Theory of biochemical information processing with transients

Manish Yadav

Fakultät für Chemie und Chemische Biologie
Technische Universität, Dortmund

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

November 2022

Vorgelegt im November 2022
von Manish Yadav

Gutachter:
Priv. Doz. Dr. Aneta Koseska
Prof. Dr. Philippe Bastiaens

The work presented in this dissertation was performed in the group of
PD Dr. Aneta Koseska at Cellular Computations and Learning (CCL) Lab, Max
Planck Institute for Neurobiology of Behaviour - CAESAR,
Ludwig-Erhard-Allee 2,
D-53175 Bonn, Germany.

Manish Yadav was affiliated to the International Max Planck Research School for
Living Matter (IMPRS-LM), Dortmund, Germany.

"We have a closed circle of consistency here: the laws of physics produce complex systems, and these complex systems lead to consciousness, which then produces mathematics, which can then encode in a succinct & inspiring way the very underlying laws of physics that gave rise to it."

Roger Penrose

Abstract

Cells in tissues and organisms operate in dynamic environments, continuously sensing and responding to time-varying chemical signals. In order to accurately interpret the complex information from their environment, biochemical networks in single cells actively process these extracellular signals in real-time. The current concept of biochemical computations places a strong focus on attractor based information processing in cells. Recent studies however have shown that cells generate completely opposite phenotypic responses depending upon frequency of the growth factor, independent of growth factor identity. This breaks down the steady-state description of biochemical information processing. Therefore, we propose to describe biochemical networks embedded in non-stationary environments as non-autonomous systems whose solutions are the dynamic input-dependent trajectories. We show that memory arising through metastable states will enable the system to integrate time-varying signals such that, inputs resulting in different phenotypic responses will be uniquely encoded in phase-space trajectories. The extracellular information of different phenotypes is spread throughout the large signaling networks and represented by characteristically different classes of phase-space trajectories. This encoded information will further be decoded downstream by early response genes (ERG) in real-time, where we show that the feed-forward structure of ERG is sufficient for this task.

Zusammenfassung

Zellen in Geweben und Organismen bewegen sich in einer dynamischen Umgebung, in der sie ständig zeitlich veränderliche chemische Signale wahrnehmen und darauf reagieren. Um die komplexen Informationen aus ihrer Umgebung genau zu interpretieren, verarbeiten biochemische Netzwerke in einzelnen Zellen diese extrazellulären Signale aktiv und in Echtzeit. Das derzeitige Konzept der biochemischen Berechnungen legt den Schwerpunkt auf die auf Attraktoren basierende Informationsverarbeitung in Zellen. Jüngste Studien haben jedoch gezeigt, dass Zellen je nach Stimulationsfrequenz mit einem Wachstumsfaktor völlig entgegengesetzte phänotypische Reaktionen hervorrufen können, unabhängig von der Identität des Wachstumsfaktors. Dies bricht die Beschreibung des stationären Zustands der biochemischen Informationsverarbeitung auf. Daher schlagen wir vor, biochemische Netzwerke, die in nicht-stationäre Umgebungen eingebettet sind, als nicht-autonome Systeme zu beschreiben, deren Lösungen die dynamischen, input-abhängigen Trajektorien sind. Wir zeigen, dass das durch metastabile Zustände entstehende Gedächtnis das System in die Lage versetzt, zeitlich veränderliche Signale zu integrieren, so dass Eingaben, die zu unterschiedlichen phänotypischen Reaktionen führen, eindeutig in Phasenraum-Trajektorien kodiert werden. Die extrazelluläre Information der verschiedenen Phänotypen ist über weitläufige Signalnetzwerke verteilt und wird durch charakteristisch unterschiedliche Klassen von Phasen-Raum-Trajektorien dargestellt. Diese kodierten Informationen werden von frühen Reaktionsgenen (ERGs) in Echtzeit dekodiert, wobei wir zeigen, dass die Feed-Forward-Struktur von ERGs für diese Aufgabe ausreichend ist.

Contents

List of Figures

List of Tables

List of Abbreviations

RTK	**R**eceptor **T**yrosine **K**inase
GPCR	**G**-**P**rotein **C**oupled **R**eceptor
GTP	**G**uanosine **Tri****p**hosphate
GDP	**G**uanosine **Di****p**hosphate
EGFR	**E**pidermal **G**rowth **F**actor **R**eceptor
PTP	**P**rotein **T**yrosine **P**hosphatase
ER	**E**ndoplasmic **R**eticulum
ODE	**O**rdinary **D**ifferential **E**quation
LSM	**L**iquid **S**tate **M**achine
ED	**E**uclidean **D**istance
ERG	**E**arly **R**esponse **G**ene
IPI	**I**nter-**P**ulse **I**nterval
BC	**B**etweenness **C**entrality
CC	**C**lustering **C**oefficient
RNN	**R**ecurrent **N**eural **N**etwork
ESN	**E**cho **S**tate **N**etwork
LSM	**L**iquid **S**tate **M**achine
SP	**S**eparation **P**roperty
AP	**A**pproximation **P**roperty

Chapter 1

Introduction

The current understanding of cell signaling had its origins in the pioneering work of Earl W. Sutherland, whose research led to a Nobel Prize in 1971. Sutherland's work suggested that cellular communication can be dissected into three stages: detection, transduction, and response [93]. A chemical signal is "detected" when the signaling molecule binds to a receptor protein located at the surface of a cell. The binding of the signaling molecule changes the activity state of receptor protein, initiating the process of transduction. The transduced signal finally triggers a specific cellular response such as migration, proliferation, apoptosis or differentiation of cells.

1.1 Extracellular information processing in cells

In multi-cellular organisms, the extracellular environment is the space outside the plasma membrane of a cell. The extracellular space consists of metabolites, ions, proteins and other macro-molecules secreted by different cell types, which further affect the functionality of cells. The signals present in the extracellular environment regulate many important process including cellular proliferation, adhesion, migration, differentiation and tissue homeostasis [46]. Cells operating in dynamic environments such as that of a developing organism or an adult tissue, continuously sense and respond to chemical signals that vary in space and time. In order to accurately interpret the complex information from their environment, the biochemical networks in single cells sense the extracellular signals in real-time via membrane-bound receptors. These receptors are organized on the plasma membrane of cells and act as first receivers of the extracellular signals. Receptors in the monomeric form interact with the membrane-bound proteins like phosphatases, which are part of the spatially-distributed receptor networks that modulate the response of receptors depending on the overall topology of the network [51].

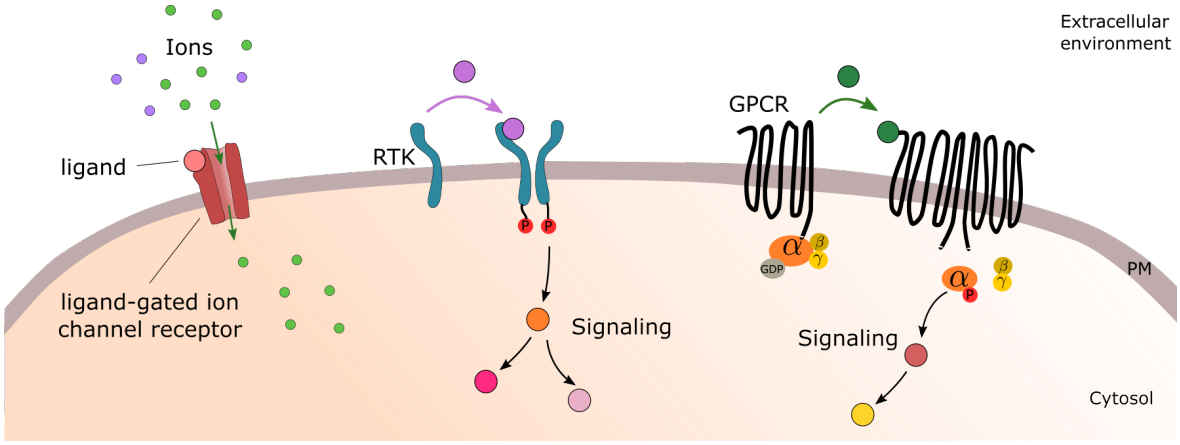

Figure 1.1: **Cell membrane-bound different type of receptors.** *(Left)* Ligand-gated ion channel (red) with an open channel allowing the flow of specific ions (green). *(Center)* In receptor tyrosine kinase (RTK), binding of a ligand (purple) causes two receptor monomers to form a dimer, resulting in phosphorylation of each monomer triggering signal transduction by activating downstream signaling proteins. *(Right)* In G-protein coupled receptor (GPCR), dimerization leads to activation and release of heterotrimeric G-protein complex (α, β and γ subunits) that relays signal from ligand to intracellular proteins. PM: Plasma membrane.

The complexity in the understanding of regulatory networks within cellular functionality increases with the intricate molecular details from receptor sensing to response of a cell. Therefore, a detailed point of view is required to investigate dynamical properties of receptor networks giving rise to complex responses. The extracellular signal sensing initiates at the level of cell surface. Membrane-bound receptors are the proteins that facilitates the interaction of cells with extracellular environment and play an essential role in translating extracellular signals into different cellular responses. Membrane-bound receptors are mainly divided by structure and function into 3 types: ion channel receptors, Receptor Tyrosine Kinases (RTKs) and G protein-coupled receptors (GPCRs) [8, 93, 3].

A ligand-gated ion channel is a type of cell membrane receptor containing a region that acts as a "gate" (Fig. 1.1 *(left)*). When a signaling molecule binds as a ligand at a specific site on the extracellular part of the ion channel receptor, the shape of the receptor changes thereby opening a channel. The ion channel allows or blocks the flow of specific ions, such as Na^+ or Ca^{2+}. Ion-channels are involved in rapid synaptic signaling between nerve cells and other electrically excitable target cells such as muscle cells.

In GPCRs, each receptor is associated with an intracellular protein called a G-protein. G-proteins are heterotrimeric, meaning that they contain three distinct subunits (α, β, and γ subunits). The binding of signaling molecule to the receptor results in conformational change of the G-protein, which releases GDP on the α subunit and

binds GTP, resulting in the G-protein being activated [99, 36]. The activation of G-protein can change the concentration of one or more small intracellular mediators, also known as 'second messengers' (Fig. 1.1 *(right)*). These intracellular mediators act in turn to alter the behavior of yet other signaling proteins in the cell and therefore, such interactions triggers a signal transduction by activating the intracellular components of downstream signaling network. GPCR-based signaling systems are extremely widespread and diverse in their functions, including roles in embryonic development and sensory reception. In humans, for example, vision, smell and taste depend on GPCRs [93, 3].

RTKs are cell-surface receptors with intracellular domains that have intrinsic enzymatic activity. Signaling molecules called growth factors (or mitogens) stimulate RTKs. The dimerization of monomeric RTKs takes place when a signaling molecule binds to their extracellular domains. The intracellular domains of each RTK dimer have kinase enzymatic activity that is active when the two subunits are brought together after binding the signaling molecule on the surface of the cell (Fig. 1.1 *(center)*). Upon dimerization, the RTK completes autophosphorylation by transferring phosphate groups to tyrosine residues within the receptor itself. The phosphate groups on the intracellular domains provide docking sites for adaptor proteins that initiate an intracellular signaling cascade [56, 105]. However, RTKs have a propensity to interact laterally and form dimers even in the absence of ligand [104, 57, 69]. Abnormal RTKs that function even in the absence of signaling molecules are associated with many kinds of cancer. For example, breast cancer patients have a poor prognosis if their tumor cells harbor excessive levels of a RTK called HER2 [93, 3].

RTKs and GPCRs are the most prominent families of receptors that recognize a variety of extracellular ligands. They both are trans-membrane proteins that have extracellular domains that bind to ligand and undergo a ligand dependent canonical activation. This activation event causes structural and functional modifications to their intracellular regions and represents the first step in receptor-mediated signal transduction.

1.1.1 Biochemical information flow from extracellular environment to intracellular signaling network

The binding of a specific signaling molecule triggers the first step in the signal transduction pathway - the chain of molecular interactions that leads to a particular response within the cell. The signal-activated receptor activates the downstream signaling molecules present in the cytoplasm, which activates the early response genes (ERG) present inside the nucleus of the cell and eventually the final cellular

response is generated [93, 3]. This chain of biochemical cascades is known as signaling networks. There are several signaling networks responsible for various cellular functionality, for instance, the cyclic AMP or cAMP-dependent network whose end results include taste and tumor-promotion, and mitogen-activated protein kinase or MAPK signaling module promotes cell division and is responsible for many forms of cancer [97].

The MAPK is an important and conserved signaling module responsible for key biological processes including cell proliferation, growth, and survival. Alterations in MAPK signaling cascades are found to result in various diseases including cancer [12]. After activation of an RTK receptor, adaptor proteins bind to the phosphate groups present on the intercellular domains of the receptor. The adaptor proteins recruit the enzyme Ras, a GTPase. Active Ras (i.e., Ras-GTP) activates MAP Kinase Kinase Kinase (MAPKKK) or RAF. MAPKKK then phosphorylates and activates MAP Kinase Kinase (MAPKK) or MEK. The cascade terminates when MAPKK phosphorylates and activates MAP Kinase (MAPK) or ERK. Or simply, activated Ras activates the protein kinase activity of a RAF kinase. The RAF kinase phosphorylates and activates a MEK Kinase which phosphorylates ERK. The phosphorylation of ERK results in an activation of its kinase activity and leads to phosphorylation of its many downstream targets involved in regulation of cell proliferation [12]. However, experimental studies further suggest that MAPK signaling cascade can be activated by interacting with multiple receptors present on the plasma membrane of cells that becomes active upon binding to their specific growth factors [79, 75].

1.2 Cellular response to sustained stimulus

Cells selectively and differently responds to specific signaling entities present in the extracellular environment which consists of a myriad of different signaling molecules, ligands, ions and cytokines. This is known as the *specificity* feature of biochemical information processing where a cell generates different responses specific to receptor stimulation [93]. Such a specific response behavior of cells to different extracellular signals can be experimentally probed by exposing cells to a sustained stimulus of different growth-factors and their response in the form of different phenotypes are observed. For instance, experiments show that PC12 cells proliferates upon stimulation with EGF, while differentiation has been reported upon stimulation with NGF [75, 79]. This shows the ability of cells to selectively sense and process different extracellular signals into specific biological responses.

1.2.1 Same signaling proteins can generate multiple cellular responses

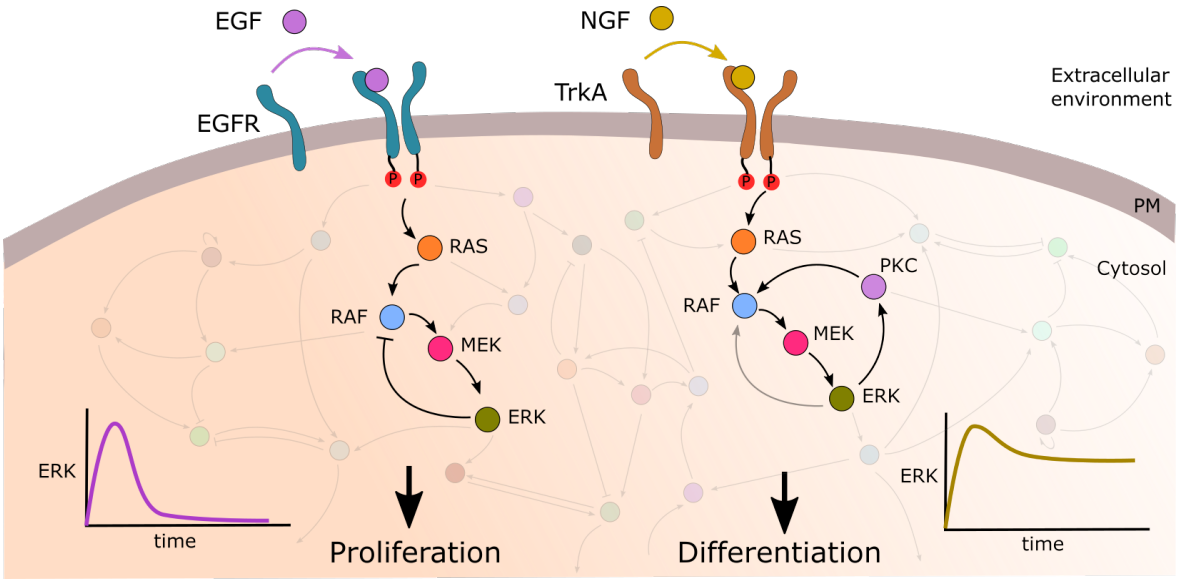

Figure 1.2: **EGF and NGF activated MAPK signaling module.** MAPK signaling module with a negative feedback from ERK to RAF for sustained EGF stimulation resulting in proliferation with a transient ERK activity over time *(left)* and a positive feedback for sustained NGF stimulation resulting in differentiation with a sustained ERK activity over time *(right)*.

Specific responses of cells to different signaling molecules have been explained using different receptor types. However, experimental studies suggests that cells recruit same signaling module to generate multiple cellular responses [12, 75, 79]. A classic example is the MAPK signaling module which is activated by different growth factors resulting in opposite cellular phenotypes.

In a study with PC12 cells, Santos et al. [79] demonstrated the details of the MAPK signaling module, which is activated through the receptor tyrosine kinases TrkA and epidermal growth factor receptor (EGFR) respectively by two different stimuli NGF and EGF. With sustained NGF stimulation, ERK has been reported to show a sustained activation, whereas sustained EGF induced a transient ERK response (schematic in Fig. 1.2). The question of different ERK dynamics originating from receptor level activity has been explained using growth-factor identity dependent manifestation of different topological features in MAPK signaling network connectivity. Using modular response analysis, (MRA) Santos et al. [79] demonstrated that in the case of EGF stimulation a negative feedback is established from ERK to RAF, explaining the transient activation of ERK for a sustained EGF stimulation. On the other hand, the sustained ERK activity is explained using a positive feedback between ERK and RAF. A positive feedback loop generates a bistable regime and Santos et al. reported such bistablity, with a switch-like ERK activation for NGF

dose-response. Santos et al. [79] further demonstrated that EGF stimulation can also result in cell differentiation. A positive feedback mechanism from ERK to RAF involves protein kinase C (PKC). Therefore, EGF stimulation in conjunction with PKC activation by phorbol-12-myristate-13-acetate (PMA) gave rise to a sustained ERK activation profile and cells differentiated as if stimulated by NGF.

The notion of conjugating cellular response with the identity of growth-factor has been questioned by Santos et al. [79] suggesting the importance of information sensing and processing mechanisms employed by membrane-bound receptor networks for generating different cellular responses. However, under the current description, different cellular responses are explained using context-dependent signaling modules with the assumption that they function in isolation from the rest of the proteins present in the intra-cellular signaling network.

1.3 Biochemical network modules as autonomous systems

Biochemical networks including membrane-bound receptor networks and intracellular signaling networks are commonly described using autonomous systems [40]. An autonomous differential equation is a system of ordinary differential equations (ODE) that does not explicitly depend on the independent variable. Autonomous systems are also called time-invariant systems, as time being an independent variable [87]. An autonomous system can be represented by a general Eq. 1.1, where x takes values in n-dimensional Euclidean space and t is time.

$$\frac{dx}{dt} = f(x(t)) \tag{1.1}$$

Since autonomous systems are time-independent, such systems reach steady state in a finite time. In other words, the partial derivative of an ODE is zero with respect to time and it's solution called steady-state can be obtained using:

$$\dot{x} = 0 \quad \Rightarrow \quad f(x(t)) = 0 \tag{1.2}$$

Not all steady-state solutions are equivalent. Some may be quickly abandoned if the system is slightly perturbed; others may function as an attractor of the system, even under a large perturbation. The stability of an equilibrium relates to the tendency of the system to return to its position when perturbed. The number of steady-states depends on the network topology of the underlying system. A system can change

its activity by hopping from one steady-state to another under the influence of an external signal, shown schematically in Fig. 1.3. However, the state of the system remains stable in this final steady-state unless it is acted upon by another external stimulus and jump to another state, if available. It is also possible that only a single equilibrium point is present, in such case the system is poised to stay in that only steady-state, even after being perturbed by an external stimulus. We elucidate these two cases in the following example of RTK coupled to phosphatase (PTP).

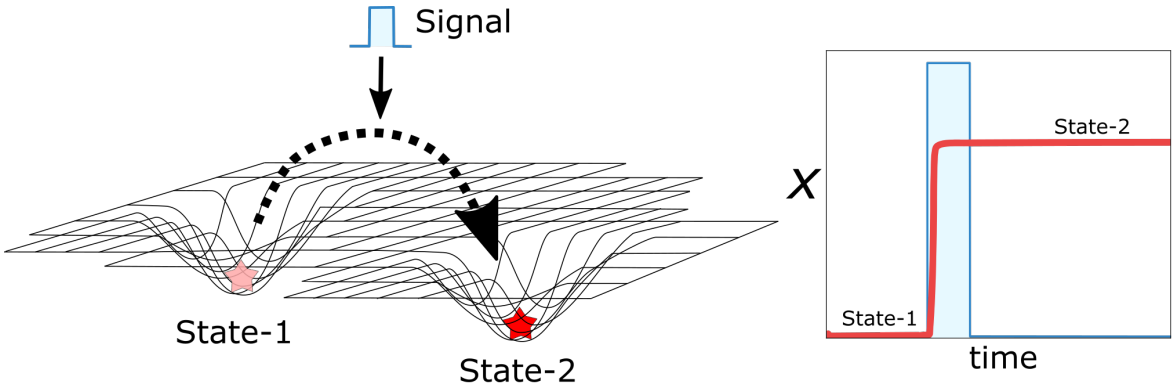

Figure 1.3: **System hopping between steady-states under the influence of external signal.** (*left*) Schematic of potential landscape with equilibrium states. An external signal triggers the system to jump from stable steady-state-1 to steady-state-2. (*right*) Activity (x) of system as state changes over time.

Rich dynamical behavior of membrane-bound RTKs appears when the activity of PTP is coupled to that of RTK. Studies suggest that PTP (PTPRG) interacts with an RTK (EGFR) in a double-nagative feedback loop or generating a toggle switch (Fig. 1.4 (a)) [6, 51, 84]. All the equilibrium points (stable and unstable) of the system can be identified using bifurcation analysis with respect to a system parameter. Bifurcation analysis of the RTK-PTP system obtained using the XPPAUT software [26] reveals a bistable regime in the system with a total of three solutions. Two are dynamically maintained low and high RTK phosphorylation states, both of which are stable and can coexist in a given parameter interval, and an unstable solution that acts as a threshold in the system (Fig. 1.4 (b)). In which stable solution the system will end up is directly related to the initial conditions, that is, the initial phosphorylation level of the receptor. In a cellular context, this might be dependent on the history of the cell in terms of prior growth factor-induced signaling states. If the system is initially in the low phosphorylation states, it can reach the high phosphorylation states by external stimulus (Fig. 1.4 (d)). However, since this is the stable-state, the system stays in this high phosphorylation state even if perturbed by an external

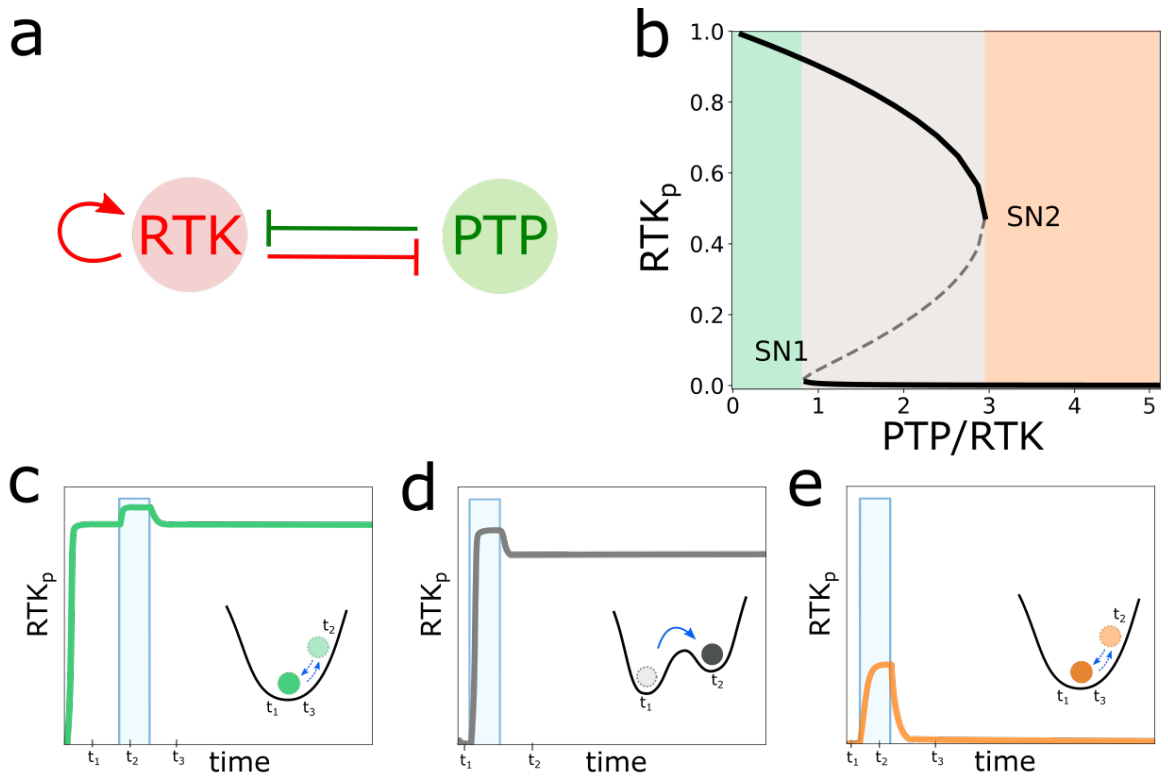

Figure 1.4: **Different steady-state regimes for RTK and PTP interation in a double-negative feedback manner.** (a) Schematic of RTK and PTP interacting in a double-negative feedback manner. (b) Bifurcation diagram of RTK and PTP toggle-switch showing response of RTK. Time-profile of RTK activity in (c) high-monostable, (d) bistable and (e) low-monostable regimes. (c) to (e) are the RTK time profiles in the presence of an input pulse (blue). Inset: schematic of system potentials with equilibrium points in the respective system parameterization regimes.

stimulus. This is regarded as a long-term memory of stimuli that a cell might have sensed sometime in the past [30].

Apart from the bistable regime, RTK-PTP system also exhibits two monostable regimes with high and low phosphorylation states for respective system parameters. In terms of equilibrium points, monostable means only single steady-state is present and system will remain in this equilibrium state irrespective of initial conditions. In the high monostable regime, system abruptly reaches the high phosphorylation state marking the irreversible pre-activation state even in the absence of external stimulus (Fig. 1.4 (c)). In the low monostable regime, an external stimulus can perturb the system but it returns back to its initial low phosphorylation state as soon as the stimulus is removed (Fig. 1.4 (e)).

The presence of a fixed number of steady-states depending on the underlying network topology has been used to explain the features of biochemical networks by approximating the biological system as autonomous system. However, cells continuously sense and respond to dynamic extracellular environment meaning that

the underlying biochemical network parameters and external environment of the biological system are dynamic over time and space. Therefore, approximation of biological systems as autonomous might not be demonstrating the full dynamical picture and the underlying information processing mechanisms of complex biological systems.

1.4 Mechanisms of cellular responsiveness to time-varying extracellular signals

Apart from sustained stimulus, experimental studies suggest that cells can also sense time-varying extracellular signals and generate various responses. Recent experimental study has shown that cell surface receptor networks can sense and integrate dynamic growth-factor signals [84]. Furthermore, cells generate various phenotypic responses depending on the temporal profiles of growth-factor signals [78]. In this section, we will discuss the mechanisms explaining cellular responsiveness to dynamic extracellular signals. Furthermore, we will also discuss if the autonomous framework used for describing biochemical systems will be able to explain cellular responsiveness to time-varying extracellular signals as well.

1.4.1 EGFR network as a complex signal sensing system

EGFR has been experimentally shown to sense the presence of extracellular growth factors, and also to interpret the complex dynamic growth factor patterns that lead to diverse and functionally opposed cellular responses including proliferation, survival, apoptosis, differentiation, and migration [52, 105]. A recent experimental study of EGFR signaling has revealed that the major protein tyrosine phosphatases (PTPs) that dephosphorylate EGFR are present on the plasma membrane and endoplasmic reticulum. Vesicular trafficking of phosphorylated EGFR unifies the interactions with these PTPs into a network architecture whose sensing dynamics dictates responsiveness to time-varying growth factor signals [84].

Experimentally established EGFR network consists of multiple feedback interactions between EGFR and PTPs that are spatially segregated, schematically shown in Fig. 1.5 (a). In the study, dynamic signatures of EGFR phosphorylation response to EGF dose were analyzed to identify the type of underlying feedback motif and the

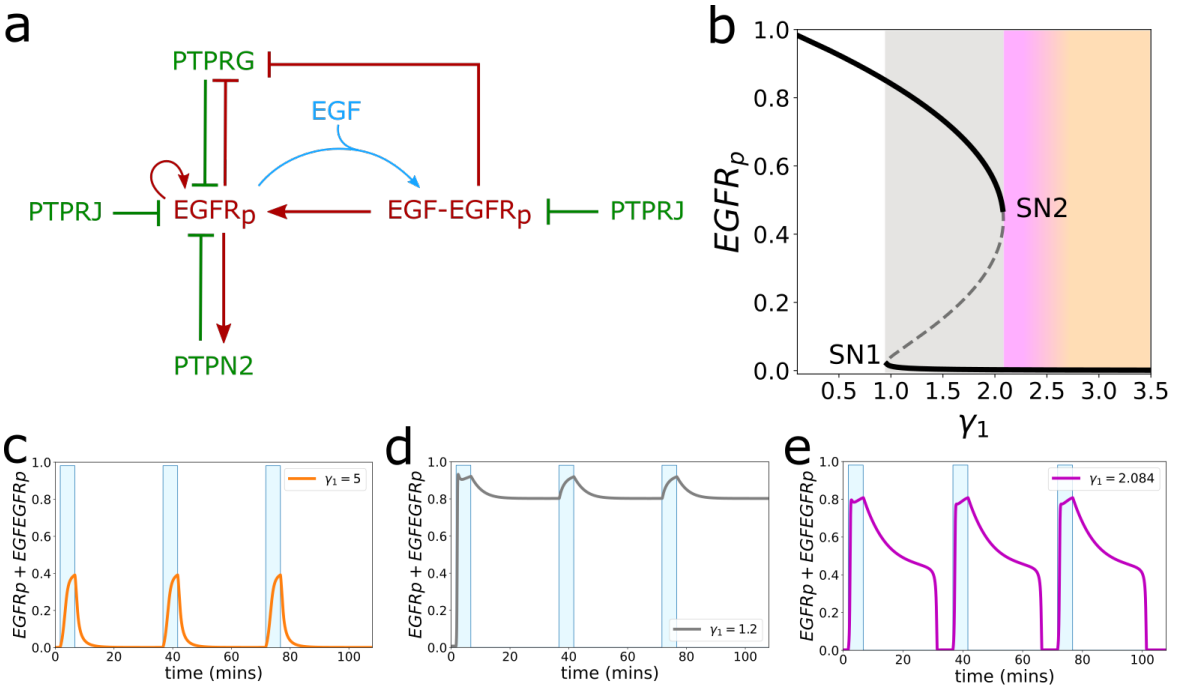

Figure 1.5: **Dynamics of EGF receptor network.** (a) Schematic diagram of EGF receptor network established through EGFR trafficking dynamics. (b) Bifurcation diagram of $EGFR_p$ in the absence of input with respect to γ_1 being the bifurcation parameter. γ_1 is proportional to $PTPRG_T/EGFR_T$. (c), (d), (e) Temporal profiles of EGFR response ($EGFRp + EGFEGFRp$) for an input signal of 3 min pulses followed by 30 min inter-pulse-intervals, when the system is respectively organized in monostable regime (orange), bistable regime (grey) and near the saddle-node ($SN2$) bifurcation point (magenta).

findings were validated using genetic perturbations. It was identified that ligandless phosphorylated EGFR (or just EGFRp) interacts with two PTPs, namely PTP Receptor Type G (PTPRG) and PTP Receptor Type J (PTPRJ) at the plasma membrane of MCF7 cells. Active PTPRG interacts with EGFRp in a double-negative feedback manner, creating a EGFRp-PTPRG toggle switch. On the other hand, PTPRJ constitutively interacts with both ligandless and liganded EGFR, and dephosphorylates them with a negative regulation. Both ligandless and liganded EGFR also interacts with an endoplasmic reticulum (ER) membrane bound PTPN2 with a negative feedback. This EGFR receptor network is modeled using the law of mass action which enabled the use of bifurcation analysis to comprehend the mechanism behind the sensing of dynamic EGF signals.

A bifurcation diagram of EGFRp response with respect to a system parameter (γ_1) proportional to $PTPRG_T/EGFR_T$ reveals multiple dynamical regimes as shown in Fig. 1.5 (b). A bistable regime (grey) exist between two monostable regimes which are characterized by basal (orange) and high (white) EGFR phosphorylation. This

region is marked by the presence of two saddle-nodes, SN1 and SN2 and the co-existence of two stable states (solid lines) that represents high and low phosphorylation state together with an unstable state (dashed line) acting as a separatrix. Parameter organization in the monostable (low) regime, an EGF signal of temporal profile 3 minute pulses followed by 30 minute inter-pulse interval (IPI) stimulation elicit a transient, reversible response that could not activate the system Fig. 1.5 (c). For organization in the bistable region, EGFR can be in either in a basal or a high phosphorylation state in the absence of any stimulus. When starting from the basal EGFR phosphorylation state, the response shows abrupt activation in presence of the first EGF pulse which sustained even after signal removal corresponding to the irreversibility in the response Fig. 1.5 (d).

Both theory and experiments demonstrated that sensing growth factor signal changes cannot be realized when the system organization is in either the bistable or the monostable regime. The irriversible nature of the bistable regime limits the responsiveness of system to temporal growth factor changes by switching it to a high EGFR phosphorylation state after the initial growth factor stimulus that leaves the system in a non-responsive state (Fig. 1.5 (d)). Organization in the monostable (low) region, on the other hand, leaves the system with a linear response that follows the growth factor profile, thus being unable to explain history-dependent response (Fig. 1.5 (c)). However, the authors have experimentally shown that the EGFR network is organized near the saddle-node bifurcation (SN2) (Fig. 1.5 (b), magenta region) for optimally sensing the time-varying external environment. As shown in Fig. 1.5 (e), when the EGFR network is critically organized near the saddle node, the EGFR response profile shows an activation followed by a transient activity before the system returns to basal state for each pulse of EGF. Such a short transient activity in EGFR phosphorylation has been shown to follow the experimental observations in MCF7 cells [84]. In a different theoretical study, Stanoev et al. [85] showed that such a critical organization near SN2 bifurcation generates a 'ghost' of saddle-node which is responsible for producing high receptor activity state for a limited amount of time after the removal of growth-factor stimulus. This explained the experimentally observed transient EGFR activity was the dynamic 'transient' memory enabled by the 'ghost' of saddle-node which endows EGFR network to sense and process dynamic extracellular environment. The authors further showed that such a dynamic transient memory enables unique processing of different time-varying growth factor signals.

1.4.2 Dynamic transient memory arising from 'criticality' enables integration of time-varying growth factor signals

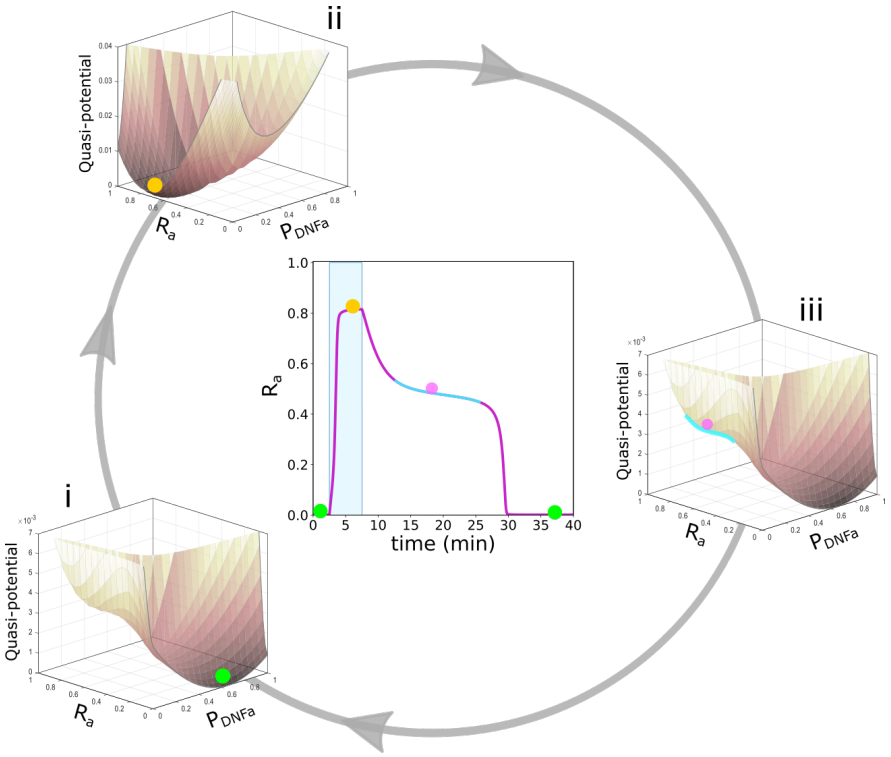

Figure 1.6: **Criticality enabled 'transient' memory in receptor activity.** (*Center*) Response of a receptor (R_a) interacting with a PTP (P_{DNFa}) in a double-negative feedback manner for the system organization near the saddle-node ($SN2$) bifurcation (Fig. 1.4 (a, b)). An external input pulse generates transitions in quasi-potential landscape of system: (i) low-monostable state, (ii) high-monostable state and (iii) metastable state (cyan).

