Publisher Name: ZenToks Books
Editorial Book: Issues on Health Science
ISBN: 978-81-958975-1-3

Chapter 15

Organic Component Identification using Syntactic Pattern Recognition

Payal Bose[1] **and Samir Kumar Bandyopadhyay**[2*]

[1]*Assistant Professor, Swami Vivekananda University, Barrackpore, Kolkata, India.*
[2]*Academic Advisor, The Bhawanipur Education Society, Kolkata, India.*

Abstract

The identification of organic components is fundamental to chemistry with implications of detection, spanning drug discovery, materials science, and environmental monitoring. Traditional analytical methods often rely on spectroscopic data and empirical rules, which can be limited in capturing complex structural detection. This paper explores the application of Syntactic Pattern Recognition (SPR) for the identification of organic components. By representing chemical structures in formal languages, SPR leverages grammatical rules to describe valid molecular arrangements. This approach allows for a robust and interpretable framework for recognizing known structures and potentially identifying novel substructures within larger molecules. It is presented here the construction of chemical grammars and present examples of their application. It also highlights the advantages of SPR in handling the inherent structural complexity of organic compounds.

Keywords: Syntactic Pattern Recognition, Organic Component Identification, Functional Groups, Formal Language Theory, Molecular Structure Analysis, Organic Compounds.

1. Introduction

The vast and diverse landscape of organic chemistry necessitates efficient and accurate methods for identifying chemical compounds and their constituent parts. Conventionally, this involves a combination of experimental techniques (e.g., NMR, IR, and Mass Spectrometry) and expert interpretation. While powerful, these methods can be time-consuming and may struggle with the intricate structural nuances of complex molecules or the discovery of previously uncharacterized entities [1].

The precise identification of molecular structural features including specific functional groups and substructures stands as a central and indispensable task across all branches of chemistry. This process is not merely an exercise in classification; it is fundamental to elucidating a compound's intrinsic properties, predicting its reactivity and understanding its biological activity [2]. In the realm of drug discovery, for instance, cheminformatics plays an indispensable role by facilitating the efficient storage, retrieval, and analysis of vast chemical datasets which inherently include complex molecular structures. Furthermore, the ability to identify synthesizable sub structural domains is of immense strategic value in devising effective retro synthetic pathways for the construction of highly challenging molecular targets [3–6].

Despite the critical importance of this task, conventional computational methods for structural analysis, such as Root Mean Square Deviation (RMSD) calculations frequently encounter significant limitations. These methods often suffer from the ”curse of dimensionality” where their performance degrades substantially with increasing molecular size and complexity. It also typically necessitates a consistent ordering of atoms between comparing structures, which can be a non-trivial and computationally intensive prerequisite. The sheer scale of the chemical universe, with estimates of virtual organic chemistry space ranging between 10^{20} and 10^{24} molecules further exacerbates the challenge of comprehensive and efficient structural analysis using traditional approaches [2, 7]. The inherent complexity and vastness of chemical space, coupled with the limitations of existing analytical techniques for large and intricate molecules, underscore a profound need for advanced computational methodologies that transcend the boundaries of single-domain expertise [8]. This situation creates a critical demand for interdisciplinary solutions that can effectively integrate sophisticated chemical knowledge with powerful computational and pattern recognition paradigms, thereby enabling a more robust and scalable approach to molecular identification [9, 10].

How To Cite: P. Bose, S.K. Bandyopadhyay. *Organic Component Identification using Syntactic Pattern Recognition.* In A. Sreenivasan (Eds). Issues on Health Science. Pages 172-177, 2025. ZenToks Books, India. ISBN: 978-81-958975-1-3

Syntactic pattern recognition offers an alternative paradigm by treating patterns as ”sentences” in a formal language. Instead of relying solely on numerical features, SPR emphasizes the structural relationships between primitives (basic elements) that form a pattern [11]. In the context of organic chemistry, these primitives can be atoms, bonds, or small functional groups, and the ”syntax” defines how these primitives combine to form valid organic structures. This approach provides a hierarchical and descriptive framework, enabling the recognition and interpretation of complex molecular architectures. This paper develops the application of SPR for organic component identification and outlining its theoretical underpinnings and practical implications [12].

