Persistent homology analysis of deconfinement transition in effective Polyakov-line model

Takehiro Hirakida\textsuperscript{a,*}, Kouji Kashiwa\textsuperscript{b}, Junpei Sugano\textsuperscript{a}, Junichi Takahashi\textsuperscript{c}, Hiroaki Kouno\textsuperscript{d}, Masanobu Yahiro\textsuperscript{a}

\textsuperscript{a}Department of Physics, Graduate School of Sciences, Kyushu University, Fukuoka 819-0395, Japan
\textsuperscript{b}Fukuoka Institute of Technology, Wajiro, Fukuoka 811-0295, Japan
\textsuperscript{c}Numerical Prediction Division, Forecast Department, Japan Meteorological Agency, Tokyo, 100-8122, Japan
\textsuperscript{d}Department of Physics, Saga University, Saga 840-8502, Japan

Abstract

The persistent homology analysis is applied to the effective Polyakov-line model on a rectangular lattice to investigate the confinement-deconfinement nature. The lattice data are mapped onto the complex Polyakov-line plane without taking the spatial average and then the plane is divided into three domains. This study is based on previous studies for the clusters and the percolation properties in lattice QCD, but the mathematical method of the analyses are different. The spatial distribution of the data in the individual domain is analyzed by using the persistent homology to obtain information of the multiscale structure of center clusters. In the confined phase, the data in the three domains show the same topological tendency characterized by the birth and death times of the holes which are estimated via the filtration of the alpha complexes in the data space, but do not in the deconfined phase. By considering the configuration averaged ratio of the birth and death times of holes, we can construct the nonlocal order-parameter of the confinement-deconfinement transition from the multiscale topological properties of center clusters.

Keywords: Lattice QCD, Deconfinement transition, Persistent homology

*Corresponding author

Email address: hirakida@phys.kyushu-u.ac.jp (Takehiro Hirakida)
1. Introduction

Topological properties of the system can play a crucial role in the classification of phase transitions. There are several ways to apply topological knowledge in mathematics to physics. Recently, the persistent homology [1, 2], which is one of the ways to introduce the topological viewpoint to physics, attracted much more attention in the classification of system’s structure [3, 4, 5]. In this article, we utilize the persistent homology to investigate phase structures in the effective Polyakov-line model which is the effective model of quantum chromodynamics (QCD) in the heavy quark mass regime.

The property of the confinement-deconfinement nature in the pure Yang-Mills theory and also QCD is a long-standing problem and thus several proposal have been stated so far; see Ref. [6]. At least in the pure Yang-Mills theory, the Polyakov-line (loop) which relates the gauge invariant holonomy can exactly describe the confinement-deconfinement transition because it can be expressed by using the one-quark excitation free-energy. The finite value of the Polyakov-line also indicates the spontaneous breaking of $\mathbb{Z}_3$ symmetry. Thus, we can regard the Polyakov-line as the order-parameter of the confinement-deconfinement transition. However, in QCD, the Polyakov-line is not an exact order parameter anymore, since the existence of dynamical quark breaks $\mathbb{Z}_3$ symmetry explicitly. Hence our understanding of the confinement-deconfinement nature is still limited.

From the topological viewpoint, it has been recently suggested that the confinement and deconfinement states of QCD at zero temperature can be clarified via the topological order [7] which is characterized by the ground-state degeneracy in the compactified space [8]. Note that the notion of the topological order is valid even if there is no symmetry that governs the transition. Analogy of the topological order in QCD has been applied to finite temperature in Refs. [9, 10, 11] by considering the non-trivial free-energy degeneracy at finite imaginary chemical potential. Therefore, it is natural to expect that the topology can bring us to deeper understanding of the confinement-deconfinement
Another interesting study based on the topological viewpoint is the study of the center clusters and the percolation [12, 13, 14]. The center clusters can be classified from the behavior of Polyakov-line. In the complex Polyakov-line plane, the spatial distribution of local Polyakov-line can have the wide spread; see Fig. 2 in Ref. [14]. The distribution of the center cluster depends on the temperature and have important information of confinement-deconfinement nature. It was shown that the fractal dimension related to the scale properties of the center cluster distribution is useful to analyze the confinement-deconfinement transition [14]. It was found that the fractal dimension changes rapidly through the transition. Note that the study has been extended to the case of the full QCD [15] where there is no exact symmetry that governs the confinement-deconfinement transition. Hence, it is naturally expected that the analyses of the multiscale properties of the center cluster distribution is important and useful to investigate the confinement-deconfinement nature of QCD. Combining the topological notion and the multiscale properties, in this study, we investigate the multiscale topological properties of spatial distribution of the center cluster by another method, namely, the persistent homology to obtain the deeper understanding of confinement-deconfinement nature.

The persistent homology is widely used to investigate the multiscale topological structure of the distribution of data set; for example, the spatial structures of matter such as the glass and liquid states of SiO$_2$ can be well classified from the persistent homology, even if these states have almost the same spatial distributions [4]. Therefore, it seems to be suitable to apply the persistent homology analysis to investigate the structure of the center clusters.

In Fig. 1, we show a schematic figure of the multiscale topological structure that the persistent homology can extract. In the actual calculation of the persistent homology, we just need the data set and analyze the data from filtration, as explained later. In the calculation, we can prepare several different data forms such as the bare data, the averaged data, and the mapped data to analyze the topological structure of data space. Therefore, we may systematically
analyze the topological structure from the persistent homology with different data forms.

In this article, as a first step to apply the persistent homology to QCD, we consider the effective Polyakov-line model as a QCD effective model in the heavy quark mass regime. This article is organized as follows. In Sec. 2 the persistent homology is explained. After giving an introductory summary of persistent homology in Sec. 2.1, some technical and mathematical explanations are given in Sec. 2.2. The formulation of the effective Polyakov-line model is shown in Sec. 3. Numerical results are shown in Sec. 4. Section 5 is devoted to summary.

2. Persistent homology

2.1. Introductory summary of persistent homology

First, we show an introduction of persistent homology. Technical and mathematical details will be given in the next subsection. We consider the three data
points in two dimensional data space, and consider a two dimensional “ball” the center of which is the data point and the radius of which is denoted by $r$. As shown in Fig. 2 (a), when $r$ is small enough, three data points are isolated each other. In this case, we regard the ball as a zero-dimensional point which is called a “zero-dimensional simplex” in the contexture of homology.

