Foreword

Reservoir computing seems simple but is difficult, it feels new but it is old, it opens horizons and is brutally limiting. I will do my best in this foreword to leave the reader with many questions — to be answered, or maybe not, in the many chapters of this richly filled book.

The basic principle of reservoir computing (RC) is simple. Given: a *training* input signal $\mathbf{u}^{\mathsf{train}}(t)$ paired with a desired target output signal $\mathbf{y}^{\mathsf{train}}(t)$. Wanted: a filter (transducer) $\mathcal{F}$ which, when fed with input $\mathbf{u}^{\mathsf{train}}(t)$, generates an output signal $\hat{\mathbf{y}}^{\mathsf{train}}(t)$ which comes close to the target $\mathbf{y}^{\mathsf{train}}(t)$. Approach: *Step 1.* Prepare a high-dimensional dynamical system $X(t)$, the *reservoir*, which can be driven by input $\mathbf{u}^{\mathsf{train}}(t)$ and in which many state variables $x_i(t)$ (where $i = 1, \ldots, N$) can be observed and recorded. *Step 2.* Drive this system with input $\mathbf{u}^{\mathsf{train}}(t)$ and record the corresponding reservoir-internal response signals $x_i^{\mathsf{train}}(t)$. *Step 3.* Find (train, learn, estimate) a *readout function* F which maps every recorded state vector $(x_1^{\mathsf{train}}(t), \ldots, x_N^{\mathsf{train}}(t))$ to an output $\hat{\mathbf{y}}^{\mathsf{train}}(t)$ which approximates the training targets $\mathbf{y}^{\mathsf{train}}(t)$. Finding such a readout F is often boils down to a simple linear regression. Exploitation: feed new input signals $\mathbf{u}(t)$ to the reservoir, observe the reservoir-internal state vectors $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$, and compute the output signal $\hat{\mathbf{y}}(t) = F(\mathbf{x}(t))$.

This basic scheme is very versatile. One can solve temporal input-output tasks for timeseries prediction, dynamical pattern generation, classification and segmentation, control, de-noising and channel equalization, rare event monitoring, and many more. One can apply RC to obtain practical engineering solutions in signal processing and control and robotics and communication technologies and machine learning and AI; one can call upon RC models as explanatory principle in theoretical neuroscience; and in mathematics one can use RC as an entry point to identify and analyse a number of interesting phenomena in high-dimensional dynamical systems. But most importantly, one can in principle use *any* kind of nonlinear, high-dimensional dynamical sytem for the reservoir $X(t)$, regardless of whether it is an experimental probe of a freshly engineered nanomaterial, a quantum dot preparation, a replica of an octopus arm made from soft plastic and suspended in water, or a digital simulation of a neural network — all to be found in the scintillating collection of reservoirs that the reader will find in this book.

But... the closer one becomes involved with RC, the more difficult it gets, and if one to embrace it in full contact, it gets almost impossibly difficult. Reservoirs are high-dimensional, input-driven, nonlinear and often stochastic dynamical systems. A full theory of reservoir dynamics would be a full theory of everything that evolves in time. Only fragmentary insights into the unbounded phenomenal richness in general dynamical systems are currently available in mathematics, theoretical physics and biology or the general complex systems sciences. Compared to what we *could* know about,

observe in, and utilize from reservoir dynamics, we currently *do* know, see and use — almost nothing. It is easy to program a recurrent neural network with 100 neurons on a digital computer, declare it a reservoir, apply the basic RC scheme on a simple modeling task, exclaim "it works!" and call it good. This is how students worldwide get hooked on RC. But when the data are noisy or incomplete or have outliers or are nonstationary or have a wandering baseline or variable amplitude or are high-dimensional or have multiple spatial or temporal scales, or when stability conditions have to be guaranteed, or when the task demands continual online learning, or when the input data consist of rare events spiking out of a zero baseline, or when there are many possible options for input and output signal re-coding (there always are), or when one's computer allows only fast experimentation with small reservoirs but one wants to extrapolate to large ones, or when one wants to automate the readout training, — to cut this short: when one gets pushed out of the comfort zone of the dozen or so ever-repeated "benchmark" tasks that pervade the RC literature — then reservoirs turn into feral beasts that take an enormous amount of patience and experience to tame. The promise of RC: one *need not* train the reservoir, turns into a problem: one *can not* train the reservoir. There is an unlimited variability in task specifics and there is an infinity of dynamical behaviors in candidate reservoir systems. In a haystack of possibilities one must find a reservoir whose native dynamics matches the demands of the task at hand. After two decades of RC research we only have the faintest inklings of how to match reservoir dynamics with task dynamics. Many of my students choose a reservoir computing theme for their graduating thesis. I dare say that, when after much trial and error they ultimately arrive at the point where "it works!", they don't understand *why* it works — and neither do I.

