Deep-Learning-Enabled Direct Detection With Reduced Computational Complexity and High Electrical-Spectral-Efficiency

Xingfeng Li, Jingchi Li, Shaohua An, Hudi Liu, William Shieh, *Fellow, IEEE*, and Yikai Su, *Senior Member, IEEE*

***Abstract*—Complex-valued double-sideband (CV-DSB) direct detection (DD) is a promising solution for high capacity and cost-sensitive data center interconnects, since it can reconstruct the optical field as a homodyne coherent receiver while does not require a costly local oscillator laser. In a carrier-assisted CV-DSB DD system, the carrier occupies a large proportion of the total optical signal power but bears no information. Reducing the carrier to signal power ratio (CSPR) can improve the information-bearing signal power and thus maximize the achievable system performance. Recently, we have proposed and demonstrated a deep-learning-enabled DD (DLEDD) scheme to reconstruct the full-field of the CV-DSB signal. In the DLEDD scheme, the optical CV-DSB signal was detected by a dispersion-diversity receiver and then recovered by a deep convolutional neural network (CNN). Nevertheless, the computational complexity of the deep CNN is the main obstacle to the application of the DLEDD scheme. In this paper, we demonstrate a 50-GBaud CV-DSB 32-ary quadrature amplitude modulation (32-QAM) signal transmission over 80-km single-mode fiber with ∼64% computational-budget reduction in the field reconstruction. This is achieved by using 1×1 convolutions to attain a sparse dimensionality (in particular the number of channels) of the deep CNN. To the best of our knowledge, we achieve the highest electrical spectral efficiency of 7.07 b/s/Hz per polarization per wavelength for a CV-DSB DD receiver without requiring a sharp-roll-off optical filter.**

***Index Terms*—1×1 convolution, complex-valued double-sideband, deep-learning-enabled direct detection, sparse dimensionality.**

I. Introduction

DRIVEN by various emerging broadband services, such as augmented/virtual reality, cloud computing, and 4K/8K high-definition video, transmission capacity of data center interconnects continues to scale up for years [1]. Among the schemes for data center interconnects, coherent detection possesses excellent performances with a high electrical spectral efficiency (ESE), but it requires one local oscillator (LO) laser compared to direct detection (DD), which significantly increases the cost of the system [2]. In addition, usually the LO laser is based on epitaxial III-V group materials, which are not compatible with mature complementary metal-oxide-semiconductor (CMOS) manufacturing process.

Complex-valued double-sideband (CV-DSB) DD can reconstruct the optical field as a homodyne coherent receiver but does not require a costly LO laser, thus it is regarded as a promising solution for high capacity and cost-effective data center interconnects. If large-scale photonic integrated circuits are implemented, III-V materials are not introduced in the standard CMOS manufacturing process. On the other hand, the CV-DSB signal can be a traditional quadrature amplitude modulation (QAM) signal with a guard band. The digital signal processing (DSP) of the CV-DSB signal is also compatible with current single carrier DSP. In [3], a dual-single sideband (SSB) modulation technique was proposed to detect a CV-DSB signal using two sharp-slope optical bandpass filters (OBPFs), at the expense of high implementation cost and low tolerance to laser wavelength drift. In [4], a carrier-assisted differential detection (CADD) receiver based on a delay interferometer structure was proposed to reconstruct the optical field of a CV-DSB signal. In [5], a 27-GBaud quadrature phase shift keying (QPSK) transmission experiment over 160-km single-mode fiber (SMF) based on CADD was conducted for the first time. Afterwards, some extended versions of CADD such as interleaved CADD [6] and parallel dual delay-based CADD [7] were presented to improve the performance of the CADD receiver. Moreover, a high-ESE silicon photonic (SiP) CADD receiver with a ∼12-dB optimum carrier-to-signal power ratio (CSPR) was reported [8], [9]. Nevertheless, the CADD receiver requires a high CSPR with a complex structure. In [10], a carrier-less DD phase retrieval receiver enabling full-field reconstruction of the CV-DSB signal was demonstrated by leveraging a chromatic dispersion (CD) element and a modified Gerchberg-Saxton (GS) algorithm. Besides, an enhanced carrier-less phase retrieval receiver relied on space-time diversity was proposed to accelerate

Manuscript received 23 November 2022; revised 9 February 2023 and 26 March 2023; accepted 28 March 2023. Date of publication 31 March 2023; date of current version 4 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1803602 and in part by the National Natural Science Foundation of China under Grant 61860206001/61835008. *(Corresponding author: Yikai Su.)*

Xingfeng Li, Jingchi Li, Shaohua An, Hudi Liu, and Yikai Su are with the State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xingfengli@sjtu.edu.cn; jingchi_lee@sjtu.edu.cn; shaohuaan@sjtu.edu.cn; liuhudi@sjtu.edu.cn; yikaisu@sjtu.edu.cn).

William Shieh is with the School of Engineering, Westlake University, Hangzhou 310030, China (e-mail: shiehw@westlake.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2023.3263640.

Digital Object Identifier 10.1109/JLT.2023.3263640

the algorithm convergence and improve detection imperfections tolerance [11], [12]. However, the required electrical bandwidth of the receiver is twice as much as that of the homodyne coherent receiver since the signal-to-signal beat interference (SSBI) need to be fully detected, which limits the ESE of the carrier-less phase retrieval receiver. In [13], to recover a dual single-sideband (dual-SSB) four-level pulse amplitude modulation signal, an asymmetric self-coherent detection (ASCD) receiver was employed based on single CD element and an iterative SSBI-cancellation algorithm. However, the received signals suffer from severe frequency-selective power fading induced by the CD element. In [14], an ASCD receiver based on Mach-Zehnder interferometers was investigated and verified via numerical simulation. Moreover, an integrated SiP ASCD receiver using Mach-Zehnder interferometers was demonstrated to recover a CV-DSB signal through a 40-km SMF [15], where a high CSPR of 15 dB was required for the SiP ASCD receiver.

