Submitted to **Energy**, June 24, 2023; revised November 28, 2023

Thermophotovoltaic emitter design with a hyper-heuristic custom optimizer enabled by deep learning surrogates

Preston Bohm, Chiyu Yang, Akanksha K. Menon, and Zhuomin M. Zhang*

George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

Micro/nanostructures hold promise for use as emitters to boost the efficiency of thermophotovoltaic (TPV) systems. For example, periodic gratings may alter the spectrum of irradiation on the photovoltaic cell to better match the spectral response of the cell. Photons with energies slightly higher than the bandgap of the semiconductor are the most desired as they generate electron-hole pairs with minimal thermalization losses. This prompts the use of gratings as selective emitters, and the choice of grating geometry has been an active research topic. Even for a one-dimensional (1D) grating, millions of possible geometries exist, and each grating requires computationally intensive full-wave simulation, *e.g.*, the rigorous coupled-wave analysis (RCWA), to calculate the spectral, directional emissivity. Since optimization algorithm performance is problem-dependent, in this work, a hyper-heuristic search enabled by a fully connected neural network surrogate of the native RCWA is employed to select an algorithm for a specific grating design problem. A comparison with existing untuned algorithms for an ideal emitter problem demonstrates that the hyper-heuristically generated algorithm yields superior performance. This algorithm is then employed for the optimization of the emitter for a full TPV system comprising a heated 1D tungsten binary grating paired with a 300 K InGaSb cell. The system is optimized for maximum power or efficiency at 2000 K and 1500 K, respectively, and the grating properties for the optimized cases are analyzed.

Keywords: Deep learning, periodic gratings, hyper-heuristic, optimization, selective emitter, thermophotovoltaics

* Corresponding author: zhuomin.zhang@me.gatech.edu

Nomenclature

c_0	speed of light in vacuum
D	diffusion coefficent of minority carriers
d	grating depth
e	elementary charge
$F(s)$	loss/objective function
f	filling ratio
h	a simple heuristic
J_{ph}	photocurrent
J_{s}	reverse saturation current
j	jth diffraction order
k	wavevector
k_0	wavevector in vacuum
k_{B}	Stefan-Boltzmann constant
N	number of diffraction orders
$N_{\mathrm{A,D}}$	acceptor, doner concetrations
N_{i}	intrinsic carrier concetrations
q''	heat flux
$P_{\max}$	output power (per unit area) at operating point
r	Fresnel reflection coefficent
$\vec{r}$	random vector
$\mathbf{R}_D$	rotation matrix
s	a set grating parameters representing a possible solution
s^*	best solution of a given population
s_{z_i}	random solution of the population
T	temperature

$U(r_0-\sigma, r_0+\sigma)$ random uniform distribution function, centered around r_0

V_{OC} open circuit voltage

w grating width

Greek

ε permittivity

ε' real part of ε

ε'' imaginary part of ε

η (thermal) efficiency

θ zenith angle

$\Theta(\omega, T)$ mean energy of a Planck ocillator

λ wavelength (in vacuum)

Λ grating period

ν wavenumber ($1/\lambda$)

ξ exchange factor

σ radius of uniform distibution

τ lifetime for minority carriers

ϕ azmuthal angle

ω angular frequency

Other

ϵ emissivity

$\hbar$ reduced Planck constant

$\mathfrak{S}$ parameter space

Superscripts

p,s *p*-polarization (TM wave), *s*-polarization (TE wave)

Subscripts

1 or 2	medium 1 or 2
e or h	electrons or holes
g	bandgap

Acronyms

AAE	adversarial autoencoder
CNN	convolutional neural networks
FCNN	fully connected neural networks
GA	genetic algorithm
k-NN	k-nearest neighbors
LC	inductor capacitor
MP	magnetic polariton
PSO	particle swarm optimization
PV	photovoltaic
RCWA	rigorous coupled-wave analysis
RMSE	root-mean-squared error
SPP	surface plasmon polariton
TE, TM	transverse electric, transverse magnetic
TPV	thermophotovoltaic

1. Introduction

Thermophotovoltaic (TPV) systems are solid-state optical heat engines with few or no moving parts that can transform essentially any heat source into electricity [1,2]. In traditional photovoltaics, the incident photons on a photovoltaic (PV) cell directly come from the sun. TPVs pair low-bandgap (0.1-0.8 eV) PV cells with heated emitters, typically at temperatures from 1300 K to 2200 K using a variety of sources, such as radioactive isotopes [3], the sun [4,5], biomass in microscale energy systems [6], and industrial waste heat [7,8]. The performance of an emitter is generally determined by the ability to radiate photons at energies just above the bandgap of the PV cell, because photons with energies below the bandgap have little chance to create electron-hole pairs and those at energies much greater than the bandgap have large thermalization losses [8,9]. Several other technologies have been proposed to improve the performance of TPV systems. Back surface reflectors give photons a "second chance" at being absorbed by reflecting unused photons, thus increasing the system efficiency [10]. Similarly, spectral filters aim to reflect non-ideal photons and are typically located in between the heated body and the cell [7,8,11]. Recently, multi-junction cells have been combined with back reflectors to achieve efficiencies of 40% [12]. Near-field micro-TPV can take advantage of evanescent waves and photon tunneling effects to surpass the blackbody limit and improve the TPV performance [13-15]. All aforementioned techniques can be combined with selective emitters that radiate photons at desirable wavelengths for a particular cell.

A plethora of materials and microstructures have been considered to optimize these metrics for different system implementations [16-18]. Dias et al. [19] provided an extensive review and outlined a photonic roadmap to achieve high-efficiency TPV systems. Tungsten has been identified as a suitable material for TPV emitters [20-25] due to its favorable bulk emissivity [26-28], good

chemical stability, and high melting temperature of around 3600 K [29]. These desirable properties of tungsten have been combined with different subwavelength features using diffraction gratings [13,21], metallodielectric multilayer films [22], microcavities [20], photonic crystals [17], and micro/nanostructured metasurfaces [23-25] to enhance the radiative properties. One-dimensional (1D) binary gratings serve as building blocks for more complex structures, and they are more feasible for manufacturing compared to other intricate designs. The design of these gratings exemplifies an inverse problem [30], where the desired spectral emissivity informs the geometric parameters. Maximizing or minimizing the performance function can be challenging due to the multimodal nature of the problem over a large search space defined by the grating period, width, and depth [31]. Parameter sweeping can be extremely computationally expensive or lead to information holes based on the resolution of the search [13].

Machine learning has been extensively used to inversely design thermal emitters [32-35] and other energy-related applications [36-38]. These studies used trained neural networks as surrogates to simplify the complex calculations and then applied inverse solution algorithms to find the optimized materials, structures, and systems based on the desired performance functions. For example, Kecebas and Sendur [32] used the adjoint method to facilitate the gradient-based topological optimization. Kudyshev et al. [33] used adversarial autoencoders (AAEs) for rapid nanophotonics design and optimization. Sullivan et al. [34] developed an inverse neural network to invert the forward problem for the identification of the material and microstructure that would best match the specified optical response. Yang et al. [35] used a normalized flow method, also known as invertible neural networks, to optimize metasurface emitters for targeted TPV cells.

While heuristic optimization methods, such as the genetic algorithm (GA) and particle swarm optimization (PSO), have been used in earlier studies of 1D binary grating design [31,39],

advanced metaheuristic optimization and hyper-heuristic search techniques [40] have received little attention in optical and photonic design. Heuristic methods find approximate solutions that are good enough for a given problem; the optimization algorithms are stochastic, and their performance may be highly problem dependent. According to the no free lunch (NFL) theorem, any given metaheuristic optimization algorithm has equivalent performance compared to any other algorithm averaged over all possible problems. Methods have been proposed to procedurally build and optimize algorithms for a specific problem or a set of similar problems; these are referred to as hyper-heuristics [40]. Hyper-heuristic techniques are similar to parameter tuning meta-optimization techniques, except that a hyper-heuristic algorithm can combine components of several different algorithms into a more powerful one. In a recent study, Hu et al. [41] used a Monte Carlo tree search algorithm to optimize a multilayer emitter, considering the photocurrent generation and efficiency of the TPV system. In recent years, more and more studies have utilized metaheuristic optimization for mechanical design and energy system optimization [42-45].

This work applies machine learning techniques, especially deep neural networks, to the optimization of 1D binary grating design. The objective is to develop an optimization algorithm, *i.e.*, customer optimizer, using a hyper-heuristic search and apply it to maximize the specific TPV efficiency or maximum output power. Deep learning is employed to significantly reduce the computation time requirement for full-wave numerical solutions without much loss of calculation accuracy. To choose the best surrogate method to replace the native rigorous coupled-wave analysis (RCWA) calculations, four data-driven models are compared: (1) decision tree, (2) k-nearest neighbors (k-NN), (3) fully connected neural networks (FCNN), and (4) hybrid ensemble neural networks (hybrid-NN). A simple grating design problem, *i.e.*, ideal emissivity optimization, that is solely based on the spectral normal emissivity function is considered. The hyper-heuristic

search is employed using the FCNN surrogate to identify a customer optimizer. Comparison is made with several untuned heuristic algorithms to demonstrate the advantage of this optimizer, which is then employed to optimize a TPV system comprising a 1D tungsten grating, heated at 2000 K and 1500 K, that is paired with an InGaSb cell at 300 K. Furthermore, the driving physics behind the properties of the gratings designed for these cases is elucidated with emissivity contours.

2. Methods

2.1. Ideal emissivity optimization

Consider a TPV system, as schematically shown in Fig. 1(a), that is composed of a binary tungsten grating and a ternary $In_{0.18}Ga_{0.82}Sb$ cell, with a band gap of 0.56 eV [8,10]. A common method for designing such a grating is to choose an ideal emissivity spectrum and then optimize the grating geometry by minimizing the mean square difference between the grating emissivity at selected emission angles and the ideal emissivity spectrum [31]. The ideal emissivity function may be simplified as a step function with a value of one (or zero) at wavelengths shorter (or longer) than the wavelength corresponding to the bandgap, λ_g, as depicted in Fig. 1(b), along with the normal emissivity of plain tungsten and the blackbody intensities at 2000 K and 1500 K (normalized to the corresponding peak value). Clearly, the InGaSb cell with λ_g around 2 μm matches well with the blackbody emission spectra and the plain tungsten already has a favorable spectral selectivity. In this study, the dielectric function of tungsten is assumed to be independent of the temperature and the values at room temperature are taken from Ref. [26].

The emissivity of the grating is calculated by subtracting the reflectance from one, according to Kirchhoff's law, since the grating structure is assumed to be opaque [46,47]. The reflectance as a function of incidence angle, wavelength, and polarization state may be calculated

by RCWA, as shown in Fig. 2, for specified grating parameters [48,49]. The plane of incidence is defined by the wavevector **k** and the *z*-axis. To obtain the desired optimizer using the hyper-heuristic search method, only the normal emissivity of the gratings is considered in the spectral region for wavelength λ from 0.6 μm to 4.0 μm, covering the main region of the blackbody distribution function. For normal incidence, the azimuthal angle ϕ is set to zero so that the plane of incidence is the *x*-*z* plane. The bandgap wavelength of InGaSb is set to λ_g = 2.2 μm in the present study. The spectral region is expanded in the TPV system optimization, considering the actual properties of the $In_{0.18}Ga_{0.82}Sb$ cell, and the angular and polarization dependence. The objective function *F* is a modified mean square error that gives equal weighting to the emissivity values above and below the bandgap. The parameter space $\mathfrak{S}$ for the grating period Λ, width *w* (determined by the filling ratio $f = w/\Lambda$), and depth *d* are determined by considering common resonance features associated with the grating as done by Nguyen-Huu and Chen [31]. In general, metallic gratings support surface plasmon polaritons (SPPs), Wood's anomaly, cavity modes, and magnetic polaritons (MPs) [47,49]. A brief discussion of these modes is given in the following section.

