Multi-objective optimization for retinal photoisomerization models with respect to experimental observables

Cite as: J. Chem. Phys. (in press) (2021); https://doi.org/10.1063/5.0060259
Submitted: 16 June 2021 • Accepted: 23 November 2021 • Accepted Manuscript Online: 23 November 2021

Rodrigo A. Vargas-Hernández, Chern Chuang and Paul Brumer

Accepted to J. Chem. Phys. 10.1063/5.0060259

Multi-objective optimization for retinal photoisomerization models with respect to experimental observables

Rodrigo A. Vargas-Hernández,[1,2] Chern Chuang,[1] and Paul Brumer[1]
[1)] *Chemical Physics Theory Group, Department of Chemistry, University of Toronto,Toronto, Ontario, M5S 3H6, Canada*
[2)] *Vector Institute for Artificial Intelligence, Toronto, Ontario M5S 1M1, Canada.*

(*Electronic mail: r.vargashernandez@utoronto.ca)

(Dated: 22 November 2021)

The fitting of physical models is often done only using a single target observable. However, when multiple targets are considered, the fitting procedure becomes cumbersome, there being no easy way to quantify the robustness of the model for all different observables. Here, we illustrate that one can jointly search for the best model for each desired observable through multi-objective optimization. To do so we construct the Pareto front to study if there exists a set of parameters of the model that can jointly describe multiple, or all, observables. To alleviate the computational cost, the predicted error for each targeted objective is approximated with a Gaussian process model, as it is commonly done in the Bayesian optimization framework. We applied this methodology to improve three different models used in the simulation of stationary state $cis-trans$ photoisomerization of retinal in rhodopsin, a significant biophysical process. Optimization was done with respect to different experimental measurements, including emission spectra, peak absorption frequencies for the *cis* and *trans* conformers, and the energy storage. Advantages and disadvantages of previously proposed models are exposed.

I. INTRODUCTION

In chemical physics, observables such as absorption spectra are often simulated with closed-form model Hamiltonians or Liouvillians. Examples studied in detail here are models of $cis-trans$ photoisomerization of retinal in rhodopsin. The parameters of such models, $\boldsymbol{\theta}$, are usually fit by minimizing a scalar error function, $f(\boldsymbol{\theta}) : \mathbb{R}^d \rightarrow \mathbb{R}$, that quantifies the accuracy of the model with respect to a target set of observables, where d is the number of parameters. Optimization of $f(\boldsymbol{\theta})$ can be achieved by gradient-based methods or by sampling algorithms that self-learn from previous iterations, e.g., genetic algorithms[1] and Bayesian optimization (BO)[2,3]. For observables that depend on quantum dynamics calculations, gradient-based methods are rarely used, due to inaccuracies in approximating the gradient of $f(\boldsymbol{\theta})$ with respect to the free parameters of the model Hamiltonian or Liouvillian. By contrast, sampling algorithms have proven to be robust search tools in chemical physics. For example, BO has been used to optimize density functionals,[4,5] generate low energy molecular conformers,[6,7] and inverse design potential energy surfaces for reactive molecular systems[8]. BO has also been successfully applied for screening chemical compounds[9–11], minimizing the energy of the Ising model[12], and has recently been used to optimize laser pulses for molecular control[13,14] .

Much effort has been devoted to mapping the search for a physical model into an optimization problem. However, in chemical physics comparing different models with respect to a single observable can lead to overfitting. To account for more realistic physical models the search-space must include M-different observables with $M > 1$. One common strategy is to adopt linear combinations of the M individual error functions, $\{f_1, \cdots, f_i, \cdots, f_M\}$, that account for the discrepancy between the model's prediction and that of each of the target observables,

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_i^M w_i f_i(\boldsymbol{\theta}), \tag{1}$$

However, mapping the search of a physical model using Eq. 1 assumes that (i) the values of the linear coefficients, w_i's, are known, and (ii) there exists a single optimal minimum that jointly describes all observables. Both are extremely strong assumptions. Furthermore, the possibility of having multiple local minima strongly depends on how the values of w_i are chosen.

Such difficulties can be circumvented by not assuming a scalar function (a certain set of w_i's) to quantify the accuracy of the model but instead by working with a vector function whose elements are the individual error functions: $\mathbf{F}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}), \cdots, f_M(\boldsymbol{\theta})]$. Here, we propose that deeper insights can be gained about the model-space by learning the boundaries of $\mathbf{F}(\boldsymbol{\theta})$ at different values of $\boldsymbol{\theta}$ and by reformulating the problem in terms of a multi-objective optimization scheme. Multi-objective optimization has been used to guide the experimental search for materials with desired properties[15–17] and to optimize chemical reactions.[18,19] This procedure was also used to fit the potential energy surface for diatomic molecules from their rotational and vibrational spectra.[20]

In this paper we apply multi-objective optimization to examine, and improve, three proposed models of retinal photoisomerization, the first step in vision. These models have generally been considered for short time dynamics arising from pulsed laser studies. However, as we have previously shown[21–23] the fact that nature operates with

Accepted to J. Chem. Phys. 10.1063/5.0060259

incoherent (solar) radiation implies that we should focus on properties in the stationary state, the approach adopted below. In particular, we examine, using previously underutilized stationary state data, the extent to which these models can provide a minimal representation of retinal isomerization in the naturally-occurring steady-state. The Pareto front results identify advantages and disadvantages of each model.

The paper is organized as follows: Section II provides an introduction t o the central tool used in this work, an algorithm to construct the Pareto front. In Section III, we apply the method to three different models for retinal photoisomerization. Sections IV and V provide results and the discussion, respectively.

II. MULTI-OBJECTIVE OPTIMIZATION

In this section we present a brief introduction to multi-objective optimization and to the concept of the Pareto front (PF).

A. Pareto Front

Multi-objective optimization considers the optimization of multiple objective functions and is usually formulated as seeking a minimum of the vector function;

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbf{F}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{\theta}} \; [f_1(\boldsymbol{\theta}), \cdots, f_M(\boldsymbol{\theta})] \quad (2)$$

Since it is unknown if a single $\boldsymbol{\theta}$ could jointly minimize all f_i's, we must consider all different values where components of $\mathbf{F}$ are minimal. To quantify the improvement of $\mathbf{F}$, we introduce the concept of *Pareto dominant points*, $\boldsymbol{\theta}_\ell$, defined as,

1. $f_i(\boldsymbol{\theta}_\ell) \leq f_i(\boldsymbol{\theta})$, for all f_i's in $\mathbf{F}$.
2. $f_i(\boldsymbol{\theta}_\ell) < f_i(\boldsymbol{\theta})$, for at least a single f_i in $\mathbf{F}$.

$Y_P = \{\mathbf{F}(\boldsymbol{\theta}_1), \cdots, \mathbf{F}(\boldsymbol{\theta}_\ell), \cdots, \mathbf{F}(\boldsymbol{\theta}_n)\}$ are the collection of all the dominant points in the function space, referred to as the PF, and $X_P = \{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_\ell, \cdots, \boldsymbol{\theta}_n\}$ are the corresponding points in the parameter space, where n is the number of Pareto dominant points.

B. Example of Pareto Front

As a pedagogical example of the PF consider the linear combination of the Fonseca-Fleming functions[39],

$$\mathcal{L}_{FF}(\mathbf{x}; \boldsymbol{\lambda}) = \lambda_1 f_1(\mathbf{x}) + \lambda_2 f_2(\mathbf{x}), \quad (3)$$

where $f_1(\mathbf{x}) = 1 - e^{-\sum_{i=1}^{2}\left(x_i - \frac{1}{\sqrt{2}}\right)^2}$ and $f_2(\mathbf{x}) = 1 - e^{-\sum_{i=1}^{2}\left(x_i + \frac{1}{\sqrt{2}}\right)^2}$. As is evident, the function $\mathcal{L}_{FF}$ also depends on the linear combination coefficients, $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]$, influencing the location of the minimizer; $\mathbf{x}^* = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}; \boldsymbol{\lambda})$. f_1 and f_2 have different minimizers: for f_1 at $\mathbf{x}_1^* = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$, and for f_2 at $\mathbf{x}_2^* = \left[-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]$.

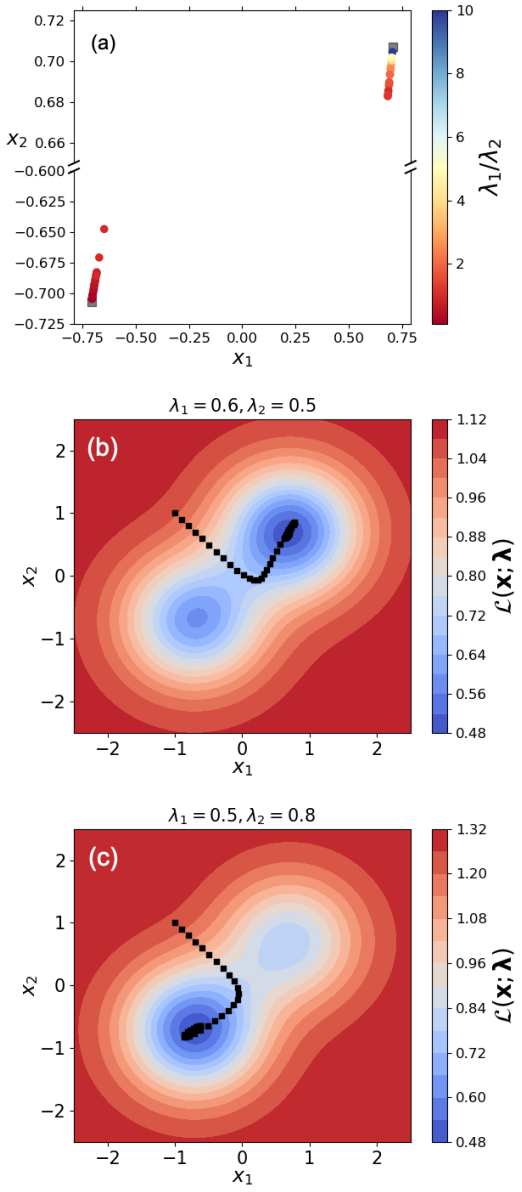


FIG. 1. (a) The symbols represent the minimizer of $\mathcal{L}_{FF}(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{2} \lambda_i f_i(\mathbf{x})$, as a function of λ_1/λ_2, and the black-square symbols are the minimizers, $\mathbf{x}^*$, for each individual f_i function. (b) and (c) Function $\mathcal{L}_{FF}(\mathbf{x}, \boldsymbol{\lambda})$, for the sample (b) $\lambda_1 = 0.6$ and $\lambda_2 = 0.5$, and (c) $\lambda_1 = 0.5$ and $\lambda_2 = 0.8$. The black symbols represent the optimization trajectory with gradient descent started with a random $\mathbf{x}$.

From Figure 1, we can observe that by varying the values of λ_1 and λ_2, the location of $\mathbf{x}^*$ moves between $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$, the minimizers of f_1 and f_2. Tuning the values $\boldsymbol{\lambda}$ is not a trivial process, but we could study the optimization of multiple functions by means of multi-objective op-

timization, Eq. 2. For the Fonseca-Fleming functions example, the Pareto dominant points connect both minima of the f_1 and f_2 functions, as these points describe where, in the function-space, f_1 and f_2 are minimal. Figure 2 depicts, in part (a), a set of points (blue) obtained by evaluating $[f_1(\mathbf{x}), f_2(\mathbf{x})]$ at various points $\mathbf{x}$. The Pareto dominant points $\mathbf{x}'$ generate a set of $[f_1(\mathbf{x}'), f_2(\mathbf{x}')]$ that give the black markers in Figure 2(a), which form the PF. In principle, since we seek minimization, the PF is the relevant quantity (a). However, the other (blue) points are shown to give an overview of the $[f_1(\mathbf{x}), f_2(\mathbf{x})]$ for all $\mathbf{x}$. The Pareto dominant points are shown as black markers in Figure 2 (b).

The PF, for this example and those obtained for photoisomerization models, were found using the multi-objective Bayesian optimization (MOBOpt) algorithm[24,25], explained below.

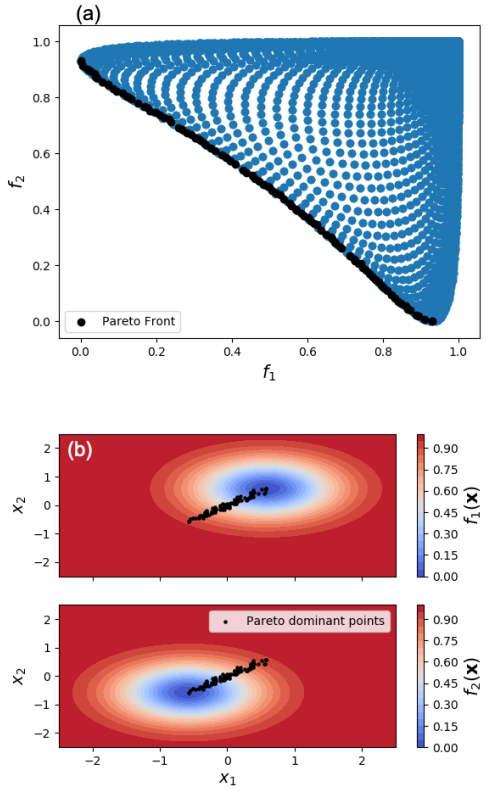


FIG. 2. The black markers represent the PF for the Fonseca-Fleming functions found with the MOBOpt algorithm. (a) The blue markers are functional values of both f_1 and f_2 functions, and (b) the $\mathbf{x}$ value for each Pareto point.

C. Optimization Algorithm

The optimization algorithm relies on approaches described in references[25,27,29]. We provide a brief overview of this algorithm in this section.

For multi-objective optimization protocol, the function $\boldsymbol{F}$ has multiple elements, $\boldsymbol{F} = [f_1(\boldsymbol{\theta}), \cdots, f_M(\boldsymbol{\theta})]$, each individual one ($f_i$) describes a different task. One of the key advantages of machine-learning algorithms is the ability to approximate any function based on few data points. Here, we approximate each f_i with an individual Gaussian process (GP), as is commonly done in single-objective BO;

$$f_i \sim \mathcal{GP}_i(\mu_i, \Sigma_i), \tag{4}$$

where μ_i is the mean and Σ_i is the covariance of the model. Gaussian process prediction for a new point $\mathbf{x}^*$ is achieved by,

$$\mu_i(\mathbf{x}^*) = \mathbf{k}(\mathbf{x}^*, \mathbf{X})^\top \left[K(\mathbf{X}_i, \mathbf{X}_i) + \sigma_n 1\right]^{-1} \mathbf{y}_i \tag{5}$$

$$\sigma_i(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) \left[K(\mathbf{X}_i, \mathbf{X}_i) + \sigma_n 1\right]^{-1} \mathbf{k}(\mathbf{x}^*, \mathbf{X}_i), \tag{6}$$

where $\mu_i(\mathbf{x}^*)$ and $\sigma_i(\mathbf{x}^*)$ are the mean and standard deviation of the posterior distribution for f_i. $\mathbf{X}_i$ (and $n \times d$ dimensional matrix) and $\mathbf{y}_i$ (an $n-$ component vector) are the training points for the task i. σ_n represent the noise for each training observations; all of our simulations are noiseless, meaning $\sigma_n = 0$. Here $K(\cdot, \cdot)$ is the covariance matrix, whose matrix elements are computed using the kernel function, $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. The entries of $\mathbf{k}(\mathbf{x}^*, \mathbf{X})$ are between $\mathbf{x}^*$ and all the training data $\mathbf{X}$. For all calculations in this work we used the Matern 2/5 kernel,

$$k_{Mat}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left[1 + \sqrt{5} r(\boldsymbol{x}_i, \boldsymbol{x}_j) + \frac{5}{3} r^2(\boldsymbol{x}_i, \boldsymbol{x}_j)\right] \exp^{-\sqrt{5} r(\boldsymbol{x}_i, \boldsymbol{x}_j)}, \tag{7}$$

where $r^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^\top M (\boldsymbol{x}_i - \boldsymbol{x}_j)$, and M is a diagonal matrix that contains different length scale constants, $\boldsymbol{\ell} = \{\ell_i\}_{i=1}^d$, for each dimension of $\boldsymbol{x}$. Its dimension, d, is the number of parameters. The kernel parameters for each $\mathcal{GP}_i$ are different given that they are constructed with different data points. The optimal kernel parameters ($\{\ell_i^*\}_{i=1}^d$) were found by maximizing the log-marginal likelihood,

$$\begin{aligned} \boldsymbol{\ell}^* &= \arg\max_{\boldsymbol{\ell}} \ \log p(\mathbf{y}|\mathbf{X}) \\ &= \arg\max_{\boldsymbol{\ell}} \left[-\frac{1}{2}\mathbf{y}^\top K(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y} - \frac{1}{2}\log|K(\mathbf{X}, \mathbf{X})| - \frac{n_0}{2}\log 2\pi\right], \end{aligned} \tag{8}$$

where n_0 is the number of training points for each individual Gaussian processes. For more details about GPs we refer the reader to Ref.[27].

The second key component in BO is the acquisition function that guides the samples through the optimization towards the minimum of the scalar function under study. For multi-objective BO optimization, the acquisition function must describe how the PF change by sampling all the objective functions at different $\boldsymbol{\theta}$. At each iteration, we used the manifold of $\{\mathcal{GP}_i\}_{i=1}^M$ to infer $X_P(t)$ and $Y_P(t)$, where t denotes the iteration number in the

optimization procedure, using the non-dominated sorting genetic algorithm II (NSGAII)[28]. At each iteration, the goal is to sample a new point that would make $X_P(t)$ and $Y_P(t)$ a more accurate approximation of the Pareto front. The proposed samples are suggested by,

$$\mathbf{x}_{next} = \arg\max_{\ell \in 1,\cdots,n} \left[q \left(\frac{d_{\ell,f} - \mu_f}{\sigma_f} \right) + (1-q) \left(\frac{d_{\ell,x} - \mu_x}{\sigma_x} \right) \right], \quad (9)$$

where the first and second terms describe the distance between $Y_P(t)$ and $\{\mathbf{y}_i\}_{i=1}^M$, and $X_P(t)$ and $\mathbf{X}$; $\{\mathbf{y}_i\}_{i=1}^M$ are the training data for all M objectives. At each iteration, the suggested point $\mathbf{x}_{next}$ is evaluated for all M objectives, making the training data $\mathbf{X}_i$ the same for all M objectives. $d_{\ell,f}$ and $d_{\ell,x}$ are the L-2 norm between the point ℓ in $Y_P(t)$ and $X_P(t)$ respectively, and all training points at iteration t. μ_x and σ_x are the average and standard deviation for all $d_{\ell,x}$. Eq. 9 ensures that we sample points closer to the Pareto front to increase the accuracy of all M Gaussian processes close to the real Pareto front. q is a hyperparameter associated with exploration-exploitation; $q \in [0,1]$. If $q \approx 1$, we will sample more points close to the approximate PF (exploitation), and when $q \approx 0$ the points are sampled from the parameter space where the functions are minimal (exploration). Lastly, at each iteration all $\{\mathcal{GP}_i\}_{i=1}^M$ are retrained by including in the training data $\mathbf{x}_{next}$ and $\{\mathbf{y}_i\}_{i=1}^M$, the suggested point evaluated for all f_i's ($y_i = f_i(\mathbf{x}_{next})$). In all cases discussed below, we found that after 150 iterations the PF did not change significantly.

For future references we denote this algorithm as the multi-objective Bayesian optimization (MOBOpt) algorithm. For more details consult Refs.[25,29].

III. $cis - trans$ ISOMERIZATION MODELS

Here we introduce the physical models used to simulate (a) the emission spectra and (b) the maximum absorption frequencies for the *cis* and *trans* retinal conformers, and (c) the energy storage, the energy difference between these two conformers in rhodopsin. We then focus on the extent to which these models can represent observed data. The primary motivation of this study is to identify important features of retinal in the steady state. We are cognizant of the fact that retinal itself spans a 49 dimensional potential surface, so we cannot anticipate quality agreement of these models with experimental data. However, the extent to which optimized results do or do not agree with experimental data points to required directions for improved modeling.

A. Single and two-mode models

The numerical simulation of minimal models has been a fruitful way of studying *11-cis*→all-*trans* photoisomerization of retinal in rhodopsin in the last decades. Some of the most well known models rely on a parametrized form of the two lowest electronic energy states to describe the isomerization[30–33]. Three of them are considered here. The simplest model uses a rotational coordinate ϕ and potentials

$$V_{0,0}^{SM} = \frac{1}{2} W_0 (1 - \cos\phi) \quad (10)$$

$$V_{1,1}^{SM} = E_1 - \frac{1}{2} W_1 (1 - \cos\phi), \quad (11)$$

where $V_{i,i}^{SM}$ are the diabatic potential energy surfaces (PESs) for the ground ($i = 0$) and excited ($i = 1$) electronic states. The coupling between different electronic states, $V_{i,j}$, is assumed to be constant, $V_{0,1} = V_{1,0} = \lambda$. For this model, the free parameters to be optimized are $\boldsymbol{\theta}_{SM} = [E_1, W_0, W_1, \lambda]$. We refer to this model as the *single-mode* (SM) model. By including an additional vibrational degree of freedom, x, each diabatic PES now includes the terms[30–33],

$$V_{0,0}^{TM} = V_{0,0}^{SM} + \frac{\omega}{2} x^2 \quad (12)$$

$$V_{1,1}^{TM} = V_{1,1}^{SM} + \frac{\omega}{2} x^2 + \kappa x. \quad (13)$$

We refer to this model as the *two-mode* (TM) model, and the coupling between the ground and excited state is defined as $V_{0,1}^{TM} = V_{1,0}^{TM} = \lambda x$. The free parameters in this model are $\boldsymbol{\theta}_{TM} = [E_1, W_0, W_1, \kappa, \lambda]$. For all the calculations presented here, we fixed the value of ω to 0.19 eV, corresponding to literature fits to short time dynamics.

We also consider a third model inspired by (but not the same as) the two-state-three-mode model reported in Ref.[34], where

$$V_{0,0}^{Omod} = \frac{\omega}{2} x^2 - W_0 \cos\phi - W_1 \cos(2\phi) \quad (14)$$

$$V_{1,1}^{Omod} = E_1 + \frac{\omega}{2} x^2 + \kappa x - W_2 \cos\phi + W_3 (1 - \cos(2\phi)) \quad (15)$$

$$V_{0,1}^{Omod} = V_{1,0}^{Omod} = \lambda \sin(2\phi). \quad (16)$$

Here $V_{i,i}^{Omod}$ are the diabatic PESs for the ground and excited states, and $V_{0,1}^{Omod}$ is the coupling between the two states. The free parameters of this model are $\boldsymbol{\theta}_{Omod} = [\omega, W_0, W_1, E_1, \kappa, W_2, W_3, \lambda]$.

The kinetic energy operators of these models are written as $T = -\frac{1}{2m}\frac{\partial^2}{\partial\phi^2} + \frac{\omega}{2}\frac{\partial^2}{\partial x^2}$, except for the SM model where only the first term exists. Here, we take $\hbar = 1$ and the total Hamiltonian is $H_s = V + T$. The parameters of all three models as reported in the literature are given in Table I.

In the following sections we present the PF for these three models with respect to different experimental measurements. They are three emission spectra excited at different wavelengths, the maximum absorption frequencies for the *cis* and *trans* conformers, and energy storage. The energy storage is defined as the energy difference between the lowest state in the *trans*-well ($\frac{\pi}{2} \leq \phi < \frac{3\pi}{2}$) and

that in the *cis*-well ($-\frac{\pi}{2} \leq \phi < -\frac{\pi}{2}$). These two states are also instrumental in defining the absorption spectra of the *cis*- and *trans*-conformers, while the specific definitions of the other five spectroscopic observables are given in the next section.

The aim of this analysis is twofold; first to compare the best sets of parameters for each model, and second to determine the extent to which the models can agree with experimental data. We emphasize, once again, that our interest lies in the stationary state that models natural excitation with sunlight[21–23]. All results were computed with the MOBOpt algorithm described in Section II C. Significantly, once having optimized the parameters in these models, any remaining deviation from experiment implies the need for improved models with additional functional forms. In the extreme case, it may imply the need for a basic atomistic model of retinal photoisomerization.

TABLE I. The literature value of the system parameters: *SM* and *TM* models are from Refs.[30–33], and *Omod* model is based on Ref.[34]. All parameters are given in [eV].

	SM	*TM*		*Omod*
E_1	2.0	2.58	ω	0.19
W_0	2.3	3.56	W_0	0.65
W_1	1.5	1.19	W_1	0.73
λ	0.065	0.19	E_1	1.68
ω	–	0.19	W_2	0.58
κ	–	0.19	W_3	-0.5
			κ	0.19
			λ	0.61
m^{-1}	1.43×10^{-3}	28.06×10^{-4}	m^{-1}	2.8×10^{-2}

B. System-Bath Couplings and Spectral Lineshape

In order to compare to experimentally measured spectra we take into account the interaction between the systems and their environment following the approach of Stock et al.[38] in writing the coupling between the system and a collection of harmonic modes accounting for the effects of the protein pocket and the vibrational degrees of freedom of retinal not included in the system Hamiltonian:

$$H_{\mathrm{s,b}} = \sum_{\mathrm{b}} S_{\mathrm{b}} \cdot \sum_{k} g_{\mathrm{b},k}(b^{\dagger}_{\mathrm{b},k} + b_{\mathrm{b},k}) \tag{17}$$

where S_{b} is a system operator, $g_{\mathrm{b},k}$ is the coupling strength of the kth mode of bath b, and $b^{\dagger}$ (b) is the corresponding creation (annihilation) operator of the bath mode.

The forms of S_{b} and the spectral density $J_{\mathrm{b}}(\omega') = \frac{\pi}{2}\sum_k g^2_{\mathrm{b},k}\delta(\omega'-\omega'_{\mathrm{b},k})$ depend on the model. We adopt the convention of Stock et al.[38] and take $S_\phi = (1-\cos\phi)|1\rangle\langle 1|$ for all three system models and $S_x = x|1\rangle\langle 1|$ for the *TM* and *Omod* models, and Ohmic bath spectral densities $J_{\mathrm{b}}(\omega') = \eta_{\mathrm{b}}\omega' \exp(-\omega'/\omega'_{\mathrm{b}})$ with parameters given in Table II.

TABLE II. The parameters for the system-bath coupling terms adopted in the calculation.