More recently we have proposed and demonstrated a deep-learning-enabled DD (DLEDD) receiver to reconstruct the full field of a CV-DSB signal in the presence of a strong SSBI under the 2-dB CSPR condition [16]. In the DLEDD scheme, the optical CV-DSB signal is detected by a dispersion-diversity receiver and then recovered by a deep convolutional neural network (CNN). The receiver diversity technique employing more CD elements can weaken the effect of the frequency-selective power fading. A deep CNN based on the residual learning technique is used to accurately recover the CV-DSB signal. We also experimentally demonstrated a 42-GBaud 16-QAM signal transmission over an 80-km SMF with a 5-dB optimum CSPR [17]. However, the deep CNN used in the DLEDD scheme requires $\sim 10^5$ real-valued multiplications to reconstruct the full-field of one symbol. The computational complexity greatly limits the application of the DLEDD scheme. To cope with the expensive computational budget in the DLEDD, we used 1×1 convolutions to realize a sparse dimensionality (in particular the number of channels) of the deep CNN [18]. With the 1×1 convolutions for all the six convolutional layers of the three residual blocks, we realized $\sim$64% computational-budget reduction in field reconstruction. Based on the simplified deep CNN, we demonstrated a 50-GBaud 16-QAM signal transmission over 80-km SMF with a net data rate of 159.2 Gb/s and a high ESE of 5.69-b/s/Hz. In this paper, we extend our work in [18] as follows: (i) we present more details on the operation principle and experimental-results analysis; (ii) we improve both the net data rate and the ESE by 25% and achieve a high capacity and a record ESE by using a 32-QAM modulation format. A review of transmission results on CV-DSB DD receivers without requiring sharp-roll-off optical filters is provided in Fig. 1. We successfully demonstrate a 50-GBaud 32-QAM signal transmission over 80-km SMF with an 8-dB optimum CSPR. Considering the frame redundancy and 25% forward error correction (FEC) overhead, we achieve a net data rate of 199 Gb/s and an ESE of 7.07 b/s/Hz per polarization per wavelength. To the best of our knowledge, this is the highest ESE per polarization per wavelength for CV-DSB DD receiver without requiring a sharp-roll-off optical filter.

The rest of this paper is organized as follows. In Section II, we elucidate the operation principle of the computation complexity reduction. Section III describes the experimental setup and DSP algorithms. Section IV presents the experimental results and discussion. Finally, we conclude this work in Section V.

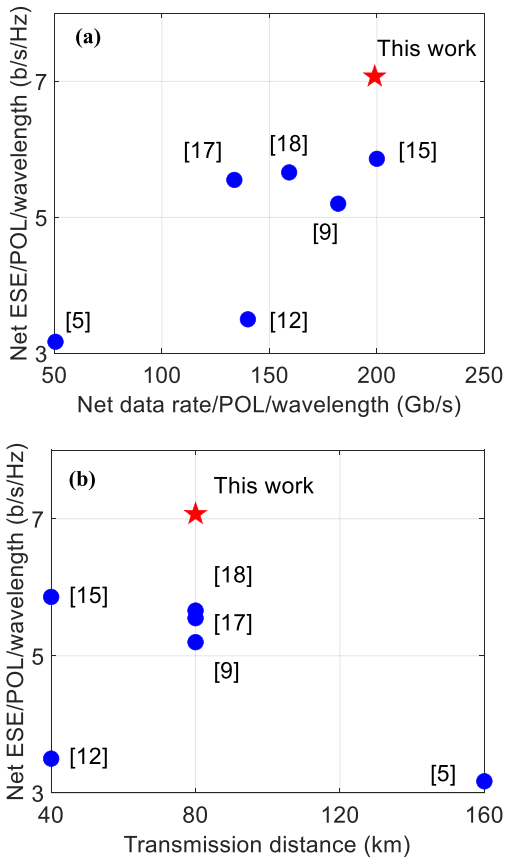

Fig. 1. Review of CV-DSB DD transmission experiments. (a) Net ESE/polarization (POL)/wavelength versus net data rate/POL/wavelength. (b) Net ESE/POL/wavelength versus transmission distance.

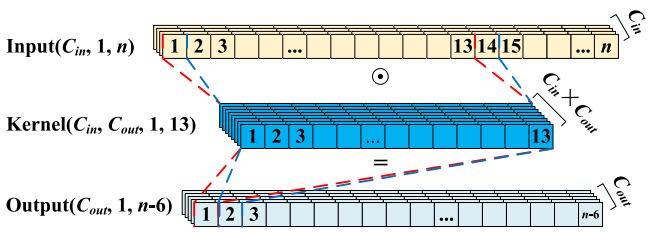

Fig. 2. Schematic of the convolutional operation.

II. Operation Principle

Fig. 2 illustrates the schematic of the convolutional operation. n is the sequence length. Thus, the shape of the sequence is $1\times n$. C_{in} and C_{out} represent the number of the input and output channels, respectively. Here, we take the kernel size of 13 as an example to illustrate the convolutional operation. Therefore, the shape of the kernel is 1×13. During the convolutional operation, a window slides along the sequence. Hence, $1\times(n\text{-}6)\times1\times13\times C_{in}\times C_{out}$ real-valued multiplications are required

