Supplementary Figure 1: Calibration for monitoring $R_{\text{ph}}$. Transmission ($T$, blue curve) and reflection ($R$, red curve) of the beam splitter including the two lenses (shown on Fig. 1), measured for calibrating the setup.
Supplementary Figure 2: Superconducting thin film characterization. a) Scanning electron microscope image of a test nanowire. The scale bar is 1µm. b) Measured critical current versus temperature. Circle, square, diamond and triangle markers are for $W_{NW}$ equal to 100nm, 80nm, 60nm and 40nm, respectively. $T_C$ is set to 7.16K; and the vertical lines mark $T = 4.2$K and $T = 2.05$K. c) $I_{C0}$ versus $W_{NW}$.
Supplementary Figure 3: Measuring \( \eta_A \). a) A sample test device used for measuring \( \eta_A \). The scale bar is 200nm. b) Transmission through grating-waveguide-grating devices with a nanowire on top. The total nanowire lengths \( (L_{NW}) \) are 5\( \mu \)m, 12.4\( \mu \)m, 16.8\( \mu \)m, 22.15\( \mu \)m, 39.8\( \mu \)m, and 57.2\( \mu \)m top to bottom. c) The mean normalized transmission versus \( L_{NW} \) as measured (circles), the best linear fit (dashed line), and expected from simulations (solid line).
Supplementary Figure 4: Transmission through cryostat. The lines show transmission measurement results (reflecting off a thick gold film) for several test devices.
**Supplementary Figure 5: Measuring $\eta_{WG}$.** a) An example of a test device: two gratings connected by different lengths of waveguide. The scale bar is 20µm. b) Average transmission versus total waveguide length for straight (blue circles) and bent (black squares) waveguides.
**Supplementary Figure 6: Grating couplers.** a) The dimensions are: the first slot width = 100nm, width of the other slots = 160nm, array pitch = 530nm, horizontal distance from the first slot to the focal point ($F$) = 14.9µm, and the opening angle = $44^\circ$. The scale bar is 2µm. b) Simulated input and output coupling efficiencies for this coupler design.
Supplementary Figure 7: Dispersion of input optics. a) Simplified schematic of the input optical setup used to focus the laser on the grating couplers (L: lens, W: cryostat window, BS: beam splitter, S: shield, IGC: input grating coupler). The fixed positions are: $x_1 = 50\text{mm}$, $x_2 = 90\text{mm}$, $x_3 = 100\text{mm}$, $x_4 = 122\text{mm}$, and $x_5 = 140\text{mm}$. b) The calculated shift (dx) in the focal point position as a result of material dispersion.
Supplementary Figure 8: Transmission through Grating-Waveguide-Grating test devices.

