

Node-Reduction Through Joint Optimization of Input and Readout Layers in Photonic Reservoir Equalization

Ruben Van Assche¹, Sarah Masaad¹, and Peter Bienstman¹, *Member, IEEE*

Abstract—Photonic reservoir computing is a machine learning paradigm in which a recurrent neural network remains fixed while only the output weights are trained. This makes it a well-suited approach for high-speed signal equalization in optical communication systems, offering a trainable, low-power, and low-complexity solution. However, achieving strong performance typically requires relatively large network sizes, as learning is confined to the output layer. To address this, we investigate the role of trainable input mappings alongside conventional output weight optimisation. Across a range of short- and mid-reach IM/DD transmission scenarios, reaching up to 200 km for a 28 Gb/s NRZ signal, improvements of over two orders of magnitude in BER are achieved. In matched comparisons with the same number of programmable complex optical weights, this enables halving the reservoir-core node count while maintaining comparable performance. Furthermore, we show that this approach effectively extends the memory of the reservoir, resulting in over three orders of magnitude improvement in memory-intensive tasks. These results also show that starting at 16 nodes a performance of at least one to two magnitudes better than both a complexity matched FFE and a Volterra filter of second order are reached.

Index Terms—Optical fiber communication, optical signal processing, optimization methods, photonic integrated circuits, reservoir computing, silicon photonics.

I. INTRODUCTION

INTENSITY-MODULATION/DIRECT-DETECTION (IM/DD) optical links face growing equalization demands as symbol rates and transmission distances increase. Chromatic dispersion, bandwidth limitations, and device nonlinearities introduce intersymbol interference (ISI) that classical feed-forward equalizers (FFEs) and Volterra-type nonlinear equalizers can address, but at rapidly increasing implementation costs [1], [2], [3]. This has motivated research

into machine-learning-based equalizers that can offer better performance–complexity trade-offs [2], [3].

Reservoir computing (RC) is attractive in this context. It consists of a fixed dynamical system (the reservoir) that maps the input waveform to a high-dimensional state. This fixed recurrent neural network is followed by a trainable linear readout layer that implements the task-specific learning [4], [5], [6]. RC performance depends on the balance between memory and nonlinear processing of past inputs [7], [8]—a property directly relevant to equalization. Photonic implementations are particularly compelling given their high bandwidth, low latency, and rich transient dynamics. Early demonstrations established photonic RC in delay-based and integrated platforms [9], [10], [11], [12], and subsequent work extended these toward increasingly capable processors [13], [14], [15], [16]. Reservoir-based equalization has since been demonstrated in both IM/DD and coherent links [3], [17], [18], [19], [20], [21], and integrated photonic neural equalizers have shown similar promise [22], [23], [24]. Despite this progress, the input transformation in most RC equalization studies is treated as fixed, random, or heuristic [10], [11], [25]. Yet input encoding choices matter: deterministic masks can outperform random masking [26], mask temporal structure strongly affects performance [27], timing relations between injected signals and reservoir delay influence memory capacity [28], [29], delayed-input strategies can improve physical RC systems [30], [31], and input training can meaningfully improve performance compared to random input weighting in photonic computing systems [32].

These results motivate further study in treating the input layer as a design variable, especially in the context of telecom equalization problems. This paper studies that problem in a four-port all-optical silicon photonic reservoir equalizer [33] with integrated optical input and readout layers. The simulated implementation targets a passive integrated photonic reservoir in which recurrent mixing is provided by multimode-interferometer-based nodes connected by waveguide delay lines, while the programmable input and readout layers are represented by passive complex optical weights. Within this fixed architecture, the central question is whether the joint optimization of input and output weights changes the node-scaling behavior strongly enough to reduce the node count required to reach a given equalization performance. The present work focuses on IM/DD equalization rather than coherent long-haul transmission. The considered non-return-to-zero on-off keying (NRZ)

Received 22 April 2026; revised 17 June 2026; accepted 8 July 2026. Date of publication 10 July 2026; date of current version 2 October 2026. This work was supported in part by European Commission in the Horizon 2020 programme through Project Nebula under Grant 871658, in part by Horizon Europe programme through Project Prometheus under Grant 101070195, in part by POSTDIGITAL+ under Grant 101169118, in part by Necho under Grant 101046329, in part by Neuropuls under Grant 101070238, in part by NEHIL under Grant 101194363, in part by RESPITE under Grant 101098717, and in part by Belgian FWO and EOS programmes. (*Corresponding author: Ruben Van Assche.*)

The authors are with Photonics Research Group, INTEC, Ghent-University-imec, 9052 Ghent, Belgium, and also with the imec, 3001 Leuven, Belgium (e-mail: ruben.vanassche@ugent.be).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2026.3712478>.

Digital Object Identifier 10.1109/JLT.2026.3712478

TABLE I

FIRST-ORDER HARDWARE ACCOUNTING FOR THE MATCHED INTRA-ARCHITECTURE COMPARISON. IN THE MATCHED COMPARISONS CONSIDERED HERE, THE (N)-NODE JOINTLY TRAINED INPUT-READOUT SYSTEM AND THE (2 N)-NODE READOUT-ONLY SYSTEM USE THE SAME TOTAL NUMBER OF PROGRAMMABLE COMPLEX OPTICAL WEIGHTS, NAMELY (2 N). THE RECURRENT-LINK COUNT IS SHOWN AS A MEASURE FOR RESERVOIR-CORE INTERCONNECT COMPLEXITY. THIS TABLE IS NOT INTENDED AS A COMPLETE LAYOUT-LEVEL AREA OR POWER MODEL.