Cavity resonances arise from the interference of the diffracted waves and the resonance wavelength for the 1D gratings maybe predicted by [20,47]:

$$\lambda_{m,p} = \frac{2}{\sqrt{\left(\frac{m}{\Lambda - w}\right)^2 + \left(\frac{p}{2d}\right)^2}} \tag{1}$$

where m = 0, 1, 2, 3… and p = 0, 1, 3, 5…, as long as they are not simultaneously zero. Cavity modes may be excited for both transverse electric (TE) and transverse magnetic (TM) waves. For

deep gratings with a high aspect ratio $d/(\Lambda - w)$, MPs may dominate cavity modes, giving a much longer resonance wavelength than $4d$, predicted by Eq. (1) [50].

Wood's anomaly causes abrupt changes in the optical properties of the grating when a diffraction order just emerges or submerges the surface of the grating, creating a surface wave. The resonance wavelength may be predicted according to [49]

$$\left(\frac{j\lambda_j}{\Lambda}\right)^2 + 2\frac{j\lambda_j}{\Lambda}\sin\theta\cos\phi - \cos^2\theta = 0 \tag{2}$$

where j signifies the jth diffraction order that satisfies the Bloch-Floquet condition given by [49]

$$k_{x,j} = k_0 \sin\theta\cos\phi + j\frac{2\pi}{\Lambda} \tag{3}$$

where $k_{x,j}$ is the x-component of the wavevector for the jth diffraction order and $k_0 = 2\pi/\lambda$ is the wavevector in vacuum or air. For normal incidence ($\theta = 0°$), Eq. (2) gives $\lambda_j = \Lambda/j$ (*i.e.*, Λ, $\Lambda/2$, $\Lambda/3$, …).

SPPs are electromagnetic waves that travel along the metal-dielectric interface. For SPPs to occur between the air-tungsten interface, $k_{x,j}$ must satisfy the following equation [27,47]:

$$k_{x,j} = \pm k_0\sqrt{\frac{\varepsilon}{\varepsilon+1}} \tag{4}$$

where ε is the dielectric function of tungsten. For large absolute ε values, Wood's anomaly and SPPs are hard to distinguish [27]. However, when $\phi = 0°$, SPPs can only be excited by TM waves, while Wood's anomaly may be excited by both TE and TM waves. For normal incidence, the longest wavelength for Wood's anomaly to occur is Λ when $j = \pm 1$. If all incidence angles are considered, the longest wavelength that Wood's anomaly could occur is 2Λ with $\theta = 90°$, $\phi = 0°$, and $j = -1$. These criteria apply to SPPs conditions within reasonable approximation. The range of

the grating width and depth may be chosen based on cavity resonances. To boost the averaged emissivity for the two polarizations at wavelengths shorter than and close to λ_g, following Ref. [31], the parameter space is chosen as follows:

$$\mathfrak{S}(\Lambda, f, d): \begin{cases} \Lambda \in [0.3\ \mu\text{m},\ 2.0\ \mu\text{m}] \\ f \in [0.1,\ 0.9] \\ d \in [0.3\ \mu\text{m},\ 2.0\ \mu\text{m}] \end{cases} \tag{5}$$

While these parameters may be continuously tuned, a practical limitation of 10 nm resolution associated with available microfabrication techniques is considered. The grating period, width, and depth are digitalized or rounded to the nearest hundredth of a micrometer. If f is chosen with an increment of 0.01, then there are 81 points from 0.1 to 0.9. The total number of grating configurations is approximately 2.4 million.

2.2. RCWA surrogates with deep learning

Due to the complex non-homogeneous nature of Maxwell's equations in 1D binary gratings, several different computational methods have been utilized to handle computing the optical properties, *e.g.*, finite difference time domain (FDTD) [32-34], RCWA [23,27,31], and effective medium theory (EMT) [47]. EMT cannot capture the effect of subwavelength features unless the geometric dimensions are much smaller than the wavelength [49]. While FDTD has advantages for modeling more complicated microstructures, it is much more computationally intensive than RCWA for 1D gratings [24]. RCWA allows for the calculation of spectral directional emissivity by taking advantage of the periodic nature of gratings, and the predicted result approaches the exact solution by increasing the number of diffraction orders used in the simulation [48,49]. The emissivity is calculated from the reflectance, which is summed up for all diffraction orders that are propagating waves, j = 0, ±1, ±2…. While only a few of them are propagating waves, to obtain

convergence of RCWA solutions, sufficient diffraction orders need to be employed in the calculation. In the present study, the number of diffraction orders (N) is set to 81, that is, the highest positive and negative diffraction orders are ±40, which allows the computed emissivity to be accurate to the third decimal point. The RCWA calculations for 100 discrete wavelengths and both polarizations take approximately 24 s for normal incidence on a PC with an Intel Core i7-6700 3.4 GHz CPU.

Despite the relatively fast converging speed with the RCWA simulation, it is still too slow for use in the hyper-heuristic search. The time constraint of RCWA motivates the development of a surrogate model to approximate RCWA simulation. Therefore, the present work explores a data-driven bottom-up approach to surrogate models that focuses on the input/output behavior of RCWA. The input data can be represented as a vector containing the period (Λ), width (w), depth (d), zenith angle (θ), azimuthal angle (ϕ), a polarization binary (TE), wavelength (λ), real part of dielectric function (ε'), and the imaginary part of dielectric function (ε''). The output data are the spectral directional emissivity (ϵ_λ). Four different classes of surrogate models are explored, namely, decision tree, k-NN, FCNN, and hybrid-NN. These models are described in the following, considering their advantages and disadvantages, for use as RCWA surrogates for normal incidence ($\theta = 0°$ and $\phi = 0°$) for the geometric parameter space $\mathfrak{S}$ given in Eq. (5).

A decision tree is a supervised learning method based on a nonparametric hierarchical model for classification and regression [51]. Tree models often have several orders of magnitude faster training cycles than other regression models, as tabular data correspond better to the axis-aligned splits seen in tree models [52]. Therefore, a decision tree model with a depth of 25 is adopted as a surrogate model and implemented using Scikit-learn in Python [53]. The input data for this study have relatively few independent dimensions (9 total), of which two collapse (θ,ϕ) as

their value is constant for normal incidence. Another widely used regression model is based on the nearest neighbors, k-NN, which is suitable to the present work as the Euclidean distance between the geometric input data provides useful information [54]. A 10-nearest-neighbor regression model is trained using Scikit-learn [53], giving uniform weighting to each neighbor selected. Scikit-learn is a tool kit in machine learning written in Python and it allows for easy tuning of the model hyper-parameters.

Neural networks can emulate any function to an arbitrary accuracy given enough training data [55]. The current trend in machine learning is the extensive utilization of recurrent neural networks (RNNs) and convolutional neural networks (CNNs). RNN and CNN models have the outputs of neurons and cycle back to feed into the inputs. The main difference is that RNNs have infinite impulse response while CNNs have a finite response. RNN and closely related long short-term memory (LSTM) models are commonly used for physics simulation surrogates [56,57], as these models typically predict the future states of a system. These models are not applicable to RCWA surrogate generation, as the input data are non-sequential and lack locality that can be exploited by convolutions. Therefore, a deep learning model consisting of an FCNN with three hidden layers of sizes 512, 1024, and 1024 is trained over 200 epochs. A diagram of FCNN model architecture is shown in Fig. 3. The three hidden layers use a LReLU activation function, while the output neuron uses sigmoid. This model is explicitly implemented in PyTorch using a Leaky Rectified Linear Unit (LReLU) activation function and sigmoid neuron as the output [58].

The fourth model used here is a 5-layer stacked ensemble model that is composed of decision trees, k-NNs, and FCNNs. Ensemble models combine the predictions of multiple individual models to make more accurate predictions than any single model could achieve alone. This is because different models may have different strengths and weaknesses, and a wider range

of patterns and relationships in the data can be captured by combining them. This hybrid model is implemented through AutoGluon utilizing bagging, stacking, and boosting to further increase the model performance [52,59].

2.3. Generation of new optimization algorithms

Metaheuristics, within the realm of optimization algorithms, are high-level strategies that iteratively explore and improve solutions by intelligently navigating complex solution spaces [42,44]. Unlike exact algorithms that guarantee optimality, metaheuristics do not provide such guarantees but aim to find near-optimal solutions efficiently. Metaheuristic algorithms can be broken down into a sequence of "building blocks", called simple heuristics. An algorithm's performance is dependent on the particular application or set of problems. Therefore, a new purpose-built algorithm can be created by re-arranging these "building blocks" or simple heuristics. Hyper-heuristic methods are techniques to procedurally build and optimize the simple heuristics composing a metaheuristic optimization algorithm [60,61]. The building blocks can be swapped out and rearranged to generate new algorithms. These simple heuristics take in a population of solutions and output a new population of solutions. The hyper-heuristic search essentially optimizes different simple heuristics to create a new metaheuristic optimization algorithm. Then the performance of this algorithm is evaluated, and the information of this performance is used in the generation of subsequent algorithms.

In the present study, a hyper-heuristic search is applied to develop an algorithm that excels at designing a binary periodic grating that closely matches an ideal emissivity spectrum. Here, the objective function is a weighted mean square difference between the spectral normal emissivity and the ideal emissivity defined in Sec. 2.1 over 100 equally spaced wavelengths averaged for TE

and TM polarization. The hyper-heuristic search technique is illustrated in Fig. 4 through a block diagram and implemented with CUSTOMHyS [61]. An algorithm cardinality of two is chosen to limit possible algorithms to being composed of 2 simple heuristics. To begin, two randomly chosen simple heuristics, $h_1^{(0)}$ and $h_2^{(0)}$, are chosen from the simple heuristic algorithm library [61]. These simple heuristics make up the first metaheuristic algorithm. The performance of this algorithm is evaluated and then the simulated annealing optimization is used to generate new heuristics, which is then evaluated and annealed for 1000 times until a total of 1001 metaheuristics are generated, *i.e.*, $h_1^{(1000)}$ and $h_2^{(1000)}$, each having its own performance indicator. The heuristic algorithm with the best performance value is chosen as the output customized optimizer composed of h_1 and h_2 for later use to optimize a similar problem of designing gratings for optimal efficiency or maximum power of the TPV systems to be discussed in Sec. 3. The specifics of this procedure are explained in the following steps:

1. Initialize the procedure and generate a metaheuristic algorithm composed of two simple heuristics, $h_1^{(m)}$, $h_2^{(m)}$. Here, m is an index from 0 to 1000, representing the step in hyper heuristic procedure.
2. Evaluate performance of algorithm
    a. Use algorithm to optimize a grating to minimize the objective function
        i. Initialize a population of eight (8) solutions, each solution is represented by prescribed grating parameters (Λ, f, d) that are randomly picked from the parameter space given in Eq. (5).
        ii. Apply the given heuristic $h_1^{(m)}$ to the population to get a new population; then, apply $h_2^{(m)}$ to the new population to get another one.

iii. Repeat Step 2.a.ii for 10 times, and then output the best solution from the 10 populations (*i.e.*, 80 solutions) based on the objective function (the smaller the better).

b. Run the optimization algorithm (Step 2.a), 100 times to collect a set of 100 optimal gratings. For each metaheuristic algorithm specified by $h_1^{(m)}$ and $h_2^{(m)}$, it takes 8×2×10×100 = 16,000 calculations of the emissivity spectra for both polarizations in order to obtain a set of 100 optimal gratings with their associated objective functions.

c. Take the 75th percentile of the objective function values from the set of optimized gratings obtained in step 2.b. This sampling represents the minimum expected performance for 75% of optimization runs. In other words, 25 out of the 100 data have a higher objective function value, (*i.e.*, 25% of the gratings are worse). Use this 75th percentile value as the performance indicator (the smaller the better) of the particular metaheuristic algorithm defined by $h_1^{(m)}$ and $h_2^{(m)}$.