Blue lines are measurements, dashed pink line is what is expected from numerical simulations of the grating couplers taking into account the measured waveguide losses, and the dotted red line is the result of scaling and shifting the numerically simulated spectra by 0.87 and 3nm respectively. The inset shows a histogram of measured transmissions at the peak wavelength (solid red line is the best Gaussian fit with standard deviation of ~8%).
Supplementary Figure 9: Outside vs inside cryostat transmission measurements. Transmission through a fixed device measured outside (black) and inside (blue) the cryostat. The red dotted lines are Gaussian fits.
Supplementary Figure 10: Transmission spectra at room temperature vs cryogenic temperature. Transmission through a fixed device measured at room temperature (RT blue curve) and submerged in superfluid helium (CT black curve). The red dotted lines are Gaussian fits.
Supplementary Figure 11: Comparison of the two approaches to estimating $\eta_C$. The solid red line and gray dashed lines show the $\eta_C^{\text{est}}$ (left axis) determined using the measurements described in the Supplementary Note 5. The dotted blue line shows the result for $\eta_C(\lambda)\eta_D^{\text{sat}}$ (right axis) obtained using the technique described in the main text.
<table>
<thead>
<tr>
<th>Description of deviations in simulation compared to nominal</th>
<th>$\lambda_R$</th>
<th>$A_{\text{peak}}$</th>
<th>$BW$</th>
<th>$R$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal ($W_{NW}=40\text{nm}$, $N_r=7$, $R_r=66$, Waveguide Width = 500nm, at 2.05K)</td>
<td>1545.4</td>
<td>98.6</td>
<td>6.7</td>
<td>0.03</td>
<td>0.3</td>
</tr>
<tr>
<td>Wider nanowire ($W_{NW}=60\text{nm}$)</td>
<td>1545.0</td>
<td>92.0</td>
<td>9.3</td>
<td>7.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Narrower nanowire ($W_{NW}=20\text{nm}$)</td>
<td>1546.0</td>
<td>82.7</td>
<td>4.8</td>
<td>15.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Nanowire misalignment relative to cavity (50nm closer to front mirror)</td>
<td>1545.3</td>
<td>97.6</td>
<td>7.5</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Nanowire misalignment relative to cavity (50nm closer to back mirror)</td>
<td>1545.5</td>
<td>97.9</td>
<td>7.2</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Front mirror hole sizes (all holes 10% smaller radius)</td>
<td>1550.9</td>
<td>91.5</td>
<td>8.8</td>
<td>7.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Front mirror hole sizes (all holes 10% larger radius)</td>
<td>1540.3</td>
<td>92.5</td>
<td>5.5</td>
<td>6.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Wider waveguide (width equal to 530nm)</td>
<td>1568.6</td>
<td>95.2</td>
<td>6.9</td>
<td>3.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Narrower waveguide (width equal to 470nm)</td>
<td>1519.7</td>
<td>96.2</td>
<td>6.8</td>
<td>2.9</td>
<td>0.2</td>
</tr>
<tr>
<td>At room temperature in air (Index of refraction for silicon is set to 3.4785, and for environment is set to 1.0)</td>
<td>1554.9</td>
<td>98.6</td>
<td>6.4</td>
<td>0.04</td>
<td>0.3</td>
</tr>
</tbody>
</table>

**Supplementary Table 1: Robustness Analysis.** $\lambda_R$ is the resonance wavelength, $BW$ is the full width at half maximum bandwidth, $A_{\text{peak}}$, $R$, and $T$ are the absorption, reflection and transmission at $\lambda_R$, respectively. See Fig. 2 of the main text for details of the nominal design.
Supplementary Note 1. Robustness Analysis

The robustness of this microcavity SNSPD design to fabrication imperfections is evaluated by taking a $W_{NW}=40\text{nm}$, $N_r=7$, $R_r=66$, $Q=231$ chip as an example, and simulating the effect of various deviations on its performance. Supplementary Table 1 summarizes these simulation results.

A general observation is that any change - like nanowire width or front mirror hole-sizes - that can significantly perturb the balance between $Q_r$ and $Q_A$ shows up as a significant increase of reflection at the expense of reduced peak absorption. Also, changes like waveguide width or temperature, that can perturb the effective index of the waveguide mode manifest themselves as a change in resonance wavelength. Note that deviations considered in Supplementary Table 1 are extreme cases, and today’s fabrication technology usually provides much better control on dimensions of photonic circuits. Nonetheless, the microcavity SNSPD continues to provide more than 90% of peak absorption efficiency in almost all simulated scenarios.
Supplementary Note 2. Calibrating $R_{\text{Ph}}$

The optical setup shown in Fig. 1 includes a beam splitter right before the cryostat windows. The reflected power ($P_{\text{R}}$) from this splitter (registered by D$_1$ on Fig. 1) was used to monitor the power incident on the first cryostat window ($P_{\text{in}}$). The rate of incident photons ($R_{\text{Ph}}$) was then calculated using $R_{\text{Ph}} = P_{\text{in}}/E_{\text{Ph}}$, where $E_{\text{Ph}}$ is the energy of a photon at the known wavelength of the laser.