Architecture	Reservoir nodes	Recurrent links	Readout weights	Input weights	Total Weights
Readout-only, $2N$ nodes	$2N$	$4N$	$2N$ complex	0	$2N$
Joint input+readout, N nodes	N	$2N$	N complex	N complex	$2N$

OOK) links are used as controlled impairment benchmarks for studying the reservoir architecture and training strategy. They are not intended to represent the full range of modern long-haul coherent systems, where higher-order modulation formats, coherent detection, carrier recovery, and additional complex-valued DSP steps such as carrier recovery are typically required.

This is architecturally relevant because, in integrated photonic reservoirs, node count directly determines the number of interferometric elements, recurrent waveguide interconnects, and routing overhead. In the matched comparisons considered here, an (N)-node jointly trained input-readout system is compared with a (2 N)-node readout-only system. Both systems use the same total number of programmable complex optical weights: (N) input and (N) readout weights in the jointly trained case, versus (2 N) readout weights in the readout-only case. The difference is therefore not an increase in the number of programmable weights, but a reallocation of the same programmable-weight budget together with a factor-of-two reduction in reservoir-core size. A first-order accounting of this comparison is shown in Table I. The matched comparison therefore keeps the programmable-weight count fixed while reducing the recurrent reservoir core. Under a common implementation of the programmable weights, the first-order tuning/control count associated with these weights is comparable between the two systems. The expected hardware saving is in the passive reservoir core, through fewer reservoir nodes, fewer recurrent interconnects, and reduced routing footprint. A complete chip-level cost and energy analysis would require a concrete layout and implementation-specific assumptions, and is left for future work.

The rest of this paper is structured as follows. In Section II the hybrid simulation framework that was used to generate high-speed telecom signals and process them is explained, along with the optimisation scheme and baselines. Next, in Section III, simulations of varying fiber lengths, launch power and reservoir sizes can be found along with their discussion. Finally, Section IV concludes the paper.

II. OPTICAL RESERVOIR SIMULATIONS

This section introduces the hybrid simulation framework and the comparison protocol used to support the architectural node-scaling claim of this paper. The optical communication link is simulated in VPIphotonics [34], while the all-optical reservoir is simulated in the Photontorch Python library [35]. This separation allows transmission impairments and the photonic equalizer to be modeled using tools that are well matched to their respective physical domains.

A. Optical Communication Model in VPIphotonics

The distorted communication signal used as input to the reservoir equalizer is generated using VPIphotonics. The considered transmitter-receiver chain is shown schematically in Fig. 1. A continuous-wave laser is intensity modulated using a Mach-Zehnder modulator (MZM) driven by the transmitted data sequence. In the VPIphotonics communication model, this MZM is used only as the transmitter modulator that generates the NRZ OOK waveform. It is modeled as an ideal external modulator with bias voltage of 1 V, $V_{\pi,DC} = 2$ V, $V_{\pi,RF} = 2$ V, zero insertion loss, 30 dB extinction ratio, and ideal electro-optic bandwidth response. These assumptions isolate the equalizer comparison from transmitter hardware nonidealities; practical transmitter impairments are therefore not the focus of the present study. The modulated optical signal then propagates through the link, where channel impairments such as chromatic dispersion, Kerr nonlinearities, and controlled noise loading are introduced.

The resulting received waveform contains the cumulative effect of transmitter modulation, optical propagation, and receiver-side noise loading. In discrete time, the detected communication signal can be written in the generic form

$$r_k = \sum_{i=0}^L h_i s_{k-i} + \eta_k + \epsilon_k, \quad (1)$$

where s_k denotes the transmitted symbol, h_i the effective linear channel response, L the channel memory, η_k additive noise, and ϵ_k residual nonlinear distortion. The equalization task is to approximate the inverse mapping

$$\hat{s}_k = F(r_k, r_{k-1}, \dots, r_{k-L}). \quad (2)$$

In general, we use NRZ OOK transmission at 28 Gb/s with 8 samples per symbol and sweep operating points over transmission distance, launch power, and receiver-side optical signal-to-noise ratio (OSNR).

B. Reservoir Model in Photontorch

The equalizer, see fig. 2, takes the abstract form

$$\mathbf{x}(t + dt) = \mathbf{R}\mathbf{x}(t) + \mathbf{w}_{in}^T \mathbf{u}(t), \quad y(t) = g(\mathbf{w}_{out}^T \mathbf{x}(t)), \quad (3)$$

where $\mathbf{x}$ is the reservoir state, $\mathbf{R}$ the fixed weight matrix of the reservoir, $\mathbf{w}_{in}$ the fixed input coupling (trainable under joint optimization), $\mathbf{u}$ the injected optical field, $\mathbf{w}_{out}$ the optical readout weights, and $g(\cdot)$ the photodetection nonlinearity. While the physical reservoir computing operates in continuous time, simulations are run in discrete time with 8 samples per symbol.

