



Sheaf Cohomology for Fragment-Sequencing in Hierarchical Block Rectangular Matrices with Spectral Gaps in the Presence of Random Effects and White Noise: The Chain

Orchidea Maria Lecian 

Department for Astronautics Engineering, Electrical and Energetics, Sapienza University of Rome, 00185 Rome, Italy

Article History

Received: October 19, 2025

Accepted: December 19, 2025

Published: December 30, 2025

Abstract

The sheaf cohomology of topological shift for the block-rectangular matrix-representation of the hierarchical Markov Model incorporates an analytic formulation of white noise and of random effects. New analytical techniques for fragment sequencing are developed. The fragment sequencing is obtained after the topological Markov chain of the adjacency matrix of the corresponding undirected graph; the presence of white noise and that of random effects are included. The paradigm consists in defining the hierarchical block rectangular matrices, from which the Topological Hidden Markov Models are issued (as clusters), with the aspects of Hidden Markov Models of 'multivariate Gaussian data' with vanishing mean; the generalized covariance matrix is studied. The model is compared with the stochastic properties of the decomposition that approximates experimental data. One of the previous results of the applications of the new method can be looked at in the analysis of the numerical simulation of the sequencing techniques. As an example, it is known that the Gojobori-Ichii-Nei model fails in reproducing the Jukes-Cantor scheme, while the Kimura matrix model succeeds in it. The difference is explained as the former model is not obtained from the Jukes-Cantor paradigm after application of the differential operators (for substituting the entries of the matrix), while the latter model is. In the present paper, the new analytical result is further accomplished, to calculate the maximal likelihood analytically in protein sequencing and in DNA-sequencing, in cases where the sequence of elements varies over time; the method solves analytically the phylogenetics computer programs. The analysed problem belongs to the nondeterministic polynomial time (NP) hard class of complexity.

Keywords:

markov chain; sheaf cohomology; hierarchical block rectangular matrices; fragment sequencing; machine learning

1. Introduction

The question raised in [1] regarding the work in [2] is addressed in the present paper. Specifically, the issue concerns establishing the well-posedness of comparing contact maps that have different bin sizes. The relevance of overlapping the contact maps is that it allows for the comparison of graphs (and related structures, such as distances.)

The further interrogation about white noise in [3] and about random effects [4] are codified within the same paradigm. As recalled from [5], solving the problem of contact-maps overlap belongs to the NP hard class of complexity [6,7]. A visual method for comparing an adjacency matrix is proposed in [8].

The question posed in [1] is addressed here analytically through the following strategy. A topological

Markov chain is constructed for the network starting from a rectangular hierarchical block matrix that incorporates white noise and random effects and encodes the cluster structure. The problem formulated in [1] is resolved by demonstrating that the topology of the manifold on which the chain is defined is endowed with a Hilbert metric. The answer to the interrogation of [1] is therefore here found to be analytical; for these purposes, the new paradigm in topology was created- the numerical methods are not used in the present paper.

The tasks of graph matching are addressed in the present guidelines for investigation using several methodologies. A spectral method for these purposes is developed in [8]. The adjacency matrix of the graphs to be compared is analyzed in the same work as far as the leading eigenvector (only) is concerned. The sequences are matched after developing the graph nodes to align them. The corresponding Markov chain is constructed, taking into account only the leading eigenvector of the adjacency matrix.

A key difference between the method proposed in the present paper and that proposed in [8] is that, in the formalism adopted here, the sequence order does not have to be re-aligned a posteriori. Indeed, the approach of [8] relies on reappreciating the paradigms of segmentation and those of grouping. In the analysis presented here, the isolated segments can be arranged into a block-matrix representation.

Another limitation of [8] is that it descends from the methods developed in [9]. In [9], graphs of the same size can only be compared. Accordingly, in the present analysis, graphs of different sizes are generated for comparison in one-to-one correspondence with the pertinent block matrix.

One more discrepancy with the work of [8] is that the point-proximity matrix is analyzed and its eigenvectors are considered, i.e., as in [10], after [11]; as a result, only the Gaussian-weighted distances between points can be calculated.

The various point sets are compared based on the patterns of eigenvectors corresponding to different sequences. This comparison is carried out by juxtaposing the immanantal polynomials of the Laplacian matrices of the line connectivity graphs [12]. Local topological spectra are considered after [13].

One attempt to codify white noise within chains relies on time-series expansion, as last shown in [3] for Markov chains. In contrast, in the present paper, the white noise of topological shifts is represented as a matrix.

In [4], the following question is posed: which pieces of information must be added to a model to obtain a Markov representation of random effects? This question is addressed in the present paper.

In contrast, the pairwise Euclidean distance [14] is applied to explain how clusters behave differently under white noise and random effects, after computing the dissimilarity matrix. Consequently, the clusters are modeled as topological Hidden Markov Models, to which the Cameron–Martin distance can be applied. Two methods of fragment sequencing are proposed, based on the new formalism developed in [15]. The first method extends the Jukes-Cantor model following [16], while the second models the fragment directly. Both methods involve constructing the fragment for the states of the probability space by applying the appropriate Morse operators to an initial pairwise sequence [17]. The filtration of the probability space is determined after the definition of Hidden Markov Models of ‘multivariate Gaussian data’ from [18] (i.e., Topological Hidden Markov Models), from which the indications from [19] and those from [20] are applied. The application of these guidelines must be compared with the new techniques developed in [21].

Furthermore, starting from [17], white noise and random effects are included within the analytical representation.

The paper is organized as follows.

In Section 2, the methodology is presented: the analytical tools for fragment sequencing are reviewed, and the theorems on the singular-value decomposition (SVD) of ordered block matrices are recalled.

In Section 3, the new results are presented. The definition of the chain from the contact map in the presence of random effects and white noise is established, and the Topological Hidden Markov Models for the clusters are constructed. Further results in sequencing are presented: the further interrogation from [22] is analytically solved, addressing how to maximize the likelihood of DNA and protein sequences when substitution rates of nucleotides or amino acids vary over time. In particular, the technique substitutes standard phylogenetics programming codes and is also theoretically extended to applications in machine learning.

In Section 4, applications are presented, including the use of Dirichlet forms to analytically obtain a vanishing mean for a multivariate distribution, and the formulation of a paradigm for model construction that is independent of the measure while remaining consistent with [21]). As a result, the question posed in [1], following [2], is addressed.

In Section 6, the Conclusions are presented.

2. Methodology

2.1. Fragment-Sequencing

This section summarizes the issue posed in [1] after [2].

Definition 1. Let a ‘bin’ be a vector of consecutive fragments; more in detail, let $\vec{a} = (a_1, a_2, \dots, a_k)$ be a row vector with k fragments. The bin with one fragment is denoted as $\vec{b} = (b_1, b_2, \dots, b_l)$.

