Nonlinear Convergent Dynamics for Temporal Information Processing on Novel Quantum and Classical Devices

Author:
Chen, Jiayin

Publication Date:
2022

DOI:
https://doi.org/10.26190/unsworks/24115

License:
https://creativecommons.org/licenses/by/4.0/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/100408 in https://unsworks.unsw.edu.au on 2024-03-29

Nonlinear Convergent Dynamics for Temporal Information Processing on Novel Quantum and Classical Devices

Jiayin Chen

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Electrical Engineering and Telecommunications

Faculty of Engineering

The University of New South Wales

Feb 2022

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Chen**

First name: **Jiayin**

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Electrical Engineering and Telecommunications** Faculty: **Faculty of Engineering**

Title: Nonlinear Convergent Dynamics for Temporal Information Processing on Novel Quantum and Classical Devices

Abstract

Reservoir computing is an emerging neuromorphic computing paradigm for temporal processing tasks that is also energy and memory-efficient. It has demonstrated promising performance on chaotic modeling, speech processing and time series prediction. This thesis presents theoretical and experimental studies aimed at expanding the toolkit for temporal information processing by utilizing uniformly convergent dynamical systems as reservoir computers. Reservoir computing offloads computations to naturally occurring or engineered nonlinear dynamical systems and typically only a simple readout mechanism is optimized to perform temporal tasks. The uniform convergence property ensures that the computation performed is asymptotically independent of the reservoir computer's initial condition.

Physical reservoir computers are hardware implementations of reservoir computers for fast signal processing. We propose two families of universal quantum reservoir computers as physical reservoir computers–the Ising quantum reservoir computers and the gate-model quantum reservoir computers–that are both based on uniformly convergent dissipative quantum dynamics. We demonstrate numerically with the Ising scheme and experimentally with the gate-model scheme, that small and noisy quantum reservoirs can tackle nonlinear temporal tasks.

The study of quantum reservoir computers is followed by a theoretical effort in broadening the applications of reservoir computers. We study a general architecture of reservoir computing, in which reservoir computers governed by different dynamics are interconnected in an output-feedback configuration. This architecture is motivated by the use of nonlinear closed-loop structures to better capture data that demonstrate nonlinear feedback phenomena, akin to the Wiener-Hammerstein feedback model for system identification. A theorem for interconnected reservoir computers to be uniformly convergent is derived. We then show that uniformly convergent reservoir computers with output feedback implement a large family of nonlinear autoregressive models. Finally, we consider the reservoir design problem and propose an efficient algorithm to optimize the reservoir internal parameters, and show the almost sure convergence to a Kuhn-Tucker point under noisy state measurements.

Declaration relating to disposition of project thesis/dissertation

I hereby grant the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public version of the thesis.

Signature **Jiayin Chen** Witness Date **11 Feb 2022**

FOR OFFICE USE ONLY Date of completion of requirements for Award

Originality Statement

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Jiayin Chen
11 Feb 2022

Copyright Statement

I hereby grant the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public version of the thesis.

Jiayin Chen
11 Feb 2022

Authenticity Statement

I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

Jiayin Chen
11 Feb 2022

Abstract

Reservoir computing is an emerging neuromorphic computing paradigm for temporal processing tasks that is also energy and memory-efficient. It has demonstrated promising performance on chaotic modeling, speech processing and time series prediction. This thesis presents theoretical and experimental studies aimed at expanding the toolkit for temporal information processing by utilizing uniformly convergent dynamical systems as reservoir computers. Reservoir computing offloads computations to naturally occurring or engineered nonlinear dynamical systems and typically only a simple readout mechanism is optimized to perform temporal tasks. The uniform convergence property ensures that the computation performed is asymptotically independent of the reservoir computer's initial condition.

Physical reservoir computers are hardware implementations of reservoir computers for fast signal processing. We propose two families of universal quantum reservoir computers as physical reservoir computers–the Ising quantum reservoir computers and the gate-model quantum reservoir computers–that are both based on uniformly convergent dissipative quantum dynamics. We demonstrate numerically with the Ising scheme and experimentally with the gate-model scheme, that small and noisy quantum reservoirs can tackle nonlinear temporal tasks.

The study of quantum reservoir computers is followed by a theoretical effort in broadening the applications of reservoir computers. We study a general architecture of reservoir computing, in which reservoir computers governed by different dynamics are interconnected in an output-feedback configuration. This architecture is motivated by the use of nonlinear closed-loop structures to better capture data that demonstrate nonlinear feedback phenomena, akin to the Wiener-Hammerstein feedback model for system identification. A theorem for interconnected reservoir computers to be uniformly convergent is derived. We then show that uniformly convergent reservoir computers with output feedback implement a large family of nonlinear autoregressive models. Finally, we consider the reservoir design problem and propose an efficient algorithm to optimize the reservoir internal parameters, and show the almost sure convergence to a Kuhn-Tucker point under noisy state measurements.

Acknowledgements

I would first like to thank Dr. Hendra Nurdin, my primary supervisor, for his invaluable support and guidence during this doctoral undertaking. His attention to details has motivated me to strike to always improve. I am also very grateful to Prof. Victor Solo and Prof. Robert Malaney for their advice and mentorship throughout the journey. I am thankful to my collaborator Prof. Naoki Yamamoto, who provided an invaluable opportunity for me to visit Japan and learn from his talented crew.

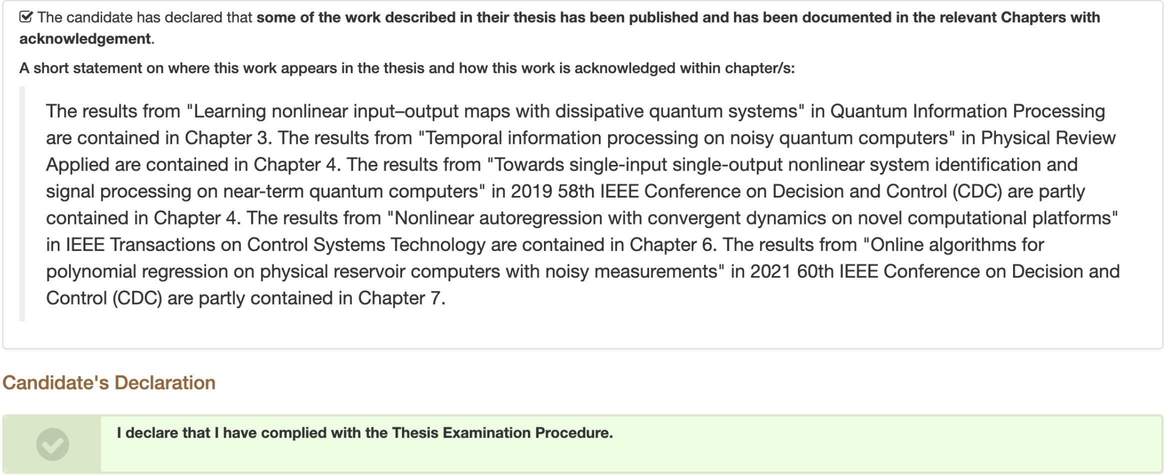
Sincere thanks are also due to Matt, Ian, Michael, Guodong, Daoyi, Elanor, Shota and Hidehiro for all the inspirations and opportunities they have kindly provided.

I would like to thank all the people who I have learned from: Nina, James, Nadia, Alfred, Thien, Harrison, Riddhi, Kev, Jen, Simon, Lachlan, Al, Dan, Joyce, Liam, Serge, Haris, Kamran, and anyone else my memory neglects.

Inclusion of Publications Statement

The results from ‘Learning nonlinear input–output maps with dissipative quantum systems’ in Quantum Information Processing are contained in Chapter 3. The results from ‘Temporal information processing on noisy quantum computers’ in Physical Review Applied are contained in Chapter 4. The results from ‘Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers’ in 2019 58th IEEE Conference on Decision and Control (CDC) are partly contained in Chapter 4. The results from ‘Nonlinear autoregression with convergent dynamics on novel computational platforms’ in IEEE Transactions on Control Systems Technology are contained in Chapter 6. The results from ‘Online algorithms for polynomial regression on physical reservoir computers with noisy measurements’ in 2021 60th IEEE Conference on Decision and Control (CDC) are partly contained in Chapter 7.

I declare that I have complied with the Thesis Examination Procedure.



☑ The candidate has declared that **some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement**.

A short statement on where this work appears in the thesis and how this work is acknowledged within chapter/s:

> The results from "Learning nonlinear input–output maps with dissipative quantum systems" in Quantum Information Processing are contained in Chapter 3. The results from "Temporal information processing on noisy quantum computers" in Physical Review Applied are contained in Chapter 4. The results from "Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers" in 2019 58th IEEE Conference on Decision and Control (CDC) are partly contained in Chapter 4. The results from "Nonlinear autoregression with convergent dynamics on novel computational platforms" in IEEE Transactions on Control Systems Technology are contained in Chapter 6. The results from "Online algorithms for polynomial regression on physical reservoir computers with noisy measurements" in 2021 60th IEEE Conference on Decision and Control (CDC) are partly contained in Chapter 7.

Candidate's Declaration

✓ **I declare that I have complied with the Thesis Examination Procedure.**

Publications

This thesis is based on the following collection of papers which have been published in or submitted to peer-reviewed journals or conferences.

Journal papers

- **Chen, J.** and Nurdin, H.I., "Learning nonlinear input–output maps with dissipative quantum systems," Quantum Information Processing 18, 198 (2019), https://doi.org/10.1007/s11128-019-2311-9.
- **Chen, J.**, Nurdin, H.I. and Yamamoto, N., 2020. "Temporal information processing on noisy quantum computers," Physical Review Applied, 14(2), p.024065.
- **Chen, J.** and Nurdin, H.I., "Nonlinear autoregression with convergent dynamics on novel computational platforms," IEEE Transactions on Control Systems Technology, doi: 10.1109/TCST.2021.3136227.
- **Chen, J.** and Nurdin, H.I., "A small-gain theorem for discrete-time convergent systems and its applications," arXiv preprint arXiv:2105.02376, 2021. Submitted for publication.

Conference papers

- **Chen, J.**, Nurdin, H.I. and Yamamoto, N., "Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers," 2019 58th IEEE Conference on Decision and Control (CDC), 2019, pp. 401-406, doi: 10.1109/CDC40024.2019.9029180.
- **Chen, J.** and Nurdin, H.I., "Online algorithms for polynomial regression on physical reservoir computers with noisy measurements," 2021 60th IEEE Conference on Decision and Control (CDC), 2021, pp. 1502-1508, doi: 10.1109/CDC45484.2021.9683633.

List of Publications Not Included in This Thesis

- Chen, J., Nurdin, H.I., “Generalized simulated annealing with sequentially modified cost function for combinatorial optimization problems,” 2019 Australian & New Zealand Control Conference (ANZCC), 2019, pp. 48-53, doi: 10.1109/ANZCC47194.2019.8945670.
- Nurdin H.I., Amini N.H. and Chen J., “Data-driven system identification of linear quantum systems coupled to time-varying coherent inputs,” 2020 59th IEEE Conference on Decision and Control (CDC), 2020, pp. 3829-3835, doi: 10.1109/CDC42340.2020.9303815.

Notation

- $\mathbb{R}$: the set of real numbers.
- $\mathbb{C}$: the set of complex numbers.
- ι: the imaginary number $\iota = \sqrt{-1}$.
- $\mathbb{Z}$: the set of integers.
- $\mathbb{Z}_-$: the set of integers less than or equal to zero.
- $\mathbb{Z}_{\geq k_0}$: the set of integers greater than or equal to $k_0 \in \mathbb{Z}$.
- $\mathbb{R}^n$: the set of column vectors of dimension n with real entries.
- $x^\top$: the transpose of $x \in \mathbb{R}^n$.
- $\| \cdot \|$: the Euclidean norm.
- (x_1, x_2): is a concatenation of vectors $x_1 \in \mathbb{R}^{n_1}$ and $x_1 \in \mathbb{R}^{n_2}$ into a column vector.
- $\mathbb{R}^{n \times m}$: the set of $n \times m$ real matrices.
- $\mathbb{C}^{n \times m}$: the set of $n \times m$ complex matrices.
- $A^\top$: the transpose of $A \in \mathbb{C}^{n \times m}$.
- $A^\dagger$: the conjugate transpose of $A \in \mathbb{C}^{n \times m}$.
- $A \succ 0$: positive definite matrix A.
- $A \succeq 0$: positive semi-definite matrix A.
- $A \prec 0$: negative definite matrix A.
- $A \preceq 0$: negative semi-definite matrix A.
- $\| \cdot \|_p$: the Schatten-p norm for complex matrices, defined as $\|A\|_p = \mathrm{Tr}(\sqrt{A^\dagger A}^p)^{1/p}$, where $\mathrm{Tr}(\cdot)$ is the trace.
- $(\mathbb{R}^n)^\mathbb{Z}$: the set of infinite sequences of vectors in $\mathbb{R}^n$. That is, $u \in (\mathbb{R}^n)^\mathbb{Z}$ with $u = \{u_k\}_{k \in \mathbb{Z}}$ and $u_k \in \mathbb{R}^n$ for all $k \in \mathbb{Z}$.

- l_n^∞: the set of bounded infinite sequences of vectors in $\mathbb{R}^n$. That is, $u \in l_n^\infty$ if $u = \{u_k\}_{k\in\mathbb{Z}}$ with $u_k \in \mathbb{R}^n$ and $\|u\|_\infty \coloneqq \sup_{k\in\mathbb{Z}} \|u_k\| < \infty$.
- $(\mathbb{R}^n)^{\mathbb{Z}_-}$: the set of left-infinite sequences of vectors in $\mathbb{R}^n$. That is, $u \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ with $u = \{u_k\}_{k\in\mathbb{Z}_-}$ and $u_k \in \mathbb{R}^n$ for all $k \in \mathbb{Z}_-$.
- $\mathbb{R}^{\mathbb{Z}}, l^\infty$ and $\mathbb{R}^{\mathbb{Z}_-}$: the set of real infinite sequences, the set of bounded real infinite sequences and the set of real left-infinite sequences, respectively.
- $D^{\mathbb{Z}}$: the set of infinite sequences taking values in a compact set $D \subset \mathbb{R}$.
- $D^{\mathbb{Z}_-}$: the set of left-infinite sequences taking values in a compact set $D \subset \mathbb{R}$.
- $P_n^{\mathbb{Z}_-}$: is the canonical projection from $(\mathbb{R}^n)^{\mathbb{Z}}$ onto $(\mathbb{R}^n)^{\mathbb{Z}_-}$ (we write $P_1^{\mathbb{Z}_-}$ as $P^{\mathbb{Z}_-}$).
- $z_n^{-\tau}$: time shift operator. For any $\tau \in \mathbb{Z}$, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $k \in \mathbb{Z}$, $z_n^{-\tau}(u)|_k = u_{k-\tau}$ (we write $z_1^{-\tau}$ as $z^{-\tau}$).
- $\otimes$: tensor product (Kronecker product).
- Tr_{H_B}: partial trace over H_B. That is, if $\rho \in H_A \otimes H_B$ where H_A and H_B are Hilbert spaces and $\mathrm{Tr}_{H_B}(\rho)$ is the partial trace over H_B.
- $\circ$: function composition.

Contents

Abstract iii

Acknowledgements iv

Inclusion of Publications Statement v

Publications vi

Notation viii

Contents x

List of Figures xvi

List of Tables xxi

1 Introduction 1

- 1.1 Reservoir computing . . . 1
- 1.2 Physical reservoir computing . . . 3
  - 1.2.1 Noisy intermediate-scale quantum technology . . . 4
  - 1.2.2 Quantum reservoir computers and related works . . . 6
- 1.3 Architectures of reservoir computers . . . 7
  - 1.3.1 Stochastic modeling . . . 9

1.4 Reservoir design 9
1.5 Summary of contributions 11
1.6 Outline 12

2 Preliminaries 15

2.1 Notation 15
2.2 Quantum systems 16
2.2.1 Closed quantum systems 17
2.2.2 Dissipative quantum systems 18
2.2.3 Quantum noise 20
2.3 Quantum computing 22
2.3.1 Quantum gates and quantum circuits 22
2.4 Reservoir computing 23
2.4.1 The uniform convergence property 25
2.4.2 Sufficient conditions for the uniform convergence property 27
2.4.3 Filters and functionals 28

3 Ising quantum reservoir computers 31

3.1 The universality property 32
3.1.1 Fading memory property 33
3.1.2 Polynomial algebra 34
3.1.3 The separation property 35
3.1.4 The universality theorem 35
3.2 A universal class of QRCs implemented by dissipative quantum systems . . 36
3.2.1 Uniform convergence 38
3.2.2 Fading memory and polynomial algebra 39

3.2.3 Separation of points 43

3.3 Numerical experiments 45

3.3.1 Overview of SA performance 47

3.3.2 SA performance under noise 49

3.3.3 Effect of different input encodings 51

3.3.4 Further comparison with ESNs 52

3.4 Conclusion and outlook 56

4 Gate-model quantum reservoir computers 58

4.1 A universal class of QRCs 60

4.1.1 Invariance under certain noise 64

4.2 A subclass implementable on noisy gate-model quantum devices 65

4.3 Realization of a subclass on current quantum hardware 66

4.3.1 Implementation schemes 66

4.3.2 Unitary trick for efficient QRC predictions 69

4.4 Proof-of-principle experiments 70

4.4.1 Quantum circuits for QRCs 72

4.4.2 Experimental implementation 74

4.4.3 QRC task performance 75

4.5 Discussion 78

4.6 Conclusion 79

5 Interconnected reservoir computers 80

5.1 Stability concepts 82

5.2 A UIOC small-gain theorem 85

5.3 Interconnected RCs for temporal tasks 91

5.3.1 Echo-state networks (ESNs) 93
5.3.2 Quantum reservoir computers (QRCs) 94
5.4 Conclusion 101

6 Nonlinear autoregression with reservoir computers 102

6.1 Uniformly convergent feedback dynamics 104
6.2 NARX(∞) models 107
6.2.1 Asymptotic stationarity and ergodicity 110
6.3 Parameter estimation 112
6.3.1 Echo-state networks (ESNs) 113
6.3.2 Lur'e systems 114
6.3.3 Quantum reservoir computers (QRCs) 116
6.4 Numerical examples 118
6.4.1 Nonlinear quantum optics 119
6.4.2 Finance 122
6.4.3 Coupled electric drive system 123
6.5 Discussion 125
6.6 Conclusion 126

7 Reservoir design 127

7.1 The reservoir design problem 129
7.1.1 Noisy state measurements 131
7.2 Simultaneous perturbation stochastic approximation 131
7.3 Algorithm and its analysis 133
7.3.1 Motivation 133
7.3.2 Constructing regressors from four noisy state measurements 136

7.3.3 Almost sure convergence 139

7.4 Numerical examples 142

7.4.1 Sensitivity of reservoir internal parameters 144

7.4.2 Performance 146

7.5 Discussion 148

7.6 Conclusion 152

8 Conclusion and Future Directions **153**

A Appendix for Chapter 4 **157**

A.1 Invariance under time-invariant readout error 157

A.2 Monte Carlo estimation 158

A.3 Implementation using QND measurements 161

A.4 Hardware specifications 163

A.5 Quantum circuits for QRCs 164

A.6 Full input-output sequential data 165

A.7 Measurement and simulation data 166

B Appendix for Chapter 5 **171**

B.1 Proof of the UIOS Theorem 5.2 171

B.2 Uniform convergence of Lur'e system 176

B.3 Trace inequality 180

B.4 Noisy QRC simulation 180

C Appendix for Chapter 6 **182**

C.1 Unitary gates for QRCs 182

C.2 Parameters of selected RCs 183

References **184**

List of Figures

1.1 Outline of the thesis. 12

2.1 From left to right are the quantum circuit diagrams of $U_3(\theta, \phi, \lambda)$, $R_X(\theta)$, $R_Y(\theta)$ and $R_Z(\theta)$. 23

2.2 From left to right are the quantum circuit diagrams of CNOT and controlled-U. 23

3.1 Typical SA outputs during the evaluation phase, for the (a) LRPO, (b) Missile (c) NARMA15 and (d) NARMA20 tasks. The leftmost, middle and rightmost panels show the outputs for timesteps 1501–1530, 2001–2030 and 2471–2500, respectively. 48

3.2 Average SA NMSE for the (a) LRPO, (b) Missile, (c) NARMA15 and (d) NARMA20 tasks, the error bars represent the standard error. For comparison, horizontal dashed lines labeled with 'Em' indicate the average performance of ESNs with m computational nodes, and horizontal dot-dashed lines labeled with 'Vo, p' indicates the performance of Volterra series with kernel order o and memory p. Overlapping dashed and dot-dashed lines are represented as dashed lines. 48

3.3 Average SA NMSE under decoherence. Error bars represent standard errors. 50

3.4 Average sum of complex modulus of off-diagonal elements in the system density operator for timesteps 1501–1550, under the (a) dephasing noise, (b) decaying noise, (c) GAD with $\lambda = 0.4$ and (d) GAD with $\lambda = 0.6$. Row $n-1$ corresponds the n-qubit SA. 51

3.5 Average NMSE for different input encodings, for the (a) LRPO, (b) Missile (c) NARMA15 and (d) NARMA20 tasks. Error bars represent standard errors. 52

3.6 Average SA NMSE as the state-space size and the number of computational nodes vary for all computational tasks. The average NMSE for E256 with the same number of computational nodes is plotted for comparison. The data symbols obscure the error bars, which represent the standard error . . 54

3.7 Average ESNs NMSE as the state-space size and the number of computational nodes vary for all computational tasks. The data symbols obscure the error bars, which represent the standard error 55

4.1 Quantum circuit interpretation of the QRC subclass described in Sec. 4.2. Here $\rho_{k-1}^{(K)}$ and σ_K are two quantum registers (i.e., groups of qubits) whereas $\rho(u_k)$ and ρ_{ϵ_K} are two single-qubit states. The unitaries $U_1^{(K)}, U_0^{(K)^\dagger}$ act on $\rho_{k-1}^{(K)}$, controlled by $\rho(u_k)$. The right-most operation (SW's) swaps the states of $\overline{\rho}_{k-1}^{(K)}$ and σ_K, controlled by ρ_{ϵ_K}. 65

4.2 Quantum circuit schematics for (a) $U_0(\boldsymbol{\theta})$ and (b) $U_1(\boldsymbol{\phi})$ employed in proof-of-principle experiments, described by (4.7). Here j_t and j_c are the target and control qubits, respectively. The unitaries $U_0(\boldsymbol{\theta}), U_1(\boldsymbol{\phi})$ consist of N_0, N_1 layers of highlighted gate operations, with each layer acting on a different qubit pair (j_t, j_c). 73

4.3 Qubit coupling maps of the IBM superconducting quantum processors. (a) The 20-qubit Boeblingen device. (b) Both the 5-qubit Ourense and Vigo devices. 74

4.4 The QRC's predicted outputs for (a) the multi-step prediction problem and (b) the map emulation problem. Rows and columns in (a) correspond to different tasks and QRCs, respectively. The first column in (b) corresponds to the multiplexed QRC. 76

4.5 The spatial multiplexing schematic. The same input sequence is injected into two distinct 5-qubit QRCs. The internal states $\mathrm{Tr}(\rho_k Z^{(i)})$ of the two QRCs are linearly combined to form a single output. 77

5.1 Schematic of the interconnected system (5.5). 85

5.2 Target outputs $y'_{l'}(k)$ and the closed-loop ESN outputs $\hat{y}'_{l'}(k)$ for $k = 501, \ldots, 540$ with (a) $l' = 1$ under a uniform random input $w'_1(k)$ and (b) $l' = 2$ under a sum of sinusoidals $w'_2(k) = \sin(2\pi k/25) + \sin(\pi k/5)$. 94

5.3 Schematic of an interconnected QRC described by (5.20). 95

5.4 Target outputs $y'_{l'}(k)$ and the closed-loop QRC outputs $\hat{y}'_{l'}(k)$ for $k = 501, \ldots, 540$ with (a) $l' = 1$ under a uniform random input $w'_1(k)$ and (b) $l' = 2$ under a sum of sinusoidals $w'_2(k) = \sin(2\pi k/25) + \sin(\pi k/5)$. 98

6.1 Schematic of RCs operating in an output-feedback configuration described by (6.1), where z^{-1} is the one-step time delay operator. 105

6.2 (a) Kerr-nonlinear cavity with two input-output ports. The top mirror is fully reflective (without loss) and the other mirrors are partially transmitting. (b) The simulated trajectory of β_k and the data employed. 120

6.3 The noisy time series β_k, the signal $\alpha_k = \mathrm{Tr}((L_2 + L_2^\dagger)\rho_k)$, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e perdiction on the first 100 validation data. (d) The ESN residual sample ACF (horizontal blue lines show the 95% CI). (e) The ESN residual Q-Q plot. (f) The QRC residual sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACf. (i) The Lur'e residual Q-Q plot. 121

6.4 The finance time series, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e prediction on validation data. (d) The ESN residual sample ACF. (e) The ESN residual Q-Q plot. (f) The QRC residual sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACF. (i) The Lur'e residual Q-Q plot. 122

6.5 Schematic of spatial multiplexing. Here, x_k and $\overline{x}_k$ are the states of the first and second RC member, respectively. The one-step ahead prediction of the multiplexed RC is $\hat{y}_k = W_1^\top h_1(x_k) + W_2^\top h_2(\overline{x}_k)$. 124

6.6 The target output, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e prediction on validation data. (d) The ESN residual sample ACF. (e) The ESN residual Q-Q plot. (f) The QRC sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACF. (i) The Lur'e residual Q-Q plot. The sample cross-correlation between inputs and residuals for (j) the ESN, (k) the QRC and (l) the Lur'e. 124

7.1 Box plot of the MSE of ESNs during (a) the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) the testing phase ($\mathrm{MSE}_{\mathrm{test}}$) for the LRPO task. The inset in (a) shows the box plot for reservoir size 50 during the training phase. 145

7.2 Box plot of the MSE of ESNs during (a) the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) the testing phase ($\mathrm{MSE}_{\mathrm{test}}$) for modeling the finance time series. 145

7.3 The average ESN MSE against iteration for the LRPO task. The first column shows the average ESN MSE during the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and the second column shows the average ESN MSE during the testing phase ($\mathrm{MSE}_{\mathrm{test}}$). The first row corresponds to a linear output, the second row corresponds to a quadratic output and the third row corresponds to a cubic output. MSPSA refers to Algorithm 2 and SPSA refers to Algorithm 1.148

7.4 The average ESN MSE against iteration for the LRPO task with a cubic readout function under Algorithm 3 (labelled as SPSA) and Algorithm 3 where $W_l^* = \left(\hat{H}(\gamma_l)^\top \hat{H}(\gamma_l)\right)^{-1} \hat{H}(\gamma_l)Y$ (labelled as USPSA). (a) shows the average ESN MSE during the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) shows the average ESN MSE during the testing phase ($\mathrm{MSE}_{\mathrm{test}}$). 149

7.5 The average ESN MSE against iteration for the finance time series task with a cubic readout function. (a) shows the average ESN MSE during the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) shows the average ESN MSE during the testing phase ($\mathrm{MSE}_{\mathrm{test}}$). 149

A.1 Quantum circuit implementing the QND measurements by coupling ancilla qubits $|0\rangle^{\otimes n}$ with the QRC system qubits $|\psi\rangle_{k-1}$. 161

A.2 Quantum circuits for the (a) 10-qubit QRC and (b) 4-qubit QRC on the Boeblingen device. 165

A.3 Quantum circuits for the (a) 5-qubit Ourense QRC and (b) 5-qubit Vigo QRC. 165

A.4 Full washout, train and test input-output sequences for (a) the multi-step ahead prediction problem and (b) the map emulation problem. The first row in (a) and (b) shows the input sequences. 166

A.5 The full target output sequences, the train and test output sequences of the four QRCs for each task on the multi-step ahead prediction problem. Each column corresponds to each n-qubit QRC outputs and each row corresponds to each task. 167

A.6 The full target output sequences, the train and test output sequences of the QRCs for each task on the map emulation problem. (a) Shows the two train output sequences and (b) shows the test output sequence. The columns (from the left to the right) correspond to the multiplexed 5-qubit QRCs, 5-qubit Ourense QRC and the 5-qubit Vigo QRC. Each row corresponds to each task. 168

A.7 Input sequence, experimental and simulation results for each qubit of the four QRCs at each time step $k = 1, \ldots, 30$, for the multi-step ahead prediction problem. 169

A.8 Experimental and simulation results for each qubit $i = 0, \ldots, 4$ and each time step $k = 1, \ldots, 24$, for the map emulation problem. Three input sequences are used in this problem, labeled as inputs I, II and III. Row i in a sub-figure corresponds to the experimental data for the i-th input sequence. Column j corresponds to the experimental data for the j-th qubit. (a) Shows the experimental data for the 5-qubit Ourense QRC. (b) Shows the experimental data for the 5-qubit Vigo QRC. 170

B.1 Schematic of the closed-loop system (B.13) consisting of the plant and the observer error dynamics. 177

C.1 Parameters of the selected ESNs. 183

C.2 Parameters of the selected Lur'e systems. 183

List of Tables

3.1 Average 4-qubit SA and E256 NMSE for the LRPO, Missile, NARMA15, NARMA20, NARMA30 and NARMA40 tasks. Results are rounded to two significant figures. The notation (± se) denotes the standard error. 53

4.1 NMSEs on the multi-step ahead prediction problem. 77

4.2 NMSEs on the map emulation problem. 77

5.1 Summary of stability concepts and their acronyms. 85

6.1 Dimension (or qubit number) N, p-value, RMSE and average RMSE ($\overline{\text{RMSE}}$) of RCs for stochastic modeling. 121

Chapter 1

Introduction

1.1 Reservoir computing

The last two decades have witnessed a profound shift in computational paradigms from von Neumann architecture digital computers to nature inspired computing systems that address the demand for energy and memory-efficient intelligent devices. On one side of the spectrum, biologically inspired systems simulated on digital computers–'a machine that thinks', such as artificial neural networks, have demonstrated strong capacity for solving temporal information processing tasks. They power text-to-speech synthesis [1], voice search [2] and sequence-to-sequence machine translation [3] and show remarkable performance in time series forecasting [4]. The success of artificial neural networks does not come without costs. Current deep artificial neural networks require vast amounts of computational resources to approach state-of-the-art performance [5,6]. Training for these networks are mostly based on gradient-based methods that incur a long training time and suffer from the vanishing and exploding gradient problems [7, 8]. Another criticism is that although artificial neural networks began as an attempt to mimic the human brain to perform tasks that digital computers had difficulties with, current models of artificial neural networks do not sufficiently reflect neuronal functions in the brain [9].

Reservoir computing (RC) has been proposed to circumvent the expensive computational costs of artificial neural networks [10, 11]. The RC framework uses an arbitrary but fixed nonlinear dynamical system, the 'reservoir', to map input signals into its high-dimensional state space. Only a simple readout mechanism is trained (often via linear regression) to approximate target output signals. The use of simple readout has connections to the biological concept of mixed selectivity, as demonstrated in monkeys [12]. This simple and fast training process drastically reduces the training cost compared with alternative schemes such as recurrent neural networks, and is the key advantage of RC. In this thesis, we focus on RCs described by discrete-time state space models. Other dynamical system models, such as the delayed dynamical systems [13], are beyond the scope of this thesis.

A prominent example of RC are the echo state networks (ESN), which has demonstrated remarkable ability to predict chaotic time series [10, 14]. Liquid state machines were independently proposed to use a spiking neural network [15] with arbitrary but fixed connectivity as the 'reservoir' to process input signals. Another attractiveness of the RC framework is that naturally occurring dynamical systems (with certain properties) can be exploited for temporal information processing. This is in sharp contrast to artificial neural networks, which are software and simulation-based computing schemes. Following the initial success of RC, different research directions have emerged:

- Hardware implementations of RC, *physical reservoir computers* (abbreviated as RCs), for realizing fast processing speed with low computational costs.
- Introduction of new architectures and applications of RCs to improve their versatility.
- Development of algorithms to optimize internal parameters of RCs to improve their performance.

This thesis pursues these research directions. We will expand on each direction in the following sections.

1.2 Physical reservoir computing

There is substantial interest in *physical RCs*, hardware implementations of RC for fast processing using less memory and energy, as opposed to software-based implementation on digital computers. The ease of implementation of the RC framework has brought forward many successful experimental demonstrations of physical RCs. For instance, a spintronic RC achieved state-of-the-art performance for spoken digit recognition [16], a photonic RC demonstrated a million words per second speech classification with a low error rate [17] and a FPGA-based RC reached an 160MHz rate for time series prediction [18]. Their training efficiency also makes RC attractive for edge computing, where computations are distributed at the 'edges' closer to the end users to reduce computation and transmission overhead. See [19, 20] for further examples of physical RCs.

Although these previous works have showcased excellent empirical performance on certain tasks, the following question requires further investigation: what are the properties for physical RCs to efficiently solve nonlinear temporal information processing tasks? Unsurprisingly, nonlinearity is necessary for RCs to act as a nonlinear map between input and output sequences; and high dimensionality determines the computational capacity of RC [21]. This computational capacity can be interpreted as the number of linearly independent functions of its stimuli the system can compute. Other properties for RC are the *uniform convergence property*, the *fading memory property* and the *separation property.* The uniform convergence property [22,23], also referred to as the echo state property [24], ensures that the computation performed by an RC is asymptotically independent of its initial condition. The fading memory property [25] states that the (infinite) output sequences of a RC should stay close if the corresponding (infinite) input sequences are close in recent times. The fading memory property is important in a broad class of real-world problems including spoken digit recognition [16] and neural modeling [15]. If a family of RCs has the separation property, then there is a member in this family with sufficiently rich dynamics to distinguish any two different input sequences.

If a family of RCs possesses the uniform convergence, fading memory and separation properties, then this family is *universal* for approximating nonlinear fading memory input-output (I/O) maps [26, 27]. Any nonlinear fading memory map can be approximated arbitrarily closely by some member of this family (in a certain norm), uniformly over time. The well-known Volterra series [25] is universal for nonlinear fading memory maps. Software-based classical RC schemes that are provably universal include the ESNs [27], linear reservoirs with polynomial readouts (LRPO) and the non-homogeneous state affine systems (SAS) [26]. The aforementioned liquid state machines have also been shown to be universal. They have been implemented on hardware based on FPGA [28, 29] and application-specific integrated circuits [30, 31].

Despite recent interest in seeking universal RC schemes, few physical RCs with the universality property has been introduced. While this search is underway, we are witnessing a spectacular growth in near-term quantum computing. Motivated by the increasing availability of quantum devices, in this thesis, we aim to develop universal physical RC schemes on current noisy quantum computers.

1.2.1 Noisy intermediate-scale quantum technology

Quantum computers make use of quantum-mechanical phenomena such as superposition and entanglement to perform computations. Data are encoded as quantum bits (qubits) which can be in a superposition of 0 and 1 binary values simultaneously, hence the capability to represent exponentially more information than the classical bits, making quantum computers promising for solving problems that are previously untenable on conventional digital computers. Followed by the seminar paper for Peter Shor [32], tremendous progress in designing quantum algorithms that offer speedups over classical counterparts has been made [33–35]. These algorithms implemented on quantum computers with 50–100 qubits have the potential to surpass today's most advanced classical supercomputers, but noise in quantum operations will render the quantum computations unreliable.

We are in the midst of noisy intermediate-scale quantum (NISQ) technology era [36], marked by noisy quantum computers consisting of roughly tens to hundreds of qubits.

NISQ quantum computing machines are not equipped with quantum error correction and are thus incapable of performing continuous quantum computation. Algorithms that offer computational speedup, such as the Grover search, can only be demonstrated for very small instances (with three qubits) on NISQ machines [37], and scaling up the number of qubits with quantum error correction is still a distant goal. While awaiting for a full-fledged quantum computer, there is a substantial interest in early applications of NISQ devices that can accelerate the development of practical quantum computers, akin to how the humble hearing aid sparked the development of integrated circuit technology [38].

Several research directions are being explored for NISQ machines. One direction is to demonstrate 'quantum supremacy', in which NISQ machines can outperform the most powerful digital supercomputers. Such tasks include sampling problems such as boson sampling [39, 40] and sampling from random quantum circuits [41]. In 2019, Google demonstrated a form of quantum supremacy on their gate-based superconducting quantum devices [42]. Gate-based quantum devices transform input data according to a unitary operation, specified as a sequence of quantum gate operations and measurements represented by a quantum circuit [43]. In the last few years, companies such as IBM [44], Rigetti Computing [45] and IonQ [46], have made gate-based NISQ devices widely available as cloud-based services for commercial, academic and public use. Another direction is to design hybrid quantum-classical algorithms using short-depth circuits. Many notable experimental demonstrations of NISQ devices employ hybrid algorithms for data classification [47] and quantum chemistry [48]. NISQ devices are also used to solve certain classes of optimization problems. A prominent example is the quantum annealing algorithm which employs quantum fluctuations to search for the solution of classical optimization problems [49]. Quantum annealing encodes problems using Ising Hamiltonians and can be implemented on the commercially available D-Wave machines [50, 51].

An emerging paradigm is to harness the computational capability of dissipative quantum systems, that is, quantum systems under the effect of noise resulting from interactions with the external environment. Dissipative quantum dynamics has been shown to realize universal quantum computation [52] and has been applied in a time-delay manner for supervised quantum machine learning without intermediate measurements [53]. More recently, quantum reservoir computers (QRCs) have been introduced to employ disordered-ensemble dissipative quantum dynamics for classical temporal information processing [54, 55]. Numerical experiments show that this QRC scheme with a small number of qubits can achieve comparable performance to classical RC schemes with a larger number of tunable parameters. However, a theoretical framework for approximating nonlinear fading memory I/O maps by QRCs is still lacking. In Chapter 3 of this thesis, we address this problem by introducing a variation of this previous QRC scheme that is universal in approximating nonlinear I/O maps with fading memory.

QRC schemes introduced in [54, 55] and Chapter 3 are based on Ising Hamiltonians and are suitable for systems such as nuclear magnetic resonance (NMR) instruments [56] and D-Wave machines. Such systems are generally for commercial use. On the other hand, gate-based NISQ machines are becoming increasingly available for the general public. However, realizing these QRC proposals in the quantum gate model remains challenging due to the large number of quantum gates required to implement the Ising Hamiltonian via Trotterization [57]. In Chapter 4 of this thesis, we address this question by proposing a universal family of QRCs implementable on gate-based NISQ devices, and present proof-of-principle experiments on IBM superconducting quantum processors [44].

1.2.2 Quantum reservoir computers and related works

QRC has attracted substantial research interests since its initial proposal in [54, 55]. Here, we take the opportunity to review some recent advances beyond the contributions of this thesis.

For processing classical information, many new quantum implementations of the RC scheme have been introduced. In [58], the authors utilize dissipative superconducting quantum computers as QRCs, where the dissipative property is solely served by the intrinsic hardware noise. This is in contrast with our QRC proposal in Chapter 4, where dissipation is engineered. Nevertheless, the scheme in [58] has not been shown to be universal. The work [59] proposes a universal family of QRCs implemented by Gaussian states and [60] demonstrates that encoding the input to quantum or classical fluctuations of a network of interacting harmonic oscillators achieve high performance comparable to ESNs in several nonlinear tasks. A QRC scheme based on an ensemble of interacting spin-1/2 particles has been proposed and demonstrated numerically using arrays of Rydberg atoms [61]. In [62], continuous-variable QRCs with a single nonlinear oscillator are proposed to reduce the measurement costs of discrete-variable QRCs. The work [63] details numerical investigations of factors that influence the performance of QRCs implemented by Ising Hamiltonians. The authors have empirically found conditions for an optimum input data injection frequency and provided insights into the choice of measurement observables for the QRC output.

QRC has also been proposed to process quantum information as opposed to classical information. In [64], a QRC based on a 2D fermionic lattice is demonstrated to recognize quantum entanglement and estimate nonlinear functions of an input quantum state, such as entropy and log-negativity. The work [65] presents a QRC for quantum state preparation. In [66], QRCs are demonstrated to induce robust quantum gate operations. Recently, Kerr networks are employed as reservoir to classify stochastic time-dependent signals for quantum state measurements [67]. In [68], a framework for QRC under continuous heterodyne measurements is described and employed to classify quantum states of the systems that are part of the same measurement chain as the QRC, providing insights into measurement-contingent advantages and disadvantages of reservoir computing in quantum regimes. For further development of QRC for both classical and quantum tasks, see the recent reviews [69, 70].

1.3 Architectures of reservoir computers

Since the initial success of RC schemes, researchers have been devising new architectures of the original RC schemes to improve their performance and increase their versatility. Inspired by deep neural network architectures, deepESN has been proposed [71–73] to concatenate multiple ESNs in a feedforward configuration. Detailed numerical experiments suggest that deep layered ESNs allow an effective diversification of temporal representation in the layers, by amplifying the effects of the richness of the dynamics measured by the entropy [71]. RCs with output feedback (i.e., a RC's output is fedback as its input at the next discrete time step) have demonstrated a remarkable performance for chaotic time series prediction [10, 74].

The fundamental RC condition, the uniform convergence property or the echo state property, has been generalized to the case of deepESN. The work [75] analyzes the contractivity of nested ESNs and derives a sufficient condition for the echo state property to hold. However, for RC with output feedback, theoretical analysis of the uniform convergence property is so far lacking. In Chapter 5, we consider a more general architecture where RCs governed by different dynamics are interconnected in an output-feedback loop. This RC architecture is motivated by the use of nonlinear closed-loop structures in block-oriented models for system identification, such as the Wiener-Hammerstein feedback model, to better capture data that exhibit nonlinear feedback phenomena [76, 77]. We then derive a small-gain sufficient condition for interconnected RCs to be uniformly convergent. This general theory opens up the potential for a vast number of RC configurations as nonlinear candidate models for system identification and temporal information processing.

After the completion of Chapter 5, we became aware of an independent development of a system identification methodology for Lur'e systems with static nonlinear output feedback [78]. This methodology guarantees the uniform convergence property has been proposed by exploiting special structures of Lur'e systems and only applies to this class of systems. In contrast, we focus on deriving a general framework for interconnected systems, potentially with different dynamics, to be uniformly convergent without specific knowledge on the system structures.

1.3.1 Stochastic modeling

Following the general theory we developed in Chapter 5, we then consider a specific RC architecture–RCs with uniformly convergent output feedback dynamics, for time series modeling and system identification.

One of the promising applications of the RC framework is chaotic time series prediction [10, 74, 79]. However, these works model chaotic systems based on the dynamical system paradigm in a deterministic way. For stochastic modeling, previous works [80, 81] have investigated the potential of RC for approximating input-output maps under stationary stochastic inputs, and considered forecasting, reconstruction and filtering problems in this context. Further substantial research efforts have been focused on testing the empirical performance of RC for modeling stochastic time series; see e.g. [82, 83]. However, a general theoretical framework that takes into account stability, stationarity, ergodicity and Birkhoff-Khinchin ergodicity [84] of RC is still missing.

In Chapter 6, we show that RCs with uniformly convergent output feedback dynamics implement *infinite-order nonlinear autoregressive models with exogenous inputs, or NARX(∞) models.* We then show that such NARX(∞) models can be represented as a *nonlinear moving average models with exogenous inputs, or NMAX(∞) models.* Through this equivalence, we establish asymptotic stationarity and (Birkhoff-Khinchin) ergodicity of the NARX(∞) models induced by RCs with uniformly convergent feedback dynamics. We then employ these models for stochastic modeling and perform validation diagnostics to evaluate the model quality. With this framework, a wide class of NARX(∞) models can be realized, making RC a versatile scheme for stochastic modeling.

1.4 Reservoir design

The performance of RC not only depends on its architecture, but also its reservoir internal parameters. This sensitivity to reservoir parameters has sparked studies in various linear and nonlinear memory capacity measures [21, 85, 86]. The work [81] extends previous capacity measures that are defined using independent input signals to correlated and stationary signals. The optimization of the reservoir internal parameters to improve the RC performance is referred to as the reservoir design problem.

The reservoir problem presents two major challenges. Firstly, reservoir design based on capacity measures is difficult since the capacity measures mentioned above are often a non-convex and nonlinear function of the reservoir internal parameters. Secondly, algorithms that optimize the reservoir internal parameters should preserve the training, energy and memory efficiency of the RC framework. Several research directions have been proposed to tackle these difficulties. One approach is to simplify the capacity measure and therefore simplifying the optimization problem. In [87] a simplified model of time-delay reservoirs is constructed as an approximation which allows a direct functional link between the reservoir internal parameters and its performance. In [81], a Taylor series expansion is used to approximate the capacity measure. These approximations are used to determine the reservoir parameters by solving a well-structured optimization problem, however, they might not adequately capture the full nonlinear model.

Another approach is to employ global optimization algorithm to optimize both the output and internal parameters of the RC by minimizing the mean-squared error between the target and predicted outputs [88, 89]. However, global optimization methods often require a lot of computational resources and exhibit slow convergence, making training of RCs inefficient. To remedy these issues, heuristic algorithms, such as the FORCE learning algorithm [90] and its variant [91] have been proposed to efficiently tune the reservoir internal parameters. Although these heuristic algorithms have demonstrated success in generating complex running and walking human motions, they provide no theoretical convergence guarantee nor robustness to noise. An efficient algorithm that has a theoretical convergence guarantee even under the effect of noise is still awaiting to be developed.

In Chapter 7, we propose an efficient algorithm to optimize the reservoir internal parameters by minimizing the mean-squared error. The proposed algorithm is a stochastic approximation algorithm based on the simultaneous perturbation stochastic approximation (SPSA) algorithm [92]. Our algorithm does not make approximations of the cost function nor the model. Furthermore, we will show that using our algorithm, the reservoir internal parameters exhibit an almost sure asymptotic convergence to a Kuhn-Tucker (KT) point even under the effect of state measurement noise. To ensure the convergence, we only need four noisy state measurements to approximate the gradient regardless of the number of optimization variables, making it efficient for high-dimensional problems.

1.5 Summary of contributions

This thesis is the synthesis of a number of research contributions and publications. Citations to the relevant publications where these contributions were presented will be provided.

1. The provision of a learning theory for dissipative quantum systems and the introduction of a universal class of QRCs governed by Ising Hamiltonians for approximating nonlinear I/O maps with fading memory are presented in Chapter 3.
2. A detailed numerical study of the aforementioned universal QRC family is also presented in Chapter 3.
3. The proposal of a universal QRC family that are suitable for implementation on gate-based NISQ machines is presented in Chapter 4.
4. Proof-of-principle experiments of the above QRC family on the remotely accessed IBM superconducting quantum processors [44] are demonstrated in Chapter 4.
5. A small-gain theorem to provide sufficient conditions for interconnected dynamical systems to be uniformly convergent is derived in Chapter 5.
6. The introduction of interconnected RCs as nonlinear models with closed-loop structures for system identification and their numerical demonstrations are presented in Chapter 5

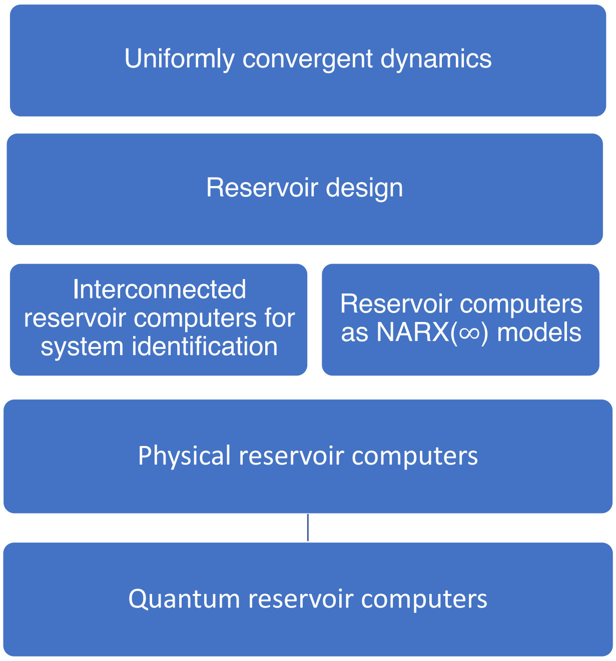

Uniformly convergent dynamics
Reservoir design
Interconnected reservoir computers for system identification
Reservoir computers as NARX(∞) models
Physical reservoir computers
Quantum reservoir computers

Figure 1.1: Outline of the thesis.

7. The introduction of NARX(∞) models induced by RCs with uniformly convergent output feedback dynamics is presented in Chapter 6.

8. The asymptotic stationarity, ergodicity and Birkhoff-Khinchin ergodicity of NARX(∞) models induced by RCs are also established in Chapter 6.

9. A stochastic approximation algorithm is designed to optimize the reservoir internal parameters under noisy state measurements and is shown to exhibit almost sure convergence to a KT point in Chapter 7.

1.6 Outline

This thesis will study nonlinear convergent dynamics for temporal information processing at various levels of abstraction as depicted in Fig. 1.1. We present the thesis in a 'bottom-up' approach. As we traverse across the thesis, we initially focus on QRCs, a specific hardware implementation of RC for approximating input-output maps, and then shift towards theoretical and numerical explorations of uniformly convergent systems for time series modeling and system identification. In Chapter 2, we first introduce the background to support ensuing development. We then detail our studies on quantum implementation of RCs:

- Chapter 3: we develop a theory of learning nonlinear deterministic input–output maps with fading memory by dissipative quantum systems. The theory identifies the properties required for a class of dissipative quantum systems to be universal, in that any input–output map with fading memory can be approximated arbitrarily closely and uniformly over all inputs by a QRC of this class. We then introduce an example universal class of dissipative quantum systems implemented by Ising Hamiltonians. Numerical experiments illustrate that with a small number of qubits, this class can achieve comparable performance to classical learning schemes with a large number of tunable parameters.
- Chapter 4: the universal class of QRCs in Chapter 3 are suitable for ensemble systems such as NMRs. In this chapter. we propose a universal class of universal QRCs that are suitable for current NISQ machines. Our proof-of-principle experiments on cloud-based superconducting quantum computers demonstrate that small and noisy quantum reservoirs can tackle high-order nonlinear temporal tasks. Our theoretical and experimental results pave the path for attractive temporal processing applications of near-term gate-model quantum computers of increasing fidelity but without quantum error correction, signifying the potential of these devices for wider applications including neural modeling, speech recognition, and natural language processing, going beyond static classification and regression tasks.

We then abstract away from specific hardware implementation to higher-level theoretical explorations of interconnected uniformly convergent systems as RCs for temporal information processing, expanding the toolbox of available nonlinear models in the literature. In particular,

- Chapter 5: we introduce interconnected RCs as candidate nonlinear models with closed-loop structures for system identification. To model a system based solely on input-output data, the candidate model with closed-loop structures should be uniformly convergent, so that it asymptotically forgets its initial condition. We derive a small-gain theorem to ensure the interconnected RCs have this property. We numerically demonstrate the use of interconnected ESNs and QRCs to model a feedback-controlled nonlinear system.

- Chapter 6: we travel from a deterministic setting to a probabilistic setting and show that RCs with uniformly convergent output feedback dynamics implement NARX(∞) models. We further establish the asymptotic stationarity and (Birkhoff-Khinchin) ergodicity of these NARX(∞) models. We highlight the versatility of this approach by employing classical and quantum RCs to model synthetic and real data sets, further exploring their potential for stochastic modeling applications.

Finally, we arrive at the top of our 'stack' in Fig. 1.1 and consider the problem of reservoir design:

- Chapter 7: we develop an algorithm to optimize the RC internal parameters under state measurement noise. The algorithm is based on the SPSA algorithm [92] and it inherits SPSA's efficiency by using only four noisy state measurements to approximate the gradient, regardless of the dimension of the optimization problem. We show that even with noisy state measurements, the reservoir internal parameters asymptotically and almost surely converge to a KT point. Numerical examples demonstrate that the proposed algorithm can mitigate the negative effect of noise.

Chapter 2

Preliminaries

This chapter develops the preliminaries required for the development of Chapters 3–7. In Sec. 2.1, the notation used throughout this thesis is introduced. Sec. 2.2 gives an overview of closed quantum systems and dissipative quantum systems. In Sec. 2.3, we introduce the basics of quantum circuits and quantum gates. Sec. 2.4.1 formally defines the uniform convergence property, the key property used throughout this thesis. Sec. 2.4.2 states sufficient conditions to guarantee the uniform convergence property for both quantum and classical dynamical systems. Finally, Sec. 2.4.3 formally introduces the unique filters and functionals a uniformly convergent system induces.

2.1 Notation

The notation used throughout this thesis is fairly standard, a summary can also be found on pages iii-iv. The set of real (complex) numbers is denoted by $\mathbb{R}$ ($\mathbb{C}$). The imaginary number is denoted by $\iota = \sqrt{-1}$. The set of integers is denoted by $\mathbb{Z}$, $\mathbb{Z}_- = \{..., -1, 0\}$ and $\mathbb{Z}_{\geq k_0}$ denotes integers greater than or equal to $k_0 \in \mathbb{Z}$. Elements of $\mathbb{R}^n$ are real column vectors. For any $x \in \mathbb{R}^n$, $x^\top$ is its transpose and $\|x\| = \sqrt{x^\top x}$ is its Euclidean norm. For any $x_1 \in \mathbb{R}^{n_1}, x_2 \in \mathbb{R}^{n_2}$, $(x_1, x_2) \in \mathbb{R}^{n_1+n_2}$ is their concatenation into a column vector.

The set of $n \times m$ real (complex) matrices is $\mathbb{R}^{n\times m}$ ($\mathbb{C}^{n\times m}$). A positive (semi-)definite matrix A is denoted by $A \succ 0$ ($A \succeq 0$). A negative (semi-)definite matrix A is denoted by $A \prec 0$ ($A \preceq 0$). For any $A \in \mathbb{C}^{n\times m}$, $A^\top$ is its transpose, $A^\dagger$ is its conjugate transpose and $\|A\|_p = \mathrm{Tr}(\sqrt{A^\dagger A}^p)^{1/p}$ denotes the Schatten-p norm, where $\mathrm{Tr}(\cdot)$ denotes the trace. A square complex matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian or self-adjoint if $A = A^\dagger$. Hermitian matrices are ubiquitous in quantum mechanics. The symbol $\otimes$ denotes a tensor product between Hilbert spaces or an algebraic tensor product between operators. Suppose that $\rho \in H_A \otimes H_B$, where H_A and H_B are Hilbert spaces. We let $\mathrm{Tr}_{H_B}(\rho)$ denote the partial trace of ρ over H_B. The partial trace is a generalization of the trace and is an operator-valued function. If $\rho = \sum_m \beta_m A_m \otimes B_m$, then $\mathrm{Tr}_{H_B}(\rho) = \sum_m \beta_m \mathrm{Tr}(B_m) A_m$.

The set of infinite sequences is denoted by $(\mathbb{R}^n)^{\mathbb{Z}}$, that is, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ with $u = \{u_k\}_{k\in\mathbb{Z}}$ and $u_k \in \mathbb{R}^n$ for all $k \in \mathbb{Z}$. The set of bounded infinite sequences is denoted by l_n^∞, i.e., $u \in l_n^\infty$ if $u = \{u_k\}_{k\in\mathbb{Z}}$ with $u_k \in \mathbb{R}^n$ and $\|u\|_\infty := \sup_{k\in\mathbb{Z}} \|u_k\| < \infty$. The set of left-infinite sequences is $(\mathbb{R}^n)^{\mathbb{Z}_-}$, that is, $u \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ with $u = \{u_k\}_{k\in\mathbb{Z}_-}$ and $u_k \in \mathbb{R}^n$ for all $k \in \mathbb{Z}_-$. For these aforementioned notations, we omit the superscript n for real sequences, i.e., $\mathbb{R}^{\mathbb{Z}} = (\mathbb{R}^1)^{\mathbb{Z}}, l^\infty = l_1^\infty$ and $\mathbb{R}^{\mathbb{Z}_-} = (\mathbb{R}^1)^{\mathbb{Z}_-}$. In the quantum setting, we often restrict ourselves to infinite (left-infinite) sequences $D^{\mathbb{Z}}$ ($D^{\mathbb{Z}_-}$) taking values in a compact set $D \subset \mathbb{R}^n$. $P_n^{\mathbb{Z}_-} : (\mathbb{R}^n)^{\mathbb{Z}} \to (\mathbb{R}^n)^{\mathbb{Z}_-}$ is the canonical projection (we write $P_1^{\mathbb{Z}_-} = P^{\mathbb{Z}_-}$). For any $\tau \in \mathbb{Z}$, $z_n^{-\tau}$ is the time shift operator such that for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $k \in \mathbb{Z}$, $z_n^{-\tau}(u)|_k := u_{k-\tau}$ (we write $z_1^{-\tau} = z^{-\tau}$). Finally, $\circ$ denotes function composition.

2.2 Quantum systems

Quantum systems are governed by the basic postulates of quantum mechanics. These postulates provide a description of nature at the atomic and subatmoic scale. This section provides an overview of quantum systems, see [43, 93] for an in-depth introduction.

2.2.1 Closed quantum systems

The state space of any isolated quantum system is described by a Hilbert space (a complex vector space equipped with inner product). A closed quantum system is completely described by its state vector [43, Chapter 2.2.1].

The quantum bit (qubit) is a fundamental unit in quantum systems. A qubit has a two-dimensional state space with $|0\rangle$ and $|1\rangle$ forming an orthonormal basis for that state space. An arbitrary one-qubit state vector can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. Intuitively, $|0\rangle$ and $|1\rangle$ act as the values 0 and 1 a calssical bit can take. The important difference here is that superpositions of $|0\rangle$ and $|1\rangle$, of the form $\alpha|0\rangle + \beta|1\rangle$, can exist. Here, $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. Physically, this means that when we measure the qubit in the Pauli-Z basis (see (2.1)), we get either $|0\rangle$ with probability $|\alpha|^2$ or $|1\rangle$ with probability $|\beta|^2$.

For a multi-qubit system, the state space is the tensor product of the state spaces of the constituting systems. That is, an N-qubit system has a 2^N-dimensional state space. An important example is a two-qubit system with four orthonormal basis states $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$. Here, $|00\rangle = |0\rangle \otimes |0\rangle$ and analogously for $|01\rangle, |10\rangle$ and $|11\rangle$. These basis states are examples of product states or separable states. Let H_1, H_2 be the Hilbert spaces of two systems. We say that a state $|\psi\rangle \in H_1 \otimes H_2$ on the composite system is a product state if $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, where $|\psi_i\rangle \in H_i$ for $i = 1, 2$. A pair of qubits can be in any superpositions of the four basis states and a notable example is the Bell state or EPR pair $\frac{|00\rangle + |11\rangle}{\sqrt{2}}$. The Bell state is an entangled state and is the key ingredient in quantum teleportation and superdense coding, see e.g. [43, Chapter 1.3.7 and Chapter 2.3]. In general, we say that a state $|\psi\rangle \in H_1 \otimes H_2$ is entangled if it cannot be written as a product state.

The evolution of a closed quantum system is described by a unitary transformation. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at time t_1 by a unitary operator U which depends only on the times t_1 and t_2, given by $|\psi'\rangle = U|\psi\rangle$. Here, $U^\dagger U = UU^\dagger = I$, where I is the identity matrix (of suitable size). For NMR [54, 56] and D-Wave systems [51], the unitary evolution often takes the form $U = e^{-\iota H(t_2 - t_1)}$, where H is an Ising Hamiltonian; see Chapter 3 for an example. For quantum computers, not all unitary operators are easy to implement. Fortunately, combinations of single-qubit and two-qubit unitary operations that can implement any arbitrary unitary operation [43, Chapter 4]. This will be the main topic of Sec. 2.3.

We consider projective measurements. A projective measurement is described by an observable M, a Hermitian matrix defined on the state space of the quantum system being observed. By the spectral theorem, we can decompose $M = \sum_m \lambda_m P_m$, where P_m is the projector onto the eigenspace of M with eigenvalue λ_m. The eigenvalues λ_m correspond to the possible outcomes of the projective measurement. Given a quantum state $|\psi\rangle$, the probability of measuring outcome λ_m is $\mathbb{P}(\lambda_m) = \mathrm{Tr}(P_m|\psi\rangle\langle\psi|)$, where $\langle\psi| = |\psi\rangle^\dagger$. The expectation value of M in the state $|\psi\rangle$ is $\langle M\rangle = \mathrm{Tr}(M|\psi\rangle\langle\psi|)$.

Projective measurements are usually performed in the Pauli matrices Z, X, Y bases, where

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 & -\iota \\ \iota & 0 \end{bmatrix}. \tag{2.1}$$

Given a N-qubit state $|\psi\rangle$, the expectation value by measuring the i-th qubit in the Z basis is $\langle Z^{(i)}\rangle = \mathrm{Tr}(Z^{(i)}|\psi\rangle\langle\psi|)$, where $Z^{(i)}$ acts on the i-th qubit as Z and acts on all other qubits as the identity I.

So far, we have discussed the dynamics of closed quantum systems. Although fancinating ideas can be achieved in such ideal systems, these ideas are tempered by noise in real systems. It is crucial to understand real quantum systems under noise in order to develop useful applications of them. This is the central topic for the following subsection.

2.2.2 Dissipative quantum systems

Dissipation is a ubiquitous phenomenon in real physical systems. A swinging pendulum interacts with its environment through friction, resulting in dampened oscillations. Similarly, quantum systems, and especially quantum computers, suffer from unwanted noise resulting from interactions with the outside environment. A prominent example of noise is cross-talk in superconducting quantum computers, which arises from unwanted interactions between the qubits and from leakage of the control signals onto qubits that are not part of the desired operation [94].

The state of a N-qubit dissipative quantum system is described by a density operator, a Hermitian matrix $\rho \in \mathbb{C}^{2^N \times 2^N}$ with $\mathrm{Tr}(\rho) = 1$ and $\rho \succeq 0$. We write $\mathcal{D}(2^N)$ to denote the set of $2^N \times 2^N$ density operators. We remark that $\mathcal{D}(2^N)$ is a compact convex set. In this thesis, we consider evolutions of a dissipative quantum system governed by completely positive trace-preserving (CPTP) maps T (also referred to as quantum channels). We remark that not all dissipative evolutions can be described by CPTP maps. A CPTP map sends a set of density operators $\mathcal{D}(2^N)$ on the Hilbert space $H_1 = \mathbb{C}^{2^N \times 2^N}$ to density operators $\mathcal{D}(2^{N'})$ on the Hilbert space $\mathbb{C}^{2^{N'} \times 2^{N'}}$ (N' may or may not be the same as N). A CPTP map preserves the physical properties of a density operator. That is, for any $\rho \in \mathcal{D}(2^N)$, $T(\rho)$ is again Hermitian, $\mathrm{Tr}[T(\rho)] = 1$ and $T(\rho) \succeq 0$. Furthermore, T is completely positive, meaning that if we introduce an extra system on a Hilbert space H_2 with arbitrary finite dimension, then $I \otimes T(A) \succ 0$ for any positive operator $A \succ 0$ defined on the composite system $H_2 \otimes H_1$, where I is the identity operator on H_2. The complete positivity property ensures that if $\rho_{H_2 \otimes H_1}$ is a density operator of the composite system $H_2 \otimes H_1$, and if T only acts on H_1, then $I \otimes T(\rho_{H_2 \otimes H_1})$ is also a valid density operator of the composite system; also see [43, Box 8.2].

An important example of a CPTP map is unitary transformation $T(\rho) = U\rho U^\dagger$ for some unitary matrix U. However, in general, $T(\rho)$ may not be related to ρ by a unitary transformation. A natural way to view dissipative quantum system is through an interaction of a principal system with an environment. Let ρ be the density operator of the principal system and ρ_{env} be the density operator describing an environment. In this view, an example of a CPTP map is $T(\rho) = \mathrm{Tr}_{\mathrm{env}}[U(\rho \otimes \rho_{\mathrm{env}})U^\dagger]$, where U acts on both the system and the environment, and $\mathrm{Tr}_{\mathrm{env}}[\cdot]$ denotes the partial trace over the environment. CPTP maps of this form will be utilized in Chapter 3 in quantum reservoir computers (QRCs) for temporal information processing. Another example of a CPTP map is a convex combination of unitary transformations $T(\rho) = aU_0\rho U_0^\dagger + bU_1\rho U_1^\dagger + c\sigma$, where $a, b, c \in [0, 1]$ and

$a+b+c=1$. Here, U_0, U_1 are two unitary matrices and σ is a fixed density operator. We will exploit CPTP maps of this form in Chapter 4 to implement QRCs on current noisy quantum computers.

Projective measurements for dissipative quantum systems are performed similarly as for closed quantum systems. A projective measurement is described by an observable $M = \sum_m \lambda_m P_m$. Given a quantum state ρ, the probability of measuring outcome λ_m is given by $\mathbb{P}(\lambda_m) = \mathrm{Tr}(\rho P_m)$ and the expectation value of M of the state ρ is $\langle M \rangle = \mathrm{Tr}(M\rho)$.

A state ρ of a composite system on a Hilbert space $H_1 \otimes H_2$ is called a product state if $\rho = \rho_1 \otimes \rho_2$, where ρ_i is a density operator on H_i for $i = 1, 2$. If $\rho = \sum_k p_k \rho_1^k \otimes \rho_2^k$ where $p_k \geq 1$, $\sum_k p_k = 1$ and ρ_i^k are density operators of H_i, then ρ is separable. We say that ρ is entangled if ρ is not separable.

2.2.3 Quantum noise

We have discussed the dynamics of dissipative quantum systems, quantum systems that are impacted noise. Here, we introduce some important examples of quantum noise, including depolarizing noise, dephasing noise, decaying noise and the generalized amplitude damping (GAD) noise [43, Chapter 8]. These quantum noise examples are also examples of CPTP maps.

The depolarizing noise replaces an N-qubit quantum state by the completely mixed state $I/2^N$ with probability p and leaves the state unchanged with probability $1-p$. That is, it depolarizes the state with probability p. A state ρ under the depolarizing noise undergoes the evolution $T_{\text{depolarize}}(\rho) = p\frac{I}{2^N} + (1-p)\rho$. The depolarizing noise has a unique fixed state $I/2^N$ such that $T_{\text{depolarize}}(I/2^N) = I/2^N$. This means that the depolarizing noise causes any state to become the completely mixed state with time.

The dephasing noise describes the loss of quantum information without loss of energy and is a noise process unique to quantum systems. A physical example of the dephasing noise is when a photon scatters randomly as it travels through a waveguide. A single-qubit quantum state ρ under the dephasing noise is described by $T_{\text{dephasing}}(\rho) = p\rho + (1-p)Z\rho Z$. For an N-qubit quantum state ρ, we apply the dephasing nooise on each qubit as a simple model to approximate its effect. That is, $T_{\text{dephasing}}(\rho) = T^{N}_{\text{dephasing}} \circ \cdots \circ T^{1}_{\text{dephasing}}(\rho)$, where $T^{j}_{\text{dephasing}}(\rho) = p_j\rho + (1-p_j)Z^{(j)}\rho Z^{(j)}$. The dephasing noise causes the off-diagonal elements of the density operator to decay exponentially to zero with time without impacting the diagonal elements.

The decaying noise is also referred to as the amplitude damping noise and it describes energy dissipation in a quantum system. A physical example of the decaying noise is when a spin system at high temperature approaches equilibrium with its environment. A single-qubit quantum state ρ under the decaying noise is governed by $T_{\text{decaying}}(\rho) = E_0\rho E_0^{\dagger} + E_1\rho E_1^{\dagger}$, where

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \text{ and } E_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}.$$

The operators E_0, E_1 are referred to as Kraus operators. The decaying noise has a unique fixed state $|0\rangle\langle 0|$, meaning that it relaxes any state to the ground state (i.e., corresponds to zero energy or zero temperature) with time. This is different from the dephasing noise in that the decaying noise impacts both the diagonal and off-diagonal elements of the density operator.

The GAD noise is a generalization of the decaying noise and it describes the effect of energy dissipation to an environment at finite temperature. A single-qubit state ρ under the GAD noise is described by $\rho = \sum_{i=0}^{3} M_i\rho M_i^{\dagger}$, where $M_i = M_i(p,\lambda)$ are the Kraus operators parameterized by a finite temperature parameter $\lambda \in [0,1]$ and a decaying rate $p \in [0,1]$. Here,

$$M_0 = \sqrt{\lambda}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, M_2 = \sqrt{\lambda}\begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix},$$
$$M_3 = \sqrt{1-\lambda}\begin{bmatrix} \sqrt{1-p} & 0 \\ 0 & 1 \end{bmatrix}, M_4 = \sqrt{1-\lambda}\begin{bmatrix} 0 & 0 \\ \sqrt{p} & 0 \end{bmatrix}.$$

When $\lambda = 1$, the GAD noise is the same as the decaying noise. The GAD noise also impacts both the diagonal and off-diagonal elements of the density operator. As for the dephasing noise, for an N-qubit quantum system, we apply the decaying and GAD noise to each qubit as a simple model to approximate their effects.

2.3 Quantum computing

Quantum computing utilizes intrinsic quantum mechanical effects, such as superposition and entanglement introduced in the previous subsection, to perform quantum computation. For an in-depth introduction to quantum computing, we refer to the excellent text [43]. In this section, we provide an overview of quantum gates and quantum circuits.

2.3.1 Quantum gates and quantum circuits

Most quantum computers implement a set of single-qubit and two-qubit gates that together can realize any arbitrary unitary operation. The most general single-qubit gate is described by the unitary operator

$$U_3(\theta, \phi, \lambda) = \begin{bmatrix} \cos(\theta/2) & -e^{-\iota\lambda}\sin(\theta/2) \\ e^{\iota\phi}\sin(\theta/2) & e^{\iota(\phi+\lambda)/2}\cos(\theta/2) \end{bmatrix}, \tag{2.2}$$

parameterized by $\theta \in [-2\pi, 2\pi], \phi \in [-2\pi, 2\pi]$ and $\lambda \in [-2\pi, 2\pi]$ [95]. Other commonly used single-qubit gates include rotational-X gates $R_X(\theta) = e^{-\iota\theta X}$, rotational-$Y$ gates $R_Y(\theta) = e^{-\iota\theta Y}$ and rotational-Z gates $R_Z(\theta) = e^{-\iota\theta Z}$, where X, Y, Z are the single-qubit Pauli operators defined in (2.1).

The $U_3(\theta, \phi, \lambda)$ gate and controlled-NOT (CNOT) gate together can implement any arbitrary multi-qubit unitary operation [43, Sec. 4.5.2]. CNOT is a two-qubit gate described by the unitary operator

$$\mathrm{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.3}$$

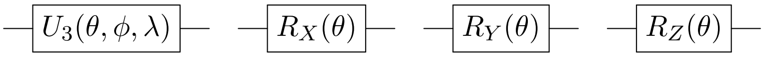


Figure 2.1: From left to right are the quantum circuit diagrams of $U_3(\theta, \phi, \lambda), R_X(\theta), R_Y(\theta)$ and $R_Z(\theta)$.

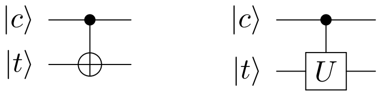


Figure 2.2: From left to right are the quantum circuit diagrams of CNOT and controlled-U.

