Task-Independent Computational Abilities of Semiconductor Lasers with Delayed Optical Feedback for Reservoir Computing

Krishan Harkhoe [ID] **and Guy Van der Sande *** [ID]

Applied Physics Research Group, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium; Krishan.Harkhoe@vub.be
* Correspondence: guy.van.der.sande@vub.be

check for updates

Abstract: Reservoir computing has rekindled neuromorphic computing in photonics. One of the simplest technological implementations of reservoir computing consists of a semiconductor laser with delayed optical feedback. In this delay-based scheme, virtual nodes are distributed in time with a certain node distance and form a time-multiplexed network. The information processing performance of a semiconductor laser-based reservoir computing (RC) system is usually analysed by way of testing the laser-based reservoir computer on specific benchmark tasks. In this work, we will illustrate the optimal performance of the system on a chaotic time-series prediction benchmark. However, the goal is to analyse the reservoir's performance in a task-independent way. This is done by calculating the computational capacity, a measure for the total number of independent calculations that the system can handle. We focus on the dependence of the computational capacity on the specifics of the masking procedure. We find that the computational capacity depends strongly on the virtual node distance with an optimal node spacing of 30 ps. In addition, we show that the computational capacity can be further increased by allowing for a well chosen mismatch between delay and input data sample time.

Keywords: semiconductor laser; feedback; delay; reservoir computing; neuromorphic computing

1. Introduction

Artificial neural networks (ANN) have played a significant role in the current artificial intelligence (AI) boom, especially with the invention of ImageNet [1] as catalyst. ANNs may be efficient and versatile during operation but they require complex and time-consuming algorithms to train the connection weights in the large network that forms the ANN. When interested in processing tasks and data where the temporal evolution is key, standard feed-forward ANNs are not sufficient and one needs to turn to recurrent neural networks (RNNs). The training of RNNs is a nonlinear problem due to feedback loops in the network and is far more involved than the training algorithms of feed-forward networks. Reservoir computing (RC) is a paradigm that solves the training issue of RNNs in an efficient way.

RC offers a framework to exploit the transient dynamics within an RNN for performing useful computation. It has been demonstrated to have state-of-the-art performance for a range of tasks that are notoriously hard to solve by algorithmic approaches, e.g., speech and pattern recognition and nonlinear control. RC simplifies the training procedure for RNNs considerably. Its training procedure only acts on the output layer which consists of a linear combination of network states to generate the desired output signals. The connections of the RNN itself, which is now referred to as reservoir, remain fixed. During training, only the connections from the network to the output

layer are adjusted. Due to this simplification, RC is very suited as a framework for neuromorphic computing activities in photonics. Today, multiple photonic RC systems exist that can provide a practical yet powerful hardware substrate for neuromorphic computing [2]. Some examples include a network of semiconductor optical amplifiers [3,4], an integrated passive silicon circuit forming a very complex and random interferometer, with nonlinearity introduced in the readout stage [5] and a semiconductor laser network based on diffractive coupling [6]. All these implementations have one thing in common: they rely on a network of photonic nodes that are spatially distributed and can be measured simultaneously.

However, the reservoir is not required to be a networked structure. In fact, any dynamical system with a high dimensional state space can be considered as reservoir substrate. We consider here specifically a semiconductor laser with delayed feedback as reservoir substrate. The concept of delay-based RC, using only a single nonlinear node with delayed feedback, was introduced some years ago by Appeltant et al. [7]. Its operation boils down to a time-multiplexing with the delay arising from propagation in the external feedback loop, limiting the resulting processing speed. The system is easily scaled by tuning the delay length and only has one single physical node reducing the hardware complexity in photonic systems. The first working prototype was developed in electronics in 2011 by Appeltant et al. [7] and several performant optical systems followed quickly after that [8,9]. Brunner et al. [10] employed off-the-shelf telecom equipment to experiment with a single-mode semiconductor laser subjected to optical feedback. The delay time in his experiments was around 80 ns, which translates to a few meters of fiber. Recently, Takano et al. [11] have presented a photonic integrated circuit consisting of a distributed-feedback semiconductor laser, a semiconductor optical amplifier (SOA), a phase modulator, a short passive waveguide and an external mirror for optical feedback. The external cavity length in this system reached 10.6 mm, corresponding to a round-trip delay time of 254 ps. Several other types of semiconductor laser have been considered such as semiconductor ring lasers using the two available directional modes [12] and vertical-cavity surface-emitting employing the two polarization modes [13]. In this paper, we will only focus on single-mode Fabry-Péro type quantum-well semiconductor lasers.

The information processing performance of a semiconductor laser-based RC system is related to its dynamical behaviour both in the absence of external input and in the presence of said input. After the very first experiment by Brunner et al. [10], other works have focused on understanding the fundamental properties of semiconductor laser-based RC for non-linear prediction tasks. In Ref. [14], it has been shown that the conditions to achieve good predictive performance are given by the injection locking, consistency and memory properties of the system. More specifically, Bueno et al. found that the lowest prediction error for a non-linear prediction task occurs at the injection locking boundary. Note that in this work the laser was operating below or close to the solitary lasing threshold. Consistency, the ability of a system to have a similar response for similar input signals, is widely regarded as key for good reservoir computing performance [14,15].