3. Use the algorithms performance and simulated annealing optimization to generate a new metaheuristic algorithm.
4. Repeat steps 2 and 3, until 1000 metaheuristic algorithms are generated.
5. Output the best metaheuristic algorithm from the set of 1000 algorithms.

The total process requires approximately 16 million grating spectra for each polarization. The grating emissivity calculations use the FCNN model, which is orders of magnitude faster than the native RCWA simulation. It would not be feasible to do the hyper-heuristic search without the deep learning surrogate.

3. Thermophotovoltaic modeling

While it is instructive to use an ideal emissivity spectrum to optimize the grating, the result may not be the best choice for actual system performance. Hence it is imperative to directly optimize the TPV system using the performance parameters (*i.e.*, the efficiency and maximum output power), even though such optimization will significantly increase the complexity and computational demand. As shown in Fig. 1(a), the system is modeled as two semi-infinite plates separated by a distance much longer than the wavelength such that the thermal radiation between them is in the far-field regime, meaning no evanescent waves contribute to the heat flux between the two surfaces. The net flux from the emitter to the cell is then calculated by integrating over all frequencies as follows [47]:

$$q'' = \int_0^\infty q''_\omega(\omega)d\omega = \frac{1}{8\pi^3 c_0^2}\int_0^\infty [\Theta(\omega,T_2)] - \Theta(\omega,T_1)]\xi(\omega)\omega^2 d\omega \tag{6}$$

where ω is the angular frequency, q''_ω is the spectral heat flux, c_0 is the speed of light in vacuum, $\Theta(\omega,T) = \dfrac{\hbar\omega}{\exp(\hbar\omega/k_B T)-1} + \dfrac{\hbar\omega}{2}$ is the mean energy of Planck's oscillator at temperature T (with k_B and $\hbar$ being the Boltzmann constant and reduced Planck's constant), and subscripts 1 and 2 denote the emitter and the cell, respectively. In Eq. (6), $\xi(\omega)$ is an exchange factor that takes into consideration of the emissivity of both bodies by a hemispherical integration considering individual polarization states [47],

$$\xi(\omega) = 4\int_0^{\pi/2}\int_0^{\pi/2}\left(\frac{1}{1/\epsilon^s_{\omega,1}+1/\epsilon^s_{\omega,2}-1} + \frac{1}{1/\epsilon^p_{\omega,1}+1/\epsilon^p_{\omega,2}-1}\right)\cos\theta\sin\theta d\theta d\phi \tag{7}$$

Here, ϵ is the spectral directional emissivity, and superscript s and p signify TE and TM waves, respectively. Due to symmetry, the integration over the azimuthal angle is taken to be 4 times the value integrated from 0 to $\pi/2$.

Integrating over all possible frequencies is not feasible, and therefore a characteristic subset is used. Here, the cutoff wavelengths are set to 0.4 μm and 10 μm, and 200 logarithmically spaced frequencies are used with trapezoidal numerical integration. Similarly, the zenith and azimuthal angle values are calculated every 5° for their trapezoidal numerical integration.

A scenario occurs when the polarization of the incident wavevector is not perpendicular to the gratings (*i.e.,* the azimuthal angle $\phi \neq 0$ or π), as illustrated in Fig. 2 where the plane of incidence is not parallel to the *x*-*z* plane. In this case, the majority of the diffracted beams ($j \neq 0$) do not lie in the plane of incidence; instead, they form a cone centered around the grating grooves [49]. When the azimuthal angle $\phi = 0$, RCWA calculations require $4N$, where N is the number of diffraction orders, unknowns to be solved for [48]. These unknowns are the coefficients of exponential terms that represent the forward and backward coupled diffracted waves in the grating region. The conical diffraction case is more computationally intense and has $10N$ unknowns to be solved [49]. No surrogate is used for the TPV performance evaluation. The use of a custom optimizer can significantly reduce the number of iterations needed in the inverse problem to optimize the objective function, *i.e.*, the efficiency or maximum power of the TPV system as described in the following.

The spectral directional emissivity of the PV cell, $In_{0.18}Ga_{0.82}Sb$, is calculated from the reflectance, which is the square of the modulus of the Fresnel reflection coefficients calculated by [46,47],

$$r^s = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \quad \text{and} \quad r^p = \frac{n_1 \cos\theta_2 - n_2 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1} \tag{8}$$

where n_1 is the refractive index of the incident medium ($n_1 = 1$ for air or vacuum), n_2 is the complex refractive index of the PV cell, and θ_1 and θ_2 are the incidence and refraction angles, respectively. To simplify the calculation, it is assumed that the dielectric function of the PV cell is independent of the doping as done in the previous studies [10,13].

Once the radiative flux on the cell is determined, the photocurrent from photons is calculated from [14]

$$J_{\text{ph}} = \int_{\omega_g}^{\infty} \frac{e}{\hbar\omega} q''_{\omega}(\omega) d\omega \tag{9}$$

where e is the element charge, and ω_g is the frequency associated with the bandgap of the cell. It is assumed that the internal quantum efficiency is 100%, *i.e.*, each absorbed photon above the bandgap generates exactly one electron-hole pair. Furthermore, non-radiative recombination losses are not considered. These assumptions provide an upper limit of the TPV performance and the inclusion of nonidealities should not significantly affect the emitter design. Such a detailed balance approach has been used previously [13,41]. The reverse saturation current may be calculated based on the diffusion process as [62]

$$J_s \approx e\left(\frac{N_i^2}{N_A}\sqrt{\frac{D_e}{\tau_e}} + \frac{N_i^2}{N_D}\sqrt{\frac{D_h}{\tau_h}}\right) \tag{10}$$

where $N_i = 2\times10^{13}$ cm^{-3}, $N_A = 1\times10^{19}$ cm^{-3}, and $N_D = 1\times10^{19}$ cm^{-3} are the intrinsic, acceptor, and donor concentrations, respectively; $D_e = 125$ cm^2 s^{-1} and $D_h = 31.3$ cm^2 s^{-1} are the diffusion coefficients for electrons and holes, and $\tau_e = 9.75$ ns and $\tau_h = 30.8$ ns are the lifetime of electrons and holes, respectively.

The open-circuit may be determined from the dark current curve, yielding [62]

$$V_{\mathrm{oc}} = \frac{k_{\mathrm{B}}T}{e}\ln(J_{\mathrm{ph}} / J_{\mathrm{s}} + 1) \tag{11}$$

The output power depends on the operating voltage, and the maximum power generated at the ideal voltage is given by [62]

$$P_{\mathrm{max}} = J_{\mathrm{ph}}V_{\mathrm{oc}}\left(1 - \frac{1}{y}\right)\left[1 - \frac{\ln(y)}{y}\right] \tag{12}$$

where $y = \ln(J_{\mathrm{ph}} / J_{\mathrm{s}})$. The efficiency of the TPV system is the ratio of the maximum power to the net heat flux [10,14]:

$$\eta = \frac{P_{\mathrm{max}}}{q''} \tag{13}$$

The tungsten grating is optimized using the custom-optimization algorithm generated by the hyper-heuristic search discussed in the previous section for maximum power and separately for efficiency at two operating temperatures, 2000 K and 1500 K.

4. Results and discussion

4.1. Comparison of RCWA surrogate models

Each surrogate model is allotted about one day of computational time for training. The decision tree and k-NN models can be completed. The FCNN converges to a point, where accuracy improvements were no longer being obtained with subsequent training epochs. The hybrid-NN model does not reach a converging point. The Autogluon implementation allows for training time budgets to be set; this is a great feature for machine learning. However, if a new model feature is estimated to take more time than what is left in the budget, the training will prematurely stop before expending the entire computational budget. The hybrid-NN model would continue to get accuracy

improvements if more training time was allotted. The main downside of these hybrid models is the long training and/or prediction time. The longer prediction time associated with these hybrid models can be overcome with teacher-student deep leaning methods [63]. That is, the output of the complex hybrid model is used to train a simpler model like FCNN. This often drastically improves prediction time while slightly decreasing accuracy. Teacher-student methods further increase training time, reiterating the fact that hybrid neural net models are viable when large training computational resources are present.

A commonly used training-to-test split ratio is 80:20. However, other splitting ratios such as 70:30, 60:40; and even 50:50 have also been used in the literature [64]. In the present study, the splitting ratio of 2:1 is used to ensure the fidelity of the surrogates. The training uses a tabular data set composed of 16 million individual RCWA calculations (or 80,000 RCWA spectra for each polarization). Each of the RCWA surrogates is tested on a data set of 8 million RCWA calculations. The combined training and testing use about 5% of the total parameter space defined in Eq. (5). Using the deep learning enabled surrogate models are much faster than the brutal force optimization if the whole grating parameter space is considered.

Table 1 presents the training time, prediction time for each calculation, and accuracy for the surrogate models explored in this study. For RCWA, the prediction time is the average time needed to compute the emissivity for a single wavelength at a given polarization. The accuracy is based on the root-mean-squared error (RMSE) for 200 data as compared with the native RCWA calculations. Each of the models varies quite significantly in their training (or generation) time. The decision tree results in the quickest training of just 110 s and a prediction time of 4.15E-07 s. On the other end, it takes 20 hours for the FCNN model to be trained, while the prediction time of the hybrid-NN is 5.09E-04 s, which is the longest among these surrogate models. Generally, the

trend of increases in model complexity translates to longer training times and slower prediction speed. However, the most complex model of the hybrid-NN has a slightly shorter training time but a higher RMSE than the hybrid-NN model. The FCNN model is the most accurate with an RMSE of 0.0045.

To illustrate the accuracy of the surrogate model, Fig. 5 compares the emissivity spectra predicted by the surrogate models with that of RCWA calculations for an example emitter with Λ = 1.5 μm, w = 0.5 μm, and d = 1.0 μm. The common stair step behavior often associated with decision trees and k-NN is seen in Figs. 5(a,b). The neural network methods generally produce smoother curves as shown in Figs. 5(c,d). As expected, the differences are the largest near the narrow peaks and resonances. While the FCNN model closely agrees with the RCWA calculations, the hybrid-NN model exhibits some deviations. Therefore, the FCNN surrogate model is used for hyper-heuristic custom optimizer generation as discussed in the subsequent section.

