

This is a repository copy of *A Substrate-Independent Framework to Characterise Reservoir Computers*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/200178/>

Version: Other

Preprint:

Dale, Matthew, Miller, Julian F. orcid.org/0000-0002-7692-9655, Stepney, Susan orcid.org/0000-0003-3146-5401 et al. (1 more author) (2018) *A Substrate-Independent Framework to Characterise Reservoir Computers*. [Preprint]

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A Substrate-Independent Framework to Characterise Reservoir Computers

Matthew Dale^{1,3}, Julian F. Miller³, Susan Stepney^{1,3}, Martin A. Trefzer^{2,3}

Abstract The Reservoir Computing (RC) framework states that any non-linear, input-driven dynamical system (the *reservoir*) exhibiting properties such as a fading memory and input separability can be trained to perform computational tasks. This broad inclusion of systems has led to many new physical substrates for RC. Properties essential for reservoirs to compute are tuned through reconfiguration of the substrate, such as change in virtual topology or physical morphology. As a result, each substrate possesses a unique “quality” – obtained through reconfiguration – to realise different reservoirs for different tasks.

Here we describe an experimental framework that can be used to characterise the quality of *any* substrate for RC. Our framework reveals that a definition of quality is not only useful to compare substrates, but can also help map the non-trivial relationship between properties and task performance. And through quality, we may even be able to predict the performance of similarly behaved substrates. Applying the framework, we can explain why a previously investigated carbon nanotube/polymer composite performs modestly on tasks, due to a *poor* quality. In the wider context, the framework offers a greater understanding to what makes a dynamical system compute, helping improve the design of future substrates for RC.

Keywords Reservoir Computing, Physical Computation, Substrate Characterisation, Dynamical Properties

¹Department of Computer Science, University of York, UK

²Department of Electronic Engineering, University of York, UK

³York Cross-disciplinary Centre for Systems Analysis
matt.dale@york.ac.uk

1 Introduction

Reservoir Computing (RC) first emerged as an alternative method for constructing and training recurrent neural networks [27, 32]. The method primarily involved constructing a random fixed recurrent network of neurons, and training only a single linear readout layer. It was found that random networks constructed with certain dynamical traits could produce state-of-the-art performance without the laborious process of training individual internal connections. The concept later expanded to encompass any high dimensional, input-driven dynamical system that could operate within specific dynamical regimes, leading to an explosion in new reservoir computing substrates¹.

In recent years, the reservoir computing model has been applied to a variety of physical systems such as optoelectronic and photonic [1, 31], quantum [9, 23, 30], disordered and self-organising [4, 29], magnetic [16, 25], and memristor-based [8] computing systems. The way each substrate realises a reservoir computer varies. However, each tends to implement, physically or virtually, a network of coupled processing units.

Each implementation is designed to utilise and exploit the underlying physics of the substrate, to embrace its intrinsic properties to improve performance, efficiency and/or computational power. As with many physical systems, each can be configured, controlled and tuned to perform a desired functionality. In all of the above examples, this requires the careful tuning of parameters in order to produce working and optimal physical reservoirs.

¹ The term “substrate” is used here to refer to any physical or virtual system that realises a reservoir computer: any dynamical system featuring configurable parameters and a method to observe system states.

In general, the abstract term *reservoir* usually represents a single, typically static, configuration of the substrate. For an artificial recurrent neural network, implemented *in silico*, this may refer to a set of trained connection weights, defined neuron types and topology. For another substrate, configuration may refer to the physical morphology, physical state, external control signals, or complexification of the driving input signal. This implies that the number of possible reservoirs realised by one substrate depends upon the number of free parameters and distinct dynamical behaviours resulting from those parameters. For unconstrained substrates, limited only by the laws of physics, this number may be vast. Yet this does not imply that every such configuration/reservoir is practical or useful.

In terms of all possible reservoirs realisable by one substrate, the vast majority may be unusable in terms of solving a task. However, some region of the substrate’s configuration space may well provide interesting reservoirs and potentially high-performing reservoirs, or even reservoirs with large generalising computing abilities.

Characterising the configuration and reservoir space (referred to below as the *behaviour space*) of usable and optimal reservoirs would help in assessing the substrate’s “quality” for reservoir computing, that is, the substrate’s ability to realise different reservoirs, and therefore its capacity as a generic reservoir computing substrate.

According to [7], all dynamical systems have an almost universal characteristic to perform useful information processing, provided a fading memory and linearly independent internal variables are present. However, each dynamical system tends to suit different tasks, and rarely, but not unattainably, will one feature a universal set of properties to perform well across many, if not all tasks. This implies that high-performing, task-specific, and potentially some good task-generalising reservoir computers can be built, leading to a vast array of new highly-efficient and powerful computing substrates.

But first, we have to discover these good reservoirs lying somewhere in their vast behaviour space. Our CHARC (CHARacterisation of Reservoir Computers) framework provides a method to do so.

2 The CHARC Framework

Dambre et al. [7] devise a quantitative measure that is independent of physical realisation, allowing anyone to compare the computational properties of a broad class of dynamical systems. However, that total capacity measure may not be informative enough to guide

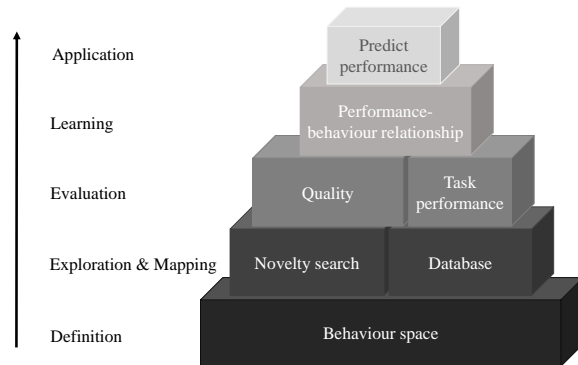


Fig. 1: Framework levels and building blocks.

substrate optimisation for specific tasks, or demonstrate how expressive the substrate is in terms of realising vastly different reservoirs.

So far, no practical framework exists to map, then utilise, the full computational expressiveness of physical or virtual substrates. That is, no experimental method has been proposed to characterise the reservoir computing *quality* of substrates, or to use measures of computational properties to configure and discover optimal reservoirs.

The challenge in creating a generic framework arises from the enormous variety of possible substrates, and from each substrate having its own set of configuration and other parameters.

To tackle this non-trivial problem we have developed the CHARC framework. The main purpose of the framework is to characterise and assess the quality of any potential RC substrate. The framework also has a secondary purpose, to exploit the quality assessment process to better understand the general relationships between computational properties and task performance.

To conceptualise and visualise the framework it is divided into a series of building blocks and levels, as shown in Fig. 1. To make use of the framework, and to assess accuracy and validity, five stages need to be completed. These are represented by the five levels in the figure: 1) *Definition*, 2) *Exploration & Mapping*, 3) *Evaluation*, 4) *Learning*, and 5) *Application*.

Each level of the framework relies on the presence of the level below. Each level consists of building blocks. In general, these blocks are adaptable, and may be improved by future work. For example: the current method for exploring the behaviour space may be improved or interchanged with another exploration technique; more dimensions may be added to the behaviour space with the creation of new property measures. In some cases, explained later, some blocks/levels can even be removed.

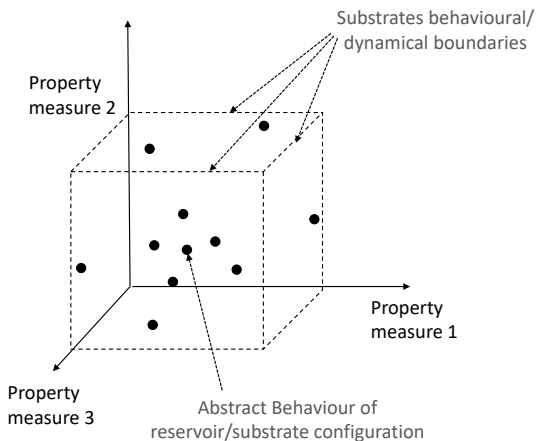


Fig. 2: Example of a 3-Dimensional Behaviour Space. Here each abstract behaviour is relative to the three chosen property measures. Given enough time to explore the space, the substrate’s dynamical/behavioural limitations become apparent.

2.1 Definition

On the first level, a *behaviour* space is defined. This abstract behaviour space represents the dynamical behaviour of the substrate when configured. To represent the n -dimensional space, n independent property measures are used, defining the axes of the space (see example in Fig. 2). It is hypothesised here that the more distinct measures applied, the better the representation of the substrate’s dynamical freedom and thus the better the accuracy of the quality measure. In the example later in this paper, we use three measures, and hence a three dimensional behaviour space.

2.2 Exploration & Mapping

The next level is *Exploration & Mapping*. To determine a true measure of quality, rather than an approximation, exploration of the behaviour space would require an exhaustive search of the substrate’s parameter space, which is infeasible. Rather than exhaustive search, an implementation of novelty search (NS) [18] is recommended. Novelty search is an open-ended genetic algorithm designed to explore a behaviour space for novel solutions until some user-defined termination criteria is met.

When applied in this setting, novelty search characterises the substrate’s search space and as a by-product outlines its dynamical boundaries, when given enough time. This experimental characterisation can then help determine the practical use, if any, of the substrate,

or whether the selected method of configuration and observation (which itself may be optimised) is appropriate.

