A Thin Film Lithium Niobate Near-Infrared Platform for Multiplexed Quantum Nodes

Daniel Assumpcao[1*†], Dylan Renaud[1*†], Aida Baradari[1], Beibei Zeng[2], Chawina De-Eknamkul[2], C.J. Xin[1], Amirhassan Shams-Ansari[3], David Barton[1], Bartholomeus Machielse[2] and Marko Loncar[1*]

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, 02138, MA, United States.
[2]AWS Center for Quantum Networking, Boston, 02135, MA, United States.
[3]DRS Daylight Solutions, San Diego, 16465, CA, United States.

*Corresponding author(s). E-mail(s): dassumpcao@g.harvard.edu; renaud@g.harvard.edu; loncar@g.harvard.edu;
†These authors contributed equally to this work.

Keywords: Visible Wavelength, Lithium Niobate, Integrated Photonics

1 Electro-Optic Shifting

Electro-optic frequency shifting of photons is a possible solution for the inhomogenous distribution of solid-state emitters.

Phase modulators can in principle shift the frequency of CW optical tones with unity efficiency through the application of a linear phase advance or retardation to the transmitted field. More formally, if am input CW tone $Ae^{j\omega t}$ with frequency ω is transmitted through a phase modulator with an applied phase $\phi(t)$, the transmitted field is given by $E_{T,ideal}(t) = Ae^{j(\omega t+\phi(t))}$. When $\phi(t) = kt$ we observe that $E_{T,ideal}(t) = Ae^{j(\omega+k)t} = Ae^{j\omega'}$: a CW tone with a shifted frequency $\omega' = \omega + k$

In practice, a limitless linear increase in phase would require a correspondingly unbounded increase in applied voltage which is nonphysical. Thus various

techniques are used to emulate a linear phase advance. In this work we explore two techniques: shearing and serrodyning. In shearing, a sinusoidal tone is applied to the phase modulator and an optical pulse is co-localized to the linear region of the phase to emulate a linear rise. This technique requires the input light be pulsed, thus causing a fundamental trade-off between the pulse linewidth and frequency shift. With serrodyning, an electrical sawtooth wave is used with an amplitude of 2π. As a 2π phase is equivalent to 0 phase, an ideal sawtooth wave is equivalent to a monotone linear increase. This method works regardless of the initial bandwidth of the input tone, but requires higher bandwidth electro-optic control.

In this section, we will model both techniques to understand their limitations.

1.1 Modeling and Limits of Spectral Shearing

With the shearing technique, a sinusoidal tone with frequency f_{RF} and modulation depth ϕ_0 is applied to the modulator

$$\phi(t) = \phi_0 sin(2\pi f_{RF} t)$$

To emulate a tone, the input optical signal is assumed to be pulsed with pulses co-localized to the rising edge: $E_{T,ideal}(t) = A(t)e^{j(\omega t+\phi(t))}$. Thus we can perform a taylor expansion:

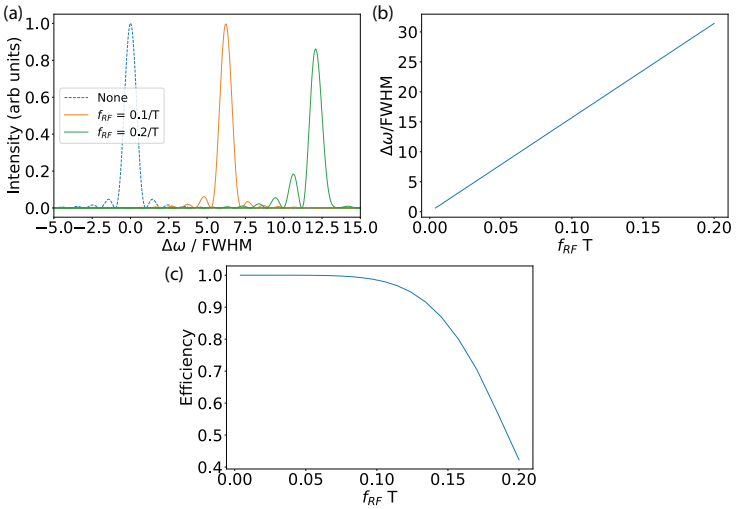
$$\phi(t) \approx 2\pi f_{RF}\phi_0 t - \phi_0 1/6(2\pi f_{RF})^3 t^3 + ...$$

where we observe that to first order, an optical frequency shift occurs: $\Delta\omega_{op} = \phi_0/t = 2\pi f_{RF}\phi_0$. It is clear that in order to maintain the approximate linearity of $\phi(t)$, it is vital that the optical pulse is limited temporally to mitigate the influence of the t^3 and higher order terms. Higher order terms will lead to a shift to other frequency bins, which can be considered to be either a shift distortion or inefficiency.

There are two figures of merit for shearing: the achievable frequency shift and the efficiency of that shift. The first is given by $\Delta\omega_{op}$ divided by the spectral linewidth of the input pulse $\Delta\nu_{pulse}$, which we define to be the full-width half max in the spectral domain for the purposes of this analysis. The second is the efficiency or distortion of the shift. As shearing is not necessarily periodic and thus the transmission optical spectrum is not intrinsically band-limited, the exact useful definition of efficiency is application dependent. For purposes of this analysis, we define efficiency as the overlap between the transmitted frequency shifted pulse and an ideal frequency shifted pulse $\epsilon = \int E_T(t)E^*_{T,ideal}(t)dt$.

Two control variables can be utilized to maximize the performance of shearing, f_{RF} and ϕ_0. f_{RF} is principal set as a function of the temporal length, T, of the pulse that one wishes to frequency shift. If the RF period $1/f_{RF}$ is small compared to T, the pulse will experience a larger nonlinear segment

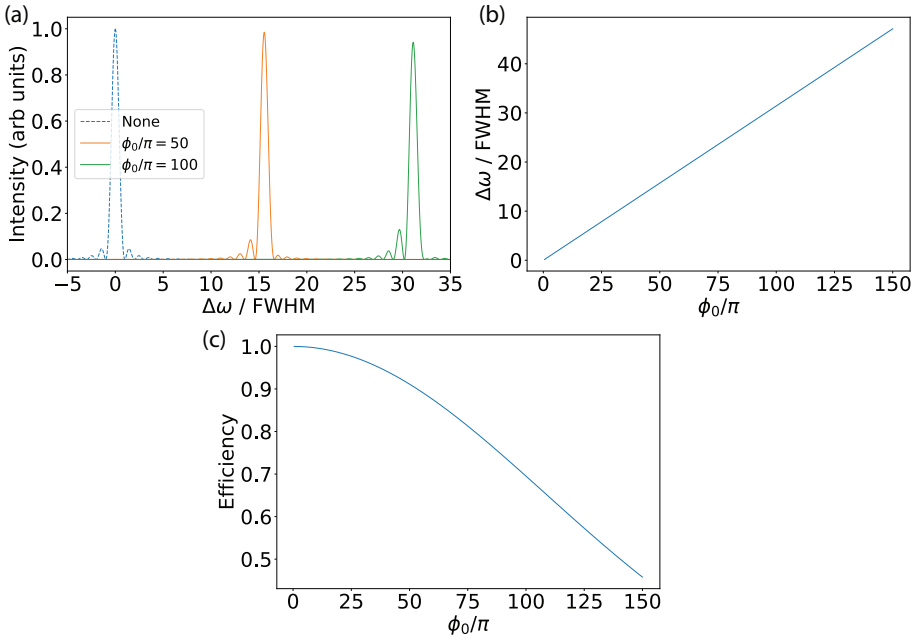
of the sinusoidal tone (t^3 terms in the Taylor expansion) and thus experience a distortion. This is also why the ratio $\Delta\omega_{op}/\Delta\nu_{pulse}$ is taken as the figure of merit as opposed to the absolute shift $\Delta\omega_{op}$, since in principle the optical pulse can be shrunk to increase the absolute magnitude of the shift, but this will increase the ν_{pulse} and thus the ratio provides a fairer metric.