4.2. Generated custom algorithm

Using the hyper-heuristic search described in Sec. 2.3, the resulting custom optimized algorithm for the grating problem is obtained that is composed of two simple heuristics h_1 and h_2. These algorithms are described in the following [60].

$$h_1(s) = s^* - \vec{r}\mathbf{R}_D(\Theta)(s - s^*) \tag{14}$$

This heuristic takes in a set of solutions, $s = (\Lambda, f, d)\times 8$, called a population, and manipulates them with a spiral dynamic heuristic. In Eq. (14), $\mathbf{R}_D(\Theta = 22.5^o)$ is the rotation matrix determined by the product of all the combinations of two-dimensional rotation matrices by utilizing the Euler-Rodrigues rotation formula, s^* is the best solution of the population, and $\vec{r}$ is a uniformly distributed random vector. The elements, r_i, of $\vec{r}$ are given by [60]

$$\vec{r} = r_i \sim U(r_0 - \sigma, r_0 + \sigma) \tag{15}$$

where U represent a uniformly chosen random number in a given range, this range is centered around $r_0 = 0.9$ with size $\sigma = 0.1$. The heuristic h_1 is followed by a differential mutation simple heuristic [60,61],

$$h_2(s) = s_{z_1} + F \cdot (s^* - s_{z_2}) + F \cdot \sum_{m=1}^{M} (s_{z_{2m+1}} - s_{z_{2m+2}}) \tag{16}$$

where $F = 1$ is the "strength" of the mutation, $M = 1$ is the number of mutations, and z_i ($i = 1, 2$) represents a random solution from the population of solutions. This heuristic essentially randomly mutates a random solution toward the best solution. Detailed discussion of the heuristics can be found from Cruz et al. [60,61].

Again, the heuristics are paired with a uniform random initializer to generate the first population of solutions. Due to the nature of the cost function evaluating the utility of the algorithms, the combination of these heuristics balances exploitation and exploration. This custom optimization algorithm should outperform most unoptimized algorithms. To verify this, this algorithm is compared to several untuned search algorithms with a random set of parameters: (1) the random search, (2) differential evolution, (3) PSO, (4) single crossover GA, and (5) firefly optimization algorithms for the optimization of an emitter according to the ideal emissivity function described previously. All of which have their simple heuristic explicitly defined by Cruz et al. [60]. The same algorithm objective function used in the hyper-heuristic search was used for this comparison. Figure 6 shows the 75th percentile of the optimization runs performed with each algorithm and shows the custom algorithm has a much better performance than the other algorithms. The custom algorithm exhibits extreme consistency as there is a much lower interquartile range at almost all algorithm iterations than the other algorithms. As discussed before,

each of the comparison algorithms can have their defining parameters tuned to improve their performance. The hyper-heuristic optimization essentially has already tried different parameters of each algorithm to generate the custom optimizer. Even if each of the comparison algorithms are meta-optimized to have their parameters tuned, the custom algorithm should still outperform the rest albeit the advantage may be less dramatic.

4.3. Optimized gratings and TPV performance

The emitter temperature of 2000 K is studied first. The figures of merit of a TPV system are the maximum power and the conversion efficiency, defined in Eq. (12) or Eq. (13), respectively. However, due to the change of spectral distribution, an optimal output power does not necessarily correspond to an optimal efficiency. Hence, the customer optimizer is used to optimize the TPV performance based on either the efficiency or the maximum power. The RCWA simulation considering conical cases is used for these calculations. Using the generated custom algorithm, the number of iterations needed is approximately 50 for a population of eight, making it feasible to use the native RCWA calculations. While it is possible to use a surrogate model for the RCWA for various incidence angles, the training would take a longer time than directly optimizing the system. The optimized grating geometries and their associated TPV efficiency and maximum power are listed in Table 2. The results for a planar tungsten emitter and the grating A optimized according to the ideal emissivity spectrum are also shown for comparison. The grating geometrics are identical for both optimal efficiency (grating B) and optimal output power (grating C), in the case with an emitter temperature of 2000 K. Note that grating A already has a fairly good performance compared with a planar tungsten emitter. The maximum power is more than doubled, though the absolute value of the efficiency is increased by 2.66% only. This is because the choice

of the ideal emitter spectrum is to match with the bandgap of the cell. Grating B differs from grating A only by a smaller filling ratio (or width). When the whole system is optimized, the performance further improves, resulting in an efficiency of 27.06% with an output power density of 7.19 W/cm^2.

When the emitter temperature is 1500 K, the geometric parameters of grating D for optimal efficiency are very different from those of grating E for optimal output power, as shown in Table 3. In this case, the overall efficiency and output power are lower than for a 2000 K emitter. Nevertheless, the enhancement over planar tungsten emitter is still significant. Furthermore, gratings D and E have quite different parameters than grating A. However, grating A seems to perform between gratings D and E in terms of either efficiency or maximum power. It is interesting to look at the emissivity spectra considering different polarizations to understand what features are the most important for TPV performance.

The spectral emissivity is plotted in Fig. 7 for the four gratings. Except in a few spectral regions, the hemispherical emissivity generally falls between the values of the TE and TM waves at normal incidence. Gratings A and B have similar emissivity with a slight shift of the peaks for TE waves. Cavity resonance with $\lambda = 2\Lambda(1-f)$ results in a peak for TE wave at 1.9 μm for grating A and 2.0 μm for grating B. Towards shorter wavelengths, TM waves exhibit a higher emissivity due to SPPs. Both gratings yield a low emissivity at wavelengths longer than 2.2 μm and thus enhance the TPV efficiency.

However, gratings C and D behave very differently. For grating D, the broader peak near 2 μm for TM waves is due to MPs that can be predicted using the LC circuit model but cannot be explained by the cavity mode of $\lambda = 4d$ [50]. Due to the small emissivity for TE waves with grating D, the output power is lower than those of gratings A and E. For grating E, there is a MP

mode in the infrared near 11 μm (not shown). The MPs may couple with the cavity modes to give multiple oscillations including those at 3.3 μm, 2.15 μm, and 1.55 μm. Hence, even though the output power is optimized, the efficiency is lower than those with gratings A and D and even lower than that of planar tungsten emitter. Since the heuristic optimization is based on the objective function only, the outcome may be partially or completely unexpected. To shed light on the directional emissivity and resonance modes, the emissivity contours are plotted and discussed in the next section.

4.4. Emissivity contours

The directional emissivity of the four gratings at $\lambda = 2$ μm are shown in polar contour plots for both TE and TM waves in Fig. 8. Grating A and B are similar while grating B has a higher emissivity since the selected wavelength is closer to its resonance peak. Cavity resonances form standing waves and result in several zones with high emissivity values, especially for TE waves as shown in Fig. 8(b). It should be mentioned that due to conical diffraction, the effects of TE and TM waves are somehow mixed when $\phi \neq 0°$ [49]. Hence, it is difficult to associate them with any particular resonances. This is especially true for gratings D as shown in Fig. 8(c,g) and grating E as shown in Fig. 8(d,h), where high emissivity occurs for both TE and TM waves in different regions. It should be noted that the reflectance of the cell increases toward higher zenith angles according to the Fresnel formulas given in Eq. (8). Therefore, the emitted photons at $\theta > 60°$ have a much lower probability to be absorbed.

Typically, resonances can be analyzed via dispersion relations for $\phi = 0°$. Figure 9 plots the emissivity contours for gratings A and B in terms of the wavenumber ν and the parallel wavevector k_x for both TE and TM waves. The horizontal bright stripes near $\nu = 5000$ cm^{-1}

($\lambda = 2.0$ μm) for TE waves are due to cavity modes as mentioned previously. Higher order resonances can also be seen. The Wood's anomaly can barely be seen only for TE waves as a diamond in Fig. 9(a,c). The corners of the diamond are located at (4500,4200), (0,9000), (4500,13000), and (9000,9000), which is extrapolated. For TM waves, there are multiple cavity resonances at large wavenumbers (or short wavelengths) that are consistent with the high emissivity spectral region shown in Fig. 7.

As shown in Fig. 10(a,c), gratings D and E are essentially featureless for TE wave, except that the values are higher at $\nu > 7000$ cm^{-1} and much lower at $\nu < 6000$ cm^{-1}. For TM waves, the MP resonance at ν near 5000 cm^{-1} results in a horizontal band seen in Fig. 10(b). At high wavenumbers, the bright line from (5750,6000) to (0,11400) is due to SPPs. This can also be seen in Fig. 7(c) by the spike near 1.5 μm. As shown in Fig. 10(d), there are several horizontal bands for grating E due to MPs and cavity resonances. On the other hand, due to the small grating period, SPPs cannot be excited in the considered spectral region. While the resonance features can be understood. In general, the algorithm's solutions can neither be designed intuitively nor based on fundamental mechanisms.

5. Conclusions

This study demonstrates that hyper-heuristic techniques allow the creation of effective optimization algorithms for the design of microscale TPV emitters. The new algorithm generation is facilitated by FCNN serving as RCWA surrogates for calculation of the normal emissivity of 1D tungsten gratings. The custom optimizer is then used in the full system evaluation to further optimize the grating parameters for specific TPV parameters and performance metrics, considering the spectral and directional emissivity of the gratings and TPV cells. The grating behavior can be

explained according to the resonance phenomena associated with cavity resonance, Wood's anomaly, SPPs, and MPs. Nevertheless, it is hard to draw any conclusions in terms of how to optimize the gratings simply based solely on these resonance mechanisms. Fortunately, the heuristic optimization is merely based on the objective function to identify the best solution.

A major takeaway from this study is that it demonstrates a pathway towards fast and reliable optimization of a complex system using a custom algorithm, which is developed from a hyper-heuristic search on a much simpler yet similar problem. Furthermore, the hyper-heuristic search is enabled by deep learning surrogates with only a subset of the input data obtained from rigorous calculations. This custom optimizer obtained from this work is recommended for emitter design of future TPV systems based on binary gratings with various emitter and cell materials. The procedure outlined here may also be adopted for optimizing other microscale energy conversion devices and systems based on advanced heuristic algorithms.

Future studies include modeling and optimization of gratings based on metals other than tungsten, TPV cells with different semiconductors and bandgaps, as well as two-dimensional gratings and other nanostructures for selective emitter designs.

Acknowledgements

This work was mainly supported by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (Grant No. DE-SC0018369). C.Y. would like to the support of the National Science Foundation (Grant No. CBET-2029892).

Author credit statement

P. Bohm: Conceptualization, Investigation, Methodology, Validation, Writing-original draft, Review and editing. C. Yang: Methodology, Validation, Review and editing. A.K. Menon: Supervision, Review and editing. Z.M. Zhang: Conceptualization, Funding acquisition, Supervision, Writing-original draft, Review and editing.

Declaration of competing interest

The authors declare no conflicts of interest.

Data availability

Data will be made available on request.