Definition 2. The contact frequency f_{ab} is defined as Equation (1)

$$f_{ab} = \sum_{i=1}^{1=k} \sum_{j=1}^{j=l} e_{a_i b_j} \quad (1)$$

with $e_{a_i b_j}$ the number of ‘ligation’ events between fragments, and, in the present case, between fragment a and fragment b .

The contact map is the matrix whose entries correspond to the pairwise ‘contact frequencies’ between two vectors of bins. The definition of the contact-map matrix is taken from [23] p. 26 is taken as

Definition 3. A contact map is a matrix \mathcal{M} whose entries $m_{ij} = 0$ iff the Euclidean distance between the two elements (of the sequence) i and j is less than or equal to a pre-assigned threshold \bar{t} .

It is worth noting that, in contrast, [1], m_{ij} typically assigns a value of 1 when the distance is less than a given threshold. An alternative definition of a contact map is issued from [24] as

Definition 4. A contact map is a square, symmetrical matrix of pairwise contact of residues;

Accordingly, this definition is well-suited for defining pairwise sequencing.

The analysis from [1] is here reappraised.

Let \vec{p} be the row vector of m consecutive bins, i.e.,

$$\vec{p} = (p_1, p_2, \dots, p_m).$$

Let \vec{q} be the row vector of n consecutive bins, i.e.,

$$\vec{q} = (q_1, q_2, \dots, q_n).$$

Clearly, vectors p and q have different dimensions.

This is a characteristic of rectangular matrices, which are used for fragment sequencing within clusters.

Definition 5. \hat{A} is the contact-map (matrix) whose entries are given in Equation (2)

$$a_{ij} \equiv f_{p_j q_i} \quad (2)$$

with $1 < i < m$, $1 < j < n$ when all the bins are in \vec{p} and those in \vec{q} .

Definition 6. Cis contact maps are contact maps obtained in the case when all the bins form \vec{p} and those form \vec{q} are from the same fragment.

Definition 7. Trans contact maps are contact maps obtained in the case when all the bins form \vec{p} and those form \vec{q} are not from the same fragment.

The hierarchical structure of clusters is more apparent after the comparison with [1]. With all the above parameterizations in place, the scenario is prepared for the definition of contact networks.

Definition 8. $N_{\vec{p}}(V, \mathcal{E}, w)$ is the contact network of \vec{p} if $V = \{p_1, p_2, \dots, p_m\}$ is the set of nodes, $\mathcal{E} \equiv \{ \{p_i, p_j\} \mid f_{p_i, p_j} > 0 \}$ is the set of edges with f_{p_i, p_j} the contact frequencies between the nodes (bins) with $1 \leq i \leq m$, and w the weighting function, i.e., an application which sends \mathcal{E} in \mathbb{R}^+ as Equation (3)

$$w : \mathcal{E} \rightarrow \mathbb{R}^+; \quad (3)$$

more in detail, the weighting function w assigns to each edge its contact frequencies as in Equation (4)

$$w(p_i, p_j) = f_{p_i, p_j}. \quad (4)$$

Remark 1. It is anticipated that the weighting functions will be generalized.

Definition 9. The cis contact-map matrix $\hat{A}(\text{cis } \vec{p})$ is the adjacency matrix of the contact network $N_{\vec{p}}(V, \mathcal{E}, w)$.

The notions of weighted graphs and those of unweighted graphs can therefore be recalled.

Definition 10. $N_{\vec{p}}(V, \mathcal{E}, w)$ is named an unweighted graph if the weight w is such that $w \rightarrow \mathcal{E} \rightarrow \{1\}$.

Definition 11. $N_{\vec{p}}(V, \mathcal{E}, w)$ is a weighted graph otherwise.

Let $\{C\}$ be the set of clusters in $N_{\vec{p}}(V, \mathcal{E}, w)$; the following definition holds

Definition 12. The set of clusters $\{C\}$ is defined as Equation (5)

$$\{C\} \equiv \{c_l\}_{l=1}^{l=k} \quad (5)$$

with $c_l \subset V, k \geq 1$.

The Shavit-Walker-Lio' (SWL) matrices \hat{B} 's are defined as

- $B(1)$ an $m \times m$ matrix whose entries $b(1)_{ij}$ are as Equations (6a) and (6b)

$$b(1)_{ij} = 1, i, j \in c, c \in C, c \subset V, \quad (6a)$$

$$b(1)_{ij} = 0 \text{ otherwise}; \quad (6b)$$

and

- $B(s)$ an $m \times m$ matrix whose entries $b(s)_{ij}$ are as Equations (7a) and (7b)

$$b(s)_{ij} = 1, i, j \in c, c \in C_{s-1}, c \subset V, \quad (7a)$$

$$b(s)_{ij} = 0 \text{ otherwise}; \quad (7b)$$

If the structure is hierarchical, clusters may be present. For this reason, Hierarchical Block Matrices (HBMs) are sought.

The HBM matrices of N can be studied here.

Definition 13. The HBM block matrix of $N_{\bar{p}}(V, \mathcal{E}, w)$ is a non-negative $m \times m$ matrix \hat{G} whose entries g_{ij} are defined as Equation (8)

$$g_{ij} \equiv \min_s \{ s \mid b_s \text{ }_{ij} = 1 \} \quad (8)$$

with $s \geq 1$.

2.2. Singular-Value Decomposition (SVD) of Ordered Block Matrices: Theorems

From [16], the underlying 'signal' matrix with ordered block structure \hat{D} is considered, i.e., $\hat{D} \in \mathbb{R}^{m \times n}$ that is starting index of J as $m_1 \times n$ to $m_M \times n$.

A rectangular 'observation' matrix \hat{Y} is considered, under the hypothesis that there are M blocks of sizes (m_1, m_2, \dots, m_M) with $\sum_{i=1}^M m_i = m$ and N blocks of sizes (n_1, n_2, \dots, n_N) with $\sum_{j=1}^N n_j = n$.

Without loss of generality, $m \leq n$ is assumed.

The observation matrix \hat{Y} can be decomposed according to the entries Equation (9)

$$y_{ij} = d_{H(i)V(j)} + \eta_{H(i)} + \nu_{V(j)} \quad (9)$$

where $H(i)$ and $V(j)$ are block membership indicators with values $(1, \dots, M)$ and $(1, \dots, N)$.

The column vectors $\eta_{H(i)}$ and $\nu_{v(j)}$ are defined as continuous random variables with vanishing means and standard deviation σ_η and σ_ν , respectively, i.e., they represent white noise.

The underlying constant signal \hat{D} is therefore isolated from the other components in Equation (9).