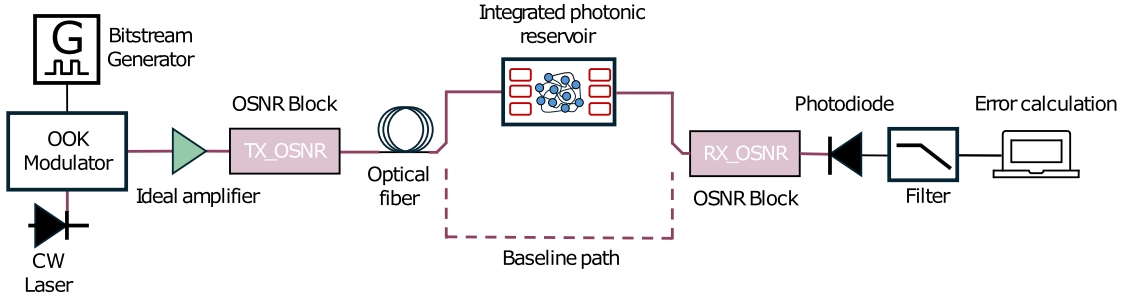


Fig. 1. Optical communications setup. Upper and lower paths respectively show the setups with and without the photonic reservoir. The lower path is used for contextual baseline comparisons. CW: continuous wave. OOK: on-off keying. OSNR: optical signal-to-noise ratio.

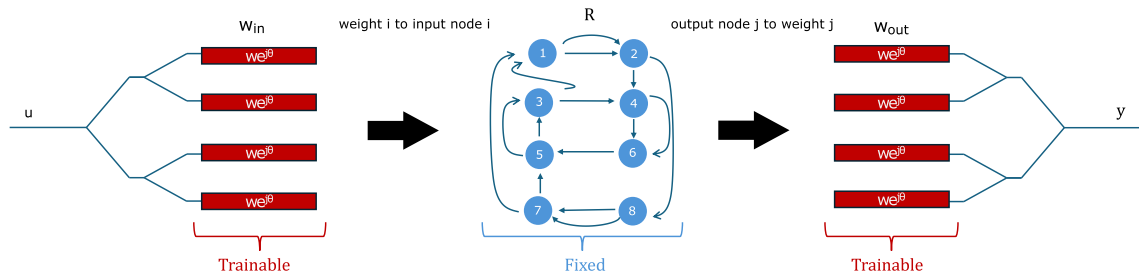


Fig. 2. Schematic overview of the four-port reservoir computing architecture. An input signal u gets transmitted to a trainable input layer with weights w_{in} . This input is used in a fixed, random reservoir, here configured in the four-port architecture. Finally the states from the reservoir are extracted and recombined into an output y by being transmitted through a trainable output layer with weights w_{out} . The black arrows represent a 1-to-1 mapping from the trainable weights to the input of the nodes (left arrow) or the output of the nodes (right arrow). This is simulated through waveguide connections with the same properties as the waveguides in the reservoir.

Crucially, R is not a trainable weight matrix: internal mixing arises physically from waveguide propagation, with each interconnect contributing attenuation, phase accumulation, and propagation delay. Fabrication-induced variations—waveguide-length errors, sidewall roughness—make each reservoir realization unique but fixed [20], [21].

The inter-node waveguides within the reservoir act as physical delay lines. In all simulations reported here, the interconnection delay is kept fixed and is chosen to be exactly half the target symbol period. This delay is not optimized during training; it is a fixed architectural parameter inherited from previous work on integrated photonic reservoirs for telecom equalization [20], [36], [37], where delay-line lengths were studied relative to the bit/symbol period and half-symbol-scale delays were found to provide an effective balance between temporal mixing and memory. Physically, such a delay allows neighbouring reservoir nodes to interfere with sub-symbol-shifted versions of the waveform, while recurrent circulation through multiple interconnects still gives access to longer symbol-scale memory. System memory is thus determined by the physical propagation network rather than trainable recurrent weights.

The equalizer is modeled in Photontorch as a passive spatially distributed photonic network following the four-port architecture [20], [21], [33]. In the simulations reported here, the reservoir is implemented as a passive SiN four-port architecture composed of multimode-interferometer-based nodes connected

by waveguides that provide delayed optical feedback and recurrent mixing. The propagation model uses an effective refractive index of 2.34 and a group index of 3.40. The interconnection waveguide length is 1.25 mm, and propagation loss is included using a waveguide loss of 25 dB/m. The physical interconnection delay is kept fixed at half a symbol period in all reported simulations.

Each reservoir node is implemented as a 3×3 multimode interferometer (MMI). Two ports receive fields from neighbouring reservoir nodes, two ports route fields onward to neighbouring nodes, one port is used for external optical input injection, and one port is used for state readout. The inter-node waveguides therefore act simultaneously as optical delay lines and recurrent interconnects. The recurrent matrix R in (3) is not a freely chosen numerical matrix; it is the effective optical transfer induced by the fixed MMI scattering matrices, waveguide connectivity, propagation loss, accumulated phase, and propagation delay. Different reservoir realizations correspond to different fixed internal phase in the recurrent waveguides, or input waveguides, with the same four-port topology and physical parameters. This four-port architecture is the evolution of the swirl reservoir topology in which every node participates in the recurrent network through exactly two inter-node input and two inter-node output connections. This avoids the asymmetric edge splitters and associated radiation losses of the original swirl layout [33], [38], while preserving recurrent signal circulation, optical input

injection, and state extraction within a compact planar topology. The simulated reservoir core is therefore described at the circuit level rather than as an abstract fully connected recurrent neural network. Optical splitting and recombination inside the reservoir are provided by the MMI scattering relations and by coherent propagation through the waveguide network. The programmable input and readout stages are treated separately from the fixed reservoir core, as described below.

C. Programmable Optical Input and Output Layers

The new trainable degree of freedom considered in this work is the optical input layer. Rather than injecting the distorted signal through a fixed coupling arrangement, the incoming optical field is distributed across the driven reservoir input channels using programmable passive complex optical weights. In an integrated implementation, such weights can be realized with Mach–Zehnder-interferometer-type weighting elements that control the effective amplitude and phase applied to each driven input channel. These elements are modeled here at the level of their effective complex transmission coefficients, subject to passive amplitude bounds, rather than by a detailed thermal or electro-optic tuning circuit.