References

[1] Daneshvar H, Prinja R, Kherani NP. Thermophotovoltaics: Fundamentals, challenges and prospects. Appl Energy 2015;159:560–75. https://doi.org/10.1016/j.apenergy.2015.08.064

[2] Burger T, Sempere C, Roy-Layinde B, Lenert A. Present efficiencies and future opportunities in thermophotovoltaics. Joule 2020;4;1660–80. https://doi.org/10.1016/j.joule.2020.06.021

[3] Wang X, Liang R, Fisher P, Chan W, Xu J. Radioisotope thermophotovoltaic generator design methods and performance estimates for space missions. J Propul Power 2020;36:593–603. https://doi.org/10.2514/1.B37623

[4] Lenert A, Bierman DM, Nam, Y, Chan WR, Celanović I, Soljačić M, et al. A nanophotonic solar thermophotovoltaic device. Nat Nanotechnol 2014;9:126–30. https://www.nature.com/articles/nnano.2013.286

[5] Datas A, Ramos A, Martí A, del Canizo C, Luque A. Ultra high temperature latent heat energy storage and thermophotovoltaic energy conversion. Energy 2016;107:542–9. http://dx.doi.org/10.1016/j.energy.2016.04.048

[6] Chen W-L, Currao G, Li Y-H, Kao C-C. Employing Taguchi method to optimize the performance of a microscale combined heat and power system with Stirling engine and thermophotovoltaic array. Energy 2023;270:126897. https://doi.org/10.1016/j.energy.2023.126897

[7] Zhang C, Tang L, Liu Y, Liu Z, Liu W, Qiu K. A novel thermophotovoltaic optical cavity for improved irradiance uniformity and system performance. Energy 2020;195:116962. https://doi.org/10.1016/j.energy.2020.116962

[8] Basu S, Chen Y-B, Zhang ZM. Microscale radiation in thermophotovoltaic devices—a review. Int J Energy Res 2007;31(6-7):689–716. https://doi.org/10.1002/er.1286

[9] Zhou Z, Jiang C, Huang H, Lijiang Liang L, Zhu G. Three-junction tandem photovoltaic cell for a wide temperature range based on a multilayer circular truncated cone metamaterial emitter. Energy 2020;210;118503. https://doi.org/10.1016/j.energy.2020.118503

[10] Bright TJ, Wang LP, Zhang ZM. Performance of near-field thermophotovoltaic cells enhanced with a backside reflector. J Heat Transfer 2014;136(6):062701. https://doi.org/10.1115/1.4026455

[11] Mao L, Ye H. New development of one-dimensional Si/SiO_2 photonic crystals filter for thermophotovoltaic applications, Renew Energy 2010;35:249–56. https://doi.org/10.1016/j.renene.2009.06.013

[12] LaPotin A, Schulte KL, Steiner MA, Buznitsky K, Kelsall CC, Friedman DJ, et al. Thermophotovoltaic efficiency of 40%. Nature 2022;604(7905):287–91. https://www.nature.com/articles/s41586-022-04473-y

[13] Watjen JI, Liu XL, Zhao B, Zhang ZM. A computational simulation of using tungsten gratings in near-field thermophotovoltaic devices. J Heat Transfer 2017;139(5):052704. https://doi.org/10.1115/1.4035356

[14] Feng D, Yee SK, Zhang ZM. Improved performance of a near-field thermophotovoltaic device by a back gapped reflector. Sol Energy Mater Sol Cell 2021;237:111562. https://doi.org/10.1016/j.solmat.2021.111562

[15] Song J, Han J, Choi M, Lee BJ. Modeling and experiments of near-field thermophotovoltaic conversion: A review. Sol Energy Mater Sol Cell 2022;238: 111556. https://doi.org/10.1016/j.solmat.2021.111556

[16] Wang Z, Kortge D, He Z, Song J, Zhu J, Lee C, et al., Selective emitter materials and designs for high temperature thermophotovoltaic applications. Sol Energy Mater Sol Cell 2022;238:111554. https://doi.org/10.1016/j.solmat.2021.111554

[17] Sakakibara R, Stelmakh V, Chan WR, Ghebrebrhan M, Joannopoulos JD, Soljačić M, et al. Practical emitters for thermophotovoltaics: a review. J Photon Energy 2019;9(3):032713. https://doi.org/10.1117/1.JPE.9.032713

[18] Meng C, Liu Y, Xu Z, Wang H, Tang X. Selective emitter with core–shell nanosphere structure for thermophotovoltaic systems. Energy 2022;239:121884. https://doi.org/10.1016/j.energy.2021.121884

[19] Dias MRS, Gong T, Duncan MA, Stuart C. Ness SC, McCormack SJ, et al. Photonics roadmap for ultra-high-temperature thermophotovoltaics, Joule 2023;7: 2209–2227. https://doi.org/10.1016/j.joule.2023.08.015

[20] Shi H, Yugami H. Thermophotovoltaic generation with selective radiators based on tungsten surface gratings. Appl Phys Lett 2004;85(16):3399–3401. https://doi.org/10.1063/1.1807031

[21] Laroche M, Arnold C, Marquier F, Carminati R, Greffet J-J, Collin S, et al. Highly directional radiation generated by a tungsten thermal source. Opt Lett 2005;30(19):2623–5. https://doi.org/10.1364/OL.30.002623

[22] Narayanaswamy A, Chen G. Thermal emission control with one-dimensional metallodielectric photonic crystals. Phys Rev B 2004;70(12):125101. https://doi.org/10.1103/PhysRevB.70.125101

[23] Wang LP, Zhang ZM. Wavelength-selective and diffuse emitter enhanced by magnetic polaritons for thermophotovoltaics. Appl Phys Lett 2012;100 (6):063902. https://doi.org/10.1063/1.3684874

[24] Zhao B, Wang LP, Shuai Y, Zhang ZM. Thermophotovoltaic emitters based on a two-dimensional grating/thin-film nanostructure. Int J Heat Mass Transfer 2013;67:637–45. https://doi.org/10.1016/j.ijheatmasstransfer.2013.08.047

[25] Liu XJ, Zhao CY, Wang BX, Xu JM. Tailorable bandgap-dependent selective emitters for thermophotovoltaic systems. Int J Heat Mass Transfer 2023;200:123504. https://doi.org/10.1016/j.ijheatmasstransfer.2022.123504

[26] Palik ED, Handbook of Optical Constants of Solids II. San Diego: Academic Press; 1998.

[27] Chen Y-B, Zhang ZM. Design of tungsten complex gratings for thermophotovoltaic radiators. Opt Commun 2007;269(2):411–7. https://doi.org/10.1016/j.optcom.2006.08.040

[28] Watjen JI, Bright TJ, Zhang ZM, Muratore C, Voevodin AA. Spectral radiative properties of tungsten thin films in the infrared. Int J Heat Mass Transfer 2013;61:106-13. https://doi.org/10.1016/j.ijheatmasstransfer.2013.01.063

[29] Lassner E, Schubert W-D. Tungsten—Properties, chemistry, technology of the elements, alloys and chemical compounds. New York: Kluwer Academic; 1999.

[30] Ertürk H, Daun K, França FHR, Hajimirza S, Howell JR. Inverse methods in thermal radiation analysis and experiment. J Heat Transfer 2023;145:050801. https://doi.org/10.1115/1.4056371

[31] Nguyen-Huu N, Chen Y-B, Lo, Y-L. Development of a polarization-insensitive thermophotovoltaic emitter with a binary grating. Opt Express 2012;20(6):5882–90. https://doi.org/10.1364/OE.20.005882

[32] Kecebas MA, Sendur K. Broadband high-temperature thermal emitter/absorber designed by the adjoint method. J Opt Soc Am B 2021;38(10):3189–98. https://opg.optica.org/josab/abstract.cfm?uri=josab-38-10-3189

[33] Kudyshev ZA, Kildishev AV, Shalaev VM, Boltasseva A. Machine-learning-assisted metasurface design for high-efficiency thermal emitter optimization. Appl Phys Rev 2020;7:021407. https://doi.org/10.1063/1.5134792

[34] Sullivan J, Mirhashemi A, Lee J. Deep learning-based inverse design of microstructured materials for optical optimization and thermal radiation control. Sci Rep 2023;13:7382. https://doi.org/10.1038/s41598-023-34332-3

[35] Yang J-Q, Xu YC, Fan K, Wu J, Zhang C, Zhan D-C. Normalizing flows for efficient inverse design of thermophotovoltaic emitters. ACS Photon 2023;10:1001–11. https://doi.org/10.1021/acsphotonics.2c01803

[36] Afzal S, Ziapour BM, Shokri A, Shakibi H, Sobhani B. Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms, Energy 2023;282:128446. https://doi.org/10.1016/j.energy.2023.128446

[37] Ye Y, Wang H, Xu B, Zhang J. An imitation learning-based energy management strategy for electric vehicles considering battery aging, Energy 2023;283:128537. https://doi.org/10.1016/j.energy.2023.128537

[38] Toopshekan A, Abedian A, Azizi A, Ahmadi E, Vaziri Rad MA. Optimization of a CHP system using a forecasting dispatch and teaching-learning-based optimization algorithm, Energy 2023; 285:128671. https://doi.org/10.1016/j.energy.2023.128671

[39] Shokooh-Saremi M, Magnusson R. Particle swarm optimization and its application to the design of diffraction grating filters. Opt Lett 2007;32(8):894–6. https://doi.org/10.1364/OL.32.000894

[40] Burke EK, Gendreau M, Hyde M, Kendall G, Ochoa G, Özcan E, et al. Hyper-heuristics: a survey of the state of the art. J Oper Res Soc 2013;64:1695–1724. https://link.springer.com/article/10.1057/jors.2013.71

[41] Hu R, Song J, Liu Y, Xi W, Zhao Y, Yu X, et al., Machine learning-optimized Tamm emitter for high-performance thermophotovoltaic system with detailed balance analysis. Nano Energy 2020;72:104687. https://doi.org/10.1016/j.nanoen.2020.104687

[42] Elaziz MA, Elsheikh AH, Oliva D, Abualigah L, Lu S, Ewees AA. Advanced metaheuristic techniques for mechanical design problems: Review, Arch Computat Methods Eng 2022;29:695–716. https://doi.org/10.1007/s11831-021-09589-4

[43] Elsheikh AH, El-Said EMS, Elaziz MA, Fujii M, El-Tahan HR. Water distillation tower: Experimental investigation, economic assessment, and performance prediction using optimized machine-learning model, J. Cleaner Production 2023;388:135896. https://doi.org/10.1016/j.jclepro.2023.135896

[44] Khajavi H, Rastgoo A. Improving the prediction of heating energy consumed at residential buildings using a combination of support vector regression and meta-heuristic algorithms, Energy 2023;272:127069. https://doi.org/10.1016/j.energy.2023.127069

[45] Hasanien HM, Alsaleh I, Tostado-Véliz M, Zhang M, Alateeq A, Francisco Jurado F, et al. Hybrid particle swarm and sea horse optimization algorithm-based optimal reactive power dispatch of power systems comprising electric vehicles, Energy 2024;286:129583. https://doi.org/10.1016/j.energy.2023.129583

[46] Modest MF, Mazumder S. Radiative Heat Transfer. 4th ed. Amsterdam: Academic Press/Elsevier; 2021.

[47] Zhang ZM, Nano/Microscale Heat Transfer. 2nd ed. Cham: Springer; 2020.

