Automatic Homology Computation
with Application to Pattern Classification

By

Marcio GAMEIRO* and Paweł PILARCZYK**

Abstract

We briefly introduce the approach to homology computation based on rectangular grids, as opposed to the usual approach based on simplices, and we explain how homology, this important topological invariant, can be effectively computed in an algorithmic way. We also point out some specific applications of this method of homology computation to the classification of patterns that appear in numerical simulations of PDEs or come from physical experiments.

§ 1.   Introduction

In numerical simulations or physical experiments, complicated spatial-temporal patterns may appear, and their complexity needs to be assessed. This gives rise to the question of finding some computable quantitative measures for their complexity. In this paper we discuss the application of algebraic topology for this purpose.

We begin with a brief introduction to the computational homology which we use to quantify the topological pattern complexity. Then, in the next section, we discuss some applications of this algorithmic tool to the classification of patterns.

§ 2.   Cubical Homology

The computation of homology of a topological space is based on its decomposition into a cellular complex, so we begin with the definition and illustrations of cellular complexes based on rectangular grids (see [5] for more details).

Let $\mathcal{H}$ denote the set of all the n-dimensional hypercubes in $\mathbb{R}^n$ of edge length 1 and vertices in $\mathbb{Z}^n$. Elements of $\mathcal{H}$ are called *full cubes*. Any intersection of two full cubes is called an *elementary cube* (or *cube* for short). It is easy to see that an elementary cube is a product of intervals of length 1 or 0, with integral endpoints, that is,

$$Q = I_1 \times I_2 \times \cdots \times I_n$$

for $I_i = \{k_i\}$ or $I_i = [k_i, k_i + 1]$, where $k_i \in \mathbb{Z}$. Note that a full cube is a particular case of an elementary cube. The *dimension* of an elementary cube is the number of intervals with length 1 in the formula above, that is, $\dim Q := \operatorname{card}\{i : |I_i| > 0\}$.

If $\mathcal{A}$ is a finite set of (full) cubes, then $|\mathcal{A}| := \bigcup_{Q \in \mathcal{A}} Q$ (the union of these cubes) is called a *(full) cubical set*. Note that (full) cubical sets are always compact polyhedra.

If A is a cubical set, then the group of q-dimensional *chains* $C_q(A)$ is defined as the free abelian group generated by all the q-dimensional elementary cubes contained in A. Note that $C_q(A) \cong 0$ if there are no q-dimensional elementary cubes contained in A. In particular, this holds true for all $q < 0$ and $q > n$. The (algebraic) generator of C_q corresponding to the elementary cube $Q \subset A$ is denoted by $\widehat{Q}$. With this definition, each chain can be perceived as a formal combination of elementary cubes with integral coefficients.
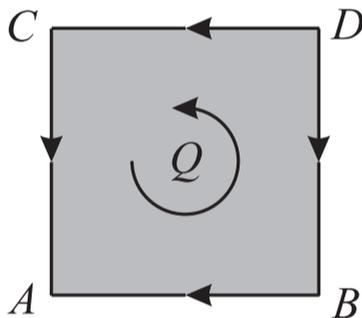

Figure 1. A 2-dimensional rectangular cell with its orientation indicated, as well as orientation of its faces. $\partial(\widehat{Q}) = -\widehat{BA} - \widehat{BD} + \widehat{DC} + \widehat{CA}$, $\partial(\widehat{BA}) = \widehat{B} - \widehat{A}$, $\partial(\widehat{A}) = 0$, etc.