Stanoev et al. [85] proposed a minimal receptor network consisting of RTK (R_a) and PTP ($P_{DNF,a}$) interacting in a double-negative feedback manner i.e. generating a toggle switch (Fig. 1.4 (a)). The effective input for the cells is the fraction of ligand-bound receptors (LR_a) that reflects the extracellular ligand concentration which directly receives the external input (L_T) from the extracellular environment. Following the law of mass action, the equation for this system can be written as:

$$\begin{aligned} \frac{dR_a}{dt} &= R_T(R_i(\alpha_1 R_i + \alpha_2 R_a \alpha_3 L R_a) - \hat{\gamma}_{DNF} P_{DNF,a} R_a) \\ \frac{dP_{DNF_a}}{dt} &= k_1(P_{DNF,i} - k_{2/1} P_{DNF,a} - \hat{\beta}_{DNF} P_{DNF,a}(R_a + LR_a)) \end{aligned} \tag{1.3}$$

This minimal network motif is modeled using law of mass action. Both the receptor and the deactivating enzyme have active (R_a, P_{DNF_a}) and inactive (R_i, P_{DNF_i})

states, and their state transition rates are described by the model equations. Therefore, mass is conserved in the system and the total protein concentrations of both species (R_T, P_{DNF_T}) are constant parameters. This allows $R_i = 1 - Ra$ and $P_{DNF_i} = 1 - P_{DNF_a}$ to be expressed as fractions from the total protein concentrations. For binding/unbinding of ligand L_T to modulate LR_a and additional term $-k_{on}R_aL_T + 0.5k_{off}LR_a$ is added to the equation of R_a, and the dynamics of LR_a was modelled with $\frac{dL_a}{dt} = K_{on}(R_a + R_i)L_T - K_{off}LR_a$.

Bistability in receptor activity is exhibited between two saddle-node bifurcation points (SN1 and SN2). A 'transient' memory is observed in the receptor activity R_a when the receptor network is critically organized near the saddle-node bifurcation point $SN2$ (magenta region in Fig. 1.5 (b)). Due to the vicinity to a saddle node, the emergent metastable state gives rise to transient temporal memory in receptor activity even after the removal of growth factor stimulus (Fig. 1.6 (b), center). A quasi-potential of the system revealed that when the system is near SN2 bifurcation point, there are two equilibrium points: monostable (low) and a metastable state, which can be thought of as a 'ghost' of saddle-node $SN2$. The "ghost" sucks trajectories, delaying the flow in phase space, rendering a metastable dynamics. The system essentially starts from the monostable (low) state in the absence of an external stimulus (Fig. 1.6 (i), green dot). In the presence of a stimulus, only monostable (high) state is available (Fig. 1.6 (ii), orange dot). The dynamics become interesting when the stimulus is removed. The receptor activity does not abruptly falls to monostable (low) state but, rather it passes through the metastable state (Fig. 1.6 (iii), magenta dot). As the name suggests, this 'ghost' of $SN2$ slows down the dynamics of receptor activity by creating a bottleneck with infinitesimally close nullclines [87]. Therefore, a single pulse generates a full transition: $(i) \rightarrow (ii) \rightarrow (iii) \rightarrow (i)$, during which the number of available steady-states of the system changes. There will be a series of such transitions in quase-potential landscape in case of multiple pulses. However, the sequence of such transitions might vary depending on the length of dynamic memory and the inter-pulse-intervals.

1.4.3 Self-organized positioning at criticality via vesicular trafficking

The EGFR network is critically organized near a saddle-node bifurcation and such a critical positioning uniquely endows the receptor network with the necessary features for processing dynamic extracellular signals [84]. However, the mechanism of such a critical organization of the receptor network is unknown. One can hypothesize that either the receptor networks have been evolved in such a way that the

total receptor concentration is positioned at 'criticality' or it is dynamically maintained. The latter case has been proposed and investigated by Stanoev et al. [85] using molecular features of the spatially extended EGFR network.

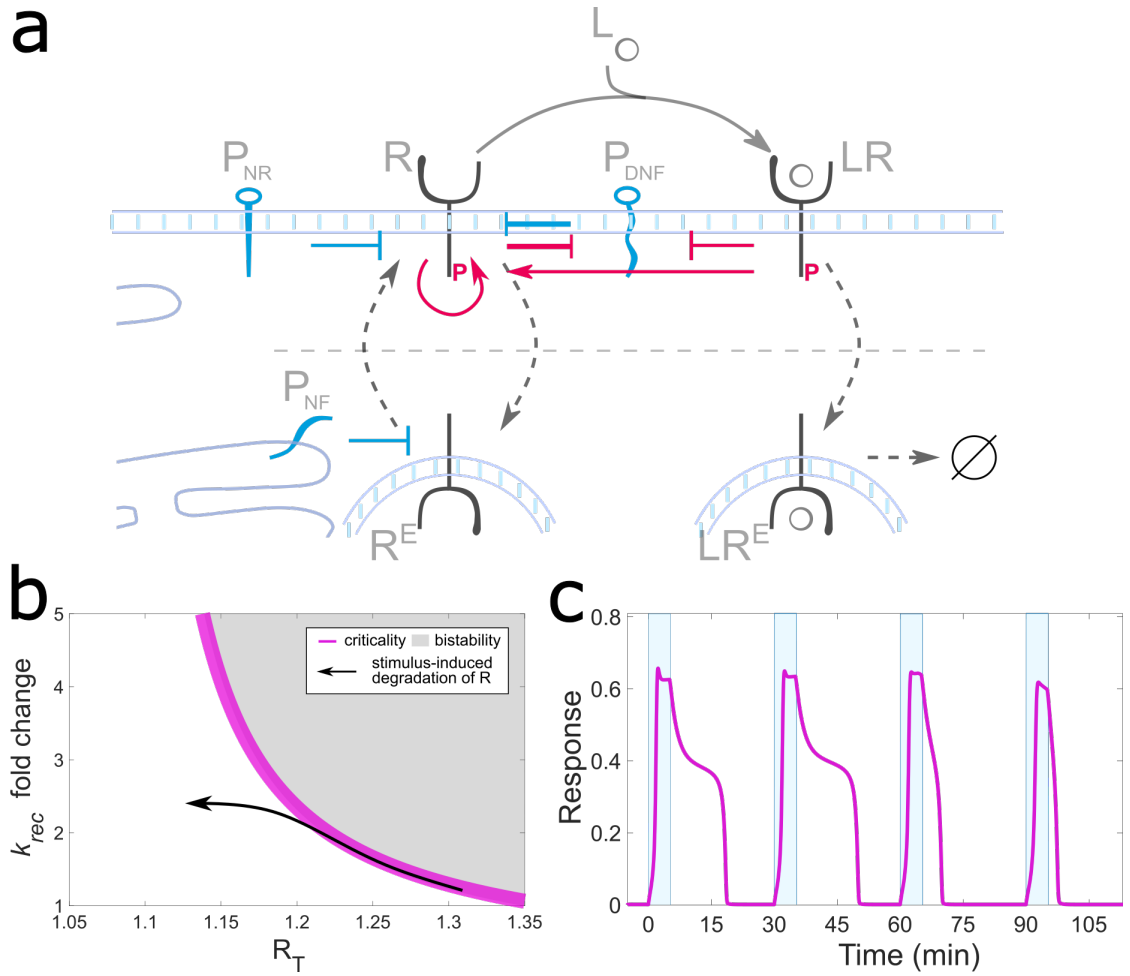

Figure 1.7: **Dynamical mechanism of self-organization at criticality of EGFR network.** (a) At the plasma membrane, ligandless EGFR (R) is coupled to PTPRG (P_{DNF}) in a double-negative feedback manner and negatively regulated by PTPRJ (P_{NR}). Activated receptors are then endocytosed in the perinuclear area (R^E), deactivated by PTPN2 (P_{NF}) and recycled back to the plasma membrane, which amounts to a spatially established negative feedback. Ligand (L) binding converts R to ligand-bound species (LR) that are internalized (LR^E) and subsequently degraded. LR promotes autocatalytic activation of R (red arrows). (b) Two-parameter (k_{rec}, R_T) bifurcation diagram depicting the positioning of the saddle-node bifurcation point SN2 (magenta curve). Bistability region is shaded with grey and monostable region in white. Black arrow represents the stimulus-induced nonlinear shift of system organization away from criticality. (c) Dynamically maintained transient memory in receptor activity generated by multiple stimulus pulses. Adapted from Stanoev et al. [85].

The EGFR activity dynamics is regulated by a four-component spatially distributed network as shown schematically in Fig. 1.7 (a). The interactions of EGFR (R) has been observed to be coupled via the vesicular trafficking with three specific membrane-associated enzymes [84]. Ligand-bound EGF receptors (LR) promote autocatalytic

activation of ligandless receptors (red arrows) [6, 16, 61, 60]. Therefore, ligand-bound EGF receptors transfer information about the dynamic extracellular environment before they are internalized and degraded (LR^E) [38, 61]. The ligandless EGF-receptors are also internalized. However, the vesicular recycling brings the internalized and deactivated ligandless EGF-receptors (R^E) back to the plasma membrane, establishing the spatially distributed EGF signal processing network.

Experimental studies have shown that ligand-bound receptors are unidirectionally internalized and degraded which decreases the EGFR concentration on the plasma membrane [61, 38]. Such unidirectional degradation of ligand-bound EGFR receptors rapidly shifts the operation of the system from critical organization into the monostable regime [84, 85]. Thus, to balance the shift of the system away from the SN bifurcation, the loss of EGFR receptors is compensated with an increase of the recycling rate of ligandless receptors. Using two-parameter bifurcation analysis Stanoev et al. showed how the position of the SN bifurcation depends on total receptor concentration R_T and receptor recycling rate k_{rec} (Fig. 1.7 (b)). Starting from critical organization (magenta), the SN positioning asymptotically approaches a minimal receptor concentration below which the receptor recycling rate can no longer sufficiently compensate for the loss of receptors from the membrane. As a result of nonlinear decrease in R_T, the system would effectively remain in the vicinity of the SN for several subsequent growth factor pulses (Fig. 1.7 (c)). The duration and dose of growth-factor pulses determines the number of pulses required to bring the system to the limit when receptor network can no longer sense and process the dynamic environment [85].

Therefore, the spatially extended EGFR network which is critically organized near a saddle-node bifurcation has been shown using experiments and numerical simulations, to sense and process a dynamic extracellular environment. Mechanistically, these studies reveal that EGF-receptor networks use a dynamic 'transient' memory enabled by the 'ghost' of a SN bifurcation to process such time-varying growth-factor signals.

1.4.4 Growth-factor signal temporal frequency dependant opposite cellular responses

The sensing of time-varying extracellular EGF signals by critically organized EGFR network has been experimentally shown by Stanoev et al. [84] in MCF7 cells. The physiological inference of dynamic growth-factor sensing behavior of cells has been demonstrated by H. Ryu et al. [78] using experiments with PC12 cells. It has been reported that stimulating PC12 cells in a microfludic device with various frequencies

of a growth-factor signal generates opposite phenotypic responses. For instance, signals of 25 ng/ml EGF with a consecutive temporal profile of 3 minutes of pulse followed by 60 minutes interval (or 3'60' in short) and 3'3' results in proliferation. Stimulation with same dose of EGF but with signals of frequencies 3'10' and 3'20', on the other hand, result in differentiation in PC12 cells. In the classic experiments with PC12 cells, proliferation has been previously reported using sustained EGF stimulation while on the other hand, differentiation was observed for sustained NGF stimulation [79]. Experiments with time-varying growth-stimulus, therefore, provide further insight into the dynamic extracellular signal processing capabilities of biochemical networks. PC12 cells are perfect to test the sensing and responsiveness of cells to different frequencies of EGF and NGF stimulations, as this cell line contains EGFR and also receptor tyrosine kinases (TrkA) [37]. H. Ryu et al. [78] reported varying differentiation ratios and also no differentiation at all in response to stimulation of PC12 cells with different frequencies of 50 ng/ml NGF signals, summarized in Table. 1.1.

Table 1.1: Phenotypes generated by temporal frequencies of EGF and NGF signals.

	Proliferation	Differentiation
EGF	[3'3', 3'60']	[3'10', 3'20']
NGF	[3'60']	[3'3', 3'10', 3'20']

Different cellular responses to various frequencies of EGF and NGF signals has been explained by H. Ryu et al. [78] using different signaling modules in the same PC12 cells. A conditioned negative feedback in the MAPK module has been used to explain the temporal profiles of ERK activation dynamics in case of EGF stimulation. A negative feedback between ERK and RAF has been directly conditioned by the receptor activity via an unknown protein and has been proposed to explain different ERK profiles corresponding to different phenotypes. However, the same MAPK proteins but with different interactions has been used to explain cellular responses in the same PC12 cells for NGF stimulation. A positive feedback between ERK and RAF is modified to be conditioned by the receptor activity via another unknown protein to explain the ERK activity dynamics in case of stimulation by different frequencies of NGF signal.

The observations made by H. Ryu et al. [78] exhibit that cellular responses can also be specific to the frequency of growth-factor signals. The study has provided a further insight into the dynamic extracellular signal processing capabilities of biochemical networks, where it was well-established that cells generate phenotypic responses specific to the identity of growth-factors. However, the explanation of

different cellular responses to various frequencies of external signals using growth-factor specific feedbacks in same signaling module is arguable. Nevertheless, the observations made by H. Ryu et al. are important and show that growth-factors can in fact, generate multiple phenotypic responses depending on the frequency of external signals.

1.4.5 *Specificity* and *generalization* depending on temporal frequency of inputs

The previous experimental observations of cells generating different phenotypes upon stimulation by different growth-factors provides the growth-factor centric definition of *specificity*. The *specificity* of cellular responses has been related to the activation of different receptors upon binding by their corresponding ligands [49]. For instance, PC12 cells resulting in proliferation and differentiation respectively when stimulated by sustained signal of EGF and NGF [79]. However, recent experimental studies reported that cells can sense time-varying extracellular signals [84, 78, 11]. Furthermore, cells are also observed to generate different phenotypic responses depending on temporal frequencies of growth-factor signals [78]. For instance, PC12 cells start differentiating when stimulated by a certain frequency of EGF signal ($3'10'$), while on the other hand, proliferation has been observed for a different frequency ($3'3'$) of same growth-factor signal. This means that not just in the identity of growth-factor signals, there is evidence of *specificity* of cellular responses depending on the temporal frequencies of just one growth-factor signals as well.

Furthermore, it has been reported that there exists multiple frequencies of same growth-factor signals that can generate same phenotypic response. For example, H. Ryu et al. [78] observed that PC12 cells result in differentiation upon stimulation by different frequencies ($3'10'$ and $3'20'$) of EGF signals and also by a range of frequencies ($3'3'$, $3'10'$ and $3'20'$) of NGF signals. This means that cells exhibit a generalized response to a range of frequencies of different growth-factor signals. In other words, contrary to the *specificity* feature, cells generate same phenotypic response for a range of dynamic extracellular signals of different ligands in *Generalization* feature, shown schematically in Fig. 1.8.

The *specificity* of cellular responses to growth-factor identity has been explained using signaling modules with different feedbacks and connectivities specific to growth-factors. However, these signaling modules are highly interconnected and generate large signaling networks contrary to the earlier concept of discrete linear pathways, which relate extracellular signals to specific gene patterns [49, 8]. Furthermore, the

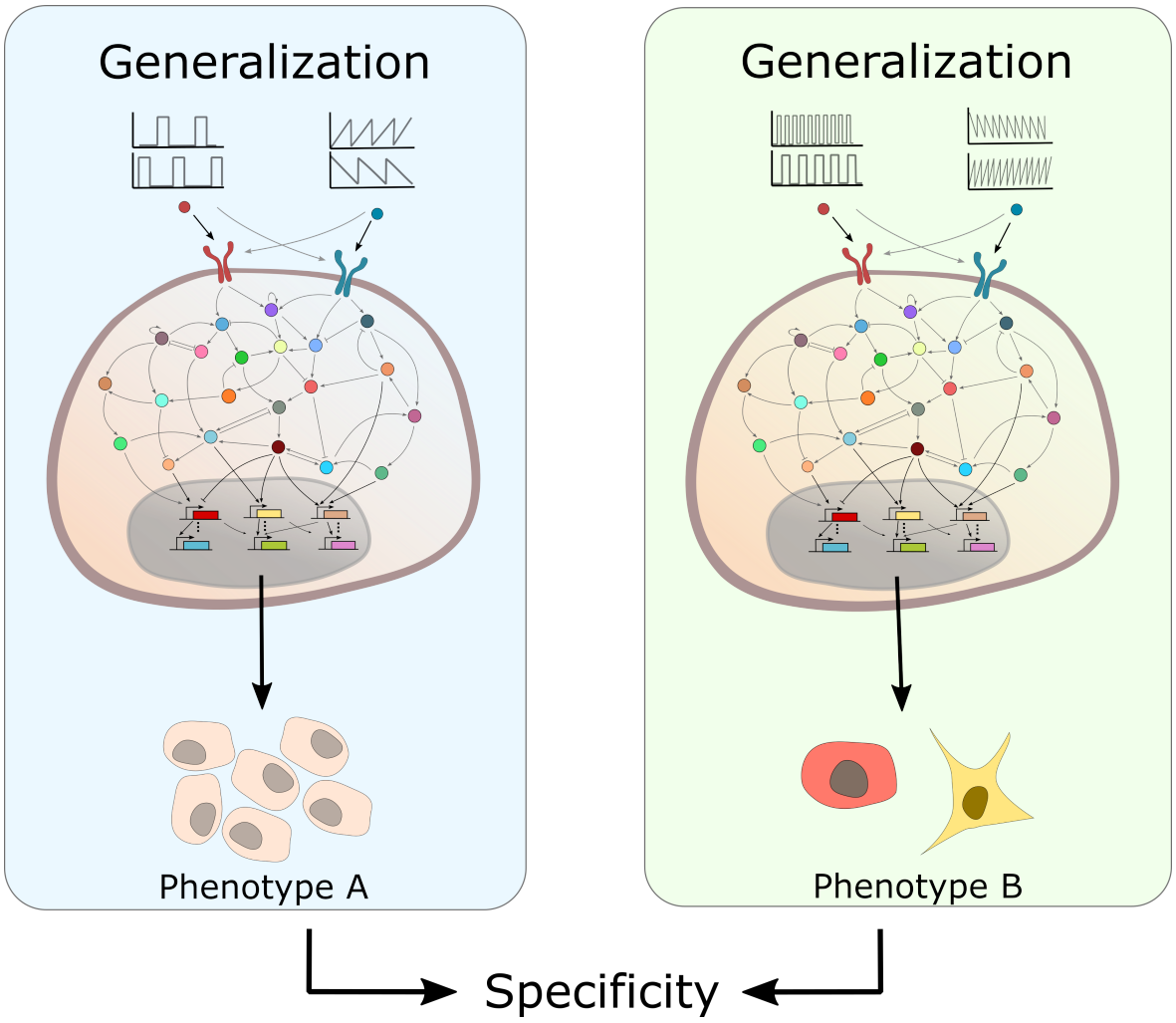

Figure 1.8: **Temporal signature of growth-factor signal dependant generalization and specificity in cell response.** Schematic showing different input signals for different receptor types resulting in a *generalized* cellular response of phenotype A and B for a range of input signals. Multiple cellular responses represent *specificity* in cellular responses depending on the temporal signature of external signals.

protein complement that mediates signal transduction downstream of RTKs is similar for all the RTK-mediated interconnected signaling modules. For example, same downstream proteins complements in the MAPK signaling module are activated by EGFR as well as TrkA receptors [79]. This means that *specificity* of cellular responses to growth-factor identity needs to be explained using the entire intracellular signaling network rather than using a selected module with context-dependent topology. Moreover, as we discussed above, cells generate different phenotypic responses depending on the temporal frequencies of extracellular signals while maintaining the two completely opposed features of *generalization* and *specificity*. This mean that these two opposed features are inherent part of time-varying information processing mechanism of cells and need to be explained using the entire intracellular network of the cell.

1.4.6 Limitations of approximating biochemical networks as autonomous systems

Experimental studies suggests that cells generate different phenotypic responses maintaining opposite information processing features of *generalization* and *specificity* depending on the temporal frequency of growth-factor signals. These experiments implement the simplified version of the complex time-varying extracellular information sensing, processing and responding features of biochemical networks that occur in cells that are continuously subjected to complex dynamic extracellular signals present in different tissues or organs. However, the authors [78, 11] described the underlying receptor and signaling networks using autonomous ordinary differential equations. By definition, autonomous systems describe only those systems that do not explicitly depend on the independent variable, as we discussed in Section 1.3.

Cells continuously sense and respond to dynamic extracellular signals which consists of changing concentrations of growth-factors, cytokines and hormones. These external dynamic signals change over time and space and act as inputs that cells sense via receptor networks present on their plasma membranes. The concentrations of receptors and signaling components vary along with their activity levels when an external signal is detected. In other words, the dynamic extracellular signals are the independent variables that result in parametric and dynamical changes in the entire biochemical network. For instance, experiments show that time-varying EGF pulses result in total EGFR concentration to decrease over time, along with the changing phosphorylation levels of EGFR and MAPK components which are present in the cytosol of the cell [84, 78, 11]. This means that biochemical networks explicitly depend on independent variables contrary to their current description as autonomous systems [40]. These independent variables are the system parameters that determine the dynamical state of biochemical network components. In other words, changing system parameters can be considered as a direct change in the dynamical properties of a system. The number of steady-states present in a system is given by its parameter organization meaning that the number and stability of steady-states will keep changing if such parameter organization is dynamic in nature [85, 45, 96]. There is also a possibility that these dynamic independent variables will result in the generation of transiently stable-states.

Furthermore, attractor-based information processing introduces two additional problems: lack of memory of previous inputs and insensitivity to upcoming time-varying signals. For instance, when the EGFR receptor network (Section 1.4.1) is parameterized in the monostable regime i.e only one steady-state is present, the system cannot

get activated in the presence of stimulus and therefore, quickly returns back to the only steady-state having no memory at all of the past stimulus (Fig. 1.5 (c)). On the other hand, in the bistable regime, a single pulse is enough to activate the system to the high phosphorylation state and the system becomes insensitive to upcoming growth-factor pulses (Fig. 1.5 (d)). This shows that the permanent memory of first pulse is hindering the system to sense more dynamic signals.

Parametric reorganization of biochemical systems because of external dynamic signals corresponds to the generation or destruction of new and transiently stable states. A sequential change in the number and stability of steady-states of membrane-bound receptors has been described as a result of extracellular time-varying signals by Stanoev et al. [85] using a minimal two-component receptor network. Contrary to the changing steady-states and stability in biochemical networks because of external dynamic signals, a fixed number of steady-states are the solutions of autonomous systems, as discussed in Section 1.3. Therefore, approximation of biological systems as autonomous might not be correctly portraying the actual underlying real-time dynamic information processing mechanisms of complex biological systems.

1.5 Current mechanisms and concepts of time-varying information processing

Steady-state formalism works for systems without any external perturbations or those with sustained stimulus. In both the cases, system reaches a stable attractor and remains there unless and until externally forced into another steady-state. As discussed in the previous section, the steady-state formalism fails when a system is perturbed by time-varying external signals and approximation of such systems as autonomous becomes invalid. In this section, we will survey the known concepts and mechanisms from various disciplines which people have attempted to solve similar problems where time-varying external signals drives the dynamics of entire system.

1.5.1 Criticality enabled integration of time-varying signals

Systems that sense time-varying signals require memory in order to integrate the signal information continuously [35]. To integrate information contained in time-varying signals such a memory must be dynamic in nature. For instance, in neuronal

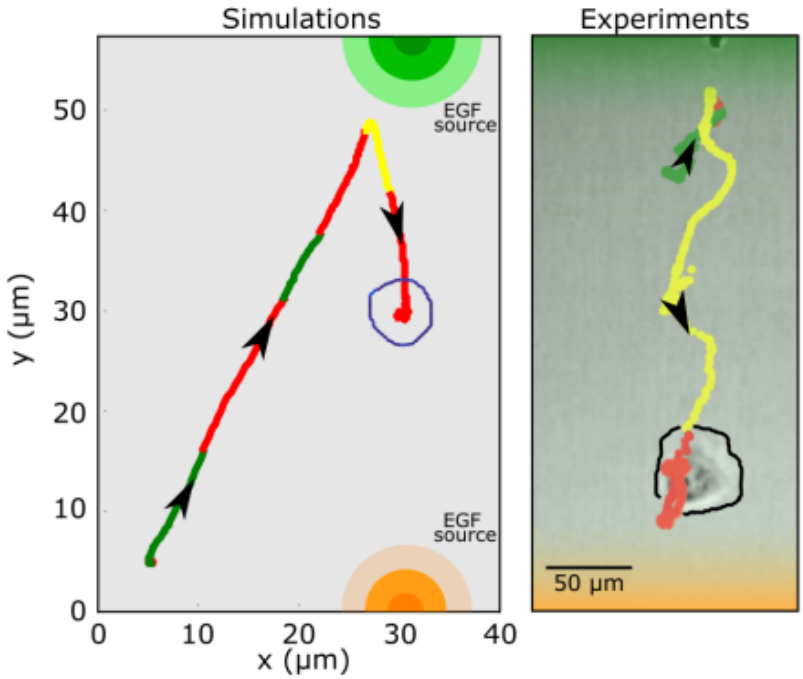

Figure 1.9: **Working memory enables history-dependent single-cell migration in a changing chemoattractant field.** (*left*) Simulation and (*right*) experiment with MCF10 cells, showing response to the sequence of EGF gradients: two signals from top (green) and a third one from bottom (orange). Red part of trajectory represents the migration of cells after removal of EGF gradients which includes transient memory. Used with permission from Nandan and Das et al. [70].

networks, the working memory has been proposed to store and process information as a stream of continuously incoming input pulses [9]. Working memory is different from short-term and long-term memories. In short-term memory, information can only be stored temporarily but not manipulated while long-term memory refers to permanent storage of information contrary to working memory. The reason that information cannot be held and manipulated in either of short and long-term memories is that, both are based on stable attractors. The criticality enabled 'transient' memory provides benefits over stable attractors based information processing. A system organized in the critical region can hold the information of past stimuli in its activity for a limited amount of time and at the same time it can also sense the upcoming external stimulus. A recent experimental study by Nandan, Das et al. [70] shows that such a criticality enabled 'transient' memory helps in directional migration of cells. The authors generated an artificial time-varying environment using a microfluidic device resembling the dynamic extracellular environment in multi-cellular organisms. When MCF10A cells were exposed to two consecutive EGF stimulus from same direction, the cells integrated both the signals and continued migrating in the direction of this signal even after EGF removal. However, the same cells quickly responded to another EGF stimulus from a completely opposite direction and started migration in that direction and again continued in that

direction even after its removal (Fig. 1.9). Using reaction-diffusion model and experimentally obtained parameters, the authors derived the dynamical mechanism of signal sensing and migrating behavior of cells where a saddle-node gives rise to a 'ghost' memory state upon signal removal for organization at criticality. Therefore, Nandan, Das et al. [70] showed the underlying dynamical mechanism for how cells to decipher complex chemical cues for migration which is based on critical organization of the system in the vicinity of a saddle-node rather than stable attractor. Such an organization helps cells to sense, integrate and resolve multiple chemical cues enabled by a transient memory and facilitate them to maintain polarization in the direction of growth-factor signal gradients even after their removal. Therefore, this 'transient' memory has all the features of a 'working' memory which equip the system with an advanced signal integration property by holding the information of past stimulus and at the same time being sensitive to upcoming extracellular stimuli.

1.5.2 Reservoir computing as a mechanism for dynamic signal processing

Reservoir computing is another framework that has been proposed for real-time dynamic information sensing and processing mechanism in neuronal networks. Reservoir computing is a computational framework developed basically for temporal or sequential data processing. This framework includes several recurrent neural network (RNN) models, including echo state networks (ESN) and liquid state machines (LSM) with recent advances in this field. In the early 2000s, ESNs [42] and LSMs [64] were independently conceptualized by W. Maass and H. Jaeger respectively as real-time computing models. A reservoir computing model consists of a reservoir for mapping inputs into a high-dimensional space and a readout for pattern analysis from the high-dimensional states in the reservoir. Maass et al. [64] asserted, using theoretical analysis and computer simulations, that two emergent macro properties as necessary and sufficient conditions for real-time computation based on transients: *separation property* (SP) and *approximation property* (AP). SP addresses the amount of separation between the trajectories of internal states of the system that are caused by two different input signals. AP addresses the resolution and recording capabilities of the readout mechanisms or its capability to distinguish and transform different internal states of the $'liquid'$ into given target outputs. Whereas SP depends mostly on the complexity of the liquid, AP depends mostly on the adaptability of the readout mechanism to the required task [64].

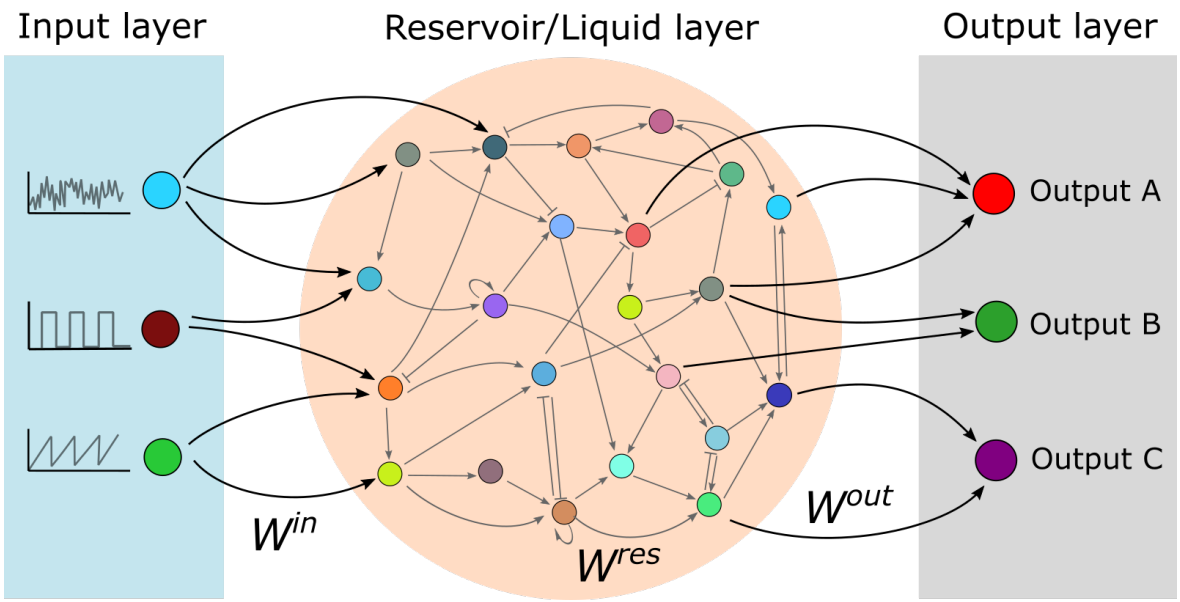

Figure 1.10: **Structure of a Liquid State Machine.** An LSM consisting of an input layer (blue) which transmits multiple inputs to a large dimensional *liquid* or reservoir layer (orange). Information encoded in the dynamics of reservoir nodes is decoded by the output layer (grey) nodes corresponding to different tasks.

The main characteristic feature of reservoir computing is that the input weights (W^{in}) and the weights of the recurrent connections within the reservoir (W^{res}) remain fixed whereas only the readout weights (W^{out}) (Fig. 1.10) are trained with a simple learning algorithm such as linear regression. In a physical implementation, this liquid state consists of all information about the current internal state of a dynamical system that is accessible to the readout nodes. In mathematical terms, this liquid state is simply the current output of some operator L^M that maps input functions $u(\cdot)$ onto functions $x^M(t)$ given by Eq. 1.4:

$$x^M(t) = (L^M u)(t) \tag{1.4}$$

The memoryless output layer or the readout map f^M transforms, at every time t, the current liquid state $x^M(t)$ into the output:

$$y(t) = f^M(x^M(t)) \tag{1.5}$$

In contrast to the liquid filter L^M, this readout map f^M is in general chosen in a task-specific manner and there may be many different readout maps that extract different task-specific information in parallel from the current output of L^M.

Maass et al. [64] showed the *separation property* of LSM using separation between

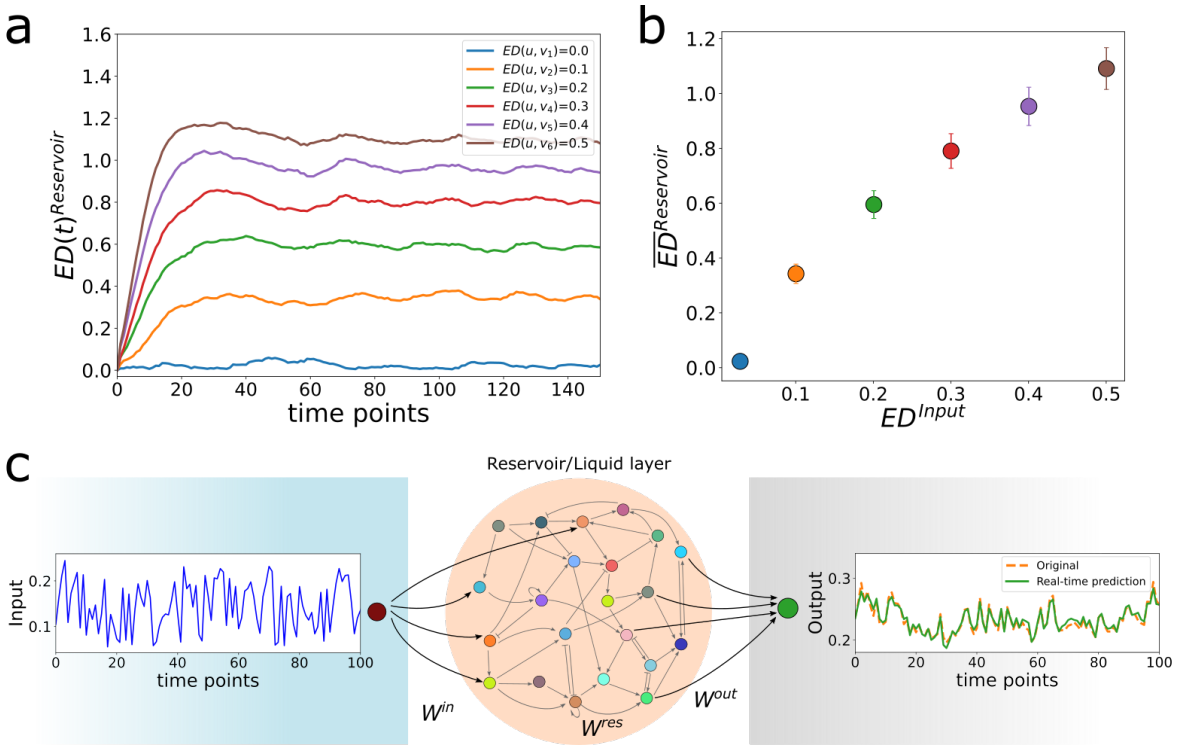

Figure 1.11: **Separation and approximation properties of Liquid State Machines.** (a) Emergent *separation property* in a LSM with sparsely connected reservoir network of 50 nodes and connection density of 0.037. Separation of full *liquid* state is shown for two different inputs u and v_i as a function of time t. Different inputs (v_i) are generated to have a differing separation with respect to a reference input u. (b) Mean of *liquid* state separation over time with respect to corresponding separation between inputs. (c) *Approximation property* analyzed using 10th order NARMA task with a uniform input of $U(0.05, 0.25)$ (blue) giving a real-time output of approximated NARMA function (green) with root-mean-square (RMS) error of just 0.2586 with respect to the original output (dashed, orange). A separate set of input drawn from the same uniform distribution was used to obtain output weights W^{out} using ridge-regression Eq. 1.8.

internal states of the '*liquid*' using computer simulations. In other words, two different signals $u(\cdot)$ and $v(\cdot)$ were randomly generated and used as input signals. The separation between the M-dimensional resulting trajectories $x_u^M(\cdot)$ and $x_v^M(\cdot)$ of the liquid states were obtained (Fig. 1.11 (a)) and the resulting average distance $||x_u^M(t) - x_v^M(t)||$ between the liquid states is found to be nonlinearly increasing with increasing separation between the input signals $||u(\cdot) - v(\cdot)||$ (Fig. 1.11 (b)). The *approximation property* of reservoir computers has been estimated using various nonlinear functions and used them for making predictions for instance, time-series forecasting of chaotic functions, handwritten text classification, speech recognition and stock price prediction [90]. A benchmark memory demanding task, 10th order Nonlinear Auto-Regressive Moving Average (NARMA) task [7] is used to analyze the *approximation property* of LSMs along with memory of previously encountered

inputs. The 10th order NARMA function is defined by Eq. 1.6:

$$y(t+1) = 0.3y(t) + 0.05y(t)\sum_{i=0}^{9} y(t-i) + 1.5u(t-9)u(t) + 0.1 \tag{1.6}$$

This task consists of training the output weights of the LSM such that an input ($u(t)$) will be mapped to obtain an approximate NARMA output in real-time. Inputs are usually drawn from a uniform distribution $U(0.05, 0.25)$. Each node dynamics can be defined by a activation functions (L) for instance, Tanh, ReLU and Sigmoid [90]. The LSM approximated real-time output $\tilde{Y}$ can be obtained using a linear combination of the *liquid* states given by the following Eq. 1.7:

$$\tilde{Y} = W^{out}X \tag{1.7}$$

Ridge regression is used to obtain the output weights W^{out}. It is a type of linear regression in which the regression coefficients are obtained from the following Eq. 1.8:

$$W^{out} = YX^T(XX^T + \gamma^2 I) \tag{1.8}$$

where X^T is the transpose of X and Y is the matrix with all expected outputs over time t, I is the identity matrix, and γ is a regularization parameter. Ridge regression favors regression coefficients with smaller absolute values. Therefore, memory dependant functions like NARMA 10th-order task can be obtained using ridge-regression method where even a minimal number of output nodes provides an accurate real-time mapping of dynamic input to output (Fig. 1.11 (c)). Like the Turing machines, the model of an LSM is based on a rigorous mathematical framework that under idealized conditions guarantees universal computational power. Turing machines have universal computational power for off-line computation on discrete inputs, while LSMs have universal computational power for real-time computing with fading memory on analog functions in continuous time. This real-time computation framework is described as non-autonomous dynamical systems.