Before placing the optical setup near the cryostat, $P_{\text{R}}$ and $P_{\text{in}}$ were measured for a fixed input laser power, and reflection and transmission were calculated as $R = P_{\text{R}}/(P_{\text{R}} + P_{\text{in}})$, and $T = P_{\text{in}}/(P_{\text{R}} + P_{\text{in}})$, respectively. The result is shown on Supplementary Figure 1. In the experiments reported here and in the main text, the $P_{\text{in}}$ was measured by $P_{\text{in}} = P_{\text{R}} \times T/R$.

For single photon detection experiments, the input laser power was attenuated to the level of pico-watts using fiber attenuators. To read this power accurately with the power meter, the laser was first turned off to read an offset level ($P_{\text{Off}}$). Turning on the laser, another power ($P_{\text{U}}$) was read. $P_{\text{in}}$ was calculated as $P_{\text{in}} = P_{\text{U}} - P_{\text{Off}}$. In all the cases the power meter sensitivity was adjusted to have $P_{\text{in}} \gg P_{\text{Off}}$.

The uncertainty in $P_{\text{in}}$ is mostly determined by the calibration of the power meter, which is known to be better than few percent ($\sim 2\%$) through its calibration report. As shown below in Supplementary Note 5, the one standard deviation uncertainty in $\eta_C$ due to device-to-device variations is $\pm 8\%$, so this $\sim 2\%$ uncertainty associated with $P_{\text{in}}$ has negligible impact on the overall uncertainty when added in quadrature.
Supplementary Note 3. Superconducting Thin Film Characterization

Several short superconducting bridges were fabricated to test the superconducting properties of the 8nm thick NbTiN coated SOI wafers used to make the SNSPDs as reported in the main text. Each device (see Supplementary Figure 2a) was laid out as a short (1µm long) nanowire connected to large gold contact pads (not shown) through an optimized width converter to avoid current crowding\(^1,2\). The devices were mounted in the same setup used to test the SNSPDs, and the experimental critical currents \((I_C)\) versus temperature were measured.

Supplementary Figure 2b shows \(I_C(T)\) versus \((1 − T/T_C)^{3/2}\) for nanowires with different widths (as expected for the temperature dependence of the Ginzburg-Landau depairing critical current\(^3\)). The solid lines are \(I_{C0}(1 − T/T_C)^{3/2}\). A \(T_C=7.16\text{K}\) makes the measured points line up as lines passing through origin. Different \(I_{C0}\) values adjust the slopes of the lines for each \(W_{NW}\) to fit the measured \(I_C\). The \(I_{C0}\) values for each \(W_{NW}\) are plotted on Supplementary Figure 2c. The solid line shows the best linear fit that includes the origin. It indicates the proportionality of \(I_{C0}\) to \(W_{NW}\). The fitted \(I_{C0}/(W_{NW} \times 8\text{nm})\) gives the value of the critical current density \(J_C(T = 0) = 7.57 \times 10^6 \text{A/cm}^2\) reported in the main text.

Based on these characterizations, \(I_C(T = 2.05\text{K}, W_{NW} = 35\text{nm})\) was expected to be \(~12.7\mu\text{A}\) for the SNSPDs reported in the main text. However, the observed critical currents are all < 8µA (see Fig. 3c), which suggests that either the U-shaped bend in those wires constricts the current flow, or there are fabrication constrictions along their rather longer length. Regardless, the U-shaped nanowires exhibit almost saturated \(\eta_D\) versus bias current (see Fig. 3c).
Supplementary Note 4. Measuring $\eta_A$

A set of test devices was placed close to the on-chip detectors to enable measurement of the absorption of the guided light by the nanowires ($\eta_A$) for TW SNSPDs. These devices are comprised of two grating couplers connected by a long waveguide (see input/output grating couplers number 3 and 4 on Fig. 1) that has a U-shaped nanowire with different lengths on top. The geometry of the nanowire (except the length) was the same as the TW SNSPDs but with no contacts (see Supplementary Figure 3a as an example).

The transmission through these devices normalized to the transmission of a similar device but with no nanowire is shown on Supplementary Figure 3b. It is clear that the absorption in the small wavelength range of interest around 1550nm is almost independent of the wavelength. These normalized transmissions, averaged over 1540nm to 1560nm and plotted versus total nanowire length (including two straight sections and a 180-degree turn) are shown by the circles of Supplementary Figure 3c. The dashed line is the best linear fit. The expected exponential decay of the optical field (as it is evanescently absorbed by the nanowire) is confirmed by the good linear fit on the logarithmic scale plot.

