nature portfolio

https://doi.org/10.1038/s41586-026-10523-6

Supplementary information

Gaussian boson sampling with 1,024 squeezed states in 8,176 modes

In the format provided by the authors and unedited

Supplementary Information for Gaussian boson sampling with 1024 squeezed states in 8176 modes

Contents

1 Squeezed light source

1.1 Squeezed light generation

Fig. S1 illustrates the optical parametric oscillator (OPO) configuration. Squeezed light at 1551.4 nm is generated through type-0 parametric down-conversion by pumping a doubly resonant OPO with 775.7 nm light below threshold. The OPO comprises a hemilithic cavity with a periodically poled potassium titanyl phosphate (PPKTP) crystal as the nonlinear medium, featuring a planar facet with an anti-reflection coating and a curved facet with a high-reflection coating. The output mirror has a partially reflective coating. Crystal dimensions and coating parameters are detailed in Fig. S2. The cavity exhibits a round-trip optical path length of approximately 40.4 mm, yielding a free spectral range of 7.4 GHz and a finesse of 59 for the fundamental mode and 29 for the pump mode. Double resonance is achieved by precise temperature tuning of the PPKTP crystal.

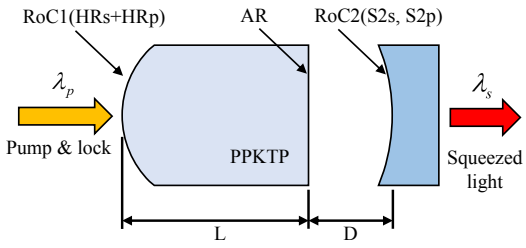


Figure S1: Schematic of the OPO.

Parameter	Value	Parameter	Value
λ_ω	1551.4 nm	ROC2	7 mm
$\lambda_{2\omega}$	775.7 nm	$S2_\omega$	90%
ROC1	12 mm	$S2_{2\omega}$	90%
HR_ω	99.978%	L	10 mm
$HR_{2\omega}$	90%	D	2.1 mm
AR	$< 0.05\%$		

Figure S2: OPO configurations.

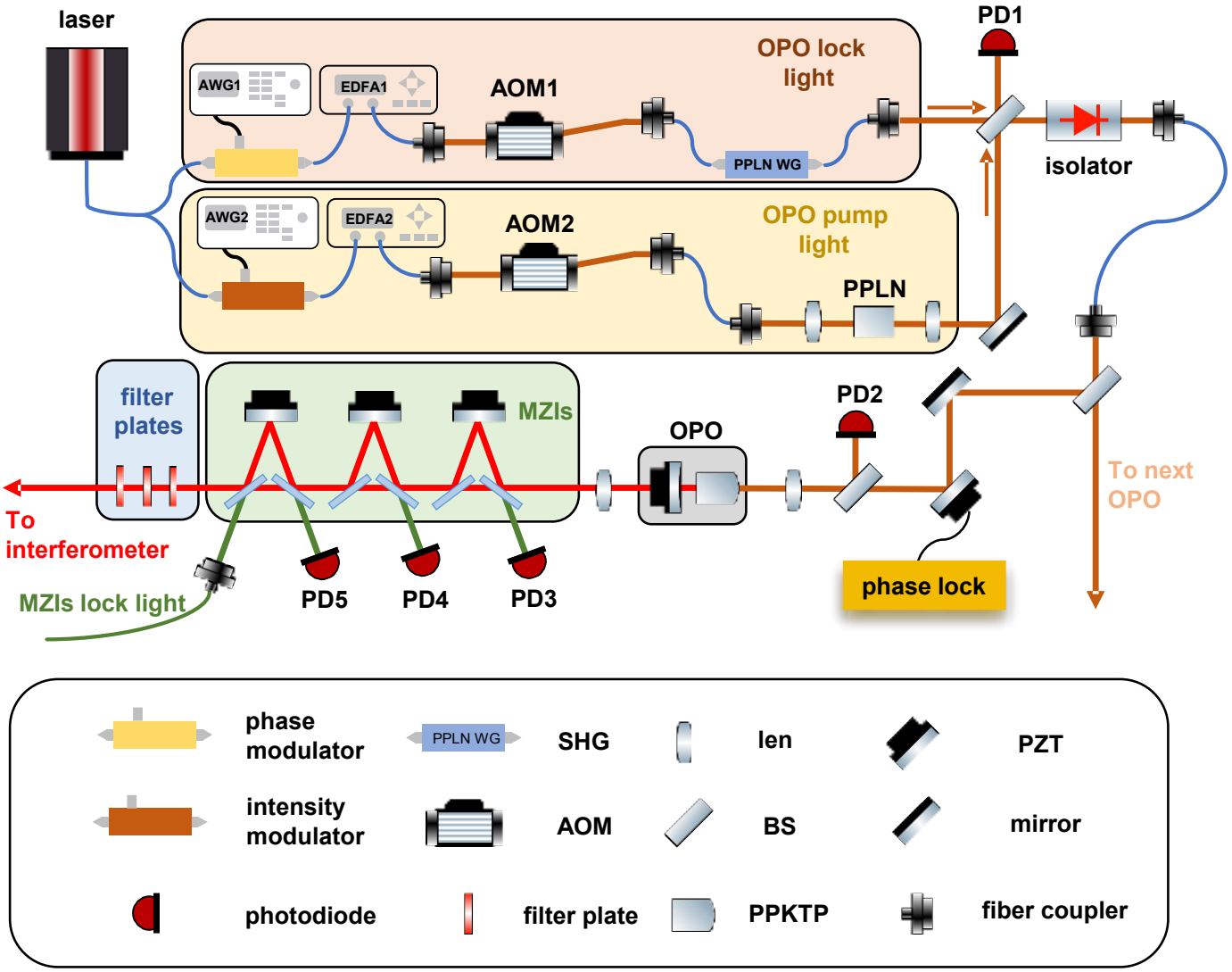


Figure S3: Setup of the squeezed light sources.

Fig. S3 depicts the detailed implementation of squeezed light generation. Rectangular pump pulses of 1.6 ns duration, centered at 775.7 nm, drive the optical parametric oscillators (OPOs) to produce down-converted photons at 1551.4 nm. The pulse duration is configured to satisfy the single temporal mode condition determined by the OPO design. A 50 ns interval between consecutive pulses aligns with the delay loops. A continuous seed laser is modulated into pulses using a 20 GHz lithium niobate electro-optic amplitude modulator (EOM) and amplified by an erbium-doped fiber amplifier (EDFA). The pump light is generated via second harmonic generation (SHG) in a periodically poled lithium niobate (PPLN) bulk crystal. Prior to entering the PPLN crystal, the amplified fundamental beam is controlled by an acousto-optic modulator (AOM), which

serves as a high-speed optical switch to regulate the number of injected pump pulses for varying sampling scales.

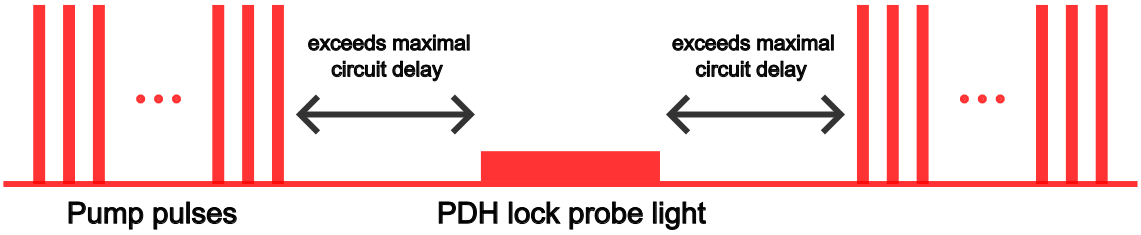


Figure S4: Temporal distribution of the pump pulses and the PDH lock probe pulse.

The Pound–Drever–Hall (PDH) technique is employed to actively stabilize the optical parametric oscillators (OPOs). The probe light is generated using a process analogous to that of the pump light. A portion of the seed laser is phase-modulated at 100 MHz using an electro-optic modulator (EOM) and amplified by an erbium-doped fiber amplifier (EDFA). The fundamental beam is shaped into 1 μs rectangular pulses and up-converted in a periodically poled lithium niobate (PPLN) waveguide to serve as the PDH lock probe. The two acousto-optic modulators (AOMs) controlling the pump and probe pulses operate at identical driving frequencies and diffraction orders, ensuring the probe and pump share the same center frequency. To prevent crosstalk and stray counts from the probe pulses, the probe and pump pulses are temporally separated, as shown in Fig. S4. The probe and pump beams are combined using a variable beam splitter, with the cost of discarding a small part of pump laser.

1.2 Theoretical simulation of the joint spectral property of the squeezed light sources

Compared to the continuous joint spectrum of spontaneous parametric down-conversion (SPDC), the resonance cavity imposes a frequency grid defined by the resonance condition. By selecting an appropriate pump linewidth, only diagonal frequency modes, such as (0,0), (+1,–1), (+2,–2), and higher orders, are retained. These modes exhibit nearly identical squeezing parameters due to the gradual variation of the phase-matching function. Cascaded narrow-band filters and unbalanced Mach-Zehnder interferometers subsequently suppress all modes except the (0,0) mode, yielding a

single-mode squeezed state as the experimental squeezed light source.

The (0,0) mode retains a fine internal spectral structure, primarily responsible for reduced spectral purity. In the low-gain regime, the joint spectrum results from the product of the transmission spectra of the two parametric beams and the pump spectrum. The -45° orientation of the pump spectrum prevents direct factorization into two independent functions, leading to a Schmidt decomposition that produces multiple frequency modes with varying amplitudes. As pump power increases, the squeezing parameters r of these modes scale proportionally. Given the exponential relationship between photon number, $n = \sinh^2(r)$ and the squeezing parameter r, spectral purity approaches unity with higher pump power.

However, the predicted increase in spectral purity does not fully align with experimental results, as the model overlooks time-ordering effects [1], which are significant in optical parametric oscillator (OPO) systems and must be included in theoretical simulations. Time ordering introduces cubic and higher odd-order terms in the relationship between the squeezing parameter and pump field intensity. This behavior can be numerically modeled using the Heisenberg equations [2].

The Heisenberg equation for the single-mode SPDC (the dispersion term is neglected because it's insignificant within a single FWHM) and the boundary conditions are:

$$\frac{\partial}{\partial z}\hat{a}_k(z,\omega) = \mathrm{i}\gamma \int \mathrm{d}\omega' \beta_k(\omega+\omega')\hat{a}_k^\dagger(z,\omega') \tag{S1}$$

$$\hat{a}_1(0,\omega) = \hat{a}_{-1}(0,\omega) \tag{S2}$$

$$\hat{a}_{-1}(L,\omega)\mathrm{e}^{-\mathrm{i}\omega L} = r\hat{a}_1(L,\omega)\mathrm{e}^{\mathrm{i}\omega L} + t\hat{a}_i(\omega) \tag{S3}$$

$$\hat{a}_o(\omega) = -t\hat{a}_1(L,\omega)\mathrm{e}^{\mathrm{i}\omega L} + r\hat{a}_i(\omega) \tag{S4}$$

Here, $\hat{a}$ denotes the annihilation operator for the down-converted photons, β represents the pump light spectrum, L is the cavity length, γ is the nonlinear coefficient, r and t are the reflectance and transmittance of the cavity mirrors, respectively, and z is the spatial coordinate along the light propagation direction within the PPKTP crystal. The parameter $k = \pm 1$ distinguishes the two counter-propagating beams within the cavity. The divergence of the Gaussian beam inside the PPKTP crystal is neglected, as its spectral impact occurs on a scale significantly larger than the free spectral range (FSR).

Furthermore, the above equations can be discretized and then solved numerically:

$$\mathbf{u}_{n,k}(z) = \hat{a}_k(z, \omega_n)$$
$$\mathbf{v}^{\dagger}_{n,k}(z) = \hat{a}^{\dagger}_k(z, \omega_n)$$

$$\begin{pmatrix} \mathbf{u}_k(z) \\ \mathbf{v}^{\dagger}_k(z) \end{pmatrix} = \mathrm{e}^{\mathrm{i}Q_k l} \begin{pmatrix} \mathbf{u}_k(z) \\ \mathbf{v}^{\dagger}_k(z) \end{pmatrix}$$
$$\begin{pmatrix} \mathbf{u}_{-1}(z) \\ \mathbf{v}^{\dagger}_{-1}(z) \end{pmatrix} = rU_{\phi} \begin{pmatrix} \mathbf{u}_1(z) \\ \mathbf{v}^{\dagger}_1(z) \end{pmatrix} + t \begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}^{\dagger}_i \end{pmatrix}$$
$$\begin{pmatrix} \mathbf{u}_o \\ \mathbf{v}^{\dagger}_o \end{pmatrix} = -tU_{\phi} \begin{pmatrix} \mathbf{u}_1(z) \\ \mathbf{v}^{\dagger}_1(z) \end{pmatrix} + r \begin{pmatrix} \mathbf{u}_i \\ \mathbf{v}^{\dagger}_i \end{pmatrix} \tag{S5}$$

Here, Q comes from the discretization of Eq. (S1), U_{ϕ} comes from the discretization of the boundary conditions Eq. (S3) and l represents the length of PPKTP. Solving the above matrix equation yields the unitary operator for the entire process:

$$\begin{pmatrix} \mathbf{u}_o(z) \\ \mathbf{v}^{\dagger}_o(z) \end{pmatrix} = \frac{r - U_{\phi}\mathrm{e}^{\mathrm{i}Q_1 l}\mathrm{e}^{\mathrm{i}Q_{-1} l}}{I - rU_{\phi}\mathrm{e}^{\mathrm{i}Q_1 l}\mathrm{e}^{\mathrm{i}Q_{-1} l}} \begin{pmatrix} \mathbf{u}_i(z) \\ \mathbf{v}^{\dagger}_i(z) \end{pmatrix} \tag{S6}$$

Finally, performing the Schmidt decomposition on $\mathbf{u}_o$ will yield the spectrum and squeezing parameter of each Schmidt mode, which further leads to the relationship between the spectral purity and the pump power.

1.3 Measurement of the spectral purity

Spectral purity of the squeezed light sources is quantitatively calibrated using unheralded second-order correlation measurements, implemented via a Hanbury-Brown-Twiss (HBT) experiment on thermal states generated by interfering two single-mode squeezed states. The measurement setup, illustrated in Fig. S5, consists of two cascaded 50:50 balanced two-mode interferometers interconnected by fiber delay loops with a time delay difference of τ. When two single-mode squeezed state pulses, separated by τ, traverse the circuit, the second pulse interferes with the first at the second interferometer, each with 50% transmission. With the interference phase correctly set,

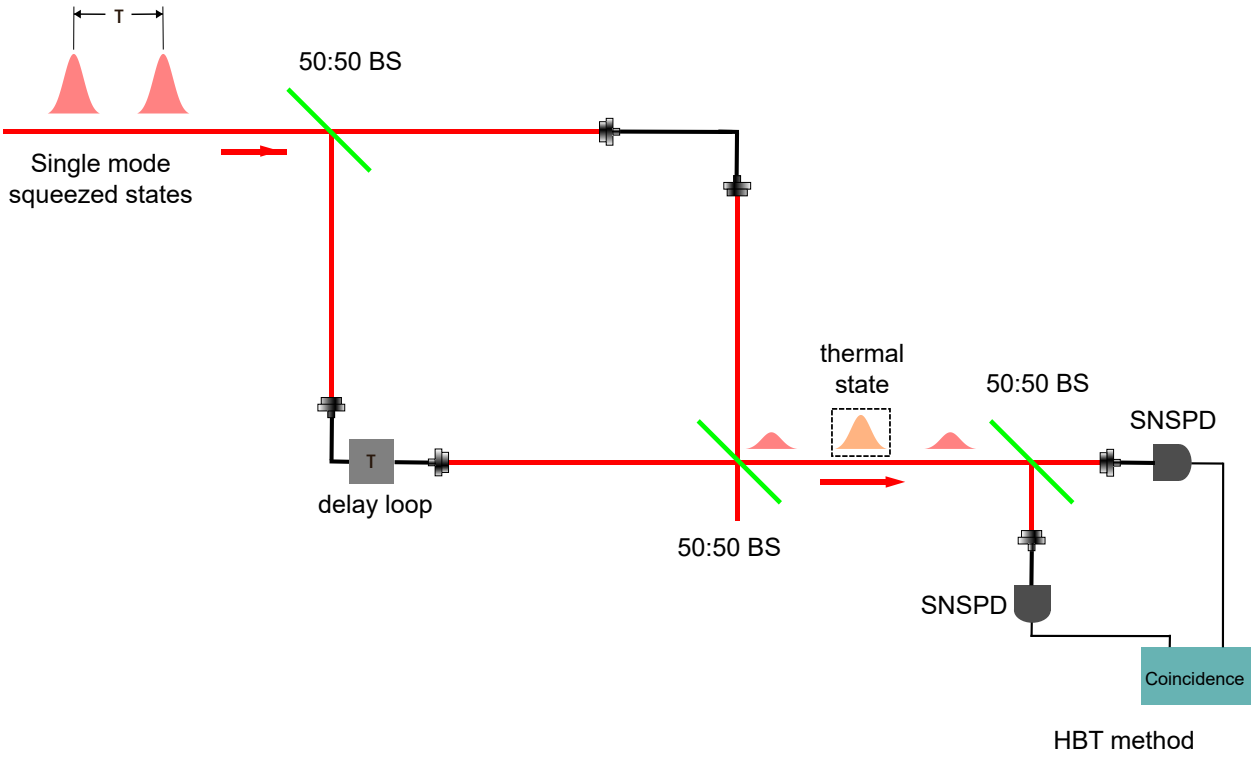


Figure S5: Measurement of the squeezed light sources' spectral purity.

two independent thermal states are produced at the output ports. As thermal states remain invariant under arbitrary transmission losses, such losses do not affect the calibration. The thermal state from one output port is selected and analyzed using a 50:50 beam splitter in an HBT configuration to measure the second-order correlation.

