Supplementary Information for Chip-to-Chip High-Dimensional Teleportation via A Quantum Autoencoder

H. Zhang,[1,*] L. Wan,[1,*] T. Haug,[2] W. K. Mok,[3] H. Cai,[4] Y. Z. Shi,[5] L. K. Chin,[1] F. K. Muhammad,[1] L. M. Xiao,[6] G. Q. Lo,[4] X. S. Luo,[7] F. Gao,[7] B. Dong,[7] S. Assad,[8] M. S. Kim,[2] L. C. Kwek,[1,3,9,†] and A. Q. Liu[1,†]

[1]*Quantum Science and Engineering Centre (QSec), Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore*

[2]*Quantum Optics and Laser Science, Imperial College London, Exhibition Road, London SW72AZ, UK*

[3]*Centre for Quantum Technologies, National University of Singapore, Block S15, 3 Science Drive 2, 117583, Singapore*

[4]*Institute of Microelectronics, A*STAR (Agency for Science, Technology and Research), 138634, Singapore*

[5]*National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiao Tong University, Shanghai 200240, China*

[6]*School of Information Science and Technology, Fudan University, Shanghai 200433, China*

[7]*Advanced Micro Foundry, 11 Science Park Road, 117685 Singapore*

[8]*Department of Quantum Science, Centre for Quantum Computation and Communication Technology, The Australian National University, Canberra, ACT 2600, Australia*

[9]*National Institute of Education, 1 Nanyang Walk, 637616 Singapore*

[*] Equal Contribution

[†] Corresponding Author: cqtklc@gmail.com, eaqliu@ntu.edu.sg

SUPPLEMENTARY NOTE 1: QUANTUM AUTOENCODER

The autoencoder is a tool for dimensional reduction as in classical data processing [1, 2]. The parameters of an autoencoder are optimized across a training set of data, to erase some bits of the original inputs but will not abandon useful information, thus affecting their reproduction. In analogy with classical autoencoder, the quantum autoencoder [3–8] is proposed to represent quantum states in a lower-dimensional space, through optimization. Similarly, the quantum autoencoder can be represented by a graphical representation of a collection of interconnected nodes, as shown in Fig. S1a. The edges connecting adjacent layers represent the weights (a unitary transformation) from one layer to the next. Specifically, the encoder shrink the space between two layers and the decoder expand the space to reconstruct the original input states. The intermediate nodes are called "bottleneck" nodes, and those won't be utilized are called "trash" nodes.

An autoencoder for compressing classical information can be symbolized as to find two transitions $\mathcal{E}$ and $\mathcal{D}$ for given inputs X, such that

$$\mathcal{E}, \mathcal{D} = \arg\min_{\mathcal{E},\mathcal{D}} \|X - (\mathcal{D} \circ \mathcal{E})X\| \tag{1}$$

Similarly, a quantum autoencoder can be formalized as to find unitary transitions that preserve quantum information with smaller "bottleneck" space. The process is quantified as evaluated by the fidelity $F\left(|\psi_i\rangle, \rho_i^{out}\right) = \langle\psi_i \left|\rho_i^{out}\right| \psi_i\rangle$ between input and output quantum states, where $|\psi_i\rangle$ is the input state and ρ_i^{out} is the density matrix of the output state. The training objective is to achieve $F\left(|\psi_i\rangle, \rho_i^{out}\right) = 1$ for all input states. Notably in quantum autoencoder, the decoder can simply be the transpose-conjugate of the encoder U, to recover the input state $|\psi_i\rangle$, because

$$U^\dagger(U|\psi_i\rangle) = (U^\dagger U)|\psi_i\rangle = |\psi_i\rangle \tag{2}$$

There are two training methods, the full-way training and the half-way training. In full-way training, the whole process of the encoder-decoder and the output state ρ_i^{out} is expressed as

$$\rho_i^{out} = (U^\dagger) Tr_A \left[U|\psi_i\rangle\langle\psi_i|U^\dagger\right] (U^\dagger)^\dagger \tag{3}$$

Where A represents the bottleneck nodes, and B represents the trash nodes. The cost function of training the quantum autoencoder can be expressed by

$$C_1 = \sum_i F\left(|\psi_i\rangle, \rho_i^{out}\right) \tag{4}$$

Alternatively, a half-way training can be adopted, which is to watch over the trash state by

$$C_2 = \sum_i F(|r\rangle, Tr_B(U|\psi\rangle\langle\psi|U^\dagger)) \tag{5}$$

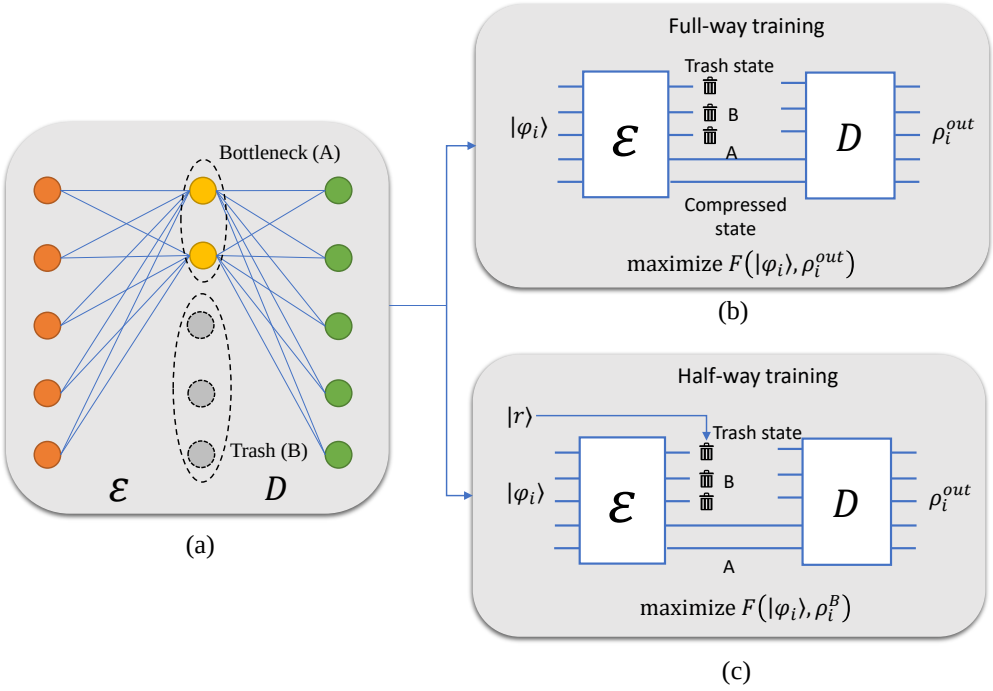


FIG. S1: Quantum autoencoder and its training strategies. (a) A graphical representation of a quantum autoencoder. The map $\mathcal{E}$ encodes a 5-bit input into a 2-bit "bottleneck" state, after which the decoder $\mathcal{D}$ reconstruct the input state to the output. The 3-bit "trash" state is also visualized, which in fact will not enter the decoder and is discarded. (b) Full-way training. (c) Half-way training.

where $|r\rangle$ is the reference state, and also serves as the input to the corresponding ports of decoder. The full-way and half-way training are shown in Figs. S1b and S1c, respectively. A quantum autoencoder that enables data compression, thereby reducing the requirements for quantum memory, quantum communication channels and the amount of quantum gates, has great potential in quantum simulation and communication in quantum networks. The original scheme performs compression on a multi-qubit state, while in our case, we apply it to compress a high-dimensional quantum state into a qubit.

