Integrated photonic reservoir computing based on hierarchical time-multiplexing structure

Hong Zhang, Xue Feng,[*] Boxun Li, Yu Wang, Kaiyu Cui, Fang Liu, Weibei Dou, and Yidong Huang

Department of Electronic Engineering, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing100084,China

*x-feng@tsinghua.edu.cn

Abstract: An integrated photonic reservoir computing (RC) based on hierarchical time-multiplexing structure is proposed by numerical simulations. A micro-ring array (MRA) is employed as a typical time delay implementation of RC. At the output port of the MRA, a secondary time-multiplexing is achieved by multi-mode interference (MMI) splitter and delay line array. This hierarchical time-multiplexing structure can ensure a large reservoir size with fast processing speed. Simulation results indicate that the proposed RC system yields better performance than previously reported ones. The achieved normalized mean square error between the system output and target sequence are 0.5% and 2.7% for signal classification and chaotic time series prediction, respectively, while the sample rate is as high as 1.3Gbps.

OCIS codes: (200.3050) Information processing; (250.4745) Optical processing devices; (250.4745) Optical processing devices; (140.3945) Microcavities.

References and links

1. M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**(3), 127–149 (2009).
2. H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication," Science **304**(5667), 78–80 (2004).
3. M. D. Skowronski and J. G. Harris, "Automatic speech recognition using a predictive echo state network classifier," Neural Netw. **20**(3), 414–423 (2007).
4. E. A. Antonelo, B. Schrauwen, and D. Stroobandt, "Event detection and localization for small mobile robots using reservoir computing," Neural Netw. **21**(6), 862–871 (2008).
5. F. Wyffels and B. Schrauwen, "A comparative study of reservoir computing strategies for monthly time series prediction," Neurocomputing **73**(10-12), 1958–1964 (2010).
6. H. Jaeger, "Short term memory in echo state networks" Technical Report 152, (2001).
7. K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. Van Campenhout, "Toward optical signal processing using photonic reservoir computing," Opt. Express **16**(15), 11182–11192 (2008).
8. 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," Nat Commun **2**, 468 (2011).
9. D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," Nat Commun **4**, 1364 (2013).
10. Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," Sci Rep **2**, 287 (2012).
11. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing," Opt. Express **20**(3), 3241–3249 (2012).
12. F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," Opt. Express **20**(20), 22783–22795 (2012).
13. R. Martinenghi, S. Rybalko, M. Jacquot, Y. K. Chembo, and L. Larger, "Photonic nonlinear transient computing with multiple-delay wavelength dynamics," Phys. Rev. Lett. **108**(24), 244101 (2012).
14. C. Mesaritakis, V. Papataxiarhis, and D. Syvridis, "Micro ring resonators as building blocks for an all-optical high-speed reservoir-computing bit-patternrecognition system," J. Opt. Soc. Am. B **30**(11), 3048–3055 (2013).

15. K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," IEEE Trans. Neural Netw. **22**(9), 1469–1481 (2011).
16. K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," Nat Commun **5**, 3541 (2014).
17. M. C. Soriano, S. Ortín, D. Brunner, L. Larger, C. R. Mirasso, I. Fischer, and L. Pesquera, "Optoelectronic reservoir computing: tackling noise-induced performance degradation," Opt. Express **21**(1), 12–20 (2013).
18. T. Baba, H. C. Nguiyen, N. Yazawa, Y. Terada, S. Hashimoto, and T. Watanabe, "Slow-light Mach-Zehnder modulators based on Si photonic crystals," Sci. Technol. Adv. Mater. **15**(2), 024602 (2014).
19. L. Alloatti, D. Korn, R. Palmer, D. Hillerkuss, J. Li, A. Barklund, R. Dinu, J. Wieland, M. Fournier, J. Fedeli, H. Yu, W. Bogaerts, P. Dumon, R. Baets, C. Koos, W. Freude, and J. Leuthold, "42.7 Gbit/s electro-optic modulator in silicon technology," Opt. Express **19**(12), 11841–11851 (2011).
20. J. Yu, X. Wang, J. Liu, Q. Yan, J. Xia, Z. Fan, Z. Wang, and S. Chen, "MMI optical coupler and switches with SOI technology" Optoelectronics, Proceedings of the sixth Chinese symposium **10**, 1109 (2003)
21. M. P. Nezhad, O. Bondarenko, M. Khajavikhan, A. Simic, and Y. Fainman, "Etch-free low loss silicon waveguides using hydrogen silsesquioxane oxidation masks," Opt. Express **19**(20), 18827–18832 (2011).
22. F. Shinobu, N. Ishikura, Y. Arita, T. Tamanuki, and T. Baba, "Continuously tunable slow-light device consisting of heater-controlled silicon microring array," Opt. Express **19**(14), 13557–13564 (2011).
23. S. J. Koester, J. D. Schaob, G. Dehlinger, and J. O. Chu, "Germanium-on-SOI infrared detectors for integrated photonic applications," IEEE J. Sel. Top. Quantum Electron. **10**, 1109 (2006).
24. D. Liang, M. Fiorentino, S. Srinivasan, J. E. Bowers, and R. G. Beausoleil, "Low threshold electrically-pumped hybrid silicon microring lasers," IEEE J. Sel. Top. Quantum Electron. **17**(6), 1528–1533 (2011).
25. S. Akiyama, T. Kurahashi, T. Baba, N. Hatori, T. Usuki, and T. Yamamoto, "A 1 V peak-to-peak driven 10-Gbps slow-light silicon Mach–Zehnder modulator using cascaded ring resonators," Appl. Phys. Express **3**(7), 072202 (2010).
26. F. Shinobu, N. Ishikura, Y. Arita, T. Tamanuki, and T. Baba, "Continuously tunable slow-light device consisting of heater-controlled silicon microring array," Opt. Express **19**(14), 13557–13564 (2011).
27. H. P. Uranus and H. J. W. M. Hoekstra, "Modeling of loss-induced superluminal an negative group velocity in two-port ring-resonator circuits," J. Lightwave Technol. **25**(9), 2376–2384 (2007).
28. J. K. S. Poon, J. Scheuer, Y. Xu, and A. Yariv, "Designing coulped-resonator optical waveguide delay lines," J. Opt. Soc. Am. B **21**(9), 1665–1673 (2004).
29. O. Boyraz, T. Indukuri, and B. Jalali, "Self-phase-modulation induced spectral broadening in silicon waveguides," Opt. Express **12**(5), 829–834 (2004).
30. http://www.coretk.com/CataLog/cata_img/FILE/1718755439/u2t/157/157_158_1068698470.pdf
31. http://literature.cdn.keysight.com/litweb/pdf/5990-9712EN.pdf
32. http://www.ti.com/lsds/ti/data-converters/high-speed-adc-greater-than-1gsps-products.page
33. http://www.libertytest.com/assetmanager/uploaded/pdf-2011928-105832-gt-1040a.pdf
34. H. Friis, "Noise figure of radio receivers," Proc. IRE **32**, 419–422, (1944).
35. A. S. Weigend and N. A. Gershenfeld, http://www-psych.stanford.edu/andreas/Time-Series/SantaFe.html#setA (1991)
36. U. Hübner, N. B. Abraham, and C. O. Weiss, "Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH3 laser," Phys. Rev. A **40**(11), 6354–6365 (1989).
37. A. Rodan and P. Tino, "Minimum complexity echo state network," IEEE Trans. Neural Netw. **22**(1), 131–144 (2011).
38. L. J. Cao, "Support vector machines experts for time series forecasting," Neurocomputing **51**, 321–339 (2003).