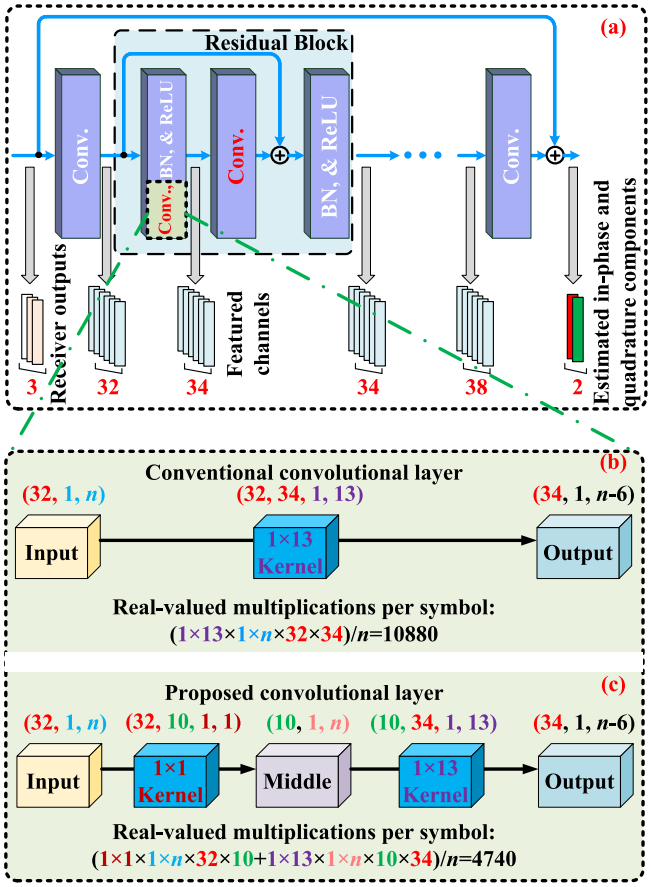

Fig. 3. (a) Schematic of the deep CNN structure in the DLEDD. Zoom-in of the (b) conventional and (c) proposed convolutional layer.

to realize a convolutional operation [19]. For simplicity, we use its approximate value, i.e., $1\times n\times 1\times 13\times C_{in}\times C_{out}$.

Fig. 3(a) depicts the schematic of the deep CNN structure in the DLEDD scheme [16]. The deep CNN depends on three stacked residual blocks. Each residual block contains two convolutional layers and a shortcut connection. The shortcut connection enables a fast-training process of the network. Two rectified linear unit (ReLU) functions are used as the activation functions to introduce nonlinear relationships. Before the ReLU activation, batch normalization (BN) is performed to improve the generalization ability of the CNN and overcome the problem of gradient disappearance in propagation. In addition, two convolutional layers are located outside of the residual blocks. Larger kernel size and more featured channel counts can improve the performance at the expense of increased computational complexity. As a trade-off between the complexity and performance, the corresponding kernel sizes of the eight convolutional layers and the shortcut connections are designed to be 13 and 1, respectively. We set 32, 34, 34, 36, 36, 38, 38, and 2 as the numbers of the featured channels after the eight convolutional layers, respectively. The optimization of the hyperparameters will be discussed in detail in Section IV.

Large kernel sizes and featured channel counts result in high computational complexity. The computational budget is expensive to implement a convolutional layer. Here we choose the first convolutional layer in the first residual block as an example to analyze the computational complexity. As shown in Fig. 3(b), 10880 real-valued multiplications per symbol are required if we use conventional method to realize the convolutional layer. In the deep CNN, convolutional operation occupies most of the computational budget. The deep CNN contains two categories of convolutional operations, i.e., eight convolutional layers with a kernel size of 13 and four shortcut connections with a kernel size of 1. The required real-valued multiplications is $1\times n\times 1\times 13\times(3\times 32+32\times 34+34\times 34+34\times 36+36\times 36+36\times 38+38\times 38+38\times 2)$ for eight convolutional layers with a kernel size of 13. The required real-valued multiplications is $1\times n\times 1\times 1\times(32\times 34+34\times 36+36\times 38+3\times 2)$ for four shortcut connections with a kernel size of 1. Thus, to realize the deep CNN, the required total real-valued multiplications are $104410\times n$, i.e., the sum of the real-valued multiplications required by two categories of convolutional operations. To reconstruct the field of one symbol, $\sim 1\times 10^5$ ($104410\times n/n\approx 1\times 10^5$) real-valued multiplications are required [16], [17].

Fig. 3(c) shows our proposed convolutional scheme. Before the convolutional layer with a large kernel size, we insert a 1×1 convolutional layer to reduce the required real-valued multiplications since 1×1 convolution has a powerful capability in dimensionality sparsification [22]. Here, 1×1 means that the kernel size of the convolutional layer is 1. Dimensionality sparsification means that the number of channels is reduced. For the 1×1 convolutional layer, the number of the output channel is set to 10, which will be discussed in detail in Section IV. As shown in Fig. 3(c), the required real-valued multiplications per symbol is reduced to 4740 after employing a 1×1 convolutional layer. If we apply the proposed convolutional method for all the six convolutional layers in the three residual blocks, the required real-valued multiplications per symbol are $(1\times 32\times 10+13\times 10\times 34+1\times 34\times 10+13\times 10\times 34)+(1\times 34\times 10+13\times 10\times 36+1\times 36\times 10+13\times 10\times 36)+(1\times 36\times 10+13\times 10\times 38+1\times 38\times 10+13\times 10\times 38)$ to realize the three residual blocks. Including additional four shortcut connections and two convolutional layers located outside of the residual blocks as introduced above, the deep CNN requires in total 36102 real-valued multiplications per symbol. Consequently, only $\sim 3.6\times 10^4$ real-valued multiplications are required to reconstruct the field of one symbol after the dimensionality sparsification. The computational complexity of the field reconstruction is reduced by $\sim 64\%$.

It should be noted that only convolutional layers with the number of both the input and output channels greater than 10 can use the 1×1 convolution layer to reduce the computational complexity. According to [20], [21], lower 10^8 real-valued multiplications per symbol are feasible for realistic signal processing if advanced field-programmable gate array (FPGA) is used. Since the CNN operates in a window-sliding manner, the consecutive signal is not required to be loaded at once. Therefore, with the increase of the sequence length, the required memory will not increase.

