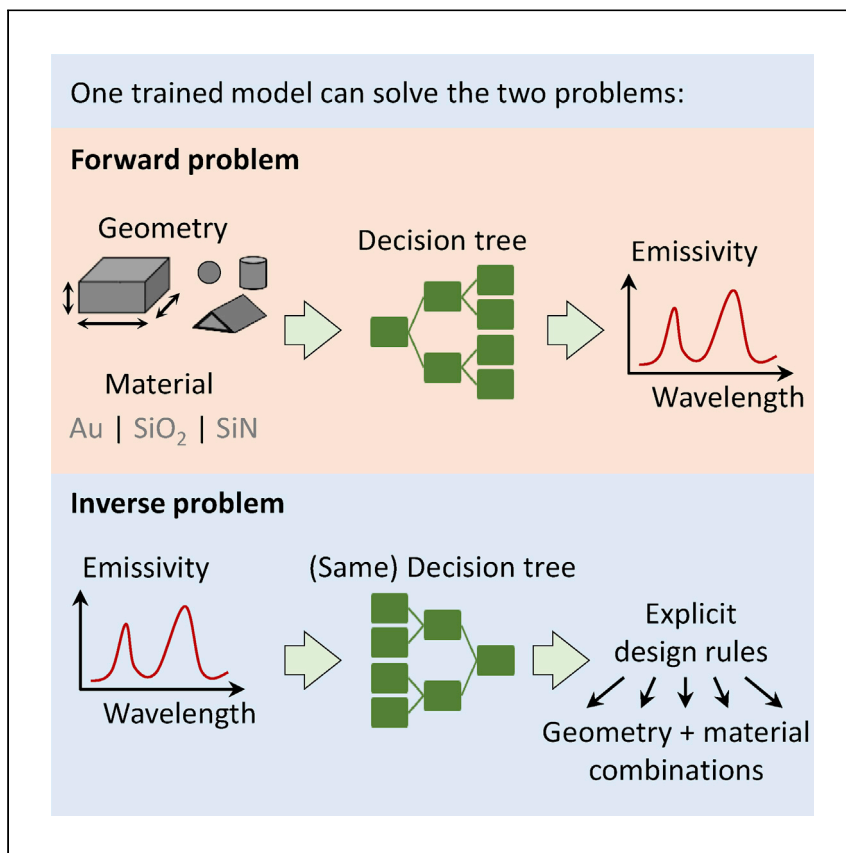**Article**

Interpretable Forward and Inverse Design of Particle Spectral Emissivity Using Common Machine-Learning Models

Inverse design is usually done by expensive, slow, iterative optimization. Elzouka et al. show how a simple machine-learning model (decision trees) can efficiently perform inverse design in one shot, while recovering design rules understandable by humans. Inverse design of the spectral emissivity of particles is used as an example.

Mahmoud Elzouka, Charles Yang, Adrian Albert, Ravi S. Prasher, Sean D. Lubner

rsprasher@lbl.gov (R.S.P.)
slubner@lbl.gov (S.D.L.)

HIGHLIGHTS

Radiative particles are ubiquitous in nature, and energy and biomedical technologies

Particle spectral emissivity is modeled and investigated using simple decision trees

The same trained model can solve both the forward and inverse problems

The highly interpretable models return explicit design rules for the inverse problem

**Cell Reports
Physical Science**

CellPress
OPEN ACCESS

Article

Interpretable Forward and Inverse Design of Particle Spectral Emissivity Using Common Machine-Learning Models

Mahmoud Elzouka,[1] Charles Yang,[1,2] Adrian Albert,[1] Ravi S. Prasher,[1,3,4,*] and Sean D. Lubner[1,*]

SUMMARY

Radiative particles are ubiquitous in nature and in various technologies. Calculating radiative properties from known geometry and designs can be computationally expensive, and trying to invert the problem to come up with designs specific to desired radiative properties is even more challenging. Here, we report a machine-learning (ML)-based method for both the forward and inverse problem for dielectric and metallic particles. Our decision-tree-based model is able to provide explicit design rules for inverse problems. Furthermore, we can use the same trained model for both the forward and the inverse problem, which greatly simplifies the computation. Our methodology shows the promise of augmenting optical design optimizations by providing interpretable and actionable design rules for rapidly finding approximate solutions for the inverse design problem.

INTRODUCTION

Controlling light-matter interactions is central to a variety of important applications such as energy harvesting,[1,2] solar-thermal desalination,[3] radiative cooling,[4,5] heating,[6] and computing.[7–9] In particular, the interaction of light with particles is ubiquitous throughout technological applications and the natural world. For example, in combustion, flame irradiance depends on soot formation. Iron oxide nanoparticles are designed to couple strongly to radio frequency photons for targeted thermal biomedical therapies,[10] while other particles can be used for optically triggered targeted drug delivery.[11] Quantum dots can be used as light emitters[12] and sensors.[13] Particle emulsions and mixtures can be used as radiative composite materials[5] and spectrally selective paints.[14] In the natural world sunlight is scattered and absorbed by raindrops, aerosols, and other particulates in our atmosphere such as carbon black, which affects the solar albedo leading to detrimental effects on climate change.[15]

Calculating these optical properties from known particle geometries and designs can be computationally expensive,[16] and trying to invert the problem to come up with a design that produces desired optical properties generally amounts to a nonlinear one-to-many problem that is very difficult to solve.[17,18] Many researchers have recently been turning to machine learning (ML) to speed up these calculations.[19–23] In this fast-emerging field of ML-accelerated optical properties calculations, the current status is as follows:

(1) Most of the geometries are surface-based geometries. Very few studies have been conducted on radiative properties of particles. Peurifoy et al.[19]

[1]Energy Storage & Distributed Resources Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[2]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA

[3]Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA 94720, USA

[4]Lead Contact

*Correspondence: rsprasher@lbl.gov (R.S.P.), slubner@lbl.gov (S.D.L.)

https://doi.org/10.1016/j.xcrp.2020.100259

conducted ML-based inverse design of core/shell spherical particles. This was a good step forward in the ML-based design of particles; however, spheres are the simplest geometry, and these capabilities must be extended to more general geometries in order to be relevant for most applications. Solving both the forward and the inverse problem for non-spherical geometries is significantly more complex, and of practical interest,[24–26] especially to better describe particles in nature, like dust.[27]

(2) A popular ML model is artificial neural networks, which have shown great promise in inverse design.[28,29] Significant work has been done to improve and understand the features learned by such models, and both the qualitative and quantitative relationships between those features and the model predictions or properties of interest.[30,31] These approaches often use gradient-based methods to evaluate the local impact of input features on outputs and properties.

(3) Separate models are typically required for the forward and direct inverse problems, complicating interpretability and inefficiently decoupling the physics that must be relearned in both cases.

In this paper, we propose a training methodology that results in a single decision tree that can rapidly solve both the forward and inverse problems with high accuracy for the spectral optical properties of diverse 3D metallic and dielectric particles. In addition to standard feature importance interpretability, this model also automatically provides learned global design rules for inverse design. We train our model on a single dataset of solid particles. The dataset consists of spectral emissivity curves numerically computed for 15,900 particles of varying shapes that can be experimentally fabricated (spheres, parallelepipeds,[32] triangular prisms,[33] and cylinders[34]), aspect ratios, sizes (nanometers to tens of microns; spanning sub-wavelength to super-wavelength regimes), and materials (SiO_2, SiN, and Au), as summarized in Figures S2–S5. Our model is able to efficiently learn the underlying unknown emissivity function, as a function of six input feature parameters as discussed later. Our training dataset samples this function with an average sampling density of 4 points per parameter, holding all other parameters constant. Emissivity is calculated by analytical solution and direct numerical simulation (DNS) of Maxwell's equations (see Experimental Procedures). According to Kirchhoff's law, the emissivity model we have built is equally valid for absorptivity, which is very important for describing natural radiation phenomena and its impact on climate change.