1.6 Non-autonomous systems

From the perspective of thermodynamics and based on the nature of the system-environment interactions, a system can classified as one of the three classes: open, closed and isolated [17]. A system is called isolated if it does not exchange neither energy nor matter with its environment (Fig. 1.12 (a)). An open system, on the contrary, exchanges both energy and matter with its environment. Finally, a closed system exchanges only energy but not the matter with its environment. Living systems cannot be seen as isolated by nature, and need to be modelled as open systems. The distinction between isolated and non-isolated systems manifests in an important way for dynamical systems, namely in the distinction between autonomous dynamical systems and dynamical systems that are non-autonomous [86]. Thus, non-autonomous systems are conceptually related to non-isolated systems such as open systems: as illustrated in Fig. 1.12 (b), the evolution of an open system depends on its own internal dynamics and also on external influences, which can come in many forms. The evolution of the state $x(t)$ over time of a non-autonomous dynamical system can be written in general as Eq. 1.9, where the function f is deterministic and explicitly depends on time. For autonomous dynamical systems the underlying assumption is that the system parameters do not change with environment and time.

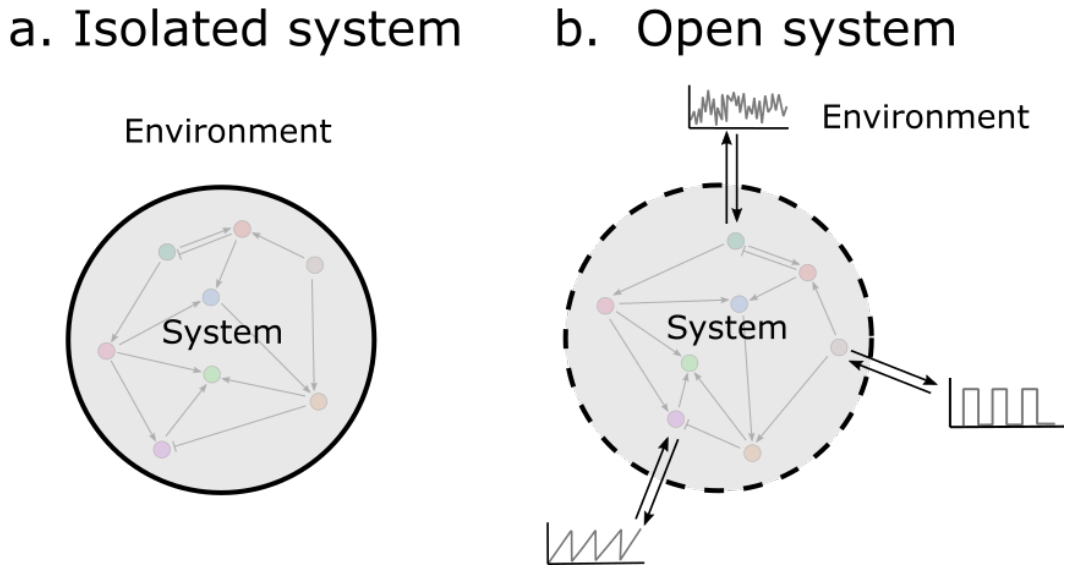

Figure 1.12: **Thermodynamically isolated and open system.** (a) An isolated system with closed boundary restricting exchange of matter and energy with its surrounding. (b) An open system which can interact with its environment by exchanging matter and energy via permeable boundary.

The formal mathematical description of non-autonomous dynamical systems has been applied to solve problems in various disciplines [50]. For instance, having applications in non-autonomous coupled oscillators [5], stability of networks [59] and also in biological sciences [21]: in bacterial growth [41], epidemiology [91], tumor drug treatment models [34] and cardiovascular systems [21]. In all these studies the non-autonomous dynamical systems are carefully modeled only in low dimensions. However, the non-autonomous dynamical systems theory has not yet been applied to explore the information processing mechanism in living systems.

1.6.1 Complete trajectories are solutions of non-autonomous systems

The theory of autonomous differential equations and their solutions is well established [87]. The solutions of an autonomous differential equation are the steady-states. Which steady-state will be occupied the system depends on the stability of the steady-states and also the initial state of the system. To describe the theory and meaning of solutions of non-autonomous biological systems, we start with the generic form of a non-autonomous ordinary differential equation which can be described using Eq. 1.9:

$$\dot{x} = f(x(t), t) \tag{1.9}$$

with $x(s) = x_0$, where s is the initial time and t is the final time. The solution of non-autonomous systems can be analytically represented by a family of solution operators given by $\{S(t,s)\}_{t\geq s}$, which depend on both the initial and final times. The solution of Eq. 1.9 at time t is denoted by $S(t,s)x_0$, meaning the solution operator $S(t,s)$ operates on initial state of the system x_0. On the contrary, the solution of autonomous equation depends only on $t-s$, meaning $S(t,s) = T(t-s)$ for a family of solution operators $\{T(t)\}_{t\in\mathbb{R}}$ of autonomous system. Therefore, the notion of steady states in autonomous system is replaced by complete trajectory in non-autonomous system. The common definitions of trajectories generated by solution operators of non-autonomous systems [1, 50, 19]:

Definition 1.1. *The continuous map:* $\mathbb{R} \to \mathbb{R}^m$ *is a complete trajectory if*

$$S(t,s)x(s) = x(t) \quad \textit{for all } t,s \in \mathbb{R}$$

However, in general there will be many complete trajectories (stable and unstable), since they are just solutions that exist for all $t \in \mathbb{R}$. A complete trajectory is a particular example of an *invariant set* in a non-autonomous equation.

Definition 1.2. *A time-varying family of sets* $\{\Sigma(t)\}$ *is invariant if*

$$S(t,s)\Sigma(s) = \Sigma(t) \quad \textit{for all } t,s \in \mathbb{R}$$

This formal definition simply says that if $x(s) \in \Sigma(s)$ then $S(t,s)x(s) = \Sigma(t)$. In other words, the full trajectory $\Sigma(t)$ of a non-autonomous system in Eq. 1.9 will be invariant if the initial state of the system $x(s)$ belongs to a $\Sigma(s)$. The set $\Sigma(s)$ contains all the initial states from where the trajectory $\Sigma(t)$ could have been generated.

1.6.2 Pullback attraction and stability

In autonomous systems, an invariant set Σ is (locally) attracting, if there exists a neighbourhood $\mathcal{O}$ of Σ such that

$$\text{dist}(S(t,0)x_0, \Sigma) \to 0 \qquad \text{as } t \to \infty \qquad \text{for all } x_0 \in \mathcal{O} \tag{1.10}$$

where the distance is Hausdorff semi-distance between two set A and B is the longest distance from a point in one set to another set, $\text{dist}(A,B)$ is defined as

$$\text{dist}(A,B) = \sup_{a\in A} \inf_{b\in B} |a-b| \tag{1.11}$$

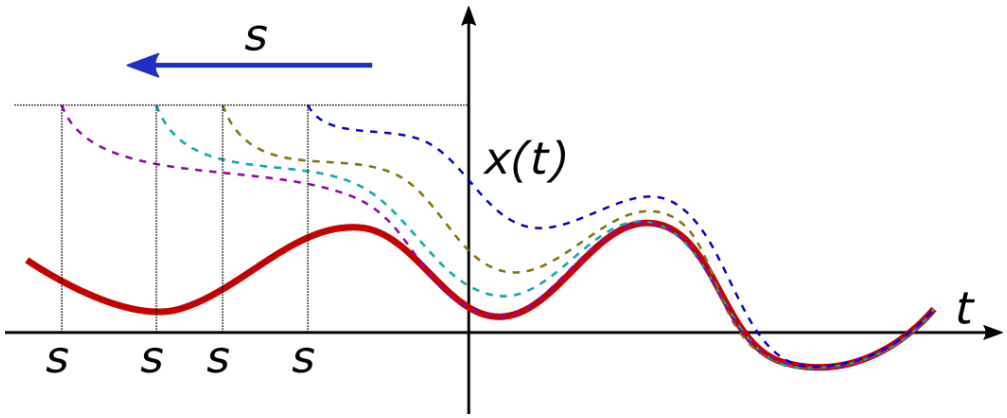

Figure 1.13: **Pullback attraction.** Starting at different initial times (s) long ago in the past, the respective trajectories (dashed) converges to a stable attractor (red) as $s \to -\infty$.

We are accustomed to considering 'time-asymptotic' behaviour as the limiting behaviour as $t \to \infty$. The implication of this idea is that if we run the experiment for long enough and then consider the state of the system at some future time, it is well approximated by one of the states in the attractor. n the autonomous case, where $S(t,s) = T(t-s)$, the concept of attraction is entirely equivalent to the existence of a neighbourhood $\mathcal{O}$ of Σ such that for each fixed $t \in \mathbb{R}$,

$$\text{dist}(S(t,s)x_0, \Sigma) \to 0 \qquad \text{as } s \to -\infty \qquad \text{for all } x_0 \in \mathcal{O} \tag{1.12}$$

This is the idea of 'pullback' attraction [50, 19, 1] that does not involve running backwards in time rather we consider taking measurements in an experiment now, at time t, which began some long ago some time s in the past with $s < t$, schematically shown in Fig. 1.13. If the experiment has been running long enough then the state of the system is approximated by one of the attracting states. Stability of pullback attractors are defined using 'Lyapunov stability', which constrains trajectories to stay close to the invariant set.

Definition 1.3. *$\Sigma(\cdot)$ is pullback Lyapunov stable if for every $t \in \mathbb{R}$ and $\epsilon > 0$ there exists a $\delta > 0$ such that for any $s < t$, $x_s \in N(\Sigma(s), \delta(t))$ implies that $S(t,s)x_s \in N(\Sigma(s), \epsilon)$.*

where the notation $N(X, \epsilon)$ denotes the ϵ-neighbourhood of a set X:

$$N(X, \epsilon) = \{y : y = x + z, x \in X, z \in \mathbb{R}^m \text{ with } |z| \leq \epsilon\} \tag{1.13}$$

All the systems that explicitly depend on time and also their dynamic environment are thermodynamically open systems. Since open systems exchange both energy and matter with their respective environment, the internal variables of the system also start to vary with the dynamics of external environment and therefore, their approximation as autonomously evolving system is no longer valid. Such systems are modeled as non-autonomous systems, and their general solutions are not constant over time rather rather the solutions are also dynamic in nature and are given by trajectories in the euclidean space of the dimension same as that of the system.

1.7 Hypothesis

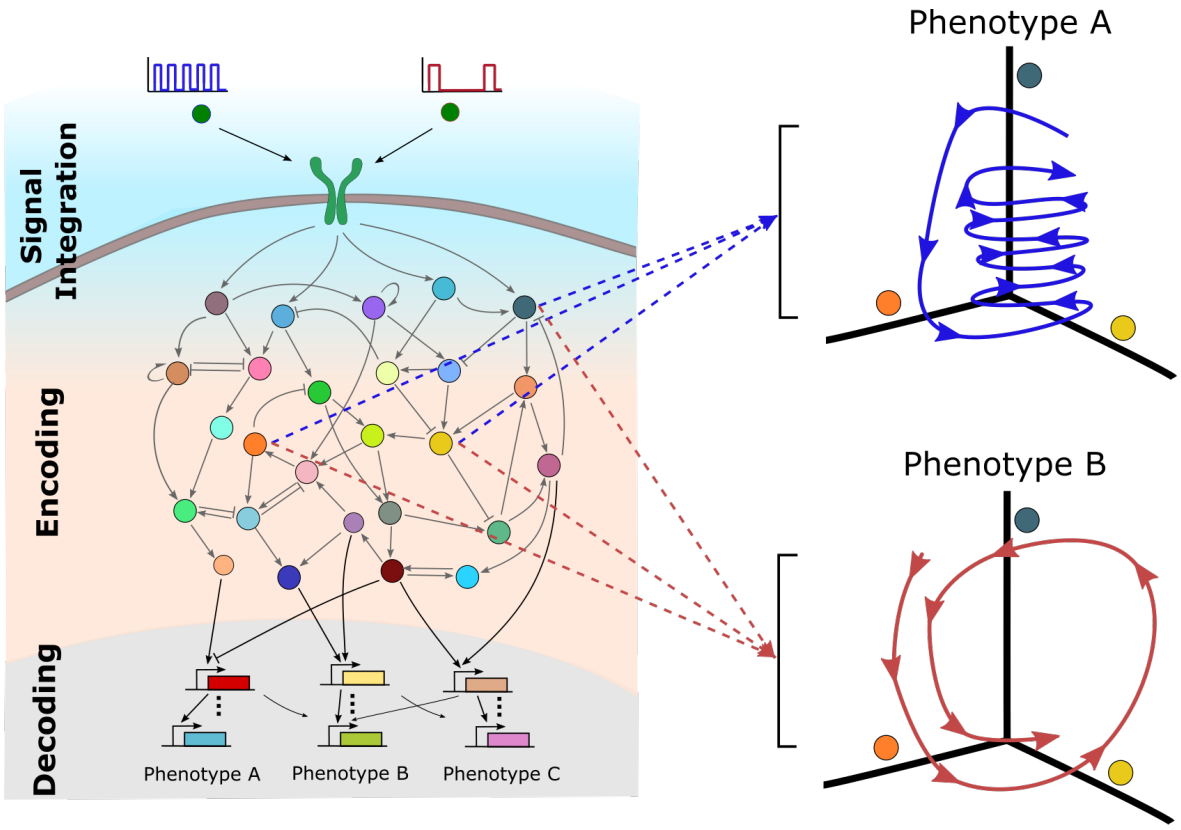

Figure 1.14: **Biochemical network as a non-autonomous system.** Schematic diagram of a biochemical network processing multiple external stimuli at the receptor level and triggering different phenotypic responses. Different external input signals are integrated differently by receptor networks and also encoded uniquely. Such information can be extracted by multiple read-outs from the signaling network and visualized in the form of multi-dimensional phase-space trajectories.

With recent advancements in microfluidic devices several independent studies have reported that cells sense and react to its time-varying extracellular environment [84, 78, 70]. Single cells, acting as an individual entities are shown to resolve conflicting spatio-temporal environment with plasticity to quickly change its response according to the information present in dynamic environment [70]. These experimental studies proves that cells are open systems that sense their extracellular environment not just for its survival but also to serve its purpose by integrating the information present in the dynamic environment and generating various responses depending upon spatio-temporal signatures of external signals. The current mechanisms to explain these dynamic behavior of cellular systems is based on the approximation that cells are autonomous systems. Such assumptions have worked to explain

the steady-state dynamics of biochemical networks with time-invariant extracellular signals. However, to accurately elucidate the dynamic environment dependant properties of biochemical networks such assumptions need to be revised accordingly.

Inspired by the transient based real-time information processing mechanism of liquid state machines, we propose a new theoretical framework where the entire intracellular network is treated as a non-autonomous system whose solutions are the signal dependent trajectories. This framework describes the multi-level processing of dynamic chemical signals with transients in real-time, schematically shown in Fig. 1.14. In this framework, the information from time-varying extracellular environment is integrated at the level of cell surface receptor networks, constructing the 'signal integration' layer. The information is then encoded in the high-dimensional phase-space trajectories of the downstream signaling network present in the cytosol of the cell, constructing the 'encoding' layer. We propose that any complex network of arbitrary structure will have real-time dynamical information processing capabilities if the system is organized near a dynamic saddle node. The phenotype specific information will be encoded in unique phase-space trajectories in such a way that trajectories belonging to same phenotype will have similar characteristic features in phase-space. On the other hand, trajectories belonging to different phenotypes will be characteristically different and remain separated in phase-space (Fig. 1.14). Furthermore, the temporal signature of external signal dependent *specificity* and *generalisation* will be an emergent property of such a dynamic information processing by transients. Information encoded in the high-dimensional phase-space transients will be interpreted into different phenotypic responses in real-time by early response genes (ERG) constructing the 'decoding' layer.

1.8 Objectives

The main objective of this thesis is to develop a new conceptual framework that describes the real-time information processing mechanism utilized by cells to sense and operate in time-varying extracellular environments. This new conceptual framework will be based on the experimentally identified features of the biochemical networks and should explain the robust real-time information processing features of biochemical networks while maintaining the completely opposed features of *generalisation* and *specificity*. First, we mathematically define and prove the theory of real-time biochemical information processing mechanism by defining biochemical networks as non-autonomous systems. We investigate the benefits of metastable states in the signal integration layer of biochemical networks over attractor-based information processing. Throughout the thesis, we address the question of benefits of 'working' memory over steady-state dynamics at different levels of information processing. We define the general principles of information processing in large arbitrary signaling network and then use an experimentally identified biochemical networks to elucidate the realistic biochemical information encoding and decoding mechanisms. Furthermore, we incorporated experimentally identified features and noise levels in the simulations to make results more closure to real experimental scenarios. We set out to use publicly available experimental data to compare the predictions made with this novel real-time information processing mechanism. Finally, we addressed the question of abundance and conservation of proteins present in different information processing levels of biochemical networks.

Chapter 2

Results

2.1 Theoretical description of information processing with transients

Biochemical networks are currently being formulated as autonomous systems with the assumption that the biological systems are isolated systems, that do not exchange matter and energy with their surroundings. The solutions of autonomous differential equations are the steady-states (Section 1.3). In other words, the implication of such assumption is that the relevant dynamics of biological processes occur near or at the steady-states, thereby placing a strong focus on attractors based dynamic information processing. The consequence of steady-state formalism is that different cellular responses are being assumed to have a correspondingly different steady-states [79, 20]. As noted above, in order to address the question of mapping dynamic signals to phenotypic outputs, the steady-state paradigm must be extended when considering time-varying signals. In the presence of external signals not only the state variables, but also the parameters of the intracellular network change over time. Therefore, we start by formally describing biochemical networks as non-autonomous systems whose solutions are the trajectories depending on the extracellular signals, as discussed in Section 1.6.

2.1.1 Describing biochemical networks as non-autonomous systems

We start by considering a biochemical network that consists of n interacting proteins. The activity of i^{th} protein at any time t is given by $x_i(t)$. Defining an n-dimensional vector $x(t)$ which consists of all the protein activity levels at any time t: $x(t) = [x_1(t), ...x_i(t), ...x_n(t)]^T$. A general protein interaction function f is defined in such a way that all the interactions for a protein will be included in this function.

Therefore, the generic form of this non-autonomous biochemical network which explicitly depends on external signal $p(t)$ is given by:

$$\dot{x} = f(x(t), p(t)) \tag{2.1}$$

with $x(s) = x_0$, where s is the initial time and t is the final time. From the theory of non-autonomous differential equations [50, 19] discussed in Section 1.6, the general analytical solution of any non-autonomous differential equation of the form Eq. 2.1 is given by a family of solution operators $\{S(t,s)\}_{t\geq s}$, which depends on both the initial and final times. Described in definitions 1.1, 1.2 and 1.3, these solution operators will result in complete, invariant and Lyapunov stable trajectories $\Sigma(t)$ which are the solutions of non-autonomous differential equations. This means that the solutions of non-autonomous biochemical networks are the n-dimensional trajectories that explicitly depends on time t.

The general form of non-autonomous differential equation given by Eq. 2.1, however, to define the underlying biochemical reactions and interactions of proteins in the large signaling network there are several ways to model them. For instance, using law of mass action kinetics, Michaelis-Menten kinetics and using biochemical systems theory (BST) equations. We use BST form of equations where the dynamics of a species is modeled using power-law expansions of the system's variables, and all the activation and deactivation entities for a protein are separated. Furthermore, a separate external signal $p(t)$ term is present in all the protein entities which directly provides input to specified signaling proteins. For a biochemical network with $i = 1,..N$ components, Eq. 2.1 takes the following general form for biochemical networks:

$$\dot{x}_i = B_i^f (1 - x_i)^{C_i^{fe}} \prod_{\substack{j=1 \\ A_{ij}=1}}^{N} x_j^{F_i^{ge}} - D_i^f x_i^{E_i^{fe}} \prod_{\substack{j=1 \\ A_{ij}=-1}}^{N} x_j^{F_i^{ge}} + (1 - x_i) I_1^l p(t) \tag{2.2}$$

where B^f, C^{fe} represent the autonomous activation and D^f, E^{fe} represent the autonomous deactivation rate constants. $A_{ij} \in [-1, 0, 1]$ is the connectivity matrix of the arbitrary signaling network. F^{ge} is the parameter controlling the rate of activation ($A_{ij} = 1$) and deactivation ($A_{ij} = -1$) of i^{th} component by the j^{th} component of the network if there exists a link between them, i.e $A_{ij} \neq 0$. We will use this general non-autonomous differential equation with external stimulus $p(t)$ for defining the

non-autonomous biochemical networks in this thesis. Furthermore, the general solutions of Eq. 2.2 will be trajectories depending on the time-varying external signal. In the next section, we will discuss the formal mathematical definitions of different solution trajectories of non-autonomous biochemical networks that will represent various cellular processes.

2.1.2 Pullback attractors of different phenotypes are separated in phase-space

We have hypothesized that the phenotype specific information will be encoded in unique phase-space trajectories in such a way that trajectories belonging to same phenotype will have similar characteristic features in phase-space while trajectories belonging to different phenotypes will be characteristically different and remain separated in phase-space (Section 1.7, Fig. 1.14).

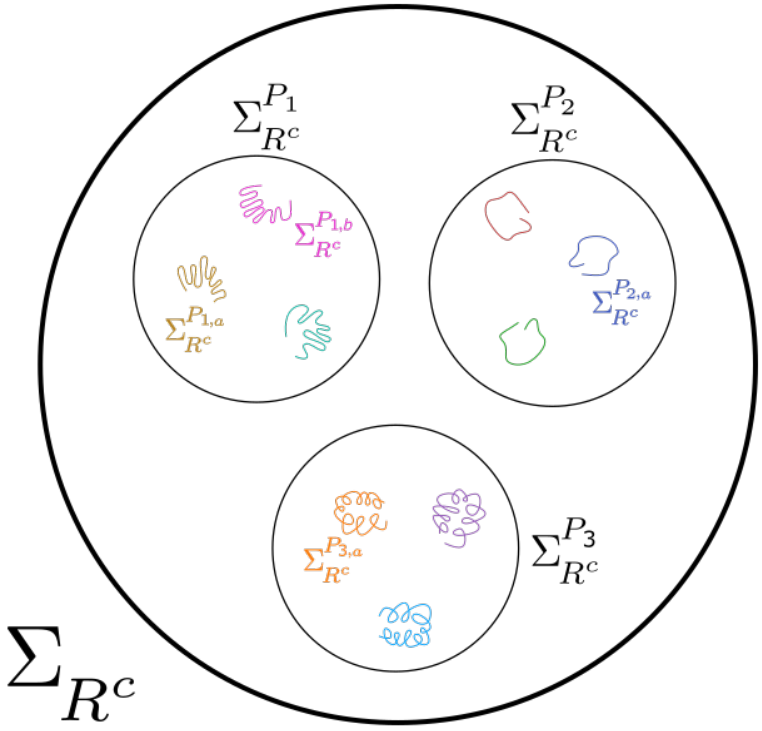

Figure 2.1: **Trajectories Σ_{R^c} generated by receptor network R^c.** Schematic of three phenotype sets ($N = 3$) with corresponding trajectories $\Sigma_{R^c}^{P_{i,j}}$. Each phenotype set $\Sigma_{R^c}^{P_i}$ has trajectories with similar characteristic features in phase-space.

Consider a general N-dimensional non-autonomous biochemical network given by Eq. 2.2 whose parametric organization is such that it generates a range of phase-space trajectories specific to external time-varying signal $p(t)$, under the process R^c.

Defining all the possible trajectories in N-dimensional phase-space generated by this process R^c as $\Sigma_{R^c}(\cdot)$. We have hypothesized that the information of different phenotypes is encoded in different phase-space trajectories (Section 1.7). In other words, trajectories corresponding to a phenotype will have similar characteristic features in phase-space for instance, curves, volume covered or time spent in specific regions of phase-space. On the other hand, trajectories of different phenotypes will be characteristically different from each other with unique features in phase-space. Let us consider that under process R^c, the non-autonomous biochemical network can process extracellular time-varying signals to generate M different phenotypes. All the trajectories resulting in a phenotype i are given by an *invariant set* $\{\Sigma^{P_i}_{R^c}(\cdot)\}$. An individual trajectory a belonging to phenotype i is $\Sigma^{P_{i,a}}_{R^c} \in \Sigma^{P_i}_{R^c}$ generated by a specific signal. Therefore, all the *invariant sets* of trajectories generated in R^c will correspond to N different phenotypes constituting $\Sigma_{R^c}(\cdot)$ set, shown schematically in Fig. 2.1 and described in the following definition:

Definition 2.1. *In process R^c, the N-dimensional non-autonomous biochemical network can uniquely sense and integrate extracellular signals to generate M different phenotypes. The generated unique trajectories sets specific to each phenotype are as follows:*

$$\left\{\{\Sigma^{P_1}_{R^c}\},\{\Sigma^{P_2}_{R^c}\},\ldots\{\Sigma^{P_i}_{R^c}\},\ldots\{\Sigma^{P_M}_{R^c}\}\right\}\subset \Sigma_{R^c} \tag{2.3}$$

such that there exists no signal which can generate a trajectory common in any two given trajectory sets:

$$\Sigma^{P_i}_{R^c}\cap\Sigma^{P_j}_{R^c}=\emptyset \quad \textit{with } i\neq j. \tag{2.4}$$

Different classes of trajectories are defined in a way that trajectories resulting in same phenotype will have similar characteristic features in phase-space as shown schematically in Fig. 2.1. Quantitatively, any two trajectories of the same phenotype will remain close to each other in N-dimensional phase-space. In mathematical terms, being close to each other means that the separation between the same class of trajectories will remain bounded by an $\epsilon \geq 0$, given by the following definition:

Definition 2.2. *If two trajectories $\Sigma^{P_{i,a}}_{R^c}(\cdot)$ and $\Sigma^{P_{i,b}}_{R^c}(\cdot)$ of same phenotype set $\Sigma^{P_i}_{R^c}$, are globally pullback Lyapunov stable and attracting for every $t\in\mathbb{R}$ then there exists an $\epsilon^{P_i}\geq 0$ such that*

$$\lim_{s\to-\infty}\left|\Sigma^{P_{i,a}}_{R^c}(t)-\Sigma^{P_{i,b}}_{R^c}(t)\right|\leq\epsilon^{P_i}.$$

This defines the criteria for trajectories generated by distinct external signals and yet generating same phenotype or same class of trajectories. Definition 2.2 also gives the formal description of *generalization* property of biochemical information processing where depending on the temporal signatures of different extracellular dynamic signals, cells result in same phenotypic responses whose information is encoded in trajectories with similar characteristics that remain close to each other in phase-space. Based on this definition, we make the following theorem defining the phase-space properties of trajectories that belong to different phenotypes.

Theorem 2.3. *If any two trajectories belonging to different phenotype sets of trajectories exist:*

$$\Sigma_{R^c}^{P_{i,a}}(\cdot) \subset \Sigma_{R^c}^{P_i} \quad \textit{and} \quad \Sigma_{R^c}^{P_{j,b}}(\cdot) \subset \Sigma_{R^c}^{P_j} \quad \textit{with } i \neq j.$$

such that

1. *Both the trajectories are globally pullback Lyapunov stable and attracting for every* $t \in \mathbb{R}$.
2. *There exist* $\epsilon^{P_i}, \epsilon^{P_j}$ *(according to Definition 2.2) and* $\epsilon^{P_{ij}} > 0$ *where* $\epsilon^{P_{ij}} = max\{\epsilon^{P_i}, \epsilon^{P_j}\}$

then

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{i,a}}(t) - \Sigma_{R^c}^{P_{j,b}}(t) \right| > \epsilon^{P_{ij}}.$$

Proof. Let four trajectories from two different phenotype sets of trajectories exist:

$$\Sigma_{R^c}^{P_{i,a}}(\cdot), \Sigma_{R^c}^{P_{i,c}}(\cdot) \subset \Sigma_{R^c}^{P_i} \quad \text{and} \quad \Sigma_{R^c}^{P_{j,c'}}(\cdot), \Sigma_{R^c}^{P_{j,b}}(\cdot) \subset \Sigma_{R^c}^{P_j}$$

Then according to definition 2.2:

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{i,a}}(t) - \Sigma_{R^c}^{P_{i,c}}(t) \right| \leq \epsilon^{P_i} \tag{2.5}$$

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{j,c'}}(t) - \Sigma_{R^c}^{P_{j,b}}(t) \right| \leq \epsilon^{P_j} \tag{2.6}$$

Assume that $\Sigma_{R^c}^{P_{j,c}}$ and $\Sigma_{R^c}^{P_{j,c'}} \in \Sigma_{R^c}^{P_i} \cap \Sigma_{R^c}^{P_j}$ and $\Sigma_{R^c}^{P_{j,c}} = \Sigma_{R^c}^{P_{j,c'}}$. Then adding Eq. 2.5 and Eq. 2.6 gives:

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{i,a}}(t) - \Sigma_{R^c}^{P_{i,b}}(t) \right| \leq \epsilon^{P_i} + \epsilon^{P_j} \tag{2.7}$$

Further, let $\epsilon^{P_i} > \epsilon^{P_j}$ and consider that $\Sigma_{R^c}^{P_{j,b}}$ exists with a $\delta > 0$ such that

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{i,a}}(t) - \Sigma_{R^c}^{P_{i,b}}(t) \right| = \epsilon^{P_i} + \delta$$

No other trajectory exists with any $0 < \delta' < \delta$. This means that the following inequality is valid:

$$\lim_{s \to -\infty} \left| \Sigma_{R^c}^{P_{i,a}}(t) - \Sigma_{R^c}^{P_{i,b}}(t) \right| > \epsilon^{P_i} \tag{2.8}$$

□

Theorem 2.3 states that if any two trajectories belonging to opposite phenotypic responses in cells then they will remain separated in phase-space. This provides a distance based quantification method for distinguishing different phenotypic information encoded in different class of trajectories. Theorem 2.3 is also the formal description of *specificity* feature of biochemical information processing in cell (Section 1.4.5). Depending on temporal signatures of different external signals, cells result in different phenotypic responses whose information is encoded in characteristically different trajectories that are separated in phase-space.

Under process R^c, the non-autonomous biochemical networks generating phenotype-specific trajectory sets given by definition 2.1, is in line with the experimental observations. Stanoev et al. [84, 85] showed that cell surface receptor networks integrate time-varying extracellular signals depending upon their temporal signatures and this signal integration property is enabled by dynamic organization of the system near 'criticality'. Furthermore, H. Ryu et al. [78] showed that cells generate similar and opposite phenotypic responses depending upon the frequency of dynamic growth-factor signals. The *generalization* and *specificity* features (Section 1.4.5) of dynamic information processing in non-autonomous biochemical networks are mathematically formulated by definition 1.2 and theorem 2.3 respectively.

2.2 Frequency dependent integration of time-varying signals by 'critically' organized receptor network

The solutions of non-autonomous biochemical networks are the phase-space trajectories generated by time-varying extracellular signals. The information of distinct cellular responses are encoded in unique trajectories with characteristic features in phase-space. In the previous section, we provided mathematical formulation of our hypothesis with definitions and theorem representing how phenotype-specific information is encoded in distinct phase-space trajectories while maintaining the features of *specificity* and *generalization*. We will now use the simulations to provide the numerical proof and elucidate the details of our hypothesis by modeling biochemical networks as non-autonomous systems.

Biochemical networks have cell surface receptors that sense and integrate the information present in extracellular time-varying signals. Experimental studies show that receptor networks integrate the information from external inputs using a dynamic memory [84]. Using a minimal receptor network, Stanoev et al. [85] showed that the dynamic memory in receptor activity arises when the system is organized near a saddle-node bifurcation point, discussed in Section 1.4.2. This dynamic memory enables the receptor to hold the information of past stimulus at the same time, the system remain sensitive to the upcoming input stimulus, to integrate information contained in time-varying signals. Therefore, we explore the signal integration properties of the saddle-node 'ghost' enabled dynamic memory in minimal receptor network (Fig. 2.2 (a)) using just two input pulses. For that, we will use just two subsequent input pulses of same duration and vary the time duration in between them which we call inter pulse interval or IPI in short.

First, we parameterize the system in bistable regime with $\hat{\gamma}_{DNF} = 2$, where the memory is permanent. Starting with two 3 minute input pulses and a fixed IPI of 15 minutes, the system senses the first pulse and reaches a high receptor activity state. After the removal of the input pulse, the system stays in the high activity state and becomes insensitive to the second stimulus (Fig. 2.2 (d)). Once activated, the system remains in the high activity state irrespective of the IPI between two pulses. This makes the total duration of receptor activity to be fixed irrespective of the IPI as shown with grey curve in Fig. 2.2 (e). In case of no memory, when the system is parameterized far in the monostable regime with $\hat{\gamma}_{DNF} = 10$ the system is not activated at all (Fig. 2.2 (f)) and therefore the total duration of receptor activity remains zeros irrespective of IPI (Fig. 2.2 (e), orange curve).

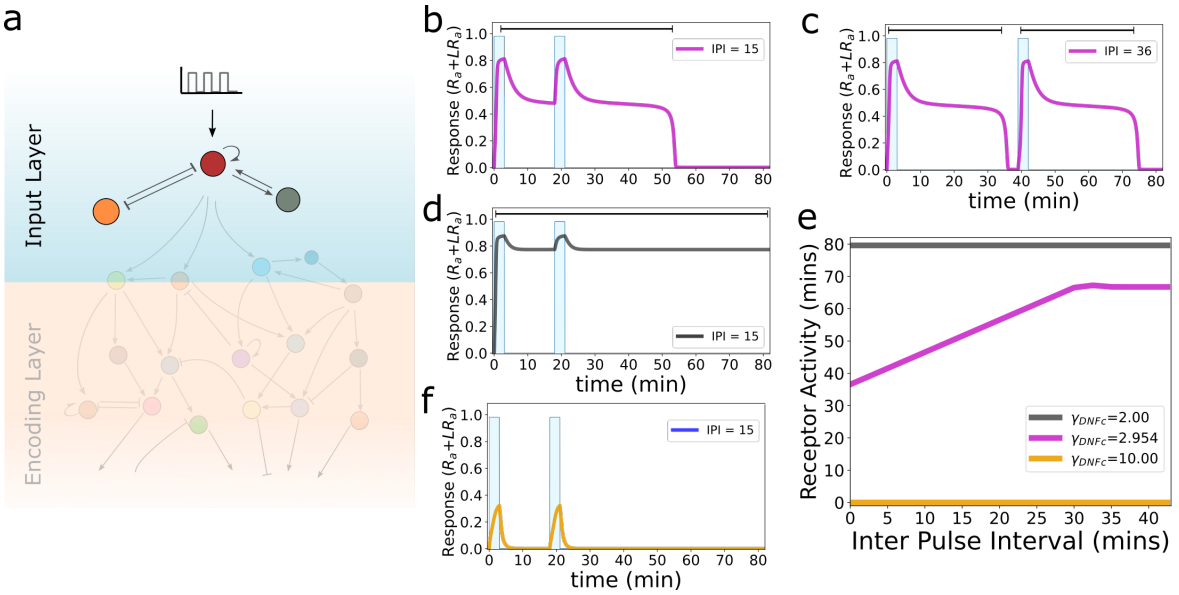

Figure 2.2: **Frequency dependent integration of time-varying signals by minimal receptor network.** (a) Schematic of a minimal receptor network constituting the input layer of intracellular network [85]. Temporal receptor response profiles for 2 consecutive input pulses of 3 minutes and 15 minutes inter-pulse interval (IPI) when the receptor network is organized in (b) criticality, (d) bistable and (f) monostable regimes. (c) Response profile for 2 pulses of 36 minute IPI when system is organized near criticality. (e) Total receptor activity for two consecutive 3 minute pulses with respect to changing IPI for system organization in aforementioned regions.

When the system is organized near 'criticality' with $\hat{\gamma}_{DNF} = 2.954$, the minimal receptor network senses the first pulse and reaches a high receptor activity state. However, after the removal of input pulse, the system is trapped in the 'ghost' state and transiently maintains the memory of first pulse. The second input pulse again make the system reach high receptor activity state followed by a memory state of $\sim$ 34 minutes, after which the system returns back to the basal state. Therefore, both the input pulses are integrated together when they are separated by an IPI of 15 minutes because of transient memory and therefore the total receptor activity duration becomes $\sim$ 52 (Fig. 2.2 (b)). However, when the IPI (36 minutes) is more than the receptor activity duration generated by a single stimulus then both the input pulses generate similar looking activity patterns having a transient memory of $\sim$ 34 minutes followed by returning to basal state. The total receptor activity, in this case rises to $\sim$ 70 minutes (Fig. 2.2 (c)). Therefore, the total receptor activity increases linearly along with increasing IPI, where both the input pulses are integrated together as shown with magenta curve in Fig. 2.2 (e). However, when the IPI becomes higher than the memory duration generated by a single stimulus then the total receptor activity generated by both pulses becomes constant, which is $\sim$ 70 minutes in this case.

This elementary analysis using two input pulses shows the feature of dynamic memory arising from the 'ghost' of saddle-node that enables the receptor network to uniquely integrate input pulses depending on the inter-pulse-interval, as described by [85].

2.2.1 Signal classes of multiple input pulses with different temporal frequencies

The 'ghost' of saddle-node bifurcation enables a transient dynamic memory in receptor activity, which enables signal integration property in the minimal receptor network, which we showed using two pulses with variable IPIs. Now, we systematically analyze how any receptor network and the downstream signaling network process the information from external signals with multiple pulses. For that, we will use a *reference* signal of multiple pulses with a fixed inter-pulse interval between any two consecutive pulses, to generate opposite classes of input signals, namely S_{sim} and S_{diff}. We will use these opposite signal classes of inputs to stimulate the biochemical network to further analyze the differences between the response profiles of the network, using response profiles generated by the *reference* signal as a reference response curve. In this way, we will build a systematic method for analyzing the extracellular information processing property of biochemical networks enabled of the dynamic memory in receptor activity.