While much attention was drawn to the specific dynamical regimes, much less attention was devoted to the actual pre-processing procedure, which implements the time-multiplexing. In delay-based RC, the laser response to data samples can be measured sequentially and these responses will form virtual nodes states. The nodes are now denoted as virtual as they exist in a time-multiplexed way rather than corresponding to real physical spatially distributed and interconnected nodes. Specifically, a lot of debate exists on what the optimal size of such a virtual node should be. In Ref. [7], a rule of thumb was suggested implying that an optimal choice of θ was around 20% of the internal timescale of the system. In the case of a semiconductor laser with delayed optical feedback and optical data injection, it is not readily clear which timescale should be taken. In Ref. [10], the timescale considered corresponds to the relaxation oscillations, and the choice for θ was 200 ps. Nguimdo et al. stated that the node distance could easily be reduced to 20 ps, claiming injection locking dynamics as the underlying reason [16]. In addition, it is not clear if it is better to fit all virtual node states exactly into one delay line length as in [7,10] or that a mismatch is beneficial as in, for

example, Ref. [9]. These considerations will have their impact on the processing speed of the RC. The effect of the parameters of the masking method on the computational performance of the laser-based RC can be task-dependent. In the worst case scenario, a bad choice of benchmark can obscure the most important trends. Therefore, we will analyse these dependencies in a task-independent way. To this aim, we will perform calculations of the computational capacity from numerical timetraces obtained from rate equations. Contrary to [14], we will focus on above solitary laser threshold behavior and limit ourselves to the zero detuning regime.

In Section 2, we introduce the rate equation model that is used for all numerical simulations in this work. We will also review the necessary pre-processing or masking procedure of delay-based reservoir computing and the training procedure. By way of example, we analyse the optimal parameters of the semiconductor laser with delayed optical feedback to tackle a chaotic time-series prediction task in Section 3. Finally, in Section 4, we analyse the computational performance of the laser-based reservoir in a task-independent way by calculating computational capacities associated to a set of polynomial target functions. We investigate how the computational capacity depends on the virtual node distance and mask length which are defined in the masking scheme.

2. Reservoir Computing with a Semiconductor Laser and Delayed Feedback

2.1. Rate Equation Model

We confine our laser model to the case of a single section Fabry-Pérot device, supporting only a single transversal mode, single longitudinal mode and a single polarization. The general rate equations are:

$$\frac{dE}{dt} = -\frac{1}{2}\left(\Gamma - g\right)E + \frac{1}{2}\xi(1 + i\alpha)\mathcal{N}E + E_{fb} + E_{inj} + \tilde{F}_k, \tag{1}$$

$$\frac{d\mathcal{N}}{dt} = J - \frac{\mathcal{N}}{T} - g\,|E|^2 - \xi\mathcal{N}\,|E|^2. \tag{2}$$

E is the slowly varying complex amplitude of the electric field. $\mathcal{N}$ is the carrier number. Γ, g, ξ and α are respectively the cavity loss, linear gain, differential gain and the linewidth enhancement factor. The last term in Equation (1) $\tilde{F}$ is a complex Gaussian white noise term with zero mean and $\langle \tilde{F}_{(t)}\tilde{F}_{(t')}\rangle = \beta\,T^{-1}(\mathcal{N} + \mathcal{N}_{thr})\delta_{(t-t')}$, with β being the spontaneous emission factor. J is the injection current with respect to the threshold current $J = (I - I_{th})/e$, with I and I_{th} being the pump and threshold pump current and e the elementary charge. T is the carrier lifetime. Two terms are added to the right hand side of Equation (1), namely:

$$E_{fb} = \eta\,E(t - \tau)e^{-i\Omega\tau} \tag{3}$$

$$E_{inj} = \mu\mathcal{E}\left(1 + e^{i[B(t)+\Phi_0]}\right). \tag{4}$$

The terms E_{fb} and E_{inj} represent the optical feedback and injection, with η and μ being the feedback and injection rates. The feedback has a delay τ and $\Omega\tau$ is the constant phase mismatch that arises from the roundtrip. Equation (4) actually models a Mach–Zehnder modulator (MZM), where $\mathcal{E}$ is the complex amplitude of the injected electrical field, $B(t)$ is the masked data and Φ_0 is the bias of the MZM. The frequency detuning between injected signal $\mathcal{E}$ and laser field E is assumed in this work to be zero. The effect of detuning on performance is discussed in Ref. [14]. A schematic illustration of these mechanisms can be seen in Figure 1. We have numerically integrated these equations using the Heun method with a time step of 0.5 ps.
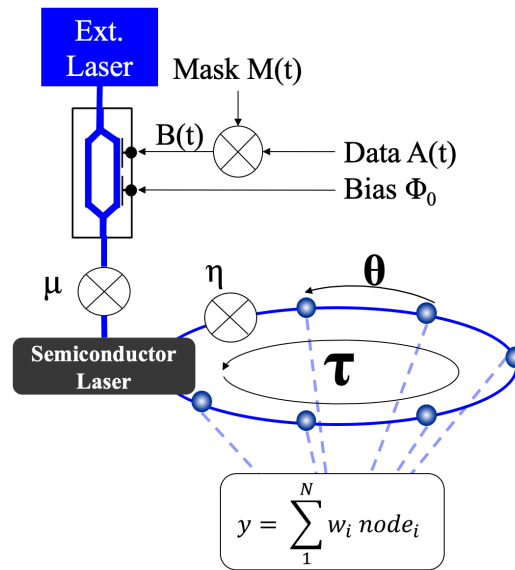
Figure 1. A schematic of the delay based RC with a semiconductor laser which is modelled by Equations (1)–(4). The data $A(t)$ is multiplied with mask $M(t)$, resulting in masked stream $B(t)$. The masked stream is modulated unto the output $\mathcal{E}$ of an external laser. The Mach–Zehnder modulator is biased by Φ_0, such that it remains in its quasi-linear regime. The modulated stream is multiplied by the injection rate μ and injected into the laser. The feedback strength in the delay line is controlled by the feedback rate η.

2.2. Pre- and Post-Processing

The input stage to the delay-based reservoir serves to preprocess the data, such that it corresponds to the timescales in the reservoir and the intended computation speed. The preprocessing of a normalised univariate dataset is done in two steps. First, each point of the dataset is sampled and held for a period τ_M, the mask length. In Ref. [7], τ_M was considered to be equal to the delay time τ such that upon injection the delay line is completely filled with responses to a single data sample. Nevertheless, in other works such as Refs. [9,17,18], a mismatch was introduced between τ and τ_M. We will consider both cases in this work. This datastream, denoted as $A(t)$ in Figure 1, is then multiplied with a mask $M(t)$, which is periodic over τ_M and has small temporal features of interval length θ, resulting in the masked datastream $B(t)$ in Figure 1. As the laser is fed with the masked stream $B(t)$, it will have different nonlinear responses to these temporal variations. These responses can be measured sequentially and then form the states of the virtual nodes. The nodes are now denoted as virtual as they now exist in a time-multiplexed way rather than corresponding to real physical spatially distributed and interconnected nodes. The time interval θ is also known as the virtual node separation or node distance. When delay and mask are synchronised, i.e., $\tau = \tau_M$, the node separation has to be chosen carefully, as it should be in the same range as a timescale or inertia of the nonlinear node [2]. If θ is much larger than the timescale of the node, the nonlinear node goes into a quasi-steady state regime for each mask feature, leading to a significant drop in state diversity. When the node separation is much shorter than the timescale of the nonlinear node, the mask features will be too fast for the node to follow leading to all virtual nodes having very similar state values and very low state diversity. In both cases, the time-multiplexing procedure will be unable to form a virtual interconnected network. In the case considered in Refs. [9,17,18], the virtual network is formed due to the mismatch between delay τ and mask τ_M, while the node distance θ is considered to be very large. The number of virtual nodes $N = \tau_M/\theta$ is equally important, as it determines the speed as well as the performance of the setup. If the number of nodes is small, the performance decreases but the system speeds up. Higher number of nodes often means a better state diversity, which improves performances, at the cost of a slower computation speed. In this work, we generate random, but fixed

masks with four discrete levels, $[0, 0.25, 0.75, 1]$. Once the mask $M(t)$ is fixed and multiplied with the datastream $A(t)$, the resulting datastream $B(t)$ is rescaled such that $0 \leq B(t) \leq \pi/2$ and the MZM bias Φ_0 is set to $\pi/4$. This is to ensure that the MZM is modulating in its quasi-linear regime. The node separation will be varied throughout the paper.

The output intensity of the semiconductor laser is sampled with a sample period corresponding to θ; the samples correspond to the end of each virtual node interval. The N samples within the τ_M interval correspond to the virtual node values that have responded to a single data sample and they define the reservoir state. These node values ($node_i$ in Figure 1) are linearly combined using weights w_i to form the output signal y. In training, the goal is to set w_i such that y approximates a desired target signal y_{exp} as well as possible in a least squares sense.

3. Timeseries Prediction

To illustrate the performance of a semiconductor laser with delayed feedback and how it scales with its system parameters, we have chosen to use a timeseries prediction as a benchmarking task. We have utilised a timeseries from the Santa Fe competition generated by a far-infrared laser operating in a chaotic regime [19]. The aim of this task is to predict the next sample in the chaotic time trace. This dataset has 9093 datapoints of which the first 3005 points are used for training and the subsequent 1005 points are used for testing the performance. The first 5 points are discarded from both stages, in order to filter out possible transients arising from turning on the injection of data. By comparing the reservoir's trained output y with the target y_{exp} for previously unseen input samples, we can quantify the performance using the normalised mean square error (NMSE), which is defined as:

$$NMSE(y, y_{exp}) = \frac{\langle ||y(n) - y_{exp}(n)||^2 \rangle}{\langle ||y_{exp}(n) - \langle y_{exp}(n) \rangle ||^2 \rangle}, \tag{5}$$

where y is the predicted value and y_{exp} is the expected value, n is the discrete time index of the input samples and the symbols $||...||$ and $\langle ... \rangle$ stand for the norm and the average respectively. The $NMSE$ is always a positive value, with lower $NMSE$ values corresponding to better performances.