CNOT operates on two qubits, a control qubit $|c\rangle$ and a target qubit $|t\rangle$. The action of CNOT is given by $|c\rangle \otimes |t\rangle \rightarrow |c\rangle \otimes |t \oplus c\rangle$, where $\oplus$ is modulo two addition. More generally, given an arbitrary single-qubit unitary operation U, a controlled-U acts on a control qubit $|c\rangle$ and a target qubit $|t\rangle$ via $|c\rangle \otimes |t\rangle \rightarrow |c\rangle \otimes U|t\rangle$ when $|c\rangle = |1\rangle$, otherwise $|c\rangle \otimes |t\rangle \rightarrow |c\rangle \otimes |t\rangle$.

In quantum computing, gates are often represented as quantum circuit diagrams. The circuit diagrams of $U_3(\theta, \phi, \lambda), R_X(\theta), R_Y(\theta)$ and $R_Z(\theta)$ are shown in Fig. 2.1. Here, the single line represents a qubit and the square box represents a single-qubit gate. The circuit diagrams of CNOT and controlled-U gates are shown in Fig. 2.2. Here, the top line represents the control qubit and the bottom line represents the target qubit.

We have mainly focused on providing an overview of quantum systems and quantum computing. As the central theme of this thesis is temporal information processing on both quantum and classical devices, we now turn to discussions on nonlinear dynamical systems (including classical and quantum systems) in the context of reservoir computing.

2.4 Reservoir computing

Reservoir computing (RC) exploits nonlinear dynamical systems for temporal information processing, processing input sequences to produce output sequences that approximate some target outputs. In this thesis, we focus on single-output reservoir computers (also abbreviated as RCs) to approximate real-valued output sequences. A straightforward extension for multiple outputs will be discussed shortly. Consider a discrete-time state space model

$$\begin{cases} x_{k+1} = f(x_k, u_k), \\ \hat{y}_k = h(x_k), \end{cases} \tag{2.4}$$

for all $k \in \mathbb{Z}$. Here, $u_k \in \mathbb{R}^n$ is the input and $\hat{y}_k \in \mathbb{R}$ is the RC output. The internal state for a classical RC is denoted by $x_k \in \mathbb{R}^N$ with $f : \mathbb{R}^N \times \mathbb{R}^n \to \mathbb{R}^N$ and $h : \mathbb{R}^N \to \mathbb{R}$. For a QRC, $x_k = \rho_k \in \mathcal{D}(2^N)$ and we write (2.4) as

$$\begin{cases} \rho_{k+1} = f(\rho_k, u_k), \\ \hat{y}_k = h(\rho_k), \end{cases} \tag{2.5}$$

where $f : \mathcal{D}(2^N) \times \mathbb{R}^n \to \mathcal{D}(2^N)$ and $h : \mathcal{D}(2^N) \to \mathbb{R}$.

The reservoir dynamics $f = f_\gamma$ is often parametrized by a parameter $\gamma \in \mathbb{R}^p$ (p is some positive integer), chosen according to the task. In Chapters 3–6, we consider a fixed f_γ by choosing elements of γ and fixing them at the onset. In Chapter 7, we consider a more general scenario in which γ can be optimized according to some criterion, a problem known as the reservoir design problem [81, 96]. The readout parameters $W \in \mathbb{R}^m$ in $h = h_W$ are optimized to match the target outputs.

We require the computations performed by an RC to be *independent* of its initial condition. The key to achieve this is the uniform convergence property, which roughly speaking, ensures that an RC asymptotically forgets its initial condition; see Sec. 2.4.1 for further details. In practice, given an input sequence $\{u_0, \ldots, u_L\}$ and a target output sequence $\{y_0, \ldots, y_L\}$, we washout the effect of the initial condition of a uniformly convergent RC using the first L_w data. We then optimize the readout parameters (or output weights) W via least squares by minimizing $\frac{1}{L_T} \sum_{k=L_w}^{L_t} |y_k - \hat{y}_k|^2$, where $L_T = L_t - L_w$. The rest of the input-output data are used to assess the estimation quality. Although we focus on single-output RCs here, it is a straightforward extension for multiple output channels. If the target output is $y_k \in \mathbb{R}^{n'}$, we employ n' sets of output weights for a single RC, with each set of output weights optimized to approximate each output channel.

As alluded to before, the key property required in the RC scheme is the uniform convergence property, which ensures that the RC asymptotically forgets its initial condition. We now formally define the uniform convergence property.

2.4.1 The uniform convergence property

Roughly speaking, the uniform convergence property ensures that an RC has a bounded reference state solution defined both backwards and forwards in time, and all other solutions asymptotically converge to this reference state solution, independent of the initial condition [22, 23]. The uniform convergence property is also referred to as the *echo-state property* in the RC literature [24, 97].

We will define the uniform convergence property with respect to a set of inputs $u \in K^{\mathbb{Z}}$, where $K \subseteq \mathbb{R}^n$ is a subset of $\mathbb{R}^n$. Of particular interests are the input set $K^{\mathbb{Z}}$ where K is compact and the set of bounded inputs l_n^{∞} and $(R^n)^{\mathbb{Z}}$. Let $\phi(k; k_0, \xi)$ denote a solution to (2.4) parameterized by $u \in K^{\mathbb{Z}}$, starting at time k_0 with initial condition $x_{k_0} = \xi \in \mathbb{R}^N$. That is, for all $k \geq k_0$, $\phi(k+1; k_0, \xi) = f(\phi(k; k_0, \xi), u_k)$ and $\phi(k_0; k_0, \xi) = \xi$. A function $\gamma : [0, \infty) \to [0, \infty)$ is a $\mathcal{K}$ function if it is continuous, strictly increasing and $\gamma(0) = 0$. A $\mathcal{K}$ function is a $\mathcal{K}_\infty$ function if it is unbounded. We say that $\beta : [0, \infty) \times \mathbb{Z}_+ \to \mathbb{R}$ is a $\mathcal{KL}$ function if $\beta(0, \cdot) = 0$, continuous and strictly increasing in the first argument, and non-increasing in the second argument with $\lim_{t\to\infty} \beta(s, t) = 0$ for all $s \in [0, \infty)$ [98]. As in [98, 99], we do not require $\beta \in \mathcal{KL}$ to be continuous or strictly decreasing in the second argument.

We first define the uniform convergence property for a classical RC.

Definition 2.1 (Uniform convergence property for classical RCs). *[22, 23] A classical RC described by* (2.4) *has the uniform convergence property (or is uniformly convergent) with respect to inputs $K^{\mathbb{Z}}$ if for any $u \in K^{\mathbb{Z}}$,*

(i) there exists a unique and bounded solution $x^* \in (\mathbb{R}^N)^{\mathbb{Z}}$ to (2.4) *that satisfies $x^*_{k+1} = f(x^*_k, u_k)$ for all $k \in \mathbb{Z}$ and $\sup_{k\in\mathbb{Z}} \|x^*_k\| < \infty$;*

(ii) there exists $\beta \in \mathcal{KL}$ *(independent of* u*) such that, for any* $k, k_0 \in \mathbb{Z}$ *with* $k \geq k_0$ *and any* $\xi \in \mathbb{R}^N$,

$$\|x_k^* - \phi(k; k_0, \xi)\| \leq \beta(\|x_{k_0}^* - \xi\|, k - k_0). \tag{2.6}$$

We remark that the uniform convergence property is a property of the reservoir dynamics f. The unique and bounded solution x^* in Definition 2.1 is called the reference state solution (determined by u and f). Equation (2.6) imposes that as $k_0 \to -\infty$, any solution $\phi(k; k_0, \xi)$ to (2.4) asymptotically converges to the reference state solution x^*, independent of its initial condition ξ. Here, 'uniform' means that for each x_{k_0}, the bound $\beta(\|x_{k_0}^* - \xi\|, k - k_0)$ in (2.6) depends on $k - k_0$ but not k_0.

The uniform convergence property for a QRC is defined analogously in terms of the Schatten-p norm $\| \cdot \|_p$.

Definition 2.2 (Uniform convergence property for QRCs). *[22, 23] An* N*-qubit QRC described by* (2.5) *has the p-uniform convergence property (or is p–uniformly convergent) with respect to inputs* $K^{\mathbb{Z}}$ *if for any* $u \in K^{\mathbb{Z}}$,

(i) there exists a unique solution $\rho^* \in \mathcal{D}(2^N)^{\mathbb{Z}}$ *to* (2.5) *that satisfies* $\rho_{k+1}^* = f(\rho_k^*, u_k)$ *for all* $k \in \mathbb{Z}$*;*

(ii) there exists $\beta \in \mathcal{KL}$ *(independent of* u*) such that, for any* $k, k_0 \in \mathbb{Z}$ *with* $k \geq k_0$ *and any density operator* $\xi \in \mathcal{D}(2^N)$,

$$\|\rho_k^* - \phi(k; k_0, \xi)\|_p \leq \beta(\|\rho_{k_0}^* - \xi\|_p, k - k_0). \tag{2.7}$$

We remark that the boundedness condition for the reference state solution ρ^* is not included in Definition 2.2. This is because the set of density operators $\mathcal{D}(2^N)$ is compact and hence the condition $\sup_{k \in \mathbb{Z}} \|\rho_k^*\|_p < \infty$ always holds. In this thesis, we are interested in quantum systems that are 2–uniformly convergent (see Chapter 3) and 1–uniformly convergent (see Chapter 4).

2.4.2 Sufficient conditions for the uniform convergence property

The uniform convergence property is a genuine condition that is not automatically satisfied for all RCs. Here, we introduce two theorems to ensure the uniform convergence property of a RC. The first theorem is useful for establishing the uniform convergence of a classical RC, and the second theorem is useful for ensuring the uniform convergence of a QRC.

Theorem 2.1. *[23, Theorem 1] A classical RC described by* (2.4) *is uniformly convergent with respect to inputs* $K^{\mathbb{Z}}$ *if there exists some* $N \times N$ *positive definite matrix* P *and some* $\theta \in (0,1)$ *(independent of* u*) such that, for any* $u \in K^{\mathbb{Z}}$,

$$\sup_{k\in\mathbb{Z}} \|f(0,u_k)\|_P < \infty, \tag{2.8}$$

and for any $x_1, x_2 \in \mathbb{R}^N$ *and any* $k \in \mathbb{Z}$,

$$\|f(x_1,u_k) - f(x_2,u_k)\|_P \leq \theta \|x_1 - x_2\|_P, \tag{2.9}$$

where $\|x\|_P := \sqrt{x^\top P x}$.

Theorem 2.2. *An* N*-qubit QRC described by* (2.5) *has the* p*–uniform convergence property with respect to* $K^{\mathbb{Z}}$ *if there exists some* $\epsilon \in [0,1)$ *(independent of* u*) such that for any* $u_k \in K$ *and any density operators* $\rho_1, \rho_2 \in \mathcal{D}(2^N)$,

$$\|f(\rho_1,u_k) - f(\rho_2,u_k)\|_p \leq \epsilon \|\rho_1 - \rho_2\|_p. \tag{2.10}$$

Proof. We first employ [23, Lemma 2] to show the existence of a solution defined and bounded on $\mathbb{Z}$. We then show its uniqueness.

Existence: [23, Lemma 2] states that if a QRC (2.5) admits a compact state-space, then there exists a (state) solution ρ^* of system (2.5) defined and bounded on $\mathbb{Z}$. Since the set of density operators forms a compact set, the existence of ρ^* follows.

Uniqueness: Suppose that ρ^* and $\hat{\rho}^*$ are two (state) solutions of (2.5) defined and bounded on $\mathbb{Z}$. From (2.10), for any $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$, we have

$$\|\rho_k^* - \hat{\rho}_k^*\|_p \leq \epsilon^{(k-k_0)} \|\rho_{k_0}^* - \hat{\rho}_{k_0}^*\|_p \leq 2\epsilon^{(k-k_0)},$$

where the last inequality follows from $\|\rho^*_{k_0} - \hat{\rho}^*_{k_0}\|_p \leq \|\rho^*_{k_0}\|_p + \|\hat{\rho}^*_{k_0}\|_p \leq 2$. Taking $k_0 \to -\infty$, we have $\|\rho^*_k - \hat{\rho}^*_k\|_p \leq 0$ and hence $\rho^*_k = \hat{\rho}^*_k$. Since this holds for any $k \in \mathbb{Z}$, it follows that the two solutions are the same.

Finally, to show that (2.7) in Definition 2.2 holds, for any initial density operator ρ, any $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$, we have

$$\begin{aligned}\|\rho^*_k - \phi(k; k_0, \rho)\|_p &= \|f(\rho^*_{k-1}, u_{k-1}) - f(\phi(k-1; k_0, \rho), u_{k-1})\|_p \\ &\leq \epsilon \|\rho^*_{k-1} - \phi(k-1; k_0, \rho)\|_p \\ &\ \ \vdots \\ &\leq \epsilon^{k-k_0} \|\rho^*_{k_0} - \rho\|_p.\end{aligned} \tag{2.11}$$

Hence, the p–uniform convergence property follows. □

2.4.3 Filters and functionals

A class of input-output (I/O) maps, mapping from $u \in K^{\mathbb{Z}}$ to $\hat{y} \in \mathbb{R}^{\mathbb{Z}}$, can be obtained from RCs described by (2.4) or QRCs described by (2.5) satisfying the uniform convergence property with respect to $K^{\mathbb{Z}}$. In particular, a uniformly convergent RC or QRC induces a unique time-invariant and causal filter $M_{f,h} : K^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ that depends on the reservoir dynamics f and the readout function h. When evaluated at any time $k \in \mathbb{Z}$, $\hat{y}_k = M_{f,h}(u)|_k := h(x^*_k)$, where x^* is the reference state solution to (2.4); also see [25,27]. For a QRC, we write $\hat{y}_k = M_{f,h}(u)|_k := h(\rho^*_k)$, where ρ^* is the reference state solution to (2.5).

If h is uniformly continuous, this filter can be constructed by taking the initial time $k_0 \to -\infty$ as follows.

Theorem 2.3. *Consider a uniformly convergent (with respect to inputs $K^{\mathbb{Z}}$) RC described by* (2.4) *or a QRC described by* (2.5). *If h is uniformly continuous, then for any $u \in K^{\mathbb{Z}}$*

and any $k \in \mathbb{Z}$*,*

$$\begin{aligned} \hat{y}_k = M_{f,h}(u)|_k &= h \circ f(x_{k-1}, u_{k-1}) \\ &= h \circ f(f(x_{k-2}, u_{k-2}), u_{k-1}) \\ &\;\;\vdots \\ &= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots), \end{aligned} \tag{2.12}$$

where the following point-wise limit

$$\begin{aligned} M_{f,h}(u)|_k &= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots) \\ &:= \lim_{k_0 \to -\infty} h \circ f(\ldots f(f(\xi, u_{k_0}), u_{k_0+1}) \ldots) \end{aligned} \tag{2.13}$$

exists and is independent of the initial condition ξ*.*

Proof. The proof for a classical RC (2.4) and a QRC (2.5) is the same. We will present the proof for a classical RC. Equation (2.12) follows from (2.4). To show that the point-wise limit (2.13) exists, fix $k \in \mathbb{Z}$ and ξ. For any $k_0 \in \mathbb{Z}$ with $k_0 \le k$, consider a solution $\phi(k; k_0, \xi)$ to (2.4). Then (2.13) can be written as

$$M_{f,h}(u)|_k = \mathcal{F}(u_{k-1}, u_{k-2}, \ldots) = \lim_{k_0 \to -\infty} h \circ \phi(k; k_0, \xi).$$

We show that $\{h \circ \phi(k; k_0, \xi)\}_{k_0 \le k}$ is a Cauchy sequence and thus (2.13) exists. Since h is uniformly continuous, it suffices to show $\{\phi(k; k_0, \xi)\}_{k_0 \le k}$ is Cauchy.

Let $k_0 \ge k_0'$ and x^* be the reference state solution to (2.4). By (2.6) in Definition 2.1, we have

$$\begin{aligned} \|\phi(k; k_0, \xi) - \phi(k; k_0', \xi)\| &\le \|x_k^* - \phi(k; k_0, \xi)\| + \|x_k^* - \phi(k; k_0', \xi)\| \\ &\le \beta(\|x_{k_0}^* - \xi\|, k - k_0) + \beta(\|x_{k_0'}^* - \xi\|, k - k_0') \\ &\le \beta(R_\xi, k - k_0) + \beta(R_\xi, k - k_0'), \end{aligned}$$

where $R_\xi = \max\{\|x_{k_0}^* - \xi\|, \|x_{k_0'}^* - \xi\|\} < \infty$. Since $\beta \in \mathcal{KL}$, for any $\epsilon > 0$ there exists $k_0^* \in \mathbb{Z}$ such that for all $k_0' \le k_0 \le k_0^*$, $\beta(R_\xi, k - k_0) + \beta(R_\xi, k - k_0') < \epsilon$. It follows that $\{\phi(k; k_0, \xi)\}_{k_0 \le k}$ is Cauchy for any fixed k and ξ.

To show that (2.13) is independent of ξ, let $\phi(k; k_0, \xi')$ be another solution to (2.4) starting at another initial condition $\xi' \neq \xi$. Mimicking the argument above gives

$$\|\phi(k; k_0, \xi) - \phi(k; k_0, \xi')\| \le \beta(\|x_{k_0}^* - \xi\|, k - k_0) + \beta(\|x_{k_0}^* - \xi'\|, k - k_0).$$

The limit (2.13) is independent of ξ now follows from taking $k_0 \to -\infty$ and uniform continuity of h. $\square$

By construction, the filter $M_{f,h}$ is time-invariant, meaning that it commutes with the time shift operator $M_{f,h} \circ z_n^{-\tau} = z^{-\tau} \circ M_{f,h}$ for any $\tau \in \mathbb{Z}$. The filter $M_{f,h}$ is also causal by construction, meaning that for any $u, v \in K^{\mathbb{Z}}$ such that $u_k = v_k$ for all $k \leq T$ where $T \in \mathbb{Z}$, we have $M_{f,h}(u)|_k = M_{f,h}(v)|_k$ for all $k \leq T$.

There is a bijective correspondence between $M_{f,h}$ and its associated functional $F_{f,h} : K^{\mathbb{Z}_-} \to \mathbb{R}$, defined as $F_{f,h}(u') \coloneqq M_{f,h}(\tilde{u}')|_0$ [25]. Here $\tilde{u}'$ are arbitrary extensions of $u' \in K^{\mathbb{Z}_-}$ to $K^{\mathbb{Z}}$. We can recover $M_{f,h}$ from $F_{f,h}$ via $M_{f,h}(u)|_k = F_{f,h}(P_n^{\mathbb{Z}_-} \circ z_n^{-k}(u))$ for any $k \in \mathbb{Z}$. This bijection will be useful for establishing properties of the filter $M_{f,h}$ by establishing certain properties of the functional $F_{f,h}$, where the latter is often simpler. For example, in Chapter 6 where we show that uniformly convergent RCs implement infinite-order nonlinear autoregressive models, we prove that the $M_{f,h}$ is measurable by showing that $F_{f,h}$ is measurable.

Chapter 3

Ising quantum reservoir computers

In this chapter, we are interested in harnessing dissipative quantum systems governed by Ising hamiltonians as QRCs (Ising QRCs) for temporal information processing. This chapter is based on the publications [100] and [101].

As alluded to in the Introduction, QRCs have been introduced to harness complex real-time dissipative quantum dynamics [54, 55]. This approach is a quantum implementation of the classical RC scheme, in which a dissipative quantum system processes an input sequence and produces an output sequence that approximates a target sequence. Here, we are interested in approximating nonlinear I/O maps with fading memory. As pointed out before, such maps can be approximated by a series expansion such as the Volterra series [102] or a universal family of classical nonlinear dynamical systems such as liquid state machines [15], ESNs [27], LRPO, and SAS systems [26]. However, a theoretical framework for the learning of nonlinear fading memory I/O maps by quantum systems is so far lacking. The main contributions of this chapter are

- The provision of a learning theory for dissipative quantum systems.
- The introduction of a universal class of QRCs governed by Ising hamiltonians for approximating nonlinear I/O maps with fading memory.

- A study of this universal class of Ising QRCs via numerical experiments.

The main tool employed here is the Stone-Weierstrass Theorem [103, Theorem 7.3.1]. This theorem has been applied to show that neural networks are universal function approximators mapping from finite-dimensional vectors to finite-dimensional vectors [104]. The work [26] uses the Stone-Weierstrass Theorem to establish a class of I/O maps induced by SAS systems is universal. In this chapter, we employ a similar approach as [26] to show that a class of I/O maps implemented by uniformly convergent QRCs is universal.

The universal class of Ising QRCs introduced in this chapter is suitable for ensemble quantum systems such as NMR [56]. When implemented on ensemble quantum systems, we can directly obtain the expectation of observables and the finite sampling error is negligible. In the next chapter, we will introduce another universal class of QRCs that are suitable for current NISQ machines, in which the finite sampling error must be taken into account.

This chapter is organized as follows. Sec. 3.1 formulates the universality theory. In Sec. 3.2, we introduce a universal class of Ising QRCs implemented by dissipative quantum systems. Sec. 3.3 numerically demonstrates the emulation performance of the proposed universal class. We also investigate the effect of input encodings and provide an in-depth comparison with the ESN. Sec. 3.4 presents a conclusion and discusses future outlook.

3.1 The universality property

Consider a QRC consisting of N qubits undergoing discrete-time dissipative evolution:

$$\rho_k = T(u_k)\rho_{k-1}, \tag{3.1}$$

for $k \in \mathbb{Z}$. Here, $\rho_k = \rho(k\tau)$ is the system density operator at time $t = k\tau$, τ is a (fixed) sampling time and $T(u_k)$ is a CPTP map for each u_k. In this chapter, we focus on real-valued inputs $u_k \in K$ for some compact set $K \subset \mathbb{R}$. The scheme can be extended to multiple input channels; see Chapters 5 and 6 for examples. In (3.1), the input sequence

$u = \{u_k\}_{k\in\mathbb{Z}}$ determines the dissipative quantum system's evolution. The output of the dissipative quantum system is in the form

$$\hat{y}_k = h(\rho_k), \tag{3.2}$$

where $h : \mathcal{D}(2^N) \to \mathbb{R}$ is a real functional of ρ_k. Equations (3.1) and (3.2) define a QRC with input sequence u and output sequence $\hat{y}$. The overall I/O map in the long time limit is in general nonlinear.

Recall from Chapter 2.4.3 that if a QRC described by (3.1) and (3.2) is uniformly convergent with respect to inputs $K^{\mathbb{Z}}$ (in this chapter, we consier 2–uniformly convergent QRCs), then it induces a unique time-invariant and causal filter $M_{T,h} : K^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$. Here, $M_{T,h}$ is determined by the input-dependent CPTP map $T(\cdot)$ and the readout function h. Furthermore, let $F_{T,h} : K^{\mathbb{Z}_-} \to \mathbb{R}$ be the corresponding functional.

Our goal is find a universal class of QRCs described by (3.1) and (3.2). As alluded before, we use the Stone-Weierstrass Theorem [103, Theorem 7.3.1] to establish universality.

Theorem 3.1. *(Stone-Weierstrass [103, Theorem 7.3.1]) Let E be a compact metric space and $C(E)$ be the set of real-valued continuous functions defined on E. If a subalgebra A of $C(E)$ contains the constant functions and separates points of E, then A is dense in $C(E)$.*

In the following, we expand on each ingredient of the Stone-Weierstrass Theorem.

3.1.1 Fading memory property

The fading memory property is a property of the time-invariant and causal filter $M_{T,h}$. We say that $M_{T,h}$ has the fading memory property with respect to a decreasing sequence $w = \{w_k\}_{k\geq 0}$, $\lim_{k\to\infty} w_k = 0$ if, for any two input sequences $u, v \in K^{\mathbb{Z}}$, $|M_{T,h}(u)|_0 - M_{T,h}(v)|_0| \to 0$ whenever $\sup_{k\in\mathbb{Z}^-} |w_{-k}(u_k - v_k)| \to 0$. In other words, if the elements of two sequences agree closely up to some recent past before time $k = 0$, then their output sequences will also be close at time $k = 0$.

Thanks to the bijection between $M_{T,h}$ and $F_{T,h}$, we can re-formulate the fading memory property as continuity of $F_{T,h}$ with respect to a weighted norm defined as follows.

Definition 3.1 (Weighted norm)**.** *For a null sequence* $w = \{w_k\}_{k\geq 0}$*, that is* $w : \mathbb{Z}_{\geq 0} \to (0, 1]$ *is decreasing and* $\lim_{k\to\infty} w_k = 0$*, define a weighted norm* $\|\cdot\|_w$ *on* $K^{\mathbb{Z}_-}$ *as* $\|u\|_w \coloneqq \sup_{k\in\mathbb{Z}^-} |u_k|\, w_{-k}$*.*

Definition 3.2 (Fading memory)**.** *A time-invariant causal filter* $M_{T,h} : K^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ *has the fading memory property with respect to a null sequence* w *if and only if its corresponding functional* $F_{T,h} : K^{\mathbb{Z}_-} \to \mathbb{R}$ *is continuous with respect to the weighted norm* $\|\cdot\|_w$*.*

To emphasize that fading memory is defined with respect to a null sequence w, we will say that $M_{T,h}$ is a w*-fading memory filter* and the corresponding functional $F_{T,h}$ is a w*-fading memory functional.* We also state the following compactness result [26, Lemma 2].

Lemma 3.1 (Compactness)**.** *For any null sequence* w*,* $K^{\mathbb{Z}_-}$ *is a compact metric space with the weighted norm* $\|\cdot\|_w$*.*

We write $(K^{\mathbb{Z}_-}, \|\cdot\|_w)$ to denote the space $K^{\mathbb{Z}_-}$ equipped with the weighted norm $\|\cdot\|_w$. This space will play the role of the compact metric space E and w-fading memory functionals will play the role of $C(E)$ in the Stone-Weierstrass Theorem 3.1.

3.1.2 Polynomial algebra

Let w be a null sequence and $F_{T,h}, F_{T',h'} : K^{\mathbb{Z}_-} \to \mathbb{R}$ be two w-fading memory functionals induced by two uniformly convergent QRCs. We can readily define their product $F_{T,h} \times F_{T',h'} : K^{\mathbb{Z}_-} \to \mathbb{R}$ and their sum $F_{T,h} + \lambda F_{T',h'} : K^{\mathbb{Z}_-} \to \mathbb{R}$ for any $\lambda \in \mathbb{R}$ as

$$(H_{T,h} \times H_{T',h'})(u) = H_{T,h}(u) \times H_{T',h'}(u), \quad (H_{T,h} + \lambda H_{T',h'})(u) = H_{T,h}(u) + \lambda H_{T',h'}(u),$$

for any $u \in K^{\mathbb{Z}_-}$.

Definition 3.3 (Polynomial algebra)**.** *Let* w *be a null sequence and* $\mathcal{F}_w$ *be a family of* w*-fading memory functionals. We say that* $\mathcal{F}_w$ *forms a polynomial algebra consisting of*

w-fading memory functionals if given any $F_{T,h}, F_{T',h'} \in \mathcal{F}_w$, we have $F_{T,h} \times F_{T',h'} \in \mathcal{F}_w$ and $F_{T,h} + \lambda F_{T',h'} \in \mathcal{F}_w$ for any $\lambda \in \mathbb{R}$.

3.1.3 The separation property

We now introduce the last ingredient in the Stone Weierstrass theorem.

Definition 3.4 (Separation property)**.** *Let w be a null sequence and $\mathcal{F}_w$ be a family of w-fading memory functionals defined on $K^{\mathbb{Z}_-}$. We say that $\mathcal{F}_w$ has the separation property or $\mathcal{F}_w$ separates points of $K^{\mathbb{Z}_-}$ if, for any $u, v \in K^{\mathbb{Z}_-}$ with $u \neq v$, there exists $F_{T,h} \in \mathcal{F}_w$ such that $F_{T,h}(u) \neq F_{T,h}(v)$.*

The separation property is a genuine condition that is often challenging to check. In this chapter, we will use a power series argument to prove the separation property for a class of fading memory functionals implemented by QRCs.

3.1.4 The universality theorem

We now state the main universality theorem as a direct consequence of the compactness Lemma 3.1 and the Stone-Weierstrass Theorem 3.1.

Theorem 3.2. *Let w be a null sequence and $\mathcal{M}_w$ be the set of w-fading memory filters. Let $\mathcal{F}_w$ be the family of corresponding w-fading memory functionals defined on $K^{\mathbb{Z}_-}$. If $\mathcal{F}_w$ forms a polynomial algebra of $C(K^{\mathbb{Z}_-}, \|\cdot\|_w)$, contains the constant functionals and separates points of $K^{\mathbb{Z}_-}$, then $\mathcal{F}_w$ is dense in $C(K^{\mathbb{Z}_-}, \|\cdot\|_w)$. That is for any w-fading memory filter M_* and any $\epsilon > 0$, there exists $M_{T,h} \in \mathcal{M}_w$ such that for all $u \in K^{\mathbb{Z}_-}$,*

$$\|M_*(u) - M_{T,h}(u)\|_\infty = \sup_{k \in \mathbb{Z}} |M_*(u)|_k - M_{T,h}(u)|_k| < \epsilon.$$

Proof. $\mathcal{F}_w$ is dense follows from Lemma 3.1 and Theorem 3.1. To prove the second part of the theorem, since $\mathcal{F}_w$ is dense in $C(K^{\mathbb{Z}_-}, \|\cdot\|_w)$, for any w-fading memory functional F_*

and any $\epsilon > 0$, there exists $F_{T,h} \in \mathcal{F}_w$ such that for all $u_- \in K^{\mathbb{Z}_-}$, $|F_*(u_-) - F_{T,h}(u_-)| < \epsilon$. For $u \in K^{\mathbb{Z}}$, notice that $P^{\mathbb{Z}_-} \circ z^{-k}(u) \in K^{\mathbb{Z}_-}$ for all $k \in \mathbb{Z}$, hence

$$|M_*(u)|_k - M_{T,h}(u)|_k| = \left| F_*(P^{\mathbb{Z}_-} \circ z^{-k}(u)) - F_{T,h}(P^{\mathbb{Z}_-} \circ z^{-k}(u)) \right| < \epsilon.$$

Since this is true for all $k \in \mathbb{Z}$, therefore, for all $u \in K^{\mathbb{Z}}$, $\|M_*(u) - M_{T,h}(u)\|_\infty < \epsilon$. □

We have introduced the universality property for approximating nonlinear I/O maps with fading memory. In the following, we construct a class of QRCs implemented by dissipative quantum systems equipped with the universality property.

3.2 A universal class of QRCs implemented by dissipative quantum systems

We now specify a class of QRCs governed by dissipative quantum dynamics that is universal in approximating fading memory I/O maps defined on $[0,1]^{\mathbb{Z}}$. The class consists of systems that are made up of N_s *non-interacting* subsystems initialized in a product state of the N_s subsystems, with subsystem K consisting of $n_K + 1$ qubits, n_K 'system' qubits and a single 'ancilla' qubit. We label the qubits of subsystem K by an index i_j^K that runs from $j = 0$ to $j = n_K$, with i_0^K labeling the ancilla qubit. The $n_K + 1$ qubits interact via the Hamiltonian

$$H_K = \sum_{j_1=0}^{n_K} \sum_{j_2=j_1+1}^{n_K} J_K^{j_1,j_2} (X^{(i_{j_1}^K)} X^{(i_{j_2}^K)} + Y^{(i_{j_1}^K)} Y^{(i_{j_2}^K)}) + \sum_{j=0}^{n_K} \alpha Z^{(i_j^K)}, \tag{3.3}$$

where $J_K^{j_1,j_2}$ and α are real-valued constants, while $X^{(i_j^K)}$, $Y^{(i_j^K)}$ and $Z^{(i_j^K)}$ are Pauli X, Y and Z operators of qubit i_j^K. The ancilla qubits for all subsystems are periodically reset at time $t = k\tau$ and prepared in the input-dependent mixed state $\rho_{i_{0,k}}^K = u_k |0\rangle\langle 0| + (1 - u_k)|1\rangle\langle 1|$ (with $u_k \in [0,1]$). The system qubits are initialized at time $t = 0$ to some density operator. The density operator ρ_k^K of the Kth subsystem qubits evolves during time $(k-1)\tau < t < k\tau$ according to $\rho_k^K = T_K(u_k)\rho_{k-1}^K$, where $T_K(u_k)$ is the CPTP map defined by $T_K(u_k)\rho_{k-1}^K = \mathrm{Tr}_{i_0^K}\left(e^{-iH_K\tau} \rho_{k-1}^K \otimes \rho_{i_{0,k}}^K e^{iH_K\tau}\right)$ and $\mathrm{Tr}_{i_0^K}$ denotes the partial trace over the ancilla qubit of subsystem K. Let $\overline{H}_K = I \otimes \cdots \otimes H_K \otimes \cdots \otimes I$ with H_k in the K-th position, the total Hamiltonian of the N_s subsystems is $H = \sum_{K=1}^{N_s} \overline{H}_K$. Writing $\rho_k = \bigotimes_{K=1}^{N_s} \rho_k^K$, the overall dynamics of the N_s non-interacting subsystem is

$$\rho_k = T(u_k)\rho_{k-1} = \bigotimes_{K=1}^{N_s} T_K(u_k)\rho_{k-1}^K. \tag{3.4}$$

We now specify an output functional h associated with this system. We will use a single index to label all system qubits from the N_s subsystems running from 1 until $n = \sum_{K=1}^{N_s} n_K$. The ancilla qubits are not used in the output. The output functional h is defined to be of the general form,

$$\begin{aligned}\hat{y}_k &= h(\rho_k) \\ &= C + \sum_{d=1}^{R} \sum_{i_1=1}^{n} \sum_{i_2=i_1+1}^{n} \cdots \sum_{i_n=i_{n-1}+1}^{n} \sum_{r_{i_1}+\cdots+r_{i_n}=d} W_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}} \langle Z^{(i_1)} \rangle_k^{r_{i_1}} \cdots \langle Z^{(i_n)} \rangle_k^{r_{i_n}}\end{aligned} \tag{3.5}$$

where C is a constant, R is an integer and $\langle Z^{(i)} \rangle_k = \mathrm{Tr}(\rho_k Z^{(i)})$ is the expectation of the operator $Z^{(i)}$. We note that the output functional h (the right hand side of the above) is a multivariate polynomial in the variables $\langle Z^{(i)} \rangle_k$ $(i = 1, \ldots, n)$. Thus, computing $\hat{y}_k$ only involves estimating the expectations $\langle Z^{(i)} \rangle_k$ and the degree of the polynomial R can be chosen as desired. If $R = 1$ then $\hat{y}_k$ is a simple linear function of the expectations.

More generally, we can choose the output functional to be a multivariate polynomial in the variables $\langle X^{(i)} \rangle_k$, $\langle Y^{(i)} \rangle_k$ and $\langle Z^{(i)} \rangle_k$ $(i = 1, \ldots, n)$. The class of QRCs with this more general output functional is also universal since this class contains members of the QRC class with the output functional (3.5). Increasing the number of observables can potentially improve the emulation ability of a QRC, at the expense of performing more measurements. Since the observables $X^{(i)}, Y^{(i)}, Z^{(i)}$ are non-commuting, quantum mechanics prevents simultaneous measurements of these observables. Therefore, to obtain $\langle X^{(i)} \rangle_k$, $\langle Y^{(i)} \rangle_k$ and $\langle Z^{(i)} \rangle_k$, we need to perform $3n$ measurements instead of n measurements. In Sec. 4.3.2 of the next chapter, we will present a method that utilizes pre-measurement single-qubit rotations to obtain the linear combination $\sum_{i=1}^{n} \sum_{M=\{X,Y,Z\}} W_i^M \langle M^{(i)} \rangle_k$ with n measurements. This means that forming the QRC prediction with a linear output $(R = 1)$ requires only n measurements instead of $3n$ measurements. For simplicity, we will consider the output functional (3.5) in this thesis. All mathematical proofs presented in this chapter and the next chapter also apply to this alternative output functional.

The class specified above with output functional 3.5 is a variant of the QRC model in [54] but is universal by the theory of the previous section. The differences are in the general form of the output and, in our model, the ancilla qubit is not used in computing the output. Also, we do not consider time-multiplexing. We remark that time-multiplexing can be in principle incorporated in the model using the same theory. However, this extension is more technical and will be pursued in a future work.

This family of QRCs exhibits three important properties. Let $H_0(2^{n_K})$ be the hyperplane of of $2^{n_K} \times 2^{n_K}$ traceless Hermitian operators. Firstly, if for each subsystem K there exists $\epsilon_K \in [0,1)$ such that for all $u_k \in [0,1]$, $\|T(u_k)|_{H_0(2^{n_K})}\|_{2-2} := \sup_{A \in H_0(2^{n_K}), A \neq 0} \frac{\|T(u_k)A\|_2}{\|A\|_2} \leq \epsilon_K$, then the QRC governed by (3.4) is 2–uniformly convergent. We only consider 2–uniformly convergent QRCs in this chapter, and will drop the prefix '2' from now on. Secondly, the family forms a polynomial algebra consisting of systems that implement fading memory maps. Lastly, uniformly convergent single-qubit QRCs with a linear output ($n = 1$, $N_s = 1$ and $R = 1$) separates points of $[0,1]^{\mathbb{Z}_-}$. These properties and an application of the Stone-Weierstrass Theorem [103, Theorem 7.3.1] guarantee the universality property for this family of QRCs.

3.2.1 Uniform convergence

We provide a sufficient condition for the QRC governed by (3.4) to be uniformly convergent.

Lemma 3.2. *An n-qubit QRC governed by* (3.4) *is uniformly convergent with respect to* $[0,1]^{\mathbb{Z}_-}$ *if, for each subsystem K there exists $\epsilon_K \in [0,1)$ such that for all $u_k \in [0,1]$,* $\|T(u_k)|_{H_0(2^{n_K})}\|_{2-2} := \sup_{A \in H_0(2^{n_K}), A \neq 0} \frac{\|T(u_k)A\|_2}{\|A\|_2} \leq \epsilon_K$.

Proof. For each of the K subsystem, let $\rho_1^K, \rho_2^K \in \mathcal{D}(2^n)$. Then $\rho_1^K - \rho_2^K$ is a traceless

Hermitian operator. For any $u_k \in K$, we have,

$$\|T(u_k)(\rho_1 - \rho_2)\|_2 \leq \|T(u_k)|_{H_0(2^{n_K})}\|_{2-2}\|\rho_1 - \rho_2\|_2 \leq \epsilon_K \|\rho_1 - \rho_2\|_2.$$

By Theorem 2.2 in Chapter 2, each of the K subsystem is uniformly convergent. Since the subsystems are non-interacting and the QRC is always initialized in a product state, hence the QRC governed by (3.4) is uniformly convergent. □

We remark that if $T_K(\cdot)$ satisfies the conditions in Lemma 3.2, then any two density operators converge uniformly to one another under $T_K(\cdot)$. This observation will be used in the following subsection.

3.2.2 Fading memory and polynomial algebra

Let $\mathcal{F} = \{F_{T,h}\}$ be the set of functionals induced by uniformly convergent QRCs governed by (3.4) and (3.5). We will show in Lemma 3.3 that the uniform convergence and continuity of $T_K(\cdot)$ are sufficient to guarantee that $F_{T,h}$ is fading memory with respect to any null sequence. In Lemma 3.5, we show that $\mathcal{F}$ forms a polynomial algebra, made of fading memory functionals. In the following, let $\mathcal{L}(\mathbb{C}^{2^n})$ be the set of linear operators on $\mathbb{C}^{2^n}$, and for a CPTP map $T(\cdot)$ and for all $u_k \in [0,1]$, define $\|T(u_k)\|_{2-2} := \sup_{A \in \mathcal{L}(\mathbb{C}^{2^n}), \|A\|_2 = 1} \|T(u_k)A\|_2$.

Lemma 3.3 (Fading memory)**.** *Consider an n-qubit QRC with dynamics* (3.4) *and output* (3.5)*. Suppose that for each subsystem K, the CPTP map $T_K(\cdot)$ satisfies the condition in Lemma 3.2, so that the QRC is uniformly convergent. Then for any null sequence w, the induced filter $M_{T,h}$ and the corresponding functional $F_{T,h}$ are w-fading memory.*

Proof. We first state the boundedness of CPTP maps [105, Theorem 2.1].

Lemma 3.4. *For any CPTP map $T : \mathcal{L}(2^{n_K}) \to \mathcal{L}(2^{n_K})$, we have $\|T\|_{2-2} \leq \sqrt{2^{n_K}}$.*

Further, $T_K(\cdot)$ defined in (3.4) is uniformly continuous. That is, for any $x, y \in [0,1]$,

$$\begin{aligned}\|T_K(x) - T_K(y)\|_{2-2} &= \sup_{\substack{A \in \mathcal{L}(\mathbb{C}^{2^n}) \\ \|A\|_2 = 1}} \|(T_K(x) - T_K(y))A\|_2 \\ &= \sup_{\substack{A \in \mathcal{L}(\mathbb{C}^{2^n}) \\ \|A\|_2 = 1}} \|\mathrm{Tr}_{i_0}^K(e^{-iH_K\tau} A \otimes (x-y) Z e^{iH_K\tau})\|_2 \\ &= |x-y| \sup_{\substack{A \in \mathcal{L}(\mathbb{C}^{2^n}) \\ \|A\|_2 = 1}} \|\mathrm{Tr}_{i_0}^K(e^{-iH_K\tau} A \otimes Z e^{iH_K\tau})\|_2 \\ &= |x-y| \|\tilde{T}_K\|_{2-2},\end{aligned} \tag{3.6}$$

where $\|\tilde{T}_K\|_{2-2} < \infty$ since $\tilde{T}_K$ is a linear operator defined on a finite dimensional normed space. Also recall that $\mathrm{Tr}(\cdot)$ is continuous, that is, for any $\epsilon > 0$, there exists $\delta_{\mathrm{Tr}}(\epsilon) > 0$ such that $|\mathrm{Tr}(A - B)| < \epsilon$ whenever $\|A - B\|_2 < \delta_{\mathrm{Tr}}(\epsilon)$ for any complex matrices A, B.

Let w be any null sequence. We will show the linear terms in the functional $F_{T,h}$ are continuous with respect to the weighted norm $\|\cdot\|_w$ defined in Definition 3.1, and the continuity property of $F_{T,h}$ follows from the fact that finite sums and products of continuous elements are also continuous.

Let $\rho_{-\infty} = \bigotimes_{K=1}^{N_s} \rho_{-\infty}^K$ and for any $u \in [0,1]^{\mathbb{Z}_-}$, let $L(u) = \mathrm{Tr}\left(Z^{(K)} \left(\overrightarrow{\prod}_{k=0}^{\infty} T(u_{-k})\right) \rho_{-\infty}\right)$ be a linear term in $F_{T,h}(u)$, where $\overrightarrow{\prod}_{k=0}^{\infty} T(u_{-k})$ is a time-ordered composition from left to right. Since the N_s subsystems are non-interacting, $L(u) = \mathrm{Tr}\left(Z^{(K)} \left(\overrightarrow{\prod}_{k=0}^{\infty} T_K(u_{-k})\right) \rho_{-\infty}^K\right)$. Now for any $u, v \in [0,1]^{\mathbb{Z}_-}$,

$$|L(u) - L(v)| = \left|\mathrm{Tr}\left(Z^{(K)} \left(\left(\overrightarrow{\prod}_{k=0}^{\infty} T_K(u_{-k})\right) \rho_{-\infty}^K - \left(\overrightarrow{\prod}_{k=0}^{\infty} T_K(v_{-k})\right) \rho_{-\infty}^K\right)\right)\right|,$$

Denote $\rho_u^K = \left(\overrightarrow{\prod}_{k=N}^{\infty} T_K(u_{-k})\right) \rho_{-\infty}^K$ and $\rho_v^K = \left(\overrightarrow{\prod}_{k=N}^{\infty} T_K(v_{-k})\right) \rho_{-\infty}^K$ for some $0 < N < \infty$, then

$$\begin{aligned}&\left\|Z^{(K)} \left(\left(\overrightarrow{\prod}_{k=0}^{\infty} T_K(u_{-k})\right) \rho_{-\infty}^K - \left(\overrightarrow{\prod}_{k=0}^{\infty} T_K(v_{-k})\right) \rho_{-\infty}^K\right)\right\|_2 \\ &\leq \left\|Z^{(K)}\right\|_2 \left(\left\|\overrightarrow{\prod}_{k=0}^{N-1} T_K(u_{-k}) - \overrightarrow{\prod}_{k=0}^{N-1} T_K(v_{-k})\right\|_{2-2} \left\|\rho_u^K\right\|_2 + \left\|\left(\overrightarrow{\prod}_{k=0}^{N-1} T_K(v_{-k})\right)(\rho_u^K - \rho_v^K)\right\|_2\right).\end{aligned} \tag{3.7}$$

Since $T_K(u_k)$ satisfies conditions in Lemma 3.2, any two density operators converge uniformly to one another. Therefore, for any $\epsilon > 0$, there exists $N(\epsilon) > 0$ such that for all

$N' > N(\epsilon)$,

$$\left\| \left(\overrightarrow{\prod}_{k=0}^{N'} T_K(v_{-k}) \right) (\rho_u^K - \rho_v^K) \right\|_2 < \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2 \left\| Z^{(K)} \right\|_2}. \tag{3.8}$$

Choose $N' = N(\epsilon) + 1$, bound the first term in the sum of (3.7) by rewriting it as a telescopic sum:

$$\begin{aligned} &\left\| \overrightarrow{\prod}_{k=0}^{N(\epsilon)} T_K(u_{-k}) - \overrightarrow{\prod}_{k=0}^{N(\epsilon)} T_K(v_{-k}) \right\|_{2-2} \\ &= \left\| \sum_{l=0}^{N(\epsilon)} \Big(T_K(v_0) \cdots T_K(v_{-(l-1)}) T_K(u_{-l}) T_K(u_{-(l+1)}) \cdots T_K(u_{-N(\epsilon)}) \right. \\ &\qquad \left. - T_K(v_0) \cdots T_K(v_{-(l-1)}) T_K(v_{-l}) T_K(u_{-(l+1)}) \cdots T_K(u_{-N(\epsilon)}) \Big) \right\|_{2-2} \\ &\le \sum_{l=0}^{N(\epsilon)} \left\| T_K(v_0) \cdots T_K(v_{-(l-1)}) \right\|_{2-2} \left\| T_K(u_{-l}) - T_K(v_{-l}) \right\|_{2-2} \left\| T_K(u_{-(l+1)}) \cdots T_K(u_{-N(\epsilon)}) \right\|_{2-2} \\ &\le 2^{n_K} \sum_{l=0}^{N(\epsilon)} \left\| T_K(u_{-l}) - T_K(v_{-l}) \right\|_{2-2}, \end{aligned} \tag{3.9}$$

where the last inequality follows from Lemma 3.4. We claim that for any $\epsilon > 0$, if

$$\|u - v\|_w = \sup_{k \in \mathbb{Z}^-} |u_k - v_k| \, w_{-k} < \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2^{n_K+1} \left\| Z^{(K)} \right\|_2 (N(\epsilon)+1) \|\tilde{T}_K\|_{2-2}} w_{N(\epsilon)}$$

then $|L(u) - L(v)| < \epsilon$. Indeed, since w is decreasing, the above condition implies that

$$\max_{0 \le l \le N(\epsilon)} |u_{-l} - v_{-l}| \, w_{N(\epsilon)} < \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2^{n_K+1} \left\| Z^{(K)} \right\|_2 (N(\epsilon)+1) \|\tilde{T}_K\|_{2-2}} w_{N(\epsilon)}.$$

Since $w_{N(\epsilon)} > 0$, for all $0 \le l \le N(\epsilon)$, $|u_{-l} - v_{-l}| < \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2^{n_K+1} \left\| Z^{(K)} \right\|_2 (N(\epsilon)+1) \|\tilde{T}_K\|_{2-2}}$. By uniform continuity of $T_K(\cdot)$ established in (3.6), we bound (3.9) by

$$2^{n_K} \sum_{l=0}^{N(\epsilon)} \left\| T_K(u_{-l}) - T_K(v_{-l}) \right\|_{2-2} < 2^{n_K} \sum_{l=0}^{N(\epsilon)} \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2^{n_K+1} \left\| Z^{(K)} \right\|_2 (N(\epsilon)+1)} = \frac{\delta_{\mathrm{Tr}}(\epsilon)}{2 \left\| Z^{(K)} \right\|_2}. \tag{3.10}$$

Since $\left\| \rho_u^K \right\|_2 \le 1$, (3.7), (3.8) and (3.10) give

$$\begin{aligned} &\left\| Z^{(K)} \right\|_2 \left(\left\| \overrightarrow{\prod}_{k=0}^{N(\epsilon)} T_K(u_{-k}) - \overrightarrow{\prod}_{k=0}^{N(\epsilon)} T_K(v_{-k}) \right\|_{2-2} \left\| \rho_u^K \right\|_2 + \left\| \left(\overrightarrow{\prod}_{k=0}^{N(\epsilon)} T_K(v_{-k}) \right) (\rho_u^K - \rho_v^K) \right\|_2 \right) \\ &< \delta_{\mathrm{Tr}}(\epsilon). \end{aligned}$$

The result now follows from the continuity of $\mathrm{Tr}(\cdot)$. □

Lemma 3.5 (Polynomial algebra)**.** *Consider a family of QRCs described by* (3.4) *and* (3.5) *satisfying the conditions in Lemma 3.2, so that the QRCs are uniformly convergent. Let* $\mathcal{F} = \{F_{T,h}\}$ *be the induced family of functionals. Then for any null sequence* w*,* $\mathcal{F}$ *forms a polynomial algebra consisting of* w*-fading memory functionals.*

Proof. Consider any two QRCs described by (3.4) and (3.5), with n_1 and n_2 system qubits respectively. Let $T^{(m)}$ be the CPTP map and $h^{(m)}$ be the readout function of QRC $m = 1, 2$. Since both QRCs satisfy the conditions in Lemma 3.2, hence they are uniformly convergent. Further, by Lemma 3.3, given any null sequence w, their induced functionals $F_{T^{(m)},h^{(m)}} \in \mathcal{F}$ are w-fading memory. We prove the result by showing that there exists uniformly convergent QRCs described by (3.4) and (3.5) satisfying the conditions in Lemma 3.2 implementing $F_{T^{(1)},h^{(1)}} \times F_{T^{(2)},h^{(2)}}$ and $F_{T^{(1)},h^{(1)}} + \lambda F_{T^{(2)},h^{(2)}}$ for any $\lambda \in \mathbb{R}$.

Let $\rho_k^{(m)} \in \mathcal{D}(\mathbb{C}^{2^{n_m}})$ be the state of the m-th QRC. Let $j_1 = 1, \ldots, n_1$ and $j_2 = 1, \ldots, n_2$ be the respective qubit indices for the two systems. For the observable $Z^{(j_m)}$ of qubit j_m,

$$\begin{aligned}
\mathrm{Tr}\left(Z^{(j_1)}\rho_k^{(1)}\right) &= \mathrm{Tr}\left((Z^{(j_1)} \otimes I)(\rho_k^{(1)} \otimes \rho_k^{(2)})\right), \\
\mathrm{Tr}\left(Z^{(j_2)}\rho_k^{(2)}\right) &= \mathrm{Tr}\left((I \otimes Z^{(j_2)})(\rho_k^{(1)} \otimes \rho_k^{(2)})\right),
\end{aligned}$$

where I is the identity operator. Therefore, we can relabel the qubit for the combined system described by the density operator $\rho_k^{(1)} \otimes \rho_k^{(2)}$ by j, running from $j = 1$ to $j = n_1 + n_2$. Using this notation, the above expectations can be re-expressed as

$$\begin{aligned}
\mathrm{Tr}\left(Z^{(j_1)}\rho_k^{(1)}\right) &= \mathrm{Tr}\left(Z^{(j)}\rho_k^{(1)} \otimes \rho_k^{(2)}\right), \quad j = j_1 \\
\mathrm{Tr}\left(Z^{(j_2)}\rho_k^{(2)}\right) &= \mathrm{Tr}\left(Z^{(j)}\rho_k^{(1)} \otimes \rho_k^{(2)}\right), \quad j = n_1 + j_2.
\end{aligned}$$

Following this idea, write out the outputs of two QRCs as follows,

$$\begin{aligned}
\hat{y}_k^{(1)} &= C_1 + \sum_{d_1=1}^{R_1} \sum_{i_1=1}^{n_1} \cdots \sum_{i_{n_1}=i_{n_1-1}+1}^{n_1} \sum_{r_{i_1}+\cdots+r_{i_{n_1}}=d_1} W_{i_1,\ldots,i_{n_1}}^{r_{i_1},\ldots,r_{i_{n_1}}} \langle Z^{(i_1)}\rangle_k^{r_{i_1}} \cdots \langle Z^{(i_{n_1})}\rangle_k^{r_{i_{n_1}}}, \\
\hat{y}_k^{(2)} &= C_2 + \sum_{d_2=1}^{R_2} \sum_{j_1=1}^{n_2} \cdots \sum_{j_{n_2}=j_{n_2-1}+1}^{n_2} \sum_{r_{j_1}+\cdots+r_{j_{n_2}}=d_2} W_{j_1,\ldots,j_{n_2}}^{r_{j_1},\ldots,r_{j_{n_2}}} \langle Z^{(j_1)}\rangle_k^{r_{j_1}} \cdots \langle Z^{(j_{n_2})}\rangle_k^{r_{j_{n_2}}}.
\end{aligned}$$

Let $n' = n_1 + n_2$ and $R = \max\{R_1, R_2\}$, then for any $\lambda \in \mathbb{R}$,

$$\hat{y}_k^{(1)} + \lambda\hat{y}_k^{(2)} = C_1 + \lambda C_2 + \sum_{d=1}^{R} \sum_{k_1=1}^{n'} \cdots \sum_{k_n=k_{n-1}+1}^{n'} \sum_{r_{k_1}+\cdots+r_{k_n}=d} \bar{W}_{k_1,\ldots,k_n}^{r_{k_1},\ldots,r_{k_n}} \langle Z^{(k_1)}\rangle_k^{r_{k_1}} \cdots \langle Z^{(k_n)}\rangle_k^{r_{k_n}},$$

where the weights $\bar{W}^{r_{k_1},\dots,r_{k_n}}_{k_1,\dots,k_n}$ are changed accordingly. For instance, if all $k_m \leq n_1$, then $\bar{W}^{r_{k_1},\dots,r_{k_n}}_{k_1,\dots,k_n} = W^{r_{i_1},\dots,r_{i_{n_1}}}_{i_1,\dots,i_{n_1}}$, corresponding to the weights in the output $\hat{y}^{(1)}_k$. Similarly, let $R = R_1 + R_2$, then

$$\hat{y}^{(1)}_k \hat{y}^{(2)}_k \quad = \quad C_1 C_2 + \sum_{d=1}^{R} \sum_{k_1=1}^{n'} \cdots \sum_{k_n=k_{n-1}+1}^{n'} \sum_{r_{k_1}+\cdots+r_{k_n}=d} \hat{W}^{r_{k_1},\dots,r_{k_n}}_{k_1,\dots,k_n} \langle Z^{(k_1)} \rangle_k^{r_{k_1}} \cdots \langle Z^{(k_n)} \rangle_k^{r_{k_n}},$$

where $\hat{W}^{r_{k_1},\dots,r_{k_n}}_{k_1,\dots,k_n}$ are changed accordingly. Therefore, $\hat{y}^{(1)}_k + \lambda \hat{y}^{(2)}_k$ and $\hat{y}^{(1)}_k \hat{y}^{(2)}_k$ again have the same form as the right hand side of the multivariate polynomial readout function (3.5).

Furthermore, the QRC implementing $\hat{y}^{(1)}_k + \lambda \hat{y}^{(2)}_k$ and $\hat{y}^{(1)}_k \hat{y}^{(2)}_k$ is governed by $T^{(1)} \otimes T^{(2)}$, again of the form (3.4). Since the two QRCs governed by $T^{(1)}$ and $T^{(2)}$ respectively are non-interacting and are always initialized in a product state, hence the QRC governed by $T^{(1)} \otimes T^{(2)}$ is uniformly convergent.

Finally, since sums and products of w-fading memory functionals is also w-fading memory, it follows that $\mathcal{F}$ forms a polynomial algebra of w-fading memory functionals. ☐

3.2.3 Separation of points

Consider a family of uniformly convergent QRCs described by (3.4) and (3.5) satisfying the conditions in Lemma 3.2. Let $\mathcal{F} = \{F_{T,h}\}$ be the induced family of functionals. To show that the family $\mathcal{F}$ separates points of $[0,1]^{\mathbb{Z}_-}$, we state the following lemma whose proof can be found in [106, Theorem 3.2].

Lemma 3.6. *Let $f(\theta) = \sum_{j=0}^{\infty} x_j \theta^j$ be a non-constant real power series, having a non-zero radius of convergence. If $f(0) = 0$, then there exists $\beta > 0$ such that $f(\theta) \neq 0$ for all θ with $|\theta| \leq \beta$ and $\theta \neq 0$.*

Lemma 3.7 (Separation of points)**.** *Consider a family of uniformly convergent QRCs described by* (3.4) *and* (3.5) *satisfying the conditions in Lemma 3.2. Their induce a family of functionals $\mathcal{F} = \{F_{T,h}\}$ separates points of* $[0,1]^{\mathbb{Z}_-}$*.*

Proof. Consider a single-qubit QRC with a single subsystem and a linear readout function ($n = 1, N_s = 1, R = 1$) governed by (3.4), interacting with a single ancilla qubit. Order an orthogonal basis of $\mathcal{L}(\mathbb{C}^2)$ as $\mathcal{B} = \{I, Z, X, Y\}$. Recall that the normal representations of a CPTP map T and a density operator ρ are given by [43]

$$\overline{T}_{i,j} = \frac{\operatorname{Tr}\left(B_i T(B_j)\right)}{2} \qquad \text{and} \qquad \overline{\rho}_i = \frac{\operatorname{Tr}(\rho B_i)}{2},$$

where $B_i \in \mathcal{B}$. Without loss of generality, let $\tau = 1$ and set $J_1^{j_1,j_2} = J \in \mathbb{R}$ for all j_1, j_2 in the Hamiltonian given by (3.3). We obtain the normal representation of the CPTP map

$$\overline{T}(u_k) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sin^2(2J)(2u_k - 1) & \cos^2(2J) & 0 & 0 \\ 0 & 0 & \cos(2J)\cos(2\alpha) & -\cos(2J)\sin(2\alpha) \\ 0 & 0 & \cos(2J)\sin(2\alpha) & \cos(2J)\cos(2\alpha) \end{pmatrix}.$$

When restricted to the hyperplane of traceless Hermitian operators,

$$\overline{T}|_{H_0(2)} = \begin{pmatrix} \cos^2(2J) & 0 & 0 \\ 0 & \cos(2J)\cos(2\alpha) & -\cos(2J)\sin(2\alpha) \\ 0 & \cos(2J)\sin(2\alpha) & \cos(2J)\cos(2\alpha) \end{pmatrix}$$

with $\left\|\overline{T}|_{H_0(2)}\right\|_{2-2} = \sigma_{\max}(\overline{T}|_{H_0(2)}) = |\cos(2J)|$. Here, $\|\cdot\|_{2-2}$ is the matrix 2-norm and $\sigma_{\max}(\cdot)$ is the maximum singular value. Choose $J \neq \frac{z\pi}{2}$ for $z \in \mathbb{Z}$, then $|\cos(2J)| \leq 1 - \epsilon$ for some $\epsilon \in (0, 1]$. By Lemma 3.2, this QRC is uniformly convergent and we choose an arbitrary initial density operator $\overline{\rho}_{-\infty} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \end{pmatrix}^T$, corresponding to $\rho_{-\infty} = |0\rangle\langle 0|$. Only taking the expectation $\langle Z \rangle$ in the output (3.5) by setting the degree $R = 1$, then this QRC induces a functional $F_{T,h}(u) = W\left[\left(\overrightarrow{\prod}_{j=0}^{\infty} \overline{T}(u_{-j})\right)\overline{\rho}_{-\infty}\right]_2 + C$, for all $u \in [0,1]^{\mathbb{Z}_-}$. Here, $W \in \mathbb{R}$ is the readout parameter and $[\cdot]_2$ refers to the second element of the vector, corresponding to $\langle Z \rangle$ given the order of the orthogonal basis elements in $\mathcal{B}$. Choose $W \neq 0$. Given two input sequences $u \neq v$ in $[0,1]^{\mathbb{Z}_-}$, consider two cases:
(i) If $u_0 \neq v_0$, choose $J = \frac{\pi}{4}$ such that $\cos^2(2J) = 0$ and $\sin^2(2J) = 1$. Then

$$F_{T,h}(u) - F_{T,h}(v) = W(u_0 - v_0) \neq 0.$$

(ii) If $u_0 = v_0$, we have $F_{T,h}(u) - F_{T,h}(v) = W \sin^2(2J) \sum_{j=0}^{\infty} \left(\cos^2(2J)\right)^j (u_{-j} - v_{-j})$.

Let $\theta = \cos^2(2J)$, then given our choice of J, we have $0 \leq \theta \leq 1 - \epsilon$ and $\sin^2(2J) \geq \epsilon$ for some $\epsilon > 0$. Consider the power series $f(\theta) = \sum_{j=0}^{\infty} \theta^j (u_{-j} - v_{-j})$, since $|u_{-j} - v_{-j}| \leq 1$,

$f(\theta)$ has a non-zero radius of convergence R such that $(-1, 1) \subseteq R$. Moreover, $f(\theta)$ is non-constant and $f(0) = 0$. The separation of points follows from invoking Lemma 3.6. $\square$

3.3 Numerical experiments

We demonstrate the performance of the universal class introduced above. We focus on single subsystem ($N = 1$) with system qubits $n = \{2, 3, 4, 5, 6\}$, and denote this subclass as SA. We will drop the subsystem index K from now on. We introduce $S > 0$ such that the Hamiltonian parameters $J^{j_1,j_2}/S$, $\alpha/S = 0.5$ and $\tau S = 1$ in (3.3) are dimensionless. As in [54], we randomly choose $J^{j_1,j_2}/S$ from $[-1, 1]$ and test the uniform convergence property by checking if 50 initial density operators converge in 500 timesteps under an input. Input $u = \{u_k\}$, with each u_k randomly uniformly chosen from $[0, 0.2]$, is applied to all tasks. We apply multitasking, in which the expectations $\langle Z^{(i)} \rangle_k$ for all k are recorded once, while the output weights in (3.5) are optimized independently for each task.

Three tasks are presented here. The first task is LRPO [26, 102],

$$\begin{cases} x_k = Ax_{k-1} + cu_k \\ y_k = \hat{h}(x_k), \end{cases}$$

where $\hat{h}$ is chosen to be a degree two multivariate polynomial, whose coefficients are randomly chosen from $[-0.1, 0.1]$. We set A to be a diagonal block matrix $A = \mathrm{diag}(A_1, A_2, A_3)$, where A_1, A_2 and A_3 are 200×200, 500×500 and 700×700 real matrices, respectively. In this setting, each linear reservoir defined by A_i evolves independently, while the output of the LRPO depends on all state elements $x_k \in \mathbb{R}^{1400}$. Elements of A_i ($i = 1, 2, 3$) are randomly chosen from $[0, 4]$. After randomly choosing elements of A_i, let $\sigma_{\max}(A_i)$ be the maximum singular value of A_i. We linearly re-scale each element of A_i by $A_i' = A_i/\sigma_{\max}(A_i) \times 0.9$ and use these re-scaled A_i' in the numerical examples. This ensures the uniform convergence and fading memory properties of the LRPO system [26].

The next two tasks test SA to emulate systems that do not strictly implement fading memory maps as defined here. We simulate a missile [107] using ‘ode45’ in MATLAB [108],

$$\begin{cases} \dot{x}_1 = x_2 - 0.1\cos(x_1)(5x_1 - 4x_1^3 + x_1^5) - 0.5\cos(x_1)\tilde{u} \\ \dot{x}_2 = -65x_1 + 50x_1^3 - 15x_1^5 - x_2 - 100\tilde{u} \end{cases}$$

where $y = x_2$ is the output, with a sampling time of $4 \times 10^{-4}s$ for $1s$ and zero initial condition. We make a change of variable $\tilde{u} = 5u - 0.5$ so that the input range is the same as in [107]. The NARMA models are often used to benchmark learning algorithms. NARMA models depend on its time-lagged outputs and inputs, specified by a delay τ_{NARMA},

$$y_k = 0.3y_{k-1} + 0.05y_{k-1}\left(\sum_{j=0}^{\tau_{\text{NARMA}}-1} y_{k-j-1}\right) + 1.5u_{k-\tau_{\text{NARMA}}}u_k + \gamma.$$

We consider $\tau_{\text{NARMA}} = \{15, 20, 30, 40\}$. For $\tau_{\text{NARMA}} = \{15, 20\}$, we set $\gamma = 0.1$. For $\tau_{\text{NARMA}} = \{30, 40\}$, γ is set to be 0.05 and 0.04 respectively.

Initial conditions of SA and all tasks are washed-out with 500 timesteps, followed by 1000 training timesteps, where output weights are optimized via ordinary least squares. We associate the output weights C and $W_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}}$ in (3.5) with *computational nodes*. While the number of computational nodes for SA can be chosen arbitrarily by varying the degree R in the output, the state-space size of SA is 4^n. This state-space size corresponds to the number of real variables to describe the system density operator. Since the density operator has unity trace, only up to at most $4^n - 1$ of these nodes are linearly independent.

We compare the SA performance with the ESN's and the Volterra series's. An ESN with state-space size m (Em) is given by [109]

$$\begin{cases} x_k = \tanh(W_r x_{k-1} + W_i u_k) \\ \hat{y}_k = W_o^\top x_k + w_c, \end{cases} \tag{3.11}$$

where $x_k \in \mathbb{R}^m$ and $W_o \in \mathbb{R}^m, w_c \in \mathbb{R}$ are output weights and $\tanh(\cdot)$ is an element-wise operation. The number of computational nodes is $m + 1$. Elements of W_r are uniformly randomly chosen from $[-2, 2]$ with $\sigma_{\max}(W_r) < 1$ to ensure the uniform convergence and the fading memory properties. The Volterra series with kernel order o and memory p (Vo, p) is $\hat{y}_k = h_0 + \sum_{i=1}^{o} \sum_{j_1,\cdots,j_i=0}^{p-1} h_i^{j_1,\cdots,j_i} \prod_{l=1}^{i} u_{k-j_l}$ [102]. The number of computational nodes is $\frac{p^{o+1}-p}{p-1}+1$. We analyze the performance of all schemes using 1000 evaluation data and the normalized mean-squared error

$$\mathrm{NMSE} := \sum_{k=1501}^{2500} |\bar{y}_k - y_k|^2 / \sum_{k=1501}^{2500} \left| y_k - \frac{1}{1000} \sum_{k=1501}^{2500} y_k \right|^2,$$

where y is the target output and $\bar{y}$ is the approximated output. For each task and each n, NMSEs of 100 convergent SA samples and ESNs are averaged for analysis.

The average NMSE for ESNs is obtained as in [54] to reflect the average ESN task performance over different choices of reservoir parameters. Let $\mathcal{S}$ denote the set of 10 points evenly spaced between $[0.01, 0.99]$. For each state-space size m and each of 100 ESNs, we linearly re-scale elements of W_r as for LRPO so that $\sigma_{\max}(W_r) = s$ for all $s \in \mathcal{S}$. For each s, the elements of W_i are randomly chosen within $[-\delta, \delta]$, where δ is chosen from the set $\mathcal{I}$ of 10 points evenly spaced between $[0.01, 1]$. Now, for the i-th ($i = 1, \ldots, 100$) ESN with parameter (m, s, δ), its NMSE is $\mathrm{NMSE}_{(m,s,\delta,i)}$. For each m, the average NMSE is $\frac{1}{|\mathcal{S}|}\frac{1}{|\mathcal{I}|}\frac{1}{100}\sum_{s\in\mathcal{S}}\sum_{\delta\in\mathcal{I}}\sum_{i=1}^{100}\mathrm{NMSE}_{(m,s,\delta,i)}$.

3.3.1 Overview of SA performance

We present an overview of SA performance in the LRPO, Missile, NARMA15 and NARMA20 tasks. The degree of the multivariate polynomial output (3.5) is $R = 1$. Fig. 3.1 shows the typical SA outputs during the evaluation phase. It is observed that as the number of system qubits n increases, the SA outputs better approximate the target outputs. This is quantitatively shown in Fig. 3.2, which plots the average SA NMSE as n increases.

From Fig. 3.2, SA with a small number of computational nodes performs comparably as ESNs and the Volterra series with a large number of computational nodes. The average NMSE of 6-qubit SA with 7 computational nodes is comparable to the average NMSE of E100 with 101 computational nodes in the LRPO task. On average, 5-qubit SA with 6 computational nodes performs better than V2, 22 with 504 computational nodes in the Missile task. Similar observations hold for the NARMA tasks. Our results are similar to the performance of the QRCs with time multiplexing reported in [54]. However, for

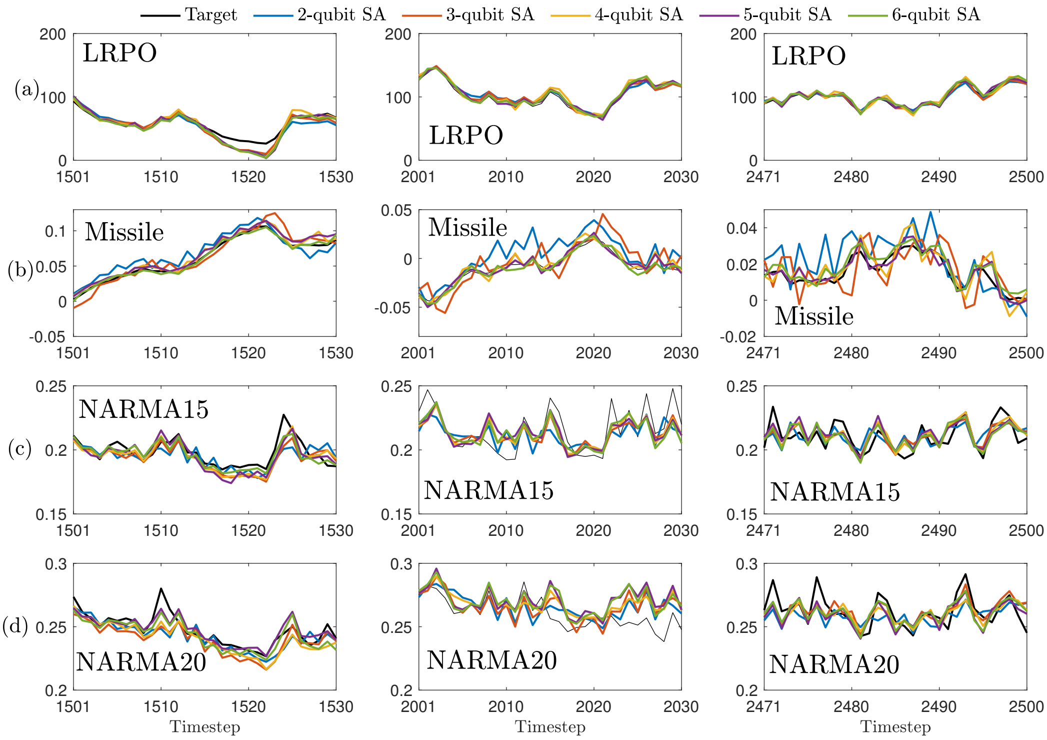


Figure 3.1: Typical SA outputs during the evaluation phase, for the (a) LRPO, (b) Missile (c) NARMA15 and (d) NARMA20 tasks. The leftmost, middle and rightmost panels show the outputs for timesteps 1501–1530, 2001–2030 and 2471–2500, respectively.