The *boundary operator* is defined on the generators $\widehat{Q}$ of C_q in the following way (compare with Figure 1). Assume $Q = I_1 \times \cdots \times I_n \subset \mathbb{R}^n$ and $d = \dim Q$. Let the non-degenerate intervals in the product that defines Q be denoted by $I_{j_1} = [k_{j_1}, k_{j_1} + 1], \ldots, I_{j_d} = [k_{j_d}, k_{j_d} + 1]$. The left-hand-side and the right-hand-side $(d-1)$-dimensional face of Q in the direction m is given by the following formulas, respectively:

$$Q_m^- := I_1 \times \cdots \times [k_{j_m}, k_{j_m}] \times \cdots \times I_n$$
$$Q_m^+ := I_1 \times \cdots \times [k_{j_m} + 1, k_{j_m} + 1] \times \cdots \times I_n$$

With this notation, the boundary operator is defined on generators of C_q as follows:

$$\partial(\widehat{Q}) = \sum_{m=1}^{\dim Q} (-1)^m \left(\widehat{Q_m^+} - \widehat{Q_m^-}\right)$$

and then this definition is extended to a homomorphism $C_q \to C_{q-1}$.
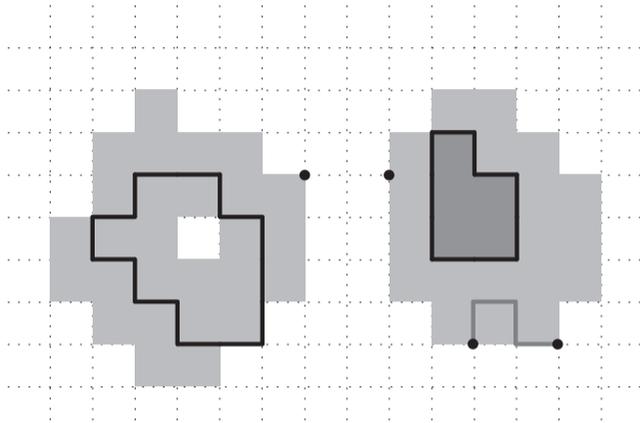

Figure 2. Geometric illustration of cycles: A 1-dimensional cycle can be understood as a closed loop; the loop at the right-hand side is a boundary of the shaded area; the loop at the left-hand side is not a boundary of any chain, so it gives rise to a homology generator. The two points at the right-hand side can be joined by a 1-dimensional chain, so they differ by a boundary, and they are equivalent in homology; the two points in the middle, however, correspond to different homology generators.

The important property of the boundary operator is that $\partial \circ \partial \cong 0$. This allows one to define the homology groups of A in the following way. Let $B_q(A) := \operatorname{im} \partial_{q+1} \subset C_q$ be the group of q-dimensional *boundaries* in A, and $Z_q(A) := \ker \partial_q \subset C_q$ be the group of q-dimensional *cycles* in A. Then the quotient group $H_q(A) := Z_q(A)/B_q(A)$ is well defined and it is called the q-th *homology group* of A. It is a finitely generated commutative group. It is a well known fact from algebra (see [5]) that every such group has a simple canonical form

$$G \cong \mathbb{Z}_{p_1^{n_1}} \oplus \cdots \oplus \mathbb{Z}_{p_k^{n_k}} \oplus F,$$

where $p_1, \ldots, p_k$ are prime numbers, $n_1, \ldots, n_k$ are positive integers, and $F = \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}$ is a free group. This factorization is unique (up to the order of factors). In particular, the rank of F in the factorization of $H_q(A)$ is denoted $\beta_q(A)$ and is called the q-th *Betti number* of A.

The definition of homology as the quotient group suggests the geometric interpretation of homology as a means of counting those cycles which are not boundaries of any

other cube or chain (see Figure 2). In general, the Betti numbers reflect the topological complexity of a cubical set. More precisely, β_0 equals the number of connected components, β_1 indicates the number of holes (tunnels in 3D), and β_2 represents the number of cavities (see Figure 3).
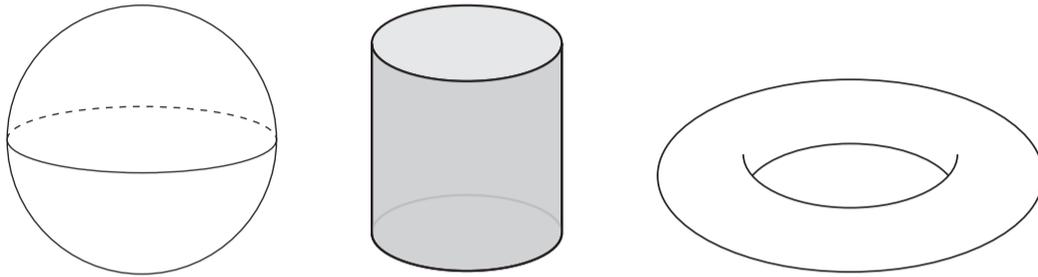