RESULTS AND DISCUSSION

Training of the ML Model

We display our entire dataset in Figure 1, where every point is a unique particle whose optical properties were calculated via DNS. Emissivities of simulated particles cover a wide range, from 0.001 up to almost 10 (Figure S1). Although it is impossible for emissivity to exceed unity for large surfaces, it is possible for finite particles with dimensions smaller than the wavelength of radiation. This phenomenon is well established in classic radiation textbooks (e.g., by Bohren and Huffman[35]) and is known in modern literature as Super-Planckian radiation and have been investigated theoretically[36,37] and experimentally.[38] The term emissivity we use in this paper is a synonym for emission/absorption efficiency, which is the ratio between the absorption cross-section averaged over the entire solid angle and the particle surface area. Bohren[39] has shown a good graphical illustration for absorption efficiency exceeding unity.

CellPress
OPEN ACCESS
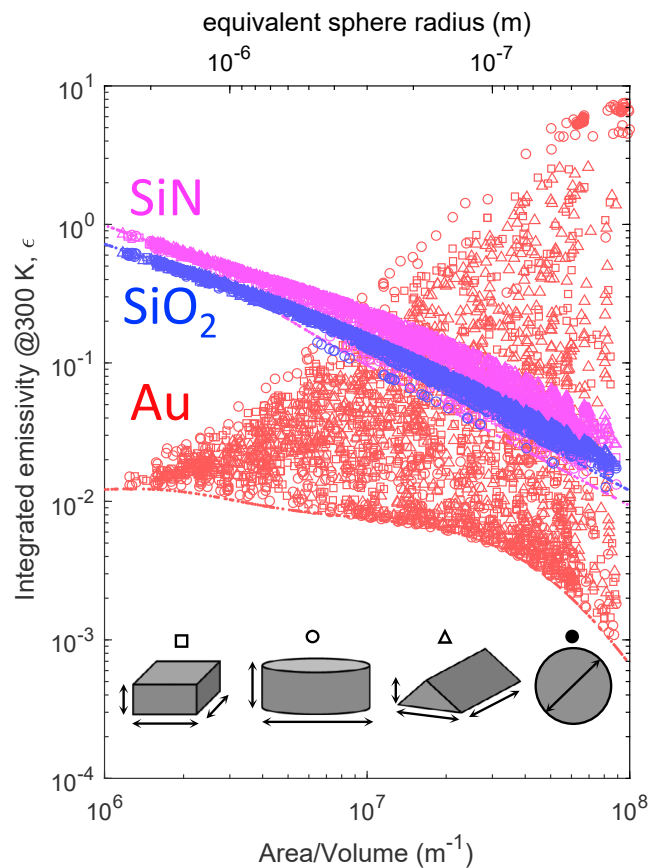
Figure 1. Integrated Thermal Emissivity for the Training Dataset
Spectrally and hemispherically integrated emissivity at 300 K for every particle in our training dataset as a function of particle area-to-volume ratio. Colors and symbol shapes represent different materials and geometries, respectively. Red, blue, and pink represent Au, SiO_2, and SiN, respectively.

We select decision trees and random forests (RF, ensembles of decision trees)[40] as our class of models. These models average and mix data during the training process so that they can predict emissivity spectra that they have never seen during training. We wish for our model to be highly interpretable, and for the same model to solve both the forward problem (i.e., predict optical properties from a given design) and the inverse problem (i.e., predict designs to produce desired optical properties). In general a random forest offers superior model accuracy and robustness as compared to a decision tree,[40] but it forfeits the direct interpretability of a decision tree. Additionally, while it is possible to retrace back up a decision tree to perform inverse design, this is not possible for a random forest because its output is an averaged ensemble vote. We wish to produce a decision tree that embodies the performance metrics of a random forest. To accomplish this, we apply a combined multiple models (CMM) method,[41] which compresses ensemble-based models (e.g., RF) into a single base model (e.g., DT) without significantly affecting model performance (see Experimental Procedures). The use of the CMM algorithm has been largely limited because it requires problems for which it is cheap to generate large amounts of synthetic unlabeled training data. For instance, it is difficult to programmatically generate images of faces, dogs, or cats (canonical computer vision datasets). However, in our inverse design problem we can easily generate a diverse set of particles

CellPress
OPEN ACCESS

Cell Reports
Physical Science
Article

simply by randomly sampling over the ranges of (self-consistent) material and geometric parameters.

We train two separate models using the same dataset based on which optical properties we wish to target. One model targets the full emissivity spectrum (array target; emissivity as a function of wavelength), and the other model targets the spectrally integrated emissivity (scalar target; integrated average of the emissivity spectrum, weighted by the blackbody distribution at 300 K); see Experimental Procedures for complete details. Training each model is a two-step process: first, we train a random forest on our data, and then we use that random forest to train a large decision tree to emulate the performance of the random forest (summarized in Figure S14). The end result is a decision tree with the performance of a random forest, which can be used for both the forward and inverse problems. During the first step of training, each particle in the training set is presented to the random forest as a length-6 array of parameterized geometric features and a material type. By carefully choosing physics-relevant parameterizations of the particles, such as the area-to-volume ratio and the largest particle dimension along a principal axis, we enable our model to efficiently learn particle emissivity distributions without requiring excess data. We use a 50/50 test-train split (i.e., train using 7,950 particles and test on the remaining different 7,950 particles). For the second step of training, we generate roughly 2,000,000 random, unlabeled particle designs. These are generated by randomly selecting self-consistent material type and (logarithmically distributed) geometric parameter values within the ranges spanned by the original dataset. We use the trained random forest to label these synthetic data by predicting their scalar or spectral emissivities. We then train a new decision tree on this much larger generated synthetic dataset. We call this new decision tree DTGEN because it is trained on generated data. DTGEN is our final trained model. DTGEN emulates the superior performance of a random forest while preserving the interpretability and retracability of a regular decision tree. Thus, even though DTGEN is trained for inference (solving the forward problem), it can immediately be used to also solve the inverse problem too. DTGEN takes approximately 11 (135) CPU-ms/sample to train and 0.005 (1.6) CPU-ms/sample to approximately solve the forward problem for scalar (spectral) targets, respectively (see Experimental Procedures). In contrast, DNS takes approximately 12 CPU hours to exactly solve the forward problem for one sample.

The models' performances on solving the forward problem for the test dataset are presented for scalar (Figures 2A and S6) and array (Figures 2B and S7) targets. The model errors (Figure 2B insets) are always below 10%, and generally below 5%. Higher error for Au compared to SiO_2 and SiN is consistent with the greater diversity of optical interactions that can occur on metallic particles, as can be seen in Figure 1. The model can learn particle emissivity functions for dielectrics faster than for metals (Figure 2C) because metals can be more optically expressive. High symmetry shapes (spheres—1 degree of freedom) are learned faster than lower symmetry shapes (cylinders—2 degrees of freedom; parallelepipeds and triangular prisms—3 degrees of freedom). This difference is most pronounced for metals, again due to their higher expressiveness.