If the radius $r$ increases and becomes larger than the half of the distance between two data points, two ball overlap as shown in Fig. 2 (b). In this case, we regard that the two data points are “connected” and a line segment newly appears in addition to the data points themselves. The line segment is a one-dimensional object and called a “one-dimensional simplex”. Note that, in Fig. 2 (b), there is a “hole” surrounded by three line segments. If $r$ becomes larger than the half of the maximum of distances between two points, three balls overlap and hole disappears as is shown in Fig. 2 (c). On the other hand, a “two-dimensional simplex”, namely, a triangle, newly appears.

As $r$ increases, the number of the simplexes in the diagram increases. In Fig. 2 (a) there are only three data points. In Fig. 2 (b), in addition to data points, three line segments appear. There are three data points, three line seg-

![Figure 2: Ball model of three data points, where $r$ denotes the radius of the ball. Corresponding complexes are also shown. (a) Three points. (b) Three points and three line segments. (c) Three points, three line segments and a triangle.](image-url)
ments and one triangle in Fig. 2 (c). The set of simplexes is called a “complex”. (For rigorous definitions of complex, see the next subsection.)

As \( r^2 \) varies, the process of changing of contents of complex is called “filtration”. (According to the terminology of persistent homology, one calls \( r^2 \) “time”, but the “time” is not real time. It is simply measuring an effective size of the overlap of balls.) The filtration process can be regarded as a coarse grain-
ing of the structure of data points. We can understand the multiscale structure of data space by seeing the filtration process. In particular, the existence of holes in the filtration process is important in the context of homology. As already been illustrated in simple examples, intuitively, the persistent homology counts holes created via the filtration. A hole was born in Fig. 2 (b) and died in Fig. 2 (c). Then, as shown in Fig. 3, we should plot a persistent diagram (PD) whose horizontal and vertical axis are the birth and death times; the birth time \( b = r_i^2 \) denotes the creation time of the holes in the filtration and the death time \( d = r_j^2 \) does the disappearing time of corresponding holes.

If the ratio of the birth time to the death one is approximately 1, we should consider the hole having the birth and death time set \((b, d)\) as a noise because its life time \(d - b\) is almost zero. Thus, we may pick up the nontrivial correlation between data points from the persistent diagram, namely, by seeing birth-death-time sets far from the diagonal line in Fig. 3. The hole which has a long life time is “persistent”, hence, meaningful.

2.2. Technical details of persistent homology

Let us express finite data points as \( P = \{ x_i \in \mathbb{R}^N \mid i = 1, \ldots, m \} \) in the \( N \)-dimensional space where \( m \) is the number of data points. For details of the computation of the persistent homology, see Ref. [16] as an example.

To discuss the topological structure of the data points, one interesting way is the calculation of the persistent homology [17, 18, 19] and then we need the geometric model [20] which characterized multiscale topological properties in \( P \).
One of the models is the $r$-ball model

$$P_r = \bigcup_{i=1}^{m} B_r(x_i),$$  \hspace{1cm} (1)

where $B_r(x_i) = \{ y \in \mathbb{R}^N : ||y - x_i|| \leq r \}$ and $r$ means the radius of each ball and its center is $x_i$. The condition of the crossing of two balls is expressed as

$$B_r(x_i) \cap B_r(y_j) \neq \emptyset,$$  \hspace{1cm} (2)

where $\emptyset$ denotes an empty set. Similarly, the condition of the crossing of $(k+1)$-
balls is expressed as
\[
\bigcap_{j=0}^{k} B_r(x_i) \neq \emptyset. \tag{3}
\]

The Čech complex \((P, r)\) is a nerve of the \(r\)-ball collection \(\Phi = \{B_r(x_i) | i = 1, \ldots, m\}\) and is defined by
\[
C(P, r) = \left\{ \{x_i_0, \ldots, x_i_k\} \left| \bigcap_{j=0}^{k} B_r(x_i_j) \neq \emptyset, \ k = 0, \ldots, m-1 \right. \right\}, \tag{4}
\]
where the set \(\{x_i_0, \ldots, x_i_k\}\) represents the \(k\)-dimensional simplex which has the points \(x_i_0, \ldots, x_i_k\) as vertexes. For example, \(\{x_i_0, \ldots, x_i_k\}\) for \(k = 1, 2, 3\) is a line segment, a triangle and a tetrahedron, respectively. In Eq. (4), for convenience, we have defined
\[
\left\{ \{x_i_0\} \left| \bigcap_{j=0}^{0} B_r(x_i_j) \neq \emptyset \right. \right\} \equiv \{\{x_1\}, \ldots, \{x_m\}\}, \tag{5}
\]
which corresponds to the collection of vertex points themselves. Note that the \((k + 1)\)-balls crossing corresponds to the \(k\)-dimensional simplex. Hence, \(k\) can be greater than the dimension \(N\) of data space, when the number \(m\) of data points is greater than \(N + 1\).

Simple examples of Čech complex are shown in Fig. 4. In the left panel, the Čech complex is given by
\[
C(P, r_i) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_1\}\} \tag{6}
\]
In this case, there is no 2-dimensional simplex, namely, a (shaded) triangle and there is a hole surrounded by 0 or 1-dimensional simplexes, namely, points and line segments. The right panel shows the case with larger \(r\). The Čech complex...
is given by

\[ \mathcal{C}(P, r_j) = \{ \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_1\}, \{x_1, x_2, x_3\} \} \quad (7) \]

In this case, the (shaded) triangle appears and the hole vanished. (Note that, in this case, the simplex with dimension higher than \( N = 2 \) never appears for any \( r \), since \( m = 3 \) is not greater than \( N + 1 = 3 \).) As is clearly seen in the samples above, in general, the following relation holds true.

\[ \mathcal{C}(P, r) \subset \mathcal{C}(P, r'), \quad r < r'. \quad (8) \]