Supplementary Fig. 1 Shearing vs RF frequency. Optical frequency spectra of the transmitted light for various RF shearing frequencies. We see that when f_{RF} is larger a larger shift occurs, but the shape of the line in the optical frequency domain becomes distorted. (b) Optical frequency shift $\Delta\omega$ versus applied RF frequency, showing an increased achieved shift with a higher RF frequency. (c) Shift efficiency versus applied RF frequency, showing a decreased efficiency at higher RF frequencies due to increased nonlinear distortions.

To numerically evaluate this effect, we compute the transmitted frequency spectra of $E_T(t)$, for a fixed $\phi_0 = 20\pi$ and a square wave optical input with temporal duration T and for a variable RF frequency f_{RF} (Supplementary Fig 2). While $f_{RF} = 0.2/T$ achieves double the frequency shift of the case of $f_{RF} = 0.1/T$, it is significantly more distorted. To quantify this trade off $\Delta\omega_{op}/\Delta\nu_{pulse}$ and the efficiency are computed as a function of $f_{RF}T$, thus showing a fundamental tradeoff between the two. As a compromise, for the experiment in this work we pick $f_{RF} = 0.1/T$ as this ensures we have maximized f_{RF} without any significant penalty to the efficiency.

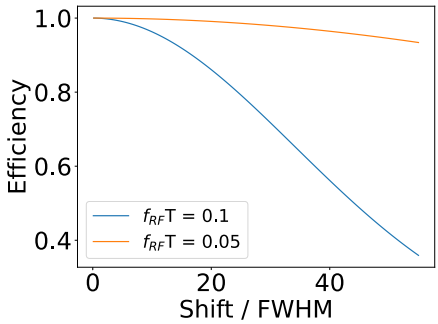
In terms of the effect of the phase modulation index ϕ_0, one might intuitively think that the goal should be to maximize ϕ_0 based on the practical limit of applicable RF voltage and minimizing V_π of the modulator. Instead we observe that increasing ϕ_0 itself also non-negligibly leads to distortion in the signal (Supplementary Fig 2). Indeed, numerically we find for the shearing parameters used in this work, if we wanted to further increase ϕ_0 to achieve $\Delta\omega_{op}/\Delta\nu_{pulse} > 100$ would induce a significant penalty to the efficiency. To understand this, we see that in the taylor expansion of $\phi(t)$, the highest order



Supplementary Fig. 2 Shearing vs drive amplitude. Optical frequency spectra of the transmitted light for various phase amplitudes ϕ_0. We see that increased ϕ_0 leads to an increased optical shift but also increased distortions. (b) Optical frequency shift $\Delta omega$ versus ϕ_0 , showing an increased achieved shift with larger amplitudes (c) Shift efficiency versus ϕ_0, showing a decreased efficiency at higher drive amplitudes due to increased nonlinear distortions.

error term is $\phi_0 1/6 f_{RF}^3 t^3 \propto \phi_0$,. This thus provides another challenge in achieving larger shifts with shearing than just increasing the voltage, and makes it challenging to achieve $\Delta\omega_{op}/\Delta\nu_{pulse} > 100$ while simultaneously achieving high efficiencies.

There are several possible ways to get around this. First we note that while the error term $\phi_0 1/6 f_{RF}^3 t^3$ has a linear dependence on ϕ_0 it has a cubic dependence on f_{RF}. Thus this term can theoretically be decreased through decreasing f_{RF} and correspondingly using a larger ϕ_0 (Supplementary Fig. 3), with the clear disadvantage that an even larger ϕ_0 is required. To contextualize this, we estimate that to achieve a shift $\Delta\omega_{op}/\Delta\nu_{pulse} > 100$ with efficiency $> 90\%$ would require $f_{RF} = 0.04/T$ and $\phi_0 = 160\pi$, corresponding to a voltage amplitude of $\approx 240V$ for the phase modulators used in this work. As an alternative strategy, one could use a more complex waveform consisting of the summation of multiple sinusoidal harmonics whose taylor expansion has higher order cancellation: e.g. $\phi_2(t) = \phi_0[sin(2\pi f_{RF} t) + 1/8 sin(2\pi(2f_{RF})t)] \approx \phi_0[3/4(2\pi f_{RF})t - 1/40(2\pi f_{RF})^5 t^5 + ...]$. These more complex waveforms can also be understood as a higher harmonic approximation to a sawtooth wave



Supplementary Fig. 3 Efficiency vs Optical Shift. Efficiency versus normalized optical shift for two different frequencies f_{RF}. Shift efficiencies above 90% for optical shifts $\Delta\omega_{op}/\Delta\nu_{pulse} > 50$ are theoretically achievable using a low f_{RF}.

and thus as a sort of hybrid between serrodyne and shearing, with the disadvantage the additional complexity required as well as a larger ϕ_0 required to achieve a given shift.

1.1.1 Suitability of Shearing for Multiplexing Quantum Nodes

In the experimental results shown, we have shown the ability to perform high efficiency shearing for $\Delta\omega_{op}/\Delta\nu_{pulse} > 10$. For compensating for the inhomogenous distribution for photons that have been emitted and entangled with the quantum memories, this is insufficient due to the large mismatch between photon lifetimes and inhomogenous distributions. For example, the silicon vacancy center in diamond has a intrinsic linewidth of ∼100 MHz but an inhomogenous distribution on the order of 10 GHz and thus requires $\Delta\omega_{op}/\Delta\nu_{pulse} > 100$. Given the analysis in the above section, it is clear that increasing this to shifts beyond $\Delta\omega_{op}/\Delta\nu_{pulse} > 100$ will require a greater than 10x increase in the achievable phi_0 to compensate for the additional distortions introduced by increasing the drive voltage. Thus it does not seem practical to use shearing for frequency shifting the transmitted entangled single photon signals without significant advances in modulators (potentially using resonant modulator designs or other advances) or on the emitter side to decrease the inhomogenous distribution. Despite this, frequency shearing on the input control side to use a common laser to address a large number of color centers with varying frequencies seems feasible. This is especially true for emission based schemes where excitation optical pulses much shorter than the emitted photon lifetime are used to excite quantum memories and thus have significantly relaxed $\Delta\omega_{op}/\Delta\nu_{pulse}$ requirements and less efficiency and purity requirements as well.

1.2 Modeling and Limits of Serrodyning

In serrodyning, a sawtooth wave with an amplitude of 2π and frequency f_{ST} is applied to the modulator. This emulates a linear wave as a phase of 2π is equivalent to a phase of 0 so assuming the drop off region of the sawtooth

is instantaneous (as in an ideal sawtooth) the sawtooth is equivalent to a linear increase in phase, and nominally frequency shifts the input light by a frequency f_{ST}. Since no temporal pulsing is required, there is no intrinsic trade off between achievable frequency shift and initial pulse linewidth that is observed with shearing - even a CW tone can be efficiently serrodyned.

