


Unsupervised machine learning techniques for exploring tropical coamoeba, brane tilings and Seiberg duality

Rak-Kyeong Seong^{✉*}

*Department of Mathematical Sciences, and Department of Physics,
Ulsan National Institute of Science and Technology,
50 UNIST-gil, Ulsan 44919, South Korea*

 (Received 13 September 2023; accepted 13 October 2023; published 8 November 2023)

We introduce unsupervised machine learning techniques in order to identify toric phases of $4d \mathcal{N} = 1$ supersymmetric gauge theories corresponding to the same toric Calabi-Yau 3-fold. These $4d \mathcal{N} = 1$ supersymmetric gauge theories are world volume theories of a D3-brane probing a toric Calabi-Yau 3-fold and are realized in terms of a type IIB brane configuration known as a brane tiling. It corresponds to the skeleton graph of the coamoeba projection of the mirror curve associated to the toric Calabi-Yau 3-fold. When we vary the complex structure moduli of the mirror Calabi-Yau 3-fold, the coamoeba and the corresponding brane tilings change their shape, giving rise to different toric phases related by Seiberg duality. We illustrate that by employing techniques such as principal component analysis and t -distributed stochastic neighbor embedding, we can project the space of coamoeba labeled by complex structure moduli down to a lower-dimensional phase space with phase boundaries corresponding to Seiberg duality. In this work, we illustrate this technique by obtaining a 2-dimensional phase diagram for brane tilings corresponding to the cone over the zeroth Hirzebruch surface F_0 .

DOI: [10.1103/PhysRevD.108.106009](https://doi.org/10.1103/PhysRevD.108.106009)

I. INTRODUCTION

The world volume theories of a D3-brane probing a toric Calabi-Yau 3-fold [1,2] form a very rich class of $4d \mathcal{N} = 1$ supersymmetric gauge theories [3–8]. These supersymmetric gauge theories are realized by a type IIB brane configuration that takes the form of a bipartite periodic graph on a 2-torus. Such bipartite graphs have been extensively studied in mathematics as dimers [9,10], and are known as brane tilings [11–13] in string theory. More recently, brane tilings have also been studied in relation to integrable systems [14,15], scattering amplitudes [16,17], and lattice gauge theories [18,19].

Under T-duality, the D3-brane probing the toric Calabi-Yau 3-fold singularity becomes a D5-brane suspended between a NS5-brane wrapping a holomorphic curve Σ [20–22]. The holomorphic curve is given by

$$\Sigma: P(x, y) = 0, \quad (1.1)$$

where $P(x, y)$ is the Newton polynomial in $x, y \in \mathbb{C}^*$ of the toric diagram Δ corresponding to the toric Calabi-Yau 3-fold. The Newton polynomial is defined as

$$P(x, y) = \sum_{(n_x, n_y) \in \Delta} c_{(n_x, n_y)} x^{n_x} y^{n_y}, \quad (1.2)$$

where $(n_x, n_y) \in \mathbb{Z}^2$ are the coordinates of the vertices of the 2-dimensional convex lattice polygon Δ , and $c_{(n_x, n_y)} \in \mathbb{C}^*$ are complex coefficients, which are complex structure moduli of the mirror Calabi-Yau 3-fold [23–25].

Brane tilings and the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theories have been extensively studied using mirror symmetry of the corresponding toric Calabi-Yau 3-fold [24,26]. The mirror geometry of the toric Calabi-Yau 3-fold is another 3-fold given by the double fibration over the complex W -plane,

$$W = P(x, y), \quad W = uv, \quad (1.3)$$

where $u, v \in \mathbb{C}$. The critical points of $P(x, y)$ are given by (x^*, y^*) and satisfy

$$\frac{\partial}{\partial x} P(x, y) \Big|_{(x^*, y^*)} = 0, \quad \frac{\partial}{\partial y} P(x, y) \Big|_{(x^*, y^*)} = 0. \quad (1.4)$$

On the W -plane, the critical points correspond to $W^* = P(x^*, y^*)$. When the toric diagram Δ contains at

*seong@unist.ac.kr

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

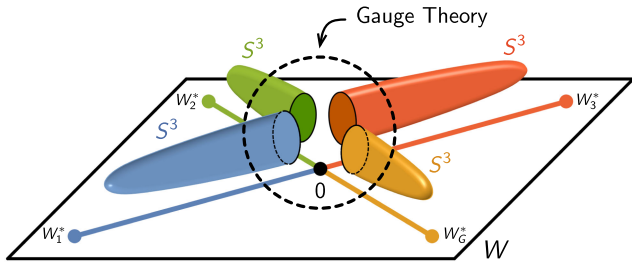


FIG. 1. The mirror geometry of a toric Calabi-Yau 3-fold is another 3-fold given by a double fibration over the complex W -plane. Along a vanishing path connecting $W = 0$ with a critical point W^* , we obtain from the double fibration an overall S^3 . The S^3 's meet at $W = 0$ and how they intersect is given by the corresponding brane tiling that represents a $4d \mathcal{N} = 1$ supersymmetric gauge theory.

least one internal point, the number of critical points of P matches the normalized area of the toric diagram, which is also the number of gauge groups G in the corresponding $4d$ theory [27].

The fiber associated to $P(x, y)$ in (1.3) corresponds to a holomorphic curve $\Sigma_W: P(x, y) - W = 0$, while the fiber associated to uv in (1.3) is a \mathbb{C}^* fibration. For generic values of the complex structure moduli $c_{(n_x, n_y)}$ in (1.2), an $S^1 \subset \Sigma_W$ shrinks to zero size at each critical point W^* , whereas additionally the S^1 from the uv -fibration vanishes at the origin $W = 0$. Overall, from this $S^1 \times S^1$ fibration over a vanishing path connecting $W = 0$ to a specific critical point W^* , we obtain an S^3 as illustrated in Fig. 1. All these S^3 's meet at $W = 0$, where the uv -fibration vanishes. D6-branes wrapping these S^3 's give rise to the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theory, and the way how these spheres intersect each other at the vanishing locus $P(x, y) = W = 0$ specifies the quiver and superpotential of the $4d$ theory. The intersection structure is precisely what is given by the holomorphic curve Σ in (1.1), which is represented by the corresponding brane tiling.

For different values of the complex structure moduli $c_{(n_x, n_y)}$, the positions of the critical points W^* change and therefore the configuration of the S^3 's change. This gives rise to a different intersection structure between the spheres and hence to a different $4d$ gauge theory phase. These phases are all represented by brane tilings and are also known as toric phases [7,28,29]. They are related by a local deformation of the brane tiling, which corresponds to Seiberg duality for $4d$ theories [30]. This local deformation is also known as urban renewal or spider move [15,31,32] in the mathematics literature and is illustrated in Fig. 2.

In the following work, we focus on brane tilings given by the coamoeba description [26,33] of the holomorphic curve Σ . When we project the holomorphic curve Σ onto a 2-torus using

$$(x, y) = (r_x e^{i\theta_x}, r_y e^{i\theta_y}) \mapsto (\theta_x, \theta_y) \in T^2, \quad (1.5)$$

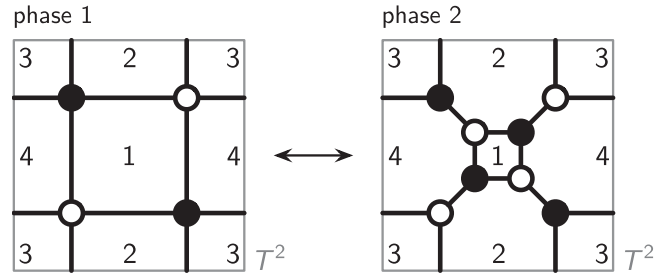


FIG. 2. A local deformation of a brane tiling known as urban renewal or spider move corresponds to Seiberg duality between the $4d \mathcal{N} = 1$ theories represented by the brane tilings connected by the local deformation.

we obtain what is known as the coamoeba projection [26,33] of Σ . The coamoeba projection identifies the locations on T^2 where the D5-brane meets the NS5-brane wrapping Σ in the T-dual description of the D3-brane probing the toric Calabi-Yau 3-fold. The skeleton graph of the coamoeba, as illustrated in Fig. 3, is precisely what we call as the brane tiling [11–13], which represents the $4d \mathcal{N} = 1$ supersymmetric gauge theory corresponding to the toric Calabi-Yau 3-fold. We note that there is a one-to-one map between a specific coamoeba and the corresponding brane tiling, and a given brane tiling uniquely identifies the associated $4d \mathcal{N} = 1$ supersymmetric gauge theory.

The choice of complex structure moduli $c_{(n_x, n_y)} \in \mathbb{C}^*$ in the Newton polynomial $P(x, y)$ determines the shape of the coamoeba on T^2 . When the complex structure moduli pass critical values, the coamoeba transforms into a new shape, which corresponds to a new toric phase related by Seiberg duality to the original $4d$ theory.

Although, it is well-understood that the complex structure moduli $c_{(n_x, n_y)} \in \mathbb{C}^*$ affect the shape of the coamoeba, it is in general not well-understood which values of the complex structure moduli correspond to which toric phase. We propose in this work a new method to explore and parameterize using complex structure moduli the phase space for brane tilings corresponding to a given toric Calabi-Yau 3-fold. Given that the number of complex structure moduli increases with increasing number of vertices in the toric diagram of the toric Calabi-Yau

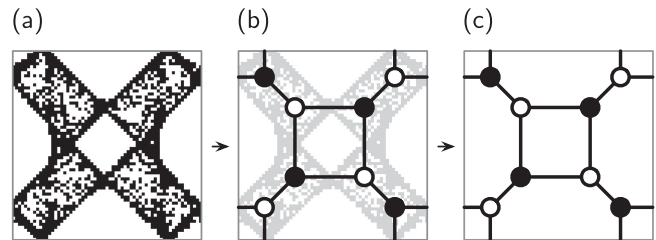


FIG. 3. (a) A coamoeba plot on T^2 (b) can be mapped to its corresponding unique skeleton graph, (c) which is a brane tiling on T^2 . The brane tiling is a bipartite graph where edges connect nodes with different colors.

3-fold, the problem of identifying regions in the space of all possible values of the complex structure moduli associated to specific toric phases is extremely challenging.

We therefore propose unsupervised machine learning techniques in order to simplify the space of all possible coamoeba corresponding to a toric Calabi-Yau 3-fold. By using principal component analysis (PCA) [34–38] and additionally t -distributed stochastic neighbor embedding (t -SNE) [39,40], the space of coamoeba labeled by complex structure moduli can be dimensionally reduced to a lower-dimensional phase space. Such a phase space obtained using machine learning allows us to visualize regions in the phase space as toric phases and boundaries between regions as phase boundaries corresponding to Seiberg duality. Our proposed method also enables us to construct probability functions in terms of the complex structure moduli, whose probability values at certain choices of the complex structure moduli identify the corresponding toric phases.