Let $\hat{J}_{a \times b}$ be an $a \times b$ matrix with all entries equal to 1. The matrices \hat{D} , $\hat{\eta}$ and $\hat{\nu}$ in Equation (9) are written as

$$\hat{D} \equiv \{d_{ij}\}$$

$$\hat{D} = \begin{bmatrix} d_{11}J_{m_1 \times n_1} & \dots & d_{1N}J_{m_1 \times n_N} \\ \dots & \dots & \dots \\ d_{M1}J_{m_M \times n_1} & \dots & d_{n_1 m}J_{m_M \times n_N} \end{bmatrix}$$

$$\hat{\eta} \equiv \{\eta_{ij}\}$$

$$\hat{\eta} = \begin{bmatrix} \eta_1 J_{m_1 \times n} \\ \dots \\ \eta_M J_{m_M \times n} \end{bmatrix}$$

$$\hat{\nu} \equiv \{\nu_{ij}\}$$

$$\hat{D} = \begin{bmatrix} \nu_1 J_{n_1 \times M} \\ \dots \\ \nu_M J_{m_M \times n} \end{bmatrix}$$

3. Results

3.1. New Definition of the Chain from the Contact Map in the Presence of Random Effects and White Noises

The request of [23] that the distances between the contact maps be Euclidean is here accomplished.

The Matrix \hat{Y} from [15] is here rewritten using its singular-value decomposition, as shown in Equation (10)

$$\hat{Y} \equiv \hat{\nu} \hat{\gamma} \hat{\nu}^T \quad (10)$$

The matrix $\hat{\nu}$ and the matrix $\hat{\nu}$ are reduced to vectors: the vector \vec{u} is recovered in the general case; differently, in the present analysis, if \hat{Y} is compact, the metric is Euclidean when that of the vector \vec{v} is Hilbert (while the definition of \vec{u} is not in general Hilbert). In this way, the chain \hat{Q} is defined on a surface endowed with a Hilbert metric. The chain associated with the adjacency matrix corresponds to a topological Markov chain.

The Euclidean distance from [14] is used to define the 'average pairwise distance'.

From [18], the Cameron-Martin metric is defined from $|\gamma|$ as Equation (11)

$$|\gamma| = \sup v^T \gamma : v^T \zeta v \leq 1 \quad (11)$$

On the vector \vec{v} of the singular-value decomposition Equation (10); the vectors \vec{v} and the vector \vec{v} are made to coincide, in the case of fragment sequencing. The role of ζ is that of the variance, from which the covariance is taken, which allows one to specify marginal distributions; the choice of ζ is specified in Theorem 2.

On the vector \vec{v} of the singular-value decomposition Equation (10), the vectors \vec{v} and the vector \vec{v} are made to coincide in the case of fragment sequencing is studied. The role of ς is that of the variance, from which the covariance is taken, which allows one to specify marginal distributions; the choice of ς is specified in Theorem 2.

The hypothesis from [18] on the expectation value and on the variance are kept as Equations (12) and (13)

$$\mathbb{E}(v) = 0, \quad (12)$$

and

$$Var v = v^T \hat{\varsigma} v. \quad (13)$$

3.2. Clusters as Hidden Markov Models

As in [18], the obtained clusters are Hidden Markov Models of ‘multivariate Gaussian data’ when the covariance matrix is fixed. It is the purpose of the next Section to fix the covariance ς .

The mean vector is newly estimated from the log-emission function from [18] in terms of the Baum-Welch backwards probabilities and for the Baum-Welch forward ones and is required to vanish in the form of [15].

In the Baum-Welch algorithm, the likelihood function L is here written from [25] on the λ collection of model parameters. The Baum-Welch backward probabilities $\alpha(t)$ and the Baum-Welch forward probabilities $\beta(t)$ are here considered for the observation of the probability space O as Equations (14a) and (14b)

$$\alpha_t(j) = Pr(O_1, O_2, \dots, O_t, s_t = j \mid \lambda), \quad (14a)$$

$$\beta_t(j) = Pr(O_{t+1}, \dots, O_T, s_t = j \mid \lambda) \quad (14b)$$

The multivariate distribution mean is calculated as Equation (15)

$$\tilde{m}_j = \frac{\sum_{t=1}^{t=T} \alpha_t(j) \beta_t(j) O_t}{\sum_{t=1}^{t=T} \alpha_t(j) \beta_t(j)} \quad (15)$$

It is now newly required to be set to zero, as expressed in Equation (16)

$$\tilde{m}_j \equiv 0. \quad (16)$$

The techniques for analytically obtaining a multivariate mean of zero are discussed in Section 4.1. In the following subsection, the paradigm developed here is shown to generate a Markov process. The covariance matrix is now fixed from [20], and the Topological Hidden Markov Models are newly built.

3.3. The new Topological Hidden Markov Models (of Clusters)

As studied from [20], a Gaussian distribution with vanishing mean is taken. The specification of the variance ς allows one to specify the marginal distributions.

Gaussian probability measures with ‘prescribed marginals’ are defined after the joint probability density \mathcal{P} with marginals $\mathcal{P}_{c_1}, \mathcal{P}_{c_2}, \dots, \mathcal{P}_{c_l}$. The c_l here used are those from the subset Equation (5).

A class of Gaussian measures with prescribed marginals is newly found after [20] as there exists the covariance matrix $\hat{\varsigma}$ and $\hat{\varsigma}$ is unique. The probability space (*states, observation, filter*) is now constructed for the Topological Hidden Markov Models of clusters. The filter is constructed based on the probability function, and the measure of the probability space is subsequently defined.

From [20] p. 139, the random vector field \vec{X} is considered, with a Gaussian distribution and with vanishing mean and positive-definite covariance $\hat{\varsigma}$. The density $p(\vec{X})$ is written as Equation (17)

$$p(\vec{X}) = (2\pi)^{-|C|/2} (\det[\hat{\varsigma}])^{-1/2} e^{-\frac{1}{2} \vec{X}^T \hat{\varsigma}^{-1} \vec{X}}, \quad (17)$$

from which the measure of the probability space is calculated.

Marginal densities $p_{\Gamma}(\vec{X})$ are defined for arbitrary subsets $\bar{\Gamma}$ of \vec{X} . The following Proposition is drawn from Proposition 1 from [20].

Proposition 1. *The zeros of the matrix ς correspond to the definition of conditional independence.*

The following new proposition is newly derived after Proposition 2 from the same source as

Proposition 2. *Let C be the simple graph of vertices $\{c_l\}$ from Equation (5). The vertices of c_l index the Gaussian random variables X .*

The particular cases of the pairwise sequence are now studied for comparison with [18]. The covariance obeys the following:

Definition 14. $\hat{\varsigma}^{-1}(\alpha, \beta) = 0$ implies that the pairwise sequence (α, β) is not in $\mathbb{E}(C)$, and $\alpha \neq \beta$.

The generalized covariance here used is $\hat{\varsigma}(\alpha, \beta)$. The generalized covariance matrix $\hat{\varsigma}$ is defined as in Equation (18)

$$\hat{\varsigma}(\alpha, \beta) = \mathbb{E}(X_{\alpha} X_{\beta}) \quad (18)$$

The following two theorems are recalled in [20] from [26].

