Supporting Information: Demonstration of a Thermally-Coupled Row-Column SNSPD Imaging Array

Jason P. Allmaras,*,†,‡ Emma E. Wollman,‡ Andrew D. Beyer,‡ Ryan M. Briggs,‡ Boris A. Korzh,‡ Bruce Bumble,‡ and Matthew D. Shaw‡

†*Department of Applied Physics and Materials Science, California Institute of Technology, Pasadena, CA, United States.*

‡*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States.*

E-mail: jallmara@caltech.edu

Device Fabrication

Devices were fabricated on a 4 inch silicon wafer. A 60 nm thick Au back reflector was patterned via optical lithography and lift-off before a 155 nm thick film of SiO_2 was sputtered to form an insulating layer. An 8 nm thick WSi film was sputtered from a compound target, after which Au contact pads and leads were patterned via lift-off. The bottom nanowire layer was pattered using electron-beam lithography and etched using an ICP RIE dry-etch of CHF_3 and O_2. After the first nanowire layer was patterned, a buffer layer of $\sim$190 nm SiO_2 was sputtered and smoothed using angled incidence Ar ion milling. The sputtering and ion

milling process was necessary to planarize the surface sufficiently to yield the second layer of nanowires without constrictions. The second nanowire layer and leads were patterned using the same process as the lower layer and a final capping layer of $\sim 65\,\mathrm{nm}$ of SiO_2 was sputtered for passivation. Optical efficiency could be enhanced by designing and depositing antireflection coatings to form an optical stack.

Two designs were fabricated using this procedure. The first was a 4×4 array using parallel nanowires on the top and bottom layers as shown schematically in Main Text Figure 1b and in the optical micrograph of Figure 1a. The total active area was $91.2\,\mu\mathrm{m}$ by $91.2\,\mu\mathrm{m}$ with each row and column consisting of four $22.8\,\mu\mathrm{m}$ by $22.8\,\mu\mathrm{m}$ units connected in series. The second design was a 4×4 cross-polarized array using perpendicular top and bottom nanowires as shown schematically in Main Text Figure 1c and in the microscope image of Figure 1b. The total active area was $91.2\,\mu\mathrm{m}$ by $91.2\,\mu\mathrm{m}$. Both devices use $160\,\mathrm{nm}$ wide nanowires with $1200\,\mathrm{nm}$ pitch. The low fill factor was selected to avoid crosstalk between adjacent pixels[1] while maintaining a uniform nanowire fill throughout the entire active area. A higher nanowire fill-factor could be achieved while avoiding crosstalk by increasing the fill-factor within a pixel but leaving additional guard space between adjacent pixels.

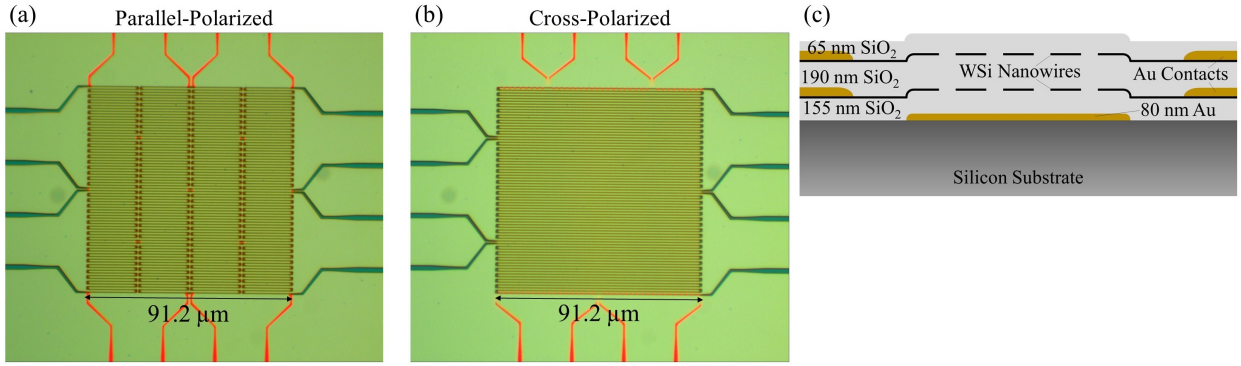

Figure 1: Optical microscope images of (a) parallel-polarized and (b) cross-polarized devices. Blue channels are the top layer while red channels are the bottom layer. (c) Schematic cross-section of the TRC devices showing layer thicknesses.