Figure 3. Betti numbers count holes. For the sphere, $\beta_0 = 1$, $\beta_1 = 0$, and $\beta_2 = 1$. For the cylinder with open ends, $\beta_0 = 1$, $\beta_1 = 1$, and $\beta_2 = 0$. For the torus surface, $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$.

Given geometric patterns in terms of bitmap images, independent of whether they come from numerical simulations, or from photographs taken from physical experiments, one can treat these images as full cubical sets if each pixel is identified with a 2-dimensional square, as illustrated in Figure 4. The same holds true for higher-dimensional bitmap data, in particular, each voxel corresponds to a 3-dimensional cube.
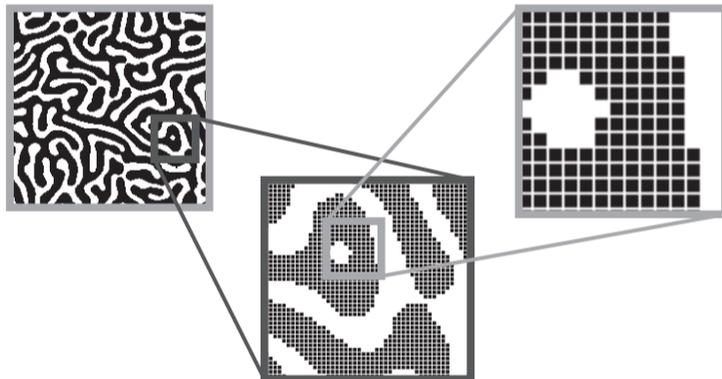

Figure 4. Pixels of a 2-dimensional computer image give rise to 2-dimensional full cubes (squares) in $\mathbb{R}^2$, as seen at the magnification of a sample black-and-white bitmap image, and together they form a cubical set.

Since the algebraic computations that follow the definition of the homology groups directly are very time-consuming, advanced algorithms based on the transformation of the matrix of the boundary homomorphism into the Smith Normal Form are used in the actual software [1]. Moreover, prior to the algebraic computations, the cubical sets

to be processed are reduced in a geometric way. The general idea of one type of the reductions is to remove cubes from the set of cubes, provided this removal does not change the homology of the corresponding cubical set. Another idea is called *free face collapses*, and its key is to remove pairs of elementary cubes such that one cube is a face of the other cube, and only of that cube. Figure 5 illustrates such cases. The reader is referred to [7, 8] for more details on these reductions. The software [1] we use in this research does all these reductions automatically to minimize the amount of algebraic data that has to be processed.

Figure 5. Some full cubes that can be removed from the cubical set are shaded dark at the left-hand side. Free faces are indicated with arrows at the right-hand side.

§ 3.   Applications to Pattern Classification

The Computational Homology software [1] described above has been applied to patterns arising in numerical simulations of the Cahn-Hilliard Equation [4] and numerical simulations of Spiral Wave Patterns [3], as well as in experiments of Spiral Defect Chaos in Rayleigh-Benard Convection [6].