Our work follows a long list of applications of machine learning techniques in string theory, beginning with the pioneering works in [41–49], and aims for explainable and interpretable results in string theory and mathematics through machine learning. We focus in this work on unsupervised machine learning techniques and the presentation of interpretable results in the context of coamoeba and brane tilings, which is in contrast to some previous attempts [50] on applying supervised machine learning techniques for the study of amoeba and brane webs [51,52]. We also note that our emphasis on the interpretation of the space of coamoeba under PCA and t -SNE as a phase space for brane tilings and corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theories is in stark contrast to the focus on numerical accuracy of estimating amoeba features in [53,54].

II. BRANE TILINGS AND COAMOEBEA

$4d \mathcal{N} = 1$ supersymmetric gauge theories corresponding to toric Calabi-Yau 3-folds can be realized in terms of a type IIB brane configuration known as a brane tiling [11–13]. A brane tiling consists of D5-branes suspended from a NS5-brane, where the NS5-brane extends along the (0123) directions and wraps a holomorphic curve Σ embedded into the (4567) directions. The coordinates (45) and (67) are pairwise combined into complex variables $x, y \in \mathbb{C}^*$, respectively. They are the complex coordinates in the Newton polynomial $P(x, y)$ and the arguments $(\arg(x), \arg(y)) = (\theta_x, \theta_y)$ are the coordinates of a T^2 . The brane configuration is summarized in Table I.

As discussed in (1.1), the holomorphic curve Σ is defined as the zero locus of the Newton polynomial $P(x, y)$ of the toric diagram Δ . The brane tiling is the skeleton graph of the coamoeba projection of Σ onto $(\arg(x), \arg(y)) \in T^2$ [26]. This is a projection of every point on the curve $\Sigma: P(x, y) = 0$ to its angular component, resulting in a doubly periodic image of the curve Σ

TABLE I. A $4d \mathcal{N} = 1$ gauge theory corresponding to a toric Calabi-Yau 3-fold can be represented by a type IIB brane configuration known as a brane tiling.

	0	1	2	3	4	5	6	7	8	9
D5	×	×	×	×	×	·	×	·	·	·
NS5	×	×	×	×			Σ		·	·

on T^2 . We refer to this image on T^2 as the coamoeba [26,33] of the toric Calabi-Yau 3-fold. The alternative projection of Σ onto $(|x|, |y|) = (r_x, r_y)$ is called the amoeba projection [55,56].

The brane tiling associated to the coamoeba is a bipartite periodic graph on T^2 . We can use the following dictionary to identify the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theory from a brane tiling [11,12]:

- (i) *White and black nodes.* Positive and negative terms in the superpotential W of the $4d \mathcal{N} = 1$ supersymmetric gauge theory correspond to white and black nodes in the brane tiling, respectively. The white and black nodes also have a clockwise and anticlockwise orientation, respectively.
- (ii) *Edges.* Bifundamental fields X_{ij} in the $4d \mathcal{N} = 1$ supersymmetric gauge theory are represented by edges in the brane tiling, which always connect nodes that have different colors. Going along the orientation of a node, one can identify the fields $X_{ij}, X_{jk}, \dots, X_{mi}$ associated to the specific superpotential term $\pm \text{Tr}(X_{ij}X_{jk}\dots X_{mi}) \in W$ in the correct cyclic order.
- (iii) *Faces.* $U(N_i)$ gauge groups correspond to faces of a brane tiling. Every edge in the brane tiling corresponding to a bifundamental field X_{ij} has neighboring faces associated to $U(N_i)$ and $U(N_j)$. The orientation of the bifundamental field X_{ij} is given by the orientation around the white and black nodes at the two opposite ends of the corresponding brane tiling edge.

In the following work, all gauge groups are considered to be $U(1)$ such that the mesonic moduli space [57–60] of the $4d \mathcal{N} = 1$ supersymmetric gauge theories is precisely the probed toric Calabi-Yau 3-fold. Under Seiberg duality, dual brane tilings keep corresponding to Abelian $4d$ theories with $U(1)$ gauge groups. Accordingly, for Abelian $4d$ theories corresponding to toric Calabi-Yau 3-folds, Seiberg duality is often referred to as toric duality in the literature [7,8,28].

As an example, Fig. 4 shows the brane tiling for the first toric phase corresponding to the Calabi-Yau cone over the zeroth Hirzebruch surface F_0 [7,61–63]. When we replace the faces as nodes, and edges as arrows along the orientation of the white and black nodes of the brane tiling, we obtain a periodic quiver [11] on T^2 . The T^2 unit cell of the periodic quiver gives the quiver diagram [5] of the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theory.

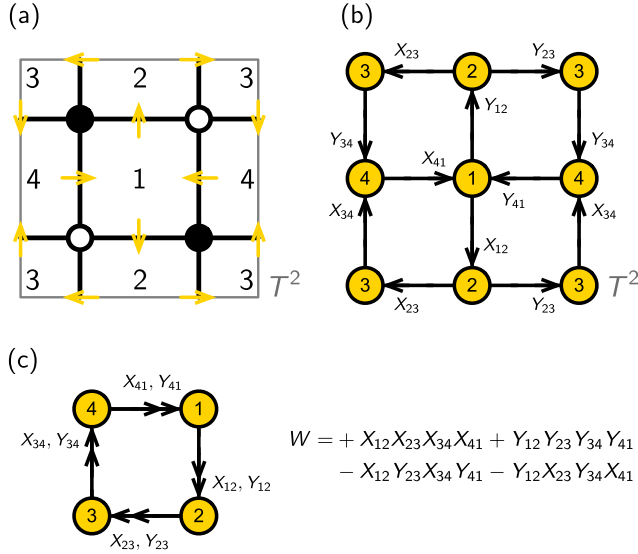


FIG. 4. (a) The brane tiling on T^2 with clockwise/anti-clockwise orientations on white/black nodes, (b) the periodic quiver on T^2 , and (c) the quiver and superpotential corresponding to the first toric phase of F_0 .

The faces of the brane tiling indicate the locations of the D5-brane suspended between the NS5-branes wrapping Σ in the brane construction in Table I. In the mirror symmetry description, the faces of the brane tiling are centered around the critical points $W^* = (x^*, y^*)$ from (1.4). Each critical point W_i^* corresponds to a gauge group, and when they are projected under the coamoeba map to T^2 , they form the centers of the brane tiling faces. The solutions to $P(x, y) = 0$ mapped to T^2 indicate the location of the holomorphic curve Σ on T^2 and is highlighted in black in the coamoeba plots as illustrated in Fig. 5. These regions surround the critical points W_i^* that are located inside the white regions of the coamoeba plot. In the following section, we review how we generate these coamoeba plots and how we illustrate them on T^2 .

III. COAMOEBA GENERATION AND REPRESENTATION

For the following work, the coamoeba projection of Σ onto T^2 plays a central role, because it directly corresponds to the brane tiling and $4d \mathcal{N} = 1$ supersymmetric gauge theory for a given toric Calabi-Yau 3-fold. From (1.2), we note that $\Sigma: P(x, y) = 0$ depends on the complex structure moduli $c_{(n_x, n_y)} \in \mathbb{C}^*$. In order to identify a correspondence between complex structure moduli $c_{(n_x, n_y)}$ and specific toric phases, we have to generate for the same toric Calabi-Yau 3-fold many coamoeba for different values of $c_{(n_x, n_y)}$.

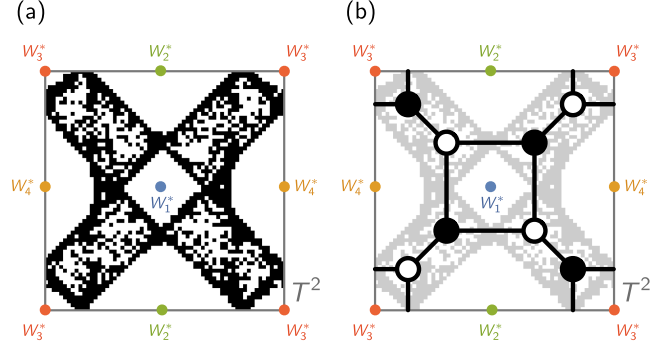


FIG. 5. Each critical point W_i^* on the W -plane corresponds to a $U(N)$ gauge group in the $4d \mathcal{N} = 1$ theory. When they are projected under the coamoeba map to T^2 , (a) they form the centers of white regions in the coamoeba plot and (b) are in the interior of the brane tiling faces.

As discussed above, the coamoeba is given by the angular projection of $\Sigma: P(x, y) = 0$ onto T^2 . In order to obtain the coamoeba plot on T^2 , one needs to solve for the set of solutions for $P(x, y) = 0$, which then can be mapped to T^2 using the coamoeba map. We make use of the Monte-Carlo method [64] and scan for solutions in the range of $\theta_x = \arg(x) \in [0, 2\pi)$ and $\theta_y = \arg(y) \in [0, 2\pi)$. We call the process of searching for solutions while varying θ_x the θ_x -scan, while the search for solutions while varying θ_y is called the θ_y -scan.

For the θ_x -scan, we first pick a random value $\theta_{x_0} \in [0, 2\pi)$ and a random value for $m_\epsilon \in [-\epsilon, \epsilon]$, where the positive parameter $\epsilon \ll 1$ is taken to be small. Here, ϵ is a measure of the overall thickness of the coamoeba after projecting Σ onto T^2 . By taking $x_0 = e^{i\theta_{x_0} + m_\epsilon}$, we can then numerically solve for y using

$$P(x = x_0, y) = \sum_{(n_x, n_y) \in \Delta} c_{(n_x, n_y)} x_0^{n_x} y^{n_y} = 0. \quad (3.1)$$

The solutions $y = y^*$ then combined with the value for x_0 can be mapped to T^2 by taking $(\theta_x^*, \theta_y^*) \equiv (\theta_{x_0}, \arg(y^*))$, where θ_{x_0} is given by $x_0 = e^{i\theta_{x_0} + m_\epsilon}$. In total, we take N_θ random values for $\theta_{x_0} \in [0, 2\pi)$ and $m_\epsilon \in [-\epsilon, \epsilon]$ in order to find corresponding solutions $y = y^*$ in (3.1). The resulting collection of solutions of the form $(\theta_{x_0}, \arg(y^*))$ gives the θ_x -scan of the coamoeba. The θ_y -scan of the coamoeba is similarly defined for a range of values for $\theta_y \in [0, 2\pi)$.