Let $u_k \in \mathbb{C}$ denote the incoming optical field launched toward the equalizer. The optical input layer maps this field to the driven reservoir input channels according to

$$\mathbf{w}_{\text{in}}^T u(t), \quad (4)$$

where $\mathbf{w}_{\text{in}}^T \in \mathbb{C}^{P \times 1}$ is the trainable vector of effective input weights and P is the number of input channels.

Each input weight is represented as

$$w_{\text{in},i} = a_i e^{j\phi_i}, \quad (5)$$

with amplitude a_i and phase ϕ_i . Because the input layer is passive, the amplitude is bounded by the maximum transmission allowed by the insertion loss of the interferometer,

$$0 \leq a_i \leq A_{\text{max}}. \quad (6)$$

Hence, the trainable optical input layer is constrained to physically realizable passive complex weights of the form

$$\mathbf{w}_{\text{in}} = \begin{bmatrix} a_1 e^{j\phi_1} \\ a_2 e^{j\phi_2} \\ \vdots \\ a_P e^{j\phi_P} \end{bmatrix}, \quad 0 \leq a_i \leq A_{\text{max}}. \quad (7)$$

The optical output (i.e. readout) stage combines the reservoir states coherently according to

$$y(t) = \mathbf{w}_{\text{out}}^T \mathbf{x}(t), \quad (8)$$

where $\mathbf{w}_{\text{out}} \in \mathbb{C}^N$ contains the complex readout weights. The resulting optical field is then photodetected, so the final electrical output depends nonlinearly on the coherent optical recombination. The processor is therefore not an optical front-end followed by a digital linear regressor, but an all-optical computation block in which both the input projection and the readout projection are implemented in the optical domain.

For the node-scaling experiments, we consider three input-node policies: *all*, *center*, and *spread*. In the *all* policy, all available input nodes are driven. The *center* and *spread* policies instead select a subset of input nodes from the full reservoir grid.

The *center* policy is designed to concentrate the driven inputs near the middle of the reservoir. It first searches for a rectangular region inside the full reservoir that can contain the required number of driven nodes. Among all such candidate rectangles, it prefers the most compact one; if several are equally compact, it chooses the one whose shape best matches the overall reservoir aspect ratio. This rectangle is then placed at the center of the reservoir grid, and the driven nodes are selected from the middle outward so that the chosen inputs remain as concentrated as possible around the geometric center.

The *spread* policy follows the opposite principle. It is designed to distribute the driven inputs as broadly as possible over the reservoir grid. Starting from a node near the outer part of the reservoir, additional driven nodes are added one by one using a greedy farthest-point rule based on Euclidean distance on the two-dimensional grid. At each step, the next node is chosen to maximize its distance from the nodes already selected, yielding a set of driven inputs that is distributed across the full reservoir footprint.

D. Training Objective, Node-Scaling Comparison, and Fairness

The trainable parameters are optimized using the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS).

For each set of weights, the VPI-generated waveform is propagated through the Photontorch reservoir. A warm-up interval is discarded to suppress transient effects. The remaining data are split into 8192 training symbols, a validation set of 32768 symbols, and a test set consisting of 10 blocks of 32768 symbols all generated with a different random seed. Receiver-side conditioning before evaluation consists of low-pass filtering at 28 GHz followed by normalization.

The training objective is the mean-squared error (MSE) evaluated at the symbol decision points. To find the optimal decision point, a discrete output-target latency sweep is performed to align the equalizer output to the target symbol sequence. This sweep ranges from 0 to 160 samples, corresponding to 20 symbol intervals at 8 samples per symbol. The selected latency is an alignment parameter of the training and evaluation pipeline; it does not modify the physical delay structure of the reservoir itself. While we use the MSE for training, for testing and validation, the bit-error rate (BER) and statistical BER [39] are used as metrics.

The central comparison of this paper is between matched variants of reservoirs with a different size:

- 1) an N -node reservoir with jointly optimized input and output (readout) weights;
- 2) a readout-only reservoir with $2N$ nodes.

The purpose is to determine whether the performance gain obtained by joint optimization is large enough to offset a factor-of-two reduction in reservoir size.

All reservoir variants use the same communication waveforms, train/test split, latency sweep, receiver-side filtering and normalization, and the same evaluation metrics.

To evaluate whether these findings remain consistent over different random reservoir initialisations, the simulations for each operating point are repeated over 10 random reservoir phase realizations. In readout-only experiments, the fixed input phases are additionally randomized using a separate random seed to avoid an artificially symmetric excitation pattern. In the simulations, a specific architecture seed determines the fixed phase realization of the reservoir interconnection waveguides. Thus, all random reservoir instances share the same four-port topology, MMI-node model, waveguide length, propagation loss, effective index, group index, and half-symbol delay, but differ in their internal accumulated phases. In readout-only experiments, a separate input-layer seed is used to randomize the fixed input phases and avoid an artificially symmetric excitation pattern.

For each latency, both the readout-only and joint input-readout optimizations use the same BFGS algorithm, latency-selection protocol, hyperparameter sweep over initialization radius, and the same maximum training budget. In practice, both scenarios were observed to reach comparably good local optima under this common protocol, although the joint optimization typically required more function evaluations. To quantify the additional offline optimisation effort, we reran representative optimisation experiments for 16-, 24-, and 32-node reservoirs using both BFGS and CMA-ES. These runs compare a gradient-based method with an evolutionary strategy at the 200 km, 15 dBm operating point, using the all-input configuration. Joint input-readout optimisation required approximately one order of magnitude more optimiser iterations than readout-only optimisation: $8.4\text{--}10.2\times$ for BFGS and $12.6\text{--}13.9\times$ for CMA-ES. This increased cost affects offline calibration/training only; after training, the optical weights are fixed during inference. It should be noted that both in the case of the joint training or the readout-only training, the optimizer always converged to a solution. The proposed method should therefore not be interpreted as reducing training complexity. Rather, it explicitly trades greater offline optimization effort for a reduction in the reservoir size required to reach a given equalization regime.