In this paper we focus on the application of computational homology to spiral wave patterns arising in simulations of the FitzHugh-Nagumo system

(1)
$$u_t = \Delta u + \epsilon^{-1} u(1-u)(u - \tfrac{v+\gamma}{\alpha})$$
$$v_t = u^3 - v.$$

with Neumann boundary conditions on the rectangular domain $\Omega = [0, 80] \times [0, 80]$. We study (1) at $\alpha = 0.75$, $\gamma = 0.06$ and we vary the parameter ϵ. Following [3] and references therein, we produce patterns by thresholding the solutions as indicated in Figure 6. Since we want to understand the time evolution of the time dependent excited media

$$E(t) = \{x \in \Omega \mid u(x, t) \geq 0.9\},$$

we create "movies" of the excited media, that is, we consider the subsets of $\Omega \times [0, \tau]$ of the form

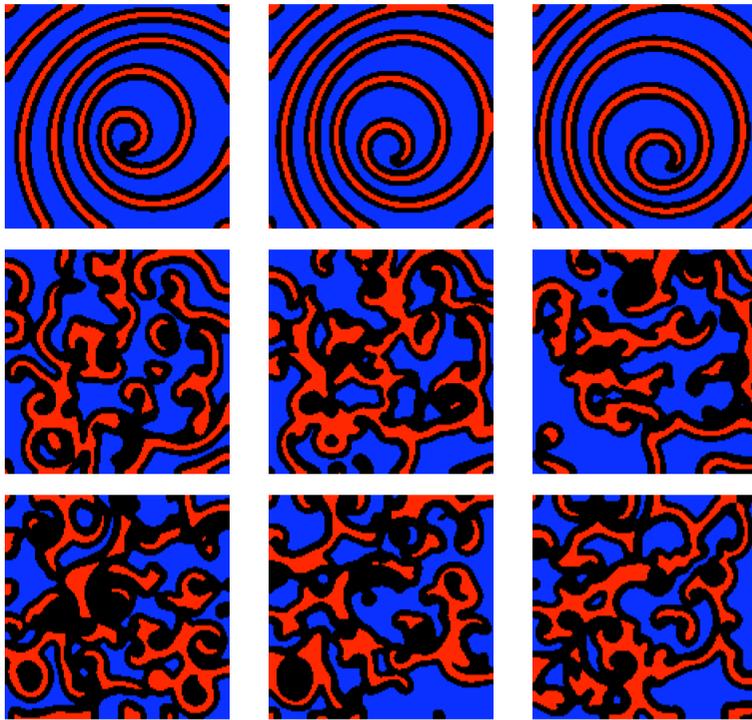$$M(\tau_1, \tau_2) := \bigcup_{t \in [\tau_1, \tau_2]} E(t),$$

Figure 6. (Color online) Patterns generated by thresholding the solutions of (1) as follows: the light gray (red) region corresponds to excited points ($u \geq 0.9$), dark gray (blue) to the quiescent region ($u \leq 0.1$), and black to the reaction zone ($0.1 < u < 0.9$). The top row refers to $1/\epsilon = 14.0$, the middle row to $1/\epsilon = 12.0$ and the bottom row to $1/\epsilon = 11.5$. Each row shows different snapshots of the same solution.

and compute the homology groups of these 3-dimensional sets. Since these sets are embedded in $\mathbb{R}^3$, their homology is completely characterized by their Betti numbers. For each value of ϵ we compute the Betti numbers $\beta_i(n, \epsilon)$ of $M\big(10(n-1), 10(n-1)+1000\big)$ for $n = 1, 2, \ldots, 10000$, thus producing time series of Betti numbers

$$B_i(\epsilon) := \{\beta_i(n, \epsilon) \mid n = 1, 2, \ldots, 10000\}, \quad \text{for} \quad i = 0, 1, 2.$$

As already pointed out in [3], $\beta_2(n, \epsilon) \equiv 0$, and $\beta_0(n, \epsilon)$ is piecewise constant and takes on fairly small values. The time series $B_1(\epsilon)$, however, proved to be quite interesting. Figure 7 shows plots of some time series. Notice that we can see a clear difference in the plots of β_1 for the two time series shown in Figure 7, thus it seems that the Betti numbers can be used to detect the fact that the corresponding patterns originated from different parameter values. Notice also that we cannot see this difference so clearly on the corresponding patterns in Figure 6.