The challenge with serrodyning is the generation and signal integrity of the electrical sawtooth wave. The sawtooth wave is a intrinsically high-bandwidth signal, which can be observed either in the time domain due to the sharp drop in voltage at the teeth of the sawtooth, or in the frequency domain where the sawtooth wave can be decomposed into a Fourier series -

$$\phi(t) = \phi_0[1/2 - 1/\pi \sum_{n=1}^{\inf} \frac{1}{n} sin(n\pi t \cdot f_{ST})]$$

where ϕ_0 is the modulation depth, and f_{ST} is the frequency of the sawtooth wave. We observe a very modest roll-off $\propto 1/n$ in the Fourier component amplitude thus requiring a large bandwidth to faithfully reproduce the sawtooth wave. Distortions in the sawtooth wave when applied to the modulator causes frequency content to partially shit to other undesired frequency bins, thus decreasing the efficiency.

For generation of the sawtooth wave, we choose to use a digital approach using a high-speed arbitrary waveform generator (AWG). This is advantageous compared to the previously utilized analog approaches since the AWG allows arbitrary modification of the signal in order to compensate for distortions due to frequency-dependent loss in transmission. One potential issue with the AWG is its temporal discretization of the sawtooth waveform, though numerically we will find that the sample clock of the utilized AWG is sufficiently fast that its effect is minimal.

From the output of the AWG, the sawtooth wave is then directed into a broadband amplifier to amplify the signal, transmitted through cables, and then applied to the modulator. When designing this signal path there are two primary concerns. First the amplifier must have enough gain and saturation power so it can fully drive a 2π phase shift on the modulator. Having a modulator with a low V_π is crucial to easily meet this requirement since achieving a high output power and simultaneously high bandwidth RF amplifier is difficult. Second, there will be frequency dependent losses through all components including the amplifier, cables, and the modulator itself which will introduce distortions into the signal. It is important to minimize these losses through through utilizing short, high efficiency cables and engineering a high-bandwidth modulator. When these are well engineered to be on the order of a few dB, the distortions from the modulator and cable can be effectively compensated via pre-distortion from the AWG. The amplifier, however, has a much sharper roll-off at its cutoff frequency that cannot be effectively compensated for, and thus it is what ultimately limits the achievable shift efficiencies particularly at larger frequency shifts (higher f_{ST}).

1.2.1 Serrodyne Modeling

To quantitatively understand the limitations of serrodyning, we model the effect of distortions of the sawtooth wave $\phi(t)$ on the frequency shift efficiency. We consider an electrical sawtooth wave with frequency f_{ST} aiming to impart an optical frequency shift of $\Delta\omega = f_{ST}$ to the input CW wave. The primary figure merit we care about is the shift efficiency (ϵ) into the target frequency bin as a function of f_{ST}. As $\phi(t)$ is periodic, the transmitted wave $E_T(t) = Ae^{j(\omega t+\phi(t))}$ is also periodic with period $1/f_{ST}$. Thus the frequency spectrum of the transmitted field is discrete ,

$$\|\mathcal{F}(E_T(t))\|^2 = \sum_{n=-inf}^{inf} A_n \delta(\omega + 2\pi n f_{ST})$$

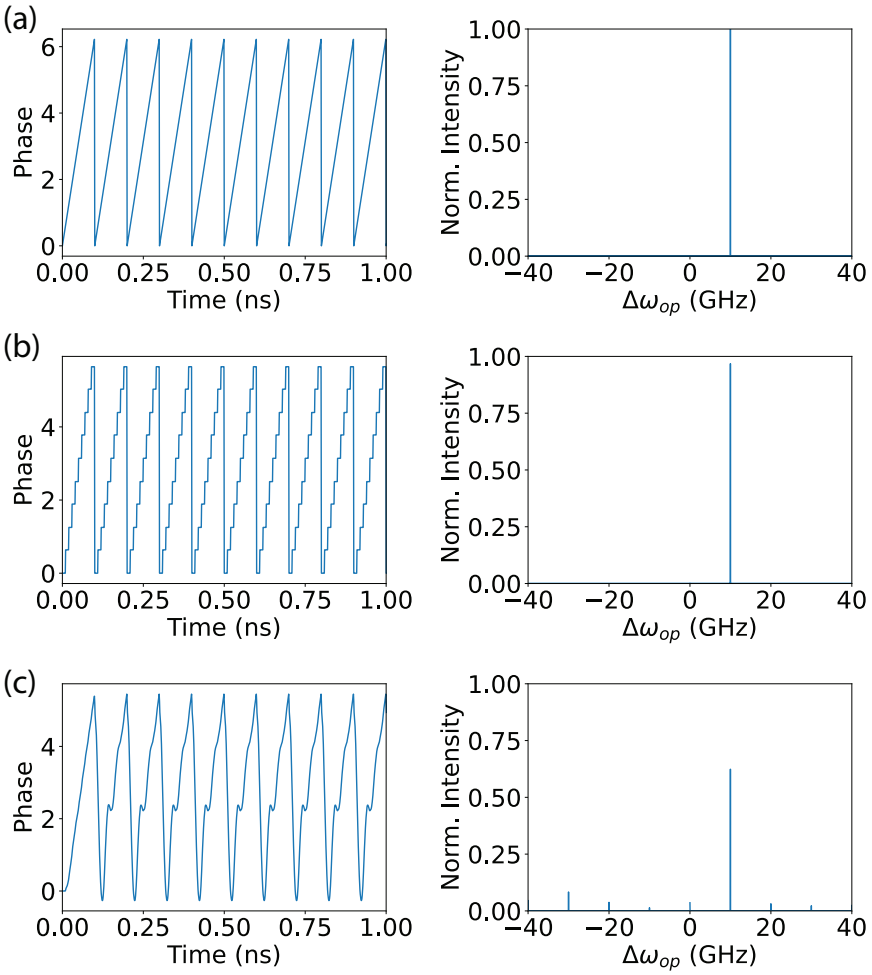
where $\mathcal{F}$ is the Fourier transform operation. As the goal is to frequency shift into the n=1 bin, the efficiency can be trivially defined as

$$\epsilon = A_1 / \sum_n A_n$$

Thus through this set of equations, the shift efficiency can be easily computed from a modeled $\phi(t)$.

In terms of modelling $\phi(t)$ we used a three step process. First $\phi(t)$ was assumed to be an ideal sawtooth wave with amplitude 2π and frequency f_{ST} (10 GHz in this initial example), and from this $\|\mathcal{F}(E_T(t))\|^2$ and the efficiency ϵ of the frequency shift are computed (Supplementary Fig. 8a). This is the "ideal" case and a shift of unity efficiency is expected. After this, we incorporate the effect of the digital emulation of the electrical waveform, namely temporal discretization. We temporally discretize the sawtooth wave at a sampling rate of $f_{AWG} \approx 100GSa/s$ With this we can recompute $\|\mathcal{F}(E_T(t))\|^2$ and the efficiency ϵ, and observe that although there is some penalty to efficiency due to this discretization, the sampling rate is high compared to f_{ST} so this penalty is not particularly significant (Supplementary Fig. 8b).