Model feature importance analysis (Figure 2D) shows that surface area-to-volume ratio is a more important feature for dielectrics than metals. Consequently Figure 1, which plots emissivity versus particle area-to-volume ratio (which represents the inverse of the characteristic particle size), shows a strong trend with the surface area-to-volume ratio for dielectrics, but not for metals. This trend can be understood as
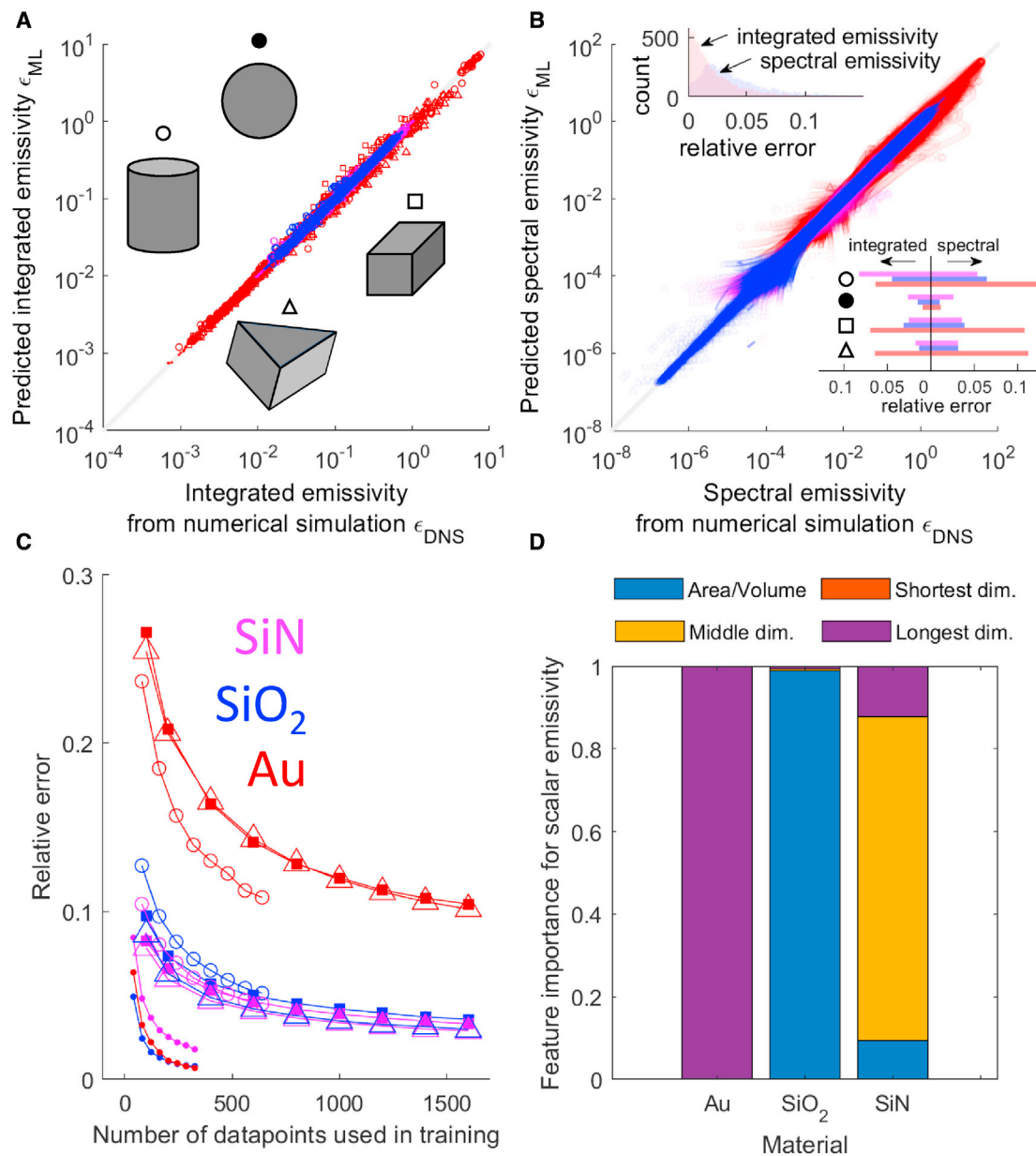
**Figure 2. Model Training, Interpretation, and Inference Performance**

Colors and symbol shapes represent different materials and geometries, respectively (for A–C, red, blue and pink represent Au, SiO₂, and SiN, respectively).

(A and B) Machine learning predicted (ε_{ML}) compared to DNS results (ε_{DNS}) for integrated emissivity (A) and spectral emissivity (B). Top inset in (B) shows relative inference error (see Experimental Procedures); bottom inset shows relative error by material and geometry type, averaged over 100 different trained models with different random test-train splits.

(C) Learning curves stratified by material and geometry type. Each point represents the average of 100 independent training runs, with random test-train splits.

(D) Relative feature importance analysis.

being due to the volumetric nature of photon absorption and emission for dielectric materials, which exhibit low attenuation.[42] SiN deviates from this trend more than SiO2, because its attenuation is slightly higher. On the other hand, the emissivity of metals has no clear correlation with particle size other than increased emissivity variability for smaller particles. This is explained by the dominating effects of surface
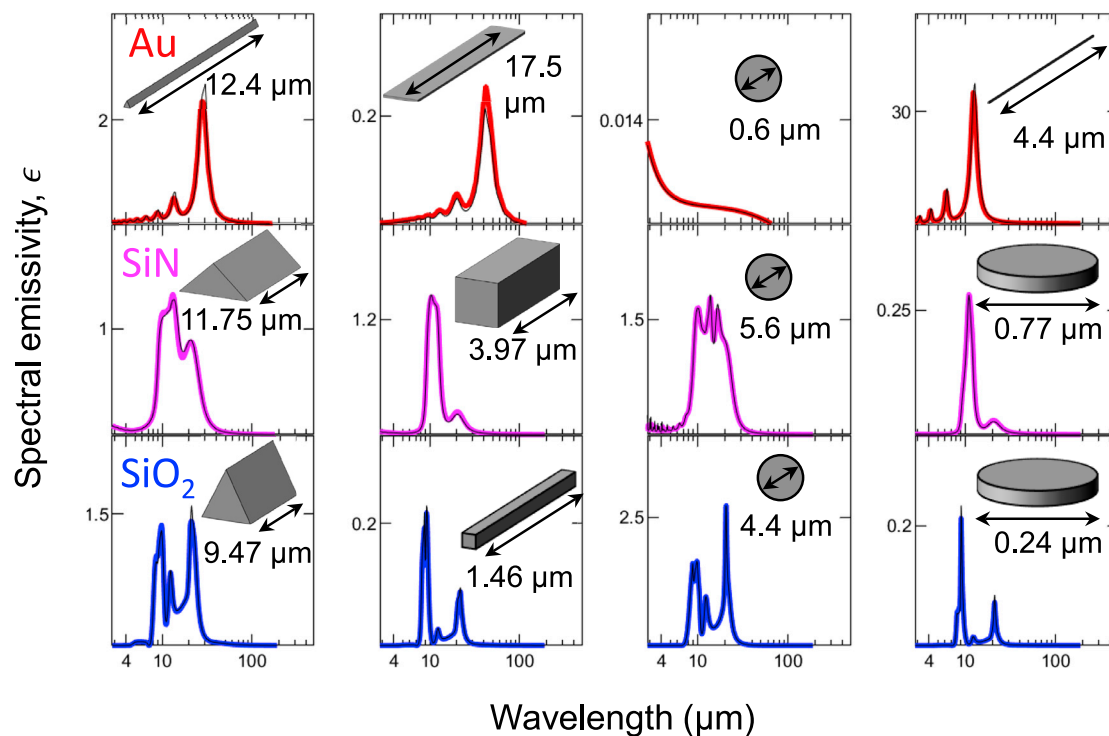
Figure 3. Model Predictions versus Numerical Simulations for Spectral Emissivity
Model predictions (colored lines) compared to ground truth (black lines; from DNS) of spectral emissivity for particles of different shapes (columns) and materials (rows). Color conventions are the same as in previous figures. These examples were chosen from among the worst scoring 20% of predictions from the test set, as shown in Figures S6 and S7.