Shown on the same figure is the solid line from the equation $T = (1 - \eta_T)\exp(-\alpha_{NW}L_{WG})$
where $\eta_T = 0.0273$ is the simulated absorption of the nanowire bend, and $\alpha_{NW} = 0.0789/\mu$m is the simulated attenuation constant of a nanowire loaded waveguide, and $L_{WG}$ is the length of the waveguide that supports the two straight nanowires ($L_{WG} = (L_{NW} - \pi R)/2, R = 175$nm).
For $L_{NW} = 57.2\mu m$ (the same as the long TW SNSPD), the measured $T$ is 0.0723, the best linear fit gives 0.0830, and the simulated value is 0.1043. These correspond to 92.7%, 91.7% and 89.6% of absorption.
Supplementary Note 5. Direct Determination of $\eta_C$

The coupling efficiency of the incident photons to the strip waveguide ($\eta_C$) was measured up to an scaling constant using a relatively simple but indirect method as described in the methods section of the main text. Here, a more involved method is presented for measuring $\eta_C$, and the results of the two approaches are compared.

As illustrated in Fig. 1 (of the main text), photons incident on the cryostat ($R_{Ph}$) reach the single photon detector after travelling through the cryostat windows, an input grating coupler, and a waveguide with a single bend. Therefore, $\eta_C$ can be written as $\eta_C = \eta_W \eta_{IGC} \eta_{WG}$, where $\eta_W$ is the loss due to cryostat windows, $\eta_{IGC}$ is the loss due to the input grating couplers, and $\eta_{WG}$ is the loss due to the waveguide used to connect the detectors to the grating couplers. It what follows each of these efficiencies is carefully examined, and $\eta_C$ is deduced at the end.

Measuring $\eta_W$ In order to measure the loss due to the cryostat windows, $\eta_W$, the incident photons were focused on the center of a highly reflective gold pad and the power meter $D_2$ (see setup of Fig. 1) monitored the reflection. The ratio of the output power ($P_{out}$) to the incident power ($P_{in}$, see $R_{Ph}$ in Fig. 1) was measured for several of the test devices, and plotted in Supplementary Figure 4.

The 100nm thick, large square gold pad made when defining the contact pads can be seen in Fig. 1, to the left of the input grating couplers. It is a highly reflective surface in the wavelength range of interest: its s-polarized reflectivity at 45° is over 98%. Considering this near unity reflectivity, the symmetry of the cryostat windows (W on Fig. 1), and also near unity transmission of the
anti-reflection coated lenses used (L of Fig. 1): \( \eta_W \simeq \sqrt{P_{\text{out}}/P_{\text{in}}} \simeq 0.74 \).

The measured \( \eta_W \) is somewhat smaller than \( \sim 0.93^2 = 0.86 \) expected for Fresnel reflections from two glass windows. Two factors contributed to this difference: (i) the inner cryostat windows have a visible thin film of vacuum pump oil on them, and (ii) the diameters of the cryostat windows and the holes in the blackbody shield (S on Fig. 1) occlude a few percent of the beam power.

The above results were measured when the cryostat was filled with superfluid helium. However, repeating the same measurement at room temperature resulted the same value of 0.74 for \( \eta_W \). This shows the negligible effect of superfluid helium on transmission of the optical beams, as expected.

**Measuring \( \eta_{WG} \)** A set of test devices were fabricated and measured to determine the loss of waveguides made with the methods described in the main text. The devices each consist of two grating couplers connected by waveguides of different lengths (see Supplementary Figure 5a as an example). The transmission through these devices was measured using the same setup of Fig. 1, but without the cryostat.