By using this relation, the filtration (increasing sequence) for the Čech complex is now defined as

\[ \mathcal{C}(P, r_0) \subset \cdots \subset \mathcal{C}(P, r_t) \subset \cdots \subset \mathcal{C}(P, r_{t_{\text{max}}}), \quad (9) \]

where \( r_0 < \cdots < r_t < \cdots < t_{\text{max}} \) and the filtration time \( t \) is defined by \( t = r^2 \). In principle, the time \( t \) is a continuous real number. However, since the data space itself is discretized in actual calculations, \( r \) and \( t \) are also treated as discretized quantities. According to the nerve theorem, the nerve of \( \Phi = \{ B_r(x_i) \} | i = 1, \cdots, m \} \), namely, \( \mathcal{C}(P, r) \) is homotopy equivalent with \( P_r \). Hence, the filtration of \( \mathcal{C}(P, r) \) has the information of the multiscale topological properties, namely, the persistent homology, in \( P \). The filtration process can be regarded as a coarse graining. By the effects of the coarse graining, a topological structure (hole) appears at some time \( t \) (at some scale \( r \)) and disappears at larger time (at larger scale). For example, in Fig. 4, we see that the hole was born at the birth time \( t_i \equiv r_i^2 \) and died at the death time \( t_j \equiv r_j^2 \).

To know the filtration, we need to check the crossing of multi balls. Checking the crossing of two balls following the above condition (2) is easy in the numerical computation. However, the numerical cost to check more than the crossing of three balls is large. In the case of simple \( r \)-ball model, the maximum of the dimension of the simplex which appears in the filtration process becomes larger
Figure 4: A schematic figure of the $r$-ball model (top) and the corresponding Čech complex (bottom) for 2-dimensional data space.

As $m$ (the number of data points) increases, and the computation time to analyze the persistent homology becomes very large. Thus, we employ the alpha complex in this study as the more convenient version of the $r$-ball model as usual in the persistent homology analysis.

To determine the alpha complex, we first assign the region $V(x_i)$ to the data points as

$$V(x_i) = \{ x \in \mathbb{R}^N | ||x - x_i|| \leq ||x - x_j||, \ 1 \leq j \leq m, \ j \neq i \},$$ \hspace{1cm} (10)

and then we have $\mathbb{R}^N = \bigcup_{i=1}^{m} V(x_i)$. The region, $V(x_i)$, is so called the Voronoi region. In the actual calculation, we consider the Voronoi region consisted of $P$. Next, we consider the intersection defined by

$$W_r(x_i) = B_r(x_i) \cap V(x_i).$$ \hspace{1cm} (11)

It means that we restrict balls, $B_r(x_i)$, in the corresponding Voronoi regions. The alpha complex $\alpha(P, r)$ is defined by the nerve of the collection $\Psi = \{ W_r(x_i) \} i =
1, \cdots, j}, namely,

$$\alpha(P, r) = \left\{ \{x_{i_0}, \cdots, x_{i_k}\} \mid \bigcap_{j=0}^{k} W_r(x_{i_j}) \neq \emptyset, \quad k = 0, \cdots, m - 1 \right\}. \quad (12)$$

A necessary and sufficient condition for the existence of the $k$-simplex in the alpha complex $\alpha(P, r)$ is

$$\bigcap_{j=0}^{k} W_r(x_{i_j}) \neq \emptyset. \quad (13)$$

If the condition mentioned above is satisfied for $r$, it holds true for $r'$ satisfying $r' > r$. Hence we get

$$\alpha(P, r) \subset \alpha(P, r'), \quad r < r'.$$ \quad (14)

As is in the case of Čech complex, the filtration for the alpha complex is defined as

$$\alpha(P, r_0) \subset \cdots \subset \alpha(P, r_t) \subset \cdots \subset \alpha(P, r_{t_{\text{max}}}), \quad (15)$$

where $r_0 < \cdots < r_t < \cdots < r_{t_{\text{max}}}$.

A schematic figure of the filtration of alpha complexes are shown in Fig. 5. In the left panel, the alpha complex is given by

$$\alpha(P, r_i) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_1\}\} \quad (16)$$

In this case, there are only 0 or 1-dimensional simplexes (points and line segments), and a hole surrounded by them is formed. The right panel shows the case with larger $r$. In this case, the alpha complex is given by

$$\alpha(P, r_j) = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_4, x_1\}, \{x_1, x_3\}, \{x_1, x_2, x_3\}, \{x_1, x_3, x_4\}\} \quad (17)$$
The 2-dimensional simplexes (shaded triangles) appear and the hole vanishes. The hole appears at the birth time \( b = r_i^2 \) and disappears at the death time \( d = r_j^2 \).

It should be remarked that 3-dimensional simplex (tetrahedron) \( \{x_1, x_2, x_3, x_4\} \), which have a larger dimension than the dimension \( N = 2 \) of data space itself, does not appear in the right panel of Fig. 5 (also in Eq. (17)). The relation \( W_r(x_2) \cap W_r(x_4) = \emptyset \) holds true and the 3-dimensional simplex never appears even if \( r \) becomes larger than \( r_j \), since \( V(x_2) \) is not neighboring to \( V(x_4) \). In general, the alpha complex defined for \( N \)-dimensional data space does not have simplexes the dimension of which is greater than \( N \), when there is no special point \( x \) which satisfies

\[
||x - x_{i_1}|| = \cdots = ||x - x_{i_j}|| = ||x - x_{i_{N+2}}||.
\]

This situation makes the computation of persistent homology \([17, 18, 19]\) in alpha complex much easier than that in \( \acute{\text{C}} \)ech complex. Actually, by using this filtration of alpha complex, we can calculate the persistent homology. Actual construction of the alpha complex in the numerical simulation, we employ DIPHA (Distributed Persistent Homology Algorithm) library \([21]\) via the homcloud-base software \([22]\).
Figure 5: A schematic figure of the filtration of the alpha complex for two dimensional data space. The left (right) panel shows the alpha complex $\alpha(P, r_i)$ ($\alpha(P, r_j)$) ($r_i < r_j$). The dashed lines denote the borders of the Voronoi region. In the left panel, the hole surrounded by $W_{r_i}(x_i)$ was born at the birth time $b = r_i^2$. In the right panel, the hole disappears at the death time $d = r_j^2$.