and localized electromagnetic modes that can be supported by small metal particles (e.g., localized plasmons), which can significantly influence the emissivity. These modes depend on surface geometry more so than the overall particle size, as indicated in Figure 2D by the larger importance in metals of the particle's longest dimension. See Experimental Procedures for a description of how feature importance was extracted by material.

We also tried fitting a variety of single- and multi-variable linear models to the data (Figures S15 and S16). In all cases, the linear model predictions had relative errors roughly one order of magnitude (absolute) worse than equivalent tree-based models. This implies that although most of the data variation can be attributed to one or two of the most important input features for a given material type, the relationship between those features and emissivity is likely quite non-linear, necessitating more complex models.

Emissivity spectra predictions for individual particles of each type are shown in Figure 3. These spectra were chosen from among the worst 20% of predictions from the test set (as measured by relative error) for each material and geometry class combination (more spectra are shown for gold in Figure S8). A single DTGEN model accurately solves the forward problem across all geometries and materials. For metals, the peak emissivity typically occurs at a wavelength roughly twice the longest dimension of the particle. For dielectrics, the wavelength of the emissivity peak is almost independent of the longest dimension and usually associated with materials' absorption bands. This is again explained by the different mechanisms of emission for dielectrics and metals and indicated in Figure 2D.

Cell Reports
Physical Science
Article

CellPress
OPEN ACCESS
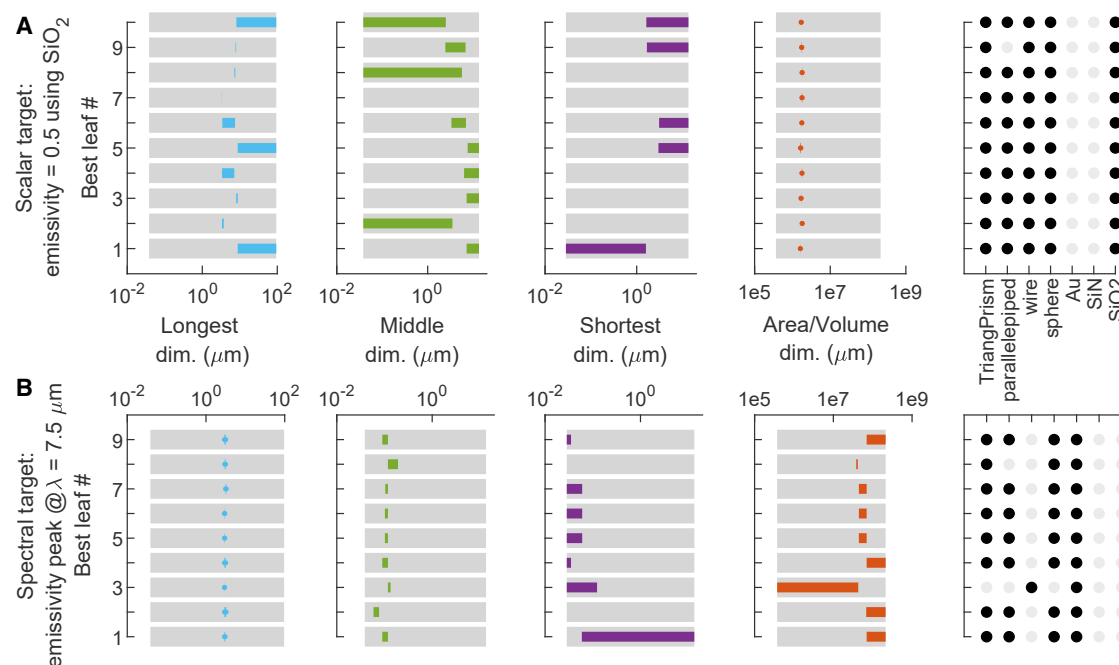
Figure 4. Design Rules for Scalar and Spectral Inverse Design

(A and B) Grayed-out regions show the range of each feature spanned by the full dataset. Colored or filled-in regions show the allowed range of each feature, as predicted by the algorithm, to produce a design that satisfies the target optical properties. Colored region is absent if the algorithm has no restriction on the feature. This figure shows for the specific case of scalar emissivity = 0.5 for SiO_2 (A) and peak emissivity at 7.5 μm (B). The same methodology is followed for inverse design for other targets. Using DNS, we have calculated the emissivities of samples generated using the design rules given in (A) leaf #1 and (B) leaf #1 and compared these results against their corresponding emissivity targets. These results are given in Figures 5A and 5D, respectively.

Inverse Design for Scalar and Spectral Emissivity Targets

To perform inverse design, we find the output label ("leaf") on DTGEN whose value (can be a scalar or an array) corresponds closest to the desired optical properties we want our particle to produce. We trace up the decision tree branch from this leaf taking the intersection of all branch-splitting criteria on all particle features encountered along the way. The result is a set of design rules for each feature (Figure 4). Having a set of design rules naturally captures the one-to-many mapping behavior of inverse design problems. We randomly sample self-consistent particle designs from these design rules and calculate their true optical properties using the same DNS scheme. The optical properties of these generated designs are then compared to the original target optical properties.