With the encoder trained, we send over random qutrits for compression. Two separate encoders are trained for two separate subspaces, and for each encoder, two random qutrits are used for validation, showing that the encoder successfully compressed the qutrit into a qubit with the trash path unoccupied. The tomography results of these qutrits represented by three paths are depicted for the group 1 ($|\varphi_1\rangle$ and $|\varphi_2\rangle$), the group 2 ($|\varphi_3\rangle$ and $|\varphi_4\rangle$, as shown in Fig. S2. The reconstructed density matrix of the starting state is shown in brown, and the density matrices of the compressed qubits are colored in red (group 1) and green (group 2). As expected, the photon occupancy of the third port is nearly zero, which means that the qutrit can be losslessly compressed to a qubit and then recovered by the decoding process. As observed from the tomography results of the initial qutrits and the compressed states, the device achieves a lossless and

high-quality compression, reducing the photon occupation of the junk mode to 0.023 ± 0.011. When we restore the qutrit by the decoder (i.e., the inverse of the trained encoder), the fidelity $F = \langle\varphi|\rho|\varphi\rangle$ between the original qutrit $|\varphi\rangle$ and the reconstructed density matrix ρ is reported to be 0.971 ± 0.013.

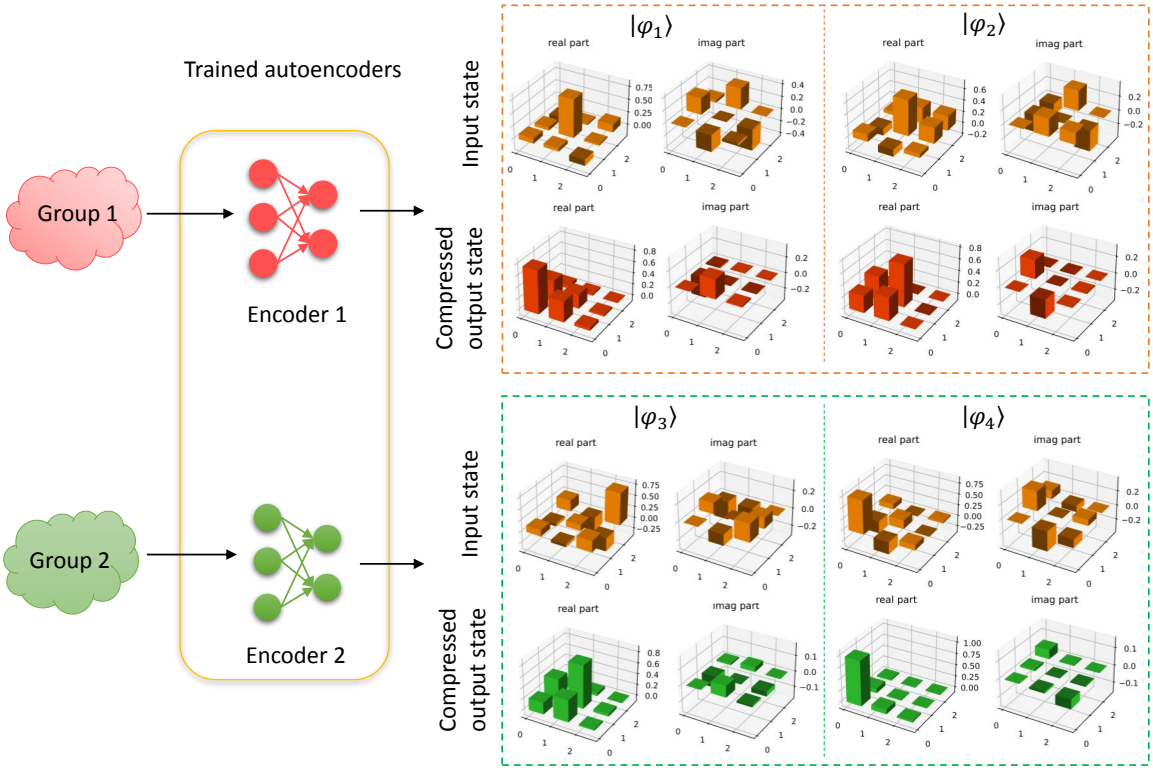


FIG. S2: Validation of trained encoders by state tomography. We trained two encoders for two respective subspaces of qutrits, and each subspace can be compressed by a common encoder. For each encoder, two exemplary qutrits are validated on it, showing their tomography results. The reconstructed density matrix of the starting state is depicted in brown, and the density matrices of the compressed qubits are in red for group 1 and are in blue for group 2.

SUPPLEMENTARY NOTE 2: PHOTON GENERATION IN SPIRALS

Non-degenerated four wave mixing process are adopted in our photon pair generation, where the pump light at a single wavelength is injected to generate the signal and idler photons with different wavelengths, in line with the phase matching and energy conservation condition of $2\beta(\omega_p) = \beta(\omega_s) - \beta(\omega_i)$ and $w_s + w_i = 2w_p$, where β is called the propagation constant and ω is the frequency of photon. For each pair of generated photons, two AMZIs are adopted for filtering out the pump light and separating the signal and idler photons. The transmission spectra of respective AMZI under different voltages applied on the