1. Introduction

Reservoir computing (RC) is a machine learning paradigm partially inspired by the brain and has been considered as a powerful tool for sequential data processing [1, 2]. In RC, the nonlinear network and the linear output layer are separated while only the output layer would be adopted for training. Thus, the tough training process in traditional recurrent neural network could be simplified. Both software and hardware implementations of RC have already been applied in many areas ranging from time series prediction, speech recognition to robotics [2–6]. Among the various implementations, photonic approaches attract more and more interests due to the advantage of high bandwidth, high power efficiency and intrinsic parallelism. In 2008, the first photonic RC is proposed based on semiconductor optical amplifier network [7]. Since then, optical or optoelectronic RC have been proposed and demonstrated with fiber optics and on-chip photonic systems [8–16]. The reported architectures of photonic RC could be sorted as network based RC and delay feedback RC.

In the network based RC, both active and passive optical networks with multi-loop circuits are utilized to form the reservoir. In [14, 15], photonic network with active nonlinear

components is proposed on semiconductor optical amplifiers (SOA) and *InGaAsP/InP* micro ring resonators, in which the nonlinearities are due to the power saturation of the SOA and the two-photon absorption (TPA) within the micro-rings, respectively. More recently, a passive network approach has been demonstrated on silicon photonic chips [16]. The passive RC network is composed with long silicon waveguide delay lines and the nonlinear nature stems from the square law detection of the electric field in various positions of the network. This implementation presents capabilities of Boolean logic operations as well as isolated spoken digit classification. Compared with active approaches, the passive network could dramatically reduce the energy consumption and can be fabricated with the CMOS compatible process, which is very significant for large scale implementation. However, only limited node numbers (9~25) have been achieved on the photonic network based RC, which is due to the highly demanding complexity and uniformity of fabrication process.

The delay-feedback RC is an effective approach to simplify the physical implementation since there is only one nonlinear element required while a large node number (50~2000) can be readily achieved by time-multiplexing [8–13]. Since the first time-delay structure introduced by Appeltant et.al [8], both optical and optoelectronic delay-feedback RCs have been demonstrated [9–13]. In these works, an input mask consisting of a series of random numbers is utilized by time-multiplexing to break the symmetry of the system so that a large virtual network could be set up while a feedback loop is adopted to achieve the recurrent nature. In [9], a continuous wave (CW) laser with output feedback is utilized to form an RC system. The convincing performance of delay-feedback RC showing in isolated spoken digit classification and chaotic series prediction indicates that large node number by the virtual network could be achieved with time-multiplexing. However, in delay-feedback RC, the contradiction between processing speed and reservoir scale is inevitable. The reservoir is introduced by simple time-multiplexing, which is equivalent to converting each input digit into a long series of masked digits for processing. As a result, the processing speed would degrade as the reservoir size increases. Therefore, the sample rate of the previously reported delay-feedback optical RCs is only 200k~13Mbps [9–13]. For a real application, such contradiction between processing speed and reservoir size has to be addressed.

In this work, an RC implementation based on a hierarchical time-multiplexing structure is proposed, which alleviates the contradiction of large reservoir size and fast processing speed while ensuring a relatively low requirement of fabrication at the same time. As we proposed, a silicon micro-ring array (MRA) is employed as the reservoir and a secondary time-multiplexing within each optical pulse interval is introduced at the output port of the MRA to multiply the node number of the reservoir. Such secondary time-multiplexing is achieved by multi-mode interference (MMI) splitter, an array of delay lines and detectors. With the help of the inserted MMI, different time delay is achieved before each detector within a single pulse interval so that the signal can be effectively detected multiple times within the time window of each optical pulse. Different to other time-delay RC systems, the delay length in this work varies within the interval between two adjacent pulses so that both fast processing speed and a large reservoir size can be achieved simultaneously. Specifically, sample rate of 1.3Gbps and node number of 480 are achieved in our simulations, while there are only an MRA composed with 5 rings and a 1×16 MMI coupler with its outputs delayed from 6 to 21ps and connected to 40GHz photodiodes. In our proposed RC, the passive structure is also helpful to avoid the additional noise from active components and therefore to achieve higher signal to noise ratio [17]. According to simulation results, error rate (the normalized mean square error between the system output and target sequence) of 0.5% for signal classification and 2.7% for chaotic time series prediction can be achieved with our proposed RC while the sample rate is as fast as 1.3Gbps. Comparison between our simulation and the reported results shows that our proposed RC system has advantages in terms of both speed and low error rate. Furthermore, our RC system could be integrated on a single silicon chip, which is very proposing for future large scale implementation.

2. Arrangement

The arrangement of our proposed RC is shown in Fig. 1(a). A CW laser followed by an intensity modulator is used to produce the incident laser pulse train. A phase modulator (PM) is used to load the input electrical signal on the incident laser pulse with a phase shift from 0 to 2π according to the input voltage level. The encoded pulses are then injected into the MRA, which consists of five serially cascaded micro-ring cavities. After the MRA, the signal is sent into an array of delay lines via an MMI and afterwards detected by photo-diode (PD) arrays. At each port of MMI, the length of the delay lines is varied to achieve different time delay.