	SM	*TM*	*Omod*
η_x	–	0.1	0.1
η_ϕ	0.15	0.15	0.15
ω_x	–	ω	ω
ω_ϕ	$\sqrt{\frac{W_0}{2m}}$	$\sqrt{\frac{W_0}{2m}}$	$\sqrt{\frac{W_0+4W_1}{2m}}$

The effect of system-bath coupling is treated within the Markovian-Redfield framework, described elsewhere[21]. Since we are primarily interested in spectroscopic observables at the stationary state, we further employ the Bloch-secular approximation, decoupling the dynamics of populations and coherences in the system energy eigenbasis. The former can be expressed as the solution to the master equation $\dot{P} = W \times P$, where $P_i(t)$ is the population of system energy eigenstate i and W_{ij} is the scattering rate from states j to i due to the system-bath coupling. Under the secular approximation, the absorption spectrum is simply the superposition of all bright states broadened by a Lorentzian lineshape:

$$A_x(\epsilon) = \sum_i f_{x,i} \cdot \frac{|W_{ii}|}{(\epsilon-\epsilon_i)^2 + W_{ii}^2} \tag{18}$$

where $f_{x,i} = |\langle i|\hat{\mu}|x\rangle|^2$ is the oscillator's strength from state x (the ground state in the *cis*- or *trans*-wells) to state i with energy $\epsilon_i = \langle i|H_{\mathrm{s}}|i\rangle$, and $W_{ii} = -\sum_j W_{ij}$ is the diagonal term of the rate matrix ensuring the conservation of population under the Bloch-secular approximation. We note that while the decoupling of population dynamics with that of the coherences employed by the Bloch-secular approximation generally introduces error in the transient regime, we are primarily interested in the stationary state, where we expect it to yield reliable results.

C. Spontaneous Emission Spectra

We also examine the fluorescence emission spectra arising from the radiative decay of the excited state. To this end, in the master equation $\dot{P} = W \times P$ we also consider the rate kernel corresponding to a photon bath with $J^{\mathrm{em}}(\omega') = 4\pi^2\omega'^3/3\epsilon_0(2\pi c)^3$, the photon spectral density from Planck's law where ϵ_0 is the vacuum permittivity and c is the speed of light, coupled to the system[22]. The spontaneous emission spectrum of a given reduced system density matrix at frequency ω' is obtained by collecting temperature-independent terms in the corresponding rate at $\omega' \approx -\omega'_{kl}$. In other words, the intensity of the emission spectrum at a certain frequency is proportional to the magnitude of the generalized rates

connecting states with that frequency difference. For Bloch-secular approximated dynamics, the spectrum is given by

$$E(\omega', t) \propto \left| \sum_{ij} W_{ij}^{\rm em} P_j(t) \delta(\omega'_{ij} + \omega') \right| \tag{19}$$

and the cumulative spectrum is,

$$E(\omega') = \int_0^\infty dt\ E(\omega', t). \tag{20}$$

In practice the δ function is replaced by a window function with a width $\Delta\epsilon$ on the order of the average nearest-neighbor energy gap of the system. Here we use the rectangular function $\delta(\omega'_{ij}+\omega') \to \Theta(\omega'_{ij}+\omega')\Theta(-\omega'_{ij}-\omega'-\Delta\omega')$.

The three experimental emission spectra with which we compare were obtained by exciting the rhodopsin sample with a continuous wave laser at various frequencies[35]. Such an excitation condition is mimicked in our simulations by narrow band incoherent excitation coupling to the photon bath. Here only the stimulated absorption and emission terms of the corresponding Redfield rate kernel are included, and the spectral density is masked with a window function $J'^{\rm em}(\omega') = J^{\rm em}(\omega')D(\omega'; \omega'_e, \delta\omega')$ centered at the excitation energy ω'_e and a finite width $\delta\omega'$. We take a Gaussian window function with $\delta\omega' = 140\ {\rm cm}^{-1}$. The secular Redfield master equation with this source term of a given excitation energy is propagated from the ground state ($P(0) = |1\rangle$) to the steady state and Eq. (20) is used to extract the emission spectrum.

In addition, to account for the inhomogeneous broadening effect and better compare to the experiments, we further convolve Eqs. (18) and (20) with a Gaussian of width 2000 ${\rm cm}^{-1}$.

IV. RESULTS

To gain insight into the relative merits of each of the three retinal models, we obtain the PF, optimizing first with respect to subsets of: the energy storage, three emission spectra excited at 472.7, 514.5, and 568.2 nm, and the maximum absorption frequency for the *cis* and *trans* conformers. The experimental values are reported in Table III. For the emission spectra we used the experimental data from Ref.[35].

For all simulations presented in the paper, we quantify the accuracy of each model with respect to the emission spectrum via the sum square error for each emission spectra, $f(E_i)$,

$$f(E_i) = \sum_{\nu_i} \Big(g(\nu_i) - \hat{g}(\nu_i) \Big)^2, \tag{21}$$

where $g(\nu)$ is the calculated emission spectra at frequency ν, and $\hat{g}(\cdot)$ are the corresponding experimental values of the emission spectra from Ref.[35]. $f(E_i)$ is dimensionless given that $g(\nu_i)$ is normalized to the peak value. E_i are the excitation frequencies $[472.7, 514.5, 568.2]$ in nm from Ref.[35], that yield the three observed emission spectra. Additionally, we included a penalty to the contributions to the emission spectra for wavelengths larger than 2×10^4 nm; $\hat{g}(\nu_i) = 0$ for $\nu_i > 2\times10^4$ nm. We quantified the accuracy of the models for the energy storage, and the maximum absorption frequencies for the *cis* and *trans* states using their absolute error. For all three models, the MOBOpt algorithm learned the PF as a function of $\Delta\boldsymbol{\theta}_i$, a correction factor for each parameter from Table I; $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^0 + \Delta\boldsymbol{\theta}_i$. For the TM model we did not include the ω parameter in the search procedure, and for all models the mass parameter in the kinetic energy operator was held fixed. In Table III we report the calculated values for all three models when $\Delta\boldsymbol{\theta}_i = 0$ (i.e., the models in the literature). The corresponding emission spectra are displayed in Figure 3. All are clearly in poor agreement with the experimental data, one of the motivations for an effort to improve the models.

For all calculations we set the number of Pareto points to 100, and in the starting of the optimization we randomly sample 50 different $\Delta\boldsymbol{\theta}$. By iteratively sampling new points with Eq. 9 we can update each objective function and reconstruct the PF using the NSGAII algorithm[28]. The maximum number of optimizations for each example was set to 200 to ensure proper exploration of the parameter space, for the majority of our simulations the Pareto Front did not change significantly after the $150th$ iteration . Additionally, one sanity check was to verify that the proposed Y_P are positive, since for this work all error functions are positive.

TABLE III. The calculated values of the energy storage, and *cis* and *trans* states maximum absorption frequencies for all three different models with parameters from Table I, as well as the experimental values for energy storage[36], and the maximum absorption frequencies for *cis* and *trans*[37].

	SM	*TM*	*Omod*	Exp.
Energy Storage [eV]	0.496	1.26	1.3	1.39
max *cis* [nm]	623.61	490.37	497.34	504.85
max *trans* [nm]	654.21	530.30	611.35	538.86

A. Pareto Front for Emission and Energy Storage

In this section we present the PF, considering only the experimental subset of the three emission spectra and the energy storage. Figure 4 displays the Y_P computed with the SM model, Figure 5 for the TM model, and Figure 6 for the *Omod* model. All . consist of three panels (a)-(c) that correspond to two dimensional projections of the full four dimensional space of $\mathbf{F}$ with a third dimension

Accepted to J. Chem. Phys. 10.1063/5.0060259

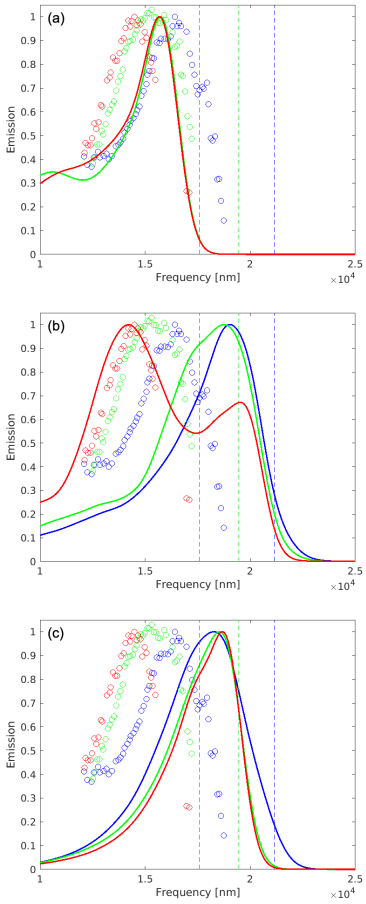


FIG. 3. Emission spectra predicted with the SM model (a), TM model (b), and the $Omod$ model (c) using the parameters reported in Table I. The symbols represent the experimental emission spectra excited at the three frequencies[35], indicated by the vertical dashed lines, and the solid curves are the predicted emission spectra.

represented by a color; $\mathbf{F} = [f(E_1), f(E_2), f(E_3), |ES - 1.39|]$.

The PF for the SM model shows that the emission spectra and the energy storage are not simultaneously minimized. Two values of $\boldsymbol{\theta}$ are identified, $\boldsymbol{\theta}^{\alpha}_{SM} = [1.58, 1.97, 0.55, 0.64]$ and $\boldsymbol{\theta}^{\beta}_{SM} = [2.02, 2.73, 0.56, 0.33]$ corresponding to the lowest mean error for the three emission spectra, and the lowest energy storage error and mean emission spectra, respectively. $\boldsymbol{\theta}^{\alpha}_{SM}$ shows an energy storage of 0.92 eV, and 1.42 eV for $\boldsymbol{\theta}^{\beta}_{SM}$, compared to the experimental value of 1.39 eV. Both are improved compared to the value in Table III. Figure 7 shows the emission spectra and the adiabatic potential energy surfaces for both SM models. Both set of spectra are in poor agreement with experiment.

Figure 5 displays the PF for the TM model. The visual impression of points gathering towards the lower left corner implies that reasonably good agreement with experiment is possible with this model. The selected points in black, corresponding to $\boldsymbol{\theta}^{\alpha}_{TM} = [1.88, 3.87, 0.641, 0.074, 0.627]$ and $\boldsymbol{\theta}^{\beta}_{TM} = [1.833, 4.477, 1.143, -0.0158, 0.351]$ give energy storage of 1.28 and 0.69 eV respectively. Whereas the emission spectra for $\boldsymbol{\theta}^{\beta}_{TM}$ is in good agreement with experiment, as shown in Figure 8 (b), its energy storage is half of the experimental value. By contrast, $\boldsymbol{\theta}^{\alpha}_{TM}$ gives very good agreement with the experimental energy storage, but is not in good agreement with the emission spectra (Figure 8 (a)).

Adiabatic potential energy surfaces for the three TM models are shown in Figure 9. Clearly there are significant differences in the structure of these surfaces and the relative positioning of the conical intersection. Specifically, for the $\boldsymbol{\theta}^{\beta}_{TM}$ and the original model, the energies of the ground electronic states remain mostly constant for $\phi = \frac{\pi}{2}$. However, this feature disappears for the $\boldsymbol{\theta}^{\alpha}_{TM}$ model. Interestingly, $\boldsymbol{\theta}^{\alpha}_{TM}$ and $\boldsymbol{\theta}^{\beta}_{TM}$ both have similar mean error for the three emission spectra, but $\boldsymbol{\theta}^{\beta}_{TM}$ has almost twice the error for the energy storage.

Potential energy surfaces for other models are not shown since we conclude below that the TM model is in best agreement with experimental data.

Figure 6 shows the PF for the $Omod$ model. The panel (b) is particularly interesting because the PF avoided the lower left corner, indicating that simultaneous agreement of these properties to experiments is not possible. For the $Omod$ model, the parameter set that best describes all emission spectra with an accurate energy storage (1.36 eV), is given by $\boldsymbol{\theta}_{Omod} = [0.2709, 0.6802, 0.5314, 1.5213, 0.2066, 0.5498, -0.636, 0.6623]$, as shown in Figure 10. While the energy storage is in excellent agreement, the emission spectra are poor. Lastly, from Figure 6 we can observe that there exist points with more accurate emission spectra but all have considerable energy storage error. The $\boldsymbol{\theta}_{Omod}$ model was the only point that has an acceptable energy storage error.

The difficulty of interpreting the PF data motivated us to simplify matters by computing the mean square error for the emission spectra and the energy storage absolute error with the PF for each model. From Figure 11 we can observe that the TM and $Omod$ models have a smaller error in the predicted emission spectra than the SM model. This approach motivates the study in the next section.

B. Pareto Front for mean Emission, Absorption and Energy Storage

In the previous section we showed that the mean error for the emission spectra is a useful metric to learning a set of viable parameters. Here, we use $\sum_{i=1}^{3} f(E_i)$ as one of the elements of $\mathbf{F}$ to quantify the accuracy of all three emission spectra. Moreover, we expand the number of target observables to include the peak absorption

Accepted to J. Chem. Phys. 10.1063/5.0060259

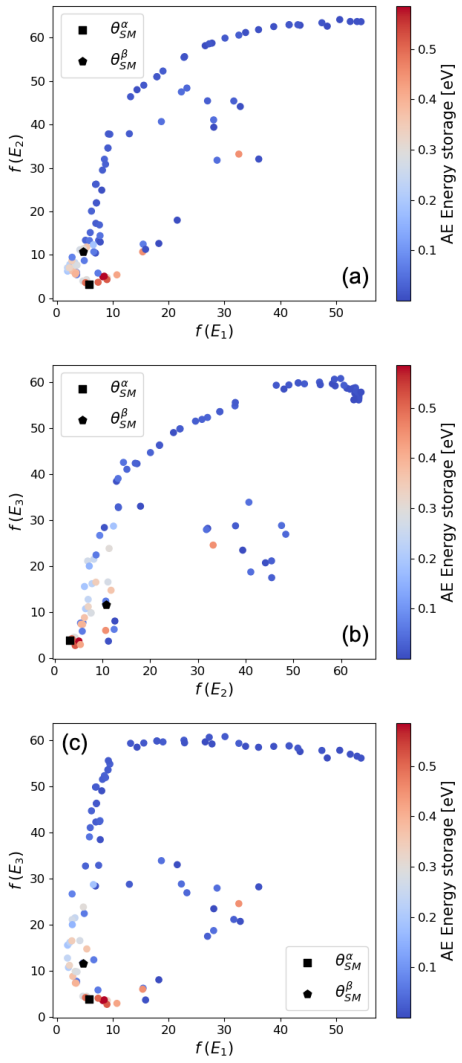


FIG. 4. PF for the SM model computed with respect to the three emission spectra and the energy storage. The energy storage accuracy was quantified with the absolute error. The black symbols indicate the selected two models $\boldsymbol{\theta}^{\alpha}_{SM}$ and $\boldsymbol{\theta}^{\beta}_{SM}$. Each panel displays the PF points for different pairs of functions, i.e., (a) $f(E_1)$ vs $f(E_2)$, (b) $f(E_2)$ vs $f(E_3)$, and (c) $f(E_1)$ vs $f(E_3)$.

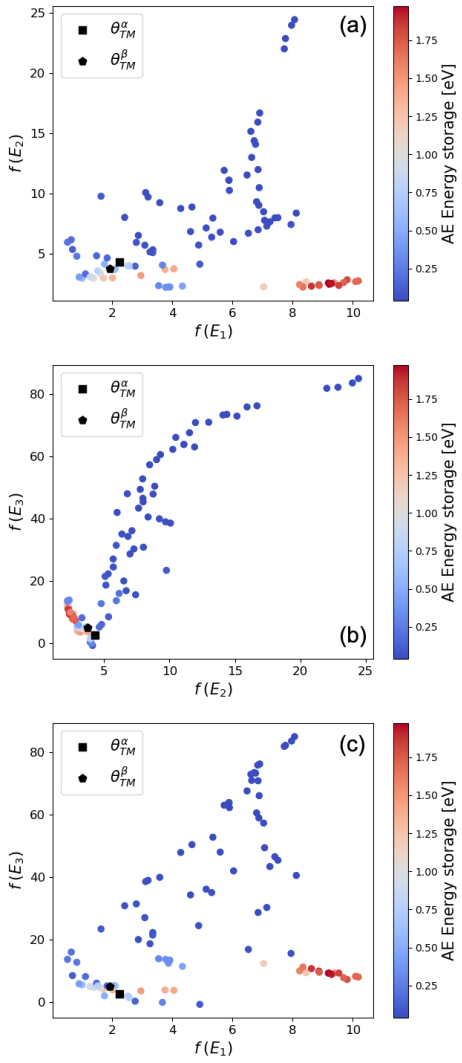


FIG. 5. PF for the TM model computed with respect to the three different emission spectra and the energy storage. The black symbols indicate the selected two models $\boldsymbol{\theta}^{\alpha}_{TM}$ and $\boldsymbol{\theta}^{\beta}_{TM}$. Note that the axes of the three panels are not on the same scale. Each panel displays the PF points for different pairs of functions, i.e., (a) $f(E_1)$ vs $f(E_2)$, (b) $f(E_2)$ vs $f(E_3)$, and (c) $f(E_1)$ vs $f(E_3)$.

frequencies for the *cis* and *trans* conformers so that $\mathbf{F} = \left[\sum_{i=1}^{3} f(E_i), |\nu_{cis} - \hat{\nu}_{cis}|, |\nu_{trans} - \hat{\nu}_{trans}|, |ES - 1.39|\right]$, where $\hat{\nu}$'s are the frequencies reported in Table III, and ES is the calculated energy storage. We follow the same procedure and use the MOBOpt algorithm to look for the PF for these four objectives.

The PF for the SM model was computed but was found to be incapable of jointly describing these observables, not even individually. The lowest error value for the *cis* and *trans* absorption frequencies is relatively large. Given this result, and the poor results reported in the previous section, the SM model is not regarded as a "contender" for a usable description of the system.

The global potential energy surface for retinal chromophore is a 49 dimensional function, which would require a vast amount of data for an accurate fitting. As noted earlier, this has been the main motivation to construct a reduced representation that accurately describes its *cis* $\rightarrow$ *trans* photoisomerization process. The main difference between the SM model and the TM and *Omod* is the extra vibration degree of freedom, typically interpreted as the bond length alternation coordinate in the retinal chromophore. According to our results, it is the key factor of why the SM lacks robustness to describe the emission spectra.

The PF for the TM model, displayed in Figure 12, shows that this model is capable of

Accepted to J. Chem. Phys. 10.1063/5.0060259

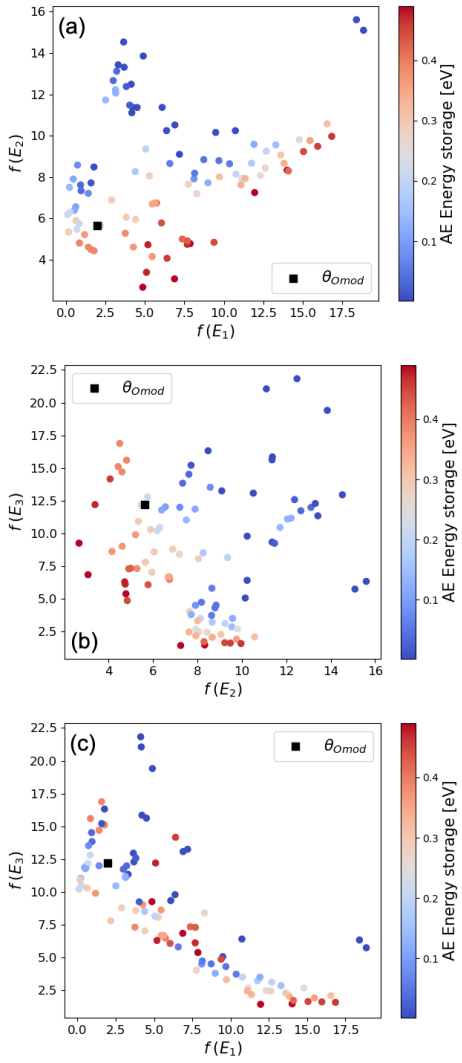


FIG. 6. PF for the *Omod* model computed with respect to the three different emission spectra and the energy storage. The black symbol indicates the selected model $\boldsymbol{\theta}_{Omod}$. Note that the axes of the three panels are not on the same scale. Each panel displays the PF points for different pairs of functions, i.e., (a) $f(E_1)$ vs $f(E_2)$, (b) $f(E_2)$ vs $f(E_3)$, and (c) $f(E_1)$ vs $f(E_3)$.

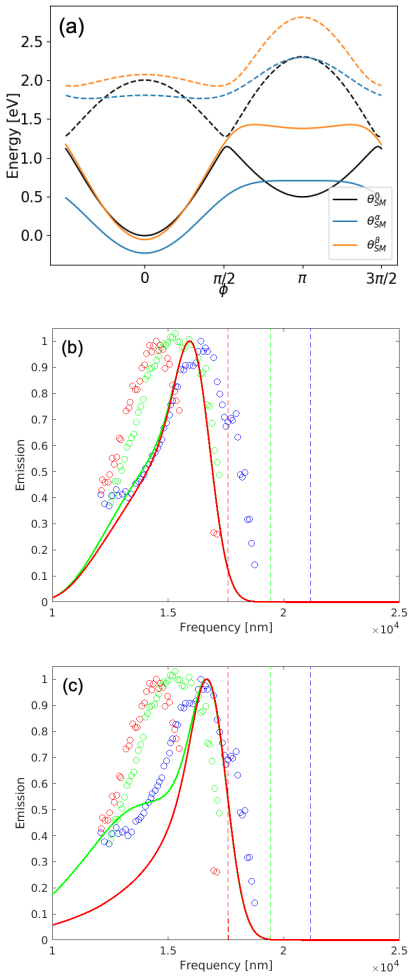


FIG. 7. (a) Adiabatic surfaces for the original SM model $\boldsymbol{\theta}^0_{SM}$ and those of $\boldsymbol{\theta}^\alpha_{SM}$ and $\boldsymbol{\theta}^\beta_{SM}$. Dashed curves represent the excited adiabatic state, and the solid curves the ground adiabatic state. (b) Predicted emission spectra of $\boldsymbol{\theta}^\alpha_{SM}$ model. (c) Predicted emission spectra of $\boldsymbol{\theta}^\beta_{SM}$ model. In panels (b) and (c) the symbols represent the experimental emission spectra excited at the three frequencies[35], indicated by the vertical dashed lines, and the solid curves are the predicted emission spectra.

better describing this set of physical observables. The best values were obtained with $\boldsymbol{\theta}^\gamma_{TM} = [2.385, 3.382, 0.8418, 0.1987, -0.00519]$. The predicted energy storage is 1.31 eV, and 516.6 and 515.7 nm are the max absorption frequencies for the *cis* and *trans* conformers, respectively. This compares to an experimental energy storage of 1.39 eV and max absorption at 504 nm and 539 nm, respectively. Figure 13 (b) displays the predicted emission spectra with these parameters, which are in far better agreement with experiment than the original TM parameters.

The biggest and perhaps most consequential difference introduced in the $\boldsymbol{\theta}^\gamma_{TM}$ model is its λ parameter, the nonadiabatic coupling, that is nearly vanishing. This leads to the similarity between the adiabatic surfaces with their diabatic counterparts, as seen in Figure 14. In contrast to a "peaked" conical intersection typically seen in a TM model, the energy gap between the two adiabatic surfaces opens up only slightly along the $|x| > 0$ direction owing to the small λ value. Consequently, the gradient of the adiabats is dominantly aligned in the ϕ direction, especially in regions close to the Franck-Condon states expected to be most emissive. In such a case one expects a relaxation trajectory that retains more memory of its initial starting point, leading to a stronger dependence of emission spectrum on the excitation frequency as observed in the experiment. We note that the situation in the actual rhodopsin is much more complex and might

Accepted to J. Chem. Phys. 10.1063/5.0060259

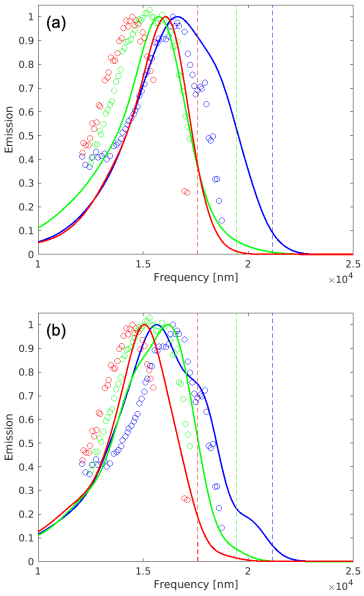


FIG. 8. Emission spectra predicted with TM model with various parameter sets. (a) $\boldsymbol{\theta}^{\alpha}_{TM}$, and (b) $\boldsymbol{\theta}^{\beta}_{TM}$. The symbols represent the experimental emission spectra excited at the three frequencies[35], indicated by the vertical dashed lines, and the solid curves are the predicted emission spectra.

involve more degrees of freedom afforded by an effective two-state two-model model.[34] However, our analysis of the $\boldsymbol{\theta}^{\gamma}_{TM}$ model which produces good agreement with experimental emission spectra sheds light on the relevant energy landscape.

Regarding the *Omod* model, our calculations, displayed in Figure 15, indicate that the TM model better describes all experimental observables. For example, the Pareto front point that best describes all the three emission spectra has an error twice as large as the $\boldsymbol{\theta}^{\gamma}_{TM}$ model, and also has a 0.5 eV error for the energy storage. These results indicate that although the *Omod* model has more free parameters that the TM model, it does not perform better in describing the three emission spectra, maximum absorption frequencies for the *cis* and *trans* states, and the energy storage.

V. DISCUSSION

There are two aspects to the results in this paper, the nature of the multi-objective optimization procedure, and the application to the retinal models. In the first category we note that the optimization of physical models to reproduce a set of target observables is a common task where machine-learning algorithms have proven to be useful. When multiple target observables are considered, the standard procedure is to linearly combine all error functions. This can lead to models with a variety of accuracy for the target observables due to possible

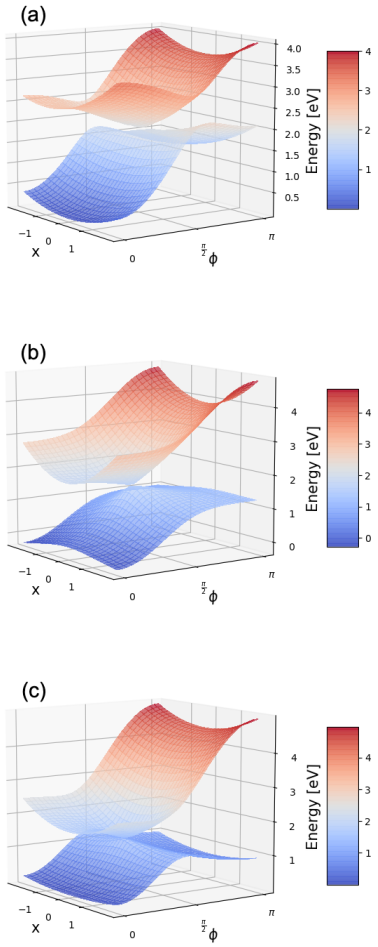


FIG. 9. Adiabatic surfaces for the selected TM models. (a) Parameters from Table I, (b) $\boldsymbol{\theta}^{\alpha}_{TM}$, and (c) $\boldsymbol{\theta}^{\beta}_{TM}$.