The laser-based RC scheme has a number of parameters that can be tuned to obtain an optimised reservoir. We have employed a Bayesian optimisation technique [20] combined with the upper confidence bound acquisition function to scan the parameter space spanned by some of the parameters that can easily be manipulated in practice. These parameters are: the pump current of the reservoir laser, J; the feedback rate, η; the injection rate, μ. Table 1 presents the parameter values used during the simulations. In this section, we chose as node distance θ a value of 20 ps and mask and delay synchronised ($\tau = \tau_M$). This is following the work of Nguimdo et al. [16]. In this work, $N = 200$ is chosen sufficiently large such that state diversity is not compromised. In this section, we do not vary N. We will analyse the effect of the value of θ later on in Section 4. We have generated only one random mask and used this for all results presented in this section.

We have performed a Bayesian optimisation over a three dimensional parameter space (feedback rate, injection rate and pump current). Figure 2 shows two-dimensional projections of the parameter space. The performance indicator $NMSE$ is colour and size coded in the scatter plot. Better performances, in other words lower $NMSE$ values, are represented by larger circle markers and a reddish hue. Worse performances, or higher $NMSE$ values, are represented by smaller markers and are on the blue side of the colour scale. All plots have the pump current along the x-axis. Figure 2a illustrates how the NMSE relates to the feedback rate η on the y-axis. Similarly, Figure 2b has the (total) injection rate μ on the y-axis. The best performances are achieved when the current is around twice the threshold pump current (see Figure 2a,b). For pump currents above this range, we have observed that dynamical behavior of the semiconductor laser becomes chaotic and unable to produce consistent responses for similar inputs. This degrades performance. We find from Figure 2a that as the pump current is increased, the feedback rate has to be lowered to achieve better performances. This could be explained as follows. An increase in pump current will increase the overall power emitted

by the laser diode. As a result, the power of the feedback signal will also increase and will be able to destabilise the laser more easily. Lowering the feedback rate will therefore reduce the feedback power and favour consistent behaviour and good performance. As the pump current increases, we observe in Figure 2b that the injection rate needs to be increased as well to stay in a regime of low NMSE and good performance. We believe the higher injection power is required here to better injection lock. The optimal parameter values obtained from the Bayesian optimisation $\mu = 98.1$ ns^{-1} and $\eta = 7.8$ ns^{-1} at $I = 2.02 I_{th}$. In this case, the lowest NMSE equals 1%. Repeating the same analysis with other randomly generated masks delivers the same optimal parameter values and a variation of the performance smaller than 0.2%.

Table 1. Parameters used in the Bayesian optimisation for a timeseries prediction task.

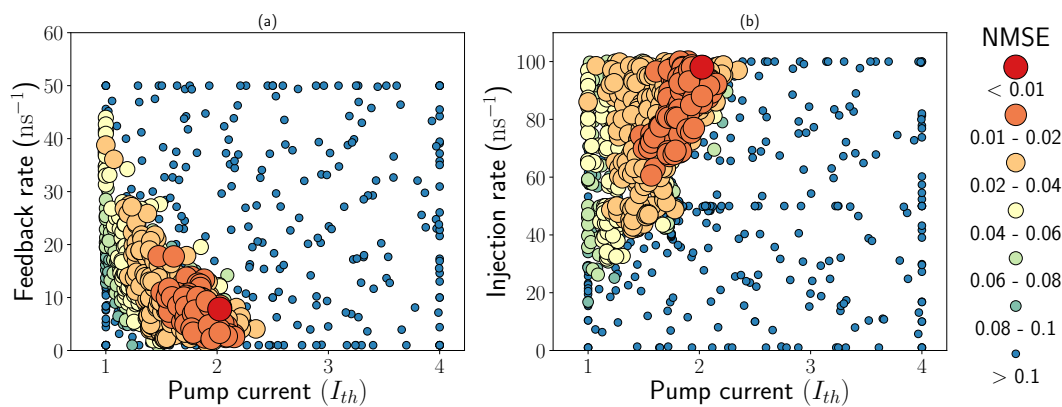| Parameters	Designation	Value Used in Bayesian Optimization
Linewidth enhancement factor	α	3.0
Loss	Γ	1 ps^{-1}
Threshold gain	g	1 ps^{-1}
Differential gain	ζ	5000 s^{-1}
Spontaneous emission factor	β	10^{-6}
Carrier-lifetime	T	1 ns
Threshold pump-current	I_{th}	16 mA
Pump-current	I	scanned over $[I_{th}; 3I_{th}]$
Feedback rate	η	scanned over $[0; 50$ ns$^{-1}]$
Injection rate	μ	scanned over $[0; 100$ ns$^{-1}]$
Amplitude of injected field	$\mathcal{E}$	200
Bias voltage of the MZM	Φ_0	$\pi/4$
Constant feedback phase	Ω	0
Node distance	θ	20 ps
Number of nodes	N	200
Delay time	τ	4 ns
Mask length	τ_M	4 ns