Design of TRC arrays requires a balance between the electrical and thermal crosstalk mechanisms in neighboring nanowires. During a detection event in these two-layer devices,

both mechanisms couple energy from the detecting nanowire to adjacent channels on both the original and second layers.[2] Capacitive coupling is the dominant mechanism of microwave coupling in this geometry, so as the thickness of the spacer layer between the nanowire layers decreases, the strength of the electrical coupling between neighboring nanowires increases. Electrical crosstalk of this form is undesirable because it is distributed across all channels of the second layer rather than localized to the detecting pixel. Thermal coupling is inherently local to where the detection occurs, but the timescale of this coupling is slow compared to electrical crosstalk. The time difference between the generation of normal domains in the two detecting nanowires, defined as the thermal coupling delay time (see Main Text Figure 1a), is on the order of a few nanoseconds for the geometries we demonstrate in this work. An optimal design would minimize the spacer thickness while still preventing electrical crosstalk.

The need to balance electrical and thermal crosstalk dictates the acceptable spacing between the two nanowire layers. Previous experiments on planar arrays of WSi nanowires demonstrated that ∼2.5 mm long nanowires experience near-unity thermal crosstalk and noticeable electrical crosstalk at a pitch of 400 nm with an estimated capacitance of 110 fF.[1] For the TRC, we conservatively desired to reduce the estimated electrical crosstalk by a factor of 3 which would reduce it to the level of the amplifier noise. For the spacer thickness of 190 nm chosen in this work, the channel to channel capacitance is estimated to be approximately 31 fF for the parallel-polarized and 24 fF for the cross-polarized devices, meeting the target capacitance requirement. The lack of evidence for electrical crosstalk in our experimental results justifies this choice of spacer thickness, but thinner dielectric layers could likely be used while still avoiding undesired electrical crosstalk and improving the thermal coupling performance.

Timing Jitter

The timing jitter of the system was characterized using a 20 MHz repetition rate 1550 nm mode-locked laser. A phase-locked loop circuit converted the electrical sync of the laser to a

10 MHz clock which acted as the timing source for the time to digital converter (TDC). The laser was focused to a Gaussian spot with diameter of approximately 85 µm FWHM. Due to non-Gaussian instrument response function (IRF) of the array channels, the timing jitter is defined as 2.355σ where σ is the sample standard deviation of the photocount timetags modulo 50 ns. Figure 2 shows the IRF and the bias current dependent timing jitter for representative channels of both layers of the array, demonstrating that the overall jitter is less than 300 ps at optimal bias regardless of which nanowire absorbs the photon. The jitter of a given pixel of the array is the combination of the two histograms of the channels forming the pixel weighted by the relative absorption efficiency of the two channels because the measured arrival time of the photon is given by the detection time of the first channel to detect the photon.

Thermal Coupling Efficiency

The coupling probability for the two array designs was characterized using a 1550 nm mode-locked laser for a range of bias currents on both top and bottom layers. All four channels of a given layer were biased at the same current and the photon flux was kept sufficiently low to make the probability of two photons being detected in the same optical pulse negligible. For each channel, a bias dependent calibration delay was applied to all timetags such that the mean delay for photons detected on each channel is zero based on the timebase of the mode locked laser. The coupling probability is shown as a function of bias current in Figure 3 for representative channels on both the bottom and top layers. The parallel-polarization device exhibits near unity coupling efficiency for a wide range of bias current combinations $(I_{\mathrm{B,Top}}, I_{\mathrm{B,Bot}})$ as is expected based on the direct overlap between the nanowire regions of the two layers (see Main Text Figure 1b). The coupling efficiency increases to near unity as both the detecting and receiving nanowire bias currents increase. The explanation is straightforward. As the detecting nanowire bias current increases, more Joule heat is released