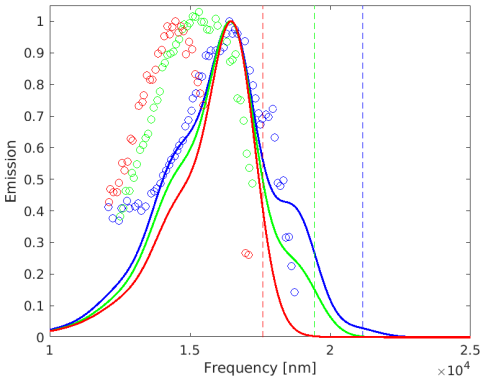


FIG. 10. Emission spectra predicted with the *Omod* model using the $\boldsymbol{\theta}_{Omod}$ parameters given in the text. The symbols represent the experimental emission spectra excited at the three frequencies[35], indicated by the vertical dashed lines, and the solid curves are the predicted emission spectra.

Accepted to J. Chem. Phys. 10.1063/5.0060259

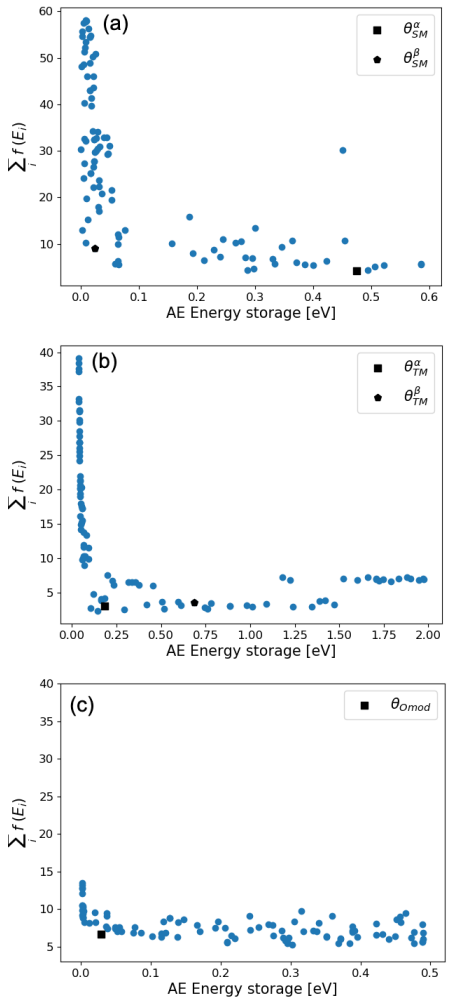


FIG. 11. PF for the average square error for the different emission spectra and the absolute error of the energy storage computed with the (a) *SM* model, (b) *TM* model, and (c) *Omod* model. Note that the axes of the three panels are not on the same scale.

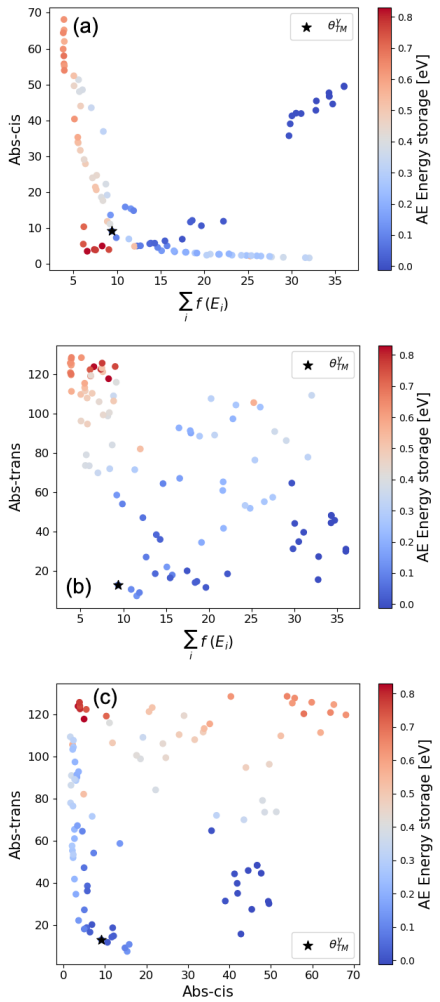


FIG. 12. PF for the *TM* model computed with respect to the sum of three different emission spectra errors, absolute difference for the maximum absorption frequency for the *cis* and *trans* states, and the absolute error for the energy storage. Note that the axes of the three panels are not in the same scale. Each panel displays the PF points for different pairs of functions, i.e., (a) $\sum_{i=1}^{3} f(E_i)$ vs $|\nu_{cis} - \hat{\nu}_{cis}|$, (b) $\sum_{i=1}^{3} f(E_i)$ vs $|\nu_{trans} - \hat{\nu}_{trans}|$, and (c) $|\nu_{cis} - \hat{\nu}_{cis}|$ vs $|\nu_{trans} - \hat{\nu}_{trans}|$.

existence of multiple local minima. This type of optimization scheme will only succeed if there is a single set of parameters jointly optimizing all the target functions, which is a strong assumption. To overcome these obstacles we used a multi-objective optimization protocol where the goal is to learn the limits of each objective through the Pareto front. The results illustrate that we can learn more about the robustness of a physical model by finding the Pareto front than would results from using a single objective optimization scheme. Furthermore, the MOBOpt algorithm used in this work relies on individual Gaussian processes for each target observable reducing number of evaluations that are required.

One of the key features of multi-objective optimization is the possibility of recognizing a joint minimizer shared across different objectives as displayed in the Pareto front in . 4, 5, and 6. This is in contrast to the Pareto front for the *cis* and *trans* conformers' maximum absorption frequencies which displays a *wall* shape[40], e.g., Figure 12, which indicates that the two functions do not share a minimizer in the parameter space. Furthermore, given the flexibility of MOBOpt algorithm and the efficacy of Gaussian processes, the present procedure is independent of the simulation procedure for each observable, and could be used for other problems besides quantum dynamics.

Retinal photoisomerization, as it occurs in nature, i.e. in the steady state induced by incoherent (e.g., solar) radiation, is a particularly important system to examine via this algorithm. In particular, models exist for this system that are parameterized to fit short time femtosecond dynamical data. Optimizing the fits to steady state data provides deep insight into the overall utility of such min-

Accepted to J. Chem. Phys. 10.1063/5.0060259

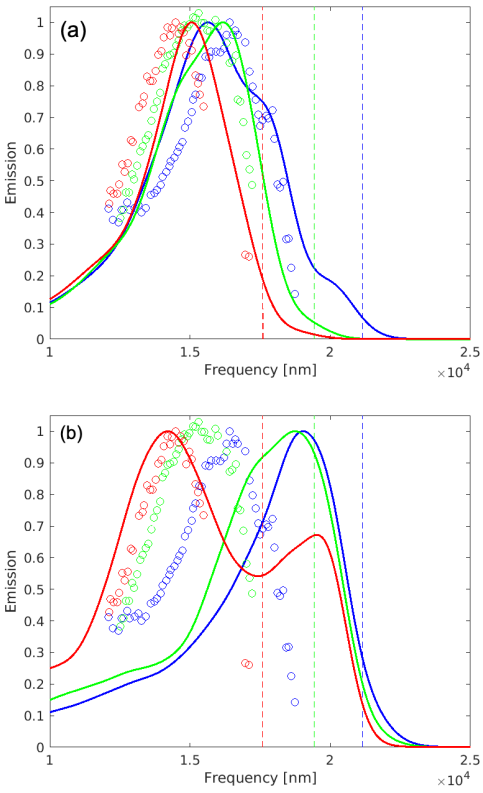


FIG. 13. Emission spectra predicted with TM model with various parameter sets. (a) $\boldsymbol{\theta}^{\gamma}_{TM}$. (b) Parameters from Table I, figure reproduced from Figure 3. The symbols represent the experimental emission spectra excited at the three frequencies[35], indicated by the vertical dashed lines, and the solid curves are the predicted emission spectra.

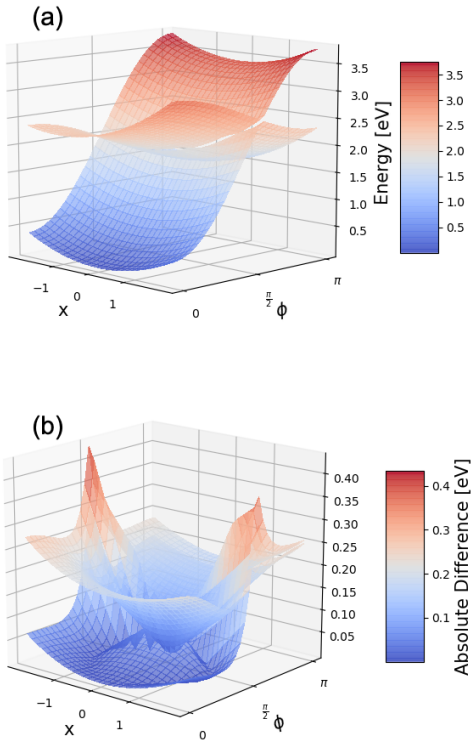


FIG. 14. (a) Adiabatic surfaces for $\boldsymbol{\theta}^{\gamma}_{TM}$.(b) Absolute difference between the adiabatic surfaces produced with $\boldsymbol{\theta}^{\gamma}_{TM}$ and parameters from Table I.

imal models, their advantages and their deficiencies for photoisomerization under natural conditions. To utilize this algorithm, we considered the emission spectra excited at three different wavelengths, combined with the maximum absorption frequencies for the *cis* and *trans* conformers and the energy storage for the retinal chromophore in rhodopsin, all evaluated in the steady state. This set of observables were simulated with three different models: single-mode, two-mode models, and a third one based on Ref.[34], accounting for the two lowest energy electronic states. We expose the character of the Pareto front by considering different combinations of all six observables. For example, we demonstrate that the SM model is incapable of fitting the emission spectra and the maximum absorption frequencies without compromising other observables. However, for the TM and the $Omod$ model the Pareto fronts show that more accurate model exists without jeopardizing the accuracy of the rest of the observables.

We found that the TM model with the $\boldsymbol{\theta}^{\gamma}_{TM}$ parameters $E_1 = 2.385$, $W_0 = 3.382$, $W_1 = 3.382$, $\omega = 0.8418$, $\kappa = 0.1987$, and $\lambda = -0.00519$ is the model that best describe all the different experimental observables considered in this study. Here the λ value is significantly different from the literature value (Table I); $E_1 = 2.58$, $W_0 = 3.56$, $W_1 = 1.19$, $\omega = 0.19$, $\kappa = 0.19$, and $\lambda = 0.19$, Most significantly, the small magnitude of λ, the nonadiabatic coupling, of the $\boldsymbol{\theta}^{\gamma}_{TM}$ model is likely responsible for the good agreement with the experimental emission spectra. This is a trait that might also be applicable to more complicated, higher dimensional models to be investigated in the future. Indeed the lack of quantitative agreement with experimental data in all the models after parameterization motivates such studies, which are ongoing in our laboratory.

ACKNOWLEDGMENTS

This work was supported by U.S. Air Force Office of Scientific Research (AFOSR) grant FA9550-20-1-0354.

DATA AVAILABILITY STATEMENT

The code for the MOBOpt algorithm can be downloaded from `https://github.com/ppgaluzio/MOBOpt`, and the code for the open quantum system dynamics is available upon request.

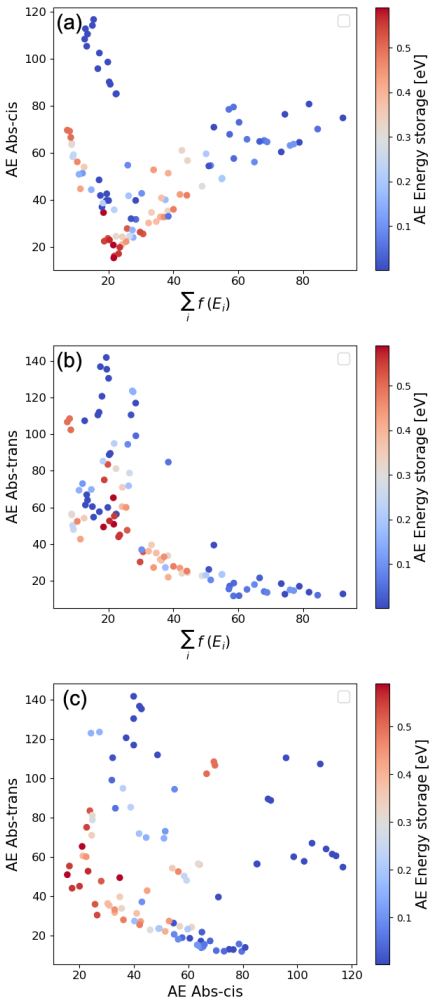


FIG. 15. PF for the *Omod* model computed with respect to the sum of three different emission spectra errors, absolute difference for the maximum absorption frequency for the *cis* and *trans* states, and the absolute error for the energy storage. Note that the axes of the three panels are not in the same scale. Each panel displays the PF points for different pairs of functions, i.e., (a) $\sum_{i=1}^{3} f(E_i)$ vs $|\nu_{cis} - \hat{\nu}_{cis}|$, (b) $\sum_{i=1}^{3} f(E_i)$ vs $|\nu_{trans} - \hat{\nu}_{trans}|$, and (c) $|\nu_{cis} - \hat{\nu}_{cis}|$ vs $|\nu_{trans} - \hat{\nu}_{trans}|$.

[1] W. Banzhaf, P. Nordin, R. Keller, and F. Francone, *Genetic programming : an introduction on the automatic evolution of computer programs and its applications* (San Francisco, Calif. : Morgan Kaufmann Publishers, 1998)

[2] J. Snoek, H. Larochelle, and R. P. Adams, *Adv. Neur. Inf. Process. Sys.* **25**, 2951(2012).

[3] B. Shahriari, K. Swersky, Z. Wang, R .P. Adams, and N. de Freitas, *Proc. IEEE* **104**, 148 (2016).

[4] J. Proppe, S. Gugler, and M. Reiher, *J. Chem. Theory Comput.* **15**, 11 (2019).

[5] R. A. Vargas-Hernández, *J. Phys. Chem. A* **124** (20) , 4053 (2020).

[6] S. Carr, R. Garnett, and C. Lo. *PMLR* **48**, 898 (2016).

[7] L. Chan, G. Hutchison, and G. Morris. *J. Cheminform* **11**, 32 (2019).

[8] R. A. Vargas-Hernández, Y. Guan, D. H. Zhang, and R. V. Krems, *New J. Phys.* **21**, 022001 (2019).

[9] T. Ueno, T. D. Rhone, Z. Hou, T. Mizoguchi, and K. Tsuda. *Mater. Discovery* **4**, 18 (2016).

[10] R. Jalem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki, and T. Saito, *Sci. Rep.* **8**, 5845 (2018).

[11] S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda, and J. Shiomi, *Phys. Rev. X* **7**, 021024 (2017).

[12] R. Tamura, and K. Hukushima, *PLoS ONE* **13**, e0193785 (2018).

[13] R. V. Krems, *Phys. Chem.Chem. Phys.* **21**, 13392 (2019).

[14] Z. Deng, I. Tutunnikov, I. Sh. Averbukh, M. Thachuk, and R. V. Krems, *J. Chem. Phys.* **153**, 164111 (2020).

[15] A. M. Gopakumar, P. V. Balachandran, D. Xue, J. E. Gubernatis, and T. Lookman, *Sci. Rep.* **8**, 3738 (2018).

[16] A. Solomou, G. Zhao, S. Boluki, J. K. Joy, X. Qian, I. Karaman, R. Arróyave, and D. C. Lagoudas, *Materials & Design* **160**, 810 (2018).

[17] Y. Chen, Y. Tian, Y. Zhou, D. Fang, X. Ding, J. Sun, and D. Xue, *J. Alloys and Compounds* **844**, 156159 (2020).

[18] F. Häse, L. M. Roch, and A. Aspuru-Guzik, *Chem. Sci.* **9**, 7642 (2018).

[19] K. F. Koledina, S. N. Koledin, A. P. Karpenko, I. M. Gubaydullin, and M. K. Vovdenko, *J. Math. Chem.* **57**, 484 (2019).

[20] I. C. Stevenson, and J. Pérez-Ríos, *Journal of Physics B: Atomic, Molecular and Optical Physics* **52**, 105002 (2019).

[21] T. V. Tscherbul, and P. Brumer, *J. Phys. Chem. A* **118**, 3100 (2014).

[22] A. Dodin, T. V. Tscherbul, and P. Brumer, *J. Chem. Phys.* **145**, 244313 (2016).

[23] P. Brumer, *J. Phys. Chem. Lett.* **9**, 2946 (2018).

[24] M. Emmerich , K. Yang , A. Deutz, H. Wang , C.M.Fonseca, A Multicriteria Generalization of Bayesian Global Optimization. In: Pardalos P., Zhigljavsky A., Žilinskas J. (eds) Advances in Stochastic and Deterministic Global Optimization. Springer Optimization and Its Applications, vol 107. Springer, Cham. (2016)

[25] P. P. Galuzio, E. Hochsteiner de Vasconcelos Segundo, L. dos Santos Coelho, and V. Cocco Mariani, *SoftwareX* **12**, 100520 (2020).

[26] K. Miettinen, *Nonlinear Multiobjective Optimization* (Springer US, 1998).

[27] C. E. Rasmussen, and C. K. I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, 2006).

[28] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, *IEEE transactions on evolutionary computation*, **6**, no. 2 (2002).

[29] P. P. Galuzio, Multi-Objective Bayesian Optimization too for Python. 2020. https://github.com/ppgaluzio/MOBOpt/wiki

[30] L. Seidner, and W. Domcke, *Chem. Phys.* **186**, 27 (1994).

[31] L. Seidner, G. Stock, and W. Domcke, *J. Chem. Phys.* **103**, 3998 (1995).

[32] S. Hahn, and G. Stock, *J. Phys. Chem. B* **104**, 1146 (2000).

[33] S. Hahn, and G. Stock, *Chem.Phys.* **259**, 297 (2000).

[34] E. Marsili, M. H. Farag, X. Yang, L. De Vico, and M. Olivucci, *J. Phys. Chem. A* **123**, 1710 (2019).

[35] G. G. Kochendoerfer, and R. A. Mathies, *J. Phys. Chem.* **100**, 14526 (1996).

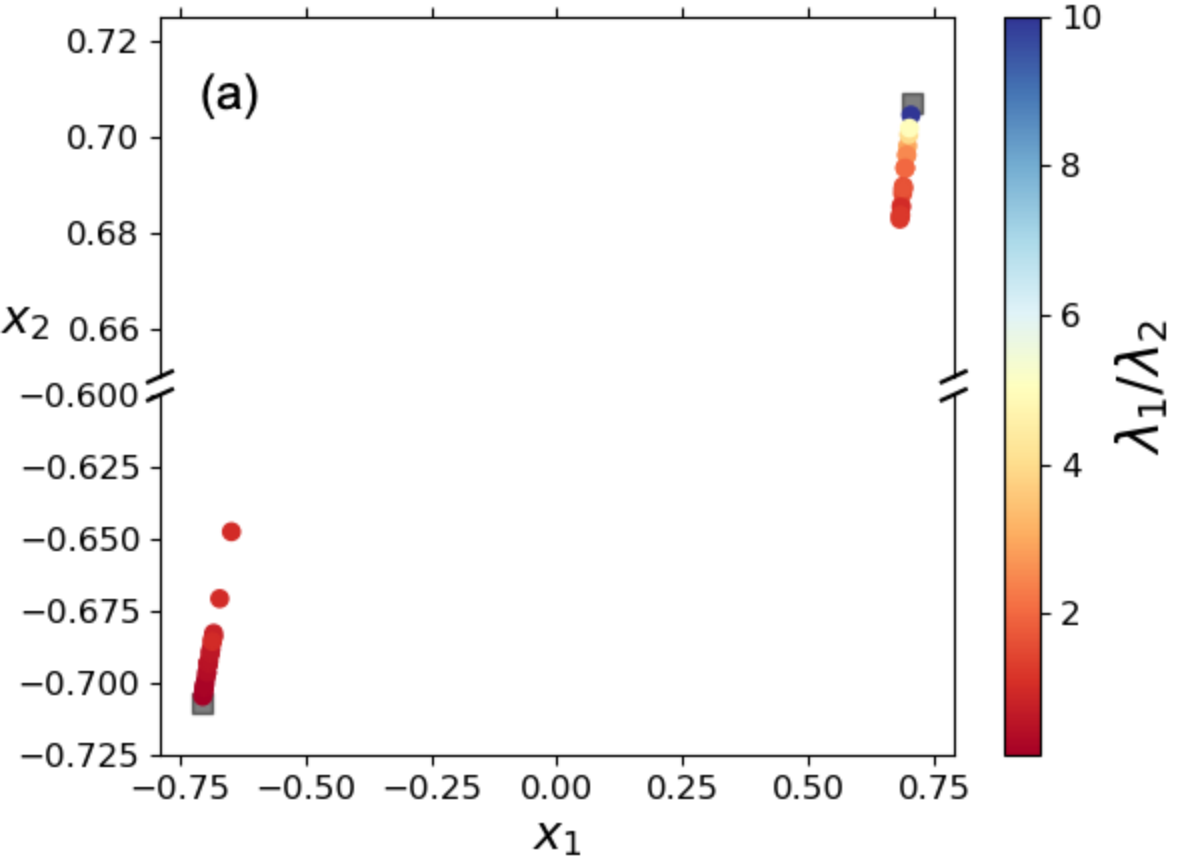
[36] G. A. Schick, T. M. Cooper, R. A. Holloway, L. P. Murray, and R. R. Birge, *Biochemistry* **26**, 2556 (1987).

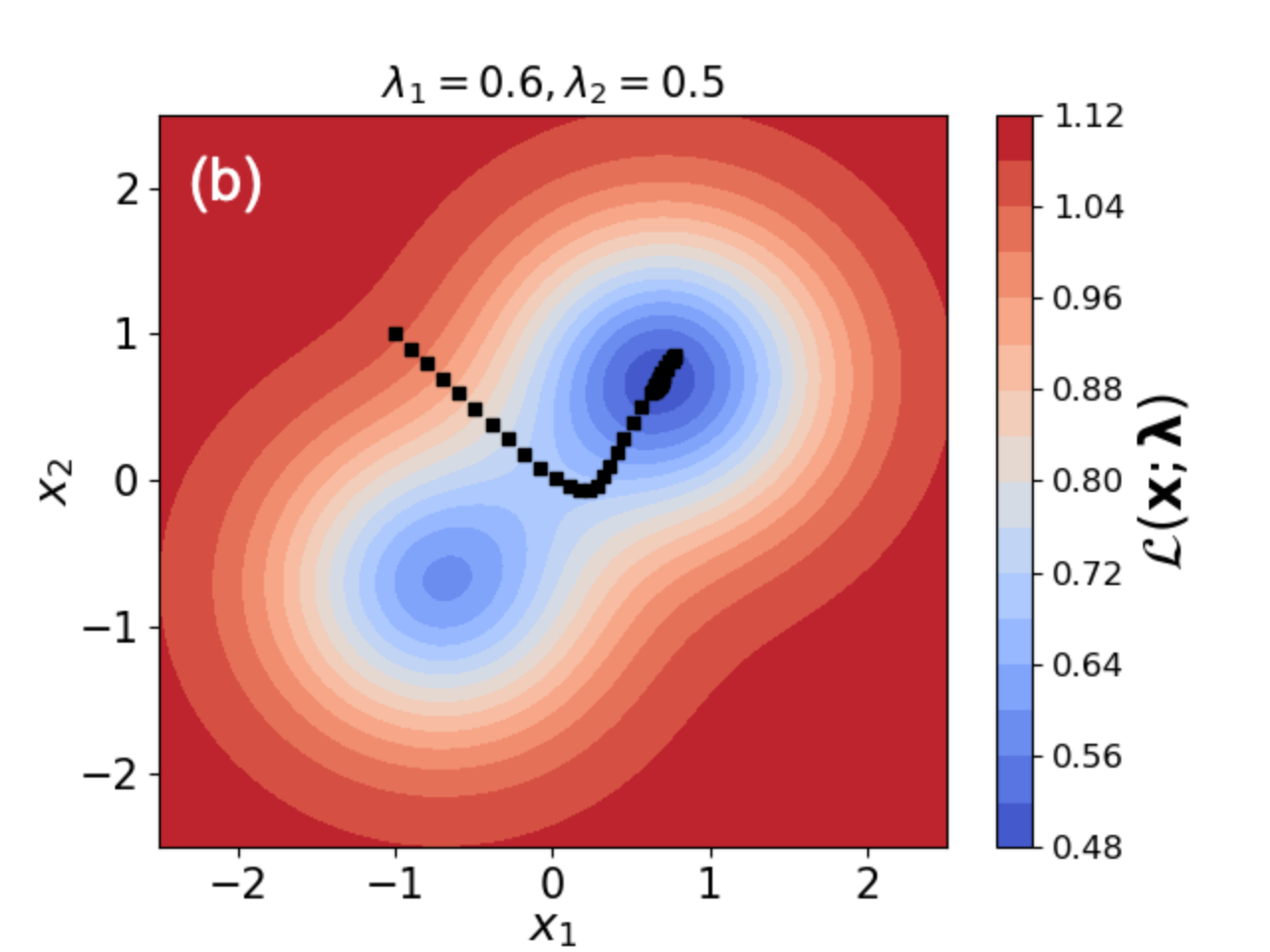
[37] R. R. Birge, C. M. Einterz, H. M. Knapp, and L. P. Murray, *Biophys. J.* **53**, 367 (1988).