Figure 2. Test results obtained from the Bayesian optimisation for the semiconductor laser with delayed feedback trained on a time-series prediction task projected in the plane of (**a**) feedback rate and pump current or (**b**) injection rate and pump current. The performance indicator *NMSE* is coded into the colour and size of the markers in the scatter plot. A better performance corresponds to a bigger marker size and a redder colour. Parameters as in Table 1 and $\theta = 20$ ps, $N = 200$, $\tau = \tau_M = 4$ ns.

4. Task-Independent Computational Performance

Benchmarking the performance of a semiconductor laser with delayed feedback as an RC by training it to perform one or several benchmark tasks is useful. However, the system parameters for which optimal performance is obtained can vary from task to task. As an alternative, a framework has been introduced to quantify any system's total information processing capacity in a general and task-independent way. This computational capacity (*CC*) is typically split into two main parts:

the capacity of the system to retain past input samples is captured by the linear memory capacity [21] and the capacity of the system to perform nonlinear computation is captured by the nonlinear memory capacity [22]. It has been shown that the total memory capacity is limited by the number of read-out degrees. In our case, the upper limit corresponds to the total number of virtual nodes N. Due to this ideal limit, a trade-off between linear and nonlinear memory capacity exists.

To measure the linear and nonlinear capacities, a series of independent and identically distributed input samples $u(n)$ drawn uniformly from the interval $[-1, 1]$ are injected into the reservoir, with n a discrete time. Then we train the RC to reconstruct a set of linear and nonlinear polynomial functions depending on past inputs $u(n - i)$, looking back i steps in the past. In our calculations, the maximum value of i is 20. We follow the approach of Dambre et al. (Ref. [22]) and take for these functions Legendre polynomials $P_d(u)$ (of degree d), due to their orthogonality in the interval $[-1, 1]$. As an example, we can train the reservoir to reproduce the target signal $y_{exp}(n)$, given by

$$y_{exp}(n) = P_3(u(n - 2))P_1(u(n - 4)). \tag{6}$$

Instead of using the NMSE as we did for the timeseries prediction task, a memory capacity C is defined to quantify the ability of the RC to reconstruct each of these functions. The memory capacity C lies between 0 and 1 and is defined as [22]:

$$C = 1 - \frac{\langle (y_{exp} - y)^2 \rangle}{\langle y_{exp}^2 \rangle}, \tag{7}$$

where $\langle . \rangle$ denotes the average over all samples used for the evaluation of C. As the Legendre polynomials are orthogonal over the distribution of the input samples, the capacities C corresponding to different functions will yield independent information. Their sum will give the total computational capacity (CC), i.e., the total information processing capacity of the RC. We will group the memory functions by their total degree, which is the sum of degrees over all constituent polynomial functions, e.g., Equation (6) has total degree 4. Within each degree group, we can sum the memory capacities C yielding the total memory capacity per degree. We will use this to evaluate the contributions of individual degrees to the total computational capacity (CC) of the RC (the sum over all memory capacities per degrees). We have used 10,005 input samples for training. The first five states of the nodes are discarded to allow for the transient. The approach of Dambre in Ref. [22] does not use a testing session. Using this procedure, a risk exists of overestimating the memory capacities C due to the use of data sets with finite length. As explained in Ref. [22], Equation (7) is plagued by a positive bias. Following Ref. [22], a cutoff capacity C_{co} is used ($C_{co} \approx 0.003$ for 10,000 test samples) and capacities below this cutoff are neglected. Typically, for the linear capacities no non-zero capacities were obtained for $i > 13$, with i being the number of time steps in the past used in constructing the nonlinear functions.

Results

We will start by analysing the total CC of our system close to the optimal parameter set that was obtained in Section 3. We fix all system parameters at the optimum, but we will vary the feedback strength. The node distance θ is fixed to $\theta = 20$ ps. The delay and mask lengths match ($\tau = \tau_M$). As calculating a large amount of CC is numerically challenging, we have reduced the number of nodes to $N = 101$ to speed up the numerical analysis. This smaller value of N compared to the one used in Section 3 leads to similar performances on the Santa-Fe benchmark. In addition, the delay time will be smaller, but we do not observe a change in dynamical regime as compared to the situation in Section 3. We have generated only one random mask and used this for all results presented in this section. Repeating the same analysis with other randomly generated masks did not change the findings. The results are shown in Figure 3. The total memory capacity is maximum around $\eta \approx 10$ ns^{-1}. So the optimum that was reached in the benchmark task is also the point in parameter space where the total

CC reaches its peak value. Note that the total *CC* does not reach its ideal value of $N = 101$. This reduction of *CC* will be discussed later. At the optimal point all memory capacities of degrees higher than one, i.e., all nonlinear capacities, have increased. The linear memory capacity (degree 1) remains mainly unaffected by the feedback strength. It is only for higher feedback strengths that the linear memory capacity degrades. These results illustrate that the Santa Fe timeseries prediction task is a very diverse task requiring both nonlinear and linear capacities.