Throughout the exploration process every behaviour is stored in a database forming a behavioural map of the substrate for later use.

2.3 Evaluation

The next level is *Evaluation*. To determine quality, the spread and number of distinct behaviours recorded in the abstract behaviour space (the database) is quantitatively measured. Quality is therefore a measure of how many distinct dynamical behaviours the substrate experimentally possesses in the defined behaviour space. The measure itself can be simple; however, any measure of quality requires some contextual reference against which the measure can be regarded as high or low, good or poor. To assess the quality of a new substrate it is recommended that an easy to define/inspect, ideally known to be high-performing, reference substrate first be assessed and evaluated. In the work here, to provide a baseline substrate to compare to, and to evaluate the framework, simulated *Echo State Networks* (ESNs) [13] are used as a *reference* substrate. In the future, if better, i.e. good generalising substrates with higher degrees of dynamical freedom, are found, the reference substrate can be replaced.

By this point, use of the framework will have resulted in a quality assessment of a particular substrate.

2.4 Learning

The learning level fulfils the secondary goal of the framework: to relate computational properties and behaviours to task performance.

As explained in [10], relating properties to expected performance is non-trivial for all task applications, as good properties for one task may be detrimental to another. Therefore, no single set of properties will always lead to high performance. However, the relationship between properties and a single task may be simpler to determine.

To model and estimate the relationships between properties and task performance a broad range of properties and performances are required.

As part of the exploration phase, diverse properties are sought and stored in the database. Then, at the evaluation level, task performance of every behaviour/reservoir in the database is assessed. This combination then leads to a basic dataset to train a learning system, for example, a neural network.

2.5 Application

At this level, the learnt relationships can be reused to predict the performance of any substrate – without the need to evaluate directly – based on the reference substrates behaviours and task performances. As a result, the cost of future task assessments can be significantly reduced by predicting task performances based on similarly behaved substrates.

3 Task-Independent Properties

In order to create the behaviour space, each dimension of the space must be defined. In the following subsections different computational properties and measures are discussed.

A potential problem in defining the behaviour space is that some properties are difficult, if not impossible, to measure across all substrates. It is therefore important to remember that any properties applied with the framework should represent the behaviour of the system independent of its implementation.

3.1 Kernel and Generalisation Rank

Kernel quality is a measure of the reservoir’s ability to produce a rich non-linear representation of the input u and its history $u(t-1), u(t-2), \dots$. Also known as the *linear separation property*, it was first introduced by Legenstein & Maass [17] to measure a reservoir’s ability to separate distinct input patterns. As many practical tasks in machine learning are linearly inseparable, reservoirs would not be able to solve such problems without some non-linear transformation of the input.

The kernel quality measure is performed by computing the rank r of an $n \times m$ matrix M , outlined in [3]. To create the matrix M , apply m distinct input streams u_i, \dots, u_m and collect the n resulting reservoir states x_{u_i} . Place the states x_{u_i} in each column of the matrix M and repeat m times. The rank r of M is computed using Singular Value Decomposition (SVD) and is equal to the number of non-zero diagonal entries in the unitary matrix. The maximum value of r is always equal to the smallest dimension of M . To calculate the effective rank, and better capture the information content, remove small singular values using some high threshold value. To produce an accurate measure of kernel quality m should be sufficiently large, as accuracy will tend to increase with m until it eventually converges.

The generalisation rank is a measure of the reservoir’s capability to generalise given similar input streams.

It is calculated using the same rank measure as kernel quality, however each input stream u_{i+1}, \dots, u_m is a noisy version of the original u_i . A low generalisation rank symbolises a robust ability to map similar inputs to similar reservoir states.

Reservoirs in ordered regimes typically have low ranking values in both measures, and in chaotic regimes both are high. In general, a good reservoir should possess a high kernel quality rank and a low generalisation rank [3]. However, in terms of matching reservoir dynamics to tasks, the right balance will vary. These two measures are important, but by themselves do not capture enough information about the reservoir’s dynamical properties.

3.2 Memory Capacity

A simple measure for the linear short-term memory capacity (MC) of a reservoir was first outlined in [14] to quantify the *echo state* property. For the echo state property to hold, the dynamics of the input driven reservoir must asymptotically wash out any information resulting from initial conditions. This property therefore implies a fading memory exists, characterised by the short-term memory capacity.

To evaluate memory capacity of an N node reservoir, we measure how many delayed versions k of the input u the outputs y can recall, or recover with precision. Memory capacity MC is measured by how much variance of the delayed input $u(t-k)$ is recovered at $y_k(t)$, summed over all delays.

$$MC = \sum_{k=1}^{2N} MC_k = \sum_{k=1}^{2N} \frac{\text{cov}^2(u(t-k), y_k(t))}{\sigma^2(u(t))\sigma^2(y_k(t))} \quad (1)$$

A typical input consists of t samples randomly chosen from a uniform distribution between $[0 \ 1]$. Jaeger [14] demonstrates that echo state networks driven by an i.i.d. signal can possess only $MC \leq N$.

A full understanding of a reservoir’s memory capacity cannot be encapsulated through a linear measure alone, as a reservoir will possess some non-linear capacity. Other memory capacity measures proposed in the literature quantify the non-linear, quadratic and cross-memory capacities of reservoirs [7].

4 Behaviour Exploration

In general, theoretically determining the computational *capacity* of a system helps us understand its limitations. One might think that this knowledge should then be

used to construct or search for reservoirs with a maximal computational capacity. However in practice such maximisation is often unnecessary, time-consuming and may in fact hinder performance. A balance between properties is essential to match reservoir dynamics to tasks.

The CHARC framework starts by exploring and mapping a vast range of dynamics for which the right balance can be selected for any task. In order for the framework to function properly and translate to many systems, this mapped space of dynamics requires substrate-independence.

To be substrate-independent, exploration must function without any prior knowledge of how to construct reservoirs far apart from each other in the observed behaviour space. Exploration cannot, therefore, be measured in the substrate parameter space: diversity in observed dynamics does not always coincide with diversity in substrate-specific parameters.

4.1 Novelty Search

In the following example of the framework, an open-ended evolutionary algorithm called novelty search (NS) [18, 19, 20] is adopted. In our implementation, novelty search is used to characterise the substrate’s behaviour space, i.e. the dynamical freedom of the substrate, by sampling its most interesting dynamical behaviours.

In contrast to objective-based techniques, a search guided by novelty has no explicit task-objective other than to maximise novelty. Novelty search directly rewards divergence from prior behaviours instead of rewarding progress to some objective goal.

Exploration without objectives has been shown, somewhat counter-intuitively, to outperform objective-based methods with deceptive task and solution spaces [26]. A deceptive objective (fitness) landscape is one where local optima are pervasive. When characterising substrate dynamics, this is of particular concern due to the high dimensionality of the substrate’s computational properties which are only partially described, i.e. measures are only approximate.

Novelty search explores the behaviour space by promoting configurations that exhibit novel behaviours. Novelty of any individual is computed with respect to its distance from others in the behaviour space. To track novel solutions, an *archive* is created holding previously explored behaviours. Contrary to objective-based searches, novelty takes into account the set of all behaviours previously encountered, not only the current population. This enables the search to keep track of (and map) lineages and niches that have been previously explored.

To promote further exploration, the archive is dynamically updated with respect to two parameters, ρ_{min} and an update interval. The ρ_{min} parameter defines a minimum threshold of novelty that has to be exceeded to enter the archive. The update interval is the frequency at which ρ_{min} is updated. Initially, ρ_{min} should be low, and then raised or lowered if too many or too few individuals are added to the archive in an update interval. Typically in other implementations, a small random chance of any individual being added to the archive is also set.

In the following implementation, a small initial ρ_{min} is selected relative to the behaviour space being explored and updated after a few hundred generations. ρ_{min} is dynamically raised by 20% if more than 10 individuals are added and ρ_{min} is lowered by 5% if no new individuals are added; these values are guided by the literature.

To maximise novelty, a selection pressure rewards individuals occupying sparsely populated regions in the behaviour space. To measure local sparsity, the average distance between an individual and its k -nearest neighbours is used. A region that is densely populated results in a small value of the average distance, and in a sparse region, a larger value. The sparseness ρ at point x is given by:

$$\rho(x) = \frac{1}{k} \sum_{i=1}^k dist(x, \xi_i) \quad (2)$$

where ξ_i are the k nearest neighbours of x .

The search processes is guided by the archive contents and the current behaviours in the population, but the archive does not provide a complete picture of all the behaviours explored. Throughout the search process the population tends to meander around existing behaviours until a new novel solution exceeding the novelty threshold is discovered. To take advantage of this local search, here all the explored behaviours are stored in a separate database D . This database stores all the information used to characterise the substrate later, and has no influence on the search, which uses only the archive.

4.2 Novelty Search Implementation

In the literature, novelty search is frequently combined with the Neural Evolution of Augmented Topologies (NEAT) [20, 28] representation; this neuro-evolutionary method focusses on adapting network topology and complexifying a definable structure. For the CHARC framework, a more generic implementation is desired: a form of evolutionary search algorithm that uses co-evolution

and speciation, but features a minimalistic implementation not based on any specific structure or representation. For these reasons, an adaptation of the steady-state Microbial Genetic Algorithm (MGA) [11] combined with novelty search is used here. The MGA is a genetic algorithm reduced to its basics, featuring horizontal gene transfer (through bacterial conjugation) and asynchronous changes in population where individuals can survive long periods.