For comparisons based on the model size, we count the number of real-valued trainable optical degrees of freedom. Each complex optical weight therefore contributes two real degrees of freedom, corresponding to its effective amplitude and phase parameters. Unless stated otherwise, degree-of-freedom counts refer only to trainable optical weights and not to fixed architectural choices such as input-node policy.

For context and benchmarking, we also implemented a linear FFE and a second-order nonlinear Volterra equalizer on the received waveform. In both cases, taps were spaced by half a bit period, matching the temporal sampling granularity used for the reservoir computer. The number of FFE taps was chosen such that the number of real-valued trainable coefficients matches the number of real-valued trainable degrees of freedom of the optical network. The second-order Volterra baseline was instead chosen to match the same memory span as the FFE and therefore has a quadratically larger number of trainable degrees of freedom.

III. RESULTS

A. Fiber-Length Sweeps At Fixed Receiver OSNR

We first compare three systems as a function of transmission distance at a fixed receiver OSNR of 30 dB: an 8-node reservoir with joint input-readout optimization, an 8-node readout-only variant, and a 16-node readout-only variant.

Fig. 3(a) shows the BER as a function of fiber length for a launch power of 0 dBm. Several trends are clear. First, joint optimization provides a substantial improvement over the 8-node readout-only variant across the full sweep. Across all lengths, the median BER of the jointly trained 8 node RC is roughly one to two orders of magnitude below that of readout trained 8 node RC. It is also clear that a gap smaller than an order of magnitude persists between the jointly trained 8 node RC and the readout trained 16 node RC. Thus, over the short-to-intermediate-length regime, jointly optimizing the optical input and readout allows an 8-node reservoir to reach the same BER order as a 16-node readout-only variant with half the amount of reservoir hardware.

Beyond approximately 180 km, all three systems deteriorate sharply and the gap between them narrows. In this regime, the underlying channel impairment becomes too strong for the smaller jointly optimized system and the 16-node readout-only system, as both move into a comparatively high-error regime. This indicates that joint optical-layer optimization can expose more useful reservoir features, but does not remove the operating-window limitations of the architecture and the memory requirements of the task. Compared to the baseline FFE and Volterra filter, it can be seen that the reservoirs with an equal number of degrees of freedom consistently outperform these.

The corresponding sweep at 15 dBm launch power is shown in Fig. 3(b). The same qualitative ranking is observed, where the jointly trained 8 node RC consistently outperforms the 8-node readout-only variant and follows the performance of the readout-only 16 node RC. It can also be seen that for this more nonlinear and difficult task, at high lengths the second-order Volterra filter with 33 taps, benefiting from its higher nonlinear computational power, does perform at the same order of magnitude as the RC's starting at 140 km, where performance in general collapses to the order of 10^{-1} . Taken together, Fig. 3(a) and Fig. 3(b) show that the benefit of trainable optical encoding is robust within the explored operating window.

B. Node Scaling at a Challenging Operating Point

To study the node-scaling question in more detail, we next fix a challenging operating point at 200 km fiber length, 15 dBm launch power, and 30 dB receiver OSNR, and examine how performance evolves with reservoir size and input-node policy. The considered policies are the earlier explained all, spread and center policies both for joint training of input and readout layer, and readout-only training.

Fig. 4(a) shows the node-evolution of the median BER over the 10 paired seed realizations. The most important observation is that the jointly optimized systems improve much more rapidly with node count than the readout-only variants. In particular, the jointly trained all configuration already reaches a much lower

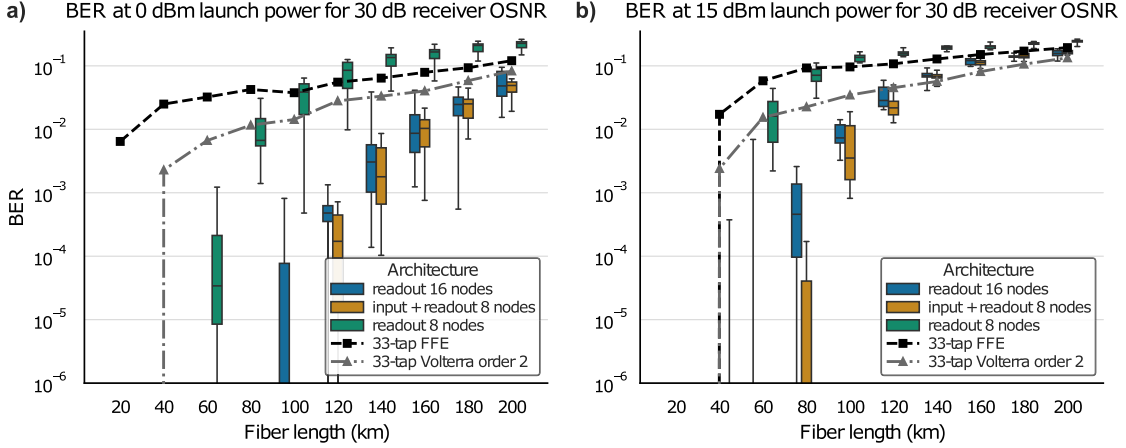


Fig. 3. BER versus fiber length at (a) 0 and (b) 15 dBm launch power and 30 dB receiver OSNR for an 8-node jointly optimized reservoir, an 8-node readout-only reservoir, and a 16-node readout-only reservoir. The jointly optimized 8-node variant consistently improves on the 8-node readout-only variant and, over a useful short-to-intermediate-length regime, approaches the BER range of the 16-node readout-only variant.