We start with generating a *reference* signal having a temporal profile of 3 minutes pulses and 10 minutes inter-pulse intervals (or 3'10' in short) as shown with red curve in Fig. 2.3 (a). We designed similar and different classes of inputs based on the following principle: each of the input pulses in *reference* signal will be probabilistically shuffled to a different time point using a shuffling probability $P_{shuffle} \in [0, 1]$. Where $P_{shuffle} = 0$ means that all of the input pulses of *reference* signal will be reshuffled to any other time point, while $P_{shuffle} = 1$ implies no reshuffling of pulses at all. Following this principle, a signal class having a temporal frequency signature similar to that of *reference* signal is generated with $P_{shuffle} = 0.95$ and we call this calss: *similar* signal class or S_{sim}. Three exemplar input signals belonging to S_{sim} class are shown with blue in Fig. 2.3 (a). In a similar fashion, a signal class having a temporal frequency signature completely different to that of *reference* signal is generated with $P_{shuffle} = 0.1$ and we call this calss: *different* signal class or S_{diff}. Three exemplar input signals belonging to S_{sim} class are shown with green in Fig. 2.3 (a). Each of these classes are generated to have 10 input signals. To quantify the difference of each of these classes with respect to *reference* signal, we use a euclidean distance

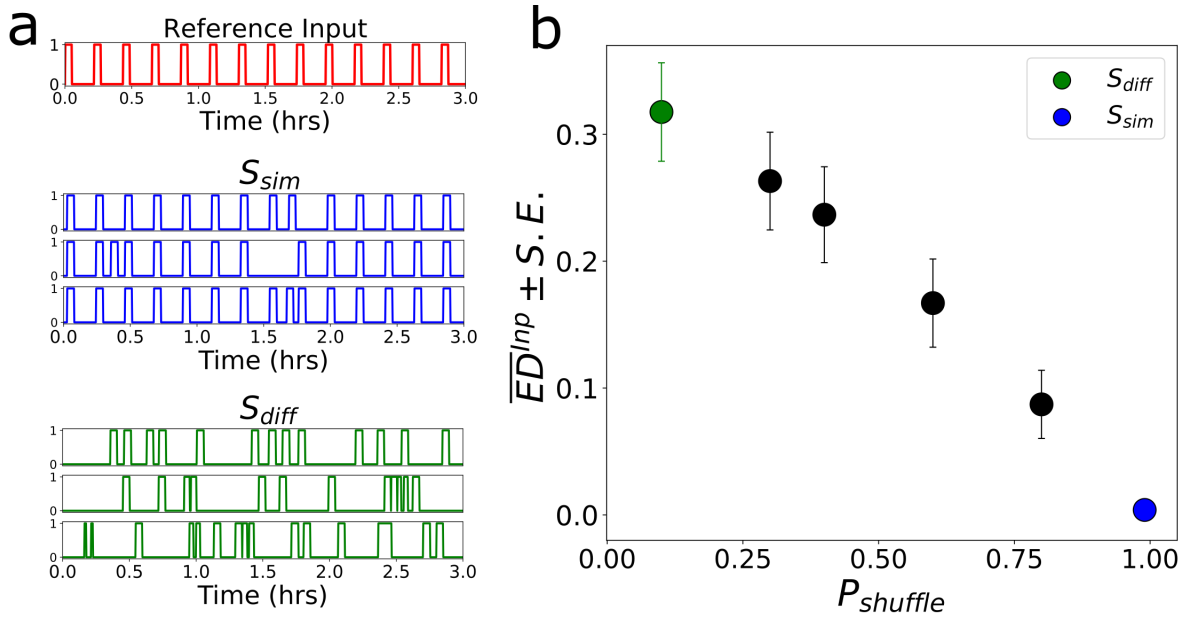

Figure 2.3: **Input classes with different temporal frequencies.** (a) Reference input (red) of 3 hours with a temporal frequency profile of 3'10' is used to generate S_{sim} (blue) and S_{diff} (green) classes of inputs with shuffling probability ($P_{shuffle}$) of 0.1 and 0.95 respectively. (b) Mean euclidean distance ($\overline{ED}^{Inp}$) of inputs calculated with respect to reference input for all generated signal classes. Each signal class contains 10 inputs.

(ED) measure. The mean separation (over time and signal repetitions in each class) between a signal class with respect to *reference* signal can be calculated using Eq. 2.9:

$$\overline{ED}^{Inp}(S^{Ref}, S^{class}) = \frac{1}{N \times T} \sum_{i=1}^{N} \sum_{t=0}^{T} \left| (S^{Ref}(t) - S_i^{class}(t))^2 \right| \tag{2.9}$$

Using the same principle as described above, various signal classes were obtained using a range of shuffling probability ($P_{shuffle} \in (0,1)$) and $\overline{ED}$ is measured for all the obtained signal classes as shown in Fig. 2.3 (b). All the classes which are similar to *reference* signal have a small $\overline{ED}^{Inp} < 0.1$ with respect to the *reference* signal, for example, S_{sim} having $\overline{ED}^{Inp} \simeq 0.004$ with respect to *reference* are shown with blue in Fig. 2.3 (b). While classes which are different from *reference* signal have a large ED with respect to the *reference* signal, for example, S_{diff} having $\overline{ED}^{Inp} \simeq 0.32$ with respect to *reference* are shown with green in Fig. 2.3 (b).

Therefore, we have systematically generated a number of signal classes and quantified them based on ED measure with respect to *reference* signal. These classes of signals will be used in the following sections to analyze the signal integration property and to quantify the separation between response profiles of receptor activity with respect to that of *reference* signal.

2.2.2 Dynamic memory enables maximal separation of response profiles for different class of signals

To analyze the time-varying information processing mechanism of receptor networks present in input layer of large biochemical networks, we use *similar* and *different* classes of signals generated using method described in the previous section. We aim to quantify the characteristic differences between response profiles of minimal receptor network generated by these classes of input signal with respect to that of *reference* signal. We compare this quantification of response profiles when the minimal receptor network is organized in 'criticality' with that of bistable and monostable organization. This further will help in analyzing if the dynamic memory in the input layer of biochemical networks arising from the saddle-node 'ghost' have benefits in signal integration and separation property over systems having permanent or no memory.

We used S_{sim} and S_{diff} classs of signals along with *reference* signal to stimulate the minimal receptor network when it is organized in bistable regime ($\hat{\gamma}_{DNF} = 0.5$). Since the system is parameterized in bistable regime, it always remains in the high receptor activity state once the first input pulse activates the system as shown in Fig. 2.4 (a) with red, blue and green inputs respectively for *reference* signal, one from each S_{sim} and S_{diff} classes respectively. The separation between the obtained response profiles for each of these class inputs with respect to that of *reference* signal can be obtained using the euclidean distance measure. Eq. 2.10 shows a general expression of ED between n-dimensional signal responses x_1 and x_2 at a given time t:

$$ED(x_1, x_2)(t) = \sqrt{\sum_{i=1}^{n}(x_{1,i}(t) - x_{2,i}(t))^2} \tag{2.10}$$

We therefore used Eq. 2.10 to calculate the ED between response profiles of S_{sim} class with respect to that of *reference* signal and took mean over the entire duration (3 hours). The overall separation of signals that are similar to *reference* signal remains below ~ 0.01 when the minimal receptor network have permanent memory with it's parameterization in the bistable regime Fig. 2.4 (c). The mean ED for S_{diff} class with respect to that of *reference* signal is also remain small ~ 0.07. This shows that, in case of bistable organization there is no separation between response profiles in either of S_{sim} and S_{diff} classes as the mean ED remains below 0.2. When the system is

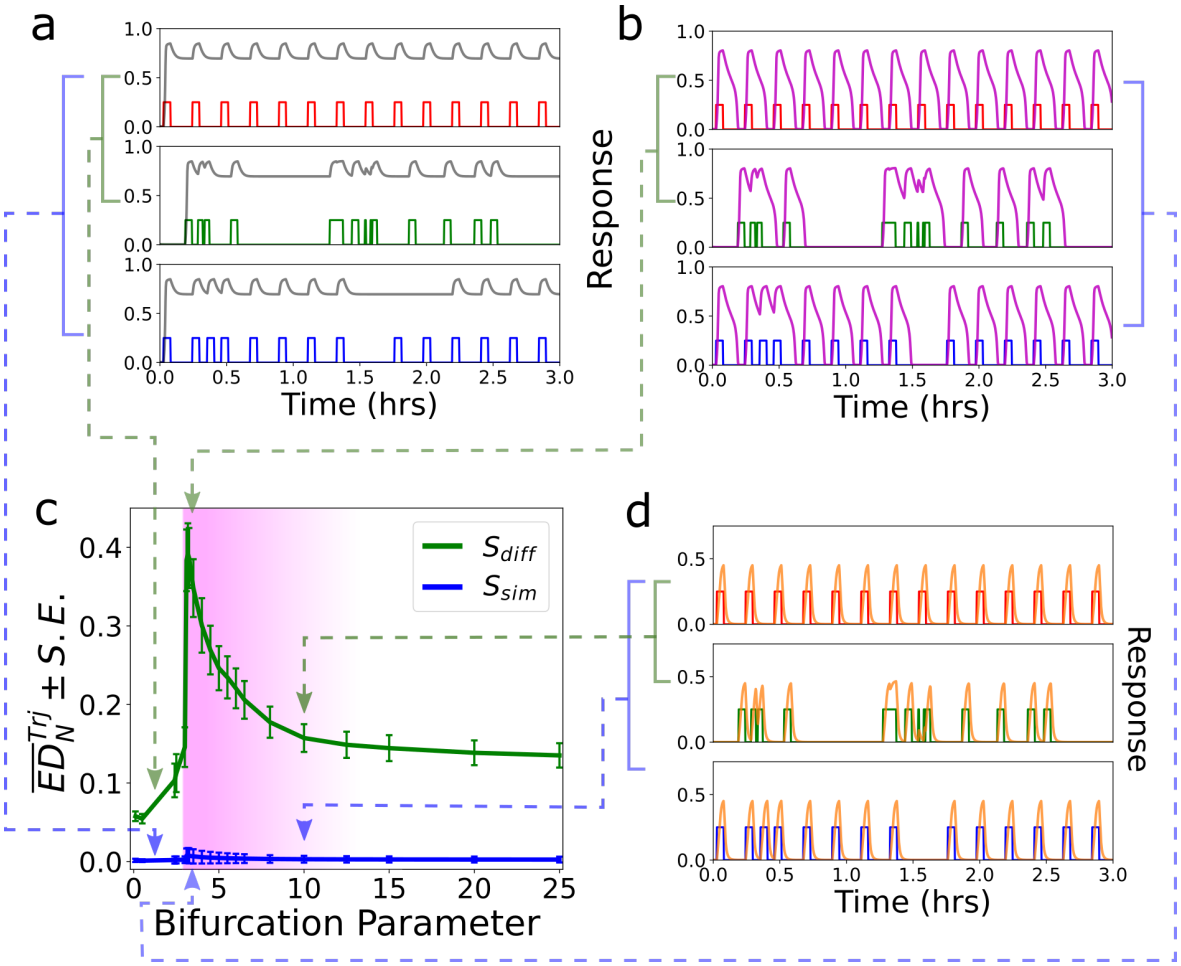

Figure 2.4: **Separation of trajectories in phase space for a minimal two component network.** Response profiles of the minimal two-component network (Fig. 2.2 (a)) for reference signal (red), one inputs from each of S_{diff} (green) and S_{sim} (blue) classes when system is organized respectively in (a) bistable regime (grey), (b) criticality (magenta) and (d) monostable regime (orange). (c) $\overline{ED}^{Trj}$ between trajectories generated by input classes S_{sim} and S_{diff} with respect to *reference* trajectory over the range of bifurcation parameter ($\hat{\gamma}_{DNF}$).

organized in the monostable regime with $\hat{\gamma}_{DNF} = 10$, the the receptor activity follows the input signal and returns back to basal state as soon as the pulse is removed irrespective of temporal signature of input pulses, as shown for *reference* signal, one from each S_{sim} and S_{diff} classes in Fig. 2.4 (d). The results show that in the case of monostable organization as well there is no separation between response profiles in either of two classes as the mean ED remains below 0.2.

However, when this minimal receptor network is parameterized in the vicinity of saddle-node bifurcation with $\hat{\gamma}_{DNF} = 3.18$, signals with small IPI are integrated together while those have large IPI, their responses are separated and are characteristically different. Since S_{sim} class have signals similar to that of *reference* signal, their response curves also look similar. On the other hand, S_{diff} class is maximally

different from *reference* signal and their response profiles in are completely different, as shown in Fig. 2.4 (b). This is reflected in the mean ED where the separation of S_{sim} class response relative to the *reference* signal response remain small ~ 0.01. While the the separation of S_{diff} class response relative to *reference* signal response reaches a maximal value of ~ 0.4. This analysis is systematically carried out for the entire range of bifurcation parameter spanning bistable, monostable and criticality regimes and mean ED is calculated using the method described above as shown in Fig. 2.4 (c).

The mean ED between the receptor response profiles generated by S_{sim} class and that of *reference* signal, remains bounded between $\simeq [0, 0.01]$ over the entire range of bifurcation parameter. This suggests that the signals with similar temporal frequencies are integrated similarly by the receptor network. In contrast, when the inputs belong to S_{diff} signal class, the separation between the receptor response activity is maximized in the critical region (shown with magenta color in Fig. 2.4 (c)). This suggests that the signals with different temporal frequencies are integrated differently by the receptor network organized at 'criticality'. Furthermore, it is also evident that the dynamic memory realized via 'ghost' of saddle-node enables efficient frequency dependant integration of time-varying signals by the receptor network, this property is completely missing when the system is parameterized in bistable or monostable regimes. Therefore, this systematic analysis shows the benefits of dynamic memory in efficient signal integration and separation property of receptor networks organized near 'criticality' over systems having permanent memory or no memory at all. However, the maximal separation achieved at 'criticality' is ~ 0.4, and the difference is relatively low when the system is parameterized in monostable regime (~ 0.15). This is due to the low dimensionality of the minimal receptor network which has only two nodes, receptor R_a and an inactivating enzyme P_{DNFa}. In the following sections, we will therefore explore the high dimensionality case with a biochemical network with several interconnected nodes.

2.3 Dynamic memory in the 'input' layer enables separation of phase-space trajectories in biochemical networks

An arbitrary downstream signaling network of 27 nodes is generated using probabilistic network generation method as explained in Section 2.3.2. The input layer comprising of minimal receptor network is interconnected with the arbitrary signaling network as shown schematically in Fig. 2.5 (a). The receptor network is modeled using law of mass action as discussed in Section 1.4.2, reflecting the response dynamics of cell surface receptors such as receptor tyrosine kinases [84, 85]. The encoding-network is modeled using biochemical systems theory (BST) [98, 95], as described in Section 2.1.1. The activation and deactivation parameters are drawn from distributions normal distributions.

The extracellular information processing and encoding property of biochemical network is analyzed using a minimal time-varying signal consisting of just two consecutive pulses of 5 minutes each with an IPI of 15 minutes. The receptor network is first parameterized in the bistable regime meaning a permanent memory at the level of input layer. Shown with grey curve in Fig. 2.5 (b), the receptor once activated remains in high activity state. This permanent memory makes the system insensitive to upcoming dynamic signals which is reflected in the activity profiles of all the biochemical network nodes, shown with different colors in Fig. 2.5 (b). However, when the receptor network is organized in the 'criticality' regime means that the dynamic memory is present at the level of input layer. Both the pulses are integrated together because of the presence of dynamic memory, receptor profiles are shown with magenta curve in Fig. 2.5 (c). The receptor activity remains high till $\sim$ 45 minutes and then rapidly returns back to the basal state. The dynamic memory effect is reflected in the response profiles of all the biochemical network nodes constituting the input and encoding layers. Finally, the receptor network is parameterized in the monostable regime where the system has no memory. The receptor follows the input signal without reaching the high activity state and quickly returns back to its basal state upon removal of external stimulus as shown with orange curve in Fig. 2.5 (d). The encoding layer nodes follow the same pattern and remain close to their respective basal states.

This arbitrary biochemical network consists of 30 nodes and therefore the activity of multiple nodes can be visualized altogether in a phase-space diagram. For that, we

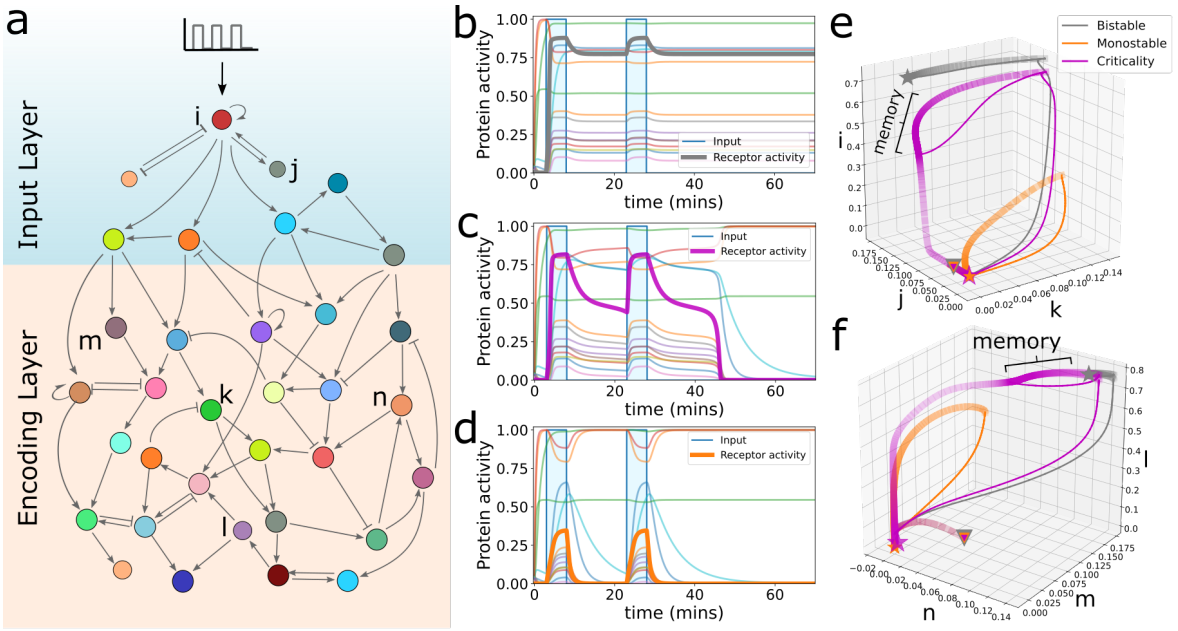

Figure 2.5: **Time-varying external signal processing in an arbitrary biochemical network enabled by the dynamic memory in 'input' layer.** (a) Schematic of an arbitrary biochemical network comprising of 30 nodes with a minimal two-component receptor network in the 'input' layer connected with a synthetically generated signaling network representing the 'encoding' layer. Protein activity temporal profiles of downstream signaling network for an input signal of two consecutive 5 minutes pulses with a 15 minutes IPI when the receptor network is organized in (b) bistable, (c) criticality and (d) monostable regimes respectively. Receptor activity is shown in thick lines of grey, magenta and orange colors respectively in (b), (c) and (d). (e, f) Phase-space trajectories generated by network nodes {i, j, k} and {l, m, n} as marked in (a) respectively. Thin curves in trajectory represent when the input stimulus is present and broad faded curves represent the when the stimulus is absent. Triangles represent initial state at t=0 and stars represent the state of system at t=65 minutes.

randomly selected nodes from the entire biochemical network irrespective of input or encoding layer, marked with {i, j, k, l, m, n} in the network scheme in Fig. 2.5 (a). Two 3D phase-space diagrams are plotted in Fig. 2.5 (e) and (f) comparing the three dynamical regimes namely bistable, monostable and criticality of receptor network shown with grey, orange and magenta curves respectively. In cases of bistable regime, since the system has permanent memory, it is stuck forever in the high activity state (grey star). However, in the monostable case, the system follows the input pulses and returns back to its basal state (orange star) as soon as the pulse is removed thereby generating small loops in phase-space. Interesting is the case when receptor is organized at 'criticality', where high activity state is achieved when input pulse is present (thin magenta curve), but when the stimulus is removed the system gets trapped in the memory region (wide faded magenta curve). From this memory state, it reaches again high activity state by second stimulus followed by trapping in the memory state and eventual returning to the basal state. This is contrary to the bistable and monostable cases where either the system is trapped permanently in

the high activity state or returns back to its basal state. The unique signals integration property enabled by the dynamic memory results in a characteristic trajectory in the high-dimensional phase-space of biochemical network.

This shows that the extracellular information sensed by the cell-surface receptors present in the input layer, traverses throughout the interconnected biochemical network. Furthermore, the dynamic memory present in the input layer, arising from the 'critical' organization of receptor network facilitates the encoding of information present in the time-varying extracellular signals into characteristic phase-space trajectories.

2.3.1 Dynamic memory enables efficient signal-to-phenotype class mapping in biochemical networks

In the previous sections, we have described the signal processing property of biochemical networks which is facilitated by the dynamic memory arising from the 'critical' organization of receptor network. Using a minimal two-pulse input, we determined the benefits of dynamic memory in encoding of information present in time-varying signals over those systems that have permanent memory and systems without memory in their 'input' layers. In this section, we go one step further to investigate how signals are encoded in the biochemical network dynamics that have similar and different temporal profiles.

We use the same arbitrary biochemical network of $N = 30$ nodes (Fig. 2.5 (a)) described in the previous section. To understand the 'similar' and 'different' information encoding property of biochemical networks, we use S_{sim} and S_{diff} classes of signals as described in Section 2.2.1. All the inputs from these two opposite classes of signals along with the *reference* signal are used to stimulate this biochemical network. The information of a dynamic signal encoded in the phase-space trajectory is quantified with respect to the trajectory generated by *reference* signal. Euclidean distance (Eq. 2.10) is used to obtain the separation between trajectories generated by a signal class with respect to that of *reference* signal in 30-dimensional phase-space.

The trajectories generated by S_{sim} signal class and that of *reference* signal remain close to each other in the 30-dimensional phase-space with $\overline{ED}^{Trj} \simeq [0, 0.1]$ over the entire range of bifurcation parameter shown with blue curve in Fig. 2.6. This suggests that the similar signals are encoded within same class of trajectories that have same characteristic features, remain close to each other in phase-space and stay bounded in a defined area in multidimensional phase-space. In contrast, when

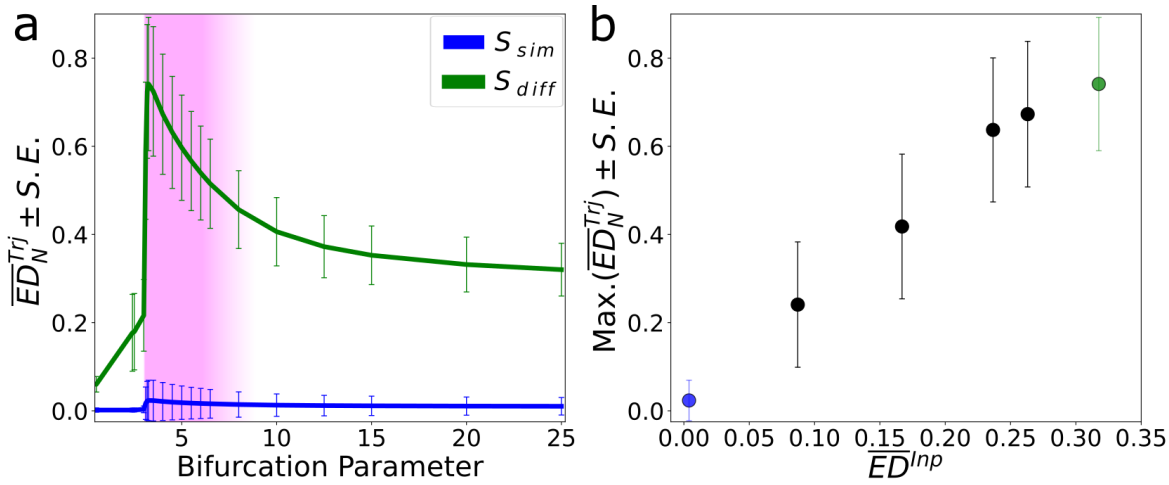

Figure 2.6: **Separation of trajectories in phase space of an arbitrary biochemical network.** (a) $\overline{ED}^{Trj}$ between trajectories generated by input classes S_{sim} (blue) and S_{diff} (green) with respect to *reference* trajectory in 30-dimensional phase-space over the range of bifurcation parameter ($\hat{\gamma}_{DNF}$). (b) Maximum value of $\overline{ED}^{Trj}$ of the trajectories over the entire bifurcation parameter range with respect to the input class $\overline{ED}^{Inp}$.

these inputs belong to different (S_{diff}) class, the separation between the trajectories is maximized in the critical region (magenta color) and remains above 0.4, shown with green curve in Fig. 2.6 (a). When the receptor network is parameterized in the bistable and monostable regimes this separation between trajectories becomes small while in the case of 'critical' organization, trajectories reach a maximal separation of $\sim$ 0.75. This principle was systematically analyzed for various classes of input signals, with a continuous increase of the $\overline{ED}^{Inp}$ from the equivalent ($\overline{ED}^{Inp}$ $\simeq$ 0.004) to the different input class ($\overline{ED}^{Inp} \simeq 0.32$). As shown in Fig. 2.6 (b), the maximal of $\overline{ED}^{Trj}$ between phase space trajectories increases linearly with increase of the separation between input signals.

Depending on the temporal signature of extracellular signals, cells generate different phenotypic responses while maintaining the completely opposed features of *specificity* and *generalization*, as discussed in detail in Section 1.4.5. *Specificity* feature of biochemical information processing arises when completely opposite phenotypic responses are generated by extracellular signals irrespective of growth factor identity. On the other hand, *generalization* emerges when extracellular signals with different temporal signatures result in same cellular response. Whether it's the same or different cellular response, all the information from extracellular environment is encoded in the biochemical network. The encoded information is then decoded by early response genes (ERG), resulting in same or different phenotypic responses accordingly. We hypothesized that the information of same phenotypic response will

be encoded in similar phase-space trajectories while one the other hand, trajectories belonging to different phenotypes will be characteristically different and remain separated in phase-space (Section 1.7). The systematic analysis provided in this section is in line of our hypothesis, that the extracellular signals are indeed encoded in the form of phase-space trajectories in high dimensional phase-space. These trajectories tend to stay close to each other in phase-space when extracellular signals are similar to each other i.e belonging to S_{sim} class. However, when these signals are different from each other (S_{diff}), the information is encoded in different trajectories which remains separated in phase-space, reflected by the large ED. Therefore, the generation of characteristically similar and different phase-space trajectories in case of S_{sim} and S_{diff} classes of signals, reflects the *generalization* and *specificity* feature of biochemical networks respectively. We can therefore conclude that these two completely opposed features of *generalization* and *specificity* are the consequences of signal processing and encoding properties of the biochemical networks enabled by the presence of dynamic memory in the 'input'-layer of biochemical networks.

2.3.2 Role of 'encoding' layer network structure in separation property

The different phenotypic information is encoded through unique phase-space trajectory classes, which can be quantified using a distance measure. In this section, we systematically test the robustness of phenotypic information encoding property of biochemical network against the network structure. In particular, we investigate whether the architecture of 'encoding' layer network affects the phenotype specific information processing and encoding property of the biochemical network. For that, we define network generation methods that changes the overall structure of the initial network controlled by rewiring probabilities as parameters. Different parameter ranges will generate a uniquely structured biochemical network. Therefore, a number of networks with different structural properties can be generated by providing a different range of parameters. First, we define these network generation methods as described below:

1. **Cascade and rewire method:** Starting with N number of nodes connected in a linear cascade (i.e node $1 \rightarrow 2$, $2 \rightarrow 3$ etc.) with activating links. If two nodes are connected then the link will be removed and one of these nodes will be connected to another random node within the network with probability $P_b \in [0, 1]$, with $P_b = 0$ represents no rewiring and $P_b = 1$ meaning all the links

will be rewired. We add more activatory (+1) and inhibitory (-1) links to the network respectively with probabilities P_p and $P_n \in [0,1]$.

2. **Probabilistic method:** Any two randomly selected nodes are connected with probabilities P_p and $P_n \in [0,1]$ respectively with activatory (+1) and inhibitory (-1), and if any node remains disconnected joining it randomly with any other node present in the network.

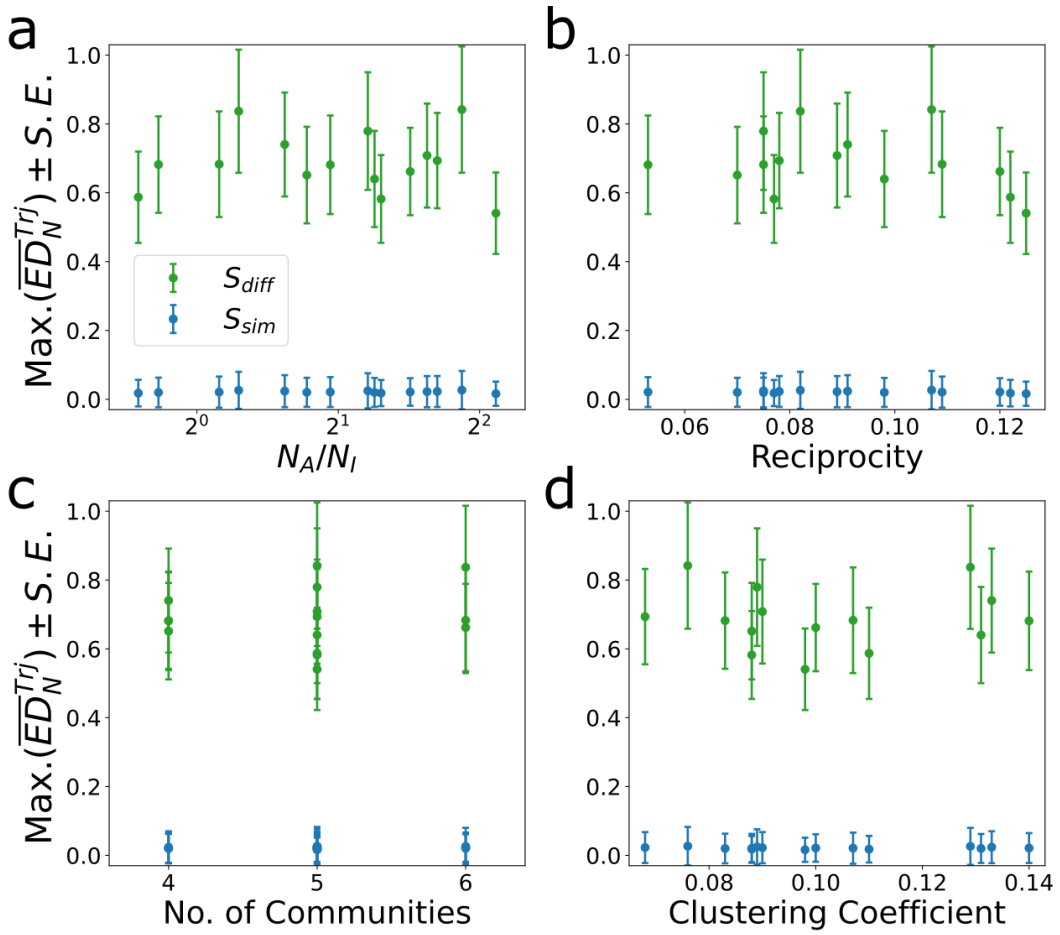

Figure 2.7: **Role of network structure in separation property.** Maximum value (over the bifurcation parameter range) of $\overline{ED}^{Trj}$ of the trajectories with respect to network structural properties of (a) ratio of activatory to inhibitory links (N_A/N_I), (b) reciprocity, (c) number of communities present and (d) clustering coefficient in 15 different generated artificial biochemical networks of size 30 nodes.

We use these two systematically defined network generation methods to obtain 15 different arbitrary 'encoding' layer networks of size $N = 27$, using a range of parameters given in Appendix A, Table A.1. These network parameters are chosen in such a way that the generated arbitrary networks have a range of the four different network properties: N_A/N_I, reciprocity, communities and clustering coefficient. N_A/N_I provides the ratio of activatory to inhibitory links, where 1 represents equal

number of links of both the kind. Reciprocity provides a measure of the likelihood of nodes in a directed network to be mutually linked (loop of 2 nodes). It is defined as the ratio of the number of links pointing in both directions to the total number of edges in the network. Third, the number of communities present in the network are examined, where a community is defined as a subset of nodes within the network such that connections between the nodes are denser than connections with the rest of the network. Last, clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. In other words, it is measure which provides if the neighbours of a node are also connected together i.e formation of triangle in the network.

The minimal receptor network is connected to these arbitrary signaling networks as described in the previous section, constituting a biochemical network of size $N = 30$ nodes. These networks are stimulated with similar (S_{sim}) and different (S_{diff}) classes of input signals along with the *reference* signal described in Section 2.2.1. The separation between trajectories in 30-dimensional phase-space is obtained for each of the signals using ED measure (Eq. 2.10) when the receptor is critically organized near the saddle-node bifurcation point. Maximum of ($\overline{ED}_N^{Trj}$) for both the signal classes, S_{sim} (blue) and S_{diff} (green) are represented against the aforementioned four different network properties.

The maximum $\overline{ED}^{Trj}$ between the trajectories of (S_{sim}) class and that of *reference* signal remains bounded between $\simeq$ [0, 0.1] for all the networks. While the maximal separation between the trajectories generated by (S_{diff}) class of signals remains > 0.6 for all the networks for entire range of all four diverse network structure properties, Fig. 2.7. In case of all four network properties there is no correlation found suggesting the dependency of encoding property of large biochemical networks on the structure of downstream signaling network. This points out the fact that irrespective of 'encoding' network structure, the trajectories remain bounded in the same area of phase-space for similar class of signals, but separate when signals have different temporal profiles, provided that the dynamic memory is present in the 'input' layer which arises when the receptor network is organized near 'criticality'.

Moreover, this systematic analysis based on different network properties suggests that the two completely opposed features of *generalization* and *specificity* are independent of downstream signaling network structure. This observation is contrary to the belief that these features arise because of particular structure of signaling networks [77, 76, 18]. Our results, on the other hand suggests that the features of *generalization* and *specificity* are rather the consequences of efficient signal processing and encoding properties of the biochemical networks enabled by the dynamic memory arising at the 'input' layer of biochemical network.

2.4 Structure of experimentally identified EGFR network

The *specificity* and *generalization* features of signaling networks have been widely argued to arise from particularly organized structures of signaling networks [77, 76, 18]. Bow-tie or hourglass structure has been proposed to be such a network structure. In this network scheme, the information is projected on a relatively small sized core part of the network, connected by diverse and redundant input and output subnetworks with various feedback control loops. Such a structure of signaling network has been proclaimed to exist in various signaling networks: epidermal growth factor receptor (EGFR) [77], toll-like receptor (TLR) [76] and also in mammalian target of rapamycin (mTOR) [18] signaling network. We will discuss here in general the EGFR signaling network that is a highly interconnected complex network (Fig. 2.8 (a)) with importance in regulating growth, survival, proliferation and differentiation in mammalian cells.

A bow-tie is a directed network which can be divided basically into three layers: input, core and output (I-C-O) layers, where the number of nodes in core (red) are less as compared to input (green) and output (magenta) layers as schematically shown in Fig. 2.8 (b). In mathematical term it can be represented as $N_C << N_{(I,O)}$, where N_I, N_C and N_O represents number of nodes in input, core and output layers. In such networks, the information from a broad input layer is compressed to a relatively less number of core nodes, which is then distributed to the nodes of the broad output layer. Recent developments in network structural analysis provides a systematic way of rearranging any arbitrary directed network into a bow-tie architecture [92, 103]. These algorithms work by the identifying strongly connected component of the network representing 'core' where all the nodes are connected to each other, then identifying the input and output layers by analyzing the direction of connections. We used this algorithm to identify the underlying bow-tie structure in the EGFR network with $N = 56$ number of nodes. EGFR network is rearranged into input, core and output layers as shown in Fig. 2.8 (c). The distribution of nodes in I-C-O layers are 21-22-13 respectively, revealing a high number of core nodes than in input and output layers i.e $N_C^{EGFR} > N_{(I,O)}^{EGFR}$, which is completely opposite to the definition of hourglass architecture that has less number of core nodes than input and output layers. Therefore, this systematic *graph*-theoretic based network structure analysis suggests that the EGFR signaling network likely does not have a bow-tie architecture.

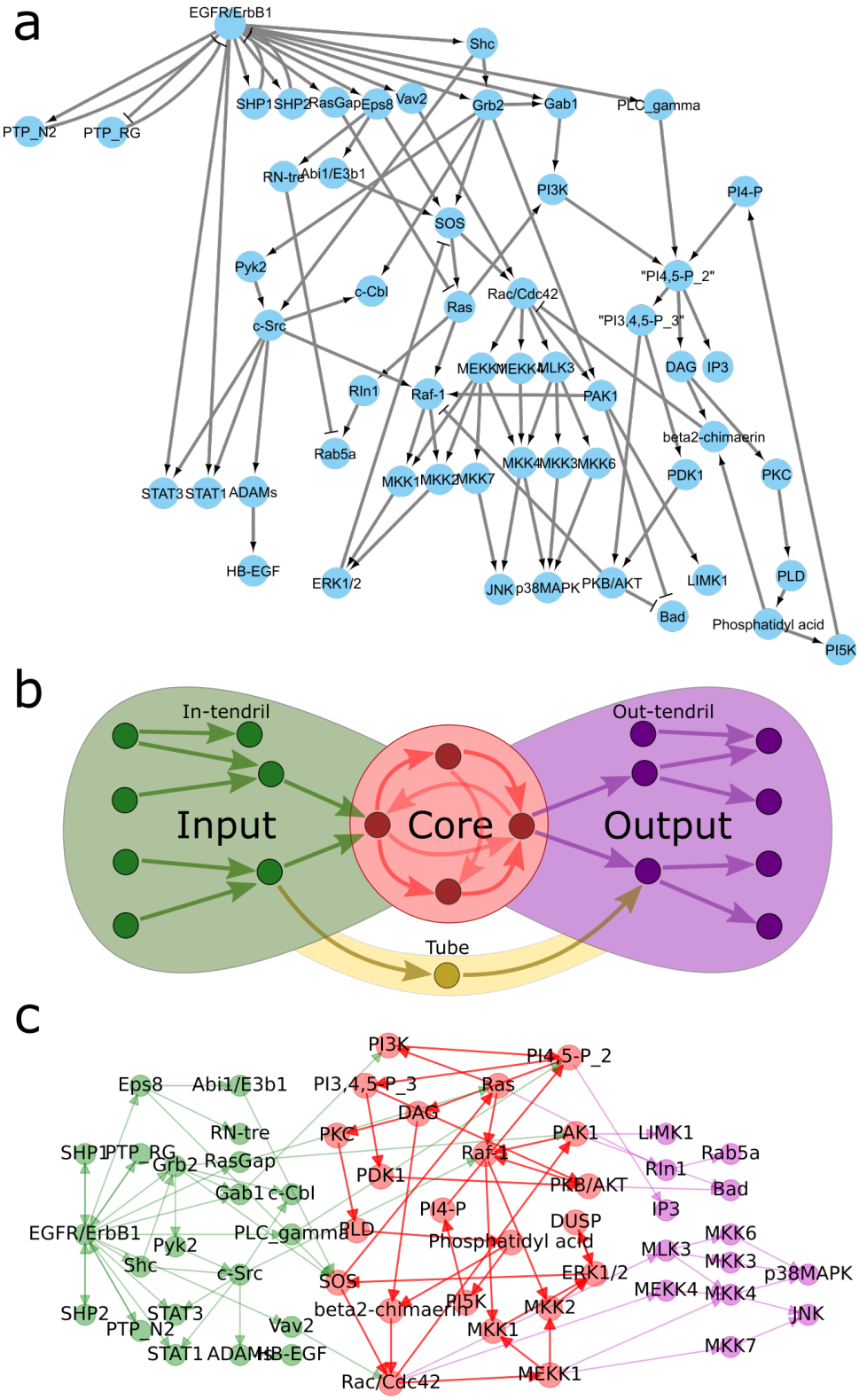

Figure 2.8: **Experimentally identified EGFR network.** (a) Network structure of experimentally identified EGFR network [77]. (b) Schematic of a typical bow-tie network with input (green), core (red) and output (purple) (I-C-O) layers. (c) EGFR network after being rearranged into 3 layers of bow-tie architecture.