Theorem 1. *The covariance $\hat{K}_{scm}(\alpha, \beta)$ is determined as Equation (19)*

$$\hat{K}_{scm}(\alpha, \beta) \equiv \hat{\xi}(\alpha, \beta), \{ \alpha, \beta \} \in \mathcal{E}(C), \text{ or } \alpha \equiv \beta \quad (19)$$

where \hat{K}_{scm} is the sample covariance matrix.

Theorem 2. *The choice is taken from Equation (20)*

$$\hat{\xi}(\alpha, \beta) \equiv \hat{I}, \{ \alpha, \beta \} \notin \mathcal{E}(C), \alpha \neq \beta, \quad (20)$$

being \hat{I} the identity matrix.

The sheaf-cohomology techniques from [17] can be used to model the scenario with noise and random effects.

3.4. Topological Framework of Contact Probabilities

The role of contact probability is inscribed within a topological framework.

In the work of Lieberman-Aiden et al. [27], massive parallel sequencing is used to demonstrate the 3-dimensional features of the complete genome with respect to proximity-based ligation. Open and closed chromatin are shown to exhibit spatial segregation.

As mentioned before, the long-range interactions between chosen pairs of loci are discriminated using Chromosome Conformation Capture (CCC), where spatially constrained ligation plays a crucial role. Hi-C is a methodology based on massive unbiased sequencing, whereas CCC does not permit unbiased genome-wide analysis.

In the work of Kalhor et al. [28], the Tethered Conformation Capture (TCC) is described as a technique for genome-wide mapping of chromatin interactions. Ibidem, the TCC is outlined to enhance the signal-to-noise ratio, which highlights inter-chromosomal interactions; diverse combinations of interactions are hypothesized to be present in cells; 3-dimensional genome-wide structures are highlighted. As a result, the statistical analysis is limited, and chromosomal interactions are investigated only in the human genome.

As discussed, only a few structural aspects that rule the organization of chromatin are nowadays reported to be understood at the genome scale. Various factors limit the understanding of Chromosome Conformation Capture (CCC). Low signal-to-noise ratios in chromosome capture experiments reduce the ability to map low-frequency interactions, and individual structures are now hypothesized to vary across the cell population.

The capture of conformational data into 3-dimensional structural models is now, therefore, an open challenge. Accordingly, theoretical folding models [27]

have been applied to genome-wide conformation capture data.

In the same source, the TCC is treated as a modified conformation capture method, in which a higher signal-to-noise ratio is calculated, allowing analysis of inter-chromosomal interactions. The resulting analysis technique is probabilistic and enables one to describe some features of the genome. As a methodology, massive parallel sequencing is performed, which links the initial contacts to the locations of the paired loci in the genome. The obtained contact maps accurately account for the observed patterns: the results are in accord with [27].

In the work of Misteli [29], a characterization of the genome is depicted. As previously discussed, the organization of the genomic sequence is described as being determined by spatial and temporal factors at three hierarchical scales: the functioning of nuclear properties, higher-order mechanisms associated with the chromosome fiber, and the spatial arrangement of genomes within the cell nucleus. Genome stability is understood to be influenced by these three factors, which also play a role in gene expression.

The three factors discussed above are pivotal for understanding large-scale DNA sequence mapping. Consequently, the cellular mechanisms that determine genome positioning and their effects on genome regulation must be comprehended to complete the sequencing process. The nature of transcription complexes is thus highly dynamic, with their behavior additionally modulated by compartmentalization.

In the work of Branco et al. [30], the compartmentalization processes are analyzed as responsible for gene expression after accounting for chromatin interactions within distal chromatin organization. More specifically, interactions between distal chromatin segments are reported to influence transcriptional regulation. The topology of chromosomes is introduced in [31], where the chromosome topology is also the capability to undergo the nuclear processes.

In the work of Haaf and Schmid [32], the topology of chromosomes is shown to follow several rules that dictate the number of attachment sites on each chromosome. The arrangement of DNA ‘families’ repeats is studied ibidem. The topological structures demonstrate patterns also in evolutionarily distant species. In turn, topological structures help define transcriptional processes. The compartmentalization of the processes regulating transcription is not yet fully understood.

In the work of Zhang et al. [33], the topology of chromosomes is explained as shaping the energy landscapes, which are described in the present paper within Markov Models and are now newly analyzed as apt to

be arranged within the framework of the Markov State Model.

In [33], the energy landscape is also found to exert a backreaction on the three-dimensional genome organization. In the same work, the energy landscape is described using a maximum-entropy approach, which yields a least-biased effective energy landscape.

In the work of Boulos et al. [34], graph theory is applied in the description of the human genome as far as chromatin interaction (HiC) is concerned. The main replication regions are shown to be in correspondence with DNA loci of maximal network centrality; furthermore, these loci are demonstrated to constitute a set of ‘interconnected hubs’ both at the chromosome level and at the scales implied for different chromosomes. The genomic mechanisms of replication and of transcription can be framed within a graph-theoretical organization, which can be exploited to validate the polymer models of the nuclear organization. The DNA sequences are represented as networks, where critical positions are assigned based on a centrality hierarchy that distinguishes among degree centrality, betweenness centrality, and eigenvector centrality. The ranking accounts for the total weights of the incident edges; within this analysis, the degree centrality [35] is a local centrality measure, the betweenness centrality [35] accounts for to which extent a vertex is located between other vertices on the geodesics of the graph (it is here recalled that for these purposes the graph must be positioned on a manifold), and the eigenvector centrality [36] discriminated the vertices which are connected with ‘well-connected’ vertices. In this theoretical framework, the three-dimensional conformations are analyzed by representing genomic loci as vertices on a plane.

In the work of Sexton et al. [37], the contact map is constructed from the *Drosophila* species. More precisely, a high-resolution contact map is written from a modified genome-wide chromosome conformation capture approach. The data analysis is presented as demonstrating that the genome exhibits a linear partition into ‘well-demarcated’ domains, which overlap extensively with active epigenetic marks and repressive ones. Intra-chromosome and inter-chromosome contacts define contact density and clusters.

In the work of Hou et al. [38], the chromosome domains are proven to be defined after epigenetic marks.

In the work of Dixon et al. [39], the 3-dimensional organization of the human genome is summarized. More specifically, megabase-sized local chromatin interaction (“topological”) domains are identified. Moreover, their boundaries are also characterized. Topological domains are assigned a directionality index, which quantifies the degree and type of interaction bias of each genomic region.

A Hidden Markov State Model is then used to identify the biased states, thereby pinpointing the locations of topological domains. As a result, genomic DNA is described as partitioned into spatial modules linked by chromatin segments, with topological boundary regions defined by regions of chromatin disorganization.

4. Discussion

4.1. The Use of Dirichlet Forms for Obtaining a Vanishing Multivariate Distribution Mean Analytically

The use of Dirichlet forms is indicated in [16] and in [40].