[48] Moharam MG, Grann EB, Pommet DA, Gaylord TK. Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. J Opt Soc Am A 1995;12(5):1068–76. https://doi.org/10.1364/JOSAA.12.001068

[49] Lee BJ, Chen Y-C, Zhang ZM. Transmission enhancement through nanoscale metallic slit arrays from the visible to mid-infrared. J Comput Theor Nanosci 2008;5:201–13. https://doi.org/10.1166/jctn.2008.2461

[50] Zhao B, Zhang ZM. Study of magnetic polaritons in deep gratings for thermal emission control. J Quant Spectrosc Radiat Transf 2014;135:81–9. https://doi.org/10.1016/j.jqsrt.2013.11.016

[51] Quinlan JR. Introduction of decision trees. Machine Learning 1986;1:81–106. https://doi.org/10.1007/BF00116251

[52] Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, et al., AutoGluon-Tabular: Robust and accurate AutoML for structured data. Preprint arXiv:2003.06505, 2020. https://doi.org/10.48550/arXiv.2003.06505

[53] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Machine Learning Res 2011;12:2825–30. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[54] Cover, T, Hart, P. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7. http://doi.org/10.1109/TIT.1967.1053964

[55] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989;2(5):359–66. https://doi.org/10.1016/0893-6080(89)90020-8

[56] Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon 2018;4(11):e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

[57] Clark K, Luong M-T, Manning CD, Le QV. Semi-supervised sequence modeling with cross-view training. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, Oct 31 – Nov 4, 2018, pp. 1914–25. https://aclanthology.org/D18-1217/

[58] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32 (NeurIPS 2019), edited by H. Wallach et al. Preprint arXiv:1912.01703, 2019. https://arxiv.org/pdf/1912.01703.pdf

[59] Dietterich TG. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science 1857, edited by J.

Kittler and F. Roli, Springer, 2000, pp. 1–15. https://link.springer.com/chapter/10.1007/3-540-45014-9_1

[60] Cruz-Duarte JM, Amaya I, Ortiz-Bayliss JC, Conant-Pablos SE, Terashima-Marín H. A primary study on hyper-heuristics to customise meta heuristics for continuous optimization. 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/CEC48606.2020.9185591.

[61] Cruz-Duarte JM, Amaya I, Ortiz-Bayliss JC, Terashima-Marín H, Shi Y. CUSTOMHyS: Customising optimisation metaheuristics via hyper-heuristic search. SoftwareX 2020;12:100628. https://doi.org/10.1016/j.softx.2020.100628

[62] Nelson J. The Physics of Solar Cells. London: Imperial College Press; 2003.

[63] Watanabe S, Hori T, Le Roux J, Hershey JR. Student-teacher network learning with enhanced features. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 5275–79, 10.1109/ICASSP.2017.7953163

[64] Joseph VR., Optimal ratio for data splitting. Stat Anal Data Min 2022;15(4):531–8. https://doi.org/10.1002/sam.11583

Table 1

Comparison of the four surrogate models in terms of the training time, prediction time, and their RMSE on the test data set. All models are trained on a set of 16 million individual data points. The RCWA prediction/computation time is listed for comparison.

Surrogate model	Training time (s)	Prediction Time (s)	RMSE
RCWA	N/A	0.12	0
Decision tree	110	4.15E−07	0.0319
k-NN	31800	5.06E−06	0.0575
FCNN	72000	2.10E−05	0.0045
hybrid-NN	54000	5.09E−04	0.0062

Table 2

Grating parameters optimized for the TPV system with an emitter at 2000 K. The performance of planar tungsten and the grating optimized based on the ideal emissivity function are also included. These gratings are identified as A (ideal emissivity), B (optimal efficiency), and C (optimal power). However, the geometric parameters for B and C are coincidently identical.

Optimization Objective	Period (μm)	Depth (μm)	Filling ratio (-)	Efficiency (%)	Maximum power (W/cm^2)
Ideal emissivity (A)	1.11	2.00	0.14	26.90	7.16
Optimal efficiency (B)	1.11	2.00	0.10	27.06	7.19
Optimal power (C)	1.11	2.00	0.10	27.06	7.19
Planar tungsten	-	-	-	24.24	3.38

Table 3

Grating parameters optimized for the TPV system with an emitter at 1500 K. The efficiency and maximum power with the planar tungsten and grating A are also listed. The grating with optimized efficiency is identified as D and that with optimal power is identified as E.

Optimization Objective	Period (μm)	Depth (μm)	Filling ratio (-)	Efficiency (%)	Maximum power (W/cm^2)
Ideal emissivity (A)	1.11	2.00	0.14	21.60	1.21
Optimal efficiency (D)	0.87	0.30	0.85	22.10	1.07
Optimal power (E)	0.30	2.00	0.63	16.89	1.25
Planar tungsten	-	-	-	19.11	0.51

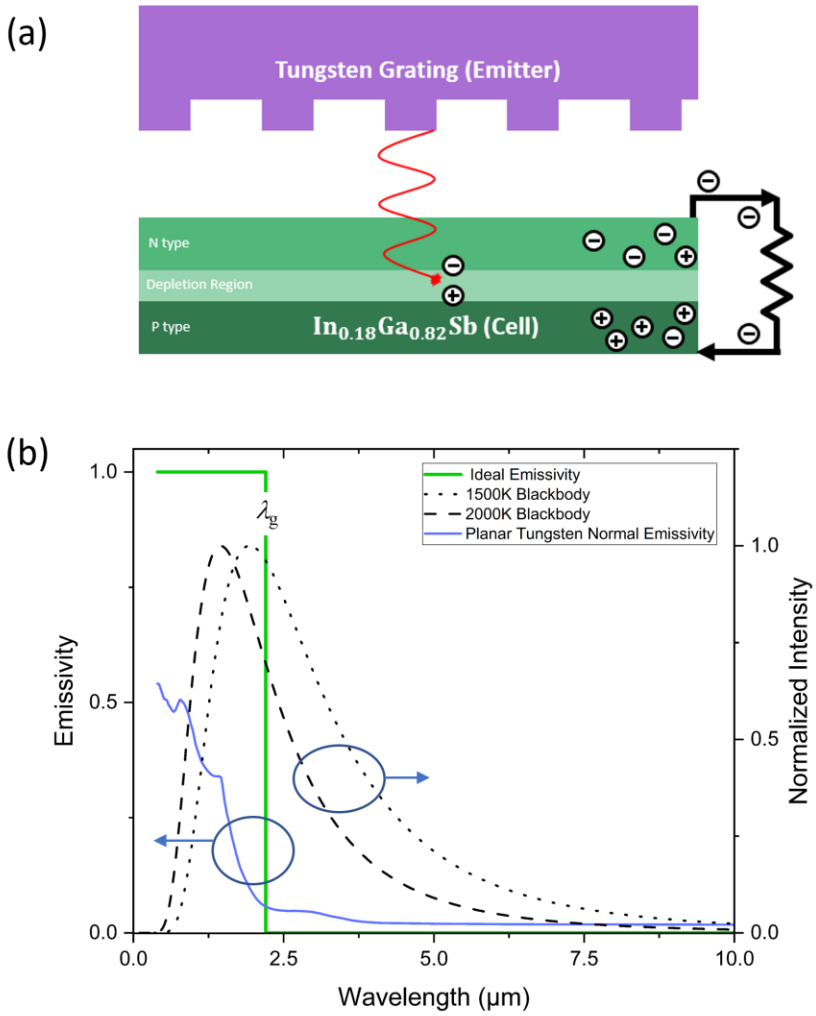

Figure 1. (a) Schematic of the thermophotovoltaic system composed of a binary 1D tungsten grating paired with an InGaSb cell. (b) The spectra for the ideal emissivity function, the normalized spectral intensities for blackbodies at 1500 K and 2000 K, and the normal emissivity of a planar tungsten surface.

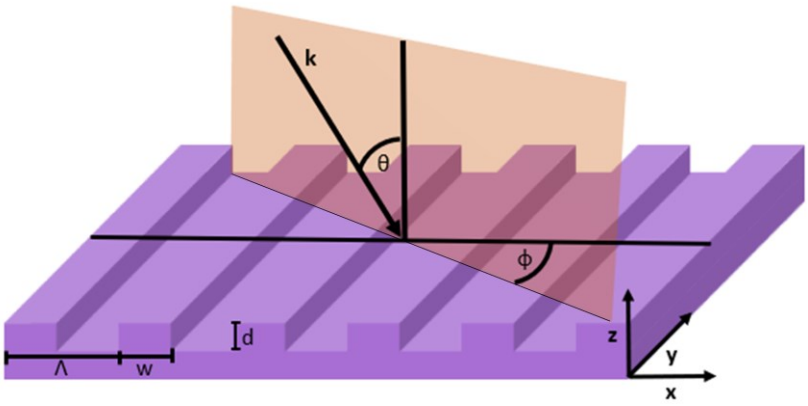

Figure 2. Schematic of a 1D binary grating characterized by three parameters: the period (Λ), the depth (d), and the width (w) or filling ratio ($f = w/\Lambda$). The emissivity is obtained from the calculated reflectance for given incident wavevector (**k**) whose direction is defined by a zenith angle (θ) and an azimuthal angle (ϕ).

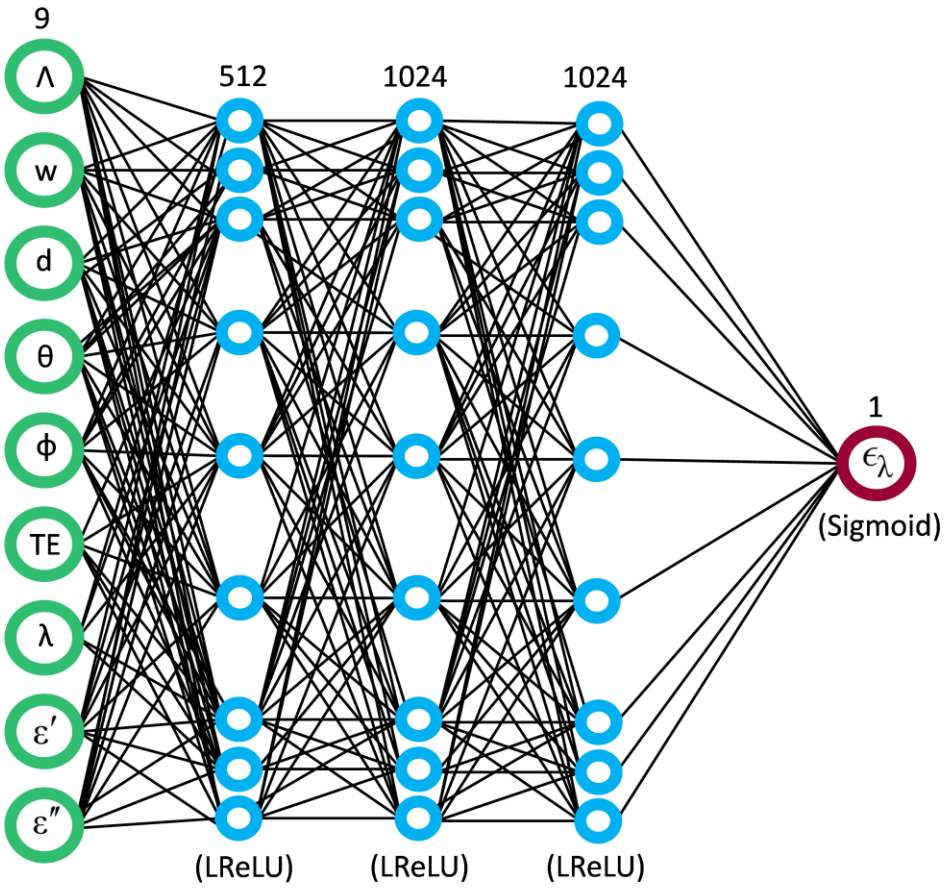

Figure 3. FCNN model architecture for RCWA surrogating. The three hidden layers use a LReLU activation function, while the output neuron uses sigmoid.