Considering state of the art photonic devices, all the components in the proposed implementation could be integrated on a monolithic silicon chip. The operating wavelength is considered as 1.55 micron so that the transmission line, modulator, micro-ring array, MMI and delay lines can be realized by silicon wire waveguide and directly fabricated on SOI wafer [18–22]. The PD array is considered as Ge detectors that could be fabricated with CMOS compatible process while the III-V laser could be bonded on the silicon chip [23, 24].
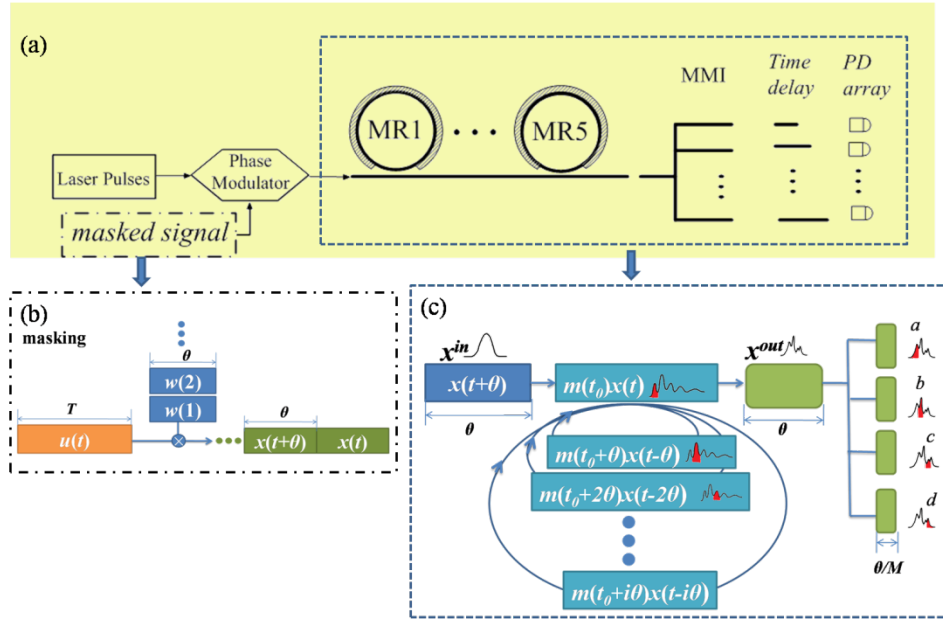

Fig. 1. (a) Schematic of the proposed optical reservoir computing system. PM is a phase modulator to load the signal on optical pulse. MR1~5 is the cascaded micro-ring cavities. (b) Schematic of the masking process. $u(t)$ is the input time dependent signal and $w(i)$ is the random input mask. (c) Schematic of the processing of the MRA (dark green) and the secondary time-multiplexing (light green). $m(t)$ is the time-domain transfer coefficient defined in Fig. 2. Superposition between the adjacent pulses is done within the MRA. Secondary time-multiplexing is made within each pulse interval at different output ports.

3. Operating principle

The operating principle of our proposed RC could be considered as a hierarchical time-multiplexing structure. In this section, the operating principle would be described in detail.

3.1 Pre-processing: the random mask

The original signal is a discrete time series denoted as $u(t)$, where $t = nT$ ($n = 1,2,3…$) and T is the sampling time interval. A digit is defined as the signal $u(nT)$ at a particular n value. The original signal is firstly multiplied with an input mask on electrical domain by time-multiplexing as shown in Fig. 1(b). The input mask $w(i)$ ($i = 1,2,…b$) is a vector with a dimension of b. Each element is randomly chosen in order to break the symmetry of the

system. Thus, each digit $u(t)$ would be stretched into a b-dimensional vector via time-multiplexing, which is similar to the preprocessing in other reported RCs [9–12]. After that, the product of the original signal $u(t)$ and the mask $w(i)$, i.e. $x_{in}(t) = x_{in}(t,i) = w(i)u(t)$, is encoded on the optical carrier by the phase modulator. The optical carrier is considered as a pulse train with duration time of θ. The masked signal would be sent into the MRA.

3.2 Micro ring array (MRA)

The central part of a reservoir computing system is the large dynamic network referred as the reservoir. In this work, an MRA consisting of five serially cascaded micro-ring cavities is utilized to provide the dynamics. As shown in previous reports [25, 26], micro-ring cavity could achieve high quality factor and real time delay. That is to say, a period of memory for RC could be obtained by the MRA. Here, we choose five micro-rings as a trade-off between memory length and system complexity. It should be mentioned that the resonant peak of each micro-ring cavity would be carefully tuned according to the operating wavelength of the CW laser. Thus, such memory could be obtained due to the strong dispersion at resonant wavelength of micro-ring cavities while the encoded optical pulse train propagates within the ring cavity [27, 28]. Obviously, such memory would result in superposition of adjacent pulses, as illustrated in Fig. 1(c). Small time width and therefore wide spectrum of laser pulses are chosen to achieve large dispersion and fast speed.

Figure 2 shows the calculated output wave shape on time domain of a single pulse propagating after five cascaded ring cavities. The adopted parameters in our calculation are shown in Table 1. It is clearly shown that the output signal has a decayed amplitude that would last for several pulse duration time. Equivalently, one input pulse is converted into several pulses after the MRA. As shown in Fig. 2, we define the transfer coefficient $m(t)$ as the electric field vector at the output port of the MRA at corresponding time spot t, which is normalized to the peak amplitude of the input pulse. As both the input pulse train interval and the detection interval are firstly considered as θ, we choose $m(t_0 + i\theta)$ ($i = 0,1,2...$) for the convenience of following elaboration. t_0 is the start-point for detection and considered as some point near the peak of the first significant pulse of the output. The sketch of a single micro-ring cavity is shown in Fig. 3 and the detailed model in calculation is shown in *Appendix. 1*.
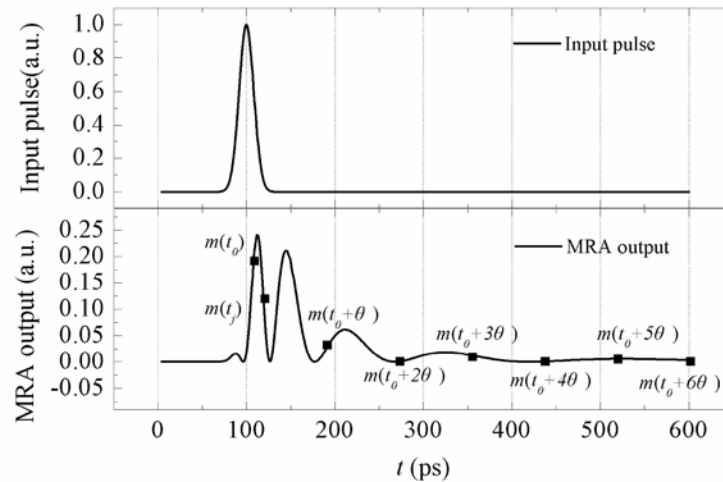

Fig. 2. Wave shape on time-domain of the input pulse and the output after MRA. $m(t)$ is the defined transfer coefficient.

