A Comparative Study of Diffusion Models and Long Short-Term Memory Networks through the Reaction–Diffusion Equation

Vincent Jing-Cheng Lo, Kotaro James Nishimura and Kazushi Ikeda

Department of Information Science, Nara Institute of Science and Technology, Ikoma, Japan
(E-mail: vincent.lo@telecom-sudparis.eu, kazushi@is.naist.jp)

Abstract: In this work, we review the similarities between two machine learning algorithms, the Long Short-Term Memory (LSTM) networks and the Denoising Diffusion Probabilistic Models (DDPMs) through the Reaction-Diffusion (RD) equation, where morphogen concentrations vary depending on a diffusion with two different time constants and a reaction between them. On one hand, the LSTM networks have the interaction between long and short term memories, which is analogous to the RD mechanism with a discrete time framework. On the other hand, the DDPMs have both the drift and diffusion terms, which correspond to the RD mechanism. These facts mean that a general mathematical equation based on the RD mechanism can describe the two models. To empirically validate the theoretical mapping, a comparative experiment is designed to measure the internal dynamics of "Reaction" and "Diffusion" across the architectures. As results, their similarity and relationship were confirmed to some extent, although they are preliminary and need more studies in the future.

Keywords: Reaction-Diffusion Equation, Diffusion Models, Long Short-Term Memory, Denoising Diffusion Probabilistic Models

1. INTRODUCTION

Understanding how complex patterns emerge from simple rules is a central question in both biology and artificial intelligence. In biological systems, the Reaction–Diffusion (RD) model provides an explanation for spontaneous pattern formation. It describes how interacting chemical substances, an activator and an inhibitor, diffuse through space and react locally, generating stable spatial patterns such as stripes, spots, or waves [1, 2]. This interaction between local activation and global inhibition has since been recognized as a general mechanism for self-organization in living systems. In parallel, modern machine learning models have revealed similar dynamics in artificial systems. Long Short-Term Memory (LSTM) networks are a class of recurrent neural architectures designed to model temporal sequences and retain information over long time series data. They achieve this by regulating how past states influence future predictions, allowing them to capture both short-term and long-term dependencies in dynamic data. In a similar perspective, Denoising Diffusion Probabilistic Models (DDPMs) construct structured data from random noise through an iterative denoising process that progressively reverses a stochastic Gaussian noise diffusion. Although all these frameworks are used in different disciplines, they share a common mathematical foundation: the dynamic balance between growth and decay, reaction and diffusion. Investigating these parallels provides an understanding of pattern formation, stability, and feedback control in both natural and artificial systems. However, the literature lacks a formal mathematical framework linking these two machine learning models. Existing studies rarely examine how the internal dynamics of artificial neural networks can be interpreted using concepts from biological Reaction–Diffusion systems. This absence leaves unexplored a potentially rich analogy between biological self-organization and artificial memory regulation.

2. BACKGROUND

2.1. Reaction-Diffusion Systems

Reaction-Diffusion systems are a theoretical model used to explain self-regulated pattern formation in biology. More specifically, in the classical Turing framework, the activator **u** and the inhibitor **v** are two interacting morphogens that obey the partial differential equations in Eq. (1) [2]. They diffuse and react with each other so that their concentrations can form stable spatial patterns such as stripes, spots, or waves.

$$
\begin{aligned}
\partial_t \mathbf{u} &= D_u \nabla^2 \mathbf{u} + R_u(\mathbf{u}, \mathbf{v}) \\
\partial_t \mathbf{v} &= D_v \nabla^2 \mathbf{v} + R_v(\mathbf{u}, \mathbf{v})
\end{aligned}
\tag{1}
$$

where D_u and D_v are the diffusion coefficients and R_u and R_v are the functions that describe the local reactions between **u** and **v**. The Reaction-Diffusion equation is a sum of a diffusion term and a reaction term.

A central mechanism underlying pattern formation in Reaction-Diffusion systems is the combination of short-range positive feedback from the activator and long-range negative feedback from the inhibitor [1].

Short-range positive feedback allows the activator to enhance its own production and diffuse slowly over a local neighborhood in the spatial domain. Hence, R_u increases when **u** increases. In other words, a small local increase in a quantity causes more of itself to be produced nearby. Because the activator diffuses slowly, it remains concentrated in small regions, enabling these regions to "stand out" from their surroundings. Even a tiny fluctuation in activator concentration can grow locally, producing the seeds of a pattern.