2. Related Works

The concept of applying formal language theory to chemical structures dates back several decades. Early works explored the use of grammars to describe and generate chemical diagrams and nomenclature. For instance, efforts have been made to derive context-free grammars for systematic organic chemical nomenclature and enabling computer-aided translation [2, 13].

More broadly, syntactic and structural pattern recognition has found applications in various domains where patterns exhibit inherent structural organization, such as image processing, speech recognition, and bioinformatics. In bioinformatics, SPR models have been successfully applied to analyze sequences and structures of DNA, RNA, and proteins, where the hierarchical arrangement of components is crucial [7, 14–18].

Pattern recognition, as a scientific discipline, is fundamentally concerned with the classification of objects into a defined number of categories or classes. This field has garnered considerable research attention over recent decades due to its widespread applicability across diverse domains, including medicine, bioinformatics, and data mining. Broadly, pattern recognition techniques can be categorized into two principal paradigms: statistical and structural, or syntactic approaches [15–17, 19].

Statistical pattern recognition typically operates on numerical feature vectors derived from data, employing probabilistic models to classify patterns. In contrast, Syntactic Pattern Recognition (SPR) analyzes data by explicitly examining structural relationships and grammatical rules. This approach is particularly well-suited for applications where patterns exhibit a definite or hierarchical structure as it utilizes symbolic data representations such as strings, trees, and graphs. SPR provides a powerful framework for understanding and interpreting the hierarchical organization of information, which is a hallmark of many complex systems [17, 20]. The intrinsic structural and hierarchical nature of organic molecules, characterized by specific arrangements of atoms, bonds, and functional groups, makes them highly amenable to analysis through SPR [21]. Unlike statistical methods that might flatten this rich relational information into fixed-dimensional feature vectors, SPR directly models and interprets these complex, nested relationships, offering a more chemically intuitive and robust framework. This direct alignment between the principles of SPR and the inherent structure of chemical entities highlights its potential for significant contributions to chemical analysis.

Syntactic Pattern Recognition applies concepts drawn from formal language theory to decipher complex patterns and thereby establishing a mathematical framework for describing pattern languages through formal grammars [8, 22, 23]. This methodology facilitates the compact representation of intricate structures using recursive rules, enabling both the generation and recognition of an infinite set of valid patterns from a finite set of specifications [16]. The inherent applicability of formal grammars to chemical structures is already evident in existing chemical representations. For instance, the Simplified Molecular Input Line Entry System (SMILES), a widely adopted chemical notation, is recognized within formal language theory as a ”word” that can be parsed by a context-free parser [8, 17, 24–26]. This foundational understanding provides a strong precedent for extending formal grammar approaches to more complex chemical identification tasks.

Pattern recognition, as a scientific discipline, is inherently focused on the classification of objects into predefined categories [27]. Its conceptual origins and theoretical advancements span several decades, with significant developments coinciding with the proliferation of computing capabilities [28]. Syntactic Pattern Recognition (SPR) emerged as a distinct and powerful paradigm specifically tailored for patterns possessing an intrinsic, discernible structure [27]. Statistical methods typically rely on numerical feature vectors of fixed dimensionality; SPR represents patterns using symbolic data structures such as strings, trees, and graphs [16].

The foundational premise in SPR involves employing formal grammars to describe the ”syntax” of various chemical pattern classes. This enables the explicit representation of hierarchical structures and the intricate relationships between elementary pattern components [29, 30]. The recognition process within SPR is then accomplished by parsing an unknown input pattern to determine whether it conforms to the rules specified by a grammar that characterizes a particular pattern class. A notable early contribution by Fu [28] introduced the concept of attributed grammars, which represented a significant step towards unifying statistical and structural pattern recognition approaches. This development acknowledged the inherent value of combining symbolic, structural information with quantitative, numerical attributes, paving the way for more comprehensive pattern analysis [27, 28].

The historical trajectory of pattern recognition reveals a clear and consistent movement towards integrating statistical and syntactic methodologies. This convergence is not merely a pragmatic combination of techniques but reflects a deeper understanding that real-world patterns often exhibit both precise structural regularities and inherent probabilistic variations or noise. Therefore, hybrid approaches which strategically leverage the strengths of both paradigms, for instance, through the use of probabilistic grammars or attributed grammars incorporating stochastic elements are proving to be essential for achieving robust and generalize solutions in complex domains [16]. For the specific challenge of organic component identification, this implies that while the core of the proposed system will be syntactic, incorporating probabilistic elements could significantly enhance its real-world performance by accounting for conformational flexibility, experimental uncertainties, or variations in chemical environments. This adaptive strategy allows the model to capture both the rigid structural rules and the more fluid probabilistic aspects of molecular behaviour [21, 31, 32].