3. Effective Polyakov-line model

One of the interesting QCD effective models is the effective Polyakov-line (EPL) model [23, 24]. The grand canonical partition function [25] is

$$Z = \int \mathcal{D}U \; e^{-(S_G + S_Q)},$$

$$(20)$$

$$S_Q = \sum_x L_Q(x),$$

$$S_G = -\kappa \sum_x \sum_{k=1}^3 \left( \text{Tr}[U_x] \text{Tr}[U_{x+k}^\dagger] + \text{c.c.} \right),$$

$$(21)$$

where $U_x$ means the Polyakov-line holonomy defined by

$$U_x = \text{diag}(e^{i\phi_1}, e^{i\phi_2}, e^{-i(\phi_1 + \phi_2)}),$$

$$(22)$$
with the real parameters \( \phi_1 \) and \( \phi_2 \), and the coupling constant \( \kappa \) in \( S_G \) relates with the inverse temperature \( \beta = 1/T \). Note that \( S_G \) and \( S_Q \) correspond to the gluon and quark actions, respectively, in QCD. Since we can consider a variety of the EPL model and thus we choose logarithmic \( L_Q \) in this study as

\[
L_Q = - \ln \left[ \det \left( 1 + e^{-\beta M} U_x \right)^{2N_f} \det \left( 1 + e^{-\beta M} U_x^\dagger \right)^{2N_f} \right],
\]

where \( M \) is the quark mass and \( N_f = 3 \) is the number of flavors. In this model, \( M \) and \( T \) should depend on the coupling constant \( \kappa \). However, it is difficult to take the line of constant physics unlike the lattice QCD simulation, and we need to solve the effective theory to obtain the relation according to Refs. [23, 24]. In this study, for qualitative study, we treat \( \beta M = M/T \) as a parameter independent of \( \kappa \), and consider the high or the low \( T \) regime by varying \( \kappa \).

Usually, the confinement-deconfinement nature in this model is characterized by the configuration average value \( \langle |L| \rangle \) of the spatial averaged Polyakov-line operator

\[
L = \frac{1}{V} \sum_x \frac{1}{3} \text{Tr}[U_x],
\]

where \( V \) is the three-dimensional volume; small \( \langle |L| \rangle \), \( \langle |L| \rangle \sim 0 \), indicates the confined phase and large \( \langle |L| \rangle \), \( \langle |L| \rangle \sim 1 \), does the deconfined phase. \( L \) is not invariant under \( Z_3 \) transformation, but is not an exact order parameter since \( S_Q \) breaks the \( Z_3 \) symmetry explicitly.

To perform the path integral of the EPL model, we use the Monte Carlo method; we generate configurations to replace the integral by the statistical sum. For the numerical simulation, we consider \( V = 24^3 \), \( \kappa \in [0.120, 0.150] \), and two cases with \( M/T = 5 \) and 10. For reader’s convenience, we here show the \( \kappa \)-dependence of \( \langle |L| \rangle \) in Fig. 6 where configuration average is taken by 50 configurations as an example. We can clearly see that this model exhibits the same tendency as QCD and there should be the phase transition between
\( \kappa = 0.135 \) and \( 0.140 \). Figure 7 shows the probability density of \( \text{Tr}[U_x] \) in the complex plane. At \( \kappa = 0.120 \) (confined phase), the probability density is widely spread in the plane. On the other hand, at \( \kappa = 0.150 \) (deconfined phase), the density is mainly distributed around the real axis and \( \text{Re}[\text{Tr}[U_x]] \gg 0 \).

![Figure 6: The \( \kappa \)-dependence of \( \langle |L| \rangle \) with \( M/T = 5, 10 \).](image)

Figure 7: The probability density of \( \text{Tr}[U_x] \) in the complex plane. The left and right panels show the results for \( \kappa = 0.120 \) and \( \kappa = 0.150 \) with \( M/T = 10 \), respectively.

### 4. Numerical results

In Refs. [12, 13, 14, 15], it is shown that the number of the percolating cluster (which has a same scale as the spatial volume) is 3 (1) in the confined (decon-
fined) phase, and the expectation value of the weight for the largest cluster normalized by the spatial volume is small in the confined phase and it rapidly grow up when the temperature becomes larger than the critical temperature. They indicate that the large structures exist at both phases. However, these results also indicate the possibility that the large hole structures may not appear in the deconfined phase since the largest cluster dominates too strongly. Therefore, by using the persistent homology, it is expected that we can extracted the large holes in the confined phase, while many tiny holes appear in the deconfined phase.

To analyze the spatial distributions of simulation data via the persistent homology analysis, we consider the following isolation procedure:

1. Map the bare data for each site to the complex Polyakov-line plane, configuration by configuration.
2. Divide the complex Polyakov-line plane into three domains as shown in Fig. 8. Note that the $Z_1$ and $Z_2$ domains are the $Z_3$-images of the $Z_0$ domain.
3. Prepare three lists, $Z_0$, $Z_1$, and $Z_2$, as data storage places. The storage places, $Z_0$, $Z_1$, and $Z_2$, become the $N^3$ dimensional list since we try to maintain the spatial information of the bare data.
4. Storage the mapped data to each list; if the data are located in the $Z_0$ domain in Fig. 8, the corresponding site is ON (full) in the $Z_0$ group and the corresponding sites are OFF (empty) in the $Z_{1,2}$ groups.

First we consider the spatial distribution of the data after the above isolation procedure in the case with $M/T = 10$. The typical examples are shown in Fig. 9 and Fig. 10 for $\kappa = 0.120$ and $\kappa = 0.150$, respectively. At $\kappa = 0.120$ (confined phase), the data are uniformly distributed, but the distribution is weighted toward the $Z_0$ domain and few data are located in the $Z_1$ and $Z_2$ domains at $\kappa = 0.150$ (deconfined phase).

Figure 11 shows the occupation rates of each domain in the lattice space. Here, we define the occupation rate as the ratio of the number of data points in
Figure 8: A schematic figure of Polyakov-line in the complex plane. To compute the persistent homology, we divide the plane into three $\mathbb{Z}_3$ domains, $\mathbb{Z}_0$, $\mathbb{Z}_1$, and $\mathbb{Z}_2$.