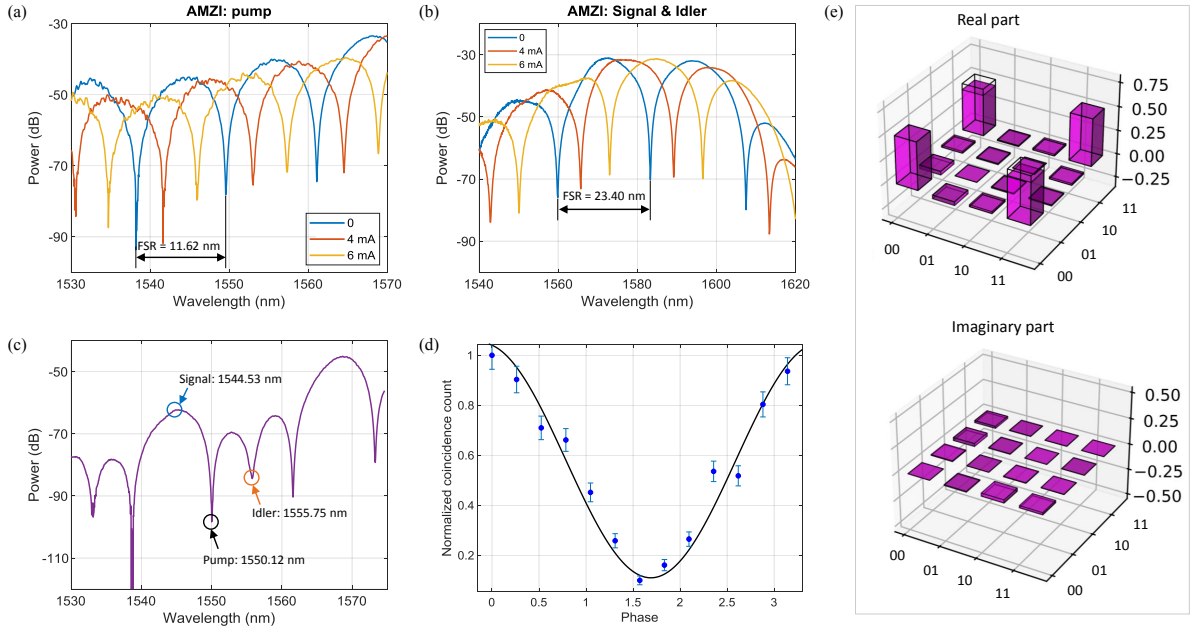
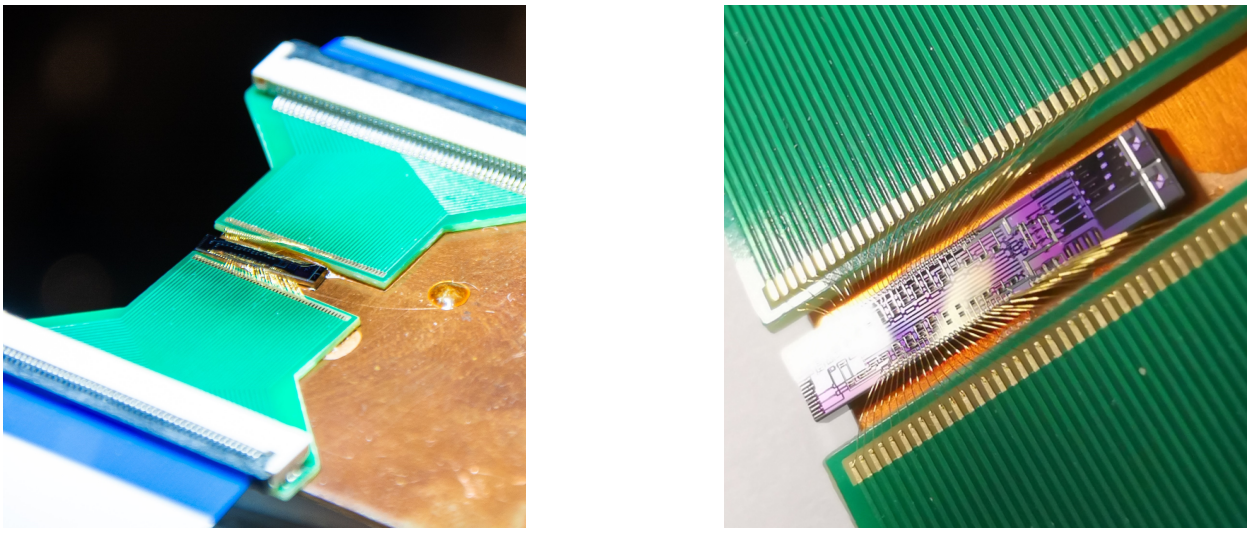


FIG. S3: Photon pair and entanglement generation in spiral structures. The measured transmission spectra for asymmetric Mach-Zehnder interferometers (AMZIs) with free spectral ranges (a) $\mathrm{FSR}_{\mathrm{pump}} = 11.62$ nm and (b) $\mathrm{FSR}_{signal} = 23.40$ nm. The characterization of AMZI with increasing currents applied on the heater is also shown. (c) AMZIs can filter the pump light and demultiplex the signal and idler photons. Signal photons are created at $\lambda_s = 1544.53$ nm and idler photons at $\lambda_i = 1555.75$ nm, with pump light at $\lambda_p = 1550.12$ nm. In the shown transmission spectra, the signal photons are used, and idler photons are filtered. Residual pump photons are removed by off-chip WDM filters with ~ 0.8 nm bandwidth and >100 dB extinction ratio. (d) Measured visibility of the heralded quantum interference. The visibility is $V = 0.894 \pm 0.083$. (e) The measured density matrix of the generated maximally entangled two-qubit states. The fidelity is reported to be 0.960 ± 0.004.

heater are shown in Figs. S3a and S3b. The free spectral ranges for the pump filter (FSRpump) and the signal/idler filter (FSRsignal) are 11.62 nm and 23.40 nm, respectively. The AMZIs can filter out the pump light and demultiplex the signal and idler photons into different waveguides, as shown in Fig. S3c. The pump, signal and idler photons are located at $\lambda_p = 1550.12$ nm, $\lambda_s = 1544.53$ nm and $\lambda_i = 1555.75$ nm, respectively. The average extinction ratio for AMZI is over 25 dB. The photon indistinguishability is estimated by the visibility of heralded two-photon quantum interference. We interfere two signal photons that are emitted from two independent spirals on an MZI (with two idler photons as heralding). Figure S3d presents the on-chip Hong-Ou-Mandel interference, with the fringe visibility of $V = 0.894 \pm 0.083$. The generated photon pair is then programmed to create a bell state $|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ as the resource for further demonstration. We implement quantum state tomography (QST) to reconstruct the density matrix



(a) (b)

FIG. S4: The packaging of the silicon photonic chip wire-bonded to a PCB. (a) The overview of the packaging. (b) Zoom in of the silicon photonic chip.

$\rho = |\phi^+\rangle \langle\phi^+|$. Figure S3e shows the measured ρ for the generated Bell state with fidelities of 0.960 ± 0.004.

SUPPLEMENTARY NOTE 3: EXPERIMENTAL SETUP

The packaging of the chip wirebonded to a PCB is shown in Fig S4. The transmitter chip and receiver chip are integrated on a microscale silicon photonic chip to form a trans-ceiver for two-way communication. Both parties holding the trans-ceiver can flexibly send or receive each other's information (as shown in Fig S5).

Figure S6 shows the detailed schematic of our experimental setup. Single photons are generated by pumping light into the long spiral-shaped waveguides, according to the principle of spontaneous four-wave mixing (SFWM). During this process, correlated pairs of photons are generated while pump photons are annihilated. A total of 6 asymmetric Mach-Zehnder interferometers (AMZIs) are employed as on-chip filters, 3 of which are used to filter the pump and another 3 are used to separate the signal and idler photons. Each AMZI consists of two multimode interference devices which function as $50:50$ beam splitters, and a pair of unbalanced waveguide arms. The optical path difference (OPD) of the two arms is designed to match the wavelength of pump laser and generated photons, so as to achieve the desired filtering efficiency. A thermo-optic tunable phase shifter is placed on one arm to align the AMZI's spectrum with the target wavelength. The chip is wire-bonded to a PCB, which provides independent control of each phase shifter, by an electronic current driver with 1kHz frequency and 12 bits resolution (QontrolTM).