In contrast, long-range negative feedback acts globally. It is implemented by the second morphogen, the inhibitor, whose role is to suppress and counteract the activator. The inhibitor diffuses much faster and spreads its effect over a

---

global spatial domain. The inhibitor rapidly propagates to prevent the activator from forming a uniform concentration everywhere.

The combination creates patterns only if the inhibitor diffuses faster, i.e. $D_v > D_u$.

2.2. Long Short-Term Memory Networks

Long Short-Term Memory networks (LSTMs) are recurrent neural networks designed to preserve information over long temporal horizons and overcome the Vanishing Gradient Problem. LSTMs achieve this through gates and the cell state $(\mathbf{C_t})_t$, a component that retains and forgets information to capture long-term dependencies at each time step. On the other hand, the hidden state $(\mathbf{h_t})_t$ describes the short-term memory and the main output of one LSTM cell for the current time step. Both states are computed using the sigmoid σ and hyperbolic tangent $\tanh$ functions to map proportions. The principle of LSTMs relies on Eqs. (2) to (7) [3] alongside Fig. 1.

The forget gate $(\mathbf{f_t})_t$ in Eq. (2) determines what information to discard from the cell state at time $t - 1$. This decision is made according to the new input $\mathbf{x_t}$ and the previous short-term memory hidden state $\mathbf{h_{t-1}}$. This mechanism maps the inhibition of the cell state.

$$\mathbf{f_t} = \sigma(\mathbf{W_f x_t} + \mathbf{U_f h_{t-1}} + \mathbf{b_f}) \tag{2}$$

The input gate $(\mathbf{i_t})_t$ determines what new information to store in the cell state with respect to Eqs. (3) and (4). This gate is processed in two parts: a sigmoid function to decide which information to update in $\mathbf{C_t}$ and the candidate cell state $\tilde{\mathbf{C}}_t$, which acts as the new information and amplification.

$$\mathbf{i_t} = \sigma(\mathbf{W_i x_t} + \mathbf{U_i h_{t-1}} + \mathbf{b_i}) \tag{3}$$
$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W_C x_t} + \mathbf{U_C h_{t-1}} + \mathbf{b_C}) \tag{4}$$

The critical equation where the dynamic balance occurs is Eq. (5), the cell state update. It combines the previous cell state, the selection of information to forget, and the selection of information to add and update.

$$\mathbf{C_t} = \mathbf{f_t} \odot \mathbf{C_{t-1}} + \mathbf{i_t} \odot \tilde{\mathbf{C}}_t \tag{5}$$

Finally, the output gate $(\mathbf{o_t})_t$ determines which part of the new cell state at time t is output as the next hidden state in Eqs. (6) and (7).

$$\mathbf{o_t} = \sigma(\mathbf{W_o x_t} + \mathbf{U_o h_{t-1}} + \mathbf{b_o}) \tag{6}$$
$$\mathbf{h_t} = \mathbf{o_t} \odot \tanh(\mathbf{C_t}) \tag{7}$$

The matrices $\mathbf{W_f}, \mathbf{W_i}, \mathbf{W_C}, \mathbf{W_o}$ and $\mathbf{U_f}, \mathbf{U_i}, \mathbf{U_C}, \mathbf{U_o}$ represent the weight parameters of the LSTM, while the vectors $\mathbf{b_f}, \mathbf{b_i}, \mathbf{b_C}, \mathbf{b_o}$ are the bias parameters. The operator $\odot$ is the Hadamard product.

The cell state structure directly mirrors the Reaction-Diffusion mechanism in a discrete form: the $\mathbf{f_t} \odot \mathbf{C_{t-1}}$ term acts as the long-range inhibition and controls stability, while the $\mathbf{i_t} \odot \tilde{\mathbf{C}}_t$ term acts as the local reaction that shows growth.
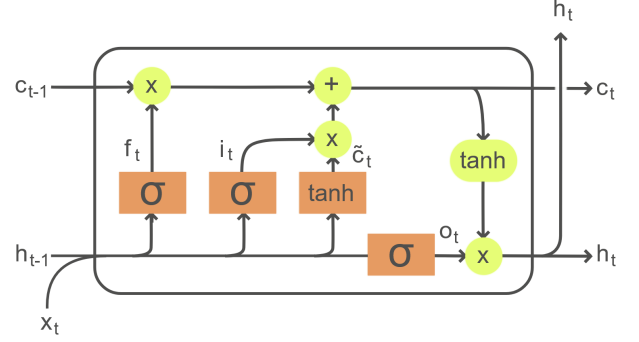

Fig. 1. LSTM cell architecture

2.3. Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models learn to construct data by reversing a noisy diffusion process. This framework fundamentally relies on a forward diffusion process and a reverse denoising process.