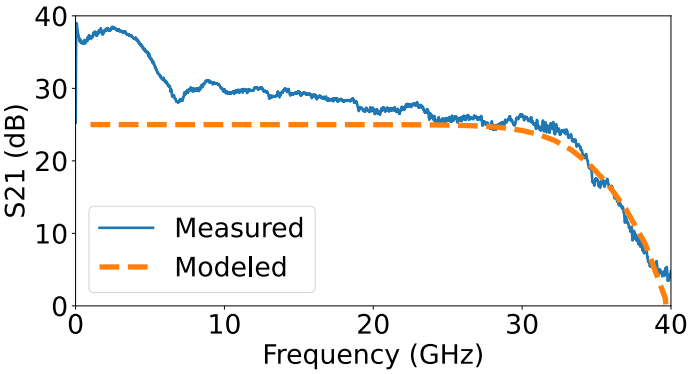
Finally we want to model the effect of our system's finite analog bandwidth on $\phi(t)$ and the serrodyne efficiency. To do so we apply a filter function $H_{sys,eff}(f)$ to $\phi(t)$ where $H_{sys,eff}(f)$ is the effective transfer filter function of the system. Determining an effective $H_{sys,eff}(f)$ is somewhat challenging, as although the overall system function H_{sys} is fairly simple to measure with standard laboratory equipment, the AWG's ability to perform pre-emphasis to partially eliminate distortions from this transfer function makes it more challenging to provide an effective $H_{sys,eff}(f)$. As discussed previously, the actual transfer function of the system H_{sys} includes a gradual decrease in transmission efficiency versus frequency due to a combination of variations in the passband of the amplifier as well as increased cable attenuation at higher frequencies. In addition to this, H_{sys} has a much steeper cutoff at 34 GHz due to cutoff of the amplifier's bandwidth (Supplementary Fig. 5). In order



Supplementary Fig. 4 Serrodyne Modeling. Applied phase vs time (left) and corresponding optical frequency spectrum (right) for three different serrodyne models: (a) ideal, (b) temporally discrete, and (c) low pass filtered. The introduction of a finite analog bandwidth and a corresponding filter function induces a significant penalty in the achieved shift efficiency

to provide a simple model for $H_{sys,eff}$, we assume that pre-emphasis on the AWG is able to effectively pre-compensate the small variations below 30 GHz and only fails to compensate for the steeper rolloff form the amplifier beyond 34 GHz. Thus we approximate $H_{sys,eff}(f)$ as be a type II Chebyshev filter whose frequency cutoff and order are set such that the rolloff matches the measured system rolloff accurately (Supplementary Fig. 5). A type II Chebyshev filter is used due to the combination of steepness in response at the cutoff frequency, to accurately mimic the measured steep cutoff, and flatness in the

passband to mimic ideal pre-compensation in the passband. We observe that the application of $H_{sys,eff}(f)$ to $\phi(t)$ leads to a significant distortion leading to a significant degradation in ϵ, thus indicating this is the most significant practical penalty in the system (Supplementary Fig. 8c).



Supplementary Fig. 5 Amplifier Fitting. Measured transmission versus frequency of the system and corresponding filter used as a numerical approximation.

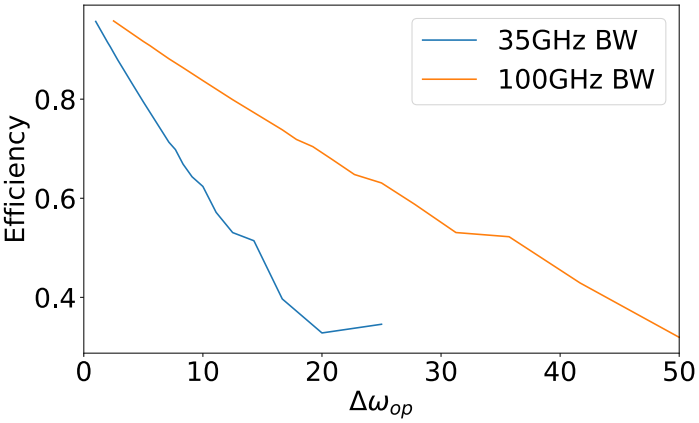
1.2.2 Serrodyne Performance

Using this model, we extract the shift efficiency ϵ as a function of optical shift frequency $\Delta\omega = f_{ST}$ (Supplementary Fig. 6). We observe we accurately match our measured serrodyne efficiencies (see main text Fig 4d). This both verifies the integrity of our model and indicates that in practice the analog bandwidth of the overall system, which is limited by the bandwidth of the broadband power amplifier, is the main limitation in the overall serrodyne efficiency.

To understand how much improvement can be expected via serrodyning with state-of-the-art high-speed electronics, we use our validated model to estimate the efficiencies achievable with serrodyning with $f_{AWG} = 250$GSa/s and an analog bandwidth of 100 GHz [1, 2]. We observe even more significant improvements in the realized efficiencies are achievable, with shifts of efficiency greater than 50% possible up to 30 GHz and greater than 90 % up to 10 GHz . This demonstrates that electro-optic frequency serrodyning is a promising technique for overcoming the inhomogenous distribution of solid-state emitters.

1.2.3 Precision and Phase Noise of Electro-Optic Frequency Shifting

The precision and associated infidelity of electro-optic frequency shifting is ultimately set by the temporal jitter (or alternatively the phase noise) associated with the electrical drive signal. One way to quantify the effect is to evaluate the effective nominal linewidth of the AWG generated electrical waveform which is determined by its phase noise, which can then be considered in



Supplementary Fig. 6 Serrodyne Efficiency Calculation. Computed achievable serrodyne frequency shift efficiency as a function of shift amount for 35GHz and 100GHz analog bandwidths.

the frequency domain to be convolved with the optical linewdith thereby introducing noise and setting a maximum frequency resolution. Given the AWG's specified phase noise of $S_v = -95$dBc/Hz at 10kHz offset (Keysight M8196), assuming a flat noise spectrum up to 10kHz we can estimate the rms frequency noise of $v_{rms} = \sqrt{\int S_v dv} = 17$Hz. This is orders of magnitude below the typical linewidths of solid-state emitters which is typically on the order of MHz, and thus not a significant infidelity. In addition, this could be further improved through use of a lower-jitter oscillator by the AWG.

1.3 Power Dissipation of Electro-Optic Shifting

One technical point of concern with electro-optic shifting is the amount of power dissipated on chip with the different methods for shifting. Too much dissipation can limit deployment of these techniques for a large number of memories or in a cryogenic environments.

For serrodyning, the power consumption can be fairly easily calculated as the modulator is terminated in a 50 Ω load. Thus the average power consumption is given by $P_{serr} = \frac{1}{T_{ST}} \int_0^{T_{ST}} v(t)^2/50\Omega$ where $T_{ST} = 1/f_{ST}$ is the period of the sawtooth wave and the voltage across the modulator is given by $v(t) = V_\pi \cdot \phi(t)/\pi$. Evaluating this integral for $\phi(t) = 2\pi(t/T_{ST} - 1/2)$ (one period of a sawtooth wave) we arrive at :

$$P_{serr} = \frac{1}{3} \cdot \frac{V_\pi^2}{50\Omega}$$

. As this is total power consumption, power dissipation is given by the proportion of this power which is lost when transmitted across the modulator's electrodes. While doing an exact calculation for this will be electrode and frequency dependent, due to the relatively high loss of the modulator's electrodes

particularly at high frequencies, a large proportion of the incident power will likely be absorbed. Thus for purposes of an upper bound, we can assume the on-chip dissipated power is equivalent to P_{serr}. For the modulator demonstrated in this work, $V_\pi \approx 1.75V$ giving a total power consumption of $\sim$20 mW. We note this is likely a pessimistic estimate, particularly at smaller f_{ST}.

