

# Defining Emergent Descriptions by Information Preservation

**Daniel Polani**

Algorithms Group/Adaptive Systems Research Group  
Dept. of Computer Science  
University of Hertfordshire, UK  
d.polani@herts.ac.uk

We propose a formalized approach for the characterization of the phenomenon of *emergence*, based on information-theoretic criteria. The proposed mechanism of description fits in well with existing approaches for the characterization of complex systems and also has ramifications towards existing algebraic models for coordinatizations of complex systems.

## 1.1 Introduction

Among the most fascinating phenomena in the field of complex systems and Artificial Life is the phenomenon of *emergence*. Informally, one speaks of emergence if some phenomenon occurs in a complex system that is not “obviously” deducible from the microdynamics of the system. Here, we must emphasize, the crucial notion is the word “obviously”. Generally, given a specific scenario, human observers will, on an intuitive level, usually agree whether emergence is present or absent in that system. It would be desirable to achieve a more objectifiable access to these phenomena and to move away from the implicit and subjective identification of emergence by human observers to a mathematically

accessible and operationally relevant characterization. However, the intuitive “emergence detection” capability of humans seems to be grounded in a “gestalt”-like perception of the phenomena in question, and the phenomenon itself turns out to offer a surprising degree of resistance to formalization.

Consequently, not a unique, formal and operational notion of emergence has “emerged”. Some approaches characterize emergence informally as “surprising” effects [19] which gain their existence only in “the eye of the beholder” [12], thus making the human observer mentioned above an indispensable ingredient in the characterization of the phenomenon.

On the other end of the spectrum, one finds highly formalized approaches, using the powerful language of category theory. Among these are [16], and, in particular, the approach developed in [18]. One of the problems remaining with definitions based on meta-models like category theory is that, while precise, in practice they are often computationally inaccessible and they make it difficult to allow for a “natural” concept of emergence to arise from the intrinsic structure of a system.

What one would like to have is a formalized concept of emergence which would be able to capture essential aspects of the “gestalt”-like detection of emergence phenomena by humans, which would be operational and thus computationally relevant and useable and which, ideally, would be forged out of the dynamics of the given system without having to resort to additional assumptions not present in the original system under consideration.

In the present paper, we propose to characterize emergence via the formal concept of *emergent description* as a decomposition of a stochastic dynamic system into approximately independent subsystems which, individually, preserve information (in sense of Shannon’s information theory) as much as possible. The following chapters will develop the rationale for this approach, introduce the formalism and give a discussion.

## 1.2 Motivation and Related Work

### 1.2.1 An Instructive Example: Collective Modes in Crystal Lattices

One of the motivations of our approach is the natural decomposition of the collective dynamics in crystal lattices into individual oscillatory modes, *phonons* [5]. This phenomenon is particularly instructive because it demonstrates, on the one hand, how “holistic”, “gestalt”-like phenomena can arise (“emerge”) in a well-understood way as collective modes from the natural dynamics of a dynamical system. Furthermore, if symmetries are present, they even pre-include an a priori dynamics decomposition structure in a universal fashion via a generalized Fourier decomposition [3, 20]. This shows how additional aspects of structure can be seamlessly worked into a definition of a canonical partition into subsystems. This case is, however, limited by the fact that it requires the systems to be linear.

### 1.2.2 Dynamics Decomposition in Nonlinear Systems

The field of synergetics develops an approach to decompose also nonlinear systems in a natural fashion [10]. Here, the natural decomposition of dynamical systems near fixed points into stable, central and unstable manifolds is reinterpreted in a heuristic way, separating *fast foliations* and *slow manifolds* in the system [14]. Slow degrees of freedom (*master modes*) of the system dynamics are said to “enslave” the fast degrees of freedom (*slave modes*), because they essentially dominate the dynamics of the system. The *master modes* can be construed as “emerging” from the system dynamics. This decomposition into individual subsystems arises naturally from the system dynamics and is therefore not due to a subjective observer as in [12]. The synergetics decomposition typically requires the systems to exhibit a splitting into different time scales. However, it turns out that under certain conditions it is even possible to decompose nonlinear (even chaotic) dynamical systems canonically into weakly coupled subsystems if they have no separate time scales [23].

It is well-known that dynamical systems can be well described using information theory using the Chaitin-Kolmogorov Entropy (e.g. [8]) measuring the information flow from microstates towards macrostates. Here, the dynamics of the system is translated into the information-theoretical language via an a priori choice of the space partition. However, it would be desirable to have an approach where the decomposition “emerges” naturally from the structure of the dynamical system and is not imposed upon it<sup>1</sup>.