The results from [16] are suited for constructing Topological Markov Models from Jukes-Cantor-inspired sequencing, i.e., as issued by [41].

On the other hand, the results from [40] are suited for the analytical solution of Equation (15) and for the implementation of the potential theory for the calculation of the rewards (which is recalled from [42] 2.2); the Cameron-Martin formula is recalled in Section 4.

With regard to the calculation of rewards, the presence of absorbing states within a fragment can be further examined.

From [43] and [44], the definition of vector fields for Dirichlet forms on Markov processes enables the analytical solution of Equation (15). In particular, [43] provides a definition of vector fields on mapping spaces for this purpose. The role of weights can be generalized for machine learning purposes, as in [45].

4.2. Generalized Constructions of the Chain of Fragment Comparison

The method of fragment comparison is described here, following [46].

From [46], the method is developed, in which a particular probability kernel is constructed with the suitable space of probability measures for the definition of the chain; the work [46] is to be implemented with the choice of the likelihood function as from [25] implementation of Equation (15). The results presented here are compatible with the most general construction [46] when a Gaussian distribution with vanishing mean is taken for the definition of the measure, i.e., the filter, of the probability space.

The comparison of the two fragments (x^n) and (y^n) here is accomplished on a chosen probability space $(\Omega, \mathcal{F}, Pr)$ with Pr assumed on normed spaces E_0 and E , respectively.

Let A and B be fixed Borel subsets of E_0 and E , respectively.

X^n is the state space $\sigma\{x_0, x_1, \dots, x_n\}$ defined on a Borel subset with its σ -algebra.

Analogously, Y^n is $\sigma\{y_0, y_1, \dots, y_n\}$. The Hidden Markov Process (x^n) is characterized by a transition kernel $\mathcal{T}(x, dx')$, and the observation process (y^n) is characterized by the transition kernel $\mathcal{T}_1^{x_1}(y, dy')$ as Equations (21a) and (21b)

$$\mathcal{T}\{x_{n+1} \in A \mid X^n, Y^n\} = \mathcal{T}(x_n, A), \quad (21a)$$

$$\mathcal{T}\{y_{n+1} \in B \mid X^n, Y^n\} = \mathcal{T}_1^{x_{n+1}}(y_n, B), \quad (21b)$$

respectively.

The kernels \mathcal{T} are probability measures for the fixed Borel subsets A and B ; \mathcal{T}_1 is chosen as in Equation (22)

$$\mathcal{T}_1^{x'} = \int_B r(x', y, y') \varpi(dy') \quad (22)$$

with $\varpi \in R(E)$, with $R(E)$ being the suitable space of probability measures.

Remark 2. *The case of undirected graphs is here newly remarked to be compatible with denumerable observations: E is covered for $\varpi(y, x', \cdot)$ univoquely.*

The Markov Process is one with Equation (23)

$$\mathcal{T}f(x, y) = \int_{E_0} \int_E f(x', y') \mathcal{T}_1^{x'}(y, dy') \mathcal{T}(x, dx') \quad (23)$$

with f in the ‘space of Borel measurable functions’ on $E_0 \times E$.

The evolution of the system is described after Equation (6) from [46]. It is noted here that the methods based on a Gaussian Markov distribution are applicable; that is, the transition kernels can induce a measure for the filter of the probability space.

The Cameron-Martin space is newly presented in [47]. The Ornstein-Uhlenbeck semigroup mapping is reappraised in [19] to implement the Ornstein-Uhlenbeck process described in [18]. An example of representing the Kantorovich–Rubinstein distance on a centered Gaussian measure over the Borel σ -field is provided in [19] to compute distances between sequences. This example applies to both pairwise and fragment sequencing. In the same work, examples are provided for which the total variation of the norm is shown to be minorizable. Additionally, the existence of a mapping operator of norm 1, as stated in Lemma 2.1 of [48], is recalled; the lemma is restated as follows:

Lemma 1. *The mapping Equation (24)*

$$v(\alpha) = D \int_0^\infty T_t \alpha dt, \quad \alpha \in L^2(X, \mu) \quad (24)$$

is defined with $v : L^2(X, \mu) \rightarrow L^2(X, \mu, H)$.

Lemma 1 allows one to define the probability space for the hierarchical block-matrix Markov shift. In contrast, the singular-value decomposition of experimental data matrices for complex non-Gaussian random variables is presented in [21], and the techniques developed in the present subsection are also applicable. Indeed, the derivation presented here, following [46], is independent of the Gaussian nature of the variables, allowing for the construction of a different chain.

4.3. Applications in Sequencing

The example from [18] can now be analyzed using the paradigms developed in the present paper to unveil the structures of Hidden Topological Markov Models. More specifically, it is demonstrated here that knowledge of the metric allows the definition of differential operators to be applied to matrices, replacing entries where necessary [17]. The metric defines the manifolds on which the graphs reside, from which the edges can be selected; the union of these edges determines the paths that describe the processes.

The question posed in [1] concerns how hierarchical block matrices can encode information about the topology of block-wise segmentation, particularly regarding the topology of neighboring regions. This investigation is motivated, among other reasons, by the tasks proposed in [22].

This section describes how the data can be represented on a topological manifold and specifies the corresponding metric. In the present case, a Hilbert metric will be determined, which defines the probability space of the process. Indeed, from the clarifications in [26], the one-dimensional segmentations are scrutinized, after which the numerical method is implemented, in which the likelihood with respect to the block boundaries is maximized. It is further noted that the likelihood can be maximized analytically.

The use of the Cameron–Martin distance enables the extraction of Markov property models from the set of “Brownian-motion-like” schemes to which the segmentation techniques may correspond.

Following [49], the problem is extended to a locally compact, connected, separable Hausdorff space equipped with a Radon measure. From [16], it is then possible to describe the time evolution of the eigenvalues of the relevant

Markov-property models using the kernels on which the Radon measure is defined; the associated Dirichlet form implies the Bochner formula.

The use of the Hilbert space, which forms an L^2 structure, is proven in [17]; the employment of this space is proven to be needed straightforwardly in the case one takes into account the prescriptions from [50].

The Markov-property models studied here are those that determine the block decomposition of the topological shift. It is noted here that the numerical model proposed in [50] can thus be solved analytically.

4.4. Applications in Machine Learning

In the work of Khan et al. [51], the methods of blockchain are addressed. Protocols for optimizing data availability are considered within the blockchain framework. The roles of blockchain and machine learning are compared, and the use of Hyperledger technology is analyzed; as a result, the integration of machine learning with blockchain distributed ledger technology is examined.

In the work of Khan et al. [52], blockchain-based platforms are applied, thereby addressing the challenges of data fluctuations. In the same work, the use of blockchain to minimize resource consumption is considered.

In the work of Khan et al. [53], the convergence of artificial-intelligence-enabled machine learning techniques, such as artificial neural networks, support vector machines, reinforcement learning, and deep learning, is analyzed. The use of adaptive control, convolutional neural networks, and recurrent neural networks in data processing is compared, and the integration of artificial intelligence with blockchain technology is proposed. In the same work, the application of artificial neural networks for assessing optimization parameters is also discussed.