For shearing, although peak to peak voltages much greater than V_π are required, the narrow-band and relatively low frequency nature of the electrical waveform ensures that the modulator electrode can be used in a lumped element configuration, significantly decreasing the power dissipation. To calculate this dissipation, We model the modulator as a lumped element R-C circuit with resistance R_{mod} and capacitance C_{mod}. In the limit where the frequency of the applied bias $f_{shear} << R_{mod}C_{mod}/(2\pi)$ the power dissipated across the resistance R_{mod} is given by

$$P_{shear} = R_{mod} \cdot (2\pi \cdot f_{shear} C_{mod} V_{RMS})^2$$

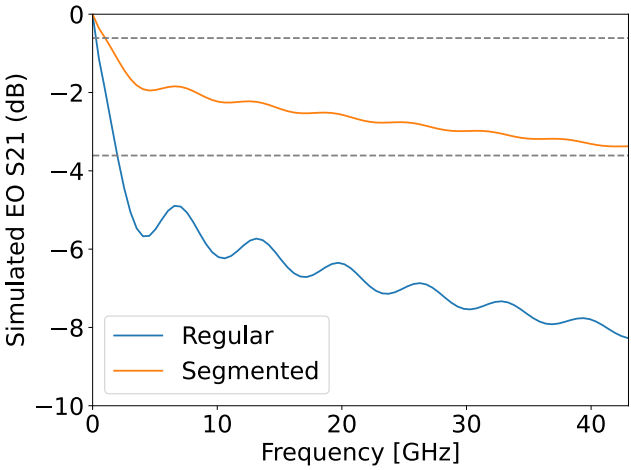
. For the shearing demonstration in this work, $R_{mod} \approx 5\Omega$, $f_{shear} = 250$ MHz, $C_{mod} = 2$pF (based on modeling of the measured electrode parameters), $V_{RMS} =\sim 35$V so we obtain $P_{shear} \approx 60$ mW.

1.4 Segmented Electrode Parameters

In order to perform electro-optic frequency conversion, it is important that the electro-optic response of the modulator is maximized at larger frequencies, particularly for shearing. In our previous work [3], our modulator utilizing a standard co-planar waveguide design had a significant initial rolloff in EO response due to a large impedance mismatch, as the center trace had to be made small in order to match the RF index to the group velocity of the waveguide (≈ 2.4). To overcome this, we chose to utilize segmented electrodes for our modulator. These segmented electrodes increase the microwave index, allowing us to better match the microwave index with the waveguide index (with the addition of a thicker buried oxide of 7 μm). In addition, this new design has the additional benefit of significantly decreased RF loss due to the breaking of the inner conductor current path (see Table 1). These factors combine to enable an enhanced EO response.

	Regular	**Segmented**
Impedance (Ω)	36	40
RF Loss (dBcm^{-1}GHz$^{-1/2}$)	1.31	0.73
RF Index	2.41	2.4

Supplementary Table 1 Modulator electrode parameters for our previous work [3] and this work



Supplementary Fig. 7 Simulated electro-optic response as a function of frequency for regular and segmented electrode modulator, demonstrating a significantly improved rolloff for the segmented design.

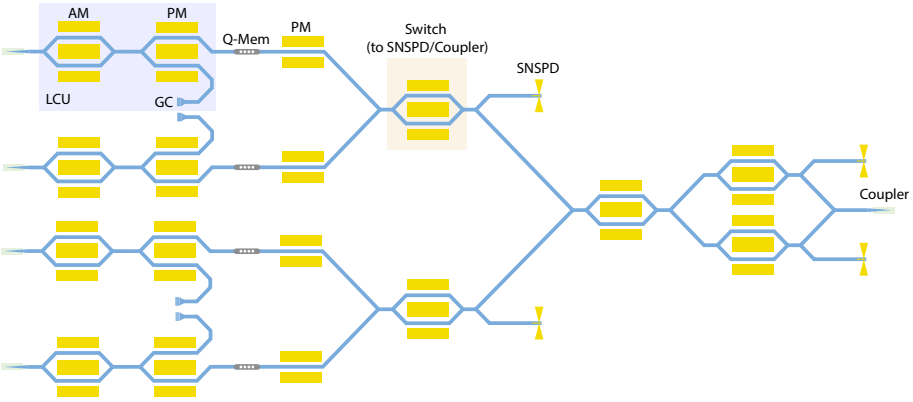
2 System Considerations

2.1 Photonic Integrated Circuit Schematic and Explanation of Requirements

A technical diagram of our proposed TFLN architecture is shown in Supplementary Fig. 8. An input CW laser light is first transmitted through a laser control unit (LCU) that is used to generate and shape the temporal and spectral profile of the optical pulse that control the quantum memory (QM). LCU consists of an amplitude modulator (AM) that carves the optical pulse out of CW light, and a phase modulator (PM) that enables frequency shifting of a common seed laser light to address different, inhomogeneously broadened, QMs. Next, optical pulse is transmitted to the QM , and spin-photon entanglement is generated through one of many schemes [4–6]. The spin-entangled photon is then passed through a Multiplexing Unit (MU) that consists of a PM, switch network, fiber couplers and on-chip photodetectors. PM is used to shift its frequency of spin-entangled photon to a common frequency shared by all of the individual quantum memories within a given memory node, thereby overcoming the inhomogeneous distribution challenge [7, 8]. The common-frequency photons are routed using a switch network either to an on-chip photon detector [9] to enable high-efficiency on-chip heralded entanglement between memories within the same node, or an output coupler so that photons from different memories can be efficiently multiplexed across the wider fiber network.

Note that we assume the use of nonlinear-optic based quantum frequency conversion (QFC) outside of the TFLN PIC to convert entangled photons from visible to telecom wavelength to enable transmission across long-distances of

fiber. Although in the future the integration of periodically poled lithium niobate (PPLN) could potentially enable nonlinear optical frequency conversion directly on the PIC [10], for the purposes of this work we assume it is performed separately from the PIC using commercially available conventional PPLN waveguides realized in bulk crystals. The reason for this is that bulk QFC systems have already demonstrated impressive performances sufficient to realize large-scale quantum networks [11, 12]. Furthermore, in the architecture proposed in this work, multiplexing is performed at visible wavelengths and therefore only a single QFC stage is required per fiber channel (as opposed to the 1000s of memories and switch elements required per fiber channel). Although the integration of thin film PPLN could enhance the performance and versatility of our platform, it may not be required to realize large-scale quantum networks proposed in this work.



Supplementary Fig. 8 Schematic of LNOI multiplexed quantum node.

2.2 Quantum Memory Compatibility

The main limitation for compatibility of various color center memories with our platform is their photon emission wavelength. Due to the broadband nature of our platform, with no resonant devices used, visible TFLN PICs can be used across a wide wavelength range with only some small adjustments required in waveguide geometry and potentially coupler geometry. We note that in this work we have demonstrated device operation down to wavelengths of 600nm (limited by measurement equipment) although other works have shown device operation at lower wavelengths such as 450nm [13]. This requirement of wavelength greater than 600nm, but less than 1100nm (preventing use of conventional silicon photonics) includes a significant portion of promising quantum memory platforms, with a non-exhaustive list included in Table 2.

In terms of other requirements for compatibility, the second main one is that

the photon length must be longer than the rise-time of the switches, otherwise they can not be maximally multiplexed to eliminate any dead time on the channel. We note that of the quantum memories listed in Table 2, the shortest lifetimes are of the solid-state emitters which have lifetimes on the order of ns [14, 15], which is significantly longer than the 100ps pulses demonstrated in this work. Thus our platform is fast enough to ensure compatibility with a large variety of candidate quantum memory platforms, including quantum dots, color centers, and trapped ions.

Quantum Memory	Type	Wavelength	Ref
SiV in Diamond	Color Center	737nm	[14]
NV in Diamond	Color Center	637nm	[14]
GeV in Diamond	Color Center	602nm	[14]
SnV in Diamond	Color Center	620nm	[14]
V_{Si}^{-} in SiC	Color Center	862-917nm	[16]
$^{171}Yb^{3+}$ in YVO	Ion	984.5nm	[17]
$^{40}Ca^{+}$	Ion	850nm	[18]
^{87}Rb	Atom	780nm	[19]
InGaAs	Quantum Dot	940nm	[20]
GaAs	Quantum Dot	850nm	[21]

Supplementary Table 2 Selection of quantum memories compatible with our visible TFLN platform. The list is non-exhaustive.