We model the thermal state as a mixing of a major mode with mean photon number $\bar{n}$ and a distinguishable minor mode with mean photon number $\bar{m}$, where $\bar{m} << \bar{n}$. $\bar{n}$ and $\bar{m}$ can be numerically solved from the measured second-order correlation function $g_2(0)$ and the click probability p at one port:

$$g_2(0) = \frac{(2+\bar{n})(2+\bar{m})(2(\bar{n}^2+\bar{m}^2)+2\bar{n}\bar{m}+3(\bar{n}^2\bar{m}+\bar{m}^2\bar{n})+\bar{n}^2\bar{m}^2)}{(1+\bar{n})(1+\bar{m})(\bar{n}\bar{m}+2(\bar{n}+\bar{m})^2} \tag{S7}$$

$$p = \frac{2(\bar{n}+\bar{m})+\bar{n}\bar{m}}{(2+\bar{n})(2+\bar{m})} \tag{S8}$$

We mainly care about the ratio of the distinguishable minor component,

$$R_{\bar{n}} = \frac{\bar{n}}{\bar{n}+\bar{m}} \tag{S9}$$

and the spectral purity, which is defined as:

$$\mathbb{P} = \frac{\bar{n}^2 + \bar{m}^2}{(\bar{n} + \bar{m})^2} \tag{S10}$$

1.4 Measured spectral purity under different squeezing level

The measurement is conducted for both interference of a single OPO, which follows the delayed scheme in Sec.1.4, and direct interference between two independent OPOs. The two cases give the nearly identical results and coincide well with the theoretical simulation(Fig. S6).

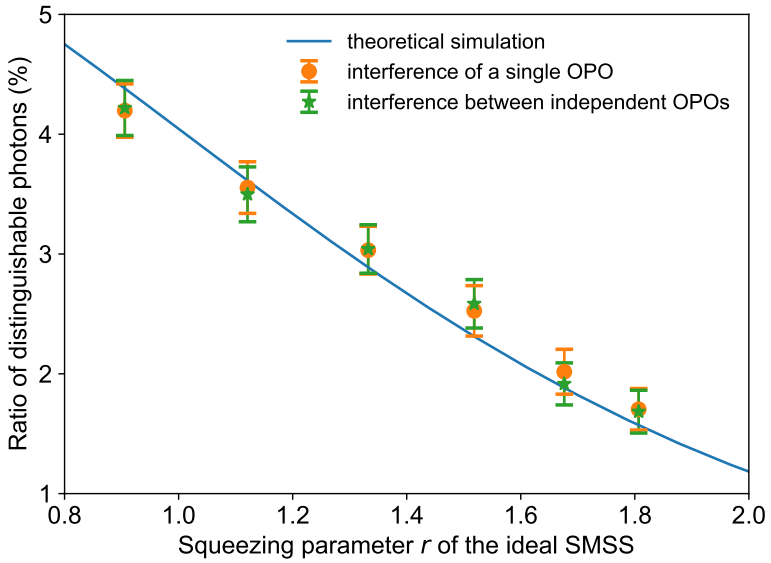


Figure S6: Measured spectral purity under different squeezing level. The error bars are obtained from repeated data collection and processing.

1.5 Measured spectral purity under different temporal delay

The circuit architecture relies on long-range temporal interference, raising concerns about whether extended temporal delays compromise interference indistinguishability. To investigate, we measured spectral purity using temporal interference with delay lengths of 50 ns, 800 ns, and 12 800 ns, the latter being the maximum delay in our circuit. Results presented in Fig. S7, show nearly identical spectral purity across all delay lengths, indicating that temporal delays within our

circuit's range do not observably degrade the indistinguishability of squeezed light pulses generated at different times.

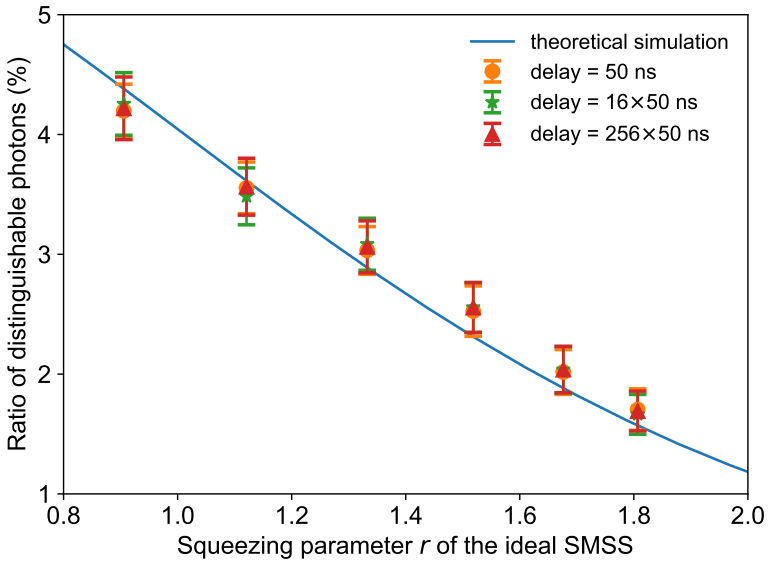


Figure S7: Measured spectral purity under different temporal delay.

2 Programmability of the system

The system's programmability, illustrated in Fig. S8, enables flexible configuration of input sequences for single-mode squeezed states (SMSSs) and the photonic circuit. By controlling the fiber electro-optic modulator (EOM) that shapes the seed laser and the free-space acousto-optic modulator (AOM) that switches the fundamental beam output from the erbium-doped fiber amplifier (EDFA), the SMSS input sequence can be tailored for sampling tasks with varying input scales and optimized repetition rates.

There are three ways that the transformation matrix can be reconfigurable:

1) The non-optical surface of the three 16-mode interferometers are tied with thermally tunable components, enabling temperature control by adjusting electric current. Thus the refractive index of fused silica can be altered, which tunes the phase inside the interferometers and reconfigures its transformation status. It is not universally programmable inside the SU(16) space. According to the measurement, this requires approximately 20 seconds for the system to stabilize under thermal

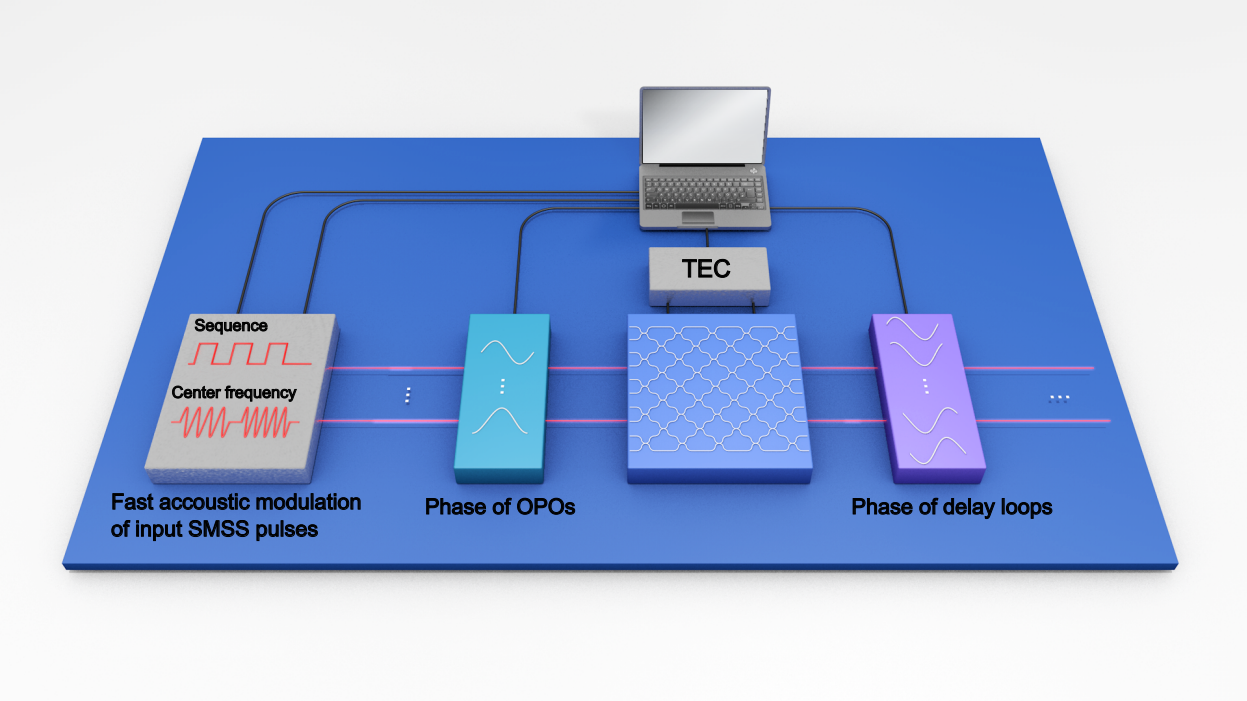


Figure S8: Programmability of the system.

control.

2) The phase of each incident OPO (4 in total) and fiber loop ($2 \times 16 = 32$ in total) can be arbitrarily set by configuring the corresponding phase-lock module setpoints, i.e, adjusting the displacement of the piezo actuators inside the optical circuit. The response latency is less than 100 μs.

3) The center frequency of each single-mode squeezed state (SMSS) pulse can be adjusted by modulating the acousto-optic modulator (AOM) driving signal, effectively altering the phase between temporal modes due to delayed interference. Though not implemented in the publicly released data, this can be realized with a fast AOM (e.g, a fiber AOM working at 400MHz center frequency whose rise/fall latency is well below 20 ns) at the seed laser stage, and the center frequency of each incident pump pulses (the temporal interval is 50 ns) can be individually set by programming the driving radio signal sequence of the used AOM with a high-speed arbitrary waveform generator.

Before the experiments, we reset the programmable parameters of the circuit and fix the parameter status during data acquisition and calibration of the transformation matrix.

3 Photon number distribution of different input scales

Photon number distribution of all the three groups of experiments are shown in Fig. S9.

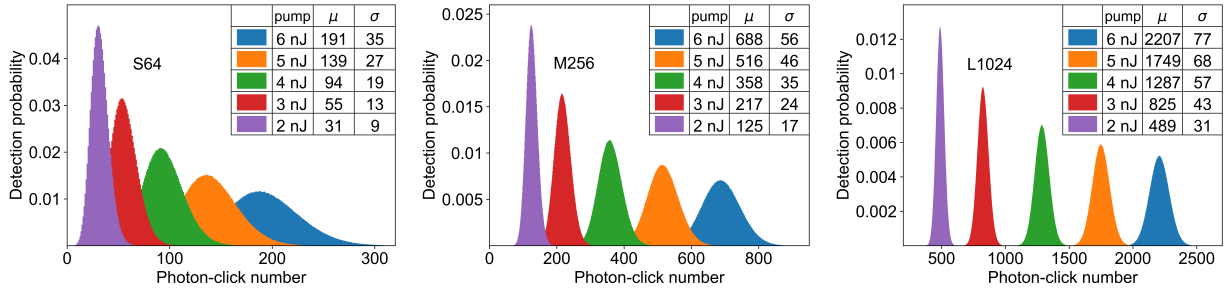


Figure S9: Experimental photon-click number distribution under different pump intensity for all the three input scale of the whole setup, with their mean μ and standard deviations σ listed in the inner panel. The level of pump intensity is represented by the energy of a pump laser pulse to generate a single SMSS pulse.

We further present the Fano scaling (the ratio between the variance and the mean) of the photon number distribution for all input scales and pump settings (Fig. S10).The experimental data agrees well with the theoretical prediction and significantly deviate from the classical mockups.

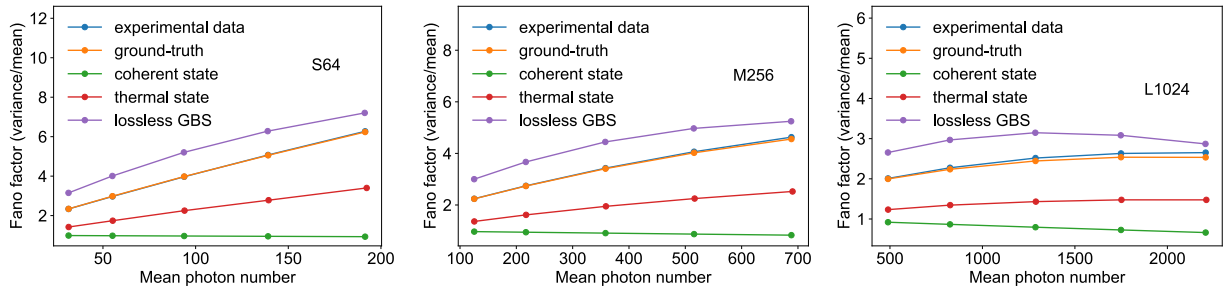


Figure S10: Fano scaling showing the variance dependence on the mean across all input scales and pump settings, with the expected reference lines for the ground-truth, coherent, thermal, and squeezed-vacuum in lossless regimes.

A slight deviation appears for those large experiments containing thousands of photons, where the experimental distribution is slightly broadener than the theoretical prediction (larger Fano factor). This is mainly due to pump power drift: suppose the mean of the distribution is N, then the

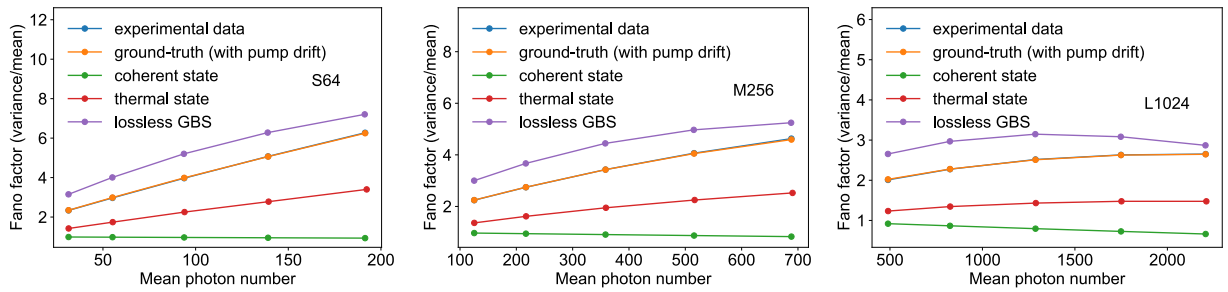


Figure S11: A uniform pump drift correction is introduced for all the three experiments, then a perfect alignment is achieved for all datasets.

width (or the variance) of the distribution scales approximately as $\sqrt{N}$, but the drift of mean scales linearly as N, then the ratio of drift and the variance itself would has a $N/\sqrt{N} = \sqrt{N}$ factor, which indicates that under the same pump power drift, the variance broadening effect would be amplified and thus more observable in the high-photon-number regime.

To support this, we introduce a uniform pump drift model for all the three scales and all the pump settings, then a perfect coincidence is observed for all these experiments (Fig. S11).

Following Fig.2b in the main text which displays the comparison with different theoretical models on a single pump setting of L1024 dataset, we present the comparison results for all pump settings (Fig. S12). The conclusions remain consistent that the experimental distribution coincides well with the groundtruth while significantly deviates from all the classical mockups. Here the distribution of TP-MPS sampler is also plotted, which nearly degrades to the squashed state model as argued in Methods and clearly performs worse than the experimental data. There is a trend that the distribution of ground-truth and the squashed state gets closer to each other as the pump intensity and thus the squeezing is increased, so in the main text we plotted the results of L1024 (3 nJ) for a more informative presentation in terms of visual distinguishability.

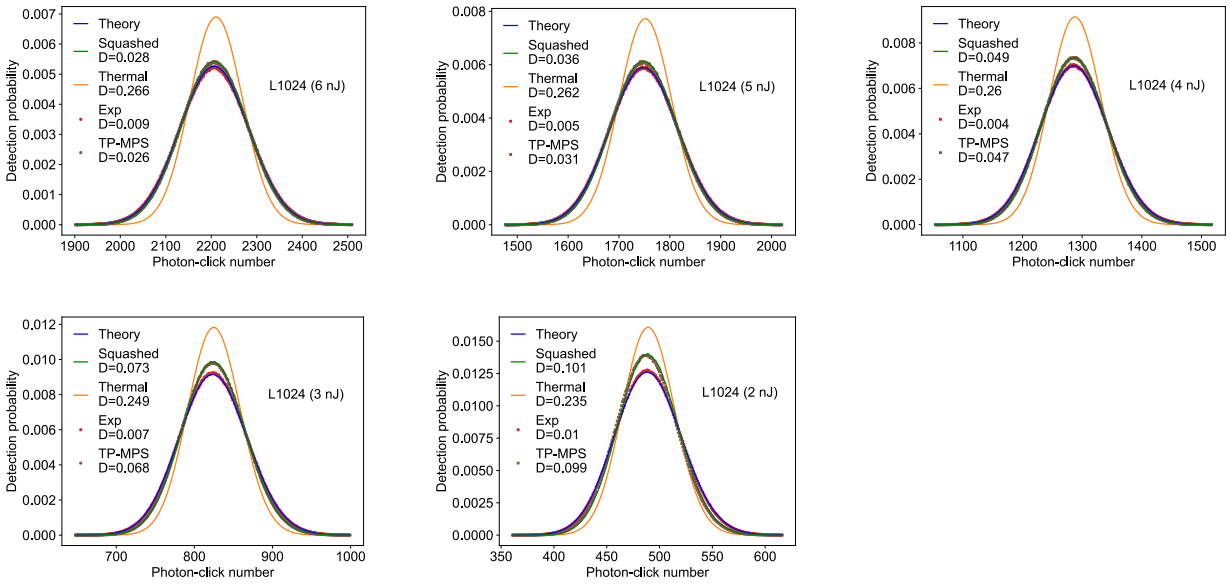


Figure S12: Comparison of the photon number distribution of the experimental results with the ground-truth and several classical mockup models, for all pump settings of L1024 dataset. D stands for the total variation distance (TVD) between the ground-truth distribution and the distribution of various samplers.