For a given choice of complex structure parameters $c_{(n_x, n_y)} \in \mathbb{C}^*$ for a Newton polynomial $P(x, y)$ of a toric Calabi-Yau 3-fold, we define the components of the coamoeba matrix \mathbf{X} as follows:

$$X_{\alpha\beta} = \begin{cases} 1 & \exists (\theta_x^*, \theta_y^*) \text{ s.t. } (\beta - 1)d_x \leq \theta_x^* < \beta d_x \text{ and } (\alpha - 1)d_y \leq \theta_y^* < \alpha d_y, \\ 0 & \text{otherwise} \end{cases}, \quad (3.2)$$

where (θ_x^*, θ_y^*) is a solution found under the θ_x - and θ_y -scans. The indices $\alpha = 1, \dots, m_y$ and $\beta = 1, \dots, m_x$ label the grid points used for the coamoeba plot, and $0 < d_x \leq 2\pi$ and $0 < d_y \leq 2\pi$ measure the grid separation, as illustrated in Fig. 7(a). The resolutions m_x and m_y for the coamoeba plot are defined by

$$d_x = \frac{2\pi}{m_x}, \quad d_y = \frac{2\pi}{m_y}. \quad (3.3)$$

Given the coamoeba matrix \mathbf{X} , we can concatenate the rows of the matrix to form a $(m_x m_y)$ -dimensional coamoeba vector as follows:

$$\mathbf{x} = (X_{11}, \dots, X_{1m_y}, X_{21}, \dots, X_{m_x,1}, \dots, X_{m_x, m_y}). \quad (3.4)$$

These coamoeba vectors live in the non-negative cone of $\mathbb{Z}^{m_x m_y}$. We note that for each choice of complex structure moduli, we have a corresponding coamoeba vector in $\mathbb{Z}^{m_x m_y}$. In the following work, we will use the coamoeba vectors as input to generate the phase diagram of coamoeba and brane tilings for a given toric Calabi-Yau 3-fold.

A. Example

Figure 6 illustrates coamoeba plots corresponding to the cone over the zeroth Hirzebruch surface F_0 , whose Newton polynomial takes the form,

$$P(x, y) = \left(x + \frac{1}{x}\right) + c_1 \left(y + \frac{1}{y}\right) + c_2, \quad (3.5)$$

where $c_1, c_2 \in \mathbb{C}^*$ are the independent complex structure moduli. We note that the complex structure moduli are nonzero such that the number of terms in the Newton

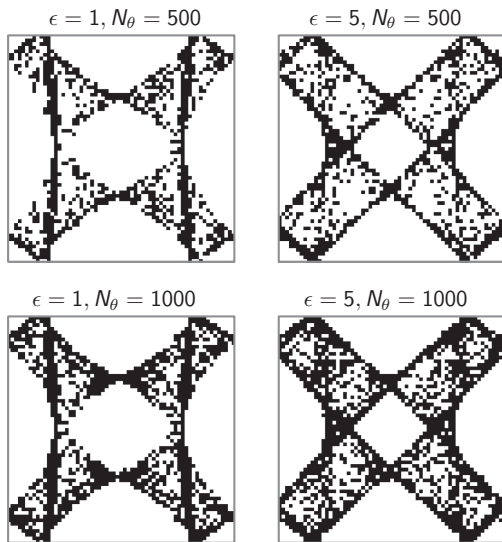


FIG. 6. Illustration of coamoeba plots corresponding to the cone over the zeroth Hirzebruch surface F_0 with the components of the complex structure moduli chosen to be $(c_{11}, c_{12}, c_{21}, c_{22}) = (-3, -6, -3, +6)$. The thickness parameter ϵ and the density parameter N_θ are varied between different plots.

polynomial is maintained and its correspondence to a specific toric Calabi-Yau 3-fold is fixed. We further note that for any toric Calabi-Yau 3-fold, with a toric diagram Δ with $|\Delta|$ vertices on \mathbb{Z}^2 , the complex coefficients in the Newton polynomial $P(x, y)$ can be rescaled such that they give precisely $|\Delta| - 3$ independent ones [27]. In the case for F_0 , this precisely gives two independent complex structure moduli chosen to be $c_1 = c_{11} + ic_{12}$ and $c_2 = c_{21} + ic_{22}$ in (3.5), where $c_{ij} \in \mathbb{R}$ such that $c_i \in \mathbb{C}^*$. Figure 6 shows the coamoeba plots for complex structure moduli set to $(c_{11}, c_{12}, c_{21}, c_{22}) = (-3, -6, -3, +6)$, with the coamoeba plot embedded in a grid defined by the resolution $m_x = m_y = 63$, giving a total of $m_x m_y = 3969$ grid points in the coamoeba plots. Accordingly, the coamoeba vectors are in the positive cone of \mathbb{Z}^{3969} . The four coamoeba plots in Fig. 6 are obtained using $\epsilon \in \{1, 5\}$ and $N_\theta = \{500, 1000\}$ and are the union of the corresponding θ_x - and θ_y -scans. We note that our use of coamoeba vectors as representations of corresponding brane tilings, $4d \mathcal{N} = 1$ theories and the choice of complex structure moduli is very much in contrast to the use of amoeba images in conjunction with supervised machine learning techniques for example in [54].

IV. DIMENSIONAL REDUCTION AND PRINCIPAL COMPONENT ANALYSIS

Given a choice of complex structure moduli for a toric Calabi-Yau 3-fold, we can obtain through the coamoeba projection the corresponding coamoeba vector defined in (3.4). The choice of complex structure moduli and the corresponding coamoeba vector are both associated to a brane tiling representing a $4d \mathcal{N} = 1$ supersymmetric gauge theory. Given that multiple choices of complex structure moduli can be associated to the same toric phase, identifying a map between toric phases and choices of complex structure moduli is a challenging problem.

In this work, we propose to dimensionally reduce the $(m_x m_y)$ -dimensional coamoeba vectors in the positive cone of $\mathbb{Z}^{m_x m_y}$ to a lower dimensional phase space, where connected regions in the phase space can be identified with toric phases of brane tilings representing $4d \mathcal{N} = 1$ supersymmetric gauge theories. We use principal component analysis (PCA) [34–37], which is one of the most fundamental and powerful techniques used for machine learning. The main idea behind PCA is to compress data by dimensional reduction with the aim of losing the least amount of information during the process. In the following section, we give a brief review of PCA applied to coamoeba vectors relating to specific choices of complex structure moduli.

A. Principal component analysis

Let us give a brief review of PCA [34–38] for coamoeba vectors associated to choices of complex structure moduli. Given a set of N coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, where for

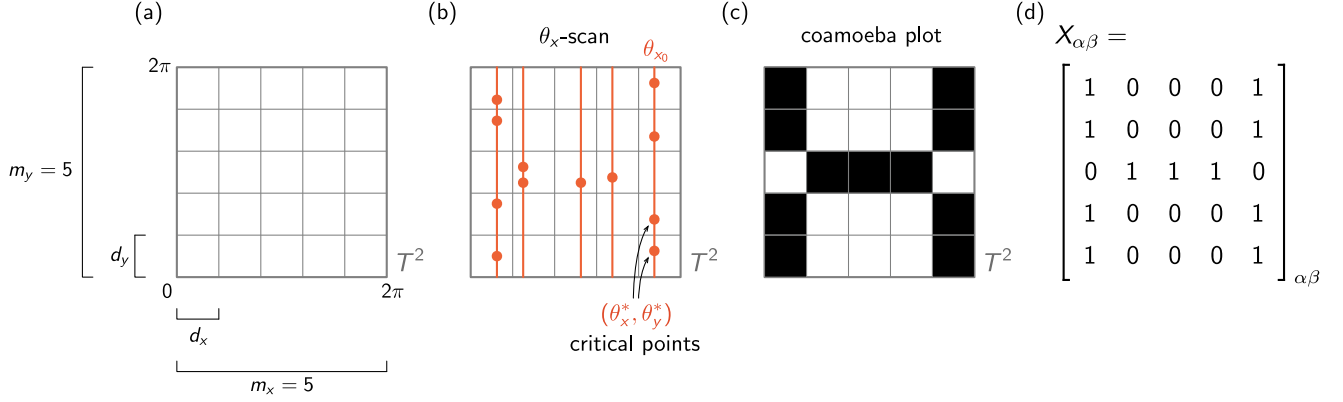


FIG. 7. A coamoeba plot consists of (a) a $m_x \times m_y$ grid with grid separation given by d_x and d_y . (b) When a critical point (θ_x^*, θ_y^*) is inside a grid point, (c) the grid point is colored black. (d) The grid can be represented in terms of a coamoeba matrix whose entries are 1 for black grid points and otherwise 0.

PCA we generalize $\mathbf{x}_a \in \mathbb{R}^d$ with $d = m_x m_y$, we can always project \mathbb{R}^d to a lower m -dimensional feature space $U = \mathbb{R}^m \subset \mathbb{R}^d$. The lower-dimensional representation of $\mathbf{x}_a \in \mathbb{R}^d$, which we call $\mathbf{z}_a = \pi_U(\mathbf{x}_a) \in U$, is given by

$$\mathbf{z}_a = \pi_U(\mathbf{x}_a) = P_\pi \mathbf{x}_a, \quad (4.1)$$

where the projection matrix from \mathbb{R}^d to U is given by

$$P_\pi = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top. \quad (4.2)$$

Here,

$$\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{d \times m}, \quad (4.3)$$

is the matrix of basis vectors \mathbf{b}_i for U . The coordinate vector based on the basis given by \mathbf{B} is $(z_{a1}, \dots, z_{am}) = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}_a$. Note that if the basis of U is orthonormal, then $\mathbf{B}^\top \mathbf{B} = I$ gives the identity matrix, leading to

$$\mathbf{z}_a = \pi_U(\mathbf{x}_a) = \mathbf{B} \mathbf{B}^\top \mathbf{x}_a, \quad (4.4)$$

where

$$(z_{a1}, \dots, z_{am}) = \mathbf{B}^\top \mathbf{x}_a, \quad (4.5)$$

are the coordinates of \mathbf{z}_a with respect to the orthonormal basis given by \mathbf{B} .

So far, we see that there are no surprises and using P_π defined in terms of \mathbf{B} , we can project any coamoeba vector $\mathbf{x}_a \in \mathbb{R}^d$ into the subspace U . This is however now the point where beautifully linear algebra meets statistics. What we mean is, we do not want to project $\mathbf{x}_a \in \mathbb{R}^d$ to any subspace U of \mathbb{R}^d , but to a special subspace \hat{U} where statistically the maximum amount of information in the original set of

coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$ is preserved. This, in other words, means we need to find \hat{U} with basis vectors $\mathbf{b}_1, \dots, \mathbf{b}_m$ such that when $\mathbf{x}_a \in \mathbb{R}^d$ are projected to \hat{U} , they have a maximized variance along $\mathbf{b}_1, \dots, \mathbf{b}_m$.