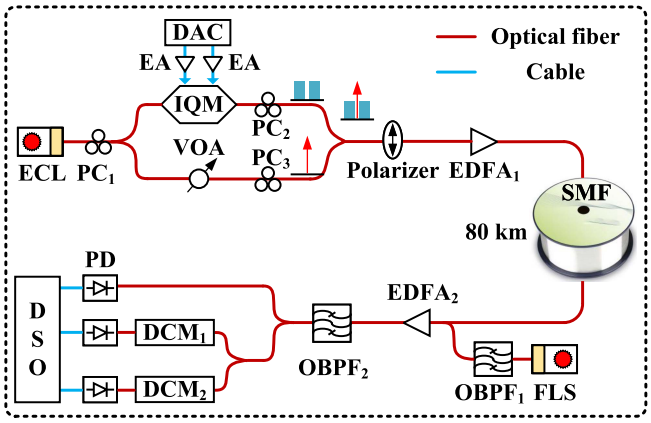

Fig. 4. Experimental setup.

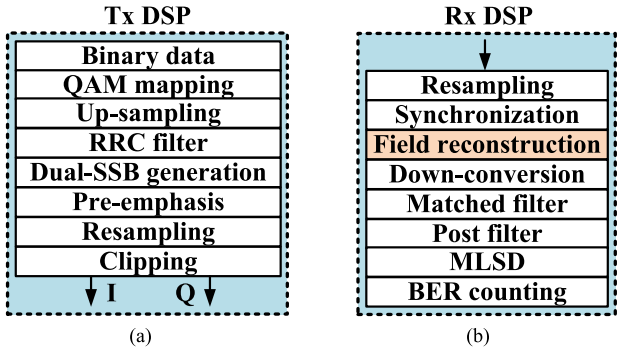

Fig. 5. DSP flow charts for (a) transmitter, and (b) receiver.

III. Experimental Setup and DSP Algorithm

To verify the feasibility of our proposed scheme, we conduct a proof-of-concept experiment in a 50-GBaud 32-QAM CV-DSB DD system.

A. Experimental Setup

We illustrate the experimental setup in Fig. 4. A 35-GHz digital-to-analog converter (DAC) (MICRAM DAC10002) operating at 100-GSa/s generates a 50-GBaud dual-SSB 32-QAM signal. After amplified by two electrical amplifiers (EAs) (SHF S807 C), the electrical 32-QAM signal drives a 35-GHz in-phase and quadrature modulator (IQM) (Fujitsu FTM7992HM) biased at its transmission null. Here, we employ the dual polarization modulator as a single polarization modulator by leveraging its high bandwidth. A continuous-wave light from an external cavity laser (ECL) with a ∼15-kHz linewidth at 1550 nm is split into two branches. One path is fed into the IQM to generate an optical dual-SSB 32-QAM signal, the other provides a carrier. Then, the dual-SSB 32-QAM signal and the carrier form a CV-DSB signal via a coupler. Here, we use two polarization controllers (PCs) and a polarizer to align the states of both the dual-SSB signal and the carrier. The CSPR of the CV-DSB signal is adjusted via a variable optical attenuator (VOA). Before launched into an 80-km SMF, the CV-DSB signal is boosted to 6 dBm by an erbium-doped fiber amplifier (EDFA).

At the receiver side, an amplified spontaneous emission (ASE) noise outputted from a fiber laser source (FLS) is injected into the CV-DSB signal. We adjust the optical signal-to-noise ratio (OSNR) of the CV-DSB signal via varying the output power of the FLS. Then, an EDFA is used to amplify the CV-DSB signal. We use two optical band-pass filters (OBPFs) to suppress the out-of-band noise. The dispersion diversity receiver is composed of three photodiodes (PDs) with a 3-dB bandwidth of 43 GHz and two dispersion compensation modules (DCMs). The dispersion values of the two DCMs are set to −;134.4, and −1008 ps/nm, respectively. After optical-to-electrical conversion, the detected electrical signals are captured by a 36-GHz digital storage oscilloscope (DSO) (LeCroy 36Zi-A) operating at 80 GSa/s.

B. DSP Algorithm

The DSP flow charts of the transceiver are shown in Fig. 5. In the transmitter-side DSP, two independent binary data streams generated by the Mersenne Twister algorithm are mapped into total 204800 32-QAM symbols for left/right sideband (LSB/RSB) [23]. Then, the 32-QAM signals are up-sampled by two times. We use two root raised cosine (RRC) filters with roll-off factors of 0.01 to generate the Nyquist-shaped LSB/RSB signals, respectively. After up-conversion for inserting guard band (−2.875 to 2.875 GHz), the LSB and RSB signals are combined to construct a dual-SSB 32-QAM signal. We use a pre-emphasis technique to compensate for the imperfect frequency response of the transmitter. Finally, the dual-SSB 32-QAM signal is resampled to 100 GSa/s and clipped for reducing its peak-to-average-power ratio (PAPR). The clipping ratio is 11 dB.

At the receiver-side DSP, the captured signals are resampled to 100 GSa/s. After frame synchronization, the three signals are processed by a deep CNN to reconstruct the field of the CV-DSB signal. In the CNN, 60% and 40% data samples are used for training and testing, respectively. 1×1 convolution layers are employed to reduce the computational complexity of convolutional layers in residual blocks. During the training process, we use an Adam optimizer to minimize the loss function of the neural network. The learning rate of the optimizer is 0.01 at the beginning and then decayed to 0.001 after 35 epochs. With 40 training epochs, the optimized network is used to recover the signal in the test set. After field reconstruction, the desired sideband of the CV-DSB signal is down-converted to zero frequency. And then, a matched RRC filter is sufficient for selecting the desired sideband as a low-pass filter. To suppress the enhanced noise, a 2-tap post filter and the maximum-likelihood sequence decision (MLSD) are implemented. We obtain a bit error ratio (BER) value by averaging the results from four consecutive BER measurements. In this case, the BER is calculated by error counting over $\sim1.6\times10^6$ bits, which are enough for testing BERs around 10^{-3}.