Laser pulses are generated by the Ultrafast Optical Clock device (PriTel) with repetition rate of 500 MHz,

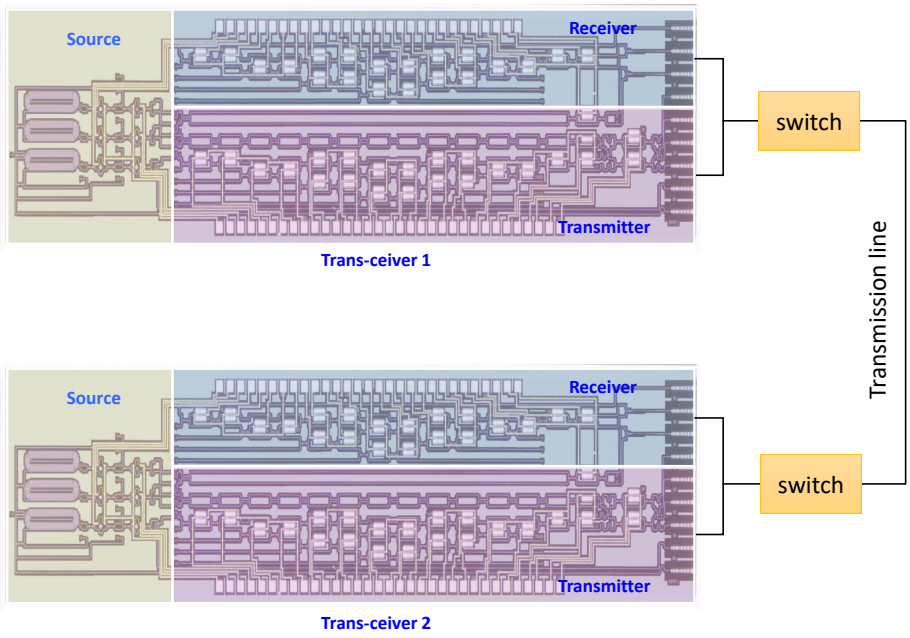


FIG. S5: The trans-ceiver that integrates both transmitter and receiver for two-way communication.

central wavelength of 1550.12 nm and bandwidth of 2 nm. The laser pulses are amplified by an Erbium-doped fibre amplifier (EDFA) and filtered off-chip via a 100G wavelength-division multiplexer (WDM) and a 2 nm bandpass filter (BPF) centered at 1550 nm to reshape the laser pulse into a FWHM of 0.5 nm and suppress the excessive noise of input laser by 100 dB. The input single pump light is coupled to the chip by a one-dimensional (1D) subwavelength grating coupler (SGC), so do the inline monitors on the chip. A polarization controller is utilized to maximize the coupling efficiency of the fiber to chip. The input power is monitored via a photodetector placed at the reflection port of the BPF. Photons are emitted via non-degenerated SFWM at C41: 1544.53 nm (signal wavelength) and C27: 1555.75 nm (idler wavelength).

Output photons are coupled out of the chip via an 8-mode edge coupling fiber array, and then filtered via a BPF and a 100G WDM. The purpose of the filtering operation is to remove the residual pump photons and enhance the photon indistinguishability. The optical signals are detected by 6 channel (D1-D6 as marked in Fig. **??**) of a 16-channel SNSPDs (PhotonSpotTM, 100 Hz dark counts, 85% efficiency). polarization controllers are placed at each of the 6 channels, before photons entering the SNSPDs. A timetagger from Swabian InstrumentTM GmbH is used as the counting logic that counts the single photon events and process them on fly. The counting device supports more than 40 million events per second. Specifically, B1 an B2 are connected by a fibre (representing transmission line), and a polarization rotator and splitter (PRS) is placed before they travelling out of the chip. The PRC is utilized to combine and convert the TE mode in two waveguides into the TE and TM modes in one waveguide, to couple them to to horizontal and vertical polarization in a single optical fibre, so that the phase difference of the teleported qubit over the long distance

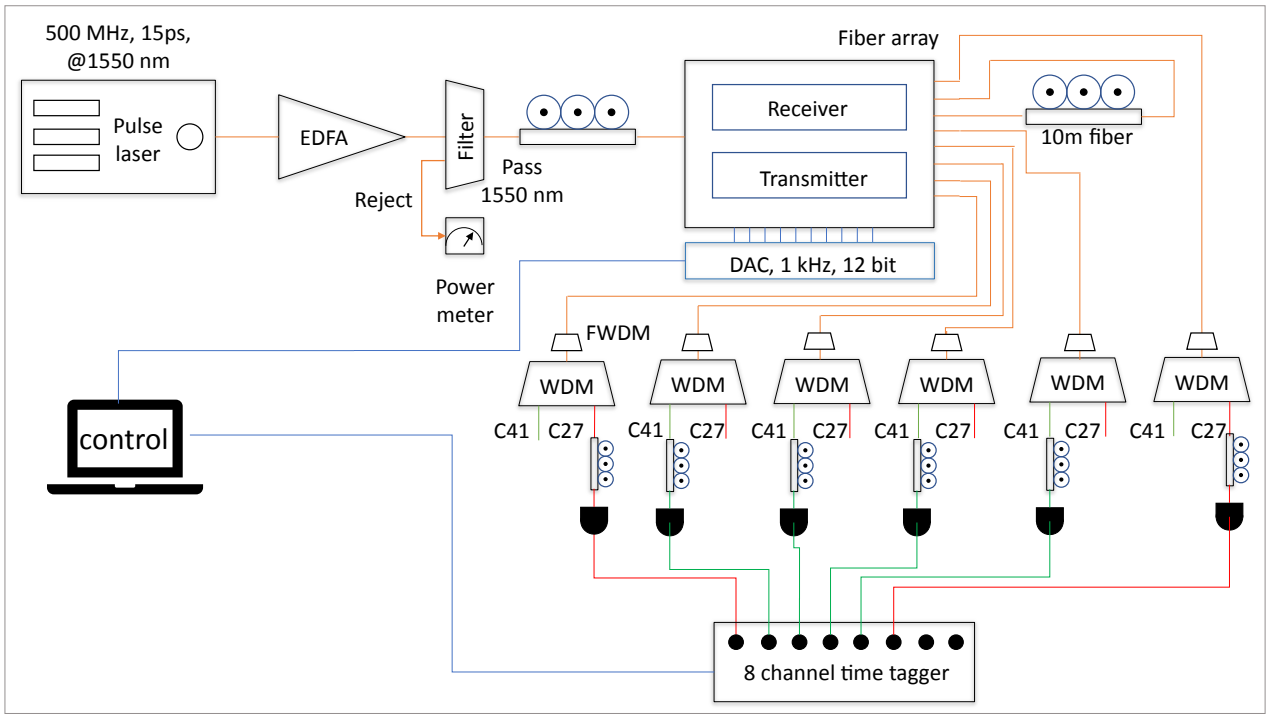


FIG. S6: The experimental setup that implements high-dimensional quantum teleportation.

transmitting can be maintained. Once the qubit is transmitted to the receiver part, the decoder will recover it to be the original qutrit.