Let us consider the first basis vector \mathbf{b}_1 of \hat{U} , that maximizes the variation of the original set of coamoeba vectors under the projection. This involves the maximization of the variance of the first coordinate z_{1a} of \mathbf{z}_a over the whole set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. Given that the basis of \hat{U} is orthonormal, we can express the first coordinate as

$$z_{1a} = \mathbf{b}_1^\top \mathbf{x}_a. \quad (4.6)$$

The variance for the first coordinate z_{1a} over $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ is then given by

$$V(z_1) = \frac{1}{N} \sum_{a=1}^N z_{1a}^2 = \frac{1}{N} \sum_{a=1}^N (\mathbf{b}_1^\top \mathbf{x}_a)^2, \quad (4.7)$$

where we note that the expression for the variance is independent of the mean of the first coordinate over the whole set $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$.

Here, let us introduce the coamoeba covariance matrix \mathbf{S} for the original set of coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, which is defined as follows:

$$\mathbf{S} = \frac{1}{N} \sum_{a=1}^N \mathbf{x}_a \mathbf{x}_a^\top. \quad (4.8)$$

The covariance matrix is symmetric and positive semi-definite and is a measure of how much the original set of coamoeba vectors is spread. The expression of the variance $V(z_1)$ of the first coordinate in (4.7) can be expressed in terms of the covariance matrix \mathbf{S} as follows:

$$V(z_1) = \mathbf{b}_1^\top \mathbf{S} \mathbf{b}_1. \quad (4.9)$$

The question now is how to maximize the variance $V(z_1)$. One option is to increase the length of the basis vector \mathbf{b}_1 . This is however not what we want given the fact that we started with the condition that \hat{U} has an orthonormal basis. Accordingly, we introduce the constraint that the norm of the basis vector satisfies $\|\mathbf{b}_1\|^2 = 1$. The result is a constrained optimization problem taking the form,

$$\begin{aligned} \max_{\mathbf{b}_1} (V(z_1)) &= \max_{\mathbf{b}_1} (\mathbf{b}_1^\top \mathbf{S} \mathbf{b}_1), \\ \|\mathbf{b}_1\|^2 &= \mathbf{b}_1^\top \mathbf{b}_1 = 1. \end{aligned} \quad (4.10)$$

In order to solve this constrained optimization problem, we write a Lagrangian function of the form

$$\mathcal{L}(\mathbf{b}_1, \lambda) = \mathbf{b}_1^\top \mathbf{S} \mathbf{b}_1 + \lambda(1 - \mathbf{b}_1^\top \mathbf{b}_1), \quad (4.11)$$

where λ is the Lagrange multiplier. By taking the partial derivatives of $\mathcal{L}(\mathbf{b}_1, \lambda)$ to zero, we obtain

$$\mathbf{S} \mathbf{b}_1 = \lambda \mathbf{b}_1, \quad (4.12)$$

and $\mathbf{b}_1^\top \mathbf{b}_1 = 1$. Here, (4.12) beautifully refers to an eigenvalue equation, where the first basis vector \mathbf{b}_1 is the eigenvector with eigenvalue λ for the coamoeba covariance matrix \mathbf{S} .

By inserting the eigenvalue equation in (4.12) with $\mathbf{b}_1^\top \mathbf{b}_1 = 1$ into the formula for the variance in (4.9), we get

$$V(z_1) = \mathbf{b}_1^\top \mathbf{S} \mathbf{b}_1 = \lambda. \quad (4.13)$$

This implies that in order to maximize the variance of the projected set of coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, we have to choose for \mathbf{b}_1 an eigenvector of the covariance matrix \mathbf{S} that has the largest eigenvalue $\lambda = \lambda_1$. This eigenvector \mathbf{b}_1 is also known as the first principal component. Using this principal component, we can then identify the optimal projected coamoeba vector using (4.6) inside the original feature space \mathbb{R}^d ,

$$\hat{\mathbf{x}}_a = \mathbf{b}_1 z_{1a} \in \mathbb{R}^d. \quad (4.14)$$

We can now extend the problem to the m -th principal component. Let us first assume that we already found the first $m - 1$ principal components $\mathbf{b}_1, \dots, \mathbf{b}_{m-1}$, which correspond to the first $m - 1$ eigenvectors of the coamoeba covariance matrix \mathbf{S} with the eigenvalues sorted from large to small as follows: $\lambda_1, \dots, \lambda_{m-1}$.

For the m th principal component, we then have to maximize the variance $V(z_m)$ for the m th component of the set of coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, under the constraint that we have orthonormal basis vectors satisfying $\mathbf{b}_m^\top \mathbf{b}_m = 1$. By solving the constrained optimization problem, we obtain the following eigenvalue equation:

$$\mathbf{S} \mathbf{b}_m = \lambda_m \mathbf{b}_m, \quad (4.15)$$

where the variance of the coamoeba vectors projected onto the m th principal component is given by the eigenvalue $V(z_m) = \lambda_m$.

Overall, in order to find the most optimal m -dimensional subspace \hat{U} of the original space \mathbb{R}^d of coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, we have to choose as basis vectors for \hat{U} eigenvectors of the covariance matrix \mathbf{S} that have the largest eigenvalues. The maximum variance from the first m principal components is given by

$$V(z_1, \dots, z_m) = \sum_{i=1}^m \lambda_i. \quad (4.16)$$

B. Unsupervised optimization

The overall optimization problem of identifying the first m principal components that maximize $V(z_1, \dots, z_m)$ can be reformulated as a loss function $J(\mathbf{b}_a)$ that needs to be minimized. Such a loss function measures the overall difference between the actual coamoeba vectors \mathbf{x}_a and the projected coamoeba vectors $\hat{\mathbf{x}}_a$ expressed in terms of \mathbb{R}^d coordinates. Following the expression for $\mathbf{z}_a \in \hat{U}$ in (4.1), we can express

$$\hat{\mathbf{x}}_a = (z_{1a}, \dots, z_{ma}, 0, \dots, 0) \in \mathbb{R}^d, \quad (4.17)$$

where (z_{1a}, \dots, z_{ma}) are the coordinates for the projected coamoeba vector \mathbf{z}_a in $\hat{U} \subset \mathbb{R}^d$. Accordingly, we can write the loss function for the original set of coamoeba vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ under PCA as

$$\begin{aligned} J(\mathbf{b}_a) &= \frac{1}{N} \sum_{a=1}^N \|\mathbf{x}_a - \hat{\mathbf{x}}_a\|^2 \\ &= \sum_{k=m+1}^d \mathbf{b}_k^\top \mathbf{S} \mathbf{b}_k = \sum_{k=m+1}^d \lambda_k. \end{aligned} \quad (4.18)$$

We can see from here that in order to minimize the loss function above, we have to identify the smallest $d - m$ eigenvalues of the coamoeba covariance matrix \mathbf{S} whose corresponding eigenvectors are orthogonal to the m principal components that span \hat{U} .

In this work, we interpret \hat{U} as a phase space for coamoeba corresponding to brane tilings and $4d$ $\mathcal{N} = 1$ supersymmetric gauge theories associated to a toric Calabi-Yau 3-fold. Every projected vector \mathbf{z}_a in \hat{U} is a projected coamoeba vector with an associated choice of complex structure moduli. If the dimension m of \hat{U} is kept small, we expect \mathbf{z}_a to cluster into connected regions in \hat{U} corresponding to coamoeba and brane tilings that form a toric phase associated to the toric Calabi-Yau 3-fold. In the

following section, we obtain \hat{U} for the Calabi-Yau cone over the zeroth Hirzebruch surface F_0 with $m = 2$. We illustrate with this example that indeed \hat{U} can be interpreted as a phase space for toric phases of F_0 .

V. A PHASE DIAGRAM FOR COAMOEBEA AND BRANE TILINGS

In this section, we overview the explicit construction of the phase diagram that we propose using PCA for coamoeba and brane tilings corresponding to the Calabi-Yau cone over the zeroth Hirzebruch surface F_0 [7,61–63].

A. Example

Let us consider the brane tilings and coamoeba for the cone over the zeroth Hirzebruch surface F_0 [7,61–63], whose Newton polynomial with complex structure moduli c_1 and c_2 is given in (3.5). Restricting ourselves to a discrete finite subset of all possible choices for the complex structure moduli, we choose the real moduli c_{ij} to take values only from $\{-9, -6, -3, 0, 3, 6, 9\}$ such that $c_i \in \mathbb{C}^*$. Under this restriction on the choices of complex structure moduli, we generate $N = 2304$ coamoeba plots with parameters $\epsilon = 5$, $N_\theta = 2000$, and $m_x = m_y = 63$. These give $N = 2304$ coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$ with $d = 3969$. We note that these hyperparameters are chosen specifically for the discussion in this work on the phase space for brane tilings, and a more detailed discussion on the role played by these hyperparameters is the subject of future work.

We use PCA to project the coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$ down to \hat{U} with $m = 2$ principal components \mathbf{b}_1 and \mathbf{b}_2 . These are the eigenvectors of the coamoeba covariance matrix \mathbf{S} defined in (4.8), whose respective eigenvalues λ_1 and λ_2 are the largest and second largest out of all eigenvalues of \mathbf{S} . This means that along $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^d$, the set of original coamoeba vectors \mathbf{x}_a when projected down to $\mathbf{z}_a = (z_{1a}, z_{2a}) \in \hat{U} \subset \mathbb{R}^d$, have the largest and second largest maximized variances $V(z_1) = \lambda_1$ and $V(z_2) = \lambda_2$ over all $a = 1, \dots, N$. The eigenvalues for our dataset of coamoeba vectors for F_0 take the following values,

$$\lambda_1 = 144.23, \quad \lambda_2 = 57.79, \quad (5.1)$$

where we express measurements up to 2 decimal points. The so-called proportional variance given by the i th eigenvalue of \mathbf{S} can be calculated using

$$\lambda_i^* = \frac{\lambda_i}{\sum_{j=1}^d \lambda_j} \times 100\%, \quad (5.2)$$

where the total variance $\sum_{j=1}^d \lambda_j$ is used in the denominator. For our analysis, the proportional variance for the first $m = 2$ principal components is given by

$$\lambda_1^* = 17.18\%, \quad \lambda_2^* = 6.88\%. \quad (5.3)$$

This means about $\lambda_1^* + \lambda_2^* = 24.06\%$ of the information in the original set of coamoeba vectors \mathbf{x}_a is preserved in total if all the vectors are projected to $\mathbf{z}_a = (z_{1a}, z_{2a}) \in \hat{U}$.