The performance of these generated designs for integrated emissivity (scalar) targets is shown in Figure 5A. To demonstrate model flexibility, we required generation of particle designs using each of the three materials for those targets satisfying $\epsilon < 1$ (because generally only metal structures have apparent emittance greater than one[39,43,44]). The design rules used to generate the SiO_2 $\epsilon = 0.5$ designs are shown in Figure 4A (leaf #1), along with the 9 next-closest leaves to the target. Because emission for dielectrics is primarily a volumetric process and does not depend on surface modes, the generated design rules are highly restrictive on the area-to-volume ratio in order to meet the target, while allowing considerable flexibility for other geometric parameters, consistent with Figure 2D showing "area/volume" being the most important feature for SiO_2. The performance of representative generated designs for spectral emissivity are shown for Au (Figure 5B) and SiO_2 (Figure 5C).
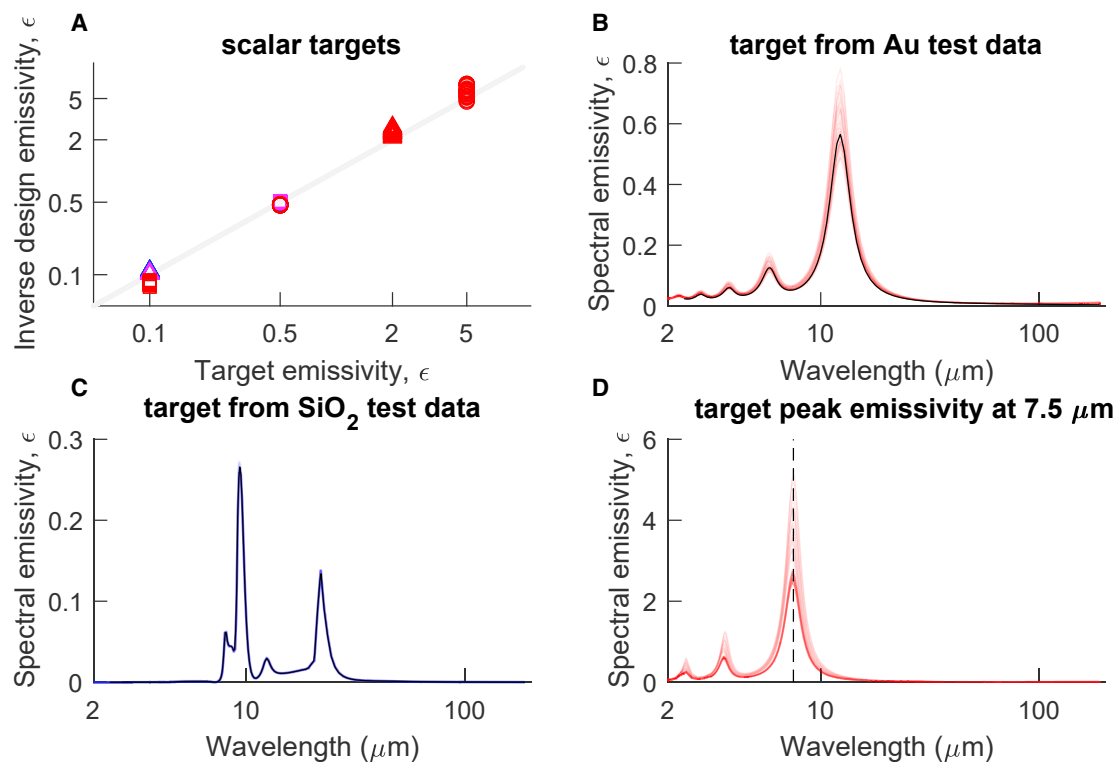
Figure 5. Inverse Design Solutions for Scalar and Spectral Target Emissivities

Black lines indicate the target optical properties and colored lines are the true optical properties of the model-generated designs, calculated via DNS. Color refer to the material, and it is consistent with previous figures.

(A) Performance of 53, 64, and 182 different designs generated to produce desired scalar emissivity targets for SiO_2, SiN, and Au, respectively. The SiO_2 ε = 0.5 designs were generated using the design rules from Figure 4A leaf #1.

(B and C) Performance of 51 Au designs generated to produce desired spectral emissivity target (B), and of 15 SiO_2 designs for a different spectral emissivity target (C). Targets randomly selected from test set. Lines overlap nearly perfectly for dielectrics.

(D) Performance of 62 designs generated to produce spectral emissivity with its peak at a target wavelength of 7.5 μm, using design rules from Figure 4B leaf #1.

Figure 5D shows the performance of designs generated to satisfy more general criteria of having the emissivity peak occur at a desired wavelength, with corresponding design rules given in Figure 4B. This time the design rules require the material to be Au and are highly restrictive on the longest dimension while being loose on other parameters, because the emissivity peak wavelength is roughly twice the longest particle dimension for Au due to standing plasmon waves, consistent with Figure 2D showing "longest dimension" being the most important feature for Au. All generated designs satisfy the target optical properties to high accuracy.

In conclusion, we have presented a decision-tree-based method for creating an interpretable ML model that can rapidly solve both the forward and inverse design problem and automatically provide intuitive design rules for generating particles with desired emissivity. We have applied this model to the realm of particle emissivity. The design space includes diverse particles, varying over shape, size, aspect ratio, and material type. A single model is able to efficiently learn the underlying emissivity function, valid for all materials and particle types spanned by the dataset. The interpretability of our models and the design rules they generate reconfirm that dielectric particles emit and absorb electromagnetic radiation mostly volumetrically, while metallic particles' interaction with light is dominated by surface modes and

depends more on the longest particle dimension than on the overall characteristic particle size. The design rules naturally capture the one-to-many mapping of the inverse design problems, allowing some flexibility to facilitate design constraints. Future work should focus on experimental validation of generated designs, extending the framework to be more generalizable, and incorporating design constraints. Our approach offers an exciting avenue for the inexpensive and interpretable discovery of novel optical metamaterial designs.

EXPERIMENTAL PROCEDURES

Resource Availability

Lead Contact
Further information and requests should be directed to and will be fulfilled by the lead contact, Ravi S. Prasher (rsprasher@lbl.gov).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
The code and data to reproduce the main results of this paper can be found at: https://github.com/mhmodzoka/DT_inverse_design.

Numerical Emissivity Calculation
Hemispherical spectrally integrated emissivity and spectral emissivity of a finite particle with arbitrary shape is calculated by solving Maxwell's equations numerically. We use the fluctuating-surface-current formulation with the boundary-element method for its efficiency in directly calculating the integrated thermal radiation over all radiation directions in a single step.[45] In this method, the radiation emitted by the particle is calculated as a result of surface current fluctuations, which represent the random thermal motion of charges within the material. This method is implemented by a free and open-source software, SCUFF-EM.[46] In the boundary-element method, we create a surface mesh for the interfaces between any two distinct media, which in our case is the interface between the particle material and the surrounding vacuum. Fluctuating surface current is assumed at each mesh point, which is the source of thermal electromagnetic radiation. The radiated power at any point in space is calculated as the magnitude of the Poynting vector, which is then integrated over the entire surface of the finite particle to calculate the total radiated power. Thermal emissivity of the particle is defined as the ratio between the calculated total radiated power and the radiated power from a hypothetical blackbody at the same temperature multiplied by the same projected geometrical surface area as the particle of interest. We have created the surface meshes using gmsh,[47] an open-source software. We created the meshes to have facets that are small compared to the shortest wavelength (1.8 μm). We display the distribution of the mesh edge sizes in Figure S9, which shows a mean value of 0.1632 μm for the square root of the surface area of the mesh surface facets. This is smaller than 10% of the shortest wavelength involved in our simulations. We performed a mesh convergence test to ensure the meshes were fine enough to achieve accurate results. Dielectric constants for Au[48], SiN,[49] and SiO2[50] were adopted from the literature.

Analytical Emissivity Calculation
For validating numerical simulations, we have compared numerical results to analytical ones for simple geometries, like infinitely wide thin films, infinitely long cylinders, and spheres. These geometries can be described using a single dimension in cartesian, cylindrical, and spherical coordinates, respectively, due to their high symmetry.