The forward diffusion process adds Gaussian noise in T time steps in order to transform a data sample $\mathbf{x_0}$ into pure noise $\mathbf{x_T}$. Each successive data point is given by Eq. (8) [4, 6].

$$\mathbf{x_{t+1}} = \sqrt{1 - \beta_t}\mathbf{x_t} + \sqrt{\beta_t}\epsilon_t \tag{8}$$

where β_t is the noise schedule that decreases with time and $\epsilon_t \sim \mathcal{N}(\epsilon_t | \mathbf{0}, \mathbf{Id})$ is the noise at time t.

This process can be modeled as a Markov chain following the distribution $q(\mathbf{x_t}|\mathbf{x_{t-1}}) = \mathcal{N}(\mathbf{x_t}; \sqrt{1 - \beta_t}\mathbf{x_{t-1}}, \beta_t\mathbf{Id})$.

The reverse denoising process is the generative part of the model that learns to reverse the forward process from $\mathbf{x_T}$ to $\mathbf{x_0}$. This process can also be approximated by a Gaussian law if the noise schedule is small. Thus, the reverse process can be modeled by the distribution $p(\mathbf{x_{t-1}}|\mathbf{x_t}, \mathbf{w}) = \mathcal{N}(\mathbf{x_{t-1}}; \mu(\mathbf{x_{t-1}}, \mathbf{w}, \mathbf{t}), \beta_t\mathbf{Id})$ where $\mu(\mathbf{x}, \mathbf{w}, \mathbf{t})$ is a deep neural network governed by the set of parameters $\mathbf{w}$.

The crucial point of DDPMs for this work is the view of such models through the lens of Continuous-Time Stochastic Partial Equations (SDEs) that highlights a deterministic drift term in $\mathbf{dt}$ and a stochastic diffusion term in $\mathbf{dw}$ [6]. The forward process corresponding to SDE is shown in Eq. (9), while the reverse process SDE is given in Eq. (10).

$$\mathbf{dx} = \underbrace{f(\mathbf{x}, \mathbf{t})\mathbf{dt}}_{\text{drift}} + \underbrace{g(\mathbf{t})\mathbf{dw}}_{\text{diffusion}} \tag{9}$$

$$\mathbf{dx} = \underbrace{[f(\mathbf{x}, \mathbf{t}) - \mathbf{g^2(t)}\nabla_{\mathbf{x}} \ln \mathbf{p(x)}]\mathbf{dt}}_{\text{drift}} + \underbrace{g(\mathbf{t})\mathbf{d\bar{w}}}_{\text{diffusion}} \tag{10}$$

where $\nabla_{\mathbf{x}} \ln p(\mathbf{x})$ is the score function, $\mathbf{w}$ a standard Brownian motion, and $\bar{\mathbf{w}}$ its inverse in time. The score function $\nabla_{\mathbf{x}} \ln p(\mathbf{x})$ is approximated by the denoising neural network ϵ_θ during training.

Both the forward and reverse DDPM processes are now expressed in the canonical form of an SDE, which is structurally identical to the Reaction-Diffusion equation when viewed as a dynamic system. Iteratively, the diffused noise is reversed by a learning denoising drift function.

3. THE REACTION-DIFFUSION PRINCIPLE AS A UNIFYING ANALOGY

In this section, we map the components of the biological RD model to the gate mechanisms of LSTMs and the stochastic differential equations of DDPMs under qualitative hypotheses and assumptions.