When we plot the coordinates z_1, z_2 corresponding to the principal components \mathbf{b}_1 (PC1) and \mathbf{b}_2 (PC2), respectively, for all coamoeba vectors, we obtain a 2-dimensional PCA score plot as shown in Fig. 8(a). We interpret this plot as a 2-dimensional phase diagram for brane tilings and coamoeba for F_0 . This is because the projected \mathbf{z}_a , each corresponding to a coamoeba defined for a particular choice of complex structure moduli whose skeleton graph is a brane tiling realizing a $4d \mathcal{N} = 1$ supersymmetric gauge theory, cluster in the 2-dimensional plot as shown in Fig. 8(a).

A closer look at the first principal component \mathbf{b}_1 (PC1) and the corresponding coordinate z_1 reveals that when plotted as a frequency histogram with bin size $\Delta z_1 = 1.25$ and with a total of $N_{\text{bins}} = 30$ bins along the z_1 -axis, there are 3 peaks in the frequency count $N(\text{bin}_h)$ of coamoeba vectors for a given bin_h along the z_1 -axis as shown in Fig. 8(b). These peaks in the frequency count along the z_1 -axis indicate clusters of projected coamoeba vectors \mathbf{z}_a in the 2-dimensional phase diagram. We call this frequency histogram along the z_1 -axis corresponding to the first principal component the $m = 1$ phase histogram for F_0 .

The locations of these clusters along the z_1 -axis are given by

$$z_1^{(1)} = 0, \quad z_1^{(2a)} = -15, \quad z_1^{(2b)} = 15, \quad (5.4)$$

where we call the clusters respectively as (1), (2a), and (2b). We note that the $m = 2$ phase diagram appears to be symmetric around $z_1 = 0$, mapping clusters (2a) and (2b) into each other.

When we take samples of projected vectors \mathbf{z}_a from the phase diagram in Fig. 8(a), the corresponding coamoeba and their brane tilings, as shown in Figs. 8(c)–8(g), contain the two known toric phases for the cone over the zeroth Hirzebruch surface F_0 . The first toric phase [8,63], whose brane tiling, quiver diagram and superpotential of the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theory are shown in Fig. 4, can be associated to cluster (1) in the phase diagram in Fig. 8(a). The second toric phase [7,29], which is Seiberg dual to the first toric phase, is identified with clusters (2a) and (2b), where as shown in Fig. 8(a) projected coamoeba vectors \mathbf{z}_a in clusters (2a) and (2b) correspond to equivalent brane tilings on the 2-torus giving the same $4d \mathcal{N} = 1$ supersymmetric gauge theory.

We can further highlight the existence of these clusters by having a closer look at the eigenvectors \mathbf{b}_1 and \mathbf{b}_2 of the coamoeba covariance matrix \mathbf{S} . We recall from (4.3) that these eigenvectors representing the two principal components form the matrix of basis vectors \mathbf{B} of \hat{U} .

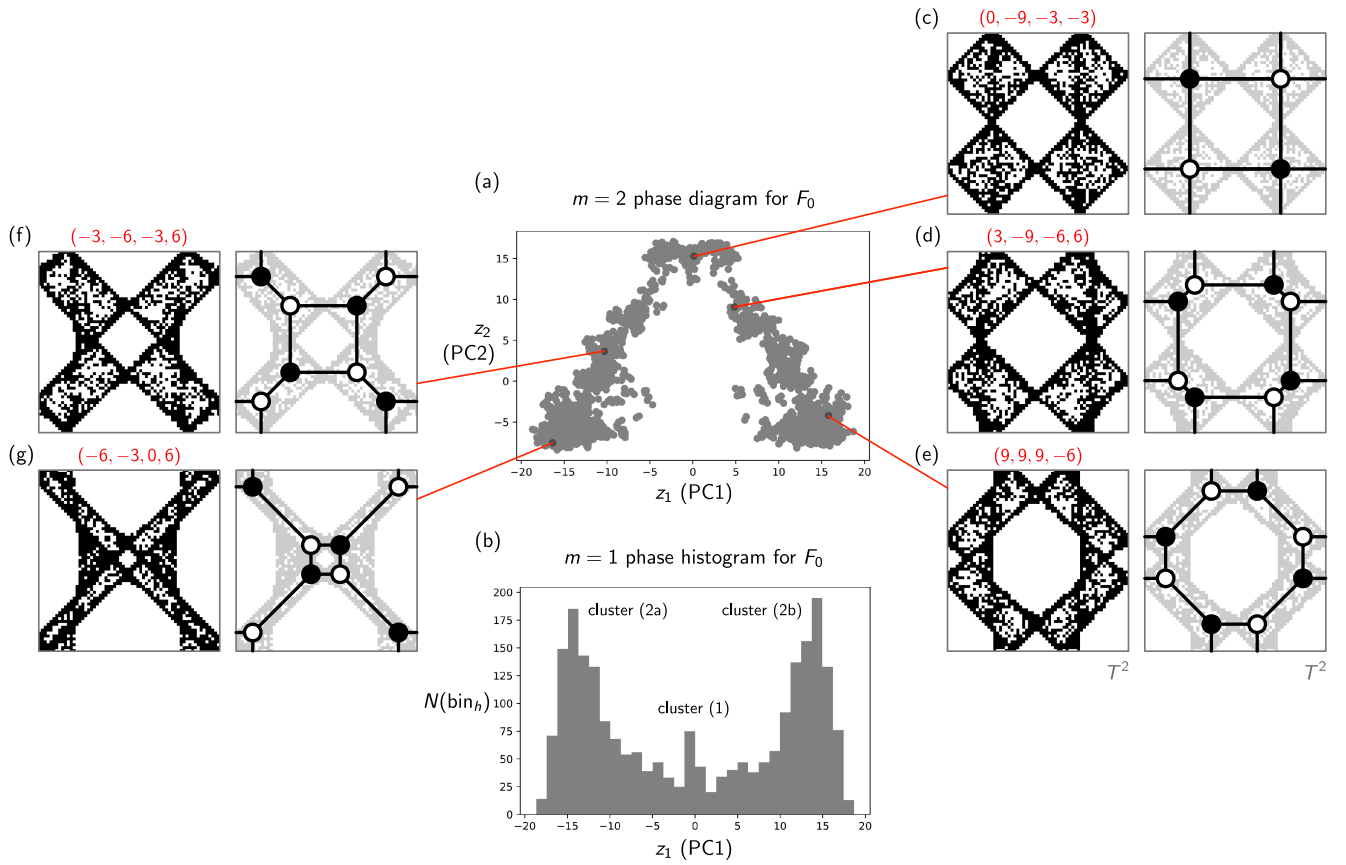


FIG. 8. (a) The $m = 2$ phase diagram for F_0 consists of points that correspond to coamoeba plots and their corresponding brane tilings for specific choices of the complex structure moduli. (b) We call the corresponding frequency histogram along the z_1 -axis corresponding to the first principal component the $m = 1$ phase histogram for F_0 . It shows three peaks corresponding to three clusters in the $m = 2$ phase diagram for F_0 . We identify these clusters with three phases of coamoeba plots and their corresponding brane tilings, called phases (1), (2a), and (2b). In (c)–(g), we select five points in the $m = 2$ phase diagram and illustrate the corresponding coamoeba plots, the associated brane tilings and the choices of complex structure moduli.

Following (4.5), given that $\mathbf{B}^\top \mathbf{B} = I$ gives the identity matrix, \mathbf{B}^\top projects the original coamoeba vectors \mathbf{x}_a to the 2 coordinates z_{1a} and z_{2a} . An alternative interpretation of the eigenvectors \mathbf{b}_1 and \mathbf{b}_2 and the matrix \mathbf{B} is that the absolute values of their components measure how correlated the $d = 3969$ components of the original coamoeba vector \mathbf{x}_a are with the $m = 2$ components of the projected vectors with coordinates (z_{1a}, z_{2a}) on the $m = 2$ phase diagram.

We can in fact plot the values of the $d = 3969$ components of the eigenvectors \mathbf{b}_1 and \mathbf{b}_2 on the unit cell of T^2 , like the coamoeba plots obtained from the coamoeba matrix defined in (3.2). The resulting heatmaps for \mathbf{b}_1 and \mathbf{b}_2 are shown in Fig. 9. Quite beautifully we can see that the large positive values (in yellow) for the components of \mathbf{b}_1 in Fig. 9(a) form the shape of a coamoeba corresponding to cluster (2b) in the $m = 2$ phase diagram in Fig. 8(a). Furthermore, the largely negative values (in dark blue) for the components of \mathbf{b}_1 form the shape of a coamoeba corresponding to cluster (2a) in the phase diagram in Fig. 8(a). This means, the first principal component \mathbf{b}_1

mainly detects coamoeba corresponding to the second toric phase of F_0 and the sign of the components of \mathbf{b}_1 mainly distinguish between clusters (2a) and (2b) in the phase diagram in Fig. 8(a). Moreover, looking at the heatmap for the second principal component \mathbf{b}_2 in Fig. 9(b) reveals that large positive values (in yellow) of the components of \mathbf{b}_2 form the shape of a coamoeba corresponding to cluster (1) in the phase diagram in Fig. 8(a). This indicates that the second principal component \mathbf{b}_2 mainly identifies the first toric phase of F_0 . Accordingly, the two principal components \mathbf{b}_1 and \mathbf{b}_2 play a vital role in distinguishing coamoeba corresponding to the two toric phases of F_0 .

In our discussion, we have identified in the $m = 2$ phase diagram for F_0 clusters of projected coamoeba vectors \mathbf{z}_a by identifying peaks in the frequency histogram of vectors along the z_1 -axis of the phase diagram. In the following section, we employ further unsupervised machine learning techniques in order to explicitly identify the clusters and discrete boundaries between them. These boundaries can be interpreted as phase boundaries between toric phases of F_0 where Seiberg duality occurs.