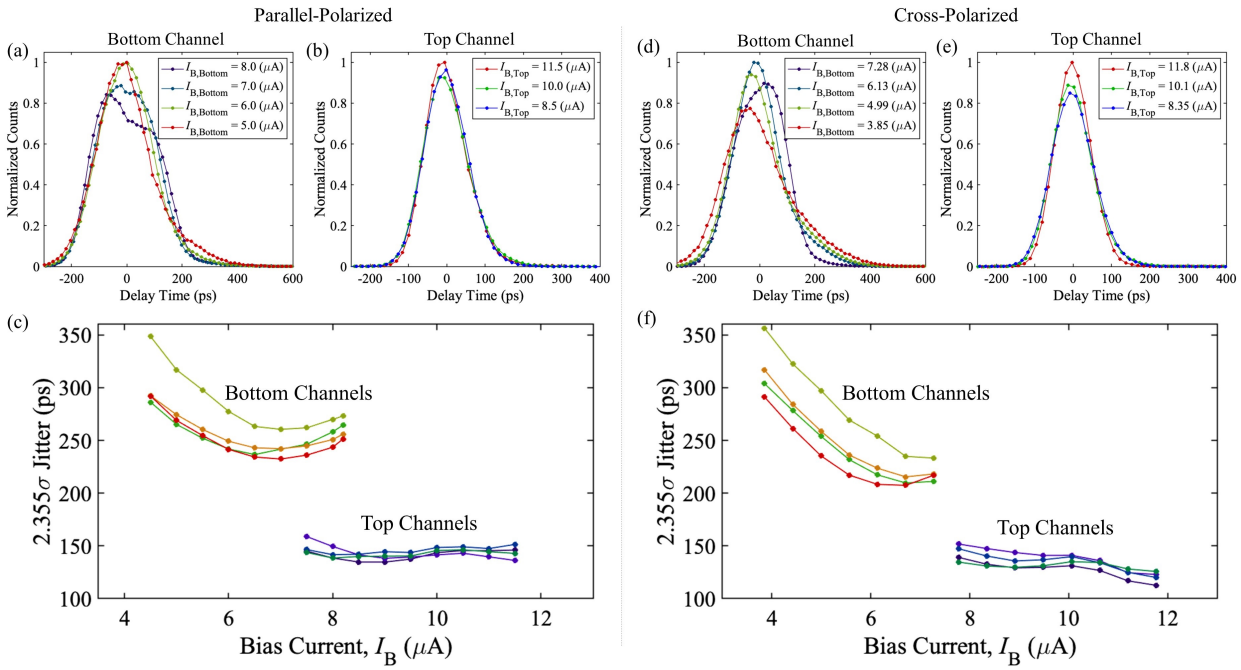

Figure 2: Jitter characterization for channels of the (a-c) parallel-polarized and (d-f) cross-polarized arrays. Jitter histograms for a representative channel of the bottom layer (a, d) show non-Gaussian IRF with a distinct asymmetry in the distribution at high bias currents. We attribute this to electrical coupling between the nanowire layers and to the back reflector which reduces the signal propagation velocity and leads longitudinal geometric jitter.[3] In contrast, the IRF of the top channels (b, e) is nearly Gaussian. To accommodate the non-Gaussian behavior of the lower layer channels, the total jitter is defined by 2.355σ (c, f), and is significantly larger for the bottom channels than the top, even when biased at the similar currents where the electrical noise contribution to the jitter from the readout electronics is similar. A device optimized for improved timing performance must manage the electrical coupling and signal propagation of the bottom layer to take full advantage of the low timing jitter of SNSPDs.

during detection which heats the dielectric layer and second nanowire layer, increasing the coupling efficiency. As the bias current on the receiving layer increases, a smaller change in nanowire temperature is required to trigger a switching event, making the nanowire more sensitive to small heat pulses generated by the source nanowire.

The cross-polarized device displays more varied behavior due to the nonuniform overlap between the two layers (see Main Text Figure 1c). Compared to the parallel-polarization device, higher bias currents are required to ensure near unity p_i ($I_{\mathrm{B,Top}}$, $I_{\mathrm{B,Bot}}$), but when both layers are biased near their switching currents, unity detection can be achieved. Furthermore, the transition region between no coupling and efficient coupling occurs over a wider range of bias currents. This is consistent with the varying overlap between the two nanowire layers.