Many contemporary RC papers that I read or review introduce their subject still with "Reservoir computing is a new approach to train neural networks ...". Well... RC may be called "new" compared to Newton's and Leibniz's calculus, but by the standards of the fast-paced innovation cycles in machine learning it is rather old. The basic RC principle has been discovered and re-discovered many times, and I continue to become aware of earlier and earlier "first" sightings. This is how it goes with most ideas that are elementary and useful.

It is not customary to cite references in a foreword, but I take this as an opportunity to give due credit to RC pathfinders. The earliest perception of the RC principle that I am aware is Kirby and Day [1990], a 1-page conference abstract that was subsequently worked out by Kevin Kirby in a paper where he gives what I consider the first concise and comprehensive account of the RC principle [Kirby, 1991], with the readout from the reservoir (which he called *context reverberation subsystem*) trained by the perceptron learning algorithm. The problem of finding a "good" reservoir is clearly identified, and a sentence in the Conclusion section reads like proph-

esy: "This may encourage molecular electronic hardware implementations." Both papers remained entirely unnoticed (the single Google Scholar cite that I saw when I queried this in 2017 was a self-citation). In the same year 1991, Lambert Schomaker, in Chapter 7 of his PhD thesis [Schomaker, 1991] (separately published in Schomaker [1992]), described how a target output signal can be obtained by learning a linear readout from a random ensemble of spiking neural oscillators. I got to know about this work by an unlikely chance: after I was appointed at the University of Groningen in 2019, Lambert became my direct senior manager and he told me about his PhD thesis in a casual conversation. I wonder how many other casual conversations with other senior colleagues worldwide would bring up similar surprises. Both Schomaker and Kirby refer back to earlier precursor ideas in their texts — clues for further studies in scientific archaeology. The next independent discovery of RC that I know about occurred in cognitive neuroscience. Peter F. Dominey described a multi-module (human) brain circuit for sequence generation which included a simplified model of prefrontal cortex as reservoir from which trainable readouts send information to the caudate nucleus [Dominey, 1995]. Up to the present day, and in close collaboration with other RC researchers, Dominey has been continuing to work out elaborate neuro-cognitive architectures with an RC core both for neuroscience modeling and for robotic / human-maching interaction applications. His chapter in this book gives a summary of a 25-year-long personal research mission.

The current RC literature mostly localizes the origin of RC in the propositions of *liquid state machines* by Wolfgang Maass and my *echo state networks* [Maass et al., 2002, Jaeger, 2001]. Wolfgang and I got to know of each other at the 2001 EU Advanced Course in Computational Neuroscience at the International Center for Theoretical Physics in Trieste, Italy, August 2001, where to our mutual surprise we found our own ideas almost identically reflected in the respective other's. We started to collaborate, soon joined by Benjamin Schrauwen who coined the term "reservoir computing" (or was it his brilliant PhD student David Verstraeten? the first published paper where this term was used seems to be Verstraeten et al. [2005]). Benjamin rapidly built up an enormously productive RC research group at the University of Gent even before he was awarded his PhD degree. I think several factors came together why reservoir computing took off only then. First, the three of us teamed up instead of defending proprietary RC islands. Second, for the first time the mathematical preconditions that make reservoirs functional were clearly spelled out through the *fading memory* and *separation property* in Wolfgang's models and the *echo state property* and an analysis of a reservoir's *memory capacity* in my work. Mathematical formulae gave authority to a wild-looking idea. Third, "it worked" really well in many demo tasks that met the taste and demands of the time — while training recurrent neural networks with other then-existing learning

algorithms was difficult, unstable, and slow. The deep learning revolution superseded RC only toward the end of the 2010's when the intricacies of gradient-descent training of recurrent neural networks finally became mastered. Reservoir computing research receded into a niche for a few years.