4 Calibration of the transformation matrix

The circuit's spatial-temporal hybrid encoding enables factorization of both input and output modes into spatial and temporal dimensions. The input comprises $M_s \times M_t$ modes, where M_s denotes the number of utilized spatial input ports of the first interferometer (or optical parametric oscillators) and M_t represents the number of temporal inputs, determined by the number of pump pulses. Given the circuit's maximum time delay, an input of M_t pump pulses generates a sequence of squeezed light pulses spanning M_t temporal modes, resulting in photons distributed across $M_t + 255$ output temporal modes. With the final interferometer's 16 spatial modes, the total output modes are $16 \times (M_t + 255)$. The interference network's transformation matrix, with input modes as rows and output modes as columns, has dimensions $M_s \times M_t$ rows and $16 \times (M_t + 255)$ columns.

The circuit's architecture enables coupling of each input mode to $256 \times 16 = 4096$ output modes. Given a uniform pump beam (no phase or frequency is modulated) and the fixed circuit, due to high symmetry of temporal encoding, the interference matrix for a single input temporal

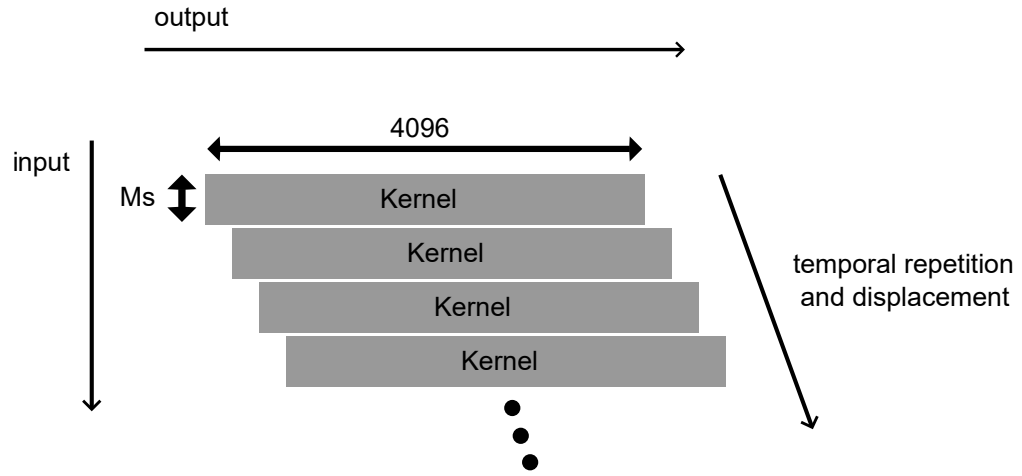


Figure S13: The entire transformation matrix can be constructed from the kernel.

mode—a kernel matrix of dimensions $M_s \times 4096$ —is sufficient to characterize the system. The full interference matrix is constructed by iteratively repeating and displacing this kernel for all input temporal modes (Fig. S13).

4.1 Measurement of the matrix amplitude

To measure the matrix amplitude (beam coupling ratio), a sequence of squeezed light pulses from a single optical parametric oscillator (OPO) is input with a temporal interval exceeding the maximum connectivity of the fiber delay loops, ensuring no internal interference occurs. Given the circuit's large scale, a single pulse input yields low count rates per output mode. Sufficient data are collected to minimize statistical fluctuations, with stray counts subtracted. The matrix amplitude is determined by first calibrating the input pulses' squeezing parameter, optimized by fitting the measured click rate and photon number distribution.

Step A : For a given squeezing r_t, coupling ratio m_i on each output mode can be calculated from the corresponding click probability:

$$m_i = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \tag{S11}$$

$$A = \sinh^4(r_t) - \frac{\sinh^2(2r_t)}{4} \tag{S12}$$

$$B = 2\sinh^2(r_t) \tag{S13}$$

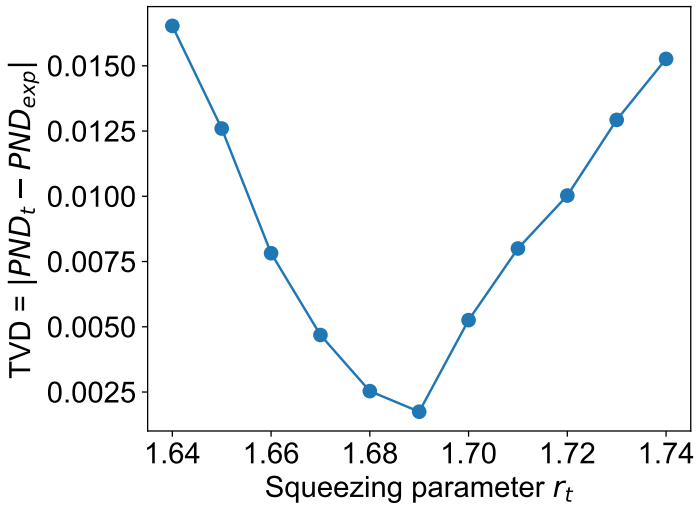


Figure S14: **Typical results to determine the input squeezing by the minimum-TVD between the measured photon number distribution.**

$$C = 1 - \frac{1}{(1 - p_i)^2} \tag{S14}$$

where p_i is the click probability of the i_{th} output mode;

Step B : Once r_t and the overall coupling ratio for all output modes M_t is determined, then the sampling model defined by the single SMSS input is well defined. Thus the coarse-grained photon number distribution PND_t can be calculated by the phase space simulation method introduced in [3].

Step C : Calculate the total variation distance(TVD) between PND_t and the photon number distribution PND_{exp} counted from the experimental data.

By the method, there will be an optimal squeezing parameter which attains the minimum TVD value (Fig. S14). Repeat step A to step C by linear search until the optimal input squeezing is found with sufficient precision.

Once the squeezing parameter of the input pulses is determined, the coupling ratios for all 4096 output modes are derived directly from the measured click probabilities. This procedure is repeated for each optical parametric oscillator (OPO) input to obtain the 4×4096 kernel matrix of amplitudes. The overall system transmission for each OPO is calculated by summing the 4096 elements in the corresponding row of the kernel matrix.

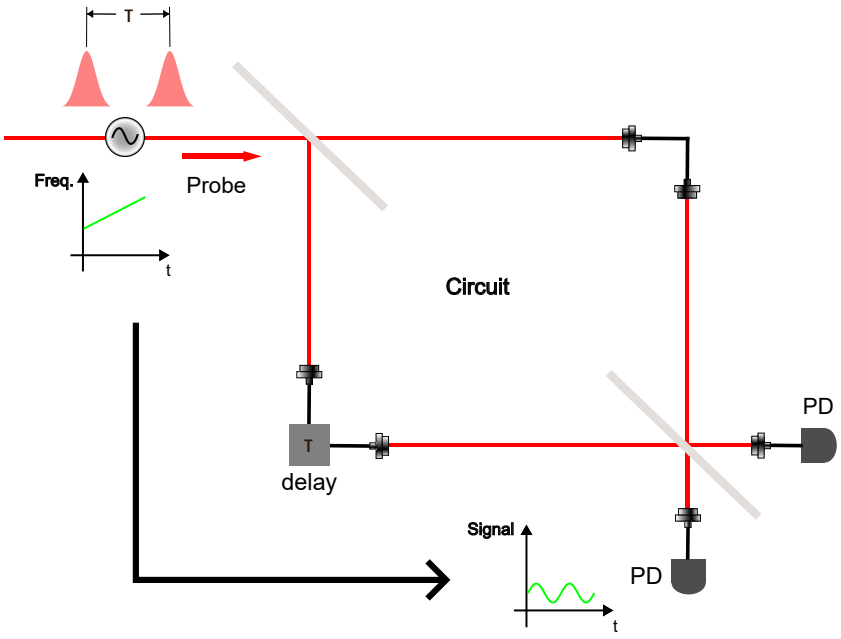


Figure S15: Frequency-swept laser pulses for phase measurement.

4.2 Measurement of the matrix phase

The interference network's phase elements are determined by analyzing two components: the internal circuit phase and the phase from squeezed light generation to circuit entry. Still we focus on how to obtain the 4×4096 kernel of the entire phase matrix.

4.2.1 Measurement of phase inside the circuit

Frequency-swept laser pulses, matched to the circuit's temporal encoding, are injected to interfere within the network and probe phase information. The temporal delay makes frequency sweeping equivalent to actively scanning the interference phase, yielding periodic interference patterns (Fig. S15). By precisely referencing the probe laser's frequency to that of the squeezed light, the phase experienced by the squeezed light within the circuit is directly extracted from the measured curves.

The phase measurement setup, illustrated in Fig. S16, employs two fiber acousto-optic modulators (AOMs) operating at $400M \pm 100$ MHz, providing a frequency sweep range of up to 200 MHz. The frequency-modulated output is amplified by an erbium-doped fiber amplifier (EDFA) and

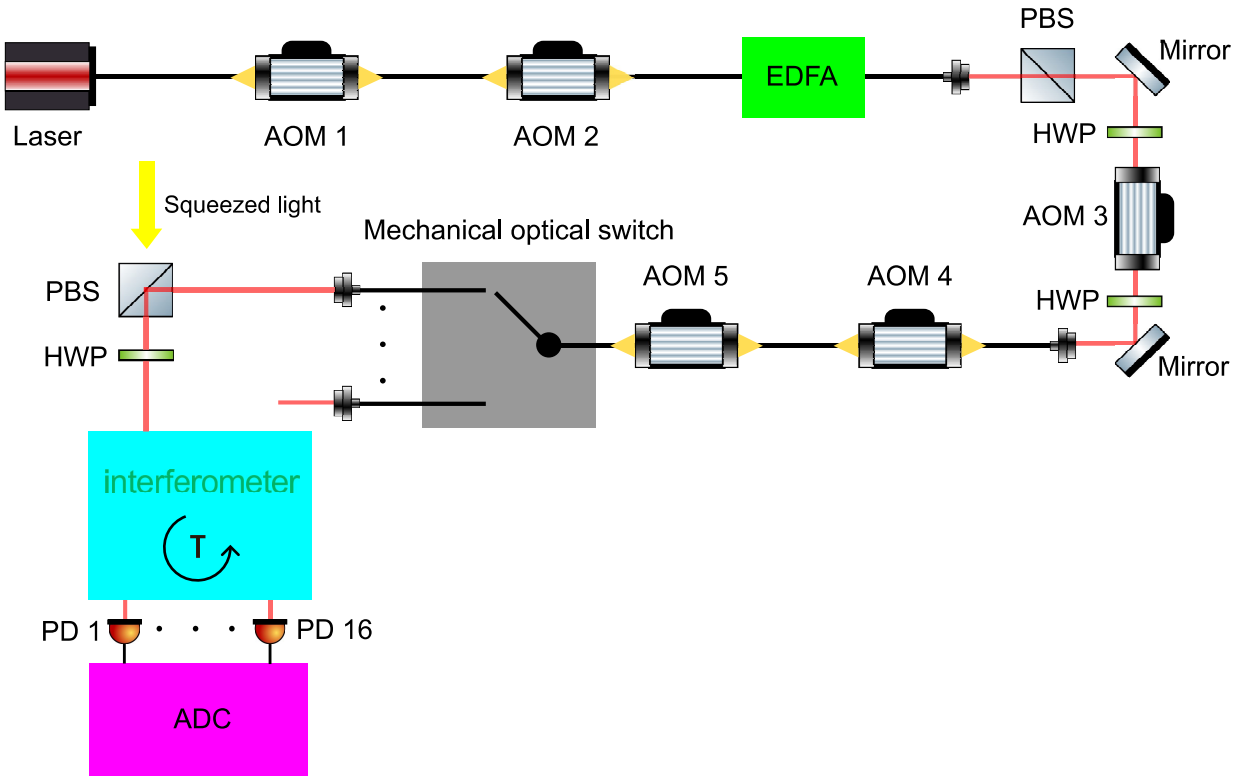


Figure S16: Setup of measurement of the phase inside the circuit.

shaped into 25 ns rectangular pulses using three additional AOMs. AOM4 and AOM5, fiber-based AOMs driven at 400 MHz with a 6 ns rise/fall time and >50 dB extinction ratio, generate high-quality pulse sequences. AOM3, a free-space AOM synchronized with AOM4 and AOM5, acts as a light switch to protect the fiber AOMs from continuous high-power laser input. The diffraction frequencies of all AOMs are calibrated to align the sweep's zero point with the squeezed light's center frequency.

The probe beam is coupled to the squeezed light's path via a polarization beam splitter before the first interferometer, followed by a half-wave plate (HWP) to rotate the polarization. During measurement, laser pulses are injected into a single input port, with a micro-mechanical light switch automating iteration across all input ports. Light signals are detected by two 8-channel photodetectors with a 200 MHz bandwidth, and their outputs are processed by two 8-channel analog-to-digital converter (ADC) boards with a 250 MHz sampling rate for data acquisition.

For each spatial input port, the circuit interconnects 256 temporal input modes. Relative to the first pulse, interference is measured between the first pulse and each of the subsequent 255 pulses individually. For each of the 255 temporal inputs, double-pulse units with matching delay intervals are repeatedly injected to generate corresponding interference patterns and extract phase information. Frequency sweeping, implemented in linear step mode and precisely referenced to the squeezed light's center frequency, scans hundreds of frequency points for each curve to produce totally millions of sinusoidal curves, each containing at least eight periods. Curve fitting enables retrieval of all interference phases.

The phase matrix is constructed using a decomposition-and-assemble approach. The interference network comprises three cascaded interferometers, M1, M2 and M3, with the phase of the first delay loop array incorporated into M2 and the second into M3 for simplicity. By determining the phase of these interferometers, the phase matrix is assembled directly from the three components. The focus remains on the 4×4096 kernel matrix K.

Without loss of generality, we index the i_{th} spatial input to be the i_{th} input port of the first interferometer M1, which corresponds to $K[i]$, the i_{th} row of the kernel. For the sake of expressional convenience, we further reshape the 4096 elements of $K[i]$ to be a matrix of size (256,16), where the row dimension represents the 256 temporal modes and the column dimension represents the 16

spatial modes of the last interferometer M3. Determined by its path inside the circuit,

$$K[i](j,k) = M1(i, j\%16) + M2(j\%16, j//16) + M3(j//16, k) \tag{S15}$$

where M1,M2,M3 are three 16×16 matrix representing the three interferometers, the first dimension for the input and the second dimension for the output; indexes i,j,k start from 0; % means the mod operation and // means the integer division operation in computer science.

The measurement results are described by a four-dimensional vector $\Phi(i,l,j,k)$, where i denotes the i_{th} input port of M1, l the temporal input dimension, j the temporal output dimension, and k the spatial output of M3. Referenced to the first temporal input, the elements of Φ are expressed as the sum of contributions from M1,M2 and M3, determined by tracing the pulse paths through the circuit to the corresponding output time and port,

$$\begin{aligned}\Phi(i,l,j,k) =&(M1(i, j\%16) + M2(j\%16, j//16) + M3(j//16, k)) \\ &-(M1(i, (j-l)\%16) + M2((j-l)\%16, (j-l)//16) + M3((j-l)//16, k))\end{aligned} \tag{S16}$$

Phase of the three 16 mode interferometers can be solved from the measurement results. For instance, one choice to yield M1,M2 and M3 is

$$(M1(i,j) - M1(i,0)) - (M1(0,j) - M1(0,0)) = \Phi(i,j,j,0) - \Phi(0,j,j,0) \tag{S17}$$

$$M2(i,j) - M2(0,j) = \Phi(0,i,16\times j + i,0) - (M1(0,i) - M1(0,0)) \tag{S18}$$

$$M3(i,j) - M3(0,j) = \Phi(0,16\times i,16\times i,j) - (M2(0,i) - M2(0,0)) \tag{S19}$$

Actually, the entire elements of Φ is over-complete to solve for M1,M2 and M3 because there are combinatorial number of topologically equivalent relations which can yield the same results as those given in (S17, S18, S19). We take average over all these equivalent expressions to figure out M1,M2 and M3 which can suppress the statistical fluctuation and refine the calibration precision. Once the elements of M1,M2 and M3 are determined, the matrix kernel of the circuit's phase can be directly calculated by S16.

4.2.2 Determine the phase of squeezed light sources

Multiple optical parametric oscillators (OPOs) serve as squeezed light sources in our experiment, each with an individual phase prior to entering the circuit, determined by the pump pulses and

the subsequent propagation of squeezed light to the first interferometer. For the i_{th} OPO, this source phase corresponds to an overall phase added to the i_{th} row of the phase matrix's kernel. One OPO is designated as the reference source, and the phases of the others are measured sequentially. This is achieved by performing sampling tasks with only the reference and one other OPO active, with all irrelevant OPOs blocked. The second-order correlation function is computed from experimental data, and the source phase is determined by linearly searching for the value that minimizes the total variation distance between the experimental statistics and theoretical predictions (a typical curve in data processing is shown in Fig. S17).