The average transmissions in the wavelength range of 1540nm to 1560nm versus total waveguide length (\( L_{WG} \)) are shown by the symbols on Supplementary Figure 5b. Blue circles are for straight waveguides, and the black squares are for waveguides with two bends (5\( \mu \)m radius of curvature; similar to the bends on the detector waveguides of Fig. 1). The lines are linear fits with a slope of \( \alpha_{WG} = -9 \text{dB/mm} \). The presence of the two bends adds an additional loss of -1dB which
indicates an additional loss of \( \eta_B = 10^{-0.05} = 0.89 \) per waveguide bend. Therefore, the total waveguide loss is expected to be \( \eta_{WG} = 10^{0.1\alpha_{WG}L_{WG}}(\eta_B)^{N_B} \), where \( N_B \) is the number of bends.

The measured \( \alpha_{WG} \) is higher than the typical values reported for silicon photonic chips. Investigation of the waveguides under an scanning electron microscope revealed a rather high line edge roughness (see edges on Supplementary Figure 3a as an example). These rough edges are known to cause waveguide losses by scattering the guided light. Inasmuch as the study of detectors is concerned, \( \eta_{WG} \) is a loss the same as the others between the nanowire detector and the incident photons. Fabrication optimization for better \( \eta_{WG} \) was not a focus for the present work. However, there are well-studied techniques for fabricating very low loss silicon waveguides that must be employed for making integrated quantum photonic circuits.

**Estimating \( \eta_{IGC} \)** It is very difficult to measure \( \eta_{IGC} \) (the loss due to the input grating couplers) as there is no direct means of accessing the optical power within the waveguides. In the following, various results of simulations and experiments are combined to yield an estimate of \( \eta_{IGC} \).

**Design of the Grating Couplers and the Input Focusing Optics** Supplementary Figure 6a shows an image of the backward focusing and tilted grating couplers (GC) used in this work. In the backward operation, light is incident on the grating in a direction counter to waveguide propagation. Therefore, the input light reflected from the backside of the substrate is directed away from the detectors, and this helps reduce coupling of the scattered photons to the detectors. The focusing grating makes the footprint small; this allows placing many elements within a single write-field of the electron beam lithography tool while avoiding stitching errors. The tilted design makes the
reflection of light approaching the coupler from the waveguide very small; this removes strong Fabry-Perot resonances between the detectors and the grating couplers (or between two couplers) and therefore greatly simplifies all the relevant analysis. The grating couplers were designed for out-coupling ~1550nm light to the far field at 45 degrees (compatible with the arrangement of the cryostat shown in Fig. 1).

To simulate the performance of the GCs, a finite difference time domain solver (Lumerical, Inc.) was used to excite the waveguide with its fundamental transverse-electric mode, and the far field of the light emitted from the coupler was monitored. The resulting output coupling efficiency, \( \eta_{\text{OGC}}^{\text{sim}}(\lambda) \), is shown by the blue circles in Supplementary Figure 6b. The output coupler shows a large bandwidth centered close to 1550nm as needed for the detector experiments.

The coupling efficiency into the waveguide is more complex as it is both a function of the grating coupler and the quality of the focused beam on the coupler. In all the experiments reported here, the input coupling setup was manually adjusted for maximum transmission (or maximum single photon counts) at 1550nm. The setup was then left intact while sweeping the wavelength for spectral measurements. Therefore, while it is fair to assume the best focus was at 1550nm, the focus at other wavelengths will have been affected by the dispersion of the focusing optics used in the setup.

Supplementary Figure 7a shows a simplified schematic of the input focusing optics. It focuses the laser light from a single mode fiber (Corning Inc. SMF-28\textsuperscript{TM}) onto the input grating coupler. A simple Gaussian beam propagation method is used to calculate the size and position
of the focal point. The index of refraction of the lenses and windows versus wavelength is known from their material specifications. The focal length of the lenses versus wavelength are calculated by having their known geometry and by using the spherical lens formula: 

\[ f = \frac{R}{n - 1} \]

where \( R \) is the lens radius of curvature and \( n(\lambda) \) is the index of refraction. In exact analogy with the experiments, the whole setup (except the cryostat windows) is moved forward and backward to accurately place the focal point on the grating coupler at 1550nm. The wavelength is then swept, and the size and location of the focal point is calculated in the wavelength range of interest. The results are: (i) the size of the focal point (5.2\( \mu \)m in radius) is almost independent of the wavelength, (ii) the position of the focal point moves by +2.5\( \mu \)m/nm change in wavelength (see Supplementary Figure 7b), and (iii) removing the cryostat windows (that is measuring outside the cryostat) gives almost the same size of the focal point with the same +2.5\( \mu \)m/nm change factor.