Information-theoretic principles have been used for the characterization of complex systems [7, 24, 4]. In [1, 2], a notion of complexity is introduced as the *persistent* information captured in an individuals’ genome during an evolution process. This shows significant parallels to the synergetics view that persistence of certain aspects of a system description over time can be important; however, it is felt that the information-theoretic formulation of synergetics has not exploited the persistence paradigm to its full potential [11].

In the crystal-lattice example or the synergetics model, the “persistence” idea is incarnated as the decomposition of the system into modes. Defining such a decomposition of the system dynamics for a general nonlinear system is non-trivial. Heuristic approaches can be used for this purpose, stemming either from nonlinear dynamics analysis [17], or clustering methods [25]. The constructive creation of a stable subdynamics (as opposed to a decomposition of an existing system) has been addressed in the context of the “homeokinesis” [9].

## 1.3 Emergent Descriptions

These developments formed the motivation for our present approach to introduce a formal characterization of emergence. It is based on information theory, is fully formal and operational in the sense that it can – in principle – be calculated

---

<sup>1</sup>In this context, studying the relation between the notion developed in Sec. 1.3.2 and the concept of generating partitions from nonlinear dynamics would be of particular interest.

for any given stochastic dynamical system. The present section introduces the concept, in Sec. 1.4 we will indicate how the stringency of our approach can be relaxed to become relevant for the study of practically relevant cases.

### 1.3.1 Concept Overview

We will define emergent descriptions as a *complete* decomposition of the system into *independent* subsystems which are *individually predictable*.

### 1.3.2 Strong Emergent Descriptions: Formal Definition

From above considerations, we see that several aspects have to be incorporated to attain a notion of emergence: an aspect of temporal persistence, a notion of partition of the system states into roughly independent subsystem states (this partition may be “holistic” in the sense that the states of each subsystem needs to encompass a state description of all constituents of the total system and thus may be “oblique” with respect of possible system constituents). Also, the notion should be capable to deal naturally with inherently nonlinear systems.

The approach we use is to mimic information-theoretically the decomposition of linear dynamical systems into independent subsystems, *eigenmodes*, whose individual dynamics, while induced by the global dynamics, can be considered separately and independently of each other. We interpret these subsystems as “emerging” from the global system dynamics.

First, we introduce some notation. Given a random variable  $X$  (uppercase), the set of its possible values will be denoted by (calligraphic)  $\mathcal{X}$  and a concrete value for  $X$  by (lowercase)  $x \in \mathcal{X}$ . Write  $P(X = x)$  for the probability that  $X$  assumes a value  $x$  or, if no confusion is possible,  $p(x)$  by abuse of notation. Similarly, write  $p(y|x)$  for the conditional probability  $P(Y = y | X = x)$  when no confusion is expected.

We introduce probabilistic maps as a generalization of deterministic maps. Let event sets  $\mathcal{X}, \mathcal{Y}$ , and a conditional probability  $p_M(y|x)$  be given. The probabilistic map  $M$  from the event set  $\mathcal{X}$  to an event set  $\mathcal{Y}$  maps any random variable  $X \in \mathcal{X}$  to a random variable  $M(X) \in \mathcal{Y}$ , such that  $P(MX = y | X = x) = p_M(y|x)$  and, conditioned on  $X$ , the variable  $M(X)$  is independent of any other random variables in the model, i.e.  $M(X)$  has a Markovian dependence on  $X$ .

For two random variables  $X$  and  $Y$ , let  $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$  denote the entropy of  $X$  and  $I(X; Y) = H(X) + H(Y) - H(X, Y)$  denote the mutual information between  $X$  and  $Y$ . For details of the properties of these notions, see e.g. [6]. We proceed now to define strong emergent descriptions.

**Definition 1.3.1 (Strong Emergent Description)** *Let  $\mathcal{S}$  be a (for sake of technical simplicity) finite set of states. Further let on  $\mathcal{S}$  a stochastic dynamics be recursively defined via a transition probability  $p(s(t+1|s(t))) \equiv P(S(t+1) = s(t+1) | S(t) = s(t))$  ( $t$  may be limited to a finite range  $\{0 \dots T\}$ ) and an initial probability distribution  $p(s(0))$ , thus defining a Markovian process with initial conditions.  $\mathcal{S}$  together with the dynamics defined by the Markov process forms*

a stochastic dynamic system which can be characterized by a (possibly finite) sequence  $S(0), \dots, S(t), \dots$  of random variables with values in  $\mathcal{S}$ .