To apply the MGA to the problem a number of adaptations are required. Caching fitness values in the standard steady-state fashion is not possible, as fitness is relative to other solutions found and stored in the growing archive. In this implementation, no individual fitnesses are stored across generations, however the same steady-state population dynamics are kept, i.e. individuals are not culled, and may persist across many generations.

An overview of the evolutionary loop is given in Fig. 3. The complete process is also outlined in pseudocode in Algorithm 1.

At the beginning of the search process, a random population is created. In the population, both the substrate configurations and the resulting behaviours B are stored. This initial population is then added to the archive A and database D .

At step 1, tournament selection with a tournament size of two is used. To ensure speciation, the first parent is picked at random and the second is chosen within some proximity to the other determined by the MGA parameter *deme size*. In this step, the fitness values (novelty) of both behaviours are calculated relative to population P and archive A . The individual with the larger distance, that is, occupying the less dense region of the behaviour space, is adjudged the winner. This elicits the selection pressure towards novel solutions. The microbial GA differs from other conventional GAs as the weaker (here, less novel) individual becomes “infected” by the stronger (more novel) one, replacing its original self in the population.

At step 2, the configurations of both behaviours are retrieved and manipulated. This constitutes the infection and mutation phase. In the infection phase, the weaker parent undergoes horizontal gene transfer becoming a percentage of the winner and loser. The genetic information of the weaker parent does not disappear in this process, as some percentage defined by the recombination rate parameter remains intact. In the mutation phase, the weaker parent undergoes multiple point-mutations, becoming the new offspring.

At step 3, the configuration of the new offspring is untested, therefore the behaviour B_{Child} of the individual needs to be updated. At steps 4a and 4b, the off-

spring’s behaviour and configuration are added to the database D and it replaces the loser in the population P .

At the last step 4c, the fitness/novelty of the offspring B_{Child} is compared to both the current population P and archive A . If the novelty of the offspring exceeds the novelty threshold ρ_{min} , the behaviour B_{Child} (configuration is not needed) is added to the archive A .

Overall, three fitness values are calculated at each generation. Two fitness evaluations occur in the selection phase and a third fitness evaluation is carried out on the offspring, in order to update the archive. The computational complexity of the fitness function is $O(nd + kn)$ using an exhaustive k -nearest neighbour search. As the dimension d of the archive/behaviour space is small ($d = 3$ property measures in the later example), the number of k -neighbours (here $k = 15$) has the dominant effect. This value of k is chosen experimentally; larger k -values improve accuracy but increase run time. As the archive size increases, run time increases proportional to archive size n . To reduce complexity, Lehman and Stanley [20] describe a method to bound the archive using a limited stack size. They find that removing the earliest explored behaviours, which may result in some limited backtracking, often results in minimal loss to exploration performance.

5 Applied to Echo State Networks

In this section, we begin by defining the behaviour space of interest and characterising the quality of the reference substrate. The chosen reference substrate is the *virtual* echo state network (ESN) substrate.

The reservoir model can represent any excitable nonlinear medium that produces a high-dimensional projection of the input $u(\cdot)$ into reservoir states $x(\cdot)$. In a conventional ESN, the reservoir state update equation $x(t)$ is represented as:

$$x(t) = f(W_{in}u(t) + Wx(t-1) + W_{fb}y(t)) \quad (3)$$

where f is the neuron activation function (typically a *tanh* function) and the weight matrices (W_{in} , W , and W_{fb}) are collections of connection weights to inputs (W_{in}), internal neurons (W), and from the output to internal neurons (W_{fb}); in many cases W_{fb} is unused. The final trained output $y(t)$ is given when the reservoir states $x(t)$ are combined with the trained readout layer W_{out} :

$$y(t) = W_{out}x(t) \quad (4)$$

A practical guide to creating and training ESNs is given in [22], including variations such as leaky-integrator ESNs.

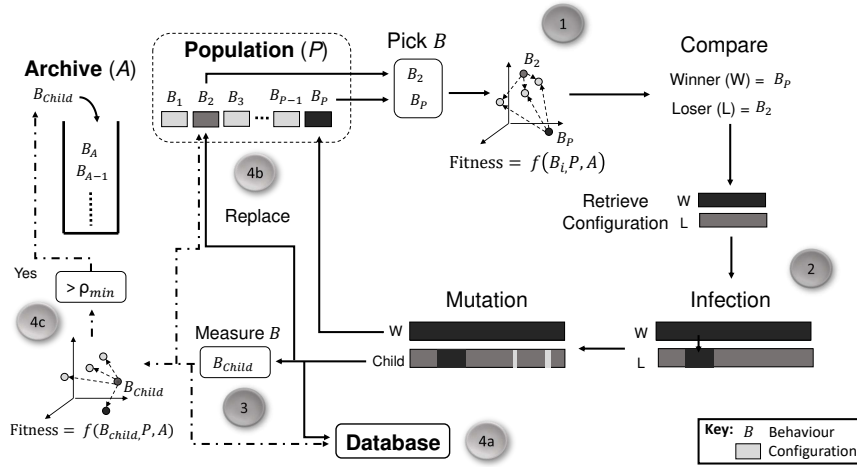


Fig. 3: Adapted microbial GA with novelty search.

Algorithm 1 Novelty search with microbial GA algorithm

```

pop ← random ▷ initial random population list
length P
A ← pop ▷ archive initialised
D ← pop ▷ database initialised
while searching do
    i :∈ 1..popSize ▷ parent 1 from pop
    j :∈ deme i ▷ parent 2 from deme
    if  $f(pop(i), A, pop) > f(pop(j), A, pop)$  then
        winner, loser ← i, j ▷ fitness is novelty
    else
        winner, loser ← j, i
    child ← infection(winner, loser)
    child ← mutation(child)
    pop(loser) ← child
    if child is sufficiently novel then
        add child to A
    add child to D
    if generation ==  $n \times update_{gen}$  then
        update novelty threshold  $\rho_{min}$ 

```

As a basic demonstration of the framework, a three-dimensional space is chosen using properties/metrics described in section 3: memory capacity (MC), kernel quality rank (KR) and generalisation rank (GR). These three measures capture different aspects of the reservoir, both chaos and order, and are simple to apply to physical systems.

To evaluate and validate the framework, multiple ESN network sizes are evolved and assessed against a control comprising random populations the same size as D . The four ESN network sizes chosen are: 25, 50, 100, and 200 node. This provides a small spectrum to assess

the framework with, from simple to more complicated reservoirs.

If novelty search performs well, i.e. better than random at covering the behaviour space, with each ESN network size, it might be possible to extrapolate that the same will hold true of other network sizes, and possibly different substrates.

5.1 Measuring Quality

To evaluate the quality of each ESN network size, a simple metric (Eqns. (5) and (6)) measures how much of the behaviour space is covered: greater coverage implies a greater degree of dynamical freedom. Statistical measures of dispersion such as standard deviation, variance, mean absolute deviation and inter-quartile range are not particularly suitable: they downplay outliers, whereas we want to push the boundaries of the region explored. Instead, the behaviour space is divided into discrete volumes, representing these ‘behaviour voxels’, and the quality measure defined counts how many ‘behaviour voxels’ are occupied: the more such behaviours, the larger the volume of space explored.

In the $3d$ example, this discretised behaviour space is captured by a cube represented by the $3d$ array $B_{i,j,k}$; where the coordinate j captures discretised memory capacity values (a continuous valued measure) and i and k capture the kernel and generalisation rank values (already discrete); the assignment of metrics to specific coordinates is arbitrary and does not affect exploration. The appropriate size of B can be deduced from measurement constraints and from the MC bounds imposed on ESNs: $0 \leq KR \leq N, 0 \leq GR \leq N$ and

$0 \leq MC \leq N$. A simple discretisation of MC is chosen, taking the integer part, so B is an $(N + 1)^3$ array.

B is defined as follows. $B_{i,j,k} = 1$ (occupied) if there is at least one individual in the database D with kernel rank $KR = i$, discretised memory capacity $MC = j$ and generalisation rank $GR = k$; otherwise $B_{i,j,k} = 0$ (unoccupied):

$$B_{i,j,k} \leftarrow \exists d \in D \text{ s.t. } KR(d) = i \quad (5)$$

$$\wedge \lfloor MC(d) \rfloor = j$$

$$\wedge GR(d) = k$$

The total space covered, θ , in an evolutionary run is the number of occupied points in $B_{i,j,k}$; $0 \leq \theta \leq N^3$:

$$\theta = \sum_{i=0}^N \sum_{j=0}^N \sum_{k=0}^N B_{i,j,k} \quad (6)$$

The behavioural voxel occupancy measure works well for this $3d$ case. For higher dimensional behaviour spaces (ones based on more properties/metrics) where the total number of voxels is considerably larger, the representation and measure may need adapting.

5.2 Experimental Parameters

In the following experiments, regardless of ESN network size etc., the same restrictions are placed on ESN parameter ranges and weights, and the same weight initiation processes is applied. For example, global parameters ranges include: an internal weight matrix (W) scaling between $[0, 2]$, scaling of the input weight matrix (W_{in}) between $[-1, 1]$, leak rate $[0, 1]$, and the sparseness of W $[0,1]$. For both random and novelty search, at creation a reservoir has each global parameter drawn from a uniform random distribution, as well as input weights and internal weights drawn uniformly from other ranges; W_{in} between $[-1, 1]$ and W between $[-0.5, 0.5]$.