Specifically for molecular structures, researchers have investigated representing molecules as strings, trees, or graphs. Graph-based representations are particularly intuitive for molecules, where atoms are nodes and bonds are edges [17]. Grammatical rules can then define valid connections and arrangements of these graph primitives. The challenge lies in developing comprehensive and computationally tractable grammars that accurately capture the vast diversity and rules of organic chemistry while being robust to variations and errors. While statistical methods have dominated much of the recent pattern recognition landscape, structural and syntactic approaches offer a complementary perspective, particularly when explicit descriptions of relationships are desired.

The primary contributions of this paper include:

- Proposing a novel and robust framework for organic component identification based on SPR grammar.
- Demonstrating the explicit encoding of intricate chemical knowledge, including context-dependent properties and reactivity, directly into the formal grammar rules.
- Providing a detailed step-by-step example of the parsing process to illustrate the inherent interpretability and transparency of the proposed approach.

3. Organic Structure

Formal language theory provides the mathematical foundation for Syntactic Pattern Recognition. A formal grammar, as rigorously defined by Noam Chomsky serves as a mathematical framework for describing the syntax of a language [23, 33]. It is formally represented as a tuple, G= (N, Σ, P, S) [34–36].

- **N (Non-terminal Symbols):** This is a finite set of abstract syntactic variables. These symbols represent categories of phrases or words within the language, such as <Functional Group>, <Ring>, or <Chain> in a chemical context. They are termed ”non-terminal” because they must be further sub-divided into other symbols (either non-terminals or terminals) during the derivation process. The set N is distinct from the final strings generated by the grammar [25].
- Σ **(Terminal Symbols):** This is a finite set of basic, irreducible symbols that constitute the actual words or characters appearing in the final strings generated by the language. In the context of organic chemistry, these could be atomic symbols (e.g., C, O, N), bond types (e.g., =, #, :), or ring closure indicators. They are ”terminal” because they cannot be rewritten further.
- **P (Production Rules):** This is a finite set of rules that dictate how strings of symbols can be rewritten. Each rule typically takes the form $\alpha \rightarrow \beta$, where α and β are strings of non-terminal and/or terminal symbols. A critical condition for these rules is that the left-hand side (α) must contain at least one non -terminal symbol. These rules govern the generation of valid patterns.
- **S (Start Symbol):** This is a distinguished non-terminal symbol (SεN) from which the derivation process always commences. It represents the highest-level syntactic category, such as <Molecule> in a chemical grammar. The language generated by the grammar, denoted L (G), comprises all strings consisting solely of terminal symbols that can be derived from the start symbol through a finite sequence of production rule applications.

The Grammar hierarchy classifies formal grammars into four types, each generating languages of increasing complexity [26]. The Rules of Chomsky are highly restricted (e.g., A→BC or A→ a), Where A, B and C are non-terminal and small letter such as a represents terminal. The above specified are grammatical rules as per Chomsky Grammar.

The Chomsky hierarchy provides a foundational understanding of the expressive power required to model different types of patterns [37]. For organic component identification, the inherent context-dependence of chemical functional groups can have properties and even the definition of a group can be influenced by its surrounding molecular environment [33]. It can suggest that context-sensitive grammars or at least context-free grammars augmented with additional mechanisms are necessary to represent chemical structure. This is because the precise chemical behaviour and structural identity of a functional group are not always independent of their molecular context. For example, the reactivity of a carbonyl group can vary significantly depending on whether it is part of an aldehyde, ketone, carboxylic acid, or ester [38, 39]. A grammar capable of capturing these contextual nuances would provide a more accurate and chemically meaningful representation.

Organic structures are inherently hierarchical and compositional. At the most fundamental level, they are composed of atoms (e.g., carbon, hydrogen, oxygen, nitrogen) connected by various types of bonds (single, double, triple, aromatic). These atoms and bonds combine to form functional groups (e.g., hydroxyl, carbonyl, amino), which in turn dictate a molecule’s chemical properties [40]. Functional groups can be arranged in chains, rings, or complex fused systems.