In fact, as it was shown in [3], the mean value of the time series of β_1 is essentially a monotonically increasing function of ϵ. The same holds true for the time series of β_0,
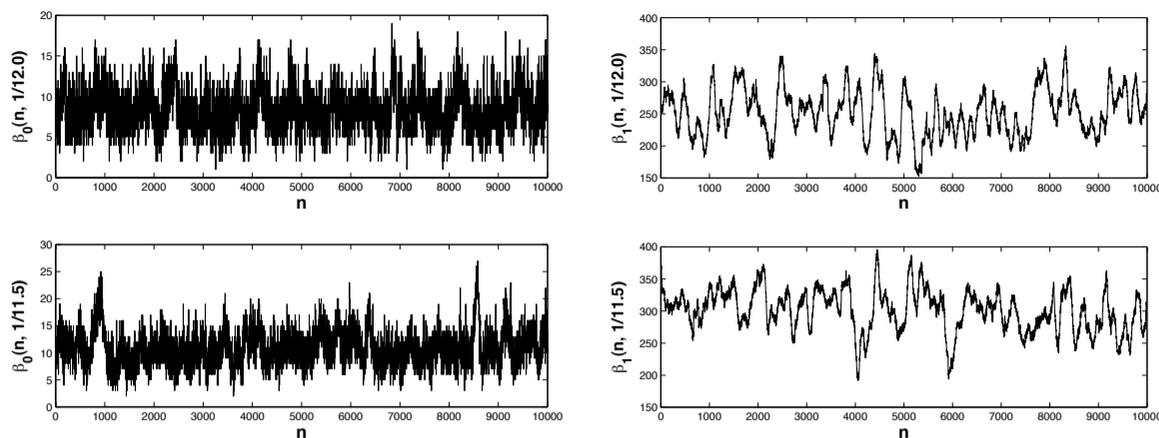
Figure 7. Time series of Betti numbers. The top row refers to $1/\epsilon = 12.0$, and the bottom row to $1/\epsilon = 11.5$. The first column shows β_0, and the second shows β_1. The time series corresponding to $1/\epsilon = 14.0$ are not plotted because they are piecewise constant and take on a limited number of small values.
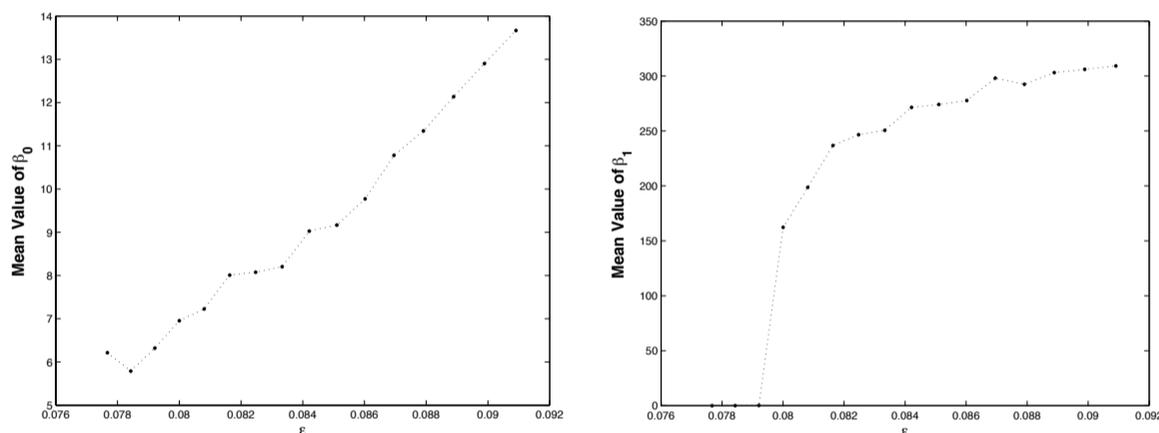

Figure 8. Mean values of the time series of β_0 and β_1 as a function of ϵ.

as it is shown in Figure 8. Therefore, it seems that it may be possible to use the Betti numbers to characterize the parameter values from the patterns.