For the evolutionary algorithm, the following MGA parameters are selected from preliminary experiments: *population size* = 200, *deme* = 40, *recombination rate* = 1, *mutation rate* = 0.2, $\rho_{min} = 3$, and ρ_{min} update = 200 generations.

To compare novelty search and the random control, 10 runs are conducted, with a limit of 2000 generations (full loops of Fig. 3) for novelty search, and 2000 randomly initialised reservoirs for random search.

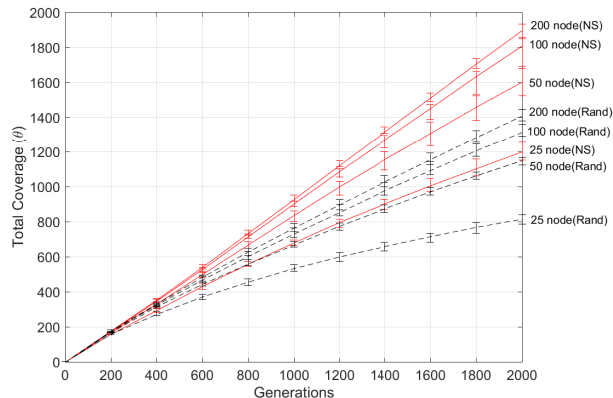


Fig. 4: Average coverage (over 10 runs) of behaviour space against number of generations. Error bars show minimum and maximum coverage.

5.3 Results: Random Versus Novelty Search

For every ESN network size, novelty search is able to explore a greater area of the behaviour space than the control (random search) in the same time. The total coverage (θ) of the behaviour space versus generations (or database size) is shown in Fig. 4. The results show that, with more generations novelty search can continue to explore an even greater area than random search, increasing linearly with the number of generations.

Fig. 5 shows all 10 runs of two network sizes (25 and 200) plotted in the behaviour space, helping visualise the difference in coverage. Random search appears to produce similar patterns in the behaviour space with different network sizes. These patterns include sparse regions that are difficult to occupy when uniformly sampling the parameter space, highlighting how deceptive the behaviour space is compared to the parameter space. Novelty search on the other hand covers the behaviour space more uniformly, both filling sparse regions and expanding beyond the region covered by random search. The difference in coverage between the two methods also becomes more distinct with an increase in network size.

Looking at Fig. 4 there appears to be a similar linear relationship between coverage (new behaviours) and generations across all ESN network sizes. However, network size affects coverage in different ways, e.g. the rate of coverage between 100 and 200 nodes does not increase as much between 25 and 50 nodes. To better understand this, each experiment is replotted using the rate r of total coverage θ per 200 generations; the number of generations at which coverage was recorded and calculated. For example, a rate $r = 1$ would in-

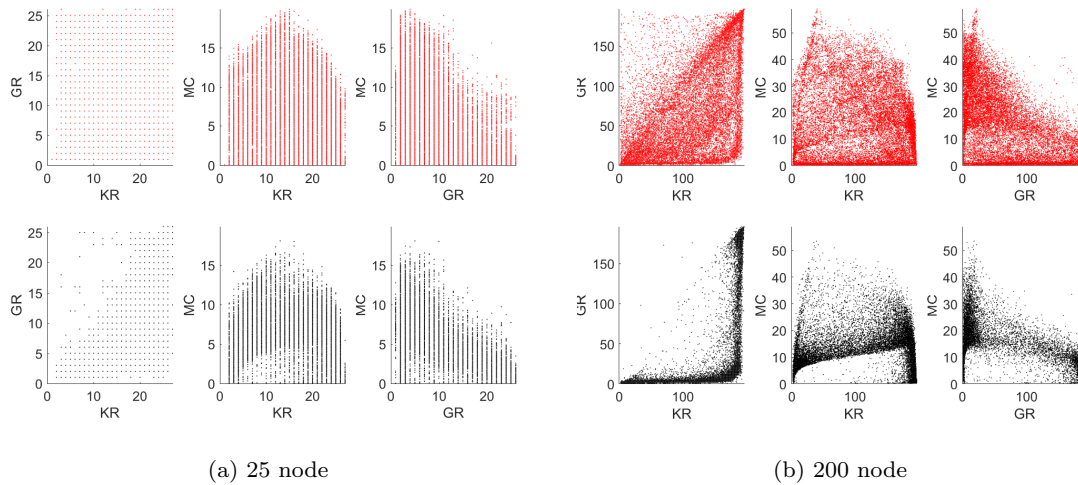


Fig. 5: Plot of all behaviours found through novelty search (Red, Top row) and all behaviours found through random search (Black, Bottom row).

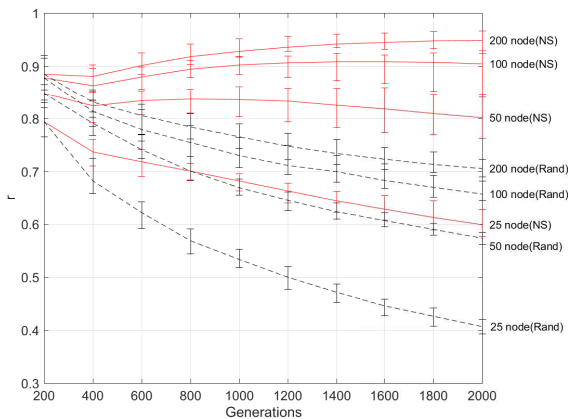


Fig. 6: A comparison of all experiments shown in fig. 4 when plotted as the rate ($r = \theta/\text{generation}$).

dicating a new behaviour was found every generation, and a rate $r = 0$ would show exploration halted. In Fig. 6, we see that smaller networks find fewer novel solutions per generation and larger networks maintain or increase slightly. This is most likely a product of larger networks having more adjustable parameters (weights). This leads to a higher dynamical degrees of freedom and therefore more distinct behaviours can be found per generation, resulting in smaller networks converging faster to $r = 0$.

6 Representation and Predicting Performance

At this stage, we have described how to define the behaviour space, how to explore/map it and how to use coverage (ϕ) as measure of substrate “quality”. Now we investigate how well the behaviour space represents computation in substrates and what relationship properties have to task performance.

Under Abstraction/Representation (A/R) theory [12], if the reservoir representation provides a faithful abstract representation of the substrate, it should be possible to *predict* how the reservoir states will evolve. In the context of the CHARC framework, we expand the A/R view, and hypothesise that if the measures of behavioural properties are substrate-independent and if relationships between properties and task performance can be learnt, then it should be possible to predict the performance of one substrate based on another substrate that exhibits similar behaviour values.

To assess whether the property measures provide a suitable representation, here we attempt to learn and model the relationship between properties and task performance. How well this relationship can be learnt will indicate how well the properties represent computation within the substrate.

6.1 Prediction Tasks

The property-performance relationship across all tasks is non-trivial. However, relationships between individual tasks and properties are sometimes simple. To predict performance, four benchmark tasks are selected

based on dissimilar requirements of reservoir properties: the common non-linear autoregressive moving average (NARMA) task with a 10-th and a 30-th order time-lag; the Santa Fe laser time-series prediction task; and the non-linear channel equalisation (NCE).

The NARMA task evaluates a reservoir’s ability to model an n -th order highly non-linear dynamical system where the system state depends on the driving input and state history. The task contains both non-linearity and long-term dependencies created by the n -th order time-lag. An n -th ordered NARMA task predicts the output $y(n + 1)$ given by Eqn. (7) when supplied with $u(n)$ from a uniform distribution of interval $[0, 0.5]$. For the 10-th order system parameters are: $\alpha = 0.3, \beta = 0.05, \delta = 0.1$, and $\gamma = 1$; for the 30-th order system: $\alpha = 0.2, \beta = 0.004, \delta = 0.001$, and $\gamma = 1$.

$$y(t + 1) = \gamma \left(\alpha y(t) + \beta y(t) \left(\sum_{i=0}^{n-1} y(t - i) \right) + 1.5u(t - 9)u(t) + \delta \right) \quad (7)$$

The laser time-series prediction task predicts the next value of the Santa Fe time-series Competition Data (dataset A)². The dataset holds original source data recorded from a Far-Infrared-Laser in a chaotic state.

The Non-linear Channel Equalisation task introduced in [15] has benchmarked both simulated and physical reservoir systems [24]. The task reconstructs the original *i.i.d* signal $d(n)$ of a noisy non-linear wireless communication channel, given the output $u(n)$ of the channel. To construct reservoir input $u(n)$ (see Eqn. 9) $d(n)$ is randomly generated from $-3, -1, +1, +3$ and placed through Eqn. 8:

$$\begin{aligned} q(n) = & 0.08d(n + 2) - 0.12d(n + 1) + d(n) \\ & + 0.18d(n - 1) - 0.1d(n - 2) \\ & + 0.091d(n - 3) - 0.05d(n - 4) \\ & + 0.04d(n - 5) + 0.03d(n - 6) + 0.01d(n - 7) \end{aligned} \quad (8)$$

$$u(n) = q(n) + 0.036q(n)^2 - 0.011q(n)^3 \quad (9)$$

Following [15], the input $u(n)$ signal is shifted +30 and the desired task output is $d(t - 2)$.