2.3 Electro-Optic Bias Stability of Thin Film Lithium Niobate

Drift in the operational bias point of lithium niobate modulators is a known issue in both bulk and thin film lithium niobate platforms whereby the bias point of a modulator will drift on slow timescales once an electrical bias is applied [22, 23]. This effect can be understood in the frequency domain as a rolloff in the electro-optic response of the modulator at low frequencies, effectively low-pass filtering the output and thereby leading to long timescale drifts in the operational point of the modulator. While a full discussion of the microscopic origin of this effect is beyond the scope of this work, recent works have shown the ability to minimize the effect of low frequency rolloff through optimization of the device and processing conditions [22, 24].

For purposes of the PIC platform presented in this platform, it is important to minimize the bias point drift so that the modulators can be biased purely using electronics. This is required since the alternative, thermal biasing, is prohibitive for operation in cryogenic environments. Thus it is important that this bias drift is minimized in our devices. To achieve this, we optimize the bias drift in our devices through A) placing the bias electrodes directly on the LN and B) applying a low temperature (300 C) overnight anneal in an ambient environment after patterning the electrodes [22, 24]. This increases the timescale of bias drifts to the order of seconds or beyond. At these time

scales, electrical feedback can be implemented to effectively lock the bias of the modulator to a given operating point.

To demonstrate the feasibility of electrical locking the bias point of the modulator, we electrically locked our TFLN amplitude modulator at the null point when generating pulses for shearing (see Fig. 5d). To do this, we applied a small sinusoidal dither on the modulator via a bias-tee network, and tapped off a small portion of the optical signal transmitted through the modulator and detected it on a photodetector. This photodetector was then demultiplexed at the dither frequency via a lock-in amplifier to provide an error signal for locking. This error signal was digitized via an analog-digital converter and read in to a program running on a host PC which implemented a basic PID loop to stabilize the error signal. We emphasize that the PID loop is implemented in software with a slow update loop (ms timescales), thus highlighting the simplicity of the approach. This feedback loop successfully prevented drift of the modulator bias for the duration of the experiment (1 day) and enabled the data shown in Fig. 5d.

We believe that this feedback approach is scalable to ensure optimal bias stability of larger PICs and specifically larger switch networks. In order to inject and pick off feedback signal for each modulator, low loss waveguide taps can be used, and potentially multiple sets of modulators can share a single tap through multiplexing of different dither frequencies, thereby minimizing the number of taps needed. On the electronics side, the relatively slow nature of the feedback loop ensures it can be effectively implemented using inexpensive microcontrollers, enabling easy scalability. The final practical challenge in implementation of feedback is co-existence of the feedback signal with the single photon signal for generating entanglement. This can be implemented either through implementing feedback at a detuned wavelength which can be filtered before the single photon detectors, or via intermittently pausing entanglement generation to allow for injection of a feedback signal. We note that both techniques are already readily implemented in state-of-the-art quantum networking experiments to lock various components (see [12]) and thus is not technically prohibitive.

2.3.1 Effect of Bias-Drift on PIC Operation

Two elements are affected by bias drift in our proposed PIC architecture - the amplitude modulator in the LCU, and the switches in the switch network. For the former, bias drifts will limit the effective extinction ratio achievable when pulse shaping with the modulator. The sensitivity of the protocol to this extinction ratio is very protocol-dependent. For the switch network, bias drifts will lead to imperfect switching whereby instead of only routing photons to a single port, the switch will partially route photons to the alternative point. We note the effect of this is an additional inefficiency in the switch as opposed to an impact on the overall fidelity of the operation, thus ensuring a

larger tolerance to bias drifts in the switch network.

2.3.2 Photorefractive Effects

One concern with short-wavelength LN modulators is photorefractive effects that can occur at mW power levels [13]. However, the optical power required to control and read out solid-state quantum emitters is orders of magnitude smaller, in the uW range and single photon range, respectively, and therefore we do not anticipate issues with the photorefractive effect.

2.4 Power Consumption and Dissipation of Switching

One major component of the power consumption and dissipation of the system is that of the switching elements. As the switches are capacitively loaded (not 50 Ω terminated) they can be modeled as an RC network where R_{tot} is given by the total resistance of network, given by the sum of the source and system resistance R_{sys} and the device resistance R_{sw}, and the capacitance is given by the device capacitance. The dynamic power consumption is given by $P_{cons} = fCV^2$: assuming a switching frequency f and with a voltage $V = V_\pi$ of the switch. The actual dynamic power dissipated on chip is given by $P_{diss} = R_{sw}/R_{tot} \cdot P_{cons}$, as it is a simple resistive power divider between the system and switch resistances.

For the switch presented in the main text, we estimate $C = 1$pF, given by estimation of the distributed capacitance using the measured transmission line parameters and telegrapher's equations, and $R_{sw} = 5\Omega$. Assuming $R_{sys} = 50\Omega$ and given the measured $V_\pi = 1.5$V we estimate a dynamic power consumption of $\approx 200\mu W$ and an on chip power dissipation of $\approx 20\mu W$ for a switching frequency of $f = 100$ MHz.

This can likely be substantially improved through only minor modifications to the geometry. First, the geometry can be shrunk to 1mm. Although this will increase $V'_\pi = 7.5V$, we estimate a decreased $R'_{sw} = 1\Omega$ and $C = 0.2$pF. Moreover, we can artificially increase the system resistance R_{sys} so that a larger proportion of power consumption is dissipiated at the system as opposed to on-chip. This is important for thermal management of the chip, particularly in cryogenic environments. The cost of increasing R_{sys} is the maximum frequency response of the system decreases as it is set by the RC time constant $R_tot \cdot C$, but since only a moderate switching frequency of 100 MHz or so is desired, R_{tot} can be increased significantly without a performance penalty. Thus we assume $R_{sys} = 5k\Omega$. This optimized set of parameters enables an on-chip power power dissipation of $\sim 0.2\mu W$. We note that the estimated total power consumption (as opposed to on-chip dissipation) of this optimized geometry actually increased to $\sim$ 1 mW primarily due to the larger V'_π due to the shorter length. This thus indicates a partial trade-off between in the switch performance between total power consumption, on-chip power dissipation, and footprint that has to be evaluated on a system by system basis.

We also note that the on-chip power dissipation, and likely the total power consumption of the switches are considerably smaller than the power consumption needed for electro-optic frequency shifting (see Section 1.3) and thus that will ultimately dominate the on-chip power dissipation for any full-scale realization of our system. If further improvements to power dissipation are required, the integration of superconducting electrodes in the modulator could potentially improve this further.