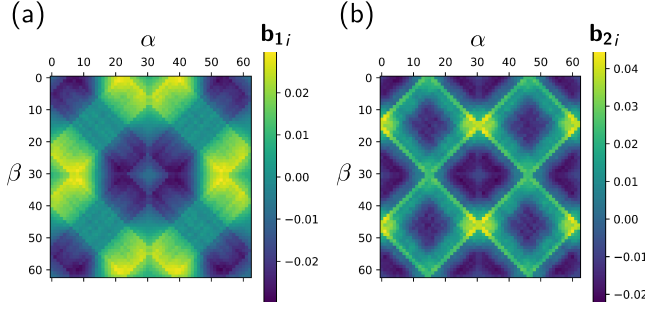


FIG. 9. The plots show the values of the $d = 3969$ components of the eigenvectors (a) \mathbf{b}_1 and (b) \mathbf{b}_2 as heatmaps on the unit cell of T^2 . In (a), they form the shape of coamoeba corresponding to phase (2a) and (2b), whereas in (b), the heatmap forms the shape of the coamoeba for phase (1). The indices $\alpha = 1, \dots, m_y$ and $\beta = 1, \dots, m_x$ label the grid points when the eigenvectors are illustrated in a $m_x \times m_y$ grid, analogous to coamoeba plots.

VI. CLUSTERS AND TORIC PHASE BOUNDARIES

In order to identify the explicit borders between toric phases in the 2-dimensional phase diagram in Fig. 8(a), we make use of a manifold learning algorithm for dimensionally reducing multidimensional data known as t -distributed stochastic neighbor embedding (t -SNE) [39,40]. Let us give a brief overview of t -SNE in terms of coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$ with $a = 1, \dots, N$.

A. t -distributed stochastic neighbor embedding

Let us first define the pairwise affinity $p_{b|a}$, which measures how similar two coamoeba vectors in the original d -dimensional space \mathbb{R}^d are,

$$p_{b|a} = \frac{\exp(-\|\mathbf{x}_a - \mathbf{x}_b\|^2/2\sigma_a^2)}{\sum_{h \neq a} \exp(-\|\mathbf{x}_a - \mathbf{x}_h\|^2/2\sigma_a^2)}. \quad (6.1)$$

The pairwise affinity $p_{b|a}$ essentially measures the likelihood that one coamoeba vector \mathbf{x}_a would pick \mathbf{x}_b as its neighbor under the assumption that neighbors are picked following a Gaussian probability distribution centered at \mathbf{x}_a . Here, the Gaussian is chosen to have a variance of σ_a . The pairwise affinity can be symmetrized as follows:

$$p_{ab} = \frac{p_{b|a} + p_{a|b}}{2N}, \quad (6.2)$$

where N is the number of coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$.

We note that the individual variances σ_a for each coamoeba vector \mathbf{x}_a are determined using binary search. The search is dependent on the value of the perplexity $Perp$ [39,40], which is defined as

$$Perp(P_a) = 2^{H(P_a)}, \quad (6.3)$$

where P_a is the probability distribution formed by the conditional probabilities given by $p_{b|a}$ defined in (6.1). $H(P_a)$ is the Shannon entropy of P_a given by

$$H(P_a) = - \sum_{b=1}^N p_{b|a} \log_2 p_{b|a}. \quad (6.4)$$

Overall, the value of the perplexity $Perp$ has the effect of emphasizing local and global features of the dimensional reduction of the coamoeba vectors to \mathbf{z}_a in m dimensions.

We can define a pairwise affinity between the dimensionally reduced coamoeba vectors \mathbf{z}_a as follows:

$$q_{ab} = \frac{(1 + \|\mathbf{z}_a - \mathbf{z}_b\|^2)^{-1}}{\sum_{h \neq g}^N (1 + \|\mathbf{z}_h - \mathbf{z}_g\|^2)^{-1}}, \quad (6.5)$$

where here we use the t -distribution with $\nu = 1$ degree of freedom. This gives us the Cauchy distribution which replaces the Gaussian distribution that is used in (6.1). Here, the pairwise affinity between the reduced coamoeba vectors \mathbf{z}_a is a measure of similarity in the dimensionally reduced space. The t -distribution is used for \mathbf{z}_a rather than the Gaussian distribution, because the ‘‘heavy’’ tails of the distribution tend to make clusters of reduced coamoeba vectors \mathbf{z}_a tighter than in other dimensional reduction techniques.

Given the probability distribution P_a defined by the pairwise affinities p_{ab} of coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$, and the probability distribution Q_a defined by the pairwise affinities q_{ab} of dimensionally reduced coamoeba vectors $\mathbf{z}_a \in \mathbb{R}^m$, we can measure the overall divergence of these two probability distributions using the Kullback-Leibler (KL) divergence [65] defined as

$$C_{\text{KL}} = \sum_{a,b=1}^N p_{ab} \log \frac{p_{ab}}{q_{ab}}. \quad (6.6)$$

The dimensional reduction of coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^d$ to $\mathbf{z}_a \in \mathbb{R}^m$ using t -SNE has the aim of reducing the KL divergence C by iteratively varying the positions of candidate coamoeba vectors \mathbf{z}_a in \mathbb{R}^m . This minimization process is done by gradient descent, where the gradient is defined as

$$\frac{\delta C_{\text{KL}}}{\delta \mathbf{z}_a} = 4 \sum_{b=1}^N \frac{(p_{ab} - q_{ab})(\mathbf{z}_a - \mathbf{z}_b)}{1 + \|\mathbf{z}_a - \mathbf{z}_b\|^2}. \quad (6.7)$$

Having reviewed t -SNE for coamoeba vectors, let us now apply t -SNE to the $N = 2304$ coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^{3969}$ that we have obtained in Sec. III for the Calabi-Yau cone over the zeroth Hirzebruch surface F_0 .

B. Example

Let us consider the $N = 2304$ coamoeba vectors $\mathbf{x}_a \in \mathbb{R}^{3969}$ obtained for the cone over the zeroth Hirzebruch surface F_0 from Sec. III. Using t -SNE, the coamoeba vectors can be dimensionally reduced to $\mathbf{z}_a \in \mathbb{R}^m$ with $m = 2$, giving the t -SNE phase diagram for F_0 shown in Fig. 10(a) with the perplexity parameter set to $Perp = 250$. In comparison to the phase diagram obtained from PCA, the t -SNE phase diagram contains three disconnected clusters S_p , which can be identified with phases $p = (1)$, (2a), and (2b), where (2a) and (2b) correspond to the same second toric phase of F_0 .

We can identify through the t -SNE phase diagram the coamoeba and the corresponding complex structure moduli corresponding to each of the three clusters S_p in Fig. 10(a), giving the following counting of coamoeba,

$$|S_{(2a)}| = 1004, \quad |S_{(1)}| = 297, \quad |S_{(2b)}| = 1003, \quad (6.8)$$

where $N = |S_{(2a)}| + |S_{(1)}| + |S_{(2b)}| = 2304$.

We define the average projected coamoeba vector for a subset S_p of vectors in a given phase p as

$$\langle \mathbf{z}_a \rangle_p = \frac{1}{|S_p|} \sum_{\mathbf{z}_a \in S_p} \mathbf{z}_a. \quad (6.9)$$

Recalling that the concatenated rows of the coamoeba matrix give the coamoeba vector, we can illustrate the averaged coamoeba plots from the average coamoeba vectors $\langle \mathbf{z}_a \rangle_p$ for the three phases S_p identified in the t -SNE phase diagram. These averaged coamoeba plots for each of the phases with the corresponding brane tilings are shown in Fig. 10(c). We note that the brane tiling for the averaged coamoeba plot for phase 1 corresponds exactly to the known first toric phase of F_0 . Moreover, the brane tilings for the averaged coamoeba plots for phases 2a and 2b corresponds exactly to the known second toric phase of F_0 .

We can identify individual coamoeba vectors \mathbf{z}_a in clusters $S_{(1)}$, $S_{(2a)}$, and $S_{(2b)}$ of the $m = 2$ t -SNE phase diagram in the $m = 2$ phase diagram obtained through PCA. This allows us to identify the phase boundaries between different phases in the $m = 2$ PCA phase diagram as illustrated in Fig. 10(b). We note that these phase boundaries in the $m = 2$ phase diagram exactly correspond to where Seiberg duality occurs between toric phases.

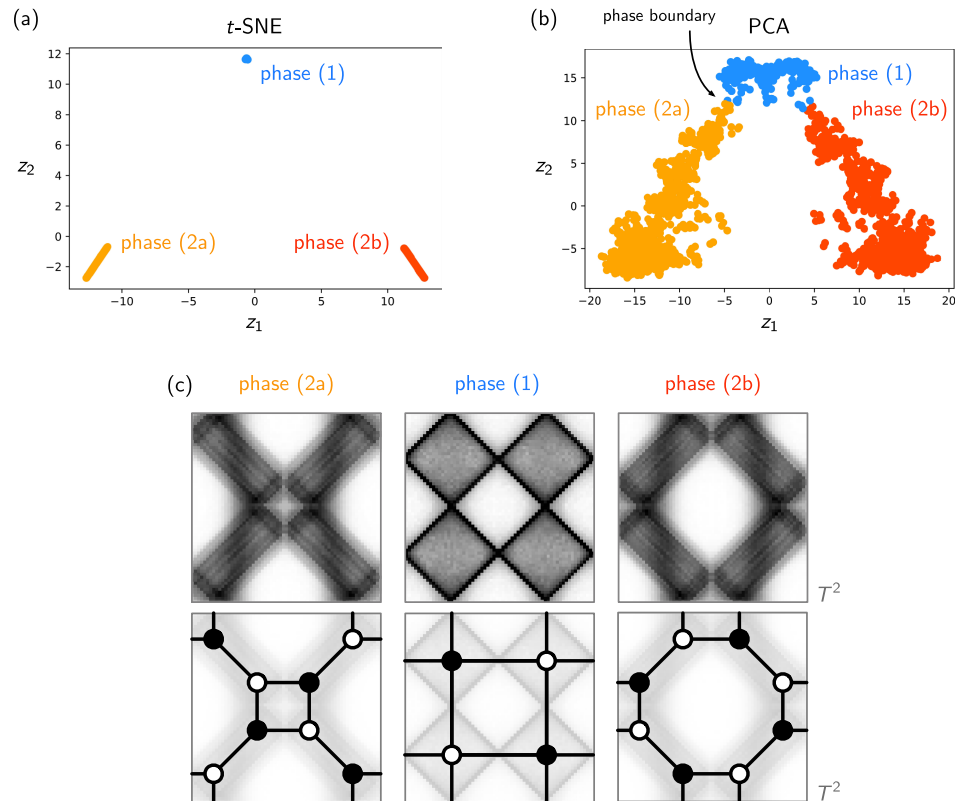


FIG. 10. (a) The $m = 2$ t -SNE phase diagram for F_0 shows three clusters S_p with sizes $|S_{(2a)}| = 1004$, $|S_{(1)}| = 297$, $|S_{(2b)}| = 1003$. The points corresponding to coamoeba for specific choices of complex structure moduli can be identified with points in the (b) $m = 2$ PCA phase diagram for F_0 . Given that through the $m = 2$ t -SNE phase diagram we know which points correspond to which cluster and phase, we can identify the phase boundaries in the $m = 2$ PCA phase diagram for F_0 . In (c), we show the averaged coamoeba plots and the corresponding brane tilings for phases (1), (2a), and (2b).