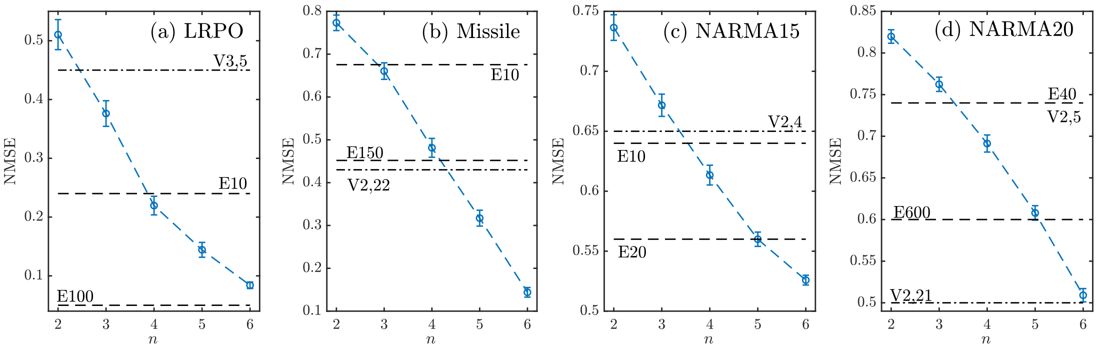


Figure 3.2: Average SA NMSE for the (a) LRPO, (b) Missile, (c) NARMA15 and (d) NARMA20 tasks, the error bars represent the standard error. For comparison, horizontal dashed lines labeled with ‘Em’ indicate the average performance of ESNs with m computational nodes, and horizontal dot-dashed lines labeled with ‘Vo, p’ indicates the performance of Volterra series with kernel order o and memory p. Overlapping dashed and dot-dashed lines are represented as dashed lines.

the small number of qubits investigated, the rate of decrease in the average NMSE is approximately linear despite the dimension of the Hilbert space increases exponentially as n increases. For the NARMA tasks, the average NMSEs for 2-qubit and 6-qubit SA are of the same order of magnitude. This suggests that a larger number of additional system qubits is required to substantially reduce the SA error.

3.3.2 SA performance under noise

We further investigate the feasibility of SA under dephasing, and GAD channel; also see [43, Chapter 8] and Sec. 2.2.3. We simulate the noise using the Trotter-Suzuki formula [57, 110], in which we divide $\tau S = 1$ into $\delta_t = \tau S/50$, and alternatively apply the unitary interaction and each noise on system and ancilla qubits with noise strength γ/S, each for a duration of δ_t. We set $\gamma/S = \{10^{-4}, 10^{-3}, 10^{-2}\}$, which are experimentally feasible for NMR ensembles [111] and some current superconducting machines [112].

The dephasing noise gives rise to the evolution $\rho \to \frac{1+e^{-2\frac{\gamma}{S}\delta_t}}{2}\rho + \frac{1-e^{-2\frac{\gamma}{S}\delta_t}}{2} Z^{(j)}\rho Z^{(j)}$ for the j–th qubit. The GAD channel gives rise to the evolution $\rho \to \sum_{l=0}^{3} M_l^{(j)}\rho(M_l^{(j)})^\dagger$ for the j–th qubit, where $M_l^{(j)} = M_l^{(j)}(\frac{\gamma}{S}, \lambda)$ $(l = 0, 1, 2, 3)$ acts on the j–th qubit and

$$M_0 = \sqrt{\lambda}\begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, M_2 = \sqrt{\lambda}\begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix},$$

$$M_3 = \sqrt{1-\lambda}\begin{pmatrix} \sqrt{1-p} & 0 \\ 0 & 1 \end{pmatrix}, M_4 = \sqrt{1-\lambda}\begin{pmatrix} 0 & 0 \\ \sqrt{p} & 0 \end{pmatrix},$$

with $\sqrt{1-p} = e^{-2\frac{\gamma}{S}\delta_t}$ and $\sqrt{p} = \sqrt{1 - e^{-4\frac{\gamma}{S}\delta_t}}$. We investigate the SA task performance for $\lambda = \{0.2, 0.4, 0.6, 0.8, 1\}$. When $\lambda = 1$, we recover the decaying noise which takes a state into the ground state $|0\rangle\langle 0|$ in the long time limit.

Fig. 3.3 plots the average SA NMSE under the dephasing, decaying and GAD with $\lambda = \{0.4, 0.6\}$ for all noise strengths. Fig. 3.3 indicates that for $\gamma/S = \{10^{-4}, 10^{-3}\}$, all noise types do not significantly degrade SA task performance. However, the impact of the noise

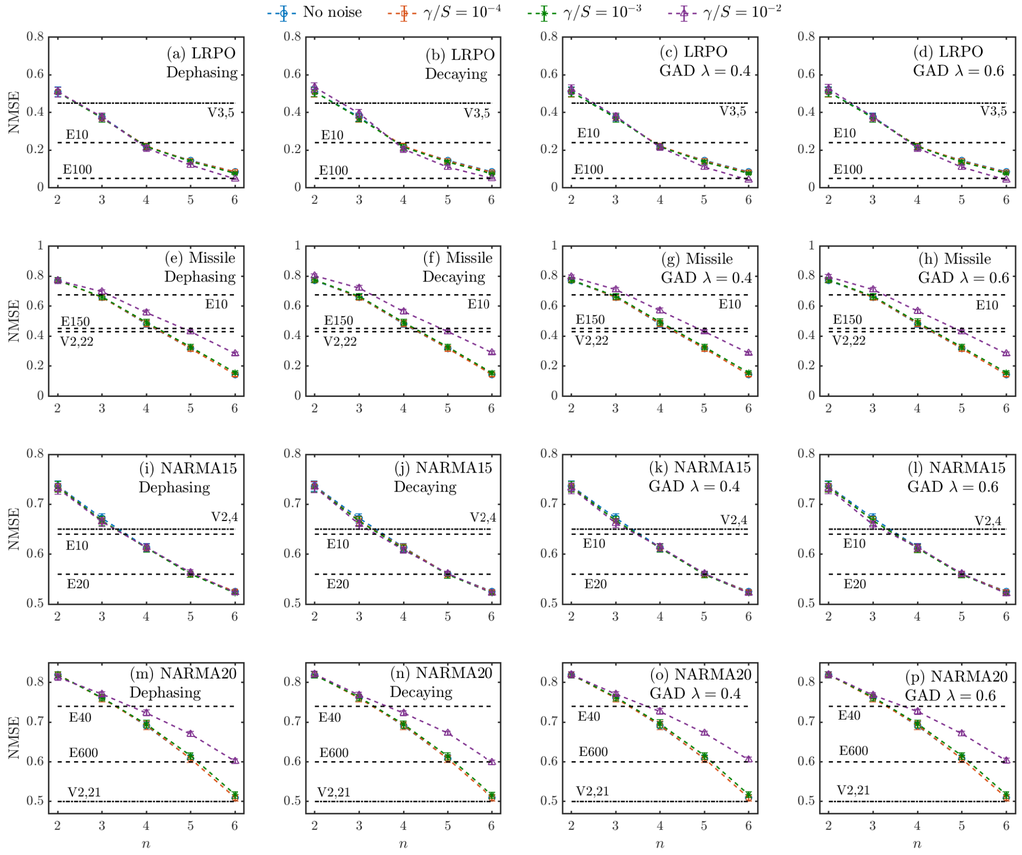


Figure 3.3: Average SA NMSE under decoherence. Error bars represent standard errors.

strength $\gamma/S = 10^{-2}$ is more pronounced, particularly for a larger number of system qubits. The same observations hold for GAD with $\lambda = \{0.2, 0.8\}$.

SA is impacted by the decaying noise and the GAD channel since the expectations $\langle Z^{(j)} \rangle_k$ in the output depend on the diagonal elements of the system density operator, which are affected by both noise. However, SA is also affected by the dephasing noise, which does not change the diagonal elements. A possible explanation is that the off-diagonal elements of the density operator become smaller and the density operator looks more like a classical probability distribution. Alternatively, this could be viewed as the off-diagonal elements contributing less to the overall computation. To support this explanation, for each n, we average the sum of the complex modulus of off-diagonal elements in the system density operator for the 100 n-qubit SA samples simulated above; see Fig. 3.4.

Fig. 3.4 shows that as λ/S increases, the average sum decreases, particularly for $\gamma/S = 10^{-2}$. Similar trends are observed for the GAD channel for all λ chosen, and the trend persists as the timestep increases to 2500. This further indicates that although the output of SA depends solely on the diagonal elements of the system density operator, nonzero off-diagonal elements play a role in the SA emulation performance. This provides a plausible explanation for the improved performance achieved by increasing the number of qubits, thereby increasing Hilbert space size and the number of non-zero off-diagonal elements. Further investigation into this topic is presented in Sec. 3.3.4.

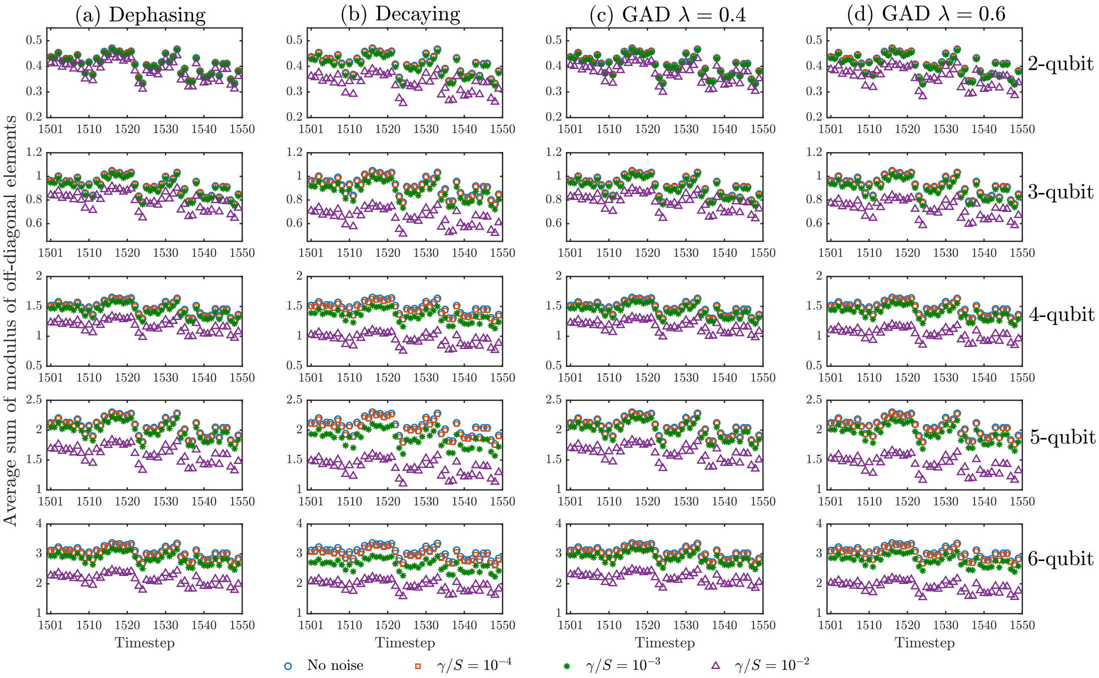


Figure 3.4: Average sum of complex modulus of off-diagonal elements in the system density operator for timesteps 1501–1550, under the (a) dephasing noise, (b) decaying noise, (c) GAD with $\lambda = 0.4$ and (d) GAD with $\lambda = 0.6$. Row $n-1$ corresponds the n-qubit SA.

3.3.3 Effect of different input encodings

Our proposed universal class encodes the input $u_k \in [0,1]$ into the mixed state $\rho_{i_0,k} = u_k|0\rangle\langle 0| + (1-u_k)|1\rangle\langle 1|$. Other input encoding possibilities include the pure state $\rho_{i_0,k} = (\sqrt{u_k}|0\rangle + \sqrt{1-u_k}|1\rangle)(\sqrt{u_k}\langle 0| + \sqrt{1-u_k}\langle 1|)$ used in the QRC model [54], the phase $\rho_{i_0,k} = \frac{1}{2}(|0\rangle + e^{-iu_k}|1\rangle)(\langle 0| + e^{iu_k}\langle 1|)$, and non-orthogonal basis state $\rho_{i_0,k} = u_k|0\rangle\langle 0| + \frac{1-u_k}{2}(|0\rangle +$

$|1\rangle)(\langle 0| + \langle 1|)$. We denote these encodings as mixed, pure, phase and non-orthogonal. For the last three encodings, the universality of the associated QRCs has not been proven.

For all encodings, the Hamiltonian parameters are sampled from the same uniform distribution, and we choose the resulting uniformly convergent QRCs. We again test the uniform convergence property by checking if 50 random initial density operators converge to the same density operator in 500 timesteps. The number of system qubits and the number of computational nodes for all encodings are the same. We average the NMSEs of 100 uniformly convergent QRCs for each encoding. Fig. 3.5 shows that for all tasks, the mixed state encoding performs better than other encodings on average. However, the average NMSE for different input encodings for all computational tasks are of the same order of magnitude. Moreover, as the number of system qubits increases, the errors of different input encodings decrease at roughly the same rate. This comparison indicates that the effect of different input encodings on the learning performance is not significant.

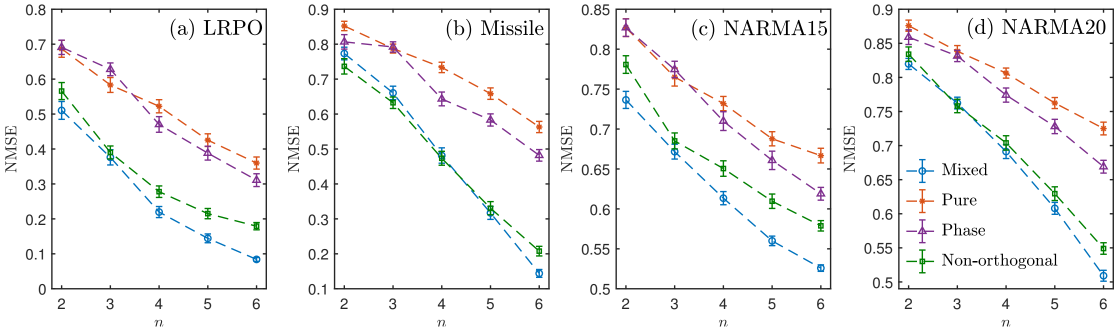


Figure 3.5: Average NMSE for different input encodings, for the (a) LRPO, (b) Missile (c) NARMA15 and (d) NARMA20 tasks. Error bars represent standard errors.

3.3.4 Further comparison with ESNs

Our numerical results so far and [54] both suggest that QRCs with a small number of qubits achieve comparable performance to classical RCs with a large number of computational nodes. However, these comparisons do not address quantum system's exponential state-space size. One can also increase the state-space size of ESNs and the number of computational nodes of SA, such that the state-space size and the number of computa-

Table 3.1: Average 4-qubit SA and E256 NMSE for the LRPO, Missile, NARMA15, NARMA20, NARMA30 and NARMA40 tasks. Results are rounded to two significant figures. The notation (± se) denotes the standard error.

Task	SA NMSE (± se)	E256 NMSE (± se)
LRPO	$0.20 \pm 1.5 \times 10^{-2}$	$0.019 \pm 7.7 \times 10^{-4}$
Missile	$0.48 \pm 2.2 \times 10^{-2}$	$0.49 \pm 3.3 \times 10^{-3}$
NARMA15	$0.61 \pm 8.0 \times 10^{-3}$	$0.32 \pm 1.6 \times 10^{-4}$
NARMA20	$0.68 \pm 1.0 \times 10^{-2}$	$0.67 \pm 3.2 \times 10^{-4}$
NARMA30	$0.67 \pm 7.1 \times 10^{-3}$	$0.67 \pm 4.0 \times 10^{-4}$
NARMA40	$0.64 \pm 5.3 \times 10^{-3}$	$0.66 \pm 5.9 \times 10^{-4}$

tional nodes are similar for both models. Here we present a further comparison between the SA model and ESNs, and provide insights on the possible advantage the SA model.

We focus on 4-qubit SA with a state-space size of 256 and 210 computational nodes (by setting $R = 6$ in (3.5)). We compare this 4-qubit SA model's average performance with the average E256 task performance. The number of computational nodes for E256 is 257 and the average NMSE of 100 convergent E256s is reported. As shown in Table 3.1, for the Missile and all the NARMA tasks, the average NMSEs for both models are of the same order of magnitude, while E256 outperforms SA in the LRPO task. This comparison suggests that when the state-space size and the number of computational nodes for both models are similar, ESNs can outperform the SA model.

We further investigate under what scenario SA may perform better. We observe that while the number of computational nodes is kept constant, increasing the state-space size of SA induces an improvement. To demonstrate this, NMSEs of 100 4-, 5- and 6-qubit SA samples are averaged for analysis. For each n-qubit SA, we vary its output degree R such that its number of computational nodes ranges from 5 to 252. The chosen degrees for 4-qubit SA are $R_4 = \{1, \ldots, 6\}$, for 5-qubit are $R_5 = \{1, \ldots, 5\}$, and for 6-qubit SA are $R_6 = \{1, \ldots, 4\}$. For comparison, we simulate 100 convergent ESNs with reservoir size 256 to perform the same tasks. For n-qubit SA, let $\mathcal{N}_n$ $(n = 4, 5, 6)$ denote the numbers of computational nodes corresponding to its output degrees R_n. The number of computational nodes $\mathcal{C}$ for E256 is set to be in $\mathcal{N}_4 \cup \mathcal{N}_5 \cup \mathcal{N}_6$. We first optimize 257 output weights for E256 via standard least squares during the training phase. When $\mathcal{C} < 257$ for E256, we select $\mathcal{C} - 1$ computational nodes (excluding the tunable constant output weight) with the largest absolute values and their corresponding state elements. These $\mathcal{C} - 1$ state elements are used to re-optimize $\mathcal{C}$ computational nodes (including the tunable constant output weight) via standard least squares. During the evaluation phase, 256 state elements evolve, only $\mathcal{C} - 1$ state elements and $\mathcal{C}$ output weights are used to compute the output.

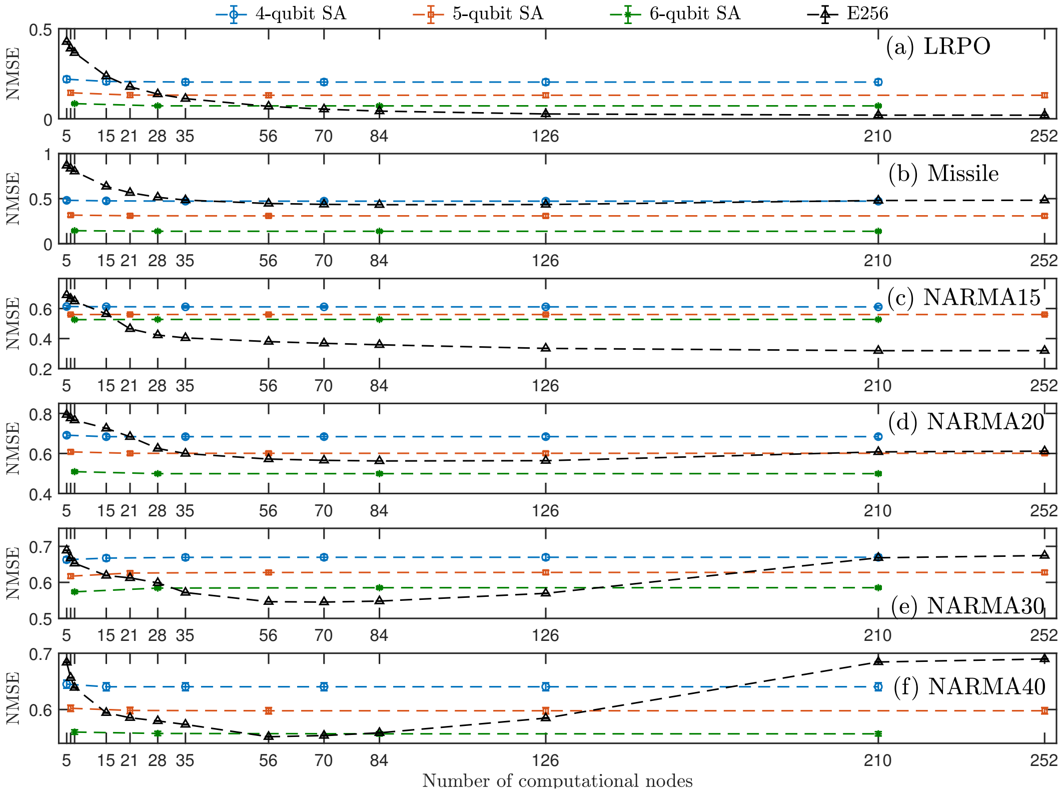


Figure 3.6: Average SA NMSE as the state-space size and the number of computational nodes vary for all computational tasks. The average NMSE for E256 with the same number of computational nodes is plotted for comparison. The data symbols obscure the error bars, which represent the standard error

Fig. 3.6 plots the 4-, 5-, and 6-qubit SA average NMSE as the number of computational nodes increases. Two important observations are that increasing the number of computational nodes does not necessarily improve SA task performance, while increasing the state-space size induces a noticeable improvement. For most tasks, despite 4-qubit SA might perform worse than E256, increasing the state-space size allows SA to outperform

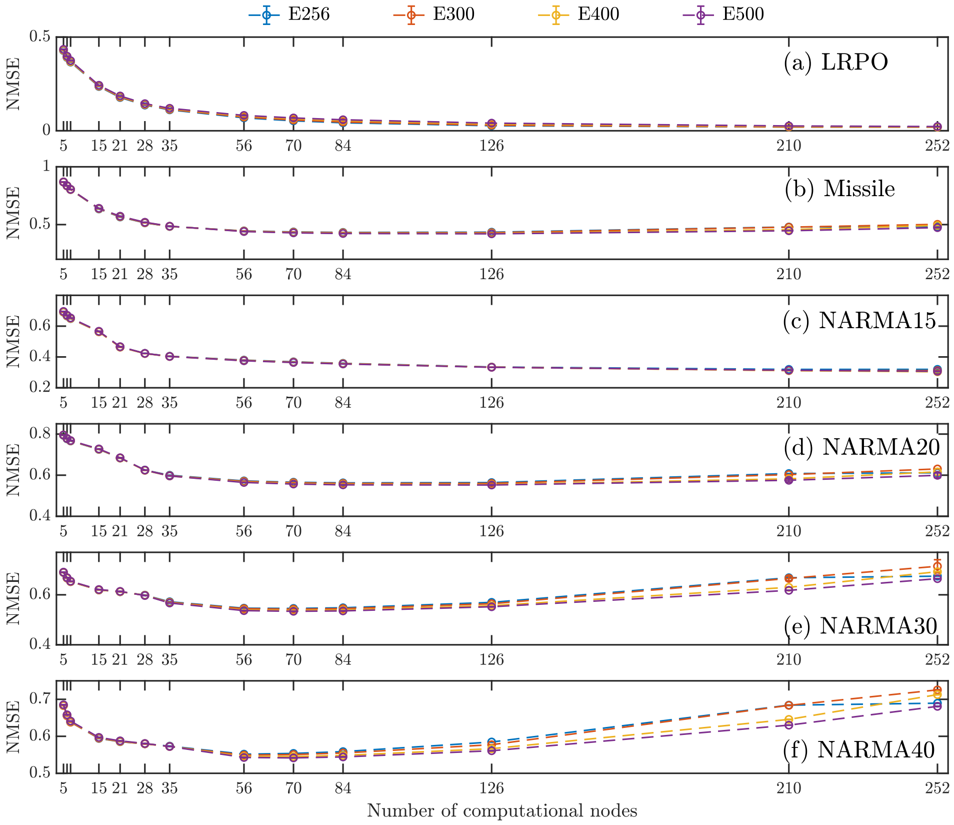


Figure 3.7: Average ESNs NMSE as the state-space size and the number of computational nodes vary for all computational tasks. The data symbols obscure the error bars, which represent the standard error

E256, without extensively increasing its number of computational nodes.

In contrary, increasing the state-space size of ESNs while keeping the number of computational nodes fixed does not induce a significant computational improvement. To numerically demonstrate this, the state-space size of ESNs is further increased to {300, 400, 500}, whose number of computational nodes is set to be the same as E256's. These computational nodes are chosen and optimized as above for E256. We average the NMSEs of 100 convergent ESNs for each reservoir size. As shown in Fig. 3.7, noticeable performance improvements for ESNs are observed as the number of computational nodes increases, but not as the state-space size varies. For the NARMA30 and NARMA40 tasks, the error increases as the number of computational nodes increases. This could be due to overfitting, occurs when too many adjustable parameters are trained on limited training data [113]. On the other hand, this observation is less significant for SA. It would be interesting to conduct further investigation into this behavior in future work.

The above observations have several implications. To improve the SA performance, one can increase state-space size while only optimizing a polynomial number of computational nodes. In contrary, enhancing emulation performance of ESNs requires more computational nodes and hence a larger state-space size. In the situation where the state-space increases beyond what classical computers can simulate with reasonable resources (such as memory), the ESN performance saturates, whereas the SA performance could be further improved by increasing the number of qubits in a linear fashion. To further test this hypothesis, the SA performance would need to be evaluated for a larger number of qubits on a physical quantum system. A possible hardware implementation is on NMR ensembles as suggested in [54]. Motivated by the availability of NISQ machines, a quantum circuit implementation of the QRC would also be attractive. This will be the topic for Chapter 4.

3.4 Conclusion and outlook

We have developed a general theory for learning nonlinear I/O maps with fading memory using QRCs implemented by dissipative quantum systems. This theory allows dissipative quantum systems (that meet certain requirements but is otherwise arbitrary) to be combined with a classical processor to learn I/O maps. We apply the theory to demonstrate a universal class of dissipative quantum systems implemented by Ising Hamiltonians.

Numerical examples suggest that with a small number of qubits and a linear output, this class can achieve comparable performance, in terms of the average normalized mean-squared error, to ESNs and the Volterra series with a large number of parameters. Moreover, numerical results indicate that increasing QRC state-space size while fixing the number of computational nodes can improve performance on some tasks, whereas increasing the ESN state-space size while fixing the computational nodes does not lead to noticeable improvement. This suggests that the possibly large Hilbert space of the quantum system presents a potential resource in this approach, particularly for state-space sizes beyond what can be simulated on a digital computer. To investigate this hypothesis, the scheme would need to be tested for a larger number of qubits on a quantum hardware. In the next chapter, we introduce a universal class of QRCs that can be implemented on current NISQ devices, paving a path towards larger-scale experiments in the future.

Chapter 4

Gate-model quantum reservoir computers

Motivated by the increasing availability of NISQ devices, we propose a gate-model QRC for temporal information processing on NISQ quantum computers. The use of quantum systems as QRCs was initially proposed in [54] to harness disordered-ensemble quantum dynamics for temporal information processing. This scheme is suitable for ensemble quantum systems such as NMR [56]. However, it remained an open problem to show this QRC class is universal. In the previous chapter, we address this problem by demonstrating that a variation of the scheme proposed in [54] is universal for nonlinear fading memory maps. However, realizing these previous QRC proposals on gate-model NISQ machines remains challenging due to the large number of quantum gates required to implement the dynamics via Trotterization [57].

This chapter is based on the publications [114] and [115]. The contribution of this chapter is twofold:

- We propose a class of QRCs endowed with the fading memory and universality properties that is not necessarily implemented by Ising Hamiltonians, circumventing the need for Trotterization required in previous proposals.

- We propose a realization of a subclass of the universal QRC class on NISQ devices and present proof-of-principle experiments on remotely accessed IBM superconducting quantum processors [44], i.e., NISQ devices not yet equipped with quantum error correction.

The QRC dynamics in this subclass can be implemented using arbitrary but fixed quantum circuits, as long as they generate non-trivial dynamics. This could be, for instance, quantum circuits that are classically intractable to simulate. The quantum circuits can be of short lengths and can be implemented using single-qubit and multi-qubit quantum gates native to the quantum hardware, without the need for precise tuning of their gate parameters. Our proof-of-principle experiments show that QRCs with a small number of noisy qubits can tackle nonlinear temporal tasks, even under current hardware limitations and in the absence of readout and process error mitigation techniques. This work serves as a theoretical and experimental realization of applying near-term gate-model quantum computers to nonlinear temporal information processing tasks, opening an avenue for time series modeling and signal processing applications of these devices.

This chapter is organized as follows. Sec. 4.1 presents our QRC proposal and the universality result. In Sec. 4.2, we then propose a subclass of the universal class suitable for implementation on current noisy gate-model quantum computers. Sec. 4.3 details two hardware realizations of the aforementioned subclass of the universal class and presents more efficient versions of both schemes. These efficient schemes could enable QRC's potential for more scalable temporal processing on gate-model quantum devices. Sec. 4.4 details our proof-of-principle experiments performed on cloud-based IBM superconducting quantum devices. We then discuss some future directions in Sec. 4.5. Lastly, concluding remarks are provided in Sec. 4.6. Detailed experimental settings are provided in Appendix A.

4.1 A universal class of QRCs

As in the previous chapter, we consider a QRC consists of N_s non-interacting subsystems initialized in a product state, each subsystem K has n_K number of qubits so that the QRC has $n = \sum_{K=1}^{N} n_K$ qubits. The QRC's system density operator ρ_k at time k evolves according to

$$\rho_k = T(u_k)\rho_{k-1} = \bigotimes_{K=1}^{N_s} T^{(K)}(u_k)\rho_{k-1}^{(K)}, \tag{4.1}$$

and the K-th subsystem density operator $\rho_k^{(K)}$ undergoes the evolution

$$T^{(K)}(u_k)\rho_{k-1}^{(K)} = (1-\epsilon_K)\left(u_k T_0^{(K)} + (1-u_k)T_1^{(K)}\right)\rho_{k-1}^{(K)} + \epsilon_K \sigma_K, \tag{4.2}$$

for input $u_k \in [0,1]$. Here, $\epsilon_K \in (0,1]$, σ_K is an arbitrary but fixed density operator, and $T_0^{(K)}$ and $T_1^{(K)}$ are two arbitrary but fixed CPTP maps. Examples of such maps include some naturally occurring quantum noise channels such as dephasing or GAD channels discussed in Sec. 2.2.3. No precise tuning or engineering of the CPTP maps $T_0^{(K)}, T_1^{(K)}$ is required for the QRC scheme and it should not generate trivial dynamics (e.g., we should not choose $T_0^{(K)} = T_1^{(K)}$).

The QRC governed by (4.1)–(4.2) is 1–uniformly convergent as shown in the following lemma.

Lemma 4.1 (Uniformly convergent)**.** *The QRC governed by* (4.1) *and* (4.2) *is 1–uniformly convergent with respect to inputs* $[0,1]^{\mathbb{Z}}$.

Proof. We first show that each subsystem governed by (4.2) is 1–uniformly convergent. For any density operators $\rho, \sigma \in \mathcal{D}(2^{n_K})$, $u_k \in [0,1]$ and $\epsilon_K \in (0,1]$, we have

$$\begin{aligned} \|T^{(K)}(u_k)(\rho-\sigma)\|_1 &= (1-\epsilon_K)\left\|\left(u_k T_0^{(K)} + (1-u_k)T_1^{(K)}\right)(\rho-\sigma)\right\|_1 \\ &\leq (1-\epsilon_K)\|\rho-\sigma\|_1 \leq 2(1-\epsilon_K), \end{aligned} \tag{4.3}$$

where the last inequality follows from [43, Theorem 9.2] and the convex combination $u_k T_0^{(K)} + (1-u_k)T_1^{(K)}$ is again a CPTP map. Applying Theorem 2.2 in Chapter 2 shows that each subsystem is 1–uniformly convergent.

Since the N_s subsystems are non-interacting and the QRC is always initialized in a product state, the QRC governed by (4.1)–(4.2) is 1–uniformly convergent. □

In this chapter, we only consider 1–uniformly convergent QRCs and will drop the prefix '1' from now on. We again specify the multivariate polynomial readout function (4.4) for the QRC introduced above. This is the same readout function as in the previous chapter; see (3.5). A simple linear form ($R = 1$) will be employed in our proof-of-principle experiments in Sec. 4.4.

$$\begin{aligned}\hat{y}_k &= h(\rho_k)\\ &= C + \sum_{d=1}^{R}\sum_{i_1=1}^{n}\sum_{i_2=i_1+1}^{n}\cdots\sum_{i_n=i_{n-1}+1}^{n}\sum_{r_{i_1}+\cdots+r_{i_n}=d} W_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}} \langle Z^{(i_1)}\rangle_k^{r_{i_1}}\cdots\langle Z^{(i_n)}\rangle_k^{r_{i_n}}\end{aligned} \tag{4.4}$$

Recall from Sec. 2.4.3 that (4.1) and (4.4) define a QRC implementing an I/O map $M_{T,h}$. This I/O map depends on the QRC dynamics T and the readout function h. We now show that $M_{T,h}$ has the fading memory property.

Lemma 4.2 (Fading memory)**.** *For any null sequence w, $M_{T,h}$ induced by a QRC described by* (4.1) *and* (4.4) *is w-fading memory.*

Proof. Using the same argument in Lemma 3.3 in Chapter 3, it follows that $M_{T,h}$ is w-fading memory if each K-th subsystem dynamics $T^{(K)}(u_k)$ is uniformly continuous with respect to the inputs $u_k \in [0,1]$ for all $K = 1,\ldots,N_s$. Let $x, y \in [0,1]$ and $A \in \mathbb{C}^{2^{n_K}\times 2^{n_K}}$,

$$\begin{aligned}\|T^{(K)}(x) - T^{(K)}(y)\|_{1-1} &= \sup_{A\in\mathbb{C}^{2^{n_K}\times 2^{n_K}},\|A\|_1=1} \left\|\left(T^{(K)}(x) - T^{(K)}(y)\right)A\right\|_1\\ &= (1-\epsilon_K)|x-y| \sup_{A\in\mathbb{C}^{2^{n_K}\times 2^{n_K}},\|A\|_1=1} \left\|T_0^{(K)}(A) - T_1^{(K)}(A)\right\|_1\\ &\le (1-\epsilon_K)|x-y|\left(\left\|T_0^{(K)}\right\|_{1-1} + \left\|T_1^{(K)}\right\|_{1-1}\right)\\ &\le 2(1-\epsilon_K)|x-y|,\end{aligned}$$

where the last inequality follows from [105, Theorem 2.1]. We remark that Lemma 3.3 in Chapter 3 is stated with respect to the Schatten–2 norm, but the same argument holds for the Schatten–1 norm. □

For any null sequence w, consider the class $\mathcal{M}_w$ of w-fading memory maps $M_{T,h}$. These w-fading memory maps may be implemented by QRCs with different number of subsystems N_s and number of qubits n. They may also be implemented by QRCs with different readout parameters and readout degree R. Our main result shows that the class $\mathcal{M}_w$ is universal for approximating nonlinear w-fading memory maps.

Theorem 4.1 (Universality). *Let w be any null sequence. For any nonlinear w-fading memory map M_* and any $\delta > 0$, there exists $M_{T,h} \in \mathcal{M}_w$ implemented by some QRC described by* (4.1) *and* (4.4) *such that for all $u \in [0,1]^{\mathbb{Z}}$, $\sup_{k\in\mathbb{Z}} |M_*(u)|_k - M_{T,h}(u)|_k| < \delta$.*

Proof. We employ the Stone-Weierstrass Theorem 3.1 as in Chapter 3. We have shown the uniform convergence in Lemma 4.1 and the fading memory property in Lemma 4.2. Let $\mathcal{F}_w = \{F_{T,h}\}$ be the corresponding induced family of w-fading memory functionals. It remains to show that $\mathcal{F}$ forms a polynomial algebra consisting of w-fading memory maps and it separates points.

The family $\mathcal{F}_w$ forms a polynomial algebra follows from Lemma 3.5 in Chapter 3 and the observation that for any two QRC dynamics $T_1(u_k) = \bigotimes_{K=1}^{N_s} T_1^{(K)}(u_l)$ and $T_2(u_l) = \bigotimes_{k=1}^{N_2} T_2^{(k)}(u_l)$, where each $T_1^{(K)}, T_2^{(K)}$ has the form (4.2), we again have $T(u_k)(\rho_1 \otimes \rho_2) = T_1(u_k)\rho_1 \otimes T_2(u_k)\rho_2$ is of the form (4.1). Furthermore, $T(u_k) = T_1(u_k) \otimes T_2(u_k)$ is again uniformly convergent when initialized in a product state. Therefore, $\mathcal{F}_w$ forms a polynomial algebra consisting of w-fading memory maps.

Constant functions can be obtained by setting $W_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}} = 0$ in (4.4). It remains to show that $\mathcal{F}_w$ separates points. We proceed similarly to Chapter 3.

Consider a single-qubit QRC with a linear readout function ($n = 1, R = 1, N_s = 1$). For the rest of this proof, we drop the subsystem index K. This QRC consists of one system qubit and one ancilla qubit denoted as ρ_a. Choose the dynamics

$$\begin{aligned}\rho_k &= T(u_k)\rho_{k-1} \\ &= (1-\epsilon)\left(u_k \mathrm{Tr}_a\left(e^{-iH}(\rho_{k-1}\otimes\rho_a^0)e^{iH}\right) + (1-u_k)\mathrm{Tr}_a\left(e^{-iH}(\rho_{k-1}\otimes\rho_a^1)e^{iH}\right)\right) + \epsilon K_{\frac{I}{2}},\end{aligned}$$

where $\rho_a^j = |j\rangle\langle j|$ for $j = 0, 1$, Tr_a denotes the partial trace over ancilla ρ_a and $\epsilon \in (0, 1)$. The map $K_{\frac{I}{2}}$ is a CPTP map defined as $K_{\frac{I}{2}}(X) = \mathrm{Tr}(X)\frac{I}{2}$ for any $X \in \mathbb{C}^{2\times 2}$. The Hamiltonian H is of the Ising type $H = J(X^{(0)}X^{(1)} + Y^{(0)}Y^{(1)}) + \alpha \sum_{j=0}^{1} Z^{(j)}$, where $X^{(j)}, Y^{(j)}$ and $Z^{(j)}$ are the Pauli X, Y and Z operators on qubit j, with $j = 0$ being the ancilla qubit.

We order an orthogonal basis for $\mathbb{C}^{2\times 2}$ as $\{I, Z, X, Y\}$. The matrix representation of the above CPTP map is

$$\begin{aligned}\overline{T}(u_l) &= |00\rangle\langle 00| \\ &+ (1-\epsilon)\begin{pmatrix} 0 & 0 & 0 & 0 \\ \sin^2(2J)(2u_l - 1) & \cos^2(2J) & 0 & 0 \\ 0 & 0 & \cos(2J)\cos(2\alpha) & -\cos(2J)\sin(2\alpha) \\ 0 & 0 & \cos(2J)\sin(2\alpha) & \cos(2J)\cos(2\alpha) \end{pmatrix}\end{aligned} \tag{4.5}$$

Since the QRC is uniformly convergent, we can choose any initial condition $\rho_{-\infty} = |0\rangle\langle 0|$ with the corresponding vector representation $\overline{\rho}_{\infty} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}^\top$. Taking a linear readout function, the QRC implements $F_{T,h}(u) = 2W\left[\left(\overrightarrow{\prod}_{j=0}^{\infty}\overline{T}(u_{-j})\right)\overline{\rho}_{-\infty}\right]_2 + C$, where $[\cdot]_2$ is the second element of the vector corresponding to $\langle Z\rangle$. Furthermore, choose $W \neq 0$.

Now given two distinct inputs $u, v \in [0, 1]^{\mathbb{Z}_-}$, suppose that $u_0 \neq v_0$. Then choose J such that $\cos^2(2J) = 0$, then $F_{T,h}(u)_0 - F_{T,h}(v)_0 = 2W(1-\epsilon)(u_0 - v_0) \neq 0$.

Suppose that $u_0 = v_0$. Choose $\epsilon \in (0, 1)$ and J such that $(1-\epsilon)\cos^2(2J) \in (0, 1-\epsilon)$. Then $F_{T,h}(u)_0 - F_{T,h}(v)_0 = 2W\sin^2(2J)(1-\epsilon)\sum_{j=0}^{\infty}\left((1-\epsilon)\cos^2(2J)\right)^j(u_{-j} - v_{-j})$.

The above is a power series of the form $f(\theta) = 2W\sin^2(2J)(1-\epsilon)\sum_{j=0}^{\infty}\theta^j(u_{-j} - v_{-j})$, where $f(\theta)$ has a nonzero radius of convergence and is non-constant since $\theta = (1-\epsilon)\cos^2(2J) \in (0, 1-\epsilon)$ and $(1-\epsilon)\sin^2(2J) \in (0, 1-\epsilon)$. Furthermore, since we assume that $u_0 = v_0$, we have $f(0) = 0$. Invoking [106, Theorem 3.2], there exists $\beta > 0$ such that $f(\theta) \neq 0$ for all $|\theta| \leq \beta, \theta \neq 0$. This concludes the proof for separation of points. The universality of $\mathcal{M}_w$ now follows from the Stone-Weierstrass Theorem. □

Besides the universality property, our proposed universal QRC class exhibits invariance properties under certain hardware imperfections; see Sec. 4.1.1 below.

4.1.1 Invariance under certain noise

The universal family of QRCs is *invariant and remains universal* under certain stationary Markovian noise processes modeled by CPTP maps. The K-th subsystem's dynamics (4.2) during some time interval $\tau(k-1) \leq t \leq \tau k$, where k is the time step and $\tau > 0$, is

$$\rho_k^{(K)} = (1-\epsilon_K)\left(u_k \mathcal{T}^{(K)} \circ T_0^{(K)} + (1-u_k)\mathcal{T}^{(K)} \circ T_1^{(K)}\right)\rho_{k-1}^{(K)} + \epsilon_K \mathcal{T}^{(K)}(\sigma_K),$$

where $\mathcal{T}^{(K)} \circ T_j^{(K)}$ is again some CPTP for $j = 0, 1$ and $\mathcal{T}^{(K)}(\sigma_K) = \sigma'_K$ is again some fixed density operator. The resulting noisy dynamics again has the form (4.2) and the form of QRC dynamics (4.1) also remains unchanged. That is, the universal family is invariant and remains universal under such stationary Markovian noise processes.

Such a noise model is the noise model adopted in the IBM Qiskit simulator [116]. The Qiskit noisy simulation approximates the hardware noise as a CPTP map being applied after the application of a unitary gate. The noise parameters are estimated during periodic calibrations on the hardware. Between two calibrations, the calibrated noise parameters remain unchanged. However, during the experiments, the underlying hardware noise could potentially be time-varying. Considering these factors, the agreement between our experimental and Qiskit noisy simulation results (see Appendix A.7 for the data) indicate the underlying hardware noise approximately preserves the QRC dynamics of the form (4.1) during the experiments. If the underlying noise is non-stationary but changes slowly, the QRC output weights can be re-trained periodically using most recently gathered data. This remains challenging to demonstrate on current cloud-accessed NISQ devices but may be possible on future NISQ machines.

Furthermore, QRC predicted outputs remain unchanged under time-invariant readout error whenever a linear readout function is used (i.e., $R = 1$ in (4.4), which is often employed in practice and in our proof-of-principle experiments). This is because time-invariant readout error introduces a time-invariant linear transformation of the measurement data and

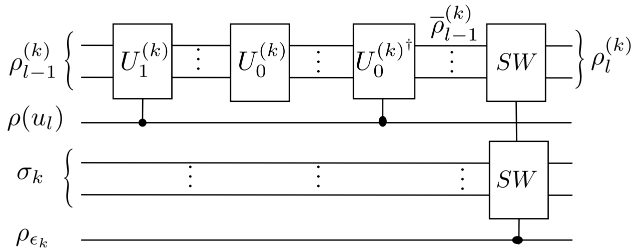


Figure 4.1: Quantum circuit interpretation of the QRC subclass described in Sec. 4.2. Here $\rho_{k-1}^{(K)}$ and σ_K are two quantum registers (i.e., groups of qubits) whereas $\rho(u_k)$ and ρ_{ϵ_K} are two single-qubit states. The unitaries $U_1^{(K)}, U_0^{(K)\dagger}$ act on $\rho_{k-1}^{(K)}$, controlled by $\rho(u_k)$. The right-most operation (SW's) swaps the states of $\overline{\rho}_{k-1}^{(K)}$ and σ_K, controlled by ρ_{ϵ_K}.

when the output weights $W_{i_1,\ldots,i_n}^{r_{i_1},\ldots,r_{i_n}}$ and C are optimized via linear regression, the resulting QRC predicted outputs $\hat{y}_k$ remain unchanged; see Appendix A.1 for the derivation.

4.2 A subclass implementable on noisy gate-model quantum devices

With a limited number of qubits and current hardware restrictions, not all QRC dynamics of the form (4.1)–(4.2) can be efficiently implemented. Here we describe a subclass of the above universal QRC class implementable on current gate-model NISQ devices.

QRCs in this subclass are governed by (4.1)–(4.2) with unitary evolutions $T_j^{(K)}(\rho_{k-1}^{(K)}) = U_j^{(K)}\rho_{k-1}^{(K)}U_j^{(K)\dagger}$ ($j = 0, 1$), where the unitaries $U_0^{(K)}$ and $U_1^{(K)}$ are arbitrary but fixed. In practice, $U_j^{(K)}$ can be implemented by native quantum gates of the NISQ devices, possibly composed of single-qubit and multi-qubit gates each parameterized by some gate parameter. These gate parameters can be chosen arbitrarily but fixed and should not generate trivial dynamics (e.g., we should not have $U_0^{(K)} = U_1^{(K)}$), thus precise tuning of these parameters is not required. In Sec. 4.4, we suggest some natural choices of $U_j^{(K)}$ tailored for the cloud-based IBM quantum devices [44]. The QRC dynamics in this subclass has a natural quantum circuit interpretation, see Fig. 4.1.

In Fig. 4.1, the state $\rho(u_k)$ encodes the input u_k as a classical mixture $\rho(u_k) = u_k|0\rangle\langle 0| + (1-u_k)|1\rangle\langle 1|$, meaning that we apply $U_0^{(K)} \rho_{k-1}^{(K)} U_0^{(K)^\dagger}$ with probability u_k, and apply $U_0^{(K)^\dagger} U_0^{(K)} U_1^{(K)} \rho_{k-1}^{(K)} U_1^{(K)^\dagger} U_0^{(K)^\dagger} U_0^{(K)} = U_1^{(K)} \rho_{k-1}^{(K)} U_1^{(K)^\dagger}$ with probability $1-u_k$. Let $\overline{\rho}_{k-1}^{(K)}$ denote the QRC's K-th subsystem state after these operations. The state ρ_{ϵ_K} is a classical mixture $\rho_{\epsilon_K} = (1-\epsilon_K)|0\rangle\langle 0| + \epsilon_K|1\rangle\langle 1|$ that encodes the rate ϵ_K at which the K-th subsystem forgets its initial conditions. That is, with probability ϵ_K, the states $\overline{\rho}_{k-1}^{(K)}$ and σ_k are exchanged, equivalent to resetting the state $\overline{\rho}_{k-1}^{(K)}$ to the fixed density operator σ_K; otherwise the state $\overline{\rho}_{k-1}^{(K)}$ is unchanged with probability $1-\epsilon_K$. We again associate the readout function (4.4) to this QRC subclass.

4.3 Realization of a subclass on current quantum hardware

We present two implementation schemes of the subclass in Sec. 4.2 on current gate-model NISQ computers. The first scheme takes into account some current hardware limitations, and the second scheme employs quantum non-demolition (QND) measurements to substantially reduce the number of circuit runs required. We further show that QRC's uniform convergence property leads to more efficient versions of both schemes. A method to efficiently obtain the QRC prediction as a function of the quantum expectation values of non-commuting observables is also introduced.

4.3.1 Implementation schemes

Here, we focus on n-qubit QRCs with a single subsystem ($N_s = 1$ in (4.1)) and drop the subsystem index K in (4.2). The case with multiple subsystems ($N_s > 1$) is a straightforward extension. We may choose $\sigma = |\psi\rangle\langle\psi|$ with an easy to prepare pure state $|\psi\rangle$. In all schemes, we initialize the QRC circuits in $|0\rangle^{\otimes n}$.

The first implementation is employed in our proof-of-principle experiments (see Sec. 4.4). We consider NISQ devices that only allow pure state preparation. Instead of realizing Fig. 4.1 that requires mixed state preparation, we implement QRCs through Monte Carlo sampling. We construct N_m circuits, such that for each circuit and at each timestep k, we apply U_0 and U_1 with probabilities $(1-\epsilon)u_k$ and $(1-\epsilon)(1-u_k)$, respectively; otherwise the circuit is set in $|\psi\rangle$ with probability ϵ. Therefore, for each N_m circuits and each time k, implementing the input-dependent QRC dynamics $T(u_k)$ in (4.1) amounts to applying the gate sequence realizing U_0 or U_1, or resetting the circuit in $|\psi\rangle$. As N_m is increased, the average of all measurements gives a more accurate estimate of the true expectation $\langle Z^{(i)}\rangle_k$ in the output (4.4), see Appendix A.2. Furthermore, some current NISQ devices do not allow qubit reset, meaning that once a qubit is measured, it cannot be re-used for computation. To estimate $\langle Z^{(i)}\rangle_k$, we re-initialize N_m circuits in $|0\rangle^{\otimes n}$ and re-apply $T(u_l)$ from time $l=1$ to time $l=k$, and only measuring $Z^{(i)}$ at the final time k. Each of the N_m circuits is run for S shots for each time k. To process a length-L input sequence under the pure state and qubit re-set limitations requires $N_m SL$ circuit runs and $N_m S(1+\cdots+L) = N_m S(L+1)L/2$ applications of $T(u_k)$.

If qubit reset is available, a more efficient scheme using QND measurements [117–120] can be realized. QND measurements allow us to measure observables during the evolution of a quantum system without increasing the uncertainty of the measurement outcomes. More precisely, QND measurements can be defined as follows. Let $\mathcal{K}$ be a set of times (we consider QND measurements in discrete-time here). At any time $k \in \mathcal{K}$, we can choose to measure some observable O_k of a quantum system. Note that one can measure a different observable at each time. Suppose that up to time k, the quantum system undergoes the unitary evolution U_k. Let $j_k(O_k) = U_k^\dagger O_k U_k$. Then a sequential measurement of $\{O_k\}_{k\in\mathcal{K}}$ is said to be a QND measurement if the commutator $[j_k(O_k), j_s(O_s)] = 0$ for all $k, s \in \mathcal{K}$. A QND measurement $\{j_k(O_k)\}_{k\in\mathcal{K}}$ is a well-defined classical stochastic process in the sense that $\{j_{k_1}(O_{k_1}), j_{k_2}(O_{k_2}), \ldots, j_{k_n}(O_{k_n})\}$ has a joint probability distribution for any $n \in \mathbb{Z}_{\geq 1}$ and any $k_1, k_2, \ldots, k_n \in \mathcal{K}$. In particular, the process $\{j_k(O_k)\}_{k\in\mathcal{K}}$ has well-defined sample paths. When a measurement is not QND, $\{j_k(O_k)\}_{k\in\mathcal{K}}$ is not a classical stochastic process. This process can be thought of as a genuine quantum stochastic process that cannot be realized on some classical probability space. For an introduction to quantum stochastic processes, see [121, 122]. In Appendix A.3, we detial this more efficient implementation using QND measurements.

Using QND measurements, we no longer need to re-run the N_m circuits from time 1 to estimate $\langle Z^{(i)} \rangle_k$. Instead we just run each of the N_m circuits S shots, meaning that for each circuit we perform a QND measurement of $Z^{(i)}$ at time k, continue running the circuit until the next measurement, and so forth. QND measurements ensure information encoded in ρ_k is retained from one timestep to the next. This scheme requires $N_m SL$ applications of $T(u_k)$ but only $N_m S$ circuit runs as opposed to $N_m SL$ runs in the first scheme. Since qubit reset was not available when our proof-of-experiments were conducted in 2019, we employed the first implementation scheme. More recently, the qubit-reset functionality has become available on some NISQ devices such as the Honeywell device [123] and the IBM devices [124]. It will be interesting to implement this QND scheme in such a device in the future.

The QRC's uniform convergence property leads to more efficient versions of both schemes. Let $M \geq 1$ be a fixed integer and suppose that we want to estimate $\langle Z^{(i)} \rangle_k$ at a sufficiently large timestep k (that depends on ϵ). Suppose we initialize N_m circuits in $|0\rangle^{\otimes n}$, re-apply and re-run $T(u_l)$ from $l = 1$ as before. We then obtain the QRC states ρ_{k-M} at timestep $k - M$ and ρ_k at timestep k. Thanks to the uniform convergence property, we can instead re-initialize the N_m circuits in $|0\rangle^{\otimes n}$ at timestep $k - M$ and from this time onwards re-apply and re-run $T(u_l)$ according to inputs $\{u_{k-M+1}, \ldots, u_k\}$. At timestep k, we have the corresponding QRC state $\tilde{\rho}_k$. By the uniform convergence property, we can make the difference between ρ_k and $\tilde{\rho}_k$ negligible by choosing M appropriately based on ϵ. If we perform repeated measurements on ρ_k and $\tilde{\rho}_k$, the estimates of $\langle Z^{(i)} \rangle_k$ and $\widetilde{\langle Z^{(i)} \rangle}_k = \mathrm{Tr}(\tilde{\rho}_k Z^{(i)})$ will also be close; see Appendix A.2 for the analysis.

The uniform convergence property can be readily exploited on current NISQ machines, leading to efficient versions of both schemes. The first scheme now requires $N_m SL$ circuit runs but only $N_m SM$ applications of $T(u_k)$. The second scheme now only needs $N_m S$ circuit runs and $N_m SM$ applications of $T(u_k)$, both are *independent* of the input length L, enabling QRC's potential for fast and scalable temporal processing. In all schemes, it is possible and perhaps advantageous to set $S = 1$ and run N_m circuits for a sufficiently large N_m. It is possible to run the N_m circuits in parallel if multiple copies of the same hardware are available. The average of N_m measurements estimates $\langle Z^{(i)} \rangle$, whose estimation accuracy increases as N_m increases; see Appendix A.2 for the analysis.

Lastly, we remark that it is possible and perhaps advantageous to set $S = 1$ in all schemes and run N_m circuits (possibly in parallel if multiple copies of the same quantum hardware are available), for a sufficiently large N_m. The average of N_m measurements at time k estimates $\langle Z^{(i)} \rangle_k$, with estimation accuracy increases as N_m increases.

4.3.2 Unitary trick for efficient QRC predictions

As alluded to in the last chapter, a more general form of the output functional as a multivariate polynomial in the variables $\langle X^{(i)} \rangle_k, \langle Y^{(i)} \rangle_k$ and $\langle Z^{(i)} \rangle_k (i = 1, \ldots, n)$ can be used as the QRC output. However, since these observables are non-commuting, we cannot measure them simultaneously. Using the first implementation scheme above, we need to prepare each of the $N_m L$ circuits 3 times and perform a total of $3N_m SLn$ measurements. Using the QND implementation scheme, we need to prepare each of the N_m circuits 3 times and perform a total of $3N_m Sn$ measurements.

In this subsection, we present a unitary trick to obtain any linear combination of the form $\sum_{i=1}^{n} \sum_{M=\{X,Y,Z\}} W_i^M \langle M^{(i)} \rangle_k$ efficiently. A crucial observation here is that a linear QRC output (with $R = 1$) is a linear combination of expectations. Once we have optimized and fixed the output weights W_i^M, we can form the linear combination as the QRC prediction without obtaining the individual expectations $\langle X^{(i)} \rangle_k, \langle Y^{(i)} \rangle_k$ and $\langle Z^{(i)} \rangle_k$. Using this trick, the first implementation scheme only requires preparation of the $N_m L$ circuits once and $N_m SLn$ measurements. The QND implementation only requires preparation of the N_m circuiits once and $N_m Sn$ measurements. That is, this trick allows us to increase the number of observables in the QRC output without increasing the number of measurements. We remark that when training the output weights, we still need to obtain the individual expectations $\langle X^{(i)} \rangle_k, \langle Y^{(i)} \rangle_k$ and $\langle Z^{(i)} \rangle_k$. Hence, this trick should only be used after training and fixing the output weights.

The unitary trick exploits the fact that $W_i^Z Z^{(i)} + W_i^Y Y^{(i)} + W_i^X X^{(i)} = C_i V_i^\dagger Z^{(i)} V_i$, where

$$\begin{aligned}
C_i &= \sqrt{(W_i^X)^2 + (W_i^Y)^2 + (W_i^Z)^2}, \\
V_i &= \begin{bmatrix} \frac{\sqrt{(W_i^X)^2+(W_i^Y)^2}}{M_{i1}} & -\frac{W_i^X - \iota W_i^Y}{M_{i2}} \\ \frac{\sqrt{(W_i^X)^2+(W_i^Y)^2}}{M_{i2}} & \frac{W_i^X - \iota W_i^Y}{M_{i1}} \end{bmatrix}, \\
M_{i1} &= \sqrt{2(C_i^2 - W_i^Z C_i)}, \\
M_{i2} &= \sqrt{2(C_i^2 + W_i^Z C_i)}.
\end{aligned}$$

In particular, we can re-express V_i as an arbitrary single-qubit rotational gate $U_3(\theta_i, \phi_i, \lambda_i)$,

$$U_3(\theta_i, \phi_i, \lambda_i) = \begin{bmatrix} \cos(\theta_i/2) & -e^{\iota\lambda_i}\sin(\theta/2) \\ e^{\iota\phi_i}\sin(\theta_i/2) & e^{-\iota(\phi_i+\lambda_i)}\cos(\theta_i/2), \end{bmatrix} \tag{4.6}$$

where $\theta_i/2 = \cos^{-1}\left(\frac{\sqrt{(W_i^Y)^2+(W_i^Z)^2}}{M_{i1}}\right)$, $\phi_i = 0$, and $\lambda_i = -\cos^{-1}\left(\frac{W_i^X}{\sqrt{(W_i^X)^2+(W_i^Y)^2}}\right)$ if $W_i^Y \geq 0$ and $\lambda_i = \cos^{-1}\left(\frac{W_i^X}{\sqrt{(W_i^X)^2+(W_i^Y)^2}}\right)$ if $W_i^Y < 0$. The single-qubit gate (4.6) can be implemented on current NISQ machines. We can now express the linear QRC output by $\hat{y}_k = \sum_{i=1}^n \sum_{S=\{X,Y,Z\}} W_i^M \langle M^{(i)} \rangle_k = \sum_{i=1}^n C_i \mathrm{Tr}(V_i^\dagger Z^{(i)} V_i \rho_k)$. To estimate the linear QRC output $\hat{y}_k$, we can apply the single-qubit rotational gates (4.6) on each of the i-th qubit before measuring all the qubits.

We remark that applying the rotational gates (4.6) at each time step before measuring all the qubits is equivalent to augmenting the QRC dynamics by the unitary $\otimes_{i=1}^n V_i$. That is, the original QRC dynamics $\rho_{k+1} = (1-\epsilon)(u_k U_0 \rho_k U_0^\dagger + (1-u_k) U_1 \rho_k U_1^\dagger) + \epsilon\sigma$ is augmented to $\rho_{k+1} = (\otimes_{i=1}^n V_i)\left((1-\epsilon)(u_k U_0 \rho_k (U_0)^\dagger + (1-u_k) U_1 \rho_k (U_1)^\dagger) + \epsilon\sigma\right)(\otimes_{i=1}^n V_i)^\dagger$.

4.4 Proof-of-principle experiments

Two temporal information processing problems are posed to learn the I/O relationship based on the given I/O data u, y. The first is the multi-step ahead prediction problem, in which we are given inputs $\{u_1, \ldots, u_L\}$ and the corresponding outputs $\{y_1, \ldots, y_L\}$. The first $L_T < L$ input-output data is the training data. The goal is to use the training data to optimize the QRC readout parameters via ordinary least squares, so that the QRC output match the target output for $L_T + 1, \ldots, L$. The second problem is the map emulation problem, that is to optimize the QRC readout parameters to emulate an I/O map using p different I/O train data of length L', so that the total number of train data is pL'. When given a previously unseen input, the task is for the QRC to approximate the corresponding output of the target I/O map.

Five nonlinear tasks are chosen to test different computational aspects of the QRC proposal. Tasks I-IV have the fading memory property. Tasks I and II test the QRC's ability to learn high-dimensional nonlinear maps. Both tasks are LRPO governed by [26, 102],

$$\begin{cases} x_k = Ax_{k-1} + Bu_k \\ y_k = \bar{h}(x_k), \end{cases}$$

where $A \in \mathbb{R}^{2000\times 2000}$ and $B \in \mathbb{R}^{2000}$. To have the uniform convergence and fading memory properties, we re-scale the maximum singular value so that $\sigma_{\max}(A) = 0.5$ for Task I and $\sigma_{\max}(A) = 0.99$ for Task II, meaning that Task II retains the initial condition and past inputs for a longer time duration. The sparsity of A determines the pairwise correlation between state elements. We set A to be a full (dense) matrix for Task I and 95% sparse for Task II using the Scipy 'sparse' command [125]. The readout function $\bar{h}$ is a degree 2 polynomial in the state elements.

Task III is SAS, a recently proposed classical reservoir computing model that achieves good performance in chaotic system modeling [26], described by

$$\begin{cases} x_k = p(u_k)x_{k-1} + q(u_k) \\ y_k = \mathbf{w}^T x_k, \end{cases}$$

where $p(u_k) = \sum_{j=0}^{4} A_j u_k^j$ and $q(u_k) = \sum_{j=0}^{2} B_j u_k^j$ are matrix-valued polynomials in the input u_k, $A_j \in \mathbb{R}^{700\times 700} \bigoplus \mathbb{R}^{700\times 700}$ and $B_j \in \mathbb{R}^{700\times 1} \bigoplus \mathbb{R}^{700\times 1}$, where $\bigoplus$ is the direct sum. To have the uniform convergence and fading memory properties, we re-scale A_j's maximum singular values so that $\sigma_{\max}(A_j) < \frac{1}{3}$ for all j. Task IV tests the short-term memory by emulating a Volterra series with kernel order 5 and memory 2 [102],

$$y_k = \mathbf{w}_c + \sum_{i=1}^{5} \sum_{j_1,\ldots,j_i=0}^{2} \mathbf{w}_i^{j_1,\ldots,j_i} \prod_{k=1}^{i} u_{k-j_k}.$$

For the first three tasks, elements of A, A_j, B, B_j and $\mathbf{w}$ are uniformly randomly sampled from $[-1,1]$. The coefficients of readout function $\bar{h}$ in Task II and the kernel coefficients $\mathbf{w}_i^{j_1,\ldots,j_i}$ and $\mathbf{w}_c$ in Task IV are also sampled from the same distribution.

Task V is a long-term memory map for testing the capability of the QRC beyond its theoretical guarantee. Task V is a missile moving with a constant velocity in the horizontal plane [107],

$$\begin{cases} \dot{x}_1 = x_2 - 0.1\cos(x_1)(5x_1 - 4x_1^3 + x_1^5) - 0.5\cos(x_1)u \\ \dot{x}_2 = -65x_1 + 50x_1^3 - 15x_1^5 - x_2 - 100u, \end{cases}$$

with $y = x_2$. This missile dynamics is simulated using the $(4,5)$ Runge-Kutta formula in MATLAB, with a sampling time of $\tau = 1/80$ for 1 second.

We implement four distinct QRCs from the subclass described in Sec. 4.2 on three IBM superconducting quantum processors [44]. Each QRC consists of a single subsystem ($N_s = 1$ in (4.1)) with a linear output function ($R = 1$ in (4.4)). Hereafter, we drop the subsystem index K. A 4-qubit and a 10-qubit QRCs are implemented on the 20-qubit Boeblingen device; qubits with lower gate errors and longer coherence times are chosen. The 5-qubit Ourense and Vigo devices are used for two distinct 5-qubit QRs. These 5-qubit quantum devices admit simpler qubit couplings but lower gate errors than the 20-qubit Boeblingen device; see Appendix A.4 for hardware specifications. Through comparison among the four QRCs, we can investigate the impact of the size of QRCs, the complexity of quantum circuits implementing the QRC dynamics and the intrinsic hardware noise on the QRC's emulation performance.

4.4.1 Quantum circuits for QRCs

We require the uniformly convergent QRCs to forget initial conditions for approximating fading memory maps. Traditionally, initial conditions are washed-out with a sufficiently long input sequence. Here we bypass the washout by choosing $\sigma = (|0\rangle\langle 0|)^{\otimes n}$ and U_0 so that $|0\rangle^{\otimes n}$ is the steady state of (4.1) under $u_k = 1$, meaning that we can initialize the QRC circuits in $|0\rangle^{\otimes n}$. Furthermore, U_0 and U_1 should be different and hardware-efficient (e.g., easy to implement on the hardware) but sufficiently complex to produce non-trivial quantum dynamics. We choose a circuit schematics (also see Fig. 4.2(a) and (b)),

$$\begin{aligned} U_0(\boldsymbol{\theta}) &= \prod_{j=1}^{N_0} \left(U_3^{(j_t)}(\boldsymbol{\theta}_{j_t}) \mathrm{CX}_{j_c j_t} U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger \right), \\ U_1(\boldsymbol{\phi}) &= \bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{0_i}) \prod_{j=1}^{N_1} \left(\bigotimes_{i=1}^{n} U_3^{(i)}(\boldsymbol{\phi}_{j_i}) \mathrm{CX}_{j_c j_t} \right), \end{aligned} \tag{4.7}$$

where $\boldsymbol{\theta}_{j_t} = (\boldsymbol{\theta}_{j_t}^0, \boldsymbol{\theta}_{j_t}^1, \boldsymbol{\theta}_{j_t}^2)$ and $\boldsymbol{\phi}_{j_i} = (\boldsymbol{\phi}_{j_i}^0, \boldsymbol{\phi}_{j_i}^1, \boldsymbol{\phi}_{j_i}^2)$ are gate parameters, each independently and uniformly randomly sampled from $[-2\pi, 2\pi]$. Here $U_3^{(i)}$ is an arbitrary rotation on single qubit i [95] with inverse $U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger = U_3^{(j_t)}(-\boldsymbol{\theta}_{j_t}^0, -\boldsymbol{\theta}_{j_t}^2, -\boldsymbol{\theta}_{j_t}^1)$, and $\mathrm{CX}_{j_c j_t}$ is the CNOT gate with control qubit j_c and target qubit j_t; also see Sec. 2.3. These quantum gates are native to the aforementioned IBM superconducting quantum processors, meaning that no further decomposition into simpler gates is required to implement these chosen gates [44]. The numbers of layers N_0 and N_1 are sufficiently large to couple all qubits linearly while respecting the coherence limits of these devices. Here, linear coupling means that qubit 0 is coupled to qubit 1 via CNOT, qubit 1 is coupled to qubit 2 via CNOT and so on, but qubit 2 might not directly be coupled to qubit 0 via CNOT. Owing to the more flexible qubit couplings in the Boeblingen device, circuits implementing the 4-qubit and 10-qubit QRCs have more gates and random parameters than the 5-qubit QRCs'.

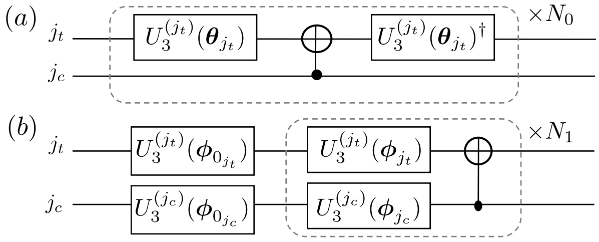

(a)
j_t
j_c
$U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})$
$U_3^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger$
$\times N_0$
(b)
j_t
j_c
$U_3^{(j_t)}(\boldsymbol{\phi}_{0_{j_t}})$
$U_3^{(j_c)}(\boldsymbol{\phi}_{0_{j_c}})$
$U_3^{(j_t)}(\boldsymbol{\phi}_{j_t})$
$U_3^{(j_c)}(\boldsymbol{\phi}_{j_c})$
$\times N_1$

Figure 4.2: Quantum circuit schematics for (a) $U_0(\boldsymbol{\theta})$ and (b) $U_1(\boldsymbol{\phi})$ employed in proof-of-principle experiments, described by (4.7). Here j_t and j_c are the target and control qubits, respectively. The unitaries $U_0(\boldsymbol{\theta}), U_1(\boldsymbol{\phi})$ consist of N_0, N_1 layers of highlighted gate operations, with each layer acting on a different qubit pair (j_t, j_c).

For the 4-qubit and 10-qubit QRCs on the Boeblingen device, we choose the number of

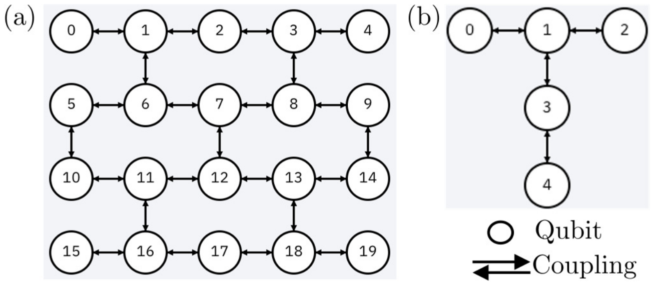


Figure 4.3: Qubit coupling maps of the IBM superconducting quantum processors. (a) The 20-qubit Boeblingen device. (b) Both the 5-qubit Ourense and Vigo devices.

layers $N_0 = N_1 = 5$ in (4.7). For the 5-qubit Ourense QRC, we implement a simpler form of (4.7), given by $U_0 = \prod_{j=1}^{4} \mathrm{CX}_{j_c j_t}$ and $U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{5} U_3^{(i)}(\boldsymbol{\phi}_i)$.

To implement a different QRC dynamics on the 5-qubit Vigo device, we choose

$$U_0(\boldsymbol{\theta}) = \prod_{j=1}^{3} \left(R_Y^{(j_t)}(\boldsymbol{\theta}_{j_t}) \mathrm{CX}_{j_c j_t} R_Y^{(j_t)}(\boldsymbol{\theta}_{j_t})^\dagger \right), \quad U_1(\boldsymbol{\phi}) = \bigotimes_{i=1}^{5} R_X^{(i)}(\boldsymbol{\phi}_i).$$

Here $R_Y^{(i)}$ and $R_X^{(i)}$ are rotational Y and X gates on qubit i, respectively. For all QRCs, natively coupled control and target qubits for the CNOT gates are chosen, meaning that a CNOT gate can be directly applied to the qubit pair without additional gate operations. See Fig. 4.3 for the device qubit coupling maps and Appendix A.5 for the circuit details.

4.4.2 Experimental implementation

We report on experiments demonstrating the first implementation scheme described in Sec. 4.3. We choose a sufficiently large $N_m = 1024$ and $\epsilon = 0.1$ for a moderate short-term memory. To estimate $\langle Z^{(i)} \rangle_k$ at time k, each of the N_m circuits implementing the QRCs on the Boeblingen device and the 5-qubit QRs are run for $S = 1024$ and $S = 8192$ shots, respectively. These shot numbers are chosen according to circuit execution times of the devices, so that the experiments can finish in a reasonable amount of time.