Each domain to the lattice volume $V$. For the calculation of the rates, we take a configuration average with 50 configurations. In the case with $M/T = 10$, the occupation rates of all domains are almost the same (about 33%) below $\kappa = 0.135$, while the $\mathbb{Z}_0$ domain occupies almost all lattice sites (90 $\sim$ 99%) in the deconfined phase. These behaviors are almost the same with the results shown in Ref. [12]. On the other hand, in the case with $M/T = 5$, the rate of $\mathbb{Z}_0$ domain is larger by about 11% than that of other domains at $\kappa = 0.120$, and the rate also increases as $\kappa$ becomes large. This difference between the two cases indicates that the explicit $\mathbb{Z}_3$ symmetry breaking is not negligible when $M/T = 5$.

From these situations, it is expected that the persistent diagram for all domains are almost the same in the confined phase, while the result in the $\mathbb{Z}_0$ domain of deconfined phase differs from that of other domains due to the difference in the occupation rate. If the center clusters with nontrivial topological structures exist in the lattice space of each domain, the persistent diagram may be different from that of data points distributed randomly. The comparison of the results of the EPL model with that of the random distributed configuration is discussed in the end of this section.
Figure 9: The spatial distributions of the data after the isolation for $\kappa = 0.120$ and $M/T = 10$. The top left, top right and bottom data show the mapped data of $Z_0$, $Z_1$ and $Z_2$.

Now we analyze the data of the EPL model by the method of the persistent homology. In this paper, since we are interested in 3-dimensional structure of the data, we search a 3-dimensional hole which is surrounded by 2-dimensional surfaces. In other words, we search the surfaces of the polyhedron the vertexes of which are data points. The typical examples of the persistent diagram are shown in Fig. 12 and Fig. 13 with $\kappa = 0.120$ and 0.150, respectively. The horizontal and the vertical axes are the birth time $b_i$ and the death time $d_i$ of the $i$-th hole, respectively.

Figure 14 summarizes typical shapes of data points in the persistent homol-
Figure 10: The spatial distributions of the data after the isolation for $\kappa = 0.150$ and $M/T = 10$. The top left, top right and bottom data show the mapped data of $Z_0$, $Z_1$ and $Z_2$.

ogy analysis with the set of these birth and death times. These shapes appear as dominant structures in the domains. In this study, we call the most generated, the second most generated, and the third most generated shapes as “dominant structures” or “typical shapes” for the domain. The shape (a) is the smallest cube on the present rectangular lattice and (b) is the chipped cube of (a). The shape (c) is the $2 \times 2 \times 2$ cube but it does not contain the data point at its center. Both (d) and (e) are cubes which can appear in the rectangular parallelepiped. In these shapes, (a), (b), (d) and (e) can be expected to appear as dominant holes of mapped data and thus those creation numbers become large;
for example, since the shape (c) can be considered as the composite of eight (b), eight (b) are always created if one (c) is created.

Let us consider the most simple structure (a) for the example of calculating the birth-death-time set \((b, d)\). This is the smallest cube in the lattice space. When the radius of the sphere at the vertex of the cube is \(r = \sqrt{1^2 + 1^2}/2 = 1/\sqrt{2}\), the half of the diagonal of the face, a hole surrounded by six faces is created. After increasing the radius, the hole disappears when the radius is \(r' = \sqrt{1^2 + 1^2 + 1^2}/2 = \sqrt{3}/2\), the half of the diagonal of the cube. Thus, we obtain the time of birth and death for the structure (a) as \((b, d) = (r^2, r'^2) = (0.5, 0.75)\).

The time of birth and death for the other structures in Fig. 14 are also obtained by the same calculation: The birth time is the square of the half of the longest diagonal of the face of the polyhedron corresponding to the hole, while the death time is the square of the half of the longest diagonal of the polyhedron. Based on the idea mentioned above, we can determine the typical
shapes in Fig. 14 inversely from their time of birth and death.

At $\kappa = 0.120$, the number of holes and their variety of shapes are almost same for each list, since $M/T$ is large and $\mathbb{Z}_3$ symmetry is approximately preserved. In this phase, the birth-death-time sets, $(1.3888 \cdots, 1.41666 \cdots)$, $(0.666 \cdots, 0.75)$ and $(1.60714 \cdots, 1.625)$, dominate the persistent diagram and these time sets are corresponding to the shape (d), (b) and (e) in Fig. 14, respectively. The shapes (d) and (e) are seemed to be constructed, because the occupation rate of each domain is about 33% and the distance between data points is moderately far.

In comparison, at $\kappa = 0.150$, mapped data are weighted toward the $\mathbb{Z}_0$ domain and thus many holes which have the fast birth and death times are created. However, mapped data are very sparse in the $\mathbb{Z}_1$ and $\mathbb{Z}_2$ domains, and thus we need long time to create holes and then the death times are not much larger than the birth times, hence, the life times $d_i - b_i$ of holes become very short. This fact can be seen from Fig. 13. In the deconfined phase, the birth-death-time sets, $(0.5, 0.75)$, $(0.666 \cdots, 0.75)$ and $(0.5, 1.0)$, dominate the persistent diagram for the $\mathbb{Z}_0$ list and these time sets are corresponding to the shape (a), (c) and (b) in Fig. 14. In the $\mathbb{Z}_1$ and $\mathbb{Z}_2$ lists, a few holes are created and then the mapped data are very sparse, and thus we do not show typical shape shown in Fig. 14. In this phase, the $\mathbb{Z}_0$ domain contains almost all lattice sites, so that the shape (a), the simplest and smallest structure in the lattice space, dominates in the persistent diagram. But this is a trivial structure. For the shapes (b) and (c), it is considered that they appear as minute structures if some data points belong to other domains. In particular, the one shape (c) with eight shapes (b) are constructed if the $2 \times 2 \times 2$ cube doesn’t contain the data point at its center. However, we may not be able to investigate these tiny structures in detail. To check them, it is needed to analyze in the larger lattice volumes, or restrict the Polyakov-line phase in smaller regions as in the percolation analysis in Refs. [12, 13, 14, 15]. Dominant structures in each phase are summarized in Table 1.