IV. Experimental Results and Discussion

In Fig. 6, we plot the optical spectra measured by an optical spectrum analyzer (OSA) (APEX AP2040C) at different

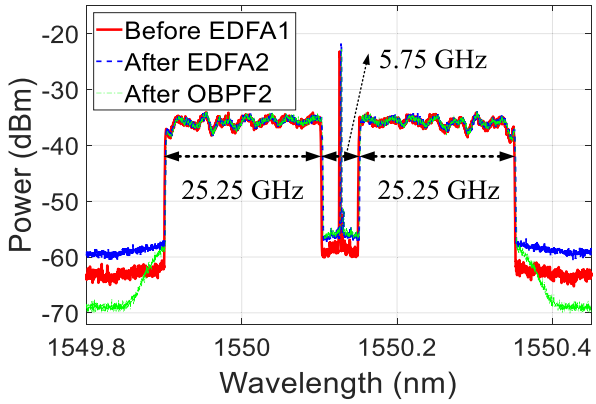

Fig. 6. Optical spectra measured at different locations.

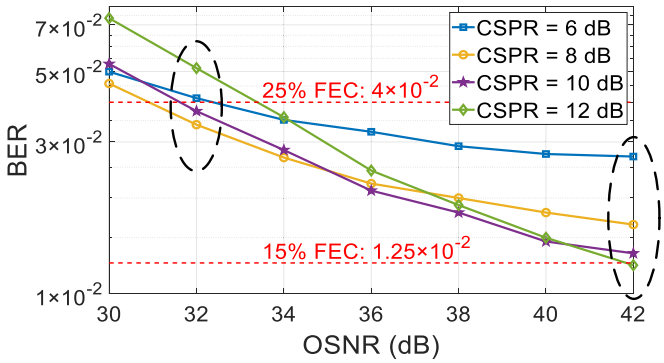

Fig. 8. BER versus OSNR for different CSPRs.

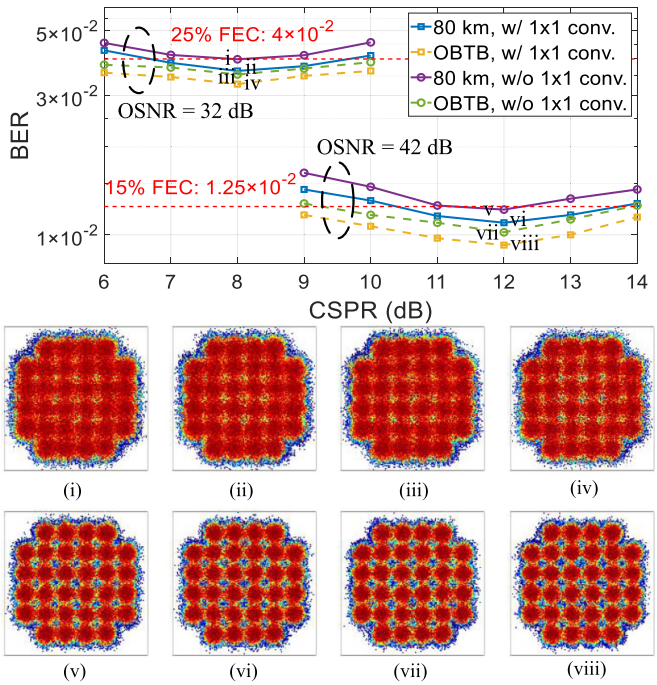

Fig. 7. BER versus CSPR for two different OSNRs in the OBTB case and after the 80-km transmission. Insets (i-iv) constellations at 32-dB OSNR and 8-dB CSPR. Insets (v-viii) constellations at 42-dB OSNR and 12-dB CSPR.

locations in the experimental setup, with a resolution of 1.12 pm. The high-frequency attenuation of the transmitter is compensated by using pre-emphasis technique. The "ripples" in the spectra are caused by the imperfection of components such as DAC. Since the signals are digitally shaped, the bandwidths of both LSB and RSB signals are 25.25 GHz. As shown, a 5.75-GHz guard band is inserted between upper and lower sidebands due to a strong SSBI enhancement induced by the zero-frequency null [16]. A guard band is widely used in the carrier assisted direct detection of the CV-DSB signal [4], [8], [15].

Fig. 7 shows the BER versus CSPR for two different OSNRs in the optical back-to-back (OBTB) case and the 80-km fiber transmission case. When the OSNR of the received signal is 32 dB, the optimum CSPR is ∼8 dB for both two cases. Compared with the conventional convolutional scheme without 1×1 convolutions, the proposed convolutional scheme performs better. This is owing to two factors: 1) the reduced redundancy information of the CNN via dimensionality sparsification; 2) the increased depth of the CNN [22], [24], [25]. Moreover, the CNN has a more powerful capability for nonlinear modeling. In the transmission case, the chromatic dispersion imposes an increased PAPR of the received signals, which results in a performance degradation due to a higher quantization noise of the DSO. When the CSPR is lower than the optimum value, the signals suffer from severe SSBI. With the increase of the CSPR, the SSBI distortion is alleviated, and thus the BER performance is improved. Nevertheless, a high CSPR sacrifices the effective OSNR, which leads to a worse BER performance. When the OSNR increases to 42 dB, the optimum CSPR is 12 dB, which is a trade-off between the SSBI and the noise. For a large OSNR, the BER performance is mainly affected by the SSBI instead of the noise. In this case, a higher CSPR is required to suppress the SSBI distortion. For both the OBTB and 80-km transmission cases with or without 1×1 convolutions in Fig. 7, the constellations of the recovered 32-QAM signals at different OSNR and CSPR are also presented in the insets. To pursue a low optimum CSPR, 32-dB OSNR is sufficient to reach the 25% FEC threshold of 4×10^{-2}. Since we artificially reduce the OSNR by injecting ASE noise, our transmission system has the potential to transmit a higher baud rate and a longer distance. In short, we successfully demonstrate a 250-Gb/s 32-QAM signal over 80-km SMF transmission in the DLEDD system. Considering the 0.5% frame redundancy of synchronization sequence and 25% FEC overhead, we achieve a high net data rate of 199 Gb/s (50 GBaud × 5 b/symbol × 99.5% / 1.25) and a record ESE of 7.07 b/s/Hz (199 Gb/s / 28.13 GHz) per polarization per wavelength in the CV-DSB DD system.