For the purpose of syntactic pattern recognition, it can define the primitives of an organic structure. These could include:

- **Atomic symbols:** C, H, O, N, S, P, Halogens, etc.
- **Bond types:** single (-), double (=), triple (≡), aromatic (ar).
- **Specialized symbols:** For charges, lone pairs, or stereo chemical indicators (e.g., R/S).

The ”patterns” aims to recognize that can range from simple functional groups (e.g., an alcohol group -OH, a carboxylic acid -COOH) to more complex substructures (e.g., benzene ring, a specific amino acid residue) or even entire molecules. The challenge is to formalize the rules by which these primitives combine to form chemically valid and meaningful structures [28, 41–44].

4. Grammar

A formal grammar G is typically defined as a 4-tuple: G= (V, T, P, S), where:

- V: A finite set of non-terminal symbols (variables representing intermediate structural components).
- T: A finite set of terminal symbols (the basic primitives, e.g., atomic symbols, bond types).
- P: A finite set of production rules (rules for replacing non-terminal symbols with strings of terminal and/or non-terminal symbols).
- S: The start symbol (representing the highest-level pattern, e.g., a complete molecule).

Each rule typically takes the form $\alpha \rightarrow \beta$, where α and β are strings of non-terminal and/or terminal symbols. The right arrow indicates → replacement.

Example Production Rules:

Let's consider a simplified grammar for identifying an alkane chain and a hydroxyl group:

Terminal Symbols (T): $\{C, H, -, O\}$ (Carbon, Hydrogen, single bond, Oxygen)
Non-terminal Symbols (V): {<Molecule>, <AlkaneChain>, <Methylene>, <Methyl>, <Hydroxyl>}
Start Symbol (S) : <Molecule>
Production Rules (P):

1. <Molecule> right arrow <AlkaneChain>
2. <Molecule> right arrow <AlkaneChain>- <Hydroxyl> (for an alcohol)
3. <AlkaneChain> right arrow <Methyl>
4. <AlkaneChain> right arrow <Methyl> - <Methylene>
5. <Methylene> right arrow C-H-H
6. <Methylene> right arrow C-H-H-<Methylene>
7. <Methyl> right arrow C-H-H-H
8. <Hydroxyl> right arrow O-H

This is a very basic example. A more comprehensive grammar would include rules for double/triple bonds, other functional groups, ring structures, branching, and stereochemistry. Representing connectivity explicitly, perhaps using a string notation or an adjacency list would be crucial for practical implementation. The parser would then analyze an input string (representing the molecule) to determine if it conforms to the defined grammar.

5. Examples

Consider the identification of ethanol CH_3CH_2OH using the simplified grammar above.

Input String (Simplified for illustration): C-H-H-H-C-H-H-O-H
Parsing Process (Derivation):

1. <Molecule> right arrow <AlkaneChain>-<Hydroxyl> (Rule 2)
2. <AlkaneChain> right arrow <Methyl>-<Methylene> (Rule 4)
3. <Methyl>-<Methylene>-<Hydroxyl>
4. <Methyl> right arrow C-H-H-H (Rule 7)
5. C-H-H-H-<Methylene>-<Hydroxyl>
6. <Methylene> right arrow C-H-H (Rule 5)
7. C-H-H-H-C-H-H-<Hydroxyl>
8. <Hydroxyl> right arrow O-H (Rule 8)
9. C-H-H-H-C-H-H-O-H (Matches input string)

This successful parsing indicates that the input structure is recognized as an alcohol, specifically containing a methyl group, a methylene group, and a hydroxyl group.

For more complex structures, the grammar would need to incorporate rules for:

- **Branching:** How side chains are attached.
- **Rings:** Rules for forming cyclic structures (e.g., cyclohexane, benzene).
- **Aromaticity:** Special rules for aromatic systems.
- **Multiple functional groups:** How different functional groups coexist and interact.

The success of SPR heavily relies on the completeness and accuracy of the defined grammar. Grammar induction techniques, which automatically learn production rules from a set of examples, could be explored to develop more robust chemical grammars.