The spectral emissivities for these 1D geometries can be calculated analytically without the need for numerical solutions. For infinitely wide thin film, we used the transfer matrix method[51] to calculate the reflectance (R) and transmittance (T) of a single thin film, averaged between the two light polarizations (i.e., transverse electric and magnetic polarizations).[52] Thermal emissivity is assumed to be equal to the absorbance of the film (A = 1 – R – T), based on Kirchoff's law. Hemispherical emissivity for a thin film can be calculated from Equation 1, where θ is the angle between the incidence angle and the normal to the film surface.

$$\epsilon(\omega) = \frac{\int_{\theta=0}^{\pi} \epsilon(\theta, \omega)\sin \theta \cos \theta d\theta}{\int_{\theta=0}^{\pi} \sin \theta \cos \theta d\theta}. \qquad \text{(Equation 1)}$$

The emissivities of a sphere and an infinitely long cylinder were calculated analytically using Mie theory.[35,53] Sphere hemispherical emissivity is calculated as the ratio between the absorption cross-section and the sphere projected cross-section,[54] as shown in Equation 2.

$$\epsilon_{sphere}(\omega) = \frac{\sigma_{abs,sphere}(\omega)}{\pi r^2}. \qquad \text{(Equation 2)}$$

Sphere emissivity is independent of angle and light polarization due to the symmetry of the sphere. Cylinder directional emissivity is also calculated as the ratio between the absorption cross-section and the cylinder projected cross-section at a given angle (as shown in Equation 3), averaged between the two distinct light polarizations (i.e., electric field parallel or perpendicular to the plane formed by the cylinder axis and the radiation direction)[42]:

$$\epsilon_{cyl}(\omega, \theta) = \frac{\sigma_{abs,cyl}(\omega, \theta)}{2r \sin \theta}, \qquad \text{(Equation 3)}$$

where $\sigma_{abs,cyl}$ is the absorption cross-section of the infinite cylinder per unit length, and θ is the angle between the radiation direction and the cylinder axis. Hemispherical cylinder emissivity can be calculated from the directional emissivity from Equation 4.

$$\epsilon_{cyl}(\omega) = \frac{\int_{\theta=0}^{\frac{\pi}{2}} \epsilon_{cyl}(\omega, \theta)\, r \sin^2 \theta\, d\theta}{\int_{\theta=0}^{\frac{\pi}{2}} r \sin^2 \theta\, d\theta}. \qquad \text{(Equation 4)}$$

Absorption cross-sections for the sphere and cylinder were calculated as the difference between extinction and scattering cross-sections.

Dataset Description

We generated the majority of the dataset by random uniform sampling over the geometric parameters describing the particle for each material and geometry class: area-to-volume ratio (A/V), the shortest dimension (ShortDim), middle dimension (MiddleDim), and the longest dimension (LongDim). The range for A/V spanned from $10^6\ \mathrm{m}^{-1}$ to $10^8\ \mathrm{m}^{-1}$. We generated 15,900 data points including three materials: gold (Au), silicon dioxide (SiO_2), and silicon nitride (SiN); four geometry classes: sphere, cylinder, parallelepiped, and triangular prism; with 500, 800, 2,000, and 2,000 data points, respectively, for each material. We have shown the full distributions of geometry parameters in Figures S2–S5.

ML Models

Random forests and decision trees were implemented in Scikit-learn[55] (version 0.22.2) with the default hyperparameters, unless otherwise specified. Random forests were trained using 200 decision tree estimators, which is a sufficiently large number to ensure good performance (impact of the number of decision tree estimators on the model performance is illustrated in Figure S12 and begins to saturate around 20 estimators). Each

decision tree estimator was trained using all the training data available for the random forest regressor. We have used the mean square error as the criterion for the decision tree estimator to determine the quality of the split while training. All experiments used a 50/50 test-train split ratio, unless specified otherwise. We performed 100 random test-train splits with these ratios to provide estimates of the models' average performances.

During training, each particle is presented to the model as a length-6 array of parameterized geometric features and a material type. In particular, these 6 features are: one-hot encodings of the geometry class and of the material type, the area-to-volume ratio, and the mutually orthogonal shortest, middle, and longest dimensions of the particle measured along edges of the smallest bounding box, as illustrated by arrows displayed on particle cartoons in Figure 1. Each training data point is therefore a length-6 input array of descriptor features with a corresponding output scalar value (spectrally integrated emissivity) or output array (emissivity spectrum spanning near to far infrared). We generated a synthetic (random forest-labeled) dataset of $250 \times 7950 = 1,987,500$ data points and used that combined with the original dataset to train DTGEN. We use spline interpolations of the calculated emissivity spectra to generate uniformly spaced 400-points-long emissivity spectra arrays for training. These 400 interpolation points were chosen in a linear spacing from 10^{13} rad/s to 0.8×10^{14} rad/s.

All continuous input and output parameters were converted to log scale for training by taking the log of the inputs and then exponentiating the outputs. Log scale features result in more accurate and stable training and performance results for systems such as ours where parameters span several orders of magnitude. However, errors and loss functions were all calculated on a linear scale. While the models were trained by minimizing mean-squared-error (MSE) of log scale features, we report the relative error (of the linear scale emissivity) as the metric by which we evaluate model performance. Relative error is more appropriate than squared error for human-interpreted (linear scale) results because our emissivity data span many orders of magnitude and MSE would disproportionately penalize errors of larger values for such linear scale features. We define relative error for integrated emissivity as $|\epsilon_{ML} - \epsilon_{DNS}|/\epsilon_{DNS}$, and for spectral emissivity as shown in Equation 5.

$$E_{rel} = \frac{\int_\omega |\epsilon_{ML} - \epsilon_{DNS}| \, d\omega}{\int_\omega \epsilon_{DNS} d\omega}. \tag{5}$$

DTGEN takes approximately 11 (135) CPU-ms/sample to train, 0.005 (1.6) CPU-ms/sample to approximately solve the forward problem, and 0.2 (30) CPU-s to approximately solve the inverse design problem for integrated (spectral) emissivity targets, respectively. Numerically solving Maxwell's equations to calculate the exact optical properties for one known particle takes about 12 CPU-h, or 43.2×10^3 CPU-s. Assuming traditional optimization algorithms require around 1,000 iterations of solving the forward problem using DNS in order to solve the inverse problem, this translates to $\sim 43.2 \times 10^6$ CPU-s. While DTGEN does not solve the inverse problem exactly, it is expected to provide solutions accurate to within the error bounds discussed above (prediction uncertainties given in Figure S13). If this level of accuracy and uncertainty are acceptable for a given application, then DTGEN may significantly reduce the computation resources and/or time required to identify a candidate solution to the inverse design problem down to 0.18 (31.31) CPU-s for integrated (spectral) emissivities. Run times are shown in Figures S10 and S11 and Table S1. In situations where very high accuracy is desired, DTGEN can still be used to rapidly converge on an approximate solution, which can then be used as a close "initial guess" for an exact DNS solver. We have used CPU-s units as they are less dependent on wall clock time or particular configurations of the computation hardware.

CellPress
OPEN ACCESS

Cell Reports
Physical Science
Article

The feature importance analysis by material was performed using the standard Gini impurity method.[55] However, to calculate feature importance by material, we traversed the decision tree and calculated feature importance for any subtree whose root node made a split based on material. To our knowledge, this is the first formulation of feature importance over categorical one-hot encodings in a decision tree, and we believe this procedure is broadly applicable for interpreting decision trees produced by our method.

Inverse Design

The purpose of the inverse design is to find particle parameters (i.e., material, geometry class, and dimensions) that can achieve a target emissivity (either integrated or spectral) with a desired criterion (e.g., a desired value or distribution).