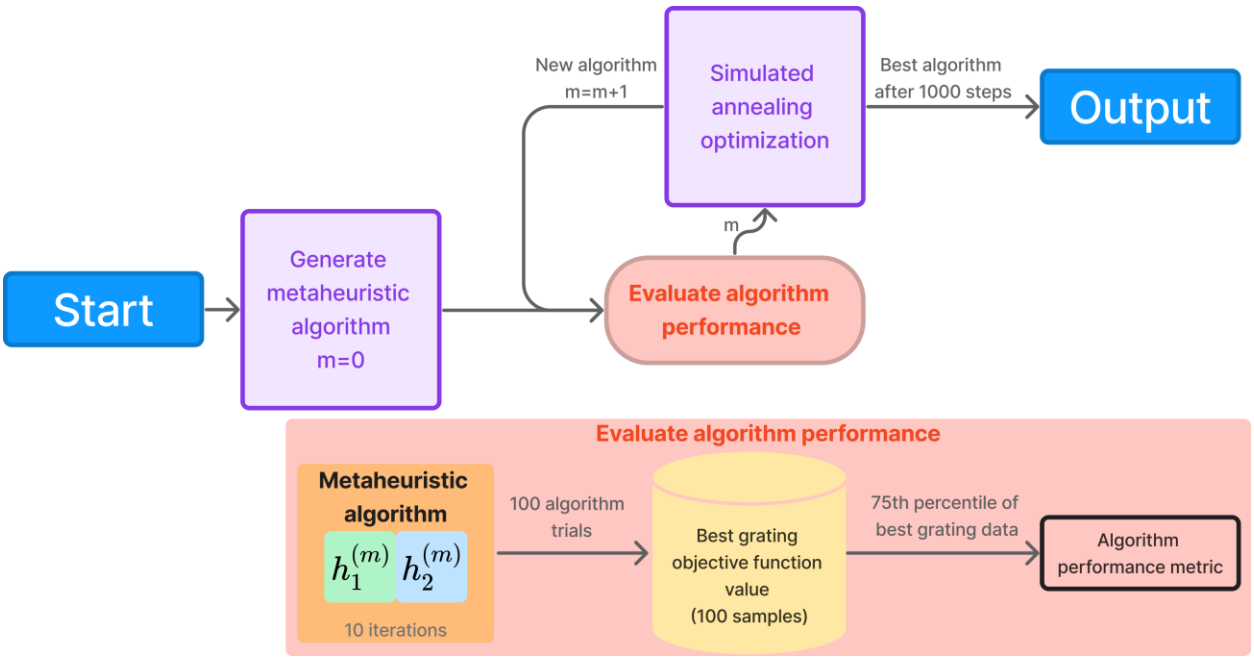

Figure 4. Hyper-heuristic framework for the creation of new or custom optimization algorithms. The stimulated annealing optimization generates new algorithms and is run 1000 times. The performance of the best grating is based on the 75th percentile of the objective function from 100 optimization runs. Each run takes 10 iterations from a random starting point. The data are then used to generate the algorithm's minimum expected performance 75% of the time. This information is then fed back into a simulated annealing algorithm to help the subsequent generation of new algorithms.

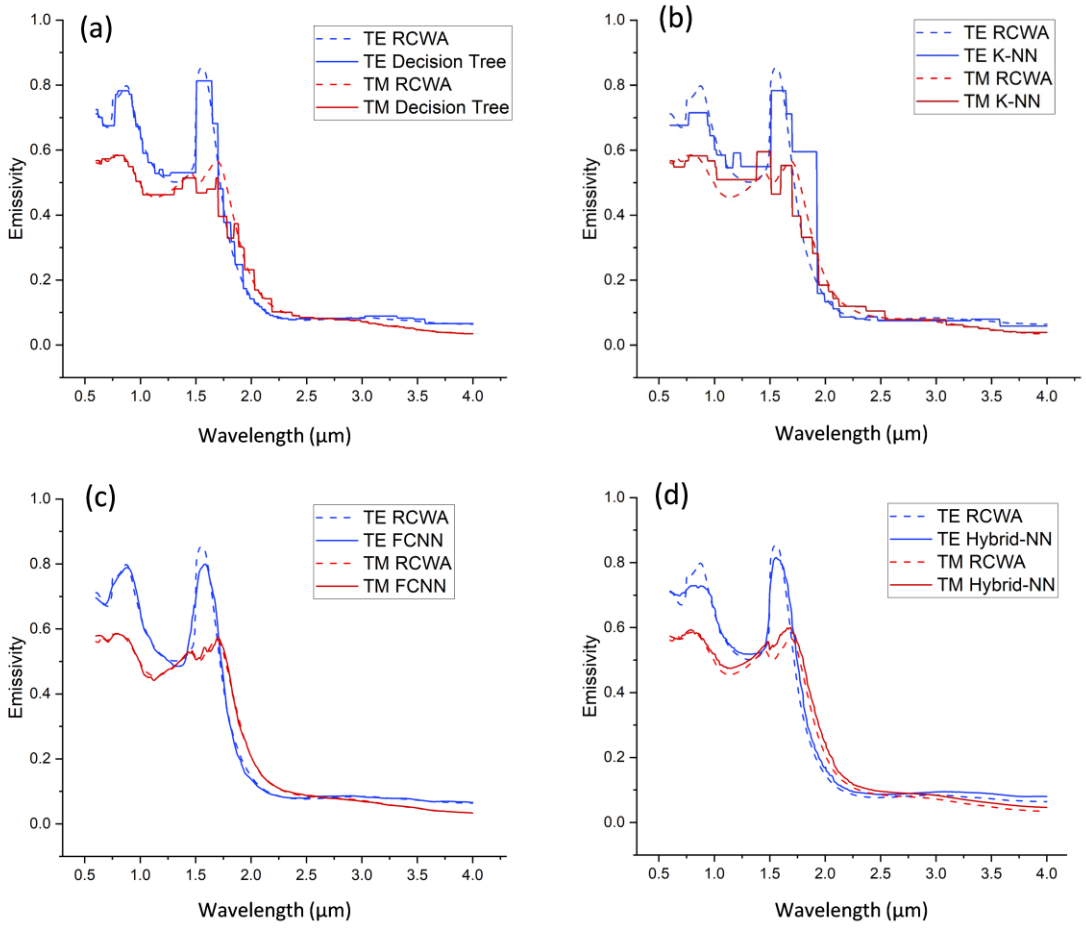

Figure 5. Comparison of the normal emissivity spectra between the surrogate model and RCWA calculations for an example grating with $\Lambda = 1.5$ μm, $f = 0.33$, and $d = 1.0$ μm. (a) decision tree model with a depth of 25, (b) k-NN model with k = 10 with uniform weighting, (c) FCNN surrogate model with three hidden layers, and (d) hybrid-NN model.

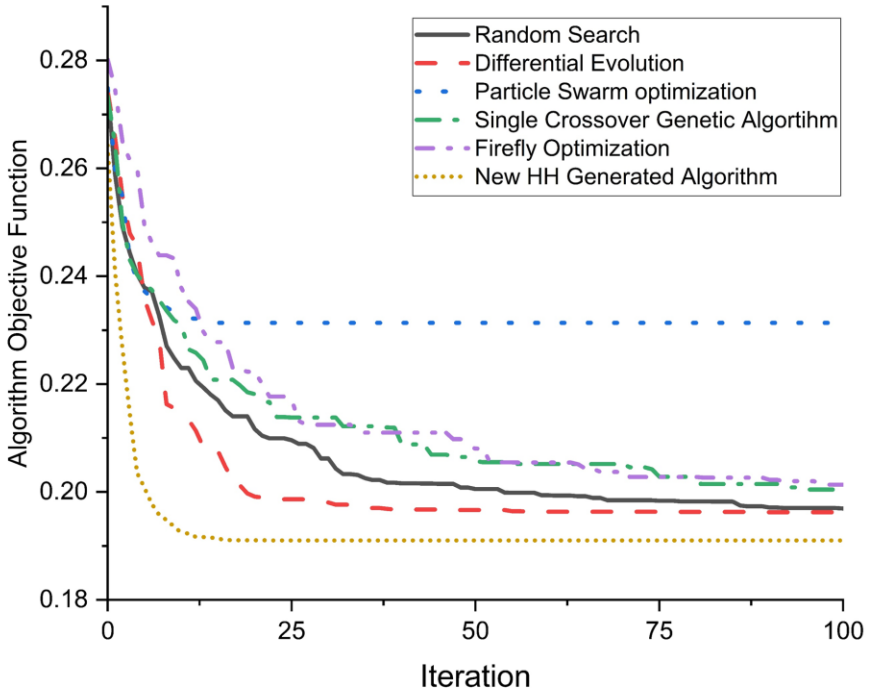

Figure 6. Comparison of the newly generated custom algorithm using the hyper-heuristic search to common optimization algorithms. Each iteration represents a population of eight solutions.

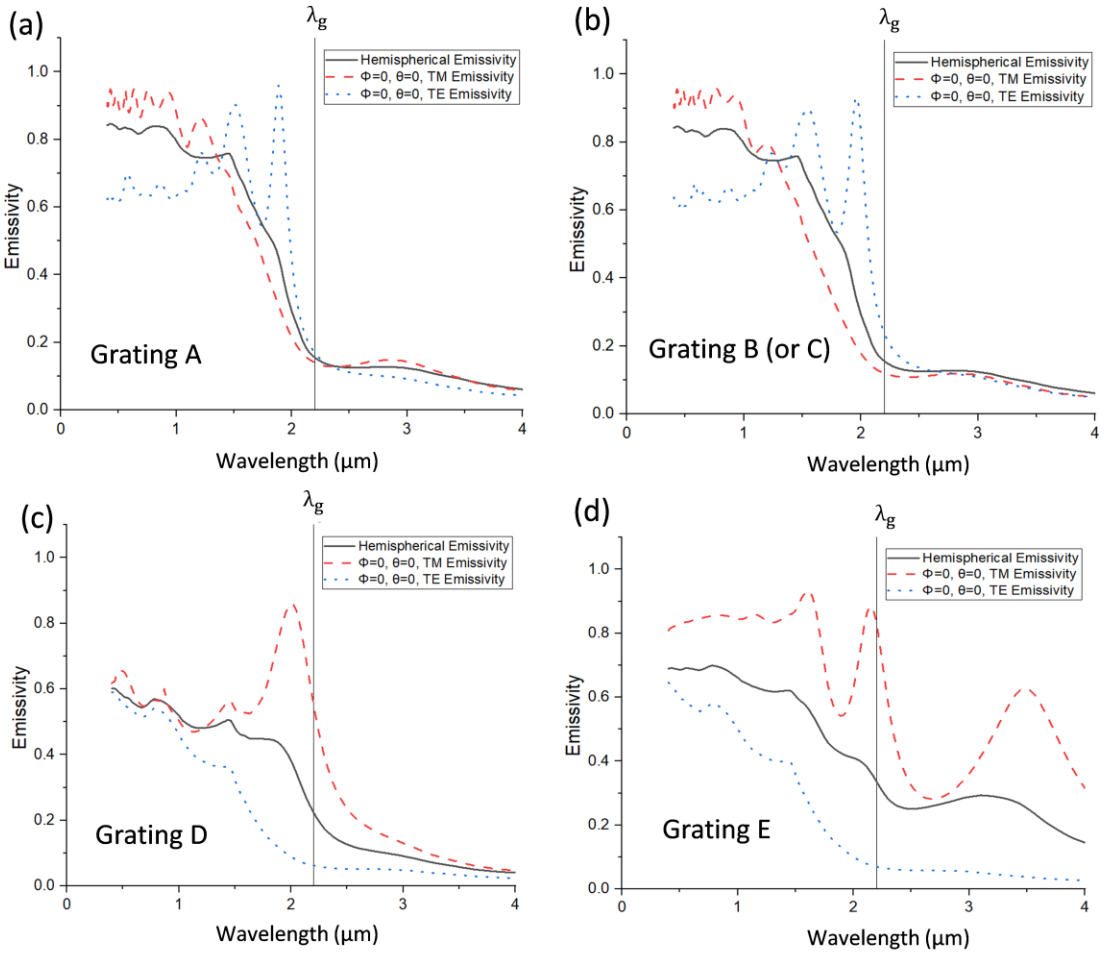

Figure 7. Normal and hemispherical spectral emissivity of the optimized gratings whose parameters are described in Tables 2 and 3. (a) Grating A optimized base on the ideal emissivity spectrum; (b) Grating B or C optimized for TPV performance with an emitter at 2000 K; (c) Grating D optmized for TPV efficiency for an emitter at 1500 K; (d) Grating E optimized for TPV output power for an emitter at 1500 K.