3.1. Discrete Activation and Forgetting

In the Reaction-Diffusion model, the concentration of morphogens evolves over continuous space and time. Using Eq. (11) instead of Eq. (5) highlights the mapping from biology to the LSTM networks.

$$\mathbf{C_t} - \mathbf{C_{t-1}} = \underbrace{-(1 - \mathbf{f_t}) \odot \mathbf{C_{t-1}}}_{\text{diffusion}} + \underbrace{\mathbf{i_t} \odot \tilde{\mathbf{C}}_\mathbf{t}}_{\text{reaction}} \qquad (11)$$

In an LSTM, in contrast to the RD model, the time is discrete. Despite this discretization, the update rule for the LSTM cell state mirrors the Reaction-Diffusion dynamic. If we look at Eq. (11), we can identify the term $-(1 - \mathbf{f_t}) \odot \mathbf{C_{t-1}}$ as the diffusion component that puts in recurrence the cell state, while the term $\mathbf{i_t} \odot \tilde{\mathbf{C}}_\mathbf{t}$ can be viewed as the reaction component. Hence, two hypotheses are made:

- The input gate and its activation coefficient $\mathbf{i_t}$ act as the activator of the system and allow reactions. In LSTM, the input gate $\mathbf{i_t}$ selectively amplifies the new signal $\tilde{\mathbf{C}}_t$, by injecting "mass" into the cell state for the next time step. This creates short-range positive feedback, allowing specific patterns to grow within the memory representation, analogous to the autocatalysis of the activator in Turing's model. For any data, it would restore the structure and patterns. This statement is stronger for periodic data due to periodicity in Turing patterns.

- The forget gate and its forget coefficient $\mathbf{f_t}$ act as the inhibitor and allow diffusion. By multiplying the previous state by a factor $\mathbf{f_t} \in [\mathbf{0}, \mathbf{1}]^\mathbf{h}$, the network applies a decay force. This is functionally equivalent to the diffusion of the inhibitor in RD. Just as the inhibitor prevents the activator from saturating the entire domain, the forget gate prevents the cell state from saturating with irrelevant information and selects the information to keep from the past, ensuring stability over long temporal sequences and creating long-range negative feedback on the cell state.

3.2. Stochastic Drift and Noise Spreading

The analogy is most explicit in DDPMs when viewed through the lens of Continuous-Time Stochastic Differential Equations [6]. The reverse generative process is defined by its stochastic diffusion term and its drift term that can be directly mapped as a diffusion and a reaction component similar to RD models. Once again, we make the following hypothesis based on qualitative assumptions:

- The learned drift term $[f(\mathbf{x}, \mathbf{t}) - \mathbf{g^2}(\mathbf{t})\nabla_\mathbf{x} \ln \mathbf{p}(\mathbf{x})]\mathbf{dt}$, specifically the score function approximation $\nabla_\mathbf{x} \ln p(\mathbf{x})$ parameterized by ϵ_θ, acts as the reaction component and the activator. Geometrically, this term represents a gradient field pointing towards the modes of the data distribution. It functions as a restorative force, actively pushing the chaotic state

$\mathbf{x_t}$ towards the structured data manifold and performs a short-range positive feedback from learning to generate images. This mirrors the role of the activator in Turing patterns, which drives the local concentration of chemical substances to form the core of a pattern.

- The term $\mathbf{g}(\mathbf{t})\mathbf{d\bar{w}}$ introduces stochastic diffusion. In the absence of this term, the reverse process would reduce to an Ordinary Differential Equation (ODE) performing deterministic gradient ascent, likely causing the system to collapse into a single, dominant mode regardless of the starting noise. The injection of noise using Brownian motion adds randomness to the generation and acts as a dispersive force, analogous to the inhibitor in biological systems. Just as rapid diffusion prevents the activator from saturating the spatial domain, the stochastic term in DDPMs prevents the drift from overfitting to a single solution, ensuring the generation of diverse, stable, and robust patterns.

4. EXPERIMENTS

To empirically validate the theoretical mapping proposed in the previous section, a comparative experiment is designed to measure the internal dynamics of "Reaction" and "Diffusion" across the three distinct architectures.

A primary challenge in this comparative study is the dimensional mismatch between systems: RD models operate on chemical concentrations, LSTMs on abstract hidden states and 1D sequences, and DDPMs on pixel intensities. Direct comparison of raw values is therefore impossible. To resolve this, we propose a dimensionless metric based on component magnitude ratios.