We apply the four QRCs to the five nonlinear tasks on the multi-step ahead prediction and map emulation problems. To implement the same washout as for the QRCs for each target map, we inject a constant input sequence $u_k = 1$ of length 50 followed by train and test inputs uniformly randomly sampled from $u_k \in [0, 1]$. This change in the input statistics leads to a transitory target output response. We remove the associated transients by discarding the first four target input-output data and the corresponding QRC experimental data, see Appendix A.6 for all data. For the multi-step ahead problem, training and testing timesteps run from $k = 5$ to $L_T = 23$ and $L_T + 1 = 24$ to $L = 30$, respectively. For the map emulation problem, $p = 2$ train input-output pairs running from $k = 5$ to $L' = 24$ are used, followed by one unseen test input-output pair with the same time steps. The number of train and test data in our proof-of-principle experiments is limited by the length of quantum circuits allowed on the IBM quantum processors. Furthermore, these cloud-based quantum processors are shared among users, making continuous experiments infeasible and durations of experiments lengthy. Yet our work indicates that despite these current limitations, NISQ devices can demonstrate learning of input-output maps and supports QRC as a viable intermediate application of NISQ machines on the road to full-fledged quantum devices equipped with quantum error correction.

To harness the flexibility of the QRC approach, a multi-tasking technique is used, in which the four QRCs are evolved and the estimates of $\langle Z^{(i)} \rangle_k$ for all time steps are recorded once, whereas the readout parameters are optimized independently for each task. We evaluate and compare the task performance of QRCs using the normalized mean-squared error between prediction $\hat{y}$ and target y, computed as

$$\mathrm{NMSE} = \sum_{k=L_T+1}^{L} |y_k - \hat{y}_k|^2 / \Delta_y^2,$$

where $\mu = \frac{1}{L-L_T} \sum_{k=L_T+1}^{L} y_k$, $\Delta_y^2 = \sum_{k=L_T+1}^{L} (y_k - \mu)^2$. While the success of experimental demonstration of hybrid quantum-classical algorithms often requires error mitigation techniques to reduce the effect of decoherence [126, 127], we remark that our results are obtained without any process or readout error mitigation.

4.4.3 QRC task performance

As the number of qubits increases, the 10-qubit Boeblingen QRC is expected to perform better than other QRCs. For the multi-step ahead prediction problem, we observe that

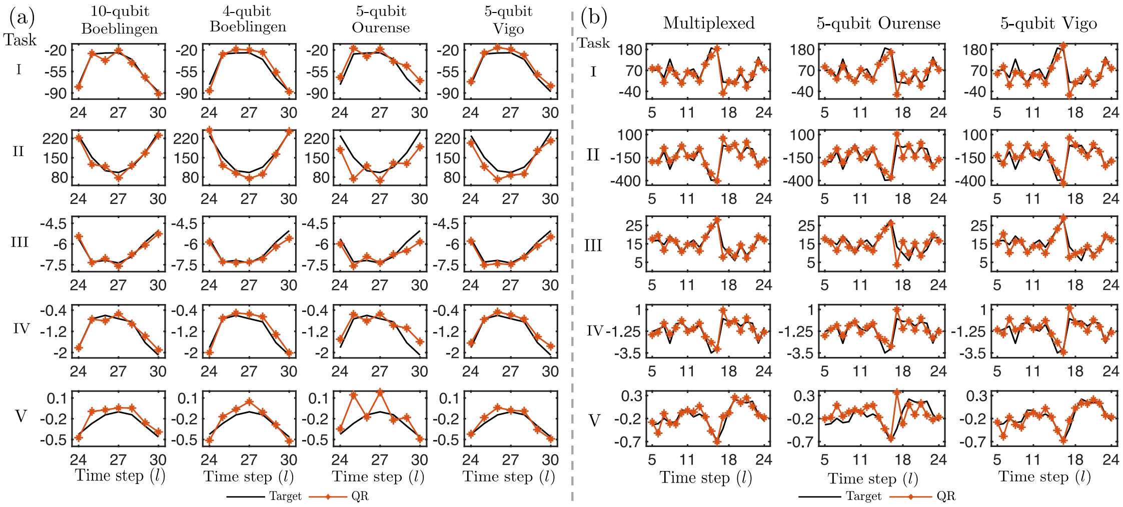


Figure 4.4: The QRC's predicted outputs for (a) the multi-step prediction problem and (b) the map emulation problem. Rows and columns in (a) correspond to different tasks and QRCs, respectively. The first column in (b) corresponds to the multiplexed QRC.

two qubits in the 10-qubit Boeblingen QRC experienced significant time-varying deviations between the experimental data and simulation results on the Qiskit simulator; see Appendix A.7. To remedy this issue, we set the corresponding readout parameters to be zeros. The resulting 10-qubit Boeblingen QRC (with NMSE<0.08) outperforms other QRCs with a smaller number of qubits on the first four tasks, and achieves an almost two-fold performance improvement on Tasks II and III; see Table 4.1 for all NMSEs on the multi-step ahead prediction problem. The 10-qubit Boeblingen QRC predicted outputs follow the target outputs relatively closely as shown in Fig. 4.4(a). The 5-qubit Ourense QRC admits very simple dynamics, in the sense that U_0 is only a product of CNOT gates and U_1 consists only of single qubit rotational gates. On the other hand, the 5-qubit Vigo QRC has more gate operations and gate parameters. The 5-qubit Ourense QRC is outperformed by the 5-qubit Vigo QRC in all tasks. Considering that the Ourense and Vigo devices have similar noise characteristics and the same qubit coupling map, this suggests the QRC performance can be improved by choosing a more complex quantum circuit, in the sense of having a longer gate sequence.

The 10-qubit Boeblingen QRC performs better on all tasks than the 5-qubit QRCs except on Task V. This could be due to the impact of the higher noise level in the Boeblingen

Table 4.1: NMSEs on the multi-step ahead prediction problem.

Task	10-qubit Boeblingen	4-qubit Boeblingen	5-qubit Ourense	5-qubit Vigo
I	0.051	0.088	0.24	0.070
II	0.072	0.12	0.68	0.22
III	0.043	0.10	0.25	0.081
IV	0.079	0.092	0.34	0.11
V	0.47	0.41	2.3	0.20

Table 4.2: NMSEs on the map emulation problem.

Task	Multiplexed QRC	5-qubit Ourense	5-qubit Vigo
I	0.20	0.26	0.32
II	0.13	0.27	0.23
III	0.16	0.46	0.26
IV	0.25	0.30	0.36
V	0.20	1.1	0.17

device and the fact that the output sequence is generated by a map that is not known to be fading memory. Our universal class of QRCs can exploit the property of spatial multiplexing as initially proposed in [55]; also see Fig. 4.5. Outputs of distinct and non-interacting 5-qubit QRCs can be combined linearly to harness the computational features of both members. Since the combined Ourense and Vigo devices have 10 qubits overall as with the 10-qubit Boeblingen QRC but with lower noise levels, it would be meaningful to combine the 5-qubit Vigo and Ourense QRCs via spatial multiplexing on the map emulation problem. The results of this multiplexing is summarized in Table 4.2.

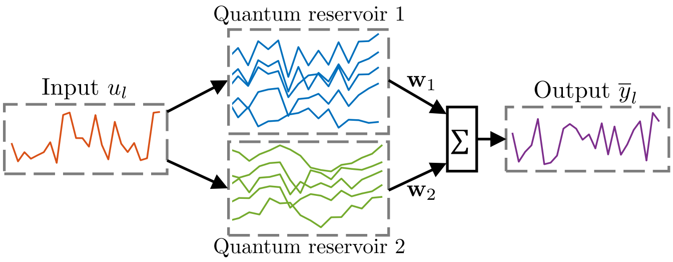


Figure 4.5: The spatial multiplexing schematic. The same input sequence is injected into two distinct 5-qubit QRCs. The internal states $\mathrm{Tr}(\rho_k Z^{(i)})$ of the two QRCs are linearly combined to form a single output.

The combination of two 5-qubit QRCs as discussed above achieves NMSE $= 0.20, 0.13, 0.16, 0.25, 0.20$ for the five tasks without any readout or process error mitigation. The predicted multiplexed QRC outputs corresponding to the unseen inputs follow the target outputs relatively closely as shown in Fig. 4.4(b). Without spatial multiplexing, the 5-qubit Ourense or the 5-qubit Vigo QRC show a worse performance in the first four tasks. The spatial multiplexed 5-qubit QRC combines computational features from the constituent QRCs and can achieve comparable performance to the individual members as well as gaining an almost two-fold performance boost on Tasks II and III. We anticipate that spatial multiplexing of QRCs with more complex circuit structures and a larger number of qubits can lead to further performance improvements.

4.5 Discussion

Current gate durations in the IBM superconducting NISQ devices can be as short as 35.5 ns for single-qubit gates and 427 ns for CNOT gates (see Appendix A.4 for all gate durations). Consider the scenario when QND measurements and a large number of identical quantum hardware are available, so that the N_m circuits can be run in parallel (recall that we can set $S = 1$). These fast gate operations suggest the QRC's potential for real-time signal processing.

However, the time it takes to measure the qubits poses a bottleneck. Current readout durations are of one to ten microseconds on these superconducting devices [128]. Such readout durations are very long considering that most qubit coherence times are only of tens to a hundred microseconds. To minimize the time required for measurements, instead of performing QND measurements and resetting the ancilla qubits at each timestep, we can allocate L ancilla qubits for each QRC 'system' qubit and defer their measurements at the last timestep. In this setting, measurements of these ancilla qubits can be performed in parallel and the QRC can operate in a microsecond time-scale. Since this requires a large number of qubits with sufficiently long qubit coherence times, real-time processing for QRCs is more plausible on full-fledged quantum machines.

Recently, there are many different quantum computing platforms being studied in the QRC context. For instance, the work [129] proposes a photonic QRC, which has the potential to implement fast input encoding and measurements. Such fast quantum devices may be more feasible to facilitate real-time QRC in the near future.

4.6 Conclusion

We have proposed a class of QRCs endowed with the universality property that is implementable on available noisy gate-model quantum hardware for temporal information processing. Our approach can harness arbitrary but fixed quantum circuits native to noisy quantum processors, without precise tuning of the circuit parameters. Our theoretical analysis is supported by proof-of-concept experiments on current superconducting quantum devices, demonstrating that small-scale noisy quantum reservoirs can perform non-trivial nonlinear temporal processing tasks under current hardware limitations, in the absence of readout and process error mitigation techniques. We also detail more efficient implementation schemes of our QRC proposal that could enable QRC's potential for fast and scalable temporal processing. It is a future work of interest to realize these more efficient protocols on quantum hardware. Our work indicates that quantum reservoir computing can serve as a viable intermediate application of NISQ devices on the road to full-fledged quantum computers.

Our approach is scalable in the number of qubits by offloading computations to noisy quantum systems and utilizing classical algorithms with a linear (in the number of qubits) computational cost to process sequential data. Guided by our theory, we applied the spatial multiplexing technique initially proposed in [55], and demonstrate experimentally that exploiting distinct computational features of multiple small noisy quantum reservoirs can lead to a computational boost. As NISQ hardware becomes increasingly accessible and the noise level is continually reduced, we hope that the quantum reservoir approach will find useful applications in a broad range of scientific disciplines that employ temporal information processing. We also hope for useful applications to be possible even with a noise level above the threshold for continuous quantum error correction.

Chapter 5

Interconnected reservoir computers

In chapters 3 and 4, we focus on specific quantum implementations of reservoir computers for approximating nonlinear fading memory I/O maps. In the following chapters, we abstract away from physical implementations to theoretical explorations of reservoir computers. In particular, in this chapter, we explore interconnected RCs in closed-loop configurations for emulating I/O maps. In the next chapter, we show that RCs with output-feedback implement stationary and ergodic infinite-order nonlinear autoregressive models.

This chapter is based on the publication [130]. We focus on approximating target I/O maps in a black-box or data-driven fashion, i.e., approximation is solely based on I/O data collected from the target I/O maps. Example nonlinear models for black-box system identification include NARMAX [131], the Volterra series [25] and block-oriented models [76, 77]. The use of closed-loop structures, such as in the Wiener-Hammerstein feedback model, is motivated by modeling systems that exhibit nonlinear feedback behavior [77]. To approximate I/O maps, we require candidate models with closed-loop structures to asymptotically forget their initial conditions. This leads to the main contributions of this chapter:

- We derive a small-gain theorem to ensure that interconnected dynamical systems are UC and induce a well-posed I/O map in the sense of [132].
- We introduce interconnected RCs as nonlinear models with closed-loop structures for emulating I/O maps, expanding the toolbox of available nonlinear models.

To derive this small-gain theorem, we introduce the uniform output convergence (UOC) and the uniform input-to-output convergence (UIOC) properties. Roughly speaking, a UOC system has a unique reference state solution with its reference output defined and bounded both backwards and forwards in time. All other outputs asymptotically converge to the reference output, independent of their initial conditions. The UIOC property adapts the uniform input-to-state convergence (UISC) property [133] to output-feedback interconnected systems. A UIOC system is UOC and the perturbation in its reference outputs are asymptotically bounded by a nonlinear gain on the input perturbation. We present a small-gain theorem for output-feedback interconnected systems to be UIOC, and that the closed-loop system induces a well-posed I/O map in the sense of [132].

Our UIOC small-gain theorem is based on a small-gain theorem for time-varying discrete-time systems in the uniform input-to-output stability (UIOS) framework, also presented herein. This latter small-gain result is used to establish the UIOC small-gain theorem for interconnected UOC systems. Small-gain criteria for time-invariant discrete-time input-to-state stable systems can be found in [134, 135], however, they do not carry over to time-varying systems [136]. On the other hand, small-gain criteria for time-varying continuous-time interconnected systems in the UIOS framework have been established [99, 137–139]. Ref. [140] develops a generalized small-gain theorem and can be applied to recover the previous results derived in [99, 137, 138]. Ref. [138] is based on [141, Lemma 3, Prop. 2.5], which concerns continuous-time systems, and is not obvious that it is immediately applicable to our setting. Here, we adopt the techniques in [140] to establish a UIOS small-gain theorem for output-feedback interconnected time-varying discrete-time systems. In particular, we establish a link to bridge the continuous-time results of [140] with our time-varying discrete-time setting.

This chapter is organized as follows. Sec. 5.1 introduces relevant stability concepts. Sec. 5.2 presents our main UIOC small-gain theorem. We also present the UIOS small-gain theorem, an intermediate result to arrive at the UIOC small-gain theorem. In Sec. 5.3, we introduce interconnected RCs as candidate models with closed-loop structures and numerically demonstrate their efficacy. We provide concluding remarks in Sec. 5.4.

Notation: we introduce some additional notations for this chapter. For a sequence u on $\mathbb{Z}$, $\|u_{[k_0,k]}\| \coloneqq \sup_{k_0 \leq j \leq k} \|u(j)\|$ for any $k_0 \in \mathbb{Z}$ and $k \geq k_0$. We write $|u_{[k_0,k]}|$ for a real-valued sequence u. For any sequences u_1, u_2 on $\mathbb{Z}$, $u = (u_1, u_2)$ is given by $u(k) = (u_1(k), u_2(k))$ for all $k \in \mathbb{Z}$.

5.1 Stability concepts

This section defines the uniform output convergence (UOC) (Definition 5.2) and the uniform input-to-output convergence (UIOC) (Definition 5.3) properties. See Table 5.1 for a summary of relevant stability definitions. We first set some preliminaries.

For $k \in \mathbb{Z}$, consider a time-varying discrete-time system,

$$\begin{cases} x(k+1) = f(k, x(k), u(k)), \\ \quad\; y(k) = h(k, x(k), u(k)), \end{cases} \tag{5.1}$$

where $x(k) \in \mathbb{R}^{n_x}$ is the state, $u(k) \in \mathbb{R}^{n_u}$ is the input and $y(k) \in \mathbb{R}^{n_y}$ is the output. Throughout this chapter, we assume that $u \in l^{\infty}_{n_u}$ (recall that $l^{\infty}_{n_u}$ is the set of infinite sequences of vectors in $\mathbb{R}^{n_u}$) and define all stability concepts with respect to inputs $u \in l^{\infty}_{n_u}$. We also assume that for each $k \in \mathbb{Z}$, $\|f(k, x(k), u(k))\| < \infty$ and $\|h(k, x(k), u(k))\| < \infty$. The last two conditions ensure that the system is non-singular at any time and for any initial condition.

The following definition of UOC adapts the uniform convergence (UC) property with respect to inputs $l^\infty_{n_u}$ ([22, Definition 3] or Definition 2.1 in Sec. 2.4.1) to systems with output of the form (5.1).

Definition 5.1. *For any* $u \in l^\infty_{n_u}$*, a solution* $x^*(k)$ *to* (5.1) *and its corresponding output* $y^*(k) = h(k, x^*(k), u(k))$ *are a reference state solution and the corresponding reference output, respectively, if they are defined for all* $k \in \mathbb{Z}$*, with* $\|x^*\|_\infty := \sup_{k\in\mathbb{Z}} \|x^*(k)\| < \infty$ *and* $\|y^*\|_\infty := \sup_{k\in\mathbb{Z}} \|y^*(k)\| < \infty$*.*

Definition 5.2. *System* (5.1) *is uniformly output convergent (UOC) if, for any input* $u \in l^\infty_{n_u}$*,*

(i) There exists a unique reference state solution x^* *with its corresponding reference output* y^**.*

(ii) There exists $\beta \in \mathcal{KL}$ *independent of* u *such that, for any* $k, k_0 \in \mathbb{Z}$ *with* $k \geq k_0$ *and* $x(k_0) \in \mathbb{R}^{n_x}$*,*

$$\|y^*(k) - y(k)\| \leq \beta(\|x^*(k_0) - x(k_0)\|, k - k_0). \tag{5.2}$$

We remark that the UOC property reduces to the the UC property when the output is the state, i.e., $y(k) = x(k)$. We emphasize that in this chapter, all gain functions (e.g., β in (5.2)) are independent of the input $u \in l^\infty_{n_u}$.

Just as UC systems induce their unique filters (see Sec. 2.4.3 for a detailed discussion on filters), so do UOC systems. Explicitly, a UOC system induces a filter (or an I/O map) $M_{f,h} : l^\infty_{n_u} \to (\mathbb{R}^{n_y})^{\mathbb{Z}}$ that depends on f and h, such that $M_{f,h}(u)|_k := y^*_k = h(x^*_k)$, where x^* and y^* are the reference state solution and the reference output. The filter $M_{f,h}$ is time-invariant and causal by construction. Furthermore, using the same argument in

Theorem 2.3, this filter can be constructed as

$$\begin{aligned}
y^*_k = M_{f,h}(u)|_k &= h(k, f(k-1, x(k-1), u(k-1)), u(k)) \\
&= h(k, f(k-1, f(k-2, x(k-2), u(k-2)), u(k-1)), u(k)) \\
&\quad \vdots \\
&= \mathcal{F}(u(k), u(k-1), u(k-2), \ldots),
\end{aligned}$$

where

$$\begin{aligned}
&\mathcal{F}(u(k), u(k-1), u(k-2), \ldots) \\
&= \lim_{k_0 \to -\infty} h(k, f(\ldots f(k_0+1, f(k_0, x(k_0), u(k_0)), u(k_0+1)) \ldots, u(k)).
\end{aligned}$$

Using (5.2) and applying a similar argument as in the proof of Theorem 2.3, we can see that the above limit exists and is independent of the initial condition $x(k_0)$. Note that Theorem 2.3 is derived for UC systems. While in Theorem 2.3 we require $h(\cdot)$ to be uniformly continuous, we do not need this condition for the above limit to exist for UOC systems. This is because (5.2) directly upper bounds $\|y^*(k) - y(k)\|$ by a $\mathcal{KL}$ gain whereas a UC system only has an upper bound for $\|x^*(k) - x(k)\|$ (see (2.6)) and $\|y^*(k) - y(k)\|$ for a UC system is not necessarily upper bounded by a $\mathcal{KL}$ gain.

The UIOC property further extends the UOC property, and ensures that the perturbation in the reference outputs is asymptotically bounded by a nonlinear gain of the input perturbation. The following definition of UIOC is a discrete-time analogue of the UISC property defined in [133, Def. 3] adapted to systems with output of the form (5.1).

Definition 5.3. *System* (5.1) *is uniformly input-to-output convergent (UIOC) if it is UOC and for any* $u, \overline{u} \in l^\infty_{n_u}$*, with the reference state solution* x^* *and its reference output* y^* *associated to* u*, and any solution* $\overline{x}(k)$ *with any initial condition* $\overline{x}(k_0)$ *and the corresponding output* $\overline{y}(k)$ *associated to* $\overline{u}$*, there exists* $\beta \in \mathcal{KL}, \gamma \in \mathcal{K}$ *such that, for all* $k_0, k \in \mathbb{Z}$ *with* $k \geq k_0$*,*

$$\|y^*(k) - \overline{y}(k)\| \leq \max\{\beta(\|x^*(k_0) - \overline{x}(k_0)\|, k - k_0), \gamma\left(\|(u - \overline{u})_{[k_0,k]}\|\right)\}. \tag{5.3}$$

If $y(k) = x(k)$*, then system* (5.1) *is uniformly input-to-state convergent (UISC) if it is UC* *and*

$$\|x^*(k) - \overline{x}(k)\| \leq \max\{\beta(\|x^*(k_0) - \overline{x}(k_0)\|, k - k_0), \gamma\left(\|(u - \overline{u})_{[k_0,k-1]}\|\right)\}. \tag{5.4}$$

We conclude this section by summarizing the aforementioned stability concepts and their acronyms in Table 5.1.

Table 5.1: Summary of stability concepts and their acronyms.

Stability concept	Acronym	Definition
Uniformly output convergent	UOC	Definition 5.2
Uniformly convergent	UC	Definition 2.1
Uniformly input-to-output convergent	UIOC	Definition 5.3
Uniformly input-to-state convergent	UISC	Definition 5.3

5.2 A UIOC small-gain theorem

In this section, we present our main UIOC small-gain theorem (Theorem 5.1). For $k \in \mathbb{Z}$, consider the interconnected system (see Fig. 5.1),

$$\begin{cases} x_1(k+1) = f_1(x_1(k), v_1(k), u_1(k)) \\ \quad y_1(k) = h_1(x_1(k), v_1(k), u_1(k)), \end{cases}$$
$$\begin{cases} x_2(k+1) = f_2(x_2(k), v_2(k), u_2(k)) \\ \quad y_2(k) = h_2(x_2(k), v_2(k), u_2(k)), \end{cases} \tag{5.5}$$

with interconnections $v_1(k) = y_2(k)$ and $v_2(k) = y_1(k)$.

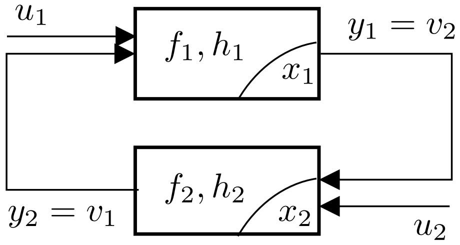


Figure 5.1: Schematic of the interconnected system (5.5).

For subsystems $j = 1, 2$, $x_j(k) \in \mathbb{R}^{n_{x_j}}$ are states, $y_j(k) \in \mathbb{R}^{n_{y_j}}$ are outputs and $u_j(k) \in \mathbb{R}^{n_{u_j}}$ are inputs with $u_j \in l^\infty_{n_{u_j}}$. Throughout this chapter, we assume that the interconnected system (5.5) is well-posed [99]. That is, for any $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$, any initial conditions $x_j(k_0)$ and any inputs u_1, u_2, there exists a unique solution $y(k) := (y_1(k), y_2(k)) \in \mathbb{R}^{n_{y_1}+n_{y_2}}$ solving the algebraic equations $y_1(k) = h_1(x_1(k), y_2(k), u_1(k))$ and $y_2(k) = h_2(x_2(k), y_1(k), u_2(k))$.

For a well-posed system (5.5), let $x(k) := (x_1(k), x_2(k)) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$ be the closed-loop solution to input $u(k) := (u_1(k),\ u_2(k)) \in \mathbb{R}^{n_{u_1}+n_{u_2}}$, starting at $x(k_0) = (x_1(k_0), x_2(k_0))$. Note that the closed-loop system is causal by definition.

We also assume that the subsystems in (5.5) are UOC. Each UOC subsystem induces an I/O map $\mathcal{F}_j : l^\infty_{n_{v_j}} \times l^\infty_{n_{u_j}} \to l^\infty_{n_{y_j}}$ defined by $\mathcal{F}_j(v_j, u_j) = y_j^*$, where y_j^* is the reference output. By construction $\mathcal{F}_j$ is causal, meaning that for any $\tau \in \mathbb{Z}$ and any $v_j \in l^\infty_{n_{v_j}}, u_j \in l^\infty_{n_{u_j}}$,

$$\Pi_\tau \circ \mathcal{F}_j \circ \Pi_\tau(v_j, u_j) = \Pi_\tau \circ \mathcal{F}_j(v_j, u_j), \tag{5.6}$$

where $\Pi_\tau(v_j, u_j) = (v_j(k), u_j(k))$ for $k \leq \tau$ and zero otherwise. We say that system (5.5) induces a well-posed closed-loop I/O map if the algebraic equations $y_1 = \mathcal{F}_2(y_2, u_2)$ and $y_2 = \mathcal{F}_1(y_1, u_1)$ have a unique bounded solution $y^*_{cl} = (y^*_{1,cl}, y^*_{2,cl}) \in l^\infty_{n_{y_1}+n_{y_2}}$, and the closed-loop I/O map $\mathcal{F}(u) = y^*_{cl}$ is causal [132], where $u = (u_1, u_2) \in l^\infty_{n_{u_1}+n_{u_2}}$. We emphasize the difference between a well-posed system (5.5) and a well-posed closed-loop I/O map induced by (5.5).

We now state our main UIOC small-gain theorem.

Theorem 5.1. *Consider a well-posed system* (5.5) *with UOC subsystems* $j = 1, 2$. *For any inputs* $u_j \in l^\infty_{n_{u_j}}, v_j \in l^\infty_{n_{v_j}}$, *let* x_j^* *and* y_j^* *be the corresponding reference state solutions and outputs. For any other inputs* $\overline{u}_j, \overline{v}_j$ *with* $\overline{u}_j \in l^\infty_{n_{u_j}}$, *let* $\overline{x}_j$ *and* $\overline{y}_j$ *be any corresponding solutions and outputs with initial conditions* $\overline{x}_j(k_0)$. *Suppose that there exists* $\beta_j \in \mathcal{KL}$ *and* $\gamma_j^y, \gamma_j^u, \sigma_j, \sigma_j^u, \sigma_j^y \in \mathcal{K}$ *such that, for all* $k_0, k \in \mathbb{Z}$, $k \geq k_0$ *and any* $\overline{x}_j(k_0) \in \mathbb{R}^{n_{x_j}}$,

$$\begin{aligned}\|y_j^*(k) - \overline{y}_j(k)\| \leq \max\{&\beta_j(\|x_j^*(k_0) - \overline{x}_j(k_0)\|, k - k_0),\\ &\gamma_j^y(\|(v_j - \overline{v}_j)_{[k_0,k]}\|), \gamma_j^u(\|(u_j - \overline{u}_j)_{[k_0,k]}\|)\},\end{aligned} \tag{5.7}$$

$$\begin{aligned}\|x_j^*(k) - \overline{x}_j(k)\| \leq \max\{&\sigma_j(\|x_j^*(k_0) - \overline{x}_j(k_0)\|),\\ &\sigma_j^y(\|(v_j - \overline{v}_j)_{[k_0,k-1]}\|), \sigma_j^u(\|(u_j - \overline{u}_j)_{[k_0,k-1]}\|)\}.\end{aligned} \tag{5.8}$$

If $\gamma_1^y \circ \gamma_2^y(s) < s$ *(or equivalently* $\gamma_2^y \circ \gamma_1^y(s) < s$ *[142, Chapter 8.1]) for all* $s > 0$*, for any* $k_0 \in \mathbb{Z}$ *and* $x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$*, the closed-loop solution and its output are bounded, i.e.,* $\sup_{k\geq k_0} \|x(k)\| < \infty$ *and* $\sup_{k\geq k_0} \|y(k)\| < \infty$*. Furthermore, the closed-loop system* (5.5) *induces a well-posed closed-loop I/O map and system* (5.5) *is UIOC.*

We remark that although Theorem 5.1 ensures the closed-loop solution $x(k)$ is bounded, in general, it does not guarantee the closed-loop system (5.5) is UISC, which would require additional conditions; see Corollary 5.2.

The main idea in the proof of Theorem 5.1 is to use the Banach fixed point theorem to prove that system (5.5) induces a well-posed closed-loop I/O map. To show that system (5.5) is UIOC, we apply a change-of-coordinate argument in which the system under consideration becomes time-varying of the form (5.10). We then apply a uniform input-to-output stability (UIOS) small-gain theorem (Theorem 5.2) to system (5.10). The UIOS small-gain theorem establishes the UIOC property of (5.5). We first define the UIOS property and state Theorem 5.2 whose full proof is given in Appendix B.1.

Definition 5.4. *System* (5.1) *is uniformly input-to-output stable (UIOS) if there exists* $\beta \in \mathcal{KL}$ *and* $\gamma \in \mathcal{K}$ *such that, for any* $u \in l_{n_u}^\infty$*,* $k, k_0 \in \mathbb{Z}$ *with* $k \geq k_0$ *and* $x(k_0) \in \mathbb{R}^{n_x}$*,*

$$\|y(k)\| \leq \max\left\{\beta\left(\|x(k_0)\|, k - k_0\right), \gamma\left(\left\|u_{[k_0,k]}\right\|\right)\right\}. \tag{5.9}$$

Theorem 5.2. *Consider a well-posed time-varying dynamical system*

$$\begin{aligned}&\begin{cases}\Delta x_1(k+1) = \tilde{f}_1(k, \Delta x_1(k), \Delta v_1(k), \Delta u_1(k))\\ \qquad \Delta y_1(k) = \tilde{h}_1(k, \Delta x_1(k), \Delta v_1(k), \Delta u_1(k)),\end{cases}\\ &\begin{cases}\Delta x_2(k+1) = \tilde{f}_2(k, \Delta x_2(k), \Delta v_2(k), \Delta u_2(k))\\ \qquad \Delta y_2(k) = \tilde{h}_2(k, \Delta x_2(k), \Delta v_2(k), \Delta u_2(k)),\end{cases}\\ &\quad \Delta v_1(k) = \Delta y_2(k),\ \ \Delta v_2(k) = \Delta y_1(k).\end{aligned} \tag{5.10}$$

For $j = 1, 2$, suppose that there exists $\beta_j \in \mathcal{KL}$ and $\gamma_j^y, \gamma_j^u, \sigma_j, \sigma_j^u, \sigma_j^y \in \mathcal{K}$ such that, for any $\Delta v_j, \Delta u_j$ with $\Delta u_j \in l^\infty_{n_{\Delta u_j}}$, $k_0, k \in \mathbb{Z}$ with $k \geq k_0$ and $\Delta x_j(k_0) \in \mathbb{R}^{n_{x_j}}$,

$$\|\Delta y_j(k)\| \leq \max\{\beta_j(\|\Delta x_j(k_0)\|, k - k_0), \gamma_j^y(\|\Delta v_{j_{[k_0,k]}}\|), \gamma_j^u(\|\Delta u_{j_{[k_0,k]}}\|)\}, \tag{5.11}$$

$$\|\Delta x_j(k)\| \leq \max\{\sigma_j(\|\Delta x_j(k_0)\|), \sigma_j^y(\|\Delta v_{j_{[k_0,k-1]}}\|), \sigma_j^u(\|\Delta u_{j_{[k_0,k-1]}}\|)\}. \tag{5.12}$$

If $\gamma_1^y \circ \gamma_2^y(s) < s$ for all $s > 0$, then for any $k_0 \in \mathbb{Z}$ and $\Delta x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$, the closed-loop solution and its output are bounded, i.e., $\sup_{k \geq k_0} \|\Delta x(k)\| < \infty$ and $\sup_{k\geq k_0} \|\Delta y(k)\| < \infty$. Furthermore, the closed-loop system (5.10) is UIOS; i.e., there exists $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}$ such that, for all $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$ and any $\Delta x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$,

$$\|\Delta y(k)\| \leq \max\{\beta(\|\Delta x(k_0)\|, k - k_0), \gamma(\|\Delta u_{[k_0,k]}\|)\}.$$

We now detail the proof for Theorem 5.1.

Proof of Theorem 5.1. For any fixed inputs $u_j \in l^\infty_{n_{u_j}}$, the I/O map induced by each subsystem is given by $\mathcal{F}_j^{u_j} : l^\infty_{n_{v_j}} \to l^\infty_{n_{y_j}}, \mathcal{F}_j^{u_j}(v_j) = y_j^*$. For any $v_j, \overline{v}_j \in l^\infty_{n_{v_j}}$, let $k_0 \to -\infty$ and take the supremum over $k \in \mathbb{Z}$ in (5.7),

$$\|\mathcal{F}_j^{u_j}(v_j) - \mathcal{F}_j^{u_j}(\overline{v}_j)\|_\infty \leq \gamma_j^y(\|v_j - \overline{v}_j\|_\infty). \tag{5.13}$$

Consider the composition $\mathcal{F}_1^{u_1} \circ \mathcal{F}_2^{u_2} : l^\infty_{n_{v_2}} \to l^\infty_{n_{v_2}}$. Applying inequality (5.13) twice, we have $\|\mathcal{F}_1^{u_1} \circ \mathcal{F}_2^{u_2}(v_2) - \mathcal{F}_1^{u_1} \circ \mathcal{F}_2^{u_2}(\overline{v}_2)\|_\infty \leq \gamma_1^y \circ \gamma_2^y(\|v_2 - \overline{v}_2\|_\infty) < \|v_2 - \overline{v}_2\|_\infty$. Therefore, $\mathcal{F}_1^{u_1} \circ \mathcal{F}_2^{u_2}$ is a strict contraction on $(l^\infty_{n_{v_2}}, \|\cdot\|_\infty)$. Its unique fixed point $y_{1,cl}^* \in l^\infty_{n_{v_2}}$ given by the Banach fixed-point theorem [143] is the reference output of subsystem $j = 1$. The corresponding reference output of subsystem $j = 2$ is $y_{2,cl}^* = \mathcal{F}_2^{u_2}(y_{1,cl}^*)$. A symmetric argument shows that $\mathcal{F}_2^{u_2} \circ \mathcal{F}_1^{u_1}$ is a strict contraction defined on $(l^\infty_{n_{v_1}}, \|\cdot\|_\infty)$.

To show that the closed-loop I/O map is causal, for any $\tau \in \mathbb{Z}$, consider $y_{1,\Pi_\tau}^* \in l^\infty_{n_{v_2}}$ the unique fixed point of $\mathcal{F}_1^{\Pi_\tau(u_1)} \circ \mathcal{F}_2^{\Pi_\tau(u_2)}$. Causality follows if $\Pi_\tau(y_{1,\Pi_\tau}^*) = \Pi_\tau(y_{1,cl}^*)$.

Re-express (5.6) as $\Pi_\tau \circ \mathcal{F}_j^{\Pi_\tau(u_j)} \circ \Pi_\tau(v_j) = \Pi_\tau \circ \mathcal{F}_j^{u_j}(v_j)$, we have

$$\begin{aligned}\Pi_\tau(y_{1,cl}^*) &= \Pi_\tau \circ \mathcal{F}_1^{u_1} \circ \mathcal{F}_2^{u_2}(y_{1,cl}^*)\\ &= \Pi_\tau \circ \mathcal{F}_1^{\Pi_\tau(u_1)} \circ \Pi_\tau(\mathcal{F}_2^{u_2}(y_{1,cl}^*))\\ &= \Pi_\tau \circ \mathcal{F}_1^{\Pi_\tau(u_1)} \circ \Pi_\tau \circ \mathcal{F}_2^{\Pi_\tau(u_2)} \circ \Pi_\tau(y_{1,cl}^*)\\ &= \Pi_\tau \circ \mathcal{F}_1^{\Pi_\tau(u_1)} \circ \mathcal{F}_2^{\Pi_\tau(u_2)} \circ \Pi_\tau(y_{1,cl}^*).\end{aligned}$$

Similarly, $\Pi_\tau(y_{1,\Pi_\tau}^*) = \Pi_\tau \circ \mathcal{F}_1^{\Pi_\tau(u_1)} \circ \mathcal{F}_2^{\Pi_\tau(u_2)} \circ \Pi_\tau(y_{1,\Pi_\tau}^*)$. Note that $\Pi_\tau \circ \mathcal{F}_1^{\Pi_\tau(u_1)} \circ \mathcal{F}_2^{\Pi_\tau(u_2)}$ is a strict contraction, by the Banach fixed point theorem we have $\Pi_\tau(y_{1,cl}^*) = \Pi_\tau(y_{1,\Pi_\tau}^*)$.

To establish closed-loop UIOC, let $x_{1,cl}^*$ be the reference state solution to subsystem $j = 1$ with respect to input $(y_{2,cl}^*, u_1)$ and analogously for $x_{2,cl}^*$. Then $x_{cl}^* = (x_{1,cl}^*, x_{2,cl}^*)$ is the closed-loop reference state solution. Let $\overline{x}(k) = (\overline{x}_1(k), \overline{x}_2(k)), \overline{y}(k) = (\overline{y}_1(k), \overline{y}_2(k))$ be any other closed-loop solution and its corresponding output to another input $\overline{u} = (\overline{u}_1, \overline{u}_2)$, starting at $\overline{x}(k_0)$. From (5.7) and (5.8), the same argument as in the proof of Theorem 5.2 shows that $\sup_{k\geq k_0} \|x_{cl}^*(k) - \overline{x}(k)\| < \infty$, $\sup_{k\geq k_0} \|y_{cl}^*(k) - \overline{y}(k)\| < \infty$.

For $j = 1, 2$ and $k \geq k_0$, define $\Delta x_j(k) = \overline{x}_j(k) - x_{j,cl}^*(k)$, $\Delta y_j(k) = \overline{y}_j(k) - y_{j,cl}^*(k)$ and $\Delta u_j(k) = \overline{u}_j(k) - u_j(k)$. Let $v_{1,cl}^* = y_{2,cl}^*$ and $v_{2,cl}^* = y_{1,cl}^*$, we have the time-varying systems with states $\Delta x_j(k)$, outputs $\Delta y_j(k)$, inputs $\Delta u_j(k)$, and interconnections $\Delta v_1(k) = \Delta y_2(k)$ and $\Delta v_2(k) = \Delta y_1(k)$,

$$\begin{aligned}\Delta x_j(k+1) &= f_j(\Delta x_j(k) + x_{j,cl}^*(k), \Delta v_j(k) + v_{j,cl}^*(k), \Delta u_j(k) + u_j(k))\\ &\quad - f_j(x_{j,cl}^*(k), v_{j,cl}^*(k), u_j(k))\\ &= \tilde{f}_j(k, \Delta x_j(k), \Delta v_j(k), \Delta u_j(k)),\end{aligned}$$

$$\begin{aligned}\Delta y_j(k) &= h_j(\Delta x_j(k) + x_{j,cl}^*(k), \Delta v_j(k) + v_{j,cl}^*(k), \Delta u_j(k) + u_j(k))\\ &\quad - h_j(x_{j,cl}^*(k), v_{j,cl}^*(k), u_j(k))\\ &= \tilde{h}_j(k, \Delta x_j(k), \Delta v_j(k), \Delta u_j(k)).\end{aligned}$$

From (5.7) and (5.8), $\Delta x_j(k)$ and $\Delta y_j(k)$ satisfy (5.11) and (5.12) in Theorem 5.2. Finally, closed-loop UIOC follows from applying Theorem 5.2 to the above interconnected systems.

□

Sometimes it is convenient to upper bound $\|y_j^*(k) - \overline{y}(k)\|$ in (5.7) by a sum instead of max of nonlinear gains. That is,

$$\begin{aligned}\|y_j^*(k) - \overline{y}_j(k)\| \leq & \beta_j(\|x_j^*(k_0) - \overline{x}_j(k_0)\|, k - k_0) \\ & + \gamma_j^y(\|(v_j - \overline{v}_j)_{[k_0,k]}\|) + \gamma_j^u(\|(u_j - \overline{u}_j)_{[k_0,k]}\|).\end{aligned} \tag{5.14}$$

Theorem 5.1 can be applied to this scenario by re-writing (5.14) in terms of max of nonlinear gains. If system (5.5) satisfies (5.14) instead of (5.7), to ensure UIOC of (5.5), the condition $\gamma_1^y \circ \gamma_2^y(s) < s$ for all $s > 0$ needs to be strengthened. We first present a lemma (first discussed in [140]) that allows us to re-write (5.14) in terms of max.

Lemma 5.1. *Given any $\lambda \in \mathcal{K}_\infty$, for any $a, b \geq 0$, it holds that $a + b \leq \max\{a + \lambda(a), b + \lambda^{-1}(b)\}$.*

Proof. Since $\lambda \in \mathcal{K}_\infty$, its inverse λ^{-1} exists and is in $\mathcal{K}_\infty$. Consider two cases. Suppose $b \leq \lambda(a)$, then $a+b \leq a+\lambda(a) \leq \max\{a+\lambda(a), b+\lambda^{-1}(b)\}$. Otherwise, $b > \lambda(a)$. Applying λ^{-1} on both sides gives $\lambda^{-1}(b) > a$ and $a+b < \lambda^{-1}(b)+b \leq \max\{a+\lambda(a), b+\lambda^{-1}(b)\}$. $\square$

Theorem 5.1 and Lemma 5.1 lead to the following Corollary.

Corollary 5.1. *Consider a well-posed system* (5.5) *with UOC subsystems. For any inputs $u_j \in l^\infty_{n_{u_j}}, v_j \in l^\infty_{n_{v_j}}$, let x_j^* and y_j^* be the corresponding reference state solutions and outputs. For any other inputs $\overline{u}_j, \overline{v}_j$ with $\overline{u}_j \in l^\infty_{n_{u_j}}$, let $\overline{x}_j$ and $\overline{y}_j$ be any corresponding solutions and outputs with initial conditions $\overline{x}_j(k_0)$. Suppose that there exists $\beta_j \in \mathcal{KL}$ and $\gamma_j^y, \gamma_j^u, \sigma_j, \sigma_j^u, \sigma_j^y \in \mathcal{K}$ such that, for all $k_0, k \in \mathbb{Z}$ with $k \geq k_0$ and any $\overline{x}_j(k_0) \in \mathbb{R}^{n_{x_j}}$,* (5.8) *and* (5.14) *hold. If there exists $\lambda_j \in \mathcal{K}_\infty$ such that for all $s > 0$,*

$$(id + \lambda_1) \circ \gamma_1^y \circ (id + \lambda_2) \circ \gamma_2^y(s) < s, \tag{5.15}$$

where id is the identity map. Then for any $k_0 \in \mathbb{Z}$ and $x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$, the closed-loop solution and output are bounded, i.e., $\sup_{k\geq k_0} \|x(k)\| < \infty$ and $\sup_{k\geq k_0} \|y(k)\| < \infty$. Furthermore, system (5.5) *is UIOC and induces a well-posed closed-loop I/O map.*

Proof. In Lemma 5.1, identify a with $\gamma_j^y(\|(v_j-\overline{v}_j)_{[k_0,k]}\|)$ and b with $\beta_j(\|x_j^*(k_0)-\overline{x}_j(k_0)\|, k-k_0) + \gamma_j^u(\|(u_j-\overline{u}_j)_{[k_0,k]}\|)$ for each $j=1,2$. Applying Lemma 5.1 to (5.14) gives

$$\begin{aligned}\|y_j^*(k)-\overline{y}_j(k)\| \le \max\{&(id+\lambda_j)\circ\gamma_j^y(\|(v_j-\overline{v}_j)_{[k_0,k]}\|),\\ &\beta_j(\|x_j^*(k_0)-\overline{x}_j(k_0)\|, k-k_0)+\gamma_j^u(\|(u_j-\overline{u}_j)_{[k_0,k]}\|)\\ &+\lambda_j^{-1}(\beta_j(\|x_j^*(k_0)-\overline{x}_j(k_0)\|, k-k_0)+\gamma_j^u(\|(u_j-\overline{u}_j)_{[k_0,k]}\|))\}\\ \le \max\{&\hat{\beta}_j(\|x_j^*(k_0)-\overline{x}_j(k_0)\|, k-k_0), (id+\lambda_j)\circ\gamma_j^y(\|(v_j-\overline{v}_j)_{[k_0,k]}\|),\\ &\hat{\gamma}_j^u(\|(u_j-\overline{u}_j)_{[k_0,k]}\|)\},\end{aligned}$$

where $\hat{\beta}_j(s,k) = 3\max\{\beta_j(s,k), \lambda_j^{-1}(2\beta_j(s,k))\}$ and $\hat{\gamma}_j^u(s) = 3\max\{\gamma_j^u(s), \lambda_j^{-1}(2\gamma_j^u(s))\}$, with $\hat{\beta}_j \in \mathcal{KL}$ and $\hat{\gamma}_j \in \mathcal{K}$. The result now follows from Theorem 5.1. $\square$

We can further apply Theorem 5.1 and Lemma 5.1 to system (5.5) with $y_j(k) = x_j(k)$. In this case, Theorem 5.1 ensures the UISC property of system (5.5), leading to the following Corollary.

Corollary 5.2. *Consider a well-posed system* (5.5) *with* $y_j(k) = x_j(k)$ *and UC subsystems. For any inputs* $u_j \in l^\infty_{n_{u_j}}, v_j \in l^\infty_{n_{v_j}}$*, let* x_j^* *be the corresponding reference state solutions. For any other inputs* $\overline{u}_j, \overline{v}_j$ *with* $\overline{u}_j \in l^\infty_{n_{u_j}}$*, let* $\overline{x}_j$ *be any corresponding solutions with initial conditions* $\overline{x}_j(k_0)$*. Suppose that there exists* $\beta_j \in \mathcal{KL}$ *and* $\gamma_j^y, \gamma_j^u \in \mathcal{K}$ *such that, for all* $k_0, k \in \mathbb{Z}$*,* $k \ge k_0$ *and any* $\overline{x}_j(k_0) \in \mathbb{R}^{n_{x_j}}$*,*

$$\begin{aligned}\|x_j^*(k)-\overline{x}_j(k)\| \le &\beta_j(\|x_j^*(k_0)-\overline{x}_j(k_0)\|, k-k_0)\\ &+\gamma_j^y(\|(v_j-\overline{v}_j)_{[k_0,k-1]}\|)+\gamma_j^u(\|(u_j-\overline{u}_j)_{[k_0,k-1]}\|).\end{aligned} \tag{5.16}$$

If (5.15) *holds, then for any* $k_0 \in \mathbb{Z}$ *and* $x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$*, the closed-loop solution is bounded, i.e.,* $\sup_{k\ge k_0}\|x(k)\| < \infty$*. Furthermore, system* (5.5) *is UISC.*

5.3 Interconnected RCs for temporal tasks

Nonlinear closed-loop model structures, such as the Wiener-Hammerstein feedback models, have been proposed to better capture nonlinear feedback phenomena of the unknown system [76, 77]. Here we introduce interconnected RCs as candidate models, expanding the toolbox of available nonlinear models for approximating I/O maps. We apply the UIOC small-gain theorem to ensure that the closed-loop RC is UOC (or UC for state-feedback interconnections), so that the estimated outputs for large times are determined by the inputs but not by the closed-loop RC's initial condition. As a concrete example, we will apply the theorem to interconnected ESNs and QRCs to emulate the output of an observed-based feedback-controlled Lur'e system modified from [144, Example 1], described by

$$\begin{cases} z(k+1) = Az(k) + B_u u(k) + B_w w(k) + 0.1G\sin(Hz(k)), \\ y(k) = Cz(k), \end{cases} \tag{5.17}$$

where $w(k) \in \mathbb{R}$ is the input and

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1.1 \end{bmatrix}, B_u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, B_w = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}, G = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}, H = \begin{bmatrix} 1 & 1 \end{bmatrix}, C = \begin{bmatrix} 0.1 & 0.5 \end{bmatrix}.$$

We implement a Luenberger observer with gain $L = P^{-1}Z \in \mathbb{R}^2$ for the Lur'e system (5.17), governed by

$$\hat{z}(k+1) = A\hat{z}(k) + B_u u(k) + B_w w(k) + \rho G \sin(H\hat{z}(k)) - P^{-1}ZC(\hat{z}(k) - z(k)),$$

where $P \succ 0$. We then consider a linear state-feedback $u(k) = -K\hat{z}(k)$ with gain $K^\top \in \mathbb{R}^2$. In Appendix B.2, we show that with $L^\top = \begin{bmatrix} 2.3258 & 2.1104 \end{bmatrix}$ and $K = \begin{bmatrix} 0.4956 & 1.006 \end{bmatrix}$, the feedback-controlled Lur'e system (5.17) is UC.

Suppose that we have inputs $w_l(k), w'_{l'}(k) \in \mathbb{R}$ and their corresponding outputs $y_l(k), y'_{l'}(k) \in \mathbb{R}$ of the unknown system for $1 \leq k \leq L$, $1 \leq l \leq M$ and $1 \leq l' \leq M'$. I/O data $w_l(k), y_l(k)$ are for parameter estimation (using $l = 1, \ldots, M_1$) and model selection (using $l = M_1 + 1, \ldots, M$) based on Akaike's final prediction error [145]. I/O data $w'_{l'}(k), y'_{l'}(k)$ are for model evaluation. For each l and l', we first washout the effect of RC's initial condition for $k = 1, \ldots, L_w$. Let $\hat{y}_l(k)$ and $\hat{y}_{l'}(k)$ be the RC's outputs under inputs $w_l(k)$ and $w_{l'}(k)$, respectively. To optimize the RC output parameters, we minimize $\sum_{l=1}^{M_1} \sum_{k=L_w+1}^{L} |y_l(k) - \hat{y}_l(k)|^2$. We estimate the model order based on FPE_l, computed as

$$\text{FPE}_l = \frac{1}{L - L_w} \sum_{k=L_w+1}^{L} |y_l(k) - \hat{y}_l(k)|^2 \frac{L - L_w + p}{L - L_w - p},$$

where p is the number of RC output parameters. For each p, we randomly generate N RCs and select a model out of Np models with the minimum $\text{FPE} := \sum_{l=M_1+1}^{M} \text{FPE}_l$. Here, FPE is the sum of all FPE_l over the model selection data. For each of the model validation data $l' = 1, \ldots, M'$, the selected model is assessed using the mean-squared error $\text{MSE}_{l'} = \frac{1}{L-L_w} \sum_{k=L_w+1}^{L} |y'_{l'}(k) - \hat{y}'_{l'}(k)|^2$.

We set $L_w = 500$, $L = 1500$, $N = 10$, $M = 10, M_1 = 8$ and $M' = 2$, with inputs $w_l(k)$ and $w'_1(k)$ sampled uniformly over $[-2, 2]$, independently for each k (persistently exciting [145] with an order of 50 estimated by the 'pexcit' Matlab command), whereas $w'_2(k) = \sin(2\pi k/25) + \sin(\pi k/5)$ as in [146].

5.3.1 Echo-state networks (ESNs)

Consider state-feedback interconnected ESNs of the form,

$$\begin{cases} x_1(k+1) = \tanh(A_1 x_1(k) + A_1^{fb} v_1(k) + B_1 w(k)), \\ x_2(k+1) = \tanh(A_2 x_2(k) + A_2^{fb} v_2(k) + B_2 w(k)), \end{cases} \tag{5.18}$$

with interconnections $v_1(k) = x_2(k) \in \mathbb{R}^{n_{x_2}}$ and $v_2(k) = x_1(k) \in \mathbb{R}^{n_{x_1}}$. Here, $w \in l_1^{\infty}$ is the input and $\tanh(\cdot)$ is applied to a vector element-wise. We choose an output $\hat{y}(k) = W_1^{\top} x_1(k) + W_2^{\top} x_2(k) + \zeta$, where $W_j \in \mathbb{R}^{n_{x_j}}$ and $\zeta \in \mathbb{R}$ is a bias term. The UC property of each ESN is guaranteed by choosing $\sigma_{\max}(A_j) < 1$ and noticing its compact state-space [22, Theorem 13]. We apply Corollary 5.2 to establish the UISC property for the interconnected ESNs (5.18). For any $k, k_0 \in \mathbb{Z}$, let $x_j(k), \overline{x}_j(k)$ be any solutions to (5.18) under inputs v_j, w and $\overline{v}_j, \overline{w}$ respectively. Let $\delta\overline{x}_j(k) = x_j(k) - \overline{x}_j(k)$, $\delta\overline{v}_j = v_j - \overline{v}_j$ and $\delta\overline{w} = w - \overline{w}$, then (5.16) in Corollary 5.2 is satisfied since

$$\|\delta\overline{x}_j(k)\| \le \sigma_{\max}(A_j)^{k-k_0} \|\delta\overline{x}_j(k_0)\| + \frac{\sigma_{\max}(A_j^{fb})}{1 - \sigma_{\max}(A_j)} \|\delta\overline{v}_{j[k_0,k-1]}\| + \frac{\|B_j\|}{1 - \sigma_{\max}(A_j)} |\delta\overline{w}_{[k_0,k-1]}|.$$

In Corollary 5.2, choose $\lambda_1(s) = \lambda_2(s) = \lambda s$ for some $\lambda > 0$. The interconnected ESNs (5.18) is UISC if

$$\frac{\sigma_{\max}(A_1^{fb})}{1 - \sigma_{\max}(A_1)} \frac{\sigma_{\max}(A_2^{fb})}{1 - \sigma_{\max}(A_2)} < \frac{1}{(1+\lambda)^2}. \tag{5.19}$$

Example 5.1. *We consider interconnected ESN* (5.18) *with* $n_{x_1} = n_{x_2} \in \{2, \ldots, 5\}$ *(i.e., with number of output parameters* $p = 2n_{x_1} + 1$*) to model the feedback-controlled Lur'e system* (5.17)*. For each ESN, elements of* A_j, A_j^{fb}, B_j *are sampled independently and uniformly over* $[-1, 1]$*. We fix* $\sigma_{\max}(A_j) = 0.5$, $\sigma_{\max}(A_2^{fb})$*, and scale* $\sigma_{\max}(A_1^{fb})$ *so that* (5.19) *holds. The minimum FPE achieved is* $\mathrm{FPE} = 0.0023$*, with* $n_{x_1} = n_{x_2} = 4$ *and* $p = 9$*. For this selected ESN,* $\sigma_{\max}(A_1^{fb}) = 0.15, \sigma_{\max}(A_2^{fb}) = 1.65$ *and* (5.19) *holds for* $\lambda = 0.003$*. This results in* $MSE_1 = 0.0012$ *and* $MSE_2 = 0.0028$ *corresponding to the evaluation data* $l' = 1, 2$*, respectively. See Fig. 5.2 for the target* $y'_{l'}(k)$ *and the closed-loop ESN outputs* $\hat{y}'_{l'}(k)$*.*

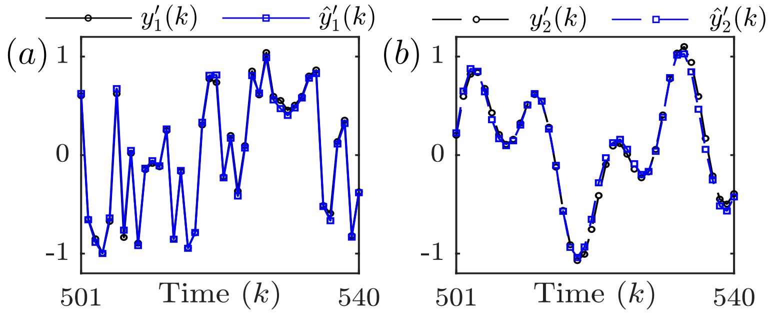


Figure 5.2: Target outputs $y'_{l'}(k)$ and the closed-loop ESN outputs $\hat{y}'_{l'}(k)$ for $k = 501, \ldots, 540$ with (a) $l' = 1$ under a uniform random input $w'_1(k)$ and (b) $l' = 2$ under a sum of sinusoidals $w'_2(k) = \sin(2\pi k/25) + \sin(\pi k/5)$.

5.3.2 Quantum reservoir computers (QRCs)

Consider output-feedback interconnected QRC of the form (also see Fig. 5.3),

$$\begin{cases} \rho_j(k+1) = \mathcal{T}_j(w(k), v_j(k))\rho_j(k) + \epsilon_\phi^{(j)}\phi_j, \\ \hat{y}_j(k) = \sum_{i=1}^{n_j} \mathrm{Tr}(Z^{(i)}\rho_j(k)), \end{cases} \tag{5.20}$$

with interconnections for $j = 1, 2$. Here $\mathcal{T}_j(w(k), v_j(k)) = \epsilon_w^{(j)}\mathcal{T}_w^{(j)}(w(k)) + \epsilon_v^{(j)}\mathcal{T}_v^{(j)}(v_j(k))$, $\epsilon_w^{(j)} + \epsilon_v^{(j)} + \epsilon_\phi^{(j)} = 1$ and $\epsilon_w^{(j)}, \epsilon_v^{(j)}, \epsilon_\phi^{(j)} > 0$. Subsystem j has n_j qubits so that $\rho_j(k)$ and ϕ_j are two $2^{n_j} \times 2^{n_j}$ density operators, with ϕ_j being fixed. For $S \in \{w, v\}$, the input-dependent CPTP maps are

$$\mathcal{T}_S^{(j)}(x)\rho_j(k) = \left[g(x)T_{S,1}^{(j)} + (1 - g(x))T_{S,2}^{(j)}\right]\rho_j(k),$$

where $g(x) = 1/(1 + \exp(-x))$ is the logistic function with a globally Lipschitz constant $L_g = 1/4$ and $T_{S,1}^{(j)}, T_{S,2}^{(j)}$ are input-independent CPTP maps. We choose the output of the closed-loop QRC as

$$\hat{y}(k) = \sum_{i=1}^{n_1} W_i^{(1)} \mathrm{Tr}(Z^{(i)} \rho_1(k)) + \sum_{i=1}^{n_2} W_i^{(2)} \mathrm{Tr}(Z^{(i)} \rho_2(k)) + \zeta,$$

where $W_i^{(j)}, \zeta \in \mathbb{R}$ ($j = 1, 2$ and $i = 1, \ldots, n_j$) are the output parameters to be optimized via ordinary least squares.

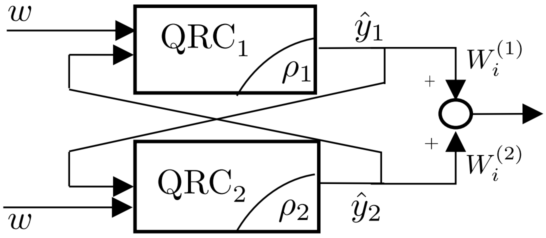


Figure 5.3: Schematic of an interconnected QRC described by (5.20).

Note that interconnected quantum systems do not generally take the form (5.20); see [147], [148] and [149, Chapter 5]. System (5.20) can describe ensembles of identical quantum systems such as NMR ensembles [54], and quantum systems that can emulate such ensembles; e.g., [115,150]. Such quantum systems have dynamics constrained by quantum mechanics, but can otherwise be viewed as deterministic systems. Since the quantum subsystems here do not interact quantum mechanically, the composite state $\rho(k)$ for (5.20) can be described by the direct sum of the subsystem density operators, $\rho(k) = \rho_1(k) \oplus \rho_2(k)$, as for interconnected classical systems. Consequently, the closed-loop system (5.20) is of the form (5.5) and Theorem 5.1 is applicable. We remark that Theorem 5.1 and its subsequent corollaries also hold for the Schatten 1–norm.

We now employ Corollary 5.1 to establish the UIOC of system (5.20). Recall from Lemma 4.1 and [43, Theorem 9.2] that for any density operators $\rho, \overline{\rho}$ and any CPTP map $\mathcal{T}$, $\|\mathcal{T}(\rho - \overline{\rho})\|_1 \leq \|\rho - \overline{\rho}\|_1$. For any $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$, let $\rho_j(k), \overline{\rho}_j(k)$ be any solutions to inputs v_j, w and $\overline{v}_j, \overline{w}$, respectively. Let $\delta\overline{\rho}_j(k) = \rho_j(k) - \overline{\rho}_j(k)$, we have

$$
\begin{aligned}
&\|\delta\overline{\rho}_j(k)\|_1 \\
&\leq (\epsilon_w^{(j)} + \epsilon_v^{(j)})\|\delta\overline{\rho}_j(k-1)\|_1 \\
&+ \epsilon_v^{(j)}\|\mathcal{T}_v^{(j)}(v_j(k-1)) - \mathcal{T}_v^{(j)}(\overline{v}_j(k-1))\|_{1-1} \\
&+ \epsilon_w^{(j)}\|\mathcal{T}_w^{(j)}(w(k-1)) - \mathcal{T}_w^{(j)}(\overline{w}(k-1))\|_{1-1} \\
&\qquad\qquad\qquad\qquad \vdots \\
&\leq (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\delta\overline{\rho}_j(k_0)\|_1 \\
&+ \epsilon_v^{(j)}\Big\{(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0-1}\|\mathcal{T}_v^{(j)}(v_j(k_0)) - \mathcal{T}_v^{(j)}(\overline{v}_j(k_0))\|_{1-1} \\
&\qquad\qquad + \ldots + \|\mathcal{T}_v^{(j)}(v_j(k-1)) - \mathcal{T}_v^{(j)}(\overline{v}_j(k-1))\|_{1-1}\Big\} \\
&+ \epsilon_w^{(j)}\Big\{(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0-1}\|\mathcal{T}_w^{(j)}(w(k_0)) - \mathcal{T}_w^{(j)}(\overline{w}(k_0))\|_{1-1} \\
&\qquad\qquad + \ldots + \|\mathcal{T}_w^{(j)}(w(k-1)) - \mathcal{T}_w^{(j)}(\overline{w}_j(k-1))\|_{1-1}\Big\}.
\end{aligned}
$$

Define $\delta\overline{v}_j = v_j - \overline{v}_j$ and $\delta\overline{w} = w - \overline{w}$. Using the fact that $\epsilon_\phi^{(j)} = 1 - (\epsilon_w^{(j)} + \epsilon_v^{(j)})$ and $\|\mathcal{T}_S^{(j)}(s) - \mathcal{T}_S^{(j)}(\overline{s})\|_{1-1} \leq L_g\|T_{S,1}^{(j)} - T_{S,2}^{(j)}\|_{1-1}|s - \overline{s}|$ for any $s \in \mathbb{R}$ and $S \in \{v, w\}$, we have

$$
\begin{aligned}
\|\delta\overline{\rho}_j(k)\|_1 \leq\ & (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\delta\overline{\rho}_j(k_0)\|_1 \\
&+ (\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{v,1}^{(j)} - T_{v,2}^{(j)}\|_{1-1}|\delta\overline{v}_{j_{[k_0,k-1]}}| \\
&+ (\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{w,1}^{(j)} - T_{w,2}^{(j)}\|_{1-1}|\delta\overline{w}_{[k_0,k-1]}| \\
\leq\ & 3\max\{(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\delta\overline{\rho}_j(k_0)\|_1, \\
&(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{v,1}^{(j)} - T_{v,2}^{(j)}\|_{1-1}|\delta\overline{v}_{j_{[k_0,k-1]}}|, \\
&(\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{w,1}^{(j)} - T_{w,2}^{(j)}\|_{1-1}|\delta\overline{w}_{[k_0,k-1]}|\}.
\end{aligned}
\tag{5.21}
$$

Furthermore, let $y_j(k), \overline{y}_j(k)$ be the outputs associated to $\rho_j(k), \overline{\rho}_j(k)$. Applying Lemma B.3 in Appendix B.3 gives

$$
|y_j(k) - \overline{y}_j(k)| = \left|\sum_{i=1}^{n_j} \mathrm{Tr}(Z^{(i)}\delta\overline{\rho}_j(k))\right| \leq n_j\|\delta\overline{\rho}_j(k)\|_1. \tag{5.22}
$$

To show that each QRC subsystem is UOC (with respect to the Schatten–1 norm), consider $\delta\overline{v}_j = \delta\overline{w} = 0$. From (5.21), we have $\|\delta\overline{\rho}_j(k)\| \leq (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\delta\overline{\rho}_j(k_0)\|_1$ with $\epsilon_w^{(j)} + \epsilon_v^{(j)} < 1$. Applying Theorem 2.2 shows that each QRC subsystem is UC with respect to the Schatten–1 norm. Furthermore, from (5.21) and (5.22), we have $|y_j(k) - \overline{y}_j(k)| \leq n_j(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\delta\overline{\rho}_j(k_0)\|_1$, and hence UOC of each QRC subsystem.

We can upper bound $\|T_{S,1}^{(j)} - T_{S,2}^{(j)}\|_{1-1} \leq 2$ for $S \in \{w, v\}$ [105, Theorem 2.1] in (5.21). Equations (5.21), (5.22) show that (5.8), (5.14) in Corollary 5.1 hold. Choose $\lambda_1(s) = \lambda_2(s) = \lambda s$ for some $\lambda > 0$ in Corollary 5.1, the closed-loop QRC (5.20) is UIOC if

$$(4\epsilon_v^{(1)}\epsilon_v^{(2)}L_g L_g n_1 n_2)/(\epsilon_\phi^{(1)}\epsilon_\phi^{(2)}) < 1/(1+\lambda)^2. \tag{5.23}$$

Example 5.2. *We consider an interconnected QRC* (5.20) *with* $n_1 = n_2 \in \{2, \ldots, 5\}$ *(i.e.,* $p = 2n_1 + 1$*) to model the feedback-controlled Lur'e system* (5.17)*. For each QRC, we fix* $\epsilon_w^{(1)} = 0.25, \epsilon_v^{(1)} = 0.1, \epsilon_\phi^{(1)} = 0.65$ *and* $\epsilon_w^{(2)} = 0.1, \epsilon_v^{(2)} = 0.45, \epsilon_\phi^{(2)} = 0.45$*, such that* (5.23) *holds for all values of* n_1 *considered here. For* $j = 1, 2$*,* ϕ_j *is chosen with its* $1, 1$*-th element* $(\phi_j)_{1,1} = 1$ *and zero otherwise. Each input-independent CPTP map is governed by a unitary matrix* $U_{S,m}^{(j)}$*, defined by* $T_{S,m}^{(j)}(\rho) = U_{S,m}^{(j)}\rho(U_{S,m}^{(j)})^\dagger$ *for* $j, m = 1, 2$ *and* $S \in \{w, v\}$*. More explicitly, we choose* $U_{w,1}^{(1)} = U_{v,2}^{(1)} = U_{w,1}^{(2)} = U_{v,2}^{(2)} = \bigotimes_{i=1}^{n_1} Z^{(i)}$*. Other unitaries are* $U_{S,m}^{(j)} = \bigotimes_{i=1}^{n_1} e^{-\iota\theta_i^{S,m,j}X^{(i)}}$*. Parameters* $\theta_i^{S,m,j}$ *are uniformly distributed on* $[-\pi, \pi]$*, independently for each* S, m, j *and* i*. The unitaries employed here are simple, more complex unitaries that entangle qubits within a QRC subsystem can also be used; e.g., see Chapter 4.*

The minimum FPE is achieved at $\mathrm{FPE} = 0.0032$ *with* $n_1 = n_2 = 5$ *and* $p = 11$*, and* (5.23) *holds for* $\lambda = 0.019$*. This selected QRC achieves* $\mathrm{MSE}_1 = \mathrm{MSE}_2 = 0.0015$*. See Fig. 5.4 for the QRC outputs* $\hat{y}'_{l'}(k)$ *against the target outputs* $y'_{l'}(k)$ *for the evaluation data* $l' = 1, 2$*.*

In practice, QRCs implemented on NISQ hardware experience decoherence; see Sec. 2.2.3 for a brief introduction. To investigate the effect of decoherence on the modeling performance measured by mean-squared error, we simulation the selected interconnected QRC under dephasing and GAD channels, and apply a multi-qubit error modeled by a depolarizing channel [116]. Output parameters of these QRCs are re-optimized under decoherence. We simulate GAD channels with finite temperature parameters $\lambda = \{0.2, 0.4, 0.6, 0.8, 1\}$*. The decoherence strengths* $\psi = \{10^{-4}, 10^{-3}, 10^{-2}\}$ *for both channels are set to be experimentally feasible in superconducting quantum devices [44]; see Appendix B.4 for detailed*

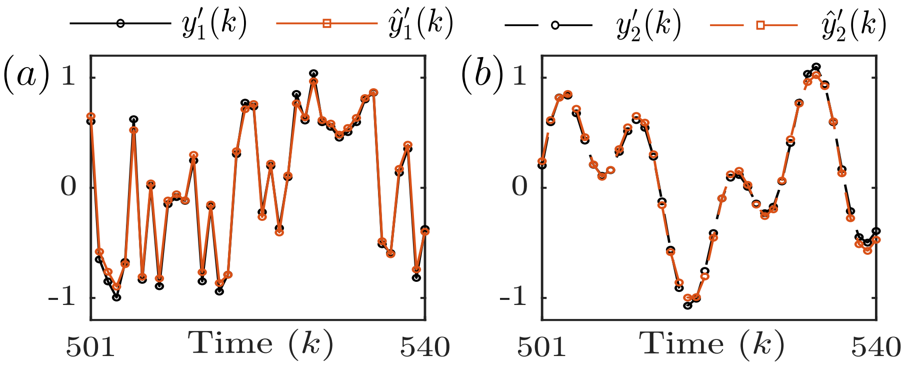


Figure 5.4: Target outputs $y'_{l'}(k)$ and the closed-loop QRC outputs $\hat{y}'_{l'}(k)$ for $k = 501, \ldots, 540$ with (a) $l' = 1$ under a uniform random input $w'_1(k)$ and (b) $l' = 2$ under a sum of sinusoidals $w'_2(k) = \sin(2\pi k/25) + \sin(\pi k/5)$.

numerical settings. The simulated decoherence is a Markovian and stationary approximation of hardware noise. The proof-of-principle and numerical experiments in Chapter 4 suggest that this approximation is reasonable for some superconducting quantum devices.

The modeling performance under the GAD noise is similar to the noiseless case. Under the GAD channels with decoherence strengths $\psi = \{10^{-4}, 10^{-3}\}$*, the selected QRC achieves the same* $\mathrm{MSE}_1 = \mathrm{MSE}_2 = 0.0015$ *as before. Under decoherence strength* $\psi = 10^{-2}$*, the selected QRC achieves* $\mathrm{MSE}_1 = \mathrm{MSE}_2 = 0.0016$ *when* $\lambda = \{0.2, 0.4, 0.6, 0.8\}$ *and* $\mathrm{MSE}_1 = 0.0015, \mathrm{MSE}_2 = 0.0016$ *when* $\lambda = 1$*. The modeling performance under the dephasing noise is the same as the noiseless case. These observations suggest that decoherence does not significantly impact the interconnected QRC's emulation ability.*

The expectation values $\mathrm{Tr}(Z^{(i)}\rho_j)$ for $j = 1, 2$ are estimated by averaging M_m measurements, whose variances decrease at a rate proportional to $1/M_m$; see Appendix A.2 for the proof. So far, we have assumed that the finite sampling noise can be made negligible by taking M_m to be very large. In practice, we must take this noise into account.