Figure 3. Total computational capacity showing the colour-coded contributions of different degrees of nonlinearity vs. the feedback strength η. Parameters as in Table 1 and $\mu = 100$ ns^{-1}, $I = 2I_{th}$, $N = 101$, $\theta = 20$ ps and $\tau = \tau_M = N\theta$.

Now, we will investigate the effects of the masking procedure on *CC*. This time-multiplexing scheme defines the number of nodes N and their separation in time θ. Together θ and N define the mask length τ_M. In Ref. [7], a rule of thumb was suggested implying that an optimal choice of θ was around 20% of the internal timescale of the system. In the case of a semiconductor laser with delayed optical feedback and optical data injection, it is not readily clear which timescale should be taken. In Ref. [10], the timescale considered was related to relaxation oscillations, and the choice for θ was 200 ps. Nguimdo et al. stated that the node distance could easily be reduced to 20 ps, claiming injection locking dynamics as the underlying reason [16]. We have opted in Section 3 and above to use that specific value. We have calculated the memory capacities per degree, while varying the node distance θ from 5 ps to 50 ps as shown in Figure 4. We observe a clear trend with the highest total *CC* at $\theta = 30$ ps. If the node distance is too short the capacity is considerably lowered. For small node distances, the node states will become highly correlated. This results in effectively having less nodes available for computation and hence a lower cap on the total computational capacity. A node distance can also be too long. In that case, the coupling between virtual nodes reduces as transients have died out. At the optimal value that we have obtained, we would like to point out that the total *CC* is also reduced and reaches only about 70% of its maximum value. This can be attributed to the previously mentioned correlation that is induced between virtual node states through the transient dynamics which reduces the node state diversity. As a side remark, we want to highlight that changing θ will change the delay time τ. The range in τ that is covered in Figure 4 goes from from 0.5 ns to 5 ns. In this entire range, we did not observe a change in dynamical behaviour of the semiconductor laser. This indicates that the change in *CC* is only due to the masking procedure and not due to a change in delay line length.
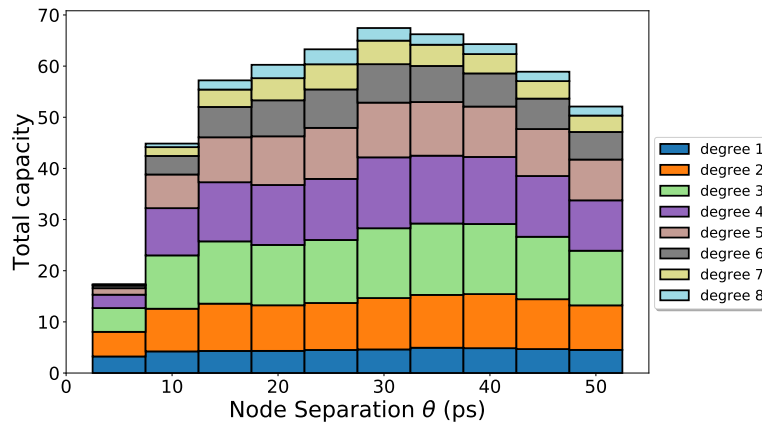
Figure 4. Total computational capacity (*CC*) showing the colour-coded contributions of different degrees of nonlinearity vs. the node distance θ. Parameters as in Table 1 and $\mu = 100$ ns^{-1}, $\eta = 10$ ns^{-1}, $I = 2I_{th}$, $N = 101$ and $\tau = \tau_M = N\theta$.