SUPPLEMENTARY NOTE 4: BELL PROJECTOR AND BELL STATE MEASUREMENT

A. Single-Qubit Operation and Teleported States

We can generate single-qubit states or implement single-qubit projective measurement arbitrarily on silicon photonic chips, by decomposing the underlying single-qubit rotations into the programmable MZI devices. The probability of observing a dual-rail photon in the $|0\rangle$ or $|1\rangle$ mode is controlled by the $\hat{U}_{\mathrm{MZI}}(\theta)$. The $\hat{U}_{\mathrm{MZI}}(\theta)$ can further be decomposed by a $\hat{U}_{\mathrm{MMI}}(\theta)$ - which realizes $50:50$ beam splitting with a $\frac{\pi}{2}$ phase difference, and $\hat{U}_{\mathrm{PS}}(\phi)$ - which controls the relative phase between interacted waveguides. The total unitary $\hat{U}$ of a single-qubit rotation is $\hat{U} = \hat{U}_{\mathrm{PS}}(\phi)\hat{U}_{\mathrm{MZI}}(\theta)$. We display the linear optical transformations

of each components in the MZI as

$$\hat{U}_{\mathrm{PS}}(\theta) = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix} \tag{6a}$$

$$\hat{U}_{\mathrm{MMI}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \tag{6b}$$

$$\hat{U}_{\mathrm{MZI}}(\theta) = \hat{U}_{\mathrm{MMI}}\hat{U}_{\mathrm{PS}}(\theta)\hat{U}_{\mathrm{MMI}} = ie^{i\frac{\theta}{2}} \begin{bmatrix} \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \end{bmatrix} \tag{6c}$$

and derive the total unitary of $\hat{U}$ as

$$\hat{U} = \hat{U}_{\mathrm{PS}}(\phi)\hat{U}_{\mathrm{MZI}}(\theta) = ie^{i\frac{\theta}{2}} \begin{bmatrix} e^{i\phi}\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \\ e^{i\phi}\cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \end{bmatrix} \tag{7}$$

The teleported input states are randomly generated qutrits which are encoded on the three path of a single photon, using the linear optic quantum circuit. The structure consists of three MZIs, and their transfer matrices are given by

$$\hat{U}_{\mathrm{M1}}(\theta_1, \phi_1) = \begin{bmatrix} e^{i(\frac{\theta_1}{2}+\phi_1)}\sin\left(\frac{\theta_1}{2}\right) & e^{i(\frac{\theta_1}{2}+\phi_1)}\cos\left(\frac{\theta_1}{2}\right) & 0 \\ ie^{i\frac{\theta_1}{2}}\cos\left(\frac{\theta_1}{2}\right) & -ie^{i\frac{\theta_1}{2}}\sin\left(\frac{\theta_1}{2}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8a}$$

$$\hat{U}_{\mathrm{M2}}(\theta_2, \phi_2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & ie^{i(\frac{\theta_2}{2}+\phi_2)}\sin\left(\frac{\theta_2}{2}\right) & ie^{i(\frac{\theta_2}{2}+\phi_2)}\cos\left(\frac{\theta_2}{2}\right) \\ 0 & ie^{i\frac{\theta_2}{2}}\cos\left(\frac{\theta_2}{2}\right) & -ie^{i\frac{\theta_2}{2}}\sin\left(\frac{\theta_2}{2}\right) \end{bmatrix} \tag{8b}$$

$$\hat{U}_{\mathrm{M3}}(\phi_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{i\phi_2} \end{bmatrix} \tag{8c}$$

Thereby the overall unitary operator is derived by

$$\hat{U} = \hat{U}_{\mathrm{M3}}(\phi_3)\hat{U}_{\mathrm{M2}}(\theta_2, \phi_2)\hat{U}_{\mathrm{M1}}(\theta_1, \phi_1) \tag{9}$$

A total of 5 free parameters can in general produce arbitrary qutrit state. Experimentally, to produce random qutrit states that can be projected into a common two-dimensional space through a common rotation matrix, we conduct the procedure reversely, i.e., we first generate random qubits analytically and apply a SU(3) to the state formed by the qubit state and a vacuum state, the resulting qutrits are from the same subspace and can be compressed by a common encoder. The arbitrary rotation SU(3) is also implemented on chip, using 8 free parameters.

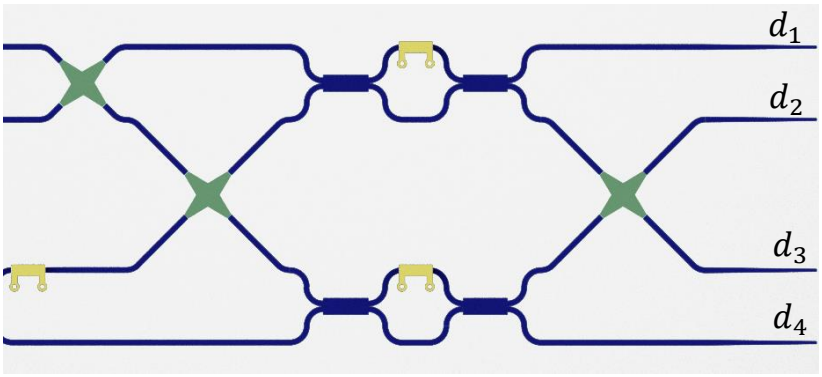


FIG. S7: The diagrams for the Bell projector designed for dual-rail qubits. The $|\phi^+\rangle$ is distinguished when joint clicks are observed at {d1, d2} or {d3, d4}, and the $|\phi^-\rangle$ are distinguished when joint clicks are observed in {d2, d3} or {d1, d4}.

TABLE S1: Coincidence event, bell state measurement result and corresponding teleported state.

	coincidence event	BSM result	teleported state
$\|\phi^+\rangle$	$d_1 d_2$ or $d_3 d_4$	$\|HH\rangle_{12} + \|VV\rangle_{12}$	$\alpha\|H\rangle_3 + \beta\|V\rangle_3$
$\|\phi^-\rangle$	$d_2 d_3$ or $d_1 d_4$	$\|HH\rangle_{12} - \|VV\rangle_{12}$	$\alpha\|H\rangle_3 - \beta\|V\rangle_3$
$\|\varphi^+\rangle$	-	$\|HV\rangle_{12} + \|VH\rangle_{12}$	$\alpha\|V\rangle_3 + \beta\|H\rangle_3$
$\|\varphi^-\rangle$	-	$\|HV\rangle_{12} - \|VH\rangle_{12}$	$\alpha\|V\rangle_3 - \beta\|H\rangle_3$

B. Bell State Measurement

It is not realistic for linear optics to realize a deterministic Bell state analysis. Three different linear-optical Bell operators, the KLM CNOT gate, the Bell projector and fusion projector can be used to identify up to two of the Bell states.The three operators are all based on post selection with different success probabilities to discriminate the Bell states. The success probability of Bell projector and fusion projector are 1/2, while KLM CNOT gate has the success probability of 1/9. In our device, we demonstrate the Bell projector $\hat{O}_{Bell}$ that has relatively higher success probability. We first consider the evolution and determination of the Bell states $|\phi^\pm\rangle$ expressed as

$$|\phi^\pm\rangle = \frac{1}{\sqrt{2}}\left(\hat{a}^\dagger_{\uparrow,1}\hat{a}^\dagger_{\uparrow,2} \pm \hat{a}^\dagger_{\downarrow,1}\hat{a}^\dagger_{\downarrow,2}\right)|\text{vac}\rangle \tag{10}$$

The first exchange operator to qubit 1 results in the transformation

$$\hat{a}^\dagger_{\downarrow,1} \leftrightarrow \hat{a}^\dagger_{\uparrow,1} \tag{11}$$