2.4.1 Comparision of EGFR signaling network with artificial bow-tie and random networks

The information in a bow-tie network is projected to the core layer which has less number of highly interconnected nodes relative to its large input and output layers. In other words, core layer has different network properties relative to the input and output layers, in terms of fraction of nodes and their connectivity within in the layer. Therefore, in order to systematically characterize the network structural properties of EGFR signaling network, it should be directly compared with synthetically generated bow-tie or hourglass networks. Additionally, we also compare the network properties of Erdős-Rényi random networks as reference networks alongside EGFR and bow-tie networks. Both the network types are generated systematically to have network density comparable to that of EGFR signaling network ($\delta_{EGFR-Net.} = 0.0295$). The density of a directed network with N number of nodes and M number of links is given by the following Eq. 2.11:

$$\delta = \frac{M}{N(N-1)} \tag{2.11}$$

We start with generating synthetic bow-tie or hourglass networks to have input and output layer nodes to be more than core nodes ($N_C << N_{(I,O)}$) using the following steps: First, starting with nodes $N \in [40, 60]$ and dividing them in three parts each for I-C-O layer. Division of N nodes should be in such a away that core should always get least number of nodes, in order to preserve the hourglass structure. Therefore, total nodes in core (N_C) is always between 5 and 25% of N. Total number of links M for each hourglass network is defined by $M \simeq \delta_{EGFR} * N(N-1)$, in order to have a comparable network density with that of EGFR signaling network. Then comes the part of specifying links within each layer and also between them. Randomly dividing (M) edges among I-C-O layers such that:

(i) each node must have atleast 1 edge within the layer.
(ii) dividing rest of the edges in a way such that 75% connections are within the core, 12.2% connections are from input to core and 12.2% connections are from core to output layer.
(iii) rest 0.6% of edges: connecting them between any two randomly selected nodes irrespective of the layer.

The final obtained adjacency matrix will therefore represents a synthetic network with bow-tie architecture. Random networks are generated in such a way to also have their density relatable to that of EGFR signaling network, using the `Networkx` package in `Python`. 100 networks are generated for each of the network type. The generated random and bow-tie networks are rearranged into I-C-O layers of bow-tie architecture using the same algorithm [92, 103] as used in the previous section for EGFR network, to further analyze the properties arising from their global architecture. First, percent of total nodes divided in these layers for both the network types are compared with that of EGFR network. The number of core nodes in synthetic bow-tie networks remains less as compared to the input and output nodes, $N_C^{Bow-tie} << N_{(I,O)}^{Bow-tie}$ as expected. Surprisingly, the distribution of EGFR signaling network nodes into I-C-O layers remain similar to that of random networks with relatively high percentage of core nodes, $N_C^{Random} > N_{(I,O)}^{Random}$ as shown in Fig. 2.9 (a). This means that if any randomly generated network is rearranged into a bow-tie/hourglass architecture having I-C-O layers, the core layer will end up having more number of nodes contrary to the very definition of hourglass architecture. This also suggests that the bow-tie networks are most fragile at the core layer, because if any of those few core nodes are damaged, then the flow of information over entire bow-tie network will be disrupted. On the other hand, the biochemical information is spread throughout all the nodes of randomly interconnected EGFR signaling network.

To further delineate the structural properties of I-C-O layer nodes, we use betweenness centrality (BC) and clustering coefficient (CC) to compare the similarities and differences with respect to that of synthetic bow-tie and Erdős-Rényi networks. BC is a way of detecting the amount of influence a node has over the flow of information in a graph. BC provides the extent to which a node lies on paths between other nodes [73]. Nodes with high betweenness may have considerable influence within a network by virtue of their control over information passing between others. They are also the ones whose removal from the network will most disrupt communications between other vertices. Therefore, as expected core nodes have high BC in case of bow-tie networks relative to its input and output nodes as show in Fig. 2.9 (b). However, Erdős-Rényi networks along with EGFR signaling network has even higher BC values relative to their respective input and output layer nodes which is starkly different to the bow-tie case.

Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. When the connections are dense, the CC is high. It provides the probability that 2 neighbors of a node are also neighbors i.e formation of a triangle in a graph [73]. Core layer of a synthetic bow-tie network have a high CC relative

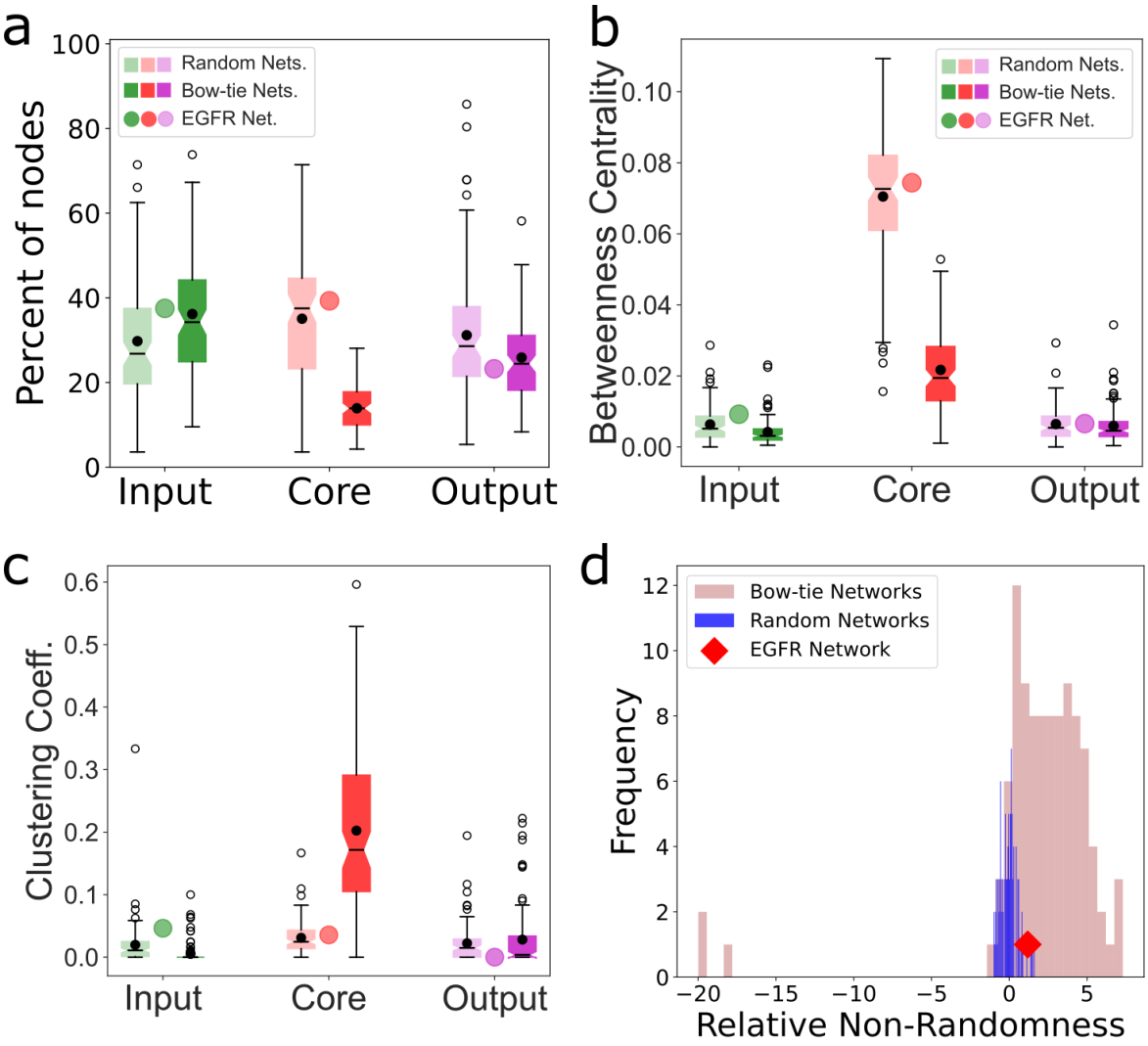

Figure 2.9: **EGFR network structure comparision with Random and Bowtie networks.** (a) Comparing distribution of nodes in input (green), core (red) and output (purple) layers in EGFR network with (100) randomly connected networks and (100) pure bow-tie networks. Using (b) Betweenness centrality and (c) clustering coefficient to further analyze the structural differences between the nodes in I-C-O layers in random and bowtie network types. (d) Using relative non-randomness to compare the structural similarity of the entire aforementioned networks with Erdős-Rényi networks which has 0 non-randomness values [106].

to its input and output layers, meaning that the bow-tie networks have uneven distribution of connections and are more densely connected with high CC at core layer as shown in Fig. 2.9 (c). On the other hand, both EGFR and Erdős-Rényi random networks has small CC in all three I-C-O layers, reflecting an even distribution of connections all over the network and also that these networks do not have any locally clustered nodes carrying majority of the information, rather it is distributed throughout the network evenly.

The analysis showed that the EGFR signaling network is similar to the structural

properties of an Erdős-Rényi network rather than that of bow-tie or hourglass network. Therefore, to further understand the global property of EGFR signaling network with that of synthetic bow-tie and Erdős-Rényi networks, we computed the degree of non-randomness present in each of these networks using a relative non-randomness measure [106]. This measure indicates to what extent a graph G is different from random graphs in terms of probability. When it is close to 0, the graph tends to be more likely generated by an Erdős-Rényi model. If a network or graph $G(n, m)$ is a set of n nodes connected by a set of m links E. The network under consideration is binary, symmetric, connected, and without self-loops. Let $A = (a_{ij})_{n\times n}$ be its adjacency matrix, $a_{ij} = 1$ if node i and j are connected and $a_{ij} = 0$ otherwise. Let λ_i be the eigenvalues of A and x_i the corresponding eigenvectors. Then, the graph non-randomness of the overall graph G can be calculated using the following Eq. 2.12:

$$R_G = \sum_{(u,v)\in E} R(u,v) \quad = \frac{1}{2}\sum_{u\in G} R(u) = \sum_{i=1}^{k} \lambda_i \tag{2.12}$$

The relative non-randomness of the overall graph $G(n, m)$ can be calculated as (Eq. 2.13):

$$R^*_G = \frac{R_G - [(n-2k)p + k]}{\sqrt{2kp(1-p)}} \tag{2.13}$$

where $p = \frac{2km}{n(n-k)}$ and k is the number of communities in graph G. Non-randomness between two graphs G_1 and G_2 can be compared. If $|R^*_{G_1}| < |R^*_{G_2}|$ then one can conclude that G_1 is more random than G_2. As expected, Erdős-Rényi(ER) networks have relative non-randomness measure close to 0, shown with blue in Fig. 2.9 (d). On the other hand, synthetic bow-tie networks (light brown) have large relative non-randomness values, meaning that the bow-tie networks are less random because of their specifically designed structure and distribution of nodes in different layers. Moreover, the relative non-randomness value of EGFR signaling network (red diamond) also remained close to 0 and thereby fall in the distribution of ER networks. These 3 network types can be compared as $|R^*_{EGFR}| \sim |R^*_{Erdos-Renyi}| < |R^*_{Bow-tie}|$.

Therefore, using these systematic methods based on various network structural analysis and measures from *graph-theory*, we found that EGFR network is structurally more similar to a randomly connected network. This implies that the dynamic extracellular information sensed and processed by the signaling network is

not compressed at any 'core' layer, rather it is distributed throughout the high dimensional *encoding* layer. This suggests that the mapping of inputs through phenotypic outputs likely occurs through signaling transients.

2.5 Experimental examples of specificity and generalization in signaling networks

In the previous section, using *graph-theoretic* methods we identified that experimentally identified EGFR signaling network is structurally more similar to a random graph. This suggested that the extracellular information sensed and processed by the cell surface receptor network is distributed throughout the signaling network, rather than being projected upon any low dimensional intermediate layer. To understand this observation in detail and to validate our generalized theory of biochemical information processing based on transients, in this section, we will use experimentally identified biochemical signaling networks.

Experimental studies have shown that stimulating PC12 cells in microfludic device with various frequencies of EGF signals generate completely opposite phenotypic responses [78, 11], discussed in Section 1.4.4. Different temporal frequencies of EGF results in proliferation and differentiation. This represents a perfect example exhibiting completely opposed signal processing features of biochemical networks: *generalization*, where same cellular response is generated when frequency of external signals are different and also *specificity*, when different frequencies of signals are interpreted differently generating opposite phenotypic responses. We therefore use the experimentally identified EGFR signaling network (Fig. 2.8 (a)) to probe numerically whether the concept of information processing with transients provides explanation for the PC12 cells observations. Further, we will particularly use the experimentally established epidermal growth factor sensing receptor network at the plasma membrane.

EGFR network has been experimentally shown to exhibit dynamic extracellular signal processing capabilities which can interpret temporally and spatially complex growth factor patterns [84, 70]. EGFR network, established on the plasma membrane of the cells processes information as a function of stimulus history through reaction kinetics and vesicular dynamics of the receptor along with spatially established interactions with protein tyrosine phosphatases (PTPs). EGFR system is dynamically organized at 'criticality' which is regulated by a spatially-distributed

network of PTPRG and PTPN2 by maintaining the vesicular trafficking of the receptors, discussed in Section 1.4.1. Therefore, considering the nodes that regulates the activity of EGFR with direct feedbacks: PTPRG, PTPN2, SHP1 and SHP2. The equations of the overall EGFR network can be written with the law of mass action, given in Appendix A.

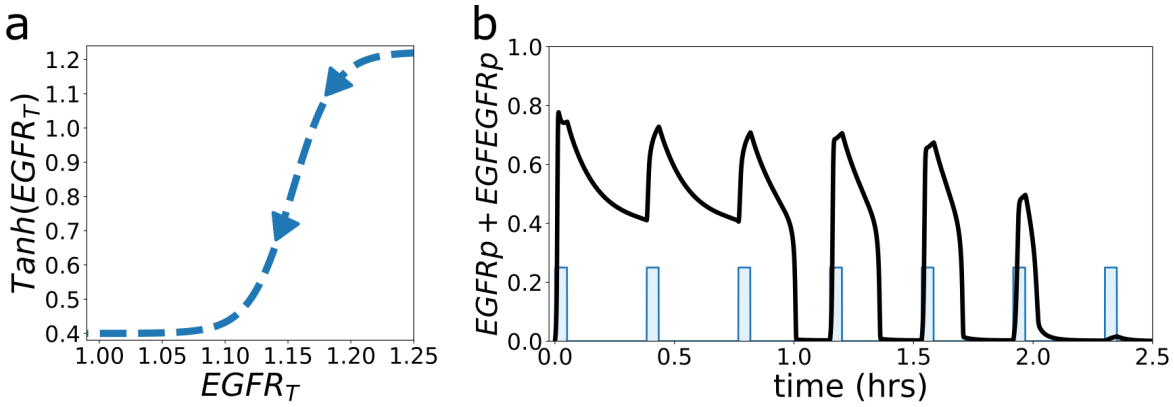

Figure 2.10: **EGFR degradation and response temporal profile with a time-varying EGF stimulation.** (a) *Tanh* function used to control the rate of $EGFR_T$ degradation after every pulse of EGF stimulation. (b) EGFR response temporal profile (black) for an input (blue) of 3 minutes consecutive pulses with 20 minutes of IPI or 3'20'.

These EGFR network equations are used to obtain the bifurcation diagram with respect to $EGFR_T$ as the bifurcation parameter. The magenta region represents the 'critical' regime of saddle-node bifurcation obtained at $EGFR_T^{SN2} = 1.23$, where the EGFR receptor network is dynamically organized [84, 85]. However, the ligand-bound receptors are unidirectionally internalized and degraded while the ligandless receptors that have been internalized are deactivated and recycled back to the plasma membrane. Furthermore, the unidirectional internalization of ligand-bound receptors decreases the EGFR concentration on the plasma membrane. In other words, this decreases the receptor's steady-state concentration upon each growth factor stimulus, effectively shifting the system towards monostable regime away from 'critical' organization near the saddle-node bifurcation [52]. Therefore, this process results in the loss of dynamic memory which arises in the receptor activity because of the 'ghost' of saddle-node bifurcation. The decrease of the total receptor concentration at the plasma membrane is modeled using a nonlinear *Tanh* function as represented by the following Eq. 2.14:

$$EGFR_T^{New} = tanh(EGFR_T^{Old}) \tag{2.14}$$

Starting from the 'critical' organization near SN2 at $EGFR_T^{SN2} = 1.23$, receptor concentration is iteratively removed from the plasma membrane at the end of every growth-factor pulse as shown in Fig. 2.10 (a). A representative response profile ($EGFR_p + EGFEGFR_p$) is shown in Fig. 2.10 (b) when the EGFR receptor network is organized near 'criticality' and the degradation follows according to stimulation by a time-varying signal having a temporal frequency of 3 minutes consecutive pulses with 20 minutes inter-pulse intervals. The dynamic memory arising from 'critical' organization of receptor network near the saddle-node enables the integration of first three input pulses altogether. The memory duration shortens after the fourth pulse and the system eventually reaches the monostable regime, where the input pulse cannot activate the system anymore and therefore it remains in the basal steady state. This mimics the experimental observations where EGFR sensing system integrates the consecutive EGF pulses and gradually shifts away from 'critically' [84, 85].

2.5.0.1 Noise in biochemical networks

In order to get reasonably more close to experimental observations and real biochemical networks, we also account for stochasticity in our simulations. For that, we performed stochastic simulations using additive Gaussian white noise, where the noise intensity was estimated from experimental measurements of temporal EGFR and ERK activity profiles. The noise estimation method we employed, is based on the functional dependency of coarse-grained correlation entropy $K_2(\epsilon)$ on the threshold parameter ϵ [47]. The correlation function $K_2(\epsilon)$ depends on the standard deviation σ of the noise in a characteristic way. Therefore, by approximation of $K_2(\epsilon)$ one can estimate the noise level σ present in experimental data. The coarse-grained correlation entropy can be approximated using Eq. 2.15:

$$K_2 \approx \ln \frac{\langle n \rangle - 1}{\langle n \rangle - 2} \tag{2.15}$$

where $\langle n \rangle$ is the average diagonal line of length greater than 1. To estimate the noise level σ one can use correlation entropy $K_{noisy}(\epsilon)$ and fit $K_{noisy}(\epsilon)\epsilon^p$ to corresponding experimental data. Therefore, one needs to estimate five free parameters κ, σ, a, b, and c for the function given by Eq. 2.16:

$$K_{noisy}(\epsilon)\epsilon^p = -c\epsilon^p g\left(\frac{\epsilon}{2\sigma}\right)\ln\epsilon + [\kappa + b\ln(1-a\epsilon)]\epsilon^p\left(1+\sqrt{\pi}\frac{\sqrt{\epsilon^2/3+2\sigma^2}-\epsilon/\sqrt{3}}{\epsilon}\right) \tag{2.16}$$

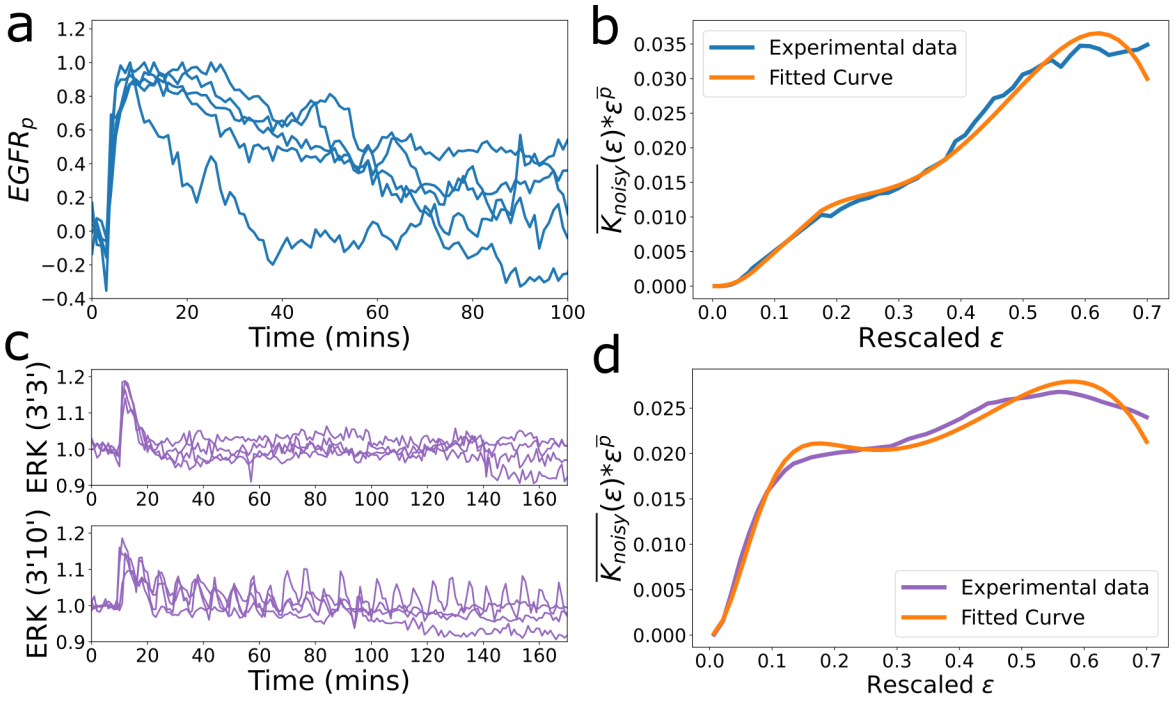

Figure 2.11: **Noise level estimation in experimental EGFR and ERK activity profiles.** (a) $EGFR_p$ response time profile of 5 cells after stimulation by a single 5-min pulse of EGF. (b) $K_{noisy}(\epsilon)$ function is fitted (in orange) for the entropy calculated from Eq. 2.15 for $EGFR_p$ data. (c) ERK response time profile with EGF stimulation of signal frequencies 3′3′ (top) and 3′10′ (bottom). (d) $K_{noisy}(\epsilon)$ function is fitted (in orange) for entropy calculated from ERK data. Average $K_{noisy}(\epsilon)$ value is taken over repetitions of cell numbers in both cases.

We used this method on single 5-minute EGF pulse experiments performed on MCF7 cells to estimate the strength of noise present at the cell surface receptor level [70]. EGFR phosphorilation time profiles used for noise estimation are shown in Fig. 2.11 (a). Each of the $EGFR_p$ response curves are used to obtained the entropy K_2 using Eq. 2.15. Then, the correlation entropy $K_{noisy}(\epsilon)$ curve is fitted to the experimental entropy using `curve_fit` function in `scipy.optimize` module of `Python`, as shown in Fig. 2.11 (b). Of all other fitted parameters, σ gives the estimate of noise present in experimental data, which is $\sigma = 0.023 \pm 0.0133$. Therefore, the amount of noise present at the receptor level is estimated to be $\mathcal{O}(10^{-2})$. We also estimated the amount of noise present in downstream signaling network as well. For that we used experimentally obtained ERK response profiles, with EGF stimulation by signals of temporal frequencies 3′3′ and 3′10′ as shown in Fig. 2.11 (c). Following the

same procedure as described above, $\sigma = 0.009 \pm 0.0099$ when the correlation entropy $K_{noisy}(\epsilon)$ curve is fitted to the experimental entropy as shown in Fig. 2.11 (d). Therefore, the amount of noise present in the downstream signaling network level is estimated to be $\mathcal{O}(10^{-3})$. The noise level estimated to be present in downstream signaling level is found to be less by an order of magnitude of $\mathcal{O}(10^{-1})$ relative to the noise present in EGFR receptor level.

2.5.1 Proliferation and differentiation class trajectories remain separated in phase-space for EGF stimulation

With input signal dependent degradation of receptors from plasma membrane and noise incorporated in receptor and downstream signaling nodes, the 56 nodes EGFR signaling is used to test signals belonging to different phenotypic classes obtained experimentally. The downstream EGFR signaling network is modeled using the same BST Eq. 2.2 discussed in Section 2.1.1. EGFR signaling network is simulated for various dynamic externals signals belonging to proliferation and differentiation classes, summarized in Table. 1.1. Starting with time-varying signals belonging to proliferation, time profiles of ERK responses corresponding to EGF stimulation with 3'3' and 3'60' signals are shown in Fig. 2.12 (a) with orange and blue curves respectively. Trajectories generated in first 1 hour for both the signals in 3D phase-space of [$EGFR_p$, ERK, MEKK1] is shown in Fig. 2.12 (b). Darker parts of the trajectories represent the presence of input pulses while the lighter parts represent the system dynamics in IPI duration, or when the input pulse is not present. Trajectories for both the signals remain close to each other and generate a characteristic shape in phase-space. Time profiles and 3D phase-space trajectories corresponding to differentiation signals [3'10', 3'20'] are shown in Fig. 2.12 (c, d) respectively with green and red curves. Both the differentiation class trajectories cover the same area of phase-space and generate a characteristic shape that is distinct from that of proliferation class. The trajectories generated in case of signals belonging to proliferation class remain in the outer periphery of the volume covered by the trajectories of differentiation class.

These characteristic differences (and similarities) between trajectories generated by signals of different (and same) phenotypic classes can be quantified using a euclidean distance (ED) given by Eq. 2.10 in Section 2.2.2. In order to compare the shapes of trajectories within the same phenotype class, the trajectories belonging to same class are first dynamically time warped to remove the temporal dependencies using dynamic time warping (DTW) method. The DTW is carried out using

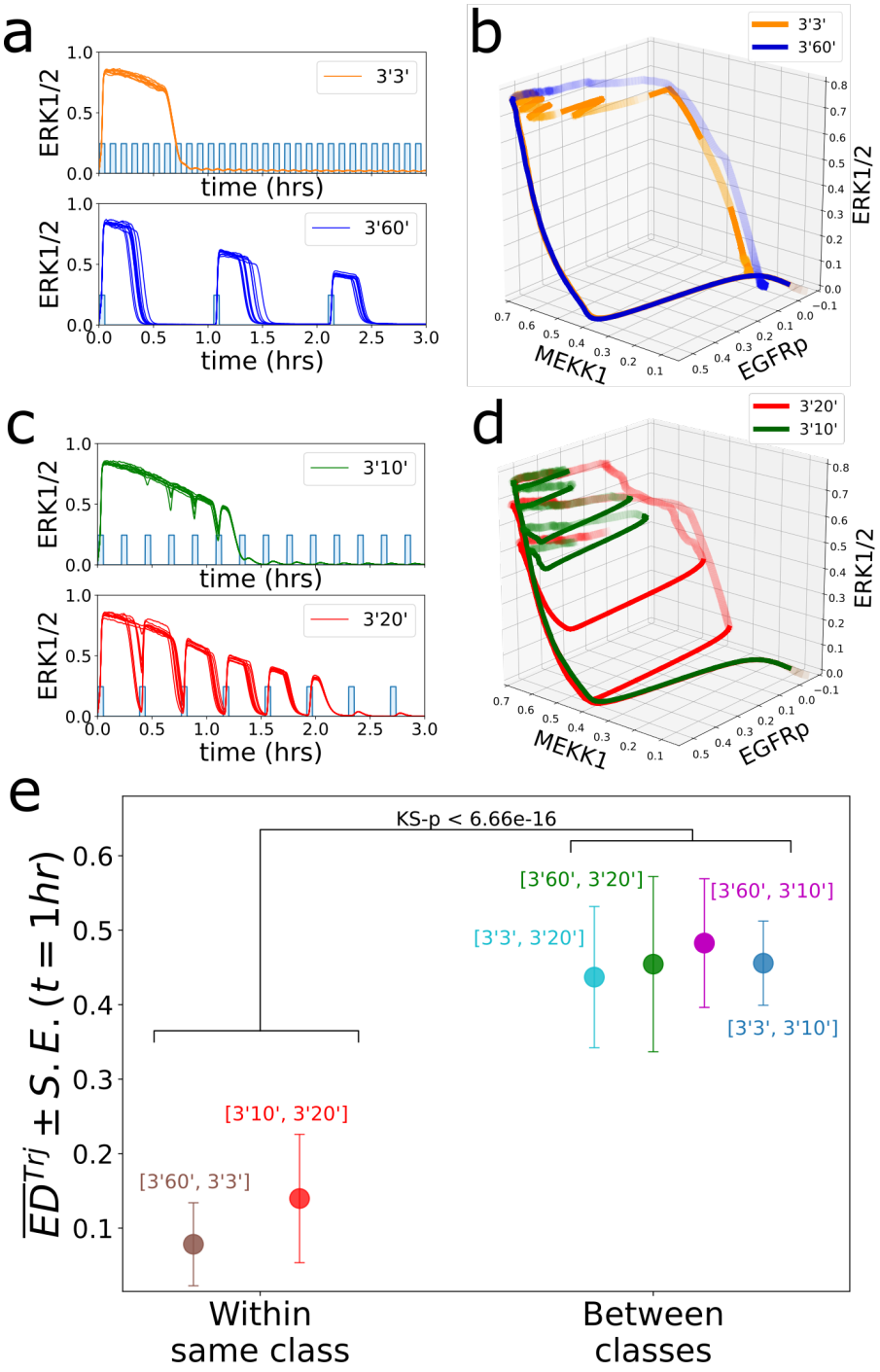

Figure 2.12: **Simulations of large EGFR network with different frequencies of EGF stimulation corresponding to proliferation and differentiation.** (a) ERK response temporal profiles for input signals 3′3′ (orange) and 3′60′ (blue) corresponding to proliferation and (b) phase-space trajectories in 3-dimensions of respective input signals in 1 hour of input stimulus. (c) ERK response temporal profiles for input signals 3′10′ (green) and 3′20′ (red) corresponding to differentiation and (d) phase-space trajectories in 3-D of respective input signals in 1 hour of input stimulus. (e) Mean euclidean distance ($\overline{ED}^{Trj}_{N=3}$) in 3-D calculated in 1 hour for all combinations of signals belonging to same class i.e within proliferation and differentiation class and across these two classes.

the `DTAIDistance` package in Python. Mean ED is calculated for 10 repetitions of each input signal and over 1 hour of duration. First, the separation is calculated between the same class of trajectories, meaning within the proliferation class and also within the differentiation class, shown in Fig. 2.12 (e). The mean ED remains low and for both cases it is < 0.3. Low ED value represents that the considered trajectories remain close to each other in phase-space. Then, the separation of trajectories for all possible combinations of signals belonging to different phenotypic classes is calculated. As shown in Fig. 2.12 (e), the mean ED remains above 0.3 for all the combination of signals. High ED represents that the considered trajectories have different characteristic curves and are separated in phase-space.

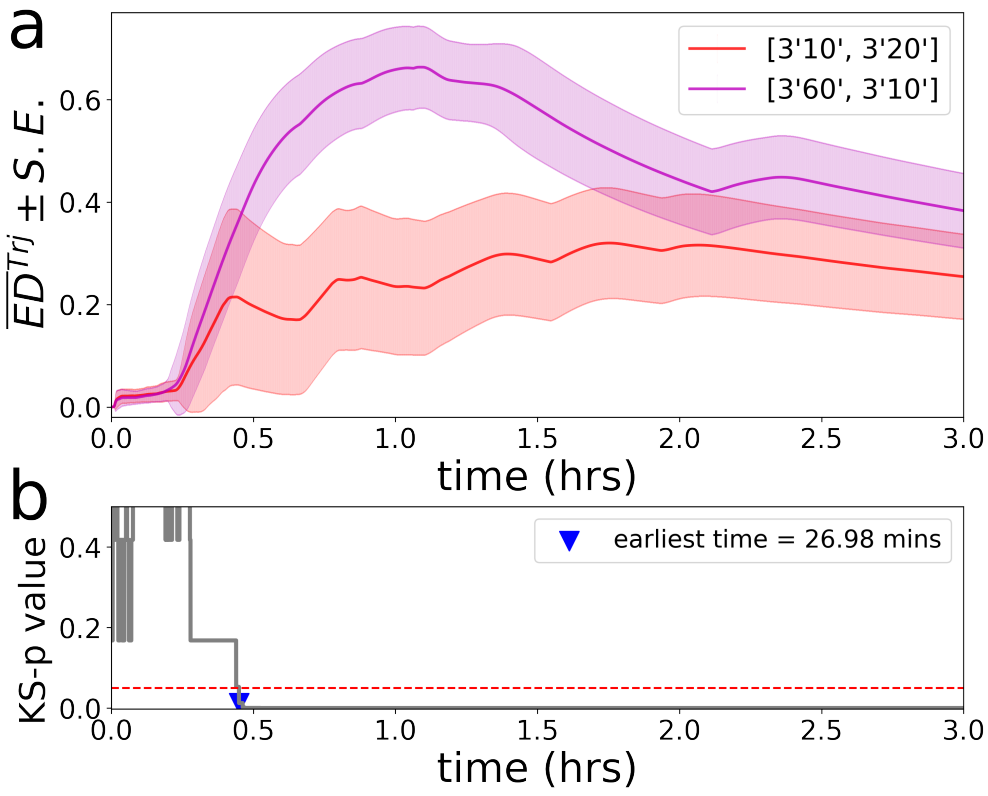

Figure 2.13: **Earliest time-point of separation between same and different phenotypic class trajectories.** (a) Mean ED over time between trajectories of inputs belonging to same phenotypic class: differentiation - 3'10' and 3'20' (magenta) while ED between that of different phenotypic classes: 3'60' (proliferation) and 3'10' (differentiation) (red). (b) p-values over time of Kolmogorov-Smirnov (KS) test between the two EDs in (a).

The results demonstrate that the phase-space trajectories generated by inputs resulting in same phenotypic responses are characteristically similar to each other. On the other hand, trajectories resulting in opposite phenotypic responses have distinctive trajectories spanning different areas of phase-space. In other words, inputs resulting

in same phenotypic outcomes will generate trajectories having similar characteristic features in phase-space.

This observation raises another important question: if the information of cellular responses are encoded via transients, then how long after the stimulation by a dynamic signal the information about a phenotype is present in the dynamics of the protein interaction network. This can found by identifying the time when these trajectories get separated in phase-space. Therefore, we obtained the ED between trajectories of same phenotypic class [3'10', 3'20'] and also between trajectories of different phenotypic classes [3'60', 3'10'] over every time point as shown in Fig. 2.13 (a) with red and purple curves respectively. As expected, the ED between the trajectories generated from different classes of inputs remain higher as compared to the ED between trajectories of same input class, but the separation is observed $\sim$ 27 minutes, shown by a blue triangle in Fig. 2.13 (b).

Therefore, the dynamic signals belonging to different phenotypic classes are integrated differently by the EGFR signaling network. This information is transmitted and encoded in the downstream signaling network in the form of characteristically different phase-space trajectories enabled by the dynamic memory in the input layer. Furthermore, information about the same phenotype is encoded in similar trajectories which have characteristically similar features and remain close to each other in phase-space. Trajectories belonging to different phenotypic classes, however remain close-by for $\sim$ 27 minutes and separate afterwards. This gives an insight into the duration a cell takes for making decision about completely opposite phenotypic response for instance, proliferation or differentiation, depending on the time varying extracellular signals.

2.5.2 Experimental evidence for the encoding of phenotypic information via phase-space trajectories

We next used the publicly available experimental data in which external dynamic signals are used to generate completely opposite phenotypic responses of proliferation and differentiation, depending on their temporal profiles [78]. The data available is only of ERK response profiles i.e 1D time series. Therefore, we use the method of reconstructing multidimensional trajectories in phase-space using time-delay embedding method [89, 48, 66]. The single dimensional observations are a projection of the multidimensional state space of the system onto the one-dimensional axis. The purpose of the time-delay embedding is to unfold the projection back to a multivariate state space that is representative of the original system

[48]. The minimum embedding dimension is obtained using *False Nearest Neighbor* (FNN) method proposed by Kennel et al. [48]. A false neighbor is a point in the data set that is a neighbor when the embedding space is too small for the attractor. However, when a large enough embedding space is achieved, all the neighbors of every attractor point in the multivariate phase-space will be the true neighbors. The optimal time-delay (τ) is obtained using the mutual information (MI) which was suggested by Fraser and Swinney [28]. The MI takes into account nonlinear correlations and if the time delayed mutual information exhibits a marked minimum at a certain value of τ, then this is a good candidate for a reasonable time delay.

ERK response profiles with time-varying external signals of 25 ng/ml EGF stimulation is used for reconstructing the trajectories in multidimensional phase-space. First, the data of each cell (total 30 cells) is denoised using `signal_savgol_filter` in `Scipy` module of `Python`. The average was taken over all cells of denoised data for each of the signal frequency i.e [3'3', 3'10', 3'20', 3'60']. MI is calculated for each signal frequency data, and obtained the time-delays of [3, 2, 3, 3] respective to signal frequencies as shown in Fig. 2.14 (a). The minimum embedding dimensions are estimated to be 3 for all signal frequencies, Fig. 2.14 (b).