Before we start the inverse design process, we define a criterion for how a given spectrum is far from a DTGEN leaf. We will call this criterion the loss function. The loss function definition depends on the desired inverse design target. For inverse design that requires a target integrated (spectral) emissivity of a given value (array), the loss function will be simply the squared error (sum of squared errors over frequency points) between the target emissivity and the emissivity predicted by DTGEN (i.e., the emissivity at the leaf of DTGEN). For the case of a desired peak emissivity to happen at a given frequency or wavelength with a desired spectral width, we would like to maximize the emissivity inside the desired spectral width, while minimize it everywhere else. Therefore, we define the loss function as the ratio between the area under the spectral emissivity curve outside and inside the desired spectral width, in frequency domain.

With the loss function defined, we can find the DTGEN leaf (or a group of leaves) which minimizes the loss function. We call this leaf the target leaf. Using this target leaf, we can extract the design rules, which are the ranges imposed by DTGEN on various features (i.e., material, geometry class and dimensions) to reach the target leaf. Design rules are determined by finding the intersection of all the split rules that happened along the way that ended at the target leaf.

We can use the design rules to generate candidate designs that should achieve an emissivity very close to the emissivity predicted at the target leaf. We generate the candidate designs by randomly choosing particle parameters within ranges dictated by the design rules, while ensuring these particle parameters are consistent with the geometry class, as will be illustrated later.

Generation of Candidate Designs from Design Rules

We start generating candidate designs by choosing materials and geometries within categories dictated by design rules for material and geometry class features, respectively. Then, we generate dimensions for candidate designs, depending on the geometry class suggested by design rules, as follows.

For spheres, we start by finding the diameter range that corresponds to the intersection of the four design rules for area to volume, longest, middle, and shortest dimensions. The diameter range that corresponds to the area-to-volume range can be calculated from the simple relationship $D = 6$ / area-to-volume. We use the resulted diameter range to generate diameters for candidate spheres.

Cylinders can be fully described by lengths and diameters, and they can be categorized into long wires and flat discs, according to whether the length is greater or smaller than the diameter, respectively. For long wires and flat discs, we start by

Cell Reports
Physical Science
Article

CellPress
OPEN ACCESS

randomly choosing length for candidate designs using longest or shortest ranges, respectively, dictated by the design rules. The diameter is chosen randomly from the intersection of the two-dimensional ranges: middle and shortest dimensions ranges for long wires, and longest and middle dimensions ranges for flat discs. Using lengths and diameters for candidate designs, we calculate their area to volume and choose only candidate designs in which the area to volume falls within the area-to-volume range dictated by the design rules.

Parallelepiped and triangular prisms can be described by the longest, middle, and shortest dimensions, and they can be generated for candidate designs by choosing them randomly from ranges dictated by the design rules, while making sure that longest > middle > shortest. We calculate the area to volume for candidate designs and choose only candidate designs in which the area to volume falls within the area-to-volume range dictated by the design rules. For the inverse design of triangular prisms, we assumed that the base height < base width < length along the extrusion axis.

AUTHOR CONTRIBUTIONS

M.E. conducted the numerical simulation, data generation, database construction, and mass data visualization, performed the literature search, and applied ML methods for data analysis. C.Y. performed the literature search, developed and implemented core code for the machine-learning model training and performance, and tested and validated inverse design methodology. The idea of using ML for inverse design was conceptualized by M.E., S.D.L., and R.S.P. and further developed by A.A., M.E., and C.Y. M.E., C.Y., S.D.L., and R.S.P wrote the paper. S.D.L. and R.S.P supervised the research.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Freedman, J.P., Wang, H., and Prasher, R.S. (2018). Analysis of Nanofluid-Based Parabolic Trough Collectors for Solar Thermal Applications. J. Sol. Energy Eng. *140*, 051008.

2. Omair, Z., Scranton, G., Pazos-Outón, L.M., Xiao, T.P., Steiner, M.A., Ganapati, V., Peterson, P.F., Holzrichter, J., Atwater, H., and Yablonovitch, E. (2019). Ultraefficient thermophotovoltaic power conversion by band-edge spectral filtering. Proc. Natl. Acad. Sci. USA *116*, 15356–15361.

3. Menon, A.K., Haechler, I., Kaur, S., Lubner, S., and Prasher, R.S. (2020). Enhanced solar evaporation using a photo-thermal umbrella for wastewater management. Nat. Sustain. *3*, 144–151.

4. Raman, A.P., Anoma, M.A., Zhu, L., Rephaeli, E., and Fan, S. (2014). Passive radiative cooling below ambient air temperature under direct sunlight. Nature *515*, 540–544.

CellPress
OPEN ACCESS

Cell Reports
Physical Science
Article

5. Zhai, Y., Ma, Y., David, S.N., Zhao, D., Lou, R., Tan, G., Yang, R., and Yin, X. (2017). Scalable-manufactured randomized glass-polymer hybrid metamaterial for daytime radiative cooling. Science 355, 1062–1066.

6. Wang, H., Kaur, S., Elzouka, M., and Prasher, R. (2019). A nano-photonic filter for near infrared radiative heater. Appl. Therm. Eng. 153, 221–224.

7. Guo, C., Xiao, M., Minkov, M., Shi, Y., and Fan, S. (2018). Photonic crystal slab Laplace operator for image differentiation. Optica 5, 251.

8. Lin, X., Rivenson, Y., Yardimci, N.T., Veli, M., Luo, Y., Jarrahi, M., and Ozcan, A. (2018). All-optical machine learning using diffractive deep neural networks. Science 361, 1004–1008.

9. Mohammadi Estakhri, N., Edwards, B., and Engheta, N. (2019). Inverse-designed metastructures that solve equations. Science 363, 1333–1338.

10. Wu, W., Wu, Z., Yu, T., Jiang, C., and Kim, W.-S. (2015). Recent progress on magnetic iron oxide nanoparticles: synthesis, surface functional strategies and biomedical applications. Sci. Technol. Adv. Mater. 16, 023501.

11. Yang, X., Liu, X., Liu, Z., Pu, F., Ren, J., and Qu, X. (2012). Near-infrared light-triggered, targeted drug delivery to cancer cells by aptamer gated nanovehicles. Adv. Mater. 24, 2890–2895.

12. Jang, E., Jun, S., Jang, H., Lim, J., Kim, B., and Kim, Y. (2010). White-light-emitting diodes with quantum dot color converters for display backlights. Adv. Mater. 22, 3076–3080.

13. Wang, Y., Ding, G., Mao, J.-Y., Zhou, Y., and Han, S.-T. (2020). Recent advances in synthesis and application of perovskite quantum dot based composites for photonics, electronics and sensors. Sci. Technol. Adv. Mater. 21, 278–302.

14. Crnjak Orel, Z.C., and Klanjšek Gunde, M. (2001). Spectrally selective paint coatings: Preparation and characterization. Sol. Energy Mater. Sol. Cells 68, 337–353.

15. Flanner, M.G., Zender, C.S., Randerson, J.T., and Rasch, P.J. (2007). Present-day climate forcing and response from black carbon in snow. J. Geophys. Res. Atmos. 112, D11202.