But RC research re-awakened and sprouted out again from this niche when RC principles were adopted first in the field of optical computing (history outline in this book's chapter by Dambre et al.) and swiftly also in other domains of *physical reservoir computing* (surveyed in the chapter by Dale et al.). Most chapters in this book are a testimonial to the refreshing new thrust that RC has given to the wider fields of *unconventional / in-materio / natural / ...* computing (I have a private list of about 15 different namings that have been branded in the last four decades or so). Materials scientists and non-digital device engineers from the most diverse makings continue to discover RC for themselves. As long as RC continues to be freshly discovered by colleagues in widening circles, there is still truth in when they say, in the introductory passages of articles, that RC is "... a new approach ...".

Ah, before I forget: there is one little technical thing that I want to point out to everyone who uses RC for the first time. So many neural-network-based RC papers state in their methods section that the spectral radius (largest absolute eigenvalue) of the network weight matrix should be less than unity to ensure the echo state property, a necessary condition to make RC work. This is a myth. A spectral radius $SR < 1$ is neither sufficient nor necessary for the echo state property [Yildiz et al., 2012], and a value much larger than 1 often gives the best performance. Please don't perpetuate this myth in your work! And while I am at it: another myth is that reservoirs work best when they are tuned to operate "at the edge of chaos", or "close to criticality". First, it's a misnomer, because the edge in question here is the edge of the echo state property, not the edge of chaos. If a reservoir slides across this edge, it doesn't necessarily (even not typically) enter a chaotic regime. Second, reservoirs are input-driven systems, and mathematicians still haven't entirely agreed how to define chaos in input-driven systems. Finally, reservoirs "close to criticality" work well only for a certain class of learning tasks — the sort of tasks which are invariably re-iterated in articles on this subject — but reservoirs far on the stable side of that edge work much better for many other tasks. Cramer et al. [2020] point a spotlight on this affair. I really can't understand why this myth remains recited so often, given the massive counter-evidence from so many practical applications where carefully optimized reservoirs come out sitting safe and far away from this edge.

RC has the elegance of simplicity, which may be explanation enough why it inspires researchers in many fields. Of course there are more substantial reasons why RC keeps blossoming, for instance because it connects the neuro- with the computing sciences in stronger than purely metaphorical

ways; or that it opens new doors for theoretical analyses of high-dimensional dynamical systems; or that materials scientists today really don't have many alternatives to make their unconventional substrates "compute".

But one should be aware that the powers of RC as a stand-alone carrier of "learning" or "computing" are decisively limited. Biological brains may be using RC in some places and some ways — it is unlikely that they don't because evolution will find and keep any trick that works — but brains use many other dynamical mechanisms and structuring principles and information encoding procedures as well, and I don't think we have an idea yet even of *how* many. From my perspective of machine learning, AI and theory of computing, the strongest weakness (what a nice oxymoron) of pure RC is its inherent blindness to hierarchical multiscale compositionality of data structures, processes and architectures. Here I understand compositionality in a strong sense which includes *bidirectional* interaction between higher modules or layers and their sub-modules or lower layers. An example are planning architectures for autonomous agents where higher planning modules generate longer-term plans that "call" lower sub-plan modules in a *top-down* direction, and stay informed about execution progress in a *bottom-up* direction of communication from the sub-modules. Another example are the Boltzmann machine or Friston's free-energy models of neural processing where higher layers send statistical biases to lower layers, and are informed about conditional feature distributions from below. In computer science, the object-oriented programming paradigm is the very manifestation of bidirectionally effective compositionality. The top-down actions can be interpreted in a variety of ways, for instance as attention control, predictive context settings, or read/write signals in working memory systems. Such top-down modulations are essential for full-fledged cognitive information processing: but, such bidirectionally effective hierarchical cognitive architectures cannot be realized by RC alone. Additional structures and algorithms are needed to coordinate intermodule communication (as in my attempt in Jaeger [2007] to design a hierarchical RC learning architecture that can discover temporal features on several timescales); or additional teacher signals for the individual modules must be created (as in Pascanu and Jaeger [2011] where we trained a kind of parser for visual text input that had a nested grammatical structure); or additional control mechanisms must be installed to modulate the reservoir dynamics "from above" (like *conceptors* [Jaeger, 2017]). It is one of the strongest strengths of today's deep learning networks that such multi-directionally organized architectures, like for instance *neural Turing machines*, can be trained by "end-to-end" gradient descent, where the requisite local training signals are automatically generated. — This said, I emphasize that RC can positively be applied with much benefit in certain multiscale learning tasks using uni-directionally coupled stacks of reservoirs where lower, typically faster reservoirs connect upwards to higher, typically slower reservoirs. Gallicchio and Micheli (in this volume) survey such archi-

tectures, which they call *deep RC* systems.