Preliminary : The input squeezing vectors of the two light sources (one is the reference OPO and the other is the measured OPO), the respective transformation matrix for the two input OPO U_{ref} and U_{target} (calibrated by the frequency-swept method introduced above), the two-order correlators C^2_{exp} counted from the two-OPO interference data.

Step A : For a given OPO input phase ϕ_t, get the transformation matrix by incorporating the input phase and merging the two single-OPO matrix : $U_{\text{2-OPO}} = (U_{\text{ref}}, U_{\text{target}}\exp(i\phi_t))$. Then calculate the theoretical two-order correlators C^2_{t} from the input squeezing and the transformation matrix.

Step B : Calculate the norm-1 distance between C^2_{t} and C^2_{exp}.

Step C : Repeat the procedure in A and B by linear searching the optimal ϕ_t which reaches the minimum distance between the theoretical calculation and the experimental statistics.

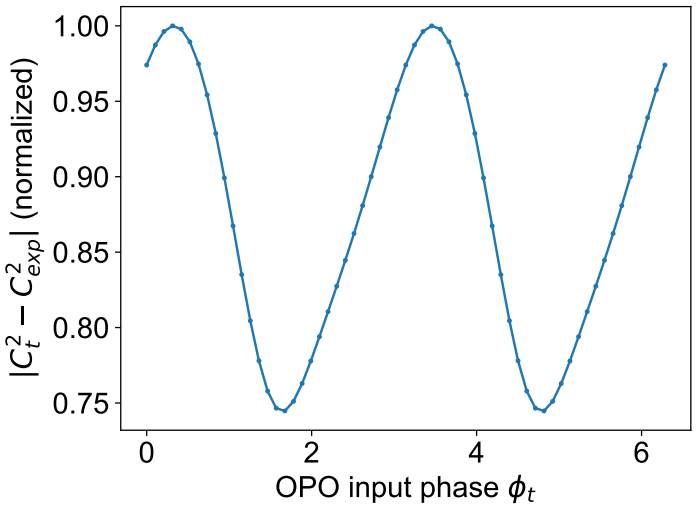


Figure S17: **Typical results to search for the OPO input phase by the minimum distance between the experimentally measured and theoretically calculated two-order correlation.**

4.2.3 Accuracy of phase stabilization and calibration

We analyzed the totally millions of direct phase measurement results, in terms of the intrinsic routing correspondence of the circuit, which yields an overall phase error of both stabilization and calibration to be within $\lambda/200$ at 1550nm (Fig. S18).

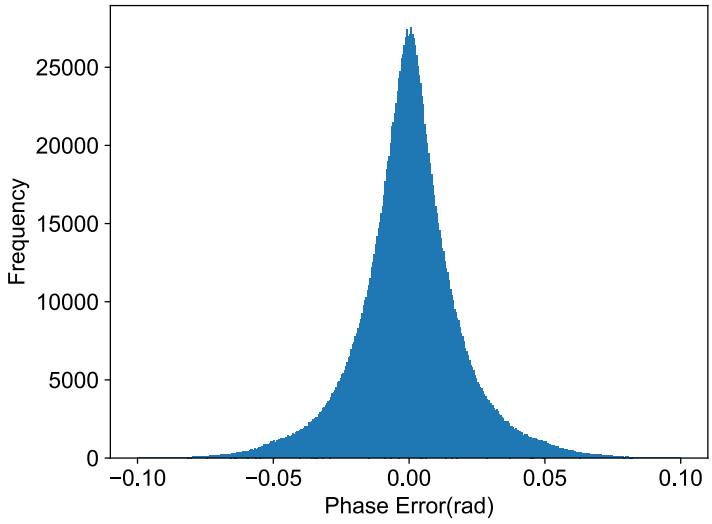


Figure S18: Analyzed phase error of both stabilization and calibration of the circuit.

5 Partial photon distinguishability model

In the main text, we model the ground-truth as a standard lossy GBS, without the photon distinguishability incorporated. In this section we describe the theoretical foundation of he model and present the validation results under the distinguishability-incorporated ground-truth.

5.1 Click probability calculation

The photon distinguishability is attributed to the non-ideal joint spectrum of spontaneous parametric down-conversion (SPDC), which generates higher-order distinguishable spectral modes except the ideal degenerate single-mode squeezed state. The physical mechanism and experimental calibration has been detailed in Sec.1.3~1.6, here we give the method to compute sampling click probability. As sketched in Fig. S19, each spectral mode, varying in squeezing parameters, independently traverses the same photonic circuit (described by the same transformation matrix) and clicks at the output. To account for the detection , we model virtual arrays of detectors at the circuit's output, with each array corresponding to the click pattern of the corresponding spectral mode. The final output results from a logical OR operation on signals from all detector arrays.

Mathematically, since these spectral modes are independent, the quantum state is the direct product of the state of different spectral modes. We denote the set of the spectral modes by the

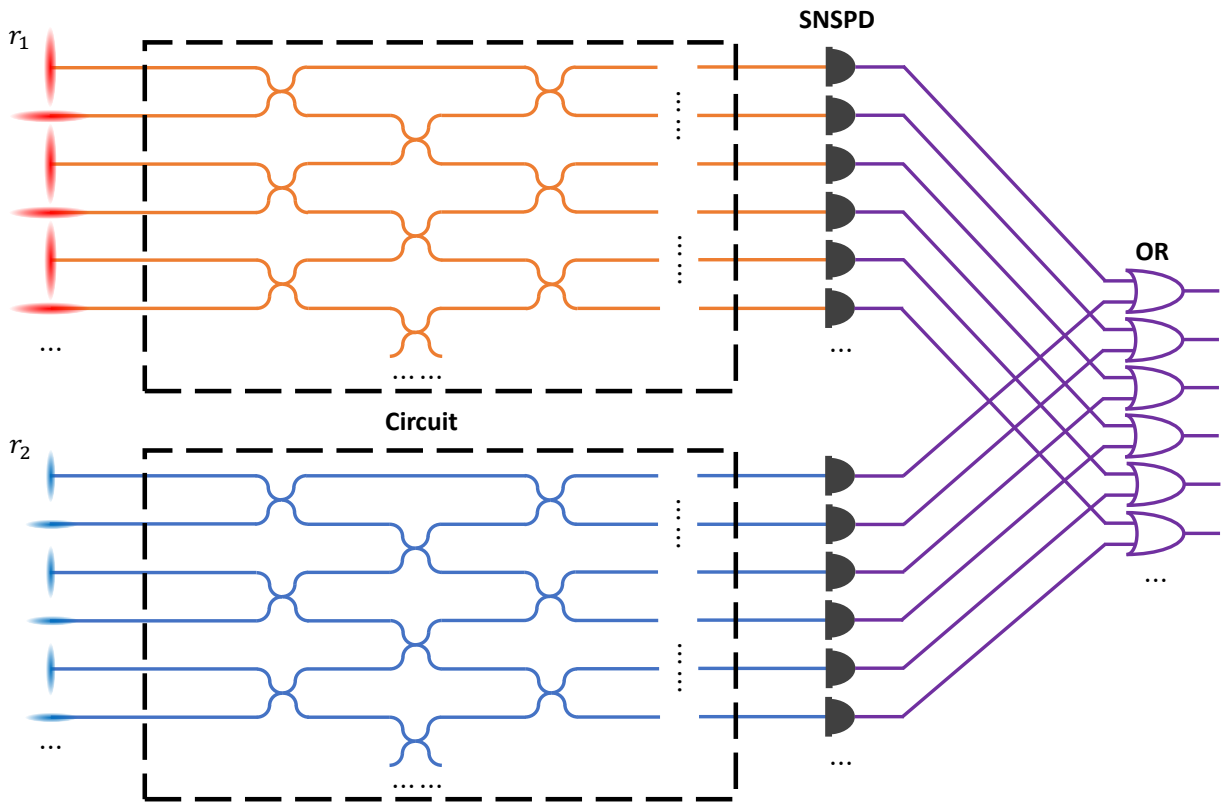


Figure S19: Sketch of the partial photon distinguishability model using threshold detectors.

index $m \in \{1, 2, \cdots, M\}$, where M is the quantity of the spectral modes. The density operator of the system is $\rho = \otimes_{m=1}^{M} \rho_m$. And the output state going through the linear interferometer in the form of covariance matrix is $V_{out} = \oplus_{m=1}^{M} V_m$. Each V_m can be independently calculated by the evolution of the light sources in the circuit.

The original GBS outcome probability of threshold detection is given by *Torontonian*

$$p(S) = \frac{\mathrm{Tor}[O_{(S)}]}{\sqrt{\det V}} \tag{S20}$$

where

$$\mathrm{Tor}(A) = \sum_{Z \in P([N])} (-1)^{|Z|} \frac{1}{\sqrt{\det[1 - A_{(Z)}]}} \tag{S21}$$

in which O is a matrix derived from the covariance matrix

$$O = 1 - V^{-1} \tag{S22}$$

the subscript (S) of O means a submatrix by indexing elements of the detection outcome S, $P([N])$

is the power set of set $[N] = \{1, 2, \cdots, N\}$, and $|Z|$ denotes the cardinality of the set. This summation is over 2^N determinants, so it has the time complexity of $O(N^3 2^N)$.

In our model, the threshold detection probability can be calculated by the modified Torontonian, mTor(O). Firstly notice that the probability of detecting a vacuum state $\Pi_0 = |0\rangle\langle 0|$ is the multiplication of the probability that detect the vacuum state in each independent mode, and $P(\Pi_1) = 1 - P(\Pi_0)$. Employing the phase space representation of the output states and an inclusive-exclusive formula, the output probability can be written as

$$p(S) = \frac{\mathrm{mTor}[O_{1_{(S)}}, O_{2_{(S)}}, \cdots, O_{M_{(S)}}]}{\prod_{m=1}^{M} \sqrt{\det V_m}} \tag{S23}$$

where

$$\mathrm{mTor}(A_1, A_2, \cdots, A_M) := \sum_{Z \in P([N])} (-1)^{|Z|} \prod_{m=1}^{M} \frac{1}{\sqrt{\det[1 - A_{m_{(Z)}}]}} \tag{S24}$$

in which O_m is a matrix derived from the covariance matrix independently

$$O_m = 1 - V_m^{-1} \tag{S25}$$

We observe that this equation is a multiplication of M single modes, so the time complexity is multiplied by a factor M.

5.2 Validation results with partial photon distinguishability incorporated

In this section we repeat the validation pipeline in the main text, with the more physical model incorporating partial photon distinguishability. That's to say, The ground-truth is modeled as a lossy and partial distinguishable GBS, which enables better agreement between the experimental observation and the theoretical model.

Fig. S20 and Fig. S21 respectively repeat the validation pipeline of Fig. 2 and Fig. 3 in the main text , with the partial photon distinguishability corrected ground-truth. All the conclusions drawn in the main text remain consistent.

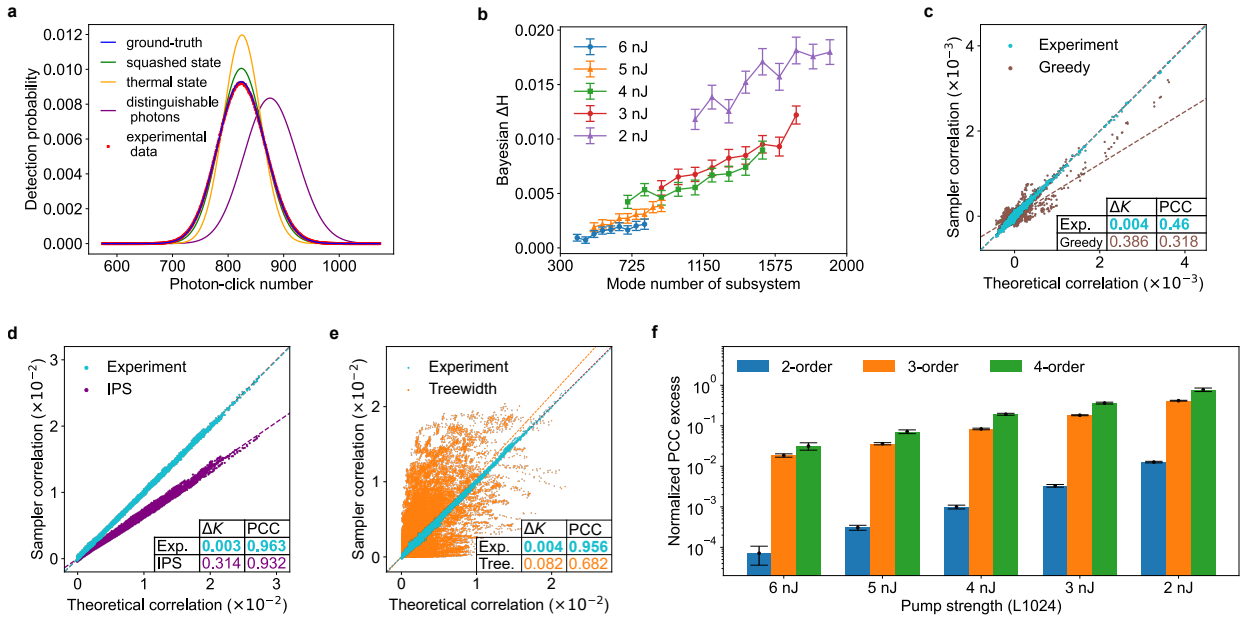


Figure S20: **a**, A comparison of the photon number distribution of the experimental results, the ground-truth theory and classical mockups, including the squashed state, the thermal state, distinguishable photons, for the L1024 group under a middle pump level at 3 nJ. **b**, Bayesian tests against the squashed state, which are computed under certain photon-click number in terms of the size of considered subsystem (25 click for 5,6 nJ and 15 click for 2,3,4 nJ). Experimental data from the S64 group are used for the Bayesian test. The subsystems for validation are randomly selected. Error bars indicate the standard error obtained from 100 groups of resampling. **c**, Scatter plot of the third-order correlation function of the greedy sampler (reproducing up to 2-order) in comparison with the experimental results. The first 4096 modes are selected for processing and each consecutive 8 modes are binned together to relieve the huge cost for counting sample statistics. **d**,**e**, Scatter plot of the two-order correlation function of the IPS sampler **d** and the treewidth sampler **e** in comparison with the experimental results. 10 million samples of the L1024 (6 nJ) dataset are used in **c**,**d**,**e** to evaluate the results. **f**, Comparison of Pearson correlation coefficient (PCC) of the 2,3,4-order correlation function between the experimental data and the squashed sampler. A normalized excess quantity $(\mathrm{PCC}_{\mathrm{exp}} - \mathrm{PCC}_{\mathrm{squash}})/\mathrm{PCC}_{\mathrm{squash}}$ is defined and plotted as the y-axis to show the comparison results more visibly. Error bars are the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

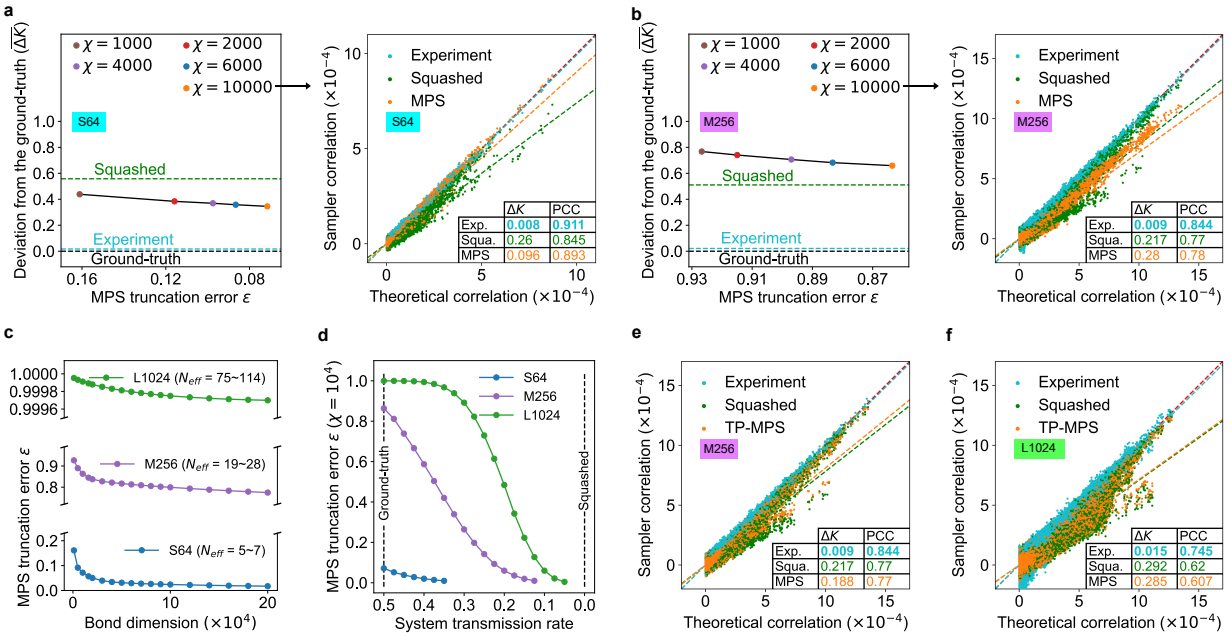


Figure S21: **a**,**b** MPS truncation error and ΔK of the two-order correlation function counted from MPS samples, for the S64 experiment **a** and the M256 experiment **b**, under progressively increasing bond dimension χ. **c**, Truncation error of MPS in approximating the ground-truth of the S64, M256 and L1024 experiments, in terms of the chosen bond dimension up to χ=2 × 10^5. The effective squeezed photon number N_{eff} of the three groups is also labeled respectively. **d**, Computational hardness analysis via parameter interpolation. With a fixed bond dimension $\chi = 10^4$, we plot the MPS truncation error for a family of states constructed by scaling down the transmission rate (x-axis) while keeping the mean photon number fixed. This creates a trajectory from the ground-truth (left dashed line) to the squashed state (asymptotic limit under excess loss, right dashed line). We select the transmission rate where the curve drops to $\epsilon \approx 0.01$ to define the "tractable proxy MPS" used in e and f. **e**,**f**, Scatter plot of the two-order correlation function counted from the experimental samples, squashed state samples, and the "tractable proxy MPS (TP-MPS)" (χ=10^4) samples, for the M256 experiment **e** and the L1024 experiment **f**. 5.5 million samples are used in **a**,**b**, and **e**. 10 million samples are used in **f**.