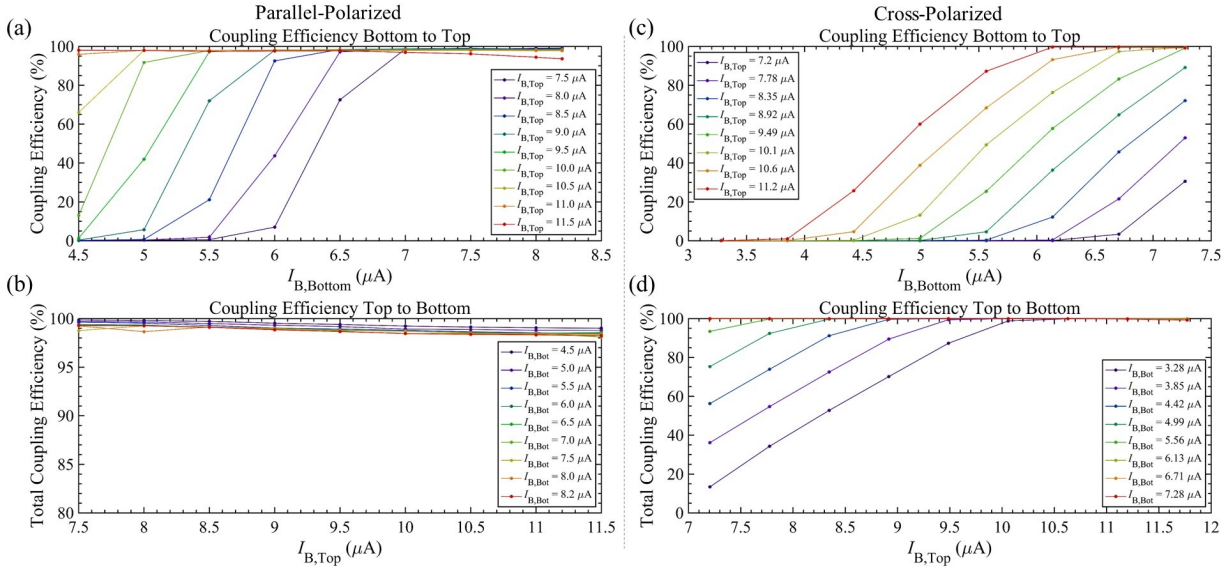

Figure 3: Coupling efficiency for representative channels of the parallel-polarized (a, b) and cross-polarized arrays (c, d). The coupling efficiency increases as both the source and receiving channel bias currents increase. Coupling from the bottom to top layer is less efficient than top to bottom due to the lower overall bias current and lower Joule heat generated during a detection. The transition from no coupling to efficient coupling is more gradual in the cross-polarized device compared to the parallel-polarized device due to the non-uniform overlap between the nanowires of the two layers. Lines are to guide the eye.

The total coupling efficiency between layers in the cross-polarized device increases monotonically as the bias current increases on either the detecting or receiving channels, but the fraction of double coincidences (two detection events within a coincidence window) and triple

coincidences (three detection events) is not constant. As seen in Figure 4a, as the bias current on the bottom channel increases, the coupling efficiency for double coincidences, shown schematically in Figure 4c, initially increases due to the increased Joule heat generated. However, this coupling efficiency actually decreases at the highest bias currents of the top channel. This occurs because the large amount of heat generated by the combination of the two channels switching can be sufficient to switch a neighboring channel which is part of a different pixel. The number of these triple coincidences, shown schematically in Figure 4d, increases as both the top and bottom bias currents increase, but the bias current of the top channel plays a more significant role than that of the bottom channel due to the larger overall bias currents involved. Triple coincidences are only considered to be valid if the the two detecting channels on the same layer are adjacent and the probability of two photons being absorbed within a coincidence window is negligibly small. Analysis of the correlations confirms that at high bias current combinations, $> 98\%$ of the triple coincidence events occur between two adjacent channels on one layer and a single channel on the second layer.

Triple coincidences are not inherently problematic in determining the location of the photon detection because at high bias current operating points, $>99.5\%$ of triple coincidences occur with the first receiving channel belonging to the second layer of the device. Thus, a proper detection location can be assigned based on the detecting and first receiving channels alone. However, to simplify an FPGA based readout system, it is desirable to have only double coincidence events. Furthermore, the possibility of single photon triple coincidence detection events makes it impossible to distinguish between blocking caused by two photon detections and a single photon triple coincidence event. In a cross-polarized device, single photon triple coincidences can be minimized by leaving guard space between adjacent pixels. In the parallel-polarized array, single photon triple coincidences make up a negligible fraction of the detection events for the large pitch arrays tested in this work.