The framework presented in this chapter can accommodate the finite sampling noise. Given a realization of the noise, the output-feedback interconnected QRC is governed by

$$\begin{cases} \rho_j(k+1) = \mathcal{T}_j(w(k), v_j(k))\rho_j(k) + \epsilon_\phi^{(j)} \phi_j, \\ \hat{y}_j(k) = \sum_{i=1}^{n_j} \mathrm{Tr}(Z^{(i)} \rho_j(k)) + e_j(k), \end{cases} \tag{5.24}$$

with interconnections $v_1(k) = \hat{y}_2(k)$ and $v_2(k) = \hat{y}_1(k)$. Here, $e_j(k) = \sum_{i=1}^{n_j} e_j^{(i)}(k)$, where $e_j^{(i)}(k)$ is the finite sampling noise when estimating the expectation value $\mathrm{Tr}(Z^{(i)}\rho_j(k))$. We remark that e_j is a bounded sequence with $e_j(k) \in [-2n_j, 2n_j]$ for all $k \in \mathbb{Z}$. This is because for any $i = 1, \ldots, n_j$, $\mathrm{Tr}(Z^{(i)}\rho_j(k)) \in [-1, 1]$ and its estimate also takes value in $[-1, 1]$. Let $z_1(k) = e_2(k)$, $z_2(k) = e_1(k)$ and $\tilde{w}_j(k) = (w(k), z_j(k))$ be a concatenation of $w(k)$ and $z_j(k)$ into a column vector for $j = 1, 2$. We can re-express (5.24) in a form such that $\tilde{w}_j(k)$ $(j = 1, 2)$ are viewed as external inputs to the QRC subsystems. That is, we can re-write (5.24) as

$$\begin{cases} \rho_j(k+1) = \tilde{\mathcal{T}}_j(\tilde{w}_j(k), \tilde{v}_j(k))\rho_j(k) + \epsilon_\phi^{(j)}\phi_j, \\ \quad \tilde{y}_j(k) = \sum_{i=1}^{n_j} \mathrm{Tr}(Z^{(i)}\rho_j(k)), \end{cases} \tag{5.25}$$

with interconnections $\tilde{v}_1(k) = \tilde{y}_2(k)$ and $\tilde{v}_2(k) = \tilde{y}_1(k)$. Each QRC subsystem is governed by $\tilde{\mathcal{T}}_j(\tilde{w}_j(k), \tilde{v}_j(k)) = \mathcal{T}_j(w(k), \tilde{v}_j(k) + z_j(k))$.

We will apply Theorem 5.1 to ensure that the interconnected QRC (5.25) under finite sampling noise is UIOC. The analysis below closely follows the analysis when there is no finite sampling noise. Let $\rho_j(k), \rho_j'(k)$ be any solutions to inputs $\tilde{v}_j, \tilde{w}_j = (w, z_j)$ and $\tilde{v}_j', \tilde{w}_j' = (w', z_j')$, respectively. We have

$$\begin{aligned} \|\rho_j(k) - \rho_j'(k)\|_1 \leq{} & (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\rho_j(k_0) - \rho_j'(k_0)\|_1 \\ & + (\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{v,1}^{(j)} - T_{v,2}^{(j)}\|_{1-1}|(\tilde{v}_j - \tilde{v}_j' + z_j - z_j')_{[k_0,k-1]}| \\ & + (\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g\|T_{w,1}^{(j)} - T_{w,2}^{(j)}\|_{1-1}|(w - w')_{[k_0,k-1]}| \\ \leq{} & (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\rho_j(k_0) - \rho_j'(k_0)\|_1 + 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(\tilde{v}_j - \tilde{v}_j')_{[k_0,k-1]}| \\ & + 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(z_j - z_j')_{[k_0,k-1]}| + 2(\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g|(w - w')_{[k_0,k-1]}| \\ \leq{} & (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0}\|\rho_j(k_0) - \rho_j'(k_0)\|_1 + 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(\tilde{v}_j - \tilde{v}_j')_{[k_0,k-1]}| \\ & + 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g\|(\tilde{w}_j - \tilde{w}_j')_{[k_0,k-1]}\| + 2(\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g\|(\tilde{w}_j - \tilde{w}_j')_{[k_0,k-1]}\|, \end{aligned} \tag{5.26}$$

where the second inequality follows from $\|T_{S,1}^{(j)} - T_{S,2}^{(j)}\|_{1-1} \leq 2$ for $S \in \{w, v\}$ [105, Theorem 2.1] and $|(\tilde{v}_j - \tilde{v}_j' + z_j - z_j')_{[k_0,k-1]}| \leq |(\tilde{v}_j - \tilde{v}')_{[k_0,k-1]}| + |(z_j - z_j')_{[k_0,k-1]}|$. Moreover, the last inequality follows from $|(z_j - z_j')_{[k_0,k-1]}| \leq \|(\tilde{w}_j - \tilde{w}_j')_{[k_0,k-1]}\|$ and $|(w - w')_{[k_0,k-1]}| \leq \|(\tilde{w}_j - \tilde{w}_j')_{[k_0,k-1]}\|$.

Let $\tilde{y}_j(k), \tilde{y}'_j(k)$ be the outputs associated to $\rho_j(k), \rho'_j(k)$. Applying Lemma B.3 in Appendix B.3 again gives

$$|\tilde{y}_j(k) - \tilde{y}'_j(k)| = \left| \sum_{i=1}^{n_j} \mathrm{Tr}\left(Z^{(i)}(\rho_j(k) - \rho'_j(k)) \right) \right| \le n_j \|\rho_j(k) - \rho'_j(k)\|_1. \tag{5.27}$$

Consider $\tilde{v}_j = \tilde{v}'_j$ and $\tilde{w}_j = \tilde{w}'_j$. From (5.26), we have

$$\|\rho_j(k) - \rho'_j(k)\|_1 \le (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0} \|\rho_j(k_0) - \rho'_j(k_0)\|_1.$$

Applying Theorem 2.2 shows that each QRC subsystem is UC with respect to the Schatten–1 norm. Furthermore, each QRC subsystem is UOC since

$$|\tilde{y}_j(k) - \tilde{y}'_j(k)| \le n_j (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0} \|\rho_j(k_0) - \rho'_j(k_0)\|_1.$$

To show that the interconnected QRC is UIOC, define

$$\begin{aligned} &\gamma_j^u(\|\tilde{w}_j - \tilde{w}'_j\|_{[k_0,k-1]}) \\ &= 2\max\{2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g\|(\tilde{w}_j - \tilde{w}'_j)_{[k_0,k-1]}\|, 2(\epsilon_w^{(j)}/\epsilon_\phi^{(j)})L_g\|(\tilde{w}_j - \tilde{w}'_j)_{[k_0,k-1]}\|\}. \end{aligned}$$

Then $\gamma_j^u \in \mathcal{K}$. From (5.26) and (5.27), we have

$$\begin{aligned} \|\rho_j(k) - \rho'_j(k)\|_1 &\le (\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0} \|\rho_j(k_0) - \rho'_j(k_0)\|_1 \\ &\quad + 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(\tilde{v}_j - \tilde{v}'_j)_{[k_0,k-1]}| + \gamma_j^u(\|\tilde{w}_j - \tilde{w}'_j\|_{[k_0,k-1]}) \\ &\le 3\max\{(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0} \|\rho_j(k_0) - \rho'_j(k_0)\|_1, \\ &\quad 2(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(\tilde{v}_j - \tilde{v}'_j)_{[k_0,k-1]}|, \gamma_j^u(\|\tilde{w}_j - \tilde{w}'_j\|_{[k_0,k-1]})\}, \\ |\tilde{y}_j(k) - \tilde{y}'_j(j)| &\le n_j(\epsilon_w^{(j)} + \epsilon_v^{(j)})^{k-k_0} \|\rho_j(k_0) - \rho'_j(k_0)\|_1 \\ &\quad + 2n_j(\epsilon_v^{(j)}/\epsilon_\phi^{(j)})L_g|(\tilde{v}_j - \tilde{v}'_j)_{[k_0,k-1]}| + n_j\gamma_j^u(\|\tilde{w}_j - \tilde{w}'_j\|_{[k_0,k-1]}). \end{aligned}$$

Therefore, conditions (5.8) and (5.14) in Corollary 5.1 hold. We can again choose $\lambda_1(s) = \lambda_2(s) = \lambda s$ for some $\lambda > 0$ in Corollary 5.1, the closed-loop QRC (5.25) is UIOC if

$$(4\epsilon_v^{(1)}\epsilon_v^{(2)}L_gL_gn_1n_2)/(\epsilon_\phi^{(1)}\epsilon_\phi^{(2)}) < 1/(1+\lambda)^2.$$

Note that this condition is the same as the condition (5.23) derived under the assumption that we can obtain the expectation values $\mathrm{Tr}(Z^{(i)}\rho_j)$ exactly without finite sampling noise.

Example 5.3. *To investigate the impact of finite sampling noise on the emulation performance, we simulate the selected QRC with* $M_m = \{10^3, 10^4, 10^5\}$*. For* $M_m = 10^3$*, the selected QRC achieves* $\mathrm{MSE}_1 = 0.0157$ *and* $\mathrm{MSE}_2 = 0.0222$*. These errors are significantly larger than before, indicating that the finite sampling noise can negatively impact the emulation performance. As* M_m *increases, the emulation performance improves. For* $M_m = 10^4$*, we have* $\mathrm{MSE}_1 = 0.0057$ *and* $\mathrm{MSE}_2 = 0.0063$*. For* $M_m = 10^5$*, we have* $\mathrm{MSE}_1 = \mathrm{MSE}_2 = 0.0024$*. These observations indicate that further investigations is needed to understand and quantify the impact of finite sampling noise on emulation performance.*

5.4 Conclusion

We present a small-gain theorem for output-feedback interconnected systems to be UOC and UIOC systems, as a discrete-time counterpart of the continuous-time results in [151]. Our proof is based on a small-gain theorem for time-varying discrete-time systems in the UIOS framework, also derived herein. The latter result bridges the gap between time-invariant and time-varying discrete-time small-gain theorems in the literature [135, 136].

Our small-gain theorems are applicable to important control problems, such as output regulation and tracking [152]. We demonstrate an application of our small-gain theorems to design parameters of interconnected reservoir computers for black-box system identification. We introduce interconnected ESNs and QRCs as candidate models equipped with closed-loop structures and demonstrate numerically their efficacy in modeling a feedback-controlled system. We further investigate the effect of decoherence and finite sampling error on the QRC's estimation quality. In the next chapter, we will consider RCs with output feedback dynamics which is a special case of the general interconnection considered in this chapter. We will explore the use of uniformly convergent output-feedback dynamics for stochastic modeling.

Chapter 6

Nonlinear autoregression with reservoir computers

The on-going quest for modeling increasingly complex systems has motivated a fruitful development in nonlinear time series modeling and system identification [153,154]. These are challenging problems that involve inferring unknown models from observed data. Substantial interest has been focused on model structures that can capture nonlinear phenomena of the unknown system. For instance, threshold model has been widely applied for ecology and hydrology times series [155] and the autoregressive conditional heteroscedastic model is useful for describing volatility clustering in financial data [156]; see [153] and references therein. For nonlinear system identification, well-known model structures include the Volterra series [25], neural networks [146], nonlinear autoregressive exogenous models [157] and block-oriented models [77].

In this chapter, we explore the use of RC for nonlinear time series modeling and system identification, and demonstrate that a wide class of nonlinear autoregressive models can be realized in this emerging computational paradigm. Together with their energy and memory efficiency, our scheme makes RC an attractive and versatile scheme for signal processing and control-oriented tasks. This chapter is based on the publication [158]. Central to our development is the uniform convergence property; see [22, 23] and Sec. 2.4.1. The main contributions of this chapter are

- We show that RCs with uniformly convergent output feedback dynamics implement *infinite-order nonlinear autoregressive models with exogenous inputs, or NARX(∞) models.*
- We establish asymptotic stationarity and ergodicity of the output of NARX(∞) models induced by RCs with uniformly convergent output feedback dynamics, and further show that their outputs are ergodic in the sense of Birkhoff-Khinchin [84, Theorem 24.1].

A key observation here is that the NARX(∞) models induced by uniformly convergent RCs can also be expressed as *infinite-order nonlinear moving average models with exogenous inputs, or NMAX(∞) models.* By exploiting this equivalence, we establish the asymptotic stationarity and ergodicity of these NARX(∞) models.

Previous works [80, 81] have investigated the potential of RC for approximating input-output maps under stationary stochastic inputs. Different from [80, 81], this chapter is concerned with developing a general theory for realizing NARX(∞) models with RCs, taking into account the stability of the model, and establishing conditions for the asymptotic stationarity and ergodicity properties of the output process of the NARX(∞) model. We then use the NARX(∞) models for time series modelling and system identification using synthetic and real-world data sets, and perform validation diagnostics to evaluate the quality of the model fit.

To highlight the versatility of our proposal, we employ QRCs, ESNs and RCs implemented by globally Lipschitz Lur'e systems to model several data sets. We cast parameter estimation for these RCs as convex optimization problems. The data sets are collected from diverse fields of interest, including nonlinear quantum optics (simulated on QuTiP [159]), finance [160] and a coupled electric drive system [161]. Numerical experiments indicate that ESNs, Lur'e systems and QRCs with only three to five parameters are able to describe these data sets.

The outline of this chapter is as follows. Sec. 6.1 presents NARX(∞) models realized by RCs with uniformly convergent output feedback dynamics and establishes stationarity and (Birkhoff-Khinchin) ergodicity of their outputs (under certain conditions). Sec. 6.3 details parameter estimation for ESNs, Lur'e systems and QRCs as NARX(∞) models. Sec. 6.4 presents applications of RCs on modeling synthetic and real data. In Sec. 6.5, we discuss the impact of RC state measurement noise on stochastic modeling. Concluding remarks and future directions are presented in Sec. 6.6.

6.1 Uniformly convergent feedback dynamics

We are interested in implementing NARX(∞) models using RCs with output-feedback (also see Fig. 6.1),

$$\begin{cases} x_{k+1} = g(x_k, u_k, y_k), \\ \hat{y}_k = h(x_k), \end{cases} \tag{6.1}$$

for all $k \in \mathbb{Z}$, where $x_k \in \mathbb{R}^N$ is the state and $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ is the input. The target output $y_k \in \mathbb{R}$ is related to the one-step ahead prediction $\hat{y}_k$ via $y_k = \hat{y}_k + e_k$, where $e_k \in \mathbb{R}$ is an external noise. Later on, we will consider e and u modeled by discrete-time stochastic processes. For now, we view e and u as sequences. We also consider an equivalent representation of (6.1) given by

$$\begin{cases} x_{k+1} = f(x_k, u_k, e_k), \\ \hat{y}_k = h(x_k), \end{cases} \tag{6.2}$$

where u and e are viewed as external inputs and

$$f(x_k, u_k, e_k) \coloneqq g(x_k, u_k, h(x_k) + e_k). \tag{6.3}$$

In system identification, u is often designed to be persistently exciting [145] to sufficiently excite all modes of the plant. In time series modeling, RCs make (one-step ahead) predictions $\hat{y}_k$ of y_k based on past observations and are not driven by input u. That is, (6.1) becomes

$$\begin{cases} x_{k+1} = g(x_k, y_k), \\ \hat{y}_k = h(x_k). \end{cases} \tag{6.4}$$

We view time series modeling as a special case of system identification and present our results on the latter problem.

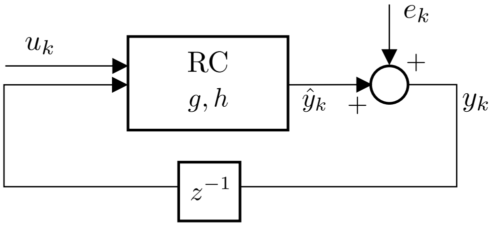


Figure 6.1: Schematic of RCs operating in an output-feedback configuration described by (6.1), where z^{-1} is the one-step time delay operator.

As alluded to in the Introduction, the uniform convergence property is central to our ensuing development. This chapter focuses on RCs that are uniformly convergent with respect to inputs $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$. We say that (6.1) is uniformly convergent if and only if (6.2) is uniformly convergent.

Recall from Sec. 2.4.3 that if an RC governed by (6.2) is uniformly convergent (e.g., if it satisfies Theorem 2.1 for classical RC or Theorem 2.2 for QRC), then it induces a unique time-invariant and causal filter $U_{f,h} : (\mathbb{R}^n)^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ such that when evaluated at any time $k \in \mathbb{Z}$, $\hat{y}_k = U_{f,h}(u, e)|_k := h(x_k^*)$, where x^* is the reference state solution to (6.2). There is a bijective correspondence between $U_{f,h}$ and its associated functional $F_{f,h} : (\mathbb{R}^n)^{\mathbb{Z}_-} \times \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}$, defined as $F_{f,h}(u', e') := U_{f,h}(\tilde{u}', \tilde{e}')|_0$ [25]. Here $\tilde{u}', \tilde{e}'$ are arbitrary extensions of $u' \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $e' \in \mathbb{R}^{\mathbb{Z}_-}$ to $(\mathbb{R}^n)^{\mathbb{Z}}$ and $\mathbb{R}^{\mathbb{Z}}$, respectively. We can recover $U_{f,h}$ from $F_{f,h}$ via $U_{f,h}(u, e)|_k = F_{f,h}(P_n^{\mathbb{Z}_-} \circ z_n^{-k}(u), P^{\mathbb{Z}_-} \circ z^{-k}(e))$ for any $k \in \mathbb{Z}$ (recall that z_n^{-k} is the time shift operator $z_n^{-k}(u)|_{k'} = u_{k'-k}$ for any $k' \in \mathbb{Z}$). This bijection will be useful for establishing statistical properties of the output of NARX(∞) models realized by uniformly convergent RCs; see Lemma 6.2 and Lemma 6.3 below.

Furthermore, if h is uniformly continuous, the filter $U_{f,h}$ can be constructed by taking the initial time $k_0 \to -\infty$ as in Theorem 2.3 in Sec. 2.4.3.

Theorem 6.1. *Consider an RC described equivalently by* (6.1) *or* (6.2) *such that* (6.2) *is uniformly convergent. Let $U_{f,h}$ be the unique time-invariant and causal filter induced by* (6.2)*. If h is uniformly continuous, then for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$, $e \in \mathbb{R}^{\mathbb{Z}}$ and any $k \in \mathbb{Z}$,*

$$\begin{aligned} y_k &= U_{f,h}(u,e)|_k + e_k \\ &= h \circ f(x_{k-1}, u_{k-1}, e_{k-1}) + e_k \\ &= h \circ f(f(x_{k-2}, u_{k-2}, e_{k-2}), u_{k-1}, e_{k-1}) + e_k \\ &\quad\vdots \\ &= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots, e_{k-1}, e_{k-2}, \ldots) + e_k, \end{aligned} \tag{6.5}$$

where the following point-wise limit

$$\begin{aligned} U_{f,h}(u,e)|_k &= \mathcal{F}(u_{k-1}, u_{k-2}, \ldots, e_{k-1}, e_{k-2}, \ldots) \\ &:= \lim_{k_0 \to -\infty} h \circ f(\ldots f(f(\xi, u_{k_0}, e_{k_0}), u_{k_0+1}, e_{k_0+1}) \ldots) \end{aligned} \tag{6.6}$$

exists and is independent of initial condition $\xi \in \mathbb{R}^N$.

In Theorem 6.1, we have written $x_{k+1} = f(x_k, u_k, e_k)$ as in (6.2). Alternatively, we can write $x_{k+1} = g(x_k, u_k, y_k)$ as in (6.1), where g and f are related via (6.3). This leads to the following Corollary.

Corollary 6.1. *Consider an RC described equivalently by* (6.1) *or* (6.2) *such that* (6.2) *is uniformly convergent. Let $U_{f,h}$ be the unique time-invariant and causal filter induced by* (6.2)*. If h is uniformly continuous, then for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$, $e \in \mathbb{R}^{\mathbb{Z}}$ and any $k \in \mathbb{Z}$,*

$$y_k = \mathcal{G}(u_{k-1}, u_{k-2}, \ldots, y_{k-1}, y_{k-2}, \ldots) + e_k,$$

where the following point-wise limit

$$\begin{aligned} &\mathcal{G}(u_{k-1}, u_{k-2}, \ldots, y_{k-1}, y_{k-2}, \ldots) \\ &:= \lim_{k_0 \to -\infty} h \circ g(\ldots g(g(\xi, u_{k_0}, y_{k_0}), u_{k_0+1}, y_{k_0+1}) \ldots) \end{aligned} \tag{6.7}$$

exists and is independent of initial condition $\xi \in \mathbb{R}^N$.

6.2 NARX(∞) models

We apply Corollary 6.1 to show that RCs described equivalently by (6.1) or (6.2), such that (6.2) is uniformly convergent, f is continuous and h is uniformly continuous, implement NARX(∞) models defined in (6.8) below. By Theorem 6.1, such NARX(∞) models can also be expressed as NMAX(∞) models, defined in (6.9) below. We first introduce a probabilistic framework.

Let $(\Omega, \Sigma, \mathbb{P})$ be a complete probability space on which all random variables are defined. For an $\mathbb{R}^m$-valued discrete-time stochastic process $\mathbf{z}$, for any $\omega \in \Omega$, $\mathbf{z}(\omega) = \{\mathbf{z}_k(\omega)\}_{k\in\mathbb{Z}}$ is a realization of $\mathbf{z}$. We view $\mathbf{z}$ as a random sequence, i.e., $\mathbf{z}$ is a stochastic process if $\mathbf{z} : (\Omega, \Sigma) \to ((\mathbb{R}^m)^{\mathbb{Z}}, (\mathcal{R}^m)^{\mathbb{Z}})$ is measurable, where $(\mathcal{R}^m)^{\mathbb{Z}}$ is the σ-algebra generated by cylindrical sets in $(\mathbb{R}^m)^{\mathbb{Z}}$; e.g., see [84, Sec. 36].

Consider RCs described by (6.1) or (6.2) under $\mathbb{R}^n$-valued and $\mathbb{R}$-valued stochastic processes $\mathbf{u}$ and $\mathbf{e}$; i.e., $\mathbf{u} : (\Omega, \Sigma) \to ((\mathbb{R}^n)^{\mathbb{Z}}, (\mathcal{R}^n)^{\mathbb{Z}})$ and $\mathbf{e} : (\Omega, \Sigma) \to (\mathbb{R}^{\mathbb{Z}}, \mathcal{R}^{\mathbb{Z}})$ are measurable. The state $\mathbf{x}_k$ and the one-step ahead prediction $\hat{\mathbf{y}}_k$ are described by a stochastic RC:

$$\begin{cases} \mathbf{x}_{k+1} = g(\mathbf{x}_k, \mathbf{u}_k, \mathbf{y}_k), \\ \quad \hat{\mathbf{y}}_k = h(\mathbf{x}_k), \end{cases}$$

where $\mathbf{y}_k = \hat{\mathbf{y}}_k + \mathbf{e}_k$. Stochasticity of $\mathbf{y}$ arises solely from the stochasticity of $\mathbf{u}$ and $\mathbf{e}$, and the maps g, h are deterministic.

For a uniformly convergent RC described by (6.1) or (6.2) with uniformly continuous h, by Corollary 6.1, $\mathbf{y}_k$ is described by a NARX(∞) model such that for all $k \in \mathbb{Z}$,

$$\mathbf{y}_k = \mathcal{G}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \ldots) + \mathbf{e}_k. \tag{6.8}$$

For each $\omega \in \Omega$, the point-wise limit

$$\mathcal{G}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \ldots) \coloneqq \lim_{k_0 \to -\infty} h \circ g(\ldots g(\xi, \mathbf{u}_{k_0}, \mathbf{e}_{k_0}) \ldots)$$

exists almost surely (a.s.) and is independent of initial condition $\xi \in \mathbb{R}^N$.

Equivalently, for a uniformly convergent RC described by (6.1) or (6.2) with uniformly continuous h, by Theorem 6.1, $\mathbf{y}_k$ is also described by a NMAX(∞) model, such that for all $k \in \mathbb{Z}$

$$\begin{aligned} \mathbf{y}_k &= U_{f,h}(\mathbf{u}, \mathbf{e})|_k + \mathbf{e}_k \\ &= \mathcal{F}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{e}_{k-1}, \mathbf{e}_{k-2}, \ldots) + \mathbf{e}_k. \end{aligned} \tag{6.9}$$

For each $\omega \in \Omega$, the point-wise limit

$$\mathcal{F}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{e}_{k-1}, \mathbf{e}_{k-2}, \ldots) := \lim_{k_0 \to -\infty} h \circ f(\ldots f(\xi, \mathbf{u}_{k_0}, \mathbf{e}_{k_0}) \ldots)$$

exists a.s. and is independent of initial condition $\xi \in \mathbb{R}^N$.

We now show that if in addition, f defined by (6.3) is continuous, then the output $\mathbf{y}$ of a NARX(∞) model given by (6.8) is a well-defined stochastic process. The proof utilizes the equivalence between the NARX(∞) model and its corresponding NMAX(∞) model. We also employ the following result.

Lemma 6.1. *[84, Theorem 13.4] Suppose that $f_1, f_2, \ldots$ are real functions measurable with respect to some sigma algebra Σ. If $\lim_{n\to\infty} f_n$ exists everywhere, then it is measurable with respect to Σ.*

Lemma 6.2. *Consider an RC described equivalently by* (6.1) *or* (6.2) *such that* (6.2) *is uniformly convergent. Let $U_{f,h} : ((\mathbb{R}^n)^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, (\mathcal{R}^n)^{\mathbb{Z}} \times \mathcal{R}^{\mathbb{Z}}) \to (\mathbb{R}^{\mathbb{Z}}, \mathcal{R}^{\mathbb{Z}})$ be the unique time-invariant and causal filter induced by* (6.2)*. Suppose that $h : \mathbb{R}^N \to \mathbb{R}$ is uniformly continuous and $f : \mathbb{R}^N \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^N$ defined by* (6.3) *is continuous. Then $U_{f,h}$ is measurable and for any stochastic processes $\mathbf{u}$ and $\mathbf{e}$, the output $\mathbf{y}$ of the NARX(∞) model defined by* (6.8) *is a well-defined stochastic process.*

Proof. Since $\mathbf{y}$ is also the output of the corresponding NMAX(∞) model (6.9), it follows that $\mathbf{y}$ is a stochastic process if $U_{f,h}$ is measurable. Recall the bijection between $U_{f,h}$ and its functional $F_{f,h}$. This bijection also implies that $U_{f,h}$ is measurable if and only if $F_{f,h}$ is measurable; see [80, Sec. II]. To this end, it suffices to show that the functional $F_{f,h} : ((\mathbb{R}^n)^{\mathbb{Z}_-} \times \mathbb{R}^{\mathbb{Z}_-}, (\mathcal{R}^n)^{\mathbb{Z}_-} \times \mathcal{R}^{\mathbb{Z}_-}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable. By Theorem 6.1, for any $u' \in (\mathbb{R}^n)^{\mathbb{Z}_-}$ and $e' \in \mathbb{R}^{\mathbb{Z}_-}$,

$$F_{f,h}(u', e') = \lim_{k_0 \to -\infty} h \circ f(\ldots f(\xi, u'_{k_0}, e'_{k_0}) \ldots),$$

where the limit exists and is independent of initial condition $\xi \in \mathbb{R}^N$. Fix $\xi \in \mathbb{R}^N$, for any $k_0 \in \mathbb{Z}_-$, define $F^{k_0}_{f,h} : (\mathbb{R}^n)^{\mathbb{Z}_-} \times \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}$ by

$$F^{k_0}_{f,h}(u', e') := h \circ f(\ldots f(\xi, u'_{k_0}, e'_{k_0}) \ldots).$$

Then $\lim_{k_o \to -\infty} F^{k_0}_{f,h}(u', e') = F_{f,h}(u', e')$ point-wise and by Lemma 6.1, $F_{f,h}$ is measurable if $F^{k_0}_{f,h}$ is measurable for all $k_0 = \{\ldots, -2, -1\}$. To show this, we write

$$F^{k_0}_{f,h}(u', e') = h \circ \overline{f}_{k_0}(\mathcal{P}^{k_0}_n(u'), \mathcal{P}^{k_0}(e')),$$

where $\mathcal{P}^{k_0}_n := \prod_{j=k_0}^{-1} P^j_n : (\mathbb{R}^n)^{\mathbb{Z}_-} \to (\mathbb{R}^n)^{-k_0}$, $\mathcal{P}^{k_0} := \prod_{j=k_0}^{-1} P^j : \mathbb{R}^{\mathbb{Z}_-} \to \mathbb{R}^{-k_0}$. Here, $P^j_n(u') = u'_j$ and $P^j(e') = e'_j$. Furthermore, $\overline{f}_{k_0} : (\mathbb{R}^n)^{-k_0} \times \mathbb{R}^{-k_0} \to \mathbb{R}^N$ is given by

$$\overline{f}_{k_0}(\mathcal{P}^{k_0}_n(u'), \mathcal{P}^{k_0}(e')) = f(\ldots f(\xi, u'_{k_0}, e'_{k_0}) \ldots).$$

Since $h, \mathcal{P}^{k_0}_n$ and $\mathcal{P}^{k_0}$ are measurable, it remains to show that $\overline{f}_{k_0}$ is measurable. To this end, to simplify notation, for any $i, j \in \mathbb{Z}_+$, let $u'_{k_0-j:-1-i}$ and $e'_{k_0-j:-1-i}$ denote the concatenation of $\{u'_{k_0-j}, \ldots, u'_{-1-i}\}$ and $\{e'_{k_0-j}, \ldots, e'_{-1-i}\}$ into a column vector, respectively. We can define $\overline{f}_{k_0}$ recursively via

$$\overline{f}_{k_0-1}(u'_{k_0-1:-1}, e'_{k_0-1:-1}) = f(\overline{f}_{k_0}(u'_{k_0-1:-2}, e'_{k_0-1:-2}), u'_{-1}, e'_{-1}).$$

Using this recursion and continuity of f, an inductive argument on k_0 shows that $\overline{f}_{k_0}$ is continuous. Hence $F_{f,h}$ and $U_{f,h}$ are measurable, and $\mathbf{y}$ defined by (6.8) is a well-defined stochastic process. □

In practice, when we are using NARX(∞) models for system identification, we are only given the input time series $\mathbf{u}$ and the output time series $\mathbf{y}$. With these data, we can produce an one-step ahead predictoin $\hat{\mathbf{y}}_k = \mathcal{G}(\mathbf{u}_{k-1}, \mathbf{u}_{k-2}, \ldots, \mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \ldots)$ of the target output data $\mathbf{y}_k$. For time series modeling, the one-step ahead prediction is $\hat{\mathbf{y}}_k = \mathcal{G}(\mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \ldots)$. In both cases, the noise process $\mathbf{e}$ cannot be observed but it is useful for evaluating how well the NARX(∞) model fits the target data. More explicitly, we can impose a statistical assumption on $\mathbf{e}$ and check if the residual $\hat{\mathbf{e}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$ has the same statistical properties of $\mathbf{e}_k$; see [153] for further detail. This suggests that the one-step ahead prediction captures useful information of the data. Throughout this chapter, we impose the following standard assumptions on $\mathbf{u}$ and $\mathbf{e}$:

Assumption 6.1. *$\boldsymbol{u}$ and $\mathbf{e}$ are independent, and $\mathbf{e}$ is identically and independently distributed (iid).*

Assumption 6.1 lays the basis for analyzing the RC residual $\hat{\mathbf{e}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$. If the RC prediction $\hat{\mathbf{y}}$ describes the target data $\mathbf{y}$ adequately, $\hat{\mathbf{e}}$ should be a proxy for $\mathbf{e}$. We will introduce some statistical tools to test this in Sec. 6.4.

6.2.1 Asymptotic stationarity and ergodicity

In this subsection, we derive conditions for which $\mathbf{y}$ defined by (6.8) is asymptotically stationary and/or ergodic given that $\mathbf{u}, \mathbf{e}$ are stationary and/or ergodic (see Definitions 6.1 and 6.2 below). We equip $\mathbb{R}^m$ with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^m)$.

Definition 6.1 (Stationarity)**.** *An $\mathbb{R}^m$–valued stochastic process $\tilde{\mathbf{z}}$ is stationary if for any $\tau, k_i \in \mathbb{Z}$, $l \in \mathbb{Z}_+$ and $H_i \in \mathcal{B}(\mathbb{R}^m)$, where $i = 0, \ldots, l$, it holds that*

$$\mathbb{P}(\tilde{\mathbf{z}}_{k_0} \in H_0, \ldots, \tilde{\mathbf{z}}_{k_l} \in H_l) = \mathbb{P}(\tilde{\mathbf{z}}_{k_0+\tau} \in H_0, \ldots, \tilde{\mathbf{z}}_{k_l+\tau} \in H_l).$$

A stochastic process $\mathbf{z}$ induces a probability measure $\mathbb{P}_{\tilde{\mathbf{z}}} = \mathbb{P} \circ \tilde{\mathbf{z}}^{-1}$ on $((\mathbb{R}^m)^{\mathbb{Z}}, (\mathcal{R}^m)^{\mathbb{Z}})$. Ergodicity of $\tilde{\mathbf{z}}$ is defined as followed.

Definition 6.2 (Ergodicity)**.** *An $\mathbb{R}^m$–valued stochastic process $\tilde{\mathbf{z}}$ is ergodic if for any $A \in (\mathcal{R}^m)^{\mathbb{Z}}$ with $(z_m^{-1})^{-1}(A) = A$, either $\mathbb{P}_{\tilde{\mathbf{z}}}(A) = 0$ or $\mathbb{P}_{\tilde{\mathbf{z}}}(A) = 1$, where $(z_m^{-1})^{-1}(A)$ is the pre-image of A under the shift operator z_m^{-1}.*

Definition 6.3 (Birkhoff-Khinchin ergodic)**.** *An $\mathbb{R}$-valued process $\tilde{\mathbf{z}}$ is Birkhoff-Khinchin ergodic if $\lim_{L\to\infty} \frac{1}{L} \sum_{k=0}^{L-1} \tilde{\mathbf{z}}_k = \mathbb{E}[\tilde{\mathbf{z}}_0]$ a.s., where $\mathbb{E}[\cdot]$ is the expectation over $\mathbb{P}$.*

Combining the discussion on [84, Page 526] and [84, Theorem 24.1], we also have the following theorem.

Theorem 6.2. *Suppose that* $\tilde{\mathbf{z}}'_k = H(z_m^{-k}(\tilde{\mathbf{z}}))$ *for all* $k \in \mathbb{Z}$, *where* $\tilde{\mathbf{z}}$ *is an* $\mathbb{R}^m$*–valued stationary and ergodic stochastic process. If* $H : (\mathbb{R}^m)^{\mathbb{Z}} \to \mathbb{R}$ *is measurable and integrable, then* $\tilde{\mathbf{z}}'$ *is Birkhoff-Khinchin ergodic.*

To establish statistical properties of the output of a NARX(∞) model defined by (6.8), we again exploit the bijective correspondence with its associated NMAX(∞) model given by (6.9). We will also apply the following result.

Theorem 6.3. *[84, Theorem 36.4] Suppose that an* $\mathbb{R}^m$*–valued stochastic process* $\boldsymbol{z}$ *is stationary and ergodic. Let* $U : ((\mathbb{R}^m)^{\mathbb{Z}}, (\mathcal{R}^m)^{\mathbb{Z}}) \to (\mathbb{R}^{\mathbb{Z}}, \mathcal{R}^{\mathbb{Z}})$ *be measurable. Then the stochastic process* $\boldsymbol{y}$ *defined by* $\boldsymbol{y}_k = U(\boldsymbol{z})|_k$ *for all* $k \in \mathbb{Z}$ *is stationary and ergodic.*

Lemma 6.3. *Consider an RC described equivalently by* (6.1) *or* (6.2) *such that* (6.2) *is uniformly convergent. Under the assumptions of Lemma 6.2, the process* $\boldsymbol{y}$ *defined by* (6.8) *is stationary (resp. ergodic) if* $\mathbf{u}$ *and* $\mathbf{e}$ *are stationary (resp. ergodic). Furthermore, suppose that* $h \circ \boldsymbol{x}_0^*$ *and* $\mathbf{e}_0$ *are integrable, where* $\boldsymbol{x}^*$ *is the reference state solution to* (6.2)*. Then under the assumptions of Lemma 6.2,* $\boldsymbol{y}$ *is Birkhoff-Khinchin ergodic if* $\mathbf{u}$*,* $\mathbf{e}$ *are stationary and ergodic.*

Proof. By (6.9), we have $\mathbf{y}_k = U_{f,h}(\mathbf{u}, \mathbf{e})|_k + \mathbf{e}_k$, where $U_{f,h}$ is the unique filter induced by (6.2). By Lemma 6.2, $U_{f,h}$ is measurable. Now $\mathbf{y}$ is stationary (resp. ergodic), given that $\mathbf{u}, \mathbf{e}$ are stationary (resp. ergodic), follows from Theorem 6.3. To show the second part of the Lemma, by Theorem 6.2, it suffices to show that $F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}), P^{\mathbb{Z}_-}(\mathbf{e})) + \mathbf{e}_0 : (\Omega, \Sigma, \mathbb{P}) \to \mathbb{R}$ is integrable, where $F_{f,h}$ is the functional induced by (6.2). Recall that

$$\begin{aligned} &\int_\Omega \left| F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}(\omega)), P^{\mathbb{Z}_-}(\mathbf{e}(\omega))) + \mathbf{e}_0(\omega) \right| \mathbb{P}(\mathrm{d}\omega) \\ &= \int_\Omega |h(\mathbf{x}_0^*(\omega)) + \mathbf{e}_0(\omega)| \mathbb{P}(\mathrm{d}\omega) \\ &\leq \int_\Omega \left(|h(\mathbf{x}_0^*(\omega))| + |\mathbf{e}_0(\omega)|\right) \mathbb{P}(\mathrm{d}\omega). \end{aligned}$$

Integrability of $F_{f,h}(P_n^{\mathbb{Z}_-}(\mathbf{u}), P^{\mathbb{Z}_-}(\mathbf{e})) + \mathbf{e}_0$ now follows from the integrability of $h \circ \mathbf{x}_0^*$ and $\mathbf{e}_0$. $\square$

Remark 6.1. *We remark that when h is uniformly continuous (as a sufficient condition in Theorem 6.1), a sufficient condition for the integrability of $h \circ \boldsymbol{x}_0^*$ is that there exists $0 < M < \infty$ such that $\|\boldsymbol{x}_0^*\| \leq M$ a.s.. This holds for RCs with compact state spaces, such as ESNs and QRCs. We also point out that $\mathbf{e}$ is stationary and ergodic due to the iid assumption on $\mathbf{e}$ (Assumption 6.1) [84, Sec. 36].*

6.3 Parameter estimation

In previous sections, we have shown that a large class of NARX(∞) models can be realized by uniformly convergent RCs of the form (6.1). In this section, we present three example RC classes, including ESNs, QRCs and RCs realized by Lur'e systems, as NARX(∞) models. ESNs are pioneering software implementation of RC and have found successful applications in chaotic system modelling [10]. Recently, there is substantial interest in physical RC, hardware implementation of RCs for real-time stochastic modeling with low energy and memory requirement. One potential hardware RC scheme is the Lur'e system, which is ubiquitous in robitics [162]. While Lur'e systems have been of interest in stability analysis and observer design [144], few stochastic modeling applications of them have been developed. Here, we demonstrate that Lur'e systems, such as one-link flexible joint robots, can implement NARX(∞) models for stochastic modelling. Another hardware implementation of RC is the QRC, implemented on current noisy quantum machines, see Chapter 4. Here, we extend the single-input QRCs introduced in Chapter 4 to multiple inputs, and employ them for stochastic modeling.

These RCs are described by (6.1) or equivalently (6.2), where f and h satisfy the assumptions in Lemma 6.3. Furthermore, ESNs and QRCs admit a compact state-space, and we will show that there exists $0 < M < \infty$ such that $\|\mathbf{x}_0^*\| \leq M$ a.s. for a uniformly convergent Lur'e system considered here. With these observations, by Lemma 6.3, these RCs realize NARX(∞) models whose output $\mathbf{y}$ and their one-step ahead prediction $\hat{\mathbf{y}}$ are stationary and ergodic. For these RCs, the one-step ahead prediction has the form $\hat{\mathbf{y}}_k = W^\top \overline{h}(\mathbf{x}_k) + W_c$, where $W \in \mathbb{R}^N, W_c \in \mathbb{R}$ are the output parameters to be estimated.

Given time series data $\{\mathbf{y}_0, \dots, \mathbf{y}_L\}$ and $\{\mathbf{u}_0, \dots, \mathbf{u}_L\}$, parameters W, W_c are optimized via least squares based on $\{\mathbf{y}_{L_1+1}, \dots, \mathbf{y}_{L_2}\}$ and $\{\mathbf{u}_{L_1+1}, \dots, \mathbf{u}_{L_2}\}$, so that $\hat{\mathbf{y}}_k$ approximates the target $\mathbf{y}_k$ (the first L_1 data are for washing out the effect of the RC's initial condition). We apply Theorem 2.1 to (6.2) to ensure the RC's convergence, the resulting optimization of W, W_c becomes a constrained least squares problem,

$$\min_{W,W_c} \frac{1}{L_t} \sum_{k=L_1+1}^{L_2} |\mathbf{y}_k - \hat{\mathbf{y}}_k|^2, \quad \text{subject to } G(W) \leq 0, \tag{6.10}$$

where $L_t = L_2 - L_1$. We will derive a convex constraint $G(\cdot)$ for each example RC model.

If $\hat{\mathbf{y}}^2$, $\mathbf{y}^2$ and $\mathbf{y}\hat{\mathbf{y}}$ are Birkhoff-Khinchin ergodic, then parameter estimation is consistent in the sense that as $L_t \to \infty$, the cost function given in (6.10) becomes $\min_{W,W_c} \mathbb{E}[|\mathbf{y}_0 - \hat{\mathbf{y}}_0|^2]$ a.s..

6.3.1 Echo-state networks (ESNs)

An ESN with state $x_k \in \mathbb{R}^N$ is governed by

$$\begin{cases} x_{k+1} = \tanh(Ax_k + Bu_k + Cy_k), \\ \hat{y}_k = W^\top x_k + W_c \end{cases} \tag{6.11}$$

where $\tanh(\cdot)$ applies to a vector elementwise. Elements of A, B, C are drawn independently and uniformly from $[-1, 1]$ (for time series modeling, we set $C = 0$). To apply Theorem 2.1, we re-express (6.11) in the form of (6.2) by substituting $y_k = \hat{y}_k + e_k = W^\top x_k + W_c + e_k$ into (6.11):

$$\begin{aligned} x_{k+1} &= f_{\text{ESN}}(x_k, u_k, e_k) \\ &:= \tanh((A + CW^\top)x_k + C(W_c + e_k) + Bu_k). \end{aligned}$$

We set $P = I$ so that $\|x\|_P = \|x\|$. Then for any $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$, $\sup_{k\in\mathbb{Z}} \|f_{\text{ESN}}(0, u_k, e_k)\| < \infty$ and (2.8) in Theorem 2.1 holds. To derive a convex constraint on W, for any $x_1, x_2 \in \mathbb{R}^N$ and any $k \in \mathbb{Z}$, we have

$$\|f_{\text{ESN}}(x_1, u_k, e_k) - f_{\text{ESN}}(x_2, u_k, e_k)\| \leq \sigma_{\max}(A + CW^\top)\|x_1 - x_2\|,$$

where $\sigma_{\max}(\cdot)$ is the maximum singular value. By (2.9) in Theorem 2.1, an ESN is uniformly convergent if $\sigma_{\max}(A + CW^\top) < 1$. We optimize W, W_c for the ESN using SeDuMi [163] through YALMIP [164] by passing the non-strict inequality constraint:

$$\begin{bmatrix} I & A + CW^\top \\ (A + CW^\top)^\top & I \end{bmatrix} \succeq 10^{-3} I. \tag{6.12}$$

6.3.2 Lur'e systems

Consider a Lur'e system with state $x_k \in \mathbb{R}^N$ described by

$$\begin{cases} x_{k+1} = Ax_k + B\tanh(u_k) + C\tanh(y_k), \\ \hat{y}_k = W^\top x_k + W_c, \end{cases} \tag{6.13}$$

where $\tanh(\cdot)$ applies to a vector elementwise. Elements of A, B, C are drawn independently and uniformly from $[-1, 1]$ (for time series modeling, we set $C = 0$). Given any $P > 0$, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$, we re-express (6.13) in the form of (6.2) by substituting $y_k = \hat{y}_k + e_k$ into (6.13),

$$\begin{aligned} x_{k+1} &= f_{\mathrm{Lur'e}}(x_k, u_k, e_k) \\ &\coloneqq Ax_k + C\tanh(W^\top x_k + W_c + e_k) + B\tanh(u_k) \end{aligned}$$

and (2.8) in Theorem 2.1 holds since

$$\sup_{k \in \mathbb{Z}} \| f_{\mathrm{Lur'e}}(0, u_k, e_k) \|_P \coloneqq M_{u,e} < \infty.$$

We show that (2.9) in Theorem 2.1 and hence the uniform convergence of a Lur'e system, holds if for some $P \succ 0$, $\delta > 0$ and $\theta \in (0, 1)$, the linear matrix inequalities (LMIs),

$$P - \delta I \prec 0,$$

$$\begin{bmatrix} -\theta P & A^\top P & \delta W C^\top & A^\top P \\ PA & -P & 0 & 0 \\ \delta C W^\top & 0 & -\delta I & 0 \\ PA & 0 & 0 & P - \delta I \end{bmatrix} \preceq 0 \tag{6.14}$$

are satisfied (we set $\delta = 0.001$ and $\theta = 0.99$ in our numerical examples).

More generally, we can consider Lur'e systems (6.13) with other globally Lipschitz functions besides $\tanh(\cdot)$. Here, we derive (6.14) for a more general class of globally Lipschitz Lur'e systems described by

$$x_{k+1} = f_{\mathrm{Lur'e}}(x_k, u_k, e_k) = Ax_k + B\Theta(u_k) + C\Phi(x_k, u_k),$$

where for any $P \succ 0$, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$, $\sup_{k\in\mathbb{Z}} \|f_{\mathrm{Lur'e}}(0, u_k, e_k)\|_P < \infty$, so that (2.8) in Theorem 2.1 holds. Suppose that there exists a $N \times N$ matrix X such that, for any $x_1, x_2 \in \mathbb{R}^N$ and $u_k \in \mathbb{R}^n$,

$$\|C\Phi(x_1, u_k) - C\Phi(x_2, u_k)\| \leq \|X(x_1 - x_2)\|. \tag{6.15}$$

Let $z = x_1 - x_2$ and $w = C\Phi(x_1, u_k) - C\Phi(x_2, u_k)$, (6.15) becomes $w^\top w - z^\top X^\top X z \leq 0$, that is,

$$\begin{bmatrix} z^\top & w^\top \end{bmatrix} \begin{bmatrix} -X^\top X & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} \leq 0. \tag{6.16}$$

Now (2.9) in Theorem 2.1 holds if there exists $\theta \in (0, 1)$ such that $(Az + w)^\top P(Az + w) - \theta z^\top P z \leq 0$, that is,

$$\begin{bmatrix} z^\top & w^\top \end{bmatrix} \begin{bmatrix} A^\top PA - \theta P & A^\top P \\ PA & P \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} \leq 0. \tag{6.17}$$

From (6.16), (6.17) holds if there exists $\delta > 0$ such that

$$\begin{bmatrix} A^\top PA - \theta P & A^\top P \\ PA & P \end{bmatrix} - \delta \begin{bmatrix} -X^\top X & 0 \\ 0 & I \end{bmatrix} \preceq 0. \tag{6.18}$$

Assuming that $P - \delta I \prec 0$ and $\delta > 0$, we claim that (6.18) is equivalent to

$$\begin{bmatrix} -\theta P & A^\top P & \delta X^\top & A^\top P \\ PA & -P & 0 & 0 \\ \delta X & 0 & -\delta I & 0 \\ PA & 0 & 0 & P - \delta I \end{bmatrix} \preceq 0. \tag{6.19}$$

To see this, under the assumptions $P - \delta I \prec 0$ and $\delta > 0$, the Schur complement shows that (6.18) is equivalent to $A^\top PA - \theta P + \delta X^\top X - A^\top P(P - \delta I)^{-1}PA \preceq 0$. The same condition is obtained by applying the Schur complement twice to (6.19).

For the Lur'e system governed by (6.13), the Lipschitz condition (6.16) holds for $X = CW^\top$, leading to (6.14). We optimize W, W_c for the Lur'e system using SeDuMi [163] through YALMIP [164] (by replacing the strict LMI $P - \delta I \prec 0$ in (6.14) with the non-strict LMI $P - \delta I \preceq -10^{-4}I$).

To show that $\|\mathbf{x}_0^*\| \leq M < \infty$ a.s. for a uniformly convergent Lur'e system (so that Lemma 6.3 holds), first note that $\|x_0^*\|_P \leq \frac{M_{u,e}}{1-\theta}$ [23, Theorem 1], where x^* is the reference state solution under u, e. Therefore, it suffices to show that $\sup_{u\in(\mathbb{R}^n)^\mathbb{Z}, e\in\mathbb{R}^\mathbb{Z}} M_{u,e} < \infty$. Let $P = LL^\top$ be the Cholesky decomposition of P for which (6.14) holds, then

$$\begin{aligned}
M_{u,e} &= \sup_{k\in\mathbb{Z}} \|f_{\mathrm{Lur'e}}(0, u_k, e_k)\|_P \\
&\leq \sup_{k\in\mathbb{Z}} \left(\|C\tanh(W_c + e_k)\|_P + \|B\tanh(u_k)\|_P\right) \\
&\leq \|C\|_P + \sup_{k\in\mathbb{Z}} \|L^\top B\tanh(u_k)\| \\
&\leq \|C\|_P + \|L^\top B\|\sqrt{n}.
\end{aligned}$$

Therefore, $\sup_{u\in(\mathbb{R}^n)^\mathbb{Z}, e\in\mathbb{R}^\mathbb{Z}} M_{u,e} < \infty$.

6.3.3 Quantum reservoir computers (QRCs)

Consider an N-qubit QRC described by

$$\begin{cases} \rho_{k+1} = (1-\epsilon)T(u_k, y_k)\rho_k + \epsilon\rho_*, \\ \hat{y}_k = \sum_{i=1}^N W_i \mathrm{Tr}(Z^{(i)}\rho_k) + W_c, \end{cases} \tag{6.20}$$

where ρ_k is a $2^N \times 2^N$ density operator, ρ_* is an arbitrary but fixed density operator, $Z^{(i)}$ is the Pauli-Z operator acting on qubit i and $\epsilon \in (0,1)$. Throughout this chapter, we set ρ_* with $(1,1)$–th element to be one and all other elements zero. The dynamics $T(u_k, y_k)$

is a CPTP map determined by u_k and y_k. We consider

$$T(u_k, y_k) = \frac{1}{n+1}\left(\sum_{j=1}^{n} g(u_k^{(j)})T_j + g(y_k)T_{n+1} + \left(n+1-\sum_{j=1}^{n} g(u_k^{(j)}) - g(y_k)\right)T_{n+2}\right),$$

where $u_k^{(j)}$ is the j-th component of $u_k \in \mathbb{R}^n$. Here, $g(\cdot) \in [0,1]$ is a continuous function which can be chosen suitably according to the task. More generally, different g's can be applied to $u_k^{(j)}$ and y_k (for time series modeling, we set $g(u_k) = 0$). In this chapter, we find that the globally Lipschitz logistic function $g(s) = \frac{1}{1+\exp(-s)}$ with Lipschitz constant $L_g = 1/4$ performs well. The input-independent CPTP maps T_l for $l = 1, \ldots n+2$ are governed by unitary matrices $U_l = U_l(\gamma_l)$ such that $T_l(\rho_k) = U_l \rho_k U_l^\dagger$. For $l = 1, \ldots, n+1$, γ_l are randomly chosen and fixed at the onset, and we fix all elements of γ_{n+2} to be π; see Appendix C.1 for the details. Such QRCs are multi-input extension of the QRCs proposed in Chapter 4 or [115]. They can be implemented on current NISQ machines using schemes proposed in Chapter 4.

To establish the uniform convergence property, we invoke Theorem 2.2 with respect to the Schatten 1–norm. That is, a QRC is 1–uniformly convergent if there exists some $\theta \in (0,1)$ (independent of u, e) such that, for any density operators ρ_j ($j = 1,2$), any $k \in \mathbb{Z}$, $u \in (\mathbb{R}^n)^{\mathbb{Z}}$ and $e \in \mathbb{R}^{\mathbb{Z}}$,

$$\|(1-\epsilon)T(u_k, y_{k,1})\rho_1 + \epsilon\rho_* - ((1-\epsilon)T(u_k, y_{k,2})\rho_2 + \epsilon\rho_*)\|_1 \leq \theta\|\rho_1 - \rho_2\|_1,$$

where $y_{k,j} = \sum_{i=1}^{N} W_i \mathrm{Tr}(Z^{(i)}\rho_j) + W_c + e_k$ is the output associated to ρ_j. An upper bound for the left-hand side of the above inequality is

$$\begin{aligned}
&\|(1-\epsilon)T(u_k, y_{k,1})\rho_1 + \epsilon\rho_* - ((1-\epsilon)T(u_k, y_{k,2})\rho_2 + \epsilon\rho_*)\|_1 \\
&\leq (1-\epsilon)\left(\|\rho_1 - \rho_2\|_1 + \frac{|g(y_1) - g(y_2)|}{n+1}\|T_{n+1} - T_{n+2}\|_{1-1}\right) \\
&\leq \left((1-\epsilon) + \frac{2L_g(1-\epsilon)}{n+1}\sum_{i=1}^{N}|W_i|\right)\|\rho_1 - \rho_2\|_1,
\end{aligned}$$

where the last inequality follows from $\|T_{n+1} - T_{n+2}\|_{1-1} \leq 2$ [105, Theorem 2.2] and $|g(y_1) - g(y_2)| \leq L_g|y_1 - y_2| \leq L_g \sum_{i=1}^{N}|W_i|\|\rho_1 - \rho_2\|_1$; see Appendix B.3. A QRC described by (6.20) is 1–uniformly convergent if there exists some $\theta \in (0,1)$ such that the following convex constraint holds:

$$\sum_{i=1}^{N} |W_i| - \frac{\theta + \epsilon - 1}{1 - \epsilon} \frac{n+1}{2L_g} \leq 0. \tag{6.21}$$

Throughout, we set $\theta = 0.99$. We optimize W using the "fmincon" command in Matlab by setting $\hat{W}, \hat{W}_c = \text{argmin}_{W,W_c} \frac{1}{L_t} \sum_{k=L_1+1}^{L_2} |\mathbf{y}_k - \hat{\mathbf{y}}_k|^2$ as the initial guess.

We remark that (6.12), (6.14) and (6.21) are only sufficient for the uniform convergence. Nevertheless, numerical experiments developed below suggest that ESNs, Lur'e systems and QRCs under these constraints are adequate in describing the data sets considered. It is an interesting and important future direction to find relaxation of these sufficient conditions.

6.4 Numerical examples

We employ ESNs, Lur'e systems and QRCs to model three time series. For a non-stationary time series, we remove its trend and seasonal components and model the remaining stationary part. Stationarity is tested based on the augmented Dickey-Fuller [165] and KPSS [166] tests within 95% confidence interval (CI). Given time series data $\{\mathbf{y}_0, \ldots, \mathbf{y}_L\}$, we randomly generate 50 ESNs, Lur'e systems with $N = 2, \ldots, 10$, and 50 QRCs with qubit number $N = 2, \ldots, 5$, so that the number of output parameters is $N + 1$.

For each ESN, Lur'e system and QRC, we washout the effect of their initial conditions using $\{\mathbf{y}_0, \ldots, \mathbf{y}_{L_1}\}$, with $L_w = L_1 + 1$ washout data points. Output parameters W, W_c are optimized via constrained least squares (see Sec. 6.3) based on $\{\mathbf{y}_{L_1+1}, \ldots, \mathbf{y}_{L_2}\}$, with $L_t = L_2 - L_1$ training data. The fitted ESNs, Lur'e systems and QRCs are selected via Akaike's final prediction error (FPE) criterion using validation data $\{\mathbf{y}_{L_2+1}, \ldots, \mathbf{y}_L\}$, defined as

$$\text{FPE} := \left(\frac{L_v + N + 1}{L_v - N + 1} \right) \text{MSE},$$

where $L_v = L - L_2$ is the number of validation data and

$$\text{MSE} := \frac{1}{L_v} \sum_{k=L_2+1}^{L} \hat{\mathbf{e}}_k^2 = \frac{1}{L_v} \sum_{k=L_2+1}^{L} (\mathbf{y}_k - \hat{\mathbf{y}}_k)^2.$$

We remark that FPE assumes the estimated output parameters are unbiased; e.g., see [167]. Although this cannot be guaranteed in general, we observe that the decrease in MSE is small as N increases and FPE prefers the lowest order $N = 2$. Furthermore, RCs selected by FPE achieve residuals that satisfy Assumption 6.1, suggesting that they are adequate in describing these time series data. We therefore choose $N = 2$ for all RCs on all time series modeling tasks. It is a future research theme to develop model selection methods for the proposed scheme. We test if the RC residuals $\hat{\mathbf{e}}_k$ are uncorrelated using sample autocorrelation (ACF) via the 'ggAcf' command in R [168], and test if $\mathbf{e}_k$ are Gaussian using the Lilliefors test [169] and Q-Q plot (via the 'lillietest' and the 'qqplot' commands in Matlab).

Throughout, we set $\epsilon = 0.9$ in (6.20) for QRCs. In practice, QRCs implemented on NISQ devices experience decoherence. To investigate its effect on stochastic modeling tasks, we simulate the selected QRCs under decoherence as in Appendix B.4. Output parameters of these QRCs are re-optimized under the constraint (6.21). We simulate GAD channels with finite temperature parameters $\lambda = \{0.2, 0.4, 0.6, 0.8, 1\}$. The decoherence strengths $\psi = \{10^{-4}, 10^{-3}, 10^{-2}\}$ for both channels are set to be experimentally feasible in superconducting quantum devices [44].

We employ the root mean-squared error (RMSE) to compare among different models on the same time series data, where $\mathrm{RMSE} := \sqrt{\mathrm{MSE}}$; see [170] for further discussions on RMSE in the context of time series modeling. We report the selected RC's RMSE and the average RMSE of 50 randomly generated RCs with the same dimension N (or qubit number) as the selected RC.

6.4.1 Nonlinear quantum optics

This example demonstrates that RCs can act as nonlinear Wiener filters to extracting the signal part of a highly noisy time series from nonlinear quantum optics. Kerr-nonlinear optical cavities are fundamental building blocks in realizing photonic sequential logic [171]. We consider a low-photon-number Kerr cavity with two input-output ports (Fig. 6.2(a) or

[171, Fig. 2(a)]), whose internal mode is governed by a Hamiltonian $H_0 = \Delta a^\dagger a + \chi (a^\dagger)^2 a^2$, where a is the annihilation operator, $\Delta = 100$ is the detuning from a reference frequency, and $\chi = -5$ governs the nonlinearity. The cavity is coupled to two incoming fields $\alpha_{\mathrm{in}j}(t)$ via operators $L_j = -\iota\sqrt{\kappa_j}a$, where $\iota = \sqrt{-1}$ and $\kappa_j = 25$ for $j = 1, 2$. Here, $\alpha_{\mathrm{in}1}(t)$ is a coherent field with a constant amplitude $\eta = 21.5$ and $\alpha_{\mathrm{in}2}(t)$ is in the vacuum state.

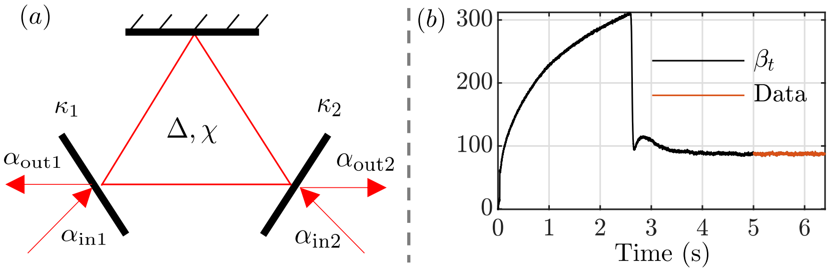


Figure 6.2: (a) Kerr-nonlinear cavity with two input-output ports. The top mirror is fully reflective (without loss) and the other mirrors are partially transmitting. (b) The simulated trajectory of β_k and the data employed.

We obtain a discretized trajectory $\beta_k = \mathrm{Tr}((L_2 + L_2^\dagger)\rho_k) + \eta_k$ from homodyne measurement on the second output port with sampling time 10^{-3} on Qutip [159]. Here, ρ_k is the state at time k and η_k is a quantum Gaussian white noise. Our goal is to employ RCs to separate the 'signal part' $\alpha_k = \mathrm{Tr}((L_2 + L_2^\dagger)\rho_k)$ from the highly noisy β_k. We simulate β_k starting in the vacuum state on a truncated Hilbert space of dimension 1000 for 6.4s and employ the data after 5s until ρ_k reaches a steady state, see Fig. 6.2(b).

We set $L_w = 99$, $L_t = 1000$ and $L_v = 300$. ESN, Lur'e and QRC with $N = 2$ achieve comparable RMSE and average RMSE, their residual sample ACFs show no autocorrelation within 95% CI and pass the Lilliefors test. See Table 6.1 for the p-values of the Lilliefors test, dimension (or qubit number) N, RMSE and average RMSE. We report 0.5 for p-values≥ 0.5.

In Fig. 6.3(a)(b)(c), we observe that the ESN, QRC and Lur'e predictions follow the signal α_k with a normalized root mean-squared error between them of $\sqrt{\frac{\sum_{k=L_2+1}^{L}(\alpha_k - \hat{\mathbf{y}}_k)^2}{\sum_{k=L_2+1}^{L}\alpha_k^2}} = 0.0047, 0.0048, 0.0040$, respectively. These suggest that these RCs can act as a nonlinear Wiener filter, separating the signal α_k from the noisy time series β_k.

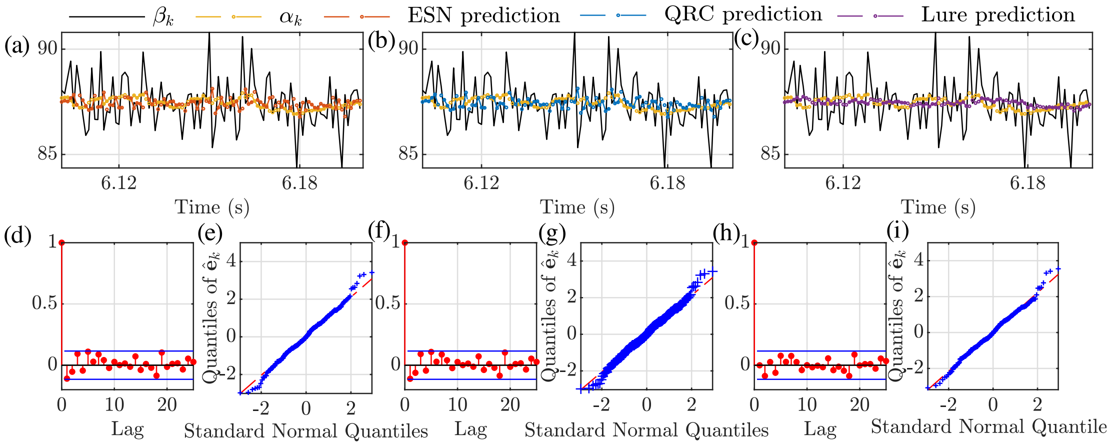


Figure 6.3: The noisy time series β_k, the signal $\alpha_k = \mathrm{Tr}((L_2 + L_2^\dagger)\rho_k)$, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e perdiction on the first 100 validation data. (d) The ESN residual sample ACF (horizontal blue lines show the 95% CI). (e) The ESN residual Q-Q plot. (f) The QRC residual sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACf. (i) The Lur'e residual Q-Q plot.

Table 6.1: Dimension (or qubit number) N, p-value, RMSE and average RMSE ($\overline{\text{RMSE}}$) of RCs for stochastic modeling.

	Optics			Finance			Coupled eletric drive system		
	ESN	Lur'e	QRC	ESN	Lur'e	QRC	ESN	Lur'e	QRC
N	2	2	2	2	2	2	2	2	2
p-value	0.5	0.5	0.5	0.5	0.23	0.5	0.5	0.29	0.44
RMSE	1.12	1.10	1.12	0.047	0.047	0.049	0.10	0.11	0.11
$\overline{\text{RMSE}}$	1.12	1.12	1.12	0.055	0.058	0.048	0.12	0.14	0.11

Fig. 6.3 also plots the RCs' residual sample ACF and Q-Q plot. For all numerical examples, we plot the residual sample ACF up to $L_v/4$ lags ($L_v/4$ is recommended for obtaining reliable estimate of sample ACF [153]). We observe that all RCs' residuals show no autocorelation and pass the Lilliefors test with p-value≥ 0.5.

For all decoherence strengths, QRC obtains RMSE $= 1.12$ as the noiseless QRC, and its residuals show no autocorrelation and pass the Lilliefors test with p-value≥ 0.5. Furthermore, QRC under decoherence can still effectively extract α_k from β_k.

6.4.2 Finance

This time series describes weekly 5/1-year adjustable rate mortgage average (2005-2020) in the US [160]. After removing trend and seasonal components using the 'mstl' R command [168], the data is highly correlated up to 300 lags. This example tests RCs' ability to model highly correlated data. We set $L_w = 100$, $L_t = 580$ and $L_v = 100$. All RCs achieve comparable RMSE. The QRC and Lur'e average RMSEs are similar to their RMSE, whereas the difference for ESN is more pronounced. All RCs achieve uncorrelated residuals (within 95% CI) and pass the Lillefors test; see Table 6.1 and Fig. 6.4. In Fig. 6.4, we plot the target and predicted time series with the trend and seasonal components. These observations suggest that these RCs are capable of modeling this highly correlated time series.

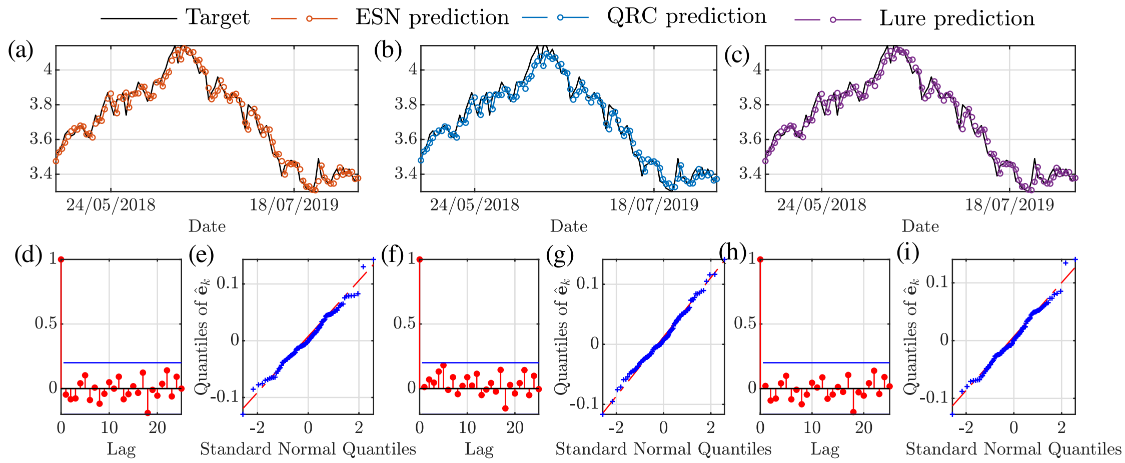


Figure 6.4: The finance time series, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e prediction on validation data. (d) The ESN residual sample ACF. (e) The ESN residual Q-Q plot. (f) The QRC residual sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACF. (i) The Lur'e residual Q-Q plot.

QRC under GAD experiences an increased RMSE. Despite this, QRC residuals under decoherence show no autocorrelation within 95% CI and pass the Lilliefors test with p-value≥ 0.35. This suggests that decoherence does not significantly impact the QRC's ability to model highly correlated data.

6.4.3 Coupled electric drive system

We test the efficacy of RCs on modeling a single-input (i.e., $u_k \in \mathbb{R}$) single-output nonlinear system consists of two electric motors driving a pulley using a flexible belt [161]. Input data is a pseudo-random binary sequence with amplitude 0.5 and only $L = 500$ (10s at a sampling time $T_s = 0.02$s) data are available, presenting a challenge for RCs.

We exploit spatial multiplexing, see Fig. 6.5. Outputs of two distinct and non-interacting RC members are combined linearly. The first member processes both $\mathbf{y}_k$ and $\mathbf{u}_k$ as described in Sec. 6.3, and the second only processes $\mathbf{u}_k$. We label each member's parameters with subscripts $1, 2$. For simplicity, we set $N = N_1 = N_2$ so that dimensions of both members (or numbers of qubits) are the same, with $N = 2$ preferred by FPE for all RCs. For the second ESN and Lur'e member, we set $C_2 = 0$ and $\sigma_{\rm m}(A_2) = 0.7$, where the latter ensures convergence. For the multiplexed QRC, we set $\epsilon_1 = 0.5$ and $\epsilon_2 = 0.9$, and set the second QRC's CPTP map $T^{(2)}(u_k)$ as

$$T^{(2)}(u_k) = g(u_k)T_1^{(2)} + (1 - g(u_k))T_2^{(2)},$$

where $T_l^{(2)}(\rho_k) = U_l^{(2)}\rho_k(U_l^{(2)})^\dagger$ for some arbitrary but fixed unitaries $U_l^{(2)}$ for $l = 1, 2$; see Appendix C.1 for the details of these unitary matrices. The second QRC has no feedback component and is uniformly convergent by construction; e.g. see Sec. 4.1. For all multiplexed RCs and QRCs, output parameters of their second member are not constrained. As before, we perform model selection (see Appendix C.2 for the selected RCs' parameters) and residual analysis for all RCs, and check whether their residuals and inputs are independent based on their sample cross-correlation via the "ggCcf" command in R [168]. We also numerically investigate the effect of decoherence on QRCs.

We set $L_w = 20$, $L_t = 400$ and $L_v = 79$. All RCs achieve comparable RMSE and average RMSE, and they pass the Lilliefors test. Residuals of ESN and QRC are uncorrelated and independent of inputs (within 95% CI); see Table 6.1 and Fig. 6.6. However, residuals of the Lur'e system exhibit autocorrelation at lag 1, and this observation persists for $N = 2, \ldots, 10$, suggesting that the Lur'e system is unsuitable for this task.

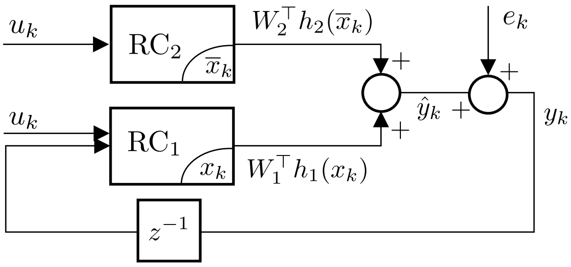


Figure 6.5: Schematic of spatial multiplexing. Here, x_k and $\overline{x}_k$ are the states of the first and second RC member, respectively. The one-step ahead prediction of the multiplexed RC is $\hat{y}_k = W_1^\top h_1(x_k) + W_2^\top h_2(\overline{x}_k)$.