² Dataset available at UCI Machine Learning Repository [33].

6.2 Experimental Set-up

At the evaluation level, the database for the reference substrate is assessed on tasks, providing a target dataset.

To model the relationships, feed-forward neural networks (FFNN) of 100-neurons are configured for regression, i.e., given behaviours from the novelty search database, predict performance on a task. The inputs to the FFNN are: MC (continuous-valued), KR and GR (discrete values). The output of the network is task performance (continuous-valued), recorded as the normalised mean squared error (NMSE) of the reservoir with the corresponding input behaviour.

To train the FFNN, the Levenberg-Marquardt algorithm [21] is used for 1000 epochs, with the training dataset set as 70% of the data (database D), 15% for validation and 15% set aside for testing. To gather statistics, 20 FFNNs are trained and tested for every test.

6.3 Prediction Results

In this section FFNNs are trained, per task and per ESN size.

If the property measures provide a good representation of the substrate, the mean prediction error of all trained models should be low and similar to each other, i.e. shows a relationship is present, not too difficult to model and holds when the substrate is changed (in this case a different size ESN). However, there will be some deviation in error between models trained with databases holding different behaviours, because having a greater or smaller behavioural range can result in an increase or decrease in complexity of the modelled relationships. For example, reservoirs in the behaviour space around $KR = GR = MC \leq 25$ tend to have similar poor performances on the NARMA-30 task because they do not meet the minimum requirement ($MC \geq 30$). This means the NARMA-30 task is easier to model with the database created by the 25 node ESNs. When databases with larger ESNs are used to model the relationship, prediction error will likely increase.

This is not always true, for some tasks to accurately model the relationship requires a greater variety of behaviours than smaller ESNs can provide (e.g. for the non-linear channel equalisation task). Therefore, an FFNN trained on a database provided by the 200 node ESNs will perform better than one provided by the smaller 25 node ESNs. This is one example of where the non-trivial problem of relating properties to performance presents itself.

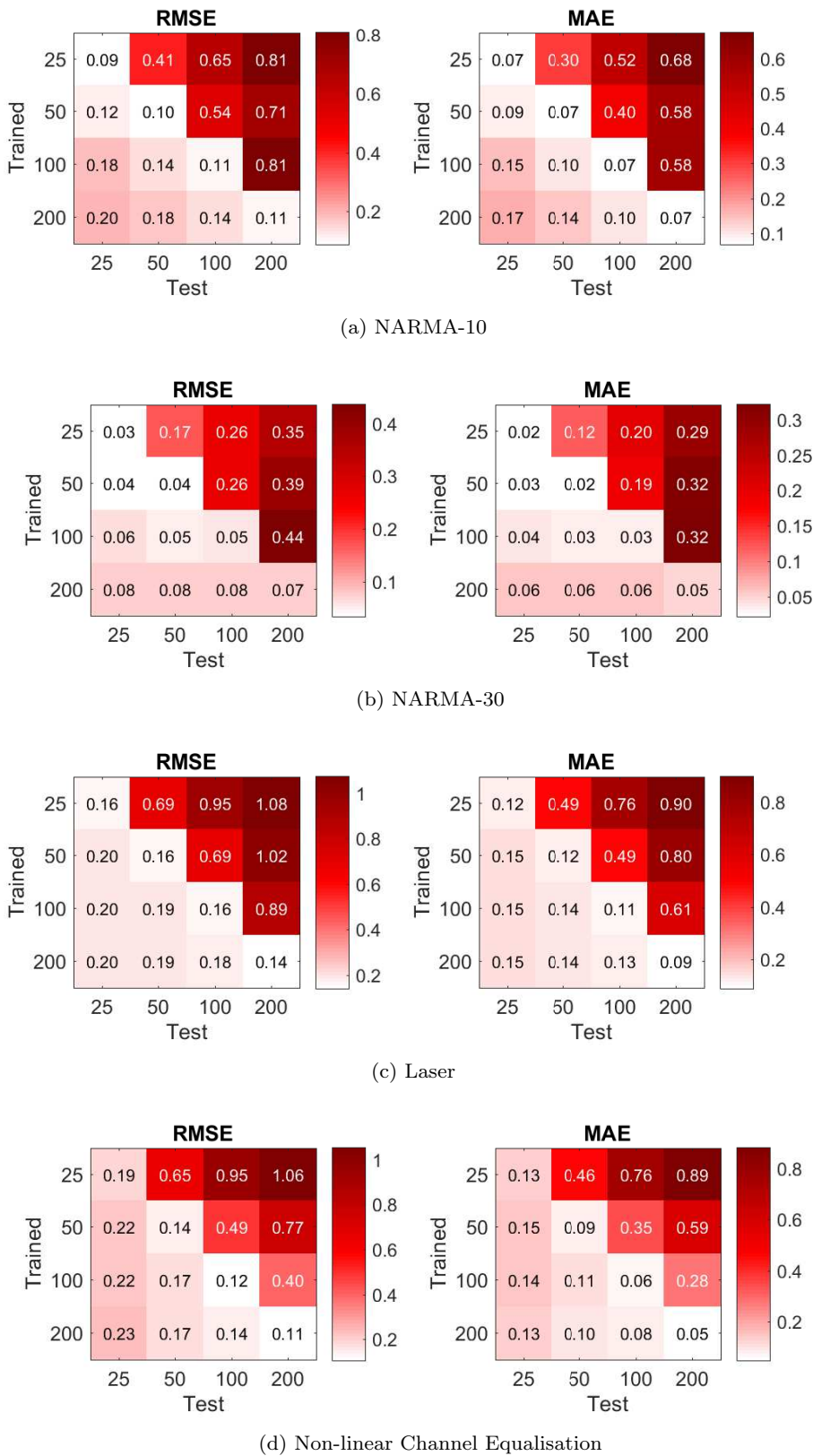


Fig. 7: Mean performance of 20 FFNNs trained and tested on each database (different sized ESN), per task .

In Fig. 7, the results of all FFNNs trained on different tasks and databases are shown as heat maps. The mean test error (either root mean square error *RMSE* or mean absolute error *MAE*) of FFNNs trained and tested using the same database are given along the diagonal, where train ESN size $y =$ test ESN size x . When focusing on these results, we tend to see a small deviation between model accuracies, however there are some exceptions, as previously discussed.

Also shown in Fig. 7 is how well the FFNNs predict the task performance of reservoirs from different databases, i.e. when train ESN size $y \neq$ test ESN size x . The generalisation to other databases, holding possibly new behaviours, varies depending again on the task. In general, if train $y \geq$ test x , the FFNNs mean prediction error is usually greater than the original, but considerably less than train $y \leq$ test x .

The most likely cause of the first case, when train $y \geq$ test x , is a loss in granularity of the learnt relationships between databases. Stretching the same number of data-points to a larger area will result in fewer examples to learn from in each region of the behaviour space, i.e., a loss in resolution that might be important to the learnt problem. This can be thought of as a sampling bias or quantisation effect. The second case is simply a result of new behaviours predicted based on no prior data, resulting in a poor prediction.

Overall, the mean prediction errors appear consistent, with small deviations as the behaviour range increases or decreases. This would suggest the behaviour space and property measures are a somewhat faithful representation of the substrate’s computational capabilities.

7 Completing the Framework

At this stage, the framework definition is almost complete. The hypothesis so far is that the behaviour space of the substrate can be explored and quality can be measured with a faithful representation of the substrates computational mechanisms.

The final part of the framework is: i) to apply each level to a new uncharacterised substrate, ii) to evaluate quality w.r.t. the reference substrate, iii) to use the predictive network from section 6 to predict task performance, and iv) to use the prediction to validate the substrate-independence of the behaviour space.

The substrate that is used to demonstrate this part of the framework is the physical substrate investigated in [4, 5, 6]. The substrate comprises a carbon nanotube-polymer composite, forming random networks of semi-conducting nanotubes suspended in a insulating polymer, deposited onto a micro-electrode array.

In previous work [4], a small amount of characterisation has been done showing that even the best substrate (1% concentration of carbon nanotubes w.r.t. weight mixed with poly-butyl-methacrylate) typically exhibits low memory capacity, despite different methods of configuration. This leads to overall modest performances, but encouraging when compared relative to size, on benchmark tasks such as NARMA-10 [4] and the Santa Fe laser time-series prediction task [5].

The challenging aspect of characterising this black-box substrate is due to its disordered structure and self-organisation during the fabrication process, making it impractical (or even impossible for the general case) to model internal workings. Originally, the CNT/polymer was proposed as a sandpit substrate to discover whether computer-controlled evolution could exploit a rich source of physical complexity to solve computational problems [2]. Due to its computational shortcomings, it provides a challenging substrate for the CHARC framework and is useful for demonstrating the process.

The training and evaluation of the substrate is conducted on a digital computer. Inputs and representative reservoir states of the substrate are supplied as voltage signals. The adaptable parameters for evolution are the number of input-outputs, input signal gain (equivalent to input weights), a set of static configuration voltages (values and location), and location of any ground connections. Configuration voltages act as local or global biases, perturbing the substrate into a dynamical state that conditions the task input signal.