Next, we analyze the $\kappa$-dependence of the persistent homology with $M/T =$
Table 1: Dominant structures for each phase.

| Phase       | Group   | Dominant structure                                      |
|-------------|---------|--------------------------------------------------------|
| Confinement | $Z_0, Z_1, Z_2$ | (d), (1.3888 · · · , 1.41666 · · · ) |
|             |         | (b), (0.666 · · · , 0.75)                              |
|             |         | (c), (1.60714 · · · , 1.625)                           |
| Deconfinement | $Z_0$   | (a), (0.5, 0.75)                                       |
|             |         | (b), (0.666 · · · , 0.75)                              |
|             |         | (c), (0.5, 1.0)                                        |
|             | $Z_1, Z_2$ | No dominant structure                                  |

10. However, the individual persistent diagram is not convenient in this case and thus we introduce the averaged ratio of the birth and death times as

$$D/B = \frac{1}{N_{\text{hole}}} \sum \frac{d_i}{b_i},$$  \hspace{1cm} (25)$$

where $N_{\text{hole}}$ is the number of holes in each list. In this study, we calculate $D/B$ except for $d_i/b_i < 1.001$, since the hole with the ratio $d_i/b_i$ close to 1 is considered as noise. We here take the statistical average $\langle D/B \rangle$ by using 50 configurations and the result is shown in Fig. 15. In the confined phase, $\langle D/B \rangle$ always becomes $\sim 1.073$ for all lists. In comparison, $\langle D/B \rangle$ of the $Z_0$ list is changed into $\sim 1.40$ and reaches $\sim 1.459$ at $\kappa = 0.150$, but $\langle D/B \rangle$ of $Z_1$ and $Z_2$ decreases toward $\sim 1.03$ in the deconfined phase. This characteristic behavior indicates the shape change in each list by the confinement-deconfinement transition.

It seems that the increase of $\langle D/B \rangle$ in deconfined phase is mainly caused by the fact that the birth time is limited since the largest percolating cluster dominates too strongly [12, 13, 14], and the trivial structure (a) dominates in large $\kappa$-limit. On the other hand, the larger holes which has the larger birth time can exist in confined phase, since the three percolating clusters dominate not so strongly.

Figure 16 shows $\langle D/B \rangle$ with $M/T = 5$. The explicit breaking of $Z_3$ symmetry is enhanced at small $\kappa$ due to the lighter quark mass, and the $\kappa$-dependence
The persistent diagram for $\kappa = 0.120$. The top left, top right, and bottom panels show the result of $Z_0$, $Z_1$, and $Z_2$ list, respectively. The diagram of the random configuration with the occupation rate 33% is also shown. The color bar expresses the creation number of each hole.

The statistic error becomes quite small for the persistent homology analysis. This results indicate that the distributions of data points for all configurations in each list are very similar.

Let us compare the results of the EPL model with that of the randomly distributed configuration. Here, we generate the configuration of data points in random so as to have an arbitrary occupation rate, and analyze it with the persistent homology. In the EPL model with $M/T = 10$, the occupation rates of all domain are about 33% at $\kappa = 0.120$, and that of the $Z_0$ domain at $\kappa = 0.150$ is about 98%. The bottom right panels in Figs. 12 and 13 show the
Figure 13: The persistent diagram for $\kappa = 0.150$. The top left, top right, and bottom panels show the result of $Z_0$, $Z_1$, and $Z_2$ list, respectively. The diagram of the random configuration with the occupation rate 98% is also shown. The color bar expresses the creation number of each hole.

typical persistent diagrams of the random configurations corresponding to the occupation rate 98%.

Comparing the results of the EPL model (top left, top right, and bottom left panels in Fig. 12) with those of the random configuration with the occupation rate 33%, we find that the number of holes with the death time $3.5 \lesssim d \lesssim 5$ in the random configuration is much less than that in the EPL model at $\kappa = 0.120$ for $M/T = 10$. This result is confirmed in all domains and all configurations. In the random configuration of the occupation rate 33%, the dominant shapes are (d), (e) and the shape with the birth-death-time set $(1.125,1.25)$ (which does not appear in Fig. 14), while the dominant ones are (d), (e) and (b) in the EPL model. As is already mentioned, the shape (c) can be considered as
the composite of eight (b). If all of the shape (b) are used to construct the shape (c), the ratio of the number of (c) to that of (b) becomes 1/8 = 0.125. However, in our calculations, the ratio is 0, and this result indicates that the shape (c) does not exist but the shape (b) exists independently. This property appears in $\kappa \leq 0.130$ for $M/T = 10$. The shape (b) has the chipped $1 \times 1 \times 1$ cube structure, and this shape is likely to appear on the surface of the clusters of data points. These results indicate existence of the more complex multiscale topological structures in the confined phase with $M/T = 10$.

Meanwhile, comparing the results of the EPL model (top left, top right, and bottom left panels in Fig. 13) with those of the random configuration with the occupation rate 98%, we find that both persistent diagrams are almost the same. This may be a trivial result because almost all lattice sites are included in the $Z_0$ domain at $\kappa = 0.150$ for $M/T = 10$ and the large holes cannot be constructed. Thus, this result is consistent with the indication mentioned in the beginning of this section. The dominant structures of the random configuration are also the same with that of the $Z_0$ domain of the EPL model at the deconfined phase. For
Figure 15: The κ dependence of ⟨D/B⟩ for M/T = 10. Cube symbol shows the results of $Z_0$, circle symbol shows the results of $Z_1$, and triangle symbol shows the results of $Z_2$. The shapes on the figure represent the typical ones which dominate the values in the confined and the deconfined phases.

the $Z_0$ domain of deconfined phase, the result of the EPL model is qualitatively consistent with that in the random distribution model.

To compare the EPL model with the random distribution model more quantitatively, we compute ⟨D/B⟩ in the random model. Figure 17 shows the occupation rate dependence of ⟨D/B⟩ for the random configuration. To compute ⟨D/B⟩ for each occupation rate, we take a statistical average with 50 configurations. As the occupation rate increasing, ⟨D/B⟩ increases monotonously.

In Fig. 18, we compare ⟨D/B⟩ in the EPL model with that of the random configuration model with the same occupation rate. In the $Z_0$ domain for both $M/T$ and in the $Z_{1,2}$ domains in $\kappa \leq 0.135$ for $M/T = 10$ and $\kappa = 0.150$ for $M/T = 5$, the value in the EPL model is larger than that of the random configuration: for example, the value in the EPL model at $\kappa = 0.120$ for $M/T = 10$ is larger than $\langle D/B \rangle \sim 1.071$ for the occupation rate 33% in the random configuration model by $\sim 0.002$, while the value in the $Z_0$ at $\kappa = 0.150$ is larger
Figure 16: The $\kappa$ dependence of $\langle D/B \rangle$ for $M/T = 5$. Cube symbol shows the results of $Z_0$, circle symbol shows the results of $Z_1$, and triangle symbol shows the results of $Z_2$.

than $\langle D/B \rangle \sim 1.452$ for the rate 98% in the random configuration by $\sim 0.006$.