In the work of Khan et al. [54], the challenge of data retention is addressed, and control systems are examined. The possibility of reshaping data analysis in the context of fog computing is also considered.

In the work of Khan et al. [55], the issue of autonomous decision-making in machine learning is examined. The aim of this study is to assess the balance between the use of artificial neural networks and Particle Swarm Optimization-enabled metaheuristic optimization methods. The hierarchy of automation is understood in terms of the artificial intelligence system, with a specific focus on cloud-native building blocks [56]. In the same work, the control plane functions are examined as being decoupled from the user planes.

In the work of Khan [57], the combination of gamification and general awareness training is explored. Gener-

ative artificial intelligence, in conjunction with gamification, is shown to replace traditional hierarchies. Moreover, generative AI and gamification-based learning and training are used to define a new metric for evaluating learners' progress, with the goal of rewarding game-based learning.

In the work of Khan et al. [58], the focus is on proposing a lightweight middleware proof of elapsed time in blockchains. The concept of a permissioned chain is explained as enabling better single-entity control, and the key aspects of blockchain technology that ensure efficiency through a lightweight topology are summarized. In the same work, the use of multithreading to enhance system scalability is also discussed.

In the work of Khan et al. [59], Deepfake technology is investigated, and a critique of the assessment measures used to evaluate model performance is provided. The features of computational effectiveness and efficiency are described. In the same work, the use of cross-model ledger technology for evaluating Deepfakes across models within resilient systems is proposed as a topic for further investigation.

In the work of Khan et al. [60], blockchains and edge computing are proposed for authentication via a scalable, lightweight system based on Hyperledger Indy. Time latency is reduced through edge computing, and a hybrid cryptographic technique enables system integration. In the same work, the use of a permissioned blockchain is shown to facilitate compliance.

5. Prospective Studies

As an example, in the case of phylogenetic analysis, the likelihood function is specified in [61].

In the present case, the approach from [25] ensures that the newly established paradigm is suitable for machine learning applications, specifically for implementing the Deep Markov Model for fragment sequencing. This follows the construction from pairwise sequencing as indicated in [17], with the application of Morse operators. Additionally, the technique for inserting gaps between residues from [5] and [24] can be further implemented.

The use of cis maps and trans maps was further developed in [62].

The use of pairwise sequencing [63] can be applied to the notion of distances as well [64].

As a comparison with [5–7], the method for sequencing developed in [65] is of linear growth in the length of the sequence.

A comparison with [21] provides an opportunity to inquire about the hypotheses from [20].

6. Conclusions

Hierarchical Block Matrices (HBMs) are employed to study the well-posedness of comparing chromatin contact maps obtained at different resolutions, such as varying bin sizes. The HBM representation offers a structured approach to align, integrate, and compare contact maps while preserving the multiscale organization captured by 3C and Hi-C measurements.

A key outcome is that clustering states can be interpreted as latent states in machine-learning models [66]. The latent structure is encoded by the resulting clusterings and can be naturally represented using Hidden Markov Models. This interpretation supports the use of Deep Markov Models for learning and inference in the presence of heterogeneous biological samples, and it clarifies how experimental noise can be accommodated without compromising the inferred latent organization.

From a methodological standpoint, Dirichlet-form techniques are employed to obtain an analytical control of the limiting behavior of multivariate distributions (in particular, to recover vanishing means under suitable assumptions). This yields a principled regularization framework that is compatible with comparisons based on overlap between contact maps.

Finally, framing contact maps as graphs emphasizes that overlap-based comparison can be extended to a broader analysis of graph structures. This includes comparisons based on distance-based summaries. Within this unified setting, noise and random effects can be consistently handled within a single probabilistic paradigm.

List of Abbreviations

CCC	Chromosome Conformation Capture
DNA	Deoxyribonucleic Acid
HBM	Hierarchical Block Matrices
HiC	High-throughput Chromosome Conformation Capture
NP	Nondeterministic Polynomial
SVD	Singular Value Decomposition
TCC	Tethered Conformation Capture

Author Contributions

The author confirms sole responsibility for the conception, design, literature review, analysis, interpretation, manuscript drafting, critical revisions, and final approval of the article.

Availability of Data and Materials

The data used in this study is available upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Funding

The study did not receive any external funding and was conducted using only institutional resources.

Acknowledgments

The author would like to acknowledge the Sapienza University of Rome for providing research facilities for this study.

AI Declaration

The author confirms that no AI tools were used to generate any content of this manuscript.

References

- [1] Y. Shavit, B. J. Walker, and P. Lio, “Hierarchical block matrices as efficient representations of chromosome topologies and their application for 3C data integration,” *Bioinformatics*, vol. 32, no. 8, pp. 1121–1129, 2016. [CrossRef]
- [2] E. Yaffe and A. Tanay, “Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture,” *Nat. Genet.*, vol. 43, pp. 1059–1065, 2011. [CrossRef]
- [3] C. Francq and M. Roussignol, “On white noises driven by hidden markov chains,” *J. Time Ser. Anal.*, vol. 18, no. 6, pp. 553–578, 1997. [CrossRef]
- [4] A. Alonso, S. Litière, and A. Laenen, “A note on the indeterminacy of the random-effects distribution in hierarchical models,” *Am. Stat.*, vol. 64, no. 4, pp. 318–324, 2010. [CrossRef]
- [5] J. R. Gonzalez, D. A. Pelta, and J. L. Verdegay, “Solving bioinformatics problems by soft computing techniques: Protein structure comparison as example”, in *Intelligent Systems and Technologies: Methods and Applications* (Studies in Computational Intelligence 217), H. N. Teodorescu, J. Watada, and L. C. Jain, Eds., Berlin/Heidelberg, Germany: Springer, 2009, pp. 123–136. [CrossRef]
- [6] A. Caprara, R. Carr, S. Istrail, G. Lancia, and B. Walenz, “1001 optimal PDB structure alignments: Integer programming methods for finding the maximum contact map overlap,” *J. Comput. Biol.*, vol. 11, no. 1, pp. 27–52, 2004. [CrossRef]
- [7] B. Carr, W. Hart, N. Krasnogor, J. Hirst, E. Burke, and J. Smith, “Alignment of protein structures with a memetic evolutionary algorithm,” in *Proc. GECCO 2002: Proc. Genet. Evol. Comput. Conf.*, San Francisco, CA, USA, 2002. [Online]. Available: <https://dl.acm.org/doi/10.5555/2955491.2955676>
- [8] A. Robles-Kelly and E. R. Hancock, “Graph matching using adjacency matrix markov chains,” 2002. [Online]. Available: https://bmva-archive.org.uk/bmvc/2001/papers/109/accepted_109.pdf