6 Mode-binned click probability of threshold detector

We apply a output mode binning method in data processing, which is particularly useful when considering high-order correlators among the system.

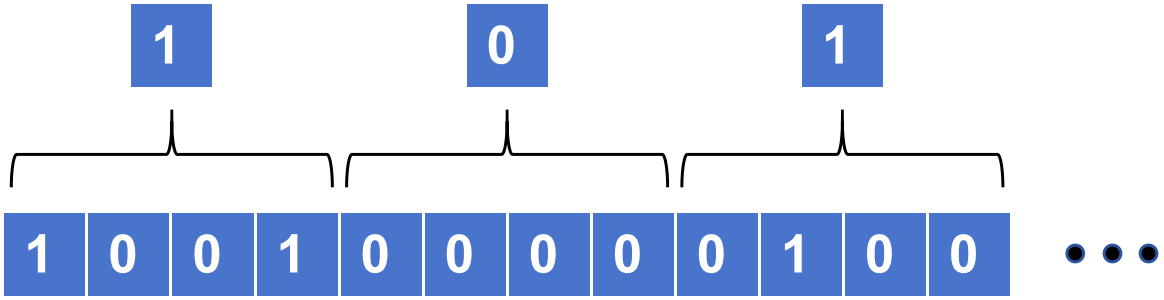


Figure S22: Multiple physical output modes are aggregated into a single binned port – a detection event in any of the constituent physical modes registers as a click on the binned port.

Assuming the GBS output is measured by M threshold detectors $[M] = \{1, 2, \cdots, M\}$. The output modes are then grouped into a set partition $\{b_1, b_2, \cdots, b_K\}$. Each element b_i is a subset of $[M]$ with K_i elements, $b_i = \{b_{i_1}, b_{i_2}, \cdots, b_{i_{K_i}}\}$, forming a grouped detector. Any click in these binned output modes of detectors will sign a click on the grouped detector (Fig. S22). The POVM of the grouped detector b_i is

$$\Pi_0^{(b_i)} = \prod_{n \in b_i} |0_n\rangle \langle 0_n|$$

$$\Pi_1^{(b_i)} = 1 - \Pi_0^{(b_i)}$$

The Glauber-P functions on the binned output port are expressed as

$$P_0^{(b_i)}(\alpha) = \prod_{n \in b_i} \delta^{(2)}(\alpha_n)$$

$$P_1^{(b_i)}(\alpha) = \frac{1}{\pi} - P_0^{b_i}(\alpha)$$

Knowing that the Husimi covariance matrix of the light is σ_Q, the Q function is

$$Q(\alpha) = \frac{\exp(-\frac{1}{2}\alpha\sigma_Q^{-1}\alpha^\dagger)}{\pi^M \sqrt{\det \sigma_Q}}$$

If a mode-binned n click event is indexed by $S = \{S_1, S_2, \cdots, S_n\}$ and define

$$\bigcup(S) = \bigcup\{b_i | i \in S\}$$

Then the probability of this event is given by

$$\begin{aligned} p(S) &= \pi^M \int \mathrm{d}^2\alpha \left(\prod_{i\in S} P_1^{(b_i)}(\alpha_i)\right)\left(\prod_{j\in[k]-S} P_0^{(b_j)}(\alpha_j)\right) Q(\alpha) \\ &= \frac{1}{\sqrt{\det\sigma_Q}} \int \mathrm{d}^2\alpha \left[\prod_{i\in S}\left(\frac{1}{\pi} - P_0^{(b_i)}(\alpha_i)\right)\right]\left(\prod_{j\in[k]-S} P_0^{(b_j)}(\alpha_j)\right) \exp\left(-\frac{1}{2}\alpha\sigma_Q^{-1}\alpha^\dagger\right) \end{aligned}$$

Since we have the δ function integral

$$\int \mathrm{d}^2\alpha_l \delta^{(2)}(\alpha_l) \exp\left(-\frac{1}{2}\alpha V \alpha^\dagger\right) = \exp\left(-\frac{1}{2}\alpha' V' \alpha'^\dagger\right)$$

where α' and V' represent vector and matrix which has done a remove of rows and columns at index l and $l + M$ from α and V. Expand the product term in the probability expression and do a Gaussian integral, we get

$$p(S) = \frac{1}{\sqrt{\det\sigma_Q}} \sum_{Z\in\{\bigcup(S_i)|S_i\in P(S)\}} \frac{(-1)^{|Z|}}{\sqrt{\det V_{(Z)}}}$$

where $P(S)$ is the powerset of S, $V = \sigma_Q^{-1}$, and $V_{(Z)}$ is the matrix with rows and columns indexed by the set $Z \bigcup\{z + M | z \in Z\}$ in V.

7 Matrix product state method

7.1 Decomposition of the output state of GBS

The matrix product state (MPS) method [4] begins by decomposing the covariance matrix $\boldsymbol{V}$ of the output Gaussian state into two components: $\boldsymbol{V} = \boldsymbol{V}_p + \boldsymbol{W}$. Here, $\boldsymbol{V}_p$ represents the covariance matrix of a pure Gaussian state, and $\boldsymbol{W}$ is a positive semidefinite matrix ($\boldsymbol{W} \succcurlyeq 0$). This decomposition is performed using semidefinite programming (SDP) under specified constraints

$$\min_{\boldsymbol{V}_p} \mathrm{Tr}[\boldsymbol{V}_p] \text{ with } \boldsymbol{V} - \boldsymbol{V}_p \succcurlyeq 0, \boldsymbol{V}_p \succcurlyeq \mathrm{i}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes \mathbf{1}_m, \tag{S26}$$

where m is the number of output modes.

To address this problem, we consider methods such as interior-point and first-order approaches. However, computational constraints, such as several hours and 128 GB of memory on a personal computer, limit the problem size. Interior-point methods are significantly slower than first-order methods for large-scale problems, and most first-order SDP solvers lack MPI support, rendering them unsuitable for supercomputer clusters. We employ the SCS solver [5] with the parameter use_indirect=True for $m < 2000$. For $m > 2000$, an analytical method is used, as described below.

Analytic method to decompose the output state We denote the transformation matrix as $\boldsymbol{M}$ and the squeezing parameter as $\vec{r}$. The covariance matrix of the output state of GBS is

$$\boldsymbol{V} = \boldsymbol{T}\,\mathrm{diag}(\mathrm{e}^{2\vec{r}}, \mathrm{e}^{-2\vec{r}})\boldsymbol{T}^T + \boldsymbol{I} - \boldsymbol{T}\boldsymbol{T}^T,$$

where

$$\boldsymbol{T} = \begin{pmatrix} \mathrm{Real}(\boldsymbol{M}) & -\mathrm{Imag}(\boldsymbol{M}) \\ \mathrm{Imag}(\boldsymbol{M}) & \mathrm{Real}(\boldsymbol{M}) \end{pmatrix}.$$

If $\boldsymbol{M}$ can be written as $\boldsymbol{M} = \boldsymbol{U}\,\mathrm{Diag}(\sqrt{\vec{\eta}})$, where $\boldsymbol{U}$ is a unitary matrix, $\boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{I}$, and $\vec{\eta}$ is the transmittance of the input entries, we can get

$$\boldsymbol{T} = \begin{pmatrix} \mathrm{Real}(\boldsymbol{U}) & -\mathrm{Imag}(\boldsymbol{U}) \\ \mathrm{Imag}(\boldsymbol{U}) & \mathrm{Real}(\boldsymbol{U}) \end{pmatrix} \mathrm{diag}(\sqrt{\vec{\eta}}, \sqrt{\vec{\eta}}).$$

Define $\boldsymbol{A}$ as

$$\boldsymbol{A} = \begin{pmatrix} \mathrm{Real}(\boldsymbol{U}) & -\mathrm{Imag}(\boldsymbol{U}) \\ \mathrm{Imag}(\boldsymbol{U}) & \mathrm{Real}(\boldsymbol{U}) \end{pmatrix}.$$

Then $\boldsymbol{T}$ can be rewritten as

$$\boldsymbol{T} = \boldsymbol{A}\,\mathrm{diag}(\sqrt{\vec{\eta}}, \sqrt{\vec{\eta}})$$

The covariance matrix of the output state is

$$\boldsymbol{V} = \boldsymbol{A}\,\mathrm{diag}(\vec{\eta}\mathrm{e}^{2\vec{r}} + 1 - \vec{\eta}, \vec{\eta}\mathrm{e}^{-2\vec{r}} + 1 - \vec{\eta})\boldsymbol{A}^T.$$

So the pure state part of the output state is

$$\boldsymbol{V}_p = \boldsymbol{A}\,\mathrm{diag}(\mathrm{e}^{2\vec{s}}, \mathrm{e}^{-2\vec{s}})\boldsymbol{A}^T,$$

where

$$\mathrm{e}^{-2\vec{s}} = \vec{\eta}\mathrm{e}^{-2\vec{r}} + 1 - \vec{\eta}.$$

$\vec{s}$ and $\boldsymbol{U}$ can be easily calculated from $\boldsymbol{M}$. However, generally the constraint $\boldsymbol{U}\boldsymbol{U}^\dagger = \boldsymbol{I}$ is not strictly satisfied. So we first use $\boldsymbol{U}$ and $\vec{s}$ to calculate $\boldsymbol{V}'$ and then perform the Williamson decomposition on $\boldsymbol{V}'$ to get $\boldsymbol{D}, \boldsymbol{S}$. Note that $\boldsymbol{D} = \boldsymbol{I}$ if $\boldsymbol{V}'$ is a pure state. For $\boldsymbol{V}'$ of experimental result, $\boldsymbol{D} \approx \boldsymbol{I}$. It means that $\boldsymbol{V}'$ is not a pure state while $\boldsymbol{V}_p = \boldsymbol{S}\boldsymbol{S}^T$ is a pure state, and $\boldsymbol{V}_p$ is a rather good approximation for the target state. Nevertheless, $\boldsymbol{V}_p = \boldsymbol{S}\boldsymbol{S}^T$ may not satisfy the constraint Eq. (S26), so we need to scale and search $\vec{s}$ globally to make $\boldsymbol{V}_p$ satisfy it. The algorithm is shown in Alg.1.

Algorithm 1: The analytic method to decompose the output state of GBS.

Data: $\vec{r}$,$\boldsymbol{T}$

Result: $\boldsymbol{V}_p$,$\boldsymbol{W}$

$\boldsymbol{V} = \boldsymbol{T}\,\mathrm{diag}(\mathrm{e}^{2\vec{r}}, \mathrm{e}^{-2\vec{r}})\boldsymbol{T}^T + \boldsymbol{I} - \boldsymbol{T}\boldsymbol{T}^T$;

$\vec{\eta} = \mathrm{diag}(\boldsymbol{T}^\dagger\boldsymbol{T})$;

$\vec{s} = -\ln(\vec{\eta}\mathrm{e}^{-2\vec{r}} + 1 - \vec{\eta})/2$;

$\boldsymbol{U} = \boldsymbol{T}/\sqrt{\vec{\eta}}$;

$\boldsymbol{V'} = \boldsymbol{U}\,\mathrm{diag}(\mathrm{e}^{2\vec{s}}, \mathrm{e}^{-2\vec{s}})\boldsymbol{U}^T + \boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^T$;

$\boldsymbol{D}, \boldsymbol{S} = \mathrm{Williamson}(\boldsymbol{V}')$;

$\boldsymbol{V}_p = \boldsymbol{S}\boldsymbol{S}^T$;

$\boldsymbol{W} = \boldsymbol{V} - \boldsymbol{V}_p$;

while $\boldsymbol{W} \succcurlyeq 0$ *is not satisfied* **do**

  scale $\vec{s}$, $\vec{s} = \vec{s} \times 1.01$ for example;

  $\boldsymbol{V'} = \boldsymbol{U}\,\mathrm{diag}(\mathrm{e}^{2\vec{s}}, \mathrm{e}^{-2\vec{s}})\boldsymbol{U}^T + \boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}^T$;

  $\boldsymbol{D}, \boldsymbol{S} = \mathrm{Williamson}(\boldsymbol{V}')$;

  $\boldsymbol{V}_p = \boldsymbol{S}\boldsymbol{S}^T$;

  $\boldsymbol{W} = \boldsymbol{V} - \boldsymbol{V}_p$;

end

return $\boldsymbol{V}_p$,$\boldsymbol{W}$;

7.2 MPS truncation error

An MPS of a given pure quantum state is written as follows:

$$|\psi\rangle \approx \sum_{n_1,\cdots,n_M=0}^{d-1} \sum_{\alpha_1,\cdots,\alpha_{M-1}=0}^{\chi-1} \Gamma_{\alpha_1}^{[1]n_1} \lambda_{\alpha_1}^{[1]} \Gamma_{\alpha_1\alpha_2}^{[2]n_2} \lambda_{\alpha_2}^{[2]} \cdots \lambda_{\alpha_{M-1}}^{[M-1]} \Gamma_{\alpha_{M-1}}^{[M]n_M} |n_1, \cdots, n_M\rangle, \tag{S27}$$

where d is the dimension of a local Hilbert space and χ is the bond dimension.

Following Ref[4], the truncation error is defined as the sum of the squared lost singular values for the bipartition at the centre, i.e.,

$$\epsilon = 1 - \sum_{\alpha_{M/2}=0}^{\chi-1} \left(\lambda_{\alpha_{M/2}}^{[M/2]}\right)^2. \tag{S28}$$

$M/2$ means that the mode number of one part of the bipartitions is $M/2$. The truncation error calculated at this partition is the largest truncation error.

Here we have a more physical explanation. The Williamson decomposition can decompose every positive definite real matrix $\boldsymbol{V} \in \mathbb{R}^{2N\times 2N}$ s.t.

$$\boldsymbol{V} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^T, \tag{S29}$$

where $\boldsymbol{S}$ is a symplectic matrix and $\boldsymbol{D}$ is a diagonal matrix $\boldsymbol{D} = \mathrm{diag}(\nu_1, \ldots, \nu_\mathrm{N}, \nu_1, \ldots, \nu_\mathrm{N})$, and ν_i are the eigenvalues of $|\mathrm{i}\Omega\mathrm{V}|$, where $||$ represents the element-wise absolute value [6].

It allows an arbitrary Gaussian covariance matrix to be decomposed into a symplectic transformation acting on the state described by the diagonal matrix $\boldsymbol{D}$.

The matrix $\boldsymbol{D}$ can always be decomposed further into a set of thermal states with mean photon number given by

$$\bar{n}_i = \frac{1}{\hbar}\nu_i - \frac{1}{2}, i = 1, \ldots, N. \tag{S30}$$

If $\boldsymbol{V}$ represents a pure state, $\boldsymbol{D}$ will be $\frac{1}{2}\hbar\boldsymbol{I}_{2N}$. However, when we bipartition a pure state into two parts, A and B, generally, they will be two mixed states, i.e., $\boldsymbol{D}$ represents a thermal state. Then, we can write it in Fock-state space and select the most important χ partitions that have the highest probability. The probability of each partition can be calculated according to the photon number distribution of the thermal state.

$$\rho(\bar{n}) = \sum_{n=0}^{\infty} \frac{\bar{n}^n}{(\bar{n}+1)^{n+1}} |n\rangle\langle n| = \sum_{n=0}^{\infty} p_{\bar{n}}^n |n\rangle\langle n|. \tag{S31}$$

Here

$$p_{\bar{n}}^{n} = \frac{\bar{n}^{n}}{(\bar{n}+1)^{n+1}}. \tag{S32}$$

$$\begin{aligned}\rho(\boldsymbol{D}) &= \sum_{n_1,\ldots,n_{M/2}} p_{\bar{n}_1}^{n_1} \cdots p_{\bar{n}_{M/2}}^{n_{M/2}} \left|n_1,\ldots,n_{M/2}\right\rangle \left\langle n_1,\ldots,n_{M/2}\right| \\ &\approx \sum_{i}^{\chi} p_i \left|n_1,\ldots,n_{M/2}\right\rangle \left\langle n_1,\ldots,n_{M/2}\right|\end{aligned} \tag{S33}$$

The truncation error can be calculated by

$$\epsilon = 1 - \sum_{i}^{\chi} p_i. \tag{S34}$$

7.3 Estimate the lower bound of required bond dimension size to simulate the experiments

The high memory cost of processing large tensors precludes direct evaluation of the truncation error for the large bond dimensions required to simulate our experiments. Instead, we estimate the necessary bond dimension by extrapolating from data obtained at tractable parameter scales, analyzing the relationship between the truncation error ε and the bond dimension χ.