A strong emergent description is then defined as a sequence of probabilistic maps (projections)  $Q_t : \mathcal{S} \rightarrow \hat{\mathcal{S}}^{(0)} \times \hat{\mathcal{S}}^{(1)} \times \dots \times \hat{\mathcal{S}}^{(k)}$ , inducing subsystems such that for all  $t \in \{0 \dots T\}$

1. the totality of  $Q_t = \hat{Q}_t^{(1)} \times \dots \times \hat{Q}_t^{(k)}$  is information conserving with respect to the total state, i.e.

$$H(S(t)) = I(S(t); Q_t(S(t))) ; \quad (1.1)$$

2. the dynamics of the individual subsystems are independent from each other:

$$\forall i \neq j : I(\hat{Q}_t^{(i)}(S(t)); \hat{Q}_t^{(j)}(S(t))) = 0 ; \quad (1.2)$$

3. and with exception of subsystem 0, the subsystem dynamics is information conserving:

$$\forall i \in \{1 \dots k\} : H(\hat{Q}_{t+1}^{(i)}(S(t+1))) = I(Q_t^{(i)}(S(t)); \hat{Q}_{t+1}^{(i)}(S(t+1))) \quad (1.3)$$

(note that  $\hat{Q}_t^{(i)}(S(t))$  is the subsystem state description). Here, subsystem  $\hat{\mathcal{S}}^{(0)}$  is the perfectly random subsystem, i.e.  $I(\hat{Q}_t^{(0)}(S(t)); \hat{Q}_{t+1}^{(0)}(S(t+1))) = 0$ . Subsystem 0 takes into account the nonpredictable part of the dynamics.

A subsystem is called trivial, if for all  $t$ :

$$H(\hat{Q}_t^{(i)}(S(t))) = 0 . \quad (1.4)$$

Note that in a strong emergent description, the random subsystem may be trivial. Apart from it, without loss of generality, one can exclude emergent descriptions with trivial subdynamics. Note also that Criterium (1.2) can be strengthened requiring the subsystems to be independent with respect to multi-information, i.e.  $\sum_i H(\hat{Q}_t^{(i)}(S(t))) - H(Q_t(S(t))) = 0$

## 1.4 Discussion and Future Work

The strong emergent description is a notion that provides a criterium to decide whether a collection of (projection) maps decomposes a stochastic dynamical system into independent subdynamics. For systems continuous in both state  $S$  and time  $t$ , it is possible to give a definition that is similar in spirit to that from Sec. 1.3.2, but technically more involved.

As example for the notion, consider the dynamics of the crystal lattice from Sec. 1.2.1. It is a deterministic dynamical system and thus a special case of a stochastic dynamical system, allowing to apply the notion from Sec. 1.3.2. Its decomposition into independent eigenmodes (for any initial state distribution) is a strong emergent description for that system, since the complete set

of eigenmodes covers the whole space, the individual modes are independent of each other and each of them individually fully predicts the future states of its subsystem.

The present concept is related, but goes significantly beyond Independent Component Analysis and deterministic annealing by introducing additional structure in form of a temporal dimension. Due to the subdynamics of the individual projection spaces, they can be interpreted as closed subsystems.

The strong emergent description is a criterium that can be applied to a very general class of systems, going beyond systems that can be studied with linear (as in the crystal lattice) or nonlinear localized analysis (as is done e.g. in synergetics). Although in this paper the space is too limited to discuss the entire spectrum of ramifications of the concept, it should be mentioned that it captures many central aspects and phenomena commonly associated with emergent phenomena. Apart from addressing the issues from Sec. 1.2, the concept of the strong emergent description opens the path towards “soft” probabilistic generalization of algebraic coordinatization models [15] which provide powerful possibilities to decompose systems in a natural fashion; it also opens new perspectives in the study of the emergence of structured information processing in agents from first principles [22, 13].

At this point, neither necessary nor sufficient conditions for the existence or uniqueness of nontrivial strong emergent descriptions in a given system are known. There are indications that, if existing, minimal descriptions might be unique under certain circumstances, but in general, this will not be the case and may not even be desirable: different strong emergent descriptions may lead to alternative valid “views” of a system.

On the practical side, it is possible to construct approximations to strong emergent descriptions, namely *weak emergent descriptions*, by optimizing (maximizing or minimizing, respectively) the terms from the three conditions in Sec. 1.3.2 using either Lagrange multipliers (similar to the multivariate bottleneck models, which also might be modified to provide possible algorithms to compute such a decomposition [21]) or multiobjective optimization. The results of such an optimization may or may not be proper strong emergent descriptions. In systems with a certain amount of mixing, one will not expect to achieve anything beyond a weak description, but for practical purposes weak emergent descriptions can be sufficiently powerful. It should also be mentioned that the limited time range incorporated in the definition in Sec. 1.3.2 allows the possibility to have different emergent descriptions at different times, allowing “paradigm” shifts in the descriptions that may be induced by the dynamics or by the observer.