An advantage of physical substrate-based reservoirs is that computational speed of a trained reservoir is limited only by interface hardware and physical response time, which can potentially be all analogue. However, the training process can take considerably longer as multiple runs for statistical tests are often needed. With this substrate in particular, there are many unstable configurations. Therefore, extra evaluations are required to measure statistical stability in order to discard unstable signals from the training process. For this reason alone, the ability to predict performance across substrates and bypass the task training process would be decidedly beneficial.

7.1 Experimental Parameters

The same GA/NS parameters applied to the virtual ESN substrate are reused with the physical substrate. These are: generations limited to 2,000; *population size* = 200; *deme* = 40; *recombination rate* = 1; *mutation rate* = 0.2; $\rho_{min} = 3$; and ρ_{min} *update* = 200 generations. Five runs are conducted here, as the time to train increases significantly.

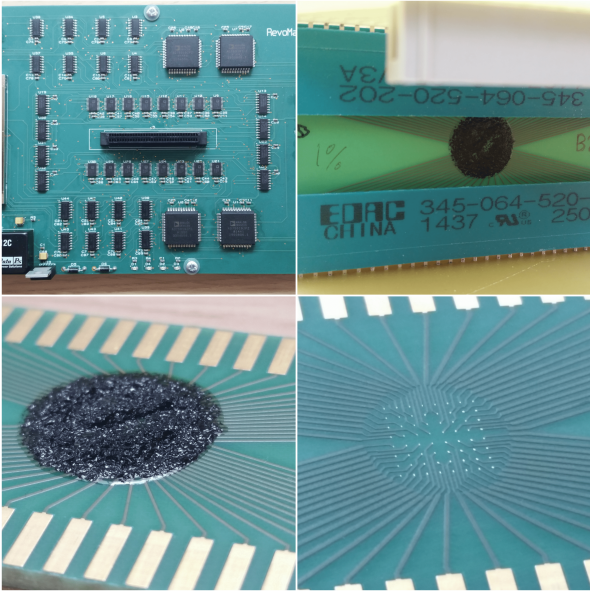


Fig. 8: Hardware Reservoir System. Micro-electrode housing, routing switch board and CNT/polymer deposited onto PCB electrode array.

In this work, a 1% carbon nanotube poly-butyl-methacrylate (CNT/PBMA) mixture substrate is investigated. The substrate was mixed and drop cast onto a micro-electrode array using the same process in [4, 5, 6]. The electrode array applied here comprises 64 electrodes (contact sizes of $100\mu\text{m}$ and spacings of $600\mu\text{m}$ between contacts) deposited onto a FR-4 PCB using a chemical process that places Nickel and then a layer of Gold (see Fig. 8).

Two National Instruments DAQ cards perform measurements and output analogue voltages; a PCI-6225 (16-Bit, 250 KS/s, with 80 analogue inputs), and PCI-6723 (13-Bit, 800KS/s, with 32 analogue outputs). Both cards communicate to a desktop PC through a session-based interface in MATLAB. The PCI-6723 supplies an additional 8 digital I/O lines to a custom routing board to program on-board switches and synchronise the cards.

7.2 Quality of Physical Substrate

The results of the evolved physical substrate suggest a poorer quality and a limited dynamical degrees-of-freedom compared to the reference ESN substrate. This is in agreement with previous work comparing the physical substrate to ESNs in [4, 5, 6].

When overlaying the explored behaviour space of the physical substrate on top of the reference substrate the difference becomes distinct (see Fig. 9). The average

coverage of the physical substrate in Fig. 10b is only a third of even the 25 node ESN experiment.

The effective number of equivalent nodes an ESN should possess to compare fairly to this physical substrate is unknown. As a rough guide, the physical substrate has up to 64 state observations at any one time, therefore it might be considered generous to compare it to a 25 node ESN, however, this assumes similar non-linear nodes, network dynamics, connectivity, etc.

In general, the search struggles to find configurations beyond a memory capacity of 5, reaching what appears to be a memory capacity limit. The bounds on the ranks are also small given only a small number of inputs are typically in use. This would suggest the substrate struggles to exhibit enough (stable) non-linear behaviour to create a strong non-linear projection, and effectively store recent input and state information.

To investigate if the substrate limits are reached, random search is also conducted (Fig. 10) and the *novelty rate* is plotted (Fig. 11).

When comparing random search to novelty search in Fig. 10b, novelty search on average appears to explore slightly faster and further for roughly 1200 generations but then begins to plateau around $\theta = 360$.

Looking at the novelty rate of the physical substrate (Fig. 11), we see that the rate of new behaviours drops considerably over the course of the evolutionary runs. For random search this drops faster, then novelty search reaches the same low novelty rate after 1600 generations, maintaining the same decrease in novelty rate as random.

Fig. 11 also shows how the physical substrate compares to different ESN sizes. The physical substrate rate r starts much lower, around $r = 0.5$ compared to the ESNs ($r \geq 0.8$). This itself begins to suggest the physical substrate has a small degree of dynamical freedom, as less than half of the 200 original random behaviours were unique.

The rate r for both random and novelty search then both decrease at a similar rate, however, this is not the case for the ESNs. Although both decrease as the behaviour space is explored/filled, the difference in r gradually increases, with random typically dropping early and continuing to get smaller.

This small difference in r across the generations, combined with a low starting and low finishing r , plus a struggle to find unique behaviours (Fig. 10a), suggests the physical substrate's dynamical limits have almost been reached. Although this shows the substrates are limited for reservoir computing, it more importantly demonstrates that the framework can approximate and quantify the dynamical boundaries of the substrate.

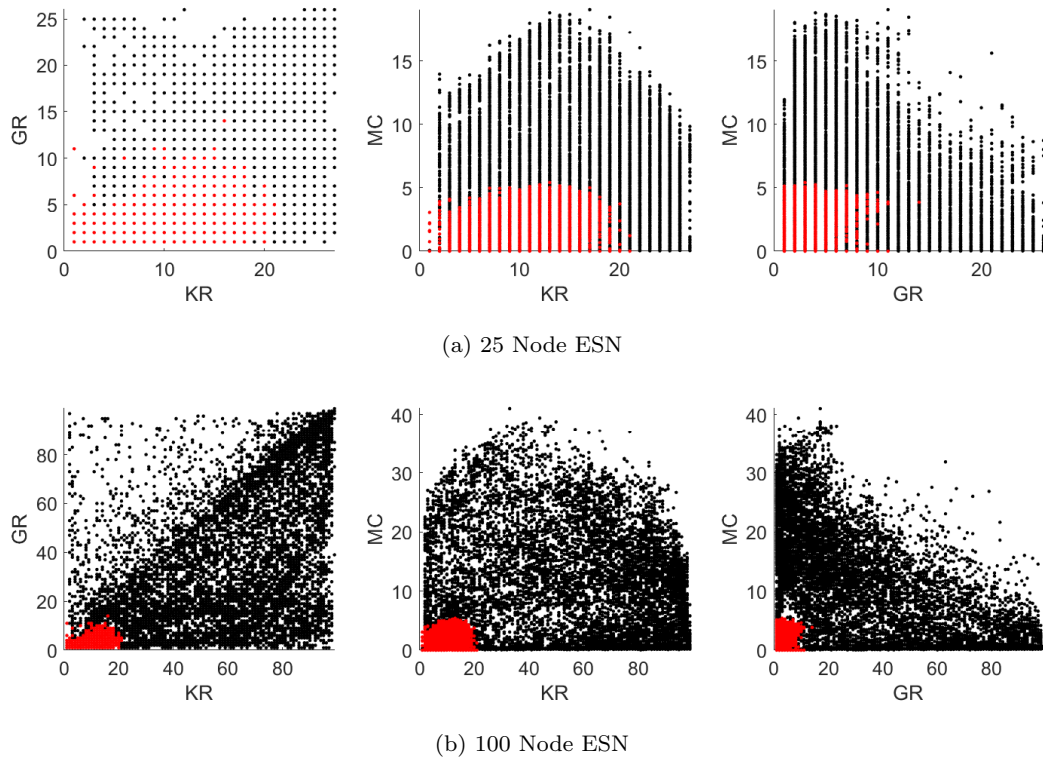


Fig. 9: Carbon nanotube composite behaviour space (Red) superimposed on ESN behaviour space (Black)

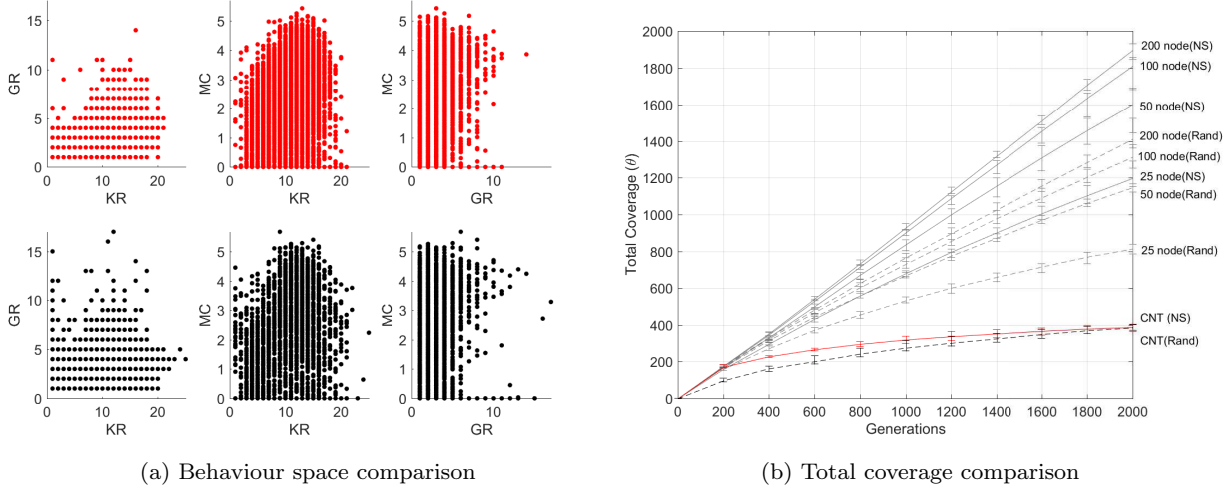


Fig. 10: Physical substrate: a) novelty search (red) versus random search (black) in the behaviour space, b) and for total coverage.