4.1. Models

- **Reaction-Diffusion**: We employ the Gray-Scott model, a canonical RD system known for generating complex Turing patterns. The system simulates the interaction of two chemical substances $\mathbf{u}$ and $\mathbf{v}$ on a 2D grid with periodic boundary conditions. In this model, the reaction functions for the activator and the inhibitor become, respectively, $R_u(\mathbf{u}, \mathbf{v}) = -\mathbf{u^2v} + \mathbf{F}(\mathbf{1} - \mathbf{u})$ and $R_v(\mathbf{u}, \mathbf{v}) = \mathbf{u^2v} - (\mathbf{F} + \mathbf{k})\mathbf{v}$ where $\mathbf{F}$ is the feed rate and $\mathbf{k}$ is the kill rate [5]. We will take a feed rate in $\mathbf{F} \in [\mathbf{0.029}, \mathbf{0.055}]$ and a constant kill rate $\mathbf{k} = \mathbf{0.062}$ with diffusion coefficients $D_u = 1 \times 10^{-5}$ and $D_v = 0.5 \times 10^{-5}$. A feed rate of 0.029 defines spot patterns, while 0.055 will produce stripes.

- **LSTM**: We implement a standard single-layer LSTM using the Eqs (2) to (7) with dimension $\mathbf{d} = \mathbf{1}$ input feature and $\mathbf{h} = \mathbf{16}$ hidden features. First of all, the network is trained on a simple sinusoidal waves dataset generated with a constant sequence length $\mathbf{s} = \mathbf{200}$ with random frequency and phase and learns to reproduce the curve shape of the waves from a given sequence. In the training we minimize the mean-squared error between the actual signal and the predicted signal output by the model at each time step. To add robustness to this study, the LSTM is also trained on a "chaos" dataset generated from one equation of the Lorenz Attractor system and another dataset based on the Van der Pol oscillator generated in dimension 1.

• **DDPM**: We utilize a standard DDPM architecture with a U-Net model for the denoising/learned drift function ϵ_θ. The model is trained on the MNIST dataset and learns to reverse a Gaussian diffusion process to construct structured digit images from pure noise. The model is trained to minimize the variational lower bound on the negative log-likelihood, approximated by the simple mean-squared error between predicted and actual noise: $||\epsilon - \epsilon_\theta(\mathbf{x_t}, \mathbf{t})||^2$. Additional trainings on the FashionMNIST dataset and CIFAR10 dataset are done to see if the results are reproducible and robust.

4.2. The Reaction Metric $\mathcal{M}_{\text{react}}$

To quantify the magnitude of the term responsible for creating, restoring, or amplifying patterns or data structures, we compute the L2 norm of the identified reaction component in each model and take the mean:

$$\mathcal{M}_{\text{react}}^{\text{RD}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}||\mathbf{u_t v_t^2} - (\mathbf{F} + \mathbf{k})\mathbf{v_t}||_{\mathbf{2}} \tag{12}$$

$$\mathcal{M}_{\text{react}}^{\text{LSTM}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}||\mathbf{i_t} \odot \tilde{\mathbf{C}}_{\mathbf{t}}||_{\mathbf{2}} \tag{13}$$

$$\mathcal{M}_{\text{react}}^{\text{DDPM}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}||\epsilon_\theta(\mathbf{x_t}, \mathbf{t})||_{\mathbf{2}} \tag{14}$$

where T is the total time length used for pattern generation depending on the model, and ϵ_θ designs the learned drift.

4.3. The Diffusion Metric $\mathcal{M}_{\text{diff}}$

Here, we measure the magnitude of the term responsible for spreading, decaying, or randomizing the pattern by using the same process as the reaction metric:

$$\mathcal{M}_{\text{diff}}^{\text{RD}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}||D_v\nabla^2\mathbf{v_t}||_{\mathbf{2}} \tag{15}$$

$$\mathcal{M}_{\text{diff}}^{\text{LSTM}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}||(1 - \mathbf{f_t}) \odot \mathbf{C_{t-1}}||_{\mathbf{2}} \tag{16}$$

$$\mathcal{M}_{\text{diff}}^{\text{DDPM}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}|||\sigma_t\mathbf{z}||_{\mathbf{2}} \tag{17}$$

where $\sigma_t\mathbf{z}$ designs the diffused noise.

4.4. Mean Dynamic Balance Ratio

By defining both metrics as magnitudes of update vectors, their ratio becomes dimensionless and directly comparable across domains and defines the Mean Dynamic Balance Ratio:

$$R = \frac{\mathcal{M}_{\text{react}}}{\mathcal{M}_{\text{diff}}} \tag{18}$$

The Mean Dynamic Balance Ratio takes values into $[0, +\infty[$. We admit hypothetically that complex and stable pattern formation requires $R \approx 1$, and such a value defines the equilibrium baseline of Reaction-Diffusion mechanism.