Fig. 8 shows the BER versus OSNR for four different CSPRs. At the 32-dB and 42-dB OSNR conditions, the corresponding optimum CSPR values are 8 dB and 12 dB, respectively, which agree well with the results in Fig. 7.

Fig. 9(a) presents the BER versus OSNR for different output channel counts of the employed 1×1 convolutional layers in residual blocks at the 8-dB CSPR. At a low OSNR value, these

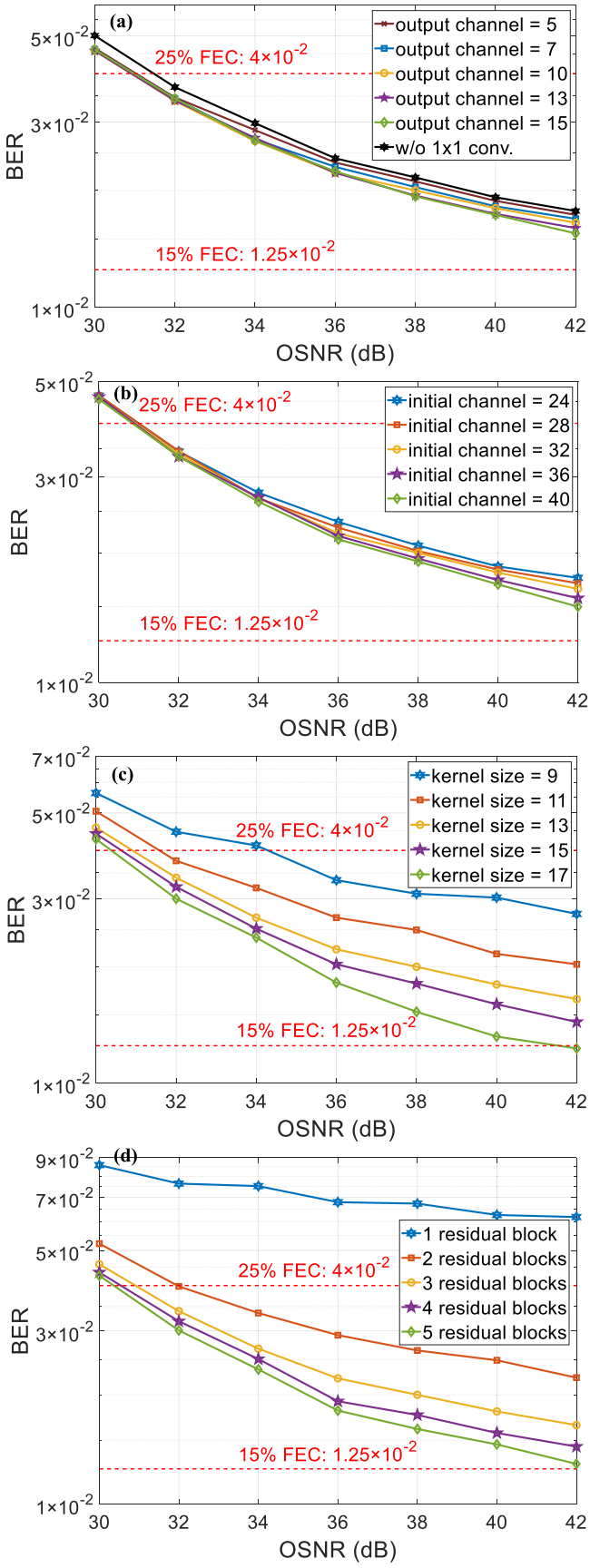

Fig. 9. BER versus OSNR for different (a) output channel counts, (b) initial channel counts, (c) kernel sizes, and (d) residual-block numbers at 8-dB CSPR.

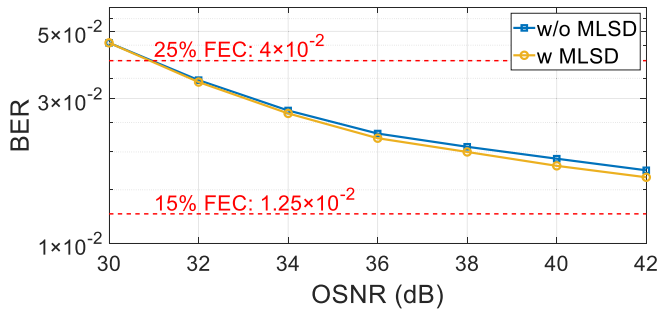

Fig. 10. BER versus OSNR with and without MLSD.

BER curves nearly overlap, since the BER is limited by the noise instead of the SSBI. The BER performance is insensitive to the CNN structure. With the increase of the OSNR, the BER curves diverge gradually. That is because that the BER is mainly limited by the SSBI at a large OSNR. Hence, the BER becomes more sensitive to the CNN structure. In addition, after using more output channels, we observe a slight improvement of the BER performance thanks to more powerful SSBI mitigation ability. Considering the trade-off between the performance and complexity, the experimental results presented in both Figs. 7 and 8 are based on ten output channels. As a result, only $\sim 3.6\times10^4$ real-valued multiplications are required after using the 1×1 convolution. The computational complexity of the field reconstruction is reduced by $\sim$64%. Furthermore, if we use five output channels, only $\sim 2.1\times10^4$ real-valued multiplications are required. The computational complexity of the field reconstruction can be potentially reduced by $\sim$79% with a slight performance degradation. The case without using the 1×1 convolution can be regarded as that the number of the output channel is zero, with the worst BER performance.