Then we swap the adjacent waveguide modes of the qubit 1 and 2 as

$$\hat{a}^\dagger_{\downarrow,1} \leftrightarrow \hat{a}^\dagger_{\uparrow,2} \tag{12}$$

and transforms the original state into

$$\frac{1}{\sqrt{2}}\left(\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\downarrow,1}\pm\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}\right)|\,\mathrm{vac}\rangle \tag{13}$$

Next a Hadamard-like operator is used to interfere each qubit, which is described by

$$\frac{1}{2\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,2}+\hat{a}^{\dagger}_{\downarrow,2})(\hat{a}^{\dagger}_{\uparrow,1}-\hat{a}^{\dagger}_{\downarrow,1})\pm(\hat{a}^{\dagger}_{\uparrow,1}+\hat{a}^{\dagger}_{\downarrow,1})(\hat{a}^{\dagger}_{\uparrow,2}-\hat{a}^{\dagger}_{\downarrow,2})\right)|\,\mathrm{vac}\rangle \tag{14}$$

Which turns the state into

$$\frac{1}{2\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\uparrow,1}+\hat{a}^{\dagger}_{\downarrow,2}\hat{a}^{\dagger}_{\uparrow,1}-\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\downarrow,1}-\hat{a}^{\dagger}_{\downarrow,2}\hat{a}^{\dagger}_{\downarrow,1})\pm(\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\uparrow,2}+\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\uparrow,2}-\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}-\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\downarrow,2})\right)|\mathrm{vac}\rangle \tag{15}$$

and the state is simplified as

$$\frac{1}{\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\uparrow,2}-\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\downarrow,2}\right)|\,\mathrm{vac}\rangle\ \text{ or }\frac{1}{\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}-\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\uparrow,2}\right)|\,\mathrm{vac}\rangle \tag{16}$$

Finally, we apply the swapping $\hat{a}^{\dagger}_{\downarrow,1}\leftrightarrow\hat{a}^{\dagger}_{\uparrow,2}$ back and produces

$$\frac{1}{\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,1}-\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\downarrow,2}\right)|\,\mathrm{vac}\rangle\ \text{ or }\frac{1}{\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}-\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\uparrow,2}\right)|\,\mathrm{vac}\rangle \tag{17}$$

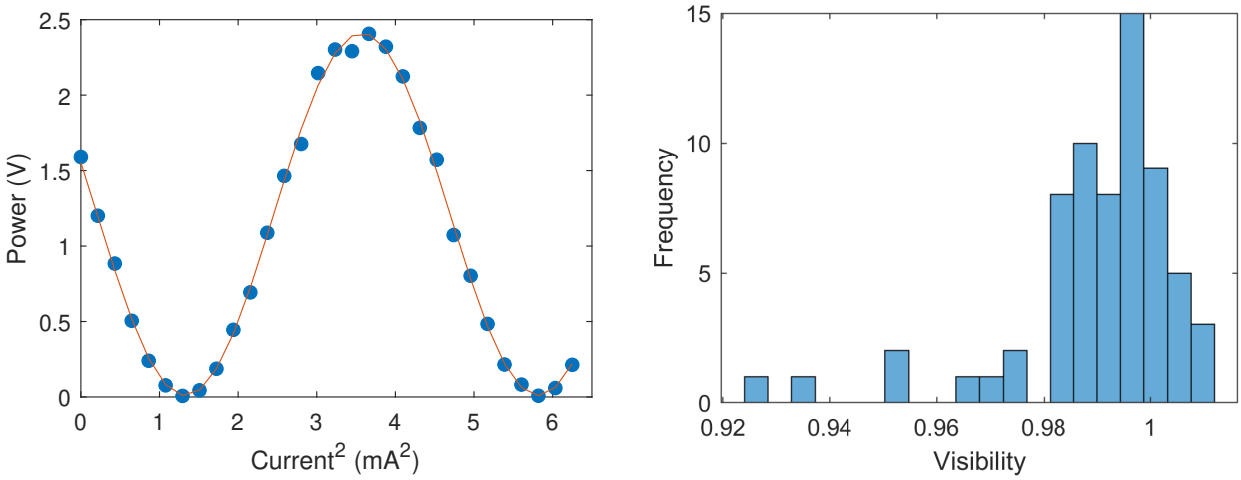
From this formula, we can determine the bell state by watching on which mode the coincidence events happens. Specifically, if the coincident event are observed in the opposite modes of the same qubit, the original state is $|\,\phi^{+}\rangle$, otherwise if the coincidence events are observed in opposite qubit and opposite modes, the original state is inferred to be $|\,\phi^{-}\rangle$. Next, we consider the evolution of another two Bell states

$$|\,\psi^{\pm}\rangle=\frac{1}{\sqrt{2}}\left(\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}\pm\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\uparrow,2}\right)|\,\mathrm{vac}\rangle \tag{18}$$

Their transformations under Bell state measurement are

$$\begin{aligned}
|\,\psi^{\pm}\rangle &= \frac{1}{\sqrt{2}}\left(\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,2}\pm\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\uparrow,2}\right)|\,\mathrm{vac}\rangle\\
&\to \frac{1}{\sqrt{2}}\left(\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\downarrow,2}\pm\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\uparrow,2}\right)|\,\mathrm{vac}\rangle\\
&\to \frac{1}{\sqrt{2}}\left(\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\downarrow,2}\pm\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\downarrow,1}\right)|\,\mathrm{vac}\rangle\\
&\to \frac{1}{2\sqrt{2}}\left(\left(\hat{a}^{\dagger}_{\uparrow,2}+\hat{a}^{\dagger}_{\downarrow,2}\right)\left(\hat{a}^{\dagger}_{\uparrow,2}-\hat{a}^{\dagger}_{\downarrow,2}\right)\pm\left(\hat{a}^{\dagger}_{\uparrow,1}+\hat{a}^{\dagger}_{\downarrow,1}\right)\left(\hat{a}^{\dagger}_{\uparrow,1}-\hat{a}^{\dagger}_{\downarrow,1}\right)\right)|\,\mathrm{vac}\rangle\\
&\to \frac{1}{2\sqrt{2}}\left((\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\uparrow,2}-\hat{a}^{\dagger}_{\downarrow,2}\hat{a}^{\dagger}_{\downarrow,2})\pm(\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\uparrow,1}-\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\downarrow,1})\right)|\,\mathrm{vac}\rangle\\
&\to \frac{1}{2\sqrt{2}}\left((\hat{a}^{\dagger}_{\downarrow,1}\hat{a}^{\dagger}_{\downarrow,1}-\hat{a}^{\dagger}_{\downarrow,2}\hat{a}^{\dagger}_{\downarrow,2})\pm(\hat{a}^{\dagger}_{\uparrow,1}\hat{a}^{\dagger}_{\uparrow,1}-\hat{a}^{\dagger}_{\uparrow,2}\hat{a}^{\dagger}_{\uparrow,2})\right)|\,\mathrm{vac}\rangle
\end{aligned} \tag{19}$$

The coincidence event happens in the same qubit and same modes. We are not able to detect them because the single photons are bunched at the same output port and would never contribute to the four-fold coincidence measurements.



(a) (b)

FIG. S8: (a) The interference fringe for the heater characteristics. (b) The histogram of visibilities for heaters. 86% of the heaters reach the visibilities over 0.98.