16. Lavrinenko, A.V., Lægsgaard, J., Gregersen, N., Schmidt, F., Søndergaard, T., Lægsgaard, J., Gregersen, N., Schmidt, F., and Søndergaard, T. (2014). Numerical Methods in Photonics (CRC Press).

17. Molesky, S., Lin, Z., Piggott, A.Y., Jin, W., Vucković, J., and Rodriguez, A.W. (2018). Inverse design in nanophotonics. Nat. Photonics 12, 659–670.

18. Jensen, J.S., and Sigmund, O. (2011). Topology optimization for nano-photonics. Laser Photonics Rev. 5, 308–321.

19. Peurifoy, J., Shen, Y., Jing, L., Yang, Y., Cano-Renteria, F., DeLacy, B.G., Joannopoulos, J.D., Tegmark, M., and Soljačić, M. (2018). Nanophotonic particle simulation and inverse design using artificial neural networks. Sci. Adv. 4, eaar4206.

20. Ma, W., Cheng, F., and Liu, Y. (2018). Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. ACS Nano 12, 6326–6334.

21. Liu, Z., Zhu, D., Rodrigues, S.P., Lee, K.-T., and Cai, W. (2018). Generative Model for the Inverse Design of Metasurfaces. Nano Lett. 18, 6570–6576.

22. Malkiel, I., Mrejen, M., Nagler, A., Arieli, U., Wolf, L., and Suchowski, H. (2018). Plasmonic nanostructure design and characterization via Deep Learning. Light Sci. Appl. 7, 60.

23. Nadell, C.C., Huang, B., Malof, J.M., and Padilla, W.J. (2019). Deep learning for accelerated all-dielectric metasurface design. Opt. Express 27, 27523–27535.

24. Luque, A., Martí, A., and Stanley, C. (2012). Understanding intermediate-band solar cells. Nat. Photonics 6, 146–152.

25. Horodynski, M., Kühmayer, M., Brandstötter, A., Pichler, K., Fyodorov, Y.V., Kuhl, U., and Rotter, S. (2020). Optimal wave fields for micromanipulation in complex scattering environments. Nat. Photonics 14, 149–153.

26. Kauranen, M., and Zayats, A.V. (2012). Nonlinear plasmonics. Nat. Photonics 6, 737–748.

27. Woodward, X., Kostinski, A., China, S., Mazzoleni, C., and Cantrell, W. (2015). Characterization of dust particles' 3D shape and roughness with nanometer resolution. Aerosol Sci. Technol. 49, 229–238.

28. Noh, J., Kim, J., Stein, H.S., Sanchez-Lengeling, B., Gregoire, J.M., Aspuru-Guzik, A., and Jung, Y. (2019). Inverse Design of Solid-State Materials via a Continuous Representation. Matter 1, 1370–1384.

29. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., and Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent. Sci. 4, 268–276.

30. Umehara, M., Stein, H.S., Guevarra, D., Newhouse, P.F., Boyd, D.A., and Gregoire, J.M. (2019). Analyzing machine learning models to accelerate generation of fundamental materials insights. npj Comput. Mater. 5, 34.

31. Lundberg, S.M., and Lee, S.I. (2017). A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 27, 4766–4775.

32. Shin, S., Elzouka, M., Prasher, R., and Chen, R. (2019). Far-field coherent thermal emission from polaritonic resonance in individual anisotropic nanoribbons. Nat. Commun. 10, 1377.

33. Zhang, Q., Li, G., Liu, X., Qian, F., Li, Y., Sum, T.C., Lieber, C.M., and Xiong, Q. (2014). A room temperature low-threshold ultraviolet plasmonic nanolaser. Nat. Commun. 5, 4953.

34. Pérez-Juste, J., Pastoriza-Santos, I., Liz-Marzán, L.M., and Mulvaney, P. (2005). Gold nanorods: Synthesis, characterization and applications. Coord. Chem. Rev. 249, 1870–1901.

35. Bohren, C.F., and Huffman, D.R. (1983). Absorption and Scattering of Light by Small Particles (Wiley).

36. Fernández-Hurtado, V., Fernández-Domínguez, A.I., Feist, J., García-Vidal, F.J., and Cuevas, J.C. (2018). Super-Planckian far-field radiative heat transfer. Phys. Rev. B 97, 045408.

37. Biehs, S.-A., and Ben-Abdallah, P. (2016). Revisiting super-Planckian thermal emission in the far-field regime. Phys. Rev. B 93, 165405.

38. Thompson, D., Zhu, L., Mittapally, R., Sadat, S., Xing, Z., McArdle, P., Qazilbash, M.M., Reddy, P., and Meyhofer, E. (2018). Hundred-fold enhancement in far-field radiative heat transfer over the blackbody limit. Nature 561, 216–221.

39. Bohren, C.F. (1983). How can a particle absorb more than the light incident on it? Am. J. Physiol. 51, 323–327.

40. Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32.

41. Domingos, P. (1998). Knowledge discovery via multiple models. Intell. Data Anal. 2, 187–202.

42. Golyk, V.A., Krüger, M., and Kardar, M. (2012). Heat radiation from long cylindrical objects. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 85, 046603.

43. Luk'yanchuk, B.S., Miroshnichenko, A.E., Tribelsky, M.I., Kivshar, Y.S., and Khokhlov, A.R. (2012). Paradoxes in laser heating of plasmonic nanoparticles. New J. Phys. 14, 093022.

44. Wei, W., Fedorov, A.G., Luo, Z., and Ni, M. (2012). Radiative properties of dense nanofluids. Appl. Opt. 51, 6159–6171.

45. Rodriguez, A.W., Reid, M.T.H., and Johnson, S.G. (2013). Fluctuating-surface-current formulation of radiative heat transfer: Theory and applications. Phys. Rev. B - Condens. Matter Mater. Physiol. 88, 1–20.

46. SCUFF-EM https://github.com/homerreid/scuff-EM.

47. Geuzaine, C., and Remacle, J.-F.F. (2009). Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. Int. J. Numer. Methods Eng. 79, 1309–1331.

48. Olmon, R.L., Slovick, B., Johnson, T.W., Shelton, D., Oh, S.-H., Boreman, G.D., and Raschke, M.B. (2012). Optical dielectric function of gold. Phys. Rev. B Condens. Matter Mater. Phys. 86, 235147.

49. Cataldo, G., Beall, J.A., Cho, H.-M., McAndrew, B., Niemack, M.D., and Wollack, E.J. (2012). Infrared dielectric

Cell Reports
Physical Science
Article

CellPress
OPEN ACCESS

properties of low-stress silicon nitride. Opt. Lett. *37*, 4200–4202.

50. Palik, E.D. (1991). Palik, Handbook of Optical Constants of Solids, Vol. 2 (Wiley).

51. Orfanidis, S.J. (2016). Electromagnetic Waves and Antennas (Rutgers University).

52. Edalatpour, S., and Francoeur, M. (2013). Size effect on the emissivity of thin films. J. Quant. Spectrosc. Radiat. Transf. *118*, 75–85.

53. Schäfer, J., Lee, S.C., and Kienle, A. (2012). Calculation of the near fields for the scattering of electromagnetic waves by multiple infinite cylinders at perpendicular incidence. J. Quant. Spectrosc. Radiat. Transf. *113*, 2113–2123.

54. Kattawar, G.W., and Eisner, M. (1970). Radiation from a homogeneous isothermal sphere. Appl. Opt. *9*, 2685–2690.

55. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res.