Future work will refine the computation of the notion, apply it to a larger range of systems, study its utility and investigate the connections of the presented concept to other notions used in the analysis of complex systems.

## Acknowledgements

The author is indebted to Chrystopher Nehaniv, Thomas Martinetz and Jan Kim for helpful discussions.

## Bibliography

- [1] C. Adami. *Introduction to Artificial Life*. Springer, 1998.
- [2] C. Adami, C. Ofria, and T. C. Collier. Evolution of biological complexity. *Proc. Natl. Acad. Sci. USA*, 97:4463–4468, 2000.
- [3] J. F. Adams. *Lectures on Lie Groups*. W. A. Benjamin, Inc., New York, Amsterdam, 1969.
- [4] Y. Bar-Yam. *Dynamics of Complex Systems*. Addison-Wesley, 1997.
- [5] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. Clarendon Press, Oxford, England, 1954.
- [6] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
- [7] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, pages 11–54, 1994.
- [8] G. Deco and B. Schürmann. *Information Dynamics: Foundations and Applications*. Springer, 2001.
- [9] R. Der. Selforganized robot behavior from the principle of homeokinesis. In H.-M. Gro, K. Debes, and H.-J. Böhme, editors, *Proc. Workshop SOAVE '2000 (Selbstorganisation von adaptivem Verhalten)*, volume 643 of *Fortschritt-Berichte VDI, Reihe 10*, pages 39–46, Ilmenau, 2000. VDI Verlag.
- [10] H. Haken. *Advanced synergetics*. Springer-Verlag, Berlin, 1983.
- [11] H. Haken. *Information and Self-Organization*. Springer Series in Synergetics. Springer, 2000.
- [12] I. Harvey. The 3 es of artificial life: Emergence, embodiment and evolution. Invited talk at Artificial Life VII, 1.-6. August, Portland, August 2000.
- [13] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. Technical Report 400, Department of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire, 2004. Accepted to *Evolvable Hardware*.
- [14] A. I. Mees. *Dynamics of feedback systems*. John Wiley & sons, Ltd., 1981.

- [15] C. L. Nehaniv. Algebraic models for understanding: Coordinate systems and cognitive empowerment. In B. G. J. P. Marsh, C. L. Nehaniv, editor, *Proceedings of the Second International Conference on Cognitive Technology: Humanizing the Information Age*, pages 147–162. IEEE Computer Society Press, 1997.
- [16] T. Nomura. Formal descriptions of autopoiesis based on the theory of category. In J. Kelemen and P. Sosik, editors, *Advances in Artificial Life (Proc. 6th European Conference on Artificial Life, Prague, September 10-14, 2001)*, LNCS, pages 700–703. Springer, 2001.
- [17] A. Penn. Steps towards a quantitative analysis of individuality and its maintenance: A case study with multi-agent systems. In D. Polani, J. Kim, and T. Martinetz, editors, *Fifth German Workshop on Artificial Life, March 18-20, 2002, Lübeck, Germany*. IOS Press Infix, Aka, 2002.
- [18] S. Rasmussen, N. Baas, B. Mayer, M. Nilsson, and M. W. Olesen. Ansatz for dynamical hierarchies. *Artificial Life*, 7:329–353, 2001.
- [19] E. M. A. Ronald, M. Sipper, and M. S. Capcarrère. Design, observation, surprise! a test of emergence. *Artificial Life Journal*, 5(3):225–239, 1999.
- [20] W. Rudin. *Fourier Analysis on Groups*. Wiley, New York, 1990.
- [21] N. Slonim, N. Friedman, , and T. Tishby. Agglomerative multivariate information bottleneck. In *Neural Information Processing Systems (NIPS 01)*, 2001.
- [22] H. Touchette and S. Lloyd. Information-theoretic limits of control. *Phys. Rev. Lett.*, 84:1156, 2000.
- [23] S. Winter. Zerlegung von gekoppelten Dynamischen Systemen (Decomposition of Coupled Dynamical Systems. Diploma thesis, Johannes Gutenberg-Universität Mainz, 1996. (In German).
- [24] D. H. Wolpert and W. G. Macready. Self-dissimilarity: An empirically observable complexity measure. In Y. Bar-Yam, editor, *Proc. of the International Conference on Complex Systems, Nashua, NH, 21-26 Sept. 1997*. NECSI, 1997.
- [25] M. Wüstel, D. Polani, T. Uthmann, and J. Perl. Behavior classification with self-organizing maps. In P. Stone, T. Balch, and G. Kraetzschmar, editors, *RoboCup-2000: Robot Soccer World Cup IV*, pages 108–118. Springer Verlag, Berlin, 2001. Winner of the RoboCup 2000 Scientific Challenge Award.