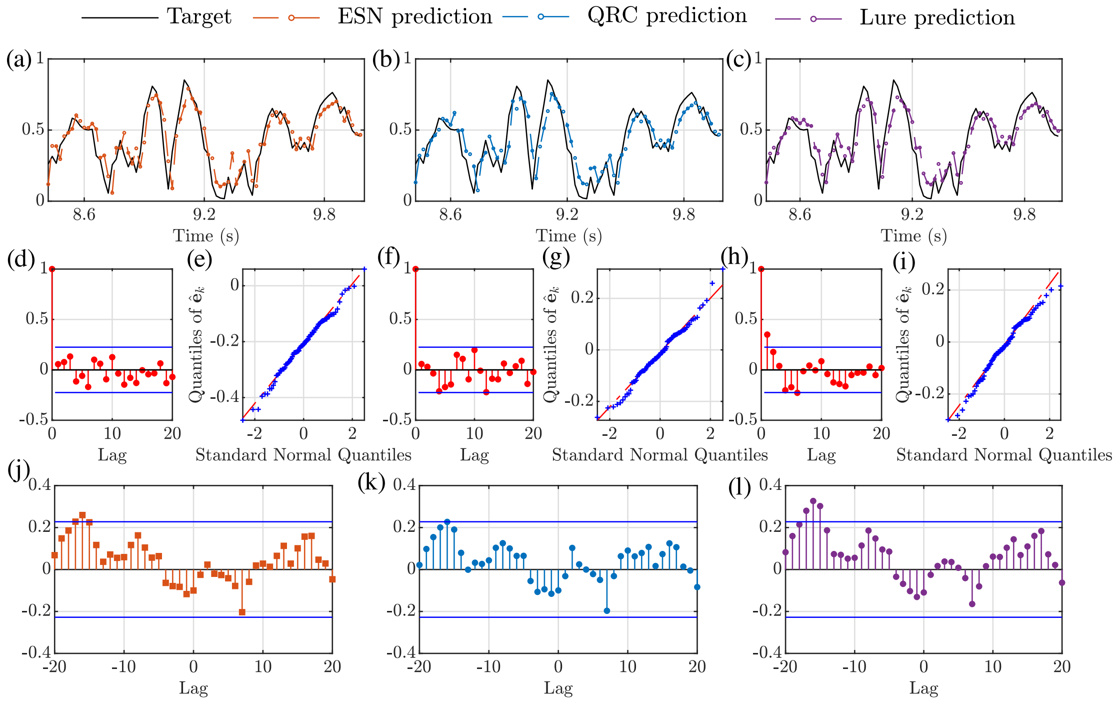


Figure 6.6: The target output, (a) the ESN prediction, (b) the QRC prediction and (c) the Lur'e prediction on validation data. (d) The ESN residual sample ACF. (e) The ESN residual Q-Q plot. (f) The QRC sample ACF. (g) The QRC residual Q-Q plot. (h) The Lur'e residual sample ACF. (i) The Lur'e residual Q-Q plot. The sample cross-correlation between inputs and residuals for (j) the ESN, (k) the QRC and (l) the Lur'e.

Under both decoherence channels, residuals of QRC show autocorrelation at lag 1. Despite this, the increase in the QRC RMSE is small, by at most 0.01 compared to the noiseless QRC.

6.5 Discussion

So far we have assumed that stochasticity only arises from the stochastic input $\mathbf{u}$ and external noise $\mathbf{e}$. For physical RCs, state measurement noise is another source of stochasticity. Suppose that we can only obtain $\hat{\mathbf{x}}_k = \mathbf{x}_k + \eta_k$, where η_k is measurement noise. In general, the framework presented here does not apply to this setting unless the RC output is a linear combination of the state elements. For a linear output $\hat{\mathbf{y}}_k = W^\top \hat{\mathbf{x}}_k$, we can write $\mathbf{y}_k = W^\top \mathbf{x}_k + W^\top \eta_k + \mathbf{e}_k$ and identify $\mathbf{e}'_k = W^\top \eta_k + \mathbf{e}_k$ as the external noise. The resulting NARX(∞) model is stationary and ergodic as long as $\mathbf{u}$ and $\mathbf{e}'$ satisfy Assumption 6.1. However, finite sampling noise for QRCs is not iid and the resulting NARX(∞) model may not be stationary.

Here, we numerically investigate the effect of finite sampling noise on the QRC's ability to model time series. We simulate the selected QRCs with $M_m = \{10^3, 10^4, 10^5\}$ number of measurement for the three tasks presented above. For the nonlinear quantum optic example, under all values of M_m, the selected QRC achieves the same RMSE $= 1.12$ and $\overline{\text{RMSE}} = 1.12$ as before. Its residuals show no autocorrelation and pass the Liliefors test with p–value=0.44, 0.50, 0.50 corresponding to each value of M_m. Furthermore, the QRC can still effectively extract the signal α_k from β_k, with normalized root mean-square errors $0.0046, 0.0048, 0.0048$ for each M_m. These observations suggest that the finite sampling noise does not impact the QRC's ability to model the nonlinear quantum optic time series. However, it shows a larger impact on the finance time series task. The QRC shows a larger error with RMSE $= 0.089, 0.067, 0.050$ and $\overline{\text{RMSE}} = 0.096, 0.071, 0.051$ for each value of M_m, compared to RMSE $= 0.049$ and $\overline{\text{RMSE}} = 0.048$ when there is no finite sampling noise. For $M_m = 10^3$, the QRC residual shows autocorrelation at lags 1–4. For larger values of M_m, the QRC residual shows no autocorrelation and passes the Liliefors test with p–value$=$ $0.34, 0.50$. For modeling the coupled electric system, we observe RMSE $= 0.19, 0.15, 0.12$ and $\overline{\text{RMSE}} = 0.21, 0.16, 0.12$ for each value of M_m. These errors are larger than when there is no finite sampling noise (RMSE $= 0.11$ and $\overline{\text{RMSE}} = 0.11$). For $M_m = 10^3$, the QRC residuals again show autocorrelation at lags 1 and 2. For larger values of M_m, the QRC is able to describe the output time series of the coupled electric system, whose residuals show no autocorrelation and pass the Liliefors test with p–value=0.50, 0.31.

6.6 Conclusion

We have introduced RCs with uniformly convergent output feedback dynamics as stationary and ergodic NARX(∞) for stochastic modeling. Our approach can harness high-dimensional nonlinear dynamical systems (with certain properties but are otherwise arbitrary) for temporal information processing, making RC an efficient and versatile scheme for control and signal processing tasks. We present and employ three example RC schemes, including the ESNs, RCs realized by Lur'e systems and QRCs, to model several synthetic and real data sets. Numerical experiments demonstrate that these RCs with a few tunable parameters are adequate in modeling these data.

We have explored the application of RC for nonlinear stochastic modeling. Many exciting problems remain open for future research, such as relaxing the current sufficient conditions for the uniform convergence property, developing suitable model selection methods for the proposed scheme and improving the modeling performance through optimizing the choice of reservoir. With this view, this work opens the potential for the emergent reservoir computing paradigm to tackle traditional challenges encountered in control and time series modeling, further bridging these scientific disciplines.

Chapter 7

Reservoir design

In previous chapters, we have explored quantum implementations of the RC scheme and different architectures of RCs for nonlinear system identification and time series modeling. So far, we have exploited the training efficiency of RCs by choosing the internal parameters of the reservoir dynamics randomly (with the uniform convergence property), and fixing them at the onset. However, it has been observed that the RC emulation performance can be significantly influenced by the internal parameters and the optimal parameters are highly dependent on the task at hand [87, 172]. These observations have motivated the reservoir design problem.

The reservoir design problem is the optimization of the reservoir internal parameters to improve the RC emulation performance. In [21, 81], it has been shown that an RC's emulation ability depends on the number of independent state variables. Generally speaking, to increase the number of independent state variables, one needs to also increase the state-space dimension of the RC. This introduces additional computational costs for software-based RCs and additional complexities for physical RCs. With this view, the reservoir design problem is important when the RC has a fixed state-space size. In this case, to improve the RC emulation performance, one can optimize the reservoir internal parameters.

The reservoir design problem has been tackled in [81, 87] using a Taylor expansion to approximate a special class of time-delay RCs to simplify the optimization problem, but this method may not fully capture the nonlinear model. Global optimization algorithms have been employed for the reservoir design problem [88, 89]. However, these algorithms often incur large computational costs. To remedy this issue, heuristic FORCE learning algorithm and its variants [90,91] have been introduced. Despite their promising empirical results, these heuristics may not provide a convergence guarantee.

Furthermore, for physical RCs, measurement noise is ubiquitous and should be taken into account in the reservoir design problem. To our best knowledge, an efficient algorithm that provides convergence guarantee even under state measurement noise is yet to be developed. This leads to the main contributions of this chapter

- We propose an efficient stochastic approximation algorithm to optimize the reservoir internal parameters.
- We show that the the proposed algorithm exhibits almost sure convergence to a Kuhn-Tucker (KT) point under the effect of state measurement noise.

The proposed algorithm is based on the simultaneous perturbation stochastic approximation (SPSA) algorithm, which only uses two noisy measurements of the cost function to approximate the gradient, regardless of the number of parameters to be optimized [92]. Our algorithm ensures convergence even when the RC output is a polynomial of arbitrary degree in the noisy state elements. Roughly speaking, the key to the convergence is to ensure that the gradient approximation is asymptotically unbiased. To achieve this, we take inspirations from the publication [173] to design regressors using four noisy state measurements, adopting the ‘double measurements’ technique used in econometrics for unbiased off-line nonlinear regression [174]. These four noisy state measurements are used to approximate the gradient to ensure that it is asymptotically unbiased. We employ our algorithm to optimize the internal parameters of ESNs under state measurement noise to emulate outputs of an LRPO and a finance time series. Numerical results suggest that the proposed algorithm can reduce the negative impact of noise on the emulation quality, particularly for higher-order RC polynomial outputs.

This chapter is organized as follows. In Sec. 7.1, we introduce the reservoir design problem and lay the assumptions on the state measurement noise. In Sec. 7.2, we introduce the SPSA algorithm. Sec. 7.3 motivates the construction of the regressors using four noisy state measurements and introduces our algorithm. We then prove its almost sure convergence to a KT point. In Sec. 7.4, we demonstrate the efficacy of our algorithm in numerical experiments. Discussions on future outlook of the reservoir design problem are presented in Sec. 7.5. Finally, concluding remarks are presented in Sec. 7.6.

7.1 The reservoir design problem

We first introduce a probabilistic framework for this chapter. Random variables are defined on a complete probability space $(\Omega, \Sigma, \mathbb{P})$ and let $\mathbb{E}[\cdot] = \mathbb{E}_{\mathbb{P}}[\cdot]$ denote the expectation over $\mathbb{P}$. As in the previous chapter, $\mathbf{z}$ is a stochastic process if $\mathbf{z} : (\Omega, \Sigma) \to ((\mathbb{R}^m)^{\mathbb{Z}}, (\mathcal{R}^m)^{\mathbb{Z}})$ is measurable, where $(\mathcal{R}^m)^{\mathbb{Z}}$ is the σ-algebra generated by cylindrical sets in $(\mathbb{R}^m)^{\mathbb{Z}}$ [84, Sec. 36]. For any random variable x, $\sigma(x)$ is the sigma algebra generated by x. Recall the definitions of stationarity (Definition 6.1) and Birkhoff-Khinchin ergodicity (Definition 6.3).

Consider an RC under a stochastic input process $\mathbf{u} : (\Omega, \Sigma) \to ((\mathbb{R}^n)^{\mathbb{Z}}, (\mathcal{R}^n)^{\mathbb{Z}})$

$$\begin{cases} \mathbf{x}_{k+1}(\gamma) = f_\gamma(\mathbf{x}_k(\gamma), \mathbf{u}_k), \\ \hat{\mathbf{y}}_k(\gamma) = h_W(\mathbf{x}_k(\gamma)), \end{cases} \tag{7.1}$$

for all $k \in \mathbb{Z}$, where $\mathbf{x}_k(\gamma) \in \mathbb{R}^N$ is the reservoir state and $\hat{\mathbf{y}}_k(\gamma) \in \mathbb{R}$ is the output. The reservoir dynamics f_γ is parametrized by $\gamma \in \mathcal{C} \subset \mathbb{R}^p$, where p is the number of parameters and $\mathcal{C}$ is a compact set. Here, we constrain the reservoir parameters in a compact set to reflect physical and numerical constraints in practical implementations.

The output function $h_W(\cdot)$ is a multivariate polynomial of degree R in the components of $\mathbf{x}_k(\gamma)$ with coefficients W. We can write $h_W(\mathbf{x}_k(\gamma)) = W^\top h(\mathbf{x}_k(\gamma))$, where $W \in \mathbb{R}^q$ and

$h(\mathbf{x}_k(\gamma)) \in \mathbb{R}^q$ whose components consist of polynomial combinations (up to degree R) of elements in $\mathbf{x}_k(\gamma)$. That is, $q = R + \binom{R}{2} + \cdots + \binom{R}{R-1} + 1$. In general, not all polynomial combinations need to be used, i.e., $q < R + \binom{R}{2} + \cdots + \binom{R}{R-1} + 1$. Since this is equivalent to setting some elements of W to be zero, for ease of notation, we assume that $h(\mathbf{x}_k(\gamma))$ contains all polynomial combinations throughout this chapter.

Given a target output process $\mathbf{y}$, the reservoir design problem is to optimize γ to minimize $F(\gamma) = \min_W \mathbb{E}\left[|\mathbf{y}_k - W^\top h(\mathbf{x}_k(\gamma))|^2\right]$. Note that $F(\gamma)$ is only a function of γ since the solution for W is the Wiener solution $W_{\text{Wiener}}(\gamma) = \mathbb{E}[h(\mathbf{x}_k(\gamma))h(\mathbf{x}_k(\gamma))^\top]^{-1}\mathbb{E}[h(\mathbf{x}_k(\gamma))\mathbf{y}_k]$ (assuming that $\mathbb{E}[h(\mathbf{x}_k(\gamma))h(\mathbf{x}_k(\gamma))^\top] \succ 0$) , also parametrized by γ.

In practice, we are often given a single realization of the target output $y = \mathbf{y}(\omega)$ and the corresponding input $u = \mathbf{u}(\omega)$ of finite length L. Suppose that $\mathbf{y}^2, \mathbf{y}\hat{\mathbf{y}}$ and $\hat{\mathbf{y}}^2$ are stationary and Birkhoff-Khinchin ergodic, then

$$\mathbb{E}\left[|\mathbf{y}_k - W^\top h(\mathbf{x}_k(\gamma))|^2\right] = \lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} |y_k - W^\top h(x_k(\gamma))|^2 \text{ a.s.},$$

where $x_k(\gamma)$ is the RC state under the (deterministic) input u. We can approximate the mean-squared error $\mathbb{E}\left[|\mathbf{y}_k - W^\top h(\mathbf{x}_k(\gamma))|^2\right]$ by $\frac{1}{L}\sum_{k=1}^{L} |y_k - W^\top h(x_k(\gamma))|^2$ and the problem is to optimize γ to minimize $F(\gamma) = \min_W \frac{1}{L}\sum_{k=1}^{L} |y_k - W^\top h(x_k(\gamma))|^2$.

To write the cost function $F(\gamma)$ in vector form, define

$$H(\gamma) = \begin{bmatrix} h(x_1(\gamma))^\top \\ \vdots \\ h(x_L(\gamma))^\top \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_L \end{bmatrix}. \tag{7.2}$$

Then $F(\gamma) = \min_W \frac{1}{L}\|Y - H(\gamma)W\|^2$. Provided that the q columns of $H(\gamma)$ are linearly independent, the solution for W is $W(\gamma) = (H(\gamma)^\top H(\gamma))^{-1}H(\gamma)^\top Y$ and hence

$$F(\gamma) = \frac{1}{L}\left\|Y - H(\gamma)(H(\gamma)^\top H(\gamma))^{-1}H(\gamma)^\top Y\right\|^2.$$

We emphasize that since we are only given a realization u of the input process $\mathbf{u}$ and y of the output process $\mathbf{y}$, we view u and y as deterministic sequences. Hence, the vector Y is also viewed as a deterministic vector throughout this chapter.

7.1.1 Noisy state measurements

Noise is ubiquitous for physical RCs. We assume that we can obtain M noisy state measurements $\hat{x}_k^j(\gamma) = x_k(\gamma) + e_k^j$, where $j = 1, \ldots, M$ and $\{e_k^j\}$ is the state noise. We consider state noise that is independent of γ, this setting is generally encountered for classical dynamical systems. However, for quantum systems, the finite sampling noise depends on the state and γ. This will be further discussed in Sec. 7.5. We impose the following assumptions on $\{e_k^j\}$ throughout this chapter:

Assumption 7.1. *The noise processes $\{e_k^j\}$ ($j = 1, \ldots, M$) are independent processes.*

Assumption 7.2. *For any $k \in \mathbb{Z}$ and $j = 1, \ldots, M$, components of e_k^j are independent with zero odd moments.*

We remark that Assumption 7.2 applies to a broad class of noise processes $\{e_k^j\}$ admitting symmetric (around zero) distributions, e.g., Gaussian distribution.

7.2 Simultaneous perturbation stochastic approximation

Under the effect of state measurement noise, we can only obtain a noisy version $\hat{F}(\gamma)$ of the cost function $F(\gamma)$. To optimize the reservoir internal parameters, an algorithm should accommodate noisy measurements as well as preserve the training efficiency of the RC framework. SPSA [92,175] is a promising candidate that uses two noisy objective function measurements to approximate the gradient regardless of the number of parameters; see Algorithm 1.

At each iteration l of the SPSA algorithm, the gradient estimate $\hat{g}_l^{\mathrm{SPSA}}(\gamma_l)$ of the true gradient $g(\gamma_l)$ (i.e., $g(\gamma_l)$ is the gradient of the noiseless cost function $F(\cdot)$ evaluated at γ_l) is

$$\hat{g}_l^{\mathrm{SPSA}}(\gamma_l) = \begin{bmatrix} \frac{\hat{F}(\gamma_l^+) - \hat{F}(\gamma_l^-)}{2c_l \Delta_{l,1}} & \cdots & \frac{\hat{F}(\gamma_l^+) - \hat{F}(\gamma_l^-)}{2c_l \Delta_{l,p}} \end{bmatrix}^\top, \tag{7.3}$$

where $\gamma_l^\pm = \gamma_l \pm c_l \Delta_l$ and $\Delta_l \in \mathbb{R}^p$ is a vector of p mutually independent zero-mean random variables $\{\Delta_{l,1}, \Delta_{l,2}, \cdots, \Delta_{l,p}\}$ that satisfy $|\Delta_{l,j}| \leq M_0$ a.s., $\mathbb{E}[|\Delta_{l,j}^{-1}|] \leq M_1$ and

$\mathbb{E}[\Delta_{l,j}^{-2}] \leq M_0'$ for $j = 1, \ldots, p$ and some $M_0, M_1, M_0' < \infty$ [92, Section III]. Often $\Delta_{l,j}$ is chosen to be distributed according to a symmetric Bernoulli distribution with support in $\{-1, 1\}$. For simplicity, we allow $\gamma_l \pm c_l\Delta_l$ to lie outside of $\mathcal{C}$ but keep $\gamma_l \in \mathcal{C}$ at all times using a projection $\Pi_{\mathcal{C}}(\cdot)$ onto $\mathcal{C}$ as in [175]. Here, $\Pi_{\mathcal{C}}(\gamma)$ is defined as the nearest point to γ on $\mathcal{C}$, where the norm is the Euclidean norm. For some physical RCs whose internal parameters are required to be strictly within the constraint set $\mathcal{C}$ at all times, we can project γ_l onto a closed set $\mathcal{C}_l \subset \mathcal{C}$ to obtain $P_l(\gamma_l)$ so that $P_l(\gamma_l) \pm c_l\Delta_l \in \mathcal{C}$. See [175, Page 890] for the explicit construction of $\mathcal{C}_l$. We remark that our results also apply to this setting when $\mathcal{C}$ is convex. We further remark that other choices of the gain sequences $\{a_l\}, \{c_l\}$ in Algorithm 1 are possible, as long as they satisfy $a_l, c_l > 0$ for all l, $\lim_{l\to\infty} a_l = 0, \lim_{l\to\infty} c_l = 0$, $\sum_{l=1}^{\infty} a_l = \infty$ and $\sum_{l=1}^{\infty}(a_l/c_l)^2 < \infty$ [92, Section III–B]. The choices presented in Algorithm 1 are found to perform well in practice [176].

Algorithm 1 Constrained SPSA algorithm [175]

```
procedure SPSA($a_{\text{SPSA}}, A_{\text{SPSA}}, \alpha_{\text{SPSA}}, c_{\text{SPSA}}, \rho_{\text{SPSA}}, K_{\text{SPSA}}$)
    for $l = 1 : K_{\text{SPSA}}$ do
        $a_l = a_{\text{SPSA}}/(l + A_{\text{SPSA}})^{\alpha_{\text{SPSA}}}$
        $c_l = c_{\text{SPSA}}/l^{\rho_{\text{SPSA}}}$
        $\gamma_l^+ = \gamma_l + c_l\Delta_l$
        $\gamma_l^- = \gamma_l - c_l\Delta_l$
        $\hat{g}_l^{\text{SPSA}}(\gamma_l) = \left[\frac{\hat{F}(\gamma_l^+)-\hat{F}(\gamma_l^-)}{2c_l\Delta_{l,1}} \quad \cdots \quad \frac{\hat{F}(\gamma_l^+)-\hat{F}(\gamma_l^-)}{2c_l\Delta_{l,p}}\right]^\top$    ▷ Gradient estimate
        $\gamma_{l+1} = \Pi_{\mathcal{C}}(\gamma_l - a_l\hat{g}_l^{\text{SPSA}}(\gamma_l))$    ▷ $\Pi_{\mathcal{C}}(\cdot)$ projects onto a compact set $\mathcal{C}$
    end for
    return $\gamma_{K_{\text{SPSA}}}$    ▷ The optimized reservoir internal parameters
end procedure
```

The attractiveness of Algorithm 1 is that under certain conditions, we have $\lim_{l\to\infty} \gamma_l = \gamma^*$ a.s., where γ^* is a KT point [175, Proposition 1], [177, Theorem 3.5]. That is, there exists $\mu_j \geq 0$ for $j = 1, \ldots, s$ such that $g(\gamma^*) + \sum_{j=1}^{s} \mu_j \frac{dq_j(\gamma*)}{d\gamma} = 0$ with $q_j(\gamma^*) = 0$. Here, $g(\cdot)$ is the gradient of the noiseless cost function and $q_j(\cdot)$ are the inequality constraints determined by the constraint set $\mathcal{C}$. That is, we can write $\mathcal{C} = \{\gamma : q_j(\gamma) \leq 0; j = 1, \ldots, s\}$.

7.3 Algorithm and its analysis

The crucial condition required for Algorithm 1 to converge a.s. to a KT point is that the noisy cost function satisfies

$$\hat{F}(\gamma_l^{\pm}) = F(\gamma_l^{\pm}) + \epsilon_l^{\pm},$$

where

$$\mathbb{E}[\epsilon_l^{+} - \epsilon_l^{-} | \gamma_l, \Delta_l] = 0 \text{ a.s.} \tag{7.4}$$

for all iterations l. This condition ensures that the gradient estimate $\hat{g}_l^{\mathrm{SPSA}}(\gamma_l)$ is asymptotically unbiased in the sense that $\lim_{l\to\infty} \mathbb{E}[\hat{g}_l^{\mathrm{SPSA}}(\gamma_l) - g(\gamma_l)|\gamma_l] = 0$ a.s.; see [92, Lemma 1]. However, under noisy state measurements, our noisy cost function does not necessarily satisfy (7.4).

7.3.1 Motivation

We provide some insight into why the noisy cost function does not necessarily satisfy (7.4). Define

$$\hat{H}_{l,\pm} = \begin{bmatrix} h(\hat{x}_1(\gamma_l^{\pm}))^\top \\ \vdots \\ h(\hat{x}_L(\gamma_l^{\pm}))^\top \end{bmatrix}, H_{l,\pm} = \begin{bmatrix} h(x_1(\gamma_l^{\pm}))^\top \\ \vdots \\ h(x_L(\gamma_l^{\pm}))^\top \end{bmatrix} \tag{7.5}$$

where $\hat{x}_k(\gamma) = x_k(\gamma) + e_k$ with measurement noise e_k for $k = 1, \ldots, L$. Assuming that $\hat{H}_{l,\pm}^\top \hat{H}_{l,\pm} \succ 0$ and $H_{l,\pm}^\top H_{l,\pm} \succ 0$, we have

$$F(\gamma_l^{\pm}) = \frac{1}{L}\|Y - H_{l,\pm}(H_{l,\pm}^\top H_{l,\pm})^{-1} H_{l,\pm}^\top Y\|^2 = \frac{1}{L}\left(Y^\top Y - Y^\top H_{l,\pm}(H_{l,\pm}^\top H_{l,\pm})^{-1} H_{l,\pm}^\top Y\right),$$

$$\hat{F}(\gamma_l^{\pm}) = \frac{1}{L}\|Y - \hat{H}_{l,\pm}(\hat{H}_{l,\pm}^\top \hat{H}_{l,\pm})^{-1} \hat{H}_{l,\pm}^\top Y\|^2 = \frac{1}{L}\left(Y^\top Y - Y^\top \hat{H}_{l,\pm}(\hat{H}_{l,\pm}^\top \hat{H}_{l,\pm})^{-1} \hat{H}_{l,\pm}^\top Y\right)$$

and

$$\epsilon_l^{\pm} = \frac{1}{L}\left[Y^\top H_{l,\pm}(H_{l,\pm}^\top H_{l,\pm})^{-1} H_{l,\pm}^\top Y - Y^\top \hat{H}_{l,\pm}(\hat{H}_{l,\pm}^\top \hat{H}_{l,\pm})^{-1} \hat{H}_{l,\pm}^\top Y\right].$$

To check if (7.4) holds, we may check if $\mathbb{E}[\epsilon_l^{\pm}|\gamma_l, \Delta_l] = 0$ a.s.. However, checking this is often difficult and cumbersome since $\epsilon_l^{\pm}$ is a nonlinear function of the noisy regressors $\hat{H}_{l,\pm}$ that involves matrix inversion. To circumvent this problem, we propose to optimize the reservoir internal parameters γ and the RC output parameters W separately in a round robin fashion. Explicitly, we first optimize W using some noisy measurements of the state, with W^* being the solution. We then optimize γ by fixing W^* and using another set of noisy state measurements. That is, our noiseless cost function becomes

$$F(\gamma) = \frac{1}{L} \|Y - HW^*\|^2$$

and we employ the noisy cost function

$$\hat{F}(\gamma) = \frac{1}{L} \left\|Y - \hat{H}(\gamma)W^*\right\|^2$$

for the optimization of γ using a stochastic approximation method.

The above modification still does not guarantee (7.4) in general. To see this, we write

$$\begin{aligned} F(\gamma_l^{\pm}) &= \frac{1}{L} \|Y - H_{l,\pm}W^*\|^2 = \frac{1}{L} \left(Y^\top Y - 2Y^\top H_{l,\pm}W^* + (W^*)^\top H_{l,\pm}^\top H_{l,\pm}W^*\right) \\ \hat{F}(\gamma_l^{\pm}) &= \frac{1}{L} \left\|Y - \hat{H}_{l,\pm}W^*\right\|^2 = \frac{1}{L} \left(Y^\top Y - 2Y^\top \hat{H}_{l,\pm}W^* + (W^*)^\top \hat{H}_{l,\pm}^\top \hat{H}_{l,\pm}W^*\right) \end{aligned}$$

and observe that

$$\epsilon_l^{\pm} = \frac{1}{L} \left[2Y^\top H_{l,\pm}W^* - 2Y^\top \hat{H}_{l,\pm}W^* + (W^*)^\top \hat{H}_{l,\pm}^\top \hat{H}_{l,\pm}W^* - (W^*)^\top H_{l,\pm}^\top H_{l,\pm}W^*\right],$$

where $\hat{H}_{l,\pm}$ and $H_{l,\pm}$ are defined in (7.5).

To check if (7.4) holds, we may check if $\mathbb{E}[\epsilon_k^{\pm}|\gamma_k, \Delta_k] = 0$ a.s.. However, this is generally nonzero since $\epsilon_l^{\pm}$ contains even degree terms of the state measurement noise. For instance, even if $h(\cdot)$ is linear, that is, consider $H_{l,\pm} = X_{l,\pm}$ and $\hat{H}_{l,\pm} = \hat{X}_{l,\pm}$, where

$$X_{l,\pm} = \begin{bmatrix} x_1(\gamma_l^{\pm})^\top \\ \vdots \\ x_L(\gamma_l^{\pm})^\top \end{bmatrix}, \hat{X}_{l,\pm} = \begin{bmatrix} \hat{x}_1(\gamma_l^{\pm})^\top \\ \vdots \\ \hat{x}_L(\gamma_l^{\pm})^\top \end{bmatrix}, E_{l,\pm} = \begin{bmatrix} (e_1^{\pm})^\top \\ \vdots \\ (e_L^{\pm})^\top \end{bmatrix}$$

with $\hat{X}_{l,\pm} = X_{l,\pm} + E_{l,\pm}$ and $e_k^{\pm}$ correspond to the state noise when measuring the states $\hat{x}_k(\gamma_l^{\pm})$. Using the tower property of conditional expectations and under Assumption 7.2, we have

$$\begin{aligned}
\mathbb{E}[\epsilon_l^{\pm}|\gamma_l, \Delta_l] &= -\frac{1}{L}\mathbb{E}\left[2Y^\top \hat{X}_{l,\pm}W^* - 2Y^\top X_{l,\pm}W^*|\gamma_l, \Delta_l\right] \\
&\quad + \frac{1}{L}\mathbb{E}\left[(W^*)^\top \hat{X}_{l,\pm}^\top \hat{X}_{l,\pm}W^* - (W^*)^\top X_{l,\pm}^\top X_{l,\pm}W^*|\gamma_l, \Delta_l\right] \\
&= -\frac{1}{L}2Y^\top \mathbb{E}\left\{\mathbb{E}\left[E_{l,\pm}|W^*, \gamma_l, \Delta_l\right] W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[E_{l,\pm}^\top|W^*, \gamma_l, \Delta_l\right] X_{l,\pm}W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top X_{l,\pm}^\top \mathbb{E}\left[E_{l,\pm}|W^*, \gamma_l, \Delta_l\right] W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[E_{l,\pm}^\top E_{l,\pm}|W^*, \gamma_l, \Delta_l\right] W^*|\gamma_l, \Delta_l\right\} \\
&= -\frac{1}{L}2Y^\top \mathbb{E}\left\{\mathbb{E}\left[E_{l,\pm}\right] W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[E_{l,\pm}^\top\right] X_{l,\pm}W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top X_{l,\pm}^\top \mathbb{E}\left[E_{l,\pm}\right] W^*|\gamma_l, \Delta_l\right\} \\
&\quad + \frac{1}{L}\mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[E_{l,\pm}^\top E_{l,\pm}\right] W^*|\gamma_l, \Delta_l\right\} \\
&= \frac{1}{L}\mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[E_{l,\pm}^\top E_{l,\pm}\right] W^*|\gamma_l, \Delta_l\right\} \text{ a.s.}
\end{aligned}$$

which is nonzero in general since the even moments of the state noise is nonzero. Here, the second equality follows since $X_{l,\pm}$ is measurable with respect to $\sigma(W^*, \gamma_l, \Delta_l)$. The third equality follows since $E_{l,\pm}$ is independent of W^*, γ_l, Δ_l and the last equality follows from Assumption 7.2 that $\mathbb{E}[E_{l,\pm}] = 0$.

To address this problem, we propose to modify the noisy cost function by

$$\overline{F}(\gamma) = \frac{1}{L}\left(Y - \hat{V}(\gamma)W^*\right)^\top \left(Y - \tilde{V}(\gamma)W^*\right), \tag{7.6}$$

with $\hat{V}(\gamma)$ and $\tilde{V}(\gamma)$ essentially replacing the role of $H(\gamma)$ defined in (7.2) and

$$\begin{aligned}
\mathbb{E}[\epsilon_l^{\pm}|\gamma_l, \Delta_l] =& \frac{1}{L}\mathbb{E}\left[2Y^\top H_{l,\pm}W^* - Y^\top \tilde{V}_{l,\pm}W^* - Y^\top \hat{V}_{l,\pm}W^*|\gamma_l, \Delta_l\right] \\
&- \frac{1}{L}\mathbb{E}\left[(W^*)^\top H_{l,\pm}^\top H_{l,\pm}W^* - (W^*)^\top \hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}W^*|\gamma_l, \Delta_l\right].
\end{aligned} \tag{7.7}$$

We want to construct regressors $\hat{V}(\gamma)$ and $\tilde{V}(\gamma)$ such that (7.7) is zero a.s.. A possible choice is to obtain $2R$ noisy state measurements $\hat{x}_k^j(\gamma) = x_k(\gamma) + e_k^j$ ($j = 1, \ldots, 2R$) for each $k = 1, \ldots L$, where R is the degree of the output function $h_W(\cdot)$. Let $\hat{x}_{k,m}^j(\gamma)$ denote the m–th component of $\hat{x}^j_k(\gamma)$, we define

$$\begin{aligned}\hat{v}_k(\gamma) =& [\hat{x}^1_{k,1}(\gamma) \quad \cdots \quad \hat{x}^1_{k,n}(\gamma) \quad \hat{x}^1_{k,1}(\gamma)\hat{x}^2_{k,1}(\gamma) \quad \hat{x}^1_{k,1}(\gamma)\hat{x}^1_{k,2}(\gamma) \quad \cdots \\ & \hat{x}^1_{k,1}(\gamma)\hat{x}^2_{k,1}(\gamma)\hat{x}^3_{k,1}(\gamma) \quad \hat{x}^1_{k,1}(\gamma)\hat{x}^2_{k,1}(\gamma)\hat{x}^1_{k,2}(\gamma) \quad \cdots \quad (\hat{x}^1_{k,n}(\gamma)\hat{x}^2_{k,n}(\gamma)\cdots\hat{x}^R_{k,n}(\gamma))]^\top, \\ \tilde{v}_k(\gamma) =& [\hat{x}^{R+1}_{k,1}(\gamma) \quad \cdots \quad \hat{x}^{R+1}_{k,n}(\gamma) \quad \hat{x}^{R+1}_{k,1}(\gamma)\hat{x}^{R+2}_{k,1}(\gamma) \quad \hat{x}^{R+1}_{k,1}(\gamma)\hat{x}^{R+1}_{k,2}(\gamma) \quad \cdots \\ & \hat{x}^{R+1}_{k,1}(\gamma)\hat{x}^{R+2}_{k,1}(\gamma)\hat{x}^{R+3}_{k,1}(\gamma) \quad \hat{x}^{R+1}_{k,1}(\gamma)\hat{x}^{R+2}_{k,1}(\gamma)\hat{x}^{R+1}_{k,2}(\gamma) \quad \cdots \\ & \cdots \quad (\hat{x}^{R+1}_{k,n}(\gamma)\hat{x}^{R+2}_{k,n}(\gamma)\cdots\hat{x}^{2R}_{k,n}(\gamma))]^\top.\end{aligned}$$

The regressors $\hat{v}_k(\gamma)$ and $\tilde{v}_k(\gamma)$ now essentially replace the role of $h(x_k(\gamma))$. Components in $\hat{v}_k(\gamma)$ and $\tilde{v}_k(\gamma)$ do not contain products of $\hat{x}^j_{k,l}(\gamma)$ for any fixed l, j of more than degree one, e.g., they do not contain $(\hat{x}^j_{k,1}(\gamma))^2$ or $(\hat{x}^j_{k,1}(\gamma))^2\hat{x}^j_{k,2}(\gamma)$. Defining the regressors

$$\hat{V}(\gamma) = \begin{bmatrix} \hat{v}_1(\gamma)^\top \\ \vdots \\ \hat{v}_L(\gamma)^\top \end{bmatrix}, \hat{V}_{l,\pm} = \begin{bmatrix} \hat{v}_1(\gamma^\pm_l)^\top \\ \vdots \\ \hat{v}_L(\gamma^\pm_l)^\top \end{bmatrix}, \tilde{V}(\gamma) = \begin{bmatrix} \tilde{v}_1(\gamma)^\top \\ \vdots \\ \tilde{v}_L(\gamma)^\top \end{bmatrix}, \tilde{V}_{l,\pm} = \begin{bmatrix} \tilde{v}_1(\gamma^\pm_l)^\top \\ \vdots \\ \tilde{v}_L(\gamma^\pm_l)^\top \end{bmatrix}, \tag{7.8}$$

then under Assumptions 7.1 and 7.2, (7.7) is zero a.s. since it no longer contains even moments of elements in e^j_k for any $j = 1, \ldots, 2R$. The caveat is that the number of measurements increases as R increases.

7.3.2 Constructing regressors from four noisy state measurements

We show that only four measurements are sufficient to ensure that (7.7) is zero. From the four noisy state measurements, we have

$$\begin{aligned}\frac{\hat{x}^1_k(\gamma) + \hat{x}^2_k(\gamma)}{2} &= x_k(\gamma) + \frac{e^1_k + e^2_k}{2}, & \frac{\hat{x}^1_k(\gamma) - \hat{x}^2_k(\gamma)}{2} &= \frac{e^1_k - e^2_k}{2}, \\ \frac{\hat{x}^3_k(\gamma) + \hat{x}^4_k(\gamma)}{2} &= x_k(\gamma) + \frac{e^3_k + e^4_k}{2}, & \frac{\hat{x}^3_k(\gamma) - \hat{x}^4_k(\gamma)}{2} &= \frac{e^3_k - e^4_k}{2}.\end{aligned}$$

These are the building blocks for $\hat{v}_k(\gamma)$ and $\tilde{v}_k(\gamma)$. Recall that $\hat{v}_k(\gamma), \tilde{v}_k(\gamma)$ replace the role of $h(x_k(\gamma))$, whose components are polynomial combinations (up to degree R) of elements in $x_k(\gamma)$. For each component $m = 1, \ldots, N$ of the state $x_k(\gamma)$, we recursively construct variables $\hat{t}^{(1)}_{k,m}(\gamma), \hat{t}^{(2)}_{k,m}(\gamma), \ldots, \hat{t}^{(R)}_{k,m}(\gamma)$ to replace $x_{k,m}(\gamma), x_{k,m}(\gamma)^2, \ldots, x_{k,m}(\gamma)^R$ appearing in $h(x_k(\gamma))$. For instance, for the first- and second-order terms:

- Let $\hat{t}^{(1)}_{k,m}(\gamma) = x_{k,m}(\gamma) + \frac{e^1_{k,m}+e^2_{k,m}}{2}$. This plays the role of $x_{k,m}(\gamma)$.
- Let $\hat{t}^{(2)}_{k,m}(\gamma) = (\hat{t}^{(1)}_{k,m}(\gamma))^2 - \left(\frac{e^1_{k,m}-e^2_{k,m}}{2}\right)^2$. This plays the role of $x_{k,m}(\gamma)^2$.

To see this, note that $(\hat{t}^{(1)}_{k,m}(\gamma))^2 = x_{k,m}(\gamma)^2 + x_{k,m}(\gamma)\left(e^1_{k,m} + e^2_{k,m}\right) + \frac{1}{4}\left(e^1_{k,m} + e^2_{k,m}\right)^2$. To eliminate the even degree terms of $e^1_{k,m}, e^2_{k,m}$, we subtract $\frac{1}{4}(e^1_{k,m} - e^2_{k,m})^2$ from the above, leading to $\hat{t}^{(2)}_{k,m}(\gamma) = x_{k,m}(\gamma)^2 + x_{k,m}(\gamma)\left(e^1_{k,m} + e^2_{k,m}\right) + e^1_{k,m}e^2_{k,m}$. More generally, for the r–th degree term, we have

$$\hat{t}^{(r)}_{k,m}(\gamma) = (\hat{t}^{(1)}_{k,m}(\gamma))^r - \sum_{\substack{d=2k,\\ 1\le k\le \lfloor r/2\rfloor}} \binom{r}{d} \hat{t}^{(r-d)}_{k,m}(\gamma)\left(\frac{e^1_{k,m} - e^2_{k,m}}{2}\right)^d. \tag{7.9}$$

Replacing $e^1_{k,m}, e^2_{k,m}$ by $e^3_{k,m}, e^4_{k,m}$ in (7.9), we can define $\tilde{t}^{(r)}_{k,m}(\gamma)$ analogously. We then define $\hat{v}_k(\gamma), \tilde{v}_k(\gamma)$ using these new variables:

$$\begin{aligned} \hat{v}_k(\gamma) = [&\hat{t}^{(1)}_{k,1}(\gamma) \quad \cdots \quad \hat{t}^{(1)}_{k,N}(\gamma) \quad \hat{t}^{(2)}_{k,1}(\gamma) \quad \hat{t}^{(1)}_{k,1}(\gamma)\hat{t}^{(1)}_{k,2}(\gamma) \quad \cdots \quad \hat{t}^{(2)}_{k,N}(\gamma) \\ &\hat{t}^{(3)}_{k,1}(\gamma) \quad \hat{t}^{(2)}_{k,1}(\gamma)\hat{t}^{(1)}_{k,2}(\gamma) \quad \cdots \quad \hat{t}^{(3)}_{k,N}(\gamma) \quad \cdots \quad \hat{t}^{(R)}_{k,N}(\gamma)]^\top, \\ \tilde{v}_k(\gamma) = [&\tilde{t}^{(1)}_{k,1}(\gamma) \quad \cdots \quad \tilde{t}^{(1)}_{k,N}(\gamma) \quad \tilde{t}^{(2)}_{k,1}(\gamma) \quad \tilde{t}^{(1)}_{k,1}(\gamma)\tilde{t}^{(1)}_{k,2}(\gamma) \quad \cdots \quad \tilde{t}^{(2)}_{k,N}(\gamma) \\ &\tilde{t}^{(3)}_{k,1}(\gamma) \quad \tilde{t}^{(2)}_{k,1}(\gamma)\tilde{t}^{(1)}_{k,2}(\gamma) \quad \cdots \quad \tilde{t}^{(3)}_{k,N}(\gamma) \quad \cdots \quad \tilde{t}^{(R)}_{k,N}(\gamma)]^\top. \end{aligned} \tag{7.10}$$

Remark 7.1. *Observe that we can write $\hat{v}_k(\gamma) = h(x_k(\gamma)) + f(x_k(\gamma), e^1_k, e^2_k)$ and $\tilde{v}_k(\gamma) = h(x_k(\gamma)) + f(x_k(\gamma), e^3_k, e^4_k)$, where $f(x_k(\gamma), e^1_k, e^2_k)$ and $f(x_k(\gamma), e^3_k, e^4_k)$ are random vectors with q components, and their m–th components have the form*

$$\begin{aligned} f(x_k(\gamma), e^1_k, e^2_k)_m &= \sum_{j=1}^{M_m} \alpha_{m,j}(x_k(\gamma))\beta_{m,j}(e^1_k, e^2_k), \\ f(x_k(\gamma), e^3_k, e^4_k)_m &= \sum_{j=1}^{M_m} \alpha_{m,j}(x_k(\gamma))\beta_{m,j}(e^3_k, e^4_k). \end{aligned} \tag{7.11}$$

Here, $\alpha_{m,j}(x_k(\gamma))$ is a polynomial of elements in $x_k(\gamma)$ and $\beta_{m,j}(e^1_k, e^2_k)$, $\beta_{m,j}(e^3_k, e^4_k)$ are polynomial of odd degrees of elements in e^1_k, e^2_k and e^3_k, e^4_k, respectively.

Remark 7.1 will be a key observation in our analysis.

Equipped with regressors $\hat{V}(\gamma)$ and $\tilde{V}(\gamma)$ defined by (7.8) and (7.10), we propose a modified constrained SPSA algorithm in Algorithm 2 and prove its almost sure convergence to a KT point in the next subsection. Incorporating the round robin method to optimize γ and W iteratively, we arrive at the overall algorithm shown in Algorithm 3.

Algorithm 2 Modified constrained SPSA algorithm

```
procedure MSPSA($W^*, a_{\text{MSPSA}}, A_{\text{MSPSA}}, \alpha_{\text{MSPSA}}, c_{\text{MSPSA}}, \rho_{\text{MSPSA}}, K_{\text{MSPSA}}$)
    for $l = 1 : K_{\text{MSPSA}}$ do
        $a_l = a_{\text{MSPSA}}/(l + A_{\text{MSPSA}})^{\alpha_{\text{MSPSA}}}$
        $c_l = c_{\text{MSPSA}}/l^{\rho_{\text{MSPSA}}}$
        $\gamma_l^+ = \gamma_l + c_l\Delta_l$
        $\gamma_l^- = \gamma_l - c_l\Delta_l$
        $\overline{F}(\gamma_l^{\pm}) = \frac{1}{L}(Y - \hat{V}_{l,\pm}W^*)^\top(Y - \tilde{V}_{l,\pm}W^*)$    ▷ $\hat{V}_{l,\pm}, \tilde{V}_{l,\pm}$ are defined in (7.8), (7.10)
        $\hat{g}_l^{\text{MSPSA}}(\gamma_l) = \begin{bmatrix} \frac{\overline{F}(\gamma_l^+) - \overline{F}(\gamma_l^-)}{2c_l\Delta_{l,1}} & \cdots & \frac{\overline{F}(\gamma_l^+) - \overline{F}(\gamma_l^-)}{2c_l\Delta_{l,p}} \end{bmatrix}^\top$    ▷ Gradient estimate
        $\gamma_{l+1} = \Pi_{\mathcal{C}}\left(\gamma_l - a_l\hat{g}_l^{\text{MSPSA}}(\gamma_l)\right)$    ▷ $\Pi_{\mathcal{C}}(\cdot)$ projects onto a compact set $\mathcal{C}$
    end for
    return $\gamma_{K_{\text{MSPSA}}}$    ▷ The optimized reservoir internal parameters
end procedure
```

Algorithm 3 Reservoir design algorithm

```
procedure RCDESIGN($a_{\text{MSPSA}}, A_{\text{MSPSA}}, \alpha_{\text{MSPSA}}, c_{\text{MSPSA}}, \rho_{\text{MSPSA}}, K_{\text{MSPSA}}, K$)
    Randomize $\gamma_0 \in \mathcal{C}$
    Construct $\hat{V}(\gamma_0)$    ▷ $\hat{V}(\gamma_0)$ is defined in (7.8), (7.10)
    $W_0^* = \arg\min_W G(\hat{V}(\gamma_0), Y, W)$
    for $l = 1 : K$ do
        $\gamma_l$=MSPSA($W_{l-1}^*, a_{\text{MSPSA}}, A_{\text{MSPSA}}, \alpha_{\text{MSPSA}}, c_{\text{MSPSA}}, \rho_{\text{MSPSA}}, K_{\text{MSPSA}}$)
        Construct $\hat{V}(\gamma_l)$    ▷ $\hat{V}(\gamma_l)$ is defined in (7.8), (7.10)
        $W_l^* = \arg\min_W G(\hat{V}(\gamma_l), Y, W)$
    end for
    return $\gamma_K, W_K^*$    ▷ The optimized internal and output parameters
end procedure
```

7.3.3 Almost sure convergence

To show that Algorithm 2 converges to a KT point a.s., we first prove the following lemma.

Lemma 7.1. *Using regressors* $\hat{V}(\gamma)$ *and* $\tilde{V}(\gamma)$ *defined in* (7.8) *and* (7.10)*, under Assumptions 7.1 and 7.2, we have* $\mathbb{E}[\epsilon_l^+ - \epsilon_l^- | \gamma_l, \Delta_l] = 0$ *a.s. for all* l*, where* $\epsilon_l^\pm = \overline{F}(\gamma_l^\pm) - F(\gamma_l^\pm)$ *and* $\overline{F}(\gamma_l^\pm)$ *is defined by* (7.6)*.*

Proof. As alluded to before, we prove a stronger property in which $\mathbb{E}[\epsilon_l^\pm | \gamma_l, \Delta_l] = 0$ a.s. for all l, where

$$\begin{aligned}\mathbb{E}[\epsilon_l^\pm|\gamma_l,\Delta_l] =& \frac{1}{L}\mathbb{E}\left[2Y^\top H_{l,\pm}W^* - Y^\top \tilde{V}_{l,\pm}W^* - Y^\top \hat{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right]\\ &-\frac{1}{L}\mathbb{E}\left[(W^*)^\top H_{l,\pm}^\top H_{l,\pm}W^* - (W^*)^\top \hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right],\end{aligned} \tag{7.12}$$

and $H_{l,\pm}$ is defined in (7.5). We examine the first summand in (7.12). Using the tower property of conditional expectations, we have

$$\mathbb{E}\left[Y^\top \hat{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right] = Y^\top \mathbb{E}\left\{\mathbb{E}\left[\hat{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right]W^*|\gamma_l,\Delta_l\right\} \text{ a.s..}$$

Recall from Remark 7.1 that $\hat{v}_k = h(x_k) + f(x_k, e_k^1, e_k^2)$. Let $\hat{v}_{k,m}(\gamma_l^\pm)$ be the m–th component of $\hat{v}_k(\gamma_l^\pm)$. The k–th row and m–th column of the matrix conditional expectation $\mathbb{E}\left[\hat{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right]$ is

$$\begin{aligned}\mathbb{E}\left[\hat{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right]_{k,m} &= \mathbb{E}\left[\hat{v}_{k,m}(\gamma_l^\pm)|W^*,\gamma_l,\Delta_l\right]\\ &= \mathbb{E}\left[h(x_k(\gamma_l^\pm))_m + f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_m|W^*,\gamma_l,\Delta_l\right]\\ &= h(x_k(\gamma_l^\pm))_m + \sum_{j=1}^{M_m}\alpha_{m,j}(x_k(\gamma_l^\pm))\mathbb{E}\left[\beta_{m,j}(e_k^1,e_k^2)|W^*,\gamma_l,\Delta_l\right]\\ &= h(x_k(\gamma_l^\pm))_m + \sum_{j=1}^{M_m}\alpha_{m,j}(x_k(\gamma_l^\pm))\mathbb{E}\left[\beta_{m,j}(e_k^1,e_k^2)\right]\\ &= h(x_k(\gamma_l^\pm))_m \text{ a.s.,}\end{aligned}$$

where the third equality follows since $h(x_k(\gamma_l^\pm))_m$ and $\alpha_{m,j}(x_k(\gamma_l^\pm))$ are measurable with respect to $\sigma(W^*,\gamma_l,\Delta_l)$, the second last equality follows since e_k^1, e_k^2 are independent of $\sigma(W^*,\gamma_l,\Delta_l)$ and the last equality follows from Assumptions 7.1 and 7.2. Therefore,

$\mathbb{E}\left[\hat{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right] = H_{l,\pm}$ and $\mathbb{E}\left[Y^\top \hat{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right] = \mathbb{E}[Y^\top H_{l,\pm}W^*|\gamma_l,\Delta_l]$ a.s.. A similar argument shows that $\mathbb{E}\left[Y^\top \tilde{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right] = \mathbb{E}[Y^\top H_{l,\pm}W^*|\gamma_l,\Delta_l]$ a.s.. As a result, the first summand in (7.12) is zero a.s..

We now examine the second summand in (7.12). Using the tower property of conditional expectations, we have

$$\mathbb{E}\left[(W^*)^\top \hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right] = \mathbb{E}\left\{(W^*)^\top \mathbb{E}\left[\hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right] W^*|\gamma_l,\Delta_l\right\} \text{ a.s..}$$

The j–row and m–column of the matrix conditional expectation $\mathbb{E}\left[\hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right]$ is

$$\begin{aligned}
&\mathbb{E}\left[\hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right]_{j,m} \\
&= \sum_{k=1}^{L} \mathbb{E}\left[(h(x_k(\gamma_l^\pm))_j + f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_j)(h(x_k(\gamma_l^\pm))_m + f(x_k(\gamma_l^\pm),e_k^3,e_k^4)_m)|W^*,\gamma_l,\Delta_l\right] \\
&= \sum_{k=1}^{L} h(x_k(\gamma_l^\pm))_j h(x_k(\gamma_l^\pm))_m \\
&\quad + \sum_{k=1}^{L} h(x_k(\gamma_l^\pm))_j \mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^3,e_k^4)_m|W^*,\gamma_l,\Delta_l\right] \\
&\quad + \sum_{k=1}^{L} h(x_k(\gamma_l^\pm))_m \mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_j|W^*,\gamma_l,\Delta_l\right] \\
&\quad + \sum_{k=1}^{L} \mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_j f(x_k(\gamma_l^\pm),e_k^3,e_k^4)_m|W^*,\gamma_l,\Delta_l\right] \text{ a.s.,}
\end{aligned}$$

where the last equality follows since $h(x_k(\gamma_l^\pm))_j$ and $h(x_k(\gamma_l^\pm))_m$ are measurable with respect to $\sigma(W^*,\gamma_l,\Delta_l)$. Expressing $f(x_k(\gamma_l^\pm),e_k^1,e_k^2)$ and $f(x_k(\gamma_l^\pm),e_k^3,e_k^4)$ as (7.11) in Remark 7.1, using Assumptions 7.1 and 7.2, we have $\mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^3,e_k^4)_m|W^*,\gamma_l,\Delta_l\right] = \mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_j|W^*,\gamma_l,\Delta_l\right] = \mathbb{E}\left[f(x_k(\gamma_l^\pm),e_k^1,e_k^2)_j f(x_k(\gamma_l^\pm),e_k^3,e_k^4)_m|W^*,\gamma_l,\Delta_l\right] = 0$ a.s.. Hence, $\mathbb{E}\left[\hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}|W^*,\gamma_l,\Delta_l\right] = \mathbb{E}\left[H_{l,\pm}^\top H_{l,\pm}|W^*,\gamma_l,\Delta_l\right]$ a.s. and

$$\mathbb{E}\left[(W^*)^\top \hat{V}_{l,\pm}^\top \tilde{V}_{l,\pm}W^*|\gamma_l,\Delta_l\right] = \mathbb{E}\left[(W^*)^\top H_{l,\pm}^\top H_{l,\pm}W^*|\gamma_l,\Delta_l\right] \text{ a.s..}$$

As a result, the second summand in (7.12) is zero a.s.. □

We now establish the almost sure convergence of Algorithm 2 by applying a simplified version of [175, Proposition 1] stated below.

Theorem 7.1. *[175, Proposition 1] Suppose that the following hold:*

1. *$F(\cdot)$ is three times continuously differentiable.*
2. *$a_l, c_l > 0$, $\lim_{l\to\infty} a_l = 0, \lim_{l\to\infty} c_l = 0, \sum_{l=1}^{\infty} a_l = \infty$ and $\sum_{l=1}^{\infty}(a_l/c_l)^2 < \infty$.*
3. *The constraints $q_j(\cdot)$ where $\mathcal{C} = \{\gamma : q_j(\gamma) \leq 0; j = 1, \dots, s\}$ are continuously differentiable and at each $\gamma \in \partial\mathcal{C}$, the gradients of the active constraints are linearly independent.*
4. *Define $\overline{\pi}(g(\gamma)) = \lim_{\delta\to 0,\delta>0}(\Pi_{\mathcal{C}}(\gamma + \delta g(\gamma)) - \gamma)/\delta$ [177, Page 191]. Let γ^* be an asymptotically stable point (which is a KT point) of the differential equation $d\gamma(t)/dt = -\overline{\pi}(g(\gamma))$, where $g(\cdot)$ is the gradient of the noiseless cost function $F(\cdot)$. Let $D(\gamma^*)$ be the domain of attraction for γ^*. Suppose that there exists a compact set $A \subset D(\gamma^*)$ such that $\gamma_l \in A$ infinitely often.*
5. *For each l, $\{\Delta_{l,j}\}$ are independently identically distributed for $j = 1, \dots, p$ and symmetrically distributed about 0 with $|\Delta_{l,j}| \leq M_0 < \infty$ a.s., $\mathbb{E}[|\Delta_{l,j}^{-1}|] \leq M_1 < \infty$.*
6. *For all l, $\mathbb{E}[\Delta_{l,j}^{-2}] \leq M_0' < \infty$, $\mathbb{E}[(\epsilon_l^{\pm})^2] \leq M_1' < \infty$ and $\mathbb{E}[F(\gamma_l^{\pm})^2] \leq M_2' < \infty$.*

Then under Algorithm 2, $\lim_{l\to\infty} \gamma_l = \gamma^*$ a.s..

We remark that Theorem 7.1 is a simplified version of [175, Proposition 1]. In [175, Proposition 1], the author includes an additional projection $P_l(\gamma_l)$ onto a closed set $\mathcal{C}_l \subset \mathcal{C}$ so that $P_l(\gamma_l) \pm c_l\Delta_l \in \mathcal{C}$. In Theorem 7.1, we allow $\gamma_l \pm c_l\Delta_l$ to lie outside of $\mathcal{C}$ but keep $\gamma_l \in \mathcal{C}$ for all l. The proof for both Theorem 7.1 and [175, Proposition 1] are the same.

Proof. The proof is identical to the proof of [175, Proposition 1], which is based on [92, Proposition 1] and [177, Theorem 5.3.1]. We sketch the proof here for completeness. The main idea is to write

$$\gamma_{l+1} = \Pi_{\mathcal{C}}(\gamma_l - a_l g(\gamma_l) - a_l b_l(\gamma_l) - a_l \xi_l(\gamma_l)),$$

where $b_l(\gamma_l) = \mathbb{E}[\hat{g}_l^{\mathrm{MSPSA}}(\gamma_l)|\gamma_l] - g(\gamma_l)$ and $\xi_l(\gamma_l) = \hat{g}_l^{\mathrm{MSPSA}}(\gamma_l) - \mathbb{E}[\hat{g}_l^{\mathrm{MSPSA}}(\gamma_l)|\gamma_l]$.

Recall from Lemma 7.1 that $\mathbb{E}[\epsilon_l^+ - \epsilon_l^- | \gamma_l] = \mathbb{E}\{\mathbb{E}[\epsilon_l^+ - \epsilon_l^- | \gamma_l, \Delta_l] | \gamma_l\} = 0$ a.s.. Using the same argument as [92, Lemma 1] shows that $\|b_l(\gamma_l)\| < \infty$ a.s. for all l and $\lim_{l\to\infty} b_l(\gamma_l) = 0$ a.s.. Furthermore, following the same argument as in [92, Proposition 1] shows that for any $\eta > 0$,

$$\lim_{k\to\infty} \mathbb{P}\left(\sup_{m\geq k} \left\| \sum_{l=k}^{m} a_l \xi_l(\gamma_l) \right\| \geq \eta \right) = 0. \tag{7.13}$$

Then the assumptions of [177, Theorem 5.3.1] are satisfied, and the result follows.

□

Remark 7.2. *In the above proof, we have used* (7.13) *instead of A5.3.2 in [177], which states that there is a* $T > 0$ *such that for each* $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{j\geq n} \max_{t\leq T} \left\| \sum_{i=m(jT)}^{m(jT+t)-1} a_i \xi_i(\gamma_i) \right\| \geq \eta \right) = 0,$$

where $m(t) = \max\{n : \sum_{l=1}^{n} a_l \leq t\}$ *for* $t \geq 0$ *and* $m(t) = 0$ *otherwise. While A5.3.2 is sufficient to show the result of Theorem 7.1, condition* (7.13) *is a stronger condition and can often be easier to verify; see [177, Page 29].*

7.4 Numerical examples

We employ the proposed reservoir design Algorithm 3 to optimize the internal and output parameters of ESNs to model the outputs of two tasks. Recall that the ESN is governed by [10]

$$\begin{cases} x_{k+1} = \tanh(Ax_k + Bu_k), \\ \hat{y}_k = h_W(x_k), \end{cases}$$

where $\sigma_{\max}(A) < 1$ to ensure the uniform convergence property. We optimize internal parameters A and B using Algorithm 2. To preserve the uniform convergence property, we set A to be diagonal and constrain its diagonal elements to $[-0.9, 0.9]$. We also constrain elements of B to $[-1, 1]$. At each iteration of the algorithm, we apply an independent additive Gaussian noise with mean zero and standard deviation $\sigma_{std} = \{0.05, 0.1, 0.2\}$ to the state to investigate the effect of noise on the optimization. This state measurement noise satisfies Assumptions 7.1 and 7.2.

The first task is the LRPO task [25, 26] described by

$$\begin{cases} x_k = A_{\mathrm{LRPO}} x_{k-1} + B_{\mathrm{LRPO}} u_k, \\ y_k = \hat{h}(x_k), \end{cases}$$

where $\hat{h}(\cdot)$ is chosen to be a degree two multivariate polynomial, whose coefficients are randomly and uniformly chosen from $[-0.2, 0.2]$. We set A_{LRPO} to be a 500×500 real matrix with $\sigma_{\max}(A_{\mathrm{LRPO}}) = 0.5$ to ensure its uniform convergence property. Elements of A_{LRPO} and B_{LRPO} are randomly and uniformly generated from $[-1, 1]$. The input sequence is randomly and uniformly chosen from $u_k \in [0, 1]$ for each k. For this task, we use the first $L_w = 100$ data points to washout the effect of ESN's initial condition. We then optimize the internal and output parameters using the next $L_{\mathrm{train}} = 500$ data points. At each iteration of Algorithm 3, the output parameter W is optimized via ordinary least squares, that is, $W_l^* = \left(\hat{V}(\gamma_l)^\top \hat{V}(\gamma_l)\right)^{-1} \hat{V}(\gamma_l) Y$. We test the ESN performance using the last $L_{\mathrm{test}} = 100$ data points. The ESN output during the testing phase is $\hat{y}_k = \hat{v}_k(\hat{\gamma}_*)^\top \hat{W}_*$, where $\hat{\gamma}_*$ and $\hat{W}_*$ are the optimized internal and output parameters.

The second task is to perform one-step ahead prediction of the finance time series employed in Chapter 6. Recall that this time series describes weekly 5/1-year adjustable rate mortgage average (2005–2020) in the US [160]. As before, we remove the trend and seasonable components using the "mstl" command in R [168]. For this task, we consider implementing a NARX(∞) model using an ESN with output-feedback (6.11) as in Chapter 6. The first $L_w = 100$ data are used to washout the effect of the ESN's initial condition. The internal and output parameters are optimized using $L_{\mathrm{train}} = 500$ data and the ESN performance is evaluated using the last $L_{\mathrm{test}} = 180$ data. At each iteration of Algorithm 3, the output parameter W is optimized by solving the following convex constrained program using CVXPY [178]:

$$\min_{W} \|\hat{V}(\gamma_l) W - Y\|^2$$
$$\text{subject to } \sigma_{\max}(A + CW^\top) \leq 0.99.$$

As shown in Sec. 6.3.1, this constraint ensures that the ESN with output-feedback is uniformly convergent and hence induces a stationary and Birkhoff-Khinchin ergodic NARX(∞) model. As for the LRPO task, the ESN output during the testing phase for the finance time series is $\hat{y}_k = \hat{v}_k(\hat{\gamma}_*)^\top \hat{W}_*$.

7.4.1 Sensitivity of reservoir internal parameters

The effect of optimizing the reservoir internal parameters depends on the sensitivity of the RC's performance to these parameters. In this subsection, we investigate the sensitivity on the LRPO and finance time series tasks. In Figs. 7.1 and 7.2, we show the box plots of the mean-squared error (MSE) for 100 randomly generated ESNs against the reservoir size (i.e., the state-space size of the ESN) and the degree of the readout function at $\sigma_{std} = 0.05$. For each ESN, we randomly and uniformly sample diagonal elements of A from $[-0.9, 0.9]$ and the elements of B from $[-1, 1]$.

In Figs. 7.1 and 7.2(a), we observe that the median ESN MSE decreases as the reservoir size increases for the two tasks. However, the MSE quartile also decreases dramatically as the reservoir size increases, suggesting that performance of ESNs with large state spaces is less sensitive to their internal parameters for the chosen tasks. This means that optimizing the ESN internal parameters when the state space is large may not be effective to improve performance for the two tasks. The same trend is observed for different readout degrees (linear, quadratic and cubic as shown in Figs. 7.1 and 7.2) and other values of σ_{std}. An exception is shown in Fig. 7.2(b), where the median MSE during the testing phase increases for the cubic readout function as the reservoir size increases from 25 to 50. This behavior could be caused by overfitting, which occurs when the model fits against the training data extremely well and becomes unable to generalize well to test data. Indeed, Fig. 7.2 suggests that the ESN has a tendency to overfit when modeling the finance time series, indicated by the steady decrease in $\mathrm{MSE}_{\mathrm{train}}$ but a slower decrease in $\mathrm{MSE}_{\mathrm{test}}$ and its large quartile as the reservoir size increases from 25 to 50. Based on these observations, we choose a reservoir size of two for the ESN, which appears to be highly sensitive to changes in the reservoir internal parameters for the two tasks, to illustrate our proposed algorithm.

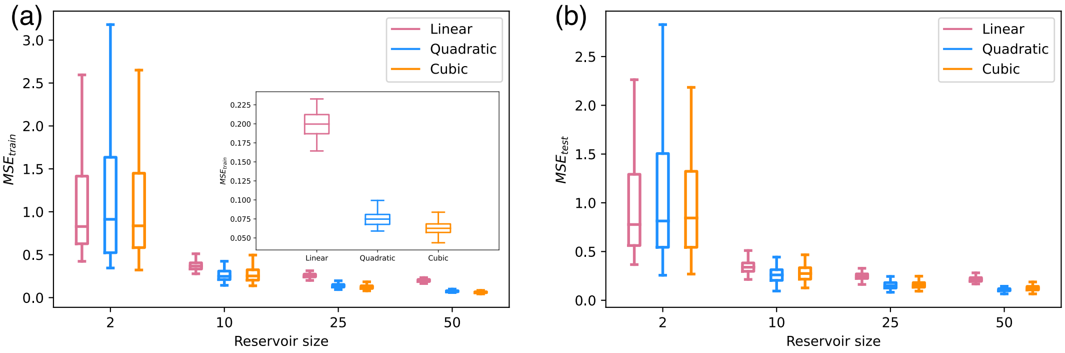


Figure 7.1: Box plot of the MSE of ESNs during (a) the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) the testing phase ($\mathrm{MSE}_{\mathrm{test}}$) for the LRPO task. The inset in (a) shows the box plot for reservoir size 50 during the training phase.

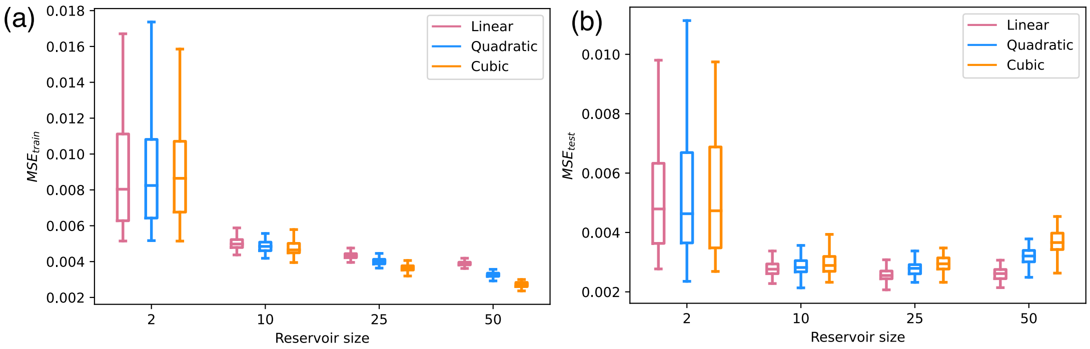


Figure 7.2: Box plot of the MSE of ESNs during (a) the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) the testing phase ($\mathrm{MSE}_{\mathrm{test}}$) for modeling the finance time series.

The above observations also suggest two directions to improve ESN performance. The first is to increase the ESN state-space size, this has also been observed in Chapter 3 on the NARMA tasks. The second is to optimize the internal parameters when the ESN state-space size is small. With this view, reservoir design is suitable in situations where increasing the reservoir size is more costly than the computational cost of the optimization algorithm employed to tune the internal parameters.

7.4.2 Performance

We employ Algorithm 3 to optimize the diagonal elements of A, elements of B and output parameters W of the ESN. We set the number of iterations to be $K = 250$ in Algorithm 3 and the number of iterations for Algorithm 2 to be $K_{\text{MSPSA}} = 500$. Furthermore, we set $\alpha_{\text{MSPSA}} = 0.7501, c_{\text{MSPSA}} = 0.1$ and $\rho_{\text{MSPSA}} = 0.25$ in Algorithm 2 for both tasks. For the LRPO task, we set $A_{\text{MSPSA}} = 10$ and $a_{\text{MSPSA}} = 0.1$. For the finance time series task, we set $A_{\text{MSPSA}} = 100$ and $a_{\text{MSPSA}} = 0.001$. These choices satisfy the condition Point #2 in Theorem 7.1. The hyper-parameters for Algorithm 2 are determined as follows. The values of $\alpha_{\text{MSPSA}}, \rho_{\text{MSPSA}}$ are set to be the same as in [92]. We then fix $c_{\text{MSPSA}} = 0.1$ for both tasks and choose $a_{\text{MSPSA}}, A_{\text{MSPSA}}$ for each task for an appropriate step size a_l. We set these values manually such that the internal parameter γ_l lies in the interior of the constraint set $\mathcal{C}$ for most iterations. We remark that these hyper-parameters can influence the performance of the algorithm on these tasks and there is no systematic procedure for selecting them yet. Nevertheless, the work [179] provides some guidelines in setting these values and some recent Python packages have implemented hyper-parameter optimization based on grid search [180].

To investigate the effect of state measurement noise on the reservoir design algorithm, we also employ Algorithm 1 to optimize the ESN internal parameters. We emphasize again that in Algorithm 1, the noisy cost function is computed using the noisy regressor $h(\hat{x}_k)$ instead of the regressors $\hat{v}_k$ and $\tilde{v}_k$ in Algorithm 2.

Fig. 7.3 plots the average ESN MSE against iteration for modeling the LRPO output, averaged over 10 random initial guesses of the internal parameters. For all readout degrees, we observe that as σ_{std} increases, the ESN performance worsens, indicating the negative impact of noise. As the readout degree increases, the ESN performance under Algorithm 1 (dashed lines in Fig. 7.3 denoted by SPSA) does not improve noticeably. This is in contrast to the ESN performance under Algorithm 2 (solid lines in Fig. 7.3 denoted by MSPSA), which shows noticeable improvement as the readout degree increases. Furthermore, as the readout degree increases, the performance discrepancy between Algorithm 1 and Algorithm 2 is more pronounced. In particular, for a cubic readout function under Algorithm 2, the ESN performance at $\sigma_{std} = 0.2$ achieves a comparable ESN performance under Algorithm 2 at $\sigma_{std} = 0.1$. This suggests that the proposed algorithm can reduce the negative impact of the noise, particularly when the readout degree and the noise variance are large.

In Algorithm 3, W^* is optimized using the regressor $\hat{V}(\gamma)$ instead of $\hat{H}(\gamma)$. This is chosen because we optimize the internal parameters using $\hat{V}(\gamma), \tilde{V}(\gamma)$ instead of $\hat{H}(\gamma)$. As a result, it is natural to optimize W^* in terms of $\hat{V}(\gamma)$. We provide numerical evidence to demonstrate that optimizing the output parameter W using $\hat{V}(\gamma)$ instead of $\hat{H}(\gamma)$ provides a better performance. To test this, we optimize the ESN internal parameters using our proposed modified SPSA Algorithm 2 as before using $\hat{V}(\gamma), \tilde{V}(\gamma)$ but optimize the output parameter as $W_l^* = \left(\hat{H}(\gamma_l)^\top \hat{H}(\gamma_l)\right)^{-1} \hat{H}(\gamma_l)Y$ at each iteration of Algorithm 3, this is referred to as USPSA. Fig. 7.4 plots the average ESN MSE for the LRPO task with a cubic readout function under these methods. It is observed that MSPSA performs better than USPSA for all values of σ_{std} considered.