In Fig. 3(a), the output channel count of the first convolutional layer is 32, which can be regarded as the initial channel count. Fig. 9(b) shows the BER versus OSNR for different initial channel counts at the 8-dB CSPR. The BER changes with the OSNR similar to that in Fig. 9(a). In the deep CNN, only two kernel sizes are employed, i.e., 1 and 13. If we vary the kernel sizes to be different from 13, the BER versus OSNR for different kernel sizes at the 8-dB CSPR is shown in Fig. 9(c) and (d) presents the BER versus OSNR for the different number of employed residual blocks. Significant difference of the BER performance can be observed when we vary the kernel sizes and the number of employed residual blocks, which can be attributed to the fact that the network structure is changed more significantly. We select the hyperparameters in Fig. 3(a) to reduce the complexity while maintaining an acceptable performance. On the other hand, this work focuses on the complexity reduction via using 1×1 convolutions. We try to keep the hyperparameters to be consistent with our previous work and vary one parameter at a time, which helps us to make a fair comparison.

Fig. 10 shows the BER versus OSNR with and without MLSD. At a low OSNR value, the BER is limited by the noise instead of the system bandwidth. With the increase of the OSNR, the MLSD becomes more effective.

TABLE I
COMPARISONS OF CV-DSB DD SCHEMES

Detection schemes	Year	Baud rate (GBaud)	Modulation format	Net data rate (Gb/s)	Receiver bandwidth (GHz)	ESE per polarization (b/s/Hz)	OSE per polarization (b/s/Hz)	Transmission distance (km)	Optimum CSPR (dB)	Real-valued multiplications per symbol
GS [12]	2021	40	16-QAM	140	40	3.5	3.5	40	$-\infty$	$\sim 10^4$
CADD [5]	2020	27	QPSK	<50.5	15.93	<3.17	<1.59	160	7	$\sim 10^3$
SiP CADD [9]	2022	60	16-QAM	182	35.25	5.2	2.6	80	12.5	$\sim 10^3$
SiP ASCD [15]	2022	60	16-QAM	200	34	5.86	2.93	40	15	$\sim 10^3$
DLEDD [17]	2022	42	16-QAM	133.7	24.11	5.55	2.78	80	5	**$\sim 10^5$**
DLEDD [18]	2023	50	16-QAM	159.2	28.13	5.66	2.83	80	5	**$\sim 3.6\times 10^4$**
This work	2023	50	32-QAM	199	28.13	**7.07**	3.54	80	8	**$\sim 3.6\times 10^4$**

Table I provides the comparison of the state-of-the-art experimental result metrics about various CV-DSB DD schemes. Compared with our previous work [17], the proposed scheme can reduce the real-valued multiplications per symbol by ∼64% without performance degradation. Moreover, compared with our work in [18], we improve both the net data rate and the ESE by 25%. As evidenced by the table, both the ESE and optical spectral efficiency (OSE) per polarization per wavelength are the highest for CV-DSB DD receiver without requiring a sharp-roll-off optical filter.

DLEDD receiver is a laser-free scheme which does not require to introduce III-V materials in the standard CMOS manufacturing process. Therefore, by leveraging the silicon photonic integration, DLEDD can realize compact footprint and low system cost. Compared with the homodyne coherent receiver, the DLEDD receiver does not require a signal equalization and carrier recovery. More importantly, the DLEDD receiver does not require a LO laser with additional high power consumption including cooling. Therefore, the power consumption of the DLEDD may be comparable to that of the homodyne coherent receiver.

V. CONCLUSION

To cope with the problem of high computational complexity in the DLEDD system, we used 1×1 convolutions to realize a sparse dimensionality (in particular the number of channels) of the deep CNN. Compared with our previous work, the proposed scheme can reduce the computational complexity of the field reconstruction by ∼64% without performance degradation. Furthermore, the computational complexity of the field reconstruction can be potentially reduced by ∼79% if we use five output channels in the 1×1 convolutional layers. We demonstrated a 50-GBaud 32-QAM signal transmission over 80-km SMF with an 8-dB optimum CSPR. To the best of our knowledge, we achieve a high net data rate of 199 Gb/s and the highest ESE of 7.07 b/s/Hz per polarization per wavelength for CV-DSB DD receiver without requiring a sharp-roll-off optical filter. We believe that the proposed DLEDD with reduced computational complexity may find its applications in high capacity and cost-effective short reach transmission scenarios.

REFERENCES

[1] Cisco annual internet report (2018–2023). 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 157–179, Jan. 2016.

[3] L. Zhang, T. Zuo, Q. Zhang, E. Zhou, G. N. Liu, and X. Xu, "Transmission of 112-Gb/s DMT over 80-km SMF by twin-SSB technique," in *Proc. Eur. Conf. Opt. Commun.*, 2015, pp. 1–3.

[4] W. Shieh, C. Sun, and H. Ji, "Carrier-assisted differential detection," *Light: Sci. Appl.*, vol. 9, no. 1, Feb. 2020, Art. no. 18.

[5] C. Sun, T. Ji, H. Ji, Z. Xu, and W. Shieh, "Experimental demonstration of complex-valued DSB signal field recovery via direct detection," *IEEE Photon. Technol. Lett.*, vol. 32, no. 10, pp. 585–588, May 2020.

[6] T. Ji, C. Sun, H. Ji, Z. Xu, Y. Peng, and W. Shieh, "Theoretical and experimental investigations of interleaved carrier-assisted differential detection," *J. Lightw. Technol.*, vol. 39, no. 1, pp. 122–128, Jan. 2021.

[7] J. Li, S. An, H. Ji, X. Li, W. Shieh, and Y. Su, "Carrier-assisted differential detection with reduced guard band and high electrical spectral efficiency," *Opt. Exp.*, vol. 29, no. 21, pp. 33502–33511, Oct. 2021.

[8] J. Li et al., "High electrical spectral efficiency silicon photonic receiver with carrier-assisted differential detection," in *Proc. Opt. Fiber Commun. Conf.*, 2022, Paper Th4B-6, pp. 1–3.

[9] J. Li et al., "Silicon photonic carrier-assisted differential detection receiver with high electrical spectral efficiency for short-reach interconnects," *J. Lightw. Technol.*, vol. 41, no. 3, pp. 919–925, Feb. 2023.