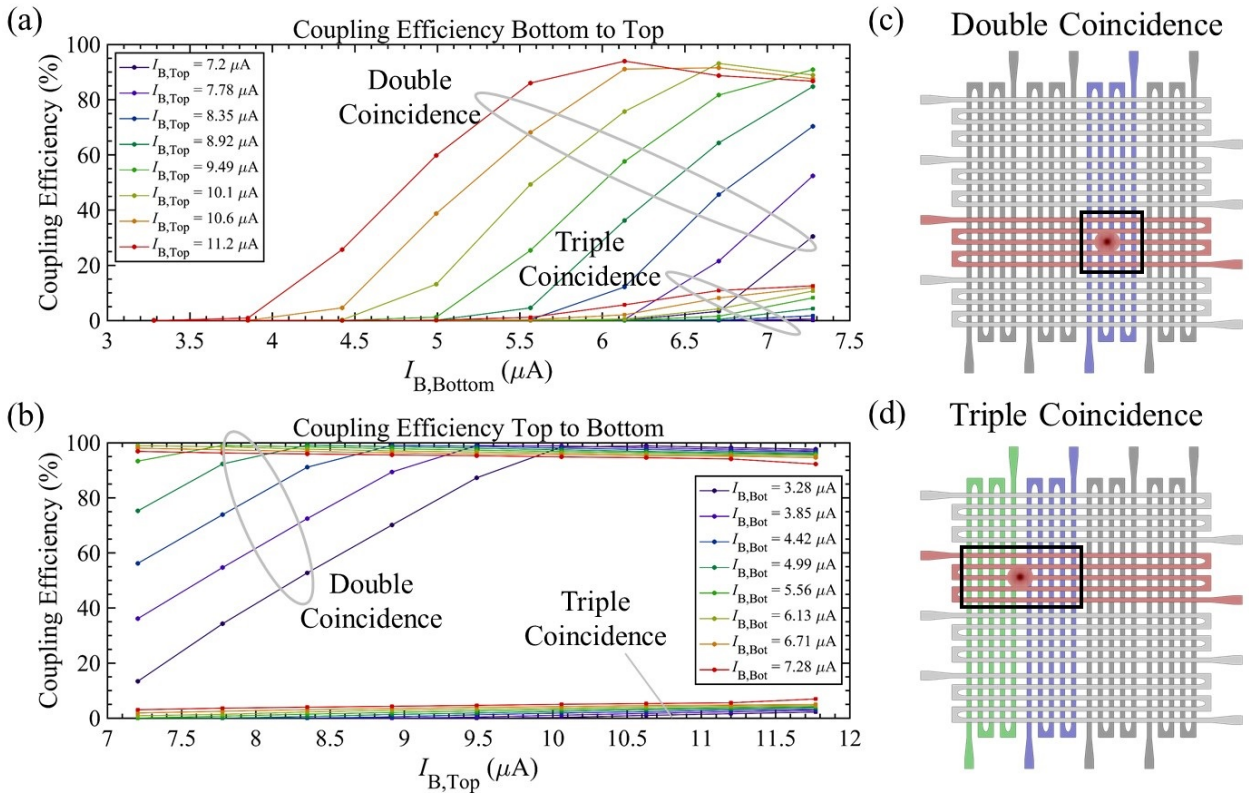

Figure 4: Cross-polarized array coupling efficiency by coincidence number for (a) a bottom channel triggering a top channel and (b) a top channel triggering a bottom channel. The number of triple coincidences is non-negligible for high bias currents on the either the top or bottom channels. Unambiguous double coincidence and single photon triple coincidence detection events are shown schematically in (c) and (d) respectively. The colored channels indicate those which register a detection during the photon detection event, and the overlap of the detecting channels indicates the determined location.

Thermal Coupling Delay Time

The thermal coupling delay time is a function of the bias currents on both the detecting and receiving channels. The average coupling delay times for various bias current configurations are shown in Figure 5 for both array designs. Due to the direct overlap between layers, the parallel-polarized device demonstrates shorter coupling delay times and narrower distributions compared to the cross-polarized device for comparable values of the bias currents. The longer delay times of the cross-polarized device require the use of a longer coincidence window when processing timetags to determine row and column coincidences.