We further note that our approach is in direct contrast to the attempts of using supervised machine learning for studying Seiberg duality in [66,67].

An interesting question one can ask is if one can predict now whether a certain choice of complex structure moduli in (3.5) leads to a specific toric phase of F_0 . We note that the clusters $S_{(1)}$, $S_{(2a)}$, and $S_{(2b)}$ obtained through the t -SNE phase diagram in Fig. 10(a) play an essential role in answering this question. In order to solve this problem, we propose the use of logistic regression [68,69], where the input parameters are chosen to be the components of the complex structure moduli c_{11} , c_{12} , c_{21} , and c_{22} for F_0 . We recall that for the set of $N = 2304$ coamoeba vectors, we chose values for $c_{ij} \in \{-9, -6, -3, 0, 3, 6, 9\}$ such that $c_i \in \mathbb{C}^*$ in Sec. III. For the purpose of using logistic regression, each component c_{ij} is scaled as follows:

$$c'_{ij} = \frac{c_{ij} - \mu(c_{ij})}{\sigma(c_{ij})}, \quad (6.10)$$

where the mean $\mu(c_{ij}) = 0$ and standard deviation $\sigma(c_{ij}) = 6.06$ for all c_{ij} over the $N = 2304$ choices of complex structure moduli.

In logistic regression, the aim is to obtain the probability that a given instance, in our case a choice of complex structure moduli given by c_{ij} , belongs to a particular class, in our case a particular subset of projected coamoeba vectors S_p . The probability is given by

$$P_p(c_{ij}) = \frac{1}{1 + e^{-l_p(c_{ij})}}, \quad (6.11)$$

where we choose $P_p(c_{ij}) = 1$ if the corresponding projected coamoeba vector $\mathbf{z}_a \in S_p$, and $P_p(c_{ij}) = 0$ if $\mathbf{z}_a \notin S_p$. Here, logit—the log odds of \mathbf{z}_a belonging to S_p with complex structure moduli components c_{ij} —is given by a linear combination of the complex structure moduli components,

$$l_p(c_{ij}) = \beta_0^p + \beta_1^p c_{11} + \beta_2^p c_{12} + \beta_3^p c_{21} + \beta_4^p c_{22}, \quad (6.12)$$

where $\beta_0^p \in \mathbb{R}$ is the intercept and all other β_i^p are real coefficients specific for a projected coamoeba vector $\mathbf{z}_a \in S_p$. In order to train the model, we randomly split the set of $N = 2304$ choices of complex structure moduli that we use to generate the coamoeba vectors in Sec. III into a train (80%) and test (20%) set. With an average test accuracy of 99.996%, we obtain the values for the intercept β_0 and coefficients β_i for each of the projected coamoeba sets S_p as summarized in Table II.

We note that the absolute magnitude $|\beta_i^p|$ of the trained coefficients for a given set S_p tells us about the strength of the relationship between the components of the complex structure moduli c_{ij} in the logit function in (6.12) and the

TABLE II. Logistic regression result in terms of the real coefficients β_i^p for the log odds of \mathbf{z}_a belonging to cluster S_p with complex structure moduli components c_{ij} . The training was done on a random train (80%) and test (20%) set split of the $N = 2304$ choices of complex structure moduli with an average test accuracy of 99.996%. We can see from the coefficients β_i^p that complex structure moduli component c_{11} corresponding to β_1^p plays the most significant role in determining the phase p for \mathbf{z}_a .

p	$ S_p $	β_0^p	β_1^p	β_2^p	β_3^p	β_4^p
(2a)	1004	-1.63	-17.51	+0.27	-0.03	+0.00
(1)	297	+3.20	+0.00	+0.00	+0.22	-0.03
(2b)	1003	-1.58	+17.71	-0.07	+0.00	+0.02

probability that the choice of complex structure moduli components c_{ij} corresponds to a projected coamoeba vector \mathbf{z}_a belonging to S_p of phase p . From the results in Table II, we see that for phases (2a) and (2b) the complex structure moduli component c_{11} with corresponding coefficient β_1^p by far plays the greatest role in determining whether the projected coamoeba vector \mathbf{z}_a corresponding to c_{ij} is in S_p . We recall that this is determined with an average test accuracy of 99.996%. A closer look reveals that

$$\mathbf{z}_a \in \begin{cases} S_{(1)} & \text{if } c_{11} = 0 \\ S_{(2a)} & \text{if } c_{11} < 0 \\ S_{(2b)} & \text{if } c_{11} > 0 \end{cases} \quad (6.13)$$

for any values of $c_{12}, c_{21}, c_{22} \in \mathbb{R}$ such that $c_i \in \mathbb{C}^*$.

In summary, we identified using t -SNE the clusters of projected coamoeba vectors $S_{(2a)}$, $S_{(1)}$, and $S_{(2b)}$ corresponding to the toric phases of F_0 . We then provided evidence through logistic regression that just the sign of the component c_{11} of the complex structure moduli for F_0 determines the phase of the corresponding coamoeba and brane tiling. This is a remarkable result given that we started with a collection of coamoeba generated for an arbitrary range of complex structure moduli centered at 0 with no knowledge about which coamoeba and choice of complex structure moduli would correspond to which toric phase of F_0 .

VII. CONCLUSIONS AND DISCUSSIONS

We made use of unsupervised machine learning techniques in order to find a phase diagram for the zeroth Hirzebruch surface F_0 . Such a phase diagram is useful because it identifies different toric phases of $4d \mathcal{N} = 1$ supersymmetric gauge theories related by Seiberg duality in terms of clusters in the phase diagram. These clusters are made up of projected coamoeba vectors \mathbf{z}_a which each refer to a coamoeba plot that corresponds to a brane tiling and a $4d \mathcal{N} = 1$ supersymmetric gauge theory. The coamoeba

plots are also associated to a specific choice of complex structure moduli in the mirror description of the toric Calabi-Yau 3-fold. The phase diagram for F_0 was obtained by using PCA and t -SNE that projected the original coamoeba vectors in $d = 3969$ dimensions to two principal components corresponding to the axes of the F_0 phase diagram. We even showed that t -SNE allows us to identify phase boundaries in the 2-dimensional phase diagram of F_0 , where phase boundaries correspond to Seiberg duality between the $4d \mathcal{N} = 1$ supersymmetric gauge theories associated to the toric phases meeting at the phase boundary. By the use of logistic regression, we gave evidence that amongst the complex structure moduli, only the value of a single real component of the moduli determines whether the corresponding coamoeba and brane tiling are in one of the two toric phases of F_0 .

We expect that a similar construction of phase diagrams is possible for different toric Calabi-Yau 3-folds that exhibit a variety of toric phases in the context of the corresponding $4d \mathcal{N} = 1$ supersymmetric gauge theories [70,71]. Moreover, we plan to report on phase diagrams constructed using the proposed unsupervised machine learning techniques for toric Calabi-Yau in higher dimensions, such as for toric Calabi-Yau 4-folds probed by D1-branes whose worldvolume theories are a class of $2d (0,2)$ supersymmetric gauge theories realized in terms of brane brick models [27,72,73]. These $2d (0, 2)$ supersymmetric gauge theories related to toric Calabi-Yau 4-folds have been found to exhibit Gaiotto-Gukov-Putrov triality [74,75] and are expected to have more elaborate phase diagrams that exhibit many interesting structures in the phase space of $2d (0, 2)$ theories.

The use of unsupervised machine learning techniques in this work in order to study the phase structure of supersymmetric gauge theories related to toric Calabi-Yau opens up a new avenue of research at the interface of supersymmetric gauge theories realized in string theory, Calabi-Yau mirror symmetry and tropical geometry, and explainable AI and unsupervised machine learning on which we hope to report on more in the near future.

ACKNOWLEDGMENTS

R.-K. S. is grateful to Per Berglund, Sebastian Franco, Sergei Gukov, Amihay Hanany, Yang-Hui He, and Cumrun Vafa for discussions and feedback about this work. He would like to thank the Simons Center for Geometry and Physics at Stony Brook University for hospitality during various stages of this work. He would also like to thank the organizers of the “Dimers: Combinatorics, Representation Theory and Physics” workshop for their hospitality at the CUNY Graduate Center in New York. He is supported by a Basic Research Grant of the National Research Foundation of Korea (No. NRF-2022R1F1A1073128). He is also supported by a Start-up Research Grant for new faculty at UNIST (No. 1.210139.01), UNIST AI Incubator Grant No. 1.230038.01, UNIST UBSI Grants No. 1.230168.01 and No. 1.230078.01, as well as the Industry Research Project No. 2.220916.01 funded by Samsung SDS in Korea. He is also partly supported by the BK21 Program (“Next Generation Education Program for Mathematical Sciences,” No. 4299990414089) funded by the Ministry of Education in Korea and the National Research Foundation of Korea (NRF).

-
- [1] W. Fulton, *Introduction to Toric Varieties*, Annals of Mathematics Studies (Princeton University Press, Princeton, NJ, 1993).
 - [2] N. C. Leung and C. Vafa, Branes and toric geometry, [arXiv:hep-th/9711013](#).
 - [3] B. R. Greene, String theory on Calabi-Yau manifolds, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 96): Fields, Strings, and Duality* (1996), pp. 543–726, [arXiv:hep-th/9702155](#).
 - [4] M. R. Douglas, B. R. Greene, and D. R. Morrison, Orbifold resolution by D-branes, *Nucl. Phys.* **B506**, 84 (1997).
 - [5] M. R. Douglas and G. W. Moore, D-branes, quivers, and ALE instantons, [arXiv:hep-th/9603167](#).
 - [6] A. E. Lawrence, N. Nekrasov, and C. Vafa, On conformal field theories in four-dimensions, *Nucl. Phys.* **B533**, 199 (1998).
 - [7] B. Feng, A. Hanany, and Y.-H. He, D-brane gauge theories from toric singularities and toric duality, *Nucl. Phys.* **B595**, 165 (2001).
 - [8] B. Feng, A. Hanany, and Y.-H. He, Phase structure of D-brane gauge theories and toric duality, *J. High Energy Phys.* **08** (2001) 040.
 - [9] R. Kenyon, An introduction to the dimer model, [arXiv:math/0310326](#).
 - [10] P. Kasteleyn, Graph theory and crystal physics, *Graph theory and theoretical physics*, 43 (1967).
 - [11] S. Franco, A. Hanany, K. D. Kennaway, D. Vegh, and B. Wecht, Brane dimers and quiver gauge theories, *J. High Energy Phys.* **01** (2006) 096.
 - [12] A. Hanany and K. D. Kennaway, Dimer models and toric diagrams, [arXiv:hep-th/0503149](#).
 - [13] S. Franco, A. Hanany, D. Martelli, J. Sparks, D. Vegh, and B. Wecht, Gauge theories from toric geometry and brane tilings, *J. High Energy Phys.* **01** (2006) 128.
 - [14] S. Franco, Dimer models, integrable systems and quantum Teichmüller space, *J. High Energy Phys.* **09** (2011) 057.
 - [15] A. B. Goncharov and R. Kenyon, Dimers and cluster integrable systems, [arXiv:1107.5588](#).