We further test the proposed algorithm on modeling the finance time series. Fig. 7.5 plots the average ESN MSE against iteration for the finance time series task with a cubic readout function, averaged over 10 random initial guesses of the internal parameters. Similar observations as for the LRPO task can be found. For this task, the advantage of our algorithm is more pronounced as σ_{std} increases. At the last iteration, the percentage decreases in the average MSE between our algorithm and Algorithm 1 are 20.2% at $\sigma_{std} = 0.2$, 9.8% at $\sigma_{std} = 0.1$ and 1.6% at $\sigma_{std} = 0.05$ during the training phase. The percentage decreases during the testing phase are 21.2% at $\sigma_{std} = 0.2$, 10.7% at $\sigma_{std} = 0.1$ and 6.2% at $\sigma_{std} = 0.05$. This again suggests that the proposed algorithm is effective at reducing the negative impact of the noise when the noise variance is large.

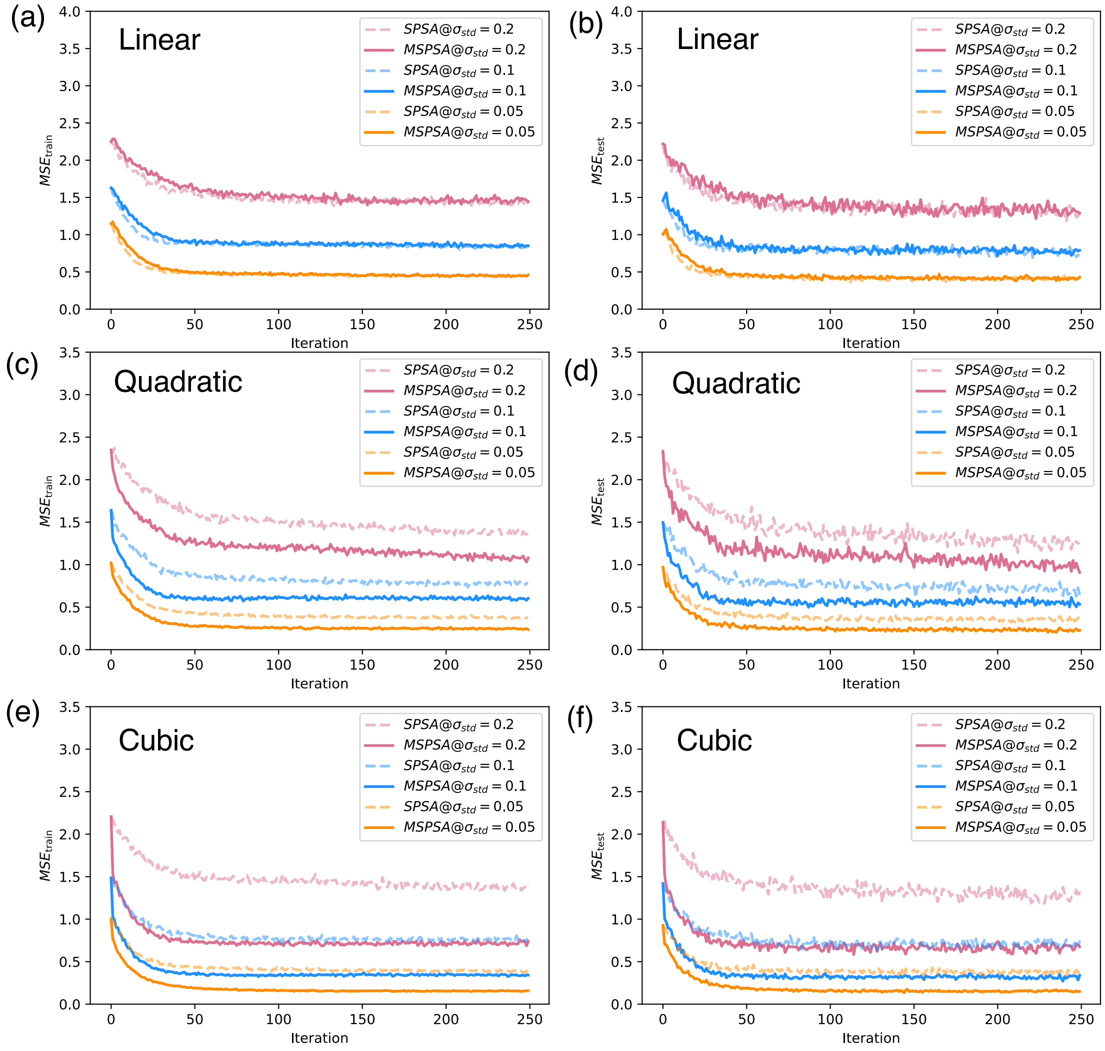


Figure 7.3: The average ESN MSE against iteration for the LRPO task. The first column shows the average ESN MSE during the training phase ($\text{MSE}_{\text{train}}$) and the second column shows the average ESN MSE during the testing phase (MSE_{test}). The first row corresponds to a linear output, the second row corresponds to a quadratic output and the third row corresponds to a cubic output. MSPSA refers to Algorithm 2 and SPSA refers to Algorithm 1.

7.5 Discussion

The work in this chapter presents a stochastic approximation approach for designing the reservoir under state measurement noise. There are several interesting extensions of the

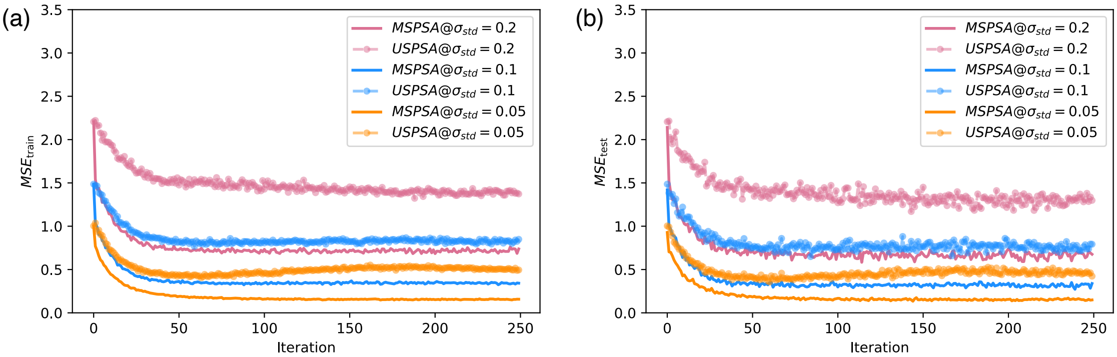


Figure 7.4: The average ESN MSE against iteration for the LRPO task with a cubic readout function under Algorithm 3 (labelled as SPSA) and Algorithm 3 where $W_l^* = \left(\hat{H}(\gamma_l)^\top \hat{H}(\gamma_l)\right)^{-1} \hat{H}(\gamma_l)Y$ (labelled as USPSA). (a) shows the average ESN MSE during the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) shows the average ESN MSE during the testing phase ($\mathrm{MSE}_{\mathrm{test}}$).

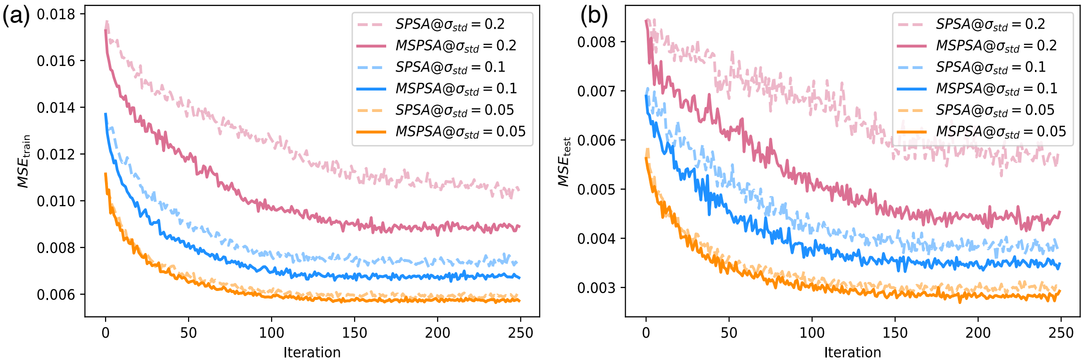


Figure 7.5: The average ESN MSE against iteration for the finance time series task with a cubic readout function. (a) shows the average ESN MSE during the training phase ($\mathrm{MSE}_{\mathrm{train}}$) and (b) shows the average ESN MSE during the testing phase ($\mathrm{MSE}_{\mathrm{test}}$).

current proposal.

Throughout this chapter, we have assumed that the state noise is independent of the internal parameter γ. Although this is a reasonable assumption for most classical RCs, for QRCs this is not true. This is because the finite sampling noise from measuring a quantum system depends on the quantum state; see Sec. 2.2. Furthermore, Assumption 7.2 may not hold for QRCs. Consequentially, Lemma 7.1 and Theorem 7.1 may not hold. It remains an open problem following from this chapter to extend the approach herein to the setting of QRC, if possible.

Another possible challenge of this extension is the barren plateau problem for quantum systems [181]. The barren plateau problem was initially recognized in hybrid quantum-classical algorithms for classical optimization problems and is derived using the Haar measure on the unitary group. Roughly speaking, the Haar measure is akin to the uniform distribution on the unitary group. A detailed discussion of the Haar measure is out of the scope of this thesis, however, excellent references are available on this topic [182–184]. Hybrid quantum-classical algorithms often employ parametrized quantum circuits of the form

$$U(\theta) = \prod_{l=1}^{L} U_l(\theta_l),$$

where $U_l(\theta_l) = \exp(-\iota\theta_l V_l)$ is a unitary gate parametrized by θ_l. Consider an objective function $F(\theta)$ expressed as the expectation value over some Hermitian observable O,

$$F(\theta) = \mathrm{Tr}\left(\rho_0 U(\theta)^\dagger O U(\theta)\right), \tag{7.14}$$

where ρ_0 is the initial state. The partial derivatives are of the form

$$\partial F(\theta)/\partial\theta_m = \iota \mathrm{Tr}\left(\rho_0 U_{m,-}^\dagger [V_m, U_{m,+}^\dagger O U_{m,+}] U_{m,-}\right),$$

where $U_{m,-} = \prod_{l=1}^{m-1} U_l(\theta_l)$ and $U_{m,+} = \prod_{l=m}^{L} U(\theta_l)$. Suppose that the circuit $U(\theta)$ is chosen such that $U_{m,-}, U_{m,+}$ or both of them match the Haar distribution up to the second moment (also referred to as unitary 2-designs), and the circuits $U_{m,-}, U_{m,+}$ are independent. In this case, the expectation value of $\partial F(\theta)/\partial\theta_m$ over the Haar measure is zero. Furthermore, the variance of $\partial F(\theta)/\partial\theta_m$ decreases exponentially as the number of qubits increases. This poses a challenge for gradient-based methods to optimize θ. Examples of such random quantum circuits include [185, 186]. In [186], the authors construct approximate unitary-2 design circuits using gates that are diagonal in the Pauli-Z and Pauli-X bases. Such approximate unitary-2 design circuits also cause the barren plateau problem.

For QRCs on gate-based quantum computers as in Chapter 4, their dynamics are described by

$$\rho_{k+1} = (1-\epsilon)(u_k U_0(\theta)\rho_k U_0(\theta)^\dagger + (1-u_k)U_1(\phi)\rho_k U_1(\phi)^\dagger) + \epsilon\sigma,$$

where $U_0(\theta) = \prod_{l=1}^{L} U(\theta_l), U_1(\phi) = \prod_{l=1}^{L'} U'(\phi_l)$ are two unitaries parametrized by θ and ϕ respectively. Here, $U(\theta_l) = \exp(-\iota\theta_l V_l)$ and $U'(\phi_l) = \exp(-\iota\phi_l V_l')$. For ease of notation, we will consider QRCs with linear outputs. The ensuing discussions also apply for polynomial outputs. We also consider optimizing the QRC internal parameters θ, ϕ and its output parameters $W = \begin{bmatrix} W_1 & \cdots & W_n \end{bmatrix}^\top$ in a round robin fashion. Here, n is the number of qubits. Let $W^* = \begin{bmatrix} W_1^* & \cdots & W_n^* \end{bmatrix}^\top$ be the optimized output parameters. For a QRC with linear output, the objective function is

$$F(\theta, \phi) = \frac{1}{T}\sum_{k=1}^{T}\left(y_k - \sum_{i=1}^{n} W_i^* \mathrm{Tr}\left(\rho_k Z^{(i)}\right)\right)^2.$$

The partial derivatives are of the form

$$\frac{\partial F(\theta, \phi)}{\partial \theta_m} = \frac{2}{T}\sum_{k=1}^{T}\left(y_k - \sum_{i=1}^{n} W_i^* \mathrm{Tr}\left(\rho_k Z^{(i)}\right)\right)\left(-\sum_{i=1}^{n} W_i^* \frac{\partial}{\partial \theta_m}\mathrm{Tr}\left(\rho_k Z^{(i)}\right)\right),$$

and similarly for $\partial F(\theta, \phi)/\partial \phi_m$. Note that the expectation value $\mathrm{Tr}\left(\rho_k Z^{(i)}\right)$ can be written as a sum of expectation values of the form (7.14). For instance, when $k = 1$,

$$\begin{aligned}&\mathrm{Tr}\left(\rho_1 Z^{(i)}\right)\\ &= (1-\epsilon)\left\{u_1 \mathrm{Tr}\left(\rho_0 U_0(\theta)^\dagger Z^{(i)} U_0(\theta)\right) + (1-u_k)\mathrm{Tr}\left(\rho_0 U_1(\phi)^\dagger Z^{(i)} U_1(\phi)\right)\right\} + \epsilon \mathrm{Tr}(\sigma Z^{(i)})\end{aligned}$$

and $\frac{\partial}{\partial \theta_m}\mathrm{Tr}\left(\rho_1 Z^{(i)}\right) = (1-\epsilon)u_1\iota\mathrm{Tr}\left(\rho_0 U_{m,-}^\dagger [V_m, U_{m,+}^\dagger Z^{(i)} U_{m,+}] U_{m,-}\right)$, where $U_{m,-} = \prod_{l=1}^{m-1} U_l(\theta_l)$ and $U_{m,+} = \prod_{l=m}^{L} U(\theta_l)$. If $U_{m,-}$, $U_{m,+}$ or both of them are unitary 2-designs and they are independent, then the expectation of $\frac{\partial}{\partial \theta_m}\mathrm{Tr}\left(\rho_1 Z^{(i)}\right)$ over the Haar measure is zero. Moreover, its variance decreases exponentially as the number of qubits increases. When this is true for all $k = 1, \ldots, T$, the partial derivative $\partial F(\theta, \phi)/\partial \theta_m$ is exponentially vanishing and poses a difficulty for our proposed gradient-based algorithm to optimize the QRC internal parameters. To make matters more complicated, it has been shown that quantum circuits comprising of only a few parametrized gates (as few as almost linear in the number of qubits) can construct (approximate) unitary 2-designs [185, 186].

7.6 Conclusion

We have introduced a stochastic approximation algorithm to optimize the reservoir internal parameters that exhibits almost sure convergence to a KT point, even under state measurement noise. Our proposed algorithm only uses four noisy state measurements to approximate the gradient, making it efficient for high-dimensional problems. Key to our algorithm is the construction of regressors using four noisy state measurements to ensure that the gradient estimate is asymptotically unbiased. Numerical examples demonstrate that our algorithm can mitigate the negative impact of state measurement noise for the ESN on certain tasks, in particular for higher-order ESN polynomial outputs.

This chapter has only considered state measurement noise that is independent of the state and the internal parameters. Although this assumption may be true for most classical dynamical systems, for quantum systems, it is not valid. It may be challenging to extend the current algorithm to QRCs. Furthermore, to investigate the reservoir design problem for QRCs, it is important to address the barren plateau problem. These directions are crucial to further improve the emulation ability of QRCs as well as to gain more insights into the feasibility of quantum systems for other applications.

Chapter 8

Conclusion and Future Directions

In this thesis, we have explored three main directions in using uniformly convergent dynamical systems as RCs for temporal information processing–quantum implementations of the RC scheme, new architectures of RC with stability guarantee for system identification and time series modeling, and an efficient algorithm for reservoir design. Although our works have demonstrated some initial promising results in theoretical analysis, experiments and numerical simulations, there are many challenging open problems to further advance these works. We will discuss some of these open challenges below.

The first main question posed for QRCs is can they take advantage of the exponentially large quantum state space to provide a speedup over their classical counterparts? Some preliminary arguments have been put forward in the literature [54, 69]. It is argued that a quantum speedup is most likely to occur from QRCs that process quantum information [69]. This is because both encoding classical data in quantum states and reconstructing classical representations of a quantum states generally require an exponential number of operations. This belief is challenged by the recent work [187], which proves that quantum kernel methods provide an exponential speedup for certain datasets with only classical access to data. However, this work considers processing of (random) input vectors for classification on a full-fledged quantum computer. This opens an opportunity to extend this approach to the QRC that processes (infinite) input sequences for time series modeling and time series classification.

Another related question is the choice of quantum circuit structure that can provide a performance advantage. This is referred to the expressive power of quantum circuits. Notable examples are Boson sampling [39] and instantaneous quantum polynomial time (IQP) circuits [188]. These works demonstrate a form of 'quantum supremacy' in which the probability distribution generated by quantum devices can not be efficiently and accurately sampled classically. More recently, this expressiveness has been studied in the quantum machine learning context; see e.g. [189, 190]. The work [189] proves that certain parametrized quantum circuits with a simple structure can outperform any classical neural network for generative tasks under the polynomial hierarchy hypothesis. In [190], the authors show that shallow alternating layered circuits are almost as expressive as hardware efficient circuits with a longer gate sequence measured by the frame potential and KL–divergence. Such shallow alternating layered circuits are particularly suitable for NISQ computers as the effect of noise is less severe for a shorter gate sequence. Taking inspirations from these works, it would be interesting to develop methods for choosing quantum circuit structures based on their expressive power for QRCs.

In this thesis, we have only considered QRCs implemented on superconducting NISQ computers for proof-of-principle experiments. As discussed in Chapter 4, we are limited by the long measurement times of these platforms, which make fast or real-time QRC difficult to realize. Nevertheless, there are many other quantum computing platforms being pursued in the QRC context. For instance, the work [129] considers QRCs implemented on photonic quantum computers, which have the potential for fast input encoding and fast measurements (e.g., the input data may be encoded using fast femtosecond lasers). A study on implementing QRCs on these fast quantum devices will be important to facilitate real-time QRCs.

Another avenue to explore is quantum–inspired RC instead of quantum implementation of the RC scheme, where the gate-model QRC in Chapter 4 is implemented as a software-based RC (for a small number of qubits on classical computers). The advantage of this quantum–inspired RC scheme is that the uniform convergence property is always guaranteed. This is closely related to the recent work [191], which proposes a convex parametrization of stable recurrent neural networks [1] and observes a significant increase in the speed of training and model performance. Furthermore, when simulated on classical computers, we can control the memory depth of the quantum-inspired RC by choosing the ϵ parameter appropriately. In the linear case, we can take $\hat{y}_k = \mathrm{Tr}(W\rho_k)$ as the output, where the output weight W is any complex Hermitian matrix of the same dimension as the density operator ρ_k.

With regard to employing interconnected RCs for system identification and time series modeling, our framework could be improved in several directions. A natural future extension would be to relax the small-gain condition in Chapter 5 and the sufficient conditions for the uniform convergence property in Chapter 6. A possible avenue to investigate this relaxation is to consider local stability analysis as some systems only operate in specific regimes. While we have derived a general theoretical framework to ensure the uniform convergence property of interconnected RCs, guidance for selecting suitable models would be useful for end-users. Many model selection methods based on information criteria have been studied in the literature; see e.g. [145, 153, 167]. Nevertheless, most of these model selection methods are developed for unconstrained optimization of the parameters, only a few tackle model selection under constraints [192–196]. While developing a model selection method for general constrained problems is challenging, we may be able to exploit the convexity in the parameter optimization problems for a number of RC schemes. Investigating model selection under convex constraints would be a natural extension of Chapter 6.

For the reservoir design problem, a natural continuation of the stochastic approximation approach presented in Chapter 7 is to optimize the internal parameters of QRCs. This may require a major extension of our current proposal, since the current framework only considers state measurement noise that is independent of the state and the internal parameters. Another challenge is the barren plateau problem for quantum circuits [181]. Since its initial recognition, many studies have proposed methods to mitigate the barren plateau problem. The work [197] proposes an initialization strategy to ensure that the gradient during the first parameter update is non-zero in expectation. However, there is no guarantee that the barren plateau problem does not occur for subsequent parameter updates. The work [198] shows that the barren plateau problem is cost function dependent and that defining a cost function with local observables leads to a polynomially vanishing gradient as opposed to an exponentially vanishing one. However, this requires the parametrized circuit to be short (in the order of the natural log of the number of qubits). In [199], the authors show that limiting entanglement can mitigate the barren plateau problem, at the expense of circuit expressiveness. A recent work [200] further shows that gradient-free algorithms do not resolve the barren plateau problem, making gradient-free algorithms inefficient to optimize gate parameters. To our best knowledge, the barren plateau problem is still a crucial challenge for the quantum community. Nevertheless, [201] shows that quantum convolutional neural networks (QCNNs) do not suffer from the barren plateau problem. The variance of the gradient vanishes no faster than polynomially, providing an analytical guarantee for the trainability of randomly initialized QCNNs. Can we take inspirations from this work to design QRCs without barren plateaus? This is a crucial to further the development of the QRC scheme.

---

[1] The authors consider stability using contraction analysis, which is closely related to the uniform convergence property and also imposes conditions that all solutions ‘forget’ their initial conditions and converge to each other. However, these concepts were derived independently and motivated distinctly, and should be distinguished. See [22] for an in-depth comparison and the relationship between the uniform convergence property and contraction analysis.

Appendix A

Appendix for Chapter 4

A.1 Invariance under time-invariant readout error

When $R = 1$ in (4.4), the QRC predicted outputs $\hat{y}_k$ remain unchanged under time-invariant readout error. Let $\mathcal{B} = \{|i\rangle\}$ be the computational basis for an n-qubit system, with $i = 1, \dots, 2^n$. The readout error is characterized by a measurement calibration matrix A whose i, j-th element $A_{i,j} = \Pr(i|j)$ is the probability of measuring the state $|i\rangle \in \mathcal{B}$ given that the state is prepared in the state $|j\rangle \in \mathcal{B}$.

We employ the readout error correction method described in [47]. For an n-qubit QRC, at each timestep k, we execute 2^n calibration circuits with each circuit initialized in one of the 2^n computational basis elements. The outcomes are used to create the measurement calibration matrix A_k. The readout error at time step l is corrected by applying the pseudo-inverse of A_k to the measured outcomes from the experiments.

For all experiments, the measurement outcomes are stored as the count of measuring each basis elements in $\mathcal{B}$. Let $\mathbf{v}_k = \begin{pmatrix} \mathbf{v}_k^1 & \cdots & \mathbf{v}_k^{2^n} & 1 \end{pmatrix}$, where $\mathbf{v}_k^i$ is the count of measuring $|i\rangle \in \mathcal{B}$ at timestep k. Let $\mathbf{z}_k = \begin{pmatrix} \overline{\langle Z^{(1)} \rangle}_k & \cdots & \overline{\langle Z^{(n)} \rangle}_k & 1 \end{pmatrix}$, where $\overline{\langle Z^{(i)} \rangle}_l$ is the finite-sampled approximation of $\langle Z^{(i)} \rangle_l$ for $i = 1, \dots, n$. Then we have $\mathbf{z}_k = \mathbf{v}_k B$, where B is a linear transformation. After applying readout error correction, we have $\mathbf{z}'_k = \mathbf{v}_k A_k^+ B$, where A_k^+ is the pseudo-inverse of A_k. Let $\mathbf{w}$ denote a column vector of the readout parameters, with C being the last element. To optimize $\mathbf{w}$ collect all measurement data in a matrix $\mathbf{v} = \begin{pmatrix} \mathbf{v}_1^\top & \cdots & \mathbf{v}_L^\top \end{pmatrix}^\top$ so that $\mathbf{z} = \begin{pmatrix} \mathbf{z}_1^\top & \cdots & \mathbf{z}_L^\top \end{pmatrix}^\top = \mathbf{v}B$, where L is the sequence length. The linear output of the quantum reservoir computer is $\hat{\mathbf{y}} = \mathbf{v}B\mathbf{w}$. Append a corresponding row and column to $A_l^\dagger$ to account for the bias term. Suppose the readout error is time-invariant, then $A^+ = A_k^+$ for $k = 1, \ldots, L$. The QRC output after readout error correction is $\hat{\mathbf{y}}' = \mathbf{v}A^+B\mathbf{w}'$. Assume that A^+ has all rows linearly independent, then ordinary least squares yields $B\mathbf{w}' = AB\mathbf{w}$. Now given test data with readout error correction, $\mathbf{v}_{\text{test}}A^+B\mathbf{w}' = \mathbf{v}_{\text{test}}A^+AB\mathbf{w} = \mathbf{v}_{\text{test}}B\mathbf{w}$. Therefore, the QRC predicted outputs are invariant under time-invariant readout error.

Back to Sec. 4.1.1.

A.2 Monte Carlo estimation

For all schemes described in Sec. 4.3, we can set $S = 1$ and run N_m Monte Carlo sampled circuits (possibly in parallel if many copies of the same hardware are available) for a sufficiently large N_m. We show that the average of all N_m measurements at timestep k estimates $\langle Z^{(i)} \rangle_k$ and its variance vanishes as N_m tends to infinity.

First consider estimating $\langle Z^{(i)} \rangle_k$ by re-initializing each N_m circuit in $|0\rangle^{\otimes n}$ and re-running them from time 1 to time k according to inputs $\{u_1, \ldots, u_k\}$. Recall that

$$\langle Z^{(i)} \rangle_k = \mathrm{Tr}(Z^{(i)}\rho_k) = \mathrm{Tr}(Z^{(i)}T(u_k)\cdots T(u_1)(|0\rangle\langle 0|)^{\otimes n}),$$

where $T(u_l)$ is the input-dependent CPTP map defined in (4.1) for $l = 1, \ldots, k$. Define independent discrete-valued random variables X_k such that

$$\Pr(X_k = 0) = (1-\epsilon)u_k, \quad \Pr(X_k = 1) = (1-\epsilon)(1-u_k), \quad \Pr(X_k = 2) = \epsilon.$$

To implement the QR, for each time k, we independently sample N_m random variables $X_{k,j}$ $(j = 1, \ldots, N_m)$ from the same distribution as X_k. Define

$$T_x = \begin{cases} T_0, & \text{if } x = 0, \\ T_1, & \text{if } x = 1, \\ K_\sigma, & \text{if } x = 2, \end{cases}$$

where $K_\sigma(\rho) = \sigma$ is a constant CPTP map that sends any density operator ρ to the constant density operator σ in (4.2). The random CPTP maps $T_{X_{k,j}}$ follow the same distribution as $X_{k,j}$ and are independent for each k and j. Furthermore, $\mathbb{E}[T_{X_{k,j}}] = T(u_k)$.

For the j-th circuit, we implement a sequence of (random) CPTP maps $T_{X_{k,j}} \cdots T_{X_{1,j}}$ so that at timestep k, the (random) QRC state is $\rho^{\boldsymbol{X}_{k,j}} = T_{X_{k,j}} \cdots T_{X_{1,j}}(|0\rangle\langle 0|)^{\otimes n}$, where $\boldsymbol{X}_{k,j} = (X_{1,j}, \ldots, X_{k,j})$. For each j-th circuit, we measure $Z^{(i)}$ and denote its random outcome by $\overline{Z^{(i)}}_{k,j}$. Note that for $j = 1, \ldots, N_m$, $\overline{Z^{(i)}}_{k,j}$ are independent (but not necessarily identically distributed) random variables taking values ± 1 (eigenvalues of $Z^{(i)}$) with conditional probabilities (conditional on the random variables $\boldsymbol{X}_{k,j}$)

$$\Pr\left(\overline{Z^{(i)}}_{k,j} = z | \boldsymbol{X}_{k,j}\right) = \operatorname{Tr}\left(\rho^{\boldsymbol{X}_{k,j}} P^{(i)}_z\right), \quad z = \pm 1,$$

where $P^{(i)}_{\pm 1}$ are the projectors such that $Z^{(i)} = P^{(i)}_{+1} - P^{(i)}_{-1}$. Consider the average of all N_m measurement outcomes, by the law of total expectation,

$$\begin{aligned}
\frac{1}{N_m}\sum_{j=1}^{N_m} \mathbb{E}\left[\overline{Z^{(i)}}_{k,j}\right] &= \frac{1}{N_m}\sum_{j=1}^{N_m} \mathbb{E}\left[\mathbb{E}\left[\overline{Z^{(i)}}_{k,j} | \boldsymbol{X}_{k,j}\right]\right] \\
&= \frac{1}{N_m}\sum_{j=1}^{N_m} \mathbb{E}\left[\operatorname{Tr}\left(Z^{(i)} \rho^{\boldsymbol{X}_{k,j}}\right)\right] \\
&= \frac{1}{N_m}\sum_{j=1}^{N_m} \operatorname{Tr}\left(Z^{(i)} \mathbb{E}[T_{X_{k,j}}] \cdots \mathbb{E}[T_{X_{1,j}}](|0\rangle\langle 0|)^{\otimes n}\right) \\
&= \frac{1}{N_m}\sum_{j=1}^{N_m} \operatorname{Tr}\left(Z^{(i)} T(u_k) \cdots T(u_1)(|0\rangle\langle 0|)^{\otimes n}\right) \\
&= \operatorname{Tr}(\rho_k Z^{(i)}) = \langle Z^{(i)} \rangle_k,
\end{aligned}$$

therefore the finite-sample estimate is unbiased. Moreover, using the fact that

$$\mathbb{E}\left[\left(\overline{Z^{(i)}}_{k,j}\right)^2\right] = \sum_{z=\pm 1} z^2 \Pr\left(\overline{Z^{(i)}}_{k,j} = z\right) = 1,$$

the variance of the average of N_m measurements is

$$\operatorname{Var}\left[\frac{1}{N_m}\sum_{j=1}^{N_m}\overline{Z^{(i)}}_{k,j}\right]=\frac{1}{N_m^2}\sum_{j=1}^{N_m}\operatorname{Var}\left[\overline{Z^{(i)}}_{k,j}\right]=\frac{1}{N_m}\left(1-\langle Z^{(i)}\rangle_k^2\right).$$

Using the uniform convergence property, to estimate $\langle Z^{(i)}\rangle_k$ for a sufficiently large k (that depends on ϵ), we re-initialize N_m circuits at time $k-M$ and run the circuits according to inputs $\{u_{k-M+1},\ldots,u_k\}$. Let $\langle\widetilde{Z^{(i)}}\rangle_k=\operatorname{Tr}(Z^{(i)}\tilde{\rho}_k)$ where

$$\tilde{\rho}_k=T(u_k)\cdots T(u_{k-M+1})(|0\rangle\langle 0|)^{\otimes n}.$$

In this setting, for the j-th circuit, we implement a sequence of (random) CPTP maps $T_{X_{k,j}}\cdots T_{X_{k-M+1,j}}$ so that the (random) QRC state at time k is

$$\rho^{\widetilde{\boldsymbol{X}}_{k,j}}=T_{X_{k,j}}\cdots T_{X_{k-M+1,j}}(|0\rangle\langle 0|)^{\otimes n},$$

where $\widetilde{\boldsymbol{X}}_{k,j}=(X_{k-M+1,j},\ldots,X_{k,j})$. Let $\widetilde{Z^{(i)}}_{k,j}$ be the random outcome of measuring $Z^{(i)}$. The conditional probabilities are

$$\Pr\left(\widetilde{Z^{(i)}}=z|\widetilde{\boldsymbol{X}}_{k,j}\right)=\operatorname{Tr}\left(\rho^{\widetilde{\boldsymbol{X}}_{k,j}}P_z^{(i)}\right),\quad z=\pm 1.$$

A similar argument as above shows that the average of all N_m measurements satisfies

$$\mathbb{E}\left[\frac{1}{N_m}\sum_{j=1}^{N_m}\widetilde{Z^{(i)}}_{k,j}\right]=\langle\widetilde{Z^{(i)}}\rangle_k,\quad \operatorname{Var}\left[\frac{1}{N_m}\sum_{j=1}^{N_m}\widetilde{Z^{(i)}}_{k,j}\right]=\frac{1}{N_m}\left(1-\langle\widetilde{Z^{(i)}}\rangle_k^2\right).$$

The uniform convergence property ensures that the bias (in mean) vanishes exponentially fast,

$$\left|\mathbb{E}\left[\frac{1}{N_m}\sum_{j=1}^{N_m}\widetilde{Z^{(i)}}_{k,j}\right]-\langle Z^{(i)}\rangle_k\right|=\left|\operatorname{Tr}\left(Z^{(i)}(\tilde{\rho}_k-\rho_k)\right)\right|\leq\|\tilde{\rho}_k-\rho_k\|_1\leq 2(1-\epsilon)^M,$$

where we have used the fact that for any Hermitian matrix A, $|\operatorname{Tr}(Z^{(i)}A)|\leq\sigma_{\max}(Z^{(i)})\|A\|_1$, with $\sigma_{\max}(Z^{(i)})=1$ denotes the maximum singular value of $Z^{(i)}$. This shows that the bias can be exponentially suppressed by choosing M appropriately based on ϵ, so that the estimates of $\langle\widetilde{Z^{(i)}}\rangle_k$ and $\langle Z^{(i)}\rangle_k$ are also close.

Back to Sec. 4.3.1.

A.3 Implementation using QND measurements

We detail the second (more efficient) implementation scheme based on QND measurements [120] described in Sec. 4.3.

To explain QND measurements, we first show that direct measurement of Z on a 'system' qubit is equivalent to coupling the qubit with an ancilla qubit via CNOT and measuring Z_{a}, the Pauli Z operator acting on the ancilla qubit 'a' [43]. To see this, let $|\psi\rangle_{\mathrm{sys}} = \alpha|0\rangle_{\mathrm{sys}} + \beta|1\rangle_{\mathrm{sys}}$ be the state of the system qubit. Prepare the ancilla qubit at the ground state $|0\rangle_{\mathrm{a}}$. We write CNOT $= |0\rangle\langle 0|_{\mathrm{sys}} \otimes I_{\mathrm{a}} + |1\rangle\langle 1|_{\mathrm{sys}} \otimes X_{\mathrm{a}}$, where I_{a} and X_{a} are the identity and Pauli X operators acting on the ancilla qubit. The system and ancilla state after applying CNOT is $|\Psi\rangle = \mathrm{CNOT}|\psi\rangle_{\mathrm{sys}} \otimes |0\rangle_{\mathrm{a}} = \alpha|00\rangle + \beta|11\rangle$.

Measurement of Z_{a} on the ancilla qubit is described by the projectors $P_+ = I_{\mathrm{sys}} \otimes |0\rangle\langle 0|_{\mathrm{a}}$ and $P_- = I_{\mathrm{sys}} \otimes |1\rangle\langle 1|_{\mathrm{a}}$. The probabilities and post-measurement system states are

$$\begin{aligned}\Pr(+) &= \langle\Psi|P_+|\Psi\rangle = |\alpha|^2, \quad \frac{\mathrm{Tr}_{\mathrm{a}}\left(P_+|\Psi\rangle\langle\Psi|P_+\right)}{\Pr(+)} = |0\rangle\langle 0|_{\mathrm{sys}},\\ \Pr(-) &= \langle\Psi|P_-|\Psi\rangle = |\beta|^2, \quad \frac{\mathrm{Tr}_{\mathrm{a}}\left(P_-|\Psi\rangle\langle\Psi|P_-\right)}{\Pr(-)} = |1\rangle\langle 1|_{\mathrm{sys}},\end{aligned}$$

where $\mathrm{Tr}_{\mathrm{a}}(\cdot)$ is the partial trace over the ancilla qubit.

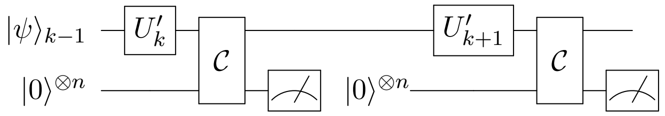


Figure A.1: Quantum circuit implementing the QND measurements by coupling ancilla qubits $|0\rangle^{\otimes n}$ with the QRC system qubits $|\psi\rangle_{k-1}$.

Now for an n-qubit QRC, we associate each system qubit in the QRC with its ancilla qubit. All n ancilla qubits are prepared in the ground state. Suppose that when restricted to pure state preparation, we have drawn N_m circuits using Monte Carlo sampling. For each of the N_m circuits and each time step k, we apply the aforementioned ancilla-coupled measurement of $Z^{(i)}$ for each system qubit in the QRC. After measuring the n ancilla qubits, we reset and re-prepare them in the ground state for measurements at next time $k+1$; see Fig. A.1.

In Fig. A.1, $|\psi\rangle_{k-1}$ denotes the state of the system (QRC) qubits and $|0\rangle^{\otimes n}$ denotes the ancilla qubits initialized in the groud state. Here we have grouped the system and ancilla qubits and represent them using single wires. The unitary operator U_k' is U_0 or U_1 with probabilities $(1-\epsilon)u_k$ and $(1-\epsilon)(1-u_k)$ and $U(k) = U_k'\mathcal{C}$, where $\mathcal{C}$ is a product of n CNOT gates each acting on the i-th system-ancilla qubit pair. Measuring $Z_{\mathrm{a}}^{(i)}$ on the i-th ancilla qubit and resetting it at each time step $k = 1, \ldots, L$ is equivalent to having L ancilla qubits associated to the i-th system qubit and measuring $Z_{\mathrm{a},k}^{(i)}$ (i.e., the k-th ancilla qubit associated to the i-th system qubit). The resulting QRC dynamics is

$$T(u_k)\rho_{k-1} = (1-\epsilon)\left(u_k T_0 + (1-u_k)T_1\right)\rho_{k-1} + \epsilon\sigma,$$

where $T_j(\rho_{k-1}) = \mathrm{Tr}_{\mathrm{A}}(U_j\mathcal{C}\rho_{k-1}\otimes(|0\rangle\langle 0|)^{\otimes n}\mathcal{C}^\dagger U_j^\dagger)$ for $j = 0, 1$, and $\mathrm{Tr}_{\mathrm{A}}(\cdot)$ is the partial trace over all n ancilla qubits denoted by 'A'.

We now show that the measured observables $Z_{a,k}^{(i)}$ commute at different times as required by QND. More generally, we will show that $Z_{\mathrm{a},k} = \bigotimes_{i=1}^{n} O_{\mathrm{a},k}^{(i)}$ $(k = 1, \ldots, L)$, where for each i we have $O_{\mathrm{a},k}^{(i)} = I^{(i)}$ (the identity operator on the i-th qubit) or $O_{\mathrm{a},k}^{(i)} = Z_{\mathrm{a},k}^{(i)}$, are QND observables. Firstly, we have the commutator $[Z_{a,k}, Z_{a,j}] = 0$ for all $k, j = 1, \ldots, L$. Denote the evolved observables in the Heisenberg picture by

$$Z_{\mathrm{a}}(l) = U(1)^\dagger \cdots U(l)^\dagger Z_{\mathrm{a},l} U(l) \cdots U(1) = U_{l:1}^\dagger Z_{\mathrm{a},l} U_{l:1},$$

where $U_{l:1} = U(l)\cdots U(1)$. For $k, j = 1, \ldots, L$ with $j < k$, we have

$$\begin{aligned}
[Z_{\mathrm{a}}(j), Z_{\mathrm{a}}(k)] &= U_{j:1}^\dagger Z_{\mathrm{a},j} U_{j:1} U_{k:1}^\dagger Z_{\mathrm{a},k} U_{k:1} - U_{k:1}^\dagger Z_{\mathrm{a},k} U_{k:1} U_{j:1}^\dagger Z_{\mathrm{a},j} U_{j:1} \\
&= U_{j:1}^\dagger Z_{\mathrm{a},j} U_{k:j+1}^\dagger Z_{\mathrm{a},k} U_{k:1} - U_{k:1}^\dagger Z_{\mathrm{a},k} U_{k:j+1} Z_{\mathrm{a},j} U_{j:1} \\
&= U_{k:1}^\dagger [Z_{\mathrm{a},j}, Z_{\mathrm{a},k}] U_{k:1} = 0,
\end{aligned}$$

where in the second last equality we have used the fact that $Z_{\mathrm{a},j}$ commutes with the future unitary operations $U_{k:j+1}$. If $j > k$, apply the same argument as above shows $[Z_{\mathrm{a}}(j), Z_{\mathrm{a}}(k)] = -[Z_{\mathrm{a}}(k), Z_{\mathrm{a}}(j)] = 0$. The commutativity of $Z_{\mathrm{a}}(j)$ and $Z_{\mathrm{a}}(k)$ for all $j, k \geq 1$ means that the sequence $\{Z_{\mathrm{a}}(j), j = 1, 2, \ldots\}$ has a joint probability distribution and constitutes a classical stochastic process. QND measurements on the sequence gives a realization of this stochastic process.

Back to Sec. 4.3.1.

A.4 Hardware specifications

The experiments were conducted on the IBM 20-qubit Boeblingen (version 1.0.0), 5-qubit Ourense (version 1.0.0) and 5-qubit Vigo (version 1.0.0) superconducting quantum processors [44]. The gate duration for an arbitrary single-qubit rotation gate U_3 [95] is $\tau_{U_3} \approx 71.1$ ns for all qubits whereas the CNOT gate durations differ for different qubits.

See Fig. A.2 for the 4-qubit and 10-qubit Boeblingen QRC quantum circuits. The circuits are chosen such that both QRCs have the same number of layers in U_0 and U_1. In this setting, the maximum duration of a circuit executed on the Boeblingen device is the same for both QRCs. The chosen qubits for the 4-qubit QRC and the 10-qubit QRC on the Boeblingen device are $Q = 0, 1, 2, 3$ and $Q = 0, 1, 2, 3, 5, 6, 7, 8, 10, 12$. These qubits were chosen due to their longer coherence times, shorter CNOT gate durations, smaller gate and readout errors. During the experiment, the maximum readout error was 10^{-2} and the maximum U_3 gate error implemented was 10^{-3}. The maximum CNOT gate error implemented was 4.3×10^{-2} and the maximum CNOT gate duration was $\tau_{\text{CNOT}} \approx 427$ ns. We assume that commuting gates can be executed in parallel. We choose $N_0 = N_1 = 5$ numbers of layers for U_0 and U_1 in the 4-qubit and 10-qubit Boeblingen QRCs. The maximum length of any input sequence (including the transient) for the multi-step ahead prediction and the map emulation problems is $L = 30$. Therefore, the maximum numbers of U_3 gate executions and CNOT gate executions is $5L = 5 \times 30 = 150$. The maximum duration of a circuit executed on the Boeblingen device was $150 \times (\tau_{U_3} + \tau_{\text{CNOT}}) \approx 150 \times (71.1 + 427) = 74.7$ μs, within the coherence times (T_1, T_2) for most qubits chosen.

The 5-qubit Ourense device achieves the same order of magnitude in readout errors, coherence times and CNOT gate durations as the 20-qubit Boeblingen device, but lower

CNOT gate errors. For the Ourense device, the maximum U_3 gate error and readout error implemented were 0.9×10^{-3} and 4.1×10^{-2}, and the maximum CNOT gate error implemented was 8×10^{-3}, a lower error compared to the Boeblingen device. The maximum CNOT gate duration implemented was $\tau_{\mathrm{CNOT}} \approx 576$ ns. For the 5-qubit Ourense QRC, the circuit implementing U_0 is longer than that for U_1. The U_0 circuit consists of four CNOT gates, and the maximum duration of a circuit executed for the 5-qubit Ourense QRC was $4L \times \tau_{\mathrm{CNOT}} \approx 70\,\mu$s, also within the coherence limits of most qubits.

The 5-qubit Vigo device is similar to the 5-qubit Ourense device. They have the same qubit couplings and share similar noise profile and hardware specifications. Rotational X and Y gates were used on this device, with gate duration $\tau = 35.5$ ns. The maximum single-qubit gate error implemented was 0.8×10^{-3} and the maximum readout error implemented was 7.8×10^{-2}. The maximum CNOT gate error and gate duration implemented was 1.3×10^{-2} and $\tau_{\mathrm{CNOT}} \approx 462.2$ ns, respectively. For this QRC, U_0 is the longer circuit consisting of three layers of single-qubit rotation Y gates and two layers of CNOT gates. Therefore, the maximum duration of a circuit implemented was $(3\tau + 2\tau_{\mathrm{CNOT}})L = (3 \times 35.5 + 2 \times 462.2) \times 30 \approx 31\ \mu$s.

Back to Sec. 4.4.

A.5 Quantum circuits for QRCs

We detail the circuits implementing the QRC dynamics in our proof-of-principle experiments presented in Sec. 4.4. The quantum circuits for the 4-qubit and 10-qubit Boeblingen QRCs are shown in Fig. A.2. The quantum circuits for the 5-qubit Ourense and 5-qubit Vigo QRCs are shown in Fig. A.3.

Back to Sec. 4.4.1.

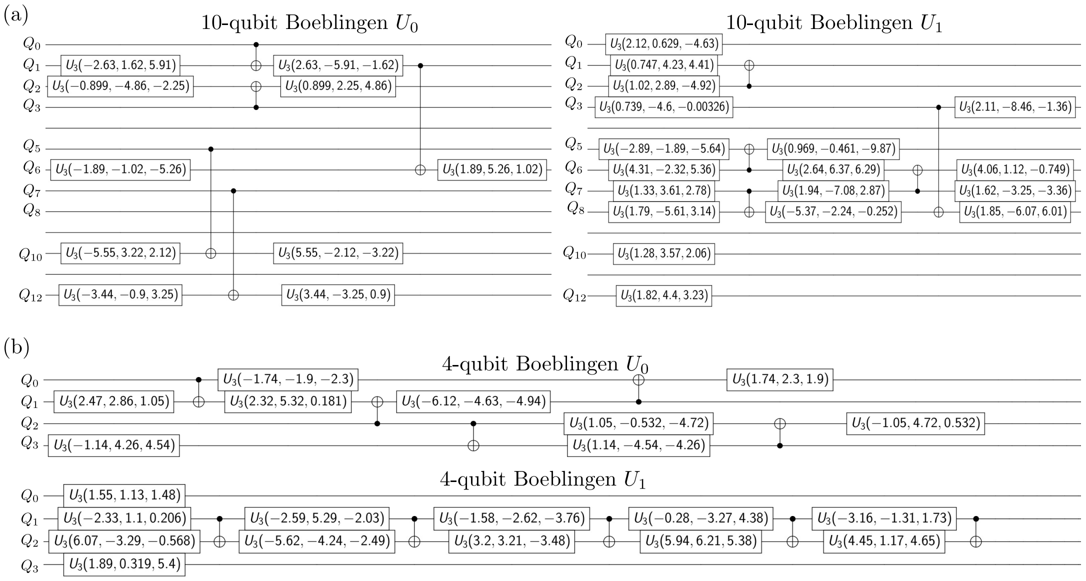

(a)
10-qubit Boeblingen U_0
Q_0
Q_1
Q_2
Q_3
Q_5
Q_6
Q_7
Q_8
Q_{10}
Q_{12}
$U_3(-2.63, 1.62, 5.91)$
$U_3(-0.899, -4.86, -2.25)$
$U_3(2.63, -5.91, -1.62)$
$U_3(0.899, 2.25, 4.86)$
$U_3(-1.89, -1.02, -5.26)$
$U_3(1.89, 5.26, 1.02)$
$U_3(-5.55, 3.22, 2.12)$
$U_3(5.55, -2.12, -3.22)$
$U_3(-3.44, -0.9, 3.25)$
$U_3(3.44, -3.25, 0.9)$
10-qubit Boeblingen U_1
$U_3(2.12, 0.629, -4.63)$
$U_3(0.747, 4.23, 4.41)$
$U_3(1.02, 2.89, -4.92)$
$U_3(0.739, -4.6, -0.00326)$
$U_3(2.11, -8.46, -1.36)$
$U_3(-2.89, -1.89, -5.64)$
$U_3(0.969, -0.461, -9.87)$
$U_3(4.31, -2.32, 5.36)$
$U_3(2.64, 6.37, 6.29)$
$U_3(4.06, 1.12, -0.749)$
$U_3(1.33, 3.61, 2.78)$
$U_3(1.94, -7.08, 2.87)$
$U_3(1.62, -3.25, -3.36)$
$U_3(1.79, -5.61, 3.14)$
$U_3(-5.37, -2.24, -0.252)$
$U_3(1.85, -6.07, 6.01)$
$U_3(1.28, 3.57, 2.06)$
$U_3(1.82, 4.4, 3.23)$
(b)
4-qubit Boeblingen U_0
$U_3(2.47, 2.86, 1.05)$
$U_3(-1.74, -1.9, -2.3)$
$U_3(2.32, 5.32, 0.181)$
$U_3(-6.12, -4.63, -4.94)$
$U_3(1.05, -0.532, -4.72)$
$U_3(1.14, -4.54, -4.26)$
$U_3(1.74, 2.3, 1.9)$
$U_3(-1.05, 4.72, 0.532)$
$U_3(-1.14, 4.26, 4.54)$
4-qubit Boeblingen U_1
$U_3(1.55, 1.13, 1.48)$
$U_3(-2.33, 1.1, 0.206)$
$U_3(6.07, -3.29, -0.568)$
$U_3(1.89, 0.319, 5.4)$
$U_3(-2.59, 5.29, -2.03)$
$U_3(-5.62, -4.24, -2.49)$
$U_3(-1.58, -2.62, -3.76)$
$U_3(3.2, 3.21, -3.48)$
$U_3(-0.28, -3.27, 4.38)$
$U_3(5.94, 6.21, 5.38)$
$U_3(-3.16, -1.31, 1.73)$
$U_3(4.45, 1.17, 4.65)$

Figure A.2: Quantum circuits for the (a) 10-qubit QRC and (b) 4-qubit QRC on the Boeblingen device.

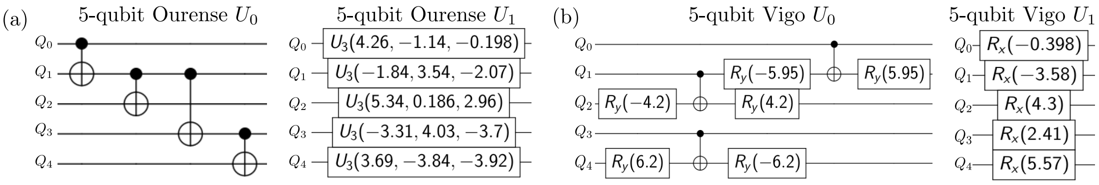

(a)
5-qubit Ourense U_0
Q_0
Q_1
Q_2
Q_3
Q_4
5-qubit Ourense U_1
$U_3(4.26, -1.14, -0.198)$
$U_3(-1.84, 3.54, -2.07)$
$U_3(5.34, 0.186, 2.96)$
$U_3(-3.31, 4.03, -3.7)$
$U_3(3.69, -3.84, -3.92)$
(b)
5-qubit Vigo U_0
$R_y(-4.2)$
$R_y(6.2)$
$R_y(-5.95)$
$R_y(4.2)$
$R_y(-6.2)$
$R_y(5.95)$
5-qubit Vigo U_1
$R_x(-0.398)$
$R_x(-3.58)$
$R_x(4.3)$
$R_x(2.41)$
$R_x(5.57)$

Figure A.3: Quantum circuits for the (a) 5-qubit Ourense QRC and (b) 5-qubit Vigo QRC.

A.6 Full input-output sequential data

In Fig. A.4, we show the full washout, train and test input-output target sequences for both the multi-step ahead prediction and the map emulation problems. Fig. A.5 plots the full target output sequences, the train and test QRC outputs on the multi-step ahead prediction problem. Fig. A.6 plots the full target output sequences, the train and test QRC outputs on the map emulation problem. In all figures, the transitory responses are indicated by dotted lines.

Back to Sec. 4.4.

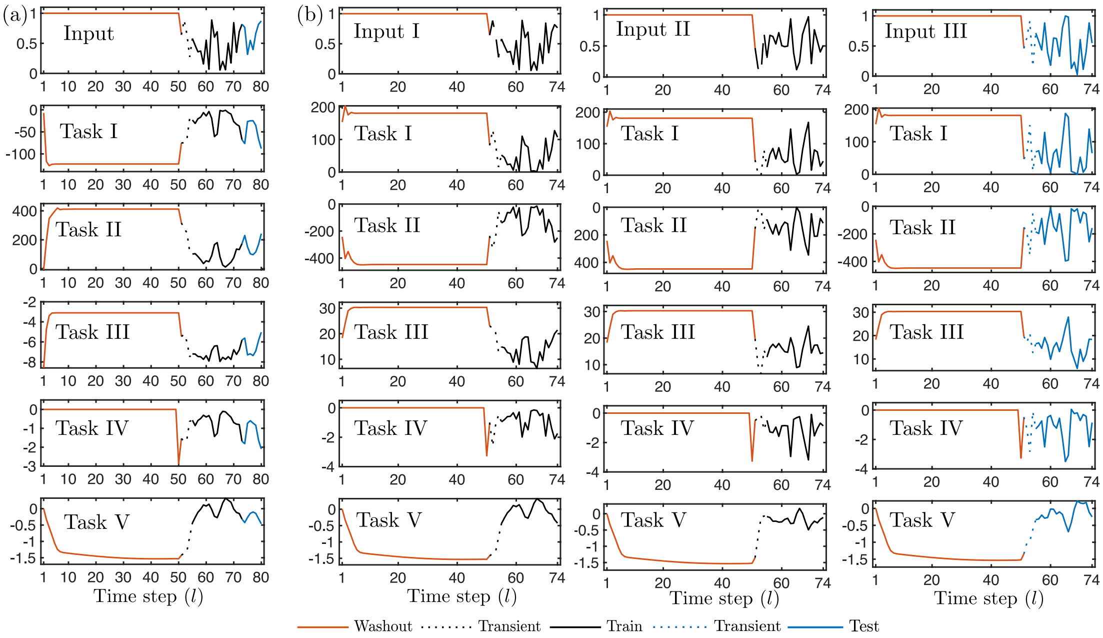


Figure A.4: Full washout, train and test input-output sequences for (a) the multi-step ahead prediction problem and (b) the map emulation problem. The first row in (a) and (b) shows the input sequences.

A.7 Measurement and simulation data

We simulate the four QRCs using the IBM Qiskit simulator under ideal and noisy conditions. The noise models used are obtained from the device calibration data. We fetched the updated device calibration data each time a job was executed on the hardware. The circuits simulated are the same as the circuits employed for the experiments and so is the number of shots. For the multi-step ahead prediction problem, the 10-qubit Boeblingen QR experienced a significant deviation from simulated results on qubits $Q = 1, 8$ (see Fig. A.7), resulting in larger NMSE $= 0.26, 0.29, 0.068, 0.15, 6.1$ for the four tasks. After setting the readout parameters for $Q = 1, 8$ to be zero, this issue was circumvented at the cost of using a fewer number of computational features. The resulting 10-qubit Boeblingen QRC still achieves performance improvement over other QRCs with a smaller number of qubits on the multi-step ahead prediction problem in the first three tasks. Although the QRC predicted outputs are invariant under time-invariant readout errors as derived in Appendix A.1. However, for the 10-qubit Boeblingen QR, the deviations in qubits $Q = 1, 8$

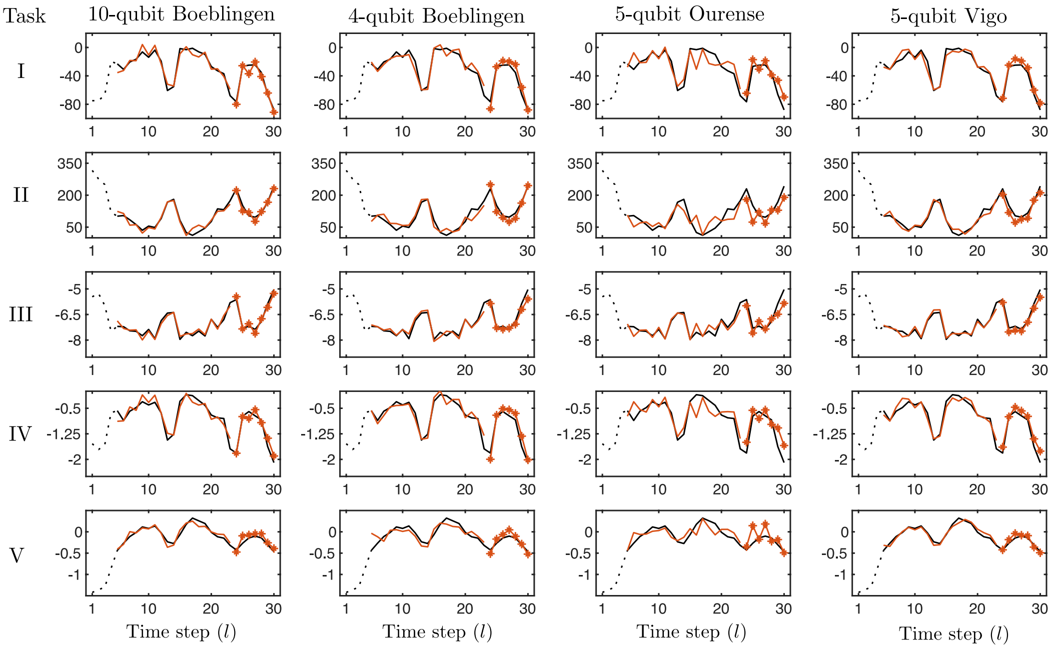


Figure A.5: The full target output sequences, the train and test output sequences of the four QRCs for each task on the multi-step ahead prediction problem. Each column corresponds to each n-qubit QRC outputs and each row corresponds to each task.

were time-varying. On the other hand, the 5-qubit Vigo device experienced almost time-invariant deviations in qubit $Q = 0$ as shown in Figs. A.7 and A.8, but this does not affect the performance of this QRC noticeably. The experimental results of the 5-qubit Ourense QRC follow the noisy simulation results closely. For the map emulation problem, the experimental results of both 5-qubit QRCs follow the simulated results closely, with an almost time-invariant shift in $Q = 0$ for the 5-qubit Vigo QRC.

Back to Sec. 4.4.

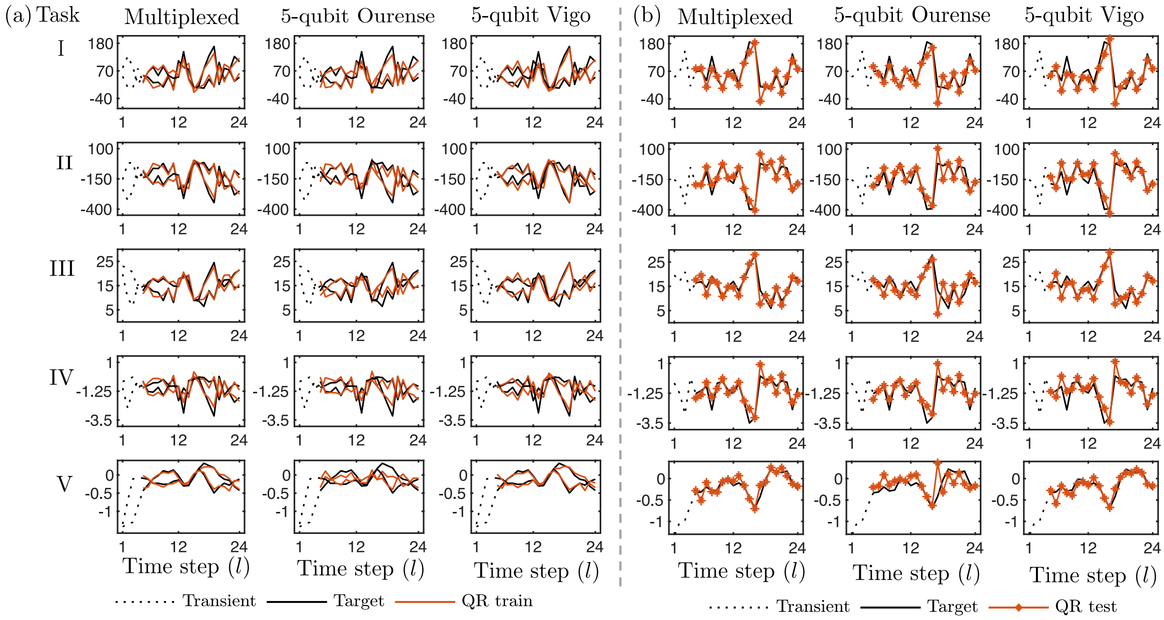


Figure A.6: The full target output sequences, the train and test output sequences of the QRCs for each task on the map emulation problem. (a) Shows the two train output sequences and (b) shows the test output sequence. The columns (from the left to the right) correspond to the multiplexed 5-qubit QRCs, 5-qubit Ourense QRC and the 5-qubit Vigo QRC. Each row corresponds to each task.

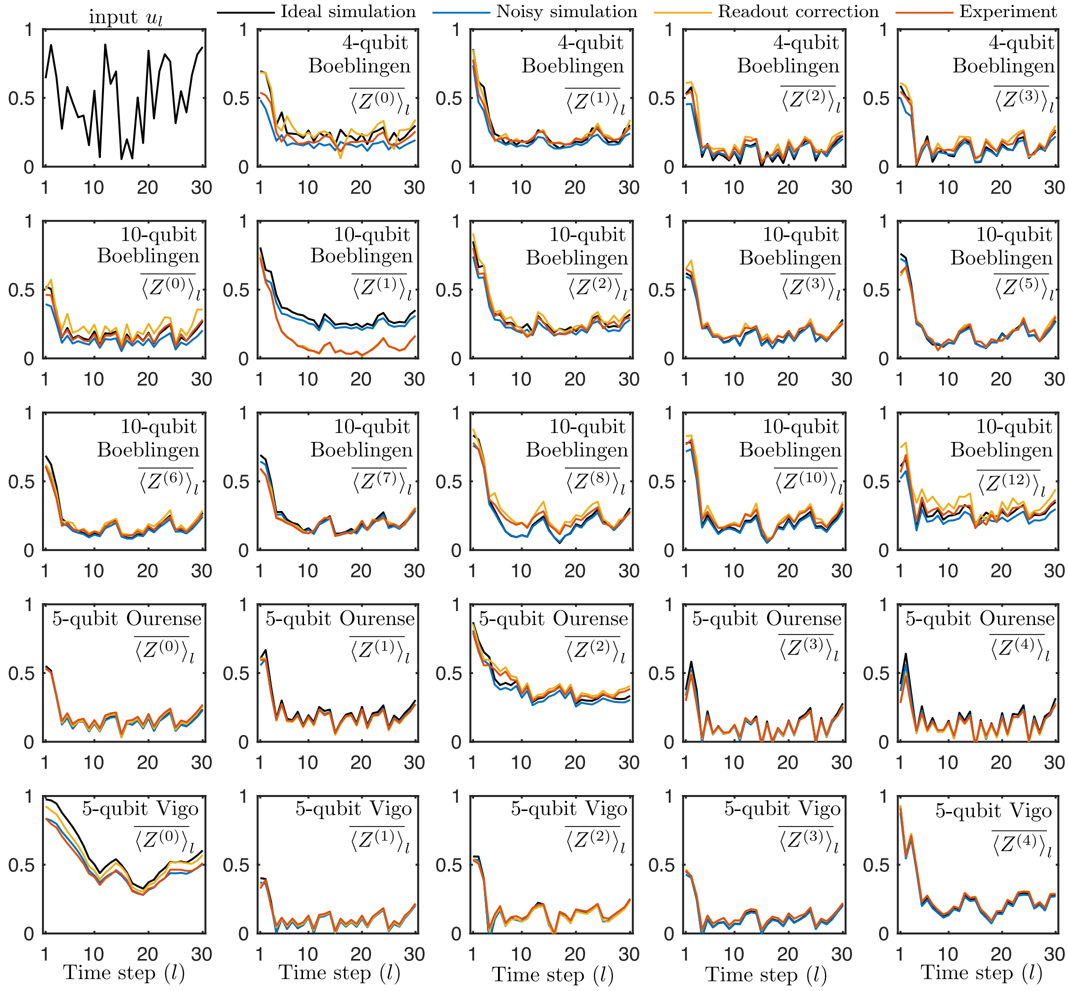


Figure A.7: Input sequence, experimental and simulation results for each qubit of the four QRCs at each time step $k = 1, \ldots, 30$, for the multi-step ahead prediction problem.

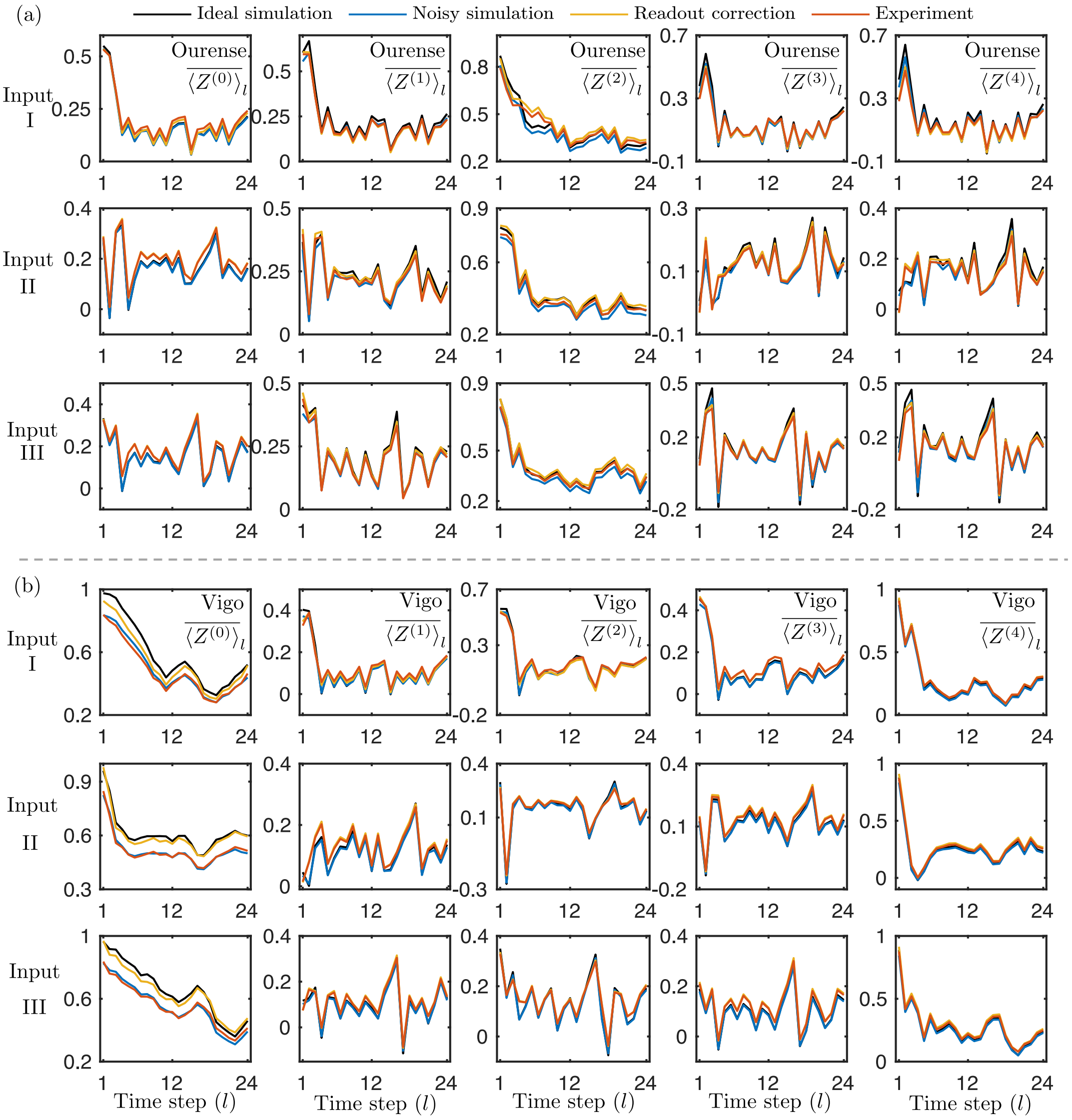


Figure A.8: Experimental and simulation results for each qubit $i = 0, \ldots, 4$ and each time step $k = 1, \ldots, 24$, for the map emulation problem. Three input sequences are used in this problem, labeled as inputs I, II and III. Row i in a sub-figure corresponds to the experimental data for the i-th input sequence. Column j corresponds to the experimental data for the j-th qubit. (a) Shows the experimental data for the 5-qubit Ourense QRC. (b) Shows the experimental data for the 5-qubit Vigo QRC.

Appendix B

Appendix for Chapter 5

B.1 Proof of the UIOS Theorem 5.2

To prove Theorem 5.2, we will apply Lemma B.1 below.

Lemma B.1. *Consider a well-posed system* (5.5)*. Suppose that for* $j = 1, 2$*, there exists* $\tilde{\beta}_j \in \mathcal{KL}$*,* $\tilde{\gamma}_j^u, \tilde{\gamma}_j^y \in \mathcal{K}$ *with* $\tilde{\gamma}_j^y(s) < s$ *for all* $s > 0$ *such that,for some* $M \in \mathbb{Z}$ *with* $M \geq 2$*, for any input* $\Delta u \in l^{\infty}_{n_{u_1}+n_{u_2}}$*, any* $k_0 \in \mathbb{Z}$ *and* $k_1 \in \mathbb{Z}_{\geq 0}$*, and any* $\Delta x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$*,*

$$\|\Delta y_j(k_0 + k_1)\| \leq \max\{\tilde{\beta}_j(\|\Delta x(k_0)\|, k_1), \tilde{\gamma}_j^y(\|\Delta y_{j_{[k_0+\lfloor k_1/M\rfloor, k_0+k_1]}}\|), \tilde{\gamma}_j^u(\|\Delta u_{[k_0,k_0+k_1]}\|)\}, \tag{B.1}$$

and

$$\sup_{k_1 \in \mathbb{Z}_{\geq 0}} \|\Delta y_j(k_0 + k_1)\| < \infty.$$

Then there exists $\hat{\beta}_j \in \mathcal{KL}$ *such that, for all* $k_0 \in \mathbb{Z}$*,* $k_1 \in \mathbb{Z}_{\geq 0}$*,*

$$\|\Delta y_j(k_0 + k_1)\| \leq \max\{\hat{\beta}_j(\|\Delta x(k_0)\|, k_1), \tilde{\gamma}_j^u(\|\Delta u_{[k_0,k_0+k_1]}\|)\}.$$

The main idea in the proof of Lemma B.1 is to apply a continuous extension argument and [140, Lemma A.2] stated below in Lemma B.2.

Lemma B.2. *Given $\delta \in \mathcal{K}_\infty$ and $T : (0,\infty) \times (0,\infty) \to \mathbb{R}_{\geq 0}$ such that, (i) for all $\epsilon > 0$, $s_1 < s_2$ implies $T(\epsilon, s_1) \leq T(\epsilon, s_2)$; (ii) for all $s > 0$, $\lim_{\epsilon \to 0^+} T(\epsilon, s) = \infty$. Then there exists $\hat{\beta} \in \mathcal{KL}$ such that, for each $s > 0$ and $t_1 \in \mathbb{R}_{\geq 0}$, there exists some $\epsilon \in A_{s,t_1} := \{\epsilon' \in (0,\infty) \mid t_1 \geq T(\epsilon', s)\} \cup \{\infty\}$ such that* $\min\{\epsilon, \delta^{-1}(s)\} \leq \hat{\beta}(s, t_1)$.