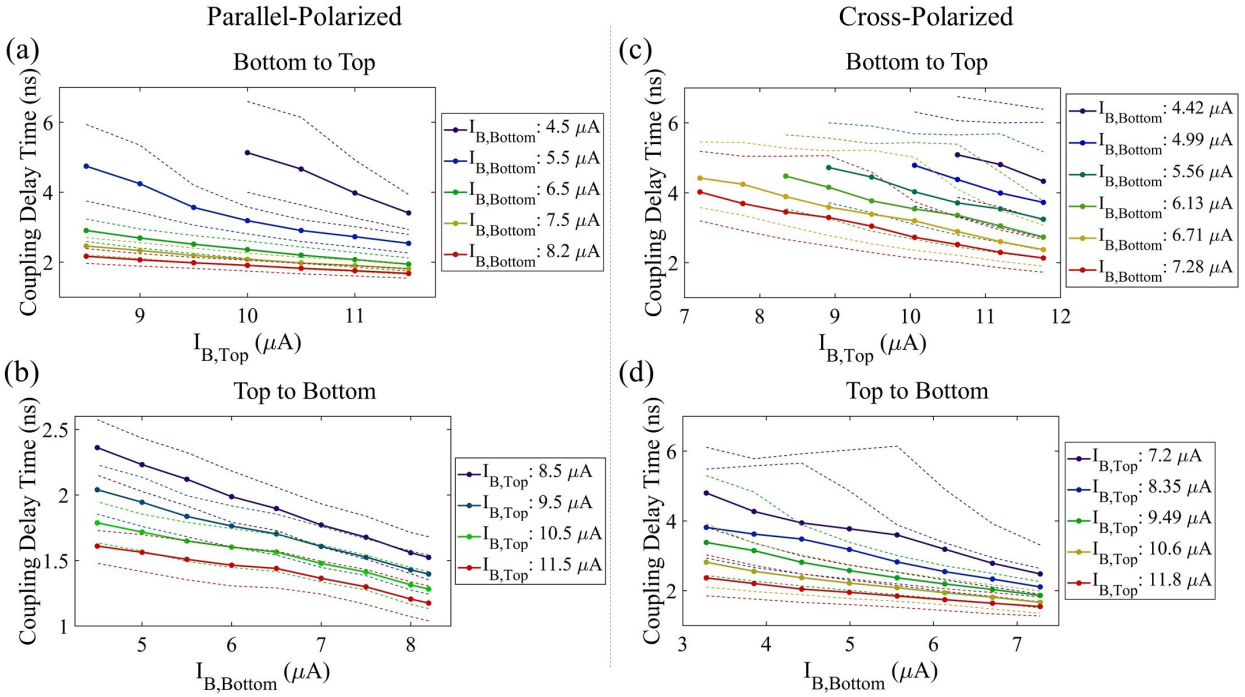

Figure 5: Thermal coupling delay times for (a, b) parallel-polarized and (c, d) cross-polarized representative channels of the arrays. Solid lines with symbols indicate the mean delay time while dashed lines indicate the 90% distribution bounds to show the width of the coupling delay time distribution. The coupling times and the width of the coupling distribution both decrease as the bias current increases on either the source or receiving channels. The cross-polarized channels show a wider distribution and longer mean coupling time than the parallel-polarized channels for comparable bias conditions.

As the bias currents increase, the delay time decreases and the distribution width becomes narrower. Despite separation distances of only a few hundred nanometers between layers,

the delay times are on the order of several nanoseconds. This is significantly slower than expected based on ballistic propagation of phonons and slower than expected based on thermal modeling using a Casimir limited thermal conductivity for thin film dielectrics. There is experimental evidence that this type of suppressed thermal conductivity in thin film SiO_2 is typical,[4,5] but more advanced modeling is needed to fully characterize the behavior, then use the results to optimize the thermal design of TRC arrays. Specialized devices can be designed which use a variety of nanowire widths, inductances, and overlap spacings in order to build a detailed experimental dataset of the coupling delay times between nanowires under different geometries. This type of device offers a new means of studying this physics and improving both engineering and fundamental models of this process. Previous work demonstrated that a complete electrothermal model of the reset dynamics of WSi nanowires requires an accurate model of the heat transfer in the surrounding dielectric material,[5] but the TRC architecture offers a way of experimentally studying this process in order to develop and validate such a model. This understanding is necessary to fully optimize TRC arrays.