If $R \ll 1$, the diffusion force is much stronger than the reaction force. The system smooths out differences faster than they can be created and leads to homogeneity in the patterns.

In LSTM, it corresponds to vanishing gradient, and the signal will converge to zero. In DDPM, the generated image will remain pure noise with no structure. On the contrary, if $R \gg 1$, the reaction force dominates and the systems amplify every small fluctuation. It leads to a single solution and cannot adapt to new patterns or data.

5. EXPERIMENTAL RESULTS

For each model and for each set of parameters or training, the magnitudes for reaction and diffusion are computed. Table 1 presents the Mean Dynamic Balance Ratio that describes the overall stability between reaction and diffusion across the total time T. In Fig. 2, we plot the Dynamic Balance Ratio over a normalized time $t \in [0, 1]$ for each model that is defined in Eq. (20).

$$\mathcal{M}_{\mathbf{i}} = \frac{1}{T}\sum_{\mathbf{t=1}}^{T}\mathcal{M}_{\mathbf{i,t}} \tag{19}$$

$$R_{\mathbf{t}} = \mathcal{M}_{\text{react},\mathbf{t}}/\mathcal{M}_{\text{react},\mathbf{t}} \tag{20}$$

Reaction-Diffusion models achieve a ratio of approximately 1.00 ± 0.03 as expected on both spot and stripe regimes. This RD ratio confirms that Turing pattern formation requires the constructive reaction force to counter the dispersive diffusion force. Hence, a value of $R = 1$ corresponds to a perfect stability. This observation is even more relevant as in Fig. 2, the Dynamic Balance Ratio (orange and blue lines) starts with transient fluctuations and spikes but rapidly converges to 1.0. It shows a static stability reached: once the Turing instability creates the patterns, the local concentration of the activator reaches a steady state where its production exactly matches its diffusion to the surroundings.

LSTM networks also achieve good results for each dataset, and the performance is close to the RD model ratio value. The slight difference between each dataset is explained by the complexity of the different systems used to generate sets of sequences for the training. Sine waves are deterministic, while the Lorenz and Van der Pol systems introduce "chaos" into the trajectories. The structure of gates in LSTM shows an active regulation between forgotten input and new input in Fig. 2 (green, red, and purple dashed lines).

Table 1. Reaction Metric, Diffusion Metric and Mean Dynamic Balance Ratio of each model and set of parameters/training.

	Parameters/Dataset	$\mathcal{M}_{\text{react}}$	$\mathcal{M}_{\text{diff}}$	R
RD	Spot Regime	0.3589	0.3551	1.0107
RD	Stripe Regime	0.3742	0.3707	1.0093
LSTM	Sine Waves	1.6732	1.6204	1.0325
LSTM	Lorenz System	1.5836	1.5254	1.0381
LSTM	Van der Pol	1.8933	1.9791	0.9566
DDPM	MNIST	0.2143	3.0274	0.0708
DDPM	FashionMNIST	0.2152	3.0277	0.0711
DDPM	CIFAR10	0.3688	5.2508	0.0702

Fig. 2. Dynamic Balance Ratio over time

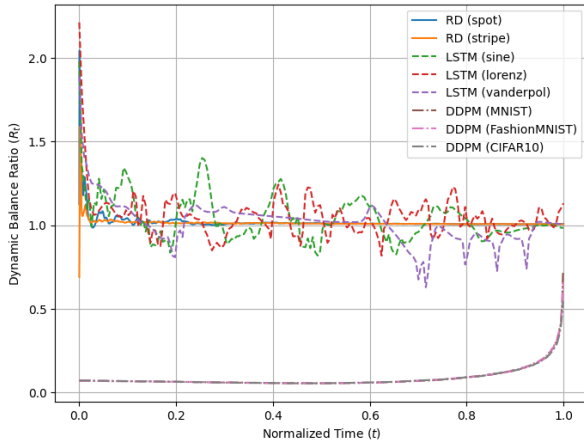**Fig. 3**. Final value of the Dynamic Balance Ratio

Unlike the flat converged line of RD, the LSTM ratio constantly oscillates around the equilibrium at $R = 1$. The peaks at $R > 1$ and the increasing parts of the ratio correspond to the moments where the input gate encodes new information to the network, while the valleys at $R < 1$ and decreasing parts correspond to the stable periods where the forget gate erases selected past information. These observations emphasize the assumption of the interplay between input and forget gates that respectively play the role of reaction and diffusion in Reaction-Diffusion systems.