Finally, we want to investigate if it is possible to remedy the reduction of the *CC* due to a reduced state diversity due to inertia of the semiconductor laser. Instead of relying on the dynamics to connect virtual nodes to each other, an overlap between delay and mask length ($\tau > \tau_M$) can be used as in Refs. [9,17,18]. However, in those cases the virtual node distance θ was chosen much longer than any timescale related to internal dynamics of the nonlinear system. In addition, contrary to those works, we decide to still use a short node distance $\theta = 20$ ps, close to the optimum obtained in Figure 4. According to Refs. [9,17,18], the mismatch and the node number should be co-primes. This was ensured by taking $N = 101$. In Figure 5, we analyse the effect of the delay τ on the *CC*. The mask length is kept constant $\tau_M = N\theta$. We observe clearly that when $\tau < \tau_M$, the *CC* is reduced further. However, for $\tau > \tau_M$, we see a slight increase in *CC*, when the delay and the mask length have a mismatch of about half a node distance. When $\tau = (N+1)\theta$, an overlap of one virtual node, the *CC* has dropped again slightly. For even longer τ, in Figure 6, we can conclude that a general trend exists to a slightly higher *CC*, but the ideal value of *CC* = *N* is never reached. From an experimental point of view, this is a very interesting result. Not much care needs to be taken in matching delay and mask lengths. In fact, a small mismatch even of several virtual nodes can increase the *CC* and the computing performance.
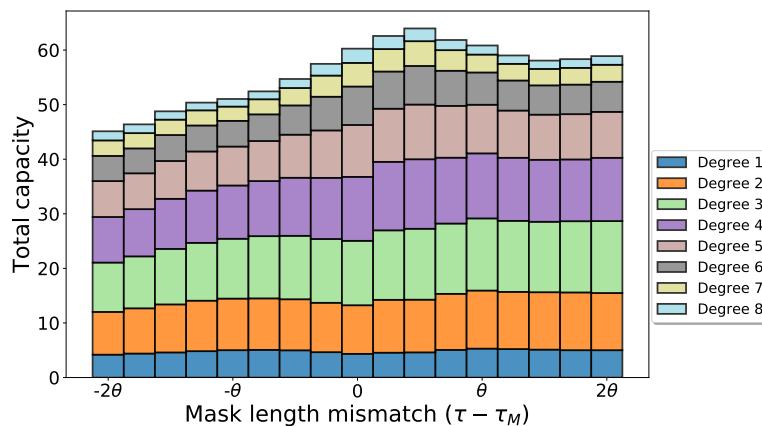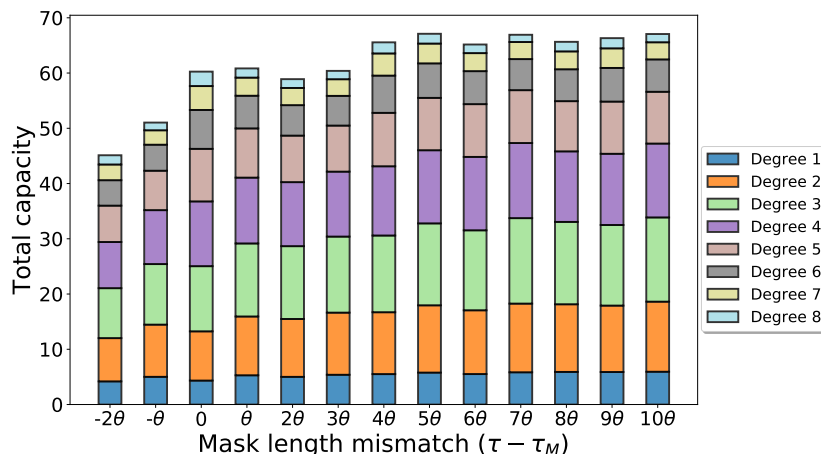

Figure 5. Total computational capacity (*CC*) showing the colour-coded contributions of different degrees of nonlinearity vs. the delay time τ. Parameters as in Table 1 and $\mu = 100$ ns^{-1}, $\eta = 10$ ns^{-1}, $I = 2I_{th}$, $N = 101$ and $\tau_M = N\theta$.

Figure 6. Total computational capacity (*CC*) showing the colour-coded contributions of different degrees of nonlinearity vs. the delay time τ. Parameters as in Table 1 and $\mu = 100 \ \text{ns}^{-1}$, $\eta = 10 \ \text{ns}^{-1}$, $I = 2I_{th}$, $N = 101$ and $\tau_M = N\theta$.

5. Conclusions

We have analysed the computational capacity of semiconductor lasers with delayed feedback used as substrates for reservoir computing. Our main focus was put on analysing the effect of the specifics of the masking procedure. We have found that an optimal node distance is found around 30 ps which maximises computational capacity in the case of a perfect match between delay length and mask length. Nevertheless, due to the fact that a virtual network is created through the dynamics of the semiconductor laser, the maximum computational capacity is never reached. A mismatch between the mask length and the delay length can be beneficial for the computational capacity. However, the effect of the mismatch is rather limited for the values of the node distance analysed. From an experimental viewpoint, this is an advantage as little care should be placed in designing an exact delay line length. The computational capacity can be further increased to its ideal theoretical maximum value, if the mismatch is combined with longer node distances as in Refs. [9,17,18]. However, this would lower the computation speed of the system significantly as it is inversely proportional to the node distance.