Figure 5 also shows the bounds containing 90% of the thermal coupling delay distribution, indicating the amount of thermal coupling jitter. This corresponds to approximately 3.29σ for a Gaussian distribution. For the parallel-polarized device biased near the switching current, the top-to-bottom thermal coupling jitter is comparable to the photon-detection jitter of the system at approximately 260 ps (See Figure 2), but for lower bias currents, thermal coupling adds significant jitter. For the cross-polarized array, even at high bias currents, for the top nanowire triggering the bottom, significant thermal coupling jitter of 600 ps persists due to the non-uniform geometry as discussed throughout the text.

Maximum Count Rate

The maximum count rate (MCR) was characterized using a 1550 nm CW source with variable attenuators and a focused spot with a FWHM diameter of 85 nm on the array. The MCR

is defined as the count rate where the efficiency of the detector drops by 3 dB compared to the low count rate efficiency. The MCR curves are shown below in Figure 6 for both array designs. The parallel-polarized array demonstrates an MCR of 14.6 Mcps while it is only 10.8 Mcps for the cross-polarized array. The parallel-polarized array was biased at 11.5 µA and 8.2 µA for the top and bottom channels respectively while the cross-polarized array was biased at 11.8 µA and 7.28 µA for the top and bottom channels respectively. Due to the polarization dependent relative efficiency of the channel layers, the maximum count rate of the cross-polarized device is slightly polarization dependent. When photons are preferentially absorbed in either the top or bottom layer, the array experiences elevated blocking loss, resulting in a reduced MCR. The MCR depends on the coincidence window chosen to analyze the timetag data. For the parallel-polarized data, the optimal window is 3.5 ns while for the cross-polarized array, the optimal window is 4.5 ns. The larger optimal window for the cross-polarized array is consistent with the longer thermal coupling delay time for these bias currents. In the MCR analysis of the cross-polarized array, triple coincidences were not included as unambiguous counts because the photon flux is not sufficiently low to eliminate the possibility of multiple photons being detected within a coincidence window. Single-photon triple coincidences lead to additional blocking loss and are the primary cause of the lower MCR in the cross-polarized array compared to the parallel-polarized device.

The thermal coupling delay times can have a direct impact on the MCR which can be achieved over the array without position ambiguity. Two factors contribute to position ambiguity. The first is due to blocking loss. During the deadtime of a channel, detections which occur on the other layer but overlapping the dead channel will not have a correlated click. In the limits of equal illumination of pixels, ideal thermal coupling, identical channel properties, and Poisson distributed photons, the array MCR (MCR_A) scales according to $MCR_A \sim N^2\, MCR_i/(2N-1)$ due to this deadtime, where MCR_i is the channel MCR. The second factor is associated with the timing uncertainty of correlated detections. Electrical and thermal timing jitter lead to a range of thermal coupling times between the

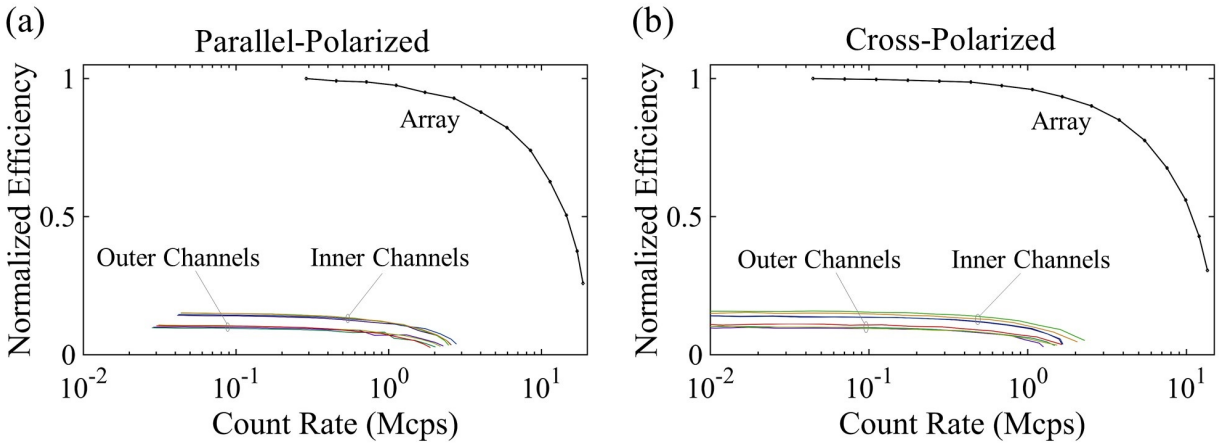

Figure 6: MCR curves for (a) parallel-polarized and (b) cross-polarized arrays. The parallel-polarized device was illuminated with TE polarized light while the cross-polarized array was illuminated with linear polarization rotated 45° with respect to the TE and TM modes. For both arrays, the two inner channels of each layer show higher count rates than the two outer channels due to unequal illumination with the laser spot. In these figures, the parallel-polarized device uses a coincidence window of 3.5 ns while the cross-polarized array uses a coincidence window of 4.5 ns.