Table 1. Simulation parameters

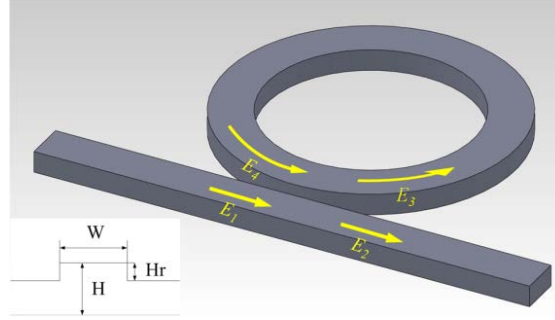| Parameters	Symbol	Value
Transmission loss	γ (dB/cm)	1.3
Effective refraction index	n_{eff}	2.82
Central wavelength	λ (nm)	1549.45
Reflection index at coupling spot	R	0.9
Micro-ring perimeter	d (μm)	650

Fig. 3. Sketch of micro-ring cavity model; subset: sketch of the interface of the shallow-ridge waveguide.

3.3 The nonlinear dynamics of the RC system

As illustrated in Fig. 1(c), after a pulse train carrying signal $x_{in}(t)$ propagates through the ring array, the output electric field is therefore $x_{out}^{RA}(t) = \sum_{i=0}^{q} m(t_0 + i\theta) \cdot x_{in}(t - i\theta)$, where q represents the length of the memory. In our calculation, q is set as 10. As can be seen in Fig. 2, most contribution comes from the first 6 intervals, i.e. $i = 0\sim5$. For a given pulse train, the length of the memory would be determined by the radius, transmission loss, the effective refraction index as well as the coupling condition of the micro-ring cavities. Here, the transfer coefficient $m(t_0 + i\theta)$ could be expressed as the complex form as $m_{t_0+i\theta} \exp(j \cdot \phi_i) = m(t_0 + i\theta), (i = 1, 2, ..., q)$, where $m_{t_0+i\theta}$ and ϕ_i are amplitude and phase, respectively. Since phase modulation is employed, the input signal on optical domain could be expressed as $x_{in}(t - i\theta) = exp(j \cdot x_{t-i\theta})$ $(i = 1,2,3...)$, where $x_{t-i\theta}$ is real. Thus, the output electric field after ring array $x^{RA}_{out}(t)$ could be expressed as:

$$x_{out}^{RA}(t) = \sum_{i=0}^{q} m_{t_0+i\theta} \cdot \exp[j \cdot (\phi_i + x_{t-i\theta})] \qquad (1)$$

After the superposition of different pulses (considering both amplitude and phase), the final output comes from the square-law detection with PD. For more clarity, we firstly consider the case that the MRA is immediately followed by a single detector (without the MMI and the delay lines). The detected output intensity could be expressed as:

$$
\begin{aligned}
X^{RA}_{out}(t) &= \left| x_{out}^{RA}(t) \right|^2 \\
&= \left[\sum_{i=0}^{q} m_{t_0+i\theta} \cdot \cos\left(x_{t-i\theta} - x_t + \Delta\phi_i \right) \right]^2 + \left[\sum_{i=0}^{q} m_{t_0+i\theta} \cdot \sin\left(x_{t-i\theta} - x_t + \Delta\phi_i \right) \right]^2
\end{aligned} \qquad (2)
$$

where $\Delta\phi_i = \phi_i - \phi_0$, which represents the phase variation between $m(t_0 + i\theta)$ and $m(t_0)$.

The above mentioned dynamics would lead to nonlinearity with the random mask. The superposition of the field would include a digit masked with different coefficient w_i. For example, $m(t_0)w(2)u(t)$ and $m(t_0 + \theta)w(1)u(t)$ would be mixed in the MRA and then detected

after. Therefore, Eq. (2) would result in nonlinearity with respect to the origin signal $u(t)$, although there is no optical nonlinear effect involved.

According to the RC principle, the system is required to have a "fading memory". The interaction among the present digit and previous digits should be adjusted so that the system could reach the edge of the stability to obtain better performance [1, 2]. Here, since the memory provided by the MRA would be attenuated in amplitude, the strength of interaction among the adjacent digits could be adjusted by the mask length b. Therefore b should be meticulously settled according to the edge of the task-dependant stability. This condition could be mathematically described as $b \times \theta \times \varepsilon \approx \tau$, where τ is the memory length of the MRA ($\tau = q \cdot \theta$) and ε (in bits) is the dynamic memory dependent on the task. In practical tasks, the dynamic memory that ensures the system with stability lasts for less than 10 bits, which could be satisfied by our system.

3.4 Secondary time-multiplexing

As shown in Fig. 2, the memory length of MRA τ lasts for about 6 bits. However, the RC principle requires that the projecting dimension, which is usually referred as the node number, be much larger than the input dimension [1, 2]. In order to extend the projecting dimension, a secondary time-multiplexing within each time step θ is adopted. As shown in Fig. 1(c), the output is detected M times instead of only once within each time step θ. After MRA, the time-domain shape of the output pulse is of high complexity, as shown in Fig. 2. Therefore, time-multiplexing within one pulse could equivalently project each output into a higher dimension and the criteria of the RC nodes could be met.