To simulate the input coupling efficiency at 1550nm, the waist (5.2\( \mu \)m in radius) of a 45\(^\circ\) incident Gaussian beam is placed on the grating coupler, and the power coupled to the waveguide is monitored. The same simulation but with the waist moved backward and forward by +2.5\( \mu \)m/nm was repeated for other wavelengths. The resulting input coupling efficiency, \( \eta_{\text{IGC}}(\lambda) \), is shown by the black squares in Supplementary Figure 6b. As expected, because of reciprocity both input and output efficiency values approach the same level when the wavelength is close to 1550nm. However, away from 1550nm, the dispersion of the input optics breaks this input/output symmetry and the expected efficiency values are different.
Estimating Efficiencies of the Grating Couplers Outside the Cryostat

A set of test devices were fabricated close to the on-chip detectors to estimate the $\eta_{IGC}$. These test devices consisted of two grating couplers connected by a long waveguide ($L_{WG}=650\mu m$) with two bends (see input/output grating couplers number 5 and 6 on Fig. 1). The $\eta_{IGC}$ is first estimated outside the cryostat. The setup was aligned for maximum transmission through the devices at 1550nm, and the output power versus wavelength ($P_{out}$) was registered. The transmission, $T_{out}(\lambda)$, is found by normalizing $P_{out}$ to the power reflected from the gold square (the same used for $\eta_w$ measurement). This is plotted on Supplementary Figure 8 using solid blue lines.

The mean transmission near 1550nm varies by $\pm 8\%$ (one standard deviation) from device to device (see inset of Supplementary Figure 8). In comparison the variation of $T_{out}(\lambda)$ for a single device measured multiple times was negligible. Therefore, $\pm 8\%$ uncertainty in the transmission is attributed to the devices and not to the experimental setup and procedures.

The dashed pink line in Supplementary Figure 8 compares several measured transmission spectra with the calculated $T_{out}(\lambda) = \eta_{IGC}^{\text{sim}}(\lambda)\eta_{OOGC}^{\text{sim}}(\lambda)10^{0.1\alpha_{WG}L_{WG}(\eta_B)^N_B}$, where $\alpha_{WG}$ and $\eta_B$ were taken to be -9dB/mm and 0.89 respectively (see above), $L_{WG} = 650\mu m$, and $N_B = 2$. The discrepancy between the average measured transmission and that simulated using the grating coupler design parameters is most likely due to the imperfect etching of the first (100nm wide) slot in the grating coupler and small differences in the width of the other slots, as determined from scanning electron microscopy. To account for this difference, the transmission spectrum calculated for the design parameters is scaled and shifted as per $\eta_{IGC}^{\text{exp-out}}(\lambda) = \eta_L\eta_{IGC}^{\text{sim}}(\lambda + \Delta \lambda)$, and
\[ \eta_{\text{OGC}}^{\exp-out}(\lambda) = \eta_{L} \eta_{\text{sim}}^{\text{OGC}}(\lambda + \Delta \lambda). \]

The dotted red line in Supplementary Figure 8 shows the plot of \( \eta_{\text{IGC}}^{\exp-out}(\lambda) \eta_{\text{OGC}}^{\exp-out}(\lambda) 10^{0.1 \alpha_{\text{WG}} L_{\text{WG}} (\eta_{B}) N_{B}} \) using \( \eta_{L} = 0.87 \) and \( \Delta \lambda = 3 \text{nm} \). Therefore, \( \eta_{\text{IGC}}^{\exp-out}(\lambda) \) and \( \eta_{\text{OGC}}^{\exp-out}(\lambda) \) are taken to be accurate estimates for the actual, measured average grating coupler efficiencies when excited outside of the cryostat.