In [3] the complexity of the patterns was characterized by means of Lyapunov exponents, computed from the time series of β_1. In this paper we try to assess the complexity by means of entropy, computed as follows: For each value of ϵ, we describe the states of the system by vectors of the form (β_0, β_1), and then we count how many times a given state (β_0, β_1) appears on the time series to compute the probability of that state. Finally, we compute the entropy at the given value of ϵ by

$$E(\epsilon) = \sum_i p_i \log(1/p_i),$$

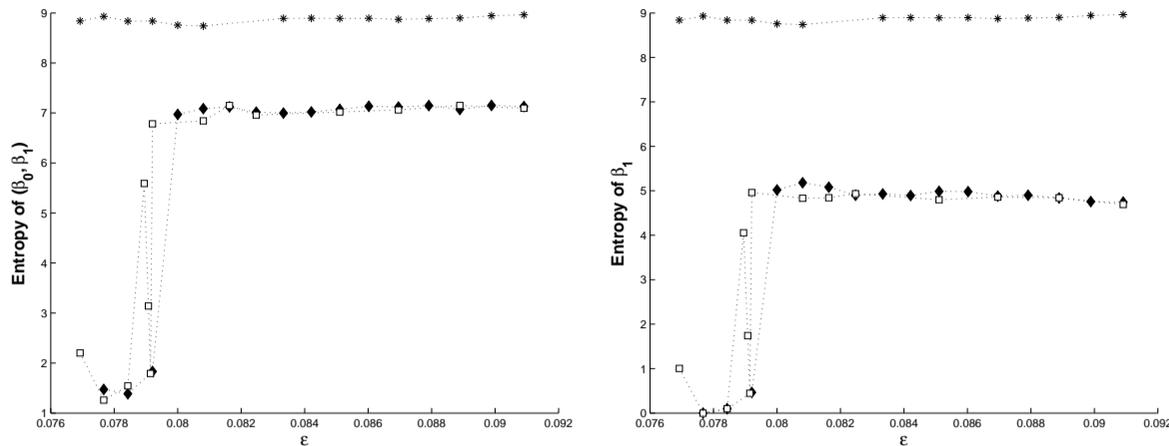where p_i denotes the probability of being in the state i, and the index i spans over all

Figure 9. Entropy as a function of ϵ. In both plots, the stars show the entropy computed directly from the solution u of the PDE. The diamonds and the squares are the entropies of time series of Betti numbers computed from patterns generated by solving the PDE with two different initial conditions. The entropies indicated by the diamonds and the squares on the left plot were computed using (β_0, β_1) to describe the states of the systems, while only β_1 was used on the right plot.
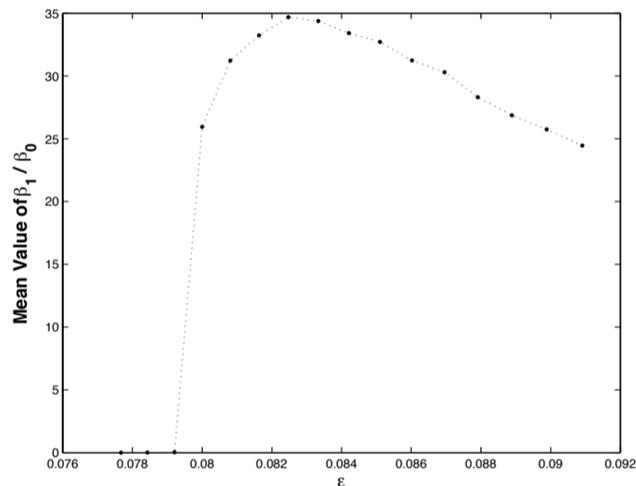

Figure 10. Mean values of the time series of β_1/β_0 as a function of ϵ. This gives basically the average number of tunnels per component, except that we are not excluding small components.