On the other hand, in the other domains and parameter sets, the values in the EPL model are smaller than those of the random configuration. These difference may come from that the EPL model has the different topological structures from the random one. In particular, in the $Z_0$ domain at $\kappa = 0.150$ for $M/T = 10$, $\langle D/B \rangle$ is somewhat different from that of the random configuration, although the persistent diagrams are almost the same in both the cases. In addition, especially around $\kappa = 0.130$ for $M/T = 5$ and $\kappa = 0.137$ for $M/T = 10$, the differences in the $Z_0$ domain become much large. These results indicate that the $Z_0$ domain with both $M/T$ has much different structures topologically there.

In Fig. 19, we compare the variance $\langle O^2 \rangle - \langle O \rangle^2$ of $D/B$ and $|L|$. For the results of $|L|$, a peak is located at $\kappa \sim 0.135$ for $M/T = 10$ and $\kappa \sim 0.125$ for $M/T = 5$. Meanwhile, the peaks in the case of $\langle D/B \rangle$ for $Z_0$ domain are located at almost the same $\kappa$ in the case of $|L|$, while the variances for the other domains behave almost the same. These results indicate that we can investigate properties of the confinement-deconfinement transition with $D/B$.

From these results, by using $\langle D/B \rangle$, we have characterized the phases. This
Figure 17: The occupation rate dependence of \( \langle D/B \rangle \) for the random configuration. The statistical error is shown, but it is tiny.

quantity is calculated after obtaining the persistent diagram, and this diagram has the information of the multiscale spatial structure of data in the entire lattice space. Thus, we can use \( \langle D/B \rangle \) as the nonlocal order parameter characterizing the confined and deconfined phases. As mentioned earlier, the statistical error of \( \langle D/B \rangle \) is much smaller than that of the expectation value of the Polyakov-line \( \langle |L| \rangle \). This may be caused by the fact that the structures for each configuration have the similar spatial distribution. Therefore, we can characterize the phases with a fewer configurations than in the case of calculating \( \langle |L| \rangle \).

5. Summary

In this study, we have investigated the confinement-deconfinement nature in the effective Polyakov-line model via the persistent homology analysis on the rectangular lattice. By using the persistent homology analysis, we can investigate the behavior of the multiscale spatial distribution of data inside the center clusters. To compute the persistent homology, we divide the complex Polyakov-line plane into three domains, \( Z_0 \), \( Z_1 \) and \( Z_2 \), and then the lattice data are mapped on the plane. We then prepare the \( Z_0 \), \( Z_1 \) and \( Z_2 \) lists to
storage the corresponding data. Our results are the following:

1. In the confined phase, the data are uniformly distributed on each list and thus we have similar persistent diagram for each list. This means that each list has same topological data structure. Typical shapes appearing in the data space are analyzed; actual shapes are depicted in Fig. 14. Comparing with the results of the random configuration with the occupation rate 33%, we find that some holes with the death time $3.5 \lesssim d \lesssim 5$, which do not appear in the random model, exist in the confined phase for the EPL model, and the one of the dominant structures is different from that of the random configuration. These results may indicate the existence of the nontrivial topological structures.

2. In the deconfined phase, the data are dense in the $Z_0$ domain, but they
become sparse in the $Z_1$ and $Z_2$ domains. Due to this property, the $Z_0$ and $Z_{1,2}$ lists have different topological structures in the data space. In the case of the $Z_0$ list, the smallest $1 \times 1 \times 1$ cube and many small structures appear. In comparison, the $Z_{1,2}$ lists have the structures which have late birth time and short life time. Comparing the result of the $Z_0$ domain with the results of the random configuration with the occupation rate 98%, the persistent diagram and the dominant structures are almost identical. The distribution of data points of the $Z_0$ domain in the deconfined phase are qualitatively described by the randomly distributed model.

3. To clearly show the phase transition, we consider the configuration averaged ratio of the birth and death time. This quantity shows the quite different behavior in the confined and the deconfined phases. In particular, the ratios of the $Z_0$ and $Z_{1,2}$ lists depart each other when the deconfined properties appear in the system. The difference comes from the structural change of the data by the confinement-deconfinement transition. Furthermore, this quantity in the EPL model differs from that of the random configuration with the same occupation rate. This difference may indicate the existence of the more complex multiscale topological structure of the domain.

4. By considering the configuration averaged ratio of the birth and death times of holes, we can construct the nonlocal order-parameter of the confinement-deconfinement transition from the topological viewpoint of the data space. The statistical error of this quantity is much smaller than that of the expectation value of Polyakov-line. This fact may indicate that the structures in the configurations have the similar spatial distribution. Therefore, by using this quantity, we can characterize the phases with a fewer configurations than that of Polyakov-line.

We can consider the following interesting and important future works:

1. In Refs. [12, 13, 14, 15], the center clusters are investigated by isolating the lattice with the narrower range of the Polyakov-line phases than ours. It
is interesting to combine this idea with our persistent homology method. In particular, to investigate more tiny structures in the deconfined phase, it is interesting to isolate the lattice into three domains according to the prescription presented in Ref. [12], and we analyze the multiscale spatial structure for each domain by using persistent homology.

2. In this study, we do not investigate so much near the phase transition point. The persistent homology analyses near the phase transition point should be necessary. In particular, the dependence of the persistent homology against the order of the phase transition is interesting.

3. The persistent homology is based on the spatial topology with the center clustering in this work. In this sense, this determination can treat not only the global feature of the spontaneous $\mathbb{Z}_3$ symmetry breaking but also the spatial distribution itself (see Figs. 9 and 10), while the ordinary determination of the confinement-deconfinement nature of QCD is done by using the averaged value of the Polyakov-line (see Fig. 6). In some cases, the present determination and the ordinary determination based on the Polyakov-line should provide the same feature of QCD, but the present determination may provide more information for the confinement-deconfinement nature of QCD if there is the topological structure in the spatial structure. Of course, we need more detailed studies on it, particularly in QCD with dynamical quarks, and hence it is our future work to be done.

