in part through the use of MIT. Nano's facilities (which is supported by the National Science Foundation under award ECCS-1542152) and Department of Chemistry Instrumentation Facility (which is supported by the National Science Foundation under award HE-9808061, DBI-972959).

References

[1] Truche, L. et al. A deep reservoir for hydrogen drives intense degassing in the Bulqizë ophiolite. Science 383, 618–621 (2024).

[2] Ellis, G. S. & Gelman, S. E. Model predictions of global geologic hydrogen resources. Sci. Adv. 10, eado0955 (2024).

[3] Mayhew, L. E., Ellison, E. T., McCollom, T. M., Trainor, T. P. & Templeton, A. S. Hydrogen generation from low-temperature water–rock reactions. Nat. Geosci. 6, 478–484 (2013).

[4]. McCollom, T. M. et al. Hydrogen generation and iron partitioning during experimental serpentinization of an olivine–pyroxene mixture. Geochim. Cosmochim. Acta 282, 55–75 (2020).

[5]. Seyfried, W. E., Jr, Foustoukos, D. I. & Fu, Q. Redox evolution and mass transfer during serpentinization: An experimental and theoretical study at 200°C, 500bar with implications for ultramafic-hosted hydrothermal systems at Mid-Ocean Ridges. Geochim. Cosmochim. Acta 71, 3872–3886 (2007).

[6] Srinivasan, P. et al. Comprehensive review of serpentinization and the experimental generation of hydrogen. in Middle East Oil, Gas and Geosciences Show (MEOS GEO) (SPE, 2025). doi:10.2118/226968-ms.

[7] Hurowitz, J. A. et al. Redox-driven mineral and organic associations in Jezero Crater, Mars. Nature 645, 332–340 (2025).

[8] Song, H. et al. An overlooked natural hydrogen evolution pathway: Ni^{2+} boosting H_2O reduction by Fe(OH)2 oxidation during low-temperature serpentinization. Angew. Chem. Int. Ed Engl. 60, 24054–24058 (2021).

Appendix A. Active Learning Methodology

Bayesian optimization campaign to enhance geological hydrogen concentration

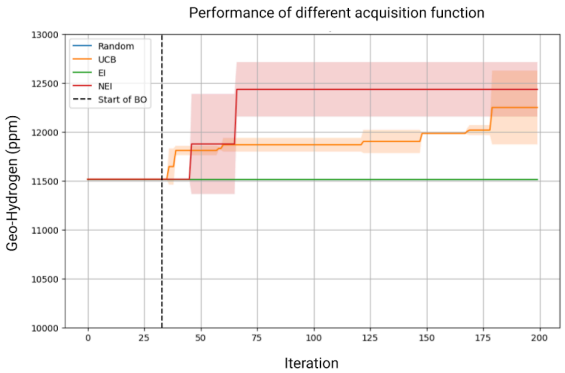
To choose the next experiment after the initialization, we employed a BO algorithm that maximized the hydrogen concentration. The BO was performed using the Botorch Python package(cite) and happened in two steps. First, a surrogate Gaussian process regression model was built over all existing data to predict the hydrogen concentration of potential formulations using thermochemical measurements from a limited dataset. The data covariates were normalized to the unit cube and outcomes were standardized (zero mean, unit variance). The surrogate model predicts the experimental output (hydrogen concentration) based on inputs such as temperature, catalyst and additive amounts. We implemented the Gaussian process model using the GPyTorch Python package.[A1]

A single-task Gaussian process model was used for the surrogate model.[A2] This algorithm learns noise actively. All other hyperparameters for the Gaussian process model were defaulted.

Next, the best acquisition function was selected based on the simulation of oracle surface results (Supplementary Figure S1), and a new set of experimental conditions were acquired. We selected a q-noisy expected improvement (qNEI) function (Supplementary Figure S1). All the sampling strategies were configured to maximize the hydrogen concentration while also effectively searching the parameter space. The qNEI function helps manage the balance between exploration (probing new regions of the parameter space) and exploitation (refining existing knowledge where improvements are deemed likely) during BO. This is achieved by modulating the selection of sampling points based on both the predicted improvement and the associated uncertainty, thus mitigating the risk of redundantly sampling the same points. The inclusion of the 'Noisy' term accounts for the inherent variability and uncertainty in the experimental measurements, a feature not typically addressed by the standard expected improvement and upper confidence bound functions (see Supplementary Fig. S1). This feature allowed the algorithm to better navigate the trade-off between exploring

new, uncertain regions and exploiting areas known to yield high outcomes, ensuring robustness against experimental noise and reducing the likelihood of repetitive sampling in less promising areas.

In order to ensure a comprehensive exploration of the parameter space and test a few human-driven hypotheses, the results from some HIL points and ChatGPT suggested points were introduced to the optimizer (Fig. 2). This strategy enhanced exploration by testing a few hypotheses and reducing the risk of entrapment in local optima. This ensured a comprehensive search of the parameter space. Once a set of experimental conditions was identified using the qNEI algorithm, the optimizer provided a set of parameters for new experimental conditions to be tested through the high throughput platform. The closed loop experiments were repeated until no improvement in the objective function was observed as shown in Fig. 2A.



Supplementary Fig. S1. Comparison of Bayesian optimization acquisition functions against random (Sobol) sampling. The figure shows the best observed simulated geo-hydrogen concentration (ppm) as a function of iteration for random Sobol sampling and Bayesian optimization (BO) using Upper Confidence Bound (UCB), Expected Improvement (EI), and q-Noisy Expected Improvement (q-NEI) acquisition functions. The dashed vertical line indicates the transition from initial random sampling (with the real experimental initialization data) to BO. Shaded regions denote variability across runs. Among the methods evaluated, q-NEI demonstrates faster convergence and achieves the highest geo-hydrogen concentrations with less number of iterations, outperforming both alternative acquisition functions and random sampling.

Appendix References

[A1] Chang, D. T. Bayesian Hyperparameter Optimization with BoTorch, GPyTorch and Ax. arXiv [cs.LG] (2019).

[A2] Rasmussen, C. E. & Williams, C. K. I. Gaussian Processes for Machine Learning. (MIT Press, London, England, 2019). doi:10.7551/mitpress/3206.001.0001.