- [9] S. Umeyama, “An eigen decomposition approach to weighted graph matching problems,” *IEEE PAMI*, vol. 10, pp. 695–703, 1988. [[CrossRef](#)]
- [10] L. S. Shapiro and J. M. Brady, “A modal approach to feature-based correspondence,” in *Proc. Br. Mach. Vis. Conf. (BMVC91)*, 1991. [[CrossRef](#)]
- [11] G. Scott and H. Longuet-Higgins, “An algorithm for associating the features of two images,” *Proc. R. Soc. Lond.*, vol. 244, no. 1309, pp. 21–26, 1991. [[CrossRef](#)]
- [12] K. Sengupta and K. L. Boyer, “Modelbase partitioning using property matrix spectra,” *Comput. Vis. Image Underst.*, vol. 70, no. 2, pp. 177–196, 1998. [[CrossRef](#)]
- [13] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker, “Indexing using a spectral encoding of topological structure,” in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, Jun. 23–25, 1999. [[CrossRef](#)]
- [14] F. Bavaud, “Euclidean distances, soft and spectral clustering on weighted graphs,” in *Proc. Eur. Conf. Mach. Learn. Knowl. Discov. Databases (ECML PKDD 2010)*, Barcelona, Spain, Sept. 20–24, 2010. [[CrossRef](#)]
- [15] T. Gong, W. Zhang, and Y. Chen, “Uncovering block structures in large rectangular matrices,” *J. Multivar. Anal.*, vol. 198, 2023, Art. no. 105211. [[CrossRef](#)]
- [16] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet Forms and Symmetric Markov Processes* (De Gruyter Studies in Mathematics, vol. 19), Berlin, Germany: De Gruyter, 2010. [[CrossRef](#)]
- [17] O. M. Lecian, “Sheaf cohomology of rectangular-matrix chains to develop deep-machine-learning multiple sequencing,” *Int. J. Topol.*, vol. 1, no. 1, pp. 55–71, 2024. [[CrossRef](#)]
- [18] A. B. Kashlak, P. Loliencar, and G. Heo, “Topological hidden markov models,” *J. Mach. Learn. Res.*, vol. 24, pp. 1–49, 2023. [[View Online](#)]
- [19] G. V. Riabov, “A representation for the Kantorovich–Rubinstein distance on the abstract Wiener space,” *Theory Stoch. Processes*, vol. 21, no. 2, pp. 84–90, 2016. [[View Online](#)]
- [20] T. P. Speed and H. T. Kiiveri, “Gaussian markov distributions over finite graphs,” *Ann. Stat.*, vol. 14, no. 1, pp. 138–150, 1986. [[CrossRef](#)]
- [21] V. B. Kulikov, A. B. Kulikov, and V. P. Khranilov, “The analysis of stochastic properties of the SVD decomposition at approximation of the experimental data,” *Procedia Comput. Sci.*, vol. 103, pp. 11–119, 2017. [[CrossRef](#)]
- [22] M. Levy-Leduc, M. Delattre, T. Mary-Huard, and S. Robin, “Two-dimensional segmentation for analyzing Hi-C data,” *Bioinformatics*, vol. 30, 2014, Art. no. i386. [[CrossRef](#)]
- [23] M. Vassura, L. Margara, P. Di Lena, F. Medri, P. Fariselli, and R. Casadio, “Fault tolerance for large-scale protein 3D reconstruction from contact maps,” in *Proc. Algorithms Bioinform.: 7th Int. Workshop (WABI 2007)*, Philadelphia, PA, USA, Sept. 8–9, 2007. [[CrossRef](#)]
- [24] G. Tradigo, “On the integration of protein contact map predictions,” in *Proc. Br. Mach. Vis. Conf. (BMVC91)*, Albuquerque, NM, USA, Aug. 2–5, 2009, pp. 1–5. [[CrossRef](#)]
- [25] F. Yang, S. Balakrishnan, and M. J. Wainwright, “Statistical and computational guarantees for the baum–welch algorithm,” *J. Mach. Learn. Res.*, vol. 18, pp. 4528–4580, 2017. [[View Online](#)]
- [26] A. P. Dempster, “Covariance selection,” *Biometrics*, vol. 28, pp. 157–175, 1972. [[CrossRef](#)]
- [27] E. Lieberman-Aiden et al., “Comprehensive mapping of long-range interactions reveals folding principles of the human genome,” *Science*, vol. 326, pp. 289–293, 2009. [[CrossRef](#)]
- [28] R. Kalhor, H. Tjong, N. Jayathilaka, F. Alber, and L. Chen, “Solid-phase chromosome conformation capture for structural characterization of genome architectures,” *Nat. Biotechnol.*, vol. 30, pp. 90–98, 2012. [[CrossRef](#)]
- [29] T. Misteli, “Beyond the sequence: Cellular organization of genome function,” *Cell*, vol. 128, pp. 787–800, 2007. [[CrossRef](#)]
- [30] M. R. Branco and A. Pombo, “Chromosome organization: New facts, new models,” *Trends Cell Biol.*, vol. 17, pp. 127–134, 2007. [[CrossRef](#)]
- [31] M. Zegal’o, E. Wiland, and M. Kurpisz, “Topology of chromosomes in somatic cells. Part 1,” *Postep. Hig Med Dosw*, vol. 60, pp. 331–342, 2006. [[View Online](#)]
- [32] T. Haaf and M. Schmid, “Chromosome topology in mammalian interphase nuclei,” *Exp Cell Res.*, vol. 192, no. 2, pp. 325–332, 1991. [[CrossRef](#)]
- [33] B. Zhang and P. G. Wolynes, “Topology, structures, and energy landscapes of human chromosomes,” *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 19, pp. 6062–6067, 2015. [[CrossRef](#)]
- [34] R. E. Boullos, A. Arneodo, P. Jensen, and B. Audit, “Revealing long-range interconnected hubs in human chromatin interaction data using graph theory,” *Phys. Rev. Lett.*, vol. 111, no. 11, 2013, Art. no. 118102. [[CrossRef](#)]
- [35] L. S. Freeman, “Centrality in social networks conceptual clarification,” *Soc. Networks*, vol. 1, 1978, Art. no. 215. [[CrossRef](#)]
- [36] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” *J. Math. Sociol.*, vol. 2, 1972, Art. no. 113. [[CrossRef](#)]
- [37] T. Sexton et al., “Three-dimensional folding and functional organization principles of the *Drosophila* genome,” *Cell*, vol. 148, 2012, Art. no. 458. [[CrossRef](#)]
- [38] C. Hou, L. Li, Z. S. Qin, and V. G. Corces, “Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains,” *Mol. Cell*, vol. 48, 2012, Art. no. 471. [[CrossRef](#)]