First, we need to determine the function form to fit the calculable data points and perform the extrapolation. In [4], it shows that for a fixed-size circuit, the bond dimension scales as $\chi = O(\text{polylog}(1/\epsilon))$ in the TVD ϵ. Here we consider the single-term polynomial hypotheses:

$$\chi = Ax^{n} \tag{S35}$$

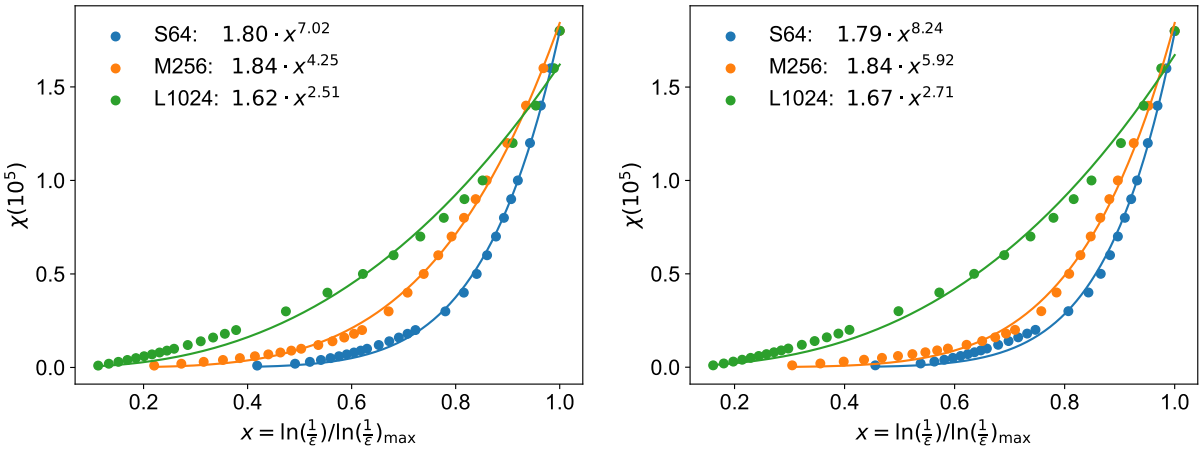
where

$$x = \frac{\ln(1/\varepsilon)}{\ln(1/\varepsilon)_{\max}}$$

Here $\ln(1/\varepsilon)_{\max}$ is the max $\ln(1/\varepsilon)$ among the obtained data points and it is introduced for normalize and better fitting performance. We also considered other more complicated polynomial models which were found to be not numerically stable to fit the data points, so the simplest single-term model was chosen.

The fitting results are shown in Fig. S23.

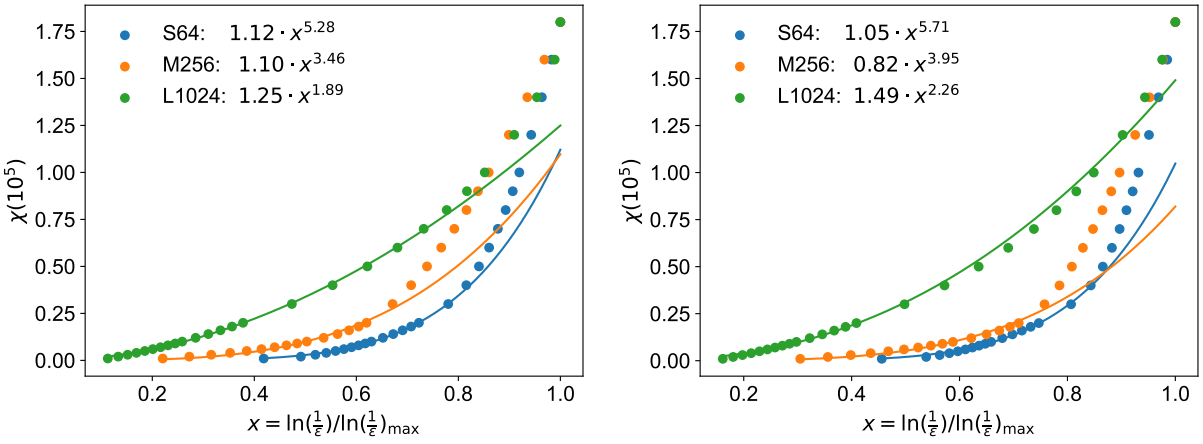
We choose the truncation error $\varepsilon = 0.05$ as the target value to determine the needed size of bond dimension, the results are listed in Table S1.



(a) Dataset (6nJ)

(b) Dataset (2nJ)

Figure S23: Fitted relation between the bond dimension χ and the truncation error ε.



(a) Dataset (6 nJ)

(b) Dataset (2 nJ)

Figure S24: The relation between bond dimension χ and truncation error ε. The points are the data we used to fit and the lines are the fitting result of the model EQ(S35) for only the first 16 points.

Then we prove that we do get the lower bound from the extrapolation method, i.e. we do not over estimate the complexity. To verify this, we pick out the last several data points from the fitting set to serve as the validation set, and then perform the fitting and extrapolation using the left data points. As shown in Fig. S24, the extrapolated value of bond dimension is clearly smaller than the real value, which indicates that our procedure here indeed gives the lower bound of the needed size of bond dimension.

Data set	$\ln(1/\varepsilon)_{\max}$	χ	$A(10^5)$	n
S64 (6 nJ)	2.68	3.88×10^5	1.80	7.02
M256 (6 nJ)	5.06×10^{-2}	6.29×10^{12}	1.84	4.25
L1024 (6 nJ)	6.87×10^{-7}	7.98×10^{21}	1.62	2.51
S64 (2 nJ)	4.01	1.64×10^4	1.79	8.24
M256 (2 nJ)	2.50×10^{-1}	4.56×10^{11}	1.84	5.92
L1024 (2 nJ)	2.95×10^{-4}	1.12×10^{16}	1.67	2.71

Table S1: The bond dimension χ at target truncation error $\varepsilon = 0.05$ and the fitting parameters for different dataset. Note that $\ln(1/0.05) = 2.996$, most χs are the result of extrapolation and they are mainly determined by the leading term.

7.4 Running time on the supercomputer

The complexity of MPS construction is given by

$$T_{\text{MPS}} = O(Md\chi^2 2^{\frac{N_{\text{eff}}}{2}}) \tag{S36}$$

To estimate the needed cost to simulate the experiments on the supercomputer, we first record the running time under moderate parameter level [7] and directly scale the needed cost for huge parameter size based on the complexity shown above, which implicitly assumes unlimited memory. We only consider the computational cost of MPS construction, though the sampling cost is also massive under large bond dimension.

The dataset of M256 (2 nJ) is used as the estimation base of moderate parameter size, with bond dimension $\chi_0 = 10000$, effective squeezed photon number $N_{\text{eff}_0} = 19.16$, mode number $M_0 = 5104$, dimension of a local Hilbert space $d_0 = 3$. The running time on an Nvidia A100 is 6 hours if only the float point operation time is included and no hard disk IO time is considered. The FP32 performance of A100 is 19.5 TFLOPS, and the Rmax of the fastest supercomputer El Capitan is 1.742 EFLOPS, so we can figure out that the time needed by El Capitan is $t_0 = \frac{19.5\times3600\times6}{1.742\times10^6} = 0.2418\,\text{s}$.

The complexity of obtaining all the matrix elements for MPS is $O(Md\chi^2 \times$ (hafnian of $\sum n_1, n_2$)). Here we replace hafnian of $\sum n_1, n_2$ with the $2^{N_{\text{eff}}/2}$, N_{eff} is the effective squeezed

photon number. So the time needed by El Capitan is

$$t = t_0 \times \frac{M d \chi^2 2^{N_{\text{eff}}/2}}{M_0 d_0 \chi_0^2 2^{N_{\text{eff}_0}/2}} \tag{S37}$$

The time needed by El Capitan for different data sets are shown in Table S2.

Dataset		M	χ	N_{eff}	T
	2 nJ		1.64×10^4	4.47	3.40 ms
	3 nJ		7.14×10^4	5.46	90.8 ms
dataset A-S64	4 nJ	4336	1.81×10^5	6.21	757 ms
	5 nJ		2.93×10^5	6.64	2.30 s
	6 nJ		3.88×10^5	6.92	4.44 s
dataset B-S64	7 nJ	4336	5.40×10^5	7.17	9.39 s
	2 nJ		4.56×10^{11}	18.02	1.07×10^7 y
	3 nJ		1.07×10^{12}	21.92	2.28×10^8 y
dataset A-M256	4 nJ	5104	2.63×10^{12}	24.89	3.86×10^9 y
	5 nJ		4.00×10^{12}	26.59	1.61×10^{10} y
	6 nJ		6.29×10^{12}	27.69	5.82×10^{10} y
dataset B-M256	7 nJ	5104	2.20×10^{13}	28.73	1.02×10^{12} y
	2 nJ		1.12×10^{16}	74.19	2.95×10^{24} y
	3 nJ		3.14×10^{18}	90.35	6.28×10^{31} y
dataset A-L1024	4 nJ	8176	4.99×10^{20}	102.23	9.74×10^{37} y
	5 nJ		1.18×10^{21}	109.12	5.93×10^{39} y
	6 nJ		$\mathbf{7.98 \times 10^{21}}$	**113.54**	$\mathbf{1.26 \times 10^{42}}$ **y**
dataset B-L1024	7 nJ	8176	3.71×10^{19}	110.82	1.06×10^{37} y

Table S2: The time needed by El Capitan without hard disk IO for different dataset. M is the mode number, χ is the bond dimension, N_{eff} is the effective photon number and T is the time needed by El Capitan. The time is calculated using the bond dimension χ at target truncation error $\varepsilon = 0.05$. Local Hilbert space dimension d is set as 3. The dataset used to claim QCA in the main text is emphasized with bold fonts.

It is worth noting that the effective squeezed photon number of data set M256(2 nJ) in the table is a little different from that we used as reference, because when do the decomposition that $V = V_p + W$, we scale $\vec{s}$ a little to make V_p satisfy the constraint Eq. (S26). So the effective squeezed photon number used to do the real MPS decomposition will be a little larger than the one we used to calculate the time. It means that the time needed by the supercomputer would be a little larger than the amount we calculated, which ensures that the running time we present here is a lower bound. Another important thing is that the squeezing parameters that we use to calculate the effective squeezed photon numbers are the main part of the squeezing parameters in the partial photon distinguishability model. It ensures that the running time we present here is a lower bound again.

8 Systematic validation results for all datasets

In this section, we provide a comprehensive and systematic validation across all experimental datasets, supplementing the representative results presented in the main text.

While the main text highlights the Bayesian test results for the S64 dataset across all pump settings (from 2 nJ to 6 nJ), here we expand this analysis to larger datasets. As discussed in the Methods, computing the exact probability required for Bayesian becomes computationally intractable for all pump settings in the M256 and L1024 datasets, particularly at higher pump powers where thousands of photons are involved.

Fig. S25 displays the Bayesian test results for the S64 (2–6 nJ), M256 (2–4 nJ), and L1024 (2–3 nJ) datasets as a function of the validation subsystem size. The trend consistently aligns with the observations in the main text: the Bayesian confidence in the experimental data increases as the subsystem expands, demonstrating that the Bayesian confidence becomes more pronounced when more global information is incorporated.

To provide a complete picture across all parameter regimes (S64, M256, L1024 under all pump settings), we use the computationally efficient metrics: the correlation function benchmark (second-, third-, and fourth-order, both with and without mode binning) and the photon number distribution benchmark (Tables S3–S17). Unlike Bayesian or XEB tests which requires the calculation of individual click probability and quickly becomes computationally prohibitive for large datasets, the two validation benchmarks can be applied to all datasets because of the relatively low calculation cost.

Having conclusively ruled out weaker mockups (such as the IPS and treewidth samplers) in the main text, here we focus our systematic analysis on the most stringent classical competitors: the squashed state and the MPS samplers. The results comprehensively covers all experiments, S64 (2,6 nJ), M256 (2,6nJ), L1024 (2-6 nJ) for MPS sampler and S64/M256/L1024 (2-6 nJ) for squashed state sampler. The bond dimension of TP-MPS sampler is set to $\chi = 10000$.

For correlation function benchmark, Pearson correlation coefficient (PCC) and the slope of linear fit (ΔK) are used for statistical metrics; for photon number distribution, the total variation distance (TVD) and the relative deviation of distribution width ($\Delta\Sigma$) are used for statistical metrics. All the metrics are evaluated between the sampler statistics and the ground-truth.

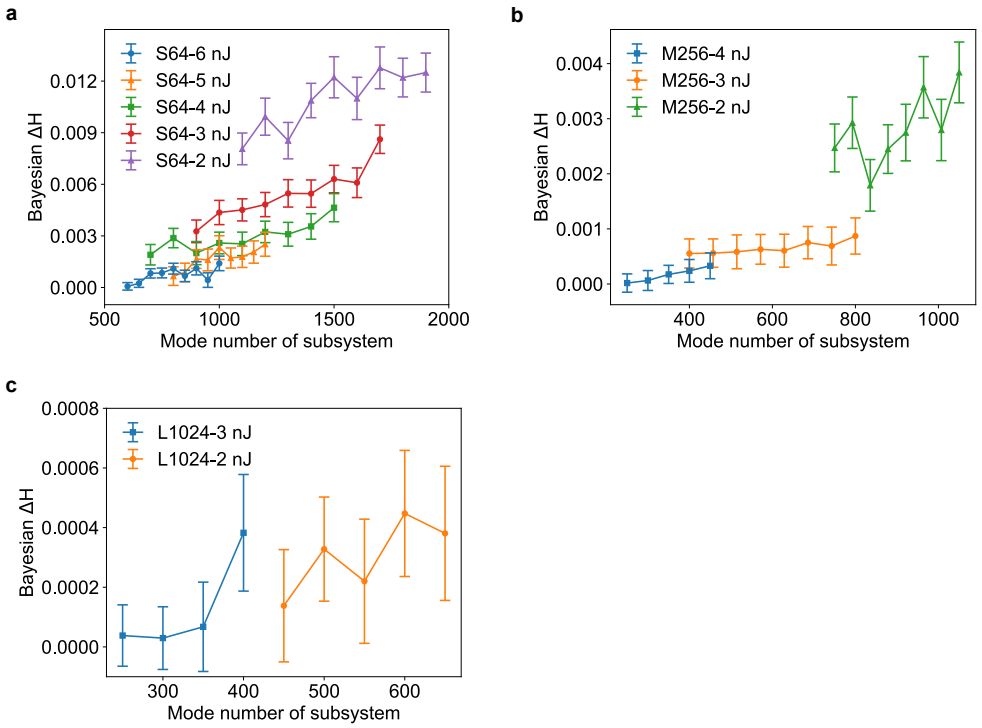


Figure S25: Bayesian tests against the squashed state, computed for specific photon-click number as a function of subsystem size . Experimental data from the S64 group (**a**), M256 group (**b**), and L1024 group (**c**) are used for the Bayesian test. For S64, the photon click number is 30 for 5,6 nJ and 15 for 2,3,4 nJ. For M256, the photon click number is 22. For L1024, the photon click number is 26. The subsystems for validation are randomly selected. Error bars indicate the standard error obtained from 100 groups of resampling.

With thoroughly examination over these test results, we find the experiment consistently outperforms both classical mockups across almost all metrics and datasets. The only nuanced regime occurs in the unbinned, low-order (2 and 3) correlation analysis of the L1024 (6 nJ) dataset, where the experimental sampler's performance appears marginally lower than the mockups. However, when accounting for the calibration error of the transformation matrix, this deviation falls well within one standard deviation. Notably, as the correlation order increases to the 4-order, the experimental data once again surpasses the classical competitors. This indicates that higher-order correlations capture more complete system information, a regime where the classical mockups begin to fail. This is because any 2(3)-order correlation corresponds to a mere $\frac{2(3)}{8176} = 0.024(0.037)\%$ fraction of the L1024 system, truncating more than 99.97(6)% modes out of inspection, therefore results in

a tremendous effective loss in system efficiency during validation.

Directly evaluating 5-order correlations is statistically infeasible due to vanishing probabilities. We instead employ the mode-binning technique (introduced in Section 6). Mode-binning not only amplifies the correlation signals but also naturally incorporates broader system information by aggregating multiple modes. It is an essential physical tool to mitigate the severe truncation problem mentioned above. As shown in the tables, the performance of the experimental sampler against the classical mockups is significantly enhanced under binning.

Furthermore, for the photon number distribution metric which conveys the global properties, the experimental sampler substantially outperforms the classical models.

Together, these systematic lines of evidence strongly reinforce our core conclusion: classical mockups (squashed state and MPS) progressively deteriorate and perform worse than the experiment as more complete global information is evaluated. This comprehensive picture solidifies the evidence in the non-simulability of our quantum processor against state-of-the-art classical algorithms.