**Estimating Efficiencies of the Grating Couplers Inside the Cryostat** The same measurement as the one described above was repeated for the chip installed inside the cryostat. For a fixed device on the chip, a new measured transmission - \( T_{\text{in}}(\lambda) \) - is compared with \( T_{\text{out}}(\lambda) \) in Supplementary Figure 9. The discrepancy is attributed to distortion of the input beam quality due to transmission through the windows, which influences only \( \eta_{\text{IGC}}(\lambda) \). A final estimate for the input grating coupling efficiency inside the cryostat is therefore:

\[ \eta_{\text{IGC}}^{\exp-in}(\lambda) = T_{\text{in}}(\lambda) / (\eta_{\text{IGC}}^{\exp-out}(\lambda) 10^{0.1 \alpha_{\text{WG}} L_{\text{WG}} (\eta_{B}) N_{B}}), \]

with \( \alpha_{\text{WG}} = -9 \text{dB/mm}, \eta_{B} = 0.89, L_{\text{WG}} = 650 \mu\text{m}, N_{B} = 2, \eta_{\text{IGC}}^{\exp-out}(\lambda) = \eta_{L} \eta_{\text{sim}}^{\text{OGC}}(\lambda + \Delta \lambda), \eta_{L} = 0.87, \) and \( \Delta \lambda = 3 \text{nm} \).

**Estimating Efficiency Changes with Temperature** Supplementary Figure 10 shows that the spectrum rigidly shifts by approximately 5nm when the sample is cooled to 2.05K, but the peak transmission value remains unchanged. The shift is attributed to the change in refractive index of silicon and silicon dioxide with temperature as well as the change in refractive index of the environment (air versus superfluid helium). Based on this data, \( \eta_{\text{IGC}}^{\exp-2K}(\lambda) = \eta_{\text{IGC}}^{\exp-in}(\lambda - 5\text{nm}) \) is taken as an estimate of the input grating coupler efficiency at the cryogenic operating temperature of the detectors.
**Discussion** Having thus measured $\eta_W$, $\eta_{WG}$, and having carefully deduced $\eta_{IGC}$, it is now possible to calculate $\eta_C$. All the single photon detectors in this work are connected to a grating coupler through a $L_{WG} = 325\mu m$ long waveguide that has a single bend. Therefore, $\eta_C = \eta_W \eta_{IGC} \eta_{WG}$ can be estimated as $\eta_C^{\text{est}}(\lambda) = \eta_W \eta_{IGC}^{\exp-2K} \eta_{B}^{L_{WG} \eta_{B}}$ with all the parameters fixed from the explained experiments. The result is the solid red line on Supplementary Figure 11. As illustrated, the $\eta_C^{\text{est}}(\lambda)$ is in agreement with the $\eta_C(\lambda) \eta_D^{\text{sat}}$ deduced using the alternative method explained in the main text (dotted blue line), assuming $\eta_D^{\text{sat}} = 1$. Repeating the same procedure for some of the other test devices of Supplementary Figure 8, it is possible to observe the variation of $\eta_C^{\text{est}}(\lambda)$ between different devices. The result is shown by gray dashed lines on Supplementary Figure 11. The variation around the peak efficiency is $\sim \pm 8\%$ in agreement with the variations reported in Supplementary Figure 8. From this it is deduced that $\eta_D^{\text{sat}}$ has, within one standard deviation, a lower bound of 0.92.

In conclusion, the directly measured $\eta_C^{\text{est}}(\lambda)$ turns out to be in almost exact agreement with $\eta_C(\lambda) \eta_D^{\text{sat}}$ deduced using a completely independent approach. This implies that, within the one standard deviation uncertainty, $\eta_D^{\text{sat}} = 1$, and the $\eta_C(\lambda)$ calibration curves are reliable. While the uncertainties associated with the methods themselves are different, the relatively large experimental device to device variability is what ultimately limits the accuracy of the final result.
Supplementary References