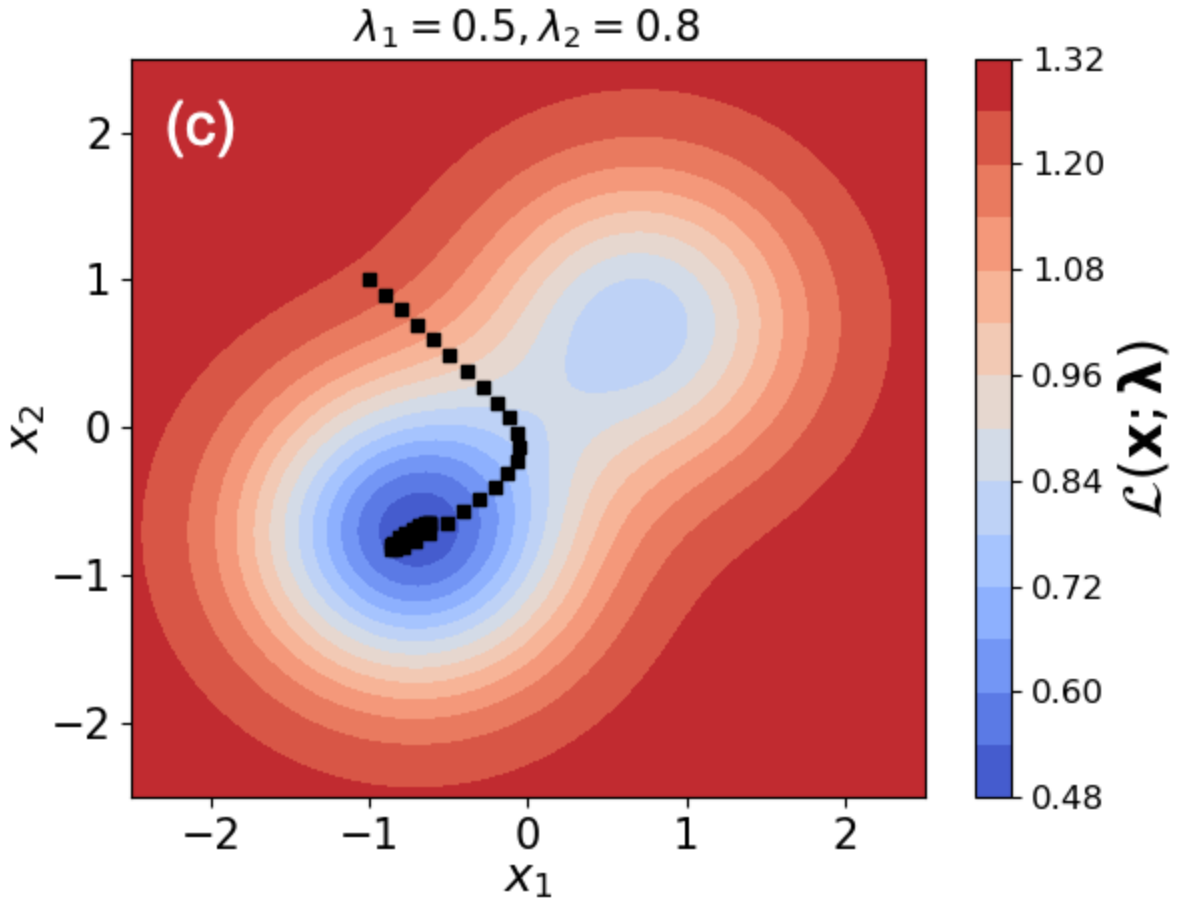
[38] B. Balzer, and G. Stock, *Chem. Phys.* **310**, 33 (2005).

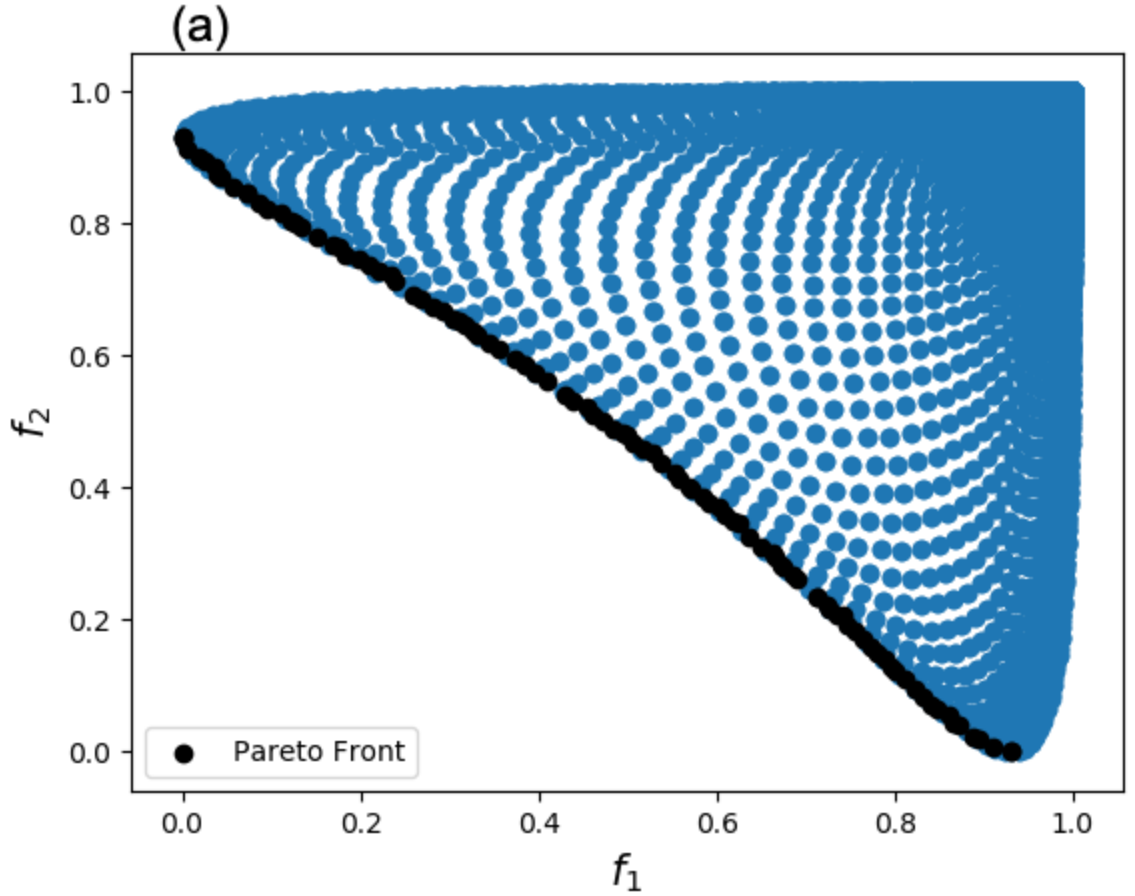
[39] C. M. Fonseca, and P. J. Fleming, *Evolutionary Computation* **3**, 1 (1995).

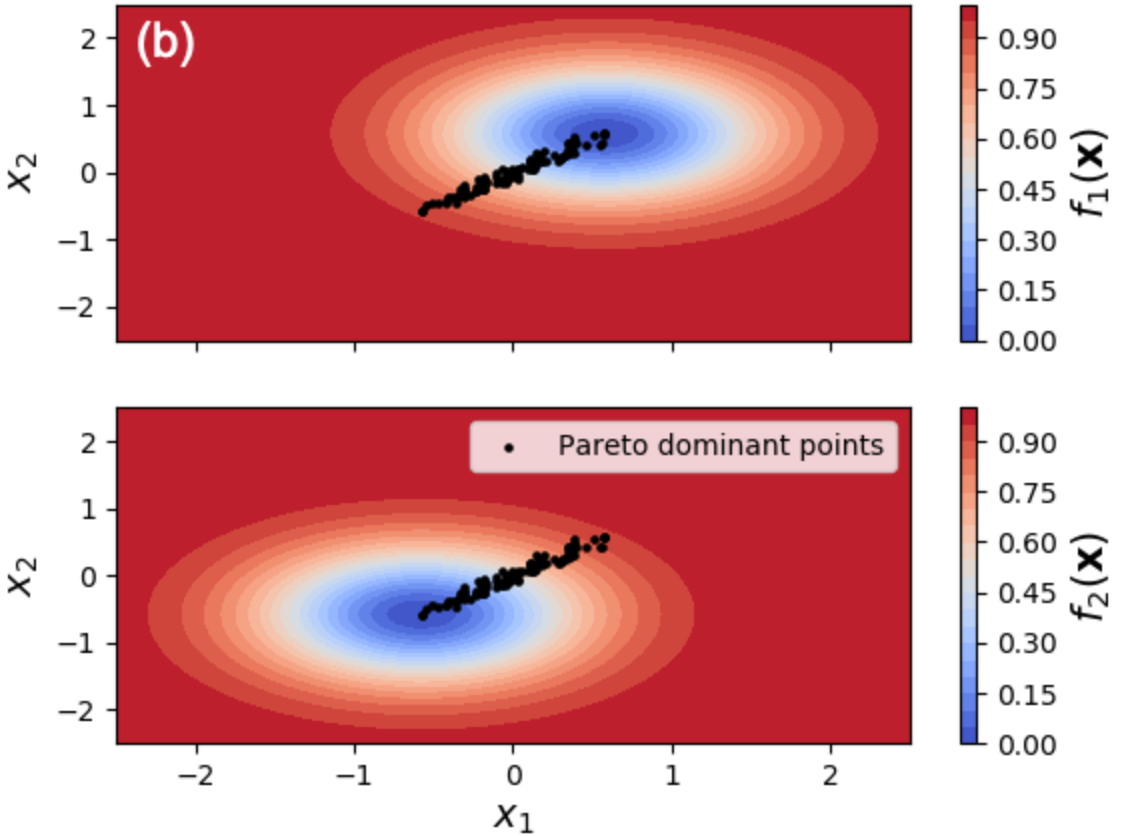
[40] When there exists a single Pareto dominant point that jointly minimizes at least two functions, the Pareto Front has a displays a sharp peak or an "L" function, for example Figure 11. On the other hand, when there is not a single point we get a set of points that lie between both minima, for example Figure 2 b. This is what we meant as "*wall*" shape.

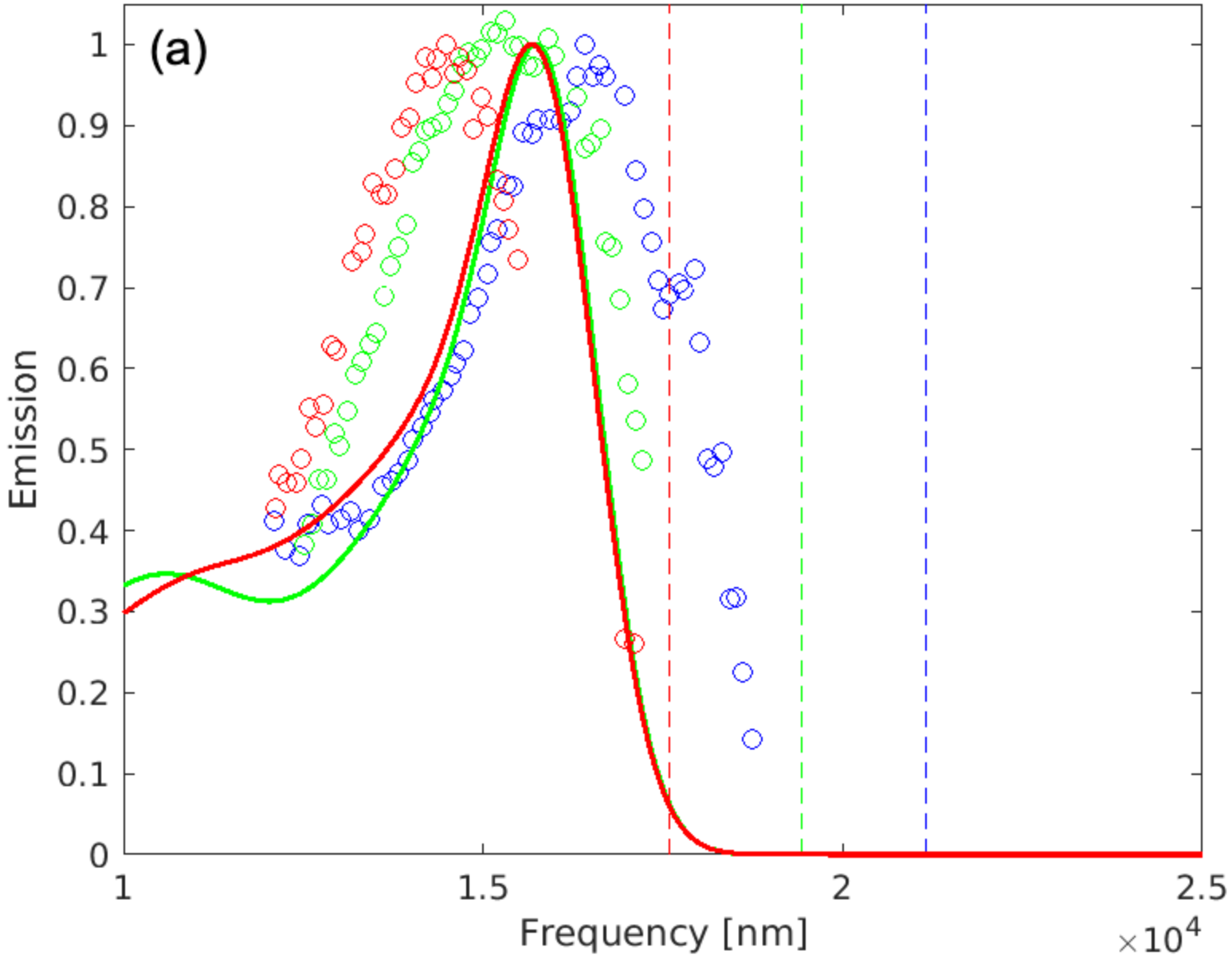

(a)
0.72
0.70
0.68
0.66
x_2
−0.600
−0.625
−0.650
−0.675
−0.700
−0.725
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
x_1
10
8
6
4
2
λ_1/λ_2

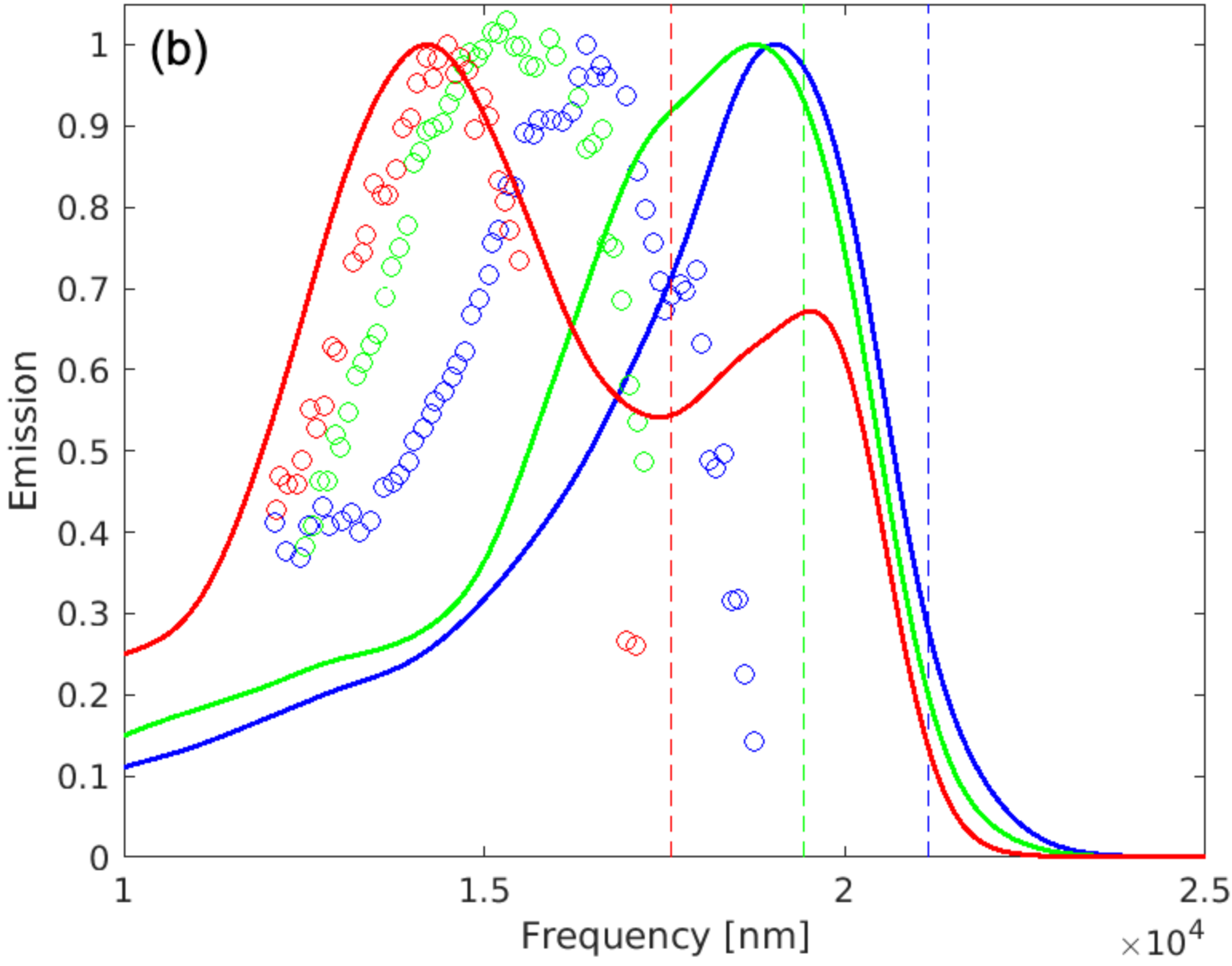

$\lambda_1 = 0.6, \lambda_2 = 0.5$
(b)
x_1
x_2
$\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda})$
1.12
1.04
0.96
0.88
0.80
0.72
0.64
0.56
0.48

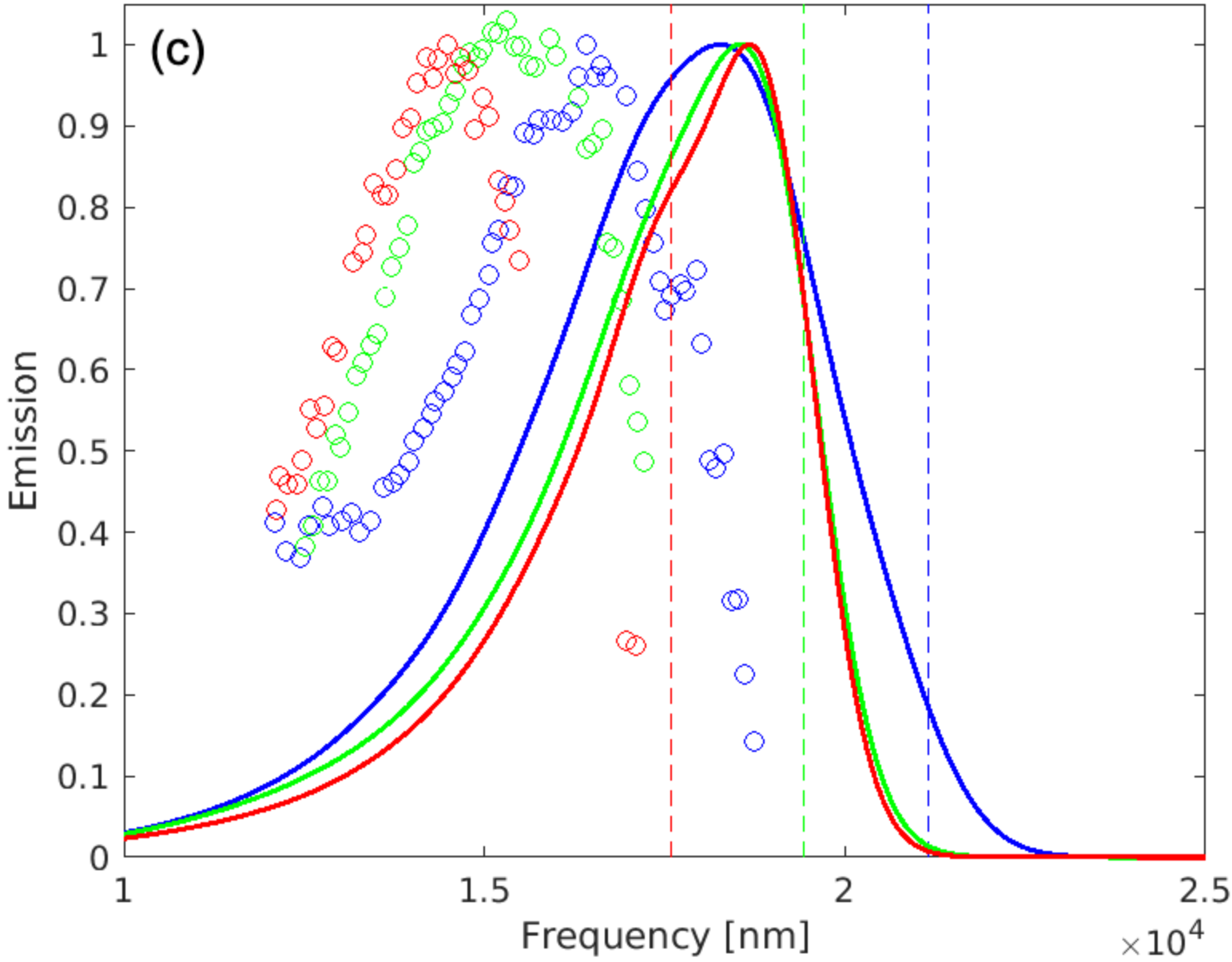

$\lambda_1 = 0.5, \lambda_2 = 0.8$
(c)
2
1
0
−1
−2
x_2
−2
−1
0
1
2
x_1
1.32
1.20
1.08
0.96
0.84
0.72
0.60
0.48
$\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda})$

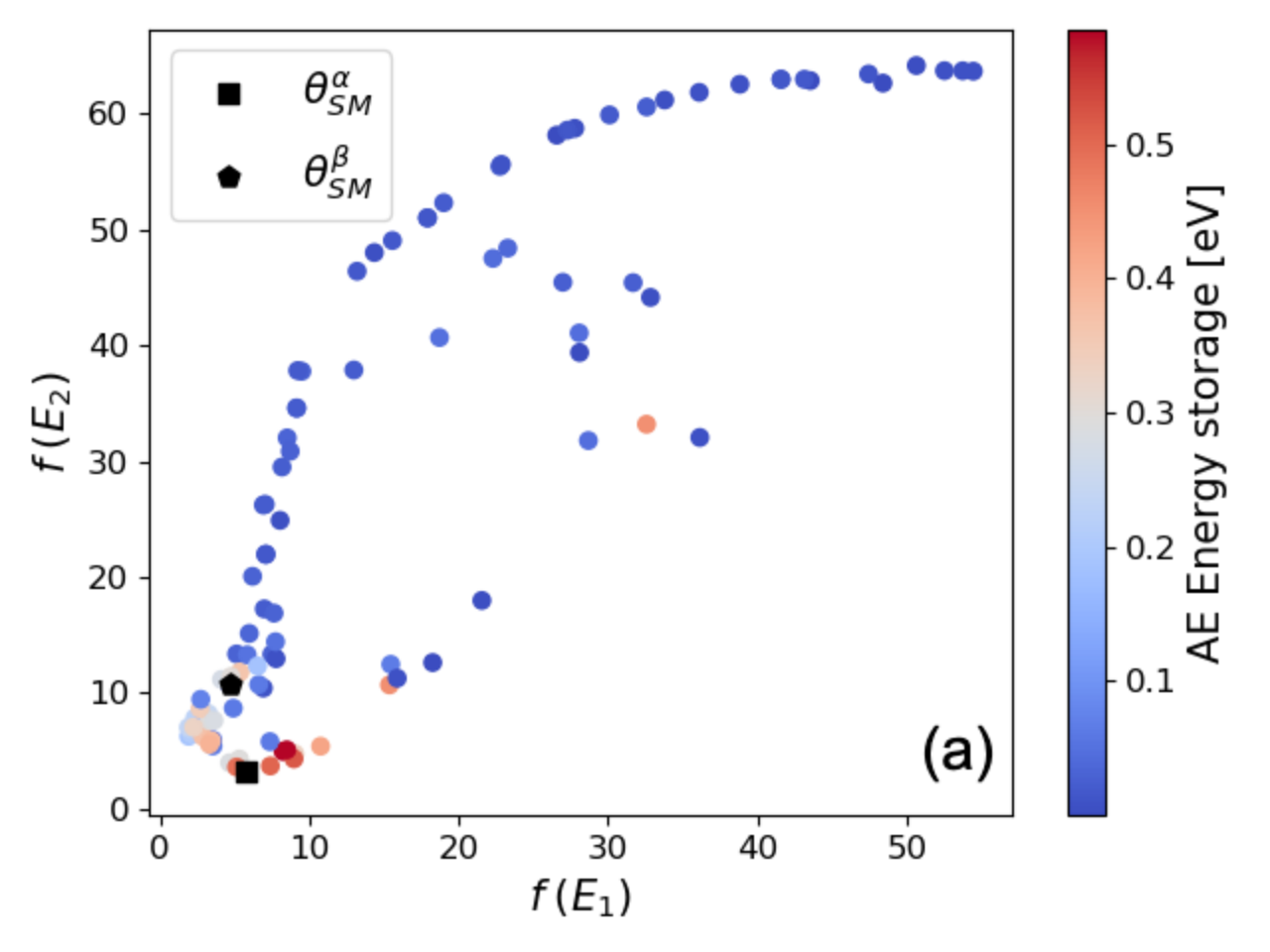

(a)
1.0
0.8
0.6
0.4
0.2
0.0
f_2
Pareto Front
0.0
0.2
0.4
0.6
0.8
1.0
f_1

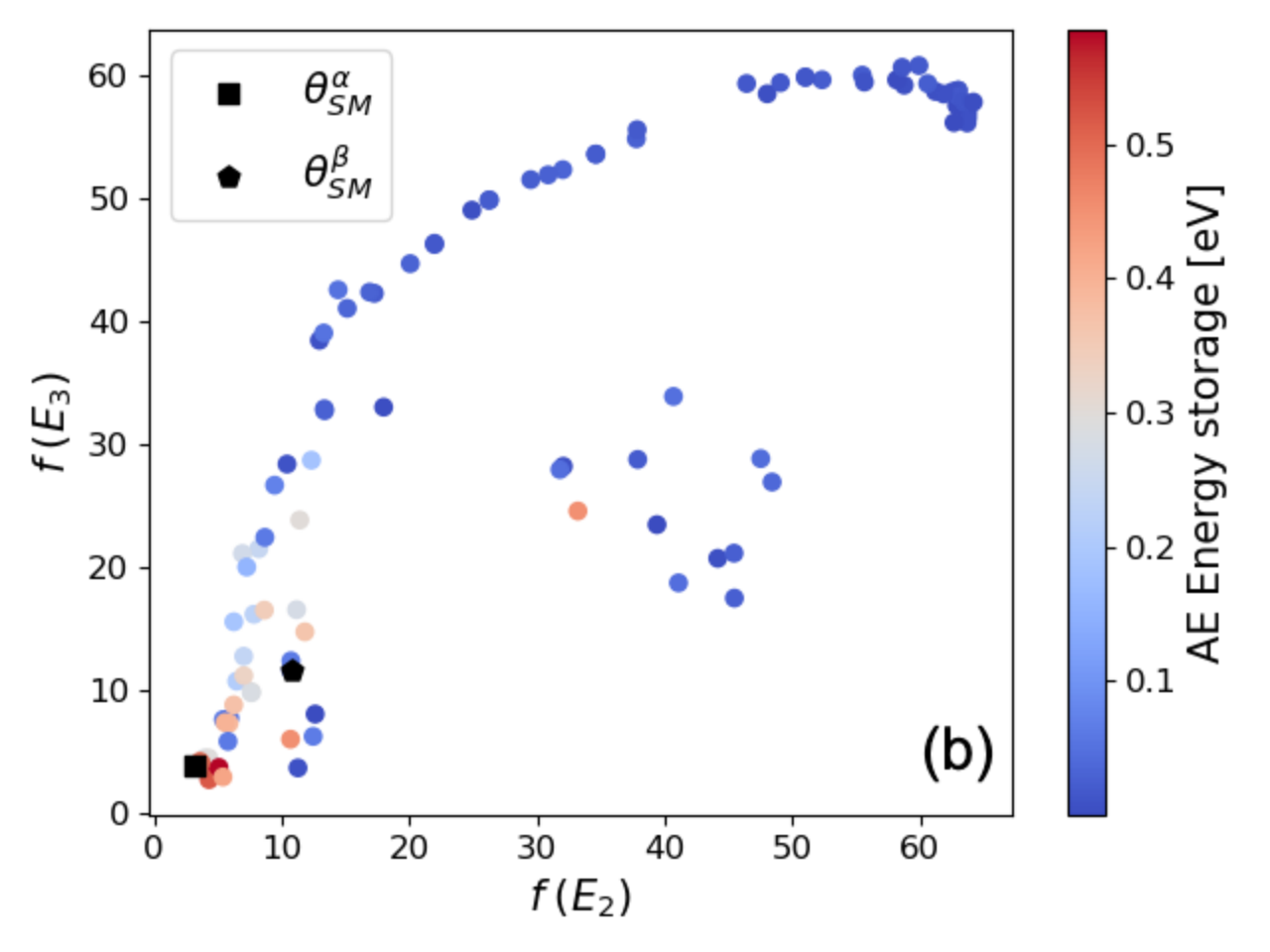

(b)
Pareto dominant points
x_1
x_2
$f_1(\mathbf{x})$
$f_2(\mathbf{x})$
0.00
0.15
0.30
0.45
0.60
0.75
0.90