Author Contributions: The work described in this article is the collaborative development of both authors. Conceptualization, G.V.d.S.; Formal analysis, K.H.; Investigation, K.H. and G.V.d.S.; Methodology, G.V.d.S.; Software, K.H.; Supervision, G.V.d.S.; Visualization, K.H.; Writing—original draft, K.H. and G.V.d.S.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1097–1105. [CrossRef]
2. Van der Sande, G.; Brunner, D.; Soriano, M.C. Advances in photonic reservoir computing. *Nanophotonics* **2017**, *6*, 561–576. [CrossRef]
3. Vandoorne, K.; Dierckx, W.; Schrauwen, B.; Verstraeten, D.; Baets, R.; Bienstman, P.; Van Campenhout, J. Toward optical signal processing using photonic reservoir computing. *Opt. Express* **2008**, *16*, 11182–11192. [CrossRef] [PubMed]
4. Vandoorne, K.; Dambre, J.; Verstraeten, D.; Schrauwen, B.; Bienstman, P. Parallel reservoir computing using optical amplifiers. *IEEE Trans. Neural Netw.* **2011**, *22*, 1469–1481. [CrossRef] [PubMed]
5. Vandoorne, K.; Mechet, P.; Van Vaerenbergh, T.; Fiers, M.; Morthier, G.; Verstraeten, D.; Schrauwen, B.; Dambre, J.; Bienstman, P. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **2014**, *5*, 2541. [CrossRef] [PubMed]

6. Brunner, D.; Fischer, I. Reconfigurable semiconductor laser networks based on diffractive coupling. *Opt. Lett.* **2015**, *40*, 3854–3857. [CrossRef] [PubMed]

7. Appeltant, L.; Soriano, M.C.; Van der Sande, G.; Danckaert, J.; Massar, S.; Dambre, J.; Schrauwen, B.; Mirasso, C.R.; Fischer, I. Information processing using a single dynamical node as complex system. *Nat. Commun.* **2011**, *2*, 468. [CrossRef] [PubMed]

8. Paquot, Y.; Dambre, J.; Schrauwen, B.; Haelterman, M.; Massar, S. Reservoir computing: A photonic neural network for information processing. In Proceedings of the SPIE Photonics Europe on International Society for Optics and Photonics, Brussels, Belgium, 12–16 April 2010; p. 77280B.

9. Paquot, Y.; Duport, F.; Smerieri, A.; Dambre, J.; Schrauwen, B.; Haelterman, M.; Massar, S. Optoelectronic reservoir computing. *Sci. Rep.* **2012**, *2*, 287. [CrossRef] [PubMed]

10. Brunner, D.; Soriano, M.C.; Mirasso, C.R.; Fischer, I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **2013**, *4*, 1364. [CrossRef] [PubMed]

11. Takano, K.; Sugano, C.; Inubushi, M.; Yoshimura, K.; Sunada, S.; Kanno, K.; Uchida, A. Compact reservoir computing with a photonic integrated circuit. *Opt. Express* **2018**, *26*, 29424–29439. [CrossRef] [PubMed]

12. Nguimdo, R.M.; Verschaffelt, G.; Danckaert, J.; Van der Sande, G. Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 3301–3307. [CrossRef] [PubMed]

13. Vatin, J.; Rontani, D.; Sciamanna, M. Enhanced performance of a reservoir computer using polarization dynamics in VCSELs. *Opt. Lett.* **2018**, *43*, 4497–4500. [CrossRef] [PubMed]

14. Bueno, J.; Brunner, D.; Soriano, M.C.; Fischer, I. Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback. *Opt. Express* **2017**, *25*, 2401–2412. [CrossRef] [PubMed]

15. Kuriki, Y.; Nakayama, J.; Takano, K.; Uchida, A. Impact of input mask signals on delay-based photonic reservoir computing with semiconductor lasers. *Opt. Express* **2018**, *26*, 5777–5788. [CrossRef] [PubMed]

16. Nguimdo, R.M.; Verschaffelt, G.; Danckaert, J.; Van der Sande, G. Fast photonic information processing using semiconductor lasers with delayed optical feedback: Role of phase dynamics. *Opt. Express* **2014**, *22*, 8672–8686. [CrossRef] [PubMed]

17. Duport, F.; Schneider, B.; Smerieri, A.; Haelterman, M.; Massar, S. All-optical reservoir computing. *Opt. Express* **2012**, *20*, 22783–22795. [CrossRef] [PubMed]

18. Duport, F.; Smerieri, A.; Akrout, A.; Haelterman, M.; Massar, S. Fully analogue photonic reservoir computer. *Sci. Rep.* **2016**, *6*, 22381. [CrossRef] [PubMed]

19. Weigend, A.S.; Gershenfeld, N.A. The Future of Time Series. In *Time Series Prediction: Forecasting the Future and Understanding the Past*; Addison-Wesley: Boston, MA, USA, 1993.

20. Snoek, J.; Larochelle, H.; Adams, R.P. Practical bayesian optimization of machine learning algorithms. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 2951–2959.

21. Jaeger, H. *Short Term Memory in Echo State Networks*; GMD-Report 152; GMD–German National Research Institute for Computer Science: Sankt Augustin, Germany, 2002. Available online: http://www.faculty. jacobs-university.de/hjaeger/pubs/STMEchoStatesTechRep.pdf (accessed on 8 September 2019).

22. Dambre, J.; Verstraeten, D.; Schrauwen, B.; Massar, S. Information Processing Capacity of Dynamical Systems. *Sci. Rep.* **2012**, *2*, 514. [CrossRef] [PubMed]