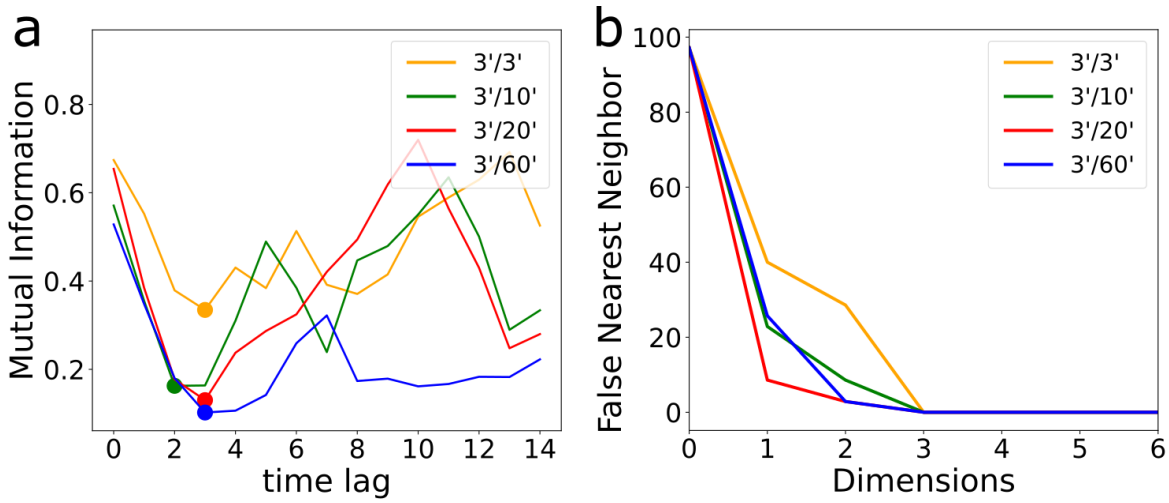

Figure 2.14: **Mutual Information (MI) and False Nearest Neighbor.** (a) For EGF stimulation, Mutual Information is calculated for first 1-hour of ERK response. First *minimums* marked with dots of same color of respective signal frequency data. (b) Minimal sufficient embedding dimension obtained using *false nearest neighbor* method.

3D phase-space trajectories are reconstructed from individual ERK time profiles using the obtained time-delays for each of the given signal frequency data. Reconstructed trajectories for the signals belonging to proliferation class i.e 3'3' and 3'60' are shown in Fig. 2.15 (a) with orange and blue trajectories respectively, where the mean trajectory is shown with thick curve and thin curves represents the individual

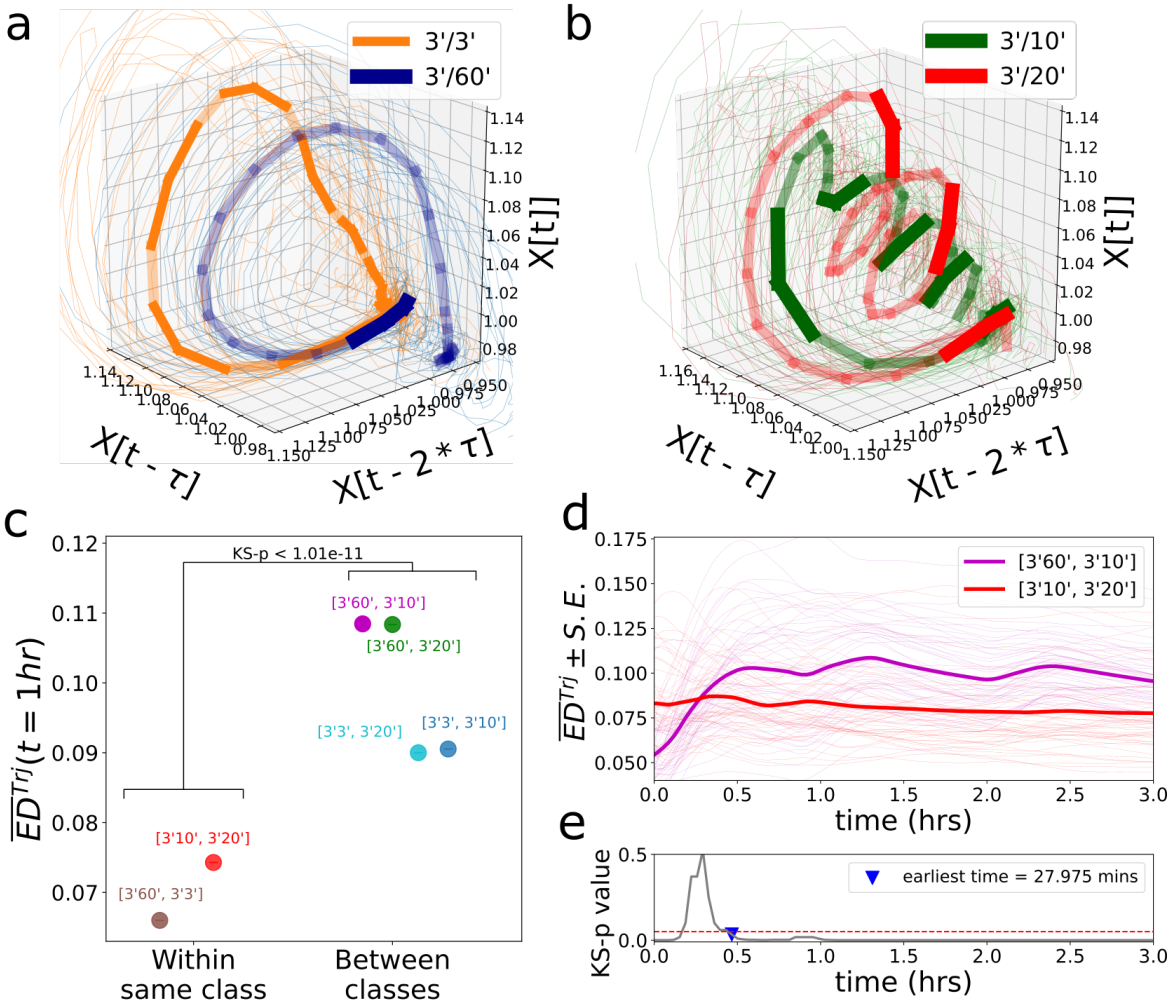

Figure 2.15: **Reconstructed trajectories from ERK experimental data resulting in Proliferation and differentiation for EGF stimulation.** (a, b) Reconstructed phase-space trajectories from experimental data for input frequencies resulting in proliferation and differentiation respectively. (c) Mean euclidean distance ($\overline{ED}^{Trj}_{N=3}$) in 3-D calculated in 1 hour for all combinations of signals belonging to same class i.e within proliferation and differentiation class and across these two classes. (d) Mean ED between 3'60' 3'10' (purple) and between 3'10' 3'20' (red) for reconstructed trajectories. (e) KS-p test over time between the two EDs in (d).

reconstructed trajectories. Similarly, reconstructed trajectories for the signals belonging to differentiation class i.e 3'10' (green) and 3'20' (red) are shown in Fig. 2.15 (b). Both the reconstructed trajectories of differentiation class signals have characteristic curves and loops in the phase-space as observed in the simulations. On the other hand, trajectories belonging to proliferation class are completely opposite and stays in the periphery of the volume covered by differentiation class trajectories. In order to quantify the separation between trajectories of same and different class of signals, we used the ED measure. Trajectories of same phenotype class are dynamically time warped to compare the shape of same trajectories and remove temporal dependencies. Mean ED is calculated with 48 different cell-combinations

over 1 hour of duration. Separation between reconstructed trajectories belonging to same class of signals: [3'3', 3'60'] i.e within proliferation class and [3'10', 3'20'] i.e within differentiation class, remains small i.e below 0.08 as shown in Fig. 2.15 (c). On the other hand, the mean ED remains above 0.09 for all the combination of signal trajectories belonging to different phenotypic classes. The ED remains small between reconstructed trajectories belonging to same class of signals while that of belonging to different phenotypic classes remain high. Separation of trajectories in both these cases are significantly different when tested with Kolmogorov-Smirnov (KS) test with p-value $< 10^{-11}$. Furthermore, we obtained the earliest time point of separation between trajectories belonging to same and different phenotypic classes. For that, ED over time is obtained between [3'60', 3'10'] shown with purple curve in Fig. 2.15 (d) and also between [3'10', 3'20'] shown with red. To obtain the earliest separation time between these two EDs, we carried out KS-p test which appears at $\sim$ 28 minutes, shown with blue triangle in Fig. 2.13 (e). The earliest time of separation between reconstructed trajectories also closely follows the observations made using stochastic simulations.

These results confirm the predictions that the phase-space trajectories generated by inputs resulting in same phenotypic responses are characteristically similar to each other, contrary to the distinct trajectories of different phenotypic class signals. All the trajectories generated by signals belonging to differentiation class have characteristic loops and curves covering a large volume of phase-space. On the the other hand, trajectories generated by signals belonging to proliferation class always remain characteristically on the periphery of the volume covered by differentiation class trajectories. These distinct features of proliferation and differentiation class trajectories are observed equivalently both in stochastic simulations and reconstructed trajectories from the experimental data. Aligned with the simulations, the quantification of separation between trajectories based on ED measure between the trajectories from different phenotypic classes always remain higher as compared to the ED between trajectories of same phenotypic class.

2.5.3 Differentiation and proliferation class trajectories stay separated in phase-space for NGF stimulation

Stimulation of PC12 cells with nerve growth factor (NGF) have been shown to generate sustained ERK activity which result in differentiation [79]. However, recent studies show that stimulating PC12 cells with different frequencies of NGF induces differentiation, as well as proliferation [78, 11]. This provides us with another example where different phenotypic responses are generated depending on temporal frequency of growth factor signal. Therefore, we tested our hypothesis by stochastic simulations and using publicly available experimental data for PC12 cells stimulation with different frequencies of NGF signals. Tropomyosin receptor kinase A (TrkA) is the high affinity receptor for NGF, which mediates neuronal differentiation, proliferation and programmed cell death. TrkA receptors have been found to remain on cell membrane for longer duration [31] as compared to EGFR receptors. To perform the simulations based on different temporal frequencies of NGF stimulus, we employ the same method used to model the decrease of receptor concentration on the plasma membrane for EGFR stimulation by dynamic EGF signals, as discussed in Section 2.5. The degradation of receptors from the cell membrane is modeled using a nonlinear *Tanh* function as represented by Eq. 2.14, however the rate of degradation of TrkA receptors from plasma membrane is decreased by $\sim 33\%$ with respect to that of EGFR case.

Using same noise levels as estimated before, we performed stochastic simulations on EGFR signaling network for different frequencies of input signals given in Table 1.1 for NGF case. Proliferation class contains only one signal with a temporal frequency of 3'60', while differentiation class contains three signals [3'3', 3'10', 3'20']. ERK time profiles for both the phenotype classes are given in Fig. 2.16 (a) and (c) respectively. Slow degradation rate of receptors from the plasma membrane of the cell results in longer and characteristically different ERK response profiles with respect to that of EGF stimulation case. 3 dimensional phase-space trajectories for both phenotype class signals are shown in Fig. 2.16 (b) and (d). Again, the trajectories in the proliferation class are completely different to those of differentiation class. The proliferation class trajectory remain characteristically on the periphery of the volume covered by differentiation class trajectories, similar to as observed in the case of EGF stimulation. The quantification of separation is obtained using the mean ED measure between dynamically time-warped trajectories. As shown in Fig. 2.16 (e), trajectories belonging to the differentiation class remain close to each other in 3D phase-space marked by a small ED between all possible combinations of signals falling within the same phenotype class. On the other hand, trajectories of

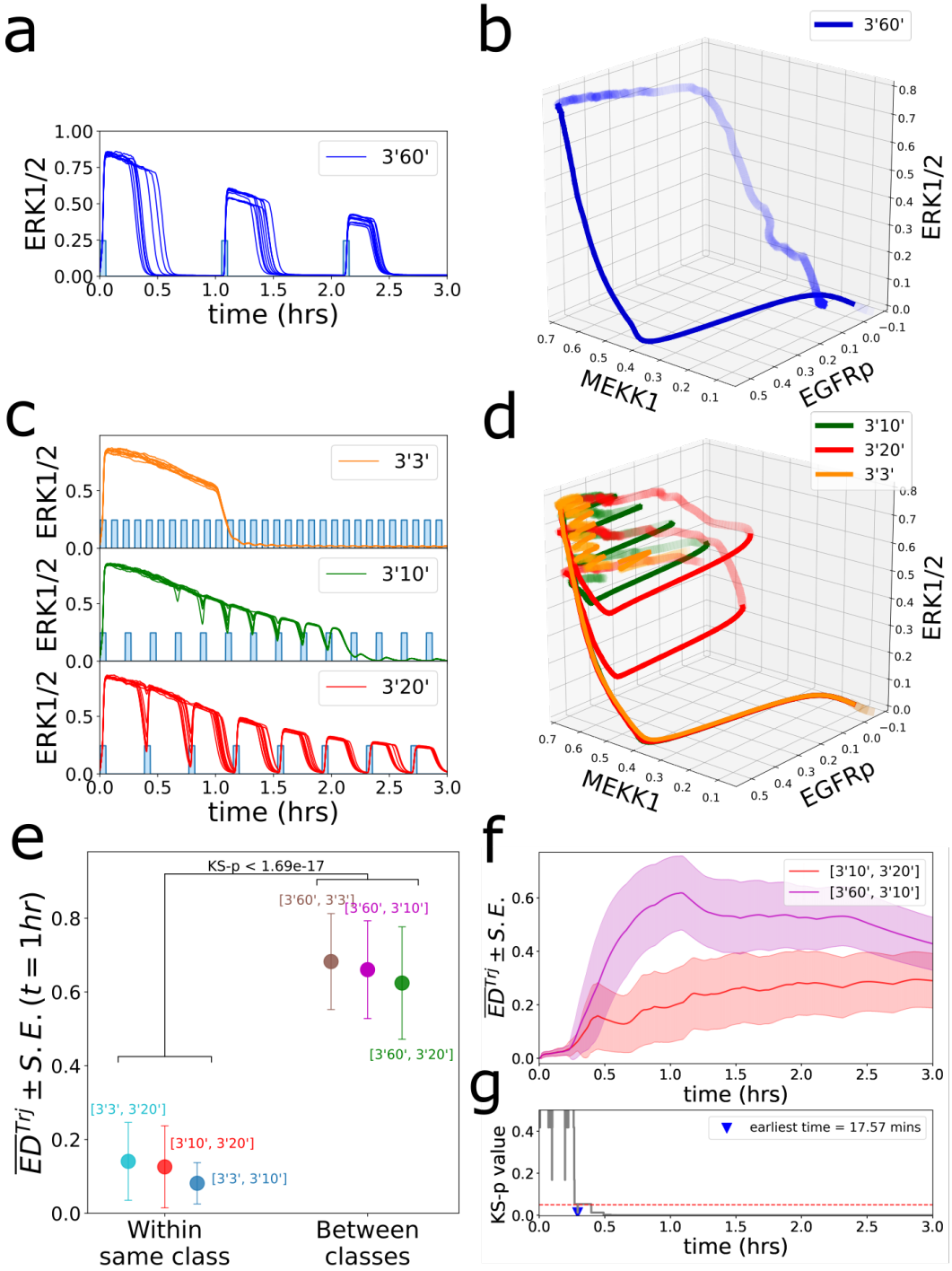

Figure 2.16: **Simulations with different frequencies of NGF stimulation corresponding to proliferation and differentiation.** ERK response temporal profiles and phase-space trajectories for input signals corresponding to (a, b) proliferation and (c, d) differentiation classes respectively. (e) Mean euclidean distance ($\overline{ED}^{Trj}_{N=3}$) in 3-D calculated in 1 hour for all combinations of signals belonging to same class i.e within proliferation and differentiation class and across these two classes. (f) Mean ED over time between trajectories of inputs belonging to same phenotypic class: differentiation - 3'10' and 3'20' (magenta) while ED between that of different phenotypic classes: 3'60' (proliferation) and 3'10' (differentiation) (red). (g) p-values over time of KS-test between the two EDs in (f).

opposite phenotype classes are maximally separated, quantified by large ED measure. Separation of trajectories in both cases are significantly different when tested with KS test with a p-value $< 10^{-17}$. Furthermore, comparison of ED measure over time between same phenotypic class trajectories [3'10, 3'20'] and between opposite phenotypic class trajectories [3'60, 3'10'], shows that trajectories start close to each other and separate quickly after $\sim$ 17.6 minutes, as shown in Fig. 2.16 (f).

These stochastic simulations mimicking NGF stimulation case based on experimental observations and estimated experimental noise demonstrate phenotypic class dependent encoding of extracellular information in phase-space trajectories of 'encoding' layer network. Slow degradation of receptors from the cell membrane resulted in longer memory duration and therefore external signals are integrated differently. This resulted in a different categorization of signals into differentiation and proliferation classes as compared to the phenotype classes in EGF stimulation case (Table 1.1). However, the features of each of these phenotype class trajectories remain same to what we observed in EGF stimulation case. Trajectories belonging to opposite phenotype class signals are characteristically distinct contrary to the same class trajectories which have similar features and curves in phase-space. The separation between trajectories is further quantified using ED measure, where maximal separation between different phenotypic class trajectories remain significantly larger than the small separation between same class trajectories. These observations are similar to what we found in both the simulations and reconstructed trajectories from experimental data in case of EGF stimulation.

2.5.3.1 Reconstructed trajectories follow phenotypic class specific features in phase-space for NGF stimulation

We next used the publicly available experimental data [78] of ERK response profiles for different frequencies of dynamic NGF signals. Generating the phase space trajectories from this NGF experimental data using the time-delay embedding method as previously. We used 50 ng/ml NGF stimulation data available for signals of temporal frequencies [3'3', 3'10', 3'20', 3'60']. Using the same procedure, first the average of each signal frequency denoised data is obtained for 30 cells. Calculation of mutual information gave time-delays of [4, 3, 3, 3], with an embedding dimension of 3 obtained using *false nearest neighbor* method.

3D phase-space trajectories are reconstructed from individual ERK time profiles using the obtained time-delays for each of the given signal frequency data. Reconstructed trajectory for the signal belonging to proliferation class i.e 3'60' is shown in Fig. 2.17 (a) with mean trajectory shown with thick curve and thin curves represent the individual reconstructed trajectories. Similarly, reconstructed trajectories for the signals belonging to differentiation class i.e 3'3' (orange), 3'10' (green) and 3'20' (red) are shown in Fig. 2.17 (b). All the reconstructed trajectories of differentiation class signals have characteristic curves and loops in phase-space as observed in simulations and also in the case of EGF stimulation. On the other hand, trajectory belonging to proliferation class is stays in the periphery of the volume covered by differentiation class trajectories as observed in the simulations as well as for the EGF stimulation case. This separation between same and different phenotype class trajectories is obtained using the mean ED measure of dynamically time-warped trajectories. Similar to the simulations, the separation is maximal between different phenotypic class trajectories which remain significantly larger ($p < 10^{-7}$, KS test) than the low separation between same class trajectories (Fig. 2.17 (c)). Furthermore, comparing the EDs over time of trajectories within the same and different phenotype classes (Fig. 2.17 (d)), reveals the earliest time-point of separation between the trajectories, which appears at $\sim$ 11 minutes (Fig. 2.17 (e)).

Reconstructed phase-space trajectories from experimental data of NGF stimulation show characteristic differences and quantified separation between those of differentiation and proliferation classes, these results closely follows all the observations made based on stochastic simulations. Furthermore, we observed consistencies not just with the NGF based simulations but also with the results obtained using stochastic simulations and reconstructed trajectories from EGF stimulation experimental data. First, same characteristic features of trajectories in 3D phase-space generated by signals of same phenotypic class. In case of trajectories generated by

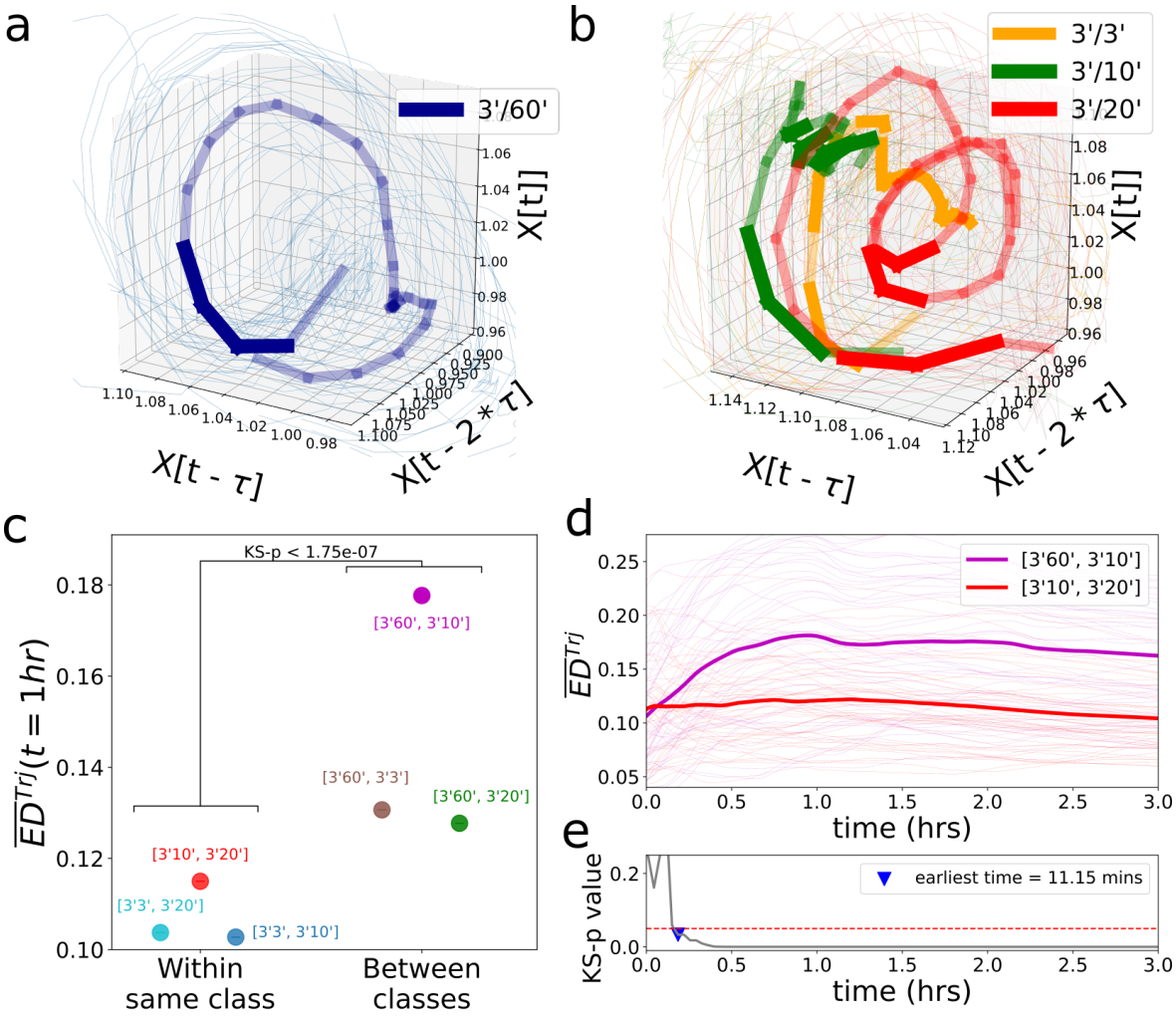

Figure 2.17: **Reconstructed trajectories from ERK experimental data resulting in Proliferation and differentiation for NGF stimulation.** (a, b) Reconstructed phase-space trajectories from experimental data for input frequencies resulting in proliferation and differentiation respectively. (c) Mean euclidean distance ($\overline{ED}^{Trj}_{N=3}$) in 3-D calculated in 1 hour for all combinations of signals belonging to same class i.e within proliferation and differentiation class and across these two classes. (d) Mean ED between 3'60' 3'10' (purple) and between 3'10' 3'20' (red) for reconstructed trajectories. (e) KS-p test over time between the two EDs in (d).

signals of differentiation class, trajectories are shown to have characteristic loops and curves covering a large volume of phase-space for both EGF and NGF stimulation. For proliferation class, trajectories remain characteristically on the periphery of the volume covered by differentiation class trajectories. Not just the characteristic observations of trajectory features, the quantification of separation between these trajectories are also consistent for both the growth-factors. The ED measure between same phenotypic class trajectories is found to be significantly lower than the separation between different class trajectories. This suggests that trajectories belonging

to the same phenotypic class elapses the same volume in 3D phase-space and remain close to each other, while they remain separated in different parts of phase-space with different characteristics when trajectories belong to opposite phenotype classes. This observation also remain consistent for both EGF and NGF stimulation. Therefore, the similarities in same phenotype class trajectories for EGF and NGF stimulation, suggests that the information of same phenotype is encoded similarly in the 'encoding' layer network for both the growth factors.

2.5.4 Conserved proliferation and differentiation class trajectories irrespective of growth factor identity

In the previous section, we used stochastic simulations and experimental data to reconstruct phase-space trajectories of differentiation and proliferation classes. Similarities are found in phase-space characteristics and in separation property between same phenotype class trajectories in both the cases of EGF and NGF stimulation. To verify these findings, we calculated the separation when the trajectories are generated by different growth-factors and categorize them accordingly if both the considered trajectories are generated by the same or different phenotypic class. We used this systematic method on stochastic simulations as well as on reconstructed trajectories from EGF and NGF stimulation experimental data.

Trajectories belonging to same phenotypic class can further be classified into two sub-categories: one belonging to proliferation class denoted by $[P^{EGF}, P^{NGF}]$ and the other, differentiation class denoted by $[D^{EGF}, D^{NGF}]$. In all the cases, we fixed the first trajectory to be that of EGF while the second one of NGF stimulation. Starting with simulations, the separation between same phenotype class trajectories remains small (≤ 0.3) for both the cases as shown in Fig. 2.18 (a) with filled circles. Then, all the trajectory combinations are taken when they belong to opposite phenotype classes and arranged into two categories: $[P^{EGF}, D^{NGF}]$ and $[D^{EGF}, P^{NGF}]$. The mean ED between different phenotype class trajectories remain high (> 0.4) shown with filled squares in Fig. 2.18 (a). The separation between different classes of trajectories remains significantly higher than those of same phenotypic classes, tested with KS test with $p < 10^{-47}$. Similarly, the separation between reconstructed trajectories from EGF and NGF stimulation experimental data is obtained for the two cases and subcategories using the same procedure as mentioned for simulations. The mean EDs between different classes of reconstructed trajectories remains significantly higher than those of same phenotypic classes with a p-value $< 10^{-21}$ for KS test between the two cases, shown in Fig. 2.18 (b).

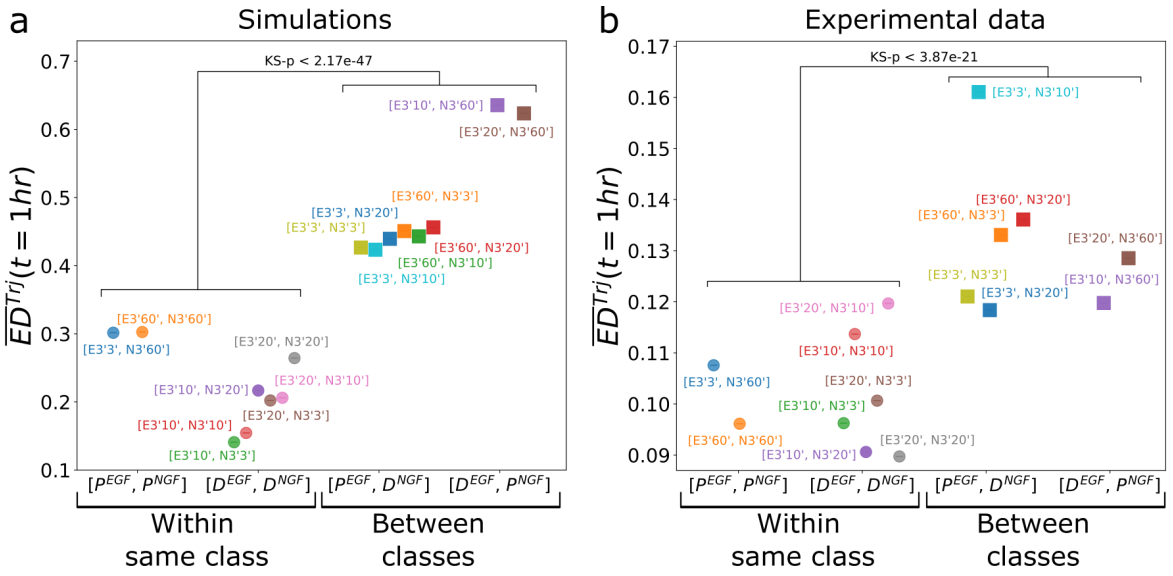

Figure 2.18: **Conserved proliferation and differentiation class trajectories and separation between them.** Mean euclidean distance ($\overline{ED}^{Trj}_{N=3}$) in 3-D calculated in 1 hour for all combinations of signals belonging to same class i.e within proliferation(P) and differentiation(D) class and between these two classes where first trajectory belonging to EGF stimulation while second that of NGF. EDs in (a) simulations and (b) reconstructed phase-space trajectories from experimental data.

This confirms that the extracellular information processed by different receptors for same phenotype class is encoded in similar phase-space trajectories. Same class of trajectories have similar characteristic features in phase-space and remain close to each other. On the other hand, opposite phenotype classes are encoded differently, with distinct characteristic features in phase-space and remain separated with each other, quantified by large ED measure. Furthermore, these characteristic features and quantification of separation between trajectories remain same even if generated by different growth-factors and respective receptors. This means that the information about same phenotype is processed similarly by different cell surface receptors and also encoded similarly in the 'encoding' layer network. This also proves the extracellular signal frequency dependent mapping of information to specific phenotype class trajectories irrespective of the growth-factor identity.

2.5.5 The information from the extracellular inputs is distributed throughout the encoding network

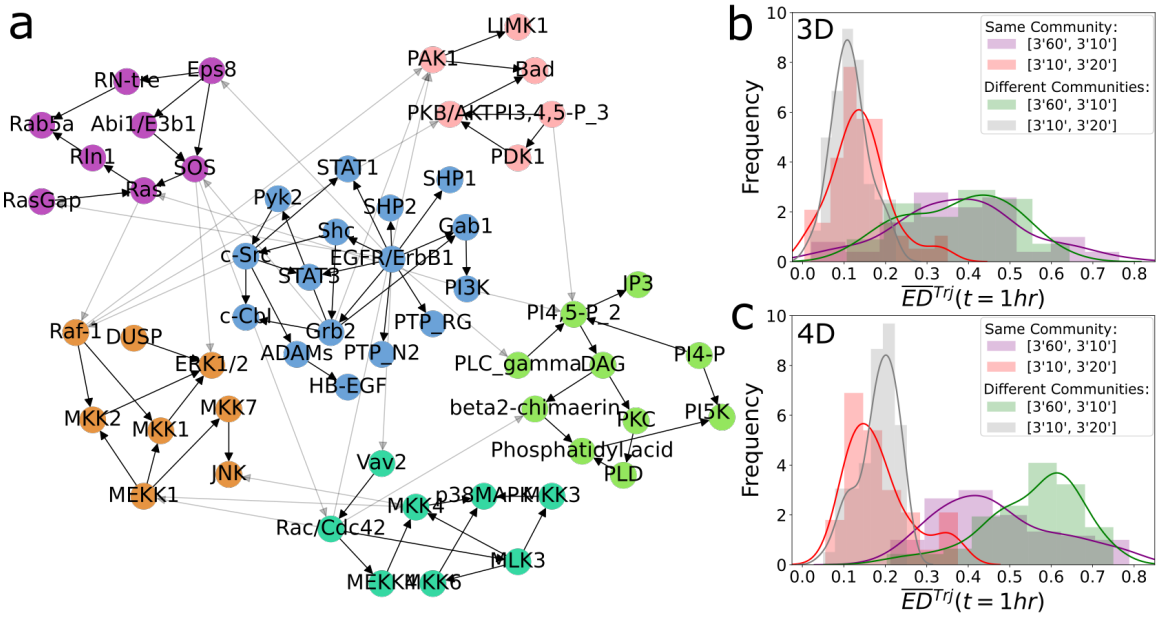

Figure 2.19: **Community structure of EGFR network.** (a) Rearranged EGFR network according to communities present within the network. ED calculation with different choices of nodes according to communities in (b) 3D and (c) 4D. Comparing EDs distribution when all the nodes are from same community (red and purple) and all from different communities (grey and green) respectively in 3D and 4D.

Using network structural analysis and *graph-theoretic* methods we found that EGFR signaling network is similar to Erdős-Rényi random networks. Extracellular information of different phenotypes is sensed by the cell surface receptors and therefore, distributed throughout the signaling network instead of passing via any intermediate bottleneck layer. The phenotypic information encoded in the phase-space trajectories of EGFR signaling network was obtained using the [$EGFR_p$, ERK, MEKK1] nodes, in the previous section. However, since the information is distributed throughout the signaling network and therefore, it can be retrieved from any node within the network. In order to investigate this, we next analyzed whether the community structure of the signaling network can be utilized to optimize the selection of nodes to extract the phenotypic information. In network science, a community is defined as a subset of nodes within the graph such that connections between the same community nodes are denser than connections with the rest of the network [32, 72, 73]. We inspected if the given number of nodes selected from different communities carry more phenotypic information than the nodes selected from the same community.

First, we rearranged the EGFR signaling network according to different communities present in the network using *greedy modularity community* algorithm. For that, we used the `Networkx` function of same name in `Python`. We identified that EGFR signaling network can be sorted into 6 different communities as shown in Fig. 2.19 (a), nodes falling under same community are shown with same color. According to different communities present in the network, we have multiple ways of selecting the nodes. We made the selection of nodes based on two categories: (A) all nodes from same community, (B) all nodes from different communities. First we selected only 3 nodes to obtain the ED separation for same phenotype class trajectories [3'10', 3'20'](red, grey) and for different phenotype classes [3'60', 3'10'](purple, green) as shown in Fig. 2.19 (b). When the nodes are selected from different communities there is negligible separation between same and different class trajectories as compared to nodes selection from same community. However, in 4D, this separation enhances when all the 4 nodes are selected from different communities of EGFR signaling network as shown in Fig. 2.19 (c). This suggests that different communities carry more phenotypic information than the same community. Therefore, the extraction of phenotypic information encoded in signaling network transients can be maximized if all the considered nodes belong to different communities of the network.

2.6 Feed-forward early response gene networks decode the information of the phenotypic states

We provided a general framework for the processing of time-varying extracellular signals at the 'input' layer level and encoding of this receptor-processed information in phenotype-specific trajectories by the 'encoding' layer network. In this section, we provide a basic framework explaining how these unique phenotype-specific trajectories in phase-space are interpreted and translated into corresponding phenotypic outputs. The first responders to the characteristic signaling network activation profiles are the immediate early genes, which are directly responsible for triggering the expression of the delayed early genes [81]. We propose that a minimal network structure that can decode the phenotypic information encoded in phase-space trajectories is the *feed-forward* network structure. In this way, we provide a general and complete framework for sensing, encoding and decoding of extracellular information by biochemical networks at 'input', 'encoding' and 'output' layers respectively.

2.6.1 Classifying extracellular dynamic signals as that of differentiation or proliferation using feed-forward network

Feed-forward structure of the 'output' network is the minimal and least complex network structure that is considered to resolve the information encoded in signaling network transients. In this section, we test the capabilities of a *feed-forward* network structure to decode the phenotypic information encoded in the phase-space trajectories of a signaling network. For that, we use a 30 node *feed-forward* 'output' layer network to resolve the different phase-space trajectories of EGFR signaling network corresponding to differentiation or proliferation, schematically shown in Fig. 2.20.

Multilayer *feed-forward* networks (FFN), by definition have several input nodes via which internal nodes of the network receives information and an output node. In this study, we used a 3-layered FFN with each layer having 10 nodes and 3 input nodes. The dynamics of the internal FFN nodes are defined by the rectified linear activation (ReLU) function given by Eq. 2.17. The output of this function linearly increases with positive input.

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0. \end{cases} \tag{2.17}$$

The full biochemical network consisting of EGFR signaling network with [ERK1/2, MKK1, Ras] (green nodes) as input for the 'output' layer (yellow nodes) is shown in Fig. 2.20 (a). The 'output' layer network is parameterized to generate an output of 1 for all those 3D phase-space trajectories corresponding to differentiation signals (3'10' and 3'20'), while 0 for that of proliferation signals (3'3' and 3'60'). As shown with red dots in Fig. 2.20 (b), both the trajectories belonging to differentiation signals (with inter-pulse interval of 10 and 20) are classified as that of differentiation while the trajectories corresponding to proliferation signals (with inter-pulse interval of 3 and 60) are classified with a low differentiation value (~ 0.1) meaning proliferation. Based on this parameterization, a range of EGFR signaling network trajectories generated by external signals of fixed 3 minute input pulses with an IPI $\in (3, 60)$ are classified as that of differentiation or proliferation shown with blue dots in Fig. 2.20 (b). It is evident that multiple trajectories are classified to result in differentiation, specially in the range with IPI $\in (8, 20)$.