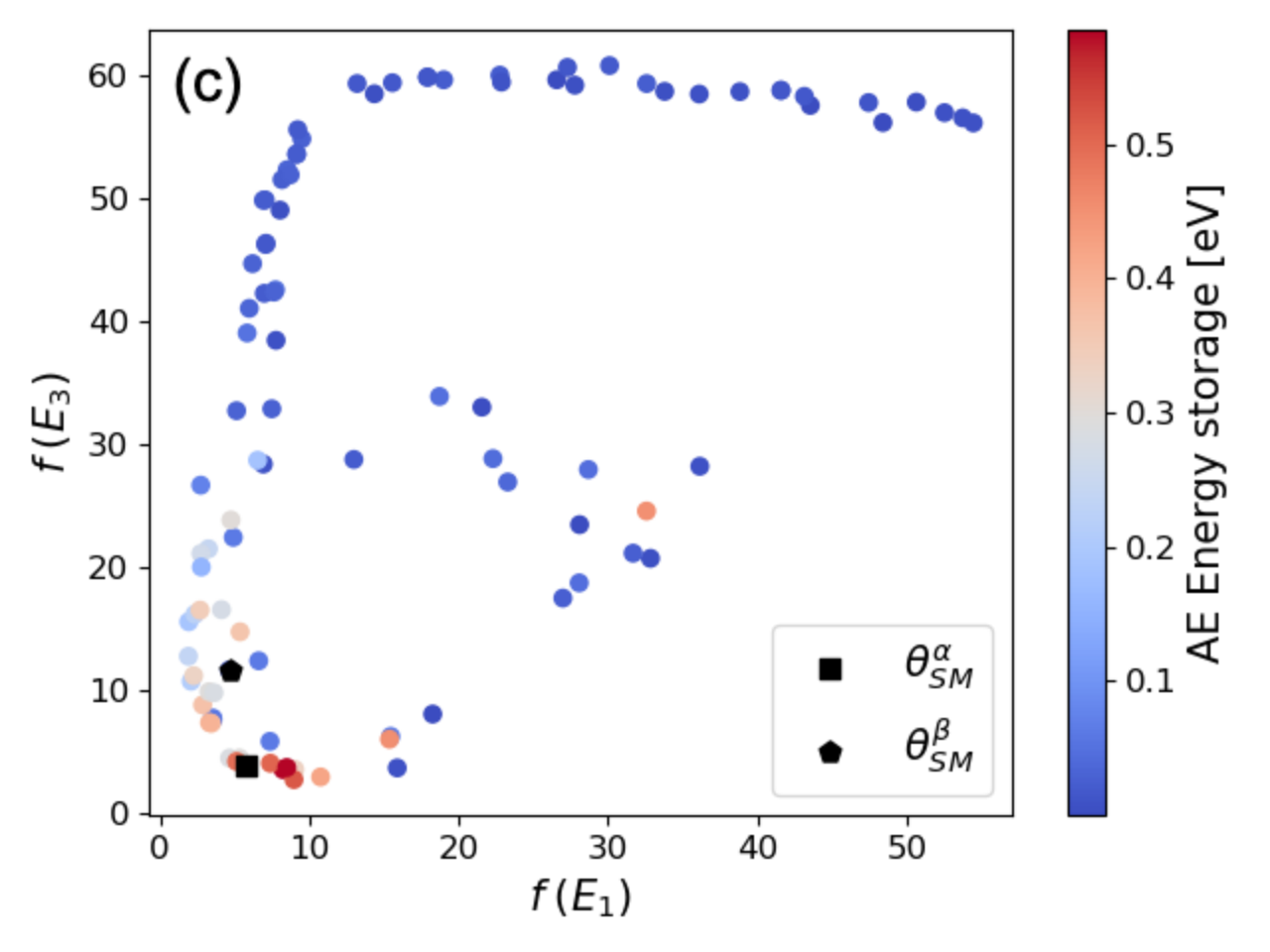

(a)
Emission
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5
Frequency [nm]
×10^4

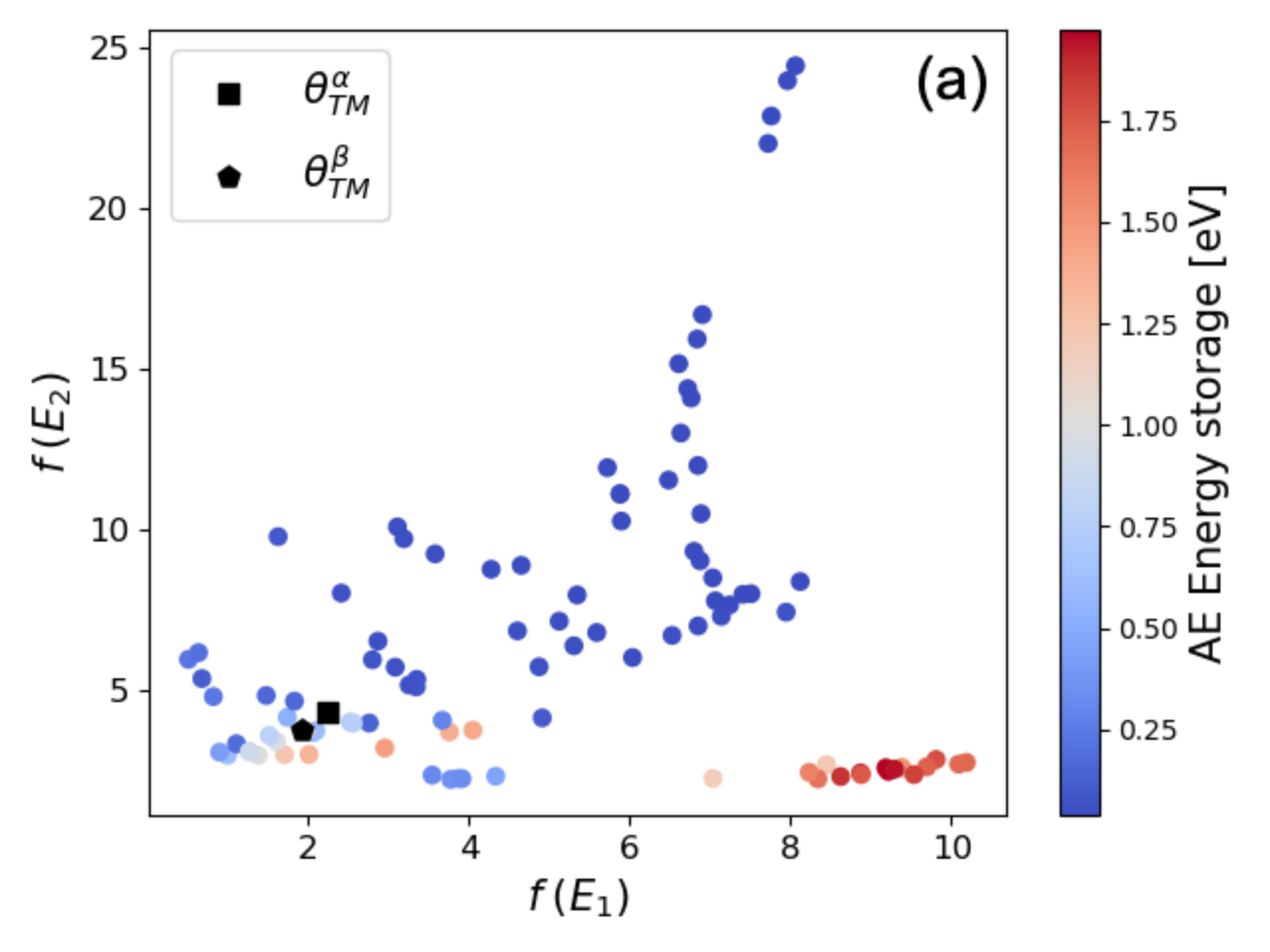

(b)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

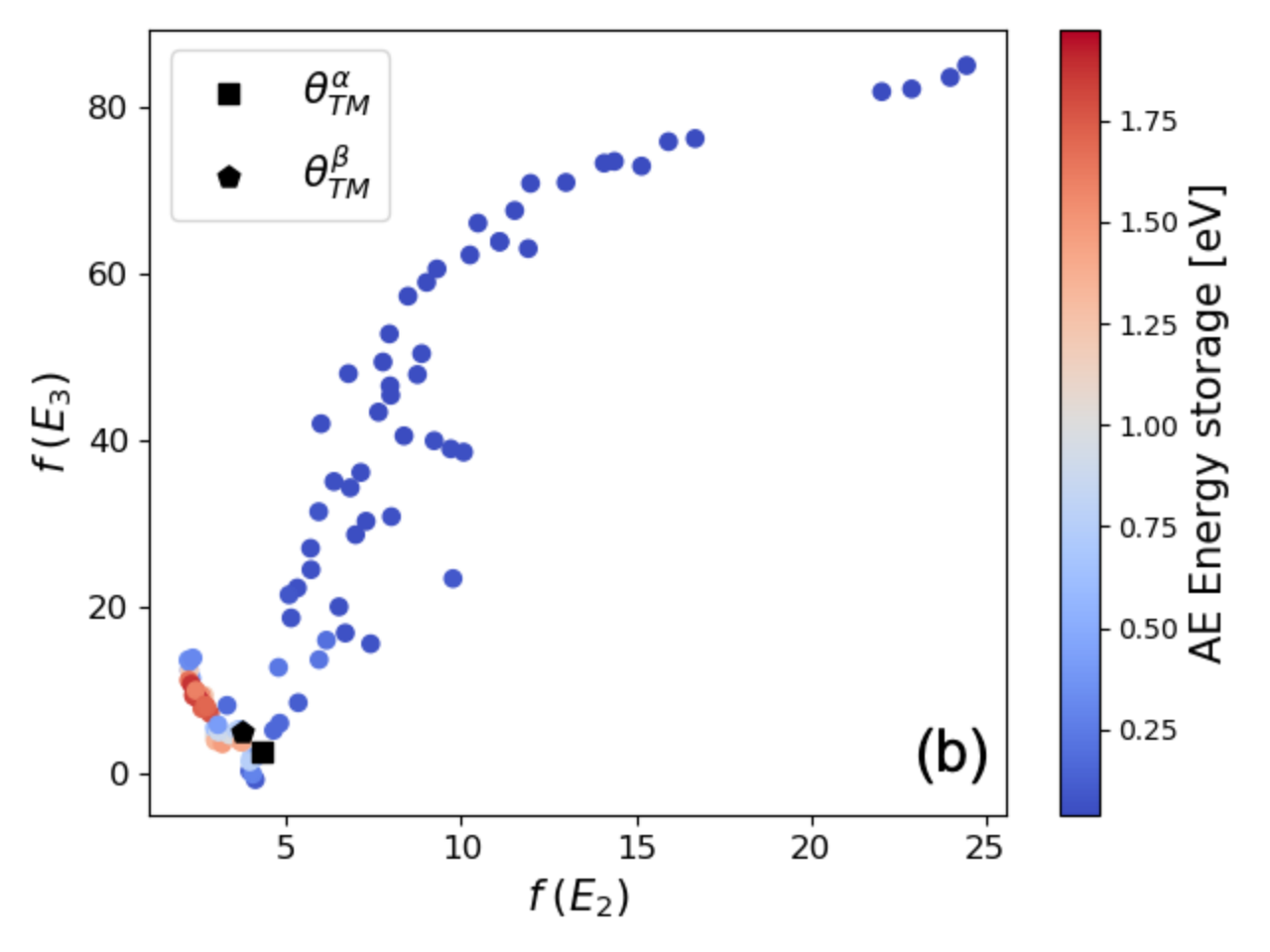

(c)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

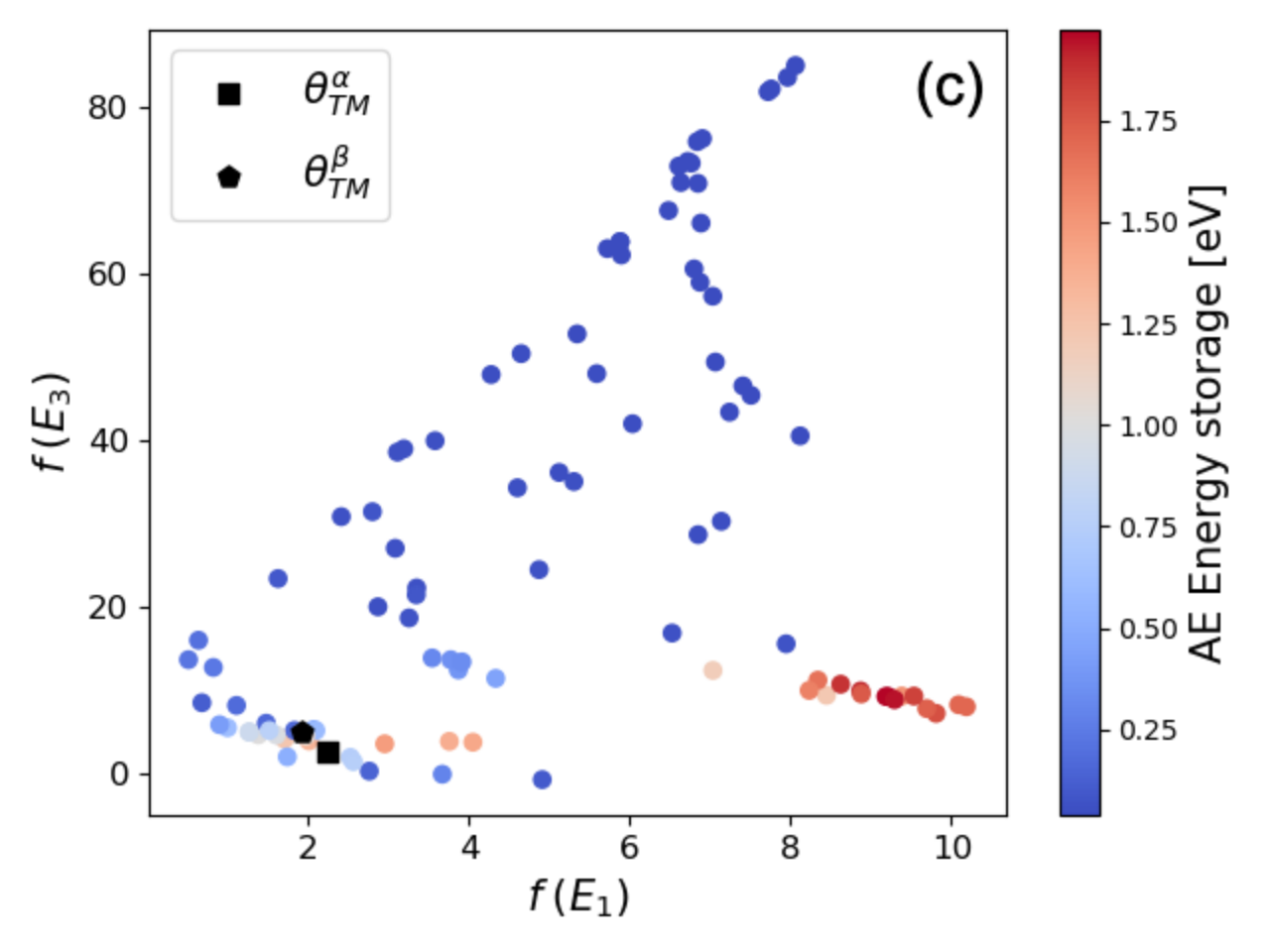

θ^{α}_{SM}
θ^{β}_{SM}
(a)
$f(E_1)$
$f(E_2)$
AE Energy storage [eV]
0
10
20
30
40
50
60
0.1
0.2
0.3
0.4
0.5

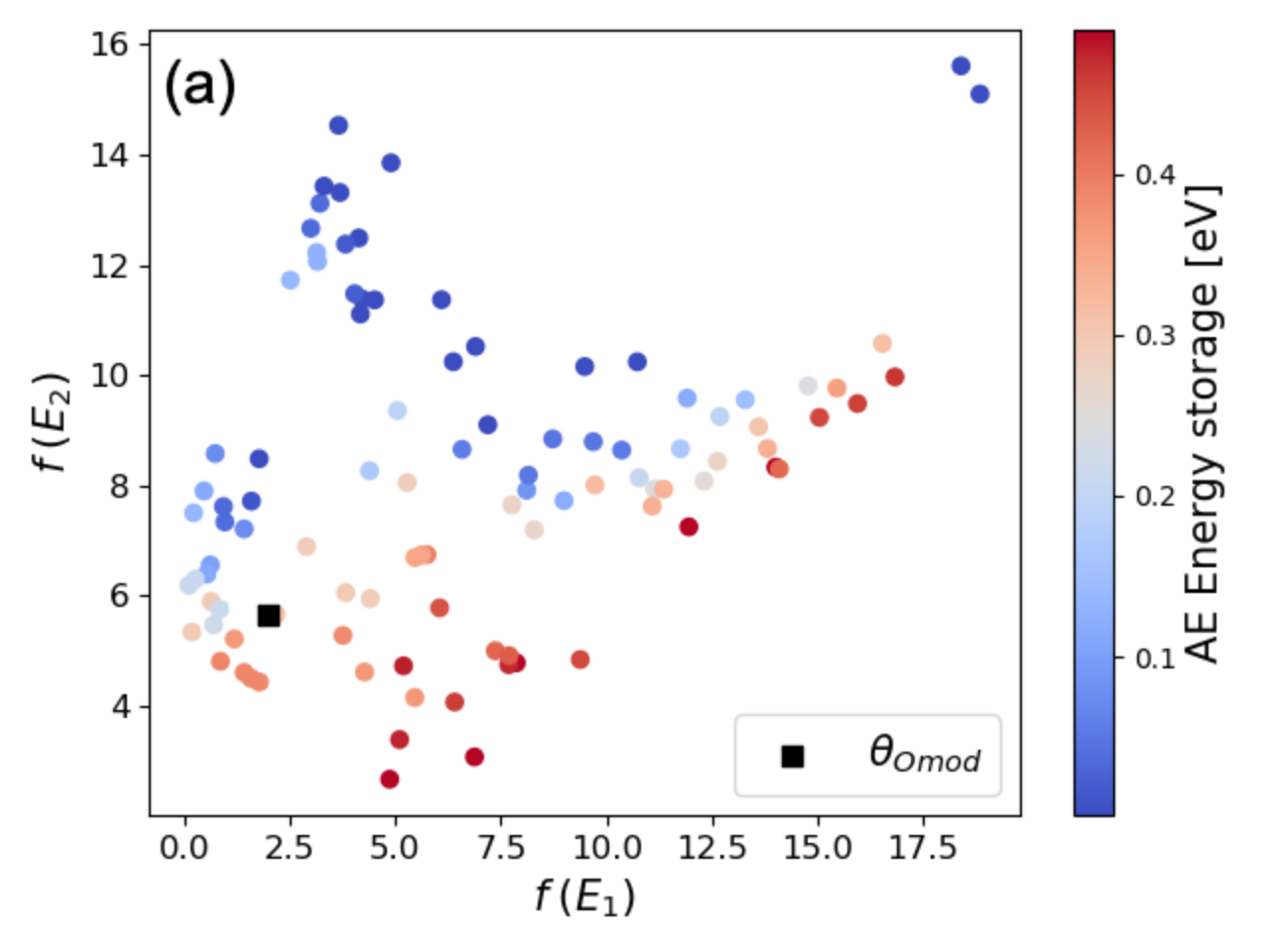

θ^{α}_{SM}
θ^{β}_{SM}
(b)
$f(E_3)$
$f(E_2)$
AE Energy storage [eV]
0
10
20
30
40
50
60
0.1
0.2
0.3
0.4
0.5

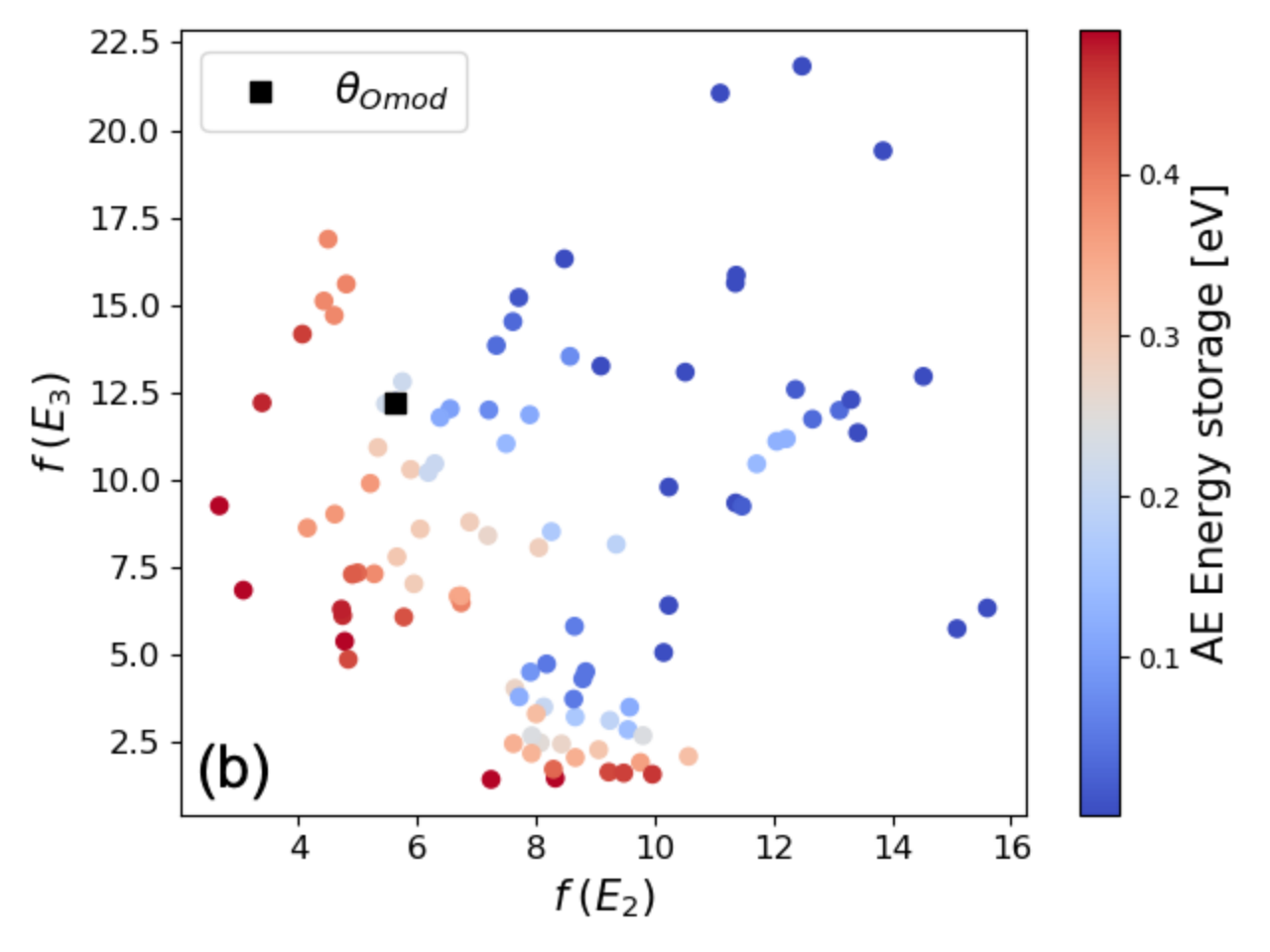

(c)
$f(E_3)$
$f(E_1)$
θ^{α}_{SM}
θ^{β}_{SM}
AE Energy storage [eV]
0
10
20
30
40
50
60
0.1
0.2
0.3
0.4
0.5