SUPPLEMENTARY NOTE 5: PHASE SHIFTER CHARACTERISTICS

In this chip design, phase shifters again are utilized extensively to control the AMZIs and MZIs. The power dissipation caused by applying electrical power to the TiN heater will heat the underneath optical waveguide, change the refractive index and induce a phase shift. In this chip fabrication, the process of deep-etched trenches are adopted and the heating efficiency is significantly improved. For heaters on MZIs, 3 mA electrical current is sufficient for a 3π phase shift range. For heaters on AMZIs, 7.5 mA induces a shift of more than one Free Spectral Range (FSR) in the transmission. From the voltage-current characteristics, the resistance of most heaters is $\sim$542.4 Ω. Fig. S8a displays the interference fringe of the heater. Fig. S8b shows the histogram of all the calibrated phase shifter visibilities, most of which exhibit visibilities over 98%. A visibility higher than 1 is just a mathematical artefact due to the optimization of the fitting model. The high visibility ensures high fidelities of quantum gate operations and guarantees the quantum computation accuracy.

There is a heater whose characteristics are worthy of detailed explanation, namely the one placed before the polarization rotator and splitter (PRS). The difficulty is due to that the two output ports of the MZI are combined together by the PRS and could not be separated. In an ideal model of MZI (not necessary to take the external phase ϕ into consideration) whose transfer function is

$$T = ie^{(i\frac{\theta}{2})} \begin{bmatrix} \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \\ \cos(\frac{\theta}{2}) & -\sin(\frac{\theta}{2}) \end{bmatrix} \tag{20}$$

If we inject calibration light from one input port and monitor at the combined two output ports, which is

equivalent to monitor a constant intensity which is

$$\|A\|^2 = \left\| ie^{(i\frac{\theta}{2})} \sin(\frac{\theta}{2}) \right\|^2 + \left\| ie^{(i\frac{\theta}{2})} \cos(\frac{\theta}{2}) \right\|^2 = 1 \tag{21}$$

Theoretically, we will get a constant intensity value no matter what the θ is. This is in line with our intuition that the intensity of the beam will not be affected by the beam splitting and combining operations. Therefore, a direct method is to use a polarization beam splitter for the combined TE (TM) beam, so that each can be distinguished for observation and calibration. Here we denote the light propagating in waveguide as A_{TE} and A_{TM}, and the light propagating in fiber (after passing PBS) as A_H and A_V. A_{TE} (A_{TM}) and A_H (A_V) are orthogonal, respectively. However, these two orthogonal basis are not aligned that there is an angle between them denoted as ψ, and we got the following equations

$$A_{TE} = \sin\frac{\theta}{2} \tag{22a}$$

$$A_{TM} = \cos\frac{\theta}{2} \tag{22b}$$

$$A_H = A_{TE}\cos\psi - A_{TM}\sin\psi \tag{22c}$$

$$A_V = A_{TE}\sin\psi + A_{TM}\cos\psi \tag{22d}$$

After simplification we get

$$\begin{aligned} \|A_H\|^2 &= \sin^2(\frac{\theta}{2})\cos^2\psi + \cos^2(\frac{\theta}{2})\sin^2\psi - 2\sin(\frac{\theta}{2})\cos(\frac{\theta}{2})\sin\psi\cos\psi \\ &= \frac{1}{4}(1-\cos\theta)(1+\cos2\psi) + \frac{1}{4}(1+\cos\theta)(1-\cos2\psi) - \frac{1}{2}\sin\theta\sin2\psi \end{aligned} \tag{23}$$

$$\begin{aligned} \|A_V\|^2 &= \sin^2(\frac{\theta}{2})\sin^2\psi + \cos^2(\frac{\theta}{2})\cos^2\psi + 2\sin(\frac{\theta}{2})\cos(\frac{\theta}{2})\sin\psi\cos\psi \\ &= \frac{1}{4}(1-\cos\theta)(1-\cos2\psi) + \frac{1}{4}(1+\cos\theta)(1+\cos2\psi) + \frac{1}{2}\sin\theta\sin2\psi \end{aligned} \tag{24}$$

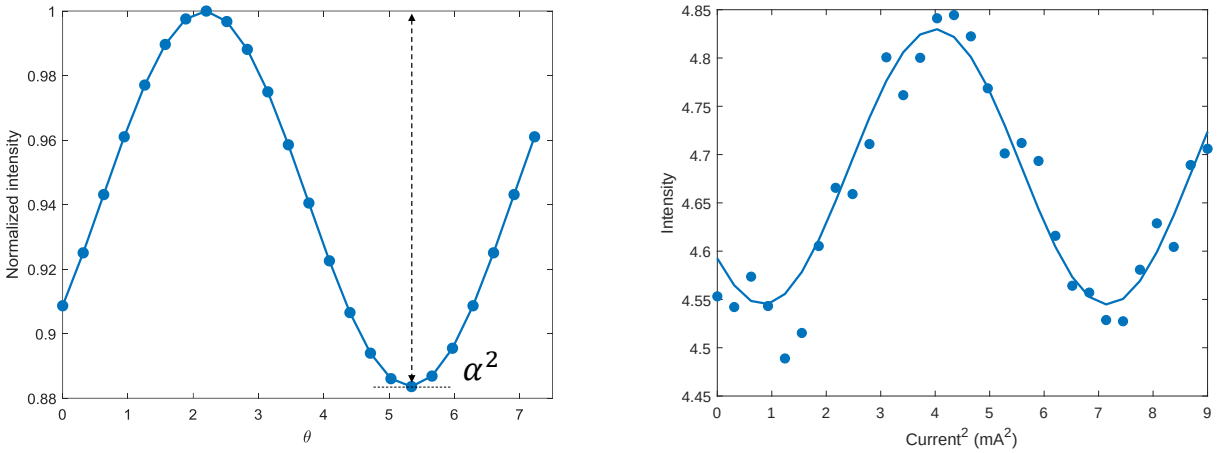
Challenge is induced by the fact that the calibration of θ and the calibration of ψ are not independently executable. However, considering that the analysis above is based on the ideal model, while in fact difference exists in the propagation loss of transverse electric (magnetic) in waveguide (although the difference is minor due to the short propagation length), we can still calibrate the heater with this slight difference. Suppose the ratio (TE/TM) of propagation loss is α, the intensity acquired at the combined port is

$$\begin{aligned} \|A_{TE}\|^2 + \alpha^2 \|A_{TM}\|^2 &= \sin^2(\frac{\theta}{2}) + \alpha^2\cos^2(\frac{\theta}{2}) \\ &= \frac{1}{2}(1+\alpha^2) - \frac{1}{2}(1-\alpha^2)\cos\theta \end{aligned} \tag{25}$$

The visibility of the calibration curve is expressed by

$$V = \frac{1+\alpha^2}{1-\alpha^2} \tag{26}$$

The results are shown in Fig. S9, and $\alpha \approx 0.97$ is estimated.



(a) (b)

FIG. S9: Calibration of the phase shifter before PRS. The loss ratio of TE to TM propagating in waveguide could be inferred from the visibility of the calibration curve. (a) The theoretical model. (b) The experimental calibration results.