Furthermore, in order to elucidate how the information contained in these phenotype-specific trajectories is decoded by the 'output' layer, we observed the responses of

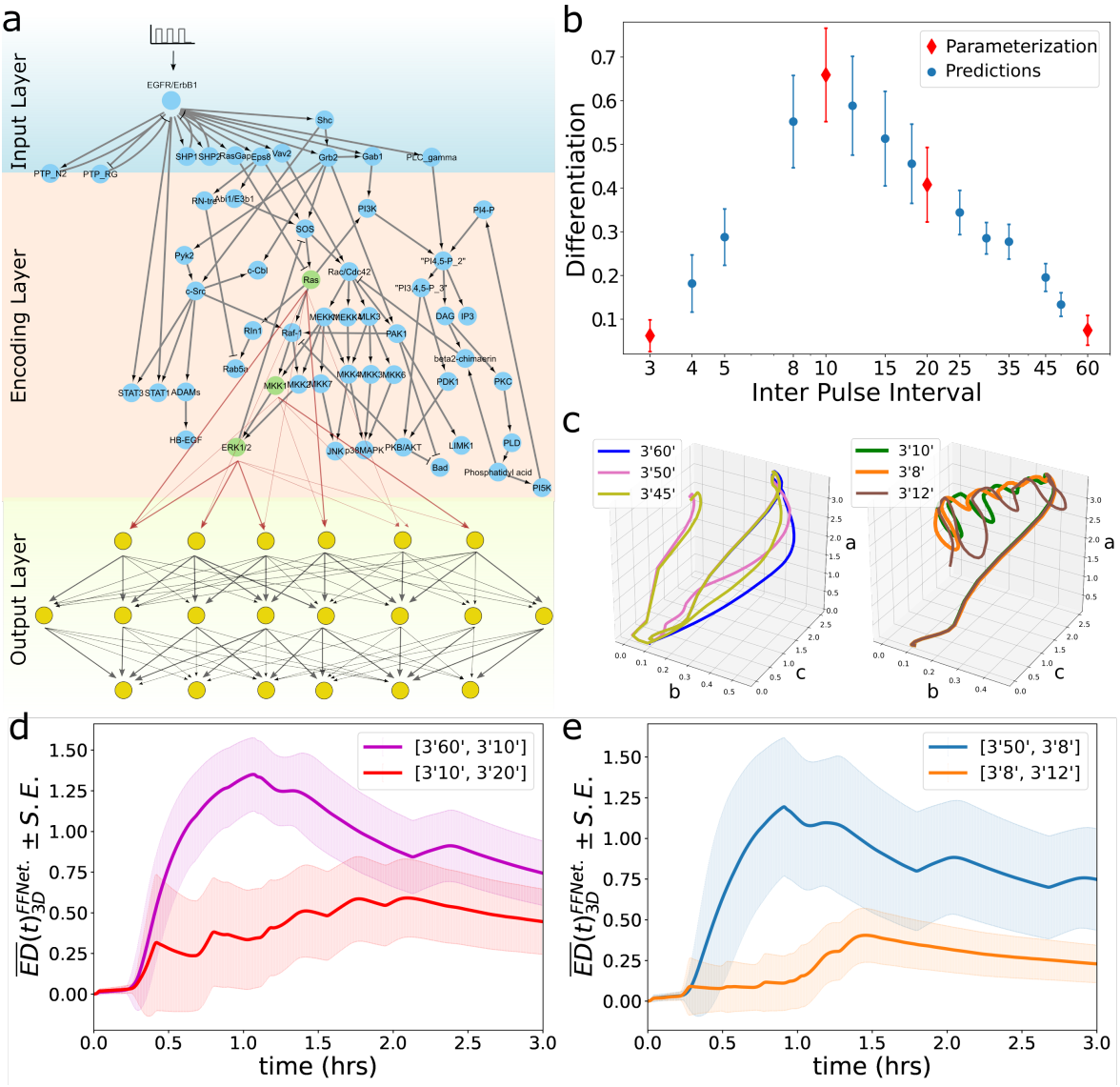

Figure 2.20: **Experimentally identified EGFR network with an arbitrary 'Output Layer' for prediction of differentiation and proliferation class dynamic input signals.** (a) Experimentally identified EGFR network with an arbitrary feedforward 'Output Layer' network (yellow nodes). (b) A 35 node feedforward network arranged in 4 layers (5-10-10-10) is used for predicting trajectories generated by 'Encoding Layer' nodes as differentiation percentage. All inputs have 3 minutes of pulse followed by mentioned IPI. Network is parameterized using trajectories of red colored input signals and blue colored ones are the predictions by 'Output Node' of feedforward network. (c) Phase-space trajectories of selected 'Output Layer' nodes for mentioned input signal frequencies of proliferation (left) and differentiation (right) classes. (d, e) $\overline{ED}^{FFNet.}$ over time between same (red, orange) and different (magenta, blue) phenotypic classes of trjectories generated by 'Output Layer' nodes [a, b, c] in panel (c).

three randomly selected FFN nodes namely $[a, b, c]$. Trajectory in the phase-space of 'output' layer nodes $[a, b, c]$ belonging to proliferation class signal 3'60' is shown with blue in Fig. 2.20 (c) (left). Two more exemplar trajectories of signals classified as that of proliferation, 3'50' (pink) and 3'45'(olive) are also plotted along-with.

This shows that just like the trajectories of signaling network, the decoded information is represented in the form of transients of the 'output' layer node dynamics. Decoded information belonging to proliferation is represented by distinct 'output' network trajectories, than those corresponding to differentiation signals [3'10', 3'8', 3'12'] shown in Fig. 2.20 (c) (right). It is also evident that the decoded information of same phenotype class is represented in similar 'output' layer node trajectories. The separation between trajectories of 'output' layer nodes $[a, b, c]$ for experimentally found signals of same and different phenotype classes is obtained using ED as shown in Fig. 2.20 (d). The 'output' layer trajectories of different class of signals [3'60', 3'10'] (magenta) are maximally separated with high ED while those of same class [3'10', 3'20'] (red) remain close-by which is reflected in small ED over the entire time. This is similar to the observations made for the signaling network trajectories when the information of different phenotypes are encoded by distinct phase-space trajectories of 'encoding' layer. Furthermore, the separation property of decoded phenotypic information is also present in the new signals classified as that of differentiation and proliferation shown in Fig. 2.20 (e), where [3'8', 3'12'] (orange) are classified result in differentiation and therefore remain together while [3'50', 3'8'] are classified to result in opposite phenotypes and therefore they are separated by large ED.

The *feed-forward* structured 'output' layer network decodes and accurately classifies all the experimentally determined time-varying signals according to the observed phenotypes. Furthermore, the 'output' layer classified signaling network trajectories of completely new dynamic external signals as that of proliferation or differentiation. Therefore, with the *feed-forward* structured 'output' layer network we predicted new external signals that will result in proliferation or differentiation. Probing into the decoding mechanism of FFN revealed that signaling network trajectories of same phenotype class generates similar responses trajectories in 'output' layer network.

2.6.2 Efficient decoding of phenotypic information by *feedforward* 'output' layer irrespective of node distribution

A three layered *feed-forward* network as 'output' layer accurately decoded the phenotypic information encoded in EGFR signaling network trajectories. The considered 'output' network had a total of 30 nodes divided equally in 3 internal layers. We next asked the question whether the node and layer distribution of the 'output' layer affect its information decoding features?

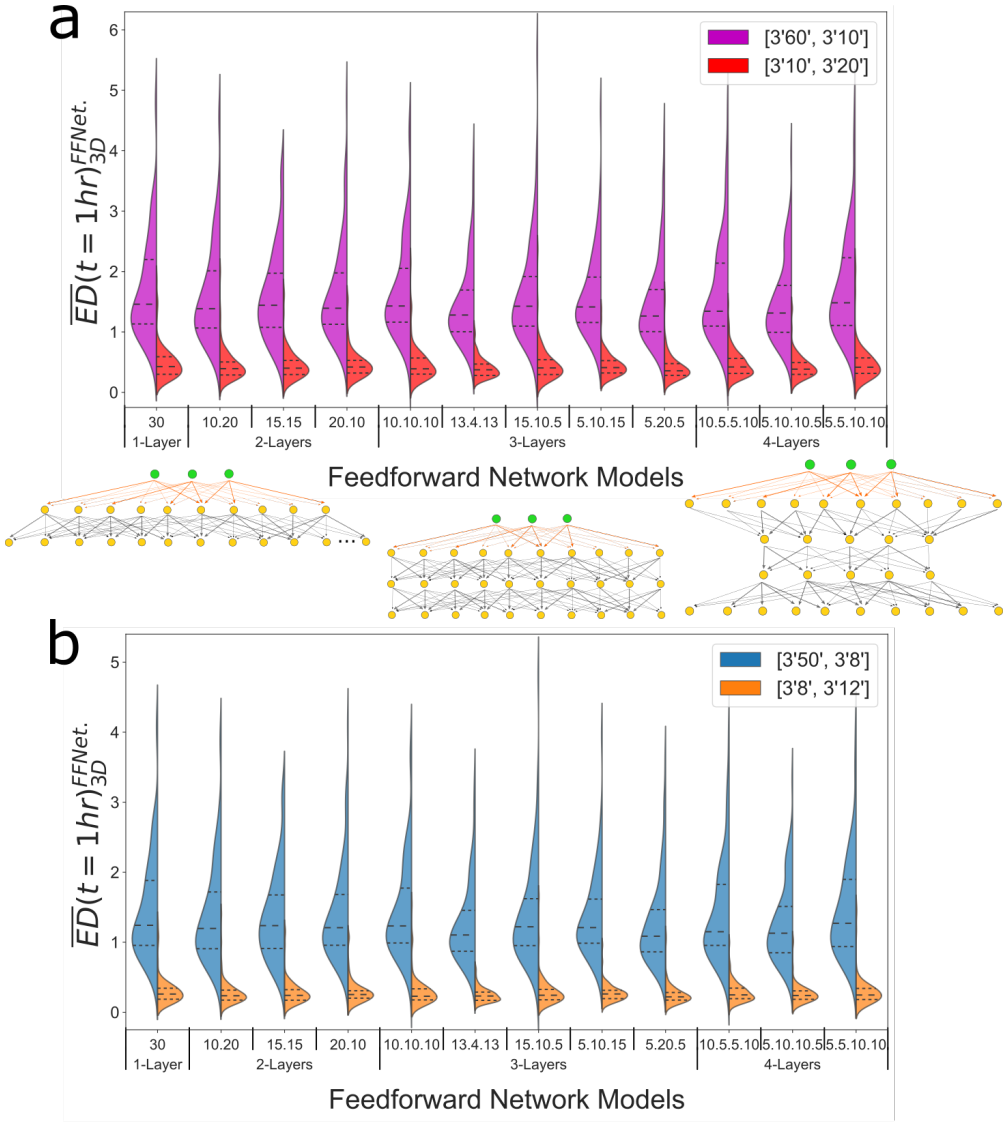

Figure 2.21: **Separation of 'Output Layer' node trajectories irrespective of nodes distribution in feedforward network structure of ERG.** (a, b) $\overline{ED}^{FFNet.}$ in 3D between same (red (a), orange (b)) and different (magenta (a), blue (b)) phenotypic classes of trajectories generated by 'Output Layer' nodes with respect to different distributions of 30 nodes in feedforward network structure of 'Output Layer'. 100 combinations of 3 randomly selected 'Output Layer' nodes are used for ED calculation.

We approached this problem from the perspective of resource-constraint, where we restricted only 30 nodes to be rearranged in different number of internal layers of 'output' network, and also by changing the ratio of nodes in these layers. This rearrangement of nodes into various layers can be categorized into 4 groups consisting of 1, 2, 3 and 4 internal layered 'output' networks. Using the same concept as defined for signaling networks, the response trajectories in phase-space of ERG will be separated for input signals generating different phenotypic responses. We previously observed that these 'output' layer trajectories remain close to each other when decoding information of same phenotypic response and separate for different phenotypic responses. We evaluated whether this feature of 'decoding' layer will arise independent of number of layers and distribution of nodes in the output network. For that, we selected 3 random nodes from each of the 'output' networks irrespective of any specific layer and calculated ED over 1 hour of response time, between the trajectories generating same phenotypic output (3'10' and 3'20') (red) and also between the trajectories generating different phenotypic outputs (3'60' and 3'10') (magenta), shown in Fig. 2.21 (a). Furthermore, we also calculated the EDs between EGFR signaling network trajectories of new signals classified as that of differentiation and proliferation in the previous section, shown in Fig. 2.21 (b).

It is evident from this analysis that the response trajectories of nodes in *feed-forward* structured 'output' layer network are also characteristically unique corresponding to those unique trajectories in signaling network. These trajectories are therefore, equivalently separated in the phase-space corresponding to different phenotypic responses and remain close to each other for same response, which we showed using euclidean distances. Therefore, the information decoding property and separation of trajectories is equivalently observed in all different arrangements of *feedforward* structured 'output' layer networks. This suggests that the number of internal layers in a *feedforward* structured GRN do not significantly contribute in the separation property.

2.6.3 Experimentally identified IEG-network decodes phenotypic information from EGFR signaling network trajectories

An artificial *feed-forward* network was employed in the previous section to decode the phenotypic information encoded in the EGFR signaling network trajectories, which gave us insight into the decoding mechanism of immediate early gene (IEG) networks. In this section, we used an experimentally identified IEG network to test if the realistic ERG networks also have a *feed-forward* structure. We will further verify that ERG networks employ the same decoding mechanism that we identified in the previous section.

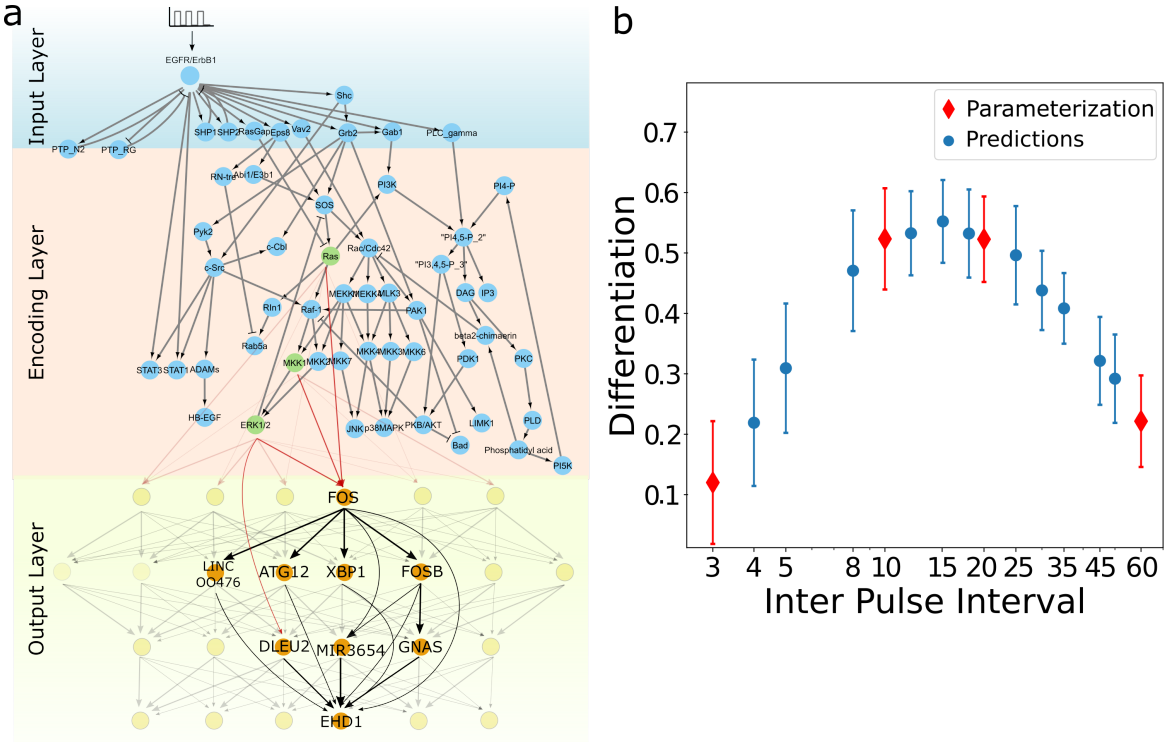

Figure 2.22: **EGFR network with experimentally identified IEG-network (MCF7_EGF1) as 'Output Layer'.** (a) Experimentally identified EGFR network with experimentally identified IEG-network (MCF7_EGF1) as 'Output Layer' (orange nodes) arranged in feedforward structure. (b) MCF7_EGF1 IEG-network is parameterized using red colored input signals to identify differentiation percentage using trajectories generated by 'Encoding Layer' nodes. Blue colored ones are the predictions by 'Output Node' of MCF7_EGF1 IEG-network.

We considerd MCF7_EGF1 network consisting of 9 immediate early genes: FOS, FOSB, XBP1, ATG12, LINC00476, DLEU2, MIR3654, GNAS and EHD1 [94]. IEGs are activated rapidly, sometimes within minutes after stimulus and transmits the information to late-response genes, eventually generating a phenotypic response. Furthermore, MCF7_EGF1 IEGs are shown to have a conserved activation network

with a *feedforward* structure, schematically shown in Fig. 2.22 (a). Therefore, we used MCF7_EGF1 as an 'output' layer of EGFR signaling network receiving inputs from [ERK1/2, MKK1, Ras]. As described in previous section, the 'output' layer network is parameterized to generate an output of 1 for all 3D phase-space trajectories corresponding to differentiation signals (3'10' and 3'20'), while 0 for that of proliferation signals (3'3' and 3'60'). MCF7_EGF1 is part of a larger gene regulator network, however, with only 9 nodes it properly classified the EGFR signaling network trajectories of experimentally determined proliferation and differentiation signals as shown with red dots in Fig. 2.22 (b). Furthermore, it was also able to classify phase-space trajectories of new time-varying signals as that of differentiation or proliferation shown with blue dots.

The classification closely follows the results obtained using generic *feed-forward* 'output' network in the previous section. Qualitatively, the classification curve shows that the EGFR signaling network trajectories generated by external signals of IPI $\in (3'8', 3'25')$ will undergo differentiation. Therefore, this systematic analysis and predictions with the experimentally identified *feed-forward* MCF7_EGF1 as 'output' layer network suggests that the phenotypic information encoded in the signaling network transients is likely decoded by minimal and sufficient *feed-forward* structured gene regulatory networks.

2.7 Signaling networks remain conserved at 'encoding' layer with a highly variable 'input' layer proteins across different cell types

In a diverse and changing extracellular environment, 'input' layer of biochemical networks plays an important role in sensing and processing different external signals present in the form of changing concentrations of growth-factors and cytokines. Therefore, cells operating in different extracellular environments must have specialized 'input' layers for sensing and processing these external signals specific to their environments, with a high cell to cell type variability. On the other hand, 'encoding' layer have a property of encoding the phenotypic information present in the dynamic extracellular environment into phase-space trajectories. In other words, 'encoding' network provides a high-dimensionality to encode the information received by the cell-surface receptors and transfer this information to GRN. This suggests that 'encoding' layer proteins act as a carrier of extracellular information between 'input' and 'output' layers, and therefore, they need not to acquire properties

specific to different extracellular environments. Therefore, 'encoding' layer proteins should be conserved across different cell types relative to 'input' and 'output' layer proteins. The 'output' layer, on the other hand, decodes the phenotype-specific information encoded in signaling transients. Furthermore, different cell types perform different functions by generating various phenotypic responses. Therefore 'output' layer likely have specialized protein species and transcription factors.

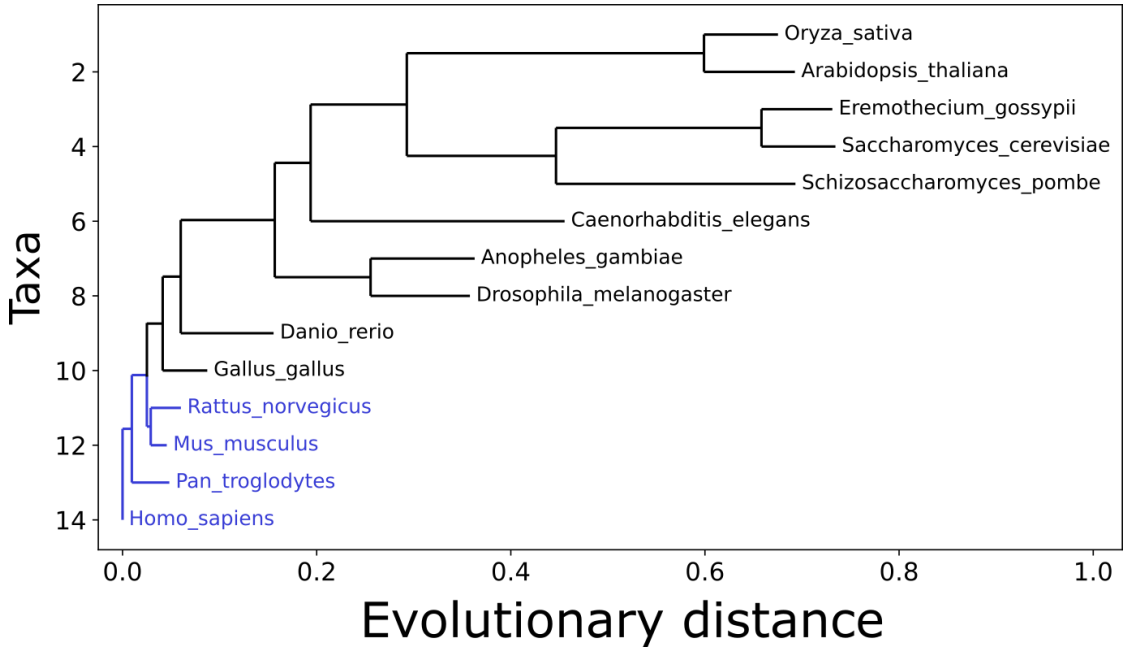

Figure 2.23: **Phylogenetic tree of 14 species considered for calculating conservation score of proteins.** Full phylogenetic tree of species used for calculating conservation score as a fraction of sum of branches in a sub-tree to sum of all branches in full tree, given by Eq. 2.18. An exemplar sub-tree for a protein conserved in given species is shown with blue. Adapted from [80].

We used the publicly available mass spectroscopy data of proteins for 11 different human cancer cell lines [88]. The cell types include A549 (lung carcinoma epithelial cells), GAMG (Glioblastoma cell line), HEK293 (exhibiting epithelial morphology), HeLa (cervical cancer cells), HepG2 (nontumorigenic cells with high proliferation rates), Jurkat (leukemia), K562 (lymphoblasts), LnCap (prostate adenocarcinoma), MCF7 (breast cancer cell line), RKO (colon carcinoma cell line) and U2OS (osteosarcoma). Homology information for 1532 proteins were obtained from NCBI database for the following 14 species: *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, *Gallus gallus*, *Anopheles gambiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Schizosaccharomyces pombe*, *Eremothecium gossypii*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Oryza sativa*. A phylogenetic tree of all these species was used to obtain a conservation score of each protein, full tree shown in Fig. 2.23.

This score reflects the evolutionary distance across the species in which the protein is conserved [80]. The conservation score (S_i) of a protein is defined as the ratio of sum of branch lengths in a pruned phylogenetic tree in which that protein is conserved to the sum of all the branch lengths in the full phylogenetic tree, given by Eq. 2.18:

$$S_i = \frac{\sum_{e \in E_i} w(e)}{\sum_{f \in E} w(f)} \tag{2.18}$$

For instance, a pruned phylogenetic tree is shown by blue color in Fig. 2.23 for a protein which is conserved across a few species. The sum of branch lengths in this pruned tree is 0.11527 and the sum of branch length of full phylogenetic tree is 2.33187, then the conservation score according to Eq. 2.18 is $S_i = 0.04943$. A protein which is conserved across all these 14 species will have a maximum conservation score of $S_i = 1$.

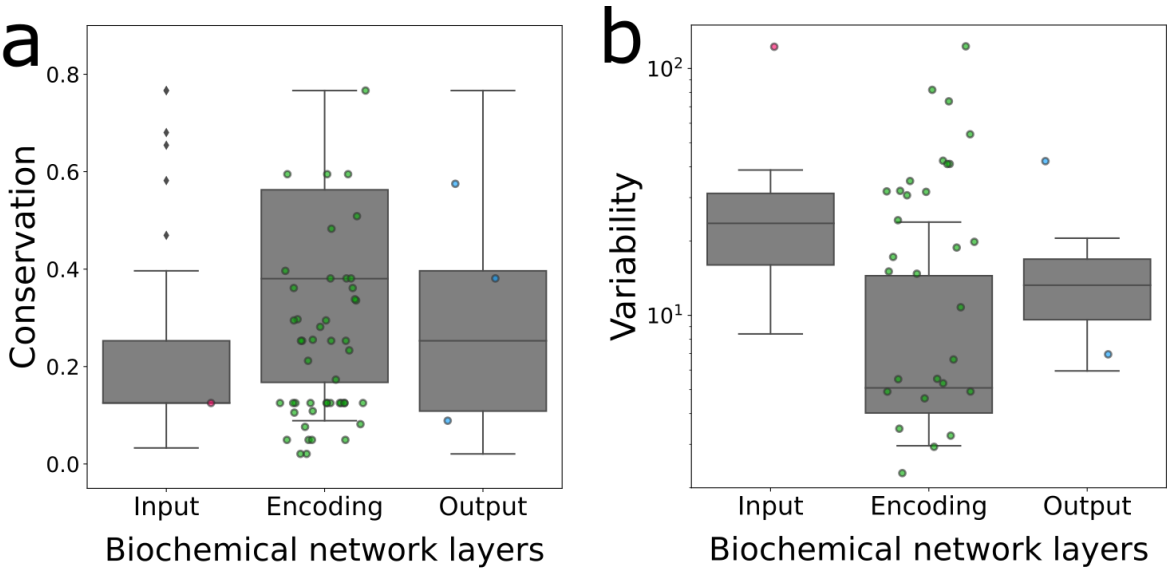

Figure 2.24: **Conservation and variability obtained for different biochemcial network layers across species.** (a) Conservation score and (b) variability of proteins relative to 'input', 'encoding' and 'output' layers of the biochemcial network across 11 different human cell lines. Conservation and variability of EGFR signaling network and MCF7_EGF1 GRN species are represented with colored circles.

Uniprot IDs are used to obtain Uniprot and Gene Ontology (GO) keywords associated with each protein from Uniprot Knowledgebase [22]. Uniprot keywords are used to categorize proteins into: membrane bound receptors, kinase, phosphatase,

GTPase and transcription factors (TF). GO keywords are used to identify adaptors. Only those proteins were considered that were unambiguously mapped to single Uniprot ID. All the proteins are further categorized into 'input', 'encoding' and 'output' layers such that : [Kinase, Phosphatase, Adaptor, GTPase] $\in$ 'encoding', membrane bound receptors in 'input' layer and TFs in 'output' layer. Using this categorization of proteins into different layers of biochemical network, we described the conservation properties of proteins in these layers. The obtained conservation scores for all the proteins are represented according to 'input', 'encoding' and 'output' layers in Fig. 2.24 (a). The 'encoding' layer have higher conservation score when compared to 'input' and 'output' layers. This suggests that the proteins comprising the 'encoding' layer remain more conserved across 14 different considered species relative to 'input' and 'output' layer proteins.

Variability between the cell types was estimated using the available label-free quantification (LFQ) of all the proteins. First, all the LFQ values were standardized by subtracting from each measurement the sample mean (over 3 repetitions of each cell type) and dividing by the sample standard deviation. Standardized LFQ values were used to obtain F-values by dividing the between cell variance by within cell variance. The mean of F-values across 11 different cell types gave abundance variability of each protein. The results show that the 'input' layer has high abundance variability than the 'encoding' and 'output' layers, shown in Fig. 2.24 (b). Whereas, the proteins comprising the 'encoding' layer have least abundance variation across 11 different cell types.

Conservation and abundance variability of EGFR signaling network along with MCF7 _EGF1 GRN proteins and TFs are also obtained using the aforementioned methods. Conservation and variability scores of EGFR signaling network is overlaid with rest of the protein data shown with colored cricles in Fig. 2.24 (a) and (b) respectively. Majority of the EGFR signaling network proteins are classified as that of 'encoding' layer, their high conservation score (green circles) aligns with rest of the 'encoding' layer proteins.

This analysis showed that the proteins comprising the 'input' layers which are responsible for sensing and processing time-varying extracellular information are less conserved with a high cell-to-cell type variability. This suggests that 'input' layer proteins have specialized sensing properties according to their organ/tissue specific environments. On the other hand, 'encoding' layer proteins encodes the phenotype-specific information in the phase-space trajectories, are evolutionarily conserved

and least variable across different cell types. This suggests that the 'encoding' network proteins have a more general property of information encoding, irrespective of cell type or any specific tissue. These observations are in agreement with the previous studies suggesting how different cell types recruit common networks to generate responses specific to cell type [67]. 'Output' layer proteins are found to be less conserved across different species with a slightly high cell-to-cell type variability as compared to that of 'encoding' layer network. This suggests their cell-type specific function of 'decoding' different information specific to various phenotypes from 'encoding' layer in real-time.

Chapter 3

Discussion

Cells operate in dynamic environments by continuously sensing, processing and responding to time-varying extracellular chemical signals. In order to accurately interpret the complex information from their environment, intracellular networks in single cells actively process these extracellular signals in real-time. The current concept explaining information processing is based on cellular dynamics near or at the steady-states which fails to explain cellular decision making in dynamic environments. Several recent studies have shown that cells respond to different frequencies of the growth factors and generate completely opposite phenotypic responses [78, 84]. The main objective of this thesis was to provide a novel framework for describing intra-cellular information processing mechanism that elucidates the processing of dynamic time-varying signals. The new framework described in this thesis, is based on experimentally identified properties of receptor and signaling networks providing a real-time transient based information processing in biochemical networks. Furthermore, the framework explains how working-memory arising at the 'input' layer level of biochemical networks enables the external signal frequency-dependent phenotypic responses of cells.

3.1 Utilizing community structure of signaling networks to control information propagation

Emergence of community structures are quite common in real complex networks [32], where individual communities play important role in maintaining the overall complexity of the entire network. However, different communities within the same network have differing properties, for instance, large social networks are found to

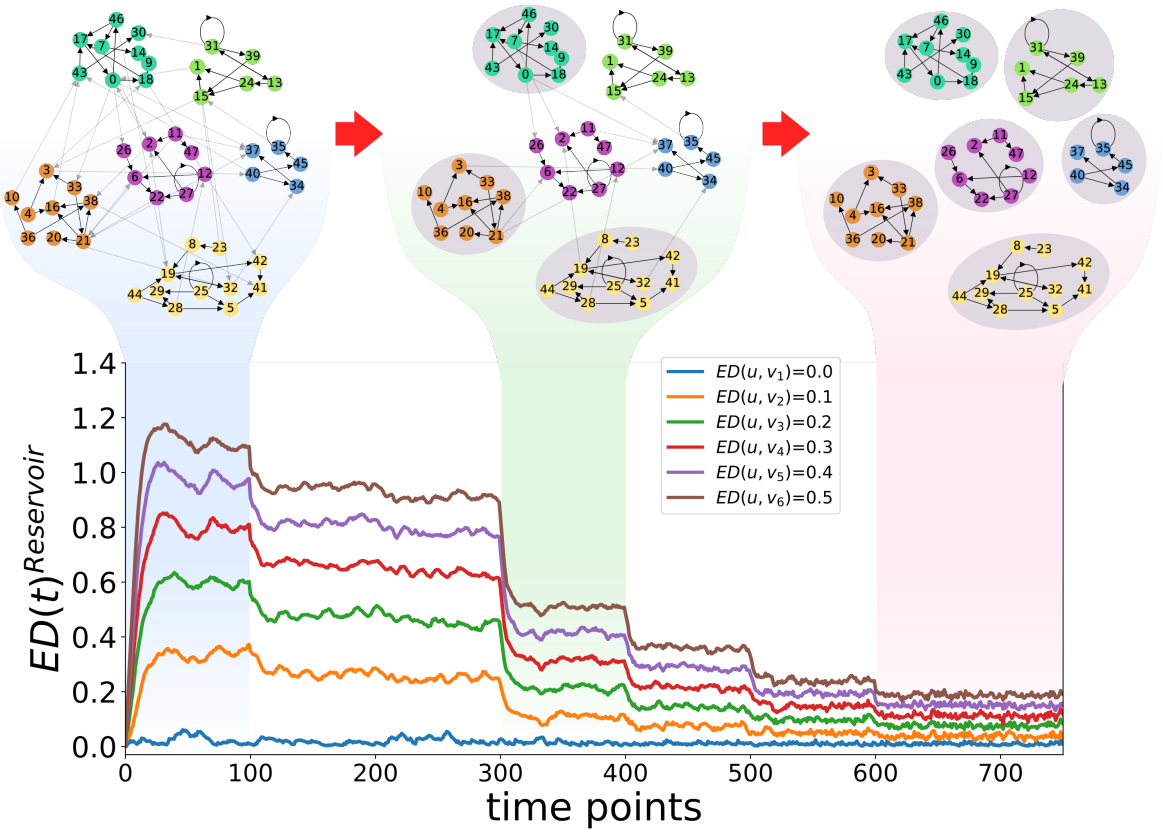

Figure 3.1: **Decrease in separation property of liquid state machines with isolated communities.** The reservoir or *liquid* is rearranged according to different communities present in the sparsely connected reservoir network of 50 nodes and connection density of 0.037. Same community nodes are clustered together and shown with same color. *(left to right)* Communities are isolated (marked with grey) by removing incoming links from other communities. *(right)* Only intra-community links (black) remain after all 6 communities are isolated from each other. The *separation property* of full *liquid* state is shown when the LSM is stimulated with an input v_i as a function of time t. *Separation property* is calculated with respect to the *liquid*-state of reference input u. Inputs v_i are generated to have differing distance ($ED(u, v_i)$) for a given reference input u. Starting with a fully connected reservoir at $t = 0$, one community is isolated after every 100^{th} time point.

have community groups based on various interests, occupation and common location [33]. Communities are often organized in a hierarchical way and are interdependent on each other for their structural organization, as the nodes are weakly connected between different communities as compared to dense connections within a same community. Furthermore, these communities play an important role in transfer of information from one end of the network to another, as they dependent on each other for exchange of information. Number of nodes and connectivity differ between different communities resulting in unique projection and processing of information within each community. For instance, we found the presence of six different communities in EGFR signaling network with the smallest community containing just 6 nodes and largest, 16 nodes (Fig. 2.19 (a)). We observed that the same

community nodes process and encode information in a similar fashion as compared to nodes belonging to different communities. The presence of unique information in different communities can be utilized to control the propagation of information in signaling networks by isolating these interdependent communities.

The recurrent structure of the reservoir/*liquid* of liquid-state-machines can be utilized to understand the changes in flow of information by deforming community structure of the reservoir. LSMs have an emergent property of *separation* of internal states of the *liquid* which enables the encoding of different inputs in different response profiles, discussed in Section 1.5.2. Therefore, to control the flow of information, one needs to control the separation of these internal states of the *liquid* by deforming the structural organization of the entire network. One way to do that is by randomly checking the effect of each link on *separation property* or it can also be achieved systematically, by exploiting the underlying community structure. The entire recurrent structure of the *liquid* can be distorted by removing the links between communities. The separation between the phase-space trajectories of different inputs decrease drastically when the communities are completely isolated from each other by removing the inter-community links, Fig. 3.1. This suggests that the information present in the *liquid* or a recurrent network can be suppressed by isolating the underlying communities. When the communities cannot share information with each other via inter-community links, the information processed and encoded by each community remains confined within itself.

Network structure based strategies to control the propagation of unwanted information in intra-cellular networks have gained a significant attention in recent years [55], but requires specific protein activity levels and kinetic parameter values for implementation. Such an unwanted flow of biochemical information within signaling networks can result in cancer which is an unchecked cell growth, as cells no longer respond to signals controlling the cellular growth and programmed cell death. The initiation, progression, invasion, and metastasis processes of cancer involve multiple molecular signaling mechanisms. One way to control the growth of cancer cells is the use of inhibitors that target specific components of these signaling networks. Methods involving the use of inhibitors are in clinical use, but the success of these agents has been limited by the resistance to inhibitor therapy that ultimately develops. Studies show that cancer cells or cell populations adapt or evolve in response to targeted therapies by rewiring molecular mechanisms to overcome the inhibitory effects of initial treatments. This rewiring involve alterations of signaling networks, such as addition or deletion of edges in the network, modification of reaction rates, and changes in molecular concentrations, all of which ultimately contribute to treatment resistance, either directly through rendering the drug ineffective or indirectly by leading to activation of alternative pro-survival or

anti-apoptotic mechanisms [100]. Cancer cells adapt their signaling circuitry as a counter-response to chronic drug treatment by taking advantage of redundancy of links in the signaling networks to maintain their functionality [39]. For instance, in breast cancer cells, EGFR activation can mediate resistance to ErbB2 antibody inhibitors and activation of ErbB2 signaling can bypass the block induced by EGFR antibody inhibitors [58].

The recurrent architecture of signaling networks provide a completely new perspective in addressing the problem of controlling the propagation of unwanted information. As described in this thesis, the recurrent structure of signaling networks have a property of information encoding in high-dimensional phase-space trajectories, similar to *separation property* of liquid-state-machines (Section 2.3). Therefore, community structures provide a completely new insight in the understanding of complex signaling networks. Systematic control of such underlying communities can prove to be helpful in controlling the overall information propagation in intra-cellular signaling networks. Specific communities can be isolated that contain proteins or transcription-factors directly responsible for activating onco-genes [43]. Furthermore, several studies has shown the presence of recurrent network architecture for information processing in large neuronal networks [54, 2]. Therefore, on a completely different scale of processing and propagation of neuronal information, the same principle can also be applied in large neuronal networks by identifying and controlling the underlying community structure for curating several neuronal diseases [23, 83, 101, 102].

3.2 Multiple receptors with dynamic memory activating the same downstream signaling network

Cells contain multiple receptor types that work simultaneously to carry out cellular functions by sensing and responding to the dynamic extracellular cues. Different receptor species are found to be associated with corresponding critical cellular processes, such as cell survival, proliferation, migration and differentiation. PC12 cells are the classic example of such a cellular behavior resulting in proliferation and differentiation upon respective activation of TrkA and EGF receptors [79]. In Humans, the family of 58 known RTKs fall into 20 subfamilies with similar molecular architecture: having a ligand binding domains in the extracellular region and a protein tyrosine kinase in the cytoplasmic region [56]. After the receptor activation on the cell surface level, the biochemical information reaches the common signaling network nodes which are present in the cytosol of the cell. The cells, therefore, employ

common downstream signaling networks to encode the dynamic extracellular information processed at the 'input' layer by the receptor-networks. The similarities and differences in the activity profiles of signaling network proteins for case of information arriving from multiple input streams relative to the case of stimulation by a single input needs a detailed and systematic study.

Changes in the information encoding property of the same downstream signaling network upon simultaneous activation by multiple cell surface receptor networks is largely unknown [44]. This problem can be analyzed by extending the formalism of biochemical networks as non-autonomous systems proposed in this thesis, for multiple receptor networks in the 'input' layer. Therefore, an extended framework addressing the simultaneous functioning of multiple receptors and activation of same downstream signaling components will provide us further insight into the information processing mechanism of simultaneous signals by different receptor types. Such a study will largely contribute in the further understanding of complex information processing mechanisms employed by cells operating continuously in dynamic environments.