6. Conclusions

Syntactic pattern recognition offers a powerful and interpretable approach to organic component identification. By formalizing the structural rules of chemistry into grammars, it enables the automated recognition of functional groups, substructures, and entire molecules. This method provides a clear, hierarchical description of chemical patterns, making it inherently suitable for the compositional nature of organic compounds.

While challenges exist in developing comprehensive and computationally efficient grammars for the vast complexity of organic chemistry, the benefits of SPR, including its descriptive power and potential for error-correcting parsing (to handle imperfect or noisy input data), warrant further exploration. Future work could focus on developing more sophisticated graph grammars to better represent molecular topology, integrating statistical methods for enhanced robustness, and exploring grammar induction techniques to automatically learn and refine chemical grammars from large datasets of known structures. Ultimately, syntactic pattern recognition holds promise as a valuable tool in the ongoing quest for efficient and intelligent chemical discovery and analysis.

References

[1] L. Cao, C. Li, and T. Mueller. The use of cluster expansions to predict the structures and properties of surfaces and nanostructured materials. *Journal of Chemical Information*, 2018.

[2] R. M. Acheson. An Introduction to the Chemistry of Heterocyclic Compounds. 1976.

[3] D. S. Wishart. *Introduction to cheminformatics*. Current Protocols in Bioinformatics, 2016.

[4] Qifeng Bai, Shuoyan Tan, Tingyang Xu, Huanxiang Liu, Junzhou Huang, and Xiaojun Yao. *MolAICal: a soft tool for 3D drug design of protein targets by artificial intelligence and classical algorithm*. Briefings in Bioinformatics, 2021.

[5] S. Bentolila. A grammar describing 'biological binding operators' to model gene regulation. *Biochimie*, 1996.

[6] W. Duch, K. Swaminathan, and J. Meller. *Artificial intelligence approaches for rational drug design and discovery*. Current Pharmaceutical Design, 2007.

[7] Attila Egri-Nagy, Chrystopher L. Nehaniv, John L. Rhodes, and Maria J. Schilstra. *Automatic analysis of computation in biochemical reactions*. Elsevier, Biosystems, 2008.

[8] I. C. Baianu. *Computer models and automata theory in biology and medicine*. Elsevier, Mathematical Modelling, 1986.

[9] Qing Lu. *Identifying molecular structural features by pattern recognition methods*. RSC Publishing, 2022.

[10] Z. G. Wang, J. Elbaz, F. Remade, R. D. Levine, and I. Willner. *All-DNA finite-state automata with finite memory*. BIOPHYSICS AND COMPUTATIONAL BIOLOGY, 2010.

[11] A. Mariusz Flasinski. SURVEY ON SYNTACTIC PATTERN RECOGNITION METHODS IN BIOINFORMATICS. *Computer Science*, 2024.

[12] M. Kadukova and S. Grudinin. Knodle: a support vector machines-based automatic perception of organic molecules from 3D coordinates. *Journal of Chemical Information and Modeling*, 2016.

[13] Minghao Guo, Wan Shou, Liane Makatura, Timothy Erps, Michael Foshey, and Wojciech Matusik. Grammar for Digital Polymer Representation and Generation. *Advanced Science*, 2022.

[14] N. S. Chang and K. S. Fu. Parallel parsing of tree languages for syntactic pattern recognition. *Pattern Recognition*, 1979.

[15] C. H. Chen, L. F. Pau, and P. S. Wang. Handbook of Pattern Recognition and Computer Vision. *World Scientific*, 1993.

[16] H. Bunke. *Attributed programmed graph grammars and their application to schematic diagram interpretation*. Pattern Analysis, IEEE Trans, 1982.

[17] J. C. Cheng and H. S. Don. A graph matching approach to 3-D point correspondences. *Pattern Recognition*, 1991.

[18] H. Bunke and A. Sanfeliu. Syntactic and Structural Pattern Recognition - Theory and Applications. *World Scientific*, 1990.

[19] K. S. Fu. *A step towards unification of syntactic and statistical pattern recognition*. Pattern Analysis, IEEE Trans, 1983.

[20] V. J. Cook. Chomsky's Universal Grammar and Second Language Learning. *Applied Linguistics*, 1985.