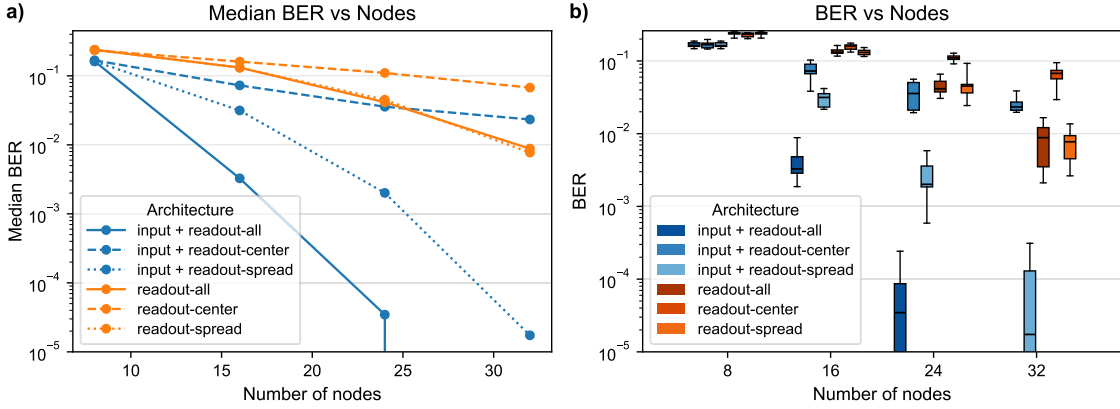


Fig. 4. (a) Median BER and (b) boxplot over 10 seeds versus reservoir size. The jointly optimized configurations, particularly input + readout-all, improve much more rapidly with node count than the readout-only variants, indicating that trainable optical encoding increases the usefulness of additional reservoir nodes in this operating regime.

error regime at 16 nodes than any of the 16-node readout-only variants, and by 24 nodes it is separated from all readout-only systems by several orders of magnitude in median BER. For all readout-only variants, performance improves only gradually with increasing reservoir size. Once the optical input layer is trained jointly with the readout, the same increase in node count becomes much more effective. The jointly trained systems make better use of each additional node, indicating that the learned optical excitation exposes a more useful subset of the reservoir's delayed interference features.

Fig. 4(b) provides boxplots of the same data to show the spread of the bit error rates over these architectures.

The same conclusion is supported when performance is plotted against trainable real-valued degrees of freedom rather than node count. Fig. 5 shows that jointly optimized input-readout architectures achieve similar or lower BER than readout-only variants at comparable numbers of trainable optical degrees of freedom. This again indicates that the observed performance is

not merely a consequence of parameter count, but of how the trainable optical resources are allocated within the architecture.

C. Effect of Input-Node Policy

The node-scaling results also show that the choice of input-node policy matters. Among the jointly optimized systems, the 'all' policy consistently provides the strongest scaling trend, while the 'spread' policy also performs well at larger node counts. By contrast, the 'center' policy improves more slowly and remains clearly above the best jointly optimized configuration. This indicates that the trainable optical front-end benefits from access to a broader set of injection channels: the input layer is most effective when it can redistribute optical amplitude and phase over multiple excitation paths and thereby reshape the delayed quadratic feature map of the reservoir.

For the readout-only variants, the differences between input-node policies are smaller and remain within a comparatively

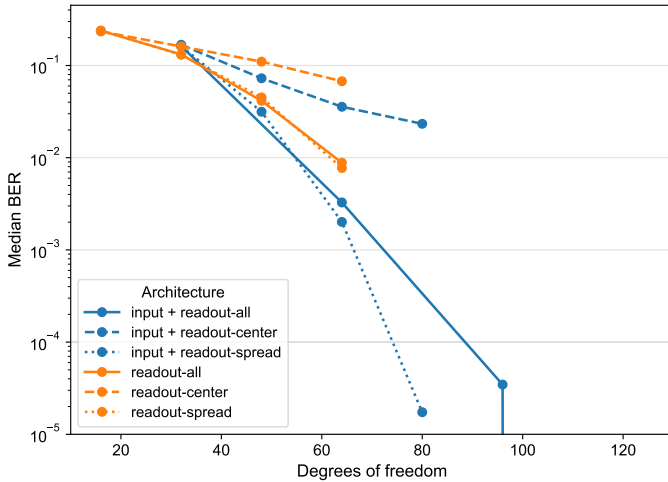


Fig. 5. Median BER versus the number of real-valued trainable degrees of freedom for the same operating point as in Fig. 4. Each complex optical weight is counted as two real degrees of freedom. Jointly optimized input–readout architectures achieve substantially lower BER than readout-only variants at comparable trainable degrees of freedom, indicating that the observed gain is not merely a consequence of parameter count but of how the trainable optical resources are allocated within the architecture.

poor error regime. This again supports the interpretation that the main gain does not arise from input placement alone, but from the ability to jointly optimize how the distorted waveform is injected into the reservoir and how the resulting states are optically recombined at the output.