7.3 Prediction transfer

At the *Application* level of the framework, we try to predict the task performance of another similarly behaved substrate based on models of the reference substrate.

To do this, all the FFNNs from the learning level (section 6) are used to predict the performance of the

carbon nanotube composite. Then, using the difference Δ between prediction errors of both substrates we measure how accurately the modelled relationships translate across substrates.

The measure Δ represents the difference between the trained *ESN substrates* mean test error and new *physical substrates* mean test error. To support the hy-

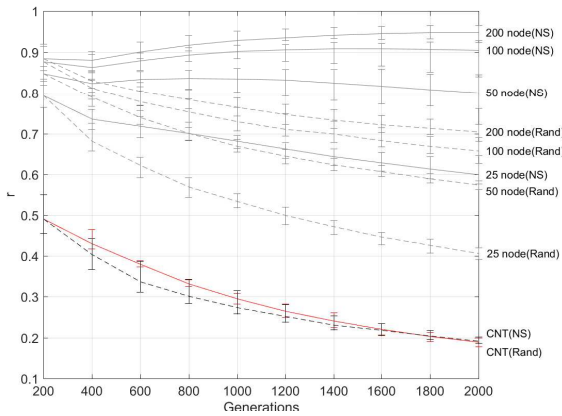


Fig. 11: The novelty rate ($r = \theta/\text{generation}$) of the physical substrate and compared with different ESN sizes.

pothesis that the behaviour space and the property measures have substrate independence, a low Δ close to zero is desired.

Prior to the prediction tests, the databases created from five runs are assessed on the four tasks, creating a five test sets per task for the physical substrate. This is done here for evaluation purposes only, if the framework is found to be substrate-independent the task evaluation step on the physical substrate can be ignored.

The mean prediction error of all FFNNs, trained with different size ESNs and tested on the substrate for all four tasks, are given in Fig. 12. The first column in each grid provides the test error of the FFNNs when predicting the physical substrates task performance.

On every task except non-linear channel equalisation, the FFNNs predict the task errors of the physical substrate almost as well as an FFNN trained and tested with the same ESN size. As seen in Fig. 12d, the FFNNs trained to predict the non-linear channel equalisation task struggle to accurately translate the learnt relationships to the physical substrate. Although the prediction error is worse than if trained and tested on the same size ESN, the error is considerably lower than if an FFNN was trained using a small ESN size then tested on a larger ESN size.

Another interesting result is that, in some cases, the FFNN trained with the 200 node ESN more accurately predicts the performance of the physical substrate than a FFNN trained with a smaller sized ESN. This is somewhat counter-intuitive because an FFNN trained with a 200 node ESN has fewer examples in the behaviour space occupied by the physical substrate, i.e. in the region within $KR = GR = MC \leq 25$, thus would likely over-generalise/under-fit this region, as it does when tested with the 25 node ESN.

All these trends are summarised in the Δ plot, Fig. 13. The average Δ (in units RMSE or MAE) is given for each task and for all FFNNs (20 per ESN size), applied to the physical substrate. For the first three tasks, small Δ 's close to zero are present in most cases, despite ESN size. However, as before, Δ increases significantly for the non-linear channel equalisation task.

The reasons why the modelled relationships of the non-linear channel equalisation task struggle to translate across substrates is still unknown and requires further investigation. However, a potential lead to explaining this is that the variation in performance across the behaviour space is very low, with reservoirs requiring very low metric values to perform well at this task. Therefore, the greater variation seen with the physical substrate is likely to be poorly modelled by the FFNNs trained with larger ESNs.

Overall, the results in this section indicate that the CHARC framework has a good level of substrate independence and that the behaviour space can accurately represent the substrate's computational properties. It also highlights again the non-trivial nature of the task-property relationship and how some tasks can be more difficult to model, or require extra thought and manipulation, than others.

These last results also show that the learning phase (and task evaluation) only needs to be performed with the reference substrate, once per task, leading to a significant reduction in time to evaluate and test new substrates on specific tasks. For example, one could evaluate the reference substrate's database on a new task, train an FFNN, then predict whether a previously explored physical substrate would be suitable, perform well on the task, or even what the best configuration would be, without having to test directly on the substrate. In the future, the learning phase may even be a shared venture, stored and shared in a repository for reference substrates, helping other practitioners quickly evaluate new substrates and their suitability for specific tasks.

8 Conclusion

A fundamental question in reservoir computing is: for a given task, what characteristics of a dynamical system or substrate are crucial for meaningful information processing? The CHARC framework tackles this question by focussing on the substrate rather than the specific task. In the process, solutions to two non-trivial problems are proposed; (i) how to characterise the quality of any substrate for reservoir computing; and (ii) how do computational properties relate to performance.

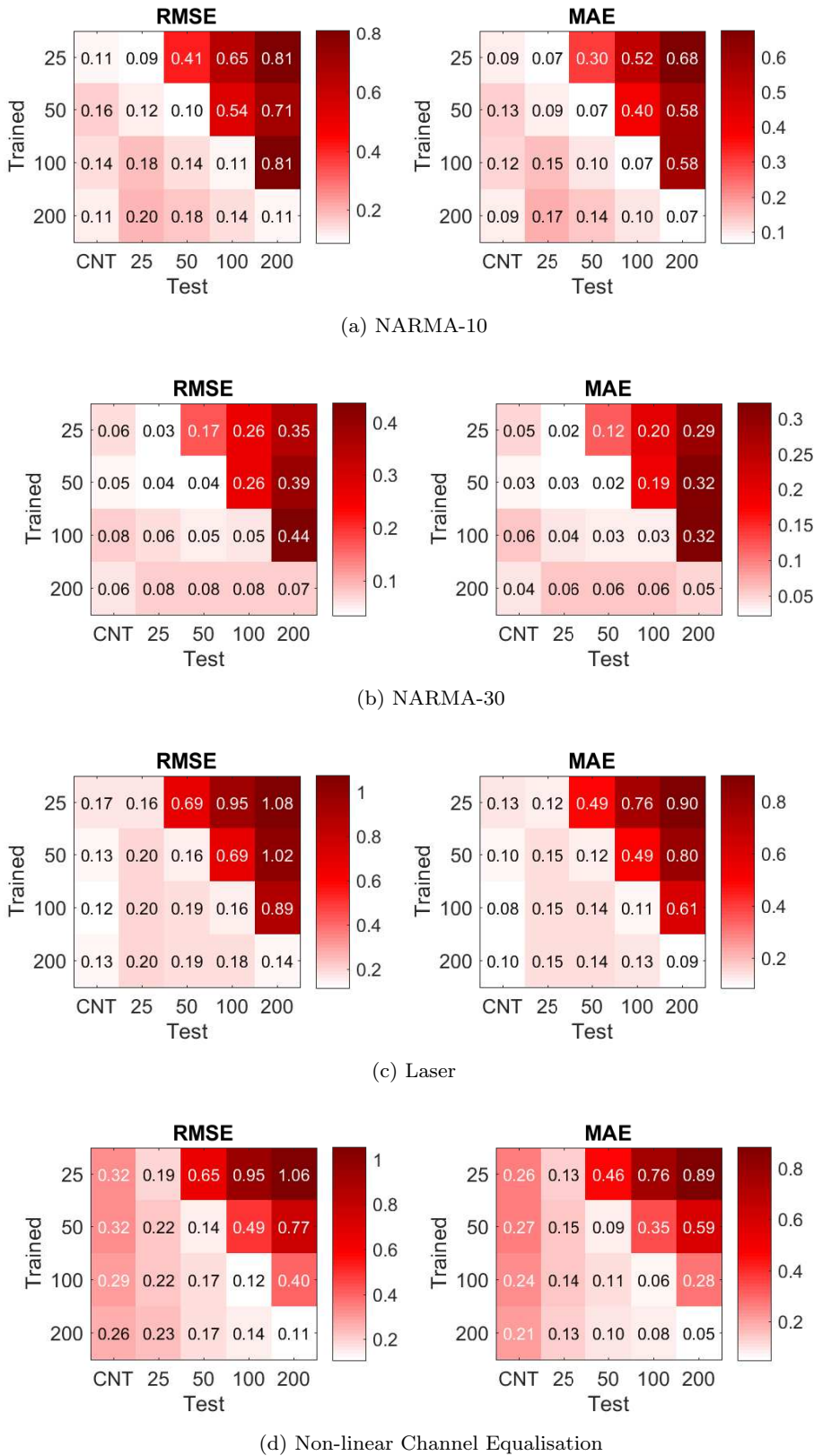


Fig. 12: FFNN prediction errors of all trained and tested ESNs, and tested on physical substrate.