Proof of Lemma B.1. Fix $k_0 \in \mathbb{Z}$. For any $k_1 \in \mathbb{Z}_{\geq 0}$, define $z_j(k_1) = \|\Delta y_j(k_0 + k_1)\|$ if $\|\Delta y_j(k_0+k_1)\| > \tilde{\gamma}_j^u(\|\Delta u_{[k_0,k_0+k_1]}\|)$ and $z_j(k_1) = 0$ otherwise. From the assumption (B.1) in Lemma B.1, we have that for some $M \in \mathbb{Z}, M \geq 2$,

$$z_j(k_1) \leq \max\{\tilde{\beta}_j(\|\Delta x(k_0)\|, k_1), \tilde{\gamma}_j^y(|z_{j_{[\lfloor k_1/M \rfloor, k_1]}}|)\}. \tag{B.2}$$

We emphasize the implicit dependence $z_j(k_1) = z_j(k_1, x(k_0), \Delta u)$. We sample and hold the left points to extend z_j to a piecewise continuous function w_j. For any $t_1 \in \mathbb{R}_{\geq 0}$, define $w_j(t_1) = \sum_{k_1=0}^{\infty} \mathbb{1}_{[k_1,k_1+1)}(t_1) z_j(k_1)$, where $\mathbb{1}_{[k_1,k_1+1)}(t_1) = 1$ if $t_1 \in [k_1, k_1+1)$ and zero otherwise. For any $\tau, \tau' \in \mathbb{R}$, let $|w_{j_{[\tau,\tau']}}| := \sup_{\tau \leq \overline{\tau} \leq \tau'} |w_j(\overline{\tau})|$. Since $\lfloor t_1 \rfloor \leq t_1 < \lfloor t_1 \rfloor + 1$, we have $w_j(t_1) = z_j(\lfloor t_1 \rfloor) = w_j(\lfloor t_1 \rfloor)$ and

$$|w_{j_{[t_1/M,t_1]}}| = |w_{j_{[\lfloor t_1/M \rfloor, \lfloor t_1 \rfloor]}}| = |w_{j_{[\lfloor \lfloor t_1 \rfloor/M \rfloor, \lfloor t_1 \rfloor]}}| = |z_{j_{[\lfloor \lfloor t_1 \rfloor/M \rfloor, \lfloor t_1 \rfloor]}}|. \tag{B.3}$$

For $s \in \mathbb{R}_{\geq 0}$, let $\beta_j(s, t_1) = \sum_{k_1=0}^{\infty} \mathbb{1}_{[k_1,k_1+1)}(t_1) \tilde{\beta}_j(s, k_1)$. Then $\beta_j \in \mathcal{KL}$ and $\beta_j(s,t_1) = \tilde{\beta}_j(s, \lfloor t_1 \rfloor)$. From (B.2) and (B.3), we have that for all $t_1 \in \mathbb{R}_{\geq 0}$,

$$w_j(t_1) \leq \max\{\beta_j(\|\Delta x(k_0)\|, t_1), \tilde{\gamma}_j^y(|w_{j_{[t_1/M,t_1]}}|)\}. \tag{B.4}$$

To apply Lemma B.2, we first show the following claims.

Claim (i): There exists $\delta \in \mathcal{K}_\infty$ such that for all $t_1 \in \mathbb{R}_{\geq 0}$ and $\Delta x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$, we have $w_j(t_1) \leq \delta^{-1}(\|\Delta x(k_0)\|)$.

Proof of Claim (i). Note that $|w_{j_{[0,\infty)}}| := \sup_{t_1 \in \mathbb{R}_{\geq 0}} w_j(t_1) = \sup_{k_1 \in \mathbb{Z}_{\geq 0}} \|\Delta y_j(k_0 + k_1)\| < \infty$. From (B.4), we have

$$|w_{j_{[0,\infty)}}| \leq \max\{\beta_j(\|\Delta x(k_0)\|, 0), \tilde{\gamma}_j^y(|w_{j_{[0,\infty)}}|)\}.$$

Since $\tilde{\gamma}_j^y(s) < s$ for all $s > 0$, it follows that for any $t_1 \in \mathbb{R}_{\geq 0}$,

$$w_j(t_1) \leq |w_{j_{[0,\infty)}}| \leq \beta_j(\|\Delta x(k_0)\|, 0). \tag{B.5}$$

Choose $\delta \in \mathcal{K}_\infty$ such that $\delta^{-1}(\|\Delta x(k_0)\|) \geq \beta_j(\|\Delta x(k_0)\|, 0)$ (e.g., $\delta^{-1} = id + \beta_j(\cdot, 0)$) gives the desired result. $\square$

Claim (ii): For any $\epsilon, r > 0$, there exists $\hat{T}_{\epsilon,r} \in \mathbb{R}_{\geq 0}$ such that for all $t_1 \geq \hat{T}_{\epsilon,r}$, $w_j(t_1) \leq \epsilon$ whenever $\|\Delta x(k_0)\| \leq r$.

Proof of Claim (ii). The proof uses (B.5) and proceeds as in [138, Lemma 2.1]. Let $\epsilon, r > 0$, if $\beta_j(\|\Delta x(k_0)\|, 0) \leq \beta_j(r, 0) \leq \epsilon$, then by (B.5), $w_j(t_1) \leq \beta_j(\|\Delta x(k_0)\|, 0) \leq \epsilon$ for all $t_1 \in \mathbb{R}_{\geq 0}$. Otherwise, since $\tilde{\gamma}_j^y$ is strictly contractive, there exists $n_{\epsilon,r} \in \mathbb{Z}_{\geq 0}$ such that the $n_{\epsilon,r}$-times composition $(\tilde{\gamma}_j^y)^{(n_{\epsilon,r})}(\beta_j(r, 0)) \leq \epsilon$. For $i = 1, \ldots, n_{\epsilon,r}$, let $\tau_i \in \mathbb{R}_{\geq 0}$ be the first time instance such that $\beta_j(r, \tau_i) \leq (\tilde{\gamma}_j^y)^{(i)}(\beta_j(r, 0))$ so that $\tau_i \leq \tau_j$ for $j = i+1, \ldots, n_{\epsilon,r}$. Define $\hat{\tau}_0 = 0$ and $\hat{\tau}_i = \max\{\tau_i, M\hat{\tau}_{i-1}\}$. We will show by induction that for $t_1 \geq \hat{\tau}_i$, $w_j(t_1) \leq (\tilde{\gamma}_j^y)^{(i)}(\beta_j(r, 0))$.

Claim (i) establishes the case for $i = 0$ (with $(\tilde{\gamma}_j^y)^{(0)} = id$). Suppose the induction hypothesis holds for $t_1 \geq \hat{\tau}_i$. For $t_1 \geq \hat{\tau}_{i+1}$, we have $t_1 \geq \tau_{i+1}$ and $t_1/M \geq \hat{\tau}_i$. From (B.4),

$$\begin{aligned} w_j(t_1) &\leq \max\{\beta_j(\|\Delta x(k_0)\|, \tau_{i+1}), \tilde{\gamma}_j^y \circ (\tilde{\gamma}_j^y)^{(i)}(\beta_j(r, 0))\} \\ &= (\tilde{\gamma}_j^y)^{(i+1)}(\beta_j(r, 0)). \end{aligned}$$

Claim (ii) follows from choosing $\hat{T}_{\epsilon,r} \geq \hat{\tau}_{n_{\epsilon,r}}$. $\square$

Let $\hat{T}_{\epsilon,r}$ be given by Claim (ii). As in [140, Proposition 2.7], define $T(\epsilon, r) = r/\epsilon + \inf\{\hat{T}_{\epsilon',r'} | r \leq r', \epsilon' \in (0, \epsilon]\}$. Then $T(\cdot, \cdot)$ satisfies the conditions in Lemma B.2. Fix $s = \|\Delta x(k_0)\| > 0$ (the case for $s = 0$ is immediate), any $t_1 \in \mathbb{R}_{\geq 0}$ and the set A_{s,t_1}. By Claim (ii), $w_j(t_1) \leq \epsilon$ for all $\epsilon \in A_{s,t_1}$. Let $\hat{\beta}_j \in \mathcal{KL}$ and $\epsilon \in A_{s,t_1}$ be given by Lemma B.2, such that $\min\{\epsilon, \delta^{-1}(\|\Delta x(k_0)\|)\} \leq \hat{\beta}_j(\|\Delta x(k_0)\|, t_1)$. Then

- If $\delta^{-1}(\|\Delta x(k_0)\|) \leq \epsilon$, then by Claim (i) we have $w_j(t_1) \leq \delta^{-1}(\|\Delta x(k_0)\|) \leq \hat{\beta}_j(\|\Delta x(k_0)\|, t_1)$.

- If $\epsilon < \delta^{-1}(\|\Delta x(k_0)\|)$, then $\epsilon < \infty$ and $w_j(t_1) \leq \epsilon \leq \hat{\beta}_j(\|\Delta x(k_0)\|, t_1)$.

Therefore, for all $t_1 \in \mathbb{R}_{\geq 0}$, $w_j(t_1) \leq \hat{\beta}_j(\|\Delta x(k_0)\|, t_1)$. In particular, for all $k_1 \in \mathbb{Z}_{\geq 0}$, $w_j(k_1) = z_j(k_1) \leq \hat{\beta}_j(\|\Delta x(k_0)\|, k_1)$. By definition of $z_j(k_1)$, we have the desired result. $\square$

We now prove Theorem 5.2. The proof adapts [138, Theorem 2.1] to discrete-time systems of the form (5.5). We first show that $\sup_{k \geq k_0} \|\Delta x(k)\| < \infty$ and $\sup_{k \geq k_0} \|\Delta y(k)\| < \infty$, then we apply Lemma B.1 with $M = 4$ to show that system (5.5) is UIOS.

Proof of Theorem 5.2. From (5.11), we have that for all $k \geq k_0$ and $j = 1, 2$,

$$\|\Delta y_{j_{[k_0,k]}}\| \leq \max\{\beta_j(\|\Delta x_j(k_0)\|, 0), \gamma_j^y(\|\Delta v_{j_{[k_0,k]}}\|), \gamma_j^u(\|\Delta u_{j_{[k_0,k]}}\|)\}.$$

Substituting $\Delta v_1 = \Delta y_2$, $\Delta v_2 = \Delta y_1$ and the bound for $\|\Delta y_{2_{[k_0,k]}}\|$ into that of $\|\Delta y_{1_{[k_0,k]}}\|$, we have

$$\begin{aligned} \|\Delta y_{1_{[k_0,k]}}\| &\leq \max\{\beta_1(\|\Delta x_1(k_0)\|, 0), \gamma_1^y \circ \beta_2(\|\Delta x_2(k_0)\|, 0), \\ &\qquad \gamma_1^y \circ \gamma_2^y(\|\Delta y_{1_{[k_0,k]}}\|), \gamma_1^y \circ \gamma_2^u(\|\Delta u_{2_{[k_0,k]}}\|), \gamma_1^u(\|\Delta u_{1_{[k_0,k]}}\|)\} \\ &\leq \max\{\beta_1(\|\Delta x_1(k_0)\|, 0), \gamma_1^y \circ \beta_2(\|\Delta x_2(k_0)\|, 0), \\ &\qquad \gamma_1^y \circ \gamma_2^u(\|\Delta u_{2_{[k_0,k]}}\|), \gamma_1^u(\|\Delta u_{1_{[k_0,k]}}\|)\}, \end{aligned} \tag{B.6}$$

where the last inequality follows from $\gamma_1^y \circ \gamma_2^y(\|\Delta y_{1_{[k_0,k]}}\|) < \|\Delta y_{1_{[k_0,k]}}\|$. A symmetric argument shows that

$$\begin{aligned} \|\Delta y_{2_{[k_0,k]}}\| \leq \max\{&\beta_2(\|\Delta x_2(k_0)\|, 0), \gamma_2^y \circ \beta_1(\|\Delta x_1(k_0)\|, 0), \\ &\gamma_2^y \circ \gamma_1^u(\|\Delta u_{1_{[k_0,k]}}\|), \gamma_2^u(\|\Delta u_{2_{[k_0,k]}}\|)\}. \end{aligned} \tag{B.7}$$

Recall that $\Delta u_j \in l_{n_{u_j}}^{\infty}$ for $j = 1, 2$. From (B.6) and (B.7), $\sup_{k \geq k_0} \|\Delta y_j(k)\| < \infty$. Substituting $\Delta v_1 = \Delta y_2$ and $\Delta v_2 = \Delta y_1$ in (5.12), it follows that $\sup_{k \geq k_0} \|\Delta x_j(k)\| < \infty$. It remains to show that system (5.5) is UIOS.

Upper bound $\|\Delta v_{j_{[k_0,k-1]}}\|$ in (5.12) by (B.6) and (B.7), using $\|\Delta x_j(k_0)\| \le \|\Delta x(k_0)\|$ and $\|\Delta u_{j_{[k_0,k]}}\| \le \|\Delta u_{[k_0,k]}\|$, we have

$$\begin{aligned}\|\Delta x(k)\| &\le 2\max\{\|\Delta x_1(k)\|, \|\Delta x_2(k)\|\} \\ &\le 2\max_{j=1,2}\left\{\sigma_j(\|\Delta x(k_0)\|), \sigma_j^y(\|\Delta v_{j_{[k_0,k]}}\|), \sigma_j^u(\left\|\Delta u_{[k_0,k]}\right\|)\right\} \\ &\le \max\left\{\overline{\sigma}(\|\Delta x(k_0)\|), \overline{\gamma}(\left\|\Delta u_{[k_0,k]}\right\|)\right\},\end{aligned} \tag{B.8}$$

where $\overline{\sigma}(s) = 2\max\{\sigma_1(s), \sigma_2(s), \sigma_1^y(\beta_2(s,0)), \sigma_2^y(\beta_1(s,0)), \sigma_1^y \circ \gamma_2^y(\beta_1(s,0)), \sigma_2^y \circ \gamma_1^y(\beta_2(s,0))\}$ and $\overline{\gamma}(s) = 2\max\{\sigma_1^u(s),\ \sigma_2^u(s), \sigma_1^y \circ \gamma_2^y \circ \gamma_1^u(s), \sigma_2^y \circ \gamma_1^y \circ \gamma_2^u(s), \sigma_1^y \circ \gamma_2^u(s), \sigma_2^y \circ \gamma_1^u(s)\}$.

Consider subsystem $j = 1$ and (5.11). For any $k_1 \in \mathbb{Z}_{\ge 0}$ and $k_0 \in \mathbb{Z}$, let $k_0 + \lceil k_1/2 \rceil$ be the initial time and $k = k_0 + k_1$. Then $k - (k_0 + \lceil k_1/2 \rceil) = \lfloor k_1/2 \rfloor$ and

$$\begin{aligned}\|\Delta y_1(k_0 + k_1)\| \le {}& \max\{\beta_1(\|\Delta x_1(k_0 + \lceil k_1/2 \rceil)\|, \lfloor k_1/2 \rfloor), \\ &\quad \gamma_1^y(\|\Delta y_{2_{[k_0+\lceil k_1/2 \rceil, k_0+k_1]}}\|), \gamma_1^u(\|\Delta u_{1_{[k_0+\lceil k_1/2 \rceil, k_0+k_1]}}\|)\} \\ \le {}& \max\{\beta_1(\|\Delta x(k_0 + \lceil k_1/2 \rceil)\|, \lfloor k_1/2 \rfloor), \\ &\quad \gamma_1^y(\|\Delta y_{2_{[k_0+\lceil k_1/2 \rceil, k_0+k_1]}}\|), \gamma_1^u(\|\Delta u_{[k_0, k_0+k_1]}\|)\}.\end{aligned} \tag{B.9}$$

Consider subsystem $j = 2$ and (5.11). Let $k_0 + \lfloor k_1/4 \rfloor$ be the initial time. For any $\lceil k_1/2 \rceil \le \overline{k}_1 \le k_1$, let $k = k_0 + \overline{k}_1$. Then $k - (k_0 + \lfloor k_1/4 \rfloor) = \overline{k}_1 - \lfloor k_1/4 \rfloor \ge \lceil k_1/2 \rceil - \lfloor k_1/4 \rfloor \ge \lceil k_1/4 \rceil$,

$$\begin{aligned}\left\|\Delta y_2(k_0 + \overline{k}_1)\right\| \le {}& \max\{\beta_2(\|\Delta x_2(k_0 + \lfloor k_1/4 \rfloor)\|, \overline{k}_1 - \lfloor k_1/4 \rfloor), \\ &\quad \gamma_2^y(\|\Delta y_{1_{[k_0+\lfloor k_1/4 \rfloor, k_0+\overline{k}_1]}}\|), \gamma_2^u(\|\Delta u_{2_{[k_0+\lfloor k_1/4 \rfloor, k_0+\overline{k}_1]}}\|)\} \\ \le {}& \max\{\beta_2(\|\Delta x(k_0 + \lfloor k_1/4 \rfloor)\|, \lceil k_1/4 \rceil), \\ &\quad \gamma_2^y(\|\Delta y_{1_{[k_0+\lfloor k_1/4 \rfloor, k_0+k_1]}}\|), \gamma_2^u(\|\Delta u_{[k_0, k_0+k_1]}\|)\}\end{aligned} \tag{B.10}$$

Note that the right-hand side of the last inequality in (B.10) does not depend on $\overline{k}_1$. Now taking $\sup_{\lceil k_1/2 \rceil \le \overline{k}_1 \le k_1}$ on both sides of (B.10) shows that

$$\begin{aligned}\|\Delta y_{2_{[k_0+\lceil k_1/2 \rceil, k_0+k_1]}}\| \le {}& \max\{\beta_2(\|\Delta x(k_0 + \lfloor k_1/4 \rfloor)\|, \lceil k_1/4 \rceil), \\ &\quad \gamma_2^y(\|\Delta y_{1_{[k_0+\lfloor k_1/4 \rfloor, k_0+k_1]}}\|), \gamma_2^u(\|\Delta u_{[k_0, k_0+k_1]}\|)\}.\end{aligned} \tag{B.11}$$

Upper bounding $\|\Delta y_{2_{[k_0+\lceil k_1/2 \rceil, k_0+k_1]}}\|$ in (B.9) using (B.11) and upper bounding $\|\Delta x(k_0 +$

$\lceil k_1/2 \rceil)\|$ in (B.9) using (B.8), we have

$$\begin{aligned}\|\Delta y_1(k_0+k_1)\| \leq \max\{&\beta_1(\overline{\sigma}(\|\Delta x(k_0)\|), \lfloor k_1/2 \rfloor),\\ &\gamma_1^y \circ \beta_2(\overline{\sigma}(\|\Delta x(k_0)\|), \lceil k_1/4 \rceil),\\ &\gamma_1^y \circ \gamma_2^y(\|\Delta y_{1[k_0+\lfloor k_1/4 \rfloor, k_0+k_1]}\|), \gamma_1^y \circ \gamma_2^u(\|\Delta u_{[k_0,k_0+k_1]}\|),\\ &\gamma_1^u(\|\Delta u_{[k_0,k_0+k_1]}\|), \beta_1(\overline{\gamma}(\|\Delta u_{[k_0,k_0+k_1]}\|), 0),\\ &\gamma_1^y \circ \beta_2(\overline{\gamma}(\|\Delta u_{[k_0,k_0+k_1]}\|), 0)\}.\end{aligned} \tag{B.12}$$

Let $\tilde{\gamma}_1^y(s) = \gamma_1^y \circ \gamma_2^y(s)$, $\tilde{\beta}_1(s, k_1) = \max\{\beta_1(\overline{\sigma}(s), \lfloor k_1/4 \rfloor), \gamma_1^y \circ \beta_2(\overline{\sigma}(s), \lfloor k_1/4 \rfloor)\}$ and $\tilde{\gamma}_1^u(s) = \max\{\gamma_1^y \circ \gamma_2^u(s), \gamma_1^u(s),\ \beta_1(\overline{\gamma}(s), 0), \gamma_1^y \circ \beta_2(\overline{\gamma}(s), 0)\}$. By a symmetric argument, we can define $\tilde{\beta}_2$ and $\tilde{\gamma}_2^u$ analogously. Here, $\tilde{\beta}_j \in \mathcal{KL}$ and $\tilde{\gamma}_j^u \in \mathcal{K}$ for $j = 1, 2$. Re-writing (B.12) in terms of $\tilde{\beta}_j$ and $\tilde{\gamma}_j$, we have

$$\|y_j(k_0+k_1)\| \leq \max\{\tilde{\beta}_j(\|x(k_0)\|, k_1), \tilde{\gamma}_j^y(\|y_{j[k_0+\lfloor k_1/4 \rfloor, k_0+k_1]}\|), \tilde{\gamma}_j^u(\|u_{[k_0,k_0+k_1]}\|)\},$$

where $\tilde{\gamma}_j^y(s) < s$ for all $s > 0$ by strict contractivity of $\gamma_1^y \circ \gamma_2^y(\cdot)$ and $\gamma_2^y \circ \gamma_1^y(\cdot)$. Recall that we have shown $\sup_{k_1 \in \mathbb{Z}_{\geq 0}} \|\Delta y_j(k_0+k_1)\| < \infty$. Invoking Lemma B.1 with $M = 4$, there exists $\hat{\beta}_j \in \mathcal{KL}$ such that for all $k_0 \in \mathbb{Z}$ and $k_1 \in \mathbb{Z}_{\geq 0}$,

$$\|\Delta y_j(k_0+k_1)\| \leq \max\{\hat{\beta}_j(\|\Delta x(k_0)\|, k_1), \tilde{\gamma}_j^u(\|\Delta u_{[k_0,k_0+k_1]}\|)\}.$$

Let $k = k_0 + k_1$, $\gamma(s) = 2\max_{j=1,2}\{\tilde{\gamma}_j^u(s)\}$ and $\beta(s,k) = 2\max_{j=1,2}\{\tilde{\beta}_j(s,k)\}$. It follows that

$$\|\Delta y(k)\| \leq 2\max\{\|\Delta y_1(k)\|, \|\Delta y_2(k)\|\} \leq \max\{\beta(\|\Delta x(k_0)\|, k - k_0), \gamma(\|\Delta u_{[k_0,k]}\|)\}$$

for all $k, k_0 \in \mathbb{Z}$ with $k \geq k_0$ and any initial condition $\Delta x(k_0) \in \mathbb{R}^{n_{x_1}+n_{x_2}}$. ☐

Back to Sec. 5.2.

B.2 Uniform convergence of Lur'e system

We show that the observer-based feedback-controlled Lur'e system (5.17) is uniformly convergent. The design is based on a convergence approach, which has been applied to

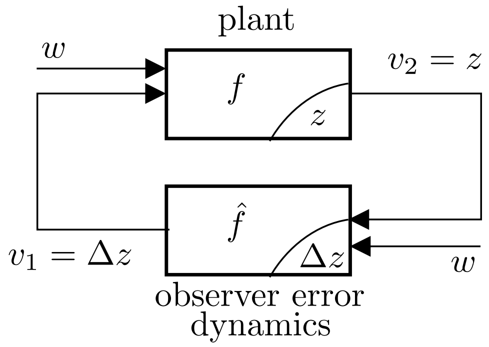


Figure B.1: Schematic of the closed-loop system (B.13) consisting of the plant and the observer error dynamics.

observer-based control in continuous-time [133]. To illustrate this for a more general class of nonlinear systems, we consider a nonlinear plant

$$z(k+1) = f(z(k), u(k), w(k)), \quad y(k) = h(z(k)),$$

with state $z(k) \in \mathbb{R}^{n_z}$, control $u(k) \in \mathbb{R}^{n_u}$, external input $w \in l^\infty_{n_w}$ and output $y(k) \in \mathbb{R}^{n_y}$. Construct an observer,

$$\hat{z}(k+1) = f(\hat{z}(k), u(k), w(k)) - L(h(\hat{z}(k)) - y(k)),$$

where $u(k) = \phi(\hat{z}(k))$ and $L(s) = 0$ if $s = 0$. In general, L can be a nonlinear function. Let $\Delta z(k) = \hat{z}(k) - z(k)$, then $u(k) = \phi(\Delta z(k) + z(k))$ and the observer error dynamics is

$$\begin{aligned} \Delta z(k+1) &= \hat{f}(\Delta z(k), z(k), w(k)) \\ &= f(\Delta z(k) + z(k), u(k), w(k)) \\ &\qquad - f(z(k), u(k), w(k)) - L(h(\Delta z(k) + z(k)) - h(z(k))), \end{aligned}$$

where $z(k), w(k)$ are viewed as inputs to the error dynamics. Consider the interconnected system (B.13) (see Fig. B.1),

$$\begin{cases} z(k+1) = f(z(k), \phi(v_1(k) + z(k)), w(k)), \\ \Delta z(k+1) = \hat{f}(\Delta z(k), v_2(k), w(k)), \end{cases} \tag{B.13}$$

with interconnections $v_1(k) = \Delta z(k)$ and $v_2(k) = z(k)$. Here, w is the input. Our goal is to employ Theorem 5.1 to design the observer gain $L(\cdot)$ and the controller $u(k) = \phi(\hat{z}(k))$ such that the above closed-loop system is UISC and hence UC. To this end, we consider the following Corollary.

Corollary B.1. *Consider a well-posed system* (B.13)*. Suppose that for any inputs* v_2, w *with* $w \in l^{\infty}_{n_w}$*, there exists* $\beta_2 \in \mathcal{KL}$ *such that, for any* $k, k_0 \in \mathbb{Z}, k \geq k_0$ *and* $\Delta z(k_0) \in \mathbb{R}^{n_{\Delta z}}$*,*

$$\|\Delta z(k)\| \leq \beta_2(\|\Delta z(k_0)\|, k - k_0). \tag{B.14}$$

Suppose that the z*-subsystem is UC and let* z^* *be the reference solution to* v_1, w*. For any other input* $\overline{v}_1, \overline{w}$ *with* $\overline{w} \in l^{\infty}_{n_w}$*, let* $\overline{z}$ *be any solution. Suppose that there exists* $\beta_1 \in \mathcal{KL}, \gamma_1^y, \gamma_1^w \in \mathcal{K}$ *such that for any* $k_0, k \in \mathbb{Z}, k \geq k_0$ *and* $\overline{z}(k_0) \in \mathbb{R}^{n_z}$*,*

$$\begin{aligned} \|z^*(k) - \overline{z}(k)\| \leq{}& \beta_1(\|z^*(k_0) - \overline{z}(k_0)\|, k - k_0) \\ &+ \gamma_1^y(\|(v_1 - \overline{v}_1)_{[k_0,k-1]}\|) + \gamma_1^w(\|(w - \overline{w})_{[k_0,k-1]}\|). \end{aligned} \tag{B.15}$$

It follows that system (B.13) *is UISC.*

Proof. By [22, Remark 5], for any inputs v_2, w, $\Delta z_2^* = 0$ is the unique and bounded reference state solution, so that the Δz-subsystem is UC. To show closed-loop UISC, let $\Delta\overline{z}(k)$ be any solution to inputs $\overline{v}_2, \overline{w}$ with initial condition $\overline{z}(k_0)$. Equation (B.14) implies that for any gains $\gamma_2^y, \gamma_2^w \in \mathcal{K}$,

$$\begin{aligned} \|\Delta\overline{z}(k)\| &= \|\Delta z^*(k) - \Delta\overline{z}(k)\| \\ &\leq \beta_2(\|\Delta z^*(k_0) - \Delta\overline{z}(k_0)\|, k - k_0) + \gamma_2^y(\|(v_2 - \overline{v}_2)_{[k_0,k-1]}\|) \\ &\quad + \gamma_2^u(\|(w - \overline{w})_{[k_0,k-1]}\|)\}. \end{aligned}$$

The above equation and (B.15) shows that (5.16) in Corollary 5.2 hold. We now show that (5.15) in Corollary 5.2 is satisfied and hence the closed-loop UISC of system (B.13). Observe that $id + \gamma_1^y \in \mathcal{K}_\infty$ with $(id + \gamma_1^y)^{-1} \circ \gamma_1^y(s) < s$ for all $s > 0$. Let $\lambda_1, \lambda_2 \in \mathcal{K}_\infty$ be arbitrary. The result follows from choosing $\gamma_2^y(s) = ((id + \lambda_1) \circ (id + \gamma_1^y) \circ (id + \lambda_2))^{-1}(s)$. □

For the Lur'e system (5.17), the observer error dynamics $\Delta z(k) = \hat{z}(k) - z(k)$ satisfies

(B.14) if there exists $Z \in \mathbb{R}^2$, $P \succ 0$, $\epsilon > 0$ and $\theta \in (0,1)$ such that

$$\begin{gathered}(P - \epsilon I) \prec 0, \\ \begin{bmatrix} -\theta P & A^\top P - C^\top Z^\top & \epsilon\rho(GH)^\top & A^\top P - C^\top Z^\top \\ PA - ZC & -P & 0 & 0 \\ \epsilon\rho GH & 0 & -\epsilon I & 0 \\ PA - ZC & 0 & 0 & P - \epsilon I \end{bmatrix} \preceq 0.\end{gathered} \tag{B.16}$$

This can be shown by applying the Schur complement in essentially the same way as in Sec. 6.3.2.

Under a linear state-feedback law $u(k) = -K\hat{z}(k)$ with gain $K^\top \in \mathbb{R}^2$, the Lur'e system becomes

$$\begin{aligned} z(k+1) &= (A - B_uK)z(k) + \rho G \sin(Hz(k)) - B_uK\Delta z(k) + B_w w(k) \\ &:= \tilde{f}(z(k), \Delta z(k), w(k)), \end{aligned} \tag{B.17}$$

where $\Delta z(k), w(k)$ are viewed as inputs. Let $z(k), \overline{z}(k)$ be any solutions starting at $z(k_0), \overline{z}(k_0)$ to inputs $\Delta z(k), w$ and $\Delta\overline{z}(k), \overline{w}$, respectively. Let $\delta\overline{z}(k) = z(k) - \overline{z}(k)$, then

$$\begin{aligned} \|\delta\overline{z}(k)\| \leq{}& \lambda_s \|\delta\overline{z}(k-1)\| + \sigma_{\max}(B_uK)\|\Delta z(k-1) - \Delta\overline{z}(k-1)\| \\ &+ \|B_w\||w(k-1) - \overline{w}(k-1)|, \end{aligned} \tag{B.18}$$

where $\lambda_s = \sigma_{\max}(A - B_uK) + \rho\sigma_{\max}(GH)$. We employ Theorem 2.1 to show that the plant (B.17) is UC. Firstly, consider $\Delta z = \Delta\overline{z}, w = \overline{w}$. From (B.18), we have $\|\delta\overline{z}(k)\| \leq \lambda_s\|\delta\overline{z}(k-1)\|$. Furthermore, note that for any $\Delta z \in l_2^\infty$ and $w \in l_1^\infty$, we have

$$\sup_{k\in\mathbb{Z}} \|\tilde{f}(0, \Delta z(k), w(k))\| \leq \|B_w\|\|w\|_\infty + \sigma_{\max}(B_uK)\|\Delta z\|_\infty < \infty.$$

By Theorem 2.1, if there exists K such that $\lambda_s < 1$, then the plant (B.17) is UC. Further, the condition $\lambda_s < 1$ also ensures that the plant satisfies (B.15) in Corollary B.1. Finally, applying Corollary B.1 shows that the closed-loop system (B.13) is UISC.

In this example, we can choose $L^\top = \begin{bmatrix} 2.3258 & 2.1104 \end{bmatrix}$ and $K = \begin{bmatrix} 0.4956 & 1.006 \end{bmatrix}$, then $\lambda_s = 0.8687$ and the linear matrix inequalities (B.16) hold for $\theta = \epsilon = 0.001$. Hence, the closed-loop system (B.13) is UISC and hence UC.

Back to Sec. 5.3.

B.3 Trace inequality

Lemma B.3. *For any $n\times n$ Hermitian matrices A and B, we have* $|\mathrm{Tr}(AB)| \leq \sigma_{\max}(A)\|B\|_1$.

Proof. Let $B = \sum_{j=1}^{n} \lambda_j v_j v_j^\dagger$ be the spectral decomposition of B, where $\{v_j\}$ forms an orthonormal basis for $\mathbb{R}^n$, where $\dagger$ is the adjoint. Let $\{e_j\}$ be the standard basis for $\mathbb{R}^n$. Then there exists a unitary matrix U such that $v_j = Ue_j$ for $1 \leq j \leq n$. Therefore,

$$|\mathrm{Tr}(AB)| = \left|\sum_{j=1}^{n} \lambda_j \mathrm{Tr}\left(e_j^\dagger U^\dagger A U e_j\right)\right| \leq \sum_{j=1}^{n} |\lambda_j| \left|(U^\dagger A U)_{jj}\right|,$$

where X_{jj} is the (j,j)-th element of a matrix X. For any Hermitian matrix A, by the min-max theorem, we have $\lambda_{\min}(A) = \min_{\sqrt{x^\top x}=1} x^\top Ax \leq A_{jj} \leq \lambda_{\max}(A) = \max_{\sqrt{x^\top x}=1} x^\top Ax$. Therefore, $|A_{jj}| \leq \max\{|\lambda_{\min}(A)|, |\lambda_{\max}(A)|\} = \sigma_{\max}(A)$. By unitary invariance of singular values, $|\mathrm{Tr}(AB)| \leq \sigma_{\max}(A)\sum_{j=1}^{n}|\lambda_j| = \sigma_{\max}(A)\|B\|_1$. □

Back to Sec. 5.3.2.

B.4 Noisy QRC simulation

We simulate decoherence by applying the Kraus operators of dephasing and GAD channels for all qubits, followed by a multi-qubit error modeled as a depolarizing channel $T_p(\rho) = (1-p_p)\rho + p_p I_{2^N}/2^N$ acting on all N qubits [116], where I_{2^N} is the $2^N \times 2^N$ identity matrix. Noise strength p_p is set to be typical controlled-not gate errors on superconducting quantum devices [44]. Typically, $p_p = 10^{-3} \sim 10^{-2}$ and we set $p_p = 10^{-2}$. The dephasing channel's Kraus operators are $M_0(\psi) = \sqrt{\frac{1+e^{-2\psi}}{2}} I_2, M_1(\psi) = \sqrt{\frac{1-e^{-2\psi}}{2}} Z$, where ψ is the decoherence strength and Z is the Pauli-Z operator [43]. GAD channels are further parametrized by a finite temperature parameter λ, with Kraus operators $M_1(\psi,\lambda) = \sqrt{\lambda}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, M_2(\psi,\lambda) = \sqrt{\lambda}\begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}, M_3(\psi,\lambda) = \sqrt{1-\lambda}\begin{bmatrix} \sqrt{1-p} & 0 \\ 0 & 1 \end{bmatrix}, M_4(\psi,\lambda) =$

$\sqrt{1-\lambda}\begin{bmatrix} 0 & 0 \\ \sqrt{p} & 0 \end{bmatrix}$, where $\sqrt{1-p} = e^{-2\psi}$ and $\sqrt{p} = \sqrt{1-e^{-4\psi}}$ [43]. We remark that some NISQ machines can achieve a nominal $\psi = 10^{-3}$ [44].

Back to Sec. 5.3.2.

Appendix C

Appendix for Chapter 6

C.1 Unitary gates for QRCs

We provide details on the unitaries $U_l(\gamma_l)$ employed in Sec. 6.4 for the QRCs. In Subsecs. 6.4.1 and 6.4.2, QRCs only process y_k and are governed by two unitaries. We set $U_0(\gamma_0) = \prod_{(m_c,m_t)\in E} C_Y^{m_c,m_t}(\gamma_0^m)$, a sequence of controlled rotational-Y gates $C_Y^{m_c,m_t}(\gamma_0^m)$. Here, m_c, m_t denote the control and target qubits for the m-th qubit pair, and γ_0^m is the rotational angle. On some NISQ devices [44], $C_Y^{m_c,m_t}(\gamma_0^m)$ is implemented by decomposing it as a sequence of controlled-NOT and single-qubit rotational-Y gates. The coupling map E denotes the set of qubit pairs on which a controlled-NOT gate can be directly applied without further decomposition. We set E to be the coupling map of the IBM Ourense device [44]. We set $U_1(\gamma_1) = \prod_{m=1}^N R_X^m(\pi)$, where $R_X^m(\pi)$ is the rotational-X gate on qubit m with angle π and N is the number of qubits.

In Subsec. 6.4.3, we employ the spatial multiplexing technique. The first QRC member processes both input $u_k \in \mathbb{R}^n$ and output $y_k \in \mathbb{R}$, and it is governed by $n+2$ unitaries. For this QRC member, we have $U_l^{(1)}(\gamma_l) = \prod_{(m_c,m_t)\in E} C_Y^{m_c,m_t}(\gamma_l^m)$ for $j = 1, \ldots, n+1$. We set $U_{n+2}^{(1)}(\gamma_{n+2}) = \prod_{m=1}^N R_X^m(\pi)$. The second QRC only processes input u_k, and it is governed by $n+1$ unitaries. We set $U_l^{(2)}(\overline{\gamma}_l) = \prod_{(m_c,m_t)\in E} C_Y^{m_c,m_t}(\overline{\gamma}_j^m)$ for $l = 1, \ldots n$ and

$U^{(2)}_{n+1}(\overline{\gamma}_{n+1}) = \prod_{m=1}^{N} R_X^m(\pi)$. For all unitaries, their angles $\gamma_l^m, \overline{\gamma}_l^m$ are uniformly randomly sampled from $[-\pi, \pi]$, independently for each m and $l = 1, \dots, n+1$.

Back to Sec. 6.4.

C.2 Parameters of selected RCs

We detail all parameters of the selected RCs in Sec. 6.4. The parameters for the selected ESNs and Lur'e systems are summarized in Figs. C.1 and C.2, respectively. Note that for the coupled electric drive system, A_1, B_1, C_1 are the system matrices for the first ESN or Lur'e member in their multiplexed configuration. Matrices A_2, B_2 describe the second ESN or Lur'e member.

For the nonlinear quantum optics time series, we have $\gamma_0^1 = -0.15$ and fix $\gamma_1^j = \pi$ for $j = 1, 2$. For the finance time series, $\gamma_0^1 = -1.81$ and $\gamma_1^j = \pi$ for $j = 1, 2$. For modeling the coupled electric drive system, all selected multiplexed QRCs are 2-qubit, i.e., each QRC member in the multiplexed QRC has 2 qubits. We set $\gamma_0^1 = 2.72, \gamma_1^1 = -0.25$, $\overline{\gamma}_0^1 = 0.84$ and $\gamma_1^j = \overline{\gamma}_2^j = \pi$ for $j = 1, 2$.

Optics		Finance	
A	$B^\top$	A	$B^\top$
$\begin{bmatrix} -0.09 & -0.01 \\ 0.03 & -0.07 \end{bmatrix}$	$\begin{bmatrix} 0.75 & -0.88 \end{bmatrix}^\top$	$\begin{bmatrix} -0.56 & -0.55 \\ 0.30 & 0.33 \end{bmatrix}$	$\begin{bmatrix} -0.37 & 0.44 \end{bmatrix}^\top$

Coupled electric drive system				
A_1	B_1	C_1	A_2	B_2
$\begin{bmatrix} -0.47 & 0.52 \\ -0.08 & -0.14 \end{bmatrix}$	$\begin{bmatrix} -0.58 \\ -0.35 \end{bmatrix}$	$\begin{bmatrix} -0.84 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} -0.16 & -0.43 \\ -0.18 & -0.50 \end{bmatrix}$	$\begin{bmatrix} 0.66 \\ -0.39 \end{bmatrix}$

Figure C.1: Parameters of the selected ESNs.

Optics		Finance	
A	$B^\top$	A	$B^\top$
$\begin{bmatrix} -0.49 & 0.18 \\ -0.10 & 0.85 \end{bmatrix}$	$\begin{bmatrix} 0.35 & -0.44 \end{bmatrix}^\top$	$\begin{bmatrix} 0.33 & -0.50 \\ 0.30 & -0.61 \end{bmatrix}$	$\begin{bmatrix} -0.86 & 0.12 \end{bmatrix}^\top$

Coupled electric drive system				
A_1	B_1	C_1	A_2	B_2
$\begin{bmatrix} -0.44 & -0.45 \\ 0.26 & 0.19 \end{bmatrix}$	$\begin{bmatrix} 0.64 \\ 0.70 \end{bmatrix}$	$\begin{bmatrix} 0.32 \\ -0.22 \end{bmatrix}$	$\begin{bmatrix} 0.51 & -0.081 \\ 0.47 & -0.087 \end{bmatrix}$	$\begin{bmatrix} -0.015 \\ 0.17 \end{bmatrix}$

Figure C.2: Parameters of the selected Lur'e systems.

Back to Sec. 6.4.

References

[1] H. Zen and H. Sak, “Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 4470–4474.

[2] H. Sak, A. Senior, K. Rao, F. Beaufays, and J. Schalkwyk, “Google voice search: faster and more accurate,” *Google Research blog*, 2015. [Online]. Available: https://ai.googleblog.com/2015/09/google-voice-search-faster-and-more.html

[3] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[4] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: Current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

[5] C. Edwards, “Growing pains for deep learning,” *Communications of the ACM*, vol. 58, no. 7, pp. 14–16, 2015.

[6] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, “A survey on deep learning: Algorithms, techniques, and applications,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–36, 2018.

[7] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning.* PMLR, 2013, pp. 1310–1318.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning.* MIT press, 2016.

[9] A. M. Zador, "A critique of pure learning and what artificial neural networks can learn from animal brains," *Nature communications*, vol. 10, no. 1, pp. 1–7, 2019.

[10] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *science*, vol. 304, no. 5667, pp. 78–80, 2004.

[11] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009.

[12] M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, "The importance of mixed selectivity in complex cognitive tasks," *Nature*, vol. 497, no. 7451, pp. 585–590, 2013.

[13] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature communications*, vol. 2, no. 1, pp. 1–6, 2011.

[14] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," *Physical review letters*, vol. 120, no. 2, p. 024102, 2018.

[15] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, pp. 2531–2560, 2002.

[16] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima *et al.*, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, no. 7664, pp. 428–431, 2017.

[17] L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification," *Physical Review X*, vol. 7, no. 1, p. 011015, 2017.

[18] D. Canaday, A. Griffith, and D. J. Gauthier, "Rapid time series prediction with a hardware-based reservoir computer," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 12, p. 123119, 2018.

[19] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 2019.

[20] F. Galán-Prado, J. Font-Rosselló, and J. L. Rosselló, "Tropical reservoir computing hardware," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2712–2716, 2020.

[21] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Scientific reports*, vol. 2, no. 1, pp. 1–7, 2012.

[22] D. N. Tran, B. S. Rüffer, and C. M. Kellett, "Convergence properties for discrete-time nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 8, pp. 3415–3422, 2018.

[23] A. Pavlov and N. van de Wouw, "Convergent discrete-time nonlinear systems: the case of PWA systems," in *2008 American Control Conference*. IEEE, 2008, pp. 3452–3457.

[24] I. B. Yildiz, H. Jaeger, and S. J. Kiebel, "Re-visiting the echo state property," *Neural networks*, vol. 35, pp. 1–9, 2012.

[25] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with volterra series," *IEEE Transactions on circuits and systems*, vol. 32, no. 11, pp. 1150–1161, 1985.

[26] L. Grigoryeva and J.-P. Ortega, “Universal discrete-time reservoir computers with stochastic inputs and linear readouts using non-homogeneous state-affine systems,” *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 892–931, 2018.

[27] ——, “Echo state networks are universal,” *Neural Networks*, vol. 108, pp. 495–508, 2018.

[28] B. Schrauwen, M. D’Haene, D. Verstraeten, and J. Van Campenhout, “Compact hardware liquid state machines on FPGA for real-time speech recognition,” *Neural networks*, vol. 21, no. 2-3, pp. 511–523, 2008.

[29] M. L. Alomar, V. Canals, A. Morro, A. Oliver, and J. L. Rossello, “Stochastic hardware implementation of liquid state machines,” in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 1128–1133.

[30] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain *et al.*, “Loihi: A neuromorphic manycore processor with on-chip learning,” *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.

[31] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura *et al.*, “A million spiking-neuron integrated circuit with a scalable communication network and interface,” *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[32] P. Shor, “Algorithms for quantum computation: discrete logarithms and factoring,” in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134.

[33] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, 1996, pp. 212–219.

[34] A. W. Harrow, A. Hassidim, and S. Lloyd, “Quantum algorithm for linear systems of equations,” *Physical review letters*, vol. 103, no. 15, p. 150502, 2009.

[35] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[36] J. Preskill, “Quantum computing in the nisq era and beyond,” *Quantum*, vol. 2, p. 79, 2018.

[37] C. Figgatt, D. Maslov, K. A. Landsman, N. M. Linke, S. Debnath, and C. Monroe, “Complete 3-qubit Grover search on a programmable quantum computer,” *Nature communications*, vol. 8, no. 1, pp. 1–9, 2017.

[38] M. Mills, “Hearing aids and the history of electronics miniaturization,” *IEEE Annals of the History of Computing*, vol. 33, no. 2, pp. 24–45, 2011.

[39] S. Aaronson and A. Arkhipov, “The computational complexity of linear optics,” in *Proceedings of the forty-third annual ACM symposium on Theory of Computing*, 2011, pp. 333–342.

[40] A. P. Lund, M. J. Bremner, and T. C. Ralph, “Quantum sampling problems, bosonsampling and quantum supremacy,” *npj Quantum Information*, vol. 3, no. 1, pp. 1–8, 2017.

[41] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, “Characterizing quantum supremacy in near-term devices,” *Nature Physics*, vol. 14, no. 6, pp. 595–600, 2018.

[42] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[43] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, ser. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, Cambridge, 2000.

[44] IBM, *IBM Quantum Experience*, 2022. [Online]. Available: https://www.ibm.com/quantum-computing/

[45] Rigetti, *Rigetti Computing*, 2022. [Online]. Available: https://www.rigetti.com/

[46] IonQ, *IonQ*, 2022. [Online]. Available: https://ionq.com/

[47] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, no. 7747, p. 209, 2019.

[48] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” *Nature*, vol. 549, no. 7671, p. 242, 2017.

[49] A. B. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, “Quantum annealing: A new method for minimizing multidimensional functions,” *Chemical physics letters*, vol. 219, no. 5-6, pp. 343–348, 1994.

[50] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer, “Evidence for quantum annealing with more than one hundred qubits,” *Nature physics*, vol. 10, no. 3, pp. 218–224, 2014.

[51] S. W. Shin, G. Smith, J. A. Smolin, and U. Vazirani, “How "quantum" is the D-Wave machine?” *arXiv preprint arXiv:1401.7087*, 2014.

[52] F. Verstraete, M. M. Wolf, and J. I. Cirac, “Quantum computation and quantum-state engineering driven by dissipation,” *Nature physics*, vol. 5, no. 9, pp. 633–636, 2009.

[53] U. Alvarez-Rodriguez, L. Lamata, P. Escandell-Montero, J. D. Martín-Guerrero, and E. Solano, “Supervised quantum learning without measurements,” *Scientific reports*, vol. 7, no. 1, p. 13645, 2017.

[54] K. Fujii and K. Nakajima, “Harnessing disordered-ensemble quantum dynamics for machine learning,” *Phys. Rev. Appl.*, vol. 8, no. 2, p. 024030, 2017.

[55] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, “Boosting computational power through spatial multiplexing in quantum reservoir computing,” *Physical Review Applied*, vol. 11, no. 3, p. 034021, 2019.

[56] M. Negoro, K. Mitarai, K. Nakajima, and K. Fujii, "Toward NMR quantum reservoir computing," in *Reservoir Computing.* Springer, 2021, pp. 451–458.

[57] H. F. Trotter, "On the product of semi-groups of operators," *Proceedings of the American Mathematical Society*, vol. 10, no. 4, pp. 545–551, 1959.

[58] Y. Suzuki, Q. Gao, K. Pradel, K. Yasuoka, and N. Yamamoto, "Natural quantum reservoir computing for temporal information processing," *arXiv preprint arXiv:2107.05808*, 2021.

[59] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, "Gaussian states of continuous-variable quantum systems provide universal and versatile reservoir computing," *Communications Physics*, vol. 4, no. 1, pp. 1–11, 2021.

[60] J. Nokkala, R. Martínez-Peña, R. Zambrini, and M. C. Soriano, "High-performance reservoir computing with fluctuations in linear networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[61] R. A. Bravo, K. Najafi, X. Gao, and S. F. Yelin, "Quantum reservoir computing using arrays of Rydberg atoms," *arXiv preprint arXiv:2111.10956*, 2021.

[62] L. Govia, G. Ribeill, G. Rowlands, H. Krovi, and T. Ohki, "Quantum reservoir computing with a single nonlinear oscillator," *Physical Review Research*, vol. 3, no. 1, p. 013077, 2021.

[63] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, "Information processing capacity of spin-based quantum reservoir computing systems," *Cognitive Computation*, pp. 1–12, 2020.

[64] S. Ghosh, A. Opala, M. Matuszewski, T. Paterek, and T. C. Liew, "Quantum reservoir processing," *npj Quantum Information*, vol. 5, no. 1, pp. 1–6, 2019.

[65] S. Ghosh, T. Paterek, and T. C. Liew, "Quantum neuromorphic platform for quantum state preparation," *Physical review letters*, vol. 123, no. 26, p. 260404, 2019.

[66] S. Ghosh, T. Krisnanda, T. Paterek, and T. C. Liew, “Realising and compressing quantum circuits with quantum reservoir computing,” *Communications Physics*, vol. 4, no. 1, pp. 1–7, 2021.

[67] G. Angelatos, S. A. Khan, and H. E. Türeci, “Reservoir computing approach to quantum state measurement,” *Physical Review X*, vol. 11, no. 4, p. 041062, 2021.

[68] S. A. Khan, F. Hu, G. Angelatos, and H. E. Türeci, “Physical reservoir computing using finitely-sampled quantum systems,” *arXiv preprint arXiv:2110.13849*, 2021.

[69] D. Marković and J. Grollier, “Quantum neuromorphic computing,” *Applied Physics Letters*, vol. 117, no. 15, p. 150501, 2020.

[70] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, “Opportunities in quantum reservoir computing and extreme learning machines,” *arXiv preprint arXiv:2102.11831*, 2021.

[71] C. Gallicchio, A. Micheli, and L. Pedrelli, “Deep reservoir computing: A critical experimental analysis,” *Neurocomputing*, vol. 268, pp. 87–99, 2017.

[72] T. Akiyama and G. Tanaka, “Analysis on characteristics of multi-step learning echo state networks for nonlinear time series prediction,” in *2019 International Joint Conference on Neural Networks (IJCNN).* IEEE, 2019, pp. 1–8.

[73] Z. K. Malik, A. Hussain, and Q. J. Wu, “Multilayered echo state machine: A novel architecture and algorithm,” *IEEE Transactions on cybernetics*, vol. 47, no. 4, pp. 946–959, 2016.

[74] P. Antonik, “Photonic reservoir computer with output feedback,” in *Application of FPGA to Real-Time Machine Learning.* Springer, 2018, pp. 91–121.

[75] C. Gallicchio and A. Micheli, “Echo state property of deep reservoir computing networks,” *Cognitive Computation*, vol. 9, no. 3, pp. 337–350, 2017.

[76] F. Giri and E. Bai, *Block-oriented Nonlinear System Identification*, ser. Lecture Notes in Control and Information Sciences. Springer London, 2010.

[77] M. Schoukens and K. Tiels, "Identification of block-oriented nonlinear systems starting from linear approximations: A survey," *Automatica*, vol. 85, pp. 272–292, 2017.

[78] N. Vervaet, "Identification of discrete-time state-space nonlinear systems with stability guarantees," Master's thesis, Eindhoven University of Technology, 2021.

[79] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, "Reservoir observers: Model-free inference of unmeasured variables in chaotic systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 27, no. 4, p. 041102, 2017.

[80] L. Gonon and J.-P. Ortega, "Reservoir computing universality with stochastic inputs," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 1, pp. 100–112, 2019.

[81] L. Grigoryeva, J. Henriques, and J.-P. Ortega, "Reservoir computing: information processing of stationary signals," in *2016 IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES).* IEEE, 2016, pp. 496–503.

[82] F. Wyffels and B. Schrauwen, "A comparative study of reservoir computing strategies for monthly time series prediction," *Neurocomputing*, vol. 73, no. 10-12, pp. 1958–1964, 2010.

[83] F. M. Bianchi, S. Scardapane, S. Løkse, and R. Jenssen, "Reservoir computing approaches for representation and classification of multivariate time series," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 5, pp. 2169–2179, 2020.

[84] P. Billingsley, *Probability and measure.* John Wiley & Sons, New Jersey, 2012.

[85] H. Jaeger, "Short term memory in echo state networks," *GMD-German National Research Institute for Computer Science, GMD-report 152*, 2002.

[86] S. Ganguli, D. Huh, and H. Sompolinsky, “Memory traces in dynamical systems,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 48, pp. 18 970–18 975, 2008.

[87] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, “Optimal nonlinear information processing capacity in delay-based reservoir computers,” *Scientific reports*, vol. 5, no. 1, pp. 1–11, 2015.

[88] A. A. Ferreira, T. B. Ludermir, and R. R. De Aquino, “An approach to reservoir computing design and training,” *Expert systems with applications*, vol. 40, no. 10, pp. 4172–4182, 2013.

[89] S. Basterrech, E. Alba, and V. Snášel, “An experimental analysis of the echo state network initialization using the particle swarm optimization,” in *2014 Sixth World Congress on Nature and Biologically Inspired Computing (NaBIC 2014)*. IEEE, 2014, pp. 214–219.

[90] D. Sussillo and L. F. Abbott, “Generating coherent patterns of activity from chaotic neural networks,” *Neuron*, vol. 63, no. 4, pp. 544–557, 2009.

[91] R. Pyle and R. Rosenbaum, “A model of reward-modulated motor learning with parallelcortical and basal ganglia pathways,” *arXiv preprint arXiv:1803.03304*, 2018.

[92] J. C. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” *IEEE transactions on automatic control*, vol. 37, no. 3, pp. 332–341, 1992.

[93] D. J. Griffiths and D. F. Schroeter, *Introduction to quantum mechanics*. Cambridge University Press, Cambridge, 2018.

[94] P. Murali, D. C. McKay, M. Martonosi, and A. Javadi-Abhari, “Software mitigation of crosstalk on noisy intermediate-scale quantum computers,” in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, 2020, pp. 1001–1016.

[95] A. W. Cross, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, "Open quantum assembly language," *arXiv preprint arXiv:1707.03429*, 2017.

[96] K. Nakajima, "Physical reservoir computing—an introductory perspective," *Japanese Journal of Applied Physics*, vol. 59, no. 6, p. 060501, 2020.

[97] M. Buehner and P. Young, "A tighter bound for the echo state property," *IEEE transactions on neural networks*, vol. 17, no. 3, pp. 820–824, 2006.

[98] Y. Lin, E. D. Sontag, and Y. Wang, "A smooth converse Lyapunov theorem for robust stability," *SIAM Journal on Control and Optimization*, vol. 34, no. 1, pp. 124–160, 1996.

[99] Z. P. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Mathematics of Control, Signals and Systems*, vol. 7, no. 2, pp. 95–120, 1994.

[100] J. Chen and H. I. Nurdin, "Learning nonlinear input–output maps with dissipative quantum systems," *Quantum Information Processing*, vol. 18, no. 7, pp. 1–36, 2019.

[101] ——, "Correction to: Learning nonlinear input–output maps with dissipative quantum systems," *Quantum Information Processing*, vol. 18, no. 11, p. 354, 2019.

[102] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series," *IEEE Trans. Circuits Syst.*, vol. 32, no. 11, pp. 1150–1161, 1985.

[103] J. Dieudonné, *Foundations of Modern Analysis.* Read Books Ltd, Redditch, 2013.

[104] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[105] D. Perez-Garcia, M. M. Wolf, D. Petz, and M. B. Ruskai, "Contractivity of positive and trace-preserving maps under l_p norms," *Journal of Mathematical Physics*, vol. 47, no. 8, p. 083506, 2006.

[106] S. Lang, *Complex Analysis*, ser. Graduate Texts in Mathematics. Springer-Verlag, 1985.

[107] X. Ni, M. Verhaegen, A. J. Krijgsman, and H. B. Verbruggen, “A new method for identification and control of nonlinear dynamic systems,” *Engineering Applications of Artificial Intelligence*, vol. 9, no. 3, pp. 231–243, 1996.

[108] J. R. Dormand and P. J. Prince, “A family of embedded Runge-Kutta formulae,” *Journal of computational and applied mathematics*, vol. 6, no. 1, pp. 19–26, 1980.

[109] H. Jaeger and H. Haas, “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communications,” *Science*, vol. 304, p. 5667, 2004.

[110] M. Suzuki, “Relationship among exactly soluble models of critical phenomena. I: 2D Ising model, dimer problem and the generalized XY-model,” *Progress of Theoretical Physics*, vol. 46, no. 5, pp. 1337–1359, 1971.

[111] L. M. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, “Experimental realization of Shor’s quantum factoring algorithm using nuclear magnetic resonance,” *Nature*, vol. 414, no. 6866, p. 883, 2001.

[112] “IBM Q 20 Tokyo,” https://www.research.ibm.com/ibm-q/technology/devices/, Accessed: 2019-04-10.

[113] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning.* Springer series in statistics New York, 2001, vol. 1, no. 10.

[114] J. Chen, H. I. Nurdin, and N. Yamamoto, “Towards single-input single-output nonlinear system identification and signal processing on near-term quantum computers,” in *2019 58th IEEE Conference on Decision and Control (CDC).* IEEE, 2019, pp. 401–406.

[115] ——, “Temporal information processing on noisy quantum computers,” *Physical Review Applied*, vol. 14, no. 2, p. 024065, 2020.

[116] IBMQ. (2021) Device backend noise model simulations. [Online]. Available: https://qiskit.org/documentation/tutorials/simulators/2_device_noise_simulation.html

[117] V. B. Braginsky, Y. I. Vorontsov, and K. S. Thorne, “Quantum nondemolition measurements,” *Science*, vol. 209, no. 4456, pp. 547–557, 1980.

[118] V. Belavkin, “Non-demolition measurement and control in quantum dynamical systems,” in *Information complexity and control in quantum physics.* Springer, 1987, pp. 311–329.

[119] V. P. Belavkin, “Nondemolition principle of quantum measurement theory,” *Foundations of Physics*, vol. 24, no. 5, pp. 685–714, 1994.

[120] V. B. Braginsky and F. Y. Khalili, *Quantum measurement.* Cambridge University Press, Cambridge, 1995.

[121] L. Accardi, A. Frigerio, and J. T. Lewis, “Quantum stochastic processes,” *Publications of the Research Institute for Mathematical Sciences*, vol. 18, no. 1, pp. 97–133, 1982.

[122] H. I. Nurdin, *Quantum Stochastic Processes and the Modelling of Quantum Noise.* London: Springer London, 2019, pp. 1–8. [Online]. Available: https://doi.org/10.1007/978-1-4471-5102-9_100160-1

[123] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. Allman, C. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer *et al.*, “Demonstration of the trapped-ion quantum ccd computer architecture,” *Nature*, vol. 592, no. 7853, pp. 209–213, 2021.

[124] IBMQ. (2021) How to measure and reset a qubit in the middle of a circuit execution. [Online]. Available: https://www.ibm.com/blogs/research/2021/02/quantum-mid-circuit-measurement/

[125] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald,

A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[126] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," *Nature*, vol. 567, no. 7749, p. 491, 2019.

[127] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Physical Review X*, vol. 7, no. 2, p. 021050, 2017.

[128] IBMQ. (2022) System properties. [Online]. Available: https://quantum-computing.ibm.com/lab/docs/iql/manage/systems/properties

[129] R. De Prins, "Photonic quantum reservoir computing," Master's thesis, Ghent University, 2021.

[130] J. Chen and H. I. Nurdin, "A small-gain theorem for discrete-time convergent systems and its applications," *arXiv preprint arXiv:2105.02376*, 2021.

[131] T. A. Johansen and B. A. Foss, "Constructing NARMAX models using ARMAX models," *International journal of control*, vol. 58, no. 5, pp. 1125–1153, 1993.

[132] J. S. Shamma and R. Zhao, "Fading-memory feedback systems and robust stability," *Automatica*, vol. 29, no. 1, pp. 191–200, 1993.

[133] A. Pavlov, N. Van De Wouw, and H. Nijmeijer, "Convergent systems: analysis and synthesis," in *Control and observer design for nonlinear finite and infinite dimensional systems*. Springer, 2005, pp. 131–146.

[134] Z. P. Jiang, Y. Lin, and Y. Wang, "Nonlinear small-gain theorems for discrete-time feedback systems and applications," *Automatica*, vol. 40, no. 12, pp. 2129–2136, 2004.

[135] Z. P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.

[136] Z. P. Jiang, Y. Lin, and Y. Wang, “Remarks on input-to-output stability for discrete time systems,” *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 306–311, 2005.

[137] Z. P. Jiang and I. Marcels, “A small-gain control method for nonlinear cascaded systems with dynamic uncertainties,” *IEEE Transactions on Automatic Control*, vol. 42, no. 3, pp. 292–308, 1997.

[138] Z. Chen and J. Huang, “A simplified small gain theorem for time-varying nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1904–1908, 2005.

[139] M. Zhu and J. Huang, “Small gain theorem with restrictions for uncertain time-varying nonlinear systems,” *Communications in Information & Systems*, vol. 6, no. 2, pp. 115–136, 2006.

[140] E. D. Sontag and B. Ingalls, “A small-gain theorem with applications to input/output systems, incremental stability, detectability, and interconnections,” *Journal of the Franklin Institute*, vol. 339, no. 2, pp. 211–229, 2002.

[141] Y. Lin, E. Sontag, and Y. Wang, “Recent results on Lyapunov-theoretic techniques for nonlinear stability,” in *Proceedings of 1994 American Control Conference-ACC’94*, vol. 2. IEEE, 1994, pp. 1771–1775.

[142] A. Isidori, *Lectures in feedback design for multivariable systems.* Springer, Switzerland, 2017, vol. 3.

[143] E. Kreyszig, *Introductory functional analysis with applications.* John Wiley & Sons, New Jersey, 1978.

[144] S. Ibrir, “Circle-criterion approach to discrete-time nonlinear observer design,” *Automatica*, vol. 43, no. 8, pp. 1432–1441, 2007.

[145] L. Ljung, *System Identification: Theory for the User.* Prentice Hall, New Jersey, 1999.

[146] S. N. Kumpati and P. Kannan, “Identification and control of dynamical systems using neural networks,” *IEEE Transactions on neural networks*, vol. 1, no. 1, pp. 4–27, 1990.

[147] H. I. Nurdin, M. R. James, and I. R. Petersen, “Coherent quantum LQG control,” *Automatica*, vol. 45, no. 8, pp. 1837–1846, 2009.

[148] J. Combes, J. Kerckhoff, and M. Sarovar, “The SLH framework for modeling quantum input-output networks,” *Adv. Phys. X*, vol. 2, no. 784, 2017.

[149] H. I. Nurdin and N. Yamamoto, *Linear Dynamical Quantum Systems: Analysis, Synthesis, and Control*, ser. Communications and Control Engineering. Springer, Switzerland, 2017.

[150] M. Mirrahimi and P. Rouchon, “Singular perturbations and Lindblad-Kossakowski differential equations,” *IEEE Transactions Automat. Control*, vol. 54, no. 6, pp. 1325–1329, 2009.

[151] B. Besselink, N. van de Wouw, and H. Nijmeijer, “Model reduction for a class of convergent nonlinear systems,” *IEEE transactions on automatic control*, vol. 57, no. 4, pp. 1071–1076, 2011.

[152] A. Pavlov, N. Van De Wouw, and H. Nijmeijer, *Uniform output regulation of nonlinear systems: a convergent dynamics approach.* Birkhäuser Boston, 2006, vol. 205.

[153] J. Fan and Q. Yao, *Nonlinear time series: nonparametric and parametric methods.* Springer Science & Business Media, New York, 2005.

[154] L. Ljung, “Perspectives on system identification,” *Annual Reviews in Control*, vol. 34, no. 1, pp. 1–12, 2010.

[155] H. Tong, *Threshold models in non-linear time series analysis.* Springer Science & Business Media, New York, 2012, vol. 21.

[156] R. F. Engle, “Estimates of the variance of US inflation based upon the ARCH model,” *Journal of Money, Credit and Banking*, vol. 15, no. 3, pp. 286–301, 1983.

[157] I. Leontaritis and S. A. Billings, “Input-output parametric models for non-linear systems part II: stochastic non-linear systems,” *International journal of control*, vol. 41, no. 2, pp. 329–344, 1985.

[158] J. Chen and H. I. Nurdin, “Nonlinear autoregression with convergent dynamics on novel computational platforms,” *IEEE Transactions on Control Systems Technology*, pp. 1–7, 2021, doi: 10.1109/TCST.2021.3136227.

[159] J. R. Johansson, P. D. Nation, and F. Nori, “QuTiP: An open-source Python framework for the dynamics of open quantum systems,” *Computer Physics Communications*, vol. 183, no. 8, pp. 1760–1772, 2012.

[160] *5/1-Year adjustable rate mortgage average in the United States*, 2020. [Online]. Available: https://fred.stlouisfed.org/series/MORTGAGE5US

[161] T. Wigren and M. Schoukens, “Coupled electric drives data set and reference models,” *Technical Report, Department of Information Technology, Uppsala University*, no. 024, 2017.

[162] S. Ibrir, W. F. Xie, and C.-Y. Su, “Observer-based control of discrete-time lipschitzian non-linear systems: application to one-link flexible joint robot,” *International Journal of Control*, vol. 78, no. 6, pp. 385–395, 2005.

[163] J. F. Sturm, “Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[164] J. Löfberg, “YALMIP : A Toolbox for Modeling and Optimization in MATLAB,” in *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[165] D. Qiu, *Augmented Dickey-Fullter test*, 2015. [Online]. Available: https://www.rdocumentation.org/packages/aTSA

[166] B. Pfaff, *KPSS test*, 2016. [Online]. Available: https://www.rdocumentation.org/packages/urca

[167] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling.* Springer, New York, 2008.

[168] R. J. Hyndman, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen, *forecast: Forecasting functions for time series and linear models*, 2022, R package version 8.16. [Online]. Available: http://pkg.robjhyndman.com/forecast

[169] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[170] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[171] C. Santori, J. S. Pelc, R. G. Beausoleil, N. Tezak, R. Hamerly, and H. Mabuchi, "Quantum noise in large-scale coherent nonlinear photonic circuits," *Physical Review Applied*, vol. 1, no. 5, p. 054005, 2014.

[172] J. Wang, T. Niu, H. Lu, W. Yang, and P. Du, "A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 337–349, 2019.

[173] J. Chen and H. I. Nurdin, "Online algorithms for polynomial regression on physical reservoir computers with noisy measurements," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 1502–1508.

[174] X. Chen, H. Hong, and D. Nekipelov, "Nonlinear models of measurement errors," *Journal of Economic Literature*, vol. 49, no. 4, pp. 901–37, 2011.

[175] P. Sadegh, "Constrained optimization via stochastic approximation with a simultaneous perturbation gradient approximation," *IFAC Proceedings Volumes*, vol. 30, no. 11, pp. 281–285, 1997.

[176] J. C. Spall, "An overview of the simultaneous perturbation method for efficient optimization," *Johns Hopkins apl technical digest*, vol. 19, no. 4, pp. 482–492, 1998.

[177] H. Kushner, D. Clark, and K. M. R. Collection, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, ser. Applied Mathematical Sciences. Springer, New York, 1978.

[178] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[179] J. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, pp. 817–823, 1998.

[180] PennyLane. (2021) Optimization using SPSA. [Online]. Available: https://pennylane.ai/qml/demos/spsa.html#spall-implementation

[181] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature communications*, vol. 9, no. 1, pp. 1–6, 2018.

[182] B. Collins and P. Śniady, "Integration with respect to the haar measure on unitary, orthogonal and symplectic group," *Communications in Mathematical Physics*, vol. 264, no. 3, pp. 773–795, 2006.

[183] S. Chatterjee, "Concentration of Haar measures, with an application to random matrices," *Journal of Functional Analysis*, vol. 245, no. 2, pp. 379–389, 2007.

[184] Z. Puchała and J. A. Miszczak, "Symbolic integration with respect to the Haar measure on the unitary groups," *Bulletin of the Polish Academy of Sciences*, vol. 65, no. 21, 2017.

[185] R. Cleve, D. Leung, L. Liu, and C. Wang, "Near-linear constructions of exact unitary 2-designs," *Quantum Information and Computation*, no. 9&10, pp. 0721–0756, 2016.

[186] Y. Nakata, C. Hirche, C. Morgan, and A. Winter, "Unitary 2-designs from random X-and Z-diagonal unitaries," *Journal of Mathematical Physics*, vol. 58, no. 5, p. 052203, 2017.

[187] Y. Liu, S. Arunachalam, and K. Temme, “A rigorous and robust quantum speed-up in supervised machine learning,” *Nature Physics*, vol. 17, no. 9, pp. 1013–1017, 2021.

[188] M. J. Bremner, R. Jozsa, and D. J. Shepherd, “Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy,” *Proc. Royal Soc. A*, vol. 467, pp. 459–472, 2010.

[189] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, “Expressive power of parametrized quantum circuits,” *Physical Review Research*, vol. 2, no. 3, p. 033125, 2020.

[190] K. Nakaji and N. Yamamoto, “Expressibility of the alternating layered ansatz for quantum computation,” *Quantum*, vol. 5, p. 434, 2021.

[191] M. Revay and I. Manchester, “Contracting implicit recurrent neural networks: Stable models with improved trainability,” in *Learning for Dynamics and Control.* PMLR, 2020, pp. 393–403.

[192] R. Kuiper, H. Hoijtink, and M. Silvapulle, “An Akaike-type information criterion for model selection under inequality constraints,” *Biometrika*, vol. 98, no. 2, pp. 495–501, 2011.

[193] F. Abramovich and V. Grinshtein, “Model selection in regression under structural constraints,” *Electronic Journal of Statistics*, vol. 7, pp. 480–498, 2013.

[194] M.-S. Oh, “Bayesian comparison of models with inequality and equality constraints,” *Statistics & Probability Letters*, vol. 84, pp. 176–182, 2014.

[195] L.-Y. Chen and J. Szroeter, “Extension of the schwarz information criterion for models sharing parameter boundaries,” *Journal of Statistical Planning and Inference*, vol. 174, pp. 68–84, 2016.

[196] J. Mulder and A. E. Raftery, “BIC extensions for order-constrained model selection,” *Sociological Methods & Research*, p. 0049124119882459, 2019.

[197] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, “An initialization strategy for addressing barren plateaus in parametrized quantum circuits,” *Quantum*, vol. 3, p. 214, 2019.

[198] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature communications*, vol. 12, no. 1, pp. 1–12, 2021.

[199] T. L. Patti, K. Najafi, X. Gao, and S. F. Yelin, “Entanglement devised barren plateau mitigation,” *Physical Review Research*, vol. 3, no. 3, p. 033090, 2021.

[200] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, “Effect of barren plateaus on gradient-free optimization,” *Quantum*, vol. 5, p. 558, 2021.

[201] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, “Absence of barren plateaus in quantum convolutional neural networks,” *Physical Review X*, vol. 11, no. 4, p. 041011, 2021.