Considering such secondary time-multiplexing, the dynamical function of Eq. (2) would now be expanded to a series of equations:

$$
\begin{aligned}
X^{j}_{out}(t) &= \left| x^{j}_{out}(t) \right|^{2} \\
&= \left[\sum_{i=0}^{q} m_{t_j + i\theta} \cdot \cos\left(x_{t-i\theta} - x_t + \Delta\phi_{i,j} \right) \right]^{2} \\
&+ \left[\sum_{i=0}^{q} m_{t_j + i\theta} \cdot \sin\left(x_{t-i\theta} - x_t + \Delta\phi_{i,j} \right) \right]^{2} \quad (j=1,2,3...)
\end{aligned}
\tag{3}
$$

where j is the jth time-multiplexing channel. $t_j = t_0 + \delta_j$, where δ_j represents the time delay of the corresponding channel after the MRA. $\Delta\phi_{i,j} = \phi_{i,j} - \phi_{0,j}$, which represents the phase variation between $m(t_j + i\theta)$ and $m(t_j)$.

To this end, the original signal $u(t)$ is now projected into a large matrix $X_{out}^{i,j}(t)$ ($i = 1,2...b$, $j = 1,2...M$), where M is the total secondary multiplexing channel number. Thus, the total node number is the product of the first time-multiplexing number b and the secondary time-multiplexing number M, i.e. $N = b \cdot M$.

It is worth mentioning that unlike other time-delay RC systems, the delay length in this work varies within the interval between two adjacent pulses so that the speed of the system would not be deteriorated by the time delay. In principle, such secondary time-multiplexing could be done by an ultra-high bandwidth detector. However, detectors with such short response time are yet beyond state of the art technology. Here we provide an alternative for practical implementation. As shown in Fig. 1(a), an array of delay lines is applied to each output interval connected with an MMI. The delay lines are considered as silicon waveguides with different lengths so that their time delays vary from each other. Each waveguide is then connected with a PD. Equivalently, each pulse would be detected for many times instead of only once within every interval time θ, as illustrated in Fig. 1(c). As an example, if 1×16 MMI is used, while the time interval of optical pulse θ (75ps) is three times as much as the response time (25ps) of the detector (40GHz), the total node extending factor would be $M = 3$

× 16 = 48. In principle the time delay after each port of the MMI could be randomly chosen within interval time θ. For simplicity, the time delay is linearly chosen from 6ps (port of MMI-1) to 21ps (port of MMI-16) in our calculation. Figures 4(a)-4(c) illustrates the relation between MRA output and two output ports of A (MMI-6) and B (MMI-16), together with the values of the different channels in the output ports. The time delay after port A and B are 11ps and 21ps, respectively. Therefore, the total node number could be easily extended with a factor of 48. As an obvious consequence, if wider bandwidth detectors are adopted, the hierarchical structure can further enlarge the total node number without deteriorating the processing speed. It should also be mentioned that the application of MMI would cause the attenuated amplitude of the optical pulse train due to power splitting and insertion loss. The insertion loss is relatively small and can be neglected while the power splitting would degrade the tolerance for noise to some extent. This issue would be discussed in *Section 4.1* with more details.
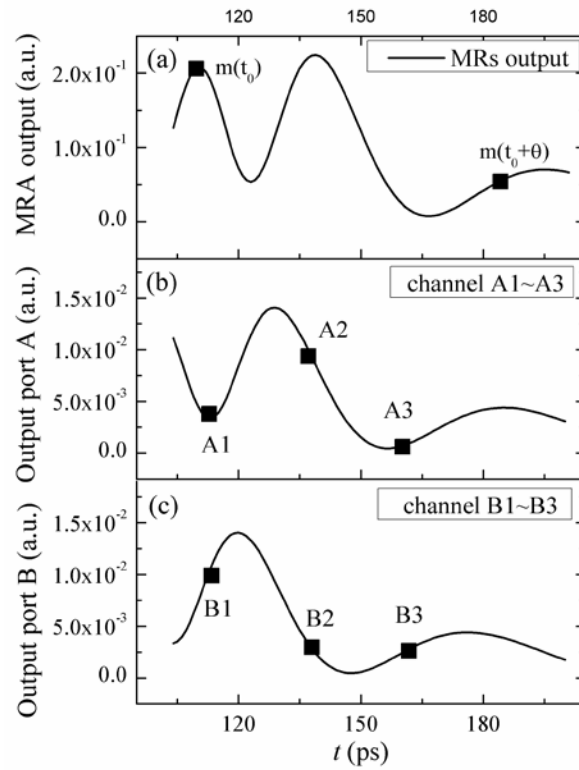

Fig. 4. Illustration of the secondary time-multiplexing. (a) Output intensity of the MRA after a single pulse input. (b) and (c) Output intensity of two different ports A and B. The output shape of each port is similar to (a) with deviations of a time delay of 11ps (port A) or 21ps (port B) and split intensity. Each port includes three channels (A1~A3 for port A and B1~B3 for port B). The value of each channels is represented by the notch.

3.5 Post-processing, training and evaluation

For convenience, each $b \times M$ output matrix $X_{out}^{i,j}$ is reshaped into a N dimensional vector X. A linear combination of the detected signal X is required as the post-processing of the system.

Therefore the final output of the system could be expressed as: $\hat{y} = \sum_{i=1}^{N} a_i X^i$, where X^i is the ith element of the vector. The weights a_i could be optimized to minimize the mean square error (MSE) between the output $\hat{y}$ and the target y over an adequate amount of data, which is known as a common training process.

$$MSE = \frac{1}{L} \sum_{n=1}^{L} (y(n) - \hat{y}(n))^2 \qquad (4)$$

Although minimizing MSE is straightforward, it is common to employ ridge regression to avoid over-fitting, which is adopted in our treatment. After the training process, the weights a_i are fixed and another set of data (the testing set) is used to evaluate the performance of the system by calculating the normalized mean square error between the target and the system's output:

$$NMSE = \frac{1}{L} \sum_{n=1}^{L} (y(n) - \hat{y}(n))^2 / \mathrm{var}(y) \qquad (5)$$

4. Simulation

In order to demonstrate the computing performance, our system has been applied to two benchmark tasks: signal classification and chaotic time series prediction. Here, the 3dB bandwidth of the PD and the phase modulator are considered as 40GHz, while the interval of the pulse train θ is set as 75ps, which is related to the rate of ~13G pulses per second. These parameters are chosen as a trade-off among processing speed, spectrum width of the pulse, reservoir size as well as the state of art devices. A white Gaussian noise is added on the final received signal to simulate the noise of the whole system. The signal to noise ratio (SNR) is defined as the intensity ratio between the output signal and the noise within the bandwidth of 40GHz. It is worth mentioning that although the system could be integrated on a silicon chip in principle, the parameters of each device are difficult to anticipate for actual simulation. Therefore, the parameters of some commercial devices are adopted in our simulations as listed in Table 2.