photon-induced detection and the thermally-coupled detection. When analyzing the channel detection times to determine the coincidence events, one must define a range of times during which two events can be correlated, known as a coincidence window. Two detection pairs which occur within the same coincidence window lead to ambiguity as to which combination of pixels were the source of detection events because two rows and two columns have four potential pixel locations. As the array size becomes large, the coincidence window of correlated detections limits the counting rate. A large thermal coupling delay indicates that the rate of temperature increase in the second nanowire is slow. Therefore, fluctuations during thermal coupling process lead to larger fluctuations in the coupling delay time, analogous to the dependence of electrical noise jitter on signal slew rate. This requires setting a larger coincidence window, which is only acceptable for low count-rate applications. Reducing the thickness of the spacer layer quickens thermal coupling and potentially leads to a smaller coincidence window, but leads to additional electrical coupling. An optimized design would reduce the thickness of the dielectric spacer to the minimum thickness where the electrical

crosstalk can still be tolerated.

Optical Efficiency

The optical efficiency of each device was measured using a 1550 nm CW laser focused to a Gaussian spot with a FWHM diameter of approximately 33 µm. Only counts which exhibit unambiguous coincidence groups are included in the efficiency measurement. This includes triple coincidences for the cross-polarized device as defined above. We find that the rigorous coupled wave analysis (RCWA) technique provides a reasonable prediction of device efficiency. The parallel polarization device exhibits 34% TE and 11% TM efficiency while the RCWA calculation predicts 34% TE and 14% TM efficiency. In the TM polarization, the model predicts a 64:36 bottom to top layer absorption ratio which is close to the experimentally measured 58:42 ratio. The cross-polarized device has 30% TE and 26% TM efficiency while the RCWA calculation predicts 30% TE and 29% TM efficiency for polarization with respect to the top layer orientation. In the TM polarization, the model predicts a 69:31 bottom to top ratio compared to the measured 72:28 ratio. The parallel-polarized device incurs additional optical coupling losses compared to the cross-polarized design due the presence of sections of non-photosensitive meander bends comprising approximately 18% of the active area. This advantage of the the cross-polarized architecture is significant for small pixel sizes, but becomes less important as the pixel size increases or the illuminating spot is not focused on an area of the array where the bends are present.

References

(1) Shaw, M. D.; Marsili, F.; Beyer, A. D.; Briggs, R. M.; Allmaras, J. P.; Farr, W. H. Superconducting nanowire single photon detectors for deep space optical communication (Conference Presentation). *Proc. SPIE* **2017**, 100960J.

(2) Verma, V. B.; Lita, A. E.; Stevens, M. J.; Mirin, R. P.; Nam, S. W. Athermal avalanche in bilayer superconducting nanowire single-photon detectors. *Appl. Phys. Lett.* **2016**, *108*, 131108.

(3) Zhao, Q.-Y.; Zhu, D.; Calandri, N.; Dane, A. E.; McCaughan, A. N.; Bellei, F.; Wang, H.-Z.; Santavicca, D. F.; Berggren, K. K. Single-photon imager based on a superconducting nanowire delay line. *Nat. Photonics* **2017**, *11*, 247.

(4) Cahill, D. G.; Fischer, H. E.; Klitsner, T.; Swartz, E. T.; Pohl, R. O. Thermal conductivity of thin films: Measurements and understanding. *J. Vac. Sci. Technol. A* **1989**, *7*, 1259–1266.

(5) Allmaras, J. P.; Kozorezov, A. G.; Beyer, A. D.; Marsili, F.; Briggs, R. M.; Shaw, M. D. Thin-Film Thermal Conductivity Measurements Using Superconducting Nanowires. *J. Low Temp. Phys.* **2018**, *193*, 380–386.