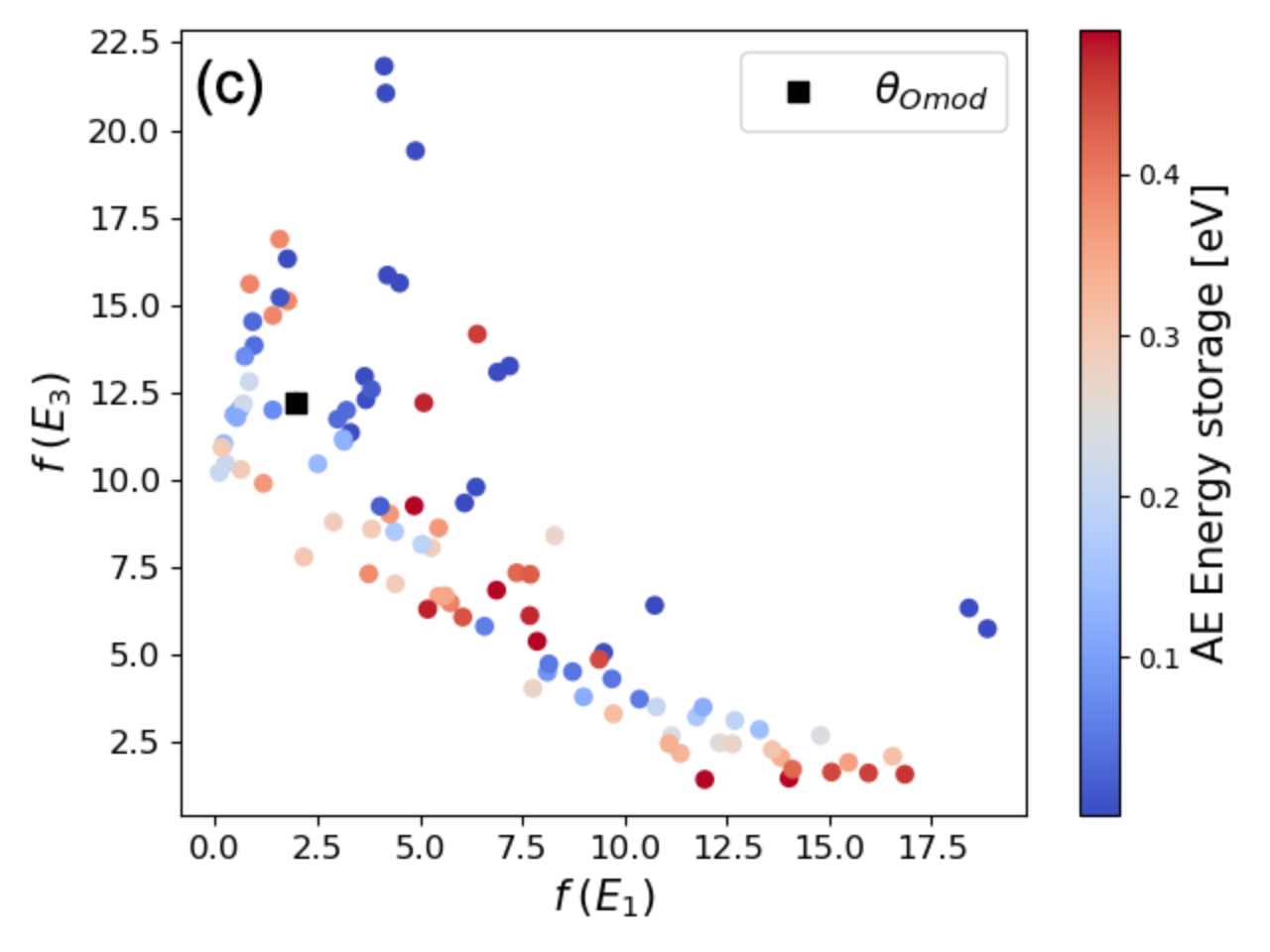

(a)
θ_{TM}^{α}
θ_{TM}^{β}
$f(E_1)$
$f(E_2)$
AE Energy storage [eV]
25
20
15
10
5
2
4
6
8
10
1.75
1.50
1.25
1.00
0.75
0.50
0.25

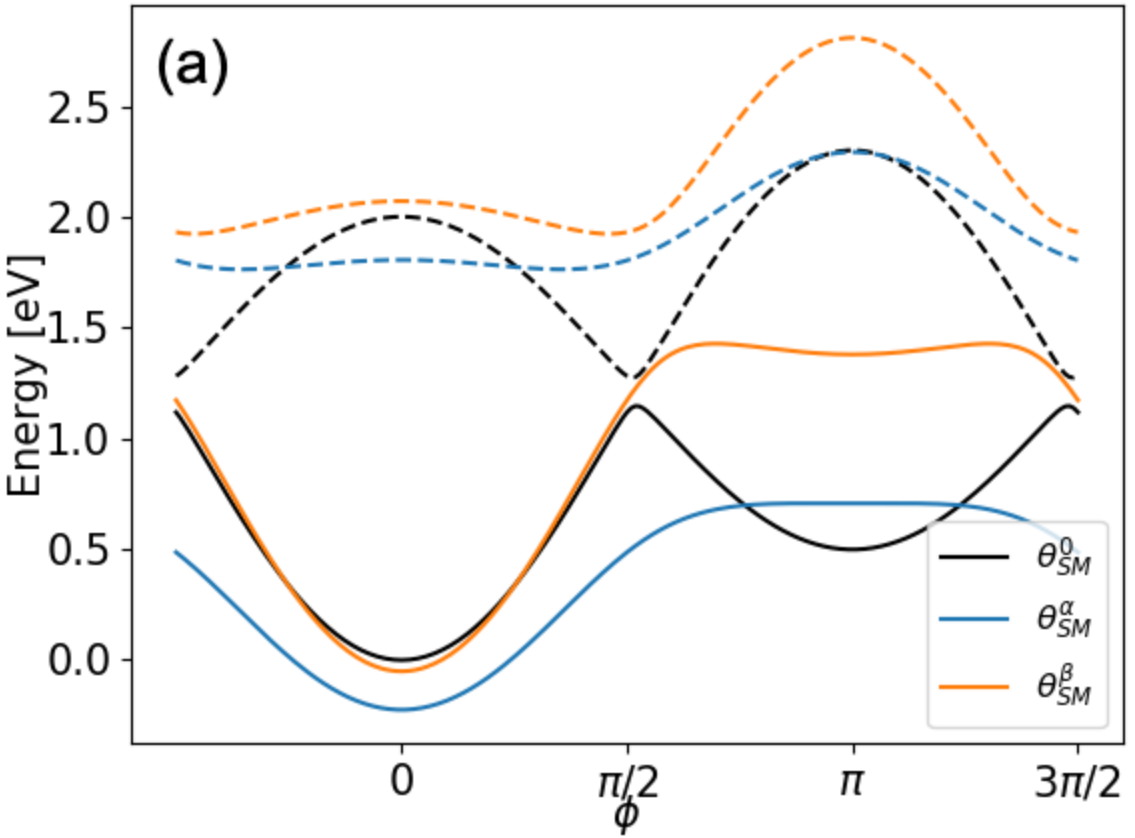

θ^{α}_{TM}
θ^{β}_{TM}
(b)
$f(E_3)$
$f(E_2)$
AE Energy storage [eV]
0
20
40
60
80
5
10
15
20
25
0.25
0.50
0.75
1.00
1.25
1.50
1.75

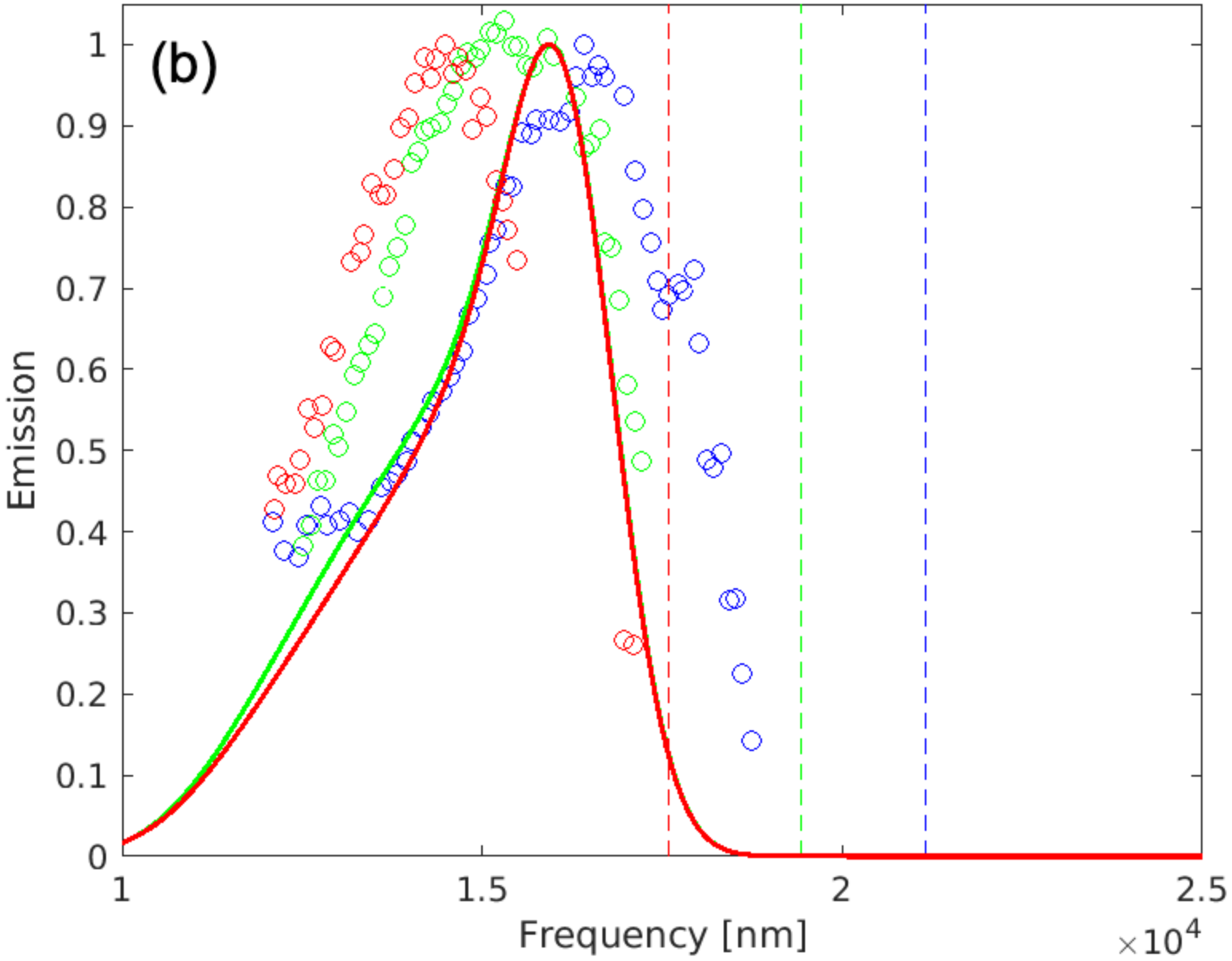

(c)
θ^{α}_{TM}
θ^{β}_{TM}
$f(E_3)$
$f(E_1)$
AE Energy storage [eV]
0
20
40
60
80
2
4
6
8
10
0.25
0.50
0.75
1.00
1.25
1.50
1.75

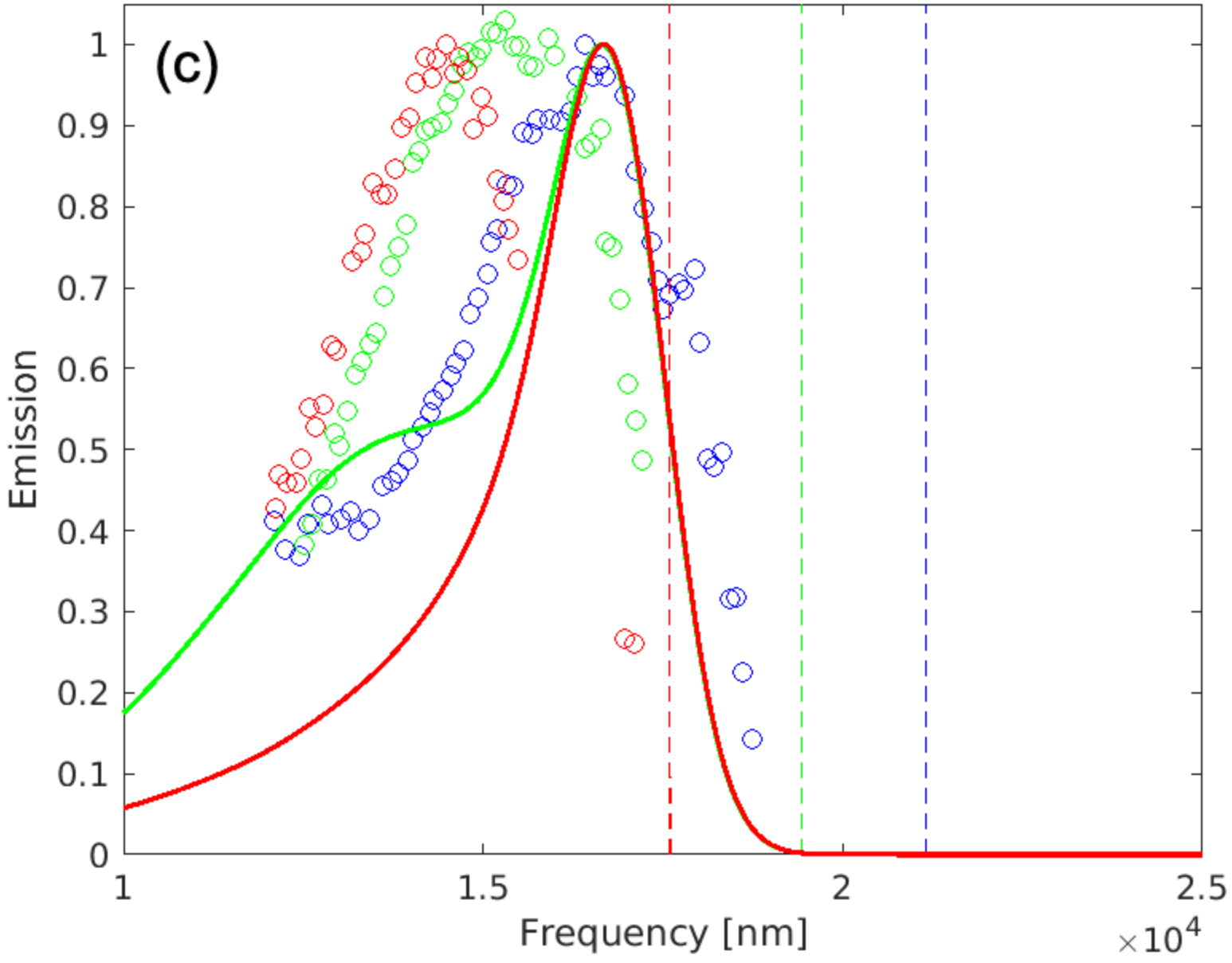

(a)
16
14
12
10
8
6
4
$f(E_2)$
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
$f(E_1)$
θ_{Omod}
0.4
0.3
0.2
0.1
AE Energy storage [eV]

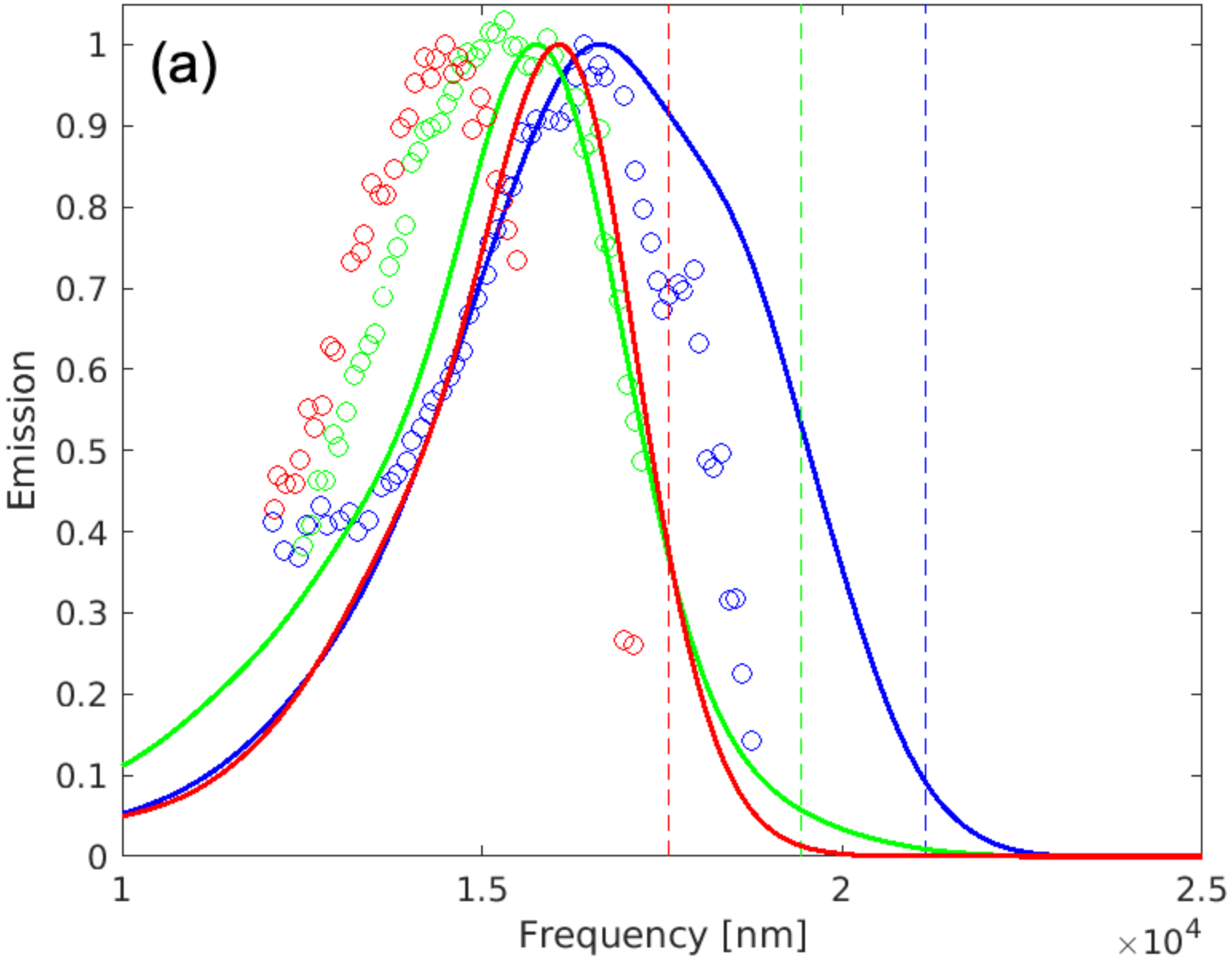

θ_{Omod}
(b)
$f(E_3)$
$f(E_2)$
AE Energy storage [eV]
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
4
6
8
10
12
14
16
0.4
0.3
0.2
0.1

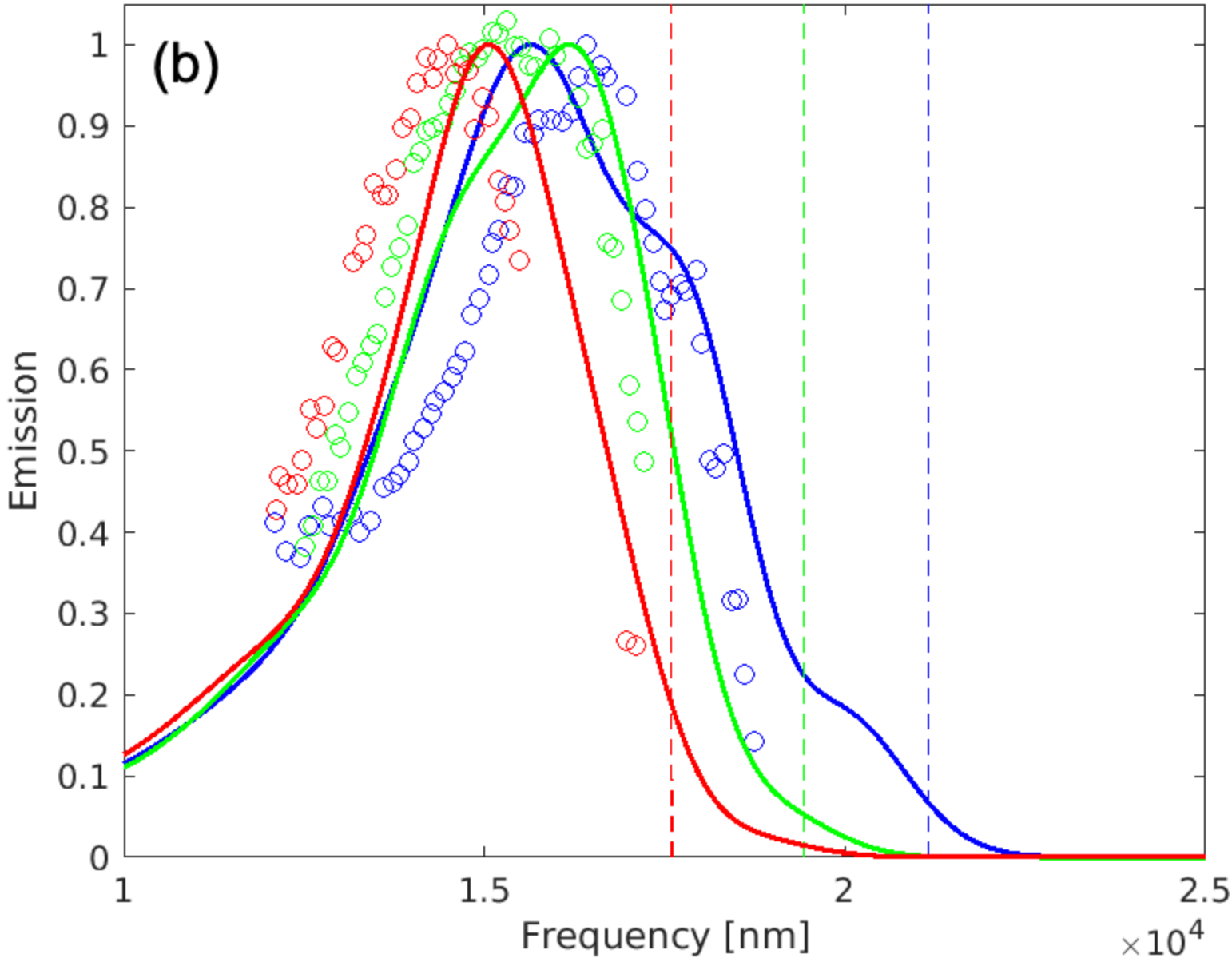

(c)
θ_{Omod}
$f(E_3)$
$f(E_1)$
AE Energy storage [eV]
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
0.4
0.3
0.2
0.1

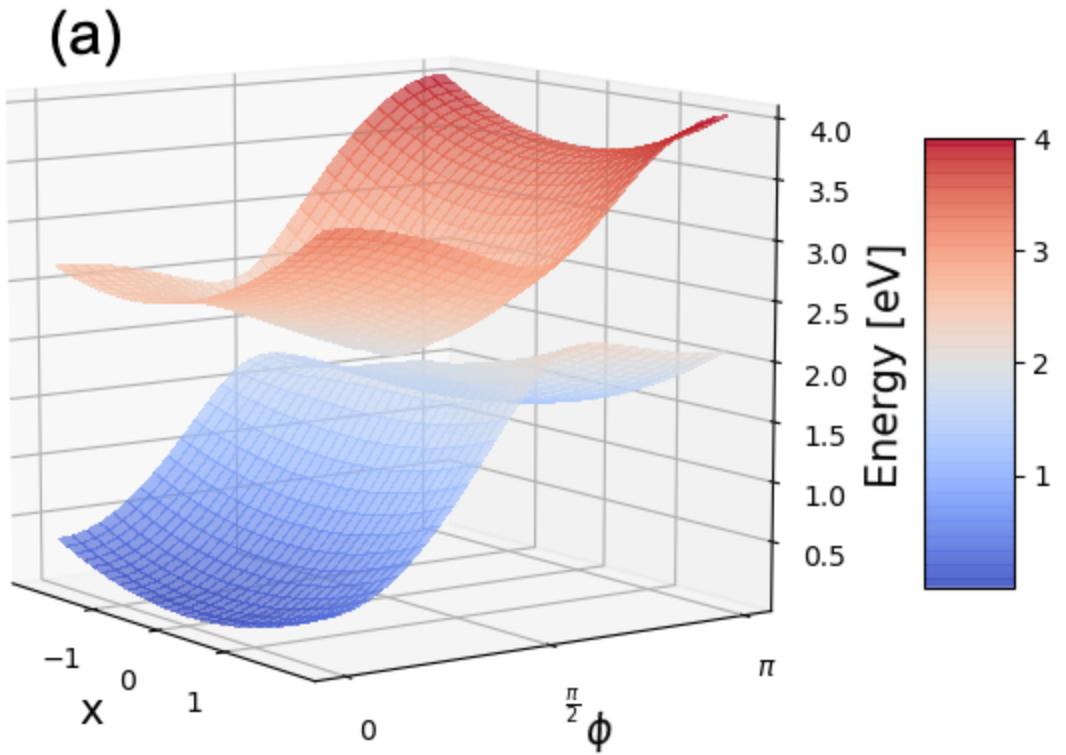

(a)
Energy [eV]
2.5
2.0
1.5
1.0
0.5
0.0
0
π/2
π
3π/2
ϕ
θ^0_{SM}
θ^α_{SM}
θ^β_{SM}