4.1 The achievable SNR with practical implementation

The performance of a RC is highly dependent on the SNR of the whole system [17]. And the power splitting of introducing MMI in our proposed RC would degrade the tolerance for noise to some extent. Thus, the achievable SNR of our proposed RC is first estimated with practical devices.

4.1.1 The maximum input power of optical signal

Generally, higher input power is helpful to achieve higher SNR. In our considered case, however, large input power would cause significant two photon absorption in the waveguide and therefore additional transmission loss would be introduced in consequence. According to [29] and taking the enhancement of micro-ring cavity into account, the average input power of the signal is chosen to be 4.3mW (corresponding to a peek power of 15mW) so that the loss induced by TPA (0.19dB/cm) is significantly smaller than the linear transmission loss(1.2dB/cm). The detailed calculation is shown in *Appendix. 2*.

4.1.2 The SNR at PD

The total insertion and transmission loss of the MRA and MMI is considered as 3dB. Thus, the input power to each PD is 0.13mW in the case of 1×16 MMI. The PD is considered as the u2t XPVD 2020 with a DC responsivity of 0.55 A/W according to the data sheet [30]. It could therefore be calculated that the average output voltage is 3.6mV (peak voltage of

13.8mV) with matched load resistance of 50 Ω. The noise power within 40GHz bandwidth is then calculated as 0.7nW, corresponding to an SNR of 26dB. Detailed calculation is given in *Appendix. 3*. It should be noted that for less multi splitting, such as 1×12 or 1×8, higher SNR could be expected.

4.1.3 The SNR at digitizers

Since the signal would be stored in digitizers, an analogue to digital converter (ADC) is required before storage. The ADC is considered as those used in the oscilloscope of Agilient (Keysight) 90000Q [31], in which the bandwidth could be as high as 65Hz. However, we cannot acquire the detailed parameters. Hence, the operating voltage is adopted from other high speed ADC. As shown in [32], the input voltage of most high speed ADC is around 1~3V. Here, we take the required signal level as 1V in our estimation. After each PD, the maximum output voltage is given as: $V_{max} = P_{max} \cdot \eta r$, where η and r are the respontivity and the match load resistance of the PD, respectively. The value is found to be 13.8mV, which is much lower than 1V. Thus, microwave power amplifier is required to match voltage between the output of PD and the input of ADC. The conversion gain of the amplifier is calculated as 37dB, which is considered to be realized with two cascaded microwave amplifiers of Giga-tronics 1040A. The noise figure of the each amplifier is considered as 8dB according to the datasheet in [33], therefore the SNR of the output signal could be calculated as ~20dB with the formula in [34]. The detail calculation method is discussed in *Appendix. 4*.

Table 2. Parameters of Devices for Practical Implement

Devices	Parameters	Symbol	Value
u2t XPVD 2020 (PD)	Responsivity	η (A/W)	0.55
	Dark Current	I_d (nA)	5
	Matched Load Resistance	r (Ω)	50
Giga-tronics 1040A	Power Gain	G(dB)	25
(Microwave Power Amplifier)	Noise Figure	NF(dB)	8
(ADC in Agilient [Keysight] 90000Q)	Input Voltage	V(mV)	1000
	Resolution	Re(bit)	8
	Bandwidth	B(GHz)	65

Till now, we have estimated that the maximum SNR of our proposed RC is about 20dB with average input power of 4.3mW. In the followed simulations, the SNR is considered as the value of <20dB.

4.2 Signal classification

The signal classification task aims to classify the square and sine waves. As shown in Fig. 5(a), a random concatenation of square and sine waves is set as input. The target output of square wave is one while that of sine wave is zero as seen in Fig. 5(b). The corresponding output of our RC system is shown in Fig. 5(c). It is clearly shown that the result presents a good coordination with the target. Over 300 digits are trained and 1000 digits are tested.

In order to investigate the influence of reservoir size and SNR, some simulations have been carried on. Figure 5(d) shows the error rate at different SNR levels with a fixed node number of 480. It could be seen that the error rate decrease from 5.5% to 0.3% with SNR varying from 16dB to 21dB. For the SNR range of 19~21dB, there is an error rate floor of ~0.3%. Furthermore, the error rate versus varied node number is shown in Fig. 5(e). The error rate would decrease from 0.9% to 0.5% as the node number increases from 120 to 480 with constant SNR of 20dB. In other reported optoelectronic or optical RC implementations, the error rate of this task is 0.15% with a sample rate of only 200kbps [9]. Our optical RC system

is competitive in error rate (0.5% vs. 0.15%) but much faster in sample rate (1.3Gbps vs. 200kbps). The detailed comparison is shown in Table 3.
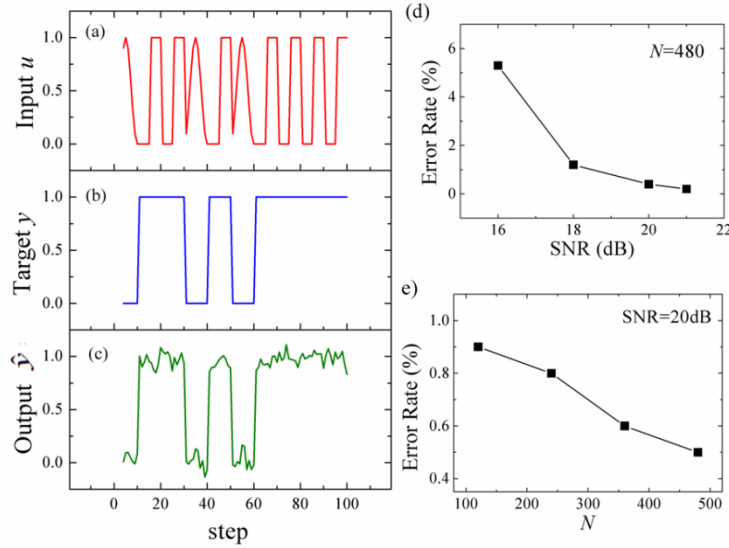

Fig. 5. Illustration and performance of signal classification task. (a) A set of input data, which is a random concatenation of square and sine waves. (b) The target output, one for square waves and zero for sine wave. (c) The output of the proposed optical RC system. (d) Relation between the error rate and SNR with constant node number of 480. (e) Relation between error rate and node number N with constant SNR of 20dB.