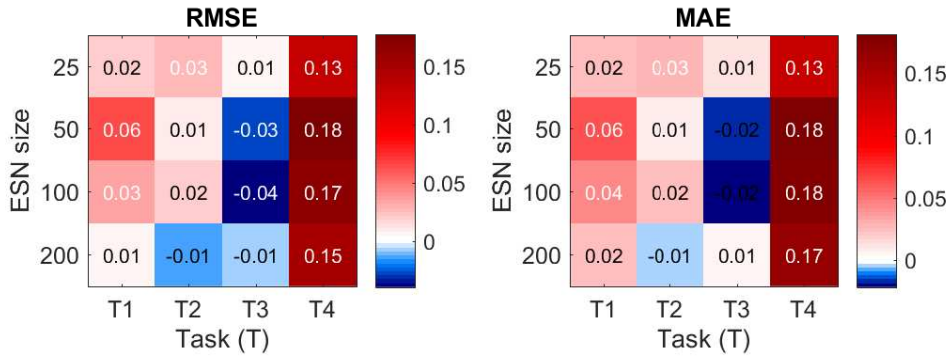


Fig. 13: The average difference Δ between prediction errors for ESNs trained/tested using same size and physical substrate, across the four tasks: T1 = NARMA-10, T2 = NARMA-30, T3= Laser, T4= Non-linear channel equalisation.).

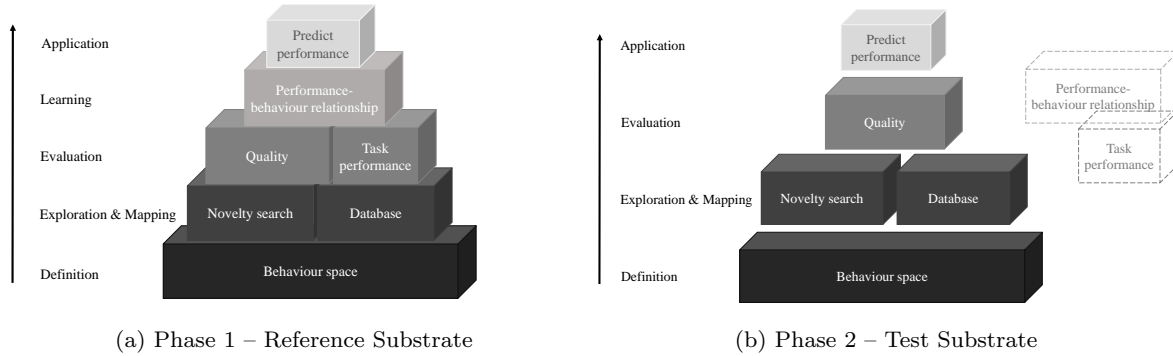


Fig. 14: Framework summary. a) Reference substrate mapped providing a reference quality and learnt relationships to predict performance. b) Test substrate is assessed for quality. No task assessment is necessary and learning phase is skipped. Performance of test substrate is predicted using learnt relationships from reference.

To fully utilise the framework, two phases must be completed. In the first phase (Fig. 14a), the lower levels of the framework are applied to the reference substrate, providing context to future quality measures on other substrates. Then the upper levels are applied to learn the transferable relationships between substrate behaviours and task performance, helping validate the framework and predicting performances of similarly behaved substrates. The second phase applies a reduced set of levels (Fig. 14b) to a new substrate for which a measure of quality is desired. The phase one learnt transferable relationships can be used to predict how well the second substrate could perform a task, without it having to be assessed directly on the task.

Throughout this work, each level of the framework has been described and tested. In some cases, adaptations and the removal of different levels/building blocks have also been discussed. This demonstrates both the flexibility and potential power of the framework where

building blocks can be changed, integrating new techniques or measures not currently available.

What was learnt whilst using the framework is that substrate limitations can be outlined (section 77.2). This helps explain why the carbon nanotube composite struggles to compete with echo state networks in previous work [4, 5].

Another feature highlighted is the non-trivial problem of relating properties to task performance (section 66.3), suggesting current measures of dynamical properties provide only a partial view of the full characteristics of reservoir systems. Therefore, in order to fully understand these systems we need to explore each property and its' relationship to others.

The CHARC framework provides a basic methodology to compare and evaluate any substrate for reservoir computing; something not achieved or widely discussed in the reservoir computing community. The advantage of having such a framework can only be speculated, but any method to compare systems, or even techniques

for perturbing or observing systems, has the chance to rapidly improve the design and implementation process. It potentially opens the reservoir computing field to a wider audience where new interesting unconventional computing substrates could appear, and even provides a goal for optimisation of current substrates.

Acknowledgements

This work was funded by a Defence Science and Technology Laboratory (DSTL) PhD studentship.

References

1. 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*, 2:468, 2011.
2. H. Broersma, F. Gomez, J. Miller, M. Petty, and G. Tufte. Nascence project: Nanoscale engineering for novel computation using evolution. *International Journal of Unconventional Computing*, 8(4):313–317, 2012.
3. L. Büsing, B. Schrauwen, and R. Legenstein. Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. *Neural Computation*, 22(5):1272–1311, 2010.
4. M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer. Evolving carbon nanotube reservoir computers. In *International Conference on Unconventional Computation and Natural Computation*, pages 49–61. Springer, 2016.
5. M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer. Reservoir computing in materio: An evaluation of configuration through evolution. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8, Dec 2016.
6. M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer. Reservoir computing in materio: A computational framework for in materio computing. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2178–2185, May 2017.
7. J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar. Information processing capacity of dynamical systems. *Scientific Reports*, 2, 2012.
8. C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu. Reservoir computing using dynamic memristors for temporal information processing. *Nature communications*, 8(1):2204, 2017.
9. K. Fujii and K. Nakajima. Harnessing disordered-ensemble quantum dynamics for machine learning. *Physical Review Applied*, 8(2):024030, 2017.
10. A. Goudarzi and C. Teuscher. Reservoir computing: Quo vadis? In *Proceedings of the 3rd ACM International Conference on Nanoscale Computing and Communication*, page 13. ACM, 2016.
11. I. Harvey. The microbial genetic algorithm. In *European Conference on Artificial Life*, pages 126–133. Springer, 2009.
12. C. Horsman, S. Stepney, R. C. Wagner, and V. Kendon. When does a physical system compute? *Proc. R. Soc. A*, 470(2169):20140182, 2014.
13. H. Jaeger. The “echo state” approach to analysing and training recurrent neural networks—with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, 148:34, 2001.
14. H. Jaeger. *Short term memory in echo state networks*. GMD-Forschungszentrum Informationstechnik, 2001.
15. H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
16. J. H. Jensen, E. Folven, and G. Tufte. Computation in artificial spin ice. In *Artificial Life Conference Proceedings*, pages 15–22. MIT Press, 2018.
17. R. Legenstein and W. Maass. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks*, 20(3):323–334, 2007.
18. J. Lehman and K. O. Stanley. Exploiting open-endedness to solve problems through the search for novelty. In *ALIFE*, pages 329–336, 2008.
19. J. Lehman and K. O. Stanley. Efficiently evolving programs through the search for novelty. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, pages 837–844. ACM, 2010.
20. J. Lehman and K. O. Stanley. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation*, 19(2):189–223, 2011.
21. K. Levenberg. A method for the solution of certain nonlinear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.
22. M. Lukoševičius. A practical guide to applying echo state networks. In *Neural Networks: Tricks of the Trade*, pages 659–686. Springer, 2012.
23. O. Obst, A. Trinchi, S. G. Hardin, M. Chadwick, I. Cole, T. H. Muster, N. Hoshcke, D. Ostry, D. Price, K. N. Pham, et al. Nano-scale reservoir computing. *Nano Communication Networks*, 4(4):189–196, 2013.
24. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar. Optoelectronic reservoir computing. *Scientific Reports*, 2, 2012.
25. D. Prychynenko, M. Sitte, K. Litzius, B. Krüger, G. Bourianoff, M. Kläui, J. Sinova, and K. Everschor-Sitte. Magnetic skyrmion as a nonlinear resistive element: A potential building block for reservoir computing. *Physical Review Applied*, 9(1):014034, 2018.
26. S. Risi, S. D. Vanderbleek, C. E. Hughes, and K. O. Stanley. How novelty search escapes the deceptive trap of learning to learn. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 153–160. ACM, 2009.
27. B. Schrauwen, D. Verstraeten, and J. Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European symposium on artificial neural networks*. Citeseer, 2007.
28. K. O. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127, 2002.
29. A. Z. Stieg, A. V. Avizienis, H. O. Sillin, C. Martin-Olmos, M. Aono, and J. K. Gimzewski. Emergent criticality in complex Turing B-type atomic switch networks. *Advanced Materials*, 24(2):286–293, 2012.
30. 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*, 547(7664):428, 2017.
31. K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman. Parallel reservoir computing using optical amplifiers. *Neural Networks, IEEE Transactions on*, 22(9):1469–1481, 2011.

-
32. D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391–403, 2007.
 33. A. Weigend. *The Santa Fe Time Series Competition Data: Data set A, Laser generated data, 1991* (accessed March, 2016).