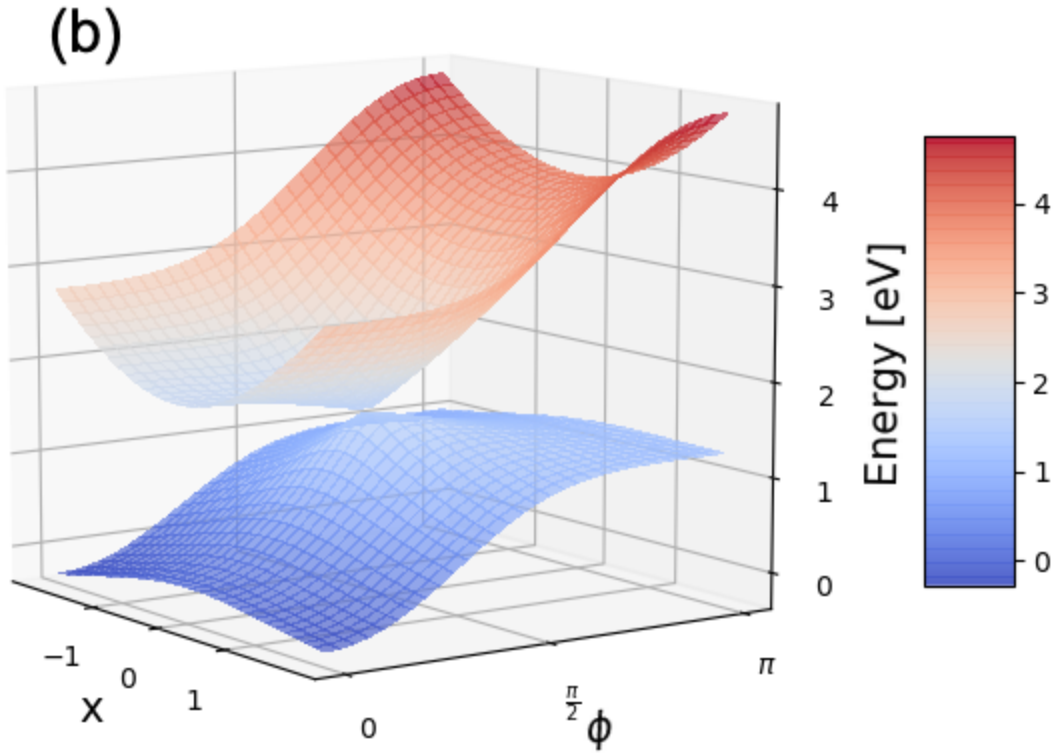

(b)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

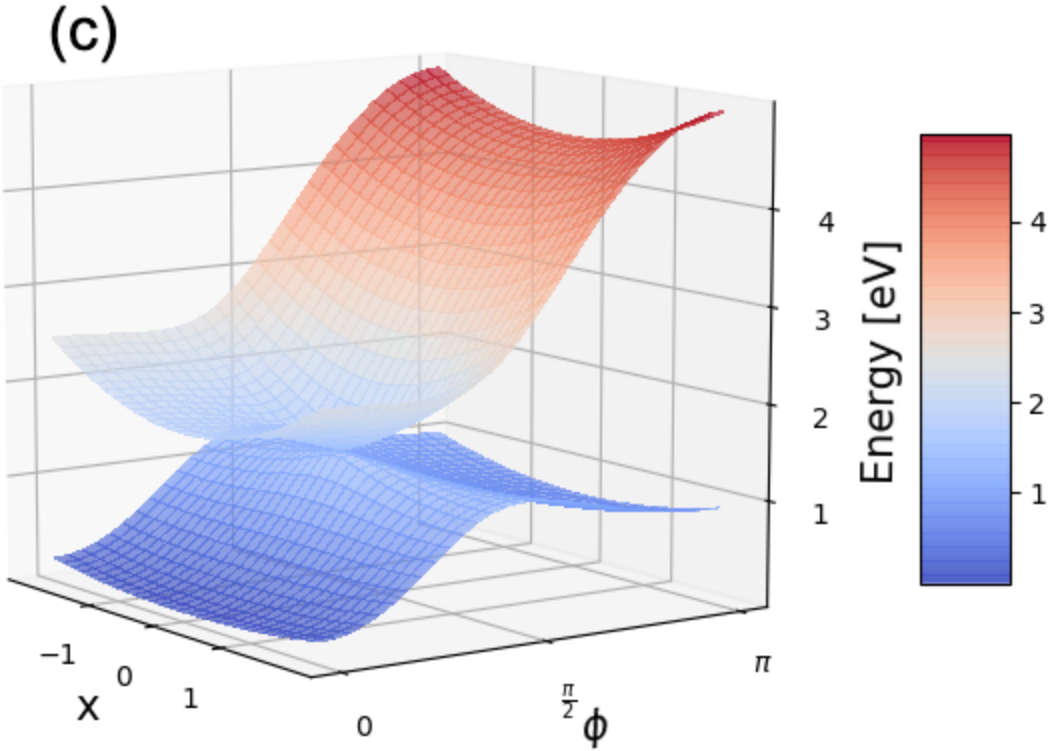

(c)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

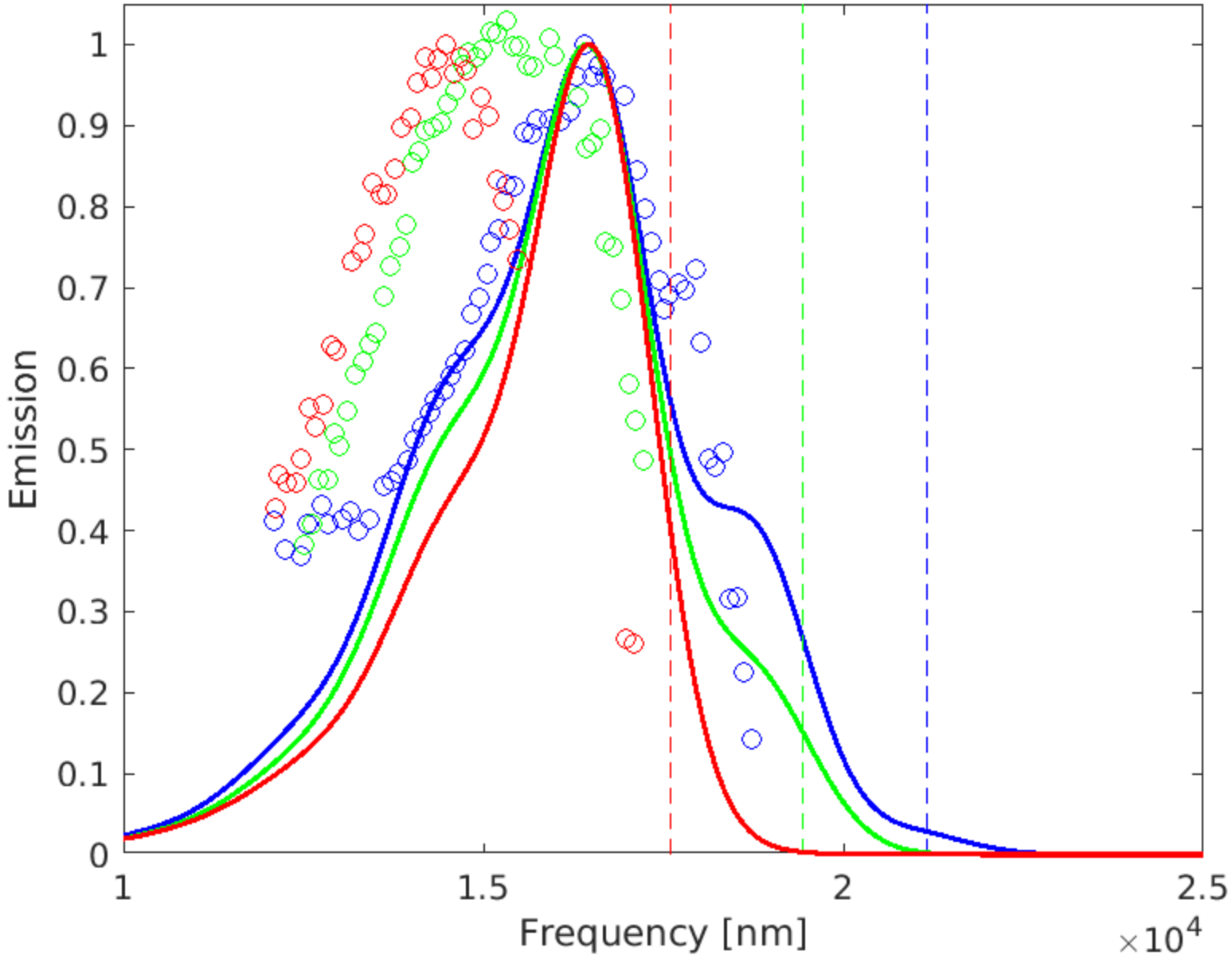

(a)
Emission
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5
Frequency [nm]
×10^4

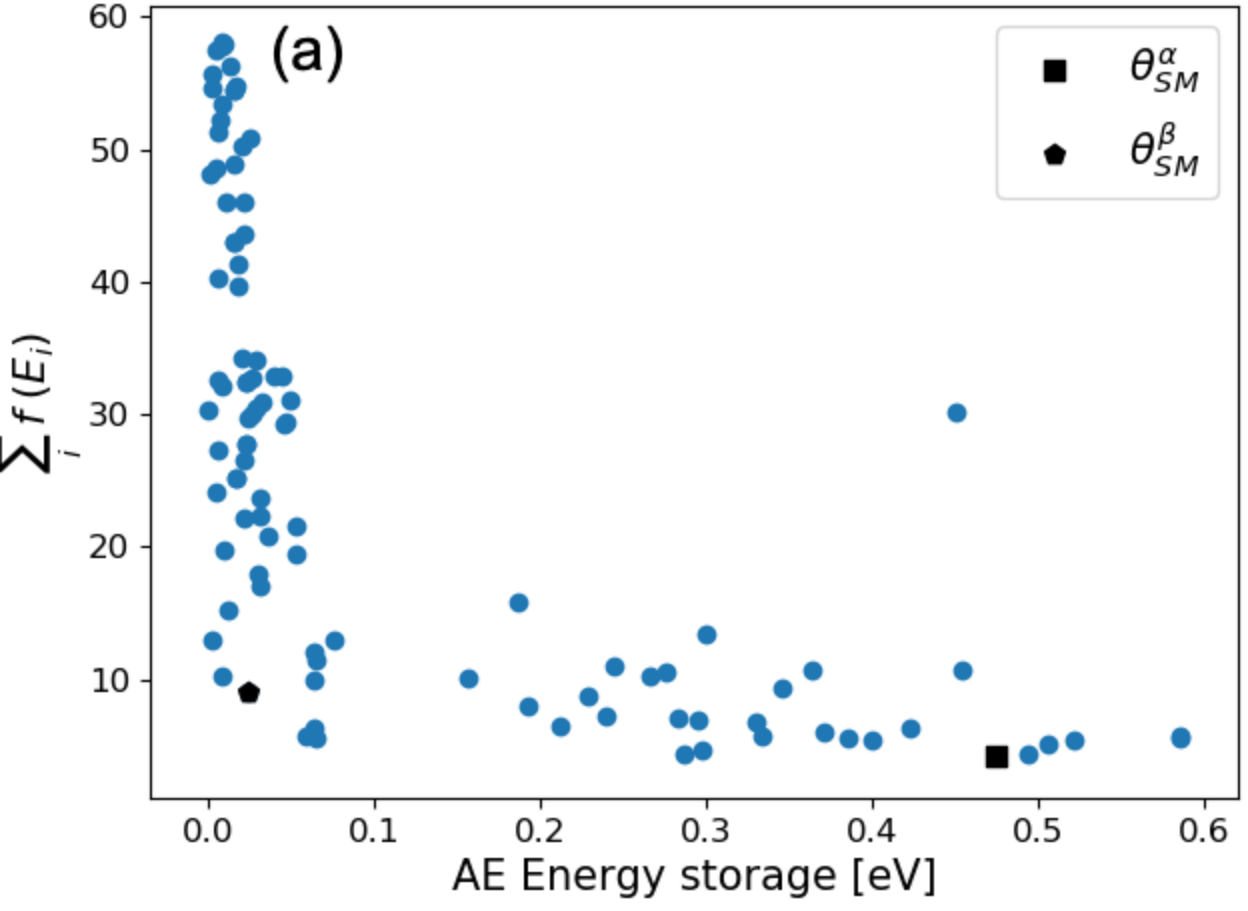

(b)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

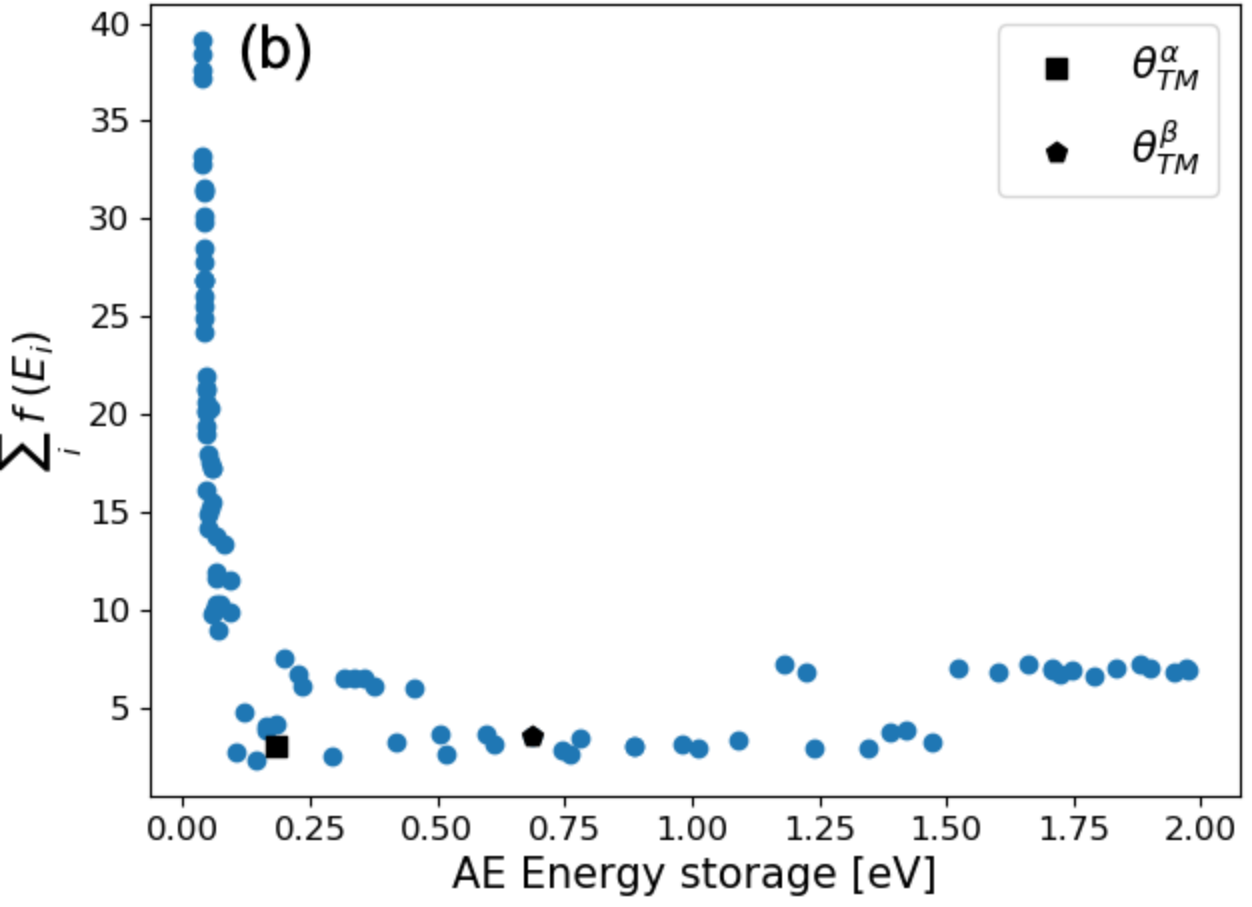

(a)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Energy [eV]
4
3
2
1
−1
0
1
x
0
$\frac{\pi}{2}$
ϕ
π