IV. CONCLUSION

Scaling computation on photonic integrated circuits is limited by higher hardware costs and increased complexity. We investigated whether a trainable optical input layer can change the node-scaling behavior of a four-port photonic reservoir sufficiently to reduce the photonic hardware required to reach a given equalization regime. Using a hybrid VPIphotonics–Photontorch framework, jointly optimized input–readout systems were compared against matched readout-only variants under identical simulation and evaluation conditions. The central result is that joint optimization provides a clear node-scaling advantage within this architecture. The jointly optimized 8-node reservoir consistently outperformed its readout-only counterpart and, over a useful short-to-intermediate fiber-length regime, approached the BER of a 16-node readout-only system. At 200 km, the advantage sharpened further: jointly optimized systems scaled substantially better with node count, with the strongest configurations well beyond the range reached by readout-only training. Interpretively, the four-port reservoir acts as a fixed delayed linear optical feature generator, with quadratic interactions arising from coherent recombination and photodetection. The trainable input layer reshapes how the fixed interference structure of the reservoir is excited and presented to the detector, giving the readout access to a more useful feature map for inverse-channel recovery. The node-scaling benefit therefore reflects a more effective use of existing reservoir dynamics, not an additional

memory capacity. For the matched comparisons considered here, an N -node jointly optimized architecture uses the same number of weights as a readout-only $2N$ -node architecture while requiring fewer reservoir nodes and recurrent interconnects. This implies a reduction of photonic hardware by a factor of 2 in reservoir-core nodes and recurrent interconnects for the matched cases studied here., not a mere redistribution of trainable weights. The benefit does carry one practical cost however: greater offline optimization effort in exchange for a smaller inference-time reservoir core.

Experimental validation, explicit area-control trade-off analysis, and extension to coherent detection, higher-order modulation formats and different reservoir architectures are natural directions for future work. The present simulations include circuit-level propagation loss, MMI-based recurrent mixing, passive programmable input/readout weights, and fixed physical delay lines, but they do not yet include a full layout-level model of thermal tuning power, fabrication tolerances, electronic control circuits, or calibration drift.

REFERENCES

- [1] S. J. Savory, “Digital filters for coherent optical receivers,” *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, Jan. 2008. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-16-2-804>
- [2] F. D. Ros et al., “Machine-learning-based equalization for short-reach transmission: Neural networks and reservoir computing,” *Proc. SPIE*, vol. 11712, 2021, Art. no. 1171205, doi: [10.1117/12.2583011](https://doi.org/10.1117/12.2583011).
- [3] S. Wang, N. Fang, and L. Wang, “Signal recovery based on optoelectronic reservoir computing for high speed optical fiber communication system,” *Opt. Commun.*, vol. 495, 2021, Art. no. 127082. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003040182100331X>
- [4] W. Maass, T. Natschlager, and H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations,” *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.
- [5] H. Jaeger, “The “echo state” approach to analysing and training recurrent neural networks—with an erratum note,” *German Nat. Res. Center Inf. Technol., Tech. Rep.* 148, 13, 2001.
- [6] M. Lukoševičius and H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” *Comput. Sci. Rev.*, vol. 3, no. 3, pp. 127–149, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1574013709000173>
- [7] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, “Information processing capacity of dynamical systems,” *Sci. Rep.*, vol. 2, 2012, Art. no. 514.
- [8] M. Inubushi and K. Yoshimura, “Reservoir computing beyond memory–nonlinearity trade-off,” *Sci. Rep.*, vol. 7, 2017, Art. no. 10199.
- [9] K. Vandoorne et al., “Toward optical signal processing using photonic reservoir computing,” *Opt. Exp.*, vol. 16, no. 15, pp. 11182–11192, Jul. 2008. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-16-15-11182>
- [10] L. Appeltant et al., “Information processing using a single dynamical node as complex system,” *Nature Commun.*, vol. 2, 2011, Art. no. 468, doi: [10.1038/ncomms1476](https://doi.org/10.1038/ncomms1476).
- [11] Y. Paquot et al., “Optoelectronic reservoir computing,” *Sci. Rep.*, vol. 2, 2012, Art. no. 287. [Online]. Available: <https://www.nature.com/articles/srep00287>
- [12] K. Vandoorne et al., “Experimental demonstration of reservoir computing on a silicon photonics chip,” *Nature Commun.*, vol. 5, 2014, Art. no. 3541, doi: [10.1038/ncomms4541](https://doi.org/10.1038/ncomms4541).
- [13] A. Katumba, J. Heyvaert, B. Schneider, S. Uvin, J. Dambre, and P. Bienstman, “Low-loss photonic reservoir computing with multimode photonic integrated circuits,” *Sci. Rep.*, vol. 8, 2018, Art. no. 2653.
- [14] M. Nakajima, K. Tanaka, and T. Hashimoto, “Scalable reservoir computing on coherent linear photonic processor,” *Commun. Phys.*, vol. 4, 2021, Art. no. 20, doi: [10.1038/s42005-021-00519-1](https://doi.org/10.1038/s42005-021-00519-1).
- [15] Y.-W. Shen et al., “Deep photonic reservoir computing recurrent network,” *Optica*, vol. 10, no. 12, pp. 1745–1751, Dec. 2023. [Online]. Available: <https://opg.optica.org/optica/abstract.cfm?URI=optica-10-12-1745>