the different states attained by the system. The left plot of Figure 9 shows the entropy computed as described above. Since in [3] only β_1 was used to compute the Lyapunov exponents, for the sake comparison, we also computed the entropy as described above, but using only β_1 instead of (β_0, β_1) to characterize the states attained by the system. The result is shown in the right plot of Figure 9. As in [3], to check our results, we computed the entropy directly from the solution values of the PDE in the following way:

We picked up a point x in the domain Ω, and computed the solution u at this point for 10000 time steps, creating the time series $\{u(x, t_k) \mid k = 0, 1, \ldots, 10000\}$. We then used this time series to compute the entropy, where the states of the system are now characterized by the values of $u(x, t_k)$, and two values are considered to describe the same state if they agree to at least the first five decimal digits. This result is also shown in Figure 9. While the entropy is computed over the states attained by the system in a finite amount of time, its value seems to asymptote to a constant for long time series.

The positive entropy values confirm the complicated spatio-temporal dynamics characterized in [3] by positive Lyapunov exponents. Notice that the Betti numbers capture the spatio-temporal complexity of the patterns, while the solution at a point $x \in \Omega$ only measures the temporal complexity. This, added to the fact that the solutions are not spatially complex for small values of ϵ (see Figure 6), may explain why the entropies from Betti numbers and from the solution u behave differently for small values of ϵ.

Another interesting fact is that the mean values of Betti numbers are increasing as a function of ϵ, while the entropy is essentially constant (except for small ϵ). We believe that this is due to the fact that, as ϵ increases, the number of components and the total number of tunnels increase, but the number of tunnels per component remains basically constant. To see this, in Figure 10 we plotted the mean values of the time series of β_1/β_0. The reason the graph is not quite constant is the existence of small components with no tunnels. If we take the quotient of β_1 by the number of components that contain tunnels (the average number of tunnels per component), then the result would be fairly constant.

§ 4. Final Remarks and Conclusion

As the results illustrated above show, the homology computation proves to be a novel and effective tool that gives a computable means for the quantitative measurement of the topological complexity of patterns.

It is worth mentioning again that the software used in this research is dimension independent. This feature allows one to analyze higher-dimensional dynamics, as well as include the time as an extra dimension to assess the complexity of patterns not only at specific time snapshots, but also their evolution in time, like it was done in this paper.

Although the currently available software is highly efficient for low dimensions, the significantly larger cost of homology computation in dimensions four or higher gives a good motivation for further development of this powerful tool.

References

[1] Computational Homology Project, `http://chomp.rutgers.edu/`.

[2] Computer Assisted Proofs in Dynamics Group, `http://capd.wsb-nlu.edu.pl/`.

[3] M. GAMEIRO, K. MISCHAIKOW AND W. KALIES, *Topological Characterization of Spatial-Temporal Chaos*, Physical Review E **70**, no. 3, 035203, (2004).

[4] M. GAMEIRO, K. MISCHAIKOW AND T. WANNER, *Evolution of Pattern Complexity in the Cahn-Hilliard Theory of Phase Separation*, Acta Materialia **53**, no. 3, 693-704, (2005).

[5] T. KACZYNSKI, K. MISCHAIKOW AND M. MROZEK, *Computational homology*, Applied Mathematical Sciences **157**, Springer-Verlag, New York, (2004).

[6] K. KRISHAN, M. GAMEIRO, K. MISCHAIKOW AND M. SCHATZ, *Homological Characterization of Spiral Defect Chaos in Rayleigh-Benard Convection*, submitted, (2005).

[7] K. MISCHAIKOW, M. MROZEK, P. PILARCZYK, *Graph approach to the computation of the homology of continuous maps*, Foundations of Computational Mathematics, Vol. 5, No. 2, 199–229, (2005).

[8] P. PILARCZYK, *Computer assisted method for proving existence of periodic orbits*, Topol. Methods Nonlinear Anal. **13**, no. 2, 365–377, (1999).

[9] E.H. SPANIER, *Algebraic topology*, McGraw-Hill, (1966).