[21] C. J. Bartel, S. L. Millican, A. M. Deml, J. R. Rumptz, W. Tumas, A. W. Weimer, S. Lany, V. Stevanović, C. B. Musgrave, and A. M. Holder. Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry. *Nature Communications*, 2018.

[22] S. B. Šegota, N. Anelić, I. Lorencin, J. Musulin, D. Štifanić, and Z. Car. *Preparation of Simplified Molecular Input Line Entry System Notation Datasets for use in Convolutional Neural Networks*. IEEE, 2021.

[23] Teresa Parodi. *Universal Grammar and Second Language Acquisition*. Encyclopaedia of Applied Linguistics, 2013.

[24] Alfred V. Aho, Monica S. Lam, Ravi Sethi, and Jeffrey D. Ullman. *Compilers: Principles, Techniques, and Tools, Pearson*. Addison-Wesley, 1986.

[25] Raj Kishor Bisht. A Survey of Applications of Finite Automata in Natural Language Processing. *International Journal on Emerging Technologies*, 2017.

[26] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 2001.

[27] K. S. Fu and A. Rosenfeld. *Pattern Recognition and Image Processing*. IEEE Transaction Computers, 1976.

[28] J. M. Mendel and K. S. Fu. Adaptive, Learning and Pattern Recognition Systems: Theory and Applications. *IEEE Transactions on Automatic Control*, 1972.

[29] M. Tsuchiya and J. Ross. Application of genetic algorithm to chemical kinetics: systematic determination of reaction mechanism and rate coefficients for a complex reaction network. *The Journal of Physical Chemistry A*, 2001.

[30] Poi Tamrakar, Abha Pathak, Pallavi Thorat, Mily Lal, and Akanksha Goel. Manisha Bhende, and Swati Sharma, Deep learning in chemical reaction prediction and synthesis Planning. In *AIP Conference Proceding, 2024*, 2024.

[31] O. Engkvist, P. O. Norrby, N. Selmi, Y. H. Lam, Z. Peng, E. C. Sherel, W. Amberg, T. Erhard, and L. A. Smyth. *Computational prediction of chemical reactions: current status and outlook*. Drug Discovery Today, 2018.

[32] D. Jha, L. Ward, A. Paul, W. Liao, A. Choudhary, C. Wolverton, and A. Agrawal. ElemNet: deep learning the chemistry of materials from only elemental composition. *Scientific Reports*, 2018.

[33] Noam Chomsky, Aspects of the Theory of Syntax, MIT Press. 2014.

[34] P. Hong, M. Turk, and T. S. Huang. Gesture modeling and recognition using finite state machines. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*, 2000.

[35] K. Oishi and E. Klavins. *Framework for engineering finite state machines in gene regulatory networks*. ACS Synthetic Biology, 2014.

[36] C. L. Giles, C. B. Miller, D. Chen, H. H. Chen, G. Z. Sun, and Y. C. Lee. Learning and extracting finite state automata with second-order recurrent neural networks. *Neural Computation*, 1992.

[37] J. Andrew GrantJames A. HaighBarry T. PickupAnthony NichollsRoger A. Sayle and Finite Lingos. State Machines, and Fast Similarity Searching. *Journal of Chemical Information and Modeling*, 2018.

[38] J. Garparić. Identification of organic compounds by chromatography after decomposition and degradation. *Journal of Chromatography A*, 1970.

[39] Omar M. Yahya. Synthesis, Characterization and Study Biological Activity of Substituted 4-Amino -3,5-Bis (2,4-dichloro phenoxy)-1,2,4-Triazole. *Rafidain Journal of Science*, 2024.

[40] Tianhong Zhang, Hongli Li, Hualin Xi, Robert V. Stanton, and Sergio H. Rotstein. *HELM: A Hierarchical Notation Language for Complex Biomolecule Structure Representation*. ACS Publication, 2012.

[41] P. Manolios and R. Fanelli. First-order recurrent neural networks and deterministic finite state automata. *Neural Computation*, 1994.

[42] M. Casey. The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction. *Neural Computation*, 1996.

[43] C. W. Omlin and C. L. Giles. Constructing deterministic finite-state automata in recurrent neural networks. *Journal of the ACM*, 1996.

[44] N. Ganesh and N. G. Anderson. *Irreversibility and dissipation in finite-state automata*. Elsevier, Physics Letters A, 2013.