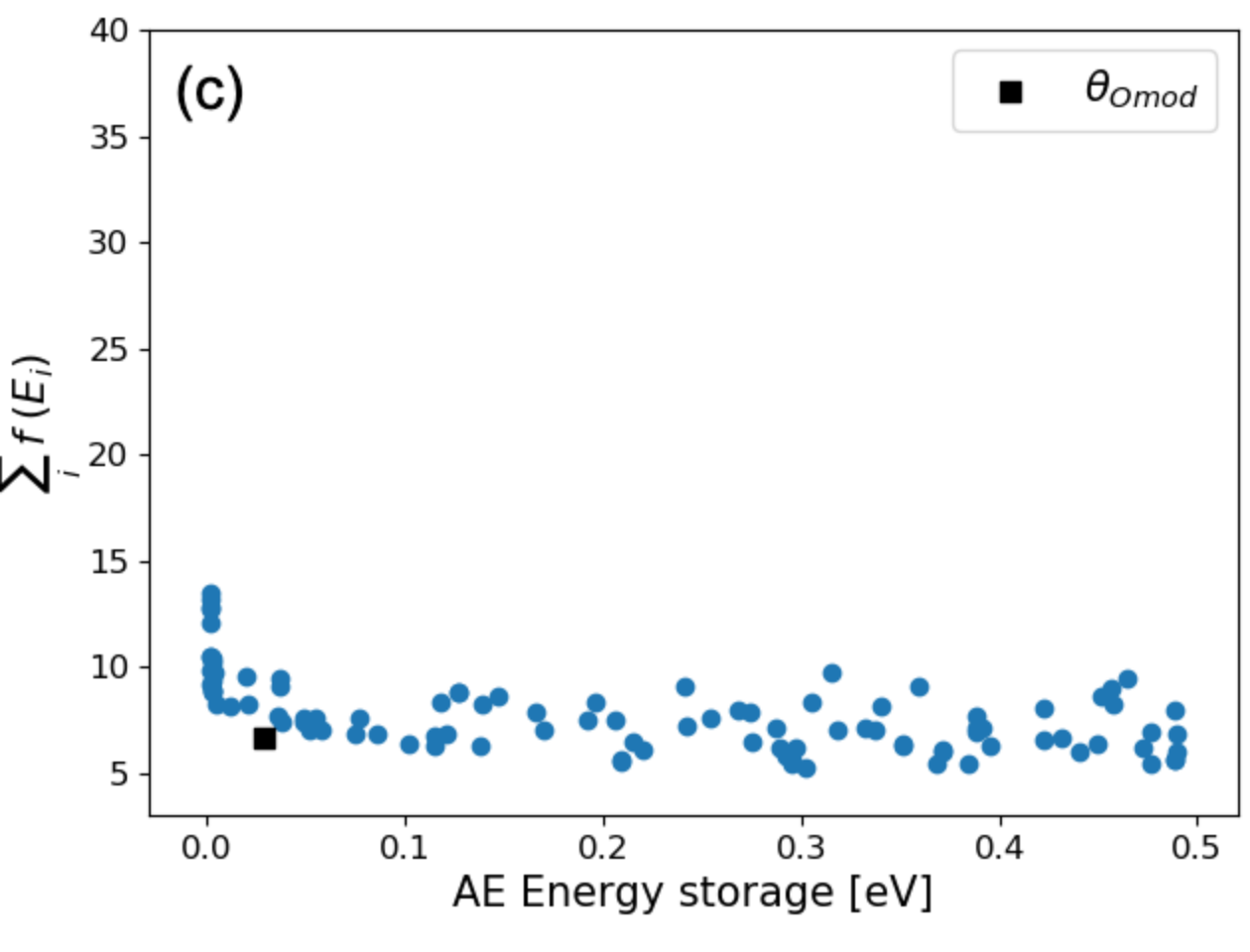

(b)
4
3
2
1
0
Energy [eV]
4
3
2
1
0
−1
0
1
x
0
$\frac{\pi}{2}$
π
ϕ

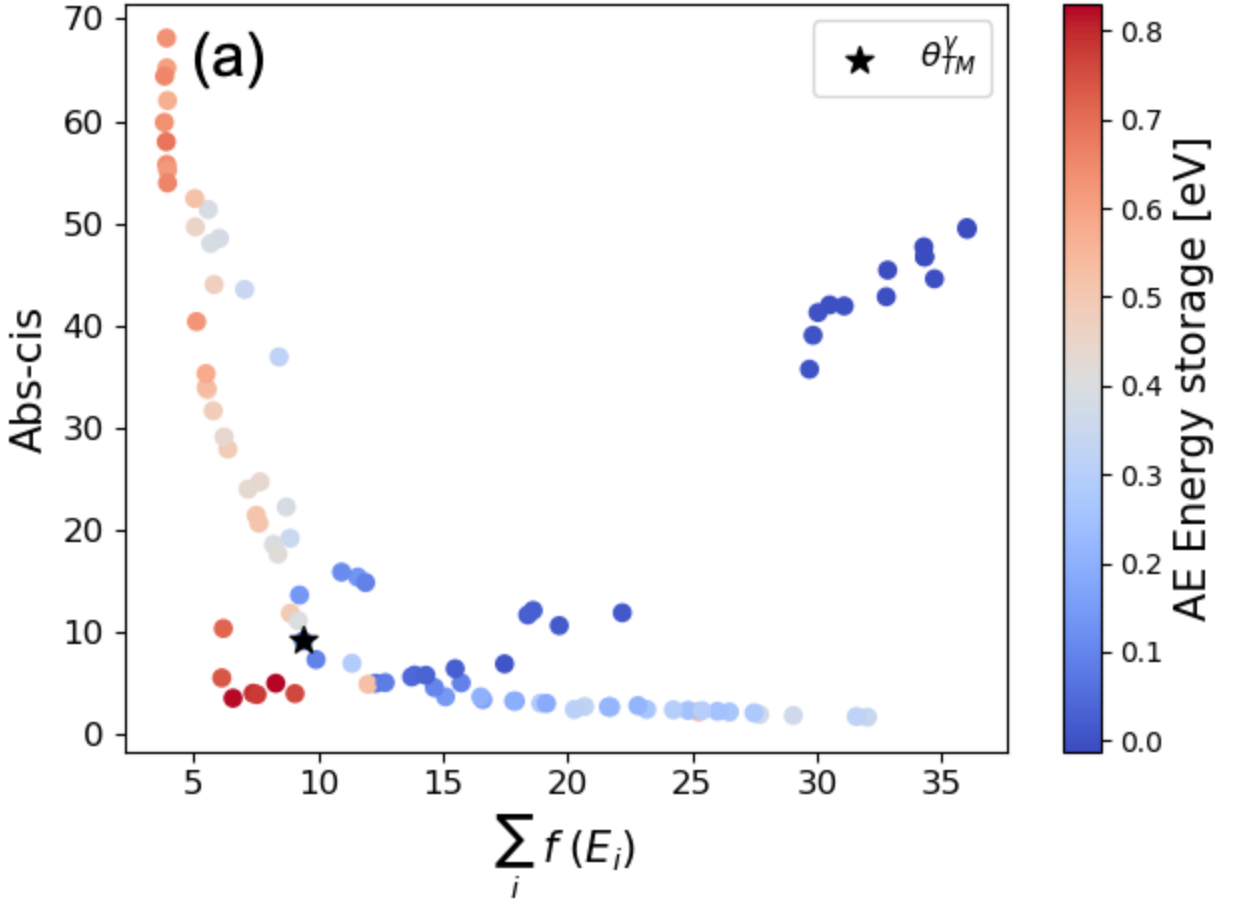

(c)
Energy [eV]
4
3
2
1
x
−1
0
1
ϕ
0
π/2
π

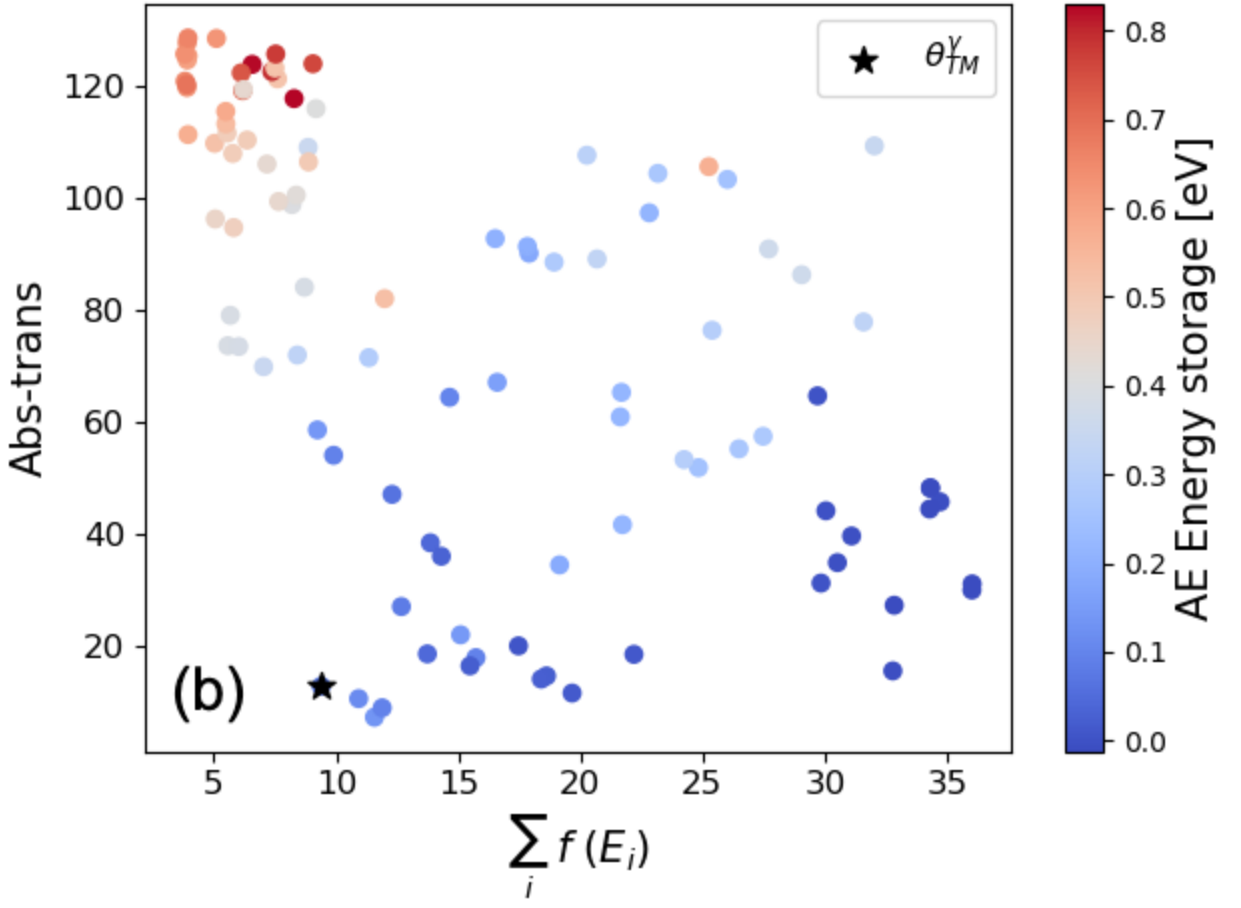

Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5

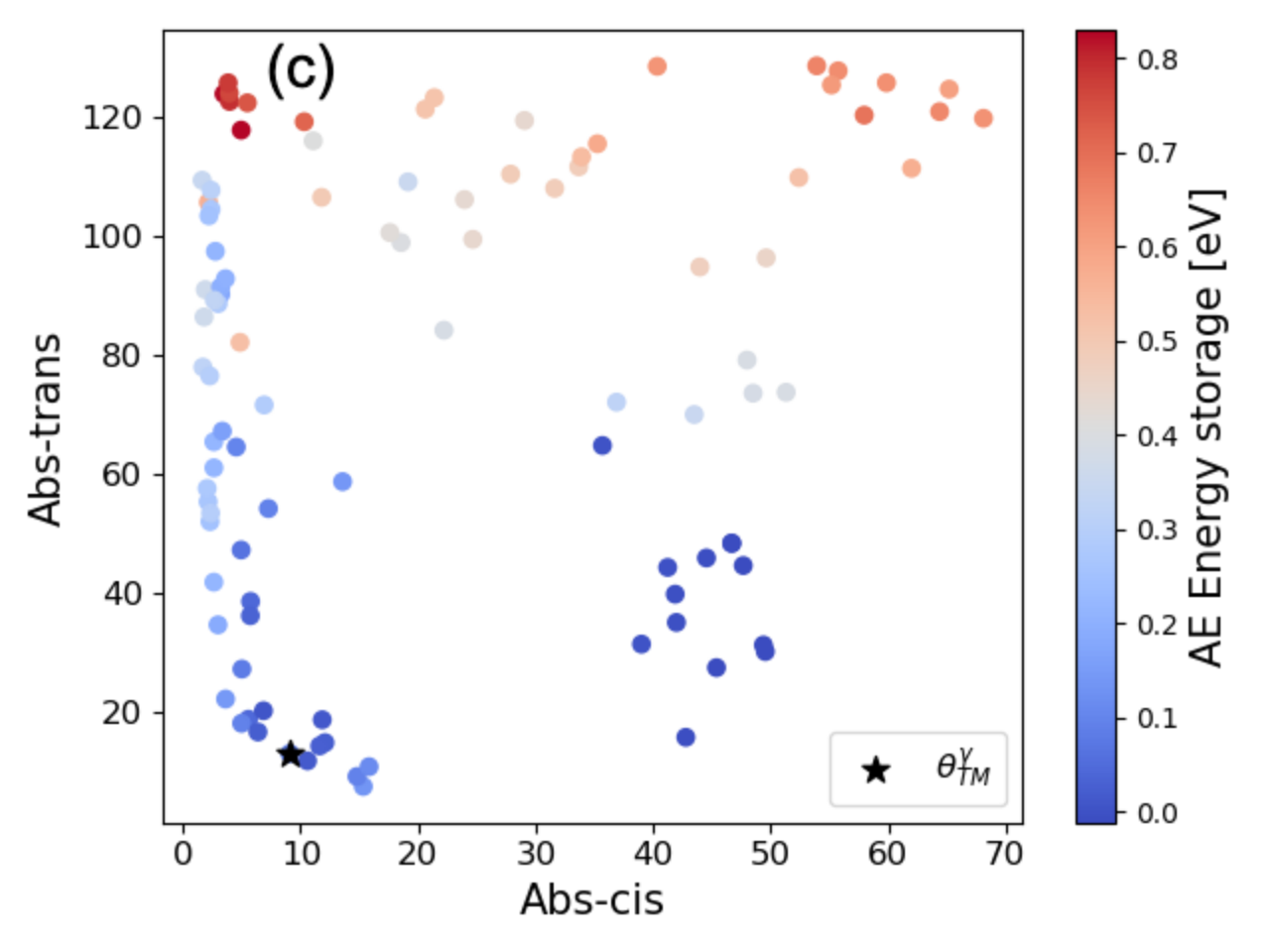

(a)
θ^{α}_{SM}
θ^{β}_{SM}
$\sum_i f(E_i)$
AE Energy storage [eV]
60
50
40
30
20
10
0.0
0.1
0.2
0.3
0.4
0.5
0.6

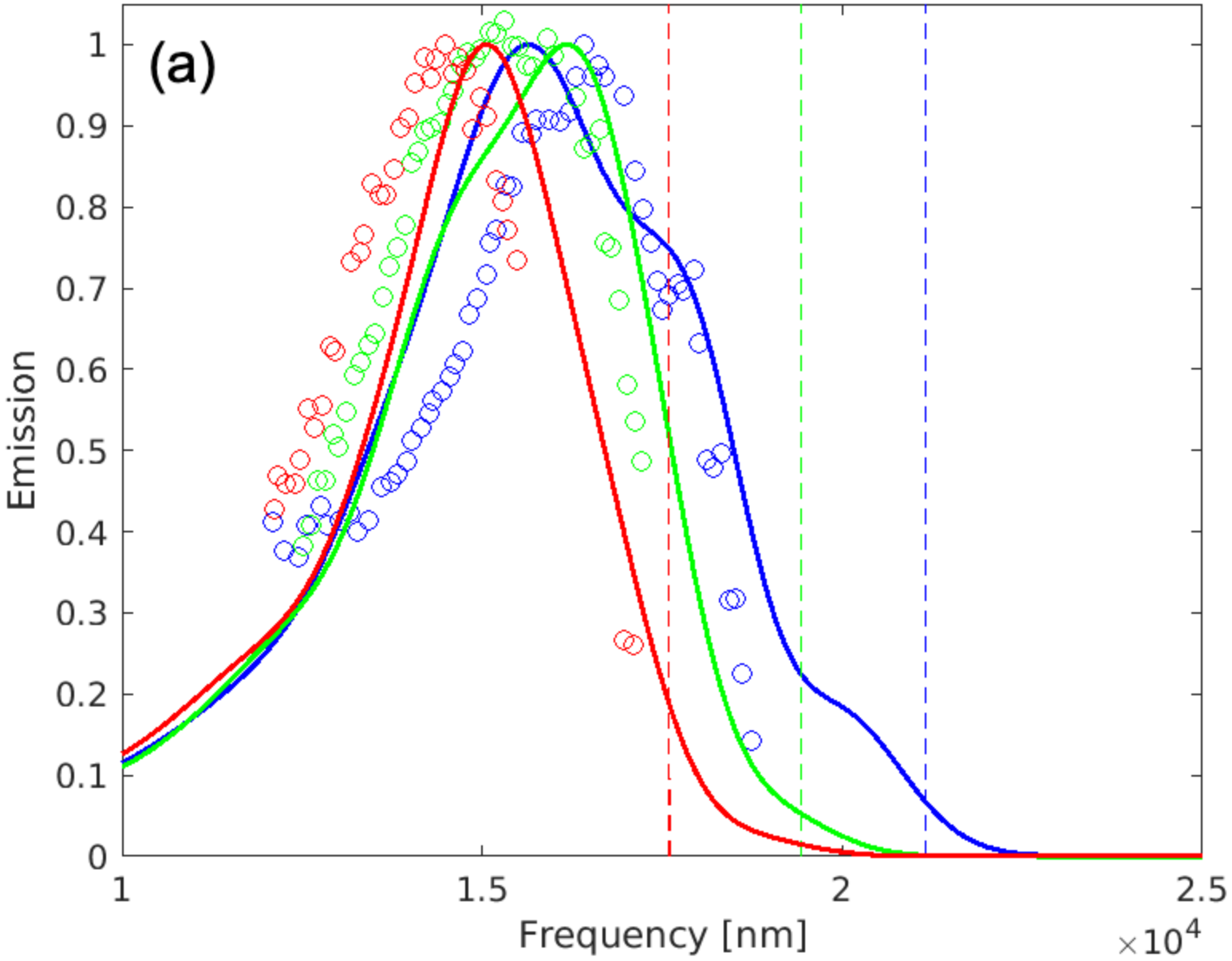

(b)
θ^{α}_{TM}
θ^{β}_{TM}
$\sum_i f(E_i)$
AE Energy storage [eV]
40
35
30
25
20
15
10
5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

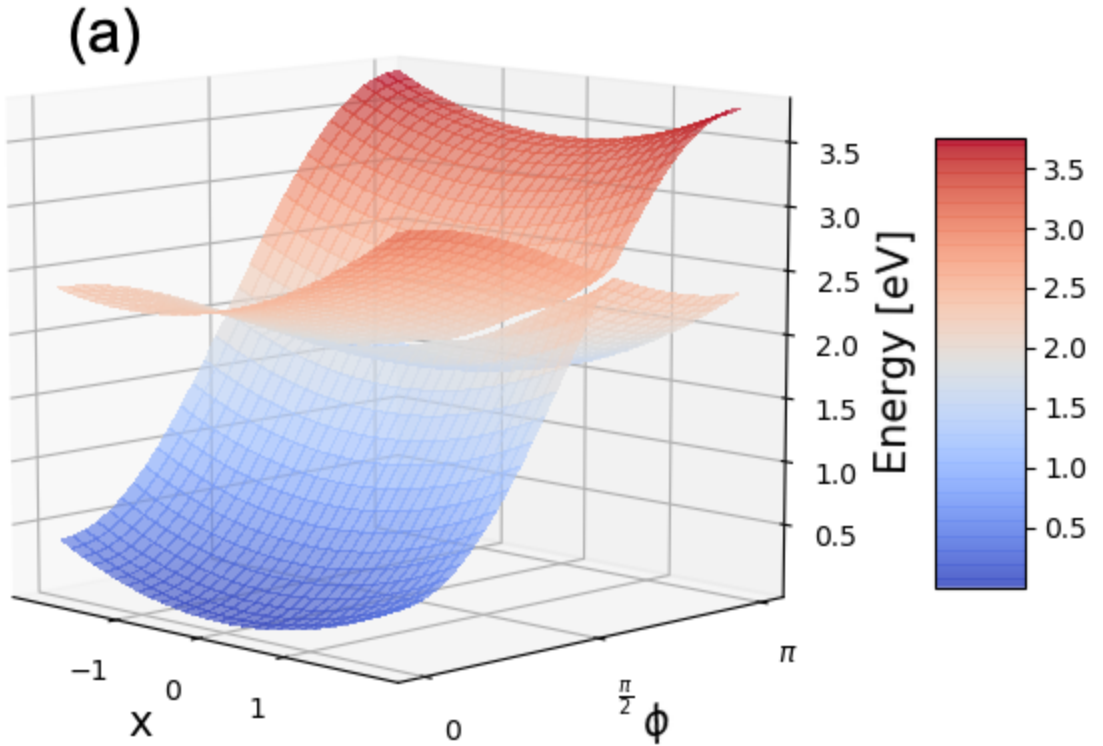

(c)
θ_{Omod}
$\sum_i f(E_i)$
AE Energy storage [eV]
40
35
30
25
20
15
10
5
0.0
0.1
0.2
0.3
0.4
0.5

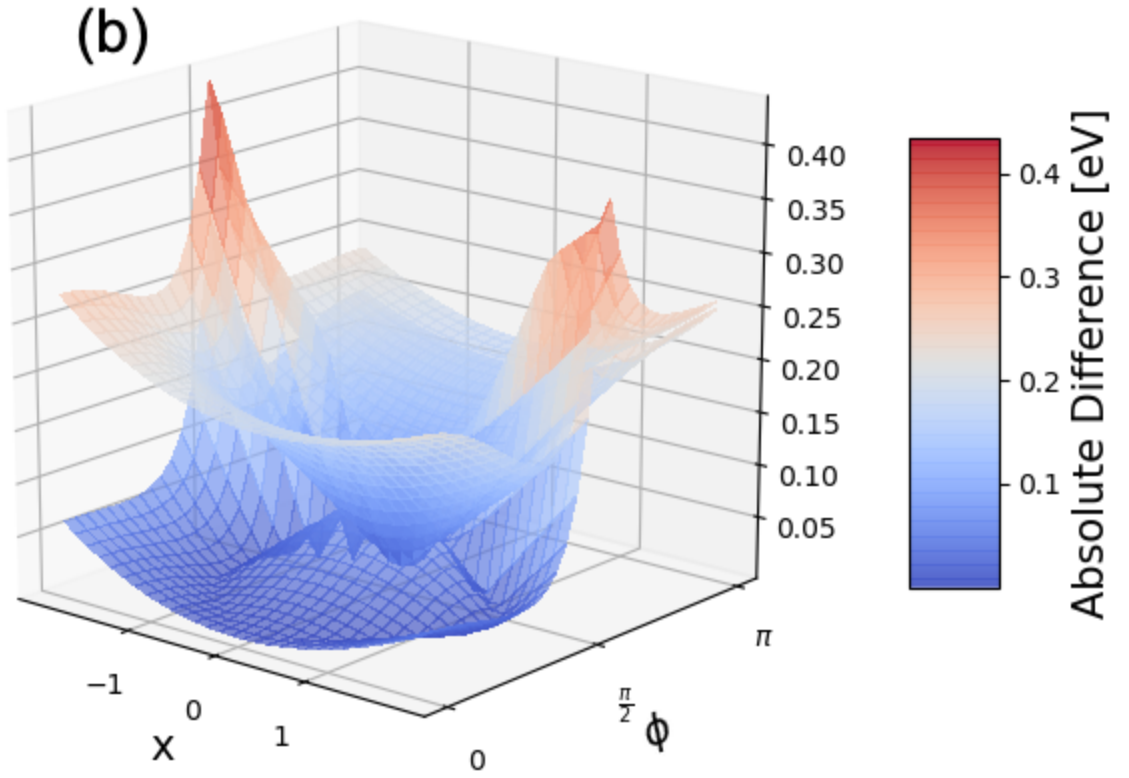

(a)
θ^{γ}_{TM}
Abs-cis
$\sum_i f(E_i)$
AE Energy storage [eV]
0
10
20
30
40
50
60
70
5
10
15
20
25
30
35
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

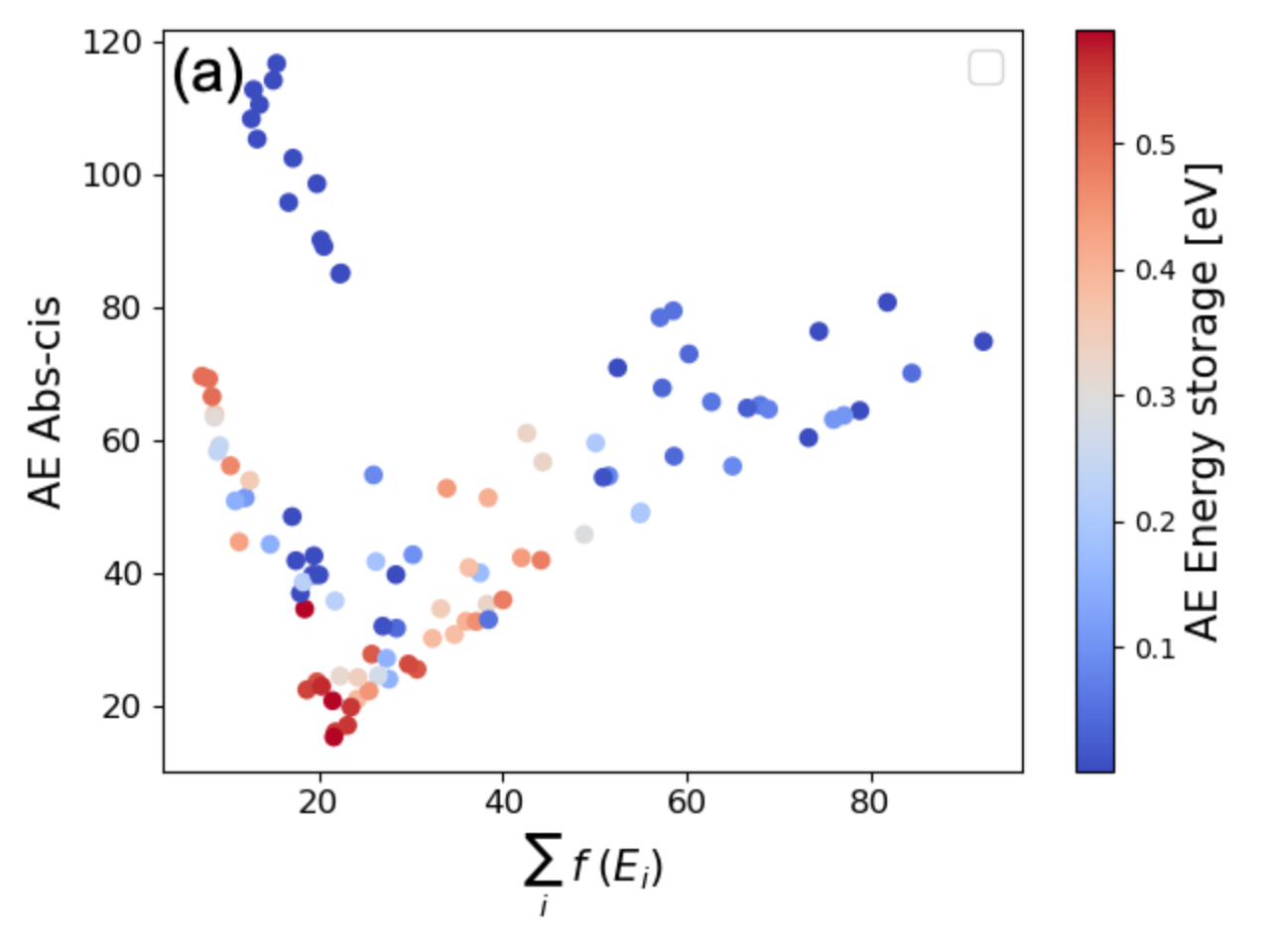

Abs-trans
120
100
80
60
40
20
(b)
5
10
15
20
25
30
35
$\sum_i f(E_i)$
θ^{γ}_{TM}
AE Energy storage [eV]
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

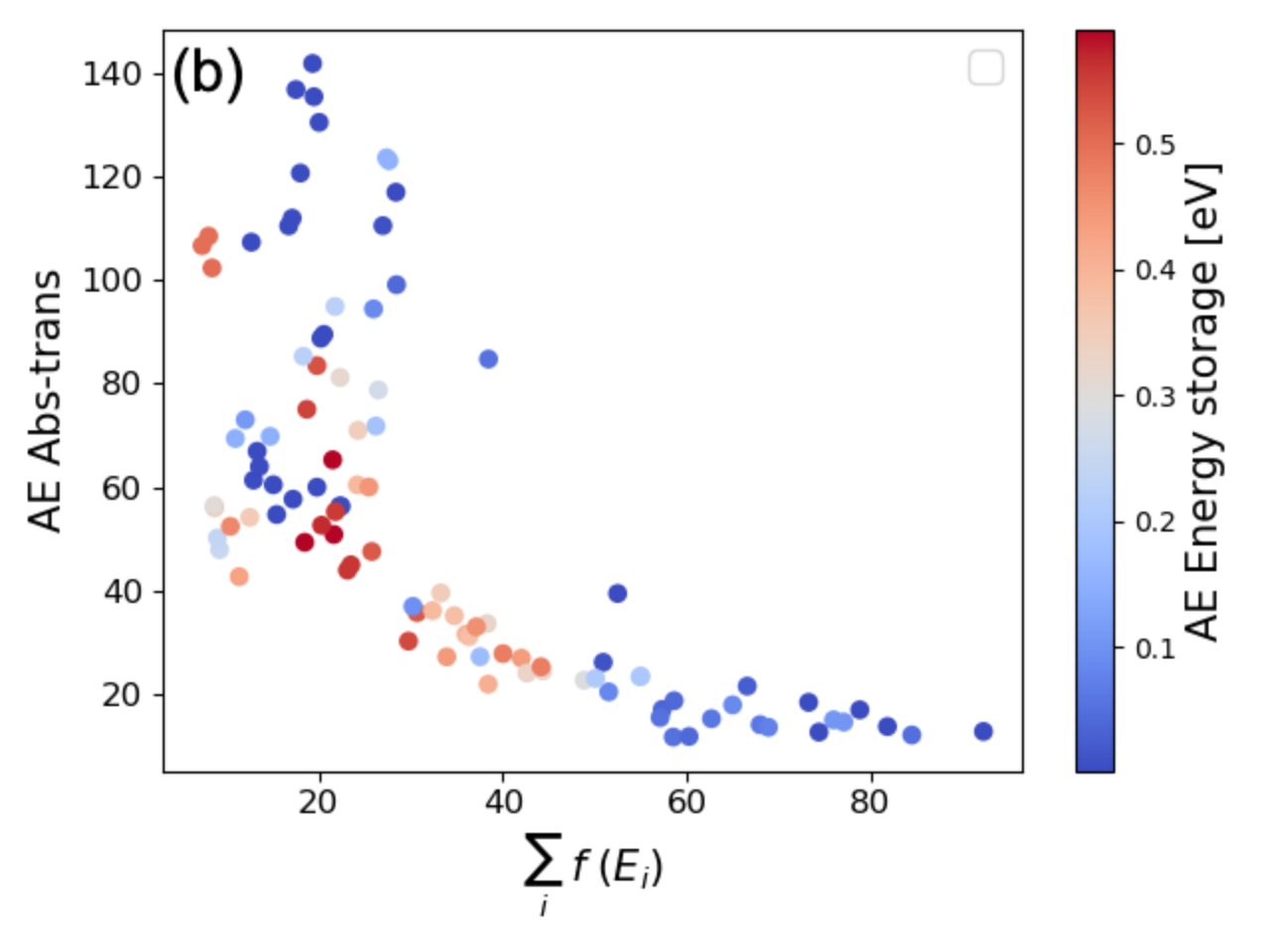

(c)
Abs-trans
Abs-cis
AE Energy storage [eV]
θ^{γ}_{TM}
0
10
20
30
40
50
60
70
20
40
60
80
100
120
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

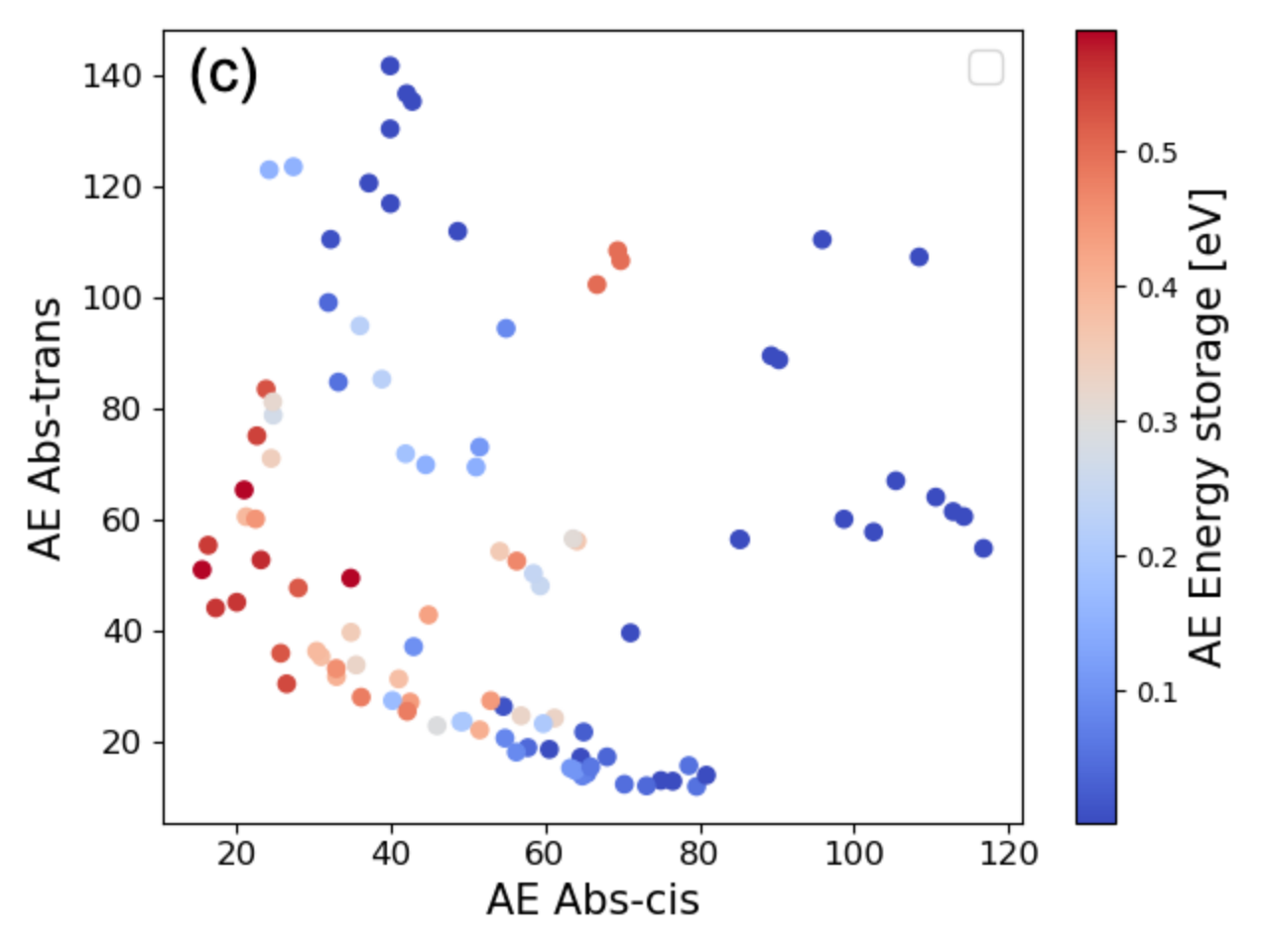

(a)
Emission
Frequency [nm]
×10^4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
1.5
2
2.5


(a)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Energy [eV]
3.5
3.0
2.5
2.0
1.5
1.0
0.5
−1
0
1
x
0
π/2
π
ϕ


(b)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
−1
0
1
x
0
$\frac{\pi}{2}$
π
ϕ
0.4
0.3
0.2
0.1
Absolute Difference [eV]


(a)
AE Abs-cis
120
100
80
60
40
20
20
40
60
80
$\sum_i f(E_i)$
AE Energy storage [eV]
0.5
0.4
0.3
0.2
0.1


(b)
AE Abs-trans
140
120
100
80
60
40
20
20
40
60
80
$\sum_i f(E_i)$
AE Energy storage [eV]
0.5
0.4
0.3
0.2
0.1


(c)
AE Abs-trans
AE Abs-cis
AE Energy storage [eV]
20
40
60
80
100
120
140
0.1
0.2
0.3
0.4
0.5