- [39] J. R. Dixon et al. “Topological domains in mammalian genomes identified by analysis of chromatin interactions,” *Nature*, vol. 485, 2012, Art. no. 376. [CrossRef]
- [40] S. Albeverio and M. Roeckner, “Classical dirichlet forms on topological vector spaces: Closability and a cameron–Martin formula,” *J. Funct. Anal.*, vol. 88, pp. 395–436, 1990. [CrossRef]
- [41] T. H. Jukes and C. R. Cantor, “Evolution of protein molecules,” in *Mammalian Protein Metabolism*, vol. III, H. N. Munro, Ed., New York, NY, USA: Academic Press, 1969, pp. 21–132. [CrossRef]
- [42] M. Fukushima, *Dirichlet Forms and Markov Processes*. Amsterdam, The Netherlands, Oxford, UK, New York, NY, USA: North-Holland, 1980. [Online]. Available: <https://www.sciencedirect.com/bookseries/north-holland-mathematical-library/vol/23/suppl/C>
- [43] K. D. Elworthy and Z.-M. Ma, “Vector fields on mapping spaces and related Dirichlet forms and diffusion,” *Osaka J. Math.*, vol. 34, pp. 629–651, 1997.
- [44] Z.-M. Ma, M. Roeckner, and T.-S. Zhang, “Approximation of arbitrary Dirichlet processes by Markov chains,” *Annales de l’I.H.P. section B*, vol. 34, no. 1, pp. 1–22, 1998. [CrossRef]
- [45] W. Tansey et al., “Vector-space markov random fields via exponential families,” in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lille, France, Jul. 7–9, 2015, Art. no. 7289472. [Online]. Available: <https://proceedings.mlr.press/v37/tansey15.html>
- [46] G. B. Di Masi and L. Stettner, “Ergodicity of hidden Markov models,” *Math. Control Signals Syst.*, vol. 17, pp. 269–296, 2009. [CrossRef]
- [47] M. Hairer, “An introduction to stochastic PDEs,” 2023, *arXiv:0907.4178*. [CrossRef]
- [48] A. A. Dorogovtsev, O. L. Izyumtseva, G. V. Riabov, and N. Salhi, “Clark formula for local time for one class of gaussian processes,” *Commun. Stoch. Anal.*, vol. 10, no. 2, pp. 195–217, 2016. [CrossRef]
- [49] P. Koskela, N. Shanmugalingam, and Y. Zhou, “Geometry and analysis of dirichlet forms (II),” *J. Funct. Anal.*, vol. 267, no. 7, pp. 2437–2477, 2014. [CrossRef]
- [50] J. Adachi and M. Hasegawa, *MOLPHY Version 2.3 Programs for Molecular Phylogenetics Based on Maximum Likelihood*. Tokyo, Japan: The Institute of Statistical Mathematics, 1996. [Online]. Available: <https://stat.sys.i.kyoto-u.ac.jp/titech/class/doc/csm96.pdf>
- [51] A. A. Khan et al., “BDLT-IoMT-a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity,” *J. Supercomput.*, vol. 81, no. 1, pp. 1–22, 2025. [CrossRef]
- [52] A. A. Khan et al., “BAIoT-EMS: Consortium network for small-medium enterprises management system with blockchain and augmented intelligence of things,” *Eng. Appl. Artif. Intell.*, vol. 141, 2025, Art. no. 109838. [CrossRef]
- [53] A. A. Khan, A. A. Laghari, S. A. Inam, S. Ullah, and L. Nadeem, “A review on artificial intelligence thermal fluids and the integration of energy conservation with blockchain technology,” *Discov. Sustain.*, vol. 6, no. 1, pp. 1–18, 2025. [CrossRef]
- [54] A. A. Khan et al., “Artificial intelligence, internet of things, and blockchain empowering future vehicular developments: A comprehensive multi-hierarchical lifecycle review,” *Human-Centric Inf. Sci.*, vol. 15, 2025, Art. no. 13. [CrossRef]
- [55] A. A. Khan et al., “ORAN-B5G: A next generation open radio access network architecture with machine learning for beyond 5G in industrial 5.0,” *IEEE Trans. Block Commun. Netw.*, vol. 8, pp. 1026–1036, 2024. [CrossRef]
- [56] M. Liyanage, A. Braeken, S. Shahabuddin, and P. Ranaweera, “OpenRAN security: Challenges and opportunities,” *J. Netw. Comput. Appl.*, vol. 214, 2023, Art. no. 103621. [CrossRef]
- [57] A. A. Khan et al., “A cost-effective approach using generative AI and gamification to enhance biomedical treatment and real-time biosensor monitoring,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–16, 2025. [CrossRef]
- [58] A. A. Khan, S. Dhabi, J. Yang, W. Alhakami, S. Bourouis, and L. Yee, “B-LPoET: A middleware lightweight Proof-of-Elapsed Time (PoET) for efficient distributed transaction execution and security on Blockchain using multithreading technology,” *Comput. Electr. Eng.*, vol. 118, 2024, Art. no. 109343. [CrossRef]
- [59] A. A. Khan, A. A. Laghari, S. A. Inam, S. Ullah, M. Shahzad, and D. Syed, “A survey on multimedia-enabled deepfake detection: State-of-the-art tools and techniques, emerging trends, current challenges & limitations, and future directions,” *Discov. Comput.*, vol. 28, no. 1, 2025, Art. no. 48. [CrossRef]
- [60] A. A. Khan et al., “A lightweight scalable hybrid authentication framework for Internet of Medical Things (IoMT) using blockchain hyperledger consortium network with edge computing,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025. [CrossRef]
- [61] M. J. Bishop and E. A. Thompson, “Maximum likelihood alignment of DNA sequences,” *J. Mol. Biol.*, vol. 190, no. 2, 1986, Art. no. 159. [CrossRef]
- [62] A. Miele and J. Dekker, “Mapping cis- and trans-chromatin interaction networks using chromosome conformation capture (3C),” in *The Nucleus (Methods in Molecular Biology, vol. 464)*, R. Hancock, Ed., Totowa, NJ, USA: Humana Press, 2008. [CrossRef]
- [63] Z. Yang and S. Kumar, “Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites,” *Mol. Biol. Evol.*, vol. 13, no. 5, 1996, Art. no. 650. [CrossRef]
- [64] M. M. Deng, C. Yu, Q. Liang, R.L. He, and S.S.-T. Yau, “A novel method of characterizing genetic sequences: Genome space with biological distance and applications,” *PLoS ONE*, vol. 6, no. 3, 2011. [CrossRef]

-
- [65] W. Gong and X.-Q. Fan, “A geometric characterization of DNA sequence,” *Phys. A: Stat. Mech. Its Appl.*, vol. 527, 2019, Art. no. 121429. [[CrossRef](#)]
- [66] A. D. Schmitt et al., “A compendium of chromatin contact maps reveals spatially active regions in the human genome,” *Cell Rep.*, vol. 17, no. 8, pp. 2042–2059, 2016. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the publisher or editors. The publisher and editors assume no responsibility for any injury or damage resulting from the use of information contained herein.