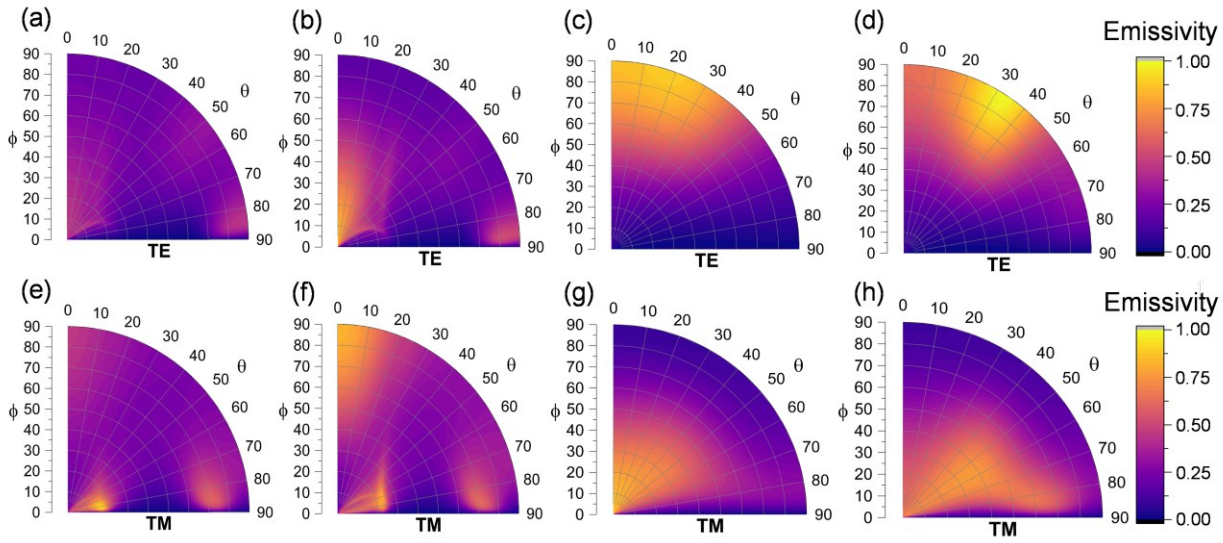

Figure 8. Calculated directional emissivity contour at $\lambda = 2$ μm: (a,b,c,d) TE waves for gratings A, B, D, and E, respectively; (e,f,g,h) TM waves for gratings A, B, D, and E, respectively.

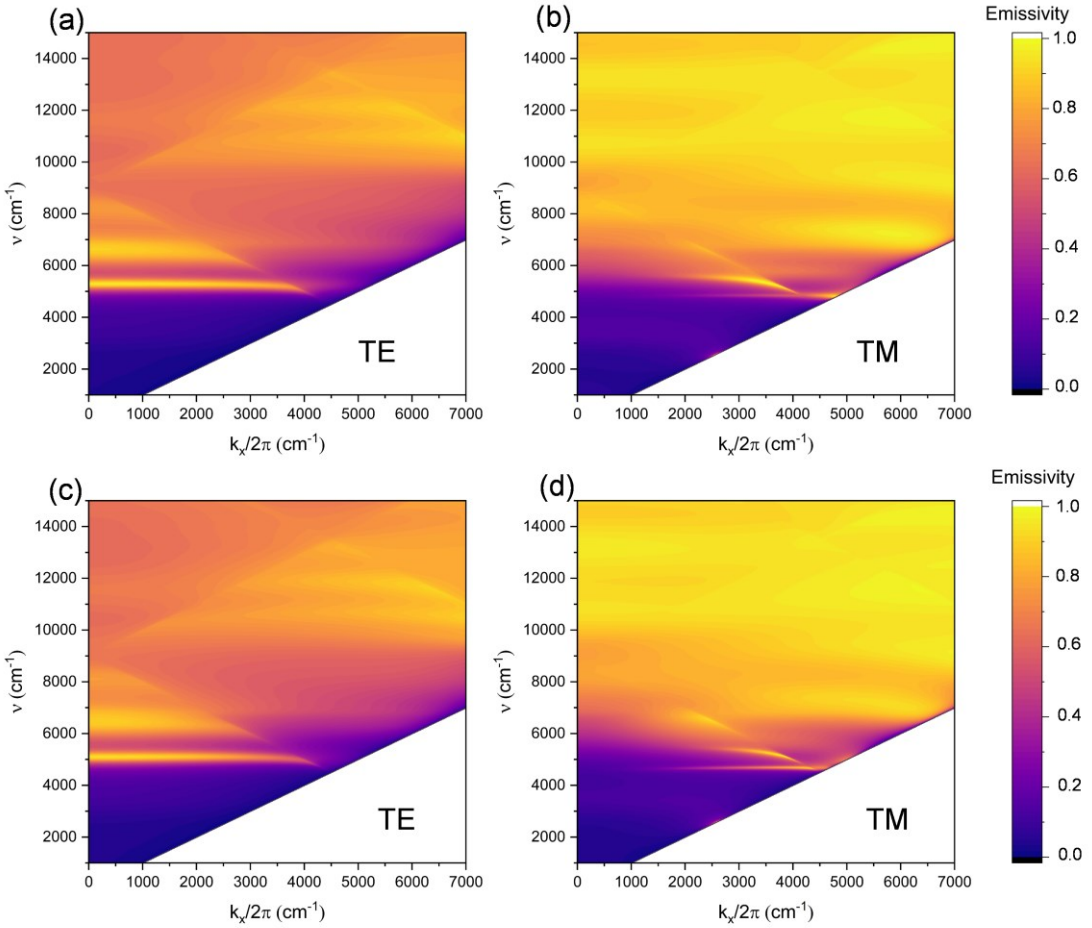

Figure 9. Spectral emissivity contour for gratings A and B for $\phi = 0°$. Note that ν is the wavenumber and $k_x = k_0 \sin\theta$ is the x-component of the wavevector in vacuum. (a) grating A for TE wave; (b) grating A for TM wave; (c) grating B for TE wave; (d) grating B for TM wave.

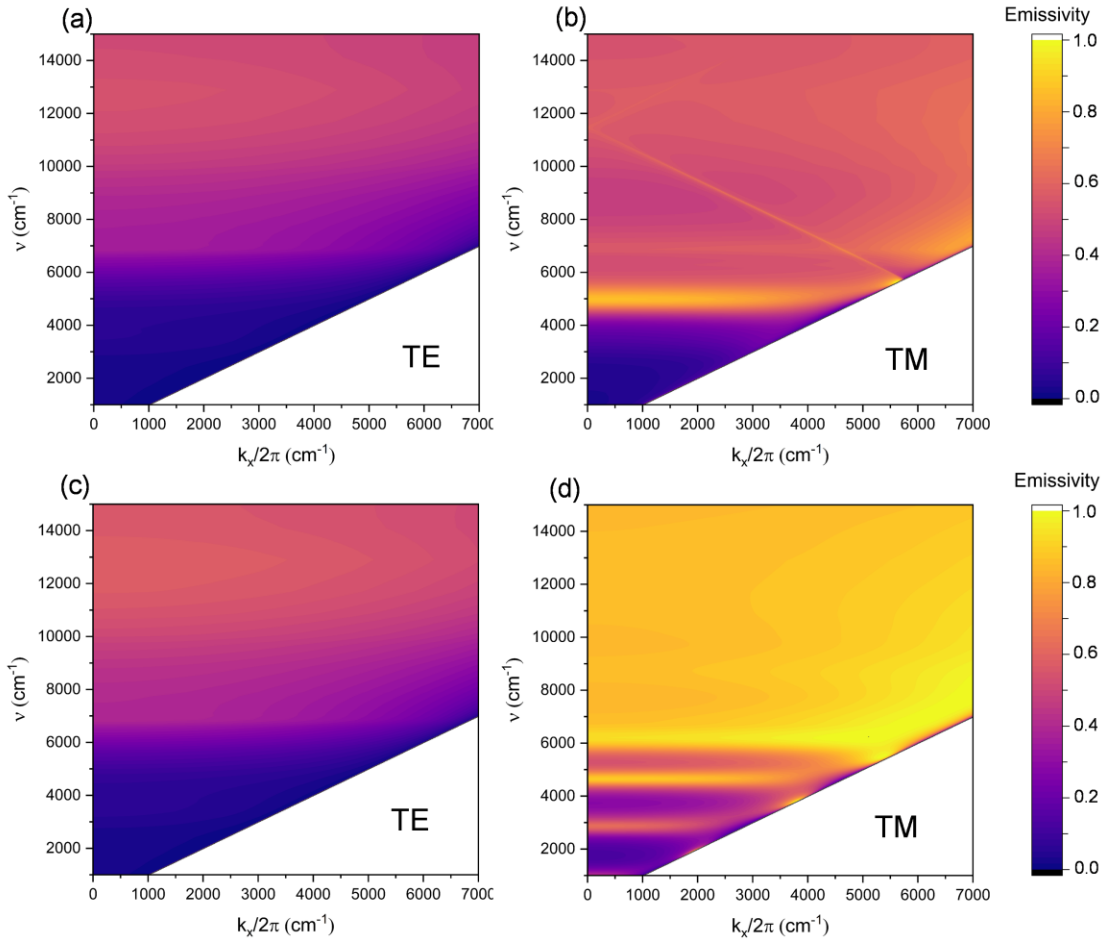

Figure 10. Spectral emissivity contour for gratings D and E for $\phi = 0°$. Note that ν is the wavenumber and $k_x = k_0 \sin\theta$ is the x-component of the wavevector in vacuum. (a) grating D for TE wave; (b) grating D for TM wave; (c) grating E for TE wave; (d) grating E for TM wave.