[10] H. Chen, N. K. Fontaine, J. M. Gené, R. Ryf, D. T. Neilson, and G. Raybon, "Dual polarization full-field signal waveform reconstruction using intensity only measurements for coherent communications," *J. Lightw. Technol.*, vol. 38, no. 9, pp. 2587–2597, May 2020.

[11] H. Chen et al., "Space-time diversity phase retrieval receiver," in *Proc. Opt. Fiber Commun. Conf.*, 2021, pp. 1–3.

[12] H. Chen et al., "140G/70G direct detection PON with >37 dB power budget and 40-km reach enabled by colorless phase retrieval full field recovery," in *Proc. Eur. Conf. Opt. Commun.*, 2021, pp. 1–4.

[13] X. Li, M. O'Sullivan, Z. Xing, M. S. Alam, M. E. Mousa-Pasandi, and D. V. Plant, "Asymmetric self-coherent detection," *Opt. Exp.*, vol. 29, no. 16, pp. 25412–25427, Jul. 2021.

[14] X. Li, M. O'Sullivan, Z. Xing, M. E. Mousa-pasandi, and D. V. Plant, "Asymmetric self-coherent detection based on Mach-Zehnder interferometers," *J. Lightw. Technol.*, vol. 40, no. 7, pp. 2023–2032, Apr. 2022.

[15] Y. Hu et al., "Transmission of net 200 Gbps/λ over 40 km of SMF using an integrated SiP phase-diverse receiver," in *Proc. Eur. Conf. Opt. Commun.*, 2022, Paper Th3B-6, pp. 1–4.

[16] X. Li, S. An, H. Ji, J. Li, W. Shieh, and Y. Su, "Deep-learning-enabled high-performance full-field direct detection with dispersion diversity," *Opt. Exp.*, vol. 30, no. 7, pp. 11767–11788, Mar. 2022.

[17] X. Li et al., "Deep-learning enabled direct detection of 42-gbaud complex-valued DSB 16-qam signal," in *Proc. Conf. Lasers Electro-Opt.*, 2022, pp. 1–2.

[18] X. Li, J. Li, S. An, H. Liu, W. Shieh, and Y. Su, "Deep-learning-enabled high electrical-spectral-efficiency direct detection with reduced computation complexity," in *Proc. Opt. Fiber Commun. Conf.*, 2023, Paper M2F-2, pp. 1–3.

[19] C. Li et al., "Convolutional neural network-aided DP-64 QAM coherent optical communication systems," *J. Lightw. Technol.*, vol. 40, no. 9, pp. 2880–2889, May 2022.

[20] P. J. Freire et al., "Performance versus complexity study of neural network equalizers in coherent optical systems," *J. Lightw. Technol.*, vol. 39, no. 19, pp. 6085–6096, Oct. 2021.

[21] Z. Que, Y. Zhu, H. Fan, J. Meng, X. Niu, and W. Luk, "Mapping large LSTMs to FPGAs with weight reuse," *J. Signal Process. Syst.*, vol. 92, no. 9, pp. 965–979, Jul. 2020.

[22] C. Szegedy et al., "Going deeper with convolutions," in *Proc. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
[23] C. Chuang et al., "Study of training patterns for employing deep neural networks in optical communication systems," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.
[24] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*.
[25] S. Arora, A. Bhaskara, R. Ge, and T. Ma, "Provable bounds for learning some deep representations," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 32, no. 1, Jun. 2014, pp. 584–592.

Xingfeng Li received the M.S. degree from Shanghai University, Shanghai, China, in 2019. He is currently working toward the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China. His research interests include short-reach data center interconnects and direct-detection systems.

Jingchi Li received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2022. He is currently with Shanghai Jiao Tong University, Shanghai, China, as a Postdoc. His research interests include optical transmission, digital signal processing, and silicon photonics direct detection systems.

Shaohua An received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2021. His research interests include high-speed optical transmission and DSP techniques in direct-detection systems.

Hudi Liu received the B.S. degree from Nankai University, Tian Jin, China, in 2020. He is currently working toward the Ph.D. degree with Shanghai Jiao Tong University, Shanghai, China. His research interests include optical computing and photonic neural network.

William Shieh (Fellow) received the M.S. degree in electrical engineering and the Ph.D. degree in physics from the University of Southern California, Los Angles, Cam USA, in 1994 and 1996, respectively. He had worked in various research institutions, such as Jet Propulsion Laboratories, Pasadena, California, and Bell Laboratories, Holmdel, New Jersey. From 2004 to 2022, he was with the Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne, VIC, Australia. In 2022, he joined Westlake University, Hangzhou, China, as the Chair Professor of optical communication and sensing. He has authored or coauthored more than 300 journal and conference papers and submitted 14 U.S. patents (nine issued). His research interests include OFDM techniques in both wireless and optical communications, few-mode fibers for optical communications and sensing, coherent optical communication systems, and optical packet switching. He was the recipient of Australian Future Fellowship, during 2011–2014. He is a Fellow of Optica.

Yikai Su (Senior Member) received the Ph.D. degree in EE from Northwestern University, Evanston, IL, USA, in 2001. He was with the Crawford Hill Laboratory of Bell Laboratories, and joined the Shanghai Jiao Tong University, Shanghai, China, as a Full Professor, in 2004. He has more than 500 publications. He gave more than 70 invited talks at conferences, including OFC, CLEO and IPS. He holds seven U.S. patents and ∼70 Chinese patents. His research interests include silicon photonic devices, fiber transmission, and optical switching. Prof. Su is an advisory board member of Advanced Optical Materials and ACS Photonics. He is the Chair of IEEE Photonics Society Shanghai chapter, and was the General Co-Chair of ACP 2012/2021 and OECC 2023. He was also a TPC Member of a large number of international conferences, including CLEO (2016-2018), ECOC (2013-2017), and OFC (2011-2013). Prof. Su is a Fellow of Optica.