Table S18 - table S20 describes how the evaluated correlators are selected for all cases. 10 million samples are used for the evaluation of L1024 group and 5.5 million samples for M256/S64 group.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
L1024 (6 nJ)		1024	8176	2598	2207	113.5	
Correlation no binning		2-order		3-order		4-order	
		PCC	ΔK	PCC	ΔK	**PCC**	**ΔK**
L1024	exp.	0.9581(3)	0.045	0.0464(3)	0.079	**0.00145(4)**	**0.097**
(6 nJ)	squashed	0.9606(3)	0.032	0.0462(3)	0.081	0.00140(4)	0.13
	TP-MPS	0.9605(3)	0.030	0.0466(3)	0.074	0.00137(4)	0.15
Correlation 8-ports binned		2-order		3-order		4-order	
		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
L1024	exp.	**0.9958(3)**	0.047	**0.199(5)**	**0.057**	**0.0221(2)**	**0.087**
(6 nJ)	squashed	0.9953(3)	0.019	0.196(4)	0.079	0.0214(2)	0.114
	TP-MPS	0.9958(3)	0.017	0.195(5)	0.074	0.0216(2)	0.105
Correlation 16-ports binned		2-order		3-order		4-order	
		PCC	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
L1024	exp.	**0.9953(3)**	**0.035**	**0.280(5)**	**0.082**	**0.0376(9)**	**0.090**
(6 nJ)	squashed	0.9951(3)	0.068	0.270(5)	0.11	0.0337(9)	0.18
	TP-MPS	0.9953(3)	0.063	0.273(5)	0.10	0.0345(9)	0.15
Photon number distribution		**TVD**			**$\Delta\Sigma$**		
L1024	exp.	**0.005**			**0.009**		
(6 nJ)	squashed	0.014			0.028		
	TP-MPS	0.013			0.025		

Table S3: Validation results summary of the L1024 (6 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
L1024 (5 nJ)		1024	8176	2119	1749	109.1	
Correlation no binning		2-order		3-order		4-order	
		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
L1024	exp.	**0.9396(3)**	0.051	**0.0374(2)**	**0.084**	**0.00120(4)**	**0.069**
(5 nJ)	squashed	0.9394(3)	0.051	0.0361(2)	0.12	0.00104(4)	0.19
	TP-MPS	0.9395(2)	0.047	0.0362(2)	0.11	0.00103(4)	0.19
Correlation 8-ports binned		2-order		3-order		4-order	
		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
L1024	exp.	**0.9940(4)**	0.056	**0.129(3)**	**0.069**	**0.0148(1)**	**0.096**
(5 nJ)	squashed	0.9928(4)	0.030	0.124(3)	0.11	0.0138(1)	0.16
	TP-MPS	0.9928(4)	0.027	0.124(3)	0.096	0.0135(1)	0.15
Correlation 16-ports binned		2-order		3-order		4-order	
		PCC	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
L1024	exp.	**0.9942(3)**	**0.039**	**0.204(3)**	**0.079**	**0.0230(5)**	**0.10**
(5 nJ)	squashed	0.9928(3)	0.092	0.188(3)	0.15	0.0201(5)	0.21
	TP-MPS	0.9931(3)	0.084	0.189(3)	0.13	0.0203(5)	0.20
Photon number distribution		**TVD**			**$\Delta\Sigma$**		
L1024	exp.	**0.003**			**0.003**		
(5 nJ)	squashed	0.018			0.037		
	TP-MPS	0.016			0.031		

Table S4: Validation results summary of the L1024 (5 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
L1024 (4 nJ)		1024	8176	1603	1287	102.2	
Correlation no binning		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9090(4)**	**0.059**	**0.0288(1)**	**0.098**	**0.00085(4)**	**0.13**
(4 nJ)	squashed	0.9035(4)	0.078	0.0263(1)	0.18	0.00073(4)	0.25
	TP-MPS	0.9050(4)	0.084	0.0266(1)	0.17	0.00073(4)	0.25
Correlation 8-ports binned		2-order		3-order		4-order	
		PCC	Δ*K*	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9925(5)**	0.060	**0.078(1)**	**0.079**	**0.0099(1)**	**0.087**
(4 nJ)	squashed	0.9907(6)	0.053	0.072(1)	0.15	0.0083(1)	0.114
	TP-MPS	0.9909(6)	0.047	0.072(1)	0.14	0.0086(1)	0.105
Correlation 16-ports binned		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9923(4)**	**0.052**	**0.124(2)**	**0.12**	**0.0125(3)**	**0.096**
(4 nJ)	squashed	0.9882(4)	0.13	0.104(1)	0.24	0.0105(3)	0.24
	TP-MPS	0.9896(4)	0.11	0.106(2)	0.22	0.0104(3)	0.25
Photon number distribution		**TVD**			**ΔΣ**		
L1024	exp.	**0.002**			**0.0017**		
(4 nJ)	squashed	0.024			0.048		
	TP-MPS	0.024			0.047		

Table S5: Validation results summary of the L1024 (4 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	N_{max}	N_{mean}	N_{eff}	
L1024 (3 nJ)		1024	8176	1058	825	90.4	
Correlation no binning		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.8514(4)**	**0.062**	**0.0191(1)**	**0.12**	**0.00046(4)**	**0.23**
(3 nJ)	squashed	0.8149(6)	0.15	0.0155(1)	0.28	0.00041(4)	0.32
	TP-MPS	0.8192(6)	0.14	0.0160(1)	0.26	0.00041(4)	0.32
Correlation 8-ports binned		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9856(6)**	**0.058**	**0.0432(4)**	**0.088**	**0.00447(5)**	**0.087**
(3 nJ)	squashed	0.9790(9)	0.11	0.0357(4)	0.25	0.00329(5)	0.114
	TP-MPS	0.9804(7)	0.10	0.0364(4)	0.23	0.00340(5)	0.105
Correlation 16-ports binned		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9869(4)**	**0.050**	**0.0702(7)**	**0.088**	**0.0050(1)**	**0.12**
(3 nJ)	squashed	0.9757(6)	0.21	0.0540(7)	0.30	0.0036(1)	0.37
	TP-MPS	0.9778(6)	0.19	0.0553(7)	0.28	0.0038(1)	0.33
Photon number distribution		**TVD**			**ΔΣ**		
L1024	exp.	**0.004**			**0.007**		
(3 nJ)	squashed	0.037			0.072		
	TP-MPS	0.034			0.068		

Table S6: Validation results summary of the L1024 (3 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
L1024 (2 nJ)		1024	8176	663	489	74.2	
Correlation no binning		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.7461(5)**	**0.064**	**0.0121(1)**	**0.12**	**0.00027(4)**	**0.23**
(2 nJ)	squashed	0.6509(8)	0.27	0.0078(1)	0.43	0.00018(4)	0.48
	TP-MPS	0.6599(8)	0.26	0.0081(1)	0.41	0.00018(4)	0.47
Correlation 8-ports binned		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9769(5)**	**0.066**	**0.0268(1)**	**0.10**	**0.00150(3)**	**0.087**
(2 nJ)	squashed	0.950(1)	0.018	0.0183(1)	0.39	0.00088(3)	0.114
	TP-MPS	0.954(1)	0.21	0.0191(1)	0.36	0.00088(3)	0.50
Correlation 16-ports binned		2-order		3-order		4-order	
		PCC	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
L1024	exp.	**0.9799(3)**	**0.057**	**0.0412(3)**	**0.10**	**0.00218(6)**	**0.15**
(2 nJ)	squashed	0.9491(6)	0.32	0.0254(3)	0.45	0.00099(6)	0.61
	TP-MPS	0.9525(6)	0.30	0.0263(3)	0.43	0.00095(6)	0.63
Photon number distribution		**TVD**				**ΔΣ**	
L1024	exp.	**0.005**				**0.01**	
(2 nJ)	squashed	0.517				0.1	
	TP-MPS	0.050				0.09	

Table S7: Validation results summary of the L1024 (2 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
M256 (6 nJ)		256	5104	997	688	27.7	
Correlation no binning		2-order		3-order		4-order	
M256 (6 nJ)		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	0.9856(1)	0.046	**0.1262(4)**	**0.069**	**0.00552(5)**	**0.096**
	squashed	0.9864(1)	0.027	0.1251(5)	0.078	0.00535(6)	0.124
	TP-MPS	0.9864(1)	0.029	0.1252(5)	0.078	0.00534(6)	0.125
Correlation 8-ports binned		2-order		3-order		4-order	
M256 (6 nJ)		**PCC**	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9968(2)**	0.051	**0.207(1)**	**0.073**	**0.01343(7)**	**0.068**
	squashed	0.9964(2)	0.023	0.202(1)	0.093	0.01289(8)	0.10
	TP-MPS	0.9964(2)	0.024	0.203(1)	0.092	0.01315(8)	0.083
Correlation 16-ports binned		2-order		3-order		4-order	
M256 (6 nJ)		**PCC**	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9907(4)**	**0.060**	**0.207(2)**	**0.088**	**0.0264(7)**	**0.082**
	squashed	0.9904(4)	0.067	0.200(2)	0.11	0.0237(7)	0.12
	TP-MPS	0.9904(4)	0.068	0.200(2)	0.11	0.0252(7)	0.10
Photon number distribution		**TVD**			**$\Delta\Sigma$**		
M256 (6 nJ)	exp.	**0.005**			**0.009**		
	squashed	0.012			0.023		
	TP-MPS	0.012			0.022		

Table S8: Validation results summary of the M256 (6 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
M256 (5 nJ)		256	5104	776	516	26.6	
Correlation no binning		2-order		3-order		4-order	
M256 (5 nJ)		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	0.9785(1)	0.049	**0.0974(4)**	**0.076**	**0.00391(5)**	**0.10**
	squashed	0.9787(1)	0.029	0.0944(4)	0.11	0.00365(5)	0.16
Correlation 8-ports binned		2-order		3-order		4-order	
M256 (5 nJ)		**PCC**	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9963(2)**	0.053	**0.182(1)**	**0.077**	**0.01047(6)**	**0.087**
	squashed	0.9955(2)	0.036	0.173(1)	0.12	0.00946(6)	0.17
Correlation 16-ports binned		2-order		3-order		4-order	
M256 (5 nJ)		**PCC**	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9916(2)**	**0.063**	**0.1983(9)**	**0.099**	**0.0150(4)**	**0.12**
	squashed	0.9898(3)	0.088	0.1852(9)	0.15	0.0127(4)	0.24
Photon number distribution		**TVD**			**$\Delta\Sigma$**		
M256 (5 nJ)	exp.	**0.006**			**0.01**		
	squashed	0.02			0.04		

Table S9: Validation results summary of the M256 (5 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
M256 (4 nJ)		256	5104	556	358	24.9	
Correlation no binning		2-order		3-order		4-order	
M256 (4 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9627(2)**	**0.054**	**0.0676(3)**	**0.083**	**0.00244(6)**	**0.10**
	squashed	0.9609(3)	0.0622(3)	0.0623(3)	0.16	0.00209(7)	0.23
Correlation 8-ports binned		2-order		3-order		4-order	
M256 (4 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9951(2)**	**0.055**	**0.1605(7)**	**0.078**	**0.00846(7)**	**0.089**
	squashed	0.9927(3)	0.059	0.1447(6)	0.17	0.00710(6)	0.24
Correlation 16-ports binned		2-order		3-order		4-order	
M256 (4 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9922(2)**	**0.061**	**0.1856(5)**	**0.095**	**0.0103(2)**	**0.13**
	squashed	0.9876(2)	0.12	0.1619(7)	0.21	0.0088(3)	0.25
Photon number distribution		**TVD**			**ΔΣ**		
M256 (4 nJ)	exp.	**0.007**			**0.013**		
	squashed	0.026			0.051		

Table S10: Validation results summary of the M256 (4 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
M256 (3 nJ)		256	5104	386	217	21.9	
Correlation no binning		2-order		3-order		4-order	
M256 (3 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9238(4)**	**0.060**	**0.0398(2)**	**0.096**	**0.00129(8)**	**0.088**
	squashed	0.9101(7)	0.11	0.0333(2)	0.25	0.00083(6)	0.41
Correlation 8-ports binned		2-order		3-order		4-order	
M256 (3 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9925(2)**	**0.056**	**0.1333(5)**	**0.082**	**0.00631(6)**	**0.096**
	squashed	0.9846(5)	0.11	0.1093(4)	0.25	0.00463(5)	0.34
Correlation 16-ports binned		2-order		3-order		4-order	
M256 (3 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9918(1)**	**0.057**	**0.1570(6)**	**0.098**	**0.0079(2)**	**0.13**
	squashed	0.9801(2)	0.20	0.1216(5)	0.30	0.0055(2)	0.39
Photon number distribution		**TVD**			**ΔΣ**		
M256 (3 nJ)	exp.	**0.007**			**0.014**		
	squashed	0.039			0.076		

Table S11: Validation results summary of the M256 (3 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
M256 (2 nJ)		256	5104	223	125	18.0	
Correlation no binning		2-order		3-order		4-order	
M256 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.8441(8)**	**0.066**	**0.0223(1)**	**0.11**	**0.00050(8)**	**0.29**
	squashed	0.792(1)	0.20	0.0156(1)	0.38	0.00021(8)	0.70
	TP-MPS	0.8075(9)	0.17	0.0169(1)	0.33	0.00034(4)	0.52
Correlation 8-ports binned		2-order		3-order		4-order	
M256 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9864(4)**	**0.061**	**0.0992(3)**	**0.096**	**0.00432(9)**	**0.11**
	squashed	0.963(1)	0.20	0.0690(3)	0.37	0.0025(1)	0.49
	TP-MPS	0.9737(9)	0.17	0.0754(3)	0.31	0.00030(1)	0.38
Correlation 16-ports binned		2-order		3-order		4-order	
M256 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9896(2)**	**0.058**	**0.1272(6)**	**0.10**	**0.0060(2)**	**0.13**
	squashed	0.9618(4)	0.31	0.0833(5)	0.41	0.0032(2)	0.53
	TP-MPS	0.9722(9)	0.26	0.0926(5)	0.35	0.00039(4)	0.44
Photon number distribution		**TVD**			**ΔΣ**		
M256 (2 nJ)	exp.	**0.006**			**0.013**		
	squashed	0.052			0.10		
	TP-MPS	0.043			0.084		

Table S12: Validation results summary of the M256 (2 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
S64 (6 nJ)		64	4336	414	191	6.9	
Correlation no binning		2-order		3-order		4-order	
S64 (6 nJ)		PCC	ΔK	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	0.9943(4)	0.038	**0.390(1)**	**0.055**	**0.0279(3)**	**0.061**
	squashed	0.9940(5)	0.038	0.383(1)	0.076	0.0263(3)	0.11
	TP-MPS	0.9943(4)	0.038	0.388(1)	0.062	0.0274(3)	0.083
Correlation 8-ports binned		2-order		3-order		4-order	
S64 (6 nJ)		**PCC**	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9989(1)**	**0.033**	**0.7976(5)**	**0.046**	**0.1178(1)**	**0.063**
	squashed	0.9983(1)	0.037	0.7861(5)	0.082	0.1116(1)	0.11
	TP-MPS	0.9987(1)	0.033	0.7958(5)	0.053	0.1160(1)	0.070
Correlation 16-ports binned		2-order		3-order		4-order	
S64 (6 nJ)		**PCC**	**ΔK**	**PCC**	**ΔK**	**PCC**	**ΔK**
	exp.	**0.9987(1)**	**0.033**	**0.8280(6)**	**0.055**	**0.1422(3)**	**0.074**
	squashed	0.9979(1)	0.057	0.8164(6)	0.096	0.1323(4)	0.14
	TP-MPS	0.9984(1)	0.034	0.8243(6)	0.061	0.1397(3)	0.085
Photon number distribution		**TVD**			**$\Delta\Sigma$**		
S64 (6 nJ)	exp.	**0.007**			**0.013**		
	squashed	0.015			0.029		
	TP-MPS	0.009			0.015		

Table S13: Validation results summary of the S64 (6 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
S64 (6 nJ)		64	4336	302	139	6.6	
Correlation no binning		2-order		3-order		4-order	
S64 (5 nJ)	exp.	**PCC** **0.9906(1)**	**Δ*K*** **0.042**	**PCC** **0.286(1)**	**Δ*K*** **0.061**	**PCC** **0.0178(3)**	**Δ*K*** **0.073**
	squashed	0.9896(1)	0.054	0.274(1)	0.10	0.0162(2)	0.16
Correlation 8-ports binned		2-order		3-order		4-order	
S64 (5 nJ)	exp.	**PCC** **0.9986(1)**	**Δ*K*** **0.037**	**PCC** **0.7561(8)**	**Δ*K*** **0.053**	**PCC** **0.1027(2)**	**Δ*K*** **0.066**
	squashed	0.9973(1)	0.054	0.736(1)	0.11	0.0941(2)	0.15
Correlation 16-ports binned		2-order		3-order		4-order	
S64 (5 nJ)	exp.	**PCC** **0.9984(1)**	**Δ*K*** **0.034**	**PCC** **0.8067(8)**	**Δ*K*** **0.057**	**PCC** **0.1282(4)**	**Δ*K*** **0.078**
	squashed	0.9967(1)	0.078	0.7859(8)	0.12	0.1140(3)	0.18
Photon number distribution		**TVD**			**ΔΣ**		
S64 (5 nJ)	exp.	**0.008**			**0.015**		
	squashed	0.02			0.038		

Table S14: Validation results summary of the S64 (5 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
S64 (4 nJ)		64	4336	219	94	6.2	
Correlation no binning		2-order		3-order		4-order	
S64 (4 nJ)	exp.	**PCC** **0.9822(2)**	**Δ*K*** **0.047**	**PCC** **0.187(1)**	**Δ*K*** **0.071**	**PCC** **0.0099(2)**	**Δ*K*** **0.086**
	squashed	0.9790(2)	0.083	0.171(1)	0.15	0.0087(2)	0.20
Correlation 8-ports binned		2-order		3-order		4-order	
S64 (4 nJ)	exp.	**PCC** **0.9977(1)**	**Δ*K*** **0.041**	**PCC** **0.679(1)**	**Δ*K*** **0.061**	**PCC** **0.0818(2)**	**Δ*K*** **0.071**
	squashed	0.9948(1)	0.084	0.640(1)	0.16	0.0705(2)	0.21
Correlation 16-ports binned		2-order		3-order		4-order	
S64 (4 nJ)	exp.	**PCC** **0.9977(1)**	**Δ*K*** **0.035**	**PCC** **0.7642(9)**	**Δ*K*** **0.063**	**PCC** **0.1062(4)**	**Δ*K*** **0.079**
	squashed	0.9941(1)	0.11	0.724(1)	0.17	0.0879(3)	0.24
Photon number distribution		**TVD**			**ΔΣ**		
S64 (4 nJ)	exp.	**0.01**			**0.019**		
	squashed	0.029			0.056		