- [16] S. Franco, Bipartite field theories: From D-brane probes to scattering amplitudes, *J. High Energy Phys.* **11** (2012) 141.
- [17] N. Arkani-Hamed, J.L. Bourjaily, F. Cachazo, A.B. Goncharov, A. Postnikov, and J. Trnka, *Grassmannian Geometry of Scattering Amplitudes* (Cambridge University Press, Cambridge, England, 2016), 10.1017/CBO9781316091548.
- [18] S.S. Razamat, Quivers and fractons, *Phys. Rev. Lett.* **127**, 141603 (2021).
- [19] S. Franco and D. Rodriguez-Gomez, Quivers, lattice gauge theories, and fractons, *Phys. Rev. Lett.* **128**, 241603 (2022).
- [20] N.C. Leung and C. Vafa, Branes and toric geometry, *Adv. Theor. Math. Phys.* **2**, 91 (1998).
- [21] A. Hanany and A.M. Uranga, Brane boxes and branes on singularities, *J. High Energy Phys.* **05** (1998) 013.
- [22] M. Aganagic, A. Karch, D. Lust, and A. Miemiec, Mirror symmetries for brane configurations and branes at singularities, *Nucl. Phys.* **B569**, 277 (2000).
- [23] K. Hori, A. Iqbal, and C. Vafa, D-branes and mirror symmetry, [arXiv:hep-th/0005247](https://arxiv.org/abs/hep-th/0005247).
- [24] K. Hori and C. Vafa, Mirror symmetry, [arXiv:hep-th/0002222](https://arxiv.org/abs/hep-th/0002222).
- [25] K. Hori, S. Katz, A. Klemm, R. Pandharipande, R. Thomas, C. Vafa *et al.*, *Mirror Symmetry*, Vol. 1 of Clay Mathematics Monographs (American Mathematical Society, Providence, RI, 2003).
- [26] B. Feng, Y.-H. He, K.D. Kennaway, and C. Vafa, Dimer models from mirror symmetry and quivering amoebae, *Adv. Theor. Math. Phys.* **12**, 489 (2008).
- [27] S. Franco, S. Lee, R.-K. Seong, and C. Vafa, Brane brick models in the mirror, *J. High Energy Phys.* **02** (2017) 106.
- [28] B. Feng, A. Hanany, Y.-H. He, and A.M. Uranga, Toric duality as Seiberg duality and brane diamonds, *J. High Energy Phys.* **12** (2001) 035.
- [29] B. Feng, S. Franco, A. Hanany, and Y.-H. He, Symmetries of toric duality, *J. High Energy Phys.* **12** (2002) 076.
- [30] N. Seiberg, Electric—magnetic duality in supersymmetric nonAbelian gauge theories, *Nucl. Phys.* **B435**, 129 (1995).
- [31] M. Ciucu, A complementation theorem for perfect matchings of graphs having a cellular completion, *J. Comb. Theory Ser. A* **81**, 34 (1998).
- [32] R.W. Kenyon, J.G. Propp, and D.B. Wilson, Trees and matchings, [arXiv:math/9903025](https://arxiv.org/abs/math/9903025).
- [33] R. Kenyon, A. Okounkov, and S. Sheffield, Dimers and amoebae, [arXiv:math-ph/0311005](https://arxiv.org/abs/math-ph/0311005).
- [34] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, *Philos. Mag. J. Sci.* **2**, 559 (1901).
- [35] H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* **24**, 417 (1933).
- [36] J.E. Jackson, *A User's Guide to Principal Components* (John Wiley & Sons, New York, 2005).
- [37] I.T. Jolliffe, *Principal Component Analysis for Special Types of Data* (Springer, New York, 2002).
- [38] M.P. Deisenroth, A.A. Faisal, and C.S. Ong, *Mathematics for Machine Learning* (Cambridge University Press, Cambridge, England, 2020).
- [39] G.E. Hinton and S. Roweis, Stochastic neighbor embedding, in *Advances in Neural Information Processing Systems* (2002), Vol. 15.
- [40] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* **9**, 2579 (2008).
- [41] Y.-H. He, Deep-learning the landscape, [arXiv:1706.02714](https://arxiv.org/abs/1706.02714).
- [42] D. Krefl and R.-K. Seong, Machine learning of Calabi-Yau volumes, *Phys. Rev. D* **96**, 066014 (2017).
- [43] F. Ruehle, Evolving neural networks with genetic algorithms to study the string landscape, *J. High Energy Phys.* **08** (2017) 038.
- [44] J. Carifio, J. Halverson, D. Krioukov, and B.D. Nelson, Machine learning in the string landscape, *J. High Energy Phys.* **09** (2017) 157.
- [45] F. Ruehle, Data science applications to string theory, *Phys. Rep.* **839**, 1 (2020).
- [46] V. Jejjala, D.K. Mayorga Pena, and C. Mishra, Neural network approximations for Calabi-Yau metrics, *J. High Energy Phys.* **08** (2022) 105.
- [47] H. Erbin and R. Finotello, Machine learning for complete intersection Calabi-Yau manifolds: A methodological study, *Phys. Rev. D* **103**, 126014 (2021).
- [48] M. Larfors, A. Lukas, F. Ruehle, and R. Schneider, Learning size and shape of Calabi-Yau spaces, [arXiv:2111.01436](https://arxiv.org/abs/2111.01436).
- [49] P. Berglund, B. Campbell, and V. Jejjala, Machine learning Kreuzer-Skarke Calabi-Yau threefolds, [arXiv:2112.09117](https://arxiv.org/abs/2112.09117).
- [50] G. Arias-Tamargo, Y.-H. He, E. Heyes, E. Hirst, and D. Rodriguez-Gomez, Brain webs for brane webs, *Phys. Lett. B* **833**, 137376 (2022).
- [51] O. Aharony and A. Hanany, Branes, superpotentials and superconformal fixed points, *Nucl. Phys.* **B504**, 239 (1997).
- [52] O. Aharony, A. Hanany, and B. Kol, Webs of (p,q) five-branes, five-dimensional field theories and grid diagrams, *J. High Energy Phys.* **01** (1998) 002.
- [53] S. Chen, Y.-H. He, E. Hirst, A. Nestor, and A. Zahabi, Mahler measuring the genetic code of amoebae, [arXiv:2212.06553](https://arxiv.org/abs/2212.06553).
- [54] J. Bao, Y.-H. He, and E. Hirst, Neurons on amoebae, *J. Symb. Comput.* **116**, 1 (2022).
- [55] G. Mikhalkin, Amoebas of algebraic varieties and tropical geometry, [arXiv:math/0403015](https://arxiv.org/abs/math/0403015).
- [56] G. Mikhalkin, Amoebas of algebraic varieties, [arXiv:math/0108225](https://arxiv.org/abs/math/0108225).
- [57] E. Witten, Phases of $N = 2$ theories in two dimensions, *Nucl. Phys.* **B403**, 159 (1993).
- [58] S. Benvenuti, B. Feng, A. Hanany, and Y.-H. He, Counting BPS operators in gauge theories: Quivers, syzygies and plethystics, *J. High Energy Phys.* **11** (2007) 050.
- [59] B. Feng, A. Hanany, and Y.-H. He, Counting gauge invariants: The plethystic program, *J. High Energy Phys.* **03** (2007) 090.
- [60] A. Butti, D. Forcella, A. Hanany, D. Vegh, and A. Zaffaroni, Counting chiral operators in quiver gauge theories, *J. High Energy Phys.* **11** (2007) 092.
- [61] E. Brieskorn, Beispiele zur differentialtopologie von singularitäten, *Inventiones Math.* **2**, 1 (1966).
- [62] F. Hirzebruch, *Singularities and Exotic Spheres* (Societe Mathematic de France, 1968).
- [63] D.R. Morrison and M.R. Plesser, Nonspherical horizons. 1, *Adv. Theor. Math. Phys.* **3**, 1 (1999).

- [64] N. Metropolis and S. Ulam, The Monte Carlo method, *J. Am. Stat. Assoc.* **44**, 335 (1949).
- [65] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [66] J. Bao, S. Franco, Y.-H. He, E. Hirst, G. Musiker, and Y. Xiao, Quiver mutations, Seiberg duality and machine learning, *Phys. Rev. D* **102**, 086013 (2020).
- [67] P.-P. Dechant, Y.-H. He, E. Heyes, and E. Hirst, Cluster algebras: Network science and machine learning, [arXiv: 2203.13847](https://arxiv.org/abs/2203.13847).
- [68] T. J. Santner and D. E. Duffy, *The Statistical Analysis of Discrete Data* (Springer Science & Business Media, New York, 2012).
- [69] A. Agresti, *Categorical Data Analysis* (John Wiley & Sons, New York, 2012), Vol. 792.
- [70] A. Hanany and R.-K. Seong, Brane tilings and reflexive polygons, *Fortschr. Phys.* **60**, 695 (2012).
- [71] J. Davey, A. Hanany, and J. Pasukonis, On the classification of brane tilings, *J. High Energy Phys.* **01** (2010) 078.
- [72] S. Franco, S. Lee, and R.-K. Seong, Brane brick models, toric Calabi-Yau 4-folds and 2d (0, 2) quivers, *J. High Energy Phys.* **02** (2016) 047.
- [73] S. Franco, D. Ghim, S. Lee, R.-K. Seong, and D. Yokoyama, 2d (0, 2) quiver gauge theories and D-branes, *J. High Energy Phys.* **09** (2015) 072.
- [74] A. Gadde, S. Gukov, and P. Putrov, (0, 2) trialities, *J. High Energy Phys.* **03** (2014) 076.
- [75] S. Franco, S. Lee, and R.-K. Seong, Brane brick models and 2d (0, 2) triality, *J. High Energy Phys.* **05** (2016) 020.