2.5 Co-packaging PIC with Quantum Memories

Heterogeneous integration of a quantum memory module with an active PIC (in this case TFLN) has often been considered as a solution to integrate active switching functionality into a passive quantum memory platform such as diamond. Although there have been a variety of works showing capabilities to achieve this and the advantage of this approach is clear, namely minimizing the net insertion loss at the quantum memory - active photonics interface, this direct integration approach has numerous practical problems. These practical issues generally stem from the fact that the quantum memories have stringent cryogenic requirements, with state-of-the art quantum memories needed to operate at 1K or even below to achieve good coherence properties. This cryogenic requirement is largely not compatible with the thermal and size requirements of the PIC. More specifically:

1. Power dissipation of switching is not insignificant. We estimate our current switches dissipate 20 μW of power. For $\sim$1000s of memories and thus switches required to maximize a quantum link's entanglement rate (see 2.6), that corresponds to 10s mW of dissipated power, which is a similar order of magnitude to the cooling power of 4K systems. In addition, the implementation of electro-optic frequency shifting will lead to the dissipation of 10s of mW of power per memory, indicating that cooling power on the order of W is required. In addition, these estimates do not include resistive heating via the cabling which likely will be significant.
2. Electrical control of the switches requires extensive electrical wiring. Assuming complete independent control of each modulator is required to achieve the desired multiplexing and local operation scheme, each modulator requires an independent control line. As a full PIC will likely need to include 1000s of modulators in order to achieve optimal temporal multiplexing, this requires 1000s of electrical wires. The heat conductivity of wires is large, and thus adding wiring from the PIC at the base of the cryostat to a controller in the external ambient environment could easily lead to a prohibitive heat load. Potentially integration of the CMOS controller within the cryostat either at the base plate or a higher temperature stage could solve this, though cryogenic compatible CMOS technology is still at early stages.
3. The footprint of the modulators is comparatively large, with lengths on the order of -mm to even -cm, and lateral density of 100s of μm limited

by optical and electrical cross-talk. Thus the total space in the cryostat occupied by these switches is quite large.

In contrast, 2D fiber arrays with pitches down to 82 μm are commercially available. Thus a more practical alternative could be to utilize high-density optical packaging and leave the active PIC outside of the cryostat. In particular, in architectures which coupling loss can be rolled into total channel loss between quantum nodes, dB level excess losses that might be incurred from this massive coupling approach can be tolerated if it enables enhanced scalability. Alternatively a hybrid solution, with limited switching in the cryostat both in number of stages and connectivity and the rest in an external PIC at ambient temperature, might provide a best of both worlds solution.

2.6 Multiplexing Theory

We construct a simple model of remote entanglement generation in order to understand the benefits of multiplexing enabled by the TFLN PIC.

Assume there are two nodes separated by a distance L each with N quantum memories with a measurement station in the middle at a distance $L/2$ from either node. At the start of the protocol which is repeated at a clock rate f_{clk}, each node probabilistically emits a spin-entangled photon with duration t_{photon}. The photons then are sent to a middle measurement node in which a partial (1 photon) or full (2 photon) bell state measurement occurs on the photon(s). The successful heralded detection of 1 (2) photons indicates entanglement was successfully generated. The probability that a given attempt leads to a heralded entanglement event is given by p_{entang}

The overall entanglement rate R_{entang} is given by the product $R_{entang} = p_{entang} \cdot f_{clk}$. In the case of $N = 1$ memories per node and no multiplexing, the maximum possible clock rate f_{clk} is determined by the time between a protocol attempt and determining whether that entanglement generation was successful. This is since there is only a single memory per node, and thus another entanglement generation attempt cannot be started as it would require re initializing that memory and thus eliminating entanglement before it is even generated. This temporal delay is determined by the combination of time it takes for the photon to arrive from the two quantum memory nodes at the central measurement node, and the time it then takes for the classical signal to propagate back indicating whether entanglement was successfully generated Thus we have $1/f_{clk,N=1} = 2 \cdot \frac{L}{2} \cdot \frac{1}{c'} = \frac{L}{c'}$ where c' is the effective speed of light in the propagation medium.

When the number of memories is increased $N > 1$, the clock rate can be increased through temporal multiplexing of memories on the channel. After a photon is entangled with one memory and sent and the node is waiting for the result, an entanglement attempt can be performed with the second memory, and so on until all of the memories at a given node have been utilized and are waiting for their respective results. Thus although the delay time between entanglement attempt and determining whether it was successful is still the

same ($\frac{L}{c'}$), the effective clock rate is increased by the number of memories n: $1/f_{clk,N=n} = \frac{L}{c'.n}$. The limit of temporal multiplexing is ultimately set by the temporal-length of the spin-entangled photon, as only one photon can exist within the channel at a given time (assuming a matched polarization and frequency). This thus sets the maximum clock rate $1/f_{clk,max} = t_{photon}$.

p_{entang} can be broken down into the product of four individual probabilities: p_{photon}: the probability a spin-entangled photon is generated and coupled into single mode fiber ($p_{photon} = p_{photon,gen} \cdot p_{photon \rightarrow fiber}$), $p_{PIC->fiber}$ the probability the photon is transmitted through the PIC into the output fiber, p_{fiber} which is the probability the photon is successfully transmitted across the fiber channel (including the efficiency of frequency conversion from visible to telecom wavelengths - $p_{fiber} = p_{vis \rightarrow telecom} \cdot p_{fiber,T}$), and $p_{protocol}$ which encompasses the probability of entanglement success due to the probabilistic nature of heralded entanglement generation itself. We then see the overall entanglement success probability is $p_{entang} = (p_{photon} \cdot p_{PIC->fiber} \cdot p_{fiber})^{n_{photon}} \cdot p_{protocol}$ where n_{photon} is the number of photons required to be detected to herald entanglement which is protocol dependent.

$p_{PIC->fiber}$ depends on the exact architecture and independent device losses of the PIC and most notably will increase with larger number of memories N as a larger switch network is required to multiplex a larger number of quantum memories. Using the architecture we propose in this work (and described in section 2.1) we determine that the full insertion loss of the PIC is given by $IL_{PIC} = 2 \cdot IL_{coupling} + (n_{switch}) \cdot IL_{switch} + IL_{pm} + IL_{\Delta\omega_{Op}}$. We note that we assume that the PIC and the quantum memory bank are not co-packaged and thus we take double the on-off chip insertion loss penalty. The number of switches n_{switch} is dependent on the number of memories that are multiplexed: $n_{switch} = log_2 N + 1$.

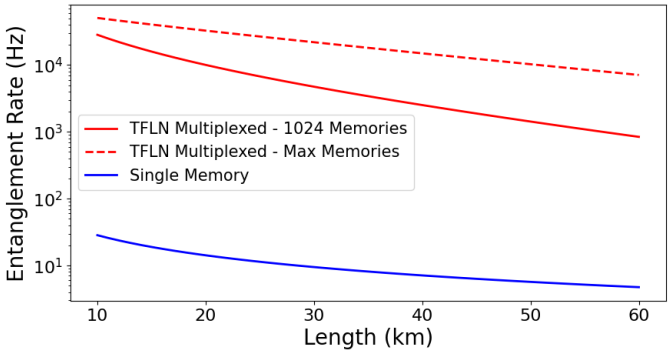
To highlight the benefits of multiplexing we compute R_{entang} as a function of N. In Table 3 we list the parameters used for this calculation. The insertion losses for the computation of $p_{PIC->Fiber}$ were all extracted from the measurements presented in this work. The rest of the parameters assumed use of the silicon-vacancy center in diamond where applicable.

For the comparison case of a non multiplxed quantum node, we fix $N = 1$, $p_{PIC} = 1$ and repeat the calculation. We observe that in spite of the additional insertion losses required to implement multiplexing, the losses of the devices in our platform are sufficiently low such that the benefits of multiplexing outweigh the excess losses. We thus predict over a 100x gain in entanglement rate through multiplexing with our LNOI PIC.

Beyond the calculations in the main text, we also compute the entanglement rate as a function of fiber length (Supplementary Fig 9). We observe that for fiber lengths up to 60km, we achieve an almost three order of magnitude improvement in entanglement rate with TFLN based multiplexing as compared to the single memory case.