RC research keeps advancing in many directions. I want to conclude with my personal favourite challenges for the next evolutionary steps in RC research:

- Coordinate RC modules in complex learning and information processing systems with the aid of additional mechanisms — this theme is also highlighted in the Conclusion of Dambre's et al. chapter in this book.

- In physical RC, find ways to realize the readout and its training directly in the non-digital material substrate, instead of delegating it to a digital host computer.

- Leverage the infinite dimensionality of spatially extended nonlinear excitable media, extending the readout combination of a finite number of reservoir signals to an infinite-dimensional integration, convolution or field transformation. In physical RC one might envision spatially continuous two-layer substrates where the bottom layer acts as reservoir and the top layer would function as a continuous version of what today are the readout weight matrices.

- Find effective ways to cope with the unpleasant properties of physical reservoirs, such as device mismatch, parameter drift, temperature sensitivity, low numerical precision and stochasticity, and partial observability. Physical reservoirs may only become practically useful when appropriate auto-calibration or homeostatic regulatory mechanisms are realized in combination with numerically robust and swiftly self-adapting readout processes.

- Rigorously analyze which abstract dynamical characteristics of input and output data and task specifications should be reflected in which characteristics of reservoir dynamics. Currently available insights are mostly distilled from experimental studies of timescale profiles or frequency spectra in input data and provide no comprehensive guides for optimising reservoir designs.

Many chapters in this collection include historical summaries of major RC research strands, and all tell enticing stories about what today's achievements are and are not. This book lets us see where we stand and invites us to imagine where we can go further. Being a veteran of the field, I feel enormously grateful for the massive labor of editors and authors to plant this landmark after thirty years of a voyage that will continue to feel fresh and young.

References

B. Cramer, D. Stöckel, M. Kreft, M. Wibral, J. Schemmel, K. Meier, and V. Priesemann. Control of criticality and computation in spiking neuromorphic networks with plasticity. *Nature Communications*, 11:2853, 2020.

P. F. Dominey. Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning. *Biological Cybernetics*, 73:265–274, 1995.

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science, 2001. URL `https://www.ai.rug.nl/minds/uploads/EchoStatesTechRepErratum.pdf`.

H. Jaeger. Discovering multiscale dynamical features with hierarchical echo state networks. Technical report 10, School of Engineering and Science, Jacobs University, 2007. URL `https://www.ai.rug.nl/minds/uploads/hierarchicalesn_techrep10.pdf`.

H. Jaeger. Using conceptors to manage neural long-term memories for temporal patterns. *JMRL*, 18:1–43, 2017.

K. Kirby. Context dynamics in neural sequential learning. In *Proc. Florida AI Research Symposium (FLAIRS)*, pages 66–70, 1991.

K. G. Kirby and N. Day. The neurodynamics of context reverberation learning. In *Proc. Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society, Vol. 12 No. 4*, pages 1781–1782, 1990.

W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

R. Pascanu and H. Jaeger. A neurodynamical model for working memory. *Neural Networks*, 24(2):199–207, 2011.

L. R. B. Schomaker. *Simulation and Recognition of Handwriting Movements: A vertical approach to modeling human motor behavior*. Phd thesis, Nijmeegs Instituut voor Cognitie-onderzoek en Informatietechnologie, Nijmegen, 1991. URL `https://repository.ubn.ru.nl/handle/2066/113914`.

L. R. B. Schomaker. A neural oscillator-network model of temporal pattern generation. *Human Movement Science*, 11:181–192, 1992.

D. Verstraeten, B. Schrauwen, and D. Stroobandt. Reservoir computing with stochastic bitstream neurons. In *Proceedings of the 16th annual Prorisc workshop*, pages 454–459, 2005.

I. B. Yildiz, H. Jaeger, and S. J. Kiebel. Re-visiting the echo state property. *Neural Networks*, 35:1–20, 2012.