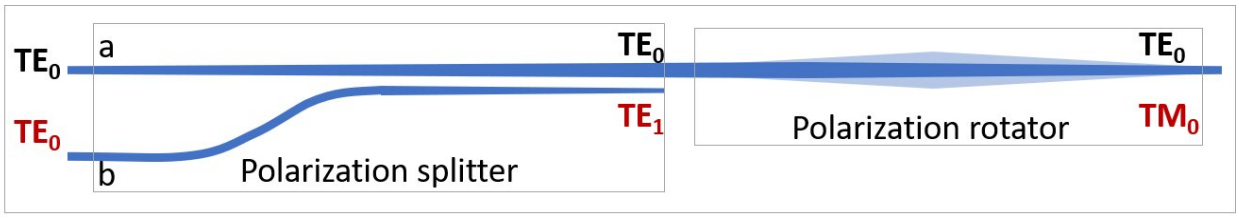


FIG. S10: Schematic of the PRS. It contains two parts, the polarization splitter and polarization rotator, to realize the conversion of qubit dual rail encoding in chip to polarization encoding in fiber and free space.

SUPPLEMENTARY NOTE 6: POLARIZATION ROTATOR AND SPLITTER

The qubit for teleportation that spans two waveguides at the encoder chip is supposed to be transferred to the decoder chip to deliver the quantum information. The intuitive way is to connect the two waveguides on encoder chip directly to the two waveguides on decoder chip, through optical fibres. However, the main concern is to maintain the phase stability of the two connecting fibres, which is rather sensitive to the environmental fluctuations.

Therefore, we propose to effectively solve this problem by combining and transferring the quantum information into a single fibre. An on-chip polarization rotator and splitter (PRS) is designed and employed to merge the two waveguides labeled with a and b (as seen in Fig. S10) into one, and convert the path state encoded on two these waveguides to the transverse electric (magnetic) modes TE_0 and TM_0 in a single waveguide. The PRS realizes the functionalities of a polarization splitter and a polarization rotator. The

cross section and coupling length is deliberately designed such that the input propagation mode TE_0 in waveguide b is first converted to TE_1 mode, and then coupled to waveguide a by a polarization splitter. The polarization rotator then rotates the TE_1 mode to TM_0 mode by the asymmetry design of waveguide cross section. The TE_0 and TM_0 are two orthogonal propagation modes in waveguide and when coupled to single mode fibres, they can occupy the two orthogonal polarization of H and V, so that the quantum information will not merge with each other and their relative phase can be maintained.

The evolution of a qubit under the structure of PRS can be written as

$$\begin{aligned} |\psi\rangle =& \alpha\,|0\rangle + \beta\,|1\rangle \\ =& \alpha\,|a\rangle\,|TE_0\rangle + \beta\,|b\rangle\,|TE_0\rangle \\ \xrightarrow{splitter}& \alpha\,|a\rangle\,|TE_0\rangle + \beta\,|a\rangle\,|TE_1\rangle \\ \xrightarrow{rotator}& \alpha\,|a\rangle\,|TE_0\rangle + \beta\,|a\rangle\,|TM_0\rangle \\ \xrightarrow{couple}& \alpha\,|fibre\rangle\,|H\rangle + \beta\,|fibre\rangle\,|V\rangle\,. \end{aligned} \tag{27}$$

where qubit state $|0/1\rangle$ is converted from $|a\rangle\,|TE_0\rangle\,/\,|b\rangle\,|TE_0\rangle$ with its waveguide position degree of freedom and propagation mode degree of freedom to the polarization degree freedom $|H\rangle\,/\,|V\rangle$.

SUPPLEMENTARY NOTE 7: LOSS CALIBRATION AND SYSTEM EFFICIENCY

The main sources of optical loss of on-chip experiments are the coupling loss from optical fibers to chip, and the insertion loss of integrated optical components. In our device, to decrease the insertion loss, the output photons are coupled out to the optical fibers using high-efficiency edge coupler arrays, the loss of which is less than 3 dB. The loss of one-dimensional (1D) subwavelength grating coupler (SGC) that used to couple input is 6 dB. However, it does not affect the overall of the photon experiment, because we can appropriately increase the pump power to compensate for this loss. The SGC loss is 4.8 dB. The propagation loss of our silicon photonic platform is $\sim$2 dB/cm, obtained by cut-back calibration on fully-etched silicon waveguide with a cross-section size of 500×220 nm^2. Similarly, we calibrate the MMI and cross that are used extensively in our design, and report losses of $\sim$0.15 dB/ea and $\sim$0.02 dB/ea, respectively.

Optical losses in the integrated chip and the experimental setup are preliminary calibrated. The total loss of each photon from generation to detection (also known as channel efficiency) is measured by the coincidence to singles ratio and averaged over all 6 channels. We measure the total loss to be 3%, which corresponds to $\sim$15.2 dB. The contributors of the loss is depicted in Fig. S2. The coupling between the input pump and the chip is not included in the total loss. Propagation loss through the entire chip is approximately 4.4 dB (0.9 cm straight waveguides and 1.3 cm spirals at 2 dB/cm). The loss of the SSC edge coupler array is 4 dB. The average transmission efficiency via the off-chip filters and the fiber connections to the detectors on all 6 channels is -2.9 dB (-2 dB for WDM, -0.9 dB for two BPFs). The insertion loss of the WDM filter

TABLE S2: Loss estimation of the transmitter chip.

Components	Loss dB	No.	Total dB
SSC Coupler	4	1	4
Waveguide	2 cm^{-1}	2.2 cm	4.4
Crossing	0.02	3	0.06
MMI	0.15	20	3
WDM	2	1	2
BPF	0.45	2	0.9
Detector	0.6	1	0.6
			14.96

varies from channel to channel, specifically C27 is 0.97 dB and C41 is 2.91 dB. The average efficiency of SNSPDs are calibrated to be 0.87 (-0.6 dB). The overall loss is summed up to be 14.96 dB.

[1] P. Baldi and K. Hornik, Neural networks and principal component analysis: Learning from examples without local minima, Neural networks **2**, 53 (1989).

[2] Y. Bengio, L. Yao, G. Alain, and P. Vincent, Generalized denoising auto-encoders as generative models, arXiv preprint arXiv:1305.6663 (2013).

[3] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Science and Technology **2**, 045001 (2017).

[4] A. Pepper, N. Tischler, and G. J. Pryde, Experimental realization of a quantum autoencoder: The compression of qutrits via machine learning, Physical review letters **122**, 060501 (2019).

[5] C.-J. Huang, H. Ma, Q. Yin, J.-F. Tang, D. Dong, C. Chen, G.-Y. Xiang, C.-F. Li, and G.-C. Guo, Realization of a quantum autoencoder for lossless compression of quantum data, Physical Review A **102**, 032412 (2020).

[6] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner, and M. Kim, Quantum generalisation of feedforward neural networks, npj Quantum information **3**, 1 (2017).

[7] L. Lamata, U. Alvarez-Rodriguez, J. D. Martín-Guerrero, M. Sanz, and E. Solano, Quantum autoencoders via quantum adders with genetic algorithms, Quantum Science and Technology **4**, 014007 (2018).

[8] Y. Ding, L. Lamata, M. Sanz, X. Chen, and E. Solano, Experimental implementation of a quantum autoencoder via quantum adders, Advanced Quantum Technologies **2**, 1800065 (2019).