In contrast and despite the hypothesis done previously, DDPMs trained on MNIST, FashionMNIST, and CIFAR10 show poor results with a ratio significantly inferior to 1. In this case, the overall magnitude of the diffused noise is approximately 14 times stronger compared to the magnitude of the learned drift. As Table 1 shows the overall Dynamic Balance Ratio, it is more interesting to look at each time step of the ratio. Fig. 2 reveals a striking phase transition behavior of the DDPM between reaction and diffusion (brown, pink, and gray dashdot lines). For $t = 0$ to $t = 0.8$, the ratio remains extremely low and is dominated by noise: it describes the moments where the generated structure is still governed by noise and no pattern is visible. However, as $t \to 1$, the ratio increases significantly and rapidly approaches the equilibrium $R = 1$. This result is expected as DDPMs do not maintain structure but emerge with a new one from pure noise. At the final time step for every dataset, the DDPM value of the ratio approaches the equilibrium. More complex images require larger time steps to generate. Hence, the final ratio in Fig. 3 for CIFAR10 is lower compared to MNIST. Finally, the sharp rise of the ratio at the end represents the emergence of fine details in the structure, moving the system from an entropic regime towards the balanced regime found in RD and LSTM.

The values of the ratio for DDPMs being lower than the equilibrium can be explained by the lack of training of the models and not enough time steps used for the image generations.
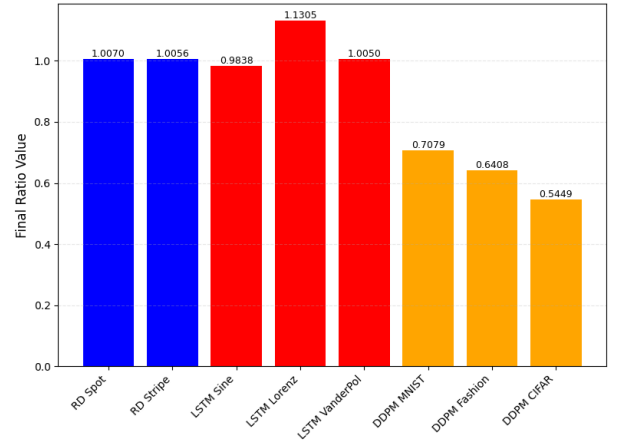
6. CONCLUSION

Our analysis demonstrates that the Turing Reaction-Diffusion principle is a universal lens for understanding the dynamics in LSTM networks and DDPMs, but it manifests in three distinct modes. Reaction-Diffusion models show static stability to create patterns. The forces of activation and inhibition converge to a steady state, locking the system into fixed spatial patterns. LSTM networks exhibit dynamic stability to create sequences by oscillating around the equilibrium ratio. The networks alternate between reactive encoding and diffusive inhibition from their gate mechanism. DDPMs manifest phase transition to create structured and detailed images from pure noise. However, some hardware limitations led the results of DDPMs to be inaccurate of the theoretical assumptions. Theoretically, a DDPM reverse process is governed by the diffused noise on the one hand and then the learned drift replaces it to make a detailed image emerging [4]. Hence the Dynamic Balance Ratio of DDPMs is an increasing and monotonic function.

REFERENCES

[1] S. Kondo and T. Miura, "Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation", *Science*, Vol 329, No. 5999, pp. 1616-1620, 2010.

[2] A. Turing, "The chemical basis of morphogenesis", *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol 237, No. 641, pp. 37-72, 1952.

[3] S. Hochreite and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, Vol 9, No. 8, pp. 1735–1780, 2018.

[4] J. Ho, A.Jain and P. Abbeel, "Denoising Diffusion Probabilistic Models", *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, 2020.

[5] J. E. Pearson, "Complex patterns in a simple system", *Science*, Vol. 261, No. 5118, pp. 189–191, 1993.

[6] C. Bishop, H. Bishop, *Deep Learning: Foundations and Concepts*, 2023.