3.3 Signaling between cells as non-autonomous systems

Cells are open systems that exchange energy and matter with the dynamic extracellular environment that are crucial for their survival and functionality. Therefore, in this thesis, we redefined the current understanding of intra-cellular biochemical networks as non-autonomous systems where trajectories in the high-dimensional phase-space are the solutions instead of steady-states (Section 1.6). However, cells do not operate individually rather they can send signals to each other and interpret the signals they receive from other cells directly or indirectly via the extra-cellular environment. Both animal and plants have cell junctions that directly connects the cytoplasms of adjacent cells where signaling molecules can pass freely between neighbouring cells. Moreover, animal cells communicate with their neighbouring cells where signaling molecules travel short distances. This type of local signaling, also kown as paracrine signaling, is especially important in embryonic development and the immune response [93]. In this case, signaling molecules are secreted by specific cells that travel only short distances and influence other cells in the local vicinity. For instance, growth factors are the compounds that stimulate nearby target cells to grow and divide. Numerous cells can simultaneously receive and respond to the growth factors produced locally by a single cell in their vicinity. This

type of signaling is spatially restricted where all the secreted signaling molecules are absorbed locally. This establishes a closed system of multiple interacting cells which generates and consumes the signaling molecules within themselves. By definition of closed systems, cells interacting in such a way can be dynamically considered as autonomous which are coupled locally (Section 1.6). However, recent experimental studies show that one cell type can alter the response of a different cell type via unidirectional interaction, thereby generating an open system [10]. For instance, during embryonic development germ cells are guided to the gonad by one or more environmental cues. In many vertebrates, the somatic cells attract migratory germ cells by secreting chemoattractant SDF-1, which binds to the CXCR4 GPCR in the migratory germ cell. This means that the response of one cell type to the dynamic environment can drive the response of another, locally present cell type in a non-autonomous manner.

In long-distance signaling which is also known as endocrine signaling, specialized cells release signaling molecules like hormones which travel via the circulatory system to different parts of the body, where they reach target cells that sense these signaling molecules and respond accordingly. Such a long-distance signaling constitutes the endocrine system comprising of internal glands that releases hormones directly into the circulatory system, regulating the distant target organs. Changing levels of hormones in the extra-cellular environment affects, in a non-autonomous way, the cellular functionality in various organs situated in different parts of the body. Immune system response is another example of such a long-distance intercellular communication. The primary task of immune system is to distinguish the antigens that belong to the body from those of foreign cells and to eliminate the invading pathogens. Eliminating pathogens typically involves several types of immune cells that coordinate their actions by communicating across wide spatial and temporal scales with myriad of cytokines [24]. To mount an efficient and accurate response, immune cells must coordinate their individual activation into global tissue-level responses: this requirement underscores the crucial relevance of cell-to-cell communication. Cytokines function by strongly binding to specific receptors on target cells and activating a cascade of downstream signalling events that culminate in the expression of a set of genes necessary for a specialized task [4]. For instance, consider a skin inflammation that occurs because of the skin being injured or invaded by pathogens [71]. Keratinocytes that reside in the epidermis sense the pathogen invasion and respond by secreting cytokines such as IL-1α. Macrophages, which are strategically positioned close to the blood vessels underneath the skin sense IL-1α and recruit other cells such as the dendritic cells, CD8+ T-cells, and neutrophils to the infected site. A short-range communication initially clusters nearby

immune cells to halt the spread of pathogens, and a long-range communication results in initially distant immune cells being recruited to the site of injury to eliminate the trapped pathogens. In this way, a complex inter-cellular network consisting of several different cell types constituting the innate-immune response is non-autonomously activated by an injury on skin or pathogens.

In another recent experimental study, it has been found that individual neutrophils exhibit extremely coordinated chemotaxis and cluster formation resembling of the swarming behaviour of insects [53]. A critical role for inter-cellular signal relay among neutrophils is mediated by the lipid leukotriene B4, which amplifies the local cell death signals to enhance the radius of highly directed interstitial neutrophil recruitment. In this newly formed environment, integrins, in concert with neutrophil-derived leukotriene B4 and other chemoattractants, promote local neutrophil interaction while forming a tight wound seal. This wound seal has borders that cease to grow in kinetic concert with late recruitment of monocytes and macrophages at the edge of the displaced collagen fibres. This reveals that the initial signals from a wound are propagated over large-range distances in a cascade of molecular events via unique inter-cellular communication signal between neutrophils in a non-autonomous manner, that allows rapid integrin-independent neutrophil recruitment through the tissue. Such a coordinated neutrophil-swarming behaviour isolates the wound or infectious site from surrounding viable tissue [53].

These are the experimentally identified examples where inter-cellular communication is established over short and long distances. The concentrations of hormones and chemoattractants has been experimentally found to change over space and time [74]. Whether they are cells responding to changing hormonal levels or the immune cells that communicate and recruit more cells for wound healing, the subsequent collective response of cells to external dynamic signals are examples of inter-cellular non-autonomous systems, where the cells not just sense and respond to external dynamic signals but they also communicate with each other simultaneously. Furthermore, network analysis suggests that the immune-cells are organized as the densest of complex networks yet studied and composed of hierarchically organized genes, proteins, and cellular components [82, 27]. All these suggests that the collective response of a complex network of immune cells to external dynamic signals also varies over space and time, which is consistent with both the experimental observations [53, 29] and theory of non-autonomous systems. This suggests that the solution of inter-cellular non-autonomous systems are the high-dimensional phase-space trajectories instead of steady-state solutions, similar to what we have discussed previously in Section 1.6 for intra-cellular networks. Further research is needed in this direction to develop a generalized theory and formalism of non-autonomous inter-cellular communication.

3.4 Specificity and generalization in neuronal networks

Neuronal networks have been extensively studied as computational systems, but they also serve as communications networks in transferring large amounts of information between different regions of brain. Once the external information is sensed at the sensory layer, the basic principles governing the subsequent flow and processing of this information by various neuronal networks is an active area of research. Recent work suggests that the structural organization of neuronal networks plays an important role in processing of different information in specific areas of the brain [54, 65, 15]. For instance, cerebral cortex is the largest site of neuronal information integration in the central nervous system responsible for cognition with a key role in attention, perception, memory, language and consciousness [93]. It has been found that the circuitry of cerebral cortex is dominated by local connections, and long-distance connections constitute only about 20% of total cortical connectivity. Furthermore, local networks conform to a canonical microcircuit that spans all cortical layers and includes recurrent excitation presumed to shape and amplify the sparse inputs from subcortical and distant cortical sources [25]. Studies involving graph theoretic methods used for investigating complex networks are providing clearer insight into the architectural organization of these networks. For example, in macaque brain, there exists a tightly integrated core circuit, spanning parts of premotor cortex, prefrontal cortex, temporal lobe, parietal lobe, thalamus, basal ganglia, cingulate cortex, insula, and visual cortex, that play a special role in higher cognition and consciousness [68]. Further studies showed that this central core of neurons has a very high density comprising of 92% nodes [2]. The nearby periphery nodes are further found to interact directly with the core neurons, in a layered structure generating an overall three layered network. The information from the periphery neurons goes to the highly interconnected and large core area for central processing and eventually reaches the output layer [54]. However, in-spite the large core network and a few neurons in the peripheral layers, such a structure is sometimes misunderstood as 'bow-tie' network where a few number of nodes in the central layer is the hallmark of such an architecture, as we discussed in Section 2.4.1. *Drosophila Melanogaster's* connectome is an another example of a large and randomly connected neuronal network. In a recent study of *Drosophila* connectome at the mesoscopic scale [83], a hierarchical structure has been found between the sensory modalities: olfactory, mechanoauditory, two-visual and the pre-motor center. The overall network structure is recurrent with interdependent sensory modules, exchanging the information locally processed within the respective subnetworks.

A recent theoretical and experimental work suggests that the real-time spatiotemporal information processing in neuronal networks emerges from the interaction between incoming stimuli and the internal dynamic state of neuronal networks [14]. These observation suggests that the information processing in large neuronal networks is similar to that in intra-cellular biochemical networks as we described in this thesis, but on a completely different scale. The information from the external dynamic environment is sensed by the receptors of the cell, constituting the input layer and the information is spread throughout the intra-cellular network, and decoded into different cellular phenotypes by the 'output' layer consisting of gene-regulatory-network. Similarly, in neuronal networks, the external information is sensed by the sensory layer which is further processed by specialized dense neuronal networks in different brain regions and as a result different voluntary or involuntary responses are generated. Similar to highly interconnected signaling networks, neuronal information is processed and encoded in the densely connected core layer of the neuronal network [14].

Furthermore, it has been shown that the cortical networks are inherently capable of processing complex spatiotemporal stimuli as a result of the interaction between external stimuli and the state of the internal network. Studies demonstrate that different stimuli elicit distinct spatiotemporal patterns of activity meaning external signal dependent different neuronal trajectories [62]. For instance, in an experimental study based on leech nervous system, stimulation by identical sensory inputs that elicit crawling and swimming behaviors are marked by completely different set of trajectories of neuronal activity separated in phase-space [13]. This suggests that cortical networks are inherently capable of processing complex spatiotemporal stimuli as a result of the interaction between external stimuli and the state of the internal network, which is in line with our conceptual framework of real-time information processing in biochemical networks based on signaling transients. Additionally, this also suggests that the sets of read-out neurons can extract information from the time-varying neural trajectories represented in high-dimensional space, meaning computations away from the steady-state dynamics of the network [63]. Therefore, the principles of information processing based on high-dimensional phase-space trajectories should remain same from intra-cellular networks to large neuronal networks which span different areas of nervous system. Subsequently, neuronal networks should be modeled in a non-autonomous way where external inputs drives the neuronal dynamics away from the steady-state organization. Analogous to phenotype-specific trajectories of non-autonomous intra-cellular networks, specific behaviors will be represented by characteristically different neuronal phase-space trajectories while a generalized response will be represented by characteristically similar trajectories.

Appendix A

A.1 EGFR receptor network equations

The equations of EGFR network corresponding to Fig. 2.10 (a), are adapted from Stanoev et al. [84]:

$$\begin{aligned}\frac{dEGFR_p}{dt} &= EGFR(\alpha_1 * EGFR + \alpha_2 EGFR_p + \alpha_3 * EGFEGFR_p) \\ &\quad - EGFR_p(\gamma_1 RG_a + \gamma_2 N2_a + \gamma_3 SHP1_a + \gamma_4 SHP2_a) \\ &\quad - k_{on} * 2 * (EGFR_p^2) * EGF + k_{off} * EGFEGFR_p \qquad \text{(A.1)} \\ \frac{dEGFEGFR_p}{dt} &= k_5 * EGFEGFR - \gamma_1 * RG_a * EGFEGFR_p - \gamma_2 * N2_a * EGFEGFR_p \\ &\quad + k_{on} * (EGFR_p^2 + EGFR^2) * EGF - k_{off} * EGFEGFR_p \qquad \text{(A.2)} \\ \frac{dEGFEGFR}{dt} &= -k_5 * EGFEGFR + \gamma_1 * RG_a * EGFEGFR_p + \gamma_2 * N2_a * \\ &\quad EGFEGFR_p - k_{off} * EGFEGFR \qquad \text{(A.3)} \\ \frac{dRG_a}{dt} &= k_1 * RG_i - k_2 * RG_a - b_1 * RG_a * (EGFR_p + 2 * EGFEGFR_p) \\ &\qquad \text{(A.4)} \\ \frac{dN2a}{dt} &= \epsilon * (k_1 * N2_i - k_2 * N2_a + b2 * (EGFR_p + 2 * EGFEGFR_p) * N2_i) \\ &\qquad \text{(A.5)} \\ \frac{dSHP1_a}{dt} &= k_{1S1} * SHP1_i - k_{2S1} * SHP1_a + b_{1S1} * SHP1_a * EGFR_p \qquad \text{(A.6)} \\ \frac{dSHP2_a}{dt} &= k_{1S2} * SHP2_i - k_{2S2} * SHP2_a + b_{1S2} * SHP2_a * EGFR_p \qquad \text{(A.7)}\end{aligned}$$

Where the parameters are as follows: $EGFR = EGFR_T - (EGFR_p + EGFEGFR + EGFEGFR_p)$. $k_d = 5.56, \alpha_1 = 0.001, \alpha_2 = 0.3, \alpha_3 = 0.7, \gamma_1 = 1.9, \gamma_2 = 0.1, \epsilon = 0.01, k_1 = 0.5, k_2 = 0.5, k_5 = 1.613, b_1 = 11, b_2 = 1.1, k_{on} = 0.001$ and $k_{off} = k_d * k_{on}$. $k_{1S1} = 0.1, k_{2S1} = 0.9, b_{1S1} = 1, k_{1S2} = 0.15, k_{2S2} = 0.85, b_{1S2} = 1$. $RG_i = RG_T - RG_a, N2_i = N2_T - N2_a, SHP1_i = SHP1_T - SHP1_a, SHP2_i = SHP2_T - SHP2_a$ with

$RG_T = 1, N2_T = 1, SHP1_T = 1$ and $SHP2_T = 1$.

A.2 Network parameters corresponding to arbitrary networks

The network parameters of the 15 different arbitrary biochemical networks generated in Section 2.3.2 are as follows:

Table A.1: Cascade and rewire method (left) and probabilistic method (right) network parameters.

Method 1	Parameters			
	P_b	$Prob(p)$	P_p	P_n
Net. 1	5/N	0.5	$1/N^2$	$1/N^2$
Net. 2	15/N	0.5	$10/N^2$	$1/N^2$
Net. 3	15/N	0.5	$10/N^2$	$8/N^2$
Net. 4	15/N	0.5	$5/N^2$	$2/N^2$
Net. 5	15/N	0.5	$5/N^2$	$2/N^2$
Net. 6	18/N	0.5	$5/N^2$	$2/N^2$
Net. 7	12/N	0.4	$5/N^2$	$2/N^2$
Net. 8	12/N	0.4	$8/N^2$	$2/N^2$

Method 2	Parameters	
	$Prob_p$	$Prob_n$
Net. 9	0.5/N	0.5/N
Net. 10	0.1/N	0.75/N
Net. 11	0.1/N	0.5/N
Net. 12	0.1/N	0.75/N
Net. 13	0.1/N	0.75/N
Net. 14	0.75/N	0.5/N
Net. 15	0.25/N	0.15/N

Bibliography

[1] Langa J A, Robinson J C, and Su´arez A. "Stability, instability, and bifurcation phenomena in nonautonomous differential equations." *Nonlinearity* 16 (2003), pp. 1277–1293.

[2] M. Ercsey-Ravasz et al. "A predictive network model of cerebral cortical connectivity based on a distance rule." *Neuron* 80 (2013), pp. 184–197.

[3] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of The Cell*. Vol. 5. Garland Science, 2008. Chap. 15.

[4] Grégoire Altan-Bonnet and Ratnadeep Mukherjee. "Cytokine-mediated communication: a quantitative appraisal of immune complexity." *Nature Reviews* 19 (2019), pp. 205–217.

[5] V. Anishchenko, T. Vadivasova, and G. Strelkova. "Stochastic self-sustained oscillations of nonautonomous systems." *Eur. Phys. J. Spec. Top.* 187 (2010), pp. 109–125.

[6] Reynolds AR, Tischer C, Verveer PJ, Rocks O, and Bastiaens PI. "EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation." *Nat Cell Biol* 5 (2003), pp. 447–453.

[7] Amir F. Atiya and Alexander G. Parlos. "New Results on Recurrent Network Training: Unifying the Algorithms and Accelerating Convergence". *IEEE TRANSACTIONS ON NEURAL NETWORKS* 11 (2000), pp. 697–709.

[8] Evren U. Azeloglu and Ravi Iyengar. "Signaling Networks: Information Flow, Computation, and Decision Making." *Cold Spring Harbor Perspectives in Biology* 7 (2015), a005934.

[9] Alan Baddeley. "Working Memory: Theories,Models, and Controversies". *Annu. Rev. Psychol.* 63 (2012), pp. 1–29.

[10] Lacy J Barton, Michelle G LeBlanc, and Ruth Lehmann. "Finding their way: themes in germ cell migration." *Current Opinion in Cell Biology* 42 (2016), pp. 128–137.

[11] Y Blum, J Mikelson, M Dobrzyński, H Ryu, M.A. Jacques, N.L Jeon, M.and Khammash, and O Pertz. "Temporal perturbation of ERK dynamics reveals network architecture of FGF2/MAPK signaling". *Mol Syst Biol.* 15 (2019), e8947.

[12] Cornelia Braicu, Mihail Buse, Constantin Busuioc, Rares Drula, Diana Gulei, Lajos Raduly, Alexandru Rusu, Alexandru Irimie, Atanas G. Atanasov, Ondrej Slaby, and Calin Ionescu andIoana Berindan-Neagoe. "A Comprehensive Review on MAPK: A Promising Therapeutic Target in Cancer." *Cancers* 11 (2019), p. 1618.

[13] K. L. Briggman, H. D. I. Abarbanel, and W. B. Kristan Jr. "Optical Imaging of Neuronal Populations During Decision-Making." *Science* 307 (2005), pp. 896–901.

[14] Dean V. Buonomano and Wolfgang Maass. "State-dependent computations: spatiotemporal processing in cortical networks." *Nature Reviews, Neuroscience* 10 (2009), pp. 113–125.

[15] Timothy J. Buschman. "Balancing Flexibility and Interference in Working Memory." *Annu. Rev. Vis. Sci.* 7 (2021), pp. 367–388.

[16] Tischer C and Bastiaens PI. "Lateral phosphorylation propagation: an aspect of feedback signalling?" *Nat Rev Mol Cell Biol* 4 (2003), pp. 971–974.

[17] H. B. Callen. *Thermodynamics and an introduction to thermostatistics*. Vol. 2. Wiley, 2014.

[18] Etienne Caron, Samik Ghosh, Yukiko Matsuoka, Dariel Ashton-Beaucage, Marc Therrien, Se´bastien Lemieux, Claude Perreault, Philippe P Roux, and Hiroaki Kitano. "A comprehensive map of the mTOR signaling network". *Molecular Systems Biology* 6 (2010), 6:453.

[19] Alexandre N. Carvalho, Jos´e A. Langa, and James C. Robinson. *Attractors for infinite-dimensional non-autonomous dynamical systems*. Vol. 182. 2013.

[20] JY Chen, JR Lin, KA Cimprich, and T. Meyer. "A two-dimensional ERK-AKT signaling code for an NGF-triggered cell-fate decision". *Mol. Cell* 45 (2012), pp. 196–209.

[21] P. Clemson, B. Petkoski, T. Stankovski, and A. Stefanovska. *Coupled nonautonomous oscillators, in Nonautonomous Dynamical Systems in the Life Sciences.* Vol. 2102. 2014.

[22] The UniProt Consortium. "UniProt: the universal protein knowledgebase in 2021." *Nucleic Acids Research* 49 (8 January 2021), pp. D480–D489.

[23] Nicolas A. Crossley, Andrea Mechelli, Petra E. Vértes, Toby T. Winton-Brown, Ameera X. Patel, Cedric E. Ginestet, Philip McGuire, and Edward T. Bullmore. "Cognitive relevance of the community structure of the human brain functional coactivation network." *PNAS* 110 (2013), p. 28.

[24] Hirad Daneshpour and Hyun Youk. "Modeling cell–cell communication for immune systems across space and time." *Current Opinion in Systems Biology* 18 (2019), pp. 44–52.

[25] R. J. Douglas, C. Koch, M. Mahowald, K. A. Martin, and H. H. Suarez. "Recurrent excitation in neocortical circuits." *Science* 269 (1995), pp. 981–985.

[26] Bard Ermentrout. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students.* SIAM, 2002.

[27] Ziv Frankenstein, Uri Alon, and Irun R Cohen. "The immune-body cytokine network defines a social architecture of cell interactions." *Biology Direct* 1:32 (2006).

[28] A. M. Fraser and H. L. Swinney. "Independent coordinates for strange attractors from mutual information". *PHYSICAL REVIEW A* 33 (1986), p. 1134.

[29] Peter Friedl, Joseph Locker, Erik Sahai, and Jeffrey E. Segall. "Classifying collective cancer cell invasion." *Nature Cell Biology* 14:8 (2012), pp. 777–783.

[30] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. "Construction of a genetic toggle switch in Escherichia coli." *Letters to nature* 403 (2000), pp. 339–342.

[31] Thangiah Geetha and Marie W. Wooten. "TrkA Receptor Endolysosomal Degradation is Both Ubiquitin and Proteasome Dependent". *Traffic* 9 (2008), pp. 1146–1156.

[32] M. Girvan and M. E. J. Newman. "Community structure in social and biological networks." *PNAS* 99 (2002), pp. 7821–7826.

[33] Mark Goldberg, Stephen Kelley, Malik Magdon-Ismail, Konstantin Mertsalov, and Al Wallace. "Finding Overlapping Communities in Social Networks." *IEEE* 11570026 (2010).

[34] I. Gyori, S. Michelson, and J. Leith. "Time-dependent subpopulation induction in heterogeneous tumors." *Bull. Math. Biol.* 50(6) (1988), pp. 681–696.

[35] Sepp Hochreiter and Jurgen Schmidhuber. "Long Short-Term Memory". *Neural Computation* 9 (1997), pp. 1735–1780.

[36] Weis W. I. and Kobilka B. K. "The molecular basis of g protein-coupled receptor activation." *Annual review of biochemistry* 87 (2018), pp. 897–919.

[37] Naoyuki Inagakil, Hans Thoenen, and Dan Lindholm. "TrkA Tyrosine Residues Involved in NGF-induced Neurite Outgrowth of PC12 Cells." *European Journal of Neuroscience* 7 (1995), pp. 1125–1133.

[38] Schlessinger J. "Ligand-induced, receptor-mediated dimerization and activation of EGF receptor." *Cell* 110 (2002), pp. 669–672.

[39] Gillies R. J. and Verduzco D.and Gatenby R. A. "Evolutionary dynamics of carcinogenesis and why targeted therapy does not work." *Nat. Rev. Cancer* 12 (2012), pp. 487–493.

[40] Tyson J., Chen K., and Novak B. "Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell." *Current Opinion in Cell Biology* 15 (2003), pp. 221–231.

[41] J.Baranyi, T.A.Roberts, and P.McClure. "A non-autonomous differential equation to modelbacterial growth." *Food Microbiology* 10 (1993), pp. 43–59.

[42] H. Jaeger. "Short term memory in echo state networks". *German National Research Center for Information Technology* GMD Report 152 (2001).

[43] Jr James E. Darnell. "TRANSCRIPTION FACTORS AS TARGETS FOR CANCER THERAPY." *Nature* 2 (2002), pp. 740–749.

[44] Kevin A. Janes, John G. Albeck, Suzanne Gaudet, Peter K. Sorger, Douglas A. Lauffenburger, and Michael B. Yaffe. "A Systems Model of Signaling Identifies a Molecular Basis Set for Cytokine-Induced Apoptosis." *Science* 310 (2005), pp. 1646–1653.

[45] A Jimenez, J Cotterell, A Munteanu, and J Sharpe. "A spectrum of modularity in multi-functional gene circuits". *Mol Syst Biol* 13 (2017), p. 925.

[46] Mouw J. K., Ou G. Q., and Weaver V. M. "Extracellular matrix assembly: a multiscale deconstruction." *Nat. Rev. Mol. Cell Biol.* 15 (2014), pp. 771–785.

[47] Urbanowicz K. and J.A. Hołyst. "Noise-level estimation of time series using coarse-grained entropy". *Physical Review. E Statistical Nonlinear and Soft Matter Physics* 67 (2003), p. 046218.

[48] M. B. Kennel, R. Brown, and H. D. I. Abarbanel. "Determining embedding dimension for phase-space reconstruction using a geometrical construction". *PHYSICAL REVIEW A* 45 (1992), pp. 3403–3411.

[49] Boris N. Kholodenko. "Cell-signalling dynamics in time and space." *MOLECULAR CELL BIOLOGY* 7 (2006), pp. 165–176.

[50] Peter E. Kloeden and Martin Rasmussen. *Nonautonomous Dynamical Systems*. Vol. 176. 2011.

[51] Aneta Koseska and Philippe IH Bastiaens. "Cell signaling as a cognitive process". *The EMBO Journal* 36 (2017), pp. 568–582.

[52] Aneta Koseska and Bastiaens P. I. H. "Processing Temporal Growth Factor Patterns by an Epidermal Growth Factor Receptor Network Dynamically Established in Space". *Annual Review of Cell and Developmental Biology* 36 (2020), pp. 359–383.

[53] Tim Laemmermann, Philippe V. Afonso, Bastian R. Angermann, Ji Ming Wang, Wolfgang Kastenmueller, Carole A. Parent, and Ronald N. Germain. "Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo." *Nature* 498 (2013), pp. 371–375.

[54] Simon B. Laughlin and Terrence J. Sejnowski. "Communication in Neuronal Networks." *Science* 301 (2003), pp. 1870–1874.

[55] Daewon Lee and Kwang-HyunCho. "Signal flow control of complex signaling networks." *Scientific Reports* 9 (2019), p. 14289.

[56] Mark A. Lemmon and Joseph Schlessinger. "Cell Signaling by Receptor Tyrosine Kinases". *Cell* 141 (2010), pp. 1117–1134.

[57] Edwin Li and Kalina Hristova. "Role of Receptor Tyrosine Kinase Transmembrane Domains in Cell Signaling and Human Pathologies." *American Chemical Society, Biochemistry* 45 (2006), p. 20.

[58] Jeremy S. Logue and Deborah K. Morrison. "Complexity in the signaling network: insights from the use of targeted inhibitors in cancer therapy." *GENES DEVELOPMENT* 26 (2012), pp. 641–650.

[59] Maxime Lucas, Duccio Fanelli, and Aneta Stefanovska. "Nonautonomous driving induces stability in network of identical oscillators." *Phys. Rev. E* 99 (2019), p. 012309.

[60] Baumdick M, Gelleri M, Uttamapinant C, Beranek V, Chin JW, and Bastiaens PIH. "A conformational sensor based on genetic code expansion reveals an autocatalytic component in EGFR activation." *Nat Commun* 9 (2018), : 3847.

[61] Baumdick M, Bruggemann Y, Schmick M, Xouri G, Sabet O, Davis L, Chin JW, and Bastiaens PI. "EGF-dependent re-routing of vesicular recycling switches spontaneous phosphorylation suppression to EGFR signaling." *Elife* 4 (2015), : e12223.

[62] Broome B. M., Jayaraman V., and Laurent G. "Encoding and decoding of overlapping odor sequences." *Neuron* 51 (2006), pp. 467–482.

[63] Rabinovich M., Huerta R., and Laurent G. "Transient dynamics for neural processing." *Science* 321 (2008), pp. 48–50.

[64] W. Maass, T. Natschlaeger, and H. Markram. "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations". *Neural Computation* 14(11) (2002), :2531–60.

[65] Nikola T. Markov, Mária Ercsey-Ravasz, David C. Van Essen, Kenneth Knoblauch, Zoltán Toroczkai, and Henry Kennedy. "Cortical High-Density Counterstream Architectures." *Science* 342 (2013), 1238406:1–13.

[66] Norbert Marwan, M. Carmen Romano, Marco Thiel, and Jürgen Kurths. "Recurrence plots for the analysis of complex systems". *Physics Reportss* 438 (2007), pp. 237–329.

[67] Kathryn Miller-Jensen, Kevin A. Janes, Joan S. Brugge, and Douglas A. Lauffenburger. "Common effector processing mediates cell-specific responses to stimuli." *Nature Letters* 448 (2007), pp. 604–609.

[68] Dharmendra S. Modhaa and Raghavendra Singh. "Network architecture of the long-distance pathways in the macaque brain." *PNAS* 107 (2010), pp. 13485–13490.

[69] Takanori Moriki, Hiroko Maruyama, and Ichi N. Maruyama. "Activation of Preformed EGF Receptor Dimers by Ligand-induced Rotation of the Transmembrane Domain." *J. Mol. Biol.* 311 (2001), pp. 1011–1026.

[70] Akhilesh Nandan, Abhishek Das, Robert Lott, and Aneta Koseska. "Cells use molecular working memory to navigate inchanging chemoattractant fields". *eLife* 11:e76825 (2022).

[71] Y. Natsuaki, G. Egawa, S. Nakamizo, S. Ono, S. Hanakawa, T. Okada, N. Kusuba, A. Otsuka, A. Kitoh, T. Honda, S. Nakajima, S. Tsuchiya, Y. Sugimoto, K. J. Ishii, H. Tsutsui, H. Yagita, Y. Iwakura, M. Kubo, L. Ng, T. Hashimoto, J. Fuentes, E. G. Yassky, Y. Miyachi, and K. Kabashima. "Perivascular leukocyte clusters are essential for efficient activation of effector T cells in the skin." *Nature Immunology* 15 (2014), pp. 1064–1069.

[72] M. E. J. Newman. "Detecting community structure in networks." *THE EUROPEAN PHYSICAL JOURNAL B* 38 (2004), pp. 321–330.

[73] M. E. J. Newman. *Networks: An Introduction.* Oxford University Press, 2010.

[74] Philipp Niethammer, Clemens Grabher, A. Thomas Look, and Timothy J. Mitchison. "A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish." *Nature Letters* 459 (2009), pp. 996–999.

[75] E O'Neill and W Kolch. "Conferring specificity on the ubiquitous Raf/MEK signalling pathway." *British Journal of Cancer* 90 (2004), pp. 283–288.

[76] K. Oda and H. Kitano. "A comprehensive map of the toll-like receptor signaling network". *Molecular Systems Biology* 2 (2006), p. 2006.0015.

[77] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano. "A comprehensive pathway map of epidermal growth factor receptor signaling". *Molecular Systems Biology* 1 (2005), p. 2005.0010.

[78] H Ryu, M Chung, M Dobrzyński, D Fey, Y Blum, Lee SS, M Peter, BN Kholodenko, NL Jeon, and O Pertz. "Frequency modulation of ERK activation dynamics rewires cell fate". *Mol Syst Biol.* 11 (2015), p. 838.

[79] Silva D. M. Santos, Peter Verveer, and P. H. Bastiaens. "Growth factor-induced MAPK network topology shapes ErK response determining PC-12 cell fate". *Nature Cell Biology* 9 (2007), pp. 324–330.

[80] Martin H. Schaefer, Jae-Seong Yang, Luis Serrano, and Christina Kiel. "Protein Conservation and Variation Suggest Mechanisms of Cell Type-Specific Modulation of Signaling Pathways". *PLOS Computational Biology* 10 (2014), Issue 6, e1003659.

[81] Andrew D. Sharrocks. "Cell Cycle: Sustained ERK Signalling Represses the Inhibitors". *Current Biology* 16 (2006), p. 14.

[82] Hao Shi, Koon-Kiu Yan, Liang Ding, Chenxi Qian, Hongbo Chi, and Jiyang Yu. "Network Approaches for Dissecting the Immune System." *iScience* 23 (2020), p. 101354.

[83] Chi-Tin Shih, Yen-Jen Lin, Cheng-Te Wang, Ting-Yuan Wang, Chih-Chen Chen, Ta-Shun Su, Chung-Chuang Lo, and Ann-Shyn Chiang. "Diverse Community Structures in the Neuronal-Level Connectome of the Drosophila Brain." *Neuroinformatics* 18 (2020), pp. 267–281.

[84] A Stanoev, A Mhamane, KC Schuermann, HE Grecco, W Stallaert, M Baumdick, Y Bruggemann, MS Joshi, P Roda-Navarro, S Fengler, Rabea Stockert, Lisaweta Roßmannek, Jutta Luig, Aneta Koseska, and Philippe I H Bastiaens. "Interdependence between EGFR and Phosphatases spatially established by vesicular dynamics generates a growth factor sensing and responding network". *Cell Syst.* 7(3) (2018), pp. 295–309.

[85] Angel Stanoev, Akhilesh P Nandan, and Aneta Koseska. "Organization at criticality enables processing of time-varying signals by receptor networks". *Mol Syst Biol.* 16(2) (2020), e8870.

[86] Aneta Stefanovska and Peter V. E. McClintock. *Physics of Biological Oscillators*. Springer: Complexity, 2020.

[87] Steven H. Strogatz. *Nonlinear Dynamics and Chaos - with applications to Physics, Biology, Chemistry and Engineering*. Vol. Second Edition. 2015.

[88] Geiger T, Wehner A, Schaab C, Cox J, and Mann M. "Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins." *Mol Cell Proteomics* 11 (2012), p. M111.014050.

[89] Floris Takens. "Detecting Strange Attractors in Turbulence". *Dynamical Systems and Turbulence. Springer: Berlin/Heidelberg, Germany* (1981).

[90] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. "Recent advances in physical reservoir computing: A review." *Neural Networks* 115 (2019), pp. 100–123.

[91] H.R. Thieme. "Uniform weak implies uniform strong persistence for non-autonomous semiflows." *Proc. Am. Math. Soc.* 127(8) (1999), pp. 2395–2403.

[92] G. Timár, A.V. Goltsev, S.N. Dorogovtsev, and J.F.F. Mendes. "Mapping the Structure of Directed Networks: Beyond the Bow-Tie Diagram". *Phys. Rev. Lett.* 118 (2017), p. 078301.

[93] Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, and Jane B. Reece. *Biology*. Vol. 7. Campbell, 2017. Chap. 11.

[94] Annalaura Vacca, Masayoshi Itoh, Hideya Kawaji, Erik Arner, Timo Lassmann, Carsten O. Daub, Piero Carninci, Alistair R. R. Forrest, Yoshihide Hayashizaki, the FANTOM Consortium, Stuart Aitken, and Colin A. Semple. "Conserved temporal ordering of promoter activation implicates common mechanisms governing the immediate early response across cell types and stimuli". *Open Biology* 8 (2018), p. 180011.

[95] J. Vera, E. B.Canto, P. Wellstead, J. R. Banga, and O. Wolkenhauer. "Power-law models of signal transduction pathways". *Cellular Signalling* 19 (2007), pp. 1531–1541.

[96] Berta Verd, Anton Crombach, and Johannes Jaeger. "Classification of transient behaviours in a time-dependent toggle switch model". *BMC Syst Biol* 8 (2014), p. 43.

[97] Bert Vogelstein and Kenneth W Kinzler. "Cancer genes and the pathways they control." *Nature medicine* 10 (2004), pp. 789–799.

[98] Eberhard O. Voit. "Biochemical Systems Theory: A Review". *Hindawi Publishing Corporation* 897658 (2013).

[99] W. Wang, Y. Qiao, and Z. Li. "Trends in pharmacological sciences." *Journal of Physics: Conference Series* 39 (2018), pp. 367–386.

[100] Zhihui Wang and Thomas S. Deisboeck. "Dynamic targetting in cancer treatment." *Frontiers in Physiology* 10 (2019), p. 96.

[101] Anna Weller, Gerard N. Bischof, Philipp Schlueter, Nils Richter, Julian Dronse, Oezguer Onur, Bernd Neumaier, Juraj Kukolja, Karl-Josef Langen an d Gereon Fink, Angela Kunoth, Yaping Shao, Thilo van Eimeren, and Alexander Drzezga. "Finding New Communities: A Principle of Neuronal Network Reorganization in Alzheimer's Disease." *BRAIN CONNECTIVITY* 11 (2021), pp. 225–238.

[102] Kai Wu, Yasuyuki Taki, Kazunori Sato, Yuko Sassa, Kentaro Inoue, Ryoi Goto, Ken Okada, Ryuta Kawashima, Yong He, Alan C. Evans, and Hiroshi Fukuda. "The Overlapping Community Structure of Structural Brain Network in Young Healthy Individuals." *PLOS One* 6 (2011), e19608.

[103] R. Yang, L. Zhuhadar, and O. Nasraoui. "Bow-Tie decomposition in directed graphs". *IEEE* 978-0-9824438-2-8 (2011).

[104] Yan Yang, Peng Xie, Yarden Opatowsky, and Joseph Schlessinger. "Direct contacts between extracellular membrane-proximal domains are required for VEGFreceptor activation and cell signaling." *PNAS* 107 (2010), pp. 1906–1911.

[105] Yosef Yarden and Mark X. Sliwkowski. "UNTANGLING THE ErbB SIGNALLING NETWORK". *NATURE REVIEWS, MOLECULAR CELL BIOLOGY* 2 (2001), pp. 127–137.

[106] Xiaowei Ying and Xintao Wu. "On Randomness Measures for Social Networks". *Conference on Data Mining, SIAM International* (2009).

Acknowledgements

Looking back at the last 3.5 years seems like life is indeed chaotic and unpredictable. I started my scientific voyage in Dortmund, and writing this thesis in Bonn while the world is still recovering from a pandemic. This journey was full of great and challenging times, bringing the experiences that become a part of me. Over these years, I am grateful that I got the opportunity to work, learn and grow with many incredible people.

First and foremost, I'd like to thank my supervisor, Dr. Aneta Koseska, for all the guidance and support from the very first day. I will always be grateful for the discussions with you, that not only gave a distinctive shape to this thesis but also helped me align my ideas and thought-process over these years. Discussions with you made me even more curious everyday and also taught me an invaluable life lesson of self-criticism.

I would like to extend my sincere thanks to TAC members, Prof. Dr. Philippe Bastiaens and Dr. Peter Beiling for invaluable suggestions, discussions, and constructive criticism throughout the development of this project.

To all the people in the CCL lab, I am fortunate to have amazing colleagues and I enjoyed working with every single one of you. Thanks to Akhilesh for your support and assistance throughout these years, and also for me not eating me. Thank you, Abhishek, Daniel and Robert for all the valuable scientific discussions and random amusing conversations. Thanks to Antje for arranging all the fun activities and creating a lively atmosphere around. Thanks to Angel for providing initial support in the early stage of this project. Thanks to all the members of Abt. 2 for the early morning discussions in group meetings. Special thanks to Sarah for your wise words, tasty food and also for creating nonsense stories. Thank you Mai, Kitso and Manual for all the creating a friendly and merry environment.

I feel fortunate and privileged to be a part of both, MPI of Molecular Physiology and MPI for Neurobiology of Behavior. Being a graduate student in the IMPRS-LM was an honor for me. A special thanks to Christa and Lucia for your support and organization throughout these years.

I'd like to take this opportunity to thank IISER Mohali for providing a unique multi-disciplinary course structure, and developing a strong scientific foundation that prepared me for my doctoral research.

Things become *easy peasy and lemon squeezy* with friends like Anuj, Ganguly, Bob and Suhani, who never cease to provide support and unwanted trash talks since 2013. Special thanks to Im Krausfeld 47 squad: Frederik, Negar and Neha for keeping me insanely sane with games evenings, parties, random political discussions and food.

I express my final gratitude to my family, Gaurav and Sonia for all the moral support and for listening to my weird world views over weekend video calls. I dedicate this thesis to my parents, Mrs. Uma Yadav and Mr. Ghanshyam Yadav, thank you for always believing in me and encouraging me unconditionally.

Manish Yadav
November 2022