4. Studying the spatial structures of the center vortex is also an interesting subject. It is considered that the center vortex goes through the loop constructed by the center projected link variables with nontrivial values, and that it is related strongly with the mechanism of the confinement-deconfinement nature [6, 26, 27]. By using the persistent homology, we may obtain the hints to describe the mechanism of the quark confinement from the viewpoint of the multiscale spatial structures of center vortices.

These results will be shown in elsewhere.
Acknowledgments

The authors are thankful to Masahiro Ishii, Akihisa Miyahara and Akira Ohno for fruitful discussions. This study is supported in part by the Grants-in-Aid for Scientific Research from JSPS (No. 18K03618 and No. 17K05446).

References

[1] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, in: Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on, IEEE, 2000, pp. 454–463.

[2] A. Zomorodian, G. Carlsson, Computing persistent homology, Discrete & Computational Geometry 33 (2) (2005) 249–274.

[3] T. Nakamura, Y. Hiraoka, A. Hirata, E. G. Escolar, Y. Nishiura, Persistent homology and many-body atomic structure for medium-range order in the glass, Nanotechnology 26 (30) (2015) 304001.

[4] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, Y. Nishiura, Hierarchical structures of amorphous solids characterized by persistent homology, Proceedings of the National Academy of Sciences 113 (26) (2016) 7035–7040.

[5] I. Donato, M. Gori, M. Pettini, G. Petri, S. De Nigris, R. Franzosi, F. Vaccarino, Persistent homology analysis of phase transitions, Physical Review E 93 (5) (2016) 052138.

[6] J. Greensite, Introduction, in: An Introduction to the Confinement Problem, Springer, 2010, pp. 1–2.

[7] M. Sato, Topological discrete algebra, ground state degeneracy, and quark confinement in QCD, Phys.Rev. D77 (2008) 045013. arXiv:0705.2476, doi:10.1103/PhysRevD.77.045013.
[8] X. G. Wen, Topological Order in Rigid States, Int.J.Mod.Phys. B4 (1990) 239. doi:10.1142/S0217979290000139.

[9] K. Kashiwa, A. Ohnishi, Topological feature and phase structure of QCD at complex chemical potential, Phys. Lett. B750 (2015) 282–286. arXiv:1505.06799, doi:10.1016/j.physletb.2015.09.036.

[10] K. Kashiwa, A. Ohnishi, Quark number holonomy and confinement-deconfinement transition, Phys. Rev. D93 (11) (2016) 116002. arXiv:1602.06037, doi:10.1103/PhysRevD.93.116002.

[11] K. Kashiwa, A. Ohnishi, Topological deconfinement transition in QCD at finite isospin density, Phys. Lett. B772 (2017) 669–674. arXiv:1701.04953, doi:10.1016/j.physletb.2017.07.033.

[12] C. Gattringer, Coherent center domains in SU(3) gluodynamics and their percolation at $T_c$, Phys. Lett. B690 (2010) 179–182. arXiv:1004.2200, doi:10.1016/j.physletb.2010.05.013.

[13] S. Borsanyi, J. Danzer, Z. Fodor, C. Gattringer, A. Schmidt, Coherent center domains from local Polyakov loops, J.Phys.Conf.Ser. 312 (2011) 012005. arXiv:1007.5403, doi:10.1088/1742-6596/312/1/012005.

[14] G. Endrodi, C. Gattringer, H.-P. Schadler, Fractality and other properties of center domains at finite temperature: SU(3) lattice gauge theory, Phys. Rev. D89 (5) (2014) 054509. arXiv:1401.7228, doi:10.1103/PhysRevD.89.054509.

[15] A. Schafer, G. Endrodi, J. Wellnhofer, Center clusters in full QCD at finite temperature and background magnetic field, Phys. Rev. D92 (1) (2015) 014509. arXiv:1506.07698, doi:10.1103/PhysRevD.92.014509.

[16] I. Obayashi, Y. Hiraoka, Persistence diagrams with linear machine learning models, arXiv preprint arXiv:1706.10082.
[17] P. Frosini, A distance for similarity classes of submanifolds of a euclidean space, Bulletin of the Australian Mathematical Society 42 (3) (1990) 407–415.

[18] V. Robins, Towards computing homology from finite approximations, in: Topology proceedings, Vol. 24, 1999, pp. 503–532.

[19] A. J. Zomorodian, Computing and comprehending topology: Persistence and hierarchical morse complexes, Citeseer, 2001.

[20] H. Edelsbrunner, E. P. Mücke, Three-dimensional alpha shapes, ACM Transactions on Graphics (TOG) 13 (1) (1994) 43–72.

[21] DIPHA plugin for python available at https://github.com/DIPHA/dipha.

[22] HomCloud available at http://www.wpi-aimr.tohoku.ac.jp/hiraoka_labo/homcloud-english.html.

[23] J. Greensite, K. Langfeld, Finding the effective Polyakov line action for SU(3) gauge theories at finite chemical potential, Phys. Rev. D90 (1) (2014) 014507. arXiv:1403.5844, doi:10.1103/PhysRevD.90.014507.

[24] J. Greensite, Comparison of complex Langevin and mean field methods applied to effective Polyakov line models, Phys. Rev. D90 (11) (2014) 114507. arXiv:1406.4558, doi:10.1103/PhysRevD.90.114507.

[25] T. Hirakida, J. Sugano, H. Kouno, J. Takahashi, M. Yahiro, Sign problem in Z_3-symmetric effective Polyakov-line model, Phys. Rev. D96 (7) (2017) 074031. arXiv:1705.00665, doi:10.1103/PhysRevD.96.074031.

[26] J. Greensite, The Confinement problem in lattice gauge theory, Prog. Part. Nucl. Phys. 51 (2003) 1. arXiv:hep-lat/0301023, doi:10.1016/S0146-6410(03)90012-3.

[27] M. Engelhardt, Generation of confinement and other nonperturbative effects by infrared gluonic degrees of freedom, Nucl. Phys. Proc. Suppl. 140 (2005) 92–105, [,92(2004)]. arXiv:hep-lat/0409023, doi:10.1016/j.nuclphysbps.2004.11.335.