Parameter	Description	Value	Notes
L	Node-Node Distance	20 km	
c'	Group velocity of fiber	$2 \cdot 10^8$ m/s	
t_{photon}	Photon Lifetime	50 ns	[4].
n_{photon}	Number of Photon Event	1	
p_{photon}	Probability of Spin-Photon Generation	0.493 × 0.4	Product of demonstrated spin-photon gate efficiencies (Ref [25]) times high μ weak coherent source.
$p_{protocol}$	Protocol success probability	0.25	Protocol probability for 1-photon heralded protocol [26]
$p_{photon\text{-}¿fiber}$	Efficiency of single photon coupling into single mode fiber	0.9 (-0.4dB)	[27]
$p_{vis \rightarrow telecom}$	Efficiency of conversion from visible to telecom wavelengths for long-distance fiber transmission	0.5 (-3dB)	[11]
$p_{fiber,\ T}$	Transmission efficiency through length $L/2$ fiber	0.5 (-3dB)	0.3 dB/km
$IL_{coupling}$	Insertion loss of coupling into LNOI chip	0.8 (-1dB)	This work
IL_{switch}	Insertion loss of LNOI switch	0.9 (-0.4dB)	This work
IL_{pm}	Insertion loss of LNOI phase modulator	0.85 (-0.7dB)	This work
$IL_{\Delta\omega_{Op}}$	Frequency shifting efficiency	0.9 (-0.5dB)	This work

Supplementary Table 3 Parameters used in multiplexing system calculation.



Supplementary Fig. 9 Entanglement rate as a function of fiber length. Computed entanglement rate between two quantum memories separated by a varying length when TFLN based multiplexing is used either with a fixed number of memories (1024) or the maximum number of memories which saturates the channel for a given length of fiber.

References

[1] Chen, X. *et al.* All-electronic 100-ghz bandwidth digital-to-analog converter generating pam signals up to 190 gbaud. *Journal of Lightwave Technology* **35** (2017). https://doi.org/10.1109/JLT.2016.2614126 .

[2] Wakita, H., Jyo, T., Nagatani, M. & Takahashi, H. 108-ghz-bandwidth compact inp-hbt baseband amplifier module with integrated dc-block functions. *IEEE Microwave and Wireless Technology Letters* **33** (2023). https://doi.org/10.1109/LMWT.2023.3243233 .

[3] Renaud, D. *et al.* Sub-1 volt and high-bandwidth visible to near-infrared electro-optic modulators. *Nature Communications* **14** (2023). https://doi.org/10.1038/s41467-023-36870-w .

[4] Nguyen, C. T. *et al.* An integrated nanophotonic quantum register based on silicon-vacancy spins in diamond. *Physical Review B* **100**, 165428 (2019). https://doi.org/10.1103/PhysRevB.100.165428 .

[5] Chan, M. L. *et al.* On-chip spin-photon entanglement based on photon-scattering of a quantum dot. *npj Quantum Information* **9** (2023). https://doi.org/10.1038/s41534-023-00717-5 .

[6] Hensen, B. *et al.* Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526** (2015). https://doi.org/10.1038/nature15759 .

[7] Evans, R. E., Sipahigil, A., Sukachev, D. D., Zibrov, A. S. & Lukin, M. D. Narrow-linewidth homogeneous optical emitters in diamond nanostructures via silicon ion implantation. *Physical Review Applied* **5**, 044010 (2016). https://doi.org/10.1103/PhysRevApplied.5.044010 .

[8] Machielse, B. *et al.* Quantum interference of electromechanically stabilized emitters in nanophotonic devices. *Physical Review X* **9**, 031022 (2019). https://doi.org/10.1103/PhysRevX.9.031022 .

[9] Colangelo, M. *et al.* Molybdenum silicide superconducting nanowire single-photon detectors on lithium niobate waveguides. *ACS Photonics* **11**, 356–361 (2024). URL https://doi.org/10.1021/acsphotonics.3c01628. https://doi.org/10.1021/acsphotonics.3c01628, doi: 10.1021/acsphotonics.3c01628 .

[10] Xin, C. J. *et al.* Spectrally separable photon-pair generation in dispersion engineered thin-film lithium niobate. *Optics Letters* **47** (2022). https://doi.org/10.1364/ol.456873 .

[11] Leent, T. V. *et al.* Long-distance distribution of atom-photon entanglement at telecom wavelength. *Physical Review Letters* **124** (2020). https://doi.org/10.1103/PhysRevLett.124.010510 .

[12] Knaut, C. M. *et al.* Entanglement of nanophotonic quantum memory nodes in a telecommunication network. *Nature* **629**, 573 (2024). https://doi.org/10.1038/s41586-024-07252-z .

[13] Celik, O. T., Ammar, N. Y., Stokowski, H. S., Park, T. & Safavi-Naeini, A. *Bias-stable sub-volt visible electro-optic modulator in thin-film lithium niobate* (2023).

[14] Ruf, M., Wan, N. H., Choi, H., Englund, D. & Hanson, R. Quantum networks based on color centers in diamond. *Journal of Applied Physics* **130** (2021). https://doi.org/10.1063/5.0056534 .

[15] Uppu, R. *et al.* Scalable integrated single-photon source. *Science Advances* **6** (2020). https://doi.org/10.1126/sciadv.abc8268 .

[16] Lukin, D. M., Guidry, M. A. & Vučković, J. Integrated quantum photonics with silicon carbide: Challenges and prospects. *PRX Quantum* **1** (2020). https://doi.org/10.1103/PRXQuantum.1.020102 .

[17] Kindem, J. M. *et al.* Control and single-shot readout of an ion embedded in a nanophotonic cavity. *Nature* **580** (2020). https://doi.org/10.1038/s41586-020-2160-9 .

[18] Krutyanskiy, V. *et al.* Entanglement of trapped-ion qubits separated by 230 meters. *Physical Review Letters* **130** (2023). https://doi.org/10.1103/PhysRevLett.130.050803 .

[19] van Leent, T. *et al.* Entangling single atoms over 33 km telecom fibre. *Nature* **607** (2022). https://doi.org/10.1038/s41586-022-04764-4 .

[20] Sund, P. I. *et al.* High-speed thin-film lithium niobate quantum processor driven by a solid-state quantum emitter. *Science Advances* **9** (2023). https://doi.org/10.1126/sciadv.adg7268 .

[21] Zhai, L. *et al.* Low-noise gaas quantum dots for quantum photonics. *Nature Communications* **11**, 4745 (2020). https://doi.org/10.1038/s41467-020-18625-z .

[22] Holzgrafe, J. *et al.* Relaxation of the electro-optic response in thin-film lithium niobate modulators. *Opt. Express* **32**, 3619–3631 (2024). URL https://opg.optica.org/oe/abstract.cfm?URI=oe-32-3-3619. https://doi.org/10.1364/OE.507536 .

[23] Sosunov, A., Ponomarev, R., Zhuravlev, A., Mushinsky, S. & Kuneva, M. Reduction in dc-drift in linbo3-based electro-optical modulator. *Photonics* **8** (2021). https://doi.org/10.3390/photonics8120571 .

[24] Renaud, D. Visible photonics in thin-film lithium niobate and transition metal dichalcogenides for classical and quantum information applications (2024) .

[25] Bhaskar, M. K. *et al.* Experimental demonstration of memory-enhanced quantum communication. *Nature* **580** (2020). https://doi.org/10.1038/s41586-020-2103-5 .

[26] Humphreys, P. C. *et al.* Deterministic delivery of remote entanglement on a quantum network. *Nature* **558** (2018). https://doi.org/10.1038/s41586-018-0200-5 .

[27] Zhang, C. *et al.* Integrated photonics beyond communications. *Applied Physics Letters* **123**, 230501 (2023). URL https://doi.org/10.1063/5.0184677. https://doi.org/10.1063/5.0184677 .