Table S15: Validation results summary of the S64 (4 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	$N_{\max}$	N_{mean}	N_{eff}	
S64 (3 nJ)		64	4336	134	55	5.5	
Correlation no binning		2-order		3-order		4-order	
S64 (3 nJ)	exp.	**PCC** **0.9592(4)**	**Δ*K*** **0.056**	**PCC** **0.1015(6)**	**Δ*K*** **0.87**	**PCC** **0.0043(2)**	**Δ*K*** **0.11**
	squashed	0.9446(5)	0.14	0.0845(6)	0.25	0.0031(2)	0.38
Correlation 8-ports binned		2-order		3-order		4-order	
S64 (3 nJ)	exp.	**PCC** **0.9960(1)**	**Δ*K*** **0.051**	**PCC** **0.530(1)**	**Δ*K*** **0.081**	**PCC** **0.0536(3)**	**Δ*K*** **0.088**
	squashed	0.9875(1)	0.15	0.455(1)	0.25	0.0404(2)	0.32
Correlation 16-ports binned		2-order		3-order		4-order	
S64 (3 nJ)	exp.	**PCC** **0.9971(1)**	**Δ*K*** **0.045**	**PCC** **0.669(1)**	**Δ*K*** **0.080**	**PCC** **0.0749(4)**	**Δ*K*** **0.094**
	squashed	0.9871(1)	0.19	0.584(1)	0.27	0.0530(3)	0.37
Photon number distribution		**TVD**			**ΔΣ**		
S64 (3 nJ)	exp.	**0.009**			**0.019**		
	squashed	0.039			0.076		

Table S16: Validation results summary of the S64 (3 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Dataset scale		input	output	N_{max}	N_{mean}	N_{eff}	
S64 (2 nJ)		64	4336	88	31	4.5	
Correlation no binning		2-order		3-order		4-order	
S64 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9108(6)**	**0.065**	**0.0539(4)**	**0.11**	**0.0020(3)**	**0.11**
	squashed	0.861(1)	0.24	0.0387(4)	0.37	0.0011(2)	0.51
	TP-MPS	0.8946(9)	0.15	0.0472(4)	0.24	0.00017(4)	0.26
Correlation 8-ports binned		2-order		3-order		4-order	
S64 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9929(2)**	**0.058**	**0.365(1)**	**0.099**	**0.0305(2)**	**0.13**
	squashed	0.9694(6)	0.26	0.262(1)	0.38	0.0189(2)	0.47
	TP-MPS	0.9851(9)	0.16	0.0317(1)	0.25	0.0265(2)	0.26
Correlation 16-ports binned		2-order		3-order		4-order	
S64 (2 nJ)		**PCC**	**Δ*K***	**PCC**	**Δ*K***	**PCC**	**Δ*K***
	exp.	**0.9956(1)**	**0.051**	**0.530(2)**	**0.095**	**0.0484(3)**	**0.11**
	squashed	0.9710(6)	0.31	0.386(1)	0.40	0.0261(3)	0.53
	TP-MPS	0.9879(9)	0.20	0.0472(2)	0.24	0.0398(3)	0.29
Photon number distribution		**TVD**			**ΔΣ**		
S64 (2 nJ)	exp.	**0.008**			**0.014**		
	squashed	0.053			0.10		
	TP-MPS	0.035			0.065		

Table S17: Validation results summary of the S64 (2 nJ) dataset. The metrics on which the experimental data performs best are emphasized with bold fonts. The uncertainty of PCC is the standard error obtained by randomly and equally partitioning the overall correlators into 100 subgroups.

Correlation no binning	2-order	3-order	4-order
L1024/M256/S64	All correlators except a tiny part of datapoints which involve temporally adjacent output modes and are affected by the dead time of single photon detectors.	350 million randomly selected correlators.	680 million randomly selected correlators.

Table S18: Detailed description of correlator selection for the no-binning case.

Correlation 8-ports binned	2-order	3-order	4-order
L1024	All correlators.	All correlators.	Totally 8176 modes are partitioned into 4 disjoint parts [1,2048], (2048,4096], (4096,6044], (6044, 8176] to count the interior correlation statistics separately and merge the results as a whole.
M256	All correlators.	All correlators.	Totally 5104 modes are partitioned into 3 disjoint parts [1,2048], (2048,4096], (4096,5104] to count the interior correlation statistics separately and merge the results as a whole.
S64	All correlators.	All correlators.	Totally 4336 modes are partitioned into 3 disjoint parts [1,2048], (2048,4096], (4096,4336] to count the interior correlation statistics separately and merge the results as a whole.

Table S19: Detailed description of correlator selection for the 8-ports-binned case. Each consecutive 8 ports are binned together.

Correlation 16-ports binned	2-order	3-order	4-order
L1024	All correlators.	All correlators.	Totally 8176 modes are partitioned into 2 disjoint parts [1,4096), [4096,8176] to count the interior correlation statistics separately and merge the results as a whole.
M256	All correlators.	All correlators.	Totally 5104 modes are partitioned into 3 disjoint parts [1,2048], (2048,4096], (4096,5104] to count the interior correlation statistics separately and merge the results as a whole.
S64	All correlators.	All correlators.	Totally 4336 modes are partitioned into 3 disjoint parts [1,2048], (2048,4096], (4096,4336] to count the interior correlation statistics separately and merge the results as a whole.

Table S20: Detailed description of correlator selection for the 16-ports-binned case. Each consecutive 16 ports are binned together.

9 Distinct unitaries implemented in the experiments

To demonstrate the programmability of our system, we implemented two distinct unitary transformations in the experiments.

Two datasets are collected and validated with different transformation matrix during the experiments : dataset A under pump settings from 2 nJ to 6 nJ which is thoroughly examined across the paper, dataset B under pump settings from 3 nJ to 7 nJ. In this section we provide an overview of dataset B together with supportive validation results.

Photon number distribution of the L1024 group of dataset B is plotted in Fig. S26. Up to 3050 photon-detection events is observed in the largest experiment.

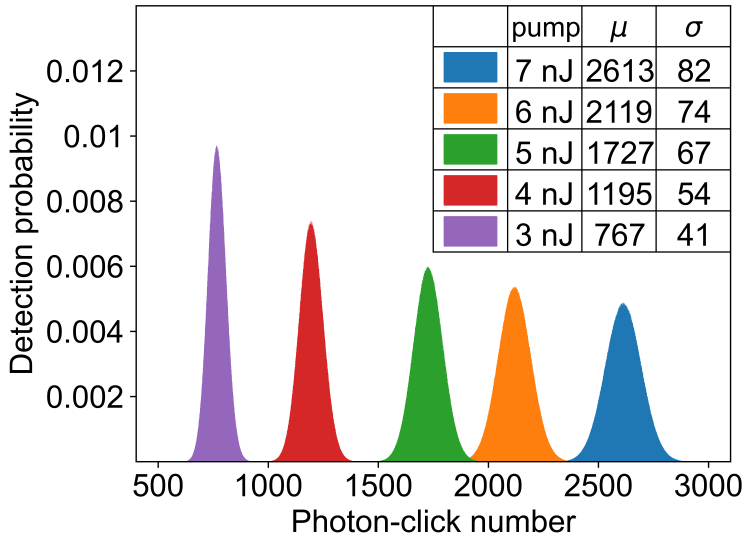


Figure S26: Photon-click number distribution for the L1024 group (dataset B) across various single-pulse energy, with their mean μ and standard deviations σ listed in the inner panel.

Fig. S27 compare the experimental photon number distribution with the theoretical prediction of the groundtruth and the mockups,the experimental distribution overlaps well with the groundtruth while strongly deviates from all the classical mockups.

Fig. S28 presents the Bayesian test results of the S64 group of dataset B. We start from a subsystem with fewer output bosonic modes, and gradually increase the subsystem size . The conclusion here is consistent with dataset A in the main text that not only all the Bayesian scores at the largest subsystem size are higher than zero which indicates that the experimental samples are

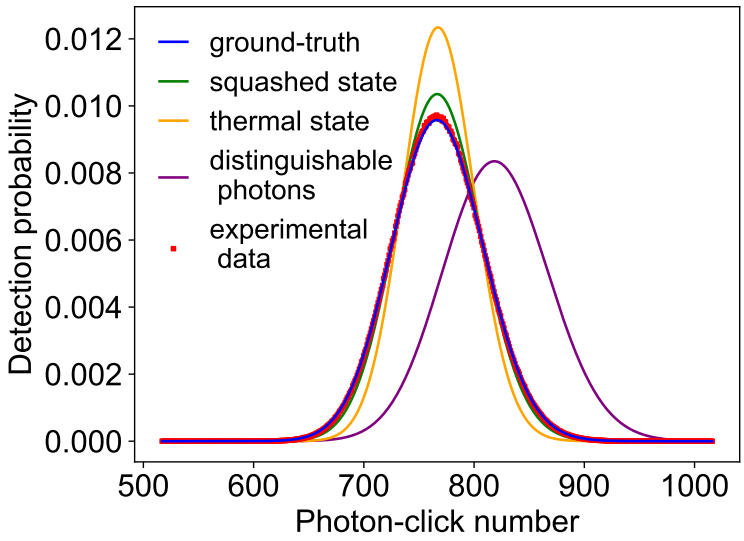


Figure S27: A comparison of the photon number distribution of the experimental results (L1024 (3 nJ) of dataset B), the ground-truth theory and classical mockups (squashed state, thermal state, distinguishable photons).

more likely generated from the ground-truth distribution, there is also a clear trend of increasing as the size of subsystem ramps up.

Fig. S29a,b,c rules out the independent pairs and singles (IPS) sampler, the treewidth sampler (treewidth = 513) and the greedy sampler, respectively.

Following the same procedure in Fig.2f of the main text, Fig. S30 shows the high-order correlation benchmark of dataset B, with the squashed state mockup for comparison. The experimental data yields better correlation values compared to the squashed sampler across all orders and pump settings, except for the 3-order correlation of the 7 nJ dataset where the difference does not exceed the standard error.

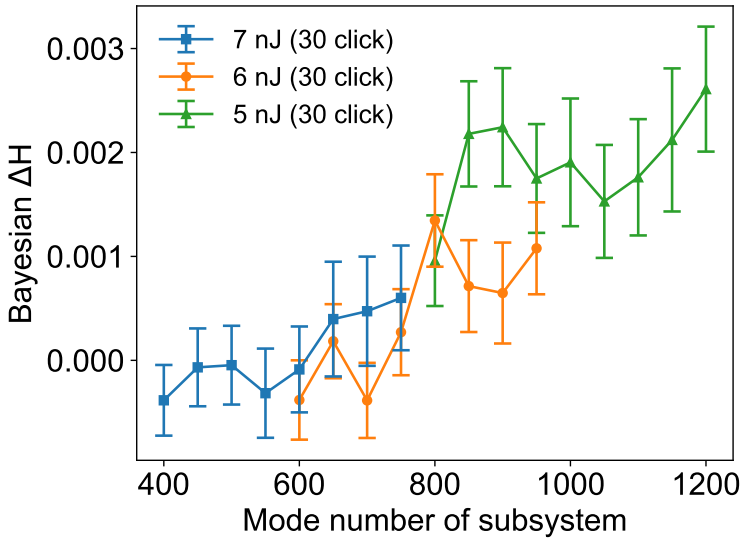


Figure S28: Bayesian tests against the squashed state, computed for specific photon-click number as a function of subsystem size. Experimental data from the S64 group (dataset B) are used for the Bayesian test limited by classical computational overhead. The subsystems for validation are randomly selected. Error bars indicate the standard error obtained from 100 groups of resampling.

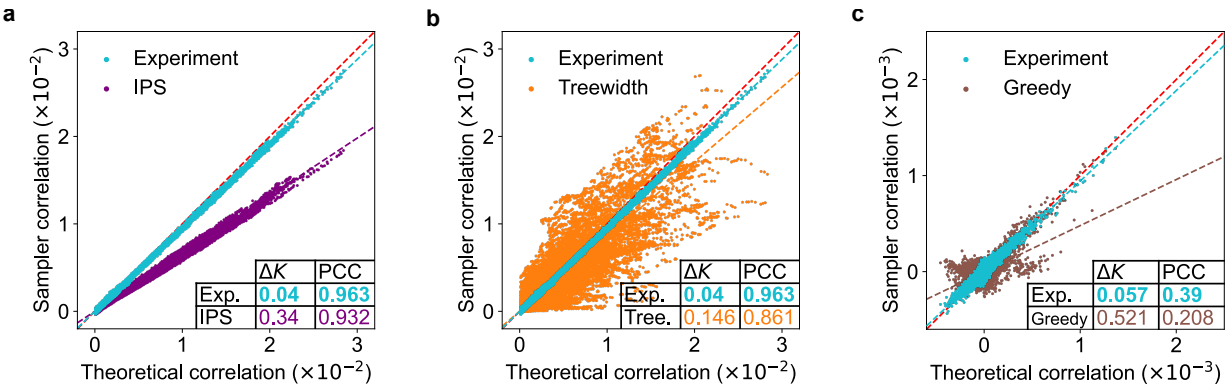


Figure S29: **a**,**b**, Scatter plot of the two-order correlation function of the IPS sampler **a** and the treewidth sampler **b** in comparison with the experimental results. **c**, Scatter plot of the third-order correlation function of the greedy sampler (reproducing up to 2-order) in comparison with the experimental results. The first 4096 modes are selected for processing and each consecutive 8 modes are binned together to relieve the huge cost for counting sample statistics. 10 million samples of the L1024 (7 nJ) subset of dataset B are used in **a**,**b**,**c** to evaluate the results.

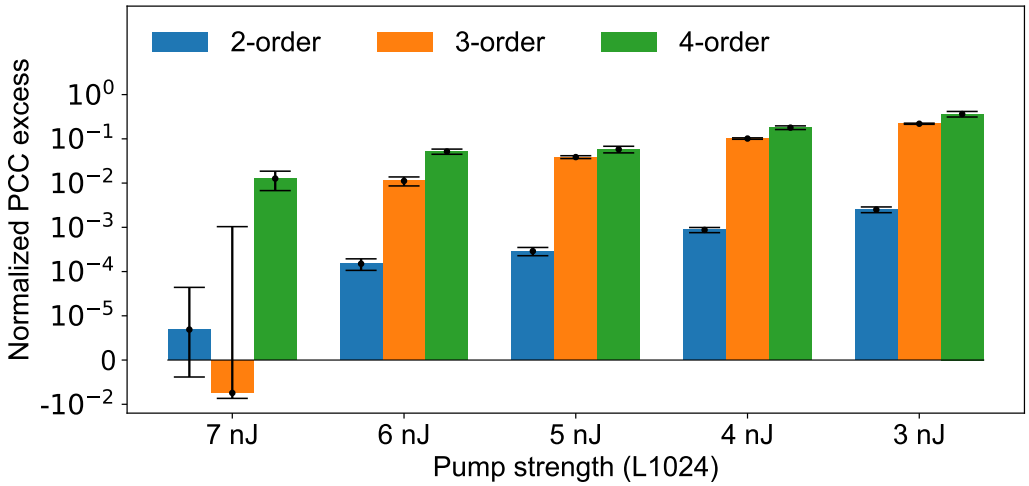


Figure S30: Comparison of Pearson correlation coefficient (PCC) for the 2,3,4-order correlation function between the experimental data and the squashed sampler. We define the normalized PCC excess as $(\mathrm{PCC}_{\mathrm{exp}} - \mathrm{PCC}_{\mathrm{squash}})/\mathrm{PCC}_{\mathrm{squash}}$, plotted on the y-axis for visual clarity of the contrast. L1024 group (10 million samples) of dataset B are shown for pump settings ranging from 3 nJ to 7 nJ.

References and Notes

[1] N. Quesada, J. E. Sipe, Effects of time ordering in quantum nonlinear optics. *Physical Review A* **90** (6), 063840 (2014), doi:10.1103/PhysRevA.90.063840, https://link.aps.org/doi/10.1103/PhysRevA.90.063840.

[2] N. Quesada, G. Triginer, M. D. Vidrighin, J. E. Sipe, Theory of high-gain twin-beam generation in waveguides: From Maxwell's equations to efficient simulation. *Physical Review A* **102** (3), 033519 (2020), doi:10.1103/PhysRevA.102.033519, https://link.aps.org/doi/10.1103/PhysRevA.102.033519.

[3] P. D. Drummond, B. Opanchuk, A. Dellios, M. D. Reid, Simulating complex networks in phase space: Gaussian boson sampling. *Physical Review A* **105** (1), 012427 (2022).

[4] C. Oh, M. Liu, Y. Alexeev, B. Fefferman, L. Jiang, Classical Algorithm for Simulating Experimental Gaussian Boson Sampling pp. 1–8, doi:10.1038/s41567-024-02535-8, https://www.nature.com/articles/s41567-024-02535-8.

[5] B. O'Donoghue, Operator Splitting for a Homogeneous Embedding of the Linear Complementarity Problem. *SIAM Journal on Optimization* **31**, 1999–2023 (2021).

[6] Decompositions — Strawberry Fields, https://strawberryfields.ai/photonics/conventions/decompositions.html.

[7] Y. Chen, S.-Q. Gong, L. Gan, Y. Liu, A. Yang, Y. Wang, C.-Y. Lu, G. Yang, FastMPS: Revisit Data Parallel in Large-scale Matrix Product State Sampling (2025), doi:10.48550/arXiv.2512.20064.