4.3 Chaotic time series prediction

The chaotic time series prediction is a much more challenging task that aims to make a single point prediction of a chaotic time series data obtained by experiments. We have adopted the same data as those used in the Santa Fe competition [35, 36]. A five-fold cross validation is used for training and testing. Figures 6(a) and 6(b) show a set of the original chaotic signal data and the output of our RC, respectively.

Similar to the previous task, our RC system is investigated with varied reservoir size and SNR. The results are shown in Figs. 6(c) and 6(d). In Fig. 6(c), error rate at different SNR level is shown with a node number fixed at 120. It is clear that the error rate shrinks down from 5.1% to 2.7% as the SNR increases from 14dB to 20dB. In Fig. 6(d), the performance with different node number is presented. The error rate changes from 4.7% to 2.7% as the node number increase from 10 to 120 at the SNR of 20dB. This result is better than that reported in [9], where the error rate is about 5.5% with a sample rate of about 13M.The error rate of the proposed RC system is close to the time-series prediction done by machine learning methods on software (below 1%) [37, 38]. It should be mentioned that the error rate of 0.5% in [37, 38] is achieved without considering noise. Although the chaotic time series prediction task is sensitive to noise, our implementation is still able to perform a low error rate with a fast prediction rate of 1.3Gbps. The comparison is also shown in Table 3.
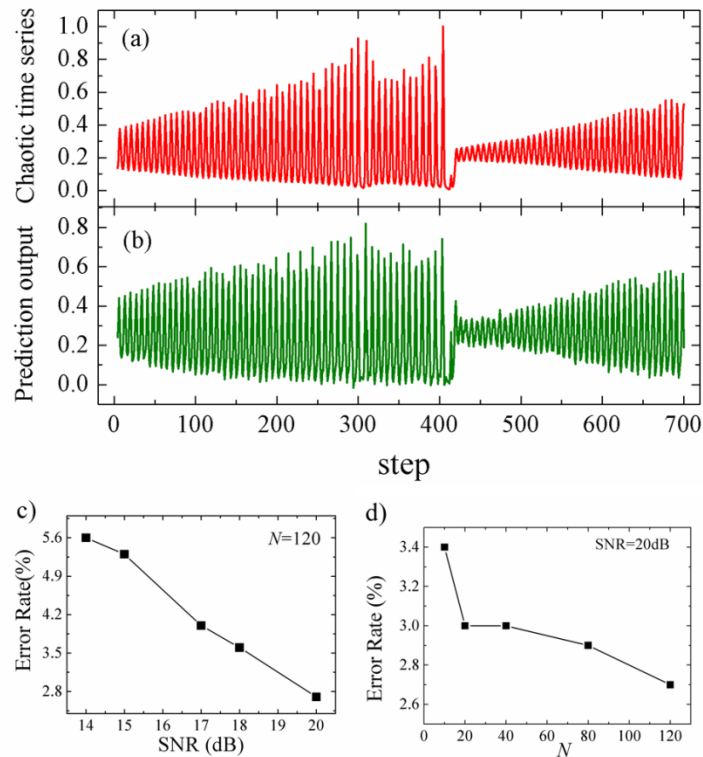
Fig. 6. Illustration and performance of chaotic prediction task. (a) A sample of target chaotic time-series data. (b) The prediction output of the proposed optical RC system. (c) Relation between error rate and SNR at constant node number of 120. (d) Relation between error rate and N at constant SNR of 20dB.

Table 3. Comparison of different RC implements

	Signal classification		Chaotic series prediction	
	Current work	Optoelectronic delay-feedback [9]	Current work	Optical delay-feedback [8]
error rate (%)	0.5	0.15	2.7	5.5
node number N	480	50	120	388
sample rate (bps)	1.3G	200k	1.3G	13M

Table 3 shows that this work has better performance considering both error rate and speed. It is worth mentioning that in all of the stimulations, the dynamical node number b is fixed at 10. The total node number N changes with extending node number. Thus, the speeding up of our proposed RC results from the proposed hierarchical structure, *i.e.* employing secondary time-multiplexing. Its nature could be described as follows. The first time-multiplexing is used to break the symmetry in a time scale comparable to the physical memory length. The physical implementation of MRA not only provides the nonlinear interactions between adjacent digits, but also presents further asymmetry within each time interval. This allows us to employ the secondary time-multiplexing within each time interval so that a large number of nodes can be obtained with a fast speed simultaneously. We speculate that the lower error rate in chaotic series prediction task results from the better SNR level in the all passive virtual network.

5. Conclusion

In conclusion, a hierarchical time-multiplexing structure is proposed for integrated photonic RC to achieve both a large node number and fast processing speed. Error rate of 0.5% for signal classification and 2.7% for chaotic time series prediction are achieved, with the sample rate of 1.3Gbps. Although the performance of the proposed RC is still dependent on properties of the devices, the hierarchical time-multiplexing structure could effectively alleviate the contradiction of processing speed and reservoir size in the delay-feedback implementation. The relatively small time delay for the extending nodes results in a better feasibility of realization on a large scale, comparing with the network-like optical implementation, in which large on-chip time delay is required. Furthermore, apart from the light source, the components employed in our proposed RC could be fabricated with CMOS compatible process and suitable for massive production. Thus, our implementation could be readily applied on single-chip integration.

It should be mentioned that although the hierarchical time-multiplexing is considered as to be achieved with the MRA in this work, it is also potential to serve as a more general method and implemented with other physical systems.

Appendix:

1. Micro-ring cavity model

A well-studied model for the ring cavity resonators is used to analyze the transfer relation of the system. r and t are the reflection and transmission index at the coupling spot, respectively. ϕ is the phase delay of one round trip in the cavity and τ is the amplitude transmission factor after propagating one round trip in the ring. $E_1 \sim E_4$ are electric field define as in Fig. 3. The relation between $E_1 \sim E_4$ could be written as:

$$E_4 = \tau E_3 \cdot \exp(i\phi) \tag{6}$$

$$E_3 = r \cdot E_4 + it \cdot E_1 \tag{7}$$

$$E_2 = r \cdot E_1 + it \cdot E_4 \tag{8}$$

Therefore:

$$\frac{E_2}{E_1} = -\exp(i\phi)\frac{\tau - r \cdot \exp(-i\phi)}{1 - r\tau \cdot \exp(i\phi)} \tag{9}$$

Transfer function of the ring cavity is calculated as:

$$Tr = \left|\frac{E_2}{E_1}\right|^2 = \frac{\tau^2 - 2r\tau\cos(\phi) + r^2}{1 - 2r\tau\cos(\phi) + r^2\tau^2} \tag{10}$$

Therefore the output of certain input pulses could be obtained as:

$\chi_{out} = \chi_{in}Tr$, where χ_{out} and χ_{in} are the output and input pulse in frequency domain. Their time domain feature could be easily obtained with inverse Fourier transformation.