- [16] D. Wang, Y. Nie, G. Hu, H. K. Tsang, and C. Huang, "Ultrafast silicon photonic reservoir computing engine delivering over 200 tops," *Nature Commun.*, vol. 15, 2024, Art. no. 10841, doi: [10.1038/s41467-024-55172-3](https://doi.org/10.1038/s41467-024-55172-3).
- [17] A. Katumba, X. Yin, J. Dambre, and P. Bienstman, "A neuromorphic silicon photonics nonlinear equalizer for optical communications with intensity modulation and direct detection," *J. Lightw. Technol.*, vol. 37, no. 10, pp. 2232–2239, May 2019.
- [18] F. Da Ros, S. M. Ranzini, H. Bülow, and D. Zibar, "Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 5, Sep./Oct. 2020, Art. no. 7701912.
- [19] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for non-linear fiber distortion compensation," *Opt. Exp.*, vol. 29, no. 20, pp. 30991–30997, Sep. 2021. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-29-20-30991>
- [20] S. Masaad, S. Sackesyn, S. Sygletos, and P. Bienstman, "Experimental demonstration of 4-port photonic reservoir computing for equalization of 4 and 16 QAM signals," *J. Lightw. Technol.*, vol. 42, no. 24, pp. 8555–8563, Dec. 2024.
- [21] R. V. Assche et al., "Real-time all-optical signal equalisation with silicon photonic recurrent neural networks," 2025, *arXiv:2503.19911*.
- [22] E. Staffoli, M. Mancinelli, P. Bettotti, and L. Pavesi, "Equalization of a 10 Gbps IMDD signal by a small silicon photonics time delayed neural network," *Photon. Res.*, vol. 11, no. 5, pp. 878–886, May 2023. [Online]. Available: <https://opg.optica.org/prj/abstract.cfm?URI=prj-11-5-878>
- [23] E. Staffoli, G. Maddinelli, and L. Pavesi, "A silicon photonic neural network for chromatic dispersion compensation in 20 Gbps PAM4 signal at 125 km and its scalability up to 100 GBPs," *J. Lightw. Technol.*, vol. 43, no. 2, pp. 557–571, Jan. 2025. [Online]. Available: <https://opg.optica.org/jlt/abstract.cfm?URI=jlt-43-2-557>
- [24] B. Wang et al., "An all-optical signal processor enabling terabit-per-second real-time equalization," *Science*, vol. 392, no. 6803, Jun. 2026, doi: [10.1126/science.ady5344](https://doi.org/10.1126/science.ady5344).
- [25] G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017, doi: [10.1515/nanoph-2016-0132](https://doi.org/10.1515/nanoph-2016-0132).
- [26] L. Appeltant, G. Van Der Sande, J. Danckaert, and I. Fischer, "Constructing optimized binary masks for reservoir computing with delay systems," *Sci. Rep. - Nature*, vol. 4, Jan. 2014, Art. no. 3629.
- [27] Y. Kuriki, J. Nakayama, K. Takano, and A. Uchida, "Impact of input mask signals on delay-based photonic reservoir computing with semiconductor lasers," *Opt. Exp.*, vol. 26, no. 5, pp. 5777–5788, Mar. 2018. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-26-5-5777>
- [28] F. Stelzer, A. Röhm, K. Lüdge, and S. Yanchuk, "Performance boost of time-delay reservoir computing by non-resonant clock cycle," *Neural Netw.*, vol. 124, pp. 158–169, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608020300125>
- [29] T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, "Role of delay-times in delay-based photonic reservoir computing, invited," *Opt. Mater. Exp.*, vol. 12, no. 3, pp. 1214–1231, Mar. 2022. [Online]. Available: <https://opg.optica.org/ome/abstract.cfm?URI=ome-12-3-1214>
- [30] L. Jaurigue, E. Robertson, J. Wolters, and K. Lüdge, "Reservoir computing with delayed input for fast and easy optimisation," *Entropy*, vol. 23, no. 12, 2021, Art. no. 1560. [Online]. Available: <https://www.mdpi.com/1099-4300/23/12/1560>
- [31] E. Picco, L. C. Jaurigue, K. Lüdge, and S. Massar, "Efficient optimisation of physical reservoir computers using only a delayed input," *Commun. Eng.*, vol. 4, 2025, Art. no. 3. [Online]. Available: <https://www.nature.com/articles/s44172-025-00340-6>
- [32] A. Skalli et al., "Model-free front-to-end training of a large high performance laser neural network," 2025, *arXiv:2503.16943*.
- [33] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "An enhanced architecture for silicon photonic reservoir computing," in *Proc. Cogn. Comput. Merging Concepts Hardware*, 2018, pp. 1–2.
- [34] VPIphotonics, "VPI design suite." Accessed: Apr. 3, 2026. [Online]. Available: <https://www.vpi Photonics.com/Tools/DesignSuite/>
- [35] F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework pytorch," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 5918.
- [36] A. Katumba, "Energy-efficient photonic neuromorphic computing for telecommunication applications," Ph.D. dissertation, Ghent Univ., Ghent, Belgium, 2019.
- [37] S. Masaad et al., "Baudrate-and reach-flexible all-optical equalization with a co-packaged photonic reservoir and receiver," 2026, *arXiv:2604.20504*.
- [38] J. Dambre et al., "Computing with integrated photonic reservoirs," in *Reservoir Comput. : Theory, Phys. Implementations, and Appl.* (Natural Computing Series), K. Nakajima and I. Fischer, Ed. Berlin, Germany: Springer, 2021, pp. 397–419, doi: [10.1007/978-981-13-1687-6_17](https://doi.org/10.1007/978-981-13-1687-6_17).
- [39] N. S. Bergano, F. W. Kerfoot, and C. R. Davidsson, "Margin measurements in optical amplifier system," *IEEE Photon. Technol. Lett.*, vol. 5, no. 3, pp. 304–306, Mar. 1993.