In our simulation, the r is fixed at 0.9, which could be obtain by setting the proper gap between the micro-ring and the transmission waveguide. τ and ϕ could be calculated as:

$$\tau = d\gamma \tag{11}$$

$$\phi = \frac{dn_{eff}}{\lambda} \tag{12}$$

where λ and d is chosen to be 1549.5nm and 650μm, respectively so that the micro-ring cavities would operate at the resonant point. The Q value of the cavity could be expressed as Eq. (13).

$$1/Q = 1/Q_e + 1/Q_0 \qquad (13)$$

where Q_0 and Q_e are the intrinsic and external quality factor, respectively. Q_0 and Q_e could be calculated as Eq. (14) and Eq. (15), respectively, where ω is the angular frequency and c is the speed of light in free space.

$$Q_0 = \frac{\omega n_{eff}}{\gamma c} \qquad (14)$$

$$Q_e = \frac{\omega n_{eff} d}{|r|^2 c} \qquad (15)$$

The interface of the waveguide is shown in the subset of Fig. 3. The height of the ridge is set as Hr=50nm, where the total height is H=220nm, according to the thickness of silicon layer of the standard SOI wafer. The width of the ridge is chosen as W=1500nm. γ=1.2dB/cm is adopted according to the reported experimental results with state of art fabrication [16]. And n_{eff}=2.82 is used according to finite element method simulation result. Q_0 and Q_e are therefore calculated to be 3.8×10^5 and 3.9×10^4, respectively, corresponding to a total Q value of 3.9×10^4. In principle, a smaller γ/n_{eff} ratio would result in longer memory of the MRA, while small deviation of these parameters would have little impact on the performance of RC. In practical implementation, fabrication induce deviation of the parameters, especially d, can be compensated by thermal tuning of the micro-rings so that the cavity could work at the resonant point.

2. TPA transmission loss calculation

As shown in [29], the additional transmission loss caused by two photon absorption could be expressed as:

$$\gamma_{TPA} = \frac{\beta}{A_{eff}} P \qquad (16)$$

where β is the TPA coefficient which is considered as 6.7×10^{-12} m/W for silicon. A_{eff} is the effective area which is calculated to be 4.0×10^{-9}cm^2 and P is the light power in the waveguide.

The power in the cavity could be expressed as:

$$P_c = P_{in} |F|^2 \qquad (17)$$

where $|F|^2$ is the enhancement factor that can be expressed as:

$$|F|^2 = \frac{c}{Ln_{eff}\omega} \cdot \frac{4/Q_e}{(1/Q_e + 1/Q_0)^2} \qquad (18)$$

Substituting Eq. (17) and Eq. (18) to Eq. (16), the proper input power could be estimated. The maximum input power of 15mW would correspond to the loss induced by TPA (0.15dB/cm), which is significantly smaller than the linear transmission loss (1.2dB/cm).

3. Noise power calculation in PD

The noise power could be given as:

$N_{pd}=N_s+N_t$, where N_s and N_t are the shot noise and the thermal noise, respectively. N_s could be given by Eq. (19), where I_d and I_c are the dark current and signal induced current,

respectively. ΔB is the operating bandwidth and e is the electron unit. r corresponds to the matched load resistance. N_t could be given by Eq. (20), where k_B is the Boltzmann constant and T is the operating temperature.

$$N_s = 2(I_d + I_c) \cdot \Delta B \cdot e \cdot r \tag{19}$$

$$N_t = 4k_B T \cdot \Delta B \tag{20}$$

In our case, $I_c = P_{average} \cdot \eta = 0.07\text{mA}$, $I_d = 5\text{nA}$, $r = 50\Omega$, $T = 290\text{K}$, $\Delta B = 40\text{GHz}$, the calculated noise power in PD is therefore: $N_{pd} = 0.7\text{nW}$.

4. Noise power calculation in microwave power amplifier

According to Friis formula [34]:

$$NF = NF_1 + \frac{NF_2 - 1}{G_1} \tag{21}$$

where NF is the noise figure of the amplifiers and G is the power gain. According to the data sheet $NF_1 = 8\text{dB} = 6.3$, $G_1 = 25\text{dB} = 316$, therefore $NF = 6.3$.

According to the definition of noise figure by Friis [34] and the IRE (forerunner of the IEEE), noise figure NF is the ratio of input and output SNR when noise power of the input signal is considered as thermal noise level at room temperature T=290K.

$$NF = \frac{S_{in} / N_{in}}{S_{out} / N_{out}} = \frac{S_{in} / (k_B T \cdot \Delta B)}{G \cdot S_{in} / (G \cdot k_B T \cdot \Delta B + N_a)} = 1 + \frac{N_a}{G \cdot k_B T \cdot \Delta B} \tag{22}$$

where S_{in} and S_{out} is the average power of the input and output signal, respectively. N_{in} is the noise power of the input signal, which is considered as the thermal noise at room temperature: $N_{in} = k_B T \cdot \Delta B$; N_a is the noise power added by the amplifier; N_{out} is the output noise power, which is the sum of the amplified input noise and the noise power added by the amplifier: $N_{out} = G \cdot k_B T \cdot \Delta B + N_a$; G is the total gain.

Therefore the noise power added by the amplifier at room temperature could be calculated as:

$$N_a = (NF - 1) \cdot G \cdot k_B T \cdot \Delta B \tag{23}$$

In our case, $G = 5406$, $\Delta B = 40\text{GHz}$, $S_{in} = 0.26\mu\text{W}$, $N_{in} = N_{pd} = 0.7\text{nW}$.

Substituting all the above, we could get the result $N_a = 4.6\mu\text{W}$.

Therefore the output SNR could be calculated as:

$$SNR_{output} = \frac{S_{out}}{N_{out}} = \frac{GS_{in}}{GN_{in} + N_a} = 167 = 22dB \tag{24}$$

Since some other factors may affect the SNR level during the transmission and acquisition process, we conservatively estimate the final SNR at the digitizer to be ~20dB.